



COMBOSA3D: combining sequence alignments with three-dimensional structures

P. M. Stothard

Department of Biological Sciences, University of Alberta, Edmonton, Alberta, T6G 2E9, Canada

Received on August 6, 2000; revised on October 1, 2000; accepted on October 6, 2000

ABSTRACT

Summary: COMBOSA3D is a program that allows sequence conservation to be viewed in its proper three-dimensional environment. It superimposes sequence alignment information onto a protein structure using a customizable color scheme, which is also applied to a textual sequence alignment for reference.

Availability: The program can be tested at <http://www.bioinformatics.org/combosa3d/>, and the source code is freely available.

Contact: stothard@ualberta.ca

Supplementary information: Additional COMBOSA3D documentation is available at the above URL.

INTRODUCTION

If you flip through any current journal that deals with molecular biology you are likely to see several sequence alignments. These figures are popular because they illustrate potential evolutionary and functional relationships among sequences in an easy to interpret format. In cases where one of the sequences has a known three-dimensional structure it can be informative to compare the alignment with the solved structure, to better understand how the local environment of the residues relates to conservation. This type of comparison often involves preparing a sequence alignment figure in which the secondary structure characteristics of the solved sequence are added manually as drawings. Another method involves adjusting the font of the solved sequences in an alignment figure to reflect particular characteristics extracted from the three-dimensional structure (Mizuguchi *et al.*, 1998). Here I describe COMBOSA3D (Coloring Of Molecules Based On Sequence Alignment), a program that combines sequence alignment information with three-dimensional structures in an alternative and informative way. Instead of adjusting the appearance of a sequence alignment figure to convey structural information, COMBOSA3D adjusts a three-dimensional structure to convey alignment information. The resulting annotated structure can be used to address questions regarding the functional significance of evolutionary and mutational differences in an intuitive way.

OVERVIEW OF COMBOSA3D

COMBOSA3D is primarily mouse and menu driven and it runs in popular web browsers, making it accessible to a wide range of researchers. The interface to the program consists of a single window divided into four resizable viewing areas, or panes. One of the panes uses the Chime molecular structure viewing plug-in (available from <http://www.mdli.com/download/>) to display three-dimensional structures, while the others are used primarily for data entry and structure manipulation. Operation of the program is divided into three basic steps, beginning with the loading of a protein structure of interest. Protein structures are obtained remotely from the Protein Data Bank (Berman *et al.*, 2000) using a Protein Data Bank ID, which is entered in the Molecule Control pane. To manipulate the orientation and display of the loaded structure a set of buttons and menus is available, along with a text area that accepts Chime and RasMol (Sayle and Milner-White, 1995) scripts. Next, the pre-aligned set of sequences that is to be compared with the loaded three-dimensional structure is entered into the Alignment text area, located in the Sequence pane. Finally, the loaded molecule's sequence (which may or may not be included in the Alignment text area) is pasted in aligned form into the Solved Sequence text area. The program then evaluates each column in the alignment entered into the Alignment text area, to determine the level of sequence similarity, and a color is assigned to the aligned residue in the solved sequence, depending on the level of sequence similarity detected. After all the columns have been evaluated the resulting color scheme is applied to the loaded structure (Figure 1A). A formatted textual alignment colored using the same color scheme is also generated to serve as a reference (Figure 1B). Pointing to a residue in the solved sequence in the reference alignment displays the residue's position in the web browser's status bar. Alignment segments of interest can be labeled on the three-dimensional structure by entering residue positions or ranges into the Molecule Control pane. Alternatively, specific residues in the three-dimensional structure can be identified in the status bar by clicking on the structure.

