BIOINFORMATICS APPLICATIONS NOTE

Vol. 28 no. 4 2012, pages 584–586
doi:10.1093/bioinformatics/btr688

Sequence analysis

Advance Access publication December 9, 2011

JDet: interactive calculation and visualization of function-related conservation patterns in multiple sequence alignments and structures

Thilo Muth[1,†], Juan A. García-Martín[2], Antonio Rausell[3], David Juan[3], Alfonso Valencia[3] and Florencio Pazos[1,*]

[1]Computational Systems Biology Group, National Center for Biotechnology (CNB-CSIC), 28049 Madrid, [2]Department of Plant Molecular Genetics (CNB-CSIC) and [3]Structural Biology and Biocomputing Programme (CNIO), 28029 Madrid, Spain

Associate Editor: John Quackenbush

ABSTRACT

Summary: We have implemented in a single package all the features required for extracting, visualizing and manipulating fully conserved positions as well as those with a family-dependent conservation pattern in multiple sequence alignments. The program allows, among other things, to run different methods for extracting these positions, combine the results and visualize them in protein 3D structures and sequence spaces.

Availability and implementation: *JDet* is a multiplatform application written in Java. It is freely available, including the source code, at http://csbg.cnb.csic.es/JDet. The package includes two of our recently developed programs for detecting functional positions in protein alignments (*Xdet* and *S3Det*), and support for other methods can be added as plug-ins. A help file and a guided tutorial for *JDet* are also available.

Contact: pazos@cnb.csic.es

Received on September 30, 2011; revised on December 1, 2011; accepted on December 6, 2011

1 INTRODUCTION

Sequence-based protein functional analysis has become a fundamental part of a number of studies. Multiple sequence alignments (MSAs) of protein families are valuable sources of structural and functional information (Pazos and Bang, 2006). These MSAs can be obtained from the vast amounts of genomic information coming from high-throughput sequencing initiatives, and there are many resources with automatically pre-compiled alignments for thousands of families.

MSAs can be used to assess the importance of the different residues for a given protein (Zuckerkandl and Pauling, 1965). The most obvious 'important' positions are those not allowed to change (fully conserved). Conserved positions were the first indicators of functionality and consequently many methods were developed for detecting them (Valdar, 2002).

Most protein families can be subdivided into subfamilies with different functional specificities, a division which is reflected in the structure of the MSA (Rausell *et al.*, 2010). This division is related to a class of positions with a subfamily-dependent conservation pattern (i.e. the conserved amino acid is different for the different subfamilies). These positions are intuitively related to sites with some importance for defining the functional specificity of the different subfamilies within the MSA and, for that reason, they are commonly known as 'specificity-determining positions' (SDPs). A plethora of methods have been developed which, based on different concepts, try to detect these positions [see Pazos and Bang (2006) and Rausell *et al.* (2010) for an exhaustive list of the most important methods].

The presence of SDPs in protein regions related to functional and interaction specificity has been recently shown to be a widespread phenomenon (Rausell *et al.*, 2010). Moreover, the role of SDPs (automatically extracted from MSAs) in controlling the functional specificity of the family has been, in many cases, experimentally demonstrated by mutating the corresponding residues [see for example Bauer *et al.* (1999); Morillas *et al.* (2003)].

Working with SDPs usually involves running different programs for extracting them from MSAs, interactively visualizing and combining their results, and mapping them onto protein structures when available. Since no single package implemented all these capabilities, we developed JDet, a user-friendly program which contains all the features we consider practical and indispensable for the daily work with SDPs and conserved positions.

2 PROGRAM AND INTERFACE

JDet is a stand-alone multiplatform application written in Java. The basic input for JDet is a MSA. This input alignment can be filtered (e.g. to remove redundancy, fragments …) according with the parameters provided by the user or using pre-compiled sets of parameters. The user can also edit and modify the alignment by introducing/removing gaps or amino acids in one or more sequences simultaneously. Two of our recently developed methods for detecting SDPs and conserved positions are included in the package and can be run within it: Xdet (Pazos *et al.*, 2006) and S3Det (Rausell *et al.*, 2010). Additionally, JDet can read pre-generated results from these programs, and from any other program for detecting SDPs and conserved positions provided these are converted to a simple format. Moreover, filters for other methods (or even the methods themselves)

*To whom correspondence should be addressed.
†Present address: Bioprocess Engineering, Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany.

