

A Computational Study of the Protein-Ligand Interactions in CDK2 Inhibitors: Using Quantum Mechanics/Molecular Mechanics Interaction Energy as a Predictor of the Biological Activity

Jans H. Alzate-Morales,* Renato Contreras,* Alejandro Soriano,[†] Iñaki Tuñon,[†] and Estanislao Silla[†]

*Departamento de Química, Facultad de Ciencias, Universidad de Chile, Santiago, Chile; and [†]Departamento de Química Física, Universidad de Valencia, Valencia, Spain

ABSTRACT We report a combined quantum mechanics/molecular mechanics (QM/MM) study to determine the protein-ligand interaction energy between CDK2 (cyclin-dependent kinase 2) and five inhibitors with the N²-substituted 6-cyclohexylmethoxypurine scaffold. The computational results in this work show that the QM/MM interaction energy is strongly correlated to the biological activity and can be used as a predictor, at least within a family of substrates. A detailed analysis of the protein-ligand structures obtained from molecular dynamics simulations shows specific interactions within the active site that, in some cases, have not been reported before to our knowledge. The computed interaction energy gauges the strength of protein-ligand interactions. Finally, energy decomposition and multiple regression analyses were performed to check the contribution of the electrostatic and van der Waals energies to the total interaction energy and to show the capabilities of the computational model to identify new potent inhibitors.

INTRODUCTION

The cyclin-dependent kinases (CDKs) play an essential role in regulating eukaryotic cell-cycle progression (1). These protein kinases are generally categorized into G1, S, and G2 phase regulators because they are present at various checkpoints in the cell cycle (2). As their name suggests, the CDKs are dependent on larger proteins known as cyclins for activation. Only as a complex can these proteins regulate cell growth and DNA synthesis properly. Partial activation occurs upon binding of these positive regulatory subunits; complete activation requires phosphorylation of the CDK subunit by the CDK-activating kinase at the conserved threonine residue. The CDK considered in this study is CDK2, which combines with cyclin E at an S-phase checkpoint known as the restriction point. In the same way, the completion of the S-phase depends on a complex of CDK2 and cyclin A (2).

The activity of the CDK-cyclin complex can be reduced by at least two major mechanisms: the phosphorylation of the CDK subunit at inhibitory sites and the binding of the specialized protein inhibitors known as CKIs or CDK inhibitors. These inhibitors compete with ATP (adenosine 5'-triphosphate) for binding to the CDK active site. However, in some cancer cells it has been shown that the CKIs are underexpressed, and medicinal chemists have made numerous efforts to replace the CKIs with synthetic inhibitors (3). Considerable progress has been made in the identification of pharmacologic agents targeting the CDKs (4). A large number of ATP-competitive inhibitors from a variety of chemical classes have been identified (2,5–7). Among noteworthy attempts to

produce such inhibitors are a series of compounds based on O⁶-cyclohexylmethylpurine or NU2058 (8) (see Fig. 1), which are competitive inhibitors of both CDK1 and CDK2 with respect to ATP. They also display good selectivity over CDK4 (9). Several authors, with the aid of iterative structure-based drug design, have carefully explored this scaffold. In this way, it has been possible to identify three kinds of characteristic interactions for this class of compounds within the active site of some CDKs. The first is the presence of the triplet of hydrogen bonds formed between the different tested compounds and the hinge region in CDK1 and CDK2. This feature induces a different orientation of these compounds inside the active site of the enzyme with respect to other inhibitors such as flavopiridol and olomoucine, and of course it has direct consequences for enzyme and cell growth inhibition (8). A second characteristic for this family of compounds is that optimum binding occurs with a moderately sized aliphatic O⁶ substituent that tightly packs against the hydrophobic patch presented by the glycine loop, centered on Val-18—an interaction promoted by the conformational constraints imposed in a cyclohexylmethyl or cyclohexenylmethyl ring.

Thus, the parent compound O⁶-cyclohexylmethylguanine (NU2058) is the preferred starting point for exploring other areas of the kinase active site (10). The third characteristic kind of interactions are those established with the so-called 'specificity surface'; this is with residues that lie outside the highly conserved ATP binding site cleft (11,12). Sequence differences between the different members of the CDK family exist in this region, and targeting them may afford selectivity. For example, CDK2 residues His-84, Gln-85, and Lys-89 are, respectively, an aspartate, a glutamine, and a threonine in CDK4. Studies on olomoucine (13) and roscovitine (14) have

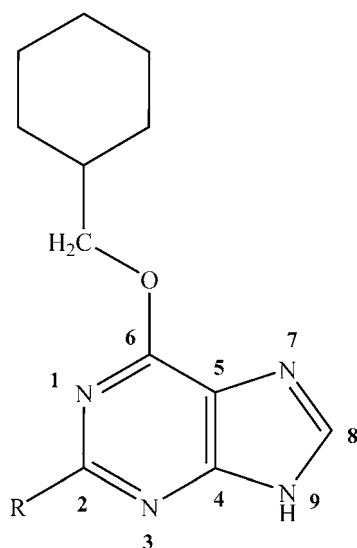
Submitted June 20, 2006, and accepted for publication October 6, 2006.

Address reprint requests to Jans H. Alzate-Morales, Tel.: 56-2-978-7272; Fax: 56-2-271-3888; E-mail: jalzate@ciq.uchile.cl; or Iñaki Tuñon, Fax: 34-96-386-4564; E-mail: ignacio.tunon@uv.es.

© 2007 by the Biophysical Society

0006-3495/07/01/430/10 \$2.00

doi: 10.1529/biophysj.106.091512



COMPOUND 2 R = 2-anilino

COMPOUND 3 R = 2-(4'-sulfamoylanilino)

COMPOUND 7 R = 2-methylamino

COMPOUND 9 R = 2-(1-methylethylamino)

COMPOUND 25 R = 2-(4'-hydroxyanilino)

FIGURE 1 Structures of the N²-substituted O⁶-cyclohexylmethylguanine derivatives of Hardcastle et al. (15). Numbering of atoms in the purine moiety is also displayed.

shown that large gains in potency, in addition to specificity, are possible by targeting this 'specificity surface'. Structural analysis had indicated that an aromatic ring at the N2 position of NU2058 would improve inhibitory activity against CDK1 and CDK2 (9). This was found to be the case with the resulting compound produced by maintaining the specificity for the inhibition of CDK1/2 over CDK4. Additional potency is conferred by the presence of a group capable of donating a hydrogen bond at the 4'-position in the aromatic ring. The resulting compound of all these previous structural-based investigations is NU6102 (compound 3 (Cp3) in Fig. 1). This inhibitor is selective and one of the most active CDK2 inhibitors described so far (9,15). Regardless of this significant progress in the structural-activity relationships for this kind of anticancer compounds, there remains the need to find a broader spectrum inhibitor that can, for example, selectively inhibit CDK1, CDK2, CDK4, and CDK6 at low nanomolar IC₅₀ concentrations. The investigation of CDK2-ligand interactions can provide further insight for developing new compounds that are still required to achieve selective cytotoxicity and valuable anticancer activity.

