BIOINFORMATICS

Vol. 19 no. 8 2003, pages 1015–1018
DOI: 10.1093/bioinformatics/btg124

3D-Jury: a simple approach to improve protein structure predictions

Krzysztof Ginalski[1], Arne Elofsson[2], Daniel Fischer[3] and Leszek Rychlewski[1,]*

[1]BioInfoBank Institute, Limanowskiego 24A, 60-744 Poznan, Poland, [2]Stockholm Bioinformatics Center, AlbaNova, Stockholm University, 10691 Stockholm, Sweden and [3]Bioinformatics, Department of Computer Science, Ben Gurion University, 84015 Beer-Sheva, Israel

Received on November 7, 2002; revised on January 9, 2003; accepted on January 14, 2003

ABSTRACT

Motivation: Consensus structure prediction methods (meta-predictors) have higher accuracy than individual structure prediction algorithms (their components). The goal for the development of the 3D-Jury system is to create a simple but powerful procedure for generating meta-predictions using variable sets of models obtained from diverse sources. The resulting protocol should help to improve the quality of structural annotations of novel proteins.

Results: The 3D-Jury system generates meta-predictions from sets of models created using variable methods. It is not necessary to know prior characteristics of the methods. The system is able to utilize immediately new components (additional prediction providers). The accuracy of the system is comparable with other well-tuned prediction servers. The algorithm resembles methods of selecting models generated using *ab initio* folding simulations. It is simple and offers a portable solution to improve the accuracy of other protein structure prediction protocols.

Availability: The 3D-Jury system is available via the Structure Prediction Meta Server (http://BioInfo.PL/Meta/) to the academic community.

Contact: leszek@bioinfo.pl

Supplementary information: 3D-Jury is coupled to the continuous online server evaluation program, LiveBench (http://BioInfo.PL/LiveBench/).

INTRODUCTION

The knowledge of the 3D structure of a protein is an extremely useful prerequisite for the understanding of the function and for the rational modification of proteins. Due to the increasing gap between the number of known protein sequences and the number of structural annotations, the problem of predicting the tertiary structure of a protein from its amino acid sequence remains an important field of research in molecular biology (Baker and Sali , 2001). Objective and community-wide assessment of the accuracy of available methods such as CASP (Moult *et al.*, 2001) or CAFASP (Fischer *et al.*, 2001) have made a significant contribution to the progress in this area and have lead to an increased interest in the development of new prediction algorithms. As a result, the number of prediction services available on the internet that participated in last year's CASP-5 and CAFASP-3 experiments has almost doubled compared to the numbers from the previous experiments, conducted three years ago. New servers diversify the set of available prediction approaches and provide added value to the community of automated structure annotation methods. Due to the increased number of available predictions, the chances of obtaining a correct model increases. However, from the user point of view, it is not easy to benefit from the large selection and it is sometimes even more difficult to select the best model.

SYSTEM AND METHODS

First attempts to benefit from the variety of available services were made by the semi-automated CAFASP-Consensus groups (Fischer *et al.*, 2001). The success of the semi-automated approach in CASP-4 lead to the development of a series of fully automated services, which are based on a similar principle of using the results of independent prediction methods, but differ in the way the information is processed.

First benchmarks within the LiveBench-2 (Bujnicki *et al.*, 2001) and LiveBench-4 experiments have indicated that fully automated meta-predictors are more accurate than any individual server used for building the consensus. Initial results were obtained with the Pcons (Lundstrom *et al.*, 2001) method, which currently has several variants that differ in the set of components and in the final pro-

*To whom correspondence should be addressed.

 © Oxford University Press 2003; all rights reserved.

cessing of the models (with Modeller; Sali and Blundell, 1993). Pcons ranks models generated by a set of servers by employing a scoring function, which takes into account the confidence of the model reported by the server and the similarity of the model to all other models. A neural network is used to translate the original confidence scores into standard scores to facilitate the comparison of different servers. This procedure requires an initial tuning of the neural network before a new server can be added to the set of servers used for consensus building.

