

In silico design of small molecule inhibitors of CDK9/cyclin T1 interaction



Jelena Randjelovic ^{a,b,*}, Slavica Eric ^b, Vladimir Savic ^{a,*}

^a University of Belgrade, Faculty of Pharmacy, Department of Organic Chemistry, Vojvode Stepe 450, 11221 Belgrade, Serbia

^b University of Belgrade, Faculty of Pharmacy, Department of Pharmaceutical Chemistry, Vojvode Stepe 450, 11221 Belgrade, Serbia

ARTICLE INFO

Article history:

Accepted 3 April 2014

Available online 13 April 2014

Keywords:

CDK9/cyclin T1

Druggability mapping

Protein–protein interactions

Accelerated molecular dynamics

MM-GBSA

ABSTRACT

In order to design a small molecule which potentially may interfere with CDK9/cyclin T1 complex formation and therefore influence its physiological role, a computational study of dynamics and druggability of CDK9 binding surface was conducted. Druggability estimates and pocket opening analyses indicated binding regions of cyclin T1 residues, Phe 146 and Lys 6, as starting points for the design of small molecules with the potential to inhibit the CDK9/cyclin T1 association. A pharmacophore model was created, based on these two residues and used to select potential inhibitor structures. Binding energies of the inhibitors were estimated with MM-GBSA. A good correlation of MM-GBSA energies and FTMap druggability predictions was observed. Amongst studied compounds a derivative of 2-amino-8-hydroxyquinoline was identified as the best potential candidate to inhibit CDK9/cyclin T1 interactions.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Two isoforms of cyclin-dependent kinase 9 (CDK9) are present in humans, differing in the structure of their N-terminus [1]. Both of them, in complex with cyclins T1, T2 and K can form the core of P-TEFb (positive transcription elongation factor b) protein complex [2], which promotes the elongation of RNA polymerase II transcripts.

P-TEFb activity is significant in different pathological states in humans, including several types of cancers, HIV-1 infection, cardiac hypertrophy and chronic inflammatory conditions [3–7]. The inhibition of P-TEFb function interferes with the general transcription levels in the cell, resulting in reduced concentration levels of various proteins. Short-lived proteins with a high turnover rate are the first to be affected by changes in cellular transcription. As tumor cells in certain cancers such as chronic lymphocytic leukemia [8], multiple myeloma [9] and osteosarcoma [10] survive by constantly synthesizing short-lived anti-apoptotic proteins, inhibition of P-TEFb function leads to apoptosis in these cells [8,9,11]. In addition,

inhibition of CDK9 activity and therefore P-TEFb function interferes with the process of angiogenesis [12].

Several viruses require P-TEFb function for their replication in cells, as they exploit the cellular transcription machinery for their own processes. HIV-1 virus recruits the CDK9/cyclin T1 complex specifically [13]. In several studies, inhibition of P-TEFb function led to suppressed HIV-1 transcription and replication without affecting the survival and proliferation of host cells [14,15].

Increased P-TEFb function plays a role in the pathogenesis of cardiac hypertrophy [16], which makes the inhibition of its function a potential therapeutic goal in cardiology [6,17].

In chronic inflammatory diseases, such as rheumatoid arthritis, CDK9 activity controls the neutrophil life-span, by maintaining the levels of short-lived Mcl1 protein [7]. Inhibition of CDK9 activity in such states could lead to reduced number of active neutrophils, reducing the resulting tissue damage.

The most prominent method of inhibiting P-TEFb function used today is to directly inhibit the activity of its catalytic unit, CDK9, using molecules binding to its ATP-binding site. More than 30 ATP-competitive CDK9 inhibitors have been developed [6] and some have reached clinical trials [18]. These inhibitors target the conserved active site common to the CDK family of proteins and as a result, are often of limited selectivity [6,19]. While certain progress in developing more selective inhibitors of this type has been made [20], alternative methods of inhibiting P-TEFb function are also explored. In one such study, HIV-1 replication was inhibited using two molecules binding to the protein surface of CDK9 under the T-loop and disrupting the function of P-TEFb/Tat complex [21].

Abbreviations: CDK9, cyclin-dependent kinase 9; aMD, accelerated molecular dynamics; MM-GBSA, molecular mechanics generalized Born surface area.

* Corresponding authors at: Vojvode Stepe 450, 11221 Belgrade, Serbia phone: +381 11 3951 235.

E-mail addresses: jelenar@pharmacy.bg.ac.rs (J. Randjelovic), vladimir.savic@pharmacy.bg.ac.rs (V. Savic).

Another method of modulating the activity of P-TEFb would be to directly interfere with the CDK9/cyclin complex formation [22]. Monomeric CDK9 is quickly degraded in the cell [22] and its protein levels depend on the expression of its cyclin partners [15]. Interfering with the protein–protein interaction of a CDK/cyclin pair has already been shown to be a viable strategy of influencing the activity of the complex. A peptide inhibitor of CDK2/cyclin A complex was developed, based on the structure of cyclin A and experimentally verified [23]. The peptide bound to the complex itself and altered its conformation, inhibiting its function *in vitro*.

Our previous work was focused on the possibility of direct inhibition of CDK9/cyclin T1 interaction [24]. We conducted a computational study of the CDK9/cyclin T1 interaction interface and designed short peptides targeting the CDK9 binding surface. In this work we report the continuation of our CDK9/cyclin T1 inhibitor design efforts, focusing on the dynamics and druggability of CDK9 binding surface, and the design of more stable, small molecule potential inhibitors of the interaction, based on structural features of cyclin T1.

2. Computational methods

2.1. Druggability analysis

Preliminary druggability analysis was conducted on the crystal structure of CDK9 from the CDK9/cyclin T1 complex (PDB ID: 3blh) [25]. Representative CDK9 structures obtained via accelerated molecular dynamics (aMD) simulations were also mapped (see next section). All druggability analyses were conducted using only the N-terminal domain of CDK9, as described in the literature [26]. The CDK9 domain sequence to map was obtained from the CATH database [27]. The structure was prepared using VMD [28], and the druggability mapping was done through the FTMap web server [29] in Protein Surface mode. Results were visualized in Pymol [30] and VMD. VMD was used for all structural alignments and for rendering the figures shown in this work.

2.2. CDK9 binding surface dynamics

The starting CDK9 structure used for all simulations in this work is the CDK9 structure from the final frame of CDK9/cyclin T1 4 ns MD simulation from our previous work [24]. Four separate simulations were conducted in this work: simulations N1 and N2 on CDK9 structure missing the first seven N-terminal residues and simulations A and B on CDK9 with a modeled N-terminus. For simulations A and B, the starting CDK9 structure was extracted from its complex with cyclin T1 and the seven missing residues of the N-terminus (sequence MAKQYDS) were added in Pymol in an extended conformation. The modeled N-terminus was subjected to a short minimization (5000 steps) in vacuum, keeping the rest of the protein fixed and then the whole structure was solvated with a rectangular TIP3P water box with the box boundaries at least 15 Å from the protein, neutralized with Cl⁻ ions and prepared for MD. For simulations A and N1, the structure was prepared using the tleap module of AmberTools [31] software, using Cornell et al. forcefield version ff10 (ff99SB [32] for proteins) with Joung and Cheatham [33] adjustments for ions in TIP3P water model (frc-mod.ionsjc.tip3p). For simulations B and N2, the preparation was done in VMD, using the CHARMM22/CMAP forcefield [34,35].

In all four simulation lines, prepared CDK9 structures were kept fixed and the solvent alone was minimized for 50 000 steps, followed by another 50 000 steps of minimization over the whole system. The systems were slowly heated to 310 K in 0.1 K increments over 93 ps and then simulated in NPT ensemble for 5 ns. MD simulation parameters were: timestep of 1 fs, periodic boundary

conditions, 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. All simulations were done in NAMD 2.9 [36]. For RMSD C_α plots of the simulations, see Fig. SI 1 in supporting information.

Final 2000 frames from the above 5 ns classical MD simulations were used in each instance to calculate the average dihedral and total potential energy of the systems, needed for aMD simulations. The energies were calculated using the NAMDEnergy plugin in VMD. For both forcefields used, parameters needed for aMD simulations were calculated according to the protocol described in the literature [37]. aMD specific parameters for all simulations in this work are listed in Table SI 1 in Supporting Information. The literature protocol is for the Cornell et al. forcefield, but secondary structure elements in CDK9 remained stable in aMD simulations with parameters calculated in the same way for the CHARMM forcefield as well. Dual boost aMD simulations were conducted as implemented in NAMD [38]. Each simulation lasted 20 million steps, using the same MD simulation parameters as the classical MD simulations described above.

For validation of the procedure used for modeling and simulating the missing N-terminus of CDK9, a CDK8 structure was submitted to the same protocol (N-terminus modeling, minimization, heating, 5 ns MD simulation, aMD simulation). Starting CDK8 structure was obtained from the crystal structure of its complex with cyclin C (PDB ID: 3rgf) [39]. Missing loops in CDK8 were modeled using MODELLER [40] offered through the ModLoop web server [41] with default parameters. The protonation of the structure was checked with PROPKA web server [42] and adjusted to pH = 7.4. To obtain structures corresponding to the missing N-terminus in CDK9, residues 1–10 of the CDK8 N-terminus were deleted. Ten, and not seven residues were removed as there is a three-residue gap in the sequence alignment of CDKs 8 and 9 in that area. CDK8 N-terminus was modeled with the same procedure as CDK9 N-terminus. The aMD simulations of CDK8 lasted 7 000 000 steps each.

Ten thousand CDK9 conformations were recorded in each aMD simulation (every two steps) and sorted according to the boost potential applied to them. Two thousand conformations with the highest boost potential were kept for clustering. Clustering was done with wordom [43] software, using a qt-like clustering algorithm. Three different clustering cutoffs were tested (1, 1.5 and 2 Å) and RMSD values were determined for all heavy atoms of the CDK9 amino acids at the interface with cyclin T1, excluding the missing seven residues of the N-terminus (amino acids 8–13, 57–59, 61, 64, 67–68, 71, 81–84, 86, 99). Representative structures (centers) of top 10 clusters at 2 Å clustering cutoff from each simulation were used for FTMap druggability analysis and subsequent docking of ligands.

2.3. Pockets opening on CDK9 binding surface

Opening of pockets on the CDK9 binding surface to cyclin T1 was estimated on heavy-atoms aligned 10 000 CDK9 conformations collected from the aMD simulations (every two steps). The analysis was done using EPOSBP software [44]. Transient pockets were obtained by clustering together pockets detected in each of the conformations, sharing at least 75% of pocket lining atoms. Pockets were visualized in VMD.