Supplementary to a tremendous amount of experimental work, a few theoretical studies on CDKs, based on classical and quantum mechanical calculations, have also been reported in the literature. For instance, Cavalli et al. have developed force field parameters based on quantum chemical calculations on a relevant model system for simulating the binding of ATP to CDKs (16). Molecular dynamics (MD) simulation with the newly calculated parameters showed that CDK1 should be dynamically more stable than CDK2, implying that differences in protein structure flexibilities among CDKs should be considered in designing their selective inhibitors. More recent computational studies on CDK2 by Sims et al. have established a binding free energy model for CDK2 inhibitors within the framework of the continuum model for the solvent (17). This approach proved to be successful in reproducing the relative inhibitory activities of flavopiridol analogs. On the basis of the validated utility of the computational approach, they suggested new putative inhibitors that were predicted to be more potent than the lead compound. In a later work, these same authors implemented the previously developed, continuum solvent-based charge optimization model with a simple, quadratic programming algorithm and the UHBD Poisson-Boltzmann solver (18). This method allowed them to compute the best set of point charges for a ligand or ligand region based on the ligand and receptor shape and the receptor partial charges by optimizing the binding free energy obtained from a continuum-solvent model (18). However, there are several limitations to such a model of binding free energy (17,19,20). In their particular case, they applied a fixed conformation model that assumes that the ligand and receptor undergo no conformational change upon binding.

Consideration of protein flexibility is indispensable for a critical evaluation of ligand-binding affinity (21–24), especially here because the conformational plasticity of the catalytic domain is a hallmark of protein kinases (25). As ever, a tradeoff between accuracy and computational expense is unavoidable. Empirical scoring functions designed for the fast estimation of binding affinities for high-throughput virtual screening often neglect conformational flexibility of the protein, entropic effects, and desolvation during the binding process (26). At the other end of the accuracy range are techniques based on MD simulations in explicit solvent that make use of free energy perturbation, thermodynamic integration, and linear interaction energy methods (27,28). Given sufficient sampling, MD methods can be used to calculate the free energy of protein/ligand binding with high accuracy, including all entropic and solvent effects as well as receptor flexibility. The main problem associated with this methodology is the computational effort required to adequately sample the full configurational space and also to parametrize each new ligand essayed. Here we investigate a strategy to overcome these two computational bottlenecks based on the use of combined quantum mechanical/molecular mechanical (QM/MM) methods. These methods are now widely used for the analysis of enzymatic reactions (29–31). In combined QM/MM

methods, the ligand/substrate species is treated explicitly by a QM model (32–35), avoiding the derivation of parameters for each new ligand.

The protein and solvent environment is represented by MM force fields, which are computationally efficient. Treating the ligand quantum mechanically and the protein molecular mechanically has the additional advantage of the inclusion of ligand polarization upon binding. Besides these two important computational aspects, there exists an additional point that is worth mentioning. With a QM definition of the ligand, a better treatment of its accessible configurations could be expected, no matter what kind of functional group it may contain. It is important to mention that some functional groups are very difficult to parametrize, in the MM approach, when effects like electronic delocalization are present. There are some recent computational studies that successfully applied the combined QM/MM method to the study of protein-ligand interactions in the HIV-1 protease system (36) and in the trypsin system (37). In this work we propose an additional step in the use of QM/MM method for inhibitors analysis. We show below that the QM/MM interaction energy can be safely used to predict the biological activity of a particular inhibitor, at least when comparing within a given family of compounds. As mentioned before, binding free energies are computationally expensive, whereas interaction energies can be much more easily obtained. In this way two of the largest inconveniences in the use of MD simulations, the need of reparametrization for each new compound and the extensive sampling of different states required to evaluate free energy changes, can be conveniently overcome. In particular, we will show the effectiveness of the proposed strategy for the set of CDK2 inhibitors presented in Fig. 1.

METHODS

Simulations

All simulations were performed using the DYNAMO software package (38). The initial coordinates for the QM/MM calculations were taken from the x-ray crystal structure of the Nu6102 inhibitor bound to the fully activated Thr-160-phosphorylated CDK2-cyclin A complex (T160pCDK2-cyclinA) (Protein Data Bank code 1H1S) (9). The amino acids Arg-297 and Leu-298 were added to the CDK2 subunit of the initial crystal structure with the aid of the program Molden (39) to have a complete protein system. To mimic the aqueous environment, an equilibrated water box with sides of 79.5 Å, centered on the mass center of each inhibitor, was used to solvate the T160pCDK2-cyclinA-ligand system. Amino acid residues and water molecules placed more than 25 Å away from the center of the box were fixed during the simulations. The inhibitors inside the ATP active site were chosen to be the QM subsystem, described at the AM1 level (40), whereas the rest of the T160pCDK2-cyclinA complex and the water molecules of crystallization and those belonging to the solvent box constitute the MM subsystem were described using the optimized potential for liquid simulations (OPLS) force field (41) and the flexible TIP3P potential (42,43), respectively.

There are no bonds between the QM and the MM subsystems; so we do not need to consider any special treatment to complete the valence of the frontier quantum atoms (44). The remaining systems studied were obtained from this initial model by simply deleting and replacing a few atoms of the inhibitor Nu6102 (see Fig. 1). Initially, the hydrogen atoms of the system

were relaxed using the conjugate gradient subroutine implemented in the DYNAMO program (45). Then the full system was minimized up to a gradient tolerance of 1.0 kJ/mol. Subsequently the system was heated up to 300 K by a sequence of MD simulations. Afterwards the system was further equilibrated during additional 100 ps using the NVT ensemble at 300 K. The production run consisted of an MD simulation of 250 ps. In all cases we used a time step of 1 fs to solve the equations of motion and a switched cutoff distance of 13.5 Å. One protein-ligand configuration was saved each 15 time steps for a posteriori energetic and structural analysis. This procedure was applied to the five inhibitors presented in Fig. 1.

Energy decomposition

The energy for a QM/MM system can be obtained as

$$E = \langle \Psi | \hat{H}_0 | \Psi \rangle + \langle \Psi | \hat{V}_{\text{QM/MM}} | \Psi \rangle + E_{\text{MM}}, \quad (1)$$

where \hat{H}_0 is the in vacuum Hamiltonian for the selected QM method, Ψ is the polarized wave function (that is, the wave function obtained in the presence of the MM field), E_{MM} represents the force field energy, and $\hat{V}_{\text{QM/MM}}$ is the coupling operator between the QM and MM subsystems and includes an electrostatic and a van der Waals term. Eq. 1 can be written in a different form to define the QM/MM interaction energy. In our partition scheme the total energy is written as the sum of the in vacuum energy of the QM subsystem, the QM/MM interaction energy, and the MM energy (46):

$$E = \langle \Psi^0 | \hat{H}_0 | \Psi^0 \rangle + [\langle \Psi | \hat{H}_0 | \Psi \rangle - \langle \Psi^0 | \hat{H}_0 | \Psi^0 \rangle] + \langle \Psi | \hat{V}_{\text{QM/MM}} | \Psi \rangle + E_{\text{MM}}, \quad (2)$$

