Novel Structural Features of CDK Inhibition Revealed by an ab Initio Computational Method Combined with Dynamic Simulations

Lucy Heady,[†] Marivi Fernandez-Serra,[‡] Ricardo L. Mancera,[§] Sian Joyce,[†,||] Ashok R. Venkitaraman,[⊥] Emilio Artacho,[#] Chris-Kriton Skylaris,[∇] Lucio Colombi Ciacchi,*[,†,○] and Mike C. Payne[†]

Theory of Condensed Matter Group, Cavendish Laboratory, University of Cambridge, J. J. Thomson Avenue, Cambridge CB3 0HE, U.K., Départment de Physique des Matériaux, Université Claude Bernard Lyon, 69622 Villeurbanne, France, Western Australian Biomedical Research Institute & School of Pharmacy, and School of Biomedical Sciences, Curtin University of Technology, GPO Box U1987, Perth WA 6845, Australia, Photonics Theory Group, Tyndall National Institute, Lee Maltings, Cork, Ireland, CR UK Department of Oncology & The Medical Research Council Cancer Cell Unit, Hutchison/MRC Research Centre, Hills Road, Cambridge CB2 2XZ, U.K., Earth Sciences Department, University of Cambridge, Cambridge CB2 3EQ, U.K., School of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, U.K., and Fraunhofer Institut für Werkstoffmechanik, Wöhlerstrasse 11, 79108 Freiburg, Germany

Received February 17, 2006

The rational development of specific inhibitors for the ∼500 protein kinases encoded in the human genome is impeded by a poor understanding of the structural basis for the activity and selectivity of small molecules that compete for ATP binding. Combining classical dynamic simulations with a novel ab initio computational approach linear-scalable to molecular interactions involving thousands of atoms, we have investigated the binding of five distinct inhibitors to the cyclin-dependent kinase CDK2. We report here that polarization and dynamic hydrogen bonding effects, so far undetected by crystallography, affect both their activity and selectivity. The effects arise from the specific solvation patterns of water molecules in the ATP binding pocket or the intermittent formation of hydrogen bonds during the dynamics of CDK/inhibitor interactions and explain the unexpectedly high potency of certain inhibitors such as 3-(3*H*-imidazol-4-ylmethylene)-5-methoxy-1,3-dihydro-indol-2-one (SU9516). The Lys89 residue in the ATP-binding pocket of CDK2 is observed to form temporary hydrogen bonds with the three most potent inhibitors. This residue is replaced in CDK4 by Thr89, whose shorter side-chain cannot form similar bonds, explaining the relative selectivity of the inhibitors for CDK2. Our results provide a generally applicable computational method for the analysis of biomolecular structures and reveal hitherto unrecognized features of the interaction between protein kinases and their inhibitors.

1. Introduction

Approximately 500 different protein kinases are encoded in the human genome.[1] The similarity of the mechanism and structure of their catalytic domains remains a major obstacle to the rational development of specific inhibitors for the treatment of human diseases ranging from cancer to autoimmunity. One important case in point involves the family of cyclin-dependent kinases (CDKs), members of which are essential for progression through different stages of the cell division cycle in all eukaryotes.[2] In particular, one family member, CDK2, is required for the events that lead to DNA replication.[3−5] CDK2 and its associated cyclins are overexpressed in cancer cells and might contribute to their deregulated growth.[6] Therefore, the inhibition of CDK2 through the insertion of small molecules into its ATP-binding pocket has long been a potential target for cancer therapies.[6−9] A number of inhibitor molecules have already been designed for this purpose, some of which are currently in clinical trials (e.g., see refs 7 and 9 and references therein). The clinical efficacy of these inhibitors critically depends not only on their potency, that is, their ability to bind

to CDK2 more stably than ATP, but also on their selectivity for CDK2 over other highly homologous members of the CDK family. Nonselective inhibition carries the risk of undesired and potentially toxic side effects.[10] Thus, along with potency, selectivity is a crucial issue in inhibitor design.[9,11,12] However, although the crystallographically determined structures of inhibitors bound to CDK2 have been used as a basis for the rational development of several different inhibitors, features of their potency in CDK2 inhibition or their selectivity for CDK2 over the closely related kinase CDK4 remain to be explained.

We address here the problem of computing relative potencies and explaining the selectivity of five CDK2 inhibitors using an approach that combines classical dynamic simulation with novel ab initio calculations linear-scalable to molecular interactions involving thousands of atoms.[13,14] First principles quantum techniques enable us to calculate the binding energies of hydrogen-bonded systems to high accuracy, whereas classical dynamical simulations allow access to large regions of the potential energy surface and long time scales. As a reference, we also performed a series of extensive docking simulations and scoring functions binding energy calculations.

Our study is based on the crystallographically determined structures of inhibitors bound to CDK2. These provide a clear picture of the binding features between the ligand and those residues that constitute the binding pocket, revealing the dominant interactions to be hydrogen bonding, electrostatic, and van der Waals forces. Potent lead compounds for new CDK inhibitors have been readily developed by taking advantage of these interactions via an analysis of the so-called structure−

* To whom correspondence should be addressed. Tel: +49 761 5142113. Fax: +49 761 5142404. E-mail: lucio@izbs.uni-karlsruhe.de.
† Theory of Condensed Matter Group, University of Cambridge.
‡ Université Claude Bernard Lyon.
§ Curtin University of Technology.
|| Tyndall National Institute.
⊥ Hutchison/MRC Research Centre.
Earth Sciences Department, University of Cambridge.
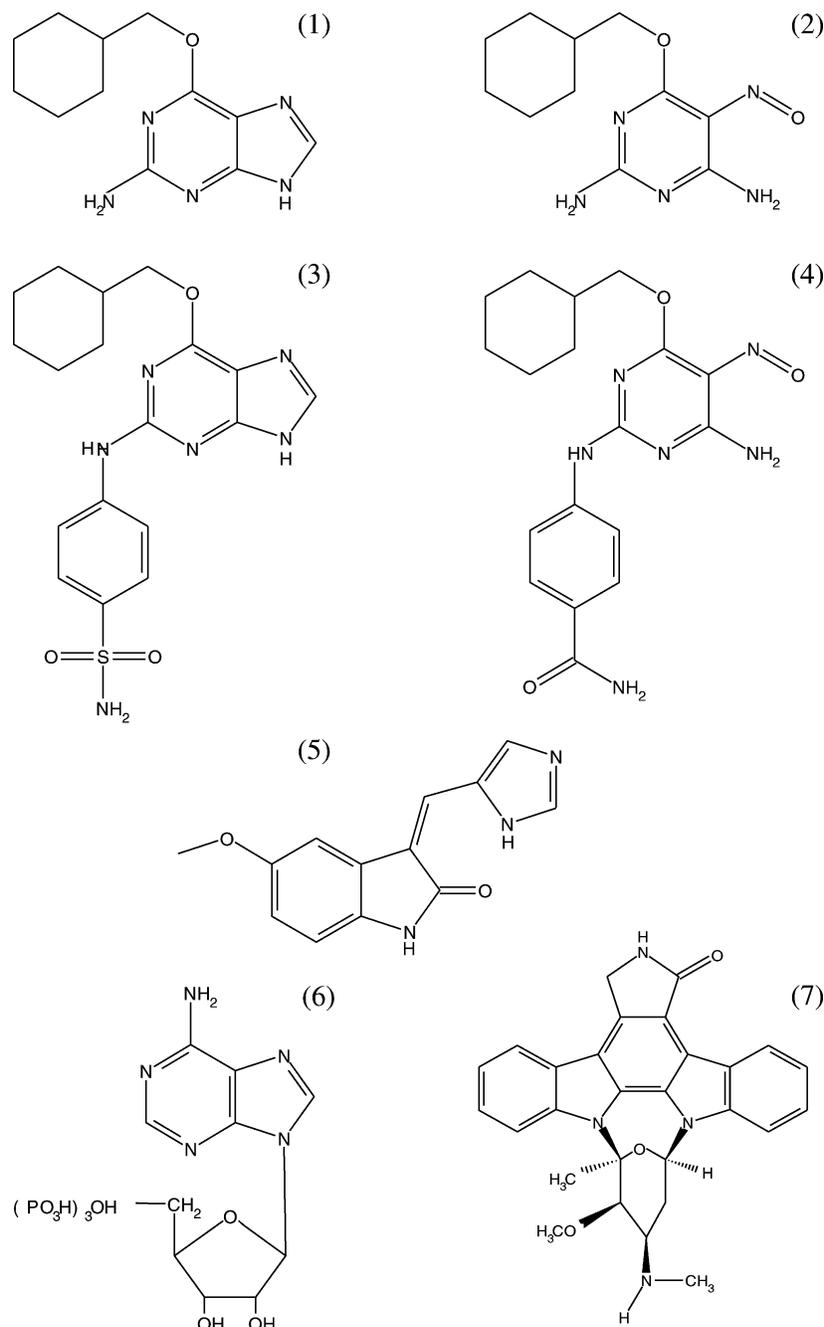∇ University of Southampton.
○ Fraunhofer Institut für Werkstoffmechanik.

Figure 1. Five CDK2 inhibitors considered in this study: NU2058 (**1**), NU6027 (**2**), NU6102 (**3**), 9d-NU6027 (**4**), and SU9516 (**5**); the natural CDK2 ligand, ATP (**6**); and staurosporine (**7**).

activity relationship.[12,15−18] In addition, computer-aided approaches such as database,[19] docking[20−22] and scoring function methods[23] provide an effective way of probing the binding modes and testing the binding strength of large arrays of molecules and thus help identify potential new lead compounds. Moreover, molecular dynamics simulations have proven very useful for the analysis of the binding modes and the structural rearrangements that are connected to the inhibition or activation of CDKs.[24−27] In this work, we extend these techniques to encompass the use of first principles methods, which permit the calculation of hydrogen-bond strengths at an accuracy of about 1 kcal/mol[28] and do not depend on any external set of empirical parameters.

