CHAPTER SIX

Virtual Screening

INGO MUEGGE
ISTVAN ENYEDY
Bayer Research Center
West Haven, Connecticut

Contents

1 Introduction, 244
2 Concepts of Virtual Screening, 244
2.1 Druglikeness Screening, 245
2.1.1 Counting Schemes, 245
2.1.2 Functional Group Filters, 246
2.1.3 Topological Drug Classification, 247
2.1.3.1 Artificial Neural Networks and Decision Trees, 247
2.1.3.2 Structural Frameworks and Side Chains of Known Drugs, 248
2.1.4 Pharmacophore Point Filter, 249
2.2 Focused Screening Libraries for Lead Identification, 250
2.2.1 Targeting Protein Families, 251
2.2.2 Privileged Structures, 251
2.3 Pharmacophore Screening, 252
2.3.1 Introduction to Pharmacophores, 252
2.3.2 Databases of Organic Compounds, 254
2.3.3 2D Pharmacophore Searching, 255
2.3.4 3D Pharmacophores, 255
2.3.4.1 Ligand-Based Pharmacophore Generation, 255
2.3.4.2 Manual Pharmacophore Generation, 256
2.3.4.3 Automatic Pharmacophore Generation, 256
2.3.4.4 Receptor-Based Pharmacophore Generation, 259
2.3.5 Pharmacophore-Based Virtual Screening, 259
2.4 Structure-Based Virtual Screening, 260
2.4.1 Protein Structures, 261
2.4.2 Computational Protein-Ligand Docking Techniques, 262
2.4.2.1 Rigid Docking, 262
2.4.2.2 Flexible Ligands, 263
2.4.3 Scaring of Protein-Ligand Interactions, 264
2.4.3.1 Force Field (FF) Scoring, 264

Burger's Medicinal Chemistry *and* Drug *Discovery*
Sixth Edition, Volume 1: Drug Discovery
Edited by Donald J. Abraham
ISBN 0-471-27090-3 © 2003 John Wiley & Sons. Inc.

2.4.3.2 Empirical Scoring, 264
2.4.3.3 Knowledge-Based Scoring, 264
2.4.3.4 Consensus Scoring, 265
2.4.4 Docking as Virtual Screening Tool, 266
2.5 Filter Cascade, 267
3 Applications, 267
3.1 Identification of Novel DAT Inhibitors through 3D Pharmacophore-Based Database Search, 267
3.2 Discovery of Novel Matriptase Inhibitors through Structure-Based 3D Database Screening, 269
4 Conclusions, 271
5 Acknowledgments, 274

1 INTRODUCTION

Virtual screening, sometimes also called *in silico* screening, is a new branch of medicinal chemistry that represents a fast and cost-effective tool for computationally screening compound databases in search for novel drug leads. The roots for virtual screening go back to structure-based drug design and molecular modeling. In the 1970s researchers hoped to find novel drugs designed rationally using a fast growing number of diverse protein structures being solved by X-ray crystallography (1, 2) or nuclear magnetic resonance (NMR) spectroscopy (3). However, only very few drugs have resulted from those early efforts. Examples include captopril as angiotensin-converting enzyme inhibitor (4) and methotrexate as dihydrofolate reductase inhibitor (5). The reasons for this somewhat disappointing drug yield lie in the low resolution of the protein structures as well as limitations in computer power and methods. Researchers have often tried *de novo* to design the final drug candidate on the computer screen. The compounds suggested have often been difficult to synthesize; initial failure in exhibiting potency has often resulted in the termination of structure-based projects. At the end of the 1980s rational drug design techniques became somewhat discredited because of the high failure rate in drug discovery projects.

In the 1990s drastic changes occurred in the way drugs are discovered in the pharmaceutical industry. High throughput synthesis (6, 7) and screening techniques (8) changed the lead identification process that is now governed not only by large numbers of compounds processed but also by fast prosecution of many putative drug targets in parallel. The characterization of the human genome has resulted in a large number of novel putative drug targets. Improved screening techniques also make it possible to look at entire gene families, at orphan targets, or at otherwise uncharacterized putative drug targets. In this environment of data explosion, rational design techniques have experienced a comeback (9). Although the exponentially growing number of solved protein structures at high resolution makes it possible to embark on structure-based design for many drug targets, virtual screening—the computational counterpart to high throughput screening—has become a particularly successful computational tool for lead finding in drug discovery. Whereas proprietary screening libraries typically hold about 10^6 compounds, this is only a tiny fraction of the conceivable chemical space for which estimates range between 10^{60} and 10^{100} compounds (10, 11). The question is, of course, which subset of this enormous space should be synthesized and screened? Virtual screening attempts to answer this question by evaluating large virtual libraries of up to 10^{12} compounds through the use of a cascade of various screening tools to reduce the chemical space. This chapter describes the different concepts and tools used today for virtual screening. They reach from the assessment of the overall "druglikeness" of a small organic molecule to its ability to specifically bind to a given drug target. The interested reader is also referred to a selection of recent books and reviews on the subject of virtual screening (10, 12–18).

2 CONCEPTS OF VIRTUAL SCREENING

The basic goal of virtual screening is the reduction of the enormous virtual chemical space of small organic molecules, to synthesize

and/or screen against a specific target protein, to a manageable number of compounds that exhibit the highest chance to lead to a drug candidate (10, 19). The major sources of information to guide virtual screening for a particular target are derived from the following questions:

1. What does a drug look like in general?
2. What is known about compounds that interact with the receptor?
3. What is known about the structure of the target protein and the protein-ligand interactions?

In the following subsections we address these three points, outlining concepts of assessing the overall druglikenss of molecules, the concentration of subsets of molecules in focused libraries, and the identification of specific leads through structure-based virtual screening techniques.

2.1 Druglikeness Screening

Many drug candidates fail in clinical trials because of reasons unrelated to potency against the intended drug target. Pharmacokinetics and toxicity issues are blamed for more than half of all failures in clinical trials. Therefore, the first part of virtual screening evaluates the *druglikeness* of small molecules, mostly independent of their intended drug target (there are specific drug classes such as those acting in the central nervous system that require specific drug profiles). Druglike molecules exhibit favorable absorption, distribution, metabolism, excretion, and toxicological (ADMET) parameters (20–24). They are synthetically feasible and possess pharmacophore features that offer the chance of specific interactions with the intended protein target. Druglikeness is currently assessed using the following types of methods: simple counting methods, functional group filters, topological filters, and pharmacophore filters. Computational techniques used to identify druglikeness include neural networks (25–27), recursive partitioning approaches (25, 28), and genetic algorithms (29). These methods are further discussed below.

Table 6.1 Typical Ranges for Parameters Related to Druglikeness[a]

Parameter	Minimum	Maximum
Log *P*	−2	5
Molecular weight	200	500
Hydrogen bond acceptors	0	10
Hydrogen bond donors	0	5
Molar refractivity	40	130
Rotatable bonds	0	8
Heavy atoms	20	70
Polar surface area [Å^2]	0	120
Net charge	−2	+2

[a]Data taken from ref. 21.

2.1.1 Counting Schemes. Database collections of known drugs [e.g., CMC (30), WDI (31), or MDDR (32)] are typically used to extract knowledge about structure and properties of potential drug molecules. Key physicochemical properties such as molecular weight, charge, and lipophilicity (33, 34) of drug collections are profiled to extract simple counting rules for relevant descriptors of ADMET-related parameters. Examples include Lipinski's "rule-of-five" (33), which limits the range for molecular weight (MW ≤ 500), computed octanol-water partition coefficient (Clog P ≤ 5), and hydrogen-bond donors and acceptors (OHs + NHs ≤ 5; Ns + Os ≤ 10). Other authors limit the number of rotatable bonds (RB ≤ 8) or rings in a molecule (number of rings ≤ 4) (34). Table 6.1 shows a list of typical boundaries of counting parameters. Figure 6.1 illustrates the profiling procedure for these counting parameters using polar surface area (PSA) (35) as a descriptor. Collections of 776 orally administered CNS drugs and 1590 orally administered non-CNS drugs that reached phase II efficacy studies were analyzed for their PSA. It was found that 90% of the non-CNS compounds have a PSA below 120 Å^2; 90% of CNS drugs have a PSA below 80 Å^2. Although it is possible that drugs have higher PSA values and are still orally bioavailable or penetrate the blood-brain barrier (as the result of active transport or other reasons), the profile suggests that it is much less likely. It is therefore a reasonable assumption in a virtual screening approach to discriminate against compounds outside the most populated descriptor space (in this case, PSA

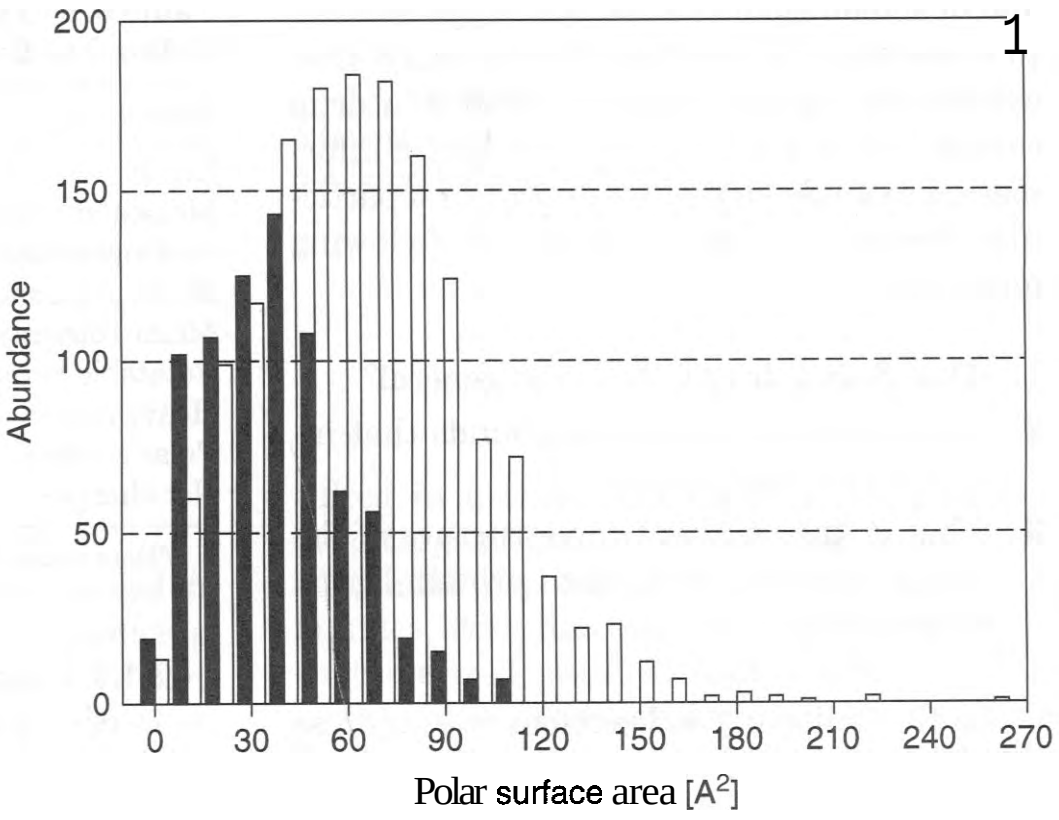


Figure 6.1. Distribution of polar surface area for 776 orally administered CNS drugs (black bars) and for 1590 orally administered non-CNS drugs (white bars) that have reached clinical phase II efficacy studies (35).

< 120 Å²), especially if the compound lies outside the optimal region for several descriptors (e.g., MW > 500 and Clog P > 5).

Simple descriptors as described above are quickly calculated and counted. Therefore, after typically removing compounds with atoms other than C, N, O, S, H, P, Si, Cl, Br, F, and I, counting schemes present the first filter in virtual screening approaches.

2.1.2 Functional Croup Filters. Reactive, toxic, or otherwise unsuitable compounds, such as natural product derivatives, are removed using specific substructure filters. Figure 6.2 shows a subset of substructures that lead to the dismissal of compounds in virtual screening. Typical reactive functional groups include, for example, reactive alkyl halides, peroxides, and carbazides. Unsuitable leads may include crown ethers, disulfides, and aliphatic methylene chains seven or more long. Unsuitable natural products may include quinones, polyenes, or cycloheximide derivatives. A list of such fragments coded in Daylight SMARTS is given, for example, by Hann and coworkers (36). It should be noted, however, that natural product derivatives are not always unsuitable leads.

Screening out compounds that contain certain atom groups associated with toxicity provides a practical and fast way to reduce large databases; however, it is only a crude approximation for eliminating potentially toxic compounds. Better descriptions of toxicity may be provided by structure-based methods to assess toxicity of compounds. They draw primarily from mutagenicity, carcinogenicity, and acute toxicity databases assembled, for instance, by the National Toxicology Program (37) and the Toxic Effect of Chemical Substances database, RTECS (38). CASETox (39), TOPKAT (40), and DEREK (41) are commercial software

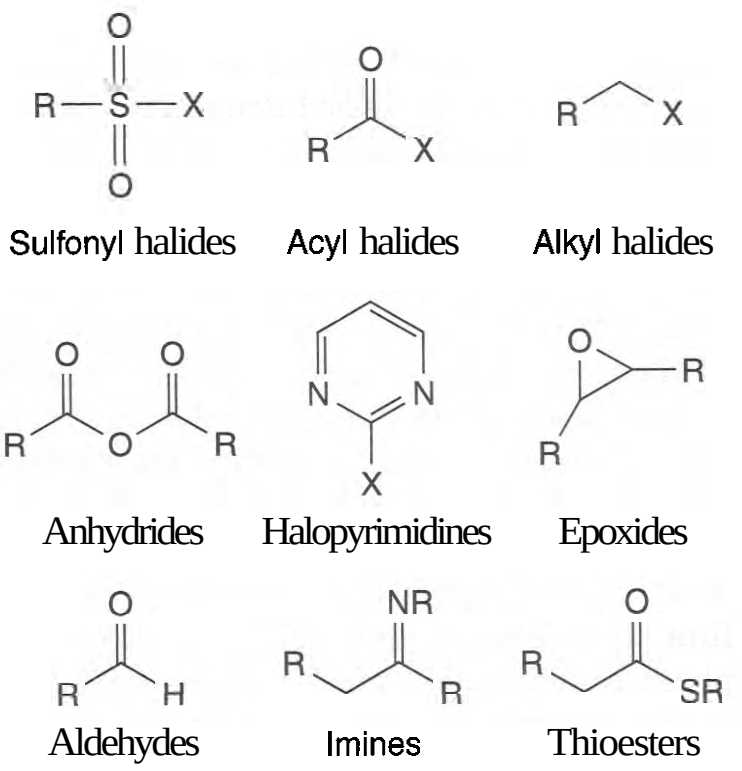


Figure 6.2. Selection of reactive functional groups that should be removed from a virtual screen (examples taken from Ref. 212).

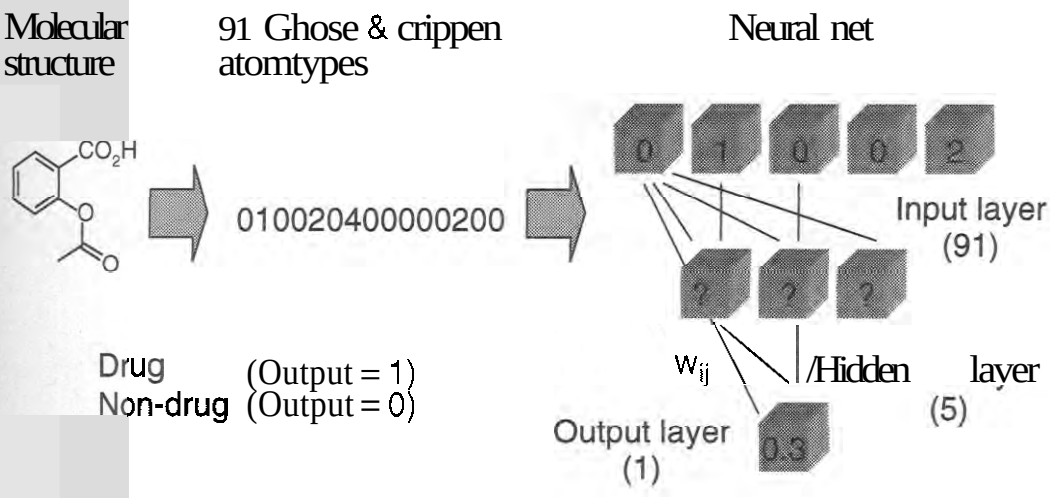


Figure 6.3. Neural network architecture for prediction of druglikeness.

products that can be used to evaluate virtual compounds for potential toxicity.

2.1.3 **Topological Drug Classification.** It is generally assumed that compounds with structural similarity to known drugs may exhibit druglike properties themselves, such as oral bioavailability, low toxicity, membrane permeability, and metabolic stability. Following this iidea, drug databases and reagent databases such as the ACD (42) as negative control (assuming they do not contain many drugs) have been analyzed to find structural features of drugs and nondrugs. Neural network approaches have been devised (25, 27) that can discriminate between drugs and nondrugs with about 80% certainty. Recursive partitioning approaches classify drugs and nondrugs with similar accuracy.

2.1.3.1 Artificial Neural Networks and Decision Trees. Figure 6.3 shows an example of a simple neural network that uses Ghose and Crippen atom types (43) to code the molecular structure. Ninety-one statistically significant atom types correspond to 91 input neurons of the neural net. Typically, five neurons in the hidden layers are used in the net design (25, 27). The single neuron output layer can vary between 0 (nondrugs) or 1 (drugs). Trained on 5000 drugs taken from the WDI and 5000 compounds labeled nondrugs taken from the ACD, the resulting neural net was shown to correctly classify about 80% of other drugs/nondrugs (27).

Recursive partitioning, also known as the decision tree approach, is another powerful method to extract knowledge from a database. Wagener and Geerestein have explored the WDI and ACD databases to train a decision tree for the discrimination of drugs and nondrugs (28). Figure 6.4 shows a partial decision tree derived by the authors. One rule derived from this partial tree is, for example, if a compound possesses no alcohol and a tertiary aliphatic amine but no methylene linker between a heteroatom and a carbon atom, it is not

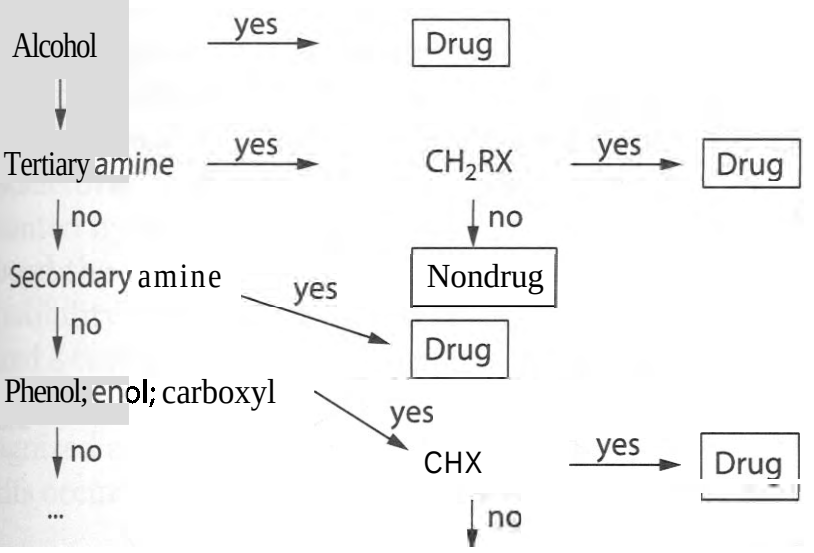


Figure 6.4. Partial decision tree from Wagener and Geerestein (28). C(n)sp^x describes a carbon with hybridization sp^x and formal oxidation number n. X refers to a heteroatom; R refers to any group linked through a carbon. The tree starts at the top left corner. Here is an example of how to read the tree: If a compound contains an alcohol, it is classified as a drug. If it does not contain an alcohol, the presence of a tertiary amine is checked. If it contains a tertiary amine and also contains (does not contain) a CH_2 group with attached heteroatom as well as another R group, it is classified as drug (nondrug).

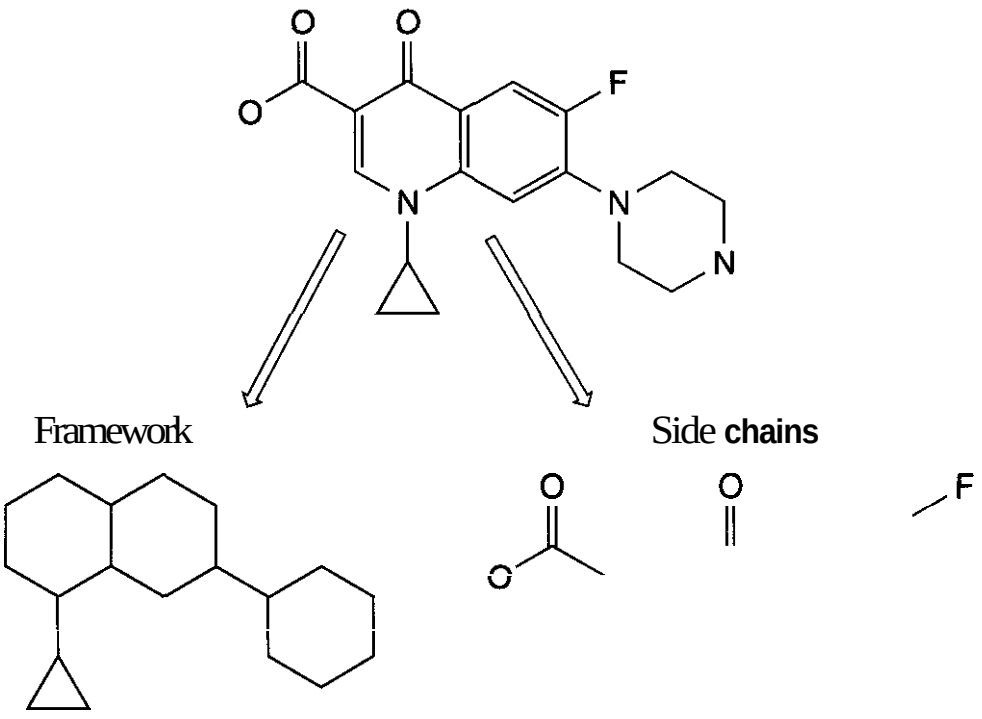


Figure 6.5. Dissection of a drug molecule into framework and side chains.

druglike. It is interesting to note that just by testing the presence or absence of hydroxyl, tertiary and secondary amines, carboxyl, phenol, or enol groups, 75% of all druglike structures in the MDDR and CMC can be recognized.

Once they are trained, neural networks and decision trees are very fast filter tools in virtual screening approaches. They are therefore applied early in the virtual screening filter cascade.

2.1.3.2 ***Structural Frameworks and Side Chains of Known Drugs.*** Databases have been mined to find structural motifs and pharmacophore features of small molecules that characterize drugs. Bemis and Murcko (44) dissected drug molecules from the Comprehensive Medicinal Chemistry (CMC) (30) database into side chains and frameworks (containing ring systems and linkers). They found that only 32 frameworks described the shapes of half the 5120 drugs in the CMC containing 1170 scaffolds. Figures 6.5 and 6.6 show the process of reducing a drug molecule to its framework and a list of the most frequently occurring frameworks in the CMC. Side chains most frequently occurring in drug molecules have also been analyzed (45). It has been found that of the 15,000 side chains cantained in the CMC, about 11,000 belong to one of only 20 side chains, including (starting with the most frequent): carbonyl, methyl, hy-

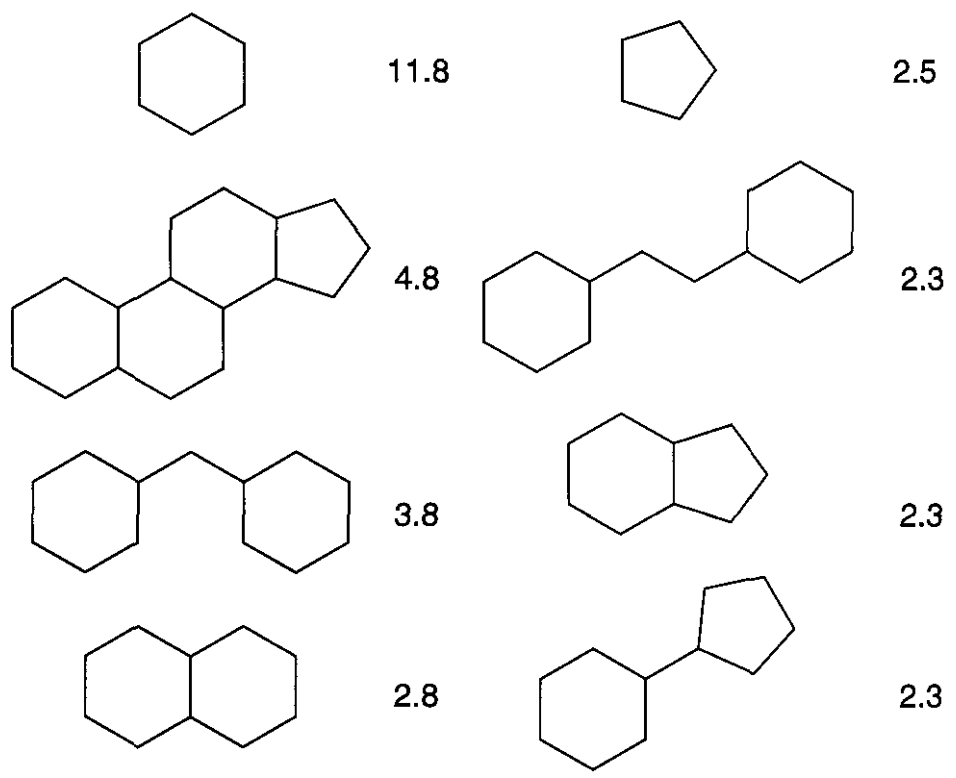


Figure 6.6. Most frequently occurring frameworks in drugs (numbers indicate percentages of occurrence in CMC database). Data are taken from Bemis and Murcko (44).

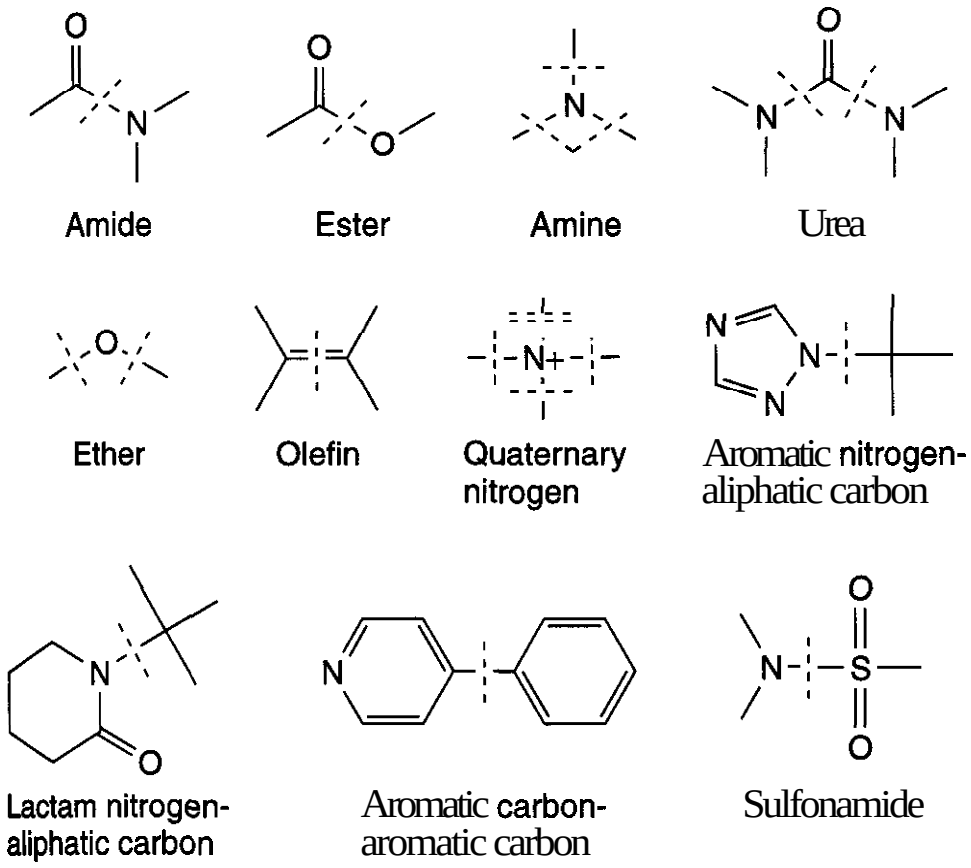


Figure 6.7. Retrosynthetic reactions in RECAP.

droxyl, methoxy, chloro, methylamine, primary amine, carboxylic acid, fluoro, and sulfone. Most molecules possess between one and five side-chains; more than 20% of the drugs stored in the CMC have two side chains per molecule.

For the analysis of virtual libraries according to the presence or absence of druglike frameworks, side-chains or structural motifs can be used for virtual screening. This idea has been extended in RECAP (retrosynthetic combinatorial analysis procedure), a technique that identifies common motifs in drugs based on fragmenting molecules around bonds formed by common reaction (46) (Fig. 6.7). Extrading rules from RECAP for virtual screening represents a possible way of addressing the questions of ease of synthesis of compounds. A similar approach to assess the occurrence of structural motifs in drug molecules was presented by Wang and Ramnarayan, who developed the concept of multilevel chemical compatibility (MLCC) between a drug database and a test molecule as a measure for druglikeness. In the MLCC method a compound is recognized as druglike if all of its topological motifs occur in other known drugs.

2.1.4 Pharmacophore Point Filter. The topological drug fragmentation approaches discussed above suggest that the occurrence of a relatively small number of frameworks (ring structures and linkers), an even smaller number of side chains, and a small number of polar groups characterize drugs very well. Although drugs and nondrugs are not completely distinguishable, it has been observed that drugs differ somewhat from nondrugs in their possession of hydrophobic moieties that are well functionalized. Non-, drugs often contain underfunctionalized hydrophobic groups (Fig. 6.8). Recent work to characterize the druglikeness of molecules focuses more on the presence of key functional groups in molecules.

