*Journal of Computer-Aided Molecular Design*, **13:** 109–121, 1999.
*KLUWER/ESCOM*



# Predicting relative binding affinities of non-peptide HIV protease inhibitors with free energy perturbation calculations

Margaret A. McCarrick* & Peter A. Kollman**
*University of California at San Francisco, San Francisco, CA 94143-0446, U.S.A.*

Received 13 June 1997; Accepted 18 August 1998

*Key words:* free energy calculations, molecular dynamics

**Summary**

The relative binding free energies in HIV protease of haloperidol thioketal (THK) and three of its derivatives were examined with free energy calculations. THK is a weak inhibitor ($IC_{50} = 15\ \mu M$) for which two cocrystal structures with HIV type 1 proteases have been solved [Rutenber, E. et al., J. Biol. Chem., 268 (1993) 15343]. A THK derivative with a phenyl group on C2 of the piperidine ring was expected to be a poor inhibitor based on experiments with haloperidol ketal and its 2-phenyl derivative (Caldera, P., personal communication). Our calculations predict that a 5-phenyl THK derivative, suggested based on examination of the crystal structure, will bind significantly better than THK. Although there are large error bars as estimated from hysteresis, the calculations predict that the 5-phenyl substituent is clearly favored over the 2-phenyl derivative as well as the parent compound. The unfavorable free energies of solvation of both phenyl THK derivatives relative to the parent compound contributed to their predicted binding free energies. In a third simulation, the change in binding free energy for 5-benzyl THK relative to THK was calculated. Although this derivative has a lower free energy in the protein, its decreased free energy of solvation increases the predicted $\Delta\Delta G_{(bind)}$ to the same range as that of the 2-phenyl derivative.

## Introduction

Human immunodeficiency virus (HIV) protease is an aspartic acid protease that has been shown to be essential for the development of mature infective virus particles [1]. It was the first HIV structural target for which the three-dimensional structure was known [1]. Since X-ray crystal structures for HIV-1 protease and its complexes with inhibitors were first reported [2], a number of molecular dynamics and free energy simulations have been reported [3, 4]. Thus far, good agreement has been obtained for the cases in which experimental inhibitor binding free energy differences have been available for comparison with calculations.

In an important contribution to structure-based drug design, Desjarlais et al. [5] used the DOCK program to search a crystallographic database for molecules complementary in shape to the HIV-1 protease active site region. One of the high-scoring structures was bromoperidol, a close relative of the antipsychotic drug haloperidol. Haloperidol displayed weak HIV-1 protease inhibition, while exhibiting much less inhibition for two other aspartic proteases, renin and pepsin [5]. The thioketal derivative of haloperidol, which we refer to as THK in this text (Figure 1), inhibits HIV-1 protease with an $IC_{50}$ of 15 $\mu$M, a 10-fold improvement over haloperidol itself [6]. However, to be a serious drug candidate, inhibition should be in the low nanomolar range. We have attempted to use free energy perturbation to assist in identifying modifications of the structure of THK which will lead to tighter binding in HIV protease.

Haloperidol thioketal has been cocrystallized with wild-type HIV-1 protease and with a mutant ($Gln^7 \rightarrow$ Lys) [5] of HIV-1 protease [6]. The crystal structure of HIV-1 wild-type with THK contained one inhibitor molecule in the active site in an orientation quite different from the DOCK prediction. Many derivatives were synthesized at the University of California at San

*Present address: Berlex Biosciences, 15049 San Pablo Avenue, Richmond, CA 94804-0099, U.S.A.
**To whom correspondence should be addressed.

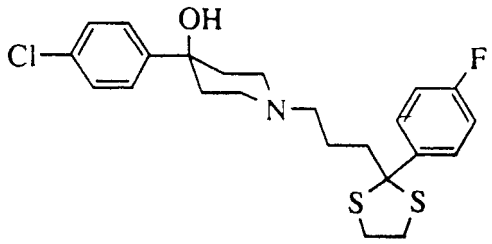


*Figure 1.* Haloperidol thioketal (THK).

Francisco (UCSF) based on this structure with little improvement in binding affinity. A crystal structure of THK complexed with the Q7K mutant protease showed a completely different mode of binding, even though the $K_i$ and $IC_{50}$ of this mutant were the same as those of wild-type HIV protease. This structure contained two inhibitor molecules per active site, both with 100% occupancy, in an unsymmetrical orientation. One of the THK molecules, which we will refer to as THK1 (labeled residue 201 in Figure 2) bound in a similar fashion to the DOCK prediction: its long axis was in approximately the same orientation, and its hydroxyl group appeared to displace the water molecule which is often observed bound to the active site aspartic acid residues in crystal structures. The second THK molecule, which we will call THK2, packed near the chlorophenyl ring of THK1, with its piperidyl ring near $Asp^{129}$ and $Asp^{130}$. A chloride ion was associated with the amino group of the piperidine ring in each of the observed binding modes.

The observation that THK binds in three very distinct ways to two essentially identical protease active sites with the same observed $IC_{50}$ makes it less than straightforward to apply free energy perturbation techniques to this system. By adding substituents which should improve the binding in only one of the modes, we hope to reduce the problem of multiple binding sites and at the same time improve binding affinity. Since we felt it had intrinsically better binding potential, the present calculations are based on the 'closed', or Q7K mutant, structure.

We wished to use free energy perturbation to predict the relative binding energies of haloperidol derivatives. To choose the sites on THK for modification, we examined the Q7K/THK crystal coordinates using MidasPlus [7]. C2 and C5 of the piperidine ring appeared to be good positions for modification. In both positions, the equatorial hydrogens were aligned so that substituents would point toward relatively open hydrophobic binding pockets. The C2 position should be less favorable, since the binding pocket is already partially filled by the fluorophenyl ring. The C5 position seemed more promising since it faced the chlorophenyl ring of the second THK molecule in the crystal structure (Figure 2), which led us to conclude that this position was favorable for a hydrophobic fragment. Of the haloperidol derivatives that have been synthesized and tested so far, very few contained either C2 or C5 substituents. A 2-phenyl derivative of haloperidol ketal showed 20% inhibition at 150 $\mu$M; the parent ketal has an $IC_{50}$ of 75 $\mu$M (Caldera, P., personal communication). Therefore, we decided to perturb haloperidol thioketal into its 2-phenyl derivative as a 'control run', since this was our only point of comparison with experiments. We found several examples of C5 substituents tested at UCSF which had varying effects: a large hydrophobic group lowered the $IC_{50}$ of haloperidol ketal, while polar groups raised it. We chose a phenyl at this position for three reasons: (i) to compare the same substitution at C2 and C5; (ii) because we thought the phenyl group would be large enough to fill the hydrophobic pocket; and (iii) because it is fairly rigid so that conformational sampling was not expected to pose a problem. In a third simulation, we chose a C5-benzyl substituent since we believed at the time that this compound would be easier to synthesize than the C5-phenyl derivative.

