

Structural bioinformatics

Con-Struct Map: a comparative contact map analysis toolJo-Lan Chung^{1,2,3}, John E. Beaver³, Eric D. Scheeff^{3,†} and Philip E. Bourne^{2,3,*}¹Department of Chemistry and Biochemistry, ²Skaggs School of Pharmacy and Pharmaceutical Sciences and³San Diego Supercomputer Center, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA

Received on May 11, 2007; revised on July 1, 2007; accepted on July 3, 2007

Associate Editor: Alfonso Valencia

ABSTRACT

Summary: Con-Struct Map is a graphical tool for the comparative study of protein structures. The tool detects potential conserved residue contacts shared by multiple protein structures by superimposing their contact maps according to a multiple structure alignment. In general, Con-Struct Map allows the study of structural changes resulting from, e.g. sequence substitutions, or alternatively, the study of conserved components of a structure framework across structurally aligned proteins. Specific applications include the study of sequence-structure relationship in distantly related proteins and the comparisons of wild type and mutant proteins.

Availability: http://pdbrs3.sdsc.edu/ConStructMap/viewer_argument_generator/singleArguments

Contact: bourne@sdsc.edu

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

1 INTRODUCTION

An increasing number of protein structures are being determined by structural genomics (Burley *et al.*, 1999) and traditional structural biology. Structure alignments allow direct comparison of protein structures in 3D space and provide a way to extract comparative information embedded in these structures. For example, structure alignments may provide some functional information for a protein with unknown function by aligning this protein against another protein with similar fold and known function. In addition, structure alignment may reveal the relationship of distantly related proteins, which is not available from the analysis of protein sequences. These alignments provide remote sequence relationship used in profiles and hidden Markov models (Kelley *et al.*, 2000; Scheeff and Bourne, 2006).

We have developed a graphical tool, Con-Struct Map, for the comparative study and visualization of multiple protein structures. Con-Struct Map detects residue contacts shared (or not shared) by multiple structures in an alignment. By exploring these conserved contacts, important features of a shared structural framework are revealed as details between sequence substitutions and structural conservation.

*To whom correspondence should be addressed.

†Present address: Razavi-Newman Center for Bioinformatics, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037, USA.

2 THE CON-STRUCT MAP

The software first generates the contact map of each of the structures and then superimposes these contact maps according to their multiple structure alignment. A protein contact map is a 2D matrix representing the distance between any two residues in a protein tertiary structure. Two residues are said to form a potential contact if the distance between them, as defined by C- α and C- β of the respective residues, is within a certain distance cutoff. Through the direct comparison of protein contact maps, Con-Struct Map is able to extract information of structural conservation from aligned protein structures.

The structure alignment software used by Con-Struct Map is the CE-MC program (Guda *et al.*, 2004). CE-MC automatically aligns multiple protein structures using the Combinatorial Extension algorithm (Shindyalov and Bourne, 1998) and Monte Carlo optimization (Guda *et al.*, 2001). Con-Struct Map also allows users to provide their own alignments generated from other software or databases. The alignment file has to be converted to Joy input format (Mizuguchi *et al.*, 1998a). For example, one can use manually curated multiple structure alignments [e.g. HOMSTRAD (Mizuguchi *et al.*, 1998b)] or use sequence-independent multiple structure alignments [e.g. MUSTA (Leibowitz *et al.*, 2001)] for disconnected protein chains. Pair-wise structure alignments can be input to software if users only want to study two structures.

Supplementary Figure 1 illustrates the Con-Struct Map output for nine non-redundant serine–threonine protein kinases with sequence identity <40%. The left panel presents the superimposed contact maps of these structures. The x and y bars show the sequence number and secondary structures (red: helix, blue: strand, gray: coils and others) for the user selected reference structure [in this case, protein kinase A (PKA), PDB id: 1CDK:A]. Within the contact map, a colored cell located at position (i, j) represents the degree of conserved contact between residues at position i and j . Different colors represent the degree of conservation. For example, the red cells represent the most conserved contacts shared by all of these protein kinases. The pink cells represent the contacts shared by 8 out of the 9 protein kinases, etc.

