JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

Cite This: J. Chem. Inf. Model. 2020, 60, 152−161

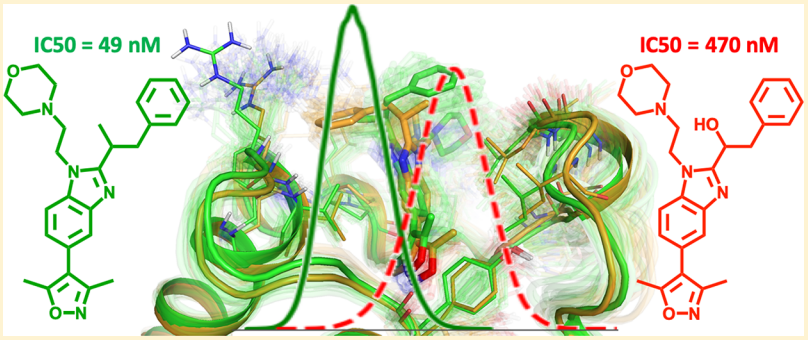
pubs.acs.org/jcim

Off-Pocket Activity Cliffs: A Puzzling Facet of Molecular Recognition

Tigran M. Abramyan, Yi An, and Dmitri Kireev*

Center for Integrative Chemical Biology and Drug Discovery, Division of Chemical Biology and Medicinal Chemistry, Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, 27599-7363

Supporting Information



ABSTRACT: While accurate quantitative prediction of ligand−protein binding affinity remains an elusive goal, high-affinity ligands to therapeutic targets are being designed through heuristic optimization of ligand−protein contacts. However, herein, through large-scale data mining and analyses, we demonstrate that a ligand's binding can also be strongly affected through modifying its solvent-exposed portion that does not make contacts with the protein, thus resulting in "off-pocket activity cliffs" (**OAC**). We then exposed the roots of the OAC phenomenon by means of molecular dynamics (**MD**) simulations and MD data analyses. We expect OAC to extend our knowledge of molecular recognition and enhance the drug designer's toolkit.

INTRODUCTION

Structure−activity relationships (SAR)[1] and structure-based design (SBD)[2,3] are the yin and yang of modern medicinal chemistry. The former leverages an intuitive concept of chemical similarity to prioritize modifications needed to obtain more active compounds based on the available structure−activity data. While accurate prediction of the ligand−protein binding affinity remains unattainable, SBD relies upon a heuristic toolkit of molecular recognition[4−6] to propose ligand modifications intended to optimize ligand−protein contacts. The symbiotic SAR/SBD association allows each of the two approaches to mitigate imperfections of the other. In particular, one of the key principles underlying SAR—"similar compounds tend to have similar biological activities"—is subject to frequent disruptions called "activity cliffs".[7−10] In general, the activity cliffs can be rationalized in terms of protein−ligand recognition: even a small ligand modification may induce a steric clash or disrupt a hydrogen bond. In turn, SBD lacks a full grasp of entropic effects in ligand-protein binding. A typical *ad hoc* SBD strategy deals with "tangible", enthalpic interactions, such as hydrogen or halogen bonds, aromatic stacking, cation-π, and others, while amorphous entropic component of binding is often left aside. In SAR, the entropy problem is implicitly addressed−the similarity principle extends to the ligand's entropic behavior just like to any other property. On the whole, SBD and SAR constitute mutually complementing components of the drug designer's toolkit: SAR implicitly account for the entropy, while SBD rationalizes ligand−protein interactions and activity cliffs.

Recently, though, we have serendipitously observed a yet understudied effect, a class of activity cliffs that do not lend themselves to an intuitive structural interpretation. The intriguing aspect of these cliffs is that they occur as a result of off-pocket ligand modifications (OLM), that is, changes in solvent-exposed parts of the ligand which are not in contact with the protein. We came across OAC while working on the development of inhibitors targeting the Tyro3, Axl, and Mer (TAM) family of receptor tyrosine kinases.[11−13] A series of compounds was synthesized to investigate structural variation at a site pointing to the solvent.[13] In general, such modifications are intended to improve the pharmacokinetic profile of a compound without affecting its inhibitory potency. To our surprise, some of these modifications have resulted in a significant potency variation. As can be seen in Figure 1 (ligand pair **1**/**1'**), addition of a solvent-exposed *N*-cyclohexylmethylamine group at the 3-position of the pyridine ring resulted in a 30-fold improvement in IC_{50} (measured in an *in vitro* Axl microfluidic capillary electrophoresis assay[13]). We were intrigued by this off-pocket effect and sought to confirm its authenticity, determine possible magnitudes, and to elucidate its origin. Understanding physical principles underlying OAC would extend our general knowledge of molecular

