Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry Letters 15 (2005) 2851–2856
Bioorganic & Medicinal Chemistry Letters

Homology modeling and docking study of cyclin-dependent kinase (CDK) 10

Miao Sun, Zesheng Li,* Yuan Zhang, Qingchuan Zheng and Chia-chung Sun

Institute of Theoretical Chemistry, State Key Laboratory of Theoretical and Computational Chemistry, Jilin University, Changchun 130023, PR China

Received 1 February 2005; revised 19 March 2005; accepted 23 March 2005
Available online 29 April 2005

Abstract—In order to understand the mechanisms of ligand binding and the interaction between the ligand and the cyclin-dependent kinase 10 (CDK10), a three-dimensional (3D) model of the CDK10 is generated based on the crystal structure of the cyclin-dependent kinase 2 (CDK2) (PDB code 1AQ1) by using INSIGHTII/Homology module. With the aid of the molecular mechanics and molecular dynamics methods, the last refined model is obtained and is further assessed by PROFILE-3D and PROSTAT, which show that the refined model is reliable. With this model, a flexible docking study is performed and the results indicate that the Lys39 and Asp94 form hydrogen bonds and have strong nonbonding interaction with adenosine 5′-triphosphate (ATP). From the docking studies, we also suggest that the Leu141, Tyr21, and Val24 in CDK10 are three important determinant residues in binding as they have strong nonbonding interaction with ATP. The hydrogen bonding interactions also play an important role for the stability of the complex. Our results may be helpful for further experimental investigations.

1. Introduction

Among the estimated 500–1000 human protein kinases,[1] a family of kinases activated by a family of cyclins, the cyclin-dependent kinases (CDKs), has been extensively studied because of their essential role in the regulation of cell proliferation, of neuronal and thymus functions and of transcription.[2–5] Cell cycle progression is tightly controlled by the activity of CDKs.[6] Ten cyclin-dependent kinases (CDK1–CDK10) are currently known, of which only CDKs1, 2, 3, 4, and 6 intervene directly in the cell cycle, while CDK7 plays only an indirect role as an activator of these CDKs.[7,2] Furthermore, CDK7, CDK8, CDK9 act as regulators of transcription.[8–13] CDK1 controls G2/M transition and CDK2, CDK3, CDK4, and CDK6 are implicated at G1/S.

In 1994, a human novel CDC2–related protein kinase (CDK10) was identified using two independent PCR-based strategies.[14,15] A study by Li et al. using a dominant negative mutant and an antisense vector led to the suggestion that CDK10 was implicated at the G2/M transition of the cell cycle.[16] Three gene isoforms of human CDK10 were shown in 2000.[17] A study by Kasten and Giordano showed that CDK10 interacts with the N-terminus of the Ets2 transcription factor, which contains the highly conserved pointed domain and transactivation domain.[18] CDK10 was identified that it is one of the 19 new genes in nonsmall cell lung cancer.[19]

The function of protein is determined essentially by its corresponding three-dimensional (3D) structure[20] and one of the major obstacles to further elucidate the molecular origin of the observed metabolic profiles of these enzymes is lack of the corresponding 3D structures. The detailed structures of CDK10 and of CDK10–substrate complex remain unknown, and these have hindered the further investigation in CDK10 and cell cycle. Three isoforms of CDK10 have been principally expressed[17] and in the present investigation we construct a 3D model of CDK10 isoform 2 and search for the binding site of substrate. The 3D features of the model are obtained by a homology modeling procedure based on the X-ray diffraction structure of CDK2 (PDB code: 1AQ1)[21] and it is known that the homology model is effective for the 3D structure construction of protein.[22] The obtained result can be used to explain substrate specificity and relate enzyme function to its structure.

