

Progress and issues for computationally guided lead discovery and optimization

William L. Jorgensen

INTRODUCTION

Since the late 1980s there have been striking advances, fueled by large increases in both industrial and NIH-funded academic research, that have revolutionized drug discovery. This period has seen the introduction of high-throughput screening (HTS), combinatorial chemistry, PC farms, Linux, SciFinder, structure-based design, virtual screening by docking, free-energy methods, absorption/distribution/metabolism/excretion (ADME) software, bioinformatics, routine biomolecular structure determination, structures for ion channels, G-protein-coupled receptors (GPCRs) and ribosomes, structure/activity relationships (SAR) obtained from nuclear magnetic resonance (SAR by NMR), fragment-based design, gene knockouts, proteomics, small interfering RNA (siRNA), and human genome sequences. The result is a much-accelerated progression from identification of biomolecular target to lead compound to clinical candidate. However, a serious concern is that the dramatic increase in drug discovery abilities and expenditures has not been paralleled by an increase in FDA approvals of new molecular entities.¹ High demands for drug safety, broader and longer clinical trials, too much HTS, too little natural products research, and effective generic drugs for many once-pressing afflictions have all been suggested as contributors.^{2–4} Numerous corporate mergers and acquisitions may have also had adverse effects on productivity through distractions of reorganization and integration. Nevertheless, one should consider what the success would have been in the absence of the striking technical advances. Certainly, progress with some critical and challenging target classes such as kinases would have been greatly diminished, and the adverse impact on many cancer patients would have been profound. Indeed, further gains in the treatment and prevention of human diseases must require even more emphasis and commitment to fundamental research. As in other discovery enterprises, the answer is to drill deeper.

The topic of this volume focuses on one of the areas in drug discovery that has seen major transformation and progress: structure- and ligand-based design. The design typically features small molecules that bind to a biomolecular target and inhibit its function. The distinction stems

from whether a three-dimensional structure of the target is available and used in the design process. Structure-based design can be carried out with nothing more than the target structure and graphics tools for building ligands in the proposed binding site. However, additional insights provided by evaluation of the molecular energetics for the binding process are central to most current structure-based design activities. Ligand-based design does not require a target structure but rather stems from analysis of structure/activity data for compounds that have been tested in an assay for the biological function of the target. One seeks patterns in the assay results to suggest potential modifications of the compounds to yield enhanced activity. The upside is that a target structure is not required; the downside is that substantial activity data are needed. My research group has focused on the development and application of improved computational methodology for structure-based design. Some of the experiences and issues that have been addressed are summarized in the following.

LEAD GENERATION

Both lead generation and lead optimization may be pursued through joint computational and experimental studies. As summarized in Figure 1.1, our approach has evolved to feature two pathways for lead generation, *de novo* design with the ligand-growing program BOMB (Biochemical and Organic Model Builder)⁵ and virtual screening using the docking program GLIDE.⁶ Fragment-based design, which involves the docking and linking together of multiple small molecules in a binding site, is another popular alternative.^{7,8} Desirable compounds resulting from *de novo* design normally have to be synthesized, whereas compounds from virtual screening of commercial catalogs are typically purchased. In both cases, it is preferable to begin with a high-resolution crystal structure for a complex of the target protein with a ligand; though the ligand is removed, it is advisable to start from a complex rather than an apo structure, which may have side chains repositioned to fill partially the binding site. An extreme example occurs with HIV-1 reverse transcriptase (HIV-RT) for which the allosteric binding site for nonnucleoside inhibitors (NNRTIs) is fully collapsed in apo structures.⁹

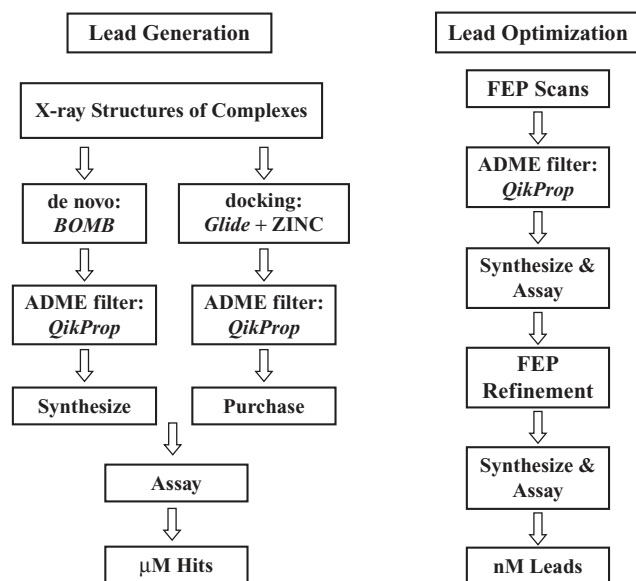


Figure 1.1. Schematic outline for structure-based drug lead discovery and optimization.

De Novo design with BOMB

BOMB is used to construct complete analogs by adding 0–4 substituents to a core that has been placed in a binding site. A thorough conformational search is performed for each analog, and the position, orientation, and dihedral angles for the analog are optimized using the OPLS-AA force field for the protein and OPLS/CM1A for the analog.¹⁰ The resultant conformer for each analog with the lowest energy is evaluated with a dockinglike scoring function to predict activity. The core may be as simple as, for example, ammonia or benzene, or it may represent a polycyclic framework of a lead series. For the example in Figure 1.2, ammonia was the original core, and it was positioned to form a hydrogen bond with the carbonyl group of Lys101. A library of molecules is then often built using a “template” that has been envisioned by the user to be complementary to the binding site and often to also be amenable to straightforward synthesis. For Figure 1.2, the template was Het-NH-34Ph-U, where Het represents a monocyclic heterocycle, 34Ph is a 3- or 4-substituted phenyl group, and U is an unsaturated hydrophobic group. The template specifies the components that constitute the desired molecules and the topology by which they are linked together.

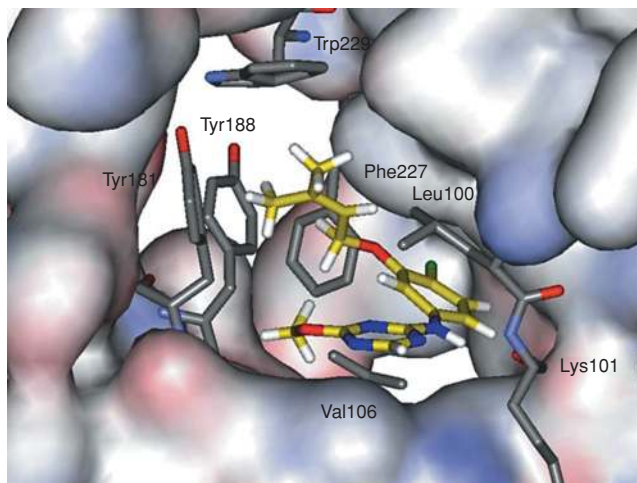


Figure 1.2. An inhibitor built using BOMB in the NNRTI binding site of HIV-RT.