Received: August 31, 2019
Published: December 2, 2019

DOI: 10.1021/acs.jcim.9b00731
J. Chem. Inf. Model. 2020, 60, 152−161

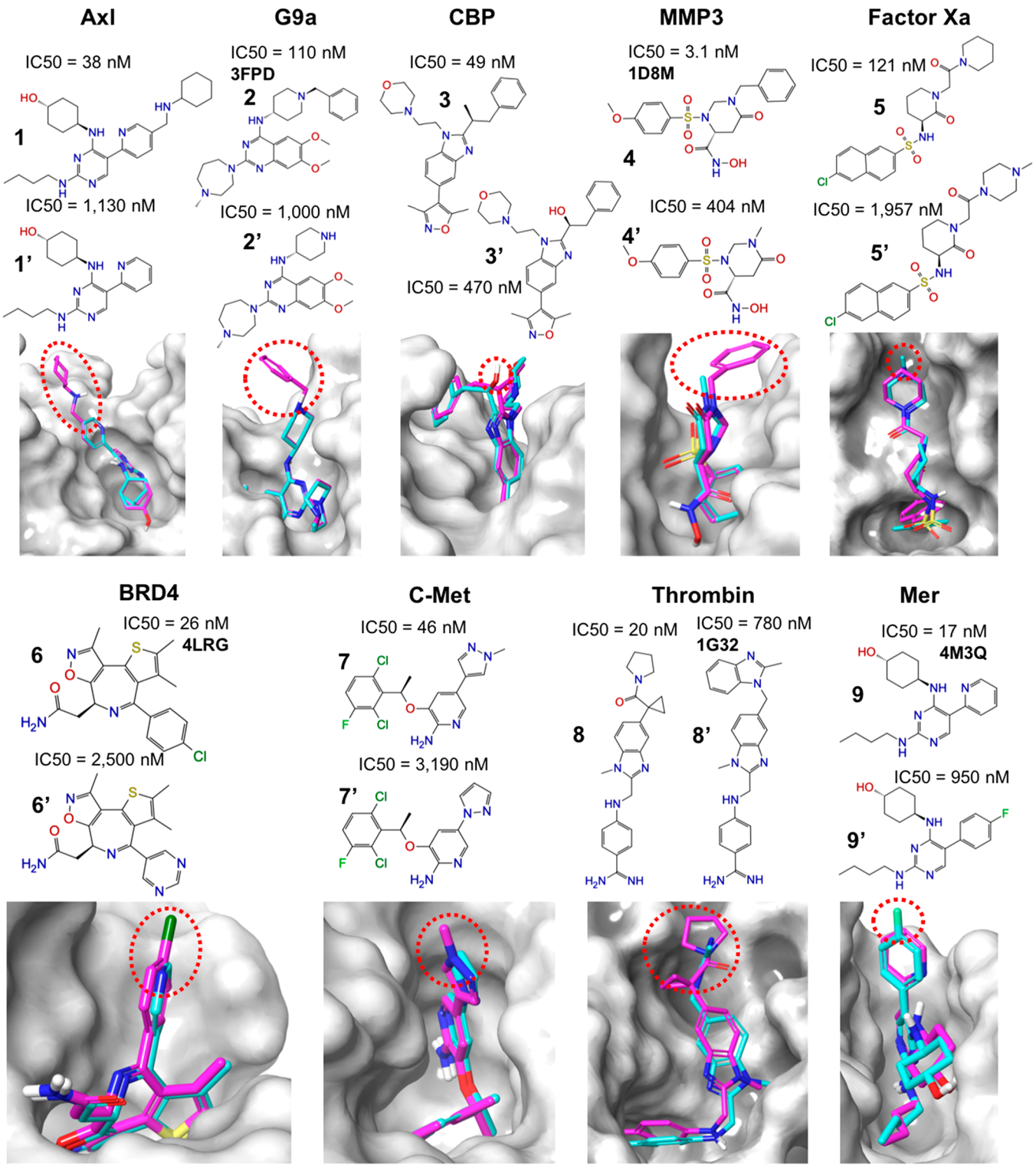


Figure 1. Off-pocket activity cliffs analyzed in this study. The data set of OAC collected for this study; the active ligand of each active/inactive pair is rendered as magenta sticks; the ligand modification regions in 3D models are encircled by a red dotted line; activities are expressed as half-maximal inhibitory concentrations (IC50). The binding modes of five ligands—**2**, **4**, **6**, **8′**, and **9**—with their respective targets, were known from the crystal structures (PDB ID: 3FPD, 1D8M, 4LRG, 1G32, and 4M3Q, respectively). The docking simulations reproduced the crystal structure binding modes of these five ligands with an RMSD < 1 Å. The coordinates of the remaining complexes shown here were obtained by aligning and minimizing the molecule on a ligand from the available crystal structure (see Methods for the remaining PDB codes); these poses are additionally supported by docking studies.

recognition. Moreover, mastering OLM can be a useful addition to the drug designer's toolkit. Since the analysis of static structures of ligand−protein complexes, and the respective interaction signatures, would not allow one to discriminate between the active and inactive compounds in an OAC pair, we employed molecular dynamics (MD) simu-

DOI: 10.1021/acs.jcim.9b00731
J. Chem. Inf. Model. 2020, 60, 152−161

lations to identify structural factors that would enable such a distinction.

RESULTS

OAC Implicate Chemically Diverse Ligands and Structurally Distinct Proteins. The first question we sought to address was whether the OAC phenomenon is real. To this end, we mined large public structural and structure−activity databases to find as many OAC occurrences as possible. BindingDB[14] was chosen as the data source since its entries contain all necessary references: ligand structure, activity/affinity data for the target protein, and PDB ID for the respective 3D structure of the ligand-protein complex. Conceptually, mining for OAC involved the following steps: (i) identify 3D structures of ligand−protein complexes with ligands having sizable solvent-exposed portions, (ii) for each ligand from the previous step, find at least one similar compound with significantly different affinity to the same protein, (iii) align the analogs from the previous step, and (iv) select only analogs differing from the initial queries only in their solvent-exposed portions. Although intuitively obvious, some of these tasks are difficult to automate and need significant human input. Hence, the whole process needs to be organized in such a way that the most automated tasks are executed first to prefilter as much data as possible, limiting the human intervention to a relatively small amount of data "in the bottom of the funnel".

In the first step of the mining process, we retrieved ligands satisfying the following criteria: (i) availability of an experimentally determined 3D structure in complex with a protein, (ii) significant potency/affinity with respect to the cocrystallized protein, expressed as K_i, K_d, IC_{50}, or EC_{50} (a threshold of 150 nM was used to define "significant"), and (iii) satisfying drug-likeness criteria (including a modified Lipinski[15] and structural[16] filters). This step resulted in a selection of 2311 ligand−protein complexes involving 1803 unique ligands and 480 proteins. Next, a semiautomated procedure was applied to detect ligands having both solvent-exposed and sizable buried parts. To this end, a ratio of the solvent-accessible surface area (SASA) of the protein-bound ligand to the SASA of free ligand was calculated. Ratios between 0.20 to 0.55 were considered corresponding to a partially buried, partially solvent-exposed ligand. Application of this filter resulted in 769 complexes (582 unique ligands and 239 proteins). The third step was mining the ChEMBL database[17] for ligands chemically similar to those selected in step two and having potency/affinity values available for the protein of interest. A total of 310 pairs of similar ligands for 187 unique proteins were selected at this step. The last step was performed essentially by visual inspection. Every active/inactive pair was examined on whether their activity data were compatible and could be used as proxies for the affinity and whether the differing fragments of the two ligands were exposed to solvent. For most pairs, original publications were used to pick an active/inactive pair with the largest possible difference in activity and with structural differences that are definitely solvent-exposed. We were able to collect a set of 9 OACs with a high degree of diversity on both the ligand and the protein side (Figure 1). The protein targets include bromodomains of BRD4[18] and CBP,[19] protein kinase domains of c-Met,[20] Axl,[13] and MerTK,[13] proteinases Thrombin[21] and Factor Xa,[22] a metalloproteinase MMP3,[23] and a methyl transferase G9a.[24] For the sake of brevity, here and throughout the text, the less active compound in an OAC pair will be referred to as the inactive one, although they all show some degree of activity and are certainly able to bind to the protein of interest.

The diversity of ligands and proteins involved in OAC so far identified emphasizes a universal character of the off-pocket effect, while the activity data show the significant impact of the off-pocket modification on the binding affinity. Indeed, the ratios of dissociation constants (K_d) for respectively active and inactive compound pairs range from 9 (for G9a inhibitors) to 130 (for MMP3 inhibitors). Here, according to the Cheng−Prusoff equation,[25] the ratio of IC_{50}s was considered equal to the ratio of K_ds. One notable uncertainty in our OAC data is that experimentally determined 3D structures (in complex with the protein of interest) are not available for at least one ligand in each OAC pair. That is, we cannot be certain whether the in-pocket portion of one ligand in an OAC pair would still remain in-pocket after the ligand modification. In a recent study, Malhotra and Karanicolas[26] outlined major factors that may cause a ligand to significantly change its pose upon a chemical modification. In particular, we tested two important factors inducing a pose switch:[26] (i) the modified ligand may not be able to adopt a similar binding pose (either due to a steric clash or disruption of a critical interaction) and (ii) the modified ligand may adopt a different binding pose because it enables new, stronger interactions. Of these two factors, the first one can be safely ruled out since added substituents do not change the interacting portion at all. However, it is indeed possible that a modification may enable new, stronger interactions, thus forcing a ligand to flip. Accordingly, we performed additional unbiased studies using Glide docking algorithm[27] to assess the risk of a significant pose change. Up to three poses per ligand were generated using Glide. We then inspected whether the highest ranked in-pocket poses for both active and inactive ligands of an OAC pair are similar. It must be noted that crystal structures were available for five of the ligands studied: **2** (PDB ID: 3FPD), **4** (1D8M), **6** (4LRG), **8** (1G32), and **9** (4M3Q). For all other ligands binding modes were determined by Glide docking. To corroborate docking accuracy, available crystal structures of **2**, **4**, **6**, **8**, and **9** were compared to docking poses and showed a remarkable agreement (with RMSD < 1 Å). Encouragingly, we also observed that in all OAC pairs the orientation of the in-pocket ligand portions for the active and inactive ligands were virtually the same (Supplementary Files ci9b00731_si_002.zip and ci9b00731_si_003.xlsx). In two complexes, **3**:CBP and **7**':c-Met, we observed an inversion of highly symmetric groups (respectively, isoxazole and dichloro-fluoro-benzene). However, these divergences are within the experimental resolution. From a qualitative viewpoint, most of the nine proteins have small ligand-binding pockets whose shapes are complementary to those of the in-pocket ligand cores. Therefore, on the whole, while experimental determination of the missing structures is beyond our reach, there is a solid rationale to assume that both ligands of an OAC pair share the same in-pocket portion.

