BIOINFORMATICS APPLICATIONS NOTE Vol. 22 no. 17 2006, pages 2164–2165
doi:10.1093/bioinformatics/btl349

Structural bioinformatics

MAGOS: multiple alignment and modelling server

N. Garnier[1,†], A. Friedrich[2,†], R. Bolze[3], E. Bettler[1,*], L. Moulinier[2], C. Geourjon[1], J. D. Thompson[2], G. Deléage[1] and O. Poch[2]

[1]Institut de Biologie et Chimie des Protéines (IBCP UMR 5086),CNRS, Univ. Lyon1;IFR128 BioSciences Lyon-Gerland;7, passage du Vercors, 69367 Lyon cedex 07, France, [2]Laboratoire de Bioinformatique et Génomique Intégratives, UMR CNRS 7104, Institut de Génétique et de Biologie Moléculaire et Cellulaire, 1 rue Laurent Fries, BP 163, 67404 ILLKIRCH Cedex, France and [3]Laboratoire d'Informatique du Parallélisme; Ecole Nationale Supérieure de Lyon; 46, allée d'Italie, 69364 Lyon cedex 07, France

Received on April 26, 2006; revised on June 9, 2006; accepted on June 22, 2006
Advance Access publication July 4, 2006
Associate Editor: Keith A Crandall

ABSTRACT

Summary: MAGOS is a web server allowing automated protein modelling coupled to the creation of a hierarchical and annotated multiple alignment of complete sequences. MAGOS is designed for an interactive approach of structural information within the framework of the evolutionary relevance of mined and predicted sequence information.

Availability: The web server is freely available at http://pig-pbil.ibcp.fr/magos

Supplementary information: The website supplies detailed explanations and illustrations of processes and results at http://pig-pbil.ibcp.fr/html/magos/help.html

Contact: e.bettler@ibcp.fr

1 INTRODUCTION

Functional protein analysis relies mainly on the use of complementary structural and evolutionary approaches. The structure-to-function relationship can be directly addressed through three-dimensional (3D) structure determination, while the sequence-to-function relationship can be understood through the analysis of conserved patterns and evolution of protein organization mainly based on amino acid sequence comparisons in the context of the multiple alignments. Homology modelling represents a powerful starting point for studies of the relationships between a sequence, its 3D structure and its function, particularly when based on multiple alignment of complete sequences (MACS), which allows the integration and visualization of essential aspects of sequence data within the context of a full-length protein family (Lecompte *et al*., 2001).

Several web-based tools are available to generate and analyse MACS or to perform homology modelling; they include MAFFT (Katoh *et al*., 2005), NPS@ (Combet *et al*., 2000), PipeAlign (Plewniak *et al*., 2003), SWISS-MODEL (Schwede *et al*., 2003), 3D-jigsaw (Bates *et al*., 2001) or Geno3D (Combet *et al*., 2002). However, until now the generation of 3D models of proteins, the creation of high-quality MACS and the mapping of molecular features on the 3D models and the MACS have not been possible at a single web-based server.

We present MAGOS, a web-based server designed to perform simultaneous evolutionary and structural studies of a protein, by allowing a simple-to-use, interconnected analysis of a MACS integrating heterogeneous (e.g. structural, functional, disease related) information and a homology 3D model.

2 PROCESS OVERVIEW

MAGOS accepts a single protein sequence in FASTA format as input and incorporates four main interconnected steps:

(1) A high-quality MACS is first computed using a modified version of the PipeAlign suite of programs, a protein family analysis tool. It includes processes ranging from the search for homologous sequences in the Uniprot (Apweiler *et al*., 2004) and PDB (Berman *et al*., 2000) databases or the UniRef90 (Apweiler *et al*., 2004) database up to the construction of a high-quality hierarchized MACS.
If not specified by the user, the PDB sequence to be used as a template for homology modelling is determined automatically. The alignment of the query and the PDB template sequence is extracted from the final multiple alignment to compute a model.

(2) The validated MACS is automatically annotated *via* MACSIMS (MACS Information Management System) (Thompson *et al*., 2006). MACSIMS integrates structural and functional information mined from external databases as well as various *ab initio* predictions. Depending on certain conservation criteria, the retrieved information can be propagated to all the sequences in the alignment, including the query sequence. At the same time, the aligned proteins are characterized according to their homology with proteins implicated in human genetic diseases.

(3) The query protein is modelled using a tuned version of Geno3D, whose main advantage is the ability to generate

*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.

