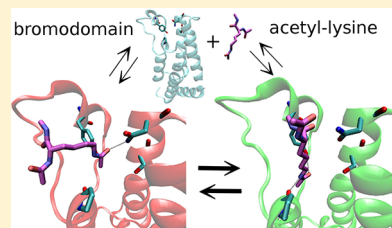


# Mechanism and Kinetics of Acetyl-Lysine Binding to Bromodomains

A. Magno,<sup>†,‡</sup> S. Steiner,<sup>†,‡</sup> and A. Caflisch<sup>\*,†</sup><sup>†</sup>Department of Biochemistry, University of Zurich, Winterthurerstrasse 190 CH-8057 Zurich, Switzerland

## S Supporting Information

**ABSTRACT:** Bromodomains are four-helix bundle proteins that specifically recognize acetylation of lysine side chains on histones. The available X-ray structures of bromodomain/histone tail complexes show that the conserved Asn residue in the loop between helices B and C is involved in a hydrogen bond with the acetyl-lysine side chain. Here we analyze the spontaneous binding of acetyl-lysine to the bromodomain TAF1(2) by the first molecular dynamics simulations of histone mark binding to an epigenetic reader protein. Multiple events of reversible association sampled along the unbiased simulations allow us to determine the pathway and kinetics of binding. The simulations show that acetyl-lysine has two major binding modes in TAF1(2) one of which corresponds to the available crystal structures and is stabilized by a hydrogen bond to the conserved Asn side chain. The other major binding mode is more buried than in the crystal structures and is stabilized by two hydrogen bonds with conserved residues of the loop between helices Z and A. In the more buried binding conformation, three of the six structured water molecules at the bottom of the binding pocket are displaced by the acetyl-lysine side chain. The kinetic analysis shows that the two binding modes interconvert on a faster time scale with respect to the association/dissociation process. The atomic-level description of the binding pathway and binding modes is useful for the design of small molecule modulators of histone binding to bromodomains.



## INTRODUCTION

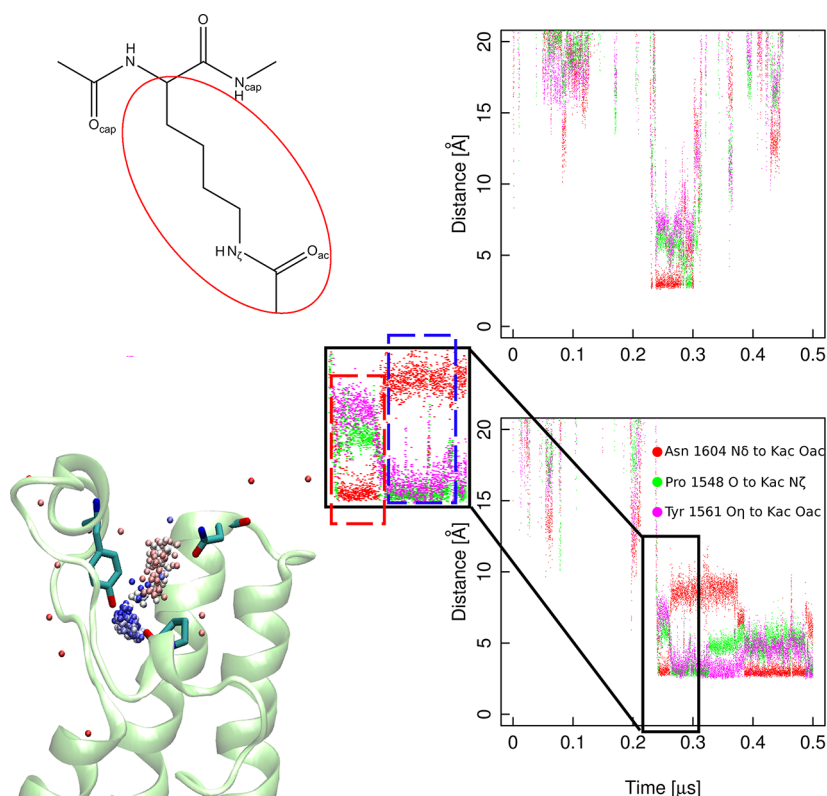
Bromodomains are protein modules of about 110 residues which recognize acetylated lysine side chains mainly in histones and are thus involved in transcriptional regulation.<sup>1,2</sup> In the human genome 46 proteins with a total of 61 different bromodomains have been identified, with up to 6 bromodomains per protein.<sup>3</sup> Bromodomains occur often in tandem with a PHD module or with other bromodomain(s), suggesting that combinatorial readout may be at work in translating the histone code.<sup>2,4,5</sup> The potential role of bromodomains in tumors and inflammation<sup>4,6,7</sup> has spurred large-scale structural studies with the ultimate goal to facilitate the discovery of small-molecule inhibitors able to interfere in the process of reading acetylated lysine.<sup>8,9</sup> In 1999, the first three-dimensional structure of a human bromodomain was solved;<sup>10</sup> since then, the crystal and/or solution structures of more than 40 human bromodomains have been reported.<sup>11,12</sup>

All available structures show a conserved four-helix bundle topology in which the ZA-loop and BC-loop connect the first two  $\alpha$  helices (called Z and A) and last two  $\alpha$  helices (called B and C), respectively.<sup>12,13</sup> The acetyl-lysine binding site is very similar in all structures of bromodomains.<sup>12</sup> There are several crystal structures with bromodomains in complex with acetyl-lysine<sup>4</sup> (e.g., PDB codes 3P1C, 3O34, and 3O35) or peptides with two acetylated lysine side chains (PDB codes 3UVW, 3UVX, 3UVY, and 3UW9). These structures share a common intermolecular hydrogen bond between the side chain of the conserved Asn in the BC loop and the acetyl oxygen of the ligand. The binding mode is further stabilized by a network of water-mediated hydrogen bonds at the bottom of the binding groove and hydrophobic contacts.<sup>14</sup> In addition to the

structural characteristics, thermodynamic parameters of histone tail binding have been measured by isothermal titration calorimetry using mono- and diacetylated histone peptide segments.<sup>4</sup> The measured dissociation constants are in the high micromolar range, and seem rather insensitive to the peptide length. In general, the dynamical character of the process of ligand binding makes it a challenging investigation subject.<sup>15</sup> Molecular dynamics (MD) simulations with explicit solvent provide a useful tool to study binding of small molecules to proteins at the atomic level of detail.<sup>15–22</sup>