Propensity for pocket opening on CDK9 interface with cyclin T1 was checked using a literature protocol [45] as implemented in Rosetta 3.5 software [46]. The analysis was conducted on the following CDK9 residues: 8–13, 57–59, 61, 64, 67, 68, 81–84, 99 (interface) and 21, 37, 76, 119, 129, 141, 146, 185, 216, 226, 232, 243, 249, 281, 302, 307, and 316 (surface). The control surface residues were selected using the literature criterion [45]: they are at

least 12 Å from the interface and at 12 Å away from each other. For each of the studied amino acids, 1000 conformations of CDK9 were generated with pocket opening around the target amino acid featuring as one of the optimization parameters. Further 1000 CDK9 conformations were generated with no bias in the optimization. For each CDK9 conformation the volume of the pocket present was measured 100 times and the maximum pocket found was reported. Fractions of CDK9 conformations containing pockets of different sizes were compared only among residues with matching secondary structure.

2.4. Ligand selection and docking

Pharmacophores used in ligand selection were defined using PocketQuery method [47], and the pharmacophore searches were done with AnchorQuery [48] and ZincPharmer [49]. The initial pharmacophore was based on cyclin T1 residues Phe 146 and Lys 6 and had five points: one aromatic, matching the side chain of Phe 146, two hydrogen bond donors on the backbone and side chain nitrogens of Lys 6, one hydrogen bond acceptor at the carbonyl oxygen of Phe 146 and a hydrophobic point describing the carbon atoms of the Lys 6 side chain. Obtained hits were manually sorted and five structures were selected. Molecular descriptors of the representative structures, needed for the PPI inhibitor likeness decision Tree 50 were calculated using e-dragon software [51].

Ligands and receptor conformations were prepared using AutoDock Tools [52] and docked with AutoDock Vina [53] software. Docking was done using a 20 Å × 20 Å × 20 Å binding site grid box centered on the binding positions of cyclin T1 residues Phe 146 and Lys 6, with the exhaustiveness parameter of 32. Top 20 poses were retained from each docking run.

2.5. MM-GBSA

In this work, MM-GBSA analysis was used for re-scoring the obtained docking results and for a more detailed analysis of ligand binding energies.

For the re-scoring of docking results, MM-GBSA was conducted on a single structure of each complex, after minimizing the complex for 2000 steps in implicit solvent as implemented in NAMD, with the solvent dielectric of 80 and concentration of ions in solvent of 0.2 M. For the analysis of ligand binding energies, each ligand–CDK9 complex was prepared for MD simulation as described for free CDK9 structure and simulated for 5 ns in explicit solvent.

Structures for the simulations were prepared using the LeaP module of AmberTools 13, with Cornell et al. forcefield ff10 for proteins and GAFF [54] forcefield for the ligands. Ligands were parametrized using the antechamber [55] module of AmberTools through acpype [56]. Ligand charges were derived through the R.E.D web server [57–59] using the multi-RBRA procedure with three conformations and six orientations per molecule. MD simulations were conducted using NAMD 2.9.

Two hundred frames from the last 2 ns of each 5 ns MD simulation were used for the analysis of ligand binding energies, sampled at 10 ps intervals. MM-GBSA analyses were done as implemented in the MMPBSA.py [60] script in AmberTools. Two different MM-GBSA models ($igb=2$ and $igb=5$) and two different salt concentrations were tested (0.1 and 0.15 M). Changes in conformational entropy upon binding were estimated using the nmode module of AmberTools on 5 structures for each of the complexes.

The same set of frames and MMPBSA.py script in AmberTools were also used for the additional prediction of binding energies using the MM-PBSA method [61] (salt concentration = 0.15 M and solute dielectric = 2).

Interaction diagrams for the complexes were mapped out using LigPlot + software [62].

2.6. Hydration

Prediction of conserved water locations around the ligand binding sites was done using the WaterDock procedure, as described in the literature [63]. The 15 Å × 15 Å × 15 Å boxes used in WaterDock procedure were centered on the docked ligand positions.

3. Results and discussion

In our previous study, a subset of cyclin T1 interface residues were identified as potentially useful starting points for the design of modulators of CDK9/cyclin T1 interactions [24]. Top three proposed peptides in that study bound to two distinct areas of the CDK9 binding surface. Peptide 1 (LQTLGF) was created to interact with the conformation of CDK9 extracted from the CDK9/cyclin T1 complex and was suggested to position in the vicinity of cyclin T1 Phe 146 hot spot residue. Peptides 2 (ESIILQ) and 3 (PRWPE), designed against a simulated conformation of free CDK9, bound in the vicinity of the CDK9 β 3- α C loop. To follow on these initial results, the main goal of this study was to design a small molecule with inhibitor properties of the CDK9/cyclin T1 interactions. Favorable starting points for the design were determined from the structure of the complex, as well as from the binding of previously designed peptide inhibitors.

Our previous conclusions were further scrutinized using other available methods of predicting druggability of a protein surface. We first conducted a preliminary druggability mapping of CDK9 N-terminal domain from the CDK9/cyclin T1 crystallographic structure used in our preceding studies (PDB ID: 3blh [25]). The employed FTMap methodology maps the protein surface with organic probes and reports locations in which different probes predominately bind [29]. FTMap found four consensus clusters of probes on the CDK9 interface surface (Fig. 1). Encouragingly, three out of four clusters, 14, 12a and 6 (numbers represent number of probes) overlapped with previously reported peptide 1 (LQTLGF) orientation on the binding surface (Fig. 1).

The largest cluster of probes (14 probes) coincided with the Phe 146 location on the interface. The only other cyclin T1 residue overlapping with the probe clusters was Lys 6 which overlapped with two clusters (with 12 and 6 probes, shown in red and gray, Fig. 1). Under the same mapping conditions, FTMap druggability mapping of the cyclin T1 surface from the crystallographic structure of the complex (PDB ID: 3blh) did not reveal any probe clusters on the binding surface involved in the CDK9/cyclin T1 interface. This may suggest that the binding surface of CDK9 presents a better target for the design of inhibitors of the aggregation of these two proteins.

None of the probe clusters identified on the CDK9 binding surface matched the literature criterion for indicating a hot spot of binding on a protein–protein interface (16 or more probes in a cluster) [64]. However, this analysis was conducted on a single CDK9 conformation without accounting for protein dynamics, so it was repeated after exploring the dynamics of free CDK9 using dual boost accelerated molecular dynamics (aMD) simulations [65,66].

Initially, CDK9 was simulated without the seven N-terminal residues whose coordinates are not resolved in the crystallographic structure of its complex with cyclin T1 (simulations N1 and N2, see Section 2). Given that aMD facilitates the exploration of a molecule's conformational space, not all conformations it provides are energetically favorable. During an aMD simulation a boost potential is applied only when the energy of the system falls beneath a predefined threshold. To roughly limit the analysis to the most favorable regions of the conformational space explored during aMD, only 20%

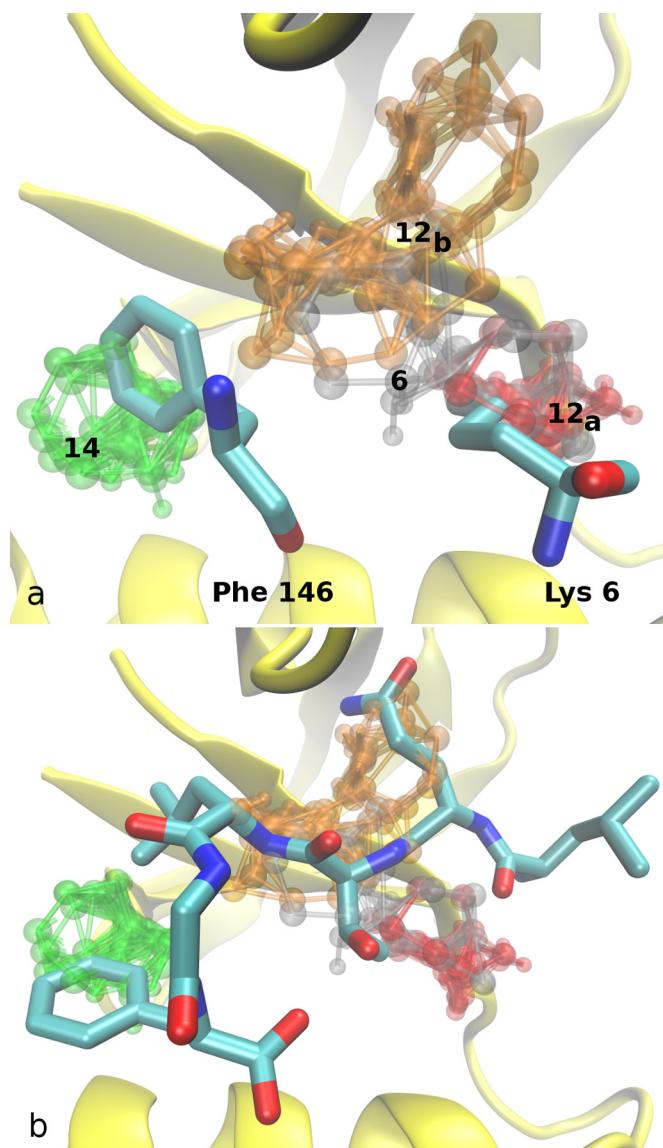


Fig. 1. Overlap of four clusters of probes (number of probes indicated) found on the CDK9 interface surface (yellow, PDB ID: 3blh) with cyclin T1 residues Phe 146 and Lys 6 (a) and with peptide 1 (LQLG) (b). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of all conformations were considered, those corresponding to the top 20% of applied boost potential during the simulation [67].

After clustering, 10 representative structures from simulations N1 and N2 were re-mapped with FTMap method. The results of this mapping differed from those obtained with the crystallographic CDK9 structure. The largest probe cluster on the binding surface of CDK9 was no longer located at the binding site of Phe 146 cyclin T1 residue. Instead, the Lys 6 binding site dominated the probe mapping, showing up as a hot spot cluster in all 10 structures examined from simulation N2 (with 18–34 probes) and in 7 out of 10 structures from simulation N1 (16–23 probes). The Phe 146 site was, instead, filled up with CDK9 residue Phe 59 and residues from the truncated N-terminus in both simulations and did not show up as a probe cluster site in any of the 20 CDK9 conformations mapped (Fig. 2).

