PROTEINS: Structure, Function, and Genetics 53:542–547 (2003)

LiveBench-6: Large-Scale Automated Evaluation of Protein Structure Prediction Servers

Leszek Rychlewski,[1]* Daniel Fischer,[2] and Arne Elofsson[3]
[1]*Bioinformatics Laboratory, BioInfoBank Institute, Poznan, Poland*
[2]*Bioinformatics, Department of Computer Science, Ben Gurion University, Beer-Sheva, Israel*
[3]*Stockholm Bioinformatics Center, Stockholm University, Stockholm, Sweden*

ABSTRACT **The aim of the LiveBench experiment is to provide a continuous evaluation of structure prediction servers in order to inform potential users about the current state-of-the-art structure prediction tools and in order to help the developers to analyze and improve the services. This round of the experiment was conducted in parallel to the blind CAFASP-3 evaluation experiment. The data collected almost simultaneously enables the comparison of servers on two different benchmark sets. The number of servers has doubled from the last evaluated LiveBench-4 experiment completed in April 2002, just before the beginning of CAFASP-3. This can be partially attributed to the rapid development in the area of meta-predictors (consensus servers). The current results confirm the high sensitivity and specificity of the meta-predictors. Nevertheless, the comparison between the autonomous (not meta) servers participating in the last CAFASP-2 and LiveBench-2 experiment and the current set of autonomous servers demonstrates that progress has been made also in sequence structure fitting functions. In addition to the growing number of participants, the current experiment marks the introduction of new evaluation procedures, which are aimed to correlate better with functional characteristics of models. Proteins 2003;53:542–547.**

Key words: protein structure prediction; consensus fold recognition; CAFASP; Meta-Server; ToolShop; LiveBench

INTRODUCTION

LiveBench belongs to a group of community-wide protein structure prediction evaluation programs, which includes CASP,[1] CAFASP[2] and EVA.[3] These programs have a common goal to assess the current accuracy of structure prediction approaches, but differ in procedural details. CASP and CAFASP evaluate blind predictions, operating on targets, which don't have a known structure at the time the prediction is being made. LiveBench evaluates only fully automated fold recognition methods and uses protein structures freshly released in the PDB[4]. Only non-trivial PDB entries are used as targets (we define a non-trivial target as one that does not exhibit strong sequence similarity to other previously known structures, as measured by Blast using an E-value cutoff of 0.001). Circa 5 new proteins extracted weekly from the PDB result in a set of over 100 non-trivial targets submitted in 6 months. The collected predictions are evaluated automatically and the summaries of the performance of participating servers are updated continuously and available at the homepage of the program http://BioInfo.PL/LiveBench/. The final summary presented in this manuscript is based on 98 targets submitted between August and December of 2002.

MATERIALS AND METHODS

The CASP experiment is known to change the evaluation procedure with every new biannual session in a search for a single "best" scoring protocol. The LiveBench experiment is no exception to this practice. This round 5 new evaluation methods were introduced, which are added to the 4 used previously (MaxSub,[5] two LGscore[6] versions and Touch).[7] The selection of the new methods is the result of a struggle to find objective criteria for comparing evaluation procedures. The rationale behind the selection relates to the purpose of using structure prediction methods by biologists. In most of the cases the user is interested in the function of the target protein. The structure prediction is used as a guide to restrict the set of potential functions of the target, by analyzing the functions found in proteins, which share the predicted fold. One of the possibilities to use this query as an evaluation criteria for fold recognition methods would be to ask if the predicted structure modeled based on a template share the same functional / structural class as the target. We have chosen the SCOP[8] database as the standard of truth for this task, as it is known to be biased towards functional and evolutionary properties of the proteins.

There are three major drawbacks of this evaluation procedure. First, servers that do not use templates for structure prediction (e.g. ab initio servers) cannot be evaluated based on the functional or structural similarity between the template and the target. Second, this evaluation method ignores the accuracy of the alignment, which

Grant sponsor: Swedish Strategic Research Foundation; Grant sponsor: Carl Trygger Foundation; Grant sponsor: Swedish National Research Council.