Here, we have employed MD simulations in conjunction with cut-based free energy profile analysis<sup>15,23,24</sup> to shed light on the spontaneous association of acetyl-lysine to the second bromodomain of the human transcription initiation factor TFIID subunit 1 (abbreviated as TAF1(2)). This bromodomain was chosen based on a previous study<sup>25</sup> in which the side chain of the conserved Asn showed two interconverting orientations. In particular, the side chain amide nitrogen of TAF1(2) Asn1604 populates a solvent exposed orientation which is rotated by about 180 degrees with respect to the orientation observed in most crystal structures. In this context there are several interesting questions which can be addressed by simulations: Is conventional MD able to reproduce the experimentally observed binding mode of the acetyl-lysine side chain? On which time scale does spontaneous binding take place? Which role does the side chain orientation of the conserved Asn play in the binding process? Are there alternative binding modes? The simulations describe with

Received: May 3, 2013



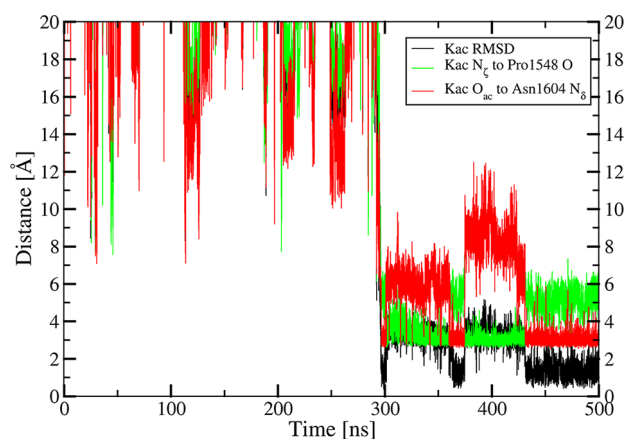
**Figure 1.** Spontaneous binding of Kac to TAF1(2). (top left) Two-dimensional structure of Kac. The red oval highlights the Kac side chain. (bottom left) Spheres representing the position of the carbonyl carbon atom of the Kac side chain. Their color reflects the simulation time from red to blue. The backbone of TAF1(2) and the side chains involved in binding are shown by a green ribbon and sticks, respectively. (right) Time series of key interactions along two independent simulations. The inset emphasizes the binding event shown in the bottom left panel and in Supporting Information Movie S1. The red and blue dashed rectangles in the inset illustrate the two most populated binding modes and correspond to the colors of the spheres in the bottom left panel.

atomistic resolution acetyl-lysine binding to the position observed in the crystal structure. Moreover, they reveal an alternative binding mode and provide kinetic information on the binding/unbinding processes.

## RESULTS AND DISCUSSION

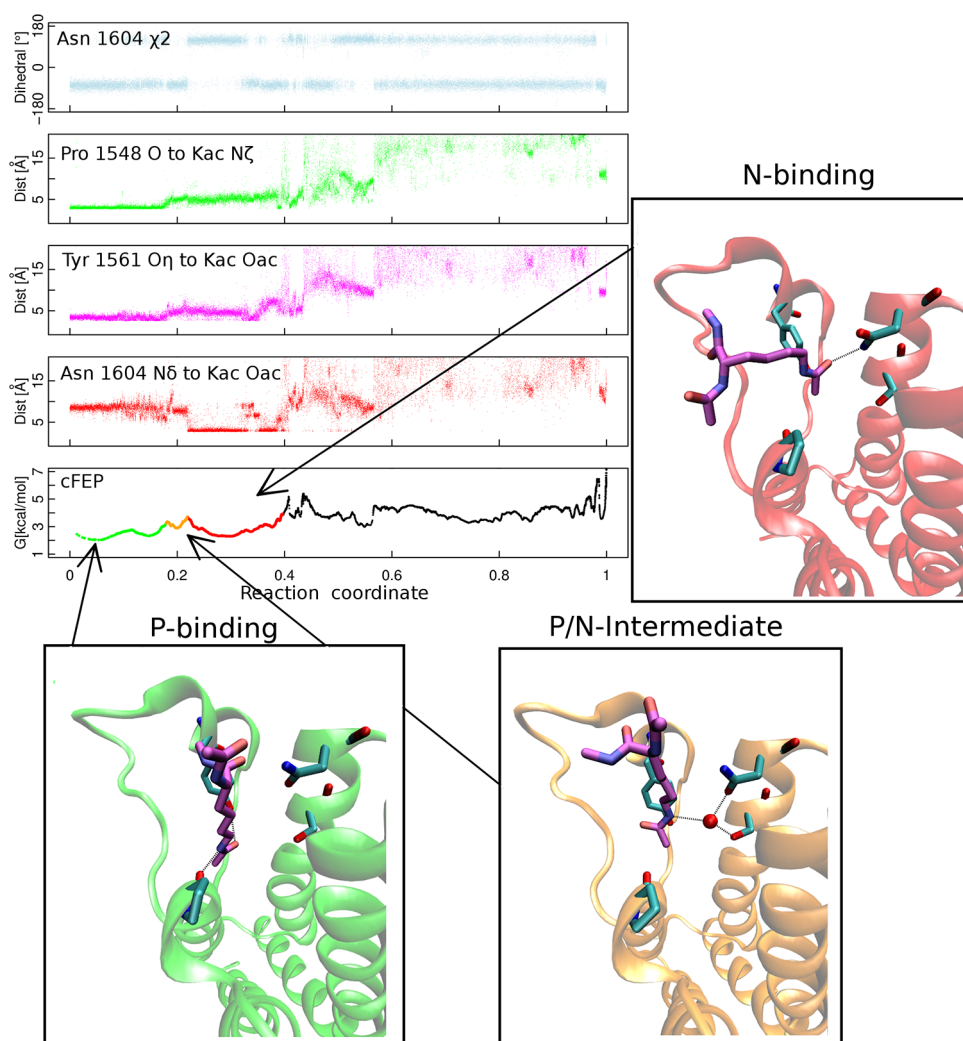
Twenty-four independent simulations of 0.5  $\mu$ s each were started with acetylated lysine (with neutral blocking groups at its termini, abbreviated as Kac, see Figure 1 for the 2D structure) positioned randomly at a distance of more than 20 Å from the binding site of TAF1(2). No artificial guiding force or bias toward the binding site was used. In 16 of the 24 MD runs at least one event of spontaneous binding to the acetyl-lysine pocket was observed. In the remaining 8 MD runs Kac did not reach the binding site and remained most of the time in the solvent or was sporadically involved in unspecific association to different regions of the TAF1(2) surface. Unless stated otherwise, the following analysis refers to the 8- $\mu$ s sampling of the 16 runs in which binding events were observed (a total of  $4 \times 10^5$  snapshots).

