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

Hwangseo Park,*[a] Min Sun Yeom,[b] and Sangyoub Lee*[a]

The design and discovery of selective cyclin-dependent kinase 4 (CDK4) inhibitors have been actively pursued in order to develop therapeutic cancer treatments. By means of a consecutive computational protocol involving homology modeling, docking experiments, and molecular dynamics simulations, we examine the characteristic structural and dynamic properties that distinguish CDK4 from CDK2 in its complexation with selective inhibitors. The results for all three CDK4-selective inhibitors under investigation show that the large-amplitude motion of a disordered loop of CDK4 is damped out in the presence of the inhibitors whereas their binding in the CDK2 active site has little effect on the loop flexibility. It is also found that the binding preference of CDK4-selective inhibitors for CDK4 over CDK2 stems from the reduced solvent accessibility in the active site of the former due to the formation of a stable hydrogen-bond triad by the Asp99, Arg101, and Thr102 side chains at the top of the active-site gorge. Besides the differences in loop flexibility and solvent accessibility, the dynamic stabilities of the hydrogen bonds between the inhibitors and the side chain of the lysine residue at the bottom of the active site also correlate well with the relative binding affinities of the inhibitors for the two CDKs. These results highlight the usefulness of this computational approach in evaluating the selectivity of a CDK inhibitor, and demonstrate the necessity of considering protein flexibility and solvent effects in designing new selective CDK4-selective inhibitors.

Introduction

Cyclin-dependent kinases (CDKs) are a family of heterodimeric serine/threonine protein kinases comprising a catalytic CDK subunit and an activating subunit (cyclin), and play critical roles in cell-cycle progression.[1] Because uncontrolled cell proliferation is a cardinal characteristic of a cancer, an effective CDK inhibitor capable of regulating tumor cell growth would be expected to be a promising therapeutic agent for the treatment of cancer.[2,3] Of the CDKs identified so far, CDK4/cyclin D and CDK6/cyclin D are two desirable targets for inhibitor design, in view of the several lines of persuasive evidence for the link between their activities and many types of cancer.[4]

During the past decade, a number of CDK inhibitors with a wide variety of potencies and specificities have been identified.[5–16] Although most of them are potent inhibitors of CDK1 and/or CDK2, several inhibitor scaffolds with a moderate selectivity for CDK4—including aminothiazoles, benzocarbozoles, and pyrimidine-based compounds[17–23]—have also been identified. The difficulty in developing a new CDK4-selective inhibitor is a consequence of the similarity in the amino acid residues constituting the active sites of the various CDKs and the lack of a three-dimensional structure of CDK4. Nonetheless, the crystal structures of other homologous CDKs in non-ligand-bound and ligand-bound forms have provided much information pertinent to the catalytic mechanism and rational design of inhibitor drugs for CDK4.[24–28] Indeed, a structure-based drug design campaign has proved successful in designing new selective CDK4 inhibitors through the use of the known structural information for CDK2 and comparison of the active site amino acid sequences of CDK2 and CDK4.[29–32]

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. Carvalli et al., for example, have developed force field parameters based on quantum chemical calculations on a relevant model system for simulating the binding of ATP to CDKs.[33,34] 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 recently, Sims et al. have established a binding free energy model for CDK2 inhibitors within the framework of the continuum solvent model.[35] This approach proved successful in reproduction of the relative in-

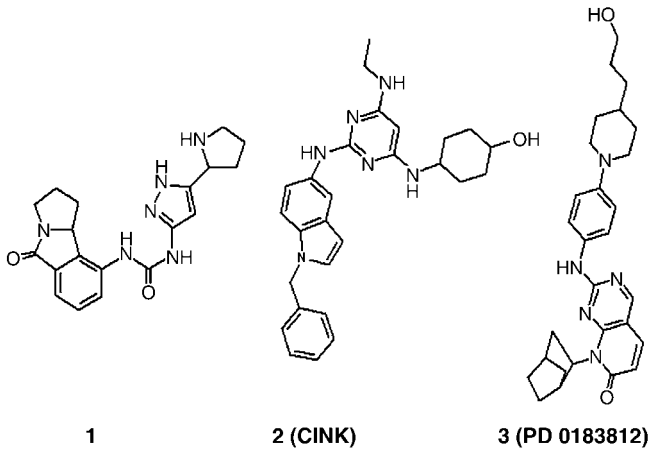
[a] Dr. H. Park, Prof. S. Lee
School of Chemistry and Molecular Engineering, Seoul National University
Seoul 151-747 (Korea)
E-mail: hwangseo@snu.ac.kr
sangyoub@snu.ac.kr

[b] Dr. M. S. Yeom
Supercomputing Research Department, Supercomputing Center
Korea Institute of Science and Technology Information
Yusung-Gu, Eoeun-Dong 52, Daejon 305-806 (Korea)
Fax: (+82) 2-889-1568

  DOI: 10.1002/cbic.200400214

hibitory activities of flavopiridol analogues. On the basis of the validated utility of the computational approach, they suggested new putative inhibitors potentially more potent than the lead compound.

Consideration of protein flexibility is indispensable to the critical evaluation of ligand-binding affinity,[36–39] and the conformational plasticity of the catalytic domain is a hallmark of protein kinases.[40] Furthermore, recent theoretical and experimental studies have shown that conformational changes of CDKs play a critical role in ligand binding.[41–44] Here we therefore investigate the characteristic dynamic properties of CDK4 that distinguish it from CDK2 in complexation with its selective inhibitors. For this purpose, comparative solution-phase MD simulations have been carried out with the aid of the X-ray structure of CDK2 and the homology-modeled CDK4, onto which the three CDK4-selective inhibitors described in Scheme 1 and Table 1 are docked. It is shown that CDK2 and



Scheme 1. *Molecular structures of the three CDK4-selective inhibitors under investigation.*

CDK4 have very different dynamic flexibilities in the disordered loop region when the CDK4-selective inhibitors are bound to their active sites. We have also examined the effects of the difference in solvent dynamics around the active sites of the two CDKs on the relative strength of protein–ligand interactions in the hope that comparative analysis of structural details would provide information pertinent to the design of new potent CDK4-selective inhibitors.

Table 1. *IC50 values of the three inhibitors under investigation against CDK2–cyclin A and CDK4–cyclin D complexes.*

Inhibitor	IC_{50} [μM]		Reference
	CDK2–cyclin A	CDK4–cyclin D	
1	25	0.21	[31]
2	0.21	0.008	[18]
3	>50	1.5	[19]

Computational Methods

Homology modeling: The peptide sequence of human CDK4, comprising 303 amino acid residues, was retrieved from the SWISS-PROT protein sequence data bank (accession number: P11802).[45] Sequence alignment between CDK4 and CDK2 was then derived by use of the ClustalW 1.82 package and the BLOSUM matrices for scoring the alignments.[44] From the best-scored sequence alignment, we obtained 3D structural models for CDK4 by use of the MODELLER program in version 6v2.[47] Since, for the purposes of our study of enzyme inhibition, it is more relevant to consider a model of an enzyme–inhibitor complex rather than that of apoenzyme, homology modeling of CDK4 was performed on the basis of the X-ray structure of CDK2 in complexation with the potent inhibitor NU6102.[25] With respect to the structures of the gap regions, the coordinates were built from a randomized distorted structure located approximately between the two anchoring regions as implemented in MODELLER 6v2. To obtain a reasonable structural model for the target, we employed an optimization method involving conjugate gradients and molecular dynamics to minimize violations of the spatial restraints. To increase the accuracy of the structures of flexible loops, the loop modeling was also performed with the enumeration algorithm.[48] Of the ten calculated structural models, the one with the lowest value of MODELLER objective function was selected for use in the subsequent docking simulations with the CDK4-selective inhibitors.

