AI-Based Methods for Cryptic Pocket Detection Are Fast and Qualitative Compared to Quantitatively Predictive Simulations

Si Zhang,[1] Justin J. Miller,[1] Gregory R. Bowman[1, *]

[1]Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, PA 19104, USA

*Corresponding author: grbowman@seas.upenn.edu

Abstract

Artificial intelligence (AI) models have advanced rapidly, driving breakthroughs in protein structure prediction, functional annotation, and conformational exploration. Among these, molecular dynamics (MD)-inspired generative models such as AlphaFlow and BioEmu show strong potential for capturing conformational ensembles. In this study, we benchmark these models alongside physics-based MD simulations to evaluate their ability to detect cryptic pockets in proteins. Identifying such transient pockets remains a vital goal in drug discovery, as they can offer new avenues for targeting proteins traditionally challenging to modulate. We also assess two specialized residue-level predictors, PocketMiner and CryptoBank. Using the interferon inhibitory domain of Zaire Ebola VP35 (VP35), TEM-1 β-lactamse with the M182T substitution (TEM β-lactamase), and their mutants, we test whether each method can detect pockets and capture the effects of point mutations known to enhance or suppress pocket formation. All methods successfully identify pockets in VP35 and distinguish between opening and closing mutants. However, in TEM, where pocket opening is subtle, the methods perform inconsistently. These results highlight the promise of AI-based and simulation-based strategies in cryptic pocket discovery while pointing to the need for further improvements to achieve robust, system-independent predictions.

Introduction

Cryptic pockets are dynamic binding sites that are usually closed but open some fraction of the time due to thermal fluctuations of a protein's structure. They are often invisible in experimentally-derived structures, which resolve what a protein typically looks like rather than more rare states, making cryptic pockets difficult to capture experimentally[1,2]. Despite these challenges, identifying such elusive pockets can create new opportunities for drug discovery by revealing binding sites on

proteins historically deemed "undruggable", providing a means to achieve specificity, or enabling allosteric regulation of functional regions[3–11].

Recent advances in computational methods have made it increasingly feasible to discover and characterize cryptic pockets, particularly through molecular dynamics (MD) simulations[12–16]. For example, Rangari et al. used MD simulations to reveal an extended cryptic pocket in the CB1 receptor located between the orthosteric site and the conserved signaling residue $D^{2.50}$, which subsequently guided the design of a novel ligand[17]. Enhanced sampling techniques have further improved the detection of rare pocket-opening motions, especially those involving slow secondary structure rearrangements. Methods such as Fluctuation Amplification of Specific Traits (FAST)[18], Metadynamics[19,20], Sampling Water Interfaces through Scaled Hamiltonians (SWISH)[21,22] have proven particularly effective in this area. In recent studies, the combination of FAST with Markov state models (MSMs), a framework commonly used to map conformational landscapes sampled by MD simulations[23], successfully identified cryptic pockets in several challenging systems. These included the 5-HT_{3A} receptor, a member of the pentameric ligand-gated ion channel family[24], and VP35[25]. A follow-up study applying the same strategy to VP35 homologs from Reston and Marburg filoviruses further demonstrated that the open state of this pocket has a biological function, providing an exciting insight into its potential regulatory role[26]. In this study, we found that running FAST simulations followed by seeding additional simulations from evenly dispersed structures discovered by FAST gave reproducible populations of the open state. We refer to this procedure as FAST+seeding in this work to distinguish it from just using FAST simulations without any follow-up simulation to ensure good statistics across the landscape.

While MD simulations are powerful for detecting cryptic pockets, their effectiveness often depends on the specific system, the timescale of the relevant pocket motions, and the choice of simulation parameters[13,22]. The significant computational cost required to achieve adequate sampling can also limit their utility in large-scale or high-throughput studies.

Rapid advances in AI-based methods for structure prediction, such as AlphaFold, Chai-1, and Botz-1, suggest that AI has the potential to accelerate cryptic pocket discovery too[27–31]. For instance, by modifying the multiple sequence alignment (MSA) input, AlphaFold can generate

distinct structure variants that may reveal pocket formation directly or serve as starting points for further exploration[32–34]. However, AlphaFold was trained primarily on experimentally determined structures from the Protein Data Bank (PDB)[35] and large sequence database, without explicit training to sample alternative conformations. As a result, its ability to capture pocket formation across the global conformational landscape remains limited[32].

A number of new AI methods have been developed to facilitate broad and efficient exploration of conformational landscapes by using simulations during training. Once such approach, AlphaFlow, repurposes the AlphaFold architecture and incorporates 1.28 million experimentally solved PDB structures[36]. It is further fine-tuned under a custom flow matching framework using the ATLAS dataset, which includes all-atom MD simulations of 1,390 proteins, each performed in triplicate for 100 ns. The resulting model serves as a sequence-conditioned generative framework for sampling protein conformations. This approach achieves faster wall-clock convergence to equilibrium-like properties (e.g., root mean square fluctuation profiles) compared with replicate MD trajectories initiated from a static PDB structure.

Another generative model, BioEmu, was developed using a large and diverse dataset that combines 200 million AlphaFold-predicted structures, over 200 ms of MD simulation data, and approximately 500,000 experimentally measured protein stabilities, integrated through a novel multi-stage training pipeline[37]. Importantly, the BioEmu dataset also includes about 0.5 million mutant sequences, providing a valuable foundation for mutational and stability analyses. The final model accurately predicts equilibrium ensembles and folding free energies, while also capturing dynamic features such as local unfolding events and conformational flexibility in intrinsically disordered proteins.

Beyond generative models, several task-specific predictors like PocketMiner and CryptoBank have been developed for residue-level cryptic pocket detection[38–40]. These models are particularly attractive because they can produce predictions within seconds. One representative structure-based model is PocketMiner, a geometric vector perceptron (GVP)-based graph neural network. It was trained on a curated dataset constructed from 38 proteins containing 39 experimentally validated cryptic pockets[13], with more than 3,000 independent MD simulations, each lasting at least 40 ns.

In total, around 0.9 million residues with simulation-derived labels were used to train the model. This enables PocketMiner to predict whether each residue in a given protein structure is likely to participate in cryptic pocket formation during an MD simulation initiated from that structure. PocketMiner achieved strong performance on the test set (ROC-AUC: 0.87) and is highly computationally efficient. A recently released sequence-based predictor, CryptoBank, was trained on 71 cryptic and 128 non-cryptic cases[41]. It attains high precision (PR AUC: 0.8) when query sequences share more than 20% identity with the CryptoBank entries.

Although these approaches have shown impressive results, a systematic benchmark comparing their strengths and limitations in cryptic pocket detection is still lacking. BioEmu has included a preliminary assessment on 34 wild-type (WT) proteins, but more in-depth evaluations remain unexplored. In this work, we aim to examine the performance of both AI-based and physics-based approaches in identifying cryptic pockets and probing mutational effects, using VP35 and TEM β-lactamase as model systems (**Figure 1**). These proteins have been extensively characterized experimentally, including detailed studies on mutation-induced conformational changes, making them well-suited for comparative analysis[9,25,26,42–46]. Specifically, we evaluate how well each method can distinguish whether mutants lead to pocket opening or closing, quantify the extent of pocket formation, and capture structural differences among mutants. Through this comparison, we highlight the current limitations of these approaches and underscore the need for further improvements.

