Metadynamics Simulations Distinguish Short- and Long-Residence-Time Inhibitors of Cyclin-Dependent Kinase 8.

Metadynamics Simulations Distinguish Short and Long Residence-Time Inhibitors of Cyclin-Dependent Kinase 8 (CDK8).

Donatella Callegari,[1] Alessio Lodola,[1] Daniele Pala,[1] Silvia Rivara,[1] Marco Mor,*[1] Andrea Rizzi,[2] and Anna Maria Capelli[2]*

[1]Dipartimento di Farmacia, Università degli Studi di Parma, Viale delle scienze 27/A, 43124 Parma, Italy; [2]Chemistry Research and Drug Design Department, Chiesi Farmaceutici S.p.A., Largo F. Belloli 11/A, 43122 Parma, Italy.

ABSTRACT

The duration of drug efficacy in vivo is a key aspect primarily addressed during the lead optimization phase of drug discovery. Hence, the availability of robust computational approaches able to predict the residence time of compounds at their target would accelerate candidate selection. Nowadays the theoretical prediction of this parameter is still very challenging. Starting from methods reported in the literature, we set up and validated a new metadynamics (META-D)-based protocol that was used to rank the experimental residence times of ten arylpyrazole cyclin-dependent kinase 8 (CDK8) inhibitors, for which target-bound X-ray structures are available. The application of reported methods, based on the detection of the escape from the first free-energy well, gave poor correlation with experimental values. Our protocol evaluates the energetics of the whole unbinding process accounting for multiple intermediates and transition states. Using seven collective variables (CVs) encoding both roto-translational and conformational motions of the ligand, a history-dependent biasing potential is deposited as a sum of constant-height Gaussian functions until the ligand reaches an unbound state. The time required to achieve this state is proportional to the integral of the deposited potential over the CVs hyperspace. Average values of this time, for replicated META-D simulations, provided an accurate classification of CDK8 inhibitors spanning short, medium, and long residence times.

INTRODUCTION

Binding affinity of a drug for its target is rarely the only requirement leading to a satisfactory *in vivo* efficacy. This is due to the fact that *in vivo* the binding process is not in its equilibrium state for the concurrence of complex physiological processes such as absorption, metabolism, distribution and elimination as well as cellular-related events (e.g., receptor internalization and degradation).[1] Whenever the equilibrium conditions are not fulfilled, it is the lifetime of a drug–target complex that represents the main determinant of the pharmacological response.[2] Optimization of this parameter is thus emerging as an effective strategy to control several pharmaceutically relevant drug properties, such as duration of action,[3] *in vivo* efficacy,[4] selectivity,[5] and safety.[6] As the lifetime of a drug-target complex depends on its dissociation rate constant (k_{off}),[7] the availability of reliable computational approaches that can predict k_{off} values, or accurately rank different compounds binding the same target would accelerate drug optimization. In principle, either the k_{off} value or the residence time ($1/k_{off}$) can be estimated by performing molecular dynamics (MD) simulations.[8] However, extensive exploration of the free-energy surface (FES) for physical dissociation of a drug from its binding site is necessary to guarantee prediction reliability[9]. It also provides information about the geometry and energetics of both minimum-energy and transition states that are instrumental to infer the free-energy barrier for drug-target dissociation (ΔG_{off}).[10] Despite the impressive boost of hardware and software performance,[11,12] the reconstruction of an unbiased binding FES is still challenging.[13] While classical MD suffers from the inability to effectively explore drug-target conformational space within a reasonable time-scale,[14] metadynamics (META-D)[15] simulations rely on an external biasing potential that might lead to a distortion of the shape and the position of the transition state (TS) region(s).

In recent years a wide range of computational approaches aimed at estimating the k_{off} have been reported in the literature.[16] They include: *i.* enhanced sampling algorithms[17] (i.e., conformational flooding,[18] steered-MD simulations[19,20] and adiabatic-bias molecular dynamics[21]) *ii.* Markov-state models;[22,23] *iii.* smoothed-potential MD methods.[24] Among them, the conformational flooding algorithm, originally introduced by Grubmüller[18] and Voter,[25] and re-adapted to the well-tempered metadynamics (*wt*-META-D) scheme[26] by Tiwary and Parrinello,[27] seems very promising. Assuming that the unbinding process can be reduced to a two-state system, where the ligand can adopt either a protein-bound or a protein-unbound configuration, the Tiwary's method was able to correctly estimate the residence time of benzamidine into the bovine trypsin active site.[28] According to this approach, the residence time can be estimated as the product of the transition time recorded in the *wt*-META-D simulation and the *acceleration factor* (α), which accounts for the energy deposited during the simulation until the transition occurs. To account for stochastic differences among single simulations, transition times from several independent *wt*-META-D simulations are collected and their distribution interpolated by a Poisson-like function to get a characteristic parameter.[28]

Prompted by this encouraging result, we used this approach to estimate the residence times of a congeneric series of biologically active compounds bound to their therapeutically relevant target.

For a set of arylpyrazole inhibitors of the cyclin-dependent kinase 8 (CDK8) both ligand-target bound X-ray structures and their experimental residence times have recently been reported in the literature.[29] In our preliminary set of simulations, a straightforward application of the Tiwary & Parrinello approach to this dataset hardly reproduced the experimental residence times. Furthermore, this protocol failed to accurately rank these inhibitors probably as a result of the complexity of their unbinding process.

In light of this, we conceived an empirical method for the estimation of residence times that: *i.* takes into account the whole unbinding process; *ii*. can be applied in a simple way to different drug-target complexes; *iii*. provides reliable estimation of relative residence times for a series of structurally-related ligands within the same binding site. The new protocol is based on META-D and uses the total amount of the deposited potential necessary to drive the ligand from the bound to a completely unbound state, both in the early and in late phases of unbinding, as a measure of ligand propensity to reside into the binding site. In our opinion, this represents a significant improvement with respect to the original methodology. In fact, the sole evaluation of the early-phases protein-ligand unbinding can lead to erroneous residence time prediction when the exit from the starting energy minimum does not represent the limiting step of the unbinding process or when multiple transition states with similar energies have to be crossed by the ligand to reach the bulk solvent. In our protocol, we used a classical implementation of META-D without the scaling factors of the well-tempered algorithm. In this way, the total biasing potential deposited over the hyperspace of collective variables (CVs) is proportional to the simulation time needed to move from the bound to the fully unbound state ($t_{META\text{-}D}$) and can be directly compared to the experimental residence time.

METHODS

Model Building

CDK8-inhibitor complexes were downloaded from the Protein Data Bank (PDB IDs: 4F6S, 4F6U, 4F6W, 4F7J, 4F7L, 4F7O, 4F7N).[29] The complexes were prepared with the Protein Preparation Wizard in Maestro version 10.4.[30] In details, part of the activation loop (sequence 177-194, which follows the key DMG motif always present in its out conformation) was missing

in all X-ray structures and was built in Prime 4.2,[31] using the homologue CDK6 (PDB ID 1BI8,[32] similarity score: 42%, crystallized with its DFG motif in out conformation) as a template structure (see Figure S1 for the alignment); missing side chains and hydrogen atoms were added and H-bond networks were optimized through an exhaustive sampling of hydroxyl and thiol groups. Tautomer states of His establishing the highest number of H-bonds were taken. Compounds **1**-**5** and **7** were modeled in their neutral state, whereas compound **6**, carrying a basic nitrogen in a piperidine ring, was modeled in its protonated form. Finally, in all complexes hydrogen atoms, protein side chains, and the ligand were energy-minimized using the OPLS2005 force field[33] implemented in Macromodel 11.[34] The refined CDK8-inhibitor complexes were *i.* solvated by TIP3P water molecules in a simulation box distant 12 Å from the protein in every direction, *ii.* neutralized by adding 5 Cl^- ions, and *iii.* equilibrated for 20 ns of MD simulation (5 ns in NPT and 15 ns in NVT ensemble) at 300 K using the Langevin thermostat.[35] All bond lengths to hydrogen atoms were constrained using M-SHAKE. Short-range electrostatic interactions were cut off at 9 Å whereas long-range electrostatic interactions were computed using the Particle Mesh Ewald method.[36] A RESPA integrator was used with a time step of 2 fs, and long-range electrostatics were computed every 6 fs. MD simulations were performed applying the OPLS2005 force field in Desmond 4.4 software.[37]

wt-META-D simulations

Equilibrated CDK8-inhibitor complexes were submitted to well-tempered metadynamics simulations using the Desmond 4.4 software (GPU implementation) on NVDIA780 graphic cards. To this end, a set of collective variables (from CV1 to CV7, see the main text for the definition), which describe the translation, rotation and conformation movements of the inhibitor in the CDK8 binding site, were used to simulate the exit of the inhibitors from their

crystallographic free-energy minima. Gaussians were deposited every 5 ps with a starting height of 0.54 kcal/mol, which gradually decreased on the basis of adaptive bias with a ΔT of 2700 K. The width of the Gaussians was defined on the basis of a series of explorative *wt*-META-D of the CDK8-compound **1** complex, which was tracked in the early phases of the unbinding process (*vide infra*). As a rule of thumb, we set the width of the Gaussians to one fifth of the variation of the CV required to drive the ligand out from the crystallographic free-energy minimum. In details, we used 0.6 Å for the distance CV1, 6° degrees for the angle CV2, 5° degrees for the torsion angle CV3, 15° degrees for the angle CV4, 15° degrees for the torsion angle CV5, 15° degrees for the angle CV6, 0.3 Å for the ligand RMSD (CV7).

