SPECIAL TOPIC — Soft matter and biological physics

Computational study of non-catalytic T-loop pocket on CDK proteins for drug development*

Huiwen Wang(王慧雯)[1], Kaili Wang(王凯丽)[1], Zeyu Guan(管泽雨)[1], Yiren Jian(简弋人)[2,4], Ya Jia(贾亚)[1], Fatah Kashanchi[3], Chen Zeng(曾辰)[1,2,†], and Yunjie Zhao(赵蕴杰)[1,‡]

[1] *Institute of Biophysics and Department of Physics, Central China Normal University, Wuhan* 430079, *China*
[2] *Department of Physics, The George Washington University, Washington, DC* 20052, *USA*
[3] *George Mason University, Laboratory of Molecular Virology, Manassas, VA* 20110, *USA*
[4] *QM Simulations Inc.*, 4464 *Willow Rd, Pleasanton, CA* 94588, *USA*

(Received 7 August 2017; revised manuscript received 25 September 2017; published online 20 November 2017)

Cyclin-dependent kinases (CDKs) are critical to the cell cycle and many other biological processes, and as such, are considered as one of the promising targets for therapy against cancer and other diseases. Most pan-CDK inhibitors bind to the highly conserved catalytic ATP-binding pocket and therefore lack the specificity to prevent side effects. It is desirable to develop drugs targeting non-catalytic pockets for specificity towards individual CDKs. Here we performed a systematic analysis of non-catalytic pockets on CDKs and identified a region underneath the T-loop, which we term TL pocket, for potential inhibitor development. Specifically, we compared the TL pockets of human CDK2 and CDK7-homolog Pfmrk of Plasmodium falciparum, a malaria-causing parasite. Molecular dynamics simulations of several short peptides revealed that this less conserved TL pocket could be used to design potentially specific inhibitors against malaria disease.

Keywords: cyclin-dependent kinases, non-catalytic, TL pocket, inhibitor design

PACS: 87.14.E−, 87.15.ap, 87.15.Qt, 87.19.X− **DOI:** 10.1088/1674-1056/26/12/128702

1. Introduction

Cyclin-dependent kinases (CDKs) are a family of protein kinases that regulate the cell cycle and many other biological processes.[1] For example, CDK1, 2, 4, and 6 are directly involved in regulating the cell cycle; CDK5 is required for proper development of the brain; CDK7, 8, and 9 are part of the elongation factor for RNA polymerase II transcription.[2–7] Structurally, CDK proteins consist of an N-terminal lobe rich in beta strands, C-helix, alpha helical, and C-terminal lobe. ATP binds to CDKs in a deep cleft between the two lobes. Cyclin binding to CDKs tightly regulates their functions. Aberrant activity of various cell cycle proteins can result in uncontrolled tumor cell proliferation. Therefore, CDK proteins are considered attractive drug targets.[8–12] For example, current therapy against malaria becomes increasingly less effective due to drug resistance in the malaria-causing parasite; it is thus highly desirable to develop new drugs targeting the parasite's CDK required for its life cycle.[13–18]

CDK activation is a two-step process. In the inactive state, T-loop arises from the C-terminal lobe to block the binding of a protein substrate at the entrance of the active site cleft. So, the first step is the binding of the regulatory protein Cyclin to decrease the T-loop flexibility. Second, the Cyclin protein can also position several essential amino-acid side chains correctly, so the phosphates of ATP can be ideally oriented for the kinase reaction. Therefore, previous CDK drug development focused on blocking ATP or breaking up the CDK/Cyclin interface.

Currently, most of the CDK inhibitors are ATP-competitive. Although with some slight differences, the catalytic ATP-binding pockets of CDK family proteins share a similar structure, the pan-CDK inhibitors typically exhibited little specificity towards individual CDKs. For example, flavopiridol is a well-known pan-CDK inhibitor and can inhibit CDK1, 2, 4, and 9.[19–22] Previous study and clinical trials indicated that flavopiridol causes the cell cycle arrest in G1 and G2 phases but also lead to tissue apoptosis or organ atrophy. Another approach for drug design is to break up the CDK/Cyclin complex. However, it is noted that the buried surface at the interface of CDK/Cyclin is extensive with hydrophobic interactions. It is difficult to design a small compound to compete directly with or break up Cyclin protein at the interface. However, recent computational studies suggest that it is possible to design small compounds to break up Cyclin protein at non-catalytic pocket via allosteric interactions.[23–29] Stephane *et al.* identified one non-catalytic pocket away from the ATP site that extends from the DFG region above the C-helix.[30] Crystal structures with ANS re-

*Project supported by the National Natural Science Foundation of China (Grant No. 11704140), the Natural Science Foundation of Hubei Province, China (Grant No. 2017CFB116), the Thousand Talents Plan (Grant No. 31103201603), and the Self-determined Research Funds of CCNU from the Colleges' Basic Research and Operation of MOE 20205170045 to YZ.

†Corresponding author. E-mail: chenz@gwu.edu
‡Corresponding author. E-mail: yjzhaowh@mail.ccnu.edu.cn

© 2017 Chinese Physical Society and IOP Publishing Ltd

vealed that two ANS molecules were bound adjacent to each other. Binding of ANS induced some structural changes in the C-helix conformation that made Cyclin binding impossible. Giulio *et al.* further developed the allosteric inhibitors of cyclin-dependent kinase 2 using this non-catalytic pocket.[31] Previously, we reported some small peptides targeting a non-catalytic TL pocket located under the T-loop of CDK protein to break up the CDK2/Cyclin complex.[32] Computational dynamical network analysis revealed that these peptides weaken the complex via allosteric interactions. Our experiments showed that upon binding to the non-catalytic pocket, these peptides break up the CDK2/Cyclin complex partially and diminish its kinase activity in vitro. Yutong *et al.* further developed the allosteric chemical small-molecule CDK2 inhibitor based on this non-catalytic pocket.[33] These results indicate that the non-catalytic pocket may be utilized as a potential drug target for other CDK proteins. However, the specificity of the non-catalytic pocket towards individual CDKs remains poorly understood.

In this paper, we present a systematic analysis of CDKs to identify all potential non-catalytic pockets. Results on sequence evolution suggest that the non-catalytic TL pocket located under the T-loop is much less conserved in both sequence and structure than the ATP-binding pocket. The case study of the malaria-causing parasite kinase Pfmrk, a sequence homolog of human CDK7, shows that the interface regions of Pfmrk/Cyclin H complex would be weakened differently from that of human CDK2/Cyclin E by certain peptide inhibitors. Taken together, we provide a scheme to identify non-catalytic pockets and further demonstrate that the non-catalytic TL pocket could be used to develop drugs with specificity towards the parasite and reduced side effects on host cells.

2. Materials and methods

2.1. Pocket detection

Potential binding cavities were detected using active site identification program DoGSiteScorer.[34,35] The program identifies all cavities on the surface of a given protein structure and then calculates the global properties of the cavity including its volume, surface, shape and chemical features (see Refs. [34] and [35] for details). DoGSiteScorer[34,35] has correctly classified druggable or undruggable cavities with an accuracy of 90% in a non-redundant data set (NRDD) and accuracy of 88% in a druggability data set (DD). Therefore, this program could reliably discover potential cavities for drug binding. PyMOL was used to visualize the protein structures and their putative cavities (www.pymol.org). The phylogenetic trees were drawn using iTOL.[36–38]

2.2. Conservation analysis

The CDK2, CDK7, and CDK9 structures were extracted from the PDB database (PDB codes: 1FIN, 1UA2, and 3MI9, respectively).[39–41] The parasite CDK Pfmrk and human Cyclin H complex was built via homology modeling by I-TASSER.[42,43]

The homologous sequences of CDK2, CDK7, and CDK9 proteins were extracted from ConSurf-DB.[44,45] The CSI-BLAST is used to search for the homologous sequences similar to the selected structure. Then, the multiple sequence alignment (MSA) of these homologous sequences is calculated by MAFFT.[46] The position-specific conservation scores of each amino acid position in the alignment were computed using the Rate4Site program.[47] This Rate4Site algorithm assigns a conservation level for each residue using an empirical Bayesian inference. Visualization of the conservation patterns on tertiary structure often provides insights on functionally important regions of the protein. Tertiary structural conservation patterns were visualized using PyMOL (www.pymol.org).

