ABSTRACT

Title of Document: PARAMETERIZATION OF AN ENERGY MODEL FOR SCORING OF ANTI-HIV DRUGS AND A COMPUTATIONAL METHOD OF LEAD COMPOUND OPTIMIZATION FOR DRUG DISCOVERY

Himan Mookherjee, Master of Science, 2005

Directed By: Professor. Richard C. Stewart. Dept. of Cell Biology & Molecular Genetics.

This project aims to parameterize an energy model with the goal of developing a fast method for predicting binding affinities of HIVP inhibitors. This method will be used for in silico compound screening to discover new potential anti-HIV drug candidates. The project also aims to develope a method of optimizing the charges of local parts of a ligand while keeping the rest of the charges roughly constant, rather than attempting to modify all of the ligand's charges towards an optimum, as done in previous approaches. The method developed here will also be computationally faster than existing approaches.

PARAMETERIZATION OF AN ENERGY MODEL FOR SCORING OF ANTI-HIV DRUGS AND A COMPUTATIONAL METHOD OF LEAD COMPOUND OPTIMIZATION FOR DRUG DISCOVERY

By

Himan Mookherjee

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Master of Science
2005

Advisory Committee:
Professor Richard C. Stewart, Chair
Professor Michael K. Gilson
Professor Sergei Sukharev

© Copyright by
Himan Mookherjee
2005

Acknowledgements

I hereby thank by advisor and committee members for their continuous guidance and help to complete my thesis. I also will also like to take this opportunity to thank my lab members, all of whom sincerely helped me learn, advance and successfully complete my thesis project.

Table of Contents

List of Tables

List of Figures

Chapter 1: Significance of the Problem

Human Immunodefficiency Virus (HIV) protease (HIVP) is an enzyme critical for the maturation of the virus. The protease plays a critical role in viral assembly. The protease cleaves the viral polyproteins gag or gag-pol at nine non-homologous sites (Table 1) to yield separate structural proteins and enzymes, including the protease itself, essential for the viral life cycle (Figure 1) (11). Hence, the most potent medications currently available for treatment of Acquired Immunodeficiency Syndrome (AIDS), caused by HIV, are the inhibitors of the viral protease (3). These agents in combination with other agents comprise highly active antiretroviral therapy (HAART).

The active site of this aspartyl protease is located at the interface of the dimers and each monomer contributes an aspartate residue, which is part of an Asp-Thr-Gly triad. The interactions of HIVP with its natural substrates represent an interesting example of recognition of asymmetric substrates by a symmetric enzyme. The natural peptide substrates bind to HIV-1 protease in an extended conformation with eight contiguous residues on the peptide, labeled P4 to P4', making contact with the eight enzyme subsites S4 to S4' (Figure 2) (12). Insight about the interactions of the substrates with HIVP has been gained from crystallographic studies of substrates bound to a form of the enzyme that has been inactivated by mutating the catalytic aspartic acids to asparagines (11). From their analysis of six crystal structures with an inactive (D25N) HIV-1 protease mutant, King *et al.* have hypothesized that the substrate specificity of

HIV-1 protease is based largely on shape, which is conserved across substrates, rather than on the detailed amino acid sequences of the substrates (13). This view is corroborated by the fact that the nine substrate sequences cleaved by the protease differ significantly in their amino acid sequence, but fill very similar volumes within the active site, when bound. The shape defined by the van der Waals volume of the substrates in the active site of the protease has been termed the "substrate envelope" (Figure 3a & 3b).

The currently available, FDA approved, HIV-1 protease inhibitors act by binding to the active site of the protease. However, they can become ineffective due to mutations in the protease that diminish the affinity of the inhibitors but do not prevent the cleavage of the natural substrates of the enzyme. The drug-resistant mutations appear due to the high replicative rate of HIV, the infidelity of the reverse transcriptase and the selective pressure of protease inhibitor therapy on the evolution of the virus. Hence, there is a continuing need for novel multi-targeted inhibitors of HIVP. The present project forms part of a larger, collaborative effort to develop protease inhibitors that effectively inhibit not only wild type but also mutant forms of HIV protease.

Structure-based virtual screening of chemical libraries or optimization of lead compounds obtained from such screenings, have become valuable tools in drug development and the development of the existing HIVP inhibitors is considered a major success of structure-based drug design (6). Structure-

based design essentially requires an understanding of the physical interactions between small molecule inhibitors and their target. Better elucidation of the factors influencing binding of HIVP and its inhibitors will facilitate the discovery of novel drug molecules, especially when combined with the extensive set of crystal structures of HIVP, both wild type and mutants, complexed with inhibitors and substrates that have become available (7). The binding affinities of many inhibitors to both wild type and mutant HIVPs have also been published. These data include the affinities of the cyclic urea inhibitors of HIVP (12-14). The binding affinities of this class of compounds span a wide range, providing the scope to optimize and test models of binding with a varied set of ligands. In addition, the experimental data for the cyclic ureas are likely to be consistent because they are from a single laboratory. The data for this set of ligands have been used extensively in the present study.

Structure-based design of HIVP inhibitors targeted against both wild type protease but also mutants will involve adequate ranking of the binding affinities of potential ligands. Binding affinity predictions of HIVP inhibitors using an empirical force field like CHARMM (8) combined with a continuum solvation model (9) have yielded promising results in a previous study (10). Although the method was not highly accurate, it did appear to have significant predictive value. However, the prior method is limited by its use of computationally expensive detailed calculations of configurational integrals for energy minima. In this project an energy model is being parameterized with the goal of developing a fast method for predicting binding affinities of HIVP inhibitors. This method will be used for *in silico* compound

screening to discover new potential anti-HIV drug candidates. After such a lead drug molecule is obtained from virtual screening, its chemical structure usually must be optimized by chemical modifications to enhance its binding affinity to the receptor.

In recent years, lead optimization procedures have been published which work by minimizing the electrostatic part of the binding free energy by adjusting the ligand's atomic charges to maximize the complementarity of its charge distribution with that of the binding site. However, the existing computational procedures optimize the charge distribution of the whole ligand, so the charge changes suggested for a given part of the ligand may be based upon the assumption that charges elsewhere in the ligand also are changed. This assumption will frequently be a poor one, since chemical modifications of a lead compound are usually local in nature. Thus, local, rather than global, charge optimization should provide more accurate guidance for ligand modification. A central element of the present project is a method of optimizing the charges of local parts of a ligand while keeping the rest of the charges roughly constant, rather than attempting to modify all of the ligand's charges towards a computed optimum, as done in previous approaches.

The method developed here will also be computationally faster than existing approaches because it will require only one Poisson-Boltzmann calculation for the free ligand and one more for the bound ligand, while existing methods require two Poisson-Boltzmann calculations for each charge center of the ligand. The design algorithm can find direct application in the HIV protease inhibitor design effort.

Chapter 2: Background

2.1 **Parameterization of an Energy Model**

An energy model, described below, is being tuned by fitting to experimental binding affinities of a training set of ligands by downhill simplex minimization (11), with the aim of using the parameterized model for scoring ligands according to their binding affinities to their target receptors. The parameterized energy model is being validated on various test systems.

Energy model

The binding energy of a ligand is the free energy change on complex formation of the ligand with its target protein:

$$L + P \xrightarrow{K} PL. \qquad (1)$$

The binding energy of a ligand can be calculated as the difference between free energy of the complex, the free ligand and the free protein:

$$\Delta G = G_{PL} - G_{L} - G_{P} = -RT\ln K, \qquad (2)$$

where K is binding affinity of the ligand, L, and GX is the free energy of the subscripted species.

Influences that stabilize the complex are dispersion forces, charge-charge complementarity, hydrogen bonding, and hydrophobicity. In the present energy model, the first three driving forces are explicitly accounted for by Lennard-Jones and electrostatic terms that form part of the empirical force field CHARMM22. Thus, the van der Waals interaction between atoms i and j is approximated as 12-6 Lennard-Jones potential

ULJ = 4□ij { (□/rij)12 - (□/rij)6 }, (3)

where, rij is the distance between atoms i and j, □is the mean radius of atoms i and j, and □ij is a constant for the atom pair. The electrostatic Coulombic energy is computed according to Coulomb's law as

UCoulombic = qiqj / Drij , (4)

where D is the dielectric constant of the medium and qi is the charge on atom i. The force field also includes a term that accounts for intrinsic energy changes associated with bond rotations. This dihedral energy term is proportional to the cosine of the dihedral angle under consideration. The hydrophobic effect is accounted for approximately via a term proportional to the molecular surface area:

WNP = C A (5)

where A is the surface area, which falls when two molecules bind to form a complex, and C is a coefficient fitted to reproduce measured solvation energies of nonpolar compounds in water. It is also possible to allow the surface area component to account for the van der Waals dispersion forces between ligand and protein, and variant models were tested in which the Lennard Jones term was omitted, as described below.

Influences that destabilize the complex are desolvation of polar groups and loss of translational, rotational, configurational entropies. The Poisson Boltzmann solvation energy accounts for desolvation penalty paid for complexation. Here, the electrostatic potential is computed by solving the three-dimensional linearized Poisson-Boltzmann equation.

$$-\varepsilon_o \nabla \bullet [D(r)\nabla\phi(r)] = \rho(r) + \varepsilon_o D(r)\kappa^2(r)\phi(r) \quad (6)$$

where $\nabla \bullet [D(r)\nabla\phi(r)]$ is the divergence of the displacement field, $\rho(r)$ is the charge density as a function of the position r, $\varepsilon_o D(r)\kappa^2(r)\phi(r)$ gives a measure of the screening effect of the salt in solution, D(r) is the position dependent dielectric constant having a particular value in the molecular region (Dint) and a second value in the solution region and the ion-exclusion region (Dsolv). Here $\kappa^2(r)$ is given by

$$\kappa^2(r) = \frac{2I(r)}{D\varepsilon_o kT} \quad (7)$$

where I(r) is the ionic strength as a function of position, k is the Boltzmann constant and T is the temperature in Kelvin. Note that the ionic strength is zero inside the solute and so $\kappa^2(r)$ is also zero there. The Poisson-Boltzmann can be solved using numerical techniques, such as the finite difference method. This method discretizes the continuum electrostatics problem onto a 3D grid, where the dielectric constants are coded onto the grid lines and the charges and ionic strength are coded onto the vertices of the finite difference grid, thereby transforming the Poisson-Boltzmann equation into a set of difference equations, which can then be solved by standard techniques of linear algebra.

Then the electrostatic energy can be calculated as:

$$W = \frac{1}{2}\sum_{i=1}^{N_{charges}} q_i\phi_i \quad . \quad (8)$$

The number of rotatable bonds (Ntor) of a ligand is considered an approximate measure of configurational entropy loss. Rotational and translational entropies are difficult to compute. Here, it is assumed that these changes are about the same for all inhibitors. As a consequence, this term cancels when one considers relative binding affinities of various ligands for the same receptor, as done here.

In summary, for a molecular species *i*, the potential energy is calculated as the sum of the Lennard-Jones, Coulombic and dihedral energy terms

$$U_i = U_{LJ,i} + U_{Coulombic,i} + U_{Dihedral,i} \quad (9)$$

and the solvation energy for the species *i* is computed as the sum of the polar, Poisson-Boltzmann component (PB) and the non-polar surface area (NP) terms:

$$W_i = W_{PB,i} + W_{NP,i}. \quad (10)$$

The total free energy for the species *i*, is then, calculated as the sum of potential and solvation energies:

$$G_i = U_i + W_i. \quad (11)$$

Treatment of multiple conformations

A detailed physical model of the binding process of a ligand to its receptor considers an ensemble of conformations of the free ligand, and of the ligand in the complex, where the free energy of each conformation can be calculated according to equation (11). The free energy of the species can then be computed as the Boltzmann sum of the free energies of the conformations of that species, as shown

in equation (12). To account for the configurational entropy change, a term proportional to the number of rotatable bonds of the ligand is added to that energy. The resulting expression for n conformations of the free ligand is:

$$G_L = -RT \ln \Sigma_n e^{-G_i/kT} + c N_{tor.} \qquad (12)$$

Similarly, for an ensemble of n conformations of the complexed ligand, the free energy of the ligand in complex (G_{PL}) is calculated as the Boltzmann sum of the free energies of all the conformations:

$$G_{PL} = -RT \ln \Sigma_n e^{-G_i/kT} \qquad (13)$$

Here it is assumed that the ligand's bonds are no longer rotatable, so the additional term of Equation (12) is omitted.

