Xenon Explores Apparent and Cryptic Binding Sites

Shinji Iida*

Department of Data Science, Kitasato University, Kanagawa, Japan

E-mail: iida.shinji@kitasato-u.ac.jp

Abstract

Molecular binding sites of a protein tell us how to target the protein from their structural information, which is required for structure-based drug design. However, regardless of their importance, some binding sites, called cryptic binding (CB) sites, are not always readily apparent in the absence of ligands. The identification of CB sites from protein structures remains challenging, and no standard method has been proposed for this purpose. In this study, inspired by the observation of X-ray protein structures with noble gases, particularly xenon, I evaluate a xenon-based CB site identification strategy from their apo states. I performed atomistic molecular dynamics simulations of apo proteins with known CB sites in the presence of xenons. The simulations have revealed that xenons are likely to occupy not only apparent binding sites but also CB sites, implying that they serve as a probe to explore binding sites.

Introduction

Proteins adopt specific three-dimensional structures that enable them to interact with other molecules, playing essential roles in biological processes. These interactions often involve concave regions on the protein surface, commonly referred to as pockets or binding sites.

Accordingly, various binding site detection algorithms have been developed to provide structural insights relevant to drug design.[1–3]

However, applying such algorithms to a single, static structure risks overlooking transiently formed binding sites. In reality, some binding sites are not visible in the unbound (apo) state but become apparent upon ligand binding (holo state). Such sites are known as cryptic binding (CB) sites—a term first introduced, to my knowledge, by Frembgen-Kesner and Elcock.[4] CB sites were later comprehensively studied by Cimermancic et al., who created the dataset, CryptoSite, and trained a classical machine learning model (e.g., support vector machine) predicting that CB sites are widespread across the human proteome.[5] More recently, machine learning-based prediction methods have been developed so far. For example, Škrhák et al. demonstrated that CB sites can be predicted by a protein language model; their results imply that embeddings from a protein language model contain rich information for CB sites.[6] Meller et al. proposed PocketMiner based on a graph neural network predicting pockets likely to open during MD simulations.[7]

Detecting CB sites requires access to the conformational ensemble of a protein—i.e., a set of structures sampled at equilibrium. Atomistic molecular dynamics (MD) simulations are a powerful tool for generating such ensembles, sampling structures according to the Boltzmann distribution.[8,9] Iida et al. proposed the CB site index, which quantifies the free energy difference associated with the solvent-accessible surface area (SASA) of an amino acid residue in its buried versus exposed state.[9] More recently, Meller et al. demonstrated that MD simulations initiated from AlphaFold2-generated structures can effectively sample cryptic site openings.[10]

In addition to such geometric or energetic criteria, probe molecules have been employed in MD simulations to identify CB sites.[11,12] Yanagisawa et al. developed a rational and systematic cosolvent set for this purpose.[11] Smith and Carlson applied MixMD protocols and successfully mapped known cryptic sites.[12]

Among such probes, a noble gas xenon shows the potential to probe CB sites due to

its ability to bind hydrophobic sites on a protein.[13] Xenon-protein interactions have been studied via X-ray crystallography and NMR spectroscopy,[14,15] although the characterisation of interactions remains challenging. Limitations include: (i) xenon cannot be firmly attached to a solid substrate or marked with a fluorescent (or similar) label, (ii) due to its symmetrical shape and the lack of charge or polarity, xenon engages in interactions that are relatively weak, and (iii) the analytical techniques commonly employed for studying protein-gas interactions, namely NMR spectroscopy and X-ray crystallography, offer detailed structural information but are less proficient in quantifying binding strengths. These limitations have hindered the study of xenon's physiological roles, such as its anaesthetic effects, solubility in blood, and influence on apoptosis.[16]

In this study, I characterise xenon-protein interactions with atomistic MD simulations by exploring apparent and CB sites on proteins. I first built a xenon model with appropriate Lennard-Jones parameters for the reproduction of xenon's physical properties. To assess its validity, I performed MD simulations for proteins that are known to be in complex with xenon. I then evaluate the ability of xenon to bind to apparent binding sites and CB sites for target proteins, compared with benzene molecules. Lastly, I discuss the limitations of xenon atom for the identification of CB sites.

