Biographics Laboratory 3 R
Basel/Switzerland
E-Mail: admin@biograf.ch
Internet: http://www.biograf.ch

Reference Manual (Macintosh and Linux)

Yeti 7.0

Photo taken by Eric Shipton in 1951
© Royal Geographical Society, London

A molecular mechanics program for proteins and small-molecule protein complexes

Authors: Angelo Vedani and Max Dobler

Copyright © 1990–2006, Biographics Laboratory 3R, Basel, Switzerland

References

Vedani, A. and Huhta, D.W. (1990). A new force field for modeling metalloproteins. *J. Am. Chem. Soc. 112*, 4759–4767

Vedani, A. and Huhta, D.W. (1991). An algorithm for the systematic solvation of proteins based on the directionality of hydrogen bonds. *J. Am. Chem. Soc. 113*, 5860–5862

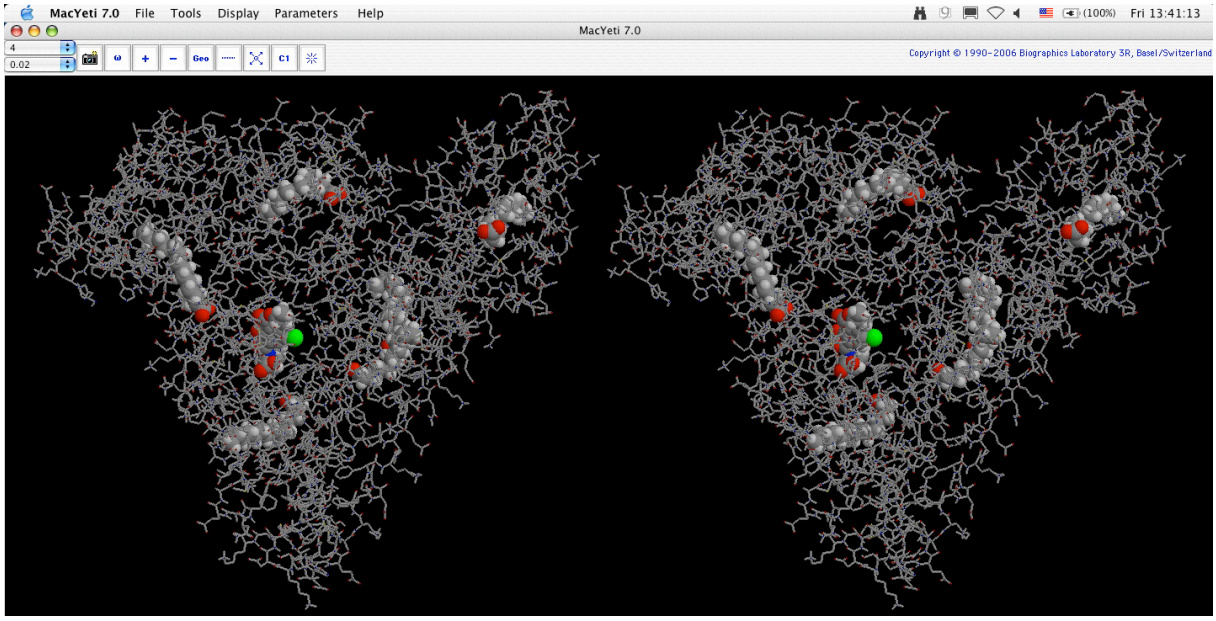
Vedani, A., Dobler, M. and Lill, M.A. (2005). Combining protein modeling and 6D-QSAR. *J. Med. Chem. 48*, 3700–3703

Lill, M.A., Dobler, M., and Vedani, A. (2006). Prediction of small-molecule binding to Cytochrome P450 3A4: Flexible docking combined with multi-dimensional QSAR. *ChemMedChem 1*, 73–81.

Table of Contents

1 Introduction to the molecular-mechanics software *Yeti*

O v e r v i e w



The modules of *Yeti* are presently dimensioned to handle up to 8,000 atoms and 1,250 residues. A maximum of 12 ligand molecules (e.g. substrates, inhibitors, co-enzymes, oligoatomic counter ions) are supported, each of which may include 200 atoms and 60 variable torsion angles. The full non-bonded list may be as large as 1,000,000. The special metal-center function (cf. below) is dimensioned for up to eight metal ions. Monoatomic ions (e.g. Cl^-) are not limited in number.

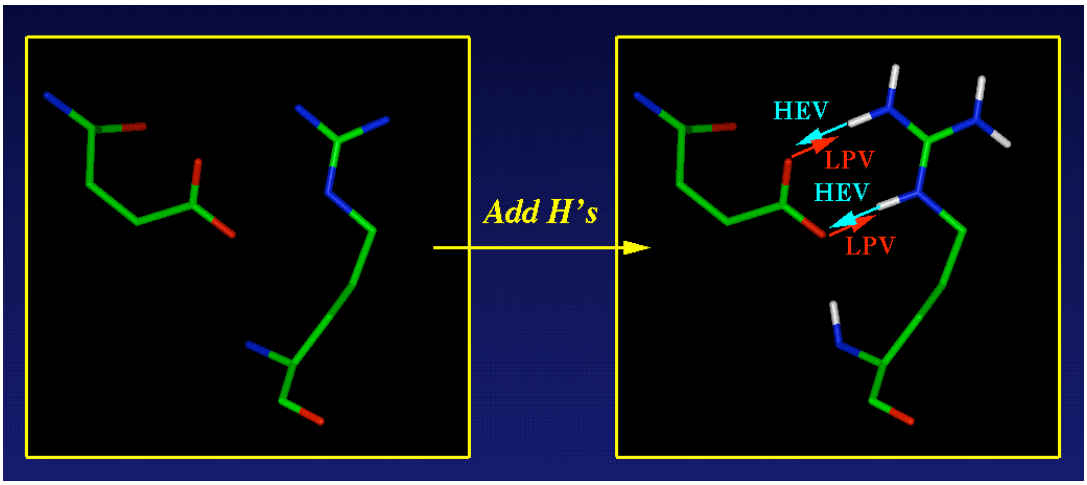
Yeti Modules

- Graphical version: *MacYeti* or *WinYeti*
  Manual or dynamic docking, protein processing including H-bond network orientation and protein solvation, minimization, all analysis options; setup of batch protocols.
- Non-graphical (batch) version: *MacYetiBackground* or *WinYetiBackground*
  cpu-intensive calculations (e.g. Monte-Carlo docking)
- Autodocking: *MacYeti_AutoDockX* or *LnxYeti_AutoDockX* (Linux)
  Automated docking of large batches of ligands, generation of 3D/4D ligand output

Functionalities in Yeti

- AddHyd: Generating polar hydrogen atoms
- Orient: Orienting H-bond networks
- Protein Solvation: Identification of structural and surface water molecules
- Minimization: Proteins and small-molecule protein complexes with a special emphasis on metalloproteins; directional force field, dynamic charge transfer
- Monte-Carlo: local/global search of binding modes; Boltzmann sampling
- Automated Docking: flexible docking and dynamic solvation

2 Adding polar hydrogen atoms



This module generates the positions of polar H atoms, i.e. H atoms attached to O, N, and S atoms (which, in general, are not included in the Protein Data Bank file). If the input file already include polar H atoms, this option remains silent.

AddHyd processes any native Protein Data Bank file, checks residues for internal consistency and generates the positions of all polar H atoms for all standard as well as for five modified amino acids (HIP: doubly protonated histidine, HIM: fully deprotonated histidine, ASH: aspartic acid, GLH: glutamic acid, CYM: deprotonated cysteine). In the *Yeti* force-field, H atoms attached to C atoms are represented as united atoms; apolar H atoms of the non-standard residues may be included explicitly.

Residues with a non-integer residue number (for example, residue insertions such as "189A", or HETATM records with H[n], n=1,2,3, ... residue numbers) will be assigned a new pure integer residue number. Disulfide bridges are automatically accepted if the separation of the two S atoms lies in the range from 1.86 Å to 2.26 Å. If a non-disulfide bridged residue is in contact with a metal center (HET), it may be renamed to CYM. Presently, only one polypeptide chain may be processed. If the protein includes more than one chain you can specify the chain of your choice. Trailer residues with residue number "zero" (e.g. "ACE 0") will be skipped. Chains may be linked together afterwards for further processing.

AddHyd will not generate any H atoms attached to non-standard residues such as substrates, inhibitors and co-factors. To generate polar H-atom positions of those molecules, you may use generally available software (e.g. *MacroModel*, *Bio*).

AddHyd contains various checks for incomplete coordinate input or multiple atom sites. If AddHyd fails to complete successfully, in general the problems are associated with some inconsistencies in the input. With the help of an editor, you may quickly locate the source of the problem. Note that ambiguous atom names (e.g. "AE1" and "AE2" in GLU or GLN residues residues indicating an uncertainty about the nature of the atom) will lead to termination of the program.

To generate the polar H-atom positions, AddHyd uses the following assumptions:

- Main-chain amide H: distance N–H: 1.01 Å, N–H direction along bisector of C–N–CA. For planar amide moieties, the H atom lies in the amide plane ("cis" or "trans" defined by the O, C, N and CA atom, respectively). For deviations from planarity, the algorithm places the amide H atom between the planar (i.e. undistorted) position and the C-N-CA' bisector, thus the N atom becomes pyramidal. It is assumed that the carbonyl moiety remains planar.

- TRP side chain N–H: distance N–H: 1.01 Å, NE1–H direction along bisector of CD1–NE1–CE2, the H atom lies in aromatic plane.

- SER side chain O–H: distance OG–H: 0.96 Å, angle CB–OG–H: 109.5°, torsion angle CA–CB–OG–-H: 180.0°

- THR side chain O–H: distance OG1–H: 0.96 Å, angle CB–OG1–H: 109.5°, torsion angle CA-CB-OG1-H:180.0°

- ASH side chain O–H: distance OD2–H: 0.96 Å, angle CG–OD2–H: 109.5°, torsion angle CB-CG-OD2-H: 180.0°. ASH = aspartic acid (protonated ASP)

- GLH side chain O–H: distance OE2–H: 0.96 Å, angle CD–OE2–H: 109.5°, torsion angle CG–CD–OE2–H: 180.0°. GLH = glutamic acid (protonated GLU)

- CYS side chain S–H: distance SG–H: 1.338 Å, angle CB–SG–H: 96.9 Å, torsion angle CA–CB–SG–H: 180.0°

- TYR side chain O–H: distance OH–H: 0.96 Å, angle CZ–OH–H: 109.5°, H atom lies in the aromatic plane.

- ASN side chain N–H2: distances ND2–H: 1.01 Å, angles CG–ND2–H: 120.0°, H atoms lie in amide plane (defined by ND2, CG, and OD1)

- GLN side chain N–H2: distances NE2–H: 1.01 Å, angles CD–NE2–H: 120.0°, H atoms lie in amide plane (defined by NE2, CD, and OE1)

- LYS side chain N–H3: distances NZ–H: 1.01 Å, angles CE–NZ–H: 109.5° torsion angles CD–CE–NZ–H: –60.0°/60.0°/180.0° (staggered conformation)

- ARG side chain N–H: distance NE–H: 1.01 Å, NE–H direction along bisector of CD–NE–CZ, H lies in plane defined by CD–NE–CZ. N–H2: distances NH–H: 1.01 Å, angles CZ–NH–H: 120.0°, H atoms lie in plane defined by NH1, CZ, and NH2.

- HID side chain N–H: distance ND1–H: 1.01 Å, ND1–H direction along bisector of CG–ND1–CE1, H atom lies in aromatic plane. HID = δ–protonated histidine.

- HIE side chain N–H: distance NE2–H: 1.01 Å, NE1–H direction along bisector of CE1–NE2–CD2, H atom lies in aromatic plane. HIE = ε–protonated histidine.

- HIP side chain N–H: distances ND1–H; NE2–H: 1.01 Å, ND1–H direction along bisector of CG–ND1–CE1; NE1–H direction along bisector of CE1–NE2–CD2, both H atoms lie in the aromatic plane. HIP = double-protonated histidine.

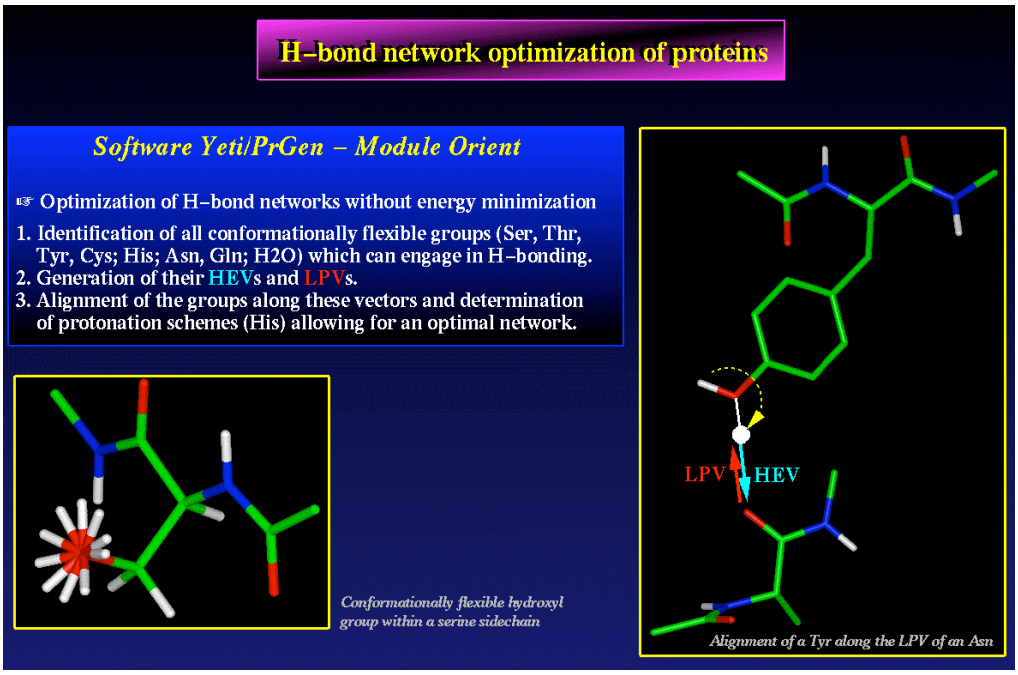
- WAT protons O–H: distances OH–H: 0.96 Å, angle H–OH–H: 104.5°, orientation: random. Solvent is properly oriented by module ORIENT (cf. below).

Note: The positions of side chain hydroxyl, ammonium, and mercaptyl H atoms of the amino-acid residues tyrosine, serine, threonine, and cysteine represent arbitrary conformations (planar and staggered, respectively) and should, therefore, be optimized prior to a full relaxation of the system. This can be accomplished using the modules ORIENT, and, for fine tuning, the [H]-type refinement of module MIN (cf. below). The H-atom positions of water molecules are generated in a random orientation and must therefore be properly oriented (H-bond networks) prior to any optimization. This can also be accomplished using the modules ORIENT and MIN.

PDB atom names ("CA", "CB", "OG1", etc.) are converted into atom types allowing to distinguish non-bonded properties. For standard amino-acid residues, these atom types are identical to those used in the software AMBER (cf. Scott C. Weiner et al., *J. Am. Chem. Soc.* **1984**, *106*, 765–784): The only difference includes sulfur-containing residues, (i.e. CYS, CYX, and MET) where no explicit lone-pairs are placed on the S atom (lone-pair directionality for H-bonding and metal-coordination, of course, is recognized in *Yeti*). In addition, *Yeti* allows for the simulation of the amino-acid residues CYM (metal-bound, deprotonated cysteine), ASH (protonated aspartic acid). and GLH (protonated glutamic acid). The *Yeti* residue types include: ALA, ASH, ASN, ASP, ARG, CYM, CYS, CYX, GLH, GLN, GLU, GLY, HID, HIE, HIM, HIP, ILE, LEU, LYS, MET, PHE, PRO, SER, THR, TRP, TYR, VAL, WAT (water), OHM (hydroxide ion), metal ions (Zn, Co, Fe, Cu, Co, Ni, Mg, Ca) as well as any protein-bound small molecules. Convention: if the 4th character of a small-molecule residue name is a "C", it is assumed to be covalently bound to the protein — otherwise, a non-covalent attachment is assumed.