**Spontaneous Binding.** The time series of some of the key interactions show that spontaneous and reversible binding of Kac to the acetyl-lysine pocket can take place in the 0.5  $\mu$ s time scale of the MD runs (Figures 1 and 2 and Supporting Information Figures S1, S2, and movie files, ct400361k\_si\_001.mpg (Movie S1) and ct400361k\_si\_002.mpg (Movie S2)). From the analysis of hydrogen bond frequencies, it emerges that the hydrogen bond between the acetyl oxygen of the Kac side chain and the side



**Figure 2.** Spontaneous binding to the crystallographic binding mode. The plot shows the time series of the RMSD from the X-ray structure (black) together with the distances that characterize the hydrogen bond observed in the crystal structures (red) and the hydrogen bond of the more buried binding mode revealed by the present MD study (green). Note the frequent interconversions between the two binding modes. The RMSD of the Kac side chain atoms was calculated upon optimal overlap of the  $C_{\alpha}$  atoms in the four helices. The reference structure for the RMSD calculation is the crystal structure of CREBBP in complex with Kac (PDB code 3P1C) as there is no crystal structure of TAF1(2) with Kac or a histone tail peptide. The MD run used for this plot is shown in Supporting Information Movie S2.

chain of the conserved Asn1604 in the BC loop is present in nearly 40% of the snapshots of the bound state. Interestingly,



**Figure 3.** Geometric variables used to analyze binding (top) and cut-based free energy profile (abbreviated as cFEP) calculated using the mean first passage time from the representative node of the P-binding mode (bottom). The reaction coordinate of the cFEP is the relative partition function (see the Methods section). The colors of the cFEP reflect the three main basins whose representatives are shown with TAF1(2) in ribbon model and Kac in sticks (insets). The unbound state is in black in the cFEP. The 400 000 MD snapshots were ordered according to the free-energy sub-basin they belong to. In this way, each of the sub-basins on the free energy profile is annotated with the corresponding values of the geometric variables. As an example, the first sub-basin on the left is the P-binding mode which has hydrogen bond distance for Pro 1548 O—Kac  $N_{\zeta}$  and Tyr1561  $O_{\eta}$ —Kac  $O_{ac}$  whereas the Asn1604  $N_{\delta}$ —Kac  $O_{ac}$  distance is about 8–9 Å.

two residues of the ZA loop are also involved in hydrogen bonds with Kac. The backbone oxygen of Pro1548 (in the so-called WPF shelf) and the side chain hydroxyl of Tyr1561 act frequently as hydrogen bond acceptor and donor, respectively.

**Two Major Binding Modes.** The trajectories with spontaneous events of association and the corresponding time series of key intermolecular distances indicate that there are two distinct binding modes (red and blue spheres, respectively, in Figure 1 bottom, left). From the unbound state, Kac adopts first a binding mode matching the one observed in several crystal structures with the hydrogen bond between the Asn1604  $N_{\delta}$  atom and the Kac side chain oxygen  $O_{ac}$  (called N-binding mode hereafter). The time series of the Kac root-mean-square deviation (RMSD) from the crystallographic binding mode shows several simulation segments during which the RMSD oscillates between 0.5 and 1.5 Å (e.g., the last 60 ns of the MD run shown in Figure 2) which confirms that the N-binding mode corresponds to the binding mode determined experimentally. Note that the structure of

TAF1(2) does not change upon Kac binding as monitored by the RMSD of the  $C_{\alpha}$  atoms from the structure at the beginning of the MD run (see Supporting Information Figure S3) and also shown in Supporting Information Movie S2.

From the N-binding mode, Kac can reach reversibly a binding conformation which is more buried than in the crystal structure and in which the oxygen atom of Pro1548 acts as hydrogen bond acceptor for the Kac side chain  $N_{\zeta}$  (called P-binding mode hereafter). Further stabilization is provided by a hydrogen bond between the hydroxyl group of Tyr1561 and the Kac side chain  $O_{ac}$ , as well as between atom  $N_{\delta}$  of Asn1604 and either of the two oxygen atoms in the backbone moiety of Kac, i.e.,  $O_{cap}$  (N-terminal capping group) or O (C-terminal group). An MD run with a spontaneous binding event and further interconversions between N- and P-binding modes is shown in Figure 1 and Supporting Information Movie S2.

Clustering of the snapshots based on pairwise RMSD of the Kac side chain atoms and consecutive calculation of the cut-based free energy profile<sup>23</sup> confirm the existence of two major



binding modes (Figure 3). It is important to note that the P-binding mode is more buried (Figure 1 bottom, left) and slightly more populated than the N-binding mode. The former includes about 50% of the MD snapshots in the bound state while the latter includes about 40%. Since the P-binding mode has not been reported as of today, we decided to run additional simulations of Kac binding to TAF1(2) with a force field different from CHARMM27.<sup>26</sup> The N- and P-binding modes are also observed in each of four 1  $\mu$ s runs with the AMBER<sup>27</sup> force field, and they show similar statistical weight (Supporting Information Figure S4).

We also wanted to exclude the idea that the more buried P-binding mode is an artifact due to the simplified model of the ligand. For this purpose, we run eight additional simulations of 1  $\mu$ s each with the Kac-Gly-Gly-Kac tetrapeptide (derived from the diacetylated histone tail H4K5/K8) using two different bromodomains (i.e., CREBBP and BRD4(1)) and with two different force fields (CHARMM27<sup>26</sup> and AMBER,<sup>27</sup> Supporting Information Table S1). The P-binding mode is populated also by the tetrapeptide ligand in each of the two bromodomains and irrespective of the force field used (Supporting Information Figure S4). Thus, the results of the control simulations provide evidence that the P-binding mode is not an artifact of the force field or the simplified ligand. It is also important to note that in the N-binding mode the RMSD from the crystal structure of the side chain of the N-terminal Kac is almost always below 2 Å (Supporting Information Figure S5). Moreover, the C-terminal Kac, i.e., the acetyl-lysine side chain that is not in contact with the conserved Asn, has a more pronounced flexibility than the N-terminal Kac (Supporting Information Figure S5) which is consistent with the larger temperature factors of the C-terminal than the N-terminal acetyl-lysine in the crystal structure of the complex with BRD4(1) (PDB code 3UVW).

