



# Discovery of a hidden transient state in all bromodomain families

Lluís Raich<sup>a</sup>, Katharina Meier<sup>b</sup>, Judith Günther<sup>c</sup>, Clara D. Christ<sup>c</sup>, Frank Noé<sup>a,d,1</sup>, and Simon Olsson<sup>a,1,2</sup>

<sup>a</sup>Department of Mathematics and Computer Science, Freie Universität Berlin, 14195 Berlin, Germany; <sup>b</sup>Computational Molecular Design, Pharmaceuticals, R&D, Bayer AG, 42096 Wuppertal, Germany; <sup>c</sup>Computational Molecular Design, Pharmaceuticals, R&D, Bayer AG, 13342 Berlin, Germany; and <sup>d</sup>Department of Chemistry, Rice University, Houston, TX 77005

Edited by David Baker, University of Washington, Seattle, WA, and approved December 1, 2020 (received for review August 17, 2020)

**Bromodomains (BDs) are small protein modules that interact with acetylated marks in histones. These posttranslational modifications are pivotal to regulate gene expression, making BDs promising targets to treat several diseases. While the general structure of BDs is well known, their dynamical features and their interplay with other macromolecules are poorly understood, hampering the rational design of potent and selective inhibitors. Here, we combine extensive molecular dynamics simulations, Markov state modeling, and available structural data to reveal a transiently formed state that is conserved across all BD families. It involves the breaking of two backbone hydrogen bonds that anchor the ZA-loop with the  $\alpha_A$  helix, opening a cryptic pocket that partially occludes the one associated to histone binding. By analyzing more than 1,900 experimental structures, we unveil just two adopting the hidden state, explaining why it has been previously unnoticed and providing direct structural evidence for its existence. Our results suggest that this state is an allosteric regulatory switch for BDs, potentially related to a recently unveiled BD-DNA-binding mode.**

bromodomains | Markov models | cryptic pockets | allosteric effects | minor conformational states

**P**roteins are highly dynamic biomolecules, often adopting multiple conformational states that can be critical to their function, regulation, and evolvability (1–3). Transitions between the different conformations of a protein may directly connect to changes in several properties, such as protein stability, interactions with binding partners, or the appearance of “cryptic” pockets that could be druggable but are hidden in native structures (4–7). Characterizing and determining such conformations are therefore essential to understanding the intricate relationship between structure, dynamics, and function of proteins, with a direct impact on drug design and protein engineering (8–10). Unfortunately, most of the conformations that a protein can adopt have low populations and are consequently difficult to characterize with conventional biophysical techniques. Indeed, minor conformational states may be masked under certain experimental conditions, making their detection challenging (11, 12). Nevertheless, several methods have emerged to overcome these limitations, including ambient-temperature X-ray crystallography (13, 14), NMR relaxation dispersion experiments (15, 16), and extensive computer simulations (17–20), all of which have been successful in the discovery of such minor but relevant states.

Human bromodomains (BDs) are important epigenetic regulators, recognizing or “reading” acetylation marks in histone tails (21). These domains are part of larger proteins that may also include enzymatic domains to modify chromatin structure or binding domains that can serve as platforms to recruit and activate transcription factors (22). Their central role make BD-containing proteins pivotal to regulate gene expression, representing crucial targets for the treatment of several diseases (23). Significant efforts have gone into characterizing the structure of BDs, aiming to discern the differences between the eight BD families. These studies revealed that all BDs share a common fold

comprised by a four-helix bundle ( $\alpha_Z$ ,  $\alpha_A$ ,  $\alpha_B$ , and  $\alpha_C$ ) with two principal loops that connect them (ZA- and BC-loops), conforming a hydrophobic pocket that is suited for acetyl-lysine binding (Fig. 1) (24). The BC-loop contains a well-conserved hydrogen bond donor or “reader”—usually an asparagine—that is key for recognition. The ZA-loop is the longest loop and the one that displays more variability in terms of sequence and structure, presenting short helices, turns, and even hairpin insertions in BD family VIII (Fig. 1). This loop is anchored to the  $\alpha_A$  helix via two conserved backbone hydrogen bonds (h-bonds), forming a cavity that is known as the ZA channel, an important structural feature for optimizing the potency and selectivity of BD inhibitors (23, 25). Characterization of BD flexibility has been, on the other hand, limited to a few studies reporting different crystallographic structures and molecular dynamics (MD) simulations that show subtle motions of the ZA-loop, as well as transitions between rotameric states of residues involved in ligand binding (26–29).

Although these insights have helped in the general understanding of BDs, the poor comprehension of their conformational ensembles, their biological relevance within BD-containing proteins, and their precise role in disease are hampering the advance of BD inhibitors into the clinic (30). A recent study illustrates this challenge through a detailed characterization of the impact of

## Significance

**Proteins dynamically shape shift, adopting different three-dimensional structures, called conformations. The conformations fulfill different functions and expose different binding surfaces to drug molecules aiming to inhibit the protein. While structural biology experiments can capture the most frequently occurring conformations, those that are more rarely visited are very difficult to capture but may be functionally important. Using extensive computer simulations, we discover an undocumented conformation present in all families of bromodomains—protein domains crucial for epigenetic regulation and associated with several diseases. We also show that available experimental data confirms the existence of this previously overlooked state. The surprising ubiquity of this new conformation makes it a likely player in biological function and opens up new drug development strategies.**

Author contributions: L.R., K.M., J.G., C.D.C., F.N., and S.O. designed research; L.R. and S.O. performed research; L.R. contributed new analytic tools; L.R., K.M., J.G., C.D.C., and S.O. analyzed data; and L.R., F.N., and S.O. wrote the paper.

Competing interest statement: K.M., J.G., and C.D.C. are employees and/or shareholders of Bayer AG.

This article is a PNAS Direct Submission.

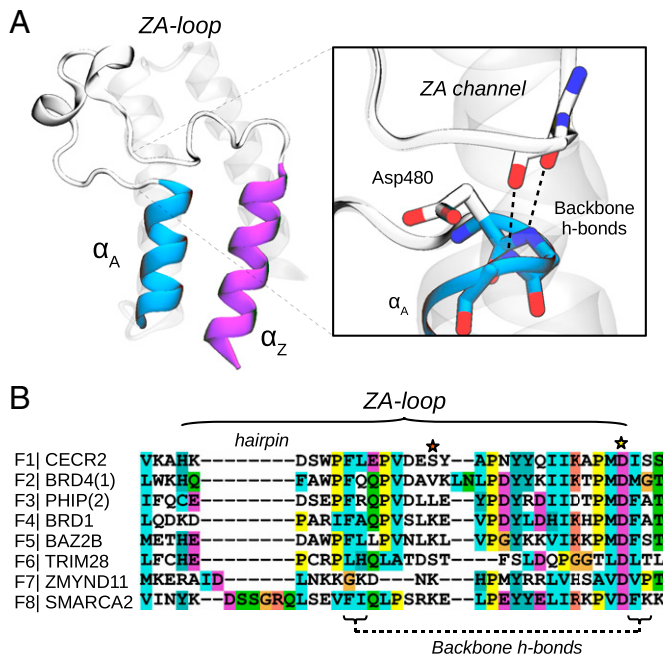
Published under the PNAS license.

