BIOINFORMATICS APPLICATIONS NOTE Vol. 18 no. 1 2002
Pages 213–214

Geno3D: automatic comparative molecular modelling of protein

Christophe Combet, Martin Jambon, Gilbert Deléage and Christophe Geourjon*

Institut de Biologie et Chimie des Protéines, IBCP-CNRS UMR 5086, Pôle Bioinformatique Lyonnais, Bioinformatics and structural NMR, 7, Passage du Vercors, 69 367 Lyon Cedex 07, France

Received on June 26, 2001; revised on August 27, 2001; accepted on September 5, 2001

ABSTRACT

Geno3D (http://geno3d-pbil.ibcp.fr) is an automatic web server for protein molecular modelling. Starting with a query protein sequence, the server performs the homology modelling in six successive steps: (i) identify homologous proteins with known 3D structures by using PSI-BLAST; (ii) provide the user all potential templates through a very convenient user interface for target selection; (iii) perform the alignment of both query and subject sequences; (iv) extract geometrical restraints (dihedral angles and distances) for corresponding atoms between the query and the template; (v) perform the 3D construction of the protein by using a distance geometry approach and (vi) finally send the results by e-mail to the user.

Availability: Free to academic users http://geno3d-pbil.ibcp.fr

Contact: c.geourjon@ibcp.fr

The 3D structure of a protein can be experimentally determined by x-ray crystallography or by Nuclear Magnetic Resonance (NMR) spectroscopy. When no experimental structure is available, homology modelling represents a starting point for the biologist involved in structure–function relationships studies. Thus tools performing homology modelling have been developed such as Swiss-model web server (Guex *et al.*, 1999), Modeller software (Sali and Blundell, 1993), 3D-JIGSAW server (Bates and Sternberg, 1999), CPHmodels server (Lund *et al.*, 1997) and FAMS server (Ogata and Umeyama, 1997).

A web server (Geno3D) that uses distance geometry, simulated annealing and energy minimization algorithms to build the protein 3D model is available at http://geno3d-pbil.ibcp.fr. In homology modelling restraints derived from a structural 3D template are used to fold the query sequence in the distance geometry step. A good candidate as this template is a homologous protein that can be identified by a sequence similarity search. This step is ensured by PSI-BLAST (Altschul *et al.*, 1997) onto PDB_select_95 (proteins sharing less than 95% pairwise identity). A special interface (Figure 1A) has been written so as to display the potential template, the BLAST E-value, the identity level as well as a link to look at the pairwise alignment between template and query sequences. In order to choose a molecular template, a special link is also provided towards the NPS@ analysis sequence–structure system (Combet *et al.*, 2000). Once the template has been chosen, pressing the special button at the beginning of the line starts the modelling process. Spatial restraints (dihedral angles and distances) are measured onto the template by a home made program that measures all possible restraints from the pairwise alignment. These restraints are further formatted in input files that will be used by the CNS program (Brünger *et al.*, 1998) as well as molecular modelling script files generated on-the-fly. Once the modelling process is finished, files containing atomic coordinates of each model (3 is the default value) best satisfying the spatial restraints and the alignment (Figure 1B) between query and template sequences are sent to the user. Another e-mail is sent containing the superimposition (Figure 1C) of the models with the template. This superimposition highlights the poorly defined regions that correspond to gaps in the alignment (Figures 1B, C). In order to check the homogeneity of the models, a matrix of binaries root mean square deviations between coordinates of equivalent α-carbons is also sent to the user (Figure 1D).

The main advantages in using a distance geometry approach in homology modelling are no *a priori* choice in loops construction and easy identification of well defined regions.

*To whom correspondence should be addressed.

ACKNOWLEDGEMENTS

This work is supported by CNRS and University Lyon 1 ('Bonus Qualité Recherche'). C.C. is a doctoral fellowship of ANRS ('Agence Nationale de Recherches sur le Sida').

