

MACROMOLECULE MODELING WITH DISCOVERY STUDIO®

Determining the three-dimensional structure and properties of a macromolecule, such as enzymes, receptors, antibodies, DNA or RNA is a fundamental component to a wide range of research activities. For example, predicting the location and characteristics of small molecule binding sites or optimizing the stability and selectivity of therapeutic biologics, all require access to precise, accurate molecular models. Discovery Studio delivers a comprehensive portfolio of market leading, validated scientific tools, able to assist in every aspect of macromolecule-based research.

MACROMOLECULE-BASED DESIGN

With Discovery Studio, you can rapidly and easily:

- Generate 3D structure models
 - Analyze and clean-up models from 3D structure repositories (e.g., PDB)
 - Build models directly from either sequence information or X-ray diffraction data
 - Verify the quality of a structure model
- Study macromolecule interactions with their partners
 - Perform protein-protein docking to predict protein binding partners
- Conduct forcefield based simulations
 - Predict protein ionization and protonate residues at a specified pH
 - Perform macromolecules simulations using explicit solvation models
 - Perform macromolecules simulations using implicit solvation or implicit membrane models
- Predict effects of amino acid mutations on macromolecule properties
 - Impact on binding affinity
 - Thermal stability
 - Protein aggregation

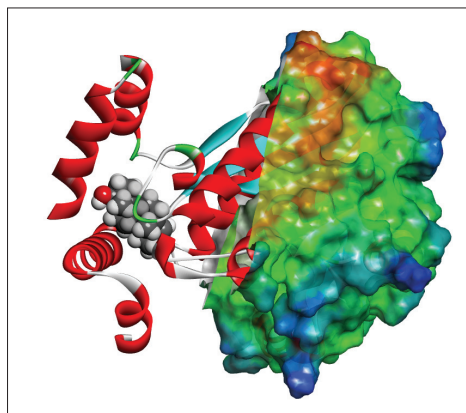


Figure 1: Example of 17beta-Hydroxysteroid Dehydrogenase Type 1 complexed with Testosterone [PDB: 1JTV], showing ribbon structure cartoon and hydrophobic surface.

MACROMOLECULES: FROM STRUCTURE DATABASES

With thousands of macromolecule structures now available both publicly and from in-house structure databases, a model may already have been deposited:

- **Query the RCSB database** directly, either via simple or advanced criteria, including ID, substructures or sequence motifs
- **Generate protein reports** to summarize information and potential problems with a retrieved structure
- **Clean Protein** either automatically, or manually:
 - Includes: standardize atom names, remove alternate conformations, insert missing main-chain or side-chain atoms
 - Optional: Remove waters, break bonds between metals and protein atoms
- **Build missing loops** using the PDB SEQRES data and optimize their conformations
- **Optimize side-chain conformations** for missing residues using CHARMM simulations¹

MACROMOLECULES: FROM SEQUENCE

An experimentally-derived structure of a macromolecule is not always available. However, it is often possible to derive a model from closely related protein homologs:

- **Template Identification**
 - **Search for templates** in the PDB using Blast or PSI-BLAST, to identify optimal templates for the target macromolecule
 - Refine template selection, by specifying and filtering by particular species
- **Align sequences quickly and accurately** with templates using multiple sequence-alignment algorithms
 - Use Align123² to align a set of sequences to existing alignment profile information
 - Use SALIGN³ to align two profiles, or to improve sequence alignments to low homology regions.
 - Predict the transmembrane helices in transmembrane proteins from sequence
 - Phylogenetic and Evolutionary Trace analysis tools are also available to determine relationships between sequences⁴ and structural conservation of amino acids

• **Superimpose proteins structures** using either ranges of residues, sequence alignment, or using C-alpha pairs.

– Even superimpose a large set of proteins from files

• Build homology models

Use the industry-standard MODELER^{5,6,7} to build homology models of target proteins automatically

– Use the LOOPER algorithm⁸ to systematically search main chain poses and rank using CHARMM^{9,10}

– Alternatively, manually copy loop conformations from a template structure onto a target model.

– Systematically search for side-chain conformation using CHARMM simulations¹

MACROMOLECULES: FROM X-RAY

Many macromolecule structures can now be sourced directly using X-ray crystallography. Based on CNX (Crystallography and NMR Explorer), a suite of refinement tools are available:

- **Generate electron density maps** from a molecular structure and its corresponding X-ray reflection data.
- **Perform full refinement** of a model structure with rigid-body minimization, simulated annealing, coordinate minimization, occupancy minimization, or B-factor minimization
- **Use HT-X PIPE** to run an automated high throughput structure determination of protein-ligand complexes

MACROMOLECULES: FROM NUCLEIC ACIDS

Rapidly, create single, double or triple-stranded DNA molecules in A-, B-, or Z-form using standard helix parameters:

- **RNA and DNA-RNA** hybrid molecules can be generated in the A-form in either single- or double-strand forms
- **Modify the model further** by toggling termini between capped and primed forms, ligate nucleic acid molecules together, or modify the sugar moieties

MACROMOLECULES: MODEL VERIFICATION

Discovery Studio provides a suite of essential tools to verify the quality of a protein model.

