

# PROTEIN MODELING AND SEQUENCE ANALYSIS IN DISCOVERY STUDIO

**Harnessing the Power of Protein Modeling.** Experimental structure determination involves difficult methods that require a significant amount of expertise and resources, and can take years to yield results. Conversely, protein modeling enables access to sensible structural models within a matter of hours, or even minutes. Template-based modeling has been shown to yield useful models in countless publications. As more and more experimental structures become available, template based protein modeling will continue to provide a greater scope of understanding to researchers, delivering accurate models quickly and efficiently. Protein Modeling and Sequence Analysis solutions in Discovery Studio provide the necessary set of tools for the construction of molecular structures, as well as macromolecular docking, in an easy to use customizable graphical user interface suitable for novice and expert users.

## ACCELRYS SCIENCE

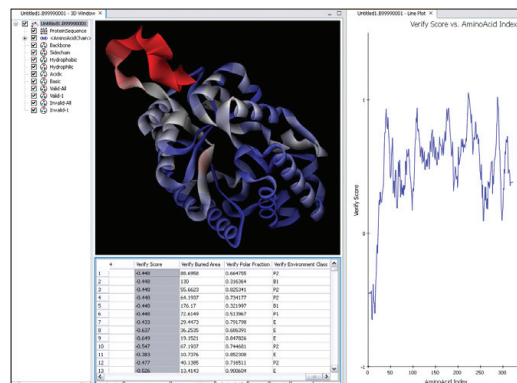
### Search for homologs using sequence similarity.

Elucidate biological function by searching for homologs using sequence similarity methods with DS Sequence Analysis.

- New functionality in the area of Antibody Modeling: a pre-compiled CDR information data file is used to automate the process of CDR identification and annotation. A sequence alignment file of the best aligned hits is provided for automated loop grafting of the CDR regions
- Access to the routinely used BLAST and PSI-BLAST algorithms, to search real time databases installed behind your firewall or at the NCBI
- Supporting tools that perform Phylogenetic and Evolutionary Trace analysis help you create dendrograms using hierarchical clustering methods

### GENERATE HIGH QUALITY HOMOLOGY MODELS QUICKLY AND ACCURATELY.

Use DS MODELER to quickly construct high quality alignments and build a collection of homology models for further analysis. Further your research with modern molecular modeling in silico based



*A xylanases hydroxylases protein model created using the build model protocol from DS MODELER from 3 protein structure templates (1b30, 1gom and 1ta3) was verified using the Profile-3D verify method. Blue (high verify scores) and red (low verify scores) residues in the protein ribbon display correspond to valid and invalid regions, respectively.*

methods by using DS MODELER in conjunction with simulations and structure based design tools in Discovery Studio.

- The gold-standard engine for automatically generating high quality homology models using spatial restraints.
- A new method for improving the sequence alignment in low homology cases by using the SALIGN method which uses sequence profile information.
- Loop modeling using the DOPE energy function. The DOPE function represents an improved energy function from potentials extracted from on a library of non-redundant high resolution crystal structures, and has been shown to provide higher quality models
- Inclusion of important ligand information during the homology modeling building process
- Creation of mutants for site directed mutagenesis studies

## TEST THE VALIDITY OF A PROTEIN MODEL

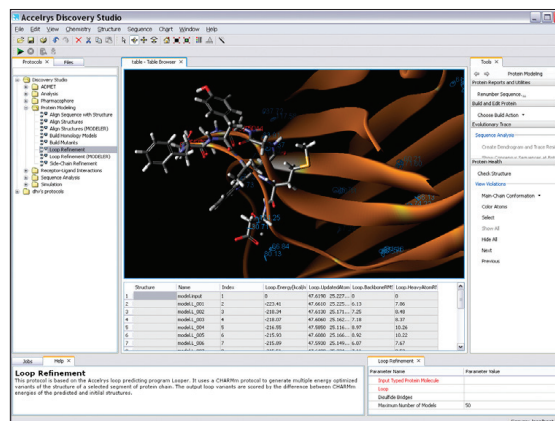
Use DS Protein Health to test the validity of a protein structure (full or partial) derived from modeling studies or experimental methods.

