J Mol Model (2013) 19:1711–1725
DOI 10.1007/s00894-012-1735-2

ORIGINAL PAPER

Computational study and peptide inhibitors design for the CDK9 – cyclin T1 complex

Jelena Randjelović · Slavica Erić · Vladimir Savić

Received: 17 September 2012 / Accepted: 13 December 2012 / Published online: 8 January 2013

Abstract Cyclin dependent kinase 9 (CDK9) is a protein that belongs to the cyclin-dependent kinases family, and its main role is in the regulation of the cell transcription processes. Since the increased activity of CDK9 is connected with the development of pathological processes such as tumor growth and survival and HIV-1 replication, inhibition of the CDK9 could be of particular interest for treating such diseases. The activation of CDK9 is initiated by the formation of CDK9/cyclin T1 complex, therefore disruption of its formation could be a promising strategy for the design of CDK9 inhibitors. In order to assist in the design of potential inhibitors of CDK9/cyclin T1 complex formation, a computational study of the CDK9/cyclin T1 interface was conducted. Ten peptides were designed using the information from the analysis of the complex, hot spot residues and fragment based design. The designed peptides were docked to CDK9 structures obtained by molecular dynamics simulations of CDK9/cyclin T1 complex and the CDK9 alone and their binding affinities were evaluated using molecular mechanics Poisson Boltzman surface area (MM-PBSA) method and steered molecular dynamics (SMD). Designed peptide sequences LQTLGF and ESIILQ, both derived from the surface of cyclin T1, as well as the peptide sequence PRWPE, derived from fragment based design, showed the most favorable binding properties and were selected for our further studies.

Keywords CDK9/cyclin T1 · Fragment based design · MM-PBSA · Protein-protein interaction · SMD

Electronic supplementary material The online version of this article (doi:10.1007/s00894-012-1735-2) contains supplementary material, which is available to authorized users.

J. Randjelović · V. Savić
Department of Organic Chemistry, University of Belgrade - Faculty of Pharmacy, Vojvode Stepe 450,
11221 Belgrade, Serbia

J. Randjelović · S. Erić (✉)
Department of Pharmaceutical Chemistry, University of Belgrade - Faculty of Pharmacy, Vojvode Stepe 450,
11221 Belgrade, Serbia
e-mail: seric@pharmacy.bg.ac.rs

Introduction

Cyclin-dependent kinase 9 (CDK9) is a serine-threonine kinase and the catalytic component of the positive transcription elongation factor b (P-TEFb) complex, which promotes the elongation of RNA polymerase II transcripts. There are two isoforms of CDK9 in humans (CDK9_42 and CDK9_55). Monomeric CDK9 molecules of both isoforms, in complex with cyclins T1, T2a, T2b and K, form distinct P-TEFb complexes [1]. In the cell, the newly synthesized CDK9 is not stable - it interacts with the cytoplasmatic proteins Cdc37, Hsp90 and Hsp70 [2] and is escorted by them to bind to its cyclin partner. CDK9 has the ability to shuttle between the nucleus and cytoplasm, while its cyclin partners are localized in the nucleus [3]. Binding of CDK9 to a cyclin leads to its partial catalytic activation, which is further enhanced by autophosphorylation of its Thr186 residue.

Inhibition of CDK9 may prove beneficial in treating several human diseases including some types of cancer, HIV-1 infection and cardiac hypertrophy [1, 4–6]. It has been shown that CDK9 inhibitors suppress the general transcription levels within the cell, mostly affecting the concentration levels of short lived proteins, such as cell survival factors, which could have an impact on the survival of cancer cells [7]. Among these anti-apoptotic proteins most heavily affected by the decreased activity of CDK9 is myeloid cell factor 1 (Mcl1), which is necessary for the survival of multiple myeloma cells. An in vitro study conducted on

multiple myeloma cells has shown that inhibition of CDK9 in P-TEFb complex led to the selective apoptosis of tumor cells [8]. The ability of most CDK inhibitors undergoing clinical trials to induce apoptosis in tumor cells has been linked to the inhibition of CDK9 [4].

CDK9 is also a potential therapeutic target in treatment of HIV-1 infection, as the HIV-1 virus recruits human P-TEFb to enable its own transcription to proceed. This interaction is specific for P-TEFb complex formed from CDK9 and cyclin T1 only [1] and is accomplished through the binding of a small viral protein Tat (transactivator of transcription) to P-TEFb. Tat binds to the cyclin T1 and to the T loop of CDK9 inducing a conformational change in the complex, which alters its substrate specificity and increases its activity [9]. An in vitro study conducted on human primary peripheral blood lymphocytes has shown that a selective inhibition of CDK9 mediated transcription has led to the inhibition of HIV-1 replication [10]. There are indications that selective inhibition of CDK9 activity could lead to inhibition of HIV replication in infected cells without affecting cell viability [11]. As HIV protein Tat can also interfere with the physiological regulatory mechanism that inhibits P-TEFb activity [12, 13], simply sequestering active P-TEFb molecules is insufficient for a lasting therapeutic effect. On the other hand, the monomeric unbound CDK9 is quickly degraded in the cell [2], which implies that inhibitors designed to disrupt the formation or lower the stability of P-TEFb complex could potentially prove to be useful in the therapy of HIV-1 infection.

The known small molecule inhibitors of P-TEFb complex act by binding at the ATP binding pocket of CDK9. As the ATP binding pocket is fairly conserved in the CDK family, the selectivity of action of known CDK9 inhibitors is limited. In addition, these inhibitors have to compete with high intracellular concentrations of ATP during binding. There has been some slight progress in improving the selectivity of CDK9 inhibitors binding to its catalytic pocket achieved by exploiting the change of conformation of the glycine-rich loop of CDK9 [14], however, compounds with increased selectivity are still needed [4, 10]. An alternative approach in designing inhibitors of CDK9 would be to interfere with the formation of P-TEFb complex [2], which could be achieved by inhibiting the CDK9/cyclin T1 interaction. A direct inhibitor of the CDK9/cyclin T1 interaction would prevent the formation of P-TEFb complex and exhibit selectivity lacking in the existing small molecule inhibitors of CDK9. Despite the inherent difficulties in targeting protein-protein interactions for drug development, inhibitors of a number of protein-protein interactions have been successfully developed [15–17], and some have even entered clinical trials [18] or have already been approved for use in therapy [19]. To the best of our knowledge, no direct inhibitors of the CDK9/cyclin T1 interaction have been designed so far.

In this work, in order to obtain information useful for the design of potential inhibitors of the CDK9/cyclin T1 interaction, we have performed a computational analysis of the CDK9/cyclin T1 interaction interface. At this stage of our research, we decided to design peptide inhibitors, as the interactions already present at the native protein-protein interface could be more naturally transferred onto peptides, and, once evaluated, the peptides themselves could be altered into peptidomimetics and eventually small molecule inhibitors with improved stability and cellular delivery.

We performed the design of a series of peptides by using the information obtained from the interface hot spot residues, as well as by applying fragment based design methods. The evaluation of the binding affinity of the designed peptides was performed by using molecular mechanics Poisson Boltzman surface area (MM-PBSA) and steered molecular dynamics (SMD).

Methods

MD simulations

The structure of the P-TEFb complex chosen for computational study and inhibitors design was the crystallographic structure of the complex [20] (PDB ID: 3BLH). Its resolution is 2.48 Å and both R and R-free values fit the recommended quality (R<0.25, R-free <0.4 and R-free - R=0.042 <0.05) [21]. None of the CDK9 or cyclin T1 residues forming the P-TEFb interface in the PDB:ID 3BLH structure are directly involved in crystal contacts, as verified by CryCo web server [22], and although the potential effects of crystal packing on the starting complex conformation could not be completely eliminated in the absence of an additional crystallographic structure of the same complex solved in a different crystal form, their influence was minimized by relaxing the complex in subsequent MD simulation. The 51–54 loop was grafted from a related crystallographic structure of P-TEFb (PDB ID: 3BLQ) [20], while the 177–181 and 260–266 loops were grafted from the P-TEFb-Tat complex structure [9] (PDB ID: 3MI9) since the structural overlap of these regions in the two CDK9 structures was satisfactory, considering that the majority of Tat-induced conformational changes of P-TEFb are localized at the cyclin T1 structure and the region around the 51–54 loop [9]. The final missing loop (residues 88–97) was modeled using MODELLER [23] offered by ModLoop online server [24]. The phosphorylated T186 residue in CDK9 was replaced with a regular threonine. The final structure was submitted to the PROPKA 3.0 [25] web server for the assignment of amino acid protonation states at pH 7.4. In accordance with PROPKA recommended protonation, the N-terminus of the cyclin T1 in the protein PDB structure,

corresponding to the Arg 5 residue of the primary sequence, was deliberately left unprotonated. While solvent exposed N-terminal amino acids of proteins at pH 7.4 are usually protonated, the chain in question here was truncated (amino acids 1–7 are not present in the crystallographic structure) and the charge was not naturally present at that position. The N-terminus of the CDK9 was also truncated to Val 8, but as this position was in the vicinity to the interface surface studied, oriented toward the solvent and highly flexible, the charge was retained. The Lys 6 of the cyclin T1 was also deprotonated, as the amino group of its side chain is in a non-polar environment. The Cys 10 from the CDK9 sequence, according to PROPKA, would have both ionized and neutral forms present at pH 7.4 (pKa=6.67), so the neutral form was chosen.

The P-TEFb structure was kept fixed and the modified loops alone were minimized for 20,000 steps in NAMD 2.8 [26], using the CHARMM22/CMAP [27, 28] forcefield. Files for the simulation were prepared in VMD software [29].

