Calculation of Relative Differences in the Binding Free Energies of HIV1 Protease Inhibitors: A Thermodynamic Cycle Perturbation Approach

M. Rami Reddy,*,† Michael D. Varney, Vince Kalish, Vellarkad N. Viswanadhan,† and Krzysztof Appelt

Agouron Pharmaceuticals, Inc., 3565 General Atomics Court, San Diego, California 92121

Received January 3, 1994®

An iterative computer-assisted drug design (CADD) method that combines molecular mechanics, dynamics, thermodynamic cycle perturbation (TCP) calculations, molecular design, synthesis, and biochemical testing of peptidomimetic inhibitors and crystallographic structure determination of the protein–inhibitor complexes has been successfully applied to the design of novel inhibitors for the HIV1 protease. The first "designer" compound in this series (**I**) was designed by replacing the C-terminal Val-Val methyl ester of a known hydroxyethylene inhibitor with a diphenhydramine amide derivative in which two phenyl groups fill the p2′ and p3′ side-chain binding pockets in the HIV1 protease. Subsequent testing showed modest inhibition (K_i = 1.67 μM). Concurrently, molecular mechanics calculations on designed analogs indicated the feasibility of replacement of a phenyl ring with an indole ring (**II**). Synthesis and biochemical testing resulted in better inhibition potency for II. X-ray crystal structure determination of HIV1 protease complexed with **I** and **II** provided structural information for subsequent design and TCP calculations. A TCP protocol was established and validated for the mutation of **I** → **II**. TCP results showed a net gain of 2.1 (±0.9) kcal/mol in replacing **II** with **I**, which agreed with experimental result within an error margin of 0.8 kcal/mol. TCP calculations for six other mutations (**I** → **III**, **II** → **III**, **IV**, **V**, **VI**, and **VII**) were performed prior to synthesis and testing. These results allowed for the prioritization of design ideas for synthesis. In all cases where experimental results are available, TCP calculations showed good agreement. These results demonstrate that the TCP approach can be used with medicinal chemistry and crystallography for screening the proposed derivatives of a lead compound prior to synthesis, thus potentially reducing the time for the discovery of new drugs.

Introduction

As part of an overall effort to develop an effective treatment for acquired immunodeficiency syndrome (AIDS), researchers have mounted an intense campaign to understand and exploit the critical pathways in the life cycle of human immunodeficiency virus 1 (HIV1).[1,2] With the availability of the X-ray structure of HIV1 protease and an ever increasing number of inhibitor complexes,[2-4] inhibition of HIV1 protease has emerged as one of the most promising lines of anti-AIDS drug design research.[4-11] To speed up the achievement of this goal, the use of computational tools[10-16] has come into focus alongside experimental methods as part of an overall strategy of computer-assisted drug design.

Recently, we have had the opportunity to initiate an iterative computer-assisted drug design (CADD) approach for the discovery of a new class of HIV1 protease inhibitors. This is summarized in Figure 1 and is inspired by the success of similar approaches used for the discovery of novel thymidylate synthase and purine nucleoside phosphorylase inhibitors.[17] In our earlier effort,[17a] however, computational analysis was much less rigorous than the present one and was limited primarily to visualization and in a few cases, to energy minimization of the protein–inhibitor complex. If, however, during the design process the relative binding affinity of any two potential inhibitors could be accurately and reliably estimated, the total number or compounds required to reach a potent and novel inhibitor would be greatly reduced. The present article describes our recent efforts in developing such a computationally based design scheme (Figure 1). We begin with the crystal structure of a known protein–inhibitor complex and analyze the active-site characteristics. Then, a new analog is designed using the crystal structure of the complex as a starting point, and the new design is examined on a graphics screen. If the initial design looks satisfactory, then a precise theoretical estimation of relative binding affinity is made. If favorable binding is indicated by these calculations, then the synthesis and biochemical testing of the new inhibitors proceed. Once completed, another cycle of crystallographic structure determination is performed if necessary, and the process of design modifications, free energy calculations, synthesis, and testing is continued in an iterative fashion to improve inhibitor binding.

Critical to the success of this rational drug-design paradigm is the ability to correctly predict the binding affinity of an inhibitor designed from a lead compound, *prior* to synthesis. A free energy simulation technique known as the thermodynamic cycle perturbation (TCP) approach[15,18,19] used in conjunction with molecular dynamics calculations offers a theoretically precise method of determining the binding free energy differences of related inhibitors. However, in the past molecular dynamics calculations[19a] and the TCP approach[15,18,19] have been used primarily to rationalize known binding affinities, with few applications focusing on predictions.[10,11,14,21-25] Here, we report our results in the HIV1 protease system, demonstrating that the TCP approach can be used

* To whom correspondence should be addressed at Gensia, Inc., 4575 Eastgate Mall, San Diego, CA 92121.
† Present address: Gensia, Inc., 4575 Eastgate Mall, San Diego, CA 92121.
® Abstract published in *Advance ACS Abstracts*, March 15, 1994.

