Bioinformatics, 32(12), 2016, 1885–1887
doi: 10.1093/bioinformatics/btw082
Advance Access Publication Date: 15 February 2016
Applications Note

OXFORD

Structural bioinformatics

Multiple structure single parameter: analysis of a single protein nano environment descriptor characterizing a shared loci on structurally aligned proteins

José Augusto Salim[1,2], Luiz Borro[1,2], Ivan Mazoni[1], Inácio Yano[1], José G. Jardine[1] and Goran Neshich[1,*]

[1]EMBRAPA Agriculture Informatics, Computational Biology Research Group, Brazilian Agricultural Research Corporation, Campinas 13083-886, Brazil and [2]State University of Campinas/UNICAMP, Campinas, 13081-970, Brazil

*To whom correspondence should be addressed.

Associate Editor: Anna Tramontano

Received on June 8, 2015; revised on January 14, 2016; accepted on February 7, 2016

Abstract

Motivation: A graphical representation of physicochemical and structural descriptors attributed to amino acid residues occupying the same topological position in different, structurally aligned proteins can provide a more intuitive way to associate possible functional implications to identified variations in structural characteristics. This could be achieved by observing selected characteristics of amino acids and of their corresponding nano environments, described by the numerical value of matching descriptor. For this purpose, a web-based tool called multiple structure single parameter (MSSP) was developed and here presented.

Results: MSSP produces a two-dimensional plot of a single protein descriptor for a number of structurally aligned protein chains. From a total of 150 protein descriptors available in MSSP, selected of >1500 parameters stored in the STING database, it is possible to create easily readable and highly informative XY-plots, where X-axis contains the amino acid position in the multiple structural alignment, and Y-axis contains the descriptor's numerical values for each aligned structure. To illustrate one of possible MSSP contributions to the investigation of changes in physicochemical and structural properties of mutants, comparing them with the cognate wild-type structure, the oncogenic mutation of M918T in RET kinase is presented. The comparative analysis of wild-type and mutant structures shows great changes in their electrostatic potential. These variations are easily depicted at the MSSP-generated XY-plot.

Availability and Implementation: The web server is freely available at http://www.cbi.cnptia.embrapa.br/SMS/STINGm/MPA/index.html. Web server implemented in Perl, Java and JavaScript and JMol or Protein Viewer as structure visualizers.

Contact: goran.neshich@embrapa.br or gneshich@gmail.com

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

1 Introduction

While sequence alignments are useful to detect mutations and conserved amino acids, structural alignments are useful to investigate conformational changes due to different reasons, such as mutations, inhibition and allosteric changes. However, some functional differences between two or more similar protein structures may not be identified and/or explained by using only this kind of information, requiring instead some complementary methods.

In many studies, the overall effects on protein function, caused by mutations, are not sufficiently understood, requiring some complementary analysis of specific environments of amino acids residues by means of observing changes in their physicochemical and structural properties. Based on that, we propose a web-based tool, which can easily pinpoint changes in selected protein descriptors.

Multiple structure single parameter (MSSP) is a graphical tool for visualizing the numerical value of a single, selected, protein descriptor for a set of structurally aligned protein chains, allowing the direct comparison of changes in that specific descriptor caused by amino acid mutations, relative to the wild-type protein. However, MSSP is not limited to compare only mutants—it can compare any set of structurally similar protein chains.

2 Web server implementation

2.1 Overview

MSSP is a new module of STING server (Neshich *et al.*, 2005), which using the protein descriptors available in STING server, makes possible the graphical visualization of single protein descriptor for an arbitrarily selected number of structurally aligned protein chains.

STING server (version Blue Star) itself is a web-based platform for a comprehensive analysis of the relationship between protein sequence, structure, function and stability. The STING server calculates and maintains a set of protein descriptors for all protein structures deposited in the Protein Data Bank (PDB; Berman *et al.*, 2000). The server can also calculate descriptors for user-modeled protein structures (www.cbi.cnptia.embrapa.br/SMS/JPD/tgzHelp.htm), which may be then loaded in any of STING modules, including MSSP. Two versions, MSSP applet (using JMol) and JS MSSP (using Protein Viewer and JS JMol), are selectable.

2.2 Data input and display

MSSP accepts a list of PDB codes with their respective chain identifiers and/or the STING TGZ files (generated by STING server for user-modeled proteins). The structures could be globally aligned using one of two available structural alignment software: A multiple (protein) structural alignment algorithm (MUSTANG) (Konagurthu *et al.*, 2006), and matching Models Obtained from Theory (MAMMTH-mult), (Lupyan *et al.*, 2005), and the result can be visualized in Jmol window (Fig. 1A). MUSTANG and Mammoth-mult also provide the sequence alignment as a result of the generated structural alignment, which is loaded in the Sequence Frame (Fig. 1B). MSSP then uses the sequence alignment to guide the plot generation, where it produces a two-dimensional (2D) line plot for the