The 3D-SHOTGUN meta-predictors (Fischer , 2002) are reminiscent of the so-called 'cooperative algorithms' known in the Computer Vision sub-area of Artificial Intelligence (Marr, 1982) The program also takes as input the models with their confidence scores. The result is a hybrid model, which is spliced from fragments of the input models and has the potential of covering more parts of the native protein than any template structure alone. Thus, 3D-SHOTGUN entails the first fold-recognition meta-predictor attempt to go beyond the simple selection of one of the input models. The 3D-SHOTGUN methods have demonstrated their capabilities since the LiveBench-4 experiment.

The 3D-Jury system, like other meta-predictors, incorporates the comparison of models as the main processing step. It follows an approach similar to that employed in the field of *ab initio* fold recognition. Recent advances in the development in this area can be accredited to the application of non-energetic constrains such as preferences for high contact order or the detection of clusters of abundant conformations (Bonneau *et al.*, 2002). The experience with *ab initio* prediction methods lead to the conclusion that averages of low-energy conformations obtained most frequently by folding simulations are closer to the native structure than the conformation with lowest energy. The direct translation of this finding into the field of fold recognition by threading methods would mean that most abundant high-scoring models are closer to the native structure than the model with highest score. This is the main rationale behind the 3D-Jury approach.

ALGORITHM

3D-Jury, takes as input groups of models generated by a set of servers, however, neglecting the assigned confidence scores. All models are compared with each other and a similarity score is assigned to each pair, which equals to the number of Calpha atom pairs that are within 3.5 Å after optimal superposition. The MaxSub tool (Siew *et al.*, 2000) is used to calculate the similarity of two models, but any other similar programs can be used as well. If this number is below 40, the pair of models is annotated as not similar and the score is set to Zero. The cutoff value of 40 was taken from previous benchmarking results (unpublished) and indicates a roughly 90% chance for both models to belong to the same fold class. The final 3D-Jury score of a model is the sum of all similarity scores of considered model pairs divided by the number of considered pairs plus one. The 3D-Jury system can operate in two modes, which differ by the allowed set of considered model pairs. The best-model-mode (3D-Jury-single) allows only one model from each server to be used in the sum, while the all-models-mode (3D-Jury-all) allows the consideration of all models of the servers:

$$3D-Jury-all(M_{a,b}) = \frac{\sum_i^N \sum_{j,a\neq i \text{ OR } b\neq j}^{N_i} sim(M_{a,b}, M_{i,j})}{1+\sum_i^N N_i}$$

$$3D-Jury-single(M_{a,b}) = \frac{\sum_i^N \max_{j,a\neq i \text{ OR } b\neq j}^{N_i} \left(sim(M_{a,b}, M_{i,j})\right)}{1+N}$$

$sim(M_{a,b}, M_{i,j})$: similarity score between model $M_{a,b}$ and model $M_{i,j}$

$3D-Jury-all$: 3D - Jury score in the all - models - mode

$3D-Jury-single$: 3D - Jury score in the best - model - mode

$M_{a,b}$: model number b from the server a

$M_{i,j}$: model number j from the server i

N : number of servers

N_i : Number of top ranking models from the server i (maximum 10)

The 3D-Jury system neglects the confidence scores assigned to the models by the servers. This does not necessary mean that the information about the original scores is lost. It can be expected that highly reliable models produced by fold recognition methods have less ambiguities in the alignments to their template structures, which would result in higher similarity between models generated on templates with the same fold and consequently in higher 3D-Jury scores.

IMPLEMENTATION

The 3D-Jury system was evaluated in the latest LiveBench-6 program. The results presented in Table 1 demonstrate that the 3D-Jury system shows very high sensitivity on the difficult targets while some well tuned sequence alignment methods generate higher quality models for the easy targets. Nevertheless, the number of correct predictions is the highest for some versions of the 3D-Jury system in both categories. A very important criterion is, however, the specificity of the reported