where Ψ^0 stands for the unpolarized (gas phase) wave function of the QM subsystem. The first term on the right side represents the gas phase energy of the quantum subsystem (usually the ligand), the second term (the one inside the brackets) the polarization energy of the quantum subsystem wave function, and the third one the QM/MM interaction energy with the polarized wave function. The sum of these two last terms is the total QM/MM interaction energy. This energy can be further decomposed considering that the coupling operator between the QM and MM subsystems ($\hat{V}_{\text{QM/MM}}$) includes an electrostatic and a van der Waals term. This last contribution is usually evaluated using the Lennard-Jones expression

$$E_{\text{vdW, QM/MM}} = \sum_{i \in \text{QM}} \sum_{j \in \text{MM}} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{R_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{R_{ij}} \right)^6 \right], \quad (3)$$

where R_{ij} is the distance between the QM and MM interaction centers and ϵ_{ij} and σ_{ij} are the Lennard-Jones parameters. As this expression does not involve electronic coordinates, this energy term does not need to be included in the SCF evaluation of the QM subsystem wave function.

The interaction energy can, then, be calculated as the energy difference between the full QM/MM system and the separated QM and MM subsystems. Taking into account that in nonpolarizable force fields the MM energy exactly cancels out, we can write

$$E_{\text{QM/MM}} = E - \langle \Psi^0 | \hat{H}_0 | \Psi^0 \rangle - E_{\text{MM}}, \quad (4)$$

which means that in vacuo single point energy calculation must be carried out for given nuclear configurations of the quantum atoms along the trajectory and then subtract this energy and the MM energy to the full QM/MM potential energy. Note that according to our definition of the polarization energy (term inside the brackets in Eq. 2) as the gas phase energy difference between the polarized and unpolarized wave functions, this is a positive contribution. A different energy decomposition scheme was used in Hensen et al. (36). Finally, it is also worth noting that the set of configurations over which the average is carried out depends on the temperature at which the simulation is performed. So, the average interaction energy is in fact temperature dependent.

RESULTS AND DISCUSSION

In this section we first present the structural features of each of the CDK2-inhibitor complexes. Then we analyze the interaction energies and the relationship to the biological activity (measured as IC_{50}). Both geometrical data and energy terms have been averaged during the production run, and the standard deviation is provided as a number in parenthesis accompanying the average value.

Ligand-active site interactions: compounds 2 (Nu6094) and 25

It is well known that several compounds such as purvalanols (47) and indirubins (48,49) pack aromatic moieties with the specificity surface of CDK2. Compound 2 (Cp2) was initially synthesized taking this property into account. Then, according to Davies et al., the aniline group of Cp2 projects out of the adenine site through a largely hydrophobic tunnel constituted by the side chains of Phe-82 and Ile-10 and packs against the kinase surface, forming a π - π stacking interaction with the peptide backbone between Gln-85 and Asp-86 (9). The final structure obtained from the MD simulations is shown in Fig. 2. As a common feature for the five inhibitors analyzed here, a triplet of hydrogen bonds is formed between the purine ring and the hinge region of CDK2: NH-9 acts as a hydrogen-bond donor to the backbone carbonyl group of Glu-81, and N-3 and 2-NH₂ accept and donate a hydrogen bond to the backbone carbonyl and amide groups of Leu-83, respectively. The averaged distances for the triplet of hydrogen bonds for each ligand are given in Table 1, being the values quite similar for the five compounds. From the analysis of geometrical data from the MD simulation, we have found other weak interactions between the Cp2 and the protein, the C-H... π weak interaction between the α -carbon hydrogen of Gln-85 and the center of the aniline ring of ligand being the most significant. The averaged distance from the hydrogen atom to the centroid of the ring is ~ 2.90 Å. There exists another weak interaction of the type C-H...O=C between one of the aniline ring hydrogens and the carboxylate group of Asp-86 (Cp2 Ph-H to Asp-86 OD2 = 3.35 Å (± 0.49 Å)). It is worth mentioning that the side chain of Lys-89 is tilted away from the aniline ring of the ligand. The NH₃⁺ group of this residue makes a stabilizing hydrogen bond with the carbonyl side-chain group of Gln-85 (Fig. 2).

The potencies of the 4'-substituted C²-anilino-O⁶-cyclohexylmethyl-purine series were evaluated by Hardcastle et al.

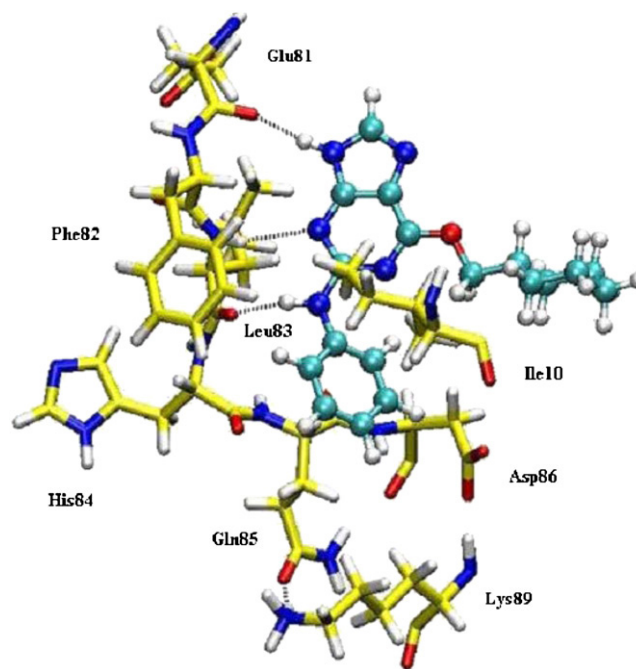


FIGURE 2 Snapshot of the CDK2-Cp2 structure. Selected residues of the CDK2 active site are shown in licorice representation, and Cp2 is presented in ball-and-stick representation, with carbon atoms colored in yellow (CDK2) and cyan (inhibitor). The hydrogen bonds are depicted as black dashed lines.

(15). They found that 4'-hydroxy substitution in this series generated compound 25 (Cp25) which was substantially more potent than the parent aniline Cp2. This result suggested that there was a hydrogen-bond acceptor in the region of the protein accessed by the 2-anilino group. An obvious candidate was the Asp-86 residue. The data obtained from our MD simulations (see Fig. 3) show the hydroxyl group of Cp25 interacting with the carboxylate moiety of Asp-86 via a water-mediated hydrogen bond (Cp25 O-TIP3-Asp-86 OD2 = 3.29 Å/ 2.62 Å (± 0.42 Å/ ± 0.10 Å)). This result is in agreement with the crystallographic data reported by Davies et al., although the hydrogen-bond distances are different from those quoted by them (9). We observed two additional hydrogen bonds between the hydroxyl group and the Lys-89 residue. One of them is formed between the oxygen atom of the hydroxyl group and the NH₃⁺ group of Lys-89 (Cp25 O-Lys-89 NZ = 3.79 Å (± 0.66 Å)). The hydroxyl group also interacts with Lys-89 via a water-mediated hydrogen bond (Cp25 O-TIP21-Lys-89 NZ = 3.31 Å/ 2.94 Å (± 0.37 Å/ ± 0.51 Å)).

