*Original paper*

# Helix Interaction Tool (HIT): A web-based tool for analysis of helix-helix interactions in proteins

Anne E. Counterman Burba*, Ursula Lehnert, Eric Z. Yu, and Mark Gerstein

Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520

**ABSTRACT**

**Motivation:** In many proteins, helix-helix interactions can be critical to establishing protein conformation (folding) and dynamics, as well as determining associations between protein units. However, the determination of a set of rules that guide helix-helix interaction has been elusive. In order to gain further insight into the helix-helix interface, we have developed a comprehensive package of tools for analyzing helix-helix packing in proteins. These tools are available at http://helix.gersteinlab.org. They include quantitative measures of the helix interaction surface area and helix crossing angle, as well as several methods for visualizing the helix-helix interaction. These methods can be used for analysis of individual protein conformations or to gain insight into dynamic changes in helix-helix interactions. For the latter purpose, a direct interface from entries in the Molecular Motions Database to the HIT site has been provided.

## 1 INTRODUCTION

Helix-helix interactions are of interest because they provide stabilization in many protein structures. Helix interactions have particularly significance in membrane proteins, where transmembrane helical segments often direct protein orientation with respect to the lipid bilayer.

Historically, initial models of packing between α-helices were developed from crystallographic studies of fibrous proteins such as α-keratin. Following the proposal of the α-helix structure by Pauling and coworkers (Pauling *et al.*, 1951), Crick delineated a 'knobs in holes' model for a helical coiled coil (Crick, 1953). Chothia and coworkers later developed the 'ridges into grooves' model of helix intercalation (based on structures of ten globular proteins of known crystallographic structure) to explain an average observed helix packing angle of ~50° (Chothia *et al.*, 1981), which differed from the 20° and -70° packing angles reported by Crick. While both of these models focus on geometric considerations, Chothia and coworkers also note that the side chains of residues forming a helix have an effect on packing: smaller residues near the center of a helix-helix contact are associated with larger helix-helix interfaces.

Membrane proteins, many of which are particularly rich in helix-helix interactions, provide an important basis set for studies of helix packing. As the number of available crystal structures for membrane proteins has increased, more detailed studies of packing and sequence effects can be performed. Helices in membrane proteins have been found to be more tightly packed than those in soluble proteins (Gerstein and Chothia, 1999; Eilers *et al.*, 2000). Among the efforts to understand packing in membrane proteins, considerable attention has been given to the occurrence of motifs [e.g., GxxxG (Russ and Engelman, 2000)] which are identified by statistical comparison of the expected and actual occurrence of the sequence motif within protein structures. Additionally, Adamanian and Liang described "polar clamp" and "serine zipper" spatial motifs which are located in regions of tight interhelical packing, indicating that interhelical hydrogen bonding can play an important role in determining packing (2002). Networks of weak $C_\alpha$—H···O hydrogen bonds have also been found in membrane protein helix-helix interfaces, and even appear to be favored in parallel right-handed helix interactions (Senes *et al.*, 2001).

In order to gain insight into helix-helix interactions (and ultimately develop predictive methodologies), it is important to mesh considerations of packing, sequence, and overall interaction geometry. In this article, we describe a comprehensive suite for performing all of the main helix geometry calculations – computing helix angles, determining residue contacts and surface areas at the helix interface, and delineating sequence motifs. The package, which we have named 'HIT,' or 'Helix Interaction Tool', is implemented at http://helix.gersteinlab.org. The site includes two methods of assessing atomic contacts between helices: a distance-based assessment and a determination based on atomic packing (Richards, 1974; Richards, 1985; Harpaz *et al.*, 1994; and Gerstein *et al.*, 1995) (Voronoi method). Where feasible, we have provided tools for visualizing packing information, allowing the user to obtain a clearer understanding of the results.

## 2 METHODS AND RESULTS

### 2.1 Web site overview

The web server that we have implemented for analyzing helix-helix interactions is general; any protein containing two or more helices that interact can be analyzed. Figure 1 shows a schematic of the layout of our web site. The home page of our server allows two options for selecting a protein coordinate file. Users can input the 4-character PDB identification code to extract a coordinate file from the Protein Data Bank (PDB; http://www.rcsb.org/pdb). Alternatively, users can upload a coordinate file that uses the PDB format. In the latter case, STRIDE (Frishman and Argos, 1995) is used to identify the start and end residues of each helix. The user is initially directed to a page confirming successful upload of the PDB file that serves as the launch page for the computational and visualization tools that we have implemented on the site. Figure 2a shows the layout of the tool selection menu. Currently available tools include a packing-based helix interaction report, visualization of the helix-helix interface using Voronoi polyhedra (Richards,

*To whom correspondence should be addressed.