0022-2623/94/1837-1145$04.50/0

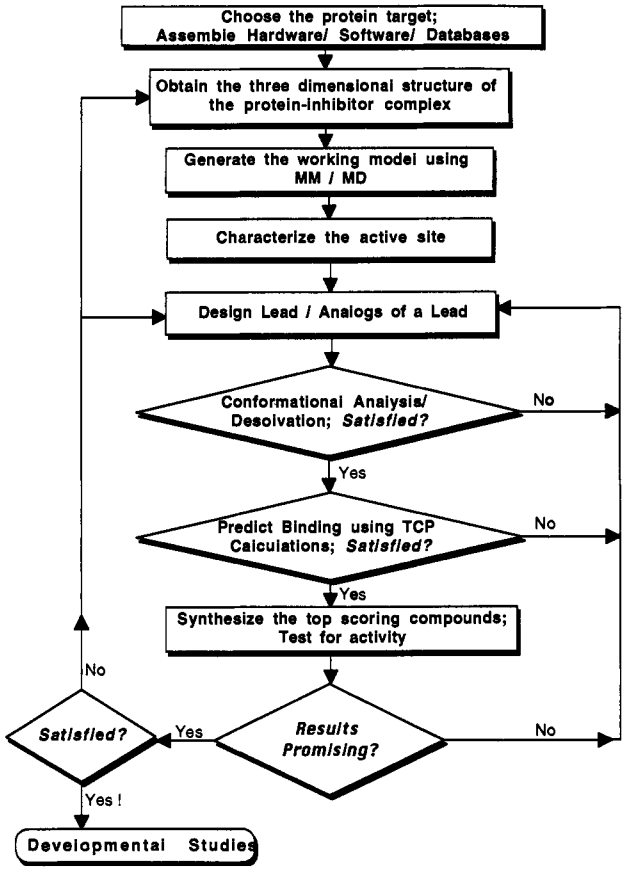


Figure 1. Flowchart for the structure-based drug-design paradigm used in the design of HIV1 protease inhibitors described in this paper. Though computational and experimental parts of the paradigm work together, only the computational aspects are highlighted as the main focus of this work. After identifying the target, all the necessary computers, graphics systems, software, and databases are assembled which will be used in the analysis of 3D structural information. After obtaining the structural information, a working model is generated by developing the molecular mechanics parameters, building the missing segments (if any) using molecular mechanics/molecular dynamics (MM/MD) simulations and removing bad contacts by energy minimization. The next step involves the characterization of the active site in terms of donor and acceptor groups for hydrogen bonds, hydrophilic and hydrophobic regions, empty pockets, etc. The information gained during the characterization process forms the basis for the design of lead compounds or their analogs to maximize binding affinity. The next step is to estimate the desolvation energy for the designed compound and to analyze its low-energy conformations in solution in relation to the putative binding conformation. This modeling process is followed by, in favorable cases, a detailed TCP analysis for predicting binding of the proposed derivative(s). If the TCP calculations indicate better binding, the compound is synthesized and tested; otherwise one returns to the design step. Finally, if promising results are obtained from biochemical testing, developmental studies are undertaken.

successfully in an iterative structure-based design program to accurately predict relative binding affinities of designed inhibitor analogs, thus eliminating the need to synthesize weak or nonoptimal inhibitors.

Methods

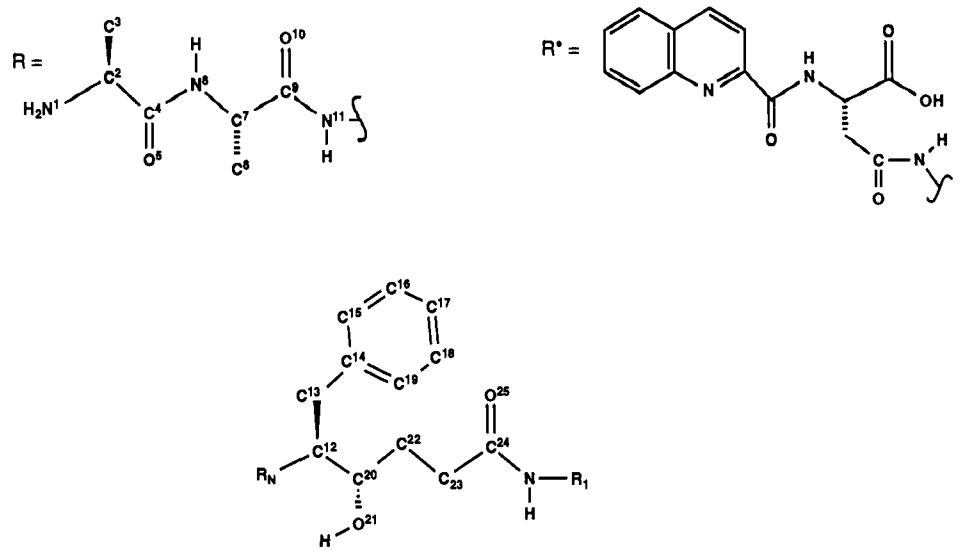
In this section, attention is focused on design considerations that led to the initial syntheses and a brief discussion of the computational approach for estimating relative binding affinities. Other experimental aspects such as synthesis and biochemical testing will be discussed elsewhere. Computational details involving validation and other design experiments along with results and discussion will be presented in the following section.

(a) Design Considerations for Theoretical Analysis. Ample evidence has indicated that peptide-based inhibitors make for poor drug candidates.[26] As a result, we focused our efforts on modifications that would render the molecules less peptidic in nature. Our design process began with the structure of the MVT-101/HIV1 protease complex[3c] (K_i = 780 nM) available from the protein data bank. A model was built of a recently reported hydroxyethylene inhibitor[5] and docked in the active site of the enzyme. Using this model as a guide, the C-terminal Val-Val methyl ester was removed and replaced with a diphenhydramine moiety. This simple scaffold served to fill the p2′ and p3′ sites in the protein with phenyl groups. The N-terminal portion of the molecule was not altered in order to preserve aqueous solubility critical to the crystallization process. Figure 2 shows the first compound (**I**) designed in this manner. This compound (**I**) was synthesized and found to be a modest inhibitor of HIV protease with a K_i of 1.67 μM. The crystal structure of the HIV protease complexed with this compound was then solved. Concurrently, compound **II** was designed and selected for synthesis based on space-filling considerations and molecular mechanics calculations which indicated better enthalpic interaction with the protein. After syntheis, biochemical testing revealed that compound **II** is a more potent inhibitor than **I** by a factor of 8 (K_i = 0.2 μM). Crystallographic analysis of the complex of **II** with HIV protease (shown in Figure 3a as the trace of protein main chain) indicated that the increased binding, relative to compound **I**, was probably due to a combination of increased hydrophobic contact in the p3′ side-chain pocket and an additional hydrogen bond from the indole NH to Gly 27 carbonyl oxygen via an intervening water molecule. This water bridge is shown in Figure 3b. The modeled configuration of inhibitors **II**–**VII** is the *R*-configuration. It was chosen because the p1′ and p2′ side-chain binding pockets are oriented in such a manner that the *S* isomer would overlap with the protein. This was based on a visual inspection of the model with the inhibitor **II** bound in the protein. Later, crystal structure analysis of the complex of inhibitor **II** confirmed that the protein binds only the *R* isomer. All these experimental aspects of our work (synthesis, crystallography, and biochemistry) will be described in a following paper.[27] Using the crystal structures of the HIV1 protease complexed with compounds **I** and **II**, we then developed a TCP[15,18,19] protocol to screen the proposed inhibitors.

(b) Thermodynamic Cycle Perturbation Approach. The TCP approach may be described as a method for computing the relative chanles of free energy for a binding process by the construction of nonphysical paths connecting the desired initial and terminal states. This approach enables the calculation of relative change in binding free energy difference ($\Delta\Delta G_{bind}$) between two related compounds, by computationally simulating the "mutation" of one to the other. The relative solvation free energy change for two substrates is computed using the first part of the cycle shown in Figure 4, represented in the following equation:

$$\Delta G_3 - \Delta G_4 = \Delta G_{aq} - \Delta G_{gas} = \Delta\Delta G_{sol} \quad (1)$$

The relative binding free energy change for the two substrates is computed using the second part of the cycle



(a)

