A Reliable Docking/Scoring Scheme Based on the Semiempirical Quantum Mechanical PM6-DH2 Method Accurately Covering Dispersion and H-Bonding: HIV-1 Protease with 22 Ligands

Jindřich Fanfrlík,[†] Agnieszka K. Bronowska,[‡] Jan Řezáč,[†] Ondřej Přenosil,[†] Jan Konvalinka,[†] and Pavel Hobza*[,†,§]

Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic and Center for Biomolecules and Complex Molecular Systems, 166 10 Prague 6, Czech Republic, Heidelberg Institute for Theoretical Studies, D-69118 Heidelberg, Germany, and Department of Physical Chemistry, Palacký University, 771 46 Olomouc, Czech Republic

Received: April 13, 2010; Revised Manuscript Received: August 18, 2010

In this study, we introduce a fast and reliable rescoring scheme for docked complexes based on a semiempirical quantum mechanical PM6-DH2 method. The method utilizes a PM6-based Hamiltonian with corrections for dispersion energy and hydrogen bonds. The total score is constructed as the sum of the PM6-DH2 interaction enthalpy, the empirical force field (AMBER) interaction entropy, and the sum of the deformation (PM6-DH2, SMD) and the desolvation (SMD) energies of the ligand. The main advantage of the procedure is the fact that we do not add any empirical parameter for either an individual component of the total score or an individual protein−ligand complex. This rescoring method is applied to a very challenging system, namely, the HIV-1 protease with a set of ligands. As opposed to the conventional DOCK procedure, the PM6-DH2 rescoring based on all of the terms distinguishes between binders and nonbinders and provides a reliable correlation of the theoretical and experimental binding free energies. Such a dramatic improvement, resulting from the PM6-DH2 rescoring of all the complexes, provides a valuable yet inexpensive tool for rational drug discovery and de novo ligand design.

Introduction

In silico drug design provides a very attractive avenue for pharmaceutical research.[1−4] The use of computational tools could dramatically reduce the number of compounds that need to undergo synthesis and testing in the process of discovering potent new drugs. Recently, in silico drug design has matured greatly and has already demonstrated several major successes.[1−4] In contrast to purely empirical or statistical methods, such as QSAR,[5−8] the more rigorous molecular structure-based approaches are critically dependent on a realistic and accurate description of the protein−ligand binding. This means that the interaction of the complexes must be described fully in terms of the binding free energy and not merely the binding energy or enthalpy. Because the complete binding free energy is difficult or impossible to calculate, it is often replaced by the scoring function, a generalized variable that correlates with the free energy and replaces it in the practical use.

In this work, we present a physicochemical approach to this problem attempting to eliminate or minimize empiricism in the calculation of the protein−ligand binding free energy. The total binding free energy can be divided into multiple separate components; to each of them, we apply the best method available to solve it efficiently. Less empiricism also means increased robustness, since the method can be applied to different systems without modifications. It also leads to a simpler workflow, because no previous knowledge of the particular system is required to set up the calculation. The detailed physical description of the system also allows the decomposition of the ligand−target interactions and provides deeper insight into the nature of the molecular recognition, which is essential for any truly "rational" drug design. We also believe that a more accurate, albeit more demanding, calculation of the binding free energy greatly reduces the set of compounds that require experimental synthesis and testing and hence reduces the overall expense of drug development.

Since such a calculation is more demanding than the simpler methods commonly used, we have limited our calculations to an evaluation of the binding free energy on structures obtained by other methods rather than integrating this computational protocol into a docking algorithm.

The most common protocol of the computational prediction of the pharmacological profile of drugs can be divided into two steps: molecular docking and virtual screening. Molecular docking, in principle, predicts the optimal geometry and orientation of the drug upon binding to the cognate molecular target. The final goal is, of course, to optimize the binding free energy. Since a large ensemble of possible configurations has to be tested, more approximate but efficient scoring functions are used at this stage. A large number of docking software packages is available. These packages use different search algorithms and methods to estimate the binding free energy needed to select the most favorable binding orientations. For a review of the docking methods and packages, see ref 9. The process of virtual screening takes the binding scores and ranks the compounds into a list of drug candidates by predicted affinity; the best-scoring candidates are then selected for further

* To whom correspondence should be addressed. Fax: +420 220 410 320. E-mail: pavel.hobza@uochb.cas.cz.
[†] Academy of Sciences of the Czech Republic and Center for Biomolecules and Complex Molecular Systems.
[‡] Heidelberg Institute for Theoretical Studies.
[§] Palacký University.

10.1021/jp1032965
Published on Web 09/14/2010

experimental testing. Here, we can benefit from a scoring function that is as accurate as possible; if it is different from the one used in docking, we call this step rescoring. The scoring function is also the measure of binding affinity used for lead optimization (ligand modification)[10] and the prediction of drug resistance (protein mutation).[11]

One of the important challenges in docking predictions is the nature of the protein target: its dynamical behavior, flexibility, the large number of degrees of freedom, and the complexity of its free energy surface. The intrinsic dynamics of the target, emerging from the presence of flexible linkers and hinge regions, side-chain motions, dynamic allostery, large domain reorientation, domain−domain interactions, ligand-enhanced protein mobility, etc., play a prominent role in the control of a drug binding.[12−15] The combination of these factors may prevent the current docking algorithms from identifying the best binder from a list of compounds.

While the calculations of the absolute values of the protein−ligand binding free energies are currently impractical, the relative free energy values of structurally related ligands can be obtained based on empirical force field molecular dynamics (MD) simulations. MD-based methods,[16−19] although accurate within the limit of the accuracy of empirical potential, are computationally demanding and time consuming, with the result being that their application for certain stages of the drug development process, such as high-throughput screening, is very limited.

In our work, we focus on the methods that apply a scoring function to a single representative geometry of the system. In our protocol, however, both the complex and the free ligand and receptor are fully relaxed, so this method accounts for part of the flexibility of the system. This approach is widely used, and previous studies, as well as our work, indicate that such a description is sufficient for most systems. The other possibility, calculation of the binding free energies from a set of structures sampling the configuration space,[20] is not discussed here.

The measure of the quality of a scoring function is how accurately it describes the binding free energies.[21,22] At first glance, the procedure appears fairly straightforward—the binding between a protein and a ligand occurs solely because of the noncovalent interactions[23] between both species. More detailed insight reveals, however, that the picture can be very complex. Figure 1A shows the complex of the HIV-1 protease and its inhibitor, atazanavir (ATV), which is a frequently studied protein−ligand complex. Figure 1B highlights several characteristic motifs of the protein−ligand binding. Besides the well-described strong electrostatic and H-bonding, there are also stacking interactions, X−H$\cdots\pi$ (X = O, N, C) interactions, halogen bonding, and aliphatic$\cdots$aliphatic dispersion-bound motifs. The first two interactions are stronger than the others but the weaker ones are often numerous. Moreover, dispersion is not weakened by solvation in contrast to electrostatic interaction (including hydrogen bonding) and thus the latter ones can often bring an equally important contribution to the overall stabilization energy. A reliable and well-balanced theoretical description of all of these motifs is therefore pivotal, and only highly accurate quantum mechanical (QM) methods can provide satisfactory results.

Recently, we have reviewed the performance of various wave function and density-functional theories (WFT and DFT) in providing reliable energies and geometries for noncovalent complexes.[23] The techniques frequently used in the description of biological systems, such as DFT methods using standard functionals, or standard semiempirical QM methods, often fail,

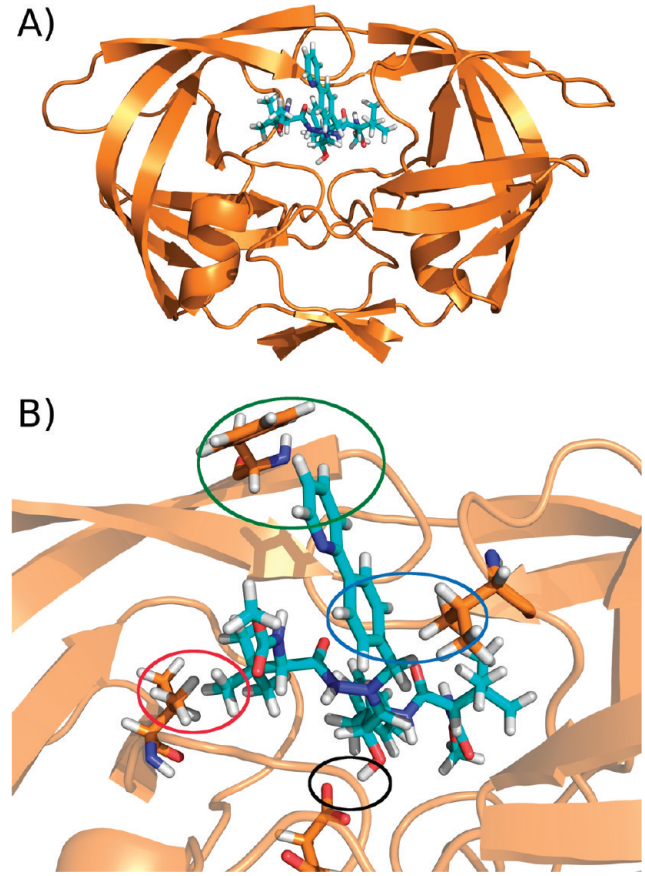


Figure 1. Crystal structure (2AQU[59]) of the HIV PR−atazanavir (ATV) complex: (A) an overall and (B) detailed view with several characteristic motifs of the protein−ligand binding highlighted.

and the application of WFT and DFT methods for drug design is limited by the size of the systems studied, which can be larger than several thousand atoms. When dealing with large systems, empirical force fields (also called molecular mechanics or MM methods) are frequently employed and are capable of satisfactorily describing many types of noncovalent complexes. As part of the review,[23] MM techniques were tested using benchmarks similar to those of the WFT and DFT methods. The MM methods demonstrated good performance with various noncovalent complexes.[23] The force-field-based methods are computationally very efficient and thus seem, superficially, to be ideally suited for use in drug discovery. The efficacy of such methods, and therefore their applications to drug discovery, are, however, seriously limited by the inability to incorporate quantum effects, principally the charge transfer between a protein and a ligand, and the wide variations in atomic charges between different structures of the protein or ligand.[24] Raha and Merz demonstrated that the charge transfer between a protein and a ligand is frequently as high as 0.1e, and for metalloproteins even several times larger (0.5e and more). This charge transfer relates to the stabilization energies of dozens of kcal/mol and MM cannot describe this stabilization correctly and attribute it to the right reasons. Another issue concerns the determination of atomic charges by MM methods. Several studies have indicated problems with the MM calculations of metalloproteins in this regard.[25] Others report problems with a variability of the polarized charges in proteins upon the binding of different ligands.[26] The issue has also arisen in simple peptides, where a strong dependence of their atomic charges on the local structure has been reported.[27] Figure 2 demonstrates the structure of the acetyl-pepstatin (ACP) in a complex with HIV-1 protease