<sup>1</sup>To whom correspondence may be addressed. Email: frank.noe@fu-berlin.de or simonols@chalmers.se.

<sup>2</sup>Present address: Department of Computer Science and Engineering, Chalmers University of Technology, 41258 Gothenburg, Sweden.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2017427118/-DCSupplemental>.

Published January 19, 2021.



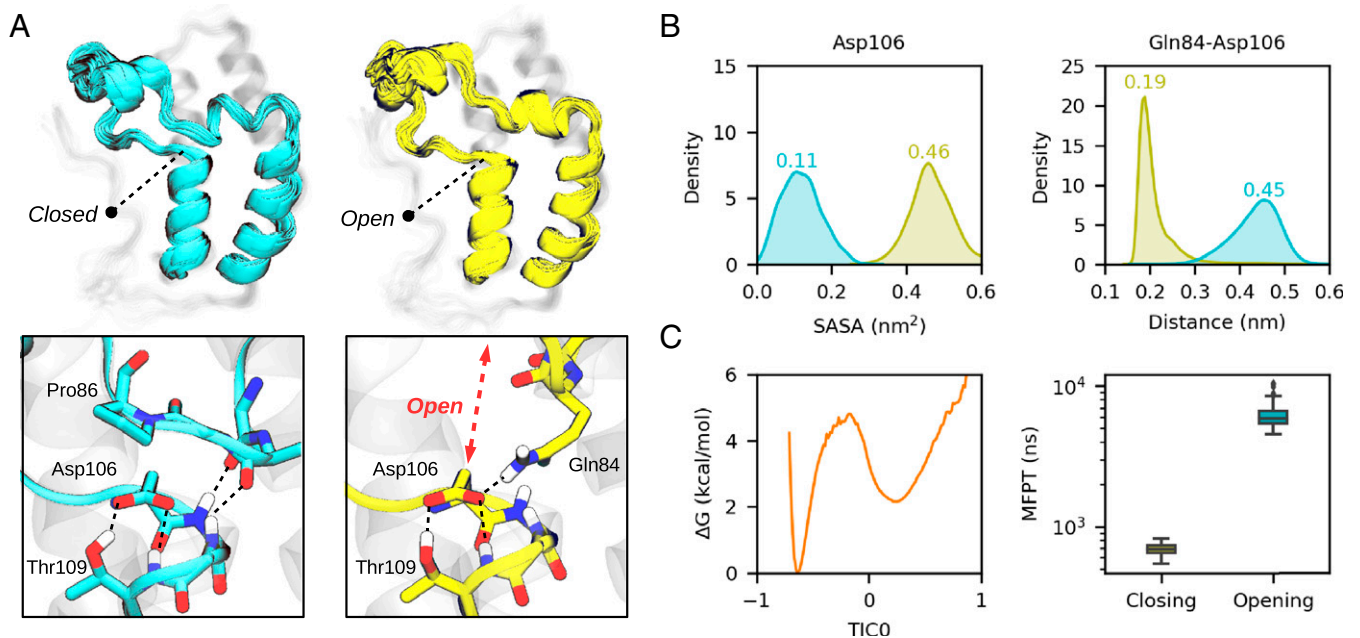
**Fig. 1.** Bromodomains have two conserved backbone h-bonds that anchor the ZA-loop with the  $\alpha_A$  helix. (A) Structure of CECR2 (PDB 3NXB) highlighting the two first helices,  $\alpha_Z$  and  $\alpha_A$ , and the long ZA-loop, detailing the ZA channel and the two conserved backbone h-bonds that anchor the loop with the  $\alpha_A$  helix. (B) Sequence alignment of the ZA-loop region for the BDs analyzed in this work, including all families. Note the presence of a hairpin insertion in family VIII and the well-conserved Asp480 (yellow star). The orange star highlights lysine residues commented below.

missense variation on BD function (31). The authors show that while specific variants did not alter the global fold of BDs in crystal structures, spectroscopic measurements in solution suggest

changes in dynamic properties that could account for the loss of protein stability and inhibitor affinity. Furthermore, another recent study demonstrated that BDs can directly bind to DNA through a basic patch on their surface and that, in fact, DNA-binding drives the association of BDs with chromatin (32). These macroscopic interactions could affect the structure and flexibility of BDs, modulating their function and properties, which opens questions on the relevance and precise role of BDs inside their biological context.

MD simulations have been pivotal to complement experiments in the understanding of protein function, focusing on their dynamics and flexibility. MD simulations, in principle, grant direct access to proteins at the full spatial and temporal resolution. However, extensive simulations may be needed to sample the rare transitions between long-lived protein conformations. The combination of high-throughput MD simulations with Markov state models (MSMs) provides a powerful framework to recover long-time dynamics from multiple short-time simulations, following a “divide-and-conquer” philosophy (33). These techniques have enabled the quantitative study of critical biological processes, such as protein folding or protein-protein association (34, 35), protein plasticity and ligand-binding kinetics (36), as well as to characterize cryptic pockets in proteins and find common states across protein families (9, 37), making them very useful to evaluate the structural flexibility of BDs.

Here, using MD simulations and MSMs, we survey representative BDs from the eight families to characterize their conformational dynamics. Although the analyzed proteins share little sequence identity, we identify two recurring, transient states in which the two backbone hydrogen bonds that anchor the ZA-loop with the  $\alpha_A$  helix break, and the acetyl-lysine binding pocket becomes occluded. The population of this state is generally smaller than that of the crystallographic state, suggesting why so far it has escaped experimental observation. Nonetheless, through a detailed analysis of more than 1,900 experimental structures, we reveal two BDs that adopt the hidden state, confirming its existence. We propose that this conformational change represents a mechanism of autoregulation in BDs, potentially related to a



**Fig. 2.** A hidden conformational state in BDs. (A) Ensemble of structures of the “closed” crystal-like state (cyan) and the “open” state (yellow) for BRD4(1). A close view of the opening region is represented below, highlighting Asp106. (B) Distribution of the Asp106 solvent accessible surface area and closest distance between Gln84-Asp106 side chains for the two metastable states. (C) MSM reweighted free energy profile along the first independent component (TIC0, arbitrary units) and boxplot of a bootstrapped distribution of the closing and opening mean first passage times.

