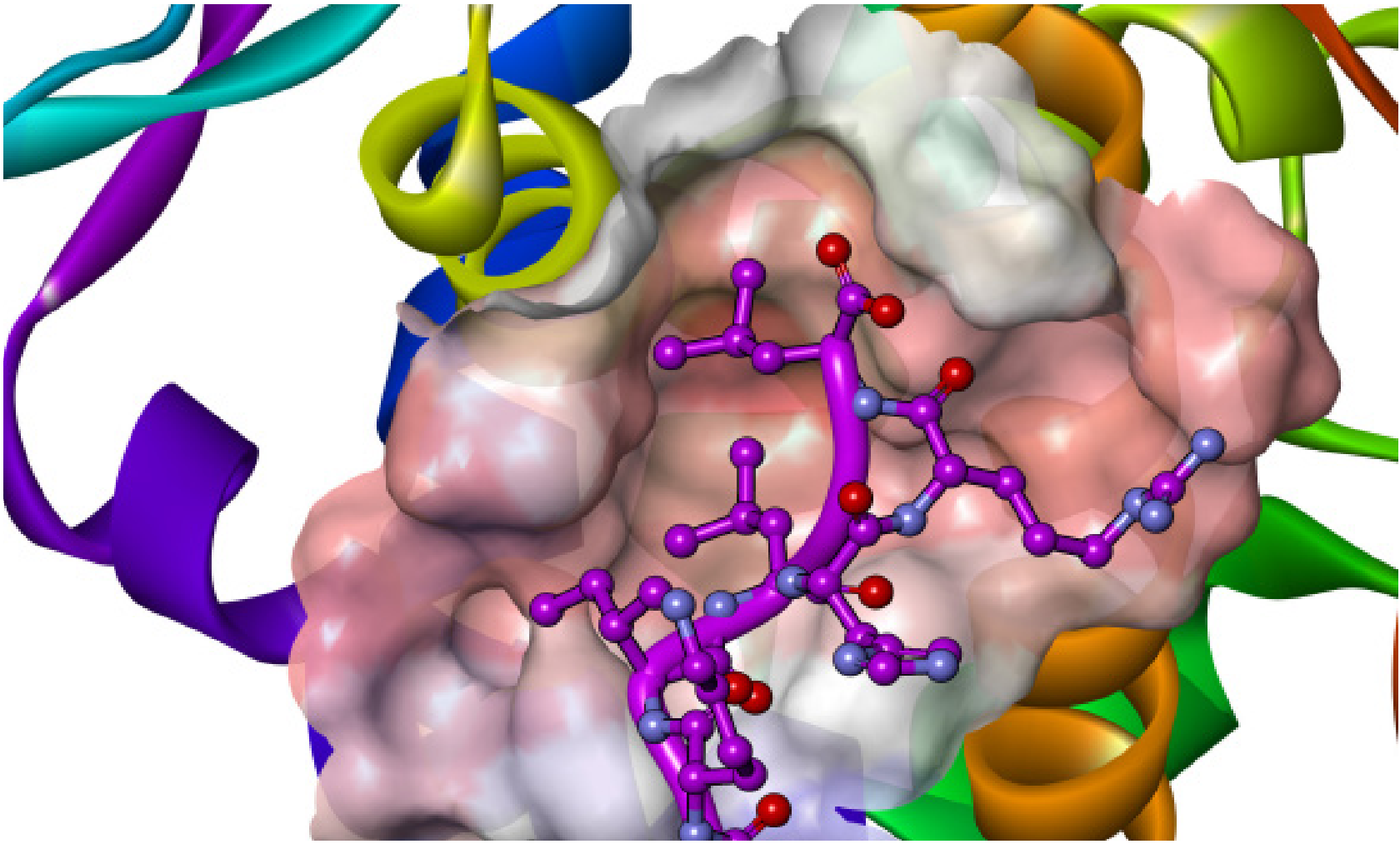
DASSAULT SYSTEMES

BIOVIA

MACROMOLECULE MODELING WITH BIOVIA DISCOVERY STUDIO®

DATASHEET



COMPREHENSIVE PROTEIN MODELING

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. BIOVIA Discovery Studio delivers a comprehensive portfolio of market leading, validated scientific tools able to assist in every aspect of macromolecule-based research.

