Acta Biochimica Polonica

Vol. 52 No. 1/2005

267–269

QUARTERLY

Communication

MOFOID — not only the protein modeling server

Pawel Szczesny[1☼], Grzegorz Wieczorek [1] and Piotr Zielenkiewicz [1, 2 ✉]

[1]*Institute of Biochemistry and Biophysics, Polish Academy of Sciences,* [2]*Plant Molecular Biology Laboratory, Warsaw University, Warszawa, Poland*

Received: 07 January, 2005; accepted: 02 March, 2005

Key words: homology modeling, protein structure prediction

MOFOID is a new server developed mainly for automated modeling of protein structures by their homology to the structures deposited in the PDB database. Selection of a template and calculation of the alignment is performed with the Smith-Waterman or Needleman-Wunsch algorithms implemented in the EMBOSS package. The final model is built and optimised with programs from the JACKAL package. The wide spectrum of options in the web-based interface and the possibility of uploading user's own alignment make MOFOID a suitable platform for testing new approaches in the alignment building. The server is available at https://valis.ibb.waw.pl/mofoid/.

There are over a million protein sequences known. Several tools developed in the last few years address the problem of predicting the function of these proteins, but sometimes visualisation of their three-dimensional structure is crucial for proper designing of an experiment or deeper understanding of its results. Experimental procedures of determining protein structure struggle to catch vast number of known sequences, but the gap remains at around two orders of magnitude. The PDB database (Berman *et al.*, 2000) at the time of submitting this paper has around 29 000 deposited structures. Thus, theoretical methods are often the only way to have at least a hint how the protein may look like.

Modeling by homology has a long history (Cozzetto *et al.*, 2005) and its progress since

★50th Anniversary of the Institute of Biochemistry and Biophysics, Polish Academy of Sciences.
✪The development of the modeling server was supported by the QLG2-CT-2001-01335 EU 5FP project.
☼Present address: Max-Planck Institute for Developmental Biology, Tuebingen, Germany
✉Correspondence to: P. Zielenkiewicz, Institute of Biochemistry and Biophysics, Polish Academy of Sciences, A. Pawińskiego 5a, 02-106 Warszawa, Poland; tel.: (48 22) 658 4703; fax: (48 22) 658 4636, e-mail: piotr@ibb.waw.pl

1994 can easily be observed every two years in the CASP (http://predictioncenter.llnl.gov/Center.html) experiment (Ginalski & Rychlewski, 2003; Tramontano & Morea, 2003), conducted every two years where new methods are thoroughly tested. Another assessment of structure prediction algorithms is the continuous evaluation of fully automated methods called LiveBench (Rychlewski & Fisher, 2005). The major involvement has been in the area of preparing the alignment, because its quality critically affects the final model. Recently the accent has shifted to the modeling approaches which do not use an alignment, as sophisticated methods for similarity searching seem to have approached their limits. The latest edition of the CASP experiment seems to show that the new challenge for structure prediction is the structure refinement (Valencia, 2005).

SERVER

MOFOID was developed with two goals in mind. One was to give the user a tool for quick estimation of the complexity of the model building problem for a particular protein sequence and to return final model, in its most reliable part, even for an unexperienced user. To accomplish that we decided to use pairwise sequence alignment methods, because most of them can identify homologs down to about 30% sequence identity level. In the MOFOID local or global alignments are calculated with the Smith-Waterman (Smith & Waterman, 1981) or Needleman-Wunsch (Needleman & Wunsch, 1970) algorithms, because both methods always return the best scoring alignment, in contrast to, e.g., the BLAST algorithm (Altschul *et al.*, 1990). The second goal was to give the user full control over how the alignment is calculated and the model built. This does not only mean the control of gap penalties, but also the ability to choose amongst many substitution matrices or to control the level of model optimisation. There is also a possibility to submit user's own alignment for prospective benchmarking of the new alignment calculation methods. And, last but not least, prediction of transmembrane helices computed with the TMHMM program (Sonnhamer *et al.*, 1998) is available on request.

EXAMPLE

To show the quality of the alignment we made a model of a phosphatase from *Bacillus stearothermophilus* (PDB: 1H2F). As a template rat biphosphatase (PDB: 1C7Z) was chosen. It was not the closest hit in the "*mofoid*" search, but it had high structural similarity (the same superfamily according to SCOP and RMSD of 2.3 Å). Sequence identity was low: 31.8% and it was at the border of usability of the sequence-sequence comparison algorithms. The alignment obtained with the Smith-Waterman algorithm was modeled with **nest** with standard parameters. Then, the obtained model was compared to the original structure using the Combinatorial Extension method (Shindyalov & Bourne, 1998). The difference between the model and the original structure was very small (RMSD 2.4 Å), almost the same as the difference between the original structure and the template used.

SUMMARY

MOFOID was developed not only as a modeling server and its modular structure supports many potential applications. Currently MOFOID is used as a framework for testing new methods for the calculation of alignments. The server is publicly available at https://valis.ibb.waw.pl/mofoid/.

METHODS

The server uses **water** and **needle** programs from the EMBOSS package (Rice *et al.*, 2000) for calculating alignments with the Smith-Waterman and Needle-Wunsch algorithms, respectively. The best alignment for given parameters (gap penalties, substitution matrix) is submitted to the **nest** program of the JACKAL (http://trantor.bioc.columbia.edu/programs/jackal/index.html) package. If requested, prediction of transmembrane helices is obtained from TMHMM program, version 2.0. In the case of a user submitted alignment, any discrepancies between the user's sequence and the sequence found in the PDB file are maintained by the modeling program. The structural alignment of the example described was obtained at the CE server, http://cl.sdsc.edu/ce.html.

We would like to thank the authors for permissions to use the JACKAL and TMHMM programs, and the open source community for the EMBOSS package.

REFERENCES

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990) *J Mol Biol.*; **215**: 403–10.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. (2000) *Nucleic Acids Res.*; **28**: 235–42.

Cozzetto D, Di Matteo A, Tramontano A. (2005) *FEBS J.*; **272**: 881–2.

Ginalski K, Rychlewski L. (2003) *Proteins.*; **53** Suppl: 410–7.

Needleman SB, Wunsch CD. (1970) *J Mol Biol.*; **48**: 443–53.

Peitsch MC. (1995) *Bio/Technology.*; **13**: 658–60.

Rice P, Longden I, Bleasby A. (2000) *Trends Genet.*; **16**: 276–7.

Rychlewski L, Fischer D. (2005) *Protein Sci.*; **14**: 240–5.

Shindyalov IN, Bourne PE. (1998) *Protein Eng.*; **11**: 739–47.

Smith TF, Waterman MS. (1981) *J Mol Biol.*; **147**: 195–7.

Sonnhamer ELL, von Heijne G, Krogh A. (1998) *Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology.*, pp 175–82, Menlo Park, CA.

Tramontano A, Morea V. (2003) *Proteins.*; **53** Suppl: 352–68.

Valencia A. (2005) *Bioinformatics.*; **21**: 277.