2.3. Molecular dynamics simulation and inhibitor evaluation

All molecular dynamics (MD) simulations were performed using the GROMACS software package.[48] We have done a benchmark simulation test in our previous CDK2 study with experimental validation.[32] Here we used the tested MD parameters in our current Pfmrk MD study. The peptide inhibitor structures were exactly the same as described in our previous work.[32] For the MD simulations, the G53a6 force field and SPC water were employed.[49] The temperature was set at 300 K. Before MD simulations, the entire systems were first minimized by a 1000-step steepest descent calculation followed by a 3000-step conjugate gradient optimization. We performed 30-ns MD simulations for each different state. The dynamical protein network is constructed as follows. A node is defined as a single amino acid, and if the distance of any two heavy atoms of a pair of different nodes is less than 4.5 Å for at least 75% of all snapshots sampled during MD simulations, then this pair of nodes was said to form an edge.[50,51] The neighboring nodes in the sequence were not considered to be in contact and thus no edge between them. The final 20 ns of the 30-ns trajectories, sampled every 100 ps, was used to construct the protein network. Moreover, we define the pairwise correlations

$$C_{ij} = \frac{\langle \Delta \boldsymbol{r}_i(t) \cdot \Delta \boldsymbol{r}_j(t) \rangle}{(\langle \Delta \boldsymbol{r}_i(t)^2 \rangle \langle \Delta \boldsymbol{r}_j(t)^2 \rangle)^{1/2}}, \tag{1}$$

where,

$$\Delta \boldsymbol{r}_i(t) = \boldsymbol{r}_i(t) - \langle \boldsymbol{r}_i(t) \rangle \tag{2}$$

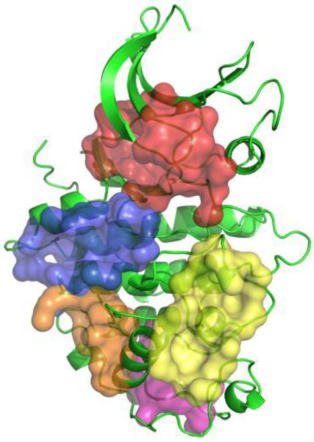
and $\boldsymbol{r}_i(t)$ is the position vector of the C_α atom of the i-th amino acid. The value of C_{ij} ranges from -1 to 1. For the interface correlation analysis, we first identify the CDK/Cyclin interface residues, then calculate the correlations for all pairs of such residues with one on CDK and the other on Cyclin, and finally, take the average of these pair correlations as the interface correlation between CDK and Cyclin. A residue is identified as

interface residue if it is solvent-exposed in either CDK or Cyclin alone but not in CDK/Cyclin complex. Solvent-exposed residues were identified by using the solvent accessible surface recognition program GETAREA.[52]

3. Results

3.1. Potential drug pockets on CDK proteins

Accurate identification of pockets is important in the study of computational drug design, therefore, we identified all the potential drug binding pockets of CDK2, CDK7, and CDK9 using DoGSiteScorer program.[34,35] There are putatively 13, 9, and 15 pockets on CDK2, CDK7, and CDK9 structural surface, respectively. The approximate volumes and surfaces of different pockets are given in Tables 1–3. To illustrate the identified pockets in CDK proteins, we highlight the top 5 pockets of CDK2 structure in Figure 1. Notably, the largest pocket is mainly located around the ATP-binding pocket (colored in red). The second largest one is located under the ATP-binding pocket (colored in blue). The third one is located under the T-loop region (colored in yellow), which is far away from the ATP-binding pocket but near CDK/Cyclin interface. The other two pockets (colored in magenta and orange) are located in the back. There are two similar pockets present in all CDKs: one is the ATP-binding pocket and the other is the pocket underneath the T-loop, which we shall term the TL pocket.



Pockets	Volume/Å^3	Surface/Å^2
Red	651.26	706.35
Blue	443.90	754.79
Yellow	330.69	593.57
Magenta	215.04	395.13
Orange	210.05	318.61

Fig. 1. (color online) The top five detected pockets (shown as surface representation) together with a cartoon representation of CDK2 protein (PDB code: 1FIN). The main shape descriptors (volume and surface) for the top 5 detected pockets are shown in the table on right.

Table 1. The shape descriptors (volume and surface) for all detected pockets on CDK2 structural surface (PDB code: 1FIN). All the pockets are identified using DoGSiteScorer[34,35] program.

Pocket	Volume/Å^3	Surface/Å^2	Pocket	Volume/Å^3	Surface/Å^2
1	651.26	706.35	8	168.38	445.05
2	443.90	754.79	9	154.30	362.75
3	330.69	593.57	10	111.23	262.27
4	215.04	395.13	11	109.95	243.25
5	210.05	318.61	12	104.96	161.33
6	207.04	452.21	13	104.77	243.60
7	181.82	448.15			

Table 2. The shape descriptors (volume and surface) for all detected pockets on CDK7 structural surface (PDB code: 1UA2). All the pockets are identified using DoGSiteScorer[34,35] program.

Pocket	Volume/Å^3	Surface/Å^2	Pocket	Volume/Å^3	Surface/Å^2
1	1733.70	2171.35	6	153.79	378.47
2	337.28	488.53	7	128.77	298.85
3	233.15	393.87	8	126.46	280.29
4	204.54	334.22	9	106.30	178.14
5	182.40	310.82			

Table 3. The shape descriptors (volume and surface) for all detected pockets on CDK9 structural surface (PDB code: 3MI9). All the pockets are identified using DoGSiteScorer[34,35] program.

Pocket	Volume/Å^3	Surface/Å^2	Pocket	Volume/Å^3	Surface/Å^2
1	831.55	1160.48	9	183.44	369.74
2	513.66	671.42	10	181.95	392.95
3	285.71	477.63	11	176.08	451.33
4	250.70	370.25	12	160.32	297.65
5	227.36	271.56	13	117.06	282.35
6	224.98	528.44	14	102.49	86.49
7	222.60	281.99	15	101.97	223.66
8	191.91	345.43			

3.2. Conservation analysis of pockets

Given that protein structures evolve,[53] we performed sequence conservation analysis to infer the structural or functional important residues. The evolutionary conservation scores were identified using the ConSurf-DB.[44,45] The continuous conservation scores are divided into a discrete scale of 9 grades with grade 1 indicating the most variable positions and grade 9 the most conserved positions. The highly conserved residues (grades 7–9), variable residues (grades 1–3), and pocket conservations of CDK proteins are listed in Tables 4–9.

We have projected the evolutionary conservation scores of each amino acid onto the CDK tertiary structures (shown in Fig. 2). For example, most of the beta strands show a pattern of high conservation and form the parallel beta-sheets with 7 or 8 beta strands. Note that these beta strands form the ATP-binding pocket. The high conservation of the beta strands indicates that the ATP-binding pocket is highly conserved. The alpha helix under beta strands located at the CDK/Cyclin interface provides several interactions to stabilize the complex structure. Therefore, the complex requires this helix to be also conserved to form stable interactions. In addition, the amino acids marked in red inside the CDK protein demonstrated the high-level conservation of the protein core. These highly conserved amino acids are important for maintaining the protein tertiary structure. Overall, the biologically functional ATP-binding pocket and CDK/Cyclin interface regions are manifested with a remarkable degree of conservation by the amino acids mostly from grade 7 to 9. On the other hand, other surface residues located at a long distance away from

the CDK/Cyclin interface are variable residues in grades 1–3, especially the residues located under the T-loop.

Previously, we reported some short peptide inhibitors targeting this non-catalytic TL pocket.[32] Molecular dynamics simulations and detailed dynamical network analysis revealed that these peptides weaken the complex formation via allosteric interactions. Our experiments also showed that upon binding to the non-catalytic TL pocket, these peptides break the CDK2/Cyclin complex partially and diminish its kinase activity in vitro. Given that this non-catalytic TL pocket is less conserved, the variation might be exploited to design specific drugs for different CDK proteins.

