Two other steric parameters worth mentioning are molar refractivity (MR) and the Verloop parameter. *Molar refractivity*,[200] the molar volume corrected by the refractive index which represents the size and polarizability of a fragment or molecule, is defined by the Lorentz-Lorenz equation:

$$\mathrm{MR} = \frac{n^2 - 1}{n^2 + 2}\frac{\mathrm{MW}}{d} \tag{2.17}$$

where n is the index of refraction at the sodium D line, MW is the molecular weight, and d is the density of the compound. The greater the positive MR value of a substituent, the larger its steric or bulk effect. This parameter also measures the electronic effect and, therefore, may reflect dipole–dipole interactions at the receptor site.

The *Verloop steric parameters*[201] are used in a program called STERIMOL to calculate the steric substituent values from standard bond angles, van der Waals radii, bond lengths, and user-determined reasonable conformations. Five parameters are involved. One (L) is the length of the substituent along the axis of the bond between the substituent and the parent molecule. Four width parameters (B_1–B_4) are measured perpendicular to the bond axis. These five parameters describe the positions, relative to the point of attachment and the bond axis, of five planes that closely surround the group. In contrast to E_s values which, because of the reaction on which they are based, cannot be determined for many substituents, the Verloop parameters are available for any substituent.

G.3 Methods Used to Correlate Physicochemical Parameters with Biological Activity

Now that we can obtain numerous *physicochemical parameters* (also called *descriptors*) for any substituent, how do we use these parameters to gain information regarding what compound to synthesize next in an attempt to optimize the lead compound? First, several (usually, many) compounds related to the lead are synthesized, and the biological activities are determined in some screen. These data, then, can be manipulated by a number of QSAR methods. I present Hansch analysis first. If you are not interested in an overview of computational methods, you can skip sections 2.2.G.3 and 2.2.G.4, pp. 68–78.

a. Hansch Analysis: A Linear Multiple Regression Analysis

With the realization that there are (at least) two considerations for biological activity, namely, lipophilicity (required for the journey of the drug to the site of action) and electronic factors (required for drug interaction with the site of action), and that lipophilicity is a parabolic function, Hansch and Fujita[202] expanded Equation 2.8 to that shown in either Equation 2.18a or 2.18b, known as the *Hansch equation*,

$$\log 1/C = -k\pi^2 + k'\pi + \rho\sigma + k'' \tag{2.18a}$$

$$\log 1/C = -k(\log P)^2 + k'(\log P) + \rho\sigma + k'' \tag{2.18b}$$

where C is the molar concentration (or dose) that elicits a standard biological response (e.g., ED_{50}, the dose required for 50% of the maximal effect; IC_{50}, the concentration that gives 50% inhibition of an enzyme or antagonism of a receptor; LD_{50}, the lethal dose for 50% of the animal population); k, k', ρ, and k'' are the regression coefficients derived from statistical curve fitting; and π and σ are the lipophilicity and electronic substituent constants, respectively. The reciprocal of the concentration ($1/C$) reflects the fact that greater potency is associated with a

lower dose, and the negative sign for the π^2 [or $(\log P)^2$] term reflects the expectation of an optimum lipophilicity, i.e., the π_0 or $\log P_0$.

Because of the importance of steric effects and other shape factors of molecules for receptor interactions, an E_s term and a variety of other shape, size, or topography terms (S) have been added to the Hansch equation:

$$\log 1/C = -a\pi^2 + b\pi + \rho\sigma + cE_s + dS + e \tag{2.19}$$

The way these parameters are used is by the application of the method of linear multiple regression analysis.[203] The best least-squares fit of the dependent variable (the biological activity) to a linear combination of the independent variables (the descriptors) is determined. Hansch analysis, also called the *extrathermodynamic method*, then, is a linear free-energy approach to drug design in congeneric series in which equations are set up involving different combinations of the physicochemical parameters; the statistical methodology allows the best equation to be selected and the statistical significance of the correlation to be assessed. Once this equation has been established, it can be used to predict the activities of untested compounds. Problems associated with the use of multiple regression analysis in QSAR studies have been discussed by Deardon.[204]

Several assumptions must be made when the extrathermodynamic method is utilized: Conformational changes in receptors can be ignored, metabolism does not interfere, linear free-energy terms relevant to receptor affinity are additive, the potency–lipophilicity relationship is parabolic or linear, and correlation implies a causal relationship. According to Martin[205] and Tute,[206] there is a balance of assets and liabilities to the extrathermodynamic method. The strengths are several-fold: (1) The use of descriptors (π, σ, E_s, MR, and so forth) permits data collected from simple organic chemical model systems to be utilized for the prediction of biological activity in complex systems, (2) the predictions are quantitative with statistical confidence limits, (3) the method is easy to use and is inexpensive, and (4) conclusions that are reached may have application beyond the substituents included in the particular analysis.

The weaknesses of this method are that (1) parameter values must be available for the substituents in the data set; (2) a large number of compounds must be included in the analysis to have confidence in the derived equations; (3) expertise in statistics and computer use is essential; (4) small molecule interactions are imperfect models for biological systems; (5) in contrast to chemical reactions in which you know the atoms that interact with the reagent, steric effects in biological systems may not be relevant, since it is often not certain which atoms in the drug interact with the receptor; (6) organic reactions used to determine the descriptors usually are studied under acidic or basic conditions when all analogs are fully protonated or deprotonated, but in biological systems, the drug may be partially protonated; (7) because QSAR is empirical, it is a retrospective technique that depends on the pharmacological activity of compounds belonging to the same structural type, and, therefore, new types of active compounds are not discovered (i.e., it is a lead optimization technique, not a lead discovery approach); and (8) like other empirical relationships, extrapolations frequently lead to false predictions.

Despite the weaknesses of this approach, it is used, and several successes in drug design attributable to Hansch analysis have been reported.[207] As pointed out in Chapter 3 (Section 3.2.E.2, p. 143), however, caution should be used when applying QSAR methods to racemic mixtures if only one enantiomer is active. Other important statistical approaches are mentioned briefly.

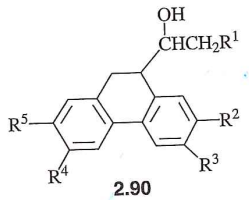
b. Free and Wilson or *de novo* Method

Not long after Hansch proposed the extrathermodynamic approach, Free and Wilson[208] reported a general mathematical method for assessing the occurrence of additive substituent effects and for quantitatively estimating their magnitude. It is a method for the optimization of substituents within a given molecular framework that is based on the (tenuous) assumption that the introduction of a particular substituent at any one position in a molecule always changes the relative potency by the same amount, regardless of what other substituents are present in the molecule. A series of linear equations of the form shown here is constructed:

$$\mathrm{BA} = \sum a_i \mathrm{X}_i + \mu \tag{2.20}$$

where BA is the magnitude of the biological activity, X_i is the ith substituent with a value of 1 if present and 0 if not, a_i is the contribution of the ith substituent to the BA, and μ is the overall average activity of the parent skeleton. These linear equations are solved by the method of least squares for the a_i and μ. All activity contributions at each position of substitution must sum to zero. The pros and cons of the Free-Wilson method have been discussed.[209] Fujita and Ban[210] suggested two modifications of the Free-Wilson approach on the assumption that the effect on the activity of a certain substituent at a certain position in a compound is constant and additive. First, that the biological activity should be expressed as log A/A_0, where A and A_0 represent the magnitude of the activity of the substituted and unsubstituted compounds, respectively, and that a_i is the log activity contribution of the ith substituent relative to H. This allows the derived substituent constants to be compared directly with other free-energy-related parameters that are additive. Second, that μ become analogous to the theoretically predicted (calculated) activity of the parent compound of the series. Both of these modifications have been widely accepted.

As an example of the Free-Wilson approach, consider the hypothetical compound **2.90**.[211] If in one pair of analogs for which R^1, R^2, R^3, and R^4 are constant and R^5 is Cl or CH_3, the methyl compound is one-tenth as potent as the chloro analog, then the Free-Wilson method assumes that every R^5 methyl analog (where R^1–R^4 are varied) will be one-tenth as potent as the corresponding R^5 chloro analog. A requirement for this approach, then, is a series of compounds that have changes at more than one position. In addition, each type of substituent must occur more than once at each position in which it is found. The outcome is a table of the contribution to potency of each substituent at each position. If the free-energy relationships of the extrathermodynamic method are linear or position specific, then Free-Wilson calculations will be successful.



2.90

The interaction model[212] is a mathematical model similar to that of the Free-Wilson additive model with an additional term ($e_X e_Y$) that is to account for possible interactions between substituents X and Y.

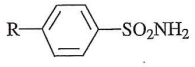
c. Enhancement Factor

One of the earliest QSAR observations resulted from a retrospective analysis of a large number of synthetic corticosteroids.[213] Examination of the biological properties of steroids prepared by the introduction of halogen, hydroxyl, alkyl, or double bond modifications revealed that each substituent affects the activity of the molecule in a quantitative sense and almost independently of other groups. The effect (whether positive or negative) of each substituent was assigned a numerical value termed the *enhancement factor*. Multiplication of the enhancement factor for each substituent by the biological activity of the unsubstituted compound gave the potency of the modified steroid.

d. Manual Stepwise Methods: Topliss Operational Schemes and Others

Because of the lack of easy access to computers by chemists in the early 1970s, Topliss[214] developed a nonmathematical, nonstatistical, and noncomputerized (hence, manual) guide to the use of the Hansch principles. This method is most useful when the synthesis of large numbers of compounds is difficult and when biological testing of compounds is readily available. It is an approach for the efficient optimization of the potency of a lead compound with the minimization of the number of compounds needed to be synthesized. The only prerequisite for the technique is that the lead compound must contain an unfused benzene ring. However, according to literature surveys at the time that this method was published, 40% of all reported compounds[215] contained an unfused benzene ring and 50% of drug-oriented patents[216] were concerned with substituted benzenes. This approach relies heavily on π and σ values and to a much lesser degree E_s values. The methodology will be outlined here; a more detailed discussion can be found in the Topliss papers.

Consider that your lead compound is benzenesulfonamide (**2.91**, R = H) and its potency has been measured in whatever screen is being used. Because many systems are $+\pi$ dependent, that is, the potency increases with increasing π values, then a good choice for your first analog would be one with a substituent having a $+\pi$ value. Because $\pi_{4\text{-Cl}} = 0.71$ and $\sigma_{4\text{-Cl}} = 0.23$ (remember, $\pi_H = \sigma_H = 0$), the 4-chloro analog (**2.91**, R = Cl) should be synthesized and tested. There are three possible outcomes of this effort, namely, the 4-chloro analog is more potent (M), equipotent (E), or less potent (L) than the parent compound. If it is more potent, then it can be attributed to a $+\pi$ effect, a $+\sigma$ effect, or to both. To determine which is important, one term could be held more or less constant and the other varied. For example, the 4-phenylthio analog ($\pi_{4\text{-PhS}} = 2.32$, $\sigma_{4\text{-PhS}} = 0.18$) would be a good test of the importance of lipophilicity, and the 4-trifluoromethyl analog ($\pi_{4\text{-CF}_3} = 0.88$, $\sigma_{4\text{-CF}_3} = 0.54$) would test the importance of electron withdrawal. If the 4-phenylthio analog is more potent than the 4-chloro analog, further increases in lipophilicity would be desirable. At this point a potency tree, termed a *Topliss decision tree*, could be constructed (Figure 2.16), and additional analogs could be made.