Docking experiments: To obtain starting structures for MD simulations of the CDK–inhibitor complexes, all the inhibitors shown in Scheme 1 were docked onto the active sites of CDK2 and CDK4 from which the inhibitor NU6102 had been removed. In this docking simulation we used the AutoDock 3.0.5 program,[49] which has been widely used in the literature, to find favorable binding modes of ligands in enzymatic active sites. It combines a rapid energy evaluation through precalculated grids of affinity potentials with various search algorithms to find suitable binding positions for a ligand. Although the protein structure has to be fixed, the program allows torsional flexibility of a ligand. The protein atom coordinates for CDK2 and CDK4 were taken from the X-ray structure in complexation with the inhibitor NU6102[25] and from the final model of homology modeling, respectively. Docking to the active sites of the two CDKs was then carried out by use of the Lamarckian genetic algorithm and the empirical scoring function, which has the following form:

$$\Delta G^{aq}_{bind} = W_{vdW} \sum_i \sum_j \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right) + W_{h\,bond} \sum_i \sum_j E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}}\right) + W_{elec} \sum_i \sum_j \frac{q_i q_j}{\varepsilon r_{ij}} + W_{tor} N_{tor} + W_{sol} \sum_i \sum_j (S_i V_j + S_j V_i) e^{(-r_{ij}^2/2\sigma^2)} \quad (1)$$

here W_{vdW}, $W_{h\,bond}$, W_{elec}, W_{tor}, and W_{sol} are weighting factors of van der Waals, hydrogen bonding, electrostatic interactions, torsional terms, and desolvation energies of the inhibitors, respectively. The hydrogen bond term has an additional weighting factor, $E(t)$, representing the angle-dependent directionality. A sigmoidal distance-dependent dielectric function as proposed by Mehler et al.[50] was used in computing the interatomic electrostatic interactions in the CDK–inhibitor complexes. In the desolvation term, S_i and V_i are the solvation parameter and the fragmental volume, respectively, of atom i.[51] From the conformations obtained from 200 independent docking runs, the results differing by less than 1.5 Å in positional root mean square deviation (rmsd) were clustered together. The

most stable configuration of enzyme–inhibitor complex was then selected for further analysis.

Molecular dynamics simulations: It has been found that the AMBER force field is appropriate for investigating dynamic properties of CDKs,[34] so MD simulations of CDK2 and CDK4 and their complexes with the inhibitors shown in Scheme 1 were carried out with the aid of the SANDER module in AMBER 7[52] and the force field reported by Cornell et al.,[53] starting from the most stable of the CDK–inhibitor complexes found in the preceding docking simulation. To be consistent with the standard AMBER force field, we computed the electrostatically derived atomic charges of all three inhibitors under study by the RESP method[54] for the fully optimized geometries at the RHF/6–31G* level of theory. Missing force field parameters for the inhibitors were estimated from similar chemical species in the AMBER database.

We used the X-ray structure of CDK2 in complexation with the inhibitor NU6102[25] to determine the protonation states of the Asp, Glu, and Lys residues. The side chains of the Asp and Glu residues, for example, were assumed to be neutral if their carboxylate oxygens of OD or OE atoms were located within 3.5 Å of a hydrogen bond-accepting group, including the backbone aminocarbonyl oxygen. Similarly, the lysine side chains were assumed to be ionized unless the NZ atom was in proximity to a hydrogen bond donating group. In this way, Glu2, Glu42, Asp206, and Lys75 were found to be neutral, while the rest of the side chains of Asp, Glu, and Lys residues were ionized. The same procedure was used for the homology-modeled CDK4 to assign the protonation states of its Asp, Glu, and Lys residues.

The all-atom models for six CDK–inhibitor complexes were neutralized by addition of counter-ions and were then immersed in a rectangular box containing 11891 TIP3P[55] water molecules for CDK2 and 13737 for CDK4. After 1000 cycles of energy minimization to remove bad steric contacts, we equilibrated all six CDK–inhibitor complex systems beginning with 20 ps equilibration dynamics of the solvent molecules at 300 K. The next step involved equilibration of the solute with a fixed configuration of the solvent molecules for 10 ps at 10, 50, 100, 150, 200, 250, and 300 K. The equilibration dynamics of the entire system was then carried out at 300 K for 100 ps. Following the equilibration procedure, 1.5 ns MD simulations were carried out with a periodic boundary condition in the NPT ensemble at 300 K by Berendsen temperature coupling[56] and at constant pressure (1 atm) by isotropic molecule-based scaling. The SHAKE algorithm,[57] with a tolerance of 10^{-6}, was applied to fix all bond lengths involving hydrogen atoms. We used a time step of 1.5 fs and a nonbonding interaction cutoff radius of 12 Å; the trajectory was sampled every 0.15 ps (100 step intervals) for analysis. Although a cutoff distance of 12 Å was used for short-range electrostatic and van der Waals interactions, all electrostatic interactions beyond this distance were accounted for by use of the particle-mesh Ewald summation with an interpolation order of 4.[58]

Results and Discussion

Homology modeling of CDK4

Figure 1 displays the sequence alignment of CDK4 with CDK2, used as the target and as the template, respectively, in comparative protein structure modeling of CDK4. We note that the amino acid sequences are completely different in the disordered loop region (residues 37–44 for CDK2 and 39–46 for CDK4), indicating a significant difference in their behavior in response to ligand binding. Structural evaluation of the final model for CDK4 with PROCHECK[59] indicates that the backbone Φ and Ψ dihedral angles of 76.6%, 21.0%, and 2.4% of the residues are located within the most favorable, the additionally allowed, and the generously allowed regions of the Ramachandran plot, respectively, with no residues in the disallowed region. This good stereochemical quality is not surprising in view of the high sequence identity (44%) and similarity (63%) between the template and the target as illustrated in Figure 1.

Figure 1. *Sequence alignment of CDK2 and CDK4. Asterisk and colon represent identity and similarity, respectively, between the corresponding residues.*

Figure 2 compares the X-ray structure of the CDK2–NU6102 complex and the final model of the CDK4–NU6102 complex. Common to both CDKs is the fact that, of the flexible loops,

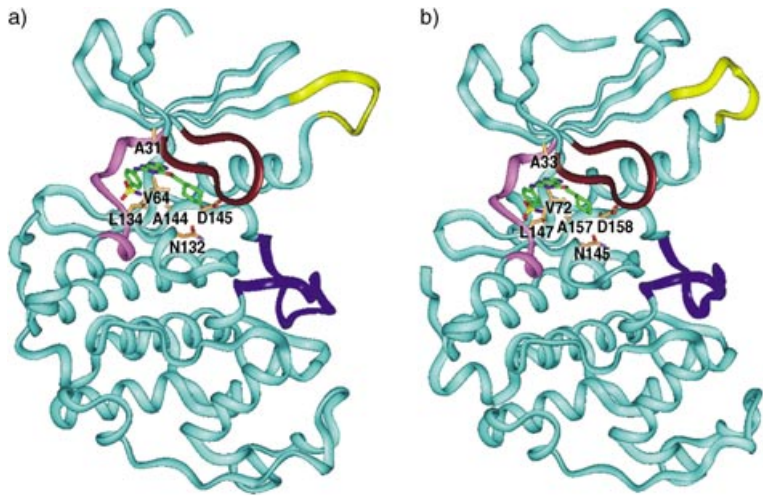


Figure 2. *Comparative views of: a) the X-ray structure of CDK2, and b) the final model of CDK4 in their complexes with NU6102. Carbon atoms of the inhibitor and the enzymatic groups are shown in green and orange, respectively. The Gly loops (brown), disordered loops (yellow), ATP-binding regions (pink), and T loops (blue) are also indicated.*

only the Gly loop is involved in a direct interaction with the inhibitor. As might be expected from the difference in amino acid sequences, however, significant conformational differences around the disordered and the T loops are observed. However, the template and the target possess very similar folding structures on the whole and are superimposable over the main chain atoms. As in the crystal structure of the CDK2–NU6102 complex, the inhibitor is bound around the ATP-binding site (consisting of residues 93–102 in CDK4). The hydrophobic and hydrophilic residues (Ala33, Val72, Asn145, Leu147, Ala157, and Asp158) in the active site of CDK4 are identical to those in CDK2 (Ala31, Val64, Asn132, Leu134, Ala144, and Asp145). The residues Val14, Ala16, and Thr19 in the N-terminal Gly loop and His95, Val96, Asp97, Arg101, and Thr102 in the ATP-binding sites of CDK4 are also components of the active site, but these are different from the corresponding Glu12, Thr14, Val17, Phe82, Leu83, His84, Lys88, and Lys89, respectively, in CDK2. This may serve as a key piece of information for the design of a new CDK4-selective inhibitor, as confirmed in part in the de novo design program of Honma et al.[30]

Docking experiments

To check the usefulness of the AutoDock 3.0.5 program in preparing a starting structure of the CDK–inhibitor complex for MD simulation, we examined its accuracy in predicting binding modes of CDK inhibitors against 20 X-ray structures of recently reported CDK2–inhibitor complexes. As shown in Table 2, the rmsd between the bound conformations (X-ray) and the most stable conformations generated with AutoDock in most cases fell under 3.0 Å, except in the cases of very small inhibitors with fewer than 15 heavy atoms, as in 1PXI and 1PXJ. Moreover, 13 of the remaining cases produce a 1.5 Å hit. These validation results indicate that the most stable AutoDock configuration of a CDK–inhibitor complex may be a reasonable choice as a starting point for investigation of its dynamic properties in aqueous solution.

