Protein/Ligand Binding Free Energies Calculated with Quantum Mechanics/Molecular Mechanics

Frauke Gräter,[†,‡] Sonja M. Schwarzl,[‡,§] Annick Dejaegere,[||] Stefan Fischer,[*,‡] and Jeremy C. Smith[§]

IWR−Computational Biochemistry and IWR−Computational Molecular Biophysics, University of Heidelberg, Im Neuenheimer Feld 368, 69120 Heidelberg, Germany, and Departement de Biologie et Génomique Structurales, UMR 7104, Ecole Supérieure de Biotechnologie de Strasbourg, Boulevard S. Brant, FR-67400 Illkirch, France.

Received: December 21, 2004; In Final Form: March 3, 2005

The calculation of binding affinities for flexible ligands has hitherto required the availability of reliable molecular mechanics parameters for the ligands, a restriction that can in principle be lifted by using a mixed quantum mechanics/molecular mechanics (QM/MM) representation in which the ligand is treated quantum mechanically. The feasibility of this approach is evaluated here, combining QM/MM with the Poisson−Boltzmann/surface area model of continuum solvation and testing the method on a set of 47 benzamidine derivatives binding to trypsin. The experimental range of the absolute binding energy ($\Delta G = -3.9$ to -7.6 kcal/mol) is reproduced well, with a root-mean-square (RMS) error of 1.2 kcal/mol. When QM/MM is applied without reoptimization to the very different ligands of FK506 binding protein the RMS error is only 0.7 kcal/mol. The results show that QM/MM is a promising new avenue for automated docking and scoring of flexible ligands. Suggestions are made for further improvements in accuracy.

Introduction

Protein/ligand recognition is a complex process, the determining factors of which are not quantitatively understood. However, steady progress is being made in the development of methods for quickly and accurately calculating protein/ligand binding affinities. Available methods differ in how the structures of the complexes are predicted (the "docking" problem) as well as in how the free energy is calculated from the predicted structure.[1,2]

As ever, a tradeoff between accuracy and computational expense is unavoidable. Empirical scoring functions designed for fast estimation of binding affinities for high-throughput virtual screening often neglect conformational flexibility of the protein, entropic effects, and (de)solvation during the binding process.[3] Typically, binding free energies calculated with these methods have average deviations from experiment of 2 kcal/mol or more, with correlation coefficients of approximately 0.5.[4−6] At the other end of the accuracy range are techniques based on molecular dynamics (MD) simulations in explicit solvent that make use of free energy perturbation (FEP), thermodynamic integration (TI), and linear interaction energy (LIE) methods.[7,8] Given sufficient sampling, MD methods calculate energetics of protein/ligand interactions with high accuracy, including all entropic and solvent effects as well as receptor flexibility. In recent applications, the deviations between calculated and experimental binding free energies are less than 1 kcal/mol, confirming the accuracy of FEP and related methods.[9−12] Although formulations for calculating absolute binding free energies have been suggested,[13] MD methods are generally restricted to the calculation of the relative binding free energies of similar ligands. In addition, their very high computational cost limits their applicability in virtual screening and drug design.[14,15]

Continuum approaches have been investigated with growing interest as a means of bridging the gap between empirical scoring functions and MD-based methods. Continuum methods approximate the solvent as a polarizable continuum, thus including solvent effects at reasonable computational cost.[16] Within the continuum framework, the electrostatic contribution to the solvation free energy can be determined by solving the Poisson−Boltzmann (PB) equation. The combination of PB with a surface area (SA) dependent term for the hydrophobic effect yields the total solvation free energy. Several studies have applied PB/SA methods to protein/ligand binding, most of which used MD simulations with an explicit solvent model to generate a conformational ensemble of the bound state.[17−22] This approach has also been found to yield high accuracy, with a root-mean-square deviation (RMSD) of ∼1 kcal/mol.[19−21] In recent work, the PB/SA continuum method was tested for the binding of six small, C_2-symmetric benzamidine-like ligands to trypsin. The resulting RMSD between the calculated and experimental values was 0.55 kcal/mol.[23] In other recent work, the PB/SA approach was extended to de novo ligand design with cyclin-dependent kinase 2 as an example target.[24] With its moderate computational expense, satisfying reliability, and broad applicability, PB/SA has become a useful tool in protein/ligand binding studies.

We investigate here whether PB/SA can be effectively combined with a quantum mechanical description of the ligand in protein/ligand binding affinity calculations. Treating the ligand quantum mechanically and the protein molecular mechanically has the advantages of (i) not requiring the derivation

* Author to whom correspondence should be addressed. Phone: +49/6221/54-8879. E-mail: stefan.fischer@iwr.uni-heidelberg.de

[†] Present address: Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Goettingen, Germany.

[‡] IWR−Computational Biochemistry.

[§] IWR−Computational Molecular Biophysics.

[||] Ecole Supérieure de Biotechnologie de Strasbourg.

10.1021/jp044185y CCC: $30.25
Published on Web 04/29/2005

of parameters for each new ligand, (ii) including quantum effects such as ligand polarization upon binding, and (iii) describing the largest part of the system classically, thus allowing fast calculations and sufficient sampling.

Both PB/SA and FEP/TI require a force field, which has typically been of the molecular mechanics (MM) type.[7−24] For nonpeptidic inhibitors, the diversity of parameters needed for a detailed description leads to using general force fields such as MMFF or GAFF.[25,26] An interesting alternative to using generic force fields is using a quantum mechanical (QM) representation of the ligand. Combined quantum mechanical/molecular mechanical (QM/MM) methods have been developed for protein calculations and are now widely used for the analysis of enzymatic reactions.[27] In the context of protein/ligand binding affinity calculations, QM techniques have so far been routinely applied only to geometry optimization of the unbound ligand and the calculation of ligand atomic charges. Electrostatic protein/ligand interactions have been calculated using the partial charges of the unbound states, assuming negligible change in the partial charges of the binding partners upon complex formation (see, for example, ref 23). However, ab initio studies also indicate that charge polarization and other quantum mechanical effects not routinely represented in MM force fields can be important in protein/ligand complexation. Examples of these are polarization and charge-transfer effects[28] and cation$-\pi$ interactions.[29] Significant polarization effects have also been found in the binding of the substrate nicotinamide adenine dinucleotide phosphate (NADPH) to dihydrofolate reductase.[30] The limitation of MM-based binding free energy calculations also becomes apparent in the poor performance of scoring functions for ligand coordination to metal ions.[31] This was emphasized in a very recent study in which ligand binding based on a QM-only strategy was reported.[32]

Whereas FEP and TI have been combined with QM for the calculation of hydration free energies[33] and tautomerisation free energies,[34] this is the first time a combined QM/MM description is incorporated into free energy calculations of protein/ligand binding. The present study aims at combining a quantum mechanical treatment of the inhibitor and classical treatment of the protein with our previously developed PB/SA free energy calculation scheme[23] and evaluating its reliability. QM/MM geometry optimization is followed by binding free energy evaluation, which involves a quantum mechanical energy component accounting for ligand strain and protein/ligand electrostatic interactions including polarization effects. To compute the (de)solvation energy with PB, the partial charges of the ligand atoms are determined by using ab initio electrostatic surface potential (ESP) fitting methods such as the CHELPG module in Gaussian94.[35] The van der Waals parameters of the ligand are assigned with simple atom-element-based rules.

Recent studies have incorporated protein flexibility into the docking procedure. These studies showed that using only a single minimum-energy structure performs poorly in comparison with methods considering an ensemble of possible conformations for a given complex.[36,37] In the latter approach, the binding free energy is then the difference of weighted averages over the free energies of the conformational ensembles in bound and unbound states. In the present study, a docking protocol based on energy minimizations was used that generates an ensemble of complex conformations. Results from docking to both rigid and flexible protein models are compared to evaluate the importance of protein flexibility in the particular model system and to assess the robustness of the QM/MM-PB/SA approach with respect to the docking protocols.

The model system used in the present work is trypsin inhibited by benzamidinium derivatives. Trypsin is selective for lysine and arginine due to binding of these side chains in the "P1" pocket of the enzyme.[38,39] Benzamidine-like ligands are a well-studied class of noncovalent trypsin inhibitors that mimic arginine and lysine by forming a salt bridge to Asp189 at the bottom of the P1 pocket.[38,11] Figure 1 shows examples of the 47 trypsin inhibitors used as ligands in the present work. These ligands were chosen because they exhibit an unambiguous protonation pattern in the unbound state and because they can be docked in only a limited number of ways. One subset (set n0) comprises six molecules of C_2-symmetry that exhibit low variability in the way they fit into the binding pocket. This subset was used in previous work.[23] The remaining 41 molecules feature one degree of freedom in the docked conformation (set n1), which is due either to their asymmetry (C_1) or to a single internal rotational degree of freedom.