Materials and Methods

Xenon's model

Lennard-Jones (LJ) potential is expressed as

$$u_{\mathrm{LJ}}(r) = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] \tag{1}$$

where r is the distance between two particles, σ is the distance at which the potential becomes zero (a.k.a the size of the particle), and ϵ is the depth of the potential well. Truncated and

shifted Lenard-Jones (TSLJ) potential is then defined as

$$u_{\mathrm{TSLJ}}(r) = \begin{cases} u_{\mathrm{LJ}}(r) - u_{\mathrm{LJ}}(r = r_{\mathrm{c}}) & (r \leq r_{\mathrm{c}}) \\ 0 & (r > r_{\mathrm{c}}) \end{cases} \tag{2}$$

where at the cut-off distance r_{c}, the potential becomes zero. The use of TSLJ potential avoids the calculation of long-range correction, which may be computationally demanding.[17] TSLJ potential is implemented in Gromacs wherein the Verlet cut-off scheme, the potential is shifted by a constant such that it is zero at the cut-off distance.[18] In this study, the parameters σ and ϵ of TSLJ potential were adopted from the literature where it was optimised to reproduce the vapour pressure and saturated liquid density (Table 1).[17] The parameters are slightly different from those in the work by Tang and Toennies.[19]

Table 1: LJ parameters of xenon

σ/nm	ϵ/kJ mol^{-1}
0.39460	2.28531

One might think that xenon is a notably larger noble gas atom compared to neon, argon, and krypton and has polarisability that enables xenon to be susceptible to change-induced dipole interaction, dipole-induced dipole interaction, and London dispersion interaction with a protein; this description aligns with reality. Nonetheless, if the xenon model expressed by the LJ parameters has successfully reproduced xenon-protein complex structures, it is reasonable to state that the model describes a crucial aspect of xenon-protein interactions.

Even though the LJ parameters correctly reproduce some physical quantities for purely xenon solution,[17] this does not guarantee that the LJ parameters correctly model the interactions of xenon with proteins. It is therefore necessary to confirm whether xenon with the LJ parameters binds to xenon-binding sites experimentally verified, and it was verified by MD simulations for four protein known to bind to xenon.

Pairs of apo and holo PDBs

Table 2 shows the pairs of apo and holo PDB IDs. The pairs were used to define each set of amino-acid residues for binding sites, apparent binding sites, and CB sites. The pairs were adopted from the previous work,[9] being selected such that the ligand binding affinity (IC50 or K_{d}) is in the order of nM in the CryptoSite dataset.[5]

Table 2: Pairs of apo and holo structures

Pair 1		**Pair 2**		**Pair 3**		**Pair 4**	
Apo	**Holo**	**Apo**	**Holo**	**Apo**	**Holo**	**Apo**	**Holo**
1b6b[20]	1kuv[21]	1bsq[22]	1gx8[23]	1fvr[24]	2oo8[25]	3cj0[26]	2brl[27]
1hag[28]	1ghy[28]	1jbu[29]	1wun[30]	1k5h[31]	2egh[32]	2ohg[33]	2ohv[33]
1m47[34]	1py2[35]	1maz[36]	2yxj[37]	1nep[38]	2hka[38]	3f74[39]	3bqm[40]
1uk3[41]	2gz7[42]	1w50[43]	3ixj[44]	2wgb[45]	2v57[45]	5j2v[46]	2wi7[47]

Definition of relative solvent accessible surface area

Solvent accessible surface area (SASA) is defined as the surface area traced with a probe whose radius r_{p} is set to 1.4 Å. Relative SASA was defined for each amino-acid residue by

$$rSASA_i := \frac{SASA_i}{SASA_i^{\mathrm{iso}}}, \tag{3}$$

where $SASA_i$ is for an i-th amino-acid residue in a protein, and $SASA_i^{\mathrm{iso}}$ is for an isolated amino-acid residue with its two nearest neighbours if present. It should be noted that $rSASA_i$ ranges from 0 to 1 where "0" corresponds to a completely buried state and "1" does to a completely exposed state. The $rSASA_i$ was computed by the function *cmd.get_sasa_relative* implemented in PyMOL[48]

Definition of binding sites

Let us define a set of contact amino-acid residues as a binding site. Suppose that apo structure and its corresponding holo structures are given. Let P and L be a set of amino-

acid residues in a protein and a set of atoms in the ligand bound to the target protein, respectively. With the cutoff distance $d_{\mathrm{c}} = 5$ Å, the binding site of the molecule was defined by a set of contact residues:

$$S_{\mathrm{BS}} = \{s \in P \mid d(\boldsymbol{r}^{\mathrm{holo}}_{\mathrm{com},s}, \boldsymbol{r}_a) < d_{\mathrm{c}}, a \in L\}, \tag{4}$$

where $d(\boldsymbol{r}^{\mathrm{holo}}_{\mathrm{com},s}, \boldsymbol{r}_a)$ is the distance between the centre-of-mass coordinates $\boldsymbol{r}^{\mathrm{holo}}_{\mathrm{com},s}$ of the side chain for an amino-acid residue s in the holo structure and the coordinates $\boldsymbol{r}_a$ of an atom a in L. The cutoff distance was set to the sum of typical Van der Waals radii (2.0 Å), yielding 5.0 Å, with an additional 1 Å buffer.