A simple pharmacophore point filter has been introduced recently (47). It is based on the assumption that druglike molecules should contain at least two distinct pharmacophore groups (47). Four functional motifs have been identified that guarantee hydrogen-bonding capabilities that are essential for the specific interaction of a drug molecule with its biological target (Fig. 6.9). These motifs can be combined to functional groups that are also referred to here as *pharmacophore* points; they include: amine, amide, alcohol, ketone, sulfone, sulfonamide, carboxylic acid, carbamate, guanidine, amidine, urea, and ester. The following main rules apply to the pharmacophore point filter (PF1):

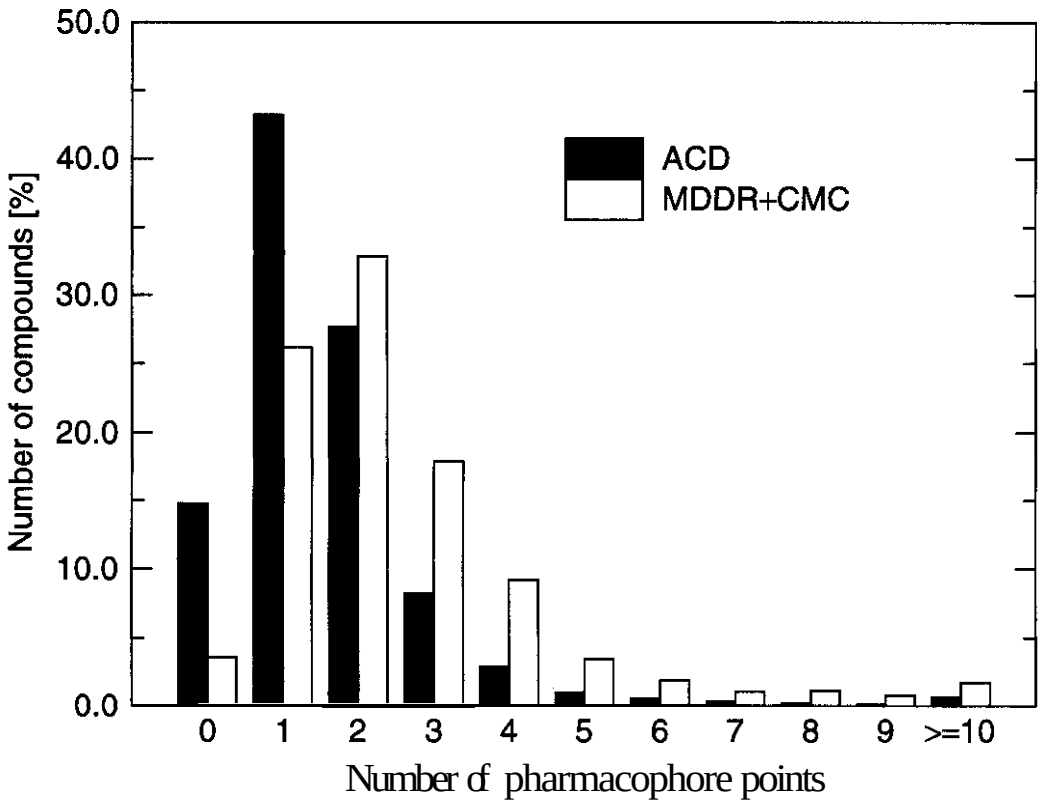


Figure 6.8. Number of **pharma**cophore points in drug databases (MDDR + CMC) and reagent **da**tabases (ACD).

- Pharmacophore points are fused and counted as one if they are separated by less than two carbon atoms.
- Molecules with less than two and more than seven pharmacophore points fail the filter.
- Amines are considered pharmacophore points but not **azoles** or diazines.
- Compounds with more than one carboxylic acid are dismissed.
- Compounds without a ring structure are dismissed.
- Intracyclic **amines** in the same ring are fused to one pharmacophore point.

The requirement of two distinct pharmacophore points neglects at least one very important class of drugs: biogenic amine-containing CNS drugs. Therefore, a second pharmacophore filter has been designed that requires only one pharmacophore point in small molecules of the type amine, amidine, guanidine, or carboxylic acid (PF2).

An analysis of drug databases and **reagent**-

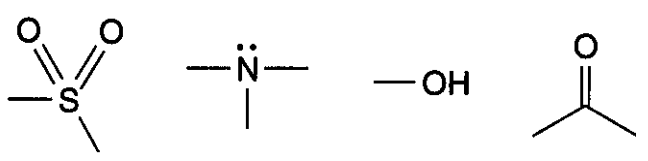


Figure 6.9. Functional motifs of drugs used to build pharmacophore points.

type databases reveal that about two thirds of drugs and nondrugs can be classified correctly by PF1. This performance is not as impressive as that of neural networks. However, as a filter for virtual screening, pharmacophore point filters offer some advantages. First, the occurrence and count of pharmacophore points can be evaluated on the building-block level of a virtual combinatorial library. No enumeration is necessary as for **druglike** neural nets. Second, the results of the pharmacophore point filter can be easily interpreted. Third, the settings of the filter can be easily adjusted (e.g., PF1 for non-CNS drugs, PF2 for CNS drugs).

2.2 Focused Screening Libraries for Lead Identification

Without the knowledge about specific drug targets it is sometimes useful to apply virtual screening for the design of focused libraries of a few thousand compounds rather than to find a small number of hits to be tested against a specific target. To save resources it may sometimes be more prudent not to run the entire HTS file against a target protein; instead, a focused library with higher chances of containing hits may be scrutinized. Those focused libraries may be designed to target specific protein families such as GPCRs, kinases, or nuclear hormone receptors. They can also be enriched with privileged structures that occur

more often in drug molecules and/or were found to inhibit members of the protein family.

2.2.1 Targeting Protein Families. Target class-directed libraries can be built from available compounds or be synthesized in combinatorial fashion. The design of target class-directed libraries relies on the identification of structural motifs in small molecules or in building blocks for combinatorial libraries that can be linked to increased activity for the target class. Functional groups that show the propensity to hit a certain target class can be found by examining ligands from the literature. Recurring motifs for GPCRs include, for example, piperazines, morpholines, and piperidines; for kinases they include, for example, heterobicyclic compounds or pyrimidines. Compounds bearing those structural motifs are thought to have a generally higher chance to be active against the respective target classes. A more rigorous approach to identify the "GPCR-likeness" of compounds or buildingblocks can be provided by a statistical analysis of druglike databases. Neural networks have been shown to be particularly useful in classifying chemical matter, such as CNS-active compounds (26, 48).

A neural network approach similar to that of Sadowski and Kubinyi (27) has been described recently to address the "GPCR-likeness" of small molecules as well as building blocks for combinatorial libraries (49). A feed-forward neural net was trained using 5000 compounds from the MDDR that target GPCRs and 5000 compounds that target other protein classes. Using the "activity-class" field of the database, about 20,000 GPCR-like and 55,000 non-GPCR-like have been identified by entries such as 5HT, leukotriene, and PAF. The resulting neural net classifies GPCR-like compounds correctly with 80% certainty. An independent test of compounds in our proprietary database that were found to hit GPCRs or other targets showed a correct prediction of GPCR-like compounds in 70% of the cases. When several virtual combinatorial libraries were analyzed, it turned out that the property of being GPCR-like could be attributed to the GPCR-likeness of the building blocks alone; that is, the GPCR-likeness of the enumerated compound correlated very well with the GPCR-likeness of the most GPCR-like building block it contained. This offers an important advantage for the design of combinatorial libraries because, for large virtual libraries, the computer costs for enumeration go with the power of the number of R groups and thus very quickly becomes impractical. For instance, for a 3-R-group library with 1000 building blocks each, the enumerated library would contain 1 billion compounds to be analyzed, whereas the building block-level analysis needs to examine only 3000 compounds. Figure 6.10 shows a list of amine building blocks extracted from the ACD that were found to be most GPCR-like by the neural net.

Not every portion of a GPCR-like molecule has to be GPCR-like. The presence of one GPCR-like moiety (building block or core structure) is sufficient to make a compound GPCR-like. Therefore, the neural network offers two different strategies for the design of GPCR-like libraries: (1) GPCR-like core + druglike building blocks (need not be GPCR-like); (2) non-GPCR-like core + GPCR-like building blocks. Virtual screening of a database of existing compounds using the described neural net can be applied to assemble a focused screening library. Alternatively, combinatorial libraries can be designed.

2.2.2 Privileged Structures. Privileged structures are structural types of small molecules that are able to bind with high affinity to multiple classes of receptors (50). An enrichment of libraries with privileged structures may increase the chance of finding active compounds. Examples of privileged structures include benzazepine analogs found to be effective ligands for an enzyme that cleaves the peptide angiotensin I, whereas others are effective CCK-A receptor ligands. Cyproheptadine derivatives were found to have peripheral anticholinergic, antiserotonin, antihistaminic, and orexigenic activity. Hydroxamate and benzamidine derivatives have been shown to be privileged structures for metalloproteases and serine proteases, respectively. For the class of 7-transmembrane G-protein-coupled receptors a large number of privileged structures has been found including, for example, diphenylmethane, diazepine, benzazepine, biphenyltetrazole, spiropiperidine, in-

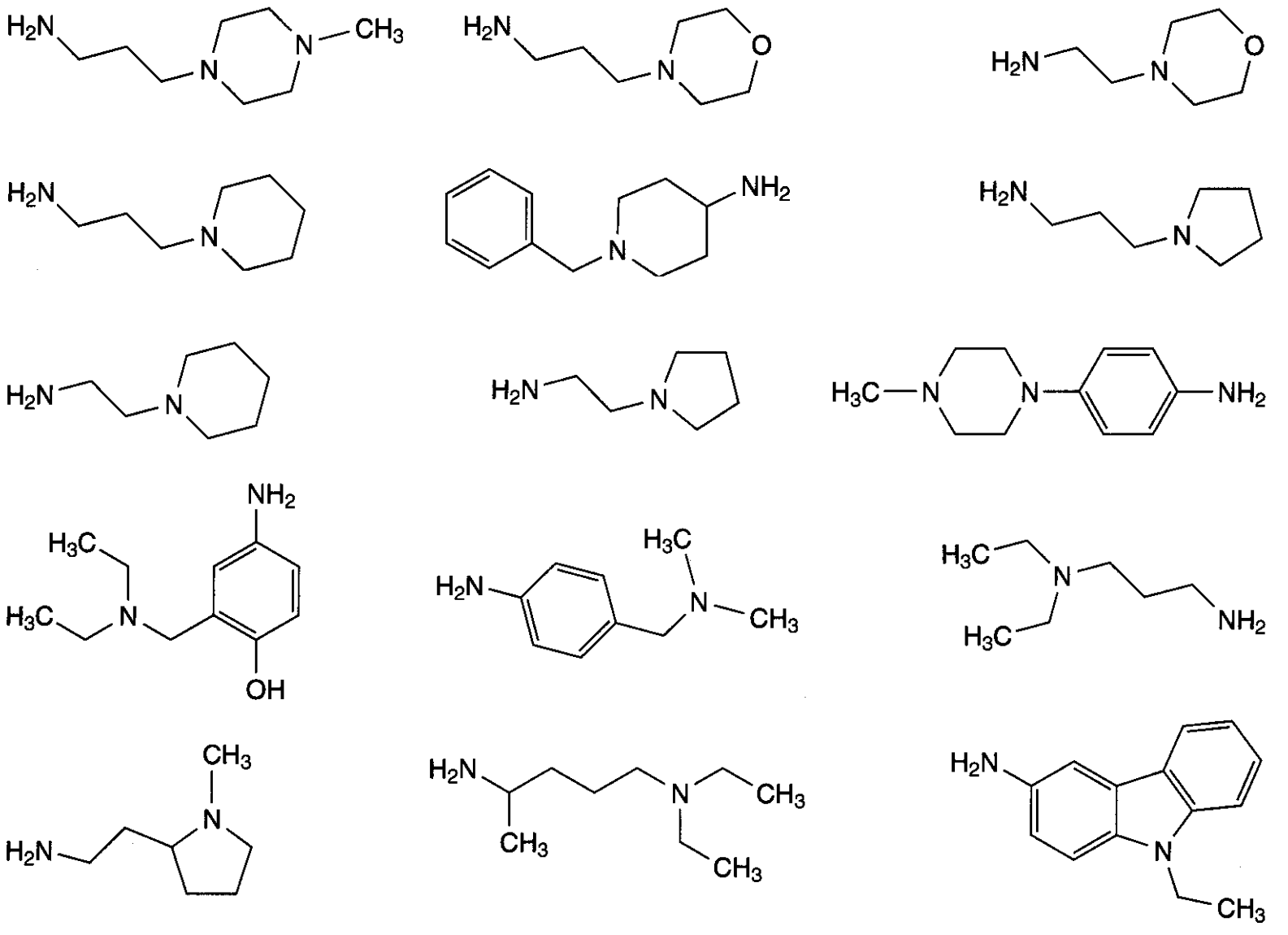


Figure 6.10. Selection of GPCR-like amines from the ACD.

dole, and benzylpiperidine (51). Some ubiquitously privileged structures have recently been identified (52). They include carboxylic acids, biphenyls, diphenylmethane, and, to a lesser extent, naphthyl, phenyl, cyclohexyl, dibenzyl, benzimidazole, and quinoline.

2.3 Pharmacophore Screening

In cases where no structural information about the target protein is given, pharmacophore models can provide powerful filter tools for virtual screening (53). Even in cases where the protein structure is available, pharmacophore filters should be applied early because they are generally much faster than docking approaches (discussed below) and can, therefore, greatly reduce the number of compounds subjected to the more expensive docking applications. For example, a pharmacophore model consisting of three pharmacophore points can be tested against about 10^6 compounds in a few minutes of computer time [disregarding the time it takes to generate three-dimensional (3D) conformations of each molecule] (10). Another interesting aspect of pharmacophores in virtual screening is 3D-pharmacophore diversity. Although the diversity concept for virtual compounds in general is not applicable because of the enormity of the chemical space, diversity in pharmacophore space is a feasible concept. Virtual libraries can therefore be optimized for covering a wide pharmacophore space.

2.3.1 Introduction to Pharmacophores. In 1894 Emil Fischer proposed the "lock-and-key" hypothesis to characterize the binding of compounds to proteins (54). This can be considered the first attempt to explain binding of small molecules to a biological target. Proteins recognize substrates through specific interactions. It is a challenge for the medicinal chemist to synthesize compounds that can capture the 3D arrangement of functional groups in a small molecule that forms the pharmacophore and that is responsible for substrate binding

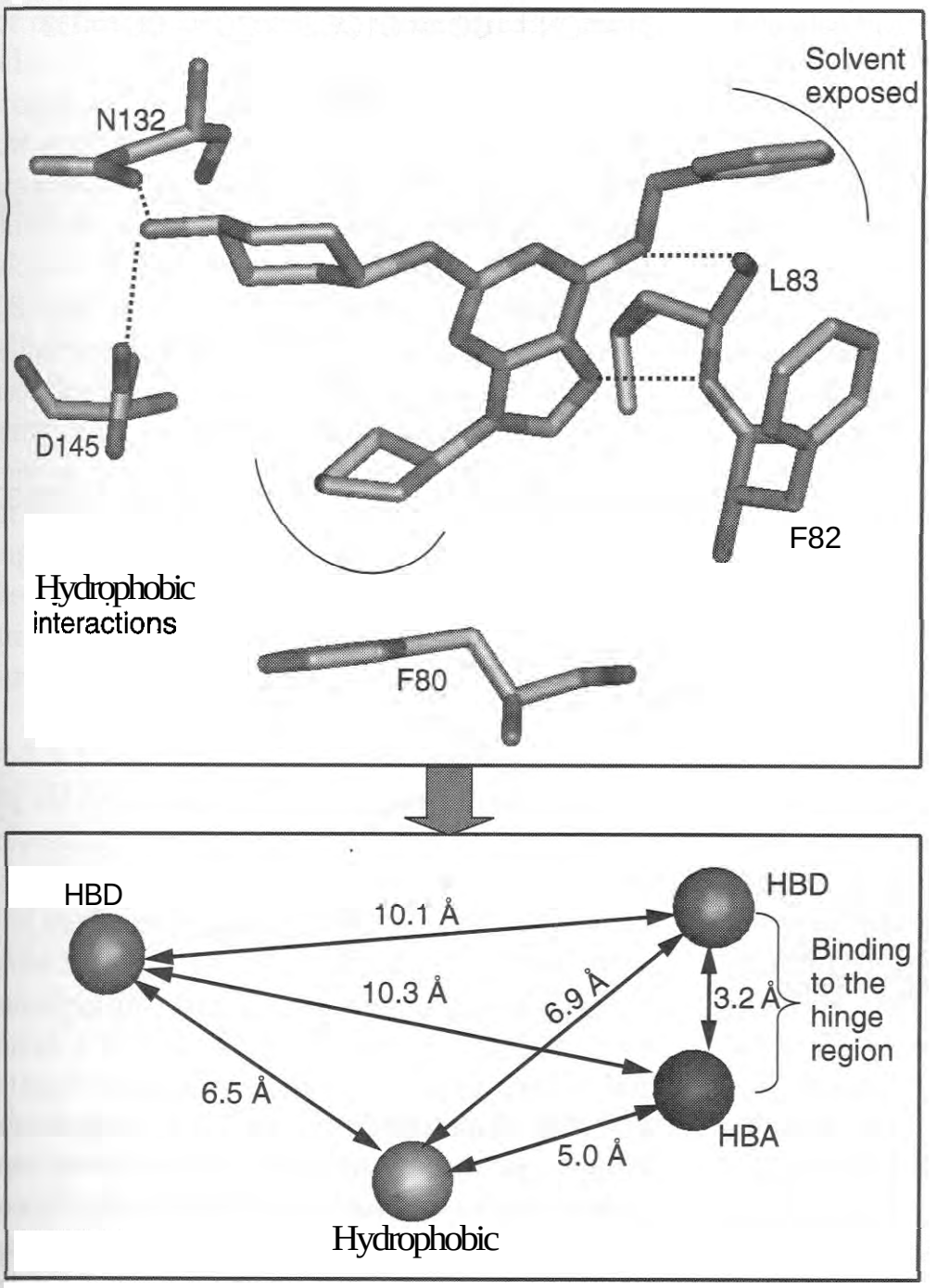


Figure 6.11. Pharmacophore derived based on the interactions between human cyclin-dependent kinase 2 and the adenine-derived inhibitor H717 as observed in the X-ray structure of the complex (PDB entry 1G5S). Dashed lines highlight hydrogen-bonding interactions. HBD, hydrogen-bond donor; HBA, hydrogen-bond acceptor. The hinge region is linking the *N*- and C-terminal domains of a kinase.

to the protein. The first definition of the pharmacophore formulated by Paul Ehrlich was "a molecular framework that carries (phoros) the essential features responsible for a drug's (pharmacon) biological activity" (55). This definition was slightly modified by Peter Gund to "a set of structural features in a molecule that is recognized at a receptor site and is responsible for that molecule's biological activity" (56). An example is shown in Fig. 6.11. An X-ray structure of CDK2 complexed with the adenine-derived inhibitor H717 (57–59) has been solved. Interactions that are essential to substrate and inhibitor binding to the enzyme will form the pharmacophore that should be captured by inhibitors binding the same way H717 does. As shown in Fig. 6.11, the inhibitor binds to the hinge region (Phe82 and Leu83) through two hydrogen bonds, to a hydrophobic region through the cyclopentyl group, and to Asp145 and Asn132 through hydrogen bonds. The pharmacophore that reflects these interactions has a hydrogen-bond donor and a hydrogen-bond acceptor pair that ensures binding to the hinge region, a hydrophobic group that corresponds to the cyclopentyl binding site, and a hydrogen-bond donor that ensures binding to Asp145 and/or Asn132. Note that in addition to distances that describe the 3D relationship among pharmacophore points, angles, dihedrals, and exclusion volumes are also used. Each additional restraint can reduce the number of hits, thus making the compound selection easier for testing. Pharmacophore hypotheses for searching can be generated using structural information from active inhibitors, ligands, or from the protein active site itself (60, 61).

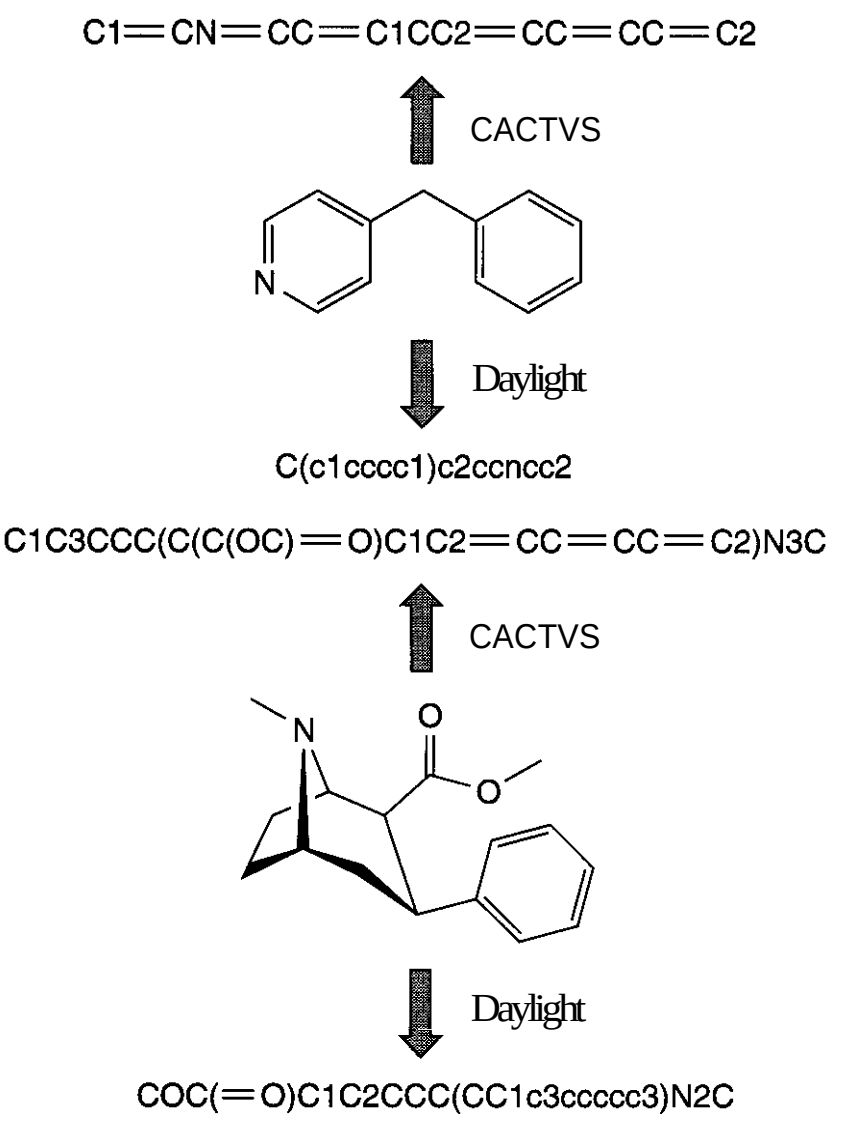


Figure 6.12. Examples of SMILES notations for two compounds obtained using CACTVS and Daylight.

2.3.2 Databases of Organic Compounds. Virtual screening is used in general for selecting potentially active compounds from databases of compounds available either in-house or from a vendor. Because virtual screening is not accurate enough to identify only active compounds as hits, it is less risky to screen databases with existing compounds rather than synthesize a new library. Nevertheless, virtual libraries that can be synthesized through combinatorial chemistry and/or rapid analoging can easily be generated using *in silico* methods. These libraries are more often generated for lead optimization and synthesis prioritization (62, 63).

There is a wealth of databases that code available compounds typically in the two-dimensional standard data (2D-SD) format including connectivity from MACCS (32, 64). The most common databases are the Available Chemicals Directory (ACD) (42), Spresi (65), Chemical Abstracts Database (66), and the National Cancer Institute Database (67, 68). Many vendors of chemicals also provide searchable databases with 2D-structure and property information of their compounds. Sometimes compounds are coded in linear representations such as the SMILES (69, 70) notation. The SMILES codes obtained using CACTVS and Daylight programs for 4-benzyl pyridine and R-cocaine are shown in Fig. 6.12.

The primary source of 3D experimental structures of organic molecules is the Cambridge Structural Database (71). Alternatively, 2D databases of organic compounds can be converted into 3D databases using several software programs (72). Each program starts with generating a crude structure that is subsequently optimized using a force field. CONCORD (73) applies rules derived from experimental structures and a univariate strain function for building an initial structure. CORINA (74) generates an initial structure by use of a standard set of bond lengths, angles and dihedrals, and rules for cyclic systems. RUBICON (75) invokes distance geometry techniques to generate 3D structures based on connectivity tables. This program also uses bond lengths and angle tables to build a matrix containing the upper and lower bounds

for distances between all atoms in the molecule. OMEGA (76) uses a torsion-driven approach for building conformers. It generates low energy conformers for each molecule by assembling it from fragments and searching through possible orientations of the subunit added. WIZARD (77) and COBRA (78), AIMB (79) and MIMUMBA (80) employ artificial intelligence techniques for generating a set of user-specified low energy conformations for a compound. MOLGEO (81) uses a depth-first approach for generating 3D structures based on connectivity using bond length and bond angle tables. IDEALIZE (82) is a molecular mechanics program that minimizes 2D structures to generate the corresponding 3D structure.

2.3.3 2D Pharmacophore Searching. Searching 2D databases is of great importance for accelerating drug discovery. Chemical suppliers provide databases of purchasable compounds that medicinal chemists search for starting material for synthesis or analogs of a lead compound. Different strategies are pursued to search a 2D database to identify compounds of interest. Exact structure search is applied to find out whether a compound is present in the database. Substructure searches identify larger molecules that contain the user-defined query, irrespective of the environment in which the query substructure occurs (83) (Fig. 6.13). Furthermore, substructure searching *can* identify all compounds in a database that share the same core structure. Biochemical data obtained from testing these compounds can be used for generating structure-activity relationships (SARs), even before synthetic plans are made for lead optimization (84). In contrast, superstructure searches are used to find smaller molecules that are embedded in the query (Fig. 6.14). One problem that arises from substructure searches is that the number of compounds identified *can* reach into the thousands. A solution to this problem is ranking the compounds based on similarity to a reference compound. Similarity searches use one or more structural descriptors for quantifying the similarity between compounds in the database and in the query (85, 86) (Fig. 6.15). A review of descriptors used in similarity searches is provided by Willett et al. (86). Beyond structural similarity, activity has also been the subject of several studies. Xue et al. showed that compounds with similar activity could be identified using mini-fingerprints (87–89), physicochemical property descriptors (90), or latent semantic structure indexing (91, 92). In addition, similarity searches *can* be combined with superstructure searches for limiting the number of compounds selected. Flexible match searches are used for identifying compounds that differ from the query structure in user-specified ways. In addition, isomer, tautomer, and parent molecule searches may be done to find in a database isomers, tautomers, or parent molecules of the query.

2.3.4 3D Pharmacophores

2.3.4.1 Ligand-Based Pharmacophore Generation. Ligand-based pharmacophores are typically used when the crystallographic, solution structure, or modeled structure of a protein cannot be obtained. When a set of active compounds is known and it is hypothesized that all compounds bind in a similar way to the protein, then common groups should interact with the same protein residues. Thus, a pharmacophore capturing these common features should be able to identify from a database novel compounds that bind to the same site of the protein as the known compounds do. The process of deriving a pharmacophore, called ***pharmacophore mapping,*** consists of three steps: (1) identifying common binding elements that are responsible for biological activity; (2) generating potential conformations that active compounds may adopt; and (3) determining the 3D relationship between pharmacophore elements in each conformation generated. To build a pharmacophore based on a set of active compounds, two methods are usually applied. One method is to generate a set of minimum energy conformations for each ligand and search for common structural features. Another method is to consider all possible conformations of each ligand to evaluate shared orientations of common functional groups. Analyzing many low energy conformers of active compounds can suggest a range of the distance between key groups that will take in account the flexibility of the ligands and of the protein. This task can be performed either manually or automatically.

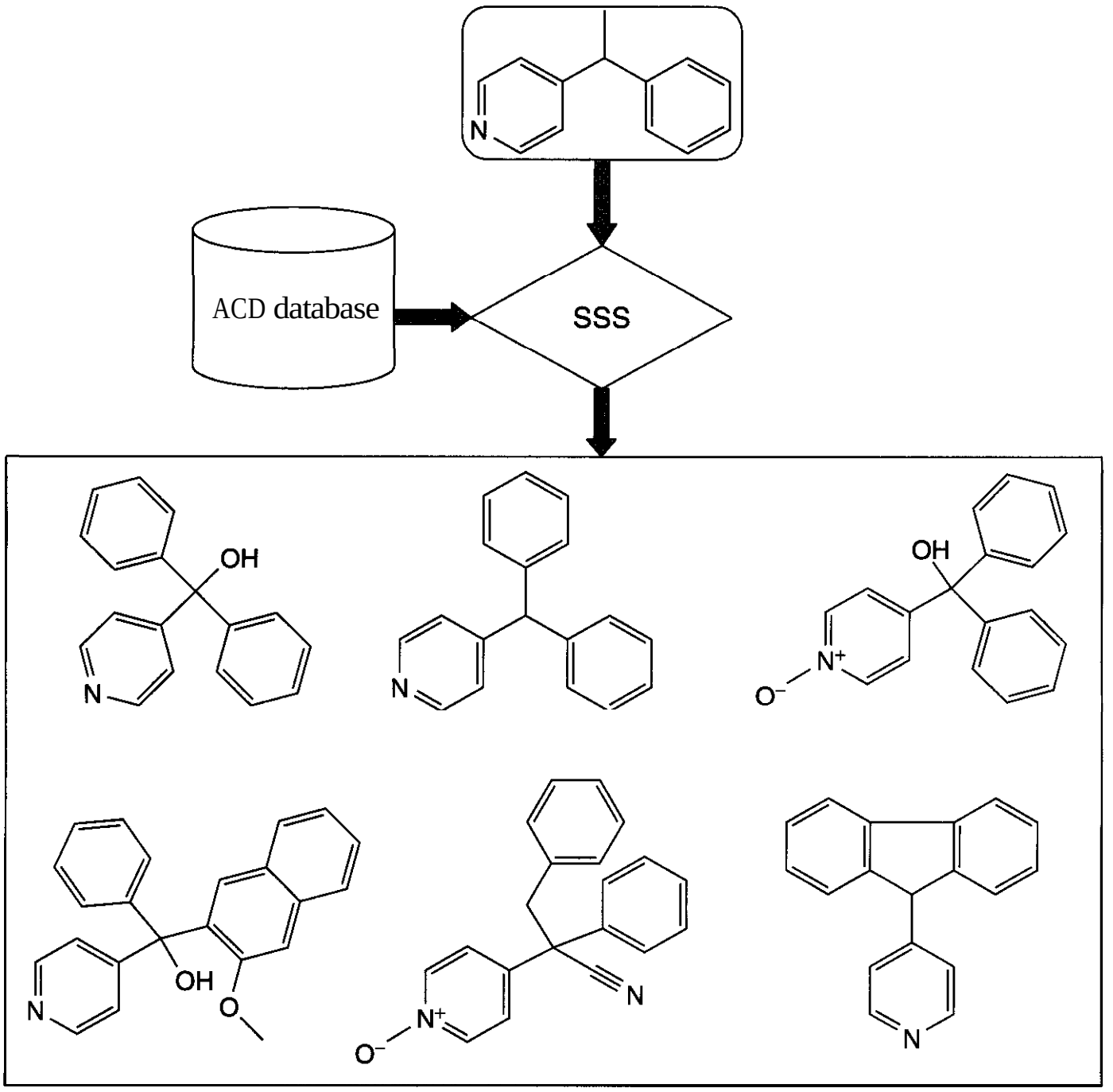


Figure 6.13. Compounds identified from the ACD database through substructure search.