The prepared P-TEFb structure was solvated with a rectangular TIP3P water box with the box boundaries at least 15 Å from the complex, neutralized with Cl^- ions and minimized for an additional 30,000 steps. Afterward, the complex was heated from 0 to 310 K, for 93 ps, in 0.1 K increments and submitted to molecular dynamics in NPT ensemble, for 4 ns. Simulation parameters were as follows: timestep of 1 fs, with particle-mesh Ewald electrostatics, Langevin thermostat set to 310 K, with a damping constant of 2/ps and pressure maintained at 1.01325 bar with a Nose-Hoover Langevin piston with a period of 200 fs, decay of 100 fs and temperature of 310 K.

To estimate the binding surface differences between the structures of CDK9 bound in P-TEFb and free in solution, the CDK9 structure was extracted from the final frame of the 4 ns MD simulation of the P-TEFb complex and prepared in the same way. The structure was solvated with a rectangular TIP3P water box with the box boundaries at least 15 Å from the complex, neutralized with Cl^- ions and minimized for 30,000 steps, then heated to 310 K in 0.1 K increments, for 93 ps and simulated in NPT ensemble for 4 ns, with the same MD simulation parameters as those used for the P-TEFb complex simulation.

Since the MM-PBSA method selected for the evaluation of binding affinities of the potential inhibitors, as implemented in Amber Tools 1.5 [30] software, requires the use of Cornell at al. forcefield in the simulations, the above simulation of the complex and CDK9 was repeated using the Cornell at al. forcefield version ff10 (ff99SB [31]) with Joung and Cheatham adjustments for ions in TIP3P water model [32] (frcmod.ionsjc_tip3p). Files for this simulation were prepared using LeaP implemented in Amber Tools 1.5 and the simulation itself was conducted in NAMD 2.8.

Simulations were terminated at 4 ns as the systems appeared to have stabilized. RMSD plots for alpha carbons for the P-TEFb and CDK9 alone 4 ns MD simulations are presented in Online resource 1 (Figs. 1 and 2), along with the potential energy plots for the same simulations (Online resource 1, Figs. 3 and 4).

CDK9 structures extracted from the final frames of the P-TEFb and CDK9 alone 4 ns simulations were checked using the WHAT IF Protein Model Check webserver [33] and found satisfactory. None of the residues had occupied a rotamer state deviating significantly from the expected probability distribution of rotamers for the corresponding amino acid.

Interface analysis

To study the differences in the binding surfaces of cyclin T1 bound and free CDK9 structures, the structures of CDK9 were extracted from the final frames of 4 ns MD simulations of P-TEFb and CDK9 alone and superimposed to the crystallographic structure of CDK9 from PDB ID: 3BLH in VMD software.

The CDK9 structures obtained from the P-TEFb and the free CDK9 MD simulations were submitted to fpocket [34] web server, in an attempt to detect hydrophobic pockets on the binding surface suitable for the binding of small molecules. The CDK9 binding surface of interest to us, throughout this work, is defined by the CDK9 amino acid residues involved in the CDK9/cyclin T1 protein interaction.

To obtain information of the energetically most important amino acid residues involved in the CDK9/cyclin T1 interaction (hot spot residues), we submitted the crystallographic structure of the P-TEFb complex (PDB ID: 3BLH) to three web servers for interface analysis: in addition to the computational alanine scan analysis of the protein-protein interface between CDK9 and cyclin T1, conducted by the Robetta web server [35, 36], an additional two algorithms: KFC2 [37] and HotPoint [38] were also used to identify the hot spot residues on the interface. The Robetta method results were used as the main selection criterion for hot spot residues, with KFC2 and HotPoint methods used to test the overall agreement of the algorithms and to obtain additional confirmation. Changes in the solvent accessible surface area of the residues (ΔSASA) upon complex formation were explicitly used as an additional parameter in choosing residues most suitable for inhibitor design and were retrieved from the ANCHOR database [39], while the interaction maps of the residues were obtained from the SCOWLP database [40, 41].

In order to estimate the potential for inhibiting the CDK9/cyclin T1 interaction without affecting CDK9s interaction with cyclins T2 and K, the sequences of cyclins T1, T2 and K were aligned using CLUSTALW [42].

The crystallographic structure of P-TEFb complex bound to HIV1 Tat protein (PDB ID: 3MI9) was also submitted to the Robetta web server for computational alanine scan analysis, in order to determine if any of the Tat protein amino acid residues interacting with CDK9 could be used in our inhibitor design. It was superimposed to CDK9 structures obtained from our MD simulations of P-TEFb and free CDK9 in VMD software, for comparison.

SEED/GANDI fragment based design

Fragment based design was conducted using the solvation energy for exhaustive docking (SEED) [43] and genetic algorithm-based de novo design of inhibitors (GANDI) [44] software. For the SEED run, the fragments to be docked were the 134 rotamers of the 19 natural amino acids present in the Penultimate Rotamer Library [45] and an added glycine structure. SEED software docks molecular fragments to the surface of a rigid protein receptor, while estimating the desolvation energy of binding in addition to the traditional vdW and electrostatic terms. The fragments were docked using default SEED parameters to the structures of CDK9 extracted from the final frame of the 4 ns MD simulations of the P-TEFb complex and CDK9 alone, with clustering cutoffs set to −5 and −8 kcal mol^{-1} and binding sites defined by the CDK9 amino acid residues from the CDK9/cyclin T1 interface. Once the favorable positions for all the fragments were determined, they were forwarded to GANDI software, to be linked into complete pentapeptides. The aforementioned 135 fragments were used both as fragments and linkers for the GANDI software run, with connection only allowed along the heavy atom – hydrogen vectors that are involved in the formation of peptide bonds. GANDI software was used to design pentapeptides, evolving 20 islands of 100 candidates each for 5000 steps, with the 2D and 3D similarity terms of the scoring function turned off and default run parameters. The same receptor files and binding sites were used as in the SEED run.

Docking

Peptides were docked to the surface of CDK9 structures using AutoDock Vina [46]. The peptides were docked to the CDK9 structures from the final frame of the 4 ns simulation of P-TEFb complex and CDK9 alone, with the binding site defined as a grid of 24×30×34 Å, centered on the interface surface. Top 20 candidates were retained for each run. Initially, each peptide was subjected to 20 docking runs, with the exhaustiveness parameter of the software (describing the number of starting random conformations tested during the docking procedure) set to 400, but on testing on a limited number of peptides, similar binding modes were recovered with exhaustiveness parameter of 64, so an additional 20 runs were performed with that parameter setup. Four randomly assembled pentapeptides were used as a control group and subjected to the same docking procedure. Files for docking were prepared using AutoDock Tools [47] and results visualized in VMD.

MM-PBSA

The binding affinities of the peptide inhibitor candidates were evaluated using MMPBSA method [48]. Each peptide-CDK9 complex was solvated with a rectangular TIP3P water box with the box boundaries at least 15 Å from the complex and neutralized with Cl^- ions. The Cornell at al. forcefield version ff10 (ff99SB for proteins) with Joung and Cheatham adjustments for ions in TIP3P water model (frcmod.ionsjc_tip3p) was used. After being minimized for 30,000 steps and heated to 310 K over 93 ps, in 0.1 K increments, the complexes were simulated for 4 ns in NPT ensemble, with rigid water molecules, particle-mesh Ewald electrostatics, Langevin thermostat set to 310 K with the damping constant of 2/ps, pressure maintained at 1.01325 bar using a Nose-Hoover Langevin piston with a period of 200 fs, decay of 100 fs and temperature set to 310 K. All files for simulations were prepared using LeaP and VMD, and the simulations were run using NAMD 2.8. From each trajectory 100 snapshots from the final nanosecond were extracted and used to calculate the free energy of binding, through the automated MM-PBSA.py procedure in AmberTools 1.5, with the salt concentration set to 0.1 M and to 0.15 M. Free energies of binding were also evaluated using a MM-GBSA approach [49], as implemented in AmberTools 1.5. Entropy of the binding was estimated using the nmode module of AmberTools 1.5, on five snapshots extracted every 200 ps from the final nanosecond of the peptide-CDK9 complex 4 ns MD simulations, and calculated at 310 K.

Peptide ligand – CDK9 interaction diagrams were mapped out using LigPlot + software [50].

SMD

Steered molecular dynamics (SMD) runs were performed as an alternative method of estimating relative binding affinity of the peptide inhibitors. Simulations were conducted on all complexes from the MM-PBSA runs. Peptide-CDK9 complexes were solvated with a rectangular TIP3P water box with the box boundaries at least 15 Å from the solutes, neutralized with Cl^- ions and simulated using the CHARMM22/CMAP force field. The complexes were minimized for 30,000 steps and then heated from 0 to 310 K, for 93 ps, in 0.1 K increments. In order to select the optimal simulation conditions, test runs were conducted with three different fixed atom schemes: a single atom (Cα of residue 20, positioned at the far side of the protein, opposite to the binding surface explored) fixed, protein backbone fixed and

"barrel back" selection fixed (the distal side of the β barrel opposite the binding surface, including backbone atoms of residues 20–25, 32–37, 44–48, 104, 106, 107 from the CDK9 structure), with different pulling speeds (5, 10, 50 and 100 Å/ns) and different spring constants (1.5, 3 and 7 kcal mol^{-1}·$Å^2$). The final protocol selected was the "barrel back" selection of fixed atoms, pulling speed of 10 Å/ns and spring constant of 1.5 kcal mol^{-1}·Å.