Keywords: Cyclin-dependent kinase; Docking; Molecular dynamics; Molecular modeling.
* Corresponding author. Tel.: +86 431 8498960; fax: +86 431 8498026; e-mail: zeshengli@mail.jlu.edu.cn

0960-894X/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2005.03.088

This article describes the modeling of the human CDK10 and CDK10 complexes with adenosine 5′-triphosphate (ATP). The investigation is made to gain further insight into the structural basis for complex of CDK10 with ATP. In order to design novel and higher affinity ligands, an understanding of the interaction between ATP and CDK10 at the molecular level would be valuable.

2. Methods

All simulations are performed on the SGI O3800 workstations using INSIGHTII software package developed by Biosym Technologies. The sequence of CDK10 is obtained from the databank in the National Biomedical Research Foundation (NBRF) of Georgetown University Medical Center (http://pir.georgetown.edu). Molecular minimization (MM) and molecular dynamics (MD) calculations are carried out with the aid of the force field CVFF.

2.1. Homology of CDK10

Homology modeling is usually the method of choice when a clear relationship of homology between the sequence of target protein and at least one known structure is found. This approach would give reasonable results based on the assumption that the tertiary structures of two proteins will be similar if their sequences are related.[23] The homology protein is searched by FASTA program. The sequence identity between CDK10 and reference protein CDK2 (PDB code 1AQ1) is 43.667%. The alignment of CDK10 (target) and CDK2 (template) is shown in Figure 1. Sixteen residues from 298 to 314 at the C-terminus are removed from the CDK10 model, since there is no good template for the fragment and these residues are far away from the catalysis domain of CDK10. For modeling 3D structure of CDK10, the program MODELER[24] was used. MODELER is an implementation of an automated approach to comparative modeling by satisfaction of spatial restraints.[25–27]

The initial structure of modeling is revised by means of refining loops and rotamers, checking bonds and adding hydrogen atoms, and then molecular mechanics (MM) and molecular dynamics (MD) simulations were used to optimize the initial modeling structure. For energy minimization, the 20,000 iterations of steepest descent (SD) calculation were performed and then the conjugated gradient (CG) calculation was carried out until 0.005 kcal mol^{-1} Å^{-1} of convergences on the gradient. After the above simulations, the homology model is obtained by carrying out the molecular dynamics (MD) calculation with the use of Discover3 software package[28] in INSIGHTII. Explicit solvent model TIP3P water is also used. The homology model is solvent with a 10 Å water cap from the center of mass of CDK10. Finally, a conjugate gradient energy minimization of full protein is performed until the root mean-square (RMS) gradient energy is lower than 0.001 kcal mol^{-1} Å^{-1}. In this step, the quality of the initial model is improved.

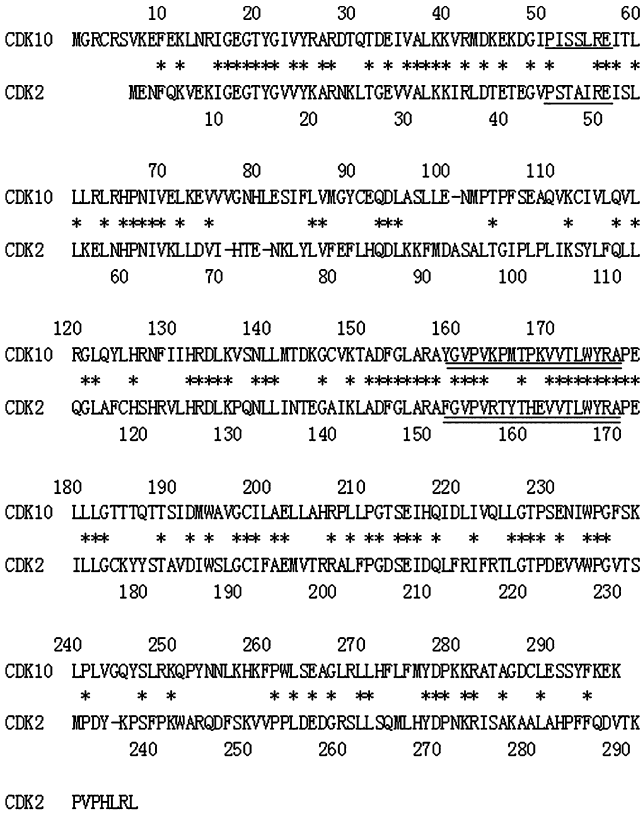

Figure 1. The alignment of 1AQ1 and CDK10.

