PROTEINS: Structure, Function, and Genetics 53:352–368 (2003)

Assessment of Homology-Based Predictions in CASP5

Anna Tramontano[1]* and Veronica Morea[2]
[1]*Department of Biochemical Sciences, "A. Rossi Fanelli," University of Rome "La Sapienza," Rome, Italy*
[2]*CNR Institute of Molecular Biology and Pathology, University of Rome "La Sapienza," Rome, Italy*

ABSTRACT **This report describes the assessment of the homology-based predictions submitted to the fifth edition of the Critical Assessment of Methods for Protein Structure Prediction (CASP5) experiment. We assessed the ability of the methods to predict the overall fold, the portions of the structure that differ substantially between the target protein and its closest structural homologue and the conformation of the side-chains. We also compared the results with those obtained in previous editions of the experiment and derived some general conclusions about the state of the art of comparative modeling methods and their usefulness for experimentalists. Proteins 2003;53:352–368.**

INTRODUCTION

Comparative modeling methods exploit evolutionary relationships between proteins to produce three-dimensional models of unknown proteins using the known structures of their homologues as templates.

Although the technique is already a few decades old,[1] the protocol used so far has, by and large, remained unchanged. In general, it consists of a stepwise approach, starting from the identification of the template(s), the sequence alignment of target and template(s), the modeling first of the conserved regions and, subsequently, of the structurally divergent regions (SDR), the assignment of side-chain conformations, and the refinement of the model.

The quality of the produced models depends on the extent of structural conservation between the target proteins and their homologous parents,[2] on our ability to detect the homology and infer the correct structural alignment on the basis of the amino acid sequences, and on the capability of predicting those features, such as the conformation of SDRs and of side-chain, which cannot be directly inherited from the parent.

In the previous edition of CASP, in which we assessed the evolutionary based models with the help of Raphael Leplae,[3] we concluded that models were usually sufficiently accurate to be helpful in designing experiments, especially because biologically important regions were, in general, quite accurately modeled and better predicted than the rest of the structure. We also highlighted the problems that needed to be addressed in future developments, namely, the correct prediction of domain orientation in multidomain proteins as well as the prediction of the conformation of SDRs and side-chains. We also noticed that rarely was a model closer to the experimental structure than its structural template.

Although there has undoubtedly been progress in the field, these problems still represent major bottlenecks in comparative modeling methods although they were expected to be positively affected by the continuous increase in available protein sequences and structures. Because the protein structure space is more finely sampled, the probability of finding a close structural homologue to a target protein increases, thus widening the scope of the approach. In parallel, the availability of more protein sequences simplifies the detection of evolutionary relationships and improves the quality of the alignment between the target and the template(s) by virtue of multiple-sequence alignments, profiles, Hidden Markov Models, and intermediate sequence searches. As our knowledge of the protein sequence and structure space becomes more complete, we should also be able to exploit the available data more effectively and derive better heuristic rules for predicting other features of protein structures not directly inheritable from the parents. The constructive interplay between experimental and theoretical laboratories should help the latter to improve prediction techniques to be able to provide, in return, useful information to the experimentalists.

In this report, we present the results of our assessment of the models submitted in the comparative modeling category in CASP5, describe the reliability and usefulness of the produced models, and address the question of how much experimentalists and modelers can benefit from each other.

RESULTS

Targets

The assignment of a target to one of the three traditional CASP categories is domain based, because different domains can have different evolutionary relationships to proteins of known structure and, therefore, be assigned to different categories. A domain is assigned to the Comparative Modeling (CM) category when its evolutionary relationship to a protein of known structure can be detected on the basis of sequence data. The targets used in the analysis presented here (39 proteins accounting for 51 domains) are listed in Table I and are classified into close (CM) and remote [CM/FR(H)] homologues, according to Kinch et al.[4]

*Correspondence to: Anna Tramontano, Department of Biochemical Sciences "A. Rossi Fanelli," University of Rome "La Sapienza," P.le Aldo Moro, 5-00185 Rome, Italy. E-mail: Anna.Tramontano@uniromal.it

Received 4 March 2003; Accepted 23 June 2003

TABLE I. Targets Analyzed in the CM Category of CASP5

Target-ID	Name	Length	Method (PDB ID)	Description	CM Domains	Category
T0130	HI0073	114	X-ray	HI0073, *H. influenzae*	**T0130**	CM/FR(H)
T0132	HI0827	154	X-ray	HI0827, *H. influenzae*	T0132	CM/FR(H)
T0133	HIP1R	312	X-ray	HIP1R N-terminal domain, rat	T0133	CM/FR(H)
T0136	TC12S	523	X-ray	Transcarboxylase 12S subunit, *P. shermanii*	T0136_1:E4–E259	CM/FR(H)
					T0136_2:E260–E523	CM/FR(H)
T0137	FABP1	133	X-ray	Fatty acid-binding protein FABP1, *E. granulosus*	**T0137**	CM
T0140	1B11	103	X-ray	1b11, synthetic protein	T0140_1:B16–B74,A75–A102 (single, composite of 2 chains)	CM
T0141	AmpD	187	NMR (ÅYA)	AmpD protein, *C. freundii*	T0141	CM/FR(H)
T0142	NITRO	282	X-ray	Nitrophorin, *C. lectularius*	**T0142**	CM
T0143	V8prot	216	X-ray	V8 protease, *S. aureus*	**T0143_1:1–20,116–216**	CM
					T0143_2:21–115	CM
T0149	yjiA	318	X-ray (ILA9)	yjiA, *E. coli*	T0149_1:A2–A202	CM/FR(H)
T0150	L30E	102	X-ray (1H7M)	Ribosomal protein L30E, *T. celer*	**T0150**	CM
T0151	SSBP	164	X-ray	Single-strand-binding protein (SSB), *M. tuberculosis* H37Rv	T0151	CM
T0152	Rv1347c	210	X-ray	Hypothetical protein Rv1347c, *M. tuberculosis* H37Rv	T0152	CM/FR(H)
T0153	DUT	154	X-ray (1MQ7)	Deoxyuridine 5′-triphosphate nucleotidohydrolase (dUTPase), *M. tuberculosis*	*T0153*	CM
T0154	PANC	309	X-ray	Pantothenate synthetase, *M. tuberculosis*	*T0154_1:A3–A187*	CM
					T0154_2:A188–A290	CM
T0155	FOLX	133	X-ray	Probable dihydroneopterin aldolase (DHNA), *M. tuberculosis*	**T0155**	CM
T0159	PROX	309	X-ray	Glycine betaine-binding periplasmic protein, *E. coli*	T0159_1:X1–X91,X234–X309	CM/FR(H)
					T0159_2:X92–X233	CM/FR(H)
T0160	VAP-A	128	X-ray	VAP-A protein, rat	T0160	CM
T0165	CAH	318	X-ray (1L7A)	Cephalosporin C deacetylase, *B. subtilis*	T0165	CM/FR(H)
T0167	yckF	185	X-ray (1M3S)	Hypothetical cytosolic protein yckF, *B. subtilis*	**T0167**	CM
T0168	GLS2	327	X-ray (1MKI)	Glutaminase, *B. subtilis*	T0168_1:A1–A68,A210–A327	CM/FR(H)
					T0168_2:A69–A209	CM/FR(H)
T0169	yqjY	156	X-ray	yqjY, *B. subtilis*	T0169	CM/FR(H)
T0172	MRAW	299	X-ray (1M6Y,1N2X)	Conserved hypothetical protein, *T. maritima*	T0172_1:A2–A115,A217–A294	CM/FR(H)
T0176	yggU	100	NMR (1N91)	Hypothetical protein yggU, *E. coli*	T0176	CM
T0177	HP0162	240	X-ray (1MW7)	Hypothetical protein HP0162, *H. pylori*	**T0177_1:A21–A77**	CM
					T0177_2:A78–A130,A206–A240	CM
					T0177_3:A131–A205	CM
T0178	DEOC	219	X-ray (1MZH)	Deoxyribose-phosphate aldolase, *A. aeolicus*	**T0178**	CM
T0179	ywhF	276	X-ray (1IY9)	Spermidine synthase homologue, *B. subtilis*	**T0179_1:A2-A57**	CM
					T0179_2:A58-A275	CM
T0182	TM1478	250	X-ray (1O0X)	TM1478, *T. maritima*	**T0182**	CM
T0183	TM1559	248	X-ray (1O0Y)	TM1559, *T. maritima*	**T0183**	CM
T0184	TM1102	240	X-ray (1O0W)	TM1102, *T. maritima*	**T0184_1:B1-B165**	CM
					T0184_2:B166-B237	CM/FR(H)
T0185	TM0231	457	X-ray (1J6U)	TM0231, *T. maritima*	T0185_1:A1-A101	CM/FR(H)
					T0185_2:A102-A298	CM
					T0185_3:A299-A446	CM/FR(H)
T0186	TM0814	364	X-ray (1O12)	TM0814, *T. maritima*	**T0186_1:A1-A44,A331-A363**	CM
					T0186_2:A45-A256,A293-A330	CM/FR(H)
T0188	TM1816	124	X-ray (1O13)	TM1816, *T. maritima*	**T0188**	CM
T0189	TM0828	319	X-ray (1O14)	TM0828, *T. maritima*	T0189	CM/FR(H)
T0190	YEDX	114	X-ray	Transthyretin-related protein, *E. coli*	**T0190**	CM
T0191	AROE	282	X-ray	Shikimate 5-dehydrogenase, *M. jannaschii*	T0191_2:A105-A247	CM
T0192	SSAT	171	X-ray	Spermidime/spermine acetyltransferase (SSAT), human	T0192:2-153,154-171 (single, composite of 2 chains)	CM
T0193	ATBP	211	X-ray	AT-rich DNA-binding protein (ATBP), *T. aquaticus*	T0193_2:A79-A187,B188-B208 (single, composite of 2 chains)	CM
T0195	YJG8	299	X-ray	Hypothetical esterase in SMC3-MRPL8 intergenic region, *S. cerevisiae*	T0195	CM/FR(H)

Domains indicated by CM/FR(H) were also analyzed in the fold recognition category (for details, see Ref. 3). Domains in bold have been selected for SDR and side-chain quality evaluation.

As is customary in CASP, the targets showing remote homology to a known structure were analyzed by both the CM and the FR assessors.