Residence time estimation by wt-META-D simulations

According to the conformational flooding method,[18,25,27] the calculated residence time (RT_{calc}) within the free-energy basin A of a molecular system can be estimated using equation 1:

$$RT_{calc} = \alpha(t) \times t \qquad \text{(eq 1)}$$

where t is the time of metadynamics simulation necessary to observe a transition from basin A to basin B, whereas $\alpha(t)$ is the integral over time of the potential deposited in the free-energy basin A, necessary to trigger spontaneous transition toward a free-energy basin B. $\alpha(t)$ can be expressed as:

$$\alpha(t) = \langle e^{\beta V(s(R),t)} \rangle_A \qquad \text{(eq 2)}$$

where the angle brackets denote an average over a metadynamics run confined to the energy basin A, $V(s,t)$ is the metadynamics bias, and β is the Boltzmann factor. The time necessary for the transition from the basin A to B corresponds to the simulation time at which an abrupt

change in the sign of the first derivative of α over the time of simulation is observed.[27] In a recent application of the method, transition time has been approximately defined as the time of simulation at which α(*t*) does not increase more than 2-fold for a certain amount of time.[21] During our META-D simulations, often a local maximum in α *vs* time function was followed by a sudden rise in the α value (Figure S2). To identify clear-cut maximum points, we selected the second way to determine the transition time *t*, assuming that the ligand reached a new energy basin (B) when α(*t*) did not increase more than 2-fold for at least 1000 ps.

To evaluate the Poisson character of a set of A → B conformational transitions, the empirical cumulative distribution of calculated residence times (RT_{calc}) was fitted to the cumulative probability function (P) calculated by equation 3:

$$P = 1 - \exp(-RT_{calc} / \tau) \quad \text{(eq 3)}$$

where the characteristic residence time τ corresponds to the characteristic parameter of a Poisson model[38] and was calculated by least-squares fitting. The null hypothesis, which is verified when an empirical distribution function is not significantly different from a Poisson one, was evaluated by the Kolmogorov−Smirnov (KS) test. If the p-value associated with the KS test was higher than the significance threshold of 0.05, the null hypothesis was not rejected.

META-D simulations

Equilibrated CDK8-inhibitor complexes were submitted to META-D simulations with no well-tempering conditions, using the same set of seven collective variables employed for *wt*-META-D simulations described above. Gaussians were deposited every 0.5 ps to obtain the full unbinding of the ligand in reasonable times of simulation (25 ns). A Gaussian height of 0.3 kcal/mol was used for CV1-7 with Gaussian widths equal to those employed in *wt*-META-D simulations. To

measure the parameter t_{META-D}, which corresponds to the simulation time necessary for the ligand unbinding, a criterion for the detection of the unbound state was developed as follows. Compound **1** and a portion of CDK8 surrounding the ligand (residues 92-103) were extracted from the X-ray structure (PDB ID 4F6S) and prepared for simulations in a water box. The N- and C-terminus of the truncated protein were capped using the Protein Preparation Wizard implemented in Maestro and hydrogen atoms were added to saturate free valences. Compound **1** was translated 30 Å apart from the center of mass of the CDK8-deriving peptide. The system was solvated by a cubic box of TIP3P water molecules with a length of 100 Å and equilibrated by MD for 2.5 ns in NPT conditions, followed by 2.5 ns in NVT conditions at 300 K. All the MD parameters were the same employed for the equilibration of the whole CDK8-inhibitor complexes. The equilibrated complex was submitted to a META-D simulation in the space of the 7 CVs described above, with the same protocol of Gaussian deposition. The ligand was free to explore the solvent bulk without any interaction with the protein atoms. Under these conditions, corresponding to an unbound state, no accumulation of the potential deposited by META-D protocol was observed and the average value of the potential showed a constant value of 0.8 kcal/mol for all points of the CV space explored throughout the entire simulation (5 ns). Therefore, also for the META-D simulations with CDK8-inhibitor complexes, we considered that the ligand reaches the unbound state when for the preceding 500 ps of simulation the average value of the biasing potential dropped to the value of 0.8 kcal/mol and this reference energy level was maintained for other 500 ps. When this condition was met, the t_{META-D} was registered. For each CDK8-inhibitor complex 14 META-D replicas were performed changing the seed to assign the initial atom velocities, and a final t_{META-D} was reported as average ± standard

error of the mean. In average this META-D protocol required 5 h for each replica on a NVIDIA GTX780 GPU card installed on a computer with an Intel Xeon processor E5-2650.

RESULTS AND DISCUSSION

Residence time estimation by wt-META-D simulations for the CDK8 dataset

Figure 1 reports the structures of the set of CDK8 inhibitors, for which both experimental residence times and X-ray structures are avalilable.[29] These compounds, showing a common pyrazol-5-yl urea scaffold, can be divided into three different sets. The first one comprises short residence time inhibitors **1**-**3** (SRT, RT < 1.4 min) characterized by the lack of a pendant chain (**1**) or the presence of small hydrophilic substituents (hydroxyethyl for **2** and morpholinoethyl for **3**). The second one is formed by medium residence time (MRT) inhibitors **4** and **5** which carry either a morpholinopropyl or a hydroxypentyl chain and exhibit residence times of 14 and 57 min, respectively. The third group includes long residence time (LRT) inhibitors **6** and **7,** which have a 1-(2-(4-(3-tert-butyl-1-p-tolyl-1*H*-pyrazol-5-ylcarbamoyl)piperazin-1-yl)ethyl) or a *tert*-butoxycarbonylaminopropyl chain, showing extremely long residence times (1626 and 1944 min, respectively).

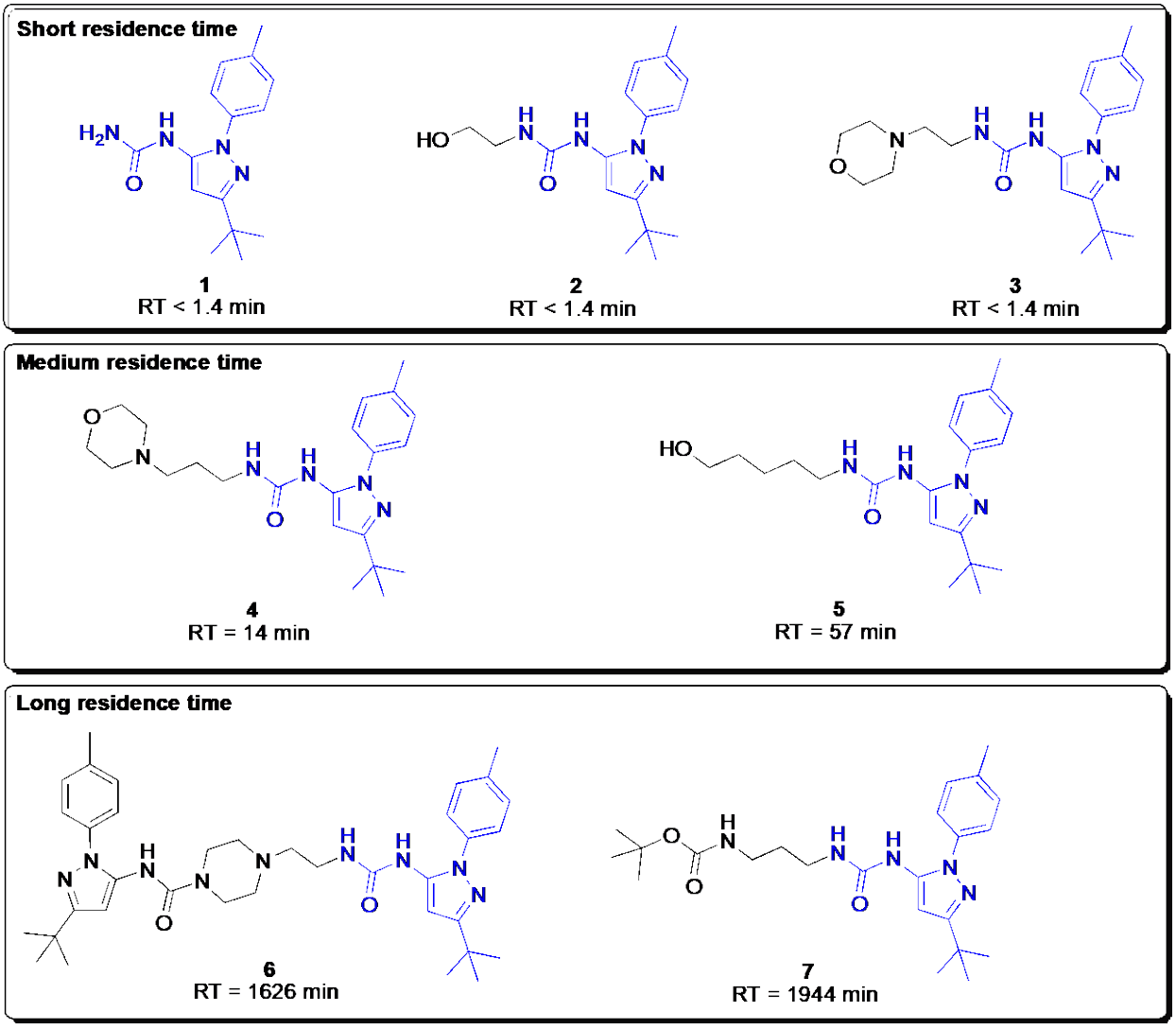

Figure 1. Chemical structures of the CDK8 inhibitors initially considered in this study and their experimental residence times. The common 1-(3-(*tert*-butyl)-1-(*p*-tolyl)-1*H*-pyrazol-5-yl)urea scaffold is depicted in blue.