A number of structural studies have demonstrated that the measured potency of inhibition can be directly correlated with the strength of the local protein/inhibitor interactions within the ATP binding pocket (see, e.g., ref 9 and references therein). This idea has led to the rational development of inhibitors via the addition of functional groups designed to encourage specific interactions such as hydrogen bonds with promising residues. Most inhibitors bind to CDK2 in a fashion similar to that of the adenine ring of ATP, forming a triplet of hydrogen bonds to the peptide backbone of residues Glu81 and Leu83, which reside in the hydrophilic hinge region at the back of the binding pocket.[29] With the goal of increasing protein/inhibitor interactions, specific functional groups have been added to O^6-cyclohexylmethylguanine (NU2058)[29] to develop its more potent variant O^6-cyclohexylmethoxy-2-(4′-sulfamoylanilino)purine (NU6102)[30] and to 2,6-diamino-4-cyclohexylmethyloxy-5-nitrosopyrimidine (NU6027)[29] to obtain a potent carboxamide derivative (the 9d variant in ref 17) (Figure 1). Structural investigations demonstrated that the higher potency of these two
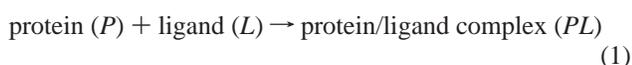
variants is due to the formation of additional interactions with the polar residue Asp86, whereas the standard triplet of hydrogen bonds is retained.[17,30] A puzzling exception is 3-(3*H*-imidazol-4-ylmethylene)-5-methoxy-1,3-dihydro-indol-2-one (SU9516), which appears to form only the standard hydrogen bond triplet but is highly potent and also presents a good degree of selectivity for CDK2 against CDK4.[31,32] Unlike potency, the issue of selectivity has proven to be much harder to explain via an analysis of the available crystal structures alone.[11,12] In the particular case of SU9516, the observed selectivity had previously been proposed to arise from an interaction with the CDK2 specific residue Lys89,[31] but no such interaction was found in a recent crystallographic study.[32] Indeed, the Lys89 residue was also a target for the two derivatives mentioned above, NU6102 and 9d-NU6027, but again, no such interaction has been observed in the crystallographic studies performed.[17,30]

The potency of inhibition can be quantitatively described by the so-called IC_{50} value, that is, the concentration of inhibitor that is required to reduce the activity of the protein by 50% with respect to a chosen reference state in the absence of inhibitor. We note that the IC_{50} value is dependent on the specific experimental conditions, in particular on the ATP concentration used in the activity assay. A more objective measure of the potency, which can be used to compare results from different assays, is the inhibition constant K_i. This is defined as the equilibrium dissociation constant for the reaction of an inhibitor with a protein to form an inhibitor/protein complex. Under the same assay conditions, K_i and IC_{50} values are related via a simple proportionality relationship[33] so that often the values can be directly compared.[17] Therefore, both IC_{50} and K_i values can be assumed to be directly related to the free energy change associated with the inhibitor binding to the protein.[30] Our ultimate aim here is to reproduce the experimentally measured relative potencies of the five inhibitors mentioned above against CDK2 by calculating the differences between their free energies of binding.

Our calculations reveal that a static analysis of the hydrogen bond patterns formed by the inhibitors in the ATP binding pocket is generally not sufficient to predict the correct rank order of inhibitor potencies. In particular, electronic structure calculations at the DFT level appear to be necessary to quantitatively account for the screening of the hydrogen-bond interactions due to the presence of water molecules in the binding cleft. Moreover, dynamical simulations reveal the profound effect of the motion of certain key residues that are involved in hydrogen bonding within the binding pocket. Only by taking into account the dynamical nature of the protein/ligand interactions are we able to rationalize the measured potency and explain the observed CDK2 selectivity of the inhibitors considered.

2. Free Energy of Binding and Potency of Inhibition

The potency of an inhibitor is measured in terms of the inhibition constant K_i, which is inversely proportional to the association constant K_a of the reaction.

$$\text{protein } (P) + \text{ligand } (L) \rightarrow \text{protein/ligand complex } (PL)$$
$$(1)$$

This process involves a change of free energy $\Delta G = -k_B T \ln K_a$, where k_B is the Boltzmann constant, and T is the absolute temperature.

Let us first split ΔG into gas-phase and solvation contributions and the former in enthalpic and entropic contributions:

$$\Delta G = \Delta G_g + \Delta G_{solv} = \Delta E_g - T\Delta S_g + \Delta G_{solv} \quad (2)$$

where ΔG_{solv} is defined as the difference $G_{solv}(PL) - G_{solv}(P) - G_{solv}(L)$, and ΔE_g is defined as $E_g(PL) - E_g(P) - E_g(L)$. The enthalpic term ΔE_g can be accurately calculated using the DFT approach, whereas both the solvation free energy contributions ΔG_{solv} and the gas-phase entropic contributions ΔS_g are in general not readily accessible within this computational scheme. However, when we are looking at the relative potencies between two inhibitors, 1 and 2, then we are only interested in the ratio K_i^1/K_i^2. This enables us to look at the difference between their associated free energy gains, $\Delta\Delta G = \Delta G^1 - \Delta G^2$, which can be directly compared with the K_i values via the relationship

$$\Delta G^1 - \Delta G^2 = k_B T \ln\frac{K_i^1}{K_i^2} \quad (3)$$

and can be written as follows:

$$\Delta\Delta G = \Delta\Delta E_g + \Delta\Delta G_{solv} - T\Delta\Delta S_g \quad (4)$$

Following ref 34, we can assume that the difference between the gas-phase entropy changes $\Delta\Delta S_g$ is negligible within the error bar associated with the calculation of the remaining terms. This appears to be justified in the present case given the large similarity between the systems considered.

Within this approximation, we are left with only two terms:

$$\Delta\Delta G = \Delta\Delta E_g + \Delta\Delta G_{solv} \quad (5)$$

We calculate the gas-phase enthalpy differences ΔE_g within the DFT approach and the solvation free energy differences ΔG_{solv} classically, using the Generalized Born/Surface Area (GBSA) model. Continuum theories have been used to compute free energies of solvation with success for over 20 years,[35−37] accounting for entropic and hydrophobic effects in a simple and computationally inexpensive manner. This allows us to avoid all of the difficulties involved in first principles simulations of polar solvents.[38−40] However, we are aware that more refined formalisms could be used to treat solvation effects to higher accuracy,[41] and in particular, we will pay great attention to the specific hydration patterns of the *PL* complexes within the binding pocket, including explicit solvent water molecules in our final set of calculations (Section 4.1). Finally, using classical force-field techniques we have checked that dispersion forces, which are severely underestimated in the standard DFT approach, contribute in a nearly equal way to the binding energy of all inhibitors to CDK2 and, therefore, do not contribute to the free energy differences $\Delta\Delta G$ (see Computational Details).

3. Static Binding Energy Calculations

3.1 Convergence of Binding Energy with the System Size. Within the approximations described in the previous section, we are, in principle, able to compute the relative differences of binding free energy $\Delta\Delta G$, provided that we succeed in calculating the total energy (enthalpy) of the protein and the protein/ligand complexes. Inclusion of the whole CDK protein in our calculation would be prohibitively expensive in terms of computer time, but this is unnecessary in the present case. Because the small inhibitors bind to a well defined and spatially limited region of the protein (the ATP binding pocket), we are able to consider only those amino acid residues that are close enough to the inhibitor to significantly contribute to the binding energy. In doing this, we must strike a delicate balance between the gain in accuracy from using a DFT approach and the loss of a realistic model due to limiting the size of the systems
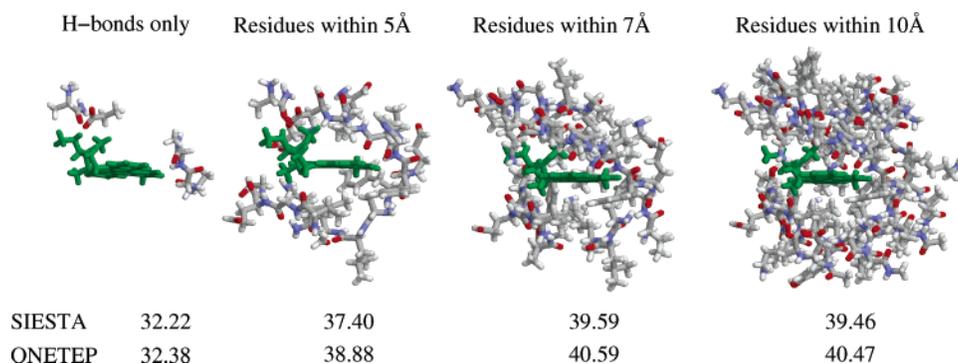
Figure 2. Convergence with CDK fragment size of binding energy values (kcal/mol) calculated using two different O(N) DFT codes in the case of staurosporine bound to the ATP pocket of CDK2.

considered. To select the smallest possible fragment of CDK2, which ensures that all important interactions in the pocket are taken into account, we tested the convergence of ΔE_g with respect to increasing fragment size.

The inhibitor used in this convergence test is staurosporine (inhibitor **7** in Figure 1). Because of its large size, the calculated binding energy of staurosporine to CDK2 is expected to be strongly influenced by long-range electrostatic interactions, and hence to be very sensitive to the size of the system used to model the binding pocket. Our starting point is the available crystal structure of staurosporine in complex with CDK2 (pdb key 1AQ1).[42] As a first model, we consider staurosporine surrounded by only those amino acids with which it makes direct hydrogen bonds (E81, L83, D86, and Q131). Because the side chains of E81 and L83 do not directly interact with the inhibitor, they are replaced by methyl groups bound to the respective C_α (Figure 2). Keeping the C_α atoms of the protein backbone constrained in their original X-ray positions, we fully minimize the geometry of the system using the DFT approach (see Computational Details). This results, in particular, in the accurate estimation of the hydrogen bond lengths and angles between the inhibitor and the protein. Three larger models are then constructed, adding to the minimized small system all amino acids within 5.0, 7.0, and 10.0 Å from the inhibitor (Figure 2). The binding energy is then calculated for all models as the difference in total energy between the protein/inhibitor complex, the isolated protein model, and the isolated inhibitor. With the exception of the smallest model, we chose not to minimize the geometry of the systems for the following reasons. First, the potential energy surface of a large number of amino acids is expected to be very flat and contain a large number of local minima so that (i) the minimization would take a long time and would be very expensive to perform at the quantum level of accuracy, and (ii) even after full minimization, it is very hard to know whether the correct global minimum has been found. Second, in principle, we should minimize both the protein/inhibitor and the isolated protein models. However, in this case, the binding energy would be affected by differences in the relative positions of amino acids far away from the inhibitor, with the risk of wrongly estimating the binding energy because of the aforementioned difficulties in finding the global energy minimum. Therefore, we assume here that the available crystal structures represent an average configuration of the amino acids reliable enough to compute long-range interactions to an accuracy within the error bars associated with all other approximations. Binding energies of systems that are fully relaxed using the classical force-field approach will be presented in Section 5.

Table 1. Binding Energies, Solvation Energies, and Relative Binding Free Energy Differences (All in kcal/mol) of Five CDK2 Inhibitors Calculated at the DFT Level of Theory Compared to the Experimentally Measured Inhibition Constants and Relative Difference of Free Energy of Binding (in Italics)[a]

inhibitor	K_i (μM) exp.	ΔE_g(HB)	ΔE_g(7 Å)	ΔG_{solv}	$\Delta\Delta G$	$\Delta\Delta G$ exp.
1	*(1.2 ± 0.3)E+1*	−20.0	−18.3	19.7	0.0	*0.0*
2	*(1.3 ± 0.2)*	−18.5	−18.4	21.4	+1.6	*−1.4*
5	*(3.1 ± 0.6)E−2*	−14.6	−15.1	19.4	+2.9	*−3.7*
4	*(2.4 ± 0.8)E−2*	−37.1	−43.4	41.4	−3.4	*−3.8*
3	*(6.0 ± 0.5)E−3*	−43.3	−48.2	32.7	−14.1	*−4.7*

[a] The binding energy values are obtained starting from the available crystal structures after geometry optimization of the hydrogen-bond distances of the protein/ligand complexes.