It is evident that the increased sophistication of sequence-based methods to detect homologies has substantially enlarged the scope of application of comparative modeling techniques, which now include proteins with a sequence identity as low as 6% with their template. This is particularly remarkable compared with the CASP4 experiment, in which none of the CM targets shared <17% sequence identity with its template.[3]

Numerical Evaluation

The parameters used to measure the similarity between the models and the target structures are described in detail elsewhere in this issue.[5]

We used GDT-TS, root-mean-square deviation (RMSD) of the Cα atom, and percent of correctly aligned residues (AL0) to evaluate the general correctness of the models. These parameters are correlated to each other, and their concomitant usage can be seen as an internal control for the consistency of our evaluation.

We used local Cα RMSD to evaluate the prediction quality of SDRs and both the percentage of correctly assigned χ_1 and χ_2 angles (where correct is defined as within 30° from the experimentally determined value) and all atom RMSD values to assess the quality of the modeled side-chains. As detailed below, SDRs and side-chains were analyzed only on a subset of models, of sufficiently high overall quality.

Criteria

Underlying the results of a CASP assessment is always the question of the criteria used for evaluating and comparing the performance of different methods and of the statistical reliability of the conclusions derived from the data.

On one side, it has been pointed out that the assessment of the quality of different methods should only be based on the comparison of models produced for the same protein targets[6]; on the other, that such an approach is bound to penalize groups submitting a small number of models, because they could not be compared with other groups on the basis of a statistically significant number of common models.[7]

In this edition of the experiment, the number of target domains for the CM category is sufficiently high to allow the evaluation of prediction methods both on the basis of their average performance on the whole range of targets and by comparing their results on the same set of target proteins, as demonstrated by *t*-test analysis (see below).

For each of the selected measures, we calculated the average and standard deviation for each target and removed models "worse" than two standard deviations with respect to the average. In other words, we did not consider predictions whose GDT-TS or AL0 were more than two standard deviations below the average and those whose RMSD Cα was more than two standard deviations above the average. We then calculated average and standard deviation for the distribution of the remaining predictions for each target. Models better than average were assigned a score equal to their *z* score normalized by the percent of predicted structure; all other models were assigned a score of 0. The group score was defined as the average score over all the submitted models and, therefore, the submission of models of average or lower than average quality penalizes the group to the same extent. The rationale of removing very low scoring predictions in calculating the average score and of not penalizing predictions performing worse than average is to encourage (or at least not discourage) the usage of less tested and less widely used techniques, bound to be riskier in their expected results. We also compared each pair of groups on the basis of the models produced on the same subset of targets by performing a paired *t*-test on the GDT-TS or AL0 values.

Assessment of the Overall Quality of the Submitted Models

Table II shows the average score obtained by each group for both the GDT-TS and the AL0 measures. The GDT-TS scores for groups submitting models for >10 of the 51 comparative modeling target domains are also plotted in Figure 1. Five groups (20: Bujnicki-Janusz; 425: VENCLOVAS; 448: Murzin; 453: Ginalski; 517: GeneSilico) produced, on average, models of higher overall quality.

We show the results of the paired *t*-test on common models for the 20 groups with the highest GDT-TS and AL0 average score in Figures 2(a) and (b), respectively. The upper right part of the matrix reports the number of common models between the groups, and the diagonal reports the total number of predictions submitted by each group. We also show in the lower left part of the matrix the difference in average GDT-TS or AL0 values for the common models. Shaded cells highlight cases in which the distributions of the values obtained by the respective groups are not statistically significant ($p < 0.05$).

Notably, the top five groups in the average score ranking are clearly distinguishable from the others. As expected, groups submitting a small number of models might have an insufficient number of common models with other groups to be compared with them. The results obtained with either measure are virtually identical and, therefore, we can conclude that the number of groups, predictions, and targets in this edition of CASP is sufficient to derive statistically significant conclusions on the state of the art for this category.

As discussed previously, the difficulty of identifying the correct template is not uniformly distributed for all targets. Therefore, we performed our analysis also on the separate sets of "easy" CM targets and CM/FR(H) targets (see Table I). The results, shown in Table II, do not highlight any significant difference in the performances of the top groups in the two sets of data.

With the exception of Murzin (group 448) who used his well-known knowledge-based personal approach to prediction, all the best performing groups used several initial models that were then either combined (Bujnicki, groups 20 and 517) or scored to select the most plausible one. The

TABLE II. GDT-TS and AL0 Score for the Set of All Targets and GDT-TS Score for the "Easy" CM and CM/FR(H) Targets Analyzed Separately

Server	Predictor #	Group Name	N. Pred	Average GDT-TS Score	Average AL0 Score	Rank GDT-TS	Rank AL0	N. Pred	Average GDT-TS Score for CM Targets	Rank GDT-TS for CM Targets	N. Pred	Average GDT-TS Score for CM/FR(H) Targets	Rank GDT-TS for CM/FR(H) Targets
	P0001	SAM-T02-human	51	0.37	0.35	87	92	30	0.33	103	21	0.43	57
	P0002	BAKER	51	0.62	0.63	21	20	30	0.35	98	21	1.00	5
	P0006	BioInfo.PL	51	0.74	0.81	8	8	30	0.60	12	21	0.93	9
*	P0007	SAM-T99	50	0.33	0.38	99	80	30	0.40	81	20	0.22	106
	P0008	CBC-FOLD	49	0.29	0.30	110	102	30	0.22	129	19	0.41	64
	P0010	Skolnick-Kolinski	51	0.60	0.57	25	33	30	0.47	49	21	0.77	15
	P0011	Floudas-C.A.	9	0.00	0.00	164	164	7	0.00	157	2	0.00	147
	P0012	ORNL-PROSPECT	51	0.64	0.67	16	18	30	0.52	24	21	0.81	13
*	P0014	FUGUE2	49	0.33	0.35	98	91	30	0.33	106	19	0.34	81
	P0015	KGI-QMW	8	0.06	0.07	152	149	1	0.00	157	7	0.07	129
	P0016	Levitt	51	0.40	0.36	75	88	30	0.34	102	21	0.48	48
	P0020	Bujnicki-Janusz	48	0.85	0.92	4	4	29	0.69	4	19	1.10	3
	P0021	Baldi	47	0.00	0.00	164	164	27	0.00	157	20	0.00	147
	P0024	Braun-Werner	45	0.42	0.45	63	63	28	0.37	92	17	0.52	42
	P0025	HMMSPECTR	51	0.42	0.32	68	97	30	0.35	97	21	0.51	44
	P0028	Celltech	51	0.66	0.74	15	10	30	0.63	8	21	0.69	23
*	P0029	BAKER-ROBETTA	51	0.64	0.60	17	27	30	0.52	26	21	0.80	14
	P0032	Pan	49	0.15	0.11	137	136	28	0.19	137	21	0.10	119
*	P0034	ESyPred3D	39	0.39	0.41	79	75	29	0.43	64	10	0.25	101
	P0035	LAMBERT-Christophe	39	0.43	0.41	62	73	28	0.45	57	11	0.38	73
*	P0038	Pcons2	49	0.28	0.24	111	116	30	0.31	108	19	0.24	102
*	P0039	Pcons3	50	0.46	0.51	51	40	30	0.43	62	20	0.49	46
*	P0040	Pmodel	50	0.58	0.59	26	28	30	0.50	31	20	0.71	19
	P0041	CBRC	51	0.50	0.51	43	39	30	0.49	40	21	0.52	40
	P0044	SHESTOPALOV	46	0.17	0.13	133	135	30	0.24	124	16	0.04	133
*	P0045	Pmodel3	50	0.50	0.46	42	57	30	0.45	56	20	0.58	32
*	P0046	Pcomb	48	0.31	0.25	105	112	29	0.25	122	19	0.39	68
	P0047	FROST-MIG	45	0.38	0.36	83	89	28	0.49	34	17	0.19	111
	P0050	DelCLAB	50	0.00	0.01	164	158	30	0.00	157	20	0.00	147
	P0051	SAMUDRALA-NEWFOLD	28	0.09	0.00	147	164	10	0.00	157	18	0.14	116
	P0052	SAMUDRALA-FOLD-RECOGNITION	49	0.01	0.00	161	164	29	0.00	157	20	0.03	137
	P0053	SAMUDRALA-COMPARATIVE-MODELLING	49	0.16	0.10	135	140	29	0.22	130	20	0.08	127
	P0054	Raghava-Gajendara	31	0.17	0.19	134	123	17	0.30	111	14	0.01	142
	P0056	Zhou-HX	51	0.36	0.50	91	42	30	0.40	82	21	0.31	90
	P0061	Tsai	6	0.00	0.00	164	164	1	0.00	157	5	0.00	147
	P0064	Aligners	23	0.09	0.14	146	133	5	0.15	142	18	0.08	126
*	P0065	SUPERFAMILY	47	0.29	0.30	108	103	29	0.29	113	18	0.30	91
	P0067	Jones	49	0.62	0.59	20	29	29	0.52	25	20	0.76	16
	P0068	Jones-NewFold	1	0.00	0.00	164	164	0	0.00	157	1	0.00	147
*	P0070	GenTHREADER	49	0.25	0.24	118	114	30	0.34	101	19	0.12	117

TABLE II. (Continued)