All these compounds bind CDK8 in a similar way. Their common 1-(3-(tert-butyl)-1-(p-tolyl)-1*H*-pyrazol-5-yl)urea scaffold lies in close proximity to the DMG motif of the enzyme (corresponding to the DFG sequence of other kinases), with the urea donating two H-bonds to the carboxylate group of Glu66 and accepting a H-bond from the backbone NH of Asp173

(Figure 2). The variable chain is projected towards the hinge region and shows a compound-dependent binding mode. While compounds **1**-**3** do not establish direct contacts with the hinge region, compounds **4**-**5** are able to form H-bonds using their morpholinopropyl and hydroxypentyl chains. Instead, the chains of compounds **6** and **7** do not interact with the hinge region, but protrude into the solvent-exposed front pocket of CDK8.

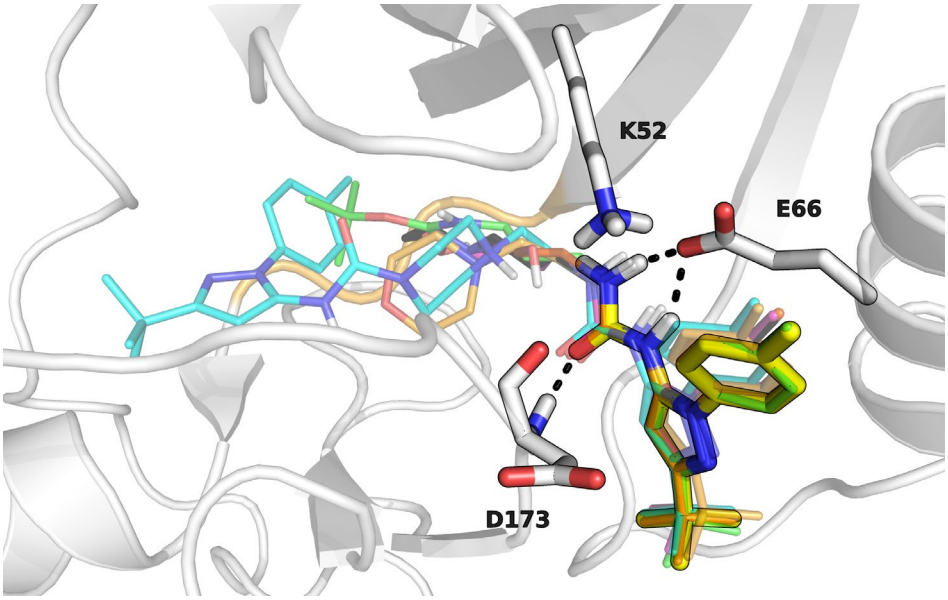

Figure 2. Active site of the crystal structure of human CDK8 in complex with compounds **1**-**7.** In evidence the interactions taken by the urea group of crystallized inhibitors with Glu66 and Asp173. Color code: compound **1** (yellow carbon atoms), **2** (red C), **3** (orange C), **4** (black C), **5** (magenta C), **6** (cyan C), and **7** (green C). Protein carbon atoms are represented in white. The protein backbone is represented by white cartoons, with the exception of the hinge region (orange cartoons).

Our first *wt*-META-D simulations were carried out with compound **1** bound to CDK8. To get a meaningful estimation of the residence time by the conformational flooding method of Tiwary and Parrinello[27] (see Methods), a series of conditions have to be satisfied: *i.* no biasing potential

should be deposited in the TS area, so that the well-to-well transition is unaffected by the potential itself; *ii.* the free-energy barrier separating the two basins involved in the transition should be significantly larger than RT (≈ 0.6 kcal/mol), *iii.* the CVs employed to drive the system out of the starting basin should clearly distinguish the two basins. When all these conditions are met, the calculated residence times for different simulations conducted on the same molecular process should adopt a Poisson distribution, in agreement with the law of rare events.[38] A time interval in the picosecond (ps) scale between two successive depositions of biasing potential may fulfill the first condition. The second condition is satisfied if no spontaneous ligand unbinding is observed after tens of ns in a classical MD simulation. In our case, compound **1** remained tightly bound to the protein active site for 30 ns in a plain MD simulation (Figure S3, Supporting Information), indicating that the barrier of unbinding should be at least as high as 7.0 kcal/mol at 300 K, according to the Eyring equation.[39] The third condition requires an iterative work based on chemical intuition and extensive testing to find out suitable CVs.

After several trials we concluded that seven CVs describing both ligand roto-translational movements and conformational changes could provide reliable results for this protein-ligand complex and may be used with the other ligands. In particular, the first three CVs describe the translational movements of the ligand center of mass (COM) with respect to the CDK8 active site. As shown in Figure 3A-C, they encode the following geometrical parameters: *a.* the distance between the COM of compound **1** and the COM of the protein hinge region (amino acids 98-103, panel A); *b.* the angle defined by the COM of the ligand, the COM of the protein hinge region and the COM of the β-strand-6 of the kinase active site (amino acids 92-98, panel B); *c.* the dihedral angle defined by these three COMs and a fourth COM taken on β-strand-5 of

the kinase active site (amino acids 80-87, panel C). A second set of three CVs (Figure 3D-E) describes the rotation of the whole ligand with respect to the CDK8 active site. They consist of: *d.* the angle formed by the major axis of inertia of the ligand (approximately defined picking two atoms of the rigid pyrazolylurea scaffold) and the COM of the protein hinge region (panel D); *e.* the dihedral angle defined by the major axis of inertia of the ligand, the COM of the protein hinge region and the COM of the β-strand-1 of the kinase active site (panel E); *f.* the angle formed by the minor axis of inertia of the ligand (approximately defined picking two atoms of the rigid phenylpyrazole scaffold) and the COM of the protein hinge region (panel D). Finally, the RMSD of the ligand with respect to the initial experimental orientation was included as the seventh CV to account for degrees of freedom orthogonal to those explored by the previous CVs.