2.2. Identification of the binding site of CDK10

The Binding-Site module is a suite of programs in INSIGHTII for identifying and characterizing protein binding sites and functional residues from protein. In this study, the ACTIVESITE-SEARCH program was used to search the protein binding site by locating cavities in CDK10 structure, which can be used to guide the protein–ligand docking experiment. Through comparing the conserved residues in CDKs and combining the searched results, we can predict the binding site of CDK10.

2.3. Docking the substrate into the binding site

Molecular docking can fit molecules together in a favorable configuration to form a complex system. The structural information from the theoretically modeled complex may help us to clarify the catalytic mechanism of enzyme. The 3D structure of the ATP was built with the BUILDER program and the geometry was optimized. For the reason of taking the interacting mode of CDK10 with ATP, the advanced docking program Affinity[29] was used to perform the automated molecular docking. The best binding structures of the ligand to the receptor based on the energy of the ligand/receptor complex were automatically found and in this procedure a combination of Monte Carlo type and stimulated annealing procedure to dock a guest molecule with a host was employed. A key feature is that the 'bulk' of the receptor, defined as atoms not in the binding site specified, held rigid during the docking process, while the binding

site atoms and ligand atoms were moveable. The potential of complex was assigned according to the CVFF force field. Nonbonding interactions were used for cell multipole approach. When docking the ATP, Mg^{2+} is considered in the complex. Finally, the docked complexes of CDK10 with ligands were selected according to the criteria of interacting energy combined with geometrical matching quality. These complexes of CDK10 from docking study were used as the starting conformation for further energetic minimization and geometrical optimization before the final models were achieved.

3. Results and discussion

3.1. Homology modeling of CDK10

By means of FASTA program, the sequence of CDK10 was compared to all the known proteins in PDB, and the results showed that 1AQ1 had the best sequence identity (43.667%) with CDK10, so 1AQ1 was used to model the 3D structure of the CDK10. All the side chains of CDK10 were set by AUTO_ROTAMER program which uses the library proposed by Ponder and Richards.[30] In order to remove the steric contacts and get a stable conformation, the energy minimizations of 20,000 iterations and dynamics simulations of 150 ps were performed. The variation of potential energy with time during the 150 ps of molecular dynamics simulations is plotted in Figure 2. From Figure 2, we can see that the potential energy falls rapidly in the first 20 ps, and then it decreases with very low deviation between two steps, and the dynamics simulations tend to equilibrium at 150 ps. Thus, we chose the conformation at 150 ps as the final 3D structure for the further study.

The final structure was further checked through PROFILE-3D and PROSTAT program. The checked results of PROFILE-3D are presented in Figure 3. From Figure 3, we know that all residues are scored positive, which means that these residues are reasonable. The PROSTAT program was used to check the dihedral angles, bond lengths, and bond angles in the structure of CDK10

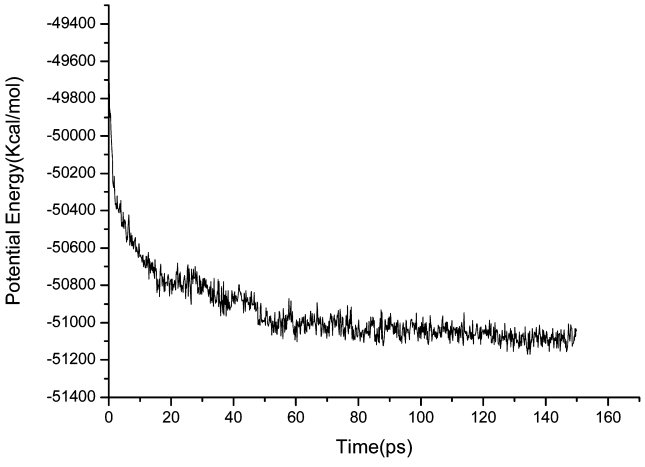

Figure 2. The potential energy as a function of simulations time (ps).