**Kinetics.** The existence of two major binding modes calls for further investigations of the time scales required for binding/unbinding and how they relate to the transitions within the bound state. For this analysis, first individual free energy basins were isolated by the cut-based free energy profile method following an iterative procedure reported previously<sup>24</sup> (see also the Methods section). Besides the two major binding modes, there is a metastable state (called P/N-intermediate and populated at about 10%) formed by MD snapshots in which the side chain of Kac points inside the binding groove but is arranged such that direct hydrogen bonds with Pro1548, Tyr1561, or Asn1604 are not possible. This binding conformation is stabilized mainly by a water-bridged hydrogen bond between Kac N $\epsilon$  and Asn1604 O $\delta$  (see bottom, right inset in Figure 3).

The mean first passage time (MFPT) between free energy basins was evaluated using representative clusters selected according to statistical weight as published previously.<sup>24</sup> The MFPT values indicate that there are two well-separated time scales (Table 1). While the two major binding modes interconvert relatively fast, i.e., with an MFPT of  $\approx 20$  ns, binding/unbinding takes place on a much slower time scale (MFPT  $\approx 300$  ns). Figure 4 illustrates these two kinetic regimes by the free energy profiles plotted together with the MFPT to the reference state. The free energy barrier to exit the bound state (located at a value of the reaction coordinate of about 0.4 in Figure 4) is significantly higher than the barriers between the three basins in the bound state. The number of direct transitions between basins is shown in Table 1. The large

**Table 1. Kinetics Data Extracted from the Simulations<sup>a</sup>**

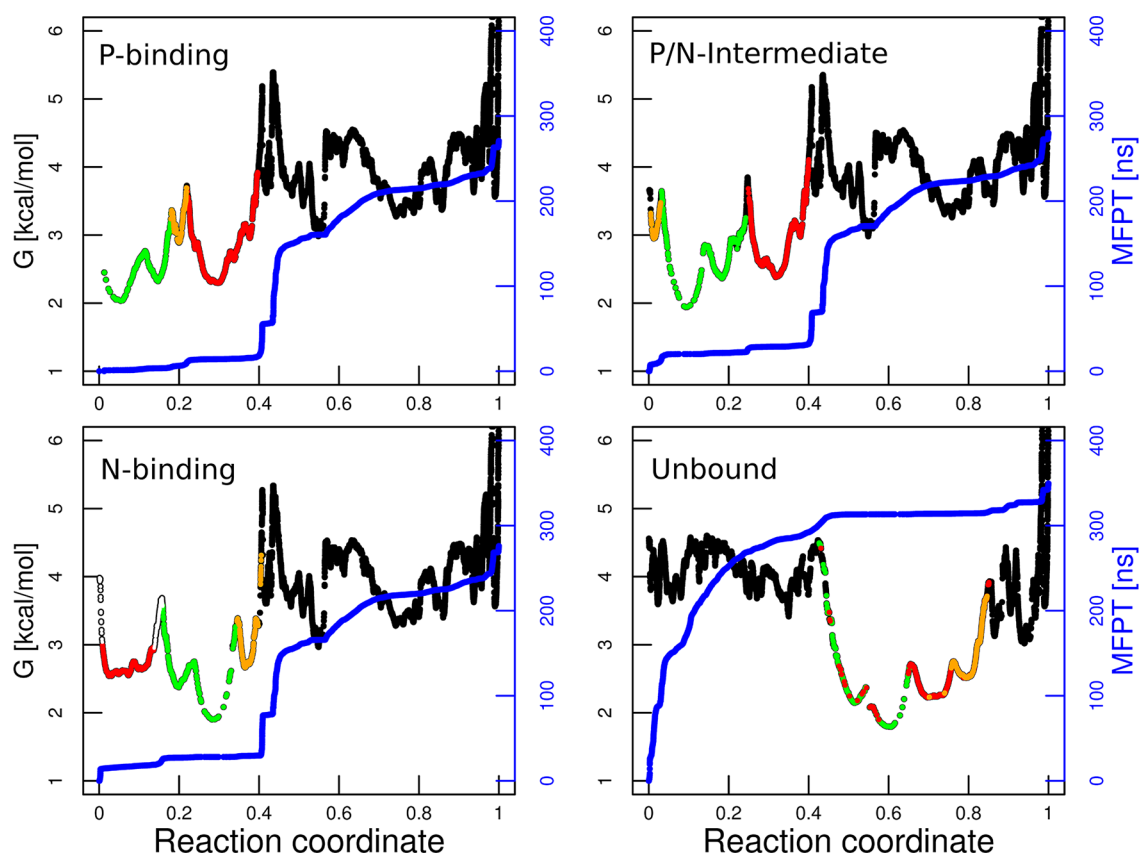
transitions (MFPT [ns]) from\to	unbound	P-binding	N-binding	P/N-intermediate
unbound		1 (220)	14 (220)	5 (220)
P-binding	2 (320)		129 (28)	310 (20)
N-binding	4 (320)	135 (15)		17 (29)
P/N-intermediate	4 (320)	312 (6)	16 (30)	

<sup>a</sup>The table shows the number of direct transitions between pairs of states as observed in the MD simulations. The MFPT values in nanoseconds are given in parentheses. Note that the MFPT values are calculated from the transition matrix, i.e., they take into account both direct and indirect transitions.

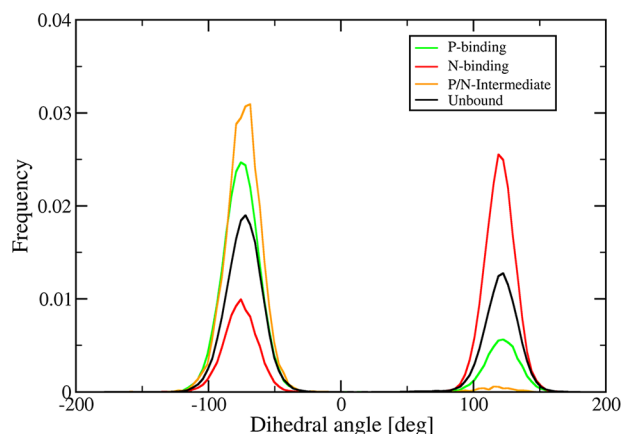
number of direct transitions between P-binding and N-binding modes compared to the much smaller number of binding/unbinding events reflects the kinetic separation of the two processes which emerges also from the MFPT calculations. It is important to note that the number of binding and unbinding events (20 and 10, respectively) determined using the definition of basins by the cut-based free energy profile agrees with the visual inspection of the trajectories and the time series presented in Supporting Information Figures S1 and S2. Such agreement provides strong evidence that the clustering and the definition of free energy basins by the cut-based free energy profile method capture the relevant features of the binding and unbinding processes.