BD-DNA-binding mode, and reveals allosteric pockets that could be leveraged to modulate both histone and DNA-binding contributions to chromatin, offering new starting points to overcome the current challenges facing the design of potent and selective BD inhibitors.

## Results

**All BD Families Share a Conserved and Hidden Metastable State.** We started our study focusing on BRD4(1), a thoroughly characterized BD of family II. We ran 64 independent, 1- $\mu$ s MD simulations in explicit solvent to build an MSM using the PyEMMA software (see *SI Appendix, Computational details and methodology* for details) (38). Analysis of the resulting model reveals a minor metastable state apart from the one that resembles the crystallographic structure, which involves the displacement of the ZA-loop from the  $\alpha_A$  helix, opening a space beneath it that increases the solvent accessible surface area of the well-conserved aspartate (Asp106, see Fig. 2). Interestingly, the conformational change disrupts the ZA channel, a structural feature that is relevant for inhibitor selectivity. In terms of interactions, the opening process involves the breaking of the two conserved backbone h-bonds, whose interactions are partially compensated by the h-bonds that Gln84 establishes with Asp106, acting as a “latch.” The free energy profile along the slowest time-lagged independent component (or TIC (39, 40), a type of collective variable) shows two clear basins, with the open state being  $\sim 2$  kcal/mol<sup>-1</sup> above the closed (Fig. 2C). The mean first passage time (MFPT) for the opening is of  $5.9 \pm 1.2$   $\mu$ s and  $688 \pm 61$  ns for the closing, indicated by medians and one SD from a bootstrapping distribution.

The fact that the opening implies the breaking of two backbone h-bonds, likely present in all BDs, prompted us to analyze the structural flexibility of other BD families, wondering whether this could be a general conformational feature across them. We selected a member of each BD family based on a sequence alignment, focusing on a set of residues that lay around the conserved aspartate that is beneath the ZA-loop. In total, ordered by families, we decided to study CECR2, BRD4(1), PHIP(2), BRD1, BAZ2B, TRIM28, ZMYND11, and SMARCA2 (see *SI Appendix, Table S1* for the Protein Data Bank [PDB] identifiers and amino acid sequences). We ran 40 to 60 independent 1- $\mu$ s MD for each BD, collecting an aggregate simulation data of 392  $\mu$ s, which represents, as far as we know, the most extensive computational study of BDs.

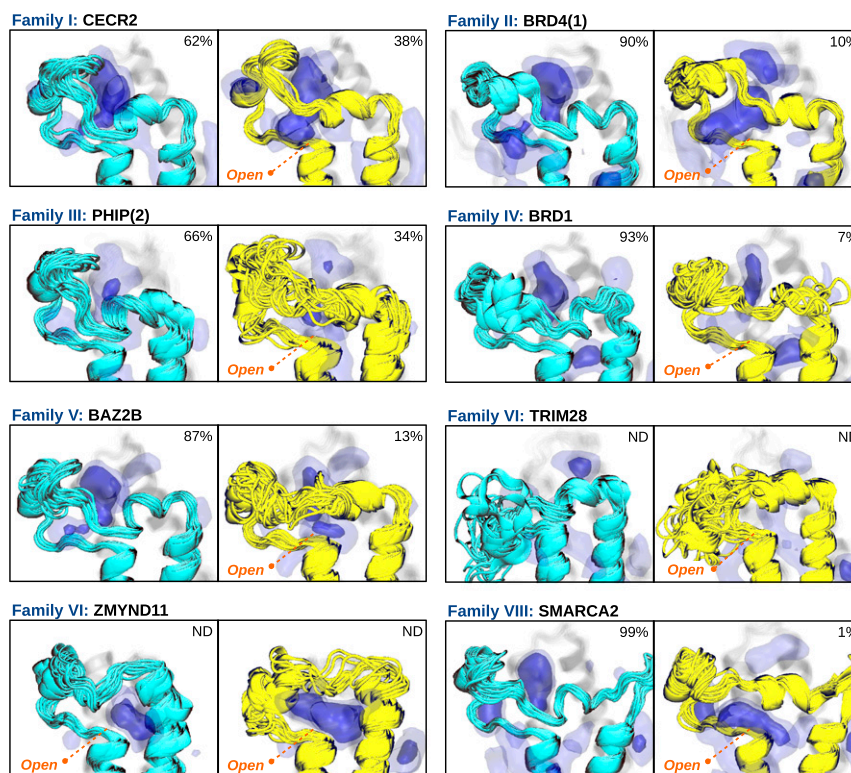
Strikingly, our results reveal that the hidden state is conserved in all families, with slight differences in their populations, exchange timescales, and structural features (Fig. 3 and *SI Appendix, Figs. S2–S9 and Tables S2 and S3*). For instance, CECR2 and PHIP(2) show a population of the open state above 30%; BRD4(1), BRD1, and BAZ2B around 10%; and SMARCA2 below 5%. ZMYND11 is a special case because it is always in the open state—that is, 100%—even though we observe a “semi-closed” state in which the extreme part of the ZA-loop partially closes. TRIM28 is another special case, as its ZA-loop is wrapped on top of the  $\alpha_A$  and  $\alpha_B$  helices in the experimental structure. It presents a quite unstable fold and a complex conformational landscape that would require an extensive simulation effort to compute state populations and rates reliably, and thus, we do not provide values in this case. Nonetheless, we have observed several trajectories showing clear metastability in the open state, confirming its kinetic relevance (*SI Appendix, Fig. S7*). In terms of timescales, CECR2 is the BD that opens fastest, with a median MFPT of  $182 \pm 7$  ns, followed by PHIP(2) with  $903 \pm 70$  ns, BAZ2B with  $2.6 \pm 0.5$   $\mu$ s, BRD4(1) with  $5.9 \pm 1.2$   $\mu$ s, BRD1 with  $8.3 \pm 4.1$   $\mu$ s, and SMARCA2 with  $11.5 \pm 1.5$   $\mu$ s (*SI Appendix, Table S3*). The closing event is generally one order of magnitude faster than the opening, all of them below 1  $\mu$ s, having SMARCA2

closing within  $68 \pm 3$  ns, followed by CECR2 within  $110 \pm 4$  ns, BAZ2B within  $377 \pm 57$  ns, PHIP(2) within  $449 \pm 51$  ns, BRD1 within  $643 \pm 117$  ns, and BRD4(1) within  $688 \pm 61$  ns. These results may explain why the hidden state has been elusive in previous studies, since published simulation timescales are usually around few hundreds of nanoseconds (41–45), which is insufficient to characterize the opening event for most of the BDs analyzed here.

