Structural Bioinformatics

VISTAL – A new 2D visualization tool of protein 3D structural alignments

Rachel Kolodny[1] and Barry Honig[1,*]

[1] Howard Hughes Medical Institute, Department of Biochemistry and Molecular Biophysics, Center for Computational Biology and Bioinformatics, Columbia University, 1130 St Nicholas Avenue, Room 815, New York, NY 10032, USA.

ABSTRACT

Summary: We offer a tool, denoted VISTAL, for two-dimensional visualization of protein structural alignments. VISTAL describes aligned structures as a series of matched secondary structure elements, colored according to the three-dimensional distance of their Cα atoms.

Availability: VISTAL can be downloaded from http://trantor.bioc.columbia.edu/~kolodny/software.html

1 INTRODUCTION

Protein structural alignment offers a powerful tool when predicting the structure and function of proteins. Structural similarity suggests shared functionality, and because structure is more conserved than sequence, it can also point to distant evolutionary relationships. Structural alignment programs find substructures in two proteins that bear high geometric similarity, and match their corresponding residues. It is often of interest to examine how closely these substructures match. Also, there are many publicly available methods for structural alignment. By comparing the alignments of different programs, we can identify the best one, effectively creating a 'Best-of-All' method (Kolodny, et al., 2005). We present a tool, called VISTAL (VISualizing Structural Alignments), for easy and meaningful two-dimensional (2D) visualizations of protein structural alignments.

The quality of a structural alignment can generally be judged by the number of matched residues (preference for long alignments), their pairwise distances in three-dimensional (3D) space (preference for small distances), and the number of gaps (preference for few). The matching of secondary structure elements (SSEs) is also important. Yet a typical presentation of structural alignment only shows the matched residues and the position of the gaps. In many structural alignment programs the Cα atoms of matched residues are closer in space than some cutoff value, yet the exact distance is not displayed. Thus, the output in an alignment in which the matched residues are very close does not differ from that in which these residues are as far as the cutoff value.

Of course, one can inspect the structural similarity of two proteins by looking at the superimposed structures in a 3D molecular viewer. There are several drawbacks to this visualization: (1) the parts in the front of the viewing window obscure the parts in the back. Thus we must rotate the structures to see the details, and the parts in the protein core may even be obscured from all view points. (2) When comparing residue pairs that are not very close or very far away, it is hard to estimate which are closer than others. (3) 3D visualization is not easily presented in a manuscript. Lastly, when only superimposing the structures we do not see the alignment itself, e.g. the position of the gaps.

VISTAL creates 2D pictures which describe the structural properties of protein alignments. A protein structure is described in VISTAL as a series of secondary structure elements (SSEs). The SSEs are calculated using DSSP (Kabsch and Sander, 1983) and passed as input to the program. Here, we draw helices as rounded fat rectangles (residues classified as H, G, and I by DSSP), strands as arrows (residues classified as E and B), coil as a thick lines (all other residues), and gaps as dashed lines. Matched residues are placed one above the other and colored according to the distance in 3D space between their Cα atoms. We use a color scale that ranges from blues for nearby distances through cyan and yellow for medium scale distances to reds for the furthest away ones. The maximum value in the default scale is 15A, a distance above the cutoff value used by most structural alignment programs. Distances greater than the maximum value are colored in pink. VISTAL calculates the distances after superimposing the two structures such that the root mean square distance (RMSD) of the matched residues is minimized (Kabsch, 1978). The output of VISTAL is in metapost language (Hobby, 1989). The metapost program then generates a postscript file with the structural alignment visualization.

2 USAGE

The input to VISTAL is: the two protein structure files (in pdb format), the chain ids, two secondary structure assignment files (in DSSP format), and an alignment file. The alignment file lists which residues are matched, and can be either in VISTAL format, or in the output format of SKA (Petrey and Honig, 2003), Structal (Subbiah, et al., 1993), CE (Shindyalov and Bourne, 1998), or BLAST(Altschul, et al., 1990). Note that BLAST does not compare the structures and thus is not appropriate for finding structural similarities in highly divergent sequences. Since aligned residues are used to superimpose the two structures, all residues listed as aligned in the input file must be in the structure files. Colors can be rescaled by specifying the maximal distance (the default is 15A), or changed to grayscale for black & white pictures. The user

*To whom correspondence should be addressed: bh6@columbia.edu.