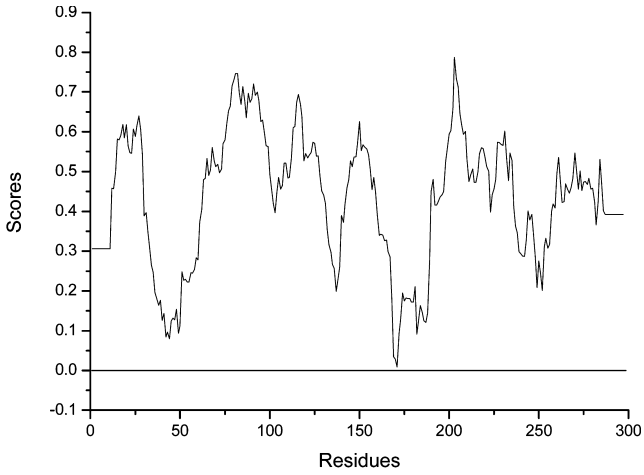

Figure 3. The evaluation of the CDK10 final structure of by PROFILE-3D program.

and 1AQ1, and the main results are listed in Table 1. From Table 1, we can see that the Φ–Ψ values of 87.5% residues in CDK10 fall within the favored regions of the Ramachandran plot, and 85.6% of 1AQ1 locate in the favored regions. From Table 1, we can also see that eight bond angles deviate from the formal values in the 297 residues for CDK10, and for 1AQ1, two bond angles have greater values than the reference values in the 298 residues. The root mean-square deviation (RMSD) of the residue C-α atoms between CDK10 and 1AQ1 is 0.61 Å, which is in the reasonable range. By the comparison of the 3D structure of CDK10 with that of 1AQ1 and the assessments mentioned above, we think that our final 3D structure of CDK10 is reliable.

The final structure of CDK10 is presented in Figure 4, which shows that the final structure includes 10 α-helices and eight β-sheets. It is well known that for the structure of 1AQ1, nine α-helices, and eight β-sheets are involved. From Figure 4, we can see that the structure of CDK10 is typical for the kinase family (i.e., bilobal) and comprises an N-terminal domain of β-sheets and a larger C-terminal domain mainly constituted by α-helices. The N-lobe of CDK10 also contains the PISSLRE helix, which correspond to the PSTALRE helix in CDK2. This helix is another feature found in all CDKs and is important for interaction with cyclins.[31] During the MD simulation, we found that the most variable

Table 1. The results of dihedral angles, bonds, and angles of 1AQ1 and CDK10 checked by PROSTAT program

Target protein	1AQ1	CDK10
%Φ–Ψ Angles in core Ramachandran region	87.5%	85.6%
Number bond distances with significant deviations	0	0
Number bond angles with significant deviations	2	8
Number residues examined	298	297

