DOI: 10.1002/chem.200700040

# A Quantum Mechanics/Molecular Mechanics Study of the Protein–Ligand Interaction for Inhibitors of HIV-1 Integrase

**Cláudio N. Alves,*[a] Sergio Martí,[b] Raquel Castillo,[b] Juan Andrés,[b] Vicent Moliner,*[b] Iñaki Tuñón,[c] and Estanislao Silla[b, c]**

**Abstract:** Human immunodeficiency virus type-1 integrase (HIV-1 IN) is an essential enzyme for effective viral replication. Diketo acids such as L-731,988 and S-1360 are potent and selective inhibitors of HIV-1 IN. In this study, we used molecular dynamics simulations, within the hybrid quantum mechanics/molecular mechanics (QM/MM) approach, to determine the protein–ligand interaction energy between HIV-1 IN and L-731,988 and 10 of its derivatives and analogues. This hybrid methodology has the advantage that it includes quantum effects such as ligand polarisation upon binding, which can be very important when highly polarisable groups are embedded in anisotropic environments, as for example in metal-containing active sites. Furthermore, an energy decomposition analysis was performed to determine the contributions of individual residues to the enzyme–inhibitor interactions on averaged structures obtained from rather extensive conformational sampling. Analysis of the results reveals first that there is a correlation between protein–ligand interaction energy and experimental strand transfer into human chromosomes and secondly that the Asn-155, Lys-156 and Lys-159 residues and the $Mg^{2+}$ ion are crucial to anti-HIV IN activity. These results may explain the available experimental data.

**Keywords:** diketo acids · HIV-1 integrase · inhibitors · molecular dynamics · quantum mechanics/molecular mechanics

## Introduction

The human immunodeficiency virus type-1 (HIV-1) virus has the capability of selectively infecting and ultimately incapacitating the immune system. The global acquired immunodeficiency syndrome (AIDS) killed 3.1 million people in 2005 and the estimated number of infected people living in the world was 39.4 million, becoming one of the most serious world health problems.[1–3]

Three enzymes are essential for the replication cycle of HIV-1: reverse transcriptase (RT), protease (PR) and integrase (IN). These enzymes are important targets in the development of anti-AIDS drugs.[4] Anti-retroviral chemotherapy in the treatment of AIDS is based on the use of a combination of drugs that inhibit RT, PR and the fusion of the virus with the host cell.[5] Because of problems of drug toxicity and the development of drug resistance, HIV-1 IN has recently received much attention as a possible target in anti-AIDS drug design because of its crucial role in the viral life

[a] Dr. C. N. Alves
Departamento de Química, Centro de Ciências Exatas e Naturais
Universidade Federal do Pará
CP 11101, 66075-110, Belém, PA (Brazil)
Fax: (+55) 3201-1635
E-mail: nahum@ufpa.br

[b] Dr. S. Martí, Dr. R. Castillo, Prof. J. Andrés, Dr. V. Moliner, Prof. E. Silla
Departament de Química Física i Analítica
Universitat Jaume I, 12071, Castellón (Spain)
Fax: (+34) 964-728-066
E-mail: moliner@uji.es

[c] Dr. I. Tuñón, Prof. E. Silla
Departament de Química Física, Universitat de Valencia
46100 Burjasot, Valencia (Spain)

Supporting information for this article is available on the WWW under http://www.chemeurj.org/ or from the author. Figure S1: Representation of the most important interactions between HIV-1 IN and the nine selected compounds not shown in Figure 1. Figure S2: Population analysis of all the compounds. Figure S3: Representation of the different relative orientations of the $Mg^{2+}$ ion between HIV-1 IN and compound **A** and compound **F**, respectively, computed at the AM1/MM level of theory. Figure S4: Time evolution of the interaction energy between the protein and the 11 inhibitors during QM/MM molecular dynamics during the last 100 ps of the simulations. Figure S5: Representation of the most important interactions between HIV-1 IN and compounds **A** and **F**, computed at the DFT/MM level of theory. Table S1: Averaged hydrogen-bond distances (Å) between the tetrazole or carboxylic ring of each ligand and the hinge region of the integrase active site.

cycle and the fact that there is no cellular homologue in humans.[6–8] IN catalyses the integration of viral DNA into the human genome, which facilitates stable viral replication and sustained infection.[4–6] In vitro studies using recombinant protein have permitted the distinction between the two main reactions catalysed by IN: cleavage of a dinucleotide pair at the 3′-end of viral DNA (named 3′-processing) and insertion of the shortened strands into the human chromosome (strand transfer).[4] Both reactions involve a nucleophilic attack of a hydroxy group on a phosphodiester bond with participation of a divalent metal ion ($Mg^{2+}$ or $Mn^{2+}$) as cofactor.[4] However, the mechanism of inhibition in both processes is still unclear.

Previous studies were designed to identify the key interactions associated with inhibitory processes.[9–16] Structure–activity relationship (SAR) studies have illustrated that an aromatic portion is crucial for the action of many inhibitors.[9–13] There is also a possible charge–charge interaction between the metal ions and the ionic or partial charges of the ligands.[9–15] Other work has also suggested a binding mode in which inhibitors bind through bidentate chelation of this ion,[11,13,14,16] and recently Pommier et al. also proposed that some inhibitors be classified as interfacial inhibitors.[17]

A number of compounds bearing a diketo acid (DKA) moiety have been independently discovered by scientists from Shionogi & Co. Ltd[18] and Merck Research Laboratories.[19] These compounds represent one of the most promising classes of IN inhibitors in terms of potency and selectivity. Shionogi's DKA, 5CITEP (with PDB code 1QS4), crystallises in the active site of the catalytic domain;[18] it is a well-known bioisoester of a carboxylic acid.[20] Several molecular dynamics (classical MD)[21] and docking[22] studies of HIV-1 IN complexed to the inhibitor 5CITEP and its derivatives have been published. Recently, Briggs and co-workers[23] performed classical MD simulations on the inhibitor L-731,988. These simulations demonstrated that the most favourable orientation of the inhibitor L-731,988 is similar to the one adopted by 5CITEP in the crystal structure. More recently, some bifunctional DKA inhibitors, characterised by the presence of two diketo acid chains, were reported[10,11,24] and their mechanism of action investigated by docking techniques.[25]

