BIOINFORMATICS

Vol. 20 no. 14 2004, pages 2228–2235
doi:10.1093/bioinformatics/bth232

Arby: automatic protein structure prediction using profile–profile alignment and confidence measures

Niklas von Öhsen[1,*,†], Ingolf Sommer[2,†], Ralf Zimmer[3] and Thomas Lengauer[2]

[1]Institute for Algorithms and Scientific Computing (SCAI), Fraunhofer Gesellschaft, Schloss Birlinghoven, Sankt Augustin, 53754, Germany, [2]Max-Planck-Institut für Informatik, Stuhlsatzenhausweg 85, Saarbrücken, 66123, Germany and [3]Institut für Informatik, Ludwig-Maximilians-Universität München, Amalienstrasse 17, München, 80333, Germany

Received on December 12, 2003; revised on February 13, 2004; accepted on February 16, 2004
Advance Access publication April 1, 2004

ABSTRACT

Motivation: Arby is a new server for protein structure prediction that combines several homology-based methods for predicting the three-dimensional structure of a protein, given its sequence. The methods used include a threading approach, which makes use of structural information, and a profile–profile alignment approach that incorporates secondary structure predictions. The combination of the different methods with the help of empirically derived confidence measures affords reliable template selection.

Results: According to the recent CAFASP3 experiment, the server is one of the most sensitive methods for predicting the structure of single domain proteins. The quality of template selection is assessed using a fold-recognition experiment.

Availability: The Arby server is available through the portal of the Helmholtz Network for Bioinformatics at http://www.hnbioinfo.de under the protein structure category.

Contact: Niklas.von-Oehsen@scai.fhg.de

INTRODUCTION

With the advent of large-scale genome sequencing programmes and the completion of the human genome project the importance of protein structure determination has become a main challenge of structural molecular biology. Theoretical methods for protein structure prediction are of particular importance since they offer the possibility of supporting experimental methods as well as of validating the scientific theory behind the principles of protein structure. Fully automated protein structure prediction is a significant challenge and has been an unreached goal of research in the field since it allows the possibility of decoupling the quality of the prediction from the expertise of the person using the method. The rising interest in this field is demonstrated by the biennial CAFASP experiments for the critical assessment of fully automated structure prediction (Fischer *et al*., 2001). Starting with 11 servers in 1998, the third CAFASP experiment had 55 participating servers in 2002. Furthermore, the prediction quality of the best servers was much more accurate than a large fraction of the manual predictions.

The improvement of server performance over the last years has several reasons. Currently, most servers are based on a homology principle which states that similar sequences will most likely have a similar structure. Thus, the performance of these servers automatically increases due to the continuous growth of the number of resolved structures. On the other hand, there are algorithmic developments that lead to more sensitive methods. One of these developments is the profile–profile alignment approach that has been introduced by Rychlewski *et al*. (2000) as an extension of the profile–sequence alignment concept by Gribskov *et al*. (1987) and is used by a number of top performing servers. We have recently developed a new scoring function for the profile–profile approach that is now used in the Arby server (von Öhsen and Zimmer, 2001; von Öhsen *et al*., 2003).

The combination of the results of several fold-recognition methods is a problem that aroused great interest recently. The big success of the meta server approach (integration of the outputs of several other servers into a new prediction) at the CAFASP3 experiment has especially focused the attention on the selection and the combination of multiple fold recognition results. While some methods analyze similarities of the three-dimensional (3D) models created by the different methods (e.g. Ginalski *et al*., 2003a) we follow a simple yet effective

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First authors.

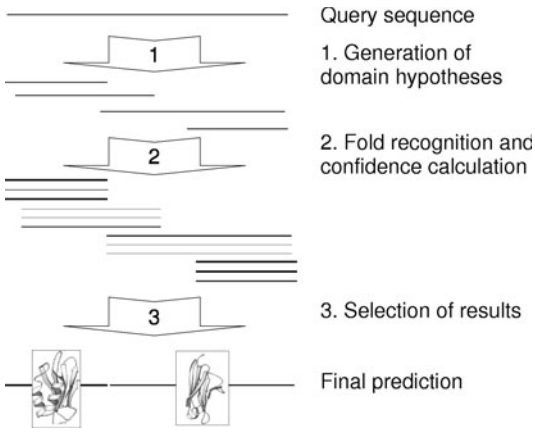

Fig. 1. The steps of the Arby server prediction pipeline.

approach based on confidence measures. These are figures indicative of the significance of the result of a database search performed by an alignment method, like the BLAST *E*-value. In a recent study on this subject, the score-gap function was identified as a possible confidence measure that yields suitable results for a range of alignment methods (Sommer *et al*., 2002). This function is used by the Arby server to identify the most significant database hits from the outputs generated by several fold recognition methods.

