WILEY InterScience® DISCOVER SOMETHING GREAT

PROTEINS STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

New Folds: Server

Automated server predictions in CASP7

James N. D. Battey,[1,2] Jürgen Kopp,[1,2] Lorenza Bordoli,[1,2] Randy J. Read,[3] Neil D. Clarke,[4] and Torsten Schwede[1,2]*

[1] Biozentrum, University of Basel, Basel, Switzerland

[2] Swiss Institute of Bioinformatics, Basel, Switzerland

[3] Department of Haematology, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, United Kingdom

[4] Genome Institute of Singapore, Singapore

ABSTRACT

With each round of CASP (Critical Assessment of Techniques for Protein Structure Prediction), automated prediction servers have played an increasingly important role. Today, most protein structure prediction approaches in some way depend on automated methods for fold recognition or model building. The accuracy of server predictions has significantly increased over the last years, and, in CASP7, we observed a continuation of this trend. In the template-based modeling category, the best prediction server was ranked third overall, i.e. it outperformed all but two of the human participating groups. This server also ranked among the very best predictors in the free modeling category as well, being clearly beaten by only one human group. In the high accuracy (HA) subset of TBM, two of the top five groups were servers. This article summarizes the contribution of automated structure prediction servers in the CASP7 experiment, with emphasis on 3D structure prediction, as well as information on their prediction scope and public availability.

Proteins 2007; 69(Suppl 8):68–82.

Key words: CASP7; automated modeling server; model quality assessment.

INTRODUCTION

Protein structure prediction has become increasingly dependent on automated approaches in response to the mounting volume of data generated by large scale genome sequencing and structural genomics efforts.[1,2] Today, numerous fully automated modeling servers (for review, see recent NAR web server issue)[3] and model databases[4–6] are offering modeling services to the biomedical research community. Ultimately, the development of automated prediction methods seeks to encode expert knowledge into software. This automation allows these methods to be applied to large datasets such as whole proteomes of different organisms, as well as providing input for other prediction efforts. Since automated algorithmic approaches are devoid of human bias, their accuracy and consistency can be assessed objectively, which is an important prerequisite for their application as services for the research community.

The CASP experiment endeavors to provide a rigorous and blind assessment of state of the art methods in structure prediction.[7] Although many predictor groups participating in CASP use computational modeling procedures, these methods require human intervention at many points in the process such as performing plausibility checks, refining certain modeling steps, and selecting final models. Although this manual intervention has given human predictors a decisive advantage over fully automated methods in the past, prediction servers have played an increasingly important role. While from CASP3 to CASP5, server predictions were assessed separately as part of the CAFASP experiments,[8,9] they were part of the regular assessment as of CASP6.[10] During CASP7, prediction targets were sent to the servers automatically by the Prediction Center at UC Davis. A time limit of 48 h was imposed, in effect simulating real life modeling situations, in

The authors state no conflict of interest.

*Correspondence to: Torsten Schwede, Biozentrum University of Basel, Swiss Institute of Bioinformatics, Klingelbergstrasse 50-70, 4056 Basel, Switzerland. E-mail: torsten.schwede@unibas.ch

Received 19 May 2007; Revised 18 July 2007; Accepted 1 August 2007
Published online 25 September 2007 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/prot.21761

which time is often at a premium and hence long waiting times for results are undesirable. The server predictions were then made publicly available on the CASP7 web site to allow predictor groups not registered as servers to use these data as input for their own predictions.

One aim of the CASP experiment is to measure the progress in the field. However, it has proven difficult to devise a suitable way to estimate the individual difficulty of a prediction target, which would allow comparing prediction success based on the different target data sets of two CASP experiments.[11] During the CASP7 meeting in Asilomar, it was suggested that the predictions based on servers with "frozen" algorithms using updated databases could serve as a baseline for measuring progress when comparing different CASP experiments. This approach would thereby allow the separation of improvements due to growth of underlying sequence and structure databases from those due to algorithmic developments.

This article provides an overview of all server methods participating in CASP7 in the various categories, including information on their prediction scope, public availability, URLs, and author contact information. Additionally, we summarize the results of servers in the assessment of the tertiary structure prediction categories and highlight some examples of successful predictions by servers.

SERVERS PARTICIPATING IN CASP7

In CASP7, 93 of 305 predictor groups participated as servers, with 68 servers in the tertiary structure prediction category, 8 servers in disorder prediction, 14 servers in domain boundary prediction, 8 in residue–residue contact prediction, and 6 in function prediction. Tables I and II summarize the results of a survey among the server groups regarding the following questions:

- What is the scope of the prediction server? Which input data are required?
- Is the server publicly available? Is the software available for local installation?
- Will the algorithm and/or databases be updated during the next 2 years?
- Contact details and URL for submission (if applicable).

The accuracy of server predictions in CASP7 has been assessed together with predictions submitted by manual predictor groups in each of the individual categories as described elsewhere: free modeling (FM),[66] template-based modeling (TBM),[67] high accuracy models (HA),[68] disorder,[69] domain boundary,[70] contact prediction,[71] and function prediction.[72]

Numerical assessment of the tertiary structure prediction categories in CASP7 is based on several criteria. 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 (GDT-TS: 1, 2, 4, and 8 Å; GDT-HA: 0.5, 1, 2, and 4 Å). AL0 is defined as the percentage of correctly aligned residues in a sequence independent superposition of the model and experimental structure of the target. HBscore was introduced as an additional measure in this round of CASP in the TBM assessment. It is defined as the number of correctly predicted hydrogen bonds relative to the total number in the target structure. For this calculation we excluded side chains of residues with more than 50% relative surface exposure in the target structure, and residues with incorrect topology or involved in steric clashes in the models. For a detailed discussion of the criteria please refer to the assessment reports of the individual categories.[66–68] In the following, we will summarize the results of servers in the assessment of the tertiary structure prediction categories, and highlight some examples of successful predictions.

SERVER PREDICTIONS IN THE TBM CATEGORY

Accuracy of server predictions in the template-based modeling category

The accuracy of server predictions has continuously increased over the last years, and in CASP7 we observed a continuation of this trend with servers performing very well. In the template-based modeling (TBM) category, 68 of 187 groups were registered as prediction servers. As described in the CASP7 TBM assessment, the top 25 groups selected based on combined *z*-scores of GDT-HA and AL0 were compared by direct head-to-head comparison of statistically significant differences of GDT-HA, AL0, and HBscore on common targets.[67] Among these 25 groups, 6 were servers with the best group (25 Zhang-server) ranked third overall.

Here, we aimed at a direct comparison of only servers in this category. We have therefore recalculated the numerical assessment, taking only into account the predictions submitted by servers. The results are presented in Figure 1 and Table III. The best performing group Zhang-server (group 25) is followed by servers developed by Soeding *et al.* (213 HHpred2; 214 BayesHH; 418 HHpred3), Elofsson and coworkers (47 Pmodeller6), Baker and coworkers (4 Robetta), and Skolnick and coworkers (307 MetaTasser).

Examples of successful server predictions

Several examples of outstanding predictions submitted by servers were observed in CASP7, e.g. for target T0321 (PDB:2h1q), which is a structural genomics target from *Desulfitobacterium hafniense* of unknown function. The protein of 250 amino acid residues forms a two-domain mixed αβ-structure. The C-terminal domain is character-

Table I
Servers Participating in CASP7

Server	CASP7 ID	Which type of prediction does this server offer?	Is the server publicly available?	Which input data are required?	Which output data are generated?	Average server response time	CASP8	Method updates	Data updates	Local installation	Reference
3D-JIGSAW	302	TBM	Y	S, A	FA, TTA	~30 min	Y	Y	Y	N	12
3D-JIGSAW_POPULUS	247	TBM, FM	N	–[c]	–[c]	–[c]	–[c]	Y	Y	Y	13
3D-JIGSAW_RECOM	420	TBM	Y	C	FA, energy curve	~1 h	Y	N	N	N	14
3Dpro	137	FM	Y	S	FA	~10 min	Y	Y	–[c]	N	15
ABIpro	139	FM	Y	S	FA	~3 h	Y	N	Y	N	15
BAKER-ROSETTADOM	497	DBP	N	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	N	16
BayesHH	214	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]
beautshot	275	TBM	N	S	FA	~2 min	Y	Y	Y	N	–[c]
Beautshotbase	347	TBM	N	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	N	–[c]
BETApro	141	CP	Y	S	–[c]	Several minutes	Y	–[c]	–[c]	Y	17
bicmkusk-serv	202	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]
Bilab-ENABLE	179	TBM	N	S	FA	~3 h	Y	Y	Y	N	18
BIME@NTU_serv	272	DP	Y	S	Propensity for disorder/order	~5 min	Y	Y	Y	N	19
CaspIta-FOX	186	TBM	Y	S	FA	~14 h	Y	Y	Y	N	–[c]
CaspIta-GOret	573	FP	Y	S	Prediction sent by email in Casp7 FN format	~1 h	Y	Y	Y	N	–[c]
Chop	595	DBP	–[c]	–[c]	Neural network prediction of domain linking regions	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]
Chop_homo	649	DBP	–[c]	–[c]	Homology based method for domain prediction	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]
CIRCLE	298	TBM	N	S	FA	~24 h	Y	Y	Y	N	–[c]
CPHmodels	494	TBM	Y	S	FA	~1 min	Y	Y	Y	Y	20
DISOPRED	470	DP	Y	S	Regions predicted to be disordered	–[c]	Y	N	–[c]	Y	21
DISpro	140	DP	Y	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	Y	22
Distill	168	TBM, FM, DP, DBP	Y	S	CA, contact maps, SS, solvent accessibility, etc.	~10 min	Y	Y	Y	Y	23
DomFOLD	240	DBP	Y	S	Domain prediction in CASP DP format	~30 min	Y	Y	Y	N	–[c]
DomSSEA	312	DBP	Y	S	Domain boundary prediction	–[c]	Y	–[c]	–[c]	N	24
DPS	310	DBP	Y	S	Domain boundary prediction	–[c]	Y	–[c]	–[c]	N	24
DRIPPRED	153	DP	Y	S	Per-residue disorder score	~1 h	Y	N	Y	N	–[c]
FAMS	351	TBM, FM	Y	S	FA	~7 days	Y	Y	Y	N	25
FAMSD	349	TBM	Y	S	FA	~2 days	Y	N	Y	N	25
FOLDpro	136	TBM	Y	S	FA, TTA	Several hours	Y	Y	Y	N	26
forecast-s	333	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]
FORTE1	257	TBM	Y	S	FA, TTA	–[c]	Y	–[c]	Y	N	27
FORTE2	316	TBM	N	–[c]	–[c]	–[c]		–[c]	Y	N	28
FPSOLVER-SERVER	511	FM	–[c]	S	FA	~2 days	N	Y	N	N	–[c]
Frankenstein	368	TBM	Y	S	FA	~3 days	Y	Y	Y	N	29
FUGMOD	319	TBM	N	S, A	FA	–[c]	N	–[c]	–[c]	N	30
FUGUE	242	TBM	Y	S, A	TTA, BB (without loops)	~10 min	Y	–[c]	Y	Y	30
FUNCTION	318	TBM	N	–[c]	FA	~24 h	N	Y	Y	N	–[c]

(*Continued*)

DOI 10.1002/prot

Table I
(Continued)