2.3.4.2 Manual Pharmacophore Generation. Manual pharmacophore generation is used when there is an easy way to identify the common features in a set of active compounds and/or there is experimental evidence that some functional groups should be present in the ligand for good activity. An example is the development of a pharmacophore model for dopamine-transporter (DAT) inhibitors (Fig. 6.16). In the first step common structural features were identified in the selected five DAT inhibitors (93–95) (Fig. 6.16, circles). Four out of five compounds were structurally rigid, whereas the 4-hydroxy piperidinol was flexible. A systematic conformational search for 4-hydroxy piperidinol identified 10 possible conformations. Measuring distances among pharmacophore elements in every inhibitor and every conformation considered led to the distance ranges among pharmacophore points shown in Fig. 6.16. Because proteins are flexible, pharmacophores should also have some flexibility built in, thus justifying the use of distance ranges.

2.3.4.3 Automatic Pharmacophore Generation. Pharmacophore generation through conformational analysis and manual alignment is a very time-consuming task, especially when the list of active ligands is large and the elements of the pharmacophore model are not obvious. There are several programs, HipHop (96), HypoGen (97), Disco (98), Gasp (99), Flo (100), APEX (101), and ROCS (102), that can automatically generate potential pharmacophores from a list of known inhibitors. The performance of these programs in automated

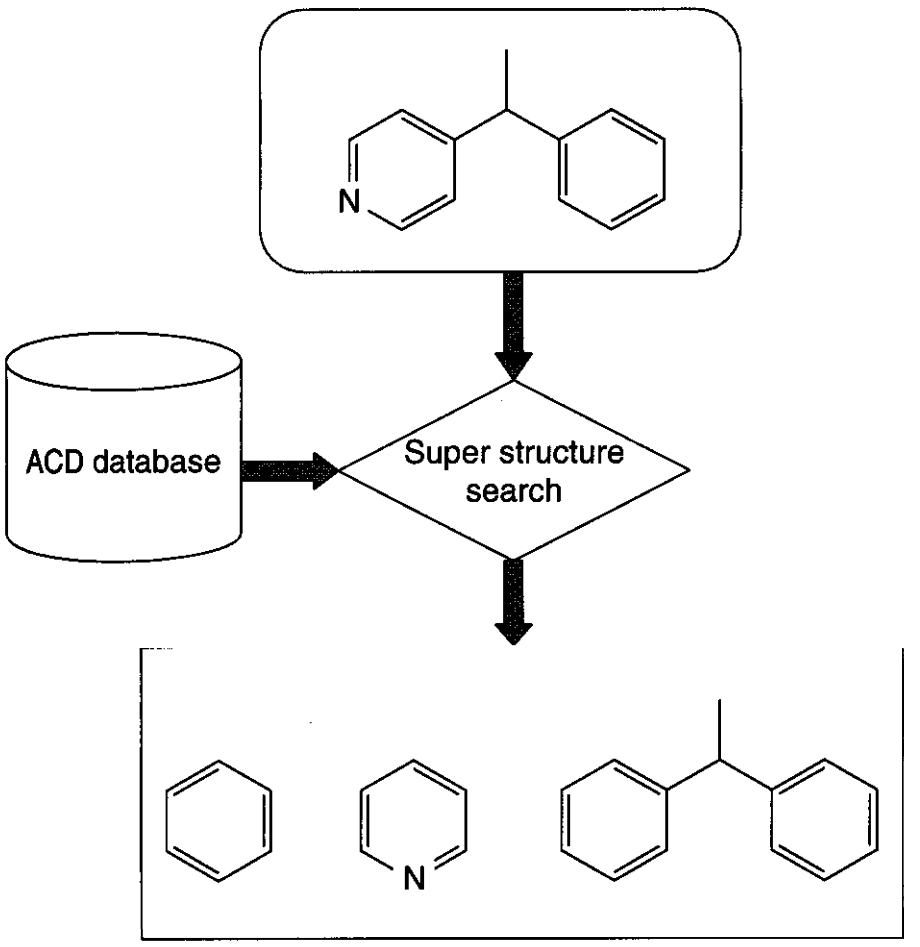


Figure 6.14. Compounds identified from the ACD database through superstructure search.

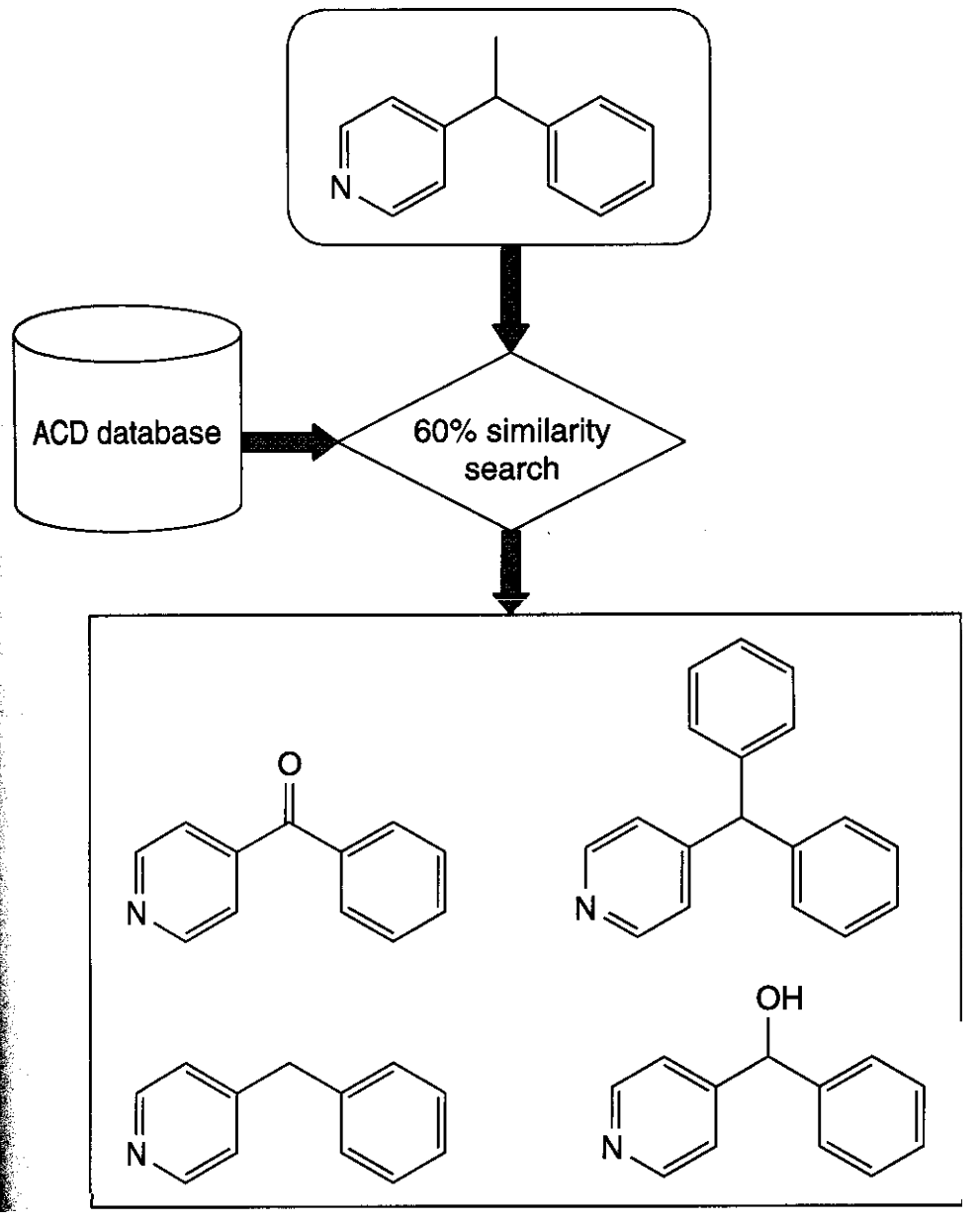
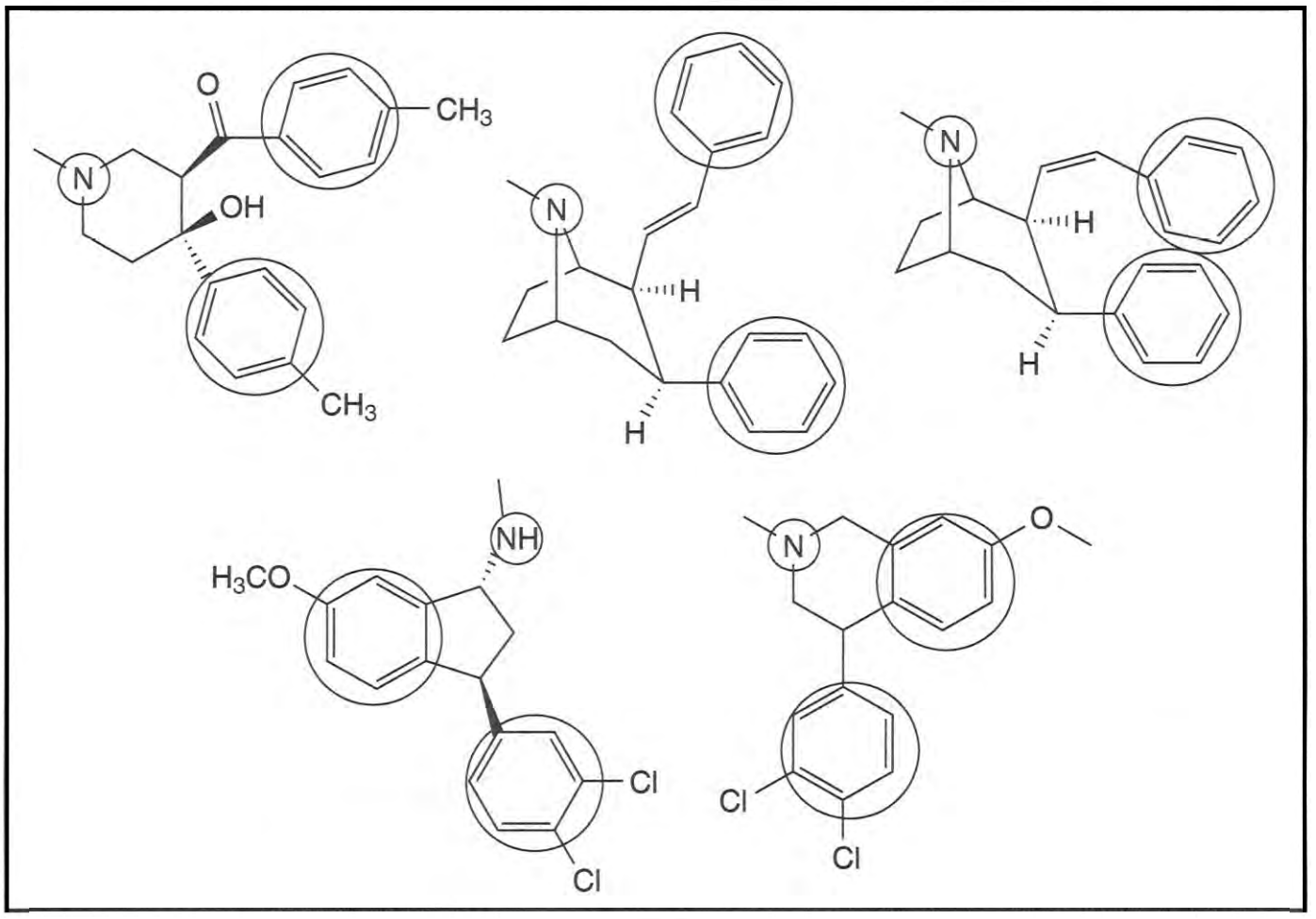


Figure 6.15. Compounds identified from the ACD database through similarity search using 60% similarity as threshold. The lower the specified similarity the higher the number of hits identified.



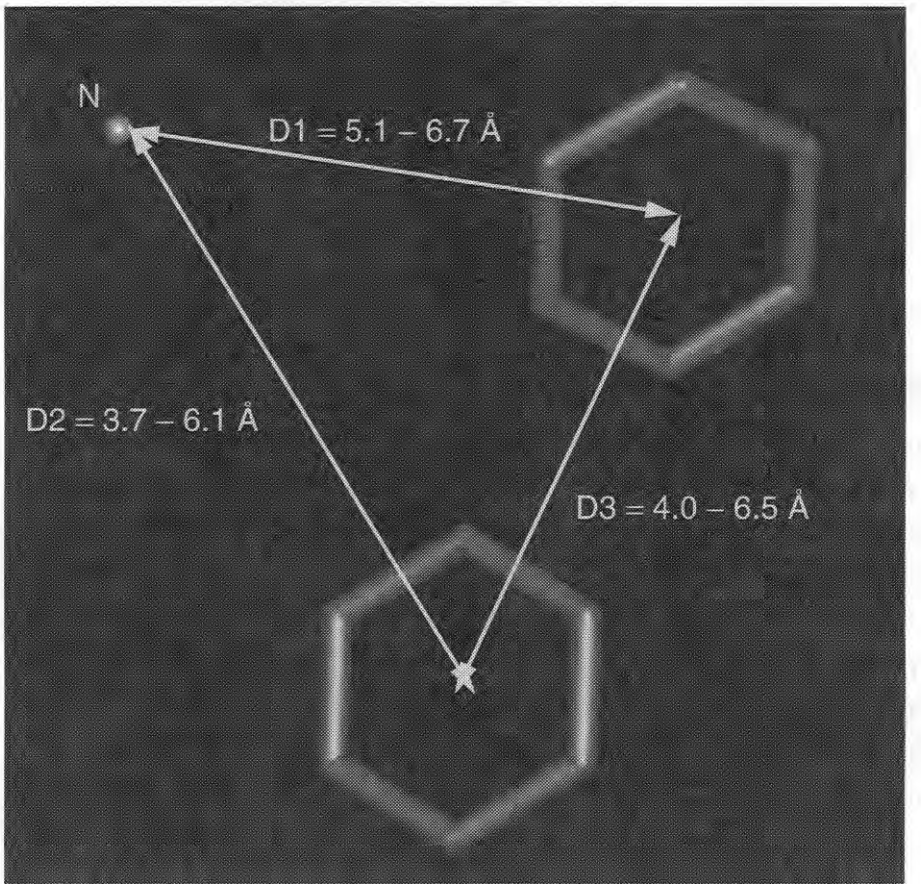


Figure 6.16. Manual pharmacophore mapping by measuring distances between pharmacophore points in every compound and conformation considered. Pharmacophore elements are highlighted with circles. All structures were built and minimized using QUANTA. Conformers of 4-hydroxy piperidinol were generated using the Grid Scan method from QUANTA, followed by clustering, to identify unique conformers.

pharmacophore generation varies depending on the training set. The use of these programs for pharmacophore generations was recently reviewed in detail (103). Here we focus on common features of these programs. All programs use algorithms that identify common pharmacophore features in the training set molecules; they use scoring functions to rank the identified pharmacophores. The following features are identified in each molecule: hydrogen-bond donors, hydrogen-bond acceptors, negative and positive charge centers, and surface accessible hydrophobic regions that can be aliphatic, aromatic, or nonspecific. Most of the programs consider ligand flexibility when generating pharmacophores because compounds might not bind to the protein in the minimum energy conformation.

2.3.4.4 Receptor-Based pharmacophore Generation. If the 3D structure of a receptor is known, a pharmacophore model can be derived based on the receptor active site. Biochemical data can be used for identifying key residues that are important for substrate and/or inhibitor binding. This information can be used for building pharmacophores targeting the region defined by key residues or for choosing among pharmacophores generated by an automated program. This can greatly improve the chance of finding small molecules that inhibit the protein because the search is focused on a region of the binding site that is crucial for binding substrates and inhibitors. Many ligands bind to proteins through nonbonded interactions such as hydrogen bonds and hydrophobic interactions. Programs such as LUDI (104–106) or POCKET (107) can use the structure of the protein to generate interaction sites or grids to characterize favorable positions that ligand atoms should occupy. Four types of interaction sites are characterized: hydrogen-bond donors, hydrogen-bond acceptors, and hydrophobic groups that can be lipophilic-aliphatic or lipophilic-aromatic. LUDI-generated interaction maps for Cerius^2 Strudure-Based Focusing (108) do not differentiate between aliphatic and aromatic interaction sites. This is based on the observation by Burley and Petsko (109) that, besides aromatic side chains, aliphatic and aromatic side chains also pack closely to form the hydrophobic core of proteins. Because proteins are not rigid, Carlson et al. (110) proposed using molecular dynamics simulation for generating a set of diverse protein conformations to include protein flexibility in the pharmacophore development. In this case distance ranges between pharmacophores are obtained by examining several conformations of the protein. This technique is similar to the one used for the generation of flexible pharmacophores (Fig. 6.16), based on active compounds, when several conformations of the compound and/or many compounds are considered for pharmacophore mapping.

2.3.5 Pharmacophore-Based Virtual Screening. Pharmacophore-based virtual screening is the process of matching atoms and/or functional groups and the geometric relations between them to the pharmacophore in the query. Examples of programs that perform pharmacophore-based searches are 3Dsearch (111), Aladdin (53), UNITY (112), MACCS-3D (113), Catalyst (114), and ROCS (102). There are also web-based applications (115,116) that can perform pharmacophore searches. Usually pharmacophore-based searches are done in two steps. First, the software checks whether the compound has the atom types and/or functional groups required by the pharmacophore; then it checks whether the spatial arrangement of these elements matches the query. The fastest approach used in the matching step is considering rigid compounds. Because molecules that are not rigid might have a conformation that matches the pharmacophore, flexibility of the ligands should be considered. Flexible 3D searches identify a higher number of hits than rigid searches do (117). However, flexible searches are more time consuming than rigid ones. There are two main approaches for including conformational flexibility into the search: one is to generate a user-defined number of representative conformations for each molecule when the database is created; the other is to generate conformations during the search. By use of the first approach, any rigid search program can be used for doing a flexible search; however, generating the database takes more time and disk space. The second approach gives more flexibility to the user, given that a larger number of conformations can be generated for each

molecule during the search. In this case the database search requires more computer resources; however, this approach will not miss conformations that fit the query but were not stored in the database. Pharmacophore queries that define distance ranges between pharmacophore elements compensate for possible conformational changes in the receptor site upon ligand binding. Also, these flexible pharmacophore queries compensate for the difference between using multiconformer databases and generating conformers during the search.

ROCS is using a shape-based superposition for identifying compounds that have similar shape. Grant and Pickup (118) showed that using atomic-centered Gaussians instead of a spherical function can dramatically reduce the time required for a shape alignment of two molecules. This improved routine allows the program to perform shape-based database searches at an acceptable speed (300–400 conformers/s).

There are several methods for generating conformers during *in silico* screening. Torsion optimization (119) is used for minimizing the root-mean-square (rms) deviation between the constraints from the pharmacophore and the corresponding distances in the compound. The "directed tweak" (120) algorithm also uses torsion optimization for minimizing the sum of the squared deviations between distances in the pharmacophore and the corresponding ones in the compound. ChemDBS-3D (121) generates low energy conformations that can match the pharmacophore using rules similar to those in WIZARD (77). The distance geometry algorithm (122) uses bond length and bond angle information for building a matrix containing upper and lower limits of distances between atoms in the organic compound. These distances can be used for building the conformation that fits the pharmacophore query. The systematic search method (123) is feasible for molecules with few rotatable bonds and thus has limited applicability.

2.4 Structure-Based Virtual Screening

In direct analogy to high throughput screening, docking and scoring techniques can be applied to computationally screen a database of hundreds of thousands of compounds against a specific target protein. Computational methods that predict the 3D structure of a protein-ligand complex are often referred to as *molecular docking* approaches (Fig. 6.17) (124). Protein structures can be employed to dock ligands into the binding site of the protein and to study their interactions (125). For virtual screening, the crucial task at hand is the fast and reliable ranking of a database of putative protein-ligand complexes according to their binding affinities. Depending on ligand and protein flexibility, sampling depth, and optimizing schemes, docking programs used today (Table 6.2) can facilitate this task within a few minutes or sometimes seconds per processor and molecule. Virtual screening as a computation task can be trivially run using parallel computing because the protein-ligand docking events are completely independent of each other. Although docking has initially been developed as a specialist modeling tool run on computer workstations, nowadays inexpensive Linux clusters or distributed computing over networked PCs can be used for virtual screening. This increases the *in silico* throughput into the realm of 100,000 compounds per day on a Linux cluster, thereby reaching the speed of today's high throughput screens. Energy functions that evaluate the

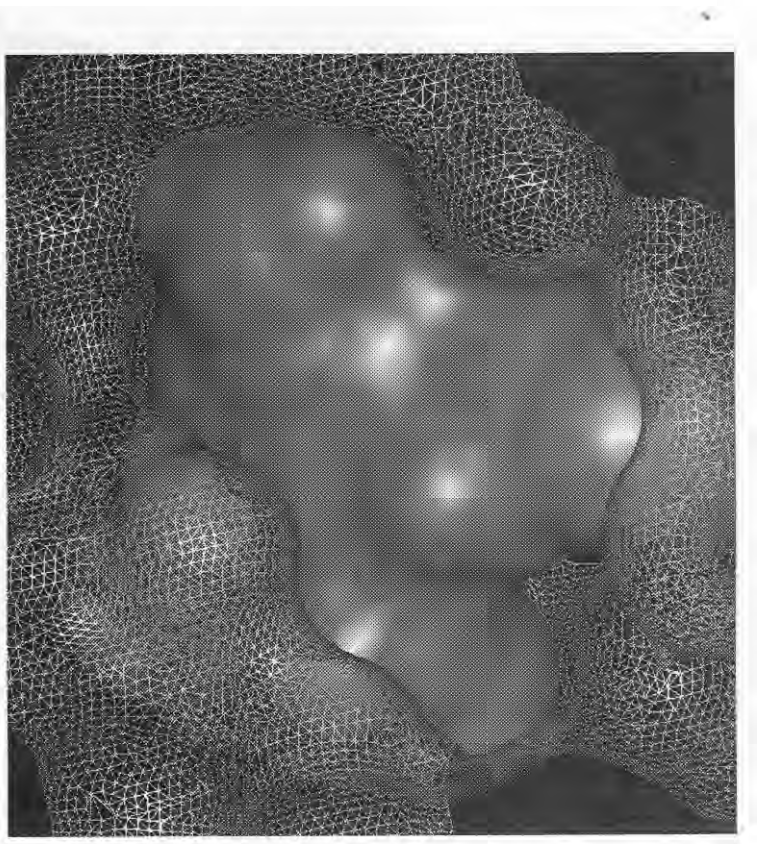


Figure 6.17. Crystal structure (PDB entry 1a4q) of the neuraminidase inhibitor zanamivir bound in the active site (213).

Table 6.2 Selection of Available Protein–Ligand Docking Software for Structure-Based Virtual Screening

Docking Program	Docking/Sampling Method	Scoring Method
GLIDE (www.schrodinger.com)	Rigid protein; multiple conformation rigid docking; grid-based energy evaluation	Empirical scoring, including penalty term for unformed hydrogen bonds; force-field scoring
DOCK (www.cmpharm.ucsf.edu/kuntz/dock.html)	Rigid protein; flexible ligand docking (incremental construction)	Force-field scoring; chemical scoring, contact scoring
FlexX (cartan.gmd.de/FlexX)	Rigid protein; flexible ligand docking (incremental construction)	Empirical scoring intertwined with sampling
DockVision (www.dockvision.com)	Monte Carlo, genetic algorithm	Various force fields
DockIT (www.daylight.com/meetings/emug00/ Dixon)	Ligand conformations generated inside binding-site spheres using distance geometry	PLP, PMF
FRED (www.eyesopen.com/fred.html)	Exhaustive sampling; rigid protein, multiple conformation rigid docking	Chemscore, PLP, ScreenScore, and Gaussian shape scoring
LigandFit (www.accelrys.com)	Monte Carlo	LIGSCORE, PLP, PMF, LUDI
Gold (www.ccdc.cam.ac.uk/prods/gold/)	Genetic Algorithm	Soft core vdW potential and hydrogen bond potentials

binding free energy between protein and ligand sometimes employ rather heuristic terms. Therefore, those functions are more broadly referred to as ***scoring functions.***

2.4.1 Protein Structures. A 3D-protein structure of the receptor at atomic resolution is necessary to start a protein-ligand docking experiment. The exponential growth of solved crystal and solution structures in recent years provides a reliable source of protein structures. The protein database (PDB) currently holds more than 18,000 protein structures. It should be noted, however, that the chances of a successful virtual screen very much depend on the quality of the available structure. The crystal structure should be well refined; typically a resolution of at least 2.5 Å is considered to be necessary (126). Small changes in structure can drastically alter the outcome of a computational docking experiment (127). Moreover, many receptor sites are flexible; they often undergo conformational changes upon ligand binding. A good example is the Tyr248 movement of carboxypeptidases upon substrate or ligand binding, which has provided the first structural perspective of Koshland's induced-fit hypothesis (128, 129). Proteins have to be studied carefully in every individual case to decide how promising a virtual screen may be.

For many protein drug targets crystal or solution structures are not available. In such cases homology models (130,131) and pseudo-receptor models (132) are often used. However, unless there is a very high conservation of receptor site residues the use of homology models for virtual screening is much riskier than using solved structures. On the other hand, the PDB contains a wealth of protein

Table 6.3 Occurrence of Selected Protein Classes Currently Identified in the Human Genome Database (GDB) and the Protein Database (PDB)

Class	GDB[a]	PDB[b]
Nuclear receptors	49	42
G-protein-coupled receptors	408	1
Kinase	945	625
Protease	190	330
Peptidase	108	128
Esterase	106	87
Reductase	210	417
Synthase	191	335
Lyase	38	70
Hydrolase	131	110
Transferase	500	467
Anhydrase	27	156
Sulfatase	26	6
Dehydrogenase	347	338
Desaturase	10	1
Phosphatase	315	184
Phosphodiesterase	63	1
Deacetylase	18	2
Transporter	238	1
Channel	271	24

[a]www.gdb.org

[b]www.rcsb.org/pdb; note that the number of structures available from the PDB often include several structures of the same protein.

structures of a wide variety of enzymes and receptors that can be used for homology modeling. Homology models can be built for a large number of protein classes coded in the human genome (Table 6.3). Because virtual screening is so inexpensive and the possible rewards, if successful, are so high it is generally warranted to run a virtual screening experiment, even if the chances of success are very small, as is often the case when homology models are employed.

2.4.2 Computational Protein-Ligand Docking Techniques. Docking ligands into a receptor site is a geometric search problem. The search has to take protein and ligand conformations as well as their relative orientations into account. The receptor conformation is typically reasonably well known. However, the bioactive conformation of the ligand is usually unknown. Nicklaus and coworkers showed that force-field energies of bioactive conformations of ligands, as represented in crystal structures of protein-ligand complexes, are typically about 25 kcal/mol[2] higher than minimum conformations in vacuum (133). Therefore, the bioactive conformation of a ligand is hard to guess and a large number of possible ligand conformations have to be considered in docking. Most docking approaches keep the receptor rigid and the ligand flexible during the docking. Although protein flexibility is sometimes included (134–136), we will not discuss protein flexibility here, given that it is currently rarely used for virtual screening because of speed limitations. Some relevant concepts of docking approaches are shortly discussed below (for a broader review we refer the reader to Ref. 125). Scoring protein-ligand complexes will be discussed separately.

2.4.2.1 ***Rigid Docking.*** Although ligand and often also protein flexibility are crucial for protein-ligand docking, the simpler rigid ligand docking is sometimes useful. Ligand flexibility can, for example, be simulated by rigidly docking an ensemble of preassigned ligand conformations that represent the relevant conformational space of the molecule. Algorithms such as clique search techniques (137) and geometric hashing (138) are often used to search for distance-compatible matches of protein and ligand features (139). Possible features include complementary hydrogen-bonding interactions, distances, or volume segments of the receptor site of the protein or the ligand.

The program DOCK uses an algorithm for rigid-body docking based on the idea of searching for distance-compatible matches. Starting with the molecular surface of the protein (140–142), a set of spheres is created inside the receptor site. The spheres represent the volume that could be occupied by a ligand molecule (Fig. 6.18). Spheres can represent the ligand also; a direct atom representation is also possible. Early versions of DOCK relied solely on rigid ligand docking. Sets of up to four distance-compatible matches were evaluated. Each set was used for an initial fit of the ligand into the receptor site. Additional compatibility matches were used to improve the fit. The position of the ligand was then optimized and scored.

