WILEY InterScience® DISCOVER SOMETHING GREAT

PROTEINS STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Template Based Assessment

Assessment of CASP7 predictions for template-based modeling targets

Jürgen Kopp,[1,2] Lorenza Bordoli,[1,2] James N.D. Battey,[1,2] Florian Kiefer,[1,2] and Torsten Schwede[1,2]*

[1] Biozentrum, University of Basel, Switzerland

[2] Swiss Institute of Bioinformatics, Basel, Switzerland

ABSTRACT

This manuscript presents the assessment of the template-based modeling category of the seventh Critical Assessment of Techniques for Protein Structure Prediction (CASP7). The accuracy of predicted protein models for 108 target domains was assessed based on a detailed comparison between the experimental and predicted structures. The assessment was performed using numerical measures for backbone and structural alignment accuracy, and by scoring correctly modeled hydrogen bond interactions in the predictions. Based on these criteria, our statistical analysis identified a number of groups whose predictions were on average significantly more accurate. Furthermore, the predictions for six target proteins were evaluated for the accuracy of their modeled cofactor binding sites. We also assessed the ability of predictors to improve over the best available single template structure, which showed that the best groups produced models closer to the target structure than the best single template for a significant number of targets. In addition, we assessed the accuracy of the error estimates (local confidence values) assigned to predictions on a per residue basis. Finally, we discuss some general conclusions about the state of the art of template-based modeling methods and their usefulness for practical applications.

Proteins 2007; 69(Suppl 8):38–56.

***Key words:** CASP; template based protein modeling; comparative modeling; model quality assessment.*

INTRODUCTION

Protein structure modeling and prediction has gained significant interest in the biological research community for its ability to provide structural models for proteins lacking experimental structures. Template-based protein models, which exploit the evolutionary relationship between a target protein and others with known experimental structures, have been used successfully in a variety of applications, such as studying the effect of mutations, designing site-directed mutagenesis experiments, predicting binding sites, and docking small molecules in structure-based drug discovery. A variety of such modeling methods have been published over the last years. As the usefulness of a protein structure model depends on the accuracy of the prediction, it is crucial to identify the most suitable method for the task at hand from amongst this growing list of resources.

The Critical Assessment of Techniques for Protein Structure Prediction (CASP)[1] provides an objective evaluation of current prediction methods. In identifying their strengths and weaknesses this experiment serves two purposes. For biologists, this assessment aids in choosing the most suitable methods to meet their needs. For researchers engaged in the development of protein structure prediction techniques, detailed scrutiny of their methods and comparison with other approaches helps pinpoint strengths and limitations, and serves as a guide for future development. To ensure objectivity, CASP is organized as a double-blind prediction experiment, i.e. at

The Supplementary Material referred to in this article can be found online at http://www.interscience.wiley.com/jpages/0887-3585/suppmat/

The authors state no conflict of interest.

*Correspondence to: Torsten Schwede, Klingelbergstrasse 50-70, 4056 Basel, Switzerland. E-mail: Torsten.Schwede@unibas.ch

Received 10 April 2007; Revised 23 July 2007; Accepted 26 July 2007

Published online 25 September 2007 in Wiley InterScience (www.interscience.wiley.com).

DOI: 10.1002/prot.21753

the time of the experiment, the predictors do not know the target structures, and the identity of the predictors is hidden from the assessors. At the end of the experiment, all predictions and assessment data are made publicly available. Accurate and appropriate assessment of protein structure prediction is not a simple standard procedure. While the main numerical assessment criteria have been well established over the series of CASPs,[1–4] progress and convergence in the prediction methods over subsequent CASP experiments require assessors to update existing criteria or even introduce new ones. This ensures that the assessment adequately appraises the overall quality of the models, as well as those features of the predictions that are relevant to their usefulness in specific scientific applications.

The assessment of the template-based modeling (TBM) category in CASP7 greatly benefits from having a broad basis for numerical and statistical analysis. With most participating groups having submitted predictions for the majority of target proteins, a sufficiently large number of diverse targets is available for comparing the various methods and testing the statistical significance of the differences between them. However, we would like to emphasize that this large number of targets, for which predictions had to be made within the 3 months of the prediction season in spring and summer 2006, represented an enormous work load for the participating predictor groups.

Predictions that could largely be built based on template structures were assessed in the TBM category. A subset of the TBM models was additionally assessed in the high accuracy (HA) category with respect to detailed structural features such as side-chain orientation and suitability for application in molecular replacement.[5] In this report, we present the results of our assessment of the models in the TBM category of CASP7 based on numerical criteria for evaluating the correctness of the overall structure and alignment, local residue interactions, accuracy of cofactor binding sites, and improvement over the best templates. Additionally, we evaluated the ability of predictors to correctly assign error estimates as per residue confidence values to their predictions. Finally, we discuss some general conclusions about the state of the art of template based modeling methods and their usefulness for practical applications.

RESULTS AND DISCUSSION

Targets, assessment units, and predictions in the TBM category

For the assessment, prediction target structures were split into assessment units (AU) and classified into three categories: free modeling (FM),[6] template-based modeling (TBM), and high-accuracy template-based models (HA).[5] Assessment units correspond to individual structural domains for single domain proteins. Multidomain proteins, for which the relative orientation could not be inferred from the template structure, were split into separate AUs. Multidomain proteins with the same relative orientation as the template were assessed as a single unit. Definition of assessment units and categorization criteria are discussed in detail elsewhere.[7] Traditionally, the term "target domain" has been used in previous CASP experiments to describe the segment of a prediction target to be assessed individually. However, depending on the context, "domain" denotes quite diverse concepts from an evolutionary, structural, or functional perspective. Therefore, we introduced the term "assessment unit" to describe the segments of a structure on which we based the assessment of 3D structure predictions. However, for historical reasons and for easier readability, the term "domain" might be used interchangeably with "assessment unit" in this manuscript.

In CASP7, the category of TBM comprises 108 out of a total of 123 assessment units, for which 15,717 predictions were submitted by 187 predictor groups. Sixty-eight groups were registered as prediction servers. Among the targets in the TBM category, 28 assessment units were also evaluated in the high-accuracy category (HA), and four overlapped with the definition of free modeling (FM) and were therefore assessed in both categories. The assessment units of the TBM category ranged in size from 526 residues for T0334, a flavin-dependent halogenase from *Nocardia aerocolonigenes*, to 36 residues for T0335, a NMR structure of protein ynzC from *Bacillus subtilis*. In Table I, we summarize the characteristics of the TBM targets. Predictions were received by the Protein Structure Prediction Center at UC Davis (in TS or AL format) and split according to assessment unit definitions. Standard numerical assessment data such as GDT-TS, GDT-HA, AL0, and RMSD values were provided to the assessors by the Protein Structure Prediction Center.

Assessment of the overall quality of the models

Visual inspection of all predictions showed that, like in previous CASP experiments, a significant number of physically impossible models were submitted. Models with more than 2% of the C_α atoms involved in clashes (C_α–C_α distance <1.9 Å) or more than 10% in bumps (C_α–C_α distance between 1.9 and 3.5 Å) and severely fragmented predictions were flagged as physically impossible. In total, 451 models were considered physically impossible. As shown in Figure 1, the majority of these models were submitted by only a few groups, while the distribution over assessment units is homogenous. Since three target structures (T0320, T0332, T0378) were methyltransferases containing topological knots,[8] predictions with knots were not penalized, provided that the structures passed the C_α–C_α distance criteria outlined before.

Numerical assessment in recent CASPs has been based on the two well-established criteria, GDT and AL0. AL0 is defined as the percentage of correctly aligned residues

Table I
TBM Assessment Units in CASP7

Target	Residues	UniProt	Description	Best template	LGA-S	Seq. id
T0283	97	Q9K5V7	JCSG target 10176605, BH3980 protein, *Bacillus halodurans*	2b2jA	54.0	6.7
T0284	250	Q9HUU1	Member of isocitrate lyase family, *Pseudomonas aeruginosa*	1oqfA	87.8	30.1
T0285	99	n/a	Extracytoplasmic domain from histidine kinase, *Cellvibrio japonicus*	1p0zG	54.6	10.1
T0286	202	A3DDK4	Lipolytic enzyme, *Clostridium thermocellum*	1esd	76.7	20.1
T0288	86	Q9NRD5	PDZ domain of PICK1, *Homo sapiens*	2fneB	93.4	32.5
T0289_1	233	Q9R1T5	Aspartoacylase, *Rattus norvegicus*	1yw4A	55.8	19.4
T0289_2	74			1vdzA	59.5	9.6
T0290	173	Q13427	Peptidyl-prolyl isomerase domain of cyclophilin G, *Homo sapiens*	1c5fM	99.2	60.5
T0291	281	P29320	Epha3 Receptor Tyrosine Kinase and Juxtamembrane Region, *Homo sapiens*	1jpaA	92.8	81.1
T0292_1	77	P51955	Nek2 Centrosomal Kinase, *Homo sapiens*	2bmcF	95.4	33.8
T0292_2	173			2acxA	85.7	28.5
T0293	198	Q86W50	Methyltransferase 10 domain-containing protein, *Homo sapiens*	1nv9A	66.2	20.9
T0295_1	180	Q8ILT8	Dimethyladenosine transferase, *Plasmodium falciparum*	1zq9B	95.2	49.2
T0295_2	95			1zq9A	96.9	41.1
T0297	211	Q97PY9	Putative platelet activating factor, *Streptococcus pneumoniae*	1bwp	72.8	22.6
T0298_1	148	O87014	Putative aspartate-semialdehyde dehydrogenase, *Pseudomonas aeruginosa*	2g17A	82.7	21.2
T0298_2	186			1pquA	88.9	16.6
T0299_1	91	Q97RI5	MCSG target APC80351, *Streptococcus pneumoniae*	2cg8C	68.6	13.7
T0299_2	89			1rjjA	60.2	9.1
T0301_1	200	Q9I5E5	JCSG target np_249484.1, *Pseudomonas aeruginosa*	1w61A	53.4	12.3
T0301_2	191			1w62A	47.0	11.8
T0302	129	Q9NS28	RGS domain of RGS18, *Homo sapiens*	1agrE	90.5	53.9
T0303_1	147	Q0I1W8	Phosphoglycolate phosphatase, *Haemophilus somnus*	2ah5A	89.0	28.4
T0303_2	77			1fezB	81.3	7.1
T0304	101	P76364	YeeU protein, *Escherichia coli*	2gnxA	55.9	8.3
T0305	280	P23470	Tyrosine receptor phosphatase gamma, *Homo sapiens*	2fh7A	95.9	53.1
T0306	95	P0AEJ8	Ethanolamine utilization protein, *Escherichia coli*	1d7qA	55.6	15.8
T0308	165	Q9H0F7	ADP-ribosylation factor-like protein 6, *Homo sapiens*	1o3yB	93.7	41.5
T0311	64	P67699	Antitoxin HigA, *Escherichia coli*	1rpeL	90.0	18.0
T0312	132	O30132	NESG target GR103, Archaeoglobus fulgidus	1xv2B	61.1	13.5
T0313	316	Q9BVG8	Kinesin-like protein KIFC3 motor domain, *Homo sapiens*	1ii6A	88.4	42.1
T0315	253	Q7A1S8	TatD deoxyribonuclease, *Staphylococcus aureus*	1j6oA	94.9	39.0
T0316_1	188	Q97T38	tRNA (5-Methylaminomethyl-2-Thiouridylate)-Methyltransferase TrmU,	1kh3C	51.0	17.4
T0316_3	90		*Streptococcus pneumoniae*	1wb3B	78.1	17.9
T0317	149	Q8BTR5	putative dual specificity phosphatase, *Mus musculus*	2esbA	92.4	34.5
T0318_1	154	P34629	Leucine aminopeptidase, *Caenorhabditis elegans*	1vhuA	38.9	4.4
T0318_2	335			1gytL	88.7	28.7
T0320_1	214	P38913	FAD synthetase, *Saccharomyces cerevisiae*	1sur	55.0	19.3
T0321_1	96	Q18YZ7	JCSG target ZP_00559375.1, *Desulfitobacterium hafniense*	1f9cA	59.3	18.8
T0321_2	148			1kxzE	48.0	8.8
T0322	128	P25734	Colonization factor antigen I subunit E, *Caulobacter crescentus*	2h4uD	85.4	20.8
T0323_1	101	Q9KC25	DNA-3-methyladenine glycosidase, *Bacillus halodurans*	1yqmA	55.7	20.8
T0323_2	116			1dizA	87.8	26.6
T0324_1	142	Q88YA8	Putative phosphoglycolate phosphatase, *Lactobacillus plantarum*	2fdrA	89.6	25.0
T0324_2	65			2ah5A	92.6	4.7
T0325	261	P59745	Protein EF3048, *Enterococcus faecalis*	1v6tA	49.1	16.1
T0326	289	Q9WZY3	Homoserine O-succinyltransferase, *Thermotoga maritima*	2ghrA	84.9	55.2
T0327	73	O31639	YjcQ protein, *Bacillus subtlis*	1lnwF	76.5	13.3
T0328	307	Q8EIU4	Putative melanin biosynthesis protein TyrA, *Shewanella oneidensis*	2gvkA	90.5	30.4
T0329_1	141	Q1GA24	Putative phosphoglycolate phosphatase, *Lactobacillus delbrueckii*	1rdfB	90.9	25.8
T0329_2	92			1rqlA	60.7	11.6
T0330_1	153	Q8KBS5	Haloacid dehalogenase-like hydrolase, *Chlorobium tepidum*	2ah5A	82.9	29.2
T0330_2	72			1lvhB	73.5	7.9
T0331	139	Q2ZZ07	Pyridoxamine 5'-phosphate oxidase-related protein, *Streptococcus suis*	1ty9A	72.3	16.4
T0332	153	Q13395	Methyltransferase Domain of Human TAR (HIV-1) RNA binding protein 1, *Homo sapiens*	1zjrA	88.8	22.9
T0333_1	206	Q8KND7	CalG3, *Micromonospora echinospora*	1rvvB	48.9	15.7
T0333_2	148			1rvvB	69.2	25.6
T0334	526	Q8KHZ8	Flavin-dependent halogenase, *Nocardia aerocolonigenes*	2ajqA	94.4	56.2
T0335	36	O31818	YnzC protein, *Bacillus subtilis*	1yluA	96.7	0.0
T0338_1	143	O60583	Cyclin T2, *Homo sapiens*	1jkw	74.6	20.3
T0338_2	113			1n4mA	60.9	10.0