Server	CASP7 ID	Which type of prediction does this server offer?	Is the server publicly available?	Which input data are required?	Which output data are generated?	Average server response time	CASP8	Method updates	Data updates	Local installation	Reference
GajdaPairings	618	CP	Y	S	Robust contact prediction	~3 days	Y	Y	Y	N	–[c]
GeneSilicoMetaServer	609	TBM, DBP, DP, MQE, SSP	Y	S, A	FA, TTA	~15 min	Y	Y	Y	N	31
GeneSilicoUnimod	461	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]
GPCPRED	154	CP	Y	S	CASP format contact prediction, 2D contact map image in HTML version	~60 min	Y	N	Y	N	32
Gtg	44	TBM	Y	S, DBI	FA	Seconds	Y	N	Y	Y	–[c]
HHpred1	212	TBM, DBP, FP	Y	S	Confidence value for homology, TTA, FA; Domain prediction; GO function prediction	~10 min	Y	Y	Y	Y	33
HHpred2	213	TBM	Y	S	Confidence value for homology, TTA, FA; domain prediction; GO function prediction	~12 min	Y	Y	Y	Y	33
HHpred3	418	TBM, DBP, FP	Y	S	Confidence value for homology, TTA, FA; domain prediction; GO function prediction	~17 min	N	Y	Y	N	33
Huber-Torda-Server	102	TBM	Y	S	Ranked list of templates, models complete only up to beta carbons	~75 min	Y	Y	Y	Y	34
karypis.srv	22	TBM	Y	S	FA	~10 min	Y	Y	Y	N	–[c]
karypis.srv.2	268	TBM, MQE	Y	S	FA	~2 days	N	Y	N	Y	35,36
karypis.srv.4	193	FM	Y	S	FA	~30 min	N	N	N	N	–
keasar-server	277	TBM	Y	S	FA	~1 h	Y	Y	N	Y	37
LOOPP	83	TBM	Y	S	FA, TTA	~5 h	Y	Y	Y	N	38–40
Ma-OPUS-DOM	229	DBP	Y	S	Domain information in CASP format	~1.5 days	Y	Y	Y	N	–[c]
Ma-OPUS-server	92	TBM, FM	Y	S	FA	~5 h	Y	Y	Y	N	–[c]
Ma-OPUS-server2	728	TBM, FM	Y	S	FA	~5 h	Y	Y	Y	N	–[c]
MBI-NTU-serv	538	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]
Meta-DP	269	DBP	Y	S	Domains and domain boundaries	~5 min	Y	N	Y	N	41
MetaTasser	307	TBM, FM	Y	S	FA	–[c]	Y	Y	Y	Y	42–44
mGen-3D	261	TBM	N	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]
MIG_FROST	586	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]
MIG_FROST_FLEX	588	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]
nFOLD	239	TBM	Y	S	BB	~30 min	Y	N	Y	N	45
NN_PUT_lab	654	TBM, DBP	Y	S	FA, metaserver	–[c]	Y	Y	Y	N	–[c]
panther2	69	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]
panther3	304	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]

(Continued)

Table I
(Continued)

Server	CASP7 ID	Which type of prediction does this server offer?	Is the server publicly available?	Which input data are required?	Which output data are generated?	Average server response time	CASP8	Method updates	Data updates	Local installation	Reference
Pcons6	46	TBM, FM, MQE	Y	S	FA, ALN, QS	5–30 min	Y	Y	Y	Y	46
PFP_HAWKINS	753	FP	Y	S	Predicted Gene Ontology annotations (MF, BP, CC)	~30 min	Y	Y	Y	Y	47
Phyre-1	468	TBM	Y	S	FA	30 min–1 h	Y	–[c]	N	Y	s
Phyre-2	469	TBM, FM	N	S	FA	~10 h	N	Y	Y	N	–[c]
Pmodeller6	47	TBM, FM, MQE	Y	S	FA, ALN, QS	5–30 min	Y	Y	Y	Y	46
POMYSL	464	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]
Possum	230	CP	Y	S	Contact pairs in Casp format	~5 min	Y	Y	Y	N	48
PROFcon-Rost	296	CP	Y	S	List of amino acid pairs ranked according to PROFcon contact score (low values = contact unlikely)	1 h	Y	N	Y	N	49
PROTINFO	28	TBM, FM	Y	–[c]	–[c]	–[c]	Y	Y	Y	Y	50
PROTINFO-AB	29	TBM, FM[b]	Y	S, A	Depends on input	3–24 h	Y	Y	Y	Y	50
Raghava-GPS-mango	598	FP	Y	S	Text	~50 s	Y	Y	Y	N	51
RAPTOR	248	TBM	Y	S	FA	~5 h	Y	Y	Y	N	52
RAPTOR-ACE	267	TBM	N	–[c]	–[c]	–[c]	Y	Y	Y	Y	53
RAPTORESS	435	TBM	N	S	FA	–[c]	N	Y	Y	N	–[c]
ROBETTA	4	TBM, FM, DBP, FL, AS	Y	S	FA, TTA, MAMMOTH hits to known structures for de novo models	~1 month	Y	N	Y	N	54
ROBETTA-GINZU	581	DBP	Y	S	Domain boundary prediction	2 h	Y	N	Y	N	16,54
ROKKY	35	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]
Rost_PROFbval	594	Bval	Y	S, MSA	Per residue normalized B-value; 2 state prediction - flexible/rigid	–[c]	Y	N	–[c]	N	55,56
Rost-ECGO	751	FP	N	S	Predicted EC class	5 min	Y	Y	N	N	–[c]
SAM_T06_server	389	TBM, FM, CP, LSP	Y	S	FA, TTA, residue-residue contact predictions, local structure predictions, MSA, HMMs, . . .	~ 12 h	Y	N	Y	N	57
SAM-T02[a]	381	TBM, SSP, MSA	Y	S	TTA, secondary structure predictions, burial predictions	~ 4 h	Y	N	Y	N	58
SAM-T99[a]	380	TBM, SSP, MSA	Y	S	TTA, MSA, SS	~ 12 h	Y	N	Y	N	59
Shub	274	TBM	N	–[c]	–[c]	–[c]	N	Y	N	N	–[c]
SP3	414	TBM	Y	S	FA, TTA	Several hours	Y	N	Y	Y	60
SP4	415	TBM	Y	S	FA, TTA	Several hours	Y	N	Y	Y	61
SPARKS2	413	TBM	Y	S	FA, TTA	Several hours	Y	N	Y	Y	60
SVMcon	138	CP	N	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	–[c]	62
UNI-EID_bnmx	383	TBM	N	–	–	–	–	–	–	N	63
UNI-EID_expm	245	TBM	N	–	–	–	–	–	–	N	64

(*Continued*)

DOI 10.1002/prot

Table I
(Continued)

Server	CASP7 ID	Which type of prediction does this server offer?	Is the server publicly available?	Which input data are required?	Which output data are generated?	Average server response time	CASP8	Method updates	Data updates	Local installation	Reference
UNI-EID_sfst	243	TBM	N	–	–	–	–	–	–	N	63
Zhang-Server	25	TBM, FM	Y	S	FA	~10 h	Y	Y	Y	N	65

The presented information about the participating servers was collected in a survey among the registered groups after the experiment. The table summarizes group name and number, type of predictions performed, public accessibility of the server, required input data and output format, and average response time. The rightmost columns indicate if the server is expected to be available at least until CASP8, if the algorithm and underlying databases will be modified in the meantime, and if the software is available for local instillation. For specific comments provided by the server authors see footnotes.
TBM, template-based modeling; FM, free modeling; DBP, domain boundary prediction; DP, disorder prediction; CP, contact prediction; FP, function prediction; SSP, secondary structure prediction; LSP, local structure prediction; MSA, multiple sequence alignment; MQE, model quality estimation; Bval, normalized B values; FL, fragment libraries; AS, computational Alanine scanning; S, protein sequence; ALN, sequence alignment; C, coordinates; DBI, database identifier; FA, full atom model; BB, backbone model; CA, alpha carbon trace; TTA, target template alignment; SS, secondary structure prediction; MSA, multiple sequence alignment; QS, quality score; s, manuscript submitted for publication.
[a]This server is obsolete and is being kept alive only for historical comparisons.
[b]Optional experimental data.
[c]No information provided by authors.

ized by an extended central β-sheet flanked by four α-helices and has been classified as TBM/FM prediction target since a significant part of the structure could not be modeled based on the available template structure. However, for the N-terminal domain, one server (415 SP4) recognized the structural similarity to the N-terminal domain of Enolases (CATH code 3.30.390.10).[73] The C-terminal domains of enolases are TIM barrels (CATH code 3.20.20.120) and do not resemble the second domain of target T0321. The submitted model by the SP4 server for domain 1 based on mandelate racemase from *Pseudomonas putida* (PDB:2mnr) as template (Fig. 2) achieved a GDT-HA of 35.2 (AL0 of 49.0), which is outstanding when compared with a GDT-HA of 24.2 (AL0 of 0.0) of the second best prediction.

Target T0356 is also a structural genomics target, the 3-octaprenyl-4-hydroxybenzoate decarboxylase (UbiD) from *Escherichia coli* (PDB:1idb). For the assessment, T0356 has been divided into three assessment units: domains 1 and 3 were assessed in the FM category; the second domain, which resembled an FMN-binding protein domain (CATH code 2.30.110.10), was assessed as a TBM target. The best available template, the structure of an archeal FMN-binding protein from *Methanobacterium thermoautotrophicum* (PDB:1eje), was used for several of the best-submitted predictions. The structural similarity was difficult to detect, and only predictions by eight groups were significantly better than the remainder (Fig. 3)—among which seven were registered as servers (212 HHpred1; 213 HHpred2; 214 BayesHH; 418 HHpred3; 92 Ma-OPUS-server; 245 UNI-EID_expm; 383 UNI-EID_bnmx). Interestingly, only one metapredictor method (675 Fams-ace), and none of the manual predictor groups made use of these server predictions.