BOMB includes a library of approximately 700 possible substituents, with code numbers from 1 to about 700, including most common monocyclic and bicyclic heterocycles and about 50 common U groups such as allyl, propargyl, phenyl, phenoxy, and benzyl derivatives. They are provided as groupings by the code numbers or the user can create a custom grouping with desired code numbers. The groupings correspond to template components such as Het, 5Het (just 5-membered ring heterocycles), 6Het, biHet, U, oPhX, mPhX, pPhX, mOPhX, pSPhX, OR, NR, SR, and C = OX. The program then builds all molecules that correspond to the template. In the example, if there were 50 Het and 20 U options, the program would build the 1,000 Het-NH-3-Ph-U and 1,000 Het-NH-4-Ph-U possibilities. This de novo design exercise with HIV-RT as the target resulted in identification of Het = 2-thiazolyl and U = dimethylallyloxy as a promising pair. Subsequent synthesis of the thiazole **1** in Figure 1.3 did provide a 10-μM lead in an MT-2 cell-based assay for anti-HIV activity. As described below, the lead was optimized to multiple highly potent NNRTIs, including the chlorotriazine in Figure 1.2 (31 nM), the corresponding chloropyrimidine (10 nM), and the cyanopyrimidine analog **2** (2 nM).^{11–14}

Some additional details should be noted. The host, typically a protein, is rigid in the BOMB optimizations except for variation of terminal dihedral angles for side chains with

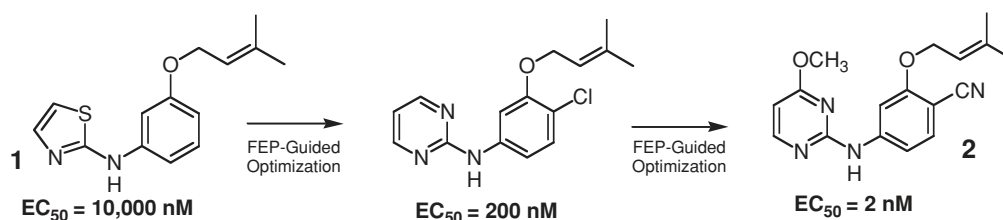


Figure 1.3. Example of a 10-μM lead proposed by BOMB that was optimized to provide numerous potent anti-HIV agents.

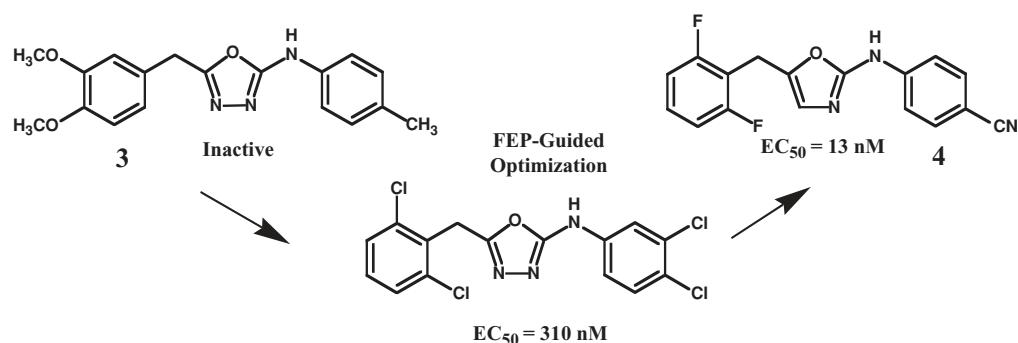


Figure 1.4. Progression of a false positive from docking to potent anti-HIV agents.

hydrogen-bonding groups, for example, the OH of tyrosine or serine and ammonium group of lysine. The current scoring function has been trained to reproduce experimental activity data for more than 300 complexes of HIV-RT, COX-2, FK506 binding protein, and p38 kinase.⁵ It yields a correlation coefficient r^2 of 0.58 for the computed versus observed log(activities). The scoring function contains only five descriptors that were obtained by linear regression, including an estimate of the analog's octanol/water partition coefficient from QikProp (QlogP),¹⁵ the amount of hydrophobic surface area for the protein that is buried on complex formation, and an index recording mismatched protein/analog contacts, such as a hydroxyl group in contact with a methyl group. Interestingly, the most significant descriptor is QlogP, which alone yields a fit with an r^2 of 0.47. Thus, the adage that increased hydrophobicity leads to increased binding is well supported, though it requires refinement for quality of fit using the host/ligand interaction energy or an index of mismatched contacts. Overdone, it also leads to ADME problems, especially poor aqueous solubility and high serum protein binding.

The results from a BOMB run include the structure for each protein/analog complex as a Protein Data Bank (PDB) file or BOSS/MCPRO Z-matrix (internal coordinate representation)¹⁶ and a spreadsheet with one row for each analog summarizing computed quantities from the BOMB calculations, including host–analog energy components and surface area changes as well as predicted properties for the analog, including log $P_{o/w}$, aqueous solubility, and Caco-2 cell permeability from QikProp, which is called as a subroutine. The processing time for Het-NH-Ph-U using ammonia as the core is approximately 15 s per analog on a 3-GHz Pentium IV processor. The required time increases roughly linearly with the number of conformers that need to be constructed. For large libraries, multiple processors are used.