Dynamics−Activity Relationships. Putatively, a variety of physical mechanisms may cause the off-pocket effect: ligand−solvent interactions, intermittent ligand−protein interactions, conformational straining, or combinations thereof. However, ultimately, all the off-pocket effects should somehow affect the in-pocket dynamics, thus stabilizing or destabilizing the persistent, localized in-pocket interactions. Hence, we sought to investigate whether, and how strongly, the dynamics of protein−ligand complexes are affected by off-pocket ligand

DOI: 10.1021/acs.jcim.9b00731
J. Chem. Inf. Model. 2020, 60, 152−161

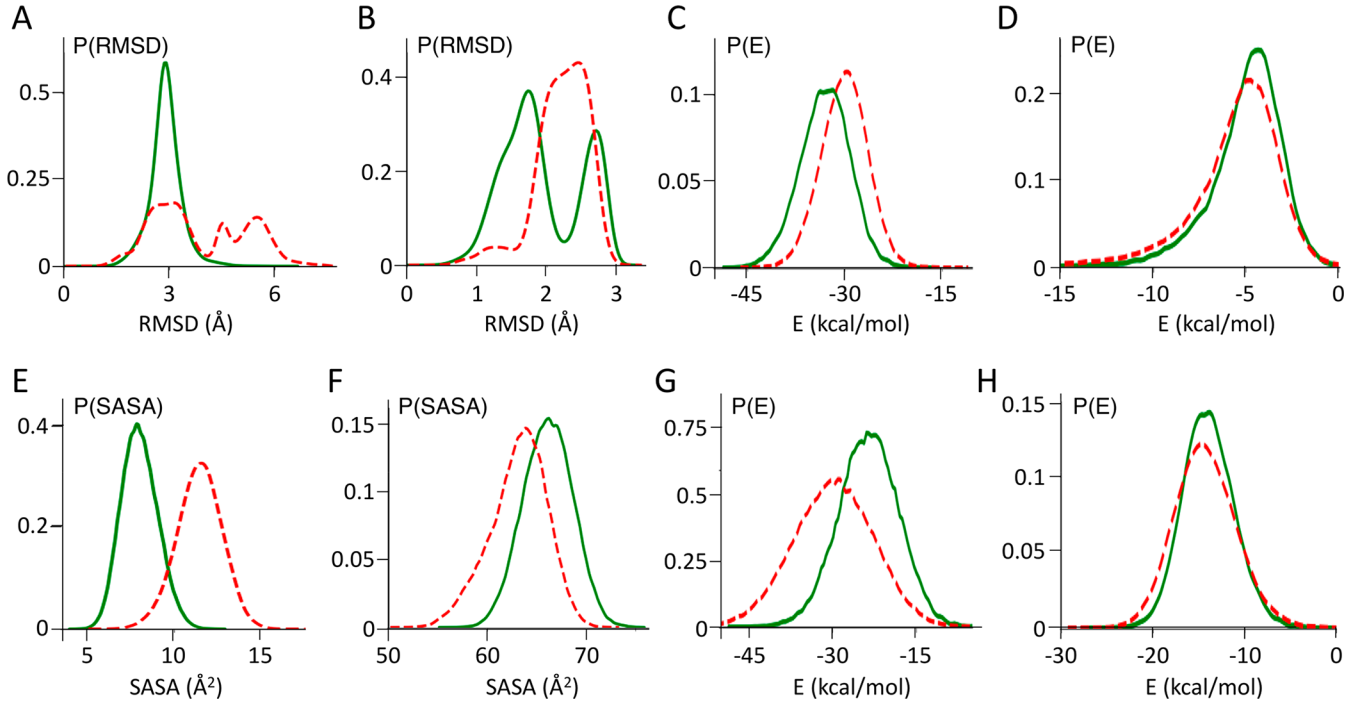


Figure 2. Structure−dynamics relationships for the p300-bromo OAC pair (ligands 3/3′). Probability density functions for (A) ligand mobility in the binding pocket expressed as RMSD with respect to the native pose in the protein coordinates; (B) conformational ligand mobility expressed as RMSD with respect to the native pose in the ligand coordinates; (C) protein−ligand vdW interaction energy; (D) protein−ligand electrostatic interaction energy; (E) the ligand's lipophilic SASA; (F) the ligand's polar SASA; (G) water−ligand electrostatic interaction energy; and (H) water−ligand vdW interaction energy. In all panels of the figure, the solid green line corresponds to the active molecule and the red dashed line—to the inactive.

Table 1. Normalized Overlaps of PDFs of MD Metrics for OAC Compound Pairs 1−9[a]

	MMP3	G9a	CBP	Axl	FXa	BRD4	c-Met	Thrombin	Mer
$RMSD_{ip}$	0.32*	0.27*	0.33*	0.64	0.67	0.27*	0.76	0.41*	0.45*
$RMSD_{conf}$	0.42*	0.44*	0.33*	0.23*	0.32*	0.28*	0.71	0.26*	0.31*
$E_{el}^{lig\text{-}prot}$	0.73	0.46*	0.78	0.80	0.65	0.42*	0.81	0.68	0.64
$E_{vdw}^{lig\text{-}prot}$	0.51	0.12*	0.86	0.61	0.79	0.39*	0.51	0.68	0.53
$SASA_{lip}$	0.00*	0.00*	0.65	0.00*	0.90	0.00*	0.58	0.49*	0.65
$SASA_{pol}$	0.27*	0.10*	0.08*	0.28*	0.26*	0.00*	0.16*	0.68	0.78
$E_{el}^{lig\text{-}water}$	0.69	0.62	0.45*	0.36*	0.56	0.40*	0.18*	0.60	0.91
$E_{vdW}^{lig\text{-}water}$	0.31*	0.60	0.79	0.07*	0.81	0.42*	0.73	0.55	0.77

[a]Asterisks highlight properties enabling a PDF overlap of 0.5 or less.

modifications. Multiple metrics expressing either the ligand's conformational diffusion or its interactions with the protein or its exposure to the solvent were calculated (Figure 2, Supplementary Figures S1−S5). The calculated properties included: (i) conformational ligand mobility and ligand mobility with respect to the protein pocket calculated as the root-mean-square deviation (RMSD) of the ligand from its native pose where, respectively, ligand and protein Cα atoms were used to align the frames, (ii) polar and lipophilic solvent-accessible surface areas ($SASA_{pol}$ and $SASA_{lip}$) of the bound ligand, and (iii) ligand−protein and ligand−solvent interaction energies, van der Waals ($E_{vdw}^{lig\text{-}prot}$), and electrostatic ($E_{el}^{lig\text{-}prot}$) terms.

Probability density functions (PDFs) for these properties were used to collectively characterize the MD trajectories of the protein−ligand complexes. Finally, a normalized overlap between the PDFs of an OAC pair was used as a measure of the discriminative power of a metric. The normalized overlap was calculated as $(\int \min(F(L), F(L'))\, dx)/(\int \max(F(L), F(L'))dx)$, where $F(L)$ is a PDF for a ligand L and x is a given structure-based metric. The overlap has values between 1 (fully overlapping functions) and 0 (no overlap). To ensure an adequate sampling of the ligand/pocket configuration space, four independent runs of unbiased MD (up to 1 μs/each) were performed in the NVT ensemble for each of the 18 ligand−protein complexes. Here and below, the OAC pair of CBP ligands **3** and **3′** (the most remarkable of OACs **1**−**9** in terms of the structure change) is used as a showcase of the essential dynamics−activity relationships for the off-pocket effect. Data and figures for the other eight OACs are available as supporting information (Supplementary Figure S1−S5).