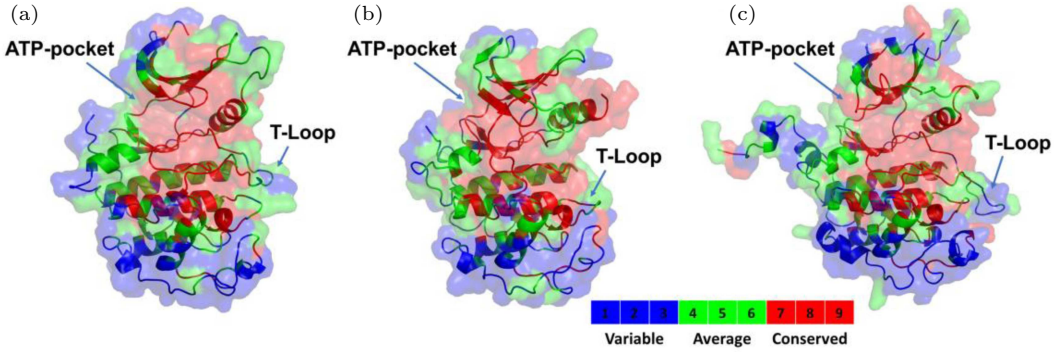


Fig. 2. (color online) Ribbon and surface representation of (a) CDK2 (PDB code: 1FIN), (b) CDK7 (PDB code: 1UA2), and (c) CDK9 (PDB code: 3MI9). Color scheme follows the conservation scores. Variable residues (conservation score from 1 to 3), average residues (conservation score from 4 to 6), and conserved residues (conservation score from 7 to 9) are colored in blue, green and red, respectively. Both the ATP-binding pocket and protein structural core are highly conserved as marked in red. The non-catalytic T-Loop pocket (TL pocket) has more sequence and structural variations.

Table 4. The conservation analysis of CDK2 structure (PDB code: 1FIN). The evolutionary conservation scores are identified using the ConSurf-DB.[44,45] The continuous conservation scores are divided into a discrete scale of 9 grades with grade 1 for the most variable positions while grade 9 the most conserved positions.

Conservation	Residues
1–3	GLN5(1), ARG22(1), LYS24(1), LEU25(1), GLY27(1), GLU28(1), VAL29(1), ASN59(1), LYS65(1), LEU67(1), LEU96(1), THR97(1), GLY98(1), LEU101(1), PRO102(1), GLU224(1), VAL225(1), GLY229(1), THR231(1), SER232(1), PRO238(1), SER239(1), LYS242(1), TRP243(1), ALA244(1), ARG245(1), GLN246(1), ASP247(1), SER249(1), LYS250(1), VAL251(1), PRO254(1), GLU257(1), ASP258(1), SER264(1), GLN265(1), HIS268(1), ASN272(1), LYS278(1), ALA279(1), ALA282(1), GLN287(1), VAL289(1), LYS291(1), VAL293(1), PRO294(1), HIS295(1), LEU296(1), LEU298(1), GLU2(2), THR39(2), GLU73(2), ALA95(2), PHE109(2), ALA140(2), GLY153(2), TYR179(2), ARG200(2), PHE213(2), ARG217(2), VAL226(2), PRO228(2), ASP235(2), LYS237(2), PHE248(2), PRO253(2), ARG260(2), LYS273(2), PRO284(2), ASP288(2), LYS6(3), GLU8(3), PRO61(3), SER94(3), THR137(3), ALA151(3), VAL156(3), ARG157(3), VAL197(3), ASP206(3), THR218(3), ASP223(3), TYR236(3), PHE240(3), ASP256(3), ARG297(3)
4–6	ASN3(4), VAL7(4), TYR19(4), ASP38(4), THR41(4), ASN74(4), LYS75(4), ALA93(4), PRO100(4), LEU103(4), ILE104(4), LEU108(4), ILE104(4), LEU108(4), ALA116(4) , SER120(4), GLU138(4), PHE152(4), THR158(4), ILE186(4), THR198(4), ARG214(4), VAL252(4), LEU255(4), LYS9(5), VAL17(5), ASN23(5), GLU40(5), GLY43(5), VAL44(5), GLU57(5), ILE70(5), HIS71(5), THR72(5), PHE82(5), LYS88(5), LYS89(5), ASP92(5), ILE99(5), SER106(5), TYR107(5), GLN113(5), LEU115(5), PHE117(5), ARG122(5), VAL154(5), CYS177(5), LYS178(5), THR182(5), ALA183(5), PHE193(5), ARG199(5), ILE209(5), LEU219(5), TRP227(5), VAL230(5), PRO241(5), SER261(5), TYR269(5), ILE275(5), PHE285(5), THR290(5), THR26(6), LEU37(6), GLU42(6), SER46(6), LEU54(6), LYS56(6), LEU58(6), LEU76(6), TYR77(6), PHE80(6), HIS84(6), GLN85(6), PHE90(6), MET91(6), LEU111(6), LEU112(6), CYS118(6), HIS121(6), LEU124(6), ILE135(6), PRO155(6), TYR159(6), HIS161(6), LEU189(6), PRO204(6), ASP210(6), LEU212(6), PHE216(6), PRO234(6), GLY259(6), LEU263(6), MET266(6), ASP270(6), SER276(6)
7–9	MET1(7), ILE10(7), GLU12(7), LYS20(7), LYS34(7), ARG36(7), PRO45(7), ILE49(7), ILE52(7), SER53(7), ILE63(7), ASP68(7), VAL69(7), LE78(7), LEU83(7), LYS105(7), VAL123(7), ASN136(7), GLY139(7), ILE141(7), LEU143(7), ILE173(7), LEU175(7), SER181(7), ALA194(7), MET196(7), ALA201(7), LEU202(7), SER207(7), GLU208(7), THR221(7), MET233(7), LEU281(7), PRO292(7), PHE4(8), THR14(8), TYR15(8), ALA21(8), VAL30(8), LEU32(8), ILE35(8), THR47(8), ALA48(8), VAL64(8), LEU66(8), VAL79(8), GLU81(8), LEU87(8), GLY114(8), LEU128(8), GLN131(8), LEU133(8), ALA144(8), GLU162(8), VAL164(8), TRP167(8), ALA170(8), LEU174(8), GLY176(8), VAL184(8), PHE203(8), LEU262(8), LEU267(8), PRO271(8), ALA277(8), ALA280(8), HIS283(8), PHE286(8), GLY11(9), GLY13(9), GLY16(9), VAL18(9), ALA31(9), LYS33(9), ARG50(9), GLU51(9), LEU55(9), HIS60(9), ASN62(9), ASP86(9), GLN110(9), HIS119(9), HIS125(9), ARG126(9), ASP127(9), LYS129(9), PRO130(9), ASN132(9), LEU134(9), LYS142(9), ASP145(9), PHE146(9), GLY147(9), LEU148(9), ALA149(9), ARG150(9), THR160(9), VAL163(9), THR165(9), LEU166(9), TYR168(9), ARG169(9), PRO171(9), GLU172(9), TYR180(9), ASP185(9), TRP187(9), SER188(9), GLY190(9), CYS191(9), ILE192(9), GLU195(9), GLY205(9), GLN211(9), ILE215(9), GLY220(9), PRO222(9), ARG274(9)

Table 5. The conservation analysis of CDK7 structure (PDB code: 1UA2). The evolutionary conservation scores are identified using the ConSurf-DB.[44,45] The continuous conservation scores are divided into a discrete scale of 9 grades with grade 1 for the most variable positions while grade 9 the most conserved positions.