ALGORITHM

We first present a brief overview of the server pipeline (Fig. 1). The server obtains from the user interface the protein sequence whose structure is to be predicted. This sequence is referred to as query or target sequence. During the first step of the computation, the Arby server generates several estimates for the domain structure of the query sequence. The resulting putative domain sequences are forwarded to the fold recognition step. Fold recognition is the computationally most expensive step: for each of the subsequences from the previous stage and all five implemented fold recognition methods, the highest scoring template in the database is computed. This approach of first segmenting the target sequence and then applying different methods of fold recognition to each segment produces a large amount of computational workload. The advantage of this method is that we can use the global version of all alignment methods during the fold recognition stage. As a previous benchmark showed, the quality of the non-PSI-Blast fold recognition methods we use is generally higher in global alignment than in local alignment mode (Sommer *et al*., 2002). This is most probably due to the additional information contained in the lengths of the protein sequences (homologous domains will more likely have similar length and length differences are penalized by global alignment methods).

The results of the fold recognition stage are then collected and annotated with confidence values. These values allow for assessing the significance of the database hits independently of the underlying fold recognition method. Furthermore, all alignments of the selected templates against the query subsequence are computed. In a final step, a selection of the most promising collection of database hits is computed. The final selection yields one or more templates together with non-overlapping subsequences of the query sequence. Each of these is predicted to be a separate structural domain of the query protein. In the following section, the steps are explained in detail.

Generation of initial domain hypotheses

The first step of the prediction algorithm seeks to identify the domain structure of the protein from the sequence alone. Since this is not an easy task (and itself subject to research, e.g. Nagarajan and Yona, 2003), a set of subsequences corresponding to potential domains is generated using database searches as well as heuristic methods. The goal is to produce a set of subsequences that contain good approximations to the actual domains of the protein. If this is achieved and a good template for a domain is found in the following fold recognition stage, the fold recognition methods will assign the highest scores to those subsequences that match the top template best and thus the original domain best.

First, the query sequence is checked for homologs in current domain sequence databases. This is performed using a PSI-Blast (Altschul *et al*., 1997) run against the ProDom database (Servant *et al*., 2002) and a scan of the InterPro database using the InterProScan tool (Apweiler *et al*., 2001). If any hits are found, then these are added to the set of subsequences, along with their complements and unions of consecutive fragments.

Furthermore, a secondary structure prediction for the complete target sequence is computed using the PSIPRED tool (Jones, 1999). The boundaries of predicted secondary structure elements are determined by identifying loop and coil regions and cutting within these at least at every 25th amino acid position. Then all fragments that are shorter than 40 amino acids are removed and the full target sequence is added to the set. Finally, duplicates and highly similar segments are removed and the size of the set is constrained to a value considered feasible in terms of computation time. For the CAFASP experiment, we chose a maximum of 50 subsequences (roughly corresponding to 12–24 h of compute time). The selection algorithm takes care that the database hits and the full sequence are not accidentally removed and iteratively throws out one of the two most similar fragments in the set.

Template identification

This step is the computationally most intensive part of the algorithm. All the fragments from the previous stage are used as query sequences for all five fold recognition methods described below to search the database of template domains.

At the time of the CAFASP3 experiment, this database was a representative set of the SCOP 1.57 of size 3951 taken from the ASTRAL server (Lo Conte *et al*., 2002; Brenner *et al*., 2000) with a maximum of 40% pairwise sequence identity. The five fold recognition methods used are the following:

- *PSI-Blast (PSI).* The PSI-Blast program has already been used in the previous stage. Since the reliability of the results in the context of fold recognition is usually very high and is also used to identify templates. For this purpose, PSI-Blast is run with three iterations against the union of the nr database [a non-redundant protein sequence database, Wheeler *et al*. (2000)] and the sequences from our template domain database. If there is a hit in the template domain database then the most significant of all hits in the database is forwarded to the next stage. The multiple alignment is retrieved from the output of the PSI-Blast run and converted into a frequency profile using a sequence weighting scheme by Henikoff and Henikoff (1994). This frequency profile is used in some of the other fold recognition methods.
- *123D threading (123D).* The basic version of the 123D fast dynamic programming threading program is described by Alexandrov *et al*. (1996). It uses the contact capacity potential approach to sequence–structure alignment. This is the only fold recognition method that is completely independent of the results from the PSI-Blast program. Thus, it serves as a fallback algorithm in case the PSI-Blast run fails due to technical or algorithmic problems (e.g. bias introduced through the sequence database).
- *123D profile threading (123D*).* An extension of the original 123D threading program that makes use of the frequency profiles on the target side. These profiles are computed from the previous PSI-Blast run. This method typically has higher predictive power than the plain 123D sequence–structure alignment version. The 18 parameters of the empirical threading scoring function of both 123D methods were adopted from Zien *et al*. (2000).
- *Profile–profile alignment using log average scoring (PPL).* This alignment method has recently been developed and shown to have considerable advantages over some established methods in terms of fold recognition performance (von Öhsen and Zimmer, 2001; von Öhsen *et al*., 2003). It is also a generalization of the common log-odds score for sequence similarity and thus based on an amino acid similarity model [in this application BLOSUM62, Henikoff and Henikoff (1992)]. Since frequency profiles are also needed for the templates when using this method, these have been precomputed (using the construction method mentioned above) and stored in the template database.
- *Profile–profile alignment using log average scoring with a secondary structure term (PPS).* This is a recent addition to the profile–profile alignment method above. The use of secondary structure predictions can have a positive effect on the alignment quality of a method (Elofsson, 2002). In our approach, we compute a secondary structure prediction using the PSIPRED program. The algorithm yields not just a discrete three-state secondary structure sequence but, alternatively, the likelihood that a corresponding secondary structure state exists is assumed at a certain position. Thus, we have a frequency profile over the three-state secondary structure alphabet available for the target. The corresponding predictions for the template sequences have been precomputed and stored in the template database. Analogous to the BLOSUM62 model used in the amino acid profile–profile score, the secondary structure similarity model from Kawabata and Nishikawa (2000) was used to compute a secondary structure profile–profile scoring term according to the log average score formula. For a position of the target sequence and a position of the template sequence, the alignment score is thus the amino acid profile–profile score plus a weighted secondary structure profile–profile score. The parameters for this method (gap insertion, gap extension, secondary structure weight) have been optimized using the linear programming approach mentioned above (Zien *et al*., 2000).

The latter four methods use global dynamic programming alignment algorithms. This penalizes length differences between the target subsequence and template sequences. Since a target subsequence that is a good approximation of a domain most likely has similar length to the correct database hit, the fragments that match the actual domain boundaries best will get higher scoring results.

The different fold recognition methods have been chosen to complement each other in several respects: the PSI-Blast program yields highly reliable results for targets for which a close homolog exists in the template database. The 123D threading and the profile–profile alignment method are both targeted at the detection of remote homologues that are missed by the PSI-Blast method. In a recent study, it was shown that the combination of these two methods can potentially yield a significant improvement in FR quality especially for hard-to-predict targets (von Öhsen *et al*., 2003). While each of the methods was able to correctly recognize approximately the same number of targets, the recognition performance could theoretically have been increased by 50 % using an optimal combination of the two methods. The combination of all these fold recognition methods therefore covers a broad range of different target characteristics with specialized applications (see the benchmark section for evaluation).

Confidence measures

The problem of identifying the method, which provides the best prediction is approached using confidence measures. A confidence measure is a function that yields a real value indicating the reliability of the computed fold prediction. Several possible functions that can be applied to the result of alignment-based template database searches have been studied recently (Sommer *et al*., 2002). One result from this study was the fact that the score-gap function is a robust confidence measure which can easily be computed and which performs well over a number of different fold recognition methods with different parameters. The score-gap function is defined as the difference in alignment score between the top hit in the template database and the next best template from a different SCOP fold. If this difference is large the hit is considered more significant, in the sense that it has a larger chance of being correct compared with the case where the difference is small. In the latter case, either the method cannot distinguish between the correct and the wrong templates or the correct template is not in the database. The score-gap function is used to calculate a confidence value for all methods except PSI-Blast. If the PSI-Blast program returns a hit, it is assigned a high confidence value, such that hits from PSI-Blast are preferred over results from the other four methods. For the remaining four methods, the score-gap function has been calibrated using a benchmark set of 2232 target sequences (Sommer *et al*., 2002). The calibration was performed by constructing a function that calculates for each score-gap value, the number of false predictions observed in the benchmark that have a greater or equal score-gap value. This value, divided by the size of the benchmark set, is an estimator for the probability that the prediction is false, given the score-gap value (a p-value). Since the data is sparse in the regions of high significance, a linear approximation was used to compute the values for large score gaps. The resulting p-values are comparable between the different methods and yield an absolute measure of reliability for the fold prediction of each target fragment for each method.