As the N-terminus of CDK9 forms a part of its binding surface to cyclin T1, we decided to model it fully, by adding the missing seven starting amino acids. Modeling the N-terminus also explored the possibility that the increased flexibility of CDK9 residues 8–13

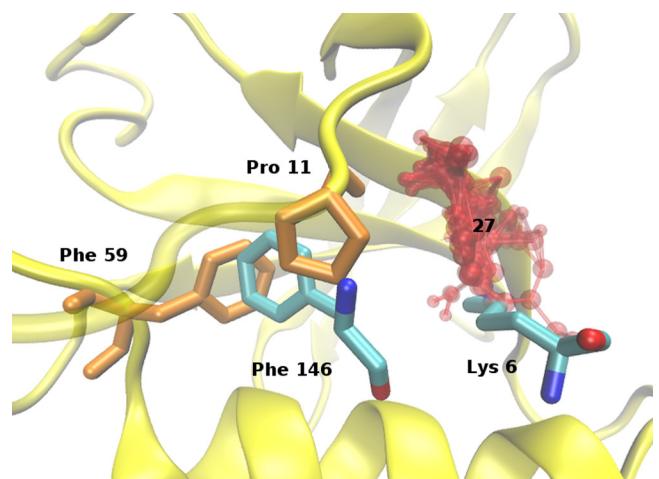


Fig. 2. FTMap mapping result of a representative CDK9 structure with truncated N-terminus. Main probe cluster location (red) and CDK9 residues overlapping the Phe 146 site (orange). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

caused by the truncation of its N-terminus is behind the observed druggability mapping results.

Two additional aMD simulations were conducted to simulate the CDK9 structure with the modeled N-terminus, using two different forcefields: Cornell et al. forcefield (simulation A) and CHARMM22/CMAP forcefield (simulation B).

During the aMD simulation B the N-terminus of CDK9 was observed to form a helical turn. In the available crystallographic structures of CDK9, in complex with cyclin T1, the N-terminus is disordered and partly absent. The alternate CDK9 isoform, CDK9_55 has an extended N-terminus of an additional 117 amino acids, but no tridimensional structure of this isoform has been solved. Among the other CDKs, an N-terminal helix was, to our knowledge, experimentally observed only in the structure of CDK8, another transcriptional CDK, in complex with cyclin C [39]. CDK8's N-terminal helix is present in the structure of the CDK8/cyclin C complex, which shows similarities to the CDK9/cyclin T1 complex structure [39]. As a result, the structure of CDK8 was used in this study to check the simulation conditions and procedure used to model the missing N-terminus of CDK9. Three aMD simulations were conducted on the free CDK8 structure, analogs of simulations A and B and a control simulation with the experimental N-terminus present. In simulations modeling it, CDK8 N-terminal helix formed fully under the conditions of simulation A and partially under the conditions corresponding to simulation B in the given time frame, indicating that the simulation conditions for its modeling were appropriate. The experimental N-terminus remained stable throughout the aMD simulation.

Ten representative structures from simulation A and B each were also mapped with the FTMap method. The inclusion of the entire CDK9 N-terminus in the simulations has led to the changes in the appearance of the binding surface. For conformations from both simulations (A and B) probe clusters appear at the binding sites of both Lys 6 and Phe 146. Out of 10 conformations from simulation A, 7 have probe clusters at the Lys 6 binding spot (2 of which are hot spot sites with 19 and 17 probes), and 4 at the Phe 146 site (no hot spots). Results are qualitatively similar for simulation B—6 out of 10 conformations have probes at the Lys 6 spot (3 hot spots, 32, 28, 17 probes) while probes appear at 5 out of 10 conformations for the Phe 146 site (one hot spot cluster, 23 probes).

Structures chosen for mapping were selected after all structures were clustered using the heavy atoms of all the CDK9 interface residues, except the modeled N-terminus (residues 1–7), which

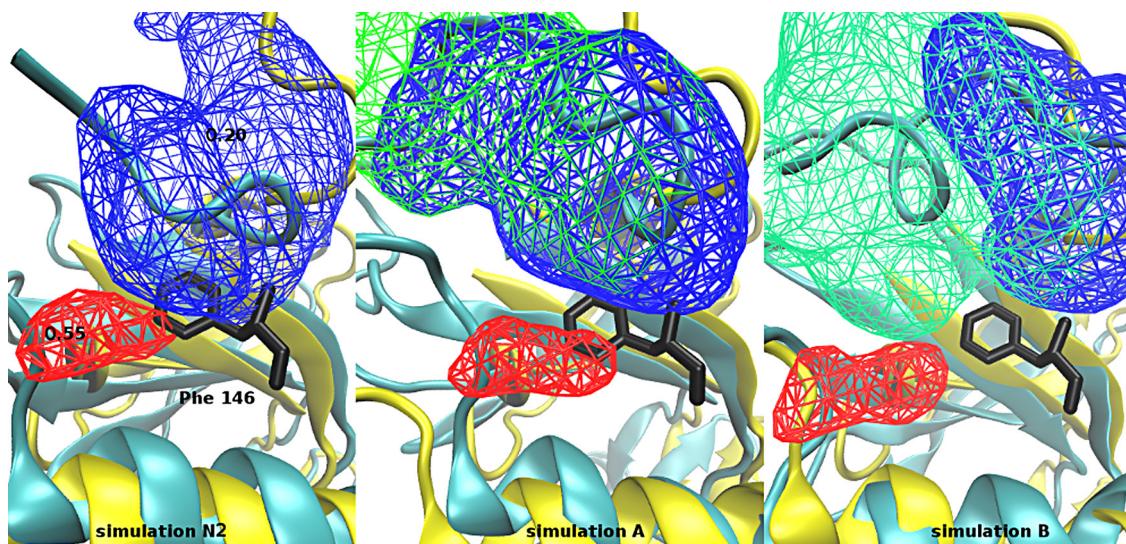


Fig. 3. Occupancy maps of CDK9 residues 59 (red, isovalue 0.55), 8–13 (blue, isovalue 0.20) and 1–13 (green, isovalue 0.20) across simulations N2, A and B. CDK9 structures from the simulations (cyan) are aligned with the CDK9 crystal structure (yellow) to show the location of Phe 146 from cyclin T1 (black). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

were excluded as their flexibility might skewer the results of RMSD clustering. However, to ensure that the results were not influenced by the choice of residues for clustering, the clustering and mapping were repeated using all the interface residues (including the modeled N-terminus) and similar results were obtained. To illustrate, for simulation B, with alternate clustering, 8 out of 10 structures featured probes at the Lys 6 binding spot (one hot spot) and 5 out of 10 had probes at the Phe 146 binding spot (no hot spots).

Taken together, the mapping results indicate that the introduction of the full N-terminus has led to the increased availability of the Phe 146 binding area on the CDK9 surface for ligand binding. Occupancy analysis conducted for residues 8–13 and residue Phe 59 of CDK9 across the four simulations corroborates this (Fig. 3). At same isovalue levels, spatial regions visited by the selected CDK9 residues show less overlap with the binding site of Phe 146 in simulations A and B, compared to simulation N2. In simulation N1, Phe 146 binding pocket is filled predominately by the truncated N-terminus (see Fig. SI 2 in Supporting Information).

Similar results were obtained when analyzing the opening of transient pockets on the CDK9 binding surface during the three

aMD simulations. Opening of a pocket at the Phe 146 binding location was almost not observed during simulations N1 and N2, the pocket rarely opened during simulation A and was short-lived, while it persisted in simulation B for about 55% of its duration. In simulations N1 and N2, the dominant pocket at the CDK9 binding surface, present for 65 and 75% of the simulation durations was formed in the region bordered by the truncated N-terminus, $\beta 3\alpha C$ loop and the loop containing residues 87–99. This pocket coincides with the binding area of peptides 2 (ESIILQ) and 3 (PRWPE) from our previous study. The missing section of the N-terminus could explain the predominance of this pocket, as it is formed in the region of its potential interactions with the loops listed. In accordance with this, pockets at this location are present only for 46% and 41% of the duration of simulations A and B and even then, take the form of narrow canals between the N-terminus and the loops. A third and final pocket that opened on the binding surface of CDK9 roughly corresponded to the binding location of Lys 6 and was present in all four simulations, N1, N2, A and B, at 34, 35, 66 and 48% of the time, respectively.

Although the above analyses had pointed at the same regions of CDK9 binding surface (regions binding Phe 146 and Lys 6 from

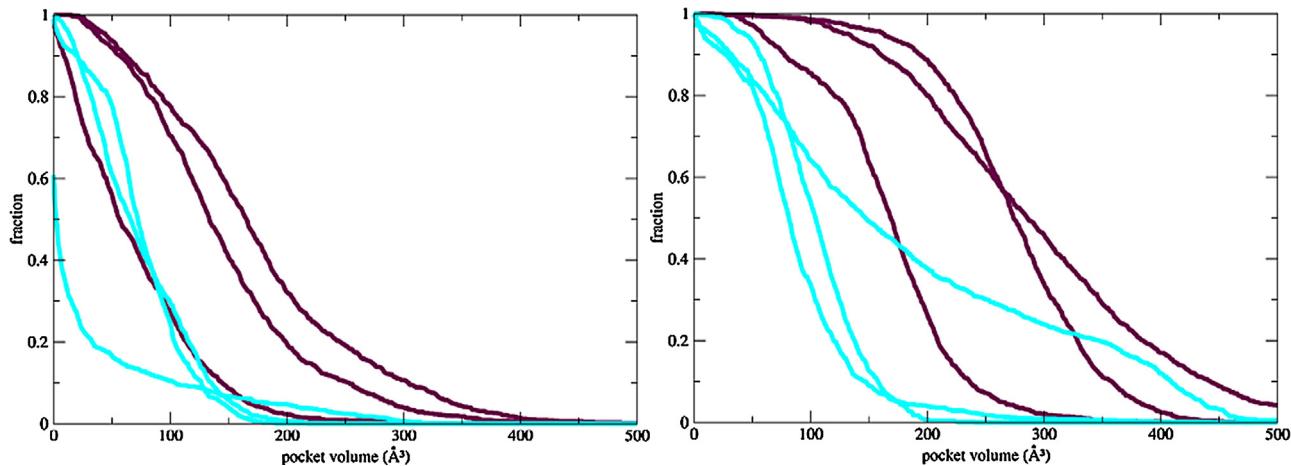


Fig. 4. Fraction of pocket-containing conformations of CDK9 at targeted interface (red) and random surface (cyan) residues, in unbiased (left) and biased (right) simulations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

cyclin T1) as most suitable for inhibitor development, an additional pocket and druggability analysis was conducted, for additional verification. The focus of this analysis was the propensity for opening of pockets around targeted residues, investigated through biased simulations [45]. The premise of the procedure is that druggable areas of protein surface are predisposed for pocket formation, and show a larger fraction of conformations containing pockets of a certain volume in the absence of ligands [45]. Simulations biased for pocket opening were conducted on CDK9 interface residues and randomly picked surface residues (see Section 2). Three interface residues, Ile 67, Leu 81 and Ile 84, showed marked shifts in pocket-containing conformations fraction in the biased simulations (Fig. 4). All three residues are located in the binding region of Lys 6 and Phe 146 residues from the cyclin T1 structure (Fig. 5), which further validates the choice of this region of CDK9 binding surface for inhibitor design. For full pocket opening results over all interface and surface residues analyzed, see Fig. SI 3 in Supporting Information.

