PROTEINS: Structure, Function, and Genetics Suppl 5:22–38 (2001)
DOI 10.1002/prot.10015

Analysis and Assessment of Comparative Modeling Predictions in CASP4

Anna Tramontano,[1]* Raphael Leplae,[2] and Veronica Morea[3]
[1] *Department of Biochemical Sciences "A. Rossi Fanelli," University of Rome "La Sapienza," Rome, Italy*
[2] *The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, United Kingdom*
[3] *Division of Structural Studies, MRC Laboratory of Molecular Biology, Hills Road, Cambridge, United Kingdom*

ABSTRACT **We describe here the results of our analysis of the comparative modeling predictions submitted to the fourth round of Critical Assessment of Structure Prediction (CASP4). On the basis of a numerical evaluation of the models, we assessed their ability to predict the overall fold correctly, the relative orientation of domains in multidomain proteins, the conformation of the side chains, the loop regions, and the biologically important residues of the targets. We also discuss the performance of automatic prediction servers and compare the results of CASP4 with those obtained in CASP3. Proteins 2001;Suppl 5:22–38. © 2002 Wiley-Liss, Inc.**

Key words: Critical Assessment of Methods of Protein Structure Prediction (CASP); comparative modeling; evaluation of protein models

INTRODUCTION

The assessor for comparative modeling in a CASP experiment is required to evaluate the quality of hundreds of protein models submitted by tens of groups either as three-dimensional (3D) structures or as implicit models (alignment to parents) in a matter of a few weeks.

The task is overwhelming, and it would be impossible if it were not for the excellent support of the scientists at the Livermore Prediction Center, who numerically evaluated each model.[1] The reliability of an assessment based only on lists of numbers such as root-mean-square deviation (RMSD) values or the percentage of correctly aligned residues is debatable, but the only way to go if the experiment is to be as wide, timely, and significant as we believe to be. The number of predictions is such that visual inspection of each model, which in our opinion is still the best form of quality assessment, is infeasible.

The role of the assessor thus becomes that of combining and analyzing the numbers, trying to make sense out of them, and checking the conclusions at various stages by visually inspecting selected models to verify that there are no obvious flaws in the criteria adopted. This does allow a few conclusions to be drawn and forms the basis of what this article aims to describe.

The assessment is thus critical; in fact, the very concept of assessor implies that some choices have to be made in how results are analyzed and presented.

An assessment has to take into account the expectations of the predictors, who invested considerable amount of time and effort in the experiment: they need to know where they stand with respect to their colleagues and fellow predictors and whether any of the novel ideas they tried actually worked. But we believe that an assessment must also meet the expectations of biologists, or other users of the models, who need to know which methods to use and which level of accuracy they can expect from it. Luckily, most of the time these two aspects of the problem coincide, and information can be provided to both predictors and end users, but this is not necessarily the rule. Here we tried to keep in mind the needs of the users foremost, leaving the description of details and technicalities of results to the numerous specialized reports that usually result from a CASP experiment.

RESULTS

Criteria

To compare and evaluate prediction methods, it is necessary to agree on a set of criteria.

In a soccer championship, each game is scored according to the number of goals. Other parameters, such as shots on target, fouls, or elegant play, although equally (or maybe more) correlated to quality of a team are completely discarded in assigning the score. Whether this is a good idea, it is accepted worldwide and seems to work reasonably well.

In comparative modeling, the community believes that no single parameter is sufficient to measure the quality of a model, so it is customary within and without CASP to use a set of parameters and combine them. This introduces ambiguity, because the combination of different parameters is arbitrary and, most importantly, because the parameters are not all independent. For instance, a model with a low RMSD value has certainly been based on a good alignment and has probably produced a better quality of side chains.

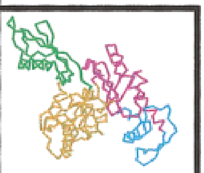
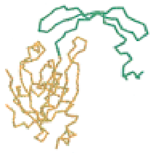
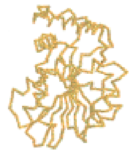
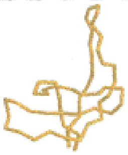
This is one reason, but not the only one, why CASP results cannot possibly be considered as equivalent to the results of a competition.

Grant sponsor: TMR Marie Curie Research Training Grant (Biotechnology Programme).

*Correspondence to: Anna Tramontano, Department of Biochemical Sciences "A. Rossi Fanelli," University of Rome "La Sapienza," 00185 Rome, Italy. E-mail: Anna.Tramontano@uniroma1.it

Received 10 April 2001; Accepted 10 August 2001

TABLE 1. Comparative Modeling Targets for CASP4

Target number Name ***Comments***	Biologically important regions	Domain boundaries	Best PDB template; extent of superposition (No of residues)/Cα RMSD (Å)	
T0089 FtsA from *Thermotoga maritima* ***Interacts with FtsZ to complete septal invagination (cell division)***	14, 16–19, 21, 84, 190, 212–217, 219, 238, 242, 257–261, 337–341, 380	Domain 1: 7–85, 167–200, 359–390	1ats 115/2.08	
		Domain 2: 239–290	1dga 48/2.53	
		Domain 3: 86–166	Not CM target	
		Domain 4: 201–238, 291–358	1dej 93/1.96	
T0090 ADP-ribose pyrophosphatase from *E. coli*	55–57, 58–59, 94, 96–97, 99, 100, 112, 115, 116, 130, 132, 133, 137, 139–141, 166, 187–189	Domain 1: 1–57	Not CM target	
		Domain 2: 58–209	1mut.M_15 100/2.49	
T0092 Hypothetical protein from *E. coli* ***Shares homology with one domain of Glycine N-methyltransferase***	The active site residues of the template are not conserved in the target.	Single domain: 1–227	1d2c 175/2.38	
T0099 SH3 domain	6, 7, 9–16, 30–33, 46, 48, 50	Single domain: 1–56	1lck 47/2.9	
T0103 Extracellular alkaline serine proteinase	34, 80, 107–111, 133–139, 193, 195, 198, 287	Single domain: 1–368	1ak9 237/1.96	

To assess which method works better in CASP, one should know which models have been produced by which method.

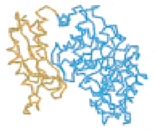
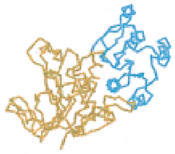
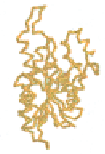
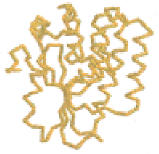
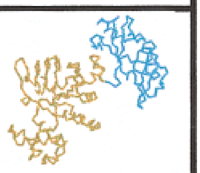
There are at least two problems to be to faced. The first is that assessment is essentially blind, that is, only at the very last minute after evaluating and scoring models is the assessor informed about the identity of the predicting group and can then see if the group has chosen to give information about the type of method used. This implies that a model is evaluated in the category of its target, not according to the method used to produce it. It is obvious that even an extraordinarily good ab initio prediction on a comparative modeling target has very few chances to be better than the comparative modeling predictions on the same target.

This has another effect: a comparative modeling group has to decide which targets are within their category on the basis of sequence data only. Protein structure databases tend to grow during the prediction season, and it is possible that database searches performed by different groups at different points in time produce different results. In any case, a group may choose to predict all the targets, only some of them, or might fail to recognize that a target can be modeled by homology. Furthermore, partial predictions can be submitted, and indeed some models only cover part(s) of the target.

And here is another difference from a soccer competition: each team has to play against all the designated teams for an allotted time; it cannot just skip or shorten matches at will. If this were possible, assigning the score in a championship would be virtually impossible.

In CASP, predictors are not required to submit models for all the targets, which implies that each model must be scored according to the difficulty of the target and to the fraction of the target predicted. Thus, both the total score and the average score achieved by each group should be considered.

TABLE 1. (Continued)

T0111 Enolase from *E. coli* ***Catalyzes the reversible dehydration of 2-phospho-D-glycerate to phosphoenolpyruvate***	38–45, 156–165, 167, 208, 245, 257–267, 289, 316, 341, 370, 371	Domain 1: 1–128 Domain 2: 129–430	1ebh 126/1.96 6enl 290/0.99	
T0112 Sorbitol dehydrogenase from silverleaf whitefly ***Reaction: L-iditol + NAD(+) = L-sorbose + NADH (related to Zn ADH)***	41, 66, 92, 96, 99, 102, 110, 152, 153, 295	Domain 1: A4–A165, A294–A351 Domain 2: A166–A293	1bxz 200/2.07 1agn 121/1.75	
T0113 HCD2 from rat ***Short chain 3-hydroxyl AcilCoA dehydrogenase***	15–20, 40–44, 64–66, 69, 91–94, 97–98, 119–120, 124, 127, 155–157, 161, 164–165, 168, 172, 200, 205, 257	Single domain: 1–255	1ahi 235/1.45	
T0117	30–36, 51, 84, 88, 104, 105, 110	Single domain: 1–197	1vtk 174/2.2	
T0121 MalK from *T. litoralis* ***ATP-binding subunit of the maltose ABC transporter***	13, 36–44, 84, 86, 88, 135, 140–148, 158, 164, 165, 173, 177, 178, 192, 193, 198, 199	Domain 1: A1–A240 Domain 2: A241–A372	1b0u 209/2.51 1b9n 114/1.99	
T0122 Tryptophan synthase alpha subunit from *Pyrococcus furiosus* ***Cleaves indole 3-glycerol phosphate to give indole + D-glyceradehyde-3-P***	10, 36, 47, 51, 86, 88, 113, 139, 161, 198, 220–222	Single domain: 1–241	1c29 233/1.59	

In conclusion, no matter how careful an assessment is, there are intrinsic reasons why results have to be treated with caution, and, although summaries such as this one can prove useful, all the above should be kept in mind and, whenever needed, reference made to the data publicly available on the CASP server (http://PredictionCenter.llnl.gov/casp4).

Targets

Any sequence showing a significant E-value (<0.02) with a protein of known structure (http://PredictionCenter.llnl.gov/casp4) after a PSI-BLAST run[2] was considered a target for comparative modeling.

The selected targets are listed in Table I. We also show in the same table:

- the location of residues known to be important for the protein's function
- the domain boundaries for multi-domain proteins
- the closest structure present in the database
- the RMSD and the extent of the structural superposition[1] between the target and the best parent.

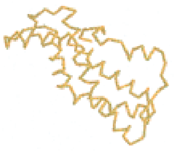
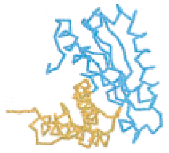
The location of residues important for the biological function of the target protein, which are probably those of interest for the end users, were identified by checking the literature data on the target (if available) or on the corresponding parent(s).

Domain boundaries were obtained by visual inspection of the target structures.

A structure-based PDB search was performed by the Livermore Prediction Center scientists,[1] and the closest parent structure was identified by using the LGA procedure.[3] Being based on knowledge of the target structure, this information was not available at the time of prediction.

From the final user's point of view, the important factor is how good the model is, not how good the result would have been if the selected parent were the only one available. This means that a model should be compared with

TABLE 1. (Continued)

T0123 Beta lactoglobulin from pig *Function unknown, probably involved in the transport of retinol and/or fatty acids*	29, 33, 34, 37–38, 40, 56, 61, 69, 84, 92, 120, 145–152, 155	Single domain: 1–260	1beb 136/2.16	
T0125 Sp18 protein, *Haliotis fulgens* ***Dimer***	1–7, 8, 11, 15, 24, 28, 31, 37, 53, 55, 59, 68, 69, 74, 85, 92, 93, 107, 100, 101, 120, 116, 123, 126, 131, 135	Single domain: 1–137	3lyn 107/2.13	
T0128 Manganese superoxide dismutase homolog from *Pyrobaculum aerophilum* ***MnSOD dismute toxic superoxide radicals to oxygen and oxygen peroxide***	37, 41, 42, 45–50, 78–79, 81–82, 90, 137–139, 143, 160, 166, 176, 178–180, 183–184, 187–188, 190	Domain 1: A12–A99	1b06 88/0.7	
		Domain 2: A100–A222	1b06 117/0.84	

Domains 1, 2, and 4 are colored orange, blue, and magenta, respectively. Domains that are not comparative modeling targets are shown in green.

the best possible parent and not to the one actually used to produce it.

Incidentally, with the continuous increase in database size and the implementation of methods that use more than one parent per target, assessing each model in relation to the parent used would be impossible anyway.