In the present method, multiple conformations of the free ligand are generated in the absence of the protein, and their energies are computed and combined as described above. Similarly, multiple conformations of the ligand-protein complex are generated, with the protein part currently held rigid, and again the energies of the conformations are computed and combined. When all the ligands in a series are modeled as binding to exactly the same protein sequence and structure, then the internal energy of the protein conformation is a constant in the calculations and will cancel when relative binding affinities are computed for the ligands.

However, when different ligands are fitted into different conformations of the protease, and/or different mutants, then the computed binding energy includes not only ligand-protein interactions but also differences in the internal energy of the protein. Some of these changes result from conformational differences that are unrelated to binding and result from arbitrary details such as the arrangement of side-chains far from the binding site, depending upon the crystal structures used. If any useful signal is to be obtained from the calculations, it is important that these rather random energy components be eliminated. This can be done by forming a model of the free protein, for each conformation used, which consists of the complexed conformation of the protein without the ligand, and then subtracting the free energy of this isolated protein from that of the complex.

2.2 Optimization of Lead Drug Compound

Computational ligand design and optimization in drug discovery

Computer-aided drug design involves the design of compounds that bind with high affinity to key regions of biologically important molecules like enzyme active sites, leading to inhibition or alteration of the activity of the target. A rational strategy for ligand design may involve *de novo* design of a potential lead compound followed by chemical optimization of the compound to enhance its

binding affinity for the targeted receptor. Several methods of de novo ligand design and/or optimization of lead compounds exist, of which some representatives are now discussed.

One way to optimize an existing ligand is to manually modify it via computer graphics software and study the new interactions of the modified ligand with the receptor using three-dimensional molecular graphics programs like QUANTA and Insight (1). However, it is difficult to test many possible changes by this approach, and hence several automated approaches have been developed.

A more detailed and automated approach (2, 3) involves the use of the multiple copy simultaneous search (MCSS) method, which is suitable for both *de novo* construction of ligands and optimization of known ligands. In the de novo design mode, the MCSS method uses molecular dynamics with the CHARMm force field (4) to search for optimal positions and orientations of small chemical fragments, each with a single functional group in most cases, in the binding site of the targeted receptor. Fragments in these favored positions are subsequently connected to form candidate ligands. An appropriate set of chemical fragments is one in which most organic molecules can be described as a collection of such groups. Functional groups that may be considered are acetonitrile, methanol, acetate, methyl ammonium, dimethyl ether, methane, acetaldehyde and isobutane. Such groups are simple enough to facilitate discovery of favorable positions in the receptor binding site but complex

enough to approximate the steric and electrostatic interactions of a chemical moiety forming part of a complete ligand.

To optimize a known ligand with the MCSS method, set of functional groups is chosen so that the desired ligand may be reconstructed. The functionality maps obtained from the MCSS method may then indicate ways of modifying an existing ligand to improve its binding affinity for the receptor. For example it may be observed that a different stereochemistry is required for formation of additional hydrogen bonds by the ligand with the receptor. Another approach may be to determine whether a fragment finds a low-energy position next to an existing ligand. Such fragments may be chemically linked to the ligand to increase affinity.

A third approach to ligand optimization involves automatically linking new groups to an existing ligand or scaffold. Implementations of such an approach are found in the programs LUDI (5) and GROW (6). LUDI makes use of statistical data from small-molecule crystal structures to determine possible binding geometries of a ligand that interacts with hydrogen bonding and hydrophobic sites of the receptor. The small molecules are then iteratively linked, and the growth process is evaluated with a simple energy function that has terms to account for hydrogen bonding, ionic interactions, lipophilic interactions, and changes in ligand entropy due to freezing of internal degrees of freedom. GROW, used for ligand design by Moon and Howe, utilizes a template set and iteratively pieces the library templates together, within a model of the target receptor, in a manner similar to LUDI. Both LUDI and

GROW can be used for ligand optimization by determining whether the existing ligand should have a different stereochemistry or by adding functional groups to the existing ligand to enhance the computed binding affinity.

The above approaches all use simplistic treatments of electrostatic interactions, which are widely believed to be important determinants of affinity. For example, they all neglect the energy cost of desolvating polar groups upon binding. Also, all the methods described above rely upon trial and error during ligand construction, rather than taking a guided approach to speed the construction of improved ligands. This tends to make the calculations costly. The main theme of the present project is to utilize a full electrostatic model, including both Coulombic interactions and desolvation penalties for polar groups, to design and/or optimize ligands. Gradients of the binding free energy with respect to atomic charges are used to guide ligand modifications and thus speed the process. The following sections describe the electrostatic model used here, and the gradient-based optimization procedure.

Detailed Description of Continuum model of molecular electrostatics

In the binding reaction of a ligand with a receptor in an aqueous solution, interactions with the solvent and intramolecular Coulombic interactions in the unbound states are exchanged for intermolecular interactions between the ligand and receptor in the bound complex. The net change in the electrostatic component of

the binding free energy is thus the result of a delicate balance between the ligand-receptor Coulombic interaction and desolvation effects. This balance is captured by a continuum electrostatics model (7) (Figure 4), which describes the ligand and receptor as low-dielectric regions with embedded charges surrounded by high dielectric solvent.

The total electrostatic energy of a molecule or a complex of molecules is given by

$$G_{elec} = \frac{1}{2}\sum_{1}^{N} q_i \phi_i \quad (14)$$

where q_i is the charge on atom *i* and ϕ_i is the electrostatic potential at atom *i*.

The electrostatic potentials at an atom can be obtained by solving the linearized Poisson-Boltzmann equation as described earlier.

Given the electrostatic potentials from a solution of the linearized PB equation, the total electrostatic free energy of assembling solute charges from infinite separation into a cavity with dielectric constant, D_{int}, surrounded by an external continuum medium with dielectric constant, D_{solv}, is given by,

$$G_{elec} = \frac{1}{2}\sum_{1}^{N} q_i \phi_i = \frac{1}{2}\sum_{1}^{N} q_i (\phi_i^C + \phi_i^R) \quad (15)$$

where N is the total number of solute charge centers, ϕ_i is the potential at the position of charge q_i, and ϕ_i^C and ϕ_i^R are the Coulomb and reaction field potentials, respectively. The reaction field potential is due to the polarization induced in the high dielectric solvent by the charges of the ligand and the protein.

The Coulombic potential is given by

$$\phi_i^c(r_j) = b_{ij} q_i = \frac{q_i}{4\pi D_{\text{int}} \varepsilon_0 r_{ij}} \qquad (16)$$

where $\phi_i^c(r_j)$ is the potential produced by charge located at position r_i, at the position r_j, b_{ij} is a proportionality constant, D_{int} is the dielectric constant inside a molecular cavity, ε_0 is the permitivity of vacuum and r_{ij} is the distance between charges at positions i and j.

The reaction field potential at position of charge q_i induced by a charge q_j is proportional to the charge q_j and opposite in sign

$$\phi_i^R = c_{ij} q_j \quad (17)$$

where c_{ij} is a proportionality constant less than zero whose value depends on the shape of the molecule or complex and position of the charges. Consequently, for both the free ligand and the complex the electrostatic free energy has a quadratic functional form

$$G_{elec} = \frac{1}{2}\left[\sum_{i=1}^{N} c_{ii} q_i^2 + \sum_{i=1}^{N} q_i \sum_{j \neq i}^{N} (b_{ij} + c_{ij}) q_j \right] \quad (18)$$

where the first term in the right-hand-side is the interaction of charge i with the reaction field potential induced by charge i itself, and the second term is due to the Coulombic and reaction field potentials produced at q_i by all other charges q_j.

Electrostatic optimization

When a ligand binds a receptor the change in Coulombic interaction is usually favorable due to the complementarity of the charge distribution of the ligand and the receptor binding site, while the desolvation of the ligand can be expected to be energetically unfavorable due to stripping of solvent molecules from the polar parts of the ligand and receptor. Consequently, in most complexation reactions, monotonically increasing the charge of the ligand makes the Coulombic interactions more favorable but makes the desolvation penalty more unfavorable. Hence, in order to maximize binding affinity the charge distribution on the ligand should be adjusted so that the gain in Coulombic interactions upon binding is maximized, while the desolvation penalty is minimized.

Recently, the Tidor and Purisima groups have developed the concept of electrostatic optimization of ligands (8-9) showing that there exists a single charge distribution that optimizes the binding energy of the ligand for a targeted receptor, assuming the ligand's conformation does not change upon binding.

This result can be understood as follows. In equation 18 the coefficient of the quadratic terms for the free ligand or the complex is given by

$$c_{ii} = \phi^{R}_{q_i=1} \quad (19)$$

where $\phi^{R}_{q_i=1}$ is the reaction field potential at charge i induced by a unit charge located at the position of q_i. This quantity is always negative based upon equation 17. As a result the electrostatic free energy of the free ligand and the complex are parabolic functions of charge with the parabolas opening downward. The change in the electrostatic free energy upon binding has a quadratic functional form, too:

$$\Delta G_{elec} = G_{elec}^{complex} - G_{elec}^{free} = \frac{1}{2}\Big[\sum_{i=1}^{N} q_i^2 (c_{ii}^c - c_{ii}^f) + \sum_{i=1}^{N} q_i \sum_{j\neq i}^{N} q_j \{(b_{ij}^c - b_{ij}^f) + (c_{ij}^c - c_{ij}^f)\}\Big] = \frac{1}{2}(\sum_{i=1}^{N} a_{ii} q_i^2 + \sum_{j\neq i}^{N} a_{ij} q_i q_j)$$

(20)

Here the coefficients with superscript "c" are for the complex, the coefficients with the superscript "f" are for the free ligand and receptor, a_{ii} denotes the difference of the c_{ii} coefficients, and a_{ij} denotes the sum of differences of the b_{ij} and c_{ij} coefficients. Normally, the ligand interacts more strongly with the solvent before it is bound rather than after, so the reaction field felt by q_i is usually smaller in magnitude in the complex than in the free ligand. This implies that $c_{ii}^c > c_{ii}^f$ and hence a_{ii} is always positive. As a consequence, the electrostatic free energy of binding ΔG_{elec} may be represented by an N-dimensional parabola that is concave up and has a unique minimum with respect to the charge q_i (Figure 5). The charges corresponding

to this minimum are expected to optimize the binding affinity of the ligand and receptor.

Traian and Purisima showed that the concept of optimal charges holds good for systems with irregular dielectric boundaries, such as a host-guest complex of 18-crown-6 ether with alkali metal ions, a complex of calcium with the calcium-binding protein carp parvalbumin, and the complex of the covalent epoxysuccinyl inhibitor CA030 with human cathepsin B (9). Kangas and Tidor showed that an *endo*-oxabicyclic transition-state analogue of the enzymatic reaction of chorismate mutase from *Bacillus subtilis,* which exhibited good electrostatic affinity for the protein, had charge distribution complementary to the enzyme active site throughout much of the binding site, but had potential for improvement at a carboxylate since that group paid a substantial desolvation penalty upon binding and did not recover significant compensatory electrostatic interactions with the enzyme (8). Their calculations showed that replacement of the carboxylate group with the isosteric nitro group should improve electrostatic binding energy by 2-3 kcal/mol due to a decrease in the desolvation penalty of the ligand and smaller losses in other electrostatic interactions. The above prediction was qualitatively, though not quantitatively, borne out by experiment (10).