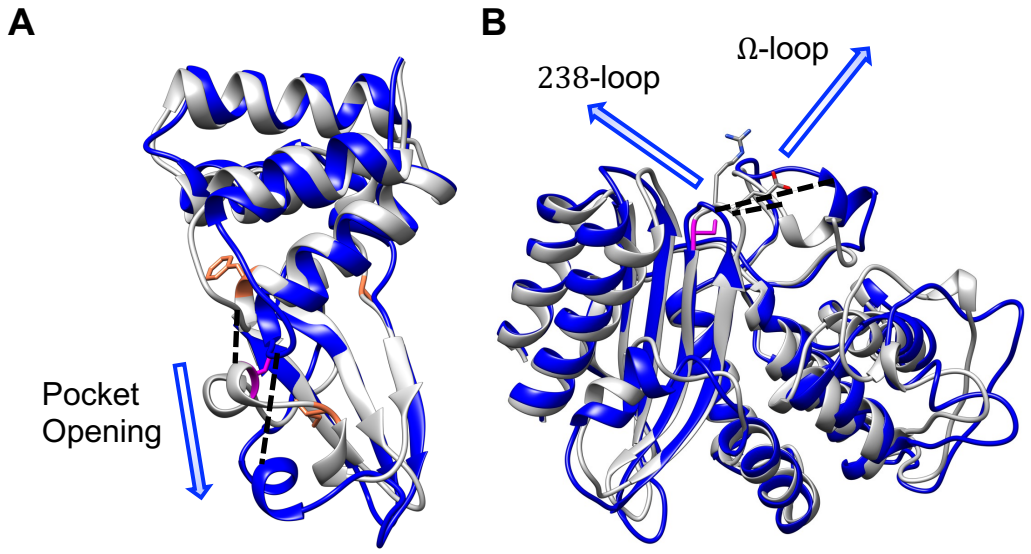

Figure 1. Cryptic pocket opening in VP35 and TEM β-lactamase. A. Crystal structure of VP35 (gray; chain A, PDB: 3L25) overlaid with a representative open conformation (blue). The Cα-Cα distance between residues G236 and A306 expands from 5.9 Å (crystal) to 12.1 Å (open), indicated by black dash lines. Residue C307, experimentally labeled and shown to undergo changes in solvent exposure associated with cryptic pocket formation, is colored in magenta in the crystal structure. Mutation sites F239, I303, and A291 are indicated with orange stick representations. **B.** The crystal structure of TEM β-lactamase (gray; PDB: 1JWP) is shown with an open conformation overlaid in blue. The Ω-loop (residue 165-175) and the adjacent 238-loop (residues 237-244) are indicated with arrows. The Cα-Cα distance between E171 and E240 increases from 5.4 Å in the crystal structure to 11.2 Å in the open conformation (black dashed line). Residue S243, mutated to cysteine in thiol labeling assays and associated with increased solvent exposure in the open state, is shown in magenta. Mutation sites E240 and R241 are displayed as sticks.

Results

AI-Based and Simulation-Based Methods Successfully Predict Whether Mutants Open or Close Pockets.

To evaluate whether computational approaches can reliably detect cryptic pocket formation and predict the effects of mutations in pocket opening, we first focus on VP35 and its mutants, F239A, I303A and A291P (**Figure 1A**). In our previous study[25], we identified a cryptic pocket in VP35 that forms when a small helix (helix 5) separates from an adjacent four-helix bundle, exposing residue C307. This pocket opening can be captured experimentally using thiol labeling assays, which quantify changes in the solvent accessibility at C307. This study, together with a recent work, demonstrates that mutants F239A and I303A enhance cryptic pocket opening, whereas A291P exhibits a distinct effect by significantly reducing pocket formation[25,47]. These findings provide a clear experimental basis for benchmarking computational predictions across both WT and mutant systems. Additionally, identifying which mutants enhance or suppress pocket opening can help prioritize candidates for further drug discovery or detailed structural studies.

We tested a range of computational methods, including two physics-based simulations (FAST and FAST+seeding MD) and four AI-based models (AlphaFlow, BioEmu, PocketMiner, and CryptoBank). For methods that output structural ensembles (MD and generative AI), we defined open vs. closed states using a Cα–Cα distance between residues G236 and A306, applying a threshold of 1.0 nm. For MD-based simulations, population distributions were calculated from MSMs. For generative AI models, we calculated the percentage of structures exceeding the threshold directly. PocketMiner and CryptoBank predict per-residue opening probabilities rather than full structures, so we compared those scores directly across systems. This unified framework allows a fair comparison of methods with different outputs and sampling strategies (**Error! Reference source not found.**).

Among the AI-based approaches, AlphaFlow struggles to identify open pockets while BioEmu consistently samples a higher fraction of open conformations across VP35 systems. The predicted open-state populations from BioEmu increase from 10.4 ± 0.2% in WT to 12.6 ± 0.4% in F239A and 18.4 ± 0.4% in I303A, with little change observed for A291P (10.2 ± 0.4%) (**Figure 2**, **Table 1**). These values were averaged from three independent runs initiated from different random seeds (**Error! Reference source not found.**). AlphaFlow, on the other hand, samples fewer than 1% open conformations in all systems. Only two slightly open structures were observed for F239A and one for I303A (**Figure 3A, B**), with similar results across all three replicates (**Error! Reference source not found.**).

To test whether AlphaFlow's performance could improve with more sampling, we increased the number of conformations from 250 to 10,000 per system. Surprisingly, the open-state fraction decreased further to less than 0.01%, despite a slight increase in the absolute number of open structures (**Error! Reference source not found.**). This result suggests that AlphaFlow struggles to quantitively detect rare pocket-opening events. The computational cost also differs significantly. A 10,000-sample AlphaFlow job required about two days per system. In comparison, BioEmu generated the same number of samples in approximately 100 minutes, while AlphaFlow needed 70 minutes for its original 250-sample run.

These results likely stem from each model's training strategy and scope. AlphaFlow, which integrates structural prediction and short-timescale dynamics sampling, may be limited by its relatively short simulation lengths and the limited protein diversity in its training set. This could constrain its ability to capture broader or rarer conformational changes. BioEmu incorporates more extensive structural, simulation, and experimental data, allowing it to model equilibrium ensembles more effectively and better reflect conformational flexibility across diverse proteins.

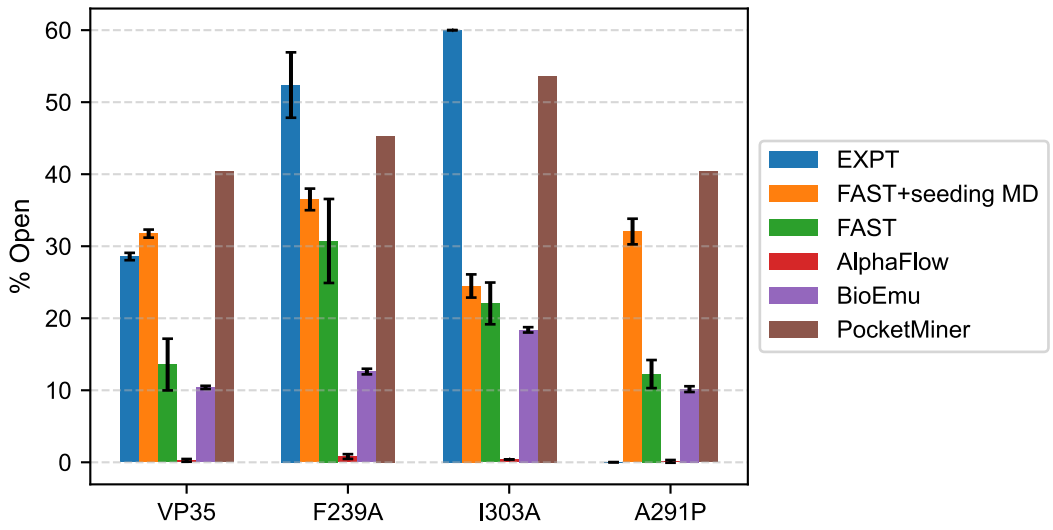

Figure 2. Analysis of cryptic pocket opening in VP35 and its mutants. Computational analysis of cryptic pocket formation in VP35 and three variants (F239A, I303A, and A291P). Open-state populations were quantified using multiple approaches. Experimental values (EXPT) are derived from equilibrium constants measured by thiol labeling assays[25]. Populations from FAST and FAST+seeding MD simulations were calculated by identifying conformations with the Cα-Cα distance between G236 and A306 exceeding 1.0 nm using MSMs. Uncertainties were estimated from a per-trajectory bootstrap procedure. Results from AlphaFlow and BioEmu represent the fraction of sampled conformations meeting the same distance criterion, with uncertainties estimated from three replicate simulations per system. PocketMiner predictions indicate the probability of C307 participating in a cryptic pocket.