1974; Richards, 1985), calculation of intersection area between helices, and a sequence motif search. Throughout the site, we have tried to incorporate a visual representation of the results using Jmol or VRML.

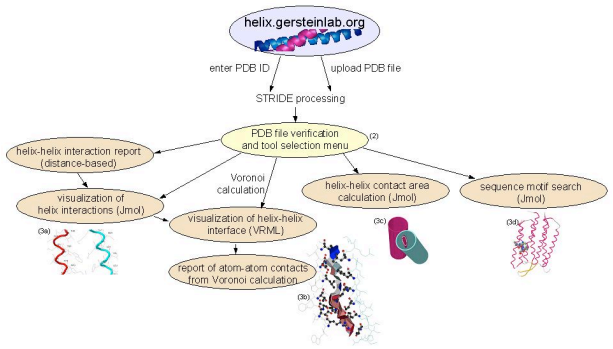


**Fig. 1.** Diagram of web site content. Numbers in parentheses indicate references to other figures. Visualizations of helix-helix interactions shown near the bottom of the figure are shown in more detail in Figure 3 [sections (a)-(d), as delineated].

## 2.2 Defining helix-helix interactions

We have used two methods to delineate atom-atom contacts between two helices: (1) a distance-based constraint and (2) a method based on considerations of atomic packing. For the distance-based constraint (method 1), we applied criteria comparable to those used by Bowie (Bowie, 1997). In this method, atoms from two helices were determined to interact if the distance between them was less than the sum of their van der Waals radii plus a threshold value of 0.6 Å. Two helices were assumed to interact if at least three van der Waals contacts were found.

On our web site, once a protein coordinate file has been received, the server performs the distance-based analysis of helix-helix interactions. A report is output to the web browser that contains a list of the interacting helix pairs, the residue numbers of residues involved in each pairwise interaction, and the number of atoms associated with each residue-residue interaction. As shown in Figure 2b, this report is given in the following format:

HELIX 1 1 GLU A 9 GLY A 31 1 23
HELIX 2 2 ASP A 36 GLY A 63 1 28
[13,57 {4}][13,60 {4}][13,61 {3}][14,61 {4}][16,57 {4}][17,54 {3}]….

where the first two lines of each record set give (in PDB format) the information on the two interacting helices, and the list of residue-residue contacts is given in the form [residue 1, residue 2 {number of contacts}]. A summary of the number of residue-residue and atom-atom contacts is also provided at the end of each record set (i.e., for each pairwise helix interaction). The report concludes with the total number of helix-helix interaction pairs. The interaction summary page also links to other analysis and visualization tools that we have developed (Figure 1).


**Fig 2.** (a) Screenshot of tool selection page. This page also verifies successful upload of the requested PDB file. (b) Screenshot of helix-helix interaction report page. The report presented here specifies residue-residue interactions and the number of atom-atom contacts for each helix-helix interaction. The links at the bottom of the web page direct the user to additional tools for analysis and visualization.

The second method that we used to determine contacts between two helices considers the partitioning of space between them using the Voronoi method (Richards, 1974; Richards, 1985; Harpaz *et al.*, 1994; and Gerstein *et al.*, 1995). Our calculations use a set of standard radii that have been optimized for calculations of packing in proteins (Tsai *et al.*, 1999). Briefly, the Voronoi method partitions space around the atoms in a molecule, constructing a polyhedron around each atom. The number of atom-atom contacts between helices determined by this method is closely correlated to the number of contacts found by the distance-constraint method, but is not identical. Part of the difference between the methods can be explained by noting that the distance-based constraint may miss some long-range atom-atom interactions, such as those associated with the favorable electrostatic interaction between two oppositely charged residues (e.g., Asp and Arg).

When determining atom-atom contacts based on packing, we considered not only the atoms comprising two interacting

helices, but also neighboring atoms that do not belong to one of the interacting helices (the "environment"). The environment surrounding each atom is important to determining the Voronoi polyhedra. For the packing calculations, we included atoms within 6.0 Å of the atoms associated with the pair of interacting helices. We chose this cutoff value of 6.0 Å by performing calculations using a series of different cutoff values; this value constitutes the threshold above which adding more atoms from the environment does not change the packing results. The report for atom-atom contacts determined using this method is accessed on the web site by clicking the "Pair interaction surface" button found on the summary page for the distance-based contact summary (described above) and entering the helix pair of interest.