The published charge optimization methods described above have yielded promising results but are based on the optimization of charges over the entire ligand. This approach may not provide an accurate indication of the local

requirements for favorable charge complementarity between the ligand and the receptor and consequently may provide inaccurate guidance for lead optimization. For instance, in the complex of HIV-1 protease with its inhibitor saquinavir, the negative charge on the catalytic aspartates of the protease means that nearby ligand atoms prefer to become more positively charged in order to maximize affinity. The optimal positive charges at these locations effectively screen the negative potential produced by the aspartyl groups, so that more distant atoms may now prefer to be more negative, depending upon their local environment. However, if the charges of ligand atoms near the aspartates are held fixed while optimizing more distant atoms, then the more distant atoms can still feel the negative potential of the aspartates and may prefer to become more positively charged as a consequence. Thus, local optimization of charges may produce quite different results from global optimization.

The methods of Tidor and Purisima require a separate PB calculation for each ligand charge and hence is computationally slow. The present design algorithm will not have these drawbacks of the published methods.

Chapter 3: Methods

3.1 Parameterization of an Energy Model

For the free ligands, twenty low-energy conformations were identified by using a conformational search method implemented in our laboratory's in-house software (15). For the sake of computational speed and simplicity, bond lengths and bond angles are held fixed. Thus, for the free ligand, only torsional angles were allowed to vary, while for the complexes, the position and orientation of the ligand also was included in the search. During the search, energies were computed with the empirical force field CHARMM, with a distance-dependent dielectric constant to account approximately for the tendency of the high-dielectric solvent to weaken electrostatic interactions. The low-energy conformations that resulted were then post-processed to yield energies according to the model described above, with fitted coefficients as described below.

For bound ligands, the method of generating conformations depended upon whether a crystal structure of the complex was available. For all but one of the cyclic ureas considered, no crystal structure is available. The missing bound conformations were therefore generated by docking these compounds into the crystal structure 1hvr, which was solved with a cyclic urea inhibitor bound. The dihedral sampling range was +/-180°, the translational search range was a cubic translational box with side of 10A, and the rotational sampling ranges were +/-180°. For ligands for which crystal structures are available, the ligands were locally optimized around the known conformations by redocking them with dihedral sampling ranges of +/-10° for each

dihedral except for hydroxyl hydrogens, which are sampled in the range of +/-180°, a cubic translation box with side 1A, and rotational sampling ranges of +/-5°. Such minimizations allow little deviations from the crystal structure conformations of the ligands but avoid interatomic clashes while using the force field of our choice, since a different force field might have been used to refine the crystal structures. During these samplings, protein is considered rigid.

For conformation, electrostatic energies and surface area energies are computed using UHBD (17). The free energy of free ligand and the ligand in the complex is computed as the Boltzmann sum of the energies of 20 conformations of the respective species, as described above, and the unbound protein's energy is approximated as described earlier. Relative binding energies of the ligands are calculated for each ligand with reference to a specific ligand, chosen arbitrarily, as the difference between the binding energy of the former and the reference ligand:

□□GL = □GL - □GReference Ligand. (21)

The relative theoretical binding energies are fitted against the relative experimental binding energies for parameterization of the energy model. The experimental binding energies are calculated from dissociation constants of ligands as shown above in equation 2. The relative experimental binding energy of a ligand is calculated with

respect to the ligand used as reference for calculation of relative calculated binding energy:

$$\Delta\Delta G_L = \Delta G_L - \Delta G_{\text{Reference Ligand}} \qquad (22)$$

The root mean square deviation (RMSD) of relative calculated and relative experimental binding energies serve as the measure of error in the theoretical affinities during optimization. It is calculated as:

(23)

An initial plot of experimental and calculated binding energies of the cyclic ureas with HIVP (Fig 6a & b) showed a poor correlation between the experimental and theoretical values and we conjectured that, although the energy terms might be basically reasonable, the final results are poor because the terms are not well balanced. Therefore, fitted scaling terms are used for computation of binding energies, as has been done by other authors in parameterizing energy models with different functional forms (20). The raw energy terms are, hence, scaled prior to computing binding affinities. For each conformation of L or PL, and the suitable conformation of P, the scaled free energy is computed as:

$$G_{scaled,i} = \alpha U_{LJ,i} + \beta U_{Coulombic,i} + \gamma U_{dihedral,i} + \delta S_{ntor} + \varepsilon W_{PB,i} + \varsigma W_{surface\ area\ energy,i} +$$

□1□ (24)

The additional coefficient, □, takes into account any error in the calculated binding affinity of the reference ligand. RMSD of scaled theoretical affinities relative to experimental are minimized using "Down Hill Simplex" algorithm (11). This minimizer was chosen because it is a robust local optimizer capable of minimizing single-valued functions in multiple dimensions. It is not constrained by conditions like monotonocity, convexity or differentiability of the function optimized. It performed reasonably fast for our purpose.

The experimental data, consisting of ligands and their affinities, were separated into training and test sets. Initially, a set of 60 cyclic ureas (Fig 7) with dissociation constants (Kds) ranging from 0.25 nM to 21000 nM (12-14) was used as the training set for parameterization of the energy model, and another set of 20 cyclic ureas with Kds ranging from 0.41 nM to 25000 nM was used in the set to test the optimized model. In addition a set of ligands, with chemical structures different from the cyclic ureas and having dissociation constants ranging from 0.02 nM to 2300 nM was added to the test set. Crystal structures of this latter set of ligands are known. The latter set of ligands contained the FDA approved drugs presently available in the market.

Several variants of the energy model were parameterized and tested for correlation of experimental and predicted binding energies. Models both including and excluding van der Waals energies, solvation energy models with protein dielectric constants of 1 and 20 and energy models with and without Poisson-Boltzmann solvation terms were tested. Moreover, different protonation states of the critical catalytic aspartate residues of HIVP were evaluated for agreement between theoretical and experimental energies. For each case the lowest test rmsd case was chosen from amongst 25 optimization runs. The RMSD and Pearson's correlation coefficient between the experimental and calculated binding energies were used to evaluate efficiencies of the energy models.

3.2 Optimization of Lead Drug Compounds Theory

In preliminary studies, I have shown that one can improve the charge distribution of ligands to optimize binding, within the context of continuum electrostatics, by a much faster, simpler and more accurate method than those previously described in the literature. From Equation 20 the derivative of electrostatic binding energy with respect to charge is,

$$\begin{aligned}\frac{\partial \Delta G_{elec}}{\partial q_i} &= q_i(c_{ii}^c - c_{ii}^f) + \sum_{i=1}^{N} q_j\left[(c_{ij}^c - c_{ij}^f) + (b_{ij}^c - b_{ij}^f)\right] \\ &= \left[q_i c_{ii}^c + \sum_{j=1}^{N} q_j\left(c_{ij}^c + b_{ij}^c\right)\right] - \left[q_i c_{ii}^f + \sum_{j=1}^{N} q_j\left(c_{ij}^f + b_{ij}^f\right)\right] \\ &= \phi_i^c - \phi_i^f\end{aligned}$$

(25)

Thus, the derivative of the change in electrostatic binding energy with respect to charge q_i is merely the change in potential at charge i upon

binding. This quantity can be computed from two linearized Poisson-Boltzmann solutions, one for the free ligand and one for the bound ligand, and it can be used to predict the change in electrostatic binding energy for a given change in q_i. Thus, it is conjectured that these derivatives can be used to guide the chemical optimization of an existing ligand. For example, it should be advantageous to replace an atom for which the change in potential upon binding is very positive with a more electronegative element, and vice versa. More particularly, one can use the derivatives from Equation 25 to predict the change in the electrostatic part of the binding energy, when a set of charges *i* is changed, as

$$\Delta\Delta G_{elec} \approx \sum_i \left(\phi_i^c - \phi_i^f\right)\Delta q_i \quad (26)$$

It is expected that the best results will be obtained for atoms with large gradients because, as can be seen from the parabolic plot of the electrostatic binding energy against charge (Figure 5), charge changes at atoms with low gradients risk overshooting the optimum and may even yield worse binding.

I tested whether it was actually possible to use energy gradients to predict the change in electrostatic binding energy of a ligand when its charge is changed. This was done by studying the electrostatic component of the binding free energy of HIV protease with inhibitors XK263 and saquinavir, as well as cytosine deaminase with one of its inhibitors, using for now the assumption that the ligand keeps the same conformation in solution as in the bound complex. The crystal structures of the complexes were obtained from the Protein Data Bank (PDB). The

PDB codes for the complexes with XK263, saquinavir and the cytosine deaminase inhibitor are 1HVR, 1HXB and 1CTT respectively. The aspartyl dyad of HIV protease was considered to be monoprotonated and the crystal water was retained in the calculations for the saquinavir. The conformations of the free ligands were taken to be the same as in the bound complexes. The partial atomic charges of the ligands were calculated using Vcharge (14), and the receptors were assigned CHARMm charges with the program QUANTA (4). The linearized Poisson-Boltzmann equation was solved with the program UHBD (15), using a solute dielectric constant of 1 initially for the HIV protease systems and 4 for the cytosine deaminase inhibitor. The solvent dielectric constant used was 78.5, since the complexation reaction was assumed to take place in an aqueous medium. The solvent was considered to have physiological ionic strength, 150 mM. An ion-exclusion zone defined with a probe of radius 2 A was used. The dielectric constant in this region is that of the solvent, but the ionic strength is zero. The cubic grid used for the HIV protease calculations had a side length of 110 grid units and a grid spacing of 0.25 Angstroms; the corresponding parameters for the cytosine deaminase system are 200 and 0.35 Angstroms.

Chapter 4: Results and Discussion

4.1 Parameterization of Energy Model

Optimization of the model for cyclic urea ligands of HIV protease

Amongst the variants of energy models compared, the model excluding van der Waals term and using protein dielectric constant of 20 for electrostatic energy calculations, gave the highest correlation between calculated and experimental binding energies. The diprotonated HIVP yielded the best-predicted binding energies. This model yielded a Pearson's correlation coefficient (r) of 0.55 between calculated and experimental relative binding energies, for the test cyclic ureas (Fig 7), when energy coefficients were optimized with a training set having cyclic urea ligands only. The coefficients from optimization of these energy terms are shown in Table 2. When the Poisson-Boltzmann energy component was omitted from the model and the coefficients were re-optimized, the resulting correlation coefficient between calculated and experimental binding energies was somewhat worse, at 0.35 (Fig 8).

Testing the model on a broader set of HIVP ligands

When the model trained on cyclic ureas was tested on a broader set of ligands, the correlation coefficient for the test set fell to 0.09 (Fig 9). Interestingly, however, a high correlation coefficient of 0.62 was found for the subset of new ligands whose scaffold was most similar to that of the cyclic ureas (Fig 10). On addition of a subset of the broader set of ligands to the training set and re-training the model, the results improved. The overall correlation coefficient was still 0.03, but removing the only outlier increased the r-value to 0.5 (Fig 11).

Testing the model on a system other than HIVP

We conjectured that part of the reason the correlations fall when the data set is expanded beyond the cyclic ureas is that the broader set of ligands was studied in different laboratories with different assays and under somewhat different experimental conditions. In contrast, all of the cyclic urea data were obtained at the same company under at least nominally uniform conditions. In order to determine whether the model optimized with cyclic ureas alone would be transferable to a very different data set that had been obtained under its own uniform set of conditions, we applied the cyclic urea-derived model to a test set of Factor Xa inhibitors (18-19), without any further parameterization. This yielded a correlation coefficient of 0.43, which is significantly better than that obtained for the broad group of HIV protease inhibitors (Fig 12).