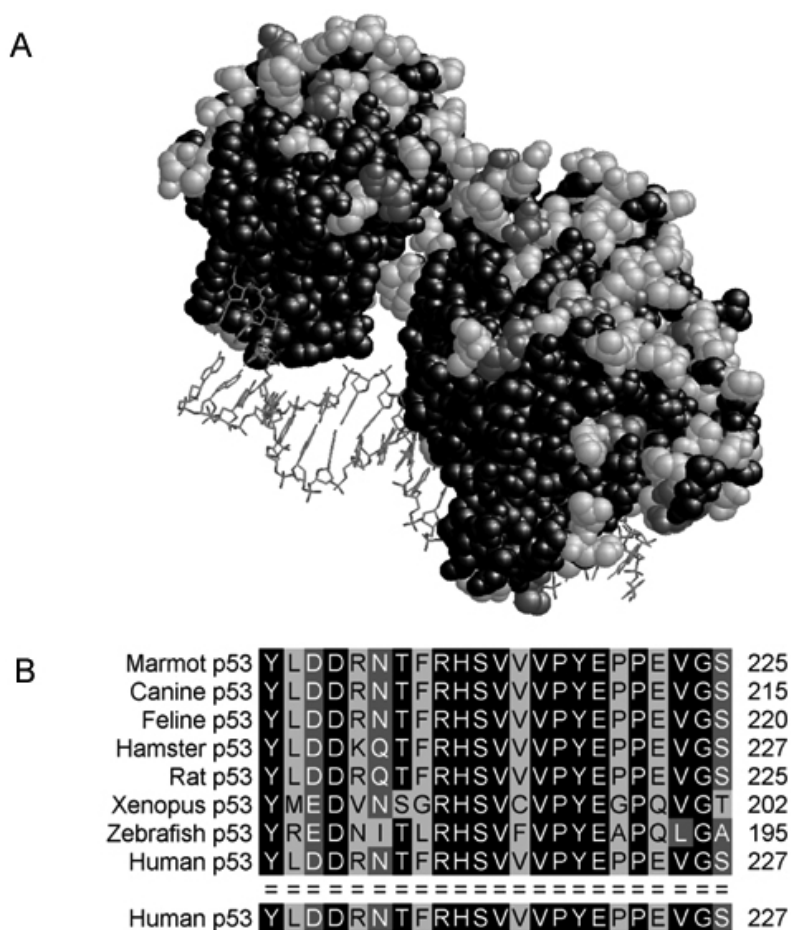


Fig. 1. Sample output of COMBOSA3D showing (A) a colored three-dimensional representation of p53 (Cho *et al.*, 1994) binding to DNA and (B) a portion of the colored reference alignment. Residues that are identical among the comparison sequences (the sequences located above the row of equal signs in (B)) are colored black, while similar residues are colored dark gray. Residues that are not identical or similar among 80% of the sequences are colored light gray.

A variety of options are available to adjust the way that COMBOSA3D calculates color schemes. The percentage threshold for the coloring of indels (amino acids in the solved sequence that align with gaps in the other sequences), identical, and similar residues can be set, and the user can define which groups of residues qualify as similar. The colors used to represent indels, identical, similar, and divergent residues can also be specified. A range of gray colors is included in the color options so that informative black and white figures can be generated. The molecule and the reference alignment can be re-colored using new settings at any time by clicking the Color Molecule and Show Alignment buttons, respectively. Once the desired view of the colored molecule is obtained, a picture of the molecule can be copied to the Clipboard (using the Clipboard button) for pasting into a graphics program, and the reference alignment can be saved as an HTML file or printed from within the web browser.

ACKNOWLEDGEMENTS

The author would like to thank Eric Martz (University of Massachusetts) for helpful comments relating to COMBOSA3D.

REFERENCES

- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
- Cho, Y., Gorina, S., Jeffrey, P.D. and Pavletich, N.P. (1994) Crystal structure of a p53 tumor suppressor–DNA complex: understanding tumorigenic mutations. *Science*, **265**, 346–355.
- Mizuguchi, K., Deane, C.M., Blundell, T.L., Johnson, M.S. and Overington, J.P. (1998) JOY: protein sequence–structure representation and analysis. *Bioinformatics*, **14**, 617–623.
- Sayle, R.A. and Milner-White, E.J. (1995) RASMOL: biomolecular graphics for all. *Trends Biochem. Sci.*, **20**, 374.