At a structural level, the hidden state shows an overall similarity between BDs, with a clear displacement of the ZA-loop from the  $\alpha_A$  helix (Fig. 3). Although this conformational change generally involves the breaking of the two backbone h-bonds (*SI Appendix, Table S4 and Figs. S10 and S11*), there are subtle differences between systems that are worth noting. For instance, CECR2 is the only BD that retains a substantial percentage of h-bond 1 in the open state, and thus, the opening process is better described by the breaking of h-bond 2. BRD1 is another interesting system in this regard, since the two key h-bonds are only transiently formed in the closed state. This happens as BRD1 switches between the crystallographic state and an alternative state in which the ZA-loop is slightly displaced toward the  $\alpha_Z$  helix, breaking the h-bonds (*SI Appendix, Fig. S12*). The loss of these key interactions is compensated by backbone amides of the ZA-loop, establishing h-bonds with the conserved aspartate to stabilize the closed state, resembling the interactions occurring in TRIM28 and in most BDs of family VIII (*SI Appendix, Figs. S7A and S9A*). Interestingly, this alternative state is also present in CECR2, PHIP(2), and BAZ2B, albeit much less populated. Finally, we note that the majority of the BDs analyzed here display sporadic opening events in which either one or the two h-bonds break. These openings occur within a microsecond timescale and are generally not metastable (*SI Appendix, Figs. S13 and S14*). This observation suggests a duality between the classical mechanisms of “fast” induced fit and “slow” conformational selection (41), both of which may be interesting for drug design.

**A Cryptic and Allosteric Pocket Opens in the Hidden State.** An intriguing aspect of the conformational change is that it always disrupts the ZA channel, reducing the accessibility to the acetyllysine pocket (*SI Appendix, Fig. S15*). This is relevant since the ZA channel is a key structural element for the design of inhibitors, conferring them potency and selectivity (46–48). Therefore, finding pockets whose targeting could stabilize the open state would be suitable for controlling BD function. To that end, we have used the MDpocket software to analyze a thousand representative structures of each metastable state for every BD (49). The results show the presence of binding pockets in the open state, just beneath the ZA-loop (Fig. 3). We classify these cryptic pockets in two groups depending on their opening region. For instance, in CECR2, BRD4(1), BRD1, and SMARCA2, the pockets mostly open in the region where the ZA-loop was lining, next to the well-conserved aspartate, while in BAZ2B and ZMYND11, the pockets open more in between the ZA-loop and the  $\alpha_Z$  and  $\alpha_A$  helices. We attribute this difference to a lysine residue found in the upper part of the ZA-loop (Fig. 1B, orange star) that reorients to interact with the exposed aspartate (*SI Appendix, Figs. S6 and S8*). We note that BRD1 also has an alternative conformational state, not included in the model, in which the pocket is displaced toward the  $\alpha_Z$  and  $\alpha_A$  helices by a lysine residue (*SI Appendix, Fig. S16*). Interestingly, the presence of this lysine in SMARCA2 does not have the same effect, which may be due to the electrostatic shielding of two charged residues that are just at the beginning of the  $\alpha_A$  helix, flanking the conserved aspartate (*SI Appendix, Fig. S9*).

The binding sites of PHIP(2) are less defined, both in the open and closed states, given the large fluctuations of the ZA-loop,



**Fig. 3.** All bromodomain families share a hidden state that opens a cryptic pocket beneath the ZA-loop. Ensemble of structures of the “closed” crystal-like state (cyan) and the “open” state (yellow) for each BD family. MDpocket frequency maps are represented by blue isosurfaces at 0.25 (light) and 0.50 (intense) isovalues, highlighting different binding sites. Percentages refer to medians of MSM populations obtained from a bootstrapped distribution (*SI Appendix, Table S2*). Populations for TRIM28 have not been determined due to its complex conformational landscape and neither have the ones of ZMYND11 since all structures are in the open state.

which partially obstructs the acetyl-lysine pocket. However, relying on the opening motion and the lack of a pocket in the central part, we can include this BD in the first group together with CECR2, BRD4(1), BRD1, and SMARCA2. Interestingly, the partial obstruction of the acetyl-lysine pocket is in agreement with the fact that PHIP(2) has been difficult to target, and thus, only a recent study reports a few hits that weakly bind the outermost region of the pocket (50). These results highlight the importance of considering structural ensembles instead of single structures to determine the druggability of a pocket (51).

TRIM28 is an exception in terms of pockets given the unusual conformation of the ZA-loop. In the closed state, for instance, it increases the solvent exposure of the acetyl-lysine pocket such that it is not anymore defined, and in the open state it fluctuates so much that any clear pocket appears in the region of interest.

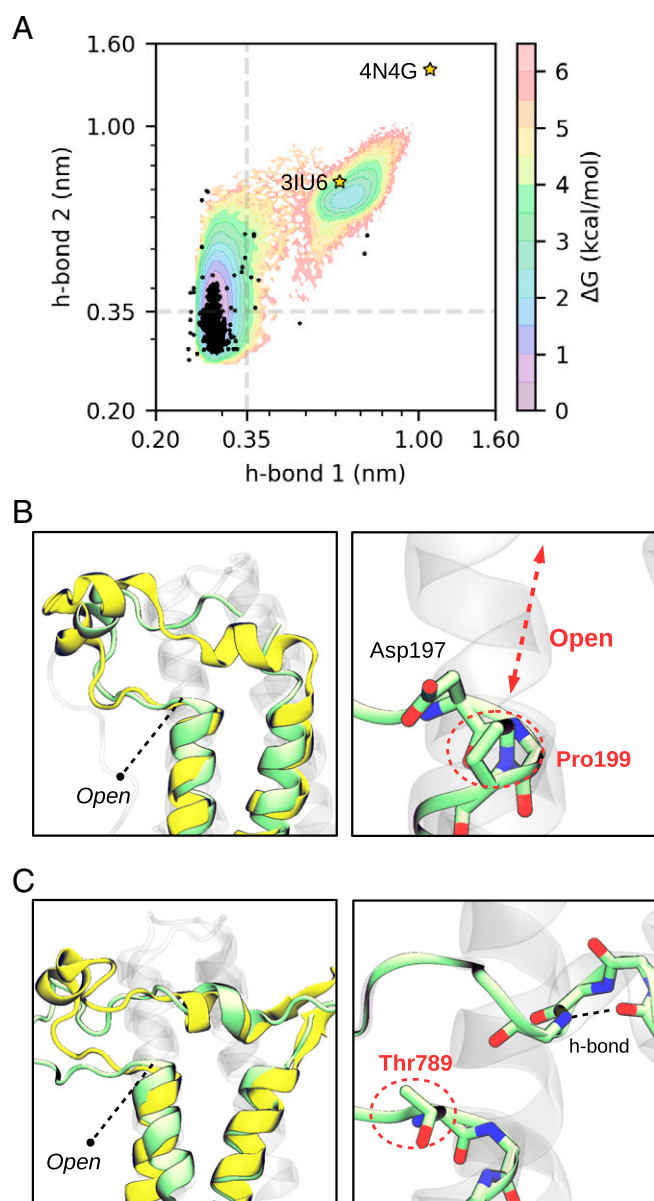
We assessed the druggability of the cryptic pockets in the open state compared to the acetyl-lysine pocket in the closed state (*SI Appendix, Fig. S17*), using a similar protocol as in the work of Porter et al. (20). We observe the distribution of scores spanning the entire range of values, with the cryptic pockets generally showing a lower druggability than the acetyl-lysine one, possibly due to their solvent exposition.