Ideally, a binding scoring function should work with proteins other than that used to develop it. Therefore, the present QM/MM-PB/SA approach was subsequently employed to calculate the absolute free energies of three small ligands binding to the FK506 binding protein (FKBP).[40] The accuracy obtained for both systems is satisfying and suggests that the presented QM/MM-PB/SA method is suitable for absolute binding free energy calculations.

Methods

Setup for Trypsin. The crystal structure for the trypsin/benzamidine complex with a resolution of 1.5 Å (PDB code 1BTY[41]) was used for the calculations. Surface water molecules were removed. Buried structural water molecules (including five in the binding site) were kept. Previous calculations showed that the protonation state of His57 does not affect binding free energy results.[23] Here, the unprotonated form was chosen. During geometry optimization, the five crystal water molecules located in the binding pocket and protein atoms within a distance of 9 Å from the ligand center were allowed to move, while the rest of the protein was fixed (185 moving atoms out of a total of 2220). The binding pocket was relaxed in the presence of benzamidine using a combined quantum mechanical/molecular mechanical (QM/MM) force field. The protein and water molecules were modeled classically using the CHARMM force field with parameter set 19,[42] while the ligand was treated with the semiempirical QM method AM1[43] as interfaced with CHARMM version c28.[44,45] AM1 has proven useful for the calculation of hydration and ligand binding free energies and polarization effects in a number of studies, including polarizable ionic solutes such as NADPH or carboxylates.[29,30,32,46] Missing van der Waals parameters for ligands with covalently attached chlorine and fluorine were adapted from the MAB force field[47] resulting in $r_{min}(\mathrm{Cl}) = 1.57$ Å, $E_{min}(\mathrm{Cl}) = -0.20$ kcal/mol, $r_{min}(\mathrm{F}) = 1.32$ Å, and $E_{min}(\mathrm{F}) = -0.12$ kcal/mol. The CHARMM parameters of sulfur in cysteine were used for the sulfur in thiophene.

Docking to Trypsin. To account for the electrostatic screening by the solvent during geometry optimization of the complex, nonuniform charge scaling (NUCS) was employed.[48] In this method, the solvent screening is approximated by dividing the partial atomic charges of the protein by scaling factors that are derived for the side chain and backbone of each residue. The scaling factors are computed such that the Coulomb interaction of a given group approximates the solvated interaction. To do

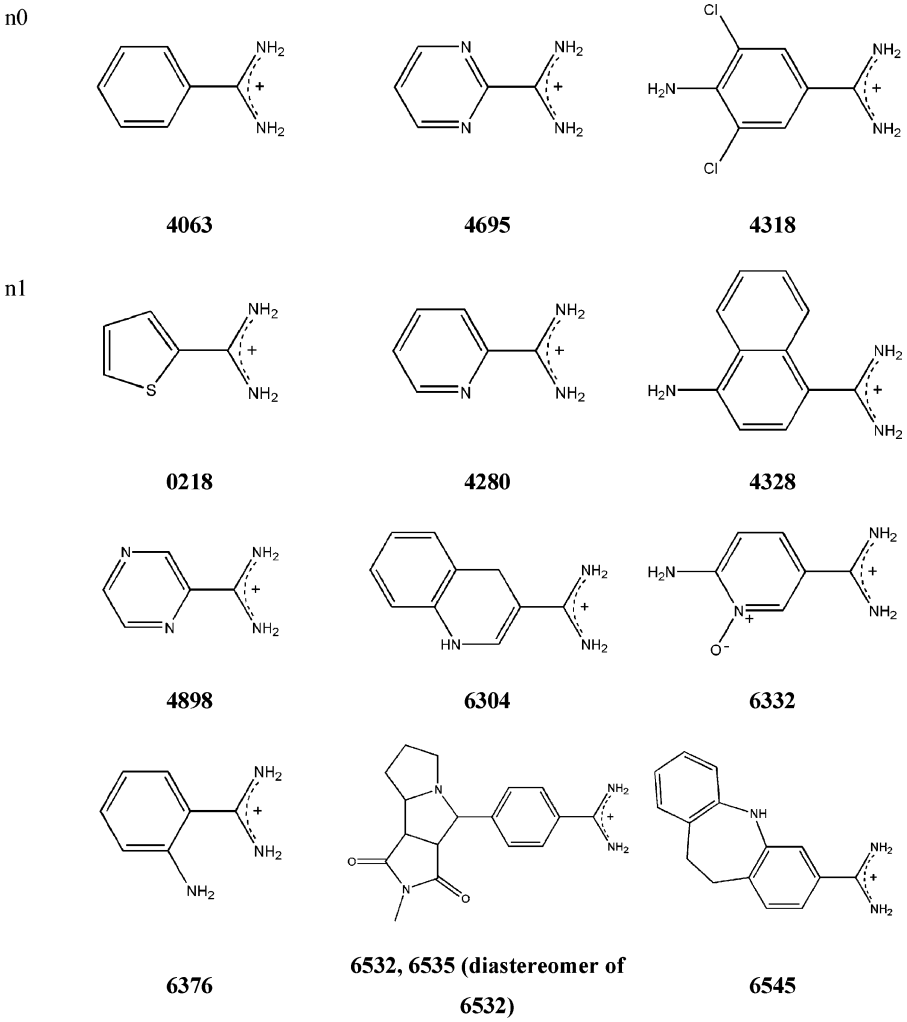

Figure 1. Structures of representative trypsin ligands (listed in Tables 1 and 2).

this, the Coulomb electrostatic interaction energies were calculated with a dielectric constant of 4. The solvated interactions were calculated by approximating the solvent by a high dielectric ($\epsilon_{solvent} = 80$, ionic strength = 100 mM) and the protein interior with $\epsilon_{prot} = 4$. The Poisson−Boltzmann equation was solved using the finite difference method (grid spacing = 0.25 Å), as implemented in the PBEQ module in CHARMM.[49] During minimization, the nonbonded interactions were smoothly brought to zero by multiplying by a cubic switching function[42] between 6 and 12 Å. The cutoff and NUCS scheme were used only during geometry optimization to approximate solvent screening for structure generation. For scoring, unscaled partial charges and no cutoff were used (see below).

A conformational ensemble of protein/ligand structures was obtained using the following docking procedure. The ligand in its cationic state was placed in the P1 pocket with its benzamidinium group located where benzamidine is found in the crystal structure. Then, the ligand was moved using a combination of ±1.5 Å translations and ±30° rotations along three orthogonal axes. Structures with steric clashes between the ligand and protein were discarded. The remaining conformations of the complex (up to 70) were energy-minimized. Those with a potential energy lower than 25 kcal/mol above the lowest energy were kept. Structures with energies deviating from each other by <0.002 kcal/mol were grouped together, and only one representative was kept. For the unbound states of trypsin and the ligands, a single conformation was chosen, obtained by energy minimization using MM or QM, respectively. To compare the influence of different docking approaches, docking was also carried out with a rigid protein with a procedure identical to that described above except that all protein atoms were kept fixed during minimization. In the rigid protein case, a single complex conformation of lowest potential energy was kept.

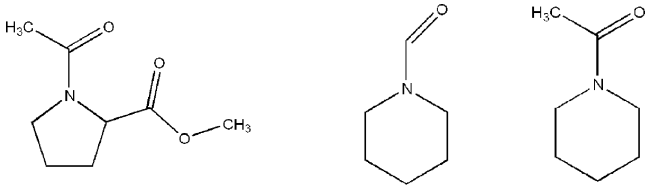

Figure 2. Structures of the FKBP12 ligands (ACPM, FOPI, PICA).

Docking to FKBP. The three FKBP ligands, ACPM, FOPI, and PICA, are shown in Figure 2. Ensembles for the structure of the FKBP/ligand complexes were taken from the complexes that have been previously obtained in ref 40 by MCSS (multiple copy simultaneous search).[50] The ensembles consisted of 3 (ACPM), 8 (FOPI), and 16 (PICA) conformations, all satisfying the experimental nuclear Overhauser effect (NOE) constraints.[40] Missing hydrogens on the ligands were placed with the SYBYL 6.8 package.[51] The complexes were then geometry-optimized using a distance-dependent dielectric permittivity ($\epsilon = r$) in combination with a dielectric constant of 4 and keeping all protein atoms fixed (to be consistent with the MCSS protocol that was used in ref 40). The FKBP ligands have internal degrees

of freedom, such that an ensemble of ligand structures in the unbound state was generated by energy minimization starting from the bound conformations. The unbound protein was kept in the same conformation as in the complex.