Limitations in template detection

TBM exploits the evolutionary relationship between a target and a template protein to infer structural similarity. In cases of high sequence identity between the target and the template, simple algorithms for sequence alignment are sufficient for identifying and aligning the best template to the target. If the similarity is low, the detection and alignment of templates require more sophisticated methods. A good template may exist for a target, yet not be detectable by simple sequence-based methods. Fold-recognition methods attempt to address the problem of detecting such remote homologs. As illustrated in the previous two examples, this problem is still far from being generally solved, and considerable performance differences can be attributed to the ability of servers to build their models on the best available templates. Here, we sought to address two issues. The first is whether a server was able to detect the best possible structural template and the second is how well it would have fared in

Table II
Contact Details for Publicly Available CASP7 Servers

Server	URL	Contact E-mail
3D-JIGSAW	http://www.bmm.icnet.uk/∼3djigsaw/	paul.bates@cancer.org.uk
3D-JIGSAW_RECOM	http://www.bmm.icnet.uk/servers/3djigsaw/recomb/index.html	
3Dpro	http://www.ics.uci.edu/∼baldig/scratch/	pfbaldi@ics.uci.edu
ABIpro	http://www.ics.uci.edu/∼baldig/scratch/	arandall@ics.uci.edu
BETApro	http://www.igb.uci.edu/?page=tools&subPage=psss	pfbaldi@ics.uci.edu
BIME@NTU_serv	http://biominer.bime.ntu.edu.tw/casp7/	cychen@mars.csie.ntu.edu.tw
CaspIta-FOX	http://protein.cribi.unipd.it/fox/	stefano.toppo@unipd.it
CaspIta-GOret	http://protein.cribi.unipd.it/go_retriever/	
CPHmodels	http://www.cbs.dtu.dk/services/CPHmodels/	lund@cbs.dtu.dk
DisoPred	http://bioinf.cs.ucl.ac.uk/disopred/	d.jones@cs.ucl.ac.uk
DISpro	http://www.ics.uci.edu/∼baldig/scratch/	pfbaldi@ics.uci.edu
Distill	http://distill.ucd.ie/distill/	gianluca.pollastri@ucd.ie
DomFOLD	http://www.biocentre.rdg.ac.uk/bioinformatics/DomFOLD/DomFOLD_form.html	l.j.mcguffin@reading.ac.uk
DomSSEA	http://bioinf.cs.ucl.ac.uk/dompred/	k.bryson@cs.ucl.ac.uk
DPS	http://bioinf.cs.ucl.ac.uk/dompred/	
DRIPPRED	http://sbcweb.pdc.kth.se/cgi-bin/maccallr/disorder/submit.pl	r.maccallum@imperial.ac.uk
FAMS	http://www.pharm.kitasato-u.ac.jp/fams/fams.html	kanouk@pharm.kitasato-u.ac.jp
FAMSD	http://www.pharm.kitasato-u.ac.jp/fams/famsd.html	
FOLDpro	http://mine5.ics.uci.edu:1026/foldpro.html	pfbaldi@ics.uci.edu
FORTE1	http://www.cbrc.jp/forte/	k-tomii@aist.go.jp
Frankenstein	https://genesilico.pl/meta2	mgajda@genesilico.pl
FUGUE	http://tardis.nibio.go.jp/fugue/; http://www-cryst.bioc.cam.ac.uk/fugue/	kenji@nibio.go.jp
GajdaPairings	https://genesilico.pl/meta2	mgajda@genesilico.pl
GeneSilicoMetaServer	https://genesilico.pl/meta2	andrzej@genesilico.pl
GPCPRED	http://sbcweb.pdc.kth.se/cgi-bin/maccallr/gpcpred/submit.pl	r.maccallum@imperial.ac.uk
gtg	http://www.bioinfo.biocenter.helsinki.fi/gtg	liisa.holm@helsinki.fi
HHpred1	http://protevo.eb.tuebingen.mpg.de/∼toolkit/hhpred1/	johannes.soeding@tuebingen.mpg.de
HHpred2	http://protevo.eb.tuebingen.mpg.de/∼toolkit/hhpred2/	
HHpred3	http://protevo.eb.tuebingen.mpg.de/∼toolkit/hhpred3/	
Huber-Torda-Server	http://www.zbh.uni-hamburg.de/wurst/	torda@zbh.uni-hamburg.de
karypis.srv	http://www.cs.umn.edu/∼karypis/servers/c7pred	karypis@cs.umn.edu
karypis.srv.2	http://dminers.dtc.umn.edu/∼rangwala/mn-fold/fp.php	rangwala@cs.umn.edu
karypis.srv.4	http://www-users.cs.umn.edu/∼deronne/c7pred/	deronne@cs.umn.edu
keasar-server	http://www.cs.bgu.ac.il/∼meshisrv/server/	keasar@cs.bgu.ac.il
LOOPP	http://cbsuapps.tc.cornell.edu/loopp.aspx	ron@cs.cornell.edu
Ma-OPUS-server	http://sigler.bioch.bcm.tmc.edu/MaLab/CASP7-server/	jpma@bcm.tmc.edu
Ma-OPUS-server2	http://sigler.bioch.bcm.tmc.edu/MaLab/CASP7-server2/	
Ma-OPUS-DOM	http://sigler.bioch.bcm.tmc.edu/CASP7-DOM/	
Meta-DP	http://meta-dp.cse.buffalo.edu	hksaini@cse.buffalo.edu
MetaTasser	http://cssb.biology.gatech.edu/skolnick/webservice/MetaTASSER/	skolnick@gatech.edu
nFOLD	http://www.biocentre.rdg.ac.uk/bioinformatics/nFOLD/nFOLD_form.html	l.j.mcguffin@reading.ac.uk
NN_PUT_lab	http://webmobis.cs.put.poznan.pl	protserv@cs.put.poznan.pl
Pcons6	http://pcons.net	bjorn@sbc.su.se
PFP_HAWKINS	http://dragon.bio.purdue.edu/pfp	thawkins@purdue.edu
Phyre-1	http://www.sbg.bio.ic.ac.uk/∼phyre/	l.a.kelley@imperial.ac.uk
Pmodeller6	http://pcons.net	bjorn@sbc.su.se
Possum	http://foo.maths.uq.edu.au/∼nick/Protein/contact.html	n.hamilton@imb.uq.edu.au
PROFcon-Rost	http://www.predictprotein.org/submit_profcon.html	mp2215@columbia.edu
PROTINFO	http://protinfo.compbio.washington.edu/protinfo_abcmfr/	ram@compbio.washington.edu
PROTINFO-AB	http://protinfo.compbio.washington.edu	admin@protinfo.compbio.washington.edu
Raghava-GPS-mango	http://www.imtech.res.in/raghava/mango/	raghava@imtech.res.in
RAPTOR	http://ttic.uchicago.edu/∼jinbo/	j3xu@tti-c.org
ROBETTA	http://robetta.org/submit.jsp	DCChivian@lbl.gov
ROBETTA-GINZU	http://robetta.org/submit.jsp	
Rost-ECGO	http://rostlab.org/services/ecgo/	amk2002@columbia.edu
Rost_PROFbval	http://rostlab.org/services/profbval/	as2067@columbia.edu
SAM_T06_server	http://www.soe.ucsc.edu/research/compbio/SAM_T06/T06-query.html	sam-info@soe.ucsc.edu
SAM-T02[a]	http://www.soe.ucsc.edu/research/compbio/SAM_T02/T02-query.html	
SAM-T99[a]	http://www.soe.ucsc.edu/research/compbio/HMM-apps/T99-query.html	
SP3	http://sparks.informatics.iupui.edu	yqzhou@iupui.edu
SP4	http://sparks.informatics.iupui.edu	
SPARKS2	http://sparks.informatics.iupui.edu	
Zhang-Server	http://zhang.bioinformatics.ku.edu/I-TASSER	yzhang@ku.edu

The presented information about the participating servers was collected in a survey among the registered groups after the experiment.
[a]This server is obsolete and is being kept alive only for historical comparisons.

DOI 10.1002/prot

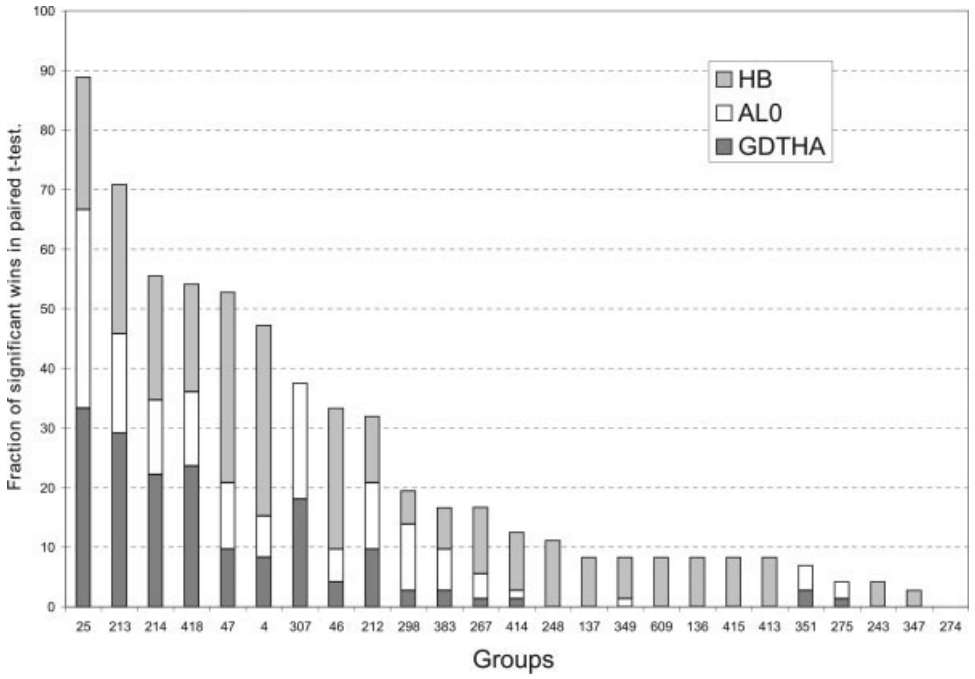

Figure 1

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

comparison with the best template identified using a purely sequence-based method.

The best-possible template is detected using a structure based search of the target against the PDB entries available up until the target's submission deadline as described elsewhere in this issue.[74] Specifically, the highest scoring structure according to the sequence-independent superposition generated by LGA[75] was defined to be the best template. "Pseudopredictions" were built based on LGA's structural alignment using a 4 Å superposition cutoff, whereby the coordinates of aligned residues were copied from the equivalent residues in the template.

For comparison, we used PSI-BLAST[76] to identify and align a template to the target sequence. The initial PSI-BLAST profile was generated for the target sequence on the NCBI nonredundant protein sequence database[77] and subsequently used to scan the PDB for templates available during the prediction window.

Using the lowest e-value as a criterion for choosing PSI-BLAST hits frequently identifies short fragments of very high similarity, but which give rise to models with a low GDT-HA due to the low coverage of the target. In the case of multidomain proteins, the e-value calculated on the basis of the whole sequence is not applicable for the individual subunits. Therefore, we filtered the PSI-BLAST hits by maximal coverage of the individual assessment units, and subsequently chose the highest ranked PSI-BLAST hits by e-value. Pseudopredictions were built by copying the backbone coordinates for all residues aligned between target and template in the PSI-BLAST alignment. No attempt was made to further improve the alignment or to model insertions or deletions.

For 41% of targets (44 of 108), the pseudopredictions built using the PSI-BLAST templates are within 10 GDT-HA units of the best structural template available. In these cases, the simple sequence search correctly identified a suitable structural template and little improvement would have been possible. For the remaining 64 targets, however, considerable improvement of prediction quality over that of the PSI-BLAST template would have been possible, namely by a margin of more than 10 GDT-HA points in 64 cases and even 20 points in 41 cases. However, it must be pointed out that for 30 of these 41 cases, the pseudopredictions based on the optimal structural template identified by LGA have GDT-HA scores below 50. These models are either incomplete or the templates are structurally divergent from the target.