**Dissociation Constant.** The multiple events of spontaneous and reversible binding allow one to calculate the dissociation constant from the ratio of the off-rate and on-rate. The off-rate is the inverse of the MFPT from any of the binding modes to the unbound state (i.e.,  $1/320$  ns<sup>-1</sup>; see Table 1). The on-rate is the inverse of the MFPT to any of the bound states multiplied by the concentration of Kac in the simulation box ( $1/(220 \text{ ns} \times 6 \text{ mM})$ ). It thus follows that the dissociation constant is about  $(220 \times 6)/320 \approx 4$  mM. This value is about 1 order of magnitude higher than the dissociation constants of mono- and diacetylated histone peptides as measured by isothermal titration calorimetry<sup>4</sup> which is probably a consequence of the smaller ligand used in the simulations than in the experiments.

**Orientation of the Side Chain Amide Group of the Conserved Asn.** The preferred orientation of the side chain of the conserved Asn1604 differs from the one observed in most bromodomain crystal structures, i.e., the N $\delta$  atom points preferentially toward the solvent ( $\chi_2 \approx -70^\circ$ , Figure 3 top) rather than the bottom of the binding site ( $\chi_2 \approx 120^\circ$ ). The solvent exposed orientation of the Asn1604  $-\text{NH}_2$  group is stabilized by a water-bridged hydrogen bond between the Asn side chain carbonyl and the side chain hydroxyl of Ser1600 in the C-terminal turn of helix B. This observation raises the question whether different binding modes of Kac prefer different orientations of the Asn1604 side chain. Here, it is important to note that based on a previous work<sup>25</sup> half of the simulation runs were started with the solvent exposed orientation of the Asn1604  $-\text{NH}_2$  group and the other half with its orientation toward the binding site (see Methods). Figure 5 shows the histograms of the Asn1604  $\chi_2$  dihedral values computed over the snapshots belonging to the three individual binding modes as well as for the segments of the trajectories during which the ligand is unbound. Note, that for the unbound conformation the histogram is characterized by two peaks corresponding to the two different orientations with a preference for the solvent exposed orientation. Importantly,



**Figure 4.** Four panels show the cut-based free energy profiles with the three equilibrium bound conformations or the unbound state as reference. Each panel is labeled with the state used as reference which is the first basin on the left. Individual basins are colored according to the same color scheme as in Figure 3, i.e., green, red, orange, and black for P-binding mode, N-binding mode, P/N-intermediate, and unbound state, respectively. The empty circles in the bottom left panel represent the nodes on the barrier between the N-basin and the unbound state. The reaction coordinate is the relative partition function (see the Methods section). The MFPT values to the representative node of the reference basin are also shown (blue lines with y-axis on the right). The MFPT is a monotonous growing function because the nodes are sorted according to their kinetic distance from the reference basin. Note that the three binding modes have approximately equal kinetic separation from the unbound conformation (i.e., 320 ns; see Table 1). Thus, there is a significant overlap of the three binding modes when plotting the profile from a node belonging to the unbound conformation (bottom right panel).



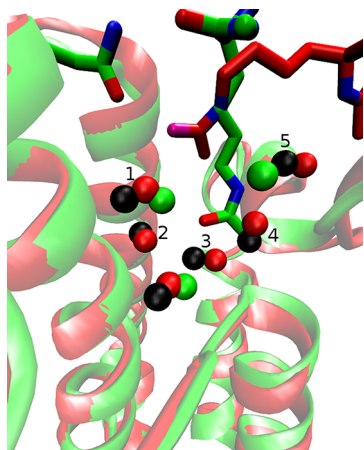
**Figure 5.** Histograms of the two main conformations of the  $\chi_2$  angle of the conserved Asn1604 for the individual binding modes (colors) and the unbound state (black). Positive values of the  $\chi_2$  angle correspond to the orientation of Asn1604  $N_\delta$  toward the bottom of the binding site while negative values to the orientation of Asn1604  $N_\delta$  toward the solvent.

as experimentally observed in several X-ray structures of the acetyl-lysine–bromodomain complex, the N-binding conforma-

tion peaks at  $\chi_2 \approx 120^\circ$ , i.e., the orientation of Asn1604  $-NH_2$  pointing toward the binding site. On the contrary, in the P-binding and the P/N-intermediate modes the orientation of the  $-NH_2$  group of Asn1604 toward the solvent is preferred, since in the former it frequently acts as donor for the O or  $O_{cap}$  atoms of Kac, while in the latter a water molecule coordinates  $O_\delta$  of Asn1604,  $O_\gamma$  of Ser1600, and  $N_\epsilon$  of Kac (see Figure 3). Furthermore, it is interesting to analyze the role of Asn1604 during the binding process. For this purpose, the N-binding mode is decomposed into two sub-basins made up of snapshots with positive and negative values, respectively, of the Asn1604  $\chi_2$  dihedral. Strikingly, 12 out of the 14 transitions from the unbound state to the N-binding state take place when the  $N_\delta$  atom of the Asn1604 points toward the solvent (Supporting Information Table S2 and Movie S1). The Kac ligand can then further penetrate the binding groove, ending up in one of the major binding modes.