*Correspondence to: Leszek Rychlewski, Bioinformatics Laboratory, BioInfoBank Institute, ul. Limanowskiego 24A, 60-744 Poznan, Poland. E-mail: leszek@bioinfo.pl

Received 17 February 2003; Accepted 15 May 2003

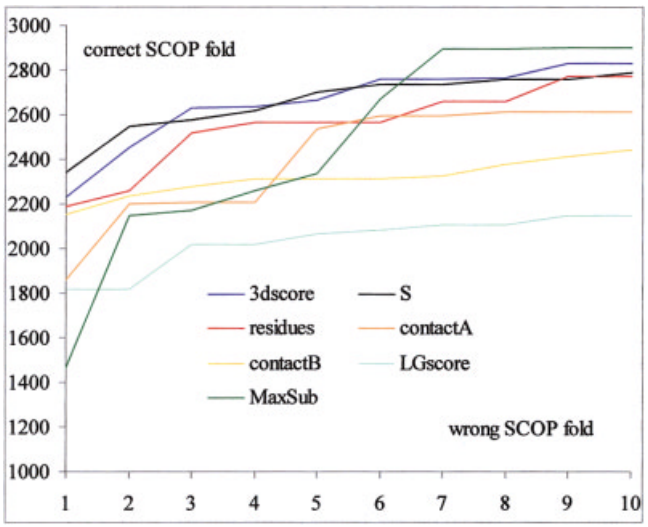

Fig. 1. Simple measures like "residues" perform as well as other measures. ROC plot of the similarity between model assessment methods and SCOP classification of the target and the template. Models collected in LiveBench-2 and LiveBench-4 for targets larger than 70 amino acids with templates classified in SCOP version 1.61 are sorted according to their quality assessed using 7 different sequence-dependent (alignment-dependent) procedures. The x-axis shows the ten first models based on templates with different SCOP fold class than the target (false positives). The y-axis sums the models with correct (identical) SCOP-fold classification (correct positives). Even the very simple and intuitive evaluation method based on the number of correctly positioned residues ("residues") shows relatively a very good agreement with the SCOP classification. Two other methods introduced on the LiveBench pages ("contactA" and "contactB"), which are based on the analysis of distances between contacting residues are also plotted. A more detailed description of the methods can be found at the LiveBench pages: http://bioinfo.pl/Meta/evaluation.html.

is often crucial for the confirmation of the predicted function (for example through comparison of aligned active site residues). Third, the assessment would be delayed as the SCOP classification of all new targets and templates has to be performed before the results can be evaluated. As a compromise we have asked the question, which of the standard model quality assessment methods correlate best with the SCOP similarity between the template and the target (Fig. 1). To our surprise very good correlation can be obtained with a simple measure, which scores models based on the number of residues that can be superimposed on the native structure within 3Å ("residues" in Figure 1).

Because of these results we will focus in this manuscript mainly on the analysis of methods evaluated using the "residues" measure. Results obtained using other measures are available from the accompanying website and a comparison of the ranking using 5 measures is shown in Figure 2. For the "residues" measures we have chosen the cutoff of 40 correct residues to define a correct model. Models that have more correctly positioned residues have a 90% chance that the corresponding template belongs to the same SCOP fold as the target (data not shown). A model, which has less than 30 correct residues, is defined as wrong. Models with between 30 and 40 correct residues are borderline models and we exclude them from the specificity analysis, for which wrong models have to be defined (see below).

9 sequence comparison servers, 12 fold recognition servers (servers which use the structural information of the template) and 13 versions of meta-predictors (servers, which use other servers as main knowledge base) are evaluated here. Information about the servers and their corresponding abbreviations are available on the LiveBench pages (http://bioinfo.pl/Meta/servers.html). The query sequences were divided into 32 easy and 66 hard targets based on the e-value of the first hit reported by PDB-Blast. If the e-value is below 0.001 the target is defined as easy. The main sensitivity ranking was calculated using the sum of correctly positioned residues for all correct models. This is a different procedure than the Z-score based procedures used frequently by the CASP assessors. The main difference is that in LiveBench missing models or wrong predictions are heavily penalized and do not obtain a minimum Z-score of zero as in CASP. Also the Z-score based assessment gives disproportional many points to correct "outliers" which are wrong but much better than other models for a given target. Often (in the difficult fold recognition category) the distribution of completely wrong models would be used as a reference distribution. From the users point of view such "outliers" are very difficult to use or validate and should not obtain additional points from the fact that other methods failed completely. The LiveBench scoring is proportional to the quality of the models (defined for example as the number of correct residues) and does not depend on the performance of other methods (in contrast to the application of Z-scores). This could also help in the global analysis of the progress in the field.