(*Continued*)

DOI 10.1002/prot

Table I
Continued

Target	Residues	UniProt	Description	Best template	LGA-S	Seq. id
T0339_1	136	Q7L670	Selenocysteine lyase, *Homo sapiens*	1eg5B	78.3	26.1
T0339_2	267			1eg5A	87.3	37.3
T0340	82	Q15599	Second PDZ domain of human NHERF-2, *Homo sapiens*	1g9oA	98.0	59.8
T0341_1	148	Q6PEB2	Haloacid dehalogenase-like hydrolase domain containing protein, *Mus musculus*	1zjjB	90.0	27.3
T0341_2	104			1wviB	93.8	21.4
T0342	122	O75223	Protein LOC79017, *Homo sapiens*	2g0qA	80.4	21.5
T0345	185	P18283	Glutathionine peroxidase 2, *Homo sapiens*	1gp1A	97.0	68.1
T0346	172	P30414	Peptidylprolyl isomerase domain of the human NK-tumour recognition protein, *Homo sapiens*	2gw2A	99.9	71.5
T0347_1	89	Q8UF59	Protein Atu1540, Agrobacterium tumefaciens	1vk1A	74.3	21.4
T0348	61	Q7NSS5	Putative Tetraacyldisaccharide-1-P 4-kinase, *Chromobacterium violaceum*	1rfs	58.1	16.2
T0349	57	Q6NAY9	Protein RPA1041, *Pseudomonas aeruginosa*	1yj7D	87.4	17.3
T0351	56	P54342	Phage-like element PBSX protein xkdW, *Bacillus subtilis*	1cs1C	62.8	4.6
T0354	122	Q7P0P8	Protein CV0518, *Chromobacterium violaceum*	2be3A	56.8	8.1
T0356_2	192	P0AAB4	3-octaprenyl-4-hydroxybenzoate decarboxylase, *Escherichia coli*	1ejeA	45.6	9.6
T0357	132	O30181	NESG Target GR101, *Archaeoglobus fulgidus*	1aco	56.3	13.5
T0358	65	P75677	Protein ykfF, *Escherichia coli*	1dgd	60.5	13.0
T0359	90	O75970	3rd PDZ domain of multiple pdz domain protein MPDZ, *Homo sapiens*	2bygA	92.9	31.0
T0360	97	Q9JY98	Protein NMB1681, *Neisseria meningitidis*	1dvoA	76.6	21.0
T0362	144	Q7V4A7	JCSG target NP_895880.1, *Prochlorococcus marinus*	2gf6A	82.9	22.7
T0363	46	Q4QNE7	NYSGRC 68057197, *Haemophilus influenzae*	2bb6A	79.9	13.2
T0364	147	Q88R33	JCSG target NP_742468.1, *Pseudomonas putida*	2av9B	79.7	14.3
T0365	207	Q8EAX1	JCSG target NP_719307.1, *Shewanella oneidensis*	1xwmA	67.0	13.1
T0366	84	O75970	12th PDZ domain of multiple pdz domain protein MPDZ, *Homo sapiens*	2fneB	93.5	26.8
T0367	123	O29944	JCSG target NP_069135.1, *Archaeoglobus fulgidus*	1ufbC	87.1	15.0
T0368	157	Q8KAL8	JCSG target NP_663012.1, *Chlorobium tepidum*	2c2lC	69.4	17.0
T0369	147	Q41IB9	JCSG target ZP_00537729.1, *Exiguobacterium sibiricum*	1rxqA	61.8	11.4
T0370	144	Q825J7	JCSG target NP_828636.1, *Streptomyces avermitilis*	1vl7B	75.4	18.8
T0371_1	162	Q11S56	Puatative HAD superfamily sugar phosphatase, *Cytophaga hutchinsonii*	1vjrA	80.7	27.7
T0371_2	121			1zjjA	65.6	20.2
T0372_1	126	Q8A1H2	Protein BT_3689, *Bacteroides thetaiotaomicron*	1ro5A	59.2	4.0
T0372_2	172			1xf8A	60.1	9.1
T0373	140	Q9HZE1	Putative transcriptional regulator protein, *Pseudomonas aeruginosa*	1s3jB	70.6	23.8
T0374	160	Q9HV14	Putative acetyltransferase, *Pseudomonas aeruginosa*	1tiqA	75.2	9.4
T0375	296	P50053	Ketohexokinase, *Homo sapiens*	2fv7B	67.5	18.9
T0376	306	Q8U6Y1	Dihydrodipicolinate synthase, *Agrobacterium tumefaciens*	1xxxB	81.3	25.0
T0378_1	89	Q7MW92	Putative RNA methyltransferase of the TrmH family, *Porphyromonas gingivalis*	1ipaA	74.1	22.0
T0378_2	142			1gz0C	89.6	27.4
T0379_1	140	Q7MWA6	Putative HAD-like family hydrolase, *Porphyromonas gingivalis*	1zd5A	83.9	27.8
T0379_2	64			2b0cA	67.2	25.5
T0380	142	Q97DI6	Pyridoxinephosphate oxidase family-related protein, *Clostridium acetobutylicum*	2fhqA	82.7	25.8
T0381_1	61	Q0SH23	Putative transcriptional regulator RHA06195, *Rhodococcus sp. RHA1*	2g7uA	99.5	45.9
T0381_2	176			2g7uD	92.7	39.8
T0382	119	Q6N8L4	MCSG target APC6185, *Rhodopseudomonas palustris*	1kpsB	56.8	9.8
T0383	125	Q97PP5	Protein SP_1558, *Streptococcus pneumoniae*	1qynB	63.2	11.8
T0384	301	Q97PV8	Gfo/Idh/MocA family Oxidoreductase, *Streptococcus pneumoniae*	1ydwA	84.2	22.2
T0385	125	O05815	Protein rv2844, *Mycobacterium tuberculosis*	1jgcB	87.2	10.9
T0386_1	206	Q6G2A9	Putative cell filamentation protein, *Bartonella henselae*	2g03A	63.8	24.5

in the 5 Å LGA sequence-independent superposition of the model and experimental structure of the target. A model residue is considered to be correctly aligned if the predicted C_α atom position falls within 3.8 Å of the corresponding experimental atom, and there is no other C_α atom of the experimental structure nearer. GDT (global distance test) identifies sets of residues in the predictions deviating from the target by not more than a specified C_α distance cutoff for different sequence-dependent superpositions, e.g. using distance cut-off values of 1, 2, 4, and 8 Å for GDT-TS calculation. It was suggested during the CASP6 meeting in Gaeta that the cut-off values applied to calculate GDT-TS may not be appropriate to detect small differences in backbone quality.[2] In our assessment, we considered the upper cut-off value of 8 Å as too lenient to discriminate the finer structural differences between models of template-based predictions and therefore decided to use GDT-HA with distance cut-off values of 0.5, 1, 2, and 4 Å in our evaluation.

Although GDT is a sequence-dependent and AL0 a sequence-independent measure, both scores are highly correlated and on average contain little complementary

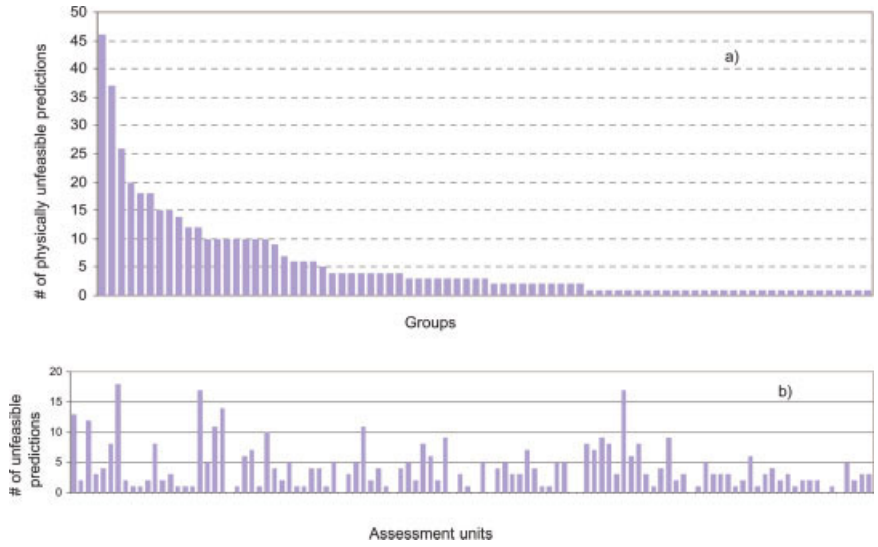

Figure 1

Distribution of physically impossible models. The majority of unfeasible models were submitted by only a few groups (a), while the distribution over assessment units is homogenous (b). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

information for comparing and contrasting the different prediction methods (Fig. 2). As both GDT and AL0 are derived from global superpositions of C_α coordinates only, they do not reflect important local structural features of a protein such as backbone geometry, packing of amino acid side-chains, and atomic interactions like hydrogen bonds or hydrophobic contacts. To complement these global C_α-based criteria, we introduced a local atomic measure termed HBscore. It counts the intersection of corresponding hydrogen bonds present in the model and the target structure: HBscore = number of (H-bonds in model $\cap$ H-bonds in target structure)/number of H-bonds in target structure. We excluded hydrogen bonds involving side-chain atoms of residues with more than 50% relative surface exposure in the target structure. In the predicted structures, hydrogen bonds were not considered if they involved amino acid residues with incorrect topology or had severe clashes with neighboring residues ($d < 1.2$ Å). Hydrogen bonds were calculated using HBPlus,[9] and relative solvent accessibility of side-chains using NACCESS (Hubbard and Thornton, 1993). The HBscore enumeration of specific H-bond interactions accounts for ambiguities arising from chemically equivalent side-chain atoms being assigned different atom names in IUPAC nomenclature (e.g., Glu OE1, OE2; Arg NH1, NH2, etc.). Figure 3 illustrates HBscore for the example of a short β-sheet. When comparing structure predictions and actual experimental structures, high scores in global criteria such as GDT and AL0 are necessary, but not by themselves sufficient indicators of accuracy. Local criteria based on specific atomic interactions provide complementary information, as illustrated in Figure 4.

Numerical evaluation and statistical significance of the results

The assessment of the individual groups was based on the predictions submitted as "Model 1." The majority of groups predicted more than one hundred assessment units and consequently, for each target more than 100

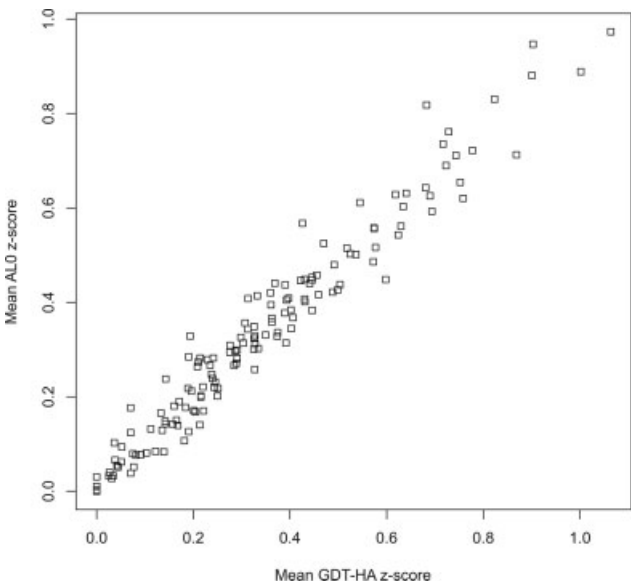

Figure 2

Correlation of GDT-HA and AL0 z-scores for groups in CASP7.

DOI 10.1002/prot

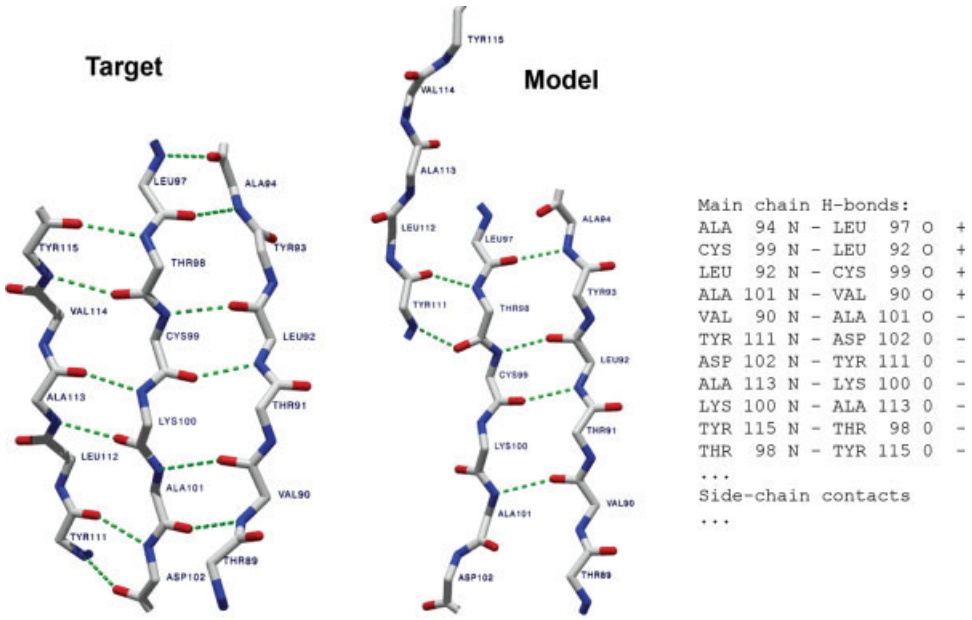

Figure 3

Illustration of HBscore on the example of a short β-sheet structure.

individual predictions were available for the assessment. For groups which submitted both unrefined and refined models, the assessment was based on the refined prediction. If predictions for a target were submitted in several fragments, the segment with the longest overlap with the assessment unit was assessed.