- Profiles-3D Verify: a method that evaluates the likelihood that an amino acid should be present within its current environment
- An interactive tool kit, that allows you to browse and correct a suggested list of structural violations which are mapped and colored to the 3D structure

## PERFORM SIDE CHAIN OR LOOP REFINEMENT ON YOUR STRUCTURE

Perform accurate CHARMM-based structural refinement of loops and side chains with DS Protein Refine.\*

- LOOPER: This algorithm has been optimized for loop refinement; it will quickly generate energy optimized variants of the structure, and provide the user with a list of proposed loop conformations that have been scored using the CHARMM Energy function
- ChiRotor: an algorithm designed to refine side-chain orientations by performing a systematic search of side-chain conformations and CHARMM energy minimizations



Loop refinement (protocol) and protein validation (Protein Health tool panel) was performed on a protein model generated from a sequence with unknown structure. The stick representation of the best loop is shown above.

## GAIN BETTER UNDERSTANDING OF THE MECHANISM OF ACTION OF A PROTEIN

Gain new insights about the mechanism of action by looking at the molecular structure of a protein, analyzing sequence conservation patterns within a family of proteins, and mapping conserved residues to the 3D structure with DS Protein Families

- Access to the Align123 program, which is based on the CLUSTAL W technology for sequence alignments
- Use of Accelrys' proprietary 3DMA algorithm for accurate structural alignments
- Additional tools like the Phylogenetic and Evolutionary Trace analysis are also accessible to help determine structural conservation of amino acids

## MODEL PROTEIN-PROTEIN INTERACTIONS:

Predict protein-protein structure interactions of novel targets rapidly and accurately with DS Protein Docking.

- Use of ZDOCK for rigid body docking
- Access to the CHARMM based RDOCK refinement program for optimization and enhanced scoring of docked poses
- Use of clustering methods to narrow the search and to help identify poses of interest that can be used for further refinement

\*Both methods are considered de novo methods and have no dependency on the starting conformation of the structure

## CONSTRUCT MOLECULAR FRAGMENTS AND PERFORM ELECTROSTATIC ANALYSIS INCLUDING PK PREDICTIONS

Construct simple molecular models with a robust set of tools. Whether it is, amino acids (proteins), nucleic acids (DNA/RNA), or any other organic compound, access all the tools you need with DS Biopolymer

- Easy-to-use interactive builders to construct molecular models quickly and accurately.
- Modeling tools commonly used in X-ray crystallography work such as real space fitting of molecules to electron-density maps, force-field based regularization of models, and more.
- The DELPHI program for calculating the electrostatic distribution of charges for macromolecules.
- A brand new algorithm which estimates the protonation state of titratable amino acids within the protein quickly and accurately. This method is based on the Generalized Born model for charge estimation, and accurately predicts pK's, pH titration curves, and overall energy of folding. This new tool has been long overdue and will certainly enable users to study macromolecular modeling in a brand new way.

## THE GOLD STANDARD IN TECHNOLOGY

**Comprehensive** – From protein sequence to well-defined docked complex, Discovery Studio includes a comprehensive suite of tools to analyze your data and provide you with reliable results. Beyond protein modeling, you can generate pharmacophore models, dock and score ligands, perform QSAR analysis, and more within a single easy-to-use environment.

**Proven history** – The core technology has undergone over a dozen years of continuous innovation and customer driven improvement, and has demonstrated dependable performance in the pharmaceutical industry with over 100 publications.

**Cutting edge** – Accelrys is incorporating new scientific tools to meet current pharmaceutical needs and we are continuously working with our customers to plan for future innovation.

**Easy-to-use interface** – DS 2.0 provides a powerful and intuitive user interface. DS 2.0 can be deployed either in a complete standalone solution for individual modelers or as part of an enterprise-level client server installation for easier protocol sharing and administration in larger modeling groups.