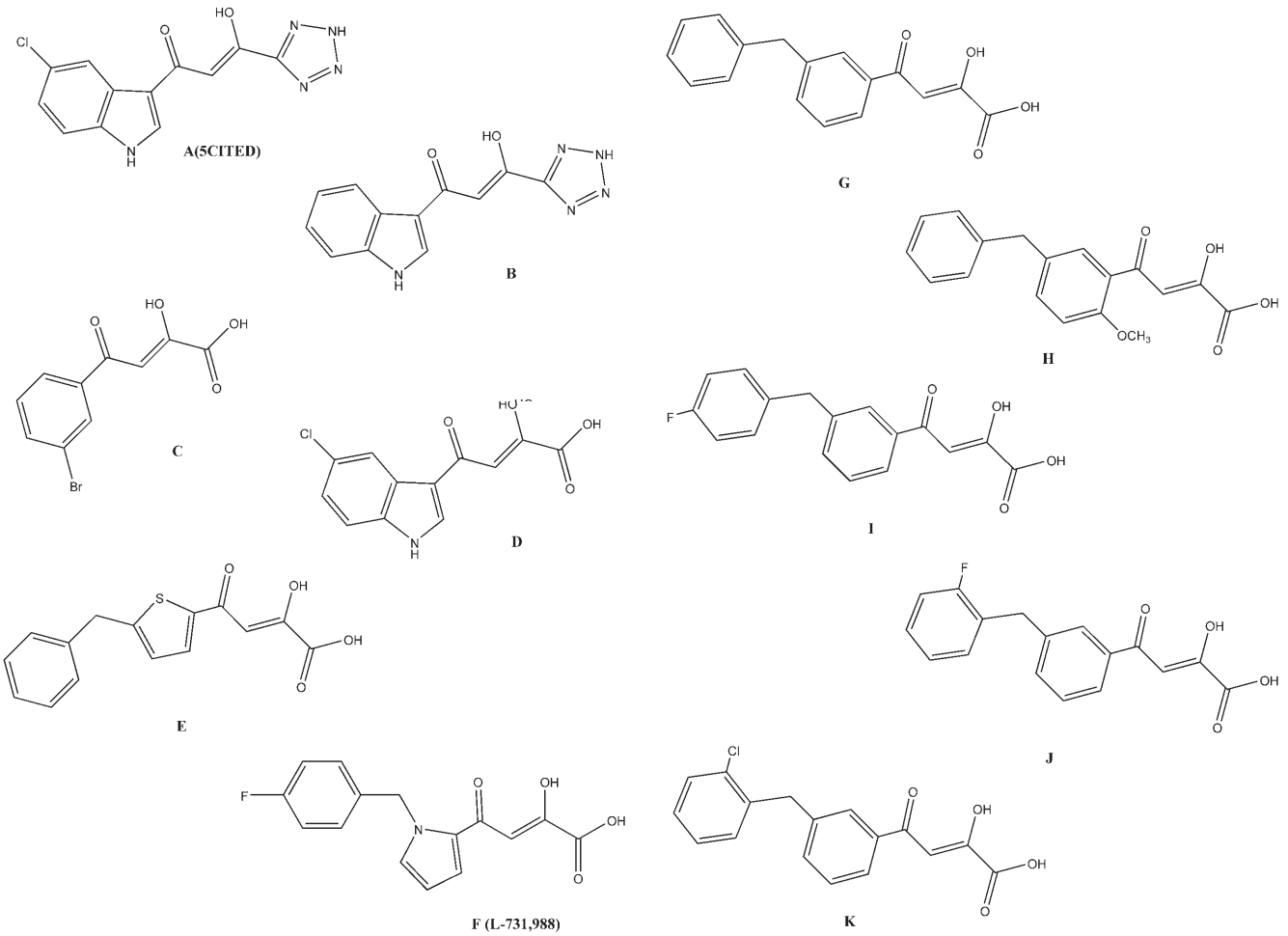
A consideration of protein flexibility may be indispensable for a critical evaluation of ligand-binding affinity.[26–29] Docking techniques usually ignore the flexibility of the receptor and only the ligand is allowed to suffer conformational changes during the docking procedure. This shortcoming can be partially overcome by considering several receptor structures conveniently identified by simulations. In any case, the possibility of cooperative structural changes between a ligand and the receptor are not incorporated into this strategy. On the other hand, MD methods can be used to calculate the free energy of protein–ligand binding, including the flexibility of the full system.[30] The main problem associated with this methodology is the computational effort required in sampling adequately the full configurational space and also in parametrising each new ligand assayed. We herein investigate a strategy to overcome these two computational bottlenecks based on the use of combined quantum mechanical/molecular mechanical (QM/MM) methods.[31–33] These methods are now widely used in the analysis of enzymatic reactions.[34,35] In hybrid QM/MM methods, the ligand/substrate species may be described by a QM model, whereas the protein and solvent environment are represented by means of MM force fields. This hybrid methodology avoids most of the work needed to obtain new force-field parameters for each new functional group, like the ones presented in these compounds. Treating the ligand quantum mechanically has the additional advantage of being able to include quantum effects such as ligand polarisation upon binding.[36] This can be very important when highly polarisable groups are embedded in anisotropic environments, as for example in metal-containing active sites. Moreover, as the largest part of the system is described classically, sufficient sampling can be achieved at a reasonable computational cost. In some recent computational studies, the combined QM/MM method was successfully applied to the study of protein–ligand interactions in the HIV-1 protease[37] and trypsin systems.[38] In this work we have employed a combined QM/MM approach to determine the protein–ligand interaction energy for L-731,988 and 10 of its derivatives or analogues acting as IN strand-transfer selective inhibitors[10,18,19,39] (Scheme 1). We chose compounds that selectively inhibit the strand-transfer reaction of IN with comparable $IC_{50}$ values. These values were obtained by using the same experimental protocol and it is expected that all the compounds bind in a similar way in the active site of the protein. A detailed analysis of the interactions of each inhibitor with the key residues inside the binding pocket is given. We further suggest that the results can be used to design more potent synthetic inhibitors.

## Methods

The initial coordinates for the QM/MM calculations were taken from the A-chain of the trimeric crystal structure of HIV-1 IN complexed to the inhibitor 5CITEP (PDB code 1QS4, compound **A** in Scheme 1).[18] This structure was chosen as it is the only one that binds the inhibitor to the IN catalytic core domain. The unresolved amino acids (141–144) were added to the 1QS4 of the initial crystallographic structure using the HYPERCHEM[40] program in order to have a complete protein system.

All the hydrogen atoms of the 1QS4–IN system were positioned in the protein according to electrostatically based p$K_a$ calculations,[41] as implemented by Field et al.;[42] no unusual amino acid ionisation states were predicted. All of the inhibitors were modelled in a deprotonated form owing to the ability of these moieties to stabilise the negative charge by electron delocalisation under physiological conditions (the N–H bond of tetrazoles or the O–H bond of carboxylic acids).[20] Initially, the hydrogen atoms of the protein were added and relaxed by means of successive minimisation steps and different search algorithms: steepest descent, conjugated gradient and LBFGSB.[43] During this process all the heavy atoms in the protein and the inhibitor were restrained using a harmonic tether with an umbrella constant of 1000 kJ mol$^{-1}$ Å$^{-2}$. In the next step, the protein was relaxed using the same restraining scheme applied to just the peptidic backbone but with a smaller constant (100 kJ mol$^{-1}$ Å$^{-2}$). Then




Scheme 1. Structures of the selected diketo acid containing HIV-1 IN inhibitors.

the optimised protein was placed in a cubic box of pre-equilibrated waters (80 Å a side) using the principal axis of the protein–inhibitor complex as the geometrical centre. Any water molecule with an oxygen atom lying within a radius of 2.8 Å of a heavy atom of the protein was deleted. The remaining water molecules in the final structure were then relaxed using several steps of the conjugated gradient algorithm followed by a LBFGSB optimisation. The final system was equilibrated using hybrid QM/MM Langevin–Verlet molecular dynamics (NVT) over 50 ps. Only the atoms of the inhibitor were treated quantum mechanically, as long as there are no covalent bonds between subsystems, using the AM1 semiempirical method.[44] Thus, both protein and water molecules have been described using the classical OPLS-AA[45] and TIP3P[46] potentials, as implemented in the DYNAMO[47] library. The resulting structure was used as a starting point for the rest of the inhibitors, just changing the substrate molecule by the different inhibitors depicted in Scheme 1. Also, any residue further than 25 Å from any of the atoms of the initial inhibitor was kept frozen in the remaining calculations (10235 mobile atoms). Afterwards, all the systems were equilibrated for 900 ps of QM/MM MD at a temperature of 300 K. The computed RMSD for the protein during the last 200 ps renders a value that is always below 0.9 Å. Furthermore, the RMS of the temperature during the different equilibration steps was always less than 2.5 K and the variation coefficient of the potential energy during the dynamics simulations was never higher than 0.3%.