RESULTS

Table 1 shows the sensitivity of the participating servers, measured as the total number of correctly positioned residues in correct models, separated into the easy and hard category. The best autonomous server performance is reported for the Sam-T02[9]server on the easy targets, and for Shotgun-INGBU[10] on the hard targets. If only the number of correctly predicted targets is taken into account, the top ranking of the Shotgun-INBGU method is confirmed in the hard category, while the results are too close in the easy category to distinguish the servers.

Table 1 shows also the specificity evaluation of the servers. The Specificity is defined as in previous LiveBench-4 experiment, as the average number of correct predictions with higher confidence score than the first to tenth false predictions (similar to the Receiver Operator Characteristics[11] for the first 10 targets). Here the ranking of the autonomous servers is lead by the Shotgun-INBGU method.

The Shotgun-INBGU server uses a meta-predictor layer on top of the 5 prediction components of the original INBGU.[12] The new layer shows much better prediction performance than the original procedure implemented in INBGU to rank all generated models. A feature of the Shotgun servers is that they assemble hybrid models by splicing fragments from different templates. The main disadvantage is that for the hardest targets, where conflict-

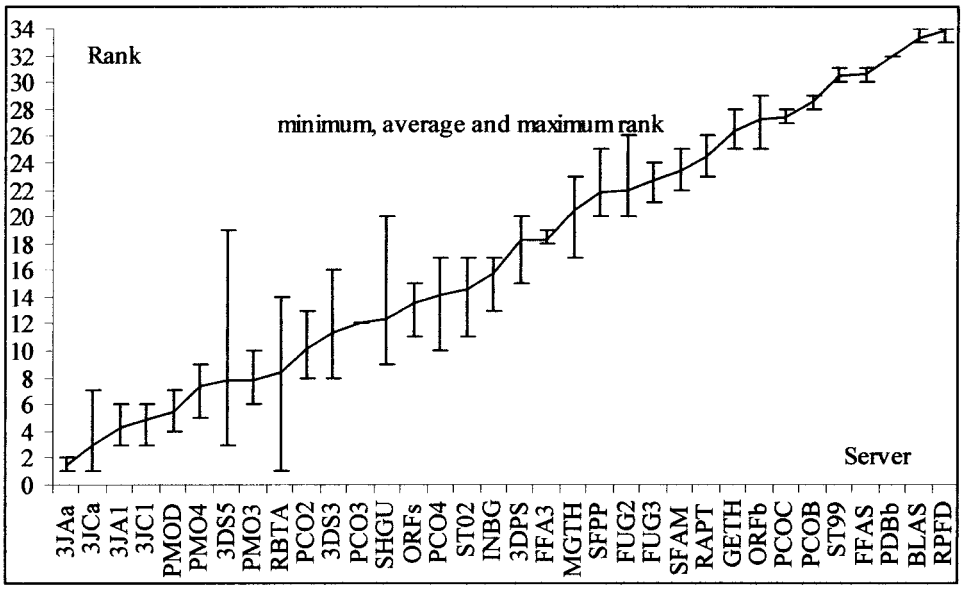

Fig. 2. Variation in the ranking of the servers. 5 new evaluation methods were used to rank the sensitivity of the servers participating in LiveBench-6 on the hard targets. The plots show the minimum (best) and the maximum rank (worst) and the average rank used to sort the 34 servers. Two contact-based evaluation methods (contactA and contactB), two rigid body superposition methods (number of correct atoms and 3dscore) and one combination of the contact-based and a rigid body evaluation method is employed. The details of the methods are described on the LiveBench homepage. Changes in ranking of almost 50% can be observed (16 positions between the best and the worst ranking obtained by 3DS5).

ing structural information exists amongst the input structures, the shotgun servers can produce "unphysical" models with sometimes overlapping residues. Because of the splicing, Shotgun-INBGU lies at the borderline between autonomous servers and meta-predictors. If this server would be ignored the picture of the performance of servers would change slightly. Servers which profit from an exhaustive utilization of sequence information (Sam-T02, FFAS-03[13] and ORFeusz)[14] would occupy the top 3 positions in the specificity ranking and the top 2 positions in sensitivity in both the easy and the hard category. This could lead to the conclusion that the current sequence structure fitting functions as implemented in the threading methods still suffer from the frozen approximation, inaccuracies of the parameters, or the structural divergence between template and target beyond the reliability limits of the force fields.