The scoring scheme adopted in our assessment was similar to the one used in previous CASP experiments.[2,3] To allow the comparison of the results for targets of different modeling difficulty, we computed *z*-scores for both GDT-HA and AL0 for each assessment unit in the following way: (i) For each assessment unit, average values and standard deviations for all predictions were calculated; (ii) For those predictions, whose scores were not more than two standard deviations below average and which were not flagged as physically impossible, we recomputed the means and standard deviations, and used these to assign *z*-scores to all predictions; (iii) Models

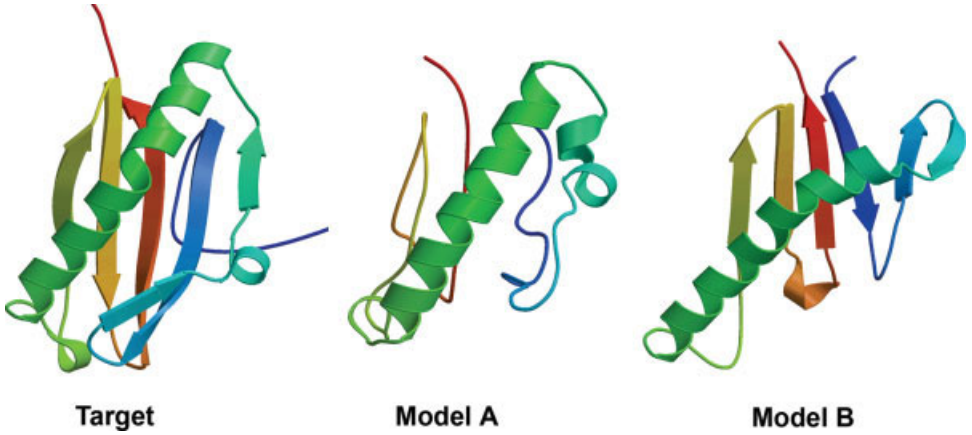

Figure 4

Complementary information assessed by GDT-TS and HBscore. While model A better resembles the global positioning of Cα atoms of the target structure (GDT-TS: 46) compared to model B (GDT-TS: 35), model B better reflects the H-bonding pattern in the central β-sheet structure (HBscore: 41) compared to model A (HBscore: 26).

Table II
Mean Values and z-Scores for Individual Prediction Groups

Group		Group name	Number of predictions	Mean GDT-HA	Mean AL0	Mean GDT-HA *z*-score	Mean AL0 *z*-score	Combined GDT-HA and AL0 *z*-score
4	s	ROBETTA	107	46.77	60.14	0.54	0.50	0.52
5		luethy	108	47.83	65.46	0.68	0.82	0.75
9		CBiS	4	5.72	0.00	0.00	0.00	n.a.
10		SAM-T06	108	44.78	56.85	0.40	0.41	0.40
11		Dlakic-MSU	10	53.24	68.94	0.26	0.17	n.a.
13		Jones-UCL	107	47.19	62.24	0.52	0.50	0.51
15		Advanced-ONIZUKA	35	20.68	10.43	0.22	0.17	0.20
16		AMBER/PB	1	54.26	88.37	0.00	0.15	n.a.
18		LUO	97	45.31	58.99	0.52	0.52	0.52
20		Baker	106	50.19	65.21	1.00	0.89	0.95
21		karypis	83	39.02	49.23	0.37	0.34	0.36
22	s	karypis.srv	106	38.28	48.87	0.23	0.27	0.25
24		Zhang	107	51.49	67.87	1.06	0.97	1.02
25	s	Zhang-Server	108	50.35	66.60	0.90	0.88	0.89
26		SAMUDRALA	106	47.87	61.55	0.69	0.63	0.66
27		SAMUDRALA-AB	106	46.81	60.09	0.57	0.56	0.57
28	s	PROTINFO	103	44.34	55.99	0.33	0.31	0.32
29	s	PROTINFO-AB	106	42.61	53.91	0.31	0.36	0.33
30		TsaiLab	52	45.86	60.48	0.22	0.22	0.22
31		Avbelj	7	15.72	0.32	0.00	0.00	n.a.
33		POEM-REFINE	19	28.66	25.75	0.47	0.43	n.a.
34		ROKKO	105	43.73	56.30	0.40	0.38	0.39
35	s	ROKKY	106	42.94	53.56	0.28	0.31	0.29
38		GeneSilico	102	48.34	63.61	0.72	0.74	0.73
40		YASARA	23	56.05	74.35	0.36	0.42	0.39
43		hu	2	54.75	69.09	0.00	0.00	n.a.
44	s	gtg	56	38.85	49.75	0.10	0.08	0.09
45		INFSRUCT	1	10.17	0.00	0.00	0.00	n.a.
46	s	Pcons6	108	46.70	60.70	0.50	0.44	0.47
47	s	Pmodeller6	108	46.98	61.06	0.58	0.52	0.55
50		SBC	105	49.05	65.18	0.78	0.72	0.75
54		PROTEO	62	8.00	1.29	0.00	0.00	0.00
60		HIT-ITNLP	104	27.98	32.74	0.04	0.05	0.05
62		Floudas	21	26.24	19.08	0.19	0.13	0.16
63		FEIG	106	36.80	48.75	0.14	0.24	0.19
64		LMM-Bicocca	29	42.37	50.29	0.21	0.14	0.18
65		Protofold	2	9.64	0.00	0.00	0.00	n.a.
66		UF_GATORS	4	16.74	18.15	0.00	0.00	n.a.
69	s	panther2	78	33.10	41.68	0.05	0.06	0.06
71		Wymore	45	40.34	51.48	0.13	0.17	0.15
74		SHORTLE	91	45.97	58.28	0.36	0.37	0.36
78		Dill-ZAP	5	26.52	10.12	0.01	0.03	n.a.
83	s	LOOPP	108	41.89	51.37	0.25	0.23	0.24
87		Pan	108	42.91	52.91	0.30	0.31	0.31
91		Ma-OPUS	107	44.82	56.86	0.37	0.33	0.35
92	s	Ma-OPUS-server	108	43.17	53.69	0.35	0.33	0.34
102	s	Huber-Torda-Server	102	39.58	47.65	0.20	0.17	0.19
103		Huber-Torda	107	42.46	52.14	0.25	0.20	0.23
105		andante	106	47.28	60.12	0.63	0.56	0.60
109		Cracow.pl	40	12.15	1.03	0.00	0.00	0.00
111		panther	82	49.27	67.44	0.21	0.26	0.24
113		Bates	108	47.41	62.36	0.63	0.60	0.62
121		Peter-G-Wolynes	24	17.35	9.26	0.07	0.18	0.12
125		TASSER	108	49.89	66.68	0.90	0.95	0.92
132		Softberry	102	39.60	51.29	0.19	0.28	0.24
135		CBSU	108	43.34	54.48	0.33	0.35	0.34
136	s	FOLDpro	108	44.97	56.46	0.45	0.38	0.41
137	s	3Dpro	107	45.42	57.02	0.49	0.42	0.45
139	s	ABIpro	107	14.38	6.56	0.03	0.03	0.03
168	s	Distill	108	26.34	30.57	0.04	0.07	0.05
170		LMU	78	44.99	55.92	0.12	0.08	0.10
174		Bystroff	58	28.71	30.35	0.08	0.08	0.08

(*Continued*)

DOI 10.1002/prot

Table II
Continued

Group		Group name	Number of predictions	Mean GDT-HA	Mean AL0	Mean GDT-HA *z*-score	Mean AL0 *z*-score	Combined GDT-HA and AL0 *z*-score
178		Bilab	108	42.85	53.83	0.33	0.26	0.29
179	s	Bilab-ENABLE	107	41.43	52.19	0.25	0.22	0.23
186	s	Casplta-FOX	107	41.32	51.76	0.22	0.20	0.21
191		Schomburg-group	22	56.25	75.99	0.57	0.49	0.53
193	s	karypis.srv.4	91	9.47	1.38	0.00	0.00	0.00
194		Scheraga	34	16.15	4.80	0.03	0.04	0.03
197		MTUNIC	103	27.91	33.33	0.07	0.12	0.10
203		forecast	103	34.68	41.59	0.16	0.14	0.15
205		NanoModel	108	40.31	51.31	0.21	0.27	0.24
208		Nano3D	63	42.51	53.89	0.29	0.30	0.29
209		NanoDesign	89	46.10	58.70	0.33	0.30	0.32
211		KIST	103	41.55	53.03	0.23	0.28	0.25
212	s	HHpred1	108	47.00	61.55	0.57	0.56	0.57
213	s	HHpred2	108	48.43	62.25	0.76	0.62	0.69
214	s	BayesHH	108	47.40	61.21	0.62	0.54	0.58
224		ricardo	4	38.99	54.45	0.66	0.72	n.a.
226		Struct-Pred-Course	2	43.86	60.91	0.10	0.00	n.a.
234		McCormack_Okazaki	10	40.74	48.92	0.23	0.21	n.a.
239	s	nFOLD	108	41.66	51.07	0.22	0.20	0.21
242	s	FUGUE	105	42.46	53.02	0.20	0.17	0.19
243	s	UNI-EID_sfst	104	46.33	61.00	0.39	0.40	0.40
245	s	UNI-EID_expm	107	47.17	61.82	0.33	0.32	0.32
247	s	3D-JIGSAW_POPULUS	104	39.50	48.07	0.14	0.15	0.15
248	s	RAPTOR	108	45.09	58.04	0.43	0.45	0.44
249		taylor	39	27.66	27.81	0.17	0.14	0.15
250		fleil	77	39.91	52.24	0.14	0.13	0.13
252		EAtorP	15	12.66	2.33	0.00	0.00	n.a.
257	s	FORTE1	108	37.65	46.62	0.14	0.14	0.14
261	s	mGen-3D	107	43.66	55.49	0.33	0.30	0.31
263		igor	45	13.60	3.54	0.00	0.03	0.02
267	s	RAPTOR-ACE	108	45.93	60.01	0.46	0.46	0.46
268	s	karypis.srv.2	108	37.12	46.51	0.19	0.22	0.20
273		BioDec	70	35.80	47.84	0.04	0.10	0.07
274	s	shub	107	44.95	58.57	0.33	0.41	0.37
275	s	beautshot	108	45.55	59.45	0.39	0.44	0.41
276		keasar	107	44.08	59.57	0.37	0.44	0.40
277	s	keasar-server	101	41.79	55.30	0.19	0.33	0.26
278		Pushchino	4	16.43	14.92	0.02	0.03	n.a.
284		Oka	4	23.78	28.93	0.02	0.03	n.a.
297		MLee	101	42.50	52.95	0.31	0.34	0.33
298	s	CIRCLE	108	46.60	61.17	0.47	0.52	0.50
302	s	3D-JIGSAW	104	38.36	46.66	0.08	0.08	0.08
304	s	panther3	16	33.54	42.68	0.03	0.08	n.a.
307	s	MetaTasser	108	45.44	60.61	0.62	0.63	0.62
316	s	FORTE2	108	37.17	45.94	0.17	0.15	0.16
318	s	FUNCTION	107	45.01	57.68	0.33	0.33	0.33
319	s	FUGMOD	100	42.99	54.14	0.24	0.22	0.23
333	s	forecast-s	98	36.48	45.83	0.16	0.18	0.17
337		AMU-Biology	98	45.76	57.80	0.41	0.37	0.39
338		UCB-SHI	103	47.20	59.59	0.50	0.43	0.46
347	s	beautshotbase	106	46.30	58.70	0.40	0.34	0.37
349	s	FAMSD	108	46.27	59.31	0.45	0.45	0.45
351	s	FAMS	108	46.21	60.03	0.44	0.44	0.44
361		Doshisha-Nagoya	7	17.90	3.82	0.00	0.00	n.a.
368	s	Frankenstein	56	35.92	42.73	0.14	0.08	0.11
380	s	SAM-T99	87	47.82	63.20	0.28	0.27	0.28
381	s	SAM-T02	104	43.17	55.00	0.24	0.24	0.24
383	s	UNI-EID_bnmx	108	46.29	60.22	0.49	0.48	0.49
389	s	SAM_T06_server	108	42.28	52.38	0.29	0.27	0.28
393		Distill_human	108	26.41	30.35	0.04	0.05	0.05
397		Tripos-Cambridge	10	58.42	74.17	0.31	0.23	n.a.
401		MIG	90	39.88	46.71	0.18	0.11	0.14
413	s	SPARKS2	108	45.10	57.62	0.39	0.38	0.38

(*Continued*)