We compared the performance of both sets of pseudopredictions, "PSI-Blast template" and "LGA template," to that of the best of all submitted server models, the best overall server (25 Zhang-server) and the best metapredictor server (47 Pmodeller6). Using the PSI-BLAST based model as baseline, we subtracted its GDT-HA value from that of the other predictions for each target and plotted the results in Figure 4 (upper panel). The Zhang-server

Table III

Statistical Significance of the Results of the 25 Highest Scoring Server Groups

	25	213	418	307	214	47	212	4	298	46	383	351	137	267	349	248	414	609	275	136	415	243	413	347	274
25		1.31(0.02) 3.20(0.00)	1.78(0.00) 3.68(0.00)	2.25(0.00) 3.08(0.00)	2.03(0.00) 4.10(0.00)	2.54(0.00) 4.36(0.00)	2.53(0.00) 4.09(0.00)	2.65(0.00) 4.47(0.00)	3.13(0.00) 4.24(0.00)	2.97(0.00) 4.96(0.00)	3.23(0.00) 4.78(0.00)	3.24(0.00) 5.09(0.00)	3.59(0.00) 6.82(0.00)	3.40(0.00) 6.25(0.00)	3.52(0.00) 5.98(0.00)	4.08(0.00) 6.43(0.00)	3.47(0.00) 6.17(0.00)	3.97(0.00) 6.45(0.00)	3.39(0.00) 4.76(0.00)	3.67(0.00) 7.18(0.00)	3.82(0.00) 6.64(0.00)	4.00(0.00) 6.18(0.00)	4.08(0.00) 7.01(0.00)	3.77(0.00) 7.10(0.00)	4.43(0.00) 6.75(0.00)
213	0.15(0.86) n = 108		0.47(0.03) 0.49(0.30)	0.98(0.10) -0.05(0.96)	0.55(0.16) 0.71(0.33)	1.27(0.02) 1.19(0.23)	1.22(0.00) 0.90(0.14)	1.36(0.01) 1.25(0.17)	1.84(0.00) 1.21(0.20)	1.64(0.00) 1.70(0.06)	1.56(0.00) 1.07(0.22)	1.92(0.00) 2.06(0.03)	2.27(0.00) 3.59(0.00)	2.03(0.00) 2.00(0.06)	2.05(0.00) 2.31(0.02)	2.77(0.00) 3.23(0.01)	2.20(0.00) 3.04(0.01)	2.36(0.00) 2.90(0.01)	1.75(0.00) 0.85(0.34)	2.36(0.00) 3.98(0.00)	2.52(0.00) 3.44(0.00)	2.20(0.00) 2.09(0.02)	2.78(0.00) 3.81(0.00)	2.38(0.00) 3.72(0.00)	2.74(0.00) 2.61(0.02)
418	0.74(0.39) n = 108	0.58(0.04) n = 108		0.49(0.38) -0.45(0.65)	0.09(0.79) 0.23(0.71)	0.81(0.16) 0.74(0.48)	0.75(0.09) 0.41(0.54)	0.88(0.11) 0.76(0.41)	1.35(0.01) 0.72(0.44)	1.16(0.04) 1.21(0.18)	1.18(0.04) 0.89(0.36)	1.43(0.01) 1.57(0.10)	1.79(0.00) 3.09(0.00)	1.55(0.01) 1.51(0.17)	1.55(0.01) 1.89(0.07)	2.29(0.00) 2.75(0.03)	1.72(0.00) 2.55(0.01)	1.92(0.00) 2.51(0.01)	1.42(0.01) 0.54(0.51)	1.89(0.00) 3.50(0.00)	2.04(0.00) 2.98(0.01)	1.85(0.00) 1.81(0.05)	2.30(0.00) 3.33(0.00)	1.93(0.00) 3.20(0.00)	2.22(0.00) 2.21(0.03)
307	5.31(0.00) n = 104	5.27(0.00) n = 104	4.63(0.00) n = 104		-0.31(0.57) 0.83(0.36)	0.44(0.39) 1.47(0.06)	0.29(0.62) 1.01(0.36)	0.54(0.31) 1.43(0.13)	0.90(0.10) 1.29(0.17)	0.85(0.13) 1.78(0.07)	0.83(0.18) 1.59(0.10)	0.91(0.08) 2.01(0.01)	1.45(0.01) 3.91(0.00)	1.14(0.03) 2.11(0.01)	1.28(0.03) 2.83(0.00)	1.93(0.00) 3.31(0.00)	1.28(0.01) 3.09(0.00)	1.64(0.00) 3.33(0.00)	1.15(0.02) 1.82(0.02)	1.50(0.02) 4.17(0.00)	1.59(0.00) 3.54(0.00)	1.56(0.01) 2.82(0.00)	1.90(0.00) 3.91(0.00)	1.55(0.01) 4.05(0.00)	2.17(0.00) 3.41(0.00)
214	0.32(0.71) n = 101	0.28(0.62) n = 101	-0.28(0.59) n = 101	**-4.76(0.00)** n = 97		0.72(0.20) 0.42(0.65)	0.78(0.17) 0.45(0.62)	0.57(0.28) 0.17(0.83)	1.18(0.01) 0.37(0.59)	1.08(0.05) 0.69(0.39)	1.32(0.04) 1.00(0.33)	1.16(0.02) 1.17(0.16)	1.47(0.02) 2.38(0.01)	1.42(0.02) 1.24(0.22)	1.45(0.01) 1.61(0.10)	2.18(0.00) 2.54(0.02)	1.60(0.00) 2.32(0.01)	1.90(0.00) 2.38(0.02)	1.25(0.01) 0.20(0.75)	1.63(0.01) 2.95(0.00)	1.83(0.00) 2.65(0.01)	1.92(0.00) 1.81(0.08)	2.17(0.00) 3.04(0.00)	1.93(0.00) 3.25(0.00)	2.29(0.00) 2.39(0.03)
47	**-1.95(0.01)** n = 107	**-1.98(0.01)** n = 107	**-2.56(0.00)** n = 107	**-7.04(0.00)** n = 103	**-2.11(0.01)** n = 100		-0.07(0.90) -0.39(0.73)	0.05(0.90) -0.03(0.97)	0.56(0.27) -0.01(0.99)	0.46(0.32) 0.51(0.56)	0.41(0.48) 0.11(0.91)	0.59(0.20) 0.81(0.27)	0.92(0.12) 2.26(0.02)	0.81(0.14) 0.81(0.37)	0.84(0.08) 1.36(0.08)	**1.48(0.01)** 1.96(0.06)	0.91(0.07) 1.79(0.04)	1.20(0.02) 1.98(0.03)	0.46(0.32) 0.10(0.90)	1.06(0.08) 2.71(0.00)	1.25(0.02) 2.26(0.02)	1.27(0.03) 1.56(0.10)	1.50(0.00) 2.53(0.00)	1.29(0.01) 2.80(0.00)	1.71(0.00) 2.18(0.04)
212	1.52(0.06) n = 108	1.37(0.01) n = 108	0.79(0.12) n = 108	**-3.83(0.00)** n = 104	1.01(0.15) n = 101	3.24(0.00) n = 107		0.14(0.79) 0.37(0.69)	0.61(0.19) 0.30(0.76)	0.40(0.45) 0.77(0.41)	0.31(0.49) 0.12(0.89)	0.67(0.20) 1.16(0.26)	1.04(0.08) 2.69(0.01)	0.81(0.09) 1.10(0.24)	0.82(0.13) 1.38(0.16)	1.55(0.01) 2.34(0.04)	0.98(0.06) 2.13(0.03)	1.18(0.02) 2.10(0.03)	0.57(0.26) 0.26(0.76)	1.14(0.06) 3.09(0.01)	1.29(0.02) 2.55(0.02)	0.96(0.03) 1.17(0.16)	1.55(0.01) 2.92(0.01)	1.13(0.02) 2.82(0.00)	1.36(0.01) 1.61(0.10)
4	**-1.91(0.03)** n = 107	**-2.08(0.01)** n = 107	**-2.67(0.00)** n = 107	**-7.16(0.00)** n = 103	**-2.57(0.00)** n = 100	-0.17(0.80) n = 108	**-3.50(0.00)** n = 107		0.39(0.41) -0.16(0.88)	0.36(0.40) 0.51(0.48)	0.31(0.56) 0.15(0.89)	0.49(0.25) 0.75(0.37)	0.97(0.10) 2.40(0.01)	0.74(0.15) 0.76(0.45)	0.72(0.13) 1.29(0.18)	1.46(0.01) 2.01(0.08)	0.82(0.11) 1.77(0.07)	0.84(0.10) 1.42(0.13)	0.50(0.26) -0.96(0.64)	1.08(0.06) 2.81(0.01)	1.16(0.03) 2.19(0.02)	1.14(0.02) 1.44(0.13)	1.44(0.01) 2.57(0.01)	1.22(0.01) 2.81(0.00)	1.64(0.00) 2.02(0.06)
298	3.10(0.00) n = 107	3.01(0.00) n = 107	2.38(0.00) n = 107	**-2.12(0.00)** n = 104	2.43(0.00) n = 100	4.88(0.00) n = 108	1.62(0.01) n = 107	5.14(0.00) n = 106		-0.13(0.77) 0.56(0.47)	-0.00(0.99) 0.63(0.52)	0.04(0.92) 0.83(0.23)	0.44(0.41) 2.45(0.01)	0.23(0.62) 0.78(0.37)	0.38(0.40) 1.56(0.08)	0.95(0.05) 2.06(0.03)	0.35(0.44) 1.74(0.03)	0.64(0.23) 2.18(0.03)	0.17(0.71) 0.29(0.72)	0.55(0.32) 2.85(0.00)	0.70(0.13) 2.27(0.02)	0.80(0.14) 1.96(0.07)	0.96(0.04) 2.64(0.00)	0.73(0.13) 2.92(0.00)	1.00(0.07) 1.97(0.06)
46	-0.27(0.70) n = 107	-0.47(0.57) n = 107	-1.09(0.19) n = 107	**-5.50(0.00)** n = 103	-0.65(0.45) n = 100	1.56(0.03) n = 108	**-1.90(0.02)** n = 107	1.71(0.05) n = 106	**-3.46(0.00)** n = 106		0.09(0.86) -0.05(0.95)	0.27(0.57) 0.32(0.71)	0.64(0.23) 1.95(0.03)	0.36(0.48) 0.25(0.80)	0.50(0.33) 0.97(0.33)	1.09(0.04) 1.49(0.18)	0.52(0.26) 1.33(0.13)	0.75(0.06) 1.35(0.10)	0.36(0.43) -0.31(0.68)	0.74(0.15) 2.35(0.01)	0.85(0.07) 1.74(0.08)	0.86(0.07) 1.26(0.18)	1.11(0.03) 2.13(0.03)	0.87(0.06) 2.29(0.01)	1.48(0.01) 1.91(0.06)
383	2.83(0.00) n = 103	2.68(0.00) n = 103	2.07(0.01) n = 103	**-2.58(0.00)** n = 101	2.33(0.01) n = 96	4.62(0.00) n = 102	1.23(0.06) n = 103	4.79(0.00) n = 102	-0.31(0.62) n = 103	3.37(0.00) n = 102		0.08(0.87) 0.31(0.70)	0.48(0.43) 2.20(0.03)	0.30(0.50) 0.52(0.49)	0.24(0.59) 0.80(0.28)	1.04(0.04) 1.67(0.07)	0.35(0.51) 1.28(0.14)	0.61(0.18) 1.42(0.07)	0.30(0.57) 0.19(0.81)	0.52(0.36) 2.34(0.01)	0.68(0.18) 1.73(0.06)	0.60(0.02) 0.94(0.03)	0.96(0.07) 2.12(0.02)	0.71(0.15) 2.35(0.00)	1.05(0.06) 1.40(0.13)
351	4.73(0.00) n = 108	4.59(0.00) n = 108	3.96(0.00) n = 108	-0.44(0.39) n = 103	4.07(0.00) n = 99	6.45(0.00) n = 105	3.13(0.00) n = 108	6.78(0.00) n = 105	1.60(0.00) n = 108	4.87(0.00) n = 105	1.89(0.00) n = 102		0.37(0.50) 1.58(0.10)	0.16(0.72) -0.11(0.88)	0.26(0.51) 0.64(0.37)	0.85(0.09) 1.17(0.18)	0.32(0.49) 1.01(0.18)	0.58(0.30) 1.11(0.21)	0.08(0.87) -0.55(0.48)	0.50(0.37) 1.99(0.02)	0.68(0.18) 1.43(0.08)	0.72(0.16) 1.02(0.24)	0.94(0.04) 1.85(0.02)	0.69(0.12) 2.02(0.01)	1.07(0.04) 1.18(0.21)
