

Systems biology

structureViz: linking Cytoscape and UCSF Chimera

John H. Morris^{1,*}, Conrad C. Huang¹, Patricia C. Babbitt^{2,1} and Thomas E. Ferrin^{1,2}

¹Department of Pharmaceutical Chemistry and ²Department of Biopharmaceutical Sciences, University of California, San Francisco, USA

Received on February 28, 2007; revised on May 29, 2007; accepted on June 16, 2007

Advance Access publication July 10, 2007

Associate Editor: Thomas Lengauer

ABSTRACT

Summary: structureViz is a Cytoscape plug-in that links the visualization of biological networks provided by Cytoscape with the visualization and analysis of macromolecular structures and sequences provided by UCSF Chimera. When combined with Cytoscape and Chimera, structureViz provides the first tool that links these two critical aspects of computational analysis in a straightforward manner. structureViz includes commands to open structures in Chimera and align them using Chimera's sequence-structure analysis tools. When a structure is opened, structureViz provides an alternative interface to Chimera: the Cytoscape Molecular Structure Navigator. This interface uses a tree-based paradigm to allow users to select and affect the display of models, chains and residues, mostly through the use of context menus.

Contact: scooter@cgl.ucsf.edu

interact. These two approaches increasingly overlap. Systems approaches now have the resolution to pose hypotheses about the interactions of individual proteins or the roles of specific metabolites in a pathway. Similarly, molecular and structural biologists are increasingly investigating the impact of regulatory pathways on transcription and how large complexes of proteins work together to perform biological function. As systems biology increases in resolution and molecular and structural biology increase in scope, we believe that there is a critical need to integrate tools commonly used within each discipline.

structureViz is an analysis tool to provide a bridge between the broad context of the systems biologist and the detailed, molecular understanding of the molecular or structural biologist. *structureViz* is a plug-in to Cytoscape (Shannon *et al.*, 2003), an open source network visualization application used to explore biological networks of various kinds and, more recently, protein similarity networks.

1 INTRODUCTION

Over the last decade, new biological data and discovery have extended our ability to interrogate and view biological processes at many levels. In particular, techniques and tools have been developed to look at biological function from a systems' perspective rather than a molecular perspective.

During this same period, structural biology has made significant advances in the number, size and complexity of the molecular structures that can be determined or modeled. Various structural genomics projects such as the Protein Structure Initiative (<http://www.nigms.nih.gov/Initiatives/PSI.htm>), improved experimental techniques, such as cryo-electron microscopy, and comparative model structure repositories such as MODBASE (Pieper *et al.*, 2006) promise to continue this growth in the availability of structural information, especially for large molecular complexes.

While systems biologists often address research questions from a top-down perspective aimed at explicitly modeling the complexity inherent in biological organisms, molecular and structural biologists contribute to systems biology via a bottom-up approach focused on identifying the 'parts list' of an organism as well as determining how the individual parts

2 structureViz DESCRIPTION

structureViz allows Cytoscape users to select nodes representing proteins in their biological networks and interactively display and analyze the 3D structures associated with those proteins in UCSF Chimera (Pettersen *et al.*, 2004). Users can associate structures from the PDB (Berman *et al.*, 2000) with their sequences by NCBI GI number using a Cytoscape node attribute file provided on the Resource for Biocomputing, Visualization and Informatics (RBVI) website (<http://www.rbvi.ucsf.edu/Research/cytoscape/structureViz/>). Alternatively, users can provide their own data by annotating their nodes with any one of the following attributes: *pdb*, *pdbFileName* or *structure*. The values of these attributes should be comma-separated lists of PDB identifiers. Once this association is made, *structureViz* can be used to open the structure(s) in Chimera, perform structural analyses (e.g. compare structures), and if desired, augment the Cytoscape network with the results of the comparisons.

3 EXAMPLE USAGE

Figure 1 shows a screenshot of a small portion of the Cytoscape network for the amidohydrolase superfamily from the Structure-Function Linkage Database (Pegg *et al.*, 2006). Nodes in this network represent proteins and the edges

*To whom correspondence should be addressed.