Server	Predictor #	Group Name	N. Pred	Average GDT-TS Score	Average AL0 Score	Rank GDT-TS	Rank AL0	N. Pred	Average GDT-TS Score for CM Targets	Rank GDT-TS for CM Targets	N. Pred	Average GDT-TS Score for CM/FR(H) Targets	Rank GDT-TS for CM/FR(H) Targets
*	P0071	mGenTHREADER	49	0.37	0.35	88	94	30	0.37	89	19	0.35	79
	P0078	rost	48	0.40	0.43	74	68	28	0.37	90	20	0.44	55
	P0080	Bilab	49	0.48	0.45	48	61	30	0.49	36	19	0.46	50
	P0081	AS2TS	27	0.48	0.40	46	78	21	0.53	23	6	0.33	84
	P0084	SBC	51	0.52	0.50	38	45	30	0.49	41	21	0.57	34
*	P0086	SUPFAM_PP	51	0.28	0.28	113	105	30	0.20	134	21	0.38	71
	P0090	Holm	35	0.38	0.44	80	64	20	0.27	117	15	0.53	39
	P0091	Tsunami	11	0.05	0.05	154	152	7	0.08	150	4	0.00	147
	P0093	Schulten-Wolynes	50	0.34	0.32	95	96	29	0.42	69	21	0.23	103
	P0096	Bates-Paul	51	0.58	0.68	27	15	30	0.51	30	21	0.69	24
	P0098	Camacho-Carlos	43	0.06	0.06	153	151	28	0.09	149	15	0.00	147
	P0099	Camacho-Carlos	49	0.41	0.50	72	46	30	0.43	63	19	0.38	72
	P0100	SBI	42	0.61	0.71	23	11	30	0.60	10	12	0.63	28
	P0102	Bioinformatics group (Dr. Lu)	3	0.00	0.00	164	164	3	0.00	157	0	0.00	147
	P0105	Sternberg	51	0.46	0.48	50	51	30	0.42	67	21	0.51	43
	P0108	CaspIta	49	0.42	0.42	64	71	30	0.48	48	19	0.34	80
	P0110	Honig Lab	50	0.72	0.76	9	9	29	0.58	16	21	0.91	10
	P0112	Friesner	50	0.55	0.53	33	36	30	0.49	38	20	0.65	26
	P0124	moe-ccg	5	0.19	0.19	127	122	4	0.19	135	1	0.17	112
	P0131	Bystroff	25	0.27	0.26	116	110	22	0.27	118	3	0.26	96
*	P0132	I-sites/Bystroff	41	0.01	0.01	160	159	24	0.02	156	17	0.00	147
*	P0135	MPALIGN	45	0.24	0.26	121	111	27	0.22	128	18	0.27	94
*	P0136	RPFOLD/Raghava-Gajendra	31	0.19	0.22	128	119	17	0.33	104	14	0.01	142
*	P0138	PROTINFO-CM	50	0.16	0.09	136	146	30	0.21	131	20	0.08	127
*	P0139	PROTINFO-FR	50	0.01	0.00	162	164	30	0.00	157	20	0.03	137
*	P0140	PROTINFO-AB	19	0.02	0.00	158	164	6	0.00	157	13	0.03	135
*	P0142	Tsunami	44	0.12	0.10	142	141	27	0.19	136	17	0.02	139
	P0144	RAPTOR	51	0.42	0.50	69	41	30	0.39	84	21	0.45	52
	P0153	CHIMERA	51	0.63	0.61	18	26	30	0.59	15	21	0.70	22
*	P0168	FAMS	50	0.35	0.31	94	99	30	0.39	83	20	0.28	93
*	P0169	FAMSD	49	0.39	0.41	77	76	30	0.38	86	19	0.41	60
	P0170	CHIMERAX	51	0.44	0.43	57	69	30	0.47	51	21	0.40	65
	P0180	evolutionaries	50	0.40	0.37	76	81	30	0.49	39	20	0.26	99
*	P0183	arby-scai	50	0.23	0.17	124	127	30	0.23	127	20	0.23	104
	P0188	harrison	21	0.24	0.21	122	121	20	0.25	121	1	0.00	147
*	P0189	SAM-T02-server	50	0.42	0.53	65	37	30	0.43	65	20	0.41	61
*	P0194	BioInfo.PL-ORFeus	50	0.45	0.55	55	34	30	0.41	76	20	0.50	45
*	P0195	ORNL-PROSPECT	50	0.32	0.31	100	100	30	0.36	93	20	0.27	95
*	P0196	BioInfo.PL-PDB-Blast	49	0.27	0.27	115	107	30	0.29	112	19	0.22	105
	P0202	Yasara-Pushchino	43	0.53	0.58	35	30	27	0.53	22	16	0.52	41
	P0203	Pushchino	51	0.51	0.49	40	47	30	0.48	44	21	0.55	37

TABLE II. (Continued)

Server	Predictor #	Group Name	N. Pred	Average GDT-TS Score	Average AL0 Score	Rank GDT-TS	Rank AL0	N. Pred	Average GDT-TS Score for CM Targets	Rank GDT-TS for CM Targets	N. Pred	Average GDT-TS Score for CM/FR(H) Targets	Rank GDT-TS for CM/FR(H) Targets
*	P0207	BioInfo.PL-ORFblast	49	0.31	0.35	101	90	30	0.39	85	19	0.20	110
*	P0208	BioInfo.PL-BasicB	50	0.44	0.50	60	43	30	0.47	53	20	0.40	66
*	P0209	BioInfo.PL-BasicC	49	0.34	0.36	97	87	30	0.30	110	19	0.39	69
	P0210	Accelrys	6	0.48	0.37	47	83	4	0.56	19	2	0.34	82
	P0214	Advanced-ONIZUKA	51	0.01	0.01	159	160	30	0.02	155	21	0.00	147
*	P0221	INBGU	50	0.38	0.42	81	70	30	0.42	71	20	0.34	83
*	P0222	3D-SHOTGUN-INBGU	42	0.52	0.54	36	35	25	0.40	79	17	0.70	21
*	P0223	3D-SHOTGUN-3DS3	46	0.66	0.67	14	17	26	0.48	42	20	0.88	12
*	P0224	3D-SHOTGUN-3DS5	46	0.67	0.69	12	14	26	0.45	59	20	0.96	8
*	P0226	FUGUE3	49	0.38	0.41	82	74	30	0.36	96	19	0.41	59
*	P0227	3D-JIGSAW	45	0.31	0.27	103	108	30	0.36	94	15	0.21	108
	P0228	Mishali-Amir	40	0.57	0.62	30	23	21	0.49	35	19	0.65	27
	P0229	3D-PSSM	50	0.43	0.45	61	62	30	0.41	75	20	0.47	49
	P0230	LIBELLULA	26	0.31	0.30	104	104	13	0.32	107	13	0.29	92
	P0231	CNBpred	16	0.01	0.00	163	163	5	0.00	157	11	0.01	144
	P0233	RAPTOR	50	0.36	0.43	90	66	30	0.33	105	20	0.42	58
	P0234	ALAX	41	0.34	0.34	96	95	27	0.41	74	14	0.20	109
	P0240	GERLOFF	4	0.58	0.63	28	22	0	0.00	157	4	0.58	33
	P0242	GeneSilico.PL-servers-only	50	0.56	0.58	32	32	29	0.53	21	21	0.60	29
	P0246	MZ-Brussels	38	0.37	0.31	84	101	28	0.47	50	10	0.10	120
	P0249	SMD-CCS	5	0.18	0.08	130	147	3	0.29	114	2	0.01	145
*	P0253	CMap23Dpro	1	0.00	0.00	164	164	0	0.00	157	1	0.00	147
	P0258	Biokol	2	0.00	0.00	164	164	1	0.00	157	1	0.00	147
	P0262	Yoon	15	0.00	0.00	164	162	12	0.00	157	3	0.00	147
	P0264	CHEN-WENDY	37	0.45	0.49	56	49	29	0.50	32	8	0.26	100
	P0265	Sasson-Iris	51	0.66	0.67	13	16	30	0.59	14	21	0.75	17
	P0267	HOGUE-SLRI	43	0.10	0.11	145	138	29	0.10	148	14	0.10	122
	P0270	Solovyev-Softberry	50	0.09	0.09	148	144	30	0.14	143	20	0.00	146
	P0271	Head-Gordon	3	0.00	0.00	164	164	0	0.00	157	3	0.00	147
	P0282	Protfinder	48	0.04	0.00	156	161	29	0.06	153	19	0.00	147
*	P0283	Protfinder	38	0.04	0.02	155	156	21	0.07	151	17	0.00	147
*	P0286	3DSN-INBGU	50	0.45	0.50	54	44	30	0.45	58	20	0.45	54
	P0288	Lomize-Andrei	51	0.52	0.58	37	31	30	0.47	52	21	0.60	31
*	P0290	FORTE1	50	0.37	0.37	89	82	30	0.40	80	20	0.31	88
*	P0291	FFAS	50	0.30	0.31	106	98	29	0.28	116	21	0.33	85
	P0292	Osgdj	10	0.00	0.02	164	155	5	0.00	157	5	0.00	147
	P0294	Wolynes-Schulten	3	0.00	0.00	164	164	1	0.00	157	2	0.00	147
	P0299	sp_cadd	4	0.68	0.64	11	19	4	0.68	7	0	0.00	147
*	P0300	Ron-Elber	39	0.00	0.00	164	164	24	0.00	157	15	0.00	147
	P0301	UCLA-DOE	41	0.47	0.44	49	65	23	0.55	20	18	0.37	75
*	P0309	FFAS03	49	0.44	0.46	59	58	30	0.42	70	19	0.48	47
	P0314	Scheraga-Harold	12	0.00	0.00	164	164	8	0.00	157	4	0.00	147

TABLE II. (Continued)