**Integrated solution** – The DS 2.0 environment, built on the Pipeline Pilot open operating platform, integrates protein modeling, pharmacophore analysis, and virtual screening as well as third-Discovery Studio party applications for an infinitely extensible virtual discovery platform. Well-tested applications including CHARMM, MODELER, Catalyst, and others are accessible in the graphical DS environment, the Pipeline Pilot scripting and protocol development environment, and from command-line prompts.

**Parallel Computing** – The DS 2.0 platform is optimized to take advantage of grid and cluster computing as well as multi-core processors to rapidly process large tasks.

## ACCELRYS IS YOUR PARTNER IN RESEARCH

**User community** – Numerous protein prediction meetings such as CASP, Protein Society, and CAPRI attract hundreds of researchers from all over the world to evaluate novel methods in protein modeling and protein structure prediction.

**Scientific Consulting** – Accelrys has dozens of experienced Ph.D.s with expertise in implementing scientific solutions for drug design that are available for short or long-term engagements to create tailored solutions or perform modeling experiments.

**Customer Support** – Accelrys customers report a 98% satisfaction rate with our support team.

**Committed to innovation** – With over 100 Ph.D.s in the field working daily with researchers in industry and academia, Accelrys is committed to delivering cutting-edge technology to our customers.

**World leading scientific advisors** – Through our in-licensing agreements, partnerships, and scientific advisors, many of the world's foremost experts in computational drug design are involved in setting our direction.

## BIOLOGICAL VALIDATION AND COMPARISON

**2004** – SALIGN validation has shown much better correlation of sequence to structure alignments than traditional sequence alignment methods.

**2007** – ZDOCK and RDOCK have been highly successful at the world renowned CAPRI (Critical Assessment of Predicted Interactions <http://capri.ebi.ac.uk/>) meetings, and having these tools in Discovery Studio make protein-protein docking a manageable task.

**2007** – A quick validation of LOOPER within DS Protein Refine shows an impressive 0.44Å back-bone RMSD in the de novo modeling of a 7-residue loop (residues 55 to 61) of the commonly used example, thymidine kinase (PDB ID: 1kim). Such a low RMSD demonstrates a very high degree of accuracy for our methods and borders experimental error.

To learn more about Discovery Studio, go to [accelrys.com/discovery-studio](http://accelrys.com/discovery-studio)

## REFERENCES:

1. Brooks, B.R., Bruccoleri, R.E.,Olafson, B.D., States, D.J., Swaminathan, S., and Karplus, M., "CHARMM: A program for macromolecular energy, minimization, and dynamics calculations," J. Comput. Chem., **1983**, 4, 187-217.
2. Chen R, Li L, Weng Z, ZDOCK: An Initial-stage Protein-Docking Algorithm. Proteins, **2003**, 52, 80 – 87
3. Eswar, N., Eramian, D., Webb, B., Shen, M., Sali. A. Protein Structure Modeling With MODELLER. Current Protocols in Bioinformatics John Wiley & Sons, Inc., **2006**, Supplement 15, 5.6.1-5.6.30
4. Li L, Chen R (joint first authors), Weng Z, RDOCK: Refinement of Rigid-body Protein Docking Predictions. Proteins, **2003**, 53, 693–707
5. Marti-Renom, M. A., Madhusudhan, M. S. and Sali, A., "Alignment of protein sequences by their profiles", Protein Science, **2004**, 13, 1071–1087
6. Morea, V., Lesk, A., and Tramontano, A. "Antibody Modeling: Implications for Engineering and Design," METHODS, **2000**, 20, 267.
7. Sali, A., Pottertone, L., Yuan, F., van Vlijmen, H., and Karplus, M., "Evaluation of comparative protein modeling by MODELLER," Proteins, **1995**, 23, 318.
8. Spassov, V., Yan, L. and Flook, P. "The Dominant Role of Side-chain Backbone Interactions in Structural Realization of Amino-acid Code. ChiRotor: a Side-chain Prediction Algorithm Based on Side-chain Backbone Interactions," Protein Science, **2007**
9. Spassov, V., Yan, L., and Flook, P., "LOOPER: A Molecular Mechanics Based Algorithm for Protein Loop Prediction." Manuscript in preparation.