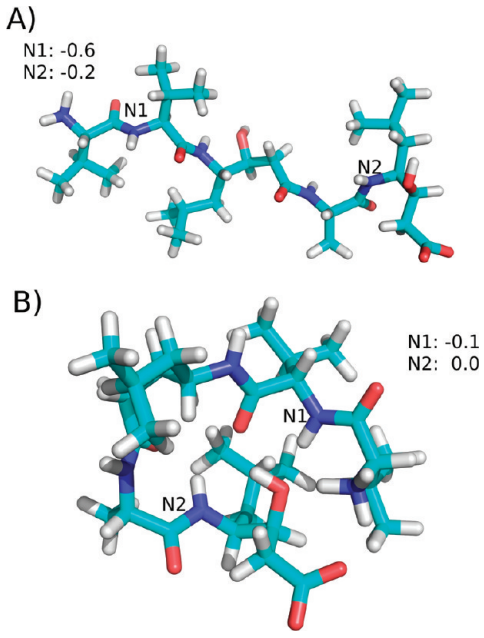


Figure 2. Comparison of the RESP charges (in e) for nitrogens N1 and N2 determined on the acetyl-pepstatin (ACP) in the geometry taken from the crystal (5HVP[58]) and presented in the water environment for conformation.

determined by X-ray crystallography, and in its free hydrated state obtained from simulation. It is evident that both structures are dramatically different, which likely indicates differences in their atomic charge distributions. To illustrate this phenomenon, the atomic charges were determined using the restrained fit to the electrostatic potential (RESP) technique,[28] and the N1 and N2 charges are shown in Figure 2. The difference is vast, calling into question the usage of the same charges for different structures. The use of a more advanced force field based on conformer-averaged charges and including polarization may alleviate these problems, but preparation of such a force field, required for each drug candidate, would be impractical. Another important consequence of quantum effects is the existence of halogen bonds that play a significant role in the binding of halogenated ligands.[29] The standard MM techniques, including the polarized ones, fail to describe this effect.[30] These quantum effects have significant demonstrable consequences on the quality of the respective calculations and should not be ignored in in silico drug design.

The difficulties described point to the conclusion that QM, rather than MM techniques, should be preferentially used for drug discovery applications. All nonempirical WFT and DFT techniques are, however, CPU-time demanding and, therefore, not good candidates for high-throughput drug design. An alternative approach would rely on the usage of either hybrid QM/MM or semiempirical QM (SQM) methods. They both cover quantum effects and have already been implemented for use in rational drug discovery.[31] Different QM/MM models have been used[32] in the molecular docking to HIV-1 protease: when the QM part contained only the ligand, the ab initio Hartree−Fock (HF), and semiempirical AM1 and PM3 methods were utilized, whereas when the QM part consisted of a ligand and nearest amino acids, only the semiempirical PM3-D technique was adopted. For the discrimination of native from non-native modes, the HF/Amber and fully empirical Amber approach exhibited comparable performance, indicating that the QM/MM technique was not successful in this case. This might be due to the use of either a limited QM part or an inefficient SQM method or both. The fully semiempirical QM approach describes the whole complex directly by using SQM methods. The use of SQM for large systems, however, depends on the introduction of linear scaling techniques. Obviously, both approaches have their advantages and disadvantages,[33,34] but we have concluded that the description of the whole complex consistently without any artificial division into the QM and MM parts is better suited for drug design purposes.

Merz et al. pioneered the use of SQM techniques in the drug design process (see above and ref 35). All of these studies were based on the standard semiempirical AM1 or PM3 Hamiltonians. Recently, the PM6 method has been also applied in virtual screening of large database of compounds.[36] As previously mentioned, both the dispersion-bonded and H-bonded noncovalent complexes are incorrectly described by these methods. Both of these failures are well-known and the first one can be eliminated by adding a separate term for dispersion attraction. Wollacott and Merz[37] constructed the binding free energy of a protein−ligand complex as a sum of the AM1 or PM3 heat of formation, solvation free energy, and the attractive term from the Lennard-Jones potential. More straightforward and internally consistent ways to solve this problem do exist, such as adding the missing dispersion term to a SQM method. In recent years, the performance of these methods has been dramatically improved, enabling the calculation of dispersion-bound noncovalent complexes. A description can be found in our recent review.[23] The simplest and most efficient method among the several different ways attempted appears to be the addition of the London dispersion energy in the form of the well-known C_6 expansion.[38,39] Exactly the same procedure has been used in the realm of the SQM methods, and the addition of the dispersion energy improved the performance of these methods considerably (for references, see also our recent review).[23]

All of our present SQM calculations were based on Stewart's PM6 semiempirical QM method,[40] which is a modern SQM method combining the recent developments in the field with parametrization for the majority of the chemical elements. This has allowed the use of a single method for almost any compound of interest. Its implementation in the MOPAC package[41] has offered a linear scaling algorithm MOZYME[42] based on the localized orbitals. In our extension of this SQM method, we have developed a dispersion correction as well as a correction for hydrogen bonds.[43,44] The resulting method, using the second generation of the corrections, named PM6-DH2, yields, to the best of our knowledge, the most accurate results for noncovalent interactions of all the SQM methods. For small model noncovalent complexes, which allow a comparison benchmark against the high-level WFT calculations, the method reaches chemical accuracy (1 kcal/mol) and the quality of the results is comparable to the much more computationally expensive WFT and DFT techniques. The dispersion correction is based on our previous work on the DFT-D methods.[39] To avoid double counting of the part of the dispersion energy already contained in the PM6 method, the dispersion correction is adapted in the following way: first, the whole correction had to be scaled, and second the overestimated dispersion contribution of the sp^3-hybridized carbon atoms had to be corrected specifically by decreasing the C_6 coefficient of all of the sp^3 carbon atoms. The H-bonding correction added a force-field-like term to each pair of atoms that could possibly form a H-bond. Unlike the core−core terms in the SQM methods, our correction was directional, depending not only on the distance but also on the H-bond angle and other geometric parameters that substantially improved the description of the hydrogen bonds. These corrections are of vital importance

for predicting the correct structure of a protein−ligand complex. More details on the PM6-DH2 method as well as a comparison with the high-quality QM results can be found in the original paper.[44]

In this study, the PM6-DH2 method was used for the first time for a rescoring procedure in rational drug design and a single protein target was considered. We demonstrate the performance of our method using a well-studied system, namely the HIV-1 protease (PR) and its complexes with 22 ligands. The methods used are, however, robust enough to describe various physical effects and thus a further extension of the methodology to describe mutated proteins (such as resistant variants of the enzyme) and other systems is straightforward (see later).

HIV PR is responsible for the cleavage of the viral polyprotein precursors of HIV into mature enzymes and structural proteins yielding infectious viral particles. Inhibition of the enzyme renders the virus noninfectious, and the PR inhibitors (PI) are thus powerful virostatics (for a review, see refs 45−47). Currently, there are nine PIs approved by the US Food and Drug Administration for clinical use. Almost 200 structures of various HIV PR−inhibitor complexes have been determined by X-ray crystallography and NMR. The history of the development of HIV PR inhibitors is a remarkable example for the power and limitations of a structure-based drug design and provides a unique data set of the structural and enzymological information for the analysis of protein−ligand interactions.

In the first step of our analysis, the putative ligand−protein binding modes were generated. Subsequently, all the obtained modes were rescored using the PM6-DH2 method. All of the atoms of the protein−ligand complex were optimized in aqueous solution. The in silico predictions were finally validated against the experimental data (X-ray crystallography and spectrophotometry) while achieving very good agreement between the predictions and experimental data.

It must be noted here that results of similar quality (measured as the correlation to experiment) can possibly be obtained with single-purpose knowledge-based empirical scoring function. The main advantage of the approach presented here is the absence of any system-specific parameters, which means that it can be applied to any protein−ligand complex without further modification.

Methods

Theoretical Description of the Protein−Ligand Binding. The formation of a protein (P)−ligand (L) complex from free subsystems in a water environment

$$P_w + L_w \leftrightarrow (P\cdots L)_w \quad (1)$$

is a crucial step in the whole drug design process, with the aim of a theoretical description being the evaluation of the binding free energy. The direct determination of the binding free energy in solution is unfeasible. It may be, however, determined using the thermodynamic cycle shown in Scheme 1 (see also ref 24).

The binding free energy, ΔG_w, can be determined from the cycle as follows:

$$\Delta G_w = \Delta G_g + \Delta G_w(PL) - [\Delta G_w(P) + \Delta G_w(L)] \quad (2)$$

where the gas-phase binding free energy (ΔG_g) is evaluated as the difference between the enthalpy and entropy terms

$$\Delta G_g = \Delta H_g - T\Delta S_g \quad (3)$$

It is unfeasible to evaluate all of the energies in eqs 2 and 3 consistently on the basis of a single theoretical method; several approximations should be adopted. The currently used approach can be described as follows:

(i) The structure of the protein−ligand complex (as determined by the DOCK code) is reoptimized using the PM6-DH2 method in a continuum water environment as implemented in the MOPAC code. The resulting electronic energy is augmented by the hydration free energy evaluated with the built-in COSMO method,[48] which considers only the electrostatic terms. These energies are assigned as $E'_w(X)$, where X means the protein−ligand complex or the isolated protein or ligand.