As can be seen in Figure 2A−H, most of the investigated metrics proved to be discriminative for the CBP OAC pair, with the $SASA_{lip}$ PDFs showing the smallest **3**/**3′** overlap (of 0.08), and the ligand−water vdW interaction energy (E_{vdw}) PDFs show the largest overlap of 0.79. As a whole, active/inactive ligands in all OAC pairs are clearly discriminated by at least two metrics (Supplementary Figures S1−S5). Taking a

DOI: 10.1021/acs.jcim.9b00731
J. Chem. Inf. Model. 2020, 60, 152−161

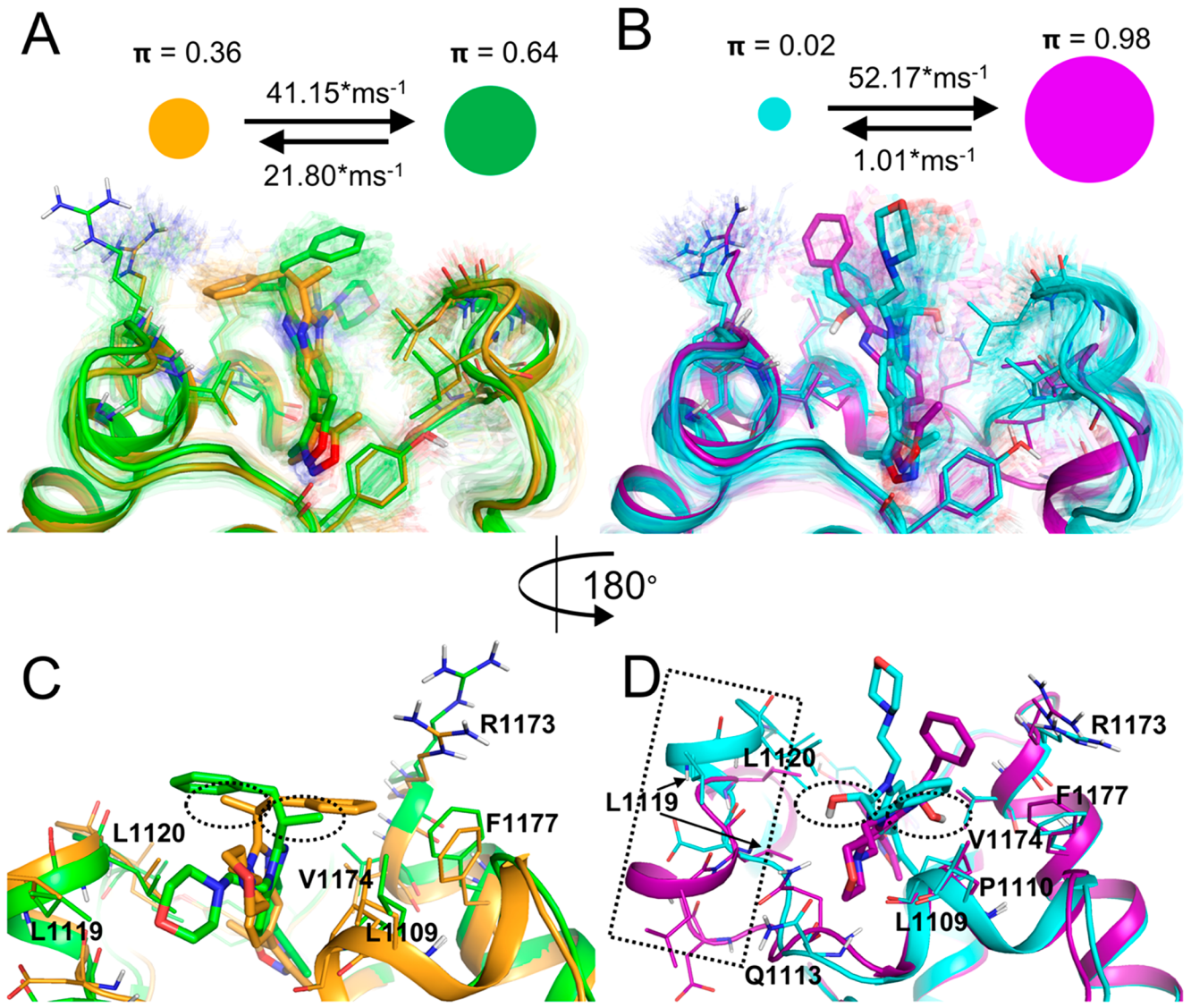


Figure 3. Markov state models for the CBP−bromodomain OAC pair (ligands 3/3′). The values from the maximum likelihood Bayesian Markov model and the transition rates between metastable states of the ligand−protein complexes for active (A) and inactive (B) ligands are shown. For each MS, the centroid and 100 representative states (transparent) with >70% probability of belonging to their respective state are displayed. Additionally, when turned by 180°, the centroid structures allow one to more clearly see the differences in orientations of the off-pocket methyl and hydroxyl groups (encircled by dashed lines) in each metastable state for respectively active (C) and inactive (D) ligands. The dashed rectangle highlights the conformational rearrangement in 1116−1121 helix of the protein in complex with the inactive.

PDF overlap smaller than 0.5 as a measure of successful discrimination (Table 1), conformational RMSD was the best PDF discriminator (eight successful discriminations) closely followed by $SASA_{pol}$ and in-pocket RMSD, with respectively 7 and 6 successful discriminations. Interestingly, $SASA_{lip}$ showed the sharpest discrimination, with near zero PDF overlap, for four OAC pairs, but failed to discriminate all other pairs (Table 1). The energy-based metrics showed the poorest performance, with 2 to 4 discernible discriminations per metric. On the ligand side, the BRD4 ligand pair (**4**/**4′**) was the "best" discriminated pair whose dynamics were successfully discriminated by all descriptors, while the Factor Xa, c-Met, and Mer OAC pairs were the most "resistant" to discrimination in our analysis (with only two discriminating metrics per pair). Remarkably, the CBP OAC pair was discriminated fairly well (by four metrics) despite the highest possible active to inactive ligand similarity (identical shape, one small-moiety difference).

While multiple MD-based metrics reveal significant dissimilarities between the dynamics of ligand−protein complexes within OAC pairs, there is not necessarily a correlation between the active-to-inactive activity change and the character of the PDF discrimination, that is, the direction and magnitude of the active's PDF shift with respect to the inactive's PDF. For example, all four PDF pairs discriminated by $SASA_{lip}$ (Supplementary Figure S4) have bell-shaped curves corresponding to simulations of complexes with active ligands clearly shifted toward the left side of the chart. On the other hand, as can be seen from the $SASA_{pol}$ PDF charts (Supplementary Figure S4), of seven discriminated OAC pair PDFs, five actives' curves are shifted leftward and two rightward. Finally, most RMSD-based PDFs show complex multimodal profiles where the curve discrimination cannot be interpreted in simple terms of a leftward or a rightward shift.

Table 1 also provides a hint on the nature of the OAC effect. It clearly shows that while the ligand−protein interaction

DOI: 10.1021/acs.jcim.9b00731
J. Chem. Inf. Model. 2020, 60, 152−161

Table 2. Mean Fraction of Long-Residence Water Molecules near the Ligand for the Nine OAC Pairs[a]

	MMP3	G9a	CBP	Axl	FXa	BRD4	c-Met	Axl	Thrombin	Mer
$\langle p(t_r > 80)\rangle_t^A$	0.38	0.42	0.33	0.40	0.47	0.36	0.49	0.40	0.51	0.45
$\langle p(t_r > 80)\rangle_t^I$	0.43	0.34	0.37	0.43	0.47	0.31	0.51	0.43	0.54	0.41

[a]Superscripts A and I denote respectively active and inactive ligands of an OAC pair.

energies for a single in-pocket configuration would be near-identical within an OAC pair of ligands (because the interaction energy is dominated by close, in-pocket interactions), properties calculated from large ensembles sampled from the configuration space by MD simulations may be significantly different.

Metastable States of Ligand−Protein Complexes. To find out how the in-pocket interactions are affected by solvent-exposed ligand modifications, we explored the differences between the free energy landscapes (FEL) of ligand−protein interactions for a pair of OAC compounds. Markov state models (MSM)[28,29] provide a powerful means for reconstructing converged discretized FELs as sets of metastable states with transition rates between them. Again, the CBP OAC pair was selected as a case for the MSM study. A total of 38 all-atom MD simulations per ligand−protein system were run, accounting for a cumulative time of >20 μs/system. Initially, four trajectories (∼1 μs each) per system were generated. In the course of the further analysis, 34 more MD trajectories were produced starting from configurations randomly picked from the initial simulations. Structural descriptors of the system for MSM included contacts between ligand atoms and protein residues lining the ligand-binding pocket (see Methods for details). Eventually, a self-consistent set of metastable states was obtained by discretization of the time-lagged independent components (TICs) calculated from the system's structural descriptors.