## Methods

The AMBER suite of programs was used in this work [8]. The AMBER 4.1 versions of SANDER and GIBBS were used for minimization, equilibration, and free energy perturbation. The force field equations and Weiner et al. [9] parameters were used. This force field uses the following equation to describe the energy of a macromolecular system:

$$\begin{aligned} U = &\sum_{\text{bonds}} k_r(r - r_0)^2 + \sum_{\text{angles}} k_\theta(\theta - \theta_0)^2 \\ &+ \sum_{\text{dihedrals}} V_n/2[1 + \cos(n\phi - \gamma)] \\ &+ \sum_{\substack{\text{nonbonded} \\ i<j}} \{\epsilon_{ij}[(R^*_{ij}/R_{ij})^{12} - 2(R^*_{ij}/R^*_{ij})^6] \\ &+ q_i q_j/\epsilon R_{ij}\} + \sum_{\text{H}-\text{bonds}} (C_{ij}/R^{12}_{ij} - D_{ij}/R^{10}_{ij}) \end{aligned} \quad (1)$$

Some additional force field parameters, summarized in Table 1, were needed to describe the haloperidol deriv-

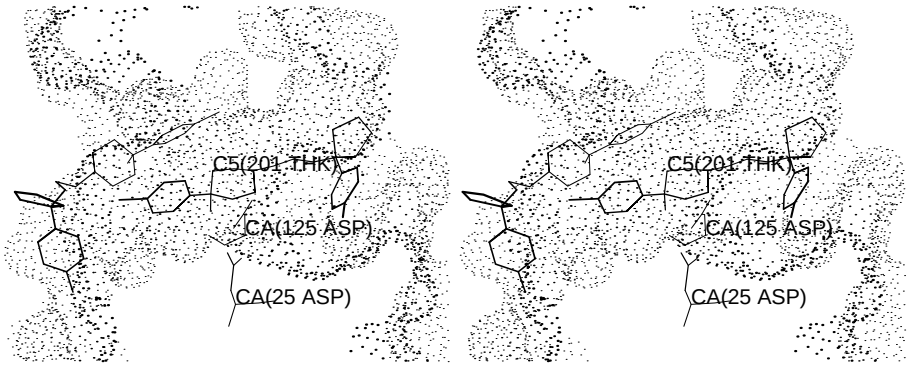


*Figure 2.* Crystal structure positions of the two molecules of THK bound in the active site of Q7K mutant HIV protease. A portion of the solvent-accessible surface of the protein is shown, with the active site Asp dyad labeled.

*Table 1.* Additional force field parameters used in this work

| Atom types | R* (Å) | | ϵ (kcal/mol) | |
|---|---|---|---|---|
| CL – organic chlorine | 2.0 | | 0.2 | |
| IM – chloride ion | 2.25 | | 0.02 | |
| F – organic fluorine | 1.65 | | 0.15 | |
| Type | $k_R$ | | $R_{eq}$ | |
| **Bond parameters** | | | | |
| CA-F | 400.0 | | 1.359 | |
| A-CL | 200.0 | | 1.727 | |
| Type | $k_\theta$ | | $\theta_{eq}$ | |
| **Angle parameters** | | | | |
| CA-CT-N3 | 63.0 | | 114.0 | |
| CT-N3-CT | 50.0 | | 113.0 | |
| S-CT-S | 32.0 | | 110.0 | |
| CA-CT-OH | 50.0 | | 109.5 | |
| CA-CA-F | 70.0 | | 120.0 | |
| CA-CA-CL | 70.0 | | 120.0 | |
| CA-CT-S | 50.0 | | 114.0 | |
| CT-C-OH | 80.0 | | 120.4 | |
| O-C-OH | 80.0 | | 126.0 | |
| Type | idivf | $V_n/2$ | ϕ | n |
| **Torsional parameters** | | | | |
| CA-CA-CT-S | 4 | 0.0 | 180.0 | 3 |
| CA-CT-S-CT | 1 | 0.242 | 0.0 | 3 |
| CT-CT-CT-S | 1 | 0.1 | 180 | 2 |

atives. Angle and torsion parameters for the thioketal ring are derived from similar parameters in the MM3 force field [10]. No lone pairs were included in the sulfurs of the thioketal. The fluorine van der Waals (vdW) parameter was from Gough et al. [11]. Others were determined by analogy. $Asp^{125}$ was assumed to be neutral and $Asp^{25}$ was anionic, as in other free energy studies of hydroxy-containing inhibitors. In this paper, residues 1–99 refer to the first monomer and residues 101–199 refer to the second monomer in the symmetrical dimer. $Asp^{25}$ accepts a hydrogen bond from the hydroxyl group of THK and its derivatives. Partial charges for neutral aspartic acid residue 125 were taken from Radmer et al. [4a]. Charges for haloperidol thioketal and two phenyl derivatives were simultaneously fitted to the 6-31G* electrostatic potentials using RESP methodology [12] so that only the $CH_2$/CH-Ph charges are different, as shown in Figure 3 and Table 2. Later, the charges for the CH-$CH_2$-Ph group of the 5-benzyl derivative were optimized, also using 6-31G*/RESP [12], but in this case the partial charges of the rest of the molecule were fixed at the same values as in the previous calculations.

For the solvent simulations, the inhibitor geometry was taken from the crystal structure and the substituent (phenyl or benzyl) atoms were built in and minimized. Next, the structure was solvated in a box of TIP3P [13] water with the solvent extending at least 10 Å beyond the inhibitor atoms in the x, y, and z directions. Periodic boundary conditions were applied with constant pressure conditions and a coupling constant of 0.2 ps. The temperature was maintained at 298 K with separate solute/solvent coupling factors of 0.2 and 0.1 ps. For the protein simulation, only the residues within approximately 10 Å of the perturbed group were allowed to move. The same moving set was used in all protein calculations: residues 7–10, 22–32, 45–54, 56, 76, and 79–88 of each chain. A spherical cap of 646 waters was centered on the inhibitor with a radius of 23 Å and a half-harmonic restraining force of 0.5 kcal/$Å^2$. Six chloride counterions in addition to the one associated