To determine the interaction energy between the inhibitor and the solvated enzymatic system, we must keep in mind the hybrid QM/MM potential energy given by Equation (1), where $V_{MM}(R_{MM})$ represents the energy of the molecular mechanics force field, $\hat{H}_0(r,R_{QM})$ is the in vacuo Hamiltonian for the selected QM method, $\Psi(r,R_{QM},R_{MM})$ is the polarised wavefunction due to the presence of an external and flexible electric field generated by the protein and the solvent molecules, and finally, $\hat{V}_{QM/MM}(r,R_{MM})$ and $V^{VW}_{QM/MM}(R_{QM},R_{MM})$ are the electrostatic and van der Waals coupling operators between the QM and MM parts (the van der Waals interaction term does not involve electronic coordinates and thus does not need to be evaluated in the SCF procedure). The interaction energy can be then calculated from an analysis of the MD trajectories as the energy difference between the full QM/MM system and the separated QM and MM subsystems. Taking into account the fact that in non-polarisable force fields the MM energy exactly cancels out, we can write Equation (2) and Equation (3), which means that in vacuo single-point energy calculations must be carried out for given nuclear configurations of the quantum atoms along the trajectory, and then this energy must be subtracted from the QM/MM energy. Note that according to our definition of the polarisation energy [third term in Eq. (3)] as the gas-phase energy difference between the polarised and unpolarised wavefunctions, this is a positive contribution.

$$E = V_{MM}(R_{MM}) + \langle\Psi(r,R_{QM},R_{MM})|\hat{H}_0(r,R_{QM})|\Psi(r,R_{QM},R_{MM})\rangle + \langle\Psi(r,R_{QM},R_{MM})|\hat{V}_{QM/MM}(r,R_{MM})|\Psi(r,R_{QM},R_{MM})\rangle + V^{VW}_{QM/MM}(R_{QM},R_{MM}) \quad (1)$$

$$E_{INT} = E - E^{\circ}(R_{QM}) - V_{MM}(R_{MM}) = V^{VW}_{QM/MM}(R_{QM},R_{MM}) + \langle\Psi(r,R_{QM},R_{MM})|\hat{V}_{QM/MM}(r,R_{MM})|\Psi(r,R_{QM},R_{MM})\rangle + \langle\Psi(r,R_{QM},R_{MM})|\hat{H}_0(r,R_{QM})|\Psi(r,R_{QM},R_{MM})\rangle - \langle\Psi(r,R_{QM})|\hat{H}_0(r,R_{QM})|(r,R_{QM})\rangle \quad (2)$$

$$E_{INT} = E_{VW}(R_{QM},R_{MM}) + E^{QM}_{ELEC}(r,R_{QM},R_{MM}) + E^{QM}_{POL}(r,R_{QM},R_{MM}) \quad (3)$$

This strategy for obtaining the interaction energy was applied in a subsequent 100 ps QM/MM MD run at 300 K, using a time step of 1 fs to solve the equation of motion. In these calculations, a switched cut-off region of 11.0–13.5 Å was used to calculate the non-bonding interaction energies. The polarised wavefunctions are conveniently stored during the MD simulations, whereas the unpolarised wavefunctions are obtained a posteriori using the saved coordinates of the QM subsystem. Next, contributions of the different residues to the total enzyme–inhibitor interaction energy were determined by using the polarised wavefunction. For this purpose we evaluated the interaction of the polarised system with the MM centres belonging to a particular residue. This procedure provides only the electrostatic contribution to the interaction energy of a given residue. Thus, the global polarisation effect must be derived from the difference between the total interaction energy and the sum of the local contribution of each residue. As stated before, this term is always positive. The described computational procedure was applied to all 11 investigated inhibitors.

Finally, to measure the goodness of the selected QM method, two protein–inhibitor complexes (**A** and **F**) were optimised by hybrid density functional theory (DFT) within the BLYP functional[47] and the 6-31G* basis set, BLYP(6-31G*)/MM methods. The main reason for choosing the AM1 Hamiltonian instead of a more accurate one lies in time limitations: performing molecular dynamics calculations that are long enough using ab initio or DFT methods combined with MM is still unaffordable. All in all, a comparison of potential energy surface explorations carried out at both levels of theory (AM1/MM and DFT/MM) can provide some useful results and validate, to some extent, the use of semi-empirical methods in this theoretical study. Analysis of the interaction energies for each residue obtained by using the AM1/MM method on different optimised systems are shown to be in a good agreement with those obtained at the BLYP(6-31G*)/MM level of theory within the framework of a QM density-fitted electrostatic interaction.[48]

There are two possible orientations for the aromatic part of the compounds **E**–**H**. Therefore, in this work we carried out simulations on compound **F** (L-731,988), which has two different orientations within the active site, and after 600 ps the orientation obtained was always the same. Thus, this orientation has been used for the remaining compounds.

Four of the eleven diketo acids studied herein do not yield a precise experimental value for the anti-HIV-1 IN activity (values reported as "less than 0.1 μM"). As a consequence, these four inhibitors (compound **H**, **I**, **J** and **K**) were not included in the regression model and, therefore, the linear regression (LR) presented in this study was constructed with only seven molecules (**A**–**G**). The resulting LR will be used on the four remaining molecules (**H**–**K**) to predict their activity and to check if the theoretical values are in fact less than 0.1 μM.

## Results and Discussion

As explained in the previous section, we have performed hybrid QM/MM molecular dynamics (QM/MM MD) simulations of 1.0 ns for the inhibitor L-731,988 and 10 of its derivatives and analogues (Scheme 1). Table 1 shows the ligand–protein interaction energies for the 11 inhibitors studied herein, as well as their decomposition into electrostatic and polarisation terms and their experimentally determined anti-HIV IN activities. The reported values are averaged over 100000 configurations from the last 100 ps of the QM/MM MD simulations. Both the geometrical data and energy terms were averaged during the simulations and the standard deviations are given in parentheses alongside the averaged values. The final structures of compounds **A** and **F** obtained from the 1.0 ns MD simulations are shown in Figure 1 (see the Supporting Information for the final structures of the other compounds) and the contributions of individual residues to the total interaction energy are displayed in Figure 2. In Figure 2, negative values correspond to stabilising effects.

Table 1. Averaged QM/MM interaction energies and experimental anti-HIV IN activities.[a] The corresponding electrostatic (Elec int) and polarisation (Pol int) interaction terms are also reported.

| Compound | Interaction energy [kJ mol$^{-1}$] | Elec int [kJ mol$^{-1}$] | Pol int [kJ mol$^{-1}$] | Activity [μM] |
|---|---|---|---|---|
| **A** | −617.3(±35.8) | −644.9 | 27.6 | 2.30(±0.10)[18] |
| **B** | −643.4(±33.8) | −683.4 | 40.0 | 1.95[10] |
| **C** | −648.6(±36.8) | −681.7 | 33.1 | 2.50[10] |
| **D** | −700.7(±41.7) | −738.6 | 37.9 | 0.52(±0.10)[10] |
| **E** | −681.3(±33.7) | −715.8 | 34.5 | 0.50(±0.22)[39b] |
| **F** | −708.1(±40.6) | −742.5 | 34.4 | 0.05[19] |
| **G** | −734.8(±40.6) | −769.4 | 34.6 | 0.01[39b] |
| **H** | −777.6(±44.0) | −818.4 | 40.8 | < 0.1[39a] |
| **I** | −707.9(±33.3) | −738.5 | 30.6 | < 0.1[39a] |
| **J** | −753.1(±36.8) | −785.4 | 32.2 | < 0.1[39a] |
| **K** | −701.0(±34.2) | −731.9 | 30.9 | < 0.1[39a] |

[a] Values have been obtained from refs. [10,18,19,39]. The values in parentheses correspond to the standard deviations.