QM/MM-PB/SA Formalism. The free energy is computed separately for the ligand L, the protein P, and the ligand/protein complex C. For each conformation *i* of the species X ∈ {P,L,C} the free energy, $G(X_i)$ is decomposed here empirically as a sum of terms, (i) intramolecular energies E_{int}, (ii) the solvation free energy G_{solv}, and (iii) an ideal-gas correction G_{ideal}. This assumes that the free energy is decomposable, i.e., that the contributing effects are additive. Although the terms are in principle nonadditive in that the free energy components are coupled, decomposition has proven useful in free energy analysis.[52,53] The decomposition is now detailed for L, P, and C.

Unbound Ligand Free Energy $G(L_i)$. As the ligand is treated quantum mechanically, $E_{int}(L) = E_{QM}(L)$. For the trypsin ligands, only a single conformation of the unbound ligand is used, i.e., the subscript *i* can be dropped. The ligand free energy is then written as

$$G(L_i) = E_{QM}(L_i^u) + G_{solv}(L_i^u) + G_{ideal}(L_i^u) \quad (1)$$

where the superscript u indicates the unbound conformation of the ligand.

Unbound Protein Free Energy G(P). For the unbound protein, a single conformation is used. $E_{int}(P)$ consists of three terms corresponding to the bonded (valence) interactions, $E_{val}(P)$, Coulombic interactions, $E_{Coul}(P)$, and van der Waals interactions, $E_{vdW}(P)$. Thus

$$G(P) = E_{val}(P^u) + E_{Coul}(P^u) + E_{vdW}(P^u) + G_{solv}(P^u) + G_{ideal}(P^u) \quad (2)$$

Complex Free Energy $G(C_i)$. $G(C_i)$ is the sum of the energy terms corresponding to the protein P and ligand L in the bound state and of the protein/ligand interactions (P/L)

$$G(C_i) = E_{val}(P_i^b) + E_{Coul}(P_i^b) + E_{vdW}(P_i^b) + \alpha E_{vdW}(P_i^b/L_i^b) + E_{QM}(L_i^b) + G_{solv}(C_i) + G_{ideal}(C_i) \quad (3)$$

The superscript b indicates that the bound conformation is used. The protein/ligand electrostatic interactions are computed using quantum mechanics (AM1) and are contained in $E_{QM}(L_i^b)$, while the van der Waals interactions ($E_{vdW}(P_i^b/L_i^b)$) are computed classically. α is an empirical scaling factor introduced to balance the van der Waals energy relative to the electrostatics (since the van der Waals parameters of the ligand were not determined self-consistently with other ligand properties). α is obtained by fitting the calculated ΔG^0 (eq 7, see below) to the experimental results and is the only fitting parameter used here. α was separately calculated for the flexible and rigid protein approaches, since the relative magnitude of the van der Waals and electrostatic energy contributions and thus the imbalance between the van der Waals and electrostatic energies depend on the docking protocol. The energy terms in eqs 1−3 are calculated as follows.

Valence Energy $E_{val}(P)$. The valence energy accounts for bonded strain in the protein and is calculated with the standard CHARMM force field as the sum of the bond stretching, angle bending, torsion angle, out-of-plane bending, and Urey−Bradley terms.[42] The bonded strain of the ligand is implicitly included in $E_{QM}(L)$.

van der Waals Energy $E_{vdW}(P/L)$ and $E_{vdW}(P)$. The van der Waals energy accounts for shape complementarity between the protein and the ligand and for steric interactions in the protein. The van der Waals terms are calculated directly after geometry optimization using the Lennard-Jones formula as implemented in CHARMM. The van der Waals contributions within the ligand are implicitly included in $E_{QM}(L)$.

Coulomb Energy $E_{Coul}(P)$. The electrostatic interactions within the protein in the bound and unbound states are calculated with Coulomb's law using standard CHARMM atomic partial charges. To be consistent with the solvation energy, it is important to use the same dielectric constant as is used for the protein interior when computing the solvation term (see below) and to use no cutoff. Thus, $E_{Coul}(P)$ was calculated with a single energy call using $\epsilon_i = 4$ and no cutoff, without reoptimizing the geometry. Electrostatic interactions within the ligand and the electrostatic protein/ligand interactions are included in $E_{QM}(L)$.

Quantum Energy $E_{QM}(L)$. The semiempirical AM1 method was chosen for the ligand, as a compromise between speed and accuracy. AM1 has been used successfully in a number of QM/MM studies on enzymatic reactions[54] and in a recent QM-only study of ligand binding.[32] The treatment of the interactions between the QM and MM regions has been described previously.[45] In the unbound state, $E_{QM}(L^u)$ corresponds to the internal energy of the ligand, $E_{int}(L^u)$. In the bound state, $E_{QM}(L^b)$ contains in addition the electrostatic protein/ligand interaction. Both unbound and bound $E_{QM}(L)$ are computed with a dielectric constant of $\epsilon = 1$. For consistency with the electrostatic solvation term, $E_{QM}(L^b)$ is taken from a single energy call without employing a nonbonded cutoff (this must be set explicitly in CHARMM for AM1) and with the partial atomic charges of the protein divided by the dielectric constant ϵ_i assigned to the protein interior (here $\epsilon_i = 4$). The latter allows the P/L electrostatic interactions to be approximated, which would have been obtained by setting $\epsilon = \epsilon_i$. A minor inconsistency remains for the electrostatics of the ligand, where the Coulomb term of the quantum Hamiltonian is computed with $\epsilon = 1$, but the interactions within the ligand are screened by the solvation term with $\epsilon = \epsilon_i$. Note that in the protein/ligand complex the electronic polarization of the ligand is explicitly calculated in the QM/MM procedure, and accordingly no internal dielectric is needed for the ligand. However, for the solvation of the ligand, the polarization of the ligand by the solvent is computed classically by using an internal dielectric of 4 for the ligand. The polarization of the ligand in solution and in the protein complex are thus not calculated with identical procedures. However, this should not introduce large errors for the small ligands considered here. A possible alternative would be to use a self-consistent reaction field instead of the classical Poisson−Boltzmann procedure for the solvation of the ligand. This will be the subject of later studies. However, the associated error is expected to be small for small ligands. The electron density of the ligand is polarized in response to charges in the environment. Thus, polarization effects are implicitly included.

Solvation Free Energy $G_{solv}(X)$. The solvation energy is split into a polar contribution $E_{solv,p}(X)$ and a nonpolar contribution $E_{solv,np}(X)$. $E_{solv,p}(X)$ is the electrostatic contribution to the solvation free energy and is calculated for each conformation of solute X as the change in the total electrostatic energies between the solvated and desolvated states

$$E_{solv,p}(X) = E_{el}(X,solv) - E_{el}(X,desolv) \quad (4)$$

$E_{el}(X)$ was computed using the linearized Poisson−Boltzmann

equation that relates the charge density, $\rho(\mathbf{r})$, to the electrostatic potential, $\Phi(\mathbf{r})$, in a medium with nonuniform dielectric permittivity, $\epsilon(\mathbf{r})$

$$\nabla\epsilon(\mathbf{r})\nabla\Phi(\mathbf{r}) - \kappa'\Phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}) \quad (5)$$

where κ' is the Debye−Hückel screening constant, which is a function of the ionic strength. A numerical solution of eq 5 using a finite difference scheme was calculated with the program UHBD.[55,56] A focusing approach with a final grid spacing of 0.25 Å was used. Ionic strengths of 100 and 0 mM were used for E_{el}(X,solv) and E_{el}(X,desolv), respectively. For E_{el}(X,solv), $\epsilon(r) = 80$ in the solvent regions and $\epsilon(r) = \epsilon_i$ in the solute regions were employed. The value used for ϵ_i was consistent with that used for E_{Coul}(P) and E_{QM}(L), i.e., $\epsilon_i = 4$. For E_{el}(X, desolv), $\epsilon(r) = \epsilon_i$ everywhere. The required partial atomic charges of the ligand were determined by ab initio electrostatic surface potential (ESP) calculations (HF/6-31G*//HF/3-21G* for trypsin ligands and HF/6-31G*//HF/6-31G* for FKBP12 ligands), using the module CHELPG in the Gaussian94 software package.[35]

When the QM/MM term for the ligand/protein interactions are combined with the purely classical electrostatic solvation term, some contributions to the binding free energy are neglected or not treated here in a fully consistent way, such as changes in the polarization of the ligand by the solvent, or the classical estimate of the ligand desolvation with a charge distribution that differs from what would be obtained with the AM1 method used in the QM/MM part. However, these approximations are not inherent to the QM/MM-PB/SA approach and could be addressed in future implementations.