*Table 2.* AMBER atom types and RESP partial charges for THK and its derivatives. See Figure 3 for atom naming scheme

| Name | AMBER type | THK | 2-Phenyl | 5-Phenyl | 5-Benzyl |
|---|---|---|---|---|---|
| **Partial charges** | | | | | |
| N1 | N3 | 0.0036 | 0.0036 | 0.0036 | 0.0036 |
| H1 | H | 0.325 | 0.325 | 0.3259 | 0.3259 |
| C2 | CT | 0.0043 | −0.0355 | 0.0043 | 0.0043 |
| H21 | HP | 0.0533 | −0.1073 (CA) | 0.0533 | 0.0533 |
| PC23 | CA | – | −0.1036 | – | – |
| PH23 | HA | – | 0.1355 | – | – |
| PC24 | CA | – | −0.1250 | – | – |
| PH24 | HA | – | 0.1479 | – | – |
| PC25 | CA | – | −0.0966 | – | – |
| PH25 | HA | – | 0.1575 | – | – |
| PC26 | CA | – | −0.1250 | – | – |
| PH26 | HA | – | 0.1479 | – | – |
| PC27 | CA | – | −0.1036 | – | – |
| PH27 | HA | – | 0.1355 | – | – |
| H22 | HP | 0.1109 | 0.1417 | 0.1109 | 0.1109 |
| C3 | CT | −0.0197 | −0.0197 | −0.0197 | −0.0197 |
| H31 | HC | 0.0007 | 0.0007 | 0.0007 | 0.0007 |
| H32 | HC | 0.0891 | 0.0891 | 0.0891 | 0.0891 |
| C4 | CT | 0.1497 | 0.1497 | 0.1497 | 0.1497 |
| O4 | OH | −0.6282 | −0.6282 | −0.6282 | −0.6282 |
| H4 | HO | 0.4291 | 0.4291 | 0.4291 | 0.4291 |
| C5 | CT | −0.0197 | −0.0197 | −0.0288 | −0.0197 |
| H52 | HC | 0.0007 | 0.0007 | 0.0025 | 0.0007 |
| H51 | HC | 0.0891 | 0.0891 | 0.0046 (CA) | 0.0891 |
| PC53 | CA | – | – | −0.1006 | – |
| PH53 | HA | – | – | 0.1374 | – |
| PC54 | CA | – | – | −0.1854 | – |
| PH54 | HA | – | – | 0.1586 | – |
| PC55 | CA | – | – | −0.0769 | – |
| PH55 | HA | – | – | 0.1487 | – |
| PC56 | CA | – | – | −0.1854 | – |
| PH56 | HA | – | – | 0.1586 | – |
| PC57 | CA | – | – | −0.1006 | – |
| PH57 | HA | – | – | 0.1374 | – |
| C6 | CT | 0.0043 | 0.0043 | 0.0043 | 0.0043 |
| H61 | HP | 0.1109 | 0.1109 | 0.1109 | 0.1109 |
| H62 | HP | 0.0533 | 0.0533 | 0.0533 | 0.0533 |
| C7 | CT | −0.0549 | −0.0549 | −0.0549 | −0.0549 |
| H71,H72 | HP | 0.0810 | 0.0810 | 0.0810 | 0.0810 |
| C8 | CT | 0.0040 | 0.0040 | 0.0040 | 0.0040 |
| H81,H82 | HC | 0.0092 | 0.0092 | 0.0092 | 0.0092 |
| C9 | CT | −0.0303 | −0.0303 | −0.0303 | −0.0303 |
| H91,H92 | HC | 0.1085 | 0.1085 | 0.1085 | 0.1085 |
| C10 | CT | −0.2497 | −0.2497 | −0.2497 | −0.2497 |
| S11,S14 | S | −0.1714 | −0.1714 | −0.1714 | −0.1714 |
| C12,C13 | CT | −0.0374 | −0.0374 | −0.0374 | −0.0374 |
| H121,H122 | H1 | 0.1041 | 0.1041 | 0.1041 | 0.1041 |
| H131,H132 | H1 | 0.1041 | 0.1041 | 0.1041 | 0.1041 |
| C101 | CA | 0.0479 | 0.0479 | 0.0479 | 0.0479 |

*Table 2. (continued)*

| Name | AMBER type | THK | 2-Phenyl | 5-Phenyl | 5-Benzyl |
|---|---|---|---|---|---|
| C102 | CA | −0.0661 | −0.0661 | −0.0661 | −0.0661 |
| H102 | HA | 0.1273 | 0.1273 | 0.1273 | 0.1273 |
| C103 | CA | −0.2214 | −0.2214 | −0.2214 | −0.2214 |
| H103 | HA | 0.1708 | 0.1708 | 0.1708 | 0.1708 |
| C104 | CA | 0.2493 | 0.2493 | 0.2493 | 0.2493 |
| F104 | F | −0.1832 | −0.1832 | −0.1832 | −0.1832 |
| C105 | CA | −0.2214 | −0.2214 | −0.2214 | −0.2214 |
| H105 | HA | 0.1708 | 0.1708 | 0.1708 | 0.1708 |
| C106 | CA | −0.0661 | −0.0661 | −0.0661 | −0.0661 |
| H106 | HA | 0.1273 | 0.1273 | 0.1273 | 0.1273 |
| C41 | CA | 0.0027 | 0.0027 | 0.0027 | 0.0027 |
| C42 | CA | −0.1923 | −0.1923 | −0.1923 | −0.1923 |
| H42 | HA | 0.1694 | 0.1694 | 0.1694 | 0.1694 |
| C43 | CA | −0.0686 | −0.0686 | −0.0686 | −0.0686 |
| H43 | HA | 0.1438 | 0.1438 | 0.1438 | 0.1438 |
| C44 | CA | 0.0041 | 0.0041 | 0.0041 | 0.0041 |
| CL44 | CL | −0.0692 | −0.0692 | −0.0692 | −0.0692 |
| C45 | CA | −0.0686 | −0.0686 | −0.0686 | −0.0686 |
| H45 | HA | 0.1438 | 0.1438 | 0.1438 | 0.1438 |
| C46 | CA | −0.1923 | −0.1923 | −0.1923 | −0.1923 |
| H46 | HA | 0.1694 | 0.1694 | 0.1694 | 0.1694 |

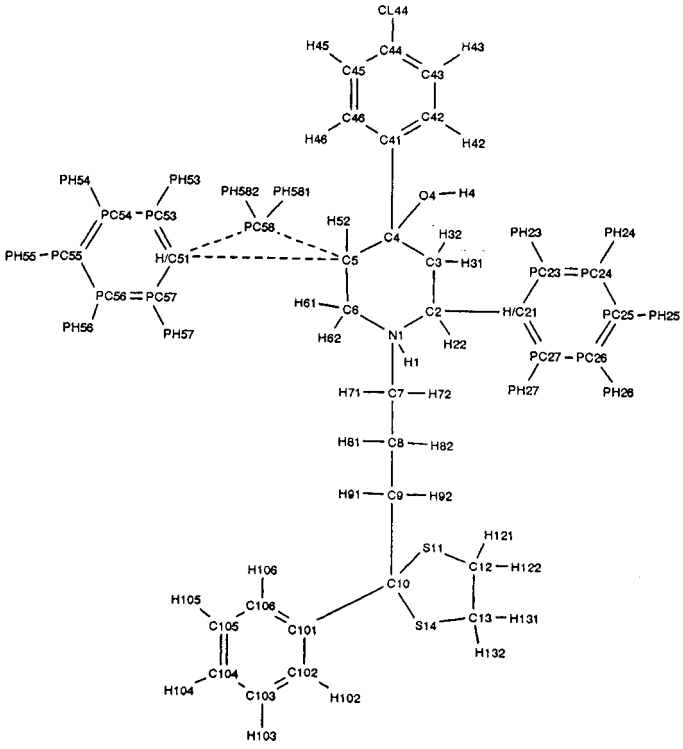


*Figure 3.* Atomic numbering scheme used for the THK derivatives in Table 2.

with THK in the crystal structure were added to reduce the overall charge to +1. This is equal to the overall charge in the solvent simulations, since the inhibitor charge is +1. All waters were allowed to move, as well as any ions associated with moving residues. A dual residue-based cutoff was used for all protein and solution calculations: the primary cutoff was 8 Å, and every 15 time steps when the pair list was updated, the energy contributions from residues between 8 and 12 Å were computed. These energies were taken to be constant between updates. The dual cutoff tended to reduce the drift of the coordinates from the crystal structure considerably, at lower computational cost than a full 12 Å cutoff.