Alignments

The fold recognition step results in a set of annotated subsequences, each containing the following information: the subsequence of the target sequence, the name of the fold recognition method used, the highest scoring template domain and the p-value denoting the significance of the hit. Hence, there are four different subsequence annotations for each subsequence from the initial set of domain hypotheses if there was no PSI-Blast hit and five otherwise. For each annotated subsequence we compute the alignment of the target subsequence against the template using the same method, store it in the annotation and pass it on to the next stage. The alignment method used here is a dynamic programming profile–profile alignment with secondary structure term without penalty for the rightmost and leftmost gaps. This variant is chosen because the goal is to obtain an optimal alignment even if the target subsequence is longer than the template domain. The resulting alignments thus typically cover only part of a longer target subsequence (Fig. 2). This is taken into account in the final selection step.

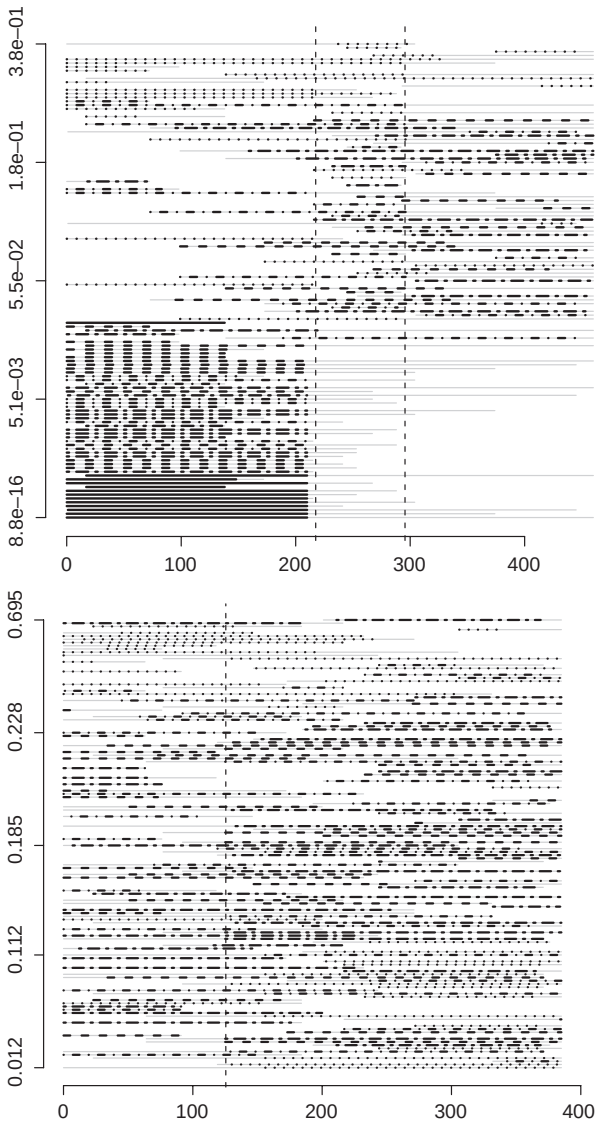

Fig. 2. Two examples of data at the end of the fold recognition stage. On the horizontal axis the sequence position in the target sequence is shown. On the vertical axis the fragments are enumerated, sorted by significance (highest significance at bottom). The target subsequence is depicted in gray, the part that is covered by the alignment against the template is black. Line styles stand for the different methods: plain, PSI-Blast; dotted, 123D; dash-dotted, 123D profile; dashed, profile–profile; double-dashed, profile–profile with secondary structure. Top: 1mu5, chain A. Three SCOP domains at positions 0–217, 218–295 and 296–459 indicated by vertical dashed lines. The first domain has a PSI-Blast hit that all other methods recognize, too. Bottom: 1miw, chain A. Two SCOP domains at positions 0–125 and 126—384, without PSI-Blast hits.