To test different pulling directions for peptides, a random acceleration molecular dynamics (RAMD) test run was conducted, as implemented in NAMD 2.8. 100 trajectories of 500 ps were generated at different random seeds, using the following parameters: ligand acceleration of 0.05 *kcal* mol^{-1}·Å·amu, minimum distance to be traveled by ligand in one step 0.006 Å and maximum distance between ligand and protein at which to terminate simulation 50 Å.

In the final SMD runs, backbone atoms of residues 20–25, 32–37, 44–48, 104, 106, 107 from the CDK9 structure were kept fixed, while all peptide atoms were pulled. The pulling direction was determined as a vector between the centers of mass for these two atom selections. Time step used in the runs was 2 fs, particle-mesh Ewald electrostatics was employed, and each run lasted 4 ns (2,000,000 steps).

Results

In this work, we first performed a computational analysis of the CDK9/cyclin T1 complex interface. Using the information of the analysis, as well as fragment based design, we have designed a series of peptides as potential inhibitors of the CDK9/cyclin T1 interaction, docked the designed peptides to the surface of CDK9 and estimated their binding affinities.

Complex preparation

Crystallographic structure of the human P-TEFb was resolved in 2008 [20] and its structure in complex with the Tat protein in 2010 [9]. The structure of the P-TEFb chosen for inhibitor design was the crystallographic structure extracted from PDB entry 3BLH (PDB ID: 3BLH). The atomic coordinates of four flexible loops in the structure of CDK9 were not resolved, including the residues 51–54, 88–97, 177–181 and 260–266. The missing loops were modeled as described in MD simulations. The phosphorylated threonine T186 from the CDK9 structure was replaced with regular threonine and the protein was protonated at pH 7.4.

The P-TEFb complex structure was then further prepared for inhibitor design through minimization, heating and MD simulation (see MD simulations.) As we were attempting to design inhibitors that would bind to the free surface of CDK9, its structure alone was extracted from the 4 ns snapshots of the equilibrated structure of the complex, solvated, neutralized and submitted to minimization, heating and MD simulation, under the same conditions as the P-TEFb complex (see MD simulations), in order to investigate the changes of the binding surface between the CDK9 complexed in P-TEFb and the "free" state we imitated.

Interface analysis

In comparison with similar CDK-cyclin complexes, P-TEFb has a smaller protein-protein interface [9]. It consists of 48 amino acids. To study the interface, the structures of CDK9 were extracted from the final frames of 4 ns MD simulations of P-TEFb and CDK9 alone structures and superimposed to the crystallographic structure of CDK9 from PDB ID: 3BLH (Fig. 1a and b). The alignment was satisfactory, with RMSD differences between the two of 2.30 Å for the backbone atoms and 2.96 Å for all atoms. On the CDK9 interface surface itself, the most prominent difference was the reorientation of two Phe residues, Phe 12 and Phe 59 (Fig. 1c).

Exploration of the interfaces using the fpocket computational algorithm did not reveal any hydrophobic pockets suitable for the direct design of small molecule inhibitors. To establish which residues of the protein complex interface contributed mostly to the free energy of binding, as hot spot residues, we performed a computational analysis of the interface using four different algorithms. Results of the interface analysis are shown in Table 1, which also numbers the individual interactions of the residues.

We extended our interface analysis to the interactions between Tat protein and CDK9 in the P-TEFb/Tat complex and found that a tryptophane residue in Tat protein, Trp 11, interacting with CDK9's T loop, is also a hot spot residue (contributes 3.82 kcal mol^{-1} to the complex binding energy).

Design of peptides

Due to the different appearance of the binding surface in the bound and "free" CDK9 structures (Fig. 1), design of potential peptide inhibitors was attempted on both CDK9 conformations. While the use of a fully flexible protein structure in the design could likely improve its predictive power, the approach used in this work does provide a degree of insight on the possible conformational changes of the binding surface, at a fraction of the computational cost. The flexibility of CDK9 is explicitly accounted for in the subsequent MD simulations of peptide-protein complexes during the binding energy evaluation stage.

Two distinct approaches were utilized in the design of the compounds listed in Table 2. In the first, 5–8 amino acid peptide sequences corresponding to the fractions of cyclin T1 primary sequence carrying the identified hot-spots were used as potential inhibitor candidates. All the hot-spot

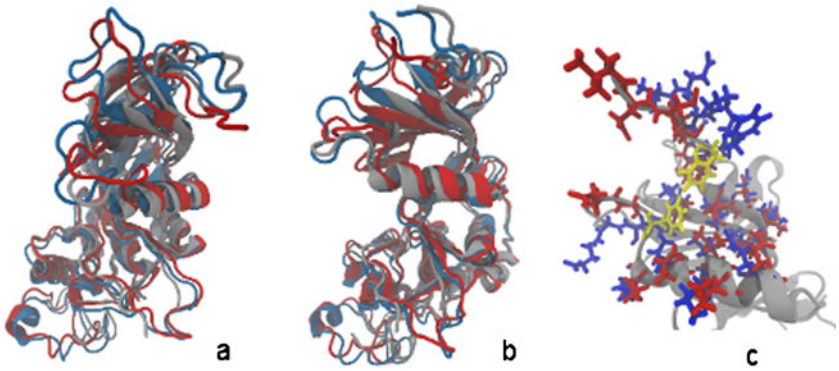

Fig. 1 **a** Alignment of CDK9 structures: crystal structure, PDB ID: 3BLH (*grey*), 4 ns MD simulation of P-TEFb (*blue*) and 4 ns MD simulation of «free» CDK9 (*red*). **b** the same alignment viewed from a different angle (90° rotation of 1a). **c** binding surface residues (liquorice representation), reoriented Phe residues on the binding surface (*yellow*)

residues chosen for the design are conserved between different cyclins interacting with CDK9 (see Fig. 5 in Online resource 1 for the sequence alignment of cyclins T1, T2 and K with marked chosen hotspots and sequence fragments used in the design). Four sequence fragments were chosen from the cyclin T1 structure in PDB ID:3BLH: residues 141–148, 89–94, 137–142, 141–146 (peptides 1–4, Table 2). Peptide 4 (LQTLGF) is a truncated version of octapeptide 1 (LQTLGFEL), containing the Phe 146 hot spot residue. Both peptides were included in our study, as one of the computational algorithms used to select hot spots (HotPoint) also identified the C-terminal leucin (Leu 148) from peptide 1 as a hot spot and this amino acid interacts with the same amino acid cluster as the hot spot Phe 146. In the second approach, the binding surfaces of the two CDK9 structures were subjected to unbiased *de novo* peptide ligand design as detailed in SEED/GANDI fragment based design. The three top scoring peptide sequences from each design run were considered for further selection (peptides 5–10, Table 2). In the fragment linking part of the fragment based design procedure on the structure of CDK9 obtained from the CDK9 alone MD simulation ("free" CDK9, in Table 2) only tetrapeptides were successfully built (peptides 5–7, Table 2).

As a negative control, four pentapeptides were also built by randomly selecting amino acids (peptides 11–14, Table 2).

The goal of our peptide design was to obtain short peptides (5–6 resides long) that would be suitable for potential subsequent transformation into non-peptide structures. Due to the small size of the peptides, in order to obtain the strongest interactions possible with the given limited molecular surface, our design considered all types of interactions equally.

Docking studies

The docking procedure is described in Docking. AutoDock-Vina software was used to perform the docking. While AutoDock itself is designed solely for the docking of small molecules, there are reports in the literature of AutoDock Vina being used to dock peptides to proteins, namely to amyloid peptides [51] and protein kinases [52]. The predicted binding energies of the top scoring peptide conformations from the docking studies are shown in Table 3.

MM-PBSA results

As scoring functions used in docking provide an approximate estimate of binding affinities, free energies of binding of peptide inhibitor candidates were more accurately determined through the MM-PBSA and molecular mechanics generalized Born surface area (MM-GBSA) approach. The procedure is detailed in MM-PBSA. and the results are shown in Table 4.

Due to high computational cost, binding entropy was estimated only for four of the peptide-CDK9 complexes,

Table 1 Excerpt from the hot spot amino acid analysis of the CDK9/cyclin T1 interface

Aminoacids	Robetta ΔΔG kcal mol^{-1}	KFC2 hot spot +/−	HotPoint hot spot +/−	ANCHOR ΔSASA Å^2	ΔSASA %	SCOWLP residue interacts with:
[a]A12 – PHE	2.21	–	–	125.8	76.7	[a]B72, 142, 143, 145, 12, 11
B73 – GLN	2.4	–	+	50	35.5	A8, A9
B93 – LYS	3.7	+	–	65.9	40.3	A61, 64, 59, 57
B137 – GLU	2.84	–	+	17.2	12.8	A59
B142 – GLN	0.75	–	+	81.5	57.8	A10, 11, 12, 13, 86, 99
B146 – PHE	2.87	+	+	91.8	55.9	A59, 67, 71, 84, 99

[a] A – CDK9 protein chain; B – cyclin T1 chain

Table 2 Sequences of designed peptides

Peptides	Sequence	Source
1	LQTLGFEL	cyclin T1 residues 141–148
2	FLAAKV	cyclin T1 residues 89–94
3	ESIILQ	cyclin T1 residues 137–142
4	LQTLGF	cyclin T1 residues 141–146
5	FWGH	SEED/GANDI, free[a]
6	KPEQ	SEED/GANDI, free
7	KPPQ	SEED/GANDI, free[b]
8	PRWPE	SEED/GANDI, bound[a,b]
9	PWKE	SEED/GANDI, bound
10	PRWKE	SEED/GANDI, bound
11	AFLIV	negative control, random
12	KIIAT	negative control, random
13	VRPHL	negative control, random
14	AAAAA	negative control, random