The method of free energy perturbation has been described in detail [14]. The perturbation of A to B must be performed in solution and in the enzyme binding site to obtain the relative free energies $\Delta G_1$ and $\Delta G_2$, respectively. The relative binding free energy $\Delta\Delta G = \Delta G_2 - \Delta G_1$. This is equivalent to the experimentally measurable quantity $\Delta G_{(bind)B} - \Delta G_{(bind)A}$, as shown in the thermodynamic cycle of Scheme 1.

$$\begin{array}{ccc} P + A & \xrightarrow{\Delta G_{(bind)A}} & PA \\ \Delta G_1 \downarrow & & \downarrow \Delta G_2 \\ P + B & \xrightarrow{\Delta G_{(bind)B}} & PB \end{array}$$

Scheme 1.

The perturbation was performed using slow growth methodology, in which the parameters change at every time step according to a coupling parameter λ, which goes smoothly from 0 to 1 (or from 1 to 0, for the reverse simulation). The total change in free energy, ΔG, is the summation of $H(\lambda_{i+1}) - H(\lambda_i)$ at each step, with the assumption that the system remains equilibrated during the course of the simulation.

$$\Delta G = \sum_i \{H(\lambda_{i+1}) - H(\lambda_i)\} \quad (2)$$

State A (at λ = 0) represents the initial state, which in this case is THK in either the protein environment or in solution. State B (at λ = 1) is the final state, which is 2-phenyl THK, 5-phenyl THK, or 5-benzyl THK in this case. The contributions to ΔG due to certain atoms or certain terms of Equation 1 can be logged separately. These components, unlike the total ΔG, are path-dependent. They also appear to converge more slowly than the total ΔG. However, they may prove qualitatively useful in interpreting the results of free energy simulations [15].

*Table 3.* Relative free energies (kcal/mol) for the simulation THK → 2-phenyl THK

| | Forward run | | | Reverse run | | |
|---|---|---|---|---|---|---|
| | Elec.[a] | vdW[b] | Total | Elec. | vdW | Total |
| Protein | 2.9 | 0.2 | **3.1** | 3.0 | −1.9 | **1.2** |
| Solvent | 2.9 | 1.3 | **4.2** | 2.6 | −0.4 | **2.2** |
| ΔΔG | | | **−1.1** | | | **−1.0** |

[a] Refers to the 50 ps electrostatic portion of the simulation only.
[b] Refers to the 200 or 400 ps vdW portion of the simulation.

In the first part of each simulation, only the bonded parameters and the vdW radii were perturbed. At the end of these simulations, a full-size substituent is present, but the newly appearing atoms have zero charges. In the second part of the simulation, only the charges were perturbed. This decoupling prevents instabilities in the trajectory resulting from the presence of charges on groups which have very small vdW radii. vdW perturbations of 200 ps and electrostatic perturbations of 50 ps were performed for all cases. For 5-phenyl THK, a separate vdW perturbation of 400 ps was performed to examine the convergence of the free energies further. For all other calculations, the difference between the results of the forward and reverse runs gives an estimate of error. The free energy differences are summarized in Table 2.

## Results

*Solvation free energies*

The change in solvation free energy for 2-phenyl THK relative to THK is 3.2 ± 1.0 kcal/mol (Table 3). Most (2.8 ± 0.2 kcal/mol) of this unfavorable free energy change appears in the electrostatic portion of the simulation. No intraperturbed group terms were included in the free energy difference, so the interaction of the phenyl group with solvent must be unfavorable. The phenyl group is adjacent to a charged tertiary amine, and it may hinder the solvation shell of the amino group. To test this hypothesis, THK with zero partial charges was perturbed to 2-phenyl THK with partial charges only on the phenyl group. In this case, the solvation free energy change had a favorable electrostatic component, supporting our previous surmise.

The solvation free energy of 5-phenyl THK relative to THK is 1.9 ± 1.0 kcal/mol (Table 4). The

*Table 4.* Relative free energies (kcal/mol) for the simulation THK → 5-phenyl THK

| | Forward run | | | Reverse run | | |
|---|---|---|---|---|---|---|
| | Elec.[a] | vdW[b] | Total | Elec. | vdW | Total |
| Protein | 0.4 | −2.1 | **−1.7** | 0.2 | −3.9 | **−3.7** |
| | | (1.2)[c] | (−0.8) | | (−3.7) | (−3.5) |
| Solvent | 1.5 | 1.4 | **2.9** | 1.3 | −0.5 | 0.8 |
| ΔΔG | | −4.6 | | | −4.5 | |
| | | (−3.4) | | | (−4.7) | |

[a] Refers to the 50 ps electrostatic portion of the simulation only.
[b] Refers to the 200 or 400 ps vdW portion of the simulation.
[c] Numbers in parentheses are for the 400 ps vdW perturbation.

*Table 5.* Relative free energies (kcal/mol) for the simulation THK → 5-benzyl THK

| | Forward run | | |
|---|---|---|---|
| | elec. | vdW | total |
| Protein | −0.2 | −1.9 | −2.1 |
| Solvent | −0.2 | −1.1 | −1.3 |
| ΔΔG (kcal) | | | −0.8 |

vdW portion of the run gave the same result ($0.5 \pm 1.0$ kcal/mol) as the previous 2-phenyl run. This is reasonable, since both cases involve growing in a phenyl group. However, the electrostatic portion of the free energy simulation yielded a smaller, but still unfavorable, result of $1.4 \pm 0.1$ kcal/mol. Again, the phenyl group is being added near a sterically hindered polar group and it may be disrupting the solvation shell.

The predicted solvation free energies may make these two THK derivatives too insoluble to be used as drugs. Their solubilities may be improved by adding polar groups, as long as they are carefully chosen to improve the binding free energy as well as the solvation free energy.

The predicted change in solvation free energy for 5-benzyl THK relative to THK is −1.3 kcal/mol (Table 5). Unlike the previous two simulations, both the electrostatic and vdW components are favorable, although the electrostatic contribution is only −0.2 kcal/mol. This simulation was run in the forward direction only.