In Figure 1 we can observe how the tetrazole ring of compound **A** (5CITEP) interacts, through hydrogen bonds, with the pocket created by the Asn155, Lys156 and Lys159 residues. These interactions favour the stabilisation of the protein–inhibitor complex, as can be observed in Figure 2a. Although not interacting directly with the ligand, Thr66 seems to stabilise the cavity created by the three aforementioned residues, in particular, interacting with Lys159. This result is in agreement with experimental data for mutations of HIV-1 IN at Thr66.[19] Another residue that seems to be essential for catalysis,[49] but that does not participate in substrate binding (as shown in Figure 2a), is the negatively charged Glu152 residue. Its role seems to be similar to that of Thr66: it interacts with Lys156 and Gln148 residues, thus stabilising the cavity structure around the tetrazole ring. Regarding the indole ring of the inhibitor, and as previously suggested by Dougherty,[15] it appears possible that an electrostatic attrac-

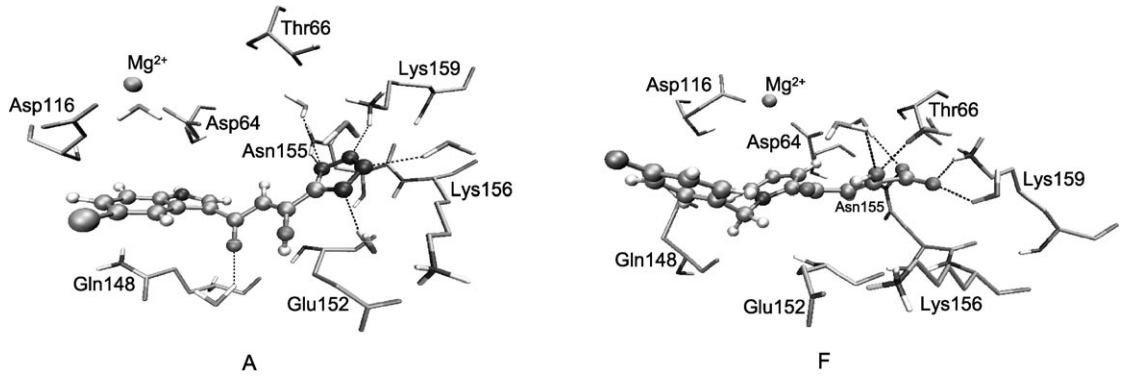



Figure 1. Representation of the most important interactions between HIV-1 IN and compounds **A** and **F**.

tion between the quadrupole moment created by the π electrons of the aromatic moiety and the $Mg^{2+}$ cation was established (population analysis of all the compounds is reported in the Supporting Information). This favourable interaction is reflected in Figure 2a. Correspondingly, the same argument can be used to explain the electrostatic repulsion observed between negative Asp64 and the indole ring. The presence of diketo or β-hydroxy keto groups and metal-dependent inhibition by DKAs and DKA-like compounds have been interpreted as being an indicator of a direct interaction of these drugs with the divalent metal in the catalytic enzyme.[11,13,14,16] This has been suggested by many studies, with a binding mode in which inhibitors bind through bidentate chelation of this ion. In our calculations the inhibitor does not chelate the metal, but a favourable interaction with the metal still exists, as demonstrated in Figure 2a, thus showing that it is not necessary for the $Mg^{2+}$ ion to be chelated by the inhibitor. In fact, Briggs and co-workers[23] have recently shown that chelation of inhibitor L-731,988 with the $Mg^{2+}$ ion is unfavourable.

These results are compatible with previous experimental and theoretical data. Thus, docking studies[50] of HIV-1 IN–viral DNA interactions show that the 3′-hydroxy moiety of the conserved deoxyadenosine is within hydrogen-bonding distance of Asp64. Mutagenesis studies show that Asn155 and Lys159 are critical for IN–DNA binding[49a,b,c] and photo-crosslinking studies[49d] have suggested that the conserved adenosine binding in the vicinity of Lys159 and Glu152, as well as the 5′-adenosine overhang, should be in contact with Gln148.

Finally, as observed in Figure 1A, some water molecules interact with the inhibitor stabilising the complex. In particular, one of them, from the coordination sphere of the $Mg^{2+}$ ion, forms a hydrogen bond with the nitrogen atom of the indole ring. Another water molecule is involved in hydrogen bonds with the keto enolic motif. Moreover, three water molecules interact with the tetrazole ring and one of them forms a hydrogen bond with Lys156. McCammon[21d] and Briggs[21c,23] and their co-workers found that water molecules are important in the stabilisation of the interaction between 5CITEP and L-731,988 and IN, respectively, in accordance with our theoretical simulations.

Compound **B** is equivalent to compound **A**, except the chlorine atom of the indole ring has been substituted by a hydrogen atom. With the exception of Asp64 and Thr66, the pattern of interactions between this inhibitor and the IN protein is quite similar that of compound **A**, as observed in Figure 2a. The new hydrogen bond established between the hydroxy group of Thr66 and a nitrogen atom of the tetrazole ring is clearly reflected in Figure 2a. Nevertheless, it is important to point out that the absence of an electronegative atom in the indole ring provokes a weaker interaction with water, however, the interactions with Lys159, Thr66 and Asp64 are of higher energy than those in compound **A** (see Figure 2a). These features are reflected in the total averaged QM/MM interaction energies reported in Table 1, for which an increase is observed for compound **B** compared with compound **A**.

In compound **C**, the tetrazole ring has been substituted by a carboxylic acid group and the indole ring by bromobenzene (see Scheme 1). The carboxylic group adopts a similar orientation to the tetrazole ring in compounds **A** and **B**, and that observed in the X-ray crystal structure of 5CITEP.[18] Briggs and co-workers[23] also found the same orientation in the structure of inhibitor L-731,988 (compound **F**, discussed below). Therefore, although they are quite different, compounds **A**, **B** and **C** have similar interaction energies with the individual residues, as observed in Figure 2a, suggesting, from the point of view of ligand–protein interactions, that it is important for there to be a functional group with a high electron density capable of interacting with the residues of the pocket: Asn155, Lys156 and, especially, Lys159. Probably the only dramatic change is observed for Lys156, which suffers a loss of interaction with the inhibitor owing to its different size and electron-density distribution. It can also be shown that the interactions with the $Mg^{2+}$ ion and the water molecules are improved. These local effects are reflected in the total ligand–protein interaction energy, which is almost equal to that calculated for compound **B** (see Table 1).

The relevance of the interactions with the residues Asn155, Lys156 and Lys159 can be checked by substituting the tetrazole ring of compound **B** by a carboxylic acid group whilst retaining the indole ring: compound **D**. Figure 2a

shows that the inhibitor–protein pattern of interactions for this inhibitor is similar to those observed for compounds **A**, **B** and **C**. However, although the strength of the interactions with the $Mg^{2+}$ cation is reduced, the interaction with Lys159 is greater than in compounds **A**, **B** and **C**, suggesting that the carboxylate group is more efficient at binding than the tetrazole ring. This, together with the favourable interactions with water molecules, is probably the key factor for the increase in the overall ligand–protein interaction (see Table 1). At this point a relationship between ligand–protein interaction and the anti-HIV IN activity of the inhibitor can be deduced, as demonstrated in Table 1, in which the experimentally measured activities of the different putative inhibitors (in μM) are reported.