- **Verify Protein (Profiles 3D)** calculates the likelihood of each residue to be found in its specific local 3D environment
- **Verify Protein (MODELER)** scores the model conformation using a statistical potential
- **Ramachandran plots** to verify the distribution of Phi and Psi angles of amino acids
- Capabilities to look at main-chain angular conformation, side-chain deviations from rotamer libraries (e.g., Ponder and Richards, Sutcliffe, and more)

FORCEFIELD-BASED MACROMOLECULE MODELING

Perform a range of model refinements including side chain optimization, minimization and detailed simulations.

- **Predict protein ionization and residue pKs**
Using a novel in-house computational method^{11,12} quickly and accurately calculate protein ionization with a CHARMM Generalized-Born (GB) solvent model^{13,14,15}
 - Predicts the pK and titration curves for each of the titratable amino acid residues
 - Protonates the residues at a given pH according to the predicted pK values
- **Simulate macromolecule structures**
 - Perform either explicit solvent or implicit solvent-based Molecular Mechanics (MM) simulations
 - Perform implicit solvent-based Molecular Dynamics (MD) simulations using CHARMM
 - Alternatively, launch a NAMD¹⁶ calculation and perform MD simulations* with explicit waters
 - Add an implicit membrane to a protein structure¹⁵ to help refine models during forcefield-based simulations
 - Calculate single point energies or perform minimizations of receptor-ligand complexes using hybrid Quantum Mechanics (QM)/MM simulations

*NAMD is available from the University of Illinois at Urbana-Champaign. <http://www.ks.uiuc.edu/Research/namd/>

PROTEIN-PROTEIN DOCKING

Predict protein-protein structure interactions of novel targets quickly and accurately:

- **Use ZDOCK**^{17,18} to comprehensively search protein-protein interaction patterns and output putative docking poses
- **Use ZRANK scoring function** to increase the accuracy of docked poses¹⁹

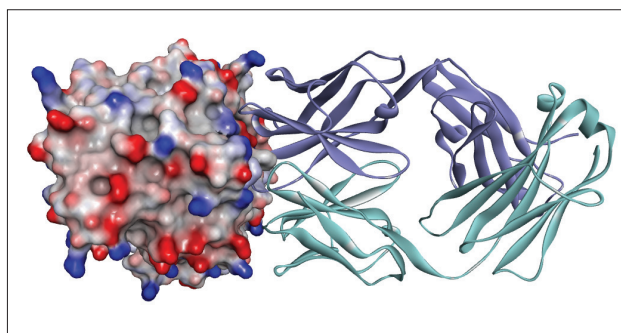


Figure 2: Example protein-protein binding interaction between the Fab domain of an antibody and its antigen [PDB: 3PGF].

PROTEIN DESIGN

- **Predict binding affinity**
 - Perform amino-acid scanning mutagenesis to evaluate the effect of single-point mutations on the binding affinity of molecular partners
 - Development based on a selection²⁰ of published experimental $\Delta\Delta G_{\text{bind}}$ energies for 380 single point mutations across 19 proteins

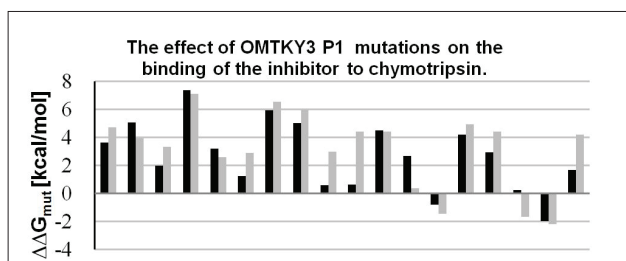


Figure 3: Calculated (in black) vs. experimental (in grey) mutation energies in the binding of turkey ovomucoid inhibitor to chymotrypsin for residue Leu18, to 18^{21,22} other amino acid types.

• Optimize protein stability

- Calculate the effect of a single-point mutation on protein stability at a given temperature
- Predict the most stabilizing two- or three-site mutations in a protein structure, among a list of specified mutations

• Spatial aggregation propensity algorithm

Use the experimentally validated spatial aggregation propensity algorithm^{23,24,25}, to predict putative sites contributing to the aggregation of proteins or regions prone to protein-protein binding²⁶

- The algorithm is licensed from the Massachusetts Institute of Technology and was developed at Prof. Trout's laboratory

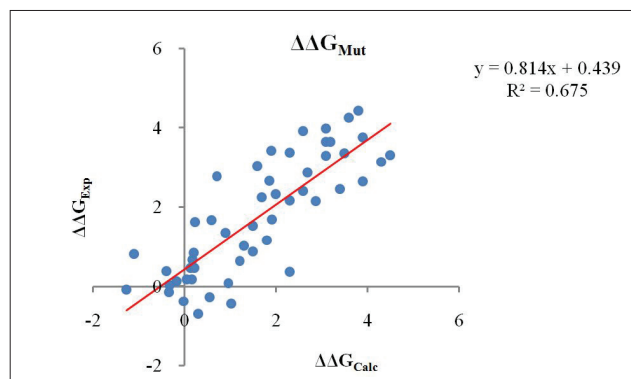


Figure 4: Example of the prediction of effect on protein stability by introducing a mutation to T4 Lysozyme²⁷.