[a] bound and free refer to the structures of CDK9 obtained from the P-TEFb (bound) and CDK9 alone (free) simulations

[b] the top-ranked peptide sequence in the fragment-based design on the corresponding (bound, free) conformation of CDK9

using just five frames from the MD simulations. At 310 K, $T\Delta S$ values obtained average at −20 to −25±10 kcal mol^{-1}. $T\Delta S$ for the best binding peptides 3 and 4 to "free" and "bound" CDK9 conformations respectively, at 310 K, are −21.6±6.0 kcal mol^{-1} and −27.8±13.8 kcal mol^{-1}, resulting in the total free energy of binding of −19.3±11.6 kcal mol^{-1} and 4.4±18.2 kcal mol^{-1}, calculated according to the MMPBSA procedure [48]. For peptide 8 (PRWPE) binding to the "free" CDK9, $T\Delta S$ is −29.3±15.7, bringing its total free energy of binding to −4.6±23.2. $T\Delta S$ values of the negative control pentapeptide 13 (VRPHL) binding to these CDK9 conformations are −27.8±14.4 kcal mol^{-1} for bound and −26.5±14.3 kcal mol^{-1} for free CDK9, bringing its total free energies of binding to 9.0±18.2 kcal mol^{-1} and 2.4±21.7 kcal mol^{-1}, respectively. The reported errors for the total free energies of binding are the sums of absolute errors of their individual components (MM-PBSA energies from Table 4 and $T\Delta S$ values).

Table 3 Docking results

Peptides	Sequence	Energies [kcal mol^{-1}][a, b]
1	LQTLGFEL	6.9/7.2
2	FLAAKV	6.9/–
3	ESIILQ	–/5.9
4	LQTLGF	6.4/–
5	FWGH	–/–
6	KPEQ	–/5.7
7	KPPQ	5.8/5.8
8	PRWPE	6.7/7
9	PWKE	6.4/6.6
10	PRWKE	–/–
11	AFLIV	–/–
12	KIIAT	–/–
13	VRPHL	6.7/7
14	AAAAA	–/–

[a] on the left are energies obtained from docking the peptide to the structure of CDK9 obtained from the P-TEFb MD simulation and on the right energies from docking to the CDK9 structure from CDK9 alone MD simulation. [b] - indicates that no docking poses were found on the CDK9-cyclin T1 binding surface

SMD results

As an additional estimate of binding affinity, constant velocity pulling steered molecular dynamics (SMD) runs were performed on the peptide-protein complexes, starting from their initial docked conformations.

Peptides were pulled from the binding surface with constant velocity and force profiles were recorded, as detailed in SMD. The final protocol selected was the "barrel back" selection of fixed atoms, pulling speed of 10 Å ns^{-1} and spring constant of 1.5 kcal mol^{-1}·Å. The "barrel back" atom fixing scheme was chosen as it provided the best compromise between artificially constraining the protein movement and preventing excessive deformation of the binding area, as the flexible N-terminus and loops kept interacting with the peptides long after they vacated the binding surface. The 10 Å ns^{-1} speed was chosen as a compromise between minimizing pulling speed and keeping the computational time reasonable. The 1.5 kcal mol^{-1}·Å spring constant results in an elastic spring and was chosen in order to allow the flexible peptides to adjust as much as possible to the pulling. Use of such a soft spring lowered the level of detail for molecular events obtained from the force profiles, but the shape of the overall force profile investigated was retained. As the surface from which the peptides are pulled away is relatively flat, the choice of pulling direction, as long as it was comparable between peptides, is not as important as in pulling ligands out of pockets on protein surface, but a random acceleration molecular dynamics (RAMD) [53] run of 100 trajectories was conducted on a peptide to test the distribution of trajectories of a peptide leaving the surface. A clustering of RAMD trajectories around a specific direction in space would have indicated the existence of a preferred direction of movement for peptides leaving the CDK9 surface, which should be chosen as the pulling direction during the SMD runs. However, as expected from the relative flatness of the binding surface on CDK9, the obtained 100 RAMD trajectories for a peptide leaving the CDK9 surface were randomly distributed in space, favoring no single direction of movement and implying that the results obtained from SMD woud not be critically biased by choice of pulling direction.

Table 4 Calculated MM-GBSA and MM-PBSA energies for peptides, without the changes in conformational entropy upon binding (TΔS) accounted for

Peptides	Sequence	MM-GBSA energies [kcal mol^{-1}]		MM-PBSA energies [kcal mol^{-1}]	
		bound[a]	free[a]	bound[a]	free[a]
1	LQTLGFEL	−23.1±5.9	−14.1±7.5	−22.2±5.7	−15.5±6.8
2	FLAAKV	−19.1±4.2	/	−21.8±5.8	/
3	ESIILQ	/	−37.6±4.6	/	−40.9±5.6
4	LQTLGF	−20.0±3.8	/	−23.4 +/− 4.4	/
5	FWGH	/	/	/	/
6	KPEQ	/	−24.0±4.1	/	−26.5±5.5
7	KPPQ	−8.3±2.4	−13.2±3.2	−9.8±3.2	−12.1±4.5
8	PRWPE	−13.8±4.0	−34.8±5.7	−14.8±3.8	−33.9±7.5
9	PWKE	−14.9±3.2	−17.7±5.0	−14.9±4.7	−17.7±5.1
10	PRWKE	/	/	/	/
11	AFLIV	/	/	/	/
12	KIIAT	/	/	/	/
13	VRPHL	−14.4±2.8	−27.7±4.8	−18.8±3.8	−24.1±7.0
14	AAAAA	/	/	/	/

[a]bound and free refer to the structure of CDK9 from P-TEFb and itself alone MD simulations, respectively

The fixed atom scheme selected for the final SMD runs and the results of the RAMD run are shown in Fig. 2a and b, respectively. Obtained force profiles for the SMD runs are shown in Figs. 3 and 4 (for force profiles of individual peptides see Online resource 1, Figs. 6 and 7).

Discussion

Interface analysis

As the main contributions to the binding energy in protein-protein interactions originate from interactions of a few significant hot-spot amino acids [54], those hot-spot residues are a promising starting point for inhibitor design. The common method to determine which amino acids of the protein-protein interface form hot spots is to systematically mutate them to alanine, whether experimentally or *in silico* [35]. Two experimental mutagenesis studies exploring the binding between CDK9 and cyclin T1 have been reported [55, 56]. In both of them the binding of cyclin T1 to CDK9 was blocked by introducing two simultaneous point mutations in the cyclin T1 sequence: Lys 93 to Leu and Glu 96 to Lys. However, to the best of our knowledge, no experimental alanine scanning mutagenesis study has been conducted on the entire CDK9/cyclin T1 interaction interface, so we have explored all of the interface residues using computational methods. As it is less expensive and faster, computational alanine scanning is routinely performed, but it is not as accurate as the experiments. To counter this, we used three computational algorithms to investigate hot spot residues of the CDK9/cyclin T1 interace, and used the consensus of their results (Table 1) to guide our study.

In addition to hot spot residues identified with Robetta method (those with a ΔΔG >2 kcal mol^{-1} upon their mutation to alanine), we included Gln 142 from the cyclin T1 chain in our analysis. While this amino acid is only selected as a hot spot in the HotPoint prediction test out of the three

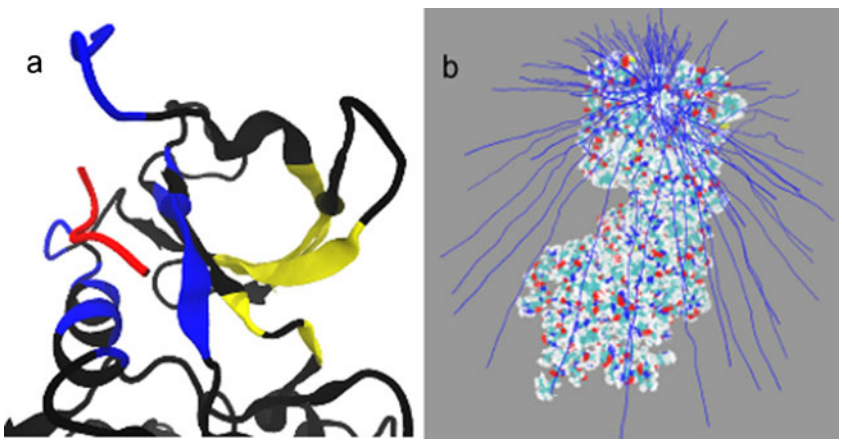

Fig. 2 SMD preparation. **a** «barrel back» choice of fixed atoms (*yellow*), binding surface residues (*blue*) of CDK9 (*black*) and bound PRWPE peptide (*red*). **b** 100 RAMD trajectories of the test RAMD simulation for verifying the validity of the chosen peptide pulling direction