2.91

What if the 4-chloro analog was equipotent with the parent compound? This could result from a favorable $+\pi$ effect counterbalanced by an unfavorable $+\sigma$ effect or vice versa. If this is the case, then the 4-methyl analog ($\pi_{4\text{-Me}} = 0.56$, $\sigma_{4\text{-Me}} = -0.17$) should show enhanced potency. Enhancement of potency by the 4-methyl analog would suggest that the synthesis

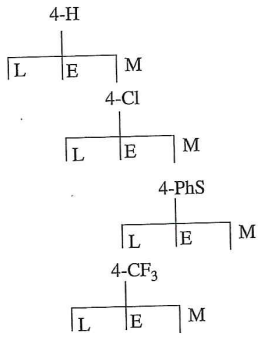


Figure 2.16 ▶ Topliss decision tree (M, more potent; E, equipotent; L, less potent)

of analogs with increasing π values and decreasing σ values would be propitious. If the 4-methyl analog is worse than the 4-chloro analog, perhaps the equipotency of the 4-chloro compound was the result of a favorable σ effect and an unfavorable π effect. The 4-nitro analog ($\pi_{4\text{-}NO_2} = -0.28$, $\sigma_{4\text{-}NO_2} = 0.78$) would, then, be a wise next choice.

If the 4-chloro analog was less potent than the lead, then there may be a steric problem at the 4 position or increased potency depends on $-\pi$ and $-\sigma$ values. The 3-chloro analog ($\pi_{3-Cl} = 0.71$, $\sigma_{3-Cl} = 0.37$) could be synthesized to determine if a steric effect is the problem. Note that the σ constant for the 3-Cl substituent is different from that for the 4-Cl one because these descriptors are constitutive. If there is no steric effect, then the 4-methoxy compound ($\pi_{4\text{-}OMe} = -0.04$, $\sigma_{4\text{-}OMe} = -0.27$) could be prepared to investigate the effect of adding a $-\pi$ and $-\sigma$ substituent. Increased potency of the 4-OMe substituent would suggest that other substituents with more negative π and/or σ constants be tried.

This analysis was based almost exclusively on π and σ values, and other factors such as steric effects have been neglected. Another way to increase both π and σ values would be by synthesizing the 3,4-dichloro analog ($\pi_{3,4\text{-}Cl_2} = 1.25$, $\sigma_{3,4\text{-}Cl_2} = 0.52$). Again, the 3,4-dichloro analog could be more potent, equipotent, or less potent than the 4-chloro compound. If it is more potent, then determination of whether $+\pi$ or $+\sigma$ is more important could be made by selection of appropriate substituents with higher π and/or σ values. If the 3,4-dichloro compound was less potent than the 4-chloro analog, it could be that the optimum values of π and σ were exceeded or that the 3-chloro group has an unfavorable steric effect. The latter hypothesis could be tested by the synthesis of the 4-trifluoromethyl analog ($\pi_{4\text{-}CF_3} = 0.88$, $\sigma_{4\text{-}CF_3} = 0.54$) which has no 3-substituent, but has a high σ and intermediate π value.

Topliss extended the operational scheme for side-chain problems when the group is adjacent to a carbonyl, amino, or amide functionality, i.e., –COR, –NHR, –CONHR, and –NHCOR, where R is the variable substituent. This approach is applicable to a variety of situations other than direct substitution on the aromatic nucleus. In this case, the parent molecule is the one where R = CH_3, and π, σ, and E_s parameters are used. Note that in the Topliss operational scheme, as in the other methods in this section, the procedure is stepwise; that is, the next compound is determined on the basis of the results obtained with the previous one.

Three other manual, stepwise methods are mentioned briefly: Craig plots,[217] the Fibonacci search method,[218] and sequential simplex strategy.[219] The Topliss decision tree approach evolved from the work of Craig, who pointed out the utility of a simple graphical plot of π versus σ (or any two parameters) to guide the choice of a substituent (Figure 2.17). Once the Hansch equation has been expressed for an initial set of compounds, the sign and

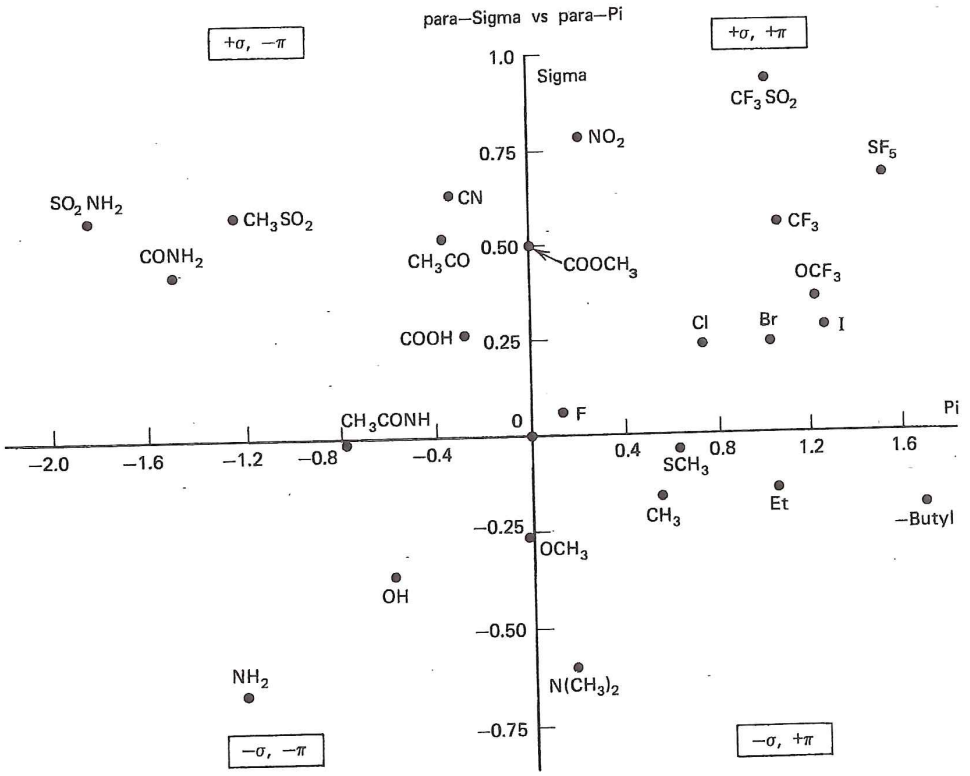


Figure 2.17 ▶ Craig plot of σ constants versus π values for aromatic substituents. [From Craig, P. N. (1980). In *Burger's Medicinal Chemistry*, (M. E. Wolff, ed.), 4th ed., Part I, p. 343. Wiley, New York. Copyright ©1980 John Wiley & Sons, Inc. This material is used by permission of John Wiley & Sons, Inc.]

magnitude of the π and σ regression coefficients determine the particular quadrant of the Craig plot that is to be used to direct further synthesis. Thus, if both the π and σ terms have positive coefficients, then substituents in the upper right-hand quadrant of the plot (Figure 2.17) should be selected for future analogs.

The Fibonacci search technique is a manual method to discover the optimum of some parabolic function, such as potency versus log P, in a minimum number of steps. Sequential simplex strategy is another stepwise technique suggested when potency depends on two physicochemical parameters such as π and σ.

e. Batch Selection Methods: Batchwise Topliss Operational Scheme, Cluster Analysis, and Others

The inherent problem with the Topliss operational scheme described above is its stepwise nature. Provided that pharmacological results can be obtained quickly, this is probably not much of a problem; however, sometimes biological evaluation is slow. Topliss[220] proposed an alternative scheme that uses batchwise analysis of small groups of compounds. Substituents were grouped by Topliss according to π, σ, π^2, and a variety of $x\pi$- and $y\sigma$-weighted combinations. The approach starts with the synthesis of five derivatives, the unsubstituted (4-H);

4-chloro; 3,4-dichloro; 4-methyl; and 4-methoxy compounds. After these five analogs have been screened, they are ranked in order of decreasing potency. The potency order determined for these analogs is, then, compared with the rankings in Table 2.9 to determine which parameter or combination of parameters is most dominant. If, for example, the potency order is 4-OCH_3 > 4-CH_3 > H > 4-Cl > 3,4-Cl_2, then $-\sigma$ is the dominant parameter. Once the parameter dependency is determined, Table 2.10 is consulted to discover what substituents should be investigated next. In the above example, 4-$N(C_2H_5)_2$, 4-$N(CH_3)_2$, 4-NH_2, 4-NHC_4H_9, 4-OH, 4-$OCH(CH_3)_2$, 3-CH_3, and 4-OCH_3 would be suitable choices. The major weakness of this approach is that it is difficult to extend the method to additional parameters unless computers are used.

A computer-based batch selection method, known as *cluster analysis*, was introduced by Hansch *et al.*[221] Substituents were grouped into clusters with similar properties according

TABLE 2.9 ▶ Potency Order for Various Parameter Dependencies [With permission from Topliss, J. G. (1977). Reprinted with permission from *J. Med. Chem.* 20, 463. Copyright © 1977 American Chemical Society.]

	Parameters									
Substituent	π	$2\pi - \pi^2$	σ	$-\sigma$	$\pi + \sigma$	$2\pi - \sigma$	$\pi - \sigma$	$\pi - 2\sigma$	$\pi - 3\sigma$	E_4^a
3,4-Cl_2	1	1-2	1	5	1	1	1-2	3-4	5	2-5
4-Cl	2	1-2	2	4	2	2-3	3	3-4	3-4	2-5
4-CH_3	3	3	4	2	3	2-3	1-2	1	1	2-5
4-OCH_3	4-5	4-5	5	1	5	4	4	2	2	2-5
H	4-5	4-5	3	3	4	5	5	5	3-4	1

[a] Unfavorable steric effect from 4-substitution.

TABLE 2.10 ▶ New Substituent Selections [With permission from Topliss, J. G. (1977). Reprinted with permission from *J. Med. Chem.* 20, 463. Copyright 1977 American Chemical Society.]

Probable operative parameters	New substituent selection
$\pi, \pi + \sigma, \sigma$	3-CF_3, 4-Cl; 3-CF_3, 4-NO_2; 4-CF_3, 2,4-Cl_2; 4-*c*-C_5H_9; 4-*c*-C_6H_{11}
$\pi, 2\pi - \sigma, \pi - \sigma$	4-$CH(CH_3)_2$; 4-$C(CH_3)_3$; 3,4-$(CH_3)_2$; 4-$O(CH_2)_3CH_3$; 4-OCH_2Ph; 4-$N(C_2H_5)_2$
$\pi - 2\sigma, \pi - 3\sigma, -\sigma$	4-$N(C_2H_5)_2$; 4-$N(CH_3)_2$; 4-NH_2; 4-NHC_4H_9; 4-OH; 4-$OCH(CH_3)_2$; 3-CH_3, 4-OCH_3
$2\pi - \pi^2$	4-Br; 3-CF_3; 3,4-$(CH_3)_2$; 4-C_2H_5; 4-$O(CH_2)_2CH_3$; 3-CH_3, 4-Cl; 3-Cl; 3-CH_3; 3-OCH_3; 3-$N(CH_3)_2$; 3-CF_3; 3,5-Cl_2
Ortho effect	2-Cl; 2-CH_3; 2-OCH_3; 2-F
Other	4-F; 4-$NHCOCH_3$; 4-$NHSO_2CH_3$; 4-NO_2; 4-$COCH_3$; 4-SO_2CH_3; 4-$CONH_2$; 4-SO_2NH_2

TABLE 2.11 ▶ Typical Members of Clusters Based on σ, π, F, R, MR, and MW [With permission from Martin, Y. C. (1979). Reprinted from *Drug Design*, Vol. VIII, E. J. Ariëns, ed., "Advances in the Methodology of Quantitative Drug Design", pp 2–72, Copyright ©1979, with permission from Elsevier.]