Virtual screening

The common alternative is to perform virtual screening on available compound collections using docking software. Many reviews and comparisons for alternative software and

scoring functions are available.^{6,17–20} There is no question that there have been many successes with docking such that, given a target structure, it is expected to be competitive with and far more cost effective than HTS and is now an important component of lead discovery programs in the pharmaceutical industry. New success stories are reported regularly in the literature and at conferences. However, it is generally accepted that correct rank-ordering of compounds for activity is beyond the current capabilities. This is not surprising in view of the thermodynamic complexity of host/ligand binding, including potential structural changes for the host on binding, which have usually been ignored, and the need for careful consideration of changes in conformational free energetics between the bound and unbound states.²¹

In our experience, docking has been a valuable complement to de novo design (Figure 1.1). When large compound collections are docked, interesting structural motifs often emerge as potential cores that may have been overlooked otherwise. Our earliest docking effort started out well, was formally a failure, and then recovered to provide an interesting lead series that yielded potent anti-HIV agents.^{5,22} Leads were sought by processing a collection of approximately 70,000 compounds from the Maybridge catalog, which was supplemented with twenty known NNRTIs. The screening protocol began with a similarity filter that retrieves 60% of the known actives in the top 5% of the screened library. The approximately 2,000 library compounds that were most similar to the known actives were then docked into the 1rt4 structure of wild-type HIV-RT, using GLIDE 3.5 with standard precision.⁶ The top 500 compounds were then redocked and scored in GLIDE extraprecision (XP) mode.²³ The top 100 of these were postscored with a molecular mechanics/generalized Born/surface area (MM-GB/SA) method that was shown to provide high correlation between predicted and observed activities for NNRTIs.²² Though known NNRTIs were retrieved well (ten were ranked in the top twenty), purchase and assaying of approximately twenty high-scoring compounds from the library failed to yield any active anti-HIV agents. Persisting, the highest-ranked library compound, the inactive oxadiazoles **3** in Figure 1.4, was pursued computationally to seek

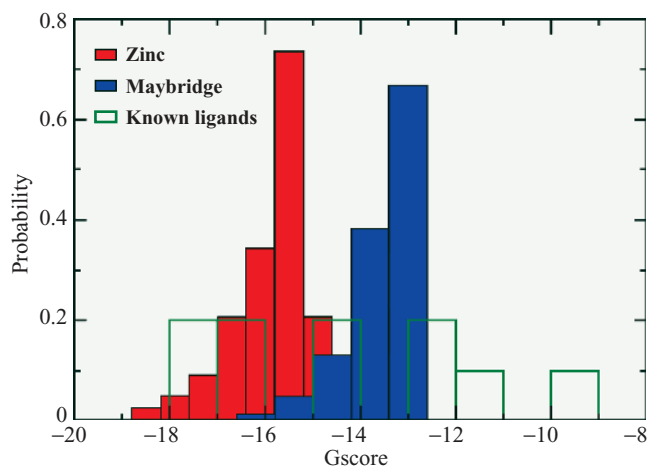


Figure 1.5. Distributions of the Glide XP scores for the top-ranked 1,000 ZINC compounds, the top-ranked 1,000 Maybridge compounds, and the 10 known tautomerase inhibitors.

constructive modifications. Specifically, the substituents were removed to yield the anilinybenzyloxadiazole core. A set of small substituents was reintroduced in place of each hydrogen using BOMB; scoring with BOMB, followed by free-energy perturbation (FEP)-guided optimization, led to synthesis and assaying of several polychloro analogs with EC_{50} values as low as 310 nM in the MT-2 HIV-infected T-cell assay.⁵ Further cycles of FEP-guided optimization led to novel, very potent NNRTIs, including the oxazole derivative **4**, as described more below.²⁴

A more recent virtual screening exercise was strikingly successful.²⁵ New protocols had evolved, including use of the much larger ZINC database of approximately 2.1 million commercially available compounds.²⁶ The goal in this case was to disrupt the binding of macrophage migration inhibitory factor (MIF) to its receptor CD74, an integral membrane protein, and a major histocompatibility complex (MHC) class II chaperone. MIF is a pro-inflammatory cytokine that is released by T-cells and macrophages. It plays a key role in a wide range of inflammatory diseases and is involved in cell proliferation and differentiation and angiogenesis.^{27,28} Curiously, MIF is also a ketonol isomerase. There is evidence that the interaction of MIF with CD74 occurs in the vicinity of the tautomerase active site and that MIF inhibition is directly competitive with MIF/CD74 binding.²⁹ The docking was performed using GLIDE 4.0 and the 1ca7 crystal structure of the complex of MIF with *p*-hydroxyphenylpyruvate.³⁰ In addition to the ZINC collection, the Maybridge HitFinder library was screened, which provided an additional 24,000 compounds. After all structures were processed using SP GLIDE, the top-ranked 40,000 from ZINC and 1,000 from Maybridge were redocked and rescored using GLIDE in XP mode.²³ GLIDE XP scoring was also shown to provide good correlation with experimental data for 10 known inhibitors of MIF's tautomerase activity.

A key observation from the docking is illustrated in Figure 1.5, which shows the distributions of GLIDE XP scores for the top-ranked 1,000 compounds from ZINC, the top-ranked 1,000 Maybridge compounds, and the ten known MIF inhibitors. Clearly, the large ZINC collection yields many compounds with much more promising XP scores than the Maybridge HitFinder library. The average molecular weights for the two sets of 1,000 compounds are 322 for ZINC and 306 for Maybridge. The variation only amounts to one additional nonhydrogen atom for the ZINC set, so the improved performance with the ZINC collection presumably results from greater structural variety. In view of the sensitivity of activity to structure, as reflected in Figures 1.3 and 1.4, it is highly unlikely that active compounds can be found in small libraries like Maybridge HitFinder unless the assays can be run with the compounds at millimolar or higher concentrations, which is often precluded by solubility limits. Even with a viable core (Figure 1.4), the chance is low that a small library will contain a derivative with a substituent pattern that yields an active in a typical assay.

Finally, the GLIDE poses for approximately 1,200 of the top-ranked compounds were displayed and 34 compounds were selected by human evaluation of the poses with input from QikProp on predicted properties and structural liabilities. The filtering included rejection of poses where the conformation of the ligand was energetically unlikely or where there were overly short intramolecular contacts and compounds with generally undesirable features such as readily hydrolyzable functional groups or substructures such as coumarins, which are promiscuous protein binders. Only 24 of the 34 selected compounds were, in fact, available for purchase, which represents a typical ratio. Ultimately 23 compounds were submitted to a protein-protein binding assay using immobilized CD74 and biotinylated human MIF with streptavidin-conjugated alkaline phosphatase processing *p*-nitrophenyl phosphate as substrate. Remarkably, eleven of the compounds were found to have inhibitory activity in the μ M regime including four compounds with IC_{50} values below 5 μ M. Inhibition of MIF tautomerase activity was also established for several of the compounds with IC_{50} values as low as 0.5 μ M. Representative active compounds are shown in Figure 1.6; optimization of several of the lead series is being pursued. Notably, these are the most potent small-molecule inhibitors of MIF-CD74 binding that have been reported to date.