Figure 3 shows the lowest-energy AutoDock configurations for compounds **1**, **2**, and **3** in the active sites of CDK2 and CDK4. The inhibitor **1** has a similar binding mode in both CDKs, as it is small enough to be fully accommodated in the active sites. This indicates that its binding preference should be explained by the relative strengths of various interaction terms. In contrast, an increase in the inhibitor length leads to differences in binding modes for the inhibitors **2** and **3** in the active sites of CDK2 and CDK4, in particular at the terminal aliphatic groups. In CDK4, the terminal hydroxy moieties of the two inhibitors each form a hydrogen bond with the ATP-binding site side chains (Arg101 and Thr102). In CDK2, on the other hand, the hydroxy group of **2** is hydrogen-bonded to the Lys129 side chain residing at the top of the C-terminal domain, while that of **3** is directed toward the backbone of Lys88. This discrepancy in the binding modes is due to the difference in the structural arrangements of the side chains of the two residues at the end of the ATP-binding site. Indeed, the Arg101 and Thr102 side chains in CDK4 are associated through a strong hydrogen bond at the entrance to the active site, while those of Lys88 and Lys89 in CDK2 are kept apart from each other and directed toward the C-terminal domain and the backbone atoms of the ATP-binding site, respectively.

Table 2. *Validation results for AutoDock 3.0.5 in the prediction of binding modes in various X-ray structures of CDK2–inhibitor complexes.*

PDB code	rmds of top score	PDB code	rmds of top score
1E1X	1.88	1H1S	1.05
1G5S	1.09	1P2A	0.95
1GIH	0.86	1PXI	6.66
1H00	2.39	1PXJ	6.97
1H01	2.69	1PXK	2.02
1H06	2.28	1PXL	1.31
1H07	1.31	1PXM	0.68
1H1P	1.49	1PXN	1.38
1H1Q	1.04	1PXO	1.47
1H1R	0.82	1PXP	0.85

MD simulation

We checked the reliabilities of the MD simulations by examining whether the protein structures in the CDK–inhibitor complexes remained stable under the simulation conditions described in the previous section. For this purpose, we calculated the rmsd from starting structures ($rmsd_{init}$) for all backbone C_α atoms as a function of simulation time, as shown in Figure 4. The $rmsd_{init}$ values remain within 1.5 Å for CDK2 and within 2.5 Å for CDK4; this demonstrates the conformational stabilities of protein structures. However, the $rmds_{init}$ value of CDK4

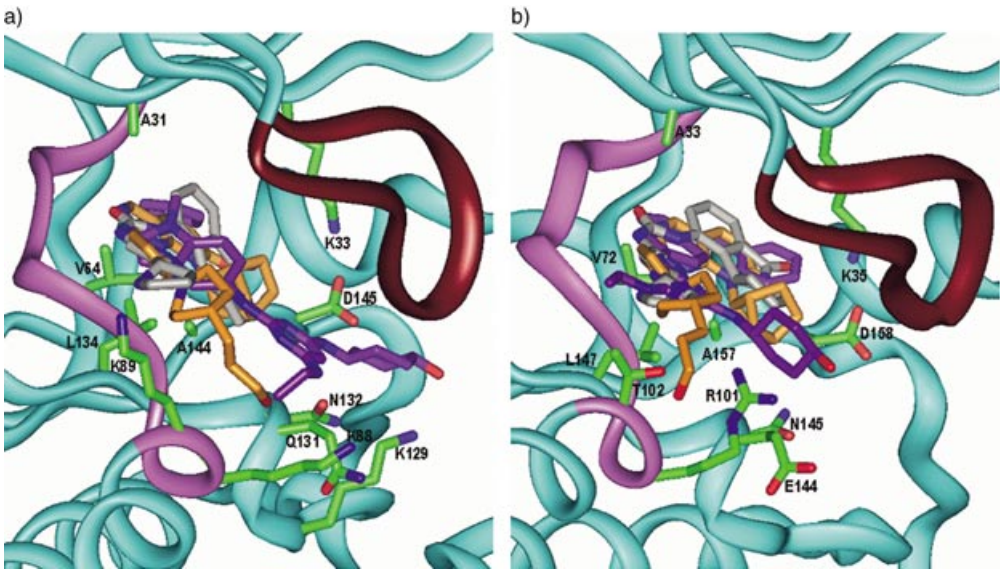


Figure 3. *Superimposition of the docked conformations of inhibitors* ***1*** *(gray),* ***2*** *(violet), and* ***3*** *(orange) in the active sites of: a) CDK2, and b) CDK4. The Gly loops and ATP-binding regions are indicated in brown and pink, respectively.*

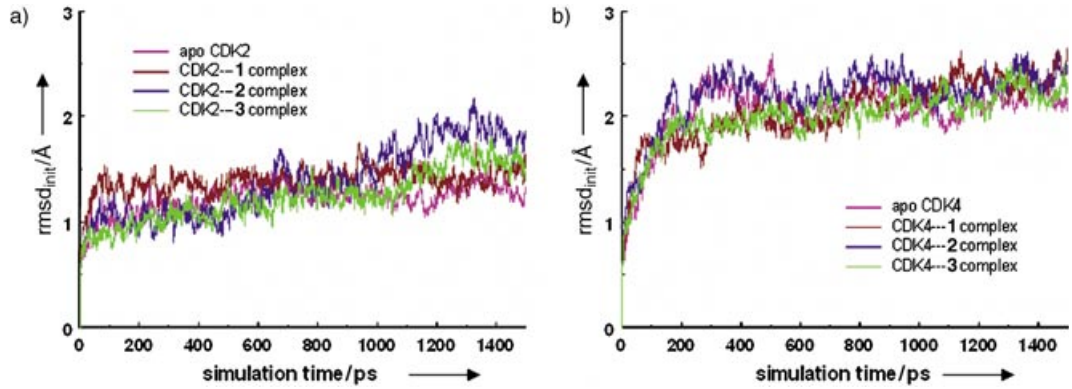


Figure 4. *Time dependence of the root mean square deviations from starting structures ($rmds_{init}$) for: a) CDK2, and b) CDK4 in complexation with the three inhibitors under investigation.*

is higher than that of CDK2 by 0.5–1.0 Å depending on the inhibitor structures, indicating that the former enzyme undergoes a larger conformational change than the latter when they are stabilized in aqueous solution in complexation with the inhibitors. This is consistent with the recent computational finding that a homology-modeled protein structure is dynamically unstable relative to a high-resolution crystal structure.[60] However, the structural relaxation of CDK4 occurs in the early stages of simulation, and the time evolution of $rmds_{init}$ for CDK4 exhibits more stable behavior than that of CDK2 during the later part of simulation. Judging from these dynamic properties, it seems necessary to carry out a more detailed analysis on the MD trajectories of CDK4–inhibitor complexes under study as compared with those of the CDK2 counterparts.

Because the ligand binding site of a CDK is located in the cleft between the small N-terminal lobe and the larger C-terminal lobe, involving the contact of the apex of the Gly loop and the starting point of the T loop, the inhibitor binding is likely to affect the relative motion of Thr14 (Ala16) with respect to Gly147 (Gly160) in CDK2 (CDK4). Figure 5 compares the probability distributions of the associated interatomic distances between CA atoms for non-ligand-bound and ligand-bound CDKs. In CDK2 we note that the most probable distance between the CA atoms of Thr14 and Gly147 decreases from 5.7 Å in the resting form to 5.1–5.3 Å in the enzyme–inhibitor complexes. Similarly, in CDK4 the distribution of the distance between the CA atoms of Ala16 and Gly160 shows a sharp peak around 5 Å in the enzyme–inhibitor complexes, whereas the distribution in the resting form has a broad spectrum between 4.5 and 6 Å. These results indicate the connection between binding of CDK inhibitors and the closing of a flap involving Gly loop and T loop. This kind of flap-closing due to inhibitor binding has also been observed in binding of fullerene-based inhibitors in the active site of HIV-1 protease.[61]

From an X-ray crystallographic analysis of CDK2, Wu et al. recently showed that the disordered loop comprising residues