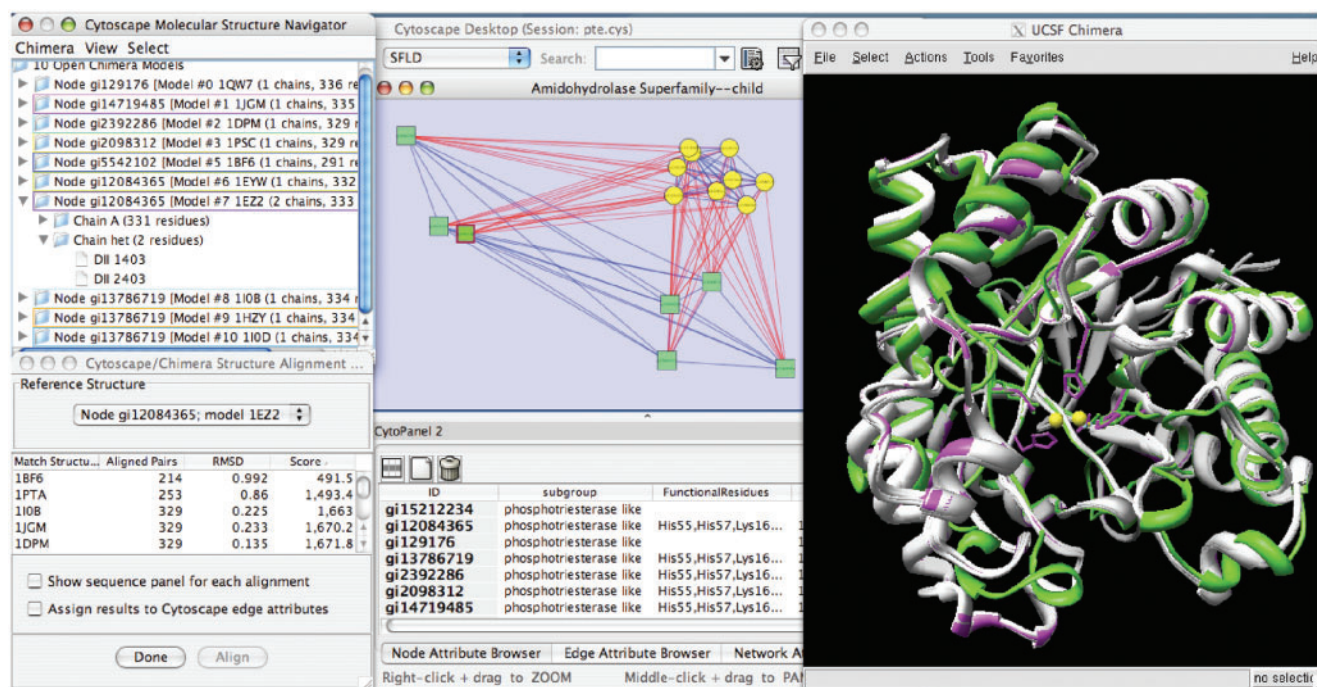


Fig. 1. *structureViz* in action. In this screenshot, a small portion of the amidohydrolase superfamily is shown in Cytoscape. The *phosphotriesterase-like* subgroup is represented as yellow round nodes, the *unknown119* subgroup nodes as green square nodes. GI#5542102, which is the only *unknown119* protein with a solved structure (1BF6) is outlined in red. Edges between groups are colored red, and edges within groups are colored blue. Ten of the available structures associated with these subgroups have been opened, aligned and their backbones displayed in UCSF Chimera. To highlight the differences between 1BF6 and the other *phosphotriesterase-like* structures, 1BF6 has been colored green, 1E2Z (a canonical phosphotriesterase structure) has been colored magenta; the other eight structures have been colored white and virtually superimpose with one another.

are weighted by their sequence similarity (BLAST *E*-value). Two subgroups of the amidohydrolase superfamily are shown: the *phosphotriesterase-like* subgroup and *unknown119*, both of which have been assigned to the *PTE* family by PFAM (Finn *et al.*, 2006). While *unknown119* proteins have very high sequence similarity to the other phosphotriesterases, they do not hydrolyze synthetic organophosphates, the functional characteristic of the phosphotriesterase family (Buchbinder *et al.*, 1998).