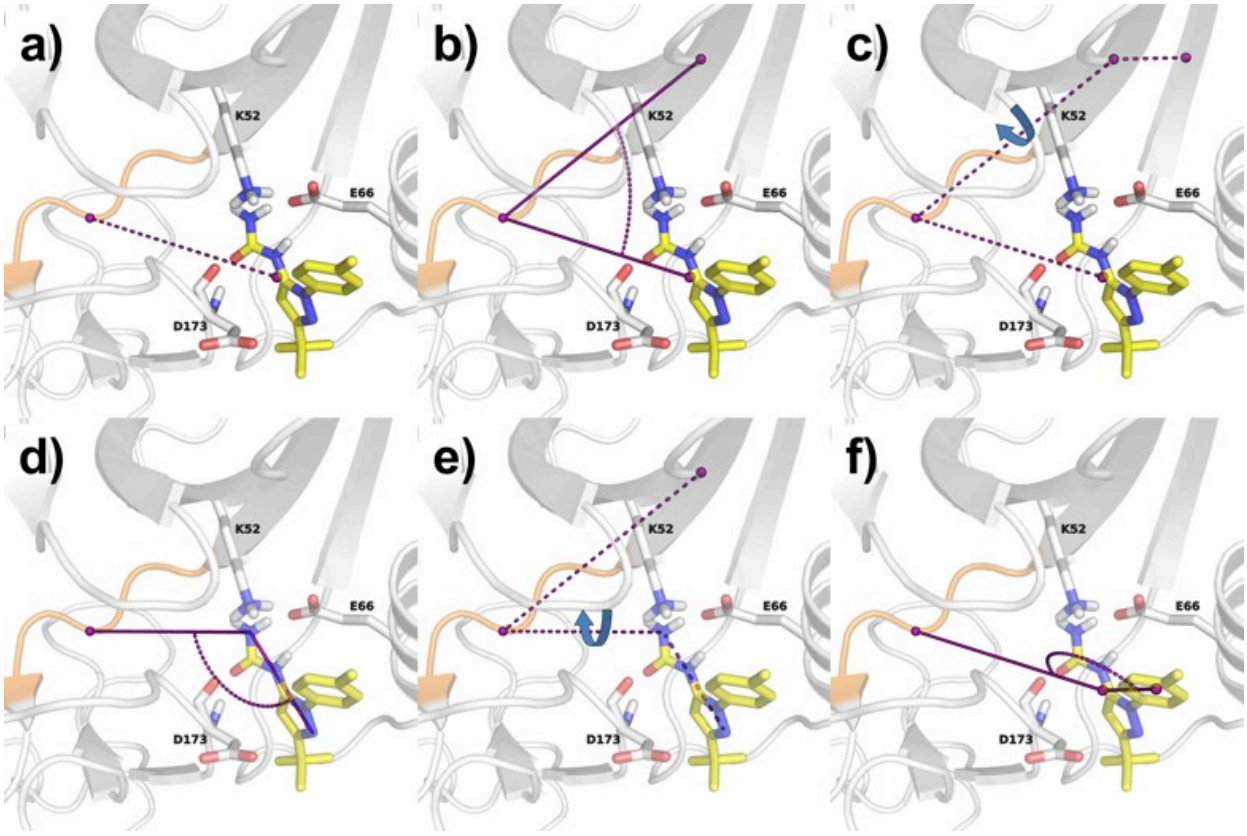

Figure 3. Graphical representation of the first six CVs employed for the metadynamics simulations involving the CDK8-compound **1** system. Key points employed to define the CVs

are depicted in purple while axes are described by plain purple lines. Panel A describes CV1, i.e., the distance between the COM of compound **1** and the COM of the protein hinge region. Panel B describes CV2, the angle defined by the COM of the ligand, the COM of the protein hinge region and the COM of the β-strand-6 of the kinase active site. Panel C describes CV3, the dihedral angle defined by these three COMs and a fourth COM taken on β-strand-5 of the kinase active site. Panel D reports CV4, the angle formed by the major axis of inertia of the ligand (purple line) and the COM of the protein hinge region. Panel E reports CV5, the dihedral angle defined by the major axis of inertia of the ligand, the COM of the protein hinge region and the COM of the β-strand-1 of the kinase active site. Panel F reports CV6, the angle formed by the minor axis of inertia of the ligand (purple line) and the COM of the protein hinge region.

Several simulations were performed for compound **1** bound to CDK8 using different combinations of CVs. According to the implementation of Bortolato *et al.*,[21] each simulation was terminated when $\alpha(t)$ did not increase more than 2-fold for at least 1000 ps and the corresponding residence times were calculated from the acceleration factor α using equation 1 (see Methods). In Figure 4 the cumulative distributions of the calculated residence times measured in twenty simulation replicas by progressively increasing the number of CVs, in the order *a* to *f*, are reported. As shown, the use of a small set of CVs (i.e., 1-3) gives rise to spread distributions, suggesting non-Poisson processes.[40] Conversely, when all CVs are used, the distribution is similar to a theoretical Poisson distribution (p-value = 0.77, Kolmogorov−Smirnov (KS) statistical test).

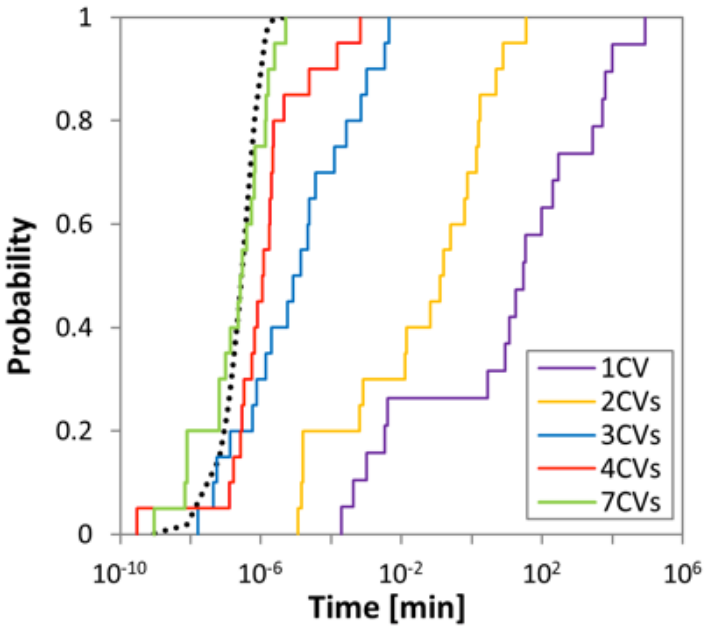

Figure 4. Empirical cumulative distribution of the residence time evaluated from twenty independent metadynamics simulations of the CDK8-compound **1** complex using an increasing number of CVs. The dotted black curve represents a theoretical Poisson distribution derived from equation 3.

To test the robustness of this approach we repeated a set of twenty simulations for the same complex either reducing the Gaussian height from 0.6 kcal/mol to 0.3 kcal/mol or increasing the time interval between two consecutive depositions of biasing potential from 5 ps to 10 ps. No significant differences were observed in the distribution of the residence time, as shown in Figure S4 (see Supporting Information).

Then, the *wt*-META-D protocol with seven CVs was applied to the other CDK8-inhibitor complexes and twenty independent *wt*-META-D simulations were carried out for each of them, and the characteristic values of the distributions (τ, see Methods, equation 3) was compared to

experimental residence times. As a result, we observed poor correlation (Figure 5). What is more, a classification as SRT, MRT and LRT inhibitors was totally wrong. Actually, the distribution curves of SRT inhibitors **1-3,** of MRT inhibitors **4-5** and of the LRT inhibitor **6** are superimposed. With respect to compounds **4-7**, the computed residence times ranged from 10^{-9} to 10^{-2} minutes, remarkably different from the experimental values.

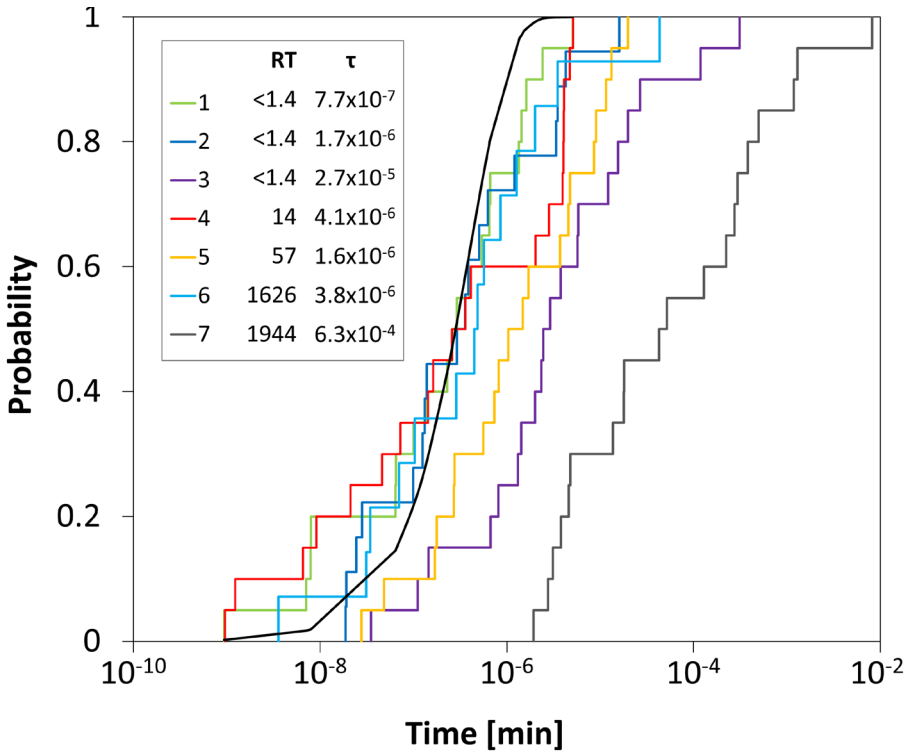

Figure 5. Empirical cumulative distribution of the residence times evaluated from twenty independent *wt*-META-D simulations with 7 CVs, for CDK8 in complex with compounds **1-7**. Color codes, experimental residence times (RT, min) and calculated ones (τ, s) are reported in the inset. The black curve represents a theoretical Poisson distribution fitted to the empirical cumulative distribution of the CDK8-compound **1** complex.

On the basis of these results we hypothesized that some conditions related to the applicability of this approach had been somehow violated. In fact, upon reconstruction of the FES for the

unbinding process of compound **1** using *wt*-META-D simulations in the space of CV1 and CV2, we observed several minimum free-energy wells. This observation supports the idea that the exit from the experimental free-energy minimum might not necessarily represent the rate-limiting step of CDK8 dissociation and that the simulated system needs to cross multiple TSs before reaching the unbound state (Figure 6).