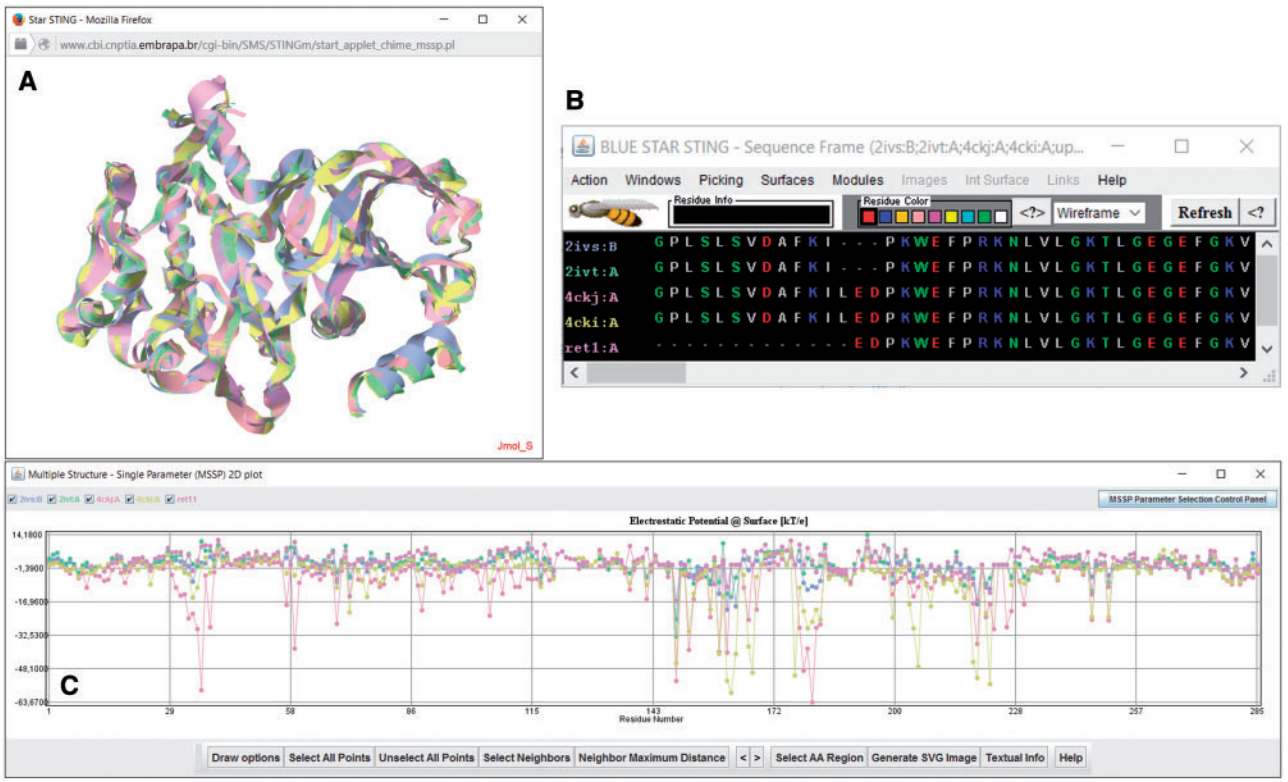

Fig. 1. MSSP plots for five chains of RET kinases (PDBid's: 2IVS:B, 2IVT:A, 4CKJ:A, 4CKI:A and modeled protein chain with M918T mutation using 2IVS:A as template). (**A**) Jmol window shows the structural alignment of selected protein chains. (**B**) Sequence frame shows the sequence alignment generated by MUSTANG, obeying hierarchically higher structural alignment. (**C**) 2D line plot for protein chains selected at the input, showing chosen protein descriptor (in this case electrostatic potential at the protein surface) along the sequences studied. The matching positions are obtained by structural alignment, which allows direct comparison of a single descriptor (Color version of this figure is available at *Bioinformatics* online.)

selected descriptor, drawing the numerical values of the descriptor on Y-axis (positioning the amino acid residue numbers in the sequence alignment on X-axis; Fig. 1C).

MSSP is freely available and does not require any installation or registration; the only requirement is the *Java Runtime Machine*, which is included in all modern web browsers.

2.3 Data export

The 2D plot can be exported as Scalable Vector Graphics, with advantages that can be scaled without loss of quality and being compatible with any modern web browser. It is also possible to export the contents (protein descriptors) as text, which then can be analyzed with any external software (Fig. 1B, Supplementary material).

Funding

The Brazilian Agricultural Research Corporation (EMBRAPA) supported this work.

Conflict of Interest: none declared.

References

Berman,H.M. *et al*. (2000) The protein data bank. *Nucleic Acids Res*., **28**, 235–242.

Konagurthu,A.S. *et al*. (2006) A multiple structural alignment algorithm. *Proteins*, **64**, 559–574.

Lupyan,D. *et al*. (2005) A new progressive-iterative algorithm for multiple structure alignment. *Bioinformatics*, **21**, 3255–3263.

Neshich,G. *et al*. (2005) The Diamond STING server. *Nucleic Acids Res*., **33**, W29–W35.