(ii) The ligand and also the protein should be desolvated prior to the binding. The desolvation free energy represents a critical term and needs to be calculated as accurately as possible. This term is usually repulsive and thus opposes binding. The desolvation of both the ligand and protein is calculated using the MOPAC code, which covers only the electrostatic terms (electrostatic interaction and mutual polarization of the solute and solvent). Nonelectrostatic terms associated with formation of a cavity to accommodate the solute, and the van der Waals interaction between solute and solvent, are neglected. This approximation works well for the studied systems, as the protein is always the same and the ligands are of similar size, which makes the differences in nonelectrostatic contributions between the ligands very small. Furthermore, the size of the ligand permits a more accurate calculation of the complete solvation free energy of the isolated ligand with a SMD method[49] based on full Hartree−Fock calculation, as implemented in the Gaussian 09 code.[50] Two different geometries of the ligand, corresponding to the bound structure in a PL complex and to a free structure in a solvent, were considered. The first was taken from the optimized P−L complex (step i), whereas the second was taken from the molecular dynamics/quenching (MD/Q) simulations using the PM6-DH2 method with a continuum solvent. Several structures obtained by quenching were optimized and the resulting structure corresponded to the average. An alternative (and much more straightforward) determination of this structure was based on the optimization of the ligand structure in a continuum solvent using the PM6-DH2 method. The solvation free energy was assigned as $G_w(X)$, where X was a protein−ligand complex or an isolated species.

(iii) The structures of the ligand and the protein in the protein−ligand complex and in the solution differed. The deformation free energy of the ligand is the energy required to "deform" the ligand from its optimal solution structure (see above) to the structure it had adopted in the complex. The PM6-DH2 electronic energy of the ligand was augmented by the SMD hydration free energy determined using the G09 code (i.e., the electrostatic as well as nonelectrostatic terms were included). The deformation free energy of the protein should be also considered and was calculated as the difference in the energy of the protein in the protein−ligand complex geometry and in the minimized geometry. The protein deformation was determined using the AMER potential.

(iv) The formation of a protein−ligand complex restricted motions of the ligand as well as the torsional motion of the protein. Furthermore, the vibrational entropy of a ligand was also restricted by the formation of a protein−ligand complex. The total entropy term was repulsive and, like the desolvation term, opposed the binding. The change of entropy accompanying the protein−ligand complex formation in a water environment

was determined as the difference between entropies of a complex and a sum of the subsystem entropies. The entropy term was repulsive since the ligand lost three degrees of rotations and vibrations upon binding, and these degrees were converted to the vibrational motions. Ligand binding led also to the loss of the configurational entropy. The evaluation of the configuration entropy is notoriously difficult and this term was either neglected or was estimated on the basis of a reduced number of the accessible rotamers upon binding. Recently, it has been shown[51] that vibrational entropy dominated configurational entropy and, consequently, it was recommended to evaluate the entropy term for the protein−ligand binding on the basis of the standard rigid rotor/harmonic oscillator approximation. In the present study, the entropic contributions were determined using the rigid rotor/harmonic oscillator approximation based on Cornell et al. empirical potential constants.[52] The structures of all of the systems were reoptimized using the empirical potential considering the continuous water, and the same method was used for the calculation of the second derivatives of the total energy. The entropies were assigned as $S(X)$, where X was either a protein−ligand complex or an isolated species.

The binding free energy in the aqueous solution, ΔG_w, was thus approximated as follows:

$$\Delta G_w \approx [E'^w(PL) - E'^w(P)^{PL} - E'^w(L)^{PL}] - [E'^w(L)^{PL} - E'^{w\text{-}G}(L)^{PL}] - [E'^{w\text{-}G}(L)^{PL} - E'^{w\text{-}G}(L)^{L\text{-}opt}] - T\Delta S + \Delta ZPVE \quad (4)$$

where

$$E'^w(PL) = E(PL) + G_w(PL) \quad (5)$$

$$E'^w(P)^{PL} = E(P)^{PL} + G_w(P)^{PL} \quad \text{(and analogously for the } E'^w(L)^{PL}\text{)} \quad (6)$$

$$\Delta S = S_w(PL) - [S_w(P) - S_w(L)] \quad (7)$$

The symbols used were explained above; the upper index PL means that the subsystem P or L is considered in the complex geometry while the upper index L-opt means a ligand optimal geometry. The upper index w indicates that a continuous water environment was taken into account and the hydration free energies were determined using the COSMO method as implemented in the MOPAC code, whereas w-G indicates that the hydration free energy was calculated using the SMD procedure as implemented in the G09 code.

To summarize, a two-step docking and rescoring scheme was introduced and applied in the present study. First, binding modes were generated using the DOCK with an empirical potential-based scoring function. Then, the obtained geometries were rescored by the procedure based on the PM6-DH2 method described above.

Systems Investigated. For the evaluation of the docking and rescoring scheme, we used the HIV-1 protease and its 22 putative or true ligands. For 11 of them, the 3-D structure of their complexes with HIV PR is available in the Protein Data Bank database: saquinavir (SQV) 3CYX,[53] indinavir (IDV) 2BPX,[54] nelfinavir (NFV) 1OHR,[55] ritonavir (RTV) 1HXW,[56] lopinavir (LPV) 1MUI,[57] amprenavir (APV) 1HPV,[58] atazanavir (ATV) 2AQU,[59] darunavir (DRV) 1T3R,[60] acetyl-pepstatin (ACP) 5HVP,[61] Boc-Phe-Psi(S)-CH(OH)CH2NH]-Phe-Gln-Phe-NH2 (SQ) 1IIQ,[62] and Boc-Phe-Psi[CH2CH2NH]-Phe-Glu-Phe-NH2 (OE) 1M0D.[63] Another 11 compounds represent "false" binders (see Table 1). The complexes were selected based on their clinical relevance and predominant types of interactions, as well as the availability of consistent experimental structure and binding data.[53−66]

Protonations. An important part of the preparation was to determine the protonation states of the titratable groups. During the docking procedure, the amine groups were protonated (N-terminal, Lys, Arg, His) and the carboxylic groups were negatively charged (Asp, Glu, C-terminal).

The inhibitors were considered to be neutral with the exception of OE,[63] which is negatively charged (a charge of −1). In the case of NFV, both the neutral and protonated forms were, however, reported in the literature[67,68] and thus both the neutral and protonated forms were considered in the calculations.

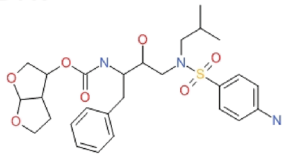
Determination of the Binding Constants. The inhibition constant values (K_i) for individual inhibitors had been determined previously[64−66] while the K_i value for ACP was determined in this study. All of the values were obtained under identical conditions for better comparison using the chromogenic peptide substrate KARVNleNphEANle-NH_2 in a spectrophotometric assay based on a decrease in the absorbance at 305 nm upon substrate cleavage.[69] The assay was carried out at 37 °C in a buffer containing 100 mM sodium acetate, with a pH of 4.7, 4 mM of EDTA, and 0.3 M of NaCl. The substrate was added to the final concentration near the K_m of the enzyme (15 μM) into a buffer containing 8 pmol of wild-type HIV-1 protease and various concentrations of a corresponding inhibitor dissolved in DMSO. The final concentration of DMSO in the reaction mixture was kept below 2.5%. The data were analyzed using the Dixon plot.[70]

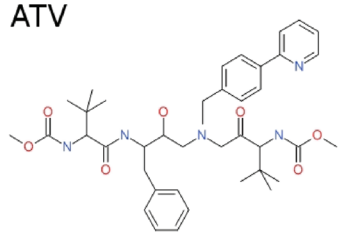
Docking. ***Docking by DOCK 6.2.*** *Preparation of the HIV-1 Protease Structure for the Docking.* An initial structure of the HIV-1 protease (homodimer) was taken from the Protein Data Bank (PDB code 2AQU).[59] In the course of the preparation for docking, the water molecules, ions, and other ligands were removed and the hydrogen atoms were added by Chimera.[71] The preparation procedure consisted of (1) a determination of the protonation state of the titratable groups by PROPKA,[72] (2) an assignment of partial charges and atomic parameters for all the atoms, (3) a determination of the location of the binding site for ligand docking, and (4) a generation of the electrostatic and van der Waals grids (the GRID module). The grid spacing was 0.25 Å while the cutoff for the nonbonded interactions was 12 Å. In all of the cases, the Amber parameters were used. The standard RESP charges[28] were assigned for each atom at the RHF/6-31G* level of theory.

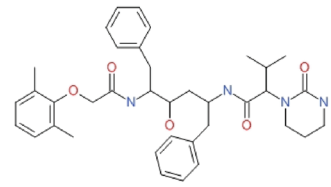
In several test cases, a few minor modifications of the protein structure have been introduced prior to the docking procedure. Namely, the protonation states of the catalytic aspartyl residues were changed manually in order to investigate the effect of different protonation states (and pH dependencies) on the docking results. We investigated the HIV-1 protease in the state of two catalytic aspartates of (1) deprotonated, (2) monoprotonated (i.e., with one of the two aspartates protonated), and (3) double-protonated states.