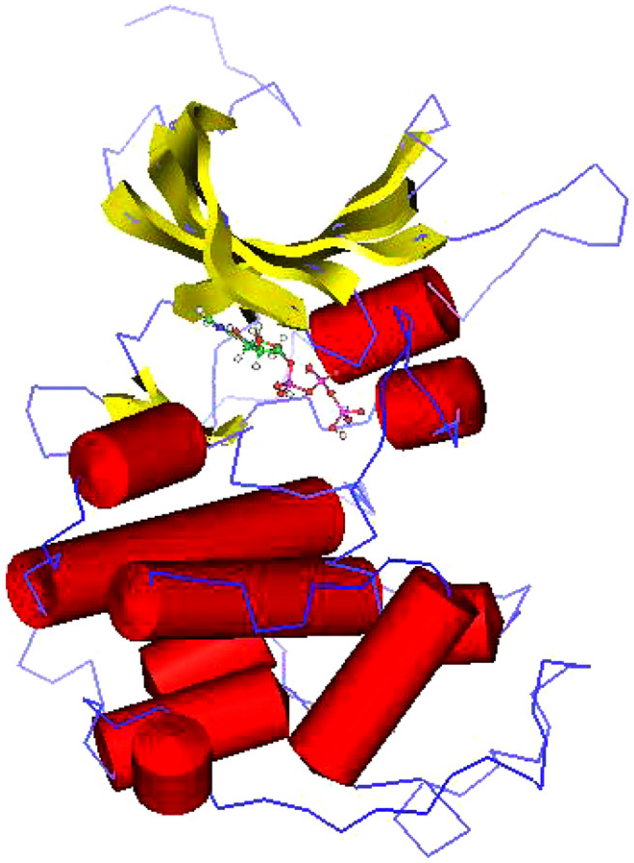

Figure 4. The final structure of CDK10. The helices are colored red and sheets are colored orange.

domain of CDK10 is the T-loop, which shows that this domain of CDKs is the most variable part of these proteins, and it might be a suitable region in which to look for drug selectivity.

3.2. Identification of the binding site of CDK10

The CDK10 is folded into the typical bilobal structure, with the smaller N-terminal domain consisting predominantly of β-sheet structure and the larger C-terminal domain consisting primarily of α-helices. There are no significant differences in the domain orientations between the inhibitor–enzyme complex and the ATP–CDK10 complex. The sequences and structures of CDK10 and CDK2 should be conserved because their main biological functions are similar. Thus, it is predicted that the ATP binds in a manner similar in both structures. Those protein kinases that bind ATP have the classic mononucleotide-binding fold, which contains a β-strand-loop-α-helix supersecondary structure, which is a characteristic feature of the phosphate-binding site.[32,33] In order to calculate the interactions between the active site of CDK10 and the substrates, the ATP-binding pocket was defined as a subset that contains residues in which any atoms are within 6.0 Å from ATP. The binding-site was searched by Binding-Site module, which can be used to guide the protein–ligand docking experiment. The residues comprising ATP pocket of CDK10 are displayed in Figure 5. From Figure 5, we

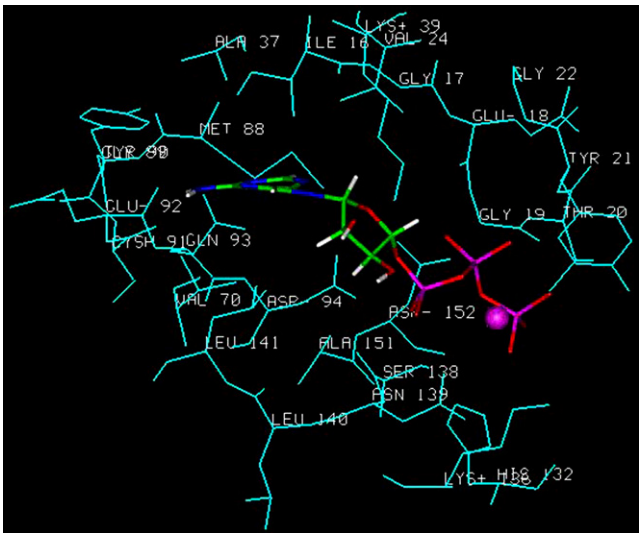

Figure 5. The binding site of CDK10.

can see that the binding pocket has β-strand-loop-α-helix supersecondary structure and an activation loop. The β-strand-loop-α-helix supersecondary structure consists of the residues of Ile16-Gly17-Glu18, Gly22-Ile23-Val24, Lys39, Val41, Leu86, in β-strand, Gly19, Thr20, Tyr21, Val70, Tyr90, Cys91, Ala151 in loop, and Leu155 in α-helix. Other CDKs and protein kinases have similar overall structures and almost superimposable catalytic clefts that include the glycine-rich ATP binding domains and the catalytic base, Lys33.[32] Through the alignment we know that the Lys33 in CDK2 is corresponding to Lys39 in CDK10. Except Gly22, Ile23, and Tyr90, other residues, which composed the binding site in CDK10 are conserved with CDK2.