On the other hand a very powerful use of the structural information of the template proteins is demonstrated by the meta-predictors. While ignoring energetic criteria, the structures of the templates are used to find similarities, structurally conserved segments in alignments with various protein hits from the database. Generally, the meta-predictors, select the models with the largest structurally consistent alignment based on the structural superposition of all models. This simple procedure has a dramatic impact on the final accuracy of the service, even without optimal implementation. This is not as clear on the easy targets where the differences between the results of servers are not that large. However on the set of hard targets all meta-predictors have higher sensitivity than any autonomous server (excluding the borderline case of Shotgun-INBGU). The best performing meta-predictors generate over 40% more correct residues than the best autonomous server (ORFeus) on this set (excluding the borderline case of Shotgun-INBGU). In the specificity evaluation meta-predictors clearly outperform autonomous servers showing a 20% increases in the ROC values.

The results of meta-predictors are also shown in Table 1. In the sensitivity evaluation, the Shotgun series of meta-predictors seem to be outperforming the single template meta-predictors, such as Pcons[15] and Pmodeller. The picture is opposite in the specificity evaluation, where the Pmodeller series seems to be leading, possibly due to the structural post-evaluations using ProQ[16]. Another meta predictor, the 3D-Jury system can operate as a meta-meta predictor, utilizing the results collected from other meta-predictors (3D-Jury-C versions). With such settings it gets very good scores in the sensitivity evaluation, mainly due to the selection of Shotgun results. The specificity results are better if only a selected set of 8 autonomous servers is allowed for model selection (3D-Jury-A version). With such settings, the 3D-Jury system generates also the highest number of correct predictions but it can hardly compete with the spliced models in the residue-based evaluation.

Robetta, a new server, is the only meta predictor that does not superimpose multiple models obtained from other services for the purpose of internal quality estimation and ranking. Robetta uses other servers such as Pcons2 only for template section and applies an *ab initio* protocol to improve the alignments and to add missing fragments to the model. Robetta did not perform very well in the LiveBench-6 experiment. Nevertheless it has obtained very good results in the CAFASP and CASP evaluation

TABLE I. Sensitivity and Specificity (ROC) of the Servers Measured on the LiveBench-6 and CAFASP-3 Targets