On the surface of CDK9, binding regions of Phe 146 and Lys 6 residues from cyclin T1 were identified by FTMap druggability mapping as favorable for ligand binding and corresponded to the binding region of peptide 1 (LQTLGF) from our previous work. These two regions were also the sites of transient pockets opening on the CDK9 binding surface and were shown to be predisposed for pocket opening, which also implies their druggability. As a result, we decided to focus on the regions of binding of cyclin T1's residues Lys 6 and Phe 146 in the design of potential small molecule inhibitors of the CDK9/cyclin T1 interaction. The

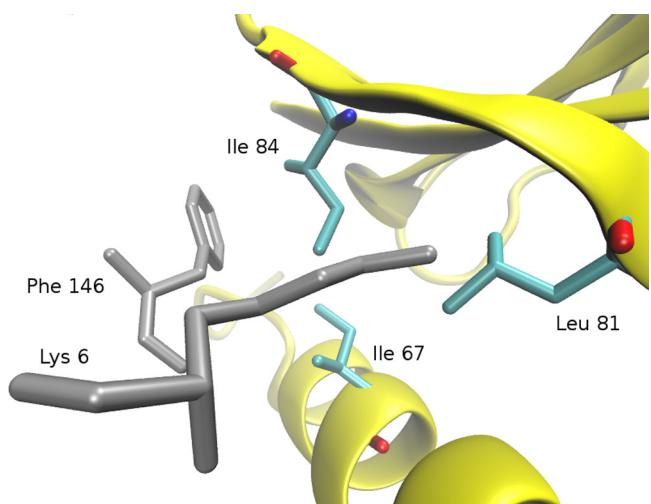


Fig. 5. CDK9 interface residues involved in pocket opening and their relative position to cyclin T1 residues Lys 6 and Phe 146.

design itself was based on defining a pharmacophore matching the features of these two residues and selecting the molecules satisfying its conditions. During the pharmacophore selection process, the choice of these two residues was further confirmed as they were listed as the top starting point for small molecule inhibitor design

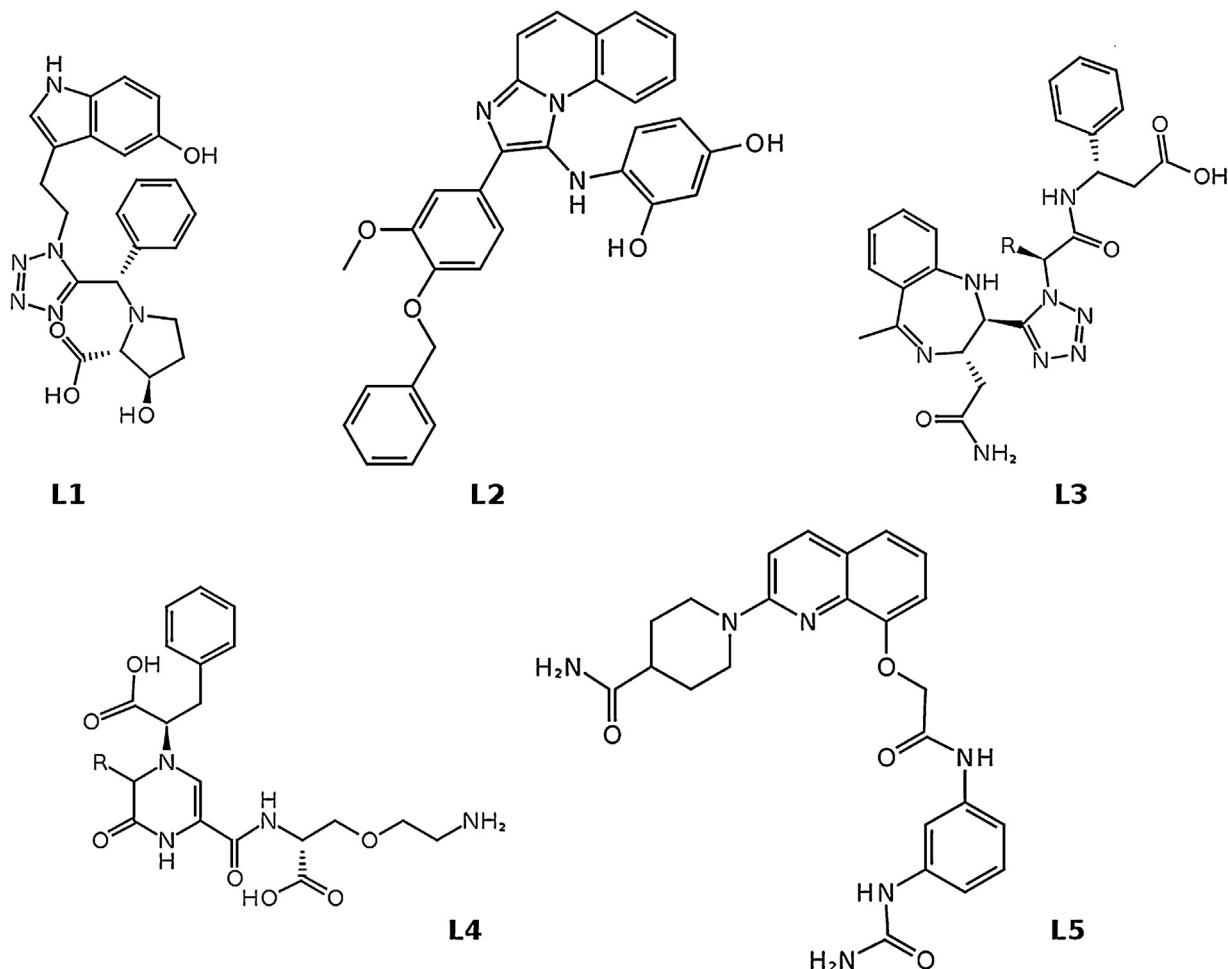


Fig. 6. Representative hits from the pharmacophore search. R groups in L3 and L4 signify substituents truncated to methyl groups to reduce steric interactions with the protein.

Table 1
Classification of potential ligands.

Ligands	RDF070m > 13.31	Ui > 4.13	Decision
L1	15.373	4.248	Yes
L2	13.834	5	Yes
L3	18.603	4.459	Yes
L4	12.178	4.087	No
L5	12.677	4.392	½

of the CDK9/cyclin T1 interaction originating from the cyclin T1 structure, selected by PocketQuery methodology.

The original pharmacophore used had five points based on the Phe 146 and Lys 6 features: aromatic core, matching the side chain of Phe 146, two hydrogen bond donors on the backbone and side chain nitrogens of Lys 6, one hydrogen bond acceptor at the carbonyl oxygen of Phe 146 and a hydrophobic point describing the carbon atoms of the Lys 6 side chain. This pharmacophore yielded three hits using AnchorQuery pharmacophore search. AnchorQuery was used as it is optimized for the discovery of protein–protein interaction inhibitors [48]. As an additional source of compounds ZincPharmer was used. To increase the number of hits, the hydrophobic point corresponding to the side chain of Lys 6 was excluded from the pharmacophore requirements and the search was repeated. Exclusion of the Lys 6 side chain hydrophobic point from the pharmacophore allowed us to retain the features of hot spot Phe 146 residue, and all H-bond acceptor and donor points. In the repeated search, 26 hits were obtained using AnchorQuery with two additional compounds chosen from ZincPharmer results for the same pharmacophore search (hits 27 and 28, see SI-structures in Supporting Information). As the obtained hits contained a number of related molecules, five groups of compounds were made based on their chemical similarity and representative compounds were selected. To reduce steric interactions with the protein surface without completely eliminating the substituents, 4-nitrophenyl and 1-methylpropyl groups in compounds L3 and L4 respectively, were truncated to methyl (Fig. 6).

Representative five compounds from each group were tested for similarity with known protein–protein interaction inhibitors using two molecular descriptors [50] (Table 1).

Two molecular descriptors used, RDF070 m (a radial distribution function descriptor) and Ui (unsaturation index), describe the molecular shape and number of unsaturated bonds respectively, and were shown to be of importance for inhibitors of protein–protein interactions [50]. Out of the five representative hits from the pharmacophore search, three passed both conditions of the decision tree. L4 did not satisfy either of the conditions and was removed from further analysis, while L5, as it satisfied one of the conditions was retained in the ligand group. The remaining four ligands were next docked to CDK9 and their binding energies were evaluated.

During the docking procedure, the dynamics of the CDK9 structure was accounted for by performing the docking on representative CDK9 conformations obtained during aMD simulations, an approach corresponding to the relaxed complex scheme described in the literature [68]. Docking was performed on representative CDK9 structures belonging to the 10 main conformation clusters from each of the four aMD simulations (N1, N2, A and B). These structures were obtained by clustering the 2000 lowest energy conformations according to the RMSD of the heavy atoms of interface residues, ignoring the modeled first seven residues from the N-terminus, as their flexibility would introduce additional noise during the clustering procedure. However, for completeness, an additional clustering was performed on structures from simulations A and B, and it included the heavy atoms of residues 1–7 along with the other interface residues. Additional 10

Table 2
Ligand poses chosen for the full MM-GBSA procedure.

Complex	Complex composition ^a	Docking binding energy [kcal mol ⁻¹]	MMGBSA energy (rescoring) [kcal mol ⁻¹]
L2_1	L2-B-2-1	-9.3	-35.42
L1_1	L1-B-2-1	-7.5	-34.92
L2_2	L2-A-1-4	-8.1	-34.40
L5_1	L5-B-1-6	-8.0	-33.50
L5_2	L5-B-2-1	-8.1	-32.98
L3_1	L3-B-2-7	-7.1	-29.86
L1_2	L1-B-1-1	-7.0	-25.58
L3_2	L3-N2-1-4	-6.7	-22.42

^a Ligand – simulation – clustering – conformation cluster number.

conformations obtained this way from simulation B were added to the pool of CDK9 conformations used for docking. With structures from simulation A, the alternate clustering provided unsatisfactory coverage of the trajectory space (top 20 clusters covered only 36.20% of trajectory space) and were not used further. In total, 50 CDK9 conformations were used for docking.

For each ligand and each subset of 10 CDK9 conformations used in docking, complexes with the top ranked ligand score were isolated and re-ranked using MM-GBSA (molecular mechanics generalized Born surface area) method on a single structure [69] (see Table SI2 in Supporting Information). Two top scoring poses for each ligand after the re-ranking were used in the full MM-GBSA procedure (Table 2).