Finally, we have superposed the closed state of our models with crystal structures that have drugs bound in the acetyl-lysine pocket, finding an excellent agreement between the pose of all drugs and the computed isosurfaces (*SI Appendix, Figs. S2–S6 and S9*), including a deeply buried drug in SMARCA2, for which a set of conserved water molecules of the pocket are known to be displaceable (48, 52).

#### Crystallographic Evidence Confirms the Existence of the Open State.

To validate our results, we analyzed all the structures that are present in the bromodomain entry (PF00439) of the Pfam database (53), accounting a total of 1,903 structures (see data repository for the complete list of PDBs). We have computed the distances of the two backbone h-bonds for all structures and projected them onto the free energy landscape of BRD4(1) obtained with the same descriptors. Remarkably, the experimental distribution of h-bonds closely overlaps with the low free energy regions of the landscape, showing a small variance for the first h-bond and a more pronounced variance for the second (see Fig. 4A and *SI Appendix, Fig. S18* for the projection onto the other BDs). This is a surprising result since the analyzed structures are of BDs with very different sequences, including organisms that are genetically far from humans, and yet the free energy landscape of a single BD encloses all this variation. Most importantly, this analysis has allowed us to detect outliers in the distribution and find two crystal structures that are stable in the open state (Fig. 4B and C). These structures are the ones of ZMYND11 and PB1(6), BDs that present substantial sequence variation—particularly in the ZA-loop region—compared to the general trend.

A close inspection of ZMYND11 reveals the presence of a proline residue (Pro199) in place of the residue that acts as donor for the second h-bond, impeding its formation. This chemical modification presumably contributes to destabilize the closed state in this BD. It is worth noting that in most crystal structures of ZMYND11 the ZA-loop is not resolved. In a notable exception (PDB 4N4G), authors proved that contacts with another crystallographic unit stabilize this flexible region, making it observable (54). This is consistent with our simulations, as we



**Fig. 4.** Distribution of h-bonds in experimental structures reveal two BDs in the open state. (A) Projection of all BD structures from the Pfam database (black dots) on a MSM reweighted free energy landscape of BRD4(1) comprising the two conserved backbone h-bonds. Axes are given in a logarithmic scale and dashed lines indicate a distance of 0.35 nm as an upper bound for h-bond formation. The stars highlight the two crystal structures that are in the open state. (B) The structure of ZMYND11 (pale green, PDB 4N4G) is compared with the open state predicted for BRD4(1) (yellow). Pro199 is highlighted next to the conserved Asp. (C) The structure of PB1(6) (pale green, PDB 3IU6) is compared with the open state predicted for SMARCA2 (yellow). Thr789, in place of the conserved Asp, is highlighted together with an internal h-bond that is formed in the short helix of the ZA-loop.

find the ZA-loop switching between the two states that are shown in Fig. 3, making it difficult to capture its electron density. Importantly, this observation provides direct evidence of the possibility to modulate BD flexibility with macromolecular contacts, suggesting that similar interactions with other biological entities—for example, DNA—could also lead to such conformational changes.

The other crystallographic evidence is a structure of PB1(6), which is also an atypical BD having an unusually short ZA-loop

(24). In comparison with SMARCA2, a member of the same family, it shows a very similar opening despite having a low sequence identity (Fig. 4C). The presence of a bulky threonine residue (Thr789) in place of the highly conserved aspartate may be one of the reasons why this BD is not stable in the closed state. We note that there are a few other BDs lacking this aspartate, and yet their crystal structures are stable in the closed state (*SI Appendix*, Fig. S19). In these BDs, the aspartate is replaced by residues like serine, alanine, or tryptophan, which represent drastic changes in terms of amino acid properties. Nonetheless, these modifications are accompanied by changes in the surrounding residues, leading to complementary interactions. This highlights that epistatic effects can compensate for the lack of the conserved aspartate, adapting local interactions to stabilize the closed state and retain BD function.

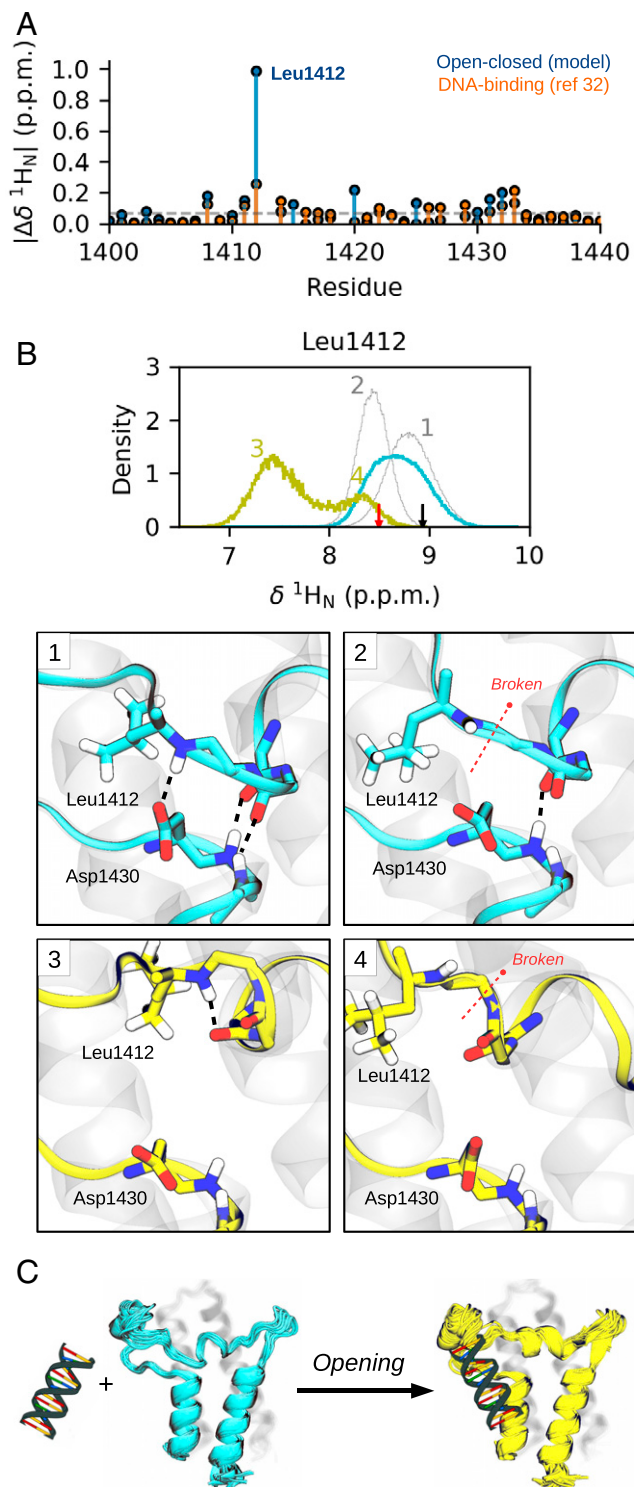
#### Chemical Shift Predictions Suggest a Link between the Hidden State and a DNA-Binding Mode.