$E_{solv,np}$(X) accounts for the unfavorable cavity formation and favorable van der Waals interactions between solute atoms and the solvent. Both these effects are assumed to depend linearly on the solvent-accessible surface area, A, and the term is thus written as

$$E_{solv,np}(X) = \gamma A(X) \quad (6)$$

where γ is an empirical coefficient that has been estimated to be 25 cal/(mol/Å^2),[57,58] the value used here. The surface area of the solutes was computed with the program CODISP,[59] which uses the Connolly algorithm[60] (probe radius 1.4 Å).

Ideal-Gas Contributions G_{ideal}. This arises from the change in translational, rotational, and vibrational degrees of freedom in the system upon binding, in particular, the creation of six additional vibrational modes in the complex. The importance of including G_{ideal} in calculations of absolute and relative binding free energies was shown in ref 23, in which the method used for calculating G_{ideal} is described. The vibrational contribution depends on the normal-mode frequencies, which were obtained here from normal-mode analyses.

Binding Free Energy. The standard binding free energy is the difference

$$\Delta G^0 = G(C) - G(L) - G(P) \quad (7)$$

For multiple conformations of the species X, the free energy is determined as the Boltzmann ensemble average

$$G(X) = \sum_i f_i G(X_i) \quad (8)$$

where f_i is the Boltzmann factor corresponding to the ith conformation and $G(X_i)$ is its corresponding free energy. f_i is determined according to

$$f_i = \frac{e^{-G(X_i)/RT}}{\sum_j e^{-G(X_j)/RT}} \quad (9)$$

where R is the universal gas constant, T is the absolute temperature, and the index j runs over all conformations. A similar formulation has been used in a recent study using the linear interaction energy (LIE) approach.[12] Note that using the definition of the Gibbs free energy in statistical thermodynamics, $G(X) = -RT \ln \sum_i e^{-G(X_i)/RT}$ computed over the ensemble of all states X_i, would not be appropriate here for two reasons. First, our docking protocol does not statistically sample the conformational space of the complex. Second, entropy contributions are estimated in the PB/SA scheme from the ideal-gas vibrational free energy instead of being calculated from a statistical ensemble of states.

For further insight into the contribution of each energy term to the binding free energy, the change in any term E_j is calculated as the Boltzmann-weighted average according to

$$\Delta E_j = \sum_{i=1}^{N_C} f_i E_j(C_i) - \sum_{i=1}^{N_L} f_i E_j(L_i) - E_j(P) = \langle E_j(C)\rangle - \langle E_j(L)\rangle - E_j(P) \quad (10)$$

where N_C is the number of complex conformations and N_L the number of unbound ligand conformations. f_i is the Boltzmann factor, computed from the free energies $G(C_i)$ of the complex (eq 9) for the first sum and from the free energies of the ligand $G(L_i)$ for the second sum. In the case of a single ligand conformation (i.e., for the trypsin ligands), $\sum_{i=1}^{N_L} f_i E_j(L_i)$ simplifies to $E_j(L)$. After all of the above equations are combined, the empirical binding free energy then becomes

$$\Delta G_{bind}{}^0 = \Delta E_{int}(P) + \alpha\Delta E_{vdW}(P/L) + \Delta E_{QM}(L) + \Delta G_{solv} + \Delta G_{ideal} \quad (11)$$

For all calculations, i.e., docking, energy minimizations, and evaluation of binding free energies, CHARMM scripts have been developed. They are easy to modify for any protein of interest and are available for download at http://www.iwr.uni-heidelberg.de/groups/biocomp/fischer.

Measurement of the Binding Constants. The experimental dissociation constants, K_d, for the trypsin/ligand complexes were measured spectrophotometrically, as described in ref 23. The experimental binding free energy is given by $\Delta G_{bind} = RT \ln K_d$. The binding constants for the FKBP12/ligand complexes are taken from ref 40.

Results

Trypsin: Flexible Protein Approach. Docking resulted in up to seven different conformations per protein/ligand complex, structurally deviating from each other by a root-mean-square (RMS) coordinate difference over all protein and ligand atoms of 0.1−1.5 Å depending on the ligand. The six C_2-symmetric ligands (set n0) form 2.0 different complexes on average, whereas the remaining 41 ligands (set n1) form 3.6 on average, reflecting the larger conformational variability of the latter group. As an illustration, two conformations found for *N*-oxo-pyridine-2-carboxamidine (ligand **6332**, set n1) are shown in Figure 3. When the *N*-oxo group is directed away from Ser195, the catalytic triad (Ser195, His57, Asp102) exhibits the same

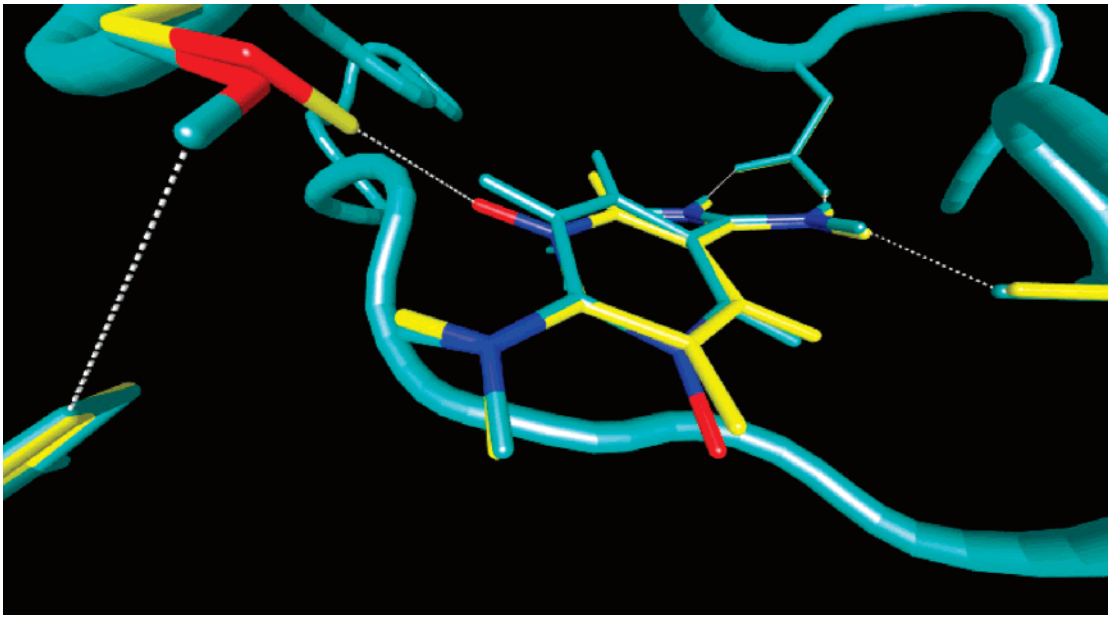

Figure 3. Overlap of two representative structures of *N*-oxo-pyridine-2-carboxamidine (ligand **6332**) bound to trypsin, showing an intact catalytic triad (cyan) or a ligand−Ser195 hydrogen bond (yellow), respectively.

TABLE 1: Energy Terms in kcal/mol for Binding to Flexible Trypsin[a]

set	ligand	ΔG_{exp}	ΔG_{calc}	$\Delta E_{vdW}(P/L)$[b]	$\Delta E_{int}(P)$[c]	$\Delta E_{QM}(L)$	$\Delta G_{solv,p}$	$\Delta G_{solv,np}$	ΔG_{tr}	ΔG_{v}
n0	**4063**	−6.32	−6.42	−22.47	6.67	6.58	−9.19	−5.96	16.50	1.45
	4695	−4.73	−7.03	−23.08	8.72	5.18	−8.89	−5.62	16.49	0.17
n1	**0218**	−6.91	−4.58	−19.58	9.61	3.71	−9.96	−5.46	16.89	0.22
	4280	−5.85	−6.66	−22.73	10.00	4.89	−9.44	−5.67	16.90	−0.60
	4328	−5.07	−5.62	−27.31	9.95	9.69	−9.99	−7.38	18.11	1.31
	4898	−4.82	−6.53	−23.20	9.60	3.98	−8.74	−5.62	16.91	0.55
	6304	−5.88	−8.09	−26.73	7.06	9.27	−11.38	−6.87	17.93	2.64
	6376	−5.01	−5.61	−23.69	8.53	7.93	−9.85	−5.98	17.22	0.24
	6532	−5.14	−3.14	−31.12	10.13	8.96	−5.91	−9.82	19.54	5.08
	6545	−5.67	−7.86	−29.70	8.01	7.65	−7.19	−8.48	18.82	3.03

[a] Each term is Boltzmann-weighted according to eq 10. Only ligands with extreme energy contributions are shown here (see Table S1 for all ligands). [b] $\Delta E_{vdw}(P/L)$ is scaled by $\alpha = 0.84$. [c] $\Delta E_{int}(P) = \Delta E_{vdW}(P) + \Delta E_{Coul}(P) + \Delta E_{val}(P)$.

hydrogen bonding pattern as in uncomplexed trypsin. In contrast, when the ligand is rotated by 180°, the hydroxyl group of Ser195 rotates to form a hydrogen bond with the oxygen atom of the *N*-oxo group. Both states are significantly occupied ($\Delta\Delta G_C$ = 0.5 kcal/mol, $f_i/f_j \approx 2{:}1$). Thus, the docking results suggest that even for benzaminidine-like ligands that are rigid more than one structure may exist for the bound state. Varying starting structures of the ligands in the unbound state converged to a single energy minimum during optimization with energy differences below 0.1 kcal/mol. Therefore, only a single structure was used to represent the unbound ligand state.