Cluster number[a]	Typical members
1	Me, H, 3,4-(OCH_2O), CH_2CH_2COOH, $CH{=}CH_2$, Et, CH_2OH
2	CH=CHCOOH
3a	CN, NO_2, CHO, COOH, COMe
3b	C + CH, CH_2Cl, Cl, NNN, SH, Sme, CH=NOH, CH_2CN, OCOMe, SCOMe, COOMe, SCN
4a	$CONH_2$, CONHMe, SO_2NH_2, SO_2Me, SOMe
4b	NHCHO, NHCOMe, $NHCONH_2$, $NHCSNH_2$, $NHSO_2Me$
5	F, OMe, NH_2, $NHNH_2$, OH, NHMe, NHEt, NMe_2
6	Br, OCF_3, CF_3, NCS, I, SF_5, SO_2F
7	CH_2Br, SeMe, $NHCO_2Et$, SO_2Ph, OSO_2Me
8	NHCOPh, $NHSO_2Ph$, OSO_2Ph, COPh, N=NPh, OCOPh, PO_2Ph
9	3,4-$(CH_2)_3$, 3,4-$(CH_2)_4$, Pr, *i*-Pr, 3,4-$(CH)_4$, NHBu, Ph, CH_2Ph, *t*-Bu, OPh
10	Ferrocenyl, adamantyl

[a] Clusters 3 and 4 contain many of the common substituents used in medicinal chemistry; hence, these clusters are further subdivided according to their cluster membership when 20 clusters have been made.

to their σ, π, π^2, E_s, F (field constant), R (resonance constant), MR (molar refractivity), and MW (molecular weight) values. Some of the clusters are shown in Table 2.11.[222] One member of each cluster would be selected for substitution into the lead compound, and the compounds would be synthesized and tested. If a substituent showed dominant potency, then other substituents from that cluster would be selected for further investigation. The important advantage of the batch selection methods is that the initial batch of analogs prepared is derived from the widest range of parameters possible so that the dominant physicochemical property can be revealed early in the lead modification process.

The initial promise of these computational methods has yet to be realized. They seem to be just additional examples of potentially exciting new approaches for which little success has been forthcoming. These early computational methods have largely been supplanted by what is known as 3D-QSAR and molecular modeling approaches.

G.4 Computer-Based Methods of QSAR Related to Receptor Binding: 3D-QSAR

Three-dimensional quantitative structure–activity relationships (3D-QSAR) permit correlations between a series of diverse molecular structures and their biological functions at a particular target. The general approach of 3D-QSAR is to select a group of molecules, each

of which has been assayed for a particular activity; align the molecules according to some predetermined orientation rules; calculate a set of spatially dependent parameters for each molecule determined in the receptor space surrounding the aligned series; derive a function that relates each molecule's spatial parameters to their respective biological property; and establish self-consistency and predictability of the derived function. A variety of computer-based methods have been used to correlate molecular structure with receptor binding, and, therefore, activity. Some are mentioned here; many more are listed in the General References at the end of the chapter.

Crippen and coworkers[223,224] devised a linear free-energy model, termed the *distance geometry* approach, for calculating QSAR from receptor binding data. The distances between various atoms in the molecule, compiled into a table called the *distance matrix*, define the conformation of the molecule. Rotations about single bonds change the molecular conformation and, therefore, these distances; consequently, an upper and lower distance limit is set on each distance. Experimentally determined free energies of binding of a series of compounds to the receptor are used with the distance matrix of each molecule in a computerized method to deduce possible binding sites in terms of geometry and chemical character of the site, thereby defining a three-dimensional pharmacophore. Although this approach requires more computational effort and adjustable parameters than Hansch analysis, it is thought to give good results on more difficult data sets.

The distance geometry approach was extended by Sheridan *et al.*[225] to treat two or more molecules as a single ensemble. The ensemble approach to distance geometry can be used to find a common pharmacophore for a receptor with unknown structure from a small set of biologically active molecules. Once the pharmacophore has been, at least, partially identified, new molecular scaffolds can be revealed which contain that pharmacophore embedded in their structure by *three-dimensional database* (or *similarity*) *searching*.[226] In this method you start from a receptor ligand, enzyme substrate, or other molecule whose pharmacophoric groups are known for a particular target. Then a database of compounds (e.g., the company's library of compounds, the CMC database, or any database of compounds) is searched to determine which ones have a similar three-dimensional structure as the pharmacophore. The top virtual "hits" are visually inspected to determine which ones might be the best candidates, and then they are tested. This was the approach taken to identify inhibitors of human immunodeficiency virus type 1 integrase (HIV-1 IN) as potential anti-AIDS drugs.[227] HIV-1 IN mediates the integration of HIV-1 DNA into host chromosomal targets and is essential for effective viral replication. From a known inhibitor of HIV-1 IN, a pharmacophore hypothesis was proposed. Based on this hypothesis, a three-dimensional search of the National Cancer Institute (NCI) database of compounds was performed, which produced 267 structures that matched the pharmacophore; 60 of those were tested against HIV-1 IN, and 19 were found to be active. The relevance of the proposed pharmacophore was tested using a small three-dimensional validation database of known HIV-1 IN inhibitors, which had no overlap with the group of compounds found in the initial search. This new three-dimensional search supported the existence of the postulated pharmacophore and also suggested a possible second pharmacophore. Using the second pharmacophore in another three-dimensional search of the NCI database, 10 novel, structurally diverse HIV-1 IN inhibitors were found.

Hopfinger[228] has developed a set of computational procedures termed *molecular shape analysis* for the determination of the active conformations and, thereby, molecular shapes during receptor binding. Common pairwise overlap steric volumes calculated from low-energy

conformations of molecules are used to obtain three-dimensional molecular shape descriptors that can be treated quantitatively and used with other physicochemical parameter descriptors.

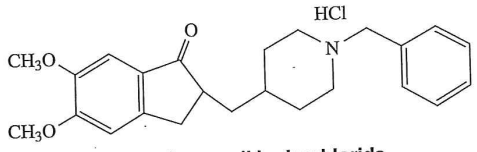
Two other descriptors for substructure representation, the atom pair[229] and the topological torsion,[230] have been described by Venkataraghavan and coworkers. These descriptors characterize molecules in fundamental ways that are useful for the selection of potentially active compounds from hundreds of thousands of structures in a database. The atom pair method can select compounds from diverse structural classes that have atoms within the entire molecule similar to those of a particular active structure. The topological torsion descriptor is complementary to the atom pair descriptor, and focuses on a local environment of a molecule for comparison with active structures.

One of the most widely used computer-based 3D-QSAR methodologies, developed by Cramer and coworkers,[231] is termed *Comparative Molecular Field Analysis* (CoMFA).[232] In this method the molecule–receptor interaction is represented by the steric and electrostatic fields exerted by each molecule. A series of active compounds is identified, and three-dimensional structural models are constructed. These structures are superimposed on one another and placed within a regular three-dimensional grid. A probe atom, with its own energetic values, is placed at lattice points on the grid, where it is used to calculate the steric and electrostatic potentials between itself and each of the superimposed structures. At each lattice point one steric value, one electrostatic value, and one inhibition value are saved for each inhibitor in the series. The results are represented as a three-dimensional contour map in which contours of various colors represent locations on the structure where lower or higher steric or electrostatic interactions would increase binding. However, because simple steric and electrostatic fields are unlikely to represent a complete description of a drug–receptor interaction, alternative and modified forms have been proposed.[233] Because it is assumed that the molecules bind with similar orientations in the receptor, which may not necessarily be the case, correct alignments are almost impossible, particularly for compounds with a large number of rotatable bonds, which limits the applicability of CoMFA. Other approaches have been developed that do not depend on a common alignment of the molecules, such as *Comparative Molecular Moment Analysis* (CoMMA),[234] EVA,[235] and WHIM;[236] these approaches provide 3D descriptors that are independent of the orientation of the molecules in space, so they do not have to be aligned. However, it is not possible to give a 3D display of the resulting model. Goodford's program called *GRID* uses a grid force field that includes a very good description of hydrogen bonding.[237] Because the energetics, as well as the shape complementarity, of a drug–receptor complex are vital to its stability, this method simultaneously displays the energy contour surfaces and the macromolecular structure on the computer graphics system. This allows both the energy and shape to be considered together when considering the design of molecules that have an optimal fit to the receptor, and it determines probable interaction sites between various functional groups on the ligand and the enzyme surface. The program *HINT* (Hydrophobic INTeractions) maps potential hydrophobic and polar interactions between a molecule and a receptor.[238]

Another useful methodology is the *hypothetical active site lattice (HASL) technique*, which creates a QSAR model from a composite lattice generated from a series of regular orthogonal 3D grids established for each molecule.[239] These points are restricted to locations embedded in the van der Waals radii of a molecule and are kept in the analysis dependent on some feature of a proximal atom (for example, hydrophobicity). Each molecule's biological activity value is then averaged over all points on its respective lattice. A composite lattice is constructed from these partial activity values by averaging over all molecules that share common points.

Using an iterative optimization scheme, the partial activity values are gradually adjusted to yield a composite lattice best fitting the molecular series. Other pharmacophore-based design algorithms include Caveat,[240] Aladdin,[241] and Spacer Skeletons.[242]

The anti-Alzheimer's drug donepezil hydrochloride (**2.92**, Aricept) was discovered using a variety of 3D-QSAR methods.[243]



donepezil hydrochloride
2.92

2.2.H Molecular Graphics-Based Drug Design

QSAR studies have relied heavily on the use of computers from the beginning for statistical calculations involving multiparameter equations. Researchers soon realized that drug design could be aided significantly if structures of receptors and drugs could be displayed on a computer terminal, and molecular processes could be observed. *Molecular graphics* is the visualization and manipulation of 3D representations of molecules on a graphics display device. The origins of molecular graphics have been traced by Hassall[244] to the project MAC (Multiple Access Computer),[245] which produced molecular graphics models of macromolecules for the first time. The potential to apply this technology to protein crystallography was quickly realized, and by the early 1970s electron density data from X-ray diffraction studies could be presented and manipulated in stick or space-filling multicolored representations on a computer terminal.[246] The number of X-ray crystal structures available in the protein data bank (PDB)[247] went from about 200 in 1990 to more than 20,000 by 2003.

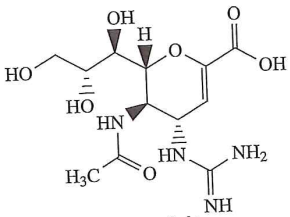
Medicinal chemists saw the potential of this approach in drug design as well. These approaches are known as *structure-based drug design* (SBDD), *computer-assisted drug design* (CADD), or *computer-assisted molecular design* (CAMD). A variety of commercial software packages are available for structure-based drug design, for example, Sybyl (Tripos), Insight II (Molecular Simulations Inc.), and Gold (Cambridge Crystallographic Data Centre). It is now possible for a synthetic chemist to carry out his or her own molecular modeling without having to become a computer scientist.