The calculated binding energy values using both the SIESTA and the ONETEP codes are reported in Figure 2, alongside the model systems used. We find that the binding energy converges to within 0.15 kcal/mol for the third model of the binding pocket, that is, when all amino acids at a distance of 7 Å from the inhibitor are explicitly considered. In the next section, models of this size will be used to compute the binding energy of all other inhibitors considered using the O(N) DFT approach (see Computational Details). As a further validation test of the O(N) DFT techniques employed, we have calculated binding energy values for the first model of Figure 2 using also the traditional plane-wave formalism. The computed binding energies obtained with the CASTEP, SIESTA, and ONETEP codes are −32.0, −32.2, and −32.4 kcal/mol, respectively. This confirms that both O(N) DFT approaches faithfully reproduce with equivalent accuracy the results of the traditional PW implementation.

3.2 Static Calculations Using the Available Crystal Structures. As mentioned in the Introduction, we consider in our study inhibitors NU2058 (**1**), NU6027 (**2**), NU6102 (**3**), the 9d-variant of NU6027 (**4**), and SU9516 (**5**) (Figure 1) (for simplicity, inhibitors are referred to by their associated number in the following text). Experimentally determined values of K_i are available for inhibitors **1**, **2**,[29] **3**,[30] and **5**.[32] Moreover, IC_{50} values for **1**−**4** have been determined in the same assay.[17] Therefore, we were able to obtain an estimate of the K_i value for **4** using the proportionality relationship between IC_{50} and K_i.[33] The experimental K_i values of all these inhibitors are reported in Table 1.

To investigate the interactions between the inhibitors and the protein, we start from the crystallographically determined structures of the *PL* complexes which are available in the RCSB Protein Data Bank. Here, one has a choice between structures where the inhibitors are bound to the active form of CDK2, that is, the Thr160-phosphorylated CDK2/cyclin A complex,

Table 2. Relative Differencies in the Free Energies of Binding Calculated with Three Different Sets of Scoring Functions after Extended Docking Simulations, Compared with the Experimental Values (in Italics)[a]

inhibitor	PLP	Böhm	ScreenScore	$\Delta\Delta G$ exp.
1	0.0	0.0	0.0	*0.0*
2	+2.1	+0.5	−0.3	*−1.4*
5	−7.2	−1.5	−1.0	*−3.7*
4	−35.5	−1.4	−1.4	*−3.8*
3	−37.0	−3.1	−3.7	*−4.7*

[a] PLP values are in arbitrary energy units, and Böhm and ScreenScore values are in kcal/mol.

or structures where they are bound to the inactive form, that is, monomeric CDK2. A major structural difference between the two forms lies in the availability (in the inactive form) versus unavailability (in the active form) of the K33 residue within the binding pocket. It has been suggested that when available K33 strongly interacts with the natural ATP ligand in a way that suppresses its turnover.[4,5] Therefore, the binding energy of inhibitors to the inactive form of CDK2 is expected to be affected by the presence of K33 in the binding pocket. Namely, the presence of the K33 ligand in the binding pocket would result in spurious interactions that are in fact absent in the actual activity assays from which the K_i values are extracted. Therefore, because our aim is to present a consistent comparison of the binding energies of the inhibitors related to their inhibition activity, we chose to consider only the interactions between the inhibitors and the binding pocket of CDK when K33 was not available. Indeed, the active CDK2/cyclin A complex is increasingly used as the reference structure in structure−activity-relationship investigations of drug activity.[30]

In the RCSB database, structures are available for **1**, **3**, and **4** bound to the Thr160-phosphorylated cyclin-A/CDK2 complex (pdb keys are 1H1P, 1H1S, 1OGU, respectively) and for **2** and **5** bound to the monomeric CDK2 form (pdb keys are 1E1X and 1PF8, respectively). To prevent the calculated binding energies from being affected by strong spurious interactions with polar residue K33, this has been substituted with an alanine residue in the two latter cases. In the available crystal structures, all inhibitors are observed to form a triplet of hydrogen bonds with E81 and L83. In addition, D86 is observed to form bonds with **3** and **4**, which leads to their increased potency.

Calculations of the binding energies of all inhibitors are performed as described in the case of staurosporine (Section 3.1), considering all amino acids within 7 Å from each inhibitor. Namely, the hydrogen-bond distances are optimized using the PW-DFT approach considering only those amino acids that are directly bound to the inhibitors in the crystal structure. Then, this optimized model is embedded in the larger pocket model, and the binding energy ΔE_g is calculated using the O(N) DFT approach without any further geometry optimization. The obtained binding energy values are reported in Table 1 along with the values of the solvation free energy differences, ΔG_{solv}, computed within the GBSA model. From these values, we are able to compute the relative free energy differences between the inhibitors, $\Delta\Delta G$, using eq 5, where we take as a reference the binding energy value of **1**, which is the least potent inhibitor considered. As a comparison, we report in Table 2 the values of binding energy differences calculated with three different sets of scoring functions after extended docking simulations, as described in the Computational Details section.

The results of the docking simulations appear to be roughly consistent with the measured rank of potencies (except in the case of inhibitor **2**). However, quantitative agreement between the measured and calculated binding energies could not be

achieved, the values of $\Delta\Delta G$ being consistently underestimated. However, the agreement between the relative energies of binding calculated from static structures within the DFT formalism and the experimentally determined relative potencies is worse. In particular, the binding energies of inhibitor **3** is highly overestimated with respect to the binding energy of inhibitor **1**. Furthermore, the wrong rank order of potency is predicted for inhibitor **2** and, more dramatically, for **5**. We note that the binding energy of **5** is far too small even when only the direct hydrogen bond interactions are considered. In this case, a strong interaction (e.g., a direct hydrogen bond) seems to be missing in the model considered. However, the recently resolved X-ray structure of this inhibitor bound to CDK2[32], from which our model was constructed, reveals no interactions other than the usual triplet of hydrogen bonds with the backbone of E81 and L83.

In general, the inconsistencies between the measured and calculated values are too large to be due to approximations, such as the neglect of gas-phase entropic contributions or the limited size of the models considered. Rather, these results suggest that the precise solvation patterns of the protein/inhibitor complexes need to be explicitly considered, and a more careful analysis of the potential energy surface of the binding modes needs to be performed. In the next section, we tackle these issues in detail by performing a dynamical analysis of the binding modes of the inhibitors bound to the active form of CDK2 in the explicit presence of water solvent molecules.

4. Dynamical Force-Field Simulations

Our investigations start with a dynamical analysis of the isolated CDK2 protein in explicit water solvent (pdb key 1HCL) and bound to its natural ligand, ATP (pdb key 1HCK).[43] After minimization and equilibration as described in the Computational Details section, we perform a 0.5 ns run for each of these two systems. We find that the TIP3P water solvent and the employed force field are able to give the correct hydration shell for residues in the interior of the protein. This is immediately visible from a comparison between the positions of the crystallization water present in the crystal structure and the positions of water molecules with long residence time around selected amino acids (e.g., see Figure 3a and b). We stress that the inputs for the dynamical simulations were prepared by stripping all crystallization water molecules from the original pdb files. Hence, this also demonstrates that 0.5 ns is a sufficient period of time to allow water molecules to diffuse into the inner cavities of the protein and give the correct hydration pattern. In the absence of the ATP ligand, water molecules are found to form a nearly planar hydrogen-bonded network inside the binding pocket (Figure 3c). This is consistent with the strong hydrophilicity of the back region of the binding pocket in an otherwise hydrophobic protein environment. The adenine ring of ATP binds into the binding pocket of CDK2 via two hydrogen bonds between the N1 and N6 atoms and the residues L83 and E81, respectively, as expected from previous experimental and theoretical investigations[25,26,43] (Figure 3a and b). These hydrogen bonds are found to be stable and present for the whole 0.5 ns simulation. We then performed dynamical simulations of about 1 ns for inhibitors **1** and **2** and of about 4 ns for inhibitors **3**−**5**. As mentioned above, there is no available crystal structure of inhibitors **2** and **5** bound to the active CDK2/cyclin A complex. Therefore, in these two cases, we constructed the initial inputs by placing the inhibitors in the binding cleft of the CDK2/cyclin A model extracted from its X-ray structure in complex with inhibitor **1**. In agreement with the crystal
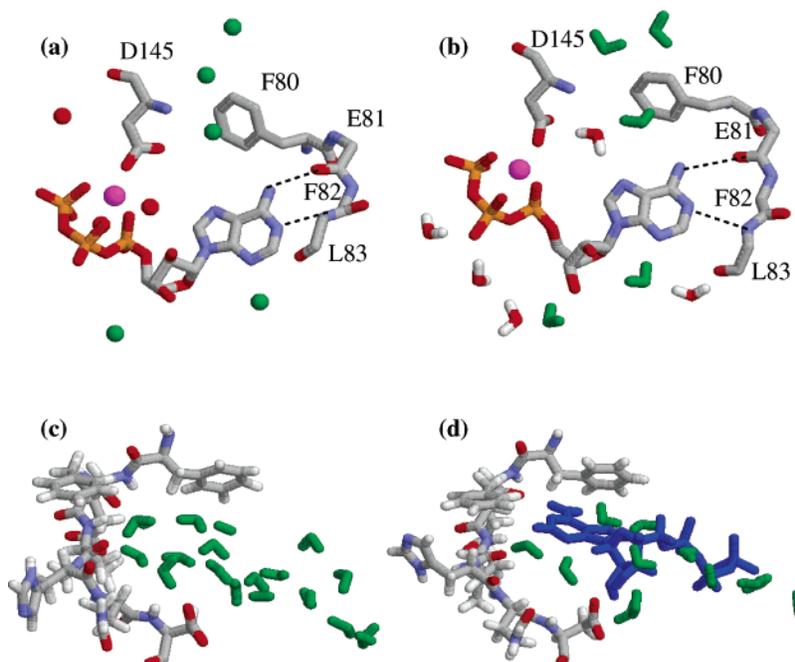
Figure 3. Comparison between the position of selected solvation water molecules visible in the crystal structure of the CDK2/ATP complex (a) and the position of water molecules with long residence time in an MD simulation of the same system (b). Corresponding molecules are highlighted in green, and the hydrogen bonds formed by ATP in the binding cleft are shown with dotted lines. In (c) and (d), a comparison is shown between the planar network of water molecules and the position of the ATP ligand within the binding cleft of CDK2.
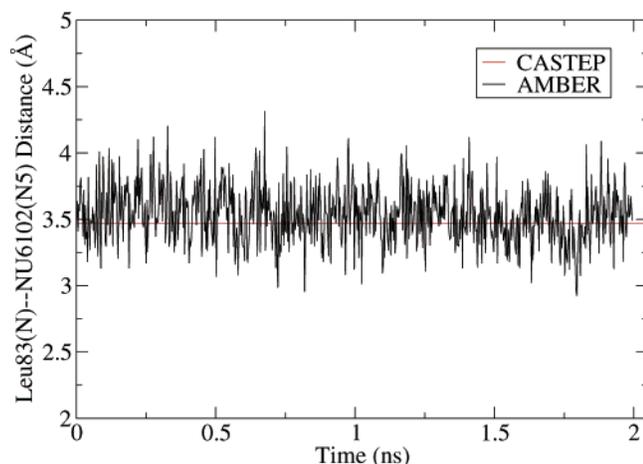