Table 1 lists the calculated binding free energies for those ligands that exhibit extreme values of some of the contributing terms (see Table S1 in the Supporting Information for the results for all trypsin ligands). The magnitude and range of the calculated binding energies (−8.1 to −3.1 kcal/mol) are close to those of the experimental values (−7.6 to −3.9 kcal/mol). Given this relatively small range of binding free energies and the large number of ligands, one does not expect to obtain a correct ranking for all ligands. Therefore, the correlation between calculated and experimental values is poor (correlation coefficient = 0.20), as can be seen in Figure 4. However, the RMS error between all calculated and experimental values is low at 1.21 kcal/mol, which is good, considering that these are absolute rather than relative binding energies.

Much can be learned about the factors driving binding to trypsin by dissecting the binding free energy into contributing terms. The van der Waals energy, ΔE_{vdW}, is found to be the main energy term favoring binding. With the most favorable van der Waals energy of −31.1 kcal/mol, ligand **6532** achieves the highest shape complementary among the set of ligands. As expected, less favorable van der Waals contributions are found for smaller ligands (ΔE_{vdW}(**0218**) = −19.6 kcal/mol). ΔE_{vdW} spans a range of 11 kcal/mol, which is significantly larger than the maximum relative binding free energy $\Delta\Delta G_{exp}^{max} = 3.7$ kcal/mol. Thus, the protein/ligand van der Waals interaction contributes here to both absolute and relative binding affinities. The polar and nonpolar contributions to the solvation free energy also drive binding, with $\Delta G_{solv,p}$ ranging from −11.4 to −5.9 kcal/mol (ligands **6304** and **6532**) and $\Delta G_{solv,np}$ varying from −9.8 to −5.5 kcal/mol (ligands **6532** and **0218**). Those ligands with relatively low solvent-accessible surface areas in the unbound state, such as those with pyrimidine as the aromatic moiety, possess less favorable nonpolar contributions to the solvation free energy, e.g., ligands **4695** and **4898** for both of which $\Delta G_{solv,np}$ is −5.6 kcal/mol. Ligand **6532**, carrying a tricyclic substituent at the para position, yields the most favorable $\Delta G_{solv,np}$ of −9.8 kcal/mol.

The protein intramolecular interactions are unfavorable for binding; their averages over all ligands are $\Delta E_{val}(P) = 0.8$ kcal/mol, $\Delta E_{Coul}(P) = 8.2$ kcal/mol, and $\Delta E_{vdW}(P) = 0.1$ kcal/mol. Hence, binding involves an increase in the protein intramolecular energy. All these terms vary by more than $\Delta\Delta G_{exp}^{max}$ (Table 1)

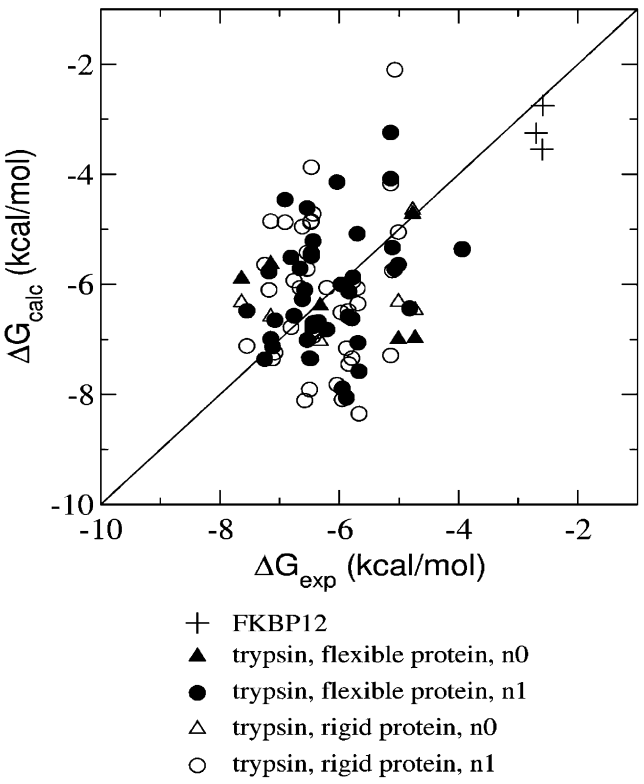

Figure 4. Calculated versus experimental binding free energies. Filled symbols are used for trypsin ligands with a flexible protein (data of Table S1). Open symbols are used for trypsin ligands with a rigid protein (data of Table S2). Crosses are used for the FKBP12 ligands (data of Table 3).

and consequently cannot be neglected. Whereas the bulky candidate **6545** induces protein internal strain upon binding, as reflected by an increase in $\Delta E_{val}(P)$ by 3.5 kcal/mol, binding of other sterically less-demanding ligands has a negligible effect on $E_{val}(P)$. An example of the latter is ligand **0218** for which $E_{val}(P)$ is -0.1 kcal/mol, but which, however, leads to an increase in the protein electrostatic energy, $\Delta E_{Coul}(P)$, of 10.4 kcal/mol.

Entropic and enthalpic contributions due to changes in the translational, rotational, and vibrational degrees of freedom must be taken into account in binding free energy calculations as indicated by the broad range of ΔG_{ideal} from 16.3 to 24.6 kcal/mol. Thus, this study confirms the importance of these terms in ligand binding. ΔG_{vib} is mostly positive (-0.6 to 5.1 kcal/mol). This is due to an increase in the vibrational zero-point energy ($\Delta H_{vib} = 4.8$ kcal/mol on average) that exceeds the increase in vibrational entropy ($-T\Delta S_{vib} = -3.4$ kcal/mol on average). Whether the latter is due to an increase in protein flexibility or simply the introduction of the six libration modes of the bound ligand could be dissected as previously described[61] and remains to be investigated.

The quantum energy term ΔE_{QM} ranges from 3.7 to 9.7 kcal/mol. ΔE_{QM} comprises the change in ligand intramolecular energy $\Delta E_{int}(L)$ and the electrostatic interaction between protein and ligand $\langle E_{Coul}(P/L)\rangle$

$$\Delta E_{QM} = \langle E_{QM}(L^b)\rangle - \langle E_{QM}(L^u)\rangle = \Delta E_{int}(L) + \langle E_{Coul}(P/L)\rangle \quad (12)$$

where $\langle\cdots\rangle$ denotes the average over the conformations of the complex according eq 10. The decomposition into $\Delta E_{int}(L)$ and $\langle E_{Coul}(P/L)\rangle$ is made possible by considering the binding as a two-step process of (i) conformational straining of the ligand and (ii) rigid binding of the strained ligand to the receptor. By the dissection of ΔE_{QM} according to eq 12, further insight into the factors governing binding can be obtained. $\Delta E_{int}(L)$ is the difference between the quantum mechanical energies of the ligand in the bound (strained) state (calculated without the protein environment) and in the unbound (relaxed) state, $\Delta E_{int}(L) = \langle E_{QM}(L^b,\text{noprot})\rangle - E_{QM}(L^u)$. The energy of the ligand in the bound conformation but in the absence of the protein is $E_{QM}(L_i^b,\text{noprot})$. The ligands are found here moderately strained, with $\Delta E_{int}(L) = 2.1$ kcal/mol on average. The electrostatic and polarization interaction between the protein and the ligands $\langle E_{Coul}(P/L)\rangle$ is 4.4 kcal/mol on average. This large unfavorable contribution is surprising at first glance, since the formation of the salt bridge with Asp189 is electrostatically favorable. However, $\langle E_{Coul}(P/L)\rangle$ is dominated by the unfavorable interaction of trypsin's net charge of $+8$ with the cationic trypsin inhibitors. $E_{Coul}(P/L)$ and $\Delta G_{solv,p}$ are of opposite sign as expected, since a large contribution to $\Delta G_{solv,p}$ is the screening of the electrostatic interactions between the protein and the ligand. The lowest ΔE_{QM} is found for ligand **0218** ($\Delta E_{QM} = 3.7$ kcal/mol), a compact thiophene derivative whose small size allows it to orient itself in the narrow P1 pocket to optimize the salt bridge with Asp189. In contrast, the benzamidinium moiety of ligand **4328** has an ortho-substituted benzyl group that prevents the ligand from fitting deep into the P1 pocket, yielding the other extreme (ΔE_{QM}(**4328**) = 9.7 kcal/mol).