Table 1. Numerical values of open-state populations and associated uncertainties for VP35 and its mutants (F239A, I303A, and A291P) derived from experimental measurements and computational methods. Experimental values (EXPT) are inferred from thiol labeling experiments[25]. Simulation-based results from FAST and FAST+seeding MD are calculated using MSMs, with open conformations defined by a Cα-Cα distance between G236 and A306 greater than 1.0 nm. AlphaFlow and BioEmu results report the fraction of conformations meeting this criterion across three replicas. PocketMiner and CryptoBank values correspond to predicted probabilities that C307, labeled in experiments, experiences a solvent exposure change associated with cryptic pocket opening.

	Cryptic Pocket Open Population (%)						
VP35 Variant	EXPT	FAST+seeding MD	FAST	AlphaFlow	BioEmu	PocketMiner	CryptoBank
WT	28.6 ± 0.5	31.8 ± 0.6	13.6 ± 3.6	0.27 ± 0.19	10.4 ± 0.2	40.4	7.9
F239A	52.4 ± 4.5	36.5 ± 1.5	30.7 ± 5.8	0.8 ± 0.33	12.6 ± 0.4	45.3	-
I303A	> 60.	24.5 ± 1.6	22.1 ± 2.9	0.4 ± 0.	18.4 ± 0.4	53.6	-

A291P	0.01 ± 0.02	32.0 ± 1.8	12.3 ± 2.0	0.13 ± 0.19	10.2 ± 0.4	40.4	-

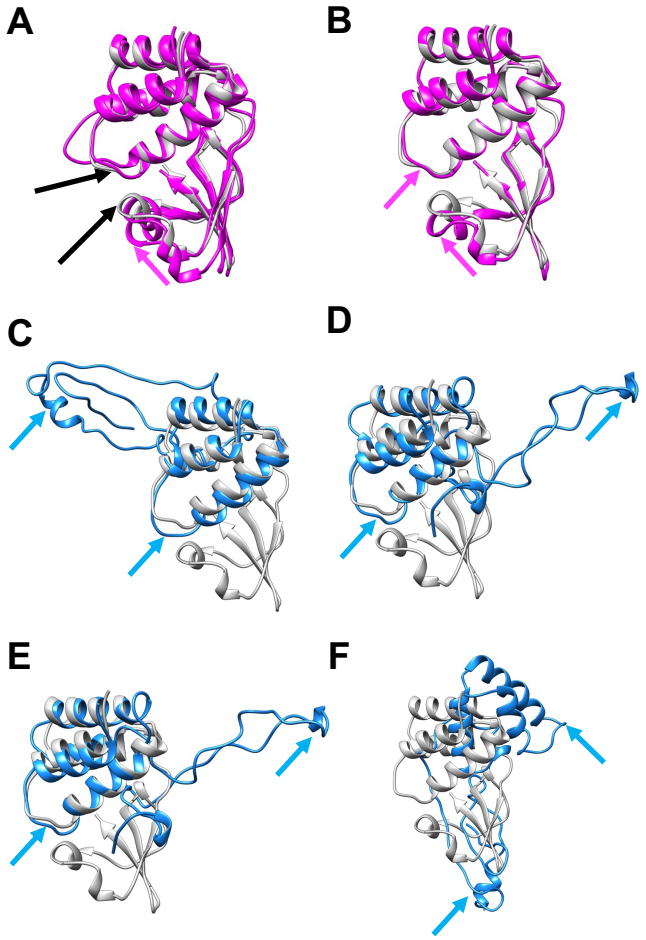

Figure 3. Structural comparison of sampled open conformations with the crystal structure of VP35 highlight that AlphaFlow only captures partial opening while BioEmu sometimes generates implausible structures. Representative open conformations generated by AlphaFlow and BioEmu are overlaid with the VP35 crystal structure (gray) for visual comparison. Residue G236 and A306, which are used to define open versus closed conformations, are indicated by black arrows in panel A. The top row shows the only three open conformations sampled by AlphaFlow: two from F239A (magenta, panel A) and one from I303A (magenta, in B). Magenta arrows highlight the positions of G236 and A306 associate with pocket opening. The bottom rows representative conformations sampled by BioEmu (blue). Panels C-F display the structures with the greatest deviation from the conformational ensembles of WT, F239A, I303A, and A291P, respectively. Blue arrows indicate the positions of G236 and A306.

To benchmark MD-based sampling, we applied the FAST adaptive sampling algorithm, previously shown to accelerate cryptic pocket discovery by guiding trajectories toward user-defined features[13,24,48]. Using this approach, combined with MSM analysis, we observed higher populations of open conformations in F239A and I303A (30.7 ± 5.8% and 22.1 ± 2.9%,

respectively), along with a moderate decrease in A291P (12.3 ± 2.0%), compared to 13.6 ± 3.6% in WT (**Figure 2**, **Table 1**).

To further enhance sampling, we conducted FAST+seeding MD simulations, launching three 40-ns replicas from 1,000 cluster centers per system (120 μs total) (See Methods). Seeding simulations from FAST runs in this manner helps to ensure most of the landscape observed is sampled well, thereby countering the intentional bias in where data is collected during adaptive sampling. These simulations captured enhanced pocket opening in F239A (36.5 ± 1.5%) and captured WT behavior well (31.8 ± 0.6%, compared to the experimental 28.6 ± 0.5%) (**Figure 2**, **Table 1**). A291P displayed a similar level of opening as WT (32 ± 1.8%), suggesting that the local structural rigidity introduced by the proline substitution may not significantly impact pocket dynamics over the current simulation timescale. Interestingly, I303A yielded a notably lower predicted open-state population (24.5 ± 1.6%) relative to the experimental value (> 60%). The discrepancy may arise from insufficient sampling or reflect a distinct conformational mechanism in this mutant that is not fully captured within our current simulation window.

Finally, we evaluated PocketMiner[13], which predicts residue-level pocket-opening probabilities from static structures. The model correctly identified increased opening propensity at F239A and I303A, and no significant change in A291P, aligning with experiments (**Table 1**). CryptoBank[41], a sequence-based predictor, could only be applied to the WT system due to its input requirement (PDB ID). It predicted a low normalized pocket-opening probability of 0.079, suggesting limited sensitivity for this system.

AI-Based Models Underestimate Relevant Open Conformations and Yield Physically Unrealistic Samples.

Although all tested methods captured the overall trend of increased pocket opening from WT to VP35 mutants, the sampled conformational landscapes varied substantially across different ensemble generation approaches. To better understand these differences, we constructed two-dimensional histograms of sampled conformations from each method: FAST+seeding MD, FAST, AlphaFlow, and BioEmu. Each plot maps the inter-residue distance between G236 and A306 (used

to define open and closed conformations) against the solvent-accessible surface area (SASA) of residue C307 (**Figure 4**).