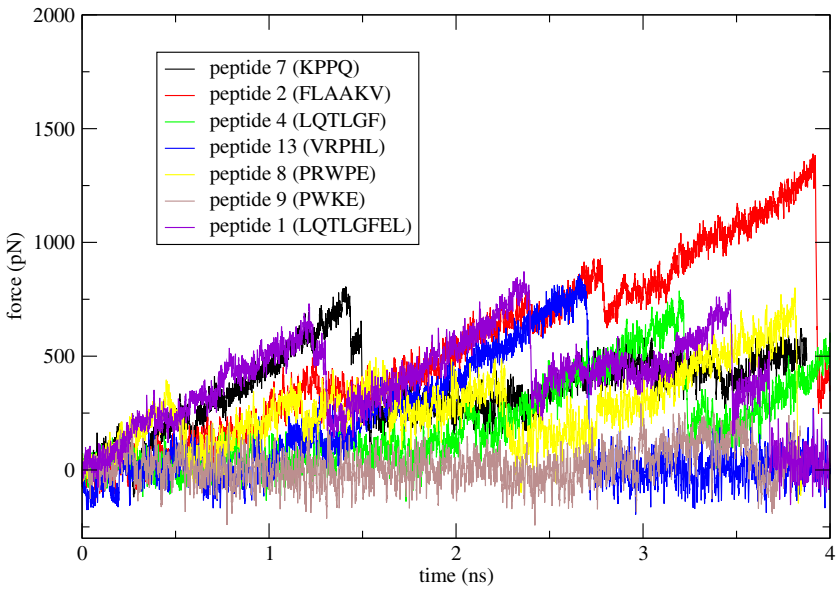

Fig. 3 Force profiles obtained in SMD runs for peptides pulled from the surface of the cyclin T1 bound CDK9 structure

conducted, it buries a large surface area upon complex formation (81.5 $Å^2$), forms contacts with six amino acids from the surface of CDK9, including one direct and one water bridged H-bond and is close to the Phe 146 hot spot in the primary sequence, so it was retained as a promising residue for inhibitor design. On the other hand, Gln 73 from the cyclin T1 sequence was excluded from further consideration in inhibitor design, even though it is identified as a hot spot ($\Delta\Delta G$=2.4 kcal mol^{-1}) in two out of three prediction tests, as it only interacts with the flexible N-terminus of CDK9. While favorable in the P-TEFb complex, this pattern of interaction would likely come with a high entropic cost if it were mimicked in a peptide inhibitor - CDK9 complex. The residues responsible for the majority of binding energy of the CDK9/cyclin T1 interaction are Lys 93, Glu 137 and Phe 146 from the cyclin T1, and Phe 12 from the surface of CDK9. Phe 12 can interact with cyclin T1 in part through its backbone atoms, forming two hydrogen bonds to a water molecule bridging the interface gap.

In light of the above analysis, computational design of potential peptide inhibitors of the CDK9 - cyclin T1 interaction was attempted, focusing on Lys 93, Glu 137 and Phe 146 as main interaction sites. These residues were chosen as they were identified as hot spots in at least two out of the three prediction tests conducted (all three in the case of Phe 146) and about 40–58 % of their surface is buried upon complex formation. The choice of Lys 93 as an important hot-spot residue is in agreement with the experimental mutagenesis studies in which its mutation, alongside the mutation of Glu96 was shown to inhibit P-TEFb complex formation. Glu96 was not considered in our interface analysis, as this position is mutated to Gly in the crystallographic structure of P-TEFb

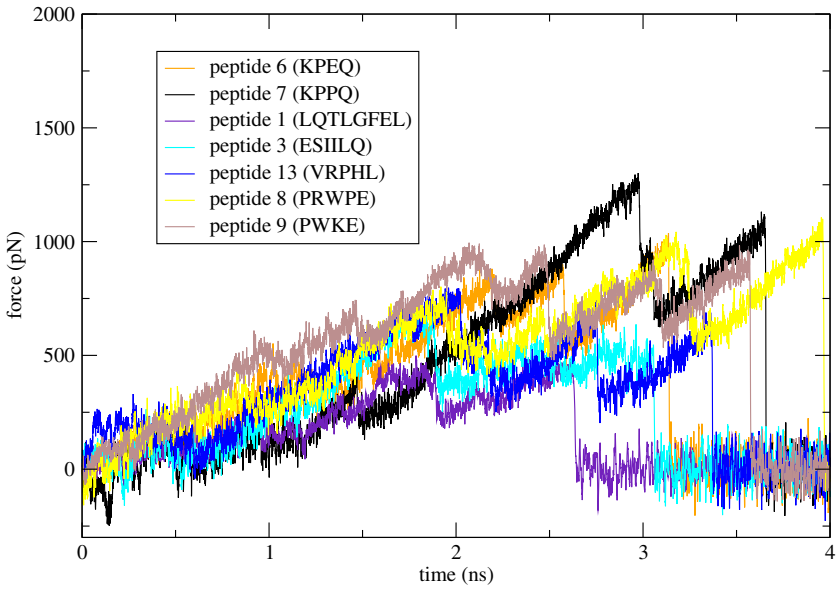

Fig. 4 Force profiles obtained in SMD runs for peptides pulled from the surface of the «free» CDK9 structure

used in our design. This mutation is natively present in some species [20] and does not disturb the structure significantly, nor does it alter ATP-binding [57].

The decision to focus on the cyclin T1 sequence region 137–148 (hot spot residues Glu 137 and Phe 146 in our design) is supported by an experimental observation that cyclin T1 segment consisting of residues 1–188 but not 1–133 was found necessary and sufficient for CDK9 binding [55]. The only residues from the 134–188 cyclin T1 sequence directy involved in the CDK9/cyclin T1 interaction interface are 134, 137, 139, 141–146, 148, 149, of which most are included in our designed peptides.

As the hot-spot residues chosen for our peptide design are conserved between cyclins T1, T2 and K (see Online resource, Fig. 5), cyclin-specific inhibition of the P-TEFb complex formation by the designed peptides is highly unlikely. However, known CDK9 inhibitors that bind to the ATP-binding pocket of CDK9 also lack this cyclin-based specificity in their action and have reached clinical trials [4].

Aside from the native hot spots originating from the surface of cyclin T1, amino acids from the surface of the Tat protein were also investigated as anchor spots for potential inhibitors, as Tat protein interacts with CDK9 in the P-TEFb/Tat complex. While most of the Tat-P-TEFb interaction is mediated through the cyclin T1 - Tat contacts, Tat does interact with the T loop of CDK9 and computational alanine scanning of this interaction indicated that tryptophane, Trp 11, from Tat contributes 3.82 kcal mol^{-1} of binding energy, which classifies it as a hot spot amino acid. Trp 11 interacts with the surroundings of the flexible T loop on the CDK9 surface, but its binding groove remains preserved in the MD simulations of the complex and the "free" CDK9 (Fig. 5), which makes it a potentially promising residue to include in the design of molecules binding to CDK9.

The Cα distance between Lys 93 hot spot from cyclin T1 and the Trp 11 from Tat is only 10.19 Å, allowing for the two residues to be linked with the addition of just three more amino acids. As a result, peptide inhibitors designed to imitate the interactions of these two residues could potentially be promising in P-TEFb inhibition. Design of such potential pentapeptide inhibitors is planned in our future studies.

Docking

Once the desired peptide candidate structures were assembled, it became necessary to generate initial peptide-protein complexes for further simulations and evaluations of the free energy of binding. For the peptides extracted from the surface of cyclin T1 (Table 2, peptides 1–4), this meant docking them to the binding surface of CDK9, while the peptides obtained through fragment based design (Table 2, peptides 5–10) already had initial conformations at the active site. These peptides were still redocked to maintain consistency. Docking of all the candidates was also needed to obtain unified preliminary values of binding energies as all the previously used programs had scored the candidates using their own scoring functions which were not directly comparable.

When presented with both the desired binding surface of the CDK9/cyclin T1 interface and the ATP binding pocket, some of the peptides docked exclusively in the ATP pocket, meaning that, in the top 10 and often even 20 docking poses, across all runs, no binding was detected on the desired interface surface. Such is the case for three out of four peptides forming the negative control (Table 3, peptides 11, 12, 14) as well as for peptides 5 (FWGH) and 10 (PRWKE) obtained by fragment based design. Peptides derived from the cyclin T1 surface (Table 3, peptides 1–4) generally interact better with the "bound" CDK9 conformation, than with its "free" one, while those obtained through fragment based design (Table 3, peptides 6–9) docked to both structures with slight preferences to the "free" CDK9

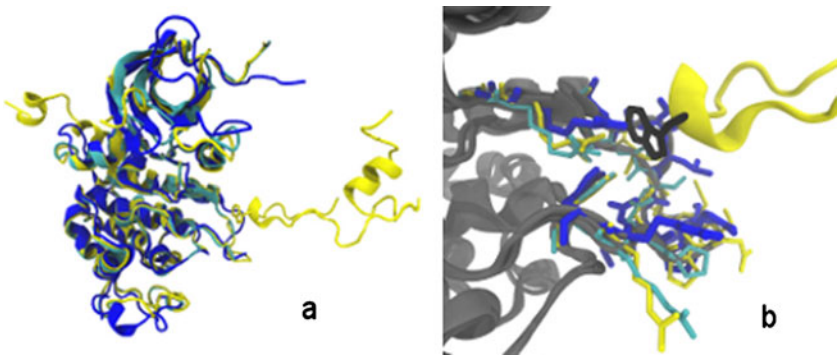

Fig. 5 CDK9/Tat interactions in P-TEFb/Tat complex. **a** Structure of P-TEFb/Tat complex (PDB ID: 3MI9) (*yellow*) superimposed to the cyclin T1 bound (*cyan*) and «free» (*blue*) CDK9 structures from our MD simulations. **b** close up of the hot spot Trp 11 residue (*black*) from Tat interacting with CDK9 (*yellow*, stick representation), cyclin T1 bound (*cyan*) and «free» (*blue*) CDK9 structures from our MD simulations

conformation. Peptide 13 (VRPHL) from the randomly assembled negative control group (peptides 11–14, Table 3) docked to both CDK9 conformations with affinities comparable to those of our designed peptides, but further energy evaluations using MM-PBSA method opposed this observation (see MM-PBSA). In all cases with successful docking, the structures were used as initial conformations for further analysis using MM-PBSA approach.

