Prediction of Ligand−Receptor Binding Thermodynamics by Free Energy Force Field (FEFF) 3D-QSAR Analysis: Application to a Set of Peptidometic Renin Inhibitors

J. S. Tokarski[‡] and A. J. Hopfinger*

Laboratory of Molecular Modeling and Design (M/C-781), University of Illinois at Chicago,
College of Pharmacy, 833 South Wood Street, Chicago, Illinois 60612-7231

A methodology is presented and applied in which the accurate estimation of ligand−receptor binding thermodynamics is achieved by formulating the calculation as a QSAR problem. When the receptor geometry is known, the free energy force field (FEFF) ligand−receptor binding energy terms can be calculated and used as independent variables in constructing FEFF 3D-QSARs. The FEFF 3D-QSAR analysis of a series of transition state inhibitors of renin was carried out. From a statistical analysis of the free energy contributions to the binding process, FEFF 3D-QSARs were constructed that reveal the change in solvation free energy upon binding and the intramolecular vacuum internal energy of the ligand in the unbound state are the most significant FEFF terms in determining the binding free energy, ΔG. Other terms, such as ligand stretching, bending, and torsion energy changes, the intermolecular van der Waals interaction energy, and change in ligand conformational entropy upon binding, are also found to make significant contributions in some FEFF 3D-QSAR ΔG models and in ΔH and ΔS binding models. Overall, a relatively small number of the thermodynamic contributions to the ligand−receptor binding process dominates the thermodynamics of binding in a given model.

INTRODUCTION

A current major goal of computational chemistry is the reliable prediction of the binding thermodynamics of a ligand to a receptor of known geometry in a (aqueous) medium. There is an urgency to achieve this goal in order to fully exploit the information being realized from the explosive growth in the availability of receptor geometries and in an increasing number of cases, ligand−receptor complex geometries, from combined recombinant techniques, spectroscopic methods, and homology modeling. The geometry of the receptor provides the fundamental information needed to define necessary, but not necessarily sufficient, requirements for ligand binding.

Quantitative structure−activity relationship (QSAR) analysis has almost exclusively been applied to structure−activity datasets where the molecular geometry of a common receptor is unknown. If the receptor geometry is known, intermolecular docking is usually performed to the exclusion of a QSAR analysis. Ligand−receptor binding modeling and QSAR analysis have, to an appreciable extent, been viewed as "noncompatible". Still, there is no reason physicochemical properties computed from ligand−receptor interaction modeling cannot be used in a 3D-QSAR. This was recognized more than 15 years ago and used to develop 3D-QSARs for anticancer anthracyclines intercalating into DNA.[1] A quantitative correlation of biological activity with calculated intermolecular binding energy and log *P*, the water−octanol partition coefficient of the ligand, was reported. Moreover, the limited capacity to both accurately and completely compute ligand−receptor binding thermodynamics has prompted an increase in the use of QSAR techniques in intermolecular modeling. Hence, a formal distinction is made here between a *receptor-independent three-dimensional (3D)-QSAR* (RI 3D-QSAR), where no receptor geometry is available, and a *receptor-dependent 3D-QSAR* (RD 3D-QSAR), in which the receptor geometry is used in computing potential QSAR independent variables. As this work was being completed another approach to RD 3D-QSAR analysis was reported. Ortiz et al.[2] constructed RD 3D-QSARs for a set of 26 inhibitors of human synovial fluid phospholipase A_2 (HSF-PLA$_2$) using a method they term comparative binding energy (COMBINE) analysis, which, as the name hints, is something of a receptor-dependent analog method to CoFMA. For each ligand in the training set intermolecular and intramolecular energies are calculated for the ligand−receptor complex, the unbound ligand and receptor. The ligands are divided into fragments for the statistical analysis of the energy terms. The receptor is divided into regions, e.g., amino acid residues. The molecular mechanics force field energy terms are divided into the intermolecular interaction energies between each fragment of the ligand and each region of the receptor, changes in the bonded (bond, angle, and torsion) and nonbonded (Leonard−Jones and electrostatic) energies of the ligand fragments upon binding to the receptor, and the bonded and nonbonded energies of the receptor regions upon binding of the ligand. All of the energy terms resulting from the partitioning process are used as descriptors in constructing a RD 3D-QSAR.

This paper reports a RD 3D-QSAR formalism called free energy force field (FEFF) 3D-QSAR analysis in which (1) all of the enthalpy and entropy contributions to the ligand−receptor interaction in a solvent medium are taken into consideration, (2) the enthalpy and entropy contributions are treated as independent variables in developing the QSAR for ligand−receptor binding, and (3) the optimum FEFF 3D-QSAR is constructed using a genetic algorithm. FEFF 3D-QSAR analysis is the intermolecular extension of molecular

---

* To whom correspondence should be addressed.
‡ Current address: The Chem21 Group, Inc., 1780 Wilson Drive, Lake Forest, IL 60045.
⊗ Abstract published in *Advance ACS Abstracts,* June 15, 1997.

shape analysis (MSA).[3] The FEFF 3D-QSAR method is applied to a set of peptidometic renin inhibitors.

METHODS

A. The FEFF Formalism. The ligand−receptor interaction can be expressed as

$$L \cdot M + \underset{\text{(unbound state)}}{R \cdot M} \overset{K}{\rightleftharpoons} \underset{\text{(bound state)}}{(LR) \cdot M} \quad (1)$$
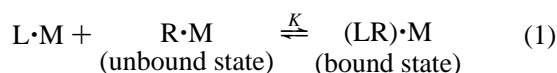
where L is the ligand, R is the receptor, M is the solvent medium, and K is the equilibrium, or binding, constant. The binding free energy, ΔG, of a ligand, L, to a receptor, R, in a solvent medium, M, is equal to the difference in free energy of the bound state and unbound states and can be expressed as

$$\Delta G = G_{LR} - (G_L + G_R) = -kT \ln K \quad (2)$$

where G_{LR} is the free energy of the bound, or complex state, G_L is the free energy of the unbound ligand, G_R is the free energy of the unbound receptor, k is the gas constant, and T is the temperature of the system. The free energy of a LRM cannot rigorously be broken into a set of component terms. However, this approximation is often reasonable and has been adopted as part of the empirical nature of the FEFF 3D-QSAR formalism.

The free energy of the ligand−receptor complex, G_{LR}, can, under the decomposition assumption, be broken into its component terms

$$G_{LR} = [G_{LR}(LL) + G_{LR}(RR) + G_{LR}(MM) + G_{LR}(LR) + G_{LR}(LM) + G_{LR}(RM)] \quad (3)$$

where $G_{LR}(XY)$ refers to the interaction between X and Y for the LR (bound, or complex) state. Each of the terms in eq 3 can be further partitioned into enthalpy and entropy contributions. Since the work term, $P\Delta V$, of binding is very small for small organic ligands binding to macromolecules at low combined solute (ligand and macromolecule) concentrations, the enthalpy terms, $H_{LR}(XY)$, can be represented by their internal energy, $E_{LR}(XY)$, contributions and we have

$$H_{LR} = E_{LR} = [E_{LR}(LL) + E_{LR}(RR) + E_{LR}(MM) + E_{LR}(LR) + E_{LR}(LM) + E_{LR}(RM)] \quad (4)$$

and for the entropy contributions, $S_{LR}(XY)$

$$S_{LR} = [S_{LR}(LL) + S_{LR}(RR) + S_{LR}(MM) + S_{LR}(LR) + S_{LR}(LM) + S_{LR}(RM)] \quad (5)$$

where, in general

$$G_{LR} = H_{LR} - TS_{LR} \quad (6)$$

Equation 5 assumes that the overall entropy change associated with ligand−receptor binding can be partitioned into specific pairwise components. In doing this the implicit assumption is made that the composite set of degrees of freedom of the system can be divided into independent sets or that certain couplings between degrees of freedom have small entropic contributions to binding free energy.

The corresponding unbound ligand, G_L, and receptor, G_R, free energies can be expressed in terms of their component

XY interactions

$$G_L = [G_L(LL) + G_L(LM) + G_L(MM)] \quad (7)$$

$$G_R = [G_R(RR) + G_R(RM) + G_R(MM)] \quad (8)$$

Each of the XY interaction free energy contributions in both eqs 7 and 8 can be broken down into their respective enthalpy and entropy contributions as done for G_{LR}. Also, as done for G_{LR}, the enthalpy contributions $H_L(XY)$ and H_R-(XY) can be well-approximated by $E_L(XY)$ and $E_R(XY)$, respectively, for tight-binding at low L and R concentrations.

Overall, we can list the internal energy and entropy changes composing ΔG in eq 2 in terms of the XY interaction contributions. These contributions, and their representations, are given in Table 1 and provide a useful reference base to use in dissecting what aspects of ligand−receptor binding are considered in a particular 3D-QSAR approach. The low concentration condition that is prefaced in Table 1 is made to avoid possible aggregation and/or multiple ligand binding (coating) to the receptor.

An analysis of current 3D-QSAR approaches indicates, perhaps not too surprisingly, that the intermolecular ligand−receptor binding interaction is of primary focus. Solvent reorganization receives virtually no attention in any 3D-QSAR method. Entropy and the unbound states of L and R are also not considered in most 3D-QSARs. Still, emerging experimental studies of ligand−receptor binding suggest that each of the six classes of interactions in Table 1 can play a major role in the binding process depending on the particular LR(M) system.[4]

Nevertheless, when the receptor geometry is available, one can theoretically evaluate all of the thermodynamic properties listed in Table 1 for the ligand−receptor complex and both the unbound ligand and receptor. Energy contributions, such as the change in intramolecular energy of the receptor and ligand upon binding, can, therefore, be evaluated when the receptor structure is known. These thermodynamic properties can be employed as independent variables to construct RD 3D-QSARs. That is the force field representation for the *complete* ligand−receptor binding process, that is a *free energy force field*, FEFF, is retooled to include a variety of scaling coefficients determined using QSAR techniques. Once the scaling coefficients are determined, the corresponding FEFF is, presumably, an accurate and reliable vehicle to predict the binding properties of hypothetical ligand analogs relative to those in the training set. Moreover, the scaled terms in the calibrated FEFF for one ligand−receptor system might be transferable to other ligand−receptor systems or at least provide a base-line as to the relative magnitudes of the various energy terms composing a trial force field in, for example, a free energy perturbation study.

The FEFF approach requires explicit binding thermodynamics information. Unfortunately, only a very small number of ligand−receptor systems are available for which binding free energies of the ligands have been measured. Nevertheless, the requirement of measured binding free energies, taken at face value, would restrict the use of a FEFF 3D-QSAR analysis to such an extent that the method would not be of practical utility. However, binding constants and/or *in vitro* inhibition constants are available for many enzyme−inhibitor systems. Binding free energies can often

Table 1. A Breakdown of the Interaction Terms, XY, for a L + R ⇌ LR Binding Process in a Solvent M at Low L and R Concentrations with the Concentration of Bound L Less Than R

chemical unit, Z	type of interaction energy, XY	change in internal energy, symbols	change in entropy, symbols
ligand L	intramolecular ligand conformational energy LL	$\Delta E_L(LL) = E_{LR}(LL) - E_L(LL)$	$\Delta S_L(LL) = S_{LR}(LL) - S_L(LL)$
ligand L	ligand solvation energy LM	$\Delta E_L(LM) = E_{LR}(LM) - E_L(LM)$	$\Delta S_L(LM) = S_{LR}(LM) - S_L(LM)$
solvent medium M	solvent reorganization energy MM	$\Delta E_M(MM) = E_{LR}(MM) - [E_L(MM) + E_R(MM)]$	$\Delta S_M(MM) = S_{LR}(MM) - [S_L(MM) + S_R(MM)]$
receptor R	intramolecular receptor conformational energy RR	$\Delta E_R(RR) = E_{LR}(RR) - E_R(RR)$	$\Delta S_R(RR) = S_{LR}(RR) - S_R(RR)$
receptor R	receptor solvation energy RM	$\Delta E_R(RM) = E_{LR}(RM) - E_R(RM)$	$\Delta S_R(RM) = S_{LR}(RM) - S_R(RM)$
ligand−receptor RL	intermolecular ligand−receptor energy LR	$\Delta E_{LR}(LR) = E_{LR}(LR)$	$\Delta S_{LR}(LR) = S_{LR}(LR)$

Table 2. Estimation of the Binding Free Energy, ΔG, and Other Thermodynamic Properties for Ligand−Receptor Systems

(1) If only the K (binding constants) are given
*Assume $P\Delta V = 0$, so that $\Delta G = \Delta A = A(B) - A(U)$, where A is the Helmholz free energy
*Assume $\Delta G = -RT \ln K$, or $-\ln K = (RT)^{-1}[G(B) - G(U)]$

$$\Delta G = -RT \ln K$$

(2) If only C_i (ligand concentration for a fixed response i) are given
*Assume same as in (1) above, and
*Assume $-\log(C_i) = a \ln K + b$, or
$$-\log(C_i) = aRT^{-1}[G(B) - G(U)] + b$$
$$= a'[G(B) - G(U)] + b$$

$$\Delta G = -(a')^{-1}[\log(C_i) + b]$$

(3) If ΔH (and ΔG plus ΔS) are given in addition to K
*Assume $\Delta H = \Delta E$ and $\Delta G = \Delta A$, that is $P\Delta V = 0$
*Assume that the force field terms can be partitioned into separable ΔE and ΔS contributions

$$\Delta E = E(B) - E(U) \text{ and}$$
$$\Delta S = T^{-1}[E(B) - G(B) - E(U) + G(U)]$$

be estimated on a relative basis from binding and inhibition constants. Consequently, *in vitro* measures of biological activity, such as IC$_{50}$'s, must be taken to reflect relative ligand binding strength (thermodynamics) and re-expressed on an energy scale if these measures are to be used as the dependent variable set in a FEFF 3D-QSAR study.

A working set of equations to compute ΔG's is given in Table 2. The estimation of ΔG in Table 2 can be extended to include other thermodynamic relationships between ΔG and binding or inhibition constants. Table 2 also outlines the extended application if, in fact, experimental ΔH, ΔG, and/or ΔS measures are available in the ligand−receptor data set.

The contributions in Table 1 represent one way in which the free energy of binding can be partitioned. Each of the constituent terms may serve as independent variables in the 3D-QSAR analysis. The free energy of binding, ΔG, could also be represented by an equivalent free energy force field expression

$$\Delta G = \alpha_1 \Delta E_{stretch} + \alpha_2 \Delta E_{bend} + \alpha_3 \Delta E_{torsion} +$$
$$\alpha_4 \Delta E_{vdW} + \alpha_5 \Delta E_{electrostatic} + \alpha_6 \Delta E_{hydrogen\ bonding} +$$
$$\alpha_7 \Delta E_{solv} + \alpha_8 T\Delta S \quad (9)$$

where $\Delta E_{stretch}$ is the change in bond stretching, ΔE_{bend} is the change in bond angle bending, $\Delta E_{torsion}$ is the change in dihedral torsional energy, ΔE_{vdW} is the change in van der Waals interaction energy, $\Delta E_{electrostatic}$ is the change in the electrostatics interaction energy, $\Delta E_{hydrogen\ bonding}$ is the change in hydrogen bonding energy, ΔE_{solv} is the change in solvation energy, and $T\Delta S$ is the change in entropy in the LR(M) system upon binding. The energy changes upon binding may

be summarized as

$$\Delta E_X = E_{LR,X} - E_{L,X} - E_{R,X} \quad (10)$$

where X represents the type of energy contribution. Equation 9 gives the partitioning of the free energy of binding with respect to the change in each of the individual free energy force field energy terms. This representation is in contrast to the terms in Table 1 which are separated according to changes in energy for each participant of the LR(M) system.

If the free energy force field is precise, then all of the coefficents, α_i, of the energy terms in eq 9 are unity. However, since no force field is exact, the energy terms may be analyzed as descriptors using a QSAR paradigm in order to scale them appropriately, in terms of the α_i, for the training set. Thus, the binding free energy contributions, such as those defined in Table 1 and eq 9, for a LR(M) system need to be determined in a FEFF 3D-QSAR analysis.