Stick (Dreiding) and space-filling (CPK) molecular models have been used extensively by organic chemists for years for small molecules, but these handheld models have major disadvantages.[248] Space-filling models often obscure the structure of the molecule, and wire or plastic models can give false impressions of molecular flexibility and tend to change into unfavorable conformations at inopportune moments. Plastic models of proteins are much too cumbersome to work with. A three-dimensional computer graphics representation of a protein that can be manipulated in three dimensions allows the operator to visualize the interactions of small molecules with biologically important macromolecules. Superimposition of structures, which is cumbersome at best with manual models, can be performed easily by molecular graphics. Also, some systems have the capability to synthesize graphically new structures by the assemblage of appropriate molecular fragments from a fragment file.

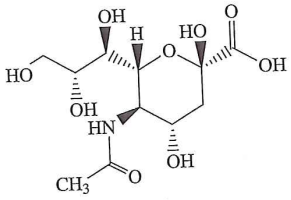
Numerous molecular graphics systems are available,[249] but the typical system, which has not changed much over the years, utilized by every major pharmaceutical company in the United States, Western Europe, and Japan, consists of a mainframe or supermini computer linked to a high-resolution graphics terminal with local intelligence. The graphics terminal may be equipped with a variety of peripheral devices such as graphic tablets, light pens, function keys, and dials to effect the molecular display and three-dimensional manipulations. The mainframe or minicomputer executes all of the molecular calculations, such as calculations of bond lengths, bond angles, and quantum chemical or force field calculations.

A variety of approaches can be taken to utilize molecular modeling for drug design; *direct design* approaches are used when the structure of the target receptor is known, and *indirect design* approaches are used when the receptor structure is not known. The basic premise in the utilization of molecular graphics is that the better the complementary fit of the drug to the receptor, the more potent the drug will be. This is the lock-and-key hypothesis of Fischer[250] in which the receptor is the lock into which the key (i.e., the drug) fits. To apply this concept most effectively, the structure of the receptor (either X-ray crystal structure or NMR solution structure) should be known; then, different drug analogs can be docked into the receptor. *Docking* is a molecular graphics term for the computer-assisted movement of a terminal-displayed molecule into its receptor. It cannot be assumed that the lowest energy structure of the molecule binds to the receptor; the bioactive conformation can be a higher energy conformation of the molecule.[251]

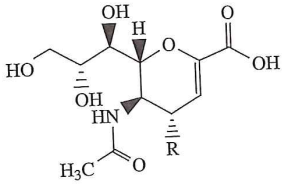
The most effective use of molecular modeling is when a high-resolution crystal structure (or NMR solution structure) of a receptor with a ligand bound is available. Molecular graphics visualization of the electron density map of this complex may reveal empty pockets in the complex that could be filled by appropriate modification of a lead compound. An important example of structure-based drug design is the discovery of zanamivir (**2.93**, Relenza), an antiviral agent used against influenza A and B infections.[252] The hemagglutinin at the surface of the virus binds to sialic acid (**2.94**) residues on receptors at the host cell surface. The virus enters the cell and replicates in the nucleus. The progeny virus particles escape the cell and stick to the sialic acid residues on the cell surface as well as to each other. *Neuraminidase* (also known as *sialidase*) is a key viral surface enzyme that catalyzes the cleavage of terminal sialic acid residues from the cell surface, which releases the virus particles to spread into the respiratory tract and infect new cells. The important feature of this enzyme that made it an attractive target for drug design is that its active site is lined with amino acids that are invariant in neuraminidases of all known strains of influenza A and B. Therefore, inhibition of this enzyme should be effective against all strains of influenza A and B. Random screening did not produce any potent inhibitors of the enzyme, although a nonselective neuraminidase inhibitor (**2.95**, R = OH) was identified. The breakthrough came when the crystal structures of the influenza A neuraminidase[253] with inhibitors bound[254] were obtained. The active site of the enzyme with **2.95** (R = OH) bound was probed computationally using Goodford's GRID program (see Section 2.2.G.4). Predictions by GRID of energetically favorable substitutions suggested replacement of the 4-hydroxyl group of **2.95** (R = OH) by an amino group (**2.95**, R = NH_2), which when protonated would form a favorable electrostatic interaction with Glu-119 (Figure 2.18a). It was apparent from the crystal structure that extension of the 4-ammonium group with a 4-guanidinium group (**2.93**) would produce an even tighter affinity because of the increase in basicity of the guanidinium group and also because it could interact with both Glu-119 and Glu-227 (Figure 2.18b).



zanamivir
2.93



2.94



2.95

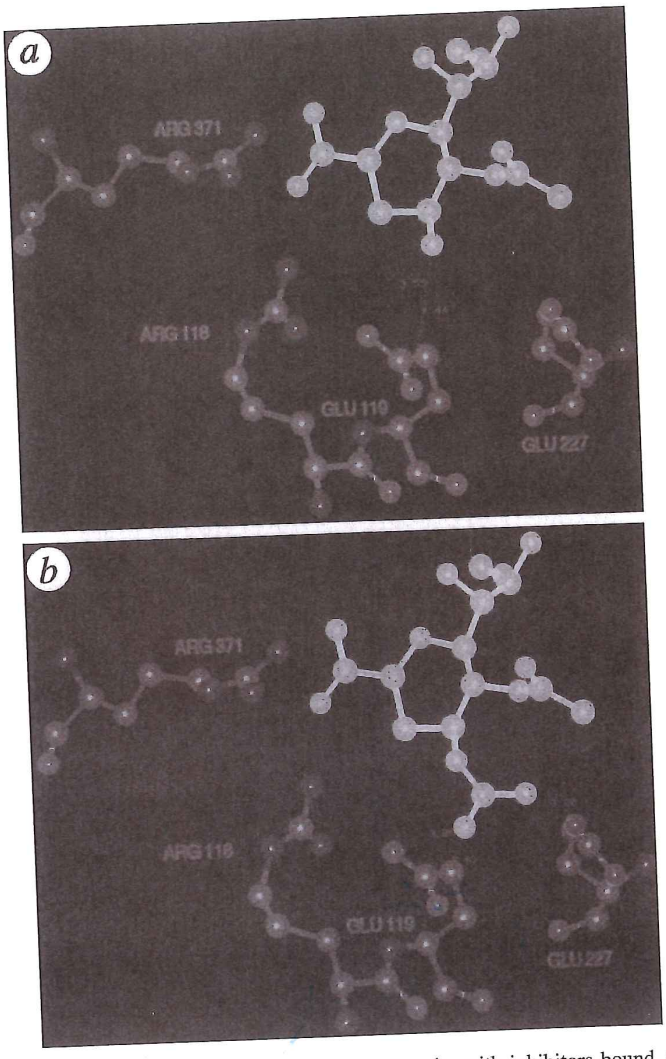


Figure 2.18 ▶ Crystal structure of neuraminidase active site with inhibitors bound. (a) Interaction of the protonated amino group of 2.95 ($R = NH_3^+$) with Glu-119. (b) Interaction of the protonated guanidinium group of **2.93** with Glu-119 and Glu-227 [Reprinted with permission from *Nature* 363, 418. Copyright ©1993 MacMillan Magazines Ltd.]. Reproduced in color between pages 172 and 173

Typically, the ideal compound is not realized so quickly. Rather, an idea for lead modification manifests from the crystal structure of the lead bound to the receptor. This modified lead is synthesized and tested; maybe only a minor improvement in potency is produced (or maybe lower potency). If there is some improvement, a new crystal structure is obtained, additional molecular modeling is carried out for further refinement of ideas, and new compounds are synthesized and tested. This process is reiterated with further rounds of design, synthesis, testing, and crystal structure until higher potency analogs are obtained. A beautiful example of how the iterative combination of molecular modeling, crystallography, and combinatorial and traditional medicinal chemistry synthesis was used to modify a lead neuraminidase inhibitor and enhance its potency 72,500-fold was described by the group at Abbott Laboratories.[255]

Sometimes even a crystal structure with the ligand bound is not sufficient. A high-resolution crystal structure of thymidylate synthase with a ligand bound did not properly account for a ligand-induced enzyme conformational change during structure-based drug design.[256] As a result the structure imparted an improper bias into the design of novel ligands.

Earlier in the chapter, the SAR of paclitaxel was described (**2.36**, Section 2.2.C, p. 23). By overlaying the molecular graphics depiction of the crystal structure of paclitaxel with those of four other natural products also found to promote stabilization of microtubules in competition with paclitaxel (Figure 2.19, Taxol), a common pharmacophore was proposed (Figure 2.20).[257] This gives a new perspective to lead modification, and permits the construction of new synthetic analogs having hybrid structures of each of the four unrelated scaffolds. Based on this pharmacophore model, **2.96** was synthesized and was shown to stabilize microtubules as well. Other 3D computer models of paclitaxel binding to microtubules have been promoted as well.[258] Without the molecular graphics capabilities, it would be very difficult to make this sort of comparison and design a new hybrid scaffold.

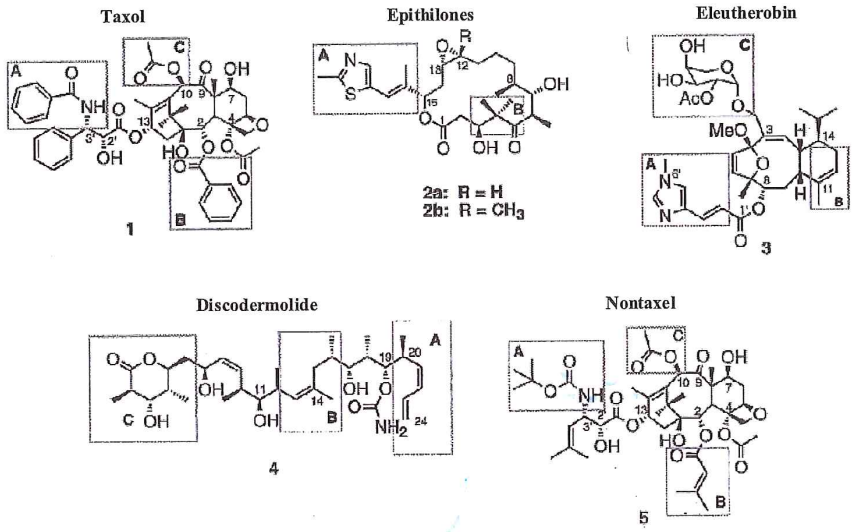


Figure 2.19 ▶ Five natural products found to promote stabilization of microtubules. The boxed sections were used to identify a common pharmacophore. [With permission from I. Ojima (1999). Reprinted with permission from *PNAS* 96, 4256. Copyright ©1999 National Academy of Science, U.S.A.]