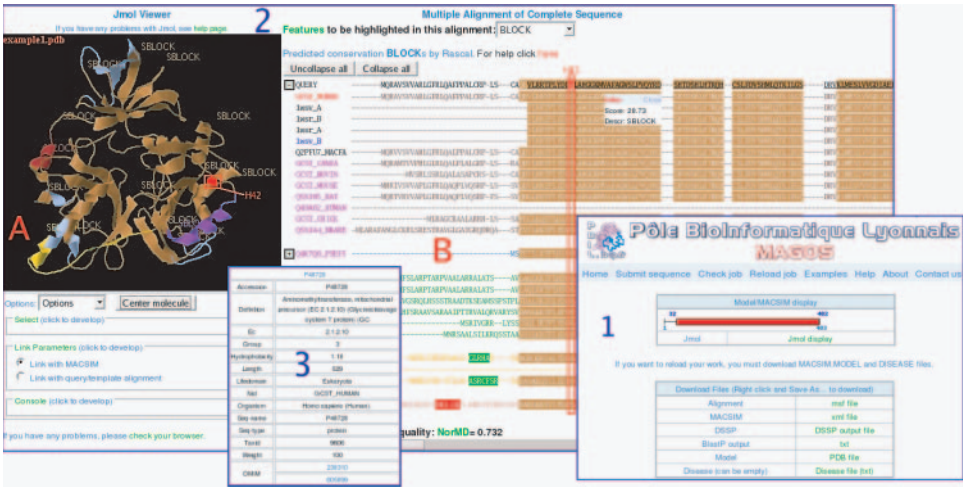

Figure 1. Screenshots of the MAGOS web server. (1) The main results page allows the download of intermediary files. The diagram represents the overlap between the modelled part (in red) and the full-length query sequence. This page also provides a link to the Jmol graphical interface (2), which is divided into two interconnected frames: the right frame (**B**) gives access to the annotated MACS interconnected with the generated homology 3D model displayed in the left frame (**A**). Several display options (e.g. zoom, background, rendering, color) are available to facilitate the interpretation of the results. All the features mined and predicted by MACSIMS can be displayed both on the query sequence in the context of the MACS and on the constructed 3D model. By default, only the first sequence of each sub-family is displayed: the sub-families can easily be expanded to show all the sequences in the clusters ('Uncollapse all' button). (3) A pop-up window accessible by clicking on the protein names contains general information about the selected protein and cross-links to external databases.

homology 3D structure models at a low rate of identity. Molecular dynamics and energy minimization steps are performed with and without geometric constraints. The model with the minimal energy and minimal RMSD is retained.

(4) The final step is the retrieval of all computed results and their interconnection through a user-friendly web interface based on the Jmol applet (http://jmol.sourceforge.net).

3 RESULTS

Figure 1 shows some screenshots of MAGOS results pages. The query sequence used in this example, the human glycine cleavage system T protein (GCST), is implicated in a genetic disease: hyperglycinemia. MAGOS data integration capabilities can be illustrated by the analysis of the known point mutations and their relative position according to structural and functional features. For example, the H42A mutation affects a residue situated in the core of a structural feature (helix) and conserved in various species (Fig. 1), even in *Escherichia coli*, suggesting an important role in catalytic functions. This mutation leads to a deficiency in GCST activity which causes a severe phenotype (Kure *et al*., 1998).

The MAGOS web-server is an extension of the Geno3D and PipeAlign servers and represents a platform for data mining and integration for novices in this field of research. Modelling from several template homologues will be added in a future release of MAGOS as well as the input of mutation information. The Jmol environment will also be upgraded in order to support new interactive functions based on residue accessibility.

ACKNOWLEDGEMENTS

The authors thank Raymond Ripp, Frédéric Plewniak and Serge Uge for stimulating discussions. This work was funded by the INSERM, the CNRS, the ULP de Strasbourg and the Décrypthon program initiated by the AFM, IBM and the CNRS. A.F. and N.G. are recipients of a fellowship from the AFM. Funding to pay the Open Access publication charges was provided by AFM.

Conflict of Interest: none declared.

REFERENCES

Apweiler,R. *et al*. (2004) UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res*., **32**, D115–D119.

Bates,P.A. *et al*. (2001) Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. *Proteins* (Suppl 5), 39–46.

Berman,H.M. *et al*. (2000) The Protein Data Bank. *Nucleic Acids Res*., **28**, 235–242.

Combet,C. *et al*. (2000) NPS@: network protein sequence analysis. *Trends Biochem. Sci*., **25**, 147–150.

Combet,C. *et al*. (2002) Geno3D: automatic comparative molecular modelling of protein. *Bioinformatics*, **18**, 213–214.

Katoh,K. *et al*. (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res*., **33**, 511–518.

Kure,S. *et al*. (1998) A missense mutation (His42Arg) in the T-protein gene from a large Israeli-Arab kindred with nonketotic hyperglycinemia. *Hum. Genet*., **102**, 430–434.

Lecompte,O. *et al*. (2001) Multiple alignment of complete sequences (MACS) in the post-genomic era. *Gene*, **270**, 17–30.

Plewniak,F. *et al*. (2003) PipeAlign: a new toolkit for protein family analysis. *Nucleic Acids Res*., **31**, 3829–3832.

Schwede,T. *et al*. (2003) SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res*., **31**, 3381–3385.

Thompson,J.D. *et al*. (2006) MACSIMS: Multiple Alignment of Complete Sequences Information Management System. *BMC Bioinformatics*, **7**, 318.