3.3. Explicit characterization of the complexes

The validated substrate-free CDK10 was used to characterize complexes of this enzyme with typical substrates. The significant compound ATP was selected for study. The phosphorylation and ATP-hydrolysis reactions play important roles in signal transduction and regulation of many proteins, especially protein kinases. The molecular structure of ATP consists of a purine base, a ribose ring, and three phosphate groups. The 3D structure of ATP was built with the BUILDER program and the geometry was further optimized by using DISCOVER_3 program.

Hydrogen bonds play an important role for structure and function of biological molecules, especially for the enzyme catalysis. The hydrogen bonds presented in the complexes are listed in Table 2 and Figure 6. We can see from Table 2 and Figure 6 that there are eight hydrogen bonds between ATP and CDK10. The purine base of ATP has a hydrogen bond with the enzyme between the NH2 group at the position of N7 and side-chain carboxyl of Cys91. The ribose hydroxyl group of ATP forms two hydrogen bonds with the side-chain carboxyl of Asp94. The γ-phosphate groups form three hydrogen bonds with the residues Thr20, His132, and Asn139 of

Table 2. Hydrogen bonds between ATP and CDK10

CDK10		ATP	Hydrogen bond	
Residues	Atom	Atom	Length (Å)	Angle (°)
Cys91	O	Purine base NH	1.95	154.25
Thr20	O	γ-Phosphate O^-	2.14	165.29
His132	OH	γ-Phosphate O^-	2.36	154.47
Asn139	NH	γ-Phosphate O^-	1.88	168.67
Lys39	H	β-Phosphate O	1.68	166.30
Tyr21	H	β-Phosphate O^-	1.78	167.18
Asp94	O	Ribose OH	1.62	157.23
Asp94	O	Ribose OH	1.62	177.03

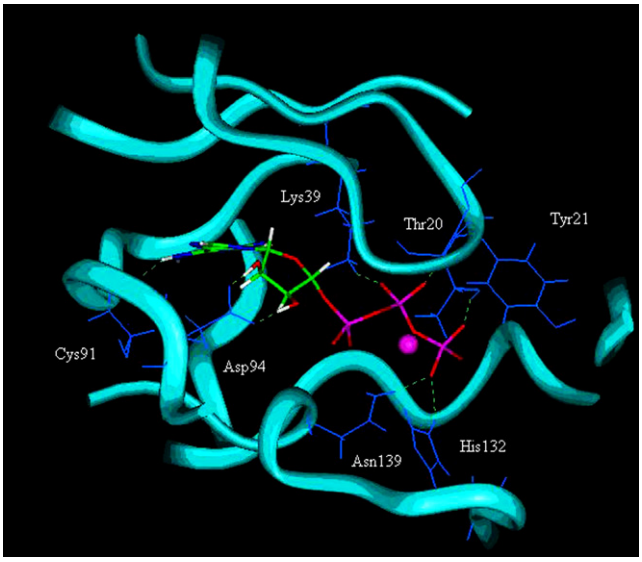

Figure 6. The hydrogen bonding interaction of complex CDK10–ATP.

CDK10, and the β-phosphate group forms two hydrogen bonds with Tyr21 and Lys39.