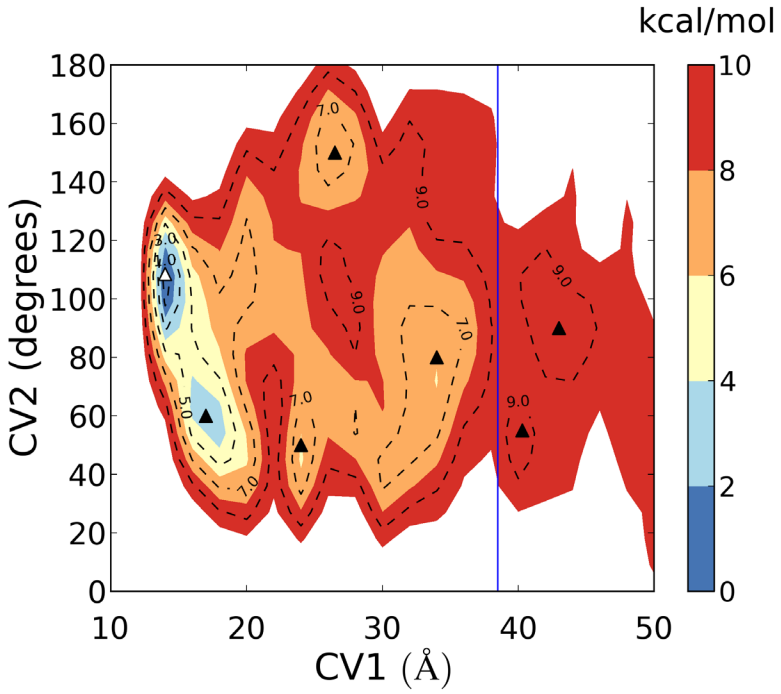

Figure 6. Free-energy surface in the CV1 and CV2 space for unbinding of compound **1** from CDK8 binding site. The white triangle identifies the values of CVs for the X-ray structure, the black triangles those for the other minima. The vertical blue line represents the boundary along CV1 separating bound and unbound states.

RT estimation by META-D simulation time

Given the complexity of the free energy landscape of CDK8-arylpyrazole inhibitor unbinding process, we decided to look for a new protocol able to reproduce the variation of experimental

residence times in an empirical way, yet taking into consideration the full process of unbinding. To be applied to the drug design process, such a method should be simple, use the smallest possible number of tunable settings, and provide a clear indication of the expected residence time in a qualitative way (compound rank order). In the framework of a semi-empirical approach, it is not required to give an absolute measure of residence times, but it is sufficient that compounds, which have high structural similarity and significantly different RT within the series, are accurately classified. According to the experiments performed, the set of seven CVs appears suitable to simulate the unbinding process of both SRT and LRT inhibitors. As for the metadynamics protocol and the RT predictor, we decided to apply a classical META-D approach, avoiding well-tempering conditions, and to measure the simulation time necessary to move from the crystallographic pose of the ligand to the completely unbound state ($t_{META\text{-}D}$). In this way, considering that the biasing potential is deposited in Gaussian functions with constant height, the simulation time spent in the bound state is proportional to the integral of deposited potential over the CV space. We assumed that those ligands, which have to cross higher or multiple energy barriers, need more energy from the metadynamics protocol to be expelled from their binding site. As a ligand could follow different unbinding paths during META-D simulations, average values of $t_{META\text{-}D}$ were collected from replicated simulations and averaged. Such an approach should be more suitable to treat complex systems due to its ability to manage multiple intermediate states and barriers along the unbinding path in spite of its empirical nature. Nevertheless, it requires an objective definition of the unbound state to measure $t_{META\text{-}D}$. While recent investigations mainly used geometrical criteria,[24,41] we looked for an energy-based definition that could be extracted from the time course of deposited META-D potentials. Such an approach does not require time-consuming analysis of META-D trajectories.

We reasoned that in a META-D simulation the biasing potential accumulated in the recursively explored regions of the CV space tends to grow proportionally to the simulation time. On the other hand, when a basin is exceptionally flat as in the case of a small molecule embedded in the bulk solvent, no accumulation of deposited energy is expected to take place and it is unlikely that the same region of the CV space is explored twice or more. We noticed that in a META-D simulation where compound **1** is freely moving in a solvent bulk (in a system also including a portion of CDK8 for the definition of the CVs - see Methods) the average value of the potential registered in the points of the CV space is approximately constant at 0.8 kcal/mol (Figure 7).

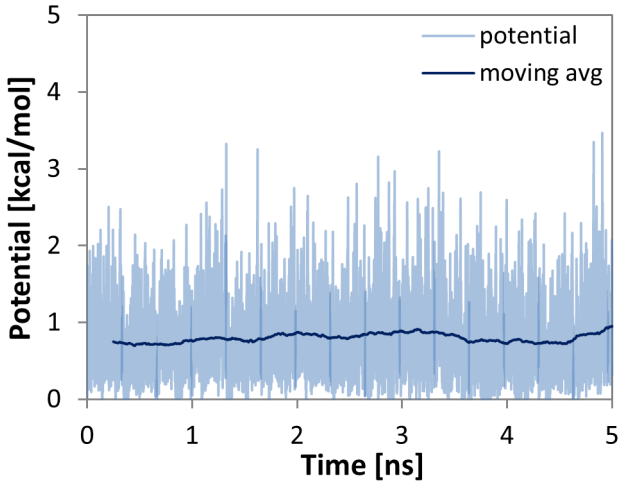

Figure 7. Biasing potential accumulated during the META-D simulation in the explored regions of the seven CV-space for compound **1** freely moving in a solvent bulk. The moving average (blue line) is calculated for the preceding 500 ps of the simulation.

We tested this new approach by simulating the full unbinding process of compound **1** from CDK8. Figure 8 reports the time course of the 500 ps-average potential deposited during the simulation (panel A), along with the time course of the CV1 (panel B), which is the distance between the ligand and the binding site COM, and offers a perspective of ligand-protein distance.

At early simulation times (0-1 ns interval) a significant growth of the average potential deposited was observed and the ligand is still in its starting basin (CV1 < 15). When compound **1** exits this basin (1-4 ns interval, 15 Å < CV1 < 30 Å) several fluctuations in the average potential were noticed. At the end of this interval, compound **1** lies outside the crystallographic binding pocket, even though it is close to the DMG motif and interacts with helix-C of CDK8. At longer times (higher than 4.45 ns, CV1 > 30 Å), compound **1** is no longer in contact with the CKD8 surface (i.e., the distance between each atom of the ligand and each atom of the protein is higher than 8 Å), being fully solvated in its unbound state (Figure 8 C). Consistently with the test simulation represented in Figure 7, the average potential decreases below 0.8 kcal/mol.