To learn more about Discovery Studio, go to accelrys.com/discovery-studio

REFERENCES:

1. Spassov, V., Yan, L., Flook, P. *Protein Science*, **2007**, *16*(3), 494-506.
2. Thompson J., Higgins D. G., Gibson T. J. *Nucl. Acids Res.*, **1994**, *22*(22), 4673-4680.
3. Marti-Renom M.A., Madhusudhan M.S., Sali, A. *Protein Science*, **2004**, *13*, 1071-1087.
4. Lichtarge O., Bourne H.R., Cohen F.E. *J Mol Biol.*, **1996**, *257*(2), 342-358.
5. N. Eswar, M. A. Marti-Renom, B. Webb, M. S. Madhusudhan, D. Eramian, M. Shen, U. Pieper, A. Sali. Comparative Protein Structure Modeling With MODELLER. *Current Protocols in Bioinformatics*, John Wiley & Sons, Inc., **2006**, *Supplement 15*, 5.6.1-5.6.30.
6. Fiser, A. Sali, A. *Methods in Enzymology*, **2003**, *374*, 463-493.
7. Sanchez, R. Sali, A. *Protein Structure Prediction: Methods and Protocols*, **2000**, 97-129.
8. Spassov, V.Z., Flook, P.K., Yan, L. *Prot. Eng., Design & Selection*, **2008**, *21*, 91-100.
9. Brooks B. R., Brooks III C. L., Mackerell A. D., Nilsson L., Petrella R. J., Roux B., Won Y., Archontis G., Bartels C., Boresch S., Caflich A., Caves L., Cui Q., Dinner A. R., Feig M., Fischer S., Gao J., Hodoscek M., Im W., Kuczera K., Lazaridis T., Ma J., Ovchinnikov V., Paci E., Pastor R. W., Post C. B., Pu J. Z., Schaefer M., Tidor B., Venable R. M., Woodcock H. L., Wu X., Yang W., York D. M., Karplus M. *J. Comp. Chem.*, **2009**, *30*, 1545-1615.
10. Brooks B. R., Brucoleri R. E., Olafson B. D., States D. J., Swaminathan S., Karplus M. *J. Comp. Chem.*, **1983**, *4*, 187-217.
11. Spassov V.Z., Yan L. *Protein Science*, **2008**, *17*, 1955-1970.
12. Spassov V.Z., Bashford D. *J. Comp.Chem.*, **1999**, *20*, 1091-1111.
13. Still, W. C., Tempczyk, A., Hawley, R. C., Hendrickson, T., *J. Amer. Chem. Soc.*, **1990**, *112*, 6127.
14. Dominy, B., Brooks, C.L. III., *J. Phys. Chem. B*, **1999**, *103*, 3765-3773.
15. Spassov, V.Z., Yan, L., Szalma, S. *J. Phys. Chem. B*, **2002**, *106*, 8726-8738.
16. Phillips J.C., Braun R., Wang W., Gumbart J., Tajkhorshid E., Villa E., Chipot C., Skeel R.D., Kale L., Schulten K. *J. Comp. Chem.*, **2005**, *26*, 1781-1802.
17. Pierce B., Weng Z. *Proteins*, **2008**, *72*(1), 270-279.
18. Chen R., Weng Z. *Proteins*, **2003**, *52*, 80-87.
19. Pierce B., Weng Z. *Proteins*, **2007**, *67*(4), 1078-1086.
20. Kortemme T, Baker, D. *PNAS*, **2002**, *99*, 14116-14121.
21. Krowarsch D., Dadlez M., Buczek O., Krokoszynska I., Smalas A.O. Otlewski J., *J. Mol. Biol.*, **1999**, *289*, 175-186.
22. Spassov V.Z., Yan L., *In preparation*, **2011**.
23. Chennamsetty N., Voynov V., Kayser V., Helk B. and Trout B. L. *J. Phys. Chem. B*, **2010**, *114*(19), 6614-6624.
24. Chennamsetty N., Voynov V., Kayser V., Helk B. and Trout B. L., *Proc. Nat. Acad. Sci.*, **2010**, *106*(29), 11937-11942.
25. Chennamsetty N., Helk B., Trout B. L., Voynov V., Kayser V. PCT/US2009/047954. Filed 19th June, **2009**.
26. Chennamsetty N., Voynov V., Kayser K., Helk B., Trout B.L. *Proteins*, **2011**, *79*, 888-897.
27. Gromiha M. M., An J., Kono H., Oobatake M., Uedaira H., Prabakaran P., Sarai A., *Nucl. Acids Res.*, **2000**, *8*, 283-285.