Since its first introduction in 1982, the DOCK software has been extended in several directions. The matching spheres can be la-

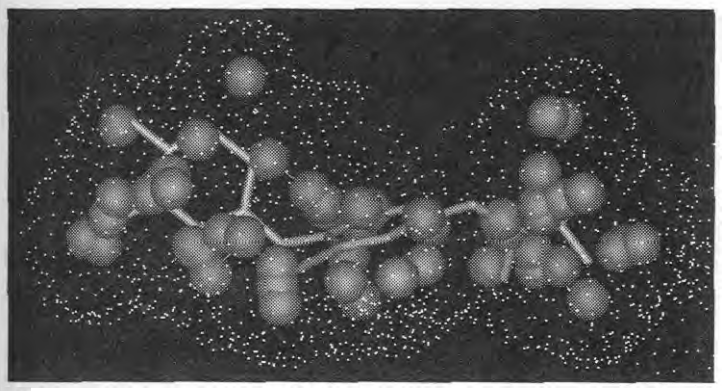


Figure 6.18. Receptor site of thyroid receptor beta filled with spheres (for sake of clarity, sphere centers are depicted; actual size of spheres is larger, so that spheres overlap) and thyronine. Crystal structure taken from the PDB (ID 1bsx).

beled with chemical properties (143) and distaice bins are used to speed up the search process (144,145).Recently, the search algorithm for distance-compatible matches was changed to the clique-detection algorithm introduced by Kuhl (139, 146). Furthermore, several scoring functions are now applied in combination with the DOCK algorithm (147–151).

2.4.2.2 Flexible Ligands. Druglike molecules are typically flexible, with usually up to eight rotatable bonds (34). Energetic differences between alternative ligand conformations are often small compared to the total binding affinity between ligand and target protein. Also, for flexible ligands it is quite common that the bioactive conformations are different from the minimum energy conformations in solution (133). Ligand flexibility is typically handled in docking approaches by combinatorial optimization protocols such as fragmentation, ensembles, genetic algorithms, or simulation techniques.

In fragmentation approaches, the ligand is dissected into pieces that are either rigid or that can be represented by small conformational ensembles. In docking approaches, typically a strategy called incremental *construction* is used to assemble fragments to whole molecules directly in the receptor site. Usually, the largest rigid moiety of the ligand (sometimes called anchor) is docked first in the receptor site. The remaining fragments are subsequently added in a buildup protocol. After each incremental buildup step, torsion angles are sampled and the growing molecule is minimized.

Ligand flexibility can be artificially included into docking by rigidly docking ensembles of pregenerated conformations of the ligand into the receptor site. Rigid docking is faster than flexible docking by use of a fragmentation approach. However, because computing time increases linearly with the number of conformations, computing time and coverage of conformational space have to be balanced. An example of rigid docking of conformation ensembles is given in Flexibase/FLOG (152). Distance geometry methods (153) are used to generate a small set of diverse conformations for each ligand in the database. A subset of up to 25 conformations per molecule is selected using rms dissimilarity criteria and then docked using a rigid-body-docking algorithm.

Different from the combinatorial approaches for docking mentioned above, simulation methods start with a given configuration of a ligand in the receptor site. Simulation techniques such as simulated annealing (154) are then applied to find energetically more favorable conformations of the ligand. To speed up the docking process, docking programs such as AutoDock (155) precalculate molecular affinity potentials of the protein on a grid. Molecular dynamics (MD) methods (see, e.g., refs. 156 and 157) and Monte Carlo simulation techniques (see, e.g., Refs. 158–162) are also frequently used in protein-ligand docking applications.

A variety of other sampling methods are applied in docking programs, including genetic algorithms, distance geometry methods, random searching, hybrid methods, and generalized effective potential methods. Genetic algorithms have been employed in programs such as Gambler (163), AutoDock (155), and GOLD (126).PRO-LEADS uses an alternative search technique called "tabu search" (164). Starting from a random structure, new structures are created by random moves. A tabu list is maintained during the optimization phase and contains the best and the most recently found binding configurations. Configurations that resemble those stored in the tabu list are rejected, except they are better than the one scoring best. The sampling performance is improved because previously sampled configurations are avoided. Finally, it should be

mentioned that multistep hybrid docking procedures have been developed that combine rapid fragment-based searching with sophisticated MC or MD simulations (165, 166).

2.4.3 Scoring of Protein-Ligand Interactions. The problem of sampling the correct binding geometry (binding mode) of a protein-ligand complex is considered to be solved in many docking programs (167). However, to identify this correct binding mode by its lowest energy or score is a different matter; this is indeed the bottleneck of docking-scoring approaches today. The most important aspect of scoring functions for virtual screening is speed. Therefore, accuracy requirements are low; most functions used do not conceptually describe binding free energies. Therefore, these functions are typically not called energy functions but scoring functions. Three main scoring strategies are typically used in docking applications for virtual screening: force field scoring, empirical scoring, and knowledge-based scoring.

2.4.3.1 Force Field (FF) Scoring. Nonbonded interaction energy terms of standard force fields are typically used in FF scoring (e.g., in *vacuo* electrostatic terms; sometimes modified by scaling constants that assume the protein to be an electrostatic continuum) and van der Waals (vdW) terms (168–171). DOCK and GREEN (172) use the intermolecular terms of the AMBER energy function (173, 174), with the exception of an explicit hydrogen bonding term (147):

$$E_{\text{nonbond}} = \sum_{i}^{\text{lig}} \sum_{j}^{\text{prot}} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + 332 \frac{q_i q_j}{D r_{ij}} \right] \tag{6.1}$$

where each term is summed up over ligand atoms i and protein atoms j. A_{ij} and B_{ij} are the vdW repulsion and attraction parameters of the 6–12 potential, r_{ij} is the distance between atoms i and j, q is a point charge at each of the atoms, and D is the dielectric constant. Intra-ligand interactions are added to the score. Up to a 100-fold gain in docking time can be achieved by precomputing these terms on a 3D grid that represents the protein during docking (155, 175). More recently, solvation terms have been added to FF scores. Examples include generalized Born/surface area approaches (176) or atomic solvation parameters (177–179).

2.4.3.2 Empirical Scoring. Empirical scoring functions are multivariate regression methods. They fit coefficients of physically motivated contributions to binding free energy in reproduction of measured binding affinities of a training set of protein-ligand complexes with known 3D structure. As an example, the docking program FlexX (180) uses a scoring function similar to that of Bohm (181,182). It calculates the sum of free-energy contributions from the number of rotatable bonds in the ligand, hydrogen bonds, ion-pair interactions, hydrophobic and pi-stacking interactions of aromatic groups, and lipophilic interactions:

$$\begin{aligned} \Delta G = \Delta G_0 &+ \Delta G_{\text{rot}} N_{\text{rot}} \\ &+ \Delta G_{\text{hb}} \sum_{\text{neutral_Hbonds}} f(\Delta R, \Delta\alpha) \\ &+ \Delta G_{\text{io}} \sum_{\text{ionic_int}} f(\Delta R, \Delta\alpha) \\ &+ \Delta G_{\text{aro}} \sum_{\text{aro_int}} f(\Delta R, \Delta\alpha) \\ &+ \Delta G_{\text{lipo}} \sum_{\text{lipo.cont}} f^*(\Delta R) \end{aligned} \tag{6.2}$$

where ΔG_0, ΔG_{rot}, AG,, ΔG_{io}, ΔG_{aro}, and ΔG_{lipo} are adjustable parameters that are fitted; $f(\Delta R, \Delta\alpha)$ is a scaling function penalizing deviations from the ideal geometry; and N_{rot} is the number of freely rotatable bonds. The interaction of aromatic groups is an addition to Bohm's original force-field design (181, 182). The lipophilic contributions are calculated as a sum of atom-pair contacts in contrast to evaluating a surface grid as in Bohm's scoring function. Bohm's scoring function and its FlexX implementation are being improved and additional terms are being tested (see, e.g., Refs. 182 and 183).

2.4.3.3 Knowledge-Based Scoring. Because the forces that govern protein-ligand interactions are so complex, an implicit approach to capture all relevant terms of protein-ligand

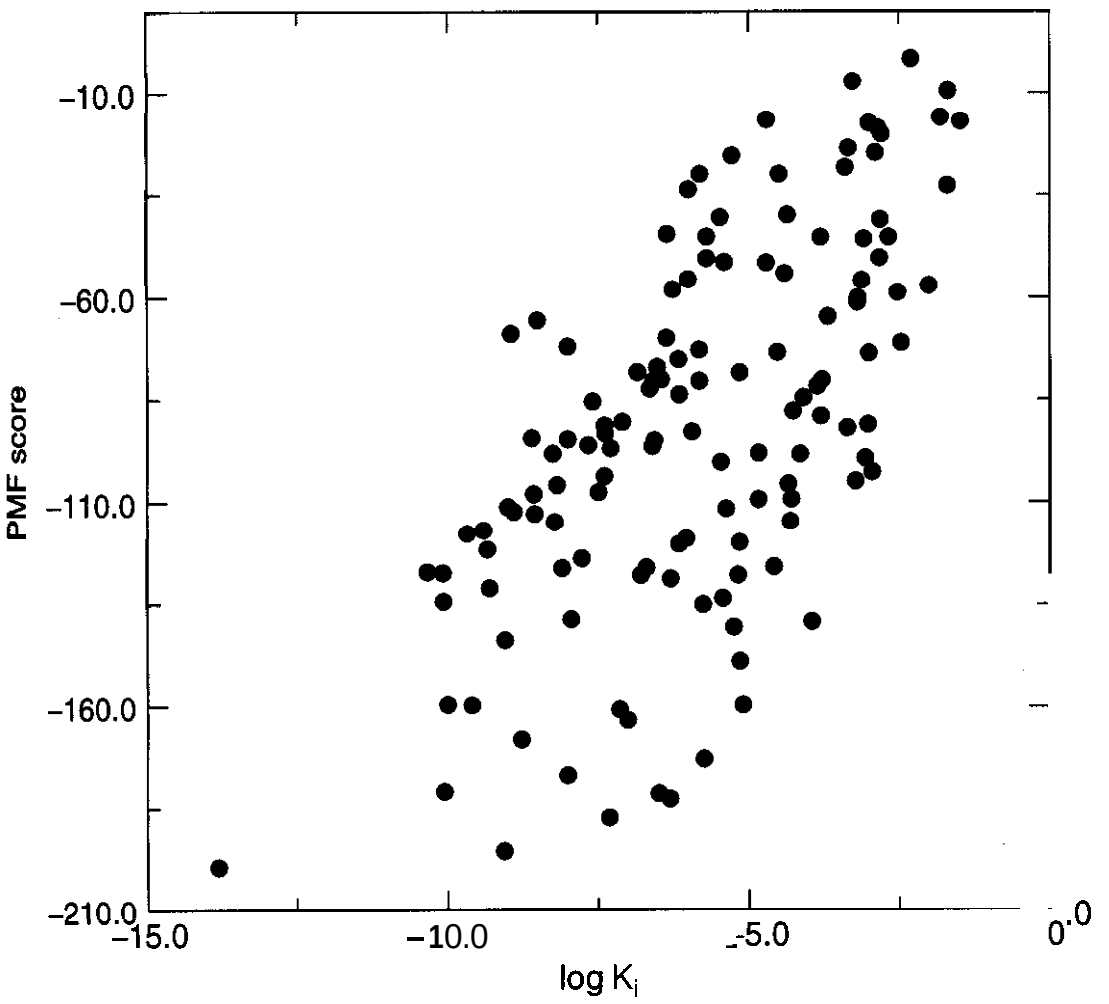


Figure 6.19. PMF score calculated for 132 protein-ligand complexes taken from the PDB without overlap to the set of 697 complexes. The PMF score was derived from Ref. 186.

binding seems very attractive. Borrowing from statistical thermodynamics of liquids, mean-field approaches derived solely from structural information have been applied to protein-ligand binding. Protein-ligand atom-pair potentials can be calculated from structural data (e.g., PDB), assuming that observed crystallographic protein-ligand complexes exhibit optimal placement. As an example, a knowledge-based scoring function was derived recently using 697 protein-ligand complexes from the PDB as knowledge base. Using 16 protein and 34 ligand atom types, a total of 282 statistically significant interaction potentials of atom pairs was derived. The final score is calculated as the sum over all protein-ligand atom-pair interactions.

$$\text{PMF_score} = \sum_{kl \to r < r_{\text{cutoff}}^{ij}} A_{ij}(r); \qquad (6.3)$$

$$A_{ij}(r) = -k_B T \ln\left[f_{\text{Vol_corr}}^{j}(r) \frac{\rho_{\text{seg}}^{ij}(r)}{\rho_{\text{bulk}}^{ij}} \right]$$

where kl is a ligand-protein atom pair of type ij; r_{cutoff}^{ij} designates the distance at which atom-pair interactions are truncated (6 Å for carbon-carbon interactions and 9 Å otherwise); all $A_{ij}(r)$ are derived with a reference sphere radius of 12 Å (184); k_B is the Boltzmann factor; T is the absolute temperature; and $f_{\text{Vol corr}}^{j}(r)$ is a ligand volume correction factor that is introduced because intraligand interactions are not accounted for (185, 186). $\rho_{\text{seg}}^{ij}(r)$ designates the number density of atom pairs of type ij at a certain atom-pair distance r. ρ_{bulk}^{ij} is the number density of a ligand-protein atom pair of type ij in a reference sphere with radius R (184). For use in docking studies, the PMF score is combined with a vdW term to account for short-range interactions (187, 188). The PMF scoring function was implemented into the DOCK4.0 program. For faster scoring it was also implemented on a grid similar to the force-field score in DOCK. Flexible docking experiments on FK506 binding protein (187), neuraminidase (127), and stromelysin (189) showed high predictive power and robustness of the PMF score. Figure 6.19 shows the predictive power of the scoring function applied to 132 protein-ligand complexes taken from the PDB.

2.4.3.4 Consensus Scoring. Consensus scoring is an approach that combines several scoring functions to find common hits. Such an approach seems desirable because of the missing robustness of current scoring functions.

Charifson et al. (163) provided a comprehensive consensus scoring study using DOCK and Gambler, in combination with 13 scoring functions: LUDI (104), ChemScore (190, 191), Score (192), PLP (193), Merck force field (194), DOCK energy score (146, 147), DOCK chemical score, Flog (152), strain energy, Poisson Boltzmann (195), buried lipophilic surface area (196), DOCK contact score (144), and volume overlap (197). Three enzymes were used as test proteins: p38 MAP kinase, inosine monophosphate dehydrogenase, and HIV protease. By comparing the performance of single-scoring functions with consensus scoring schemes involving two or three scoring functions, the authors found that false positives (inactive compounds that have high predicted scores) were significantly reduced in the latter case. The authors estimated that a consensus scoring approach would consistently provide hit rates between 5 and 10% (5–10 out of 100 compounds tested to show low μM activity) for enzymes with reasonably buried binding sites. A comparison of the different scoring functions revealed that ChemScore, PLP, and DOCK energy score performed best as single-scoring functions and also in consensus combination. Consensus scoring experiments reported by Bissantz et al. found that docking/consensus scoring performances varied widely among targets (198). Stahl and Rarey suggested that the combinations of FlexX and PLP scores are ideal for consensus scoring for a variety of targets including COX-2, ER, p38 MAP kinase, gyrase, thrombin, gelatinase A, and neuraminidase (199).

2.4.4 Docking as Virtual Screening Tool. A virtual screening protocol is schematically shown in Fig. 6.20. The necessary steps include: protein structure preparation, ligand database preparation, docking calculation, and postprocessing.

The protein has to be prepared only once for a virtual screening experiment unless different protein conformations are considered. The receptor site needs to be determined and charges have to be assigned. The protein structure and the receptor site have to be modeled as accurately as possible. Determining protein surface atoms and site points as well as the assignment of interaction data, such as

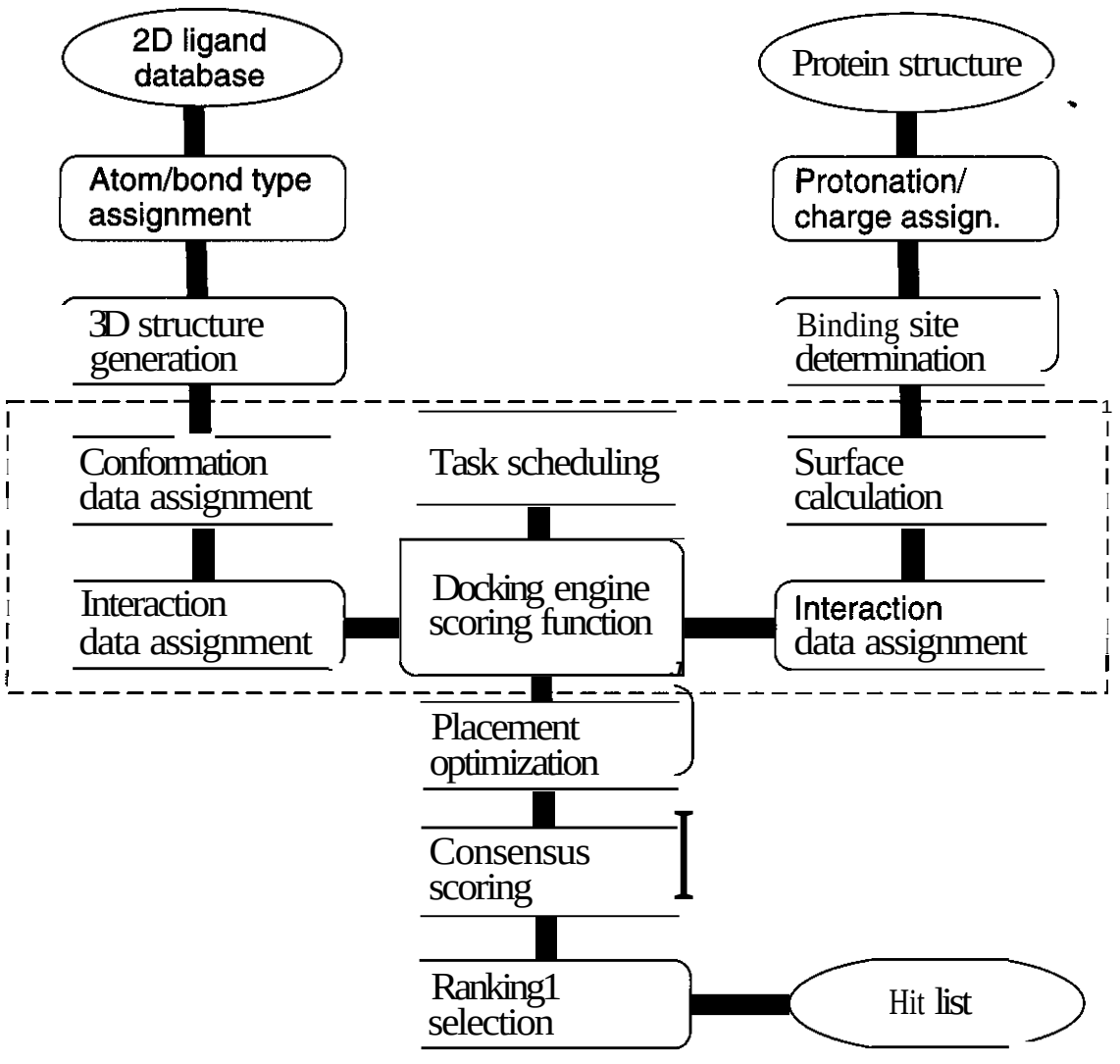


Figure 6.20. Flowchart of docking as virtual screening tool in the example of FlexX.

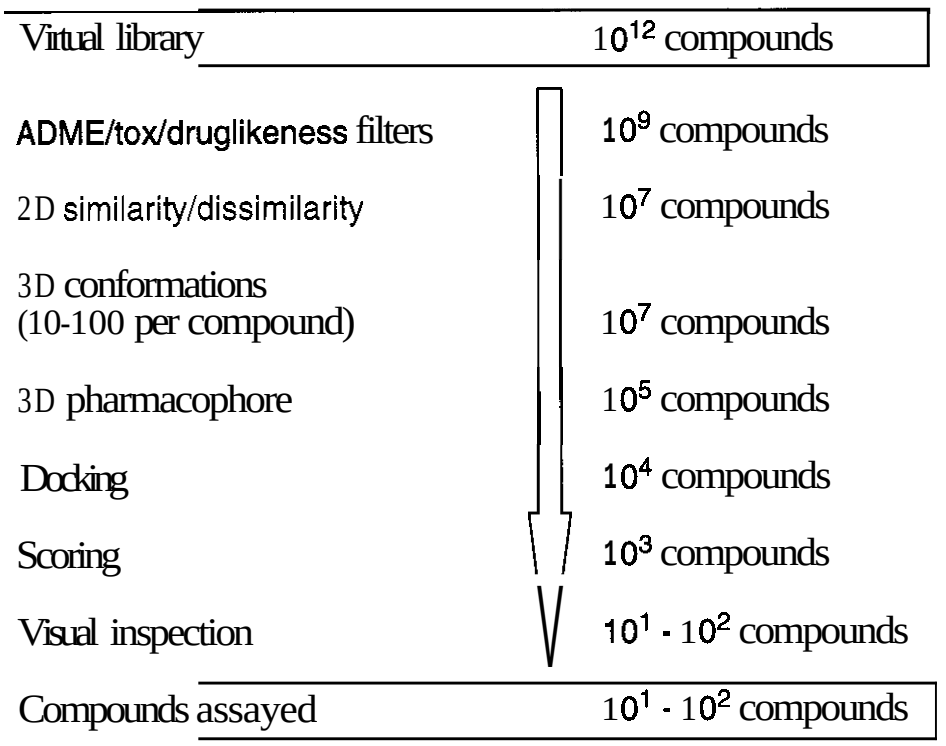


Figure 6.21. Virtual screening filter cascade.

marking hydrogen-bond donors/acceptors, and so forth, are sometimes internally included in the docking software (e.g., in FlexX) and sometimes done separately (e.g., DOCK).

Because of the large number of molecules, manual steps in the preparation of ligand databases obviously have to be avoided. Starting typically from 2D structures, bond types have to be checked, protonation states must be determined, charges must be assigned, and solvent molecules removed. 3D coordinates can be generated using a program such as CONCORD or CORINA (74) (see Section 2.3.2). Next, site points for hydrogen-bonding interactions have to be assigned and rotational barriers must be calculated. These tasks are sometimes included in the docking program (e.g., FlexX).

The docking calculation is typically done for one ligand at a time. Depending on optimization and sampling parameters as well as on the flexibility of the compound, typically between a few seconds and a few minutes of CPU time are needed to dock a ligand. Because the individual docking events are independent of each other, they can run on parallel hardware. Task schedulers that distribute ligand docking on available CPUs are used in many docking programs.

Postprocessing steps of hits may include refinement of placement using MD techniques, specific pharmacophore-based filters that penalize certain features, such as unformed hydrogen bonds or other constraints that were not met in the primary scoring function. Because of the limitations of scoring functions, a postscoring protocol can be used to reach consensus about hits (discussed above). The recognition of known active ligands mixed within the database can be used to find an appropriate threshold for separating the top-ranking compounds from the rest of the database.

2.5 Filter Cascade

Virtual screening is the process of reducing a given database as quickly and efficiently as possible to a small number of putative lead compounds for a given drug discovery project. The techniques described above form a cascade of different filter functions that are ordered by their speed. Fast ADMET filters are followed by 2D and 3D pharmacophore filters and finally by docking and scoring methods. Figure 6.21 shows a scheme of a possible virtual screening filter cascade.

3 APPLICATIONS

3.1 Identification of Novel DAT Inhibitors through 3D Pharmacophore-Based Database Search

The dopamine transporter (DAT) is a 12-transmembrane helix protein that plays a critical role in terminating dopamine neurotrans-

mission by taking up dopamine released into the synapse. There is no experimental structure available for DAT. However, an extensive SAR of DAT inhibitors (mostly cocaine analogs) is available. DAT is involved in several diseases such as drug addiction and attention deficit disorder (200). For example, ritalin [(±)-threo-methylphenidate], a DAT inhibitor, is marketed for treating attention deficit disorders in children (200, 201). Until recently, all efforts in synthesizing DAT inhibitors were focused on creating analogs around the tropane, piperazine, methylphenidate, and 2,3-dihydro-5-hydroxy-5*H*-imidazo[2,1-*a*]isoindole cores. It was shown that, despite structural differences, DAT inhibitors share one or more common 3D pharmacophore models (95, 202, 203). In an effort to identify new chemical cores for developing DAT inhibitors with new pharmacological profiles, a pharmacophore-based 3D database search was proposed (95). For this purpose a pharmacophore model was derived based on two known potent DAT inhibitors R-cocaine and WIN-35065–2 (Fig. 6.22) (95). The common binding elements of these compounds are a ring N that may be substituted, a carbonyl oxygen, and an aromatic ring that can be defined by the position of its center (Fig. 6.22). Because both compounds have some flexibility, a systematic conformational search was performed to obtain all possible conformations these compounds can have when bound to DAT. To identify structurally diverse conformers, clustering of the generated conformers was done. Measuring distances among chosen pharmacophore elements in the generated conformer led to distances shown in Fig. 6.22.

Recently, analysis of several large chemical databases showed that the NCI database has by far the highest number of unique compounds (204). Thus this database provides a large number of unique synthetic compounds and natural products and is an excellent re-

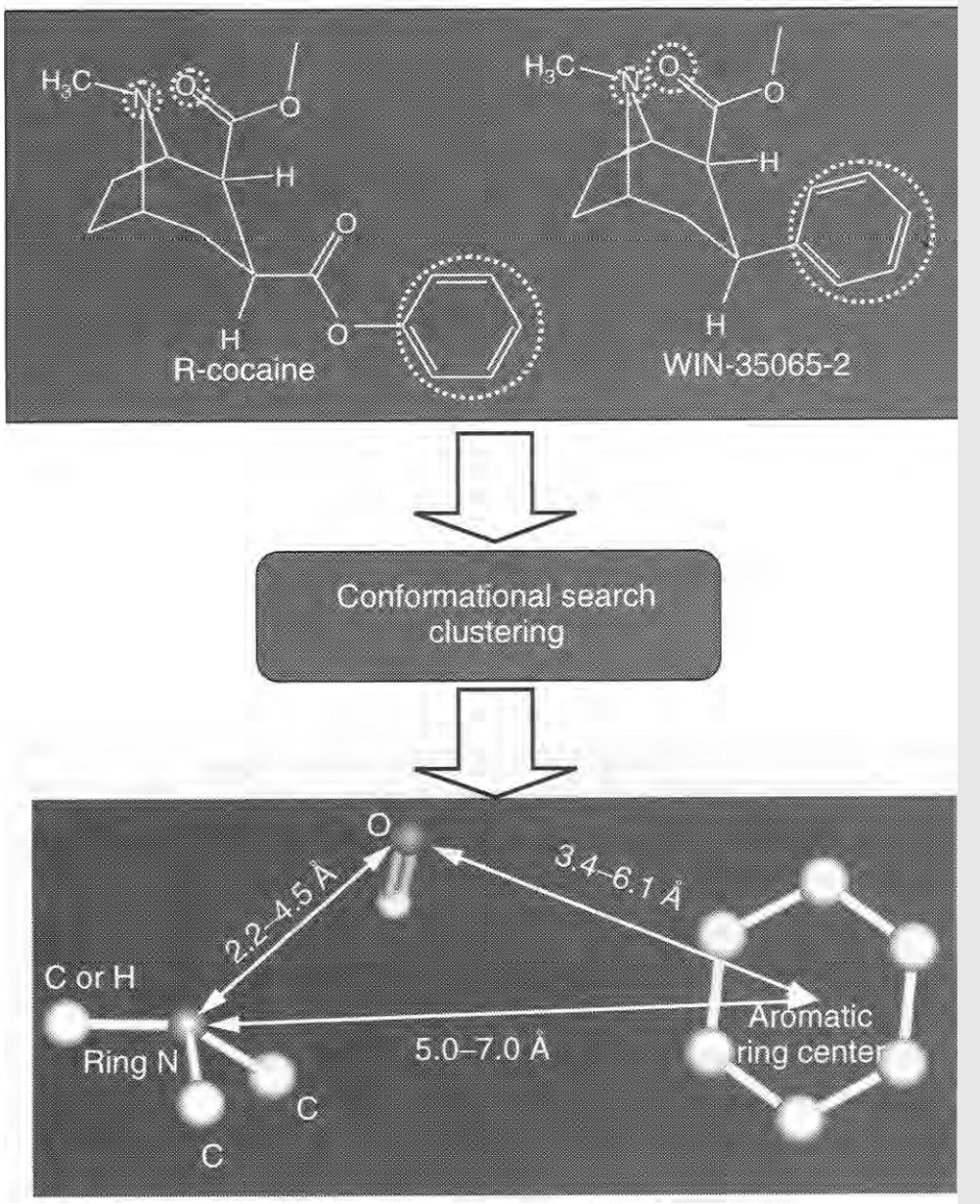


Figure 6.22. Pharmacophore proposed for identifying DAT inhibitors. The pharmacophore was obtained based on two known DAT inhibitors, R-cocaine and WIN-35065-2. Distance ranges between pharmacophore points were obtained through systematic search of all possible conformations that the two compounds may adopt when bound to DAT.

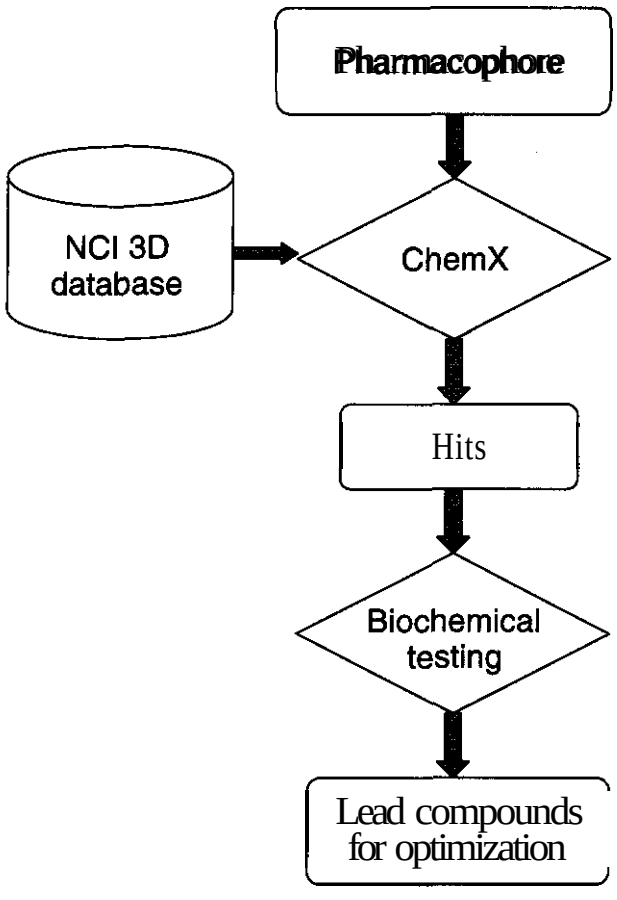


Figure 6.23. Flowchart showing steps used in lead identification using pharmacophore-based 3D database searching.

source for drug lead discovery. Using the 3D pharmacophore from Fig. 6.22, the NCI 3D-database (67) of 206,876 "open compounds" was searched using the program Chem-X (205). The strategy used for identifying leads through virtual screening is shown in Fig. 6.23. During the search each compound was first checked as to whether it had the pharmacophore elements and second as to whether it had any acceptable conformation matching the distance requirements. Up to 3 million conformations were examined for each compound. A total of 4094 compounds, 2% of the database, were identified as "hits." This number was further reduced using filters such as molecular weight, structural novelty, simplicity, diversity, and hydrogen-bond acceptor nitrogen. Seventy compounds were selected for testing in biochemical assays. Forty-four compounds displayed more than 20% inhibition at 10 μM in the [^{3}H]mazindol binding assay, from which three compounds were chosen for deriving an SAR (Fig. 6.24). These results suggested that the 3D pharmacophore-based database search is an efficient tool for identifying novel DAT inhibitors.