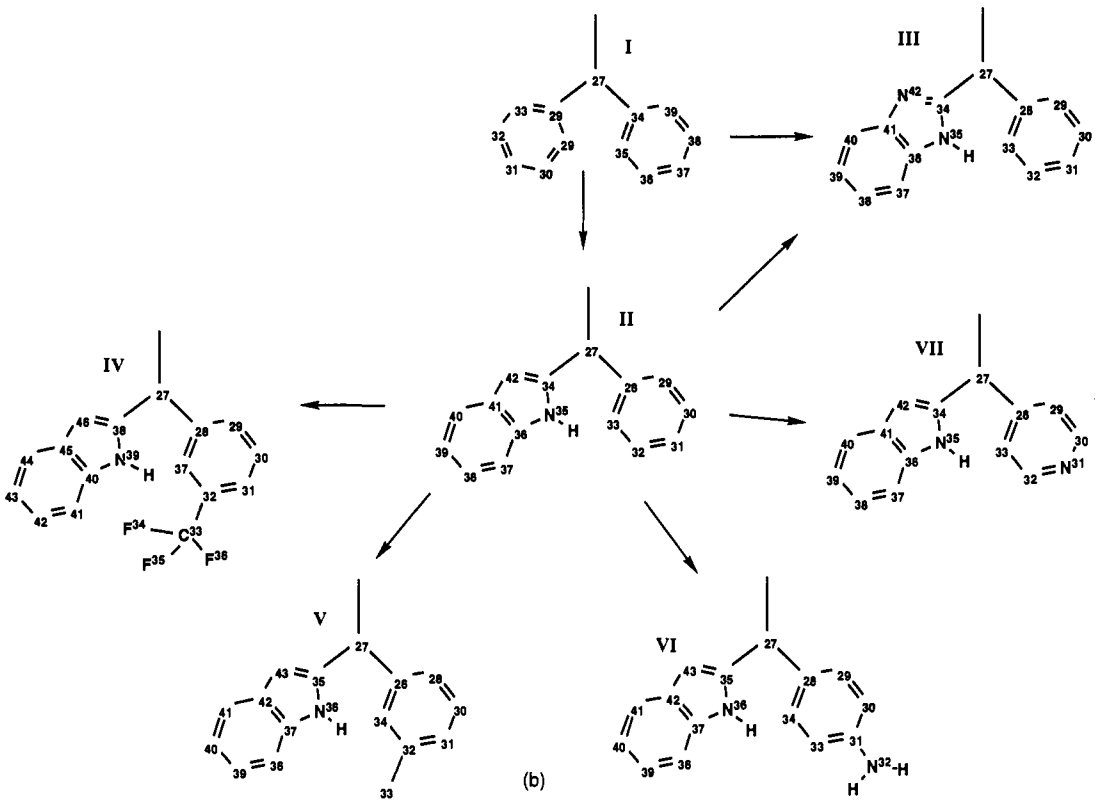


Figure 2. Compounds analyzed in the present study by the TCP approach. All the compounds (**I–VII**) are described by the basic structure shown in a, with R_N = R. The substituents at position R_1 that distinguish various compounds are indicated in b. Additional design modifications at the N-terminal group involved the replacement of NH_2-Ala-Ala with asparagine–quinoline group (R*) in the case of compounds with R_1 = **II**, **IV** and **V** (i.e., R_N = R*). These N-terminal-modified variants are referred as **II***, **IV*** and **V*** in the text. The arrows correspond to the "mutations" involved in the free energy calculations by the TCP approach in conjunction with molecular dynamics simulations.

shown in Figure 4, represented by the following equation:

$$-k_B T \ln(k2/k1) = \Delta G_2 - \Delta G_1 = \Delta G_{com} - \Delta G_{aq} = \Delta\Delta G_{bind} \quad (2)$$

where the experimentally measured binding constants $k1$ and $k2$ refer to the reactions involving S1 and S2 inhibitors, respectively, and their corresponding free energy differences are ΔG_1 and ΔG_2. Here, k_B is the Boltzmann constant and T is the absolute temperature.

The free energy change for converting S1 into S2 is computed by perturbing the Hamiltonian of reactant (initial) state S1 into that of the product (final) state S2. This transformation is accomplished through a parametrization of terms comprising the interaction potentials of the system with a change of state variable that maps onto

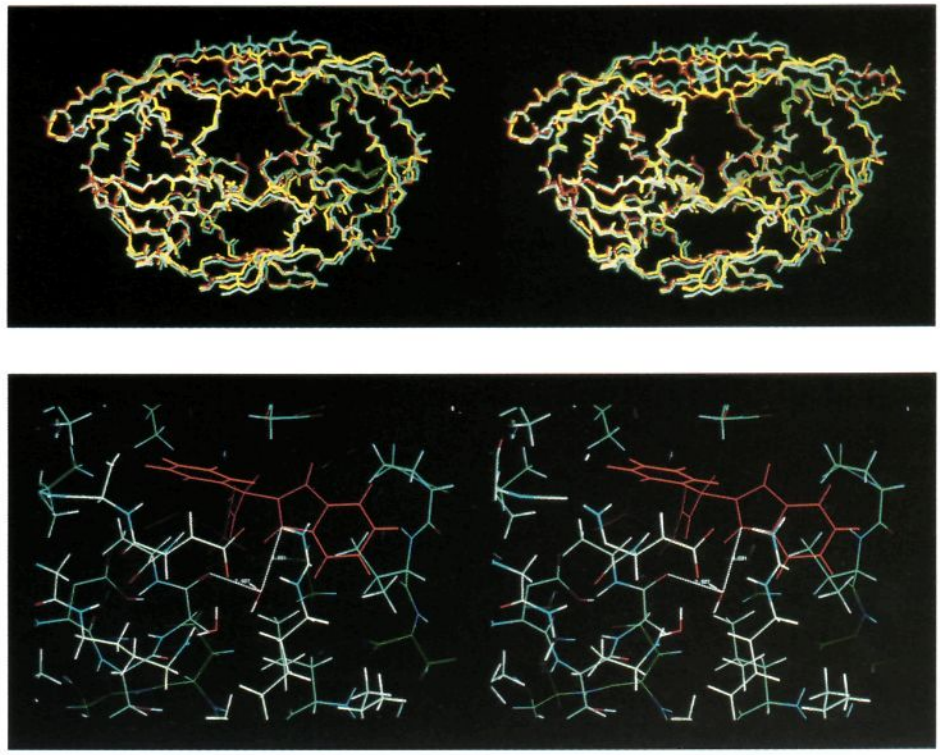


Figure 3. (a) Stereoview of comparison of the main chain of the X-ray structure of the HIV1 protease complex with compound **II** (red) with the main chains of the minimized complex (yellow) and a 20 ps average dynamical structure of the same complex of HIV1 protease (green). (b) Stereoview of the active-site geometry of the crystal structure (in half-bond color) of the HIV1 protease complexed with compound **II** (with the indole and phenyl groups shown in red) as revealed by X-ray crystallography.