Definition of apparent and cryptic binding sites

Here, apparent and CB sites were defined according to an apo state and its corresponding holo state. The definition was used for evaluation of probe-binding events.

The $rSASA_i$ of each residue in S_{BS} was calculated for the apo structure and the ligand-free holo structure (i.e., the ligand was removed but the protein structure was kept), and then the difference is given by

$$\Delta rSASA_i = rSASA_i^{\text{holo-free}} - rSASA_i^{\mathrm{apo}} \tag{5}$$

For a tolerance of exposure ϵ, if $\Delta rSASA_i > \epsilon$, then the i-th residue was considered to be in a CB site. Namely, a CB site was defined by

$$S_{\mathrm{CB}} = \{i \in S_{\mathrm{BS}}, \mid \Delta rSASA_i > \varepsilon\}, \tag{6}$$

Because $\Delta rSASA_i > 0$ indicates that the i-th residue is exposed in the holo structure but is less exposed in the apo structure, it represents an essential aspect of CB sites. The tolerance ϵ was set to 0.1 in this study. The set of residues in a CB site were provided in Table S3.

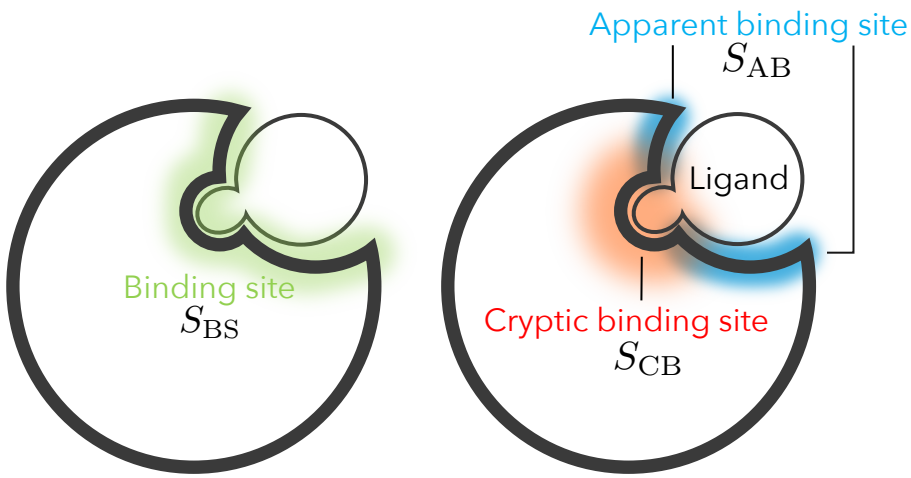


Figure 1: A binding site consists of two regions: apparent and cryptic binding (CB) sites. The left panel shows a binding site bound to a ligand, while the right panel shows the two regions, an apparent and a CB site, in the binding site. Apparent and CB sites are coloured by blue and orange, respectively. Since the definition of CB sites in this study was based on changes in solvent exposure ($\Delta rSASA$), it enables a clear distinction between CB sites and apparent binding sites.

Now, the apparent binding site was defined by a set of contact amino-acid residues except for those in a CB site:

$$S_{\mathrm{AB}} = S_{\mathrm{BS}} \setminus S_{\mathrm{CB}}. \tag{7}$$

Definition of xenon clusters on protein via HDBSCAN

The clusters of xenons (or benzenes) were identified by HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise), which can detects clusters with different densities and an arbitrary shape and removes outliers.[49] Below is the procedure of clustering:

1. From MD trajectories, the snapshots of xenon or benzene were extracted and then flattened, so that a single, aggregated coordinates were obtained.

2. HDBSCAN was applied to the single structure, generating xenon/benzene clusters c_i ($i = 1, 2, ..., N_{\mathrm{clusters}}$). The only parameter optimised was *min_clust_size* which was

set to the number of xenons in the single structure weighted by 0.0025. This weight was selected by the elbow method (See Figure S4 and S5 in the supporting information). The other parameters were set to the default values (See Table S1).

Hit rate of apparent and cryptic binding sites

To evaluate the bindability of xenon/benzene to apparent and CB sites, hit rate was defined by the following steps:

1. The coordinates of probes (i.e., xenon/benzene) obtained from MD simulations were clustered as described in the previous section. The clusters c_i ($i = 1, 2, ..., N_{\text{clusters}}$) were sorted in an descending order based on the number of points in each cluster. The largest cluster was considered to be the most plausible one. The top n_{top} clusters were used for further evaluation.

2. The top n_{top} clusters were mapped onto the corresponding apo structure, which was aligned with the MD snapshots beforehand.