Understanding the biological relevance of conformational states in a protein is crucial for the success of inhibitor design. However, this is not an easy task given the number of possible interactions that can occur in a cellular context. A recent study has demonstrated that SMARCA2 interacts with DNA through a basic surface patch, whose electrostatic contribution drive BD association to chromatin (32). Reported chemical shifts upon BD-DNA binding show significant perturbations in the region of the ZA-loop and the  $\alpha_A$  helix, adjacent to where we observe the conformational change. This observation prompted us to compute the chemical shift of  $^1\text{H}_N$ ,  $^{13}\text{C}\alpha$ , and  $^{15}\text{N}$  atoms along the process, using the CamShift empirical predictor (*SI Appendix*, Figs. S20 and S21) (55). Intriguingly, comparison of the predicted  $^1\text{H}_N$  shifts with the experimental values reveal a clear perturbation for Leu1412 (Fig. 5A), which is hydrogen bonded with the conserved Asp1430 (Fig. 5B). This interaction can break both in the closed and the open state, changing the local environment of Leu1412 and lowering its  $^1\text{H}_N$  shift. Therefore, it seems plausible that the experimental perturbation also relates to the breaking of the same h-bond, either in the closed or the open state, likely due to direct contacts with DNA. These results suggest that SMARCA2 could undergo a population shift from the closed state to the open as a consequence of the interactions with DNA (Fig. 5C), modulating its flexibility and function in an allosteric manner. This hypothesis opens the door for future experiments to further describe the nature of the open state and its potential relevance to DNA binding.

#### Discussion

In this work, we have used extensive MD simulations and Markov state modeling to characterize the dynamics and flexibility of BDs. Our results reveal a hidden and ubiquitous state in which the ZA-loop displaces from the  $\alpha_A$  helix, disrupting the ZA channel and opening a pocket beneath it. The conservation of this state across families, persisting along evolutionary pressures, suggests that it has a pivotal role in BD function and regulation.

Crystallographic structures projected onto a free energy map show an excellent agreement between the distribution of experimental structures and simulations. These results highlight that the conformational change is governed by a few conserved interactions among BDs, reinforcing previous studies that demonstrate the correspondence between ensembles of crystallographic structures and protein plasticity, both in solution and during enzyme catalysis (56, 57). Furthermore, we report that two crystallographic structures, one of ZMYND11 and one of PB1(6), are stable in the open state. These structures directly support the existence of the minor state, demonstrating that the ZA-loop can be displaced from the  $\alpha_A$  helix and the protein core can be partially exposed to the solvent. Moreover, these structures reveal subtle changes of conserved residues that could help in the rationalization of



**Fig. 5.** Chemical shift predictions of SMARCA2 along the conformational change and experimental chemical shift changes of a BD-DNA binding mode. (A) CamShift  $^1\text{H}_N$  predictions along the conformational change of SMARCA2 and experimental shifts upon its binding to DNA (5'-CTCAATTGGT-3'), obtained from Morrison et al. (32). The numbering of residues is that of PDB 5DKC, with a zoom in on residues 1,400 to 1,440. (B) Distribution of Leu1412  $^1\text{H}_N$  shift for the closed and open states (cyan and yellow, respectively). The cyan distribution can be split in two Gaussian distributions (shown in gray) that correspond to closed states with the Leu1412-Asp1430 h-bond formed or broken (panels 1 to 2). The yellow distribution has two clear modes that correspond to open states in which Leu1412 amide is placed or not in between the two backbone carbonyls (panels 3 to 4). The black arrow indicates

disease-related mutations. One of these changes is Pro199 in ZMYND11, which abrogates one of the key h-bonds that stabilize the closed state. This observation suggests that BDs could be engineered in the open state with a single site directed mutation, possibly switching between biologically active and inactive forms, with applications in target validation. Another change is Thr789 in PB1(6), which replaces the well-conserved aspartate, and whose mutation is in tumor suppressor genes in BDs of the same family (e.g., PB1(5) and the D705G variant) (58).

The open state represents a potential target for future drug design campaigns, given the presence of cryptic pockets with allosteric effects (59, 60). These pockets open in slightly different regions depending on the BD, offering alternative possibilities for interactions compared to those present in the acetyl-lysine pocket, being promising to overcome selectivity problems. Interestingly, one of the crystal structures of ZMYND11 (PDB 4N4I) shows a polyethylene glycol molecule bound beneath the predicted pocket, in the hydrophobic core of the protein, suggesting that the novel pockets may be indeed ligandable (*SI Appendix, Fig. S8D*). This evidence further opens the question of whether different molecules can act as allosteric modulators in a cellular context. More generally, we envisage the possibility of controlling BD function via the design of specific inhibitors targeting these pockets.

Finally, we have proposed that the hidden state possibly connects to a recently unveiled BD-DNA-binding mode, which displays significant chemical shift perturbations in the opening region. We hypothesize that such biologically relevant interactions could shift the population between the states, stabilizing the open state to occlude the acetyl-lysine pocket and regulate BD function. Most importantly, our detailed chemical shift analyses of the opening process reveal signals that may be particularly sensitive for future validation via NMR relaxation measurements (29, 61, 62), independently of our DNA-binding hypothesis.

Altogether, we believe that our results significantly contribute to the general understanding of BD structure, dynamics, and function, providing a mechanism of potential autoregulation that deepens in BD biology and paves the way toward the design of future inhibitors with higher potency and selectivity.

## Materials and Methods

We performed MD simulations at 300 K and 1 bar using OpenMM (63), with the ff14SB (64) force field parameters to describe the solute and the TIP3P (65) water model to describe the solvent. We ran multiple 1- $\mu\text{s}$  trajectories for each bromodomain and subsequently used the PyEMMA (38) software to build MSMs from this data. MDpocket (49) was used to detect small molecule binding sites, and Plumed (66) was used to predict chemical shifts via the CamShift model (55). Bromodomain structures of the Pfam database (PF00439) were downloaded from the PDB and superimposed with Theseus (67). Further methodological details, convergence tests, and analyses are available in the *SI Appendix*.