B. The Renin-Inhibitor Training Set. The FEFF 3D-QSAR analysis approach is applied in this study to the ligand−receptor system, renin and a series of inhibitory peptides. Renin, an aspartic acid protease, belongs to a widely distributed family of enzymes that play an important role in fungi, plants, vertebrates, and retroviruses. This class of homologous enzymes includes pepsin, chymosin, cathepsins D and E, related microbial enzymes such as endothiapepsin and penicillopepsin, and the human immunodeficiency virus (HIV) protease. There exists high sequence homology among these aspartic proteases, especially among the residues which comprise the structural core of the active site. This family of aspartic proteases is named as such, because they all possess two aspartyl residues at their active site which are involved in the mechanism of peptide cleavage. The binding cleft and active site are at the junction of two structurally similar domains (in HIV protease they are identical). The active-site aspartic acid residues are located centrally in the binding cleft with the carboxyl side chains and surrounding main chain scaffolding related by an approximate interdomain 2-fold axis. A network of hydrogen bonding among the highly conserved residues near the active site keeps these two aspartic acid carboxylates essentially coplanar. The aspartic proteases also share in common a flap region which can open and close upon substrate and inhibitors.

The X-ray crystal structure of recombinant human renin was first determined by Sialecki and co-workers at 2.5 Å resolution.[5] The crystal of recombinant human renin in complex with a transition state analog inhibitor was determined by Rahuel et al. to a resolution of 2.4 Å.[6] Dhanaraj and co-workers have also crystallized human renin with a bound peptide inhibitor and defined the structure of the complex at 2.8 Å resolution.[7]

Table 3. Chemical Structures of the Renin Inhibitory Peptides Used in the FEFF 3D-QSAR Analysis
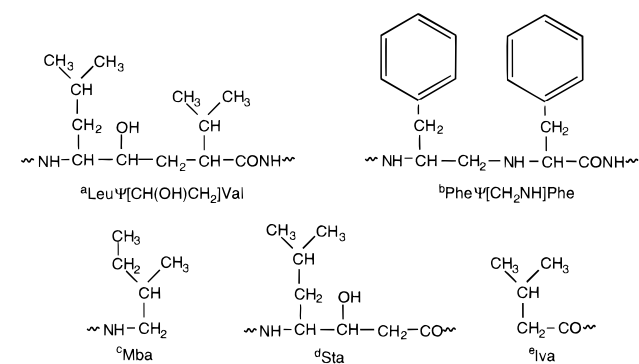
| compd	peptide
U80631E	Ac-Phe-His-Leu-Ψ[CH(OH)CH₂]Val-Ile-NH₂[a]
U77646E	Ac-Pro-Phe-His-Leu-Ψ[CH(OH)CH₂]Val-Ile-NH₂
U77647E	Ac-D-Pro-Phe-His-Leu-Ψ[CH(OH)CH₂]Val-Ile-NH₂
U73777E	Ac-Phe-His-Phe-Ψ[CH₂NH]Phe-NH₂[b]
U71909E	Ac-Pro-Phe-His-Phe-Ψ[CH₂NH]Phe-NH₂
U77451E	Ac-Pro-Phe-His-Phe-Ψ[CH₂NH]Phe-Mba[c]
U72407E	Ac-Phe-His-Sta-Ile-NH₂[d]
U72408E	Ac-Pro-Phe-His-Sta-Ile-NH₂
U72409E	Ac-His-Pro-Phe-His-Sta-Ile-NH₂
U77455E	Iva-His-Pro-Phe-His-Sta-Ile-Phe-NH₂[e]

Table 4. Thermodynamic and Binding Parameters[a] of the Renin Inhibitory Peptides[b] Used in the FEFF 3D-QSAR Analysis

compd	K_d, μm	$-\Delta H°$, kcal/mol	$-\Delta S°$, entropy units	$\Delta G°$, kcal/mol
		[Leu-Val-OH]		
U80631E	0.37 ± 0.06	14.28 ± 0.7	75.7 ± 2.7	9.2 ± 0.06
U77646E	0.0054 ± 0.0013	28.75 ± 0.6	131.1 ± 2.0	11.5 ± 0.1
U77647E	0.0013 ± 0.0002	20.33 ± 0.5	105.5 ± 1.6	12.4 ± 0.2
		PheΨ[CH₂NH]Phe		
U73777E	0.22 ± 0.01	14.20 ± 0.3	76.3 ± 0.8	9.4 ± 0.1
U71909E	0.029 ± 0.22	13.70 ± 0.5	78.4 ± 2.2	10.6 ± 0.1
U77451E	0.0025 ± 0.0015	26.70 ± 0.4	125.3 ± 1.4	12.2 ± 0.1
		[Statine]		
U72407E	0.204 ± 0.0012	26.10 ± 0.8	114.8 ± 2.7	9.5 ± 0.1
U72408E	0.098 ± 0.013	14.69 ± 0.9	79.6 ± 3.2	9.9 ± 0.1
U72409E	0.023 ± 0.0001	22.63 ± 0.5	108.0 ± 1.8	10.8 ± 0.2
U77455E	0.0017 ± 0.0001	21.36 ± 0.2	108.9 ± 0.7	12.4 ± 0.04

[a] Thermodynamic and binding parameters measured by Epps et al.[8] [b] The renin inhibitory peptides are grouped by class based on their putative transition-state isosteres at the P_1–P_1' positions, namely, Leu-Val alcohol compounds, reduced Phe-Phe compounds, and statine compounds.

Experimental thermodynamic binding properties, namely, ΔH, ΔS, ΔG, and K_d, were reported for a series of 13 renin inhibitory peptides which were made and tested at Upjohn.[8] Several types of inhibitors were synthesized by Epps and co-workers, by reason of the similarity of the structures to that of angiotensinogen. The structures of the ten inhibitors included in our modeling study are given in Table 3. The thermodynamic binding data of these ten inhibitors are given in Table 4. These inhibitors are transition state analogs because they contain a substructure at the position corresponding to the angiotensinogen cleavage site that mimics the transition state of the natural substrate. The analogs that were modeled in this study include a "hydroxyethylene", "reduced amide", and a statine transition-state isostere. The reduced amide isostere analogs contain a secondary amine at the isosteric replacement (see Table 3). Statine is a rare,
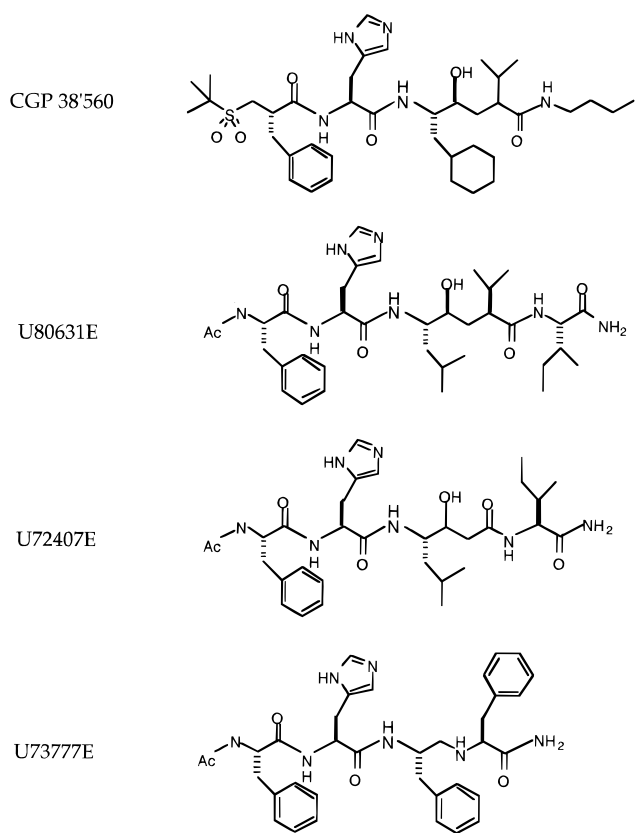
Figure 1. The chemical structures of CGP 38'560 and a representative renin inhibitor of each class of transition state isostere.

natural amino acid found in pepstatin, an inhibitor of pepsin.

An attempt to predict the binding affinities of these analogs based only on the ΔH values would prove to be unsuccessful. Compound U72407E (see Table 4) has a much more favorable ΔH of binding than U72408E. Yet, both compounds have about the same ΔG and K_d measures. The binding entropy, ΔS, of U72407E is much larger than that of U72408E which negates the more favorable ΔH of binding of U72407E. Thus, this renin-inhibitory peptide data set provides a challenge to applying a RD 3D-QSAR analysis.

1. Thermodynamic Properties of Binding. The dissociation constants, K_d's, of the renin inhibitors were determined over a temperature range of 8−37 °C by the flourescence displacement assay.[8] Duplicate assays at two different concentrations of inhibitors were performed, and van't Hoff plots of the temperature dependence of the K_d's were constructed from the assay data. The values for ΔH and ΔS were calculated from the slope and intercept of the plots, and the results are given in Table 4 together with the K_d's at 37 °C. ΔG_{37} values were calculated from $\Delta G = \Delta H - T\Delta S$ using the appropriate data. Ten out of the 13 inhibitors in the Upjohn study were modeled based on the structural diversity and range of activity of the ligands (see Tables 3 and 4).

2. Building the Molecules. Rahuel et al.[6] cocrystallized a hydroxyethylene isostere transition state analog, CGP 38'560 (see Figure 1), bound to human recombinant renin at 2.4 Å resolution. The coordinates of the complex have been deposited with the Brookhaven Protein Data Bank[9] under the pdb entry 1RNE. This particular structure of renin was selected as a starting model for the receptor geometry in this analysis for two reasons. Firstly, the geometry of the enzyme with a bound ligand is a more appropriate target

for docking the Upjohn inhibitors than attempting to start with the crystal structure of the unbound enzyme. Secondly, the cocrystallized ligand in 1RNE has similar structural features to the Upjohn inhibitors, a property that assists in modeling the ligand−receptor complex.

Most parts of the protein structure are well-defined by the electron density pattern. Individual missing amino acid residues were added to the crystal structure of renin where necessary with amino acid fragments from the library of structures in the Chemlab-II molecular modeling program.[10] The scheme of completing the enzyme structure is discussed in an earlier work.[11]

The appropriate number of hydrogens were "grown" on all atoms of the PDB structure. An option in the molecular modeling package QUANTA[12] was used to assign protons to the crystal structure. In order to check for bad steric contacts of this edited structure, 2000 steps of molecular dynamics (MD) simulation using the program MOLSIM[13] were performed. Necessary side chain rotations were performed to relieve the bad contacts.[11]

AMBER partial charges[14] were assigned to all atoms of the enzyme structure. All ionizable residues were assigned the charge state which is normally present at the pH of the binding assay experimental conditions (7.4) at which the binding thermodynamics were also measured.[15] One exception to this assignment was the two active-site aspartic acid residues which were treated separately. The assignment of the protonation states of these residues is considered in a following section. The N-termimus and C-terminus were both modeled as ionized. The total charge of the enzyme, if the two active-site aspartic acid groups are considered ionized, is −8.0. Lone pair electrons were not modeled explicitly. Water molecules located in the crystal structure of the renin complex were not explicitly included in the modeling.

The structures of the peptide inhibitors (see Table 3) were also built using standard fragments of the Chemlab-II library and standard bond lengths and angles. The transition state analogs have an additional chirality due to the presence of the isostere replacement fragments. These chiral compounds were modeled as the S-isomer analogs based upon discussions with Epps of the Upjohn company.[16] Wherever possible the assignment of AMBER partial atomic charges of the isostere replacements of the analogs were based on similar fragments found among the standard amino acid structures. Semiempirical AM1[17] calculations were performed on the isostere replacements to substantiate the choice of partial charges and/or to establish missing charge assignments. AMBER partial atomic charges on the N-terminal and C-terminal atoms of the inhibitors were adopted from available similar amino acid fragments and scaled accordingly to produce the overall desired charge of neutralilty, or +1.0, where appropriate.

3. Docking the Ligands. Ligands of aspartic proteases have generally been found to bind in an extended β-strand conformation.[18] A starting bound conformation and orientation, i.e., intermolecular alignment, for the modeled Upjohn inhibitors in the active site of renin was established using experimental information. Conformation and alignment information was based upon the bound structure of inhibitor CGP 38'560, which was cocrystallized with recombinant human renin. The structure of CGP 38'560, a hydroxyethylene isostere inhibitor, contains a number of identical amino
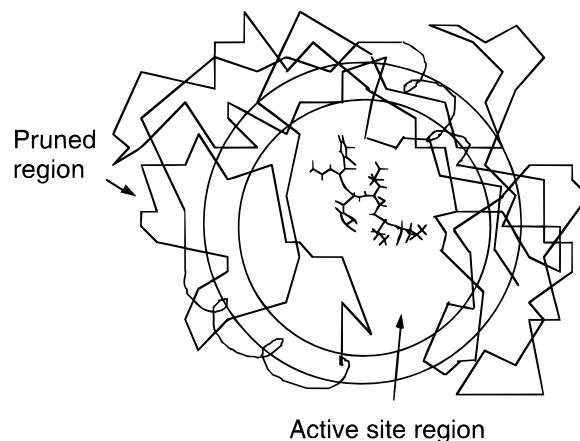
Figure 2. A C_α-trace representation of the receptor. A ligand is docked in the active site. A scaled-down, or pruned, receptor model includes all amino acid residues within a defined cutoff from any atom of the ligand. Two possible cutoffs are shown.

acid residues to the Upjohn inhibitors. Figure 1 shows the structures of CGP 38'560 and a representative Upjohn inhibitor from each class of transition state isostere studied in this work. Thus, the atomic coordinates, and ϕ, ψ, and χ torsion angle values, from the crystal structure of CGP 38'560 were partially used in the initial docking of the Upjohn inhibitors.

Additional information was extracted from the structures of similar ligands cocrystallized with various homologous aspartic proteases. The crystal structure of a hydroxyethylene isostere inhibitor bound to mouse renin[7] and several statine isostere inhibitors cocrystallized with the fungal aspartic protease, endothiapepsin,[18] provided supplementary conformational data which assisted in docking the Upjohn analogs in the active site of the crystal structure of renin. In only a few situations was there a need to relieve bad contacts between ligand and receptor in these initial dockings.

4. Determination of the Effective Size of a Receptor Model. The entire model of the renin structure, including protons, consists of 5177 atoms. This large number of atoms makes multiple full-scale ligand−receptor MD simulations computationally uneconomical. Thus, modeling approximations were sought to scale down the renin structure to a more manageable size. The analysis was restricted to those amino acid residues of the enzyme near the active-site region. Numerous crystallographic studies of aspartic protease-inhibitor complexes reveal that although binding of the inhibitors is accompanied by significant motion of the flaps of the protease, the active-site structure of the protease in its complexes with inhibitors remains almost unchanged.[19] This observed structural feature of the protease justifies the use of a receptor model centered at the active site.

A receptor model was created by pruning away all residues greater than 10 Å away from any atom of the bound ligand. If any one non-hydrogen atom of a residue was within the spherical cutoff, then that entire residue was included in the model. Figure 2 shows a conceptual representation of this receptor model. The pruning operation resulted in a receptor model that was comprised of a number of unconnected peptide fragments. The number of fragments in the receptor model was kept to a minimum by not "cutting" those fragments separated by less than or equal to four residues in the original renin sequence. The largest inhibitor from each class of transition-state isostere was docked in the active site

in order to create the largest size receptor model necessary. The single resulting receptor model was adopted in all production run simulations of the complex and unbound receptor. This size of receptor model was found to be a good compromise between computational efficiency and accuracy.[11]

A fictitious mass of 5000, assigned to all main chain atoms of the entire receptor model, regardless of their distance from the active site, was found necessary to maintain the integrity of the model. These constraints, which serve as "momentum reservoirs", were used for all simulations. The use of fictitious masses is virtually the same as using Cartesian constraints, particularly when the masses are chosen to be very large. Fictitious masses provides a convenient way to "tune in" MD simulation motions in the pruned protein model that are similar to those of the complete parent protein by selectively varying the mass values. In addition, the initial MD evaluation studies of the complex indicated that trajectories longer than 20 ps did not improve the convergence of the total potential energy, but in some cases increased the divergence of the model from the crystal structure. The choice of 10−20 ps MD simulations was found to be the best compromise in producing trajectory geometries and energies that remained close to the X-ray structure of renin for reasonable amounts of CPU time.

