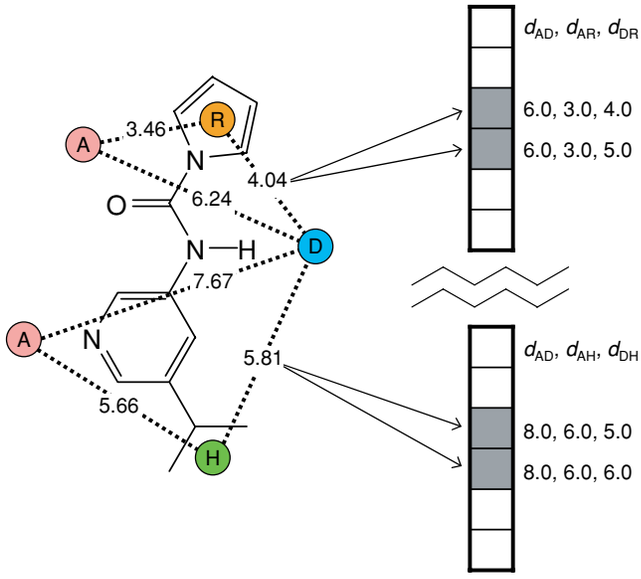

Figure 9.9. Encoding of pharmacophore triplets into a bit string with distance bins that are 1Å wide. To prevent loss of information due to the partitioning of distances, each triplet sets not only the bit whose midpoint is closest to the distances in that triplet but also neighboring bits in each distance dimension. For simplicity, neighboring bits are shown for only the last dimension in each triplet.

PHARMACOPHORE FINGERPRINTS

A summary of the pharmacophoric information contained in a structure may be represented in the form of a bit string that encodes the presence, absence, and sometimes the counts of various arrangements of pharmacophore features. These bit strings are referred to by various names, including pharmacophore fingerprints, pharmacophore keys, and pharmacophore tuplets.[15–17] The CHEM-X software system[18] is most closely identified with this concept, though many other packages contain their own implementations.

Pharmacophore fingerprints are usually based on either triplets of features (three-point pharmacophores) or quartets of features (four-point pharmacophores). Each bit in a fingerprint corresponds to a particular set of n pharmacophoric feature types (acceptor, donor, hydrophobe, etc.) separated by a specific set of $[n(n-1)]/2$ interfeature distances. To limit the number of distinct bits, the distance coordinate is divided into bins of predefined width, so a range of distances, and thus a range of pharmacophores, is mapped to a single bit. The bins may be of unequal width to allow greater discrimination in certain regions of the distance coordinate, and the upper limit bin may be defined to include all distances greater than some value, or it may simply ignore distances above that value.[18]

Figure 9.9 illustrates the encoding of pharmacophore triplets into a bit string characterized by distance bins that are 1Å wide. To prevent loss of information due to the partitioning of distances, a given triplet sets not only the bit whose midpoint is closest to the distances in that triplet but also neighboring bits in each distance dimension. Observe that certain combinations of distances are physically impossible,[2,4,10] so many bits will never be set.

A pharmacophore fingerprint may be created from a conformational ensemble, so it represents the pharmacophore space that is conformationally accessible to a particular molecule. The fingerprint is normally a logical OR (i.e., the union) of the bits set by different conformers, so information to map individual bits back to their source conformers is not retained. Whether a pharmacophore fingerprint comes from a single structure or multiple conformers, it may be used in precisely the same manner as 2D bit string representations[76–79] in applications involving similarity, diversity, clustering, and so on.[80–84]

Pharmacophore fingerprints are perhaps most powerful when used in the context of 3D database screening. When a pharmacophore query is posed, the features and distances in that query can be translated into a fingerprint, with multiple bits being set, as necessary, to account for tolerances on matching the interfeature distances. The pattern of bits set by the query creates a necessary condition for matching the pharmacophore, so if fingerprints have been created for a database of molecules to be searched, very fast logical bit operations may be performed to rapidly eliminate molecules that cannot possibly match the pharmacophore. In general, not every molecule that satisfies the fingerprint query will actually match the pharmacophore, but a majority of false positives can be eliminated, which may result in a drastic reduction in overall searching time.

3D DATABASE SCREENING

By far the most common role a pharmacophore model ultimately plays is that of 3D database query. If the pharmacophore model accurately embodies the key interactions required for binding, molecules in a database that match the query, the so-called *hits*, should have a greater-than-average chance of being active. The degree to which this advantage is actually observed depends on a number of additional factors, including the matching criteria, the presence/absence of steric constraints, and the quality of the database.

Each molecule in a 3D database may be represented by a set of precomputed conformers or by a single, low-energy structure. In the latter case, conformers for a given molecule can be generated during the database screen using systematic or stochastic searching,[7,85,86] or the structure may be flexed in an attempt to fit the query.[87,88] The primary advantage of storing precomputed conformers is searching speed, which may be one to two orders of magnitude faster than when conformers are generated during the search. However, speed comes at the expense of disk space, which grows with the number of conformers stored, although not necessarily in a linear fashion due to compression of redundant data.