3.2 Discovery of Novel Matriptase Inhibitors through Structure-Based 3 D Database Screening

Matriptase is a trypsinlike serine protease that was proposed to be involved in tissue remodeling, cancer invasion, and metastasis (206). Potent and selective matriptase inhibitors not only would be useful for further elucidation of the role matriptase has in biological systems but also may be used for the treatment and/or prevention of cancers. Hepatocyte growth factor activator inhibitor 1 (HAI-1) is a natural inhibitor of matriptase. Thus, by analyzing interactions in the complex of matriptase with HAI-1, crucial interactions that an inhibitor should capture can be identified. In consequence, the strategy for identifying inhibitors was to first build the matriptase-HAI-1 Kunitz domain 1 complex, identify binding regions on matriptase, screen the NCI 3D database for hits that capture binding groups of HAI-1 to matriptase, and in the end, biochemical testing (Fig. 6.25). The structure of matriptase was obtained from PDB entry 1EAW (207). Homology modeling, as implemented in MODELLER (208, 209), was chosen to build the 3D structure of the Kunitz domain 1 from KSPI. The complex of matriptase with HAI-1 Kunitz domain 1 was built using a combination of manual docking and molecular dynamics refinement with the program CHARMM (210). The obtained binding mode of HAI-1 Kunitz domain 1 to matriptase (Fig. 6.26) suggests that three regions might be important for inhibitor binding. The S1 binding site Asp185, which is characteristic of trypsinlike serine proteases, is the specificity pocket used to recognize substrates with Arg or Lys as P1 residue. The anionic site, defined by Asp96, Asp60.A, and Asp60.B, is the site at which Arg258 from HAI-1 binds. A hydrophobic region defined by Ile41 and Tyr60.G might also be important for specificity of future matriptase inhibitors.

Thus, the active site used for in *silico* .screening with the program DOCK constitutes all three binding regions. Energy scoring was used for ranking docked compounds. The top 2000 compounds were considered for selecting potential inhibitors. Given that

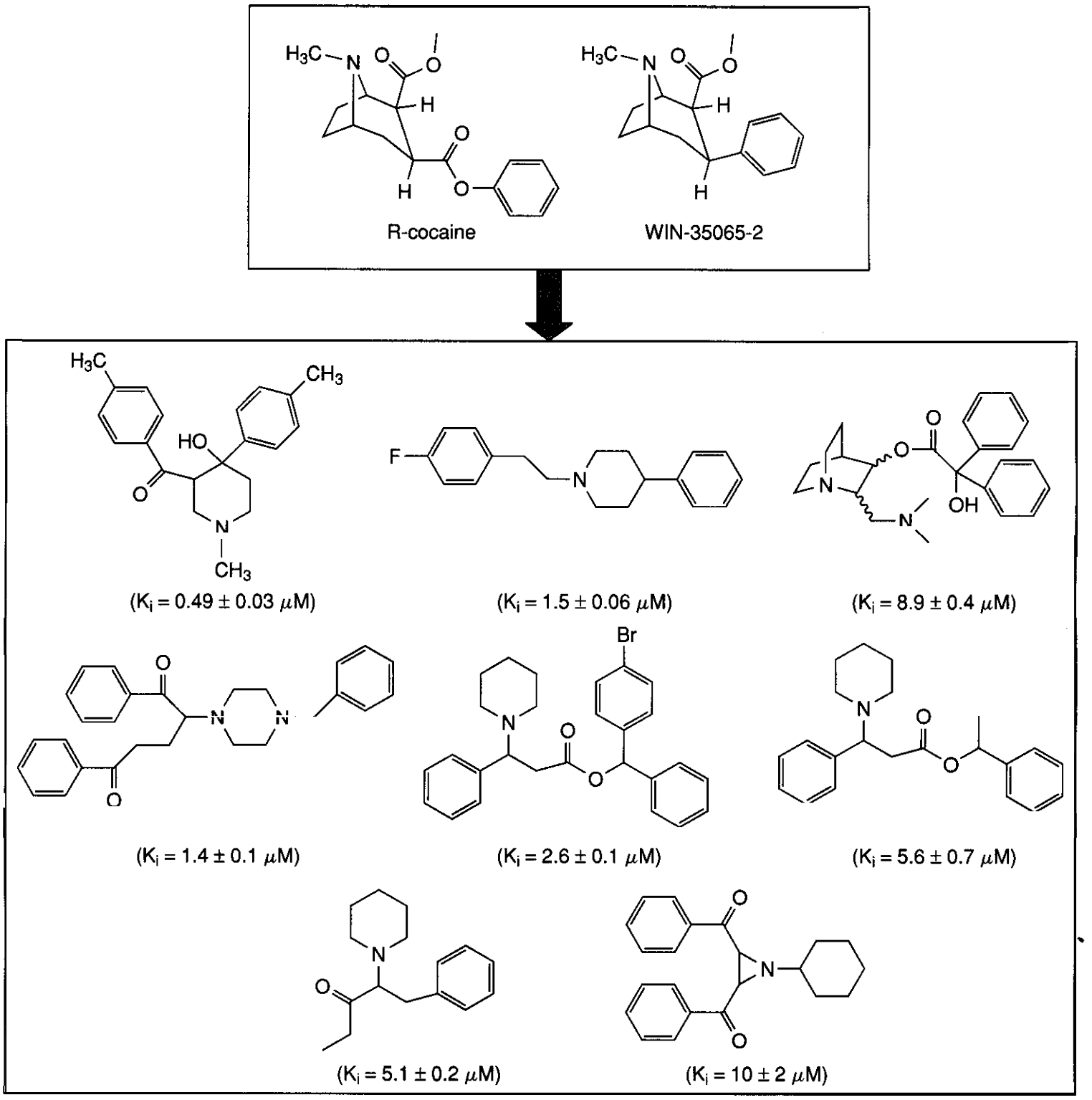


Figure 6.24. Selected DAT inhibitors identified from the NCI database.

matriptase prefers positively charged residues in the P1 position, inhibitors should also have positively charged groups to bind efficiently to Asp185 from the S1 site of matriptase (Fig. 6.27). Note that a more efficient way of doing the virtual screening presented above is to do a pharmacophore search first followed by docking. Thus, 69 compounds were selected for biochemical testing at 75 μM inhibitor and matriptase concentration. Initial screening showed that 50% of compounds tested produced more than 70% inhibition of enzymatic activity when the ratio was one inhibitor molecule for one protein molecule (Table 6.4). It should be noted that screening results at single dose and IC_{50} depend on the protein concentration, whereas K_i is concentration independent. From the hits in the screening step bis-benzamidines were chosen for K_i determination (Table 6.5) because this class of compounds could bind to both the S1 site and the anionic site. These results show that combining a pharmacophore hypothesis with a structure-based database search can provide an efficient way of identifying leads for a drug design project.

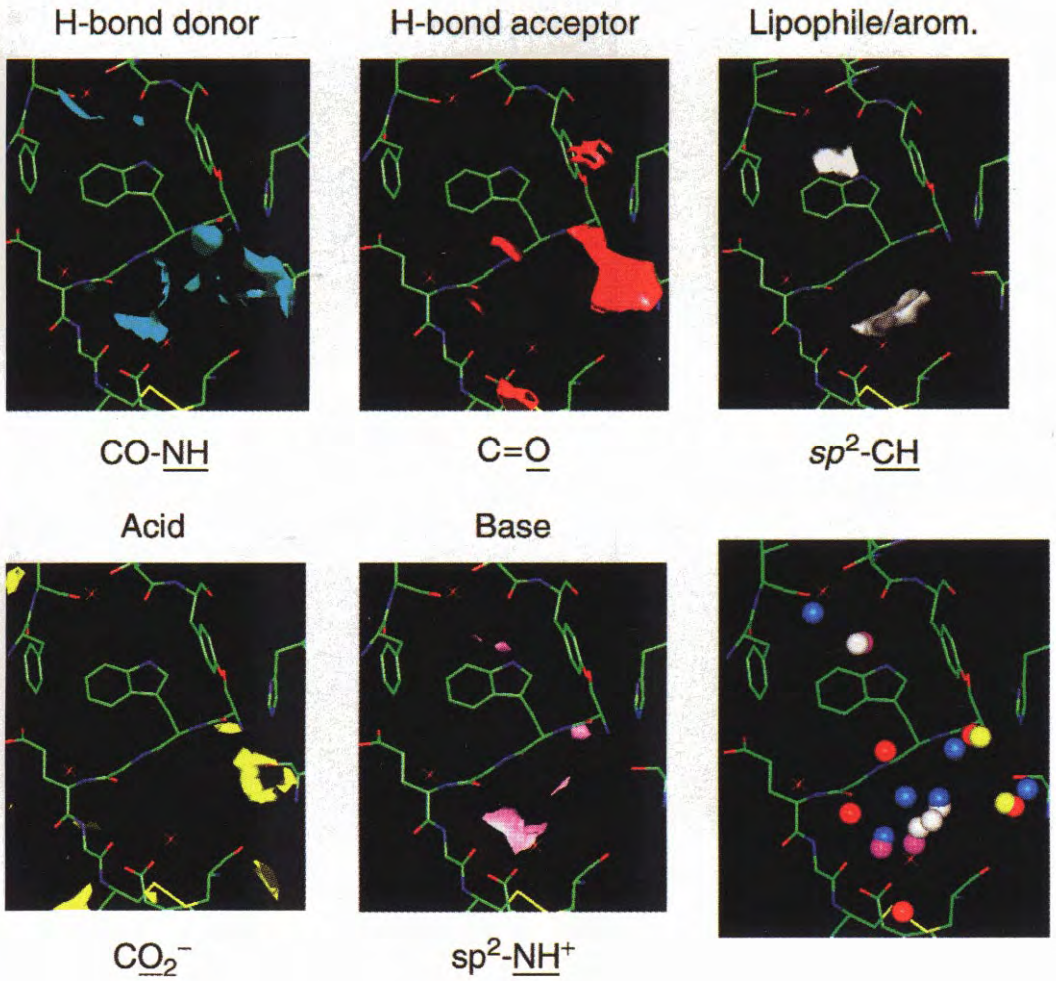


Figure 5.5. GRID probes on Factor Xa site and the combined resultant complementary site points that can be used for pharmacophore fingerprint calculations (lower right).

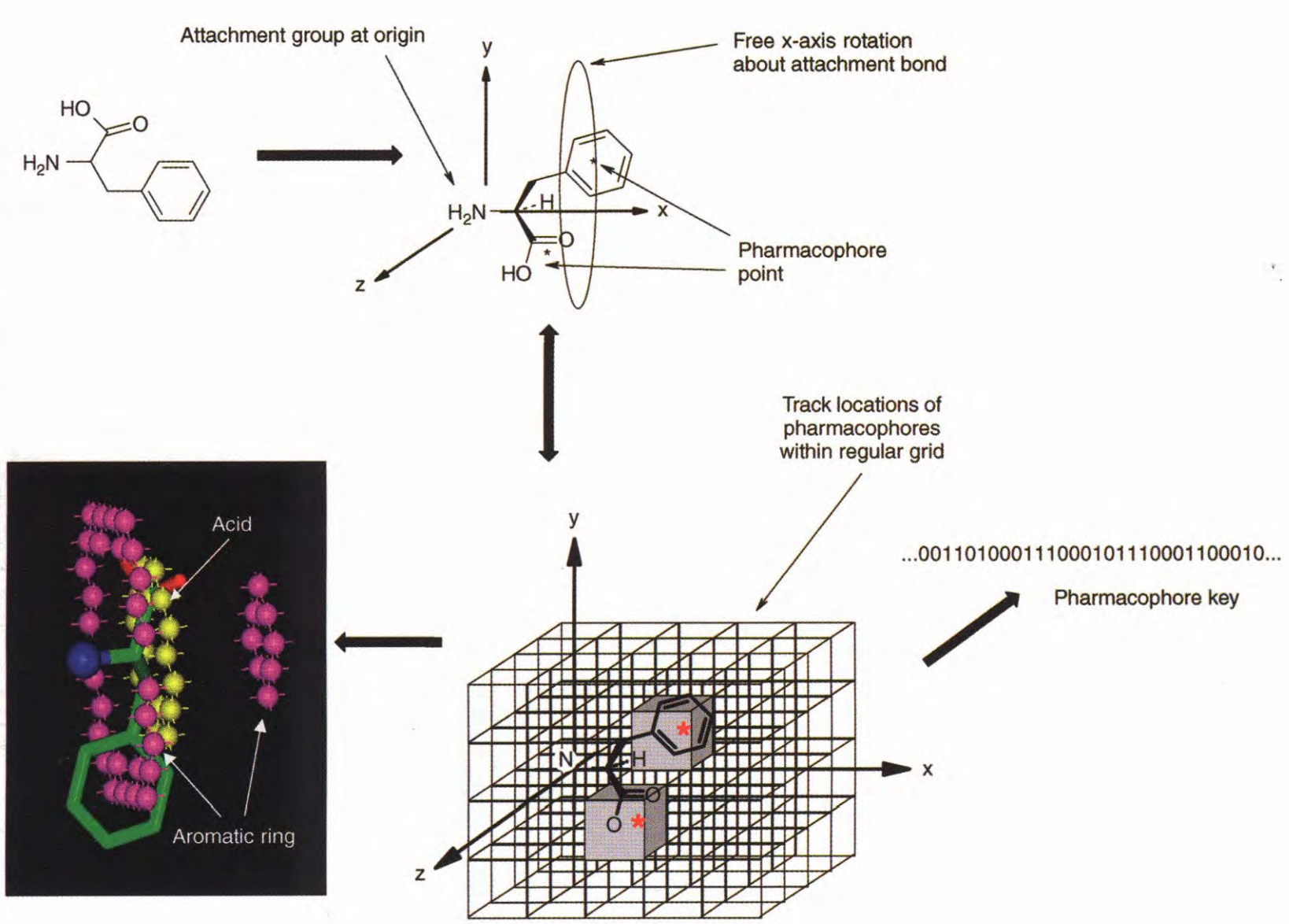


Figure 5.6. Overview of the Gridding and Partitioning (GaP) procedure as applied to monomers, exemplified using phenylalanine as a potential primary amine. This molecule thus contains two pharmacophoric groups (the aromatic ring and the carboxylic acid). During the conformational analysis the locations of these pharmacophoric groups are tracked within a regular grid. [Reproduced from A. R. Leach and M. M. Hann, *Drug Discovery Today*, **5, 326–336** (2000), with permission of Elsevier Science.]

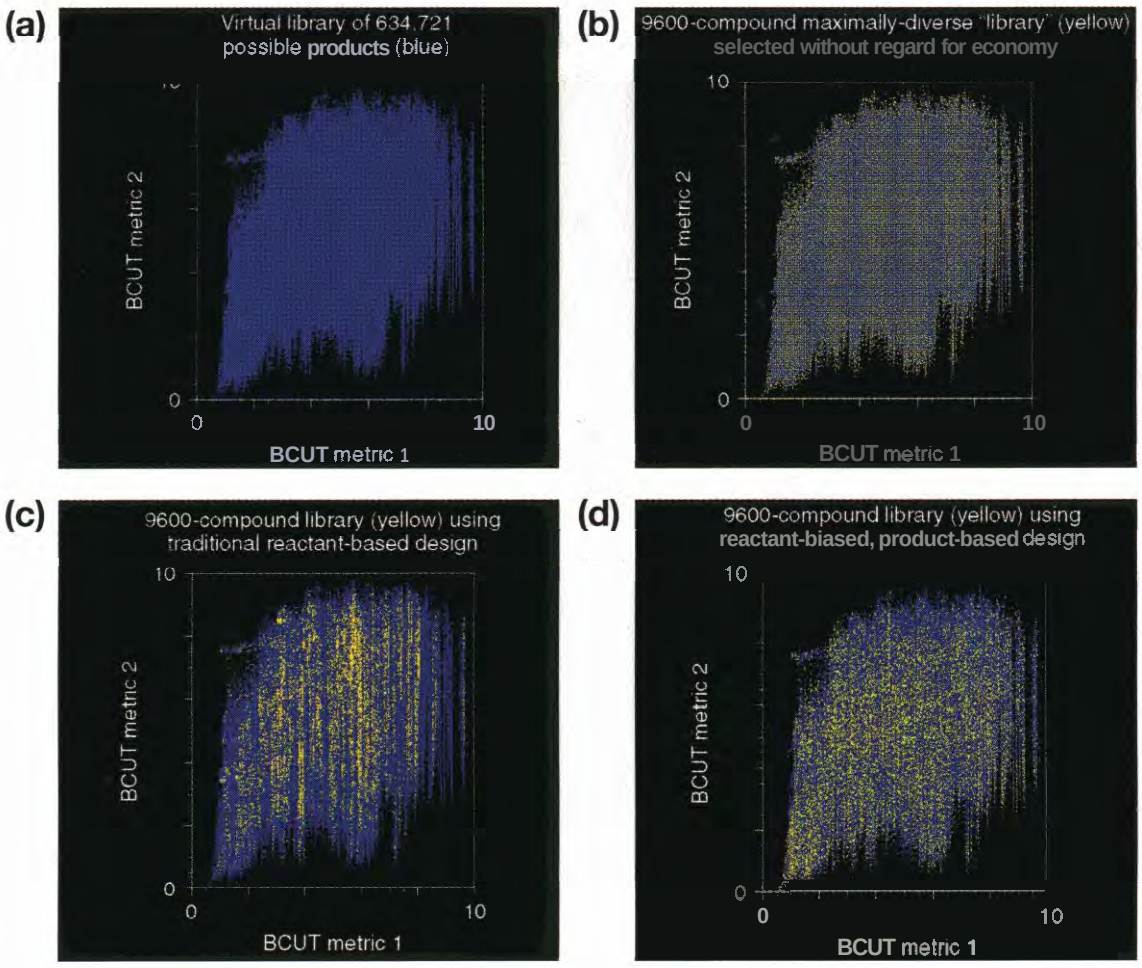


Figure 5.13. (a)A virtual library of 634,721 allowed combinatorial AB products (after filtering out products that failed Lipinski's Rule of 5 "druglike" criteria) shown in a BCUT chemistry space specifically chosen to best represent the diversity of the virtual library. (b) The maximally diverse 9600-compound subset of the virtual library, illustrating the results of purely product-based "library design." Although providing the maximal diversity, synthesis of these 9600 AB products would require the use of 347 As and 1024 B's--clearly unacceptable from the perspective of synthetic economy (number of reactants and robotic control). (c) The 9600-compound library resulting from the traditional, purely reactant-based library design strategy of selecting the 80 most diverse As and the 120 most diverse B's. Although providing user-selected synthetic economy, the diversity of these 9600 AB products is clearly quite poor. (d) The 9600-compound library resulting from the reactant-biased, product-based (RBPB) algorithm developed by Pearlman and Smith (see Refs. 31, 87c and text). The algorithm selected a different set of 80 As and a different set of 120 B's, thus providing the same level of user-selected synthetic economy, while also providing substantially greater diversity than could be achieved using a purely reactant-based library design strategy.

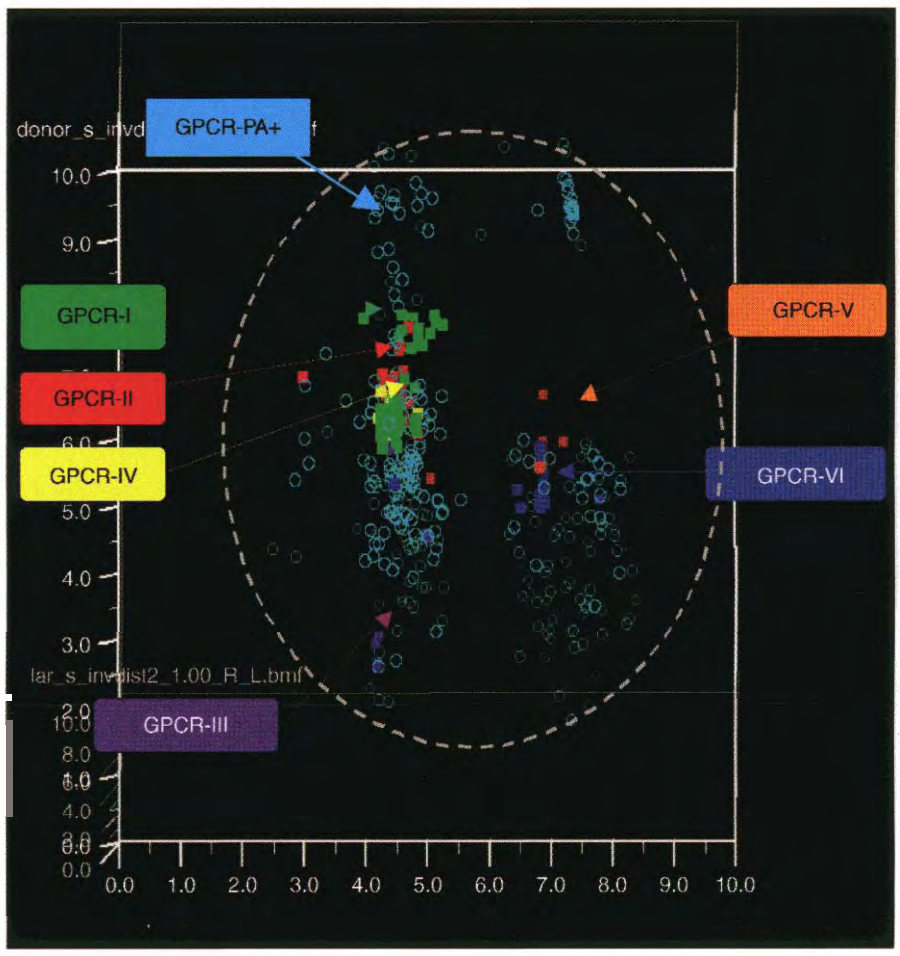


Figure 5.25. The 3D subspace most receptor relevant for members of the GPCR-PA+ family of receptors. Points indicate coordinates of 187 published ligands of various GPCR-PA+ receptors. Some have been color-coded by receptor for illustrative purposes. See Refs. 32e,i and text for further details.

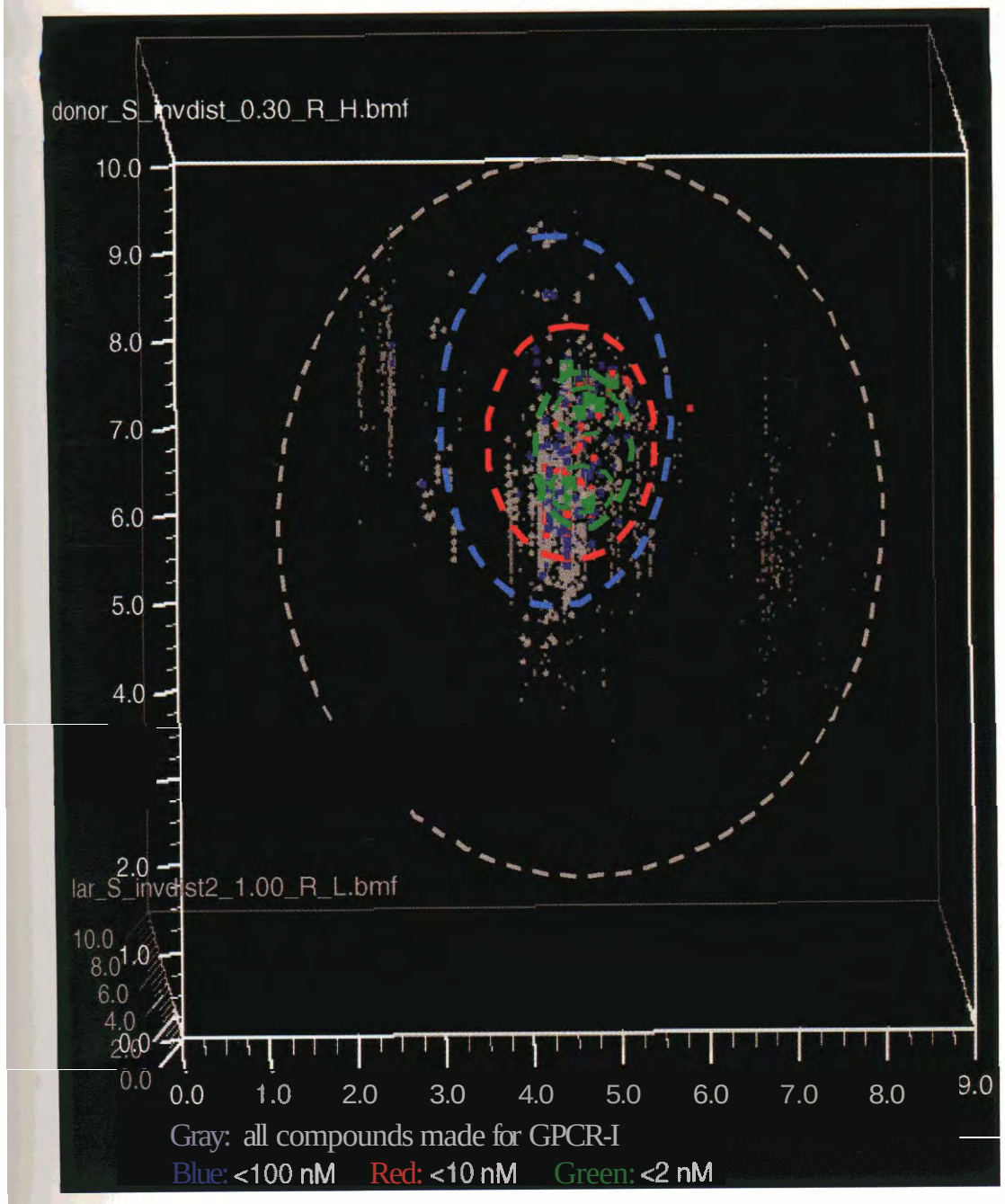


Figure 5.26. The same **3D subspace** as in Fig. 5.25, rotated slightly to provide a better viewing perspective. Points indicate coordinates of about 2000 combinatorial products selected from 14 different libraries. Color-coding indicates affinity for the GPCR-1 receptor. See Refs. 32e,i and text for further details.

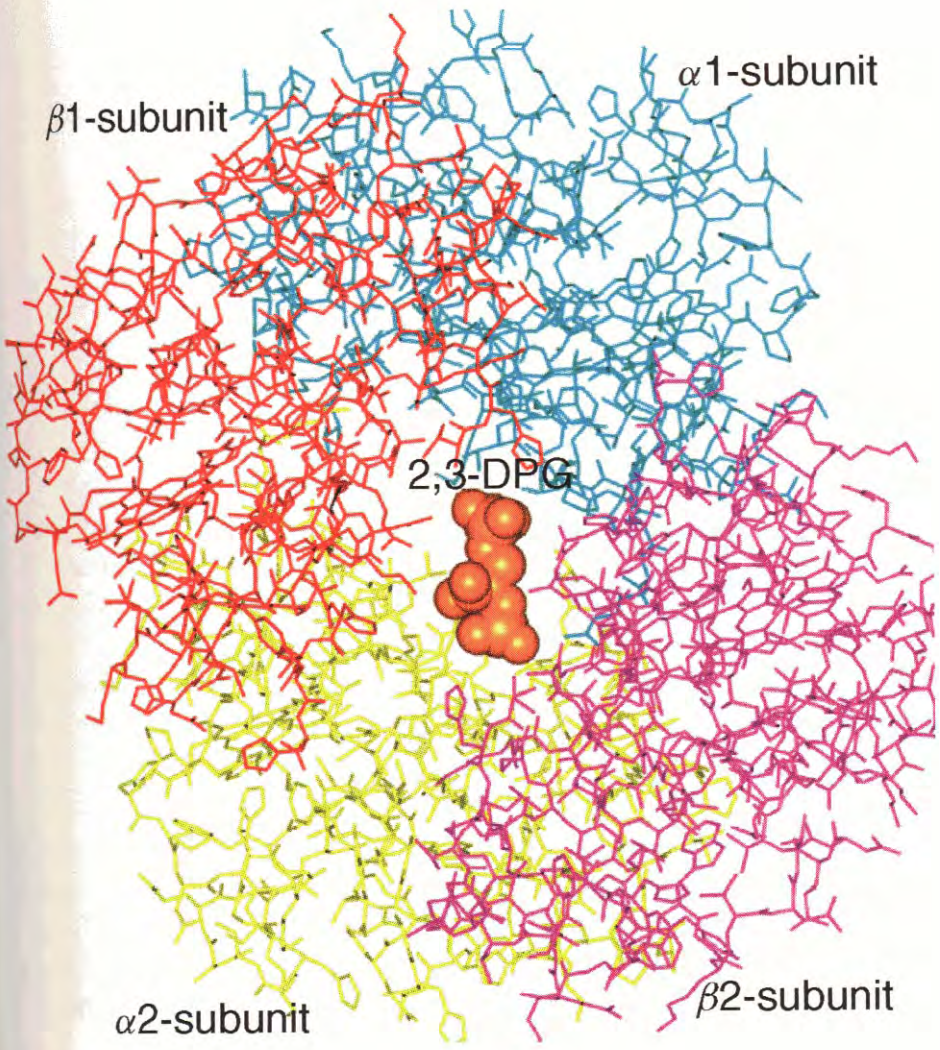


Figure 10.2. View of (4) (2,3-DPG) binding site at the mouth of the β-cleft of deoxy hemoglobin.