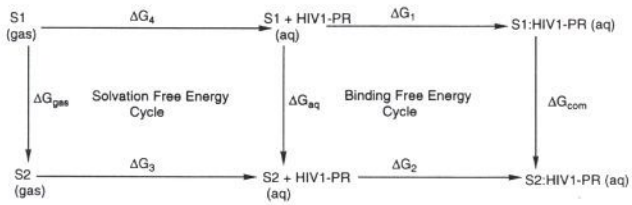


Figure 4. Thermodynamic cycles used in this study. Each reaction shown in the two cycles is reversible; the change in free energy is computed by taking the difference between free energy of the state at the end of the arrow and at the start of that arrow. S1 and S2 refer to the initial and final states of a mutation, which correspond to different pairs of inhibitors described in Figure 2.

reactant and product states when that variable is 0 and 1, respectively. The total free energy change for the mutation from the initial to the final state is computed by summing "incremental" free energy changes over several windows visited by the state variable changing from 0 to 1.

Computational Details

Structural Comparison. Initially, an energy minimization (500 steps of steepest descent followed by 2000 steps of conjugate gradient optimization) of the HIV1-PR/**II** complex was performed. This was followed by a 20 ps MD simulation for equilibration. The average "dynamical" structure of the complex was computed from the MD simulation. For time steps of 1 and 2 fs in MD simulations, the rms deviations from the crystal structure were 1.01 and 1.05 Å for backbone atoms and 1.5 Å and 1.55 for side-chain atoms, respectively. Figure 3a shows the structural comparison of the averaged "dynamical" structure from a 20-ps dynamics trajectory (with a time step of 2 fs) with the X-ray structure of the dimer. Figure 3a clearly shows larger deviations in flap region (with RMS deviations of 1.30 and 1.75 Å for backbone and side-chain atoms, respectively) when compared to the rest of the protein. This is primarily due to the flexibility of the protein in the flap region. Nevertheless, the dynamical structure is a good model for calculating relative free energy changes between two similar inhibitors. Since both time steps (1 and 2 fs) yielded good agreement with the X-ray structure, we used the larger time step of 2 fs for all free energy calculations reported, in the interest of saving *computer time*.

Model and Parameters. The X-ray structures of HIV1 protease complexes with inhibitors **I** and **II** were solved in the orthorhombic space grop $P2_12_12_1$ with the unit cell dimensions of a = 66.4 Å, b = 92.4 Å, and c = 28.8 Å. The resolutions of data were 2.6 and 2.5 Å, respectively. Both the structures were refined using Xplor[20] to the crystallographic R factor of 0.163 and 0.158, respectively. Details of these X-ray crystallographic analyses will be published elsewhere.

All molecular dynamics, mechanics, and TCP calculations were carried out on IBM 580 workstations with the AMBER program[28] using an all atom force field[28] and SPC/E model potential[29,30] to describe water interactions. Electrostatic charges and parameters for the standard residues were taken from the AMBER database. For nonstandard solute atoms, partial charges were obtained by fitting with CHELP[31] from ab initio 3-21G*//6-31G* basis set level wave functions calculated with Gaussian88.[32] These charges are shown in Table 1. All equilibrium bond lengths, bond angles, and dihedral angles for nonstandard residues were taken from *ab initio* optimized geometries. Missing force field parameters were estimated from similar

Table 1. ESP Fitted ab Initio Atomic Charges (in Electron Units) for the Common Part of All the Ligands and Various Fragments Shown and Labeled in Figure 2 (a and b)[a]

(a) Atomic Charges for the Common Part of the Ligands

atom	charge	atom	charge	atom	charge	atom	charge
N1	−0.7150	H	0.1300	H	0.0480	H	−0.0120
H	0.2520	H	0.1300	C8	−0.0980	O21	−0.5800
H	0.2520	C14	−0.1000	H	0.0380	H	0.3000
C2	0.0350	C15	−0.1500	H	0.0380	C22	−0.3299
H	0.0480	H	0.1500	H	0.0380	H	0.1200
C3	−0.0980	C16	−0.1500	C9	0.6160	H	0.1200
H	0.0380	H	0.1500	O10	−0.5040	C23	−0.1721
H	0.0380	C17	−0.1500	N11	−0.6300	H	0.1000
H	0.0380	H	0.1500	H	0.3400	H	0.1000
C4	0.6160	C18	−0.1500	C12	0.5400	C24	0.6820
O15	−0.5040	H	0.1500	H	−0.5000	O25	−0.5300
N6	−0.4630	C19	−0.1500	C13	−0.4000	N26	−0.6100
H	0.2520	H	0.1500			H	0.2900
C7	0.0350	C20	0.2420				

(b) Atomic Charges for Ligand Fragments Identified and Labeled in Figure 2b

I		II		III		IV		V		VI		VII	
atom	charge	atom	charge	atom	charge	atom	charge	atom	charge	atom	charge	atom	charge
C27	0.0330	C27	0.0459	C27	0.2000	C27	0.1200	C27	0.1200	C27	0.1200	C27	0.1200
H	0.0870	H	0.1000	H	0.1000	H	0.1000	H	0.1000	H	0.1000	H	0.1000
C28	0.1000	C28	0.1000	C28	0.1000	C28	0.1000	C28	0.1000	C28	0.1000	C28	0.2500
C29	−0.1300	C29	−0.1300	C29	−0.1300	C29	−0.1300	C29	−0.1300	C29	−0.1300	C29	−0.4000
H	0.1300	H	0.1300	H	0.1300	H	0.1300	H	0.1300	H	0.1300	H	0.1500
C30	−0.1300	C30	−0.1300	C30	−0.1300	C30	−0.1300	C30	−0.1300	C30	−0.2800	C30	0.4800
H	0.1300	H	0.1300	H	0.1300	H	0.1300	H	0.1300	H	0.1400	H	0.0200
C31	−0.1300	C31	−0.1300	C31	−0.1300	C31	−0.0800	C31	−0.2500	C31	0.4600	N31	−0.6500
H	0.1300	H	0.1300	H	0.1300	H	0.1350	H	0.1350	N32	−0.9000	H	0.0000
C32	−0.1300	C32	−0.1300	C32	−0.1300	C32	−0.1950	C32	0.2300	H	0.3600	C32	0.4800
H	0.1300	H	0.1300	H	0.1300	C33	0.9100	C33	−0.2400	H	0.3600	H	0.0200
C33	−0.1300	C33	−0.1300	C33	−0.1300	F34	−0.2750	H	0.0800	C33	−0.2800	C33	−0.4000
H	0.1300	H	0.1300	H	0.1300	F35	−0.2750	H	0.0800	H	0.1400	H	0.1500
C34	0.1000	C34	0.1371	C34	0.4500	F36	−0.2750	H	0.0800	C34	−0.1300	C34	0.1000
C35	−0.1300	N35	−0.4650	N35	−0.7200	C37	−0.0800	C34	−0.2500	H	0.1300	N35	−0.6000
H	0.1300	H	0.3430	H	0.4000	H	0.1350	H	0.1350	C35	0.1000	H	0.3800
C36	−0.1300	C36	0.1650	C36	0.3800	C38	0.1000	C35	0.1000	N36	−0.6000	C36	0.4200
H	0.1300	C37	−0.2320	C37	−0.3500	N39	−0.6000	N36	−0.6000	H	0.3800	C37	−0.3500
C37	−0.1300	H	0.1400	H	0.1600	H	0.3800	H	0.3800	C37	0.4200	H	0.1600
H	0.1300	C38	−0.1170	C38	−0.1000	C40	0.4200	C37	0.4200	C38	−0.3500	C38	−0.1000
C38	−0.1300	H	0.1090	H	0.0900	C41	−0.3500	C38	−0.3500	H	0.1600	H	0.0900
H	0.1300	C39	−0.1170	C39	−0.1000	H	0.1600	H	0.1600	C39	−0.1000	C39	−0.1000
C39	−0.1300	H	0.1090	H	0.0900	C42	−0.1000	C39	−0.1000	H	0.0900	H	0.0900
H	0.1300	C40	−0.2800	C40	−0.3400	H	0.0900	H	0.0900	C40	−0.1000	C40	−0.2000
		H	0.1600	H	0.2000	C43	−0.1000	C40	−0.1000	H	0.0900	H	0.1600
		C41	0.2700	C41	0.3800	H	0.0900	H	0.0900	C41	−0.2000	C41	0.1500
		C42	−0.3480	N42	−0.6200	C44	−0.2000	C41	−0.2000	H	0.1600	C42	−0.4000
		H	0.2000			H	0.1600	H	0.1600	C42	0.1500	H	0.2000
						C45	0.1500	C42	0.1500	C43	−0.4000		
						C46	−0.4000	C43	−0.4000	H	0.2000		
						H	0.2000	H	0.2000				