A rough measure of how well groups selected the parent is given in Table II. There we list the RMSD between the Cαs of the target and those of the chosen parent(s) in the superimposed region. We only list values for those predictions where the selected template corresponds to one of the templates listed by the Livermore Prediction Center Web site.

Figure 1 shows a plot of the RMSD for Cα atoms obtained by each group on multidomain targets after optimal superposition of the predicted and experimental structures compared with that obtained by superimposing each domain separately.

Almost invariably, the prediction of the complete structure has a higher RMSD value than the individual domains. This observation, not unexpected, points out once more the difficulty of predicting the relative position of domains even in related proteins. We always analyzed the predictions according to the superposition of domains rather than of the complete structure; if we did otherwise, a high RMSD for a poorly predicted domain could be masked by a low RMSD for a larger one.

Scoring Scheme

The aim of our scoring scheme was to evaluate the models of each target by using a number of different measures, each normalized by the distribution for that measure over predictions for that target.

Predictors can submit more than one model, but as announced beforehand (http://PredictionCenter.llnl.gov/casp4), we only analyzed the one designated as model 1. The first analysis we performed aimed at evaluating the quality of the overall folds. The selected measures were GDT-TS (*GDT*),[1] RMSD for all Cα atoms of the core (*rms*), the percentage of correctly aligned residues (*al*0), the Cα RMSD of biologically important regions (*rmsb*), and the Cα RMSD for those regions where the target differs substantially from its parent, in the following simply called loops (*rmsl*).[1]

GDT-TS is defined as:

$$(\%C\alpha \text{ within } 1\text{ Å} + \%C\alpha \text{ within } 2\text{ Å} + \%C\alpha \text{ within } 4\text{ Å} + \%C\alpha \text{ within } 8\text{ Å})/4 \quad (1)$$

where % Cα within 1 Å is the percent of aligned residues within 1 Å after superposition of model and target. It should be noted that, contrary to RMSD values, GDT-TS does not explicitly penalize models where one or more regions are predicted very incorrectly, although, as we will see, our scoring method does if other groups have produced better models for those regions.

To analyze the details of the models, we used the RMSD for all side-chain atoms (*rmssc*), all side-chain atoms defined as reliable (*rmsscr*), all side chains in the core (*rmsscc*), and all side chains in structurally divergent regions (*rmsscl*). Definitions of these parameters can be found in Ref. 1.

In all cases we used the following simple rule to assign a score to a model. Let X be the selected parameter

$$(X \in \{GDT, rms, al0, rmsb, rmsl, rmssc, rmsscr, rmsscc, rmsscl\}), \quad (2)$$

TABLE II. Selection of Templates

Id	Group	T0089	T0090	T0092	T0099	T0103	T0111	T0112	T0113	T0117	T0121	T0122	T0123	T0125	T0128
001	Shortle	—	—	—	—	—	—	—	—	—	—	—	—	—	—
002	Lomize-Andrei	—	—	—	—	—	—	—	—	—	—	—	—	2.13	—
012	Levitt	2.28	2.64	—	—	—	—	—	—	2.41	2.49	—	2.06	2.38	—
017	Yang-Ansuei	2.41	2.64	2.28	—	1.98	—	—	—	—	2.49	1.63	2.16	—	—
018	Raghava-GPS	—	—	—	—	—	—	—	—	—	—	—	—	—	—
022	InforMax	—	—	—	—	—	—	—	—	—	—	—	2.16	—	0.78
023	Jones	2.41	2.64	2.16	—	—	—	—	—	—	2.49	—	2.16	2.19	—
028	Ram-Samudrala	—	2.61	—	2.68	—	—	—	—	—	—	—	—	2.38	—
031	BioInfo.PL	—	2.61	2.16	—	—	—	—	—	2.23	2.49	—	2.13	2.13	—
032	Wolynes	—	2.64	2.16	2.68	2.01	1.05	—	—	2.31	—	1.6	2.13	2.38	—
035	Rose-Group	—	—	—	—	—	—	—	—	—	—	—	—	—	—
042	Honig-Barry	—	2.64	2.16	2.74	—	1.23	—	—	—	2.49	—	—	2.19	—
044	Walts-Wondrous	—	—	—	—	—	—	—	—	—	—	—	2.16	—	—
047	kitasato-univ.	—	2.61	—	2.74	—	—	—	—	2.2	2.49	1.63	2.16	—	—
055	Bystroff	—	—	—	—	—	—	—	—	—	—	—	—	—	—
058	Harrison-Weber	—	2.61–2.64	2.16	2.9	—	—	2.19	1.68	2.31	2.49	—	—	2.19	—
065	Torda-Andrew	—	—	—	2.55	—	—	—	1.39	—	2.49	—	2.16	—	—
077	rost	—	—	—	2.48–2.74	—	—	—	1.46	—	2.49	1.59–1.63	2.17	—	0.97
080	Skolnick-Kolinski	—	—	—	—	—	—	—	—	—	—	—	—	—	—
086	Bass-Michael	—	—	—	2.67	—	—	—	—	—	—	1.57	2.16	—	—
088	ORNL-PROSPECT	—	2.61	—	2.48	—	1.23	—	—	—	2.49	—	2.13	2.13	0.79
090	Hogue-Feldman	—	—	—	2.74	—	—	—	—	—	—	—	—	—	—
094	SAM-T2K	—	—	—	2.65	—	—	—	—	—	—	—	—	—	—
095	blundell-tl	—	2.61	2.16	2.74	—	—	—	—	2.6	2.49	—	2.17	2.19	—
118	Dlakic-Mensur	—	—	—	2.51	—	—	—	—	—	—	—	—	—	—
125	Sternberg-3D-JIGSAW	—	2.64	—	2.65–2.74	—	—	2.25–2.44	1.54–1.61	2.24	2.49	1.63	2.11–2.13	2.13–2.19	—
126	Sternberg	2.41	2.64	2.16	2.65–2.74	1.93	—	2.25–2.44	1.54–1.61	2.24	2.49	1.57–1.63	2.11–2.13	2.13–2.19	—
133	CBC-FOLD	—	2.61	—	5.36	—	—	—	—	—	—	—	—	—	—
152	Yoon	—	—	—	—	—	—	—	—	—	—	—	—	—	—
155	TUDELFT	—	—	—	5.48	—	—	—	—	—	—	—	4.34	—	—
161	GNM-FR	—	—	—	2.74	—	—	—	—	—	—	—	—	—	—
169	Dunbrack	—	2.64	2.16	2.74	—	—	2.3	—	—	2.49	1.63	2.16	2.19	—
179	Sausage	—	—	—	—	—	—	—	—	—	—	—	2.16	—	—
186	SDSC1	—	—	—	—	1.94	—	—	—	—	2.49	—	2.16	—	—
187	SDSC2:Reddy-Bourne	—	—	—	2.68–2.9	—	—	—	1.39	—	—	—	—	—	—
191	Lee-Jung	—	2.61	—	2.55	—	—	2.25	—	—	2.49	—	—	2.13	—
197	Godzik	2.41	2.64	—	2.65	—	—	—	—	2.6	2.49	1.6	—	2.19	0.84
216	Isites-Server	—	—	—	—	—	—	—	—	—	—	—	—	—	—
218	LAMBERT-Christophe	—	2.61	—	—	—	—	2.44	—	—	2.49	1.63	—	—	—
223	Braun-UTMB	2.41	2.64	2.16	2.74	—	—	—	—	—	2.49	—	2.16	2.38	—
237	Sali-Andrej	—	2.61	2.16	—	—	1.23	—	—	—	—	—	—	—	—
241	Vajda	2.41	2.64	—	2.74	—	—	—	—	—	2.49	1.6	2.17	2.38	—
243	Dill-Ken	—	—	—	—	—	—	—	—	—	—	—	—	—	—
255	BinToHes	—	2.64	—	—	—	1.13	—	—	—	2.49	—	2.16	—	—
278	Flake&mates	—	—	—	2.71	—	—	—	—	—	—	—	—	—	—
279	Bateman	—	2.61	—	—	—	—	2.44	—	—	—	—	—	—	—
281	Mohan	—	—	—	—	—	—	—	—	—	—	—	—	—	—
312	HANRAM-CDFD	—	2.64	—	—	—	—	—	—	—	—	—	—	—	—
330	Zemla-Joanna	—	—	—	2.74	—	—	—	1.39	—	—	1.63	2.11	2.19	—
341	Lai	—	—	—	2.55	—	—	—	—	—	—	—	—	—	—
342	SBI-AT	2.41	—	—	—	1.97	—	—	—	—	2.49	1.63	2.11	2.19	—
352	zhu	—	—	—	—	—	—	—	—	—	—	—	—	—	—
354	baker	—	2.61	—	—	—	—	—	—	—	—	1.57	—	2.13	—
359	Cafasp-consensus	—	2.64	2.16	2.65	—	—	—	—	—	—	—	—	—	—
363	Moult	—	2.64	—	—	—	—	—	—	—	—	—	2.09	—	—
375	Ho-Kai-Ming	—	—	—	—	—	—	—	—	—	—	—	—	—	—
381	SBfold	—	5.22	—	3.17	—	1.23	—	—	—	2.49	—	2.16	2.33	—
382	SBauto	—	2.61	—	3.17	2	1.23	2.25	—	—	2.49	1.6	—	—	1.23
383	HeadGordon-Teresa	—	—	—	—	—	—	—	—	—	—	—	—	—	—
384	Murzin	—	2.61	—	—	—	—	—	—	—	—	—	—	—	—
390	Taylor	—	2.64	—	—	—	—	—	—	—	—	—	—	—	—
393	Skolnick-Kolinski-THD	—	—	—	2.55	—	—	—	—	—	—	—	—	—	—
406	VENCLOVAS	2.28–2.41	2.64	2.16	2.3–2.51	—	1.23	2.3–2.44	1.53	—	2.49	1.63	2.11–2.16	—	—
414	Friesner	—	2.64	2.44	2.51	1.94	1.23	—	—	—	2.49	1.56	2.11	—	—
426	koehl	—	2.64	2.16	2.3	—	—	—	—	—	—	—	—	—	—
429	CHEN-WENDY	—	—	2.16	2.55	—	—	—	1.39	2.31–2.6	2.49	—	2.16	2.13–2.38	—
432	LMGDD	—	—	—	—	—	1.23	—	—	—	—	—	—	—	—
440	Deleage-Geourjon	—	—	—	—	1.94	—	—	—	—	2.49	—	—	—	—
444	MOE-CCG	—	—	—	—	—	—	—	—	—	—	—	—	—	—
447	MSI	—	—	—	2.74	—	—	—	—	—	2.49	1.63	2.11–2.17	2.13	—

TABLE II. (Continued)

Id	Group	T0089	T0090	T0092	T0099	T0103	T0111	T0112	T0113	T0117	T0121	T0122	T0123	T0125	T0128
455	NIH-Garnier	—	—	—	—	—	—	—	—	—	—	—	—	—	—
457	SBI-GR	—	—	—	2.74	—	—	—	—	2.6	2.49	—	—	2.13	—
459	mprabha	—	—	—	—	—	—	—	—	—	—	—	—	—	—
465	YASARA	—	—	—	—	—	—	—	—	—	—	—	—	—	—
468	SBI-jz	—	—	—	2.51	—	—	—	—	—	—	1.63	2.12	—	—
471	Chodera-John	—	—	—	—	—	—	—	—	—	—	—	—	—	—
473	Mushegian	—	—	—	—	—	—	—	—	—	—	—	—	—	—
486	Shoshana-Wodak	—	—	—	—	—	—	—	—	—	—	1.63	2.16	2.19	—
489	FCLD	—	—	—	—	—	—	—	—	—	—	—	—	—	—
498	Kollman-Baker	—	—	—	—	—	—	—	—	—	—	—	—	—	—
500	FAMS	—	2.64	—	2.71	—	—	—	—	—	2.49	1.59	—	—	—
512	ELAN	—	—	—	—	—	—	—	—	—	—	—	—	—	—
526	Ginalski	—	—	—	—	—	—	—	—	—	—	1.63	2.17	2.19	—
535	shankari	—	—	—	—	—	—	—	—	—	—	—	—	—	—

$C\alpha$ RMSD ranges for the target and the selected parent(s) in the superimposable regions are shown in Å only for predictions using one or more templates found by the ProSup procedure.