5. Determination of the Protonation State of the Active-Site Aspartic Acid Residues. A review of the experimental and theoretical work on aspartic proteases in the native state or in complexes clearly shows that there is not a consensus choice for the protonation states of the active-site aspartates.[11] Examination of X-ray crystal data reveals that the two active-site aspartates are most likely involved in important nonbonded interactions with bound ligands. If the correct protonation state is not assigned for these important recognition-site residues, then the computed thermodynamic binding energies may not be reasonable. Thus, a molecular modeling analysis was performed to determine which protonation state yields reliable calculated binding energies.[11] Different mono- and diprotonation states of the two active-site aspartic acid residues were examined for the enzyme alone and in complex with a representative of each of the three types of transition state analogs from the Upjohn data set. This analysis included the state in which one of the two aspartates was negatively charged, and the other was neutral. The scenarios where both aspartates were ionized and both were neutral were also explored. The protonation states of some of the inhibitors was also explored because the protonation state of the secondary nitrogen of the reduced amide transition-state isostere is uncertain. Therefore, we modeled the secondary amine nitrogen in both the neutral and ionized states in conjunction with the various protonation models of the active-site aspartates.

The structural and thermodynamic results accumulated over the course of test MD simulations of the various protonation state models, coupled to an analysis of experimental findings, led to the selection of the most plausible protonation states.[11] The protonation state consisting of a monoprotonated active-site model where the lone proton is situated at the outer carboxylate oxygen of Asp226 was chosen to be a suitable representative of the bound receptor for three reasons; (a) the intermolecular interaction energy is the most favorable for this model, (b) a hydrogen bonding network is maintained over the course of the MD simulations

that nearly reproduces the interatomic distances present between hydrogen bonding partners found in a number of aspartic protease crystal complexes, and (c) an analysis of the available renin-ligand X-ray crystal structures indicates that each of the active-site aspartate oxygens has an intramolecular hydrogen bonding partner, except for the Asp226 outer oxygen. This observation reinforces the decision to model this oxygen as neutral. Coincidentally, this protonation state has been proposed by experimentalists[6,20,21] and from results of theoretical calculations.[22] Figure 8a shows a representation of the two active-site aspartates in the presence of inhibitor U77646E for the selected protonation state. Interatomic distances between the active-site aspartatyl carboxylate oxygens and hydroxyl oxygen of isostere of U77646E are given in ref 11.

An examination of various crystal aspartic protease complexes consisting of ligands with the reduced amide isostere revealed that the secondary nitrogen of the inhibitor and outer carboxylate oxygen of Asp226 are very likely hydrogen bonding partners. The protonation state model consisting of a positively-charged inhibitor and negatively-charged active-site aspartates reproduced this close ligand−receptor contact and yielded one of the lowest energy complexes in the modeling study for this class of inhibitors.[11] This model also has one of the most favorable interaction energies. In addition, a feasible hydrogen bonding network is maintained over the course of the MD simulations that nearly replicates the average ligand−receptor interatomic distances observed in an analysis of available X-ray crystal renin/reduced amide isostere type inhibitor complexes. Therefore, this protonation state model was chosen for the reduced amide isostere analogs. This protonation state is in agreement with that proposed by experimentalists[20,21] and from results of a molecular modeling study[23] for this class of inhibitors.

Overall, the renin receptor model is assumed to possess different protonation states for the active-site aspartates depending on the bound ligand. However, the overall total charge of each of the complexes is identical which is important when making comparisons of calculated intermolecular interaction energies of different ligand−receptor systems.

A Coulomb potential is used to describe charge−charge interactions in our force field. A constant molecular dielectric of 3.5 is used to represent the medium of the interactions. This choice of molecular dielectric is likely realistic for short-range interactions where waters can not "fit between" the interacting pair. Longer-range interactions are probably overestimated using a molecular dielectric of 3.5 since waters, and/or solute groups, are likely to be located between the interacting charge-pair. However, the inverse distance dependence diminishes the overall contributions of long-range charge interactions and tends to minimize problems in dielectric representation.

C. Free Energy Force Field Terms. The free energy force field is composed of a set of molecular mechanics based free energy terms that may be calculated by a variety of methods. A summary of the methods used in this FEFF 3D-QSAR analysis is given in the following sections.

1. The Nonscaled FEFFs of the Bound and Unbound States. The most straightforward means of calculating the nonscaled FEFF energy terms is to input an acceptable molecular geometry, select the form of the FEFF, and

compute the corresponding nonscaled potential energy. However, the input geometry will need to be refined for the specific FEFF, charge state, and/or size of the model receptor. The quality and corresponding predictiveness of a FEFF 3D-QSAR will, in part, be dependent upon the parent, nonscaled force field and the ligand, receptor, and complex molecular geometries used in the analysis. There is no foolproof way to select the parent, nonscaled force field or the "correct" molecular geometries. In principle, all the energy terms evaluated in a free energy force field should be ensemble averages as the complex, the unbound receptor, and unbound ligand are dynamic entities. Few ligand−receptor modeling studies approximate ensemble averages. In most cases a single stable structure is used to evaluate thermodynamic parameters of binding.

An energy-space sampling approach of the bound and unbound states of the ligand−receptor system was formulated as part of this study. This sampling method assumes that one can model the ligand−receptor binding process, described as

$$L \cdot M + \quad R \cdot M \quad \rightleftharpoons \quad (LR) \cdot M$$
$$\text{(unbound state)} \qquad \text{(bound state)}$$
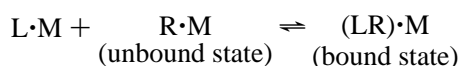
by beginning with the solvated ligand−receptor complex, $(LR) \cdot M$. This energy-space sampling approach actually models the "unbinding" process.

The geometry of the ligand−receptor complex is needed to initiate sampling. The X-ray crystal structure of the complex, or a model of the ligand docked in the active site region of a receptor of known geometry, is a suitable starting point. A MD simulation of the complex at a designated initial temperature and for a specified amount of time is performed. The resultant lowest energy structure of this simulation is identified (the lowest energy assembly is currently selected although a variety of other structures could be chosen) and then used as the starting geometry for a subsequent MD simulation at a lower temperature. The lowest energy structure from the second MD simulation is the starting point of a subsequent simulation run at a lower temperature than the previous simulation. This process of performing coupled MD simulations at progressively lower temperatures is repeated until a user-defined final temperature of evaluation is reached.

The lowest energy structure from the final MD simulation is "broken up" into its individual components, i.e., the ligand is removed unchanged from the receptor. The separated bound ligand and bound receptor structures are subjected to independent MD simulations at the lowest temperature of evaluation to begin the process of modeling these molecules in their respective unbound states. The lowest energy structures found for both the unbound ligand and unbound receptor from the simulation at lowest temperature are then used as starting structures for an MD simulation performed at the next higher temperature of evaluation. This process of "warming" up the unbound ligand and receptor is repeated until the highest temperature of evaluation is reached. The lowest energy structure found at the highest temperature for both the ligand and receptor is then subjected to an MD simulation at each of the lower temperatures of evaluation in order to allow the unbound ligand and receptor to equilibrate at each of these temperatures. A diagram showing the general sampling scheme as a function of temperature is shown in Figure 3.
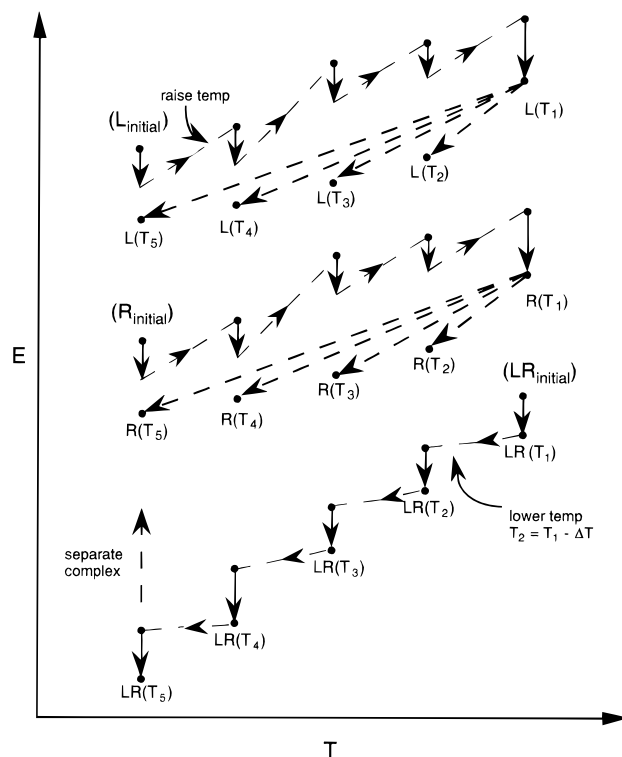
Figure 3. A representation of the MD simulation sampling scheme as a function of temperature (T). The modeling of the "unbinding" process begins with a MD simulation at T_1 on the starting geometry of the complex, $LR_{initial}$. Subsequent MD simulations are run at T_{i+1} on the lowest energy structure from the previous T_i MD simulation where $T_{i+1} = T_i − \Delta T$. Once the lowest temperature of evaluation, T_5 in this particular example, is reached, the complex, $LR(T_5)$, is separated into its components. The disassociated ligand, $L_{initial}$, and receptor, $R_{initial}$, are then "warmed" up using a reverse sampling protocol to that of the complex. Both unbound $L(T_1)$ and $R(T_1)$ are then "cooled" down at each of the lower temperatures of evaluation. The resultant energies (E) are used to calculate the FEFF terms.

This sampling approach allows the complex and both unbound ligand and receptor each to examine different conformational states as part of the geometry optimization process. This is especially useful in modeling the complex where at the higher simulated temperatures the docked ligand is allowed to find its optimal intermolecular alignment in the active site. As the simulated temperature is lowered the kinetic energy of bound ligand and receptor is also lowered which increases the probability for the interactants to form a "tighter" complex.

An advantage of this sampling scheme is that the kinetic and potential energies of each of the participants of the binding process can be balanced against one another at different temperatures. Statistical analyses of these energies will reveal the simulation temperature at which the calculated thermodynamic binding parameters best correlate with the experimental values.

In this study all energy minimization and MD calculations used a force field in which the nonbonded, electrostatic, torsional, bond stretching, and bond angle bending force field parameters of the AMBER program[14] were adopted into a MM2 force field[24] potential function representation to compute the parent, nonscaled energy terms. Missing force field parameters (torsional, bond stretching, and bond angle bending) were taken and adjusted from the set proposed by Hopfinger[25] and the MM2 force field. The "combining" of

LIGAND−RECEPTOR BINDING THERMODYNAMICS

J. Chem. Inf. Comput. Sci., Vol. 37, No. 4, 1997 **799**

AMBER and MM2 parameters is accomplished by linear scaling. The most similar set of atoms to those of the missing AMBER parameter is identified for a parameter which has both AMBER and MM2 values. The unknown AMBER parameter value is then scaled against the known MM2 value in the same ratio as the parameter having both AMBER and MM2 values. Since the ultimate goal of our work is to develop a refined custom force field for a specific ligand−receptor system, limitations in this linear scaling are considered in the force field refinement. Energy minimization and MD simulations were performed using the MOLSIM program.[13]

Selection of the AMBER force field as the main component in the nonscaled, reference potential has been done as a reasonable first choice. However, other force fields could also be used as initial, nonscaled reference potentials.

2. Solvation Energy. In addition to the contributions typically included in a molecular mechanics force field, a comprehensive FEFF must take into account solvation. Biological receptor-ligand interactions compete with water interactions due to the fact that both the receptor and the ligand are solvated before complexation and experience a change in solvation upon complex formation.

The hydration shell model proposed by Hopfinger[26a,b] was used to estimate the solvation energy terms. The solvation model was included with the other molecular mechanics force field terms to comprise the potential energy function used in the MD simulations of the sampling scheme. In energy minimizations, the solvation energetics were calculated on the structure of the complex, unbound ligand, and unbound receptor only after their vacuum geometries were energy-minimized.

The hydration shell model assumes that a characteristic sphere can be centered about each atom, or solute group, of the solute molecule. The size of the sphere, which defines the hydration shell, is dependent upon the solvent molecule and the particular solute group. Removal of a solvent molecule from the hydration shell results in a change in solvation free energy. The size of the hydration shell and the size and shape of the solvent molecule dictate how many solvent molecules can occupy the hydration shell. Intersections of the van der Waals volumes of neighboring solute atoms with a particular hydration shell results in an excluded volume which determines how many solvent molecules are removed from the hydration shell when a solute is in a particular conformation. Thus, the hydration shell is sensitive to conformation via excluded hydration shell volumes.

The hydration shell model is a four parameter system: n, the maximum number of solvent molecules which can occupy the hydration shell; Δf, the change of free energy associated with removal of one solvent molecule from the hydration shell; R_v, the effective radius of the hydration shell; and V_f, the free volume of packing associated with one solvent molecule in the hydration shell.

The absence of explicit water molecules during the simulation analysis can lead to the generation and sampling of artifact states relative to the actual binding mode. On the other hand, the inclusion of explicit waters introduces questions regarding assignments of waters and the extent of sampling needed to generate equilibrium/steady state ensembles. These unknowns can also lead to artifact states. The use of an implicit solvation model, like a hydration shell

scheme, seems to be both a good compromise, and a diagnostic, to evaluate solvation effects.

3. Entropy. The conversion of an internal energy force field to a FEFF requires the addition of an entropy measure (force field term). An approximate approach to estimating entropy contributions of the ligand, receptor, and their complex is to compute a nonscaled, parent entropy for each of these three molecular entities. Then, in the spirit of the FEFF 3D-QSAR approach, these entropies are scaled as part of the fitting process to derive the best QSAR model.

A group additive property (GAP) method of calculating conformational entropies, and its application to modeling synthetic polymers, has been given by Koehler and Hopfinger.[27] The methodology, however, is applicable to any molecule or set of molecules. The GAP concept assumes that some intrinsic contribution to any composite property, $P(i)$, is associated with the ith structural group of the molecule. The composite property, P, is simply taken to be the sum of the $P(i)$ composing the molecule

$$P = \sum_{i=1}^{m} P(i) \qquad (11)$$

where the index m corresponds to the number of structural groups in the molecular unit.

The Koehler and Hopfinger GAP method is labeled torsion angle unit, TAU, theory. A TAU is defined as

$$G_i \xrightarrow{\theta_i} G_{i+1}$$
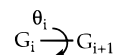
where G_i and G_{i+1} are structural groups connected by a bond about which the torsion angle θ_i occurs. TAUs permit a global molecular property of a molecule to be computed as a scalar sum of the individual molecular properties of the constituent TAUs. That is, the estimation of global molecular properties is formulated in terms of a GAP model.

The present work represents the first application of the TAU method as part of modeling a ligand−receptor (protein) system. The entropy contributions are limited to the torsion angle conformational entropy. A dataset of TAU entropy values has been reported by Hopfinger and Koehler.[27] The sum of the TAU entropy terms, assigned to each torsion angle of a molecule, approximates the maximum first-nearest group neighbor conformational entropy available to the entire molecule. Appropriate TAU values were selected for each type of torsion angle found in the structures of the inhibitors and the renin receptor model allowing the corresponding total molecular conformational entropies to be computed. The values for the sum of the TAU entropy values for each amino acid residue and the requisite isostere fragments of the peptidometic ligands are given in Table 5.

The conformational entropy of the renin receptor model is based upon only those amino acid residues in contact with the ligand. If any atom of an amino acid residue of the receptor model falls within 6 Å of any atom of the bound ligand, then that amino acid residue was included in the calculation of the conformational entropy of the receptor. Therefore, the conformational entropy of the receptor, which can change upon binding, is dependent on the size and conformation of each individual bound ligand.

The conformational entropy, as calculated by the sum of the TAU entropy values, represents the maximum confor-

Table 5. The Sum of the TAU Entropy Values for Amino Acid Residues and Other Structural Fragments

amino acid[a]	$\sum$TAU[b]	$\sum$TAU/n[c]
Gly	2.46	0.82
Ala	6.76	4.74
Val	11.64	9.65
Leu	14.70	10.67
Ile	14.70	10.67
Ser	6.31	2.10
Thr	10.64	6.43
Cys	5.26	1.75
Met	13.89	7.51
Phe	3.65	1.33
Tyr	4.30	1.54
His	3.65	1.33
Asp	5.15	1.71
Asn	4.14	1.38
Glu	9.45	3.15
Gln	8.44	2.81
Arg	12.88	4.29
Lys	20.21	9.62
Pro	0.58	0.19
Trp	3.65	1.33
N-terminal acetate	4.33	4.33
C-terminal amide	0.00	0.00
reduced amide isostere[d]	8.49	2.83
hydroxyethylene isostere[d]	11.02	3.67
statine[d]	9.26	3.09
Iva[d]	12.88	10.07
Mba[d]	15.80	11.04

[a] [b] The sum of the TAU entropy values (cal/mol-K) for the amino acid residues (including ϕ, ψ torsion angles) and non-amino acid structural fragments. [c] TAU entropy values normalized by respective number of potential wells. [d] Chemical structure is given in Table 3.