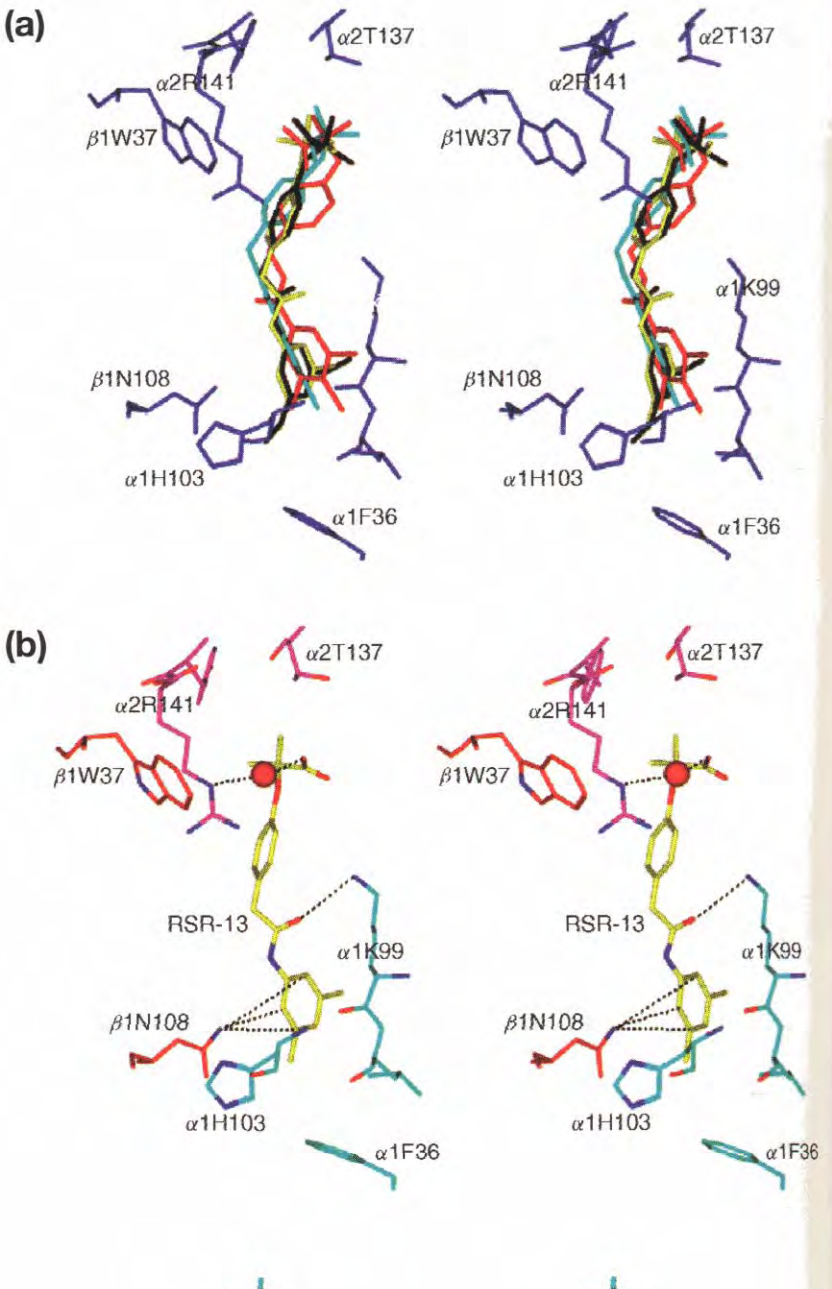


Figure 10.3. Stereoview of allosteric binding site in deoxy hemoglobin. A similar compound environment is observed at the symmetry-related site, not shown here. (a) Overlap of four right-shifting allosteric effectors of hemoglobin: (**6a**) (RSR13, yellow), (**6b**) (RSR56, black), (**7a**) (MM30, red), and (**7b**) (MM25, cyan). The four effectors bind at the same site in deoxy henoglobin. The stronger acting RSR compounds differ from the much weaker MM compounds by reversal of the amide bond located between the two phenyl rings. As a result, in both RSR13 and RSR56, the carbonyl oxygen faces and makes a key hydrogen bonding interaction with the amine of αLys99. In contrast, the carbonyl oxygen of the MM compounds is oriented away from αLys99 amine. The αLys99 interaction with the RSR compounds appear to be critical in the allosteric differences. (b) Detailed interactions between RSR13 (**6a**) and hemoglobin, showing key hydrogen bonding interactions that help constrain the T-state and explain the allosteric nature of the compound and those of other related compounds.

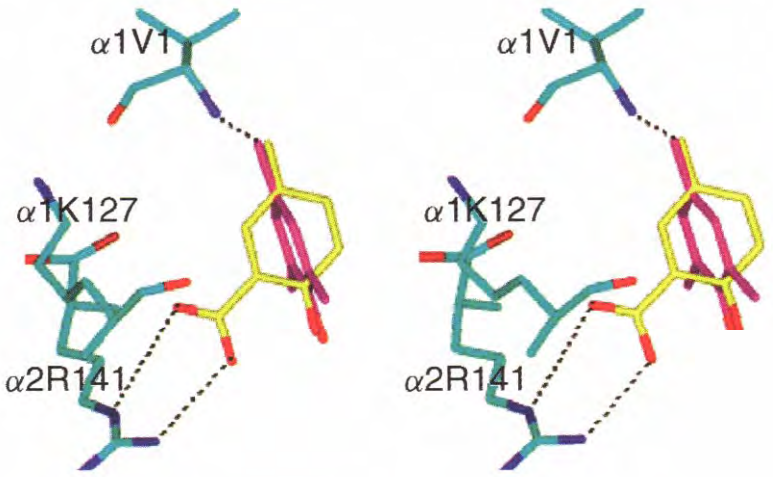


Figure 10.4. Stereoview of superimposed binding sites for (**8b**) (5-FSA, yellow) and (**8a**) (DMHB, magenta) in deoxy hemoglobin. A similar compound enviroment is observed at the symmetry-related site and therefore not shown here. Both compounds form a Schiff base adduct with the α1Val1 N-terminal nitrogen. Whereas the m-carboxylate of 5-FSA forms a salt bridge with the α2Arg141 (opposite subunit), this intersubunit bond is missing in DMHB. The added constraint to the T-state by 5-FSA that ties two subunits together shifts the allosteric equilibrium to the right. On the other hand, the binding of DMHB does not add to the T-state constraint. Instead, it disrupts any T-state salt- or water-bridge interactions between the opposite a-subunits. The result is a left shift of the oxygen equilibrium curve by DMHB.

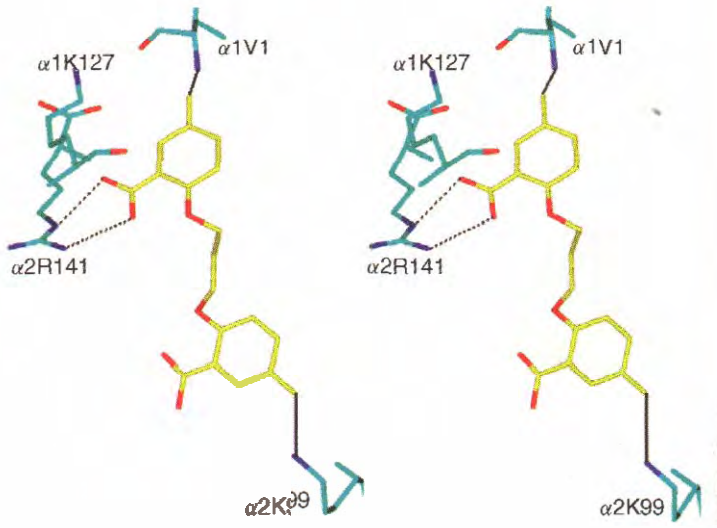
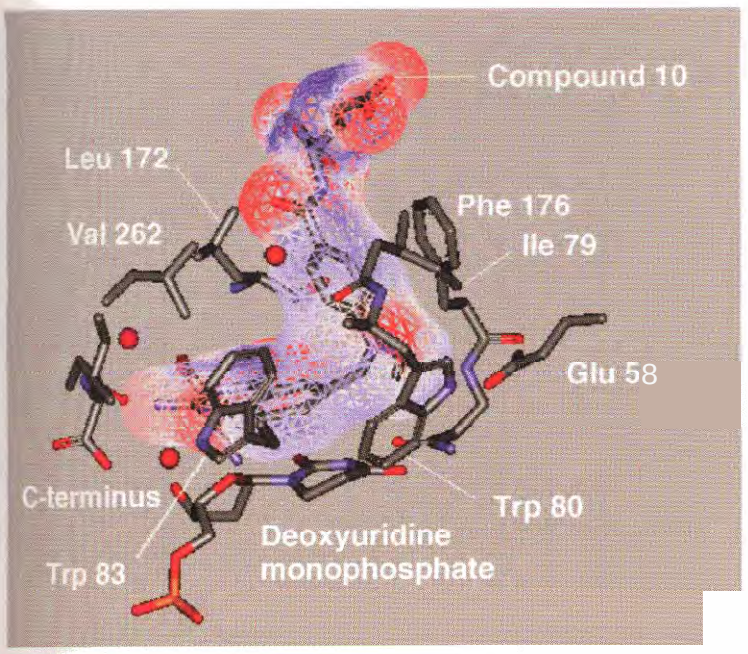


Figure 10.6. Stereoview of the binding site for (**9**) (n = 3, TB36, yellow) in deoxy Hb. A similar compound environment is observed at the symmetry-related site, not shown here. One aldehyde is covalently attached to the N-terminal α1Val1, whereas the second aldehyde is bound to the opposite subunit, α2Lys99 ammonium ion. The carboxylate on the first aromatic ring forms a bidentate hydrogen bond and salt bridge with the guanidinium ion of α2Arg141 of the opposite subunit. The effector thus ties two subunits together and adds additional constraints to the T-state, resulting in a shift in the Hb allosteric equilbrium to the right. The magnitude of constraint placed on the T-state by the crosslinked αLys99 varies with the flexibility of the linker. Shorter bridging chains form tighter crosslinks and yeild larger shifts in the allosteric equilibrium.



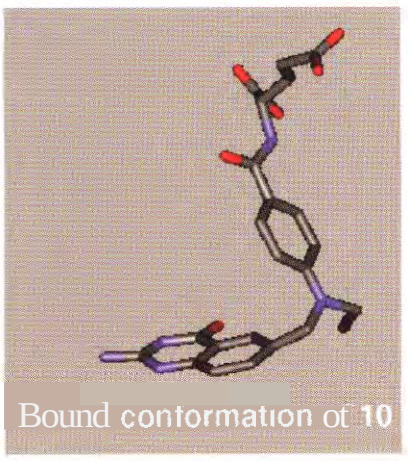


Figure 10.6. Binding site for (10) (*N10-propynyl-5,8-*dideazafolate), within the active site of **thymidylate** synthase from Escherichia coli. The surface of the inhibitor is shown in the **left** view. The red spheres in the left view are tightly bound water molecules.

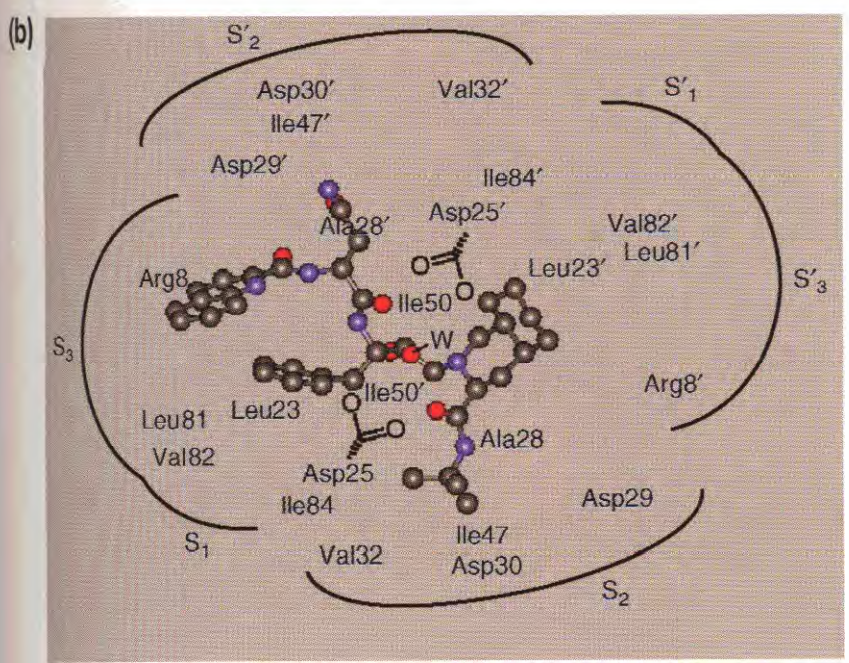


Figure 10.9. **(b)** Active site with bound **(31)** [saquinavir (PDB code 1HXB)]. Note the **asym-**metry of inhibitor binding. The flap water that is shown very close to saquinavir is labeled W.

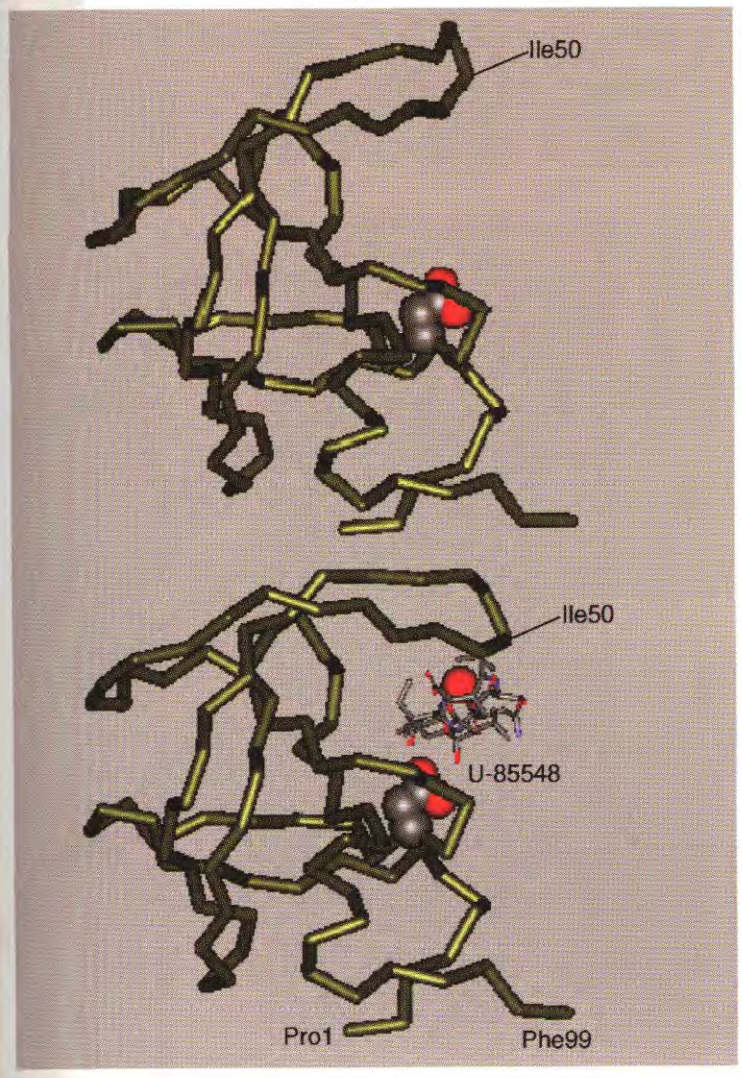


Figure 10.10. Comparison of the structures of **HIV-**P **apoenzyme** monomer (top, PDB code **3PHV**) and the wmplex between HN-P and (32) (U-85548; bottom, PDB wde **8HVP**). The inhibitor is shown as a ball and stick structure. Note the rearrangement of the flap residues; **Ile50** is indicated for reference. The van der Waals surface of Asp25 is shown in both structures. The flap water (red ball) is also shown between **Ile50** and U-85548. In the bottom structure, the locations of the N and **C** termini of HN-P are noted.

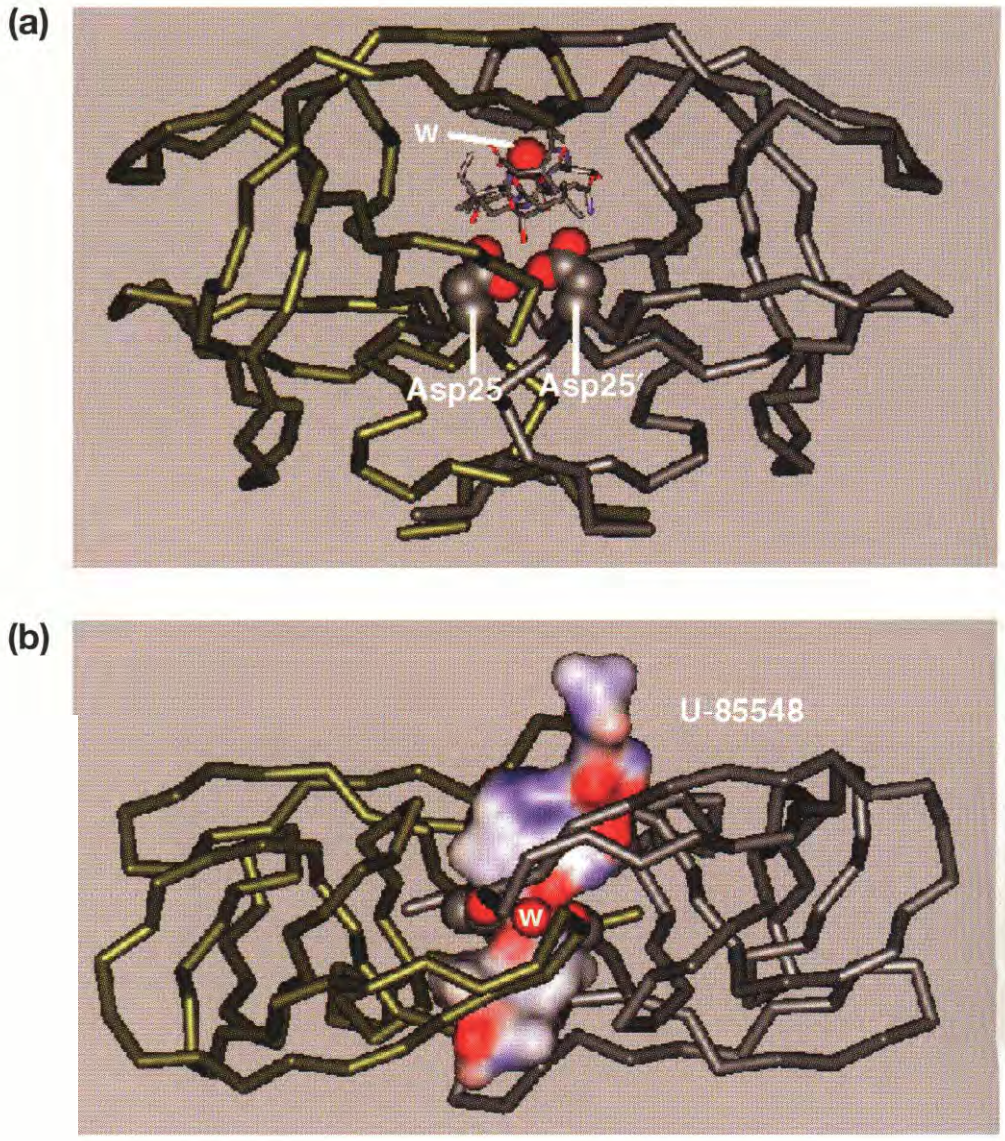


Figure 10.11. Orthogonal views of the complex between HIV-P and (32) (U-85548). The view in panel a is rotated approximately 90° (around the long axis of the protein) from the view in panel b. Van der Waals surfaces of Asp25, Asp25′, and the flap water (W) are shown. In panel b, the solvent-accessible surface of the inhibitor is shown.

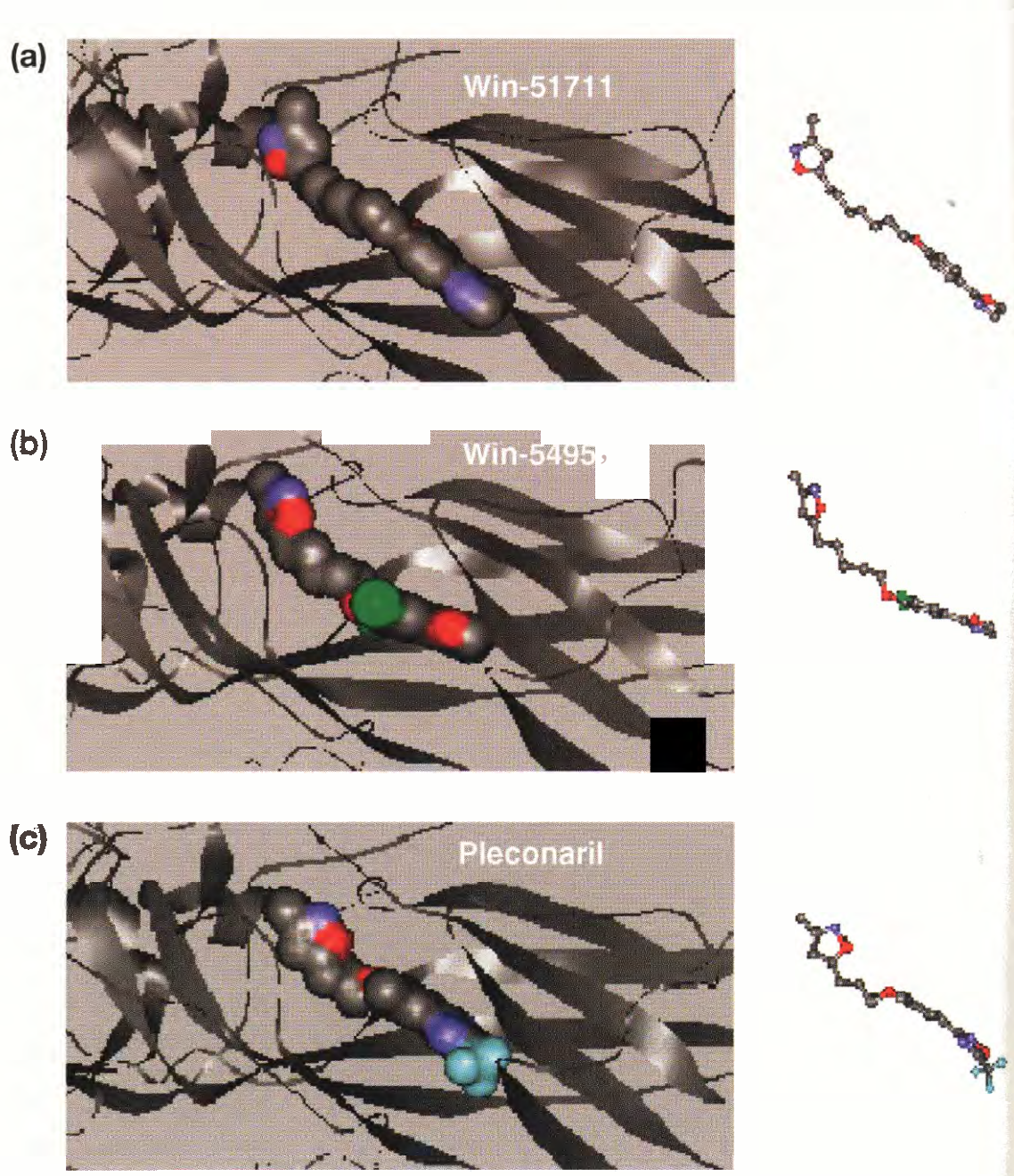


Figure 10.17. Structure of rhinovirus capsid protein VP1 showing the bound conformation of antiviral isoxazole compounds (78) [disoxaril, WIN-51711: panel a, top], (79) [WIN-54954: panel b, middle], and (80) [pleconaril, WIN-63843: panel c, bottom]. The PDB codes for the X-ray structural model coordinates used to create these views are: 1PIV (for **78**), 2HWE (for **79**), and 1C8M (for 80). On the left side of each panel, the inhibitors are shown as van der Waals surfaces, and the protein as a ribbon diagram. On the right side, the structures of the inhibitor alone are shown, from the same view, as ball and stick representations.

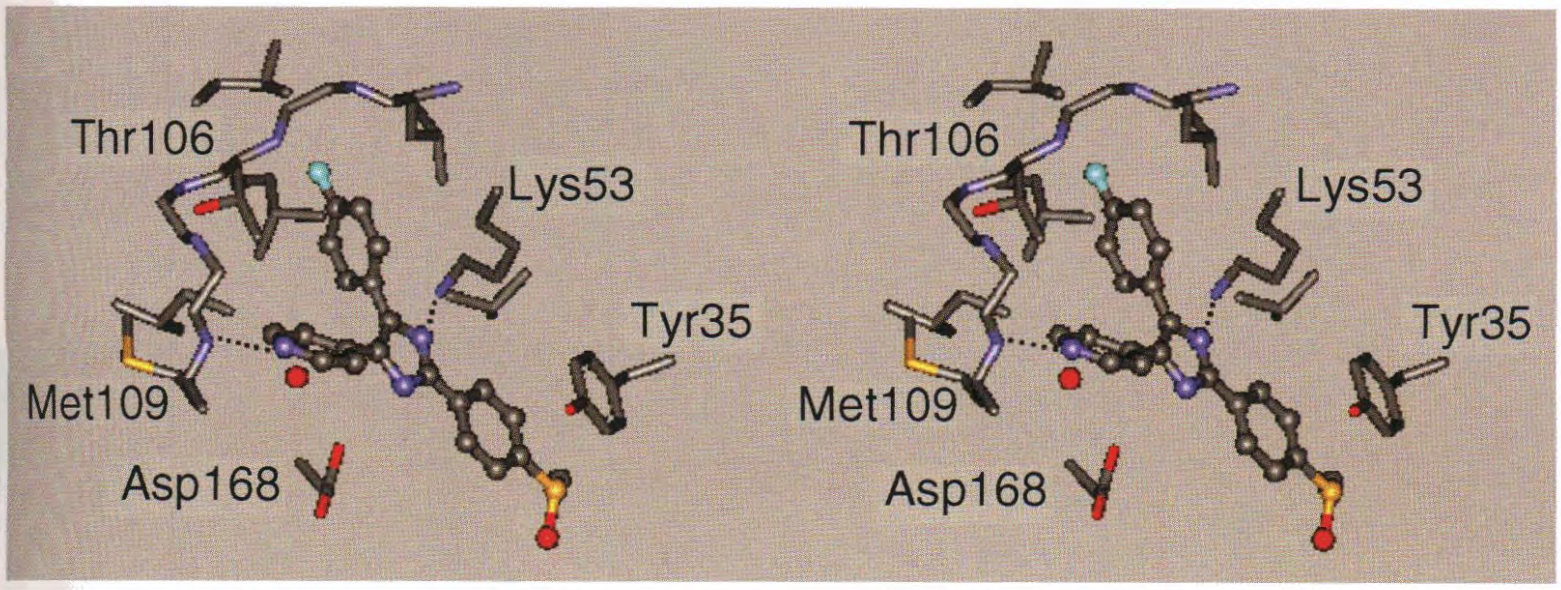


Figure 10.18. Binding of SB203580 (shown as a ball and stick structure) in the active site of MAPK p38α. In addition to the side chains of the labeled residues, the protein backbone between Leu104 and Met109 is shown, as well as several aliphatic side chains and a water molecule (red sphere). **Hydrogen bonds (dotted lines) are shown between the backbone amide of Met109 and the inhibitor's pyrimidinyl nitrogen, and between the ϵ-amino of Lys53 and the inhibitor's imidazole N3. This figure is based on the PDB coordinate set 1A9U (187).**

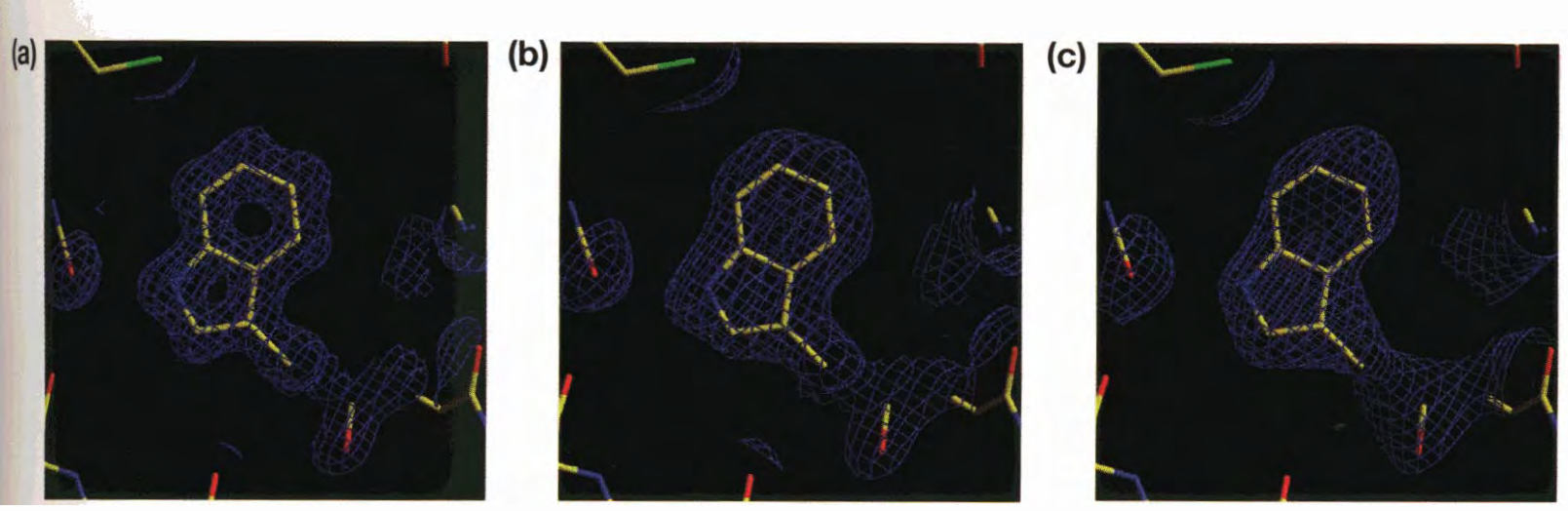


Figure 11.6. Three density maps at differing resolutions: a, 1.3 **A:** h, 2.1 **A:** *c.* 3.0 Å