LiveBench-6									CAFASP-3								
EASY			HARD			ROC			EASY			HARD			ROC		
NAME	SUM	N	NAME	SUM	N	NAME	ROC	F	NAME	SUM	N	NAME	SUM	N	NAME	ROC	F
3DS5	3003	27	3JCa	2018	26	3JA1	49.0	47	3DS5	1620	15	3JC1	815	12	3JCa	42.8	41
3JCa	3002	27	3JAa	2007	29	PMO3	49.0	43	3JCa	1607	15	RBTA	788	13	3JC1	42.4	41
3JC1	2916	27	3DS5	1945	25	PMOD	47.8	39	3DS3	1604	15	3JCa	777	11	3JAa	41.8	41
3DS3	2864	25	3JA1	1890	28	PMO4	46.1	34	3JC1	1587	15	3JAa	759	11	3DS5	41.6	40
ST02	2827	26	3JC1	1839	23	3JCa	45.9	35	PMOD	1497	15	3DS5	756	10	3JA1	41.0	41
PCO3	2745	27	PMO3	1775	25	PCO2	45.6	35	3JAa	1486	15	3JA1	713	10	3DS3	41.0	41
PMO4	2739	26	PMOD	1756	26	3JC1	45.1	38	SHGU	1471	14	3DS3	666	9	PMOD	40.0	40
RBTA	2731	25	3DS3	1690	24	3DS5	45.0	24	3JA1	1461	15	RAPT	664	10	PMO3	39.7	37
PMOD	2720	26	PMO4	1670	24	3Jaa	44.9	38	PMO3	1449	15	PMO3	656	9	PCO2	39.4	38
PMO3	2717	27	SHGU	1649	22	PCO3	44.6	27	ORFs	1417	15	PMOD	643	9	SHGU	38.7	38
3JA1	2702	26	PCO2	1608	24	3DS3	43.2	33	3DPS	1414	15	FUG3	636	10	PCO3	37.7	31
SHGU	2686	25	PCO3	1593	21	SHGU	42.9	35	RAPT	1402	15	PCO2	623	9	INBG	37.0	37
PCO4	2683	26	PCO4	1454	22	ORFs	42.8	38	PCO3	1401	15	SHGU	598	8	FUG3	36.9	30
FFA3	2648	26	RBTA	1439	20	ST02	40.1	37	RBTA	1399	15	FUG2	591	9	3DPS	36.8	33
FUG3	2647	25	ORFs	1413	20	PCO4	38.3	27	FUG3	1395	15	ST02	574	9	FUG2	36.7	30
ORFs	2629	27	ST02	1366	19	FFA3	37.2	19	PDBb	1372	15	FFA3	574	8	ORFs	36.6	34
RAPT	2555	25	INBG	1343	21	RAPT	37.0	32	PCO2	1370	15	PCO3	543	7	FFA3	36.4	30
3Jaa	2626	26	FFA3	1213	18	INBG	36.8	23	MGTH	1353	15	ORFs	534	7	MGTH	35.1	17
SFPP	2553	24	3DPS	1157	16	FUG2	35.6	13	FUG2	1333	15	INBG	471	7	FFAS	34.6	32
FUG2	2543	24	RAPT	1147	18	FUG3	35.3	11	FFAS	1324	14	3DPS	466	7	RAPT	34.0	29
PCO2	2515	26	FUG2	1134	19	SFPP	34.6	17	FFA3	1301	14	MGTH	407	6	GETH	33.5	29
INBG	2514	24	FUG3	1111	17	MGTH	34.0	22	ST99	1288	14	FFAS	401	5	ST99	33.5	30
3DPS	2513	24	SFPP	1087	16	SFAM	32.7	11	ST02	1287	13	PCOB	385	6	PDBb	33.0	33
MGTH	2420	24	MGTH	1081	16	ORFb	32.5	8	INBG	1280	14	SFPP	289	5	ORFb	32.6	30
ORFb	2404	22	SFAM	1030	16	ST99	31.0	25	ORFb	1234	14	ST99	239	5	PCOB	31.1	26
GETH	2360	23	ORFb	975	16	PCOC	30.9	5	GETH	1146	14	GETH	162	3	ST02	29.9	29
SFAM	2359	21	PCOC	917	15	GETH	30.0	25	PCOB	1129	12	ORFb	122	2	SFAM	29.4	27
FFAS	2317	23	GETH	908	14	3DPS	27.0	12	SFPP	1105	12	SFAM	120	2	SFPP	22.2	6
PDBb	2265	23	PCOB	845	12	PCOB	25.4	15	SFAM	1047	11	PDBb	45	1	RPFD	19.0	19
ST99	2171	22	FFAS	723	11	FFAS	24.6	9	RPFD	180	3	RPFD	0	0			
PCOC	2159	24	ST99	674	12	PDBb	21.4	4									
PCOB	2089	24	PDBb	277	5	RPFD	10.2	3									
RPFD	1139	13	BLAS	96	2	BLAS	8.0	6									
BLAS	712	10	RPFD	42	1												

The targets are divided into EASY and HARD (see text). The NAME columns contain the four-letter abbreviation of the participating server. Meta-predictors are indicated in black color. Servers that use the sequence information only are marked in blue. Servers marked in red use structural information about the template protein in the fitting function. SHGU, colored in brown, is a borderline case, which uses five INBGU components and a meta-predictor like jury level. The SUM column shows the sum of all correctly positioned residues in correct models. The N column shows the number of correct models (with > 39 correct residues). The F column shows the number of correct models with higher confidence score than the first false prediction (< 30 correct residues). The ROC column approximates the specificity of the confidence score. The reported value is defined as the average number of correct predictions with higher confidence score than the first to tenth false predictions (as in LiveBench-4). Other evaluation results are available at http://bioinfo.pl/LiveBench/ and http://bioinfo.pl/cafasp/. Information about the servers is available at http://bioinfo.pl/Meta/servers.html.