Table 6. The Sum of the TAU Entropy Values for the Renin Inhibitors and Associated Receptor[a]

compd	TAU (L)	TAU/n (L)	TAU (R)	TAU/n (R)
U80631E	62.46	41.24	401.29	227.25
U77646E	63.04	41.43	455.69	262.80
U77647E	63.04	41.43	437.58	256.42
U73777E	26.19	12.07	338.10	185.86
U71909E	26.77	12.26	392.50	221.41
U77451E	42.57	23.30	448.55	257.40
U72407E	49.71	31.20	374.45	212.41
U72408E	50.29	31.39	409.45	227.18
U72409E	53.94	32.53	443.55	258.63
U77455E	66.14	39.60	452.09	258.58

[a] TAU and TAU/n (normalized) entropy values are reported in cal/mol-K for ligand (L) and associated receptor (R).

mational entropy that can be lost by the receptor and/or ligand upon binding. A working assumption, in using TAU entropies in FEFF 3D-QSAR analysis, is that the total change in conformational entropy upon binding is equal to the sum of the appropriately scaled maximum entropies of the ligand and receptor. The values for the sum of the TAU entropy values for each ligand and its associated receptor model are given in Table 6.

D. Construction of FEFF 3D-QSARs. Each of the FEFF terms computed in modeling the ligand−receptor binding process are considered as potential descriptors for the prediction of the thermodynamic binding properties. The individual nonscaled energy terms of the training set are the initial independent variables of a 3D-QSAR database. The dependent variables are the observed binding properties. The construction of a FEFF 3D-QSAR requires scaling all the contributions to the free energy of binding. This calibrated

force field can then be employed to predict the free energy of binding of new, untested compounds.

A general form of the free energy change in binding, ΔG, in terms of a possible free energy force field, FEFF, is given by eqs 12−16

$$\Delta G = G(B) - G(U) \qquad (12)$$

$$G(B) - G(U) = \Delta G(L \cdot M) + \Delta G(R \cdot M) + G(LR \cdot M) \qquad (13)$$

$$\Delta G(L \cdot M) = \sum_i [\sum_j \{\chi(L)[a_{ij}\Delta(r_{ij}^{-6}) + b_{ij}\Delta(r_{ij}^{-12})] + \delta(L)q_iq_jd_{ij}^{-1}\Delta(r_{ij}^{-1})F(\Delta(r_{ij}))\} + \alpha(L)\{c_i(\sum_j\Delta(s_{ij})) + f_i\}] + \sum_u[p_u\Delta(\cos(h_u\theta_u)) + \beta_u(L)e(\theta_u)h_u^{-1}] + K(L)[\sum_w k_w\Delta(\phi_w^2)] + \Gamma(L)[\sum_w\Delta(e(\phi_w))] + K'(L)[\sum_x k_x\Delta(r_x^2)] + \Gamma'(L)[\sum_x\Delta(e(r_x))] \qquad (14)$$

$$\Delta G(R \cdot M) = \sum_k [\sum_l \{\chi(R)[a_{kl}\Delta(r_{kl}^{-6}) + b_{kl}\Delta(r_{kl}^{-12})] + \delta(R)q_kq_ld_{kl}^{-1}\Delta(r_{kl}^{-1})F(\Delta(r_{kl}))\} + \alpha(R)\{c_k(\sum_l\Delta(s_{kl})) + f_k\}] + \sum_v[p_v\Delta(\cos(h_v\theta_v)) + \beta_v(R)e(\theta_v)h_v^{-1}] + K(R)[\sum_{w'} k_{w'}\Delta(\phi_{w'}^2)] + \Gamma(R)[\sum_{w'}\Delta(e(\phi_{w'}))] + K'(R)[\sum_{x'} k_{x'}\Delta(r_{x'}^2)] + \Gamma'(R)[\sum_{x'}\Delta(e(r_{x'}))] \qquad (15)$$

$$G(LR \cdot M) = \sum_i [\sum_k \{\chi(LR)[a_{ik}\Delta(r_{ik}^{-6}) + b_{ik}\Delta(r_{ik}^{-12})] + \delta(LR)q_iq_kd_{ik}^{-1}\Delta(r_{ik}^{-1})F(\Delta(r_{ik}))\} + \alpha(LR)\{c_i(\sum_k s_{ik}) + f_i\}] \qquad (16)$$

where $\Delta G(L \cdot M)$ is the change in free energy of binding for the solvated ligand, $\Delta G(R \cdot M)$ is the change in free energy of binding for the solvated receptor, and $G(LR \cdot M)$ is the intermolecular free energy of binding in solvent M. The terms, and symbols used in eqs 12−16, are defined in Table 7. Of course, the choice of energy terms and computational model/potential function for computing the energy contributions of a ligand−receptor FEFF is user-defined. The FEFF given by eqs 14−16 is only one possible representation. Equation 12 can also be represented by the free energy force field expression given by eq 9. The introduction of scaling coefficients to refine the FEFF is prescribed by the user. The set of scaling coefficients included in the FEFF represented by eqs 14−16 are defined in Table 8. In this study scaling was limited to binding constituent energy components, given in eq 9, as opposed to the terms within the force field contributions to the constituent binding energy components as given by eqs 14−16.

Most of the terms found in eqs 14−16 would be computed using the potential functions from a "typical" molecular mechanics force field. The term $\{c_i(\sum_j\Delta(s_{ij})) + f_i\}$ represents the solvation energy contribution determined from a solvated surface area, or volume, model. The conformational entropy, in the FEFF of eqs 14 and 15, is determined by the term $\beta\epsilon(\theta)h^{-1}$ (see Table 8) where $e(\theta)h^{-1}$ is the torsion angle contribution to the entropy computed using the TAU model. The vibrational entropy associated with bonds and bond angles were not calculated in this study but are included here to complete the description of the FEFF formalism.

The large majority of ligand−receptor systems have two troublesome properties with respect to developing statistical

LIGAND−RECEPTOR BINDING THERMODYNAMICS

J. Chem. Inf. Comput. Sci., Vol. 37, No. 4, 1997 **801**

Table 7. Thermodynamic and FEFF Symbols and Terms Used in Calibration of a FEFF

B - bound state
U - unbound state
L - ligand
R - receptor
M - solvent
G - LRM system Gibb's free energy
$F(\Delta(r_{ij}))$ - functional form of the distance-dependent molecular
    dielectric $= \Delta(r_{ij}^{-1})$ in the initial analyses
i and j - ligand nonbonded atom-pair indices
k and l - receptor nonbonded atom-pair indices
i and k - ligand−receptor atom-pair indices
$\Delta(r_{ij}^{-m}) = 1/r_{ij}^m(B) - 1/r_{ij}^m(U)$; difference in atom-pair
    functional distance-dependence between B and U states
a, b - 6−12 atom-pair potential parameters
q - partial atomic charges
$d_{ij}^{-1} = (d_i + d_j)^{-1}$ where d_i and d_j are atomic dielectrics of
    atoms i and j
c - solvation model parameter measuring ΔG per unit
    exposure factor
f- solvation model parameter measuring intrinsic G of
    solvation
s_{ij} - the exposure factor of atom i eliminated from solvent
    interaction by atom j
p, h - torsional barrier, p, and rotational periodicity, h,
    of a torsion angle potential
$\Delta(\cos(h_u\theta_u))$ - difference in the normalized torsion potential
    of torsion angle θ_u between the B and U states =
    $\cos(h_u\theta_u[B]) - \cos(h_u\theta_u[U])$
$k_{w(w')}$ - force constant for bond angle ϕ_w or $\phi_{w'}$
$k_{x(x')}$ - force constant for bond r_x or $r_{x'}$
$e(\theta_u)$ - intrinsic entropy of torsion angle, θ_u, of the ligand
$e(\theta_v)$ - intrinsic entropy of torsion angle, θ_v, of the receptor
$e(\phi_{w(w')})$ - vibrational entropy of bond angle, ϕ_w, of the ligand,
    or $\phi_{w'}$, of the receptor
$e(r_{x(x')})$ - vibrational entropy of bond r_x of the ligand, or $r_{x'}$
    of the receptor

Table 8. Scaling Coefficients Used in Calibration of a FEFF

$\chi(L)$ - scaling coefficient of the ligand van der Waals interaction
    energy term
$\chi(R)$ - scaling coefficient of the receptor van der Waals interaction
    energy term
$\chi(LR)$ - scaling coefficient of the ligand−receptor van der Waals
    interaction energy term
$\delta(L)$ - scaling coefficient of the ligand electrostatic energy term
$\delta(R)$ - scaling coefficient of the receptor electrostatic energy term
$\delta(LR)$ - scaling coefficient of the ligand−receptor interaction
    electrostatic energy term
$\alpha(L)$ - scaling coefficient of the ligand solvation energy term
$\alpha(R)$ - scaling coefficient of the receptor solvation energy term
$\alpha(LR)$ - scaling coefficient of the ligand−receptor interaction
    solvation energy term
$\beta_u(L)$ - scaling coefficient of the uth ligand torsion angle entropy
    using a TAU model
$\beta_v(R)$ - scaling coefficient of the vth receptor torsion angle entropy
    using a TAU model
$K(L)$ - scaling coefficient of the total change in the ligand bond
    angle energy
$\Gamma(L)$ - scaling coefficient of the total change in the ligand bond
    angle vibrational entropy
$K(R)$ - same as $K(L)$, but for the receptor bond angles
$\Gamma(R)$ - same as $\Gamma(L)$, but for the receptor bond angles
$K'(L)$ - same as $K(L)$, but for ligand bonds
$\Gamma'(L)$ - same as $\Gamma(L)$, but for ligand bonds
$K'(R)$ - same as $K'(L)$, but for receptor bonds
$\Gamma'(R)$ - same as $\Gamma'(L)$, but for receptor bonds

approaches to performing FEFF 3D-QSAR analysis. First, the number of analogs, m (observations), is usually small compared to the number of possible independent variables, n (FEFF energy terms which can be scaled). Secondly, many of the energy terms are interrelated to one another. A

methodology for optimizing a FEFF 3D-QSAR using a calibration approach appropriate for datasets having these two properties is the "funnel" strategy. In this approach all FEFF terms are considered as descriptors in a stepwise multiple linear regression (MLR) construction of trial QSARs. The genetic function algorithm (GFA) optimization method[28] is employed to construct trial QSARs. Statistical measures of significance including the correlation coefficient, r^2, cross-validated r^2, xv-r^2, the F-statistic, F, and lack-of-fit measure, LOF,[29] are calculated to test for the robustness of the models. In each case the cross-correlation descriptor matrix is examined to eliminate trial QSARs in which pairs of energy terms have cross-correlation coefficients greater than 0.50. In addition, attention is given to the sign of the coefficient of the energy descriptors in the trial QSARs. If a trial QSAR contains a FEFF term that has been assigned a scaling coefficient of opposite sign to that of the nonscaled term and does not admit to a plausible explanation, then that QSAR is disregarded. Descriptor use in a GFA analysis as a function of the number of crossovers is also monitored as an indication of statistical significance.

If calibration of the entire FEFF is considered as the most general, all-inclusive approach to construction of RD 3D-QSARs, then the "funneling" of FEFF terms may be defined as a specific case, or subset, of the calibration approach. In the calibration approach all FEFF contributions and/or the parameters that define these terms are assigned scaling coefficients. Consequently, all of these contributions are accounted for in the final FEFF 3D-QSAR. In the "funnel" approach each FEFF term, or a combination of appropriate FEFF terms, is considered as a "whole" descriptor in the construction of FEFF 3D-QSARs. In this "funneling" approach all the FEFF terms are "filtered" to separate out those terms that are most significant in predicting the free energy of binding, and, consequently, the problem of overfitting is circumvented by constructing the FEFF 3D-QSAR in increments.

E. Computational Details. The sampling of the energy-geometry space of the renin−inhibitor system began with the ligand docked in the active site of the enzyme. An initial MD simulation of 20 ps sampling time using a time step of 0.5 fs was performed. An average temperature of 200 K was held constant during the simulation by coupling the system to a temperature bath with a relaxtion time of 0.01 ps.

Structural models of the complex as well as the associated energy terms were retained every 0.2 ps of the simulation. The structure of lowest total potential energy for the simulation at 200 K was used as the starting state for a MD simulation at 100 K consisting of 10 ps of sampling time using a time step of 0.5 fs. A structural model of the complex, and its associated energy terms, was retained every 0.1 ps of the simulation. The structure of lowest total potential energy for the simulation at 100 K was subsequently used as the starting structure for a 10 ps MD simulation at 50 K. Again the lowest energy structure from the 50 K simulation was used as the starting state for a 10 ps MD simulation evaluated at 25 K. The lowest energy configuration of the complex from the 25 K simulation serves as the starting state for a 10 ps simulation evaluated at 10 K. The lowest energy structure from each simulation temperature is the representative of the complex state from which all bound FEFF terms are derived.

The sampling scheme for the unbound state began with the bound conformations of the individual ligand and corresponding receptor geometry from the lowest energy complex state sampled during the 10 K MD simulation. The bound ligand was separated from the complex. Separate MD simulations at 10 K were subsequently performed on the disassociated ligand and corresponding receptor geometry employing the same conditions as used for sampling the complex. Because of the potential added flexibility of the ligands, an attempt was made to more fully explore their conformational spaces. Fifty picosecond simulations using a time step of 1 fs was employed for sampling the phase space of each of the ligands. The lowest energy conformation from the simulation at 10 K for each unbound ligand, and its associated unbound receptor geometry, was subsequently used as the starting configuration of a MD simulation at 25 K. This process of "heating up" the unbound ligand and receptor was repeated at 50, 100, and 200 K.

Finally, the lowest energy conformational states of the unbound receptor and ligand obtained at the 200 K simulation was subsequently used in separate MD simulations evaluated at 100, 50, 25, and 10 K. The duration of the simulations for the receptor in this "cooling down" step was 5 ps, whereas for the ligands the sampling time was extended to 50 ps. The lowest energy structures of each ligand and receptor from the 200 K and subsequent "cooled down" simulations were then chosen as the representative conformations of the unbound state at each respective simulation temperature. The free energy contributions from both the intermolecular and intramolecular interactions to ligand−receptor binding were accumulated during all of the simulations.

If the total energy of one of the complexes was significantly higher than the other ligand−receptor complexes, and this energy did not correlate with the trend in relative experimental free energies of binding, then the questionable simulation was rerun with a different set of initial random velocities. In most simulations the resultant energies using a different set of initial random velocities were found to be similar. However, in a few cases a lower energy state was found. The structure of lower energy was then used as the starting state of ensuing MD simulations in the MD sampling scheme.

Schemes were developed to, hopefully, sample a wide variety of possible conformations of the unbound ligand. In addition to using the bound ligand conformation present in low energy complexes as starting points in unbound ligand MD simulations, the following sampling schemes were employed for conformational analyses of the unbound ligand:

(1) Starting from the N-terminal residue of the ligand, systematic torsion angle searching was done for the dipeptide of the ligand. The lowest energy dipeptide conformation was retained, the next residue was added, and the conformational search was repeated for the "dipeptide" of the C-terminal side of the molecule. This process was repeated until the entire ligand was built and subjected to these sequential "dipeptide" conformational searches. The resulting lowest energy conformation was then used as the starting structure in a MD simulation of 50 ps at each of the temperatures employed in the ligand−receptor complex simulations.