Table II
Continued

Group		Group name	Number of predictions	Mean GDT-HA	Mean AL0	Mean GDT-HA *z*-score	Mean AL0 *z*-score	Combined GDT-HA and AL0 *z*-score
414	s	SP3	108	46.07	59.13	0.46	0.42	0.44
415	s	SP4	108	45.44	58.31	0.43	0.40	0.42
416		honiglab	100	46.57	60.06	0.42	0.45	0.43
418	s	HHpred3	108	47.81	61.45	0.69	0.59	0.64
420	s	3D-JIGSAW_RECOM	104	39.32	47.35	0.09	0.08	0.08
427		CADCMLAB	102	21.78	18.45	0.03	0.03	0.03
435	s	RAPTORESS	108	43.70	57.44	0.31	0.41	0.36
437		osgdj	11	15.81	5.53	0.00	0.00	n.a.
439		Sternberg	107	45.85	60.11	0.45	0.45	0.45
443		fais	79	35.67	43.46	0.21	0.28	0.25
453		Deane	18	17.99	14.93	0.09	0.09	n.a.
464	s	POMYSL	52	11.69	1.70	0.00	0.01	0.01
468	s	Phyre-1	104	41.08	52.90	0.20	0.21	0.20
469	s	Phyre-2	107	43.19	55.77	0.30	0.33	0.31
474		PUT_lab	72	28.73	28.74	0.08	0.05	0.06
483		Hirst-Nottingham	13	15.35	2.20	0.00	0.00	n.a.
490		Iwyrwicz	106	43.49	55.20	0.29	0.30	0.30
494	s	CPHmodels	60	47.06	61.96	0.11	0.13	0.12
495		largo	2	52.36	76.88	1.02	1.08	n.a.
501		Bristol_Comp_Bio	4	60.84	79.70	0.06	0.06	n.a.
509		SEZERMAN	66	30.18	32.69	0.03	0.03	0.03
511	s	FPSOLVER-SERVER	103	9.67	0.81	0.00	0.00	0.00
527		chaos	16	37.19	47.85	0.07	0.21	n.a.
536		Chen-Tan-Kihara	103	44.08	55.72	0.36	0.36	0.36
550		ZIB-THESEUS	94	29.90	30.93	0.05	0.09	0.07
556		LEE	106	49.16	62.17	0.87	0.71	0.79
559		GSK-CCMM	4	66.42	87.97	0.45	0.49	n.a.
564		ShakSkol-AbInitio	13	28.86	29.12	0.53	0.57	n.a.
568		CHIMERA	107	49.21	64.85	0.73	0.76	0.74
586	s	MIG_FROST	47	29.20	28.49	0.07	0.04	0.05
588	s	MIG_FROST_FLEX	2	38.56	44.13	0.32	0.08	n.a.
599		KORO	31	21.48	15.14	0.43	0.57	0.50
601		LTB-WARSAW	86	43.29	55.09	0.28	0.29	0.28
609	s	GeneSilicoMetaServer	100	46.57	59.99	0.43	0.41	0.42
610		Dlakic-DGSA	3	51.11	73.71	0.00	0.22	n.a.
614		Brooks_caspr	21	48.72	63.11	0.60	0.45	0.52
638		Soeding	1	29.93	36.36	1.36	1.29	n.a.
640		jive	99	40.61	51.98	0.24	0.28	0.26
641		tlbgroup	14	51.54	68.97	0.18	0.24	n.a.
650		Schulten	15	43.25	52.10	0.43	0.36	n.a.
651		verify	108	48.35	63.41	0.68	0.64	0.66
654	s	NN_PUT_lab	103	43.00	53.86	0.24	0.25	0.24
658		hPredGrp	107	49.15	64.19	0.72	0.69	0.71
659		CHEN-WENDY	32	63.93	81.46	0.39	0.31	0.35
664		CIRCLE-FAMS	108	48.95	64.51	0.74	0.71	0.73
671		fams-multi	108	48.51	62.92	0.64	0.63	0.64
673		ProteinShop	6	18.32	2.74	0.00	0.18	n.a.
675		fams-ace	108	49.64	65.79	0.82	0.83	0.83
677		UAM ICO BIB	96	41.95	52.80	0.36	0.39	0.38
683		MUMSSP	15	64.17	82.80	0.34	0.38	n.a.
698		MQAP-Consensus	108	49.04	64.01	0.75	0.65	0.70
705		Akagi	101	36.96	46.28	0.17	0.19	0.18
706		TENETA	106	39.39	49.01	0.18	0.18	0.18
710		Ligand-Circle	94	46.13	59.29	0.54	0.61	0.58
721		ROBETTA-late	3	33.48	44.76	0.35	0.40	n.a.
728	s	Ma-OPUS-server2	71	42.47	52.02	0.29	0.28	0.29
735		EBGM	13	29.60	36.03	0.01	0.05	n.a.
736		dokhlab	18	31.10	29.24	0.13	0.16	n.a.
746		CDAC	4	13.25	0.82	0.00	0.00	n.a.
757		SSU	16	22.33	13.34	0.11	0.17	n.a.
781		SCFBio-IITD	2	27.14	0.00	0.00	0.00	n.a.
794		MerzShak	4	27.22	27.86	0.70	0.70	n.a.

n.a.: Groups with less than 20 predictions were not included in the final ranking.