Figure 4. Evolution of the hydrogen bond distance between inhibitor **3** and the peptide nitrogen of Leu 83 during an MD dynamical simulation compared with the distance calculated at the DFT level (horizontal line).
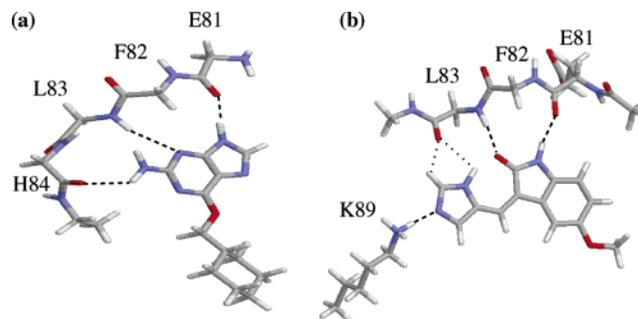

Figure 5. Binding modes of inhibitors **1** (a) and **5** (b) in the ATP binding cleft of CDK2 as obtained from MD simulations. The hydrogen bonds formed are indicated with dotted lines.

structures, all inhibitors are found to form hydrogen bonds in the hydrophilic hinge region at the back of the binding pocket. These bonds are present for the entirety of the simulations, and the average bond distances are in good agreement with the optimized bond lengths obtained after full geometry minimization within the DFT approach (Figure 4). Inhibitors **2**−**4** remain bound to the O atom of E81 and the N and O atoms of L83. In the case of inhibitor **1**, however, after ∼35 ps of dynamics at 300 K, the initial hydrogen bond with the O atom of L83 breaks and is immediately replaced by a new hydrogen bond with the nearby lying O atom of H84 (Figure 5). This binding mode is mantained for the rest of the simulation, which is stopped after 1 ns. Inhibitor **5** is also observed to form two strong hydrogen bonds with the O atom of E81 and the N atom of L83 but does not form a strong hydrogen bond with the O atom of L83. Indeed, the donor-H-acceptor angle of this bond is ∼101°, which indicates the presence of a weaker interaction. In addition, the

O atom of L83 interacts via hydrogen bonding with the C1 atom of the inhibitor, whereas an internal hydrogen bond between the N1 atom and the O1 of the inhibitor is present in the minimized structure of the protein/inhibitor complex (Figure 5).

Interestingly, hydrogen bonds formed with residues other than E81 and L83 are found to be considerably less stable during dynamics. More specifically, the donor−acceptor distances show an intermittent behavior as the dynamics simulation evolves, indicating that these bonds can be reversibly formed and broken at room temperature. In particular, the hydrogen bonds donated by inhibitor **3** to the carboxylic group of the D86 residue, which are visible in the crystallographic structure of the CDK2/inhibitor complex, are in fact found to be present for only ∼21% of the simulation time (Figure 6 and Table 4). In contrast, the corresponding interaction between D86 and inhibitor **4** is present for the whole simulation time. However, we observed that D86 binds to the inhibitor alternatively switching between the OD1 and the OD2 atoms of its carboxylic group (Figure 6). Furthermore, we find that additional hydrogen bonds, which were not indentified in the available X-ray structures, can temporarily form during the dynamics. One dramatic example is the bond that we see forming between inhibitor **5** and the NH_3^+ group of K89. In the crystal structure,[32] the distance
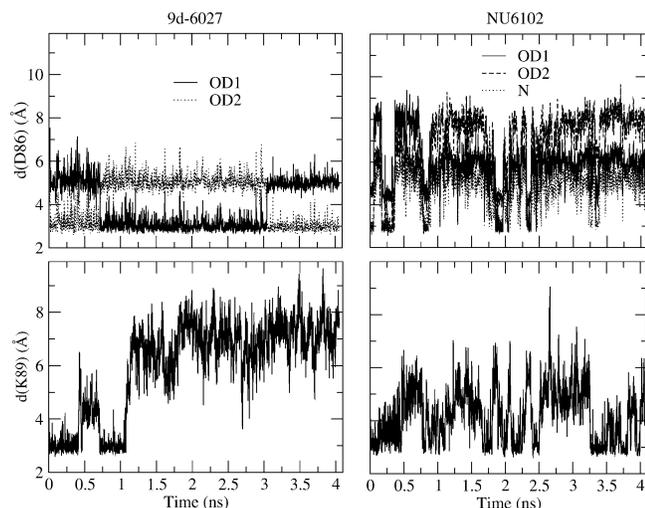
Figure 6. Evolution of the distances of hydrogen bonds formed between inhibitors **4** (left) and **3** (right) and residues D86 (top) and K89 (bottom) of CDK2.

Table 3. Binding Energies (kcal/mol) Calculated at the DFT Level of Theory[a] as Well as the Contribution to the Binding Free Energy due to Solvation ΔG_{solv}^{wat}

inhibitor	ΔE_g(HB)	ΔE_g(7 Å)	ΔE_g(7 Å dry)	ΔG_{solv}^{wat}
1	−16.2	−12.5	−17.4	10.7
2	−17.3	−13.8	−18.5	11.4
5	−50.3	−34.3	−51.3	9.7
4	−57.5	−36.7	−59.1	15.4
3	−74.6	−63.6	−81.7	19.4

[a] The values were calculated after MD simulations and geometry minimization of the protein/inhibitor complexes using the AMBER force-field and DFT optimization of the hydrogen bond distances, considering three models of the binding pocket as described in the text in Section 4.1.

between the NZ atom of K89 and the N2 atom of the inhibitor is 7.1 Å. After about 0.3 ns of simulated dynamics at room temperature, however, this distance suddenly decreases to about 3 Å, indicating the formation of a hydrogen bond (Figure 7). This bond is then repeatedly broken and reformed as the dynamics simulation proceeds and is present for ∼20% of the time over a total simulated time of more than 4 ns (see next section). In an analogous fashion, K89 is observed to form intermittent hydrogen bonds also with inhibitors **4** and **5** (Figure 6). Intermittent hydrogen bonding thus appears to be a key element of the interaction between CDK2 and its inhibitors and needs to be taken into account when calculating the binding energies of the protein/inhibitor complexes.

4.1 Calculation of Binding Energies from the AMBER Structures. To quantitatively account for the intermittent hydrogen bonds, we first performed a full structural minimization within the classical force field starting from a snapshot of the dynamics when all hydrogen bonds are present at the same time. A model is then extracted from the minimized structure, including only the inhibitor and those amino acids that are in direct hydrogen-bond contact with the inhibitor. As done in the case of the models extracted from the crystal structures, the side chains of the selected amino acids are replaced by methyl groups whenever they do not participate in any hydrogen bond. The geometry of this small inhibitor/pocket model system (HB model) is fully minimized using the PW-DFT approach, keeping the C_α atoms of the backbone fixed. Finally, the binding energy of the inhibitor to the small system is calculated as in Section 3.2 (ΔE_g(HB) values in Table 3). This minimal pocket model is then embedded in a larger model, comprising all amino acids

Table 4. Hydrogen Bonds Visible in the Crystal Structures (X-ray) and during Force-Field MD Simulations (MD)[a]

inhibitor	H-bond (X-ray)	H-bond (MD)	length (Å) (X-ray)	length (Å) (DFT)	t (%)
1	E81(O)	E81(O)	2.73	2.86	100
	L83(N)	L83(N)	3.00	3.61	100
	L83(O)	*H84(O)*	2.96	2.95	100
2	E81(O)	E81(O)	2.79	2.81	100
	L83(N)	L83(N)	2.95	3.12	100
	L83(O)	L83(O)	2.83	2.85	100
5	E81(O)	E81(O)	3.00	2.77	100
	L83(N)	L83(N)	2.94	2.92	100
	L83(O)	L83(O)	4.75	3.06	100
		K89(NZ)		2.77	20
4	E81(O)	E81(O)	3.03	2.79	100
	L83(N)	L83(N)	3.24	3.54	100
	L83(O)	L83(O)	2.92	3.07	100
	D86(OD2)	D86(OD1,OD2)	2.79	2.70	100
		K89(NZ)		2.53	27
3	E81(O)	E81(O)	2.97	2.76	100
	L83(N)	L83(N)	3.15	3.27	100
	L83(O)	L83(O)	2.97	2.99	100
	D86(N)	D86(N)	3.74	3.18	21
	D86(OD2)	D86(OD1,OD2)	2.72	2.58	21
		K89(NZ)		2.65	26

[a] Donor−acceptor pairs and hydrogen bond distances after DFT optimization are reported together with the time percentage for which each bond is present in the MD simulations. The residues making hydrogen bonds not visible in the X-ray structure are reported in italics. OD1 and OD2 are the oxygen atoms of the carboxyl group of aspartic acid, NZ is the nitrogen atom of the NH_3^+ group of lysine, and O and N are the peptide oxygen and nitrogen atoms of the indicated residues.
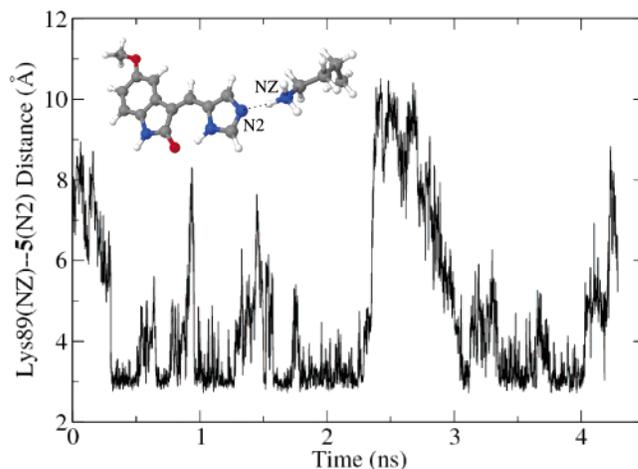

Figure 7. Evolution of the distance of the hydrogen bond formed between inhibitor **5** and residue K89 of CDK2.

and water molecules within 7.0 Å from the inhibitor, in the geometry obtained after full minimization within the force-field potential. An additional force-field minimization of this larger pocket model is performed keeping fixed the C_α atoms and all atoms of the smaller HB model. This is necessary to relax any possible steric clash between the added amino acids and water molecules and the elements of the HB model after the minimization at the DFT level.[44] After this minimization, the binding energy within the larger pocket model is calculated as the difference in the DFT total energy between this system, the isolated inhibitor, and the isolated pocket model, without relaxing the geometries of the individual components (ΔE_g(7 Å) values in Table 3).

