

Biopolymer™

PREDICT, BUILD, VISUALIZE, AND ANALYZE BIOPOLYMER STRUCTURES



Biopolymer is fully integrated with SYBYL®. It provides an intuitive graphical user interface to all tools needed for building, manipulating, and analyzing structures of peptides, proteins, DNA, RNA, and carbohydrates. For proteins, the Molecular Spreadsheet supports automated calculation of structure-related properties for the individual residues, which can subsequently be displayed as graphed or mapped onto the molecular display. Biopolymer accurately identifies potential binding sites for ligand-receptor interactions.

Applications

- Protein PDB preparation utility to clean up structures for structure-based drug design
- Create models of peptides, proteins, RNA, DNA, and polysaccharides
- Virtual mutagenesis experiments
- Pattern searching of the Protein Data Bank
- Identify and assess structural irregularities in protein models derived by theoretical or experimental methods
- Refine structures derived from crystallography, NMR, or modeling
- Locate ligand binding pockets on a biomolecule
- Identify protein-protein interaction surfaces

The Problem

The function of a biomolecule is facilitated by the molecule's three-dimensional structure. Structures are determined by a variety of experimental and *in silico* methods, but it requires visualization and analysis to judge the reliability of a predicted structure model.

Minimal changes in the sequence can have dramatic effects on a biomolecule's structure. It is important to understand what impact a specific change in sequence will have on the overall fold. The calculation of specific parameters for individual residues and their representation in graphs and on molecular displays can help to understand the effect of such changes.

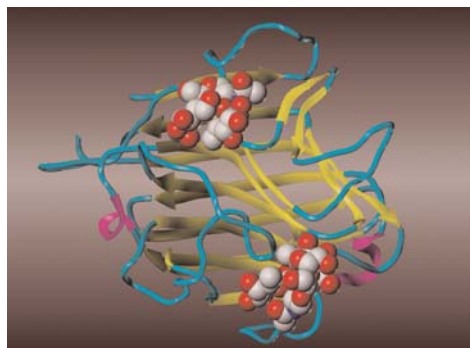
Also, biomolecules typically interact with other molecules to exhibit a specific biological behavior. Designated cavities or binding sites are responsible for the

binding of ligands. From the original three-dimensional structure determination it is not always obvious where such binding sites are or there may be additional and competitive binding sites in a protein. Automated structure analysis can help to pinpoint all potential binding locations.

Analysis and Visualization

Biopolymer's specialized analysis and visualization tools make it easy to access the information inherent in biomolecular structures. A special workflow dialog allows the user to prepare, analyze and resolve problems with protein structures. Biomolecules can be searched to locate secondary structure elements, standard torsions can be measured, and structures can be aligned by homology. An RMSD fit of specified sequences in two different structures can be performed. When multiple structures are available, for example from NMR-based structure determinations, regions that have the lowest conformational variability can be automatically located and fit.

Simplified biomolecular display options that highlight secondary structure



The methyl glycoside of the Lewis b human blood group determinant complexed with Lectin IV of *Griffonia simplicifolia*. The protein is represented as tubes color-coded by secondary structure type. The tetrasaccharides are shown as space-filled atoms.

elements include tubes and ribbons. Z-clipping and depth cuing make it possible to focus on regions of interest and are easily adjusted. Biopolymer works seamlessly with MOLCAD™ (licensed separately) to assess surface properties and display binding sites. MOLCAD creates electron density and Connolly solvent-accessible surfaces of molecules onto which it maps lipophilicity, electrostatic potential, hydrogen bonding sites, and other properties. Cavities, surfaces between molecules, and intramolecular channels can also be visualized.

Features

- Streamlined protein structure analysis and preparation dialog
- Simplified ribbon, tube, C α or backbone displays
- C α or backbone atom displays
- Align structures by homology
- Ramachandran analysis

Biopolymer Spreadsheet

Additional analysis tools are available in the molecular spreadsheet to evaluate the quality of protein structures derived by computational, NMR, X-ray or other methods. Within the Biopolymer Spreadsheet each row represents one residue of the structure under consideration. Columns are automatically calculated that describe the geometry and physical properties of each residue. Each residue is checked for non-standard bond angles, bond lengths, C α chirality, and peptide bond planarity¹. Side chain rotamers are compared to the most probable conformations determined by X-ray studies².

Other built-in residue attributes include energy, solvent accessible surface area, secondary structure, number of hydrogen

Complementary Software

- MOLCAD™ for visualizing surface features and physical properties essential for molecular recognition
- Advanced Protein Modeling for homolog finding, sequence alignment, and comparative protein modeling
- Surfex-Dock™ for receptor-based virtual screening and molecular docking
- RACHEL™ for *de-novo* ligand design

bonds, C α - C α distances, standard torsions (ϕ , ψ , χ , ω), transfer and solvation free energies, C α temperature factor, and the number of nearest neighbors. These built-in metrics are readily supplemented by user-defined residue attributes.