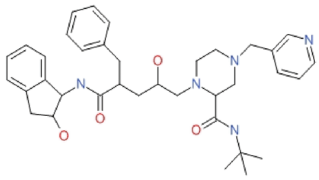
Preparation of the Ligand Molecules. The set of 11 HIV-1 inhibitors was used in the calculations (see Table 1). In addition, the set of 11 "nonbinders" was used to evaluate the procedure. For each compound, the atomic charges were assigned by the Antechamber module of AMBER, using the RESP method (HF/

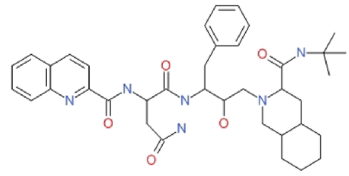
TABLE 1: Chemical Structures of the HIV-1 Protease Inhibitors and Other Ligands Used in the Docking Study

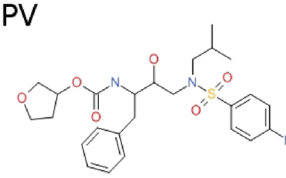
DRV 

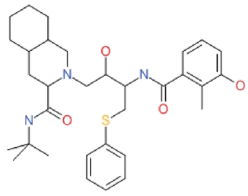
ATV 

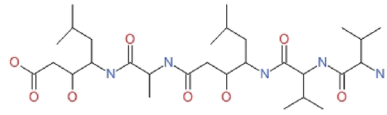
LPV 

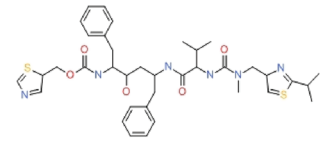
IDV 

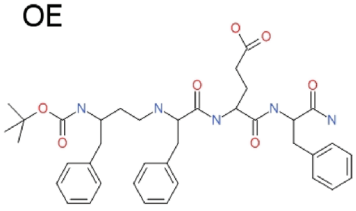
SQV 

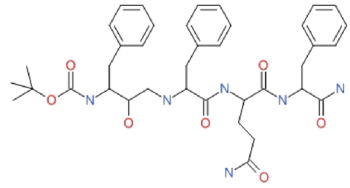
APV 

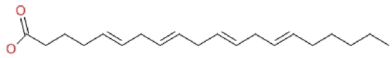
NFV 

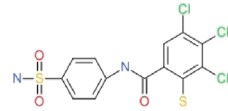
ACP 

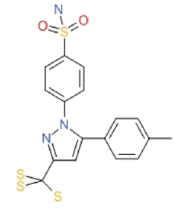
RTV 

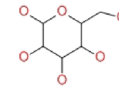
OE 

SQ 

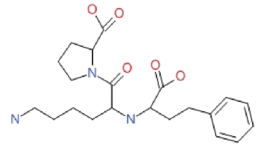
ARA 

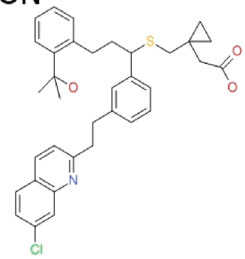
BOS 

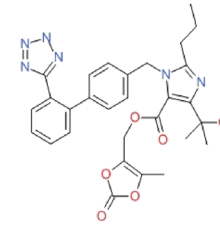
CEL 

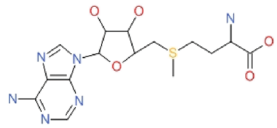
GAL 

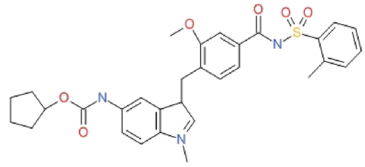
HEX 

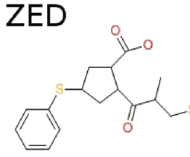
LSN 

MON 

OLM 

SAM 

ZAF 

ZED 

6-31G* level).[28] The compounds were parametrized (which included assigning atom types, fixing bonds, torsional angles, etc.) using Amber GAFF[73] parameters.

Testing the Docking Protocol and the DOCK Scoring Procedure. In order to establish a reliable docking and scoring procedure, a calibration set of several small molecules was

created and tested. It included three sets of compounds: the "true" binders (inhibitors), small molecule ligands, which should not bind with high affinity to the HIV-1 protease but should sterically fit into the catalytic pocket (e.g., galactose, cytosine, zebularine), and several compounds not fitting sterically into the binding pocket and/or too hydrophobic (e.g., oleic acid, *n*-octanol, *n*-hexanol). The calibration procedure was used to test and possibly to modify the screening protocol. All of the inhibitors were expected to have scores placing them fairly high on the "scoring test list". Small nonbinders were expected to be docked successfully into the binding site, but in terms of docking scores they were supposed to be placed fairly low on the list. The hydrophobic ligands, such as *n*-hexanol, were expected to fit sterically into the binding pocket, however, with low affinity scores, since the electrostatic energy was to supposed to be unfavorable. Finally, the very large compounds from this set were expected to be placed low on the scoring list, as they are too bulky to fit in the catalytic site. Several combinations of docking parameters (such as different minimizer options, numbers of ligand orientations and conformations, etc.), space sampling (number of allowed bumps, etc.), and scoring (contact score, energy score, "Amber score"), treatment of ligand (flexible or rigid) were tested, and the results have been compared in order to optimize the docking protocol. For flexible docking, the allowable_bump parameter was set to 12, anchor size to 5, max_orientations to 1500, and max_conformations to 100.

On the basis of the results of the testing procedure as well as the computational costs, we opted for the rigid docking as the method for prescreening the ligands and the selection of binding modes. This method had a similar performance to flexible docking, yet it was not only much faster than flexible docking but also much more suitable for fast high-throughput screening, which is the long-term goal of our approach.

All of the generated modes (complexes of HIV-1 protease with the tested compounds) were subjected to MM energy minimization and short MD simulation. In order to sample the conformational space efficiently, each ligand was docked and its 100 best modes were saved. The modes were then clustered, using rmsd on heavy (i.e., non-hydrogen) atoms as a selection criterion; the ligand modes were regarded as unique when the rmsd value was <1.5 Å. Finally, the top-scoring mode (DOCK6 scoring function) from every cluster was subjected to a crude ligand molecular-mechanical energy minimization (2500 steps, both ligand and active site flexible, AMBER force field). DOCK energy (measured as vdW and electrostatics interactions) was applied to choose the most energy-favorable ligand binding mode. The top 15 modes were used for further evaluation using the PM6-DH2 method. To validate the procedure, these modes (or ligand−protein complexes) were inspected and compared to the available structural data (crystal or NMR structures).

An all-atom model was used for docking by DOCK. In the DOCK 6.2 program suite, several different scoring methods, including energy, contact, continuum, and flexible ligand scoring, are available.[74] In the present study, we used the "rigid" ligand with the grid-based energy scoring as a primary scoring scheme. It was based on the nonbonded terms of the molecular mechanic force field (AMBER). Force field scores are approximate molecular mechanics interaction energies consisting of vdW and electrostatic components. During the docking preparation, the AM1-based BCC charges[75] as well as HF/6-31G*-based RESP charges[28] were assigned to both the protein and the ligands. A linear distance-dependent dielectric function ($\varepsilon(r) = 4r$) was used to mimic the solvent effects, and the cutoff for steric and electrostatic interactions was set to 12 Å.

Docking by AutoDock. In addition to DOCK, the AutoDock 4.2 code[76] has been employed. The docking was performed using the rigid protein model. The Gasteiger and AM1-BCC atomic charges were used. All of the parameters were set to the default or obtained automatically during the docking, except for the box dimensions, which were enlarged to 60 × 60 × 60 points because of the size of the ligands. The box was then centered in the HIV-1 PR cavity, where ligands bind. The Lamarckian Genetic Algorithm was used to obtain suitable ligand conformations. As recommended, 20 unique structures for each ligand have been assigned, because of the number of ligands' torsional degrees of freedom.

Rescoring. The two best-scoring complexes of each ligand obtained from the docking were selected and fully optimized using the PM6-DH2 method with the COSMO continuum water solvent model[48] implemented in the MOPAC code.[41] For the optimization, the FIRE procedure[77] was used with the following convergence criteria: ΔE of 0.03 kcal/mol, maxGrad 6.0, and rmsGrad 3.0. A deprotonated form of the HIV PR was used for PM6-DH2 rescoring.

The isolated ligand structure (in a water environment) was determined by MD/Q simulations using the PM6-DH2 method (with the continuum being water, temperature being 500 K, and using an Andersen thermostat). The total simulation time was 50 ps.

The hydration free energies of the ligands investigated were evaluated using the SMD method (HF/6-31G* level) implemented in the G09.[50]

The entropies of the protein−ligand complexes and of the isolated subsystems were determined by the rigid rotor-harmonic oscillator approximation based on the Cornell et al. empirical force field (with the Nucleic Acid Builder (NAB)[78] program) with the continuum solvent included. We used the FF03 force field[79] for the protein and the GAFF force field for the ligands combined with the generalized Born solvation model.[73] The complexes as well as the isolated species were optimized by the L-BFGS TNCG algorithm to the rms gradient of 10^{-6} and reoptimized by the Newton−Raphson method in the NAB.[78]

Data Analysis. The performance of the scoring schemes was evaluated in terms of their ability (i) to identify the near-native ligand binding modes (rmsd ≤2 Å) among a set of generated binding modes, and (ii) to rank different ligands correctly with respect to their interaction energies. The latter ability depends on the former, since it cannot be expected that the correct binding affinities would be obtained from the nonnative protein−ligand modes.

Results and Discussion

Docking and Scoring by DOCK. First, we have employed the DOCK procedure in order to test whether the scoring functions of DOCK are able to model the binding modes of those ligands that have been experimentally shown to form complexes with HIV PR. As a testing set of ligands, we used 11 known inhibitors of HIV protease, including nine clinically used compounds (saquinavir (SQV),[53] indinavir (IDV),[54] nelfinavir (NFV),[55] ritonavir (RTV),[56] lopinavir (LPV),[57] amprenavir (APV),[58] atazanavir (ATV),[59] darunavir (DRV),[60] acetyl-pepstatin (ACP),[61] a class-specific inhibitor of aspartic proteases, and two pseudopeptide inhibitors of HIV PR: Boc-Phe-Psi[(S)-CH(OH)CH2NH]-Phe-Gln-Phe-NH2 (SQ)[62] and Boc-Phe-Psi[CH2CH2NH]-Phe-Glu-Phe-NH2 (OE).[63] Furthermore, we

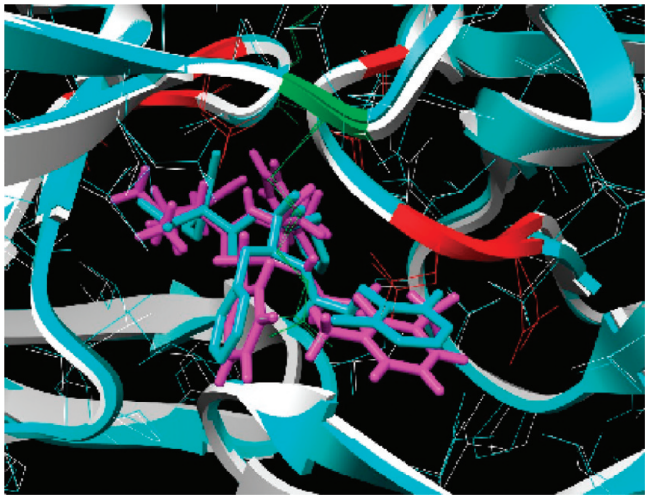


Figure 3. Comparison of the native and best binding modes obtained by DOCK for the HIV PR− LPV complex. The rmsd of the heavy atoms of LPV was 1.36 Å. The native geometry of the HIV PR−LPV complex represented by 1MUI[57] is compared with the geometry of the complex obtained by docking LPV into the HIV PR structure obtained from 2AQU.[59]