[a] Hydrogens are not numbered and carbons are not explicitly labeled in Figure 2b.

chemical species within the AMBER database. All parameters are available from the author upon request.

For computing solvation free energy, the solute was solvated with SPC/E water using the AMBER box option, and all the solvent 10.0 Å away from any of the solute atoms were removed. Any water molecules located less than 2.5 Å from the solute atom were removed. Aqueous-phase dynamics simulations were carried out in a rectangular box using periodic boundary conditions in all directions. Newton's equations of motion for all the atoms were solved using the Verlet algorithm[33] with a 2-fs time step,[10] and SHAKE was used for constraining all bond lengths.[34] Constant temperature (N, P, and T ensemble) was maintained by velocity scaling of all atoms in the system. Nonbonded interaction energies were calculated using 10.0-Å residue-based cutoff.

For the protein complex simulations (the second cycle of Figure 4), it was necessary to generate all the hydrogen atom coordinates first (for the all-atom force field to be applied) since they are not determined by X-ray crystallography. We used the EDIT module of AMBER to add hydrogens for the protein dimer and crystallographic waters. In all calculations, the water molecule shown in Figure 3b is present as part of the environment. One of the aspartic acids in the catalytic dyad (Asp 124) was protonated in all simulations. The histidine protonation at one or both ring nitrogens was deduced from hydrogen bonding and other features of the environment. The total charge on HIV1 protease was +5 e. No counterions or changes in the customary charge of protein residues were used. While such an electrostatic model is far from ideal, alternatives sometimes adopted have their own drawbacks. The development and testing of better methods is an active and crucial area of research. The entire system was immersed in a 25.0-Å-radius sphere of solvent from the center of mutating groups, which was subject to a half-

harmonic restraint near the boundary to prevent evaporation. During the simulation, all atoms of the protein were fixed beyond 25.0 Å. All nonbonded interactions involving the inhibitors and the charged residues of the protein were computed with infinite cutoff. A 10.0-Å nonbonded residue-based cutoff was used for other residues of the system. The algorithm for the complex simulation was identical to the solvent simulation, except for the absence of periodic boundary conditions in the former.

Results and Discussion

The mutation of compound **I** to compound **II** shown in Figure 2 was performed initially in order to validate our protocol. Since this mutation involves significant changes in the ligand structure, we have used the "thread" method[23,24] to accomplish nonphysical transformation of one molecule into a related molecule. A single topology is defined for those atoms which are identical in both molecules in the sense that force constants and equilibrium geometries are the same (partial charges can vary). For the portion of the molecule which must be transformed, both the starting (reactant) and ending (product) topologies are defined with their correct geometries, one beginning and the other ending the simulation entirely as dummy atoms. Dummy atoms are identical to real atoms except for their Lennard-Jones parameters and charges which are set to zero. At intermediate points during the transformation, all atoms in both topologies have fractional Lennard-Jones parameters and charges. Molecules with both topologies interact with the environment, but not with each other. For the present mutation (**I** → **II**), the phenyl group of molecule **I** and indole group of molecule **II** are threaded together at C_α positions: one is real and the other is dummy; both interact with the environment and not with each other.

In all free energy simulations, the system was initially minimized (using 500 steps of steepest descent and 2000 steps of conjugate gradient methods) and then equilibrated for 20 ps. A two-stage procedure was used to obtain relative free energy differences from the molecular dynamics simulations. During the first stage, the charges of the reactant atoms are turned off while the Lennard-Jones parameters of the product atoms are turned on. During the second stage, the Lennard-Jones parameters of the reactant atoms are turned off while the charges of the product atoms are turned on. This procedure has been used earlier[23,24] to obtain better convergence. Each stage of the simulation was performed using 101 windows, each window comprising 1 ps of equilibration and 2 ps of data collection except for the last 10 windows which needed longer simulations of 3 ps of equilibration and 6 ps of data collection to compute the free energy differences using the thread method because shorter simulations possibly lead to problems in convergence.[10] Thus, a molecular dynamics simulation of 726 ps run was needed for the complete mutation. Each mutation followed the double-wide sampling procedure, and the results reported are based on the averages from the backward and forward simulations of the mutation. On the average, about 45 min of CPU time per picosecond of molecular dynamics simulation was needed for the protein–inhibitor complex on an IBM 580 workstation. For the solvent simulations, the corresponding CPU time needed was about 10 min per picosecond of dynamics simulations.