Discussion

Optimization of the model for cyclic ureas ligands of HIV protease

Due to approximation of van der Waals energy as 12-6 Lennard Jones potential, small changes in interatomic distances can yield large changes in van der Waals energy. Consequently, even for a well-docked ligand, this energy term can have a large positive value due to proximity of a ligand and receptor atom. On optimization, the large van der Waals energies tend to force the Lennard-Jones term to be assigned small, negative coefficients, which are physically meaningless (data not shown). Therefore, van der Waals energies were removed from the model and the surface area energies were allowed to take into account dispersion forces, hydrophobic effects and steric complementarity. The other energy term with a small negative coefficient was dihedral energy. However, the dihedral energy term is small and, hence, makes little contribution to the energy model. Therefore, the physically meaningless coefficients of dihedral energies are not a cause of concern. An interior protein dielectric constant of 20 reduced noise in Poisson-Boltzmann solvation energy calculations and, hence, gave better, predicted binding energies than a protein dielectric constant of 1. Though computation of Poisson-Boltzmann solvation energy is computationally expensive, inclusion of this term appears justified

by the improvement in the predictive power of the energy model. Comparing the chemical structures of the various cyclic urea ligands with that of the outlier in Fig 7, shows that there exist compounds with very similar chemical structures (Fig 13), but while the outlier has a Kd of 25000 nM, the structurally similar compounds have Kds ranging from 1.4 to 51, suggesting that the experimental data for the outlier may be flawed.

Testing the model on a broader set of HIVP ligands

The same basic physical forces govern binding of any ligand to its receptor. Hence, theoretically, an energy model optimized for the HIVP-cyclic-urea system should be valid for any receptor-ligand system. Such a general energy model should be expected to be good for predicting binding energies of any inhibitor of HIVP. The poor prediction we initially obtained for the binding energies of non-cyclic-urea molecules necessitated probing factors that might be leading to loss of general applicability of the above energy model. One possibility we considered was that binding of structurally different ligands might change receptor conformations differently. Hence, upon complexation with ligands with widely different chemical structures, the receptor internal energy change is considerable. There is wide variation in chemical structures of the non-cyclic-ureas. Consequently, conformations of the protein vary considerably between complexes of the non-cyclic-ureas**.** As a result, receptor internal energy change due to ligand binding needs to be taken into consideration adequately. As explained earlier, the receptor internal energy change

upon complexation is approximated in our calculations as the energy of the receptor conformation in the crystal structure of the receptor-ligand complex. This approximation may be inadequate, leading to the poor correlations between the experimental and calculated binding energies of the non-cyclic-ureas using a model optimized with cyclic urea training set. It was not required to consider protein internal energy changes for complexes with cyclic ureas because all the cyclic ureas were docked into the same crystal structure obtained from a complex of a cyclic urea with HIVP. Due to similar binding modes of the ligands, all the cyclic ureas were binding well in the aforesaid receptor. Also, relative binding energies with respect to a cyclic urea reference ligand were being considered in the calculations. Consequently, changes in protein internal energy were largely cancelled in case of cyclic urea ligands. Due to similar reasons, the theoretical binding affinities of the cyclic-urea-like ligands, amongst the non-cyclic-ureas, are in good agreement with their respective experimental values. Thus, the binding modes of the cyclic-urea-like ligands are similar to that of the cyclic ureas. Notably, a carbonyl O-atom of the cyclic ureas or cyclic-urea-like molecules interact with the flap Ile residues of HIVP. For other ligands a water molecule mediates the interaction of the ligand with the same Ile residues. Due to similar binding modes, change in receptor internal energy for the cyclic ureas and the cyclic-urea-like ligands can be expected to be close.

In order to address this problem, an attempt was made to use ligand conformations obtained by docking of non-cyclic-ureas in the context of the receptor structure solved with a bound cyclic urea. In principle, this would put all the

ligands on the same footing with regard to the internal energy of the protein. However, a very low correlation coefficient of 0.03 was obtained between experimental and predicted binding energies of the non-cyclic-ureas. Here it needs to be mentioned again that same calculations for cyclic-urea-like ligands yielded a correlation coefficient of 0.62 between the predicted and experimental binding energies. However, it should be noted that the above problem was partially resolved when some of the non-cyclic-ureas were included in the training set, in addition to the original cyclic ureas (Fig 11).

The catalytic aspartates of HIVP play an important role in ligand binding. In their complexes with HIVP, different ligands may induce different protonation states of these aspartates, but the protonation states of the critical aspartate residues are not known with any certainty. For cyclic ureas, the present calculations show that the diprotonated HIVP gives the highest correlation between experimental and predicted binding energies. However, the aspartate protonation states suitable for the ligands from crystals structures may be different from that of the cyclic ureas. The issues of these protonation states will be considered further in future work.

Unlike the Kds of the cyclic ureas, the Kds of non-cyclic-ureas were obtained from different sources. The experimental conditions, like pH, temperature and ionic strength, used for determination of Kds of non-cyclic ureas may be different. This may be a source of inconsistency between experimental and

calculated binding energies. For instance, lack of knowledge of the pH used for Kd determination make it hard to consider the correct ionization state of the ligands and proteins during binding energy calculations.

4.2 Optimization of Lead Drug Compounds

I tested whether changes in ligand charges guided by electrostatic gradients (Equation 10) would in fact lead to a more favorable change in the electrostatic energy upon binding, as computed with full solutions of the PB equation. Thus, predicted changes in electrostatic binding energies (Equation 11) were compared with electrostatic binding energy (Equation 9) changes computed from the full solutions for the PB equation after the ligand modifications. In this process, I focused on the atoms at which the electrostatic gradients were high for reasons explained above. I first looked at small artificial changes of single-atom charges, then at models of actual atom substitutions and small chemical additions that required recalculation of the partial charges of all the ligand atoms. Then I looked at serial modifications of a ligand, in which successive modifications were done according to the predictions based on the atomic electrostatic potential of the preceding version of the ligand. For most of the above cases, the changes in ligand charges guided by electrostatic gradients improved the electrostatic binding energy computed with full solutions for the PB equation. Moreover, the quantitative changes

in electrostatic binding energies from the predictions agreed reasonably well with the full electrostatic energy changes (Tables 3-7 & Figures 14-18).

I began by making a series of artificial charge changes at one atom of XK263 with a gradient of 5.34 kcal/mol/e. The agreement is better when the magnitude of the change in charge is not very small or not very large (Table 3 & Figure 14). With very small changes in charge, discrepancies arise due to rounding errors in electrostatic energy calculations using numerical methods. With large changes in charge, the electrostatic gradient of the atom changes significantly due to addition of the charge, as expected based upon the parabolic form of $\Delta G_{elec,}$ thereby giving rise to above-mentioned discrepancies.

I then made a series of essentially isosteric hydrogen to halogen changes at high-gradient (4.4-8.2 kcal/mol/e) atoms in XK263. The changes in the electrostatic binding energies predicted from the gradients agree quite well with the changes obtained by a full electrostatic recalculation, (Table 4 and Figure 15), with a R^2 value of 0.92. Most of the changes yield a more favorable electrostatic binding energy with respect to the original ligand, as expected according to the predictions.

Next, functional group modifications involving additional atoms were made in the HIV protease inhibitor saquinavir, according to predicted changes in electrostatic binding energies computed on the basis of the gradient of the

original ligand atom. The changes in the electrostatic binding energies predicted from the gradients agree quite well with the changes obtained by a full electrostatic recalculation, (Table 5 and Figure 16), with a R^2 value of 0.83. The electrostatic binding energies for the modified ligands did indeed become more favorable with respect to the original ligand, in accordance with predictions. For the modification of H14 of saquinavir to a hydroxyl group, the calculations were done for four different positions of the hydroxyl hydrogen. From such calculations it is observed that such rotatable bonds need to be in a particular orientation in order to improve the binding energy of the ligand.

Functional group modifications involving additional atoms were made for the cytosine deaminase inhibitor, too. Again, reasonable agreement was found between the changes in electrostatic binding affinities predicted based on the atomic electrostatic potentials and the computed changes in full electrostatic binding energies (Table 6 and Figure 17). The estimates and actual changes in electrostatic binding energies agree with a R^2 value of 0.86. The electrostatic binding energies for the modified ligands were more favorable with respect to the original ligand, in accordance with predictions, as in the above cases.

Sets of simultaneous multiple modifications in the cytosine deaminase inhibitor were not observed to improve the electrostatic binding energy markedly more than any of the isolated modifications (Table 6). This may be because, on making multiple changes at once, one change can alter the actual charge

and the optimum charge at another site, and thus yield unexpected results. In an attempt to resolve this problem, I made ligand modifications serially rather simultaneously. Serial modifications involve using the derivatives to guide an initial modification of the ligand, generating updated charges for the modified ligand, recomputing the atomic potentials of the modified ligand, and using these potentials to guide further modifications. The net outcome of the serial mutations turned out to be better than isolated modifications, as can be seen by comparing the improvement in electrostatic energy of the serial mutations (Table 7 and Figure 18) with the simultaneous multiple mutations done on the same inhibitor (Table 6). Also the estimates and actual changes in electrostatic binding energies agree with a R^2 value of 0.71.

Tables and Figures

Table

Peptide sequences	Cleavage domains
Cleavage sites in gag	
SQNY * PIVQ	MA-CA
ARVL * AEAM	CA-p2
ATIM * MQRG	p2-NC
QANF * LGKI	NC-p1
PGNF * LQSR	p1-p6
Cleavage sites in pol	
SFNF * PQIT	TF-PR
TLNF * PISP	PR-RT
AETF * YVDG	RT-RH
RKIL * FLDG	RH-IN

Table 1: The nine non-homologous sites at which HIV protease cleaves the viral polyproteins gag or gag-pol to yield the structural proteins and enzymes essential for the viral life cycle. The cleavage sites are denoted by an asterisk. NC, nucleocapsid; MA, matrix; CA, capsid; TF, trans frame peptide; PR, protease; RT, reverse transcriptase; IN, integrase;RH, RNAse H. p1, p2, p6 are structural proteins.

Energy Terms	**Coefficients**
Coulombic	1.08
Dihedral	-0.26
Poisson-Boltzmann Solvation	2.03
Surface area energy	0.57
Number of Rotatable Bonds	0.05
Offset(η)	-2.53

Table 2: Coefficients from optimization with a training set of cyclic urea ligands only

Change in charge	Predicted change in electrostatic binding energy	Full relative binding energy after artificial charge modification of ligand
-0.01	-0.053	-0.082
-0.05	-0.267	-0.266
-0.1	-0.535	-0.174

Table 3: Comparison of predicted changes in electrostatic binding energy based on the gradient at H66 of the original ligand and changes in electrostatic binding energies from full electrostatic calculations for the on artificial charge changes of H66 of the HIV protease inhibitor XK263. The calculations were done with a protein dielectric constant of 1.

Isosteric Modifications	Original gradient (Kcal/mol/e)	Change in charge due to modifications (e)	Predicted electrostatic binding energy	Full relative binding energy of modified ligand
H66→Br	5.35	-0.277	-1.56	-1.16
H66→Cl	5.35	-0.258	-1.37	-0.91
H66→F	5.35	-0.339	-1.68	-0.74
H47→Br	8.19	-0.303	-2.03	-2.03
H47→F	8.19	-0.332	-1.44	-0.86
H48→Br	3.88	-0.303	-0.74	-0.52
H48→F	3.88	-0.332	-0.05	0.29
H50→Br	6.99	-0.319	-3.78	-3.36
H50→F	6.99	-0.335	-5.51	-4.66
H57→Br	5.73	-0.277	-1.36	-0.68
H57→F	5.73	-0.339	-1.16	-0.36
H58→Br	4.39	-0.277	-1.08	0.16
H58→F	4.39	-0.339	-0.90	0.38
H67→Br	4.79	-0.277	-1.28	-0.59
H67→F	4.79	-0.339	-1.22	-0.33
H69→Br	7.75	-0.277	-2.58	-1.44

H69→F	7.75	-0.339	-3.26	-2.60
H71→Br	7.46	-0.292	-1.65	-0.84
H71→F	7.46	-0.331	-1.06	-0.33
H72→Br	6.16	-0.292	-1.28	-0.90
H72→F	6.16	-0.331	-0.64	-0.37

Table 4: Comparison of predicted changes in electrostatic binding energy based on the gradients of the atoms of the original ligand vs. actual changes in electrostatic binding energies from full electrostatic calculations for the modified ligand of the HIV protease inhibitor xk263 due to isosteric hydrogen to halogen changes. The calculations were done with a protein dielectric constant of 1.