Atomic partial charges are assigned to all atoms of standard amino acids. Those were taken from: Weiner, S.C. et.al., *J. Am. Chem. Soc.* **1984**, *106*, 765–784). For ions (residue type "HET"), atomic partial charges represents its formal charge; it may be dynamically altered when using the metal-center function (cf. below). Any single-atom ion (e.g. metals) should carry the residue name "HET". In addition, its atom name should include the formal charge, e.g. "ZN2+", "K1+", "CL1-", "CA2+". Of course, the effective/formal charge will be read from the coordinate file and modified by the charge-transfer function in module MIN, if enabled. The atomic partial charges assigned to the water molecule (O: –0.6; H: +0.3) represent a compromise between various quantum-mechanical and semi-empirical approaches (i.e. AMPAC, Gaussian: 86/STO–3G, 3–21G, 6–31G, 6–31G*, 6–31G**, 6–311+G**). Please note that the metal-center function was calibrated using these specific charges for the water molecule.

3 Orienting H-bond networks



Module ORIENT orients polar H atoms: the orientation of hydroxyl and mercaptyl-H atoms of the residues Tyr, Ser, Thr, water, and Cys residues is based on the directionality of H-bonds and is initiated at the center of the protein and proceeds radially towards the surface of the protein.

The orientation of a polar H atom is based on the presence of unambiguous partners (donors or acceptors). Such unambiguous partners include carbonyl and carboxyl O, imidazole N, and thioether S as acceptor moieties as well as main and side-chain amide N, guanidinium N, imidazole N as donors. The orientation of "ambiguous" groups (which can function both as H-bond donors or acceptors) such as hydroxyl O, sulfhydryl S, amine N, and water O may become unambiguous if they engage in H-bonds with other unambiguous partners.

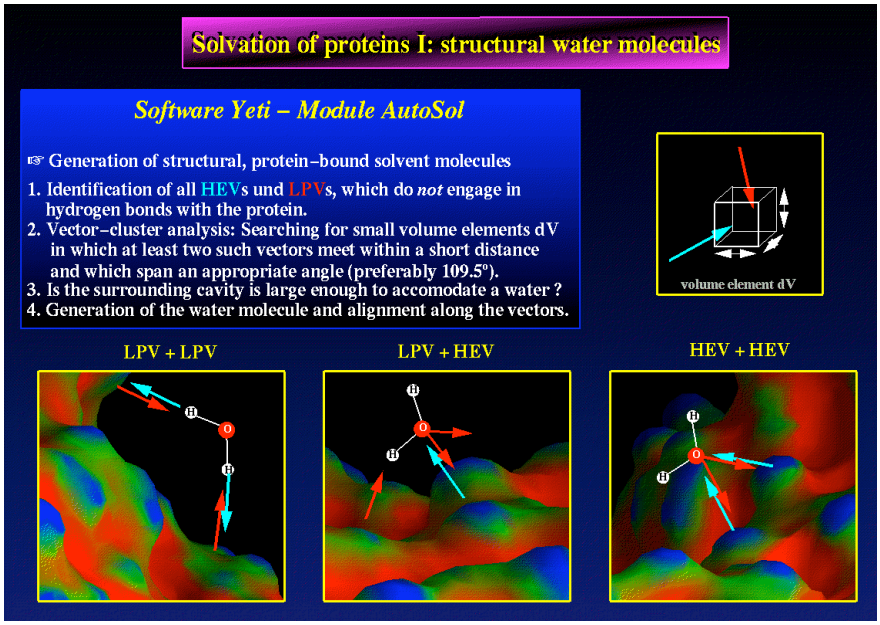
For orienting tyrosine hydroxyl-H atoms, only two orientations of the hydroxyl-H atom with respect to the aromatic ring plane are considered, cis: torsion angle CE1–CZ–OH–HOH (PDB notation) = 0°, or trans: torsion angle = 180°.

For orienting Ser and Thr hydroxyl-H atoms or Cys mercaptyl-H atoms, three staggered orientations of the H atom with respect to the amino-acid side chain (+60°, –60°, and 180°) are analyzed.

Water molecules are oriented based on distance, linearity, and directionality of the resulting H-bonds. Again, the water molecules are processed in the order of their increasing distance from the center of mass of the protein. The orientation of a metal-bound water molecule is often problematic and might lead to an unfavorable orientation with respect to the metal. Although module MIN will correct this within a few cycles, it is recommended to check the proper orientation.

4 Solvating proteins and protein complexes

(cf. Vedani, A., Huhta, D.W., *J. Am. Chem. Soc.* **1991**, *113*, 5860–5862)



Module Solvate (formerly: AutoSol) generates an oriented water shell around proteins or small-molecule protein complexes. The water shell comprises both internal and surface water molecules. The algorithm within Solvate includes the following steps:

1. Linearity of H-bonds is represented by H-extension vectors (HEVs), H-bond directionality by lone-pair vectors (LPVs). HEVs originate at H-bond donors; their endpoints mark ideal positions for H-bond acceptor atoms relative to a H-bond donor. LPVs originate at H-bond acceptors; their endpoints mark ideal positions for H-bond donor atoms relative to a H-bond acceptor. Length and orientation of HEVs and LPVs were derived from analyses of H-bonds in small-molecule crystal structures.

   To identify free HEV and LPV, able to engage in H-bonds with additional water molecules, it is necessary to delete vectors associated with already existing hydrogen bonds. Ideally, the HEV and LPV of a particular H-bond would overlap entirely, but this is rarely observed in protein structures. LPVs and HEVs are therefore extended to cones, representing the experimental distribution width. Criteria for identifying existing H-bonds are then based on the *Yeti* force field.

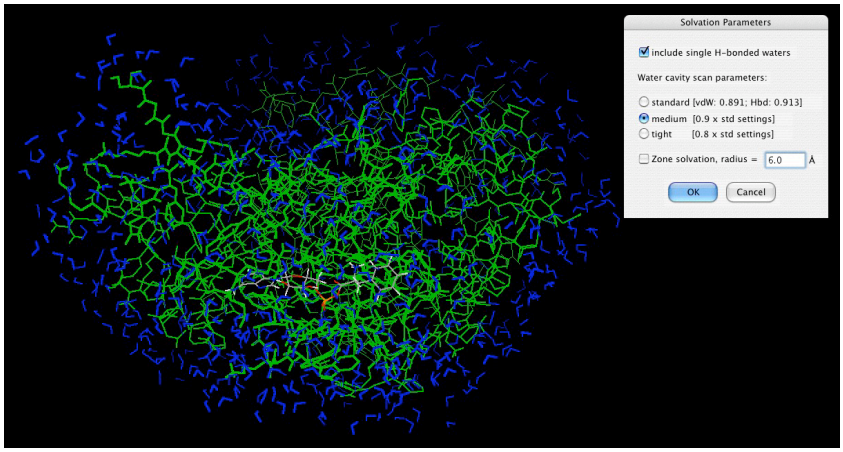
2. In a next step, the remaining vectors are analyzed for clustering. Probable sites for water molecules (i.e. the position of their O atom) are identified in regions of 3-D space where at least two vector endpoints (HEV or LPV) are located within a short distance and subtend an appropriate angle. The maximum vector separation is limited to 1.4 Å, half the length of a water···water H-bond. An initial position for the water O atom is then assigned to the midpoint of the two vector tips. Ideally, the angle subtended by these two vectors would lie in the range

from 104.5° (H–O–H angle; water molecule donates two H-bonds to the protein) to 114.4° (lone-pair [sp^3]–O–lone-pair [sp^3] angle; water molecule accepts two H-bonds from the protein). Solvate allows a maximal deviation of 45° from these values. As a third criterion, the program evaluates the deviation of the two resulting water-protein H-bonds from ideal directionality; a maximum deviation of 30° is accepted.

Next, the environment of this putative water O atom position is scanned for close contacts using van-der-Waals and H-bond parameters as defined in the *Yeti* force field. If unreasonably short contacts are encountered, the water site is rejected.

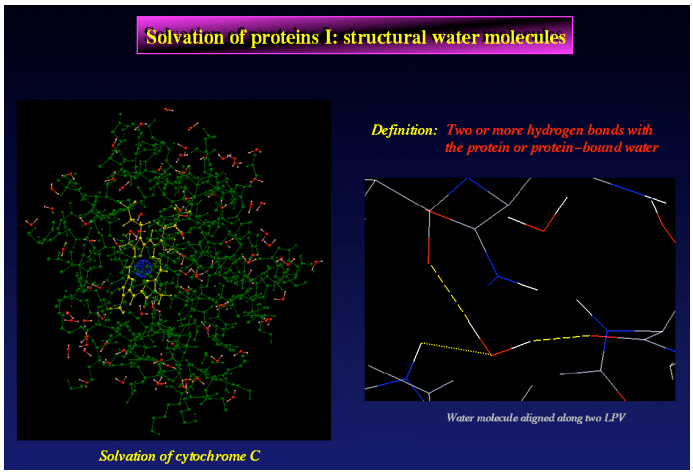
3. Finally, the water molecule is oriented along the the HEV (water lone pairs) and/or LPV (water H atoms) that were used to generate its O atom position.

Because each H-bond donor and acceptor of the protein (or protein complex) is systematically analyzed, Solvate is able to identify internal protein-bound solvent. As we showed in an earlier study on native human carbonic anhydrase I (cf. Vedani et al., *J.Am.Chem.Soc.* **1989**, *111*, 4075–4081), internal solvent is of utmost importance for molecular-mechanical simulations. Presently, the major limitation of Solvate is the solvation of larger hydrophobic regions exposed to the solvent, since only few HEV or LPV originate at side chains of hydrophobic residues. However, this limitation can be removed by combining Solvate with algorithms that place molecules in a box of pre-equilibrated water. The size and structure of a solvent shell generated by Solvate is most sensitive to the parameters defining the acceptable cavity, i.e. the van-der-Waals and H-bond parameters of the *Yeti* force field. Default values for the scaling parameters (0.891; 0.913), defining the minimal van der Waals and H-bond distances generate a maximum number of water molecules with acceptable force-field energies.

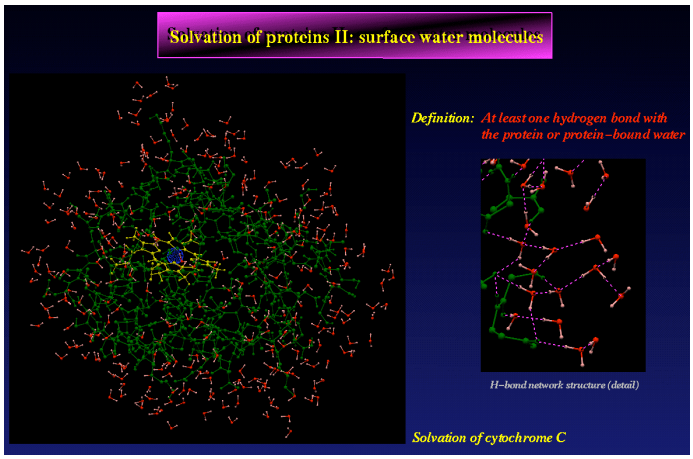


The user has the option to select smaller parameters, i.e. to accept water molecules identified in somewhat tighter cavities. During the refinement, most of these will relax if the parameters are not selected smaller than 0.713 for vdW and 0.730 for H-bond interactions, respectively (which corresponds to 80% of the default values given above).

Finally, if selected, the program will saturate all remaining "single" vectors by attaching a water molecule to their endpoints.



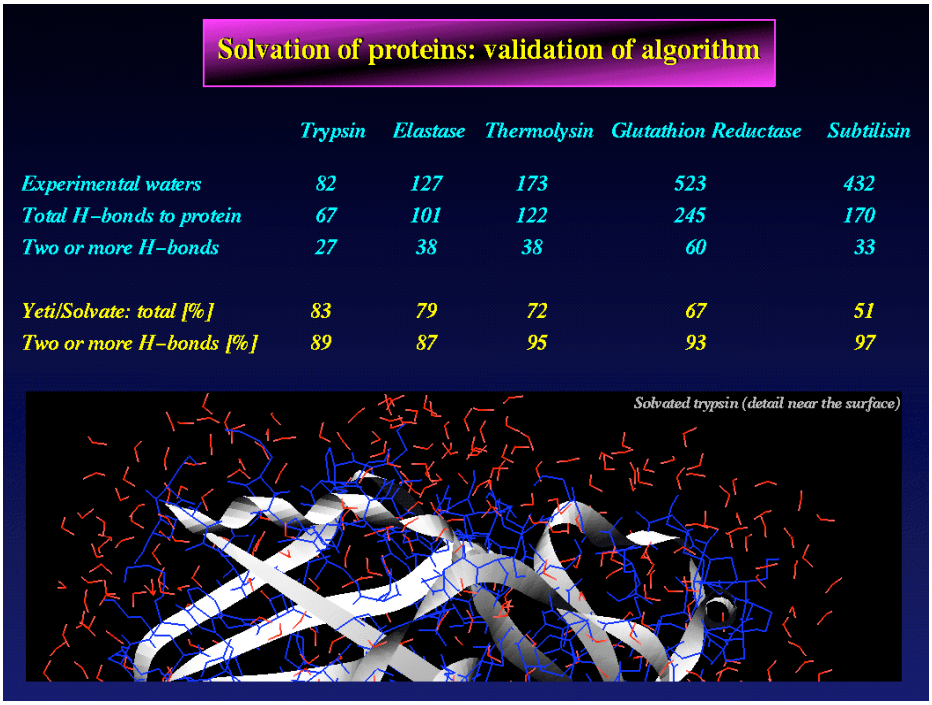
4. This procedure may be repeated (in combination with molecular mechanic optimizations of the solvent) until the desired level of solvation is reached. A possible shortcut includes the generation of n (n=1,2,3,...) solvent shells by Solvate in one single step without intermittent refinement.



Validation of the Solvate algorithm

Details and validation of the solvation algorithm are described in *J. Am. Chem. Soc. 113*, 5860–5862.

Extract: The program was tested by reproducing experimental solvent found in the high-resolution protein crystal structures of trypsin (1.7 Å resolution), elastase (1.65 Å), thermolysin (1.6 Å), glutathione reductase (1.54 Å) and subtilisin (1.2 Å). Based on these five proteins, Solvate is able to reproduce an average of 70% of the crystallographic solvent within 1.5 Å (≈ half the length of a H-bond) from the experimental position. Internal protein-bound water molecules are reproduced at the 92% level.



Summary: validation of solvation algorithm (cf. above)

5 Optimizing the structure of proteins and protein complexes

(cf. Vedani, A., Huhta, D.W., *J. Am. Chem. Soc.* **1991**, *113*, 5860–5862;
Vedani, A. et al. *J. Am. Chem. Soc.* **1989**, *111*, 4075–4081)

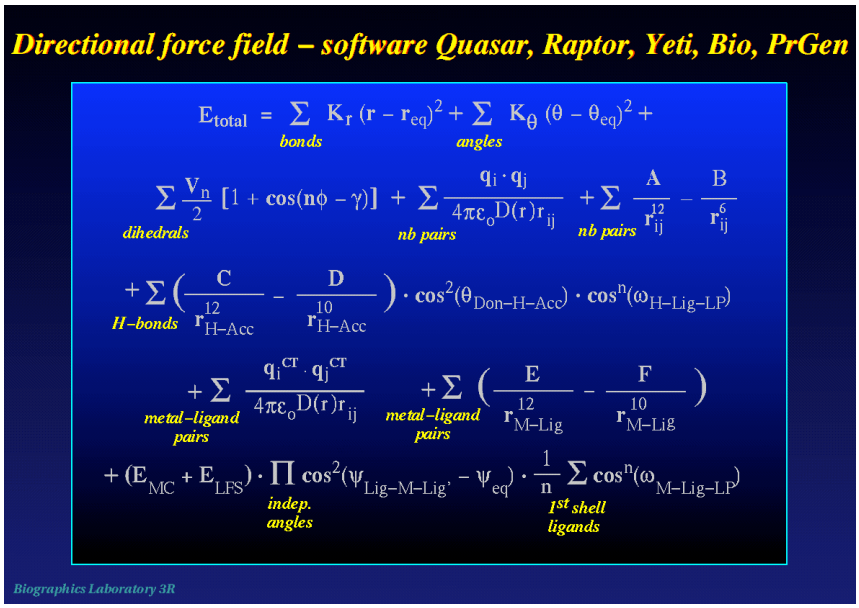
Module MIN performs the molecular mechanics optimization. Special features include a directional potential function for H-bonds, directional potential functions for metal-ligand interactions, directional and dynamic ligand-metal charge transfer for metal-bound residues, ligand-field stabilization for transition-metal ions as well as the compatibility with the AMBER software and force field (U. Chandra Singh, Peter K. Weiner, Jim W. Caldwell, and Peter A. Kollman; University of California, San Francisco, 1986).