In this example, *structureViz* was used to open 10 structures associated with these two subgroups in Chimera (several structures are known for the *phosphotriesterase-like* proteins, and one is known for *unknown119*). In addition to the standard Chimera interface, *structureViz* provides an alternate, simpler, tree-oriented interface to allow users to interactively explore, manipulate and change the visual representation of the structures in Chimera. If a comma-separated list of residue identifiers is available as a node attribute named *FunctionalResidues*, *structureViz* provides the capability to specifically select and change the depiction of those residues, e.g. to show additional detail. *structureViz* also provides an interface to Chimera's sequence-structure analysis tools (Meng *et al.*, 2006), which have been used to align the canonical phosphotriesterase structure (pdb: 1E2Z) with each of the other chosen structures. The results have been displayed in a table, which shows a close alignment for the *phosphotriesterase-like*

subgroup (RMSD of $0.22 \text{ \AA} \pm 0.05$ over 329 amino acid pairs) while the structure from *unknown119* shows an RMSD of 1.0 \AA over 207 pairs. This result is in agreement with the sequence comparison and provides further evidence that 1BF6 (and perhaps its closest subgroup members) is distinct from the *phosphotriesterase-like* subgroup. *structureViz* provides the user with the ability to record these results in the Cytoscape network by creating additional edges analogous to the sequence similarity edges already present.

4 SIGNIFICANCE

structureViz, along with the node annotation file, provide important new tools for both the systems biologist and the structural biologist. With *structureViz*, multiple levels of abstracted biological data are integrated into a single interface for visualization and analysis. This takes advantage of two powerful, yet typically disparate fields of study: structural biology and systems biology, and hence can inspire new ways of thinking about biology. The systems biologist is now able to much more easily explore the possible structural implications of protein-protein partners, or neighbors in a metabolic pathway. It also provides an important ability to explore the possible functional and evolutionary significance of sequence relationships between proteins as demonstrated in our example. For the structural biologist, *structureViz*, along with Cytoscape can

provide useful information that can suggest targets for experimental verification of structural relationships, or targets for solving the structure through NMR, crystallography, or modeling. *structureViz* is a Cytoscape plug-in, and therefore builds upon the entire Cytoscape framework and its broad range of plug-ins, yet *structureViz* is the first tool to link the critical areas of systems and structure in a straightforward manner.

5 CONCLUSION

structureViz is a useful link between a network view provided by a tool like Cytoscape and a detailed molecular view provided by a tool like UCSF Chimera. Our future plans are to enhance this linkage in a number of ways, including the ability to perform sequence comparisons in order to construct new edges in the network and access to Chimera's growing modeling capabilities, including direct access to MODBASE structures. As these capabilities become available, the annotation file provided on the web will be augmented so that users will be able to annotate their networks with important structural links.

UCSF structureViz is written in java and the jar file and sample data files are available from the UCSF RBVI web site at: <http://www.rbvi.ucsf.edu/Research/cytoscape/structureViz>. Users can also link to a detailed step-by-step tutorial on how to use *structureViz* from the same site. Chimera is available for download from <http://www.rbvi.ucsf.edu/chimera>

and Cytoscape is available for download from <http://www.cytoscape.org/>. Cytoscape and Chimera must be installed locally prior to using *structureViz*.

ACKNOWLEDGEMENT

This work was supported by NIH grant P41 RR01081.

Conflict of Interest: none declared.

REFERENCES

- Berman, H.M. *et al.* (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 957–959.
- Buchbinder, J.L. *et al.* (1998) Biochemical characterization and crystallographic structure of an Escherichia coli protein from the phosphotriesterase gene family. *Biochemistry*, **37**, 5096–5106.
- Finn, R.D. *et al.* (2006) Pfam: clans, web tools and services. *Nucleic Acids Res.*, **34**, D247–251.
- Meng, E.C. *et al.* (2006) Tools for integrated sequence-structure analysis with UCSF Chimera. *BMC Bioinformatics*, **7**, 339.
- Pegg, S.C. *et al.* (2006) Leveraging enzyme structure-function relationships for functional inference and experimental design: the structure-function linkage database. *Biochemistry*, **45**, 2545–2555.
- Pieper, U. *et al.* (2006) MODBASE: a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res.*, **34**, D291–295.
- Pettersen, E.F. *et al.* (2004) UCSF Chimera – a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.
- Shannon, P. *et al.* (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.