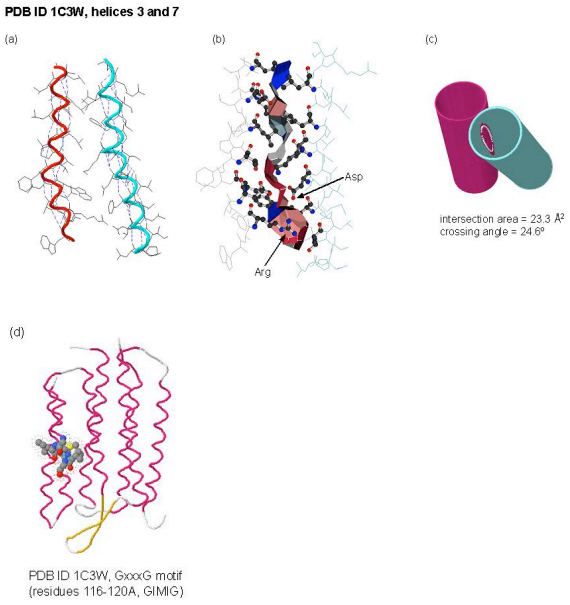


**Fig 3.** Visualization of helix-helix interactions. Parts (a)-(c) show different representations of the helix-helix interaction for helices 3 and 7 of bacteriorhodopsin (PDB identification code: 1C3W). Part (a) shows hydrogen bonding within the two helices, and allows user manipulation to view general geometric characteristics. Part (b) focuses specifically on the helix-helix interface; the surface depicted between the helices is formed by the shared polyhedra faces derived from the Voronoi packing calculation. Part (c) shows a representation of the interacting helices as cylinders; this allows a calculation of intersection area. Part (d) shows the visualization of a sequence motif search: all occurrences of a user-selected motif (in this case, GxxxG) are found in the protein sequence and visually highlighted (individual occurrences can be highlighted separately by the user).

### 2.3 Helix-helix interface analysis using Voronoi polyhedra

The user can also choose to consider only a specific helix pair from the perspective of atom packing at the interface. Two interacting helices are selected, and the packing calculation is performed. Faces of the Voronoi polyhedra that are shared by atoms of the two interacting helices are culled from the computational results, and displayed using a VRML viewer. The composite of these polyhedra faces comprises the helix-helix interface. The user is also provided with the option to upload a 20x20 matrix to color-code the helix-helix interface according to chemical or physical properties associated with each residue-residue interaction. Figure 3b shows an example of the VRML output for the interface of helices 3 and 7 in PDB entry 1EHK. Seven residues in helix 3 and six residues in helix 7 are involved in the interface. This helix pair provides an example of the distinction between the distance-constraint method and the Voronoi contact analysis: a long-range electrostatic interaction between Asp and Arg is reported by the Voronoi method, but missed in the distance-constraint analysis.
A detailed report of the atomic interactions and volumes for the selected helix pair can be accessed via a link from the visualization page. At the end of each report, we provide a summary of the number of atom-atom contacts (defined as shared Voronoi polyhedron faces) and the total area of shared Voronoi polyhedron faces. The sum of areas of polyhedron faces shared by the two interacting helices provides a rough quantitative measure of the size of the interface region.

### 2.4 Helix-helix contact area

In order to obtain another quantitative measure of helix-helix contacts, we have incorporated a calculation of contact area. In our method, each helix of a helix pair is represented as a cylinder of fixed radius. The endpoints of the cylinders are calculated using the HelixTips program (included as part of the software available via http://geometry.molmovdb.org). The intersection area of the cylinders is then computed and reported on the results page (Figure 3c). For example, the crossing angle of helices 3 and 7 from PDB entry 1C3W (bacteriorhodopsin) is 24.6° and the intersection area is calculated to be 23.3 Å. A detailed discussion and comparison of contact area calculations is provided elsewhere (Yu and Gerstein, submitted).

### 2.5 Sequence motifs

Finally, another area of interest is interaction motifs in helices (e.g., GXXXG). By entering a motif or selecting from a list of common motifs, the user can search for the positions of the selected motif in the protein. These residue positions are listed on the subsequent results page, and the user can visualize their location in the three-dimensional protein structure either individually or *en masse* (if multiple occurrences of the motif are found). In Figure 3d, the result page obtained by searching for the GxxxG motif in PDB entry 1C3W (bacteriorhodopsin) is shown as an example. For this protein, the GxxxG motif occurs only once, as sequence GIMIG (residues 116 to 120). The motif search feature is particularly helpful if the user wishes to perform a quick visual check of motif location (for instance, whether a motif is located at a helix terminus or at a helix-helix interface). A report of helix and atomic contacts is also provided for each motif.