We found that overall the active (**3**) and inactive (**3′**) CBP-bromodomain ligands significantly differ from each other both in the binding modes corresponding to the respective metastable states, as well as in time scales of the interstate transitions. No significant overlap was found between the dominant binding modes of the two ligands and contacts made with the binding site residues. As can be seen in Figure 3, the active ligand oscillates between two metastable states with significant equilibrium probabilities of 0.36 and 0.64 (for respectively "orange" and "green" states), while the inactive ligand engages the target in one dominant state that is structurally distinct from both "orange" and "green" states.

Intuitively, the solvent-exposed ligand's phenyl group is expected to form a stable cation-π interaction with the protein's R1173. However, only in less frequently visited "orange" state the active molecule features this interaction. The phenyl group, as well as the adjacent methyl group (the off-pocket modification responsible for a higher ligand's affinity; Figure 3A, C), are very mobile and are scanning the nearby hydrophobic side chains (L1119, L1120, L1174, and F1177) to protect themselves from exposure to bulk water. This behavior results in a counterintuitive, more frequently visited "green" state with the phenyl group oriented toward the solvent, in the opposite direction from R1173. On the whole, the dynamics of the protein-bound active ligand tends to minimize the free energy of hydration for its solvent-exposed portion. For the same reason, the ligand would be primed for keeping its oxazol-benzimidazole core tightly bound to its pocket. The inactive ligand displays significantly different dynamics. First, its solvent-exposed phenyl ring does form a stable cation-π interaction with R1173. This leaves the adjacent hydroxy group (responsible for its reduced activity; Figure 3B, D) permanently exposed to bulk water. Consequently, the in-pocket core is not entropically penalized for leaving the pocket, which is a likely explanation of the inactive's reduced affinity. The latter point is illustrated by the observation that the inactive's oxazol-benzimidazole core is less stable in the pocket compared to the active's core. In several trajectories, it flipped around its longest axis and even left the pocket completely. All the above supports the idea that the ligands' solvent-exposed portions may propagate their dynamics into the pocket, thus affecting the respective ligand−protein affinities.

Ligand−Solvent Interactions. The analysis of water molecules, their clusters, mobility, and interactions with the ligand may help to better understand likely causes of OAC. In particular, we investigated residence times of water molecules near the ligand as a proxy metric to the ligand's propensity for inducing stable water structures. The rationale for using such a proxy is that, if a particular water molecule resides for a long time near the ligand, while not directly interacting with the ligand, it is likely to be involved in a stable water cluster. Such ligand-bound clusters might significantly affect its dynamics. Here, for every water molecule that encounters the ligand at least once our algorithm calculated its residence time (t_r) on the ligand (a 6 Å distance threshold was used as a proximity criterion, to account for water molecules from the second hydration layer). Additionally, an average distance between a water molecule and the ligand during its residence on the ligand was calculated. These data provide rich, though indirect, evidence about the water structure. One way to express this evidence is to plot PDF for residence times observed (Supplementary Figure S6). On the whole, these PDFs for different OAC pairs look quite alike. Short stays (of less than 40 ps) are prevalent, representing ∼80% of stays. On the other hand, all PDFs have long tails with the longest stays of up to hundreds of nanoseconds. For four active/inactive compound pairs (Axl, c-Met, MMP3, and Thrombin), the t_r PDFs show significant differences; four other PDF pairs (BRD4, Factor Xa, G9a, and CBP) show smaller, but discernible differences; and, for one pair (MerTK), the respective PDFs are indistinguishable. A complementary way to assess the ligand's propensity to coordinate the nearby water molecules would be to calculate the fraction of long-residence water molecules around the ligand at any moment in time. For instance, Table 2 shows the values of mean fraction of water molecules residing near the ligand for longer than 80 ps ($\langle p(t_r > 80)\rangle_t$), where the fraction is averaged over all MD trajectory frames. This metric shows notable active/inactive differences for most OAC compound pairs (the only unresolved OAC pair is that of Factor X ligands). Similar to the metrics analyzed in previous sections, the sign of the $\langle p(t_r > 80)\rangle_t$ change does not correlate with the sign of the activity change.

DOI: 10.1021/acs.jcim.9b00731
J. Chem. Inf. Model. 2020, 60, 152−161

DISCUSSION

The primary focus of this study was to reveal the existence and physical signatures of the OAC effect. At the phenomenological level, we indeed identified a number of dynamic and structural features that are able to clearly distinguish between the active and inactive ligands of each OAC pair. At first sight, our findings may seem to challenge the previously formulated molecular recognition hypotheses: Fischer's "lock-and-key" concept,[30] Koshland's "induced fit" hypothesis,[31] and the "conformational selection" model.[32] Yet, the existence of OAC is rather a reminder that ligand−protein binding is an intricate process controlled by subtle balance between interaction energies and entropies of the three players involved (the ligand, the protein, and the solvent) which cannot always be represented by an intuitive mechanistic model. It is also possible to outline basic physical principles underlying the effect. For instance, in perturbation-theory terms,[33] the Helmholtz binding free energy change ($\Delta\Delta F_{\mathrm{bind}}$) on transforming the active compound of an OAC pair into the inactive one can be expressed as

$$\Delta\Delta F_{\mathrm{bind}} = \Delta\Delta F_t^{PL} - \Delta\Delta F_t^{L} \tag{1}$$

where $\Delta\Delta F_t^{PL}$ is the free energy change due to active-to-inactive transform in the binding site and $\Delta\Delta F_t^{L}$ is the free energy change due to active-to-inactive transform in solvent. Then, the first term of eq 1 can be written as

$$\Delta\Delta F_t^{PL} = \Delta\Delta F_t^{PL_{\mathrm{in}}} + \Delta\Delta F_t^{PL_{\mathrm{out}}} \tag{2}$$

where $\Delta\Delta F_t^{PL_{\mathrm{in}}}$ is the contribution to $\Delta\Delta F_t^{PL}$ from the in-pocket portion of the ligand and $\Delta\Delta F_t^{PL_{\mathrm{out}}}$ is the contribution from the outer, solvent-exposed portion of the ligand. For each portion, the potential energy, in addition to the regular force field terms, would also include a term due to an instantaneous force applied by the other portion to the connection site

$$\begin{aligned}\Delta\Delta F_t^{PL_{\mathrm{in}}} &= -k_B T \log \\ &\quad \langle\langle e^{-\beta((U_{\mathrm{in}}^{PL}(x)+U_{F_{\mathrm{out}}}^{PL}(t))-(U_{\mathrm{in}}^{PL'}(x)+U_{F_{\mathrm{out}}}^{PL'}(t)))}\rangle_{\mathrm{in}}\rangle_{F_{\mathrm{out}},F'_{\mathrm{out}}} \\ &= -k_B T \log \langle e^{-\beta(U_{F_{\mathrm{out}}}(t)-U_{F'_{\mathrm{out}}}(t))}\rangle_{F_{\mathrm{out}},F'_{\mathrm{out}}}\end{aligned} \tag{3}$$

where $\beta = \frac{1}{k_B T}$, k_B, the Boltzmann constant, T, temperature, $(U_{\mathrm{in}}^{PL}(x)$ and $U_{\mathrm{in}}^{PL'}(x)$ are the potential energies for in-pocket portions of L and L', $U_{F_{\mathrm{out}}}^{PL}(t)$ and $U_{F_{\mathrm{out}}}^{PL'}(t)$ are stochastic processes corresponding to instantaneous external forces drawn from the distributions $P^{L}(\overline{F_{\mathrm{out}}})$ and $P^{L'}(\overline{F_{\mathrm{out}}})$, and the angle brackets denote an equilibrium average over the accessible microstates inside the pocket and over the above force distributions. Since for any in-pocket configuration $U_{\mathrm{in}}^{PL}(x) = U_{\mathrm{in}}^{PL'}(x)$, then $\Delta\Delta F_t^{PL_{\mathrm{in}}}$, that is, the determining component of $\Delta\Delta F_{\mathrm{bind}}$ depends solely on the external forces applied by the solvent-exposed ligand portion to the in-pocket one. Such an outcome is consistent with our dynamics-activity relationships where SASA descriptors showed the best discrimination between the active's and inactive's dynamics. Additionally, in the MSM study, differing metastable states of the solvent-exposed ligand portions were causing differences in in-pocket poses.