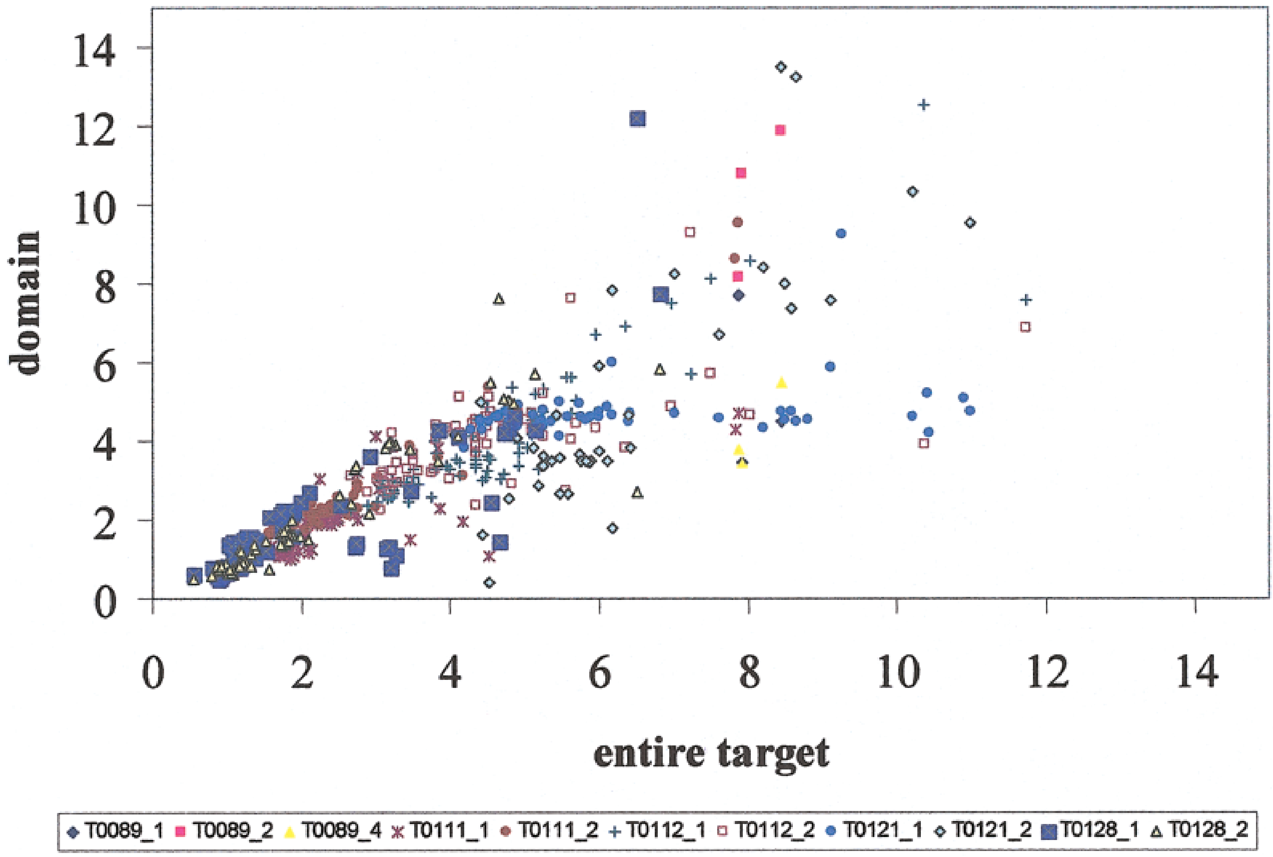


Fig. 1. Scatter plot of the $C\alpha$ RMSD between target and predicted structure (x axis) versus that of the corresponding domains (y axis). Only values below 15 Å are shown. The number following the target name indicates the domain.

$\bar{X}$ and $\sigma(X)$ the average and standard deviation of X over all predictions for a given target. We first excluded predictions with very "bad" values of X:

$$\text{If } X \in \{GDT, al0\}, \text{ eliminate all } X < \bar{X} - 2 * \sigma(X)$$

$$\text{If } X \in \{rms, rmsb, rmsl, rmssc, rmsscr, rmsscc, rmsscl\}, \text{ eliminate all } X > \bar{X} + 2 * \sigma(X)$$

We should mention here that the number of excluded predictions was very low (around 5%).

We then recalculated $\bar{X}$ and $\sigma(X)$ over all remaining predictions and assigned the score as:

$$\text{Score}(X) = \frac{X - \bar{X}}{0.5 * \sigma(X)} * \%\ \text{predicted}$$

Where % predicted is the percent of the structure (or of the biologically important region or of the loop) that is present in the prediction.

TABLE III. Group Scores for Correctness of the Overall Fold

Id	Group	N dom	GDT-TS tot	GDT-TS av	RMS CaC tot	RMS CaC av	al0 tot	al0 av	Biol tot	Biol av	Loop tot	Loop av	Sum tot	Sum av	Rank tot	Rank av	Sum of ranks
406	VENCLOVAS	18	26.23	1.46	21.70	1.21	23.46	1.30	11.74	0.65	19.27	1.07	102.40	5.69	1	5	6
354	baker	19	34.03	1.79	17.20	0.91	22.68	1.19	12.51	0.66	15.87	0.84	102.29	5.38	2	6	8
126	Sternberg	20	28.67	1.43	18.05	0.90	21.53	1.08	13.15	0.66	16.46	0.82	97.86	4.89	3	12	15
031	BioInfo.PL	19	31.72	1.67	16.26	0.86	21.91	1.15	11.34	0.60	11.52	0.61	92.76	4.88	4	13	17
342	SBI-AT	20	22.59	1.13	18.72	0.94	21.85	1.09	12.12	0.61	11.95	0.60	87.24	4.36	5	15	20
197	Godzik	20	18.92	0.95	17.98	0.90	17.56	0.88	10.57	0.53	19.54	0.98	84.57	4.23	6	16	22
088	ORNL-PROSPECT	20	32.37	1.62	14.30	0.71	14.59	0.73	10.21	0.51	12.13	0.61	83.60	4.18	7	17	24
042	Honig-Barry	20	23.23	1.16	17.83	0.89	15.89	0.79	10.48	0.52	15.43	0.77	82.86	4.14	8	18	26
260 S	mGen THREADER	20	20.15	1.01	16.59	0.83	19.33	0.97	12.76	0.64	10.35	0.52	79.19	3.96	9	21	30
259 S	Gen THREADER	20	15.33	0.77	16.05	0.80	18.81	0.94	13.32	0.67	11.91	0.60	75.43	3.77	10	26	36
169	Dunbrack	17	18.05	1.06	13.25	0.78	17.33	1.02	8.43	0.50	10.87	0.64	67.93	4.00	19	19	38
044	Walts-Wondrous	20	17.03	0.85	18.87	0.94	16.17	0.81	9.09	0.45	12.36	0.62	73.53	3.68	11	28	39
047	kitasato-univ.	18	18.63	1.03	13.43	0.75	12.29	0.68	11.84	0.66	13.35	0.74	69.53	3.86	16	23	39
223	Braun-UTMB	19	20.90	1.10	13.30	0.70	15.40	0.81	9.78	0.51	12.04	0.63	71.43	3.76	14	27	41
382	SBauto	20	23.01	1.15	12.75	0.64	15.37	0.77	7.11	0.36	14.97	0.75	73.20	3.66	12	29	41
111 S	SAM-T99	20	18.51	0.93	18.21	0.91	15.47	0.77	10.52	0.53	10.36	0.52	73.06	3.65	13	30	43
237	Sali-Andrej	12	15.81	1.32	12.31	1.03	10.23	0.85	8.99	0.75	11.16	0.93	58.49	4.87	30	14	44
107 S	bioinbgu-seqpmprf	20	18.11	0.91	15.47	0.77	15.43	0.77	9.34	0.47	11.63	0.58	69.98	3.50	15	32	47
170	DNAmining.com/p-map	16	12.14	0.76	15.54	0.97	12.73	0.80	7.44	0.47	14.43	0.90	62.29	3.89	26	22	48
132 S	Sternberg-3DPSSM	19	16.84	0.89	11.15	0.59	14.83	0.78	13.63	0.72	11.99	0.63	68.44	3.60	18	31	49
103 S	Zhou-HX	20	16.92	0.85	14.67	0.73	14.42	0.72	10.54	0.53	12.69	0.63	69.24	3.46	17	33	50
384	Murzin	4	15.84	3.96	7.62	1.91	4.02	1.00	2.68	0.67	6.17	1.54	36.32	9.08	53	1	54
093 S	bioinbgu	20	19.07	0.95	14.33	0.72	14.00	0.70	9.44	0.47	10.49	0.52	67.32	3.37	20	37	57
357	Fischer-Daniel	20	17.98	0.90	14.15	0.71	13.92	0.70	12.28	0.61	8.92	0.45	67.24	3.36	21	39	60
094	SAM-T2K	20	21.80	1.09	14.03	0.70	11.32	0.57	9.12	0.46	10.26	0.51	66.53	3.33	22	41	63
359	Cafasp-consensus	20	19.96	1.00	14.40	0.72	13.59	0.68	9.79	0.49	8.61	0.43	66.36	3.32	23	42	65
106 S	bioinbgu-seqpprf	20	17.25	0.86	14.48	0.72	14.86	0.74	8.77	0.44	10.73	0.54	66.09	3.30	24	43	67
218	LAMBERT-Christophe	17	15.23	0.90	8.91	0.52	15.71	0.92	8.51	0.50	9.58	0.56	57.94	3.41	31	36	67
158 S	PDB-Blast	15	10.09	0.67	9.38	0.63	11.06	0.74	10.60	0.71	10.64	0.71	51.77	3.45	36	34	70
137	Zhou-HX	19	13.45	0.71	14.79	0.78	12.69	0.67	9.33	0.49	11.34	0.60	61.60	3.24	27	44	71
526	Ginalski	4	6.77	1.69	5.13	1.28	5.53	1.38	4.17	1.04	2.52	0.63	24.13	6.03	68	3	71
077	rost	20	13.77	0.69	15.14	0.76	11.77	0.59	11.14	0.56	10.75	0.54	62.57	3.13	25	48	73
405	josé	6	9.44	1.57	5.70	0.95	3.73	0.62	3.15	0.53	7.66	1.28	29.69	4.95	62	11	73
500 S	FAMS	16	15.16	0.95	9.95	0.62	9.66	0.60	7.49	0.47	11.24	0.70	53.50	3.34	33	40	73
023	Jones	19	18.06	0.95	13.73	0.72	12.46	0.66	8.06	0.42	8.00	0.42	60.32	3.17	28	46	74
331	Levy	17	13.00	0.76	11.14	0.66	8.82	0.52	10.36	0.61	10.46	0.62	53.79	3.16	32	47	79
381	SBfold	20	20.44	1.02	9.92	0.50	13.82	0.69	9.19	0.46	5.57	0.28	58.94	2.95	29	51	80
426	koehl	4	5.42	1.36	3.53	0.88	6.27	1.57	2.00	0.50	3.23	0.81	20.46	5.11	77	10	87
429	CHEN-WENDY	16	13.73	0.86	10.67	0.67	7.28	0.45	9.39	0.59	8.52	0.53	49.60	3.10	38	50	88
032	Wolynes	19	13.90	0.73	10.41	0.55	9.37	0.49	7.76	0.41	11.18	0.59	52.62	2.77	35	56	91
279	Bateman	3	4.57	1.52	3.75	1.25	1.70	0.57	2.42	0.81	3.42	1.14	15.87	5.29	82	9	91
498	Kollman-Baker	1	1.67	1.67	1.16	1.16	1.70	1.70	1.82	1.82	1.19	1.19	7.54	7.54	90	2	92
447	MSI	13	12.41	0.95	8.46	0.65	9.15	0.70	6.54	0.50	5.30	0.41	41.85	3.22	49	45	94
108 S	bioinbgu-prfseq	20	15.05	0.75	11.11	0.56	9.50	0.48	9.34	0.47	8.11	0.41	53.12	2.66	34	61	95
465	YASARA	6	6.03	1.01	5.86	0.98	5.47	0.91	2.34	0.39	2.96	0.49	22.67	3.78	70	25	95
002	Lomize-Andrei	1	1.96	1.96	1.35	1.35	1.13	1.13	0.00	0.00	1.41	1.41	5.86	5.86	92	4	96
095	blundell-tl	14	12.57	0.90	8.12	0.58	6.84	0.49	7.48	0.53	8.55	0.61	43.56	3.11	47	49	96
486	Shoshana-Wodak	9	9.44	1.05	7.21	0.80	6.25	0.69	2.54	0.28	4.82	0.54	30.26	3.36	60	38	98
393	Skolnick-Kolinski-THD	19	12.61	0.66	9.61	0.51	8.03	0.42	5.47	0.29	14.20	0.75	49.92	2.63	37	63	100
118	Dlakic-Mensur	1	1.30	1.30	0.82	0.82	1.83	1.83	1.12	1.12	0.30	0.30	5.37	5.37	94	7	101
312	HANRAM-CDFD	1	1.47	1.47	1.47	1.47	0.81	0.81	0.84	0.84	0.69	0.69	5.29	5.29	95	8	103
012	Levitt	19	12.57	0.66	10.24	0.54	11.08	0.58	6.48	0.34	8.00	0.42	48.38	2.55	40	65	105
150	Chandonia-Cohen	13	9.13	0.70	8.49	0.65	6.43	0.49	6.42	0.49	6.72	0.52	37.20	2.86	51	54	105
017	Yang-Ansuei	20	12.33	0.62	11.69	0.58	8.39	0.42	8.52	0.43	8.24	0.41	49.18	2.46	39	70	109
173	Barton	15	7.93	0.53	9.35	0.62	5.82	0.39	8.64	0.58	8.27	0.55	40.00	2.67	50	59	109
104 S	bioinbgu-gonp	20	13.81	0.69	11.69	0.58	8.19	0.41	7.55	0.38	6.83	0.34	48.06	2.40	41	71	112
133	CBC-FOLD	19	9.47	0.50	11.91	0.63	8.57	0.45	10.02	0.53	6.90	0.36	46.86	2.47	43	69	112
363	Moult	18	9.92	0.55	12.40	0.69	8.65	0.48	6.38	0.35	7.35	0.41	44.70	2.48	45	67	112
028	Ram-Samudrala	20	16.57	0.83	9.05	0.45	7.16	0.36	6.94	0.35	8.04	0.40	47.76	2.39	42	72	114
395 S	FFAS	17	10.70	0.63	8.17	0.48	9.66	0.57	4.79	0.28	8.90	0.52	42.23	2.48	48	66	114
105 S	bioinbgu-gonpm	20	11.27	0.56	11.17	0.56	5.51	0.28	8.27	0.41	8.97	0.45	45.19	2.26	44	75	119
241	Vajda	19	12.24	0.64	9.52	0.50	7.45	0.39	5.53	0.29	9.10	0.48	43.84	2.31	46	73	119
341	Lai	1	1.02	1.02	0.80	0.80	0.60	0.60	0.00	0.00	1.56	1.56	3.98	3.98	100	20	120
389 S	123D+	11	6.63	0.60	6.86	0.62	6.13	0.56	4.85	0.44	4.79	0.44	29.26	2.66	63	60	123
344	PDB-ISL	1	0.75	0.75	1.27	1.27	1.38	1.38	0.09	0.09	0.31	0.31	3.80	3.80	101	24	125
330	Zemla-Joanna	10	6.13	0.61	7.62	0.76	2.30	0.23	4.57	0.46	5.86	0.59	26.49	2.65	67	62	129
361	GMD-SCAI	18	7.00	0.39	9.29	0.52	6.39	0.36	9.20	0.51	5.25	0.29	37.13	2.06	52	77	129