DOI 10.1002/prot

Table III

Statistical Significance of the Results of the 25 Highest Scoring Groups

	24	20	125	25	675	556	5	50	568	664	38	658	698	213	651	26	418	671	307	113	105	214	710	212	27
24		G: 0.74(0.24) A: 1.73(0.12) H:-**6.87(0.00)**	G: **1.46(0.00)** A: 1.23(0.15) H: 3.79(0.00)	G: 1.10(0.00) A: **1.34(0.00)** H:-0.38(0.50)	G: 1.73(0.00) A: **1.69(0.02)** H: 2.77(0.00)	G: 1.62(1.00) A: 4.08(0.00) H:-**3.50(0.00)**	G: 2.75(0.00) A: 1.77(0.01) H: 2.06(0.00)	G: 2.25(0.00) A: 3.24(0.00) H:-**2.15(0.00)**	G: 2.09(0.00) A: 2.56(0.00) H: 0.58(0.36)	G: 2.31(0.00) A: 3.13(0.00) H:-**1.71(0.02)**	G: 2.71(0.00) A: 3.45(0.00) H:-**3.34(0.00)**	G: 2.38(0.00) A: 3.37(0.00) H: 1.00(0.16)	G: 2.15(0.00) A: 3.55(0.00) H:-0.30(0.63)	G: 2.33(0.00) A: 4.38(0.00) H:-0.36(0.67)	G: 2.73(0.00) A: 3.75(0.00) H: 0.23(0.73)	G: 2.73(0.00) A: 4.69(0.00) H:-**1.80(0.02)**	G: 2.81(0.00) A: 4.83(0.00) H: 0.20(0.83)	G: 2.88(0.00) A: 4.44(0.00) H: 2.34(0.00)	G: 3.30(0.00) A: 4.36(0.00) H: 4.88(0.00)	G: 3.30(0.00) A: 4.64(0.00) H:-1.36(1.00)	G: 3.44(0.00) A: 6.22(0.00) H:-0.13(0.87)	G: 3.11(0.00) A: 5.30(0.00) H:-0.39(0.65)	G: 4.37(0.00) A: 5.61(0.00) H: 3.35(0.00)	G: 3.48(0.00) A: 5.19(0.00) H: 0.66(0.27)	G: 3.67(0.00) A: 5.66(0.00) H: 2.06(0.01)
20	n = 105		G: 0.66(0.32) A:-0.58(0.67) H: 10.68(0.00)	G: 0.37(0.59) A:-0.43(0.71) H: 8.50(0.00)	G: 0.65(0.31) A:-0.57(0.59) H: 9.89(0.00)	G: 0.84(0.28) A: 2.28(0.09) H: 3.38(0.00)	G: 1.96(0.00) A:-0.10(0.93) H: 5.96(0.00)	G: 1.44(0.03) A: 1.44(0.21) H: 4.60(0.00)	G: 1.35(0.05) A: 0.88(0.51) H: 7.55(0.00)	G: 1.58(0.03) A: 1.40(0.26) H: 5.27(0.00)	G: 1.97(0.00) A: 1.73(0.14) H: 3.49(0.00)	G: 1.44(0.04) A: 1.25(0.37) H: 7.68(0.00)	G: 1.36(0.03) A: 1.55(0.19) H: 6.55(0.00)	G: 1.60(0.04) A: 2.70(0.04) H: 6.69(0.00)	G: 1.95(0.00) A: 1.84(0.12) H: 7.13(0.00)	G: 1.96(0.00) A: 2.89(0.02) H: 5.09(0.00)	G: 2.08(0.01) A: 3.16(0.01) H: 7.25(0.00)	G: 2.22(0.00) A: 3.01(0.02) H: 9.40(0.00)	G: 2.53(0.00) A: 1.82(0.13) H: 11.72(0.00)	G: 2.54(0.00) A: 2.83(0.02) H: 5.52(0.00)	G: 2.68(0.00) A: 4.45(0.00) H: 6.70(0.00)	G: 2.37(0.00) A: 3.56(0.01) H: 6.91(0.00)	G: 3.16(0.00) A: 3.48(0.01) H: 10.39(0.00)	G: 2.83(0.00) A: 3.54(0.00) H: 8.13(0.00)	G: 2.84(0.00) A: 3.96(0.00) H: 8.90(0.00)
125	n = 106	n = 105		G:-0.32(0.48) A: 0.10(0.91) H:-**4.25(0.00)**	G: 0.13(0.76) A: 0.41(0.58) H:-**1.14(0.04)**	G: 0.23(0.87) A: 2.86(0.00) H:-**7.37(0.00)**	G: 1.24(0.00) A: 0.48(0.57) H:-**1.75(0.01)**	G: 0.67(0.17) A: 1.75(0.05) H:-**5.86(0.00)**	G: 0.82(0.20) A: 1.35(0.09) H:-**3.09(0.00)**	G: 0.87(0.08) A: 1.93(0.02) H:-**5.49(0.00)**	G: 1.37(0.01) A: 2.35(0.01) H:-**7.12(0.00)**	G: 0.98(0.08) A: 2.52(0.00) H:-**2.74(0.00)**	G: 0.64(0.21) A: 2.31(0.01) H:-**4.17(0.00)**	G: 0.80(0.17) A: 3.14(0.00) H:-**4.19(0.00)**	G: 1.32(0.01) A: 2.45(0.01) H:-**3.40(0.00)**	G: 1.36(0.01) A: 3.49(0.00) H:-**5.60(0.00)**	G: 1.27(0.03) A: 3.59(0.00) H:-**3.63(0.00)**	G: 1.40(0.01) A: 3.18(0.00) H:-1.43(0.06)	G: 1.87(0.00) A: 3.08(0.00) H: 0.92(0.05)	G: 1.81(0.00) A: 3.31(0.00) H:-**5.22(0.00)**	G: 2.00(0.00) A: 5.03(0.00) H:-**4.08(0.00)**	G: 1.49(0.01) A: 3.76(0.00) H:-**4.02(0.00)**	G: 2.04(0.00) A: 4.36(0.00) H:-0.60(0.40)	G: 2.34(0.00) A: 4.66(0.00) H:-**2.76(0.00)**	G: 2.34(0.00) A: 4.66(0.00) H:-**1.74(0.01)**
25	n = 107	n = 106	n = 107		G: 0.61(0.14) A: 0.36(0.63) H: 3.26(0.00)	G: 0.47(0.39) A: 2.71(0.00) H:-**3.13(0.00)**	G: 1.63(0.00) A: 0.42(0.51) H: 2.46(0.00)	G: 1.14(0.02) A: 1.76(0.04) H:-1.36(0.05)	G: 0.99(0.01) A: 1.23(0.12) H: 0.96(0.16)	G: 1.20(0.00) A: 1.80(0.02) H:-1.26(0.11)	G: 1.57(0.00) A: 2.20(0.01) H:-**2.93(0.00)**	G: 1.38(0.00) A: 2.29(0.01) H: 1.54(0.00)	G: 1.07(0.01) A: 2.20(0.00) H: 0.12(0.84)	G: 1.19(0.03) A: 3.05(0.00) H: 0.06(0.94)	G: 1.64(0.00) A: 2.42(0.00) H: 0.72(0.24)	G: 1.59(0.00) A: 3.32(0.00) H:-1.42(0.06)	G: 1.66(0.00) A: 3.50(0.00) H: 0.62(0.47)	G: 1.67(0.00) A: 3.06(0.00) H: 2.61(0.00)	G: 2.19(0.00) A: 3.01(0.00) H: 5.44(0.00)	G: 2.19(0.00) A: 3.19(0.00) H:-0.96(0.19)	G: 2.31(0.00) A: 4.87(0.00) H: 0.20(0.81)	G: 1.92(0.00) A: 3.91(0.00) H: 0.26(0.76)	G: 3.40(0.00) A: 4.49(0.00) H: 4.23(0.00)	G: 2.40(0.00) A: 3.91(0.00) H: 1.44(0.08)	G: 2.58(0.00) A: 4.40(0.00) H: 2.34(0.00)
675	n = 106	n = 105	n = 106	n = 107		G:-0.16(0.74) A: 2.22(0.01) H:-**6.53(0.00)**	G: 1.11(0.01) A:-0.06(0.91) H: 0.63(0.15)	G: 0.70(0.12) A: 1.70(0.02) H:-**4.78(0.00)**	G: 0.34(0.43) A: 0.76(0.31) H:-**2.26(0.00)**	G: 0.59(0.20) A: 1.31(0.10) H:-**4.53(0.00)**	G: 1.15(0.02) A: 2.11(0.01) H:-**6.20(0.00)**	G: 0.75(0.12) A: 1.77(0.05) H:-**1.89(0.00)**	G: 0.54(0.25) A: 1.92(0.01) H:-**3.18(0.00)**	G: 0.68(0.23) A: 2.58(0.00) H:-**3.25(0.00)**	G: 1.10(0.02) A: 2.08(0.01) H:-**2.60(0.00)**	G: 1.06(0.02) A: 3.14(0.00) H:-**4.77(0.00)**	G: 1.06(0.03) A: 3.03(0.00) H:-**2.66(0.00)**	G: 1.09(0.02) A: 2.52(0.00) H:-0.58(0.35)	G: 1.59(0.00) A: 2.72(0.00) H: 2.09(0.00)	G: 1.76(0.00) A: 3.17(0.00) H:-**4.28(0.00)**	G: 1.71(0.00) A: 4.45(0.00) H:-**3.15(0.00)**	G: 1.15(0.02) A: 3.10(0.00) H:-**2.68(0.00)**	G: 2.47(0.00) A: 3.52(0.00) H: 0.31(0.63)	G: 1.81(0.00) A: 3.45(0.00) H:-**1.86(0.01)**	G: 2.07(0.00) A: 4.28(0.00) H:-0.93(0.18)
556	n = 105	n = 106	n = 105	n = 106	n = 105		G: 1.12(0.03) A:-**2.38(0.01)** H: 5.58(0.00)	G: 0.68(0.21) A:-0.94(0.36) H: 1.47(0.06)	G: 0.50(0.30) A:-**1.49(0.04)** H: 4.19(0.00)	G: 0.75(0.14) A:-0.88(0.34) H: 1.90(0.02)	G: 1.03(0.03) A:-0.73(0.45) H: 0.21(0.79)	G: 0.92(0.04) A:-0.48(0.51) H: 4.73(0.00)	G: 0.62(0.26) A:-0.51(0.59) H: 3.23(0.00)	G: 0.76(0.15) A: 0.42(0.68) H: 3.31(0.00)	G: 1.13(0.03) A:-0.44(0.63) H: 3.82(0.00)	G: 1.13(0.03) A: 0.61(0.50) H: 1.71(0.03)	G: 1.24(0.02) A: 0.88(0.39) H: 3.88(0.00)	G: 1.24(0.02) A: 0.37(0.81) H: 5.95(0.00)	G: 1.24(0.02) A: 0.46(0.55) H: 8.85(0.00)	G: 1.69(0.00) A: 0.55(0.58) H: 2.15(0.02)	G: 1.71(0.01) A: 2.16(0.00) H: 3.33(0.00)	G: 1.84(0.00) A: 1.05(0.24) H: 3.38(0.00)	G: 1.34(0.01) A: 1.71(0.13) H: 7.39(0.00)	G: 2.72(0.00) A: 1.26(0.22) H: 4.76(0.00)	G: 1.59(0.00) A: 1.73(0.09) H: 5.59(0.00)
5	n = 107	n = 106	n = 107	n = 108	n = 107	n = 106		G:-0.49(0.30) A: 1.27(0.16) H:-**4.06(0.00)**	G:-0.67(0.11) A: 0.79(0.28) H:-**1.48(0.02)**	G:-0.43(0.35) A: 1.38(0.08) H:-**3.72(0.00)**	G: 0.06(0.90) A: 1.86(0.04) H:-**5.50(0.00)**	G:-0.25(0.62) A: 1.99(0.03) H:-0.97(0.16)	G:-0.57(0.16) A: 1.77(0.02) H:-**2.36(0.00)**	G:-0.44(0.42) A: 2.63(0.01) H:-**2.40(0.00)**	G:-0.01(0.98) A: 1.86(0.01) H:-**1.77(0.01)**	G: 0.01(0.99) A: 2.99(0.01) H:-**3.87(0.00)**	G: 0.03(0.95) A: 3.08(0.00) H:-**1.84(0.01)**	G: 0.10(0.83) A: 2.63(0.00) H: 0.39(0.49)	G: 0.53(0.25) A: 2.86(0.00) H: 2.90(0.00)	G: 0.56(0.22) A: 2.77(0.01) H:-**3.42(0.00)**	G: 0.72(0.16) A: 4.54(0.00) H:-**2.26(0.00)**	G: 0.26(0.63) A: 3.46(0.00) H:-**1.97(0.01)**	G: **1.49(0.01)** A: 4.29(0.00) H: 1.10(0.08)	G: 0.77(0.17) A: 3.49(0.00) H:-1.02(0.12)	G: 0.97(0.08) A: 4.02(0.00) H:-0.06(0.93)
50	n = 104	n = 103	n = 104	n = 105	n = 104	n = 103	n = 105		G:-0.06(0.91) A:-0.32(0.73) H: 2.54(0.00)	G: 0.12(0.80) A: 0.02(0.98) H: 0.36(0.61)	G: 0.55(0.23) A: 0.54(0.48) H:-1.00(0.14)	G: 0.26(0.54) A: 0.61(0.50) H: 3.14(0.00)	G:-0.01(0.98) A: 0.32(0.65) H: 1.81(0.01)	G: 0.06(0.92) A: 1.32(0.19) H: 1.49(0.06)	G: 0.57(0.17) A: 0.56(0.48) H: 2.40(0.00)	G: 0.53(0.32) A: 1.68(0.10) H: 0.06(0.93)	G: 0.52(0.36) A: 1.78(0.10) H: 2.03(0.01)	G: 0.60(0.21) A: 1.55(0.11) H: 4.16(0.00)	G: 0.94(0.06) A: 0.97(0.31) H: 6.87(0.00)	G: 1.16(0.03) A: 1.44(0.18) H: 0.61(0.23)	G: 1.15(0.03) A: 1.97(0.03) H: 1.66(0.03)	G: 0.66(0.22) A: 2.47(0.02) H: 1.82(0.03)	G: 1.94(0.00) A: 2.15(0.04) H: 5.55(0.00)	G: 1.20(0.03) A: 2.59(0.00) H: 2.80(0.00)	G: 1.37(0.02) A: 2.59(0.00) H: 3.78(0.00)
568	n = 107	n = 105	n = 106	n = 107	n = 106	n = 105	n = 107	n = 104		G: 0.22(0.58) A: 0.57(0.49) H:-**2.29(0.00)**	G: 0.52(0.25) A: 0.93(0.36) H:-**4.14(0.00)**	G: 0.37(0.37) A: 1.13(0.10) H: 0.44(0.53)	G: 0.08(0.85) A: 0.96(0.27) H:-0.89(0.22)	G: 0.24(0.66) A: 1.52(0.08) H:-0.94(0.24)	G: 0.63(0.14) A: 1.13(0.18) H:-0.36(0.57)	G: 0.60(0.25) A: **2.09(0.03)** H:-**3.49(0.00)**	G: 0.78(0.08) A: 2.27(0.03) H:-0.39(0.64)	G: 0.78(0.08) A: 1.86(0.02) H: 1.77(0.00)	G: **1.28(0.02)** A: 1.97(0.01) H:-**1.95(0.01)**	G: **1.32(0.02)** A: 1.97(0.04) H:-0.82(0.21)	G: 1.32(0.01) A: 3.53(0.00) H:-0.60(0.48)	G: 1.01(0.07) A: 2.68(0.01) H: 2.70(0.00)	G: 2.29(0.00) A: 3.29(0.01) H: 0.37(0.62)	G: 1.59(0.01) A: 3.25(0.00) H: 1.33(0.08)	G: 1.58(0.01) A: 3.49(0.00) H: 1.55(0.06)
664	n = 107	n = 106	n = 107	n = 108	n = 107	n = 106	n = 106	n = 105	n = 107		G: 0.31(0.50) A: 0.30(0.74) H:-**1.74(0.02)**	G: 0.07(0.87) A: 0.33(0.71) H: 2.74(0.00)	G:-0.14(0.76) A: 0.38(0.61) H: 1.39(0.05)	G:-0.01(0.98) A: 1.25(0.18) H: 1.33(0.12)	G: 0.43(0.33) A: 0.57(0.50) H: **1.86(0.01)**	G: 0.38(0.46) A: 1.50(0.16) H:-0.19(0.81)	G: 0.46(0.39) A: 1.70(0.07) H: **1.85(0.03)**	G: 0.56(0.19) A: 1.45(0.08) H: **4.07(0.00)**	G: **0.99(0.05)** A: 1.37(0.12) H: **6.70(0.00)**	G: **0.99(0.04)** A: 1.39(0.14) H: 0.30(0.64)	G: 1.10(0.02) A: 3.05(0.00) H: **1.43(0.04)**	G: 0.77(0.16) A: 2.22(0.01) H: 1.75(0.05)	G: **1.88(0.00)** A: 2.60(0.01) H: **4.75(0.00)**	G: **1.20(0.01)** A: 2.11(0.02) H: **2.71(0.00)**	G: **1.32(0.02)** A: 2.47(0.02) H: **3.57(0.00)**
38	n = 101	n = 102	n = 101	n = 102	n = 101	n = 102	n = 102	n = 99	n = 101	n = 102		G:-0.16(0.73) A: 0.19(0.85) H: **4.36(0.00)**	G:-0.44(0.34) A:-0.09(1.00) H: 3.01(0.00)	G: 0.43(0.43) A: 0.83(0.41) H: **3.03(0.00)**	G: 0.08(0.87) A: 0.19(0.83) H: **3.52(0.00)**	G: 0.07(0.91) A: 1.22(0.26) H: **1.58(0.03)**	G: 0.09(0.88) A: 1.35(0.20) H: **3.58(0.00)**	G: 0.21(0.63) A: 1.21(0.18) H: **5.66(0.00)**	G: 0.33(0.53) A: 0.32(0.74) H: **8.32(0.00)**	G: 0.64(0.25) A: 1.30(0.21) H: **1.95(0.03)**	G: 0.77(0.07) A: **2.76(0.00)** H: **3.19(0.00)**	G: 0.30(0.60) A: 1.90(0.06) H: **3.30(0.00)**	G: **1.46(0.01)** A: 1.87(0.09) H: **6.45(0.00)**	G: 0.99(0.06) A: 1.87(0.06) H: **4.70(0.00)**	G: 1.07(0.07) A: **2.35(0.04)** H: **5.32(0.00)**
658	n = 104	n = 103	n = 104	n = 105	n = 104	n = 103	n = 105	n = 102	n = 104	n = 105	n = 100		G:-0.26(0.56) A:-0.16(0.85) H:-**1.42(0.02)**	G: 0.04(0.94) A: 0.25(0.19) H:-1.29(0.10)	G: 0.29(0.50) A: 0.11(0.90) H:-0.87(0.18)	G: 0.43(0.37) A: 1.41(0.12) H: **2.71(0.00)**	G: 0.54(0.31) A: 1.82(0.11) H:-0.73(0.37)	G: 0.39(0.40) A: 0.95(0.27) H: 1.19(0.07)	G: 0.86(0.09) A: 0.62(0.29) H: **3.95(0.00)**	G: 0.63(0.09) A: 1.04(0.31) H:-**2.41(0.00)**	G: 0.91(0.06) A: **2.59(0.02)** H:-1.27(0.12)	G: 0.53(0.34) A: 1.67(0.09) H:-1.10(0.25)	G: **1.75(0.00)** A: 2.06(0.09) H: 2.55(0.00)	G: **1.32(0.01)** A: 2.15(0.03) H: 0.13(0.88)	G: **1.42(0.01)** A: 2.35(0.03) H: 1.55(0.23)
698	n = 106	n = 105	n = 106	n = 107	n = 106	n = 105	n = 107	n = 104	n = 106	n = 107	n = 101	n = 105		G: 0.18(0.73) A: 1.00(0.26) H:-0.06(0.95)	G: 0.56(0.16) A: 0.18(0.80) H: 0.54(0.28)	G: 0.52(0.33) A: 1.13(0.29) H:-1.49(0.07)	G: 0.64(0.23) A: 1.35(0.16) H: 0.50(0.59)	G: 0.60(0.19) A: 0.89(0.32) H: **2.55(0.00)**	G: **1.10(0.04)** A: 0.75(0.42) H: **5.32(0.00)**	G: **1.13(0.02)** A: 2.69(0.00) H:-1.09(0.09)	G: **1.24(0.02)** A: **1.88(0.03)** H: 0.20(0.82)	G: 0.86(0.10) A: 2.27(0.05) H: 0.33(0.73)	G: **1.98(0.00)** A: 1.87(0.06) H: **3.93(0.00)**	G: **1.41(0.01)** A: 2.06(0.07) H: 1.34(0.15)	G: **1.47(0.01)** A: 2.06(0.07) H: **2.27(0.01)**
213	n = 107	n = 106	n = 107	n = 108	n = 107	n = 106	n = 108	n = 105	n = 107	n = 108	n = 102	n = 105	n = 107		G: 0.45(0.43) A:-0.64(0.52) H: 0.66(0.46)	G: 0.37(0.52) A: 0.20(0.85) H: **1.60(0.02)**	G: 0.47(0.37) A: 0.45(0.31) H: 0.55(0.04)	G: 1.02(0.06) A:-0.11(0.91) H: **2.65(0.00)**	G: **1.08(0.04)** A: 0.01(0.99) H: **5.49(0.00)**	G: 0.55(0.14) A: 0.14(0.91) H:-1.02(0.26)	G: **1.79(0.00)** A: 1.75(0.08) H: 0.01(0.98)	G: **1.21(0.00)** A: 0.66(0.34) H: 0.31(0.58)	G: 1.32(0.03) A: 1.07(0.36) H: **3.65(0.00)**	G: 1.23(0.02) A: 0.86(0.14) H: **1.38(0.01)**	G: 1.22(0.03) A: 1.22(0.29) H: **2.31(0.00)**
651	n = 106	n = 105	n = 106	n = 107	n = 106	n = 105	n = 107	n = 104	n = 106	n = 107	n = 101	n = 104	n = 106	n = 107		G: 0.00(0.99) A: 1.10(0.26) H:-**1.99(0.02)**	G: 0.02(0.97) A: 0.73(0.43) H: 0.10(0.91)	G: 0.06(0.89) A: 0.61(0.52) H: **1.98(0.01)**	G: 0.45(0.42) A: 0.83(0.39) H: **4.60(0.00)**	G: 0.57(0.22) A: **2.56(0.01)** H:-**1.63(0.01)**	G: 0.72(0.16) A: 1.55(0.10) H:-0.36(0.68)	G: 0.22(0.70) A: 2.11(0.08) H: 0.25(0.80)	G: **1.48(0.02)** A: 1.50(0.14) H: **3.03(0.00)**	G: 0.76(0.17) A: 1.88(0.13) H: 0.74(0.42)	G: 0.93(0.12) A: 1.76(0.04) H: **1.76(0.04)**
26	n = 105	n = 106	n = 105	n = 106	n = 105	n = 106	n = 106	n = 103	n = 105	n = 106	n = 102	n = 103	n = 105	n = 106	n = 105		G: 0.12(0.84) A: 0.26(0.80) H: 2.17(0.00)	G: 0.01(0.98) A: 0.34(0.73) H: **4.13(0.00)**	G: 0.42(0.46) A:-0.50(0.60) H: **6.97(0.00)**	G: 0.58(0.26) A: 1.56(0.13) H: 0.44(0.59)	G: **1.48(0.01)** A: 0.99(0.38) H: **5.67(0.00)**	G: 0.89(0.13) A: 0.71(0.56) H: **3.05(0.00)**	G: **1.09(0.01)** A: 0.64(0.56) H: **3.83(0.00)**
418	n = 107	n = 106	n = 107	n = 108	n = 107	n = 106	n = 108	n = 105	n = 107	n = 108	n = 102	n = 105	n = 107	n = 108	n = 107	n = 106		G:-0.02(0.97) A:-0.54(0.58) H: **2.09(0.00)**	G: 0.55(0.33) A:-0.36(0.71) H: **4.88(0.00)**	G: 0.53(0.42) A:-0.31(0.79) H:-1.57(0.09)	G: 0.60(0.27) A: 1.29(0.20) H:-0.55(0.38)	G: 0.09(0.78) A: 0.22(0.71) H:-0.19(0.71)	G: **1.33(0.04)** A: 0.60(0.61) H: **3.18(0.00)**	G: 0.74(0.08) A: 0.41(0.53) H: 0.63(0.10)	G: 0.84(0.17) A: 0.85(0.46) H: **1.69(0.03)**
671	n = 106	n = 105	n = 106	n = 107	n = 106	n = 105	n = 107	n = 104	n = 106	n = 107	n = 101	n = 104	n = 106	n = 107	n = 106	n = 105	n = 107		G: 0.49(0.36) A: 0.09(0.92) H: 2.50(0.00)	G: 0.51(0.34) A: 0.19(0.85) H: **5.69(0.00)**	G: 0.49(0.25) A: 1.50(0.06) H: **2.66(0.00)**	G: 0.20(0.67) A: 0.89(0.29) H: **2.25(0.01)**	G: 1.08(0.07) A: 0.60(0.57) H: 0.68(0.31)	G: 0.65(0.23) A: 0.65(0.53) H: **1.37(0.04)**	G: 0.74(0.16) A: 1.04(0.31) H: 0.59(0.38)
307	n = 103	n = 102	n = 104	n = 104	n = 104	n = 102	n = 104	n = 101	n = 103	n = 104	n = 98	n = 102	n = 104	n = 104	n = 103	n = 102	n = 104	n = 103		G: 0.16(0.78) A: 0.52(0.58) H:-**6.50(0.00)**	G: 0.16(0.78) A: **1.66(0.03)** H:-**5.38(0.00)**	G:-0.33(0.54) A: 0.77(0.38) H:-**4.95(0.00)**	G: 0.78(0.25) A: 1.02(0.42) H:-**1.59(0.01)**	G: 0.25(0.67) A: 0.93(0.39) H:-**4.02(0.00)**	G: 0.53(0.38) A: 1.55(0.15) H:-**3.18(0.00)**
113	n = 107	n = 106	n = 107	n = 108	n = 107	n = 106	n = 108	n = 105	n = 107	n = 108	n = 102	n = 105	n = 107	n = 108	n = 107	n = 106	n = 108	n = 107	n = 104		G: 0.14(0.83) A: 1.61(0.15) H: 1.18(0.18)	G:-0.27(0.67) A: 0.62(0.46) H: 1.35(0.19)	G: 0.52(0.38) A: 0.69(0.53) H: **4.38(0.00)**	G: 0.22(0.73) A: 0.73(0.57) H: **2.40(0.01)**	G: 0.36(0.54) A: 1.18(0.27) H: **3.36(0.00)**
105	n = 105	n = 106	n = 105	n = 106	n = 105	n = 106	n = 106	n = 103	n = 105	n = 106	n = 102	n = 103	n = 105	n = 106	n = 105	n = 106	n = 106	n = 105	n = 102	n = 106		G:-0.37(0.49) A:-0.84(0.36) H: 0.39(0.56)	G: 0.09(0.10) A:-0.50(0.65) H: **3.90(0.00)**	G: 0.15(0.78) A:-0.91(0.34) H: **1.43(0.02)**	G: 0.23(0.67) A:-0.45(0.65) H: **2.21(0.00)**
214	n = 100	n = 99	n = 100	n = 101	n = 100	n = 99	n = 101	n = 99	n = 100	n = 101	n = 95	V = 98	n = 100	n = 101	n = 100	n = 99	n = 101	n = 100	n = 97	n = 101	n = 99		G: 1.21(0.06) A: 0.54(0.63) H: **3.17(0.00)**	G: 0.77(0.16) A: 0.47(0.59) H: 0.97(0.18)	G: 0.54(0.41) A: 0.21(0.87) H: **1.93(0.02)**
710	n = 92	n = 91	n = 92	n = 93	n = 93	n = 91	n = 93	n = 91	n = 92	n = 93	n = 89	n = 91	n = 92	n = 93	n = 92	n = 91	n = 93	n = 92	n = 91	n = 93	n = 91	n = 87		G:-0.51(0.39) A:-0.12(0.91) H:-**2.40(0.00)**	G:-0.36(0.54) A: 0.43(0.72) H:-**1.69(0.02)**
212	n = 107	n = 106	n = 107	n = 108	n = 107	n = 106	n = 108	n = 105	n = 107	n = 108	n = 102	n = 105	n = 107	n = 108	n = 107	n = 106	n = 108	n = 107	n = 104	n = 108	n = 106	n = 101	n = 93		G: 0.11(0.85) A: 0.37(0.74) H: 0.82(0.26)
27	n = 104	n = 105	n = 104	n = 105	n = 104	n = 105	n = 105	n = 102	n = 104	n = 105	n = 101	n = 102	n = 104	n = 105	n = 104	n = 105	n = 105	n = 104	n = 101	n = 105	n = 105	n = 98	n = 90	n = 105	