The Con-Struct Map Web interface allows the user to select the criteria that define a contact. That is, the type of atoms making the contact (C- α , C- β) and the distance cutoff between these atoms. In Supplementary Figure 1, two residues are said to form a potential contact if the distance between their C- β atoms is 8 Å or less. When the user clicks a contact shown on the left panel,

all of the corresponding residues across the aligned structures will be shown in the middle of the right panel. The upper part of the right panel gives a close-up view of this contact and other neighborhood contacts. Con-Struct Map also allows the user to further visualize a specific contact on any of these aligned structures. The upper part of the right panel can be switched to display the 3D structure by selecting the 3D viewer option at the top of this panel. The user can click a cell and choose one of the aligned protein structures as a template. The 3D Viewer will map this contact on the selected template structure.

The Con-Struct Map results allow rapid discovery of essential interactions in a protein family (present in all structures). It also allows the discovery of differences between structural subsets (these will be contacts present in some structures, but not all). In the example of the protein kinases, Con-Struct Map detects a variety of key shared interactions that have been previously described in both the protein kinase family and its distant relatives (Kannan and Neuwald, 2005; Kornev *et al.*, 2006; Scheeff and Bourne, 2005). Here, we use the example of the interactions of PHE 185 of protein kinase A (Bossemeyer *et al.*, 1993), a residue essential for regulation of activity in the family, which is part of the distinctive DFG motif. Con-Struct Map detects two universal (and non-chain contiguous) interactions of PHE 185 within the structure set, to LEU 95 and TYR 164 (Supplementary Fig. 1). Both are part of a proposed hydrophobic 'spine' essential for stabilizing protein kinases (Kornev *et al.*, 2006). Con-Struct Map also detects several interactions of PHE 185 conserved in only some of the structures, one of which is with GLU 91. GLU 91 forms an ion pair with LYS 72 in the active site, stabilizing it for interaction with ATP. The right panel of the Con-Struct Map output lists all the residues corresponding to the contact between PHE 185 and GLU 91 of protein kinase A (PKA, PDB code: 1CDK). Two kinases lack this contact, and these structures have a missing @ symbol in the list (Supplementary Fig. 1). The interacting residues in PKA may be compared to those in cyclin dependent kinase 2 (CDK2, PDB ID: 1B38:A) (Brown *et al.*, 1999), which lacks this contact, by selecting these structures as the template and clicking 3D Viewer. Comparison of the output reveals that in the CDK2 structure, the residue corresponding to GLU 91 has been rotated out of the active site. This conformational shift is one way in which protein kinases may be switched to an inactive state (Huse and Kuriyan, 2002). Thus, Con-Struct Map rapidly enables the detection of key structural shifts relevant to protein function.

Con-Struct Map has been used to study the structural impact of point-mutations by comparing the sequence and structure of wild type and mutant proteins (Krebs and Bourne, 2004). It also has the potential to be used in the study of protein-protein or protein-ligand interactions if the user provides the structure alignment of protein complexes.

Con-Struct Map was written in Java using the Molecular Biology Toolkit (MBT) (Moreland *et al.*, 2005). MBT provides a uniform data model for the description of biological structures and thus enables fast development of applications for protein structure analysis and visualization. MBT has been used for the development of applications including the Protein Kinase Resource Viewer (Smith *et al.*, 1997), ProteinWorkshop [www.pdb.org/robohelp/viewer/proteinworkshop.htm], Ligand Explorer [ligpro.sdsc.edu] and EpitopeViewer (Beaver *et al.*, 2007).

Con-Struct Map was implemented as a Java enabled web browser application. It is launched with Java Web Start and requires that Java Runtime Environment 1.4.2 or higher be installed. All other software components are automatically downloaded, installed and kept up to date by Java Web Start.