Table 2. Relative Binding Free Energy and Solvation Free Energy Differences (kcal/mol) Obtained from TCP Simulations for the Inhibitors Shown in Figure 2[a]

	mutation	$\Delta\Delta G_{solv}$ (calcd)	$\Delta\Delta G_{bind}$ (calcd)	$\Delta\Delta G_{bind}$ (exp)
1.	**I** → **II**	−3.1 ± 0.4	−2.3 ± 0.6	−1.3 ± 0.30
2.	**II** → **I**	3.5 ± 0.4	1.9 ± 0.6	1.3 ± 0.30
3.	**I** → **III**	−5.0 ± 0.6	1.2 ± 0.8	0.7 ± 0.20
4.	**II** → **III**	−2.0 ± 0.5	1.3 ± 0.6	1.95 ± 0.31
5.	**II** → **IV**	−1.0 ± 0.4	0.2 ± 0.5	−0.16[b] ± 0.24
6.	**II** → **V**	0.04 ± 0.2	0.4 ± 0.5	−0.06[b] ± 0.26
7.	**II** → **VI**	−3.2 ± 0.4	1.1 ± 0.6	
8.	**II** → **VII**	−0.9 ± 0.3	0.8 ± 0.5	

[a] The "mutations" indicated by "→" correspond to the following computational reactions indicated in Figure 2 (performed in the aqueous phase). For each mutation indicated by X → Y, the free energy difference is reported as $\Delta G(Y) - \Delta G(X)$. [b] These numbers are calculated from the binding data on a different N-terminal group, asparigine–quinoline replacing NH_2-Ala-Ala in the compounds **II**, **IV**, and **V**, given in Table 3.

In the case of **I** → **II**, the availability of X-ray structures for both complexes (of **I** and **II**) made it possible to test the reversibility of the mutation by starting with the X-ray structure of the complex with **II** and performing the mutation **II** → **I**. Table 2 shows the results of the TCP calculations. As may be seen from the table, TCP results for forward (**I** → **II**) and backward (**II** → **I**) mutations are in excellent agreement. Relative to compound **I**, compound **II** was predicted to bind with better potency (of ca. 2.1 kcal/mol), and the agreement with experimental result is seen to be very good, thus validating our protocol for TCP calculations. Though compound **II** has a greater desolvation penalty than compound **I**, it more than compensates with favorable interactions in the complex such as an additional hydrogen bond shown in Figure 3b and hydrophobic interactions, as expected initially. Obviously, the hydrogen-bond bridge via the water molecule shown in Figure 3b contributes to greater stability of the complex with **II** relative to complex with **I**, though that water is present in both cases, as in all other structures. The convergence of the mutation was tested by comparing the relative change in the free energies of compound **II** to **I** with two different lengths of molecular dynamics simulations.

Next, compound **III** was designed by the replacement of a carbon with nitrogen in the indole (of compound **II**) as a way of enhancing electrostatic interaction with the main-chain part of Gly 48. However, the desolvation penalty for compound **III** is expected to be greater than that of compound **II** because of the hydrophilic nitrogen in **III**, and the effect of these opposing tendencies on binding affinity is hard to predict without a TCP calculation. We therefore performed two mutations (**I** → **III** and **II** → **III**) prior to the synthesis and testing of compound **III**. For the mutation **I** → **III** we used the thread method as in the former mutation because of the large change involved. For the other mutation **II** → **III**, a single topology was used for both molecules **II** and **III**. In this commonly used method, appropriate reactant atoms will change to appropriate product atoms which includes changes in the geometry, partial charges, and van der Waal's parameters.

In Table 2, we report calculated relative differences in free energy of binding, and the corresponding experimental values computed from the binding data are shown in Table 3. Examination of Tables 2 and 3 reveals that this set of TCP calculations shows good agreement with measured binding data. Note that both mutation methods for

Table 3. Binding Constants for the Protease Inhibitors Shown in Figure 2

compd	binding constant (K_i) (μM)	compd	binding constant (K_i) (μM)
I	1.67 ± 0.25	IV[a]	0.033 ± 0.011
II	0.20 ± 0.07	V[a]	0.039 ± 0.014
II[a]	0.043 ± 0.01	VI	
III	5.37 ± 1.45	VII	

[a] These numbers are measured on the basis of a different N-terminal group, asparagine–quinoline replacing NH_2-Ala-Ala in the compounds II, IV, and V.

handling topological changes yielded satisfactory results. Relative to compound III, compound I is predicted to bind with better potency (of ca. 1.2 kcal/mol) and compound III is predicted to bind with lower potency (of ca. 1.3 kcal/mol) than compound II. Error bars were estimated for each window by dividing the window statistics into four groups and computing the standard deviation for the indicated free energy change. The root mean square of these window errors is reported in Table 2 as a measure of the statistical uncertainity in the result for each mutation.

Since compound III was not as potent as II, we focused on the other phenyl ring in the p2′ pocket of the active site, in order to improve on our lead compound II. In the crystal structure of complexes with I and II, there was some unfilled space that could be reached from the phenyl ring. Hydrophobic interactions could be obtained by suitable additions at the meta position of the phenyl ring in the p2′ pocket (IV and V) whereas hydrogen bonds or good electrostatic interactions could be obtained with Asp 129 by the addition of polar groups at the para position of the phenyl ring (VI and VII). On the basis of intial modeling of several analogs, we selected four candidates for TCP calculations and performed corresponding mutation experiments (II → IV, V, VI, and VII) using the TCP approach. The starting structures were obtained using the coordinates of the crystal structure of HIV1-PR complexed with compound II. These mutations (II → IV, V, VI, and VII) involve relatively minor changes in the ligand structures; hence, we employed a single topology for the reactant and product molecules as in the earlier mutation (II → III).

From Table 2 it may be seen that the desolvation penalty relative to compound II is higher for compounds VI (by 3.2 kcal/mol), IV (by 1.0 kcal/mol), and VII (by 0.9 kcal/mol), whereas it is slightly lower for compound V because of the hydrophobic methyl substituent. In Table 2 a negative $\Delta\Delta G$ equates to more favorable solvation free energy for compound S2 (or Y) and therefore a greater desolvation penalty upon binding. Therefore, it is important to take the desolvation penalty into consideration in the design of new analogs of a lead compound. However, a design modification that *increases* the desolvation penalty might improve binding if the additional polar groups form good hydrogen bonds and/or electrostatic interactions with the enzyme, as in the case of I → II. For another example, in the case of JG365 inhibitor[9] of the HIV1 protease, the deletion of one amino acid, valine, *decreases* binding potency though it costs *less* to desolvate the inhibitor without valine.[10] This is because the main-chain atoms of valine form good hydrogen bonds with the protease. In the cases of compounds IV and V, $\Delta\Delta G_{bind}$ predicted from TCP calculation is close to zero. The presence of the trifluoromethyl group (compound IV) improves binding; however, this could not overcome the desolvation penalty. We still felt that there is some possibility of improving binding with these compounds because of the error bounds on these free energy estimates. We therefore made some design modifications at the N-terminal position involving the replacement of NH_2-Ala-Ala with the aparagine–quinoline moiety. The rationale for replacing the N-terminal NH2-Ala-Ala with asparagine–quinoline was based on the desire to obtain compounds that had *in vitro* antiviral activity. Since the terminal amine would likely be protonated at neutral pH, which would preclude passive transport into cells, it was replaced with the uncharged asparagine–quinoline group. These modified structures (II*, IV*, and V*) were synthesized and tested. No TCP calculations were performed for these N-terminal modified variants. Experimental results (reported as free energy difference between II* and IV* or V* in Table 2, i.e., II* → IV*, II* → V*) show slight improvement in binding. Though TCP predictions (II → IV, II → V) are based on the original structures (reported as free energy difference between II and IV or V in Table 2), these are comparable to the experimental results because each member of a given pair (II → IV, II* → IV*, II → V, or II* → V*) has the same N-terminal moiety as the other member of that pair. As may be seen from Table 2, TCP results indicate a slight net loss of binding when II is replaced with IV and V. Discrepancy between TCP calculations and experimental results could due to the modifications at the N-terminal site as well as errors both in the experimental measurements and in calculations. Overall, TCP results are in good agreement with experiments. In the case of compounds VI and VII, TCP calculations indicated weaker binding (by 1 kcal or more) relative to compound II, primarily because of the larger desolvation penalty on these ligands, and hence these compounds were not synthesized.