137	2.19(0.02) n = 107	2.01(0.01) n = 107	1.40(0.06) n = 107	**-3.15(0.00)** n = 103	1.56(0.03) n = 100	3.85(0.00) n = 108	0.64(0.36) n = 107	4.24(0.00) n = 106	-1.02(0.13) n = 106	2.59(0.00) n = 106	-0.89(0.31) n = 102	**-2.60(0.00)** n = 105		-0.24(0.63) -1.57(0.08)	-0.04(0.94) -0.97(0.27)	0.55(0.27) -0.23(0.81)	-0.07(0.89) -0.53(0.54)	0.19(0.74) -0.55(0.58)	0.45(0.40) -1.09(0.21)	0.11(0.68) 0.46(0.43)	0.31(0.55) -0.01(0.99)	0.35(0.62) -0.66(0.48)	0.55(0.20) 0.34(0.69)	0.30(0.52) 0.59(0.43)	0.87(0.09) -0.03(0.96)
267	1.73(0.04) n = 107	1.50(0.05) n = 107	0.98(0.22) n = 107	**-3.54(0.00)** n = 103	1.30(0.11) n = 100	3.51(0.00) n = 108	0.20(0.76) n = 107	3.70(0.00) n = 106	**-1.43(0.01)** n = 106	2.02(0.01) n = 106	-1.18(0.08) n = 102	**-3.01(0.00)** n = 105	-0.55(0.38) n = 106		0.18(0.70) 0.68(0.33)	0.72(0.09) 1.21(0.11)	0.08(0.84) 0.96(0.17)	0.37(0.42) 1.18(0.19)	-0.06(0.90) -0.48(0.57)	0.34(0.51) 2.04(0.01)	0.45(0.30) 1.41(0.07)	0.52(0.25) 0.81(0.34)	0.72(0.09) 1.78(0.01)	0.46(0.26) 1.85(0.02)	0.84(0.05) 1.07(0.16)
349	2.97(0.00) n = 105	2.97(0.00) n = 105	2.30(0.00) n = 105	**-2.09(0.00)** n = 101	2.67(0.00) n = 99	4.65(0.00) n = 104	1.55(0.02) n = 105	4.91(0.00) n = 104	0.02(0.98) n = 104	3.41(0.00) n = 104	0.11(0.87) n = 101	**-1.66(0.00)** n = 103	0.99(0.13) n = 104	1.40(0.01) n = 104		0.60(0.18) 0.66(0.44)	-0.02(0.96) 0.30(0.68)	0.25(0.64) 0.45(0.60)	-0.01(0.99) -0.67(0.39)	0.18(0.73) 1.33(0.08)	0.33(0.52) 0.77(0.36)	0.44(0.35) 0.31(0.71)	0.59(0.14) 1.15(0.09)	0.39(0.26) 1.48(0.02)	0.58(0.16) 0.24(0.75)
248	1.41(0.12) n = 108	1.26(0.14) n = 108	0.68(0.43) n = 108	**-3.75(0.00)** n = 104	1.07(0.21) n = 101	3.18(0.00) n = 107	-0.11(0.89) n = 108	3.40(0.00) n = 107	**-1.74(0.02)** n = 107	1.72(0.05) n = 107	-1.50(0.07) n = 103	**-3.31(0.00)** n = 106	-0.72(0.31) n = 107	-0.31(0.62) n = 107	**-1.72(0.01)** n = 105		-0.64(0.20) -0.22(0.79)	-0.46(0.38) -0.27(0.78)	-0.44(0.34) -0.97(0.28)	-0.40(0.42) 0.75(0.38)	-0.25(0.62) 0.21(0.82)	-0.40(0.44) -0.59(0.55)	0.01(0.99) 0.58(0.46)	-0.22(0.63) 0.78(0.37)	0.11(0.81) -0.04(0.96)
414	1.93(0.01) n = 107	1.84(0.01) n = 107	1.25(0.06) n = 107	**-3.37(0.00)** n = 104	1.46(0.04) n = 100	3.61(0.00) n = 108	0.45(0.49) n = 107	3.88(0.00) n = 106	**-1.24(0.04)** n = 106	2.20(0.00) n = 106	-0.97(0.18) n = 102	**-2.85(0.00)** n = 105	-0.16(0.81) n = 106	0.18(0.77) n = 106	-1.13(0.06) n = 104	0.45(0.54) n = 107		0.34(0.41) 0.18(0.83)	0.04(0.93) -0.97(0.14)	0.20(0.69) 1.01(0.19)	0.35(0.28) 0.44(0.42)	0.41(0.44) 0.01(0.99)	0.61(0.02) 0.81(0.04)	0.29(0.46) 0.83(0.19)	0.73(0.08) -0.02(0.98)
609	2.80(0.00) n = 99	2.33(0.00) n = 99	1.80(0.01) n = 99	**-2.62(0.00)** n = 96	2.07(0.01) n = 92	4.34(0.00) n = 98	0.92(0.11) n = 99	4.17(0.00) n = 99	-0.65(0.32) n = 99	2.71(0.00) n = 98	-0.54(0.47) n = 98	**-1.97(0.00)** n = 98	0.42(0.56) n = 98	0.68(0.28) n = 98	-0.53(0.43) n = 96	0.84(0.23) n = 99	0.69(0.25) n = 98		-0.22(0.68) -1.04(0.21)	-0.10(0.85) 0.92(0.26)	0.25(0.49) 0.82(0.28)	0.09(0.84) -0.33(0.84)	0.31(0.50) 0.61(0.48)	0.06(0.90) 0.91(0.24)	0.47(0.38) 0.18(0.84)
275	5.00(0.00) n = 90	5.23(0.00) n = 90	4.81(0.00) n = 90	0.36(0.40) n = 86	5.20(0.00) n = 86	6.76(0.00) n = 89	3.92(0.00) n = 90	7.10(0.00) n = 89	2.01(0.00) n = 89	5.58(0.00) n = 89	2.19(0.00) n = 87	0.66(0.17) n = 88	4.05(0.00) n = 89	3.75(0.00) n = 89	2.81(0.00) n = 90	4.59(0.00) n = 90	3.79(0.00) n = 89	2.98(0.00) n = 84		-0.41(0.42) 1.19(0.13)	0.29(0.58) 1.39(0.10)	0.16(0.75) 0.52(0.47)	0.46(0.29) 1.84(0.03)	0.06(0.83) 1.64(0.00)	0.74(0.01) 0.98(0.10)
136	2.20(0.02) n = 108	2.05(0.00) n = 108	1.46(0.03) n = 108	**-3.04(0.00)** n = 104	1.58(0.02) n = 101	3.95(0.00) n = 107	0.67(0.32) n = 108	4.23(0.00) n = 107	-0.98(0.16) n = 107	2.59(0.00) n = 107	-0.87(0.31) n = 103	**-2.52(0.00)** n = 105	0.03(0.93) n = 107	0.54(0.41) n = 107	-0.98(0.11) n = 105	0.78(0.26) n = 108	0.28(0.66) n = 107	-0.43(0.50) n = 99	**-4.19(0.00)** n = 90		0.15(0.77) -0.54(0.52)	0.27(0.59) -1.07(0.17)	0.41(0.34) -0.17(0.80)	0.19(0.66) 0.02(0.98)	0.82(0.10) -0.47(0.52)
415	2.22(0.01) n = 108	2.07(0.01) n = 108	1.49(0.04) n = 108	**-3.07(0.00)** n = 104	1.63(0.04) n = 101	3.95(0.00) n = 107	0.70(0.28) n = 108	4.19(0.00) n = 107	-0.94(0.08) n = 107	2.51(0.00) n = 107	-0.83(0.23) n = 103	**-2.48(0.00)** n = 106	0.09(0.88) n = 107	0.48(0.40) n = 107	-0.92(0.14) n = 105	0.81(0.26) n = 108	0.31(0.42) n = 107	-0.26(0.63) n = 99	**-3.64(0.00)** n = 90	0.02(0.97) n = 108		0.11(0.84) -0.45(0.83)	0.26(0.47) 0.37(0.56)	-0.07(0.87) 0.43(0.56)	0.04(0.92) -1.08(0.12)
243	4.09(0.00) n = 101	3.84(0.00) n = 101	3.28(0.00) n = 101	**-1.49(0.02)** n = 97	3.71(0.00) n = 96	5.68(0.00) n = 100	2.39(0.00) n = 101	5.75(0.00) n = 100	0.91(0.13) n = 100	4.40(0.00) n = 100	0.88(0.02) n = 99	-0.68(0.20) n = 99	2.04(0.01) n = 100	2.37(0.00) n = 100	0.95(0.13) n = 100	2.39(0.00) n = 101	2.14(0.00) n = 100	1.67(0.02) n = 93	**-1.72(0.01)** n = 86	2.09(0.01) n = 101	1.96(0.00) n = 101		0.24(0.64) 1.02(0.26)	-0.05(0.91) 1.13(0.11)	0.50(0.35) 0.71(0.42)
413	2.58(0.00) n = 108	2.42(0.00) n = 108	1.84(0.01) n = 108	**-2.66(0.00)** n = 104	2.03(0.01) n = 101	4.32(0.00) n = 107	1.05(0.11) n = 108	4.57(0.00) n = 107	-0.58(0.34) n = 107	2.91(0.00) n = 107	-0.39(0.59) n = 103	**-2.14(0.00)** n = 106	0.46(0.42) n = 107	0.83(0.15) n = 107	-0.54(0.33) n = 105	1.16(0.10) n = 108	0.61(0.07) n = 107	-0.05(0.94) n = 99	**-3.31(0.00)** n = 90	0.38(0.49) n = 108	0.36(0.40) n = 108	**-1.52(0.03)** n = 101		-0.28(0.45) 0.02(0.98)	0.03(0.94) -0.82(0.20)
347	4.27(0.00) n = 105	3.99(0.00) n = 105	3.42(0.00) n = 105	-0.91(0.06) n = 101	4.09(0.00) n = 98	6.16(0.00) n = 104	2.51(0.00) n = 105	6.39(0.00) n = 104	1.00(0.06) n = 104	4.68(0.00) n = 104	1.21(0.06) n = 101	-0.40(0.43) n = 103	2.12(0.00) n = 105	2.55(0.00) n = 104	1.33(0.01) n = 102	2.98(0.00) n = 105	2.43(0.00) n = 104	1.74(0.01) n = 97	**-1.53(0.00)** n = 88	2.17(0.00) n = 105	2.02(0.00) n = 105	0.34(0.57) n = 99	1.61(0.00) n = 105		0.46(0.11) -0.70(0.16)
274	5.64(0.00) n = 96	5.26(0.00) n = 96	4.76(0.00) n = 96	0.47(0.33) n = 92	5.30(0.00) n = 90	7.45(0.00) n = 96	3.99(0.00) n = 96	7.61(0.00) n = 95	2.48(0.00) n = 95	6.16(0.00) n = 95	2.81(0.00) n = 92	1.01(0.06) n = 94	3.71(0.00) n = 96	4.02(0.00) n = 96	2.60(0.00) n = 94	4.10(0.00) n = 96	3.98(0.00) n = 95	3.31(0.00) n = 90	0.13(0.75) n = 82	3.74(0.00) n = 96	3.84(0.00) n = 96	1.79(0.00) n = 91	3.09(0.00) n = 98	1.51(0.00) n = 95	

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, AL-0, and HBscore values along with the associated P-values in parentheses. Cells above the diagonal in the top right part of the table provide values for GDT-HA (upper half of each cell) and AL0 (lower half of each cell) comparing rows with columns. Cells below the diagonal in the lower left part of the table provide values for HBscore (upper half of each cell) comparing columns with rows, and the number of common models in the pair wise comparison (lower half of each cell). Statistically significant differences between groups (P-values < 0.05) are shaded in gray.