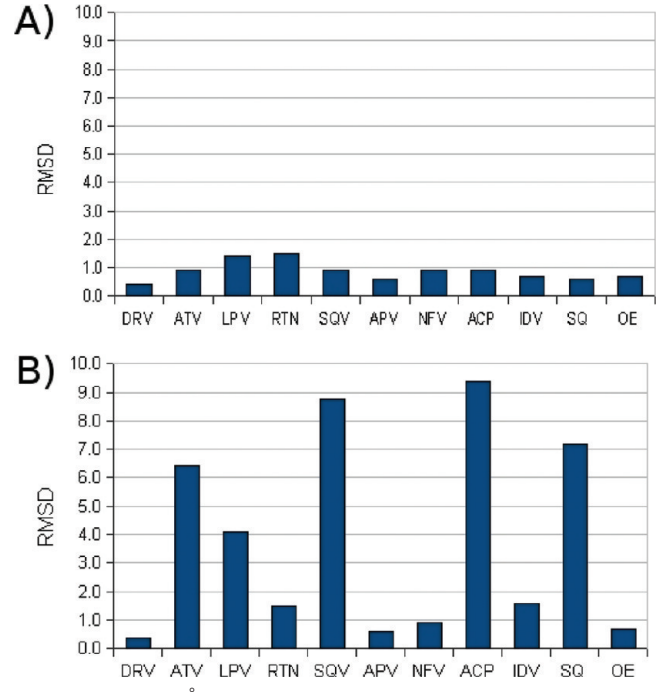


Figure 4. Rmsd (Å) of the docked structures: (A) the lowest reached rmsd and (B) the rmsd of the structures with the highest score.

used another 11 compounds representing "false" binders (for the complete compound list, see Table 1).

Both of the scoring functions implemented in DOCK[73] that we tested were able to reproduce well the "real" binding modes experimentally determined by protein crystallography: SQV 3CYX,[53] IDV 2BPX,[54] NFV 1OHR,[55] RTV 1HXW,[56] LPV 1MUI,[57] APV 1HPV,[58] ATV 2AQU,[59] DRV 1T3R,[60] ACP 5HVP,[61] SQ 1IIQ,[62] and OE 1M0D.[63] Surprisingly, flexible docking implemented in the latest versions of DOCK did not demonstrate any performance improvement in discriminating the "real" binding modes from nonnative ones. Nevertheless, for all the scoring functions, DOCK found the "true" binders (i.e., near-native), which means that their rmsd were lower than 2 Å. For a comparison, see Figure 3, which shows the native and the best binding mode obtained by the DOCK binding modes for all 11 of the ligands. However, the "true" binding modes did not correspond to the DOCK global energy minimum for five out of the 11 tested known HIV PR inhibitors (ATV, LPV, SQV, ACP, and SQ; see Figure 4B). It should be mentioned here that in all five of these cases the "true" binding mode had the second highest score and the difference between the score of the DOCK energy minimum and "true" binding mode was small (systematically below 10%). DOCK was also able to distinguish all of the real inhibitors of HIV PR (scores from −70 to −100 kcal/mol, see Figure 5A and Table 2) from the nonbinders (greater than −60 kcal/mol, specifically ARA, BOS, CEL, GAL, HEX, LSN, MON, OLM, SAM, ZAF, and ZED compounds had scores of −46.3, −38.3, −32.0, −28.6, −17.0, −38.3, −50.9, −56.4, −57, −61.4, and −44 kcal/mol, respectively). However, DOCK failed to rank the binders correctly. This is reflected in a very low correlation coefficient of 0.39 (we intentionally use R instead of R^2 in order to be consistent with the previous paper on HIV protease, see later). In Figure 5A, the scores of the HIV PR inhibitors are plotted against the experimental binding free energies. The three investigated protonation states of the catalytic aspartates, de-, mono-, and double-protonated (see Methods), yielded very similar results in our docking procedure. To summarize it, DOCK showed great potential and efficacy in the prediction of protein−ligand binding modes and in recognizing the real binders and was therefore used as an input in our rescoring scheme.

Docking by AutoDock. In contrast to DOCK, the performance of AutoDock[76] was less satisfactory. Only in the cases of DRV, APV, and NFV did the program generate "near-native" modes. The best rmsd for DRV was 2.0 Å, for APV 1.7 Å and for NFV 1.4 Å. For DRV and APV, the best fit structures were not the same as the best scored ones but lie in the first 10 of the 20 docked structures. As docking using Gasteiger charges did not give satisfying results, AM1-BCC charges were used, but with no improvement of docking reliability. Because of these poor results, AutoDock was not employed in our rescoring protocol, and only the binding modes obtained by DOCK were further rescored.

Full AMBER Rescoring. The structures found by DOCK were rescored using the AMBER force field.[78] The full energy minimization was performed in continuous solvent and the total score was constructed as a sum of interaction enthalpy, interaction entropy, and the deformation and desolvation energies of the ligand. The AMBER scores are plotted in Figure 5B. These results show that the "bare" AMBER rescoring failed to rank the binders correctly and that the obtained results are comparable to those obtained by the previously described methods. Nevertheless, it seems likely that a better sampling of the configuration space would improve the results.

PM6-DH2 Rescoring. The geometries generated by DOCK were optimized using the continuum solvent model (COSMO) as implemented in the MOPAC code[48] and rescored. Herein, however, only the electrostatic term has been included. First, the binding modes obtained by DOCK were rescored only with the interaction enthalpies obtained by PM6-DH2, disregarding the desolvation and deformation energies of the ligand and the entropy terms. When considering only the enthalpic term, the real binding modes generated by DOCK were already distinguished from the nonbinding ones. All the real binding modes had larger (more favorable) interaction energies than the nonnative modes. In the cases of ATV, LPV, SQV, ACP, and

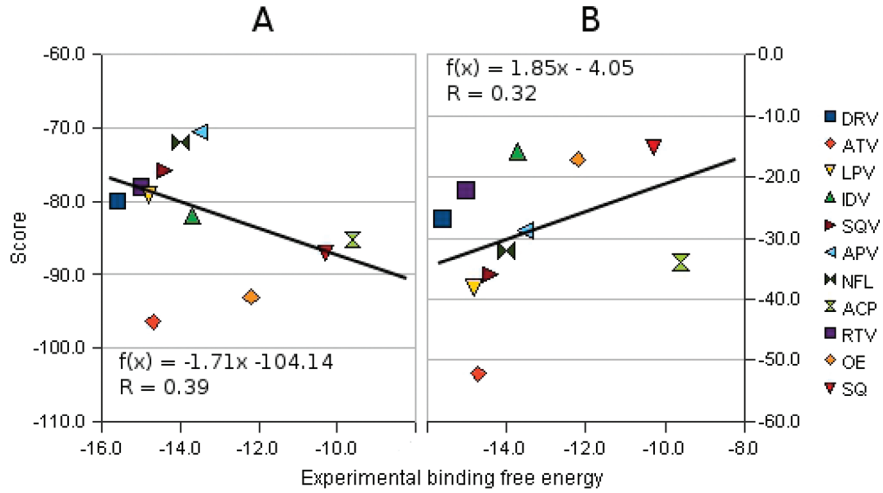


Figure 5. DOCK (A) and AMBER (B) scores for the HIV PR binders plotted against the binding free energies calculated from the experimental K_i values (Table 2); both sets of values in kcal/mol.

TABLE 2: Inhibition Constants (K_i) (nM) and the Derived Experimental Binding Free Energies ΔG_w (kcal/mol) of the Inhibitors of HIV-1 Protease[63,64] and This Study

inhibitor	K_i	ΔG_w
SQV	0.04	−14.4
IDV	0.12	−13.7
NFV	0.07	−14.0
RTV	0.02	−15.0
LPV	0.02	−14.8
APV	0.18	−13.5
ATV	0.02	−14.7
DRV	0.01	−15.6
ACP	116	−9.6
SQ	33	−10.3
OE	1.53	−12.2

SQ, where DOCK did not recognize the correct modes, the "true" binding modes were more stable than the global minimum of DOCK by about 12, 17, 12, 11, and 2 kcal/mol, respectively. Despite the enthalpic term alone distinguishing the correct binding modes from the nonnative ones, the correlation between the enthalpic term and the experimental ΔG_w was rather poor (Figure 6A). Obviously, the neglected terms were crucial, and despite the quality of the enthalpy term, the resulting scoring is poor. Experimental data from isothermal calorimetry show that, in the cases of all of the first-generation HIV-1 inhibitors, the affinity is entropy-driven.[80] When one of the two missing terms was added, the correlation improved significantly. Figure 6B,C shows the correlations where, in addition to the PM6-DH2 enthalpic term, the deformation and desolvation energies of the ligand and/or entropy terms were also added. In both cases, the correlation was improved, from 0.3 to 0.77 and 0.64, respectively. The full scores based on all the terms including the PM6-DH2 enthalpy, AMBER entropy, and the deformation and desolvation free energies are shown in Figure 7A. A high correlation between the predictions and experimental binding data with a correlation coefficient of 0.84 can be seen. The improved correlation enables the correct ranking of the ligands, which might lead to a quick and effective optimization of a lead compound. This demonstrates that taking all the relevant terms into consideration is mandatory for a reliable scoring as well as that the correlation coefficients, and hence the reliability of the predictions, are considerably poorer where some of these terms have been omitted. The correlation found is in fact very good if we take into account that (i) the structures of the studied protein−ligand complexes were obtained by docking and not taken from crystallography and that (ii) the scoring function does not contain any empirical parameter adjustable to a ligand or a protein, as a result of which exactly the same procedure can be used for diverse protein−ligand complexes. The very high correlation (0.93) between the experimental IC_{50} data and the results by free energy simulations with polarizable quantum-mechanics-based ligand charges was found[81] for HIV-1 protease. However, the comparison was performed only for five structurally similar ligands (e.g., single modification, as in DRV vs APV) with similar binding energies.

In this study, the deformation and desolvation energies of ligands were obtained by using an MD/Q technique. Optimization using only PM6-DH2 in a water environment instead of full MD/Q was also attempted. However, the structures obtained did not correspond to their global minima, and the values obtained from the MD/Q simulations of the ligand followed by an averaging of the optimized structures were more reliable. For most ligands, however, the difference in deformation energy is rather small and the correlation coefficient is only slightly lower ($R = 0.79$).