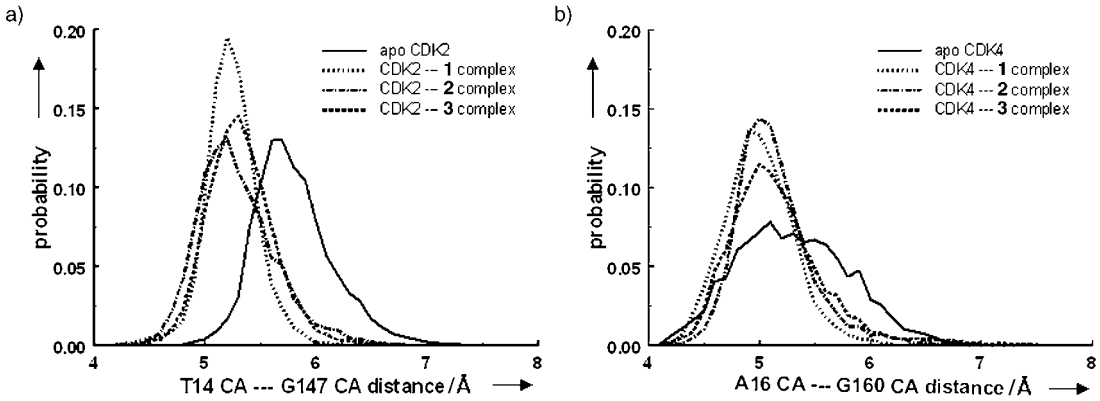


Figure 5. *Calculated probability distribution functions for the distances: a) between the CA atoms of Thr14 and Gly147 in CDK2, and b) between those of Ala16 and Gly160 in CDK4.*

37 through 44 is exceptionally flexible, considering the high temperature factor and the absence of electron density at the apex of the loop.[28] However, such a high-amplitude motion seems to be damped out upon binding of inhibitors, as can be judged from the decreases in the associated temperature factors in a variety of X-ray crystallographic data for CDK2–inhibitor complexes. Moreover, the loop flexibility has been shown to decrease substantially as the inhibitor potency increases to nanomolar levels.[25] The specificity of a CDK inhibitor may thus be related to the difference in malleability changes at the disordered loops of CDK2 and CDK4 upon binding of the inhibitor. Figure 6 compares the calculated B-factors of the C_α atoms of the two CDKs in both non-ligand-bound and ligand-bound forms by the following relation:

$$B_i = {}^8/_3\pi^2\langle\Delta r_i\rangle^2 \qquad (2)$$

where $\langle\Delta r_i\rangle$ is the RMS atomic fluctuation of the C_α atom of residue *i*. Consistent with the X-ray crystal structures of CDK2 and its various enzyme–inhibitor complexes, the calculated B-factors of CDKs show a major peak in the region of residues 40–43. In the resting forms, the disordered loops of CDK2 and CDK4 have B_i values that differ by less than 10 Å^2 (see Figure 6a). In the complexes with the CDK4-selective inhibitors, however, the disordered loop of CDK2 shows much more flexibility than that of CDK4, with associated difference in B_i values of 50–250 Å^2. This implies that, upon binding of the

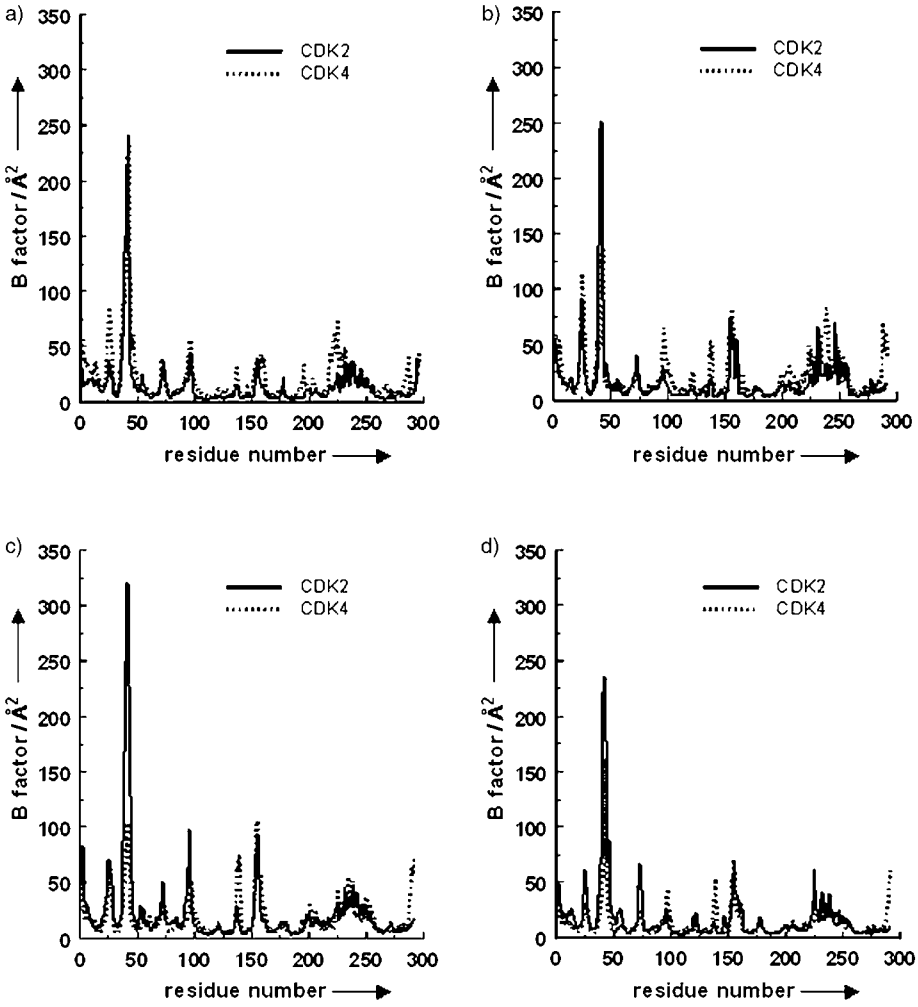


Figure 6. *The calculated B-factors of C_α atoms of CDK2 and CDK4 in the resting form (a), and in complexation with the CDK4-selective inhibitors **1** (b), **2** (c), and **3** (d). For ease of comparison, the residues of CDK4 are renumbered to correspond to the amino acid sequence of CDK2.*

