BIOINFORMATICS

Vol. 00 no. 00
Pages 1–2

CMView: Interactive contact map visualization and analysis

Corinna Vehlow [1], Henning Stehr [2], Matthias Winkelmann [2], Jose M. Duarte [3], Lars Petzold [2], Juliane Dinse [1] and Michael Lappe [2]

[1]Department of Simulations and Graphics, O.v.G. University Magdeburg, Germany
[2]Max Planck Insitute for Molecular Genetics, Ihnestr. 63-73, 14195 Berlin,Germany
[3]Laboratory of Biomolecular Research, Paul Scherer Institut, 5232 Villigen PSI, Switzerland

ABSTRACT

Summary: Contact maps are a valuable visualization tool in structural biology. They are a convenient way to display proteins in two dimensions and to quickly identify structural features such as domain architecture, secondary structure and contact clusters. We developed a tool called CMView which integrates rich contact map analysis with 3D visualization using PyMol. Our tool provides functions for contact map calculation from structure, basic editing, visualization in contact map and 3D space and structural comparison with different built-in alignment methods. A unique feature is the interactive refinement of structural alignments based on user selected subsets of contacts.
Availability: CMView is available for Linux, Windows and MacOS and can be downloaded from http://www.bioinformatics.org/cmview/. The source code is licensed under the GNU General Public License.
Contact: lappe@molgen.mpg.de, stehr@molgen.mpg.de

1 INTRODUCTION

The tertiary structure of a protein is determined by non-covalent residue interactions. An all-atom distance map is a lossless representation of the 3D coordinates (save chirality). Distance data can be further reduced to a binary residue contact map while still allowing complete reconstructions within 2Å RMSD (Duarte *et al.*, 2010). The native fold is retained in reconstructions using sparse subsets comprising as low as 10% of native contacts (Sathyapriya *et al.*, 2009). Contact maps are a convenient way to highlight structural features like domain architecture, secondary structure and contact clusters and they display unique information about the sequence seperation of contacting residues which is not easily visible in 3D representations. The study of contact maps has been a valuable source of insight in experimental and computational protein structure analysis. They have for example been used to measure the dissimilarity of structures (**?**), to analyze protein-protein interaction patterns (de Melo *et al.*, 2007) and to study protein folding (Vendruscolo and Domany, 2000). Here we present a tool which combines the strengths of contact map and 3D visualization for protein analysis.

Some other tools that provide some of the features of our application have previously been developed. *Structer+Dotter* (Sonnhammer and Wooton, 1998) contains modules for the generation (*Structer*) and simple visualization (*Dotter*) of distance maps and contact maps from PDB files. *SeqX* (Biro and Fordos, 2005) integrates frequency counts of residue combinations. With *Protmap2D* (Pietal *et al.*, 2007) and *Con-StructMap* (Chung *et al.*, 2007) contact maps of two conformations can be compared side-by-side. *Con-Struct Map* also allows to compare non-sequence-identical proteins by loading an alignment from a file.

2 CMVIEW FEATURES

CMView is a stand-alone Java application for interactive visualization, analysis and manipulation of protein contact maps. It integrates protein analysis in contact map and 3D space via an interface to the molecular viewer PyMol (DeLano, 2002). CMView is open source software licensed under the GNU General Public License (GPL). It is available for Mac OS X, Linux, Windows and other platforms supporting Java 6.

Contact maps can be imported from text files (Casp RR, or native CSV format) or calculated from 3D coordinates in PDB and Casp TS format or directly loaded from the PDB website. The contact definition can be specified in terms of contact type (all-atom, C-alpha, C-beta) and contact threshold (distance cutoff in Å). The main application window (Figure 1, right) shows the contact map and the various menu options for editing and analysis. If 3D coordinates are available, the 3D structure will be shown in a separate PyMol window (Figure 1, left). The full set of advanced visualization features of PyMol is available. The two views are connected such that contacts selected by the user in the contact map window can be shown in the 3D window. The distance map, contact density and triangle inequality relations (here called common neighborhoods) are available as colored overlays in the contact map window.

2.1 Selecting, editing and export

CMView offers various functions to manipulate the contact map. Contacts can be selected individually or by using rectangle select, fill select, diagonal select and neighbor select tools that resemble similar functions known from graphics applications. Selected contacts can be deleted, highlighted in different colors and shown in the 3D structure. Individual contacts can also be added to the contact map.

As an additional feature, *CMView* implements the Cone-Peeling Algorithm (Sathyapriya *et al.*, 2009), which tries to identify a minimal subset of contacts, which is sufficient to maintain the native fold when reconstructing the 3D structure from contacts using distance geometry. Contact maps can be saved as text files in Casp RR or

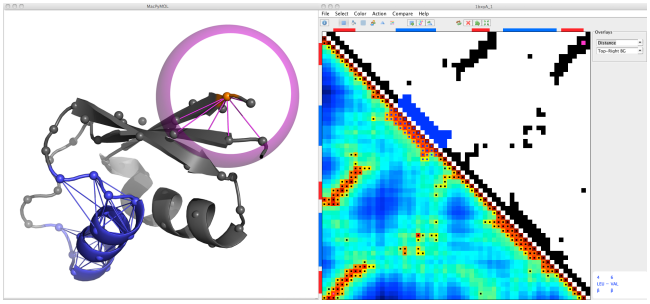

Fig. 1. Screenshot of *CMView*: 3D structure (left), contact map (right upper) and distance map (right lower) of Ribosomal Protein L30 (PDB code 1bxy). Contacts within the n-terminal alpha-helix are shown in blue. Contacts between the two termini are shown in purple.

native CSV format. The scene, including highlighted contacts can also be exported as a PNG image.