Conservation	Residues
1–3	ARG30(1), ASN33(1), LEU107(1), VAL108(1), PRO111(1), GLU235(1), CYS241(1), THR248(1), SER251(1), PHE252(1), PRO253(1), PRO256(1), HIS258(1), HIS259(1), ILE260(1), ASP266(1), ASP267(1), CYS281(1), THR287(1), LYS291(1), SER296(1), GLN306(1), ARG309(1), ASN311(1), GLU13(2), LYS32(2), ASN35(2), GLN36(2), ILE37(2), SER70(2), GLY76(2), LEU78(2), LYS84(2), SER106(2), SER112(2), MET189(2), THR223(2), GLU234(2), GLN236(2), SER242(2), ASP245(2), TYR246(2), VAL247(2), GLY254(2), ILE255(2), SER262(2), ALA263(2), GLY265(2), LEU269(2), GLN273(2), GLY274(2), LEU277(2), ALA282(2), GLN288(2), ASN297(2), ARG298(2), GLY300(2), PRO303(2), PRO72(3), LEU119(3), GLU147(3), ASN148(3), VAL150(3), ASP216(3), GLU227(3), THR228(3), ASP239(3), PHE249(3), ALA264(3), LYS293(3), THR302(3), CYS305(3)
4–6	LYS14(4), LEU15(4), ASP16(4), GLY82(4), HIS83(4), SER85(4), ASN86(4), ASN105(4), THR110(4), GLU126(4), SER161(4), GLY163(4), ASN166(4), ALA168(4), LEU207(4), VAL210(4), ARG224(4), THR233(4), LYS250(4), LEU257(4), PRO308(4), PRO310(4), TYR27(5), ASP31(5), PHE81(5), LYS103(5), ASP104(5), HIS113(5), ILE114(5), ALA116(5), MET118(5), GLN123(5), TYR127(5), GLN130(5), TRP132(5), PHE162(5), SER164(5), ARG188(5), VAL192(5), MET196(5), LEU208(5), LEU219(5), PHE261(5), ILE284(5), LEU307(5), PHE17(6), THR25(6), THR34(6), LEU65(6), GLN67(6), GLU68(6), LEU69(6), PHE93(6), GLU99(6), VAL100(6), ILE101(6), LEU109(6), TYR117(6), THR121(6), LEU125(6), LEU128(6), HIS131(6), PRO165(6), ARG167(6), ALA187(6), GLY193(6), VAL199(6), LEU203(6), ARG209(6), PRO214(6), ASP220(6), PHE226(6), LEU229(6), MET240(6), ASP270(6), ILE272(6), PHE278(6), THR285(6), TYR294(6), PRO299(6), GLY304(6)
7–9	LEU18(7), GLU20(7), LYS28(7), ARG57(7), LEU60(7), LYS64(7), ILE74(7), ASP79(7), ALA80(7), ILE87(7), SER88(7), LEU89(7), PHE91(7), MET94(7), GLU95(7), THR96(7), ILE102(7), LYS115(7), LEU122(7), ILE133(7), LEU134(7), LEU145(7), ASP146(7), GLY149(7), LEU151(7), TYR169(7), HIS171(7), GLY191(7), LEU206(7), PHE212(7), SER217(7), ASP218(7), LEU222(7), THR231(7), PRO238(7), LEU243(7), PRO244(7), LEU268(7), LEU275(7), ASN279(7), LEU290(7), PRO301(7), PHE23(8), ALA29(8), ILE40(8), LYS42(8), ILE43(8), ASN56(8), THR58(8), ALA59(8), ILE63(8), ILE75(8), LEU77(8), VAL90(8), ASP92(8), LEU98(8), GLY124(8), LEU138(8), ASN141(8), LEU143(8), LEU153(8), ALA154(8), GLN172(8), TRP177(8), LEU183(8), LEU184(8), PHE185(8), GLY186(8), VAL194(8), ALA204(8), PRO211(8), LEU213(8), TRP237(8), LEU271(8), PHE276(8), ALA286(8), ALA289(8), MET292(8), PHE295(8), GLY19(9), GLY21(9), GLN22(9), ALA24(9), VAL26(9), VAL38(9), ALA39(9), LYS41(9), ARG61(9), GLU62(9), LEU66(9), HIS71(9), ASN73(9), ASP97(9), MET120(9), HIS129(9), HIS135(9), ARG136(9), ASP137(9), LYS139(9), PRO140(9), ASN142(9), LEU144(9), LYS152(9), ASP155(9), PHE156(9), GLY157(9), LEU158(9), ALA159(9), LYS160(9), TPO170(9), VAL173(9), VAL174(9), THR175(9), ARG176(9), TYR178(9), ARG179(9), ALA180(9), PRO181(9), GLU182(9), TYR190(9), ASP195(9), TRP197(9), ALA198(9), GLY200(9), CYS201(9), ILE202(9), GLU205(9), GLY215(9), GLN221(9), ILE225(9), GLY230(9), PRO232(9), PRO280(9), ARG283(9)

Table 6. The conservation analysis of CDK9 structure (PDB code: 3MI9). The evolutionary conservation scores are identified using the ConSurf-DB.[44,45] The continuous conservation scores are divided into a discrete scale of 9 grades with grade 1 for the most variable positions while grade 9 the most conserved positions.

Conservation	Residues
1–3	ARG37(1), LYS40(1), GLY42(1), LEU118(1), LYS120(1), ASN179(1), GLU251(1), ASP257(1), LEU261(1), LEU265(1), GLU266(1), LEU267(1), VAL268(1), LYS269(1), GLY270(1), GLN271(1), LYS276(1), ASP277(1), LYS280(1), ALA281(1), TYR282(1), ALA301(1), ASP307(1), ASN311(1), TRP316(1), ASP323(1), LEU324(1), LYS325(1), GLY326(1), MET327(1), LEU328(1), SER329(1), THR330(1), HIS331(1), LEU332(1), MET335(1), PHE12(2), SER17(2), LYS18(2), GLU20(2), ARG39(2), GLN43(2), LYS44(2), LYS74(2), ASN80V, ILE82(2), VAL119(2), LEU123(2), SER124(2), GLN131(2), ALA173(2), ASP205(2), SER226(2), ALA239(2), GLN243(2), PRO250(2), VAL252(2), ASN255(2), ASN258(2), TYR262(2), GLU263(2), LYS264(2), VAL275(2), ARG278(2), ASP285(2), TYR287(2), LEU289(2), ASP293(2), LYS294(2), VAL297(2), GLN302(2), ASP308(2), ASP313(2), SER317(2), ASP318(2), MET320(2), PRO342(2), GLU9(3), CYS10(3), GLU15(3), LYS21(3), GLU76(3), THR87(3), GLY97(3), VAL162(3), SER175(3), LYS178(3), PRO182(3), ASN183(3), ARG184(3), ASN232(3), LEU244(3), THR249(3), ARG273(3), LYS274(3), LEU279(3), VAL283(3), PRO286(3), GLU337(3), TYR338(3), PRO341(3)
4–6	VAL8(4), LEU22(4), ALA23(4), ASN54(4), GLY58(4), ARG86(4), SER98(4), VAL117(4), THR122(4), ILE126(4), TYR138(4), ARG159(4), ASP160(4), PHE174(4), LEU176(4), TRP223(4), THR224(4), LEU240(4), ARG284(4), SER322(4), PHE336(4), ALA340(4), PRO11(5), LYS24(5), PHE34(5), HIS38(5), MET52(5), GLU53(5), LYS56(5), PHE59(5), ALA111(5), SER115(5), ASN116(5), GLU125(5), ARG128(5), MET130(5), ASN135(5), ARG142(5), LYS144(5), ALA177(5), GLN181(5), ARG204(5), PRO208(5), LEU212(5), ARG225(5), GLN235(5), LYS272(5), ASP290(5), LEU298(5), THR333(5), SER334(5), LEU339(5), ASP14(6), VAL16(6), GLU32(6), THR41(6), LEU51(6), LEU64(6), ILE69(6), GLN71(6), LEU72(6), LEU73(6), CYS85(6), ILE99(6), LEU101(6), PHE105(6), GLU107(6), HIS108(6), GLY112(6), LEU113(6), PHE121(6), VAL129(6), LEU133(6), LEU137(6), TYR139(6), ILE140(6), ILE157(6), SER180(6), ASN187(6), GLU203(6), GLY207(6), PRO209(6), ALA215(6), MET219(6), MET222(6), GLN230(6), HIS236(6), SER242(6), CYS245(6), VAL256(6), ILE292(6), ILE304(6), ASP305(6), PHE314(6), PRO319(6), ARG344(6)
7–9	CYS13(7), ILE25(7), GLN27(7), PHE30(7), LYS35(7), GLU55(7), GLU57(7), ILE61(7), LYS68(7), VAL78(7), GLU83(7), ILE84(7), LYS88(7), TYR100(7), PHE103(7), CYS106(7), LEU114(7), LYS127(7), LEU134(7), ASN143(7), ILE145(7), LEU146(7), THR158(7), GLY161(7), LEU163(7), TYR185(7), LEU199(7), LEU201(7), PRO227(7), ILE228(7), THR233(7), GLU234(7), LEU238(7), SER247(7), TRP253(7), PRO254(7), TYR259(7), GLU260(7), ALA288(7), LEU295(7), ASP299(7), LEU310(7), PHE315(7), PRO321(7), ARG343(7), TYR19(8), ALA36(8), LEU47(8), LYS49(8), VAL50(8), PRO60(8), THR62(8), ALA63(8), ILE67(8), VAL79(8), LEU81(8), VAL102(8), ASP104(8), GLY136(8), MET150(8), ALA153(8), VAL155(8), LEU165(8), ALA166(8), ARG188(8), VAL190(8), TRP193(8), PRO196(8), LEU200(8), GLY202(8), ILE210(8), CYS217(8), ALA220(8), MET229(8), LEU291(8), PRO300(8), SER306(8), ALA309(8), HIS312(8), GLY26(9), GLY28(9), THR29(9), GLY31(9), VAL33(9), VAL45(9), ALA46(9), LYS48(9), ARG65(9), GLU66(9), LEU70(9), HIS75(9), ASN77(9), ASP109(9), LEU110(9), MET132(9), HIS141(9), HIS147(9), ARG148(9), ASP149(9), LYS151(9), ALA152(9), ASN154(9), LEU156(9), LYS164(9), ASP167(9), PHE168(9), GLY169(9), LEU170(9), ALA171(9), ARG172(9), TPO186(9), VAL189(9), THR191(9), LEU192(9), TYR194(9), ARG195(9), PRO197(9), GLU198(9), TYR206(9), ASP211(9), TRP213(9), GLY214(9), GLY216(9), ILE218(9), GLU221(9), GLY231(9), GLN237(9), ILE241(9), GLY246(9), ILE248(9), LEU296(9), ARG303(9)