**Data Availability.** All molecular trajectories and notebooks to reproduce the results of this paper are available in Zenodo at <https://doi.org/10.5281/zenodo.4343283>.

**ACKNOWLEDGMENTS.** We acknowledge funding from the European Commission (ERC CoG 772230 "ScaleCell"), the Deutsche Forschungsgemeinschaft (CRC1114/C03, NO825/2-2), the MATH+ Berlin Mathematics Research Center (AA1x6, EF1x2), the Alexander von Humboldt Foundation (postdoctoral fellowship to S.O.), and Bayer AG (LSC project: "QuAnTaStics"). We thank Jacopo Negroni, Joseph Bluck, Michael Edmund Beck, Sikander Hayat, Jeremie Mortier, Holger Steuber, and Nicolas Werbeck for feedback and discussions at different stages during this study. We also thank Emma A. Morrison and Catherine A. Musselman for sharing raw chemical shift data.

the CamShift prediction for PDB 5DKC, and the red arrow indicates the mean prediction for PDB 2DAT. (C) Hypothesis of a closed-to-open population shift as a consequence of BD interactions with DNA.

1. N. Tokuriki, D. S. Tawfik, Protein dynamism and evolvability. *Science* **324**, 203–207 (2009).
2. E. Campbell *et al.*, The role of protein dynamics in the evolution of new enzyme function. *Nat. Chem. Biol.* **12**, 944–950 (2016).
3. M. A. María-Solano, E. Serrano-Hervás, A. Romero-Rivera, J. Iglesias-Fernández, S. Osuna, Role of conformational dynamics in the evolution of novel enzyme function. *Chem. Commun. (Camb.)* **54**, 6622–6634 (2018).
4. M. Karplus, J. Kuriyan, Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6679–6685 (2005).
5. V. Oleinikovas, G. Saladino, B. P. Cossins, F. L. Gervasio, Understanding cryptic pocket formation in protein targets by enhanced sampling simulations. *J. Am. Chem. Soc.* **138**, 14257–14263 (2016).
6. D. Beglov *et al.*, Exploring the structural origins of cryptic sites on proteins. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E3416–E3425 (2018).
7. A. Kuzmanic, G. R. Bowman, J. Juárez-Jimenez, J. Michel, F. L. Gervasio, Investigating cryptic binding sites by molecular dynamics simulations. *Acc. Chem. Res.* **53**, 654–661 (2020).
8. G. Bouvignies *et al.*, Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature* **477**, 111–114 (2011).
9. G. R. Bowman, P. L. Geissler, Equilibrium fluctuations of a single folded protein reveal a multitude of potential cryptic allosteric sites. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 11681–11686 (2012).
10. C. R. Knoverek, G. K. Amarasinghe, G. R. Bowman, Advanced methods for accessing protein shape-shifting present new therapeutic opportunities. *Trends Biochem. Sci.* **44**, 351–364 (2019).
11. M. Wong, G. Khirich, J. P. Loria, What's in your buffer? Solute altered millisecond motions detected by solution NMR. *Biochemistry* **52**, 6548–6558 (2013).
12. M. Fischer, B. K. Shoichet, J. S. Fraser, One crystal, two temperatures: Cryocooling penalties alter ligand binding to transient protein sites. *ChemBioChem* **16**, 1560–1564 (2015).
13. J. S. Fraser *et al.*, Hidden alternative structures of proline isomerase essential for catalysis. *Nature* **462**, 669–673 (2009).
14. J. S. Fraser *et al.*, Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 16247–16252 (2011).
15. A. G. Palmer 3rd, Chemical exchange in biomacromolecules: Past, present, and future. *J. Magn. Reson.* **241**, 3–17 (2014).
16. T. R. Alderson, L. E. Kay, Unveiling invisible protein states with NMR spectroscopy. *Curr. Opin. Struct. Biol.* **60**, 39–49 (2020).
17. K. M. Hart, C. M. W. Ho, S. Dutta, M. L. Gross, G. R. Bowman, Modelling proteins' hidden conformations to predict antibiotic resistance. *Nat. Commun.* **7**, 12965 (2016).
18. S. Chen *et al.*, The dynamic conformational landscape of the protein methyltransferase SETD8. *eLife* **8**, e45403 (2019).
19. J. Juárez-Jiménez *et al.*, Dynamic design: Manipulation of millisecond timescale motions on the energy landscape of cyclophilin A. *Chem. Sci.* **11**, 2670–2680 (2020).
20. J. R. Porter *et al.*, Cooperative changes in solvent exposure identify cryptic pockets, switches, and allosteric coupling. *Biophys. J.* **116**, 818–830 (2019).
21. M. Schiedel *et al.*, Chemical epigenetics: The impact of chemical- and chemical biology techniques on bromodomain target validation. *Angew. Chem. Int. Ed. Engl.* **58**, 17930–17952 (2019).
22. N. Zaware, M. M. Zhou, Bromodomain biology and drug discovery. *Nat. Struct. Mol. Biol.* **26**, 870–879 (2019).
23. N. H. Theodoulou, N. C. O. Tomkinson, R. K. Prinjha, P. G. Humphreys, Progress in the development of non-BET bromodomain chemical probes. *ChemMedChem* **11**, 477–487 (2016).
24. P. Filippakopoulos *et al.*, Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell* **149**, 214–231 (2012).
25. K. Cheung *et al.*, BET N-terminal bromodomain inhibition selectively blocks Th17 cell differentiation and ameliorates colitis in mice. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 2952–2957 (2017).
26. J. Zhu *et al.*, Binding motifs in the CBP bromodomain: An analysis of 20 crystal structures of complexes with small molecules. *ACS Med. Chem. Lett.* **9**, 929–934 (2018).
27. S. Steiner, A. Magno, D. Huang, A. Caflich, Does bromodomain flexibility influence histone recognition? *FEBS Lett.* **587**, 2158–2163 (2013).
28. G. Poncet-Montange *et al.*, Observed bromodomain flexibility reveals histone peptide- and small molecule ligand-compatible forms of ATAD2. *Biochem. J.* **466**, 337–346 (2015).
29. F. M. Ferguson *et al.*, Binding hotspots of BAZ2B bromodomain: Histone interaction revealed by solution NMR driven docking. *Biochemistry* **53**, 6706–6716 (2014).
30. A. G. Cochran, A. R. Conery, R. J. Sims 3rd, Bromodomains: A new target class for drug development. *Nat. Rev. Drug Discov.* **18**, 609–628 (2019).
31. L. Lori *et al.*, Effect of BET missense mutations on bromodomain function, inhibitor binding and stability. *PLoS One* **11**, e0159180 (2016).
32. E. A. Morrison *et al.*, DNA binding drives the association of BRG1/hBRM bromodomains with nucleosomes. *Nat. Commun.* **8**, 16080 (2017).