The results of paired Student's *t*-test on common targets are reported in the form of the mean of the differences in GDT-HA (G), AL-0 (A), and HBscore (H) values along with the associated *P*-values in parentheses. Statistically significant differences between groups (*P*-values < 0.05) are shaded in gray. Number of common targets in the comparison are reported in the lower left half of the table.

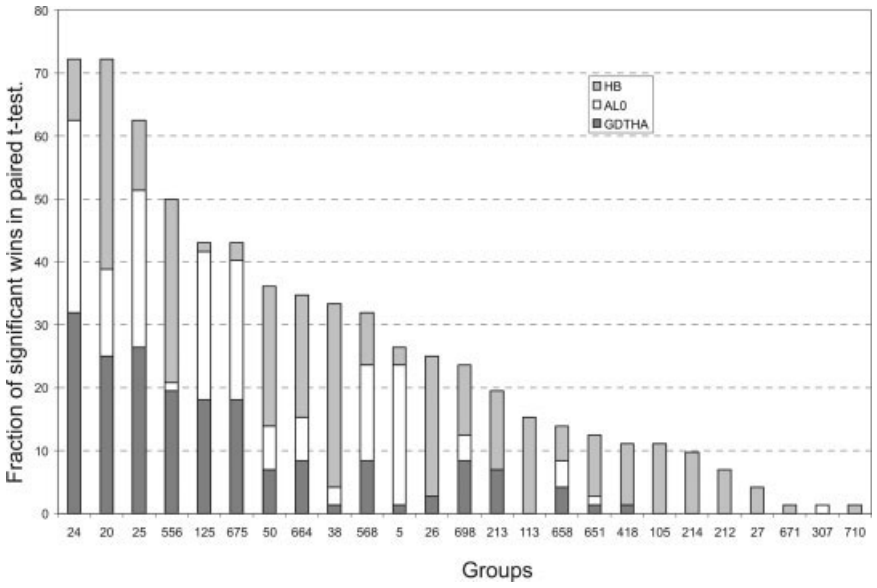

Figure 5

Head-to-head comparison for the top 25 groups showing the fraction of statistically significant wins (Student's t-test P-value < 0.05) on common targets.

that were worse than average, i.e. had negative *z*-scores, and those flagged as physically impossible were assigned *z*-scores of 0. Setting the *z*-scores of models of average or worse quality to zero ensures that the submitting group is not excessively penalized in the overall scoring, thereby encouraging the application and development of innovative and somewhat riskier methods. For each group, we calculated mean *z*-scores for GDT-HA and AL0, as well as a combined mixed *z*-score as the average of both. Table II summarizes the results for each predictor group: The number of models assessed, the mean values, and mean *z*-scores for GDT-HA and AL0, and the combined GDT-HA/AL0 *z*-score. Prediction groups registered as servers are marked with "s." From this list, we selected the 25 highest scoring groups based on the combined *z*-score for a more detailed assessment.

The results of the top 25 groups were compared by direct head-to-head comparison on common targets using paired Student's *t*-tests, as introduced in CASP5.[3] Note that these comparisons were based on the raw scores of GDT-HA, AL0, and HBscore values for each prediction (Table III). For models worse than average (i.e. negative *z*-scores) the raw scores were set to target average (corresponding to $z = 0$). HBscore values for all assessment units for the top 25 groups are provided as Supplementary Materials (Table S-I). Models flagged as physically impossible were omitted from the head-to-head comparison. Finally, the number of statistically significant wins over the other groups (Student's *t*-test *P*-value < 0.05) on common targets was calculated for all three measures, and summed up for each group. Figure 5 shows the fraction of statistically significant wins in the head-to-head comparison for the top 25 groups. The six groups ranked top according to the combined *z*-scores were the same ranked highest in the head-to-head comparison: 24 (Zhang), 20 (Baker), 25 (Zhang-Server), 556 (LEE), 125 (TASSER), and the meta-predictor group 675 (Fams-ace). Groups 24 and 20 produced on average models of higher quality. Remarkably, the automated protein modeling server of group 25 generated on average models of nearly comparable quality to the two leading manual predictor groups. In contrast to earlier CASP experiments, the methods of all top scoring groups were highly automated computational approaches. This reflects on one side the results of ongoing method development in recent years, but partly may also be a sign of the time constraints of manual groups during the modeling season caused by the relatively large number of prediction targets in CASP7.

During the meeting in Asilomar, it became clear that the top scoring methods—although producing models of comparable backbone quality—differ significantly in their algorithmic approaches, computational requirements, modeling of side chain packing, and atomic interactions. This indicates possible directions for further development of the individual methods.

Improvement over the best single template

TBM procedures rely on the detection of and correct alignment to homologous template structures. Consequently, the resulting structure models are generally

DOI 10.1002/prot

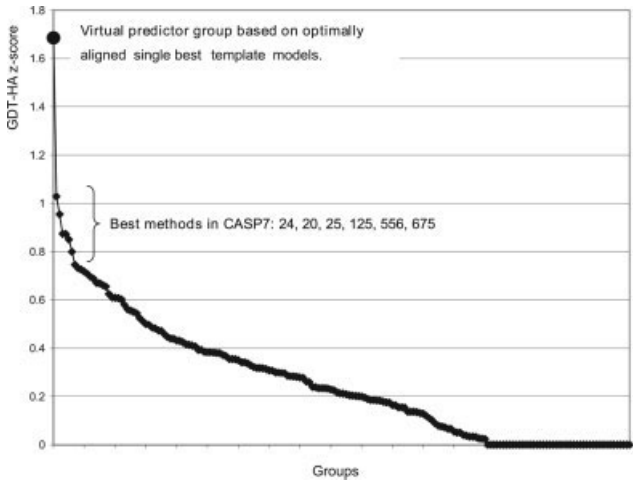

Figure 6

GDT-HA z-scores for all CASP7 methods in comparison with a "virtual predictor group" based on optimally aligned single best template models.

closer to the template than to the target. Recently, several methods have been claimed by their developers to be able to improve over the template. We have therefore assessed whether the predictions submitted to CASP7 showed improvement over the best available single template structure. For this purpose, the PDB was searched for suitable templates available at the end of the prediction period for each target, using LGA[10] and Mammoth[11] as described elsewhere.[7] Based on the structural alignments generated by a 4 Å LGA sequence-independent superposition, we generated pseudo-predictions by copying the backbone coordinates of the templates. No coordinates were assigned to unaligned residues in insertions and deletions. A "virtual predictor group" using these pseudo-predictions would on average outperform all other methods by far, as shown in Figure 6. However, for individual targets, some groups succeeded in building models better than the pseudo-prediction based on the single best template. The best group in this respect (24, Zhang) managed to achieve a higher GDT-TS score than the virtual group in more than half the assessment units and a higher GDT-HA score in approximately one-third of cases. Figure 7 illustrates the fraction of targets for which each group predicted more accurate models than the best single template (plotted on the positive *y*-axis) and targets predicted worse than the best single template (negative *y*-axis).

For a small number of targets, the submitted predictions were significantly better than the best single template model. The most remarkable example was target T0283, where the best prediction showed an improvement of 20.4 GDT-HA units. Several different effects may account for the observed improvements over single template pseudo-predictions, e.g. including information from multiple templates, modeling of insertions, deletions, structurally diverse regions, and refinement to improve the overall quality of the model. A more detailed discussion is provided elsewhere in this issue.[12] Overall however, most of the observed improvements are rather small. In Figure 7, predictions with differences of less than 2.0 GDT-HA units, between the model and the best template are shaded in black. For the majority of targets, the observed differences are still small compared to the overall modeling error—too small to make a significant difference in most biological applications.

Comparison between CASP6 and CASP7

Comparisons between different rounds of CASP are difficult as targets pose very diverse challenges to the predictors, and the modeling difficulty can only be roughly estimated by a combination of various parameters.[1,13] Overall, no large improvement in general model accuracy has been observed between CASP6 and CASP7 when comparing average GDT-TS values as a function of modeling difficulty (data not shown). To assess more subtle improvements, we performed a comparison between CASP6 and CASP7 by evaluating the ability of the methods to improve their models over the best available single template. We applied the "best template model" procedure described above to the CASP6 targets classified as comparative modeling or homologous fold recognition targets (CM and FR/H) based on the best template structures used for the CASP6 assessment.[2,14] Figure 8 shows the average fraction of predictions that boast an improvement of more than 0.0, 0.5, 1.0, 2.0, 4.0, and 8.0 GDT-HA units over the template, for the best ten groups in CASP6 and

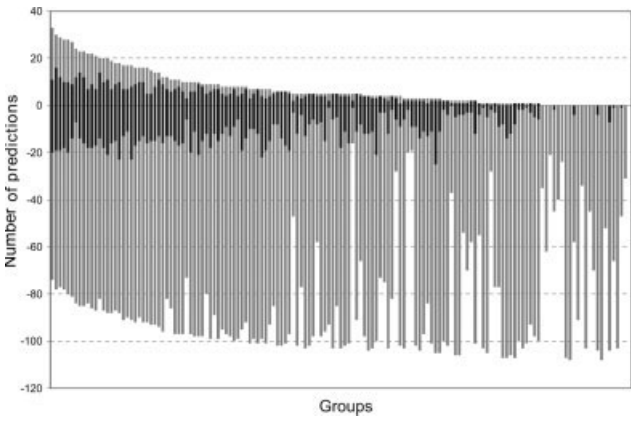

Figure 7

Performance relative to the single best template. For each group, the number of targets predicted more accurately than expected for a model based on the best single template is plotted on the positive y-axis, targets predicted worse are plotted negative. Small differences less than 2.0 GDT-HA units are shaded in black.

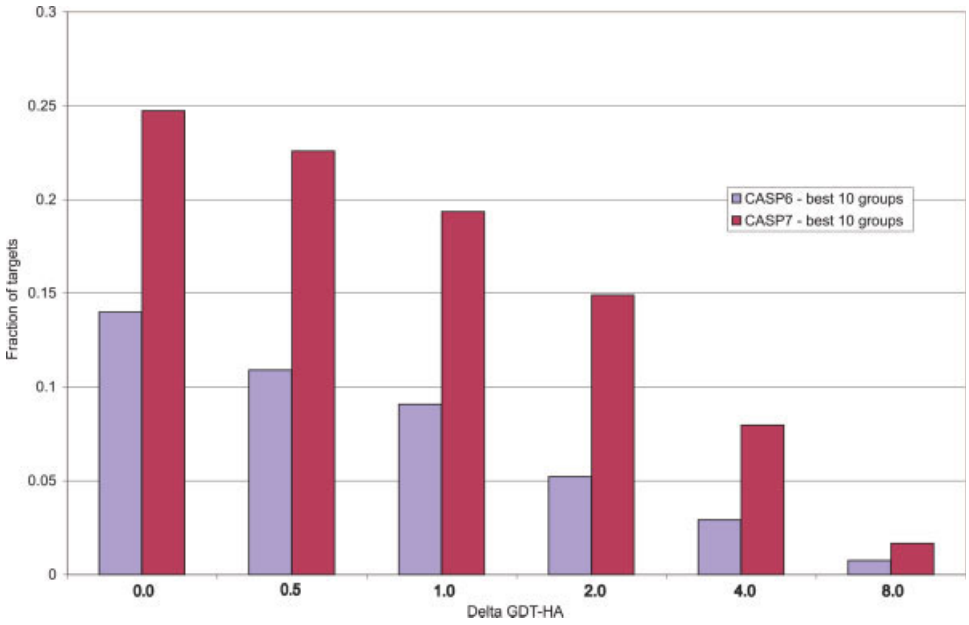

Figure 8

Performance relative to the single best template in comparison between CASP6 and CASP7. The average fractions of predictions, which achieved an improvement compared to the best template model of at least 0.0, 0.5, 1.0, 2.0, 4.0, and 8.0 GDT-HA units are shown for the best 10 groups according to this criterion in both CASP6 (blue) and CASP7 (purple). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

CASP7. Assuming that both CASP6 and CASP7 experiments were of comparable difficulty—which is consistent with a Wilcoxon Rank Sum Test ($P = 0.31$) of the target difficulty based on the scale used in CASP6[2]—we observe that a higher number of groups were able to generate a larger fraction of predictions showing improvement over the best available templates in CASP7.