The electrostatic free energies show relatively small hystereses (the differences between the forward and reverse simulations) of 0.2–0.3 kcal/mol. However, the hystereses of the vdW perturbations are consistently 2 kcal/mol. These runs are obviously not yet converged, but, as is discussed later, $\Delta\Delta G_{bind}$ seems to be closer to convergence than either $\Delta G_1$ or $\Delta G_2$.

*Protein binding free energies*

Figure 4 illustrates the structure of THK bound in the THK1 site, with positions 2 and 5 marked.

*2-Phenyl THK.* The predicted binding free energy change ($\Delta\Delta G_{bind}$) for mutating THK to 2-phenyl THK is $-1.1 \pm 2.0$ kcal/mol after correcting for the change in solvation free energy (Table 3). The mutation is disfavored by 2.2 kcal/mol in protein and 3.2 kcal/mol in solvent. The electrostatic portion of the protein run gives a free energy difference of $3.0 \pm 0.1$ kcal/mol, nearly the same as the solvent value. The vdW contribution is responsible for the more favorable protein free energy. The hysteresis is only 0.1 kcal/mol for the electrostatic run, but is 2.1 kcal/mol for the vdW simulation.

The root mean square (rms) deviation from the crystal structure is about 1.1 Å after the forward simulation (see Figure 5 for a stereoview of this structure) and 1.3 Å at the end of the reverse simulation. The rms deviation for backbone C, N, and O atoms is 0.9 Å. Examination of the structure shows no significant conformational changes except at the tip of the flap region. During the simulation, a water molecule inserted itself between the amino group of THK and the chloride ion that hydrogen bonds to $Ile^{50}$ and $Ile^{150}$, causing a shift in their positions. The inhibitor molecule moves quite a bit in the binding site during equilibration and simulation phases, probably because the binding site is much larger than the inhibitor and there are relatively few interactions holding it in place. In spite of the large deviations from the crystal structure orientation, the hydrogen bond between $Asp^{25}$ and the hydroxy group of the inhibitor remained through the entire run.

Although there is no experimental ΔΔG to directly compare with, the 2-phenyl derivative of haloperidol ketal is a poorer inhibitor than haloperidol ketal. While the $IC_{50}$ of the latter is 75 μM, at 150 μM the 2-phenyl derivative is less than 20% inhibitory. Our calculations suggest that the 2-phenyl derivative should bind somewhat better than THK; however, our error bar is quite large, within the range of the inferred experimental value.

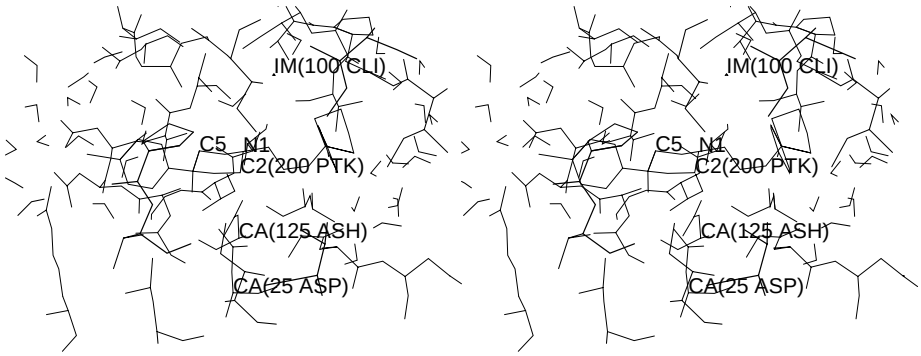


*Figure 4.* Stereoview of a snapshot from the equilibration trajectory of THK, showing the C2 and C5 positions of the piperidine ring.

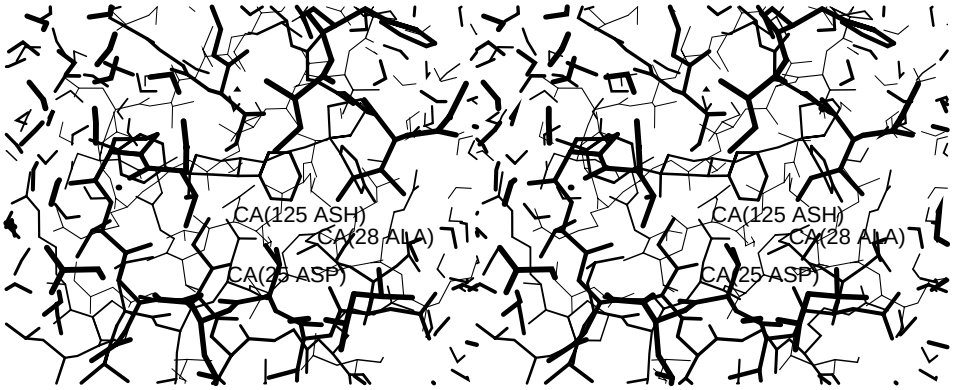


*Figure 5.* Stereoview of a snapshot from the equilibration trajectory at the 2-phenyl THK.

*5-Phenyl THK.* The 5-phenyl THK perturbation is predicted to have a $\Delta\Delta G_{bind}$ of $-4.5 \pm 2.0$ relative to THK according to our calculations (Table 4). Again, the hysteresis is 4 kcal/mol, and follows the same trend in protein as in solution. To try to achieve better convergence of the vdW portion of the run, we increased the length of the simulation from 200 to 400 ps in each direction. The first half of the run, from $\lambda = 0.0$ to 0.5, was 100 ps long, as in the previous run. The second half was slowed down so that the perturbation from $\lambda = 0.5$ to 1.0 occurred over 300 ps. This was done because the slope of $\Delta G$ versus $\lambda$ is very flat in the first half of the simulation, but increases steeply near the endpoint of the run. The forward and reverse runs differ by 2.5 kcal/mol for the 400 ps runs, which is larger than the 1.8 kcal/mol hysteresis in the 200 ps runs. Although the hysteresis is still large, these calculations are consistent with the 200 ps vdW runs in predicting that 5-phenyl THK will bind far more tightly than THK. The average value for $\Delta\Delta G_{(bind)}$ is $-4.0 \pm 2$. Figure 6 has a stereoview of the bound 5-phenyl THK structure.

The electrostatic portion of the run is about 1 kcal/mol more favorable in protein than in solvent. $Asp^{30}$ may be important in this difference, since it has a large favorable electrostatic component. Its carboxyl oxygens are oriented toward the phenyl hydrogens, which carry small positive charges. This is also true of the interaction of $Asp^{130}$ with the 2-phenyl group in the previous simulation, but the magnitude of the component is 0.5 kcal/mol smaller. Placing hydrogen bond donors on the C5-phenyl ring may further improve binding by interacting with $Asp^{30}$. The main contribution to favorable binding of this derivative is the vdW component, which is about 3.5 kcal/mol better in the protein than in solution. This may reflect the sum of a number of small favorable hydrophobic interactions with the binding pocket, a better fit than the C2-phenyl group.