Conclusion

In summary, successful application of a computer-assisted drug design paradigm is described in this report. As part of this paradigm, the TCP approach has been useful in guiding the synthesis of a series of promising inhibitors for the HIV1 protease. More importantly, once validated, this approach was used in a predictive sense to prioritize design ideas and eliminate the need to synthesize poor inhibitors, thus accelerating the drug design cycle. In all cases where experimental data is available, predictions based on this approach have been shown to be correct within an error margin of 1 kcal/mol. This study also brought to light the role played by solvation free energy in binding. The energetic cost of desolvation for the addition of polar groups to an inhibitor has to be compensated and overcome by stronger ligand–protein interactions, if the goal is to design a stronger inhibitor. Careful design, modeling, and TCP calculations enable the achievement of that goal. We are continuing to develop this drug-design paradigm in both the HIV and other drug discovery areas.

References

(1) Tomaselli, A. G.; Howe, J. W.; Sawyer, T. K.; Wlodawer, A.; Heinrickson, R. L. The Complexities of AIDS: An Assessment of the HIV Protease as a Therapeutic Target. *Chimica Oggi* **1991**, *9*, 6–27.

(2) (a) Wlodawer, A.; Erickson, J. W. Structure-Based Inhibitors of HIV-1 Protease. *Annu. Rev. Biochem.* **1993**, *62*, 543–85. (b) Appelt, K. Crystal Structures of HIV-1 Protease Complexes. *J. Comput. Aided Mol. Des.* In press.

(3) (a) Wlodawer, A.; Miller, M.; Jaskolski, M.; Sathyanarayana, B. K.; Baldwin, E.; Weber, I.; Selk, L.; Clawson, L.; Schneider, J.; Kent, S. Conserved Folding in Retroviral Proteases: Crystal Structure of a Synthetic HIV-1 Protease. *Science* **1989**, *245*, 616–621. (b) Navia, M. A.; Fitzgerald, P. M. D.; McKeever, B. M.; Leu, C.; Heimbach, J. C.; Herber, W. K.; Sigal, I. S.; Darke, P. L.; Springer, J. P. Three Dimensional Structure of Aspartyl Protease from Human Immunodeficiency Virus HIV-1. *Nature (London)* **1989**, *337*, 615–620. (c) Miller, M.; Schneider, J.; Sathyanarayana, B. K.; Toth, M. V.; Marshall, G. R.; Clawson, L.; Selk, L.; Kent, S. B. H.; Wlodawer, A. Structure of Complex Synthetic HIV-1 Protease with a Substrate-Based Inhibitor at 2.3 Å Resolution. *Science* **1989**, *246*, 1149–1152.

(4) Erickson, J. W.; Neidhart, D. J.; VanDrie, J.; Kempf, D.; Wang, X. C.; Norbeck, D. W.; Plattner, J. J.; Rittenhouse, J. W.; Turon, M.; Wideburg, N.; Kohlbrenner, W. E.; Simmer, R.; Helfrich, R.; Paul, D. A.; Knigge, M. Design, Activity and 2.8 Å Crystal Structure of a C2 symmetric Inhibitor Complexed to HIV-1 Protease. *Science* **1990**, *249*, 527–533.

(5) Dreyer, G. B.; Metcalf, B. W.; Tomaszek, T. A., Jr.; Carr, T. J.; Chandler, A. C., III, Hyland, L.; Fakhoury, S. A.; Magaard, V. W.; Moore, M. L.; Strickler, J. E.; Debouck, C.; Meek, T. D. Inhibition of Human Immunodeficiency Virus 1 Protease in vitro: Rational Design of Substrate Analogue Inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 9752–9756.

(6) Huff, J. R. HIV Protease: A Novel Chemotherapeutic Target for AIDS. *J. Med. Chem.* **1991**, *34*, 2305–2314.

(7) Swain, A. L.; Miller, M.; Green, J.; Rich, D. H.; Schneider, J.; Kent, S. B. H.; Wlodawer, A. X-ray Crystallographic Structure of a Complex Between a Synthetic Protease of Human Immunodeficiency Virus 1 and a Substrate Based Hydroxyl Ethylamine Inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 8805–8809.

(8) Meek, T. D. Inhibitors of HIV1 protease. *J. Enzyme Inhib.* **1992**, *6*, 65–98.

(9) Rich, D. H. Sun, C.-Q.; Varaprasad, J. V. N.; Pathiasseril, A.; Toth, M. V.; Marshall, G. R.; Clare, M.; Mueller, R. A.; Houseman, K. Effect of Hydroxyl Group Configuration in Hydroxyethyl Amine Dipeptide Isosteres on HIV Protease Inhibition. Evidence for Multiple Binding Modes. *J. Med. Chem.* **1991**, *34*, 1222–1225.

(10) Rami Reddy, M.; Viswanadhan, V. N.; Weinstein, J. N. Relative Free Energy Differences in the Binding Free Energies of Human Immunodeficiency Virus 1 Protease Inhibitors: A Thermodynamic Cycle Perturbation Approach. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 10287–10291.

(11) Rao, B. G.; Tilton, R. F.; Singh, U. C. Free Energy Perturbation Studies on Inhibitor Binding to HIV-1 Proteinase. *J. Am. Chem. Soc.* **1992**, *114*, 4447–4452.

(12) Kuntz, I. D. Structure-Based Strategies for Drug Design and Discovery. *Science* **1992**, *257*, 1078–1082.

(13) Desjarlais, R. L.; Seibel, G. L.; Kuntz, I. D.; Furth, P. S.; Alvarez, J. C.; Ortiz De Montellano, P. R.; DeCamp, D. L.; Babe, L. M.; Craik, C. S. Structure-based design of Nonpeptide Inhibitors Specific for the Human Immunodeficiency Virus 1 Protease. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 6644–6648.