Selection of fragments

At this step we have a maximum of 200–250 subsequences annotated with a predicted fold, a confidence measure and an alignment against the template. From these annotated subsequences, we want to select few non-overlapping subsequences to be predicted as the domains of the target protein. This is done by optimizing a heuristic quality score.

The quality score is defined per sequence position. It consists of several terms: the first term is the score gap confidence of the selected annotated subsequences as described earlier. Additionally, the alignment against the template is analyzed and a secondary structure quality score is determined which evaluates the correspondence of the predicted secondary structure with the known secondary structure of the template. Finally, bonus and penalty terms for consenting or contradicting fold predictions in the set of annotated subsequences are added. Sufficiently overlapping sequence annotations with the same SCOP fold predicted by different methods get a bonus, whereas annotations with disagreeing SCOP folds are penalized.

A default positional score is chosen for positions of the query sequence that are not covered by any template-annotated subsequence such that the occurrence of these is effectively penalized. The selection of non-overlapping annotated subsequences that optimizes the sum of positional quality scores over all positions of the query sequence is determined using dynamic programming.

Along with the final result of the selection algorithm all data available as input to the selection stage are given as appendix to the output for manual inspection. For each subsequence and fold recognition method this includes the template with its alignment against the query subsequence as well as the score-gap confidence values.

RESULTS

The performance of the Arby server is analyzed based on the comparison with other servers during the CAFASP3 experiment and by a separate evaluation on a fold recognition benchmark.

Analysis of the CAFASP3 results

The Arby server took part in the CAFASP3 experiment in 2002. Predictions were computed for all of the 67 target proteins. All domains from these targets have been separated into different classes according to the difficulty of the prediction. The category of interest for our server is the FR category, since the methods we use are specialized for the detection of remote homology relationships. The evaluation of the CAFASP3 results was presented at the fifth CASP community meeting (December 2002, Asilomar), was published recently (Fischer *et al*., 2003) and is available on the Web (http://www.cs.bgu.ac.il/~dfischer/CAFASP3). The evaluation data includes scores from the MaxSub program

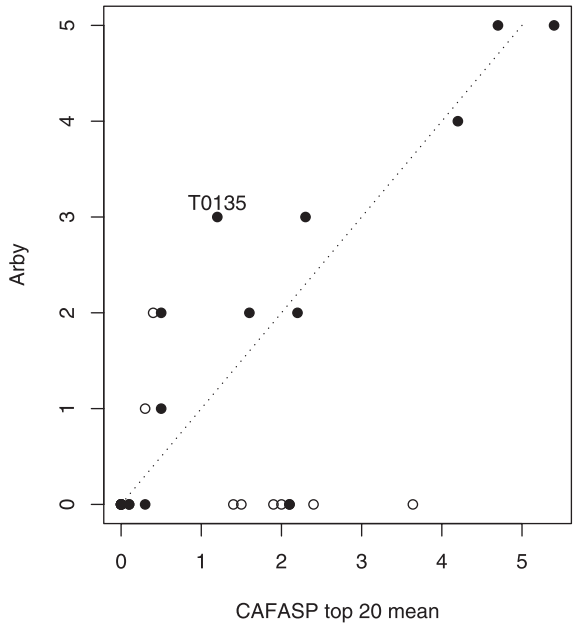

Fig. 3. MaxSub scores for the CAFASP3 targets from the fold recognition category. Horizontal axis: Arby Score, vertical axis: mean of the Top 20 CAFASP servers for this target. Solid: single domain target, Other: multidomain target. The Arby prediction for target T0135, a boiling stable protein, was particularly good.

(Siew *et al*., 2000) indicating the quality of the superposition between the model and the resolved structure available for each server and each target.

In Figure 3, the MaxSub scores of the Arby server and the mean of the top 20 CAFASP servers for a specific target are plotted against each other for each target. The identity is indicated by the dotted line. For the single domain proteins, most of the targets are above the diagonal, indicating that the Arby server outperforms the average of the top 20 CAFASP servers for this target. However, there are several targets in the plot, which were completely missed by the Arby server (on the x-axis). These targets are mostly domains from multidomain proteins. This was caused by the final selection function that is the last step in the server process. We are currently evaluating several different alternatives for improving this part of the server functionality.