Table 7. The pocket conservation score of CDK2 structure (PDB code: 1FIN). The first and second columns represent the average conservation score of the pocket and all residues in the pocket.

Pocket	Residues
1 (7.9 ± 1.2)	ILE10(7), GLU12(7), GLY13(9), THR14(8), TYR15(8), GLY16(9), VAL17(5), VAL18(9), ALA31(9), LEU32(8), LYS33(9), LYS34(7), ILE35(8), THR47(8), GLU51(9), LEU55(9), ILE63(7), VAL64(8), PHE80(6), GLU81(8), PHE82(5), LEU83(7), HIS84(6), GLN85(6), GLN131(8), ASN132(9), LEU134(9), ALA144(8), ASP145(9), PHE146(9), GLY147(9), LEU148(9)
2 (4.9 ± 2.9)	LEU67(1), LYS88(5), MET91(6), ASP92(5), SER94(3), ALA95(2), THR97(1), GLY98(1), ILE99(5), PRU100(4), LEU101(1), ILE104(4), LYS129(9), PRO130(9), GLN131(8), THR165(9), TRP167(8), TYR168(9), GLU195(9), MET196(7), VAL197(3), THR198(4), ARG199(5), ARG200(2), ALA201(7), PRO254(1)
3 (6.3 ± 2.4)	THR158(4), TYR159(6), THR160(9), GLU162(8), VAL163(9), ARG169(9), ILE173(7), LEU174(8), LEU175(7), GLY176(8), CYS177(5), TYR180(9), GLU208(7), ILE209(5), LEU212(6), ASP235(2), LYS237(2), PHE240(3)
4 (4.5 ± 3.2)	LEU219(5), GLY220(9), THR221(7), PRO222(9), ASP223(3), VAL226(2), TRP243(1), ARG245(1), SER264(1), LEU267(8), HIS268(1), TYR269(5), ASP270(6)
5 (4.9 ± 2.6)	ALA194(7), GLU195(9), THR198(4), ARG200(2), ALA201(7), LEU202(7), PHE203(8), PRO204(6), ARG214(4), ARG217(2), THR218(3), VAL251(1), VAL252(4)
6 (5.3 ± 2.6)	LEU115(5), ALA116(4), HIS119(9), SER120(4), THR182(5), ALA183(5), ILE186(4), PRO271(8), ASN272(1), ARG274(9), ILE275(5), SER276(6), ALA277(8), LYS278(1)
7 (3.9 ± 2.0)	LEU101(1), PRO102(1), ILE104(4), LYS105(7), SER106(5), LEU108(4), PHE193(5), VAL197(3), LEU255(4), ASP256(3), GLY259(6), PHE285(5), ASP288(2), VAL289(1), THR290(5), PRO292(7)
8 (5.9 ± 2.9)	ILE52(7), LEU55(9), LYS56(6), GLU57(5), LEU58(6), ASN59(1), HIS60(9), ILE63(7), VAL64(8), LYS65(1), LEU66(8), LEU67(1), ASP68(7), VAL69(7)
9 (6.6 ± 1.4)	GLU12(7), TYR15(8), GLY16(9), VAL17(5), LYS34(7), ILE35(8), ARG36(7), LEU37(6), ASP38(4), GLU42(6), GLY43(5), PRO45(7)
10 (2.8 ± 1.8)	PHE90(6), ALA93(4), SER94(3), THR97(1), GLY98(1), ILE99(5), PRO100(4), LEU103(4), VAL293(1), PRO294(1), HIS295(1), ARG297(3)
11 (5.6 ± 1.8)	LEU124(6), PHE152(4), VAL154(5), PRO155(6), VAL156(3), TYR180(9), SER181(7), THR182(5)
12 (6.6 ± 2.4)	GLU172(9), CYS177(5), TYR179(2), TYR180(9), SER181(7), ALA183(5), VAL184(8), PRO271(8)
13 (6.1 ± 2.0)	MET1(7), PHE4(8), LYS6(3), TYR19(4), LEU32(8), LYS34(7), TYR77(6)

Table 8. The pocket conservation of CDK7 structure (PDB code: 1UA2). The first and second columns represent the average conservation score of the pocket and all residues in the pocket.

Pocket	Residues
1 (7.8 ± 1.5)	LEU18(7), GLU20(7), GLY21(9), GLN22(9), PHE23(8), ALA24(9), VAL26(9), ALA39(9), LYS41(9), GLU62(9), LEU65(6), ILE75(8), PHE91(7), ASP92(8), PHE93(6), MET94(7), GLU95(7), THR96(7), ASP97(9), GLU99(6), ILE102(7), LYS103(5), ASP104(5), TRP132(5), ILE133(7), LEU134(7), HIS135(9), ARG136(9), ASP137(9), LEU138(8), LYS139(9), PRO140(9), ASN141(8), ASN142(9), LEU144(9), ALA154(8), ASP155(9), PHE156(9), GLY157(9), LEU158(9), LYS160(9), SER161(4), PHE162(5), ARG167(6), ALA168(4), TYR169(7), HIS171(7), GLN172(8), VAL173(9), VAL174(9), THR175(9), ARG176(9), TRP177(8), TYR178(9), ARG179(9), ALA180(9), LEU183(8), VAL194(8), ASP195(9), TRP197(9), ALA198(9), CYS201(9), GLU205(9), ARG209(6), VAL210(4), PRO211(8), PHE212(7), LEU213(8), PRO214(6), GLN221(9), ILE225(9)
2 (3.8 ± 2.5)	THR223(2), LEU229(6), GLN236(2), TRP237(8), GLY254(2), ILE255(2), PRO256(1), LEU257(4), HIS258(1), LEU269(2), ASP270(6), ILE272(6), GLN273(2), PHE276(8), LEU277(2), PHE278(6), ASN279(7), ALA282(2)
3 (5.6 ± 2.6)	GLU182(9), ALA187(6), ARG188(5), MET189(2), TRP237(8), PRO238(7), ASP239(3), MET240(6), SER242(2), LEU243(7), PRO244(7), ASN279(7), PRO280(9), CYS281(1)
4 (3.7 ± 2.3)	PRO111(1), SER112(2), ILE114(5), LYS115(7), LEU207(4), ALA263(2), ALA264(3), GLY265(2), ASP267(1), LEU268(7), TYR294(6), ASN297(2), PRO299(6)
5 (5.5 ± 2.3)	GLU13(2), LYS14(4), TYR27(5), LYS28(7), ALA29(8), ASP31(5), LYS32(2), ILE40(8), LYS42(8), PHE81(5), SER88(7)
6 (5.7 ± 2.3)	GLU126(4), HIS129(9), GLN130(5), TRP132(5), ILE133(7), LEU134(7), VAL192(5), THR285(6), ALA286(8), THR287(1)
7 (4.5 ± 2.9)	PHE185(8), PHE226(6), THR231(7), PRO232(9), THR233(4), GLU234(2), TRP237(8), MET240(6), CYS241(1), TYR246(2), VAL247(2), THR248(1), PHE249(3)
8 (5.3 ± 2.1)	GLU95(7), THR96(7), ASP97(9), ILE101(6), TYR117(6), LEU145(7), ASP146(7), GLU147(3), ASP148(3), GLY149(7), GLY304(6), CYS305(3), LEU307(5), PRO308(4), ARG309(1), PRO310(4)
9 (5.2 ± 2.3)	ALA116(5), LEU119(3), MET120(9), GLY149(7), PRO301(7), THR302(3), PRO303(2), GLY304(6), LEU307(5)