DOI 10.1002/prot

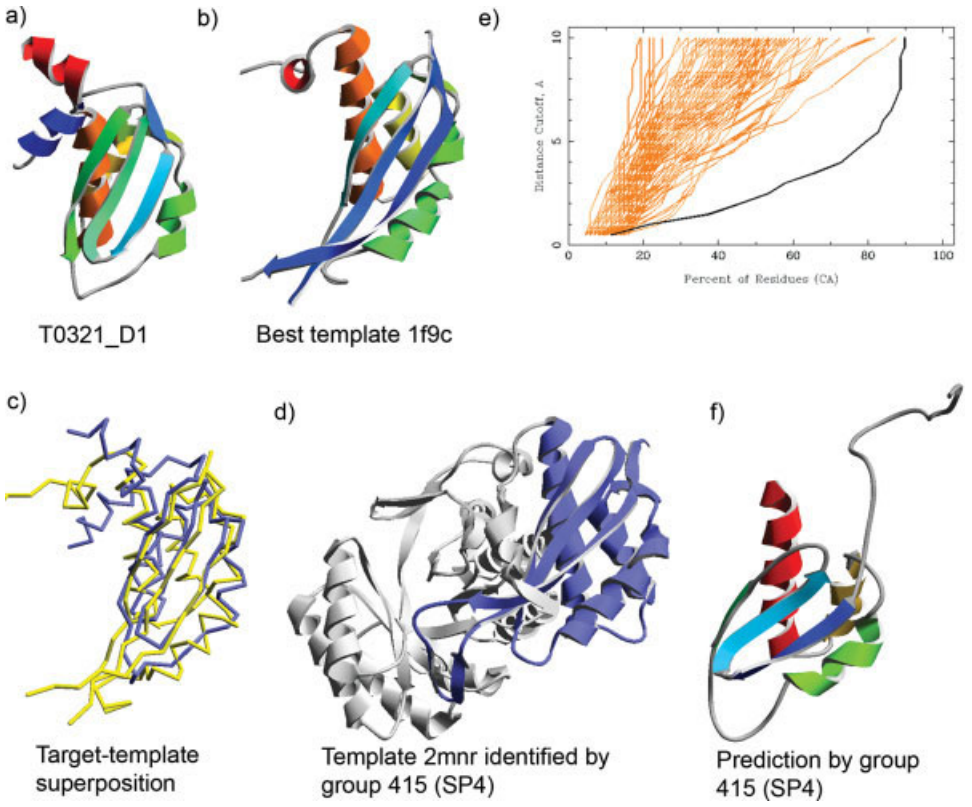

Figure 2

Prediction example of target T0321 domain 1. For discussion, see the main document.

was able to improve over the PSI-BLAST template by a margin of over 10 GDT-HA points in 46 cases, and by a margin of more than 20 points in 23 cases. It was also able to build models of comparable quality to the "LGA-predictor" in the majority of cases. Only in 23 cases, the server failed to get within 10 GDT-HA points of the LGA-based pseudopredictor, and the predictions were more than 20 GDT-HA points lower in eight cases.

As is apparent from Figure 4 (lower panel), the performance of the PSI-BLAST approach is only weakly correlated with the overall structural similarity of the best available template (correlation coefficient [GDT-HA(LGA-pseu-

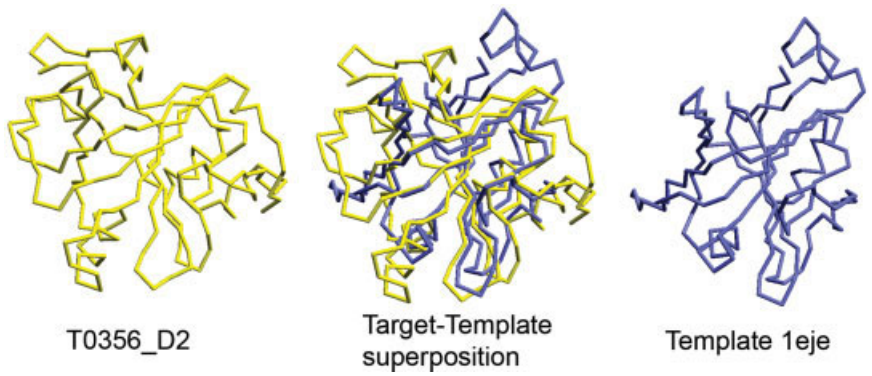

Figure 3

Prediction example of target T0356 domain 2. For discussion, see main document.

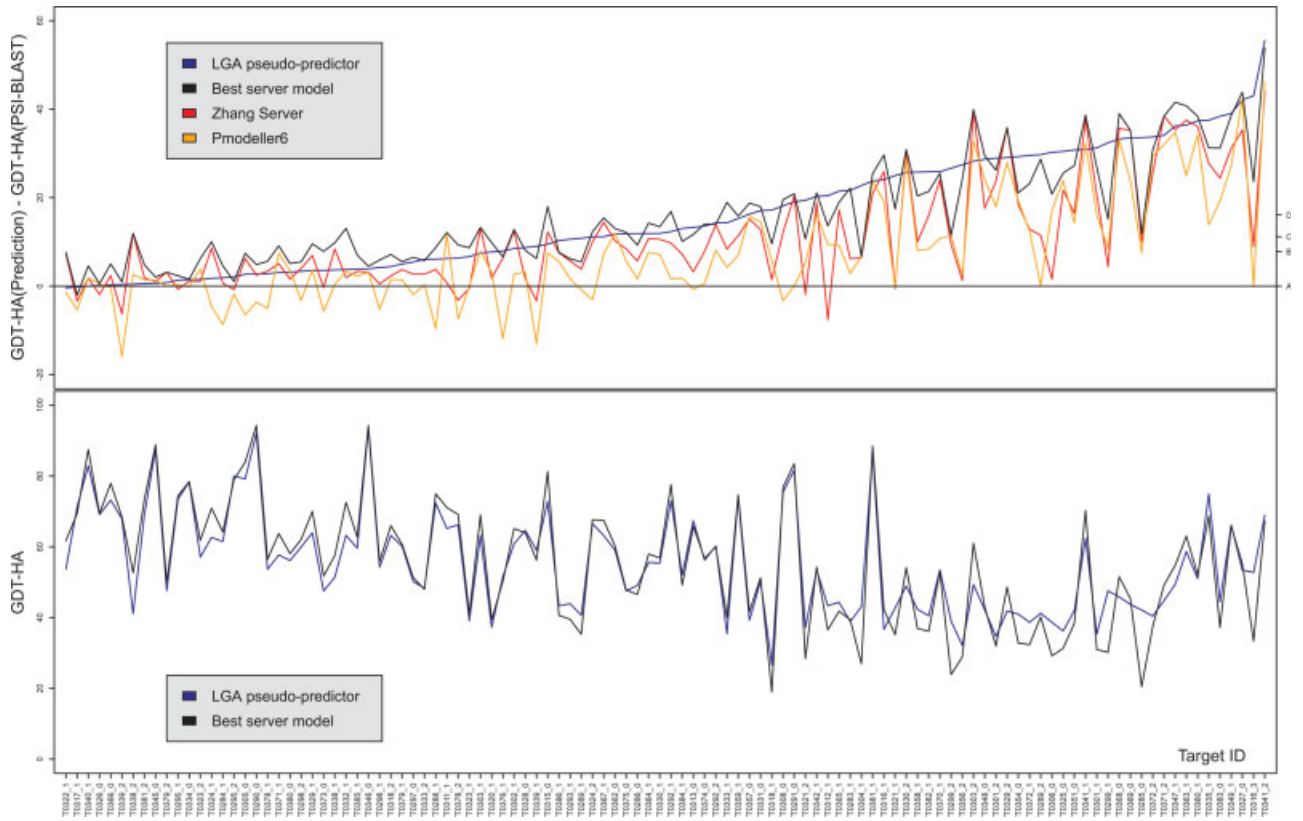

Figure 4

Server performance compared with the two pseudopredictors. Above: For each target, the performance of the different predictors is plotted relative to the GDT-HA baseline defined by the "PSI-BLAST" pseudopredictor, i.e., values on the vertical axis are GDT-HA values minus the GDT-HA of the pseudoprediction based on PSI-BLAST. Points below the x-axis performed worse than a naïve PSI-BLAST alignment would have fared. Targets are ordered by increasing difference in GDT-HA between pseudopredictions based on templates identified by LGA and PSI-BLAST, respectively. Predictions by group 25 (Zhang-server) are shown in red, 47 (Pmodeller6) in orange, and the "LGA pseudopredictor" in blue. The best of all models submitted by any server group are shown in black. For comparison, the average performance of the methods is indicated on the right side of the plot: (A) PSI-BLAST pseudo predictor, (B) Pmodeller6, (C) Zhang-server, and (D) LGA-pseudopredictor. Below: For reference, the absolute GDT-HA value is plotted for the "LGA pseudopredictor" (blue) and the best server submission per target in retrospect (black).

domodel) − GDT-HA(PSI-BLAST-pseudomodel] versus GDT-HA(LGA-pseudomodel) = −0.42). There were numerous cases where good structural templates were available (as identified by LGA), but were missed by PSI-BLAST.

Some cases, in which good templates existed but could not be detected easily, have been identified and rationalized. One example of such a notable improvement over the PSI-BLAST performance is T0349_D1, a NMR structure of the hypothetical protein RPA1041 from *Pseudomonas aeruginosa*. A model based on the best available template, a secretory protein of the YscJ/FliF family part of the *E. coli* type III secretion system (PDB:1yj7, chain D), has a GDT-HA score of 66.2. The template detected by PSI-BLAST, the structure of the L-aminopeptidase D-ala-esterase/amidase from *Ochrobactrum anthropi* (PDB: 1b65), has a high target coverage of 72%, but an unfavorable e-value and is structurally unrelated. Consequently, the resulting model achieves a GDT-HA value of only 27.2. The reason for PSI-BLAST's inability to detect a better template may be due to the meager yield of hits during the profile creation step, which consequently leads to a poor profile for further scanning. It is noteworthy that some of the better performing servers, e.g., HHpred1, work by matching profiles generated from the target and the potential templates. Presumably, this two-sided approach allows the sequence gap to be bridged successfully where the one-sided PSI-BLAST method fails.