As the *Yeti* minimizer operates mainly in internal coordinate space, the protein backbone is not altered during minimization — a choice which has both advantages and disadvantages. Consequently for the protein, only the conformation of the amino-acid side chains is optimized while all other entities (ions, solvent and small molecules) are fully optimized, i.e. translation, rotation and conformation.

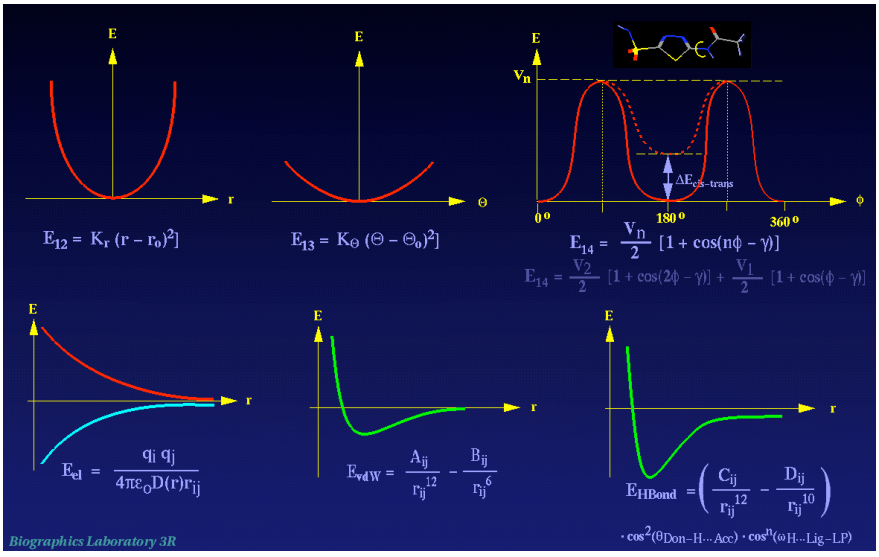
If there is evidence for larger conformational changes of the protein backbone (e. g. an induced-fit mechanism), a full relaxation in cartesian coordinate space should be considered. For this task, we recommend to use the software *AMBER* (stand-alone or, for example, as incorporated in *MacroModel*). However, it must be pointed out that a full relaxation of the protein in absence of an appropriate solvent shell may lead to unrealistic conformational changes, particularly on inner/outer surfaces of the protein.

5.1 The Yeti force field

Derivation and validation of the *Yeti* potential function are described in *J. Am. Chem. Soc.* **1991**, *113*, 5860–5862 and *J.Am.Chem.Soc.* **1985**, *107*, 7653–7658. It is based on the *AMBER* force field but includes directional terms for H-bonds and metal-ligand interaction as well as dynamic metal-ligand charge transfer:



Shape and equation of the various force-field terms are depicted below. 1–2 (bond distances) and 1–3 (bond angles) terms are not altered during a minimization with *Yeti* — *Bio 3.0* allows for the full minimization and includes the metal function.

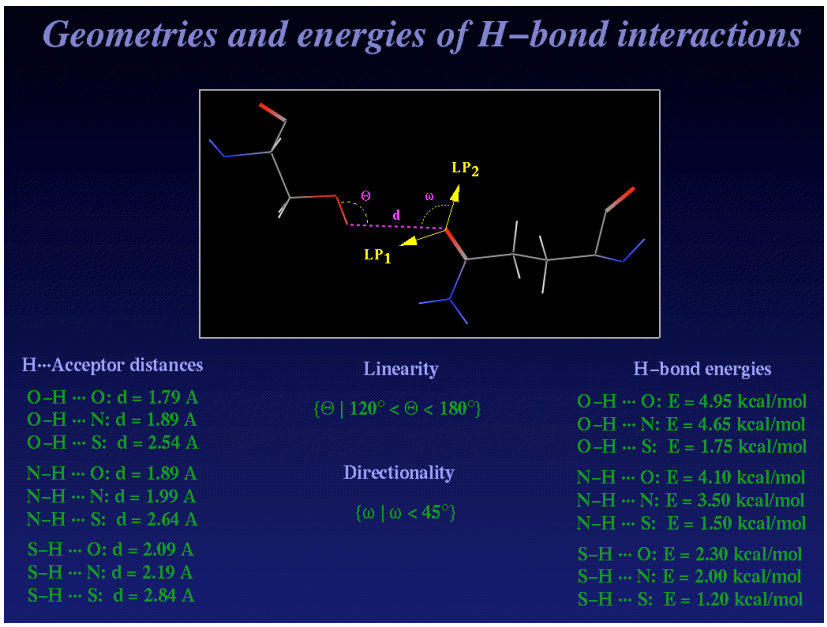


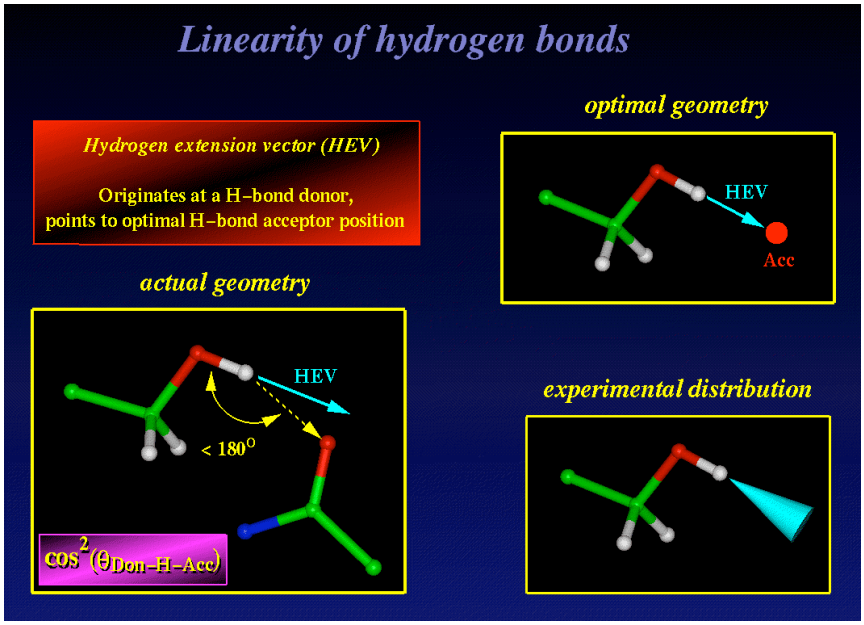
The *Yeti* force field has been calibrated using a dielectric parameter D(r)=2r. As in AMBER, van-der-Waals and electrostatic terms for 1–4 interactions are weighted with a factor 0.5. The *Yeti* metal function is not shown here but will be discussed in detail below.

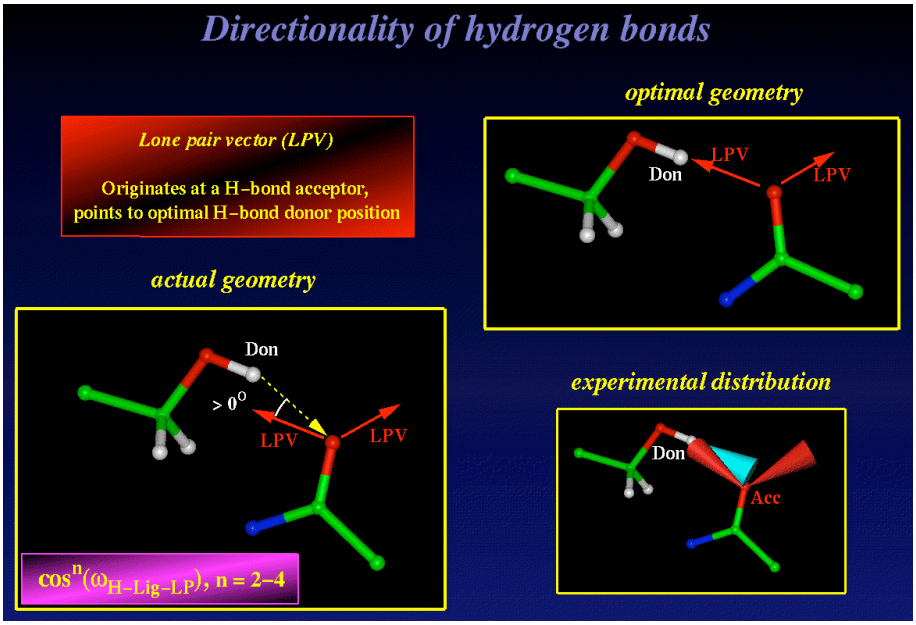
5.2 Treatment of H-bonds with a directional function

Directionality of H-bonds was first discussed by Kroon and co-workers in 1975. The analysis of 45 small-molecule crystal structures (including 195 O–H···O H-bonds) showed a preference for the donor-H atom to cluster in a plane along the bisector of the R–O–H angle at the acceptor O atom (cf. *J. Mol. Struct.* **1975**, *24*, 109–129). In 1984, Murray-Rust and Glusker analyzed the spatial geometry of H-bonds in small-molecule crystal structures with particular reference to preferred directions at O acceptor fragments. Using data retrieved from the CSD, they showed that H-bond donors are concentrated in directions commonly ascribed to the lone-pair orbitals of the O acceptor atom (cf. *J. Am. Chem. Soc.* **1984**, *106*, 1018–1025). Similarly, Kennard and Taylor studied H-bonds involving carbonyl and hydroxyl O atoms as H-bond acceptors (cf. *Acc. Chem. Res.* **1984**, *17*, 320–326). Dunitz and Vedani analyzed hydroxyl and sulfonamide O, as well as N acceptor atoms in aromatic five-and six-membered rings (cf. *J. Am. Chem. Soc.* **1985**, *107*, 7653–7658). In 1984, Baker and Hubbard published a detailed study on the geometry of H-bond in high-resolution protein structures (cf. *Prog. Biophys. Molec. Biol.* **1984**, *44*, 97–179).

In 1985, we proposed an extended potential function to allow for directionality of H-bonds in molecular-mechanical calculations. This function includes as an additional variable (besides the H···Acc distance and the linearity of the H-bond) the deviation of the H-bond from the closest lone-pair direction at the acceptor atom (angle H···Acc–LP). For details, refer to *J. Am. Chem. Soc.* **1985**, *107*, 7653–7658, or *J. Am. Chem. Soc.* **1989**, *111*, 4075–4081. The exponent *m* is most critical for the evaluation of this "explicit" H-bond term (Note that electrostatic contributions [for Don···Acc and H···Acc interactions] and van-der-Waals contributions [for Don···Acc interactions] are calculated separately) *m* defines the penalty for the deviation of the actual H-bond from the closest lone-pair direction at the acceptor fragment; *m* was calibrated for each individual H-bond acceptor to give the best possible agreement with the experimental distribution.







The H-bond function in *Yeti* was calibrated using small-molecule mimics for which structural, and preferably energy data was available. All structural information (except the water cluster which was treated *ab initio*) were retrieved from the CSD:

1. O–H···O H-bonds were calibrated using a water dimer and a homodromic water pentagon. Energies were calibrated using experimental data (from the literature) as well as by *ab initio* calculations using a 6-31G* basis set.

2. N–H···O H-bonds were calibrated using the N-methyl-acetamide dimer and geometric data for >N-H···O=C< H-bonds found in accurate small-molecule crystal structures.

3. O–H···N H-bonds were calibrated using the the water-imidazole system.

4. N–H···N H-bonds were calibrated using the N-methyl-acetamide-imidazole pair.

5. H-bonds involving S atoms as H-bond donors or acceptors were calibrated using the systems: 1. cysteine–water (S–H···O); 2. cysteine–imidazole (S–H···N); 3. water-cysteine (O–H···S); 4. N-methylacetamide-cysteine (N–H···S) and 5. cysteine-methionine (S–H···S)

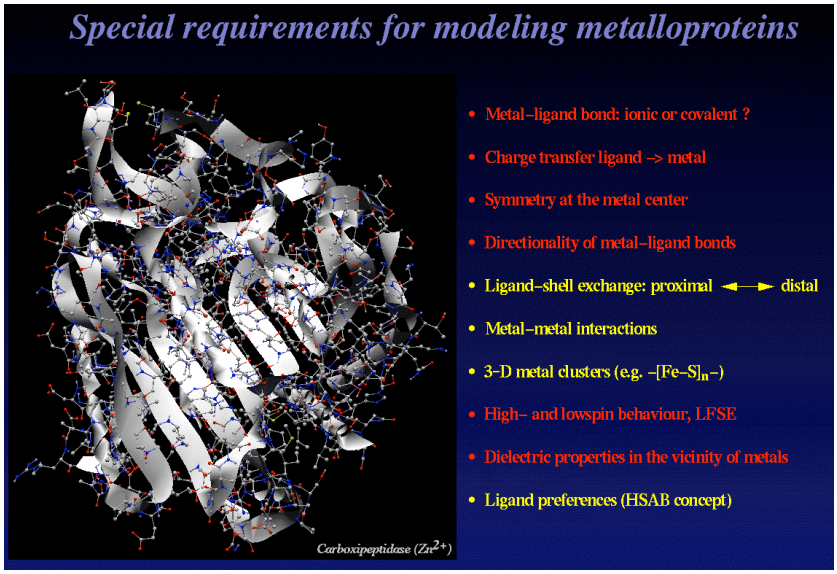
The correction terms for linearity and directionality are disabled if the distance term yields a positive energy. The coefficients A' and C' are calculated from H-bond distances and well depths assigned in the parameter file. For further details, please refer to the publication "A new force field for modeling metalloproteins" published in J. *Am. Chem. Soc.* **1990**, *112*, 4759–4767.

Definition of lone-pairs:
carbonyl O — two sp^2; n=3
carboxyl O — two sp^2; n=4
hydroxyl O — three sp^2/sp^3; n= 3
water O — three sp^2/sp^3; n= 3
ether O — two pseudo* sp^2; n=4
N in aromatic 5-membered rings — one sp^2; n=4
N in aromatic 6-membered rings — one sp^2; n=4
sulfhydryl S two intermediate; n=3
thioether S two intermediate; n=3

* The two pseudo lone pairs to mimic the broad experimental distribution

For phosphate and sulfonamide O atoms no lone pairs can be defined; here, the angle subtended at the acceptor atom is used a a criterion instead.

5.3 Metal centers: symmetry, directionality and charge transfer



To achieve a compromise between a “bonded” approach (metal-ligand interaction is defined as a covalent bond) and a “non-bonded” approach (metal-ligand interaction is simulated by electrostatic and van-der-Waals interaction instead), we have developed an empirical potential function for modeling metal centers in macromolecules: It includes two major terms, one describing the radial behaviour of the metal-ligand interactions, the other analyzing the first ligand sphere at the metal.

Apart from the metal-ion and ligand-atom type, the radial term of this new function varies solely by the metal-ligand distance. The summation extends over all potential metal-ligand pairs (the term *ligand* presently applies to all O, N, and S atoms/ions, capable of metal-coordination, i.e. having at least one free lone pair). This non-bonded type approach allows for mobility of all ligands between various shells, i.e. the metal can change both number and arrangement of its proximal ligands during a refinement.