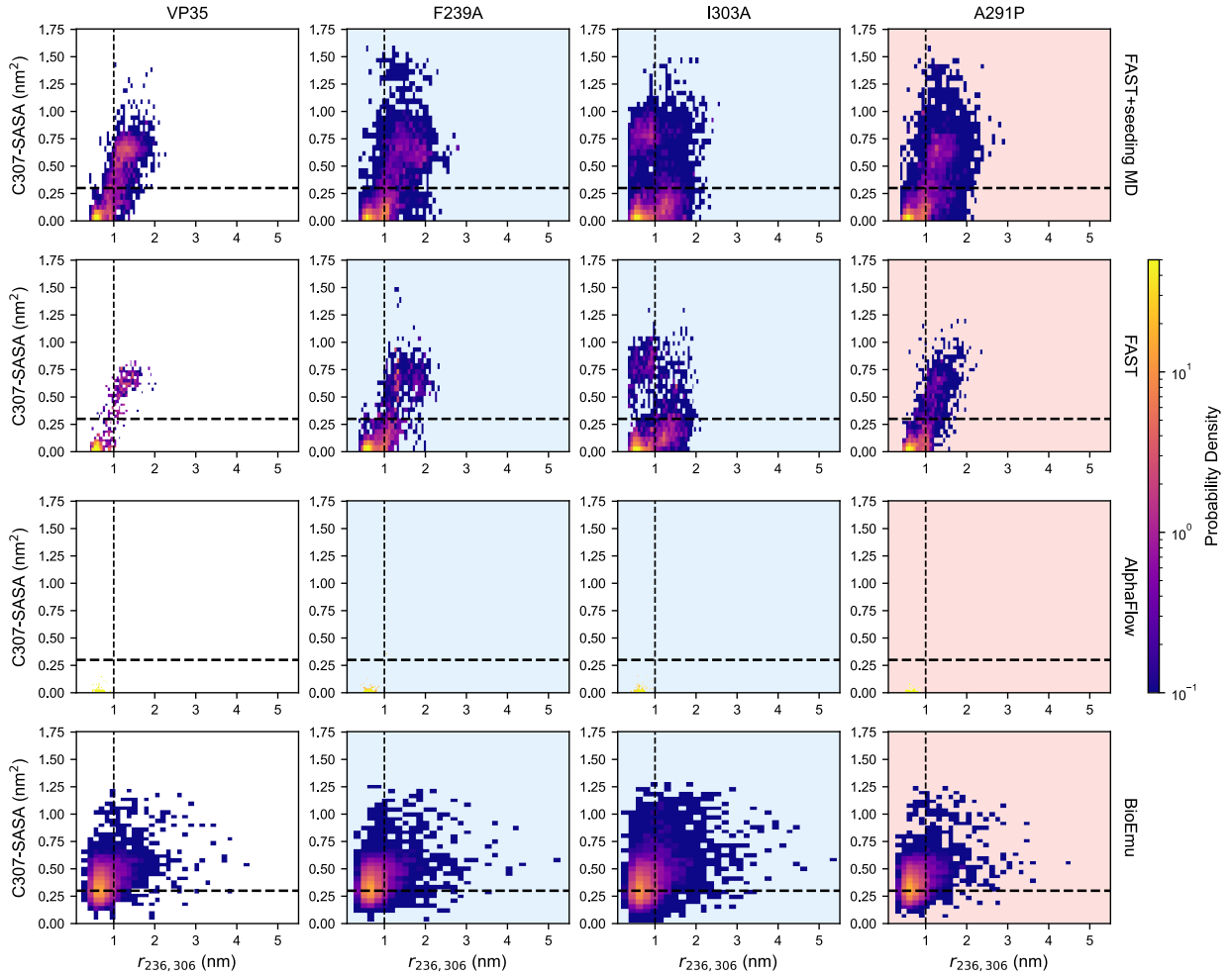

Figure 4. Two-dimensional histograms of sampled conformations generated by each computational method. Each panel shows the inter-residue distance between G236 and A306 (x-axis) plotted against the SASA of residue C307 (y-axis) for VP35 and its three mutants: F239A, I303A, and A291P. Rows correspond to different methods: FAST+seeding MD (top), FAST (second), AlphaFlow (third), and BioEmu (bottom). Background shading highlights whether a mutant opens more or less than WT: light blue indicates variants that open more than WT (F239A and I303A), whereas light pink marks the variants that open less or tend to remain closed (A291P). Logarithmic normalization (LogNorm) is applied to highlight both high- and low-density regions. Dashed lines at x = 1.0 nm and y = 0.3 nm^2 indicate the distance threshold for pocket opening and (partial) solvent exposure of C307, respectively.

AlphaFlow produced conformations largely similar to the crystal structure, with minimal sampling of open states. Only F239A and I303A exhibit sparse sampling of open states, with just two and one slightly open conformations, respectively (**Figure 3A, B**). WT and A291P display no detectable pocket opening. BioEmu generates broader ensembles across all systems (**Figure 4**). Although the 2D histograms do not reveal dramatic visual differences across systems, I303A shows a notably higher number of conformations with increased SASA at residue C307. This may

be attributed to the I303A mutation, which replaces a bulky hydrophobic isoleucine with a smaller alanine. The substitution likely increases the flexibility of the surrounding loop, particularly near the mobile short helix 5, even in closed states where the distance remains below the 1.0 nm threshold.

A subset of BioEmu-generated conformations showed extreme structural deviations not observed in other methods. Across all systems, a small fraction (4-15%) of BioEmu samples had G236-A306 distances exceeding 3.5 to 6 nm (**Error! Reference source not found.**). By contrast, the maximum distance observed in the seeded MD simulations was only 3.0 nm. Structural inspection of the most extreme outliers (**Figure 3C-F**) revealed that the three parallel β-sheets were completely unfolded and extended outward into the solvent. We deemed these structures unrealistic as they are incongruent with other structural data, both computational methods and experiments that would likely report on such conformations if they were really so well populated.

Simulation-based methods (i.e., FAST+seeding MD and FAST) also revealed greater conformational flexibility within the closed-state ensemble of I303A (**Figure 4**). This was especially evident in the seeded MD, which sampled a denser and broader range of conformations within the closed state, yet with increased solvent exposure. Such enhanced flexibility and partial exposure may explain the discrepancy between the predicted and experimental open-state populations in I303A (**Table 1**). For F239A and A291P, the FAST+seeding MD results resembled those of the WT, primarily exploring open and solvent-exposed conformations while showing limited sampling closed but exposed states. This suggest that F239A and A291P maintain pocket dynamics more like WT, whereas I303A may undergo alternative or more complex conformational transitions that are not fully captured in our current simulations.

Taken together, all methods distinguished open from closed mutants, but with varying levels of resolution and sampling depth. AI models offer rapid predictions and initial assessments in seconds (PocketMiner or CryptoBank) to hours (AlphaFlow or BioEmu), although they are less effective at resolving detailed structural differences. In contrast, both FAST+seeding MD and FAST sampled the conformational landscapes more thoroughly, capturing the structural features associated with pocket opening. Notably, the FAST simulations, with a total sampling time of 8

μs per system, achieved comparable insights to the 120 μs of FAST+seeding MD simulations, underscoring the efficiency of goal-oriented adaptive sampling.

Computational Methods Diverge in Capturing Weak Pocket Opening in TEM

To further evaluate the predictive performance of various ensemble generation and pocket detection methods, we tested their ability to detect cryptic pocket opening in TEM variants. TEM is known to contain a Ω-loop cryptic pocket, characterized by motions in the Ω-loop (residue 165-175) and the adjacent 238-loop (residues 237-244) (**Figure 1B**). This pocket exhibits only weak opening in WT, making TEM a valuable test case for assessing each method's ability to detect subtle cryptic conformational changes.

To quantify pocket opening, we used the Cα–Cα distance between residues E171 and E240 as the opening metric, with a threshold of 1.0 nm to define open versus closed states. Experimental thiol labeling measurements served as ground truth for evaluating model performance[43]. We mainly compared predictions from MD (WT only, conventional trajectories without FAST+seeding), FAST, AlphaFlow, BioEmu, PocketMiner, and CryptoBank (**Error! Reference source not found.**).

Both AlphaFlow and BioEmu predicted open-state populations for WT reasonably close to experiment. AlphaFlow yielded 0.13 ± 0.19% open conformations, and BioEmu gave 2.2 ± 0.1%, compared to the experimental value of 1.1 ± 0.2% (**Figure 5**, **Table 2**). FAST simulations, however, produced a higher open-state population of 9.9 ± 1.5%. This overestimation likely arises from FAST's objective-driven sampling strategy, which preferentially directs exploration toward pocket-opening conformations within limited simulation time. To provide a more direct comparison, we reanalyzed previously reported long-timescale conventional MD simulations (~90.6 μs) initiated from the crystal structure[44] (See Methods). Our analysis reproduced the experimental value more closely (0.56 ± 0.3%), supporting the consistency of physics-based sampling when sufficient statistics are available.