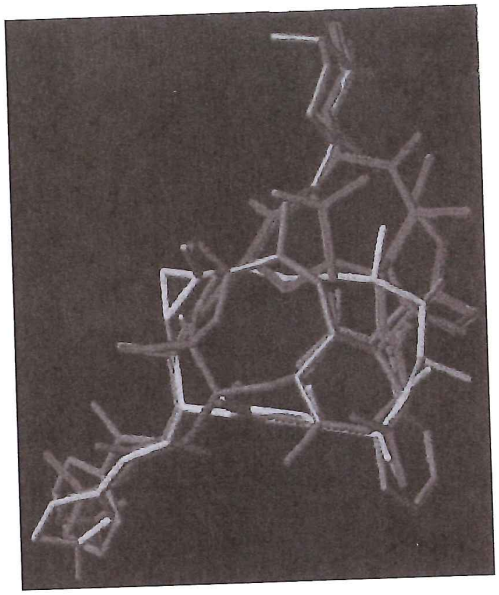
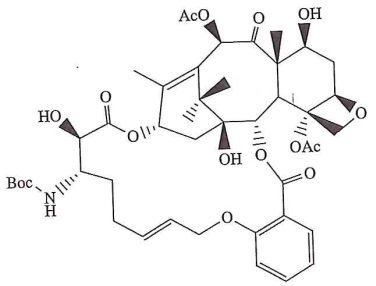


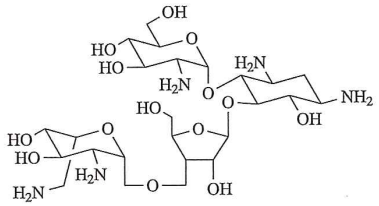
Figure 2.20 ► Common pharmacophore based on the composite of boxed sections in Figure 2.19. [With permission from I. Ojima (1999). Reprinted with permission from *PNAS* 96, 4256. Copyright ©1999 National Academy of Sciences, U.S.A.]. Reproduced in color between pages 172 and 173



2.96

Kuntz *et al.*[259] reported on an algorithm called *DOCK* that was designed to fit small molecules into their macromolecular receptors for lead discovery.[260] This shape-matching method, which was originally restricted to rigid ligands (receptor-bound molecules) and receptors, was modified[261] for flexible ligands where a ligand is approximated as a small set of rigid fragments. Ideally, a high-resolution structure (X-ray crystal structure or NMR spectral structure) of the receptor *with a ligand bound* should be available. The ligand is removed from the binding site in the graphic display, then DOCK fills the binding site with sets of overlapping spheres, where a set of sphere centers serves as the negative image of the binding site. When a crystal structure of a receptor is available, but without a ligand bound, DOCK characterizes the entire surface of the receptor with regard to grooves that could potentially form target binding sites, which are filled with the overlapping spheres. Next DOCK matches X-ray or computer-derived structures of putative ligands to the image of the receptor on the

basis of a comparison of internal distances. Then the program searches 3D databases of small molecules and ranks each candidate on the basis of the best orientations that can be found for a particular molecular conformation.[262] Various databases are available to search, such as the Cambridge Structural Database (CSD), a compendium of >200,000 small molecules whose crystal structures are known and the Fine Chemicals Directory (FCD) distributed by Molecular Design Limited, but the best ones to use are those containing commercially available compounds, such as the Available Chemicals Directory (ACD), so that any virtual hits can be purchased and assayed to determine quickly the effectiveness of the search. The drawbacks of this approach are the assumptions that binding is determined primarily by shape complementarity and that only small changes in the shape of the receptor occur on ligand binding. An important advantage, though, is that this method is not limited to docking of known ligands. A library of molecular shapes can be scanned to determine which shapes best fit a particular receptor binding site. In fact, DOCK was used to identify fullerenes as potential inhibitors of HIV-1 protease.[263] The high-resolution NMR structure of the aminoglycoside antibiotic paromomycin (**2.97**, Humatin) bound to the A site of the bacterial ribosomal RNA was used to perform a DOCK search of the CSD and the National Cancer Institute 3D database (a total of 273,000 compounds).[264] The compounds that emerged from this search formed the basis for the design of seven composite structures with additional features added to suppress resistance. As a result, several of these compounds were found to have enhanced activity *in vitro* and *in vivo* against a variety of pathogenic bacteria resistant to aminoglycosides.



2.97

The program *LUDI* uses statistical analyses of nonbonded contacts in crystal packings of organic molecules to establish a set of rules that define the possible nonbonded contacts between proteins and ligands.[265] Using these rules it also can search databases to find structures that fit a particular binding site in a protein based not on shape, as in DOCK, but on physicochemical properties, such as hydrogen bonding, ionic interactions, and hydrophobic interactions.

Another program for *de novo* molecular design, GrowMol[266] (called AlleGrow[267] in the latest version), approaches the problem of receptor binding from a different direction. Instead of docking known molecules into the binding site, it generates molecules with steric and chemical complementarity to the three-dimensional structure of the receptor binding site by evaluating each new atom according to its chemical complementarity to the nearby receptor atoms. The program also connects a newly grown atom to a previously grown atom in the growing structure to make ring systems. The principal "liability" to this method is that it generates too many diverse structures, and it is necessary to evaluate each one visually and determine which ones are best to try first. Often the decision comes from a synthetic perspective or a knowledge of potential oral bioavailability. Having too many choices, of course, also can be an asset, and a variety of other criteria could be set up to search the database of newly generated compounds for specific beneficial properties.

In the modified method for docking flexible ligands into a receptor described above, an X-ray structure of the receptor is not necessary to characterize the shape of the receptor binding site. Rather, the receptor binding site can be deduced from the shapes of active ligands. This technique, which is useful for identification of the pharmacophore geometry, is called *receptor mapping*.[268] It too is founded on the premise that receptor topography is complementary to that of drugs, but in this case the structure of the lock is deduced from the shape of the keys that fit it. A variety of receptor mapping techniques have been described. An approach termed *steric mapping*[269] uses molecular graphics to combine the volumes of compounds known to bind to the desired receptor. This composite volume generates an enzyme-excluded volume map, which defines that region of the binding site available for binding by drug analogs and, therefore, not occupied by the receptor itself. The same procedure is, then, carried out for similar molecules that are inactive. The composite volume is inspected for regions of volume overlap common to all of the inactive analogs. These are the enzyme-essential regions, sites required by the receptor itself and unavailable for occupancy by ligands. Any other molecule that overlaps with these regions should be inactive. Drug design, then, would involve the synthesis of compounds with the appropriate pharmacophore that filled the enzyme-excluded regions and that avoided the enzyme-essential regions. Another approach that does not require the structure of the target receptor, known as *homology modeling*, deduces the topography of the unknown receptor site from that of a known related receptor structure.[270]

A major improvement in the use of molecular modeling came when high-throughput crystallography was coupled with combinatorial chemistry approaches.[271] Because structure-based drug design usually involves targets whose structures are already known, but which have different ligands bound, only the part of the structure where the ligand binds needs to be resolved. Software such as AutoSolve (Astex Technology) analyzes and interprets electron density data automatically without the need for an expert crystallographer, so hundreds of receptor complex crystals can be analyzed in just a few days.[272] With solid-phase synthetic methodology to make many analogs rapidly, the two processes produce large numbers of crystal structures with various ligands bound very rapidly.

The initial expectation for structure-based drug design—that potent receptor binders would be designed rapidly leading to the discovery of many new drugs—has not yet become a reality. Several problems with this approach may contribute to its less than optimal effectiveness. Table 2.12 lists various advantages to the use of molecular modeling approaches and its many limitations. Although the ease of visualization is appealing, the main problems are (1) that the structure of the molecular model may be completely different from the actual structure in the living organism; (2) even if the structure were correct, the resolution of the structure is insufficient to make an accurate assessment of ligand binding; and (3) the bioactive conformation of the ligands is not known, so the appropriate small molecules may not be used in docking experiments. Another important reason why there has not been a large increase in the number of new drugs being developed by molecular modeling techniques derives from the fact that pharmacokinetics are ignored by this method. Prior to the drug candidate interacting with a receptor, it must be properly absorbed, it must reach the receptor without metabolic or chemical degradation (unless it is a prodrug; see Chapter 8), excretion must be at an appropriate rate, and the drug candidate and metabolites must not be toxic or lead to undesirable side effects.

Because of the uncertainty involved with this method, the process of molecular modeling, synthesis, testing, and molecular modeling again needs to undergo many iterations. Structure-based drug design has to be taken as yet another tool available to the medicinal chemist; it is not the answer to drug discovery, but it can be an important part of the process.

TABLE 2.12 ▶ Advantages and Limitations to the Use of Molecular Modeling in Lead Modification

Advantages

- ▶ Proteins can be visualized in 3D and every amino acid can be located.
- ▶ The structure can be manipulated so that it can be observed from any direction in 3D.
- ▶ Particular regions, e.g., the binding site, can be enlarged for better viewing.
- ▶ The physicochemical properties, e.g., hydrophobic, polar, positive or negative charge, etc., of each part of the receptor can be viewed.
- ▶ Distances between groups can be determined.
- ▶ Small molecules can be docked into various regions to determine their fit and interactions.
- ▶ Residues that are most suitable to mutate for mechanism studies can be determined.

Limitations

- ▶ The coordinates from an X-ray crystal structure or NMR solution structure are required.
- ▶ Crystals are obtained by crystallization of proteins under nonphysiological conditions, such as at low or high pH, well below 37°C, and in the presence of additives, such as buffers or detergents. Are the proteins really in the same conformation as in the living cell?
- ▶ Crystal structures represent the thermodynamically most stable conformation of the protein under these nonphysiological conditions. Therefore, the crystal structure may depict the protein in a conformation very different from that in a living cell.
- ▶ Often crystal structures with ligands bound are obtained by soaking the ligand into the preformed crystal. If binding of the ligand in solution results in a conformational change, it is highly unlikely that it will occur in the crystalline state because the crystal packing forces will favor the preexisting conformation.
- ▶ The protein structure is considered to be rigid, but small conformational changes of side chains can induce large changes in the size, shape, and interaction pattern of binding pockets. Typically, when a small molecule binds to a protein, there is some movement of side chains.
- ▶ Resolutions of crystal structures are generally in the range of 2–2.5 Å; some at 1.5–2.0 Å; rarely more resolved (although <1.0 Å has been reported).[a] Therefore, there is *much* uncertainty as to the exact position of each atom. A rule of thumb is that the positional error of atoms is about one-sixth of the resolution,[b] so a structure at 2.4 Å resolution has an uncertainty of every atom of 0.4 Å.
- ▶ Small molecules in the ground state are generally energy minimized to give the lowest energy conformers prior to docking them into the structure, but a ligand does not have to bind in the lowest energy conformation, and it can be quite different from the ground state conformation. Also, solvent effects generally are not taken into account.
- ▶ For highly flexible molecules with several torsional angles, there may be many different geometries having the same conformational energy, but significantly different shapes.
- ▶ You tend to believe what you see in your molecular model and think it is accurate! This leads to many wrong assumptions.

[a] For example, Betzel, C.; Gourinath, S.; Kumar, P.; Kaur, P.; Perbrandt, M.; Eschenburg, S.; Singh, T. P. *Biochemistry* **2001**, *40*, 3080.
[b] Böhm, H.-J.; Klebe, G. *Angew. Chem. Int. Ed. Engl.* **1996**, *35*, 2588.