While on the whole we made a significant step toward understanding the OAC effect and its physical origins, this general understanding would not be easy to directly translate into practical drug design, that is, to predict specific off-pocket modifications that may improve the ligand's affinity. Indeed, a myriad of various factors can change a $P^{L}(\overline{F_{\mathrm{out}}})$ distribution, often in unpredictable ways. For instance, the solvent-exposed parts in an OAC pair of ligands may differ in size, their capacities to coordinate nearby water molecules, and interactions with the intrapocket ligand core. Larger solvent-exposed portions, or those that coordinate larger numbers of water molecules, would experience higher water friction that would slow down their diffusion in the accessible configuration space, hence reducing the force magnitudes. It is also possible that same extra-pocket entity would have different effects on different intrapocket cores. Moreover, entropic effects resulting from off-pocket modifications do lend themselves to an intuitive visual representation as is the case for currently known protein−ligand interactions[4] including the entropy-based hydrophobic effect.[6] The latter, for example, can be readily wrapped into a designer's rule of thumb such as "fill a hydrophobic protein pocket with a hydrophobic ligand fragment". For off-pocket ligand modifications, such visual link between the nature of a modification and its effect on activity may not exist at all or may depend on a variety of factors.

Nevertheless, off-pocket ligand modifications constitute a very promising addition to the current drug designer's toolkit. It may be particularly useful in achieving a higher affinity or selectivity to a protein with a small or relatively shallow binding pocket, where the potential of intrapocket modifications is limited (e.g., CBP bromodomain,[19] compounds **3** and **3′**). Machine learning techniques might be used to identify the significance of every individual factor and establish quantitative structure−activity relationships (QSAR),[34,35] or quantitative dynamics-activity relationships, specific to OAC. However, such exploratory data analyses would require a significantly larger OAC data set than the one collected in this study. In the meantime, off-pocket ligand modifications can be guided by local affinity-dynamics relationships similar to those we described. This would make a strong case for more widespread use of MD simulations in practical drug design.

CONCLUSIONS

In summary, we reported a yet understudied aspect of molecular recognition where a ligand's binding can be strongly affected through off-pocket modifications where modified sites are solvent-exposed and do not make any contacts with the protein. We showed that the off-pocket effect arises from the differing dynamics of the solvent-exposed ligand portions. We examined how ligand−solvent interactions affect the ligand's dynamics inside the binding pocket thus altering the free-energy landscape of the ligand−protein contacts. This off-pocket effect extends our knowledge of molecular recognition and enhances the current drug discovery toolkit.

METHODS

Data Mining. The data mining process was automated using the Pipeline Pilot (v.18.1.100.11) environment and scripting language (3dsbiovia.com). The BindingDB database[14] (dated by 1/5/2018) was used as a source of ligand structures with associated affinities and PDB codes for respective protein−ligand complexes. The 3D structures of ligand-protein complexes for structural analysis were automatically downloaded from the Protein Data Bank[36] in PDB format

DOI: 10.1021/acs.jcim.9b00731
J. Chem. Inf. Model. 2020, 60, 152−161

using Pipeline Pilot. The active ligands for the initial analyses were defined as ligands having K_i, K_d, IC_{50}, or EC_{50} < 150 nM. Based on this definition, 1803 unique ligands were selected against 480 unique proteins (for a total of 2311 ligand−protein complexes with 3D structures). The ratio of SASA of the protein-bound ligand to SASA of the unbound ligand was used to identify ligands only partially buried within the binding pockets. Ligand−protein complexes with the above ratio of 0.20−0.55 (769 complexes, 582 unique ligands, 239 proteins) were selected as potentially implicated in off-pocket activity cliffs (OAC). Similarity search in the ChEMBL database[17] was then used to identify inactive ligands similar to the previously identified actives. Eventually, 3D structures of ligand-protein complexes were visually inspected using the PyMOL Molecular Graphics System v. 1.8.6.2 (Schrödinger, LLC).

Docking. The 3D structures of the ligand-protein complexes with the following PDB codes were used as the starting structures for the molecular docking simulations: 4NR7 (CBP bromodomain), 4LRG (BRD4 bromodomain), 2WGJ (c-Met kinase domain), 4M3Q (MerTK kinase domain), 3SW2 (Factor Xa), 1G32 (Thrombin), 3FPD (G9a HMT catalytic unit), 1D8M (MMP3). The 3D structure of the Axl kinase domain was homology-modeled using the MerTK structure (PDB: 4M3Q) as a template in Maestro modeling suite (release 2016-2; Schrödinger, LLC). The binding modes for the G9a:**2** (3FPD), MMP3:**4** (1D8M), BRD4:**6** (4LRG), Thrombin:**8** (1G32), Mer:**9** (4M3Q) were known from the respective crystal structures. Protein Preparation Wizard[37] available through Maestro (release 2018-4; Schrödinger, LLC) was used to prepare the complexes. In addition to the default settings, where necessary the missing side chains and missing loops were added with Prime. Water molecules known from the respective crystal structures to mediate ligand binding (e.g., CBP bromodomain) were kept for further analysis. Hydrogen bond assignment was optimized using PROPKA at pH 7.0. To avoid unnatural clashes between atoms, restrained minimization with heavy atoms convergence at RMSD 0.3 Å was performed. The active and inactive molecules with correct stereochemistry, according to each study, were drawn in 2D sketcher in Maestro. The Receptor Grid Generation module was used to generate the grids for docking with its center being the molecule available in the respective crystal structure. Glide standard precision[27] docking using default settings in Maestro was used for docking of the active and inactive molecules to the respective protein. The output coordinates of the top ranked Glide poses are provided in Supplementary files.

Molecular Dynamics Simulations. All molecular dynamics (MD) simulations were performed in Gromacs 2016.4 simulation package using CHARMM22 protein force field.[38] Similar to the molecular docking simulations the 3D structures of the ligand−protein complexes with the following PDB codes were used as starting structures for the MD simulations: 4NR7 (CBP bromodomain), 4LRG (BRD4 bromodomain), 2WGJ (c-Met kinase domain), 4M3Q (MerTK kinase domain), 3SW2 (Factor Xa), 1G32 (Thrombin), 3FPD (G9a HMT catalytic unit), and 1D8M (MMP3). The 3D structure of the Axl kinase domain was homology-modeled using the MerTK structure (PDB: 4M3Q) as a template in Maestro modeling suite (release 2016-2; Schrödinger, LLC). The 3D structures for the compounds, for which the ligand-protein crystal structure was unavailable, were obtained by aligning them on the aforementioned ligand−protein complexes in Maestro. The force field parameters for all ligands were generated using SwissParam.[39] The ligand−protein complexes were minimized in vacuum using steepest decent algorithm for 5000 steps or until the maximum force of 1000 $kJ{\cdot}mol^{-1}{\cdot}nm^{-1}$ was reached. The molecular systems were then solvated in TIP3P water,[40] counterions were added for system neutrality, and NaCl was added by replacing water molecules to mimic 0.15 M physiological conditions. Solvent minimization was then performed, followed by a two-step equilibration, during which all heavy atoms of the system, excluding those of water and counterions, were restrained: 0.1 ns in the NVT ensemble using the modified Berendsen thermostat[41] set at constant 300 K, and 1 ns in NPT ensemble at constant 1 atm and 300 K using the Parinello−Rahman pressure coupling.[42] All simulations were conducted using the Leapfrog integrator in periodic boundary conditions. The 12−6 Lennard-Jones potential was used to describe the vdW interactions, and the nonbonded cutoff distance was set at 0.1 nm. The long-range electrostatic interactions were calculated using the particle mesh Ewald method.[43] Bonds involving hydrogen atoms were constrained using the linear constraint solver algorithm (LINCS).[44] The production simulations were conducted in NVT ensemble with all atoms free to move. For each of the 18 individual ligand−protein complexes, 2−4 μs cumulative trajectories were obtained from four independent simulation seeds (more than 40 μs of total MD sampling). For the MD data analysis, all frames containing unbound ligands (RMSD > 7.5 Å) were removed. Gromacs trajectory analysis tools (gmx rmsd, gmx sasa, gmx energy, gmx hbond), MDTraj, as well as custom bash and python scripts were used for data analysis and matplotlib for plotting. Molecular visualization and generation of graphics were performed using PyMOL and VMD.[45] All MD trajectories are available upon request.