Server	Predictor #	Group Name	N. Pred	Average GDT-TS Score	Average AL0 Score	Rank GDT-TS	Rank AL0	N. Pred	Average GDT-TS Score for CM Targets	Rank GDT-TS for CM Targets	N. Pred	Average GDT-TS Score for CM/FR(H) Targets	Rank GDT-TS for CM/FR(H) Targets
	P0326	INFORMAX	15	0.13	0.11	141	137	12	0.16	141	3	0.00	147
	P0327	Rokko	10	0.45	0.43	53	67	1	0.48	46	9	0.45	53
	P0329	Dunbrack	46	0.57	0.63	29	21	28	0.58	17	18	0.56	35
	P0331	SRBI	37	0.07	0.06	150	150	24	0.11	146	13	0.00	147
*	P0334	MZ-Brussels	15	0.29	0.24	109	115	11	0.26	120	4	0.39	70
*	P0344	ITSM	24	0.28	0.17	112	126	22	0.30	109	2	0.02	141
	P0347	SDSC2:Reddy-Bourne	38	0.26	0.17	117	128	24	0.36	95	14	0.10	123
	P0349	Shortle	1	0.00	0.00	164	164	0	0.00	157	1	0.00	147
	P0351	Huber-Torda	51	0.27	0.25	114	113	30	0.35	99	21	0.16	114
	P0359	GEM	51	0.46	0.46	52	59	30	0.45	54	21	0.46	51
*	P0362	PSPT	48	0.37	0.35	85	93	30	0.38	87	18	0.35	78
	P0368	José	51	0.60	0.52	24	38	30	0.60	13	21	0.60	30
	P0370	nexxus-delrio	2	0.50	0.69	45	12	2	0.50	33	0	0.00	147
	P0373	Brooks	51	0.23	0.18	123	124	30	0.12	144	21	0.39	67
	P0377	THW-FR	46	0.11	0.14	144	131	28	0.12	145	18	0.10	124
*	P0378	PILOT	39	0.22	0.21	126	120	26	0.24	123	13	0.16	115
	P0384	Bass-Michael	35	0.13	0.10	139	143	24	0.17	138	11	0.05	132
	P0391	Lund-Ole	39	0.12	0.14	143	132	26	0.16	139	13	0.03	134
	P0397	CIRB	51	0.18	0.14	129	130	30	0.23	125	21	0.11	118
	P0400	SPAM1	49	0.23	0.22	125	118	29	0.16	140	20	0.32	87
	P0401	Doniach	2	0.00	0.00	164	164	1	0.00	157	1	0.00	147
	P0403	sk-lab	2	0.00	0.00	164	164	1	0.00	157	1	0.00	147
	P0411	Roland-Sundar	1	0.00	0.00	164	164	1	0.00	157	0	0.00	147
	P0417	CBSU	42	0.40	0.37	73	84	28	0.42	72	14	0.37	74
	P0419	luethy	51	0.13	0.10	140	142	30	0.21	133	21	0.02	140
	P0423	Taylor	51	0.29	0.23	107	117	30	0.28	115	21	0.31	89
	P0424	Jun_Zhu	14	0.07	0.09	149	145	11	0.07	152	3	0.10	121
	P0425	VENCLOVAS	19	0.88	1.01	2	3	13	0.70	3	6	1.26	1
	P0427	FISCHER	51	0.76	0.69	7	13	30	0.61	9	21	0.97	6
	P0429	keasar	6	0.24	0.14	120	134	3	0.23	126	3	0.26	98
	P0435	Fujita	10	0.86	0.88	3	5	1	0.48	46	9	0.90	11
	P0437	Ho-Kai-Ming	51	0.07	0.05	151	153	30	0.10	147	21	0.03	136
	P0440	Biogen	18	0.35	0.27	92	109	6	0.40	78	12	0.33	86
	P0441	Meller-Adamczak	2	0.17	0.07	132	148	1	0.35	100	1	0.00	147
	P0447	Cam-Biochem	50	0.51	0.47	41	53	29	0.48	45	21	0.55	38
	P0448	Murzin	17	1.08	1.06	1	2	3	0.85	1	14	1.13	2
	P0450	TOME	47	0.62	0.62	19	25	30	0.57	18	17	0.73	18
	P0453	Ginalski	51	0.85	0.88	5	6	30	0.68	6	21	1.09	4
	P0459	Shakhnovich-Eugene	10	0.00	0.00	164	164	9	0.00	157	1	0.00	147
	P0464	ATOME	48	0.37	0.42	86	72	30	0.44	61	18	0.26	97
	P0465	POMI	24	0.54	0.47	34	54	10	0.51	29	14	0.56	36
	P0470	sezerman	17	0.18	0.10	131	139	10	0.26	119	7	0.05	131

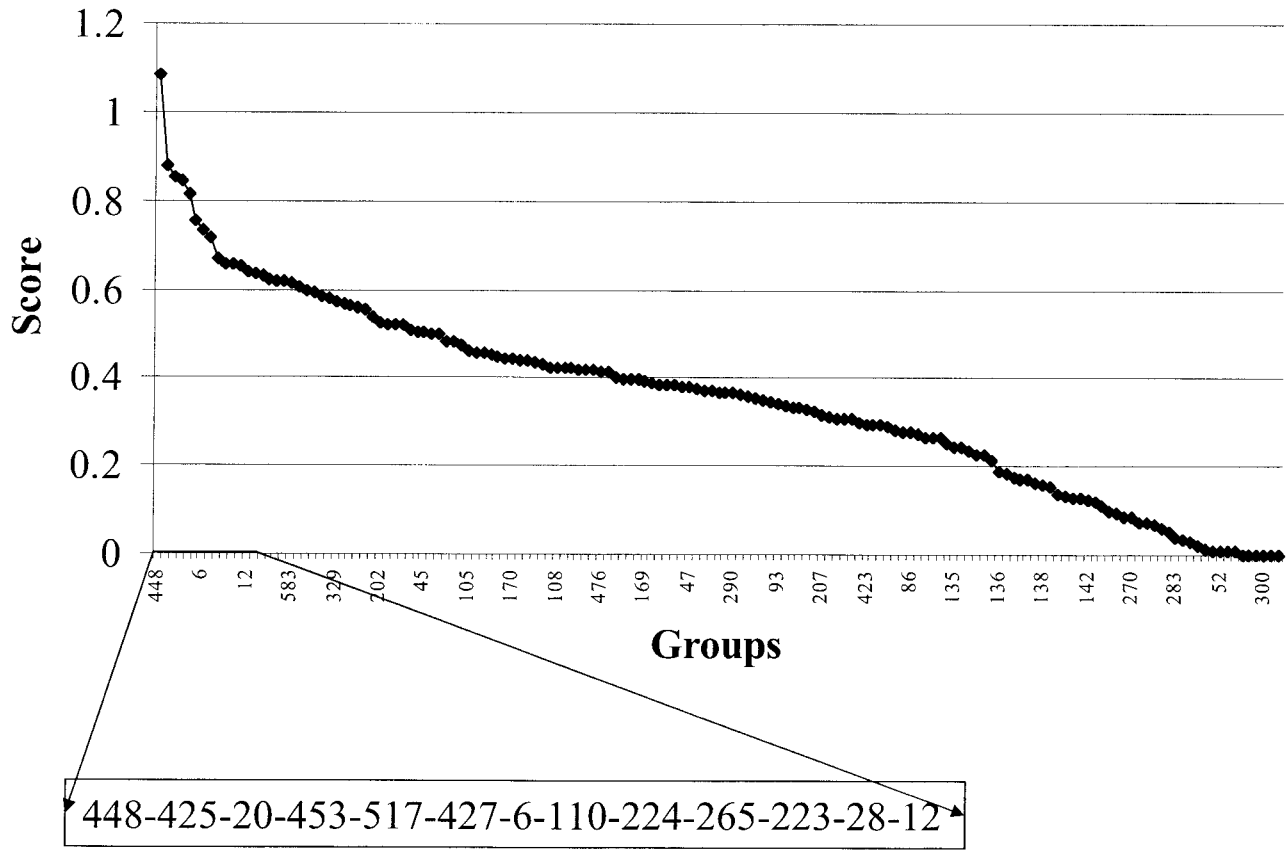

Fig. 1. Average GDT-TS score. Plot of the average GDT-TS score for groups submitting models for more than ten target domains.

initial models were obtained by using different automatic servers (Ginalski, Bujnicki) or different starting points in the multiple-sequence alignment (Venclovas).

As is discussed below, differences in performance could only be clearly detected in terms of general quality of the models, whereas the results obtained for structurally divergent regions and side-chain placement are quite similar for most groups, implying that the differences in the overall quality of the models have to be largely attributed to a better target-template alignment.

Automatic Versus Human Predictions

Another experiment runs in parallel with CASP: the Critical Assessment of Fully Automated Methods for Structure Prediction (CAFASP).[8]

It is of interest to compare the results of completely automatic methods with those of human predictors, because the former can be used by the community at large.

The CAFASP organizers submit models produced by servers to the CASP system, and they guarantee that there has been no human intervention in the construction of the models. A different matter is to decide which percentage of the models submitted via the normal CASP route is generated by automatic methods. One rough estimate of this number can be obtained by assuming that identical models are most likely automatically generated or at least generated by using publicly available methods. An all-against-all comparison of all the almost 7000 submitted models is unfeasible within a reasonable time frame, but we reasoned that the percent of models having identical Cα RMSD values with respect to the target structure should give a reasonable ballpark figure for the number of models where human intervention was limited.

The data shown in Figure 3 indicate that many models are either automatically generated or, at least, produced by using standard and generally available methodologies. As expected, the effect is more pronounced on easiest targets.

The submission of these "trivial" models has at least two effects: 1) the distinction between CAFASP and CASP submissions is rather artificial; therefore, any general conclusion about the relative performance of servers and "human" predictors derived from the comparison of the two experiments has to be taken with very much care. The second is that any statistical analysis of the data will suffer from the bias introduced in the data set by the over-representation of some of the models.

Nevertheless, we have to note here that CAFASP submissions achieved significant success, certainly more pronounced than in the previous CASP experiment.[3] In the CM category, the 3D-shotgun meta-predictor (groups 223 and 224) and Baker's ROBETTA server (group 29) did better than most human predictors.

In a way, the method used by the most successful servers mirrors that used by the most successful human predic-

	20	425	517	453	448	427	110	6	28	100	223	153	265	224	12	67	450	368	29	583
20	**48**	19	47	48	14	48	47	48	48	41	43	48	48	43	48	46	46	15	48	14
425	-0.1	**19**	18	19	4.0	19	19	19	19	18	19	19	19	19	19	19	19	7	19	3
517	-0.4	-0.7	**50**	50	17	50	49	50	50	41	45	50	50	45	50	48	46	15	50	14
453	-0.9	-0.5	0.0	**51**	17	51	50	51	51	42	46	51	51	46	51	49	47	15	51	14
448	-2.1	-4.0	-0.7	-0.7	**17**	17	17	17	17	9	15	17	17	15	17	16	13	2	17	2
427	-2.2	-3.0	-1.3	-1.3	0.0	**51**	50	51	51	42	46	51	51	46	51	49	47	15	51	14
110	-2.2	-2.4	-1.7	-1.7	-1.9	-0.4	**50**	50	50	41	45	50	50	45	50	49	46	15	50	14
6	-2.6	-3.4	-1.7	-1.7	-1.5	-0.4	0.1	**51**	51	42	46	51	51	46	51	49	47	15	51	14
28	-3.2	-4.4	-2.3	-2.2	-4.3	-1.0	-0.6	-0.6	**51**	42	46	51	51	46	51	49	47	15	51	14
100	-2.5	-4.4	-1.7	-1.7	-4.0	-0.9	-0.3	-0.7	-0.5	**42**	38	42	42	38	42	41	41	13	42	13
223	-5.1	-6.3	-4.1	-4.0	-3.9	-2.9	-2.3	-2.6	-1.9	-1.6	**46**	46	46	46	46	45	42	13	46	9
153	-3.7	-3.0	-3.0	-3.0	-4.9	-1.8	-1.4	-1.3	-0.8	-0.6	0.7	**51**	51	46	51	49	47	15	51	14
265	-4.4	-4.0	-4.6	-4.5	-11.5	-3.3	-3.0	-2.9	-2.3	-0.2	0.7	-1.5	**51**	46	51	49	47	15	51	14
224	-4.7	-6.4	-3.6	-3.6	-1.3	-2.4	-1.8	-2.1	-1.4	-1.5	0.5	-0.2	-0.2	**46**	46	45	42	13	46	9
12	-4.2	-4.2	-3.4	-3.4	-4.8	-2.1	-1.8	-1.7	-1.1	-0.7	0.3	-0.4	1.2	-0.2	**51**	49	47	15	51	14
67	-4.1	-5.2	-3.7	-3.6	-2.9	-2.6	-1.9	-2.3	-1.7	-1.4	0.4	-0.5	-0.5	-0.1	-0.1	**49**	45	14	49	13
450	-4.9	-5.1	-4.4	-4.1	-8.4	-2.5	-2.8	-2.3	-1.8	-1.2	-0.7	-1.9	0.6	-0.6	-0.6	-0.3	**46**	15	46	14
368	-3.9	-5.1	-3.2	-3.2	-4.6	-1.9	-1.3	-1.5	-0.9	-0.1	0.4	-0.1	1.4	-0.1	0.2	0.5	1.1	**15**	51	14
29	-4.9	-5.1	-4.1	-4.0	-2.7	-2.8	-1.8	-2.4	-1.8	-2.4	0.1	-1.0	0.5	-0.4	-0.6	0.4	-0.1	18.0	**51**	14
583	-0.7	-0.6	-0.7	-0.3	-1.7	1.0	-0.3	1.5	1.2	-0.8	3.4	1.5	5.5	5.4	0.4	1.7	0.7	22.9	2.4	**14**