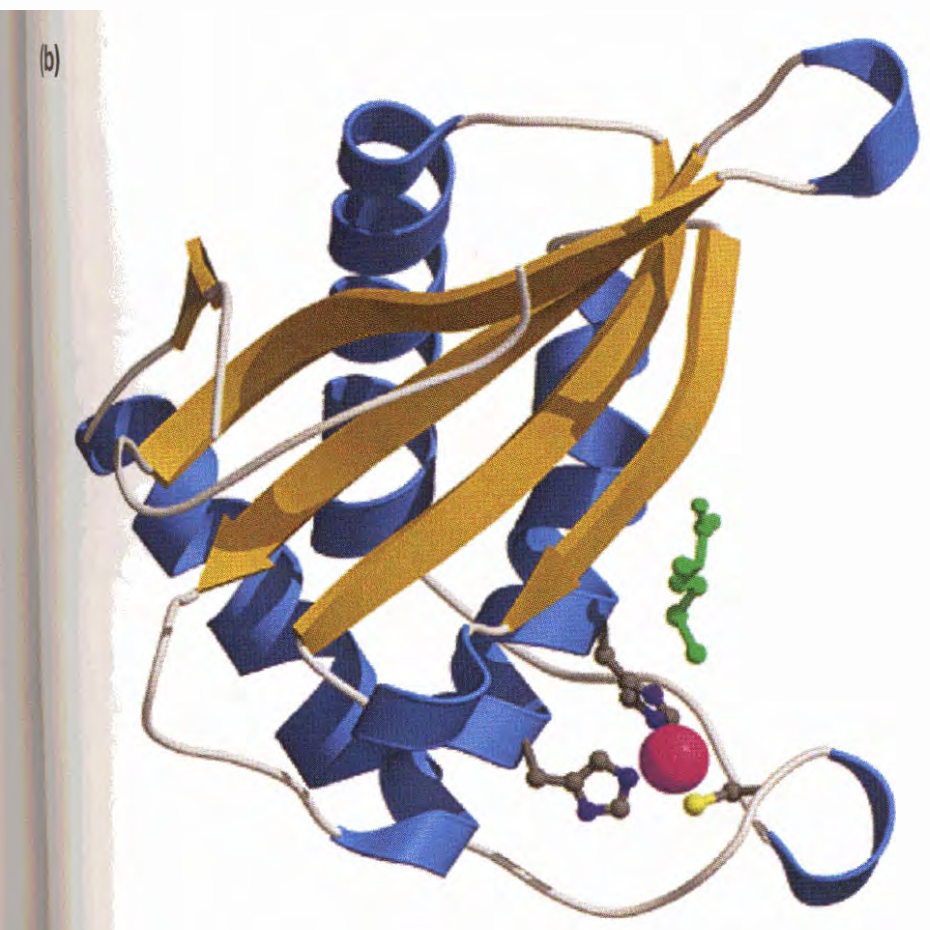


Figure 11.7. (b) **Structure of the LuxS monomer highlighting the bound zinc ion (magenta) and methionine (green).**

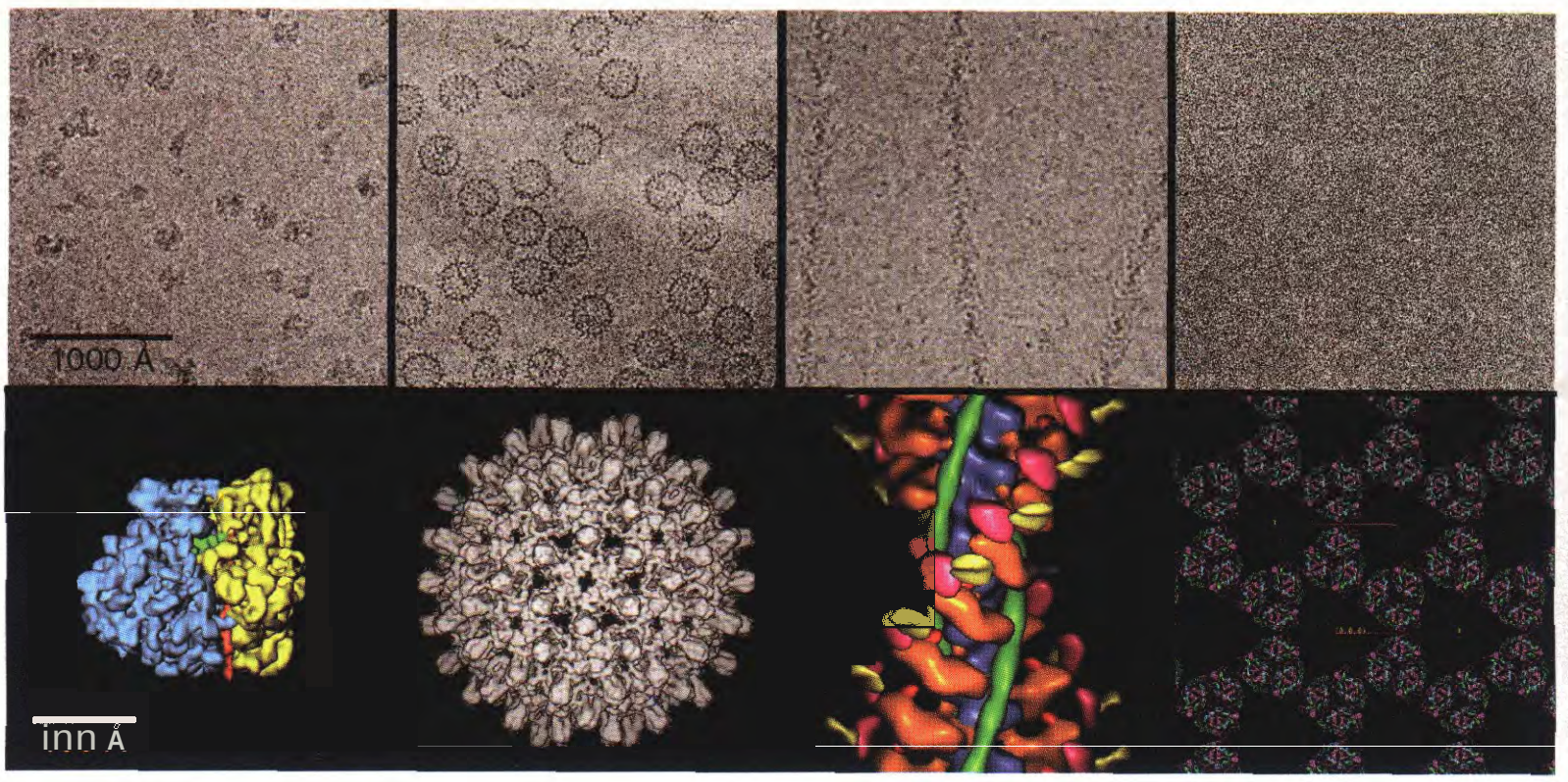


Figure 14.6. Examples of macromolecules studied by cryo-EM and 3D image reconstruction and the resulting 3D structures (bottom row) after cryo-EM analysis. All micrographs (top row) are displayed at above 170,000× magnification and all models at about 1,200,000× magnification. (a) A single particle without symmetry: The micrograph shows 70S E. coli ribosomes complexed with mRNA and fMet-tRNA. The surface-shaded density map, made by averaging 73,000 ribosome images from 287 micrographs has a resolution (FSC) of 11.5 Å. The 50S and 30S subunits and the tRNA are colored blue, yellow, and green, respectively. The identity of many of the subunits is known as some RNA double helices are clearly recognizable by their major and minor grooves (e.g., helix 44 is shown in red). [Courtesy of J. Frank (SUNY, Albany), using data from Gabashvili et al. (86).] (b) A single particle with symmetry: The micrograph shows hepatitis B virus cores. The 3D reconstruction, at a resolution of 7.4 Å (DPR), was computed from 6384 particle images taken from 34 micrographs. [From Böttcher et al. (44).] (c) A helical filament: The micrograph shows actin filaments decorated with myosin S1 heads containing the essential light chain. The 3D reconstruction, at a resolution of 30–35 Å, is a composite in which the differently colored parts are derived from a series of difference maps that were superimposed on f-actin. The components include: f-actin (blue), myosin heavy chain motor domain (orange), essential light chain (purple), regulatory light chain (white), tropomyosin (green), and myosin motor domain *N*-terminal beta-barrel (red). [Courtesy of A. Lin, M. Whittaker, and R. Milligan (Scripps Research Institute, LaJolla, CA).] (d) A 2D crystal, light-harvesting complex LHCII at 3.4-Å resolution. The model shows the protein backbone and the arrangement of chromophores in a number of trimeric subunits in the crystal lattice. In this example, image contrast is too low to see any hint of the structure without image processing (see also Fig. 14.3). [Courtesy of W. Kühlbrandt (Max-Planck-Institute for Biophysics, Frankfurt, Germany).]

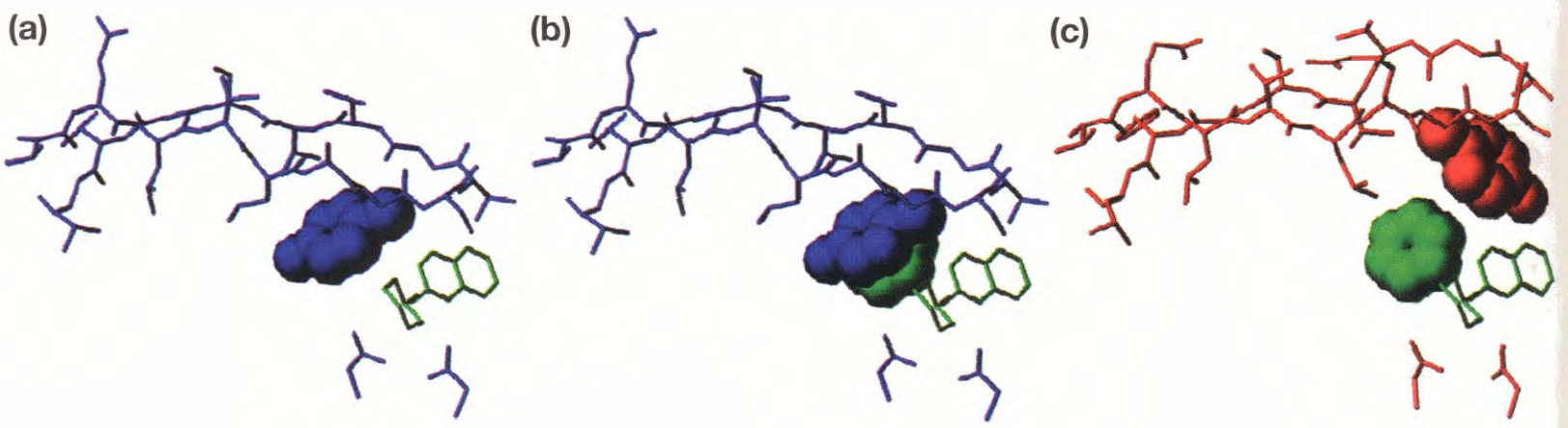


Figure 15.35. GRAB peptidomimetics in action.

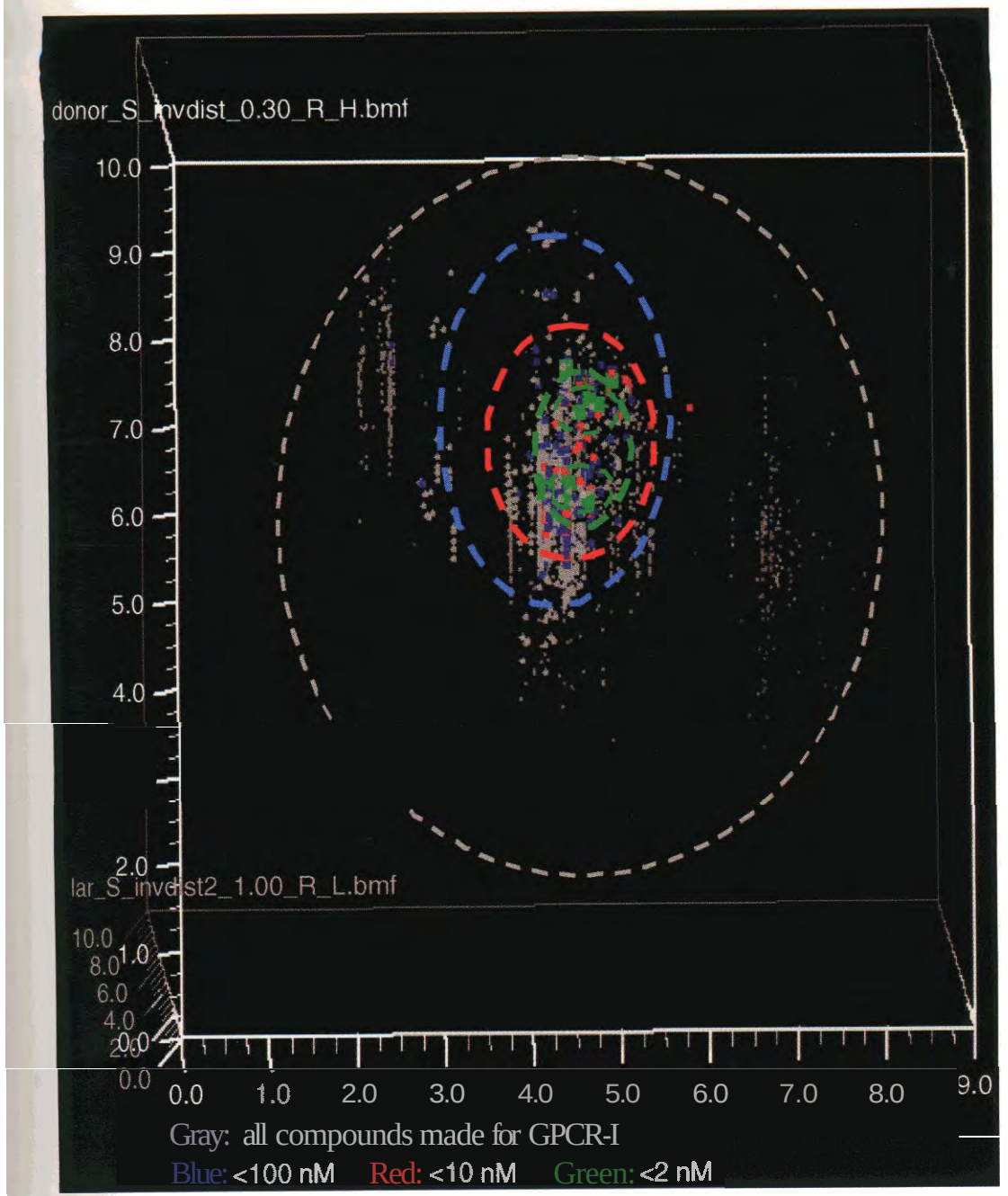


Figure 5.26. The same **3D subspace** as in Fig. 5.25, rotated slightly to provide a better viewing perspective. Points indicate coordinates of about 2000 combinatorial products selected from 14 different libraries. Color-coding indicates affinity for the GPCR-1 receptor. See Refs. **32e,i** and text for further details.

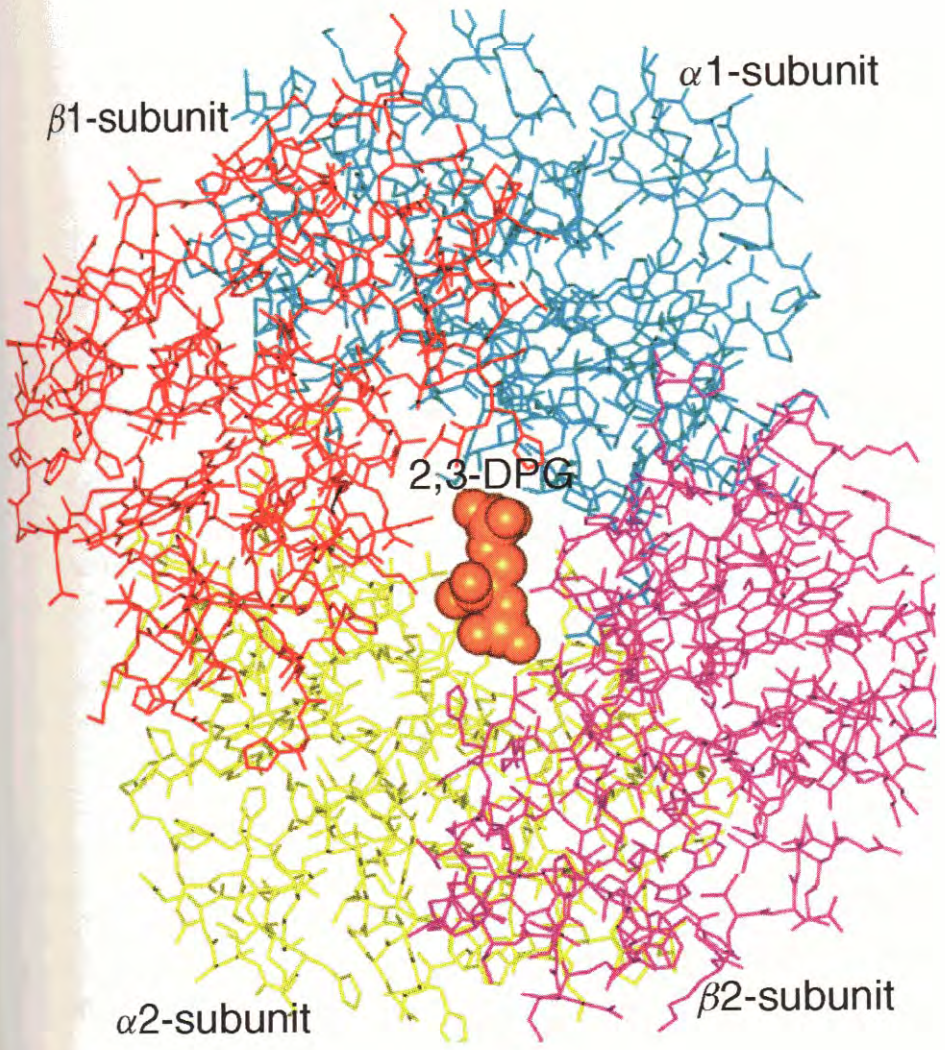


Figure 10.2. View of (4) (2,3-DPG) binding site at the mouth of the β-cleft of deoxy hemoglobin.

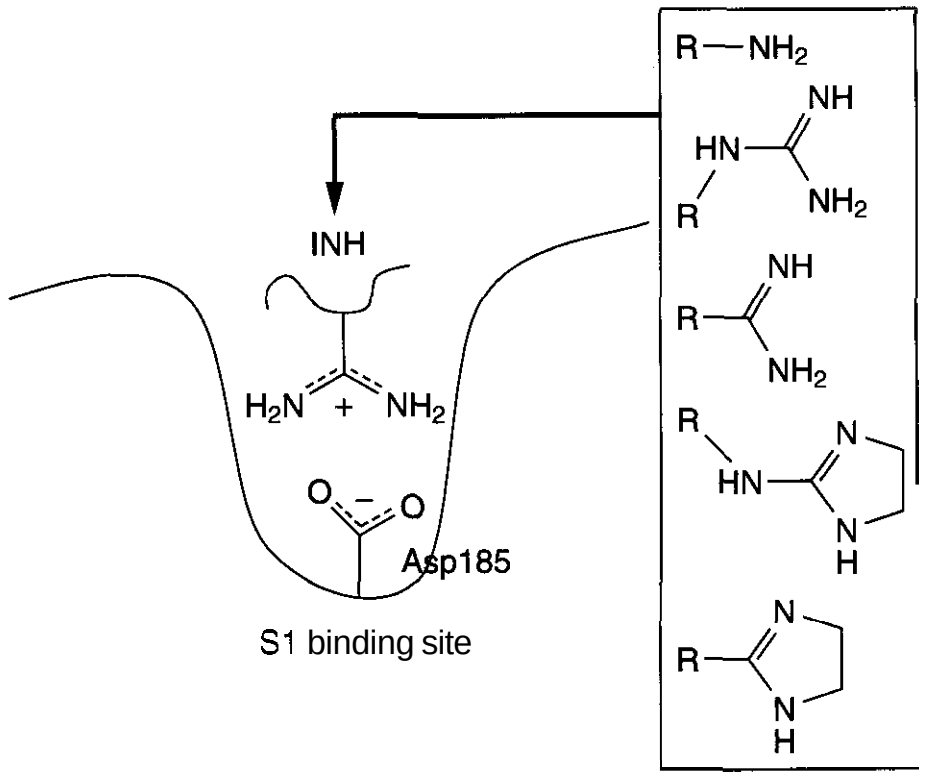


Figure 6.27. Potential functional groups in inhibitors necessary to block the specificity pocket (S1 binding site) of matriptase. The choice reflects the observation that matriptase prefers substrates with Lys or Arg as P1 residue.

ity needs to be included in high-throughput docking. Scoring functions have to improve to make consistently correct predictions of putative protein-ligand binding affinities. Scoring functions, calibrated to reproduce experimental data, have unreliable performance outside their training set. Thus, ***de novo*** methods using terms describing the thermodynamics of binding should replace the first generation of scoring functions. In consequence, some of the speed gained from low cost parallel computing should be invested into higher accuracy scoring rather than higher throughput. One way of increasing throughput is to keep the number of compounds docked as small as possible by using every bit of knowledge one has to prefilter the database, mainly based on pharmacophore information. In some cases this is easy. For instance the necessity of having certain features like salt bridges formed on ligand binding [e.g., in influenza virus neuraminidase (211)] or other prevalent information (e.g., hinge region binding for many ATP competitive kinase inhibitors) greatly helps to reduce the number of compounds subjected to docking experiments.

Table 6.4 Results Obtained from the Initial Screening of Compounds against Matriptase (206)[a]

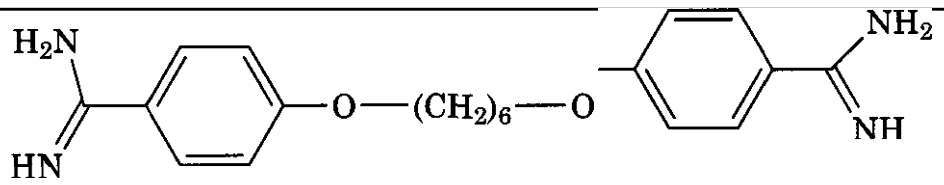
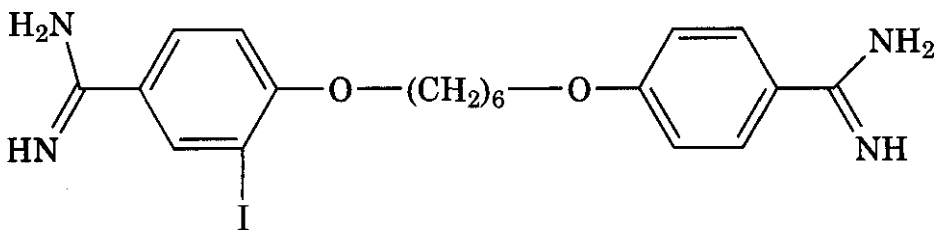
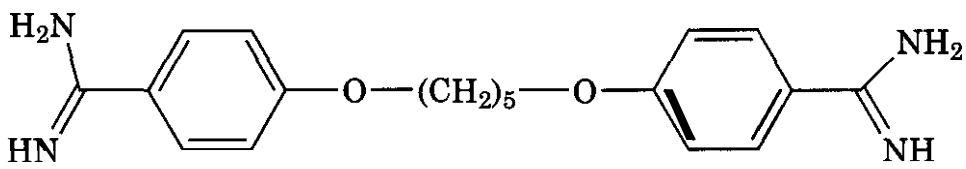
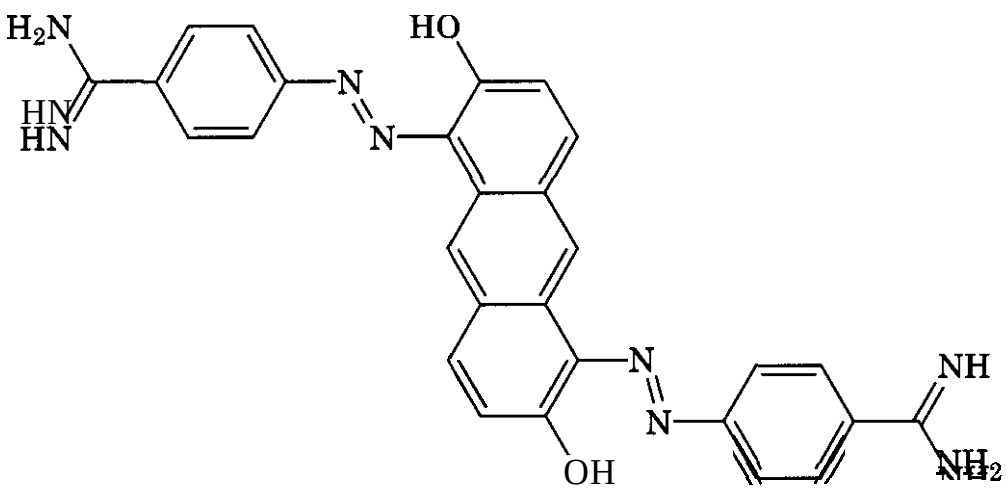
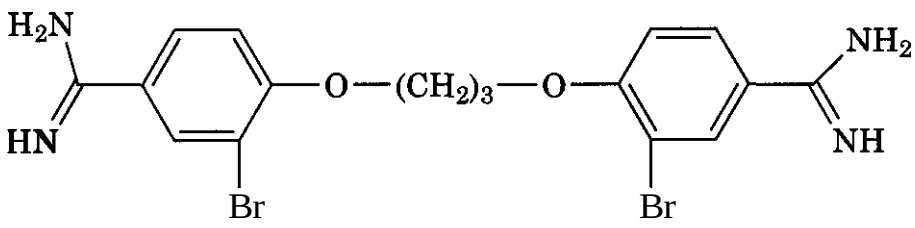
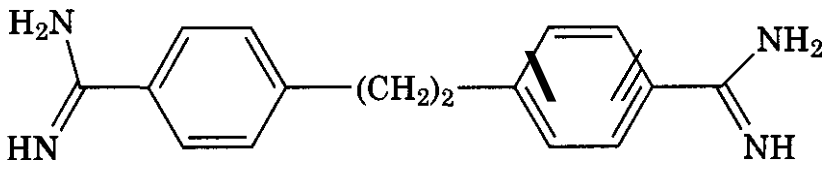
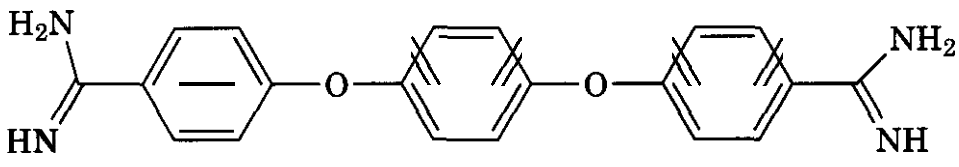
% Inhibition	Number of Compounds
Over 95%	15
90–94%	4
70–89%	15
40–69%	13
Below 39%	17
High absorbency	3
Increased activity	**3**

[a]Testing was done at 75 μM compound and protein concentration. The ratio between compound and protein molar concentration was 1:1.

The missing robustness of many structure-based docking/scoring techniques opens the questions of when should one apply it and when should one retreat to pharmacophore-based virtual screening. In many cases it makes sense to prescreen virtual libraries using pharmacophore techniques, particularly if one uses shape representations of the receptor site, such as volume-exclusion spheres, a pharmacophore search can be a very effective prefilter. Also, in cases where receptor-site flexibility is problematic, pharmacophore searching may be less restrictive (unless one tries to deal with protein flexibility in the docking routine—a task that is not easy, usually not applied today, and another future direction of development in virtual screening).

The above tools and pathways show a simple and inexpensive way of discovering novel lead chemical matter for drug discovery programs. However, there are many hurdles to overcome to make virtual screening successful. The properties of druglikeness may not be understood sufficiently enough, resulting in poor pharmacokinetics of the compounds: ex-

Table 6.5 K_i Values Obtained for Tested bis-Benzamidines against Matriptase

Compound	Structure	K_i (nM)
1		924
2		191
3		1,160
4		4,500
5		535
6		>10,000
-		208

isting SARs that lead to the generation of pharmacophore models may bias the pharmacophore toward a narrow segment of compounds; structural information of the target protein is often not available; and homology models may not be precise enough. Current scoring functions are often not robust enough to separate actives from inactives. Compounds identified may not be easy to synthesize. Hits may not be selective or patentable.

On one hand, there are obviously many risks involved in virtual screening, many assumptions made, and a positive outcome not at all guaranteed in each and every case. On the other hand, however, the overall process is extremely cost effective and fast. Even if successful in only a few cases, virtual screening can produce leads that may otherwise not have surfaced and so add immense value to a drug discovery program. Especially in cases

where high throughput screening cannot identify a viable lead chemical matter, virtual screening applied to vendor databases or combinatorial libraries to be synthesized presents a cost-effective alternative. Mainly because of its speed, cost effectiveness, ease of setup, and increasing robustness, we expect virtual screening to become a mainstream approach throughout the pharmaceutical industry.

5 ACKNOWLEDGMENTS

The authors thank Dr. Matthias Rarey for valuable discussions.