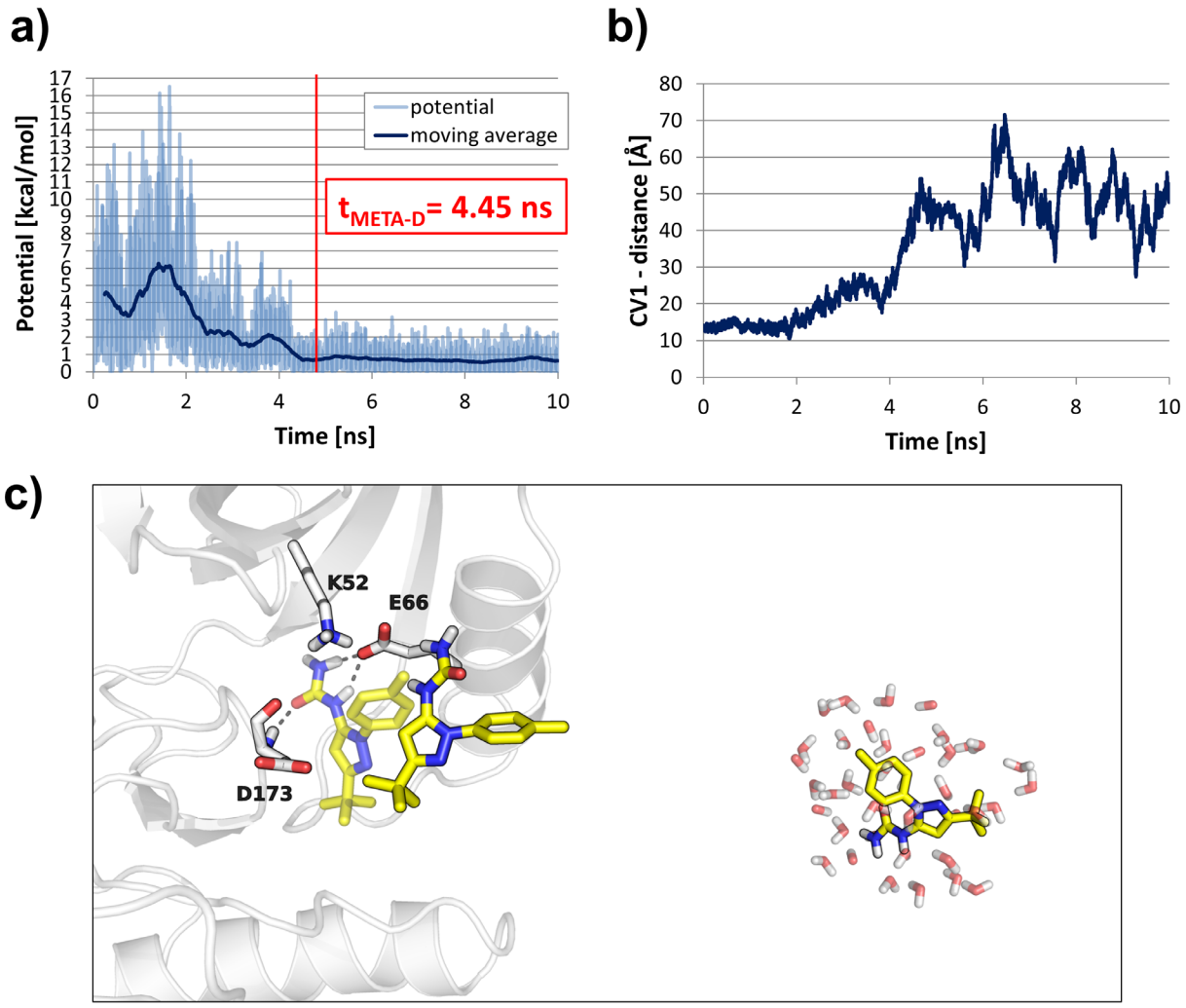

Figure 8. META-D simulation of compound **1** unbinding from CDK8. (a) time course of the average potential deposited during the simulation. (b) time course of CV1. (c) representative configurations assumed by compound **1**: in its starting pose (semi-transparent atoms), interacting with CDK8 surface and in the solvent bulk (on the right).

The unbinding simulation of compound **1** was repeated fourteen times. Visual inspection of the unbinding trajectories revealed that it had followed approximately the same route, moving straight towards the solvent, rolling over the surface of helix-C. The calculated residence times resulted normally distributed around the mean value ($t_{META-D} = 4.87 \pm 0.41$), which suggests that

under these conditions the cumulative probability of the transition is no longer Poissonian. While this new approach does not adhere to the physical and statistical requirements of the Tiwary & Parrinello method, the limited spread of the t_{META-D} values gathered for compound **1** indicates a good level of reproducibility of the META-D protocol.

Subsequently, we used this approach to simulate the unbinding of compound **7**, one LRT inhibitor of the dataset. In Figure 9 the time course of the average potential deposited (panel A), along with the time course of CV1 (panel B) and some key geometries of bound and unbound states for compound **7** (panel C) are reported. Also in this case, in the early phase of the META-D simulation (i.e., 0-5 ns interval, CV1 < 15 Å), an increase in the average potential deposited in the crystallographic free-energy minimum is observed. As soon as compound **7** leaves the starting basin (5-10 ns interval, 15 Å < CV1 < 35 Å) there is a decrease in the average potential, with quite large fluctuations. At these values of CV1, compound **7** is outside the CDK8 pocket proximal to the DMG motif (Figure 9 C), but it remains in contact with the protein surface. At longer simulation times (higher than 14.74 ns, CV1 > 30 Å), compound **7** is no more in contact with CKD8 (i.e., the distance between each atom of the ligand and each atom of the protein is higher than 8 Å), being fully solvated in its unbound state (Figure 9 C). Also in this case, the average potential decreases below the threshold of 0.8 kcal/mol in the last period of the simulation. The unbinding of compound **7** from CDK8 was replicated fourteen times. The resulting mean residence time (t_{META-D} = 17.98 ± 3.90) was significantly higher than that obtained for the SRT inhibitor **1** (t_{META-D} = 4.87 ± 0.41). Similarly to what was observed for compound **1**, compound **7** left the CDK8 active site moving straight away from the DMG motif towards the solvent rolling over the surface of helix-C in all replicated simulations.

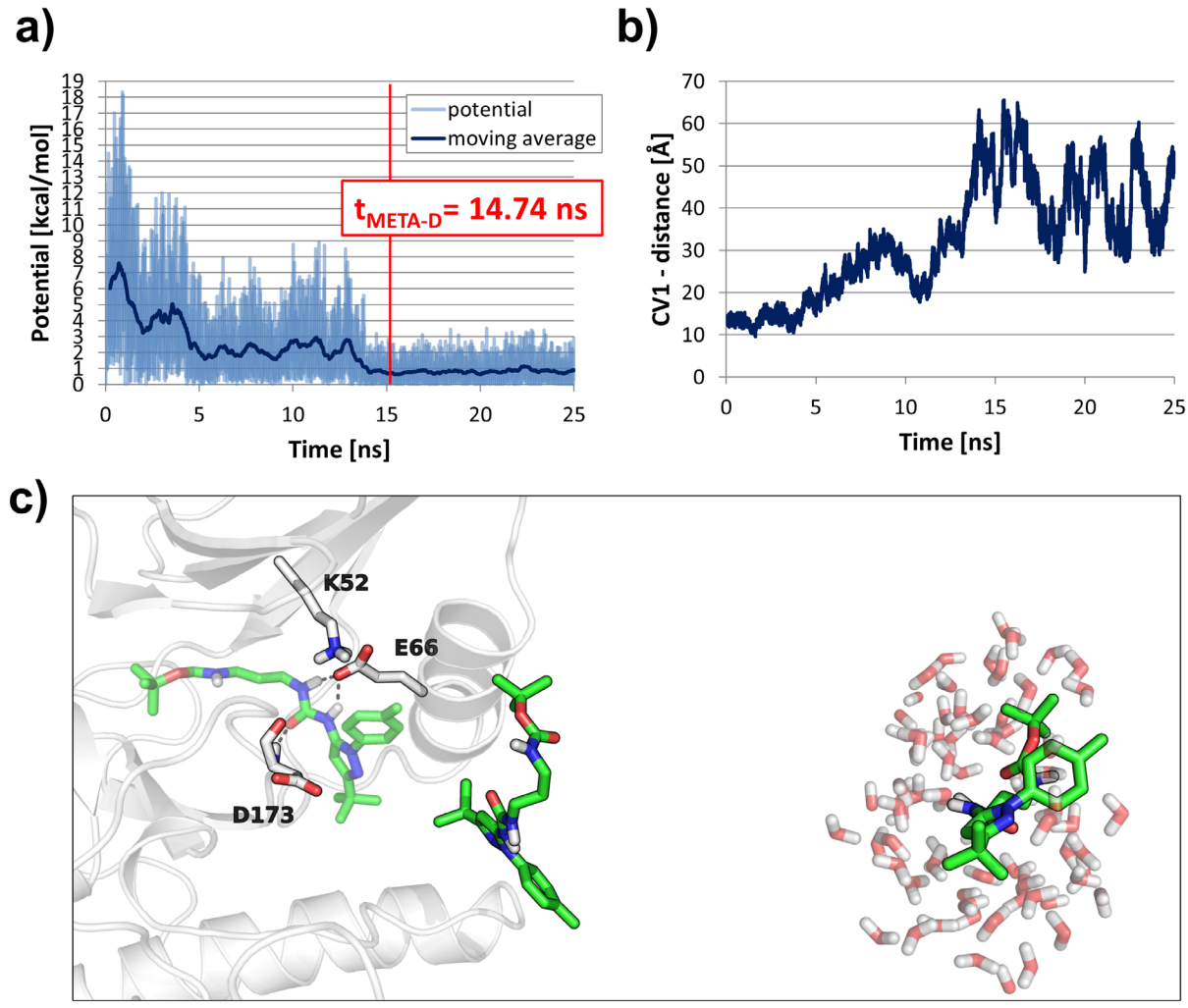

Figure 9. Representative unbinding simulation of compound **7** from CDK8. (a) time course of the average potential deposited during the simulation. (b) time course of CV1. (c) representative configurations assumed by compound **7** (green carbon atoms): in its starting minimum (semi-transparent atoms), bound to the CDK8 surface and in the solvent bulk.

Encouraged by these results, we extended our approach to the whole set of crystallized CDK8–inhibitor complexes and we performed fourteen independent META-D runs for each derivative. In Table 1 the average t_{META-D} values for compounds **1**-**7** and their standard errors of the mean are reported. The method was able to correctly discriminate LRT inhibitors from all the others.

While compounds **6** and **7** require more than 15 ns of META-D to reach their unbound state, compounds **1**-**5** took less than 12 ns of META-D to reach this configuration. Moreover, a separation can be envisaged between MRT (compounds **4**-**5**) and SRT inhibitors (compounds **1**-**2**), even if the SRT morpholino derivative **3** has a relatively high t_{META-D}.

Table 1. Experimental residence time (RT) and calculated t_{META-D} (averages±s.e.m. n=14) for CDK8 inhibitors **1**-**7**.