We note that all values of ΔE_g(7 Å) are smaller than the corresponding values of ΔE_g(HB). This indicates that the direct hydrogen bond interactions between the inhibitors and the
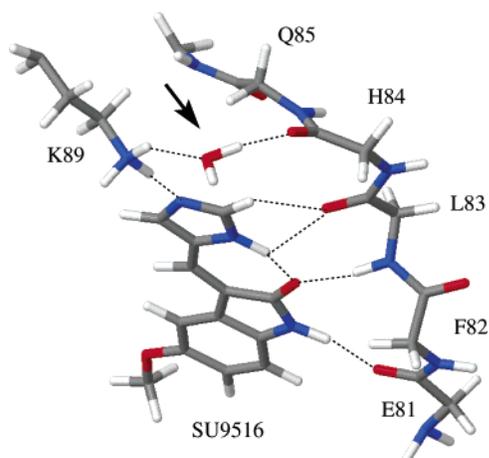
Figure 8. Water molecule (indicated with an arrow) bridging the NH_3^+ group of K89 with the peptide O atom of H84 near the SU9516 inhibitor. The calculated binding energy decreases by 5 kcal/mol when this single water molecule is included in the DFT total energy calculations (see text: Section 4.1).

Table 5. Ratios between the Binding Energies Calculated in the HB and 7 Å Models and Binding Energy Values (kcal/mol) of Systems Composed by the Inhibitors and Different Subsets of Amino Acids[a]

inhibitor	$\Delta E_g(7\ \text{Å})/\Delta E_g(\text{HB})$	ΔE_g^{tri}	$\Delta E_g^{\text{tri,86}}$	$\Delta E_g^{\text{tri,89}}$	$\Delta E_g^{\text{tri,86,89}}$
1	0.772	−16.2			
2	0.798	−17.3			
5	0.682	−17.6		−50.3	
4	0.638	−12.8	−23.5		−57.5
3	0.853	−18.3	−42.0	−47.0	−74.6

[a] As described in the text in Section 4.1.

protein are substantially screened by the presence of the surrounding amino acids and solvation water molecules. Interestingly, when all water molecules are stripped off from the 7 Å model, the computed binding energies (ΔE_g(7 Å dry) in Table 3) are slightly larger than the ΔE_g(HB) values, as found in Section 4.2 for the systems extracted from the crystal structures. Furthermore, in the case of inhibitor **5**, we computed the binding energies of systems composed of all amino acids of the HB model and all of the water molecules within 4.0, 6.0, and 7.0 Å from the inhibitor. The computed ΔE_g values (in kcal/mol) for the HB model and these three systems are −50.3, −50.0, −49.6, −49.1, indicating that the presence of water molecules alone does not contribute substantially to the observed screening of the interactions passing from the HB to the 7 Å systems. The combination of these two results leads to the hypothesis that the observed screening effect must arise from the simultaneous presence of both the surrounding amino acids and the solvation water molecules.

Indeed, the binding energy of inhibitor **5**, calculated by adding residues 10 and 86 as well as the backbone of residues 84 and 85 to the HB system, is −44.1 kcal/mol, whereas the binding energy value of the same system after adding a single water molecule, which bridges the NH_3^+ group of K89 and the peptide group between residues 84 and 85 (Figure 8), is decreased dramatically to −39.1 kcal/mol. This drop in binding energy is associated with a substantial change in the polarization of the hydrogen bonds upon the inclusion of further protein residues and the water molecule, as can be seen by analysis of the atomic charges that best fit the electrostatic potential computed from first principles. Namely, calculated atomic charges on the acceptor N atom of **5** changes from −0.23 in the HB model to 0.01 in the presence of the additional residues and to 0.10 after adding the water molecule, whereas the charge on the donor H atom of K89 changes from 0.19 to −0.02 to −0.26. This is indicative of profound changes in the local electrostatics and, thus, in hydrogen bond strength due to the rearrangements of electronic charge in the three systems considered. This result highlights the importance of calculating binding energies at a quantum level of accuracy and reveals water-mediated polarization effects[45] as a novel feature of the interaction between CDK2 and its inhibitors.

The values of the binding energy reported in Table 3 represent an upper limit to the actual binding energy of the system. In fact, because many of the hydrogen bonds are repeatedly broken and formed during the motion of the protein at room temperature, their contributions to the binding energy need to be weighted by the fraction of time for which they are present (Table 4). For example, for inhibitor **3**, a total of six hydrogen bonds are formed: the triplet of stable bonds with E81 and L83, a pair of intermittent bonds with D86, and an intermittent bond with K89. Let t^{tri}, $t^{\text{tri,86}}$, $t^{\text{tri,89}}$, and $t^{\text{tri,86,89}}$ be the fractions of time for which the bonds with the triplet only, the triplet and residue 86, the triplet and residue 89, and the triplet, residue 86 and residue 89 are present, respectively. These values can be extracted from an analysis of the variation of the hydrogen-bond distances and angles during each simulation. For the presence of a bond, we define a cutoff distance of 4.0 Å between donor and acceptor and a minimum angle of 90° between the donor, the hydrogen, and the acceptor. We stress the importance of including a condition on the hydrogen bond angle because hydrogen-bond strength is considerably reduced when this angle deviates from the ideal linear conformation.[28] In the case of inhibitor **3**, we obtained $t^{\text{tri,86}} = 0.13$, $t^{\text{tri,89}} = 0.18$, and $t^{\text{tri,86,89}} = 0.08$, and $t^{\text{tri}} = 0.61$.

Now we need to split the computed values of ΔE_g into the individual contributions corresponding to each binding mode. We define ΔE_g^{tri}, $\Delta E_g^{\text{tri,86}}$, $\Delta E_g^{\text{tri,89}}$, and $\Delta E_g^{\text{tri,86,89}}$ as the binding energies of the inhibitor to the pocket in the presence of the hydrogen-bond triplet only, the triplet and residue 86, the triplet and residue 89, and the triplet together with both residue 86 and 89, respectively. Each of these contributions are calculated separately via static total energy calculations of small models extracted from the geometry of the HB system previously obtained. At this point, we make the assumption that the ratios between the contributions of the individual bonding modes are the same in the HB and in the 7 Å pocket models. In other words, the individual contributions calculated taking into account only the minimal pocket model are scaled by the same factor ΔE_g(7 Å)/ΔE_g(HB) (0.85 in the case of inhibitor **3**). With this assumption, we can finally calculate the binding energy ΔE_g as follows:

$$\Delta E_g = \frac{\Delta E_g(7\text{Å})}{\Delta E_g(\text{HB})} \cdot (t^{\text{tri}}\Delta E_g^{\text{tri}} + t^{\text{tri,86}}\Delta E_g^{\text{tri,86}} + t^{\text{tri,89}}\Delta E_g^{\text{tri,89}} + t^{\text{tri,86,89}}\Delta E_g^{\text{tri,86,89}})$$

(6)

The computed values of binding energy of the individual binding modes of all five inhibitors are reported in Table 5.

Because explicit water molecules are included in our models, the values of ΔE_g for the five inhibitors calculated with the procedure described above already contain the enthalpic contributions to the solvation of the *PL* and the *P* systems. The remaining contributions to the solvation free energy, that is, the surface area contributions of the *PL* and *P* models, as well as the solvation free energy of the isolated ligands, are calculated

Table 6. Computed Differences of Binding Free Energy (kcal/mol) and Relative Potencies Compared with Experimental Values (in Italics) Obtained from the Measured Inhibition Constants

inhibitor	$\Delta\Delta G$	$\Delta\Delta G$ exp.	$e^{\Delta\Delta G/k_B T}$	K_i/K_i^{NU2058} exp.
1	0.0	*0.0*	1.0	*1.0*
2	−0.6	*−1.4*	3.8E − 1	*(1.1 ± 3.2)E − 1*
5	−4.9	*−3.7*	3.5E − 4	*(2.6 ± 0.8)E − 3*
4	−3.6	*−3.8*	2.9E − 3	*(2.0 ± 0.8)E − 3*
3	−5.3	*−4.7*	1.8E − 4	*(5.0 ± 1.3)E − 4*

within the GBSA method to obtain the solvation free energy differences ΔG_{solv}^{wat}, where the superscript indicates that these values refer to the models including the explicit water solvent (reported in Table 3). The obtained enthalpic and solvation contributions can now be used to compute a final set of free energy differences $\Delta\Delta G$ (Table 6). The inferred K_i ratios are now found to agree with the experimentally measured values within 1 order of magnitude for all of the inhibitors considered, including the delicate case of inhibitor **5**.

5. Conclusions

Binding Modes and Binding Energies. The main conclusion that can be drawn from our investigations is that the potency of inhibition of small molecules that bind to the ATP pocket of CDK2 is largely dependent on the dynamical nature of the protein/inhibitor interactions. In particular, the motion of both the inhibitor and particular residues within the binding pocket is responsible for the intermittent formation and rupture of hydrogen bonds. Such effects are not easy to identify by using only static methods such as X-ray crystallography for structure investigation. Moreover, even if the orientation of the inhibitor in the binding pocket remains the same, differences in the conformation of the backbone of the protein can lead to substantial differences in the magnitude of the calculated hydrogen-bond strengths (cf. Tables 1 and 3). The dynamical nature of the hydrogen-bond interactions is likely to be the main reason for the quantitative disagreement between the experimental binding energy values and the values obtained using scoring functions after rigid docking simulations (see Table 2 and ref 20).[20] Notably, the binding free energy differences calculated with this method are underestimated with respect to the experimental values, which are consistent with the neglect of additional intermittent interactions within the protein active site. However, the scoring functions employed are carefully parametrized to account for hydrophobic and solvation effects in a simple but robust manner. Indeed, they are capable of giving a better qualitative estimate of the rank of potencies between the considered inhibitors than simplistic DFT binding energy calculations after geometry optimization of protein/inhibitor complexes extracted from the crystal structures, in the absence of explicit solvent molecules (Table 1).

Our results stress the importance of including a large number of amino acids in the models for the ATP binding pocket (Figure 2) because long-range electrostatic effects contribute substantially to the computed binding energies. In general, when no explicit water molecules are included in the calculations, the binding energy increases with the increasing number of amino acids included in the calculations (Figure 2 and Tables 1 and 3), converging within 0.15 kcal/mol when all amino acids within 7 Å from the inhibitor are considered. A major change to the computed binding energy is due to the explicit presence of water molecules within the binding pocket. In this case, hydrogen-bond interactions are substantially screened because of water-mediated hydrogen-bond depolarization effects so that the binding energy is reduced up to ~35% of its original value

(Tables 3 and 5). Because relatively few water molecules can be identified by crystallographic methods, this stresses the importance of using MD methods to obtain a correct solvated model of the protein/ligand complex. Moreover, given the complex nature of the screening effects observed, this indicates that calculations of binding energies necessarily require the use of quantum techniques such as DFT.