As the docking and re-ranking of ligands was conducted without explicit water molecules present, prior to the full MM-GBSA energy evaluation, each of the ligand poses was checked for potential overlap with conserved water locations on the binding surface, as predicted using the WaterDock [63] method. According to this analysis ligand L5 in complex L5_1 was within 1.4 Å of two such water sites, but the clashes were resolved during the MD simulation prior to the MM-GBSA energy evaluation (Fig. 7). The remaining ligand poses showed no overlap with WaterDock predicted conserved water locations (see Fig. SI 4 in Supporting Information).

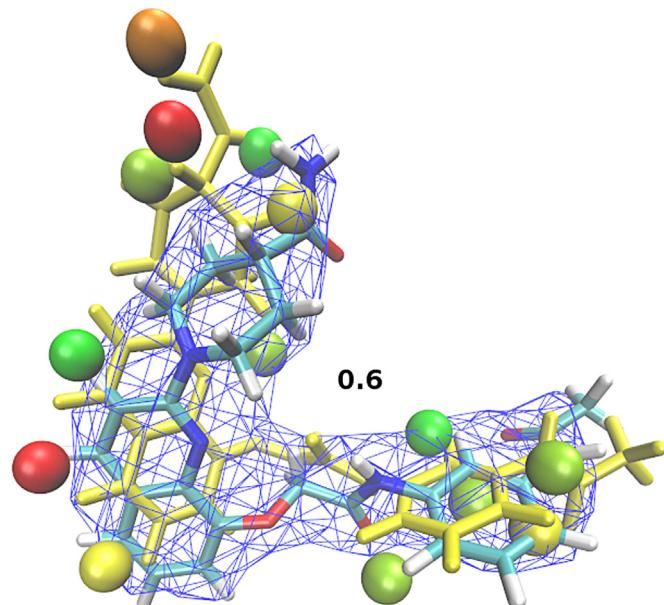


Fig. 7. Predicted conserved water locations (red spheres) around ligand L5 in complex L5_1. Docked conformation of ligand L5 (yellow) is in the vicinity of these water molecules, while the structure from the MD simulation (atom colored) is not. Occupancy is shown at isovalue 0.6 (blue mesh). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

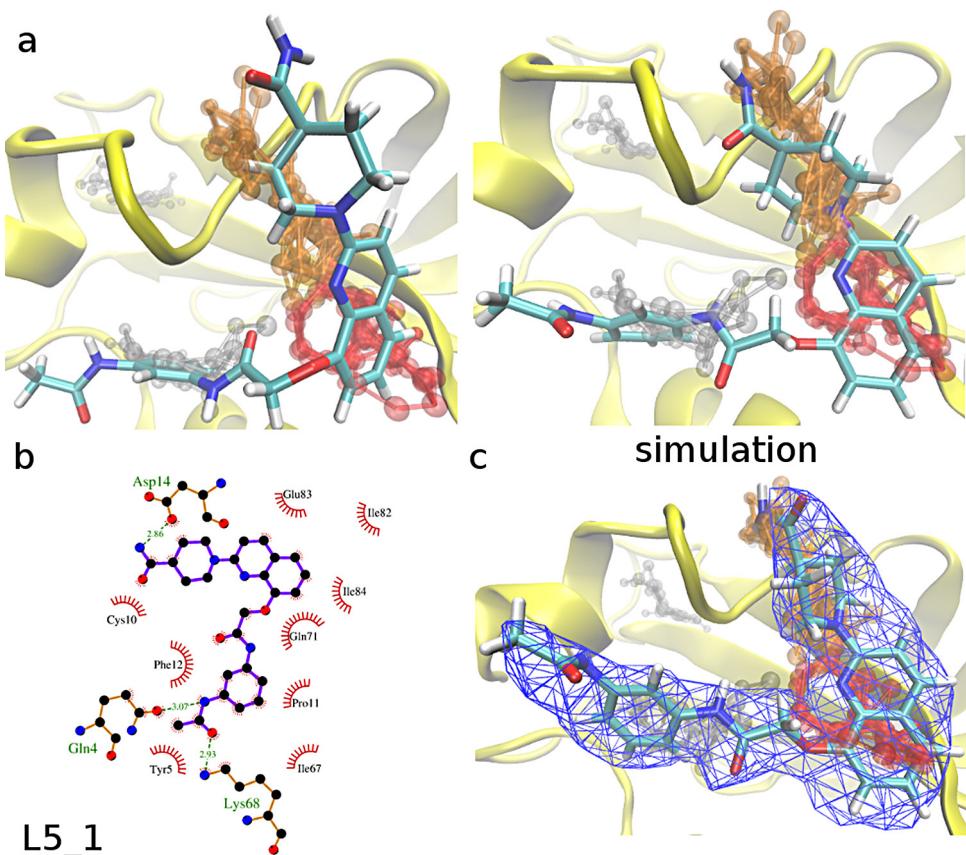


Fig. 8. Analysis of complex L5_1 (a) Overlap with FTMap predicted probe clusters (hot spot cluster shown in red) of the docked conformation of L5 (left) and conformation from the MD simulation (right) (b) interaction diagram of the L5_1 complex (c) occupancy map for ligand L5 during the simulation, at isovalue 0.6 (blue mesh).

Results of the full MM-GBSA energy analysis are shown in **Table 3**.

Changes in conformational entropy were comparable between ligands (about 20 kcal mol⁻¹) but had large errors (± 10 kcal mol⁻¹) associated with them, resulting from just five MD snapshots being

used in the calculations. As a result, the entropy terms were ignored and ligand ranking was based on the energy terms alone. Ligand L5 had the most favorable binding, followed by ligands L1 and L2. Complexes of ligand L3 showed the least favorable binding energy in the analysis. The same ligand ranking (L5 > L1, L2 > L3) was also obtained when binding energies were predicted using the MM-PBSA method (see Table SI 3 in Supporting information). MM-GBSA was chosen for the evaluation of ligand binding energies over MM-PBSA due to its speed and performance in ranking small molecule ligands [70].

When the intermolecular interaction of the ligands with CDK9 is viewed in terms of vdW and electrostatic energy (**Table 4**), ligand L5 shows the most favorable vdW interactions with the protein surface. Electrostatic interactions are significant with the more polar ligands L1 and L3, but their contribution to binding is abolished by the high desolvation penalty. This is especially true for the negatively charged ligand L3. Ligand L2 is at the other extreme, with only negligible electrostatic contributions to binding.

Table 3
MM-GBSA results.

Complex	MM-GBSA binding energy, entropy term excluded [kcal mol ⁻¹]
L5.1	-36.35 ± 3.93
L5.2	-35.65 ± 3.55
L1.1	-33.28 ± 4.55
L2.1	-32.60 ± 3.83
L2.2	-28.49 ± 2.94
L3.2	-26.42 ± 4.34
L1.2	-26.05 ± 3.95
L3.1	-21.70 ± 3.12

Table 4
Individual terms of MM-GBSA binding energy.

Complex	E_{vdW} [kcal mol ⁻¹]	E_{el} [kcal mol ⁻¹]	E_{GB} [kcal mol ⁻¹]	E_{surf} [kcal mol ⁻¹]	E_{total} [kcal mol ⁻¹]
L5.1	-46.17 ± 3.54	-40.19 ± 27.80	55.63 ± 23.91	-5.62 ± 0.33	-36.35 ± 3.93
L5.2	-45.45 ± 4.06	-25.96 ± 12.83	41.66 ± 11.26	-5.90 ± 0.34	-35.65 ± 3.55
L1.1	-37.22 ± 3.63	-59.58 ± 5.95	68.98 ± 4.94	-5.46 ± 0.40	-33.28 ± 4.55
L2.1	-42.98 ± 2.91	-16.26 ± 6.82	32.57 ± 5.19	-5.93 ± 0.30	-32.60 ± 3.83
L2.2	-39.04 ± 3.30	-2.01 ± 9.08	17.67 ± 8.82	-5.11 ± 0.36	-28.49 ± 2.94
L3.2	-35.01 ± 3.42	-140.98 ± 10.07	153.94 ± 9.22	-4.36 ± 0.35	-26.42 ± 4.34
L1.2	-30.39 ± 3.31	-52.01 ± 10.70	60.64 ± 9.66	-4.28 ± 0.35	-26.05 ± 3.95
L3.1	-35.95 ± 3.25	-133.92 ± 12.29	153.29 ± 10.95	-5.12 ± 0.31	-21.70 ± 3.12

E_{vdW} and E_{el} – van der Waals and electrostatic contributions to binding energy. E_{GB} and E_{surf} – electrostatic and nonpolar contributions to the solvation free energy. E_{total} – final calculated MM-GBSA binding energy.