2.2.I Epilogue

On the basis of what was discussed in this chapter, it appears that even if you uncover a lead, it may be a fairly slow and random process to optimize its potency. The cost to get a drug on the market has increased from $4 million in 1962 to $350 million in 1996, $500 million in 2000,[273a] and about $800 million in 2003.[273b] Between 1960 and 1980, the time for development of a compound from synthesis to the market almost quadrupled, but the time has remained fairly constant since 1980 at about 12–15 years of research. The main cause for the increase in the length of time to bring a drug on the market occurred in 1962 as a result of the devastating effects of the drug thalidomide, a hypnotic drug shown to cause severe fetal limb abnormalities (phocomelia) when taken in the first trimester of pregnancy (see Chapter 3, Section 3.2.E.2). This tragedy led to the passing of the Harris-Kefauver Amendments to the Food, Drug, and Cosmetic Act in 1962, which required sufficient pharmacological and toxicological research in animals before a drug could be tested in humans; the data of the animal studies had to be submitted to the FDA in an application for approval of an investigational new drug (IND) before human testing could begin. After 1–5 years (average 2.6 years) of animal testing, three phases of clinical (human) trials were adopted (lasting from 4 to 10 years; see the first paragraph of this chapter for a description of the phases of clinical trials) before a new drug application (NDA) could be submitted for commercial approval of a new drug.[274]

It has been estimated that, in 1950, 7000 compounds had to be isolated and tested for each one that made it to the market; by 1979 that number rose to 10,000 compounds, and now it is greater than 20,000 compounds. There are only about 6000 known drugs in the Comprehensive Medicinal Chemistry (CMC) database of the estimated 10^{60} possible compounds that could be drug-like, and these 6000 drugs interact with only about 120 targets[275a] (40% receptors, 40% enzymes, and the rest ion channels and other) or <1% of the human *proteome* (the expressed proteins). Prior to the sequencing of the human genome, it was estimated that the number of potential drug targets may be between 5000 and 10,000,[275b] but it is now thought that it may only be 600–1500.[275a] Therefore, *genomics* (identifying and analyzing new gene targets from a genome) and *proteomics* (identifying and analyzing proteins expressed by the genes in the genome) have become very important aspects to drug discovery.[276] Once a new target from the proteome is identified, *bioinformatics*, in which databases of known proteins are scanned to find known proteins with similar structures to that of the new target, is employed. When the similarities are known, inhibitors of the known protein can be tested with the new target protein. In addition to these biological methodologies, which appear to be increasing the rate of lead discovery, other rational approaches to lead discovery and lead optimization, based on chemical and biochemical principles, must be used. Between 1995 and 2000 it was estimated at Bristol-Myers Squibb that there had been a threefold to fourfold increase in new drug candidates going into development, a 50% lower chemistry staff requirement per drug candidate, and a 40% reduction in lead optimization time, believed to be the result of combinatorial approaches.[277] Other companies have not enjoyed the predicted success of combinatorial chemistry, and some have even dropped their combinatorial chemistry groups and returned to only traditional medicinal chemistry efforts. The importance of combinatorial chemistry to drug discovery will not be known for at least 10 years when we find out if there is a direct link of this approach to new drugs entering the market or if it is just another false hope. However, in 2002, for the first time in the United States, the market share of nongeneric drugs was surpassed by that of generic drugs. Also in that year the number of new chemical entity approvals by the FDA, normally in the twenties or thirties per year, hit a 20-year low of only 16, although the R&D spending by the pharmaceutical industry had tripled in the previous decade.[278] Maybe Thomas Edison said it best: "I have not failed. I've just found 10,000 ways that won't work."

2.3 General References

Combinatorial Chemistry

Books and Reviews

Affleck, R. L. Solutions for library encoding to create collections of discrete compounds. *Curr. Opin. Chem. Biol.* **2001**, *5*, 257–263.
Barnes, C.; Balasubramanian, S. Recent developments in the encoding and deconvolution of combinatorial libraries. *Curr. Opin. Chem. Biol.* **2000**, *4*, 346–350.
Czarnik, A. W. Encoding methods for combinatorial chemistry. *Curr. Opin. Chem. Biol.* **1997**, *1*, 60–66.
Czarnik, A. W. (Ed.) *Solid-Phase Organic Synthesis*, Wiley, New York, Vol. 1, 2001 (annual series).
Fenniri, H. *Combinatorial Chemistry. A Practical Approach*, Oxford University Press, Oxford, UK, 2000.
Gordon, E. M.; Kerwin, J. F. (Eds.) *Combinatorial Chemistry and Molecular Diversity in Drug Discovery*, Wiley, New York, 1998.
Jung, G. (Ed.) *Combinatorial Chemistry*, Wiley-VCH, Weinheim, 1999.
Jung, G. (Ed.) *Combinatorial Peptide and Nonpeptide Libraries: A Handbook*, Wiley-VCH, Weinheim, 1996.
Wilson, S. R.; Czarnik, A. W. (Eds.) *Combinatorial Chemistry: Synthesis and Application*, Wiley, New York, 1997.
Yan, B., Czarnik, A. W. (Eds.) *Optimization of Solid-Phase Combinatorial Synthesis*, Marcel Dekker, New York, 2002.

Journals

Combinatorial Chemistry & High Throughput Screening
Journal of Combinatorial Chemistry
Journal of Medicinal Chemistry
Molecular Diversity

Webpages

http://www.5z.com
http://www.warr.com/ombichem.html (yes, "ombichem")
http://www.combi-web.com

Peptidomimetics

Bursavich, M. G.; Rich, D. H. *J. Med. Chem.* **2002**, *45*, 541.
Giannis, A.; Kolter, T. *Angew. Chem. Int. Ed. Engl.* **1993**, *32*, 1244.
Ripka, A. S.; Rich, D. H. *Current Opin. Chem. Biol.* **1998**, *2*, 441.

Pharmacokinetics

Hardman, J. G.; Limbird, L. E.; Gilman, A. G. *Goodman and Gilman's The Pharmacological Basis of Therapeutics*, 10th ed., McGraw-Hill, New York, 2001.

Smith, D. A.; van de Waterbeemd, H.; Walker, D. K.; Mannhold, R.; Kubinyi, H.; Timmerman, H. *Pharmacokinetics and Metabolism in Drug Design*, Wiley, New York, 2000.

QSAR

Books

Devillers, J. (Ed.) *Comparative QSAR*, Taylor and Francis, Washington, DC, 1998.
Hansch, C.; Leo, A. *Exploring QSAR*, Vol. 1, *Fundamentals and Applications in Chemistry and Biology*, ACS Publ., Washington, DC, 1995.
Kubinyi, H. *QSAR: Hansch Analysis and Related Approaches*, VCH, Weinheim, 1993.
Kubinyi, H. (Ed.) *3D-QSAR in Drug Design. Theory, Methods and Applications*, ESCOM, Leiden, Netherlands, 1993.
Leo, A.; Hansch, C.; Hoekman, D. *Exploring QSAR*, Vol. 2, *Hydrophobic, Electronic, and Steric Constants*, ACS Publ., Washington, DC, 1995.
Lipkowitz, K. B., Boyd, D. B. (Eds.) *Reviews in Computational Chemistry*, Wiley-VCH, New York (entire series of volumes).

Journals

Journal of Chemical Information and Computer Science
Journal of Medicinal Chemistry

Molecular Graphics

Books and Reviews

Antel, J. *Curr. Opin. Drug Discovery Dev.* **1999**, *2*, 224.
Bohacek, R. S.; McMartin, C.; Guida, W. C. *Med. Res. Rev.* **1996**, *16*, 3.
Böhm, H.-J. *Prog. Biophys. Mol. Biol.* **1996**, *66*, 197.
Böhm, H.-J.; Stahl, M. *Curr. Opin. Chem. Biol.* **2000**, *4*, 283.
Gubernator, K.; Böhm, H.-J. (Eds.) *Structure-Based Ligand Design*, Wiley-VCH, Weinheim, 1998.
Kirkpatrick, D. L.; Watson, S.; Ulhaq, S. *Combinatorial Chemistry. High Throughput Screening* **1999**, *2*, 211.
Kuntz, I. D.; Meng, E. C.; Shoichet, B. K. *Acc. Chem. Res.* **1994**, *27*, 117–123.
Ooms, F. *Curr. Med. Chem.* **2000**, *7*, 141.

Journals

Annual Reports in Combinatorial Chemistry and Molecular Design
Journal of Computer-Aided Molecular Design
Journal of Medicinal Chemistry
Journal of Molecular Graphics

Software

Sybyl™ (Tripos, Inc.)
Insight II™ (MDL, Inc.)

Molecular Conceptor™ Courseware (Synergix, Ltd.)
CaChe™ Software (Fujitsu, Inc.)

Webpages

http://www.netsci.org/Resources/Software/Modeling/CADD

Computer-Based Drug Design Methodologies[a]

Active Site Analysis

MCSS

Caflish, A.; Miranker, A.; Karplus, M. Multiple copy simultaneous search and construction of ligands in binding sites: application to inhibitors of HIV-1 aspartic proteinase. *J. Med. Chem.* **1993**, *36*, 2142–2167.

Stultz, C. M.; Karplus, M. MCSS functionality maps for a flexible protein. *Proteins* **1999**, *37*, 512–529.

GRID

Boobbyer, D. N. A.; Goodford, P. J.; McWhinnie, P. M.; Wade, R. C. New hydrogen-bond potentials for use in determining energetically favorable binding sites on molecules of known structure. *J. Med. Chem.* **1989**, *32*, 1083–1094.

Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857.

Wade, R. C.; Goodford, P. J. Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 1. Ligand probe groups with the ability to form two hydrogen bonds. *J. Med. Chem.* **1993**, *36*, 140–147.

Wade, R. C.; Goodford, P. J. Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 2. Ligand probe groups with the ability to form more than two hydrogen bonds. *J. Med. Chem.* **1993**, *36*, 148–156.

Applications of MCSS and GRID

Bitetti-Putzer, R.; Joseph-McCarthy, D.; Hogle, J. M.; Karplus, M. Functional group placement in protein binding sites: a comparison of GRID and MCSS. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 935–960.

Powers, R. A.; Shoichet, B. K. Structure-based approach for binding site identification on AmpC β-lactamase. *J. Med. Chem.* **2002**, *45*, 3222–3234.

MCSS Applied to Structure-Based Ligand Design

Joseph-McCarthy, D.; Tsang, S. K.; Filman, D. J.; Hogle, J. M.; Karplus, M. Use of MCSS to design small targeted libraries: application to picornavirus ligands. *J. Am. Chem. Soc.* **2001**, *123*, 12758–12769.

Stultz, C. M.; Karplus, M. Dynamic ligand design and combinatorial optimization: designing inhibitors to endothiapepsin. *Proteins* **2000**, *40*, 258–289.

[a] Many thanks to Dr. Haitao Ji for compiling these methodologies.

GRID Applied to 3D-QSAR

Bohm, M.; Klebe, G. Development of new hydrogen-bond descriptors and their application to comparative molecular field analysis. *J. Med. Chem.* **2002**, *45*, 1585–1597.

Davis, A. M.; Gensmantel, N. P.; Johansson, E.; Marriott, D. P. The use of the GRID program in the 3D-QSAR analysis of a series of calcium-channel agonists. *J. Med. Chem.* **1994**, *37*, 963–972.

Cruciani, G.; Watson, K. A. Comparative molecular field analysis using GRID force-field and GOLPE variable selection methods in a study of inhibitors of glycogen phosphorylase b. *J. Med. Chem.* **1994**, *37*, 2589–2601.

Pastor, M.; Cruciani, G.; Mclay, I.; Pickett, S.; Clemente, S. Grid-INdependent Descriptors (GRIND): a novel class of alignment-independent 3-D molecular descriptors. *J. Med. Chem.* **2000**, *43*, 3233–3243.

GRID Applied to Selectivity Analysis

Kastenholz, M. A.; Pastor, M.; Gruciani, G.; Haaksma, E. E.; Fox, T. GRID/CPCA: A new computational tool to design selective ligand. *J. Med. Chem.* **2000**, *43*, 3033–3044.

X-SITE

Laskowski, R. A.; Thornton, J. M.; Humblet, C.; Singh, J. X-SITE: use of empirically derived atomic packing preferences to identify favourable interaction regions in the binding sites of proteins. *J. Mol. Biol.* **1996**, *259*, 175–201.