TABLE III. (Continued)

Id	Group	N dom	GDT-TS tot	GDT-TS av	RMS CaC tot	RMS CaC av	al0 tot	al0 av	Biol tot	Biol av	Loop tot	Loop av	Sum tot	Sum av	Rank tot	Rank av	Sum of ranks
453	Noguchi	8	5.77	0.72	5.10	0.64	2.30	0.29	2.46	0.31	6.36	0.80	21.99	2.75	73	57	130
444	MOE-CCG	7	3.58	0.51	3.99	0.57	5.44	0.78	3.38	0.48	3.86	0.55	20.24	2.89	78	53	131
457	SBI-GR	18	10.10	0.56	7.85	0.44	5.42	0.30	3.73	0.21	7.89	0.44	34.98	1.94	54	78	132
155	TUDELFT	8	5.80	0.72	5.42	0.68	5.02	0.63	2.80	0.35	2.33	0.29	21.37	2.67	75	58	133
468	SBI-jz	9	6.55	0.73	4.57	0.51	5.88	0.65	3.20	0.36	3.11	0.35	23.32	2.59	69	64	133
390	Taylor	18	8.84	0.49	9.63	0.53	3.11	0.17	4.66	0.26	7.29	0.41	33.53	1.86	56	79	135
535	shankari	5	4.65	0.93	2.88	0.58	1.99	0.40	2.76	0.55	1.65	0.33	13.93	2.79	84	55	139
191	Lee-Jung	20	8.81	0.44	8.09	0.40	4.53	0.23	6.30	0.31	6.40	0.32	34.14	1.71	55	85	140
458	strauss	1	0.54	0.54	0.90	0.90	1.20	1.20			0.80	0.80	3.43	3.43	105	35	140
058	Harrison-Weber	17	4.38	0.26	7.78	0.46	6.12	0.36	4.50	0.26	7.54	0.44	30.32	1.78	59	82	141
065	Torda-Andrew	19	8.47	0.45	6.77	0.36	4.89	0.26	6.81	0.36	5.69	0.30	32.63	1.72	57	84	141
090	Hogue-Feldman	2	1.11	0.56	0.77	0.38	0.11	0.06	1.00	0.50	2.85	1.42	5.84	2.92	93	52	145
414	Friesner	19	5.75	0.30	5.04	0.27	6.58	0.35	7.50	0.39	6.30	0.33	31.16	1.64	58	87	145
086	Bass-Michael	19	7.90	0.42	7.30	0.38	2.54	0.13	4.85	0.26	7.59	0.40	30.19	1.59	61	88	149
125 S	Sternberg-3D-JIGSAW	18	4.84	0.27	5.34	0.30	2.42	0.13	6.84	0.38	7.34	0.41	26.79	1.49	65	90	155
186 S	SDSC1	17	6.41	0.38	7.02	0.41	5.53	0.33	4.30	0.25	3.38	0.20	26.64	1.57	66	89	155
536	Fox-Sheppard	8	4.57	0.57	3.83	0.48	3.15	0.39	3.85	0.48	1.23	0.15	16.63	2.08	81	76	157
255	BinToHes	20	8.59	0.43	4.97	0.25	2.89	0.14	5.72	0.29	5.21	0.26	27.38	1.37	64	95	159
022	InforMax	4	1.39	0.35	2.40	0.60	1.83	0.46	2.68	0.67	0.85	0.21	9.16	2.29	88	74	162
127 S	ssPsi/Elofsson-Arne	16	4.65	0.29	3.67	0.23	5.64	0.35	4.50	0.28	3.81	0.24	22.26	1.39	72	93	165
161	GNM-FR	17	7.02	0.41	5.15	0.30	4.45	0.26	1.90	0.11	3.75	0.22	22.27	1.31	71	97	168
375	Ho-Kai-Ming	4	2.45	0.61	2.40	0.60	0.75	0.19	0.42	0.11	1.30	0.33	7.32	1.83	91	80	171
440	Deleage-Geourjon	6	2.38	0.40	1.78	0.30	2.76	0.46	2.28	0.38	0.75	0.13	9.95	1.66	86	86	172
278	Flake&mates	17	4.29	0.25	5.35	0.31	0.58	0.03	4.64	0.27	6.55	0.39	21.41	1.26	74	99	173
473	Mushegian	1	0.55	0.55	0.00	0.00	0.00	0.00	1.44	1.44	0.48	0.48	2.47	2.47	106	68	174
229	UCLA-DOE	19	3.18	0.17	6.60	0.35	2.28	0.12	2.30	0.12	6.94	0.37	21.29	1.12	76	101	177
080	Skolnick-Kolinski	7	2.12	0.30	4.67	0.67	1.46	0.21	0.87	0.12	0.82	0.12	9.93	1.42	87	92	179
187	SDSC2:Reddy-Bourne	3	0.00	0.00	0.82	0.27	0.00	0.00	1.79	0.60	2.55	0.85	5.16	1.72	96	83	179
152	Yoon	6	2.51	0.42	1.92	0.32	1.28	0.21	2.97	0.49	0.00	0.00	8.67	1.44	89	91	180
010	Pan	9	2.23	0.25	2.69	0.30	2.89	0.32	1.65	0.18	2.55	0.28	12.00	1.33	85	96	181
052	MRIT-Onizuka	18	5.66	0.31	4.73	0.26	2.11	0.12	1.42	0.08	6.24	0.35	20.16	1.12	79	102	181
001	Shortle	2	1.97	0.99	0.00	0.00	1.63	0.81			0.00	0.00	3.60	1.80	103	81	184
274	Tsigelny	18	4.42	0.25	4.32	0.24	0.72	0.04	5.47	0.30	4.68	0.26	19.61	1.09	80	104	184
401	Reva-Boris	15	1.85	0.12	2.31	0.15	2.63	0.18	4.64	0.31	4.31	0.29	15.74	1.05	83	105	188
027	SHESTOPALOV	3	0.00	0.00	1.96	0.65	0.00	0.00	1.65	0.55	0.53	0.18	4.14	1.38	99	94	193
179 S	Sausage	4	1.86	0.46	0.85	0.21	0.66	0.17	1.06	0.27	0.00	0.00	4.43	1.11	97	103	200
352	zhu	5	1.35	0.27	0.00	0.00	1.70	0.34	1.37	0.27	0.00	0.00	4.42	0.88	98	106	204
035	Rose-Group	1	1.30	1.30	0.00	0.00	0.00	0.00			0.00	0.00	1.30	1.30	108	98	206
432	LMGDD	2	0.57	0.28	0.16	0.08	0.00	0.00	0.00	0.00	1.68	0.84	2.40	1.20	107	100	207
280 S	Elber-Meller-2000	9	0.64	0.07	0.94	0.10	0.07	0.01	0.57	0.06	1.56	0.17	3.78	0.42	102	108	210
512	ELAN	5	0.33	0.07	0.00	0.00	0.96	0.19	0.70	0.14	1.56	0.31	3.56	0.71	104	107	211
273	WXW	6	0.00	0.00	0.05	0.01	0.00	0.00	0.00	0.00	1.01	0.17	1.06	0.18	109	111	220
329	Tatsuya	1	0.39	0.39	0.00	0.00	0.00	0.00			0.00	0.00	0.39	0.39	113	109	222
459	mprabha	3	0.00	0.00	0.46	0.15	0.00	0.00	0.00	0.00	0.00	0.00	0.46	0.15	111	112	223
045	Del-Carpia-Yoshimori	8	0.54	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.54	0.07	110	114	224
003	Gerloff	1	0.25	0.25	0.00	0.00	0.00	0.00			0.00	0.00	0.25	0.25	115	110	225
055	Bystroff	10	0.42	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.42	0.04	112	116	228
216 S	Isites-Server	7	0.00	0.00	0.39	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.39	0.06	114	115	229
471	Chodera-John	1	0.00	0.00	0.00	0.00	0.00	0.00	0.14	0.14	0.00	0.00	0.14	0.14	116	113	229
018	Raghava-GPS	8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	117	117	234
220 S	valencia-cnb-pred	2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	118	117	235
243 S	Dill-Kern	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	119	117	236
248	BMERC	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	120	117	237
281	Mohan	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	121	117	238
383	HeadGordon-Teresa	2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	122	117	239
489	FCLD	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	123	117	240

Groups identified by S in the second column of the table are CAFASP participants. For each group, we report the number of evaluated domains (**N dom**), the total and average score according to GDT-TS (**GDT-TS tot** and **GDT-TS av**), RMSD of the *C*α of the core residues (**RMS CaC tot** and **RMS CaC av**), the percent of correctly aligned residues (**al0 tot** and **al0 av**), the RMSD of the structurally divergent regions (**loop tot** and **loop av**). The sums of the average and total scores are reported in the columns labeled **Sum av** and **Sum tot,** respectively. The ranking according to the last two values and their sum are reported in the last three columns (**Rank tot, Rank ave,** and **Sum of ranks,** respectively).

The % predicted value is different for each prediction and for each of the considered regions. However, most predictions of the less difficult targets were nearly complete. For example, the percentage of Cα predictions including 80% or more of the structure was >90% for all targets except T0092, T0090_2, T0089_1 (ca. 70%) and T0121_1, and T0103 (ca. 40%).