MM-PBSA

MM-GBSA and MM-PBSA binding energies obtained for the binding of the designed peptides to CDK9 are in good agreement. As MM-PBSA generally gives more accurate energies, this discussion is focused on MMPBSA results, but MM-GBSA calculations were conducted to obtain additional information and to test agreement between the two.

Due to the fact that only five frames from MD simulations were used to estimate the binding entropy of the peptides, the numerical values obtained for the changes in conformational entropy upon binding (TΔS) are not reliable, which prohibits the direct comparison of free energies of binding between the peptides. Nevertheless, the calculated changes in conformational entropy upon binding between different peptides are comparable and in the range of −20 to −28 kcal mol^{-1} (±10–15 kcal mol^{-1}), allowing us to rank the peptides according to their calculated MM-PBSA energies without the entropic contribution. Direct inclusion of numerically unreliable TΔS values into the evaluation of binding affinities and ranking of the peptides was avoided as it would introduce significant errors in the prediction. Therefore, Table 4 and the following discussion refer to MM-PBSA calculated energies without the changes in conformational entropy upon binding (TΔS values) accounted for.

Among the peptides predicted to bind to the surface of the CDK9 obtained from the 4 ns P-TEFb MD simulation, peptides derived from the cyclin T1 surface (Table 4, peptides 1–4) were shown to bind better than those obtained with fragment based design (Table 4, peptides 5–10). The only peptide from the randomly built negative control group to bind to the CDK9 was peptide 13 (VRPHL), with the binding energy of −18.8±3.8 kcal mol^{-1}. Peptide 4 (LQTLGF) showed the most favorable binding energy of −23.4±4.4 kcal mol^{-1}, reproducing the binding mode of the Phe 146 hot spot from the cyclin T1 surface, with an additional hydrogen bond formed by the free carboxylate of its Phe residue. Interactions of peptide 4 and CDK9 residues are shown in Fig. 6. Peptide 4's glutamine residue (L-Q-TLGF), corresponding to residue 142 in the cyclin T1 sequence, forms three hydrogen bonds to CDK9 residues Glu 83 and Ile 84, further stabilizing the binding (Fig. 6). This justifies our decision to include cyclin T1's Gln 142 residue in peptide design, despite it not being a hot spot.

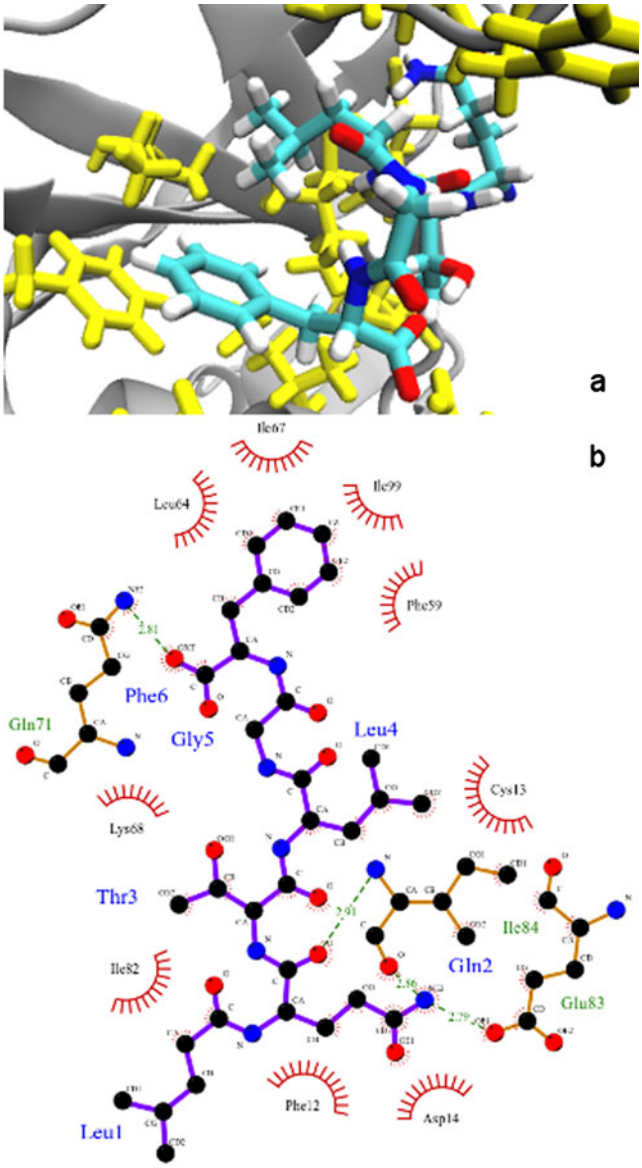

Fig. 6 Peptide 4 (LQTLGF) bound to the CDK9 structure from P-TEFb simulation. **a** binding conformation. **b** peptide 4 - CDK9 interaction diagram mapped out by LigPlot+, hydrogen bonds in *green*

De novo designed peptides (Table 4, peptides 5–10) performed somewhat better when binding to the surface of the "free" CDK9, but the best binding energy of −40.9±5.6 kcal mol^{-1} was obtained with peptide 3 (ESIILQ), a cyclin-T1 surface derived peptide. Figure 7 shows the interactions between peptide 3 and the surrounding CDK9 residues. The majority of peptide 3's predicted interactions with the surface of CDK9 are localized around peptide 3's C-terminus, where its free carboxylate forms a network of hydrogen bonds with CDK9 residues Phe 12, Cys 13 and Arg 86 (Fig. 7b).

The most promising inhibitor candidate obtained through fragment based design was peptide 8 (PRWPE) which is

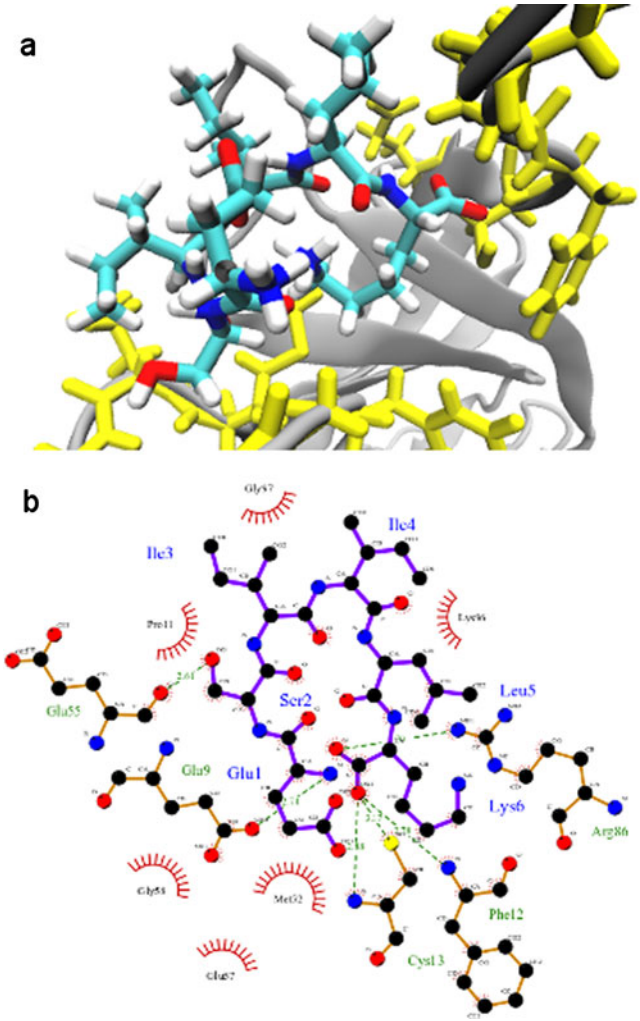

Fig. 7 Peptide 3 (ESIILQ) bound to the "free" CDK9. **a** binding conformation. **b** peptide 3 - CDK9 interaction diagram mapped out by LigPlot+, hydrogen bonds in *green*