Modifications	Original gradient at the atoms	Predicted relative electrostatic binding energy	Full relative binding energy of modified ligand
H14→OH	21.15	-10.14	-7.21
H46→NH_3^+	-3.11	-1.01	-2.09
H48→NH_3^+	-12.21	-1.18	-4.39
H48→NH_3^+ + H49→NH_3^+	-12.21 (atom 48) -21.53 (atom 49)	-10.72	-18.53
O34→NH_3^+	-15.61	-12.30	-11.51
H50→OH	-5.45	3.58	0.40
H14→OH (rotamer1)	21.15	6.98	2.78
H14→OH (rotamer2)	21.15	3.25	1.81
H14→OH (rotamer3)	21.15	7.85	2.83
H14→OH (rotamer4)	21.15	1.40	-2.12

Table 5: Comparison of predicted changes in electrostatic binding energy based on the gradients of the atoms of the original ligand vs. actual changes in electrostatic binding energies from full electrostatic calculations for the modified ligand of the HIV protease inhibitor saquinavir due to functional group modifications involving additional atoms. The calculations were done with a protein dielectric constant of 1.

Modification number	Modifications	Predicted relative electrostatic binding energy	Full relative electrostatic binding energy
A	C2O3→SO_2	-4.86	-3.03
B	C5H18H19→SO_2	-0.08	-0.53
C	H19→F	-1.53	-0.85
D	H27→OH	-0.04	-0.47
E	O12→CH_2 + H26→NH_3^+	-7.72	-4.60
G	H18→F + H19→F	-0.31	-0.30
F	B + D	-0.63	-1.26
H	D + G	-0.89	-1.03
I	C + D	-2.14	-1.54
J	A + B + E	-2.39	-3.14
K	A + B + D + E	-4.24	-3.61
L	A + D + G + E	-3.31	-3.23

Table 6: Comparison of predicted electrostatic binding energy based on the gradient of the atoms of the original ligand and actual changes in electrostatic binding energies from full

electrostatic calculations for the modified ligand of the cytosine deaminase inhibitor 3,4-dihydrozebularine due to functional group modifications. The calculations were done with a protein dielectric constant of 4.

Modification	Predicted relative electrostatic binding energy	Full relative electrostatic binding energies after modifications
C2O3→SO_2	-4.86	-3.03
C2O3→SO_2 + H19→F	-3.24	-5.06
C2O3→SO_2 + H19→COO^-	-3.45	-5.72
C2O3→SO_2 + H19→$COCH_3$	-3.15	-5.75
C2O3→SO_2 + H19→COO^- + O12→CH_2 + H26→NH_3^+	-8.30	-14.35

Table 7: Comparison of predicted electrostatic binding energy based on the gradient of the atoms of the original ligand and actual changes in electrostatic binding energies from full electrostatic calculations for the serial modifications of the cytosine deaminase inhibitor 3,4-dihydrozebularine. The calculations were done with a protein dielectric constant of 4.

Figures

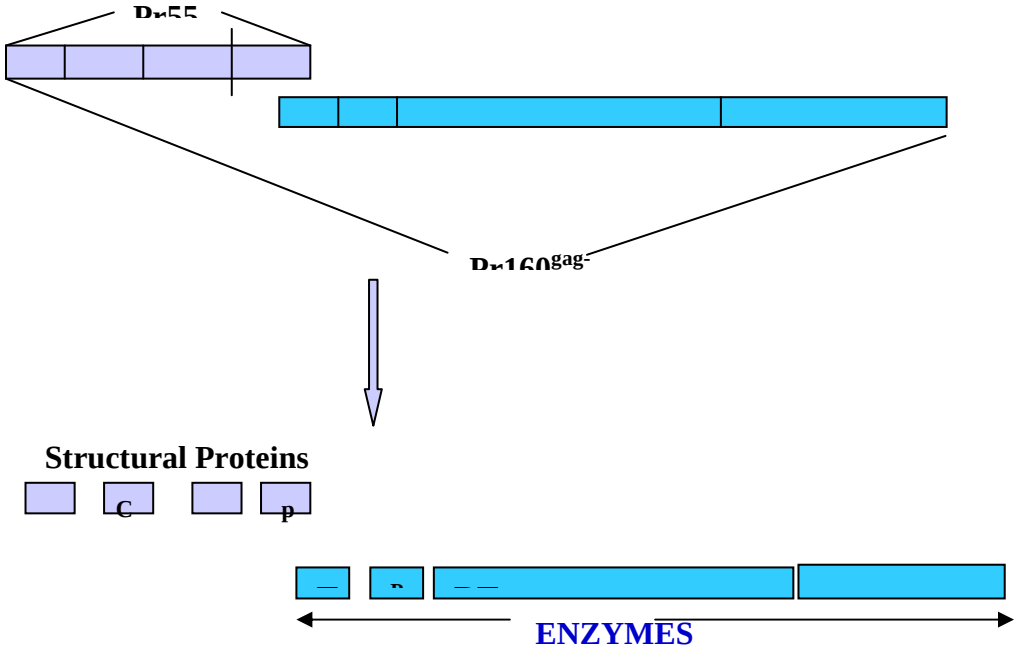


Figure 1: The protease cleaves the viral polyproteins gag or gag-pol to yield structural proteins and enzymes, including itself, essential for the viral life cycle.

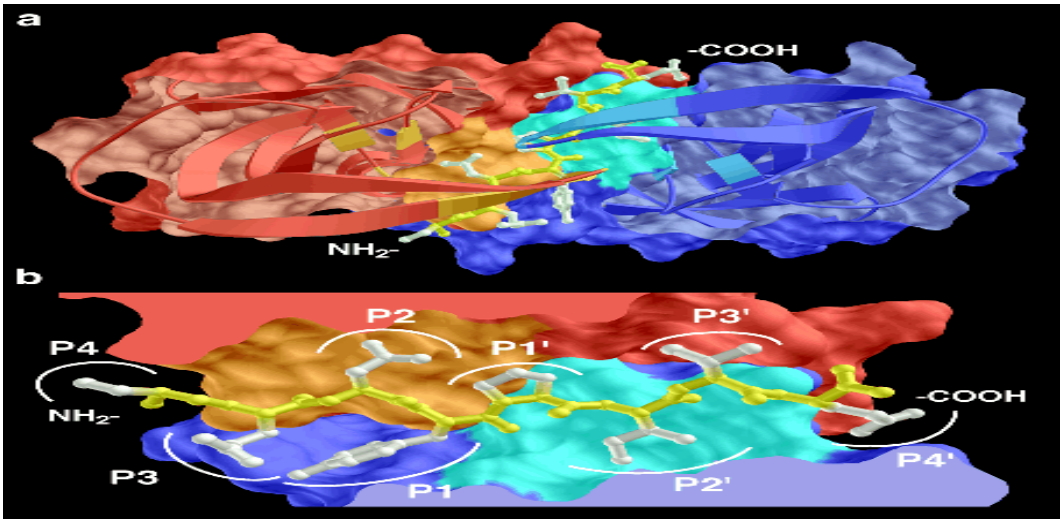


Figure 2: Complex of HIV-1 protease with a natural substrate. HIV-1 protease is represented by molecular surfaces and ribbon diagrams with subunit A in red and B in blue. A ball-and-stick representation of a natural polyprotein substrate (SQNYPIVQ) is bound in the active site. (a) View of the entire complex. (b) Close-up of the active site, with the eight peptide side chains (denoted P4 to P1 and P1' to P4') labeled. (Figure from reference 12)

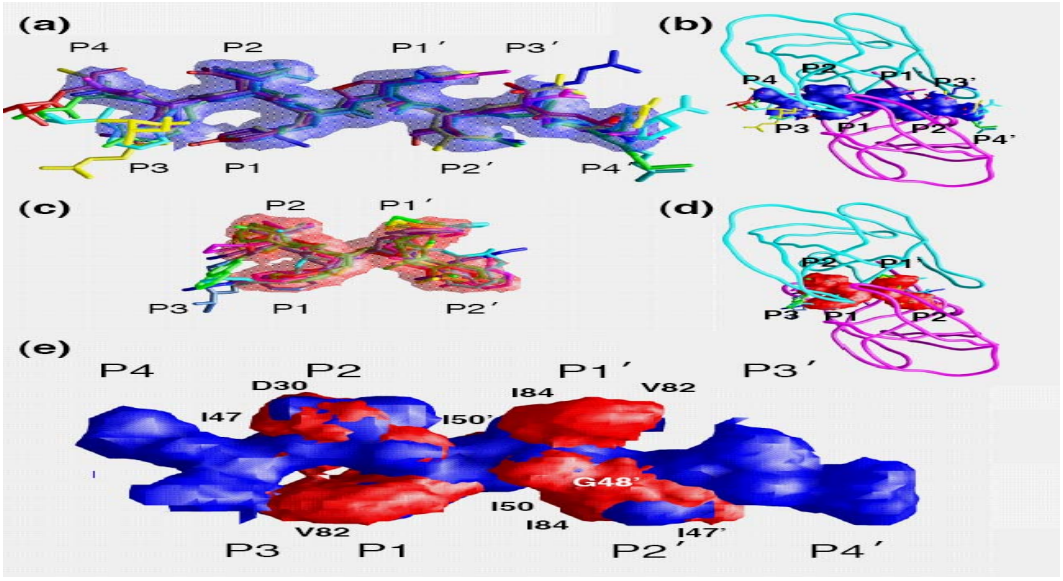


Figure 3: Substrate and inhibitor envelopes of HIV-1 protease. a) The substrate envelope from the overlapping van der Waals volume of four or more substrate peptides. The colors of the substrate peptides are: red, matrix-capsid; green, capsid-p2; blue, p2-nucleocapsid; cyan, p1-p6; magenta, reverse-transcriptase-ribonucleaseH; and yellow, rnaseH-integrase. b) The substrate envelope as it fits within the active site of HIV-1 protease. The alpha carbon trace is of the CA-p2 substrate peptide complex. c) The inhibitor envelope calculated from overlapping van der Waals volume of five or more of eight inhibitor complexes. The colors of the inhibitors are: yellow, Nelfinavir (NFV); gray, Saquinavir (SQV); cyan, Indinavir (IDV); light blue, Ritonavir (RTV); green, Amprenavir (APV); magenta, Lopinavir (LPV); blue, Atazanavir (ATV) and red, TMC114 d) The inhibitor envelope as it fits within the active site of HIV-1 protease e) Superposition of the substrate envelope (blue) with the inhibitor envelope (red). Residues that contact the inhibitors where the inhibitors protrude beyond the substrate envelope and confer drug resistance when they mutate are labeled. (Figure drawn from reference 13)