(2) α-Helix ($\Phi = -60°$, $\Psi = -40°$), β-sheet ($\Phi = -130°$, $\Psi = 120°$), and "random" coil conformations were used as starting structures in 50 ps MD simulations at each of the

Table 9. Definition of FEFF Terms Used in Constructing FEFF 3D-QSARs

$\Delta E_{stretch}$	change in stretching energy upon binding
ΔE_{bend}	change in bending energy upon binding
$\Delta E_{torsion}$	change in torsional energy upon binding
ΔE_{vdW}	change in van der Waals energy upon binding
$\Delta E_{electrostatic}$	change in electrostatic energy upon binding
$\Delta E_{E1,4}$	change in 1,4 interaction energy upon binding
ΔE_{hb}	change in hydrogen bonding energy upon binding
ΔE_{solv}	change in solvation energy upon binding
$\Delta E_{stre+bend}$	sum of changes in stretching and bending energies
$\Delta E_{stre+bend+tor}$	sum of changes in stretching, bending and torsion energies
ΔE_{el+hb}	sum of changes in electrostatic and hydrogen bonding energies
$\Delta E_{el+hb+E1,4}$	sum of changes in electrostatic, hydrogen bonding, and 1,4 interaction energies
$E_{LR}(LL,RR,LR)$	ligand−receptor complex energy
$E_{LR}(LR)$	intermolecular ligand−receptor energy
$E_{LR,vdW}$	van der Waals intermolecular ligand−receptor energy
$E_{LR,electrostatic}$	electrostatic intermolecular ligand−receptor energy
$E_{LR,hb}$	hydrogen bonding intermolecular ligand−receptor energy
$E_{LR,el+hb}$	sum of electrostatic and hydrogen bonding intermolecular ligand−receptor energies
$E_{LR,el+hb+vdW}$	sum of electrostatic, hydrogen bonding, and van der Waals intermolecular ligand−receptor energies
$\Delta E_L(LL)$	change in intramolecular ligand energy upon binding
$E_{LR}(LL)$	intramolecular energy of bound ligand
$E_L(LL)$	intramolecular energy of unbound ligand
$\Delta E_R(RR)$	change in intramolecular receptor energy upon binding
$E_{LR}(RR)$	intramolecular energy of bound receptor
$E_R(RR)$	intramolecular energy of unbound receptor
$\Delta S_L(LL)$	torsion angle conformational entropy change of ligand upon binding
$\Delta S_L(\langle LL \rangle)$	normalized torsion angle conformational entropy change of ligand upon binding
$\Delta S_R(RR)$	torsion angle conformational entropy change of receptor upon binding
$\Delta S_R(\langle RR \rangle)$	normalized torsion angle conformational entropy change of receptor upon binding
$\Delta S_{L+R}(L+R)$	sum of torsion angle conformational entropy changes of ligand and receptor upon binding
$\Delta S_{L+R}(\langle L+R \rangle)$	sum of normalized torsion angle conformational entropy changes of ligand and receptor upon binding
$E_{LR}(LRM)$	ligand−receptor complex solvation energy
$\Delta E_L(LM)$	change in ligand solvation energy upon binding
$E_{LR}(LM)$	bound ligand solvation energy
$E_L(LM)$	unbound ligand solvation energy
$\Delta E_R(RM)$	change in receptor solvation energy upon binding
$E_{LR}(RM)$	bound receptor solvation energy
$E_R(RM)$	unbound receptor solvation energy

complex simulation temperatures. The random coil structure was taken as the lowest energy conformation realized in a 50 ps, 800 K, simulation using the 10 K bound conformation as the starting structure.

The lowest-energy conformation of the ligand observed from all of these sampling procedures was taken to represent the unbound ligand state.

The most straightforward approach to calculating a non-scaled FEFF term from the MD simulation sampling scheme is to use the corresponding FEFF energy terms of the lowest energy state for the complex, E_{LR}, unbound ligand, E_L, and unbound receptor, E_R, found at a particular simulation temperature. The FEFF terms which were used in this FEFF 3D-QSAR analysis are given in Table 9. These particular FEFF terms are defined on the basis of constraining equivalent types of scaling coefficients (Table 8) of the ligand and, separately, the receptor, to be identical to one another. The FEFF terms that represent energy changes upon binding, as described in Table 9, were calculated by eq 10 where E represents an energy value for a FEFF term or combination of terms.

LIGAND−RECEPTOR BINDING THERMODYNAMICS

J. Chem. Inf. Comput. Sci., Vol. 37, No. 4, 1997 **803**

Table 10. Prediction of ΔG as a Function of Simulation Temperature Using One-, Two-, and Three-Descriptor FEFF 3D-QSARs

temp (K)	3D-QSAR[a]	$r^{2\,b}$	xv-$r^{2\,c}$	F[d]	LOF[e]
	Best One-Descriptor Models				
200	$0.05(\pm0.01)E_L(LL) + 7.55(\pm0.66)$	0.77	0.69	26.65	0.47
100	$0.06(\pm0.01)E_L(LL) + 8.90(\pm0.53)$	0.67	0.53	16.49	0.66
50	$0.04(\pm0.01)\Delta S_R(\langle RR \rangle) + 1.26(\pm2.38)$	0.67	0.57	16.19	0.66
25	$0.04(\pm0.01)\Delta S_R(\langle RR \rangle) + 1.26(\pm2.38)$	0.67	0.57	16.19	0.66
10	$0.04(\pm0.01)\Delta S_R(\langle RR \rangle) + 1.26(\pm2.38)$	0.67	0.57	16.19	0.66
	Best Two-Descriptor Models				
200	$0.06(\pm0.008)\,E_L(LL) - 0.05(\pm0.02)\Delta E_{solv} + 7.74(\pm0.57)$	0.85	0.77	20.61	0.43
100	$0.06(\pm0.01)E_L(LL) + 0.04(\pm0.01)\Delta S_L(LL) + 7.10(\pm0.79)$	0.84	0.69	17.91	0.49
50	$-0.09(\pm0.02)E_{LR,vdW} + 0.04(\pm0.01)\Delta E_{torsion} + 4.10(\pm0.02)$	0.80	0.71	14.02	0.60
25	$0.02(\pm0.01)E_L(LL) + 0.03(\pm0.01)\Delta S_R(\langle RR \rangle) + 2.06(\pm2.29)$	0.75	0.62	10.40	0.75
10	$0.02(\pm0.01)E_L(LL) + 0.04(\pm0.01)\Delta S_R(\langle RR \rangle) + 2.26(\pm2.32)$	0.75	0.62	10.40	0.75
	Best Three-Descriptor Models				
200	$0.06(\pm0.01)E_L(LL) - 0.06(\pm0.02)\Delta E_{solv} + 0.01(\pm0.009)\,\Delta E_{torsion} + 7.56(\pm0.55)$	0.89	0.80	16.37	0.53
100	$0.07(\pm0.01)E_L(LL) - 0.07(\pm0.02)\Delta E_{solv} + 0.02(\pm0.01)\Delta E_{stre+bend+tor} + 9.61(\pm0.46)$	0.87	0.69	13.45	0.63
50	$-0.10(\pm0.02)E_{LR,vdW} - 0.03(\pm0.02)\Delta E_{torsion} - 0.02(\pm0.01)\Delta E_{stretch} + 3.70(\pm1.29)$	0.85	0.54	11.38	0.22
25	$0.06(\pm0.01)\Delta S_R(\langle RR \rangle) + 0.10(\pm0.04)\Delta E_{solv} - 0.01(\pm0.005)\Delta E_{el+hb+E1,4} - 5.11(\pm3.58)$	0.82	0.61	9.27	0.26
10	$0.04(\pm0.01)E_L(LL) - 0.07(\pm0.03)\Delta E_{solv} - 0.07(\pm0.02)E_{LR,vdW} + 5.59(\pm1.47)$	0.83	0.34	9.51	0.25

[a] The values in parentheses define the 95% confidence limits. [b] Correlation coefficient. [c] Cross-validated r^2 (leave-one-out method).[31] [d] F-statistic. [e] LOF is the lack-of-fit measure[29] calculated using a smoothing factor of 0.5.

ΔE values for each of the simulation temperatures, namely, 200, 100, 50, 25, and 10 K, were computed using the lowest energy structures found in the MD simulations at each of the respective temperatures of evaluation. The MD trajectories provide multiple measures of the energetics of the ligand−receptor system. The resultant time-averaged energies were also evaluated as representatives of E_{LR}, E_L, and E_R in this analysis.

Another approach to calculating the nonscaled FEFF contributions was also explored. In this method the energy of the starting complex structure was optimized using 100 steps of steepest descent minimization followed by 300 conjugate gradient steps minimization. The initial minimization step size was 0.1 Å, and the energy convergence criteria was 0.01 kcal. The energies of both the unbound receptor and ligand were obtained using the same minimization protocol as for the complex. The starting conformation for the energy minimizations of the unbound receptor and unbound ligand were taken from the energy-minimized complex. The energy-minimized structures are chosen as the representative state of the complex, unbound receptor and unbound ligand in this approach.

All of the above approaches to searching for bound and unbound structures were compared with each other, and with the available crystal structure of the receptor, to determine the optimal scheme for generating reliable FEFF 3D-QSAR models.

RESULTS

The nonscaled FEFF terms that were found to provide the most significant statistical results were based on the complex and receptor structures of lowest E_{LR} and E_R values, respectively, that had been sampled in the multiple temperature MD simulation scheme. The ligand nonscaled FEFF terms were based on the ligand conformation with the lowest value of E_L which had been found from the various sampling schemes performed on the unbound ligands.

Since the experimental enthalpic, entropic, and free energy changes of enzyme-inhibitor binding are available, an attempt was made to predict each of these thermodynamic properties. Individual FEFF 3D-QSAR models were constructed for ΔG,

ΔH, and ΔS using the data collected from each simulation temperature, namely, 200, 100, 50, 25, and 10 K. GFA-MLR analysis, employing the WOLF program,[30] was used to construct trial FEFF 3D-QSARs. In all QSAR modeling the number of ligands in the analysis is $N = 10$.

All FEFF terms of Table 9 were considered as independent variables in the construction of FEFF 3D-QSAR models to predict the experimental values of ΔG listed in Table 4. The most significant one-, two-, and three-terms models, as ranked by the statistical criteria defined in the Methods section, are listed in Table 10. Significant models containing ΔE_{solv}, the change in solvation energy for the ligand−receptor system, $E_L(LL)$, the intramolecular energy of the unbound ligand, and/or a TAU term are especially prevalent at each of the simulation temperatures of evaluation. Stretching, bending, and/or torsion energy changes upon binding as well as $E_{LR,vdW}$, the intermolecular van der Waals interaction energy, are also found to add to the significance of the FEFF 3D-QSAR models. The higher term models show variability in the choice of significant descriptor terms as a function of temperature.

The results of a GFA optimization analysis for predicting ΔG, an example of which is shown in Figure 4 for the FEFF terms calculated at 200 K, reinforce the choice of terms found to be significant by the MLR stepwise regression analysis. FEFF 3D-QSAR model optimization is supported by the plateauing of the curves of descriptor usage, the y-axis, as a function of model evolution and optimization (crossover), the x-axis. The FEFF term, $E_L(LL)$, is found to be the most often used descriptor in building ΔG FEFF 3D-QSAR models as the number of crossovers increases in the evolution of the GFA optimization. The high usage of entropic FEFF terms as a function of increasing number of crossovers is also observed. A crossover versus usage plot reveals the relative significance of the descriptors. The more times a descriptor is used in generating new models, the greater its relative role in explaining variance in the biological activity.

Although only the optimal 3D-QSAR equations in terms of r^2, F, and LOF are shown in Table 10, other regression equations were found that are nearly as significant and which contain, in some cases, different energy terms as descriptors.
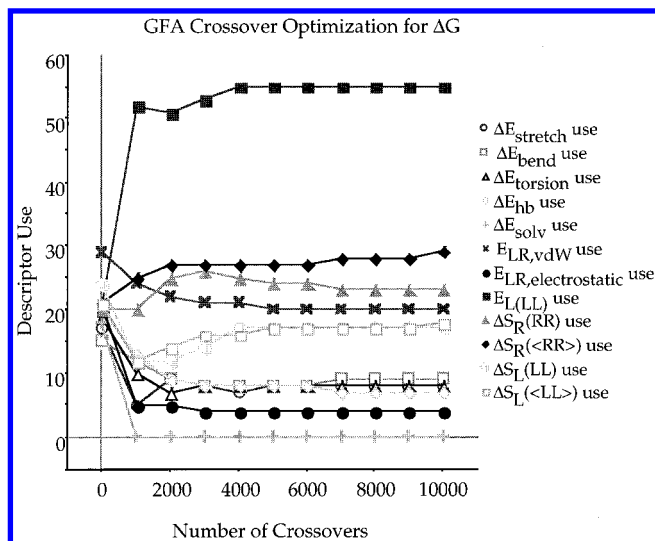
Figure 4. A graph showing the GFA crossover optimization for ΔG. The descriptor (FEFF terms) usage of the energy terms is plotted as a function of the number of crossovers in the GFA analysis. The FEFF terms in this figure are defined in Table 9 and were calculated at 200 K. Not all terms are shown but rather the most often used terms in model evolution. A smoothing factor of 1.0 was used.

It is important to be aware of the correlation between the descriptor terms in the GFA-MLR analysis. A correlation matrix of a select number of energy terms from the MD simulations run at 200 K is shown in Table 11. The correlation matrix is dependent on simulation temperature. The results of the GFA analysis, shown in Figure 4, indicate that ΔE_{solv} is not extensively used in model generation. However, in the simple stepwise MLR build-up of ΔG 3D-QSARs this energy term is found in several significant models.

The path of evolutionary optimization may be influenced by high correlations among some of the independent variables. Selective elimination of less significant free energy terms that are highly correlated (r > 0.50) with more significant terms as well as the elimination of terms that are assigned coefficients of questionable thermodynamic plausibility from the pool of independent variables (the GFA basis set) should improve the quality and the intrepretation of a GFA experiment. The results of this refined GFA-MLR analysis are shown in Figure 5. The use of ΔE_{solv} is increased in this GFA analysis as compared to the previous results which included all of the FEFF terms. In addition, the terms that are present in the optimized FEFF 3D-QSARs using the entire GFA basis set are again found to be frequently used in model evolution in the refined GFA analysis. These observations support the selective removal of cross-correlated energy terms in constructing the basis sets for GFA-MLR analysis.

Efforts to predict the observed enthalpy of binding yielded a number of significant models. The most significant one-, two-, and three-term FEFF 3D-QSAR models are listed in Table 12. All of the FEFF terms, excluding the TAU entropy terms, were considered as independent variables. The experimental ΔH values, listed in Table 4, were used as the dependent variables. An inspection of the 3D-QSAR models in Table 12 reveals that ΔE_{solv}, the change in solvation energy for the ligand−receptor system, is the single most dominant factor in determining ΔH. The calculation of ΔE_{solv} is independent of simulation temperature. Consequently, ΔE_{solv}

contributions were based on the lowest energy structures from the 10 K simulation. Analysis of the 100 and 200 K FEFF 3D QSARs shows that ΔE_{solv} and $E_L(LL)$, the intramolecular energy of the unbound ligand, explain a significant amount of the variation in the experimental ΔH values. $E_{LR,vdW}$, the intermolecular van der Waals interaction energy, as well as stretching, bending, and/or torsion energy changes upon binding are also found in some of the significant two-term models. A highly significant three-term model containing ΔE_{solv}, $E_L(LL)$, and $\Delta E_{stretch}$, the change in stretching energy upon binding, is constructed from an analysis of the energy terms calculated during the MD simulations run at 25 K.

The significance of the terms found in the FEFF 3D-QSAR models for predicting ΔH is reinforced by the results of the GFA optimization. In the GFA analysis of FEFF terms calculated at 25 K (see Figure 6) ΔE_{solv} and to a lesser extent $\Delta E_{stretch}$ are consistently used in model generation during crossover production in the evolution of the GFA optimization. The usage of other terms, such as $E_L(LL)$, ΔE_{bend}, and $\Delta E_{el+hb+E1,4}$ (see Table 9 for definition), is less but still significant.