A 48th ligand (**6535**, shown in Figure 1) could not be adequately treated. Its calculated binding free energy differs by 4.4 kcal/mol from the experimental value. This high inaccuracy was traced back to a particularly high increase in the quantum mechanical energy ($\Delta E_{QM} = 10.8$ kcal/mol). This reflects an exaggerated amount of conformational strain in the bound ligand and indicates that a correct structure for the complex could not be found. This suggests that excessive ligand strain may be a useful criterion for identifying potentially poorly docked ligands.

Trypsin: Rigid Protein Approach. The binding free energies calculated using a single complex conformation, keeping all protein atoms fixed, and after refitting the empirical scaling parameter ($\alpha = 0.61$) are given in Table 2 (see Table S2 of the Supporting Information for the results for all trypsin ligands) and are plotted versus the experimental values in Figure 4. The observed RMSD of those calculated from the experimental binding free energies for the set of 47 ligands is 1.29 kcal/mol (correlation coefficient = 0.30). Therefore, at least in the case of the relatively rigid system of trypsin and benzamidine derivatives, the fixed protein approximation does not lead to a significant loss of accuracy in the binding affinity calculations. Although, in the flexible protein calculations, several bound conformations were found for some ligands; the inclusion of these conformations does not clearly improve the results.

Because in the rigid protein case the protein cannot adapt to the ligands, the van der Waals energy term $E_{vdW}(P/L)$ is less favorable than when the protein is flexible, ranging from -21.6 to -14.3 kcal/mol. A significant difference can be also found for the vibrational free energy contribution, which is mostly favorable with the protein rigid ($\Delta G_{vib} = -3.6$ to 0.9 kcal/mol), in contrast to when the protein is flexible. However, ΔG_{vib} for the rigid and flexible protein correlate well with each other (correlation coefficient = 0.8, data not shown), i.e., the size of the flexible region does not significantly alter the relative vibrational free energy contributions. Contributions from intramolecular terms, electrostatic interactions, and solvation

TABLE 2: Energy Terms in kcal/mol for Binding to Rigid Trypsin[a]

set	ligand	ΔG_{exp}	ΔG_{calc}	$\Delta E_{vdW}(P/L)^b$	$\Delta E_{QM}(L)$	$\Delta G_{solv,p}$	$\Delta G_{solv,np}$	ΔG_{tr}	ΔG_v
n0	**4063**	−6.32	−7.03	−16.63	8.67	−6.43	−6.07	16.50	−3.07
	4695	−4.73	−6.47	−16.94	9.24	−5.98	−5.76	16.49	−3.52
n1	**0218**	−6.91	−4.87	−14.24	8.52	−6.97	−5.51	16.89	−3.56
	4280	−5.85	−6.48	−16.66	8.83	−6.52	−5.86	16.91	−3.18
	4328	−5.07	−2.10	−16.49	10.88	−6.20	−7.25	18.11	−1.15
	4898	−4.82	−6.43	−16.98	8.65	−5.80	−5.76	16.91	−3.45
	6304	−5.88	−7.16	−18.95	9.07	−5.61	−7.05	17.83	−2.45
	6376	−5.01	−3.87	−16.09	9.20	−5.24	−6.03	17.79	−3.50
	6532	−5.14	−4.17	−18.24	8.41	−3.87	−9.74	18.41	0.85
	6545	−5.67	−8.35	−21.42	8.45	−5.15	−8.50	19.27	−1.02

[a] Each term is Boltzmann-weighted according to eq 10. The same ligands are shown as in Table 1 (see Table S2 for all ligands). [b] $\Delta E_{vdw}(P/L)$ is scaled by $\alpha = 0.61$.

TABLE 3: Energy Terms in kcal/mol for Binding of FKBP Ligands

ligand	ΔG_{exp}	ΔG_{calc}	$\Delta E_{vdW}(P/L)^b$	$\Delta E_{QM}(L)$	$\Delta G_{solv,p}$	$\Delta G_{solv,np}$	ΔG_{tr}	ΔG_v
ACPM	−2.59	−3.56	−15.18	−0.87	4.44	−7.34	−2.39	17.79
FOPI	−2.59	−2.76	−11.55	−2.29	3.98	−5.40	−4.15	16.65
PICA	−2.70	−3.26	−12.98	−1.89	4.46	−5.67	−4.15	16.96

[a] Each term is Boltzmann-weighted according to eq 10. [b] $\Delta E_{vdw}(P/L)$ is scaled by $\alpha = 0.61$.

effects differ insignificantly from those found with the flexible protein approach.

To investigate how well the hybrid QM/MM treatment performs relative to a pure MM treatment of the protein and ligand, calculations assuming the protein as rigid were performed using the MAB force field[47] (which autogenerates MM parameters for each ligand) and the modeling software MOLOC,[62] as previously described.[23] While this approach proved successful in the prediction of binding affinities of 6 small and rigid ligands (RMSD = 0.55 kcal/mol), it gives significantly worse results for the 47 trypsin ligands used here (RMSD = 1.85 kcal/mol, $\alpha = 0.81$, data not shown). Differences in the energy contributions between the AM1/CHARMM and the MAB force field are mainly in terms involving QM treatment (electrostatic P/L interaction and vibrational free energy).

FKBP12. While the three FKBP12 ligands have essentially identical binding affinities (−2.7 to −2.6 kcal/mol), they differ significantly from the binding affinities of the trypsin ligands. By using the same QM/MM-PB/SA setup as for rigid trypsin, the RMSD of calculated versus experimental binding free energies is 0.66 kcal/mol when the ligands are allowed to be flexible and 1.45 kcal/mol when the ligands are kept rigid. To test the transferability of the binding free energy calculation to other protein/ligand systems, the scaling factor for the van der Waals protein/ligand interaction energy was not refitted. Instead, the value obtained for trypsin/benzamidine assuming the protein as rigid has been used ($\alpha = 0.61$). The energy contributions and total binding free energies corresponding to a flexible ligand treatment are listed in Table 3. Unlike trypsin, the FKBP ligands are neutral, such that there is no global repulsion between FKBP and the ligands. Therefore, the electrostatic contributions are of the opposite sign as compared to those of the trypsin ligands, and the protein/ligand electrostatic interaction is favorable ($\Delta E_{QM}(L)$ ranges from −0.9 to −2.3 kcal/mol). Accordingly, in the absence of screening of repulsive P/L interactions, the electrostatic contribution to the solvation free energy ($\Delta E_{solv,p}$ = 4.3 kcal/mol on average) has a positive sign, reflecting the unfavorable burying of the polar moieties.

Discussion

In the present study, the suitability of combining a quantum mechanical/molecular mechanical force field with Poisson−Boltzmann/surface area terms for ligand binding calculations has been evaluated. The general usefulness of a method can be tested by varying the target protein. As a first step in this direction, calculations have been performed for three ligands binding to the protein FKBP, yielding an RMSD of only 0.66 kcal/mol. One possible reason for the smaller error for the FKBP complexes compared to the trypsin system may be due to the fact that the bound FKBP/ligand structures were taken from experimental results rather than obtained by docking. This would then suggest that, once reliable structures are at hand, binding affinities can be predicted more accurately using the QM/MM-PB/SA approach. Moreover, the FKBP results indicate that the QM/MM-PB/SA setup and empirical parameters are transferable to other protein/ligand systems. The present method requires ∼30 min CPU time per ligand on a single 1.2 GHz processor. Consequently, with a cluster of PCs it is possible to perform high-throughput ligand searches for a given target.