The first three compounds in Figure 1.6 were ranked in 285th, 696th, and 394th place by the XP scoring, so they were not “high in the deck.” However, prior de novo structure building with BOMB had indicated that 6–5 fused bicyclic cores should be promising, so the selections were biased in this direction. The compound ranked first with XP GLIDE was also purchased and assayed; it turned out to be the 250- μ M inhibitor in Figure 1.6. In addition, the compounds ranked 26th and 32nd were purchased and found to be inactive. Overall, it is expected that contributors to the success with the virtual screening in this case were

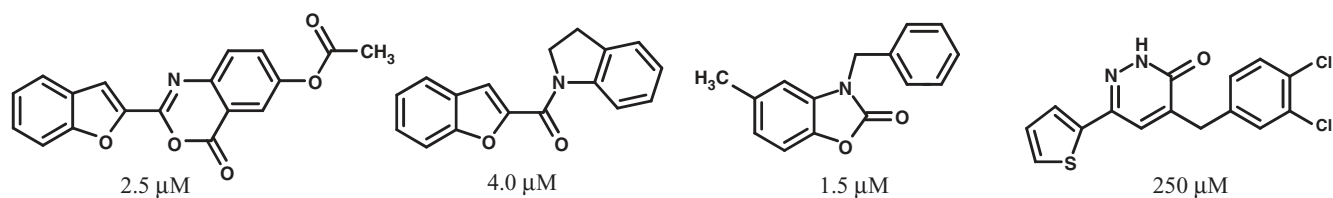


Figure 1.6. Structures and IC_{50} values for some inhibitors of MIF-CD74 binding discovered by virtual screening.

improvements with GLIDE 4.0 and the XP scoring, use of the large ZINC library, the relatively small binding site and consequently small number of rotatable bonds for potential inhibitors, and the human filtering.

ADME ANALYSES

As indicated in Figure 1.1, as one pursues leads it is important to be aware of potential pharmacological liabilities. The significance of this issue became increasingly apparent in the 1990s because of high failure rates for compounds in clinical trials that could be ascribed to ADME and toxicity problems.³¹ This led to the introduction of Lipinski's rules and recognition that compounds developed in the post-HTS era frequently tended to be too large and hydrophobic, which is accompanied by solubility and bioavailability deficiencies.³² In this atmosphere, more effort was placed on quantitative prediction of molecular properties beyond $\log P_{o/w}$ using statistical procedures such as regression analyses and neural networks, which were trained on experimental data.^{33,34} The typical regression equation is a linear one, Equation (1.1), where the sum is over molecular descriptors i that have values c_i for the given structure and the coefficients a_i are determined to minimize the error with the experimental data:

$$\text{property} = \sum_i a_i c_i + a_0. \quad (1.1)$$

In Figure 1.1, the choice for ADME analyses is QIKPROP, which was among the earliest programs to predict a substantial array of pharmacologically relevant properties.

Version 1.0, which was released in March 2000, provided predictions for intrinsic aqueous solubility, Caco-2 cell permeability, and hexadecane/gas, octanol/gas, water/gas, and octanol/water partition coefficients. The required input for QIKPROP is a three-dimensional structure of an organic molecule, and it mostly uses linear regression equations with molecular descriptors such as surface areas and hydrogen-bond donor and acceptor counts. By version 3.0 from 2006, the output covered eighteen quantities, including $\log BB$ for brain/blood partitioning, $\log K_{hsa}$ for serum albumin binding, hERG K^+ channel blockage, primary metabolites, and overall percentage human oral absorption.¹⁵ The prediction of primary metabolites is based on literature precedents and recognition of corresponding substructures; for example, methyl ethers and tolyl methyl groups are typically metabolized to the alcohols. Execution time with QIKPROP is negligible because the most time-consuming computation is for the molecule's surface area. Average root-mean-square (rms) errors for most quantities are about 0.6 log unit, as in Figure 1.7.

To gauge acceptable ranges of predicted properties, QIKPROP 3.0 was used to process approximately 1,700 known neutral oral drugs,¹³ which were compiled by Proudfoot.³⁵ For submission to QIKPROP, the original two-dimensional structures were converted to three-dimensional structures and energy-minimized with BOSS using the OPLS/CM1A force field.^{10,16} Some key results from the analyses are summarized as histograms in Figures 1.8 and 1.9. Consistent with the $\log P_{o/w}$ limit of 5 in Lipinski's rules,³² 91% of oral drugs are found to have

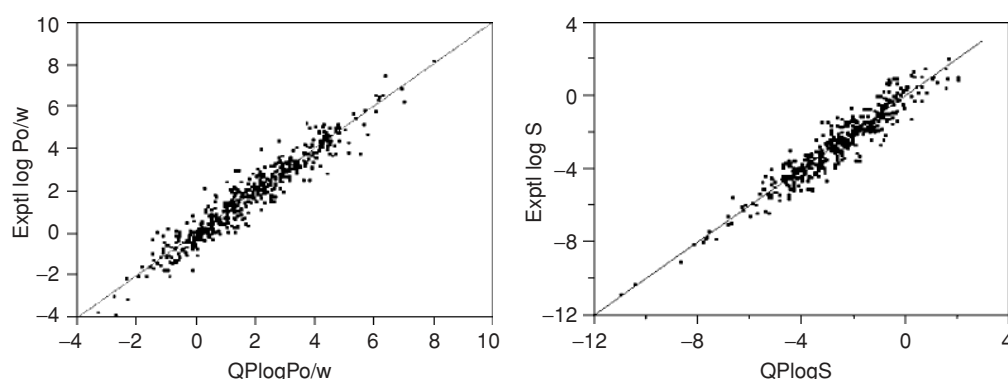


Figure 1.7. Experimental data and QikProp 3.0 results for 400–500 octanol/water partition coefficients (left) and aqueous solubilities (right). S is aqueous solubility in moles per liter. Correlation coefficients r^2 are 0.92 and 0.90 and the rms errors are 0.54 and 0.63 log unit, respectively.