**Conserved Water Molecules in the Acetyl-Lysine Binding Site.** Six water molecules stabilized by hydrogen bonds with polar groups in the bottom of the acetyl-lysine binding site seem to be conserved in most crystal structures of bromodomains.<sup>8</sup> In the N-binding mode, the six conserved water molecules are arranged in the same fashion as observed in the crystal structures. In contrast, in the P-binding mode, water

molecules 2, 3, and 4 (nomenclature of Figure 1 of ref 8) are replaced by the Kac side chain which, as mentioned above, is more buried than in the N-binding mode (Figure 6). This



**Figure 6.** Structured water molecules are conserved in the N-binding mode while three of them are displaced by the tip of the Kac side chain in the P-binding mode. The representative snapshots of the N-binding mode (red) and P-binding mode (green) are aligned structurally to the crystal structure of apo TAF1(2) (PDB code 3UV4) using only the  $C_{\alpha}$  atoms of the bromodomain helices. The bromodomain backbone, Kac, and structured water molecules are shown (ribbons, sticks, and spheres, respectively) for the two binding modes. For the crystal structure, only the structured water molecules are shown (black spheres) to avoid overcrowding of the figure. The Asn1604 side chain of the P-binding mode representative is also shown (carbon atoms in green). The labeling of the structured water molecules is the same as in ref 8. Water molecules 2–4 are displaced in the P-binding mode.

simulation result is useful for the design of small molecule modulators of histone binding to bromodomains. In particular, some of the conserved water molecules need not necessarily be present in high-throughput docking campaigns. Moreover, optimization of known binders could focus on the replacement of one or more of the structured water molecules by functional groups with similar hydrogen bonding properties.

## CONCLUSIONS

We have analyzed the pathway and kinetics of Kac binding to the TAF1(2) bromodomain by multiple MD simulations of spontaneous and reversible binding. The MD trajectories reveal two major binding modes of Kac which differ in the bromodomain residues providing the anchoring interactions. From the unbound state, Kac reaches first a position in which it is stabilized by a hydrogen bond between its side chain carbonyl oxygen and the side chain amide of the conserved Asn. This binding mode is identical to the binding mode observed in several crystal structures of bromodomains with Kac or short peptide segments from histone tails. From the crystallographic binding mode Kac further penetrates the binding site and reaches a more buried binding mode in which the Kac side chain is involved in a hydrogen bond with the carbonyl group of the conserved Pro in the so-called WPF shelf. The WPF shelf is a highly conserved motif of the bromodomain sequence and has been highlighted as an important hotspot to increase binding affinity of small molecule inhibitors.<sup>28,29</sup> Moreover, the binding of diacetylated histone tails involves the WPF shelf which interacts with the acetyl-lysine side chain that is not in contact with the conserved Asn.<sup>4,30</sup> Furthermore, the six

structured water molecules observed in most crystal structures are present in the crystallographic binding conformation, while three of them are replaced by the Kac side chain in the more buried binding mode. Importantly, in several crystal structures of bromodomains in complex with inhibitors (e.g., PDB codes 4HXX, 4HXO, 4IR3, 4A9I, and 4ALC) the carbonyl oxygen of the Pro of the WPF shelf is involved in a polar interaction with the inhibitor, e.g., with an aromatic CH of the [1,2,4]triazolo-[4,3-*a*]pyridine moiety in the complex structure with BRD4(1) (PDB code 4HXO).<sup>31</sup> The crystal structure of the complex between BRD4(1) and 6,7-dihydrothieno[3,2-*c*]pyridin-5(4*H*)-yl(1*H*-imidazol-1-yl)methanone (PDB code 4HXX) is of particular interest as the 6,7-dihydrothieno[3,2-*c*]pyridin moiety is almost as buried as acetyl-lysine in the P-binding mode (Supporting Information Figure S6). Moreover, one of the structured water molecules is displaced by the dihydrothieno ring.<sup>31</sup>

The two binding modes observed in the MD simulations are distinct and separated by a free energy barrier which is smaller than the barrier of unbinding. In other words, upon binding there are several transitions between the two binding modes which can interconvert about 1 order of magnitude faster than the time required for full dissociation which is about 300 ns. The simulation methods used here, unbiased MD with explicit solvent (carried out on conventional compute clusters) and cut-based free energy profile analysis, can be applied to determine the binding mode(s) and kinetics of small molecules to proteins thus providing dynamical information which is useful to complement the static picture obtained by conventional experimental techniques. Although the histone code may consist of multiple post-translational modifications, the mechanism and kinetics of acetyl-lysine binding to bromodomains is a first step toward the in-depth understanding of the role of bromodomains in transcriptional regulation.

## METHODS

**Simulation Protocols.** The coordinates of the TAF1(2) bromodomain (PDB code 3UV4) were downloaded from the protein database (URL [www.pdb.org](http://www.pdb.org)).<sup>32</sup> To reproduce neutral pH conditions, the side chains of aspartates and glutamates were negatively charged, those of lysines and arginines were positively charged, and the histidine side chains were neutral. Thirteen crystal water molecules inside or close to the binding site were kept while the remaining water molecules were deleted. Subsequently, the structure was solvated in a water box whose size was chosen to have a minimal distance of 12 Å between the boundary and any atom of the protein. The simulation system contained sodium and chloride ions to approximate an ionic strength of about 150 mM and to compensate for the total charge of the bromodomain. Preliminary MD simulations were carried out with CHARMM,<sup>33</sup> while the production runs were carried out with GROMACS 4.5.4<sup>34</sup> using the CHARMM PARAM27 all atom force field<sup>26,35</sup> and the TIP3P model of water.<sup>36</sup>

Periodic boundary conditions were applied, and electrostatic interactions were evaluated using the particle-mesh Ewald summation method.<sup>37</sup> The van der Waals interactions were truncated at a cutoff of 10 Å. The temperature of 310 K was kept constant by an external bath with velocity rescaling,<sup>38</sup> and the pressure was kept close to 1 atm by the Berendsen barostat.<sup>39</sup> The LINCS algorithm was used to fix the covalent bonds involving hydrogen atoms. The integration time step was 2 fs, and snapshots were saved every 20 ps.



Multiple independent MD runs of spontaneous binding of Kac to TAF1(2) were carried out. To mimic the context of the longer histone sequence, the N-terminus and C-terminus of Kac were neutralized by capping with acetyl and *N*-methylamide, respectively (Figure 1). Two different starting structures were generated for TAF1(2) by swapping the coordinates of Asn1604 O<sub>δ</sub> and N<sub>δ</sub> atoms in the X-ray structure (before hydrogen generation). Thus two different  $\chi_2$  dihedrals of Asn1604 were represented. Six starting positions for the Kac molecule were generated by randomly placing Kac with its center of mass  $\approx 20$  Å away from the center of mass of Asn1604 using the package PyMOL.<sup>40</sup> For each of the six Kac starting positions and each of the two TAF1(2) starting structures, two simulations with different seeds were started for a total 24 independent runs of 0.5  $\mu$ s each (Supporting Information Table S1). Each 0.5  $\mu$ s run took about 1 week on 64 cores of Xeon 5500-series processors.