Through the use of only the 47 trypsin ligands with their narrow range of binding affinities, the correlation of calculated and experimental binding free energies is low (correlation coefficient = 0.20). However, augmenting the data set by the three additional FKBP12 complexes significantly improves the correlation (correlation coefficient = 0.56). Thus, the correlation between experimental and calculated binding affinities is higher for the QM/MM-PB/SA approach than for the empirical scoring functions that exhibit correlation coefficients of ∼0.5 for a much broader range of binding affinities.[4−6] To assess the reliability of the structure prediction, calculations were performed with both flexible and rigid docking. There is a small decrease in the RMSD on including conformational diversity, from 1.29 kcal/mol (rigid) to 1.18 kcal/mol (flexible). However, the minimization-based docking protocol used here is still rather crude, even when including protein flexibility.

Through the use of the present approach, the absolute binding free energies determined for 47 trypsin inhibitors exhibit a root-mean-square deviation from experiment of only 1.21 kcal/mol. The level of accuracy obtained with the present method is thus comparable to the accuracy expected for much more expensive MD-based free energy perturbation or thermodynamic integration schemes and is similar to conventional MM-PB/SA methods.[19−21] The combined QM/MM approach clearly outperforms the less computationally expensive purely MM-based

approach using the MAB force field (RMSD 1.85 kcal/mol, see Results). This level of accuracy is encouraging, since in contrast to the purely MM-based methods that have been extensively developed and refined over the past years the approach presented here is a first attempt to combine continuum solvent calculations with hybrid QM/MM calculations and leaves considerable scope for improvement. The difference between calculated and experimental results can be expected to be partially caused by the docking protocol, which does not fully sample the conformational space of the complex. Docking might be improved by using a side chain rotamer library for the target protein or docking schemes such as Dock,[63] FlexE,[64] or GOLD.[65]An alternative approach is to use snapshots from molecular dynamics simulations as coordinates, a method that has yielded binding free energies to within ∼1 kcal/mol using the MM-PB/SA method, i.e., without QM ligands.[19,20] Other reasons for the disagreement with experiment might arise from the approximations involved in combining a QM/MM treatment for the protein/ligand interactions with a purely classical treatment of solvation, from approximations inherent to the AM1 method, and from the lack of a protein polarization term when using charged ligands. In addition to optimizing docking, the accuracy of the QM/MM-PB/SA approach may be improved by optimizing the implicit solvent model and by using higher-level ab initio quantum methods.

The present study has shown that the combined quantum mechanical/molecular mechanical method provides a useful approach to atomic-detail structure-based ligand binding calculations. It thereby opens up a new field of applications by permitting accurate computational studies of binding events in which quantum mechanical effects play a major role. Examples of these are complexes with an extensive π-system in the ligand and thus large polarization effects, with cation$-\pi$ interactions, or with metal binding involving charge-transfer effects.[31] Another future application area might be the affinity prediction of ligands covalently binding to their target or binding involving proton transfer between the inhibitor and the protein. The latter type of system would require the extension of the quantum mechanical region so as to include protein atoms.

The QM/MM-PB/SA approach presented here has the major advantage over purely classical-mechanics-based (MM-PB/SA) methods in that it does not require derivation of force field parameters for the ligand. The free energy calculation procedure can thus be fully automated. This permits atomic-detail physically based binding affinity calculations for a wide range of ligands with satisfying accuracy and within reasonable CPU time.

Acknowledgment. We thank Thomas Tschopp, Hoffmann-LaRoche AG, Basel, Switzerland, for providing us with experimental binding affinities for the trypsin ligands.

Note Added in Proof: After manuscript submission, a combined QM/MM study by Hensen et al. was published demonstrating the importance of polarization effects on HIV-1 protease inhibition (Hensen C.; Hermann, J. C.; Nam, K.; Ma, S.; Gao, J.; Holtje, H. D. A combined QM/MM approach to protein−ligand interactions: Polarization effects of the HIV-1 protease on selected high affinity inhibitors. *J. Med. Chem.* **2004**, *47*, 6673−6680).

Supporting Information Available: Calculated binding free energies and contributing terms for binding to flexible and rigid trypsin for the complete ligand sets n0 and n1. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Wang W.; Donini O.; Reyes C. M.; Kollman P. A. Biomolecular simulations: Recent developments in force fields, simulations of enzyme catalysis, protein−ligand, protein−protein, and protein−nucleic acid noncovalent interactions. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 211−243.

(2) Gohlke H.; Klebe G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew. Chem., Int. Ed.* **2002**, *41* (15), 2644−2676.

(3) Gohlke H.; Klebe G. Statistical potentials and scoring functions applied to protein−ligand binding. *Curr. Opin. Struct. Biol*. **2001**, *11* (2), 231−235.

(4) Wang R.; Lu Y.; Wang S. Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem*. **2003**, *46*, 2287−2303.

(5) Böhm H. Prediction of binding constants of protein ligands: A fast method for the prioritization of hits obtained from de novo design or 3D database search programs. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 309−323.

(6) Wang R.; Lai L.; Wang S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 11−26.

(7) Simonson T.; Archontis; G.; Karplus M. Free energy simulations come of age: Protein−ligand recognition. *Acc. Chem. Res.* **2002**, *35*, 430−437.

(8) Aqvist J.; Luzhkov V. B.; Brandsdal B. O. Ligand binding affinities from MD simulations. *Acc. Chem. Res.* **2002**, *35*, 358−365.

(9) Reddy M. R.; Erion M. D. Calculation of relative binding free energy differences for fructose 1,6-bisphosphatase inhibitors using the thermodynamic cycle perturbation approach. *J. Am. Chem. Soc*. **2001**, *123*, 6246−6252.

(10) Essex J. W.; Severance D. L.; Tirado-Rives J.; Jorgenson W. L. Monte Carlo simulations for proteins: Binding affinities for trypsin−benzamidine complexes via free-energy perturbations. *J. Phys. Chem. B* **1997**, *123*, 9663−9669

(11) Talhout R.; Villa A.; Mark A. E.; Engbert J. B. *J. Am. Chem. Soc.* **2003**, *125*, 10570−10579.

(12) Van Lipzig M. M. H.; ter Laak A. M.; Jongejan A.; Vermeulen N. P. E.; Wamelink M.; Geerke D.; Meerman J. H. N. Prediction of ligand binding affinity and orientation of xenoestrogens to the estrogen receptor by molecular dynamics simulations and the linear interaction energy method. *J. Med. Chem.* **2004**, *47*, 1018−1030.

(13) Boresch S.; Tettinger F.; Leitgeb M.; Karplus M. Absolute binding free energies: A quantitative approach for their calculation. *J. Phys. Chem. B* **2003**, *107*, 9535−9551.

(14) Oostenbrink B. C.; Pitera J. W.; van Lipzig M. M.; Meerman J. H.; van Gunsteren W. F. Simulations of the estrogen receptor ligand-binding domain: Affinity of natural ligands and xenoestrogens. *J. Med. Chem.* **2000**, *43*, 4594−4605.

(15) Pearlman D. A.; Charifson P. S. Improved scoring of ligand-protein Interactions using OWFEG free energy grids. *J. Med. Chem.* **2001**, *44*, 502−511.

(16) Simonson T. Macromolecular electrostatics: Continuum models and their growing pains. *Curr. Opin. Struct. Biol.* **2001**, *11*, 243−252.

(17) Kasper P.; Christen P.; Gehring H. Empirical calculation of the relative free energies of peptide binding to the molecular chaperone DnaK. *Proteins* **2000**, *40* (2), 185−192.

(18) Ma C.; Baker N. A.; Joseph S.; McCammon J. A. Binding of aminoglycoside antibiotics to the small ribosomal subunit: A continuum electrostatics investigation. *J. Am. Chem. Soc.* **2002**, *124*, 1438−1442.

(19) Wang J.; Morin P.; Wang W.; Kollman P. A. Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. *J. Am. Chem. Soc.* **2001**, *123*, 5221−5230.

(20) Huo S.; Wang J.; Cieplak P.; Kollman P. A.; Kuntz I. D. Molecular dynamics and free energy analyses of cathepsin D−inhibitor interactions: Insight into structure-based ligand design. *J. Med. Chem.* **2002**, *45*, 1412−1419.

(21) Archontis G.; Simonson T.; Karplus M. Binding free energies and free energy components from molecular dynamics and Poisson−Boltzmann calculations. Application to amino acid recognition by aspartyl-tRNA synthetase. *J. Mol. Biol.* **2001**, *306*, 307−327.

(22) Swanson J. M. M.; Henchman R. H.; McCammon J. A. Revisiting free energy calculations: A theoretical connection to MM/PBSA and direct calculation of the association free energy. *Biophys. J.* **2004**, *86*, 67−74.