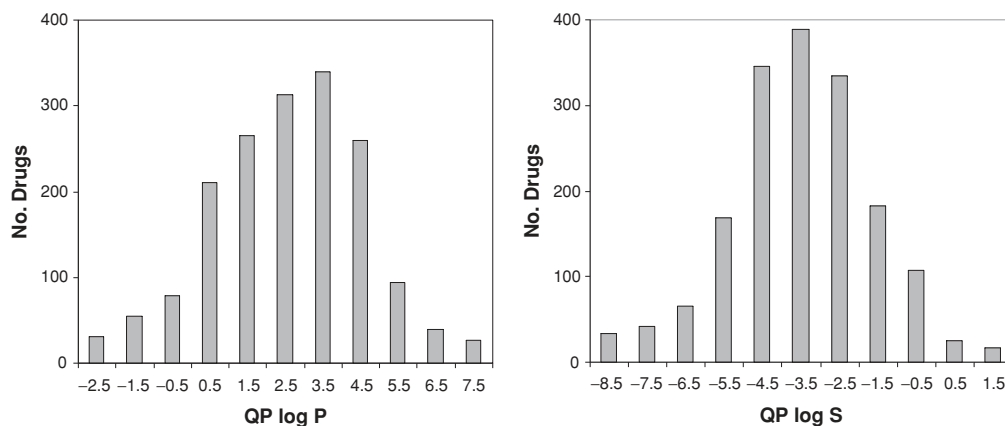
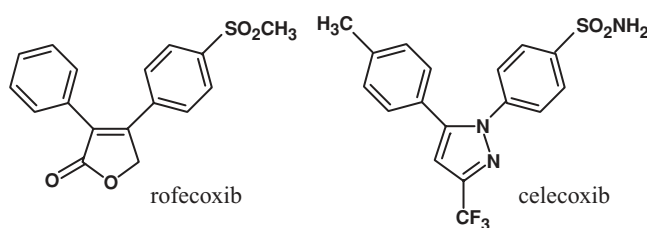


Figure 1.8. QikProp distributions for $\log P_{o/w}$ (left) and $\log S$ (right) for 1712 oral drugs.

QLogP values below 5.0. However, values below zero are uncommon, presumably because of poor cell permeability, and the “sweet” range for $\log P_{o/w}$ appears to be 1–5. For aqueous solubility, 90% of the QLogS values are above -5.7 , that is, S is greater than $1 \mu\text{M}$. QLogS values less than -6 or greater than -1 are undesirable. The QikProp results also state that 90% of oral drugs have cell permeabilities, P_{Caco} , above 22 nm/s and no more than six primary metabolites. These quantities and limits address important components of bioavailability, namely, solubility, cell permeability, and metabolism.

For our design purposes (Figure 1.1), a compound is viewed as potentially ADME challenged if it does not satisfy all components of a “rule-of-three”: predicted $\log S > -6$, $P_{\text{Caco}} > 30 \text{ nm/s}$, and maximum number of primary metabolites of 6. For central nervous system (CNS) activity requiring blood-brain barrier penetration, an addendum is that QLogBB should be positive. Also, some caution is warranted for a compound with no metabolites because of possible clearance problems.¹⁷ A further note is that QLogP and QLogS are correlated with an r^2 of 0.68, so there would be some redundancy in invoking limits on both. Among reasons for preferring solubility, there are quite a few examples of relatively small drugs that

have $\log P_{o/w}$ values greater than 5 but have acceptable solubility, for example, meclizine, prozapine, clocinnazoline, bepridil, denaverine, bopindolol, phenoxybenzamine, and terbinafine. Of course, compounds with reactive functional groups, for example, those that are readily hydrolyzable or strongly electrophilic, are flagged by QikProp and normally eliminated from inclusion in a lead structure. For example, in rofecoxib (Vioxx) concern could be expressed for possible nucleophilic attack and ring opening at the furanone carbonyl and for Michael addition to the α,β -double bond; metabolic oxidation at the allylic methylene group is also expected to yield the 5-hydroxy derivative (Scheme 1). For celecoxib (Celebrex), metabolic oxidation to the benzylic alcohol is noted by QikProp, and an “alert” is given that



Scheme 1.

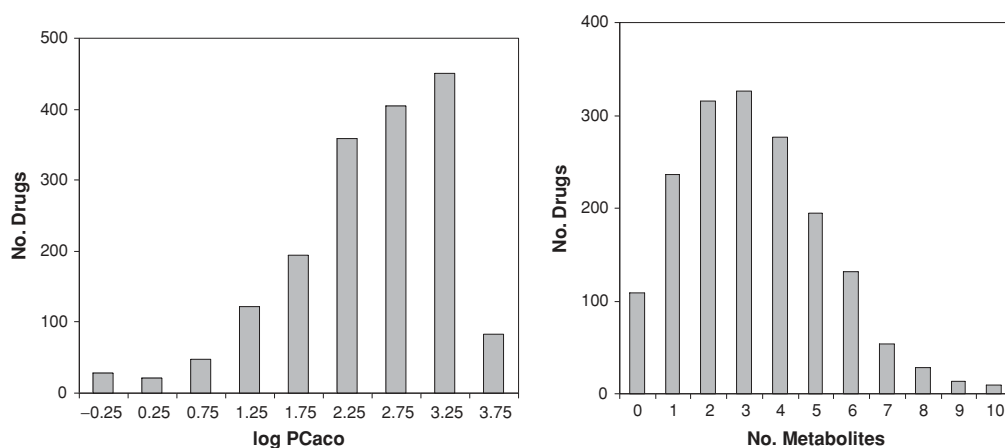


Figure 1.9. QikProp distributions for $\log P_{\text{Caco}}$ (left) and number of primary metabolites (right) for 1,712 oral drugs. P_{Caco} is the Caco-2 cell permeability in nm/s .

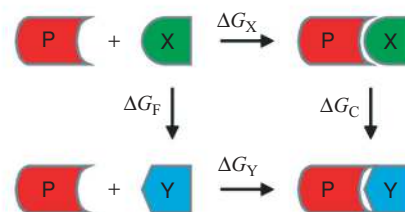
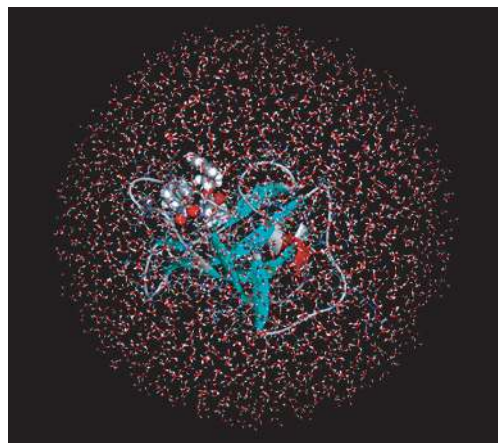


Figure 1.10. (Left) A protein/ligand complex surrounded by approximately 1,000 water molecules in a spherical shell or “cap.” (Right) Thermodynamic cycle used to compute relative free energies of binding, $\Delta\Delta G_b$. P is a host and X and Y are two ligands.