An example for the quality of the predictions generated by the Arby server is target T0135 where the Arby server provided the best CAFASP prediction according to the CAFASP MaxSub and GDT_TS scores (Fischer *et al*., 2003).

In the results table of the CAFASP3 assessment by D.Fischer, the sum of the scores plotted in Figure 3 was calculated for each server and the servers were ranked according to their results. To improve the stability of the ranking, $n-1$ ranks were calculated by taking the highest (i.e. minimal) rank that a server reached in the n rankings that are defined by leaving one of the n targets out of the calculation of the sum.

The Arby server reached a $(n-1)$-rank of 13 in the ranking of the FR category which corresponds to a real rank of 17 of a total of 51 servers participating. It should be noted that the 16 different servers doing better originate from 7 working groups (some of these took part with several variations of the same method). About 7 of the 16 servers which outperformed the Arby server were meta servers, i.e. servers that make use of structure predictions generated by other servers. Contrary to this, the Arby server belongs to the group of individual servers that calculate the prediction independently of other servers.

In order to remove the effect of the domain prediction part, we recalculated the results table using only the single domain proteins from the targets of the FR category (data not shown). While most of the ranking remained stable, the Arby server reached the third rank in this table [corresponding $(n-1)$-rank was 2], outperformed by the ORFeus and 3D shotgun servers (Ginalski *et al*., 2003b; Fischer, 2003).

An independent assessment of the CASP5 results by M. Levitt (Fischer *et al*., 2003) with a different evaluation scheme showed the Arby server as the best of all individual (i.e. not meta) servers in the FR category with three meta servers outperforming Arby. The position of Arby according to this ranking is one of two cases in which the two evaluation methods by D.Fischer and M.Levitt yield significantly different results. As mentioned in the paper on the CAFASP assessment, the reasons for this remain elusive. A possible explanation might be the high variability of the Arby server performance on single versus multidomain targets. This variability emphasizes differences in how the two evaluation methods combine the results of different targets into a single score.

Fold recognition benchmarks

We selected a number of test targets for the Arby server according to the following protocol in order to assess the fold recognition performance of the implemented methods: all protein chains that contain a domain from the 40% ASTRAL subset of SCOP 1.63 were considered. Chains that contain a domain present in SCOP 1.57 were removed from this set. From the resulting 1231 protein chains, 160 were selected in order to downscale the computation requirements. These 160 chains contain 237 ASTRAL domains that were used for the evaluation. For a protein in the test set and each domain d from the ASTRAL set that it contained and each fragment f delivered by the initial domain hypothesis generation, we calculated the offset error defined by $|a_d - a_f| + |b_d - b_f|$ where a and b denote starting and ending index with respect to the full sequence. Then we selected the fragment f_d with minimal offset error (and the longer one, if the error is the same). For the fragments f_d, we analyzed the templates that the different fold recognition methods returned. Thus, the performance of these methods can be assessed independently of the domain selection algorithm.

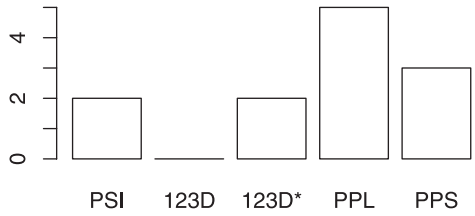

Fig. 4. Number of predictions uniquely improved by the methods.

Diversity of fold-recognition methods. In this section the question whether all five fold recognition methods contribute to the fold recognition performance of the ensemble shall be examined. For each fold recognition method, we counted for how many target domains the method returned a template, i.e. at least from the correct SCOP fold and is related to the target domain on a higher SCOP level (e.g. SCOP superfamily instead of SCOP fold) than the templates returned by the other methods. Thus, each count stands for a case in which the template identification ability of the ensemble of the five methods is actually improved by the single method.

Figure 4 shows that for all methods except for the plain 123D threading there exist targets that no other method could predict with the same quality (the 123D method is included for stability reasons rather than predictive performance, see above). Therefore, it is potentially advantageous to use a combination of the results of all the methods. In the following section, the question is addressed whether the confidence measure approach of the Arby server is effective in this regard.