Molecular Docking

DOCK, GOLD, and FlexX are used as virtual screening tools when the 3-D structure of the binding site of the receptor is known.

Fast Shape Matching (DOCK)

Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **1982**, *161*, 269–288.

Incremental Construction (FlexX, Hammerhead, Surflex)

Jain, A. N. Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *J. Med. Chem.* **2003**, 46, 499–511.

Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using as incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470–489.

Welch, W.; Ruppert, J.; Jain, A. N. Hammerhead: fast fully automated docking of flexible ligands to protein binding sites. *Chem. Biol.* **1996**, *3*, 449–462.

Tabu Search (Pro_Leads)

Baxter, C. A.; Murray, C. W.; Clark, D. E.; Westhead, D. R.; Eldridge, M. D. Flexible docking using Tabu search and an empirical estimate of binding affinity. *Proteins* **1998**, *33*, 367–382.

Genetic Algorithm (GOLD, AutoDock 3.0)

Jones, G.; Wilett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748.

Morris, G. M.; Goodsell, D. S.; Halliday, R.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639–1662.

Monte Carlo Simulations (ICM, MCDOCK, QXP)

Abagyan, R. A.; Totrov, M. M. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J. Mol. Biol.* **1994**, *235*, 983–1002.

Liu, M.; Wang, S. MCDOCK: a Monte Carlo simulation approach to the molecular docking problem. *J. Comput-Aided Mol. Des.* **1999**, *13*, 435–451.

McMartin, C.; Bohacek, R. S. QXP: powerful, rapid computer algorithms for structure-based drug design. *J. Comput-Aided Mol. Des.* **1997**, *11*, 333–344.

Simulated Annealing (AutoDock 2.4)

Goodsell, D. S.; Olson, A. J. Automated docking of substrates to proteins by simulated annealing. *Proteins* **1990**, 8, 195–202.

Postdocking Treatments

Free-Energy Perturbation (FEP)

FEP is an accurate method, but it is very time consuming.

Kollman, P. A. Advances and continuing challenges in achieving realistic and predictive simulations of the properties of organic and biological molecules. *Acc. Chem. Res.* **1996**, *29*, 461–469.

Kollman, P. A. Free-energy calculations—applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417.

Toba, S.; Damodaran, K. V.; Merz, Jr., K. M. Binding preferences of hydroxamate inhibitors of the matrix metalloproteinase human fibroblast collagenase. *J. Med. Chem.* **1999**, *42*, 1225–1234.

OWFEG

OWFEG can be used to simplify the FEP method.

Pearlman, D. A. Free energy grids: a practical qualitative application of free energy perturbation to ligand design using the OWFEG method. *J. Med. Chem.* **1999**, *42*, 4313–4324.

Thermodynamic Integration (TI)

Guimaraes, C. R. W.; Bicca de Alencastro, R. Thermodynamic analysis of thrombin inhibition by benzamidine and *p*-methylbenzamidine via free-energy perturbations: inspection of intraperturbed-group contributions using the finite difference thermodynamic integration (FDTI) algorithm *J. Phys. Chem. B.* **2002**, *106*, 466–476.

McDonald, J. J.; Brooks, C. L. Theoretical approach to drug design. 2. Relative thermodynamics of inhibitor binding by chicken dihydrofolate reductase to ethyl derivatives of trimethoprim substituted at 3-, 4-, and 5-positions. *J. Am. Chem. Soc.* **1991**, *113*, 2295–2301.

Reddy, M. R.; Viswanadhan, V. N.; Weinstein, J. N. Relative differences in the binding free energies of human immunodeficiency virus 1 protease inhibitors: a thermodynamic cycle-perturbation approach. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 10287–10291.

Scoring Methods for Virtual Screening

Force Field Scoring Functions

DOCK

Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **1982**, *161*, 269–288.

GOLD

Jones, G.; Wilett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748.

VALIDATE

Head, R. D.; Smythe, M. L.; Opera, T. I.; Waller, C. L.; Green, S. M.; Marshall, G. R. VALIDATE: a new method for the receptor-based prediction of binding affinities of novel ligands. *J. Am. Chem. Soc.* **1996**, *118*, 3959–3969.

Empirical Free-Energy Scoring Functions

LUDI

Bohm, H. J. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J. Comput-Aided Mol. Des.* **1994**, *8*, 243–256.

Chemscore

Eldrige, M.; Murray, C. W.; Auton, T. A.; Paolini, G. V.; Lee, R. P. Empirical scoring functions: I. the development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425–445.

FlexX

Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using as incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470–489.

Score

Wang, R.; Liu, L.; Lai, L.; Tang, Y. Score: a new empirical method for estimating the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425–445.

Fresno

Rognan, D.; Lauemoller, S. L.; Holm, A.; Buus, S.; Tschinke, V. Predicting binding affinities of protein ligands from three-dimensional models: application to peptide binding to class I major histocompatibility proteins. *J. Med. Chem.* **1999**, *42*, 4650–4658.

Knowledge-Based Scoring Functions

PMF

Muegge, I.; Martin, Y. C. A general and fast scoring function for protein–ligand interactions: a simplified potential approach. *J. Med. Chem.* **1999**, *42*, 791–804.

DrugScore

Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based scoring function to predict protein–ligand interaction. *J. Mol. Biol.* **2000**, *295*, 337–356.

AFMoC

Gohlke, H.; Klebe, G. DrugScore meets CoMFA: adaptation of fields for molecular comparison (AFMoC) or how to tailor knowledge-based pair-potentials to a particular protein. *J. Med. Chem.* **2002**, *45*, 4153–4170.

SmoG2001

Ishchenko, A. V.; Shakhnovich, E. I.; Small molecule growth 2001 (SmoG2001): An improved knowledge-based scoring function for protein–ligand interactions. *J. Med. Chem.* **2002**, *45*, 2770–2780.

Consensus Scoring

Charifson, P. S.; Corkey, J. J. Murcko, M. A.; Walters, W. P. Consensus scoring: a method for obtaining improved hit rates from docking databases of three-dimensional structure into proteins. *J. Med. Chem.* **1999**, *42*, 5100–5109.

Terp, G. E.; Johansen, B. N.; Christensen, I. T. ; Jorgensen, F. S. A new concept for multidimensional selection of ligand conformations (MultiSelect) and multidimensional scoring (MultiScore) of protein-ligand binding affinities. *J. Med. Chem.* **2001**, *44*, 2333–2343.

De Novo Lead Design

Site-Point Connection

LUDI

Bohm, H. J. LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 593–606.

CLIX

Lawrence, M. C.; Davis, P. C. CLIX: a search algorithm for finding novel ligands capable of binding proteins of known three-dimensional structure. *Proteins* **1992**, *12*, 31–41.

Fragment-Joining Method

SAR by NMR

Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Discovering high-affinity ligands for proteins. *Science* **1997**, *278*, 497–499.

Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **1996**, *274*, 1531–1534.

MCSS/HOOK

Eisen, M. B.; Wiley, D. C.; Karplus, M.; Hubbard, R. E. HOOK: a program for finding novel molecular architectures that satisfy the chemical and steric requirements of a macromolecule binding site. *Proteins* **1994**, *19*, 199–221.

CAVEAT

Lauri, G.; Bartlett, P. A. CAVEAT: a program to facilitate the design of organic molecules. *J. Comput-Aided Mol. Des.* **1994**, *8*, 51–66.

PRO-LIGAND

Clark, D. E.; Frenkel, D.; Levy, S. A.; Li, J.; Murray, C. W.; Robson, B.; Waszkowycz, B.; Westhead, D. R. PRO-LIGAND: an approach to *de novo* molecular design. 1. Application to the design of organic molecules. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 13–32.

Waszkowycz, B.; Clark, D. E.; Frenkel, D.; Li, J.; Murray, C.W.; Robson, B.; Westhead, D. R. PRO-LIGAND: an approach to *de novo* molecular design. 2. design of novel molecules from molecular field analysis (MFA) models and pharmacophores. *J. Med. Chem.* **1994**, *37*, 3994–4002.

NEWLEAD

Tschinke, V.; Cohen, N. C. The NEWLEAD program: a new method for the design of candidate structures from pharmacophoric hypotheses. *J. Med. Chem.* **1993**, *36*, 3863–3870.

Sequential Buildup (Atom Grow)

LEGEND

Honma, T.; Hayashi, K.; Aoyama, T.; Hashimoto, N.; Machida, T.; Fukasawa, K.; Iwama, T.; Ikeura, C.; Ikuta, M.; Suzuki-Takahashi, I.; Iwasawa, Y.; Hayama, T.; Nishimura, S.; Morishima, H. Structure-based generation of a new class of potent CDK4 inhibitors: new *de novo* design strategy and library design. *J. Med. Chem.* **2001**, *44*, 4615–4627.

Nishibata, Y.; Itai, A. Automatic creation of drug candidate structures based on receptor structure. Starting point for artificial lead generation. *Tetrahedron* **1991**, *47*, 8985–90.

Nishibata, Y.; Itai, A. Confirmation of usefulness of a structure construction program based on three-dimensional receptor structure for rational lead generation. *J. Med. Chem.* **1993**, *36*, 2921–8.

LeapFrog from SYBYL

GROW

Moon, J. B.; Howe, W. J. Computer design of bioactive molecules: a method for receptor-based *de novo* ligand design. *Proteins* **1991**, *11*, 314–328.

GROWMOL (more recent version is called AlleGrow; http://www.bostondenovo.com)

Bohacek, R. S.; McMartin, C. Multiple highly diverse structures complementary to enzyme binding sites: results of extensive application of a *de novo* design method incorporating combinatorial growth. *J. Am. Chem. Soc.* **1994**, *116*, 5560–5571.

SPROUT

Gillet, V. J.; Newell, W.; Mata, P.; Myatt, G.; Sike, S.; Zsoldos, Z.; Johnson, A. P. SPROUT: Recent developments in the *de novo* design of molecules. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 207–217.

Random Connections

MCSS/DLD

Stultz, C. M.; Karplus, M. Dynamic ligand design and combinatorial optimization: designing inhibitors to endothiapepsin. *Proteins* **2000**, *40*, 258–289.

CONCERTS

Pearlman, D. A.; Murcko, M. A. CONCERTS: dynamic connection of fragments as an approach to *de novo* ligand design. *J. Med. Chem.* **1996**, *39*, 1651–1663.

Virtual Combinatorial Screening

Methods for Virtual Combinatorial Screening

Legion from SYBYL

PRO_SELECT

Liebeschuetz, J. W.; Jones, S. D.; Morgan, P. J.; Murray, C. W.; Rimmer, A. D.; Roscoe, J. M.; Waszkowycz, B.; Welsh, P. M.; Wylie, W. A.; Young, S. C.; Martin, H.; Mahler, J.; Brady, L.; Wilkinson, K. PRO_SELECT: combining structure-based drug design and array-based chemistry for rapid lead discovery. 2. The development of a series of highly potent and selective factor Xa inhibitors. *J. Med. Chem.* **2002**, *45*, 1221–1232.

Murray, C. W.; Clark, D. E.; Auton, T. R.; Firth, M. A.; Li, J.; Sykes, R. S.; Waszkowycz, B.; Westhead, D. R.; Young. S. C. PRO_SELECT: Combining structure-based drug design and combinatorial chemistry for rapid lead discovery. 1. Technology. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 193–207.