As discussed earlier, the 5-phenyl derivative is predicted to be less water-soluble than the parent compound. If this compound indeed proves to be a tight-binding inhibitor, it must be modified further to increase its hydrophilicity. The structure of the 5-phenyl THK / HIV protease complex at the endpoint of

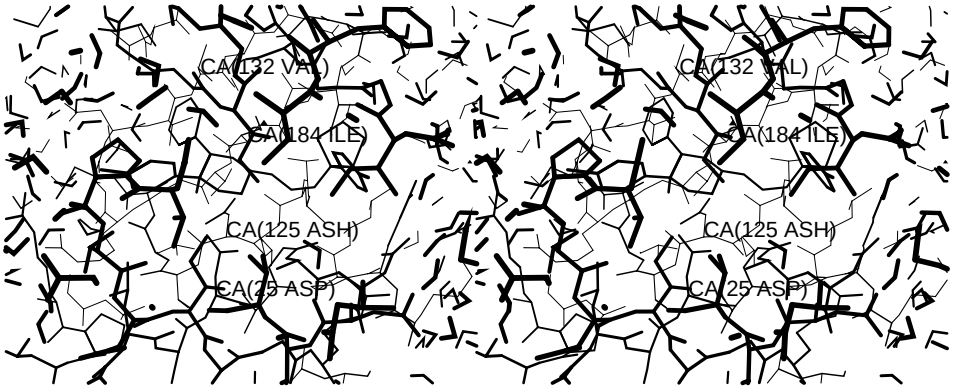


*Figure 6.* Stereoview of a snapshot from the equilibration trajectory at the 5-phenyl THK.

the perturbation was examined to search for favorable locations for polar group addition. The meta positions of both the fluorophenyl ring and the 5-phenyl ring are somewhat close to the amide hydrogens of $Asp^{30}$ and $Asp^{129}$, respectively. Hydroxy or hydroxymethylene groups on these two rings could accept hydrogen bonds. Hydroxymethylene groups appear to be better suited to span the distance to these two amide hydrogens judging from the static structure at least. In addition, the thioketal ring is exposed to solvent, so the addition of polar or even charged groups here should have little effect on $\Delta\Delta G_{(bind)}$, since the groups would be equally well solvated whether bound in HIV protease or free in solution.

*5-Benzyl THK.* After the 2-phenyl and 5-phenyl simulations were analyzed, we felt that a 5-benzyl group might have a more favorable free energy of binding than the two phenyl derivatives. The phenyl derivatives suffered from steric interactions with $Ala^{28}$ or $Ala^{128}$. It was thought that the benzyl group would have less steric interactions with $Ala^{128}$ since it has an additional methylene group reducing steric bulk near the piperidine and allowing more conformational flexibility. Also, this compound was thought at the time to be easier to synthesize. The predicted relative binding free energy for the mutation of THK to 5-benzyl THK is $-2.1$ kcal/mol, for the forward simulation in the protein only (Table 5). See Figure 7 for a stereoview of the 5-benzyl structure. Compared with the two forward simulations of 5-phenyl THK, the benzyl derivative is predicted to interact better with the protein by 0.4–1.3 kcal/mol. However, the predicted solution behavior is dramatically different from that of the two phenyl compounds. The solvent simulations for 5-benzyl THK predict a solvation free energy of $-1.3$ kcal/mol compared to THK, while the 5-phenyl solvation free energy is $+2.9$ in the forward run. Because 5-benzyl THK is predicted to have a lower free energy than THK in both solvent and protein, the overall effect on inhibition is predicted to be rather small. The calculated $\Delta\Delta G_{bind}$ of 5-benzyl THK is lower than that of 2-phenyl and 5-phenyl THK, but it is still expected to have a $\Delta\Delta G$ of $-0.8$ kcal/mol relative to the parent compound. Because the SHAKE algorithm failed frequently in the reverse run, we were not able to complete it.

Although the reverse run was not done, the margin of error for the 5-benzyl simulation is probably at least as large as, if not larger than, the 4 kcal/mol hysteresis observed for the previous phenyl runs. In addition to the sampling difficulties for the protein and the THK framework, the benzyl group has an additional rotatable bond, with three low-energy rotamers in theory. In the protein, the binding pocket is constrictive enough so that only one rotamer appears to fit reasonably well. This corresponds to the lowest energy rotamer in the gas phase. In solution, all three rotamers should be sampled according to their relative energies in solution. Since the predicted $\Delta\Delta G_{(bind)}$ was not very impressive compared with 5-phenyl THK, we did not attempt to rigorously sample the solution conformations.

## Discussion

Free energy calculations have great potential in structure-based drug design; however, the limitations of the method are significant at this point. The simple additive potential energy function used in molecular dynamics (MD) and free energy calculations has been

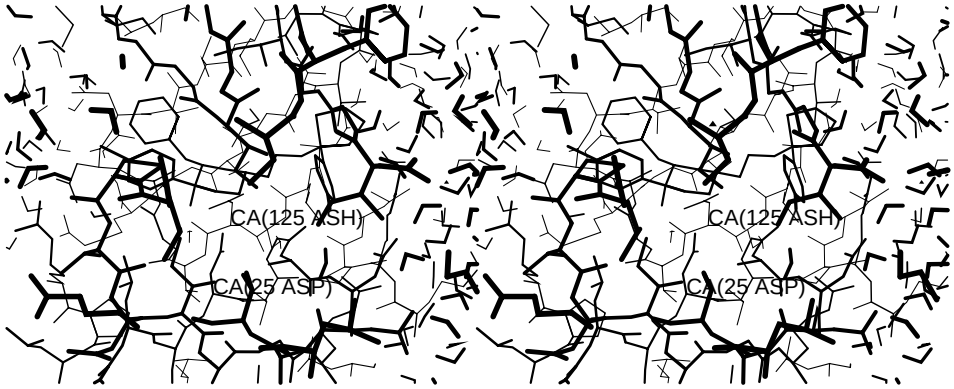


*Figure 7.* Stereoview of a snapshot from the equilibration trajectory at the 5-benzyl THK.