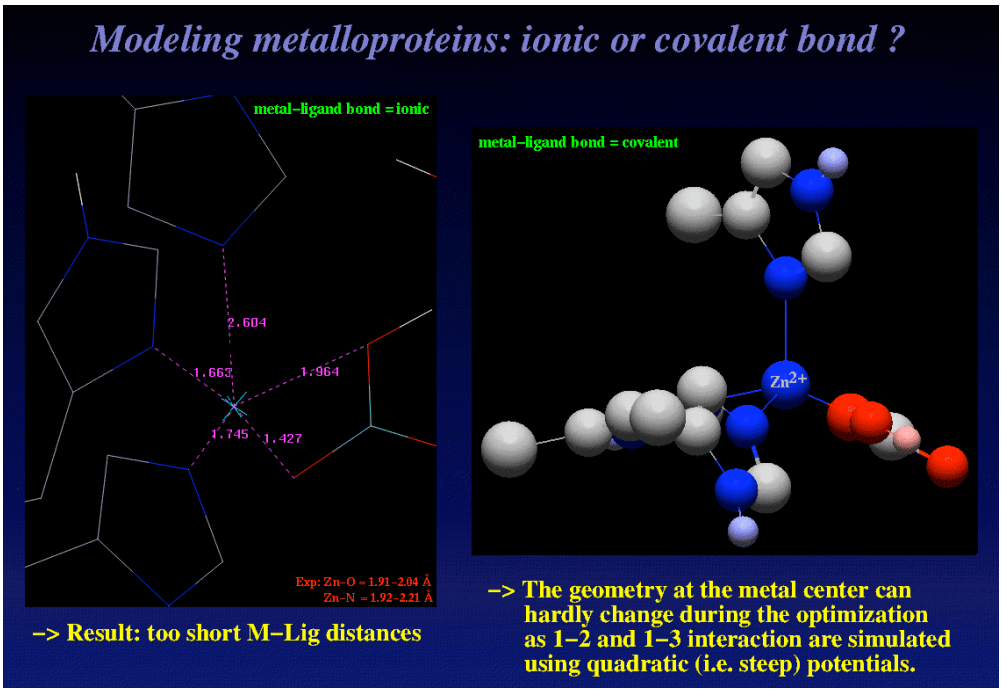
The coefficients A" and C" depend on the equilibrium distance, r(0), and on the well-depth, ε(0), for each M···L interaction, both of which depend on the coordination type as well as on the nature of the atoms involved. Values for r(0) were derived from data retrieved from the CSD, those assigned to ε (0) were estimated from semi-empirical calculations on small-molecule model systems.

One disadvantage of this approach is that axial and equatorial ligands a not treated differently. The analysis of the data retrieved from the CSD shows that (4+1), (3+2), and (4+2) distorted polyhedra are fairly common. A previous metal-center function (*Yeti* V4.3 and older) treated equatorial ligands (“narrow” experimental distribution) with a 10/12-type function, whereas axial ligands (broader distribution) were treated with a 6/12-type function. We feel, however, that for macromolecular applications an unbiased determination of coordination number and type is more important than than a slightly different treatment of axial and equatorial ligands. Moreover, the calibration of such subtleties would require higher resolved structural data than is presently available for metalloproteins.

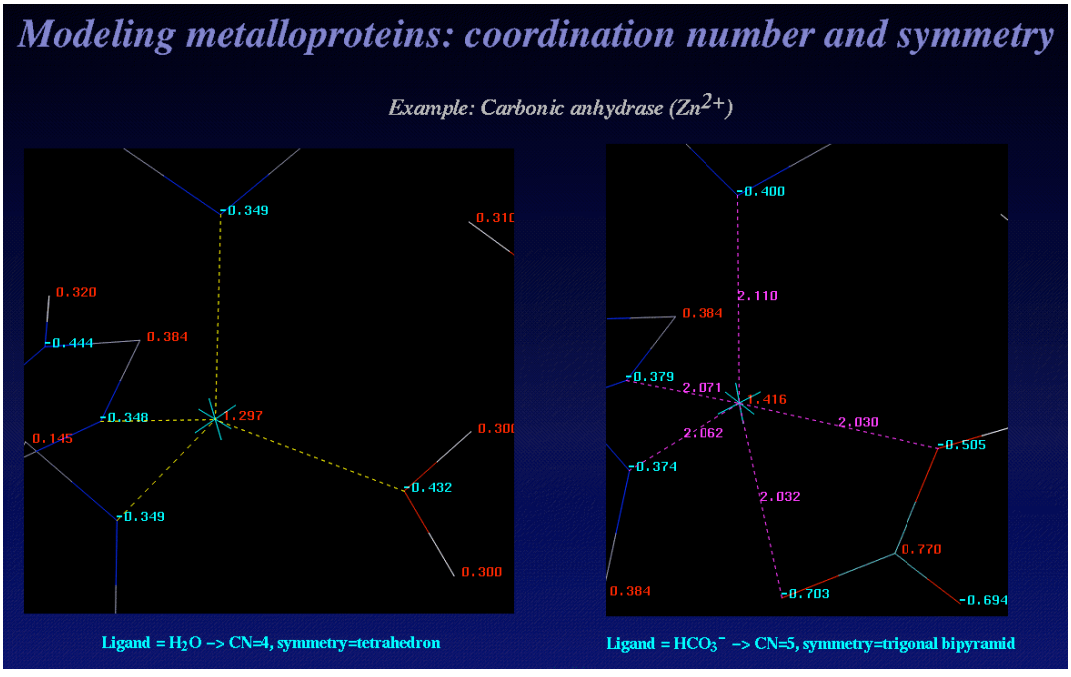
The directional term of the metal-center function analyzes the first ligand shell at the metal. Its energy depends on the symmetry at the metal center, the directionality of the metal-ligand bonds, and (in case of a transition metal) the ligand-field stabilization (LFSE). A weighting factor allows variation of the ratio of radial/directional energy. Best results were obtained by assigning a weight from 0.50–0.75 to the radial term and a corresponding weight of 0.50-0.25 to the directional term. By assigning a weight of 1.0, the directional term can be totally disabled.

Optimal values for the reference angles used in the symmetry term are given by symmetry or were obtained by analysis of structural data retrieved from the CSD. Although all Lig'-Met-Lig" angles are explicitly evaluated for this term, the summation extends only over the independent angles (tetrahedron and square plane 5 out of 6 angles; square pyramid and trigonal bipyramid 7/10; octahedron 9/15, respectively). Values for the coefficient "m" in the directional term are identical to the ones used in the H-bond function (cf. above). Values for the ligand-field stabilization can be obtained from theoretical considerations or experimental data.

A special metal-center function would seem to be essential for modeling metalloproteins as the character of a metal-ligand bond cannot simply be represented by of the the two extremes — bonded (100% covalent) or non-bonded (100% ionic):



In addition, coordination number and symmetry play a crucial role when modeling protein-bound transition-metal ions such as Zn, Co, Ni, Cu, and Fe. Here, the special function in *Yeti* allows to estimate rather subtle differences between the various coordination polyhedra. Presently, the following types are included: tetrahedron, square plane, square pyramid, trigonal bipyramid, and octahedron.

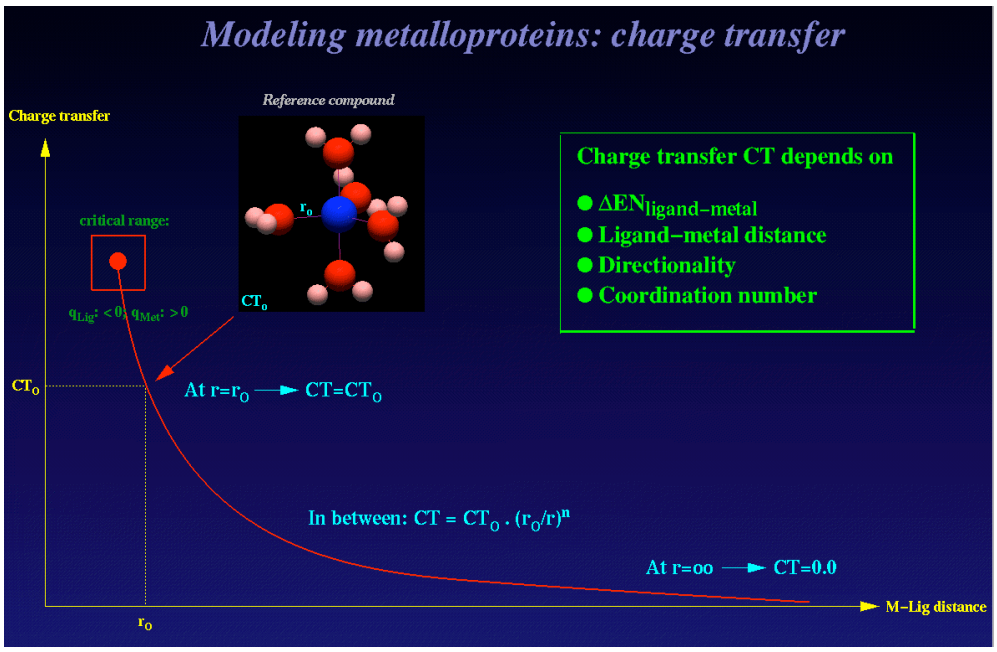


Charge Transfer

Probably the most important parameter for modeling metal-ligand interactions is the charge distribution between metal ion and ligand atoms. It is obvious that using formal charges for the metal ion will overestimate the electrostatic interactions. The weight of the electrostatic term (corresponding to the ratio of ionic/covalent character for a particular metal-ligand bond) is critical for modeling transition-metal elements where the metal-ligand bond has a finite covalent character.

We have attempted to develop an empirical function for ligand-metal charge transfer, which, based on theoretical or experimental evidence, allows dynamic adjustment of the charge distribution between metal and ligand atoms during the refinement. The main idea behind this function is that a reasonable estimate for the charge distribution between metal and ligands can be derived for small, well-characterized systems. As a first approximation we have used the difference in electronegativity to estimate the extent of charge transfer. This yields atomic charges in our model systems that are in qualitative agreement with those calculated by semi-empirical methods (MNDO as implemented in AMPAC; distributed by QCPE, University of Indiana, Bloomington, In.) using zinc parameters obtained from the literature (cf. *J. Computer-Aided Molecular Design* **1989**, *3*, 23–37).

Based on a theoretical or experimental value for a particular metal-ligand bond in a model system, our function controls the amount of charge transfer depending on the metal-ligand distance by using two reference points: At a distance equal to the reference distance, the amount of charge transfer corresponds to the theoretical or experimental value derived for the model compound; at infinite separation charge transfer vanishes (i.e. pure electrostatic/van-der-Waals interactions prevail). In between, the amount of charge transfer falls off exponentially. The steepness of this exponential decay can also be used to control the amount of charge transfer among four-, five-, and six-coordinate species, for which different mean metal-ligand separations are observed:



For distances significantly shorter than the reference distance (as they might occur at the beginning of a molecular-mechanical refinement), this function would lead to an unrealistic charge distribution by transferring too much charge from the ligand atom to the metal ion. In such cases, the charge transfer is disabled and, instead, electrostatic interactions involving the metal ion are damped until a reasonable geometry is obtained (usually after one iteration cycle).

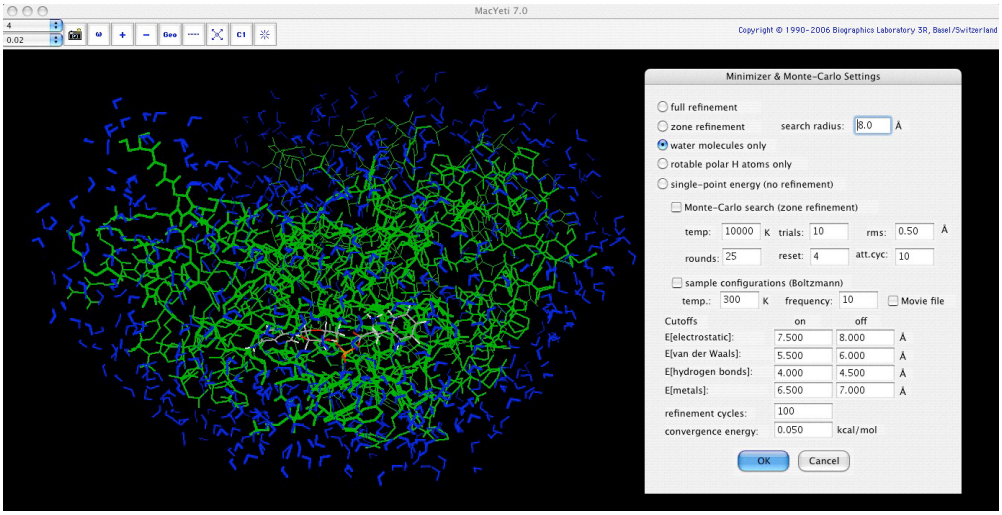
The metal-center and charge-transfer functions in *Yeti* were calibrated using the model systems ML(n) with M = Zn(II), Co(II), Ni(II), Cu(II), Fe(II), Fe(III), Mg(II) and Ca(II); L = H2O, NH3, H2S, OH–, SH–; and n=4,5,6, (except for L=H2S, SH– and M=Zn(II) where n=4,5, since no corresponding six-coordinate structures were found in the database) and mean M–L distances obtained from the search of the CSD. Detailed info is stored in the file yeti.par.

As a special feature, directionality of the ligand-metal charge transfer can be enabled. If selected, the amount of charge transfer depends not only on the metal and ligand-atom type and their actual separation, but also on the relative orientation of the ligand lone pair with respect to the metal center.

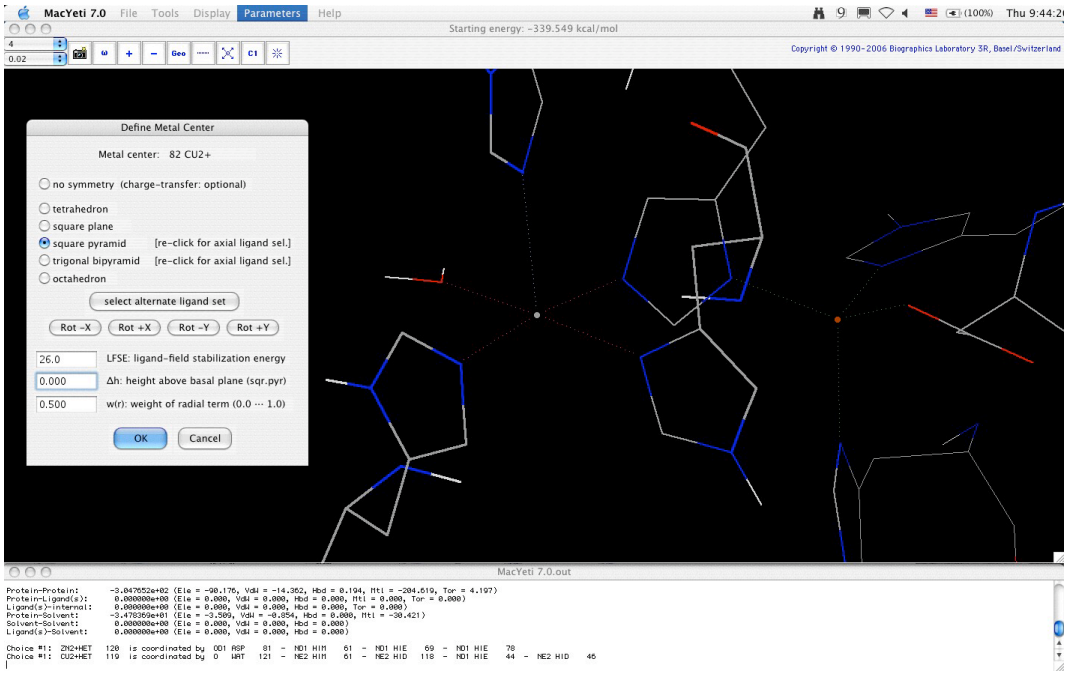
For further details, please refer to the publications "A new force field for modeling metalloproteins" (*J. Am. Chem. Soc.* **1990**, *112*, 4759–4767) and "An empirical potential function for metal centers" *J. Comp. Chem.* **1986**, *7*, 701–710.