Figure 2a shows the interaction of inhibitor **D** with Lys156 is markedly higher than with compound **C**. This is reflected in the total ligand–protein interaction (Table 1). From a comparison of their structures, it seems that the $Mg^{2+}$ ion and its coordination sphere, together with the Gln148 and Gln152 residues, are displaced away from the pocket. The average distance between the $Mg^{2+}$ ion and the centre of mass of the ring increases from 6.27 Å in compound **A** to 8.86 Å in compound **D** (other key averaged distances for all the compounds are reported in the Supporting Information). This cooperative movement creates some extra room in the cavity allowing the Lys156 to bind the ligand, thus increasing the intensity of its interaction and stabilising the protein–inhibitor complex.

Compound **E** presents a quite different moiety to interact with the $Mg^{2+}$ ion and its surroundings; the indole ring is substituted by 2-benzylthiophene (see Scheme 1). The absence of the bicyclic ring confers a certain flexibility on this side of the system. The consequence of this change is that, although the strong interaction with Lys159 remains almost invariant, the effect on the interactions established on the ring side of the inhibitor suffers slight changes. Thus, the unfavourable interaction with Asp64 and the stabilisation offered by the $Mg^{2+}$ ion appear again (Figure 2a). Rotation of the inhibitor around the methylene group forced by steric impediments with the Gln148 residue is observed in this complex. These changes, although provoking a reduction in the ligand–protein interaction, do not imply a change in the activity of the inhibitor (Table 1).

A similar analysis was performed on the inhibitor L-731,988: compound **F**. Figure 1 and Figure 2a show very strong interactions between the inhibitor and the Lys159, Asn155 and Lys156 residues. Furthermore, the interaction with the $Mg^{2+}$ ion is much stronger than in the previous compounds. The averaged distance between the centre of mass of the benzene ring and the cation is negligible, chang-

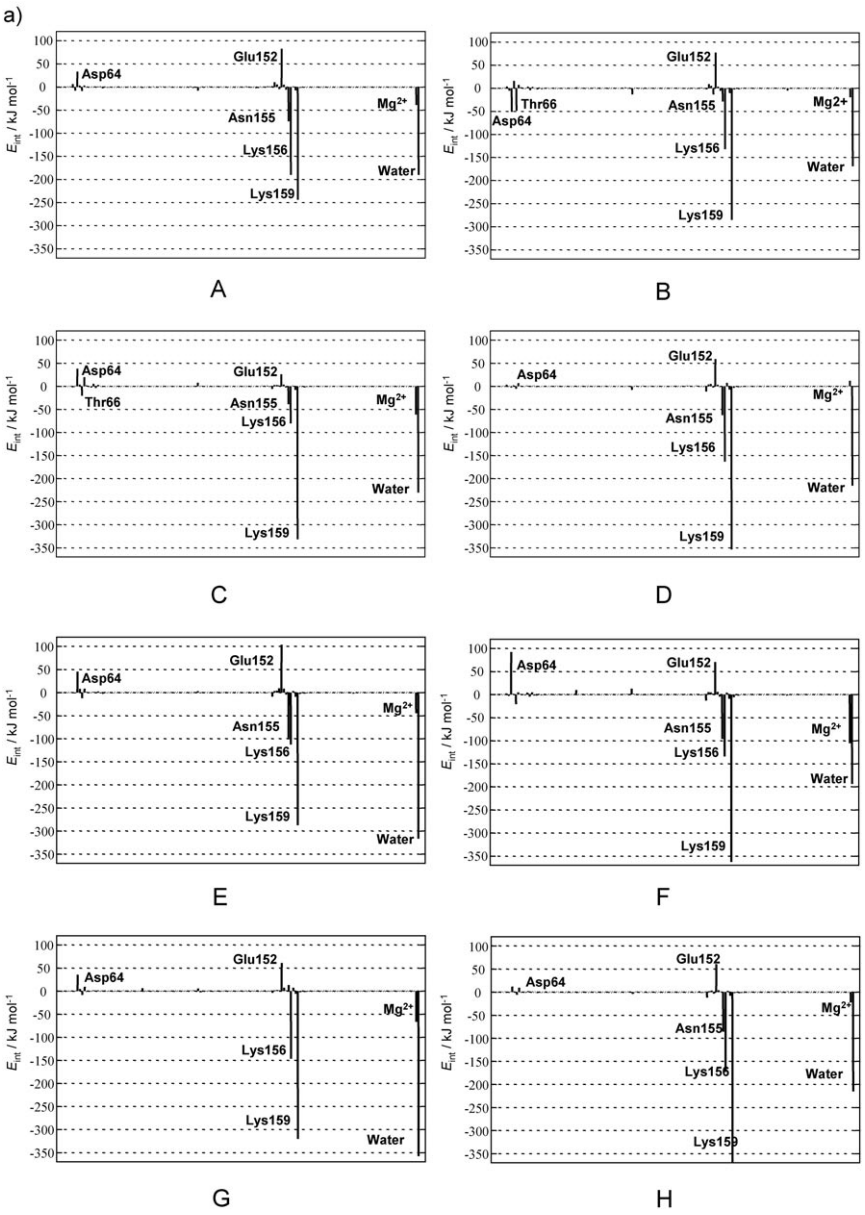



Figure 2. Contributions of individual amino acid residues to inhibitor binding of a) compounds **A**–**H** computed at the AM1/MM level, b) compounds **I**–**K** computed at the AM1/MM level and c) **A** and **F** computed at the DFT/MM level. Water refers to the effect of all the water molecules in the active site.

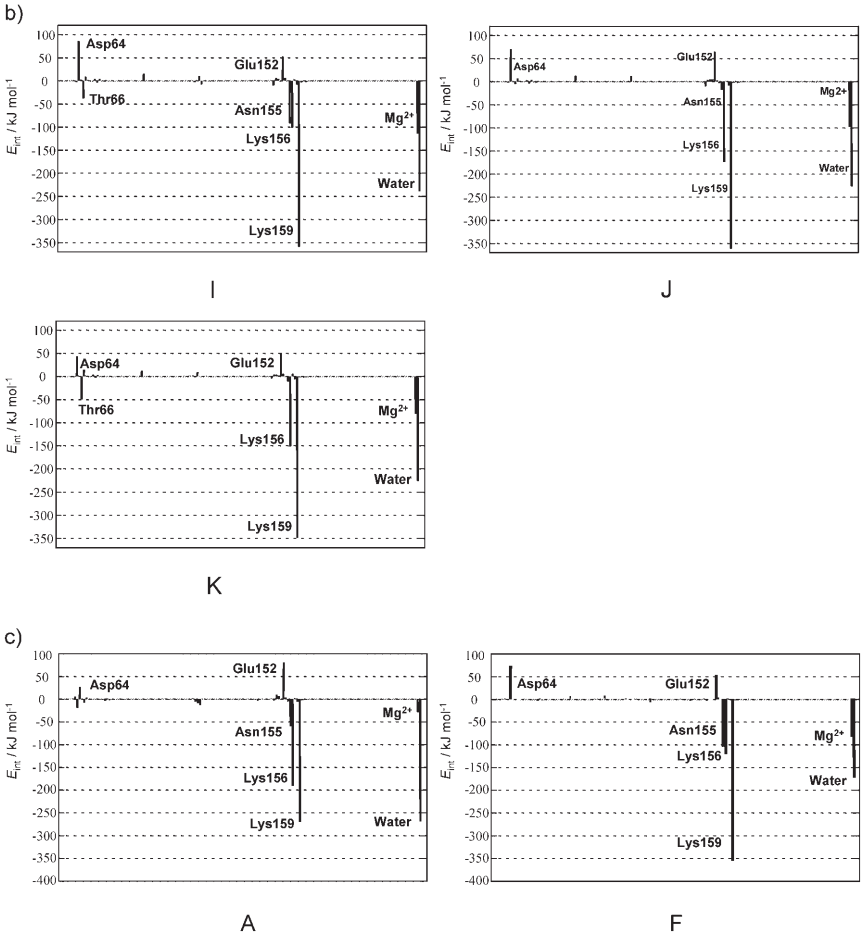
ing from 6.27 Å in compound **A** to 5.61 Å in compound **F**. Nevertheless the different relative orientations of the $Mg^{2+}$ ion and the aromatic pyrrole ring in compound **A** and **F** may be responsible for the different behaviour: the favourable cation–π interaction described above is optimised in the latter case (a representation of the different relative orientations can be found in the Supporting Information). This result is also evident in the total ligand–inhibitor interactions listed in Table 1; compound **F** exhibits a higher total energy than the previous compounds and, as a consequence, it also exhibits higher anti-HIV activity. Analysis of the complete protein–**F** structure reveals that this inhibitor adopts an orientation in which it may physically block the possible interface of the integrase–DNA complexes, in agreement with the model of Pommier et al.[17]