REFERENCES

1. D. J. Abraham, *Intra-Science Chem. Rept.*, 8, 1 (1974).
2. M. Perutz, *Protein Structure. New Approaches to Disease and Therapy*, Freeman, New York, 1992.
3. S. W. Fesik, *J. Med. Chem.*, 34,2937 (1991).
4. D. W. Cushman, H. S. Cheung, E. F. Sabo, and M. A. Ondetti, *Biochemistry*, 16,5484 (1977).
5. L. F. Kuyper, B. Roth, D. P. Baccanari, R. Ferone, C. R. Bedell, J. N. Champness, D. K. Stammers, J. G. Dann, F. E. Norrington, D. J. Blaker, and P. J. Goodford, *J. Med. Chem.*, 25, 1120 (1982).
6. M. A. Gallop, R. W. Barrett, W. J. Dower, S. P. A. Fodor, and E. M. Gordon, *J. Med. Chem.*, 37,1233 (1994).
7. E. M. Gordon, R. W. Barrett, W. J. Dower, S. P. A. Fodor, and M. A. Gallop, *J. Med. Chem.*, 37, 1385 (1994).
8. M. W. Lutz, J. A. Menius, T. D. Choi, R. G. Laskody, P. L. Domanico, A. S. Goetz, and D. L. Saussy, *Drug Discovery Today*, 1,277 (1996).
9. H. J. Bohm and G. Klebe, *Angew. Chem. Int. Ed. Engl.*, 35,2589 (1996).
10. W. P. Walters, M. T. Stahl, and M. A. Murcko, *Drug Discovery Today*, 3,160 (1998).
11. Y. C. Martin, *Perspect. Drug Discovery Des.*, 7/8, 159 (1997).
12. A. C. Good and J. S. Mason in K. B. Lipkowitz and D. B. Boyd, Eds., *Reviews in Computational Chemistry*, Vol. 7, VCH, New York, 1995, p. 67.
13. G. Klebe, *Virtual Screening: An Alternative or Complement to High Throughput Screening?*, Kluwer/Escom, Leiden, 2000.
14. H. J. Bohm and G. Schneider, *Virtual Screening for Bioactive Molecules*, Wiley-VCH, Weinheim, 2000.
15. A. C. Good, S. R. Krystek, and J. S. Mason, *Drug Discovery Today*, 5 (Suppl.), S61 (2000).
16. T. Langer and R. D. Hoffmann, *Curr. Pharm. Design*, 7, 509 (2001).
17. B. Waszkowycz, T. D. J. Perkins, R. A. Sykes, and J. Li, *IBM Systems J.*,40,360 (2001).
18. A. Good, *Curr. Opin. Drug Discovery Dev.*, 4, 301 (2001).
19. R. F. Burns, R. M. A. Simmons, J. J. Howbert, D. C. Waters, P. G. Threlkeld, and B. D. Gitter, *Exploiting Molecular Diversity, Symposium Proceedings*, Vol. 2, Cambridge Healthtech Institute, San Diego, 1995, p. 2.
20. W. P. Walters, Ajay, and M. A. Murcko, *Curr. Opin. Chem. Biol.*, 3,384 (1999).
21. W. P. Walters and M. A. Murcko, *Methods Principles Med. Chem.*, 10, 15 (2000).
22. D. E. Clark and S. D. Pickett, *Drug Discovery Today*, 5, 49 (2000).
23. B. L. Podlogar, I. Muegge, and L. J. Brice, *Curr. Opin. Drug Discovery Dev.*, 4, 102 (2001).
24. I. Muegge, *Chem. Eur. J.*,8, 1976 (2002).
25. Ajay, W. P. Walters, and M. A. Murcko, *J. Med. Chem.*, 41,3314 (1998).
26. Ajay, G. W. Bemis, and M. A. Murcko, *J. Med. Chem.*, 42,4942 (1999).
27. J. Sadowski and H. Kubinyi, *J. Med. Chem.*, 41,3325 (1998).
28. M. Wagener and V. J. vanGeerestein, *J. Chem. Inf. Comput. Sci.*, 40,280 (2000).
29. V. J. Gillet, P. Willett, and J. Bradshaw, *J. Chem. Inf. Comput. Sci.*, 38,165 (1998).
30. Comprehensive Medicinal Chemistry is available from MDL Information Systems Inc., San Leandro, CA 94577 and contains drugs already on the market.
31. World Drug Index is available from Derwent Information, London, UK. Website: www.derwent.com.
32. MACCS-II Drug Data Report is available from MDL Information Systems Inc., San Leandro, CA 94577 and contains biologically active compounds in the early stages of drug development.
33. C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, *Adv. Drug Delivery Rev.*, 23, 3 (1997).
34. T. I. Oprea, *J. Cornput.-Aided Mol. Des.*, 14, 251 (2000).

J. Kelder, P. D. J. Grootenhuis, D. M. Bayada, L. P. C. Delbressine, and J. P. Ploemen, Pharm. Res., **16, 1514** (1999).

M. Hann, B. Hudson, X. Lewell, R. Lifely, L. Miller, and N. Ramsden, J. Chem. *Inf.* Comput. Sci., **39,897 (1999).**

National Toxicology Program. http://ntp-server.niehs.nih.gov.

RTECS C2(96–4); National Institute for Occupational Safety and Health (NIOSH), U.S. Department of Health and Human Services: Washington, DC, **1996.** URL: http://www.ccohs.ca.

G. Klopman and H. S. Rosenkranz, *Mutat.* Res., **305, 33** (1994).

K. Enslein, V. K. Gombar, and B. W. Blake, *Mutat.* Res., **305, 47 (1994).**

Lhasa Ltd., School of Chemistry, University of Leeds, Leeds, UK. URL: http://www.chem.leeds.ac.uk/LUK/derek/index.html.

Available Chemicals Directory is available from MDL Information Systems Inc., San Leandro, CA **94577** and contains specialty bulk chemicals from commercial sources. Website: http://www.mdli.com.

V. N. Viswanadhan, A. K. Ghose, G. R. Revankar, and R. K. Robins, *J.* Chem. *Inf. Comp.* Sci., **29,163 (1989).**

G. W. Bemis and M. A. Murcko, J. Med. Chem., **39,2887 (1996).**

G. W. Bemis and M. A. Murcko, J. Med. Chem., **42,5095** (1999).

X. Q. Lewell, D. B. Judd, S. P. Watson, and M. M. Hann, J. Chem. *Inf.* Comput. Sci., **38, 511 (1998).**

I. Muegge, D. Brittelli, and S. L. Heald, J. Med. Chem., **44, 1841 (2001).**

B. L. Podlogar and I. Muegge, Curr. Top. Med. Chem., **1,257 (2001).**

R. M. Brunne, G. Hessler, and I. Muegge in K. C. Nicolaou, R. Hanko, and W. Hartwig, Eds., Handbook of Combinatorial Chemistry, Vol. **2,** Wiley-VCH, Weinheim, **2002,** p. **761.**

B. E. Evans, K. E. Rittle, M. G. Bock, R. M. DiPardo, R. M. Freidinger, W. L. Whitter, G. F. Lundell, D. F. Veber, P. S. Anderson, R. S. L. Chang, V. J. Lotti, D. J. Cerino, T. B. Chen, P. J. Kling, K. A. Kunkel, J. P. Springer, and J. Hirshfield, J. Med. Chem., **31, 2235** (1988).

J. S. Mason, I. Morize, P. R. Menard, D. L. Cheney, C. Hulme, and R. F. Labaudiniere, *J.* Med. Chem., **42,3251 (1999).**

P. J. Hajduk, M. Bures, J. Praestgaard, and S. W. Fesik, J. Med. Chem., **43,3443** (2000).

53. J. H. Van Drie, D. Weininger, and Y. C. Martin, J. Cornput.-Aided Mol. Des., **3,225 (1989).**

54. E. Fischer, Ber. Dtsch. Chem. *Ges.*, **27, 2985 (1894).**

55. P. Ehrlich, Ber. Dtsch. Chem. Ges., **42, 17 (1909).**

56. P. Gund, Prog. Mol. *Subcell.* Biol., **5, 117 (1977).**

57. F. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Schimanouchi, and M. J. Tasumi, J. Mol. Biol., **112, 535 (1977).**

58. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Baht, H. Weissig, I. N. Shindyalov, and P. E. Bourne, Nucleic Acids Res., **28, 235 (2000).**

59. M. K. Dreyer, D. R. Borcherding, J. A. Dumont, N. P. Peet, J. T. Tsay, P. S. Wright, A. J. Bitonti, J. Shen, and S.-H. Kim, J. Med. Chem., **44, 524 (2001).**

60. G. R. Marshall, C. D. Barry, H. E. Bosshard, R. A. Dammkoehler, and D. A. Dunn in E. C. Olson and R. E. Christoffersen, Eds., Computer Assisted Drug Design, American Chemical Society, Washington, **1979,** p. **205.**

61. J. R. Sufrin, D. A. Dunn, and G. R. Marshall, Mol. Pharmacol., **19,307 (1981).**

62. R. D. Cramer, D. E. Patterson, R. D. Clark, F. Soltanshahi, and M. S. Lawless, J. Chem. *Inf.* Comput. Sci., **38,** 1010 (1998).

63. D. Horvath in A. K. Ghose and V. N. Viswanadhan, Eds., Combinatorial Library Design and Evaluation. Principles, Software Tools, and Applications in Drug Discovery, Marcel Dekker, New York, **2001,** p. **429.**

64. A. Dalby, J. G. Nourse, W. D. Hounshell, A. K. I. Gushurst, D. L. Grier, B. A. Leland, and J. Laufer, J. Chem. *Inf.* Comput. Sci., **32, 244 (1992).**

65. Spresi Chemical Database, InfoChem GMBH, Grobenzell, Germany and Daylight Chemical Information Systems, Irvine, CA (2002).

66. The Chemical Abstracts Database, Chemical Abstracts Service, 2540 Olentangy River Road, PO Box **3012,** Columbus, OH **(2002).**

67. G. W. A. Milne, M. C. Nicklaus, J. S. Driscoll, S. Wang, and D. W. Zaharevitz, J. Chem. *Inf.* Comput. Sci., **34, 1219 (1994).**

68. G. W. A. Milne and J. A. Miller, J. Chem. *Inf.* Comput. Sci., **26, 154 (1986).**

69. D. Weininger, J. Chem. *Inf.* Comput. Sci., **28, 31 (1988).**

70. D. Weininger, A. Weininger, and J. L. Weininger, J. Chem. *Inf.* Comput. Sci., **29, 97 (1989).**

71. F. H. Allen, S. Bellard, M. D. Brice, B. A. Cartwright, A. Doubleday, H. Higgs, T. Hummelink-Peters, O. Kenard, W. D. S. Motherwell, J. R. Rodgers, and D. G. Watson, *Acta Crystallogr. B,* **35,2331 (1979).**

72. R. L. DesJarlais in P. S. Charifson, Ed., ***Practical Application of Computer-Aided Drug Design,*** Vol. **1,** Marcel Dekker, New York, **1997,** p. 73.

73. A. RusinkoIII, J. M. Skell, R. Balducci, C. M. McGarity, and R. S. Pearlman, CONCORD, University of Texas, Austin, TX and Tripos Associates, St. Louis, MO **(1988).**

74. J. Gasteiger, C. Rudolph, and J. Sadowski, ***Tetrahedron Comput. Methodol.,* 3, 537 (1990).**

75. D. Weininger, Rubicon, Daylight Chemical Information Systems, Irvine, CA (1995).

76. Omega, Open Eye, Santa Fe, NM **(2002).**

77. D. P. Dolata, A. R. Leach, and K. Prout, *J. Cornput.-Aided Mol. Des., 1,* **73** (1987).

78. A. R. Leach and K. Prout, ***J. Comput. Chem.,* 11,1193 (1990).**

79. W. T. Wipke and M. A. Hahn, *Artif. Intell. Appl. Chem.,* **306,136** (1986).

80. G. Klebe and T. Mietzner, ***J. Cornput.-Aided Mol. Des.,* 8,583** (1994).

81. E. V. Gordeeva, A. R. Katritzky, V. V. Shcherbukhin, and N. S. Zifirov, ***J. Chem. Inf. Comput. Sci.,* 33, 102 (1993).**

82. S. K. Kearsley, D. J. Underwood, R. P. Sheridan, and M. D. Miller, *J. Cornput.-Aided Mol. Des.,* **8, 153** (1994).

83. J. M. Barnard, ***J. Chem. Inf. Comput. Sci.,* 33, 532 (1993).**

84. I. J. Enyedy, J. Wang, W. A. Zaman, K. M. Johnson, and S. Wang, ***Bioorg. Med. Chem. Lett.,* 12, 1775 (2002).**

85. G. M. Downs and P. Willett in K. B. Lipkowitz and D. B. Boyd, Eds., ***Reviews in Computational Chemistry,*** VCH, New York, **1996,** p. **1.**

86. P. Willett, J. M. Barnard, and G. M. Downs, ***J. Chem. Inf. Comput. Sci.,* 38,983 (1998).**

87. L. Xue, J. W. Godden, and J. Bajorath, *J. Chem. Inf. Comput. Sci.,* **39,881** (1999).

88. L. Xue, J. W. Godden, and J. Bajorath, ***J. Chem. Inf. Comput. Sci.,* 40, 1227 (2000).**

89. L. Xue, F. L. Stahura, J. W. Godden, and J. Bajorath, ***J. Chem. Inf. Comput. Sci.,* 41, 394** (2001).

90. S. K. Kearsley, S. Sallamack, E. M. Fluder, J. D. Andose, R. T. Mosley, and R. P. Sheridan, ***J. Chem. Inf. Comput. Sci.,* 36, 118 (1996).**

91. R. D. Hull, E. M. Fluder, S. B. Singh, R. B. Nachbar, S. K. Kearsley, and R. P. Sheridan, ***J. Med. Chem.,* 44, 1185 (2001).**

92. R. D. Hull, S. B. Singh, R. B. Nachbar, R. P. Sheridan, S. K. Kearsley, and E. M. Fluder, ***J. Med. Chem.,* 44, 1177 (2001).**

93. B. T. Hoffman, T. Kopajtic, J. L. Katz, and A. H. Newman, ***J. Med. Chem.,* 43, 4151 (2000).**

94. A. P. Kozikowski, M. K. E. Saiah, K. M. Johnson, and J. S. Bergmann, ***J. Med. Chem.,* 38, 3086 (1995).**

95. S. Wang, S. Sakamuri, I. J. Enyedy, A. P. Kozikowski, O. Deschaux, B. C. Bandyopadhyay, S. R. Tella, W. A. Zaman, and K. M. Johnson, ***J. Med. Chem.,* 43,351 (2000).**

96. D. Barnum, J. Greene, A. Smellie, and P. Sprague, ***J. Chem. Inf. Comput. Sci.,* 36, 563 (1996).**

97. P. W. Sprague in K. Muller, Ed., ***Perspectives in Drug Discovery and Design,*** Vol. **1,** ESCOM Science Publishers B.V., Leiden, **1995,** p. **1.**

98. Y. C. Martin, M. G. Bures, E. A. Danaher, J. DeLazzer, I. Lico, and P. A. Pavlik, ***J. Comput.-Aided Mol. Des.,* 7, 83 (1993).**

99. D. Beusen, Alignment of angiotensin II receptor antagonists using GASP, Tripos Technical Notes 1(4), November **1996.**

100. Flo96, Thistlesoft, Colebrook, CT **(1996).**

101. V. Golender, B. Vesterman, O. Eliyahu, A. Kardash, M. Kletzin, M. Shokhen, and E. Vorpagel, Conference Proceedings Barcelona 10th European Symposium on Structure-Activity Relationships, Vol. **10,** Prous Science, Barcelona, Spain, **1994,** p. **246.**

102. ROCS, Open Eye, Santa Fe, NM **(2002).**

103. O. F. Guner, ***Pharmacophore Perception, Development, and Use in Drug Design,*** International University Line, La Jolla, CA, **2000.**

104. H.-J. Bohm, *J. Cornput.-Aided Mol. Des., 6,* **593 (1992).**

105. H. Bohm, ***J. Cornput.-Aided Mol. Des.,* 6, 61 (1992).**

106. H. Bohm, ***J. Cornput.-Aided Mol. Des.,* 8, 623 (1994).**

107. R. Wang, Y. Gao, and L. Lai, ***J. Mol. Model.,* 6, 498 (2000).**

108. Cerius2, Accelrys, San Diego, CA (2002).

109. S. K. Burley and G. A. Petsko, *J. Am. Chem. Soc.,* **108, 7995** (1986).

110. H. A. Carlson, K. M. Masukawa, K. Rubins, F. D. Bushman, W. L. Jorgensen, R. D. Lins,

J. M. Briggs, and J. A. McCammon, *J. Med. Chem.,* **43,2100 (2000).**

R. P. Sheridan, R. Nilakantan, A. Rusinko III, N. Bauman, K. S. Haraki, and R. Venkataraghavan, *J. Chem. Inf. Comput. Sci.,* **29,255 (1989).**

UNITY version **6.6;** Tripos Associates, St. Louis, MO. Website: http://www.tripos.com.

MDL Information Systems, Inc., San Leandro, CA.

J. Greene, S. Kahn, H. Savoj, P. Sprague, and S. Teig, *J. Chem. Inf. Comput. Sci.,* **34, 1297 (1994).**

W.-D. Ihlenfeldt, J. H. Voigt, B. Bienfait, F. Oellien, and M. C. Nicklaus, *J. Chem. Inf. Comput. Sci.,* **42, 46 (2002).**

X. Fang and S. Wang, *J. Chem. Inf. Comput. Sci.,* **42, 192 (2002).**

K. S. Haraki, R. P. Sheridan, R. Venkataraghavan, D. A. Dunn, and R. McCulloch, *Tetrahedron Comput. Methodol.,* **3,565 (1990).**

J. A. Grant, M. A. Gallardo, and B. T. Pickup, *J. Comput. Chem.,* **17, 1653 (1996).**

T. E. Mook, D. R. Henry, A. G. Ozkabak, and M. Alamgir, *J. Chem. Inf. Comput. Sci.,* **34, 184 (1994).**

T. Hurst, *J. Chem. Inf. Comput. Sci.,* **34, 190 (1994).**

N. W. Murrall and E. K. Davies, *J. Chem. Inf. Comput. Sci.,* **30,312 (1990).**

D. E. Clark, P. Willett, and P. W. Kenny, *J. Mol. Graph.,* **11, 146 (1993).**

D. E. Clark, G. Jones, and P. Willett, *J. Chem. Inf. Comput. Sci.,* **34, 197 (1994).**

J. M. Blaney and J. S. Dixon, *Perspect. Drug Discovery Des.,* **15, 301 (1993).**

I. Muegge and M. Rarey in D. B. Boyd and K. B. Lipkowitz, Eds., *Reviews in Computational Chemistry,* Vol. **17,** Wiley-VCH, New York, **2001,** p. **1.**

G. Jones, P. Willett, R. C. Glen, and A. R. Leach, *J. Mol. Biol.,* **267, 727 (1997).**

I. Muegge, *Med. Chem. Res.,* **9,490 (1999).**

D. E. Koshland, *Proc. Natl. Acad. Sci. USA,* **44, 98 (1958).**

D. W. Christianson and W. N. Lipscomb, *Acc. Chem. Res.,* **22, 62 (1989).**

C. Sander and R. Schneider, *Proteins: Struct., Funct., Genet.,* **9, 56 (1991).**

T. L. Blundell, B. L. Sibanda, M. J. E. Sternberg, and J. M. Thornton, *Nature,* **326, 347 (1987).**

132. A. Vedani, P. Zbinden, J. P. Snyder, and P. A. Greenidge, *J. Am. Chem. Soc.,* **117, 4987 (1995).**

133. M. C. Nicklaus, S. Wang, J. S. Driscoll, and G. W. A. Milne, *Bioorg. Med. Chem.,* **3, 411 (1995).**

134. H. A. Carlson and J. A. McCammon, *Mol. Pharmacol.,* **57,213 (2000).**

135. M. A. Knegtel, I. D. Kuntz, and C. M. Oshiro, *J. Mol. Biol.,* **266,424 (1997).**

136. E. A. Sudbeck, C. Mao, R. Vig, T. K. Venkatachalam, L. Tuel-Ahlgren, and F. M. Uckun, *Antimicrob. Agents Chemother.,* **42, 3225 (1998).**

137. C. Bron and J. Kerbosch, *Commun. ACM,* **16, 575 (1973).**

138. Y. Lamdan and H. J. Wolfson in Proceedings of the IEEE International Conference on Computer Vision, **1988,** p. **238.**

139. F. S. Kuhl, G. M. Crippen, and D. K. Friesen, *J. Comput. Chem.,* **5, 24** (1984).

140. F. M. Richards, *Annu. Rev. Biophys. Bioeng.,* **6, 151 (1977).**

141. M. L. Connolly, *J. Appl. Crystallogr.,* **16, 548 (1983).**

142. M. L. Connolly, *J. Appl. Crystallogr.,* **18, 499 (1985).**

143. B. K. Shoichet, R. M. Stroud, D. V. Santi, I. D. Kuntz, and K. M. Perry, *Science,* **259, 1445 (1993).**

144. B. K. Shoichet, D. L. Bodian, and I. D. Kuntz, *J. Comput. Chem.,* **13,380 (1992).**

145. E. C. Meng, D. A. Gschwend, J. M. Blaney, and I. D. Kuntz, *Proteins: Struct., Funct., Genet.,* **17,266 (1993).**

146. T. J. A. Ewing and I. D. Kuntz, *J. Comput. Chem.,* **18,1175 (1997).**

147. E. C. Meng, B. K. Shoichet, and I. D. Kuntz, *J. Comput. Chem.,* **13,505 (1992).**

148. E. C. Meng, I. D. Kuntz, D. J. Abraham, and G. E. Kellogg, *J. Cornput.-Aided Mol. Des.,* **8, 299 (1994).**

149. D. A. Gschwend and I. D. Kuntz, *J. Comput.-Aided Mol. Des.,* **10, 123 (1996).**

150. B. K. Shoichet, A. R. Leach, and I. D. Kuntz, *Proteins: Struct., Funct., Genet.,* **34, 4 (1999).**

151. X. Q. Zou, Y. X. Sun, and I. D. Kuntz, *J. Am. Chem. Soc.,* **121,8033 (1999).**

152. M. D. Miller, S. K. Kearsley, D. J. Underwood, and R. P. Sheridan, *J. Cornput.-Aided Mol. Des.,* **8, 153 (1994).**

153. T. F. Havel, I. D. Kuntz, and G. M. Crippen, *Bull. Math. Biol.,* **45, 665 (1983).**

154. S. Kirkpatrik, C. D. J. Gelatt, and M. P. Vecchi, *Science*, 220,671 (1983).

155. D. S. Goodsell and A. J. Olson, *Proteins: Struct., Funct., Genet.*, 8, 195 (1990).

156. T. P. Lybrand in D. B. Boyd and K. B. Lipkowitz, Eds., *Reviews in Comptational Chemistry*, Vol. 1, VCH, New York, 1990, p. 295.

157. J. A. Given and M. K. Gilson, *Proteins: Struct., Funct., Genet.*, 33,475 (1998).

158. T. N. Hart and R. J. Read, *Proteins: Struct., Funct., Genet.*, 13, 206 (1992).

159. C. McMartin and R. S. Bohacek, J. *Comput.-Aided Mol. Des.*, 11, 333 (1997).

160. A. Wallqvist and D. G. Covell, *Proteins: Struct., Funct., Genet.*, 25, 403 (1996).

161. R. Abagyan, M. Totrov, and D. Kuznetsov, J. *Comput. Chem.*, 15,488 (1994).

162. J. Apostolakis, A. Pluckthun, and A. Caflisch, J. *Comput. Chem.*, 19, 21 (1998).

163. P. S. Charifson, J. J. Corkery, M. A. Murcko, and W. P. Walters, J. *Med. Chem.*, 42, 5100 (1999).

164. C. A. Baxter, C. W. Murray, D. E. Clark, D. R. Westhead, and M. D. Eldridge, *Proteins: Struct., Funct., Genet.*, 33, 367 (1998).

165. J. Wang, P. A. Kollman, and I. D. Kuntz, *Proteins: Struct., Funct., Genet.*, 36, 1 (1999).

166. D. Hoffmann, B. Kramer, T. Washio, T. Steinmetzer, M. Rarey, and T. Lengauer, *J. Med. Chem.*, 42,4422 (1999).

167. J. S. Dixon, *Proteins: Struct., Funct., Genet.*, Suppl., 1, 198 (1997).

168. P. D. J. Grootenhuis and P. J. M. vanGalen, *Acta Crystallogr. D*, 51, 560 (1995).

169. M. K. Holloway, J. M. Wai, T. A. Halgren, P. M. D. Fitzgerald, J. P. Vacca, B. D. Dorsey, R. B. Levin, W. J. Thompson, L. J. Chen, S. J. deSolms, N. Gaffin, A. K. Ghosh, E. A. Giuliani, S. L. Graham, J. P. Guare, R. W. Hungate, T. A. Lyle, W. M. Sanders, T. J. Tucker, M. Wiggins, C. M. Wiscount, O. W. Woltersdorf, S. D. Young, P. L. Darke, and J. A. Zugay, *J. Med. Chem.*, 38,305 (1995).

170. M. K. Holloway, *Perspect. Drug Discovery Des.*, 9/10/11, 63 (1998).

171. N. S. Blom and J. Sygusch, *Proteins: Struct., Funct., Genet.*, 27,493 (1997).

172. N. Tomioka and A. Itai, J. *Cornput.-Aided Mol. Des.*, 8,347 (1994).

173. S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, Jr., and P. Weiner, *J. Am. Chem. Soc.*, 106,765 (1984).

174. S. J. Weiner, P. A. Kollman, D. T. Nguyen, and D. A. Case, J. *Comput. Chem.*, 7,230 (1986).

175. P. J. Goodford, J. *Med. Chem.*, 28,849 (1985).

176. D. Qui, P. S. Shenkin, E. P. Hollinger, and W. C. Still, *J. Phys. Chem.*, 101, 3005 (1997).

177. D. Eisenberg and A. D. McLachlan, *Nature*, 319,199 (1986).

178. P. F. W. Stouten, C. Frommel, H. Nakamura, and C. Sander, *Mol. Simul.*, 10, 97 (1993).

179. S. Vajda, Z. Weng, R. Rosenfeld, and C. DeLisi, *Biochemistry*, 33, 13977 (1994).

180. M. Rarey, B. Kramer, T. Lengauer, and G. Klebe, J. *Mol. Biol.*, 261, 470 (1996).

181. H.-J. Bohm, J. *Cornput.-Aided Mol. Des.*, 8, 243 (1994).

182. H. J. Bohm, *J. Comput.-Aided Mol. Des.*, 12, 309 (1998).

183. B. Kramer, G. Metz, M. Rarey, and T. Lengauer, *Med. Chem. Res.*, 9, 463 (1999).

184. I. Muegge, *Perspect. Drug Discovery Des.*, 20, 99 (2000).

185. I. Muegge and Y. C. Martin, J. *Med. Chem.*, 42, 791 (1999).

186. I. Muegge, J. *Comput. Chem.*, 22,418 (2001).

187. I. Muegge, Y. C. Martin, P. J. Hajduk, and S. W. Fesik, J. *Med. Chem.*, 42, 2498 (1999).

188. I. Muegge and B. Podlogar, *Quant. Struct.-Act. Relat.*, 20,215 (2001).

189. S. Ha, R. Andreani, A. Robbins, and I. Muegge, *J. Cornput.-Aided Mol. Des.*, 14, 435 (2000).

190. M. D. Eldridge, C. W. Murray, T. R. Auton, G. V. Paolini, and R. P. Mee, J. *Cornput.-Aided Mol. Des.*, 11,425 (1997).

191. C. W. Murray, T. R. Auton, and M. D. Eldridge, J. *Cornput.-Aided Mol. Des.*, 12,503 (1998).

192. R. X. Wang, L. Liu, L. H. Lai, and Y. Q. Tang, J. *Mol. Model.*, 4, 379 (1998).

193. D. K. Gehlhaar, G. M. Verkhivker, P. A. Rejto, C. J. Sherman, D. B. Fogel, L. J. Fogel, and S. T. Freer, *Chem. Biol.*, 2,317 (1995).

194. T. A. Halgren, J. *Comput. Chem.*, 17, 520 (1996).

195. B. Honig and A. Nicholls, *Science*, 268, 1144 (1995).

196. D. R. Flower, J. *Mol. Graphics Modell.*, 15,238 (1998).

197. T. R. Stouch and P. C. Jurs, *J. Chem. Inf. Comput. Sci.*, 26, 4 (1986).

198. C. Bissantz, G. Folkers, and D. Rognan, J. *Med. Chem.*, 43, 4759 (2000).

199. M. Stahl and M. Rarey, J. *Med. Chem.*, 44, 1035 (2001).

J. G. Millichap, *Ann.* N. *Y. Acad. Sci.*, **205,321 (1973).**

J. M. Swanson and M. Kinsbourne in G. H. Hale and M. Lewis, Eds., ***Attention and Cognitive Development,*** Vol. **1,** Plenum, New York, **1979,** p. **249.**

M. Froimowitz, K. S. Patrick, and V. Cody, *Pharm. Res.*, **12, 1430** (1995).

I. J. Enyedy, W. A. Zaman, S. Sakamuri, A. P. Kozikowski, K. M. Johnson, and S. Wang, ***Bioorg. Med. Chem. Lett.*, 11, 1113 (2001).**

J. H. Voigt, B. Bienfait, S. Wang, and M. C. Nicklaus, ***J. Chem. Inf. Comput. Sci.*, 41, 702 (2001).**

Chem-X (version 96), Oxford Molecular Group, Inc., Hunt Valley, MD (company no longer exists) **(2001).**

I. J. Enyedy, S.-L. Lee, A. H. Kuo, R. B. Dickson, C.-Y. Lin, and S. Wang, **J. *Med. Chem.*, 44, 1349 (2001).**

207. R. Friedrich, P. Fuentes-Prior, E. Ong, G. Coombs, M. Hunter, R. Oehler, D. Pierson, R. Gonzalez, R. Huber, W. Bode, and E. L. Madison, **J. *Biol. Chem.*, 277,2160 (2002).**

208. A. Sali, L. Potterton, F. Yuan, H. van Vlijmen, and M. Karplus, ***Proteins: Struct., Funct., Genet.*, 23,318 (1995).**

209. A. Sali, ***Proteins: Struct., Funct., Genet.*, 6,437 (1995).**

210. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, **J. *Comput. Chem.*, 4, 187** (1983).

211. M. N. Janakiraman, C. L. White, W. G. Laver, G. M. Air, and M. Luo, ***Biochemistry,* 33,8172 (1994).**

212. G. M. Rishton, ***Drug Discovery Today, 2,* 382 (1997).**

213. N. R. Taylor, A. Cleasby, O. Singh, T. Skarzynski, A. J. Wonacott, P. W. Smith, S. L. Sollis, P. D. Howes, P. C. Cherry, R. Bethell, P. Colman, and J. Varghese, **J. *Med. Chem.*, 41, 798 (1998).**