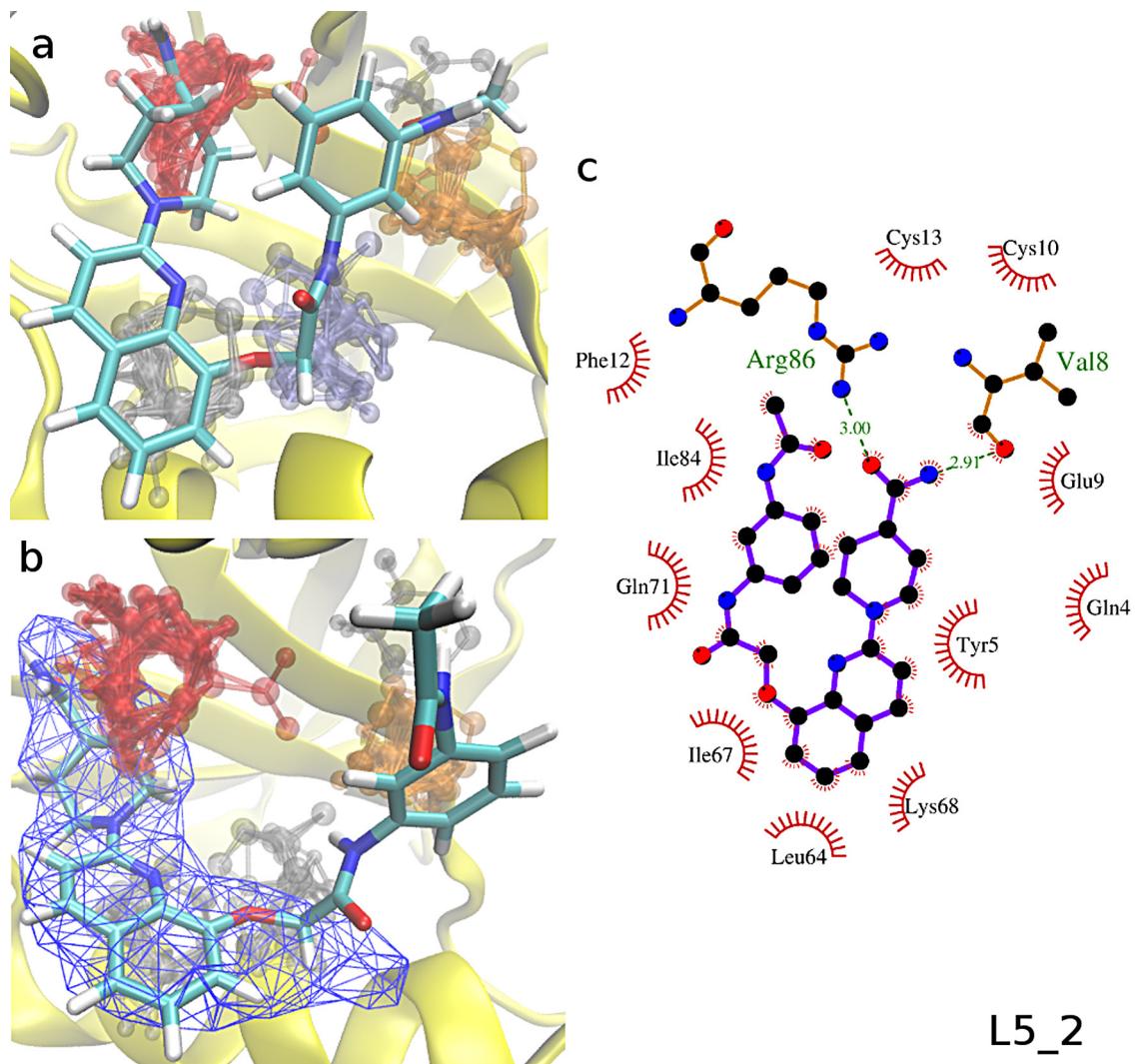


Fig. 9. Analysis of complex L5_2. Overlap with FTMap predicted probe clusters (hot spot cluster shown in red) of the docked conformation of L5 (a) and conformation from the MD simulation (b) with the occupancy map for ligand L5 during the simulation, at isovalue 0.6 (blue mesh) (c) interaction diagram of complex L5_2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

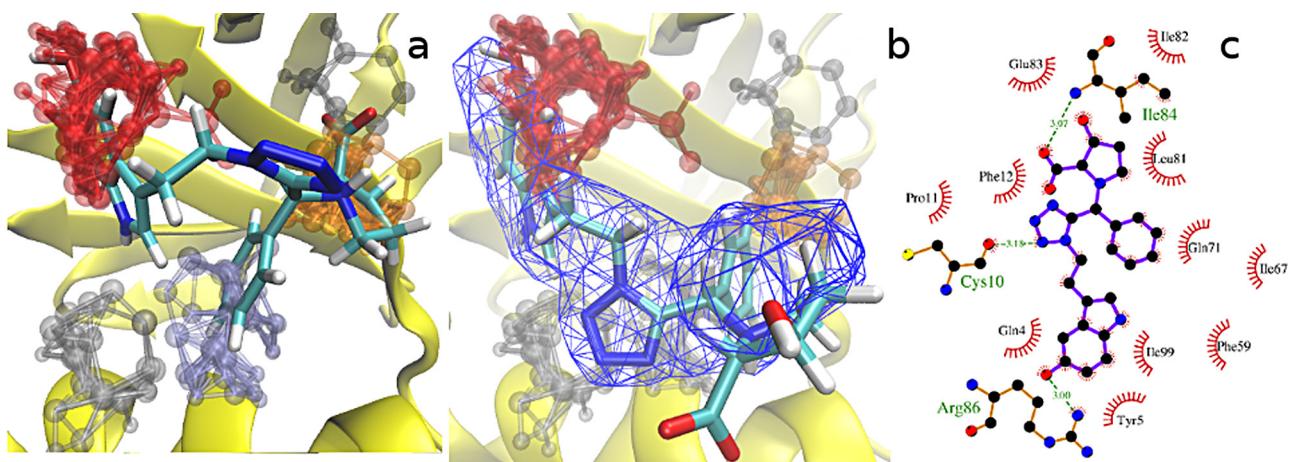


Fig. 10. Analysis of complex L1_1. Overlap with FTMap predicted probe clusters (hot spot cluster shown in red) of the docked conformation of L1 (a) and conformation from the MD simulation occupancy map for ligand L1 during the simulation, at isovalue 0.6 (blue mesh) (b) (c) interaction diagram of the starting complex. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

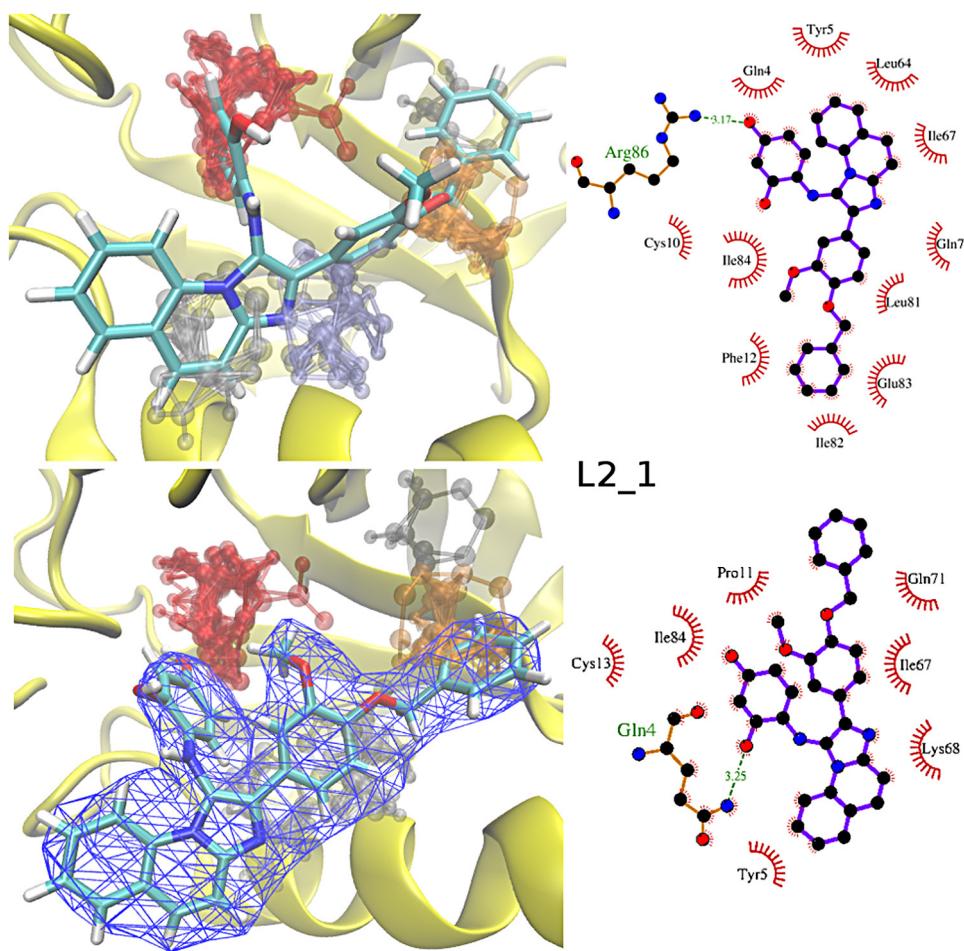


Fig. 11. Analysis of complex L2_1. On the left, docked (above) and conformation of L5 during the simulation (below) with an occupancy map for L5 (blue mesh, isovalue 0.6) and the overlap with FTMap probe clusters. On the right, interaction diagrams for the complex in the docked (above) and a typical conformation during the simulation (below). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Out of the four top scored complexes, three featured the same conformation of CDK9. This conformation is the representative of the largest cluster of energy-filtered structures from simulation B, indicating a relatively accessible conformational state of the protein in solution.

To ascertain the degree of agreement between the studied orientations of the ligands and the favorable areas of binding on the CDK9 surface, as predicted by FTMap, FTMap mapping procedure was repeated on CDK9 conformations from the eight selected complexes. Ligand overlap with hot spot probe cluster regions as described by FTMap correlated well with the obtained MM-GBSA binding energies, indicating that FTMap analysis can be used as an additional filter for proposed ligand orientations prior to more time consuming binding analyses.

For the top scored complex, L5_1, the ligand conformation's overlap with the FTMap predicted binding regions improves considerably during the MD simulation, in agreement with the favorable binding energy obtained (Fig. 8).

In its second orientation studied, the most important interaction of ligand L5 with CDK9 is a hydrogen bond to Arg 86. This interaction is analogous to those seen in the complexes of peptides 2 (ESIILQ) and 3 (PRWPE) from our previous study and persists throughout the MD simulation (Fig. 9). The region of interactions with Arg 86 also corresponds to the FTMap hot spot cluster overlapping with the orientation of ligand L5. Apart from it, the ligand only overlaps two weak clusters (shown in gray, Fig. 9).

Interactions with Arg 86 are also seen in the third and fourth ranked complexes in the MM-GBSA analysis, with ligands L1 and L2. Once again, the region of these interactions matches a FTMap hot spot cluster. In complex L1_1, interactions with Arg 86 are the only ones retained after the ligand alters its orientation during MD (Fig. 10).

After the initial re-scoring of docked poses, complex L2_1 was the top ranked ligand-CDK9 complex, but was ranked only fourth in the final analysis. A possible explanation for this is that the hydrogen bond to Arg 86 present in the starting conformation of the complex is lost during the MD simulation and replaced with interactions with the CDK9 N-terminus, leading to the observed drop in binding energy (Fig. 11). At the same time, the ligand loses its overlap with the FTMap hot spot cluster marking this interaction region.

As seen above, interactions with Arg 86 are a reoccurring feature of most of the ligand-CDK9 complexes examined, including complexes of peptides 2 and 3 from our previous work, contributing significantly to the predicted binding energies. However, care must be taken in assigning importance to this interaction, as the charged nature of Arg 86 allows it to form nonspecific electrostatic and ion-dipole interactions.