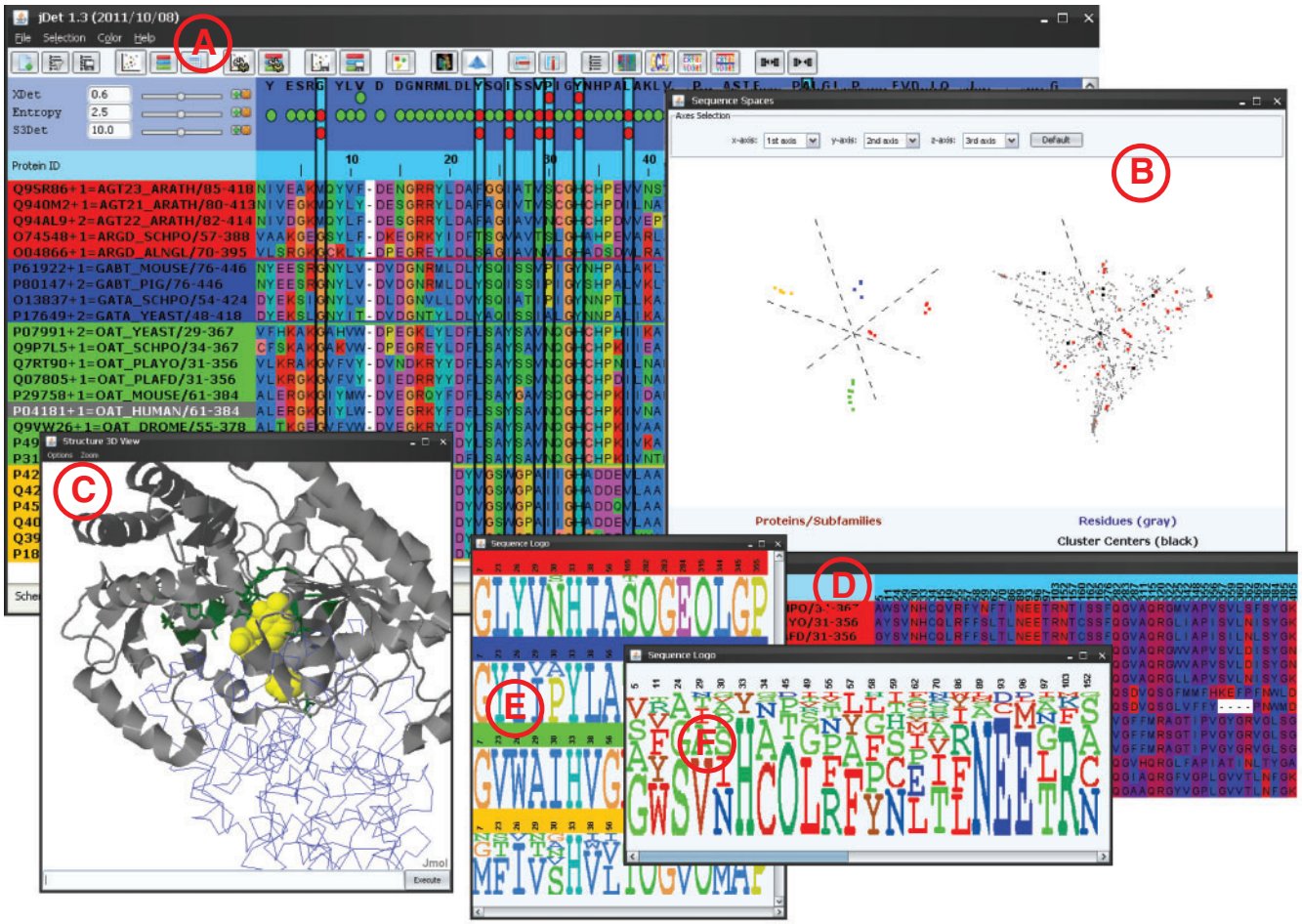

Fig. 1. Screenshots of the JDet interface. (**A**) Main window with an alignment and the results of three methods loaded into it. The subfamilies automatically detected are shown with different colors on the left. (**B**) 3D interactive views of the protein and residue spaces generated by S3Det. (**C**) Jmol window showing the 3D structure of one of the proteins of the alignment. Panels (**A**)–(**C**) are communicated with each other: proteins or residues selected in any of them become selected in the other two. (**D**) Subalignment showing only the currently selected proteins and positions, and with amino acids colored according to their polarity. (**E**) Sequence logos for the selected positions/proteins calculated for each subfamily independently. (**F**) Global sequence logos for the selected positions/proteins, and amino acids colored according with their polarity.

can be incorporated into JDet as plug-ins. With this feature, support for any method can be added to JDet by developers. As an example, JDet includes a plugin for importing the results of the widely used methods. Conseq and Consurf (Armon *et al.*, 2001), which adds to the other methods commented above that can be read by the program.