To determine the key residues that comprise the binding pocket of the model, the interaction energies of the substrates with each individual amino acid in the enzyme were also calculated. Significant binding-site residues in the models were identified by the total interaction energy between the substrates and each amino acid residues in the enzyme. This identification, compared with a definition based on the distance from the substrate, can clearly show the relative significance for every residue. The total interaction energies (lower than −1.0 kcal/mol) between the ATP and each individual amino acid of CDK10 are listed in Table 3. As seen from Table 3, Asp94, Lys39, Leu141, Tyr21, Val24, Tyr20, and Tyr90 have strong interaction energies with ATP. The residue Asp94 has the strongest interaction with ATP and the interaction energy is −35.434 kcal/mol. The residue Lys39 is another key residue, which has strong interaction with ATP, which forms a hydrogen bond with CDK10 and the interaction energy with CDK is −13.466 kcal/mol. Both of the residues Lys39 and Asp94 have strong electrostatic interaction with ATP. They play important roles in the combination between the ATP and CDK10. On the other hand, the

Table 3. The total interaction energy (E_{total}), van der Waals energy (E_{vdw}) and electrostatic energy (E_{ele})

Residues	E_{vdw} (kcal/mol)	E_{ele} (kcal/mol)	E_{total} (kcal/mol)
Asp94	2.199	−37.633	−35.434
Lys39	1.007	−14.473	−13.466
Leu141	−5.176	0.385	−4.791
Tyr21	−3.398	−1.268	−4.666
Val24	−3.496	−0.757	−4.253
Thr20	−2.619	−0.845	−3.464
Tyr90	−2.667	−0.432	−3.099
Ile16	−2.702	−0.166	−2.868
Lys136	−1.324	−1.428	−2.752
Cys91	−1.297	−1.434	−2.731
Ser138	−2.931	0.268	−2.663
Glu92	−1.108	−1.435	−2.543
His132	−2.332	−0.166	−2.498
Gly19	−2.447	0.106	−2.341
Ala37	−2.08	−0.257	−2.337
Met88	−2.184	−0.127	−2.311
Ala151	−1.796	−0.513	−2.309
Asp152	−4.953	2.942	−2.011
Glu18	−1.79	0.135	−1.655
Val70	−1.69	0.059	−1.631
Val161	−1.414	−0.052	−1.466
Asn139	0.516	−1.858	−1.342
Sum	−43.682	−58.949	−102.631

residues of Leu141, Tyr21, and Val24 in CDK10 are three important determinants in binding as they have strong interactions with the ligand. As shown in Tables 2 and 3, these results can serve as a guide to selection of candidate sites for further experimental studies of site-directed mutagenesis.

4. Conclusions

In this work, we have constructed a three-dimensional model of the CDK10 by INSIGHTII/Homology module. After energy minimization and molecular dynamics simulations, this refined model structure is obtained. The last refined model is further assessed by PROFILE-3D and PROSTAT, and the results show that this model is reliable. The stable structure is further used to perform the docking of ATP. Through the docking studies, the model structures of the ligand–receptor complex are obtained. The docking results indicate that conserved amino acid residues in CDK10 play an important role in maintaining a functional conformation and are directly involved in binding to donor and acceptor substrates. The interactions of the CDK10 and ATP proposed in this study are useful to understand the potential mechanisms of the CDK10 and ATP. As is well known, hydrogen bonds play an important role for structure and function of biological molecules, especially for the enzyme catalysis. The residues of Lys39 and Asp94 are important for strong hydrogen bonding interaction with ATP and play a major role in catalysis of CDK10. Furthermore, these residues, as well as the others in Table 2 and in Table 3, are suggested as candidates for further experimental studies of structure–function relationships.

Acknowledgements

This work is supported by the National Science Foundation of China (20333050, 20073014), Doctor Foundation by the Ministry of Education, Foundation for University Key Teacher by the Ministry of Education, Key subject of Science and Technology by the Ministry of Education of China, and Key subject of Science and Technology by Jilin Province.