*Table 6.* Residues with differential interactions greater than 0.5 kcal/mol with thioketal analog compared to parent

| | 5-Phenyl THK | 2-Phenyl THK | 5-Benzyl THK |
|---|---|---|---|
| Favorable ($\Delta\Delta G < -0.5$) | $Asp^{130}$ | $Asp^{29}$ | $Asp^{125}$ |
| | $Gly^{149}$ | $Asp^{30}$ | $Gly^{127}$ |
| | $Ile^{84}$ | $Ile^{84}$ | $Gly^{148}$ |
| | $Val^{132}$ | $Cl^-$ | $Cl^-$ |
| Unfavorable ($\Delta\Delta G > 0.5$) | $Asp^{25}$ | $Asp^{25}$ | $Asp^{25}$ |
| | $Ala^{128}$ | $Ala^{28}$ | $Ala^{128}$ |
| | $Cl^-$ | | |

shown to give reasonable results for intermolecular and conformational energies, and for free energies of binding and solvation [14]. One possible source of significant error is the truncation of nonbonded force and energy evaluation at some cutoff distance, usually 8 Å. The inclusion of long-range electrostatic effects has been demonstrated to greatly decrease the drift of the average coordinates away from the crystal structure [16]. Our use of a dual 8 Å / 12 Å cutoff reduces somewhat the errors associated with truncating the long-range electrostatic forces, and the rms motion of the atoms from the crystal structure at the end of the reverse simulations is ∼1–1.5 Å, suggesting that the structure is well maintained. The problem of sampling enough configurations to obtain reliable, converged free energies is far more difficult. To 'prove' convergence, one must compare the results of many long simulations, although the difference between 'forward' and 'reverse' simulations or between several different trajectories can provide a lower limit estimate of the error in a given free energy difference. Because these simulations are too time-consuming at present and an exact answer is not as important as a qualitative trend, we think our results are of interest even though we did not achieve fully converged free energies. Even at 400 ps simulation times, the vdW energies do not appear to be close to convergence, based on hysteresis. The most logical reason for this is that when one makes such large structural changes as adding a phenyl group, the relaxation time of the surroundings is significantly larger than the ∼400 ps of our longer simulation. Thus, one has significant hysteresis.

In all of the simulations which were run in the forward and reverse directions, the hysteresis is large and follows the same trends. The hysteresis for the vdW portion of the simulations is much greater than that of the electrostatic portion, which is understandable for changing a hydrogen to a phenyl or benzyl group. The energies for the reverse vdW runs (changing phenyl or benzyl back to hydrogen) are consistently 2 kcal/mol lower than the corresponding forward simulation, for solvent and protein runs. Because the solvent and protein simulations exhibit the same trend in spite of

very different perturbed group environments, we think that the observed hysteresis is largely due to the different nature of growing versus shrinking a group in a condensed phase. If this is the case, the forward and reverse numbers should converge as the simulation time increases and the surroundings have more time to adjust to the new atomic positions and radii. We cannot estimate the time required for convergence from our calculations, but it is apparently much longer than 400 ps. We felt it was impractical to run longer simulations for these large systems, because the main purpose of our calculations was to predict the trends.

Because free energy calculations are so expensive in terms of computer resources, we did not consider other distinct binding modes. Nonetheless, the one we have chosen is reasonable because it is not only close to the DOCK prediction, but is also consistent with other protease inhibitor binding modes determined by X-ray crystallography and the most buried inhibitor position in the observed crystal structure [6]. We made the assumption (which is by no means certain) that the most important binding mode of haloperidol thioketal and its derivatives was the one that resembled the DOCK prediction. For inhibitors with multiple binding modes, the choice of which bound orientation to begin with is crucial. Our reasoning for discarding the crystallographic information from the first (wild-type) crystal structure was simply that it was not consistent with the observed structure–activity relationships of haldol derivatives. Of the two molecules of THK bound to the protease in the second crystal structure [6], we chose the bound orientation that appeared to provide the best scaffold for adding groups to fill binding pockets. This orientation also had the advantages of being less solvent-exposed and directly hydrogen-bonded to an active site aspartate residue. Another assumption which we made of necessity was that the derivatives would bind in the same orientation as the parent compound or, more precisely, that the DOCK-like binding mode was the most favorable one for these compounds. Because of all these uncertainties, the comparison of our results with eventual experiments will be a stringent test not only of our methods, but of our choice of starting structure as well. Our intuition based on observing the crystal structure, that the 5-phenyl group would have more binding energy in the protein than the 2-phenyl group, was supported by the free energy calculations. Until the synthesis, testing, and possible cocrystallization of these derivatives with HIV protease are accomplished, we will not know if the assumptions we made were correct.

We studied the 2-phenyl thioketal inhibitor and calculated a relative binding free energy of $-1.1 \pm 2.0$ kcal/mol relative to thioketal, whereas the experimental value suggests the relative free energy is $\sim$0.5 kcal/mol. Although the calculated free energy is of the wrong sign, considering the large error bars, the two are not inconsistent with each other. On the other hand, the calculated relative free energy for 5-phenyl thioketal is $-4 \pm 2$ kcal/mol; this molecule is clearly predicted to be significantly more favorably bound than thioketal, and its 2-phenyl or 5-benzyl derivatives.

In order to be useful for drug design, free energy calculations must be predictive, rather than just striving to reproduce experimental results. In this case, we are perturbing from an inhibitor for which some binding information is available to an unknown derivative. We predict that of the three derivatives examined, 5-phenyl THK is by far the most promising candidate for synthesis, testing, and further modification. Syntheses of the compounds, with emphasis on the 5-phenyl derivative, are in progress.

In our free energy calculations, we broke down the total interaction free energy into components to assess which protein/solvent residues are responsible for the differential interaction of thioketal and its analog with the protein. We have usefully interpreted these components before [14], but there is some controversy in the literature on them. The reason for the controversy (as discussed elsewhere [17]) is that the components, unlike the total free energy, are not path-independent.

Notwithstanding the path dependence of the components, we think they can be useful in a qualitative way in molecular design. Thus, in Table 6, we note these residues whose differential free energies with thioketal and its analog in the protein are $>$ 0.5 kcal/mol in magnitude. As one can see, the differential interaction with $Asp^{15}$ and either $Ala^{28}$ or $Ala^{128}$ is unfavorable for all three analogs. The $Cl^-$ that replaces the flap water [6] interacts unfavorably with the 5-phenyl, but favorably with the 2-phenyl and 5-benzyl analogs. A key feature which enables 5-phenyl to interact favorably and 5-benzyl not to is the better interaction of the former with $Ile^{184}$ and $Val^{132}$.

## Conclusions

We have used free energy calculations to suggest modifications in the structure of THK for better inhibition of HIV-1 protease. Based upon inspection of a crystal

structure of bound THK, we chose C5 of the piperidine ring as a promising site for the addition of large hydrophobic substituents. This insight will not often be useful, since rarely are there two copies of a ligand bound to different locations in a protein structure. However, similar insights can be garnered by soaking the crystal structure with nonpolar solvents, as has been demonstrated by Allen et al. [18]. Experiments and examination of the crystal structure suggested that adding a phenyl group to C2 would lead to decreased binding relative to THK. Our simulations predict a relatively small effect on binding, favoring 2-phenyl THK binding over the parent compound. In contrast, we predict a large favorable effect on the binding free energy for adding a 5-phenyl group. This agrees with our expectation that a large hydrophobic group can be accommodated in this position, overlapping with a second binding mode of THK seen in the crystal structure. The 5-benzyl compound, which was expected to be simpler to synthesize, was predicted to bind somewhat better than THK itself, but not as well as 5-phenyl THK. The 5-benzyl derivative is predicted to be more soluble, while the phenyl derivatives are predicted to be less soluble than THK. Component analysis was used to understand the results in more detail. 5-Phenyl THK appears to make more and better vdW contacts with residues lining the hydrophobic binding pocket than the other two derivatives or the parent compound. All three derivatives appeared to have unfavorable steric interactions with $Ala^{28}$ or $Ala^{128}$. Although there are large error bars associated with these results, it is hoped that they will be successful in predicting a new, tighter-binding HIV-1 protease inhibitor.