the primary sulfonamide group can be associated with sulfa allergies and indiscriminant metal chelation.³⁶

Overall, for the 1,712 oral drugs, 278 violate one or more of the four Lipinski rules ($MW < 500$, $\log P_{o/w} < 5$, H-bond donors ≤ 5 , H-bond acceptors ≤ 10) with QPlogP used for $\log P_{o/w}$. There are 178, 82, and 18 oral drugs with one, two, and three violations, respectively. The group with two violations includes macrolides such as erythromycin and azithromycin and some other well-known drugs like atorvastatin, amiodarone, chloramphenicol, ketoconazole, and telmisartan. These examples all fail one member of the rule-of-three, either the solubility limit or number of primary metabolites, for example, respectively, atorvastatin and the macrolides. The group with three rule-of-five violations includes the HIV-protease inhibitors ritonavir and saquinavir, which are known to have low bioavailability. There are exceptions to such rules because they are based on 90th-percentile limits. Nevertheless, in all stages of lead generation, it would be imprudent to ignore property distributions for known drugs such as those in Figures 1.8 and 1.9.

LEAD OPTIMIZATION

It is assumed that inhibitory potency increases with increasing biomolecule-inhibitor binding. So, on the computational side, the key for lead optimization is accurate prediction of biomolecule-ligand binding affinities. There are many approaches, but the potentially most accurate ones are the most rigorous.¹⁷ At this time, the best that is done is to model the complexes in the presence of hundreds or thousands of explicit water molecules using Monte Carlo (MC) statistical mechanics or molecular dynamics methods (Figure 1.10).¹⁷ Classical force fields¹⁶ are used, and extensive sampling is performed for key external (translation and rotation) and internal degrees of freedom for the complexes, solvent, and any counterions. FEP and thermodynamic integration (TI) calculations then provide for-

mally rigorous means to compute free-energy changes.³⁷ For biomolecule/ligand affinities, perturbations are made to convert one ligand to another using the thermodynamic cycle in Figure 1.10. The conversions involve a coupling parameter, λ , that causes one molecule to be smoothly mutated to the other by changing the force field parameters and geometry.³⁸ The difference in free energies of binding for the ligands X and Y then comes from $\Delta\Delta G_b = \Delta G_X - \Delta G_Y = \Delta G_F - \Delta G_C$. Two series of mutations are performed to convert X to Y unbound in water and complexed to the biomolecule, which yield ΔG_F and ΔG_C .

Absolute free energies of binding are not obtained, but for lead optimization it is sufficient to assess the effects of making changes or additions to a core structure in the same spirit as synthetic modifications. Though the MC or MD plus FEP or TI calculations are rigorous, the accuracy of the results is affected by many issues, including the use and quality of force fields; missing energy terms, such as instantaneous polarization effects; and possible inadequate configurational sampling, which may be associated with, for example, infrequent conformational changes that are beyond the duration of the simulations. In the author's experience, more approximate methods are not accurate enough to provide satisfactory guidance in lead optimization.

The idea of using such calculations for molecular design goes back more than twenty years, at least to the report of the first FEP calculation for conversion of a molecule X to molecule Y in 1985³⁸ and to the earliest application of FEP calculations for protein-ligand binding by Wong and McCammon.³⁹ A final comment from McCammon's review on computer-aided molecular design in *Science* in 1987 was perspicacious: “The attentive reader will have noticed that no molecules were actually designed in the work described here.”⁴⁰ The situation has remained basically unchanged since the late 1980s. As the convergence of FEP calculations was investigated, it was apparent that they were too computationally intensive for routine use in molecular

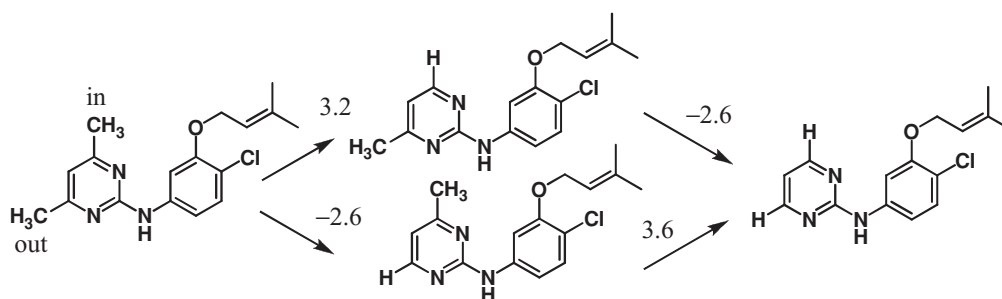


Figure 1.11. MC/FEP results for $\Delta\Delta G_b$ (kcal/mol) established a strong preference for a single methyl group oriented “in” toward Tyr188 in the NNRTI binding site. The precision of the calculations is reflected in the cycle’s small hysteresis, 0.4 kcal/mol.

design given the computer resources available before ca. 2000. In 1985, the ethane-to-methanol FEP calculation in a periodic cube with 125 water molecules required two weeks on a Harris-80 minicomputer,³⁸ and the Wong/McCammon MD simulation for the trypsin-benzamidine complex covered only 29 ps but was performed on a Cyber 205 “supercomputer.”³⁹

Thus, until recently the application of FEP or TI calculations on protein-ligand systems predominantly featured retrospective calculations to reproduce known experimental inhibition data and generally addressed small numbers of compounds. Kollman was a strong advocate of the potential of free-energy calculations for molecular design, and he and Merz reported a rare, prospective FEP result on the binding of a phosphoramidate versus phosphinate inhibitor with thermolysin.^{41,42} Pearlman also advanced the technology, though publications in 2001 and 2005 were still retrospective and confined to a simple congeneric series of 16 p38 kinase inhibitors.^{43,44} In addition, Reddy and Erion have been steady contributors; they have used FEP calculations to evaluate contributions of heteroatoms and small groups to the binding of inhibitors to gain insights on directions for improvement.^{45,46} Our own computations on protein/ligand binding began to appear in 1997 using MC/FEP methodology.^{47,48} Many issues and systems were subsequently addressed, including substituent optimization for celecoxib analogs,⁴⁹ COX-2/COX-1 selectivity,⁵⁰ and heterocycle optimization for inhibitors of fatty acid amide hydrolase.⁵¹ An additional series of publications used MC/FEP calculations to compute the effects of HIV-RT mutations on the activity of NNRTIs.^{52–55} The latter work included predictions for the structures of the complexes of efavirenz and etravirine with HIV-RT, which were subsequently confirmed by x-ray crystallography.^{52,54,56} Confidence in the predicted structures came from agreement between the FEP results and experimental activity data.