2.2 Secondary Structure Analysis

CMView can assign secondary structure based on the annotation in PDB files or by running DSSP (Kabsch and Sander, 1983). To analyze the specific interactions within or between secondary structure elements, contacts can be selected based on individual secondary structure elements or by secondary structure type (e.g. all helix-helix contacts).

2.3 Pairwise Comparison

A key feature of *CMView* is the pairwise structural comparison of two proteins or conformations. For this purpose, a second structure can be loaded and aligned via one of the following methods: Needleman-Wunsch sequence alignment (Moustafa, 2007), SADP contact-based structural alignment (Jain and Lappe, 2007) or Dali structural alignment (Holm and Sander, 1995). The comparison view allows quick identification of shared and unique contacts. Common contacts are shown in black and contacts which are unique to one structure are shown in pink (for the first structure, e.g. a predicted structure) and green (for the second structure, e.g. the native structure). The two structures are also superimposed in the PyMol window, doing a best fit on the residues which are in contact in both structures. The 3D alignment can be interactively refined by selecting a subset of contacts and recalculating the superposition based on the subset of selected contacts. This feature allows the comparision of different alignments based on shared substructures in cases where a global rigid-body alignment is not optimal. To our knowledge, CMView is the only application that allows such an alignment of substructures in an interactive fashion.

3 CONCLUSION

CMView combines the strengths of rich contact map analysis with traditional 3D visualization in a single application. As a tool for contact map generation, modification and analysis it is the most feature complete application to date. Special emphasis has been put on integrating tools for the analysis of secondary structure interaction patterns and for the pairwise comparison of structural models or related proteins. In particular, structural alignments can be interactively refined based on different subsets of shared contacts. These unique features make CMView a valuable tool for structural analysis, protein modeling, assessment of structure predictions and education in structural biology.

4 FUTURE DIRECTIONS

Future work will include integration of a reconstruction engine to transform modified contact maps back into 3D structures to give direct visual feedback how added or deleted contacts affect the structure. This will allow CMView to be used as an interactive, contact-based protein modeling tool. Preview snapshots of future versions will be available from the authors upon request.

ACKNOWLEDGEMENT

We thank Brijnesh J. Jain, Dan M. Bolser and Ioannis Fillipis for helpful comments and suggestions. Funding for this project was provided by the Max-Planck-Society.

REFERENCES

Biro, J. C. and Fordos, G. (2005). Seqx: a tool to detect, analyze and visualize residue co-locations in protein and nucleic acid structures. *Bmc Bioinformatics*, **6**, 170.

Chung, J. L., Beaver, J. E., Scheeff, E. D., and Bourne, P. E. (2007). Con-struct map: a comparative contact map analysis tool. *Bioinformatics*, **23**(18), 2491–2492.

de Melo, R., Ribeiro, C., Murray, C., Veloso, C., da Silveira, C., Neshich, G., Meira, W. J., Carceroni, R., and Santoro, M. (2007). Finding protein-protein interaction patterns by contact map matching. *Genetics and Molecular Research*, **6**(4).

DeLano, W. (2002). The PyMOL molecular graphics system - http://www.pymol.org.

Duarte, J. M., Sathyapriya, R., Stehr, H., Filippis, I., and Lappe, M. (2010). Optimal contact definition for reconstruction of contact maps. *BMC Bioinformatics*, **11**, 283–283.

Holm, L. and Sander, C. (1995). Dali: a network tool for protein structure comparison. *Trends in biochemical sciences*, **20**(11), 478–480.

Jain, B. J. and Lappe, M. (2007). *Joining Softassign and Dynamic Programming for the Contact Map Overlap Problem*, volume 4414 of *Lecture Notes in Computer Science*, pages 410–423. Springer Berlin / Heidelberg.

Kabsch, W. and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**(12), 2577–2637.

Moustafa, A. (2007). Jaligner: Open source java implementation of smith-waterman - http://jaligner.sourceforge.net.

Pietal, M. J., Tuszynska, I., and Bujnicki, J. M. (2007). Protmap2d: visualization, comparison and analysis of 2d maps of protein structure. *Bioinformatics*, **23**(11), 1429–1430.

Sathyapriya, R., Duarte, J., Stehr, H., Filippis, I., and Lappe, M. (2009). Defining an essence of structure determining residue contacts in proteins. *PLoS Comput Biol*, **5**(12), e1000584.

Sonnhammer, E. L. L. and Wooton, J. C. (1998). Dynamic contact maps of protein structures. *Journal of Molecular Graphics and Modelling*, **16**(1), 1–5.

Vendruscolo, M. and Domany, E. (2000). Protein folding using contact maps. *Vitam Horm*, **58**, 171–212.