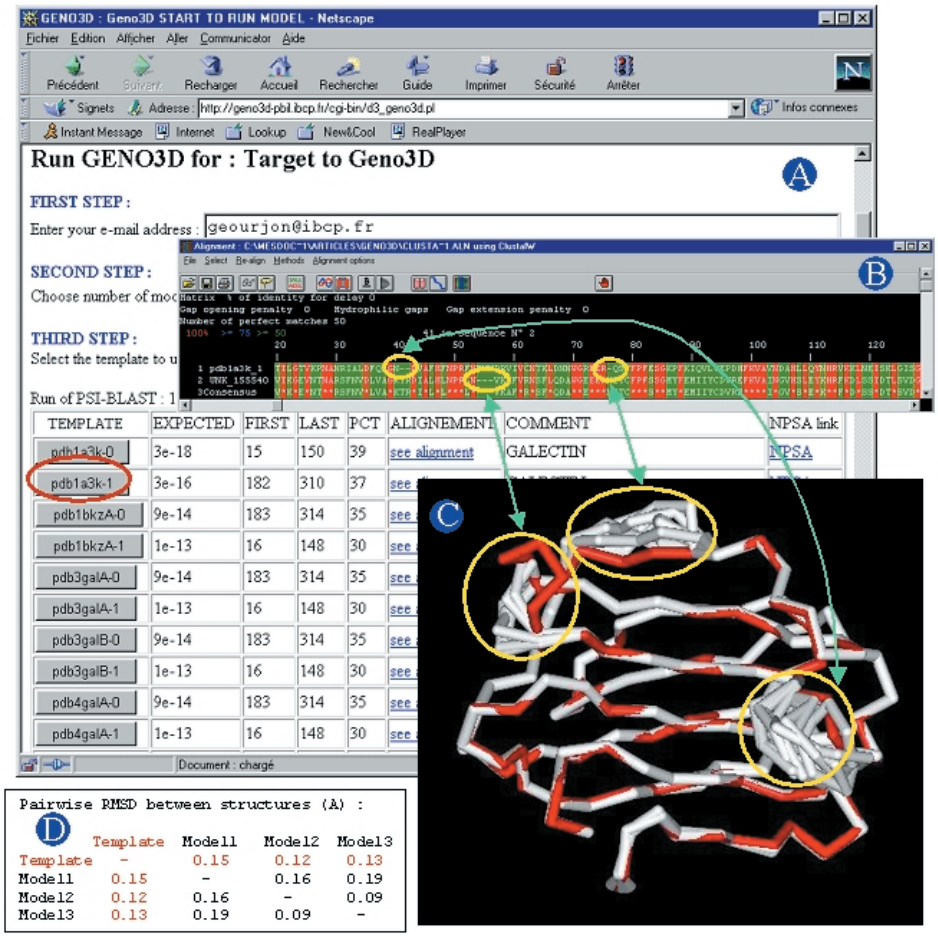

Fig. 1. General interface to Geno3D. The sequence of the galectin protein (SWISS-PROT ID: LEG8_MOUSE) has been pasted into the Geno3D query form. The similarity search was performed with PSI-BLAST onto the PDB protein sequences. The PSI-BLAST interface to template selection is shown (A). The carbohydrate recognition domain (PDB code: 1a3k) was chosen as the template for modelling (encircled button). The alignment between both sequences was loaded within the ANTHEPROT (Deléage *et al.*, 2001) package (B). The superimposition of the three models generated by Geno3D with the template (dark trace) is given in the 3D-structure panel (C). The pairwise RMSD between all structures is given as a table (D). Circles and arrows indicate the correspondence between gaps in the alignment and the variable regions in models.

REFERENCES

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

Bates,P.A. and Sternberg,M.J.E. (1999) Model building by comparison at CASP3: using expert knowledge and computer automation. *Proteins: Struct. Funct. Genet.*, **3** (Suppl.), 47–54.

Brünger,A.T., Adams,P.D., Clore,G.M., Delano,W.L., Gros,P., Grosse-Kunstleve,R.W., Jiang,J.S., Kuszewski,J., Nilges,M., Pannu,N.S., Read,R.J., Rice,L.M., Simonson,T. and Warren,G.L. (1998) Crystallography & NMR: a new software suite for macromolecular structure determination. *Acta Cryst.*, **54**, 905–921.

Combet,C., Blanchet,C., Geourjon,C. and Deléage,G. (2000) NPS@: Network Protein Sequence analysis. *Trends Biochem. Sci.*, **25**, 147–150.

Deléage,G., Combet,C., Blanchet,C. and Geourjon,C. (2001) ANTHEPROT: an integrated protein sequence analysis software with client/server capabilities. *Comput. Biol. Med.*, **31**, 259–267.

Guex,N., Diemand,A. and Peitsch,M. (1999) Protein modelling for all. *Trends Biochem. Sci.*, **24**, 364–367.

Lund,O., Frimand,K., Gorodkin,J., Bohr,H., Bohr,J., Hansen,J. and Brunak,S. (1997) Protein distance constraints predicted by neural networks and probability density functions. *Protein Eng.*, **10**, 1241–1248.

Ogata,K. and Umeyama,H. (1997) Prediction of protein side-chain conformations by principal component analysis for fixed main chain atoms. *Protein Eng.*, **10**, 353–359.

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