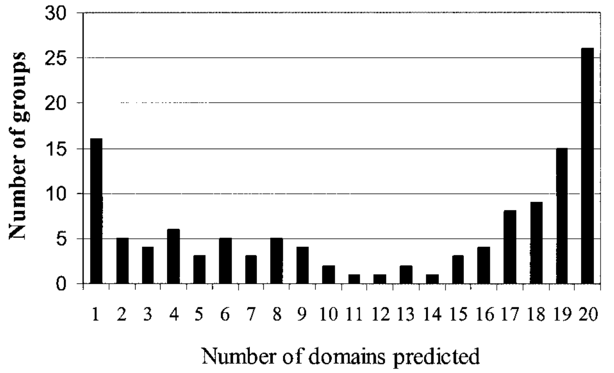


Fig. 2. Histogram of number of groups according to the number of targets predicted. Each domain of multidomain proteins was considered as a separate prediction.

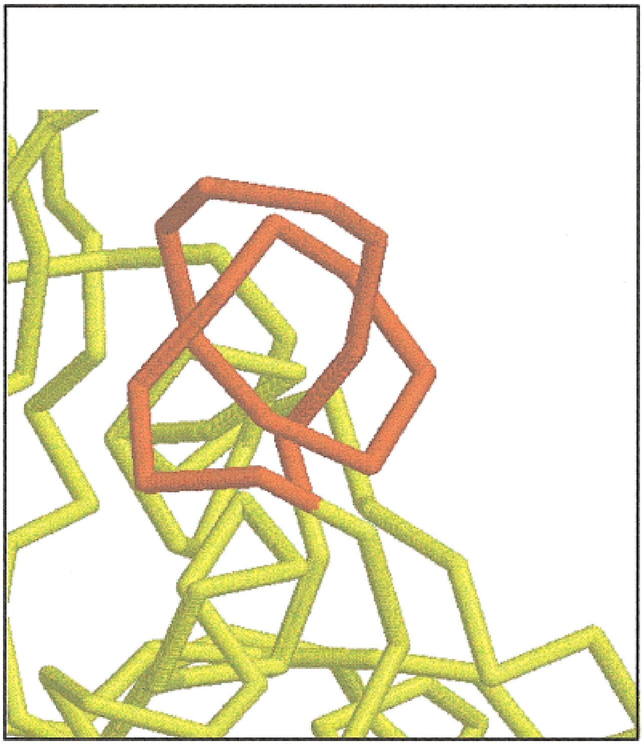


Fig. 3. One of the predictions for target 103. The region in red (residues 325–340) contains an "impossible" structure. Nevertheless, the parameters for this prediction are as follows: *GDT* = 38.6; *rms* = 9.56 *Å*; *al*0 = 126/368. The corresponding values of a good and "reasonable" prediction (group 042) are as follows: *GDT* = 37.7; *rms* = 9.21 *Å*; *al*0 = 126/368.

Assessment of Correctness of the Overall Fold

The results for all groups are shown in Table III where average and total scores for each considered parameter are reported. The last columns report the sum of the total and average score per group and the respective ranking. In general, they are different, because not all groups predicted all available targets (Fig. 2).

The "best" groups score well by using either ranking system, with some exceptions. One notable exception is group 384-Murzin (we use CASP4 group Id followed by the identifier chosen by the groups), who submitted only four predictions but achieved a very high score.

Three groups ranked among the first 15 for both total and average score, implying that their method performs well on as wide a range as possible of targets in the present experiment. Of these, groups 354-baker also obtained interesting results in fold recognition and novel

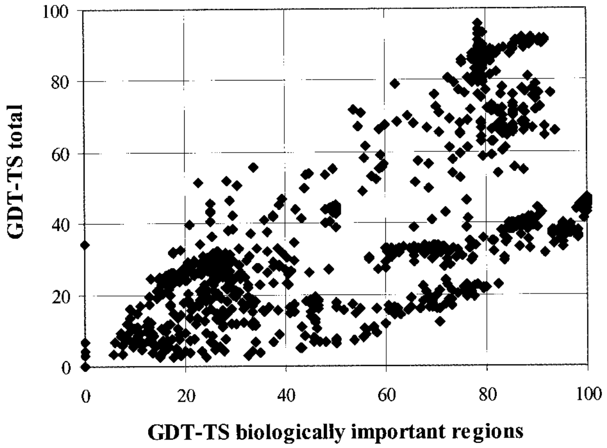


Fig. 4. Scatter plot of the GDT-TS values between target and predicted structure (*y* axis) versus that of the corresponding biologically important residues (*x* axis).

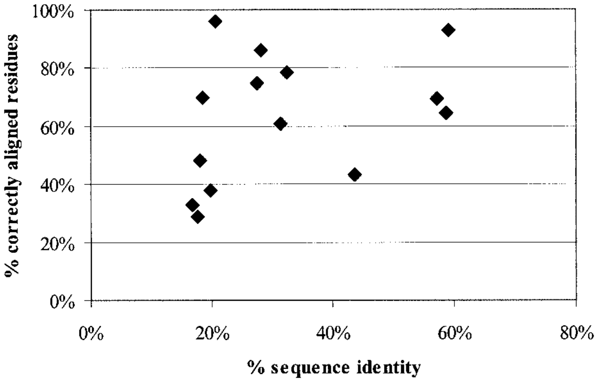


Fig. 5. Scatter plot of the percentage of correctly aligned residues versus the percentage of sequence identity between target and template for the best prediction for each target.

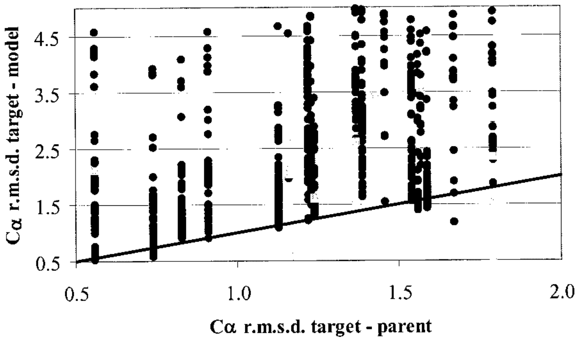


Fig. 6. Scatter plot of the RMSD between target and parent with respect to target and model. Only predictions with RMSD lower than 5 Å between template and model are shown. Squares correspond to predictions spanning <50% of the superimposable regions.

fold targets (as did group 384-Murzin).[4,5] Groups 406-VENCLOVAS and 126-Sternberg were asked to report on their comparative modeling results and methods.[6,7]

During the December 2000 meeting in Asilomar, we also asked other groups (chosen primarily on the basis of methods) to briefly discuss some specific aspect of the predictions (groups 137-Zhou-HX, 170-DNAmining.com/p-map, and 259-GenTHREADER for alignment and 354-baker, 393-Skolnick-Kolinski-THD, 31-BioInfo.PL for loop building).

It should be mentioned that the differences in the loop average scores of Table III are not very discriminating, and only for targets T0111 and T0128 were there complete loop predictions with a total Cα RMSD lower than 3 Å (and even in these easy cases only groups 126-Sternberg and 255-BinToHes and groups 161-GNM-FR, 237-Sali-Andrej and 429-CHEN-WENDY obtained values lower than 2 Å for T0111 and T0128, respectively). This finding suggests that no group is clearly better in predicting difficult loops and that a method able to consistently predict these regions with good accuracy is still an elusive goal in comparative modeling.

It is important to mention that scoring models using RMSD and/or GDT values may fail to highlight major problems with the models. Figure 3 shows one detail of a model that is clearly incorrect, because it contains impossible "knot" structures, although the corresponding RMSD and GDT values were extremely good. We tried to identify these cases (we found very few of them) and set their score to zero but cannot guarantee that none escaped our attention. Hopefully, for groups who submitted models for several targets, the effect of problems in a single prediction does not significantly alter the conclusions. Because the number of predictions submitted to future CASPs is not likely to decrease, some automatic method to detect these cases should be devised.

Figure 4 shows a plot of the GDT-TS values obtained for the total structure versus that for biologically important regions, after superposition of the whole model. It is apparent that, on average, the latter were predicted better (higher GDT-TS value). This is clearly an important aspect of the experiment. In our opinion, however, this reflects more the intrinsic better conservation of these regions than specific aspects of the methods.

By its own nature, comparative modeling exploits the evolutionary constraints posed by the biological function upon a protein and is, therefore, expected to work better on such regions. Questionnaires posed to the predictors toward the end of the experiment (data not shown) indicated that the groups and/or the methods devoted no special care to the prediction of these regions. The effect shown in Figure 4 is most likely due to the fact that biologically important regions are easier to align because of the pattern of conservation of functional residues and, in any case, more structurally conserved.

Alignment Quality

A pressing question in CASP is about general trends of prediction methods, but it is difficult to derive general conclusions using data including partial predictions on a limited number of targets.

We address here the issue of the quality of present sequence alignment methods, which poses the problem of having to take into account that a correct partial alignment is better than a complete alignment that includes the correct partial alignment but also contains incorrectly aligned residues. Our scoring system is designed to deal with this problem: if a prediction leaves out a part of the structure, it is penalized if that region is correctly predicted by other groups and rewarded if most groups were unable to predict it. This is very useful for discussing relative performance of groups on a set of targets but cannot be used to derive general conclusions.

One way to see what alignment methods are able to achieve is to show the percent of correctly aligned residues as a function of the percentage of sequence identity between each target and its best template for the best complete prediction for each target (Fig. 5). It is apparent from these data that alignment quality is still a problem and, more importantly, that its quality does not correlate with sequence identity between target and template. This should be a concern, especially because structural genomics projects plan to use sequence identity as a criterion to select candidate targets for protein structure determination and assume that comparative modeling can provide reliable models for the remaining proteins. Even a threshold of 50% would still fail to provide satisfactory models for some proteins.

Were the Models Any Better Than the Closest Structural Parent?

CASPs experiments provide a unique opportunity to evaluate whether a 3D model provides more information than simply aligning the target with the parent sequence. In other words, if the alignment is optimal and the parent correctly selected, one expects the model to be at least as good as the structure superposition of target and parent structures. If the modeling procedure adds more information than just the alignment, then the model should be closer to the target than the parent.

Figure 6 shows a scatter plot of the Cα RMSD after optimal superposition of the target and its closest parent structure (*x* axis) versus the Cα RMSD of target and model (*y* axis). In most cases, the structural similarity between model and target is worse, or even much worse, than that between target and parent. The factors primarily responsible for this effect are the selection of a nonoptimal parent structure and the errors in the alignment. In previous CASPs, the database of known structures probably was not populated sufficiently to allow predictors to make very different choices in the selection of the parent. As the database grows, also the selection of the parent is becoming a discriminating factor among the various groups.

As shown in Figure 6, some predictions have improved on the parent, but this only happened for target T0128_2 (where groups 126-Sternberg, 237-Sali-Andrej, 342-SBI-AT, 406-VENCLOVAS, and 526-Ginalski achieved a Cα

TABLE IV. Ratio Between the *Cα* RMSD for the Superposition Between Target and Model and Target and Best Template