Table 9. The pocket conservation of CDK9 structure (PDB code: 3MI9). The first and second columns represent the average conservation score of the pocket and all residues in the pocket.

Pocket	Residues
1 (6.9 ± 2.2)	LYS24(5), ILE25(7), VAL33(9), LYS35(7), VAL45(9), ALA46(9), LEU47(8), LYS48(9), GLU66(9), ILE69(6), LEU70(9), GLN71(6), LEU73(6), LYS74(2), HIS75(9), GLU76(3), VAL78(7), VAL79(8), ASN80(2), LEU81(8), ILE82(2), GLU83(7), ILE84(7), LEU101(6), PHE103(7), ASP104(8), PHE105(6), CYS106(7), GLU107(6), HIS108(6), ASP109(9), GLY112(6), LEU113(6), ASN116(5), ALA153(8), ASN154(9), LEU156(9), THR158(7), LYS164(9), ALA166(8), ASP167(9), PHE168(9), GLY169(9), HIS331(1)
2 (5.3 ± 2.4)	PHE174(4), SER175(3), LEU176(4), ALA177(5), LYS178(4), GLN181(5), PRO182(3), ASN183(3), TYR185(7), GLU198(9), GLU203(6), ASP205(2), TYR206(9), GLY207(6), PRO208(5), PRO209(6), ILE210(8), TRP253(7), PRO254(7), ASN255(2), VAL256(6), ASN258(2), TYR259(7), ASP299(7), PRO300(8), ALA301(1)
3 (5.4 ± 2.4)	LEU137(6), TYR138(4), HIS141(9), ARG142(5), PHE174(4), LEU176(4), PRO208(5), PRO209(6), LEU212(5), PRO300(8), ALA301(1), ARG303(9), ILE304(6), ASP305(6), SER306(8), ASP307(1)

Table 9. Continued.

Pocket	Residues
4 (7.8 ± 1.1)	GLU32(6), VAL33(9), LYS48(9), LYS49(8), VAL50(8), LEU51(6), PRO60(8), THR62(8), ALA63(8), GLU66(9), ILE67(8), LEU101(6), GLY169(9)
5 (4.4 ± 2.9)	CYS245(6), GLY246(9), SER247(7), THR249(3), GLU251(1), VAL252(2), LYS269(1), GLN271(1), LYS272(5), ARG273(3), LEU296(9), VAL297(2), LEU298(5), ASP299(7)
6 (3.8 ± 2.7)	GLU107(6), HIS108(6), ARG128(5), MET132(9), ARG159(4), ASP160(4), GLY161(7), VAL162(3), ASP323(1), LEU324(1), LYS325(1), LEU328(1), HIS331(1)
7 (3.6 ± 1.9)	LEU114(7), VAL119(2), LYS120(1), PHE121(6), THR122(4), LEU123(2), ILE126(4), MET222(6), TRP223(4), ARG225(5), TYR282(1), VAL283(3), ARG284(4), ASP285(2)
8 (7.2 ± 2.2)	LEU110(9), ALA111(5), LEU114(7), SER115(5), LYS151(9), ALA152(9), ALA153(8), THR191(9), TRP193(8), TYR194(9), GLU221(9), ARG225(5), SER226(2), PRO227(7)
9 (4.7 ± 2.9)	ALA220(8), THR224(4), SER226(2), PRO227(7), ILE228(7), MET229(8), LEU244(3), VAL275(2), ARG278(2), LEU279(3), TYR282(1), LEU296(9)
10 (6.6 ± 2.4)	ARG65(9), LYS68(7), ILE69(6), LEU72(6), ILE145(7), ARG172(9), ALA173(2)
11 (8.2 ± 1.3)	ILE61(7), THR62(8), ARG65(9), GLU66(9), ARG148(9), PHE168(9), GLY169(9), LEU170(9), ALA171(9), ARG172(9), TYR185(7), ASN187(6), ARG188(8), VAL189(9), ARG204(5), TYR206(9)
12 (5.4 ± 1.8)	LYS35(7), LYS44(2), PHE105(6), CYS106(7), GLU107(6), THR158(7), ARG159(4), ASP160(4)
13 (2.9 ± 2.8)	SER247(7), ILE248(9), THR249(3), PRO250(2), TYR262(2), GLU263(2), LEU265(1), LEU267(1), VAL268(1), LYS269(1)
14 (8.4 ± 0.7)	LEU192(9), TRP193(8), TYR194(9), ARG195(9), PRO196(8), LEU200(8), TRP213(9), CYS217(8), ILE218(9), GLU221(9), PRO227(7), ILE228(7), MET229(8), GLN237(9), ILE241(9)
15 (7.0 ± 2.4)	ASN187(6), ARG188(8), VAL189(9), ARG195(9), LEU199(7), LEU200(8), GLY202(8), GLU234(7), LEU261(1)

3.3. Specific peptide inhibitors for the non-catalytic TL pocket on Pfmrk for potential malaria therapy

We can explore the variability of non-catalytic binding pockets of CDKs to not only reduce cross interactions for inhibitors against different CDKs of the same organism but also potentially design inhibitors against a specific parasitic organism while leaving the host organism intact. To this end, we focus on CDK7-homolog Pfmrk from Plasmodium falciparum, a malaria-causing parasite that kills a significant number of lives in the world, especially in the developing countries.[54] Given the emergence and spread of drug-resistant malaria, new drugs against malaria are urgently needed. Pfmrk is an attractive drug target due to its role in regulating the parasite's cellular proliferation. It was reported that Pfmrk forms a stable complex with human Cyclin H and stimulates kinase activity.[55] We thus aim to design inhibitors targeting the non-catalytic TL pocket that can break Pfmrk/Cyclin H complex while leaving CDK2/Cyclin E of the host human cell unaffected.

To visualize the variability of the ATP-binding pocket and the non-catalytic TL pocket between human and parasite Plasmodium, we constructed the phylogenetic tree of these two pockets as shown in Fig. 3 with human CDK2 and Pfmrk of Plasmodium falciparum highlighted in red and blue, respectively. Indeed, the separation displayed in Fig. 3 indicates that the non-catalytic TL pocket possesses more variability between CDK2 and Pfmrk.

