Bioinformatics, 31(14), 2015, 2394–2396
doi: 10.1093/bioinformatics/btv139
Advance Access Publication Date: 12 March 2015
Applications Note

OXFORD

Structural bioinformatics

VisualCNA: a GUI for interactive constraint network analysis and protein engineering for improving thermostability

Prakash Chandra Rathi†, Daniel Mulnaes† and Holger Gohlke*

Institute for Pharmaceutical and Medicinal Chemistry, Department of Mathematics and Natural Sciences, Heinrich-Heine-University, Universitätsstr. 1, Düsseldorf, Germany

*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.
Associate Editor: Janet Kelso

Received on March 28, 2014; revised on January 12, 2015; accepted on March 5, 2015

Abstract

Summary: Constraint network analysis (CNA) is a graph theory-based rigidity analysis approach for linking a biomolecule's structure, flexibility, (thermo)stability and function. Results from CNA are highly information-rich and require intuitive, synchronized and interactive visualization for a comprehensive analysis. We developed VisualCNA, an easy-to-use PyMOL plug-in that allows setup of CNA runs and analysis of CNA results linking plots with molecular graphics representations. From a practical viewpoint, the most striking feature of VisualCNA is that it facilitates interactive protein engineering aimed at improving thermostability.
Availability and Implementation: VisualCNA and its dependencies (CNA and FIRST software) are available free of charge under GPL and academic licenses, respectively. VisualCNA and CNA are available at http://cpclab.uni-duesseldorf.de/software; FIRST is available at http://flexweb.asu.edu.
Contact: gohlke@uni-duesseldorf.de

1 Introduction

Structural flexibility (and rigidity) is important for a protein's function and stability. Being able to accurately predict protein flexibility is thus instrumental in protein-science and -engineering as well as drug design. Our group developed the Python-based software package Constraint Network Analysis (CNA) (Pfleger *et al.*, 2013a,b) for characterizing biomolecular flexibility both at the global and residue level. CNA functions as a front- and back-end to the graph theory-based rigidity analysis software Floppy Inclusions and Rigid Substructure Topography (FIRST) (Jacobs *et al.*, 2001) and has been used to predict protein thermostability, identify structural weak spots (Radestock and Gohlke, 2008, 2011; Rathi *et al.*, 2012), and link protein flexibility and function, including allosteric regulation (C. Pfleger and H. Gohlke, unpublished results). CNA models a protein as a body-and-bar constraint network in which bodies (atoms) are connected by bars (covalent and non-covalent interactions). A rigidity analysis is then performed using the *pebble game* algorithm (Jacobs and Thorpe, 1995), resulting in a decomposition of the network into rigid and flexible regions. By removing non-covalent constraints from the network in the order of increasing strength, CNA simulates thermal unfolding. From the unfolding trajectory, CNA calculates several global and local flexibility indices (Pfleger *et al.*, 2013a,b). From the global indices, phase transition points are identified at which the network switches from rigid to flexible. These points are used to predict the thermostability of a protein and identify structural weak spots. Local indices describe intrinsic biomolecular flexibility at the level of bonds and can be compared with quantitative data from experiments on biomolecular mobility.

The output from CNA is highly information-rich. For a comprehensive analysis, the data needs to be visualized as plots (showing global and local flexibility indices) as well as 3D graphics representations of the biomolecule, the constraint network, and the decomposition into rigid and flexible regions. Furthermore, the speed of CNA allows performing real-time rigidity analyses on biomolecules.

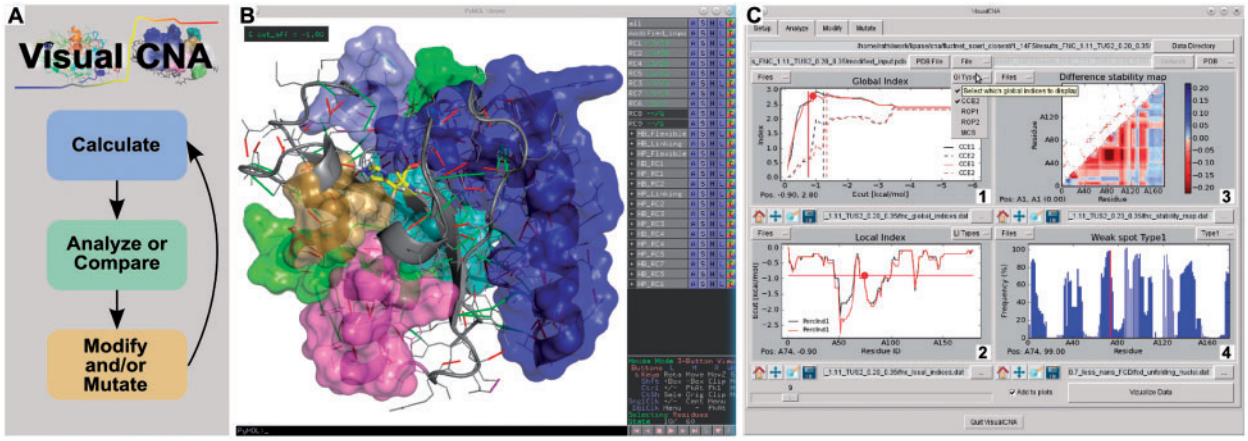

Fig. 1. (**A**) Illustration of VisualCNA's iterative work flow for optimization of protein thermostability. (**B**) PyMOL window showing the 3D protein structure at the melting point. Rigid clusters are shown as uniformly colored semi-transparent bodies. Constraints due to hydrogen bonds, salt bridges and hydrophobic contacts are shown as red, magenta and green sticks, respectively. A mutation is shown in yellow stick representation. Flexible regions are shown in grey. (**C**) The VisualCNA *Analyze* panel showing a comparison of multiple graphs from wild-type (black) and mutant (red) analyses. 1: Global indices with transition points indicated as vertical lines. 2: Local index with a red circle indicating the mutation and a horizontal red line showing the unfolding state. 3: Difference stability map between wild-type and mutant. 4: Likelihood of a residue of being a structural weak spot with the mutant shown in red