5.4 Refinement options — Examples

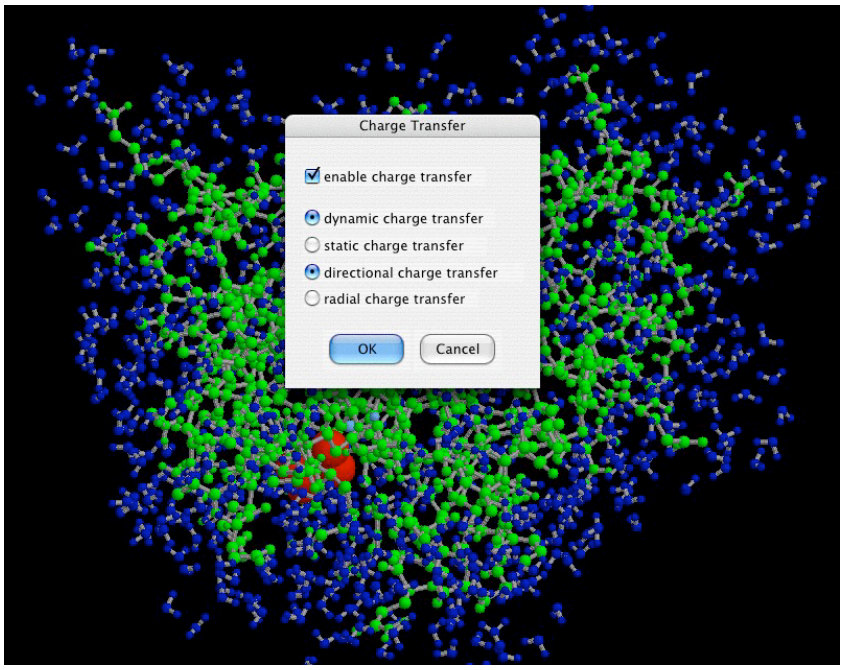
Presently, the following refinement options are supported in MIN: 1. full refinement, 2. Zone refinement [any atom/molecules located within the defined distance from any atom of the selected molecule(s)], 3. variable H atoms (this option is particularly useful after adding H atoms to a protein as it does not modify any of the non-H atoms), 4. solvent. Of course, you may select to calculate the energy of a given system without refining its structure (single-point energy). Within the zone refinement option, a Monte-Carlo search protocol may be enabled (cf. below and example 3).



When modeling a protein-bound metal ion, you may enable both the metal-center function as well as the charge transfer. Presently, the following symmetries are supported in *Yeti*: 1. tetrahedron, 2. square plane, 3. square pyramid, 4, trigonal bipyramid and 5. the octahedron. In addition, you may define the ligand-field stabilization energy (LFSE), and the height of the metal above the basal plane (sqpy):

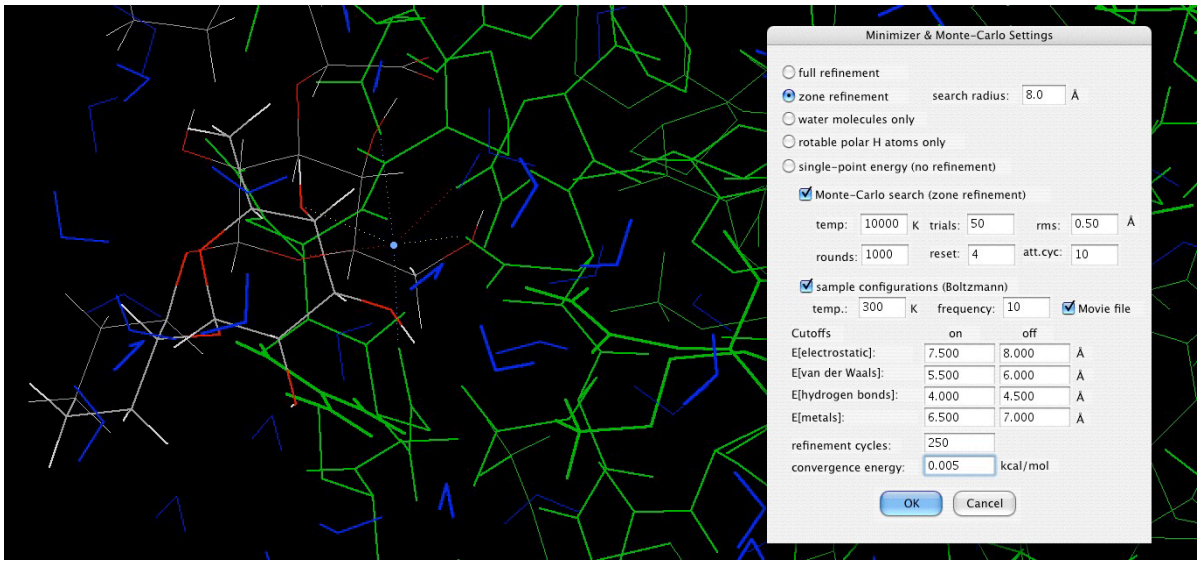


When activating the charge-transfer function, you may use the directional (default) or the radial mode. In the latter, the deviation of the metal ion from the closest lone pair at the donor fragment is ignored. If you are modeling a portion of the protein remote from any metal, use static (it saves time), otherwise dynamic (default).



Refinement using Monte-Carlo search/minimization

When optimizing conformation, position and orientation of a protein-bound small molecule, the resulting structure strongly depends on the starting point (problem of local minima). It is, therefore, recommended to enable a Monte-Carlo search during minimization:



With this option, the small molecule is randomly repositioned, reoriented and assigned a new conformation in each Monte-Carlo round. Initial changes in torsion angles [default: up to ±60°; scaled based on the number of atoms that are affected by a torsion], orientation [default: up to ±10°] and position [default: up to ±0.5Å] are automatically adjusted during the Monte-Carlo procedure to obtain a maximum number of low-energy structures. The *rms* criterion defines the minimum structural distance between parent and child structure (before minimization). The number of starting structures (default = 10) generated at the beginning of each round is controlled by the *trials* box. Those structures are first checked for sufficient structural distance from the parent structure (*rms* box) and then subjected to a Boltzmann criterion defined by the temperature T (*Temp.* box; def. =10,000 K):

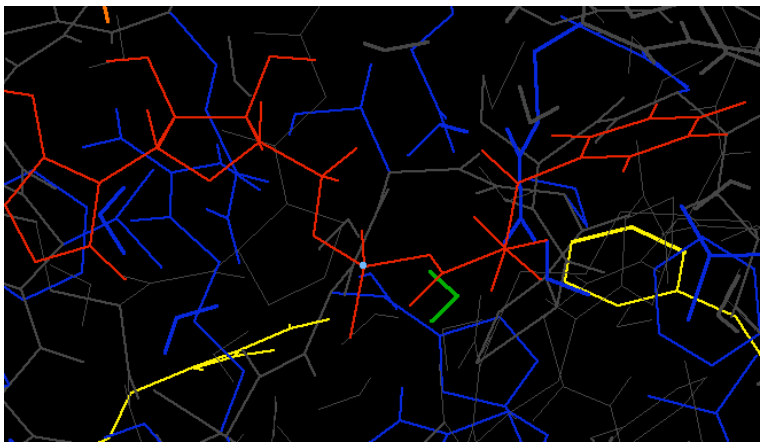
$$bf = e^{-\Delta E/RT} \quad \text{with } \Delta E = E(\text{child}) - E(\text{parent}); \; R = 1.986 \cdot 10^{-3} \text{ kcal/(mol·K)}$$

Thereafter, a random number (*rn:* 0.0···1.0) is drawn and the structure is accepted for minimization if $rn < bf$. If, after n trials, no structure meets this requirement, the lowest-energy structure is used instead. The *temperature* controls the energy range ΔE for (immediately) accepting a starting structure as follows:

T = 100 K corresponds to a ΔE=RT of 0.2 kcal/mol	
300	0.6
1,000	2.0
5,000	9.9
10,000	19.9
25,000	49.7
50,000	99.3

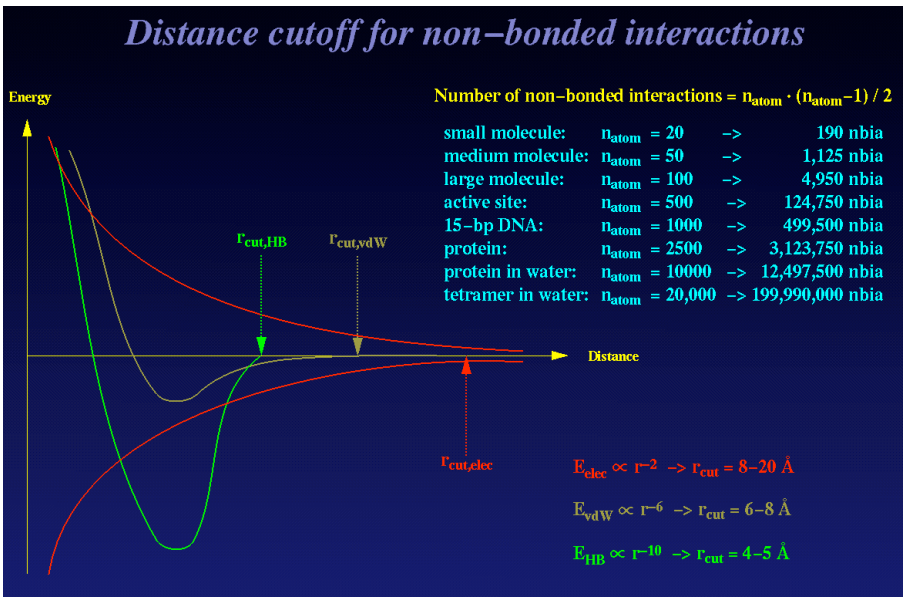
The Monte-Carlo concept is based on a global search procedure, i.e. a new search originates from the previous structure — not from the original position. If, after *n* search steps no new minimum is found, the algorithm restores the previously best structure. The default of *n=4* can be altered in the minimizer dialog (cf. reset box). The optimal number of Monte-Carlo rounds depends on the size of the molecule to be sampled and the tightness of the binding pocket. A starting value of 25 (default) would seem to be reasonable. The initial *i* structures (trials box, default = 10) are generated starting from an identical parent structure. After the final round, the structure with the lowest total energy is restored.

Please note that a Monte-Carlo search can only be initiated through the zone refinement option. If you wish to optimize the whole protein during such a task, just set the search radius to 100 Å. During the minimization, all residues that are located within the search distance of any of the original residue(s) [to be selected] will be optimized along with the small molecule subject to the Monte-Carlo search. If no small molecule can be located in the zone, the Monte-Carlo search will be disabled; if more than one small molecules are present, only the last (in sequence) will be subject to a Monte-Carlo search. Initial changes in torsion angles, orientation and position can be changed in the file yeti.par. When using large ligand molecules, the torsional changes should be reduced to small values (e.g. ±2–5°) or even set to zero, as torsional changes in one part of the molecule may result in large motions in another. This does not jeopardize the procedure as the main effects are obtained by changes in (rigid-body) rotation and translation. By default, the molecular center of mass is used as center of global rotation during the Monte-Carlo search. If the molecule binds to an asymmetric pocket or is coordinated to a metal center, this might not lead to the desired results. In such cases, the center of rotation should be manually selected (after defining the MC parameters) using the *move molecule* button. The new center of rotation will then be marked by a blue dot:



An alternative consists in the Boltzmann sampling option. Here (after refinement of the active zone), the Metropolis criterion is applied. The sampled structures are written to disk and named according to the corresponding Monte-Carlo round in which they were obtained, e.g. MC-1.pdb_ext, MC-7.pdb_ext, MC-44.pdb_ext. etc. A frequency of 10 (default) would seem to be reasonable. A movie file (to be viewed by any corresponding software, e.g. *VMD*) may optionally be written. Note that an individual temperature is assigned for the Boltzmann sampling.

Non-bonded cutoff: *Yeti* uses two switches to allow for a smooth cutoff for non-bonded interactions between the <switch-on> distance and the <switch-off> distance (cf. Brooks, B.R. et al. *J. Comp. Chem.* **1983**, *4*, 187–217). They allow for a smooth cutoff for non-bonded interactions between the <switch-on> distance and the <switch-off> distance. Separate the two switch distances by 0.5 Å to 2.0 Å. For example, set the <switch-on> distance to 9.5 ≈ and the <switch-off> distance to 10.0 Å. Note, that <switch-off> must always be larger than <switch-on>. New defaults can be saved in the file *yeti.par*. It is further suggested to set (El) >(vdW) >(H-Bond). For modeling protein-bound metal ions, (El) should be set to 15.0 Å if the metal-center function (cf. above) has not been activated.



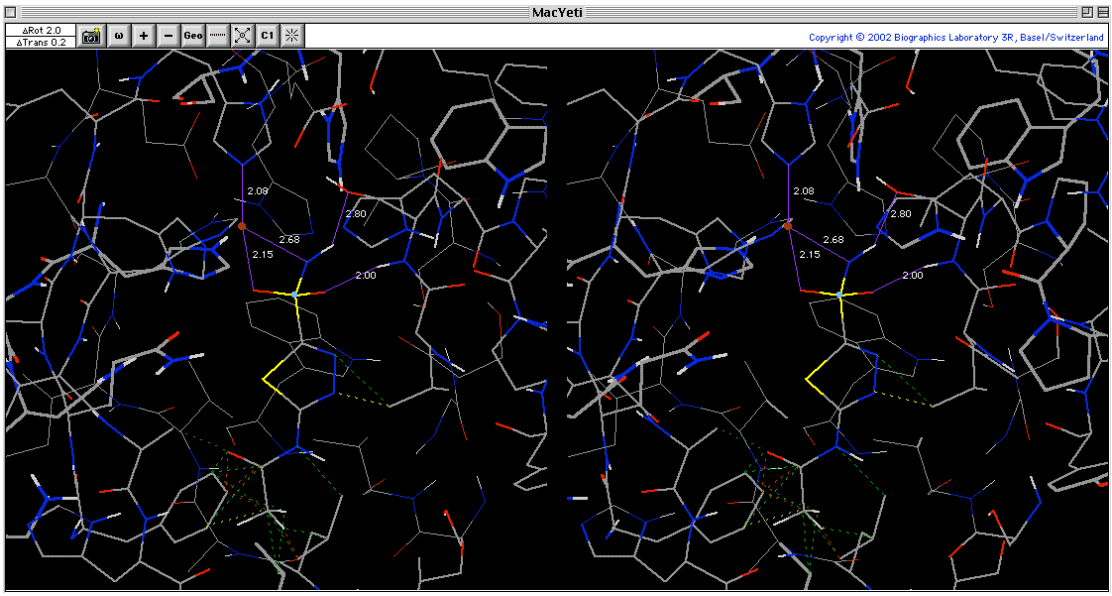
Apart from the energy, the *rms* derivatives (for torsional, rotational, and translational degrees of freedom, respectively) reflect the actual state of refinement. If all *rms* derivatives drop below the convergence values, the program will terminate. The *rms* convergence derivatives are defined in kcal/(mol·deg) for torsional and rotational degrees of freedom and in kcal/(mol·Å) for translational degrees of freedom, respectively. Lowest accepted values are 0.001 kcal/(mol·deg) for torsional degrees of freedom, 0.005 kcal/(mol·deg) for rotational degrees of freedom, and 0.010 kcal/(mol·Å) for translational degrees of freedom, respectively. These values are defined in the file yeti.par.

The convergence energy is used as the second termination criteria: if the change in energy of two consecutive cycles is smaller than the convergence energy, the program will terminate. Smallest accepted value is 0.0001 kcal/mol. Again, this values is defined in the file *yeti.par* but may be altered in the minimization dialog.

Maximal torsional, rotational, and translational shifts constitute the third convergence criterion. The program will terminate if no shift is larger than the specified value. These values, too, are defined in the file yeti.par.