TABLE II. Progress in the Field of Automated Protein Structure Prediction Measured by the Difference Between the Performance in LiveBench-6 of the Old (the OLD column) and the New (the NEW column) Versions of the Servers

		EASY		HARD		ROC	
OLD	NEW	ΔSUM	ΔN	ΔSUM	ΔN	ΔROC	ΔF
ST99	ST02	656	4	692	7	9.1	12
FFAS	FFA3	331	3	490	7	12.6	10
INBG	SHGU	172	1	306	1	6.1	12
SFAM	SFPP	194	3	57	0	1.9	6
PCOB	PCOC	70	0	72	3	5.5	−10
AVERAGE		285	2	323	4	7	6

Five pairs of servers are shown. The values are directly extracted from the values presented in Table I. ΔSUM shows the difference in the total number of correctly positioned residues in correct models. ΔN shows the difference in the number of correct models. Columns ΔSUM and ΔN are separated in EASY and HARD targets as in Table I. ΔROC shows the difference in the specificity (ROC) score. ΔF shows the difference between the numbers of correct models with higher confidence than the first false prediction. The average differences are shown in the bottom row (AVERAGE). Sam-T02 shows the biggest progress relative to its predecessor.

(Table 1). Table 1 includes the ranking of servers, which would be obtained if the current LiveBench evaluation would be conducted on the results collected for the CASP-5/CAFASP-3 targets. There are some differences between both sets of targets. The CASP team divides the targets into domains and the classification into easy, hard and other categories is defined on the domain level. The ranking on the hard targets corresponds to the results obtained on domains classified in CASP in the fold recognition category. The easy targets correspond to CASP comparative modeling domains excluding those, which could be predicted by blast (blast E-value cutoff of 0.001 was used as in regular LiveBench). The targets from the CASP new fold category are also excluded. The specificity analysis is conducted on the results collected for full targets proteins (not domains) because the reliability scores reported by servers relate to one single target and not to parts of it.

CONCLUSIONS AND DISCUSSION

As mentioned earlier Table 1 shows quite big variations in the ranking of some methods. For example RAPTOR[17] did not perform as well in the sensitivity analysis on the LiveBench set as it did on the CASP set, where it scored as high as some of the meta-servers in the fold recognition sensitivity analysis. Figure 2 shows how the sensitivity ranking on hard targets is affected by the choice of the model evaluation methods. The data indicates that the ranking is very flexible even if the performance is measured on the same set of targets. Particular large variation can be seen in the methods that use fragment based approaches, i.e. ShotGun and ROBETTA.[18] Thus the benchmarking experiments have more in common with Olympic games than with detailed investigations of the utility of prediction modules introduced in current fold recognition strategies.

Nevertheless some conclusions can be drawn from the results. An obvious conclusion is the confirmation that meta-predictors, even if being quite early in the development, provide more reliable access to structure predictions than autonomous servers. One of the reasons could be very technical. Many autonomous servers are likely to miss one or two predictions because missing templates in their fold database or because of peculiarities of their scoring function. Meta predictors will locate and ignore those result files easily. However, it is clear from earlier studies (Lundstrom, et al 2000) and the improved performance of Shotgun-INBGU over INBGU that this is not the main reason for the improvement. The main reason for the improved specificity is due to the "consensus" analysis performed by all the meta-server methods. The second important conclusion is that there is progress in the development of autonomous servers despite the "parasitic" nature of the meta-predictors. This can be evaluated best when comparing old and new version of servers from the same family on the same set of targets (Table 2).