An automated analysis identifies residues that are in conformationally disallowed or borderline regions of the Ramachandran plot. Another tool locates polar residues that are buried, or non-polar residues that are exposed. Hydropathy plots can be based on any of several standard scales.

Biopolymer employs SYBYL's sophisticated graphics capabilities to reveal regions of a protein structure that may require further refinement. Structure, spreadsheet, and graphs are linked for locating and exploring correlations between attributes. Interesting or unusual residues can be colored-coded in a structure, and any value computed in the Molecular Spreadsheet can be visually mapped onto 3D tube or ribbon drawings by varying either the tube color or thickness.

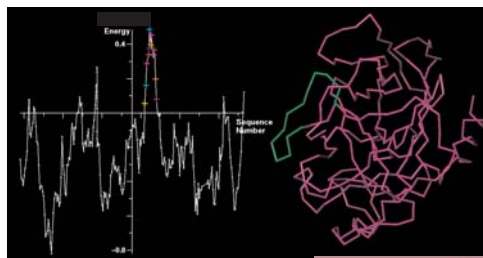
Features

- Extensive set of built-in geometric and physical attributes for evaluating the quality of protein structures
- Hydropathy profiling using any of several standard scales
- Automatic identification of buried/exposed residues
- Interactively linked structural display, molecular spreadsheet, and graphs to visualize location of anomalous structural elements
- 3D tube and ribbon renderings of protein structure on which properties can be mapped by thickness or color

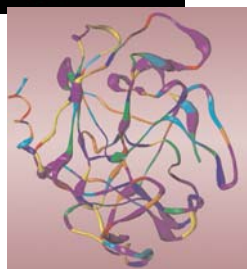
Binding Site Identification

Biopolymer provides tools leading to the identification of potential binding sites within or at the surface of biological targets.

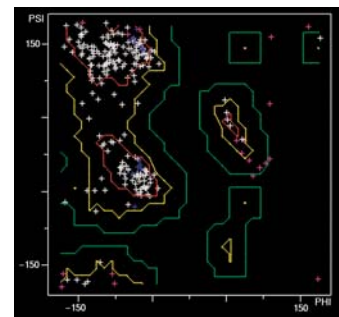
The site identification has two modes. In the first, protein pockets are rapidly identified by solvating the structure to locate regions where solvent-spheres tend to cluster. In the second mode, a detailed analysis of protein structure is performed



The thickness of a rendered tube shows the residue-by-residue energy of the model. The tube's colored surface indicates the probability of each residue's side chain conformation occurring in X-ray structures (purple and blue are low probability, orange and red are high probability).



These images show the results of a Biopolymer Spreadsheet evaluation of a theoretical model of prostate-specific antigen (2PSA). A plot of residue-based energies reveals a loop (highlighted in green) that may require further refinement.



An automated Ramachandran analysis identifies backbone torsions in borderline or disallowed regions.

based on properties correlated with likely hot-spots for binding. A molecular spreadsheet is created automatically containing every atom in a biomolecule, and for each of these atoms logs the solvent exposure and distance from the center of the biomolecule. The hydrogen-bonding character of each atom can also be determined. Additional properties can be calculated for each atom, such as hydrophobicity and local surface curvature, using separately licensed MOLCAD.

Flexible selection tools within Biopolymer make it possible to identify potential binding sites by mixing and matching key criteria such as depth, exposure, temperature factor, and hydrogen-bonding capability. For example, one can highlight the hydrophobic portions of the surface farthest from the center of a protein.

These oily surface patches could indicate a potential protein-protein interaction site. Alternatively, the solvent accessible atoms with the highest curvature might provide evidence for a ligand-binding pocket. Based on the distance between the selected atoms, a hierarchical clustering tool groups them into individual pockets or patches.

Whatever the selection criteria, Biopolymer highlights selected atoms directly on the protein. When used with MOLCAD, Biopolymer generates a surface representation of the binding site onto which properties such as hydrophobicity and local curvature can be color-mapped.

Features

- Flood-fill-based visual identification of protein pockets
- Calculation of atom-based solvent accessibility for detection of deep pockets
- Hierarchical clustering that groups atoms with similar properties into pockets or patches
- Color-coded highlighting of atoms matching surface criteria

Features (with MOLCAD)

- Calculation of hydrophobicity and local curvature
- Generation of binding-site surfaces color-coded by property

Model Building

Biopolymer uses a residue-based approach to simplify construction of proteins, polysaccharides, and nucleic acids. A sequence of residues is entered or selected from the menu, conformational states are set, and the resulting model structure is instantly available for investigation. Dictionaries support Biopolymer's residue-based approach. The large number of built-in residues includes non-standard amino acids and is easily augmented with custom residues and blocking groups. A special dialog interface simplifies creation of new residues and entry of key information

maintained with each residue, such as definitions of torsion angles and conformational states.