TABLE 1 Hydrogen-bond-averaged distances (Å) between the purine ring of each ligand and the hinge region of the CDK2 active site

Fragment	Cp 2	Cp 25	Cp 3	Cp 9	Cp 7
Glu-81 C=O—H-N9	1.96 (± 0.16)	1.88 (± 0.14)	1.92 (± 0.15)	1.90 (± 0.15)	1.91 (± 0.15)
Leu-83 N-H—N3	2.49 (± 0.25)	2.62 (± 0.21)	2.46 (± 0.21)	2.60 (± 0.23)	2.72 (± 0.30)
Leu-83 C=O—H-N2	1.94 (± 0.17)	1.94 (± 0.17)	1.90 (± 0.17)	1.97 (± 0.19)	2.01 (± 0.21)

The values in parentheses correspond to the standard deviations.

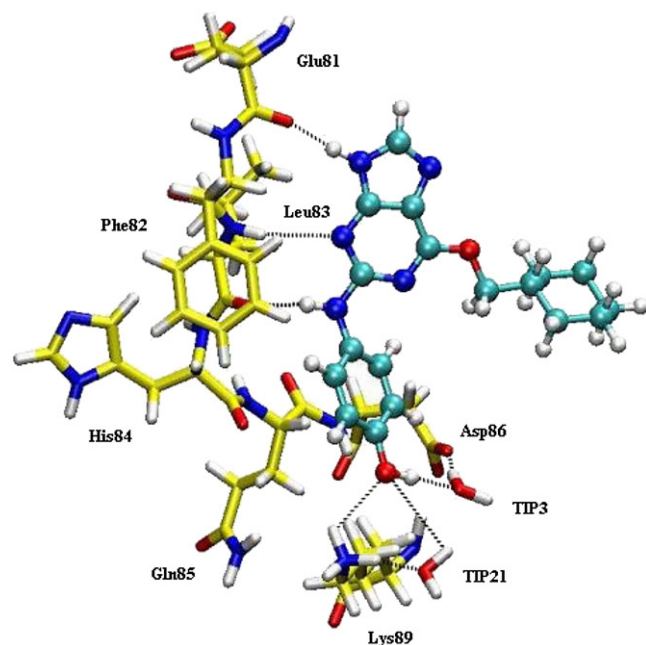


FIGURE 3 Snapshot of the CDK2-Cp25 structure. Representation is made as in Fig. 2.

Compound 3 (Nu6102)

This compound presents a particular interaction pattern within the active site of CDK2 due to the presence of the sulfonamide group. Hardcastle et al. reported that this compound interacts with Asp-86 through the NH_2 group of the sulfonamide, which donates a hydrogen bond to the side-chain oxygen of Asp-86 (Cp3 N26 to Asp-86 OD2 = 2.9 Å) and through one sulfonamide oxygen that accepts a hydrogen bond from the backbone nitrogen of Asp-86 (Cp3 O24 to Asp-86 NH = 3.1 Å) (15). They have further suggested that the high activity of Cp3 is due to the two additional H-bonds formed with Asp-86. However, we have found that the interactions formed for this compound inside the active site of CDK2 are different from those reported before. A representative snapshot of this protein-inhibitor complex as obtained in our MD simulation is provided in Fig. 4. In addition to the triplet of hydrogen bonds established with the hinge region, reported in Table 1, we have found other CDK2-Cp3 hydrogen bonds. The NH_2 group of the sulfonamide donates a hydrogen bond to each of the side-chain oxygen atoms of Asp-86 (Cp3 O₂S-N to Asp-86 OD1 = 3.78 Å (± 0.40 Å) and Cp3 O₂S-N to Asp-86 OD2 = 3.38 Å (± 0.45 Å)). One sulfonamide oxygen atom accepts a hydrogen bond from the NH_3^+ side-chain group of Lys-89 (Cp3 S=O to Lys-89 NZ = 2.80 Å (± 0.15 Å)).

According to Davies et al. (9), although the Cp3 aniline group packed closely with the specificity surface, the sulfonamide group did not form the proposed hydrogen bond with Lys-89. We have found that this interaction does occur and that the sulfonamide group is positioned in a geometry that

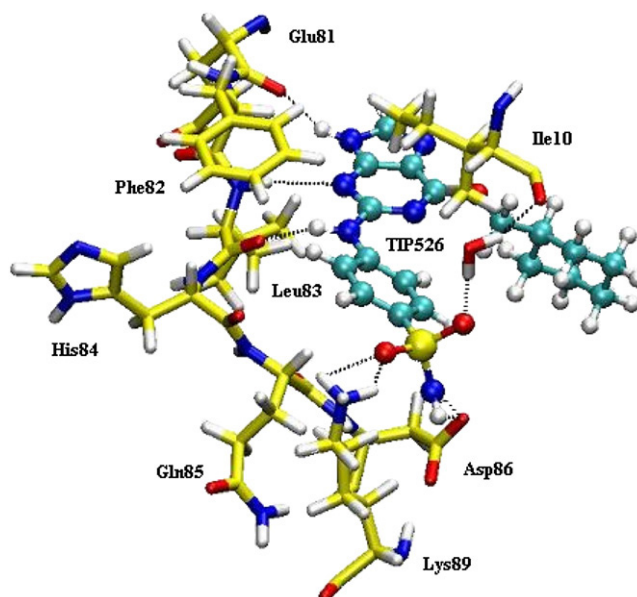


FIGURE 4 Snapshot of the CDK2-Cp3 structure. Representation is made as in Fig. 2.

facilitates this unreported protein-ligand hydrogen bond. Finally, the last hydrogen-bond interaction found for this compound, and not yet reported, occurs between the other sulfonamide oxygen atom and the backbone carbonyl group of residue Ile-10. The hydrogen bond is this time mediated by a water molecule from the solvent water box (Cp3 S=O-TIP526-Ile-10 C=O = 3.07 Å/3.39 Å (± 0.66 Å/ ± 0.76 Å)).

Compounds 9 and 7

These compounds belong to the nonaromatic series studied by Hardcastle et al. (15). They proved to be more potent CDK1 and CDK2 inhibitors than the parent compound Nu2058 (8), and both form a triplet of hydrogen bonds within the CDK2 ATP binding site (see Table 1). Compound 9 (Cp9) is the isopropyl N^2 -monosubstituted derivative of the series studied and forms a weak C-H $\cdots\pi$ interaction with the peptide backbone between Gln-85 and Asp-86 through one of the isopropyl hydrogen atoms. The protein-ligand complex is shown in Fig. 5. The isopropyl group establishes a nonpolar interaction with the side chain of Ile-10 and a weak hydrogen-bond interaction of the type C-H \cdots O=C, with one of the side-chain oxygens of Asp-86 (Cp9 isopropyl-H to Asp-86 OD2 = 3.78 Å (± 0.49 Å)). On the other hand, compound 7 (Cp7) is the methyl N^2 -monosubstituted derivative in this series. This compound establishes the typical triplet of hydrogen bonds with the hinge region of CDK2 (Table 1) and a nonpolar interaction with the side chain of residue Ile-10 (Fig. 6). The lack of a bulky group in this ligand, like the isopropyl in Cp9, causes a diminished interaction with Ile-10.