LiveBench clearly fails to declare winners and losers of the structure prediction server community and it is not our intention to change this. The program operates more like a periodic customer report watched mainly by the participants, who use the data to trace procedural errors in their prediction algorithms. One of the main conclusions of the first LiveBench cycles was that the completeness of the fold library is an important factor severely influencing the outcome of the evaluation results. Nevertheless, it is understandable that not all algorithm developers spend time with routine database updates, which cannot be regarded as scientifically challenging. This is especially true now when such errors can be partially fixed by meta-predictors, which diversify the sources of models. The future focus may shift from providing complete structure prediction solution to specialized services. In addition to the popular meta-predictors, such services could include loop modeling servers or model quality estimation servers, which would be clearly of benefit for the users. The LiveBench program will continue providing useful training data for such projects.

ACKNOWLEDGMENTS

This work is not supported by the Polish State Committee for Scientific Research. AE was supported by grants from the Swedish Strategic research foundation, the Carl Trygger foundation and the Swedish national research council. We would also like to thank all developers of

protein structure prediction servers who participate in the LiveBench program.

REFERENCES

1. Moult J, Fidelis K, Zemla A, Hubbard T. Critical assessment of methods of protein structure prediction (CASP): Round IV. Proteins 2001;45 Suppl 5:2–7.
2. Fischer D, Elofsson A, Rychlewski L, Pazos F, Valencia A, Rost B, Ortiz AR, Dunbrack RL, Jr. CAFASP2: The second critical assessment of fully automated structure prediction methods. Proteins 2001;45 Suppl 5:171–183.
3. Eyrich VA, Marti-Renom MA, Przybylski D, Madhusudhan MS, Fiser A, Pazos F, Valencia A, Sali A, Rost B. EVA: continuous automatic evaluation of protein structure prediction servers. Bioinformatics 2001;17:1242–1243.
4. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S, Fagan P, Marvin J, Padilla D, Ravichandran V, Schneider B, Thanki N, Weissig H, Westbrook JD, Zardecki C. The Protein Data Bank. Acta Crystallogr D Biol Crystallogr 2002;58:899–907.
5. Siew N, Elofsson A, Rychlewski L, Fischer D. MaxSub: an automated measure for the assessment of protein structure prediction quality. Bioinformatics 2000;16:776–785.
6. Cristobal S, Zemla A, Fischer D, Rychlewski L, Elofsson A. A study of quality measures for protein threading models. BMC Bioinformatics 2001;2:5.
7. Bujnicki JM, Elofsson A, Fischer D, Rychlewski L. LiveBench-2: Large-scale automated evaluation of protein structure prediction servers. Proteins 2001;45 Suppl 5:184–191.
8. LoConte L., Ailey B, Hubbard TJ, Brenner SE, Murzin AG, Chothia C. SCOP: a structural classification of proteins database. Nucleic Acids Res 2000;28:257–259.
9. Karplus K, Karchin R, Barrett C, Tu S, Cline M, Diekhans M, Grate L, Casper J, Hughey R. What is the value added by human intervention in protein structure prediction? Proteins 2001;Suppl 5:86–91.
10. Fischer D. 3D-SHOTGUN: A Novel, Cooperative, Fold-Recognition Meta-Predictor. Proteins 2003;51:434–441
11. Swets JA, Dawes RM, Monahan J. Better decisions through science. Sci Am 2000;283:82–87.
12. Fischer D. Hybrid fold recognition: combining sequence derived properties with evolutionary information. Pac Symp Biocomput 2000;119–130.
13. Rychlewski L, Jaroszewski L, Li W, Godzik A. Comparison of sequence profiles. Strategies for structural predictions using sequence information. Protein Sci 2000;9:232–241.
14. Pas J, Wyrwicz LS, Grotthuss M, Bujnicki JM, Ginalski K, Rychlewski L. ORFeus: Detection of distant homology using sequence profiles and predicted secondary structure. Nucleic Acids Res 2003;
15. Lundstrom J, Rychlewski L, Bujnicki J, Elofsson A. Pcons: A neural-network-based consensus predictor that improves fold recognition. Protein Sci 2001;10:2354–2362.
16. Wallner B, Elofsson A. Can correct protein models be identified? Protein Science 2003; Forthcoming.
17. Xu J, Li M, Lin.G., Xu Y, Kim D. Protein Threading By Linear Programming. Pac Symp Biocomput 2003;
18. Simons KT, Kooperberg C, Huang E, Baker D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. J Mol Biol 1997;268:209–225.