Cpd.	**RT [min]**	**t_{META-D} [ns]**
1	<1.4	4.87±0.41
2	<1.4	7.60±0.61
3	<1.4	9.62±0.76
4	14	10.85±1.28
5	57	11.20±0.94
6	1626	>25
7	1944	17.98±1.04

Considering that the pKa of alkylmorpholines is about 7.4,[42] we wondered whether the uncertain classification of compound **3**, modelled in its neutral state, could be related to its ionization state. Therefore, we carried out fourteen META-D simulations of compound **3** in its protonated form. For consistency, we did the same for compound **4**. The t_{META-D} gathered from these simulations is marginally affected by the protonation state suggesting that the morpholine amino group of

compound **3** (**t_{META-D}** = 10.39 ± 1.39 ns**)** and compound **4** (**t_{META-D}** = 11.48 ± 1.42 ns**)** does not significantly interact with CDK8 active site. Molecular simulations with fixed protonation states for titrable groups are poor approximations of complex proton-exchange equilibria. Molecular dynamics algorithms implementing simulations at constant pH[43,44] may better fit to treat cases with ligands and protein residues having weak basic or acidic groups.

Subsequently, we investigated whether the analysis of META-D unbinding trajectories could provide valuable insights as to how structural fragments can influence ligand unbinding. Differently from Schneider *et al.,*[29] who suggested on the analysis of the X-ray structures that the mapping of the front pocket can lead to long residence time (compounds **6** and **7)**, our META-D simulations indicate that either the ureido (**6**) or the carbamoyl (**7**) fragment in the pendant chain establish H-bonds interactions with Lys 52, a key polar residue proximal to the DMG motif. This interaction delays ligand exit accounting for the prolonged occupancy of the binding site (Figure S5, Supporting Information). In light of that, it might be possible to rationally design LRT inhibitors that specifically target this residue.

We also checked conformational stability of the DMG motif during the unbinding of compounds **1**-**7**. Analysis of the trajectories revealed that this critical fragment of the activation loop maintains its starting (i.e., "DMG-out") conformation, regardless of the type of inhibitor bound within the CDK8 active site. While in our simulations full unbinding of **1**-**7** does not require a change in the activation state of the kinase to occur, this may simply be due to the short simulation times.

We finally investigated the ability of the META-D approach to correctly classify other three-arylpyrazole CDK8 inhibitors included in the dataset reported by Schneider *et al.*, for which the

experimental residence times are reported, but lack X-ray structures. The methyl derivative **8**, the hydroxypropyl derivative **9** and the hydroxybutyl derivative **10** are shown in Figure 10. Compounds **8** and **9** belong to the SRT class, having very short residence times (< 1.4 min), whereas compound **10** is classified as a MRT inhibitor, having a residence time of 7 min. Compounds **8**-**10** were docked into the CDK8 kinase binding site taken from the X-ray structure complex with compound **1** using Glide software[45] **(**see Supporting Information for details). The top-ranked docking solutions of the modeled compounds showed their arylpyrazole moiety in the same spatial area as that of compound **1**, with the urea group making H-bond interactions with Glu66 and Asp173, similarly to the other derivatives (Figure 2). The hydroxybutyl chain of **10** was longer enough to undertake a H-bond with the hinge region of CDK8 (Figure S6). Subsequently, these complexes were submitted to the META-D protocol described for other CDK8 inhibitors. The resulting $t_{META\text{-}D}$ values are reported in Figure 10 along with those previously obtained for compounds **1**-**7**. As shown, compound **8** and **9** were correctly ranked as SRT inhibitors ($t_{META\text{-}D}$ of 6.67 ± 0.54 ns and of 7.43 ± 0.57 ns, respectively), whereas compound **10** was fairly well assigned to the MRT class ($t_{META\text{-}D}$ of 10.75 ± 0.99 ns). These results indicate that the method has a reasonable predictive power for this class of inhibitors and could be used for prospective drug discovery of newly designed analogues.

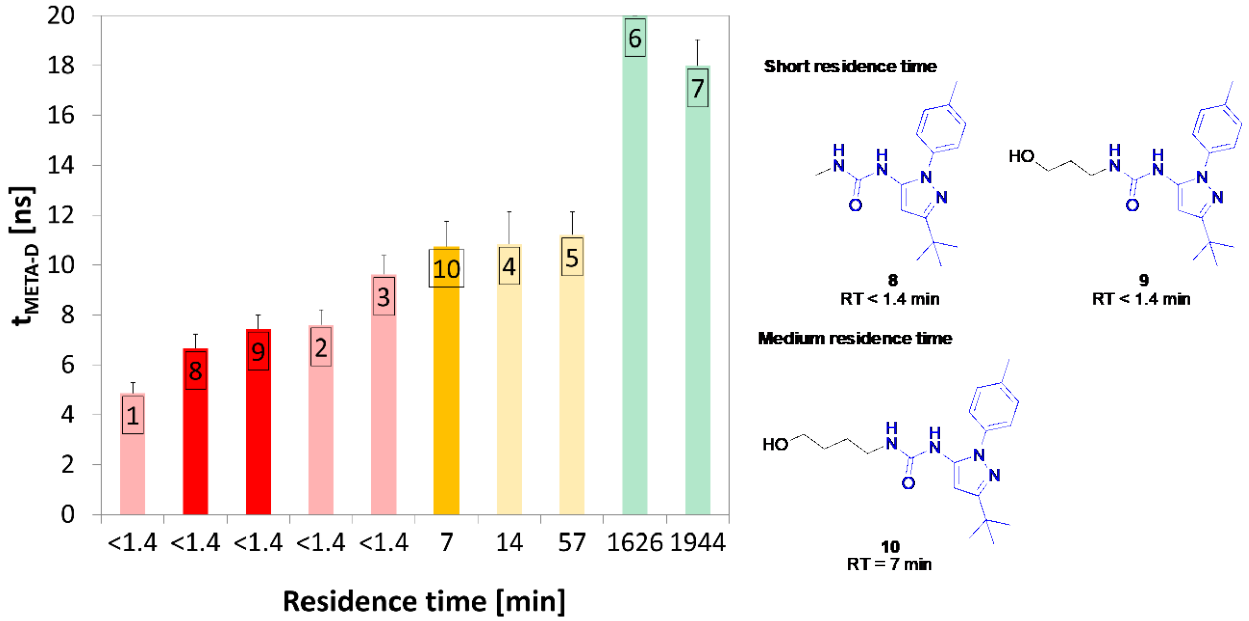

Figure 10. Calculated t_{META-D} vs experimental residence time for CDK8 inhibitors **1**-**10**. The chemical structures of compounds **8** -**10** are also reported.

CONCLUSIONS

The residence time of a protein-ligand complex is a crucial parameter affecting biological effects in vivo. Despite the availability of powerful computer resources and a wide range of computational techniques, its prediction is still a challenging task. In this scenario, we used the conformational flooding approach developed by Tiwary and Parrinello to rank a set of arylpyrazole CDK8 inhibitors for which both the X-ray structures of the enzyme-inhibitor complexes and the experimental RT of the inhibitors were available. This approach failed to correctly rank the residence time of this data set, probably due to the high degree of complexity of the system under study. We thus developed a novel META-D protocol, which is rooted into the energetics of the whole unbinding process. Our approach relies on two properties directly derived from the original implementation of the META-D algorithm i.e., the lack of

accumulation of the deposited potential to objectively define the simulation completion and the simulation time ($t_{META\text{-}D}$) to perform meaningful comparison among ligands binding the same target.

Although more investigations are necessary to reinforce the general applicability of our method, the proper ranking of residence times of arylpyrazole CDK8 inhibitors, spanning more than 3 orders of magnitude, suggests that this approach can be exploited in drug discovery projects aimed at optimizing the experimental residence time of a congeneric series of ligands.

ASSOCIATED CONTENT

The Supporting Information is available free of charge on the ACS Publications website. Details of the MD simulations of the CDK8-inhbitor complexes, CDK8-CDK6 sequence alignment, and details of docking compounds **8** - **10** into the CDK8 structure taken from the complex with compound **1.**

AUTHOR INFORMATION

Corresponding Author

* Phone: +39 0521 905062. Fax: + 39 0521 905006. E-mail: alessio.lodola@unipr.it

* Phone: +39 0521 905059. Fax: + 39 0521 905006. E-mail: marco.mor@unipr.it

Present Addresses

[†]Chemistry Research and Drug Design Department, Chiesi Farmaceutici S.p.A., Largo F. Belloli 11/A, 43122 Parma, Italy.

Author Contributions

All the authors contributed to the manuscript and have given approval to its final version.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

Dr. Pratyush Tiwary (Dept. of Chemistry, Columbia University, NY, USA) is kindly acknowledged for helpful discussions.

REFERENCES

(1) Copeland, R. A.; Pompliano, D. L.; Meek, T. D. Opinion - Drug-target Residence Time and Its Implications for Lead Optimization. *Nat. Rev. Drug Discovery* **2006**, *5*, 730−739.

(2) Tummino, P. J.; Copeland, R. A. Residence Time of Receptor- Ligand Complexes and Its Effect on Biological Function. *Biochemistry* **2008**, *47*, 5481−5492.

(3) Vanderheyden, P. M.; Fierens, F. L.; Vauquelin, G. Angiotensin II Type 1 Receptor Antagonists. Why Do Some of Them Produce Insurmountable Inhibition? *Biochem. Pharmacol.* **2000**, *60*, 1557–1563.