FEP-guided optimization of azines as NNRTIs

With this preparation, large increases in computer resources, the hiring of synthetic chemists, and collab-

oration with biologists, FEP-guided lead optimization projects were initiated in 2004. Early successes in the optimization of potent NNRTIs are reflected in Figures 1.2 and 1.3 for the Het-NH-3-Ph-U series.^{11–13} MC/FEP calculations were used to optimize the heterocycle and the substituent in the 4-position of the phenyl ring. The calculations are run with MCPRO and all use the OPLS/CM1A force field for the ligands and OPLS-AA for the protein.^{10,16} This quickly led to selection of 2-pyrimidinyl and 2-(1,3,5)-triazinyl for the heterocycle and chlorine or a cyano group at the 4-position. These combinations yielded NNRTIs with EC_{50} values near 200 nM.

Extensive FEP calculations then focused on optimization of substituents for the heterocycle.¹³ For the 2-pyrimidines, the immediate question concerned whether 4,6-disubstitution would be favorable or if mono substitution at the 4- or 6-position is preferred. In complexes with HIV-RT, the 4- and 6-positions are not equivalent; for example, in Figure 1.2, the methoxy group could be directed toward the viewer (“out”) or away (“in”), as shown. From display of structures of the complexes, the preferences for in or out were not obvious. This was clarified by MC/FEP results, which showed a strong preference for having a single small substituent on the pyrimidine ring and that the substituent would be oriented “in” (Figure 1.11). Synthesis of a variety of such mono-substituted pyrimidines and triazines yielded ten NNRTIs with EC_{50} s below 20 nM.^{11–13} There was good correlation between the FEP results and the observed activities.^{11,13} The methoxypyrimidine **2** in Figure 1.3 (2 nM) was the most potent, although it was also relatively cytotoxic (CC_{50} = 230 nM). The corresponding 1,3,5-triazine is also a potent anti-HIV agent (11 nM) and has a large safety margin (CC_{50} = 42 μ M).

Heterocycle scans

FEP results also established the orientation of the methoxy methyl group in the pyrimidine and triazine derivatives shown in Figure 1.2, that is, pointing toward Phe227 rather than Tyr181. This suggested the possibility of cyclizing the methoxy group back into the azine ring to form 6–5 and 6–6 fused heterocycles in the manner indicated in Scheme 2.

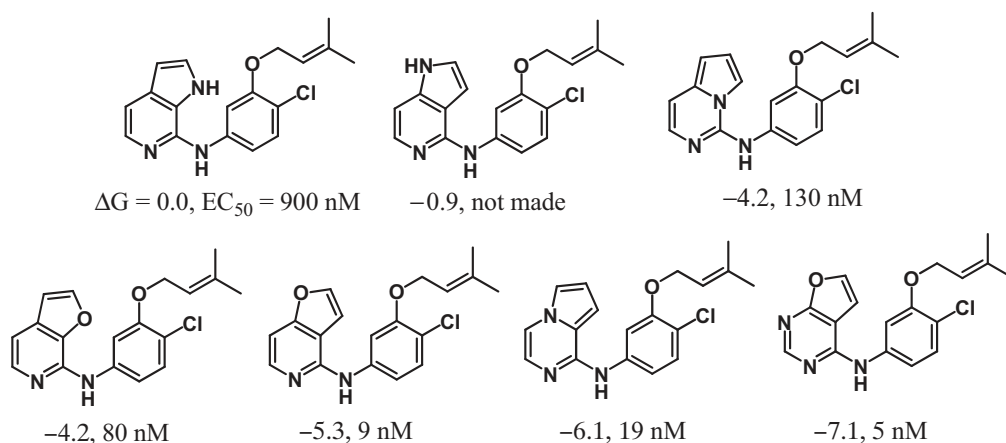
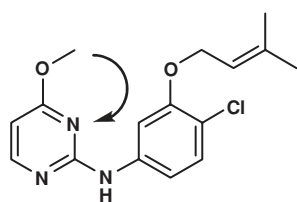


Figure 1.12. FEP results for relative ΔG_b (kcal/mol) and experimental anti-HIV activities (nM).



Scheme 2.

The decision on which analogs to pursue was driven by the prospective FEP results shown in Figure 1.12. Subsequent synthesis and assaying of the series of 6–5 compounds showed close parallel between the predicted and observed activities.¹¹ The illustrated furanopyrimidine derivative was predicted and observed to be the most potent; it is a highly novel and potent (5 nM) NNRTI. The results highlight the accuracy of the FEP predictions and again the sensitivity of activity to structure. The pyrrolopyrimidine (130 nM) and pyrrolopyrazine (19 nM) pair is particularly striking. After the fact, analyses showed a larger dipole moment for the bound pyrrolopyrazine and more negative charge on the pyrazinyl nitrogen leading to stronger hydrogen bonding with the backbone NH of Lys101.¹⁴

This procedure can be referred to as a heterocycle scan, which is clearly a powerful lead-optimization strategy.⁵¹ It is also an area where computation is far easier than synthesis, so computational screening to focus the synthetic options is very beneficial. This is particularly true for polycyclic heterocycles, as in Figure 1.12, where there are many options and the synthetic challenges can be great. In this example, heteroaryl halides were needed for reaction with substituted anilines; several were not previously reported and required considerable synthetic effort.¹⁴ Even with the notion of pursuing bicyclic heterocycles, in the absence of the FEP results, the synthetically less accessible ones might have been skipped.

Changing heterocycles in the center of a structure is also often challenging from a synthetic standpoint. For example, synthesis of the oxadiazoles and oxazole in Figure 1.4 requires fundamentally different procedures for the ring

construction.²⁴ This corresponds to a change in chemotype and there can be a significant delay as a viable synthetic route is found for the new target. In the case of this U-5Het-NH-*p*PhX series, FEP calculations were carried out for eleven alternative five-membered-ring heterocycles (5Het) by perturbation from the corresponding thiophene.²⁴ Remarkably, the only one that was predicted to be more active than the oxadiazole was the 2,5-substituted oxazole. The prediction was confirmed and provided a major step forward for the optimization of this series, as shown in Figure 1.13. It is noted that the approximately eight-fold activity improvement, which corresponds to a $\Delta \Delta G$ of about 1.2 kcal/mol, is less than the computed $\Delta \Delta G$ of 2.5 kcal/mol. This is a common pattern that likely results from the use of a cell-based assay, so the comparison is not with actual binding data (K_d). Moreover, it is also probable that the computed electrostatic interactions in the complexes are not properly damped because of the lack of explicit polarization effects.