(14) (a) Ferguson, D. M.; Radmer, R. J.; Kollman, P. A. Determination of the Relative Binding Free Energies of Peptide Inhibitors to the HIV-1 Protease. *J. Med. Chem.* **1991**, *34*, 2654–2659. (b) Tropshaw, A. J.; Hermans, J. Application of Free Energy Simulations to the Binding of a Transition State Analogue Inhibitor to HIV Protease. *Protein Eng.* **1992**, *5*, 29–33.

(15) (a) Beveridge, D. L.; DiCapua, F. M. Free Energy via Molecular Simulation. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431–492. (b) McCammon, J. A. Computer-aided Molecular Design. *Science* **1987**, *238*, 486–491.

(16) Harte, W. E., Jr.; Beveridge, D. L. Prediction of the Protonation State of the Active Site Aspartyl Residues in HIV-1 Protease-Inhibitor Complexes vis Molecular Dynamics Simulation. *J. Am. Chem. Soc.* **1993**, *115*, 3883–3886.

(17) (a) Appelt, K.; Bacquet, R. J.; Bartlett, C. A.; Booth, C. L.; Freer, S. T.; Fuhry, M. A. M.; Gehring, M. R.; Hermann, S. M.; Howland, E. F.; Janson, C. A.; Jones, T. R.; Kan, C.; Kathardekar, V.; Lewis, K. K.; Marzoni, G. P.; Matthews, D. A.; Mohr, C.; Moomaw. E. W.; Morse, C. A.; Oatley, S. J.; Ogden, R. O.; Reddy, M. R.; Reich, S. H.; Schoettlin, W. S.; Smith, W. W.; Varney, M. D.; Villafranca, J. E.; Ward, R. W.; Webber, S.; Webber, S. E.; Welsh, K. M.; White, J. Design of Enzyme Inhibitors Using Iterative Protein Crystallographic Analysis. *J. Med. Chem.* **1991**, *34*, 1925–1934. (b) Montgomery, J. A.; Shri Niwas.; Rose, J. D.; Secrist, J. A.; Sudhakar Babu, Y.; Bugg, C. E.; Erion, M. E.; Guida, W. C.; Ealick, S. E. Structure Based Design of Inhibitors of Purine Nucleoside Phosphorylase. 1. 9-(Arylmethyl) Derivatives of 9-Deazaguanine. *J. Med. Chem.* **1993**, *36*, 55–69.

(18) Zwanzig, R. W. High Temperature Equation of State by a Perturbation Method. I. Non-polar Gases. *J. Chem. Phys.* **1954**, *22*, 1420–1426.

(19) (a) Van Gunsteren, W. F.; Weiner, P. K. *Computer Simulation of Biomolecular Systems*; ESCOM Science: Leiden, Netherlands, 1989. (b) Hermans, J. *Molecular Dynamics & Protein Structure*; Hermans, J., Eds.; Polycrystal: West Springs, IL, 1985. (c) Foley, C. K.; Pederson, L. G.; Charifson, P. S.; Darden, T. A.; Wittinghofer, A.; Oai, E. F.; Anderson, M. W. Simulation of the Solution Structure of the H-ras p21-GTP Complex. *Biochemistry* **1992**, *31*, 4951–4959.

(20) Brunger, A. T.; Kuriyan, J.; Karplus, M. Crystallographic R Factor Refinement by Molecular Dynamics. *Science* **1987**, *235*, 458–461.

(21) Cummins, P. L.; Ramnarayan, K.; Singh, U. C.; Gready, J. E. Molecular Dynamics/Free Energy Perturbation Study on the Relative Affinities of the Binding of Reduced and Oxidized NADP to Dihydrofolate Reductase. *J. Am. Chem. Soc.* **1991**, *113*, 8247–8256.

(22) Hausheer, F. H.; Singh, U. C.; Saxe, J. D.; Flory, J. P.; Tufto, K. B. Thermodynamic and Conformational Characterization of 5-Methylcytosine- versus Cytosine-Substituted Oligomers in DNA Triple Helices: Ab Initio Quantum Mechanical and Free Energy Perturbation Studies. *J. Am. Chem. Soc.* **1992**, *114*, 5356–5362.

(23) Rami Reddy, M.; Bacquet, R. J.; Zichi, D.; Matthews, D. A.; Welsh, K. M.; Jones, T. R.; Freer, S. Calculation of Solvation and Binding Free Energy Differences for Folate-Based Inhibitors of the Enzyme Thymidylate Synthase. *J. Am. Chem. Soc.* **1992**, *114*, 10117–10122.

(24) Singh, U. C.; Benkovic, S. J. A Free Energy Perturbation Study of the Binding of Methotrexate to Mutants of Dihydrofolate Reductase. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 9519–9523.

(25) Gao, J.; Kuczera, K.; Tidor, B.; Karplus, M. Hidden Thermodynamics of Mutant Proteins: A Molecular Dynamics Analysis. *Science* **1989**, *244*, 1069–1072.

(26) Greenlee, W. J. Renin Inhibitors. *Med. Res. Rev.* **1990**, *10*, 173–236.

(27) Varney, M. D.; K. Appelt, K.; Kalish, V.; Rami Reddy, M.; Tatlock, J.; Wu, B. W.; Palmer, C. Unpublished results.

(28) (a) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagoha, G.; Profeta, S., Jr.; Weiner, P. K. A New Force Field for Molecular Mechanical Simulation. *J. Am. Chem. Soc.* **1984**, *106*, 765–784. (b) Singh, U. C.; Weiner, P. K.; Caldwell, J. K.; Kollman, P. A. *AMBER Version 3.0*; University of California at San Francisco: San Francisco, 1986.

(29) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271.

(30) Rami Reddy, M.; Berkowitz, M. The Dielectric Constant of SPC/E Water. *Chem. Phys. Lett.* **1989**, *155*, 173–176.

(31) Chirlian, L. E.; M. M. Francl, M. M. Atomic Charges Derived from Electrostatic Potentials: A Detailed Study. *J. Comput. Chem.* **1987**, *8*, 894–905.

(32) Frisch, M. J.; Head-Gordon, M.; Schlegel, H. B.; Raghavachari, K.; Binkley, J. S.; Gonzalez, C.; Defrees, D. J.; Fox, D. J.; Whiteside, R. J.; Seeger, R.; Melius, C. F.; Baker J.; Martin, R.; Kahn, L. R.; Stewart, J. J. P.; Fluder, E. M.; Topiol, S.; Pople, J. A. *Gaussian 88*; Gaussian, Inc.: Pittsburgh, PA, 1988.

(33) Verlet, L. Computer Experiments on Classical Fluids. I. Thermodynamic Properties of Lennard-Jones Molecules. *Phys. Rev.* **1967**, *159*, 98–103.

(34) Ryckart, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.