3. For each amino-acid residue in the apo structure, if its centre-of-mass point was within 4 Å of the closest xenon atom in any of the top n_{top} clusters, then the residue was considered to be probe-binding, and the set of the probe-interacting residues is denoted as S_C. The use of only the top one cluster ($n_{top} = 1$) to identify probe-interacting residues and determine binding site identification is the most stringent approach, as it limits the spatial regions considered to only the largest and most probable aggregation of probe atoms identified by the clustering method

4. If there is any overlap between S_C and a predefined set of apparent S_{AB} or CB site residues S_{CB} (i.e., $S_C \cap S_{\text{type}} \neq \emptyset, \text{type} = \text{AB or CB}$), then the xenons were considered to be bindable (a *hit*) to the CB site; otherwise they were considered unbindable.

5. Step 1 - 4 was repeated for each system. The overall hit rate was then calculated by dividing the total number of hits by the total number of systems.

Free energy mapping of xenon positions onto protein

The evaluation of the xenon model requires the identification of free-energetically stable sites on a protein. For this, the density $\rho(\boldsymbol{x})$ of xenon around a protein within a voxel was computed with the *Density object* of MDAnalysis.[50,51] The free energy was computed by

$$G(\boldsymbol{x}) = -k_{\mathrm{B}}T \ln \frac{\rho(\boldsymbol{x})}{\rho_{\max}}, \tag{8}$$

where $\boldsymbol{x} \in \mathbb{R}^3$ denotes Cartesian coordinates in a simulation box and k_{B} and T are Boltzmann constant and temperature, respectively. The division by the maximum density $\rho_{\max}$ sets the most stable voxel to be zero, and thus $G(\boldsymbol{x}) \geq 0$. It should be noted that each MD frame were superimposed before this analysis.

Molecular dynamics simulation

For known protein-xenon complexes, five independent 100-ns NPT simulations at 300 K were performed. Note that the simulations were initiated from a protein structure where xenons are not bound. For two systems (PDB IDs: 1c10 and 5hw1), it was observed that xenon atoms rarely bound to the known xenon-binding sites within 100 ns $\times$ 5 runs because of the requirement of relatively large conformational changes. To enhance conformational changes, five independent 100-ns NPT simulations at 400 K were also carried out for the two systems, being checked if xenons correctly find the known xenon-binding sites. For each of the sixteen protein systems with CB sites, five independent 100-ns MD simulations were performed in the presence of xenon.

The following procedure was used to set up each system: (1) protonation states of histidine residues were assigned using *pdb2gmx* command. (2) The resultant PDB files

were processed with *tleap*, generating topology and coordinate files in Amber format (i.e., .prmptop and .inpcrd). Disulphide bonds were correctly formed in this step. (3) These Amber-format files were converted to Gromacs format (i.e., .top and .gro) using ParmEd[52] (`https://github.com/ParmEd/ParmEd`). (4) These systems were resolvated with dodecahedron water box using *gmx* tools. (5) Xenons (or benzenes) were inserted using *gmx insert-molecule*, replacing some water molecules; the concentration of xenons/benzenes was set to 0.1M. Although this concentration was not systematically optimised, Figure S2 and S3 indicated that the concentration did not destabilise protein structures, showing that 0.1 M was deemed a reasonable compromise. (6) Ions were added to neutralise the system. (7) The system, including protein, water, ions, and xenons/benzenes, was energy-minimised with positional restraints on protein backbones, then equilibrated using NVT and NPT ensembles (500 ps each). (8) Five independent production runs of 100 ns each were then performed for each system under NPT ensemble.

Computational settings were as follows: Time step was set to 2 fs, and an interaction table for cut-off scheme[53] was used whose list was updated every 20 fs. The leap-frog algorithm was used for an integrator, and LINCS[54] was for hydrogen-bond constraints (lincs-iter=1, lincs-order=4). V-rescale[55] was for temperature control for two groups, protein and non-protein ($\tau_T = 0.1$, compressibility $= 4.5 \times 10^6$). Berendsen pressure coupling[56] was used for the first box equilibration, and Parrinelo-Rahman coupling for all production runs to keep 1 atm pressure ($\tau_p = 2.0$, compressibility $= 4.5 \times 10^6$). Short-range electrostatic and Van der Waals interaction cut-off values were set to 10 Å. Amber ff14SB[57] was employed for protein force field and TIP3P model[58] for water molecules. The particle-mesh Ewald (PME) method[59] was used for long range interactions calculation. Snapshots were stored for every 10 ps. Gromacs 2022.4[60] was used for all simulations. The force field parameters of benzene were assigned with GAFF2 via Acpype.[61]