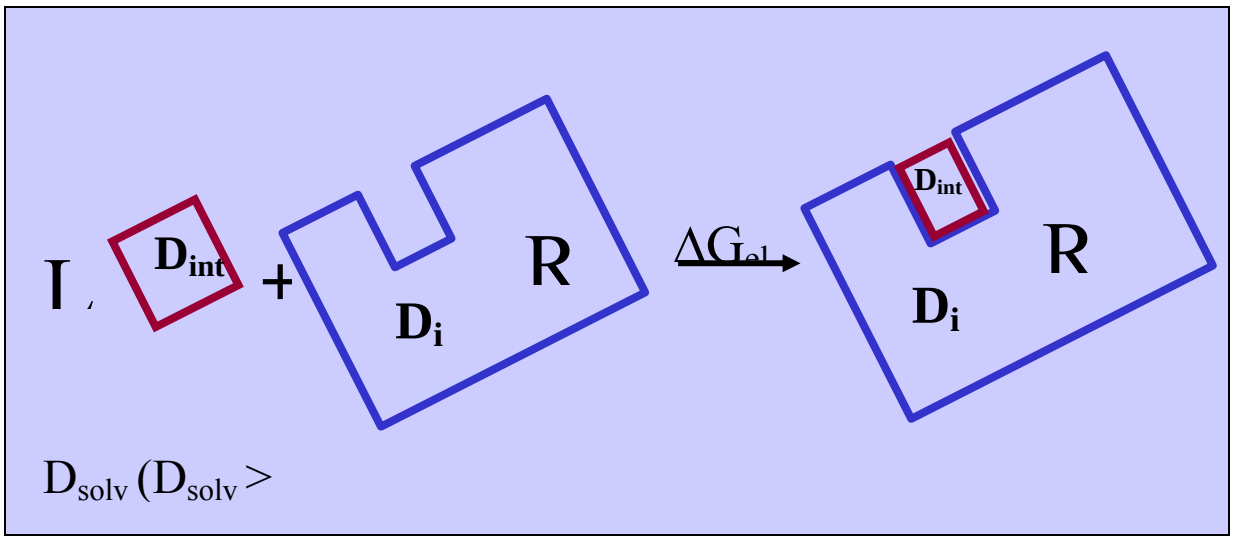
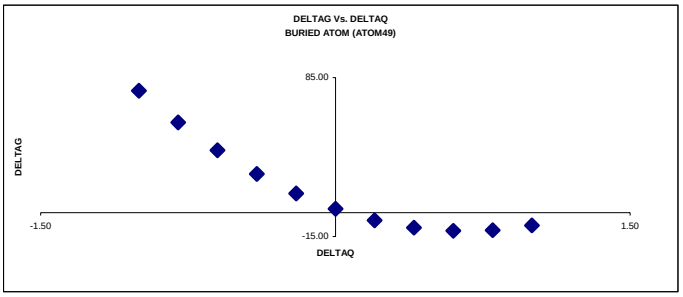


Figure 4: In the binding reaction of a ligand with a receptor in an aqueous solution, interactions with the solvent and intramolecular Coulombic interactions in the unbound states are exchanged for intermolecular interactions between the ligand and receptor in the bound complex. L, R and RL represent the free ligand, the free receptor and the complex respectively. D_{int}, the dielectric constant in the cavities representing the free ligand, free receptor and the complex, is low in comparison with that of the aqueous medium, D_{solv}. A charge embedded in the ligand is denoted by q_i while a charge embedded in the receptor is denoted by q_j.

a)



b)

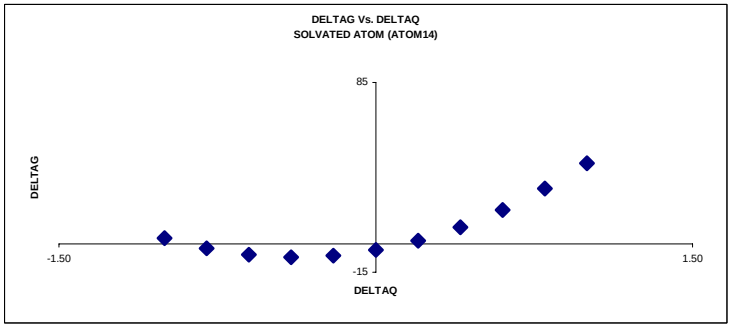
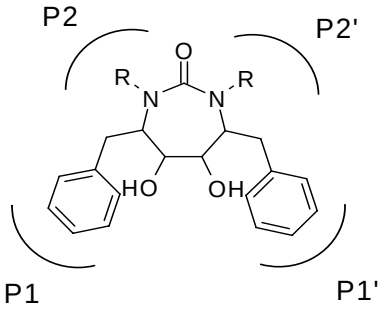


Figure 5: Variation of computed electrostatic free energy of binding, based upon full solutions of the linearized PB equation, with variation of charge on (a) buried atom (atom 49) and (b) solvated atom (atom 14) of saquinavir-HIV-1 protease complex.

A)


P2
P2'
O
R
N
N
R
HO
OH
P1
P1'

B)

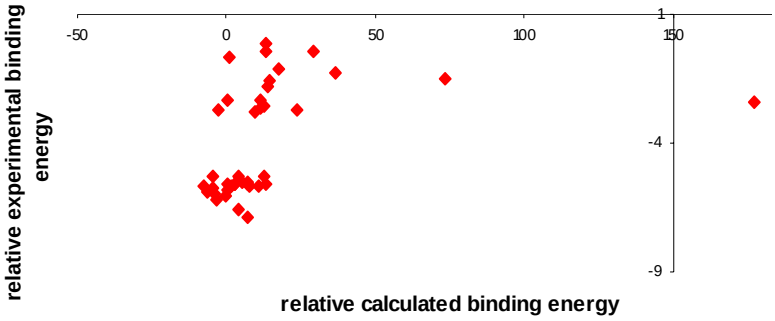
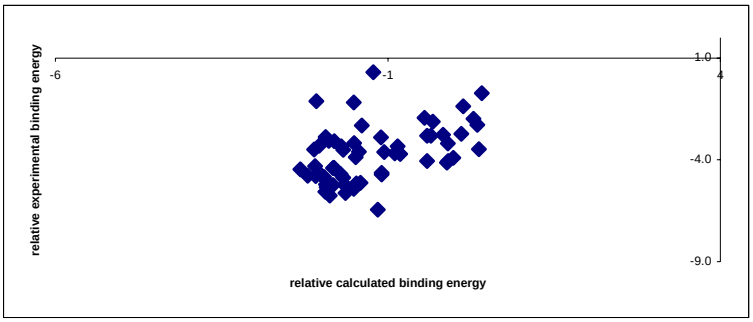

relative experimental binding energy
-50
0
50
100
150
1
-4
-9
relative calculated binding energy

Fig 6. a. Fig 1 Cyclic urea scaffold showing the substituents corresponding to the S1, S2, S1' and S2' binding pockets of HIVP. b. Experimental vs. Calculated Relative Binding Energies without scaling. RMSD between experimental and calculated binding energies was as high as 34.07.

a.



b.

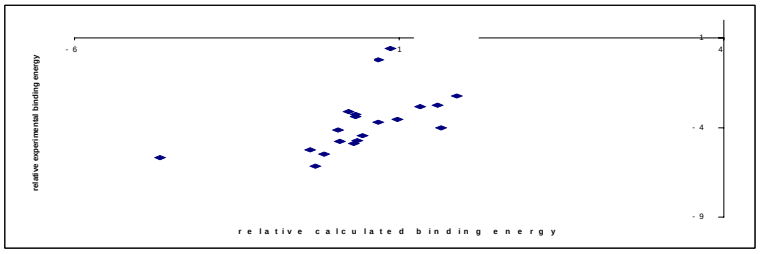


Fig 7: Plot of experimental binding energies Vs. calculated binding energies for training set (a) and test set (b) of cyclic urea ligands, when Coulombic, Dihedral, PB-solvation and Solvent accessible surface terms were included in the energy model. The result for lowest test set RMSD has been shown. Pearson's correlation coefficients of 0.4 and 0.55 were obtained between calculated and experimental binding energies for the training and test sets respectively.

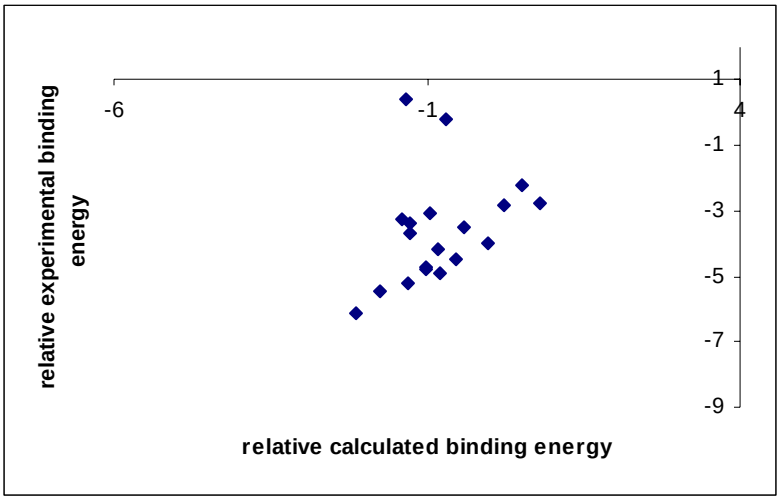


Fig 8: Plot of experimental binding energies vs. calculated binding energies for test set of cyclic urea ligands, when PB-solvation was excluded from the energy model. A correlation coefficient of 0.35 was obtained between calculated and experimental binding energies for this test set.

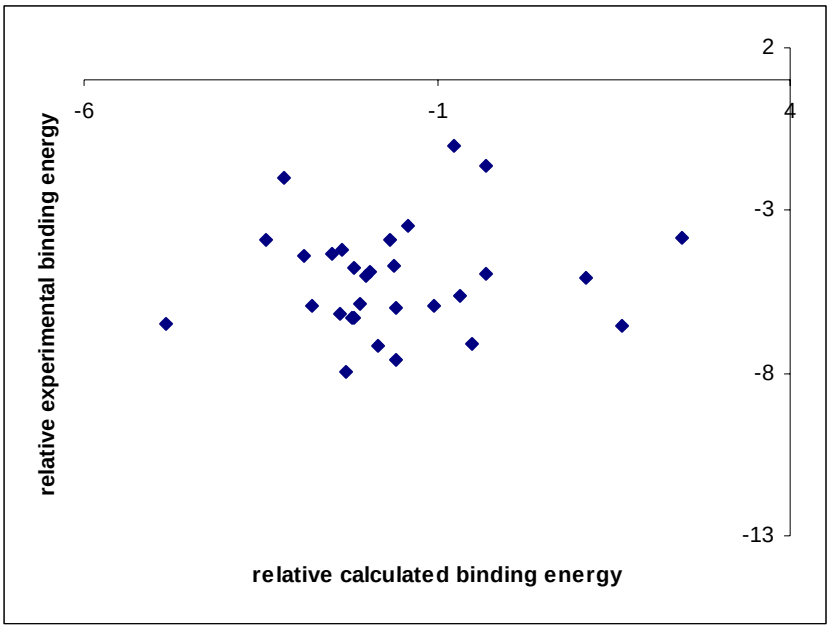


Fig 9: Plot of calculated vs. experimental binding energies of non-cyclic urea test ligands, when energy model includes Coulombic, dihedral, solvent accessible surface area and PB-solvation terms. Correlation coefficient is 0.09.