MACROMOLECULES: FROM STRUCTURE DATABASES

With thousands of macromolecule structures now resolved experimentally, a model may already have been deposited:

- Query the RCSB database directly, either via ID, substructures or sequence motifs
- Generate protein reports to summarize information and identify potential problems from a retrieved structure
- Clean Protein either automatically, or manually
  - Capabilities: standardize atom names, select alternate conformations, insert missing main-chain or side-chain atoms, adjust terminal residues, and more
- Build missing loops using the PDB SEQRES data and optimize their conformations
- Optimize side-chain conformations for missing side-chain atoms ChiRotor CHARMm simulations1
- Superimpose proteins structures using either ranges of residues, sequence alignment, or using C-alpha pairs
  - Even superimpose a large set of proteins from files

MACROMOLECULES: FROM SEQUENCE

If an experimentally-derived structure is not available, it is often possible to derive a model from closely related homologs:

- Template Identification
  - Search the PDB using BLAST or PSI-BLAST, to identify optimal template models based on sequence similarity
  - Refine template selections by species
- Align sequences quickly and accurately with templates using multiple sequence-alignment algorithms
- Use structure alignment to align template sequences
- Use Align123[2] to align model sequence to the structure profile
  - Include secondary structure matching when aligning sequences using Align123[2]
- Annotate and analyze sequences:
  - Predict the transmembrane helices in membrane proteins
  - Detect antibody variable, constant domains and CDR loops Use phylogenetic and Evolutionary Trace analysis tools to determine relationships between sequences and structural conservation of amino acids

Use phylogenetic and Evolutionary Trace analysis tools to determine relationships between sequences and structural conservation of amino acids:

- Use the industry-standard MODELER[4,5,6] to automatically build homology models of target proteins and refine the models using CHARMm based methods
- Build model with multiple templates, copy ligands and important waters
- Build model with user specified constraints to keep symmetry between different monomers
- Use the LOOPER algorithm[7] to systematically search loop conformations and rank using CHARMm[8,9]
- Graft loop conformations from a template structure onto a target model
- Systematically search for side-chain conformation using CHARMm simulations[1]

MACROMOLECULES: FROM X-RAY

Many macromolecule structures can now be solved 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 and 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, ligating nucleic acid molecules, or modifying sugar moieties

MACROMOLECULES: MODEL VERIFICATION

BIOVIA 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 environment
- Verify Protein (MODELER) scores the model conformation using a statistical potential function
- Ramachandran plots draws the distribution of Phi and Psi angles of amino acids based on statistical likelihood
- Inspect Internal Coordinates: main-chain torsions, 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
- Quickly and accurately calculate protein ionization[10,11] with a CHARMm Generalized-Born (GB) solvent model[12,13,14]
  - Predict isoelectric point of a protein
  - Predict pK values and titration curves for every titratable amino acid residue
  - Protonate the residues at a given pH according to the predicted pK values and optimize hydrogen positions

Simulate macromolecule structures:

- Perform either implicit solvent- or explicit solvent-based Molecular Dynamics (MD) simulations using CHARMm c41b1
- Launch a NAMD[15] calculation and perform MD simulations with explicit waters
- Analyze simulations, plot temperature or energy versus time, or calculate RMSD or RMSF for all or selected frames
- Add an implicit membrane to a protein structure[14] to represent membrane bound models in simulations
- Compute single point energies or minimize receptor-ligand complexes using hybrid Quantum Mechanics (QM)/MM

PROTEIN-PROTEIN DOCKING

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

- Use ZDOCK[16,17] to comprehensively search protein-protein interaction patterns and putative docking poses
- Cluster poses based on their spatial proximity and filter poses based on known interface residues
- Use ZRANK scoring function to increase the accuracy of docked poses[18]
- Analyze protein binding interfaces and generate reports for different types of interactions