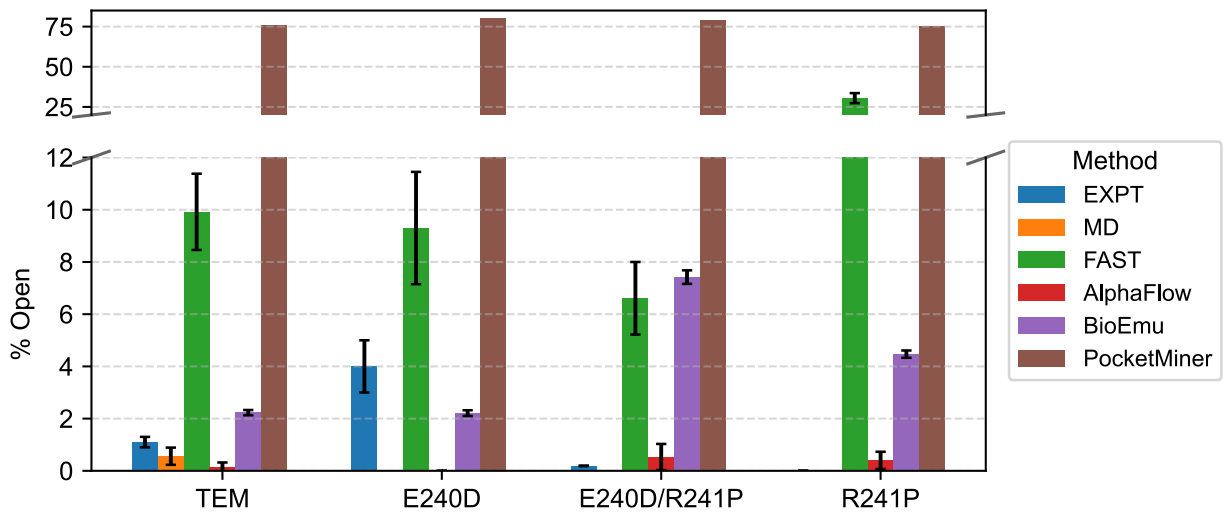

Figure 5. Analysis of cryptic pocket opening in TEM and its mutants. Computational analysis of cryptic pocket formation in TEM and three variants (E240D, E240D/R241P, and R241P). Open-state populations were quantified using multiple approaches. Experimental values (EXPT) are taken from Ref.[43]. Populations from MD and FAST simulations were calculated by identifying conformations with the Cα-Cα distance between E171 and E240 exceeding 1.0 nm using MSMs. Uncertainties were estimated from a per-trajectory bootstrap procedure. Results from AlphaFlow and BioEmu represent the fraction of sampled conformations meeting the same distance criterion, with uncertainties estimated from three replicate simulations per system. PocketMiner predictions indicate the probability of S243 participating in a cryptic pocket.

Table 2. Numerical values of open-state populations and associated uncertainties for TEM and its mutants (E240D, E240D/R241P, and R241P), derived from both experimental measurements and computational methods. Experimental values (EXPT) are taken from Ref.[43]. Simulation-based results from MD and FAST are calculated using MSMs, with open conformations defined by a Cα-Cα distance between E171 and E240 greater than 1.0 nm. AlphaFlow and BioEmu results report the fraction of conformations meeting this criterion across three replicate simulations. PocketMiner values reflect the predicted probability that residue S243 (mutated to cysteine and labeled in thiol labeling assays) undergoes a change in solvent exposure consistent with cryptic pocket opening.

	Ω-loop Pocket Open Population (%)						
TEM Variant	EXPT	MD	FAST	AlphaFlow	BioEmu	PocketMiner	CryptoBank
WT	1.1 ± 0.2	0.56 ± 0.3	9.9 ± 1.5	0.13 ± 0.19	2.2 ± 0.1	76.3	2.4
E240D	4 ± 1	-	9.30 ± 2.2	0.0 ± 0.0	2.2 ± 0.1	80.3	-
E240D/R241P	0.19 ± 0.01	-	6.61 ± 1.4	0.53 ± 0.5	7.4 ± 0.3	79.0	-
R241P	0.002 ± 0.005*	-	30.5 ± 3.1	0.40 ± 0.33	4.5 ± 0.1	75.0	-

*Determined to not have an Ω-loop pocket as the pocket populations is less than the unfolded population.

Subtle changes in flexibility proved harder to capture. For example, in the case of E240D, which shortens the side chain from glutamate to aspartate, none of the tested methods effectively detected increased pocket dynamics, including FAST. For the E240P/R241P double mutant, AI-based models incorrectly predicted enhanced opening, while FAST reproduced the experimental trend. For R241P, which experimentally exhibits an extremely lower pocket-opening rate than its

unfolding rate (suggesting the absence of a stable cryptic pocket), all methods struggled. FAST significantly overpredicted the open-state population. Among the AI-based predictors, PocketMiner captured the general trend across TEM variants but showed minimal variation between WT and mutants. This suggests that while the model reflects the overall pattern, it may lack the sensitivity to these subtle structural rearrangements. CryptoBank, which only provides predictions for WT due to input format limitations, yielded a very low normalized pocket-opening probability of 0.024, consistent with a predominantly closed conformation.

To further investigate the sampling behavior of each method, we visualized the conformational ensembles generated for TEM and its mutants (**Figure 6**), using the same 2D projection scheme applied in the VP35 analysis. Each plot maps the E171-E240 inter-residue distance (x-axis) against the SASA of S243 (y-axis), a proxy for pocket exposure. Notably, residue S243 was mutated to cysteine and labeled in previous thiol labeling assays[43]. As observed for VP35, AlphaFlow generated a limited conformational ensemble across all three replicates (**Error! Reference source not found.**). Most sampled structures remained near the closed, crystal-like state, with only a few minor pocket-opening conformations observed among all variants (**Figure 7A-C**). This again suggests that AlphaFlow struggles to capture rare events, particularly when the opening is subtle and infrequent. BioEmu sampled a broader range of conformations but also produces a small number of largely unfolded or extended structures, with E171-E240 distances exceeding ~4.0 nm (**Figure 7D-F, Error! Reference source not found.**). These conformations were not found in any MD-based methods. In contrast, FAST explored a more continuous and physically interpretable conformational landscape, effectively targeting the pocket-opening region. These simulations captured intermediate states with partial pocket opening, especially in the double mutant E240P/R241P. However, their tendency to overestimate open-state populations for weakly opening variants (e.g., R241P) underscores a tradeoff between sampling efficiency and quantitative accuracy.

These results highlight the challenges of detecting weak and transient cryptic pockets, particularly when such openings are rare or short-lived. While AI-based methods offer computational speed and broad ensemble coverage, they may overlook subtle conformational rearrangements. Conversely, physics-based simulations such as FAST and conventional MD can capture these

transitions when sufficient sampling is achieved, though they risk overestimation when biasing features dominate.