## 3 SUMMARY AND FINAL NOTES

One interesting aspect of helix-helix interactions is how they change with protein dynamics. There is a server—MolMovDB (http://www.molmovdb.org) [Gerstein and Krebs, 1998, Flores *et al.,* 2006] –that serves as a database for coordinate files that capture changes in protein conformation and that provides several tools for analysis of protein motion. Each database entry can now be analyzed for helix-helix interactions via a link on the MolMovDB entry reporting page to our helix analysis server.

Our motivation in developing the tools for helix analysis described here and making them web-accessible was to facilitate analysis of helix-helix interactions in membrane proteins (although the tools themselves are general, and not restricted to membrane proteins). We have created a gallery of the structures of known membrane proteins, with links to our analysis tools for each protein. Work is ongoing to expand our web resource to accumulate and present information specific to membrane proteins and transmembrane helix interactions.

## ACKNOWLEDGEMENTS

We thank the National Institutes of Health (NIH) for financial support. We also acknowledge helpful contributions from Thomas Royce and Sara Nichols to the early stages of this work.

## REFERENCES

Adamian, L. and Liang, J. (2002) Interhelical hydrogen bonds and spatial motifs in membrane proteins: polar clamps and serine zippers. *Proteins: Struct. Funct. Genet.* 47, 209-218.

Bowie, J. U. (1997) Helix packing in membrane proteins. *J. Mol. Biol*. 272, 780-789.

Crick, F. H. C. (1953) The packing of α-helices: simple coiled coils. *Acta Cryst*. 6, 689-697.

Eilers, M., Shekar, S. C., Shieh, T., Smith, S. O. and Fleming, P. J. (2000) Internal packing of helical membrane proteins. *Proc. Natl. Acad. Sci. USA* 97, 5796-5801.

Flores, S., Echols, N., Milburn, D., Hespenheide, B., Keating, K., Lu, J., Wells, S., Yu, E. Z., Thorpe, M., Gerstein, M. (2006) The database of macromolecular motions: new features added at the decade mark. *Nucleic Acids Res.* 34, D296-301.

Frishman, D. and Argos, P. (1995) Knowledge-based secondary structure assignment. *Proteins:*
*structure, function and genetics*, 23, 566-579.

Gerstein, M., Tsai, J. and Levitt, M. (1995) The volume of atoms on the protein surface:
calculated from simulation, using Voronoi polyhedra. *J. Mol. Biol.* 249, 955-966.

Gerstein, M. and Krebs, W. G. (1998) A database of macromolecular motions. *Nucleic Acids Res.*, 26, 4280-4290.

Gerstein, M. and Chothia, C. (1999) Signal transduction: proteins in motion. *Science*, 285, 1682-1683.

Harpaz, Y., Gerstein, M. and Chothia, C. (1994) Volume changes on protein folding. *Structure*,
2, 641-649.

Pauling, L., Corey, R. B., and Branson, H. R. (1951) The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. USA* 37, 205-211.

Richards, F. M. (1974) The interpretation of protein structures: total volume, group volume
distributions and packing density. *J. Mol. Biol.* 82, 1-14.

Richards, F. M. (1985) Calculation of molecular volumes and areas for structures of known geometry. *Methods Enzymol.* 115, 440-464.

Russ, W. P. and Engelman, D. M. (2000) The GxxxG motif: a framework for transmembrane
helix-helix association. *J . Mol. Biol.* 296, 911-919.

Senes, A., Gerstein, M., Engelman, D. M. (2000) Statistical analysis of amino acid patterns in
transmembrane helices: the GxxxG motif occurs frequently and in association with beta-
branched residues at neighboring residues. *J. Mol. Biol.* 296, 921-936.

Senes, A., Ubarretxena-Belandia, I., and Engelman, D. (2001) The $C_\alpha$—H···O hydrogen bond: a determinant of stability and specificity in transmembrane helix interactions. *Proc. Natl. Acad. Sci. USA* 98, 9056-9061.

Tsai, J., Taylor, R., Chothia, C., and Gerstein, M. (1999). The Packing Density in Proteins: Standard Radii and Volumes. *J. Mol. Biol.* 290, 253-266.

Yu, E. Z. and Gerstein, M. (submitted for publication) Fast and Accurate Prediction of Helix-Helix Surface Contacts: No Side Chain Information Needed.