The agreement between MM-GBSA and FTMap holds true for all the examined complexes. Of the four complexes with the least favorable MM-GBSA binding energy, three show very little overlap with FTMap clusters. This is true for both of the complexes of ligand

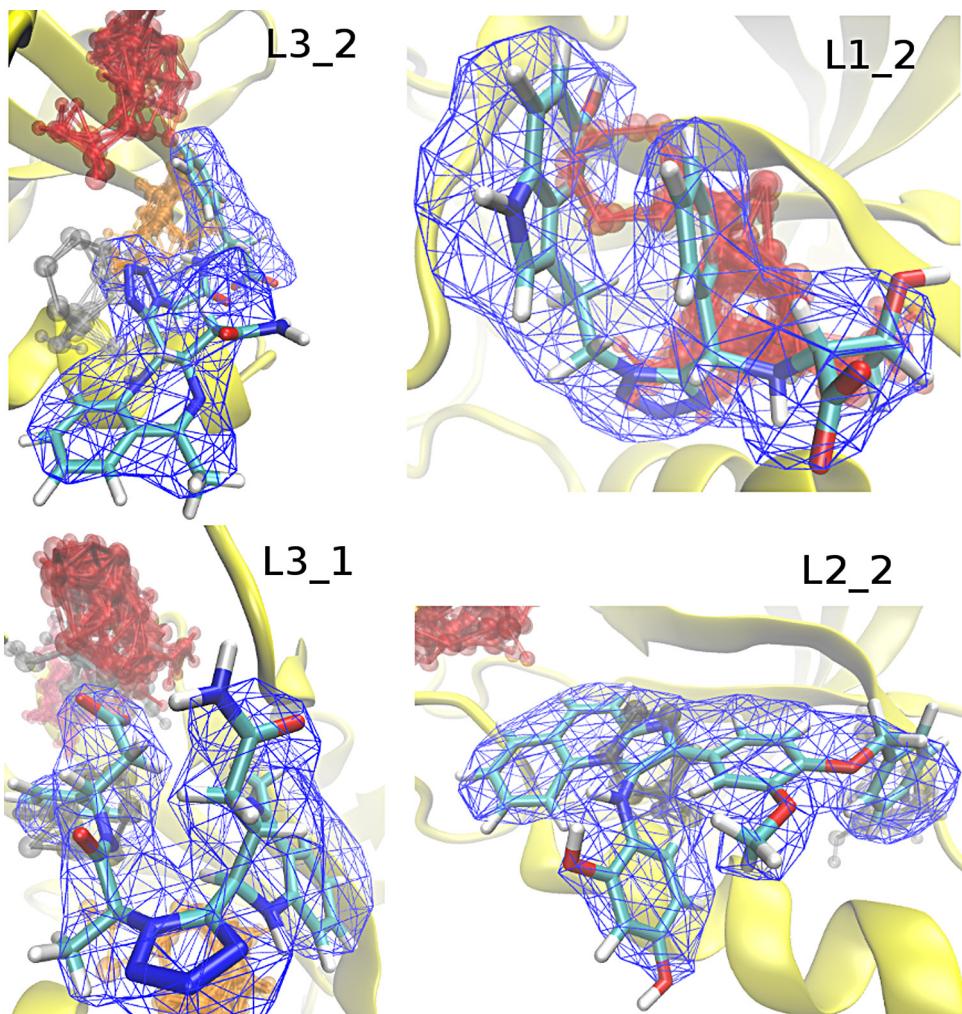


Fig. 12. Poor overlap of ligand positions in complexes L1_2, L2_2, L3_1 and L3_2 during MD simulations with FTMap hot spot clusters (red). Occupancy maps are shown for all ligands (blue mesh, isovalue 0.6). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

L3, as well as for the complex L2_2 (Fig. 12). In the fourth poorly ranked complex, L1_2, the ligand grazes a hot spot FTMap probe cluster in its vicinity, but does not fill its location.

4. Conclusions

In order to design small molecule inhibitors of CDK9/cyclin T1 interaction, a computational study of CDK9 binding surface dynamics and druggability was conducted. Druggability estimates and pocket opening analyses indicated binding regions of cyclin T1 residues Phe 146 and Lys 6 as promising starting points for the design of small molecules with the potential to inhibit the CDK9/cyclin T1 interaction. Structures of potential small molecule inhibitors were selected using a pharmacophore based on these two cyclin T1 residues and their energies of binding to CDK9 surface evaluated with MM-GBSA. Among structures tested, ligand L5 shows the most favorable MM-GBSA and MM-PBSA binding energies and the best overlap with FTMap predicted regions of ligand binding. Based on good agreement of computational predictions, this derivative of 2-amino-8-hydroxyquinoline is proposed as a small molecule that potentially inhibits the CDK9/cyclin T1 interaction. Further experimental studies are needed to verify the biological activity and selectivity of L5.

Acknowledgements

This 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.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.jmgm.2014.04.002>.

References

- [1] S.M. Shore, S.A. Byers, W. Maury, D.H. Price, Identification of a novel isoform of CDK9, *Gene* 307 (2003) 175–182.
- [2] J. Peng, Y. Zhu, J.T. Milton, D.H. Price, Identification of multiple cyclin subunits of human P-TEFb, *Genes Dev.* 12 (1998) 755–762.