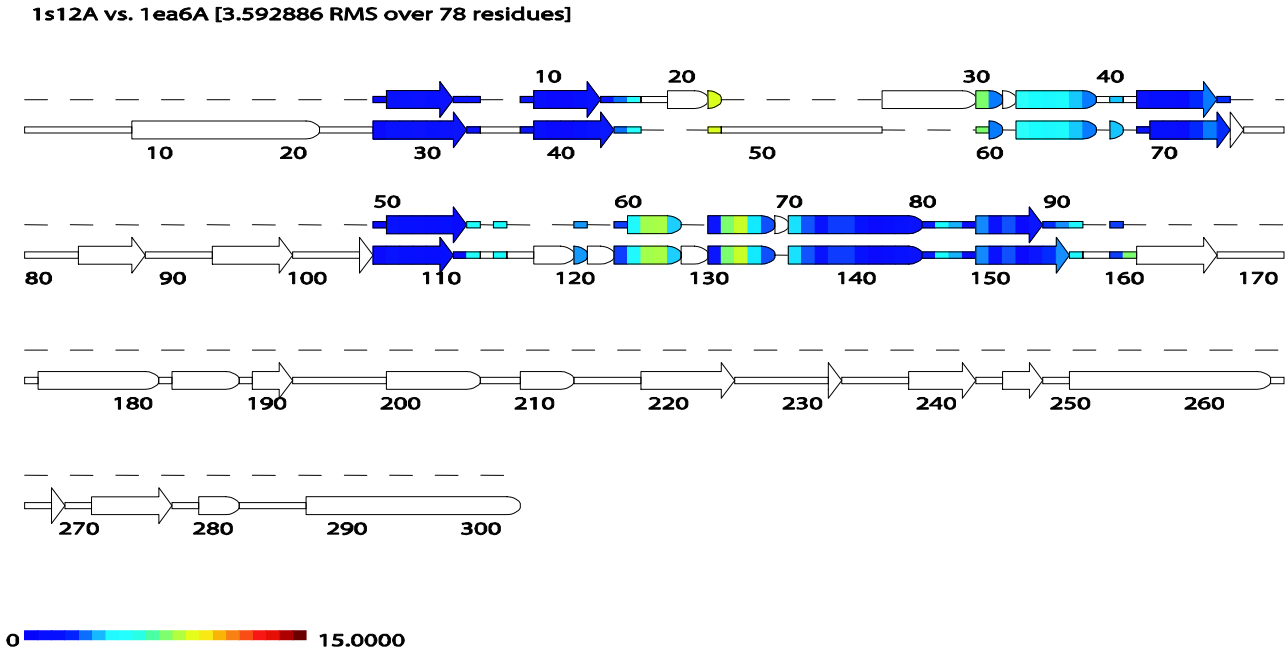

Figure 1: VISTAL visualization of the structural alignment by SKA of TM1457, a hypothetical protein from T.maritima (1s12A) and DNA mismatch repair protein PMS2 (1ea6A). The aligned residues are colored according to the distance in 3D space of their C-α atoms. We see that the two structures share many structural features.

needs to store the output of VISTAL in a temporary file, and pass it as input to the metapost program, which generates a postscript file with the same name. VISTAL alignments can also be displayed within a web-page. For this, the output file must be converted from postscript format to a web-friendly format such as jpeg; this can be done in a command line using the linux program convert. VISTAL generates metapost files with images that are 500 postscript points wide, wrapping longer alignments into multiple lines. In a web-page presentation, it can be of interest to compare multiple structural alignments by displaying them consecutively, using the horizontal scroll bar if necessary. For this, we offer a version of VISTAL that generates metapost files that create multiple postscript files, each of a fixed width; these can then be converted to jpeg files and placed in a wide table, one next to the other. See *http://trantor.bioc.columbia.edu/~kolodny/software/vistal_example.html* for an example.

3 AN EXAMPLE

We visualize the structural alignment found by SKA (Petrey and Honig, 2003) of TM1457, a hypothetical protein from *T.maritima* (PDB code 1s12A) and the DNA mismatch repair protein PMS2 (PDB code 1ea6A). Using the VISTAL view, we can easily examine the accuracy of this structural alignment. Namely, there are two beta-strands that align well, followed by a helix that aligns less well, two strands that align well, and then a helix that aligns less well in its N-terminal end. Finally there is a pair of well-aligned strands. 1s12A has a helix, after the two initial aligned strands, which is not aligned, and 1ea6A has two strands after the initial four aligned SSEs which are not aligned. All aligned residues are colored in blues and cyan-yellows, indicating that the matched residues are positioned nearby in space. We see that the two proteins share many structural features, even though TM1457 was labeled “new fold” in the recent CASP experiment. For more details regarding the significance of this example see (Kolodny).

ACKNOWLEDGEMENTS

We are grateful to Mickey Kosloff for many helpful comments. The development of VISTAL was supported in part by NIH grants GM30518 and GM074958.

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool, J Mol Biol, 215, 403-410.

Hobby, J.D. (1989) A METAFONT-like System with PostScript Output, Tugboat, the TeX User's Group Newsletter, 10

Kabsch, W. (1978) A discussion of the solution for the best rotation to relate two sets of vectors, Acta Crystallographica Section A, 34, 827-828.

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

Kolodny, R., Koehl, P. and Levitt, M. (2005) Comprehensive Evaluation of Protein Structure Alignment Methods: Scoring by Geometric Measures, Journal of Molecular Biology, 346, 1173-1188.

Kolodny, R., Petrey, D., Honig, B. Protein Structure Comparison: Implications for the Nature of “Fold Space” and Structure and Function Prediction, Current Opinion in Structural Biology, *In Press*.

Petrey, D. and Honig, B. (2003) GRASP2: Visualization, Surface Properties, and Electrostatics of Macromolecular Structures and Sequences. In, Methods in Enzymology. Academic Press, 492-509.

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.

Subbiah, S., Laurents, D.V. and Levitt, M. (1993) Structural similarity of DNA-binding domains of bacteriophage repressors and the globin core, Curr Biol, 3, 141-148.