The deformation free energy of the protein should be considered and was calculated as the difference in the energy of the protein in the protein−ligand complex geometry and in the minimized geometry. The consideration of the protein deformation changed the results only slightly as anticipated (see Figure 7B and cf. Figure 7A). The protein deformation might, however, be of crucial importance when various proteins or mutated variants of the protein are considered.

In the algorithm used in this study, the three basic terms considered were calculated at different levels. In particular, the enthalpy terms were determined at the PM6-DH2 level, whereas the entropy terms were evaluated using the rigid rotor−harmonic oscillator approximation with the AMBER characteristics. The deformation energies were calculated in a similar manner to the enthalpy term at the PM6-DH2 level, and the desolvation energies were determined using the SMD method, considering the electrostatic as well as nonelectrostatic terms. This was realized by replacing the MOPAC COSMO (electrostatic term only) desolvation with the SMD one. This inconsistent treatment was introduced because the SMD method is better suited for the description of solvation energy in neutral as well as charged systems. When the desolvation energy of ligand was determined consistently using the MOPAC procedure, the correlation was similar to the previous case, with the difference being minor.

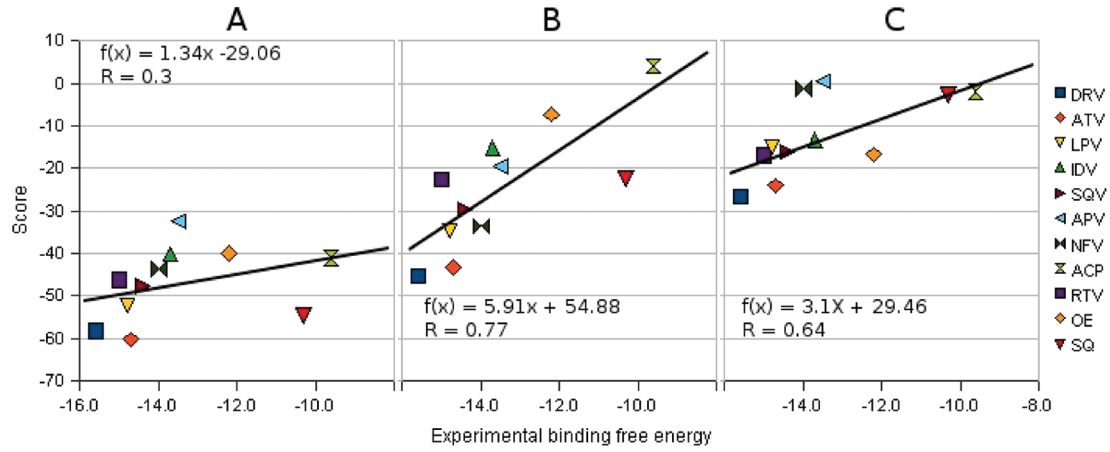


Figure 6. Partial scores for the HIV PR binders plotted against the binding free energies calculated from the experimental K_i values (Table 2): (A) with only binding enthalpy determined by PM6-DH2 considered and (B) with the vibrational entropy being omitted; (C) with the deformation free energy of the ligands being omitted; all in kcal/mol.

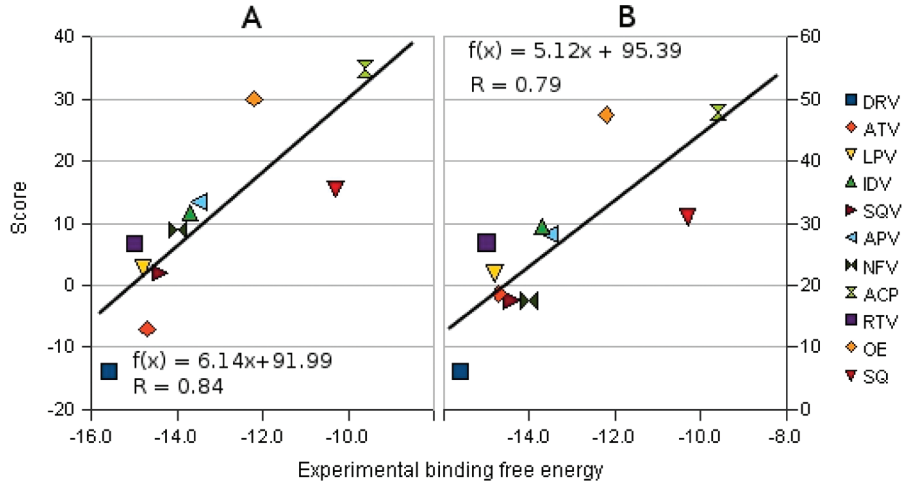


Figure 7. PM6-DH2 (SMD/HF) scores for the HIV PR binders plotted against the experimental binding free energies: (A) with the deformation energy of the protein not being considered; (B) with the deformation energy of the protein being considered.

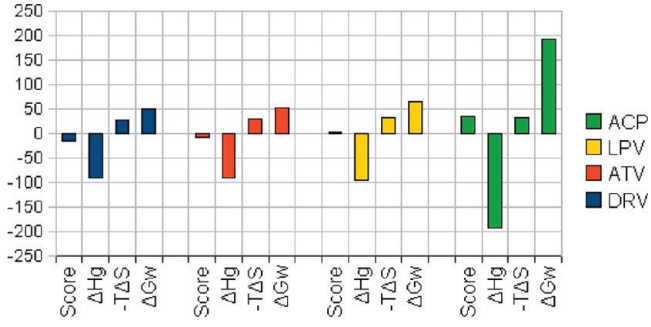


Figure 8. Decomposition of the scores into computationally accessible physical terms: the vacuum interaction enthalpy (ΔH_g), desolvation and deformation free energies (ΔG_w), and entropy contributions ($-T\Delta S$) into total scores in kcal/mol.

This finding indicates that for the set of ligands considered, the desolvation energy is determined sufficiently well already using merely the electrostatic term. The agreement between both procedures also indicates that using the theoretically more reliable solvent model also provides reliable data, and in those cases where charged, polar and nonpolar ligands are involved, it can be used with confidence.

Figure 8 shows the single contributions to the total PM6-DH2 scoring function. The final score is constructed as a sum of the vacuum interaction enthalpy, the vacuum interaction entropy term, and the desolvation and deformation free energies of the ligand. The free energy contributions shown in Figure 8 do not correspond to the contributions of the binding free energy obtained by ITC, and a direct comparison is not possible. The ITC data show the "entropic" and "enthalpic" contributions as de facto a global parameter of the complex formation while total scoring was constructed differently, as described above.

Among the ligands shown in Figure 8, only one, zwitterion ACP, is charged, whereas the others are electroneutral. We found that the vacuum interaction enthalpies represented the largest (attractive) term, reaching −100 kcal/mol. For the majority of ligands, the values of this term were in the range of −100 to −70 kcal/mol. The remaining contributions were smaller and systematically repulsive, with the desolvation and deformation free energies being larger. In the case of the zwitterionic system (ACP), the interaction enthalpy as well as the deformation energies and desolvation free energies of the ligand were considerably larger than those terms for neutral ligands, and they were found to compensate roughly. After addition of the repulsive interaction entropy term, the final score was thus slightly repulsive. We can conclude that the interaction enthalpy term is the most important, but the other two terms cannot be omitted. The very large stabilization enthalpy for the charged ligands can be compensated by the large repulsive desolvation and deformation free energy term.

The protonated and neutral forms of another PR inhibitor, NFV, yielded similar results (the protonated form of NFV had

SCHEME 1: Thermodynamic Cycle Used for the Estimation of the Binding Free Energy of the Protein−Ligand Complex in Water Environment

	ΔG_g				
P	+	L	↔	(P…L)	gas phase
↓ ΔG_w(P)		↓ ΔG_w(L)		↓ ΔG_w (PL)	hydration
	ΔG_w				
P_w	+	L_w	↔	$(P…L)_w$	water environment

a more negative score, by about 4 kcal/mol) and therefore only the results for the docking of neutral NFV are shown.

Timing. Rescoring of a single protein−ligand complex takes approximately 3 or 4 days on a single processor. Most of the time is spent on geometry optimization. Evaluation of the score on the minimized geometry is relatively fast and does not take more than a few hours. We are working on speeding up the optimization procedure by using QM/MM (with ∼1000 atoms in the QM region treated by SQM method); this will be again followed by single-point SQM calculation of the whole system. This setup allows us to calculate the score in less than 1 day.

Further Outlook. We are currently using this protocol in several other projects, and the preliminary results suggest that it is highly transferable and performs very well even for other systems. In a study of CDK2 with a set of 12 inhibitors,[82] the correlation of the calculated score with experiment is even closer than in this study. We have used the same protocol for the evaluation of the novel modifications of HIV protease inhibitors[83] too, and significant improvement of the total score was achieved for several ligand modifications. The best compound is being synthesized and tested.

Conclusion

(i) The rescoring by the PM6-DH2 method improved the docking results dramatically. The enthalpic term calculated at the PM6-DH2 level (i.e., while ignoring the desolvation and deformation energies of the ligand as well as the entropy terms) already discarded the nonnative binding modes generated by DOCK. All of the real binding modes had larger (more favorable) interaction energies than the nonnative ones.

(ii) The total scores based on all the terms including the PM6-DH2 enthalpy, AMBER entropy, and the deformation and desolvation energies show a high correlation between the predictions and the experimental binding data. The most efficient ligand had the most negative score while the least efficient ligand had the least negative score. The interaction enthalpy is systematically attractive and the most important term. The other two terms, which are repulsive, can, however, never be omitted. Whenever one of these two terms was neglected, much worse correlation resulted. The magnitude of these terms varies significantly for individual ligands. The main advantage of this procedure is that we do not add any empirical parameter either for an individual component of the total score or for an individual protein−ligand complex. Consequently, the identical procedure can be used for diverse protein−ligand complexes.

(iii) Docking by DOCK followed by rescoring by PM6-DH2 thus represents a very promising tool for fast and reliable rational drug discovery and is currently being used in our laboratory for a broad spectrum of drug design projects.