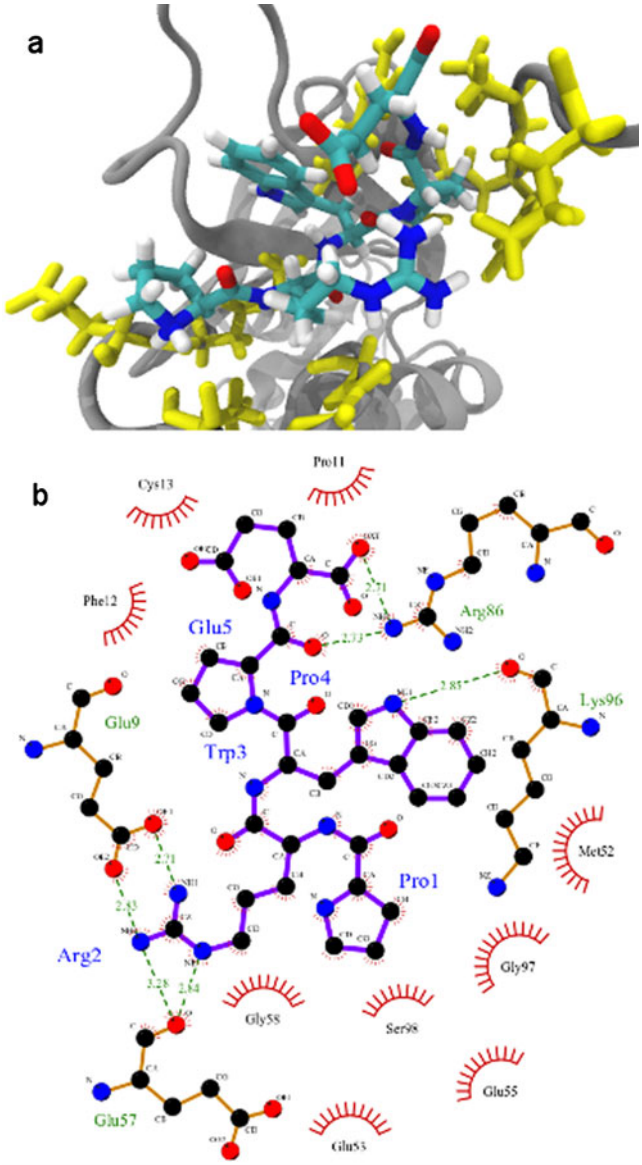

Fig. 8 Peptide 8 (PRWPE) bound to the "free" CDK9. **a** binding conformation. **b** peptide - CDK9 interaction diagram mapped out by LigPlot+, hydrogen bonds in *green*

amongst the highest ranking candidates when binding to the free CDK9 structure (calculated binding energy of −33.9± 7.5 kcal mol^{-1}). Interactions of peptide 8 and CDK9 residues are presented in Fig. 8. Two arginine residues figure prominently in the predicted interactions between peptide 8 and CDK9 surface: peptide 8's arginine residue (forming a hydrogen bond network with CDK9 residues Glu 9 and Glu 57) and CDK9 residue Arg 86, which interacts with peptide 8's C-terminus (Fig. 8b).

The differences in binding surfaces of CDK9 structures from the P-TEFb MD simulation and CDK9 alone MD simulation have influenced the binding of the designed peptides. As a result, peptide 4 (LQTLGF) was the best potential inhibitor candidate when the binding was performed on the structure of CDK9 from the P-TEFb simulation, but it failed to bind to the free CDK9 structure. On the other hand, peptide 3 (ESIILQ), interacted with the free CDK9 better than any of the other peptides tested, but failed to bind to the CDK9 structure in P-TEFb bound conformation. Of the peptides predicted to bind to both conformations of CDK9, peptide 8 (PRWPE) was the most promising potential inhibitor candidate, preferentially binding to the free form of CDK9.

SMD

In the literature, steered molecular dynamics (SMD) [58] has been used to estimate the order of relative binding affinity of small molecule ligands [59] and proteins [60]. Coupled with the Jarzynsky's equality, it can be used for more precise estimates of free energy [58], but we have only

used the method for approximate estimates of relative binding affinity of our designed peptides.

Binding energies obtained using MM-PBSA are not directly comparable with the force profiles of SMD simulations. Different force fields were used for the two methods in this study and in addition, the molecular events described in both can differ significantly, depending on the evolution of the SMD trajectories. During the 4th and final nanosecond of classical MD simulations from which data was obtained for the MM-PBSA runs, the peptides had remained bound to the CDK9 surface, while in SMD runs, due to the highly flexible N-terminus of the CDK9 in the vicinity of the binding surface, additional interactions of the peptides and CDK9 could form during the pulling process. This is most visible in the interaction of peptide 7 (KPPQ) and the "free" CDK9 structure. In the MM-PBSA run, the binding energy of this peptide-protein pair is evaluated to just −12.1 ±4.5 kcal mol^{-1}, but its force profile shows the highest peak of all the designed peptides (Fig. 3), indicating that it possesses the highest binding affinity. However, the peak on the force profile corresponds to a later event during the pulling, in which the peptide has left the binding surface and interacts with the N-terminus and a flexible loop of the CDK9 (Fig. 9). Therefore, the force needed to pull peptide 7 (KPPQ) directly from its binding surface is only described in the initial segment of the SMD force profile graph, and is in agreement with the low MM-PBSA predicted binding energy. The same behavior is observed with peptide 2 (FLAAKV) while binding to the CDK9 structure from the P-TEFb simulation, and to one degree or another, presents a common feature in our runs, due to the flexible N-terminus of the protein in the vicinity of the binding surface. While keeping this terminus fixed would eliminate this behavior, it would, at the same time, interfere with the interactions studied, as it would artificially constrain a region of the protein directly involved in the interaction.

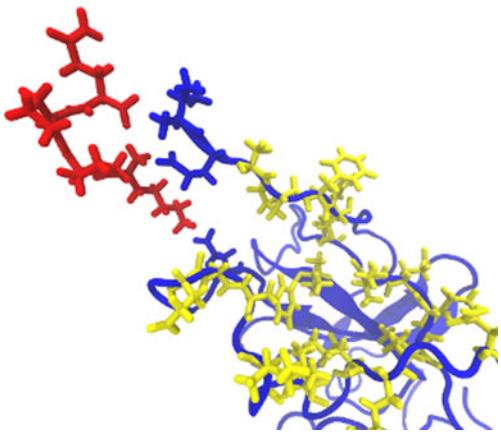

Fig. 9 Interaction corresponding to the highest peak of the SMD force profile of peptide 7 (KPPQ) in *red*, interacting with CDK9 (*blue*). Binding surface residues are displayed in *yellow*

As a result of these additional interactions forming between the peptides and CDK9, and perhaps due to the fact that the binding surfaces with which the peptides initially interacted were lacking sufficient overlap, the SMD force profiles obtained for the peptides could not be adequately used to estimate the differences in binding affinities between the peptides. Due to their flexibility, while being pulled, the peptides easily formed additional interactions with the protein surface, differing from case to case. This influenced the obtained force profiles, as the force peaks corresponding to these additional interactions altered the force profiles and masked the peaks expected from the initial interactions existing at the binding surface. We have, therefore, relied on the MM-PBSA predicted energies of binding as our main selection criterion for the peptide sequences that could be useful in the design of potential direct peptide inhibitors of CDK9/cyclin T1 interaction.

Conclusions

We have performed a computational study of the CDK9/cyclin T1 protein-protein interaction and identified two peptide sequences LQTLGF and ESIILQ, both derived from the surface of cyclin T1 as potential inhibitors of the interaction, predicted to bind to the surface of CDK9. In addition, peptide sequence PRWPE, obtained via fragment based design has also shown promising binding to CDK9 surface. While both MM-PBSA and SMD studies were conducted on the designed peptides – CDK9 complexes, the peptide sequences were selected based on their predicted binding energy calculated through the MM-PBSA approach only, as the inherent flexibility of the peptides and the N-terminus region bordering the binding surface have imposed limitations on the reliability of our SMD results. The results obtained would serve as the guidance for synthesis and biological evaluation of potential inhibitors of CDK9/cyclin T1 interaction and, along with additional computational studies, could contribute to the further exploration of P-TEFb/Tat complex as target for inhibitor design.

Acknowledgments The work makes use of results produced by the High-Performance Computing Infrastructure for South East Europe's Research Communities (HP-SEE), a project co-funded by the European Commission (under Contract Number 261499) through the Seventh Framework Programme HP-SEE (http://www.hp-see.eu/). The authors would like to thank the Ministry of Education and Science, Republic of Serbia for funding (project number 172009). J.R. would like to thank the Ministry of Education and Science, Republic of Serbia for PhD scholarship.

References

1. Romano G, Giordano A (2008) Role of the cyclin-dependent kinase 9-related pathway in mammalian gene expression and human diseases. Cell Cycle 7(23):3664–3668