Control simulations were carried out with the tetrapeptide Kac-Gly-Gly-Kac to investigate potential differences between the isolated Kac and histone tails. As in the runs with Kac, the tetrapeptide was capped with acetyl and *N*-methylamide groups at the N-terminus and C-terminus, respectively. The coordinates of the tetrapeptide and of the bromodomain BRD4(1) were taken from the PDB structure 3UVW. For CREBBP, the tetrapeptide was placed into its binding site after structural overlap of the CREBBP coordinates (PDB entry code 3P1C) onto the ones of BRD4(1). From each of these two starting structures, two independent 1  $\mu$ s runs were performed with the CHARMM PARAM27 force field. In addition, two independent 1  $\mu$ s simulations with the AMBER force field were carried out for each of the two bromodomains. An overview of all simulations is given in Supporting Information Table S1.

**Ligand Parametrization.** For the simulations with the CHARMM PARAM27 force field, Kac was parametrized using the CHARMM general force field<sup>41</sup> for small molecules which is compatible with CHARMM PARAM27. For the simulations performed with the AMBER99 SB<sup>27</sup> force field, the parameters of the Kac side chains were taken from a previous computational study.<sup>42</sup>

**Clustering and Cut-Based Free Energy Profile.** The 16 MD trajectories with one or more binding events were used for further analysis (a total of 400 000 snapshots from 8  $\mu$ s of MD). The snapshots were aligned structurally to the first frame of one of the MD trajectories using only the C <sub>$\alpha$</sub>  atoms of the helical segments of TAF1(2). Subsequently, the snapshots were clustered using a tree-based algorithm<sup>43</sup> implemented in CAMPARI.<sup>44</sup> As a measure of pairwise similarity, the root-mean-square deviation (RMSD) of the acetyl-lysine side chain heavy atoms including C <sub>$\alpha$</sub>  was computed without further alignment. The clustering was carried out with eight levels of the tree using 8 and 1 Å as RMSD thresholds for the coarsest and finest clustering level, respectively. This led to a total number of 149 024 clusters 126 552 of which contained only one snapshot (called singletons). The number of singletons seems large, but it has to be noted that in the unbound state, where the peptide moves freely around the bromodomain, the RMSD between two relative positions of Kac is almost always larger than 1 Å and thus the unbound state consists mainly of singletons. Other clustering methods which are able to partially solve the problem of the singletons were also tested but led to worse kinetic separation of the different states. Upon clustering, a free energy profile was computed using a method based on the equilibrium kinetic network which preserves the free energy

barriers.<sup>23,24,45</sup> This approach emulates the cuts in flow-networks, and the computed profile is therefore named cut-based free energy profile. The nodes and links of the equilibrium kinetic network are the clusters (determined by the tree-based algorithm as explained above) and the direct transitions between them sampled along the MD runs, respectively.<sup>46</sup> For each node, the set of nodes is partitioned into two groups A and B using the mean first passage time (MFPT) to the reference node as an order parameter. The free energy is related to the maximum flow between sets A and B and calculated as  $G = -kT \ln(Z_{AB}/Z)$ , where  $Z_{AB}/Z$  is the relative partition function which represents the statistical weight of the transitions between sets A and B (see refs 23 and 24 for details). The result is a one-dimensional profile along the reaction coordinate  $Z_A/Z$  (i.e., the relative partition function representing the statistical weight of set A) which preserves the barrier height between the free energy basins. The evaluations of cut-based free energy profiles, interatomic distances, and dihedral angles were performed with the simulation analysis package WORDOM.<sup>47,48</sup>

**Identification of Basins.** The free energy basins were isolated from the cut-based free energy profile using a previously published procedure.<sup>24</sup> First the most populated cluster is used as reference to isolate the first basin which consists of the set of clusters on the left side of the first barrier on the profile. Due to possible overlap on the right side of the barrier (i.e., similar kinetic distance from two or more basins to the reference cluster), the isolation of each basin requires a separate profile<sup>24</sup> (Supporting Information Figures S7–S10). In other words, the procedure used for the first basin is iterated for the second basin which is isolated using as a reference cluster the largest cluster that is not in the first basin. Clusters on top of barriers are not assigned to any basin to avoid spurious fluctuations, i.e., shortcuts along the trajectories. Transitions between basins were calculated from the MD trajectories when the system passed from one basin to another either by a direct transition or via the not assigned snapshots (the barrier region).

**Analysis of Hydrogen Bonds.** Hydrogen bonds were analyzed using GROMACS.<sup>34</sup> A hydrogen bond between a donor and an acceptor atom was considered to be formed if the donor–acceptor distance was lower than 3.5 Å and the acceptor–donor–hydrogen angle lower than 30°.

## ■ ASSOCIATED CONTENT

### ● Supporting Information

Simulation protocol (Table S1); transition matrix computed after splitting of the N-binding mode according to the  $\chi_2$  dihedral angle of the conserved Asn (Table S2); time series of key interatomic distances for all the performed simulations (Figures S1–S4); time series of the root-mean-square displacement from the 3P1C crystal structure along a 500 ns simulation segment (Figure S3, bottom); time series of the root-mean-square displacement of Kac-Gly-Gly-Kac in complex with BRD4(1) and CREBBP from the 3UVW crystal structure (Figure S5); structural overlap of the representative snapshots of P- and N-binding mode with the structure of the complex between BRD4(1) and inhibitor reported in 4HXX (Figure S6); geometric annotation and cut-based free energy profile projections with representative nodes of the P-binding mode, N-binding mode, P/N-intermediate, and unbound state as reference nodes (Figure S7–S10); Movies S1 and S2 as mentioned in the text. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: caflisch@bioc.uzh.ch. Phone: (+41 44) 635 55 21.  
Fax: (+41 44) 635 68 62.

### Author Contributions

†A.M. and S.S.: These authors contributed equally to this work.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We thank Drs. Danzhi Huang and Andreas Vitalis for interesting discussions. We also thank Armin Widmer for continuous support with the program WITNOTP. The simulations were carried out on the Schrödinger cluster at the University of Zurich. This work was supported by grants of the Swiss National Science Foundation and the Swiss Cancer League (Krebsliga).