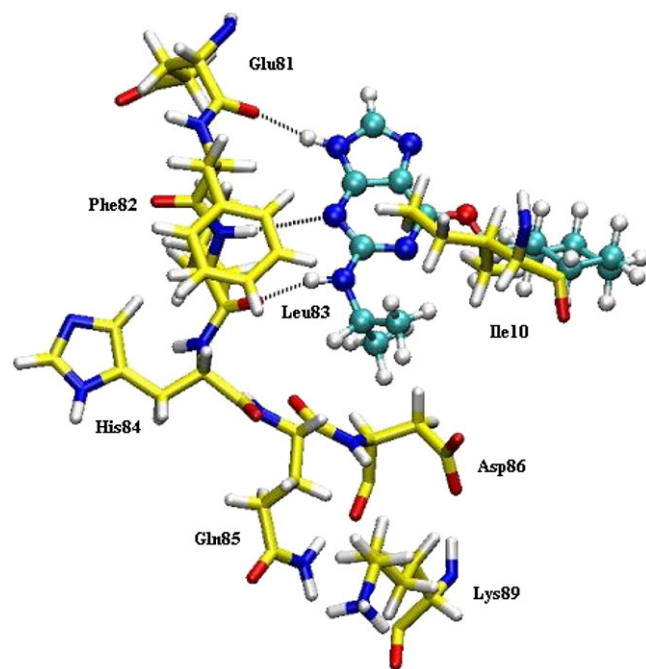


FIGURE 5 Snapshot of the CDK2-Cp9 structure. Representation is made as in Fig. 2.

Analysis of interaction energies

Averaged interaction energies for each of the inhibitors with CDK2, together with their electrostatic and van der Waals components, are reported in Table 2. The biological activity is also provided as the common logarithm of IC_{50} (15). The potency of each ligand is obviously related to its binding free energy but assuming that, within a family of compounds, solvation/desolvation energies, enzyme deformation energy, and entropic changes are proportional to the magnitude of the interactions established between the inhibitor and the protein, one may then express the activity as a function of the interaction energy:

$$\text{Log}IC_{50} = \alpha \Delta E_{\text{QM/MM}} + \beta. \quad (5)$$

The result of a least-square fit of Eq. 5 using data of Table 2 is shown in Fig. 7. The high value obtained for the correlation coefficient ($R = 0.96$) makes us confident in our simulation procedure. In addition, it clearly indicates that interaction energies can be safely used as a predictor of the

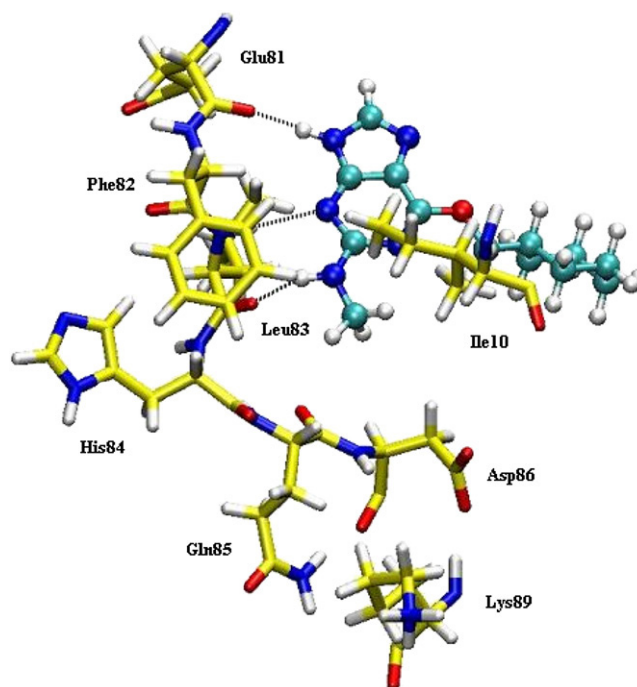


FIGURE 6 Snapshot of the CDK2-Cp7 structure. Representation is made as in Fig. 2.

biological activity of a given inhibitor, at least when comparing them with a series of compounds belonging to the same family. The interest in this kind of correlations is obvious considering the large difference in computational effort needed to evaluate interaction energies or binding free energies. The use of QM/MM procedures has the additional advantage of avoiding extensive reparametrization for each new ligand considered in the study. An interesting practical conclusion can also be drawn from the study of the quality of the fit as a function of the length of the sampling used in the evaluation of the interaction energies. If instead of using the full 250-ps simulation to average the interaction energy, one averages them only over the last 100 ps, the correlation coefficient is slightly reduced up to 0.95, indicating that the sampling, and then the computational effort, could be reduced without a significant loss of quality.

It is also interesting to relate the different interaction energies and their decomposition to the structural features of the different inhibitor-protein complexes. It can be seen from Table 2 that the magnitude of the electrostatic contribution is

TABLE 2 Averaged QM/MM interaction energies and its components (kcal/mol) for the CDK2 inhibitors studied

	Compound 2	Compound 25	Compound 3	Compound 9	Compound 7
Interaction energy	-73.4 (± 2.9)	-86.2 (± 3.5)	-124.5 (± 8.2)	-70.2 (± 4.0)	-65.5 (± 3.1)
Electrostatic energy	-17.5 (± 2.4)	-30.2 (± 3.4)	-65.8 (± 9.0)	-23.4 (± 3.5)	-22.8 (± 3.3)
vdW energy	-55.9 (± 2.2)	-56.1 (± 2.5)	-58.6 (± 3.0)	-46.8 (± 2.4)	-42.7 (± 2.4)
Log IC_{50}	-0.013	-1.161	-2.268	0.079	0.699

Biological activities are reported as the common logarithm of IC_{50} (micromolar) (15).