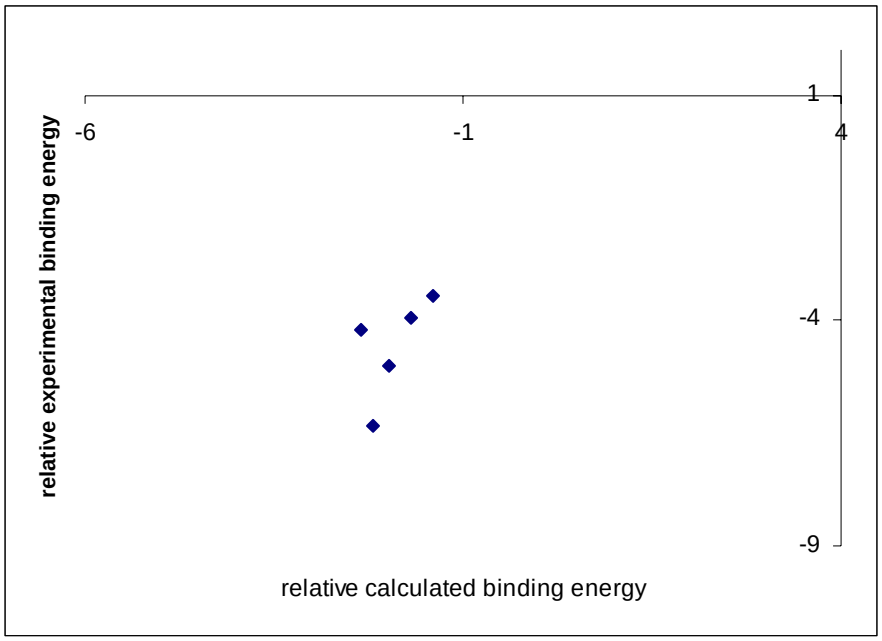
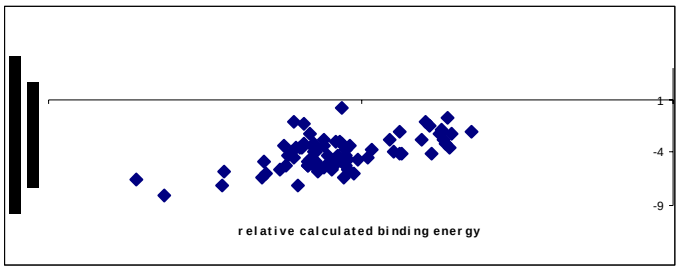


Fig 10: Plot of calculated vs. experimental binding energies of cyclic-urea-like test ligands, when energy model includes Coulombic, dihedral, solvent accessible surface area and PB-solvation terms. Correlation coefficient for these cyclic-urea-like ligands is 0.62.

a)



b)

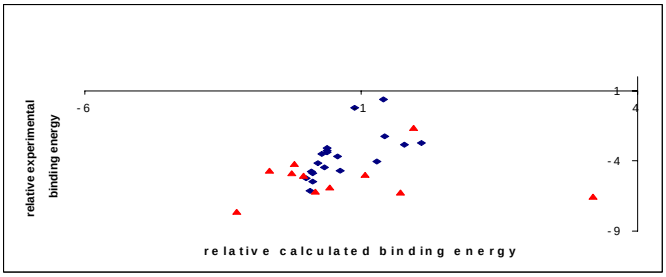


Fig 11: Plot of experimental binding energies vs. calculated binding energies for training set (a) and test set (b) of cyclic urea ligands and ligands other than cyclic ureas, when energy model ss optimized with a set of ligands having both cyclic ureas and the non-cyclic-urea ligands. Correlation coefficients of 0.59 and 0.03 are obtained between calculated and experimental binding energies for the training and test sets respectively. Taking off the outlier from the test set increases the correlation coefficient to 0.50.

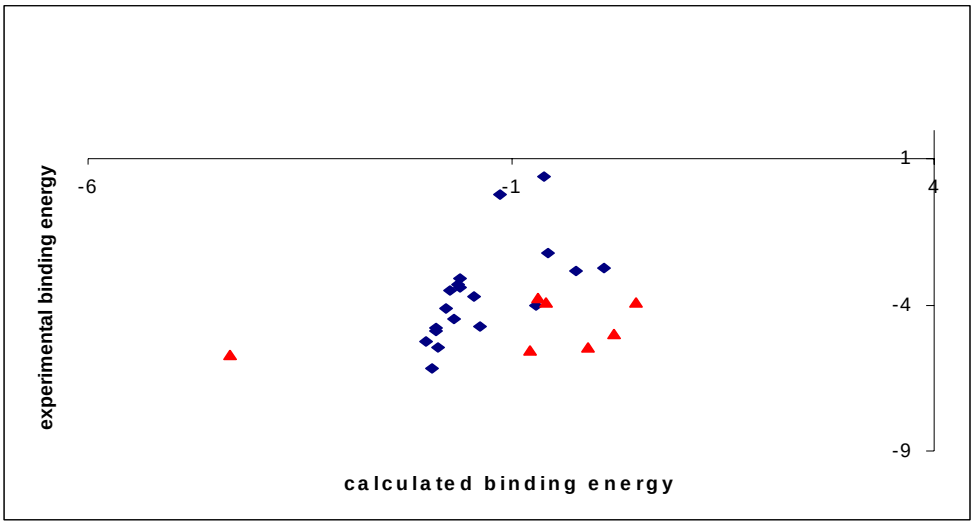
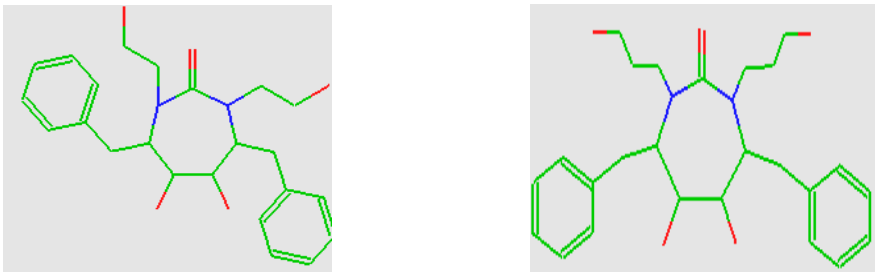


Fig 12: Plot of calculated vs. experimental binding energies of Factor Xa inhibitors (red) using energy model optimized with cyclic urea ligands and ligands other than cyclic ureas. Data are also shown for a test set of cyclic urea ligands (blue). The latter set is taken from the known crystal structures of HIVP-inhibitor complexes. Pearson's correlation coefficient for the Factor Xa inhibitors is 0.43. Errors in prediction of binding energies of Factor Xa inhibitors compare well with the error in estimates of binding energies of HIVP cyclic urea inhibitors.

a) b)



Kd : 25000 nM Kd: 51 nM

Fig 13: A representative cyclic urea ligand with structure very similar to the outlier marked "A" in Fig 3. The ligand in panel (a) is the outlier marked "A" while the one in panel (b) is the cyclic urea which structurally similar to the outlier. The K_d values (nM) of the ligands are given at the bottom of the chemical structures.

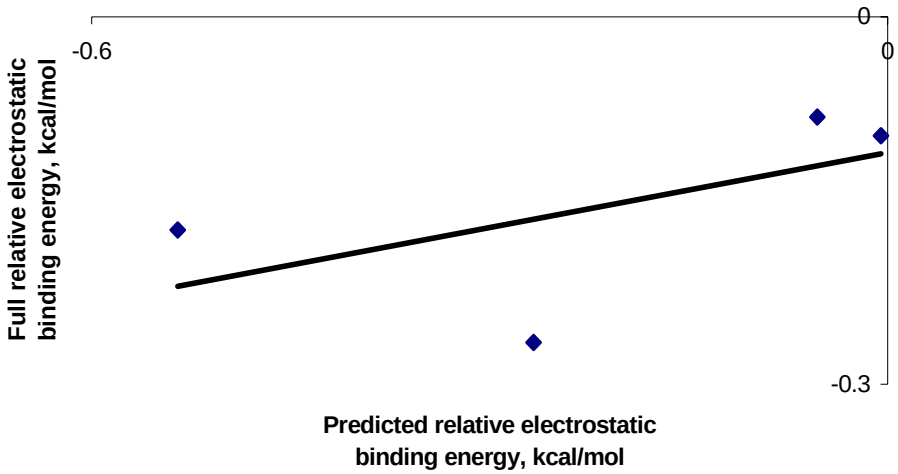


Figure 14: Predicted changes in electrostatic binding energy based on the gradient of H66 of the original ligand versus changes in electrostatic binding energies from full electrostatic calculations for the artificial charge changes of H66 of the HIV protease inhibitor XK263.

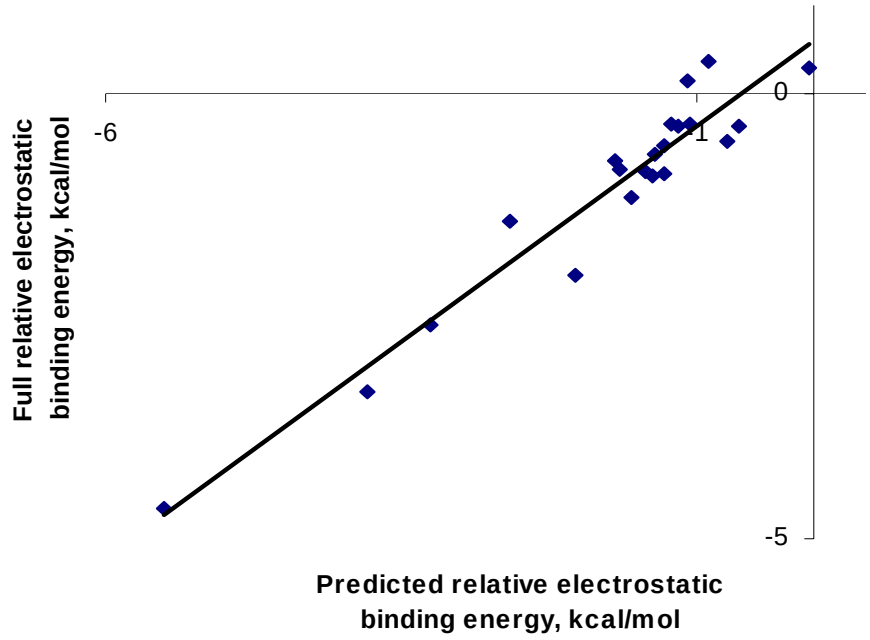


Figure 15: Predicted changes in electrostatic binding energy based on the gradients of the atoms of the original ligand versus actual changes in electrostatic binding energies from full electrostatic calculations for the modified ligand of the HIV protease inhibitor xk263 due to isosteric hydrogen to halogen changes. The estimates and actual changes in electrostatic binding energies agree with a R^2 value of 0.92.

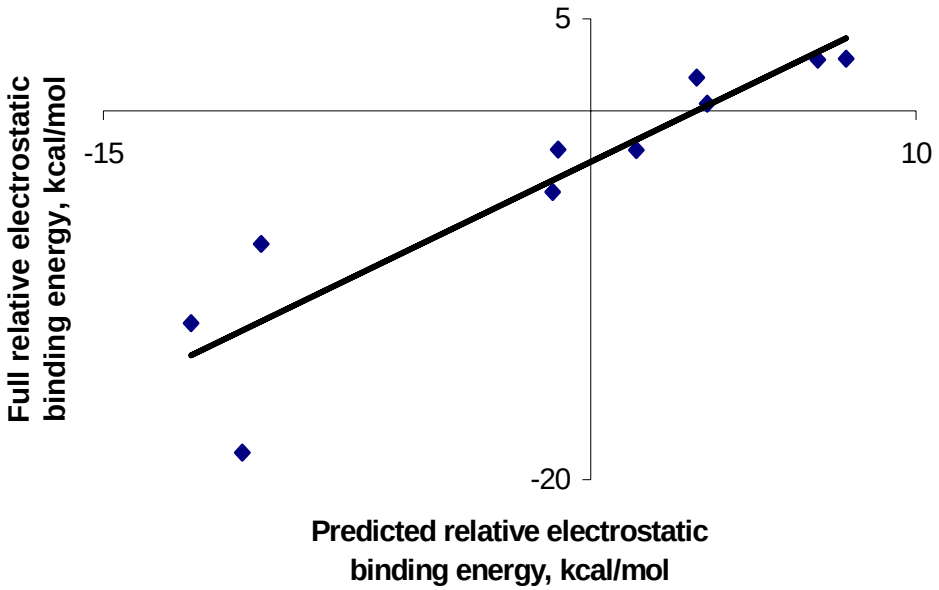


Figure 16: Predicted changes in electrostatic binding energy based on the gradients of the atoms of the original ligand versus actual changes in electrostatic binding energies from full electrostatic calculations for the modified ligand of the HIV protease inhibitor saquinavir due to functional group modifications involving additional atoms. The estimates and actual changes in electrostatic binding energies agree with a R^2 value of 0.83.