## REFERENCES

- (1) Zeng, L.; Zhou, M.-M. *FEBS Lett.* **2002**, *513*, 124–128.
- (2) Filippakopoulos, P.; Knapp, S. *FEBS Lett.* **2012**, *586*, 2692–2704.
- (3) Arrowsmith, C. H.; Bountra, C.; Fish, P. V.; Lee, K.; Schapira, M. *Nat. Rev. Drug Discov.* **2012**, *11*, 384–400.
- (4) Filippakopoulos, P.; et al. *Nature* **2010**, *468*, 1067–1073.
- (5) Jenuwein, T.; Allis, C. D. *Science* **2001**, *293*, 1074–1080.
- (6) Zuber, J.; et al. *Nature* **2011**, *478*, 524–528.
- (7) Dawson, M. A.; Kouzarides, T.; Huntly, B. J. *N. Engl. J. Med.* **2012**, *367*, 647–657.
- (8) Hewings, D. S.; Rooney, T. P. C.; Jennings, L. E.; Hay, D. A.; Schofield, C. J.; Brennan, P. E.; Knapp, S.; Conway, S. J. *J. Med. Chem.* **2012**, *55*, 9393–9413.
- (9) Furdas, S. D.; Carlino, L.; Sippl, W.; Jung, M. *Med. Chem. Commun.* **2012**, *3*, 123–134.
- (10) Dhalluin, C.; Carlson, J. E.; Zeng, L.; He, C.; Aggarwal, A. K.; Zhou, M.-M. *Nature* **1999**, *399*, 491–496.
- (11) Mujtaba, S.; Zeng, L.; Zhou, M. *Oncogene* **2007**, *26*, 5521–5527.
- (12) Filippakopoulos, P.; Picaud, S.; Mangos, M.; Keates, T.; Lambert, J.-P.; Barsyte-Lovejoy, D.; Felletar, I.; Volkmer, R.; Müller, S.; Pawson, T.; Gingras, A.-C.; Arrowsmith, C. H.; Knapp, S. *Cell* **2012**, *149*, 214–231.
- (13) Vidler, L. R.; Brown, N.; Knapp, S.; Hoelder, S. J. *Med. Chem.* **2012**, *55*, 7346–7359.
- (14) Owen, D. J.; Ornaghi, P.; Yang, J.-C.; Lowe, N.; Evans, P. R.; Ballario, P.; Neuhaus, D.; Filetici, P.; Travers, A. A. *EMBO J.* **2000**, *19*, 6141–6149.
- (15) Huang, D.; Caflisch, A. *PLoS. Comput. Biol.* **2011**, *7*, e1002002.
- (16) Seco, J.; Luque, F. J.; Barril, X. *J. Med. Chem.* **2009**, *52*, 2363–71.
- (17) Guvench, O.; MacKerell, A. D. *J. PLoS. Comput. Biol.* **2009**, *5*, e1000435.
- (18) Basse, N.; Kaar, J. L.; Settanni, G.; Joerger, A. C.; Rutherford, T. J.; Fersht, A. R. *Chem. Biol.* **2010**, *17*, 46–56.
- (19) Pietrucci, F.; Marinelli, F.; Carloni, P.; Laio, A. *J. Am. Chem. Soc.* **2009**, *131*, 11811–11818.
- (20) Buch, I.; Giorgino, T.; De Fabritiis, G. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 10184–10189.
- (21) Huang, D.; Caflisch, A. *ChemMedChem* **2011**, *6*, 1578–1580.
- (22) Dror, R. O.; Pan, A. C.; Arlow, D. H.; Borhani, D. W.; Maragakis, P.; Shan, Y.; Xu, H.; Shaw, D. E. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 13118–13123.
- (23) Krivov, S. V.; Karplus, M. *J. Phys. Chem. B* **2006**, *110*, 12689–12698.
- (24) Krivov, S. V.; Muff, S.; Caflisch, A.; Karplus, M. *J. Phys. Chem. B* **2008**, *112*, 8701–8714.
- (25) Steiner, S.; Magno, A.; Huang, D.; Caflisch, A. *FEBS Lett.* **2013**, *587*, 2158–2163.
- (26) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., III. *J. Comput. Chem.* **2004**, *25*, 1400–1415.
- (27) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712–725.
- (28) Chung, C. W.; et al. *J. Med. Chem.* **2011**, *54*, 3827–3838.
- (29) Chung, C. W.; Dean, A. W.; Woolven, J. M.; Bamborough, P. J. *Med. Chem.* **2012**, *55*, 576–586.
- (30) Moriniere, J.; Rousseaux, S.; Steuerwald, U.; Soler-Lopez, M.; Curtet, S.; Vitte, A.-L.; Govin, J.; Gaucher, J.; Sadoul, K.; Hart, D. J.; Krijgsvel, J.; Khochbin, S.; Muller, C. W.; Petosa, C. *Nature* **2009**, *461*, 664–668.
- (31) Zhao, L.; et al. *J. Med. Chem.* **2013**, *56*, 3833–3851.
- (32) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- (33) Brooks, B. R.; et al. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
- (34) Spoel, D. V. D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701–1718.
- (35) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L. J.; J., D. E.; et al. *J. Phys. Chem. B* **1998**, *102*, 35863616.
- (36) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926935.
- (37) Darden, T.; York, D.; Pedersen, L. G. *J. Chem. Phys.* **1993**, *98*, 10089.
- (38) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.
- (39) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (40) The PyMOL Molecular Graphics System, version 1.3r1; Schrödinger, LLC, New York, 2010.
- (41) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. *J. Comput. Chem.* **2010**, *31*, 671–690.
- (42) Liu, H.; Duan, Y. *Biophys. J.* **2008**, *94*, 4579–4585.
- (43) Vitalis, A.; Caflisch, A. *J. Chem. Theory Comput.* **2012**, *8*, 1108–1120.
- (44) Vitalis, A.; Steffen, A. and Lyle, N.; Mao, A. H.; Pappu, R. V. *CAMPARI v1.0*. <http://sourceforge.net/projects/campari> (accessed April 22, 2013).
- (45) Krivov, S. V.; Karplus, M. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 13841–13846.
- (46) Rao, F.; Caflisch, A. *J. Mol. Biol.* **2004**, *342*, 299–306.
- (47) Seeber, M.; Cecchini, M.; Rao, F.; Settanni, G.; Caflisch, A. *Bioinformatics* **2007**, *23*, 2625–2627.
- (48) Seeber, M.; Felling, A.; Raimondi, F.; Muff, S.; Friedman, R.; Rao, F.; Caflisch, A.; Fanelli, F. *J. Comput. Chem.* **2011**, *32*, 1183–1194.