Acknowledgment. We thank Dr. David A. Evans from the University of Leeds for editorial help and a great deal of inspiring discussions, suggestions, and comments; Milan Kožíšek for the determination of the inhibition constant for acetyl-pepstatin; Klára Šašková and Martin Lepšík for fruitful comments; and Prof. Steve Homans for providing access to the computing facilities. This work was a part of the research project No. Z40550506 of the Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, and it was supported by Grants No. LC512, MSM6198959216, and 1M0508 from the Ministry of Education, Youth and Sports of the Czech Republic. The support of Praemium Academiae, Academy of Sciences of the Czech Republic, awarded to P.H. in 2007, is also acknowledged. It was also supported by Korea Science and Engineering Foundation (World Class University Program R32-2008-000-10180-0).

References and Notes

(1) Cho, Y.; Ioerger, T. R.; Sacchettini, J. C. Discovery of novel nitrobenzothiazole inhibitors for Mycobacterium tuberculosis ATP phosphoribosyl transferase (HisG) through virtual screening. *J. Med. Chem.* **2008**, *51*, 5984–5992.

(2) Kiss, R.; Kiss, B.; Konczol, A.; Szalai, F.; Jelinek, I.; Laszlo, V.; Noszal, B.; Falus, A.; Keseru, G. M. Discovery of novel human histamine H4 receptor ligands by large-scale structure-based virtual screening. *J. Med. Chem.* **2008**, *51*, 3145–3153.

(3) Cavasotto, C. N.; Orry, A. J.; Murgolo, N. J.; Czarniecki, M. F.; Kocsi, S. A.; Hawes, B. E.; O'Neill, K. A.; Hine, H.; Burton, M. S.; Voigt, J. H.; Abagyan, R. A.; Bayne, M. L.; Monsma, F. J., Jr. Discovery of novel chemotypes to a G-protein-coupled receptor through ligand-steered homology modeling and structure-based virtual screening. *J. Med. Chem.* **2008**, *51*, 581–588.

(4) Kolb, P.; Rosenbaum, D. M.; Irwin, J. J.; Fung, J. J.; Kobilka, B. K.; Shoichet, B. K. Structure-based discovery of beta(2)-adrenergic receptor ligands. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 6843–6848.

(5) Bohm, H. J. The development of a simple empirical scoring function to estimate the binding constant for a protein ligand complex of known 3-dimensional structure. *J. Comput. Aided Mol. Des.* **1994**, *8*, 243–256.

(6) Jain, A. N. Scoring noncovalent protein-ligand interactions: A continuous differentiable function tuned to compute binding affinities. *J. Comput. Aided Mol. Des.* **1996**, *10*, 427–440.

(7) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical scoring functions 0.1. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput. Aided Mol. Des.* **1997**, *11*, 425–445.

(8) Muegge, I.; Martin, Y. C. A general and fast scoring function for protein-ligand interactions: A simplified potential approach. *J. Med. Chem.* **1999**, *42*, 791–804.

(9) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nat. Rev. Drug. Discov.* **2004**, *3*, 935–949.

(10) Kapetanovic, I. M. Computer-aided drug discovery and development (CADDD): In silico-chemico-biological approach. *Chem. Biol. Interact.* **2008**, *171*, 165–176.

(11) Almerico, A. M.; Tutone, M.; Lauria, A. Docking and multivariate methods to explore HIV-1 drug-resistance: a comparative analysis. *J. Comput. Aided Mol. Des.* **2008**, *22*, 287–297.

(12) Bakan, A.; Bahar, I. The intrinsic dynamics of enzymes plays a dominant role in determining the structural changes induced upon inhibitor binding. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 14349–1454.

(13) Evans, D. A.; Bronowska, A. K. Implications of fast-time scale dynamics of human DNA/RNA cytosine methyltransferases (DNMTs) for protein function. *Theor. Chim. Acta* **2010**, *125*, 407–418.

(14) MacRaild, C. A.; Daranas, A. H.; Bronowska, A.; Homans, S. W. Global changes in local protein dynamics reduce the entropic cost of carbohydrate binding in the arabinose-binding protein. *J. Mol. Biol.* **2007**, *368*, 822–832.

(15) Stoeckmann, H.; Bronowska, A.; Syme, N. R.; Thompson, G. S.; Kalverda, A. P.; Warriner, S. L.; Homans, S. W. Residual ligand entropy in the binding of p-substituted benzenesulfonamide ligands to bovine carbonic anhydrase II. *J. Am. Chem. Soc.* **2008**, *130*, 12420–12426.

(16) Aqvist, J.; Medina, C.; Samuelsson, J. E. New method for prediction binding-affinity in computer-aided drug design. *Protein Eng.* **1994**, *7*, 385–391.

(17) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Hui, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. Calculating structures and free

energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33*, 889–897.

(18) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices. *J. Am. Chem. Soc.* **1998**, *120*, 9401–9409.

(19) Lamb, M. L.; Jorgensen, W. L. Computational approaches to molecular recognition. *Curr. Opin. Chem Biol.* **1997**, *1*, 449–457.

(20) Wang, J. M.; 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.

(21) Wang, R.; Lu, Y.; Wang, S. Application of frameless stereotaxy in craniotomy procedures: Clinical evaluation. *J. Med. Chem.* **2003**, *46*, 2287–2303.

(22) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* **2006**, *49*, 6177–6196.

(23) Riley, K. E.; Pitonak, M.; Jurecka, P.; Hobza, P. Stabilization and structure calculations for noncovalent interactions in extended molecular systems based on wave function and density functional theories. *Chem. Rev.* **2010**, *110*, 0000.

(24) Raha, K.; Merz, K. M., Jr. Large-scale validation of a quantum mechanics based scoring function: Predicting the binding affinity and the binding mode of a diverse set of protein-ligand complexes. *J. Med. Chem.* **2005**, *48*, 4558–4575.

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

(26) Zhou, T.; Huang, D.; Caflisch, A. Is quantum mechanics necessary for predicting binding free energy. *J. Med. Chem.* **2008**, *51*, 4280–4288.

(27) Valdes, H.; Pluhackova, K.; Pitonak, M.; Rezac, J.; Hobza, P. Benchmark database on isolated small peptides containing an aromatic side chain: comparison between wave function and density functional theory methods and empirical force field. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2747–2757.

(28) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges—the RESP model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.

(29) Lu, Y.; Shi, T.; Wang, Y.; Yang, H.; Yan, X.; Luo, X.; Jiang, H.; Zhu, W. Halogen Bonding-A Novel Interaction for Rational Drug Design. *J. Med. Chem.* **2009**, *52*, 2854–2862.

(30) Sedlak, R. Personal communication.

(31) Menikarachchi, L.; Gascon, J. A. QM/MM Approaches in Medicinal Chemistry Research. *Curr. Top. Med. Chem.* **2010**, *10*, 46–54.

(32) Fong, P.; McNamara, J. P.; Hillier, I. H.; Bryce, R. A. Assessment of QM/MM Scoring Functions for Molecular Docking to HIV-1 Protease. *J. Chem. Inf. Model.* **2009**, *49*, 913–924.

(33) Cavalli, A.; Carloni, P.; Recanatini, M. Target-related applications of first principles quantum chemical methods in drug design. *Chem. Rev.* **2006**, *106*, 3497–3519.

(34) Khandelwal, V.; Lukacova, D.; Comez, D. M.; Kroll, S.; Raha, S.; Balaz, S. A combination of docking, QM/MM methods, and MD simulation for binding affinity estimation of metalloprotein ligands. *J. Med. Chem.* **2005**, *48*, 5437–5447.

(35) Raha, K.; Peters, M. B.; Wang, B.; Yu, N.; Wollacott, A. M.; Westerhoff, L. M.; Merz, K. M., Jr. The role of quantum mechanics in structure-based drug design. *Drug Discov. Today* **2007**, *12*, 725–731.

(36) Zhou, T.; Caflisch, A. High-throughput virtial screening using quantum mechanical probes: discovery of selective kinase inhibitors. *Chem. Med. Chem.* 10.1002/cmdc.201000085.

(37) Wollacott, A. M.; Merz, Jr., K. M. Assessment of semiempirical quantum mechanical methods for the evaluation of protein structures *J. Chem. Theory Comput.* **2007**, *3*, 1609−1619.

(38) Antony, J.; Grimme, S. Density functional theory including dispersion corrections for intermolecular interactions in a large benchmark set of biologically relevant molecules. *Phys. Chem. Chem. Phys.* **2006**, *8*, 5287–5293.

(39) Jurecka, P.; Cerny, J.; Hobza, P.; Salahub, D. Density functional theory augmented with an empirical dispersion term. Interaction energies and geometries of 80 noncovalent complexes compared with ab initio quantum mechanics calculations. *J. Comput. Chem.* **2007**, *28*, 555–569.

(40) Stewart, J. J. P. *J. Mol. Model.* **2007**, *13*, 1173–1213.

(41) Stewart, J. J. P. Stewart Computational Chemistry, Colorado Springs, CO, MOPAC2009; http://OpenMOPAC.net.

(42) Stewart, J. J. P. Application of the PM6 method to modeling proteins. *J. Mol. Model.* **2009**, *15*, 765–805.

(43) Rezac, J.; Fanfrlik, J.; Salahub, D.; Hobza, P. Semiempirical Quantum Chemical PM6Method Augmented by Dispersion and H-Bonding Correction Terms Reliably Describes Various Types of Noncovalent Complexes. *J. Chem. Theory Comput.* **2009**, *5*, 1749–1760.

(44) Korth, M.; Pitonak, M.; Rezac, J.; Hobza, P. A Transferable H-Bonding Correction for Semiempirical Quantum-Chemical Methods. *J. Chem. Theory Comput.* **2010**, *6*, 344–352.

(45) Wlodawer, A. Rational approach to AIDS drug design through structural biology. *Annu. Rev. Med.* **2002**, *53*, 595–614.

(46) Mastrolorenzo, A.; Rusconi, S.; Scozzafava, A.; Barbaro, G.; Supuran, C. T. Inhibitors of HIV-1 protease: Current state of the art 10 years after their introduction. from Antiretroviral drugs to antifungal, antibacterial and antitumor agents based on aspartic protease inhibitors. *Curr. Med. Chem.* **2007**, *14*, 2734–2748.