The observed improvements can be attributed to several factors. Methods that combine multiple template information have made substantial progress in recent years, and none of the top ranked groups in CASP7 was using single template approaches for model building. Multiple template and fragment-based methods can also make effective use of the increased coverage of structure space by structural genomics and individual structure elucidation efforts. Additionally, methods developed for refining template-free models may account for the observed improvement in cases with only limited template structure information. One of the most astonishing examples was target T0283, where two predictor groups have submitted models that were of significantly better quality than the remainder (Fig. 9): Group 20 (Baker) with a GDT-HA of 59.3, and AL0 of 77.3, and group 13 (Jones) with a GDT-HA of 45.1 and AL0 of 62.9. The best available template structure was the archaeal ammonium transporter Amt-1 from *Archaeoglobus fulgidus* (PDB: 2b2j) with an RMSD of 2.54 Å, sharing only 6.7% sequence identity with the target. The best models have significantly lower RMSD values of 1.78 Å and 2.37 Å, respectively. Both groups describe their methods as fragment assembly approaches with a subsequent refinement step. As illustrated in Figure 9(d), the prediction by group 20 is characterized by remarkably accurate local interactions and packing of side chains for a prediction of such low similarity to the closest template and indicates a successful atomic refinement of the model.

There is no optimal method for generating pseudo-predictions for this analysis. Different parameters can be used for identifying structural templates for a target protein, generating structural superpositions, deriving structural alignments, and building the pseudo-models. Pseudo-predictions built using different protocols can differ and data derived from them may vary. Nevertheless, the improvement observed from CASP6 to CASP7 as shown in Figure 8 is stable and largely independent of parameter choice. It should be noted that in most cases the amount of improvement observed for individual targets is relatively small compared to the overall modeling error. This may explain why improvement is only observed by using a "best template model" as internal reference point for each target, while none is detected when using the classical overall difficulty scale. A detailed discussion on progress over previous CASP experiments is provided elsewhere in this issue.[12]

DOI 10.1002/prot

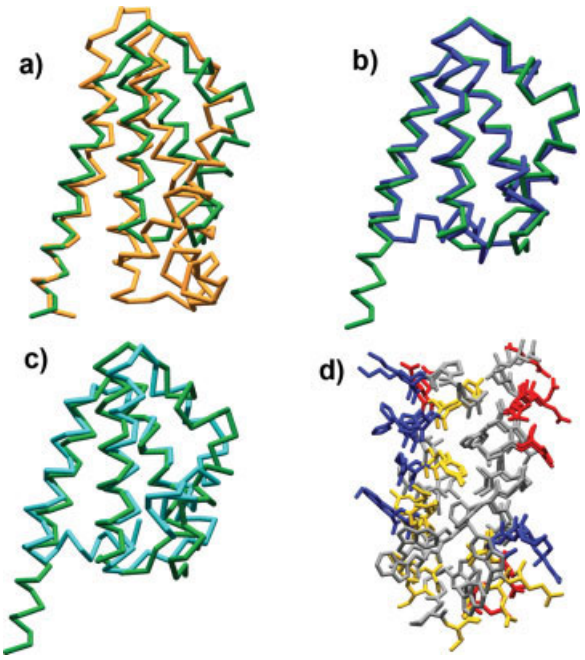

Figure 9

Examples of model quality: (a) Superposition of target structure T0283 (green) and best template (PDB: 2b2j, orange), with an RMSD of 2.54 Å sharing 6.7% sequence identity. The models submitted by two groups were of significantly better quality than the remainder: Group 20 (Baker) with a GDT-HA of 59.3, and AL0 of 77.3 shown in (b) in dark blue. The model by group 13 (Jones) achieved a GDT-HA of 45.1 and AL0 of 62.9 and is shown in (c) in light blue. (d) Detailed view of the side chain packing of the target structure and the model submitted by group 20.

Accuracy of binding site predictions

Active sites or cofactor binding sites in protein models are of great interest for biologists using protein structures or models in their daily work. For several prediction targets, the CASP organizers released the target sequence together with information about a ligand bound in the target structure, which should enable the predictors to model functionally important residues in the binding site more accurately. For the assessment of this aspect, we superposed the models onto the target structure based on only the C_α positions of residues interacting with the ligand in the crystal structure. We evaluated the quality of the modeled binding site using an atomic contact score (ACS), which considers interactions between the nonhydrogen atoms of the protein and the ligand [Eq. (1)]:

$$\mathrm{ACS} = \frac{\sum_{i,k} (\mathrm{Cont}_{i,k}^{\mathrm{target}} \cdot \mathrm{Cont}_{i,k}^{\mathrm{model}}) - \sum_{i,k} \mathrm{Clash}_{i,k}^{\mathrm{model}}}{\sum_{i,k} \mathrm{Cont}_{i,k}^{\mathrm{target}}} \quad (1)$$

with

$$\mathrm{Cont}_{i,k} = \begin{cases} 1 & \text{if } 2.0\text{Å} \le r_{i,k} \le 4.0\text{Å} \\ 0 & \text{otherwise} \end{cases},$$

$$\mathrm{Clash}_{i,k} = \begin{cases} 1 & \text{if } r_{i,k} \le 1.5\text{Å} \\ 0 & \text{otherwise} \end{cases}.$$

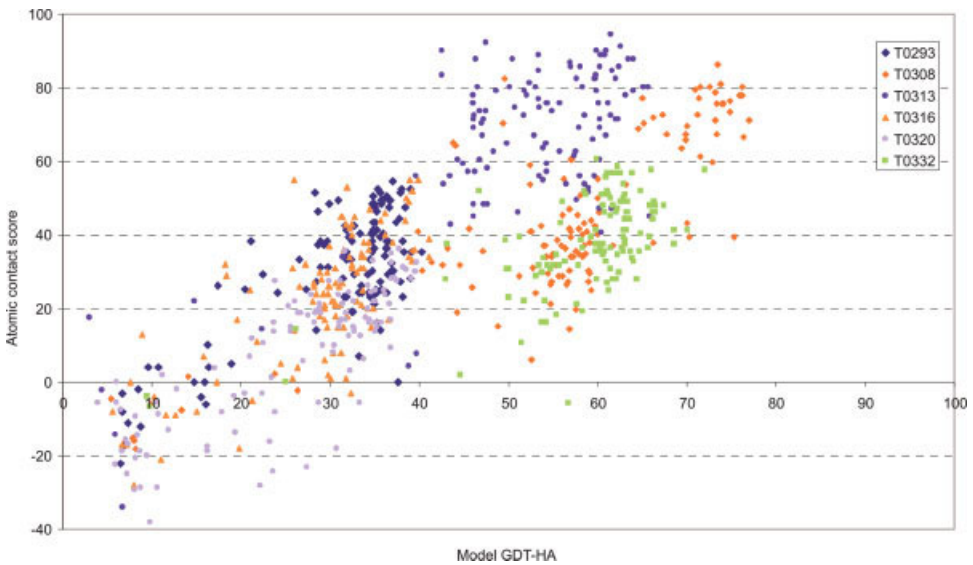

Figure 10

Accuracy of cofactor binding site predictions of six TBM targets. The fraction of correctly modeled atomic interactions in the binding sites (see text) is plotted against the overall model accuracy GDT-HA.

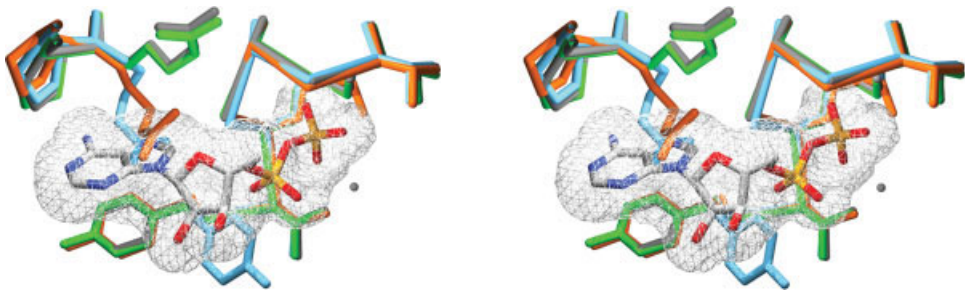

Figure 11

Superposition of experimental and predicted ADP binding sites of target T0313, a human KIFC3 motor domain (stereo view). The experimental structure with the ADP ligand and its solvent accessible surface are shown in gray, the best prediction by group 186 in green, and predictions by groups 20 in orange and 24 in light blue.

We evaluated the fraction of correctly modeled atomic contacts in the predicted binding sites by enumerating specific contacts between the protein atoms (i) and the ligand atoms (*k*) using NCONT.[15] The score in Eq. (1) penalizes interactions, which are predicted shorter than 1.5 Å by classifying them as clashes. Figure 10 illustrates the percentage of correctly modeled contacts in six prediction targets with bound cofactors: *S*-adenosylhomocysteine in T0293 and T0332, GTP in T0308, ADP in T0313, *S*-adenosylmethionine in T0316, and FAD in T0320. Except for T0316, information about the bound ligand was provided along with the prediction target sequence. We would like to emphasize that this analysis of binding site accuracy is based on six examples and therefore has limited statistical power. It only allows for a qualitative description, but not a quantitative assessment of the ability of individual groups to accurately model cofactor binding sites.

The ADP binding site of T0313, a human KIFC3 motor domain, is formed by 12 residues, nine of which are shown in Figure 11. The experimental structure with the ADP ligand and its solvent accessible surface are shown in gray. The best predictions (Fig. 11, green) reproduce the atomic interactions formed by 18 backbone and 24 side chain atoms very well. Models using human mitotic spindle kinesin Eg5 (PDB: 1ii6, chain A) with bound ADP as single template reproduce both main chain and side chain geometry successfully. However, numerous models with accurate backbone geometry were submitted, which fail to form an intact ADP binding site. Figure 11 shows two examples in which the side chain of Arg 9 protrudes into the ADP binding site (orange and light blue), and the stacking interaction by Tyr 92 is modeled incorrectly (light blue).

Overall, the accuracy of the predicted binding sites varies significantly between target structures. Compared to the other examples, T0313 represents a relatively simple modeling task as the alignment is unambiguous. For T0313, the best groups manage to reproduce more than 90% of the ligand–protein interactions, while even in the best predictions for T0320 this fraction is lower than 40% (Fig. 10). While there is a general trend for models with inaccurate backbone geometry to have incorrectly modeled binding sites, it does not hold for models with a GDT-HA above 40. In fact, for T0313, the best five binding site models vary in GDT-HA from 42.6 to 62.6. Therefore, global C_α-based measures such as GDT-HA cannot be used as the only criterion to indicate biological relevance of a model. Also, on average no significant difference in prediction accuracy of the binding site was observed between targets for which the bound cofactor was announced with the prediction target, and T0316 for which this information was not directly available to the predictors. In conclusion, it appears that modeling biologically relevant features of the target proteins as accurately as possible has not received the same level of attention by all predictor groups in CASP7.

Model quality estimates

The practical application of protein models strongly depends on their quality. However, at the time of model generation, the correct answer is unknown and the accuracy of the model must therefore be estimated beforehand. We have assessed a posteriori the ability of individual CASP7 modeling groups to assign realistic error estimates to their predictions. For all targets in the TBM category, we calculated the model error for each predicted amino acid residue as the Cartesian distance between the model C_α coordinates and the experimental target structure in a global superposition with LGA[10] using a 4 Å cut-off. For each participating group (i.e. groups who submitted at least two different values in the *B*-factor column for more than 10 targets), the accuracy of the error estimates ("Model *B*-factor") was analyzed using a log-linear correlation between the estimates and the real error (Fig. 12). Additionally, the results of a random model predictor were added to the analysis for comparison. To compile the data of this null model, "Model *B*-factor" values were