Table 1. The performance of the 3D-Jury system in LiveBench-6

EASY			HARD			ROC		
Name	Sum	All	Name	Sum	All	Name	Mean	First
3DS5	3003	27	3JCa	2018	26	3JA1	49.0	47
3JCa	3002	27	3JAa	2007	29	**PMO3**	49.0	43
3JC1	2917	27	**3DS5**	1945	25	**PMOD**	46.8	38
3DS3	2864	25	3JA1	1890	28	**PMO4**	46.1	34
ST02	2827	26	3JC1	1883	24	3JCa	46.0	35
PCO3	2745	27	**PMO3**	1775	25	**3DS5**	45.0	24
PMO4	2739	26	**PMOD**	1756	26	3JC1	44.9	38
RBTA	2731	25	**3DS3**	1690	24	3JAa	44.9	38
PMO3	2717	27	**PMO4**	1670	24	**PCO3**	44.6	27
3JA1	2702	26	**SHGU**	1649	22	**PCO2**	44.6	34
SHGU	2686	25	**PCO2**	1608	24	**3DS3**	43.2	33
PCO4	2683	26	**PCO3**	1593	21	**SHGU**	42.9	35
FFA3	2648	26	**PCO4**	1454	22	ORFs	42.8	38
FUG3	2647	25	**RBTA**	1439	20	ST02	40.4	37
ORFs	2629	27	ORFs	1413	20	**PCO4**	38.3	27
3JAa	2626	26	ST02	1366	19	FFA3	37.3	19
SFPP	2553	24	INBG	1343	21	INBG	36.8	23
FUG2	2543	24	FFA3	1213	18	FUG2	35.6	13
PMOD	2521	25	3DPS	1157	16	FUG3	35.3	11
INBG	2514	24	FUG2	1134	19	RAPT	34.9	28
3DPS	2513	24	FUG3	1111	17	SFPP	34.6	17
MGTH	2420	24	SFPP	1087	16	MGTH	34.0	22
ORFb	2404	22	MGTH	1081	16	SFAM	32.7	11
RAPT	2392	25	SFAM	1030	16	ORFb	32.5	8

The table shows the sensitivity of several structure prediction servers on 32 easy (EASY) and 64 difficult (HARD) targets and the specificity score (ROC; Swets *et al.*, 2000) computed on all 96 targets. For each of the three evaluations only the top 25 servers are shown. The evaluated servers are indicated in the NAME column using a fourletter abbreviation code (please view the original LiveBench pages for more information about the servers). The four 3D-Jury versions are marked with shaded background. 3JA1 and 3JAa use a set of eight threading servers for consensus building while 3JC1 and 3JCa use all prediction servers, including other meta-predictors. 3JA1 and 3JC1 use the best-model-scoring mode (only one model per server is used for consensus building) while 3JAa and 3JCa use the all-models-scoring mode (all models from the servers are used for consensus building). Other meta-predictors or servers that produce models from spliced fragments of several structural templates are shown in bold (PMO[X] and PCO[X] belong to the Pcons series; SHGU and 3DS[X] belong to the 3D-SHOTGUN series; RBTA indicates Robetta). The ALL column reports the number of correct models generated for easy or difficult targets by each server. A correct model is defined as a model where at least 40 C-alpha atoms (correct atoms) can be superimposed on the native structure within 3.0 Å. The SUM column sums the number of correct atoms over all correct models for each server. The sensitivity ranking is based on the SUM column. The FIRST column reports the number of correct predictions with higher confidence score than the first wrong prediction (less than 31 correct atoms). 'Borderline' predictions, between 31 and 39 correct atoms, are ignored. The MEAN column shows the average number of correct predictions obtained with a higher confidence score than the first 1–10 false predictions. The specificity ranking is based on the MEAN column. The exact ranking of all servers is subject to frequent changes and many differences cannot be regarded as significant.

confidence score. The best results are obtained with the 3D-Jury system operating in the best-model-mode on a set of eight servers (ORFeus Pas *et al.*, 2003; SamT02, Karplus *et al.*, 2001; FFAS03, Rychlewski *et al.*, 2000; mGenTHREADER, Jones, 1999; INBGU, Fischer, 2000; RAPTOR, Xu *et al.*, 2003; FUGUE-2, Shi *et al.*, 2001; 3D-PSSM, Kelley *et al.*, 2000). The score obtained with this setting is reported as default on the Meta Server pages (http://BioInfo.PL/Meta/), which is the current interface to the 3D-Jury system. The significance cutoff of 50 has been chosen, which results in a prediction accuracy of above 90%. As the main difference to other consensus methods, the interface enables the selection of servers used for consensus building and the selection between the two score summing modes by the user.