The main limitation of MD methods is that the accessible simulation time is very short. Therefore, the fractions of time associated with each individual binding modes that enter into the calculations of the relative free energy of binding should be considered as indicative values.[46] Given this limitation, the approximations made on the solvation and entropic terms contributing to the binding energies appear to be less severe. Although gas-phase entropic terms (vibrational, internal rotational, and conformational) necessarily need to be explicitly considered in order to compute absolute values of binding free energies, this is not necessary when comparing relative values of binding free energies among different inhibitors of the same protein.[34] As mentioned in Section 3, this approximation is justified by the chemical similarities of the systems considered. Also, the underestimation of direct van der Waals interactions within the DFT techniques employed appears not to be crucial in the present case. Dispersion forces can represent large fractions of absolute binding energies in biological systems[47−49] and are indeed of the order of 45 kcal/mol in our minimized CDK2/ligand complexes (i.e., at the maximum van der Waals contact between the hydrophobic and aromatic residues in the binding pocket and the ligands). However, when the differences of van der Waals energies are computed as averages of relatively long dynamic runs at 300 K, they are found to be the same for all of the inhibitors considered in this work (within the error bar of a few kcal/mol intrinsic to the MM method), and thus, they do not contribute substantially to the calculated $\Delta\Delta G$ values.

As far as the calculation of solvation free energies is concerned, this is in general a very difficult task, requiring, in principle, accurate formalisms and an explicit treatment of solvent molecules.[41] Here, we have used the widely used continuum GBSA model[37] to compute the solvation free energies (apart from the final set of calculations, where explicit water molecules are included in the *PL* and *P* models; see Section 4.1), expecting errors in the differences between the solvation free energies calculated with the same method to be negligible compared with all other approximations.

Specificity of Inhibition. Considering the approximations inherent to the methods employed and the limited statistics accessible to atomistic MD simulations, the agreement between calculated and measured relative potencies is remarkably good. It has to be noted, however, that direct calculations of inhibition constants is still out of the reach of this and analogous computational schemes. The power of our method lies in the ability to predict and quantify all interactions present between a given inhibitor and, in this case, CDK2. In particular, in the case of inhibitor **5**, an analysis of the computed hydrogen bond strengths immediately revealed that a strong interaction was absent from the previously considered structural model. With the help of dynamical simulations, this interaction has been identified in the form of an intermittent hydrogen bond with Lys89, which was not visible in the crystal structure.[32] The same residue has also been observed to form similar intermittent interactions with inhibitors **3** and **4**. The presence of such interactions is particularly important because Lys89 is a residue specific to CDK2. In CDK4, for instance, a much shorter

threonine residue is present in the same position (Thr89), which is not expected to be able to form hydrogen bonds with small inhibitors, unless they are specifically modified for this purpose.[50] Therefore, there is an increase in binding energy between the inhibitors and the protein that is specific to CDK2, and it is not expected to be present in the case of CDK4. This is consistent with the observed, yet so far unexplained, selectivity patterns of inhibitors **3**[30] and **5**.[32]

Recent inhibitor profiles have targeted the Lys 89 residue for this very reason, placing a highly hydrophilic group at a favorable position.[51] However, we note that the gain in binding enthalpy from adding hydrophilic groups targeting Lys89 is counterbalanced by the penalty of an increased solvation energy. Inhibitor **5** is an interesting case, having a considerably lower solvation energy than other pyridine and pyrimidine-based inhibitors yet still being capable of interaction with Lys89. In principle, it is also conceivable that Lys89 could form an intermittent hydrogen bond with the N1 atom of inhibitor **1**, which could help explain why also in the case of this inhibitor a moderate selectivity against CDK4 has been observed.[30] However, this selectivity could also arise from factors not directly connected to local interactions within the ATP binding cleft or from more subtle effects, for example, differences between the hydrophobicity patterns of the two proteins.[52]

Summary. The relative differences of binding free energies $\Delta\Delta G$ between five inhibitors of CDK2 have been calculated by using a combination of linear-scaling DFT techniques and classical MD simulations. The $\Delta\Delta G$ values obtained after direct structural optimization of protein/inhibitor complexes extracted from the available crystal structures are found not to correlate well with the measured potencies of inhibition. In comparison, extensive docking simulations followed by free energy calculations using scoring function methods provided a better qualitative estimation of the rank of potencies among the inhibitors, although a quantitative agreement could not be achieved. Molecular dynamics simulations using classical force-field techniques revealed that some of the hydrogen bonds formed between the inhibitors and the ATP binding cleft of CDK2 are in fact intermittent, being repeatedly broken and formed at room temperature. These dynamical effects are taken into account in the O(N) DFT binding energy calculations by weighting each interaction by the percentage of the time for which it is actually present. Moreover, a combined effect due to the presence of solvation water molecules and neighboring amino acids leads to a substantial screening of the protein/ligand hydrogen-bond interactions, which can be accurately quantified using DFT techniques. Despite the approximate nature of the methods and the limited statistics available, $\Delta\Delta G$ values calculated taking these effects into account show a much better agreement with the experimentally measured potencies of these inhibitors. Finally, the simulations reveal the presence of intermittent hydrogen-bond interactions between the inhibitors and the Lys89 residue, which have not been identified in the crystallographic studies performed to date. Because Lys89 is a CDK2 specific residue (in CDK4 a much shorter Thr residue is present in position 89), the existence of such an interaction offers an explanation for the observed, yet so far not understood specificity of some of the inhibitors for CDK2 against CDK4.

Taken together, our findings reveal new factors underlying the relationship between the structure and activity of CDK inhibitors. They demonstrate that polarization effects arising from the specific hydration patterns in the ATP binding pocket of CDK2 contribute to the differential potency of its ihibitors. Such polarization effects may be a general feature in other kinase/inhibitor complexes, which must be taken into account for rational drug design. Our results also demonstrate a contribution of intermittent hydrogen-bonding interactions, not previously recognized in static crystallographic studies, to the potency of CDK2 inhibitors and their relative selectivity over CDK4. Finally, this work provides a generally applicable computational method that can be adapted to the investigation of other kinase/inhibitor structures and that could be a useful tool in rational drug design.

6. Computational Details

Density Functional Theory Calculations. Quantum mechanical calculations of total energies were performed both with traditional plane-wave (PW),[53] and with novel linear-scaling (O(N))[13,14] implementations of density functional theory (DFT), using the PBE gradient corrected exchange-correlation functional.[54] The CASTEP[55] code was employed for the PW calculations of smaller systems, using ultrasoft pseudopotentials[56] and expanding the wave functions at the Γ-point of the Brillouin zone up to a kinetic energy cutoff of 450 eV. The modeled molecules were placed in cubic simulation cells with edge lengths of 25 Å, which was sufficient to avoid any spurious interactions between the system and its periodicallly repeated images in all cases considered. The computing cost of the PW approach increases with the cube of the number of atoms and becomes prohibitively expensive for systems with more than a few hundred atoms. In contrast, O(N) methods have recently been developed where the increase in computing cost is linear with the number of atoms and allow us to extend the DFT calculations to much larger systems. O(N) DFT calculations were performed using both the widely used SIESTA code[13] and the newly developed ONETEP code.[14] In both cases, the interactions between electrons and nuclei were described by norm-conserving pseudopotentials.[57,58] SIESTA calculations were performed by placing the systems in cubic supercells with edge lengths of up to 65 Å, avoiding any spurious interactions between the simulated system and its periodically repeated images. The employed basis set consists of a DZP basis of localized atomic orbitals, which were variationally optimized to accurately describe hydrogen bonds in biomolecular systems.[59] Because of the atomic basis set, calculations of binding energies needed to be corrected from the so-called basis set superposition error (BSSE) using the counterpoise method of Boys and Bernardi.[60] In the ONETEP code, the DFT equations are solved using a minimal set of strictly localized functions that are optimized in situ in a basis set, which is formally equivalent to a plane-wave expansion.[61] Therefore, the same levels of accuracy and rapidity of convergence of traditional PW codes are achieved,[62] and there is no superposition error to be corrected during the calculation of binding energies.[63] The ONETEP calculations were performed using a kinetic energy cutoff of 600 eV and placing the systems in a cubic simulation supercell with an edge length of about 80 Å. In all calculations, charge neutrality of the systems was imposed by protonation or deprotonation of residue groups of the protein models that were not involved in any direct interaction with the inhibitors.

We note that our calculations provide an important comparison between a well established but computationally costly DFT implementation and two recent linear-scaling methods for the quantum mechanical treatment of very large systems (up to thousands of atoms) based on two very different mathematical formalisms. Our results show that the differences in the computed binding energies between the three different codes are less than 1.5 kcal/mol for systems including about 1000 atoms (see Section 4.1), that is, within the error bar associated with the other approximations employed (supercell method, pseudopotentials, exchange−correlation functionals). For instance, the binding energy values of one of the inhibitors calculated with the CASTEP, SIESTA, and ONETEP codes are −16.2, −16.1, and −16.4 kcal/mol, respectively (other comparisons between the three codes have been presented in Section 3.1). Hence, the results obtained with any of the codes are entirely equivalent at the standard accuracy level reached by DFT techniques. Nevertheless, to ensure full

consistency of the presented results, the reported values of the differences in binding energies are the ones calculated with the ONETEP code. The large volume calculations necessary at the exploratory and setup stages of the work (concerning size convergence, tests on charged states, choice of protonation sites, etc.) were performed using the SIESTA code.

Classical Force-Field Simulations. The AMBER 7.0[64] suite of programs was used for all of the classical force-field simulations. Point charges on the atoms of the inhibitors were calculated as a best fit to the electrostatic potentials computed within the DFT formalism in a 1.0 Å thick spherical region outside the van der Waals radii of the atoms. All other parameters for the inhibitors were determined using the Antechamber facility making use of the gaff force field provided. Solvation free energies were calculated within the standard pairwise generalized Born model, taking into account entropic contributions using the LCPO method to calculate the molecular surface area. The classical simulations with the explicit solvent were carried out using the ff99[65] force field based on the parameters of Cornell et al.[66] and the TIP3P water solvent. The models for the classical simulations were prepared by reading into the Leap module of AMBER the crystal structures of the protein/inhibitor complexes after having stripped all crystallographic water molecules and counterions present. Each model was then protonated, solvated in a tetragonal box with edge lengths ensuring a mininum 20 Å distance from the system and its periodically repeated images, and neutralized by adding chloride counterions. The dynamical simulations were performed following a standard three-step equilibration procedure. (i) The system was first fully minimized to release any unwanted steric clashes, and then the solvent was equilibrated in a 0.015 ns run, slowly increasing the temperature to 300 K at the constant pressure of 1 atm and keeping the solute frozen. (ii) A series of 5 minimization runs was then performed, in which the harmonic constraints on the solute were progressively released and eventually eliminated. (iii) Subsequently, the system was equilibrated in a 0.05 ns run to reach conditions of constant temperature (300 K) and pressure (1 atm). After equilibration, we carried out the production run for a time that varied from 0.5 to 4.0 ns, depending on the system considered. Only the production runs were taken into account in the analyses of the hydrogen-bond patterns (Section 5.1). We used an Ewald cutoff sphere of 12.0 Å and a time step of 2 fs for the integration of the equations of motion, employing the SHAKE algorithm to keep fixed all bonds involving H atoms. A harmonic constraint of 5.0 kcal/mol/Å^2 was kept on the amino acids immediately before and after any protein sequences unavailable in the crystal structure (typically in the region from residue 36 to residue 43 of the CDK2/inhibitor complexes).