The main interface of the program (Fig. 1A) shows the MSA and the results of the different programs loaded. The program can simultaneously show the results of an unlimited number of programs/methods, which are piled up in the top bar and whose score thresholds can be changed in order to control the number of predicted positions shown. The user can freely select a set of positions (columns) in the alignment, those predicted by a given method or those concomitantly predicted by two or more of them (i.e. looking for consensus or 'stable' predictions). Proteins (rows) can be selected too, and the subalignment defined by the subsets of selected proteins and positions can be exported (Fig. 1D). Amino acids can be colored by a number of color schemas included in the program, and by any schema provided by the user.

If a protein in the alignment, or a close homolog, has a known 3D structure, that can be interactively displayed in the integrated Jmol viewer (www.jmol.org) (Fig. 1C). The program can automatically look for the most suitable structure for a given protein of the alignment, or the user can 'force' the program to use a given one. Positions selected in the alignment window are highlighted in the structure and vice versa (positions—i.e. columns—can be selected by clicking the corresponding residues in the Jmol window).

S3det detects not only SDPs but also the subfamily composition of the MSA as well (Rausell *et al.*, 2010). These subfamilies are highlighted with different colors in the alignment window (Fig. 1A). Moreover, within S3det proteins and residues are represented as points in a multidimensional space. 3D projections of that space can be interactively inspected and rotated (Fig. 1B). Selections of proteins and residues (points) in these 3D spaces are automatically translated to the alignment and 3D structure windows, and vice

versa. In this sense, the three main windows (Fig. 1A–C) are fully connected and selections in any of them are automatically reflected in the other two. In this way, it is possible for example to inspect the conservation pattern of a residue with a given structural characteristic (e.g. binding a cofactor) and its position in the sequence space, and all the other combinations.

The program incorporates other features we found very useful when working with SDPs and MSAs in general, such as the generation of sequence logos for a selected subset of proteins and positions (Fig. 1E and F).

3 CONCLUSION

Based on our own experience working with subfamilies and SDPs, we have implemented in JDet all the features we consider useful for that activity. Previously, these operations had to be done manually or by integrating diverse software. Although some of these features are individually present in existing resources oriented to SDPs [e.g. Treedet server (Carro *et al.*, 2006)] or to MSAs in general [e.g. Jalview (Waterhouse *et al.*, 2009)], no single package integrates all of them. For this reason, we think this software would be really useful for people working in extracting functional features from MSAs.

ACKNOWLEDGEMENTS

We want to thank the members of the Computational Systems Biology Group (CNB-CSIC) for comments and suggestions.

Funding: This work has been funded in part by the project BIO2010-22109 from the Spanish Ministry for Science and Innovation.

Conflict of Interest: none declared.

REFERENCES

Armon,A. *et al.* (2001) ConSurf: an algorithmic tool for the identification of functional regions in proteins by surface mapping of phylogenetic information. *J. Mol. Biol.*, **307**, 447–463.

Bauer,B. *et al.* (1999) Effector recognition by the small GTP-binding proteins Ras and Ral. *J. Biol. Chem.*, **274**, 17763–17770.

Carro,A. *et al.* (2006) TreeDet: a web server to explore sequence space. *Nucleic Acids Res.*, **34**, W110–W115.

Morillas,M. *et al.* (2003) Identification of conserved amino acid residues in rat liver carnitine palmitoyltransferase I critical for malonyl-CoA inhibition. Mutation of methionine 593 abolishes malonyl-CoA inhibition. *J. Biol. Chem.*, **278**, 9058–9063.

Pazos,F. and Bang,J.-W. (2006) Computational prediction of functionally important regions in proteins. *Curr. Bioinformatics*, **1**, 15–23.

Pazos,F. *et al.* (2006) Phylogeny-independent detection of functional residues. *Bioinformatics*, **22**, 1440–1448.

Rausell,A. *et al.* (2010) Protein interactions and ligand binding: from protein subfamilies to functional specificity. *Proc. Natl Acad. Sci. USA*, **107**, 1995–2000.

Valdar,W.S. (2002) Scoring residue conservation. *Proteins*, **48**, 227–241.

Waterhouse,A.M. *et al.* (2009) Jalview Version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, **25**, 1189–1191.

Zuckerkandl,E. and Pauling,L. (1965) Molecules as documents of evolutionary history. *J. Theor. Biol.*, **8**, 357–366.