	425	453	517	20	448	6	28	110	427	96	12	583	329	100	2	224	228	265	223	450
425	19	19	19	19	4	19	19	19	19	19	19	19	19	19	19	18	18	19	14	19
453	-1.5	51	51	48	17	51	51	50	51	51	51	51	46	45	51	43	43	51	39	47
517	-2.2	0.0	51	48	17	51	51	50	51	51	51	51	46	45	51	43	43	51	39	47
20	-0.3	-0.1	-0.7	48	14	48	48	47	48	48	48	48	43	44	48	40	40	48	36	46
448	-3.9	-0.3	0.2	-3.3	17	17	17	17	17	17	17	17	15	12	17	15	15	17	14	13.0
6	-4.1	-1.0	-0.9	1.0	0.1	51	51	50	51	51	51	51	46	45	51	43	43	51	39	47
28	-5.6	-2.3	-2.3	-2.3	-4.7	-1.4	51	50	51	51	51	51	46	45	51	43	43	51	39	47
110	-3.2	-2.4	-2.5	-1.8	-2.1	-1.5	-0.3	49	49	49	49	49	46	43	49	41	41	49	37	45
427	-5.3	-2.6	-2.6	-2.2	-2.8	-1.7	-0.3	-0.1	51	51	51	51	46	45	51	43	43	51	39	47
96	-5.2	-3.3	-3.3	-2.7	-5.2	-2.3	-1.0	-0.6	-0.7	51	51	51	46	45	51	43	43	51	39	47
12	-5.8	-3.4	-3.4		-4.6	-2.4	-1.0	-1.0	-0.8	-0.1	51	51.0	46.0	45	51	43	43	51	39	47
583	9.6	1.7	1.6	1.0	9.7	4.6	4.0	1.6	4.1	4.1	3.2	14	14.0	14	14	8	8	14	6	14
329	-4.6	-3.5	-3.2	-4.2	-6.0	-2.3	-1.3	-2.1	-0.9	-0.4	-0.8	-3.2	45	41	45	37	37	45	33	42
100	-2.6	-3.0	-3.2	-2.6	-13.6	-1.9	-0.7	-1.0	-0.7	-0.4	-0.7	-3.2	-1.0	42	42	35	35	42	31	41
2	-6.1	-3.8	-3.8	-3.3	-1.4	-2.8	-1.5	-1.3	-1.2	-0.5	-0.4	-3.8	-0.6	-0.6	51	43	43	51	39	47
224	-8.9	-4.2	-4.1	-3.8	-3.0	-3.5	-1.8	-1.4	-1.3	-0.5	0.0	-4.1	-0.9	-2.6	0.0	42	42	42	37	39
228	-0.7	-7.9	-7.6	-6.2	-9.3	-6.4	-4.8	-4.3	-4.6	-4.0	-3.7	-7.6	0.0	-2.8	-5.4	-2.6	27	28	23	24
265	-5.3	-6.2	-6.2	-5.1	-12.3	-5.3	-3.9	-3.9	-3.6	-2.9	-2.9	-6.2	-1.7	-3.0	-2.4	-1.4	-1.4	49	38	45
223	-9.3	-7.1	-7.2	-6.6	-5.5	-6.7	-4.6	-4.2	-4.2	-3.1	-3.1	-7.2	-2.1	-6.3	-2.7	-2.9	-2.9	-1.4	36	34
450	-7.4	-5.1	-5.3	-4.9	-4.2	-3.9	-2.7	-3.0	-2.6	-2.0	-1.4	-5.3	-1.2	-3.1	-1.2	-2.3	-2.3	1.3	1.7	46

Fig. 2. Paired *t*-test on the distribution of GDT-TS (**a**) and AL0 (**b**) values. The data are shown only for groups submitting models for >10 target domains and ranking among the top 20 groups in GDT-TS (a) or AL0 (b) average score. The diagonal contains the number of target domains predicted by each group. The top right half of the matrix contains the number of common targets predicted by the pair of groups indicated by the specific row and column. The lower left part of the matrix contains the difference of average GDT-TS (a) or AL0 (b) values on the targets predicted by both groups. A cell is shaded if the difference in the distribution of values for the models submitted by the groups is not statistically significant ($p < 0.05$).

tors, in that both combine, or select among, more than one prediction. This is an interesting conclusion in our view, because it might imply that standard comparative modeling techniques are able to reproduce the conformational space of the proteins in the neighborhood of the native structure. Whether the native structure is indeed among them or can be obtained by refining the predictions cannot be concluded by the present analysis. Essentially, no group made a convincing attempt in optimizing the models. We would really like this step of protein structure prediction not to be neglected in future editions of CASP.

Improvement Over Template

The analysis described so far highlights whether a correct template was identified and which percentage of

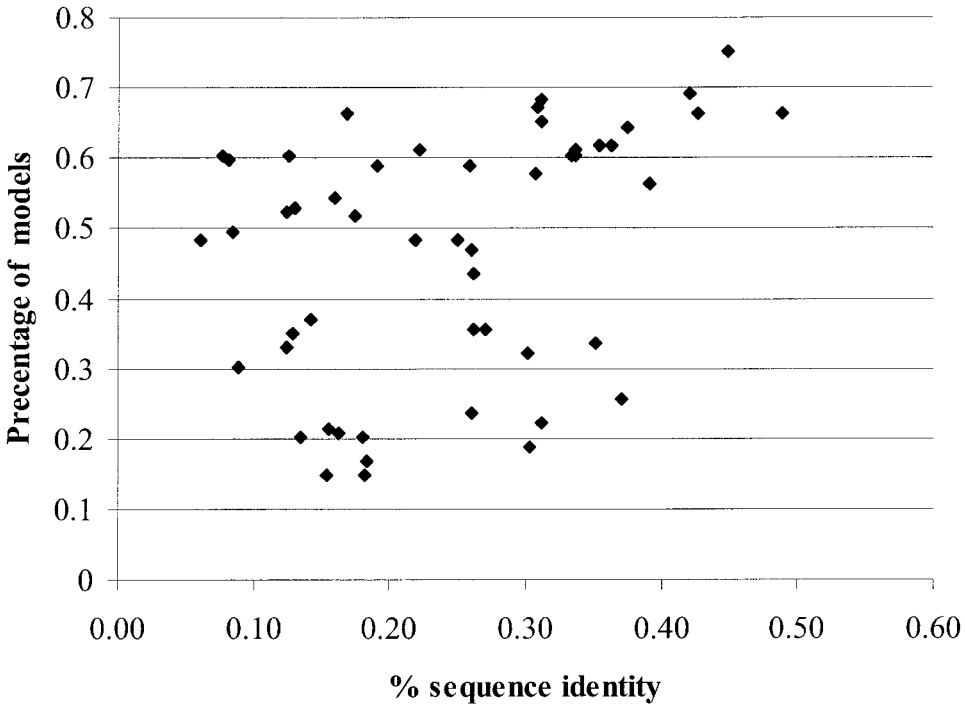

Fig. 3. Identical models. Percentage of models with identical (to the second decimal digit) Cα RMSD in Å with the template as a function of the sequence identity between target and template.

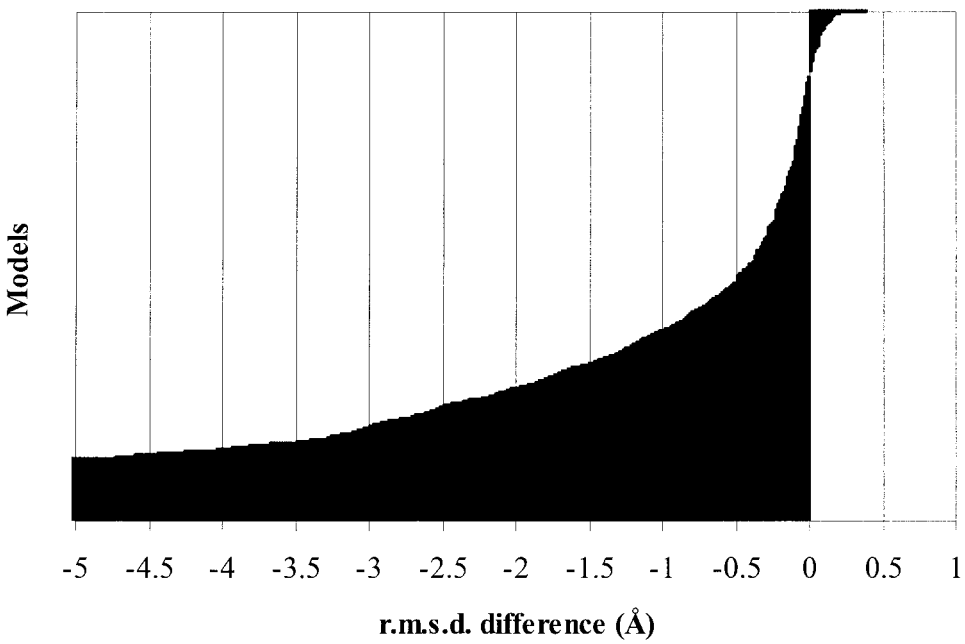

Fig. 4. Improvement over template. Difference between the Cα RMSD of the target-template and target-model superpositions.

the structurally correct correspondence between amino acids of the target and the template was identified.