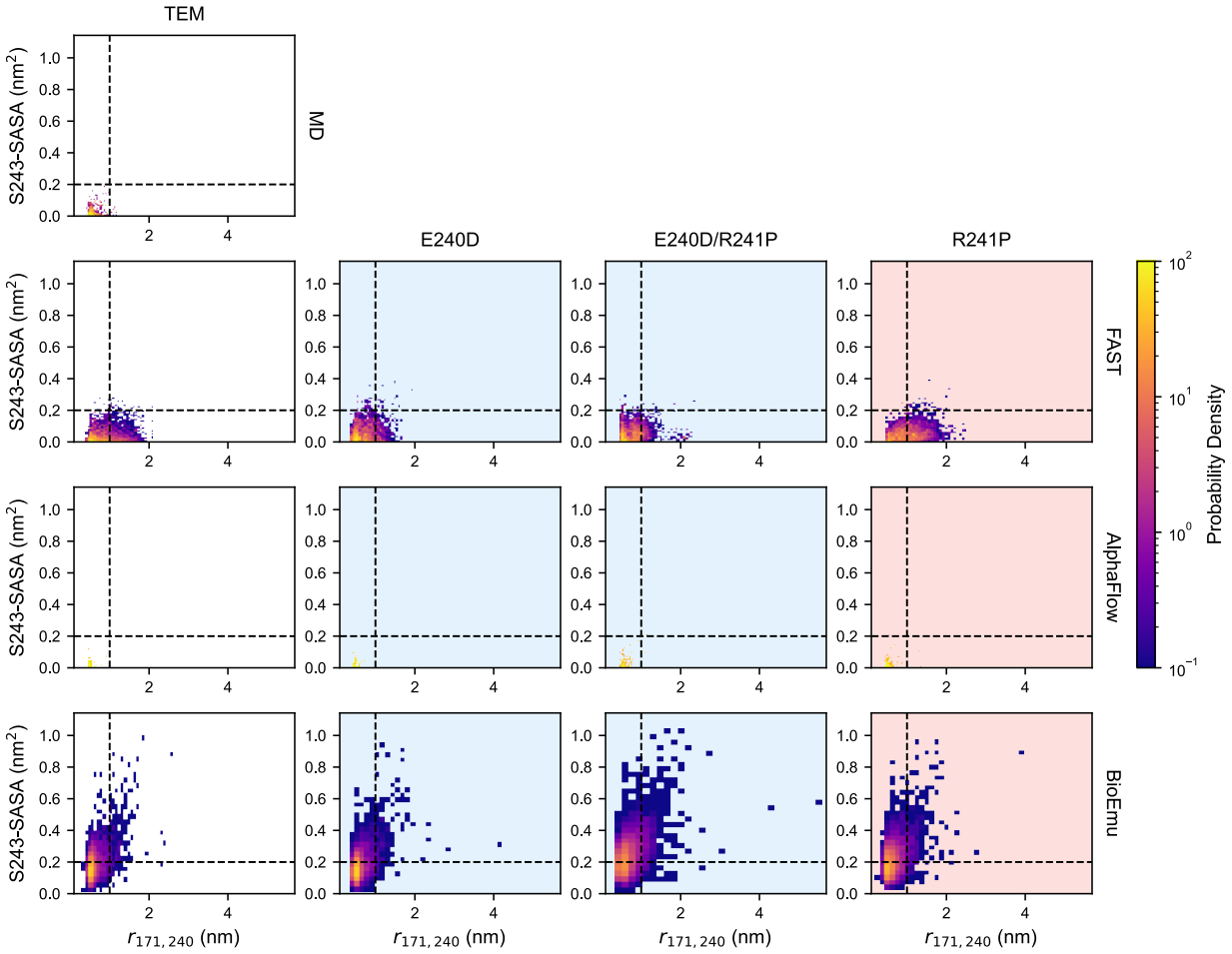

Figure 6. Two-dimensional histograms of sampled conformations from each computational ensemble generation method. Each plot shows the inter-residue distance between residues E171 and E240 (x-axis) versus the SASA of residue S243 (y-axis) for TEM and its three mutants: E240D, E240D/R241P, and R241P. Rows represents different methods: MD (top), FAST (second), AlphaFlow (third), and BioEmu (bottom). Only WT MD simulations initiated from the crystal structure are shown. Background shading highlights whether a mutant opens more or less than WT: light blue indicates variants with increased pocket opening (E240D and E240D/R241P), whereas light pink marks the variant showing no detectable pocket opening (R241P). A logarithmic normalization (LogNorm) is applied to highlight both high- and low-density regions. Dashed lines at x = 1.0 nm and y = 0.2 nm^2 are included in each subplot to indicate the distance threshold for pocket opening and (partial) solvent exposure of S243, respectively.

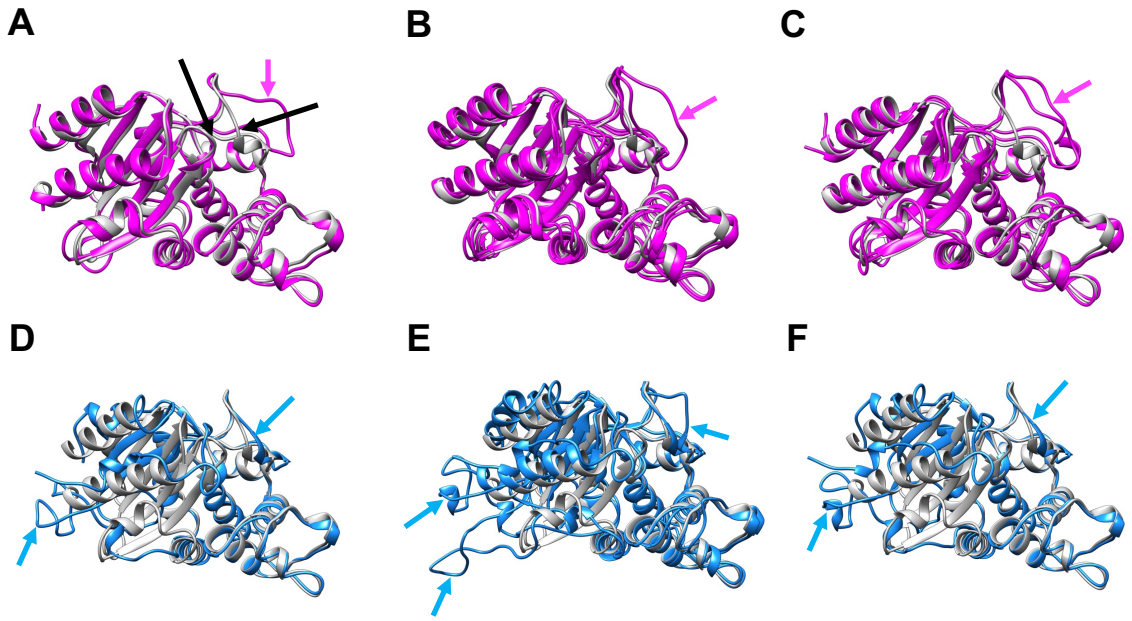

Figure 7. Structural comparison of sampled open conformations with the crystal structure of TEM. Representative open conformations generated by AlphaFlow and BioEmu are overlaid with the TEM crystal structure (gray) for visual comparison. Residue E171 and E240, which define open versus closed conformations, are indicated by black arrows in panel A. The top row shows the only six open conformations sampled by AlphaFlow: one from WT (magenta, panel A), three from E240D/R241P (magenta, panel B), and two from R241P (magenta, panel C). Magenta arrows highlight the position of E171, which is associated with Ω-loop pocket opening. The bottom row shows BioEmu-sampled conformations (blue) with inter-residue distances greater than ~4.0 nm. Panels D-F display these highly open structures from E240D, E240D/R241P, and R241P, respectively. Blue arrows indicate the positions of E171 and E240.

Discussion

We systematically benchmarked AI-based and physics-based approaches for identifying cryptic pockets and evaluating mutational-induced effects on pocket opening in two distinct systems: VP35 and TEM β-lactamase. The methods tested included goal-oriented adaptive sampling (FAST), FAST+seeding MD, and recent AI models such as AlphaFlow, BioEmu, PocketMiner, and CryptoBank. While prior studies have offered some comparative insights in WT proteins[49], detailed, side-by-side evaluations across multiple mutations have been lacking.

Our results show that all methods could distinguish between mutants in VP35 known to favor open versus closed pocket opening, though with varying levels of resolution and sampling depth. AI models offered rapid predictions, with outputs generated in seconds (PocketMiner and CryptoBank) to a few hours (AlphaFlow and BioEmu). These tools are useful for fast initial screening and provide qualitative insights, but they were generally less sensitive to subtle structural differences

across variants. In addition, BioEmu frequently generated conformations that appeared physically implausible and were not observed with the other methods. In contrast, simulation-based, i.e., FAST and FAST+seeding MD, sampled the conformational landscapes more comprehensively, capturing key features linked to pocket formation.

In VP35, FAST simulations achieved comparable insights using just 8 μs of total sampling per system, in contrast to the 120 μs required by FAST+seeding MD, underscoring the efficiency of adaptive sampling. However, in system like TEM, where pocket opening is very weak, all methods showed limited ability to fully recapitulate the experimental observations.

