database programs that are available include ChemFinder from CambridgeSoft, ChemFolder from ACDLabs, ChemWindow from Softshell, and Aura-Mol from Cybula (110).

4 CHEMICAL PROPERTY ESTIMATION SYSTEMS

The design and screening of drug candidates is increasingly being conducted in *silico.* This is made possible by improvements in programs for property calculation and estimation. Here, the term property calculation refers to the generation of some topological (depending only on the 2D structure), topographical (depending on the 3D conformation), or physicochemical property of a molecule—directly from the structure. The term property estimation refers to the generation of some property as a function of other properties—either through a regression equation, a formula, neural network calculation, or some other indirect means.

The distinction between calculation and estimation is important because some properties, like molecular weight, polar surface area, molecular connectivity values, counts of chemical functional groups, partial charges, and other quantum mechanical descriptors, can be calculated precisely and de *novo* from the structure alone. Most of these properties have some fixed definition or algorithm that enables their calculation to be performed unambiguously, with little or no error. What error is present is usually systematic or deterministic. A second class of properties, including Log*P* and other additive-constitutive properties, may be calculated by fragment additivity with various correction terms. These properties differ from de novo properties because they are approximations to the true (sometimes measured) values. Often, there are multiple approaches to their calculation. The errors in the calculation of these properties are statistical or stochastic. A third class of properties includes those that can only be estimated from other properties, using a regression analysis, neural network, or other linear or nonlinear function of variables. The errors in these properties can be complex and difficult to determine. For all these reasons, it is important to carefully consider the use of any given property for drug discovery purposes. Too often, properties are calculated simply because they are available, then used in a QSAR analysis, and possibly applied to future predictions—all without proper consideration of their precision, accuracy, and relevance to the chemical problem.

Given this caveat, it must be noted that there are a multitude of programs available for the calculation of properties of structures. Some programs compute only a single property, like Log*P*. Others calculate a series of values in a given genre of property, like molecular connectivity (111) or BCUT descriptors (112). Still others compute a vast range of properties that include topological, topographical, and physicochemical descriptors alike. It is beyond the scope of this chapter to detail all the programs and vendors that provide property calculation and estimation software. Many of the calculations are provided as part of molecular modeling and QSAR program systems. Some programs and vendors whose products are solely for property calculation are described below.

4.1 Topological Descriptors

Descriptors based on the 2D structure or simply on the connectivity matrix of a structure have long been used for chemical similarity and for property correlations. Because they often lack any relationship to mechanism, these descriptors are best used within a congeneric series or at least a set of similar structures. They may be empirically useful for cluster analysis and chemical library design, because they are effective at representing structure differences and similarities. A few programs and providers of topological descriptors include the following:

- Barnard Chemical Information—provides chemical Fingerprint Generation Pack—to compute fragment-based fingerprints for cluster and diversity analysis (113)
- DRAGON—implementation of about 1400 descriptors of Todeschini and Consonni (114) including constitutional, topological, autocorrelation, geometrical and functional