Combinatorial Library Design

Drug-Likeness

MoSELECT

Gillet, V. J.; Khatib, W.; Willett, P.; Fleming, P. J.; Green, D. V. Combinatorial library design using a multiobjective genetic algorithm. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 375–385.

Gillet, V. J.; Willett, P.; Fleming, P. J.; Green, D. V. Designing focused libraries using MoSELECT. *J. Mol. Graph Model* **2002**, *20*, 491–498.

REOS

Walters, W. P.; Murcko, M. A. Prediction of "drug-likeness." *Adv. Drug Deliv. Rev.* **2002**, *54*, 255–271.

Molecular Diversity

Martin, Y. C. Challenges and prospects for computational aids to molecular diversity. *Perspectives in Drug Discovery and Design* **1997**, 7/8, 159–172.

DiverseSolution from SYBYL

Cramer, R. D.; Clark, R. D.; Petterson, D. E.; Ferguson, A. M. Bioisosterism as a molecular diversity descriptor: steric fields of single "topomeric" conformers. *J. Med. Chem.* **1996**, *39*, 3060.

Patterson, D. E.; Cramer, R. D.; Ferguson, A. M.; Clark, R. D.; Weinberger, L. E. Neighborhood behavior: a useful concept for validation of molecular diversity descriptors. *J. Med. Chem.* **1996**, *39*, 3049.

FlexSim-S

Briem, H.; Lessel, U. F. *In vitro* and *in silico* affinity fingerprints: finding similarities beyond structural classes. *Perspective in Drug Discovery and Design* **2000**, *20*, 231–244.

LASSOO

Koehler, R. T.; Dixon, S. L.; Villar, H. O. LASSOO: A generalized directed diversity approach to the design and enrichment of chemical libraries. *J. Med. Chem.* **1999**, *42*, 4695–4704.

Others

Andrews, K. M.; Cramer, R. D. Toward general methods of targeted library design: topomer shape similarity searching with diverse structures as queries. *J. Med. Chem.* **2000**, *43*, 1723–1740.

Makara, G. M. Measuring molecular similarity and diversity: total pharmacophore diversity. *J. Med. Chem.* **2001**, *44*, 3563–3571.

Mason, J. S.; Morize, I.; Menard, P. R.; Cheney, D. L.; Hulme, C.; Labaudiniere, R. F. New 4-point pharmacophore method for molecular similarity and diversity applications: overview of the method and applications, including a novel approach to the design of combinatorial libraries containing privileged substructures. *J. Med. Chem.* **1999**, *42*, 3251–3264.

Mount, J.; Ruppert, J.; Welch, W.; Jian, A. N. Ice Pick: A flexible surface-based system for molecular diversity. *J. Med. Chem.* **1999**, *42*, 60–66.

Srinivasan, J.; Castellino, A.; Bradley E. K.; Eksterowicz, J. E.; Grootenhuis P. D. J.; Putta, S.; Stanton R. Evaluation of a novel shape-based computational filter for lead evolution: application to thrombin inhibitors. *J. Med. Chem.* **2002**, *45*, 2494–2500.

Synthetic Accessibility

Gasteiger, J.; Pfortner, M.; Sitzmann, M.; Hollering, R.; Sacher, O.; Kostka, T.; Karg, N. Computer-assisted synthesis and reaction planning in combinatorial chemistry. *Perspective in Drug Discovery and Design* **2000**, *20*, 245–264.

Gillet, V. J.; Nicolotti, O. Evaluation of reactant-based and product-based approaches to the design of combinatorial libraries. *Perspective in Drug Discovery and Design* **2000**, *20*, 265–287.

Lewell, X. Q.; Judd, D. B.; Watson, S. P.; Hann, M. M. RECAP—retrosynthetic combinatorial analysis procedure: a powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 511–522.

Ligand Optimization

ADME (Pharmacokinetics)

Property-Based Ligand Design

van De Waterbeemd, H; Smith, D. A.; Beaumont, K.; Walker, D. K. Property-based design: optimization of drug absorption and pharmacokinetics. *J. Med. Chem.* **2001**, *44*, 1313–1333.

VolSurf

Cruciani, G.; Pastor, M.; Guba, W. VolSurf: a new tool for the pharmacokinetic optimization of lead compounds. *Eur. J. Pharm. Sci.* **2000**, *Suppl 2*, S29–39.

ChemGPS

Alifrangis, L. H.; Christensen, I. T.; Berglund, A.; Sandberg, M.; Hovgaard, L.; Frokjaer, S. Structure-property model for membrane partitioning of oligopeptides. *J. Med. Chem.* **2000**, *43*, 103–113.

Crivori, P.; Cruciani, G.; Carrupt, P. A.; Testa, B. Predicting blood–brain barrier permeation from three-dimensional molecular structure. *J. Med. Chem.* **2000**, *43*, 2204–2216.

Cruciani, G.; Pastor, M.; Mannhold, R. Suitability of molecular descriptors for database mining. A comparative analysis. *J. Med. Chem.* **2002**, *45*, 2685–2694.

Egan, W. J.; Merz Jr., K. M.; Baldwin, J. J. Prediction of drug absorption using multivariate statistics. *J. Med. Chem.* **2000**, *43*, 3867–3877.

Ooms, F.; Weber, P.; Carrupt, P. A.; Testa, B. A simple model to predict blood–brain barrier permeation from 3D molecular fields. *Biochim. Biophys. Acta* **2002**, *1587*, 118–125.

Oprea, T. I.; Zamora, I.; Ungell, A. L. Pharmacokinetically based mapping device for chemical space navigation. *J. Comb. Chem.* **2002**, *4*, 258–266.

QSAR

2D-QSAR

Hansch method

Domine, D.; Devillers J.; Chastrette, M. A nonlinear map of substituent constants for selecting test series and deriving structure–activity relationships. 1. aromatic series. *J. Med. Chem.* **1994**, *37*, 973–980.

Domine, D.; Devillers, J.; Chastrette, M. A nonlinear map of substituent constants for selecting test series and deriving structure–activity relationships. 2. aliphatic series. *J. Med. Chem.* **1994**, *37*, 981–987.

Manallack, D. T.; Ellis, D. D.; Livingstone D. J. Analysis of linear and nonlinear QSAR data using neural networks. *J. Med. Chem.* **1994**, *37*, 3639–3654.

3D-QSAR

CoMFA (Comparative Molecular Field Analysis)

Cramer, R. D. Topomer CoMFA: a design methodology for rapid lead optimization. *J. Med. Chem.* **2003**, *46*, 374–388.

Cramer, R. D.; Paterson, D. E.; Bunce, J. D. Comparative molecular field analysis (CoMFA). I. Effect of shape on binding of steroids to carried proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.

Robinson, D. D.; Winn P. J.; Lyne P. D.; Richards, W. G. Self-organizing molecular field analysis: a tool for structure–activity studies. *J. Med. Chem.* **1999**, *42*, 573–583.

So, S.-S.; Karplus, M. Three-dimensional QSAR from molecular similarity matrices and genetic neural networks. 1. method and validations. *J. Med. Chem.* **1997**, *40*, 4347–4359.

So, S.-S.; Karplus, M. Three-dimensional QSAR from molecular similarity matrices and genetic neural networks. 2. applications. *J. Med. Chem.* **1997**, *40*, 4360–4371.

CoMSIA (Comparative Molecular Similarity Analysis)

Klebe, G.; Abraham, U.; Mietzner, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* **1994**, *37*, 4130–4146.

CoMMA (Comparative Molecular Moment Analysis)

Silverman, B. D.; Platt, D. E. Comparative molecular moment analysis (CoMMA): 3D-QSAR without molecular superposition. *J. Med. Chem.* **1996**, *39*, 2129–2140.

COMBINE (Comparative Binding Energy)

Ortiz, A. R.; Pisabarro, M. T.; Gago, F.; Wade R. C. Prediction of drug binding affinities by comparative binding energy analysis. *J. Med. Chem.* **1995**, *38*, 2681–2691.

Perez, C.; Pastor, M.; Ortiz, A. R. Gago F. Comparative binding energy analysis of HIV-1 protease inhibitors: incorporation of solvent effects and validation as a powerful tool in receptor-based drug design. *J. Med. Chem.* **1998**, *41*, 836–852.

GRID/GOLPE

Cho, S. J.; Tropsha A. Cross-validated R^2-guided region selection for comparative molecular field analysis: a simple method to achieve consistent results. *J. Med. Chem.* **1995**, *38*, 1060–1066.

Cruciani, G.; Watson, K. A. Comparative molecular field analysis using GRID force-field and GOLPE variable selection methods in a study of inhibitors of glycogen phosphorylase b. *J. Med. Chem.* **1994**, *37*, 2589–2601.

4D-QSAR

Klein, C. D.; Hopfinger, A. J. Pharmacological activity and membrane interactions of antiarrhythmics: 4D-QSAR/QSPR analysis. *Pharm. Res.* **1998**, *15*, 303–311.

5D-QSAR

Vedani, A.; Dobler, M. 5D-QSAR: the key for simulating induced fit? *J. Med. Chem.* **2002**, *45*, 2139–49.

Molecular Superposition

FlexS

Lemmen, C.; Lengauer, T.; Klebe, G. FLEXS: a method for fast flexible ligand superposition. *J. Med. Chem.* **1998**, *41*, 4502–20.

SQ

Miller, M. D.; Sheridan, R. P.; Kearsely, S. K. SQ: A program for rapidly producing pharmacophorically relevant molecular superpositions. *J. Med. Chem.* **1999**, *42*, 1505–1514.

QXP

McMartin, C.; Bohacek, R. S. QXP: powerful, rapid computer algorithms for structure-based drug design. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 333–344.

Pharmacophore Elucidation and Pharmacophore-Based Database Screening

Martin, Y. C. 3D database searching in drug design. *J. Med. Chem.* **1992**, 35, 2145–2154.

DISCO from SYBYL

Marriott, D. P.; Dougall, I. G.; Meghani, P. Liu Y.-J.; Flower, D. R. Lead generation using pharmacophore mapping and three-dimensional database searching: application to muscarinic M3 receptor antagonists. *J. Med. Chem.* **1999**, *42*, 3210–3216.

Martin, Y. C. DISCO: what we did right and what we missed. *IUL Biotechnology Series* **2000**, *2*, 49–68.

Martin, Y. C.; Bures, M. G.; Danaher, E. A.; DeLazzer, J.; Lico, I.; Pavlik, P. A. A fast new approach to pharmacophore mapping and its application to dopaminergic and benzodiazepine agonists. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 83–102.

GASP from SYBYL

Holliday, J. D.; Willett, P. Using a genetic algorithm to identify common structural features in sets of ligands. *J. Mol. Graph. Model.* **1997**, *15*, 221–232.

Jones, G.; Willett, P.; Glen, R. C. GASP: genetic algorithm superimposition program. *IUL Biotechnology Series* **2000**, *2*, 85–106.

RECEPTOR from SYBYL

Josien, H.; Convert, O.; Berlose, J.-P.; Sagan, S.; Brunissen, A.; Lavielle, S.; Chassaing, G. Topographic analysis of the S7 binding subsite of the tachykinin neurokinin-1 receptor. *Biopolymers* **1996**, *39*, 133–147.

UNITY from SYBYL

Catalyst

Kurogi, Y.; Guner, O. F. Pharmacophore modeling and three-dimensional database searching for drug design using catalyst. *Curr. Med. Chem.* **2001**, *8*, 1035–1055.