Van der Waals Dispersion Interactions. The standard DFT technique employed systematically underestimates direct interactions due to dispersion forces. We thus carefully checked that the dispersion interactions between the protein and the ligands in the binding pocket are roughly the same for all inhibitors considered so that they do not contribute significantly to the computed free energy differences $\Delta\Delta G$. An estimate of the maximum dispersion interactions was obtained from the values of the van der Waals energies of the minimized *PL* complexes and the isolated *P* and *L* systems in the same atomic configurations as those in the *PL* complex, computed with the classical Amber force field in an implicit solvent. For the NU2058, NU6027, SU9516, 9d-NU6027, and NU6102 inibitors, we compute net van der Waals interaction energies of 41.4, 42.0, 40.1, 54.2, and 55.8 kcal/mol, respectively. These large values are consistent with other quantum chemical and empirical calculations of dispersion energies in aromatic systems.[47,48] We note, however, that the dispersion interactions are considerably affected by the Brownian motion of the ligand and the residues within the binding pocket in water solution at 300 K. We therefore computed the average of the van der Waals binding energies over the configurations of the *PL* complexes during the dynamics at 300 K in an explicit solvent for the most weakly and most strongly bound inhibitors, that is, SU9516 and NU6102. The differences $\Delta E_{vdw}(PL) - \Delta E_{vdw}(P) - \Delta E_{vdw}(L)$ between those

inhibitors were found to be less that ~4 kcal/mol, which is within the error bar of the method. We therefore conclude that direct dispersion interactions do not contribute in a substantial way to the different potencies among the inhibitors; therefore, the differential values of binding energy can be safely computed within the DFT scheme.

Scoring Functions Calculations. Docking simulations were carried out using a recently introduced quantum stochastic tunneling docking method.[23,67] This is a very efficient hybrid optimization method that allows for the quantum and stochastic tunneling through high energy barriers and the nonlocal exploration of the potential energy surface. Two hundred simulations were performed for each ligand−protein complex under the same conditions as those described in ref 23, using the PLP scoring function[68] to represent the potential energy surface. A large number of binding modes were generated, which were then re-scored using the Böhm[69] and ScreenScore[70] scoring functions. The free energy of binding (for each scoring function) of each ligand−protein complex reported corresponds to the best energy found among all generated binding modes within 1.0 Å RMSD of the binding mode visible in the crystal structure. This precaution ensured that the best estimate of the free energy binding corresponded to the experimentally observed binding mode.

References

(1) Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. The protein kinase complement of the human genome. *Science* **2002**, *298*, 1912−1934.
(2) Morgan, D. O. Principles of CDK regulation. *Nature* **1995**, *374*, 131−134.
(3) Loog, M.; Morgan, D. O. Cyclin specificity in the phosphorylation of cyclin-dependent kinase substrates. *Nature* **2005**, *434*, 104−108.
(4) Brown, N. R.; Noble, M. E. M.; Lawrie, A. M.; Morris, M. C.; Tunnah, P.; Divita, G.; Johnson, L. N.; Endicott, J. A. Effects of phosphorylation of threonine 160 on cyclin-dependent kinase 2 structure and activity. *J. Biol. Chem.* **1999**, *274*, 8746−8756.
(5) Russo, A. A.; Jeffrey, P. D.; Pavletich, N. P. Structural basis of cyclin-dependent kinase activation by phosphorylation. *Nat. Struct. Biol.* **1996**, *3*, 696−700.
(6) Vermeulen, K.; Van Bockstaele, D. R.; Berneman, Z. N. The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer. *Cell Proliferation* **2003**, *36*, 131−149.
(7) Huwe, A.; Mazitschek, R.; Giannis, A. Small molecules as inhibitors of cyclin-dependent kinases. *Angew. Chem., Int. Ed.* **2003**, *42*, 2122−2138.
(8) Wadler, S. Perspectives for cancer therapies with CDK2 inhibitors. *Drug Resist. Updates* **2001** *4*, 347−367.
(9) Noble, M. E. M.; Endicott, J. A.; Johnson, L. N. Protein kinase inhibitors: insights into drug design from structure. *Science* **2004**, *303*, 1800−1805.
(10) Bain, J.; McLauchlan, H.; Elliott, M.; Cohen, P. The specificities of protein kinase inhibitors: an update. *Biochem. J.* **2003** *371*, 199−204.
(11) Park, H.; Yeom, M. S.; Lee, S. Loop flexibility and solvent dynamics as determinants for the selective inhibition of cyclin-dependent kinase 4: comparative molecular dynamics simulation studies of CDK2 and CDK4. *ChemBioChem* **2004**, *5*, 1662−1672.
(12) Ikuta, M.; Kamata, K.; Fukasawa, K.; Honma, T.; Machida, T.; Hirai, H.; Suzuki-Takahashi, I.; Hayama, T.; Nishimura, S. Crystallographic approach to identification of cyclin-dependent kinase 4 (CDK4)-specific inhibitors by using CDK4 mimic CDK2 protein. *J. Biol. Chem.* **2001** *276*, 27548−27554.
(13) Soler, J. M.; Artacho, E.; Gale, J. D.; García, A.; Junquera, J.; Ordejón, P.; Sánchez-Portal, D. The SIESTA method for *ab initio* order-n materials simulation. *J. Phys.: Condens. Matter* **2002** *14*, 2745−2779.

(14) Skylaris, C.-K.; Haynes, P. D.; Mostofi, A. A.; Payne, M. C. Introducing ONETEP: linear-scaling density functional simulations on parallel computers. *J. Chem. Phys.* **2005** *122*, 084119.

(15) Hardcastle, I. R.; Arris, C. E.; Bentley, J.; Boyle, F. T.; Chen, Y. H.; Curtin, N. J.; Endicott, J. A.; Gibson, A. E.; Golding, B. T.; Griffin, R. J.; Jewsbury, P.; Menyerol, J.; Mesguiche, V.; Newell, D. R.; Noble, M. E. M.; Pratt, D. J.; Wang, L. Z.; Whitfield, H. J. N^2-substituted O^6-cyclohexylmethylguanine derivatives: potent inhibitors of cyclin-dependent kinases 1 and 2. *J. Med. Chem.* **2004**, *47*, 3710−3722.

(16) Gibson, A. E.; Arris, C. E.; Bentley, J.; Boyle, F. T.; Curtin, N. J.; Davies, T. G.; Endicott, J. A.; Golding, B. T.; Grant, S.; Griffin, R. J.; Jewsbury, P.; Johnson, L. N.; Mesguiche, V.; Newell, D. R.; Noble, M. E. M.; Tucker, J. A.; Whitfield, H. J. Probing the ATP ribose-binding domain of cyclin-dependent kinases 1 and 2 with O^6-substituted guanine derivatives. *J. Med. Chem.* **2002** *45*, 3381−3393.

(17) Sayle, K. L.; Bentley, J.; Boyle, F. T.; Calvert, A. H.; Cheng, Y. Z.; Curtin, N. J.; Endicott, J. A.; Golding, B. T.; Hardcastle, I. R.; Jewsbury, P.; Mesguiche, V.; Newell, D. R.; Noble, M. E. M.; Parsons, R. J.; Pratt, D. J.; Wang, L. Z.; Griffin, R. J. Structure-based design of 2-arylamino-4-cyclohexylmethyl-5-nitroso- 6-aminopyrimidine inhibitors of cyclin-dependent kinases 1 and 2. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3079−3082.

(18) Legraverend, M.; Tunnah, P.; Noble, M.; Ducrot, P.; Ludwig, O.; Grierson, D. S.; Leost, M.; Meijer, L.; Endicott, J. Cyclin-dependent kinase inhibition by new C-2 alkynylated purine derivatives and molecular structure of a CDK2−inhibitor complex. *J. Med. Chem.* **2000** *43*, 1282−1292.

(19) Gray, N. S.; Wodicka, L.; Thunnissen, A.-M. W. H.; Norman, T. C.; Kwon, S. J.; Espinoza, F. H.; Morgan, D. O.; Barnes, G.; LeClerc, S.; Meijer, L.; Kim, S. H.; Lockhart, D. J.; Schultz, P. G. Exploiting chemical libraries, structure, and genomics in the search for kinase inhibitors. *Science* **1998** *281*, 533−538.

(20) Otyepka, M.; Krystof, V.; Havlicek, L.; Siglerova, V.; Strnad M.; Koca, J. Docking-based development of purine-like inhibitors of cyclin-dependent kinase 2. *J. Med. Chem.* **2000** *43*, 2506−2513.

(21) Gabb, H. A.; Jackson, R. M.; Sternberg, M. J. E. Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J. Mol. Biol.* **1997** *272*, 106−120.

(22) Ducrot, P.; Legraverend, M.; Grierson, D. S. 3D-QSAR CoMFA on cyclin-dependent kinase inhibitors. *J. Med. Chem.* **2000** *43*, 4098−4108.

(23) Mancera, R. L.; Källblad, P.; Todorov, N. P. Ligand-protein docking using a quantum stochastic tunneling optimization method. *J. Comput. Chem.* **2004**, *25*, 858−864.

(24) Kriz, Z.; Otyepka, M.; Bartova, I.; Koca J. Analysis of CDK2 active-site hydration: a method to design new inhibitors. *Proteins* **2004** *55*, 258−274.

(25) Bartova, I.; Otyepka, M.; Kriz, Z.; Koca, J. Activation and inhibition of cyclin-dependent kinase-2 by phosphorylation; a molecular dynamics study reveals the functional importance of the glycine-rich loop. *Protein Sci.* **2004** *6*, 1449−1457.

(26) Otyepka, M.; Kriz, Z.; Koca, J. Dynamics and binding modes of free CDK2 and its two complexes with inhibitors studied by computer simulations. *J. Biomol. Struct. Dyn.* **2002** *20*, 141−154.

(27) Cavalli, A.; Dezi, C.; Folkers, G.; Scapozza, L.; Recanatini, M. Three-dimensional model of the cyclin-dependent kinase 1 (CDK1): ab initio active site parameters for molecular dynamics studies of CDKs. *Proteins* **2001**, *45*, 478−485.

(28) Ireta, J.; Neugebauer, J.; Scheffler M. On the accuracy of DFT for describing hydrogen bonds: dependence on the bond directionality. *J. Phys. Chem. A* **2004**, *108*, 5692−5698.

