

# MODexplorer: an integrated tool for exploring protein sequence, structure and function relationships

Jan Kosinski<sup>1,†</sup>, Alessandro Barbato<sup>1,†</sup> and Anna Tramontano<sup>1,2,3,\*</sup><sup>1</sup>Department of Physics, <sup>2</sup>Center for Life Nano Science @Sapienza, Istituto Italiano di Tecnologia and <sup>3</sup>Istituto Pasteur, Fondazione Cenci Bolognietti, Sapienza University, 00185 Rome, Italy

Associate Editor: Alfonso Valencia

## ABSTRACT

**Summary:** MODexplorer is an integrated tool aimed at exploring the sequence, structural and functional diversity in protein families useful in homology modeling and in analyzing protein families in general. It takes as input either the sequence or the structure of a protein and provides alignments with its homologs along with a variety of structural and functional annotations through an interactive interface. The annotations include sequence conservation, similarity scores, ligand-, DNA- and RNA-binding sites, secondary structure, disorder, crystallographic structure resolution and quality scores of models implied by the alignments to the homologs of known structure.

MODexplorer can be used to analyze sequence and structural conservation among the structures of similar proteins, to find structures of homologs solved in different conformational state or with different ligands and to transfer functional annotations. Furthermore, if the structure of the query is not known, MODexplorer can be used to select the modeling templates taking all this information into account and to build a comparative model.

**Availability and implementation:** Freely available on the web at <http://modorama.biocomputing.it/modexplorer>. Website implemented in HTML and JavaScript with all major browsers supported.

**Contact:** [anna.tramontano@uniroma1.it](mailto:anna.tramontano@uniroma1.it)

**Supplementary information:** Supplementary data are available at *Bioinformatics* online

Received on September 19, 2012; revised on January 11, 2013; accepted on February 4, 2013

## 1 INTRODUCTION

Exploring the sequence, structure and function relationships between a protein and its homologs is a powerful strategy to transfer functional annotations such as, for example, ligand-binding sites and/or to detect the existence of alternative conformational states. Likewise, in homology modeling, a thorough survey of all homologs with known structure often can lead to building much better models than obtained by automatically selecting the closest homolog as a template.

Current tools either provide only a general overview of sequence–structure–function relationships [e.g. GeneSilico Metaserver (Kurowski and Bujnicki, 2003) or MESSA (Cong and Grishin, 2012)] or focus on selected aspects, such as

ligand-binding sites [e.g. firestar (Lopez *et al.*, 2011)]. A platform for the in-depth integrated analysis of the sequence, structure and function relationships is still missing. Consequently, the information contained in remote relationships may be missed or exploited only through tedious procedures.

Here we describe MODexplorer, a web server that integrates sequence analysis and structure comparison with functional annotations. It provides a more complete and detailed view of the sequence, structural and functional diversity within and between protein families than existing tools. If the structure of the query protein is not known, it permits to select a template for homology modeling and to automatically build the model.

## 2 TOOL DESCRIPTION

A snapshot of the user interface of MODexplorer is shown in Figure 1. A detailed description of integrated software and databases, along with the parameters and references, is included in Supplementary Data and available on the server. MODexplorer accepts as input a protein sequence, PDB code or PDB file. MODexplorer then

- Creates the multiple sequence alignment (MSA) of the query protein family using HHblits (Remmert *et al.*, 2012).
- Generates alignments to proteins of known structure using HHSearch (Söding, 2005) and the PDB database filtered at 70% sequence identity.
- For every HHSearch hit, retrieves and aligns related PDB chains, i.e. chains with a sequence similarity >70%. Such ‘redundant’ chains can be useful. For example, they might have been solved in complex with different ligands, represent alternative conformational states or have a better structural quality.
- Displays alignments to the HHSearch hits and their related PDB chains both schematically with a BLAST-like bar diagram and as multiple sequence alignments including query and hit sequences, and their homologs.
- Graphically shows annotations on the alignments. The annotations include ligand- and DNA-/RNA-binding sites, secondary structure (predicted for the query and calculated for chains of known structures), disorder (predicted for the sequence query and estimated by B-factor/missing residue annotations for the structure query and known hit structures), HHSearch similarity scores and QMEAN (Benkert *et al.*, 2011) scores of models built based on the alignments.

\*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.