For difficult models, based on distant evolutionary relationships, this is in itself an achievement, whereas models of proteins clearly homologous to known structures should be able to predict features of the modeled protein not directly inheritable from the parent. A view of the ability of current methods to do so can be obtained by comparing the RMSD between the Cα of selected models and their templates with the RMSD between target and template. The results of this analysis, performed only for the core region (defined as the set of buried residues, i.e., of

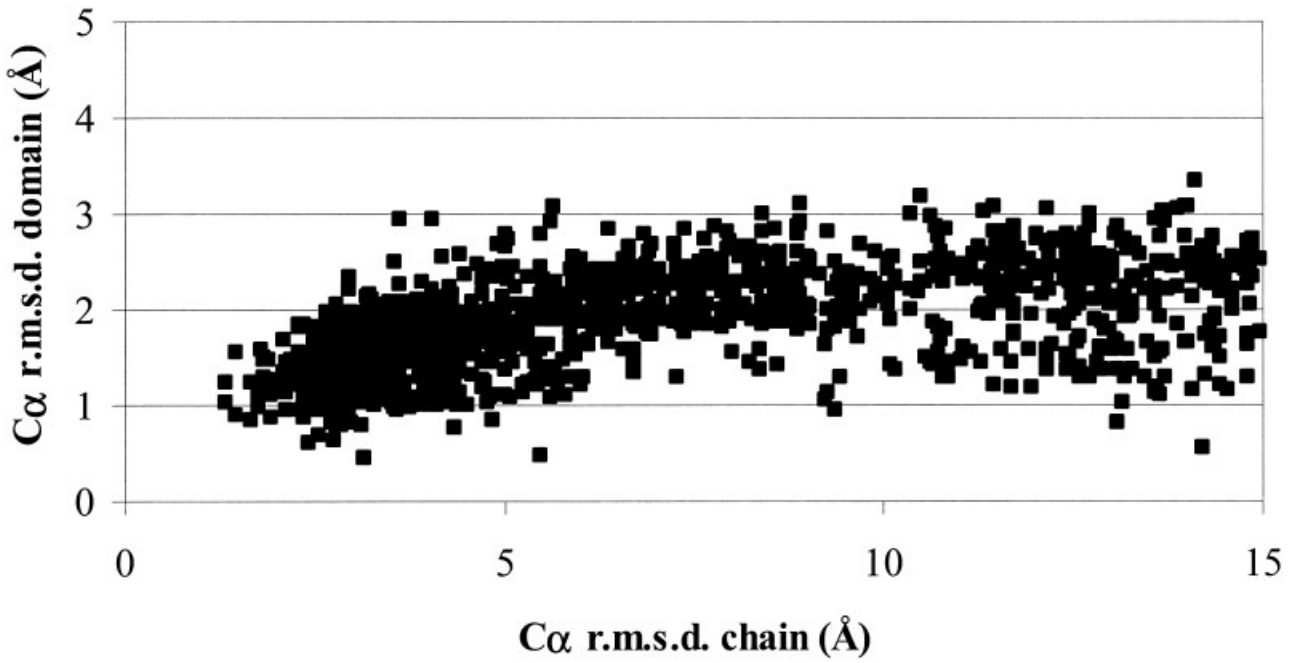

Fig. 5. Domain orientation. Scatterplot of the Cα RMSD between target and predicted structure (*x* axis) versus that of the corresponding domains (*y* axis). Only values < 15 Å are shown.

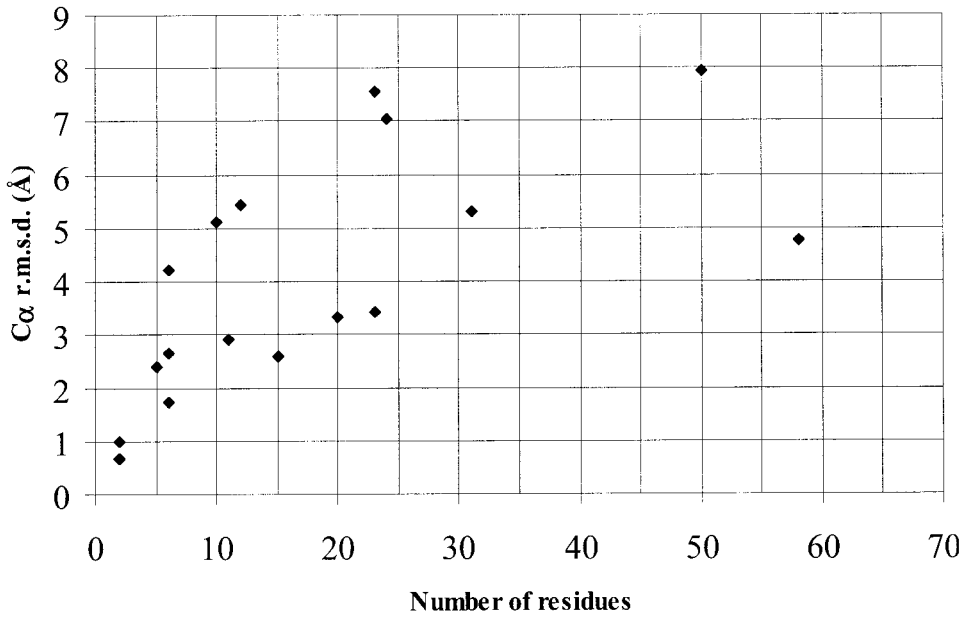

Fig. 6. Prediction of structurally divergent regions. RMSD values are given for the Cα atoms of structurally divergent regions after superposition of the core of models and targets as a function of the total number of SDR residues in each target.

residues whose solvent-accessible surface area is <20%) of selected structures is shown in Figure 4.

We are forced to draw the disappointing conclusion that, similarly to what observed in previous editions of the experiment, no model resulted to be closer to the target structure than the template to any significant extent (the maximum improvement is never >0.4 Å). On the other hand, this has to be expected if no attempt to optimize the starting models is made.

All the data used in the assessment, including the number of correctly aligned residues, are derived from the structural superposition of the model and target structures; therefore, it is technically impossible to separate the effect of errors in sequence alignment from those due to

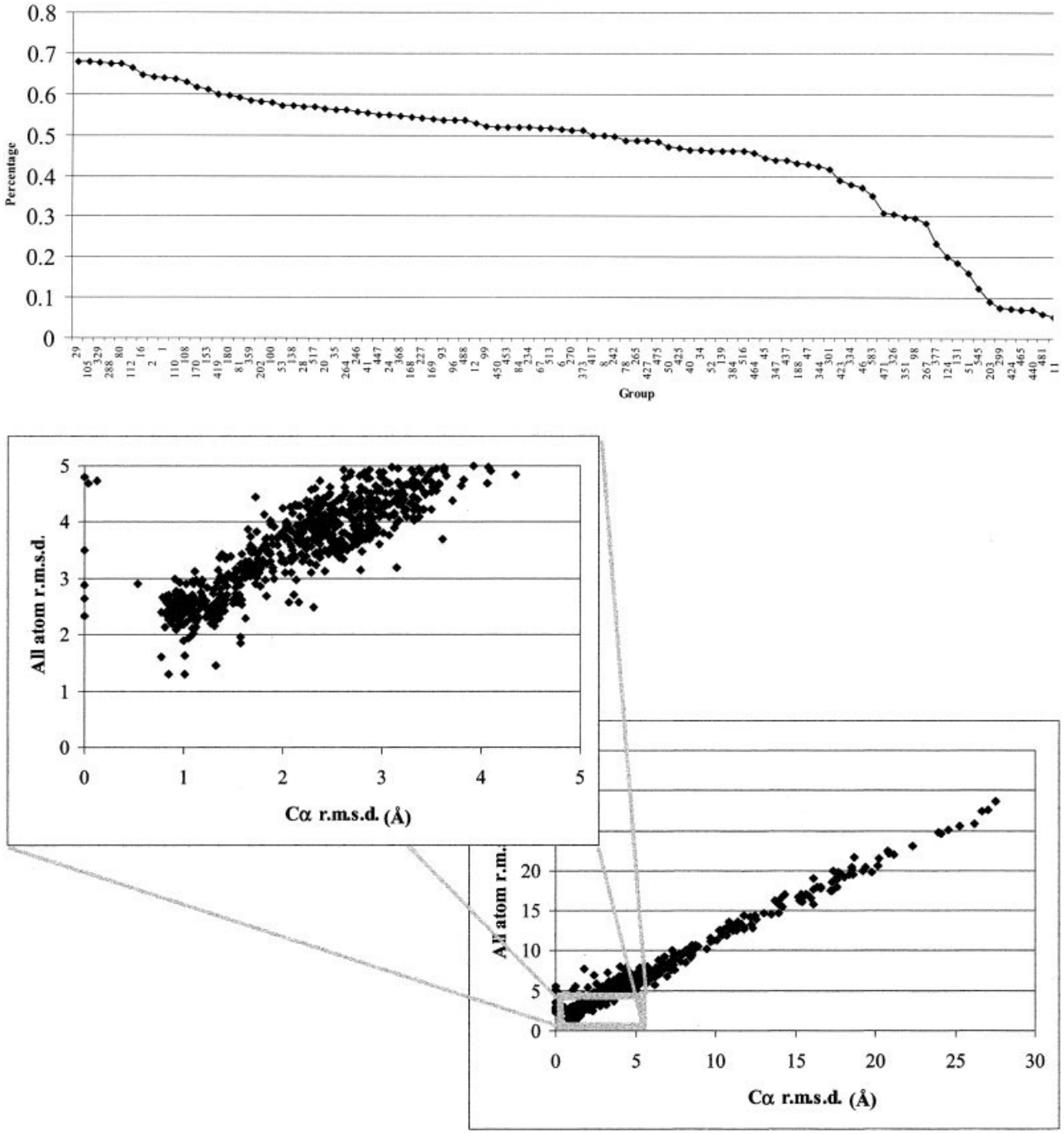

Fig. 7. Prediction of side-chains. **a:** Percentage of correctly assigned χ_1 angles for each group. Angles are considered correct if they are within 30° from the experimental values. **b:** Scatterplot of the Cα RMSD versus all atom RMSD for the selected subset of targets highlighted in Table I.

structural divergence between target and template: However, the results presented in Figure 4 strongly suggest that the former is the most relevant.

In cases in which the target protein is formed by more than one domain and both domains are CM targets, it is possible to compare the quality of the models of the domains with that of the entire chain. The comparison, shown in Figure 5, confirms our previous finding[3] that the prediction of domain orientation is a difficult task, because the models of the complete proteins invariably have a larger Cα RMSD with the target structure than the individual domains.