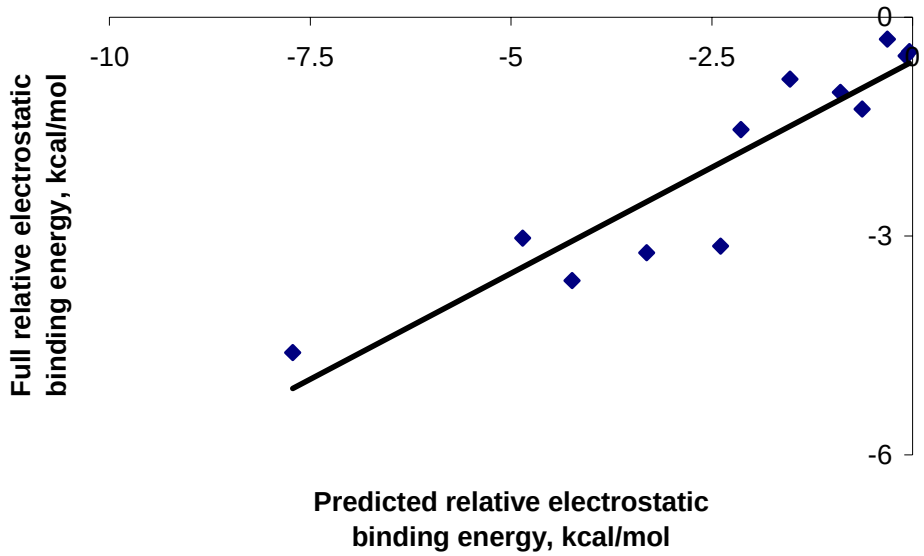


Figure 17: Predicted changes in electrostatic binding energy based on the gradient of the atoms of the original ligand versus actual changes in electrostatic binding energies from full electrostatic calculations for the modified ligand of the cytosine deaminase inhibitor 3,4-dihydrozebularine due to functional group modifications involving additional atoms. The estimates and actual changes in electrostatic binding energies agree with a R^2 value of 0.86.

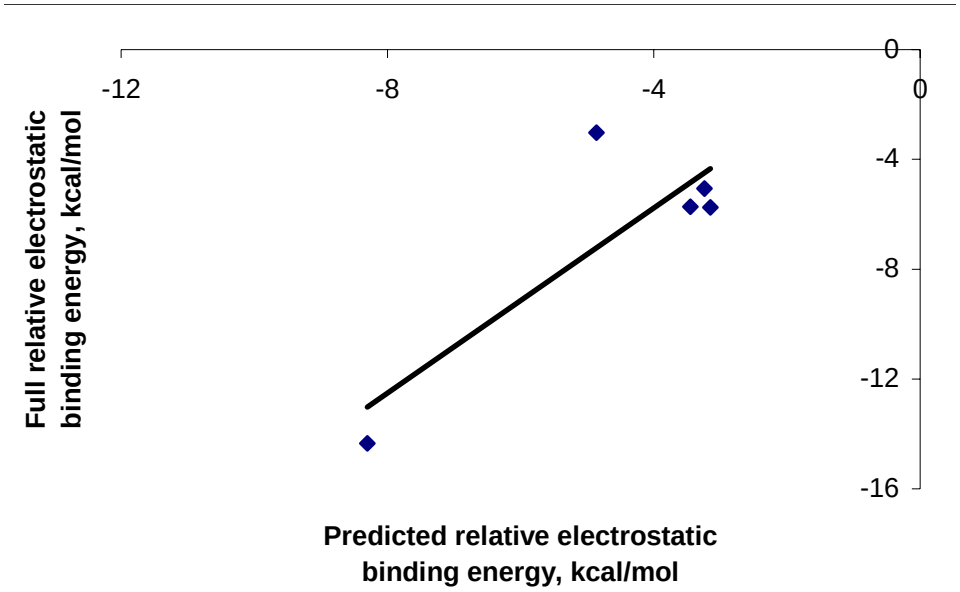


Figure 18: Predicted changes in electrostatic binding energy based on the gradient of the atoms of the original ligand and actual changes in electrostatic binding energies from full electrostatic calculations for the serial modifications of the cytosine deaminase inhibitor 3,4-dihydrozebularine. The estimates and actual changes in electrostatic binding energies agree with a R^2 value of 0.71.

References

1 Oroszlan. S. & Luftig, R.B. Retroviral Proteinases. Curr. Topics Microbiol. Immunol. 157. 153-185.

2 DeBouck. C. The HIV-1 protease as a therapeutic target for AIDS. AIDS Res. Hum. Retroviruses. 8. 153-164.

3 Kakuda, T.N., Struble, K.A., Piscitelli,S.C. Protease inhibitors for the treatment of human immunodeficiency virus infection. Am. J. Health Syst. Pharm. 55:233-254, 1918.

4 Ji, J.P., Loeb, L.A. Fidelity of HIV-1 reverse transcriptase copying RNA in vitro. Biochem. 31:954-958, 1992.

5 Larder, B.A. Viral resistance and the selection of antiretroviral combinations. AIDS 10(Suppl 1):S28-S33, 1997.

6 Wlodawer, A., Vondrasek, J. Inhibitors of HIV-1 protease: a major success of structure-assisted drug design. Annu. Rev. Biophys. Biomol. Struct. 27:249-284, 1998.

7 Prabu-Jeyabalan, M., Nalivaika, E., Schiffer, C.A. How does a symmetric dimmer recognize an asymmetric substrate? A substrate complex of HIV-1 protease. J. Molec. Biol. 301:1207-1220, 2000.

8 Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States D.J., Swaminathan, S., Karplus, M. CHARMm: A program for macromolecular energy minimization and dynamics calculations. J. Comput. Chem. 4:187-217,1983.

9 Gilson, M.K., Honig, B. Calculation of the total electrostatic energy of a macromolecular system: Solvation energies, binding energies and conformational analysis. Proteins: Struct. Func. Gen. 4:7-18, 1988.

10 Mardis, K.L., Luo, R., Gilson, M.K. Interpreting trends in the binding of cyclic ureas to HIV-1 protease. J. Molec. Biol. 309:507-517, 2001.

11 S.A. Teukolsky, W.T. Vetterling, B.P. Flannery. Numerical Recipes in C: The art of Scientific Computing. Cambridge University Press, New York, second edition 1992.

12 Lam, P.Y.S., Ru, Y., Jadhav, P.K., Aldrich, P.E., Delucca, G. V., Eyermann, C.J., Chang, C.H., Emmett, G., Holler, E.R., Daneker, W.F., Li, L.Z., Confalone, P.N., McHugh, R.J., Han, Q. & Li, R.H. et al. Cyclic HIV protease inhibitors: synthesis, conformational analysis, P2/P2' structure-activity relationship, and molecular recognition of cyclic ureas. J.Med. Chem. 39, 3514-3525. 1996

13 Rodgers, J.D., Lam, P.Y.S., Johnson, B.L., Wang, H.S., Ko, S.S. Seitz, S.P., Trainor, G.L., Anderson, P.S. Klabe, R.M. Bacheler, L.T., Cordova, B., Garber, S., Reid, C., Wright, M.R., Chang, C.H. & Erickson-Viitanen, S. Deisgn and selection of DMP 850 and DMP 851: the next generation of cyclic urea HIV protease inhibitors. Chem Biol. 5:597-608. 1998.

14 Hodge, C.N., Aldrich, P.E., Bacheler, L.T., Chang, C.H., Eyermann, C.J., Garber, S., Grubb, M., Jackson, D.A., Jadhav, P.K., Kornat, B., Lam, P.Y.S., Maurin, M.B., Meek, J.L., Otto, M.J. & Rayner, M.M. et al. Improved cyclic urea inhibitors of the HIV-1 protease: synthesis, potency, resistance profile, human pharmacokinetics and X-ray crystal structure of DMP 450. Chem. Biol. 3:301-314. 1996.

15 David, L., Luo R. and Gilson M.K. Ligand-receptor docking with the Mining Minima optimizer. J. Coomputer-Aided Molecular Design. 15:157-171. 2001.

16 Davis M.E., McCammon J. A. Solving the finite difference linearized Poisson-Boltzmann equation: a comparison of relaxation and conjugate gradient methods. J. Comp. Chem. 10(3):386-391. 1989

17 Jenwitheesuk E., Samudrala R. Improved prediction of HIV-1 protease-inhibitor binding energies by molecular dynamics simulations. BMC Structural Biology. 3:2. 2003.

18 Maignan S., Guilloteau J., Pouzieux S., et. al. Crystal structures of human factor Xa complexed with potent inhibitors. J. Med. Chem. 43:3226-3232. 2000.

19 Maignan S., Guilloteau J., et al. Molecular structures of human factor Xa complexed with ketopiperazine inhibitors: preference for a neutral group in the S1 pocket. J. Med. Chem. 46:685-690. 2003.

20 Morris G. M., Goodsell D. S., Halliday R. S., Huey R., Hart W. E., Belew R. K., Olson A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J. Comp Chem. 19 (14):1639-1662. 1998.

21. Accelrys.

22. Andrew Miranker and Martin Karplus. Functionality maps of binding sites: A multiple copy simultaneous search method. Proteins: Structure, Function and Genetics. 1991. 11: 29-34.

23. Amedeo Caflisch, Andrew Mirankar and Martin Karplus. Multiple copy simultaneous search and construction of ligands in binding sites: Application to inhibitors of HIV-1 aspartic proteinase. J. Med. Chem. 1993. 36: 2142-2167.

24. B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M.Karplus. CHARMm: A program for macromolecular energy minimization and dynamics calculations. J. Comput. Chem. 1983, 4, 187-217.

25. J.H. Bohm. The computer program LUDI: A new method for de novo design of enzyme inhibitors. J. Comp. Aid. Mol. Des.1992. 6: 61-78.

26. J.B. Moon, J.W. Howe. Computer design of bioactive molecules: A method for receptor-based de novo ligand design. Proteins:Struct., Funct., and Genet. 1991. 11:314-328.

27. Michael K. Gilson and Barry Honig. Calculation of the total electrostatic energy of a macromolecular system: Solvation energies, binding energies and conformational analysis. Proteins: Struct. Func. Gen. 1988, 4:7-18.

28. Erik Kangas and Bruce Tidor. Electrostatic complementarity at ligand binding site: application to chorismate mutase. J.Phys. Chem. B 2001, 105, 880-888.

29. Traian Sulea and Enrico O. Purisima Optimizing ligand charges for maximum binding affinity. A solvated interaction energy approach. J. Phys. Chem. B, 2001, 105, 889-899.

30. Ajay Mondal and Donald Hilbert. Charge optimization increases the potency and selectivity of a chorismate mutase inhibitor. JACS. 2003. 125 (19). 5598-5599.

31. Moses Prabu-Jeyabalan, Ellen Nalivaika and Celia A. Schiffer. How does a symmetric dimer recognize an asymmetric substrate? A substrate complex of HIV-1 protease. JMB. 2000, 301, 1207-1220.

32. Stoffler, D., Sanner, M.F., Morris, G.M., Olson, A.J., and Goodsell, D.S. Evolutionary analysis of HIV-1 protease inhibitors: Methods for design of inhibitors that evade resistance. Proteins. 2002. 48:63-74.

33. Nancy M King, Moses Prabu-Jeyabalan, Ellen A Nalivaika, Celia A. Schiffer. Combating susceptibility to drug-resistance: Lessons from HIV-1 protease. (Manuscript submitted)

34. Michael K. Gilson, Hillary S.R. Gilson and Michael J. Potter. Fast assignment of accurate partial atomic charges: An electronegativity equalization method that accounts for alternate resonance forms. J. Chem Inf. Comput. Sci. 2003, 43, 1982-1997.

35. Malcom E. Davis, Jeffry D. Madura, Brock A. Luty, J. Andrew McCammon. Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics Program. Comput. Phys. Commun. 1991, 62:187-197.