Performance improvement by confidence measures. For further analysis, we used the four alignment methods plus PSI-Blast that were introduced in the section 'template identification'. Two other, 'virtual', fold recognition methods were constructed as follows:

- *Best confidence (ARBY).* Always selects the method that returns the result with highest confidence value from the original methods. This is the method that is used by the Arby server.
- *Optimal (OPT).* With knowledge of the correct fold recognition result, always selects the result from the original methods that suggests the most closely related template to the query domain (in terms of SCOP relatedness). The performance of this method is an upper bound for the performance of the best possible fold recognition method constructed from the output of the original five methods.

If the PSI-Blast method produces a hit for a domain then this is given priority over the other methods. For the evaluation of the performance of the confidence measure approach, we therefore looked only at those domains from the benchmark set that returned no PSI-Blast hit. Of the total 237 domains, the PSI-Blast program returned a hit in 163 cases

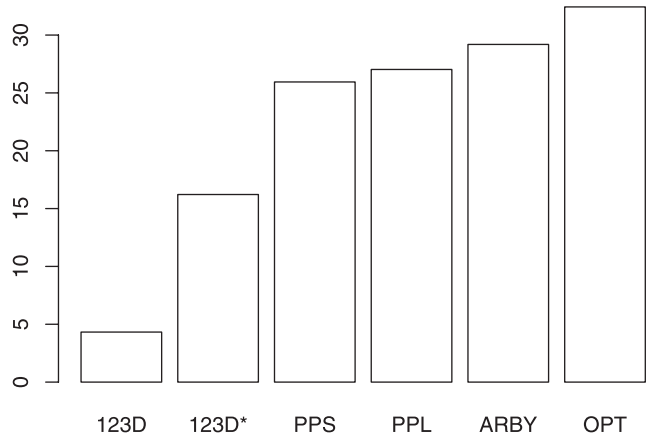

Fig. 5. Beyond PSI-Blast: number of domains without PSI-Blast hit that were correctly recognized (correct SCOP fold). The total number of domains in this set is 74. Of these, 54 had a template from the correct SCOP fold in the template database (i.e. were recognizable). See text for labels.

(68.8%). About 3 of the associated 163 templates did not have the correct SCOP fold (i.e. fold recognition failures). The performance of the other fold recognition methods on the remaining 74 domains is shown in Figure 5.

The plot shows that the single most powerful prediction method used by the Arby server is the PPL method. The confidence measure approach is able to combine the diverse methods and improve over their single performances by a noticeable margin. The OPT method shows that there is still room for improvement for the combination of the different fold recognition results.

CONCLUSION

By construction, the Arby server is best suited for proteins for which remote homologs exist in the structural databases but are hard to detect. The server combines several fold recognition methods that are based on dynamic programming alignment and make use of secondary and 3D structure information. The diversity of these methods leads to an improvement in template identification by the ensemble. The combination of the results by the use of confidence measures is shown to be successful in that it is able to improve over the performance of the best performing individual methods.

The results from the CAFASP3 experiment show that the Arby server is currently among the top ten servers for predicting the structure of proteins that do not have a close homologue (e.g. PSI-Blast hit) in a structural database. In a CASP5 assessment by M.Levitt it is mentioned as the best performing individual server in this category. The Arby server currently reaches its best prediction performance for single domain proteins. Future work will be directed towards the optimization of the performance on multidomain proteins which is currently limited by the performance of the last stage of the server. The Arby server is available via the portal of the Helmholtz Network for Bioinformatics at http://www.hnbioinfo.de

ACKNOWLEDGEMENTS

The authors thank Theo Mevissen and Mario Albrecht for their help during the development of the server and Joannis Apostolakis for his helpful comments on the manuscript. Part of this work has been supported by the BMBF project ‘Helmholtz Network for Bioinformatics’ (01SF9984/3) and the BMBF project ‘Development of Microbalance Array/Mass Spectrometry as a Tool for Functional Proteomics’ (0312708).

REFERENCES

Alexandrov,N., Nussinov,R. and Zimmer,R. (1996) Fast protein fold recognition via sequence to structure alignment and contact capacity potentials. In Hunter,L. and Klein,T.E., (eds), *Pacific Symposium on Biocomputing’96*. World Scientific Publishing Co., Singapore, pp. 53–72.

Altschul,S.F., Madden,T.L., Schäffer,A.A., Zang,J., Zang,Z., Miller,W. and Lipman,D.V. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

Apweiler,R., Attwood,T.K., Bairoch,A., Bateman,A., Birney,E., Biswas,M., Bucher,P., Cerutti,L., Corpet,F., Croning,M.D.R., *et al.*, (2001) The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res.*, **29**, 37–40.