DISCUSSION

The 3D-Jury system follows a simple protocol that can be easily reproduced and incorporated into other fold recognition programs. This addition is likely to boost the quality of the predictions. However the system does not guarantee that the correct model will be selected from a set of preliminary models, especially if the correct solution is an outlier and is provided by only a single server.

In contrast to some meta-predictors (i.e. 3D-SHOTGUNS, Pmodeller or ROBETTA; Simons *et al.*,

1997) the 3D-Jury system is not capable of improving the model (template) structures. It can only change the final ranking of the reported models. Nevertheless, because of its versatility, it can be easily placed on top of methods that modify the template structures as an additional jury module. This is currently possible via the Meta Server interface, where some fragment splicing methods are available.

REFERENCES

Baker,D. and Sali,A. (2001) Protein structure prediction and structural genomics. *Science*, **294**, 93–96.

Bonneau,R., Ruczinski,I., Tsai,J. and Baker,D. (2002) Contact order and *ab initio* protein structure prediction. *Protein Sci.*, **11**, 1937–1944.

Bujnicki,J.M., Elofsson,A., Fischer,D. and Rychlewski,L. (2001) LiveBench-2: Large-scale automated evaluation of protein structure prediction servers. *Proteins*, **45** (Suppl. 5), 184–191.

Fischer,D. (2000) Hybrid fold recognition: combining sequence derived properties with evolutionary information. *Pac. Symp. Biocomput.*, 119–130.

Fischer,D. (2002) 3D-SHOTGUN: a novel, cooperative, fold-recognition meta-predictor. *Proteins*.

Fischer,D., Elofsson,A., Rychlewski,L., Pazos,F., Valencia,A., Rost,B, Ortiz,A.R. and Dunbrack,R.L.Jr. (2001) CAFASP2: the second critical assessment of fully automated structure prediction methods. *Proteins*, **45** (Suppl. 5), 171–183.

Jones,D.T. (1999) GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. *J. Mol. Biol.*, **287**, 797–815.

Karplus,K., Karchin,R., Barrett,C., Tu,S., Cline,M., Diekhans,M., Grate,L., Casper,J. and Hughey,R. (2001) What is the value added by human intervention in protein structure prediction? *Proteins*, **Suppl. 5**, 86–91.

Kelley,L.A., MacCallum,R.M. and Sternberg,M.J. (2000) Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J. Mol. Biol.*, **299**, 499–520.

Lundstrom,J., Rychlewski,L., Bujnicki,J. and Elofsson,A. (2001) Pcons: a neural-network-based consensus predictor that improves fold recognition. *Protein Sci.*, **10**, 2354–2362.

Marr,D. (1982) *Vision*. Freeman, San Francisco.

Moult,J., Fidelis,K., Zemla,A. and Hubbard,T. (2001) Critical assessment of methods of protein structure prediction (CASP): Round IV. *Proteins*, **45** (Suppl. 5), 2–7.

Pas,J., Wyrwicz,L.S, Grotthuss,M., Bujnicki,J.M., Ginalski,K. and Rychlewski,L. (2003) ORFeus: detection of distant homology using sequence profiles and predicted secondary structure. *Nucleic Acids Res.*.

Rychlewski,L., Jaroszewski,L., Li,W. and Godzik,A. (2000) Comparison of sequence profiles. Strategies for structural predictions using sequence information. *Protein Sci.*, **9**, 232–241.

Sali,A. and Blundell,T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.*, **234**, 779–815.

Shi,J., Blundell,T.L. and Mizuguchi,K. (2001) FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties. *J. Mol. Biol.*, **310**, 243–257.

Siew,N., Elofsson,A., Rychlewski,L. and Fischer,D. (2000) MaxSub: an automated measure for the assessment of protein structure prediction quality. *Bioinformatics*, **16**, 776–785.

Simons,K.T., Kooperberg,C., Huang,E. and Baker,D. (1997) Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J. Mol. Biol.*, **268**, 209–225.

Swets,J.A., Dawes,R.M. and Monahan,J. (2000) Better decisions through science. *Sci. Am.*, **283**, 82–87.

Xu,J., Li,M., Lin,G., Xu,Y. and Kim,D. (2003) Protein threading by linear programming. *Pac. Symp. Biocomput.*, 264–275.