Bioinformatics, 31(9), 2015, 1484–1486
doi: 10.1093/bioinformatics/btu841
Advance Access Publication Date: 24 December 2014
Applications Note

OXFORD

Structural bioinformatics

RRDistMaps: a UCSF Chimera tool for viewing and comparing protein distance maps

Jonathan E. Chen[1], Conrad C. Huang[2,*] and Thomas E. Ferrin[2]

[1]Lowell High School, San Francisco, CA and [2]Resource for Biocomputing, Visualization, and Informatics, University of California, San Francisco, CA, USA

*To whom correspondence should be addressed.

Associate Editor: Burkhard Rost

Received and revised on November 7, 2014; accepted on December 16, 2014

Abstract

Motivation: Contact maps are a convenient method for the structural biologists to identify structural features through two-dimensional simplification. Binary (yes/no) contact maps with a single cutoff distance can be generalized to show continuous distance ranges. We have developed a UCSF Chimera tool, *RRDistMaps*, to compute such generalized maps in order to analyze pairwise variations in intramolecular contacts. An interactive utility, *RRDistMaps*, visualizes conformational changes, both local (e.g. binding-site residues) and global (e.g. hinge motion), between unbound and bound proteins through distance patterns. Users can target residue pairs in *RRDistMaps* for further navigation in Chimera. The interface contains the unique features of identifying long-range residue motion and aligning sequences to simultaneously compare distance maps.

Availability and implementation: *RRDistMaps* was developed as part of UCSF Chimera release 1.10, which is freely available at http://rbvi.ucsf.edu/chimera/download.html, and operates on Linux, Windows, and Mac OS.

Contact: conrad@cgl.ucsf.edu

1 Introduction

The rapid improvement of technology in recent decades has enhanced our knowledge of protein structure, as visualization software has revolutionized our understanding of molecular mechanisms. Recent applications such as *CMView* (Vehlow *et al.*, 2011) examine the nature of structural differences or changes in proteins through the use of residue–residue (RR) contact maps (Holm and Sander, 1993). RR contact maps are powerful two-dimensional (2D) representations of protein 3D structure that plot patterns of spatial interactions, e.g. pairs of amino acids with α-carbons <8 Å apart (Wu *et al.*, 2008). 2D contact maps provide a complementary view to 3D molecular visualization as they are unaffected by rotation or translation and can be easily compared and superimposed.

Contact maps are typically shown as square plots with markers denoting interactions between residues; residue pairs that fall outside of the interaction threshold are completely unmarked. However, longer-range interactions may be used to identify larger-scale motions such as hinge movements. Instead of using a single cutoff threshold for contacts, displaying a color-coded *distance* map can help users visualize longer-range interactions. As an extension to the molecular visualization application UCSF Chimera (Pettersen *et al.*, 2004), we developed *RRDistMaps*, a tool to interactively compute and display distance maps for individual proteins and to compare the distance maps of pairs of similar proteins.

2 Description

RRDistMaps is invoked from the **Tools/Structure Comparison** menu in UCSF Chimera and displays a list of the molecular chains displayed in the 3D visualization window and an area for showing a distance map. The user can select one or more chains and click the **Calculate Map** button. If a single chain is selected, the RR distances for that chain are computed and displayed as a grayscale distance map along with a color-coding legend to its right. Moving the mouse over the distance map displays the residue pair under the cursor in the status line. Selecting residue pairs by clicking the mouse or

sweeping out a rectangle automatically highlights the corresponding residues in the 3D visualization window. A green rectangle in the legend marks the distances for which the color-coding is applied; residue pairs whose RR distances fall outside of the rectangle are displayed in dark gray. When only short distances are used for color-coding, the distance maps are effectively contact maps.

For multi-protein comparison, *RRDistMaps* identifies corresponding amino acids in the different chains using sequence alignment. This limits use of proteins with sequences similar enough to be aligned correctly, and without topological rearrangements. In general, however, distance map comparisons are only appropriate for proteins with the same basic architecture. For two chains, *RRDistMaps* uses the Needleman–Wunsch algorithm (Needleman and Wunsch, 1970); for more than two chains, *RRDistMaps* uses MUSCLE (Edgar, 2004) via an online RBVI web service (Huang *et al.*, 2014). In both cases, the sequence alignment is displayed using Chimera's *Multalign Viewer* tool. Only positions in the sequence alignment where amino acids are present for all chains are used. For corresponding amino acid pairs, the average α-carbon distance and standard deviation are computed using values from all chains. The user can choose to display only the average distance or standard deviation, but the most useful display is the plot which color-codes on both attributes, with residue pairs with high standard deviation shown in color and low in grayscale. As with the single-chain interface, the user can manipulate both the distance map and the color-coding legend using the mouse.

3 Example usage

Phosphoglycerate kinase (PGK) is a key enzyme in glycolysis. The enzyme catalyzes the addition of a phosphate to ADP to make ATP. During catalysis, PGK moves from an open to a closed conformation via a hinge motion in order to bring ligands together and to shield the reaction from water molecules. In this example (Fig. 1), we use *RRDistMaps* on PGK in both closed and open forms to analyze the change in protein conformation.