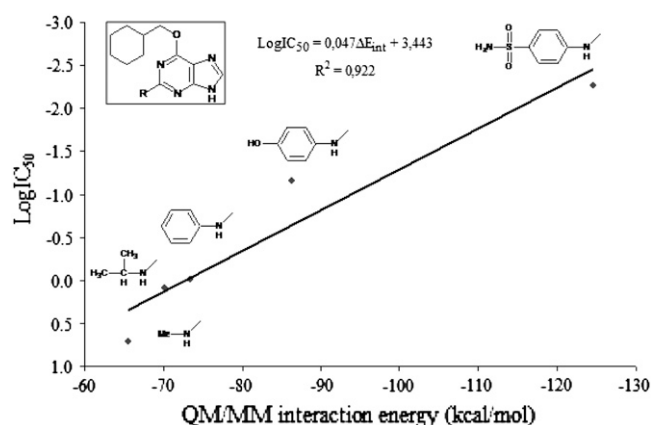


FIGURE 7 Graph of LogIC_{50} versus QM/MM interaction energy (kcal/mol) for the five N^2 -substituted O^6 -cyclohexylmethylguanine derivatives studied and equation obtained by a least-square fit.

similar for compounds 9 and 7, as result of the very similar hydrogen-bond interactions presented by these ligands with the hinge region in CDK2. The van der Waals interaction energy is different for these two compounds due to the difference in the substitution pattern at the N^2 position of the purine ring. The main consequence arising from this difference is that Cp9 bears a more polarizable isopropyl group at the N^2 position, thereby promoting a more favorable interaction with the Ile-10 residue than Cp7, which bears a methyl group at that position. This example clearly shows that the activity can be improved by the optimization of the van der Waals interactions. The opposite is found for Cp2 and Cp25. These compounds show similar van der Waals contributions to the interaction energy, and their electrostatic energy values are different (Table 2). The difference in the electrostatic energy for these two compounds can be attributed to a significant interaction between the 4'-substituent group in the phenyl ring of the aniline group and some amino acids in the specificity surface. For Cp2, the larger van der Waals energy contribution (in absolute value), when compared to the corresponding values for Cp9 and Cp7, is in agreement with the expected interaction pattern with the specificity surface at the active site of CDK2. That means that the presence of an aniline ring in the C-2 position of the purine ring is more favorable than a nonaromatic group like methyl (Cp7) or propyl (Cp9).

The electrostatic energy component for Cp2 is smaller, in absolute value, than the corresponding values found for Cp7 and Cp9. This can be attributed to a less favorable interaction between the purine ring of this compound and the hinge region of CDK2. However, this low contribution is compensated for by an increase in the van der Waals interaction energy. Cp25, on the other hand, shows an increase in the electrostatic energy component due to the hydrogen-bond interactions formed by the $-\text{OH}$ group in the 4'-position of the aniline ring with the residues of the specificity surface of the enzyme. It is also interesting to mention that Cp3 is the

only one that has an absolute value of the electrostatic energy component greater than the van der Waals term. This result can be attributed to the multiple hydrogen-bond interactions that this compound can establish with residues Asp-86, Lys-89, and Ile-10, which are present in the CDK2 specificity surface. Comparing the van der Waals energy value for this compound with those evaluated for Cp2 and Cp25, we can conclude that the van der Waals energy component has a similar weight in these compounds, but this energy component is slightly improved due to the interaction formed with the amino acids in the specificity surface and, consequently, to the more effective packing with the hydrophobic surface within the active site of the CDK2. The final result for Cp3 is an improved interaction and then a larger inhibitor activity.

Finally, to evaluate if our interaction energy partition is a reliable model to represent the interaction between each ligand and the active site of CDK2, a two-variable regression analysis was performed. The two independent variables are the electrostatic energy and the vdW energy components, and the dependent variable is the logarithm of the IC_{50} inhibitory concentration obtained from the experiments. The resulting regression equation is

$$\text{LogIC}_{50} = \gamma + \alpha * \Delta E_{\text{elec}} + \beta * \Delta E_{\text{vdW}} \\ = 4.973 + 0.038 * \Delta E_{\text{elec}} + 0.082 * \Delta E_{\text{vdW}}, \quad (6)$$

where ΔE_{elec} and ΔE_{vdW} , are the electrostatic and the van der Waals interaction energy components, respectively. The standard deviation values for the parameters γ , α , and β involved in the equation above are 1.46, 0.011, and 0.031, respectively. This equation can be used to predict LogIC_{50} values, which are quoted in Table 3. We have plotted these predicted LogIC_{50} values against the experimental LogIC_{50} values. The comparison yields a very good correlation coefficient $R = 0.976$ (see Fig. 8). Interestingly, Eq. 6 displays quite different values for the slopes associated with the electrostatic and van der Waals interaction energies, suggesting that an improvement of this last contribution could be more effective, in terms of inhibition, than a similar energetic improvement in the electrostatic component. Obviously, another question is that obtaining differences in the van der Waals interaction energy may be more difficult than for the electrostatic part, as suggested by the ranges covered by these

TABLE 3 Results of multiple regression analysis (see Eq. 5) for the compounds belonging to the N^2 -substituted-6-cyclohexylmethoxypurine family studied

Compound	Experimental IC_{50} (μM)	Experimental LogIC_{50} values	Calculated IC_{50} values (μM)	Calculated LogIC_{50} values
3	0.0054 (± 0.0010)	−2.268	0.0043	−2.368
25	0.069 (± 0.001)	−1.161	0.158	−0.800
2	0.970 (± 0.030)	−0.013	0.497	−0.303
9	1.200 (± 0.200)	0.079	1.680	0.225
7	5.000 (± 0.300)	0.699	3.820	0.582

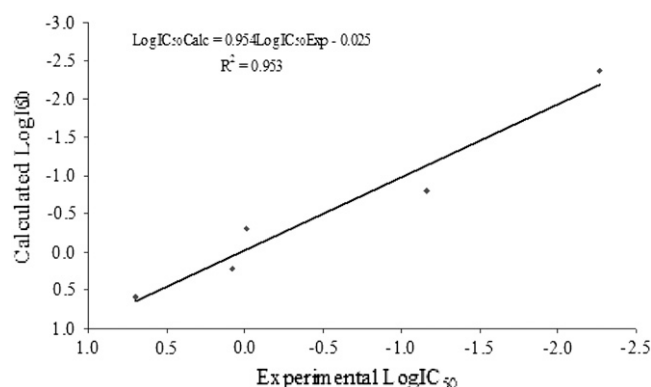


FIGURE 8 Graph of calculated LogIC_{50} values (obtained by means of Eq. 5) versus experimental LogIC_{50} values for the five compounds studied and equation obtained by least-square fit.

two components of the interaction energies in the five cases analyzed here. As shown in Table 2, the maximum difference in the van der Waals interaction energy among the inhibitors studied in this work is 16.9 kcal/mol, whereas in the case of the electrostatic interaction energy this maximum difference amounts up to 48.3 kcal/mol.

We have also tried to correlate LogIC_{50} versus ΔE_{elec} or ΔE_{vdW} separately. The correlation coefficients obtained are significantly worse, 0.89 and 0.82, respectively. This suggests that correlations using only one of those components are not sufficient to describe the behavior of the inhibitors within the CDK2 active site due to the important role that each property has in the global interaction between the inhibitor and the amino acids at the active site. Instead of using those descriptors separately, the whole interaction energy (Eq. 5) or a multiple regression analysis with respect to both terms (Eq. 6) can be used to predict the IC_{50} for a new compound structurally related to the scaffold studied here in a quantitative manner.