Id	Group	T0089_2	T0099	T0111_1	T0112_2	T0113	T0121_1	T0128_1	T0128_2
012	Levitt	—	1.04	—	—	—	1.04	—	1.03
017	Yang-Ansuei	—	1.01	—	—	—	1.01	—	—
023	Jones	—	—	—	—	—	1.08	—	1.01
028	Ram-Samudrala	—	—	1.01	—	—		—	
031	BioInfo.PL	—	—	—	—	—	1.04	—	
032	Wolynes	—	—	—	—	—	1.03	—	
044	Walts-Wondrous	—	1.05	—	—	—	1.04	—	
047	kitasato-univ.	—	—	—	—	—	1.04	—	
065	Torda-Andrew	—	1.06	—	—	—	1.11		
077	rost	—	—	—	—	—	1.08	1.07	1.01
086	Bass-Michael	—	—	—	—	—	1.08	—	—
088	ORNL-PROSPECT	—	—	—	—	—	1.04	—	—
090	Hogue-Feldman	—	1.02	—	—	—		—	—
093	bioinbgu	—	—	—	—	—	1.04	—	—
094	SAM-T2K	—	—	—	—	1.02	1.04	—	—
095	blundell-tl	—	1.04	—	—	—	1.07	—	—
103	Zhou-HX	**1.29**	—	—	—	—	1.04	—	—
104	bioinbgu-gonp	—	1.03	—	—	—	1.06	—	—
105	bioinbgu-gonpm	—	1.04	—	—	—	1.05	—	—
106	bioinbgu-seqpprf	—	1.03	—	—	—	1.04	—	—
107	bioinbgu-seqpmprf	—	1.03	—	—	—	1.04	—	—
108	bioinbgu-prfseq	—	—	—	—	—	1.04	—	—
111	SAM-T99	—	1.01	—	—	—	1.04	—	—
118	Dlakic-Mensur	—	1.05	—	—	—	—	—	—
125	Sternberg-3D-JIGSAW	—	1.03	—	—	—	—	—	—
126	Sternberg	—	—	—	—	—	1.04	—	**1.22**
132	Sternberg-3DPSSM	—	1.06	—	—	—	1.04	—	—
133	CBC-FOLD	—	—	—	—	1.04	1.05	1.07	—
137	Zhou-HX	—	1.01	—	—	—	1.04	—	1.04
150	Chandonia-Cohen	—	—	—	—	—	1.04	—	—
152	Yoon	—		—	—	—		—	1.07
155	TUDELFT	—		—	—	—		1.05	1.01
158	PDB-Blast	—		—	—	—	1.04	—	—
161	GNM-FR	—	1.11	—	—	—		—	—
169	Dunbrack	—	1.08	—	—	—	1.04	—	—
170	DNAmining.com/p-map	—	1.03	—	—	—	1.04	—	—
173	Barton	—	1.01	—	—	—	1.06	1.09	—
186	SDSC1	—	—	—	—	—	1.04	—	—
197	Godzik	—	—	—	—	—	1.02	—	—
218	LAMBERT-Christophe	—	—	—	—	—	1.06	—	—
223	Braun-UTMB	—	1.06	—	—	—	—	—	—
229	UCLA-DOE	—	—	—	—	—	1.03	—	—
237	Sali-Andrej	—	—	—	—	—	—	—	**1.35**
241	Vajda	—	1.02	—	—	—	1.06	—	1.01
259	Gen THREADER	—	1.06	—	—	—	1.04	—	—
260	mGen THREADER	—	—	—	—	—	1.04	—	—
330	Zemla-Joanna	—	1.05	—	—	—	—	—	—
331	Levy	—	—	—	—	—	1.04	—	—
341	Lai	—	1.04	—	—	—	—	—	—
342	SBI-AT	—	—	—	—	—	1.02	—	**1.16**
354	baker	—	—	—	—	—	1.04	—	—
357	Fischer-Daniel	—	—	—	—	—	1.04	—	—
359	Cafasp-consensus	—	—	—	—	—	1.04	—	—
361	GMD-SCAI	—	—	—	—	—	—	—	1.15
363	Moult	—	—	—	—	—	1.04	1.04	1.01
381	SBfold	—	—	—	—	—	1.06	—	—
382	SBauto	—	—	—	—	—	1.07	—	—
389	123D+	—	—	—	—	1.01	1.04	—	—
401	Reva-Boris	—	1.05	—	—	—	—	—	—
406	VENCLOVAS	—	—	1.01	1.02	1.02	1.04	—	**1.22**
440	Deleage-Geourjon	—	—	—	—	—	1.02	—	—
447	MSI	—	1.06	—	—	—	1.04	—	—
453	Noguchi	—	1.01	—	—	—	—	—	—
465	YASARA	—	1.08	—	—	—	—	1.05	—
468	SBI-jz	—	1.03	—	—	—	1.05	—	—
500	FAMS	—	—	—	—	—	1.04	—	—
526	Ginalski	—	—	—	—	—	—	**1.34**	
536	Fox-Sheppard	—	—	—	—	—	1.05	—	—
	average	1.29	1.04	1.01	1.02	1.02	1.05	1.06	1.12

Only ratios above 1 are shown. Values above 1.15 are in bold.

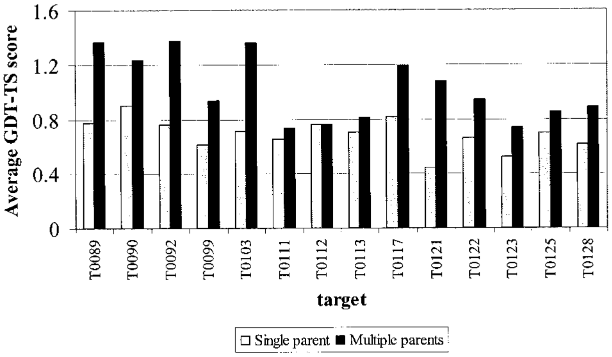


Fig. 7. Average GDT score for predictions using single or multiple parent structures.

RMSD value 15% lower than that with the closest template). One exception is the prediction of target T0089_2 of group 103-Fugue-Cam where the RMSD of the target with the closest template (1dga_A) is 1.67 Å over 26 residues and this group obtained a prediction with an RMSD value of 1.18 Å for 22 residues. Usually, however, the improvement is marginal (Table IV).

In general, methods using multiple parents perform better than those based on a single parent in general fold correctness (Fig. 7), although groups using both techniques are present among the highest ranking ones (data not shown).

Assessing the Correctness of the Details of the Models

In our experience as modelers, most of the targets were far from easy. Thus, it would be unreasonable to expect the most difficult predictions to be correct in their fine details.

We selected all the models having a Cα RMSD value lower than 5 Å and only used those to examine the details of the models.

A serious problem with CASP experiments is the limited number of available targets, and the need to reduce that number even further in analyzing the details only makes this worse. Hence, even more so, the results presented here should be treated with caution, and the reader should keep in mind that their statistical significance is necessarily limited. Table V shows the score, calculated as described above, for the following parameters:

- RMSD for all side-chain atoms
- RMSD for all side-chain atoms defined as reliable
- RMSD for all side chains in the core
- RMSD for all side chains in structurally divergent regions

It is hard to derive clear conclusions from these data for a number of reasons: the method of combining these scores with those obtained in Table III is very arbitrary, the number of evaluated models is rather small, and the measurements are all correlated (Fig. 8).

At the December 2000 meeting in Asilomar, we asked groups 237-Sali-Andrej, 197-Godzik, and 42-Honig-Barry to cover the issue of the accuracy of side-chain modeling. From this discussion and from our own attempts to derive conclusions from these data, one thing became apparent: neither the RMSD values or the percentage of correct side-chain angles (defined as within 30° from the experimental value) seemed appropriate to evaluate the details of a model on the basis of what we believe interesting to the end users. One would like to know whether a method is able to reproduce characteristics of the side chains that can effectively guide experiments or theories about the protein under study, and these are obviously interactions between groups of atoms. Rather than analyzing whether a side chain is correctly positioned in the reference space of the protein's main chain, it would be important to establish whether it is properly located with respect to other side chains.

The timing of the experiment did not allow us to develop and appropriately test different criteria, but we strongly believe that this should be accomplished before the next experiment takes place.

Servers

Another experiment run in parallel with CASP4 extended its scope: this experiment, named CAFASP2,[8] evaluated automatic methods of predicting protein structures using CASP4 targets.

All targets were processed through prediction servers that registered for the CAFASP experiment. Server developers or curators were then asked to re-submit these same predictions to CASP4 by using the correct format. The identity between the automatic and reformatted predictions is guaranteed by the CAFASP organizers.

In the CASP4 comparative modeling assessment, the automatic predictions were subjected to the same blind evaluation, together with all other predictions. Only after the process was completed were the assessors informed which groups were the publicly available servers.

Obviously, other CASP predictions might have been obtained by using automatic servers/programs, but only for those highlighted in Table III and V is it practically certain that there was no human intervention.

It is apparent that some of the servers perform as well as the best groups (260-mGenthreader and 259-GenThreader,[9] 111-SAM-T99,[10] and 93-bioinbgu[11] score among the first 20 groups; see Table III). The quality of their performance can be considered the base level for comparative modeling, because they provide good alignments to correctly selected templates. For example, servers 103-Zhou-HX, 108-bioinbgu-prfseq, 259-GenThreader, and 260-mGenThreader provided alignments for all comparative modeling targets and in all cases the percentage of residues correctly aligned was higher than the average for all groups for that prediction.

A Few Examples

Lists of numbers are the only way to describe comprehensively the quality of so many models; however, it is important to have a feeling for their structural significance. In Figure 9 we show the prediction with the highest