Markov State Models. MSM analyses[28,29] for ligands **3** and **3′** in complex with p300 bromodomain were performed using the PyEMMA program (version 2.5.4).[46] A total of 38 independent simulation trajectories were generated for each ligand−protein complex. In the first round of simulations, three 1-μs-long simulations were performed. Next, 33 more 200 ns simulations were run from configurations randomly selected from the first-round trajectories. An array of ligand−protein contacts (231 contacts for each molecular system) was used to describe a ligand pose in the protein's binding pocket and used as input to MSM. A "contact" was defined as a pair of atoms within 5 Å from each other. To facilitate comparative analyses of MSM results related to complexes with ligands **3** and **3′**, all trajectories were merged and processed in a joint feature space. This joined data set was then transformed into "kinetic" coordinates using the time-lagged independent component analysis (TICA).[47] The four slowest TICs representing ∼80% of the cumulative kinetic variance were selected for the next step of the analysis. The key initial steps in MSM building are space and time discretization. Space discretization, that is, generation of the systems' microstates, was performed using K-means clustering.[48] Multiple groupings (into 50, 200, 300, 500, and 1000 clusters) were generated and tested on their capacity to produce MSM with the best predictive power. It was shown that 200 and 300 state MSMs give the optimal performance for the molecular complexes with **3** and **3′**, respectively. First, such MSMs produce the highest mean score when the variational approach with cross-validation (score_cv method in PyEMMA) is applied (Table S1), using VAMP2 score method and 10-fold cross-validation at 80 ns lag time. Second, such MSMs showed better

DOI: 10.1021/acs.jcim.9b00731
J. Chem. Inf. Model. 2020, 60, 152−161

convergence of implied time scales in terms of the lag time (Supplementary Figure S7).

The combined clustered data set was then divided back into individual simulated systems (active and inactive), and Bayesian Markov models were estimated, in which statistical errors of 100 transition matrix samples with 95% confidence interval at each lag time were calculated. Implied relaxation time scales (ITSs) as a function of different lag times for the transition between these microstates were calculated and plotted (Supplementary Figure S7). The metastable states were validated using Chapman−Kolmogorov (CK) tests, which showed that the MSMs estimated at 80 ns were consistent with the simulation data with 95% confidence interval (Supplementary Figure S8). The PCCA+ algorithm[49] was applied to each system to merge the resulting clusters into metastable states (MSs). Further conformational and kinetic analyses were performed on the states that had more than 70% of likelihood of belonging to a particular MS.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.9b00731.

Figures S1−S8 and Table S1 as mentioned in the text (PDF)

Archive of rotatable (PDB format) OAC docked poses (ZIP)

Table of Glide scores (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: dmitri.kireev@unc.edu.

ORCID

Tigran M. Abramyan: 0000-0002-7224-6072

Dmitri Kireev: 0000-0001-8479-8555

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the University Cancer Research Fund (UCRF), a philanthropic gift to the CICBDD, and an award RX03712105 from the Eshelman Institute for Innovation. We acknowledge UNC Longleaf supercomputer cluster and their staff for support. We would also like to acknowledge useful discussions with Ara M. Abramyan (NIDA/NIH) on the MSM analysis. We thank Paul Brennan (SGC-Oxford) for sharing unpublished data on CBP bromodomain binding.

■ REFERENCES

(1) Selassie, C. D. History of Quantitative Structure-Activity Relationships. In *Burger's Medicinal Chemistry and Drug Discovery*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2003; pp 1−48.

(2) Jhoti, H.; Leach, A. R. *Structure-Based Drug Discovery*, 2nd ed.; Springer: Dordrecht, 2007.

(3) Nicholls, A. The Character of Molecular Modeling. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 103−105.

(4) Bissantz, C.; Kuhn, B.; Stahl, M. A Medicinal Chemist's Guide to Molecular Interactions. *J. Med. Chem.* **2010**, *53*, 5061−5084.

(5) Williams, D. H.; Stephens, E.; O'Brien, D. P.; Zhou, M. Understanding Noncovalent Interactions: Ligand Binding Energy and Catalytic Efficiency from Ligand-Induced Reductions in Motion within Receptors and Enzymes. *Angew. Chem., Int. Ed.* **2004**, *43*, 6596−6616.

(6) Zurcher, M.; Diederich, F. Structure-Based Drug Design: Exploring the Proper Filling of Apolar Pockets at Enzyme Active Sites. *J. Org. Chem.* **2008**, *73*, 4345−4361.

(7) Stumpfe, D.; Bajorath, J. Exploring Activity Cliffs in Medicinal Chemistry. *J. Med. Chem.* **2012**, *55*, 2932−2942.

(8) Maggiora, G. M. On Outliers and Activity Cliffs Why QSAR Often Disappoints. *J. Chem. Inf. Model.* **2006**, *46*, 1535.

(9) Bajorath, J. Representation and Identification of Activity Cliffs. *Expert Opin. Drug Discovery* **2017**, *12*, 879−883.

(10) Stumpfe, D.; Hu, Y.; Dimova, D.; Bajorath, J. Recent Progress in Understanding Activity Cliffs and Their Utility in Medicinal Chemistry. *J. Med. Chem.* **2014**, *57*, 18−28.

(11) Liu, J.; Yang, C.; Simpson, C.; Deryckere, D.; Van Deusen, A.; Miley, M. J. J.; Kireev, D.; Norris-Drouin, J.; Sather, S.; Hunter, D.; Korboukh, V. K. K.; Patel, H. S. S.; Janzen, W. P. P.; Machius, M.; Johnson, G. L. L.; Earp, H. S. S.; Graham, D. K. K.; Frye, S. V. V; Wang, X. Discovery of Novel Small Molecule Mer Kinase Inhibitors for the Treatment of Pediatric Acute Lymphoblastic Leukemia. *ACS Med. Chem. Lett.* **2012**, *3*, 129−134.

(12) Zhang, W.; DeRyckere, D.; Hunter, D.; Liu, J.; Stashko, M. A.; Minson, K. A.; Cummings, C. T.; Lee, M.; Glaros, T. G.; Newton, D. L.; Sather, S.; Zhang, D.; Kireev, D.; Janzen, W. P.; Earp, H. S.; Graham, D. K.; Frye, S. V.; Wang, X. UNC2025, a Potent and Orally Bioavailable MER/FLT3 Dual Inhibitor. *J. Med. Chem.* **2014**, *57*, 7031−7041.

(13) Zhang, W.; Zhang, D.; Stashko, M. A.; DeRyckere, D.; Hunter, D.; Kireev, D.; Miley, M. J.; Cummings, C.; Lee, M.; Norris-Drouin, J.; Stewart, W. M.; Sather, S.; Zhou, Y.; Kirkpatrick, G.; Machius, M.; Janzen, W. P.; Earp, H. S.; Graham, D. K.; Frye, S. V.; Wang, X. Pseudo-Cyclization through Intramolecular Hydrogen Bond Enables Discovery of Pyridine Substituted Pyrimidines as New Mer Kinase Inhibitors. *J. Med. Chem.* **2013**, *56*, 9683−9692.

(14) Liu, T.; Lin, Y.; Wen, X.; Jorissen, R. N.; Gilson, M. K. Binding DB: A Web-Accessible Database of Experimentally Determined Protein-Ligand Binding Affinities. *Nucleic Acids Res.* **2007**, *35*, D198−201.

(15) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Delivery Rev.* **1997**, *23*, 3−25.

(16) Walters, W. P.; Murcko, A. A.; Murcko, M. A. Recognizing Molecules with Drug-like Properties. *Curr. Opin. Chem. Biol.* **1999**, *3*, 384−387.