ACKNOWLEDGEMENTS

This work was supported by NIH grants GM63208 and GM08326. We thank Dr Chittibabu Guda for his help on using the CE-MC program. We also like to thank Dr Song Yang and Dr Werner G. Krebs for their help with software development. Funding to pay Open Access charges was provided by NIH Grant GM63208.

Conflict of Interest: none declared.

REFERENCES

- Beaver, J.E. *et al.* (2007) EpitopeViewer: a Java application for the visualization and analysis of immune epitopes in the Immune Epitope Database and Analysis Resource (IEDB). *Immunome Res.*, **3**, 3.
- Bossemeyer, D. *et al.* (1993) Phosphotransferase and substrate binding mechanism of the cAMP-dependent protein kinase catalytic subunit from porcine heart as deduced from the 2.0 Å structure of the complex with Mn²⁺ adenylyl imidodiphosphate and inhibitor peptide PKI(5-24). *EMBO J.*, **12**, 849–859.
- Brown, N.R. *et al.* (1999) Effects of phosphorylation of threonine 160 on cyclin-dependent kinase 2 structure and activity. *J. Biol. Chem.*, **274**, 8746–8756.
- Burley, S.K. *et al.* (1999) Structural genomics: beyond the human genome project. *Nat. Genet.*, **23**, 151–157.
- Guda, C. *et al.* (2001) A new algorithm for the alignment of multiple protein structures using Monte Carlo optimization. *Pac. Symp. Biocomput.*, **6**, 275–286.
- Guda, C. *et al.* (2004) CE-MC: a multiple protein structure alignment server. *Nucleic Acids Res.*, **32**, W100–W103.
- Huse, M. and Kuriyan, J. (2002) The conformational plasticity of protein kinases. *Cell*, **109**, 275–282.
- Kannan, N. and Neuwald, A.F. (2005) Did protein kinase regulatory mechanisms evolve through elaboration of a simple structural component? *J. Mol. Biol.*, **351**, 956–972.
- Kelley, L.A. *et al.* (2000) Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J. Mol. Biol.*, **299**, 499–520.
- Kornev, A.P. *et al.* (2006) Surface comparison of active and inactive protein kinases identifies a conserved activation mechanism. *Proc. Natl Acad. Sci. USA*, **103**, 17783–17788.
- Krebs, W.G. and Bourne, P.E. (2004) Statistical and visual morph movie analysis of crystallographic mutant selection bias in protein mutation resource data. *J. Bioinform. Comput. Biol.*, **2**, 61–75.
- Leibowitz, N. *et al.* (2001) MUSTA – a general, efficient, automated method for multiple structure alignment and detection of common motifs: application to proteins. *J. Comput. Biol.*, **8**, 93–121.
- Mizuguchi, K. *et al.* (1998a) JOY: protein sequence-structure representation and analysis. *Bioinformatics*, **14**, 617–623.
- Mizuguchi, K. *et al.* (1998b) HOMSTRAD: a database of protein structure alignments for homologous families. *Protein Sci.*, **7**, 2469–2471.
- Moreland, J.L. *et al.* (2005) The Molecular Biology Toolkit (MBT): a modular platform for developing molecular visualization applications. *BMC Bioinformatics*, **6**, 21.
- Scheeff, E.D. and Bourne, P.E. (2005) Structural evolution of the protein kinase-like superfamily. *PLoS Comput. Biol.*, **1**, e49.
- Scheeff, E.D. and Bourne, P.E. (2006) Application of protein structure alignments to iterated hidden Markov model protocols for structure prediction. *BMC Bioinformatics*, **7**, 410.
- Shindyalov, I.N. and Bourne, P.E. (1998) Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.*, **11**, 739–747.
- Smith, C.M. *et al.* (1997) The protein kinase resource. *Trends Biochem. Sci.*, **22**, 444–446.