TABLE V. Group Scores for Details of Models

Id	Group	N dom	CRMSC tot	CRMSC ave	CRMSC rel tot	CRMSC rel ave	CRMSC core tot	CRMSC core ave	CRMSC LSH tot	CRMSC LSH ave	Ave tot	Ave ave	Rank ave	Rank tot	Sum of ranks	Glob	TOT
406	VENCLOVAS	14	5.23	0.37	4.04	0.37	5.37	0.38	2.66	0.19	4.33	0.33	1	4	5	6	11
126	Sternberg	12	2.54	0.21	4.14	0.21	2.57	0.21	1.40	0.12	2.66	0.19	8	13	21	15	36
042	Honig-Barry	11	2.57	0.23	4.35	0.23	2.57	0.23	1.37	0.12	2.71	0.21	7	11	18	26	44
354	baker	13	1.03	0.08	4.77	0.08	1.03	0.08	1.95	0.15	2.19	0.10	12	26	38	8	46
342	SBI-AT	12	1.98	0.16	4.41	0.17	2.05	0.17	0.43	0.04	2.22	0.13	11	20	31	20	51
237	Sali-Andrej	9	3.37	0.37	3.66	0.37	3.46	0.38	1.50	0.17	3.00	0.33	4	5	9	44	53
197	Godzik	14	1.27	0.09	4.59	0.09	1.29	0.09	1.61	0.12	2.19	0.10	13	25	38	22	60
429	CHEN-WENDY	10	2.98	0.30	4.55	0.30	3.03	0.30	2.14	0.21	3.17	0.28	3	6	9	88	97
002	Lomize-Andrei	1	2.08	2.08	5.37	2.09	2.19	2.19	2.30	2.30	2.99	2.17	5	1	6	96	102
169	Dunbrack	12	0.80	0.07	4.65	0.07	0.80	0.07	0.00	0.00	1.56	0.05	33	35	68	38	106
077	rost	11	1.83	0.17	4.37	0.17	1.89	0.17	0.10	0.01	2.05	0.13	15	21	36	73	109
465	YASARA	5	2.17	0.43	3.01	0.43	2.20	0.44	1.24	0.25	2.16	0.39	14	2	16	95	111
223	Braun-UTMB	11	0.14	0.01	4.66	0.01	0.14	0.01	1.33	0.12	1.57	0.04	32	40	72	41	113
031	BioInfo.PL	12	0.12	0.01	4.70	0.01	0.12	0.01	0.18	0.02	1.28	0.01	50	48	98	17	115
023	Jones	11	1.40	0.13	4.23	0.13	1.44	0.13	0.31	0.03	1.85	0.10	18	24	42	74	116
047	kitasato-univ.	10	0.69	0.07	4.11	0.07	0.69	0.07	0.00	0.00	1.37	0.05	43	34	77	39	116
363	Moult	12	5.51	0.46	3.80	0.48	5.79	0.48	1.54	0.13	4.16	0.39	2	3	5	112	117
447	MSI	12	1.69	0.14	4.61	0.14	1.74	0.15	2.21	0.18	2.56	0.15	9	15	24	94	118
526	Ginalski	5	0.67	0.13	4.60	0.13	0.67	0.13	0.66	0.13	1.65	0.13	28	19	47	71	118
0.12	Levitt	10	3.08	0.31	4.04	0.31	3.16	0.32	1.30	0.13	2.89	0.27	6	8	14	105	119
095	blundell-tl	10	1.72	0.17	4.96	0.17	1.75	0.17	0.72	0.07	2.29	0.15	10	17	27	96	123
500	FAMS	10	0.48	0.05	4.96	0.05	0.49	0.05	0.95	0.09	1.72	0.06	23	31	54	73	127
032	Wolynes	9	1.39	0.15	3.82	0.15	1.42	0.16	0.66	0.07	1.82	0.13	19	18	37	91	128
088	ORNL-PROSPECT	13	0.04	0.00	4.46	0.00	0.00	0.00	0.13	0.01	1.16	0.00	54	51	105	24	129
381	SBfold	10	0.50	0.05	5.08	0.05	0.50	0.05	0.75	0.07	1.71	0.06	24	33	57	80	137
382	SBauto	10	0.09	0.01	4.92	0.01	0.09	0.01	0.05	0.01	1.29	0.01	49	50	99	41	140
384	Murzin	1	0.00	0.00	5.75	0.00	0.00	0.00	0.00	0.00	1.44	0.00	40	53	93	54	147
028	Ram-Samudrala	8	1.13	0.14	4.72	0.14	1.13	0.14	0.00	0.00	1.75	0.11	22	23	45	114	159
094	SAM-T2K	5	0.12	0.02	3.92	0.02	0.12	0.02	0.35	0.07	1.13	0.04	55	42	97	63	160
468	SBI-jz	6	0.97	0.16	4.58	0.16	1.01	0.17	0.69	0.11	1.81	0.15	20	16	36	133	169
155	TUDELFT	7	1.09	0.16	3.64	0.29	2.07	0.30	0.00	0.00	1.70	0.18	25	14	39	133	172
218	LAMBERT-Christophe	10	0.00	0.00	4.58	0.00	0.00	0.00	0.13	0.01	1.18	0.00	53	52	105	67	172
086	Bass-Michael	8	2.20	0.27	2.89	0.33	2.27	0.28	0.62	0.08	1.99	0.24	16	9	25	149	174
359	Cafasp-consensus	5	0.00	0.00	4.48	0.00	0.00	0.00	0.00	0.00	1.12	0.00	56	53	109	65	174
241	Vajda	11	0.67	0.06	4.66	0.06	0.67	0.06	0.71	0.06	1.68	0.06	26	30	56	119	175
426	koehl	1	0.00	0.00	6.12	0.00	0.00	0.00	0.00	0.00	1.53	0.00	35	53	88	87	175
017	Yang-Ansuei	12	0.56	0.05	4.56	0.05	0.58	0.05	0.61	0.05	1.58	0.05	31	37	68	109	177
279	Bateman	2	0.00	0.00	5.31	0.00	0.00	0.00	0.12	0.06	1.36	0.02	46	46	92	91	183
486	Shoshana-Wodak	8	0.29	0.04	3.86	0.04	0.29	0.04	0.50	0.06	1.24	0.04	51	38	89	98	187
444	MOE-CCG	5	0.73	0.15	4.27	0.15	0.77	0.15	0.23	0.05	1.50	0.12	36	22	58	131	189
457	SBI-GR	11	0.82	0.07	4.45	0.07	0.82	0.07	0.38	0.03	1.62	0.06	29	29	58	132	190
278	Flake&mates	5	1.14	0.23	3.96	0.23	1.20	0.24	1.11	0.22	1.85	0.23	17	10	27	173	200
191	Lee-Jung	8	0.88	0.11	3.57	0.11	0.88	0.11	0.38	0.05	1.43	0.09	41	27	68	140	208
022	InforMax	4	1.06	0.26	3.63	0.27	1.17	0.29	0.00	0.00	1.46	0.21	38	12	50	162	212
393	Skolnick-Kolinski-THD	9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	61	53	114	100	214
133	CBC-FOLD	1	0.02	0.02	2.65	0.01	0.09	0.09	0.00	0.00	0.69	0.03	60	43	103	112	215
065	Torda-Andrew	9	0.59	0.07	4.24	0.07	0.59	0.07	0.02	0.00	1.36	0.05	45	36	81	141	222
058	Harrison-Weber	9	0.00	0.00	5.06	0.00	0.00	0.00	0.61	0.07	1.42	0.02	42	45	87	141	228
090	Hogue-Feldman	1	0.00	0.00	6.41	0.00	0.00	0.00	0.00	0.00	1.60	0.00	30	53	83	145	228
390	Taylor	10	0.09	0.01	4.88	0.01	0.05	0.00	0.13	0.01	1.29	0.01	48	49	97	135	232
161	GNM-FR	8	0.01	0.00	4.37	0.00	0.01	0.00	1.81	0.23	1.55	0.06	34	32	66	168	234
125 S	Sternberg-3D-JIGSAW	11	0.03	0.00	4.84	0.00	0.03	0.00	0.99	0.09	1.47	0.02	37	44	81	155	236
255	BinToHes	7	0.00	0.00	4.68	0.00	0.00	0.00	1.12	0.16	1.45	0.04	39	39	78	159	237
535	shankari	6	0.46	0.08	3.71	0.08	0.48	0.08	0.25	0.04	1.22	0.07	52	28	80	157	237
330	Zemla-Joanna	8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	61	53	114	129	243
414	Friesner	10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	61	53	114	145	259
498	Kollman-Baker	1	0.20	0.20	6.17	0.20	0.20	0.20	0.48	0.48	1.76	0.27	21	7	28	240	268
186 S	SDSC1	9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	61	53	114	155	269
187	SDSC2:Reddy-Bourne	5	0.00	0.00	5.22	0.00	0.00	0.00	0.24	0.05	1.36	0.01	44	47	91	179	270
440	Deleage-Geourjon	3	0.00	0.00	4.12	0.00	0.00	0.00	0.00	0.00	1.03	0.00	57	53	110	172	282
375	Ho-Kai-Ming	1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	61	53	114	171	285
512	ELAN	2	0.00	0.00	6.65	0.00	0.00	0.00	0.00	0.00	1.66	0.00	27	53	80	211	291
152	Yoon	3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	61	53	114	180	294
179 S	Sausage	3	0.14	0.05	3.50	0.06	0.15	0.05	0.02	0.01	0.95	0.04	59	41	100	200	300
432	LMGDD	2	0.00	0.00	4.01	0.00	0.00	0.00	0.00	0.00	1.00	0.00	58	53	111	207	318
459	mprabha	1	0.00	0.00	5.40	0.00	0.00	0.00	0.00	0.00	1.35	0.00	47	53	100	223	323

Only predictions with $C\alpha$ RMSD lower than 5 Å are considered. Groups identified by S in the second column of the table are CAFASP participants. For each group we report the number of evaluated domains (**N dom**), the total and average score according to RMSD for all side-chain atoms (**CRMSC tot** and **CRMSC ave**), RMSD for all reliable side-chain atoms (**CRMSC rel tot** and **CRMSC rel ave**), RMSD for all core side-chain atoms (**CRMSC core tot** and **CRMSC core ave**), RMSD for all side-chain atoms of the structurally divergent regions (**CRMSC LSH tot** and **CRMSC LSH ave**). The averages of the total and average scores are reported in the columns labeled **Ave tot** and **Ave ave,** respectively. The ranking according to the last two values and their sum are reported in the **Rank ave, Rank tot** and **Sum of rank** columns. The **Glob** column reports the group ranking for the quality of the overall fold (see Table II). The last column (**TOT**) is the sum of the last two columns.

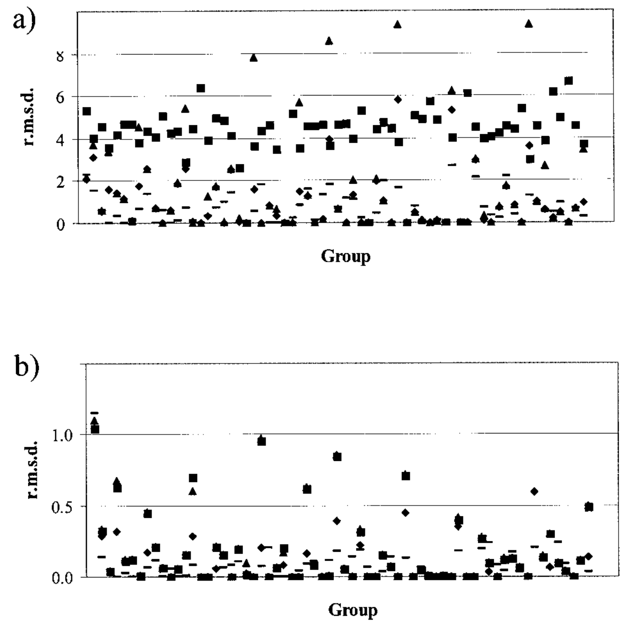


Fig. 8. Total (**a**) and average (**b**) score for RMSD of the side chain atoms of all (◆), reliable (■), core (▲), and loop (−) residues for groups submitting 3D predictions.

GDT-TS score superimposed on the respective target for each domain.

The groups that scored best for overall fold in our scheme are 406-VENCLOVAS, 354-baker, and 126-Sternberg (see Table III). Some of their predictions can be seen in the figure. In particular, note the prediction of target 92 obtained by group 406-VENCLOVAS. This group chose not to predict the N-terminal part and the 169–206 region of this protein, and these regions are indeed the most divergent between the target and its closest parent. The same group operated a similar (equally appropriate) choice on target 103.

CASP4 Versus CASP3

One important question that a CASP experiment should address is whether there has been clearly recognizable progress with respect to the previous one. Again, this poses the problem of evaluating predictions of targets with varying degrees of difficulty. Our scoring system is designed to take into account the difficulty of each target, on the basis of the predictors' average performance. Consequently, it cannot trace differences if the quality of predictions increases together with the difficulty of targets. We should mention, however, that the difficulty of targets between CASP3 and CASP4 does not seem to be substantially different when measured by the percentage identity between target and parent sequences (Fig. 10).

Under the hypothesis that the two sets of targets are equivalent, predictions can be compared with our scoring method, so we repeated the same analysis for CASP3, limiting ourselves to parameters that measure the correctness of overall folds. We excluded the CASP4 score for biologically important regions and loops, because data for the former are not available for CASP3 targets and the second depends too much on the individual features of the targets. The comparison is shown in Figure 11.

Notably, even if based on a completely different method of analysis, the groups selected by Jones and Kleywegt[12] in CASP3 for the overall fold (074, 136, 019) score among the first five in our analysis.

In some cases, it was possible to recognize groups participating in both CASP3 and CASP4 (Fig. 12), although there were some ambiguities (e.g., the Sternberg group participated as Sternberg but also as Sternberg-Jigsaw and Sternberg 3D PSSM in CASP4).

We are well aware of all the caveats of comparing predictions on different targets, but we are forced to conclude that, in first approximation, there has been no major improvement in comparative modeling between CASP3 and CASP4.

DISCUSSION

In CASP2, the number of participating groups was 72, which submitted 947 models denoted as first[13]; in CASP3, the numbers were 98 groups submitting 2261 first models.[14] In this experiment, there were 163 groups, and 4922 first models were deposited.[15]

Making all these predictions amounts to a huge amount of work, compared to which the workload of an assessor is minimum. In this report we have tried to convey the results of a very large experiment to a wider audience. It is, however, a unique and almost embarrassing aspect of CASP experiments that the people who produced the results are not those who draw the conclusions!

On the other hand, although all the data are available via the CASP WWW server (http://PredictionCenter.llnl.gov/casp4), the assessors are expected to comment on the results in a way that is meaningful to the widest possible audience.

We hope to have convinced the reader that CASPs should not be considered competitions, not for psychological reasons, but for serious technical ones connected with the complexity of the analysis of the data.

We do, however, believe that the field of protein structure prediction has benefited enormously from the CASP experiments and will continue to do so, if it is clear what can be learned from each of the experiments. To this end, we summarize here the results of our assessment of the CASP4 predictions.

Overall quality of the models

The models submitted to CASP4 are reasonable approximations of the target folds, but they are rarely closer to the experimental structure than their structural parents. There is still a long way to go before such models can be considered useful surrogates of experimental structures. It would certainly be impossible to use them for drug design or maybe even for rational mutagenesis of regions other than the active sites, especially when they are based on low-similarity parent structures.