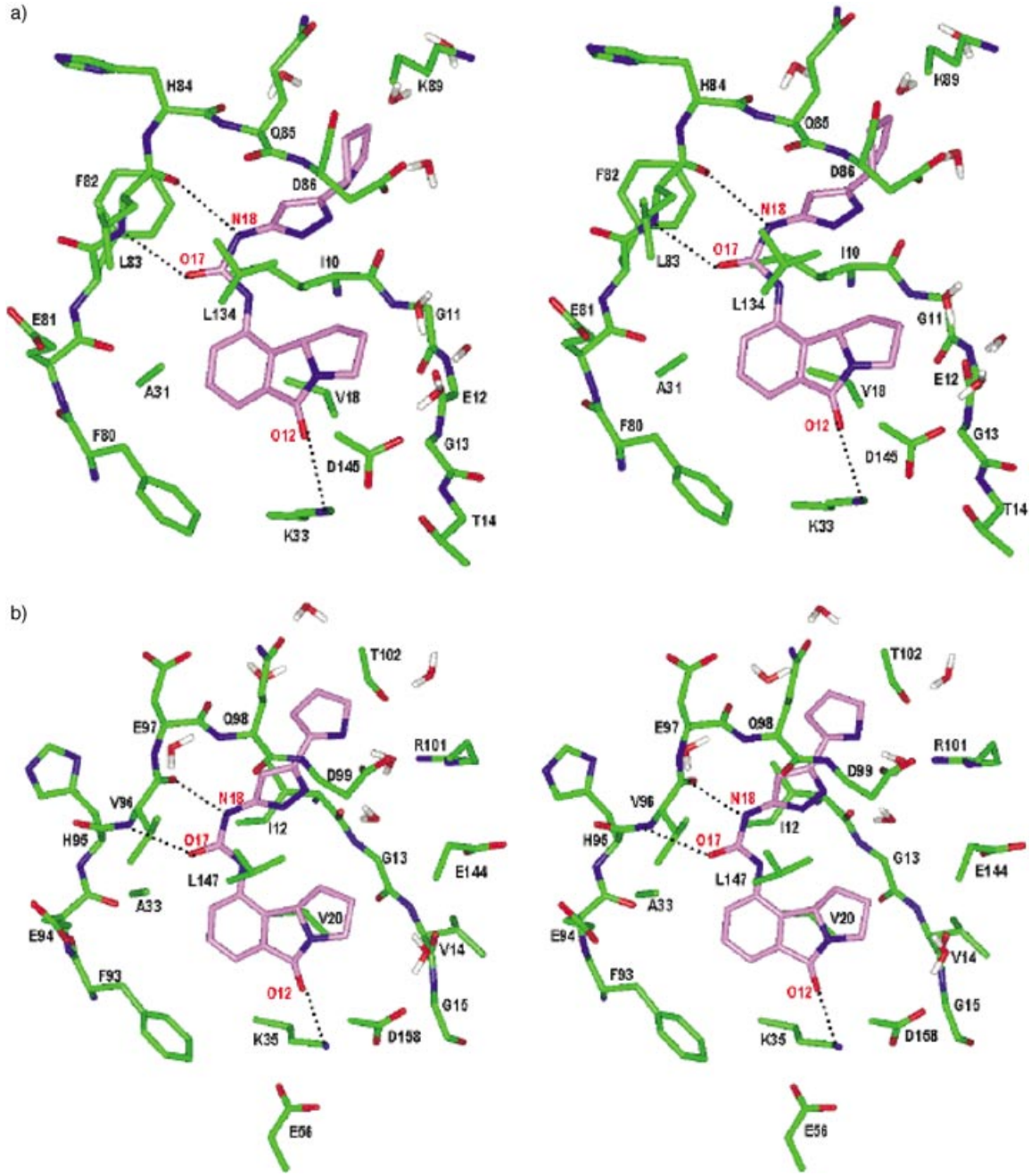


Figure 7. *Stereoviews of representative MD trajectory snapshots of: a) CDK2–**1**, and b) CDK4–**1** complexes, including solvent molecules found near the active site. The inhibitor atoms involved in intermolecular hydrogen bonds are designated by red letters. Each dotted line indicates a hydrogen bond.*

CDK4-selective inhibitors, high-amplitude motions of the flexible residues of CDK4 should be damped out to a greater extent than in CDK2. It can therefore be argued that a potent CDK4-selective inhibitor, when bound to the active site in CDK4, should keep the disordered loop more motionally restricted than in other homologous CDKs. This hypothesis may be supported further by the experimental finding that a high-amplitude motion of the disordered loop preceding the PSTAIRE helix (residues 45–51) is responsible for the binding of cyclin, which is indispensable for kinase activities of all CDKs.[24]

Figure 7 compares representative MD trajectory snapshots of CDK2 and CDK4 in complexation with the inhibitor **1**. Consistent with the X-ray crystal structures reported by Ikuta et al.,[31] the inhibitor amide moiety forms a bidentate hydrogen bond with the backbone amide group of Leu83 in CDK2 and Val95 in CDK4, both located in the middle of the ATP-binding site. A stable hydrogen bond is also established between the inhibitor carbonyl oxygen (O12) and the side chains of Lys33 in CDK2 and Lys35 in CDK4. Another structural feature common to the two enzyme–inhibitor complexes is that the

terminal bicyclic group of the inhibitor is stabilized by interaction with the terminal residues of ATP-binding site, with the Gly loop, and with water molecules from bulk solvent. In view of the overall similarity in the binding modes of **1** in the active sites of CDK2 and CDK4, differential strengths of the corresponding interactions between protein and inhibitor groups are believed to be responsible for the observed difference in inhibitory activities toward the two CDKs.

Table 3 lists the dynamic time averages and standard deviations of the interatomic distances associated with the hydrogen bond interactions in the two CDKs in complexation with the three inhibitors under investigation. We note that the bidentate hydrogen bonds involving the amide group of **1** are dynamically stable in both CDKs, with similar hydrogen bond lengths in most parts of the simulation. In contrast, the N–H···O hydrogen bonds between the lysine side chain at the bottom of the active site and the O12 atom of the inhibitor have a much higher dynamic stability in the CDK4–**1** complex than in the CDK2–**1** complex. For example, the Lys35···O12 hydrogen bond is maintained for 94% of the simulation time with an associated interatomic distance of 1.92 Å on average in the former complex, as compared to the corresponding 81% of residence time and 2.06 Å of average hydrogen bond length in the latter. In this comparison, we use 2.2 Å as the length defining a hydrogen bond, as suggested by Jeffrey.[62] It is thus apparent that the experimentally observed selectivity of **1** in favor of binding to CDK4 is largely due to the formation of a stronger Lys35···O12 hydrogen bond in CDK4 in relation to the Lys33···O12 counterpart in CDK2. Because both Lys33 of CDK2 and Lys35 of CDK4 reside in proximity to the disordered loop, the increased strength of the Lys35···O12 hydrogen bond accounts for the decrease in dynamic flexibility of the loop on going from CDK2–**1** to CDK4–**1** complexes.

Consistent with the low inhibitory activity of **2** against CDK2—with an associated IC_{50} value above 50 μM[18]—one end of the inhibitor **2** is exposed to bulk solvent without a significant interaction with the backbone atoms of the ATP-binding site. As shown in Figure 8a, solvent molecules diffuse into the active site to interact with protein and inhibitor groups, preventing tight binding of the inhibitor in the CDK2 active site. In contrast, the approach of solvent molecules to the ATP-binding site is restricted in the CDK4–**2** complex, in which the benzyl, indole, and pyrimidine moieties of **2** are bound to its backbone and side chain atoms (Figure 8b). Such a reduction in solvent accessibility on going from CDK2 to CDK4 active sites can be attributed to the differences in the structure and arrangement of the side chains at the ends of the ATP-binding sites. While the Lys88 and Lys99 side chains in the CDK2–**2** complex are exposed to bulk solvent, those of Asp99, Arg101, and Thr102 in the CDK4–**2** complex form a cyclic hydrogen bond at the top of the ATP-binding site, protecting **2** from solvation in the CDK4 active site. As a consequence of the reduced solvent accessibility, the hydrophobic side chains of Ala157 and Phe159 approach the inhibitor from the C-terminal domain to form a hydrophobic pocket accommodating the inhibitor benzyl group, further facilitating tight binding of **2** in the CDK4 active site.

Despite the reduced solvent accessibility in the CDK4 active site, one water molecule (Wat8948 in Figure 8b) is found between the inhibitor and the ATP-binding site in the CDK4–**2** complex. As shown in Table 3, this structural water molecule receives and donates hydrogen bonds from the inhibitor amine moiety (connecting the indole and pyrimidine rings) and to the backbone carbonyl oxygen of Val96, respectively. Such a water-bridged hydrogen bond is established in the early stage of simulation and maintained during the entire course of simulation. From this dynamic stability and the difference in binding modes of **2** in the active sites of the two CDKs, it is obvious that the solvent-mediated hydrogen bond plays a role in positioning the inhibitor in such a way as to be stabilized in the CDK4 active site rather than being repelled to bulk solvent as in the CDK2–**2** complex. Thus, the limited solvent accessibility and the formation of a solvent-bridged hydrogen bond may be invoked to explain the selectivity of **2** in favor of binding to the CDK4 active site.

The structural features revealed in typical MD trajectory snapshots of CDK2 and CDK4 ligand-bound with the inhibitor **3** are also consistent with its selective inhibition of CDK4.[19] As shown in Figure 9a, the side chains of Lys88 and Lys89 at the top of the active site in the CDK2–**3** complex are far part from each other, facilitating the diffusive intrusion of solvent molecules into the active site. In contrast, the side chains of Asp99, Arg101, and Thr102 in the CDK4–**3** complex are associated through a stable cyclic hydrogen bond, forming a blockade to the active site entrance in a similar way as in the CDK4–**2** complex (Figure 9b). The number of water molecules around the inhibitor decreases on going from CDK2–**3** to CDK4–**3** complexes, so that the inhibitor should be bound more tightly in the latter than in the former, presumably due to a reduced potency-lowering effect of solvent molecules. The *n*-propanol and cyclohexamine moieties of **3**, for example, are oriented along the narrow active site gorge involving the hydrogen-bond triad, Lys22, and Gln98 in the CDK4–**3** complex, while

Table 3. *Dynamic time average and standard deviation of the interatomic distances[a] associated with the hydrogen bond interactions in the two cdks in complexation with the three inhibitors under investigation.*

Complex	Hydrogen bond	Average	Standard deviation
CDK2–1	L83 H···O18	2.11	0.19
	L83 O···H19	1.90	0.13
	K33 HZ···O12	2.06	0.28
CDK4–1	V96 H···O18	2.03	0.21
	V96 O···H19	1.94	0.17
	K35 HZ···O12	1.92	0.16
CDK4–2	V96 O···Wat8948 H	1.97	0.48
	Wat8948 O···H11	2.25	0.42
CDK2–3	K33 HZ···O11	2.28	0.40
CDK4–3	K35 HZ···O11	1.97	0.32
CDK4–1	D99 OD···T102 HG	1.74	0.14
	D99 OD···R101 HH	1.81	0.12
CDK4–2	D99 OD···T102 HG	1.71	0.11
	D99 OD···R101 HH	2.05	0.23
CDK4–3	D99 OD···T102 HG	1.73	0.10
	D99 OD···R101 HH	1.94	0.18

[a] All interatomic distances are given in Å.

