so the details of each method shall not be repeated here. Rather, the basic tasks that are common to most or all of these methods are presented, with a discussion of various alternatives for achieving a given task and, where applicable, any recognized standard in the field.

PREPARING LIGANDS

Identification of the correct spatial relationships among key ligand/receptor interaction points is ultimately governed by the accuracy of the structures from which pharmacophore models are derived. Thus once an appropriate set of ligands has been identified, realistic models of 3D chemical structure must be developed. Doing so requires not only a procedure for generating a set of low-energy conformers for each ligand but also decisions about ionization and tautomeric states and appropriate treatment of stereochemistry.

Methods for sampling the thermally accessible conformational states of a ligand is a topic that deserves far more attention than can be paid here, but in the context of pharmacophore methods, conformational sampling techniques are usually divided into two classes: those that are appropriate for pharmacophore model development and those that are appropriate for searching large 3D databases. Pharmacophore software frequently provides separate methods to address each of these situations, with thorough sampling and full minimization of structures being done[46,47] for pharmacophore model development and faster, less rigorous approaches[48–50] being used for large databases.

When developing a pharmacophore model from a set of actives, the goal of conformer generation is to produce an ensemble of structures that each ligand can adopt under biological conditions, with a granularity fine enough to ensure that at least one structure is reasonably close to the bioactive conformation. Whether this can be achieved depends on both the initial set of structures generated and the force field (or Hamiltonian) that's employed to minimize them. If the initial sample contains no structure sufficiently close to the bioactive conformation, it is unlikely that subsequent minimization will dramatically improve the situation. Conversely, even if the initial sample contains a reasonable facsimile of the bioactive structure, an inferior force field may drive that structure to a somewhat distant local minimum. As noted previously, pharmacophore packages frequently provide a means for conformer generation, but MacroModel MCMM (Monte Carlo Multiple Minimum)[46] is generally recognized as a standard for thorough exploration and sampling of conformational states along a given potential energy surface, while MMFF (Merck Molecular Force Field)[51] and OPLS (Optimized Potential for Liquid Simulations)[52] are routinely used for energetics and minimization.

A structure's stability and interactions with the receptor are affected by its ionic character, so identification of ionic centers in a ligand is an important consideration. When there is a priori knowledge about the correct ionization state (e.g., a particular secondary amine in the structure is known to be protonated), it is common practice to assign ionization states explicitly based on that knowledge. In other cases, software may be called on to either neutralize all ligands or assign the most probable ionization states based on a set of rules.[53] Whether one starts with neutral or ionized structures is less of an issue when the pharmacophore feature mapping procedure automatically recognizes ionizable centers even if they are expressed in neutral form.

It is worth noting that although docking software routinely generates multiple ionic states for each ligand, doing so within the context of common pharmacophore perception is not a trivial matter. The difficulty stems from the fact that each ligand and its conformers are normally associated with only a single connection table (i.e., the set of atoms and the bonds that connect them), which allows a single set of mappings to be defined between the atoms and the pharmacophore features in a given ligand. When additional ionic states are introduced, each has a different connection table, which requires additional sets of atom→feature mappings, not to mention additional sets of conformers. When perceiving common pharmacophores, then, somewhat arbitrary decisions may be required regarding which ionic form to report for a given parent ligand when a pharmacophore is matched equally well by child structures with different ionization states.

Tautomerization raises many of the same questions as ionization, although less attention is normally paid to this issue, in part because the most commonly recognized tautomeric states are usually just assumed (keto preferred over enol, amide preferred over amidic acid, etc.). There are certainly cases where the location of a proton may be less obvious (e.g., imidzoles, pyrazoles, and triazoles), so if a particular choice is made, the possible consequences should be considered. For example, if a pharmacophore model contains a hydrogen bond donor feature that maps to a proton whose location is in question, one may wish to give equal consideration to the corresponding model that results when the proton is moved to an equally probable location. As in the case of ionization states, a more general treatment of tautomers requires a means for dealing with different connection tables and their associated atom→feature mappings and conformers.

From a mechanical perspective, treating different stereoisomers of a given parent ligand is somewhat less challenging because their structures can share a single connection table. Consequently, their conformers may be readily combined and treated as a single ligand if necessary. Thus if a pair of enantiomers has been resolved and the activities of both isomers are known, they can, and should, be treated as separate ligands. But if the observed activity is based on a racemate, the most sensible approach is to merge the conformers from the two enantiomers. If one enantiomer is far less active than the other, as is often the case, and the data set contains other ligands whose stereochemistry is not in question, the structures of the