A number of meaningful correlation models for the binding entropy are given in Table 13. The experimental ΔS values, listed in Table 4, multiplied by the temperature of the binding assay (310 K) were used as the dependent variables. The TAU entropy terms, all solvation energy-related terms, stretching, bending, and torsion energy changes and the intramolecular energy of the unbound ligand, were considered as independent variables. The intramolecular energy of the unbound ligand was included in the ΔS 3D-QSAR analysis, because it was found to be the most important FEFF term in predicting ΔG. Individually, the FEFF term, ΔS_R-($\langle RR \rangle$), explains a fair portion of the variation in the experimental ΔS values. This term is constant at all simulation temperatures for the chosen free energy force field representation. In the two-term and three-term FEFF 3D-QSAR models either ΔE_{solv}, $E_L(LL)$, one of the TAU terms and/or an energy term related to the stretching and bending energy changes are found to add to the significance of the prediction of ΔS. Conformational entropy, the intramolecular stability of the unbound ligand, and the binding solvation energy as well as changes in valence geometry energy contributions are, therefore, important in explaining the entropic contribution to the inhibitor−renin binding process.

The significance of the terms found in the FEFF 3D-QSAR models for predicting ΔS is reinforced by the evolution of the GFA optimization. For example, ΔE_{solv}, TAU entropy terms, $E_L(LL)$, and stretching and bending energy changes for the LR system upon binding are observed to be the most often used descriptors in the optimization evolution of FEFF 3D-QSAR models for binding entropy in a GFA analysis of the 25 K FEFF terms as shown by the crossover versus descriptor usage plot in Figure 7.

The resultant FEFF 3D-QSAR models for ΔG, ΔH, and ΔS indicate that FEFF terms found to be significant in the individual ΔH and ΔS models are also found to be collectively significant in the FEFF 3D-QSAR models for ΔG. Thus, the QSARs actually reflect the relationship $\Delta G = \Delta H - T\Delta S$, which is an encouraging representation of self-consistency among the models.

Analogues were considered outliers when the difference in predicted and observed energies exceed 2.0 standard

Table 11. Cross-Correlation Matrix of Select Energy Terms Evaluated at Simulation Temperature 200 K

energy term[a]	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1.00																			
2	0.29	1.00																		
3	−0.11	−0.52	1.00																	
4	−0.70	−0.16	0.12	1.00																
5	0.21	−0.18	−0.04	−0.18	1.00															
6	−0.61	0.13	0.07	0.19	0.03	1.00														
7	−0.19	0.14	−0.13	0.20	−0.95	−0.18	1.00													
8	−0.13	−0.24	0.47	0.03	−0.29	−0.03	0.31	1.00												
9	0.71	0.88	−0.43	−0.47	−0.03	−0.21	0.01	−0.24	1.00											
10	0.74	0.71	0.03	−0.46	−0.05	−0.20	−0.04	−0.03	0.89	1.00										
11	−0.07	−0.09	−0.59	0.13	0.11	−0.30	0.20	0.08	−0.10	−0.42	1.00									
12	−0.35	0.13	0.26	0.20	−0.62	0.28	0.39	−0.12	−0.08	0.05	−0.69	1.00								
13	0.33	0.19	−0.32	0.02	0.71	−0.01	−0.60	−0.18	0.30	0.17	0.35	−0.76	1.00							
14	−0.29	−0.14	0.24	−0.02	−0.77	−0.02	0.68	0.16	−0.25	−0.15	−0.30	0.75	−0.99	1.00						
15	0.65	0.42	−0.05	−0.72	0.16	−0.19	−0.23	−0.38	0.63	0.67	−0.33	0.05	−0.07	0.06	1.00					
16	0.52	−0.27	−0.05	−0.57	0.51	−0.45	−0.33	0.25	0.06	0.05	0.49	−0.84[b]	0.39	−0.39	0.24	1.00				
17	0.59	0.31	−0.40	−0.06	0.35	−0.38	−0.31	−0.66	0.52	0.38	0.05	−0.26	0.61	−0.57	0.27	0.03	1.00			
18	0.60	0.38	−0.38	−0.06	0.26	−0.35	−0.23	−0.66	0.58	0.44	0.00	−0.20	0.57	−0.52	0.31	−0.02	0.99	1.00		
19	0.74	−0.03	−0.30	−0.49	0.56	−0.57	−0.40	−0.29	0.35	0.23	0.42	−0.75	0.61	−0.59	0.42	0.75	0.65	0.61	1.00	
20	0.73	−0.04	−0.33	−0.47	0.57	−0.57	−0.41	−0.36	0.33	0.19	0.41	−0.73	0.60	−0.58	0.40	0.72	0.68	0.63	0.99	1.00

[a] Energy term identification (definition of energy terms is given in Table 9): $1 = \Delta E_{stretch}$, $2 = \Delta E_{bend}$, $3 = \Delta E_{torsion}$, $4 = \Delta E_{vdW}$, $5 = \Delta E_{electrostatic}$, $6 = \Delta E_{E1,4}$, $7 = \Delta E_{hb}$, $8 = \Delta E_{solv}$, $9 = \Delta E_{stre+bend}$, $10 = \Delta E_{stre+bend+tor}$, $11 = \Delta E_{el+hb+E1,4}$, $12 = E_{LR,vdW}$, $13 = E_{LR,electrostatic}$, $14 = E_{LR,hb}$, $15 = \Delta E_L(LL)$, $16 = E_L(LL)$, $17 = \Delta S_L(LL)$, $18 = \Delta S_L(\langle LL \rangle)$, $19 = \Delta S_R(RR)$, $20 = \Delta S_R(\langle RR \rangle)$. [b] Appears to be a chance correlation as these energy terms have correlations of 0.74, 0.52, 0.24, and 0.30 at 100 K, 50, 25, and 10 K, respectively.
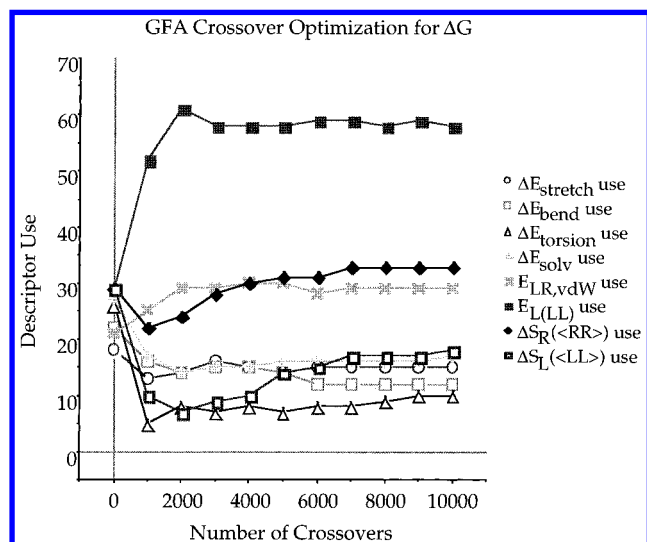
Figure 5. A graph showing the GFA crossover optimization for ΔG after selective removal of highly cross-correlated energy terms that were calculated at 200 K. The descriptor (FEFF terms) usage of the energy terms is plotted as a function of the number of crossovers in the GFA analysis. Only significant terms are shown. A smoothing factor of 1.0 was used.

deviations from the mean. An analysis of the most optimal three-term ΔG, ΔH, and ΔS QSAR models, in terms of r^2, reveals no outliers.

DISCUSSION

Several critical features inherent to modeling ligand−receptor systems were identified and considered over the course of this study. These critical features, their impact on modeling, and our means of dealing with them are summarized in Table 14 and discussed below and throughout the paper.

Firstly, the geometric representation of the system using a scaled-down enzyme model appears to promote the geometry of the receptor to drift from the X-ray structure unless constrained. The FEFF hydration shell solvation

model may incorrectly balance, over the course of a MD simulation, the solvation of two interacting hydrophilic groups with the opposing competition for the formation of hydrogen bonds between these groups, which requires some desolvation. More generally, solvation energetics may balance poorly against the rest of the FEFF and result in fatal energy drifting in an MD trajectory. Therefore, the solvation model was not included during the MD simulation. Rather, solvation energies, using the hydration shell solvation model, were calculated on the representative structures from the vacuum MD trajectories from which all other FEFF terms were also selected to construct QSAR models.

The protonation state of the two active-site aspartates in the complex was left unchanged when modeling the unbound receptor. Different protonation states of the unbound receptor were examined using the crystal structure of the bound enzyme as the initial geometry. Several protonation models appeared more stable than others in trial MD simulations. These stable models included the completely ionized state as well as the monoprotonated state where the lone proton resides at either of the two inner carboxylate oxygens. However, no one particular protonation state stood out as the most favorable. Therefore, whatever protonation state the receptor possessed in the complex was retained when modeling the unbound state.

With respect to the energy-geometry sampling schemes, it was found that the energy-minimized and time-averaged energies each proved to be less reliable in constructing FEFF 3D-QSARs than those taken from the lowest energy structure sampled from the multiple temperature MD simulation scheme. The results of the FEFF 3D-QSAR analysis of a series of transition state analog inhibitors of renin indicate that sampling both the ligand and receptor conformational spaces is critical for explaining the observed binding thermodynamics. In this study it was not sufficient to assume a stable intermolecular alignment of the ligand by docking it in the receptor, checking for bad steric contacts, and then performing an energy calculation. Energy minimization of

Table 12. Prediction of ΔH as a Function of Simulation Temperature Using One-, Two-, and Three-Descriptor FEFF 3D-QSARs

temp (K)	3D-QSAR[a]	$r^{2\,b}$	xv-$r^{2\,c}$	F^d	LOF[e]
	Best One-Descriptor Models				
200	$0.50(\pm0.18)\Delta E_{\text{solv}} - 25.30(\pm2.26)$	0.50	0.32	7.92	21.14
100	$0.50(\pm0.18)\Delta E_{\text{solv}} - 25.30(\pm2.26)$	0.50	0.32	7.92	21.14
50	$0.50(\pm0.18)\Delta E_{\text{solv}} - 25.30(\pm2.26)$	0.50	0.32	7.92	21.14
25	$0.50(\pm0.18)\Delta E_{\text{solv}} - 25.30(\pm2.26)$	0.50	0.32	7.92	21.14
10	$0.50(\pm0.18)\Delta E_{\text{solv}} - 25.30(\pm2.26)$	0.50	0.32	7.92	21.14
	Best Two-Descriptor Models				
200	$0.58(\pm0.15)\Delta E_{\text{solv}} - 0.12(\pm0.06)E_{\text{L}}(\text{LL}) - 18.53(\pm3.79)$	0.69	0.38	7.70	19.38
100	$0.61(\pm0.15)\Delta E_{\text{solv}} - 0.16(\pm0.07)E_{\text{L}}(\text{LL}) - 21.033(\pm2.55)$	0.71	0.34	8.49	18.20
50	$0.59(\pm0.16)\Delta E_{\text{solv}} - 0.13(\pm0.06)\Delta E_{\text{stre+bend+tor}} - 25.71(\pm1.93)$	0.68	0.39	7.56	19.64
25	$0.45(\pm0.14)\Delta E_{\text{solv}} + 0.23(\pm0.10)\Delta E_{\text{stretch}} - 23.34(\pm2.03)$	0.71	0.50	8.56	18.00
10	$0.56(\pm0.17)\Delta E_{\text{solv}} - 0.16(\pm0.10)E_{\text{LR,vdW}} - 12.46(\pm8.81)$	0.62	0.38	5.71	23.59
	Best Three-Descriptor Models				
200	$0.63(\pm0.15)\Delta E_{\text{solv}} - 0.13(\pm0.06)E_{\text{L}}(\text{LL}) + 0.03(\pm0.02)\Delta E_{\text{stre+bend}} - 17.33(\pm3.81)$	0.75	0.24	5.95	25.30
100	$0.54(\pm0.13)\Delta E_{\text{solv}} - 0.12(\pm0.06)\Delta E_{\text{stre+bend}} + 0.25(\pm0.20)E_{\text{LR,vdW}} - 9.55(\pm7.66)$	0.78	0.33	7.17	22.08
50	$0.78(\pm0.17)\Delta E_{\text{solv}} + 0.14(\pm0.07)\Delta E_{\text{stretch}} - 0.20(\pm0.07)E_{\text{L}}(\text{LL}) - 23.46(\pm1.83)$	0.79	0.45	7.42	21.34
25	$0.59(\pm0.11)\Delta E_{\text{solv}} + 0.26(\pm0.07)\Delta E_{\text{stretch}} - 0.16(\pm0.05)E_{\text{L}}(\text{LL}) - 23.73(\pm1.39)$	0.88	0.75	15.26	11.65
10	$0.64(\pm0.18)\Delta E_{\text{solv}} - 0.11(\pm0.09)E_{\text{L}}(\text{LL}) + 0.13(\pm0.10)E_{\text{LR,vdW}} - 15.49(\pm8.97)$	0.69	-0.01	4.46	31.13

[a] The values in parentheses define the 95% confidence limits. [b] Correlation coefficient. [c] Cross-validated r^2 (leave-one-out method).[31] [d] F-statistic. [e] LOF is the lack-of-fit measure[29] calculated using a smoothing factor of 0.5.

Table 13. Prediction of $T\Delta S$ as a Function of Simulation Temperature Using One-, Two-, and Three-Descriptor FEFF 3D-QSARs

| temp (K)	3D-QSAR[a]	$r^{2\,b}$	xv-$r^{2\,c}$	F^d	LOF[e]
	Best One-Descriptor Models				
200	$-0.17(\pm0.06)\Delta S_{\text{R}}(\langle\text{RR}\rangle) + 10.33(\pm15.40)$	0.48	0.25	7.32	28.01
100	$-0.17(\pm0.06)\Delta S_{\text{R}}(\langle\text{RR}\rangle) + 10.33(\pm15.40)$	0.48	0.25	7.32	28.01
50	$-0.17(\pm0.06)\Delta S_{\text{R}}(\langle\text{RR}\rangle) + 10.33(\pm15.40)$	0.48	0.25	7.32	28.01
25	$-0.17(\pm0.06)\Delta S_{\text{R}}(\langle\text{RR}\rangle) + 10.33(\pm15.40)$	0.48	0.25	7.32	28.01
10	$-0.17(\pm0.06)\Delta S_{\text{R}}(\langle\text{RR}\rangle) + 10.33(\pm15.40)$	0.48	0.25	7.32	28.01
	Best Two-Descriptor Models				
200	$0.63(\pm0.16)\Delta E_{\text{solv}} - 0.18(\pm0.06)E_{\text{L}}(\text{LL}) - 26.40(\pm4.00)$	0.73	0.47	9.36	21.53
100	$0.67(\pm0.16)\Delta E_{\text{solv}} - 0.23(\pm0.08)E_{\text{L}}(\text{LL}) - 30.64(\pm2.70)$	0.74	0.43	10.18	20.24
50	$-0.11(\pm0.03)\Delta S_{\text{L+R}}(\text{L+R}) + 0.25(\pm0.10)\Delta E_{\text{torsion}} + 14.34(\pm13.19)$	0.69	0.46	7.68	24.76
25	$0.39(\pm0.18)\Delta E_{\text{solv}} - 0.08(\pm0.04)\Delta S_{\text{R}}(\text{RR}) + 1.49(\pm16.62)$	0.66	0.38	6.89	26.65
10	$0.39(\pm0.18)\Delta E_{\text{solv}} - 0.08(\pm0.04)\Delta S_{\text{R}}(\text{RR}) + 1.49(\pm16.62)$	0.66	0.38	6.89	26.65
	Best Three-Descriptor Models				
200	$0.68(\pm0.16)\Delta E_{\text{solv}} - 0.19(\pm0.06)E_{\text{L}}(\text{LL}) + 0.03(\pm0.03)\Delta E_{\text{stre+bend}} - 25.16(\pm4.03)$	0.78	0.35	7.06	28.28
100	$0.67(\pm0.15)\Delta E_{\text{solv}} - 0.23(\pm0.07)E_{\text{L}}(\text{LL}) - 0.08(\pm0.06)\Delta E_{\text{stre+bend}} - 33.17(\pm3.30)$	0.80	0.40	7.84	26.06
50	$0.87(\pm0.19)\Delta E_{\text{solv}} - 0.27(\pm0.09)E_{\text{L}}(\text{LL}) + 0.16(\pm0.08)\Delta E_{\text{stre}} - 34.03(\pm2.08)$	0.79	0.44	7.31	27.52
25	$0.64(\pm0.15)\Delta E_{\text{solv}} - 0.21(\pm0.08)E_{\text{L}}(\text{LL}) + 0.28(\pm0.10)\Delta E_{\text{stre}} - 34.75(\pm1.95)$	0.82	0.59	9.22	22.84
10	$0.30(\pm0.20)\Delta E_{\text{solv}} - 0.12(\pm0.14)\Delta E_{\text{bend}} - 0.09(\pm0.04)\Delta S_{\text{R}}(\text{RR}) - 7.47(\pm20.83)$	0.71	0.18	4.79	37.72

[a] The values in parentheses define the 95% confidence limits. [b] Correlation coefficient. [c] Cross-validated r^2 (leave-one-out method).[31] [d] F-statistic. [e] LOF is the lack-of-fit measure[29] calculated using a smoothing factor of 0.5.