SERVER PREDICTIONS IN THE HA CATEGORY

One might expect that servers will do well when evolutionary relationships and sequence alignments are clear, as is generally the case for the structures in the HA/TBM category. To enter this category, it is necessary that there be a good template, and in most cases, such templates

DOI 10.1002/prot

were successfully identified. Indeed, the trend for servers to perform well continued in this category.

To obtain a rough overall ranking of the groups submitting predictions for the HA/TBM category, they were sorted by the sum of the scores for GDT-HA, prediction of side-chain χ_1/χ_2 angles, and suitability for use as models to solve the target crystal structures by molecular replacement.[68] Of the top 25 groups, four are servers. Two of the servers are among the top five, with similar overall scores placing them at positions 4 (186 CaspItaFOX) and 5 (4 Robetta). However, in contrast to the results for the general TBM category, Zhang-server did not appear among the top groups, coming in at position 43. This ranking is strongly influenced by the performance in rotamer prediction and molecular replacement, where CaspItaFOX and Robetta were both highly ranked. A number of servers did extremely well judged by the more traditional GDT-HA score, but less well against the other two criteria. In fact, when the predictors are ranked by GDT-HA alone, servers are found at positions 3 (136 FOLDpro), 4 (25 Zhang-server), and 5 (137 3Dpro).

SERVER PREDICTIONS IN THE FM CATEGORY

Servers might be expected to be at a greater disadvantage in FM than in TBM. For one thing, human insight can still be useful in modeling protein structures, and it would seem that there are more opportunities for such intervention in FM than in TBM. Second, the time limit for server predictions might be more of a constraint for FM targets than for TBM. Nevertheless, servers did rather well on FM targets in CASP7.

Although the best group in FM was clearly a human group (20 Baker), the next set of groups includes servers. On the basis of the pairwise GDT-TS comparison, the Zhang-server (Group 25) was the third best FM predictor, behind only the human-aided predictions of Baker and Zhang himself (Group 24). By visual assessment, two servers (25 Zhang-server and 4 Robetta) were in a four-way tie for second place according to the criteria and scoring scheme applied in the FM category of CASP7.[66] Combining GDT-TS assessment and visual assessment, there were six groups that were among the top 20 by GDT-TS and among the top 10 by visual assessment. Among these six groups, three were servers (Zhang-server, Robetta, and the metaserver group 47 Pmodeller6).

A particularly striking server prediction was the Robetta model 2 for target T0350, which has an antiparallel three-stranded sheet and three flanking helices that lie side by side against one surface of the sheet. Although a relatively simple αβ-structure, the arrangement of the sheet and helices is unusual and appears to be a true FM target. Nevertheless, many groups did astonishingly well on this target, and of the models that were judged best by visual assessment, Model 2 from Robetta had the highest GDT-TS rank. For two other targets (T0287 and T0314), the Zhang-server had the single best model by GDT-TS.

DISCUSSION

In the recent years, structure prediction in CASP had been dominated by human predictors using computational modeling procedures, which require manual intervention at many steps in the process. One of the recurring problems in CASP assessments has always been discerning the improvements contributed by human intervention from the purely computational element. In the assessment of CASP2, Thornton and coworkers used the automated modeling service SWISS-MODEL as reference to establish a baseline of what could be achieved by fully automated prediction.[78,79] During the following experiments, numerous predictors registered their methods as servers and were assessed separately as part of the CAFASP experiment series.[8,9] Starting with CASP6, server predictions were assessed as part of the main CASP experiment.[10] The general opinion in the community has been that the "human plus machine predictions" are superior to automated ones.[8] However, it appears that with CASP7, this view might have to be revised as 6 of the top 25 groups in the TBM category assessment were server predictors. Overall, in both CASP5 and CASP6 servers provided the best models (or tied with humans) for 7% of targets as measured by GDT-TS.[7] In contrast, for the 123 target domains in CASP7, the best model (or tied with humans) was submitted by a server in 29% of the cases. The best prediction server (25 Zhang-server) was ranked third over all, i.e. it outperformed all but two of the participating groups in the TBM category.

With the emergence of many new protein structure prediction servers, based on diverse methods with different strengths and weaknesses, the question of selecting the most appropriate server for each target became prevailing. Metaservers—methods that use the results of other servers as input to generate their predictions—were expected to have the capacity to outperform all individual autonomous servers, and challenge most human expert predictors.[80] To illustrate to what extent metaservers were able to select favorable models from the pool of the CASP7 server models, we included in Figure 4 the best-performing metaserver (47 Pmodeller6), the best individual server (25 Zhang-server), and the best of all submitted server models. Remarkably in CASP7, the best individual autonomous server (25 Zhang-server) outperformed the best metaserver (47 Pmodeller6), as well as the best manual metapredictor group (675 Famsace).[67] Comparing all server methods among themselves (Fig. 1), the top ranking groups are not metamethods,

indicating that the development of individual servers was fruitful in the recent years, and has significantly contributed to advancing the field of protein structure prediction.

There has been much debate on how to objectively compare human and server predictions. Much effort is expended during CASP by human predictors to gather information from a diverse set of resources such as scientific literature and specialist databases. This information may contribute greatly to the quality of a model for example by improving template selection and alignment or identifying ligand-binding sites. As this work, however, is very time consuming, it is only feasible for a relatively small set of targets. By contrast, prediction servers are fully automated and do not have these limitations, therefore their performance can be evaluated using a larger sample size than would be possible with human predictors. Several projects have been initiated with the aim of continuous, large-scale assessment, such as LiveBench[80] and EVA,[81] which use a sample size not tractable for nonautomated prediction methods. The CASP7 experiment comprised the relatively large number of 100 prediction targets, providing a solid basis for the numerical and statistical analysis. In this respect, CASP7 reflects a "real life" situation, where one is faced with the problem of modeling structures for an exponentially growing number of protein sequences. Although the large number of targets led to a technical advantage for automated methods over human predictors, the server predictions for all targets were publicly available early in the prediction window. Thus, most of the "routine" work had already been done automatically and human predictors could focus their efforts on improving the automated predictions using expert knowledge.

The gap between human predictors and servers is closing as automated prediction servers have come of age. However, the fundamental limitations for both human and server predictors remain in modeling of loops and effective refinement techniques. The observed progress in automated, reproducible, and scalable prediction methods in CASP7 holds the promise for further improvements in the future.

ACKNOWLEDGMENTS

We thank the team at the Prediction Center at UC Davis, especially Andriy Kryshtafovych, for their professional support, and Michael Tress for providing LGA template information. We also thank Anna Tramontano, Alfonso Valencia, Burkhard Rost, Tim Hubbard, and John Moult for fruitful and encouraging discussions. We thank all server groups who participated in the survey for this manuscript, and last but not least, we acknowledge all participating experimental groups, CASP7 predictors, and server developers, without whom CASP would not be possible.

REFERENCES

1. Friedberg I, Jaroszewski L, Ye Y, Godzik A. The interplay of fold recognition and experimental structure determination in structural genomics. Curr Opin Struct Biol 2004;14:307–312.
2. Marsden RL, Lewis TA, Orengo CA. Towards a comprehensive structural coverage of completed genomes: a structural genomics viewpoint. BMC Bioinformatics 2007;8:86.
3. Fox JA, McMillan S, Ouellette BFF. A compilation of molecular biology web servers: 2006 update on the Bioinformatics Links Directory. Nucleic Acids Res 2006;34(Suppl 2):W3–W5.
4. Pieper U, Eswar N, Braberg H, Madhusudhan MS, Davis FP, Stuart AC, Mirkovic N, Rossi A, Marti-Renom MA, Fiser A, Webb B, Greenblatt D, Huang CC, Ferrin TE, Sali A. MODBASE, a database of annotated comparative protein structure models, and associated resources. Nucleic Acids Res 2004;32(Suppl 1):D217–D222.
5. Kopp J, Schwede T. The SWISS-MODEL repository: new features and functionalities. Nucleic Acids Res 2006;34(Suppl 1):D315–D318.
6. Castrignano T, De Meo PD, Cozzetto D, Talamo IG, Tramontano A. The PMDB protein model database. Nucleic Acids Res 2006;34 (Database Issue):D306–D309.
7. Moult J. A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction. Curr Opin Struct Biol 2005;15:285–289.
8. Fischer D, Barret C, Bryson K, Elofsson A, Godzik A, Jones D, Karplus KJ, Kelley LA, MacCallum RM, Pawowski K, Rost B, Rychlewski L, Sternberg M. CAFASP-1: critical assessment of fully automated structure prediction methods. Proteins 1999;(Suppl 3):209–217.
9. Fischer D, Rychlewski L, Dunbrack RL, Jr, Ortiz AR, Elofsson A. CAFASP3: the third critical assessment of fully automated structure prediction methods. Proteins 2003;53(Suppl 6):503–516.
10. Moult J, Fidelis K, Rost B, Hubbard T, Tramontano A. Critical assessment of methods of protein structure prediction (CASP)–round 6. Proteins 2005;61(Suppl 7):3–7.
11. Kryshtafovych A, Venclovas C, Fidelis K, Moult J. Progress over the first decade of CASP experiments. Proteins 2005;61(Suppl 7):225–236.
12. Bates PA, Kelley LA, MacCallum RM, Sternberg MJ. Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. Proteins 2001;45(Suppl 5):39–46.
13. Offman MN, Fitzjohn PW, Bates PA. Developing a move-set for protein model refinement. Bioinformatics 2006;22:1838–1845.
14. Contreras-Moreira B, Fitzjohn PW, Offman M, Smith GR, Bates PA. Novel use of a genetic algorithm for protein structure prediction: searching template and sequence alignment space. Proteins 2003;53(Suppl 6):424–429.
15. Cheng J, Randall AZ, Sweredoski MJ, Baldi P. SCRATCH: a protein structure and structural feature prediction server. Nucleic Acids Res 2005;33(Web Server Issue):W72–W76.
16. Kim DE, Chivian D, Malmstrom L, Baker D. Automated prediction of domain boundaries in CASP6 targets using Ginzu and RosettaDOM. Proteins 2005;61(Suppl 7):193–200.
17. Cheng J, Baldi P. Three-stage prediction of protein β-sheets by neural networks, alignments and graph algorithms. Bioinformatics 2005;21(Suppl 1):i75–i84.
18. Ishida T, Nishimura T, Nozaki M, Inoue T, Terada T, Nakamura S, Shimizu K. Development of an ab initio protein structure prediction system ABLE. Genome Inform 2003;14:228–237.
19. Su CT, Chen CY, Ou YY. Protein disorder prediction by condensed PSSM considering propensity for order or disorder. BMC Bioinformatics 2006;7:319.
20. Lund O, Nielsen M, Lundegaard C, Worning P. CPHmodels 2.0: X3M a Computer Program to Extract 3D Models. In: CASP5: Proceedings of the 5th meeting on the critical assessment of techniques for protein structure prediction, 1–5 December 2002, Asilomar, CA.