Thus, interactive structural modifications and/or editing of the constraint network followed by a re-analysis of the biomolecule's flexibility can be performed iteratively. With this in mind, we developed VisualCNA, an intuitive, easy-to-use graphical interface for CNA built as a plug-in for the molecular viewer PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0.3 Schrödinger, LLC). VisualCNA supports scientists interested in the rigidity analysis of biomolecules by a synchronized and interactive visualization of the CNA output. It also enables interactive protein engineering for improving thermostability by iteratively mutating identified weak spots, performing a subsequent rigidity analysis, and automatically comparing CNA output from wild-type and mutant structures.

2 Implementation

VisualCNA is implemented in the Python programming language as a PyMOL plug-in for Linux operating systems. It uses the external modules NumPy, SciPy, Matplotlib, Biopython, tkintertable, and Open Babel. All modules and the PyMOL source code are packaged alongside VisualCNA for easy installation. VisualCNA requires CNA and FIRST for rigidity analysis, which are distributed independently. A user manual is also distributed with VisualCNA, and tutorial videos are available online.

3 Description

The VisualCNA GUI consists of four panels: *Setup*, *Analyze*, *Modify* and *Mutate*. The *Setup* panel allows preparing variations of thermal unfolding simulations, i.e. based on a single network derived from a single input structure (Radestock and Gohlke, 2008), an ensemble of networks derived from a single structure using definitions of fuzzy non-covalent constraints (Pfleger and Gohlke, 2013) or a structural ensemble (Rathi *et al.*, 2012).

In the *Analyze* panel (Fig. 1C), CNA output from multiple thermal unfolding simulations can be shown simultaneously, which helps comparing wild-type and mutants. CNA results are shown as interactive plots of global and local flexibility indices and weak spots. In parallel, an interactive 3D protein structure (Fig. 1B) is visualized in PyMOL in terms of states corresponding to the steps of the thermal unfolding trajectory. The trajectory can be played as an interactive movie and is linked with the flexibility indices by annotations in the plots. Clicking either the plots or the structure changes the appearance of both to focus on the selected residue and/or the corresponding unfolding state. Constraints are grouped by their associated rigid cluster or flexible region to aid visualization and selection. Constraints about to break in a given state are grouped, too, facilitating the identification of residues that could be mutated to stabilize these.

The *Modify* panel contains several ways to modify the constraint network of the protein, which is useful when modeling the effect of ligands, ions or non-standard residues. An interactive table of constraints uses check boxes to enable or disable constraints, and a search box allows navigation. User defined constraints can be added by specifying atom pairs in text fields or the 3D structure.

The *Mutate* panel is central to the interactive protein engineering capability of VisualCNA. After loading an alignment of multiple sequences, the residue conservation and substitution frequencies are calculated for each residue. Clicking a bar in the conservation plot and a mutation in the substitution frequency plot then mutates the corresponding residue and updates the constraint network. The mutation is done using the PyMOL mutation tool, which allows the user to select an appropriate rotamer for the mutant. The new structure can be automatically submitted for unfolding simulation and compared with the wild-type. In this way, the effect of point mutations can be iteratively analyzed to optimize the protein structure toward increased thermostability (Fig. 1A).

In summary, the CNA approach derives maximal advantage from information on biomolecular flexibility by linking results from rigidity analyses to relevant structural characteristics. With VisualCNA, an intuitive, easy-to-use graphical interface is available that makes CNA studies amenable to non-bioinformaticians interested in rigidity analysis of biomolecules and interactive protein engineering.

Acknowledgements

This work was supported by scholarships to P.C.R. and D.M. by the Ministry of Innovation, Science and Research of North Rhine-Westphalia and Heinrich-Heine-University within the CLIB-Graduate Cluster Industrial

Biotechnology. We thank Christopher Pfleger, Daniel B. Ciupka, and Anuseema Bhadauriya for testing and helpful suggestions.

Conflict of Interest: none declared.

References

Jacobs,D.J. and Thorpe,M.F. (1995) Generic rigidity percolation: the pebble game. *Phys. Rev. Lett.*, **75**, 4051–4054.

Jacobs,D.J. *et al.* (2001) Protein flexibility predictions using graph theory. *Proteins Struct. Funct. Bioinf.*, **44**, 150–165.

Pfleger,C. and Gohke,H. (2013) Efficient and robust analysis of biomacromolecular flexibility using ensembles of network topologies based on fuzzy noncovalent constraints. *Structure*, **21**, 1725–1734.

Pfleger,C. *et al.* (2013a) Global and local indices for characterizing biomolecular flexibility and rigidity. *J. Comp. Chem.*, **34**, 220–233.

Pfleger,C. *et al.* (2013b) Constraint network analysis (CNA): a python software package for efficiently linking biomacromolecular structure, flexibility, (thermo)stability, and function. *J. Chem. Inf. Model.*, **53**, 1007–1015.

Radestock,S. and Gohlke,H. (2008) Exploiting the link between protein rigidity and thermostability for data-driven protein engineering. *Eng. Life Sci.*, **8**, 507–522.

Radestock,S. and Gohlke,H. (2011) Protein rigidity and thermophilic adaptation. *Proteins Struct. Funct. Bioinf.*, **79**, 1089–1108.

Rathi,P.C. *et al.* (2012) Thermostabilizing mutations preferentially occur at structural weak spots with a high mutation ratio. *J. Biotechnol.*, **159**, 135–144.