Results and discussion

Xenon model reproduces xenon-protein complex structures

The xenon model must reproduce xenon-protein complex structures. In order to verify the quality of the xenon model, I performed MD simulations of four proteins that are known to have xenon-binding sites. Mapping the free energy iso-surfaces onto protein structures, I visualised probable xenon-binding sites (Figure 2) and counted the number of xenon-binding sites that were free-energetically stable sites (Table 3). It was found that, at 300 K, xenon-binding sites overlapped with free energetically stable sites (See check marks in Figure 2). However, some xenon-binding sites for the proteins of PDBID 1c10 and 5hw1 did not accommodate xenon atoms (See cross marks in Figure 2), because the sites are buried inside the proteins and were unreachable during the simulations.

To make such buried sites reachable during short MD simulations, I also performed MD simulations at 400 K. It was found that xenon atoms were also stabilised in the sites unexplored at 300K. Moreover, even though xenon interacted with proteins nonspecifically, some of the most free-energetically stable sites corresponded to the known xenon-binding site in each X-ray structure (Figure 2). These results demonstrate the reproducibility of the model for xenon-protein interactions.

One may think that the simulations at 400 K caused the disruption of the whole structure, leading to the artefact that xenon appeared to be located in the xenon-binding sites; however, since the whole structure of each protein was retained during the simulations, such artefact did not occur (Figure S1).

Xenon binds to apparent binding sites

With the xenon model, I evaluated the ability of xenon to bind to apparent binding sites with the baseline of benzene. Table 4 demonstrates that xenon and benzene can bind to apparent binding sites. If the hit rate was evaluated only for the top 1 cluster, xenon probed apparent

Table 3: Counts of known xenon-binding sites (n_{xs}) and the known sites corresponding to free-energetically stable xenon-binding sites (n_{s}).

PDB ID	$n_{\mathrm{s}}/n_{\mathrm{xs}}$
1c10[13]	1/2
1c10 (400K)	2/2
1c1m[13]	1/1
1c3l[13]	1/1
5hw1[62]	1/3
5hw1 (400K)	3/3