In compound **G**, the pyrrole ring in **F** has been substituted by a benzene aromatic system (see Scheme 1). This new inhibitor, by comparison with **F**, has lost a cation–π interaction, whereas the interaction with water is much stronger than in the previous compound (Figure 2a). This result is also evident from the total ligand–inhibitor interaction energies listed in Table 1, for which an increase is observed when comparing **G** with previous compounds and, following the observed relationship between interaction energy and activity, it exhibits a higher anti-HIV activity.

The averaged distance between the centre of mass of the aromatic ring and the cation is negligible, although following a reasonable trend considering the changes in the interaction energies observed for compounds **A**–**G**: from 6.27 Å in **A** to 5.61 Å in **F** and 5.41 Å in compound **G**.

The analogues **H**, **I, J** and **K** are formed by the introduction of 2-methoxy, 4′-fluoro, 2′-fluoro and 2′-chloro into **G**, respectively (see Scheme 1). The 2-methoxy group in compound **H** causes steric impediments with the $Mg^{2+}$ cation and, consequently, a lost cation–π interaction. However, the interaction with Lys159 is markedly higher than with previous inhibitors and the unfavourable interaction with Asp64 is negligible (see Figure 2a). As a consequence, this compound exhibits the highest total ligand–inhibitor interaction energy, which is in accordance with a very high anti-HIV activity[39b] (see Table 1). Compounds **I**, **J** and **K** also show very strong interactions between the inhibitor and the Lys159 and Lys156 residues (not so dramatic in **I**), whereas the interaction with Asn155 is variable. A common feature is the strong interaction with the $Mg^{2+}$ ion, much stronger in compound **I** than in **J** and **K** (see Figure 2b). This is because a water molecule in the coordination sphere of the $Mg^{2+}$ ion forms a hydrogen bond with the fluorine atom of the inhibitor.

Based on the criterion of a direct relationship between protein–inhibitor interaction and anti-HIV IN activity, the 11 compounds can be divided into two groups: the first group of inhibitors exhibit very high interaction energies (compounds **D**–**K**) and the second group have interaction energies below 680 kJ mol$^{-1}$, that is, they have moderate activity (**A**, **B** and **C**). As observed in the analysis of individual interactions (Figure 2a,b), the influence of Asn155, Lys156 and Lys159 seems to be crucial, the interaction established between the inhibitor and the latter being especially important. These interactions, together with those established with water molecules in the cavity created by the protein and the $Mg^{2+}$ ion, might be responsible for the anti-HIV activity. In order to check the reliability of the analysis of these interactions, the time evolution of the protein–inhibitor interaction energy during the course of the QM/MM molecular dynamics simulations was studied (see the Supporting Information). The resulting plots show no im-



portant fluctuations in any of the compounds, the first three compounds present higher interaction energies than the second group at any time of the simulation and thus supporting the conclusions obtained from the analysis of the static interactions. Furthermore, much more expensive calculations at the BLYP/6-31G*/MM level of theory were carried out for compounds **A** and **F**. In particular, full structural optimisation (reported in the Supporting Information) followed by an analysis of protein–inhibitor interactions (see Figure 2c) show that both AM1/MM and DFT/MM calculations render similar results, thus confirming the reliability of the former.

Other residues, such as Gln148, Thr66, Asp64 and Glu152, also seem to play an important role in the anti-HIV-1 IN activity. This result is in agreement with experimental studies that have shown the Lys159, Lys156, Asp64, Asp116, Gln148 and Glu152 residues to be in close contact with DNA and to be critical for IN function,[44] whereas Thr66 and Asn155 residues are associated with drug resistance.[17]

Finally, it is evident that the biological activity of each ligand is clearly related to its binding energy. Within a family of compounds, the solvation/desolvation energies, the enzyme deformation energies and the entropic changes are proportional to the magnitude of the interactions established between the inhibitor and the protein and one may then express the activity as a function of the interaction energy.

Analysis of the averaged QM/MM total ligand–protein interaction energies of compounds **A**–**G** reveals a good linear correlation between this magnitude and the biological anti-HIV-1 activity, provided as the $\log(1/IC_{50})$ index (correlation coefficient of 0.90). This correlation is shown in Figure 3, in

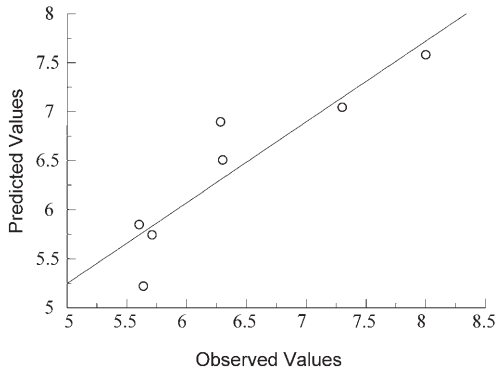



Figure 3. Experimental versus predicted anti-HIV-1 IN activities measured in terms of $\log(1/IC_{50})$.

which the experimental and predicted activities of the different compounds are compared. The model expressed as Equation (4) was obtained. The squared correlation coefficient, $R^2$, is a measure of the fit of the regression model, the Fisher test, $F$, reflects the ratio of the variance explained by the model and the variance due to the error in the model (high values for the $F$ test indicate the significance of the equation), SEP is the standard error of prediction and $N$ is the number of tested inhibitors. The fit of the model [Eq. (4)] is shown in Table 2 and Figure 3. Analysis of the results shows a satisfactory agreement between the observed and calculated values.