- [3] G. Romano, A. Giordano, Role of the cyclin-dependent kinase 9-related pathway in mammalian gene expression and human diseases, *Cell Cycle Georget. Tex.* 7 (2008) 3664–3668.
- [4] V. Krystof, S. Baumli, R. Fürst, Perspective of cyclin-dependent kinase 9 (CDK9) as a drug target, *Curr. Pharm. Des.* 18 (2012) 2883–2890.
- [5] S. Wang, P.M. Fischer, Cyclin-dependent kinase 9: a key transcriptional regulator and potential drug target in oncology, virology and cardiology, *Trends Pharmacol. Sci.* 29 (2008) 302–313.
- [6] V. Krystof, I. Chamarád, R. Jorda, J. Kohoutek, Pharmacological targeting of CDK9 in cardiac hypertrophy, *Med. Res. Rev.* 30 (2010) 646–666.
- [7] K. Wang, P. Hampson, J. Hazeldine, V. Krystof, M. Strnad, P.M. Pechan, J. Cyclin-dependent kinase 9 activity regulates neutrophil spontaneous apoptosis, *PLoS ONE* 7 (2012) e30128.
- [8] R. Chen, M.J. Keating, V. Gandhi, W. Plunkett, Transcription inhibition by flavopiridol: mechanism of chronic lymphocytic leukemia cell death, *Blood* 106 (2005) 2513–2519.
- [9] S.M. Manohar, M.J. Rathos, V. Sonawane, S.V. Rao, K.S. Joshi, Cyclin-dependent kinase inhibitor, P276-00 induces apoptosis in multiple myeloma cells by inhibition of Cdk9-T1 and RNA polymerase II-dependent transcription, *Leuk. Res.* 35 (2011) 821–830.
- [10] S.F. Scrafe, P. Kierstan, J. Borgognoni, L.-Z. Wang, S. Denny, J. Wayne, C. Bentley, A.D. Cansfield, P.S. Jackson, A.M. Lockie, N.J. Curtin, D.R. Newell, D.S. Williamson, J.D. Moore, Transient treatment with CDK inhibitors eliminates proliferative potential even when their abilities to evoke apoptosis and DNA damage are blocked, *Cell Cycle Georget. Tex.* 7 (2008) 3898–3907.
- [11] G. Polier, J. Ding, B.V. Konkimalla, D. Eick, N. Ribeiro, R. Kohler, M. Giaisi, T. Efferth, L. Desaubry, P.H. Krammer, M. Li-Weber, Wogonin and related natural flavones are inhibitors of CDK9 that induce apoptosis in cancer cells by transcriptional suppression of Mcl-1, *Cell Death Dis.* 2 (2011) e182.
- [12] V. Krystof, L. Rárová, J. Liebl, S. Zahler, R. Jorda, J. Voller, P. Cankar, The selective P-TEFb inhibitor CAN508 targets angiogenesis, *Eur. J. Med. Chem.* 46 (2011) 4289–4294.
- [13] J. Wimmer, K. Fujinaga, R. Taube, T.P. Cujeć, Y. Zhu, J. Peng, D.H. Price, B.M. Peterlin, Interactions between Tat and TAR and human immunodeficiency virus replication are facilitated by human cyclin T1 but not cyclins T2a or T2b, *Virology* 255 (1999) 182–189.
- [14] D. Salerno, M.G. Hasham, R. Marshall, J. Garriga, A.Y. Tsygankov, X. Grana, Direct inhibition of CDK9 blocks HIV-1 replication without preventing T cell activation in primary human peripheral blood lymphocytes, *Gene* 405 (2007) 65–78.
- [15] Y.-L. Chiu, H. Cao, J.-M. Jacque, M. Stevenson, T.M. Rana, Inhibition of human immunodeficiency virus type 1 replication by RNA interference directed against human transcription elongation factor P-TEFb (CDK9/CyclinT1), *J. Virol.* 78 (2004) 2517–2529.
- [16] M. Sano, M. Abdellatif, H. Oh, M. Xie, L. Bagella, A. Giordano, L.H. Michael, F.J. DeMayo, M.D. Schneider, Activation and function of cyclin T-Cdk9 (positive transcription elongation factor-b) in cardiac muscle-cell hypertrophy, *Nat. Med.* 8 (2002) 1310–1317.
- [17] M. Sano, M.D. Schneider, Cyclin-dependent kinase-9: an RNAPII kinase at the nexus of cardiac growth and death cascades, *Circ. Res.* 95 (2004) 867–876.
- [18] M. Malumbres, P. Pevarello, M. Barbacid, J.R. Bischoff, CDK inhibitors in cancer therapy: what is next? *Trends Pharmacol. Sci.* 29 (2008) 16–21.
- [19] J. Węsierska-Gądek, M. Maurer, N. Zulehner, O. Komina, Whether to target single or multiple CDKs for therapy? That is the question, *J. Cell. Physiol.* 226 (2011) 341–349.
- [20] V. Krystof, P. Cankar, I. Frysová, J. Slouka, G. Kontopidis, P. Dzubák, M. Hajdúch, J. Srovnal, W.F. de Azevedo Jr., M. Orság, M. Paprskárová, J. Rolcik, A. Látr, P.M. Fischer, M. Strnad, 4-Arylazo-3,5-diamino-1H-pyrazole CDK inhibitors: SAR study, crystal structure in complex with CDK2, selectivity, and cellular effects, *J. Med. Chem.* 49 (2006) 6500–6509.
- [21] R. Van Duyne, I. Guendel, E. Jaworski, G. Sampey, Z. Klase, H. Chen, C. Zeng, D. Kovalsky, M.H. El Kouni, B. Lepene, A. Patanarut, S. Nekhai, D.H. Price, F. Kashanchi, Effect of mimetic CDK9 inhibitors on HIV-1-activated transcription, *J. Mol. Biol.* 425 (2013) 812–829.
- [22] B. O'Keffe, Y. Fong, D. Chen, S. Zhou, Q. Zhou, 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 (2000) 279–287.
- [23] C. Gonudeau, S. Gerbal-Chaloin, P. Bello, G. Aldrian-Herrada, M.C. Morris, G. Divita, Design of a novel class of peptide inhibitors of cyclin-dependent kinase/cyclin activation, *J. Biol. Chem.* 280 (2005) 13793–13800.
- [24] J. Randjelović, S. Erić, V. Savić, Computational study and peptide inhibitors design for the CDK9 – cyclin T1 complex, *J. Mol. Model.* 19 (2013) 1711–1725.
- [25] S. Baumli, G. Lolli, E.D. Lowe, S. Troiani, L. Rusconi, A.N. Bullock, J.E. Debreczeni, S. Knapp, L.N. Johnson, The structure of P-TEFb (CDK9/cyclin T1), its complex with flavopiridol, regulation by phosphorylation, *EMBO J.* 27 (2008) 1907–1918.
- [26] D.R. Hall, D. Kozakov, S. Vajda, Analysis of protein binding sites by computational solvent mapping, *Methods Mol. Biol. Clifton NJ* 819 (2012) 13–27.
- [27] I. Sillitoe, A.L. Cuff, B.H. Dessimailly, N.L. Dawson, N. Furnham, D. Lee, J.G. Lees, T.E. Lewis, R.A. Studer, R. Rentzsch, C. Yeats, J.M. Thornton, C.A. Orengo, New functional families (FunFams) in CATH to improve the mapping of conserved functional sites to 3D structures, *Nucleic Acids Res.* 41 (2013) D490–D498.
- [28] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, *J. Mol. Graph.* 14 (33–38) (1996) 27–28.
- [29] R. Brenke, D. Kozakov, G.-Y. Chuang, D. Beglov, D. Hall, M.R. Landon, C. Mattos, S. Vajda, Fragment-based identification of druggable “hot spots” of proteins using Fourier domain correlation techniques, *Bioinformatics* 25 (2009) 621–627.
- [30] The PyMOL Molecular Graphics System, Version 1.5.0.1, Schrödinger, LLC.
- [31] D.A. Case, T.A. Darden, T.E. Cheatham III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B.P. Roberts, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, P.A. Kollman, AMBER 11, University of California, San Francisco, 2010.
- [32] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, Comparison of multiple Amber force fields and development of improved protein backbone parameters, *Proteins* 65 (2006) 712–725.
- [33] I.S. Joung, T.E. Cheatham, Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations, *J. Phys. Chem. B* 112 (2008) 9020–9041.
- [34] A.D. Mackerell Jr., D. Bashford, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. State, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of proteins, *J. Phys. Chem. B* 102 (1998) 3586–3616.
- [35] A.D. Mackerell Jr., M. Feig, C.L. Brooks, 3rd 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 (2004) 1400–1415.
- [36] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kalé, K. Schulten, Scalable molecular dynamics with NAMD, *J. Comput. Chem.* 26 (2005) 1781–1802.
- [37] P.R.L. Markwick, J.A. McCammon, Studying functional dynamics in biomolecules using accelerated molecular dynamics, *Phys. Chem. Chem. Phys. Pccp* 13 (2011) 20053–20065.
- [38] Y. Wang, C.B. Harrison, K. Schulten, J.A. McCammon, Implementation of accelerated molecular dynamics in NAMD, *Comput. Sci. Discov.* 4 (2011) 015002.
- [39] E.V. Schneider, J. Böttcher, M. Blaeser, L. Neumann, R. Huber, K. Maskos, The structure of CDK8/CycC implicates specificity in the CDK/Cyclin family and reveals interaction with a deep pocket binder, *J. Mol. Biol.* 412 (2011) 251–266.
- [40] A. Fiser, R.K. Do, A. Sali, Modeling of loops in protein structures, *Protein Sci. Publ. Protein Soc.* 9 (2000) 1753–1773.
- [41] A. Fiser, A. Sali, ModLoop: automated modeling of loops in protein structures, *Bioinform. Oxf. Engl.* 19 (2003) 2500–2501.
- [42] M.H.M. Olsson, C.R. Søndergaard, M. Rostkowski, J.H. Jensen, PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions, *J. Chem. Theory Comput.* 7 (2011) 525–537.
- [43] M. Seeber, M. Cecchini, F. Rao, G. Settanni, A. Caflisch, Wordom: a program for efficient analysis of molecular dynamics simulations, *Bioinformatics* 23 (2007) 2625–2627.
- [44] S. Eyrisch, V. Helms, What induces pocket openings on protein surface patches involved in protein–protein interactions? *J. Comput. Aided Mol. Des.* 23 (2009) 73–86.
- [45] D.K. Johnson, J. Karanicolas, Druggable protein interaction sites are more predisposed to surface pocket formation than the rest of the protein surface, *PLoS Comput. Biol.* 9 (2013) e1002951.
- [46] A. Leaver-Fay, M. Tyka, O.F. Lange, J. Thompson, R. Jack, K. Kaufman, P.D. Renfrew, C.A. Smith, W. Sheffler, I.W. Davis, S. Cooper, A. Treuille, D.J. Mandell, F. Richter, Y.-E.A. Ban, S.J. Fleishman, J.E. Corn, D.E. Kim, S. Lyskov, M. Berroondo, S. Mentzer, Z. Popović, J.J. Havranek, J. Karanicolas, R. Das, J. Meiler, T. Kortemme, J.J. Gray, B. Kuhlman, D. Baker, P. Bradley, ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules, *Methods Enzymol.* 487 (2011) 545–574.
- [47] D.R. Koes, C.J. Camacho, Small-molecule inhibitor starting points learned from protein–protein interaction inhibitor structure, *Bioinform. Oxf. Engl.* 28 (2012) 784–791.
- [48] D. Koes, K. Khouri, Y. Huang, W. Wang, M. Bista, G.M. Popowicz, S. Wolf, T.A. Holak, A. Dömling, C.J. Camacho, Enabling large-scale design. Synthesis and Validation of Small Molecule Protein–Protein Antagonists, *PLoS ONE* 7 (2012) e32839.
- [49] D.R. Koes, C.J. Camacho, ZINCPharmer: pharmacophore search of the ZINC database, *Nucleic Acids Res.* 40 (2012) W409–W414.
- [50] C. Reynès, H. Host, A.-C. Camproux, G. Laconde, F. Leroux, A. Mazars, B. Deprez, R. Fahraeus, B.O. Villoutreix, O. Sperandio, Designing focused chemical libraries enriched in protein–protein interaction inhibitors using machine-learning methods, *PLoS Comput. Biol.* 6 (2010) e1000695.
- [51] I.V. Tetko, J. Gasteiger, R. Todeschini, A. Mauri, D. Livingstone, P. Ertl, V.A. Palyulin, E.V. Radchenko, N.S. Zefirov, A.S. Makarenko, V.Y. Tanchuk, V.V. Prokopenko, Virtual computational chemistry laboratory—design and description, *J. Comput. Aided Mol. Des.* 19 (2005) 453–463.
- [52] G.M. Morris, R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, A.J. Olson, AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility, *J. Comput. Chem.* 30 (2009) 2785–2791.
- [53] O. Trott, A.J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, *J. Comput. Chem.* 31 (2010) 455–461.
- [54] J. Wang, R.M. Wolf, J.W. Caldwell, P.A. Kollman, D.A. Case, Development and testing of a general amber force field, *J. Comput. Chem.* 25 (2004) 1157–1174.

- [55] J. Wang, W. Wang, P.A. Kollman, D.A. Case, Automatic atom type and bond type perception in molecular mechanical calculations, *J. Mol. Graph. Model.* 25 (2006) 247–260.
- [56] A.W.S. Silva, da, W.F. Vranken, ACPYPE – AnteChamber PYthon Parser interfacE, *BMC Res. Notes* 5 (2012) 367.
- [57] E. Vanquelef, S. Simon, G. Marquant, E. Garcia, G. Klimerak, J.C. Delepine, P. Cieplak, F.-Y. Dupradeau, R.E.D. Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments, *Nucleic Acids Res.* 39 (2011) W511–W517.
- [58] C.I. Bayly, P. Cieplak, W. Cornell, P.A. Kollman, A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model, *J. Phys. Chem.* 97 (1993) 10269–10280.
- [59] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G.A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H.P. Hratchian, A.F. Izmaylov, J. Bloino, G. Zheng, J.L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J.A. Montgomery Jr., J.E. Peralta, F. Ogliaro, M. Bearpark, J.J. Heyd, E. Brothers, K.N. Kudin, V.N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J.C. Burant, S.S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J.M. Millam, M. Klene, J.E. Knox, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, R.L. Martin, K. Morokuma, V.G. Zakrzewski, G.A. Voth, P. Salvador, J.J. Dannenberg, S. Dapprich, A.D. Daniels, Ö. Farkas, J.B. Foresman, J.V. Ortiz, J. Cioslowski, D.J. Fox, Gaussian 09, Gaussian, Inc., Wallingford, CT, 2009.
- [60] B.R. Miller, T.D. McGee, J.M. Swails, N. Homeyer, H. Gohlke, A.E. Roitberg, MMPBSA.py: an efficient program for end-state free energy calculations, *J. Chem. Theory Comput.* 8 (2012) 3314–3321.
- [61] P.A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D.A. Case, T.E. Cheatham III, Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models, *Acc. Chem. Res.* 33 (2000) 889–897.
- [62] R.A. Laskowski, M.B. Swindells, LigPlot+: multiple ligand-protein interaction diagrams for drug discovery, *J. Chem. Inf. Model.* 51 (2011) 2778–2786.
- [63] G.A. Ross, G.M. Morris, P.C. Biggin, Rapid and accurate prediction and scoring of water molecules in protein binding sites, *PLoS ONE* 7 (2012) e32036.
- [64] D. Kozakov, D.R. Hall, G.-Y. Chuang, R. Cencic, R. Brenke, L.E. Grove, D. Beglov, J. Pelletier, A. Whitty, S. Vajda, Structural conservation of druggable hot spots in protein–protein interfaces, *Proc. Natl. Acad. Sci.* (2011).
- [65] D. Hamelberg, J. Mongan, J.A. McCammon, Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules, *J. Chem. Phys.* 120 (2004) 11919–11929.
- [66] D. Hamelberg, C.A.F. de Oliveira, J.A. McCammon, Sampling of slow diffusive conformational transitions with accelerated molecular dynamics, *J. Chem. Phys.* 127 (2007) 155102.
- [67] P.R.L. Markwick, C.F. Cervantes, B.L. Abel, E.A. Komives, M. Blackledge, J.A. McCammon, Enhanced conformational space sampling improves the prediction of chemical shifts in proteins, *J. Am. Chem. Soc.* 132 (2010) 1220–1221.
- [68] R.E. Amaro, R. Baron, J.A. McCammon, An improved relaxed complex scheme for receptor flexibility in computer-aided drug design, *J. Comput. Aided Mol. Des.* 22 (2008) 693–705.
- [69] G. Rastelli, A. Del Rio, G. Degliesposti, M. Sgobba, Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA, *J. Comput. Chem.* 31 (2010) 797–810.
- [70] T. Hou, J. Wang, Y. Li, W. Wang, Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations, *J. Chem. Inf. Model.* 51 (2011) 69–82.