33. J. H. Prinz *et al.*, Markov models of molecular kinetics: Generation and validation. *J. Chem. Phys.* **134**, 174105 (2011).
34. F. Noé, C. Schütte, E. Vanden-Eijnden, L. Reich, T. R. Weikl, Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 19011–19016 (2009).
35. N. Plattner, S. Doerr, G. De Fabritiis, F. Noé, Complete protein-protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling. *Nat. Chem.* **9**, 1005–1011 (2017).
36. N. Plattner, F. Noé, Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. *Nat. Commun.* **6**, 7653 (2015).
37. M. M. Sultan, K. Kiss, V. S. Pande, Towards simple kinetic models of functional dynamics for a kinase subfamily. *Nat. Chem.* **10**, 903–909 (2018).
38. M. K. Scherer *et al.*, PyEMMA 2: A software package for estimation, validation, and analysis of Markov models. *J. Chem. Theory Comput.* **11**, 5525–5542 (2015).
39. G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis, F. Noé, Identification of slow molecular order parameters for Markov model construction. *J. Chem. Phys.* **139**, 015102 (2013).
40. C. R. Schwantes, V. S. Pande, Improvements in Markov State Model construction reveal many non-native interactions in the folding of NTL9. *J. Chem. Theory Comput.* **9**, 2000–2009 (2013).
41. Y. Zhou, M. Hussain, G. Kuang, J. Zhang, Y. Tu, Mechanistic insights into peptide and ligand binding of the ATAD2-bromodomain via atomistic simulations disclosing a role of induced fit and conformational selection. *Phys. Chem. Chem. Phys.* **20**, 23222–23232 (2018).
42. J. Su *et al.*, A computational insight into binding modes of inhibitors XD29, XD35, and XD28 to bromodomain-containing protein 4 based on molecular dynamics simulations. *J. Biomol. Struct. Dyn.* **36**, 1212–1224 (2018).
43. B. K. Allen *et al.*, Identification of a novel class of BRD4 inhibitors by computational screening and binding simulations. *ACS Omega* **2**, 4760–4771 (2017).
44. Q. Wang *et al.*, Computational study on the selective inhibition mechanism of MS402 to the first and second bromodomains of BRD4. *Proteins* **87**, 3–11 (2019).
45. Q. Wang *et al.*, Selective inhibition mechanism of RVX-208 to the second bromodomain of bromo and extraterminal proteins: Insight from microsecond molecular dynamics simulations. *Sci. Rep.* **7**, 8857 (2017).
46. N. Bharatham, P. J. Slavish, W. R. Shadrick, B. M. Young, A. A. Shelat, The role of ZA channel water-mediated interactions in the design of bromodomain-selective BET inhibitors. *J. Mol. Graph. Model.* **81**, 197–210 (2018).
47. M. Brand *et al.*, Small molecule inhibitors of bromodomain-acetyl-lysine interactions. *ACS Chem. Biol.* **10**, 22–39 (2015).
48. C. L. Sutherland *et al.*, Identification and development of 2,3-Dihydropyrido[1,2-a]quinazolin-5(1H)-one inhibitors targeting bromodomains within the switch/sucrose nonfermenting complex. *J. Med. Chem.* **59**, 5095–5101 (2016).
49. P. Schmidtke, A. Bidon-Chanal, F. J. Luque, X. Barril, MDpocket: Open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics* **27**, 3276–3285 (2011).
50. O. B. Cox *et al.*, A poised fragment library enables rapid synthetic expansion yielding the first reported inhibitors of PHIP(2), an atypical bromodomain. *Chem. Sci.* **7**, 2322–2330 (2016).
51. L. R. Vidler, N. Brown, S. Knapp, S. Hoelder, Druggability analysis and structural classification of bromodomain acetyl-lysine binding sites. *J. Med. Chem.* **55**, 7346–7359 (2012).
52. M. Aldeghi *et al.*, Large-scale analysis of water stability in bromodomain binding pockets with grand canonical Monte Carlo. *Commun. Chem.* **1**, 19 (2018).
53. R. D. Finn *et al.*, Pfam: The protein families database. *Nucleic Acids Res.* **42**, D222–D230 (2014).
54. H. Wen *et al.*, ZMYND11 links histone H3.K36me3 to transcription elongation and tumour suppression. *Nature* **508**, 263–268 (2014).
55. K. J. Kohlhoff, P. Robustelli, A. Cavalli, X. Salvatella, M. Vendruscolo, Fast and accurate predictions of protein NMR chemical shifts from interatomic distances. *J. Am. Chem. Soc.* **131**, 13894–13895 (2009).
56. R. B. Best, K. Lindorff-Larsen, M. A. DePristo, M. Vendruscolo, Relation between native ensembles and experimental structures of proteins. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 10901–10906 (2006).
57. F. Yabukarski *et al.*, Assessment of enzyme active site positioning and tests of catalytic mechanisms through X-ray-derived conformational ensembles. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 33204–33215 (2020).
58. S. A. Forbes *et al.*, COSMIC: Mining complete cancer genomes in the catalogue of somatic mutations in cancer. *Nucleic Acids Res.* **39**, 945–950 (2011).
59. K. M. Hart *et al.*, Designing small molecules to target cryptic pockets yields both positive and negative allosteric modulators. *PLoS One* **12**, e0178678 (2017).
60. M. A. Cruz *et al.*, Discovery of a cryptic allosteric site in Ebola's 'undruggable' VP35 protein using simulations and experiments. *bioRxiv* [Preprint] (2020). <https://doi.org/10.1101/2020.02.09.940510>. Accessed 13 January 2021.
61. S. Olsson, H. Wu, F. Paul, C. Clementi, F. Noé, Combining experimental and simulation data of molecular processes via augmented Markov models. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 8265–8270 (2017).
62. S. Olsson, F. Noé, Mechanistic models of chemical exchange induced relaxation in protein NMR. *J. Am. Chem. Soc.* **139**, 210–210 (2017).
63. P. Eastman *et al.*, OpenMM 4: A reusable, extensible, hardware independent library for high performance molecular simulation. *J. Chem. Theory Comput.* **9**, 461–469 (2013).
64. J. A. Maier *et al.*, ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).
65. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
66. G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi, PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **185**, 604–613 (2014).
67. D. L. Theobald, P. A. Steindel, Optimal simultaneous superpositioning of multiple structures with missing data. *Bioinformatics* **28**, 1972–1979 (2012).