References and notes

1. Mitch, K.; Jessie, E.; Vincent, M., et al. *Genome Biol.* **2002**, *9*, 1–12.
2. Morgan, D. O. *Annu. Rev. Cell Dev. Biol.* **1997**, *13*, 261–291.
3. *Progress in Cell Cycle Research: Editions "Life in Progress"*; Meijer, L., Jezequel, A., Roberge, M., Eds.; Station Biologique deRoscoff, 2003; Vol. 5, pp 235–248.
4. Vogt, P. K.; Reed, S. I. *Current Topics in Microbiology and Immunology*; Springer, 1998, 169pp.
5. Pavletich, N. P. *J. Mol. Biol.* **1999**, *287*, 821–828.
6. Norbury, C.; Nurse, P. *Annu. Rev. Biochem.* **1992**, *61*, 441–470.
7. Harper, J. W.; Adams, P. D. *Chem. Rev.* **2001**, *101*, 2511–2526.
8. Hengartner, C. J.; Myer, V. E.; Liao, S. M., et al. *Mol. Cell.* **1998**, *2*, 43–53.
9. Serizawa, H.; Makela, T. P.; Conaway, J. W., et al. *Nature* **1995**, *374*, 280–282.
10. Garriga, J.; Peng, J.; Parreno, M., et al. *Oncogene* **1998**, *17*, 3093–3102.
11. Akoulitchev, S.; Chuikov, S.; Reinberg, D. *Nature* **2000**, *407*, 102–106.
12. Wei, P.; Garber, M. E.; Fang, S. M., et al. *Cell* **1998**, *92*, 451–462.
13. Tassan, J. P.; Jaquenoud, M.; Leopold, P., et al. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 8871–8875.
14. Grana, X.; Claudio, P. P.; De Luca, A., et al. *Oncogene* **1994**, *9*, 2097–2103.
15. Brambilla, R.; Draetta, G. *Oncogene* **1994**, *9*, 3037–3041.
16. Li, S.; MacLachlan, T. K.; De Luca, A., et al. *Cancer Res.* **1995**, *55*, 3992–3995.
17. Sergere, J. C.; Thuret, J. Y.; Le Roux, G., et al. *Biochem. Biophys. Res. Commun.* **2000**, *279*, 738–739.
18. Kasten, M.; Giordano, A. *Oncogene* **2001**, *20*, 1832–1838.
19. Singhal, S.; Amin, K. M.; Kruklitis, R., et al. *Cancer Biol. Ther.* **2003**, *2*, 291–298.
20. Lai, L. H. In *Protein Structural Prediction and Molecular Design*; Beijing University Press, 1993; 467, pp 1–15.
21. Lawrie, A. M.; Noble, M. E.; Tunnah, P.; Brown, J. A., et al. *Nat. Struct. Biol.* **1997**, *4*, 796.
22. Luthy, R.; Bowie, J. U.; Eisenberg, D. *Nature* **1992**, *356*, 83–85.
23. Kroemer, R. T.; Doughty, S. W.; Robinson, A. J., et al. *Protein Eng.* **1996**, *9*, 493–498.
24. Sali, A.; Blundell, T. L. *J. Mol. Biol.* **1993**, *234*, 779–815.
25. Sali, A.; Overington, J. P. *Protein Sci.* **1994**, *3*, 1582–1596.
26. Sali, A.; Potterton, L.; Yuan, F., et al. *Proteins* **1995**, *23*, 318–326.
27. Sali, A. *Curr. Opin. Biotechnol.* **1995**, *6*, 437–451.
28. Discover 3 User Guide, San Diego: MSI, USA, 1999.
29. Affinity User Guide, San Diego: MSI, USA, 1999.
30. Ponder, J. W.; Richards, F. M. *J. Mol. Biol.* **1987**, *193*, 775–791.
31. Sharma, P.; Steinbach, P. J.; Sharma, M.; Amin, N. D.; Barchi, J. J.; Pant, H. C. *JBC* **1999**, *274*, 9600–9606.
32. Schulz, G. E.; Schirmer, R. H. *Nature* **1974**, *250*, 142–144.
33. Schulz, G. E.; Schiltz, E.; Tomasselli, A. G.; Frank, R.; Brune, M.; Wittinghofer, A.; Schirmer, R. H. *Eur. J. Biochem.* **1986**, *161*, 127–132.