 DOI 10.1002/prot

21. Ward JJ, McGuffin LJ, Bryson K, Buxton BF, Jones DT. The DISOPRED server for the prediction of protein disorder. Bioinformatics 2004;20:2138–2139.
22. Cheng J, Sweredoski MJ, Baldi P. Accurate prediction of protein disordered regions by mining protein structure data. Data Min Knowledge Discov 2005;11:213–222.
23. Bau D, Martin AJ, Mooney C, Vullo A, Walsh I, Pollastri G. Distill: a suite of web servers for the prediction of one-, two- and three-dimensional structural features of proteins. BMC Bioinformatics 2006;7:402.
24. Bryson K, Cozzetto D, Jones DT. Computer-assisted protein domain boundary prediction using the DomPred server. Curr Protein Pept Sci 2007;8:181–188.
25. Ogata K, Umeyama H. An automatic homology modeling method consisting of database searches and simulated annealing. J Mol Graph Model 2000;18:258–272, 305–256.
26. Cheng J, Baldi P. A machine learning information retrieval approach to protein fold recognition. Bioinformatics 2006;22:1456–1463.
27. Tomii K, Akiyama Y. FORTE: a profile-profile comparison tool for protein fold recognition. Bioinformatics 2004;20:594–595.
28. Tomii K, Hirokawa T, Motono C. Protein structure prediction using a variety of profile libraries and 3D verification. Proteins 2005;61 (Suppl 7):114–121.
29. Kosinski J, Gajda MJ, Cymerman IA, Kurowski MA, Pawlowski M, Boniecki M, Obarska A, Papaj G, Sroczynska-Obuchowicz P, Tkaczuk KL, Sniezynska P, Sasin JM, Augustyn A, Bujnicki JM, Feder M. FRankenstein becomes a cyborg: the automatic recombination and realignment of fold recognition models in CASP6. Proteins 2005;61(Suppl 7):106–113.
30. Shi J, Blundell TL, Mizuguchi K. FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. J Mol Biol 2001;310: 243–257.
31. Kurowski MA, Bujnicki JM. GeneSilico protein structure prediction meta-server. Nucleic Acids Res 2003;31:3305–3307.
32. MacCallum RM. Striped sheets and protein contact prediction. Bioinformatics 2004;20(Suppl 1):I224–I231.
33. Soding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res 2005;33(Web Server Issue):W244–W248.
34. Torda AE, Procter JB, Huber T. Wurst: a protein threading server with a structural scoring function, sequence profiles and optimized substitution matrices. Nucleic Acids Res 2004;32 (Web Server Issue):W532–W535.
35. Rangwala H, Karypis G. Profile-based direct kernels for remote homology detection and fold recognition. Bioinformatics 2005;21: 4239–4247.
36. Rangwala H, Karypis G. Building multiclass classifiers for remote homology detection and fold recognition. BMC Bioinformatics 2006;7:455.
37. Kalisman N, Levi A, Maximova T, Reshef D, Zafriri-Lynn S, Gleyzer Y, Keasar C. MESHI: a new library of Java classes for molecular modeling. Bioinformatics 2005;21:3931–3932.
38. Meller J, Elber R. Linear programming optimization and a double statistical filter for protein threading protocols. Proteins 2001;45: 241–261.
39. Teodorescu O, Galor T, Pillardy J, Elber R. Enriching the sequence substitution matrix by structural information. Proteins 2004;54:41–48.
40. Tobi D, Elber R. Distance-dependent, pair potential for protein folding: results from linear optimization. Proteins 2000;41:40–46.
41. Saini HK, Fischer D. Meta-DP: domain prediction meta-server. Bioinformatics 2005;21:2917–2920.
42. Pandit SB, Zhang Y, Skolnick J. TASSER-Lite: an automated tool for protein comparative modeling. Biophys J 2006;91:4180–4190.
43. Zhang Y, Skolnick J. Automated structure prediction of weakly homologous proteins on a genomic scale. Proc Natl Acad Sci USA 2004;101:7594–7599.
44. Zhou H, Pandit SB, Lee SY, Borreguero J, Chen H, Wroblewska L, Skolnick J. Analysis of TASSER based CASP7 protein structure prediction results. Proteins 2007;69(Suppl 8):90–97.
45. Jones DT, Bryson K, Coleman A, McGuffin LJ, Sadowski MI, Sodhi JS, Ward JJ. Prediction of novel and analogous folds using fragment assembly and fold recognition. Proteins 2005;61 (Suppl 7):143–151.
46. Wallner B, Larsson P, Elofsson A. Pcons.net: protein structure prediction meta server. Nucleic Acids Res 2007;35 (Web Server Issue):W369–W374.
47. Hawkins T, Luban S, Kihara D. Enhanced automated function prediction using distantly related sequences and contextual association by PFP. Protein Sci 2006;15:1550–1556.
48. Hamilton N, Burrage K, Ragan MA, Huber T. Protein contact prediction using patterns of correlation. Proteins 2004;56:679–684.
49. Punta M, Rost B. PROFcon: novel prediction of long-range contacts. Bioinformatics 2005;21:2960–2968.
50. Hung LH, Ngan SC, Liu T, Samudrala R. PROTINFO: new algorithms for enhanced protein structure predictions. Nucleic Acids Res 2005;33(Web Server Issue):W77–W80.
51. Raghava GPS. MANGO: prediction of genome ontology (GO) class of a protein from its amino acid and dipeptide composition using nearest neighbor approach. In: CASP7: Proceedings of the 7th meeting on the critical assessment of techniques for protein structure prediction, 26–30 November 2006, Asilomar, CA.
52. Xu J, Li M, Kim D, Xu Y. RAPTOR: optimal protein threading by linear programming. J Bioinform Comput Biol 2003;1:95–117.
53. Bu D, Li S, Gao X, Yu L, Xu J, Li M. Consensus approaches for protein structure prediction. In: Zhang YQ, Jagath CR, editors. Machine Learning in Bioinformatics. New York: Wiley; 2007.
54. Chivian D, Kim DE, Malmstrom L, Bradley P, Robertson T, Murphy P, Strauss CE, Bonneau R, Rohl CA, Baker D. Automated prediction of CASP-5 structures using the Robetta server. Proteins 2003;53 (Suppl 6):524–533.
55. Schlessinger A, Rost B. Protein flexibility and rigidity predicted from sequence. Proteins 2005;61:115–126.
56. Schlessinger A, Yachdav G, Rost B. PROFbval: predict flexible and rigid residues in proteins. Bioinformatics 2006;22:891–893.
57. Karplus K, Katzman S, Shackleford G, Koeva M, Draper J, Barnes B, Soriano M, Hughey R. SAM-T04: what is new in protein-structure prediction for CASP6. Proteins 2005;61(Suppl 7):135–142.
58. Karplus K, Karchin R, Draper J, Casper J, Mandel-Gutfreund Y, Diekhans M, Hughey R. Combining local-structure, fold-recognition, and new fold methods for protein structure prediction. Proteins 2003;53(Suppl 6):491–496.
59. Karplus K, Barrett C, Cline M, Diekhans M, Grate L, Hughey R. Predicting protein structure using only sequence information. Proteins 1999;37(Suppl 3):121–125.
60. Zhou H, Zhou Y. Fold recognition by combining sequence profiles derived from evolution and from depth-dependent structural alignment of fragments. Proteins 2005;58:321–328.
61. Liu S, Zhang C, Liang S, Zhou Y. Fold Recognition by concurrent use of solvent accessibility and residue depth. Proteins 2007;68:636–645.
62. Cheng J, Baldi P. Improved residue contact prediction using support vector machines and a large feature set. BMC Bioinformatics 2007;8:113.
63. Poleksic A, Danzer JF, Hambly K, Debe DA. Convergent island statistics: a fast method for determining local alignment score significance. Bioinformatics 2005;21:2827–2831.
64. Debe DA, Danzer JF, Goddard WA, Poleksic A. STRUCTFAST: protein sequence remote homology detection and alignment using novel dynamic programming and profile-profile scoring. Proteins 2006; 64:960–967.
65. Zhang Y. Template-based modeling and free modeling by I-TASSER in CASP7. Proteins 2007;69(Suppl 8):108–117.
66. Jauch R, Yeo H, Kolatkar PR, Clarke ND. Assessment of CASP7 structure predictions for template free targets. Proteins 2007;69 (Suppl 8):57–67.

67. Kopp J, Bordoli L, Battey JND, Kiefer F, Schwede T. Assessment of CASP7 predictions for template-based modeling targets. Proteins 2007;69(Suppl 8):38–56.
68. Read RJ, Chavali G. Assessment of CASP7 predictions in the high accuracy template-based modeling category. Proteins 2007;69 (Suppl 8):27–37.
69. Bordoli L, Kiefer F, Schwede T. Assessment of disorder predictions in CASP7. Proteins 2007;69(Suppl 8):129–136.
70. Tress M, Cheng J, Baldi P, Joo K, Lee J, Seo J-H, Lee J, Baker D, Chivian D, Kim D, Ezkurdia I. Assessment of predictions submitted for the CASP7 domain prediction category. Proteins 2007;69(Suppl 8):137–151.
71. Izarzugaza JMG, Graña O, Tress ML, Valencia A, Clarke ND. Assessment of intramolecular contact predictions for CASP7. Proteins 2007; 69(Suppl 8):152–158.
72. López G, Rajas A, Tress M, Valencia A. Assessment of predictions submitted for the CASP7 function prediction category. Proteins 2007;69(Suppl 8):165–174.
73. Greene LH, Lewis TE, Addou S, Cuff A, Dallman T, Dibley M, Redfern O, Pearl F, Nambudiry R, Reid A, Sillitoe I, Yeats C, Thornton JM, Orengo CA. The CATH domain structure database: new protocols and classification levels give a more comprehensive resource for exploring evolution. Nucleic Acids Res 2007;35 (Database Issue): D291–D297.
74. 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.
75. Zemla A. LGA: a method for finding 3D similarities in protein structures. Nucleic Acids Res 2003;31:3370–3374.
76. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 1997;25:3389–3402.
77. Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Edgar R, Federhen S, Geer LY, Kapustin Y, Khovayko O, Landsman D, Lipman DJ, Madden TL, Maglott DR, Ostell J, Miller V, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Sirotkin K, Souvorov A, Starchenko G, Tatusov RL, Tatusova TA, Wagner L, Yaschenko E. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 2007;35 (Database Issue):D5–D12.
78. Martin AC, MacArthur MW, Thornton JM. Assessment of comparative modeling in CASP2. Proteins 1997;29(Suppl 1):14–28.
79. Peitsch MC. ProMod and Swiss-model: internet-based tools for automated comparative protein modelling. Biochem Soc Trans 1996;24:274–279.
80. Rychlewski L, Fischer D. LiveBench-8: the large-scale, continuous assessment of automated protein structure prediction. Protein Sci 2005;14:240–245.
81. Koh IY, Eyrich VA, Marti-Renom MA, Przybylski D, Madhusudhan MS, Eswar N, Grana O, Pazos F, Valencia A, Sali A, Rost B. EVA: evaluation of protein structure prediction servers. Nucleic Acids Res 2003;31:3311–3315.

DOI 10.1002/prot