(29) Arris, C. E.; Boyle, F. T.; Calvert, A. H.; Curtin, N. J.; Endicott, J. A.; Garman, E. F.; Gibson, A. E.; Golding, B. T.; Grant, S.; Griffin, R. J.; Jewsbury, P.; Johnson, L. N.; Lawrie, A. M.; Newell, D. R.; Noble, M. E. M.; Sausville, E. A.; Schultz, R.; Yu, W. Identification of novel purine and pyrimidine cyclin-dependent kinase inhibitors with distinct molecular interactions and tumor cell growth inhibition profiles. *J. Med. Chem.* **2000**, *43*, 2797−2804.

(30) Davies, T. G.; Bentley, J.; Arris, C. E.; Boyle, F. T.; Curtin, N. J.; Endicott, J. A.; Gibson, A. E.; Golding, B. T.; Griffin, R. J.; Hardcastle, I. R.; Jewsbury, P.; Johnson, L. N.; Mesguiche, V.; Newell, D. R.; Noble, M. E. M.; Tucker, J. A.; Wang, L.; Whitfield, H. J. Structure-based design of pure purine-based cyclin-dependent kinase inhibitor. *Nat. Struct. Biol.* **2002**, *9*, 745−749.

(31) Lane, M. E.; Yu, B.; Rice, A.; Lipson, K. E.; Liang, C.; Sun, L.; Tang, C.; McMahon, G.; Pestell, R. G.; Wadler, S. A novel CDK2-selective inhibitor, SU9516, induces apoptosis in colon carcinoma cells. *Cancer Res.* **2001**, *61*, 6170−6177.

(32) Moshinsky, D. J.; Bellamacina, C. R.; Boisvert, D. C.; Huang, P.; Hui, T.; Jancarik, J.; Kim, S.-H.; Rice, A. G. SU9516: biochemical analysis of CDK inhibition and crystal structure in complex with CDK2. *Biochem. Biophys. Res. Commun.* **2003**, *310*, 1026−1031.

(33) Cheng, Y.-C.; Prusoff, W. H. Relationship between the inhibition constant K_i and the concentration of inhibitor which causes 50% inhibition (IC_{50}) of an enzymatic reaction. *Biochem. Parmacol.* **1973**, *22*, 3099−3108.

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

(35) Cramer C. J.; Truhlar, D. J. Implicit solvation models: equilibria, structure, spectra, and dynamics. *Chem. Rev.* **1999**, *99*, 2161−2200.

(36) Jayaram, B.; Sprous, D.; Beveridge, D. L. Solvation free energy of biomacromolecules: parameters for a modified generalized Born model consistent with the Amber force field. *J. Phys. Chem. B* **1998**, *102*, 9571−9576.

(37) Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phy. Chem.* **1994**, *98*, 1978−1988.

(38) Fenández-Serra, M. V.; Artacho, E. Network equilibration and first principles liquid water. *J. Chem. Phys.* **2004** *121*, 11136−11144.

(39) Schwegler, E.; Grossman, J. C.; Gygi, F.; Galli, G. Towards an assessment of the accuracy of density functional theory for first principles simulations of water. II. *J. Chem. Phys.* **2004** *121*, 5400−5409. Grossman, J. C.; Schwegler, E.; Draeger, E. W.; Gygi, F.; Galli, G. Towards an assessment of the accuracy of density functional theory for first principles simulations of water. I. *J. Chem. Phys.* **2004** *120*, 300−311.

(40) VandeVondele, J.; Mohamed, F.; Krack, M.; Hutter, J.; Sprik, M.; Parrinello, M. The influence of temperature and density functional models in ab initio molecular dynamics simulation of liquid water. *J. Chem. Phys.* **2005**, *122*, 014515.

(41) Mancera, R. L. A new explicit hydration penalty score for ligand-protein interactions. *Chem. Phys. Lett.* **2004**, *399*, 271−275.

(42) Lawrie, A. M.; Noble, M. E. M.; Tunnah, P.; Brown, N. R.; Johnson, L. N.; Endicott, J. A. Protein kinase inhibition by staurosporine revealed in details of the molecular interaction with CDK2. *Nat. Struct. Biol.* **1997**, *4*, 796−801.

(43) Schulze-Gahmen, U.; De Bondt H. L.; Kim, S.-H. High-resolution crystal structures of human cyclin-dependent kinase 2 with and without ATP: bound waters and natural ligand as guides for inhibitor design. *J. Med. Chem.* **1996**, *39*, 4540−4546.

(44) In particular, a steric clash between all inhibitors and the Phe 82 residue is evident after including the small model minimized at the DFT level in the larger pocket model obtained after force-field minimization from a MD snapshot. This is due to the much shorter van der Waals distance between hydrophobic residues obtained with classical potentials with respect to the corresponding distances obtained at the GGA-DFT level. For this reason, because the contributions of direct van der Waals interactions with the Phe 82 residue are expected to be about the same for all inhibitors, this residue is substituted with an alanine residue in all binding energy calculations presented.

(45) (a) Greatbanks, S. P.; Gready, J. E.; Limaye, A. C.; Rendell, A. P. Comparison of enzyme polarization of ligands and charge-transfer effects for dihydrofolate reductase using point-charge embedded ab initio quantum mechanical and linear-scaling semiempirical quantum mechanical methods. *J. Comput. Chem.* **2000**, *21*, 788−811. (b) Muzet, N.; Guillot, B.; Jelsch, C.; Howard, E.; Lecomte, C. Electrostatic complementarity in an aldose reductase complex from ultra-high-resolution crystallography and first-principles calculations. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 8742−8747. (c) Sulpizi, M.; Laio, A.; VandeVondele, J.; Cattaneo, A.; Rothlisberger, U.; Carloni, P. Reaction mechanism of caspases: insights from QM/MM Car-Parrinello simulations. *Proteins* **2003**, *52*, 212−224.

(46) We note that an accurate determination of the time fractions of different binding modes would require the calculation of their associated Boltzmann factors, which is computationally feasible only at the classical level. However, the accuracy of this approach relies on the precision of the calculated energy values of different binding configurations. These, in turn, are likely to be affected by subtle electronic effects, which are difficult to describe using a fixed-parameter force-field approach. Therefore, our choice here is to use the force field to obtain an estimate of the binding mode time fractions directly from MD simulations, while all energy values are computed at the DFT level.

(47) Daabkowska, I.; Jurecka, P.; Hobza, P. On geometries of stacked and H-bonded nucleic acid base pairs determined at various DFT, MP2, and CCSD(T) levels up to the CCSD(T)/complete basis set limit level. *J. Chem. Phys.* **2005**, *122*, 204322.

(48) Elstner, M.; Hobza, P.; Frauenheim T.; Suhai, S.; Kaxiras, E. Hydrogen-bonding and stacking interactions of nucleic acid base pairs: A density-functional-theory based treatment. *J. Chem. Phys.* **2001**, *114*, 5149−5155.

(49) Barratt, E.; Bingham, R. J.; Warner, D. J.; Laughton, C. A.; Phillips, S. E. V.; Homans, S: W. van der Waals interactions dominate ligand-protein association in a protein binding site occluded from solvent water. *J. Am. Chem. Soc.* **2005**, *127*, 11827−11834.

(50) Honma, T.; Yoshizumi, T.; Hashimoto, N.; Hayashi, K.; Kawanishi, N.; Fukasawa, K.; Takaki, T.; Ikeura, C.; Ikuta, M.; Suzuki-Takahashi, I.; Hayama, T.; Nishimura, S.; Morishima, H. A novel approach for the development of selective CDK4 inhibitors: library design based on locations of CDK4 specific amino acid residues. *J. Med. Chem.* **2001**, *44*, 4628−4640.

(51) Hamdouchi, C.; Zhong, B.; Mendoza, J.; Collins, E.; Jaramillo, C.; De Diego, J. E.; Robertson, D.; Spencer, C. D.; Anderson, B. D.; Watkins, S. A.; Zhang, F.; Brooks, H. B. Structure-based design of a new class of highly selective aminoimidazo[1,2-α]pyridine-based inhibitors of cyclin dependent kinases. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 1943−1947.

(52) Kelly, M. D.; Mancera, R. L. Comparative analysis of the surface interaction properties of the binding sites of CDK2, CDK4 and ERK2. *ChemMedChem* **2006**, *1*, 366−375.

(53) Payne, M. C.; Teter, M. P.; Allan, D. C.; Arias, T. A.; Joannopoulos, J. D. Iterative minimisation techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients. *Rev. Mod. Phys.* **1992**, *64*, 1045−1097.

(54) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.

(55) Segall M. D.; Lindan, P. J. D.; Probert, M. J.; Pickard, C. J.; Hasnip, P. J.; Clark, S. J.; Payne, M. C. First-principles simulation: ideas, illustrations and the CASTEP code. *J. Phys.: Condens. Matter* **2002**, *14*, 2717−2744.

(56) Vanderbilt, D. Soft self-consistent pseudopotential in a generalized eigenvalue formalism. *Phys. Rev. B* **1990**, *41*, 7892−7895.

(57) Hamann, D. R.; Schlüter, M.; Chiang, C. Norm-conserving pseudo-potentials. *Phys. Rev. Lett.* **1979**, *43*, 1494−1497.

(58) Troullier, N.; Martins J. L. Efficient pseudopotentials for plane-wave calculations. *Phys. Rev. B* **1991**, *43*, 1993−2006.

(59) Anglada, E.; Soler, J. M.; Junquera, J.; Artacho, E. Systematic generation of finite-range atomic basis set for linear-scaling calculations. *Phys. Rev. B* **2002**, *66*, 205101.

(60) Boys, S. F.; Bernardi, F. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **1970**, *19*, 553−566.

(61) Skylaris, C.-K.; Mostofi, A. A.; Haynes, P. D.; Diéguez, O.; Payne, M. C. Nonorthogonal generalised Wannier function pseudopotential plane-wave method. *Phys. Rev. B* **2002**, *66*, 035119.

(62) Mostofi, A. A.; Haynes, P. D.; Skylaris, C.-K.; Payne, M. C. Preconditioned iterative minimisation for linear-scaling electronic structure calculations. *J. Chem. Phys.* **2003**, *119*, 8842−8848.

(63) Skylaris, C.-K.; Haynes, P. D.; Mostofi, A. A.; Payne, M. C. Using ONETEP for accurate and efficient O(N) density functional calculations. *J. Phys.: Condens. Matter* **2005**, *17*, 5757−5769.

(64) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E.; Wang, J.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *AMBER 7;* University of California, San Francisco, CA, 2002.

(65) Wang, J. M.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, *21*, 1049−1074.

(66) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmayer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force-field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(67) Todorov, N. P.; Mancera, R. L.; Monthoux, P. H. A new quantum stochastic tunnelling optimisation method for protein−ligand docking. *Chem. Phys. Lett.* **2003**, *369*, 257−263.

(68) Gehlhaar, D. K.; Verkhivker, G. M.; Rejto, P. A.; Sherman, C. J.; Fogel, D. B.; Fogel, L. J.; Freer, S. T. Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. *Chem. Biol.* **1995**, *2*, 317−324.

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

(70) Stahl, M.; Rarey, M. Detailed analysis of scoring functions for virtual screening. *J. Med. Chem.* **2001**, *44*, 1035−1042.