Our calculations, although more qualitative than quantitative, are one of very few free energy predictions involving large structural changes. These predictions are currently being tested experimentally. Although there are now a number of HIV protease inhibitors in the market, an advantage of the 5-phenyl thioketal, should our prediction be borne out, is that its small size and hydrophobicity may allow it to cross the blood-brain barrier to attack AIDS viruses in the brain [19].

## Acknowledgements

M.A.M. is grateful to the Merck Academic Development Program for postdoctoral funding. We would like to thank the San Diego Supercomputing Center and the Pittsburgh Supercomputing Center for generous allocations of computer time, and the Computer Graphics Laboratory for visualization (RR01081). This research was supported by grants from NIH (GM-29072 to P.A.K. and to Paul Ortiz de Montellano, P.I.).

## References

1. McQuade, T.J., Thomasselli, A.G, Liu, L., Karacostas, V., Moss, B., Sawyer, T.K., Heinrikson, R.L. and Tarpley, W.G., Science, 247 (1990) 454.
2. a. Navia, M.A., Fitzgerald, P.M.D., McKeever, B.M., Leu, C.-T., Heimbach, J.C., Herber, W.K., Sigal, I.S., Darke, P.L. and Springer, J.P., Nature, 337 (1989) 615.
   b. Wlodawer, A., Miller, M., Jaskolski, M., Sathyanarayana, B.K., Baldwin, E., Weber, I.T., Selk, L.M., Clawson, L., Schneider, J. and Kent, S.B.H., Science, 245 (1989) 616.
   c. Swain, A.L., Miller, M.M., Green, J., Rich, D.H., Schneider, J., Kent, S.B.H. and Wlodawer, A., Proc. Natl. Acad. Sci. USA, 87 (1990) 8805.
   d. Miller, M., Schneider, J., Sathyanarayana, B.K., Toth, M.V., Marshall, G.R., Clawson, L., Selk, L., Kent, S.B.H. and Wlodawer, A., Science, 246 (1989) 1149.
   e. Jaskolski, M., Tomasselli, A.G., Sawyer, T.K., Staples, D.G., Heinrikson, R.L., Schneider, J., Kent, S.B.H. and Wlodawer, A., Biochemistry, 30 (1991) 1600.
3. For a review, see McCarrick, M.A. and Kollman, P.A., Methods Enzymol., 241 (1994) 370.
4. a. Ferguson, D.M., Radmer, R.J. and Kollman, P.A., J. Med. Chem., 34 (1991) 2654.
   b. Tropsha, A. and Hermans, J., Protein Eng., 5 (1992) 29.
   c. Rich, D.H., Sun, C.Q., Vara Prasad, J.V.N., Pathiasseril, A., Toth, M.V., Marshall, G.R., Clare, M., Mueller, R.A. and Houseman, K., J. Med. Chem., 34 (1991) 1222.
   d. Reddy, M.R., Viswanadhan, V.N. and Weinstein, J.N., Proc. Natl. Acad. Sci. USA, 88 (1991) 10287.
   e. Rao, B.G., Tilton, R.F. and Singh, U.C., J. Am. Chem. Soc., 114 (1992) 4447.
   f. Chen, X and Tropsha, A., J. Med. Chem., 38 (1995) 42.
5. 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. and Craik, C.S., Proc. Natl. Acad. Sci. USA, 87 (1990) 6644.
6. Rutenber, E., Fauman, E.B., Keenan, R.J., Fong, S., Furth, P.S., Ortiz de Montellano, P.R., Meng, E., Kuntz, I.D., DeCamp, D.L., Salto, R., Rose, J.R., Craik, C.S. and Stroud, R.M., J. Biol. Chem., 268 (1993) 15343.
7. Ferrin, T.E., Huang, C.C., Jarvis, L.E. and Langridge, R., J. Mol. Graphics, 6 (1988) 13.
8. Pearlman, D.A., Caldwell, J.W., Case, D.A., Ross, W.S., Cheatham III, T.E., DeBolt, S., Ferguson, D., Seibel, G. and Kollman, P.A., Comput. Phys. Lett., 91 (1995) 1.
9. a. Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S. and Weiner, P., J. Am. Chem. Soc., 106 (1984) 765.
   b. Weiner, S.J., Kollman, P.A., Nguyen, D.T. and Case, D.A., J. Comput. Chem., 7 (1986) 230.
10. Allinger, N.L., Yuh, Y.H. and Li, J.H., J. Am. Chem. Soc., 111 (1989) 8551.

11. Gough, C., DeBolt, S. and Kollman, P.A., J. Comput. Chem., 13 (1992) 963.
12. Bayly, C., Cieplak, P., Cornell, W.D. and Kollman, P.A., J. Phys. Chem., 97 (1993) 10269.
13. Jorgensen, W., Chandrasekhar, J., Madura, J., Impey, R. and Klein, M., J. Chem. Phys., 79 (1983) 926.
14. Several such reviews are:
    a. Straatsma, T.P. and McCammon, J.A., Annu Rev. Phys. Chem., 43, (1992) 407.
    b. Richards, W.G., Proc. R. Soc. Edinburgh Sec. B, Biol. Sci., 99 (1992) 105.
    c. Kollman, P.A., Chem. Rev., 93 (1993) 2395.
    d. Beveridge, D.L. and DiCapua, F.M., Annu. Rev. Biophys. Biophys. Chem., 18 (1989) 431.
15. Singh, S., Pearlman, D.A. and Kollman, P.A., J. Biomol. Struct. Dyn., 11 (1993) 303.
16. Cheatham III, T.E., Miller, J.L., Fox, T., Darden, T.A. and Kollman, P.A., J. Am. Chem. Soc., 117 (1995) 4193.
17. Sun, Y., Veenstra, D.L. and Kollman, P.A., Protein Eng., 9 (1995) 273.
18. Allen, K.N., Bellamacina, C.R., Ding, X.C., Jeffrey, C.J., Mattos, C., Petsko, G.A. and Ringe, D., J. Phys. Chem., 100 (1996) 2605.
19. Driscoll, M., Seddiqui, H., Ford, J., Kelley, J., Roth, H., Mitsuya, M., Tanaka and Marquez, V., J. Med. Chem., 39 (1996) 1619.