It is known that Cyclin binding to CDK modulates the enzymatic kinase activity of CDK. A stable interface of CDK/Cyclin is thus required. Weakening the interface could be a strategy for inhibitor design. Indeed, it was the approach used in our previous studies[32] where MD simulations were performed to monitor the interface motion caused by peptide inhibitors placed in the TL pocket. The interface stability is measured by the interface correlation via dynamical network analysis (see Methods for details). Larger interface correlation indicates stronger interface coupling and thus more stable complex. As shown on the top panel of Fig. 5, the interface correlation of CDK2/Cyclin E was considerably weakened by peptides DAALT and YAALQ. These computational results were indeed consistent with the experimental observation that DAALT and YAALQ partially disrupted the CDK2/Cyclin E complex formation and reduced the kinase activity.[32]

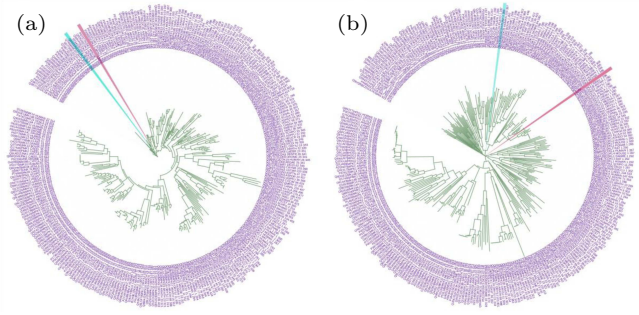


Fig. 3. (color online) Phylogenetic tree of the ATP-binding pocket (a) and the non-catalytic T-loop (TL) pocket (b). Human CDK2 and parasite CDK Pfmrk are highlighted in red and blue. The non-catalytic TL pocket is much less conserved in both sequence and structure than ATP-binding pocket.

Previous research showed that highly coupled residues with common secondary structure elements will produce a correlation value great than 0.7, while some other secondary structure elements with strong interactions can also achieve correlation value around $0.5 \sim 0.6$. Residues across an interface typically result in correlations ranging from 0.3 to 0.4.[32] With the experimental validation of our computational studies on the interface stability of human CDK2/Cyclin E complex, which showed that a decreased interface correlation corresponds to weakened interface stability,[32] we further probed how these same peptide inhibitors might impact the interface

stability of Plasmodium Pfmrk/Cyclin H complex. To this end, we first built the Pfmrk/Cyclin H complex via homology modeling as shown in Fig. 4 (see Methods). We then carried out MD simulations and dynamical network analysis on Pfmrk/Cyclin H complex. Changes in the interface correlation due to the peptide inhibitors located at the TL pocket are shown on the bottom panel of Fig. 5. The average correlation value of the interface residues in the absence of peptide is 0.36, compared with values of 0.34, 0.32, 0.34, 0.31, 0.34, 0.31 in the presence of peptides DAALT, YAALQ, RAALG, FAALA, KAALE, and RAALW, respectively. The interface correlation in the presence of RAALG, FAALA, and RAALW peptides were higher than control in CDK2/Cyclin but lower than control in Pfmrk/Cyclin. These results suggest that RAALG, FAALA, and RAALW should disrupt Pfmrk/Cyclin interface while keeping CDK2/Cyclin intact and thus become specific towards inhibiting malaria Pfmark. Since FAALA and RAALW achieved the lowest interface correlations in Pfmrk simulations, they are therefore the best candidates among the five peptides studied to partially disrupt the malarial Pfmrk/Cyclin complex formation and reduce its kinase activity. Therefore, taken together with the previous results on CDK2/Cyclin E complex,[32] these computational results appear to suggest that peptides FAALA or RAALW could be inhibitors more specific towards the parasitic Pfmrk.

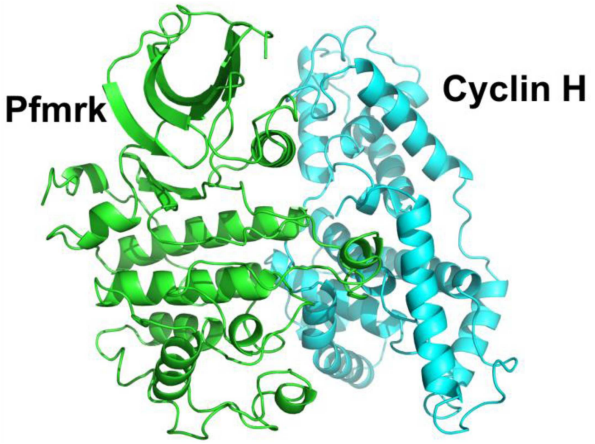


Fig. 4. (color online) Ribbon representation of the parasite CDK Pfmrk and human Cyclin H complex. The structure was built via homology modeling by I-TASSER using human CDK7 and Cyclin H complex. Pfmrk is colored in green. Cyclin H is colored in light blue.

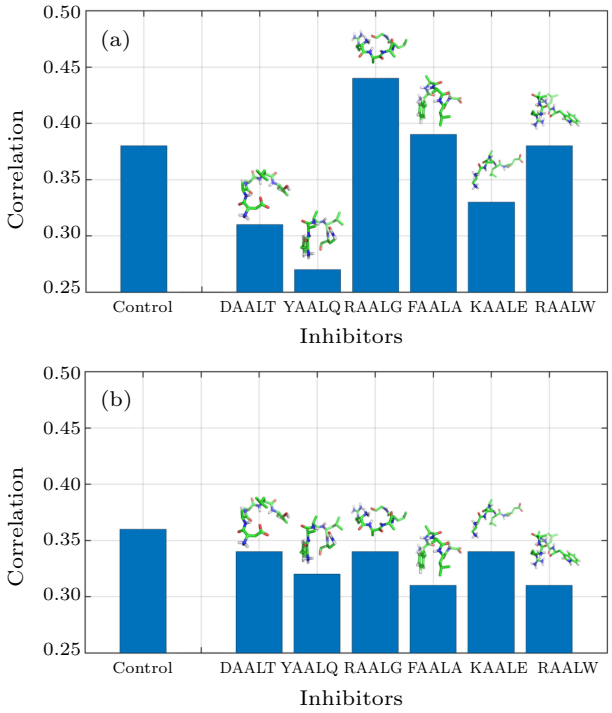


Fig. 5. (color online) Correlation strength of CDK/Cyclin interface with and without 5mer peptides located at the TL pocket of CDK2/Cyclin E (a) and Pfmark/Cyclin H (b). The correlation was computed by the dynamical network analysis of MD simulations. The most decreased interface correlations were produced by DAALT and YAALQ on CDK2/Cyclin but FAALA and RAALW on Pfmrk/Cyclin showing specificity to some degree. The computational results for CDK2/Cyclin E are consistent with previous experiments (refer to the main text).

4. Discussion

Cross-interaction due to the lack of specificity is a common problem in drug design. Targeting less-conserved non-catalytic residues offers a further opportunity to design drugs with desired specificity to reduce the risk of side effects. This approach could be applied to other kinase proteins. Some studies have extended this approach to non-kinase proteins. For example, Hagel *et al.* developed non-catalytic initiators for Hepatitis C virus (HCV) protease.[56] In this study, we performed a systematic analysis of CDK sequences and identified all potential binding pockets with a particular focus on the non-catalytic TL pocket and its potential for designing a new class of inhibitors distinct from the traditional APT-competitive inhibitors. We demonstrated, as an example, that this variable TL pocket could indeed be explored to design specific peptide inhibitors against kinase Pfmrk from malaria-causing parasite Plasmodium falciparum with minimal impact on human CDKs of the host cell. While our predictions on Pfmrk are only computational so far, which must be further verified experimentally, the results are tantalizing enough to merit further optimization on the peptide inhibitors based on the subtle difference of the TL pocket of CDKs among different organisms.

In summary, focusing only on the CDK branch of an entire kinome for different organisms, we provided a computational approach that combines structure modeling, pocket detection, and evolutionary conservation analysis to identify non-catalytic pockets for designing inhibitors of the desired specificity. It could be useful to extend this approach to the entire kinome.

Author contributions

Huiwen Wang and Kaili Wang performed most computational analysis under the supervision of Ya Jia. Fatah Kashanchi helped with the research design; Yunjie Zhao performed MD simulations; Zeyu Guan and Yiren Jian helped

with the conservation analysis; Yunjie Zhao and Chen Zeng supervised the overall study and wrote the paper.