From our analysis, it appears that one of the steps of the modeling procedure that does require some improvement

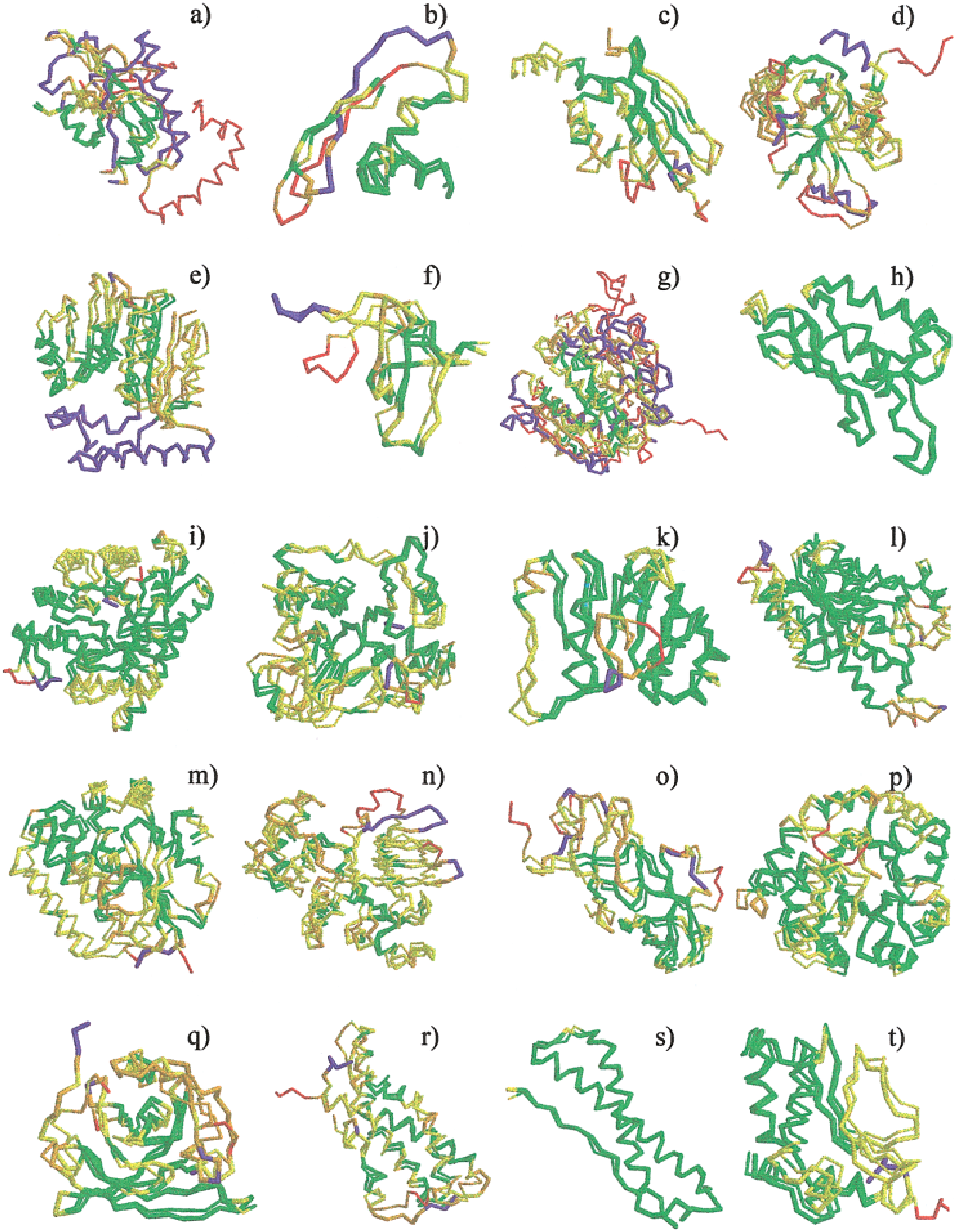


Fig. 9. Predictions with the highest value of GDT-TS for each of the target domains. Target and model structures are shown in thick and thin lines, respectively. Regions deviating <2.0, 4.0, and 8.0 Å are shown in green, yellow, and orange, respectively. Regions deviating >8.0 Å are shown in purple for the target and red for the model. Targets/groups are as follows: **a:** target 89 dom. 1/group 354-baker; **b:** target 89 dom. 2/group 126-Sternberg; **c:** target 89 dom. 4/group 390-Taylor; **d:** target 90 dom. 2/group 23-Jones; **e:** target 92/group 406-VENCLOVAS; **f:** target 99/group 237-Sali-Andrej; **g:** target 103/group 218-LAMBERT-Christophe; **h:** target 111 dom. 1 406-VENCLOVAS; **i:** target 111 dom. 2/group 32-Wolynes; **j:** target 112 dom. 1/group 406-VENCLOVAS; **k:** target 112 dom. 2/group 406-VENCLOVAS; **l:** target 113/group 406-VENCLOVAS; **m:** target 117/group 31-BioInfo.PL; **n:** target 121 dom. 1/group 381-SBfold; **o:** target 121 dom. 2/group 384-Murzin; **p:** target 122/group 526-Ginalski; **q:** target 123/group 17-Yang-Ansuei; **r:** target 125/group 95-blundell-tl; **s:** target 128 dom. 1/group 12-Levitt; **t:** target 128 dom. 2/group 526-Ginalski.

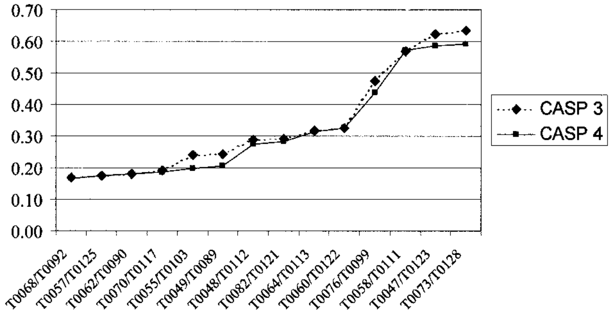


Fig. 10. Percentage sequence identity between targets and templates in CASP3 and CASP4.

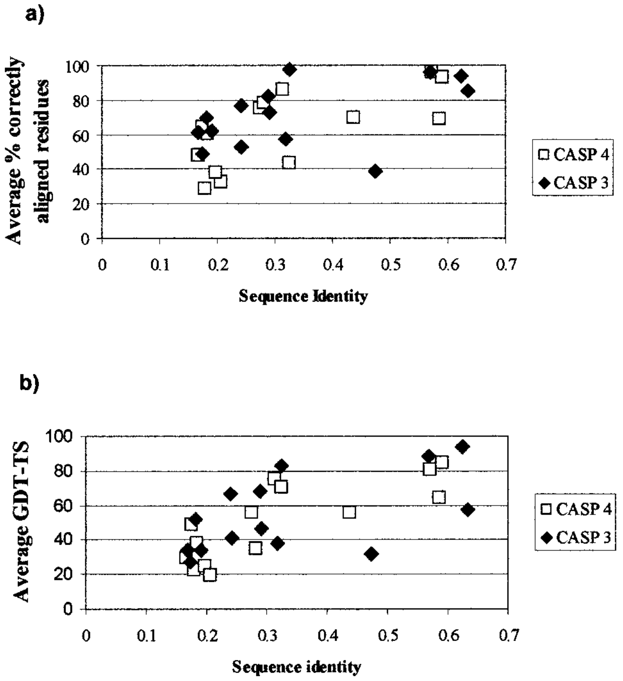


Fig. 11. Average values of the percentage of correctly aligned residues (**a**) and GDT-TS (**b**) in CASP3 and CASP4 as a function of the percentage of sequence identity between the target and its closest template.

is the development of strategies to identify the best parent structure available. The other aspect that requires attention is, obviously, the prediction of regions whose structure deviates substantially from that of the parent.

A feature of one of the most successful methods (group 406-VENCLOVAS) worth noting is the capacity to detect some of these "unpredictable" regions. From the viewpoint of a protein model user, having no model is better than having the wrong model, and methods able to detect local refolding of proteins are extremely useful.

Alignment quality

The alignment still represents a major problem in comparative modeling. It is still rare, even for the best methods, to achieve an accuracy above 80% for targets with sequence identity lower than 50% and, more importantly, it does not seem reasonable to use sequence identity as a measure of the expected accuracy of alignments.

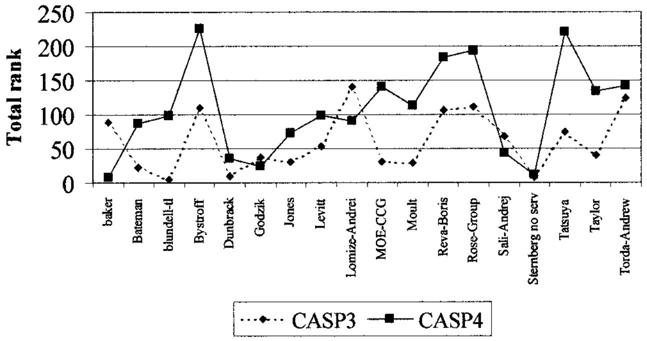


Fig. 12. Comparison of ranking in CASP3 and CASP4 for selected groups.

Domain orientation

In general, the methods assessed do not seem to achieve a reasonable degree of accuracy in predicting variations in the relative orientation of domains with respect to target proteins.

Target/parent versus target/model

We have once more to conclude that rarely is a model closer to the experimental structure than its structural parent.

Biologically important regions

As discussed above, biologically important regions are, on average, predicted better than other parts of the model. Although as we have said, this is due more to evolution than to prediction methods, it is still important to bear in mind. It is, in fact, the underlying reason why we believe that comparative modeling, with all its pitfalls and problems, is still an invaluable tool in modern biology.

Performance of automatic servers

A number of servers selected and aligned the target to the parent as well as the best "human" groups, and this is certainly important. This is, nonetheless, not true for all of them. The wide accessibility of publicly available servers offers the end user many choices but not necessarily sufficient information to select the most appropriate tool for the problem at hand. We believe that results of CASPs, as well as those from publicly available evaluation servers (see for example http://maple.bioc.columbia.edu/eva/, http://bioinfo.pl/LiveBench/, http://www.sanger.ac.uk/Users/lp1/MaxBench/), should be taken into serious consideration by the users.

CASP3 vs CASP4

Finally, we must comment on the comparison of the CASP3 and CASP4 results. It is clear that there has been only a marginal improvement, if any, between the two experiments. As a scientific community, we can either discard the problem or blame this unpleasant result on the dangers of comparing different experiments. However, we should honestly recognize that there has not been an overwhelming effort toward improving comparative modeling techniques in the last couple of years. As is clear from Figure 12, this also applies to individual groups, with the notable exception of the Baker group.

One last comment directed to the end users of models: Because of the time limit imposed on the predictors, CASP experiments are not necessarily representative of what one should expect as the standard performance of a predictor in real life. CASP results tend more toward lower, rather than average, quality of models. Certainly, all the CASP predictors would not, in real life, build a model as quickly as possible, trying to be faster than the experimentalists working on it, regardless of the difficulty involved. Certainly, they would take the necessary time, check every possible aspect of the problem, and verify the biological implications of the resulting model, thus substantially increasing the chances of its being correct.

In real life, the closest things to CASPs are large-scale prediction efforts connected, for example, with genome projects, which are becoming increasingly widespread and popular. In this respect, the results of CASP experiments are very important indicators of state of the art of prediction methods.

ACKNOWLEDGMENTS

A.T. thanks G. Tondinelli, F. De Palo, E. Fraschini, L. Di Marcoberardino, V. Tartagni, C. Di Menno, J. Clench, and C. Moroni for their continuous help and support. We also thank J. Clench for revising the manuscript, G. D'Alfonso for helpful discussions, the Protein Structure Prediction Center Members of the Lawrence Livermore National Laboratory, and especially Adam Zemla, for providing the data for the analysis and Arthur Lesk for useful comments on the manuscript. V.M. thanks the TMR Marie Curie Research Training Grant (Biotechnology Programme) for support.

REFERENCES

1. Zemla A, Venclovas C, Fidelis K, Moult J. Processing and evaluation of predictions in CASP4. Proteins 2001;Suppl 5:13–21.
2. 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: 3389–402.
3. Zemla A. LGA program: a method for finding 3-D similarities in protein structures 2000; http://PredictionCenter.llnl.gov/local/lga.
4. Bonneau R et al. Rosetta in CASP4: Progress in ab initio protein structure prediction. Proteins 2001;Suppl 5:119–126.
5. Murzin AG, Bateman A. CASP2 knowledge-based approach to distant homology recognition and fold prediction in CASP4. Proteins 2001;Suppl 5:76–85.
6. Bates PA, Kelley LA, MacCallum RM, Sternberg MJE. Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. Proteins 2001; Suppl 5:39–46.
7. Venclovas C. Comparative modeling of CASP4 target proteins: combining results of sequence search with 3D structure assessment. Proteins 2001;Suppl 5:47–54.
8. Fischer D, et al. CAFASP2: The second critical assessment of fully automated structure prediction methods. Proteins 2001;Suppl 5:171–183.
9. Jones DT. GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. J Mol Biol 1999;287: 797–815.
10. Karplus K, Barrett C, Hughey R. Hidden Markov models for detecting remote protein homologies. Bioinformatics 1998;14:846–56.
11. Fischer D. Hybrid fold recognition: combining sequence derived properties with evolutionary information. In: Pacific Symp. Biocomputing, Hawaii. Haway: World Scientific; 2000. p 119–130.
12. Jones TA, Kleywegt GJ. CASP3 comparative modeling evaluation. Proteins 1999;Suppl 3:30–46.
13. Moult J, Hubbard T, Bryant S, Fidelis K, Pedersen J. Critical assessment of methods of protein structure prediction (CASP): round II. Proteins 1997;Suppl 3:2–6.
14. Moult J, Hubbard T, Fidelis K, Pedersen J. Critical assessment of methods of protein structure prediction (CASP): round III. Proteins 1999;Suppl 3:2–6.
15. Moult J, Fidelis K, Zemla A, Hubbard T. Critical assessment of methods of protein structure prediction (CASP): round IV. Proteins 2001;Suppl 5:2–7.