Brenner,S.E., Koehl,P. and Levitt,M. (2000). The ASTRAL compendium for protein structure and sequence analysis. *Nucleic Acids Res.*, **28**, 254–256.

Elofsson,A. (2002) A study on protein sequence alignment quality. *Prot. Struct. Funct. Genet.*, **46**, 330–339.

Fischer,D. (2003) 3D-SHOTGUN: a novel, cooperative, fold-recognition meta-predictor. *Prot. Struct. Funct. Genet.*, **51**, 434–441.

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

Fischer,D., Rychlewski,L., Dunbrack,R.L., Ortiz,A.R. and Elofsson,A. (2003) CAFASP3: the third critical assessment of fully automated structure prediction methods. *Prot. Struct. Funct. Genet.*, **53**, 503–516.

Ginalski,K., Elofsson,A., Fischer,D. and Rychlewski,L. (2003a) 3D-Jury: a simple approach to improve protein structure predictions. *Bioinformatics*, **19**, 1015–1018.

Ginalski,K., Pas,J., Wyrwicz,L.S., Grotthuss,M.V., Bujnicki,J.M. and Rychlewski,L. (2003b) ORFeus: detection of distant homology using sequence profiles and predicted secondary structure. *Nucleic Acids Res.*, **31**, 3804–3807.

Gribskov,M., McLachlan,A.D. and Eisenberg,D. (1987) Profile analysis: Detection of distantly related proteins. *Proc. Natl Acad. Sci., USA*, **84**, 4355–4358.

Henikoff,S. and Henikoff,J.G. (1992) Amino acid substitution matrices from protein blocks. *Proc. Natl Acad. Sci., USA*, **89**, 10915–10919.

Henikoff,S. and Henikoff,J.G. (1994) Position-based sequence weights. *J. Mol. Biol.*, **243**, 574–578.

Jones,D.T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.*, **292**, 195–202.

Kawabata,T. and Nishikawa,K. (2000) Protein structure comparison using the markov transition model of evolution. *Prot. Struct. Funct. Genet.*, **41**, 108–122.

Lo Conte,L., Brenner,S.E., Hubbard,T.J.P., Chothia,C. and Murzin,A.G. (2002) SCOP database in 2002: refinements accommodate structural genomics. *Nucleic Acids. Res.*, **30**, 264–267.

Nagarajan,N. and Yona,G. (2003) A multi-expert system for the automatic detection of protein domains from sequence information. In Miller,W., Vingron,M., Istrail,S., Pevzner,P. and Waterman,M., (eds), *RECOMB 2003, Proceedings of the Seventh Annual International Conference on Computational Biology*. The Association for Computing Machinery, NY, 224–234.

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.

Servant,F., Bru,C., Carrere,S., Courcelle,E., Gouzy,J., Peyruc,D. and Kahn,D. (2002) Prodom: automated clustering of homologous domains. *Brief. Bioinformatics*, **3**, 246–251.

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.

Sommer,I., Zien,A., von Öhsen,N., Zimmer,R. and Lengauer,T. (2002) Confidence measures for protein fold recognition. *Bioinformatics*, **18**, 802–812.

von Öhsen,N., Sommer,I. and Zimmer,R. (2003) Profile–profile alignment: a powerful tool for protein structure prediction. In Altman,R.B., Dunker,A.K., Hunter,L., Jung,T.A. and Klein,T.E., (eds), *Pacific Symposium on Biocomputing 2003*. World Scientific Publishing Co. Pte. Ltd., Singapore, pp. 252–263.

von Öhsen,N. and Zimmer, R. (2001) Improving profile–profile alignment via log average scoring. In Gascuel,O. and Moret,B.M.E., (eds), *Proceedings of the First International Workshop (WABI 2001) on Algorithms in Bioinformatics*, Aarhus, Denmark, August, Springer-Verlag, Berlin, Heidelberg, New York, pp. 11–26.

Wheeler,D.L., Chappey,C., Lash,A.E., Leipe,D.D., Madden,T.L., Schuler,G.D., Tatusova,T.A. and Rapp,B.A. (2000) Database resources of the National Center for Biotechnology Information. *Nucleic Acids. Res.*, **28**, 10–14.

Zien,A., Zimmer,R. and Lengauer,T. (2000). A simple iterative approach to parameter optimization. *J. Comput. Biol.*, **7**, 483–501.