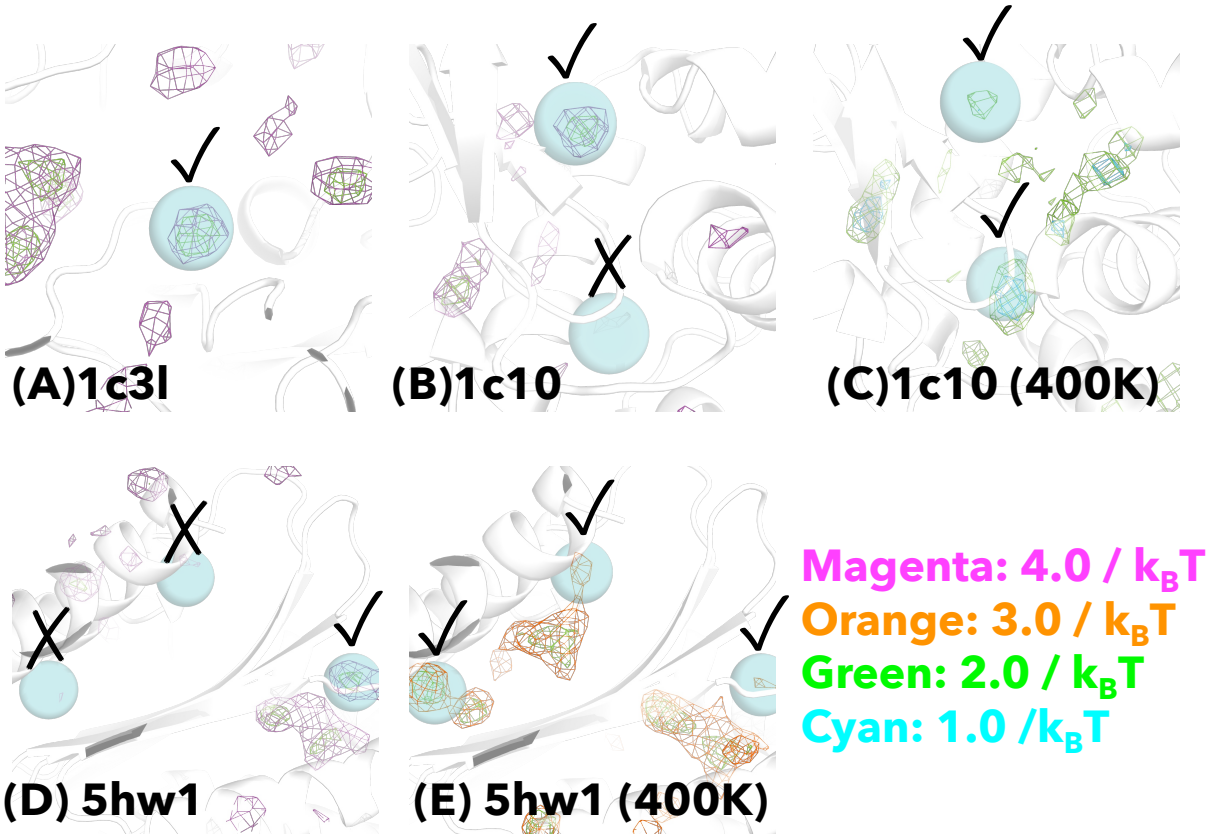


Figure 2: Free energy mapping of xenons on target proteins. Each cyan sphere represents a xenon atom in a crystal structure. Each mesh surface indicates iso-free energy surface of $F(\boldsymbol{x})$: $4.0/k_{\mathrm{B}}T$ for magenta, $3.0/k_{\mathrm{B}}T$ for orange, $2.0/k_{\mathrm{B}}T$ for green, and $1.0/k_{\mathrm{B}}T$ for cyan. The cross and check marks indicate non-explored xenon-binding sites and explored sites with high probability, respectively.

binding sites for four out of sixteen target proteins and benzene did for two out of sixteen. As more clusters were considered, the hit rate increased for both xenon and benzene. This suggests that (i) xenon and benzene can bind to apparent binding sites. and that (ii) xenon are more likely to bind to apparent binding sites than benzene. In fact, earlier studies also suggested that xenon is capable of binding to hydrophobic pockets.[13,14] The table clearly demonstrates that an apparent binding site was probed by xenon, and this result coincides with those earlier works. To my knowledge, this work offers one of the first comparative evaluations of xenon's binding to apparent sites versus the baseline benzene.

Table 4: Hit rate of apparent binding sites.
Hit was evaluated for sixteen target proteins.

	Hit rate	
Used top cluster ID	Xenon	Benzene
1	4/16	2/16
$\leq$2	6/16	3/16
$\leq$3	9/16	4/16
$\leq$4	10/16	6/16
$\leq$5	12/16	6/16
$\leq$6	12/16	6/16
$\leq$7	12/16	7/16
$\leq$8	12/16	8/16
$\leq$9	14/16	8/16
$\leq$10	14/16	9/16

Xenon binds to cryptic binding sites.

Similar to the analysis in the previous section, I examined the ability of xenon to bind to CB sites as a comparison with the baseline benzene. Table 5 shows that if the top 1 cluster was only considered, xenon probed CB sites of four target proteins out of sixteen, whereas benzene probed five target proteins out of sixteen. With the top 1 cluster, benzene was thus superior to xenon; however, if top $\leq$ 2 clusters were used for the assessment of binding, the hit rate of xenon exceeded that of benzene. As compared with Table 4, benzene struggled

to bind to CB sites. It implies that xenon would be a better probe to identify CB sites than benzene.

I visualised the clusters mapped onto holo structures in Figure 3. This figure shows that xenon clusters overlapped with ligand binding sites, visually indicating that xenon is able to explore CB sites.

Let us now examine each figure in detail. In Figure 3A, the 1st, 2nd, and 3rd xenon clusters overlapped with the entire ligand. In Figure 3B, it is found that the exposure of amino-acid residues (yellow coloured surface) opens the CB site, and the site became accessible to xenon atoms in the 3rd cluster. In Figure 3C, the aromatic ring was traced by the 2nd xenon cluster. In Figure 3D, while the site where fluorine of the tip of the ligand is located was not probed by xenon (labelled by "Not explored"), the 5th and 1st cluster covered some aromatic rings. In Figure 3E, the four rings were probed by the 1st and 2nd clusters. In Figure 3F, the overall ligand conformation was probed by the 1st and 2nd clusters. Noting that all the simulations in this study were started from an apo structure, this result would ensure that xenon can probe CB sites even when only an apo structure is known. While the cases above are success examples, failure cases would also be informative to point out limitations, which is discussed in the section Limitation.