In view of the synthetic challenges, only two alternatives were synthesized, the thiadiazole and thiazole analogs,

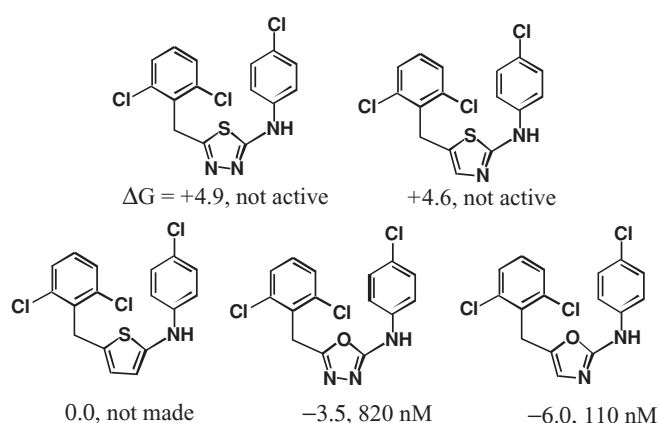


Figure 1.13. Heterocycle scan in the U-5Het-Ph X series; FEP results for ΔG_b (kcal/mol) relative to the thiophene analog and experimental anti-HIV activity (nM).

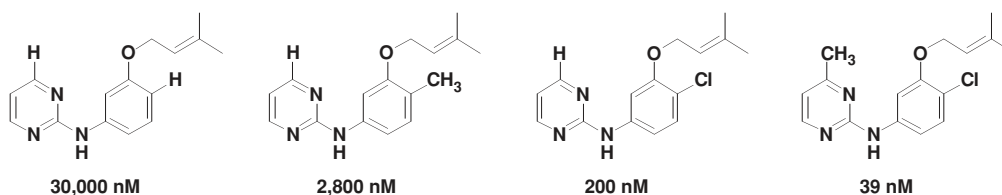


Figure 1.14. The power of chlorine and methyl scans; experimental EC_{50} values for anti-HIV activity in an MT-2 cell assay.

which were both predicted and found to be inactive (Figure 1.13). It is noted that the MT-2 assay is run to a maximum concentration of 100 μ M; the thiadiazole showed no activity or cytotoxicity up to this concentration, whereas the thiazole has a CC_{50} of 24 μ M and no anti-HIV activity to this point. Overall, this provides another example of the sensitivity of activity to structure and the desirability of rigorous computational guidance. Graphical display of modeled complexes is inadequate to gauge relative potency. In retrospect, the results indicate that the longer C-S bonds in the 2,5-disubstituted sulfur-containing heterocycles cause crowding of the dichlorobenzyl group and Tyr181, and the nitrogen in the 3-position has an electrostatically unfavorable interaction with Glu138.

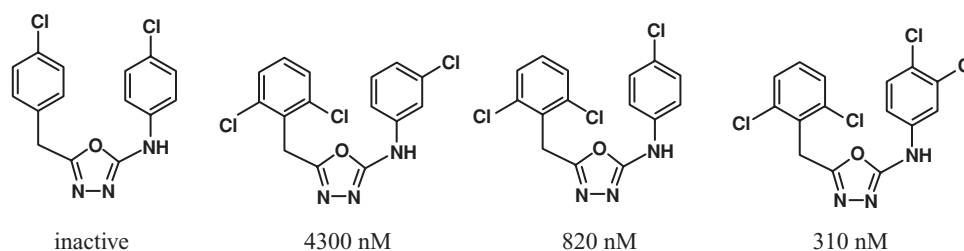
An interesting aside is that in the original publication, it was thought that the 2,5-disubstituted thiazole in Figure 1.13 showed weak activity with an EC_{50} of 3.1 μ M, which was out of line with the FEP results.²⁴ It was subsequently found that instead of the 2,5-isomer, the 2,4-isomer (S and N interchanged in the structure in Figure 1.13) was the actual compound that had been synthesized and assayed, as confirmed by a crystal structure. The two isomers are not unequivocally distinguishable by NMR. An alternative synthetic route was then pursued to yield the 2,5-isomer, which is indeed inactive, as predicted by the FEP calculations.

Small group scans

In addition to the heterocycle scans, small group scans are highly informative. These are performed routinely with BOMB to build the structures and provide initial scoring, followed by refinement with FEP calculations. A standard protocol with BOMB is to replace each hydrogen of a core, especially aryl hydrogens, with ten small groups that have been selected for difference in size, electronic char-

acter, and hydrogen-bonding patterns: Cl, CH₃, NH₂, OH, CH₂NH₂, CH₂OH, CHO, CN, NHCH₃, and OCH₃. This is generally adequate to define likely places for beneficial substitution of hydrogen by the least polar groups, Cl, CH₃, and OCH₃. The situation with the polar groups is less clear because of the competition for the ligand between hydrogen bonding in the complex versus unbound in water. As long as some hydrogens appear viable for substitution, a chlorine and/or methyl scan using FEP calculations is then desirable to obtain quantitatively reliable predictions. The potential value of using both a chlorine and methyl scan is well illustrated by the results in Figure 1.14; knowing the optimal position for a methyl group and a chlorine provides an activity boost from 30 μ M to 39 nM in this case.^{11–13}

A chlorine scan was also particularly helpful in evolving the inactive oxadiazole **3** in Figure 1.4 into potent anti-HIV agents. **3** had emerged in third place after the docking exercise and embedded among known, potent NNRTIs. The docking pose and the structure of the complex as built by BOMB also looked reasonable, although the score from BOMB was modest because of poor accommodation of the methoxy groups in the vicinity of Tyr181 and Tyr188. Assuming that the tricyclic core might be viable, the substituents were removed and a chlorine scan was performed using MC/FEP simulations.^{5,24} The predicted changes in free energy of binding for replacing each hydrogen by chlorine are summarized in Figure 1.15; again formally equivalent positions become nonequivalent in the complexes. The scan indicated that the most favorable positions for introduction of chlorines were at C3 and C4 in the phenyl ring and at C2 and C6 in the benzyl ring. A series of polychloro analogs were then synthesized and the activities were found to closely parallel the predictions (Scheme 3). The core and, for example, the 4,4'-dichloro analog were inactive; however, the illustrated



Scheme 3.