DOI 10.1002/prot

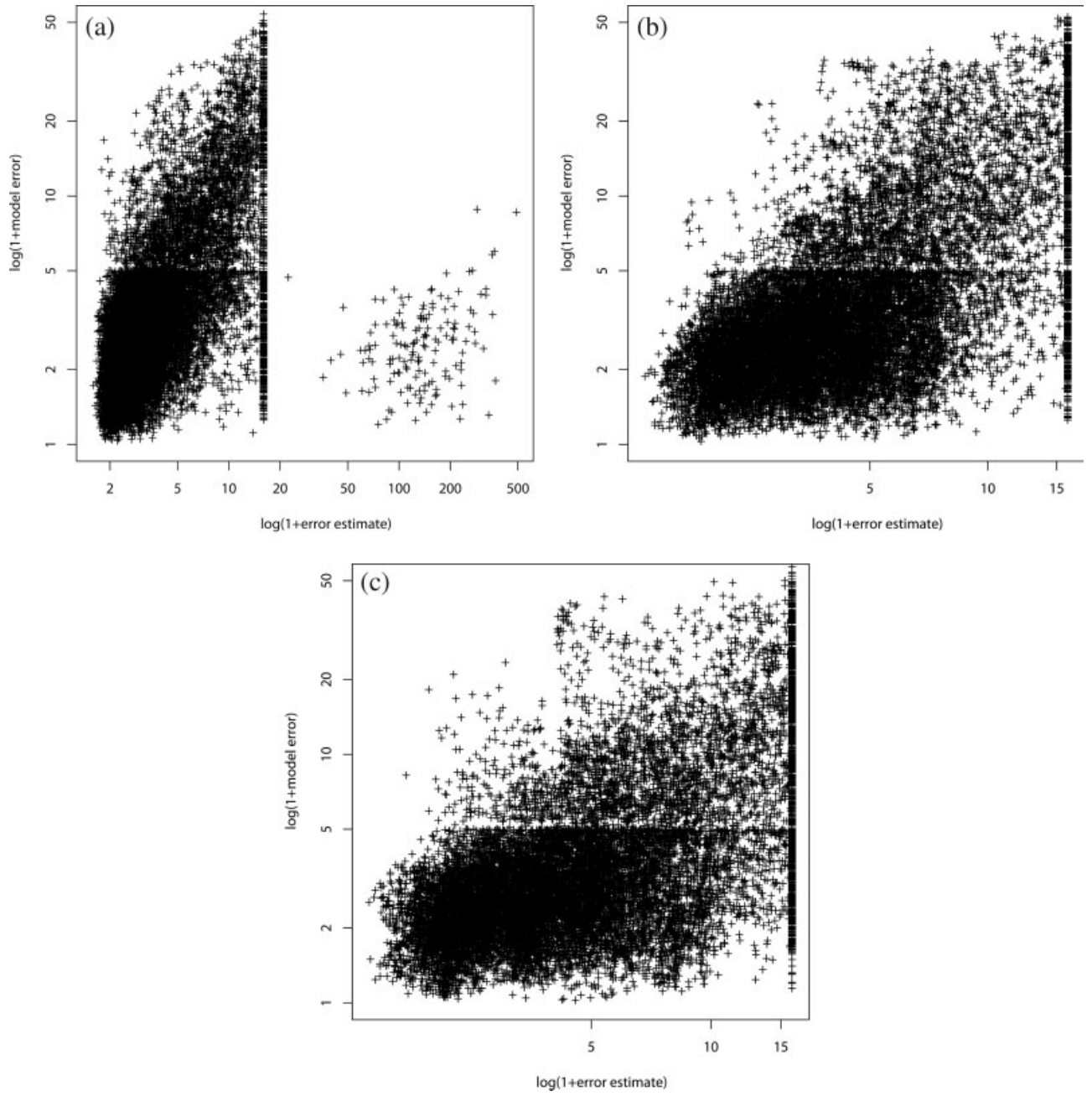

Figure 12

Log-linear correlation between the estimated model error ("Model B*-Factor") and the actual error of the model for (a) group 50, (b) group 46, and (c) group 47.*

randomly chosen from the list of model-target distances. Since linear correlation analysis is sensitive to outliers, the prediction results were additionally analyzed using receiver operator characteristic curves (ROC) (Fig. 13). We classified a residue as correctly modeled if its C_α position error is less than 3.8 Å and incorrectly if its C_α position error is greater than or equal to 3.8 Å. For each group, the "Model *B*-factor" of the predictions were reranked between 0 and 1 and the enrichment of correctly identified model errors was plotted as the false positive rate (FPR) versus the true positive rate (TPR) by varying the discrimination threshold between 0 and 1. The TPR is defined as the number of true positives (TP, the number of correctly identified model errors) over the total number of model errors (positives, *P*), and the FPR the number of false positives (FP, identified as errors in the model, but in reality modeled correctly) over the total number of correctly modeled residues (negatives, *N*). The

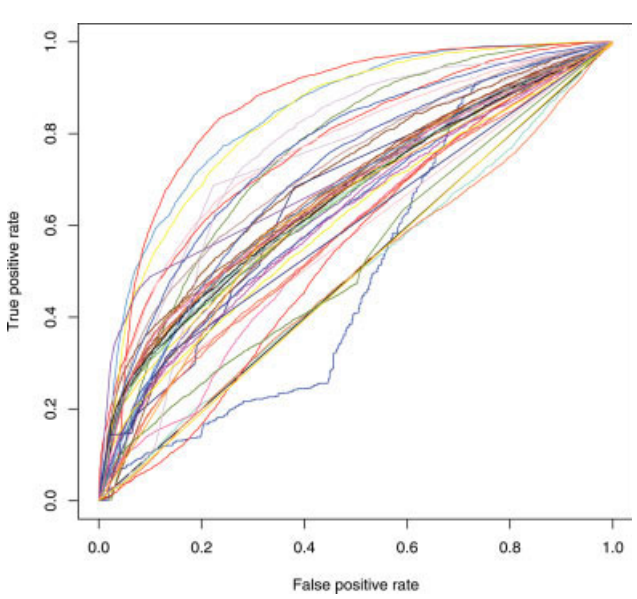

Figure 13

ROC curves analyzing the accuracy of the estimated model error ("Model B-*Factor") to correctly identify incorrect residues defined as C_α error greater than or equal to 3.8 Å. Groups with highest ROC AUC are 50 (red), 46 (light blue), and 47 (yellow).*

area under the curve (AUC) is used as a measure for the accuracy in correctly identifying model errors.

The results of the correlation analysis and the ROC curves for each of the 60 predictor groups, which submitted model error estimates, are listed in Table IV sorted by decreasing AUC. The best-performing methods according to ROC curves by Eloffson and coworkers (50, SBC; 46, Pcons6; 47, Pmodeller6) display also the highest linear correlation coefficients ranging from 0.530 to 0.620. Although many groups were using their own metric, the best groups provided model error estimates based on an absolute metric in Ångstrom as specified by the CASP format. In summary, although only 32% of the groups have provided confidence values for their predictions, the results of the present CASP experiment are encouraging. Regarding the type of methods used, it appears that consensus-based approaches[16] using server predictions submitted to CASP7 as input outperformed other approaches based solely on physics, statistical measures, and traditional model quality estimation programs (MQEP).[17]

CONCLUSIONS

From the perspective of a method developer, the aim of the assessment of template-based protein structure prediction in CASP is to establish the state of the art in the field, identify progress, and pinpoint bottlenecks in areas where further research is required. From the perspective of the life science community, template-based protein structure prediction and modeling has come of age and is widely used today as a scientific research tool. Therefore, an increasingly important aspect of the assessment is also to evaluate to what extent today's prediction methods meet the accuracy requirements of different scientific applications.

In CASP7, we could base our evaluation on a large number of predictions, which provided a solid basis to

Table IV

Assessment of Model Error Estimates

No.	Groups	ROC (AUC)	Correlation (*r*)	No.	Groups	ROC (AUC)	Correlation (*r*)	No.	Groups	ROC (AUC)	Correlation (*r*)
1	050	0.858	0.528	21	248	0.665	0.288	41	401	0.608	0.184
2	046	0.844	0.616	22	347	0.662	0.259	42	698	0.600	0.154
3	047	0.832	0.573	23	658	0.660	0.270	43	416	0.596	0.250
4	005	0.771	0.526	24	413	0.660	0.263	44	137	0.584	0.124
5	214	0.765	0.434	25	609	0.659	0.251	45	136	0.583	0.149
6	275	0.762	0.445	26	415	0.656	0.260	46	654	0.583	0.098
7	038	0.753	0.287	27	728	0.655	0.243	47	453	0.583	0.145
8	368	0.743	0.396	28	414	0.653	0.254	48	020	0.576	0.234
9	004	0.742	0.492	29	074	0.651	0.263	49	083	0.570	0.088
10	060	0.722	0.393	30	087	0.649	0.264	50	261	0.541	0.084
11	297	0.705	0.491	31	267	0.649	0.227	51	659	0.534	0.002
12	025	0.704	0.326	32	490	0.643	0.105	52	338	0.527	0.027
13	274	0.699	0.299	33	021	0.641	0.224	53	013	0.503	0.074
14	212	0.684	0.322	34	337	0.635	0.206	54	063	0.498	0.013
15	103	0.676	0.299	35	494	0.634	0.377	55	420	0.495	0.008
16	677	0.676	0.299	36	427	0.629	0.296	56	474	0.494	0.020
17	213	0.672	0.280	37	105	0.626	0.200	57	071	0.493	0.009
18	092	0.671	0.283	38	651	0.626	0.226	58	483	0.491	0.010
19	091	0.669	0.253	39	319	0.622	0.187	59	203	0.488	0.034
20	418	0.665	0.254	40	024	0.621	0.172	60	614	0.487	0.041

For the 60 groups providing model error estimates for their predictions, the accuracy of residue-based error estimates was assessed by ROC and log-linear correlation using differences between the individual model and the target structure as reference.

DOI 10.1002/prot

assess the differences between the participating prediction methods. We have adapted the numerical assessment criteria to account for the scientific progress in the field of TBM: we applied a global distance test with stricter cutoff values (GDT-HA 0.5, 1, 2, 4 Å) to focus more on the finer details of the predictions. Additionally, to complement GDT and AL0 scores, which are based solely on global superpositions of C_α atoms, we introduced HBscore as local atomic measure evaluating how well hydrogen bond interactions in the target structure are reproduced in the model. Local atomic measures such as HBscore can discriminate between predictions with otherwise similar C_α structures. We have observed significant differences in the accuracy of modeling atomic interactions of backbone hydrogen bonds and side chain packing, indicating areas for further improvement for many of the participating methods.

Overall, the top scoring groups relied on highly automated computational approaches with limited manual intervention. Remarkably, one automated modeling server produced models of nearly comparable quality to the two leading manual predictor groups. We analyzed the ability of different methods to generate predictions that improve over a model based on a single best template structure. Compared to CASP6, a higher number of methods were able to achieve improvement over the best template. It appears that several methods make effective use of multiple template structures. In some cases with limited template information, the observed improvement over template can be attributed to successful application of fragment based modeling or model refinement methods. Although the observed improvement over the best template model is a promising step in the right direction, it is mostly very limited and often cannot be considered as biologically relevant. The fact that no group would outperform a "virtual predictor" submitting models based on the single best template for each target indicates that template identification and alignment are by no means solved problems and constitute a major bottleneck in TBM, besides the challenging question of model refinement.

For regions of the protein, which are of functional importance, such as active sites or ligand binding pockets, accurate reproduction of local interactions such as hydrogen bonds and side chain conformations is essential. For six CASP7 target structures with bound cofactors, we observed considerable differences between models in terms of local model accuracy, even when their C_α structures are similar. Since the accuracy of functional regions is a limiting factor for the scientific usefulness of the predicted structure, improvements in this area would be a big benefit for the life science community.

According to Henry A. Bent*, "... a model must be wrong, in some respects—else it would be the thing itself. The trick is to see ... where it's right."[18] In other words, accurate estimates of the errors of a model are an essential component of any predictive method—protein structure prediction not being an exception. Therefore, we have (for the first time in CASP) evaluated the accuracy of the expected model errors provided by the predictors for their models. Unfortunately, only one-third of all predictor groups provided these for their predictions. Clearly, consensus-based methods gave the most accurate error estimates. From our point of view, confidence measures are an essential part of a prediction method both from a methods development and practical application perspective, and should therefore be an integral component of future assessments.

*Deliberately quoted out of context.

ACKNOWLEDGMENTS

We are grateful to the Prediction Center, especially Andriy Krystafovych, for providing data for the analysis and fast and profound support. We wish to thank our fellow assessors Randy Read and Neil Clarke for the constructive collaboration, and Ernest Feytmans for advice on the statistical analysis. We are grateful to the CASP6 assessors B. K. Lee, Alfonso Valencia, and Roland Dunbrack for valuable advice on the CASP assessment, and to Michael Tress for providing template information for CASP 6 and 7 targets. We gratefully acknowledge support of our group by the Swiss Institute of Bioinformatics. We thank all experimental groups and participating predictors, without whom CASP would not be possible. Last but not least, we thank John Moult, Burkhard Rost, Tim Hubbard, and especially Anna Tramontano for continuous encouragement and valuable discussions.

REFERENCES

1. Kryshtafovych A, Venclovas C, Fidelis K, Moult J. Progress over the first decade of CASP experiments. Proteins 2005;61(Suppl 7):225–236.
2. Tress M, Ezkurdia I, Grana O, Lopez G, Valencia A. Assessment of predictions submitted for the CASP6 comparative modeling category. Proteins 2005;61(Suppl 7):27–45.
3. Tramontano A, Morea V. Assessment of homology-based predictions in CASP5. Proteins 2003;53(Suppl 6):352–368.
4. Wang G, Jin Y, Dunbrack RL, Jr. Assessment of fold recognition predictions in CASP6. Proteins 2005;61(Suppl 7):46–66.
5. Read RJ, Chavali G. Assessment of CASP7 predictions in the high accuracy template-based modeling category. Proteins 2007;69(Suppl 8): 27–37.
6. Jauch R, Yeo H, Kolatkar PR, Clarke ND. Assessment of CASP7 structure predictions for template free targets. Proteins 2007;69(Suppl 8): 57–67.
7. Clarke ND, Ezkurdia I, Kopp J, Read RJ, Schwede T, Tress M. Domain definition and target classification for CASP7. Proteins 2007; 69(Suppl 8):10–18.
8. Virnau P, Mirny LA, Kardar M. Intricate knots in proteins: function and evolution. PLoS Comput Biol 2006;2:e122.
9. McDonald IK, Thornton JM. Satisfying hydrogen bonding potential in proteins. J Mol Biol 1994;238:777–793.
10. Zemla A. LGA: a method for finding 3D similarities in protein structures. Nucleic Acids Res 2003;31:3370–3374.

11. Ortiz AR, Strauss CE, Olmea O. MAMMOTH (matching molecular models obtained from theory): an automated method for model comparison. Protein Sci 2002;11:2606–2621.
12. Kryshtafovych A, Fidelis K, Moult J. Progress from CASP6 to CASP7. Proteins 2007;69(Suppl 8):194–207.
13. Venclovas C, Zemla A, Fidelis K, Moult J. Assessment of progress over the CASP experiments. Proteins 2003;53(Suppl 6):585–595.
14. Tress M, Tai CH, Wang G, Ezkurdia I, Lopez G, Valencia A, Lee B, Dunbrack RL, Jr. Domain definition and target classification for CASP6. Proteins 2005;61(Suppl 7):8–18.
15. The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr 1994;50(Pt 5):760–763.
16. Wallner B, Elofsson A. Identification of correct regions in protein models using structural, alignment, and consensus information. Protein Sci 2006;15:900–913.
17. Cozzetto D, Kryshtafovych A, Ceriani M, Tramontano A. Assessment of predictions in the model quality assessment category. Proteins 2007;69(Suppl 8):175–183.
18. Bent HA. Uses (and abuses) of models in teaching chemistry. J Chem Educ 1984;61:774–777.

DOI 10.1002/prot