References

[1] Endicott J A and Noble M E 1998 *Structure* **6** 535
[2] Enserink J M and Kolodner R D 2010 *Cell. Div.* **5** 11
[3] Ubersax J A, Woodbury E L, Quang P N, Paraz M, Blethrow J D, Shah K, Shokat K M and Morgan D O 2003 *Nature* **425** 859
[4] Loog M and Morgan D O 2005 *Nature* **434** 104
[5] Holt L J, Tuch B B, Villen J, Johnson A D, Gygi S P and Morgan D O 2009 *Science* **325** 1682
[6] Paglini G and Caceres A 2001 *Eur. J. Biochem.* **268** 1528
[7] Demetrick D J, Zhang H and Beach D H 1994 *Cytogenet. Cell Genet.* **66** 72
[8] Otto T and Sicinski P 2017 *Nat. Rev. Cancer* **17** 93
[9] Dachineni R, Ai G, Kumar D R, Sadhu S S, Tummala H and Bhat G J 2016 *Mol. Cancer Res.* **14** 241
[10] Shukla D, Meng Y, Roux B and Pande V S 2014 *Nat. Commun.* **5** 3397
[11] Liu H, Liu K, Huang Z, Park C M, Thimmegowda N R, Jang J H, Ryoo I J, He L, Kim S O, Oi N, Lee K W, Soung N K, Bode A M, Yang Y, Zhou X, Erikson R L, Ahn J S, Hwang J, Kim K E, Dong Z and Kim B Y 2013 *J. Biol. Chem.* **288** 25924
[12] Martin M P, Alam R, Betzi S, Ingles D J, Zhu J Y and Schonbrunn E 2012 *Chembiochem* **13** 2128
[13] Doerig C, Abdi A, Bland N, Eschenlauer S, Dorin-Semblat D, Fennell C, Halbert J, Holland Z, Nivez M P, Semblat J P, Sicard A and Reininger L 2010 *Biochim. Biophys. Acta* **1804** 604
[14] Peng Y, Keenan S M and Welsh W J 2005 *J. Mol. Graph. Model* **24** 72
[15] Leete T H and Rubin H 1996 *Parasitol. Today* **12** 442
[16] Wei Y and Lu-Hua L 2016 *Chin. Phys. B* **25** 018702
[17] Zhang Y H, Peng J H and Zhang Z Y 2015 *Chin. Phys. B* **24** 126101
[18] Sun Z H and Jiang F 2010 *Chin. Phys. B* **19** 110502
[19] Kaur G, Stetler-Stevenson M, Sebers S, Worland P, Sedlacek H, Myers C, Czech J, Naik R and Sausville E 1992 *J. Natl. Cancer Inst.* **84** 1736
[20] Arguello F, Alexander M, Sterry J A, Tudor G, Smith E M, Kalavar N T, Greene Jr J F, Koss W, Morgan C D, Stinson S F, Siford T J, Alvord W G, Klabansky R L and Sausville E A 1998 *Blood* **91** 2482
[21] Chao S H, Fujinaga K, Marion J E, Taube R, Sausville E A, Senderowicz A M, Peterlin B M and Price D H 2000 *J. Biol. Chem.* **275** 28345
[22] Lanasa M C, Andritsos L, Brown J R, Gabrilove J, Caligaris-Cappio F, Ghia P, Larson R A, Kipps T J, Leblond V, Milligan D W, Janssens A, Johnson A J, Heerema N A, Buhler A, Stilgenbauer S, Devin J, Hallek M, Byrd J C and Grever M R 2015 *Leuk. Res.* **39** 495
[23] Xing S, Li F, Zeng Z, Zhao Y, Yu S, Shan Q, Li Y, Phillips F C, Maina P K, Qi H H, Liu C, Zhu J, Pope R M, Musselman C A, Zeng C, Peng W and Xue H H 2016 *Nat. Immunol.* **17** 695
[24] Zhao Y, Zeng C and Massiah M A 2015 *Plos One* 10
[25] Zhao Y, Zeng C, Tarasova N I, Chasovskikh S, Dritschilo A and Timofeeva O A 2013 *Transcription* **4** 227
[26] Zhao Y, Huang Y, Gong Z, Wang Y, Man J and Xiao Y 2012 *Sci. Rep.* **2** 734
[27] Zhao Y, Gong Z and Xiao Y 2011 *J. Biomol. Struct. Dyn.* **28** 815
[28] Wang J, Zhao Y, Zhu C and Xiao Y 2015 *Nucleic. Acids. Res.* **43** e63
[29] Liu Q, Chen W, Chen C, Zhao Y and Zeng C 2017 *Chemical Journal of Chinese Universities* **38** 1185
[30] Betzi S, Alam R, Martin M, Lubbers D J, Han H, Jakkaraj S R, Georg G I and Schonbrunn E 2011 *ACS Chem. Biol.* **6** 492
[31] Rastelli G, Anighoro A, Chripkova M, Carrassa L and Broggini M 2014 *Cell. Cycle.* **13** 2296
[32] Chen H, Zhao Y, Li H, Zhang D, Huang Y, Shen Q, Van Duyne R, Kashanchi F, Zeng C and Liu S 2014 *PLoS One* **9** e109154
[33] Hu Y, Li S, Liu F, Geng L, Shu X and Zhang J 2015 *Bioorg. Med. Chem. Lett.* **25** 4069
[34] Volkamer A, Kuhn D, Grombacher T, Rippmann F and Rarey M 2012 *J. Chem. Inf. Model.* **52** 360
[35] Volkamer A, Griewel A, Grombacher T and Rarey M 2010 *J. Chem. Inf. Model.* **50** 2041
[36] Letunic I and Bork P 2011 *Nucleic. Acids. Res.* **39** W475
[37] Letunic I and Bork P 2007 *Bioinformatics* **23** 127
[38] Letunic I and Bork P 2016 *Nucleic. Acids. Res.* **44** W242
[39] Jeffrey P D, Russo A A, Polyak K, Gibbs E, Hurwitz J, Massague J and Pavletich N P 1995 *Nature* **376** 313
[40] Lolli G, Lowe E D, Brown N R and Johnson L N 2004 *Structure* **12** 2067
[41] Tahirov T H, Babayeva N D, Varzavand K, Cooper J J, Sedore S C and Price D H 2010 *Nature* **465** 747
[42] Yang J, Yan R, Roy A, Xu D, Poisson J and Zhang Y 2015 *Nature Methods* **12** 7
[43] Roy A, Kucukural A and Zhang Y 2010 *Nature Protocols* **5** 725
[44] Goldenberg O, Erez E, Nimrod G and Ben-Tal N 2009 *Nucleic. Acids. Res.* **37** D323
[45] Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T and Ben-Tal N 2016 *Nucleic. Acids. Res.* **44** W344
[46] Katoh K and Standley D M 2014 *Methods. Mol. Biol.* **1079** 131
[47] Pupko T, Bell R E, Mayrose I, Glaser F and Ben-Tal N 2002 *Bioinformatics* **18** S71
[48] Pronk S, Pall S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts M R, Smith J C, Kasson P M, van der Spoel D, Hess B and Lindahl E 2013 *Bioinformatics* **29** 845
[49] Oostenbrink C, Villa A, Mark A E and van Gunsteren W F 2004 *J. Comput. Chem.* **25** 1656
[50] Sethi A, Eargle J, Black A A and Luthey-Schulten Z 2009 *Proc. Natl. Acad. Sci. USA* **106** 6620
[51] Zhao Y, Jian Y, Liu Z, Liu H, Liu Q, Chen C, Li Z, Wang L, Huang H H and Zeng C 2017 *Sci. Rep.* **7** 2876
[52] Fraczkiewicz R and Braun W 1998 *J. Comput. Chem.* **19** 319
[53] Ingles-Prieto A, Ibarra-Molero B, Delgado-Delgado A, Perez-Jimenez R, Fernandez J M, Gaucher E A, Sanchez-Ruiz J M and Gavira J A 2013 *Structure* **21** 1690
[54] Crompton P D, Moebius J, Portugal S, Waisberg M, Hart G, Garver L S, Miller L H, Barillas-Mury C and Pierce S K 2014 *Ann. Rev. Immunology* **32** 157
[55] Waters N C, Woodard C L and Prigge S T 2000 *Molecular and Biochemical Parasitology* **107** 45
[56] Hagel M, Niu D, Martin T St, Sheets M P, Qiao L, Bernard H, Karp R M, Zhu Z, Labenski M T, Chaturvedi P, Nacht M, Westlin W F, Petter R C and Singh J 2011 *Nat. Chem. Biol.* **7** 22