PROTEIN DESIGN

- Identify mutation sites for disulfide bridges enables molecular biologists and antibody engineers to improve the stability of a novel biologic
  - Browse sites and zoom in to review predicted bridges
  - Convert a predicted site into a disulfide bridge
- Identify sequence motifs associated with possible post-translational modification sites in biotherapeutics
  - Motifs include: Aspartic Acid isomerization, Cleavage (D), Cleavage (N), N-Glycosylation, O-Glycosylation, C-Glycosylation, Deamidation, Glycation (Lys), Hydroxylation, Oxidation (MW) and Pyroglutamate
  - Additionally, calculate biophysical properties, such as isoelectric points, molecular charge, molar extinction coefficient, hydropathy and Antigenic sites

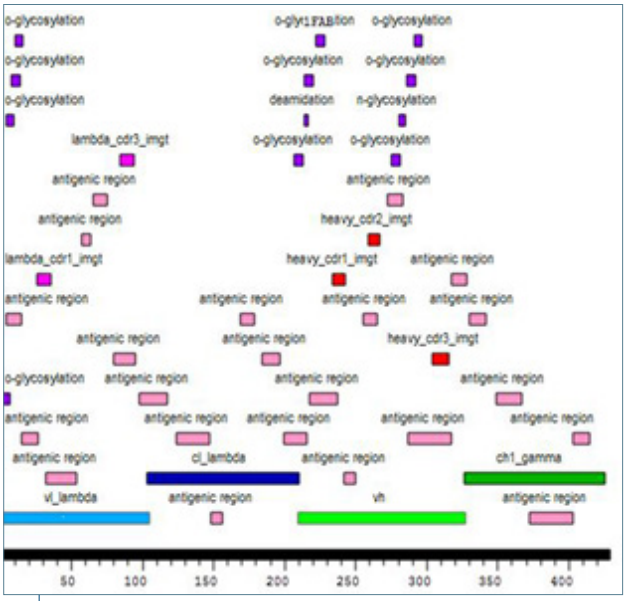


Figure 1: Example post translational modification sites identified in the Fab region of an IgG antibody [PDB: 1FAB]

- Predict mutation effects on binding affinity or stability
  - Perform amino-acid scanning to evaluate the effect of mutations on protein binding affinity or protein stability
  - Perform double, triple or multi-site mutations to identify the best multiple mutation for protein binding or stability
  - Automatically mutate selected residues to all 20 standard amino acid types and predict the mutation effects on binding affinity to help in the affinity maturation process
- Temperature and pH Dependent calculations[19]
  - Improve the quality of your stability predictions by accounting for temperature or pH dependent effects
  - Perform an in-silico titration exercise to scan stability or binding affinity across the solution pH band

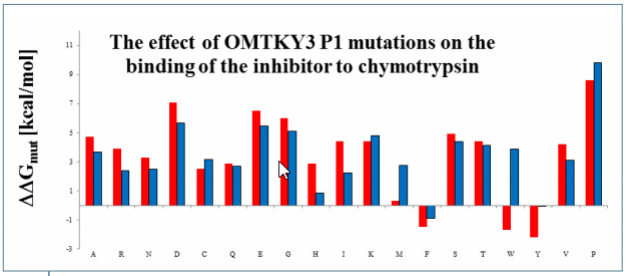


Figure 2: Example post translational modification sites identified in the Fab region of an IgG antibody [PDB: 1FAB]

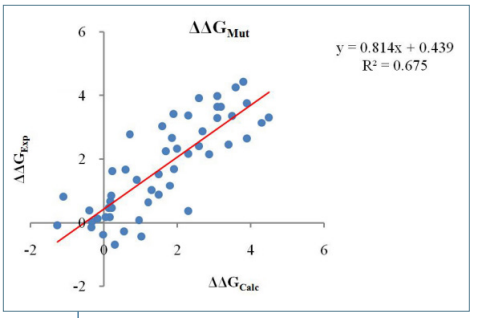


Figure 3: Example of the prediction of effect on protein stability by introducing a mutation to T4 Lysozyme.