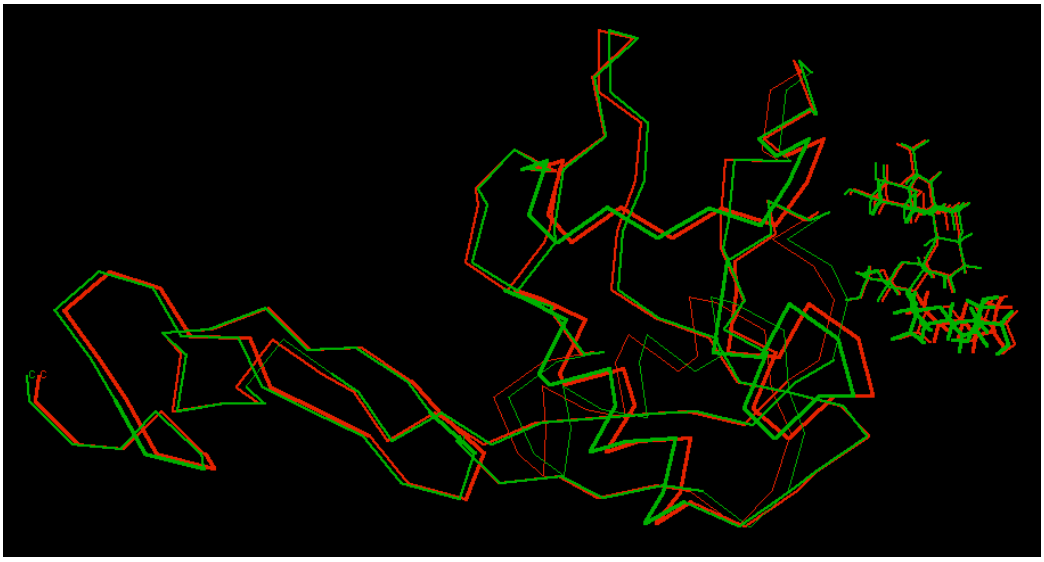
Docking

The docking of a small molecule can be interactively performed using multi-color bump-check and, optionally, dynamic mimization. During the docking process, all interatomic distances are updated. Dynamic docking allows to interactively change position and orientation of the small molecule during energy minimization. The new position is accepted "on the fly" and minimization continued therefrom. This option is disabled during a Monte-Carlo search.



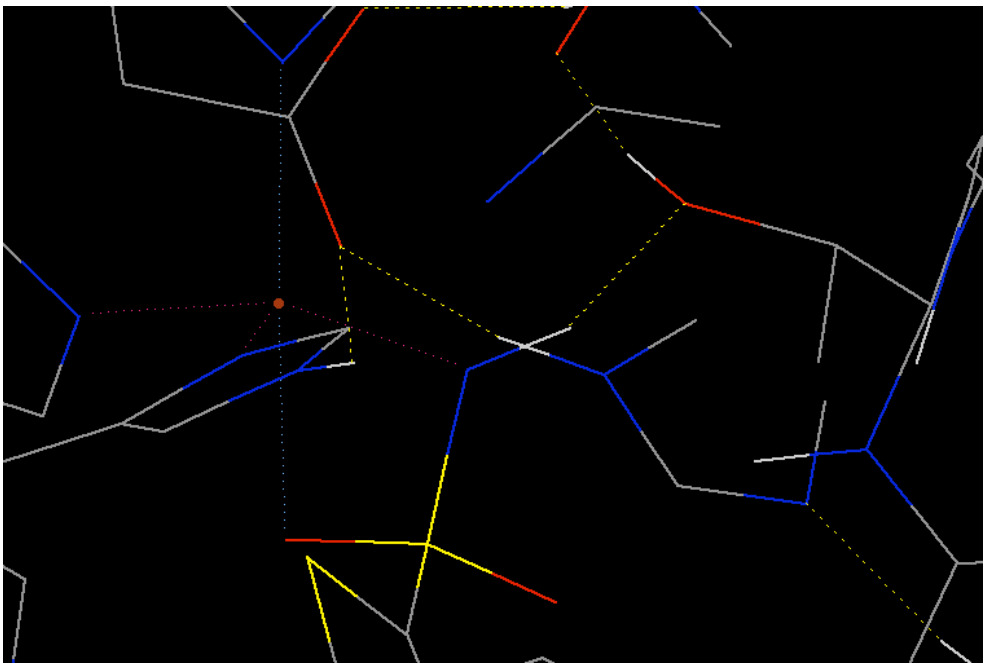
Display options

MacYeti currently includes eight display types: four "line options" [I: colored by atom; II: colored by groups (protein; red; ligands: green; solvent: blue; metals: light blue); III colored by derivatives (during minimization); IV: alpha-carbon tracing], ball-and-stick, two cpk as well as two stereo representations: side-by-side and anaglyph (requires red/green or red/blue goggles).

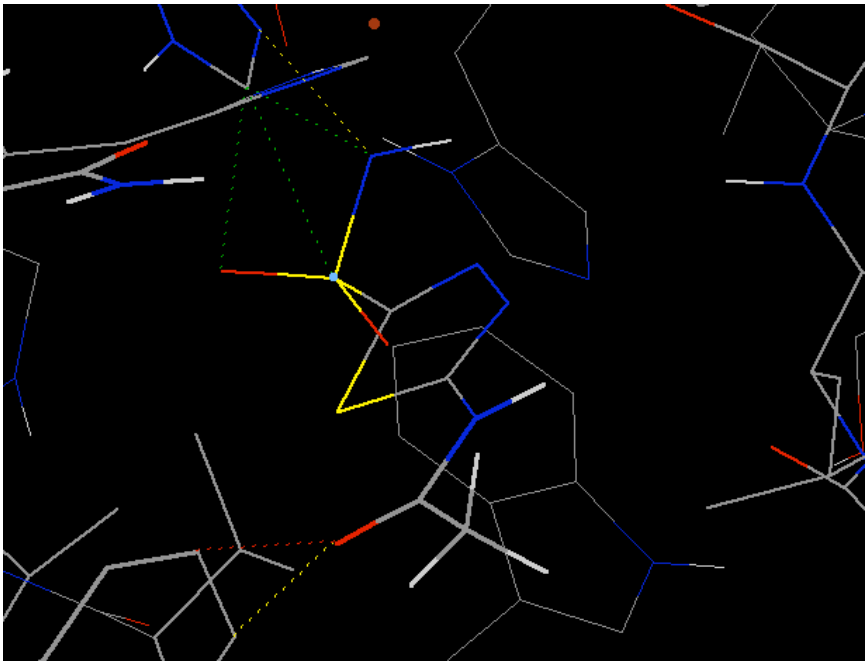


Other tools

Other useful display tools include the dynamic H-bond option and the multi-color bump check. Both can be used during docking and minimization. As they are rather cpu intensive, it is suggested to turn them off when computing speed is an issue. Labels that can be toggled on and off include residue labels and atomic partial charges. Any residue can be located with the "find residue"function in the display menu: the structure automatically centers and zooms on this residue. Clipping (i.e. slicing) and deleting residues are self-explainable functions. Procedures requiring user actions specifically ask for input in the frame bar.



H-bond interactions (shown as yellow dashed lines)



Multi-color bump check: red > yellow > green

The Yeti parameter file

This file includes all force-field parameters as well as some refinement settings:

```
CUT:
        switchee        7.5  8.0          ! switch-on switch-off cut-off for electrostatic interactions
        switchvdw       5.5  6.0          ! switch-on switch-off cut-off for vdW interactions
        switchhb        4.0  4.5          ! switch-on switch-off cut-off for H-Bond interactions
        switchmet       6.5  7.0          ! switch-on switch-off cut-off for metals
        dielectric      2.0               ! dielectric factor (do not change this value)
        fact14          0.5               ! factor to multiply 1-4 interactions
END

CONV:
        convtor         0.05               ! Convergence criteria for torsions
        convrot         0.05               ! Convergence criteria for rotations
        convtrans       0.5                ! Convergence criteria for translations
        shifttor        0.5                ! Shift criteria for torsions
        shiftrot        1.0                ! Shift criteria for rotations
        shifttrans      0.01               ! Shift criteria for translations
        convenerg       0.05               ! Convergence criteria for energy gain
        conhgrad        50                 ! Switch to conjugate gradient after # of cycles
        update          5000.0             ! Energy change before updating NB list
        maxcycle        100                ! maximum number of refinement cycles
        montecarlo      15.0 30.0 1.0      ! Monte-Carlo maximum initial steps: tor/rot/trans
        ehv_weights     1.0 1.0 1.0 1.0 1.0  ! Weigths for E[Lig*] components: El/vdW/HBd/Met/Tor (w: >= 0.0)

END

MIN:
        stepsize        1.0
        minstep         0.1
        maxstep         8.0
        ntrial          3
END
```

Force-field parameters should only be altered with great care as the *Yeti* force-field has been calibrated with the default settings. In particular, this applies to the dielectric parameter [d(r) = 2r] and the 1–4 weight. The parameters in the MIN section should also not be changed as they influence the step-length optimizer.

The Yeti dat file

This file includes the color definitions as well as the initial settings for display:

```
DISPLAY:
        resolution      83              ! screen resolution in dpi (dots per inch)
        background      black           ! background color
        mode            stereo          ! display mode at startup time (mono or stereo)
        stereo          120             ! stereo separation in mm
END

ATOM COLORS:
        H       lightgrey
        C       mediumgrey
        N       blue
        O       red
        P       orange
        S       yellow
        F       yellow
        CL      brightgreen
        BR      brown
        I       violet
END
```

Refinement in Batch Mode

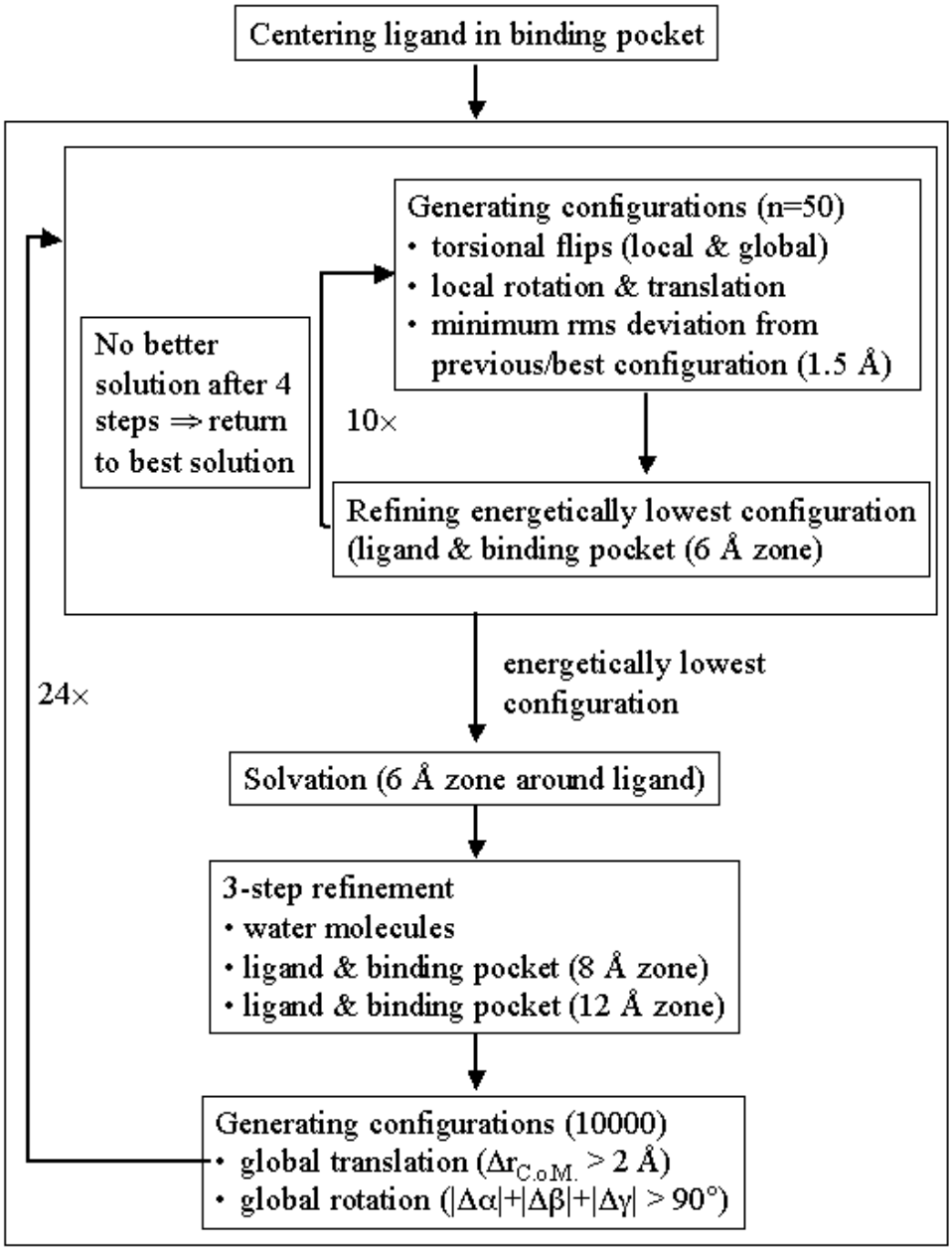
Yeti 7.0 allows to run any minimization interactively (Cmd+M) or in batch mode (Cmd+K). In batch, the program then asks for a directory name in which all necessary data files are prepared. Those include the coordinate input file (fixed name: *coord. inp*), the batch command file (*min.com*) and, if necessary, the file *zone.dat*. The output is written to the files *coord.out* (coordinates) and *yeti.log* (all info). All you have to do, is to drag *YetiBackground 7.0* (nongraphical executable) into this folder and double-click it. Of course, you may edit *min.com* or reuse it, for example, to run a new job. For the latter, make sure that you rename *coord.out* and *yeti.log* beforehand as these files would otherwise be overwritten. Refinement in batch is 3–4 times faster and hence suggested for longer minimizations, e.g. for Monte-Carlo simulations.

Yeti 7.0 Batch command file

```
MIN
        1     8.000                    | zone refinement, radius
    7.500     8.000                    | electrostatic cutoffs
    5.500     6.000                    | van-der-Waals cutoffs
    4.000     4.500                    | H-bond cutoffs
    6.500     7.000                    | metal cutoffs
    2.000                              | dielectric factor
    0.500                              | weight of 1-4 nb
      100                              | number of cycles
    0.050                              | convergence energy
    0.050     0.050     0.500          | convergence derivatives tor/rot/trs
    0.500     1.000     0.010          | convergence shifts tor/rot/trs
       10  5000.000                    | initial SD cycles; nb list update
        0     0.000                    | variable freezing: 1 = yes, 0 = no; treshold
END

METALS
        0                              | number of metal centers
END

CHARGE TRANSFER
        0                              | affected metals
        1                              | 1 = active; 0 = inactive
        1                              | 1 = dynamic; 0 = static
        1                              | 1 = directional; 0 = radial
END

MONTE-CARLO
        1                              | active
    10000                              | temperature
       25                              | trials
    0.250                              | rms
      500                              | rounds
        4                              | reset
   60.000                              | initial torsion
    5.000                              | initial rotation
    0.250                              | initial translation
     3030                              | center of rotation
        1          1                   | Boltzmann sampling; movie file (1 = yes, 0 = no)
      300         10                   | temperature, frequency
       20                              | number of attenuated cycles
END
```

Automated Docking

Yeti 7.0 includes a special module which allows for the fully automated docking of a batch of compounds in the background: *Yeti_AutodockX_7.0*. This special executable runs under Unix (Macintosh) or Linux (PC) operating systems. In contrast to other approaches in the field, *Yeti_AutodockX_7.0* allows for flexible docking and dynamic solvation: The former implies that the binding pocket (typically 8–12 Å around the small molecule) is co-optimized during the Monte-Carlo protocol. Dynamic solvation refers to the refreshing of the solvation shell around the ligand molecule during each Monte-Carlo step. With a *Unix script*, a directory tree can be generated and populated with individual molecules extracted from a single file. A second script controls the automated docking in batch. *Yeti_AutodockX_7.0* generates a 4D data set — an ensemble of orientations and conformations — suited as input for both the *Quasar* and *Raptor* technology.



To identify the most probable binding modes, *Yeti_AutodockX_7.0* samples structurally and energetically feasible arrangements within the binding pocket using flexible docking combined with an extended Monte-Carlo search protocol (cf. scheme on page 28) based on a Metropolis criterion and by allowing an adaptation of any protein side chain (typically) within 12 Å of the ligand molecule.