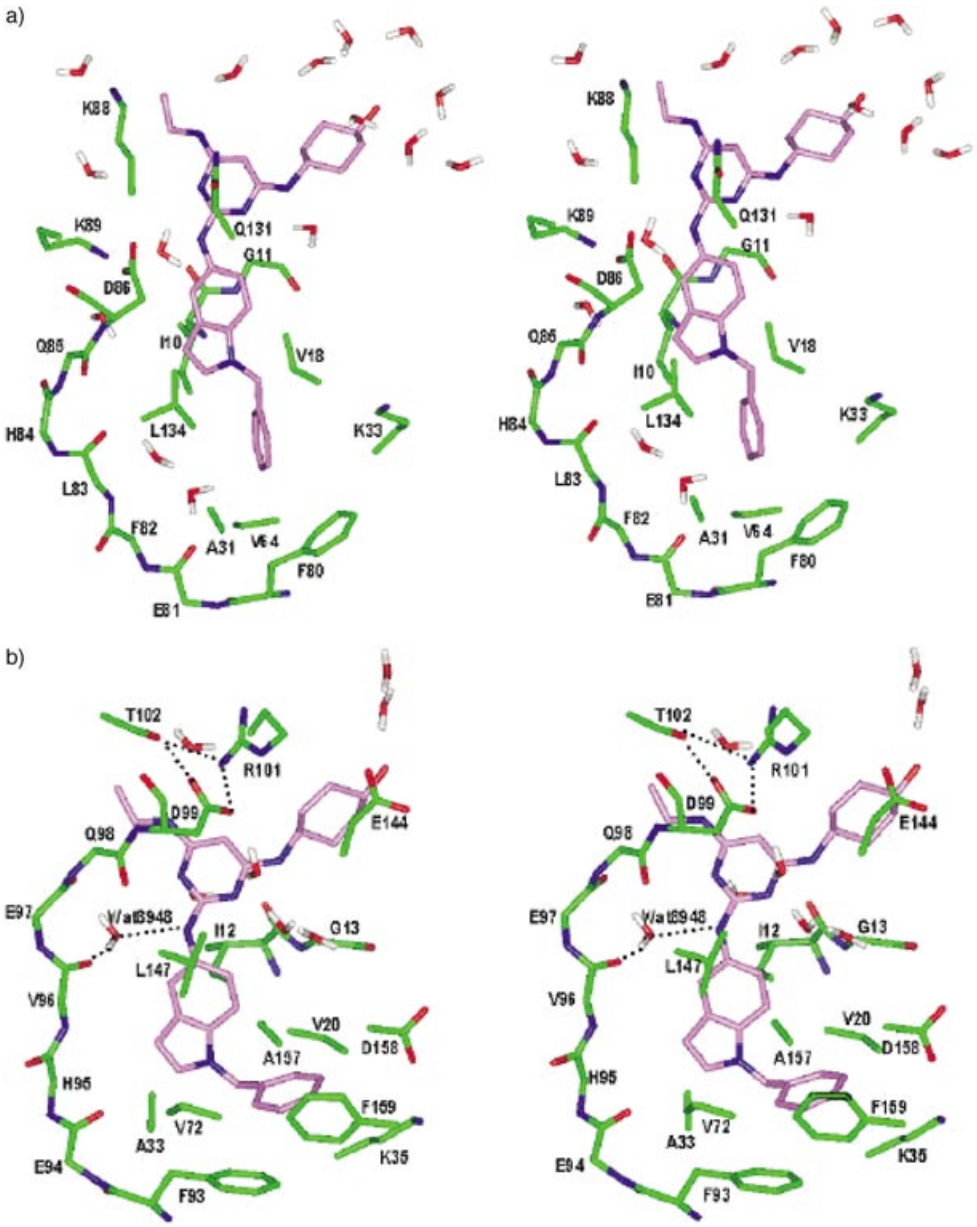


Figure 8. *Stereoviews of representative MD trajectory snapshots of: a) CDK2–**2**, and b) CDK4–**2** complexes including solvent molecules found near the active site. Each dotted line indicates a hydrogen bond.*

they are exposed to bulk solvent in the CDK2–**3** counterpart. From the similarities in the binding modes of the [2,3-d]pyridopyrimidine backbone of **3** and its lateral position in the active sites of the two CDKs, however, the solvent effects on the inhibitory selectivity should be less critical for **3** than **2**. This may account for the relatively low selectivity of the former in relation to the latter.[18,19]

As a consequence of the differences in solvent dynamics and binding modes of **3** in the active sites of the two CDKs, the hydrogen bond between the inhibitor carbonyl oxygen (O11) and the side chain of the lysine residue at the bottom of active site is established more strongly in the CDK4–**3** complex than in the CDK2–**3** counterpart (see Table 3). Indeed, the O···H–N hydrogen bond length in the former is kept shorter than in the latter for 79% of the simulation time with the associated difference in average length of 0.31 Å. This result is consistent with that for the inhibitor **1**, implying that the relative strength of a hydrogen bond of an inhibitor group with Lys35 in CDK4 with respect to that with Lys33 in CDK2 can be correlated with the inhibitor selectivity for CDK4. Both lysine residues are located at the bottom of the ATP-binding site and the starting point of the disordered loop, so that the strength of the hydrogen bonds can be affected by the solvent-induced changes in lateral position of an inhibitor and the loop flexibility in the CDK–inhibitor complexes.

We finally address the dynamic stabilities of the hydrogen-bond triad involving Asp99, Arg101, and Thr102 of CDK4 in its complexes with the three inhibitors. It is noteworthy that such a triad cannot be formed in CDK2 because the side chains of Lys88 and Lys89 are kept apart from each other and are directed toward the C-terminal domain and the backbone atoms of the ATP-binding site, respectively. In all three cases of CDK4–inhibitor complexes, the triad is formed in the initial stage of simulation and maintained during most of the simulation time. Another common feature is that some trajectory snapshots exhibit a short, strong hydrogen bond between Asp99 and Thr102 with associated interatomic distances of $\approx$1.5 Å, suggesting the possible involvement of a low-barrier hydrogen bond[62] in the CDK4 active site. As can be seen in Table 3, on the other hand, the stability of the Asp99···Arg101 hydrogen bond is affected substantially by a change of inhibitor. The hydrogen bond remains stable in the CDK4–**1** complex, but becomes less stable on moving to the CDK4–**2** and CDK4–**3** complexes. If we compare the binding modes of the three inhibitors shown in the representative MD trajectory snapshots, we may deduce that the Asp99···Arg101 hydrogen bond becomes unstable with increasing size of the inhibitor groups residing at the top of active site. This substantiates the important roles played by the hydrogen-bond triad in ligand binding.