Table 5: Hit rate of cryptic binding sites

	Hit rate	
Used top cluster ID	Xenon	Benzene
1	4/16	5/16
≤2	7/16	5/16
≤3	10/16	5/16
≤4	11/16	5/16
≤5	11/16	5/16
≤6	11/16	5/16
≤7	11/16	5/16
≤8	11/16	5/16
≤9	12/16	6/16
≤10	12/16	6/16

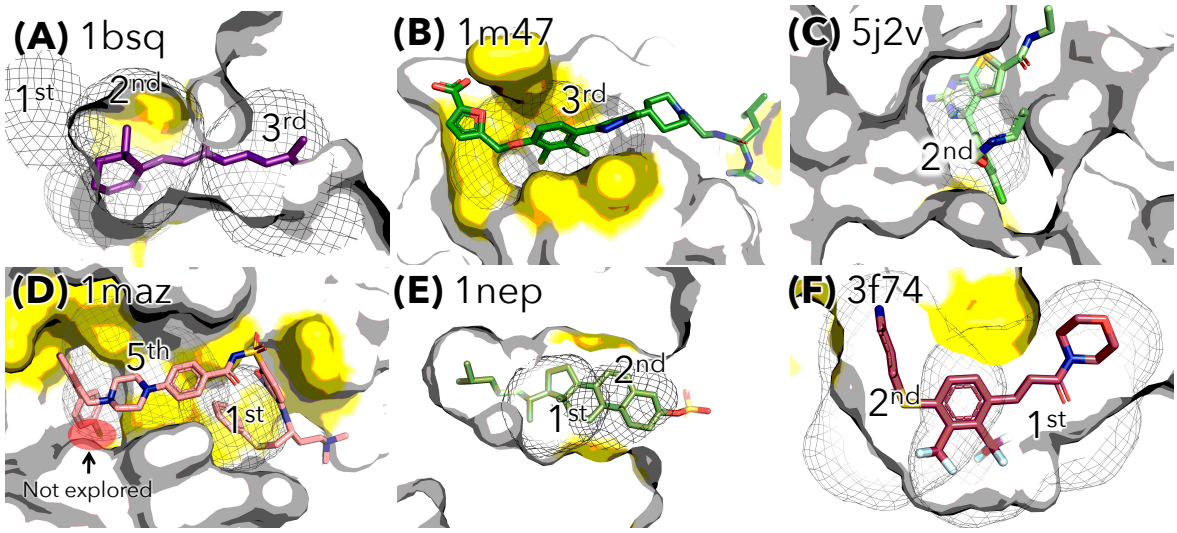


Figure 3: HDBSCAN clusters of xenon atoms mapped onto each holo structure corresponding to its apo structure whose PDB ID is written next to the Figure label (A)-(F). The yellow region indicates amino-acid residues judged as those in a CB site. Each mesh represents an occupied region by xenon atoms in a cluster on which cluster ID is labelled.

Limitation

Although xenon is likely to bind to apparent and CB sites, it has limitations as a probe. Firstly, xenon binds not only to CB sites but also to other pockets, leading to the increase in false positive rate for the identification of CB sites. To address this, prior studies have introduced filtering algorithms. However, developing such a method is beyond the scope of the present study; instead, I focused on assessing xenon's intrinsic binding preferences for apparent and CB sites.

Secondly, large conformational changes were not induced by the inclusion of xenon atoms in canonical MD simulations (Figure 4). In Figure 4A, the protein surface is so flat that xenon struggled to settle, and thus the strategy in this work failed to identify the apparent and CB sites. In Figure 4B, the CB site in the apo state is covered with a loop and the loop movement requires movement of a helix, which was not observed in MD simulations. In Figure 4C, the CB site in the apo state is covered with a helix, so that the removal of helix is necessary for the exposure of CB site; however, such large conformational changes