Table 14. Major Modeling Approximations, Corresponding Resulting Problems, and Adopted Solutions

approximation	problem	solution
1. FEFF Representation		
A. Hydration shell model is used to estimate solvation energetics.	Incorrect balance between solvation energy and the rest of the FEFF.	Only include solvation energetics for the state of the LR(M) used in the QSAR.
B. Entropy is modeled by a group additive model.	Entropy is independent of temperature and of maximum possible value.	The FEFF 3D-QSAR fitting will scale the entropy with respect to temperature.
2. Scale-down, or prune, the receptor geometry.	The scaled-down receptor geometry drifts away from the experimental form.	Use heavy masses on atoms of receptor fragments in the pruned receptor model.
3. Protonation/ionization states of the LR(M), R, and L.	Multiple protonation/ionization states are usually possible for LR(M) and R and sometimes for L without a clear selection choice.	Use all experimental data as constraints. For possible states which survive, use the state giving the global energy minimum for the most active ligands. Keep the overall protonation/ionization state constant for the LR(M) across the set of ligands.
4. Sampling schemes used to explore the geometry-energy states of the LR(M), R, and L.	The sampling schemes may be incomplete particularly with respect to bound ligand alignment, unbound ligand conformer state, and change in geometry of the receptor for bound and unbound states.	Use experimental data on receptor geometry for bound and unbound ligand states. Use geometries of homolog bound ligands as guides. Perform temperature-dependent MD simulation scans and independent conformational searches for the unbound ligand.

this "approximate" representation of the complex did not provide meaningful thermodynamic binding models. It was necessary to adopt a sampling scheme for modeling the ligand–receptor system that would allow each participant of the binding process to assume a relevant stable state in both the bound and unbound conformations. The temper-
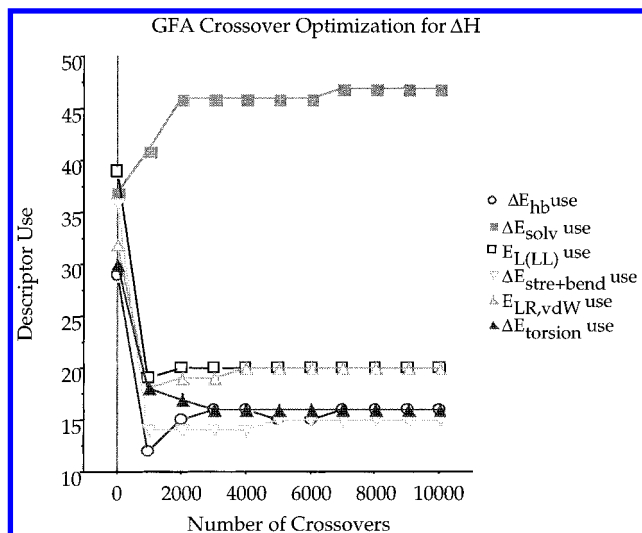
Figure 6. A graph showing the GFA crossover optimization for ΔH after selective removal of highly cross-correlated energy terms that were calculated at 25 K. The descriptor (FEFF terms) usage of the energy terms is plotted as a function of the number of crossovers in the GFA analysis. Only significant terms are shown. A smoothing factor of 1.0 was used.
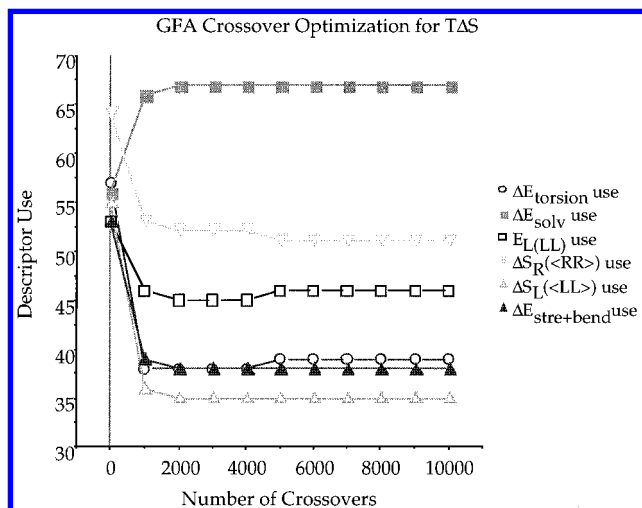
Figure 7. A graph showing the GFA crossover optimization for $T\Delta S$ after selective removal of highly cross-correlated energy terms that were calculated at 25 K The descriptor (FEFF terms) usage of the energy terms is plotted as a function of the number of crossovers in the GFA analysis. Only significant terms are shown. A smoothing factor of 1.0 was used.

ature-dependent MD simulation sampling scheme permits both sampling of LR, L, and R phase-spaces, and corresponding energy minimization, to achieve the most comprehensive analysis of the binding process.

One strategy of sampling ligand conformational space within the active site of the receptor developed in this study can be extended to the alignment sampling of any ligand−receptor system. When docking a series of structurally related compounds in a receptor of known geometry, smaller analogs should be selected to first sample the accessible active-site space. The larger analogs can then be aligned in the bound state based on the templates of the smaller ligands. Smaller ligands have greater alignment freedom at the receptor site than their larger counterparts at a fixed temperature. Hence, the small ligands can sample more phase-space for less computational resources than can large ligands. Conversely, one can first model the largest *active* ligand in the active site to limit the amount of intermolecular

sampling necessary for docking. An active large analog likely adopts a binding alignment relevant for all analog ligands, yet its determination requires less geometric exploration than for small ligands.

The goal of this study has been to develop a general formalism to calculate all energy contributions to a LR(M) binding process in order to predict the corresponding thermodynamic binding properties. A statistical analysis of all free energy contributions to the binding process, including those from entropic and solvation processes, was performed in an effort to construct a FEFF 3D-QSAR and thereby understand what thermodynamic interactions dominate in a particular LR(M) system. In essence, this methodology "filters" those terms which are not important in the binding process but allows the "funneling" of significant contributions.

Typical ligand−receptor systems are data sets where the number of analogs, m, (observations) is usually small compared to the number of possible independent variables, n (FEFF energy terms which can be scaled), and where many of the energy terms are interrelated to one another. Our approach to optimizing a FEFF 3D-QSAR for datasets having these two properties is highly dependent upon the use of a GA. In particular, the genetic function approximation, GFA,[28] appears to have some specific features well suited to treat FEFF 3D-QSAR problems. The GFA uses Friedman's LOF statistical fit measure[29] which prevents overfitting of the data by using too many independent variables. Thus, the LOF measure is one way to deal with the large number of available energy terms in the descriptor basis set. Both multivariate linear regression, MLR, and partial least squares, PLS, fitting techniques can be used to estimate the LOF measures in GFA.
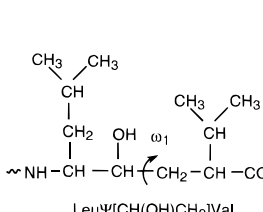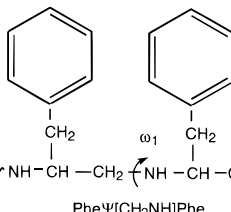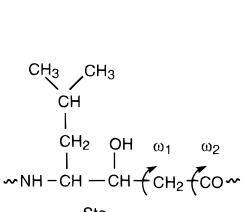
The GFA also includes power splining[29] of the independent variables. One use of power splines is to identify which independent variables and/or compounds need to be subpartitioned to yield significant correlations. The specific observations which are responsible for the necessity of subpartitioning can also be identified. Power splining can be used as a feedback diagnostic tool regarding the selection of energy terms to be scaled in evolving FEFF 3D-QSAR models.

The results of this study, ongoing additional applications of FEFF 3D-QSAR analysis, and reported experimental findings[4] lead us to advocate the development of a unique FEFF for each LR(M) system. Correspondingly, the individual FEFFs will not likely be transferable from one LR-(M) system to another. A "universal" FEFF is not likely to be realized, particularly if accurate binding thermodynamics are desired. It is also unlikely that general rules regarding the relative importance of the individual FEFF energy terms will be found. More likely is that the repetitive application of the FEFF 3D-QSAR strategy to an increasing number of LR(M) systems will lead to a better perspective on the relative uncertainty in estimating each of the FEFF energy terms and provide a basis for force field refinement.

For this particular LR(M) system it is found that the change in solvation energy upon binding and the intramolecular vacuum energy of the ligand in the unbound state are the most significant terms in determining binding free energy. Other terms such as stretching, bending, and torsion energy changes, the intermolecular van der Waals interaction energy, and an entropy term describing the change in

Table 15. The Values of the Φ, Ψ Angles of the Lowest-Energy Unbound Ligand Conformer State[a]

compd	P_5	P_4	P_3	P_2	$P_1{}^{b}$	$P_{1'}$	$P_{2'}$	state[c]	E_L(LL) (kcal/mol)
U80631E			−53, −33	−73, −13	−85, −64, −179,	−56, −50	−118, −44	α-helix	38.0
U77646E		−73, 55	−86, 76	−76, 64	−81, −60, −174,	−58, −64	−38, −67	α-helix	69.1
U77647E		−12, 13	58, 40	−143, −42	−141, −61, 178,	−58, −48	−97, −42	α-helix	69.0
U73777E			−86, 75	−33, −52	−81, 60, 179,	−57, 115		β-sheet	39.0
U71909E		−73, 57	−77, 72	−79, 73	−99, 64, 169,	−174, 113		see text	75.9
U77451E		−71, 62	62, 57	−8, −88	−84, 55, −177,	−64, 139		β-sheet	85.4
U72407E			−71, 66	−179, −48	−110, 60, −168, 72	−110, 89		β-sheet	26.1
U72408E		−75, 77	−65, 15	−96, −33	−131, 56, −174, 64	−138, 86		β-sheet	55.9
U72409E	−112, 115	−75, 135	53, 63	−56, 107	−142, 59, −179, 22	−98, 83		β-sheet	64.2
U77455E	−144, 133	−80, 51	−168, −63	−77, 151	−83, 57, 180, 85	−87, 93	−159, 83	see text	89.3

[a] P defines the binding site locations of each residue as defined by Schechter and Berger.[35] [b] Third and fourth angles listed are for ω_1 and ω_2 which are defined for the transition state isosteres above. [c] Defines the starting conformer state used in MD simulations to arrive at the lowest energy conformation.

conformational flexibility of the LR(M) system upon binding are also found to make significant contributions in some FEFF 3D-QSAR models. These significant QSAR energy terms are not independent of one another or of the other FEFF terms. The geometries that these significant QSAR energy terms are based upon are influenced by the other FEFF contributions to the overall LR(M) binding process. Interestingly, the best ΔG 3D-QSAR model (three FEFF terms) is found from the FEFF terms calculated at 200 K, whereas the best ΔH and ΔS 3D-QSAR models (three FEFF terms) are found at 25 K. One explanation for this observation is that the enthalpic and entropic contributions are both temperature dependent, and the optimal balance of these terms is not entirely captured in our representation of the FEFF.

The finding that only some of the FEFF terms are significant in a particular LR(M) binding process should not be considered unusual. Results of experimental work done by Chervenak and Toone[4] who studied the binding of different ligand−receptor systems in H_2O, and in D_2O, reveal that the major contribution to LR(M) binding free energy can be of a different thermodynamic source, as defined in Table 1, depending on the LR(M) system. For example, solvent reorganization provides anywhere from 25 to 100% of the observed enthalpy of binding for the range of systems investigated.[4] In a computational study by Vajda et al.[32] of a series of MHC receptor−peptide complexes it is found that both the internal energy and entropy of the ligand contribute significantly to the binding process. All of the above-mentioned types of thermodynamic contributions are found to be significant in our FEFF 3D-QSAR analyses. The finding of Holloway and co-workers[33] that the intermolecular van der Waals and electrostatic interactions between HIV-1 protease and a series of inhibitors could explain the variance in the observed binding affinities of these ligands is another example where only a small number of FEFF terms can predict a significant amount of the binding affinity differences among a training set of ligand analogs.

Ortiz et al.[2] found in their study that the total calculated ligand−enzyme binding energies correlate weakly with the experimental activities for the phospholipase training sets. When, however, the PLS method, coupled to the GOLPE variable selection procedure[34] is applied to the components of the binding energies, a significant regression equation is obtained in which activity is correlated with the interaction energies of parts of the ligands and key regions of the enzyme. Ortiz and co-workers found that many of the energy terms included in the binding energy were not important (no significant correlation) in determining the differences in experimental activity. One of the typical regression binding models derived by the COMBINE analysis contains 46 selected energy variables that are identified to be responsible for the observed differences in inhibitory activity.

ΔG for the renin−inhibitor training set investigated in this study is found to be a complex function of both enthalpic and entropic contributions. It was encouraging to find that those FEFF terms independently found to be significant in predicting ΔH and ΔS were jointly found in the independent construction of the FEFF 3D-QSAR models to predict ΔG. Epps et al.,[8] as part of their experimental analysis of the renin−inhibitory peptide system, determined that ΔG did not correlate with individually ΔH or with ΔS, suggesting that neither ΔH nor ΔS is dominant in the binding process for the set of inhibitors. However, ΔH and ΔS were found to be highly correlated ($r = 0.988$). Interestingly, our FEFF 3D-QSAR analyses to predict ΔH and ΔS led to similar two-descriptor models for these two thermodynamic properties, each containing ΔE_{solv} and E_L(LL), at both 100 and 200 K. Each FEFF 3D-QSAR model explains about the same amount of variance, r^2, relative to the experimental values of ΔH and ΔS. The approximately linear relationship between ΔH and ΔS is interpreted by Epps et al.[8] to be indicative of the interaction of water with the unbound peptide inhibitor in solution. These workers also concluded from CD spectra analysis, and the experimental binding data, that the relative conformational stability of each ligand in

Ligand−Receptor Binding Thermodynamics

J. Chem. Inf. Comput. Sci., Vol. 37, No. 4, 1997 **809**

Table 16. The Values of the χ_1 and χ_2 Angles of the Lowest-Energy Unbound Ligand Conformer State

compd	Iva[a]	P_5	P_4	P_3	P_2	P_1	$P_{1'}$	$P_{2'}$	Mba[b]
U80631E				179, 99	63, 99	−60, 178	−59, −	−55, −174	
U77646E				−61, 168	−66, −117	−56, 178	−57, −	−59, 174	
U77647E				−51, 147	−53, −100	−60, 178	−62, −	−60, −180	
U73777E				178, 44	−67, 32	−175, 112	−59, −70		
U71909E				−174, 61	−56, 58	177, −105	−66, 130		
U77451E				−53, −37	−171, −5	−55, −65	−177, −88		−175, −176, 67
U72407E				−174, 50	179, −18	−58, −177	−56, 180		
U72408E				−169, 81	177, 22	−62, 172	−58, −177		
U72409E		−171, −28		−175, 58	−174, 93	−67, 177	−59, −178		
U77455E	−161, −178	−178, −78		−179, 90	178, 42	−64, −178	−54, −178	−174, 84	

[a] Iva (N-terminal group) torsion angles as defined below. [b] Mba (C-terminal group) torsion angles as defined below.