Structurally Divergent Regions and Side-Chains

It is generally accepted that the prediction of detailed features of a model can meet with some success only when the backbone of the model is predicted with sufficient

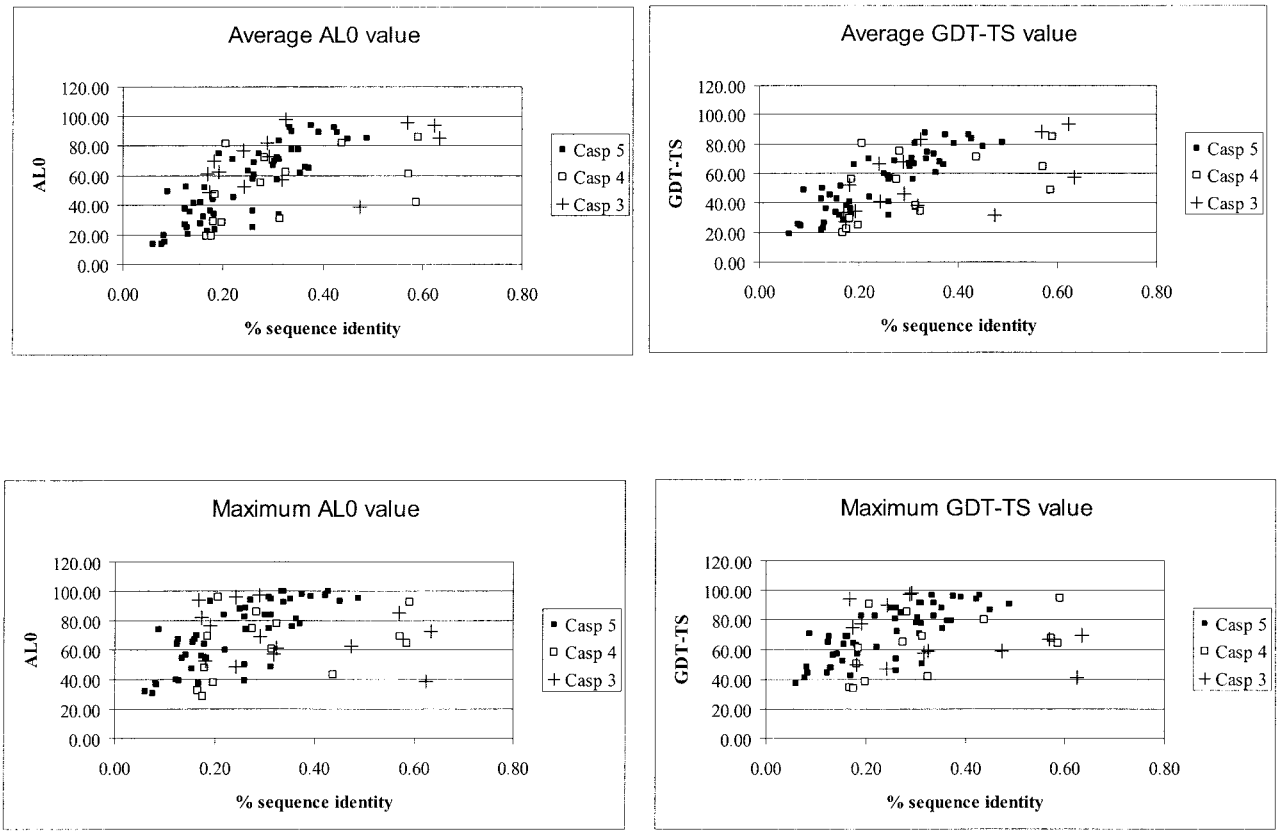

Fig. 8. Comparison between CASPs. Comparison between average and best GDT-TS and AL0 values obtained by the groups participating in CASP3, CASP4, and CASP5 as a function of the sequence identity between target and template.

accuracy. Therefore we selected targets, models, and specific residues according to the following protocol:

- The structure of the target must have been determined by *x*-ray crystallography.
- If more than one molecule is present in the original submission of the structure, the molecules should differ by no more than 0.8 Å RMSD values on their Cα atoms.
- The analyzed models must have a GDT-TS value > 80.
- For the analysis of the side-chains, we only considered buried residues with a B-factor < 40.
- For the analysis of SDRs, we only considered residues with a B-factor < 40 and not involved in crystal packing.

The SDR regions satisfying the above criteria included 304 residues, whereas 758 nonglycine residues from 17 domains were available for side-chain analysis.

The quality of the prediction of structurally divergent regions is much lower than that of the rest of the target molecule. In Figure 6 we show the RMSD between Cα atoms of SDRs for the best prediction for each target, plotted as a function of the number of residues in SDRs for the target. These values are not particularly exciting (only for targets where the number of structurally divergent residues is < 5 the RMSD value is < 1.0 Å), and the *t*-test conducted on the common sets of models shows that there is no significant difference between any pair of groups (data not shown).

The analysis of the side-chain prediction was conducted by first calculating the percent of correctly predicted χ_1 and χ_2 angles [Fig. 7(a)] and subsequently the all-atom RMSD for the selected subset of residues [Fig. 7(b)].

We believe that the percent of correct χ angles is not a sufficient indicator of the quality of side-chain predictions, but it is certainly a necessary one. The first angle is predicted "correctly" in about 80% of the cases and the behavior for χ_2 is very similar. Unfortunately, the low number of side-chains that meet the requirements for inclusion in our analysis does not allow us to derive very sound conclusions about the general features of the correctly and incorrectly predicted angles. Visual inspection shows that different groups predict correctly different subsets of side-chains. We are tempted to speculate that the correct or incorrect positioning of one side-chain is the main factor affecting the quality of the prediction for the neighboring ones; therefore, it is difficult to derive conclusions about the most frequent cause of error or about the expected quality of the prediction of specific subsets of residues. It should also be mentioned here that the difference in the performance of individual groups is not statistically significant (data not shown).

Figure 7(b) shows a plot of the RMSD values for the side-chain atoms as a function of the RMSD value of the corresponding Cα atoms; although it is undoubtedly true that the values are clearly correlated, it is also clear that

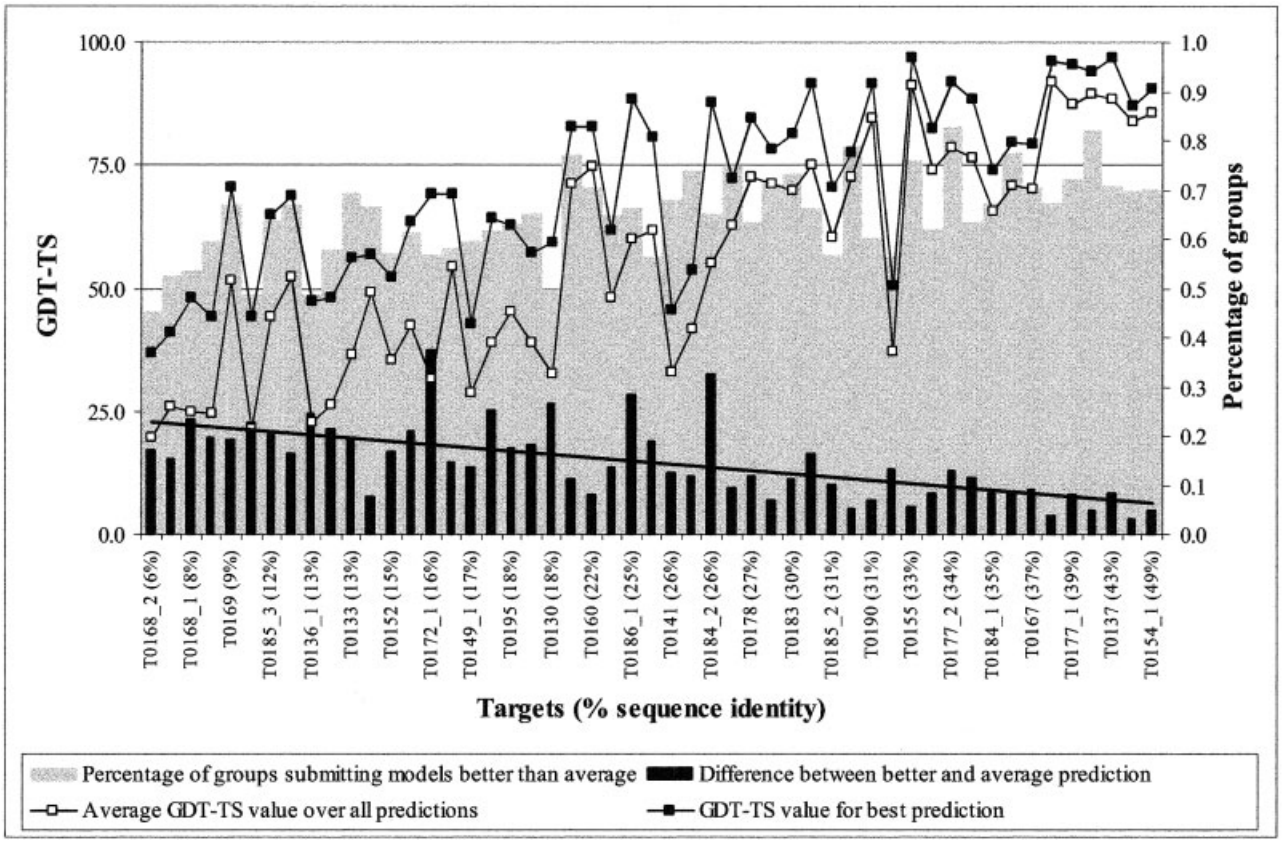

Fig. 9. Quality of the models. GDT-TS values for the best (filled squares) and average (open square) models. The shaded area indicates the percentage of predictions achieving a GDT-TS value better than average (axis on the right). The filled bars are the difference between the best and average GDT-TS value for each target. On the *x* axis, the percentage of sequence identity between target and template is shown in parentheses after the name of the target.

they are not very satisfactory, because even for the buried side-chains of the best predicted structures the all-atom RMSD is rarely < 2.0 Å.

Comparison With Previous Experiments

Figure 8 shows the comparison between the best and the average models produced in the CASP3, CASP4, and CASP5 experiments in terms of GDT-TS and AL0.

The values for the three experiments, plotted as a function of the percent of sequence identity between the target and its closest structural template, show that although the improvement is not striking, the scope of CM methods has certainly increased and that the quality of the predictions has become more constant over the targets and among the groups.

Indeed, a few interesting conclusions can be derived from Figure 9 in which we plotted the maximum and average value of GDT-TS for each target together with the percent of groups submitting models better than average.

When the sequence identity between the target and template is $>25\%$, the difference between the best and the average models is, in most cases, quite modest [Fig. 10(a)–(c)], indicating that the models reflect more the structural similarity between target and template than the specific method used. In addition, the percentage of groups producing models better than average is well above 50% in most cases.