CONCLUSIONS

In this study we have demonstrated that the combined QM/MM methods provide a useful approximation to calculate the interactions established at the CDK2 active site by a series of compounds belonging to the N^2 -substituted- O^6 -cyclohexylmethoxypurine family. This method allows us to describe with some detail the nature of the specific interactions involved in the protein-ligand binding. This fact makes the method extensible to other groups of CDK2 inhibitory molecules and, in fact, to other interesting biological systems. Within the approach used here, it is further possible to obtain an acceptable and representative picture of the real system in which the evaluation of the different contributions from the enzymatic and solvent environments to the protein-ligand binding in the active site of the protein becomes a simpler task. The consideration of flexibility of almost the whole protein and the ligand in the calculations performed with the

combined QM/MM methods allows us to achieve a wider sampling of the configurational space. A huge set of possible structures at a relevant temperature in the MD simulations can be incorporated.

With respect to the specific interactions formed inside the CDK2 active site by each of the compounds studied, we can conclude that the enhanced potency observed for Nu6102 (Cp3) results from the hydrogen-bond interactions formed between the sulfonamide group and residues Asp-86, Lys-89, and Ile-10. There is an additional gain in potency that depends on favorable packing between the aniline aromatic ring and the hydrophobic surface in the CDK2 binding site. It is worth mentioning that this is the first time to our knowledge that the interactions of a compound with Lys-89 are reported.

The statistical validation of the proposed method was performed by means of multiple regression analysis. This analysis allows us to conclude that there exists a high degree of confidence in the data presented for the calculated interaction energy components of each inhibitor using the combined QM/MM described before. Taking into account the values of the two variables proposed and their correlation with the biological activity values reported in the literature, it seems possible to predict the IC_{50} value for a new compound. The evaluation of the averaged interaction energies has allowed us to show the strong correlation existing with the biological activity of this series of inhibitors. In fact, the QM/MM interaction energy could be used as a predictor of the biological activity, at least, within this family of compounds. The use of QM/MM interaction energies has a double advantage: i) extensive reparametrization of new ligands is avoided, and ii) the computational effort is considerably smaller than for the evaluation of binding free energies. Energy decomposition shows that in all compounds, except Cp3, the major contribution to the total interaction energy is made by the van der Waals energy component. In some compounds, the hydrogen bonds facilitate the interactions with the hydrophobic surface of the enzyme.

J.H.A.M. thanks DAAD (Deutscher Akademischer Austausch Dienst, Germany) for financial support through a doctoral fellowship and Departamento de Postgrado y Postítulo (U. de Chile) for partial research grant PG/95/2004, and Millennium Nucleus for Applied Quantum Mechanics and Computational Chemistry, grant P02-004-F, Mideplan-Conicyt. We are indebted to Dirección General de Investigación for project BQU2003-04168 and Generalitat Valenciana for projects GV04B-021, GV04B-131, and GRUPOS04/08, which supported this research.

REFERENCES

1. Morgan, D. O. 1995. Principles of CDK regulation. *Nature*. 374:131–134.
2. Sielecki, T. M., J. F. Boylan, P. A. Benfield, and G. L. Trainor. 2000. Cyclin-dependent kinase inhibitors: useful targets in cell cycle regulation. *J. Med. Chem.* 43:1–18.
3. Losiewicz, M. D., B. A. Carlson, G. Kaur, E. A. Sausville, and P. J. Worland. 1994. Potent inhibition of CDC2 kinase activity by the flavonoid L86–8275. *Biochem. Biophys. Res. Commun.* 201:589–595.