These findings highlight the complementary strengths of AI- and physics-based approaches. We anticipate that such comparative studies will guide the strategic integration of both methods in future efforts toward cryptic pocket discovery and drug design. Moreover, our results point to areas needing improvements, such as enhancing the structural realism of AI-generated conformations.

Methods

Molecular Simulations

System preparation. The protein structures of VP35 (chain A, PDB ID: 3FKE) and TEM (PDB ID: 1JWP) used in this study were obtained from the Protein Data Bank (PDB)[35]. Point mutations, including single and double substitutions, were introduced using PyMOL[50].

Each protein system was solvated in a rhombic dodecahedral box containing TIP3P water molecules[51], ensuring a minimum distance of 1.0 nm between any protein atom and the box boundary. System topologies were generated with the Amber03 force field[52] in GROMACS[53]. To neutralize the system, Na^+ and Cl^- counterions were added to achieve a salt concentration of 0.1 mol/liter. Energy minimization was carried out using the steepest descents algorithm with an initial step size of 0.01 nm and terminized when the maximum force fell below 100 $kJ \cdot mol^{-1} \cdot nm^{-1}$ or after 500 steps. Equilibration followed for 0.1 ns using a 2 fs time step. All bonds were constrained using the LINCS algorithm[54] and long-range electrostatic interactions were treated with the Particle Mesh Ewald (PME) method[55] with a Fourier spacing of 0.12 nm. A cutoff of 0.9 nm was applied for both Coulomb and van der Waals interactions. The system temperature was maintained

at 300 K using the velocity-rescaling thermostat[56]. During both equilibration and production runs, Newton's equations of motions was integrated with the leap-frog algorithm.

FAST simulations. We employed the FAST algorithm[18], which iteratively guides unbiased MD simulations toward conformations that maximize or minimize a chosen order parameter. For VP35 and TEM (both WT and mutant systems), FAST simulations were conducted using pocket volume, calculated by the LIGSITE pocket detection algorithm[57], as the feature to be maximized. Following the setup described in previous work[25], each system underwent 10 rounds of adaptive sampling, with 10 parallel trajectories per round, each trajectory running for 80 ns, resulting in a total sampling time of 8 μs per system.

For analysis, the features were defined as the SASA of each residue's side chain, calculated using the Shrake-Rupley algorithm[58] implemented in MDTraj[59], with a drug-sized probe (2.8 Å sphere). Conformations were clustered based on per-residue SASA values using a hybrid k-centers/k-medoids algorithm implemented in Enspara[60], applying a 2.7 Å distance cutoff for VP35 and 3.0 Å for TEM, followed by five rounds of k-medoids updates. A Markov time of 2 ns was selected based on implied timescales convergence to compute populations with MSMs. The resulting cluster centers were subsequently used to generate the 2D conformational plots shown in the main text.

FAST+seeding MD simulations. After completing the FAST simulations, conformations were clustered based on Cα RMSD, and 1000 representative structures were used as starting points for additional MD simulations. For each cluster center, three independent simulations were launched, resulting an aggregate simulation time of 120 μs. For the WT of VP35, the simulations were previously performed on the Folding@home distributed computing platform[61,62] and reported in our earlier study[25]. The VP35 mutant simulations were newly generated on our local high-performance computing cluster, following the same simulation protocol. A corresponding MSM analysis was then carried out for each mutant to extract populations and conformational distributions, using a 3.0 Å distance cutoff and a Markov time of 5 ns (**Error! Reference source not found.**).

For TEM, we reanalyzed previously reported MD trajectories initiated from the crystal structure[44] to ensure consistency with our current analysis framework. A Markov time of 25 ns was used for the MSM population calculations (**Error! Reference source not found.**).

Experimental Measurements

The experimentally measured results for VP35 and its mutants used in this study were obtained from Ref.[25]. For each system, the open-state population was calculated based on the experimentally determined equilibrium constants describing the exposure of residue C307, using a two-state model.

For TEM, the open-state populations were directly taken from Ref.[44]. Detailed descriptions of the experimental procedures can be found in the respective references.

Data availability

The AlphaFlow- and BioEmu-generated data, along with all associated analysis files, have been deposited in the Open Science Framework (OSF) database at https://osf.io/vtx4h. Referenced structures include PDB IDs 3FKE and 1JWP. The starting WT structures used for FAST and MD simulations, as well as the resulting MSM data, are also available in the same directory. Additional FAST and MD simulation trajectories and related files can be provided upon reasonable request.

Code availability

FAST, Enspara, and PocketMiner software packages are freely available on GitHub at https://github.com/bowman-lab/fast, https://github.com/bowman-lab/enspara, and https://github.com/Mickdub/gvp.

Acknowledgements

This work was funded by the National Institutes of Health through NIGMS R35GM152085, and NSF MCB 2218156.

Competing interests

GRB consults for VidaVinci and Delphia Therapeutics. The remaining authors declare no competing interests.

REFERENCES

1. Fraser, J. S. *et al.* Accessing protein conformational ensembles using room-temperature X-ray crystallography. *Proc. Natl. Acad. Sci. U. S. A.* 108, 16247–16252 (2011).
2. Garman, E. 'Cool' crystals: macromolecular cryocrystallography and radiation damage. *Curr. Opin. Struct. Biol.* 13, 545–551 (2003).
3. Ostrem, J. M., Peters, U., Sos, M. L., Wells, J. A. & Shokat, K. M. K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions. *Nature* 503, 548–551 (2013).
4. Vajda, S., Beglov, D., Wakefield, A. E., Egbert, M. & Whitty, A. Cryptic binding sites on proteins: definition, detection, and druggability. *Curr. Opin. Chem. Biol.* 44, 1–8 (2018).
5. Arkin, M. R. & Whitty, A. The road less traveled: modulating signal transduction enzymes by inhibiting their protein–protein interactions. *Curr. Opin. Chem. Biol.* 13, 284–290 (2009).
6. Hardy, J. A. & Wells, J. A. Searching for new allosteric sites in enzymes. *Curr. Opin. Struct. Biol.* 14, 706–715 (2004).
7. Lu, S., Ji, M., Ni, D. & Zhang, J. Discovery of hidden allosteric sites as novel targets for allosteric drug design. *Drug Discov. Today* 23, 359–365 (2018).
8. Nussinov, R. & Tsai, C. J. Allostery in Disease and in Drug Discovery. *Cell* 153, 293–305 (2013).
9. Hart, K. M. *et al.* Designing small molecules to target cryptic pockets yields both positive and negative allosteric modulators. *PLoS One* 12, e0178678 (2017).
10. Kotzampasi, D. M. & Cournia, Z. Structural and allosteric effects of the membrane-bound PI3Kα H1047R mutation and opportunities for drug discovery. *Commun. Chem.* 8, 338 (2025).
11. Lazou, M., Kozakov, D., Joseph-McCarthy, D. & Vajda, S. Which cryptic sites are feasible drug targets? *Drug Discov. Today* 29, 104197 (2024).
12. Shan, Y. *et al.* How does a small molecule bind at a cryptic binding site? *PLoS Comput. Biol.* 18, e1009817 (2022).
13. Meller, A. *et al.* Predicting locations of cryptic pockets from single protein structures using the PocketMiner graph neural network. *Nat. Commun.* 14, 1177 (2023).
14. Sen Gupta, P. S., Panda, S. K., Nayak, A. K. & Rana, M. K. Identification and Investigation of a Cryptic Binding Pocket of the P37 Envelope Protein of Monkeypox Virus by Molecular Dynamics Simulations. *Journal of Physical Chemistry Letters* 14, 3230–3235 (2023).
15. Kuzmanic, A., Bowman, G. R., Juarez-Jimenez, J., Michel, J. & Gervasio, F. L. Investigating Cryptic Binding Sites by Molecular Dynamics Simulations. *Acc. Chem. Res.* 53, 654–661 (2020).
16. Bemelmans, M. P., Cournia, Z., Damm-Ganamet, K. L., Gervasio, F. L. & Pande, V. Computational advances in discovering cryptic pockets for drug discovery. *Curr. Opin. Struct. Biol.* 90, 102975 (2025).
17. Rangari, V. A. *et al.* A cryptic pocket in CB1 drives peripheral and functional selectivity. *Nature* 640, 265–273 (2025).
18. Zimmerman, M. I. & Bowman, G. R. FAST Conformational Searches by Balancing Exploration/Exploitation Trade-Offs. *J. Chem. Theory Comput.* 11, 5747–5757 (2015).
19. Herbert, C. *et al.* Molecular mechanism of SSR128129E, an extracellularly acting, small-molecule, allosteric inhibitor of fgf receptor signaling. *Cancer Cell* 23, 489–501 (2013).