Residues can be mutated, deleted, and inserted, and biopolymer chains can be broken or joined. Proteins can be cyclized through the backbone or cross-linked via disulfide bridges. When the three-dimensional structure of a protein is unknown, Biopolymer offers sequence alignment tools to investigate whether two or more sequences are related, either structurally or functionally. The Needleman-Wunsch³ algorithm is provided for automatically aligning sequences, and is more discriminating than the local alignment techniques employed by BLAST and FASTA. The complete comparative modeling workflow using the latest scientific techniques is supported by Advanced Protein Modeling (licensed separately).

Features

- Build complete protein backbone from $C\alpha$ trace
- Mutate, insert or delete residues
- Automatic side chain construction
- Secondary structure prediction methods
- Automatically set standard torsions and conformational states
- Multiple sequence alignment
- Sequence editor with customizable annotation schemes, interactively linked to the 3D display

Structure Refinement

Biopolymer includes both knowledge-based and energy-based methods for refining model structures. The protein database within Biopolymer is a compilation of the highest resolution structures from the Protein Data Bank⁴. This database can be searched for patterns of residues, secondary structures, sequences, or $C\alpha$ distances in order to refine structural models. For example, if a protein loop is not resolved by experimental techniques or modeling, candidate loop conformations can be obtained from the database using

Biopolymer's loop search capabilities.

Random tweak⁵ is an alternative method for building loops that generates random phi and psi angles for a polypeptide fragment. The phi and psi angles are modified to satisfy distance constraints derived from the structural gap.

Analysis tools within the molecular spreadsheet allow candidate loops to be compared and prioritized on the basis of sequence similarity, geometric fit, torsion and distance measurements, and steric interactions with the protein. Interactive displays link spreadsheet data, graphs, and structural models to facilitate interpretation of results.

Biopolymer and SYBYL together provide accurate energy-based modeling of protein or nucleic acid complexes as well as small organic substrates. A selection of force fields is available for geometry optimization and molecular dynamics, whether for large biomolecular assemblies, for small molecules, or for some combination of these structures. Validated force fields include Tripos⁶, MMFF94⁷, and AMBER⁸. Special routines make it possible to optimize strained or poorly resolved sections of biopolymers such as surface loops while preserving high quality regions of the models.

Features

- Validated Tripos, MMFF94, and AMBER force fields for energy refinement
- Customizable database of high-quality protein structures for searching
- Loop search and random tweak for building protein loops

Hardware and Software Requirements

Biopolymer is accessible from SYBYL as well as from the system command line. SYBYL and Biopolymer run on workstations operating under IRIX[®] (SGI[®]) or Linux[®] (x86).

References

1. Weiner, S.J.; Kollman, P.A.; Case, D.A.; Singh, U.C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. "A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins." *J. Am. Chem. Soc.* 1984, 106, 765-784.
2. Schrauber, H.; Eisenhaber, F.; Argos, P. "Rotamers: To Be or Not To Be? An Analysis of Amino Acid Side-Chain Conformations in Globular Proteins." *J. Mol. Biol.* 1993, 230, 592-612.
3. Needleman, S.B.; Wunsch, C.D. "A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins." *J. Mol. Biol.* 1970, 48, 443.
4. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. "The Protein Data Bank." *Nucleic Acids Res.* 2000, 28, 235. <http://www.rcsb.org/pdb/>
5. Fine, R.M.; Wang, H.; Shenkin, P.S.; Yarmush, D.L.; Levinthal, C. "Predicting Antibody Hypervariable Loop Conformations. II: Minimization And Molecular Dynamics Studies Of MCPC603 From Many Randomly Generated Loop Conformations Proteins." 1986, 1, 342.
6. Clark, M.; Cramer III, R.D.; Van Opdenbosch, N. "Validation of the General Purpose Tripos 5.2 Force Field." *J. Comp. Chem.* 1989, 10, 982.
7. i) Halgren, T. "Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94." *J. Comput. Chem.* 1996, 17, 720.
ii) Halgren, T. "MMFF VI. MMFF94S Option for Energy Minimization Studies." *J. Comp. Chem.* 1999, 20, 720-729.
8. Cornell, W.D.; Cieplak, P.; Bayly, C.I.; Gould, I.R.; Merz Jr., K.M.; Ferguson, D.M.; Spellmeyer, D.C.; Xov, T.; Caldwell, J.W.; Kollman, P.A. "A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules." *J. Am. Chem. Soc.* 1995, 117, 5179.



| WWW.TRIPOS.COM | | | | CONTACT_US@TRIPOS.COM | | |
|--|--------------------------|-----------------------------|-----------------------------|--------------------------|------------------------------------|--|
| AUSTRALIA +61 (7) 5439 9775 | CANADA +1 450 4334500 | FRANCE +33 1 69 59 29 49 | GERMANY +49 89 45 10 300 | JAPAN +81 3 5166 1721 | UNITED KINGDOM +44 1 908 650000 | UNITED STATES 800-323-2960 +1-314-647-1099 |
| ©Tripos, Inc. 2006 All rights reserved. Printed in USA All trademarks are the property of their respective owners. | | | | | | |