After centering a compound in the binding pocket, 50 [default] new ligand configurations are generated with local translation, rotation and combined local/global torsional changes. The lowest-energy entity out of these is refined and serves as template for the next Monte-Carlo step. If no lower-energy configuration is found after 4 [default] Monte-Carlo trials, the best previous configuration is restored. This permits a global searching without spending too much time for sampling ligand configurations in unfavorable regions of the binding pocket. As water molecules are known to play an essential role in ligand binding — mediating protein-ligand interaction, solvent-accessible binding pockets — the lowest-energy configuration emerging from the Monte-Carlo search [default = 25 rounds] is solvated and refined using an extended active-site region. An algorithm based on the directionality of hydrogen bonds, previously developed for the systematic solvation of proteins (cf. pages 10–11 or *J. Am. Chem. Soc.* **1991**, *113*, 5860–5862), is used for solvating thee complex. This constitutes the first binding mode.

Thereafter, 10,000 [default] configurations are generated using global rotation and translation in order to identify a novel binding mode (cf. scheme on page 28). For, example, this algorithm allows to flip a molecule by 180 degrees about any axis. Then, a new round of Monte-Carlo search is initiated (cf. above) and the resulting configurations is checked against all other entities that have already been accepted (*rms* and symmetry check). This procedure is then repeated until the desired number of unique binding modes [default = 25] are collected. For each accepted configuration, the coordinates of the ligand-protein complex are retained. Moreover, the ensemble of sampled ligand structures are composed into a 4D ligand data set. For this purpose, the following files are generated:

1.	3D_Ensemble.pdb_low	lowest-energy configuration only
2.	4D_Ensemble.pdb_all	all identified configurations
3.	4D_Ensemble.pdb_geo	all gemetrically distinct configurations
4.	4D_Ensemble.pdb_srt	same as above but sorted based on energy
5.	4D_Ensemble.pdb_erg	all configurations within a given energy window
6.	4D_Ensemble.pdb_sel	lowest *n* configurations [typically 4–8]

When automatically docking a batch of ligands, the corresponding 4D data files can be combined in a single file using the [Unix] *cat* command:

```
cat */4D_Ensemble.pdb_sel > All_Ligand_Ensemble.pdb_sel
```

which would generate the file All_Ligand_Ensemble.pdb_sel with all ligands present. This file can then be automatically processed by the *Quasar* or *Raptor* software. For more details, please refer to the corresponding example.

Yeti_AutodockX_7.0 processes the file *yeti.ctrl* which includes all pertinent information controlling the automated docking. Extract:

```
AUTO_DOCK:                                                          default
10          | sampling rounds                                       25
10000       | flip orientation: trials                              10000
120.0       | flip orientation: max. rotation each                  120.0
0.75        | flip orientation: min. rotation (frac. of max.)       0.75
2.0         | flip orientation: max. translation                    2.0
2.0         | minimal rms deviation for geometric acceptance        2.0
10.0        | energy window for structure acceptance                10.0
4           | select the "n" lowest-energy orientations             4
0.25        | fraction of global search during MC sampling          0.25
0.5         | fraction of local search during MC sampling           0.5
0.0         | max. constraint [kcal/mol] for grid outliers          100.0
2           | number of preferable hydrogen bonds                   0
2.0         | H-bond amplification                                  2.0
439         | H-bond constraint atom #1 (atom number)               integer
2041        | H-bond constraint atom #2 (atom number)               integer
END

MONTE_CARLO:
10000.0     | T for Boltzmann criterion                             10000
6.0         | zone refinement radius                                6.0
250         | max. refinement cycles/round                          250
0.01        | convergence energy                                    0.1
50          | trials/refined structure                              50
0.5         | min. rms distance from any other structure            0.5
1           | rounds per sampling round                             10
4           | global search reset                                   4
15.0        | torsional fluctuation                                 15.0
90.0        | rotational fluctuation                                90.0
1.0         | translational fluctuation                             0.5
10          | attenuated refinement (ligand)                        10
1.0         | weight of E(elec) for E[Lig*]                         1.0
1.0         | weight of E(vdW) for E[Lig*]                          1.0
1.0         | weight of E(Hbd) for E[Lig*]                          1.0
1.0         | weight of E(Met) for E[Lig*]                          1.0
1.0         | weight of E(tors) for E[Lig*]                         1.0
1.0         | weight of E(Lig-Prot) for E[Lig*]                     1.0
0.0         | weight of E(Lig,internal) for E[Lig*]                 0.0
0.0         | weight of E(Lig-Solv) for E[Lig*]                     0.0
END

MINIMIZING_A:
6.0         | zone solvation radius                                 6.0
8.0         | zone refinement radius                                8.0
100         | max. refinement cycles/round                          100
0.01        | convergence energy                                    0.05
END

MINIMIZING_B:
12.0        | zone refinement radius                                8.0
100         | max. refinement cycles/round                          100
0.01        | convergence energy                                    0.05
END
```

```
METAL:
            0           | metal? (1=yes, 0=no)
END
```

When modeling metal-bound ligand molecules, the charge-transfer function is automatically activated. The metal-center may be enabled as well but this would require a different input for each ligand. Such input can be produced with the graphical version or with a Unix script file. Example:

```
METAL:
1           | metal? (1=yes, 0=no)
1           | activate MC function
th          | coordination type
0.5         | weight of metal center
0.0         | LFSE
0.0         | ave. length of M-L bond
0.0         | height above basal plane
510         | metal center
179         | equ lig. #1
203         | equ lig. #2
253         | equ lig. #3
519         | equ lig. #4
0           | axi lig. #1
0           | axi lig. #2
END
```

Please note that for running *Yeti_AutodockX_7.0*, all files must feature Unix line endings. Of course, all provided files comply with this condition. Any other files may be changed using corresponding software (many of which freely available on the Internet) or with the following Unix command:

Mac to Unix: *tr '\r' '\n' < input_file > output_file*

Description of the parameters in the *yeti.ctrl* file

- sampling rounds: number of different orientations to be identified (max = 25).
- flip-orientation trials: number of attempts to identify a unique new orientation within a given round.
- maximal rotation: max. rotation about the x, y, and z axis. The effective rotation is generated through a random number.
- fraction of maximum: minimal sum of the rotations about all three axes to be accepted. Example: If max=120° and frac=0.75, then, the sum of all rotations must be larger than 90°.
- max. translation: since pure rotation may not identify a new, low-energy binding mode, translation along the three axes is useful
- minimal *rms* for geometrical acceptance: defines the "uniqueness" of a binding mode. Nonetheless, the file *4D_Ensemble.pdb_all* stores all binding modes, regardless of this criterion
- energy window: defines the content of the file *4D_Ensemble.pdb_erg*.
- selection: selects the *n* energetically lowest orientations; defines the content of the file *4D_Ensemble.pdb_sel*.
- fraction of global search: during global search, only the minimum conformations (± 15° [default]) are sampled — 0/180° for sp^2–sp^2 and +60/–60/180 for sp^3–sp^3 — thus, significantly reducing the cpu time spent during minimization. In global search mode, the low-energy minima are preset randomly.
- local search: as global search but with the difference that the closest minimum to the actual conformation is preset.
- free search: the search is executed about the current position. Its frequency is calculated as f(free) = 1.0 – f(global) – f(local).
- grid constraint: A grid is defined about the template molecule (which is otherwise solely used to "identify" the binding pocket). If a new molecule is found to lie completely outside this grid, the full penalty is applied — for smaller deviations, a fraction thereof. This option might be useful when docking to relatively open binding sites. In the final minimization round, this constraint is not applied.
- H-bond amplification: this option allows to put emphasis on preferred H-bonding groups at the protein to "saturated" during the Monte-Carlo search. Up to six such interactions may be defined and overweighted according to the amplification factor. The *integer* refers to the atom (line) number of the specific H-bond donor or acceptor in the input file. In the final minimization round, this enhancement is not applied.

- Monte-Carlo temperature: defines the *Metropolis criterion*.

- zone refinement radius: amount of the binding pocket (about the current ligand position) to be included in the refinement = flexible docking. If any atom of an amino-acid residue lies within this distance of any atom of the ligand molecule, the whole amino-acid residue is optimized.

- max. minimization cycles during a single Monte-Carlo step.

- convergence energy in kcal/mol.

- number of trials per refinement step.

- total number of Monte-Carlo rounds to identify a single orientation.

- global reset: defines global or local search (min = 1).

- torsional fluctuation: randomly controlled variation about the current torsion (or preset value in global and local mode; cf. above).

- rotational fluctuation: global rotation about the current orientation.

- translational fluctuation: global translation about the current position.

- attenuated refinement (of the small molecule): to avoid the newly attained position (child) simply be minimized back to the original position (parent), the refinement of the small molecule may be attenuated for a finite number of cycles. This allows the protein to adapt its conformation to the small molecule (induced fit). The default of *n=10* implies that during the first optimization round, the calculated shifts are reduced to 10%, in the second round to 20% etc. After 10 rounds, free refinement occurs.

- weights I: for the selection of the binding mode during the Monte-Carlo search, individual contributions to the total energy may be weighted. The corresponding results are listed as E[Lig*]. In the final minimization rounds, unit weights are applied. Do not set w(van der Waals) to 0.0; you may wish to enhance E(Hbds) or E(int) to improve hydrogen bonding or minimize internal strain, respectively. Please note that these weights are not used for minimization purposes — here, the pure force field is applied — but solely for selecting an individual orientation.

- weights II: these control the classes of interactions (ligand-protein, ligand-solvent, ligand-ligand) which shall be considered for selecting an individual orientation.

- zone solvation radius: solvation of the binding pocket around the small molecule.

- zone refinement radius I: first round of unconstrained minimization.

- zone refinement radius II: second round of unconstrained minimization.

Setup environment for analyzing a batch of molecules

First, create a directory *Template* and populate it with the following files:

1. *coord.inp* the protein plus a small-molecule template (pdb_ext format)

2. *MacYetiLibrary* just a copy of the standard library

3. *yeti.ctrl* the control file, cf. above

4. *GetSummary* optional; allows to retrieve the most important information from the yeti.log and other files:

```
head -72 yeti.log > Autodock_Summary
echo "" >> Autodock_Summary
fgrep Structure MC* >> Autodock_Summary
echo "" >> Autodock_Summary
fgrep "rms to" yeti.log >> Autodock_Summary
echo "" >> Autodock_Summary
fgrep "ligand ori" yeti.log >> Autodock_Summary
echo "" >> Autodock_Summary
fgrep "symmetry flag = " yeti.log >> Autodock_Summary
echo "" >> Autodock_Summary
fgrep Orientation yeti.log >> Autodock_Summary
echo "" >> Autodock_Summary
fgrep Accepting yeti.log >> Autodock_Summary
echo "" >> Autodock_Summary
fgrep Rejecting yeti.log >> Autodock_Summary
echo "" >> Autodock_Summary
```

Next, generate the file *SetupJob* as follows:

```
cp -r Template L01
fgrep L01 Lig108.pdb_ext > ./L01/Ligand.pdb_inp

cp -r Template L02
fgrep L02 Lig108.pdb_ext > ./L02/Ligand.pdb_inp

cp -r Template L03
fgrep L03 Lig108.pdb_ext > ./L03/Ligand.pdb_inp

cp -r Template L34
fgrep L34 Lig108.pdb_ext > ./L34/Ligand.pdb_inp

cp -r Template L35
fgrep L35 Lig108.pdb_ext > ./L35/Ligand.pdb_inp

cp -r Template L36
fgrep L36 Lig108.pdb_ext > ./L36/Ligand.pdb_inp

.
.

cp -r Template X38
fgrep X38 Lig108.pdb_ext > ./X38/Ligand.pdb_inp
```

For each of the ligands in the file *Lig108.pdb_ext*, one corresponding entry must be present in *SetupJob*.

Then, make it executable: *chmod +x SetupJob* and execute it: *./SetupJob*. This generates and populates one directory per compound and copies the coordinates of the very ligand into the file *Ligand.pdb_inp*.

Next, generate the file *MasterRunYeti* correspondingly:

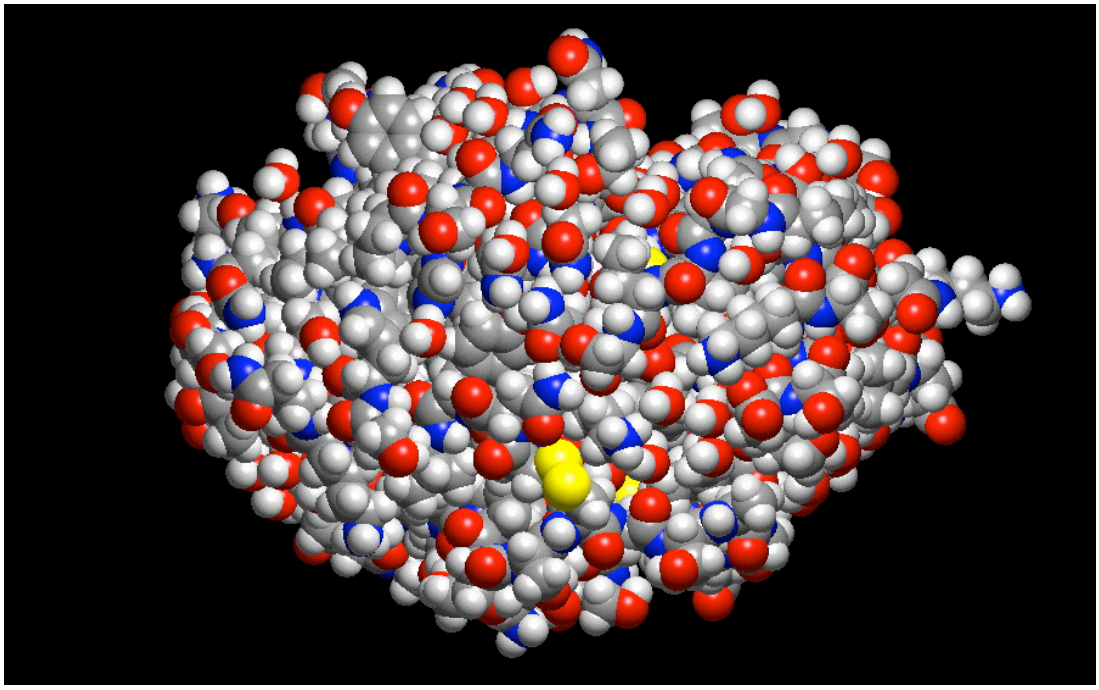
```
cd L01
../MacYeti_AutodockX_7.0.x
./GetSummary
rm yeti.log

cd ../L02
../MacYeti_AutodockX_7.0.x
./GetSummary
rm yeti.log

cd ../L03
../MacYeti_AutodockX_7.0.x
./GetSummary
rm yeti.log

cd ../L34
../MacYeti_AutodockX_7.0.x
./GetSummary
rm yeti.log

cd ../L35
../MacYeti_AutodockX_7.0.x
./GetSummary
rm yeti.log

cd ../L36
../MacYeti_AutodockX_7.0.x
./GetSummary
rm yeti.log

.
.
.

cd ../X38
../MacYeti_AutodockX_7.0.x
./GetSummary
rm yeti.log
```

Make it executable: *chmod +x MasterRunYeti* and execute it: *./MasterRunYeti*.

After completion, you may combine all ligand output files into a single file, e.g.