4. Senderowicz, A. M., and E. A. Sausville. 2000. Preclinical and clinical development of cyclin-dependent kinase modulators. *J. Natl. Cancer Inst.* 92:376–387.
5. Hardcastle, I. R., B. T. Golding, and R. J. Griffin. 2002. Designing inhibitors of cyclin-dependent kinases. *Annu. Rev. Pharmacol. Toxicol.* 42:325–348.
6. Knockaert, M., P. Greengard, and L. Meijer. 2002. Pharmacological inhibitors of cyclin-dependent kinases. *Trends Pharmacol. Sci.* 23: 417–425.
7. Toogood, P. L. 2002. Progress toward the development of agents to modulate the cell cycle. *Curr. Opin. Chem. Biol.* 6:472–478.
8. Arris, C. E., F. T. Boyle, A. H. Calvert, N. J. Curtin, J. A. Endicott, E. F. Garman, A. E. Gibson, B. T. Golding, S. Grant, R. J. Griffin, P. Jewsbury, L. N. Johnson, et al. 2000. Identification of novel purine and pyrimidine cyclin-dependent kinase inhibitors with distinct molecular interactions and tumor cell growth inhibition profiles. *J. Med. Chem.* 43:2797–2804.
9. Davies, T. G., J. Bentley, C. E. Arris, F. T. Boyle, N. J. Curtin, J. A. Endicott, A. E. Gibson, B. T. Golding, R. J. Griffin, I. R. Hardcastle, P. Jewsbury, L. N. Johnson, et al. 2002. Structure-based design of a potent purine-based cyclin-dependent kinase inhibitor. *Nat. Struct. Biol.* 9:745–749.
10. Gibson, A. E., C. E. Arris, J. Bentley, F. T. Boyle, N. J. Curtin, T. G. Davies, J. A. Endicott, B. T. Golding, S. Grant, R. J. Griffin, P. Jewsbury, L. N. Johnson, et al. 2002. Probing the ATP ribose-binding domain of cyclin-dependent kinases 1 and 2 with O⁶-substituted guanine derivatives. *J. Med. Chem.* 45:3381–3393.
11. Hanks, S. K., and T. Hunter. 1995. The eukaryotic protein kinase super family: kinase (catalytic) domain structure and classification. *FASEB J.* 9:576–596.
12. Gray, N., L. Detivaud, C. Doering, and L. Meijer. 1999. ATP-site directed inhibitors of cyclin dependent kinases. *Curr. Med. Chem.* 6: 850–875.
13. Schulze-Gahmen, U., J. Brandsen, H. D. Jones, D. O. Morgan, L. Meijer, J. Vesely, and S. H. Kim. 1995. Multiple modes of ligand recognition: crystal structures of cyclin dependent protein kinase 2 in complex with ATP and two inhibitors, olomoucine and isopentenyladenine. *Proteins: Struct. Funct. Genet.* 22:378–391.
14. De Azevedo, W. F., S. Leclerc, L. Meijer, L. Havlicek, M. Strnad, and S. H. Kim. 1997. Inhibition of cyclin-dependent kinases by purine analogues. Crystal structure of human CDK2 complexed with roscovitine. *Eur. J. Biochem.* 243:518–526.
15. Hardcastle, I. R., C. E. Arris, J. Bentley, F. T. Boyle, Y. Chen, N. J. Curtin, J. A. Endicott, A. E. Gibson, B. T. Golding, R. J. Griffin, P. Jewsbury, J. Menyerol, et al. 2004. N²-substituted O⁶-cyclohexylmethylguanidine derivatives: potent inhibitors of cyclin-dependent kinases 1 and 2. *J. Med. Chem.* 47:3710–3722.
16. Cavalli, A., C. Dezi, G. Folkers, L. Scapozza, and M. Recanatini. 2001. Three-dimensional model of the cyclin-dependent kinase 1 (CDK1): ab initio active site parameters for molecular dynamics studies of CDKs. *Proteins: Struct. Funct. Genet.* 45:478–485.
17. Sims, P. A., C. F. Wong, and J. A. McCammon. 2003. A computational model of binding thermodynamics: the design of cyclin-dependent kinase 2 inhibitors. *J. Med. Chem.* 46:3314–3325.
18. Sims, P. A., C. F. Wong, and J. A. McCammon. 2004. Charge optimization of the interface between protein kinases and their ligands. *J. Comput. Chem.* 25:1416–1429.
19. Gould, C., and C. F. Wong. 2002. Designing specific protein kinase inhibitors: insights from computer simulations and comparative sequence/structure analysis. *Pharmacol. Ther.* 93:169–178.
20. Wong, C. F., P. H. Hünenberger, P. Akamine, N. Narayana, T. Diller, J. A. McCammon, S. Taylor, and N. H. Xuong. 2001. Computational analysis of PKA-balanol interactions. *J. Med. Chem.* 44:1530–1539.
21. Teague, S. J. 2003. Implications of protein flexibility for drug discovery. *Nat. Rev. Drug Discov.* 2:527–541.
22. Lin, J. H., A. L. Perryman, J. R. Schames, and J. A. McCammon. 2002. Computational drug design accommodating receptor flexibility: the relaxed complex scheme. *J. Am. Chem. Soc.* 124:5632–5633.
23. Carlson, H. A. 2002. Protein flexibility and drug design: how to hit a moving target. *Curr. Opin. Chem. Biol.* 6:447–452.
24. Carlson, H. A., and J. A. McCammon. 2000. Accommodating protein flexibility in computational drug design. *Mol. Pharmacol.* 57: 213–218.
25. Huse, M., and J. Kuriyan. 2002. The conformational plasticity of protein kinases. *Cell.* 109:275–282.
26. Gohlke, H., and G. Klebe. 2001. Statistical potentials and scoring functions applied to protein-ligand binding. *Curr. Opin. Struct. Biol.* 11: 231–235.
27. Aqvist, J., V. B. Luzhkov, and B. O. Brandsdal. 2002. Ligand binding affinities from MD simulations. *Acc. Chem. Res.* 35:358–365.
28. Simonson, T., G. Archontis, and M. Karplus. 2002. Free energy simulations come of age: protein-ligand recognition. *Acc. Chem. Res.* 35: 430–437.
29. Soriano, A., E. Silla, I. Tuñón, and M. F. Ruiz-López. 2005. Dynamic and electrostatic effects in enzymatic processes. An analysis of the nucleophilic substitution reaction in haloalkane dehalogenase. *J. Am. Chem. Soc.* 127:1946–1957.
30. Roca, M., J. Andrés, V. Moliner, I. Tuñón, and J. Bertrán. 2005. On the nature of the transition state in catechol O-methyltransferase. A complementary study based on molecular dynamics and potential energy surface explorations. *J. Am. Chem. Soc.* 127:10648–10655.
31. Martí, S., J. Andrés, V. Moliner, E. Silla, I. Tuñón, and J. Bertrán. 2003. Preorganization and reorganization as related factors in enzyme catalysis: the chorismate mutase case. *Chem. Eur. J.* 9:984–991.
32. Gao, J., and X. Xia. 1992. A priori evaluation of aqueous polarization effects through Monte Carlo QM-MM simulations. *Science.* 258: 631–635.
33. Gao, J. 1995. Methods and applications of combined quantum mechanical and molecular mechanical potentials. *Rev. Comput. Chem.* 7: 119–185.
34. Field, M. J., P. A. Bash, and M. A. Karplus. 1990. A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comput. Chem.* 11:700–733.
35. Martí, S., M. Roca, J. Andrés, V. Moliner, E. Silla, I. Tuñón, and J. Bertrán. 2004. Theoretical insights in enzyme catalysis. *Chem. Soc. Rev.* 33:98–107.
36. Hensen, C., J. Hermann, K. Nam, S. Ma, J. Gao, and H. D. Höltje. 2004. A combined QM/MM approach to protein-ligand interactions: polarization effects of the HIV-1 protease on selected high affinity inhibitors. *J. Med. Chem.* 47:6673–6680.
37. Gräter, F., S. M. Schwarzl, A. Dejaegere, S. Fischer, and J. C. Smith. 2005. Protein/ligand binding free energies calculated with quantum mechanics/molecular mechanics. *J. Phys. Chem. B.* 109:10474–10483.
38. Field, M. J., M. Albe, C. Bret, F. Proust-De Martin, and A. Thomas. 2000. The dynamo library for molecular simulations using hybrid quantum mechanical and molecular mechanical potentials. *J. Comput. Chem.* 21:1088–1100.
39. Schaftenaar, G., and J. H. Noordik. 2000. Molden: a pre- and post-processing program for molecular and electronic structures. *J. Comput. Aided Mol. Des.* 14:123–134.
40. Dewar, M. J. S., E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart. 1985. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* 107:3902–3909.
41. Jorgensen, W. L., D. S. Maxwell, and J. Tirado-Rives. 1996. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* 118: 11225–11236.

42. Jorgensen, W. L., J. Chandrasekar, J. D. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.
43. Steinbach, P. J., and B. R. Brooks. 1994. New spherical cutoff methods for long-range forces in macromolecular simulation. *J. Comput. Chem.* 15:667–683.
44. Gao, J., and D. G. Truhlar. 2002. Quantum mechanical methods for enzyme kinetics. *Annu. Rev. Phys. Chem.* 53:467–505.
45. Hestenes, M. R., and E. Stiefel. 1952. Methods of conjugate gradients for solving linear systems. *J. Res. Nat. Bureau of Standards.* 49:409–436.
46. Martí, S., V. Moliner, and I. Tuñón. 2005. Improving the QM/MM description of chemical processes: a dual level strategy to explore the potential energy surface in very large systems. *J. Chem. Theory Comput.* 1:1008–1016.
47. Gray, N. S., L. Wodicka, A. M. W. H. Thunnissen, T. C. Norman, S. Kwon, F. N. Espinoza, D. O. Morgan, G. Barnes, S. LeClerc, L. Meijer, S. H. Kim, D. J. Lockhart, and P. G. Schultz. 1998. Exploiting chemical libraries, structure, and genomics in the search for kinase inhibitors. *Science.* 281:533–538.
48. Hoessel, R., S. Leclerc, J. A. Endicott, M. E. M. Nobel, A. Lawrie, P. Tunnah, M. Leost, E. Damiens, D. Marie, D. Marko, E. Niederberger, W. Tang, et al. 1999. Indirubin, the active constituent of a Chinese antileukaemia medicine, inhibits cyclin-dependent kinases. *Nat. Cell Biol.* 1:60–67.
49. Davies, T. G., T. Tunnah, L. Meijer, D. Marko, G. Eisenbrand, J. A. Endicott, and M. E. M. Noble. 2001. Inhibitor binding to active and inactive CDK2: the crystal structure of CDK2-cyclin A/indirubin-5-sulphonate. *Structure.* 9:389–397.