- Spatial Aggregation Propensity and Developability
  - Use the experimentally validated Spatial Aggregation Propensity and Developability algorithms, licensed from the Massachusetts Institute of Technology and developed at Prof. Trout's laboratory[21,22,23,24]
  - Identify the size and location of regions on antibodies prone to aggregation
  - Based on aggregation propensity and molecular net charge calculated using our pK prediction method[10,11,] rank proteins for their long term shelf stability and developability

To learn more about BIOVIA 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. Lichtarge O., Bourne H.R., Cohen F.E. J *Mol Biol.*, **1996**, 257(2), 342-358.
4. 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.
5. Fiser, A. Sali, *A. Methods in Enzymology*, **2003**, 374, 463-493.
6. Sanchez, R. Sali, *A. Protein Structure Prediction: Methods and Protocols*, **2000**, 97-129.
7. Spassov, V.Z., Flook, P.K., Yan, L. *Prot. Eng., Design & Selection*, **2008**, 21, 91-100.
8. 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. , Caflisch 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.
9. Brooks B. R., Bruccoleri R. E., Olafson B. D., States D. J., Swaminathan S., Karplus M. J. *Comp. Chem.*, **1983**, 4, 187-217.
10. Spassov V.Z, Yan L *Protein Science*, **2008**, 17 ,1955-1970.
11. Spassov V.Z., Bashford D. J. *Comp.Chem.*,**1999**, 20, 1091-1111.
12. Still, W. C., Tempczyk, A., Hawley, R. C., Hendrickson, T., J. *Amer. Chem. Soc.*, **1990**, 112, 6127.
13. Dominy, B., Brooks, C.L. III., J. *Phys. Chem. B*, **1999**, 103, 3765-3773.
14. Spassov, V.Z., Yan, L., Szalma, S. J. *Phys. Chem. B*, **2002**, 106, 8726-8738.
15. 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.
16. Pierce B., Weng Z. *Proteins*, **2008**, 72(1), 270-279.
17. Chen R., Weng Z. *Proteins*, **2003**, 52, 80-87.
18. Pierce B, Weng Z. *Proteins*, **2007**, 67(4), 1078-1086.
19. Spassov V.Z., Yan L., *Proteins*, **2013**, 81(4):704-14.
20. Krowarsch D., Dadlez M., Buczek O., Krokoszynska I., Smalas A.O. Otlewski J., J. *Mol. Biol.*, **1999**, 289, 175-186.
21. Chennamsetty N., Voynov V., Kayser V., Helk B. and Trout B. L. J. *Phys. Chem. B*, **2010**, 114(19), 6614-6624.
22. Chennamsetty N., Voyonov V., Kayser V., Helk B. and Trout B. L., *Proc. Nat. Acad. Sci.*, **2010**, 106(29), 11937-11942.
23. Chennamsetty N, Helk B., Trout B. L, Voyonov V., Kayser V. PCT/US2009/047954. Filed 19th June, **2009**.
24. Chennamsetty N., Voynov V., Kayser K., Helk B., Trout B.L. *Proteins*, **2011**, 79, 888-897.

Our **3D**EXPERIENCE Platform powers our brand applications, serving 12 industries, and provides a rich portfolio of industry solution experiences.

Dassault Systèmes, the **3D**EXPERIENCE Company, provides business and people with virtual universes to imagine sustainable innovations. Its world-leading solutions transform the way products are designed, produced, and supported. Dassault Systèmes' collaborative solutions foster social innovation, expanding possibilities for the virtual world to improve the real world. The group brings value to over 170,000 customers of all sizes in all industries in more than 140 countries. For more information, visit **www.3ds.com**.

3DEXPERIENCE®

DASSAULT SYSTEMES | The **3D**EXPERIENCE® Company

Dassault Systèmes Corporate
Dassault Systèmes
175 Wyman Street
Waltham, Massachusetts
02451-1223
USA

BIOVIA Corporate Americas
BIOVIA
5005 Wateridge Vista Drive,
San Diego, CA
92121
USA

BIOVIA Corporate Europe
BIOVIA
334 Cambridge Science Park,
Cambridge CB4 0WN
England