*cat */3D-Ensemble.pdb_low > ../All.pdb_low*
*cat */4D-Ensemble.pdb_sel > ../All.pdb_sel*

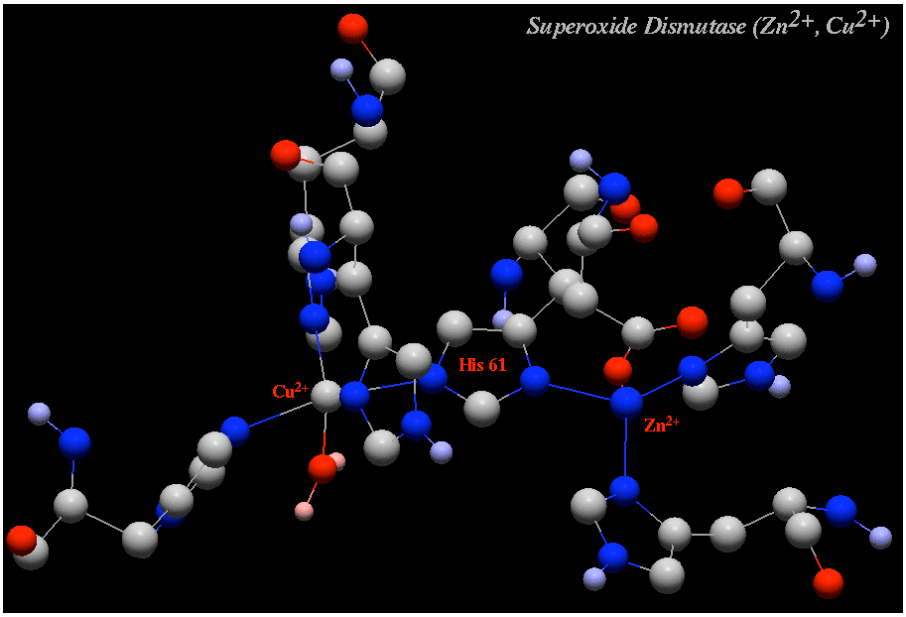
Example I: Bovine trypsin



Protocol: Read 3ptn.pdb, Generate polar hydrogens, Orient, activate MC function {octahedron, w=0.75, LFSE=0.0}, Select zone refinement mode {6.0 Å around metal}, Minimize, Write (ext.pdb) file (e.g. 3ptn.out). Results:

```
Ligand-metal charge transfer: CA2+ 2045
---------------------------------------
Lig:  468  OE1 GLU    70  r: 2.356 dLp:  57.6 CT->M:  0.004 q(L): -0.702
Lig:  491  O   ASN    72  r: 2.239 dLp:  51.4 CT->M:  0.017 q(L): -0.483
Lig:  519  O   VAL    75  r: 2.397 dLp:  42.1 CT->M:  0.017 q(L): -0.483
Lig:  561  OE2 GLU    80  r: 2.325 dLp:  44.0 CT->M:  0.014 q(L): -0.692
Lig: 2088  O   WAT   711  r: 2.433 dLp:  63.2 CT->M:  0.003 q(L): -0.597
Lig: 2208  O   WAT   714  r: 2.326 dLp:   3.6 CT->M:  0.051 q(L): -0.549
Met: 2045  CA2+HET   480                      CT<-L:  0.106 q(M):  1.894

Final energy : -2.239855e+03 kcal/mol
-------------------------------------
1-4 El :   1.295960e+03 | 1-4 vdW:   2.824332e+02 | Torsion:   2.490372e+02
nb El  :  -2.662757e+03 | nb vdW :  -8.518734e+02 | H-Bond :  -5.212571e+02
CA2+ 2045: E[rad]: -29.778, E[dir]:  -1.646, Symm=oh : 0.416 - Dirc: 0.348

Refinement summary
------------------
Optimization cycles         :              37
Improvement of total energy :         -72.797 kcal/mol
Change of 1-4 El  energy    :          -0.144 kcal/mol
Change of 1-4 vdW energy    :          -4.553 kcal/mol
Change of Torsion energy    :          -5.014 kcal/mol
Change of nbd El. energy    :         -57.284 kcal/mol
Change of nbd vdW energy    :           7.288 kcal/mol
Change of H-bond  energy    :         -13.447 kcal/mol
Change of Met-Lig energy(1) :          -1.016 kcal/mol
Change of Met-Ctr energy(1) :           1.375 kcal/mol
```

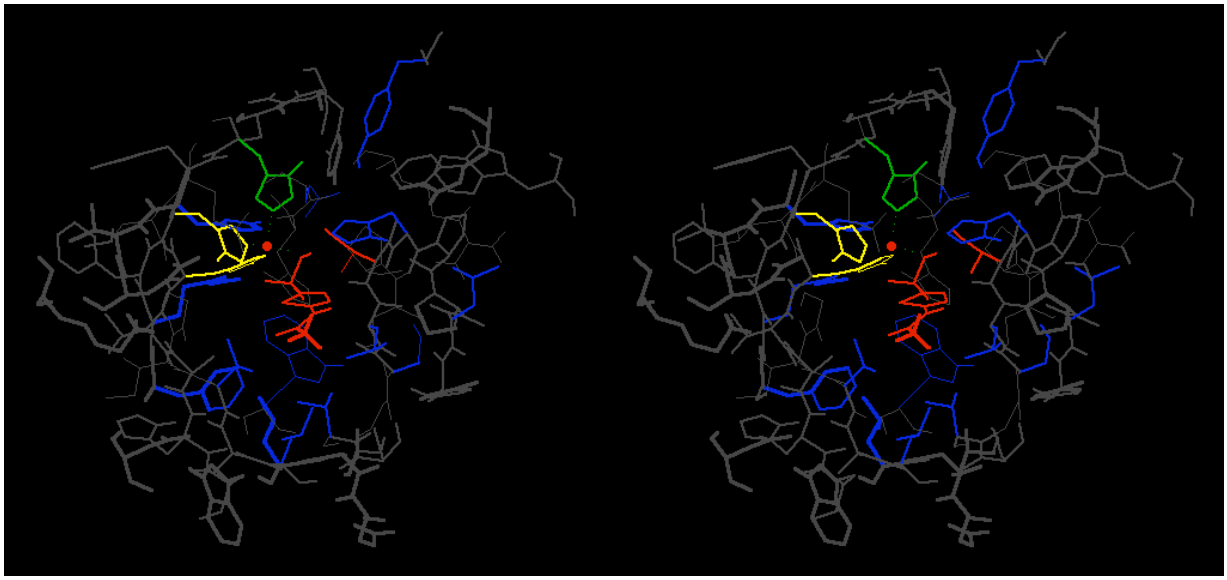
Example II: Cu/Zn-Superoxide Dismutase



Protocol: read sod.pdb, activate MC function {Cu: square pyramid; mark His 46 as the axial ligand (by clicking the button 4 times), w=0.5, LFSE=27.64; Zn: tetrahedron, w=0.5, LFSE=0.0}, Minimize, Write file (e.g. sod.out). Results:

```
Ligand-metal charge transfer: CU2+   81
---------------------------------------
Lig:    6  ND1 HIE    44  r: 2.118 dLp:  16.7 CT->M:  0.149 q(L): -0.378
Lig:   21  NE2 HID    46  r: 2.613 dLp:   5.0 CT->M:  0.032 q(L): -0.495
Lig:   32  NE2 HIM    61  r: 2.080 dLp:  13.3 CT->M:  0.183 q(L): -0.344
Lig:   77  NE2 HID   118  r: 2.114 dLp:   6.0 CT->M:  0.175 q(L): -0.352
Lig:   83  O   WAT   121  r: 2.033 dLp:   4.9 CT->M:  0.207 q(L): -0.393
Met:   81  CU2+HET   119                      CT<-L:  0.747 q(M):  1.253

Ligand-metal charge transfer: ZN2+   82
---------------------------------------
Lig:   30  ND1 HIM    61  r: 2.024 dLp:  13.1 CT->M:  0.173 q(L): -0.354
Lig:   41  ND1 HIE    69  r: 2.018 dLp:   5.6 CT->M:  0.193 q(L): -0.334
Lig:   53  ND1 HIE    78  r: 2.042 dLp:   2.2 CT->M:  0.178 q(L): -0.349
Lig:   65  OD1 ASP    81  r: 1.921 dLp:  35.3 CT->M:  0.094 q(L): -0.612
Met:   82  ZN2+HET   120                      CT<-L:  0.638 q(M):  1.362

Final energy : -3.373294e+02 kcal/mol
-------------------------------------
1-4 El :   1.677086e+01 | 1-4 vdW:   5.702437e+00 | Torsion:   4.190165e+00
nb El  :  -1.069598e+02 | nb vdW :  -2.075614e+01 | H-Bond :   1.930751e-01
CU2+ 81: E[rad]: -60.552, E[dir]: -89.705, Symm=sqp: 0.946 - Dirc: 0.938
ZN2+ 82: E[rad]: -47.830, E[dir]: -38.383, Symm=th : 0.940 - Dirc: 0.830

Refinement summary
------------------
Optimization cycles         :              27
Improvement of total energy :        -117.558 kcal/mol
...
Change of Met-Lig energy(2) :           0.332 kcal/mol
Change of Met-Ctr energy(2) :          -8.118 kcal/mol
```

Example III: Carbonic-anhydrase bound acetazolamide (Monte-Carlo search)



Protocol: read CAB+AAA-MC.start activate MC function {Zn: tetrahedron, w=0.5, LFSE=0.0}, select zone refinement (6Å around the small molecule), activate Monte-Carlo search (25 rounds, 10 trials, rms=0.5Å, reset=4), first select the active molecule by clicking it, then o.k. Next, select the center of rotation (the deprotonated N atom) by using the "select molecule button", Minimize, Write file (e.g. CAB+AAA-MC.out).

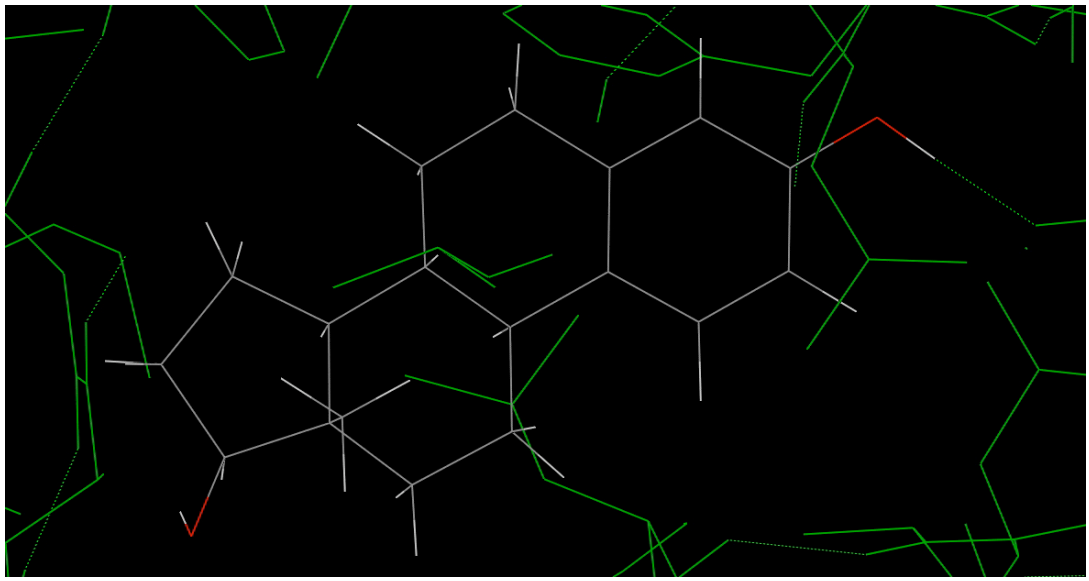
Results:

```
.
.
.
Monte-Carlo: restoring best coordinate set (cycle 10): E[tot] = -598.114;
             rms from starting structure = 0.222

Ligand-metal charge transfer: ZN2+  510
---------------------------------------
Lig:  179  NE2 HID    94  r: 2.063 dLp:   7.5 CT->M:  0.152 q(L): -0.375
Lig:  203  NE2 HID    96  r: 2.084 dLp:   2.9 CT->M:  0.144 q(L): -0.383
Lig:  253  ND1 HIE   119  r: 2.026 dLp:  10.5 CT->M:  0.170 q(L): -0.357
Lig:  517  O   AAA   300  r: 2.175 dLp:  60.5 CT->M:  0.009 q(L): -0.288
Lig:  519  NM  AAA   300  r: 2.065 dLp:  19.3 CT->M:  0.102 q(L):  0.000
Met:  510  ZN2+HET   261                      CT<-L:  0.577 q(M):  1.423

Final energy : -5.981141e+02 kcal/mol
-------------------------------------
1-4 El :   2.172855e+02 | 1-4 vdW:   5.584037e+01 | Torsion:   4.041453e+01
nb El  :  -4.958091e+02 | nb vdW :  -2.586570e+02 | H-Bond :  -5.413934e+01
ZN2+ 510: E[rad]: -59.336, E[dir]: -43.713, Symm=th : 0.911 - Dirc: 0.934

Energy Partitioning
-------------------
Protein-Protein:       -5.219266e+02
Protein-Ligand(s):     -7.292951e+01
Ligand(s)-internal:    -3.259020e+00
Protein-Solvent:        0.000000e+00
Solvent-Solvent:        0.000000e+00
Ligand(s)-Solvent:      0.000000e+00
```

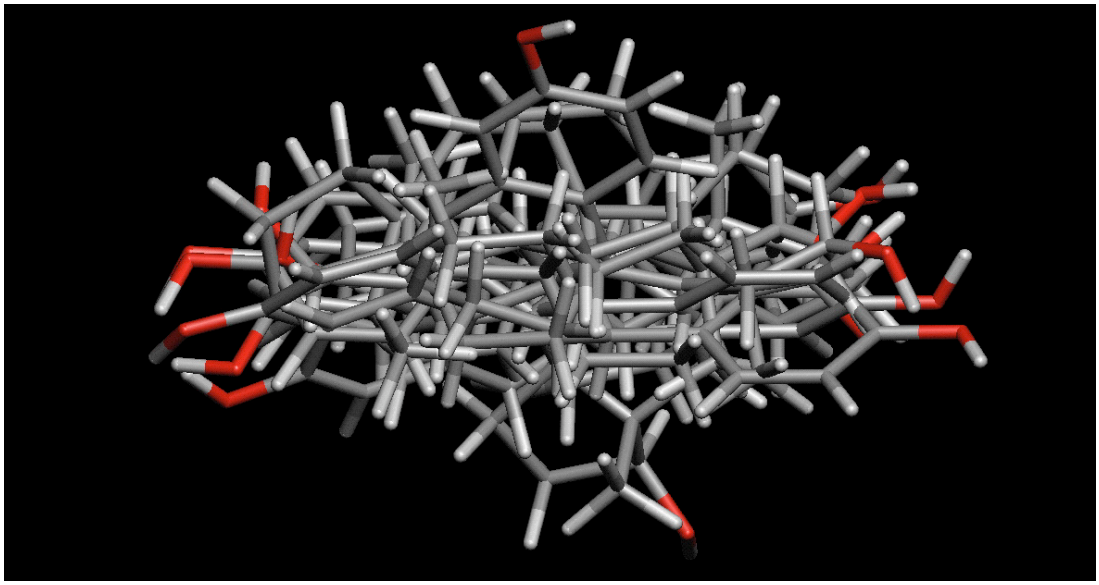
Example IV: Small-molecule binding to the estrogen receptor (Autodocking)

Simply execute the batch command file (*yeti.ctrl*) found in the folder *Example IV* by opening a Unix shell and starting the program as follows:

./YetiBackgroundX &



Resulting lowest-energy configuration



Resulting 4D ensemble

Please note that the graphical version of *Yeti* cannot by default display and process 4D data sets. Please use *Bio*, *Quasar*, *Raptor* or *PrGen* for this purpose.

6 References

Primary reference: Vedani, A. and Huhta, D.W.
A new force field for modeling metalloproteins.
J. Am. Chem. Soc. **1990**, *112*, 4759–4767.

Secondary references: Lill, M.A., Dobler, M., and Vedani, A.
Prediction of small-molecule binding to Cytochrome P450 3A4: Flexible docking combined with multi-dimensional QSAR.
ChemMedChem **2006**, *1*, 73–81.

Vedani, A. and Huhta, D.W.
An algorithm for the systematic solvation of proteins based on the directionality of H-bonds.
J. Am. Chem. Soc. **1991**, *113*, 5860–5862

Vedani, A., Huhta, D.W. and Jacober, S.P.
Metal coordination, H-bond network formation, and protein-solvent interactions in native and complexed human carbonic anhydrase.
J. Am. Chem. Soc. **1989**, *111*, 4075–4081

Vedani, A., Dobler, M. and Dunitz, J.D.
An empirical potential function for metal centers.
J. Comput. Chem. **1986**, *7*, 701–710

Vedani, A. and Dunitz, J.D.
Lone-pair directionality in H-bond potential functions for molecular-mechanics calculations.
J. Am. Chem. Soc. **1985**, *107*, 7653–7658