2. O'Keeffe B, Fong Y, Chen D, Zhou S, Zhou Q (2000) Requirement for a kinase-specific chaperone pathway in the production of a Cdk9/cyclin T1 heterodimer responsible for P-TEFb-mediated tat stimulation of HIV-1 transcription. J Biol Chem 275:279–287
3. Napolitano G, Licciardo P, Carbone R, Majello B, Lania L (2002) CDK9 has the intrinsic property to shuttle between nucleus and cytoplasm, and enhanced expression of cyclin T1 promotes its nuclear localization. J Cell Physiol 192(2):209–215
4. Kryštof V, Baumli S, Fürst R (2012) Perspective of cyclin-dependent kinase 9 (CDK9) as a drug target. Curr Pharm Des 18 (20):2883–2890
5. Wang S, Fischer PM (2008) Cyclin-dependent kinase 9: a key transcriptional regulator and potential drug target in oncology, virology and cardiology. Trends Pharmacol Sci 29:302–313
6. Krystof V, Chamrad I, Jorda R, Kohoutek J (2010) Pharmacological targeting of CDK9 in cardiac hypertrophy. Med Res Rev 30:646–666
7. Chen R, Keating MJ, Gandhi V, Plunkett W (2005) Transcription inhibition by flavopiridol: mechanism of chronic lymphocytic leukemia cell death. Blood 106:2513–2519
8. Manohar SM, Rathos MJ, Sonawane V, Rao SV, Joshi KS (2011) Cyclin-dependent kinase inhibitor, P276-00 induces apoptosis in multiple myeloma cells by inhibition of Cdk9-T1 and RNA polymerase II-dependent transcription. Leukemia Res 35(6):821–830
9. Tahirov TH, Babayeva ND, Varzavand K, Cooper JJ, Sedore SC, Price DH (2010) Crystal structure of HIV-1 Tat complexed with human P-TEFb. Nature 465:747–751
10. Salerno D, Hasham MG, Marshall R, Garriga J, Tsygankov AY, Graña X (2007) Direct inhibition of CDK9 blocks HIV-1 replication without preventing T-cell activation in primary human peripheral blood lymphocytes. Gene 405(1–2):65–78
11. Chiu YL, Cao H, Jacque JM, Stevenson M, Rana TM (2004) Inhibition of human immunodeficiency virus type 1 replication by RNA interference directed against human transcription elongation factor P-TEFb (CDK9/CyclinT1). J Virol 78:2517–2529
12. Barboric M, Yik JHN, Czudnochowski N, Yang Z, Chen R, Contreras X, Geyer M, Peterlin BM, Zhou Q (2007) Tat competes with HEXIM1 to increase the active pool of P-TEFb for HIV-1 transcription. Nucleic Acids Res 35(6):2003–2012
13. Sedore SC, Byers SA, Biglione S, Price JP, Maury WJ, Price DH (2007) Manipulation of P-TEFb control machinery by HIV: recruitment of P-TEFb from the large form by Tat and binding of HEXIM1 to TAR. Nucleic Acids Res 35(13):4347–4358
14. Baumli S, Endicott JA, Johnson LN (2010) Halogen bonds form the basis for selective P-TEFb inhibition by DRB. Chem Biol 17 (9):931–936
15. Domling A (2008) Small molecular weight protein-protein interaction antagonists: an unsurmountable challenge? Curr Opin Chem Biol 12(3):281–291
16. Yin H, Hamilton AD (2005) Strategies for targeting protein–protein interactions with synthetic agents. Angew Chem Int Edit 44:4130–4163
17. Veselovsky AV, Archakov AI (2007) Inhibitors of protein-protein interactions as potential drugs. Curr Comput-Aid Drug 3:51–58
18. Gandhi L, Camidge DR, Ribeiro de Oliveira M, Bonomi P, Gandara D, Khaira D, Hann CL, McKeegan EM, Litvinovich E, Hemken PM et al (2011) Phase I study of Navitoclax (ABT-263), a novel Bcl-2 family inhibitor, in patients with small-cell lung cancer and other solid tumors. J Clin Oncol 29:909–916
19. Kuritzkes D, Kar S, Kirkpatrick P (2008) Maraviroc. Nat Rev Drug Discov 7:15–16
20. Baumli S, Lolli G, Lowe ED, Troiani S, Rusconi L, Bullock AN, Debreczeni JÉ, Knapp S, Johnson LN (2008) The structure of P-TEFb (CDK9/cyclin T1), its complex with flavopiridol and regulation by phosphorylation. EMBO J 27:1907–1918
21. Davis AM, Teague SJ, Kleywegt GJ (2003) Application and limitations of X-ray crystallographic data in structure-based ligand and drug design. Angew Chem Int Edit 42(24):2718–2736
22. Eyal E, Gerzon S, Potapov V, Edelman M, Sobolev V (2005) The limit of accuracy of protein modeling: influence of crystal packing on protein structure. J Mol Biol 351:431–442
23. Fiser A, Do RK, Sali A (2000) Modeling of loops in protein structures. Protein Sci 9(9):1753–1773
24. Fiser A, Sali A (2003) ModLoop: automated modeling of loops in protein structures. Bioinformatics 19(18):2500–2501
25. Olsson MHM, Søndergard CR, Rostkowski M, Jensen JH (2011) PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. J Chem Theory Comput 7(2):525–537
26. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K (2005) Scalable molecular dynamics with NAMD. J Comput Chem 26:1781–1802 http://www.ks.uiuc.edu/Research/namd/
27. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE III, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102:3586–3616
28. MacKerell AD Jr, Feig M, Brooks CL III (2004) Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. J Comput Chem 25:1400–1415
29. Humphrey W, Dalke A, Schulten K (1996) VMD - visual molecular dynamics. J Mol Graphics 14:33–38 http://www.ks.uiuc.edu/Research/vmd/
30. Case DA, Darden TA, Cheatham TE III, Simmerling CL, Wang J, Duke RE, Luo R, Walker RC, Zhang W, Merz KM, Roberts BP, Wang B, Hayik S, Roitberg A, Seabra G, Kolossváry I, Wong KF, Paesani F, Vanicek J, Liu J, Wu X, Brozell SR, Steinbrecher T, Gohlke H, Cai Q, Ye X, Wang J, Hsieh M-J, Cui G, Roe DR, Mathews DH, Seetin MG, Sagui C, Babin V, Luchko T, Gusarov S, Kovalenko A, Kollman PA (2010) AMBER 11. University of California, San Francisco
31. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins 65:712–725
32. Joung S, Cheatham TE III (2008) Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. J Phys Chem B 112:9020–9041
33. Rodriguez R, Chinea G, Lopez N, Pons T, Vriend G (1998) Homology modeling, model and software evaluation: three related resources. Bioinformatics 14(6):523–528
34. Le Guilloux V, Schmidtke P, Tuffery P (2009) Fpocket: an open source platform for ligand pocket detection. Bioinformatics 10:168
35. Kortemme T, Kim DE, Baker D (2004) Computational alanine scanning of protein-protein interfaces. Sci STKE 219:pl2
36. Kim DE, Chivian D, Baker D (2004) Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res 32(Web server issue):W526–W531
37. Zhu X, Mitchell JC (2011) KFC2: a knowledge-based hot spot prediction method based on interface solvation, atomic density and plasticity features. Proteins 79(9):2671–2683
38. Tuncbag N, Keskin O, Gursoy A (2010) HotPoint: hot spot prediction server for protein interfaces. Nucleic Acids Res 38(Web server issue):W402–W406
39. Meireles LMC, Dömling AS, Camacho CJ (2010) ANCHOR: a web server and database for analysis of protein–protein interaction

binding pockets for drug discovery. Nucleic Acids Res 38(suppl 2):W407–W411
40. Teyra J, Doms A, Schroede M, Pisabarro MT (2006) SCOWLP: a web-based database for detailed characterization and visualization of protein interfaces. Bioinformatics 7(1):104
41. Teyra J, Paszkowski-Rogacz M, Anders G, Pisabarro MT (2008) SCOWLP classification: Structural comparison and analysis of protein binding regions. Bioinformatics 9:9
42. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23(21):2947–2948
43. Majeux N, Scarsi M, Apostolakis J, Ehrhardt C, Caflisch A (1999) Exhaustive docking of molecular fragments with electrostatic solvation. Proteins 37:88–105
44. Dey F, Caflisch A (2008) Fragment-Based de Novo Ligand Design by Multiobjective Evolutionary Optimization. J Chem Inf Model 48(3):679–690
45. Lovell SC, Word JM, Richardson JS, Richardson DC (2000) The Penultimate Rotamer Library. Proteins 40:389–408
46. Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. J Comput Chem 31:455–461
47. AutoDock Tools http://autodock.scripps.edu/resources/adt/index_html Accessed 8 June 2012
48. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S et al (2000) Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. Acc Chem Res 33:889–897
49. Onufriev A, Bashford D, Case DA (2004) Exploring protein native states and large-scale conformational changes with a modified generalized Born model. Proteins 55:383–394
50. Laskowski RA, Swindells MB (2011) LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. J Chem Inf Model 51:2778–2786
51. Viet MH, Ngo ST, Lam NS, Li MS (2011) Inhibition of Aggregation of Amyloid Peptides by Beta-Sheet Breaker Peptides and Their Binding Affinity. J Phys Chem B 115(22):7433–7446
52. Kisseljova K, Kuznetsov A, Baudy-Floc'h M, Järv J (2011) Effect of two simultaneous aza-β3-amino acid substitutions on recognition of peptide substrates by cAMP dependent protein kinase catalytic subunit. Bioorg Chem 39(4):133–137
53. Lüdemann SK, Lounnas V, Wade RC (2000) How do substrates enter and products exit the buried active site of cytochrome P450cam? 1. Random expulsion molecular dynamics investigation of ligand access channels and mechanisms. J Mol Biol 303(5):797–811
54. Bogan AA, Thorn KS (1998) Anatomy of hot spots in protein interfaces. J Mol Biol 280(1):1–9
55. Fraldi A, Licciardo P, Majello B, Giordano A, Lania L (2001) Distinct regions of cyclinT1 are required for binding to CDK9 and for recruitment to the HIV-1 Tat/TAR complex. J Cell Biochem 81(S36):247–253
56. Bieniasz PD, Grdina TA, Bogerd HP, Cullen BR (1999) Recruitment of cyclin T1/P-TEFb to an HIV type 1 long terminal repeat promoter proximal RNA target is both necessary and sufficient for full activation of transcription. PNAS 96(14):7791–7796
57. Baumli S, Hole AJ, Noble MEM, Endicott JA (2012) The CDK9 C-helix exhibits conformational plasticity that may explain the selectivity of CAN508. ACS Chem Biol 7(5):811–816
58. Izrailev S, Stepaniants S, Isralewitz B, Kosztin D, Lu H, Molnar F, Wriggers W, Schulten K (1998) In: Deuflhard P, Hermans J, Leimkuhler B, Mark A E, Reich S, Skeel RD (eds) Computational molecular dynamics: challenges, methods, ideas, volume 4 of Lecture notes in computational science and engineering. Springer, Berlin, Heidelberg, New York, pp 39–65
59. Colizzi F, Perozzo R, Scapozza L, Recanatini M, Cavalli A (2010) Single-molecule pulling simulations can discern active from inactive enzyme inhibitors. J Am Chem Soc 132(21):7361–7371
60. Cuendet MA, Michielin O (2008) Protein-protein interaction investigated by steered molecular dynamics: the TCR-pMHC complex. Biophys J 95(8):3575–3590