The difference between the best and average model can be quite high for more difficult targets [Fig. 10(d)–(f)] and this is taken to imply that although the detection of distant similarity and the correctness of the alignment in these cases can be difficult, the best groups use superior methods or procedures and we expect these to become more widely available and used in the future.

Biological Implications

The increase in the scope of the method and of its reliability certainly depends on the availability of a larger set of sequence and structure data provided by the experimental biologists to the computational community, but (as it can be seen from the analysis of the data reported in Fig. 9) experimental biologists are receiving a valuable return for their efforts.

Clearly, the quality of a model that can be considered satisfactory depends on the type of applications the model will be used for; therefore, it is difficult to establish a threshold above which a model can be considered of practical use in general. The results presented here once again indicate that rarely a model can be used for applications such as drug design, but, in our experience, it is most

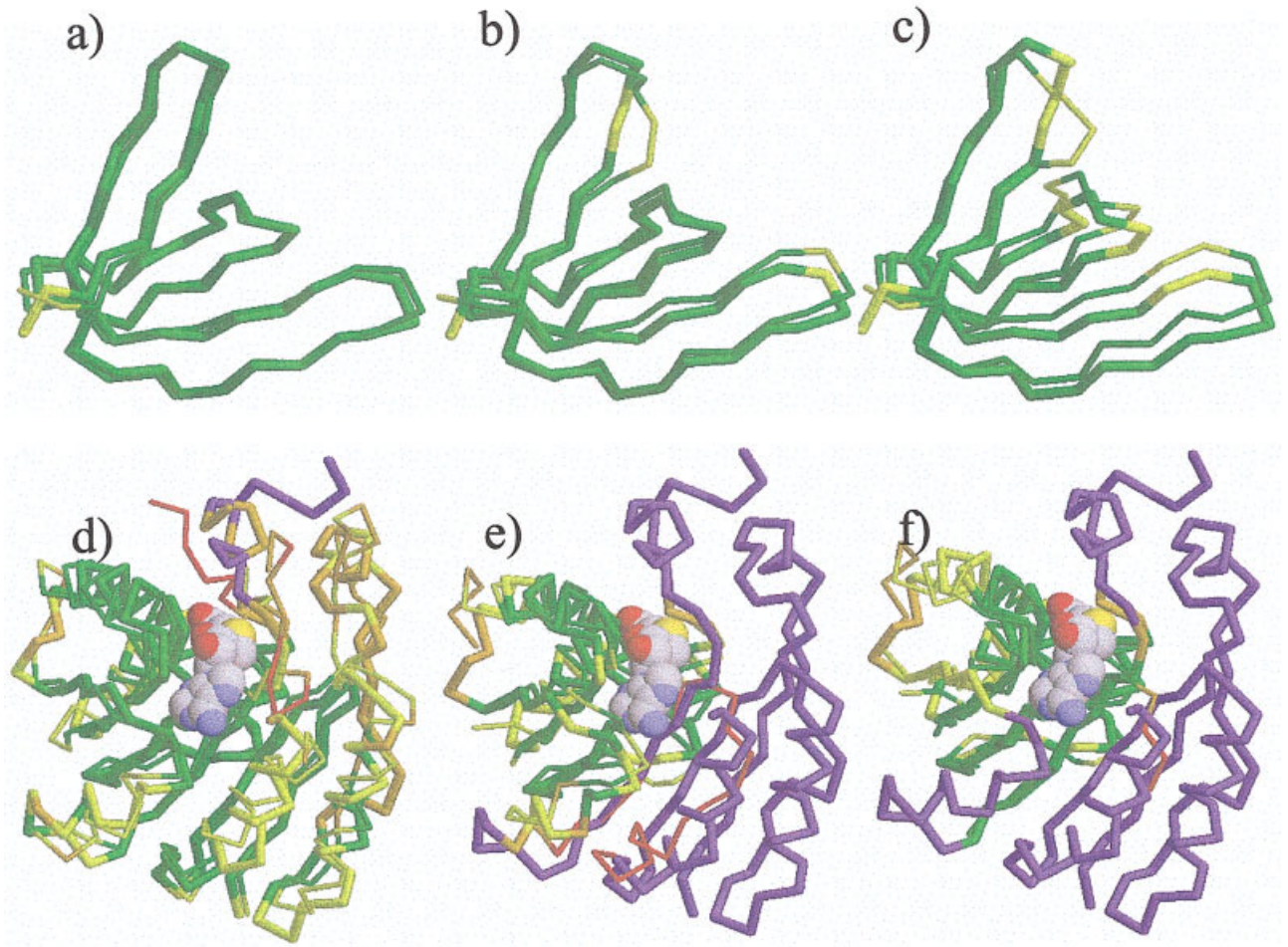

Fig. 10. Examples of the quality of the models. **a–c:** The superposition of three models (thin lines) with the experimental structure (thick lines) of target 179_1 (38% sequence identity with the best template). The model in (a), submitted by group 448, has the highest GDT-TS value; the other models (submitted by groups 270 and 329) have GDT-TS values close to the average and are representative of the quality of most of the predictions. **d:** The best model for target 172_1 (16% sequence identity with the best template) submitted by group 517. The models in **e–f,** submitted by groups 135 and 56, have GDT-TS close to the average and are typical average predictions for this difficult target. Note that in all cases the regions surrounding the ligand, S-adenosyl-L-homocysteine, are well predicted. Regions deviating $<$ 2.0, 4.0, and 8.0 Å from the experimental structure 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.

common for a three-dimensional model to be used to design molecular biology experiments, to explain biochemical differences, and to guide cloning and purification design. Models derived from templates sharing at least 25% sequence identity with the target probably represent most of the cases of practical interest in the laboratory and, in these cases, most methods (and therefore groups) can provide experimentalists with models of respectable quality, which can be instrumental in guiding experimental protocols.

Even in the most difficult cases, although the quality of the models depends more heavily on the methods used and more sophisticated procedures are required to obtain models of overall reasonable quality, most of the methods are still able to predict with reasonable accuracy the functionally important regions of the protein of interest [Fig. 10(d)–(f)].

On the basis of these considerations, we believe that evolutionary based structure prediction methods are sufficiently mature and robust to be added to the suite of tools that computational biology has made available to the biological community.

CONCLUSIONS

The assessment in CASP is meant to highlight what went right, what went wrong, and why. However, we believe that, in the genomic era, we should also address the issue of how much can be expected from structure prediction methods.

Comparative modeling has certainly enlarged its scope and is able to produce models, although only partially accurate, for proteins having very distant relationships with proteins of known structure. This is most likely due to the exploitation of the availability of a larger set of sequences that can be used to "bridge" the sequence distance between target and template.

It is clear that the pairwise sequence identity between target and template is no longer an effective parameter for

describing the expected quality of a model, and other measures should be derived.

Sad notes are once again those regarding the poor performance in predicting features not directly inheritable from the parent and in obtaining a model that is closer to the native structure than the template used to build it. Probably not very comforting for the ego of the predictors might be the observation that the performance of different methods is leveling out, especially for easiest targets (but CASP is not a competition so this is an irrelevant issue!). However, we stress that the latter does indeed represent very good news for the users of comparative modeling methods, who can rely on the fact that most groups, and several publicly available methods, can provide them with models of sufficient quality to guide their experiments and to interpret their results in many cases of practical importance.

From the point of view of computational biologists, we think that CASP5 results are giving some clear hints about the direction in which future efforts should be focused.

Traditional methods for comparative modeling were based on the assumption that each of the modeling steps (template selection, alignment, structurally divergent regions, and side-chain placement) can be optimized separately, whereas it is quite likely that a better approach would be to optimize all the parameters simultaneously. The computational complexity of such an approach is clearly beyond our present capabilities; however, the analysis of the successful predictions in the most recent CASP experiment has provided us with both a hint that this is indeed the road to follow and a method to search for an approximate solution.

Indeed, the most successful groups constructed several models for each target protein and selected the most likely one only at the end of the complete model-building procedure. In other words, rather than optimizing each of the steps of the comparative modeling procedure independently, they chose to funnel into each subsequent step also suboptimal intermediate results and to evaluate several final models at the atomic level. This represents a first-degree approximation to a full multiparameter optimization procedure.

The choice of the final model clearly depends on the method used for assessing the quality of the several alternative ones, and it is clear that in the near future, the efforts of the community should be focused on this aspect of the problem. The ability to select the most likely model out of a set is also relevant for other methods of protein structure prediction, such as fold recognition methods or methods for predicting new folds. We are already witnessing a progressive merging of the different prediction communities, and it can be easily foreseen that they will effectively cross-fertilize each other.

In conclusion, although the field is mature enough to fulfill the expectations raised by structural genomics projects and by a large number of applied projects, more work is required to satisfy the intellectual needs of the community that developed around CASP and is, commendably, testing and challenging itself every 2 years.

ACKNOWLEDGMENTS

We thank the members of the Lawrence Livermore National Laboratory, and especially Adam Zemla, Krzysztof Fidelis, and Ceslovas Venclovas for providing the data for the analysis, John Moult for continuous support, and our fellow assessors for many useful discussions. We are very grateful to the predictors and the structural biologists who provided the targets. Finally, we thank the colleagues of our department for putting up with us during the frantic days of CASP assessment.

REFERENCES

1. Browne WJ, North AC, Phillips DC, Brew K, Vanaman TC, Hill RL. A possible three-dimensional structure of bovine alpha-lactalbumin based on that of hen's egg-white lysozyme. J Mol Biol 1969;1:65–86.
2. Chothia C, Lesk A. The relation between the divergence of sequence and structure in proteins. EMBO J 1986;4:823–826.
3. Tramontano A, Leplae R, Morea V. Analysis and assessment of comparative modeling predictions in CASP4. Proteins 2001;Suppl 5:22–38.
4. Kinch LN, Qi Y, Hubbard TJP, Grishin NV. CASP 5 target classification. Proteins 2003;Suppl 6:340–351.
5. Zemla A. CASP 5 target analysis. Proteins 2003;Suppl 6.
6. Marti-Renom M, Fiser A, Rost B, Sali A. Reliability of assessment of protein structure prediction methods. Structure 2002;3:435–440.
7. Moult J, Fidelis K, Zemla A, Hubbard T, Tramontano A. The significance of performance ranking in CASP5, response to Marti-Renom et al. Structure 2002;3:291–292.
8. Fischer D. CAFASP. Proteins 2003;Suppl 6:503–516.