(47) Pokorná, J.; Rezacova, P.; Machala, L.; Konvalinka, J. Current and novel inhibitors of HIV protease. *Viruses* **2009**, *1*, 1209–1239.

(48) Klamt, A.; Schuurmann, G. COSMO—a new approach to dielectric screening in solvents with explicit expressions for the screening energz and ist gradient. *J. Chem. Soc., Perkin Trans.* **1993**, *2*, 799–805.

(49) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113*, 6378–96.

(50) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, Revision A.1*; Gaussian, Inc.: Wallingford, CT, 2009.

(51) Chang, C. E.; Chen, W.; Gilson, M. K. Ligand configurational entropy and protein binding. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1534–1539.

(52) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A 2nd generation froce-field for the simulation of proteins, nucleic-acids, and organic-molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179.

(53) Liu, F.; Kovalevsky, A. Y.; Tie, Y.; Ghosh, A. K.; Harrison, R. W.; Weber, I. T. Effect of flap mutations on structure of HIV-1 protease and inhibition by saquinavir and darunavir. *J. Mol. Biol.* **2008**, *381*, 102–115.

(54) Munshi, S.; Chen, Z.; Li, Y.; Olsen, D. B.; Fraley, M. E.; Hungate, R. W.; Kuo, L. C. Rapid x-ray diffraction analysis of HIV-1 protease-inhibitor complexes: inhibitor exchange in single crystals of the bound enzyme. *Acta Crystallogr. D, Biol. Crystallogr.* **1998**, *54*, 1053–1060.

(55) Kaldor, S. W.; Kalish, V. J.; Davies II, J. F.; Shetty, B. V.; Fritz, J. E.; Appelt, K.; Burgess, J. A.; Campanale, K. M.; Chirgadze, N. Y.; Clawson, D. K.; Dressman, B. A.; Hatch, S. D.; Khalil, D. A.; Kosa, M. B.; Lubbehusen, P. P.; Muesing, M. A.; Patick, A. K.; Reich, S. H.; Su, K. S.; Tatlock, J. H. Viracept (nelfinavir mesylate, AG1343): A potent, orally bioavailable inhibitor of HIV-1 protease. *J. Med. Chem.* **1997**, *40*, 3979–3985.

(56) Kempf, D. J.; Marsh, K. C.; Denissen, J. F.; McDonald, E.; Vasavanonda, S.; Flentge, C. A.; Green, B. E.; Fino, L.; Park, C. H.; Kong, X. P. ABT-538 is a potent inhibitor of human-immunodeficiency-virus protease and has high oral bioavailability in humans. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 2484–2488.

(57) Stoll, V.; Qin, W.; Stewart, K. D.; Jakob, C.; Park, C.; Walter, K.; Simmer, R. L.; Helfrich, R.; Bussiere, D.; Kao, J.; Kempf, D.; Sham, H. L.; Norbeck, D. W. X-ray crystallographic structure of ABT-378 (lopinavir) bound to HIV-1 protease. *Bioorg. Med. Chem.* **2002**, *10*, 2803–2806.

(58) Kim, E. E.; Baker, C. T.; Dwyer, M. D.; Murcko, M. A.; Rao, B. G.; Tung, R. D.; Navia, M. A. Crystal-structure of HIV-1 protease in complex with VX-478, a potent and orally bioavailable inhibitor of the enzyme. *J. Am. Chem. Soc.* **1995**, *117*, 1181–1182.

(59) Clemente, J. C.; Coman, R. M.; Thiaville, M. M.; Janka, L. K.; Jeung, J. A.; Nukoolkarn, S.; Govindasamy, L.; Agbandje-McKenna, M.; McKenna, R.; Leelamanit, W.; Goodenow, M. M.; Dunn, B. M. Analysis of HIV-1CRF_01 A/E protease inhibitor resistance: Structural determinants for maintaining sensitivity and developing resistance to atazanavir. *Biochemistry* **2006**, *45*, 5468–5477.

(60) Surleraux, D. L. N. G.; Tahri, A.; Verschueren, W. G.; Pille, G. M. E.; de Kock, H. A.; Jonckers, T. H. M.; Peeters, A.; De Meyer, S.; Azijn, H.; Pauwels, R.; de Bethune, M. P.; King, N. M.; Prabu-Jeyabalan, M.; Schiffer, C. A.; Wigerinck, P. B. T. P. Discovery and selection of TMC114, a next generation HIV-1 protease inhibitor. *J. Med. Chem.* **2005**, *48*, 1813–1822.

(61) Fitzgerald, P. M.; McKeever, B. M.; VanMiddlesworth, J. F.; Springer, J. P.; Heimbach, J. C.; Leu, C. T.; Herber, W. K.; Dixon, R. A.; Darke, P. L.; Fitzgerald, P. M.; McKeever, B. M.; VanMiddlesworth, J. F.;

Springer, J. P.; Heimbach, J. C.; Leu, C. T.; Herber, W. K.; Dixon, R. A.; Darke, P. L. Crystallographic analysis of a complex between human-immunodeficiency-virus type-1 protease and acetyl-pepstatin at 2.0-A resolution. *J. Biol. Chem.* **1990**, *265*, 14209–19.

(62) Dohnalek, J.; Hasek, J.; Duskova, J.; Petrokova, H.; Hradilek, M.; Soucek, M.; Konvalinka, J.; Brynda, J.; Sedlacek, J.; Fabry, M. Hydroxyethylamine isostere of an HIV-1 protease inhibitor prefers its amine to the hydroxy group in binding to catalytic aspartates. A synchrotron study of HIV-1 protease in complex with a peptidomimetic inhibitor. *J. Med. Chem.* **2002**, *45*, 1432–8.

(63) Petrokova, H.; Duskova, J.; Dohnalek, J.; Skalova, T.; Vondracková-Buchtelova, E.; Soucek, M.; Konvalinka, J.; Brynda, J.; Fabry, M.; Sedlacek, J.; Hasek, J. Role of hydroxyl group and R/S configuration of isostere in binding properties of HIV-1 protease inhibitors. *Eur. J. Biochem.* **2004**, *271*, 4451–61.

(64) Urban, J.; Konvalinka, J.; Stehlíková, J.; Gregorová, E.; Majer, P.; Soucek, M.; Andreánsky, M.; Fábry, M.; Strop, P. Reduced-bond tight-binding inhibitors of HIV-1 protease—fine tuning of the enzyme subsite specificity. *FEBS Lett.* **1992**, *298*, 9–13.

(65) Majer, P.; Urban, J.; Gregorová, E.; Konvalinka, J.; Novek, P.; Stehlíková, J.; Andreansky, M.; Sedlacek, J.; Strop, P. Specificity mapping of HIV-1 protease by reduced bond inhibitors. *Arch. Biochem. Biophys.* **1993**, *304*, 1–8.

(66) Konvalinka, J.; Litera, J.; Weber, J.; Vondrásek, J.; Hradílek, M.; Soucek, M.; Pichová, I.; Majer, P.; Strop, P.; Sedlacek, J.; Heuser, A. M.; Kottler, H.; Kräusslich, H. G. Configurations of diastereomeric hydroxyethylene isosteres strongly affect biological activities of a series of specific inhibitors of human-immunodeficiency-virus proteinase. *Eur. J. Biochem.* **1997**, *250*, 559–566.

(67) Kozisek, J.; Bray, J.; Rezacova, P.; Saskova, K.; Brynda, J.; Pokorna, J.; Mammano, F. Molecular analysis of the HIV-1 resistance development: Enzymatic activities, crystal structures, and thermodynamics of nelfinavir-resistant HIV protease mutants. *J. Mol. Biol.* **2007**, *374*, 1005–1016.

(68) Ode, H.; Neya, S.; Hata, M.; Sugiura, W.; Hoshino, T. Computational simulations of HIV-1 proteases-multi-drug resistance due to nonactive site mutation L90M. *J. Am. Chem. Soc.* **2006**, *128*, 7887–7895.

(69) Richards, A. D.; Phylip, L. H.; Farmerie, W. G.; Scarborough, P. E.; Alvarez, A.; Dunn, B. M.; Hirel, P. H.; Konvalinka, J.; Strop, P.; Pavlickova, L.; Kostka, V.; Kay, J. Sensitive soluble chromogenic substrates for HIV-1 proteinase. *J. Biol. Chem.* **1990**, *265*, 7733–7736.

(70) Dixon, M. Determination of enzyme-inhibitor con- stants. *Biochem. J.* **1953**, *55*, 170–171.

(71) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF chimera—A visualization system for exploratory research and analysi. *J. Comput. Chem.* **2004**, *25*, 1605–1612.

(72) Li, H.; Robertson, A. D.; Jensen, J. H. Very fast empirical prediction and rationalization of protein pK(a) values. *Proteins* **2005**, *61*, 704–721.

(73) Wang, J. M.; 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.

(74) Ewing, T. J. A.; Kuntz, I. D. Critical evaluation of search algorithms for automated molecular docking and database screening. *J. Comput. Chem.* **1997**, *18*, 1175–1189.

(75) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002**, *23*, 1623–1641.

(76) Goodsell, D. S.; Morris, G. M.; Olson, A. J. Automated docking of flexible ligands: Applications of AutoDock. *J. Mol. Recognit.* **1996**, *9*, 1–5.

(77) Bitzek, E.; Koskinen, P.; Gahler, F.; Moseler, M.; Gumbsch, P. Structural relaxation made simple. *Phys. Rev. Lett.* **2006**, *97*, 170201.

(78) Case, D. A.; Darden, T. A.; Cheatham, III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. *AMBER 10*; University of California: San Francisco, 2008.

(79) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24*, 1999–2012.

(80) Freire, E. Do enthalpy and entropy distinguish first in class from best in class? *Drug Discov. Today* **2008**, *13*, 869–874.

(81) Das, D.; Koh, Y.; Tojo, Y.; Ghosh, A. K.; Mitsuya, H. Prediction of Potency of Protease Inhibitors Using Free Energy Simulations with Polarizable Quantum Mechanics-Based Ligand Charges and a Hybrid Water Model. *J. Chem. Inf. Model.* **2009**, *49*, 2851–2862.

(82) Dobes, P.; Fanfrlik, J.; Rezac, J.; Otyepka, M.; Hobza, P. Manuscript in preparation.

(83) Fanfrlik, J.; Rezac, J.; Hobza, P. Manuscript in preparation.

JP1032965