molecular graphs or structural formulas are "two-dimensional," these methods are referred to as 2D-QSAR. Most of the 2D-QSAR methods are based on graph theoretical indices, which have been extensively studied by Randíc (19) and Kier and Hall (20–22). They include, for example, molecular connectivity indices (19, 20), molecular shape indices (23, 24), topological (25) and electrotopological state indices (26–29), and atom-pair descriptors (30, 31). Sometimes, topological descriptors are also combined with physicochemical properties of molecules. Although these structural indices represent different aspects of molecular structures, and, what is important for QSAR, different structures provide numerically different values of indices, their physicochemical meaning is frequently unclear. The successful applications of topological indices combined with multiple linear regression (MLR) analysis have been summarized by Kier and Hall (20, 21, 28).

The third group of methods is based on descriptors derived from spatial (three-dimensional) representation of molecular structures. Correspondingly, these methods are referred to as three-dimensional or 3D-QSAR; they have become increasingly popular with the development of fast and accurate computational methods for generating 3D conformations and alignments of chemical structures. The early examples of 3D-QSAR include molecular shape analysis (MSA) (32), distance geometry (33, 34), and Voronoi techniques (35). The first method uses shape descriptors and multiple linear regression analysis, whereas the latter methods apply atomic refractivity as structural descriptors and the solution of mathematical inequalities to obtain the quantitative relationships. These two methods have been applied to the study of structure-activity relationships of many data sets by Hopfinger (e.g., Refs. 36, 37) and Crippen (e.g., Refs. 38, 39), respectively.

Perhaps the most popular example of 3D-QSAR is the comparative molecular field analysis (CoMFA), developed by Cramer et al. (40), which has elegantly combined the power of 3D molecular modeling and partial least-square (PLS) optimization technique (41, 42) and found wide applications in medicinal chemistry and toxicity analysis (see below). Most of 3D-QSAR methods require 3D alignment of all molecules according to a phannacophore model or based on ligand docking to a receptor-binding site. Descriptors in the case of CoMFA (40, 43) and CoMFA-like methods such as COMBINE (44), COMSiA (45), and QsiAR (46) represent electrostatic, steric, and hydrophobic field values (to name but a few examples) in the grid points surrounding molecules.

Finally, QSAR methods can also be classified by the type of the correlation methods used in model development. Linear methods include linear regression or MLR, PLS (41, 42, 47), or principal component regression (PCR), whereas nonlinear methods can be exemplified, for example, by k-Nearest Neighbors (kNN) (48, 49) and artificial neural networks (50) methods. An example of the linear methods is provided by the ADAPT system, which employs topological indices as well as other calculable structural parameters (e.g., steric and quantum mechanical parameters), and the MLR method for QSAR analysis. It has been extensively applied to QSAR/QSPR studies in analytical chemistry, toxicity analysis, and other biological activity prediction (51–54). Parameters derived from various experiments through chemometric methods have also been used in the study of peptide QSAR (55), where PLS analysis was employed. The latter technique has been used almost exclusively in 3D-QSAR, where the number of descriptors characterizing molecular fields may exceed the number of compounds by orders of magnitude.

There has been a great deal of interest, especially more recently, in the use of data mining methods to extract the information from large and/or chemically inhomogeneous data sets. Examples of these methods include pattern recognition (56, 57), automated structure evaluation (58, 59), neural network (60–62), and machine learning (63–65). Recent trends in QSAR studies also include developing optimal QSAR models through variable selection, that is, by selecting a subset of available descriptors in either MLR, PLS, or nonlinear classification or artificial neural networks (ANN) analysis as applied either in 2D- (66–72) or in 3D-QSAR (73). These methods employ either generalized simulated annealing