(23) Schwarzl S. M.; Tschopp T. B.; Smith J. C.; Fischer S. Can the calculation of ligand binding free energies be improved with continuum solvent electrostatic and an ideal-gas entropy correction? *J. Comput. Chem.* **2002**, *23*, 1143−1149.

(24) Sims P. A.; Wong C. F.; McCammon J. A. A computational model of binding thermodynamics: The design of cyclin-dependent kinase 2 inhibitors. *J. Med. Chem.* **2003**, *46* (15), 3314−3325.

(25) Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parametrization, and performance of MMFF94. *J Comput. Chem.* **1996**, *17*, 490−519.

(26) Wang J., Wolf R. M., Caldwell J. W., Kollman P. A., Case D. A. Development and testing of a general AMBER force field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.

(27) Gao J.; Truhlar D. G. Quantum mechanical methods for enzyme kinetics. *Annu. Rev. Phys. Chem.* **2002**, *53*, 467−505.

(28) Auffinger, P.; Hays, F. A.; Westhof, E.; Ho, P. S. Halogen bonds in biological molecules. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16789−16794.

(29) Mo Y.; Subramanian G.; Gao J.; Ferguson D. M. Cation−π interactions: An energy decomposition analysis and its implication in δ-opioid receptor−ligand binding. *J. Am. Chem. Soc.* **2002**, *124*, 4832−4837.

(30) Garcia-Viloca M.; Truhlar D. B.; Gao J. Importance of substrate and cofactor polarization in the active site of dihydrofolate reductase. *J. Mol. Biol.* **2003**, *327*, 549−560.

(31) Ishchenko A. V.; Shakhnovich E. I. Small molecular growth 2001-(SmoG2001): An improved knowledge-based scoring function for protein−ligand interactions. *J. Med. Chem.* **2002**, *45*, 2770−2780.

(32) Raha K.; Merz K. M. A quantum mechanics-based scoring function: Study of zinc ion-mediated ligand binding. *J. Am. Chem. Soc.* **2004**, *126*, 1020−1021.

(33) Erion M. D.; Reddy M. R. Calculation of relative hydration free energy differences for heteroaromatic compounds: Use in the design of adenosine deaminase and cytidine deaminase inhibitors. *J. Am. Chem. Soc.* **1998**, *120*, 3295−3304.

(34) Yang W.; Biletti-Putzer R.; Karplus M. Chaperoned alchemical free energy simulations: A general method for QM, MM, and QM/MM potentials. *J. Chem. Phys.* **2004**, *120*, 9450−9453.

(35) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian 94*, revision D.4; Gaussian, Inc.: Pittsburgh, PA, 1995.

(36) Claussen H.; Buning C.; Rarey M.; Lengauer T. FlexE: Efficient molecular docking considering protein structure variations. *J. Mol. Biol.* **2001**, *308*, 377−395.

(37) Österberg F.; Morris G. M.; Sanner M. F.; Olson A. J.; Goodsell D. S. Automated docking to multiple target structures: Incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins* **2002**, *46*, 34−40.

(38) Krieger M.; Kay L. M.; Stroud R. M. Structure and specific binding of trypsin: Comparison of inhibited derivatives and a model for substrate binding. *J. Mol. Biol.* **1974**, *83*, 209−218.

(39) Mangel W. F.; Singer P. T.; Cyr D. M.; Umland T. C.; Toledo D. L.; Stroud R. M.; Pflugrath J. W.; Sweet R. M. Structure of an acyl-enzyme intermediate during catalysis: (guanidinobenzoyl)trypsin. *Biochemistry* **1990**, *29*, 8351−8357.

(40) Sirockin F.; Sich C.; Improta S.; Schaefer M.; Saudek V.; Froloff N.; Karplus M.; Dejaegere A. Structure activity relationship by NMR and by computer: A comparative study. *J. Am. Chem. Soc.* **2002**, *124*, 11073−11084.

(41) Katz B. A.; Finer-Moore J.; Mortezaei R.; Rich D. H.; Stroud R. M. Episelection: Novel K_i nanomolar inhibitors of serine proteases selected by binding or chemistry on an enzyme surface. *Biochemistry* **1995**, *34*, 8264

(42) Brooks B. R.; Bruccoleri R. E.; Olafson B. D.; States D. J.; Swaminathan S.; Karplus M. CHARMM: A program for macromolecular energy, minimization and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.

(43) Dewar M. J. S.; Zoebisch E. G.; Healy E. F.; Stewart J. P. Development and use of quantum mechanical molecular models. 76. AM1: A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.

(44) Stewart J. J.P. Calculation of the geometry of a small protein using semiempirical methods. *THEOCHEM* **1997**, *401*, 195−205.

(45) Field M. J.; Bash. P. A.; Karplus M. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comput. Chem.* **1990**, *11*, 700−733.

(46) Gao J., Xia X. A priori evaluation of aqueous polarization effects through Monte Carlo QM-MM simulations. *Science* **1992**, *258*, 631−635.

(47) Gerber P. R.; Müller K. MAB; A generally applicable molecular force field for structure modelling in medicinal chemistry. *J. Comput.-Aided Mol. Des.* **1995**, *6*, 251−268.

(48) Schwarzl S. M.; Huang D.; Smith J. C.; Fischer S. Non-uniform charge scaling (NUCS): A practical approximation of solvent electrostatic screening in proteins. *J. Comput. Chem.*, in press.

(49) Im W.; Beglov D.; Roux B. Continuum solvation model: Electrostatic forces from numerical solutions to the Poisson−Boltzmann equation. *Comput. Phys. Commun*. **1998**, *111*, 59−75.

(50) Miranker A.; Karplus M. Functionality maps of binding sites: A multiple copy simultaneous search method. *Proteins* **1991**, *11*, 29−34.

(51) *SYBYL*, version 6.8; Tripos Associates: St. Louis, MO

(52) Brady G. B.; Sharp K. A. Decomposition of interaction free energies in proteins and other complex systems. *J. Mol. Biol.* **1995**, *254*, 77−85.

(53) Dill K. A. Additivity principles in biochemistry. *J. Biol. Chem.* **1997**, *272*, 701−704.

(54) For examples, see: Castillo R.; Silla E.; Funon I. Role of protein flexibility in enzymatic catalyis: quantum mechanical-molecular mechanical study of the deacylation reaction in class A β-lactamases. *J. Am. Chem. Soc.* **2002**, *124*, 1809−1816. Hutter M. C.; Helms V. The mechanism of phosporylation of natual nucleosides and anti-HIV analogues by nucleoside diphosphate kinase is independent of their sugar substituents. *ChemBioChem* **2002**, *3*, 643−651.

(55) Davis M. E.; Madura J. D.; Luty B. A.; McCammon J. A. Electrostatics and diffusion of molecules in solution: Simulations with the University of Houston Brownian dynamics program. *Comput. Phys. Commun.* **1991**, *62*, 187−197.

(56) Madura J. D.; Briggs J. M.; Wade R. C.; Davis M. E.; Luty B. A.; Ilin A.; Antosiewicz J.; Gilson M. K.; Bagheri B.; Scott L. R.; McCammon J. A. Electrostatics and diffusion in solution: Simulation with the University of Houston Brownian dynamics program. *Comput. Phys. Commun.* **1995**, *91*, 57 95.

(57) Sharp K. A.; Nicholls A.; Fine R. F.; Honig B. Reconciling the magnitude of the microscopic and macroscopic hydrophobic effect. *Science* **1991**, *252*, 106−109.

(58) Tuñón I.; Pascual-Ahuir J. L. Molecular surface area and hydrophobic effect. *Protein Eng.* **1992**, 5715−5716.

(59) Fischer S. CODISP, a program for computing continuum dispersion interactions. Unpublished work.

(60) Connolly M. L. Solvent-accessible surfaces of proteins and nucleic acids. *Science* **1983**, *221*, 709−713.

(61) Fischer S.; Smith J. C.; Verma C. S. Dissecting the vibrational entropy change on protein/ligand binding: Burial of a water molecule in bovine pancreatic trypsin inhibitor. *J. Phys. Chem. B* **2001**, *105*, 8050−8055.

(62) Gerber P. R. *MOLOC*; http://www.moloc.ch/.

(63) Oshiro C. M.; Kuntz I. D.; Dixon J. S. Flexible ligand docking using a genetic algorithm. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 113−130.

(64) Rarey M.; Kramer B.; Lengauer T.; Klebe G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470−489.

(65) Jones G.; Willett P.; Glen R. C.; Leach A. R.; Taylor R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727−748.