



Centurion
UNIVERSITY
Shaping Lives...
Empowering Communities...

Protein preparation using Discovery Studio



Centurion
UNIVERSITY

Shaping Lives...
Empowering Communities...

Protein Preparation

- Protein preparation is the process of transforming a macromolecular structure into a form more suitable for a computational experiment. The specific set of tasks required depends on the experiment and the source of the structure. Tools for several common prep tasks are described below. For a demonstration of how these can be combined, see the *Preparing a Protein* example.

Processing Alternate Locations

- Protein structures rarely describe whole molecules. Instead, they are usually missing atoms that could not be resolved well. In addition, many structures contain *multiple* copies of atoms called alternate (“alt”) locations or alternate conformations that can indicate that part of the structure is moving. As the resolution of X-ray crystal structures improves, alt locations can be expected to become ubiquitous.



Centurion
UNIVERSITY

Shaping Lives...
Empowering Communities...

Adding Hydrogens and Optimizing Hydrogen Bonds

- Traditionally, most protein structures have been generated with either no explicit hydrogens or only polar hydrogens. However, having all atoms explicitly represented and positioned to make appropriate hydrogen bonds is sometimes necessary. The function `OEPlaceHydrogens` does this by adding and placing hydrogens and by “flipping” certain functional groups (e.g., sidechain amides and imidazoles) as required.

Splitting Macromolecular Complexes

- One common step in preparing a macromolecular complex for use as input to a calculation is to separate components based on their functional roles. For example, a **SZMAP** calculation on a four-chain protein crystal structure might require one of four copies of the ligand to be extracted and the protein-complex associated with that ligand binding site to be separately extracted, ignoring any³ waters, buffers, etc.



Centurion
UNIVERSITY

*Shaping Lives...
Empowering Communities...*

- Function to separate the input molecule into the ligand for the first binding site, the protein-complex (including cofactors) for that site, the binding site waters, and everything else. In this example, the ligand is identified automatically and the binding site contains components near the ligand.
- If a ligand cannot be located, the entirety of each protein and everything nearby defines each binding site. If the input molecule is empty or there is some other problem that prevents processing, `OESplitMolComplex` will return false.



Centurion
UNIVERSITY
*Shaping Lives...
Empowering Communities...*

Discovery Studio 3.5 Visualizer

File Edit View Chemistry Structure Sequence Chart Scripts Tools Window Help

Macromolecules Simulation Receptor-Ligand Interactions Pharmacophores Small Molecules X-ray My Tools Search product fea

New [Icons] Display Style...

Tools Welcome 1gra

- <Cell>
- 1GRA
 - A
 - A
 - A
 - Hetatm
 - Water
 - Active Sites
 - Protein Groups
 - Ligand Groups
- HBond Monitor1

Use the Rotate tool to rotate the selected objects. Press Shift to rotate in the Z-axis or press Ctrl to rotate selected objects

Enable Ad



Centurion
UNIVERSITY

*Shaping Lives...
Empowering Communities...*

H-bond optimization (ProtAssign)

The hydrogen bonding (H-bond) network was optimized by sampling 180 flips of the terminal chi angle for Asn, Gln, and His, which significantly changes the spatial H-bonding capabilities of the side chains, but does not appreciably change the fit to the electron density. Neutral and protonated states of His, Asp, and Glu were also sampled along with the two His tautomers (proton on either the Nd or Ne nitrogen). In addition, hydrogens on hydroxyls and thiols were sampled to optimize the H-bond network. The ProtAssign algorithm can be run in a “standard” mode that typically takes a few seconds or an “exhaustive” mode where many more states are considered, which can run for minutes or hours depending on the complexity of the H-bond networks.

Protein minimization (Impref)

In the recommended protein preparation protocol, after H-bond optimization, the entire structure is allowed to relax using the Impref module of Impact [61] and the OPLS_2005 force field [62–64]. The options explored for protein minimization included hydrogen only or all-atom with a termination criterion based on the root-mean-square deviation (RMSD) of the heavy atoms relative to their initial location.