RRDistMaps aligns two PGK sequences using the Needleman–Wunsch algorithm and displays the alignment in *Multalign Viewer*. A distance map is computed for each conformation using only residues for which the sequence alignment has identified a corresponding residue. These distance maps are shown in Figure 1b and are visually similar. To help identify differences, the two maps are combined to compute both the average distance and standard deviation for corresponding residue pairs. These residue pairs are color-coded such that pairs whose distances do not change much between the two conformations are colored from white (close together) to gray (far apart) while pairs whose distances vary greatly between the two structures are colored from red (short average distance) to blue (long average distance). Thus, residues that stay fixed relative to each other appear as grayscale while those that move are colored.

Figure 1a shows three windows: the 3D view, the *RRDistMaps* dialog and *Multalign Viewer*. The *RRDistMaps* dialog shows several structural features in the color-coded combined distance map. The white diagonal represents the zero distance from a residue to itself. The slightly darkened white diagonals perpendicular to the main diagonal are antiparallel beta strands; white diagonals parallel to the main diagonal are parallel beta strands. These can be seen in the single-protein distance and contact maps as well. The red and blue elements mark residue pairs that have the highest variations in distance in the two conformations. Sweeping out regions in the distance map automatically highlights the selected residues in both the sequence alignment and the 3D graphical view. In the case of PGK, examining the large, boldly colored regions reveals that these pairs correspond to different segments of the protein undergoing hinge motion. Our example was selected because the hinge motion can easily be identified visually using distance maps, but in general more complex algorithms may be required to identify subtle movements.

4 Significance and conclusion

Contact maps are frequently used for analyzing and comparing protein structures. *RRDistMaps* couples contact maps with the 3D visualization and modeling capabilities of UCSF Chimera. In addition, using color-coded *distance* maps rather than contact maps enables users to explore additional information using all pairwise inter-residue distances

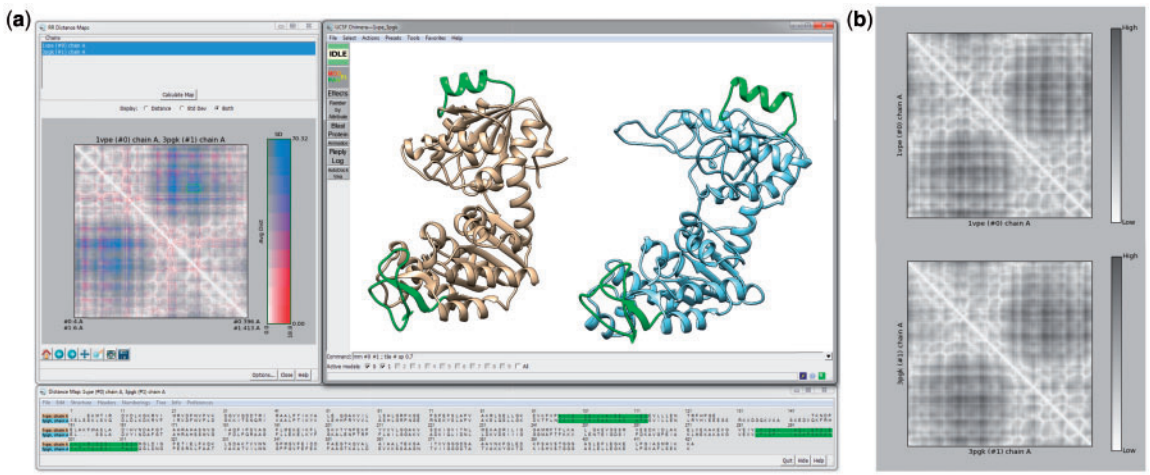

Fig. 1. Using *RRDistMaps* to analyze PGK, in closed form and open form bound to ADP. (**a**) A UCSF Chimera 3D view (top right) displaying the closed conformation (PDB:1VPE) and the open (PDB:3PGK) conformation. The sequence alignment between the proteins is shown in the *Multalign Viewer* window (bottom). The average distance and standard deviation for all amino acid pairs are shown in the *RRDistMaps* dialog (top left); color coding using both distance and standard deviation data helps to highlight residue pairs of interest. Amino acids highlighted show hinge movement between the two conformations. In contrast, (**b**) shows that the individual distance maps for both the closed form (top) and open form (bottom) are virtually indistinguishable visually when compared side-by-side

instead of just a subset. Automatic highlighting of amino acids in the 3D and sequence views for a distance map region also aids users in integrating multiple information sources. The addition of *RRDistMaps* extends the structure analysis and comparison capabilities in UCSF Chimera. Planned improvements include allowing users to set alignment parameters such as gap penalties and offering Chimera's *MatchMaker* tool as an alternative alignment method that incorporates secondary structure information.

Acknowledgement

We thank Elaine Meng for careful proofreading of the manuscript.

Funding

This work was supported by National Institutes of Health [grant NIGMS P41-GM103311] and the UCSF-Lowell Summer Internship Program.

Conflict of Interest: none declared.

References

Edgar,R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792–1797.

Holm,L. and Sander,C. (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.*, **233**, 123–138.

Huang,C.C. *et al.* (2014) Enhancing UCSF Chimera through web services. *Nucleic Acids Res.*, **42**, 478–484.

Needleman,S.B. and Wunsch,C.D. (1970) A general method applicable to the search for similarities in the ammo acid sequence of two proteins. *J. Mol. Biol.*, **48**, 443.

Pettersen,E.F. *et al.* (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.

Vehlow,C. *et al.* (2011). CMView: interactive contact map visualization and analysis. *Bioinformatics,* **27**, 1573–1574.

Wu,S. *et al.* (2008) A comprehensive assessment of sequence-based and template-based methods for protein contact prediction. *Bioinformatics,* **24**, 924–931.