Iva

Mba

a

b

c

d
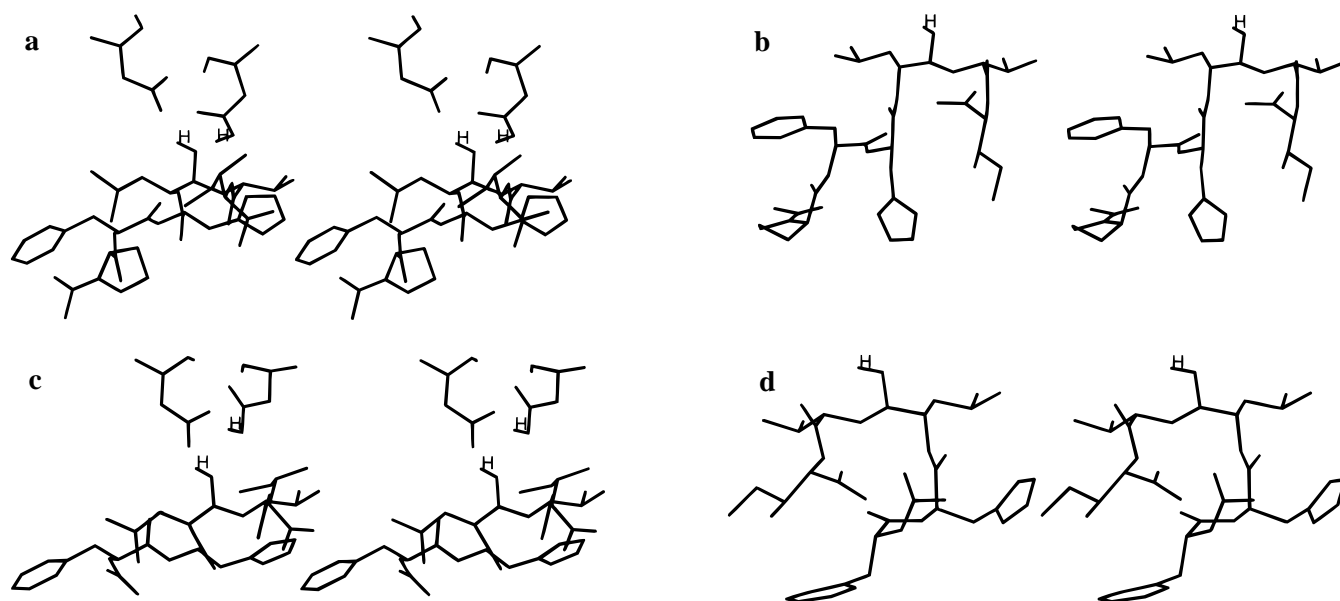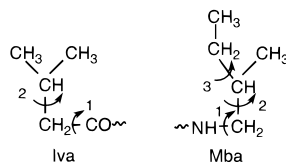
Figure 8. Stereo stick model representations of the bound conformation (along with the two active-site aspartates of the receptor model) and the unbound conformation for each of two ligands (only the hydrogen of the transition state isostere hydroxyl group of the ligand and the hydrogen of the aspartyl carboxylate are explicitly shown and labeled); (a) bound conformation of U77646E; (b) unbound conformation of U77646E; (c) bound conformation of U80631E; and (d) unbound conformation of U80631E.

solution was the most important property governing the corresponding differences in binding free energies for these analogs.

Epps and co-workers[8] also propose that if the inhibitors exist as ordered, rather than random, structures in solution, then differences in the types of ordered conformations would contribute to the changes in the binding free energies. That is, the structural differences are also expressed as differences in corresponding conformational energies. Since some peptide inhibitors are approximately of the same size, the relative intramolecular energy differences may very well correlate with the type of structures they adopt. The intramolecular energy of the unbound ligand, $E_L(LL)$, determined from a conformation found from extensive searching using a number of different sampling schemes, was found to be a significant FEFF 3D-QSAR descriptor in predicting ΔH, ΔS, and ΔG. For some ligands the unbound conformation resembles a right-handed α-helical conformation and for other ligands either an extended β-sheet conformation or a "nonordered" conformation. Tables 15

and 16 list the torsion angles of the lowest energy conformation sampled for each inhibitor in the unbound state as found in a 200 K MD simulation. The starting conformation that led to the lowest energy structure for all of the unbound hydroxyethylene isostere type inhibitors was a right-handed α-helical conformation. An extended β-sheet conformation led to the lowest energy unbound conformation for all of the statine isostere type analogs and compound U73777E. The sequential-building conformational analysis technique provided the starting geometry that led to the lowest energy conformation for compound U71909E. Heating the bound conformation of U77451E up to 800 K and then "cooling" it down at each of the temperatures of evaluation led to the lowest energy unbound conformation for this ligand.

Differences between the bound and unbound conformations of each of the ligands may play a role in their binding affinity. Figure 8 shows the bound and unbound conformations of U80631E and U77646E. The structural difference between these hydroxyethylene transition state isostere inhibitors is that U77646E contains an additional residue,

namely an N-terminal proline. This structural difference may be responsible for the increased β-structure character of the unbound conformation adopted by U77646E as compared to that of U80631E. Proline is not a β-former residue, but the unbound conformation adopted by U77646E is closer in conformer space to a β structure than to an α-structure. The observation that the bound state is an extended β-strand conformation may partially explain the higher activity of U77646E compared to U80631E. The greater the similarity between the bound and unbound conformations infers smaller entropic and/or enthalpic energy binding penalties.

Interestingly, $E_L(LL)$, and not $\Delta E_L(LL)$, is found in the best correlation equations. $\Delta E_L(LL)$, expressed as $[aE_B(LL) - bE_L(LL)]$ where $a \neq b$ and are regression coefficients, can be incorporated into significant, but not the best, QSAR models. Thus, the contribution from $E_B(LL)$ appears not to be sufficiently significant to overcome the penalty of an additional degree of freedom in a model and is deleted in the GFA optimization. Still, it has been pointed out by a reviewer that ordered conformations may not be realized in solution and that our QSAR models may simply be ligand-size dependent correlations as reflected in $E_L(LL)$. Inspection of the training set (post-analysis) indicates evidence for this point of view. For example, U77455E is approximately two "residues" larger than U72409E and has a more than 10-fold better (lower) K_d (see Table 4). On the other hand, U72409E and U77646E are both hexapeptides but differ in K_d by a factor of 4, $-\Delta H°$ by 6 kcal/mol, and $\Delta S°$ by about 23 esu. Thus, there is also a significant structural specificity inherent to the training set. Overall, the training set for this FEFF 3D-QSAR appears to possess both ligand size and structural components in terms of binding to renin.

An analysis of the optimal FEFF 3D-QSARs suggests that models to predict ΔS are as significant as the models constructed to predict ΔH but slightly less significant than those to forecast ΔG. Clearly, if one is to be successful in predicting the binding affinities of the inhibitors found in Table 4, an account of entropy must be included in the analysis.

Overall, FEFF 3D-QSAR analysis includes all contributions to the free energy of binding, including solvation and entropy in constructing 3D-QSAR models. Furthermore, all members of the binding process, namely, the complex, the unbound ligand, and unbound receptor, are included in the analysis. By measuring the binding thermodynamic properties at different temperatures one can test the significance of the analysis as a function of simulation temperature (balancing enthalpy and entropy contributions). The statistical analysis of all the free energy contributions to the binding process filters through only those terms that are found to be most significant and packages them in a FEFF 3D-QSAR model. The resultant FEFF 3D-QSARs may be used to predict the free energy of ligands outside the training set.

The automation of FEFF 3D-QSAR analysis into a bench tool would be useful in pharmaceutical CAMD applications. This is a current objective in our laboratory.

ACKNOWLEDGMENT

REFERENCES AND NOTES

(1) Hopfinger, A. J.; Nakata, Y.; Max, N. In *Intermolecular Forces*; Pullman, B., Ed.; Reidel-Dordrecht: Netherlands, 1981; p 431.

(2) Ortiz, A. R.; Pisabarro, M. T.; Gago, F.; Wade, R. C. Prediction of Drug Binding Affinities by Comparative Binding Energy Analysis. *J. Med. Chem.* **1995**, *38*, 2681−2691.

(3) Hopfinger, A. J. A QSAR Investigation of Dihydrofolate Reductase Inhibition by Baker Triazines Based Upon Molecular Shape Analysis. *J. Am. Chem. Soc.* **1980**, *102*, 7196−7206.

(4) Chervenak, M. C.; Toone, E. J. A Direct Measure of the Contribution of Solvent Reorganization to the Enthalpy of Ligand Binding. *J. Am. Chem. Soc.* **1994**, *116*, 10533−10539.

(5) Sialecki, A. R.; Hayakawa, K.; Fujinaga, M.; Murphy, M.; Fraser, M.; Muir, A. Carilli, C.; Lewicki, J. A.; Baxter, J. D.; James, M. N. G. Structure of Recombinant Human Renin, a Target for Cardiovascular-Active Drugs, at 2.5 Å Resolution. *Science* **1989**, *243*, 1346−1351.

(6) Rahuel, J.; Priestle, J. P.; Grütter, M. G. The Crystal Structures of Recombinant Glycosylated Human Renin Alone and in Complex with a Transition State Analog Inhibitor. *J. Struct. Biol.* **1991**, *107*, 227−236.

(7) Dhanaraj, V.; Dealwis, C. G.; Frazao, C.; Badasso, M.; Sibanda, B. L.; Tickle, I. J.; Cooper, J. B.; Driessen, H. P. C.; Newman, M.; Aguilar, C.; Wood, S. P.; Blundell, T. L.; Hobart, P. M.; Geoghegan, K. F.; Ammirati, M. J.; Danley, D. E.; O'Connor, B. A.; Hoover, D. J. X-ray Analyses of Peptide-Inhibitor Complexes Define the Structural Basis of Specificity for Human and Mouse Renins. *Nature* **1992**, *357*, 466−472.

(8) Epps, D. E.; Cheney, J.; Schostarez, H.; Sawyer, T. K.; Prairie, M.; Krueger, W. C.; Mandel, F. Thermodynamics of the Interaction of Inhibitors with the Binding Site of Recombinant Human Renin. *J. Med. Chem.* **1990**, *33*, 2080−2086.

(9) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protetin Databank, a Computer Based Archival File for Macromolecular Structures. *J. Mol. Biol.* **1977**, *112*, 535−542.

(10) Pearlstein R. A. CHEMLAB-II Users Guide, Version 11.1; Molecular Simulations Inc.: 16 New England Executive Park, Burlington, MA 01803, 1991.

(11) Tokarski, J. T.; Hopfinger, A. J. Constructing Protein Models for Ligand−Receptor Binding Thermodynamics Simulations: An Application to a Set of Renin Inhibitors. *J. Chem. Inf. Comput. Sci.* **1997** *37*, 779−791.

(12) QUANTA Version 3.3; Molecular Simulations Inc.: 16 New England Executive Park, Burlington, MA 01803, 1993.

(13) Doherty, D. C. MOLSIM User Guide; The Chem21 Group, 1780 Wilson Dr., Lake Forest, IL 60045, 1994.

(14) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T. An All Atom Force Field for Simulations of Proteins and Nucleic Acids. *J. Comput. Chem.* **1986**, *7*, 230−252.

(15) Epps, D. E.; Schostarez, H.; Argoudelis, C. V.; Poorman, R. A.; Hinzmann, J.; Sawyer, T. K.; Mandel, F. An Experimental Method for the Determination of Enzyme-Competitive Inhibitor Dissociation Constants form Displacement Curves: Application to Human Renin Using Flourescence Energy Transfer to a Synthetic Dansylated Inhibitor Peptide. *Anal. Biochem.* **1989**, *181*, 172−181.

(16) Epps, personal communication, 1994.

(17) Stewart J. J. P.; Seiler, F. K. QCPE #455 (Ver 4.0), 1987.

(18) Blundell, T. L.; Cooper, J.; Foundling, S. I.; Jones, D. M.; Atrash, B.; Szelke, M. On the Rational Design of Renin Inhibitors: X-ray Studies of Aspartic Proteinases Complexed with Transition-State Analogues. *Biochemistry* **1987**, *26*, 5585−5590.

(19) Swain. A. L.; Gustchina, A.; Wlodawer, A. A Comparison of Three Inhibitors of Human Immunodeficiency Virus Protease. *Adv. Exp. Med. Biol.* **1992**, *306*, 433−441.

(20) Hyland, L. J.; Tomaszek, T. A. Jr.; Roberts, G. D.; Carr, S. A.; Magaard, V. W.; Bryan, H. L.; Fakhoury, S. A.; Moore, M. L.; Minnich, M. D.; Culp, J. S.; DesJarlais, R. L.; Meek, T. D. Human Immunodeficiency Virus-1 Protease. 1. Initial Velocity Studies and Kinetic Characteriztion of Reaction Intermediates by ^{18}O Isotope Exchange. *Biochemistry* **1991**, *30*, 8441−8453.

(21) Hyland, L. J.; Tomaszek, T. A, Meek, T. D. Human Immunodeficiency Virus-1 Protease. 2. Use of pH Rate Studies and Solvent Kinetic Isotope Effects to Elucidate Details of Chemical Mechanism. *Biochemistry* **1991**, *30*, 8454−8463.

(22) Goldblum, A.; Rayan, A.; Fliess, A.; Glick, M. Extending Crystallographic Information with Semiempirical Quantum Mechanics and Molecular Mechanics: A Case of Aspartic Proteinases. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 270−274.

(23) Harte, W. E.; Beveridge, D. L. Molecular Dynamics of HIV-1 Protease. *J. Am. Chem. Soc*. **1993**, *115*, 3883−3886.

(24) Allinger, N. L. Conformational Analysis. 130. MM2. A Hydrocarbon Force Field Utilizing V_1 and V_2 Torsional Terms. *J. Am. Chem. Soc.* **1977**, *99*, 8127−8134.

(25) Hopfinger, A. J. In *Conformational Properties of Macromolecules*; Academic Press: New York, 1973; p 38.

(26) (a) Hopfinger, A. J. In *Conformational Properties of Macromolecules;* Academic Press: New York, 1973; p 71. (b) Forsythe, K. H.; Hopfinger, A. J. The Influence of Solvent on the Secondary Structures of Poly(L-Alanine) and Poly (L-Proline). *Macromolecules* **1973**, *6*, 423−437.

(27) Koehler, M. G.; Hopfinger, A. J. Molecular Modelling of Polymers: 5. Inclusion of Intermolecular Energetics in Estimating Glass and Crystal-melt Transition Temperatures. *Polymer* **1989**, *30*, 116−126.

(28) Rogers, D.; Hopfinger, A. J. Application of Genetic Function Approximation to Quantitative Structure−Activity Relationships and Quantitative Structure-Property Relationships. *J. Chem. Inf. Comput. Sci*. **1994**, *34*, 854−866.

(29) Rogers, D. G/SPLINES: An Hybrid of Friedman's Multivariate Adaptive Regression Splines (MARS) Algorithm with Holland's Genetic Algorithm. In *The Proceedings of the Fourth International Conference on Genetic Algorithms*; San Diego, 1991; p 38.

(30) Rogers, D. WOLF Reference Manual Version 5.5; Molecular Simulation Inc.: 1994.

(31) Cramer III, R. D.; Bunce, J. D.; Patterson, D. E.; Frank, I. E. Crossvalidation, Bootstrapping, and Partial Least Squares Compared with Multiple Regression in Conventional QSAR Studies. *Quant. Struct. Act. Relat.* **1988**, *7*, 18−25.

(32) Vajda, S.; Weng, Z.; Rosenfeld, R.; DeLisi, C. Effect of Conformational Flexibility and Solvation on Receptor-Ligand Binding Free Energies. *Biochemistry* **1994**, *33*, 13977−13988.

(33) Holloway, M. K.; Wai, J. M.; Halgren, T. A.; Fitzgerald, P. M. D.; Vacca, J. P.; Dorsey, B. D.; Levin, R. B.; Thompson, W. J.; Chen, L. J.; deSolms, S..; Gaffin, N.; Ghosh, A. K.; Giuliani E. A.; Graham, S. L.; Guare, J. P.; Hungate, R. W.; Lyle, T. A.; Sanders, W. M.; Tucker, T. J.; Wiggins, M.; Wiscount, C. M.; Woltersdorf, O. W.; Young, S. D.; Drake, P. L.; Zuagy, J. A. *A priori* Prediction of Activity for HIV-1 Protease Inhibitors Employing Energy Minimization in the Active Site. *J. Med. Chem*. **1995**, *38*, 305−317.

(34) Baroni, M.; Constantino, G.; Cruciani, G.; Riganelli, D.; Valigi, R.; and Clementi S. GOLPE: An Advanced Chemometric Tool for 3D-QSAR Problems. *Quant. Struct. Act. Relat*. **1993**, *12*, 9−20.

(35) Schechter, I.; Berger, A. On the Size of the Active Site in Proteases. I. Papain. *Biochem. Biophys. Res. Commun*. **1967**, *27*, 157−162.