20. Benabderrahmane, M., Bureau, R., Voisin-Chiret, A. S. & Santos, J. S. D. O. Cryptic Pockets Repository through Pocket Dynamics Tracking and Metadynamics on Essential Dynamics Space: Applications to Mcl-1. *J. Chem. Inf. Model.* 61, 5581–5588 (2021).
21. Oleinikovas, V., Saladino, G., Cossins, B. P. & Gervasio, F. L. Understanding Cryptic Pocket Formation in Protein Targets by Enhanced Sampling Simulations. *J. Am. Chem. Soc.* 138, 14257–14263 (2016).
22. Borsatto, A., Gianquinto, E., Rizzi, V. & Gervasio, F. L. SWISH-X, an Expanded Approach to Detect Cryptic Pockets in Proteins and at Protein-Protein Interfaces. *J. Chem. Theory Comput.* 20, 3335–3348 (2024).
23. Husic, B. E. & Pande, V. S. Markov State Models: From an Art to a Science. *J. Am. Chem. Soc.* 140, 2386–2396 (2018).
24. Haloi, N., Karlsson, E., Delarue, M., Howard, R. J. & Lindahl, E. Discovering cryptic pocket opening and binding of a stimulant derivative in a vestibular site of the 5-HT 3A receptor. *Sci. Adv.* 11, (2025).
25. Cruz, M. A. *et al.* A cryptic pocket in Ebola VP35 allosterically controls RNA binding. *Nat. Commun.* 13, 2269 (2022).
26. Mallimadugula, U. L., Cruz, M. A., Vithani, N., Zimmerman, M. I. & Bowman, G. R. Opening and closing of a cryptic pocket in VP35 toggles it between two different RNA-binding modes. *Elife* 14, (2025).
27. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589 (2021).
28. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* 630, 493–500 (2024).
29. Wohlwend, J. *et al.* Boltz-1 Democratizing Biomolecular Interaction Modeling. *bioRxiv* https://doi.org/10.1101/2024.11.19.624167 (2024) doi:10.1101/2024.11.19.624167.
30. Boitreaud, J. *et al.* Chai-1: Decoding the molecular interactions of life. Preprint at https://doi.org/10.1101/2024.10.10.615955 (2024).
31. Aranganathan, A., Gu, X., Wang, D., Vani, B. P. & Tiwary, P. Modeling Boltzmann-weighted structural ensembles of proteins using artificial intelligence–based methods. *Curr. Opin. Struct. Biol.* 91, 103000 (2025).
32. Meller, A., Bhakat, S., Solieva, S. & Bowman, G. R. Accelerating Cryptic Pocket Discovery Using AlphaFold. *J. Chem. Theory Comput.* 19, 4355–4363 (2023).
33. Vani, B. P., Aranganathan, A., Wang, D. & Tiwary, P. AlphaFold2-RAVE: From Sequence to Boltzmann Ranking. *J. Chem. Theory Comput.* 19, 4351–4354 (2023).
34. del Alamo, D., Sala, D., Mchaourab, H. S. & Meiler, J. Sampling alternative conformational states of transporters and receptors with AlphaFold2. *Elife* 11, (2022).
35. Berman, H. M. *et al. The Protein Data Bank*. *Nucleic Acids Research* vol. 28 http://www.rcsb.org/pdb/status.html (2000).
36. Jing, B., Berger, B. & Jaakkola, T. AlphaFold Meets Flow Matching for Generating Protein Ensembles. http://arxiv.org/abs/2402.04845 (2024).
37. Lewis, S. *et al.* Scalable emulation of protein equilibrium ensembles with generative deep learning. *Science (1979).* 389, (2025).
38. Cimermancic, P. *et al.* CryptoSite: Expanding the Druggable Proteome by Characterization and Prediction of Cryptic Binding Sites. *J. Mol. Biol.* 428, 709–719 (2016).

39. Škrhák, V., Novotný, M., Feidakis, C. P., Krivák, R. & Hoksza, D. CryptoBench: cryptic protein–ligand binding sites dataset and benchmark. *Bioinformatics* 41, (2024).
40. Škrhák, V., Novotný, M., Riedlova, K. & Hoksza, D. Cryptic binding site prediction with protein language models. in *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* 2883–2888 (IEEE, 2023). doi:10.1109/BIBM58861.2023.10385497.
41. Febrer Martinez, P., Fröhlking, T., Borsatto, A. & Gervasio, F. L. CryptoBank: A Resource for the Identification and Prediction of Cryptic Sites in Proteins. Preprint at https://doi.org/10.1101/2025.04.23.650184 (2025).
42. Leung, D. W. *et al.* Structure of the Ebola VP35 interferon inhibitory domain. *Proc. Natl. Acad. Sci. U. S. A.* 106, 411–416 (2009).
43. Knoverek, C. R. *et al.* Opening of a cryptic pocket in β-lactamase increases penicillinase activity. *Proc. Natl. Acad. Sci. U. S. A.* 118, 1–7 (2021).
44. Bowman, G. R. & Geissler, P. L. 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).
45. Bowman, G. R., Bolin, E. R., Hart, K. M., Maguire, B. C. & Marqusee, S. Discovery of multiple hidden allosteric sites by combining Markov state models and experiments. *Proc. Natl. Acad. Sci. U. S. A.* 112, 2734–2739 (2015).
46. Zimmerman, M. I. *et al.* Prediction of New Stabilizing Mutations Based on Mechanistic Insights from Markov State Models. *ACS Cent. Sci.* 3, 1311–1321 (2017).
47. Mishra, P. & Bowman, G. R. Crowding does not suppress the opening of a cryptic pocket. Preprint at https://doi.org/10.1101/2025.11.17.688865 (2025).
48. Zimmerman, M. I. *et al.* SARS-CoV-2 simulations go exascale to predict dramatic spike opening and cryptic pockets across the proteome. *Nat. Chem.* 13, 651–659 (2021).
49. Lewis, S. *et al.* Scalable emulation of protein equilibrium ensembles with generative deep learning. Preprint at https://doi.org/10.1101/2024.12.05.626885 (2024).
50. Schrödinger LLC. *The PyMOL Molecular Graphics System, Version~1.8*. (2015).
51. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935 (1983).
52. Duan, Y. *et al.* A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* 24, 1999–2012 (2003).
53. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2, 19–25 (2015).
54. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* 18, 1463–1472 (1997).
55. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98, 10089–10092 (1993).
56. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126, 014101 (2007).
57. Hendlich, M., Rippmann, F. & Barnickel, G. LIGSITE: Automatic and efficient detection of potential small molecule-binding sites in proteins. *J. Mol. Graph. Model.* 15, 359–363 (1997).
58. Shrake, A. & Rupley, J. A. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.* 79, 351–371 (1973).

59. McGibbon, R. T. *et al.* MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* 109, 1528–1532 (2015).
60. Porter, J. R., Zimmerman, M. I. & Bowman, G. R. Enspara: Modeling molecular ensembles with scalable data structures and parallel computing. *J. Chem. Phys.* 150, 044108 (2019).
61. Shirts, M. & Pande, V. S. Screen Savers of the World Unite! *Science (1979).* 290, 1903–1904 (2000).
62. Voelz, V. A., Pande, V. S. & Bowman, G. R. Folding@home: Achievements from over 20 years of citizen science herald the exascale era. *Biophys. J.* 122, 2852–2863 (2023).