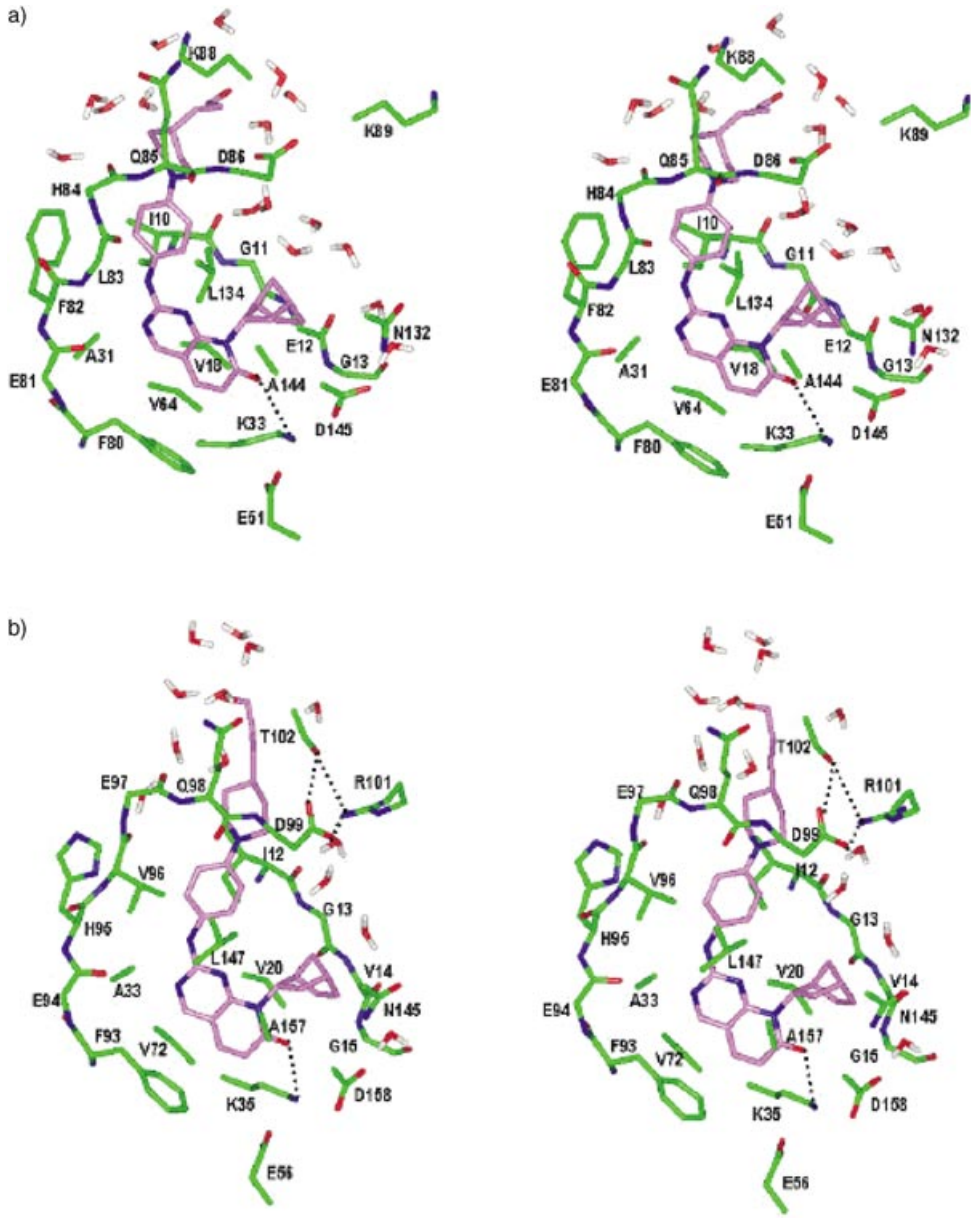


Figure 9. *Stereoviews of representative MD trajectory snapshots of: a) CDK2–**3**, and b) CDK4–**3** complexes including solvent molecules found near the active site. Each dotted line indicates a hydrogen bond.*

Conclusion

The influences of protein structure flexibility and environmental solvent molecules on protein–ligand association are well appreciated. In this study, we have investigated the selective inhibition of CDK4 from such perspectives by means of a combined computational protocol involving homology modeling, docking experiment, and solution-phase MD simulations. The overall folding structure of the homology-modeled CDK4 is very similar to that of CDK2, with little difference in local loop structures. We find that the high-amplitude motion of the disordered loop of CDK4 is damped out in the presence of its selective inhibitors, whereas their binding in the CDK2 active site has an insignificant effect on the loop flexibility. In the active site of CDK4, the terminal residues of the ATP-binding site (Asp99, Arg101, and Thr102) form a stable hydrogen-bond triad, making the entrance of the active-site gorge narrow, whereas in CDK2 the corresponding residues are kept apart from each other. Due to this structural feature of CDK4, the diffusive intrusion of solvent molecules into the active site is restricted, and the CDK4-selective inhibitors are kept stable in the active site. In contrast, the inhibitors bound in the CDK2 active site are exposed to the bulk solvent, which has the effect of weakening the associated enzyme–inhibitor interactions. Consistently with the differences in loop flexibility and solvent accessibility, the relative strengths of hydrogen bonds between the inhibitors and the side chains of the lysine residues at the bottoms of the active sites prove to be a significant determinant for the inhibitor selectivity between CDK4 and CDK2. The results found in this study may serve as a key piece of information for the structure-based design/discovery of new CDK4-selective inhibitors.

Acknowledgements

This work was supported by Grant No. R02-2002-000-00006-0 from the Basic Research Program of the Korea Science & Engineering Foundation.

Keywords: cyclin-dependent kinase · drug design · inhibitors · molecular dynamics · protein structures · solvent effects

[1] J. W. Harper, P. D. Adams, *Chem. Rev.* **2001**, *101*, 2511 – 2526.
[2] Y. Dai, S. Grant, *Curr. Opin. Pharmacol.* **2003**, *3*, 362 – 370.
[3] D. W. Fry, M. D. Garret, *Curr. Opin. Oncol. Endocr. Metab. Invest. Drugs* **2000**, *2*, 40 – 59.
[4] E. G. Nabel, *Nature Rev. Drug Discov.* **2002**, *1*, 587 – 598.
[5] A. Huwe, R. Mazitschek, A. Giannis, *Angew. Chem.* **2003**, *115*, 2170 – 2187; *Angew. Chem. Int. Ed.* **2003**, *42*, 2122 – 2138.
[6] M. Knockaert, P. Greengard, L. Meijer, *Trends Pharmacol. Sci.* **2002**, *23*, 417 – 425.