Table 2. Theoretical anti-HIV-1 activity, predicted by Equation (4), compared with experimental data, expressed in terms of $\log(1/IC_{50})$.

| Compound | Activity | |
|---|---|---|
| | Experimental | Predicted |
| **A** | 5.64 | 5.16 |
| **B** | 5.70 | 5.69 |
| **C** | 5.60 | 5.79 |
| **D** | 6.28 | 6.83 |
| **E** | 6.30 | 6.44 |
| **F** | 7.30 | 6.98 |
| **G** | 8.00 | 7.52 |
| **H** | >7.00 | 8.37 |
| **I** | >7.00 | 6.97 |
| **J** | >7.00 | 7.88 |
| **K** | >7.00 | 6.84 |

$$\log(1/IC_{50}) = -7.18 - 0.02E_{INT} \quad (4)$$

$R^2 = 0.82$, $F = 23.47$, SEP $= 0.42$, $N = 7$

## Conclusion

This work provides an insight into the activity of potential inhibitors against human immunodeficiency virus type-1 integrase (HIV-1 IN). Hybrid QM/MM MD calculations have been carried out on L-731,988 and 10 of its derivatives to determine the protein–ligand interaction energies, and the correlation between this magnitude and the activity of the putative pharmacophoric compounds has been established. The results suggest that the stronger the interaction energy is, the higher the anti-HIV-1 activity of the inhibitor. The methodology has also allowed contributions of individual residues to the total enzyme–inhibitor interaction energy to be determined. It was found that those residues in the active site of the enzyme, such as Asp64, Thr66, Gln148, Glu152, Asn155, Lys156 and Lys159, as well as the $Mg^{2+}$ ion and water molecules, determine the anti-HIV-1 IN activity. These results are in accordance with the available experimental data published in the literature. Nevertheless, we have observed that a very high interaction energy between Lys-159 and the inhibitor, together with the electrostatic attraction established between the positive charge of the $Mg^{2+}$ ion and the quadrupole moment of the aromatic ring, is an important feature in compounds exhibiting the highest anti-HIV activity. In fact, this is the first time that the importance of the interaction between the magnesium cation and the aromatic ring of the inhibitor has been pointed out. In this regard, the presence of an electron-donating substituent in the ring (for example, a methoxy group), maintaining the flexibility that allows the inhibitor to approach the cation (changing the position of the pyrrole ring substituent or

even removing it) should improve the properties of the inhibitor. Nevertheless, an electron donor can increase the $pK_a$ of the carboxylate group, which would reduce the ligand–protein interaction.

In conclusion, we have succeeded in quantifying the different interactions between the HIV-1 IN protein and several inhibitors. This kind of theoretical study will find future applications in the enzyme-targeted drugs field. In particular, in any attempt to modify a putative inhibitor in order to improve its activity, the contribution of all individual residues has to be taken into account.

## Acknowledgements

We thank the DGI for project BQU2003-04168-C03, Universitat Jaume I - BANCAIXA Foundation for projects P1.1B2005-13, P1.1B2005-15 and P1.1B2005-27 and the "Programa hispano-brasileño de cooperación interuniversitaria" of the Spanish Ministry of Culture and Education for project CAPES no. 0014-13/05. C.N.A. would like to thank the CNPq for financial support and for their warm hospitality during his research stay at the Departament de Ciències Experimentals, Universitat Jaume I. We are also grateful to Dr. M. Field and Dr. P. Amara for supporting us in the $pK_a$ calculations and discussions. The authors also acknowledge the Servei d'Informatica, Universitat Jaume I for generous allocation of computer time.