(4) Sykes, D. A.; Dowling, M. R.; Charlton, S. J. Exploring the Mechanism of Agonist Efficacy: A Relationship between Efficacy and Agonist Dissociation Rate at the Muscarinic M3 Receptor. *Mol. Pharmacol.* **2009**, *76*, 543–551.

(5) Guo D, Dijksteel GS, van Duijl T, Heezen M, Heitman LH, IJzerman AP. Equilibrium and Kinetic Selectivity Profiling on the Human Adenosine Receptors. *Biochem Pharmacol.* **2016**, *105*, 34-41.

(6) Guo, D.; Hillger, J. M.; IJzerman, A. P.; Heitman, L. H. Drug-Target Residence Time--a Case for G Protein-Coupled Receptors. *Med Res Rev* **2014**, *34*, 856–892.

(7) Copeland, R. A. The Drug-Target Residence Time Model: A 10-Year Retrospective. *Nat. Rev. Drug. Discov.* **2016**, *15*, 87-95

(8) Pan, A. C.; Borhani, D. W.; Dror, R. O.; Shaw, D. E. Molecular Determinants of Drug-Receptor Binding Kinetics. *Drug Discov. Today* **2013**, *18*, 667−673.

(9) Russo, S.; Callegari, D.; Incerti, M.; Pala, D.; Giorgio, C.; Brunetti, J.; Bracci, L.; Vicini, P.; Barocelli, E.; Capoferri, L.; Rivara, S.; Tognolini, M.; Mor, M.; Lodola, A. Exploiting Free-Energy Minima to Design Novel EphA2 Protein-Protein Antagonists: From Simulation to Experiment and Return. *Chemistry* **2016**, *22*, 8048-8052.

(10) Cavalli, A.; Spitaleri, A.; Saladino, G.; Gervasio, F. L. Investigating Drug-Target Association and Dissociation Mechanisms Using Metadynamics-Based Algorithms. *Acc. Chem. Res.* **2015**, *48*, 277–285.

(11) Harvey, M. J.; Giupponi, G.; Fabritiis, G. D. ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. *J. Chem. Theory Comput*. **2009**, *5*, 1632−1639.

(12) Stone, J. E.; Phillips, J. C.; Freddolino, P. L.; Hardy, D. J.; Trabuco, L. G.; Schulten, K. Accelerating Molecular Modeling Applications with Graphics Processors. *J. Comput. Chem*. **2007**, 28, 2618−2640.

(13) De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem.* **2016**, *59*, 4035-4061.

(14) Hamelberg, D.; Mongan, J.; McCammon, J. A. Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *J. Chem. Phys.* **2004**, *120,* 11919.

(15) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A*. **2002**, *99*, 12562−12566.

(16) Valsson, O.; Tiwary, P.; Parrinello, M. Enhancing Important Fluctuations: Rare Events and Metadynamics from a Conceptual Viewpoint. *Annu. Rev. Phys. Chem.* **2016**, *67*, 7.1–7.26

(17) Rocchia, W.; Masetti, M.; Cavalli, A. Enhanced sampling methods in drug design. *In Physico-Chemical and Computational Approaches to Drug Discovery*, first edition; Luque, J., Barril, X., Eds.; The Royal Society of Chemistry: Cambridge, United Kingdom, 2012; pp 273−301.

(18) Grubmüller, H. Predicting Slow Structural Transitions in Macromolecular Systems: Conformational Flooding. *Phys. Rev. E* **1995**, *52*, 2893−2906.

(19) Isralewitz, B.; Gao, M.; Schulten, K. Steered Molecular Dynamics and Mechanical Functions of Proteins. *Curr. Opin. Struct. Biol*. **2001**, *11*, 224−230.

(20) Capelli, A. M.; Bruno, A.; Entrena Guadix, A.; Costantino G. Unbinding Pathways From the Glucocorticoid Receptor Shed Light on the Reduced Sensitivity of Glucocorticoid Ligands to a Naturally Occurring, Clinically Relevant Mutant Receptor. *J. Med. Chem.* **2013**, *56*. 7003-7014.

(21) Bortolato, A.; Deflorian, F.; Weiss, D. R.; Mason, J. S. Decoding the Role of Water Dynamics in Ligand-Protein Unbinding: CRF1R as a Test Case. *J. Chem. Inf. Model.* **2015**, *55*, 1857–1866.

(22) Shukla, D.; Hernandez, C. X.; Weber, J. K.; Pande, V. S. Markov State Models Provide Insights into Dynamic Modulation of Protein Function. *Acc. Chem. Res*. **2015**, *48*, 414−422.

(23) Buch, I.; Giorgino, T.; De Fabritiis, G. Complete Reconstruction of an Enzyme-Inhibitor Binding Process by Molecular Dynamics Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 10184−10189.

(24) Mollica, L.; Decherchi, S.; Zia, S. R.; Gaspari, R.; Cavalli, A.; Rocchia, W. Kinetics of Protein-Ligand Unbinding via Smoothed Potential Molecular Dynamics Simulations. *Sci. Rep*. **2015**, *5*, 11539.

(25) Voter, A. F. Hyperdynamics: Accelerated Molecular Dynamics of Infrequent Events. *Phys. Rev. Lett.* **1997**, *78*, 3908–3911.

(26) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100*, 020603.

(27) Tiwary, P.; Parrinello, M. From Metadynamics to Dynamics. *Phys. Rev. Lett.* **2013**, *111*, 230602.

(28) Tiwary, P.; Limongelli, V.; Salvalaglio, M.; Parrinello, M. Kinetics of Protein-Ligand Unbinding: Predicting Pathways, Rates, and Rate-Limiting Steps. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, E386–E391.

(29) Schneider, E. V.; Böttcher, J.; Huber, R.; Maskos, K.; Neumann, L. Structure-Kinetic Relationship Study of CDK8/CycC Specific Compounds. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 8081–8086.

(30) Maestro, version 10.4; Schrödinger, LLC, New York, 2015.

(31) Prime, version 4.2; Schrödinger, LLC: New York, 2015.

(32) Russo, A. A.; Tong, L.; Lee, J. O.; Jeffrey, P. D.; Pavletich, N. P. Structural Basis for Inhibition of the Cyclin-Dependent Kinase Cdk6 the Tumor Suppressor p16ink4a. *Nature* **1998**, *395*, 237-243.

(33) Banks, J. L.; Beard, H. S.; Cao, Y.; Cho, A. E.; Damm, W.; Farid, R.; Felts, A. K.; Halgren, T. A.; Mainz, D. T.; Maple, J. R.; Murphy, R.; Philipp, D. M.; Repasky, M. P.; Zhang, L. Y.; Berne, B. J.; Friesner, R. A.; Gallicchio, E.; Levy, R. M. Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J. Comput. Chem.* **2005**, *26*, 1752−1780.

(34) MacroModel, version 11; Schrödinger, LLC, New York, 2015.

(35) Loncharich, R. J.; Brooks, B. R.; Pastor, R. W. Langevin Dynamics of Peptides: The Frictional Dependence of Isomerization Rates of N-acetylalanyl-N'-methylamide. *Biopolymers* **1992**, *32*, 523-535.

(36) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N Log (N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

(37) Desmond Molecular Dynamics System, version 4.4; D. E. Shaw Research: New York, 2015; Maestro−Desmond Interoperability Tools, version 4.4; Schrödinger: New York, 2015.

(38) Salvalaglio, M.; Tiwary, P.; Parrinello, M. Assessing the Reliability of the Dynamics Reconstructed from Metadynamics. *J. Chem. Theory Comput*. **2014**, *10*, 1420−1425.

(39) Eyring, H. The Activated Complex in Chemical Reactions. *J. Chem. Phys.* **1935**, *3*, 107–115.

(40) Valsson, O.; Tiwary, P.; Parrinello, M. Enhancing Important Fluctuations: Rare Events and Metadynamics from a Conceptual Viewpoint. *Annu. Rev. Phys. Chem*. **2016**, *67*, 159-184.

(41) Tiwary, P.; Mondal, J.; Morrone, J. A.; Berne, B. J. Role of Water and Steric Constraints in the Kinetics of Cavity-Ligand Unbinding. *Proc. Natl. Acad. Sci. USA*. **2015**, *112*, 12015-12019.

(42) Hall Jr, H. K. Potentiometric Determination of the Base Strength of Amines in Non-protolytic Solvents. *J. Phys. Chem.* **1956**, *60*, 63–70.

(43) Swails, J. M.; York, D. M.; Roitberg, A. E. Constant pH Replica Exchange Molecular Dynamics in Explicit Solvent Using Discrete Protonation States: Implementation, Testing, and Validation. *J Chem. Theory. Comput.* **2014**, *10*, 1341-1352.

(44) Donnini, S.; Tegeler, F.; Groenhof, G.; Grubmuller, H. Constant pH Molecular Dynamics in Explicit Solvent with λ-Dynamics. *J Chem. Theory. Comput.* **2011**, *7*, 1962–1978.

(45) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shaw, D. E.; Shelley, M.; Perry, J. K.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy, *J. Med. Chem.* **2004**, *47*, 1739–1749.

TABLE OF CONTENTS (TOC)