[7] I. R. Hardcastle, B. T. Golding, R. J. Griffin, *Annu. Rev. Pharmacol. Toxicol.* **2002**, *42*, 325 – 348.
[8] E. A. Sausville, *Trends Mol. Med.* **2002**, *8*, S32 – S37.
[9] P. M. Fischer, *Curr. Opin. Drug Discov. Dev.* **2001**, *4*, 623 – 634.
[10] T. M. Sielecki, J. F. Boylan, P. A. Benfield, G. L. Trainor, *J. Med. Chem.* **2000**, *43*, 1 – 18.
[11] P. Polychronopoulos, P. Magiatis, A. L. Skaltsounis, V. Myrianthopoulos, E. Mikros, A. Tarricone, A. Musacchio, S. M. Roe, L. Pearl, M. Leost, P. Greengard, L. Meijer, *J. Med. Chem.* **2004**, *47*, 935 – 946.
[12] E. W. Yue, S. V. DiMeo, C. A. Higley, J. A. Markwalder, C. R. Burton, P. A. Benfield, R. H. Grafstrom, S. Cox, J. K. Muckelbauer, A. M. Smallwood, H. Chen, C. H. Chang, G. L. Trainor, S. P. Seitz, *Bioorg. Med. Chem. Lett.* **2004**, *14*, 343 – 346.
[13] C. Sanchez-Martinez, C. Shih, G. Zhu, T. Li, H. B. Brooks, B. K. R. Patel, R. M. Schultz, T. B. DeHahn, C. D. Spencer, S. A. Watkins, C. A. Ogg, E. Considine, J. A. Dempsey, F. Zhang, *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3841 – 3846.
[14] D. C. Kim, Y. R. Lee, B. S. Yang, K. J. Shin, D. J. Kim, B. Y. Chung, K. H. Yoo, *Eur. J. Med. Chem.* **2003**, *38*, 525 – 532.
[15] M. Anderson, J. F. Beattie, G. A. Breault, J. Breed, K. F. Byth, J. D. Culshaw, R. P. A. Ellston, S. Green, C. A. Minshull, R. A. Norman, R. A. Pauptit, J. Stanway, A. P. Thomas, P. J. Jewsbury, *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3021 – 3026.
[16] K. L. Sayle, J. Bentley, F. T. Boyle, A. H. Calvert, Y. Cheng, N. J. Curtin, J. A. Endicott, B. T. Golding, I. R. Hardcastle, P. Jewsbury, V. Mesguiche, D. R. Newell, M. E. M. Noble, R. J. Parsons, D. J. Pratt, L. Z. Wang, R. J. Griffin, *Bioorg. Med. Chem. Lett.* **2003**, *13*, 3079 – 3082.
[17] P. L. Toogood, *Med. Res. Rev.* **2001**, *21*, 487 – 498.
[18] R. Soni, T. O'Reilly, P. Furet, L. Muller, C. Stephan, S. Zumstein-Mecker, H. Fretz, D. Fabbro, B. Chaudhuri, *J. Natl. Cancer Inst.* **2001**, *93*, 436 – 446.
[19] D. W. Fry, D. C. Bedford, P. H. Harvey, A. Fritsch, P. R. Keller, Z. Wu, E. Dobrusin, W. R. Leopold, A. Fattaey, M. D. Garrett, *J. Biol. Chem.* **2001**, *276*, 16617 – 16623.
[20] D. J. Carini, R. F. Kaltenbach III, J. Liu, P. A. Benfield, J. Boylan, M. Boisclair, L. Brizuela, C. R. Burton, S. Cox, R. Grafstrom, B. A. Harrison, K. Harrison, E. Akamike, J. A. Markwalder, Y. Nakano, S. P. Seitz, D. M. Sharp, G. L. Trainor, T. M. Sielecki, *Bioorg. Med. Chem. Lett.* **2001**, *11*, 2209 – 2211.
[21] G. Zhu, S. E. Conner, X. Zhou, C. Shih, T. Li, B. D. Anderson, H. B. Brooks, R. M. Campbell, E. Considine, J. A. Dempsey, M. M. Faul, C. Ogg, B. Patel, R. M. Schultz, C. D. Spencer, B. Teicher, S. A. Watkins, *J. Med. Chem.* **2003**, *46*, 2027 – 2030.
[22] J. F. Beattie, G. A. Breault, R. P. A. Ellston, S. Green, P. J. Jewsbury, C. J. Midgley, R. T. Naven, C. A. Minshull, R. A. Pauptit, J. A. Tucker, J. E. Pease, *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2955 – 2960.
[23] T. A. Engler, K. Furness, S. Malhotra, C. Sanchez-Martinez, C. Shih, W. Xie, G. Zhu, X. Zhou, S. Conner, M. M. Faul, K. A. Sullivan, S. P. Kolis, H. B. Brooks, B. Patel, R. M. Schultz, T. B. DeHahn, K. Kirmani, C. D. Spencer, S. A. Watkins, E. L. Considine, J. A. Dempsey, C. A. Ogg, N. B. Stamm, B. D. Anderson, R. M. Campbell, V. Vasudevan, M. L. Lytle, *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2261 – 2267.
[24] P. D. Jeffrey, A. A. Russo, K. Polyak, E. Gibbs, J. Hurwitz, J. Massague, N. P. Pavletich, *Nature* **1995**, *376*, 313 – 320.
[25] T. G. Davies, 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, V. Mesguiche, D. R. Newell, M. E. M. Noble, J. A. Tucker, L. Wang, H. J. Whitfield, *Nature Struct. Biol.* **2002**, *9*, 745 – 749.
[26] W. J. J. Prize, *J. Mol. Biol.* **1999**, *287*, 821 – 828.
[27] T. G. Davies, P. Tunnah, L. Meijer, D. Marko, G. Eisenbrand, J. A. Endicott, M. E. M. Noble, *Structure* **2001**, *9*, 389 – 397.
[28] S. Y. Wu, I. McNae, G. Kontopidis, S. J. McClue, C. McInnes, K. J. Stewart, S. Wang, D. I. Zheleva, H. Marriage, D. P. Lane, P. Taylor, P. M. Fischer, M. D. Walkinshaw, *Structure* **2003**, *11*, 399 – 410.
[29] T. Honma, *Med. Res. Rev.* **2003**, *23*, 606 – 632.
[30] T. Honma, T. Yoshizumi, N. Hashimoto, K. Hayashi, N. Kawanishi, K. Fukasawa, T. Takaki, C. Ikeura, M. Ikuta, I. Suzuki-Takahashi, T. Hayama, S. Nishimura, H. Morishima, *J. Med. Chem.* **2001**, *44*, 4628 – 4640.
[31] M. Ikuta, K. Kamata, K. Fukasawa, T. Honma, T. Machida, H. Hirai, I. Suzuki-Takahashi, T. Hayama, S. Nishimura, *J. Biol. Chem.* **2001**, *276*, 27548 – 27554.
[32] G. Wu, M. Vieth, *J. Med. Chem.* **2004**, *47*, 3142 – 3148.
[33] A. Cavalli, M. D. Vivo, M. Recanatini, *Chem. Commun.* **2003**, 1308 – 1309.
[34] A. Cavalli, C. Dezi, G. Folkers, L. Scaponzza, M. Recanatini, *Proteins* **2001**, *45*, 478 – 485.
[35] P. A. Sims, C. F. Wong, J. A. McCammon, *J. Med. Chem.* **2003**, *46*, 3314 – 3325.
[36] S. J. Teague, *Nature Rev. Drug Discov.* **2003**, *2*, 527 – 541.
[37] H. A. Carlson, J. A. McCammon, *Mol. Pharmacol.* **2000**, *57*, 213 – 218.
[38] C. A. Carlson, *Curr. Opin. Chem. Biol.* **2002**, *6*, 447 – 452.
[39] J. H. Lin, A. L. Perryman, J. R. Schames, J. A. McCammon, *J. Am. Chem. Soc.* **2002**, *124*, 5632 – 5633.
[40] M. Huse, J. Kuriyan, *Cell* **2002**, *109*, 275 – 282.
[41] Z. Kriz, M. Otyepka, I. Bartova, J. Koca, *Proteins* **2004**, *55*, 258 – 274.
[42] I. Bartova, M. Otyepka, Z. Kriz, J. Koca, *Protein Sci.* **2004**, *13*, 1449 – 1457.
[43] O. Villacanas, J. J. Perez, J. Rubio-Martinez, *J. Biomol. Struct. Dyn.* **2002**, *20*, 347 – 358.
[44] M. Otyepka, Z. Kriz, J. Koca, *J. Biomol. Struct. Dyn.* **2002**, *20*, 141 – 154.
[45] A. Bairoch, R. Apweiler, *Nucl. Acids Res.* **1999**, *27*, 49 – 54.
[46] J. D. Thompson, D. G. Higgins, T. J. Gibson, *Nucl. Acids Res.* **1994**, *22*, 4673 – 4680.
[47] A. Sali, T. L. Blundell, *J. Mol. Biol.* **1993**, *234*, 779 – 815.
[48] A. Fiser, R. K. G. Do, A. Sali, *Protein Sci.*, **2000**, *9*, 1753 – 1773.
[49] G. M. Morris, D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, A. J. Olson, *J. Comput. Chem.* **1998**, *19*, 1639 – 1662.
[50] E. L. Mehler, T. Solmajer, *Protein Eng.* **1991**, *4*, 903 – 910.
[51] P. F. W. Stouten, C. Frömmel, H. Nakamura, C. Sander, *Mol. Simul.* **1993**, *10*, 97 – 120.
[52] D. A. Case, D. A. Pearlman, J. W. Caldwell, T. E. Cheatham, III, W. S. Ross, C. Simmerling, T. Darden, K. M. Merz, Jr., R. V. Stanton, A. Cheng, J. J. Vincent, M. Crowley, V. Tsui, R. Radmer, Y. Duan, J. Pitera, I. Massova, G. L. Seibel, U. C. Singh, P. Weiner, P. A. Kollman, *AMBER 7*, **2002**, San Francisco, University of California.
[53] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **1995**, *117*, 5179 – 5197.
[54] C. A. Bayly, P. Cieplak, W. D. Cornell, P. A. Kollman, *J. Phys. Chem.* **1993**, *97*, 10269 – 10280.
[55] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1993**, *79*, 926 – 935.
[56] H. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, *J. Chem. Phys.* **1984**, *81*, 3684 – 3690.
[57] J. P. Ryckaert, G. Ciccotti, H. C. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327 – 341.
[58] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089 – 10092.
[59] R. A. Laskowski, M. W. MacArthur, D. S. Moss, J. M. Thornton, *J. Appl. Cryst.* **1993**, *26*, 283 – 291.
[60] H. Fan, A. E. Mark, *Protein Sci.* **2004**, *13*, 211 – 220.
[61] Z. Zhu, D. I. Schuster, M. E. Tuckerman, *Biochemistry* **2003**, *42*, 1326 – 1333.
[62] G. A. Jeffrey, *An Introduction to Hydrogen Bonding*, Oxford University Press, Oxford, **1997**.

Received: June 25, 2004
Early View Article
Published online on October 26, 2004