[1] UNAIDS, Report AIDS Epidemic Update, **2005,** http://www.unaids.org.
[2] H. Jaffe, *Science* **2004**, *305*, 1243–1244.
[3] E. Asante-Appiah, A. M. Skalka, *Adv. Virus Res.* **1999**, *52*, 351–369.
[4] P. O. Brown in *Retroviruses* (Eds.: J. C. Cooffin, S. H. Hughes, H. E. Varmus, Cold Spring Harbor Press, Plainview, NY, **1997**.
[5] S. M. Hammer, *N. Engl. J. Med.* **2005**, *353*, 1702–1710.
[6] D. Esposito, R. Craigie, *Adv. Virus Res.* **1999**, *52*, 319–333.
[7] T. K. Chiu, D. R. Davies, *Curr. Top. Med. Chem.* **2004**, *4*, 965–977.
[8] N. J. Anthony, *Curr. Top. Med. Chem.* **2004**, *4*, 979–990.
[9] M. C. Nicklaus, N. Neamati, H. Hong, A. Mazumder, S. Sunder, J. Chen, G. W. Milne, Y. Pommier, *J. Med. Chem.* **1997**, *40*, 920–929.
[10] G. C. G. Pais, X. Zhang, C. Marchand, N. Neamati, K. Cowasange, E. S. Svaraçovskaia, V. K. Pathak, Y. Tang, M. Nicklaus, Y. Pommier, T. R. Burke, Jr., *J. Med. Chem.* **2002**, *45*, 3184–3194.
[11] C. Marchand, X. Zhang, G. C. G. Pais, K. Cowasange, N. Neamati, T. R. Burke, Jr., Y. Pommier, *J. Biol. Chem.* **2002**, *277*, 12596–12603.
[12] H. Yuan, A. L. Parrill, *Bioorg. Med. Chem.* **2002**, *10*, 4169–4183.
[13] C. Marchand, A. A. Johnson, R. G. Karki, G. C. G. Pais, X. C. Zhang, K. Cowansage, T. A. Patel, M. C. Nicklaus, T. R. Burke, Y. Pommier, *Mol. Pharmacol.* **2003**, *64*, 600–609.
[14] N. Neamati, H. Hong, J. M. Owen, H. E. Winslow, J. L. Christensen, H. Zhao, T. R. Burke, Jr., G. W. Milne, Y. Pommier, *J. Med. Chem.* **1998**, *41*, 3202–3209.
[15] D. A. Dougherty, *Science* **1996**, *271*, 163–168.
[16] a) J. A. Grobler, K. Stillmock, B. H. Hu, M. Witmer, P. Felock, A. S. Espeseth, A. Wolfe, M. Egbertson, M. Bourgeois, J. Melamed, J. S. Wai, S. Young, J. Vacca, D. J. Hazuda, *Proc. Natl. Acad. Sci.* U.S.A. **2002**, *99*, 6661–6666; b) N. Neamati, Z. Lin, R. G. Karki, A. Orr, K. Cowansage, D. Strumberg, G. C. G. Pais, J. H. Voigt, M. C. Nicklaus, H. E. Winslow, H. Zhao, J. A. Turpin, J. Yi, A. M. Skalka, T. R. Burke, Jr., *J. Med. Chem.* **2002**, *45*, 5661–5670; c) W. G. Verschueren, I. Dierynck, K. I. E. Amssoms, L. L. Hu, P. M. J. G. Boonants, G. M. E. Pille, F. F. D. Daeyaert, K. Hertogs, D. L. N. G. Surleraux, P. B. T. P. Wigerinck, *J. Med. Chem.* **2005**, *48*, 1930–1940.
[17] Y. Pommier, A. A. Johnson, C. Marchand, *Nat. Rev. Drug Discov.* **2005**, *4*, 236–248.
[18] Y. Goldgur, R. Craigie, G. H. Cohen, T. Fujiwara, T. Yoshinaga, T. Fujishita, H. Sugimoto, T. Endo, H. Murai, D. R. Davies, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 13040–13043.
[19] D. J. Hazuda, P. Felock, M. Witmer, A. Wolfe, K. Stillmock, J. A. Grobler, A. Espeseth, L. Gabryelski, W. Schleif, C. Blau, M. D. Miller, *Science* **2000**, *287*, 646–650.
[20] R. J. Herr, *Bioorg. Med. Chem.* **2002**, *10*, 3379–3393.
[21] a) A. Brigo, K. W. Lee, G. I. Mustata, J. M. Briggs, *Biophys. J.* **2005**, *88*, 3072–3082; b) M. C. Lee, J. Deng, J. M. Briggs, Y. Duan, *Biophys. J.* **2005**, *88*, 3133–3146; c) M. L. Barreca, K. W. Lee, A. Chimirri, J. M. Briggs, *Biophys. J.* **2003**, *84*, 1450–1463; d) H. Ni, C. A. Sotriffer, J. A. McCammon, *J. Med. Chem.* **2001**, *44*, 3043–3047; e) C. Laboulasis, E. Deprez, H. Leh, J. F. Mouscadet, J. C. Brochon, M. L. Bret, *Biophys. J.* **2001**, *81*, 473–489.
[22] a) C. A. Sotriffer, H. Ni, J. A. McCammon, *J. Am. Chem. Soc.* **2000**, *122*, 6136–6137; b) F. Beierlein, H. Lanig, G. Schurer, A. H. C. Horn, T. Clark, *Mol. Phys.* 2003, 101, 2469–2480; c) J. R. Schames, R. H. Henchman, J. S. Siegel, C. A. Sotriffer, H. Ni, J. A. McCammon, *J. Med. Chem.* **2004**, *47*, 1879–1881; d) R. Dayam, T. Sanchez, N. Neamati, *J. Med. Chem.* **2005**, *48*, 9008–9015; e) M. L. Barreca, S. Ferro, A. Rao, L. De Luca, M. Zappala, A.-M. Monforte, Z. Debyser, M. Witvrouw, A. Chimirri, *J. Med. Chem.* **2005**, *48*, 7084–7088; f) R. Dayam, T. Sanchez, O. Clement, R. Shoemaker, S. Sei, N. Neamati, *J. Med. Chem.* **2005**, *48*, 111–120; g) J. Deng, K. W. Lee, T. Sanchez, M. Cui, N. Neamati, J. M. Briggs, *J. Med. Chem.* **2005**, *48*, 1496–1505; h) V. Nair, G. Chi, R. Ptak, N. Neamati, *J. Med. Chem.* **2006**, *49*, 445–447; i) J. Deng, T. Sanchez, N. Neamati, J. M. Briggs, *J. Med. Chem.* **2006**, *49*, 1684–1692; j) M. L. Barreca, L. De Luca, N. Iraci, A. Chimirri, *J. Med. Chem.* **2006**, *49*, 3994–3997.
[23] A. Brigo, K. W. Lee, F. Fologolari, G. I. Mustata, J. M. Briggs, *Protein J.* **2005**, *59*, 723–741.
[24] Y. Q. Long, X. H. Jiang, R. Dayam, T. Sanchez, R. Shoemaker, S. Sei, N. Neamati, *J. Med. Chem.* **2004**, *47*, 2561–2573.
[25] R. Di Santo, R. Costi, A. Roux, M. Artico, A. Lavecchia, L. Marinelli, E. Novellino, L. Palmisano, M. Andreotti, R. Amici, C. M. Galluzzo, L. Nencioni, A. T. Palamara, Y. Pommier, M. Marchand, *J. Med. Chem.* **2006**, *49*, 1939–1945.
[26] S. J. Teague, *Nat. Rev. Drug Discovery* **2003**, *2*, 527–541.
[27] J. H. Lin, A. L. Perryman, J. R. Schames, J. A. McCammon, *J. Am. Chem. Soc.* **2002**, *124*, 5632–5633.
[28] H. A. Carlson, *Curr. Opin. Chem. Biol.* **2002**, *6*, 447–452.
[29] H. A. Carlson, J. A. McCammon, *Mol. Pharmacol.* **2000**, *57*, 213–218.
[30] T. Simonson, G. Archontis, M. Karplus, *Acc. Chem. Res.* **2002**, *35*, 430–437.
[31] A. Warshel, M. Levitt, *J. Mol. Biol.* **1976**, *103*, 227–249.
[32] M. J. Field, P. A. Bash, M. Karplus, *J. Comput. Chem.* **1990**, *11*, 700–733.
[33] J. Gao, *Acc. Chem. Res.* **1996**, *29*, 298–305.
[34] M. García-Viloca, J. Gao, M. Karplus, D. G. Truhlar, *Science* **2004**, *303*, 186–195.
[35] S. Martí, M. Roca, J. Andres, V. Moliner, E. Silla, I. Tuñón, J. Bertrán, *Chem. Soc. Rev.* **2004**, *33*, 98–107.
[36] J. Gao, X. Xia, *Science* **1992**, *258*, 631–635.
[37] C. Hensen, J. Hermann, K. Nam, S. Ma, J. Gao, H. D. Höltje, *J. Med. Chem.* **2004**, *47*, 6673–6680.
[38] F. Gräter, S. M. Schwarzl, A. Dejaegere, S. Fischer, J. C. Smith, *J. Phys. Chem. B* **2005**, *109*, 10474–10483.
[39] a) J. S. Wai, M. S. Egbertson, L. S. Payne, T. E. Fisher, M. W. Embrey, L. O. Tran, J. Y. Melamed, H. M. Langford, J. P. Guare, L. G. Zhuang, V. E. Grey, J. P. Vacca, M. K. Holloway, A. M. Naylor-Olsen, D. J. Hazuda, P. J. Felock, A. L. Wolfe, K. A. Stillmock, W. A. Schleif, L. J. Gabryelski, S. D. Young, *J. Med. Chem.* **2000**, *43*, 4923–4926; b) L. C. Zhuang, J. S. Wai, M. W. Embrey, T. E. Fisher, M. S. Egbertson, L. S. Payne, J. P. Guare, J. P. Vacca, D. J. Hazuda, P. J. Felock, A. L. Wolfe, K. A. Stillmock, M. V. Witmer, G. Moyer, W. A. Schleif, L. J. Gabryelski, Y. M. Leonard,

J. J. Lynch, S. R. Michelson, S. D. Young, *J. Med. Chem.* **2003**, *46*, 453–456.
[40] HyperChem Release 7.5 for Windows Molecular Modeling System, Hypercube, Inc, Florida (USA), **2002**.
[41] See the following references in which the method is explained and its accuracy compared with the accuracy of other methods: a) M. K. Gilson, *Proteins Struct. Funct. Genet.* **1993**, *15*, 266–282; b) J. Antosiewicz, J. A. McCammon, M. K. Gilson, *J. Mol. Biol.* **1994**, *238*, 415–436.
[42] M. J. Field, P. Amara, L. David, D. Rinaldo, personal communication.
[43] R. H. Byrd, P. Lu, J. Nocedal, C. Zhu, *J. Scientific Comput.* **1995**, *16*, 1190–1208.
[44] M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, J. J. P Stewart, *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
[45] W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, *J. Am. Chem. Soc.*, **1996**, *118*, 11225–11236.
[46] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*, 926–935.
[47] a) A. Becke, *Phys. Rev. A* **1988**, 38, 3098–3100; b) C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, 37, 785–789.
[48] a) M. J. Field, *A Practical Introduction to the Simulation of Molecular Systems*, Cambridge University Press, Cambridge, **1999**; b) M. J. Field, M. Albe, C. Bret, F. Proust-de Martin, A. Thomas, *J. Comput. Chem.* **2000**, *21*, 1088–1100.
[49] a) T. M. Jenkins, D. Esposito, A. Engelman, R. Cragie, *EMBO J.* **1997**, *16*, 6849–6878; b) D. Eposito, R. Cragie, *EMBO J.* **1998**, *17*, 5832–5843; c) T. S. Heuer, P. O. Brown, *Biochemistry* **1997**, *36*, 10655–10665; d) T. S. Heuer, P. O. Brown, *Biochemistry* **1998**, *37*, 6667–6678.
[50] A. A. Adesokan, V. A. Roberts, K. W. Lee, R. D. Lins, J. M. Briggs, *J. Med. Chem.* **2004**, *47*, 821–828.

Received: January 10, 2007
Published online: June 15, 2007