At its most basic, a pharmacophore model is just an arrangement of feature points whose relative locations are defined by a set of interpoint distances, internal coordinates, or Cartesian coordinates. A query is created only after conditions on matching the pharmacophore are stipulated. Most pharmacophore packages support user-defined tolerances on matching distances, positions, angles, and so on, but a number of general guidelines have been established that rely on a combination of experimental data and common sense.[20,89] For example, observed variations in hydrogen bond distances X–Y and hydrogen bond angles X–H–Y within crystallographic complexes may be used to conclude that the positional tolerance on matching hydrogen bond acceptors and donors should be about 2Å.[20] It is also possible to use known actives and inactives to derive suitable constraints on matching interfeature distances.[21] Positional tolerances between 1Å and 2Å for various types of features are typical, but much stricter criteria are sometimes used.[86,88]

When automated common pharmacophore perception is employed, it is tempting to argue that matching tolerances should be inferable from positional variations in the superimposed ligand features. However, such variations are really characteristics of the ligands themselves and of the conformational sampling method; they are not indicative of the receptor's flexibility, promiscuity, and so on. For example, consider a common pharmacophore model that is derived from a set of rigid, congeneric ligands. When the ligands are superimposed on the pharmacophore model, there should be essentially no variations in the feature locations from ligand to ligand. If a database query were then posed with matching tolerances consistent with those tiny variations, it is unlikely that any additional actives would be found, unless the database contained molecules with the same rigid framework. This sort of restriction eliminates the possibility of scaffold hopping,[90] an advantage that pharmacophore-based searching is naturally assumed to offer.

As shown in Figure 9.10, a point-based pharmacophore query may be embellished with any number of additional characteristics and constraints, such as a distance between a point and a plane, an angle between planes, or a cone of revolution about a hydrogen bond axis. However, before any of these conditions can be applied, a suitable match to the feature points must be found, which nearly always involves identifying sets of interfeature distances in a database structure that are consistent with the locations of the feature points in the pharmacophore model. Thus, the primary criterion for a match is that a structure must contain all n features in the pharmacophore model and that a particular mapping of those features to the pharmacophore yields an $n \times n$ distance matrix whose elements are sufficiently close to the corresponding elements in the pharmacophore distance matrix. As noted previously, the user normally stipulates how closely the points must match, either by specifying tolerances on the interfeature distances themselves

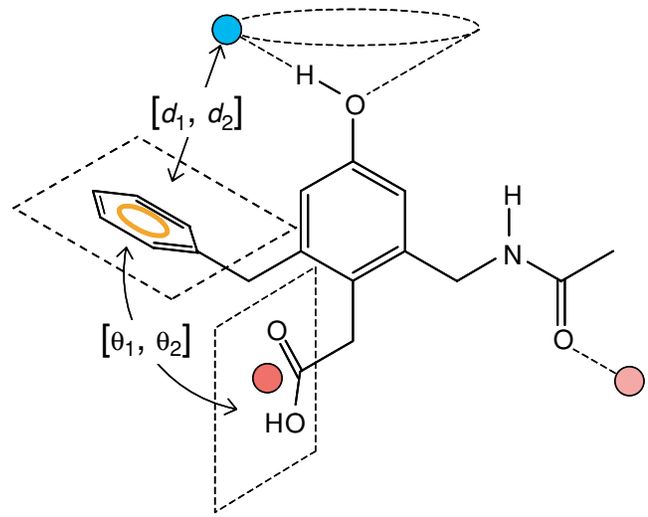

Figure 9.10. A four-point pharmacophore model with additional constraints and features. The angle between two planes and the distance between a hydrogen donor and a plane must lie within specified ranges, while the hydrogen bond donor sweeps out a cone of revolution.

or by specifying positional tolerances on the feature points after least squares alignment.[58,59]

Many pharmacophore packages allow *partial matching*, wherein only m of n points in a query must be matched. There may be user-imposed requirements to match specific points, or matching any subset of m may be sufficient. In either case, the algorithm must be modified to cycle through different subsets of m points in the pharmacophore model and attempt matching on the associated $m \times m$ submatrices. Partial matching is frequently invoked out of necessity when the pharmacophore model contains more points than can reasonably be expected to be matched by any molecule that does not contain the same chemical scaffold as the ligand(s) from which the model was derived. Thus a "kitchen sink" approach may be taken when developing a pharmacophore model, with the database screen being used to determine which parts of the model actually occur in other molecules. This convenience comes with a price, though, because the combinatorics of matching m of n points is governed by the binomial coefficient $n!/[m!(n-m)!]$, so when $m \approx n/2$, a search can be exceedingly slow if n is too large. Furthermore, certain subsets of feature points may correspond to very ordinary pharmacophores, which may cause an inordinate number of database molecules to be matched.

In practice, the matching algorithms just described are normally invoked only after performing one or more rapid prescreens to eliminate molecules that cannot possibly satisfy the query. A prescreen may involve only 2D criteria, such as rejection of molecules that are missing any required feature in the pharmacophore, or it may be 3D in nature. If the database contains precomputed conformers and their associated pharmacophore fingerprints, the strategy described in the previous section may be employed to