(17) Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P. ChEMBL: A Large-Scale Bioactivity Database for Drug Discovery. *Nucleic Acids Res.* **2012**, *40*, D1100−D1107.

(18) Gehling, V. S.; Hewitt, M. C.; Vaswani, R. G.; Leblanc, Y.; Côté, A.; Nasveschuk, C. G.; Taylor, A. M.; Harmange, J.-C.; Audia, J. E.; Pardo, E.; Joshi, S.; Sandy, P.; Mertz, J. A.; Sims, R. J.; Bergeron, L.; Bryant, B. M.; Bellon, S.; Poy, F.; Jayaram, H.; Sankaranarayanan, R.; Yellapantula, S.; Bangalore Srinivasamurthy, N.; Birudukota, S.; Albrecht, B. K. Discovery, Design, and Optimization of Isoxazole Azepine BET Inhibitors. *ACS Med. Chem. Lett.* **2013**, *4*, 835−840.

(19) Hay, D. A.; Fedorov, O.; Martin, S.; Singleton, D. C.; Tallant, C.; Wells, C.; Picaud, S.; Philpott, M.; Monteiro, O. P.; Rogers, C. M.; Conway, S. J.; Rooney, T. P. C.; Tumber, A.; Yapp, C.; Filippakopoulos, P.; Bunnage, M. E.; Müller, S.; Knapp, S.; Schofield, C. J.; Brennan, P. E. Discovery and Optimization of Small-Molecule Ligands for the CBP/P300 Bromodomains. *J. Am. Chem. Soc.* **2014**, *136*, 9308−9319.

(20) Cui, J. J.; Tran-Dubé, M.; Shen, H.; Nambu, M.; Kung, P.-P.; Pairish, M.; Jia, L.; Meng, J.; Funk, L.; Botrous, I.; et al. Structure Based Drug Design of Crizotinib (PF-02341066), a Potent and Selective Dual Inhibitor of Mesenchymal−Epithelial Transition Factor (c-MET) Kinase and Anaplastic Lymphoma Kinase (ALK). *J. Med. Chem.* **2011**, *54*, 6342−6363.

DOI: 10.1021/acs.jcim.9b00731
J. Chem. Inf. Model. 2020, 60, 152−161

(21) Nar, H.; Bauer, M.; Schmid, A.; Stassen, J.-M.; Wienen, W.; Priepke, H. W.; Kauffmann, I. K.; Ries, U. J.; Hauel, N. H. Structural Basis for Inhibition Promiscuity of Dual Specific Thrombin and Factor Xa Blood Coagulation Inhibitors. *Structure* **2001**, *9*, 29−37.
(22) Shi, Y.; O'Connor, S. P.; Sitkoff, D.; Zhang, J.; Shi, M.; Bisaha, S. N.; Wang, Y.; Li, C.; Ruan, Z.; Michael Lawrence, R.; Klei, H. E.; Kish, K.; Liu, E. C.-K.; Seiler, S. M.; Schweizer, L.; Steinbacher, T. E.; Schumacher, W. A.; Robl, J. A.; Macor, J. E.; Atwal, K. S.; Stein, P. D. Arylsulfonamidopiperidone Derivatives as a Novel Class of Factor Xa Inhibitors. *Bioorg. Med. Chem. Lett.* **2011**, *21*, 7516−7521.
(23) Pikul, S.; Dunham, K. M.; Almstead, N. G.; De, B.; Natchus, M. G.; Taiwo, Y. O.; Williams, L. E.; Hynd, B. A.; Hsieh, L. C.; Janusz, M. J.; Gu, F.; Mieling, G. E. Heterocycle-Based MMP Inhibitors with P2′ Substituents. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 1009−1013.
(24) Chang, Y.; Zhang, X.; Horton, J. R.; Upadhyay, A. K.; Spannhoff, A.; Liu, J.; Snyder, J. P.; Bedford, M. T.; Cheng, X. Structural Basis for G9a-like Protein Lysine Methyltransferase Inhibition by BIX-01294. *Nat. Struct. Mol. Biol.* **2009**, *16*, 312−317.
(25) Yung-Chi, C.; Prusoff, W. H. Relationship between the Inhibition Constant (KI) and the Concentration of Inhibitor Which Causes 50 per Cent Inhibition (I50) of an Enzymatic Reaction. *Biochem. Pharmacol.* **1973**, *22*, 3099−3108.
(26) Malhotra, S.; Karanicolas, J. When Does Chemical Elaboration Induce a Ligand To Change Its Binding Mode? *J. Med. Chem.* **2017**, *60*, 128−145.
(27) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; 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.
(28) Pande, V. S.; Beauchamp, K.; Bowman, G. R. Everything You Wanted to Know about Markov State Models but Were Afraid to Ask. *Methods* **2010**, *52*, 99−105.
(29) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. Markov Models of Molecular Kinetics: Generation and Validation. *J. Chem. Phys.* **2011**, *134*, 174105.
(30) Fischer, E. Einfluss Der Configuration Auf Die Wirkung Der Enzyme. *Ber. Dtsch. Chem. Ges.* **1894**, *27*, 2985−2993.
(31) Koshland, D. E. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **1958**, *44*, 98−104.
(32) Boehr, D. D.; Nussinov, R.; Wright, P. E. The Role of Dynamic Conformational Ensembles in Biomolecular Recognition. *Nat. Chem. Biol.* **2009**, *5*, 789−796.
(33) Pohorille, A.; Jarzynski, C.; Chipot, C. Good Practices in Free-Energy Calculations. *J. Phys. Chem. B* **2010**, *114*, 10235−10253.
(34) Wainberg, M.; Merico, D.; Delong, A.; Frey, B. J. Deep Learning in Biomedicine. *Nat. Biotechnol.* **2018**, *36*, 829−838.
(35) Cherkasov, A.; Muratov, E. N.; Fourches, D.; Varnek, A.; Baskin, I. I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y. C.; Todeschini, R.; Consonni, V.; Kuz'Min, V. E.; Cramer, R.; Benigni, R.; Yang, C.; Rathman, J.; Terfloth, L.; Gasteiger, J.; Richard, A.; Tropsha, A. QSAR Modeling: Where Have You Been? Where Are You Going To? *J. Med. Chem.* **2014**, *57*, 4977−5010.
(36) Westbrook, J.; Feng, Z.; Chen, L.; Yang, H.; Berman, H. M. The Protein Data Bank and Structural Genomics. *Nucleic Acids Res.* **2003**, *31*, 489−491.
(37) Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. *J. Comput.-Aided Mol. Des.* **2013**, *27*, 221−234.
(38) Kia, A.; Darve, E. The Accuracy of the CHARMM22/CMAP and AMBER Ff99SB Force Fields for Modelling the Antimicrobial Peptide Cecropin P1. *Mol. Simul.* **2013**, *39*, 922.
(39) Zoete, V.; Cuendet, M. A.; Grosdidier, A.; Michielin, O. Swiss Param: A Fast Force Field Generation Tool for Small Organic Molecules. *J. Comput. Chem.* **2011**, *32*, 2359−2368.
(40) Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A* **2001**, *105*, 9954−9960.
(41) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Di Nola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(42) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(43) Essmann, U.; Perera, L.; Berkowitz, M.; Darden, T.; Lee, H.; Pedersen, L. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(44) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(45) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(46) Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noé, F. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory Comput.* **2015**, *11*, 5525−5542.
(47) Schwantes, C. R.; Pande, V. S. Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9. *J. Chem. Theory Comput.* **2013**, *9*, 2000−2009.
(48) Kanungo, T.; Mount, D. M.; Netanyahu, N. S.; Piatko, C. D.; Silverman, R.; Wu, A. Y. An Efficient K-Means Clustering Algorithm: Analysis and Implementation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 881−892.
(49) Röblitz, S.; Weber, M. Fuzzy Spectral Clustering by PCCA+: Application to Markov State Models and Data Classification. *Adv. Data Anal. Classif.* **2013**, *7*, 147−179.

DOI: 10.1021/acs.jcim.9b00731
J. Chem. Inf. Model. 2020, 60, 152−161