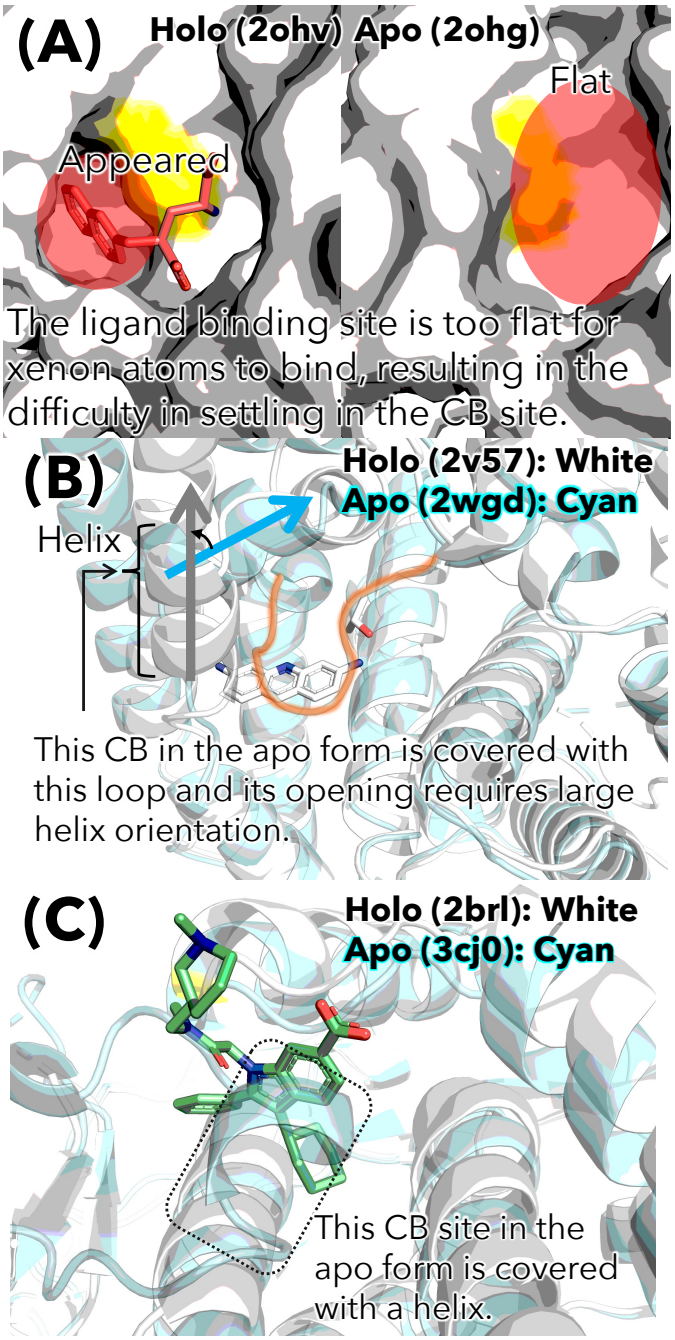


Figure 4: Cases where xenon failed to bind to CB sites. (A): Holo and apo structures are depicted by surface representation and the yellow region indicates amino-acid residues judged as those in a CB site. (B and C): The holo and apo PDB structures are coloured by white and cyan, respectively.

were not observed in MD simulations.

To observe such large conformational changes, the use of an enhanced sampling method would be required for efficient exploration of CB sites via xenon. In fact, Kuzmanic et al. suggested the effectiveness of an enhanced sampling method with cosolvent probes to explore CB sites.[8] Although the present study has several limitations, it has the following significance: (i) it provides the evaluation of the xenon model with the LJ parameters for xenon-protein interactions and (ii) it investigates the ability of xenon to bind to apparent binding and CB sites for various proteins.

Conclusion

In this study, I demonstrated that xenon, through atomistic MD simulations, can probe both apparent and CB sites on proteins. The key finding is that xenon, even in simulations from apo protein structures, is able to identify CB sites if large structural rearrangement is not required. Compared to benzene, xenon achieved higher hit rates in identifying CB sites when more than one cluster was considered, indicating its superior suitability as a probe.

I should mention that this work is not the first work focusing on xenon as a probe for binding site identification. For example, Vithani et al. utilised xenon as a probe and performed MD simulations for a protein KRAS to identify its CB sites, stating that "we believe xenon might make a reasonable generic probe of cryptic sites within any protein complex, including KRAS.".[63]

The results in this study suggest that xenon is not only a viable but a potentially superior probe for detecting cryptic binding sites in atomistic MD simulations. Its effectiveness, even from apo structures, highlights its utility in scenarios where ligand-bound information is unavailable. Future work could involve leveraging curated datasets such as CryptoBench[64] for more systematic benchmarking. Moreover, it is crucial to compare computational results to those of experiments: A plausible collaboration would involve the combination between

MD and NMR-based strategy, which can identify the population of CB sites in equilibrium.[65]

Taken together, this study reinforces the role of xenon as a powerful tool for CB site identification and provides a foundation for its broader application in computational drug discovery.

Data availability

The Gromacs topology file of xenon is publicly available at `https://gist.github.com/physicshinzui/718b101964d44a1d58dfd4cc8bd15860`. All the MD simulation data (xtc format trajectory files and their reference PDBs) are also publicly available at `https://doi.org/10.5281/zenodo.15288813`. Due to the data size, all the water molecules in each MD data were removed.

Acknowledgement

This work was supported by Kitasato University Academic Research Encouragement Fund. The computation was performed using the Research Center for Computational Science, Okazaki, Japan (Project: 24-IMS-C128). I thank Dr. Yoshifumi Fukunishi for fruitful discussions.

References

(1) Le Guilloux, V.; Schmidtke, P.; Tuffery, P. Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics* **2009**, *10*, 168.

(2) Hendlich, M.; Rippmann, F.; Barnickel, G. LIGSITE: automatic and efficient detection of potential small molecule-binding sites in proteins. *J. Mol. Graph. Model.* **1997**, *15*, 359–63, 389.

(3) Fukunishi, Y.; Nakamura, H. Prediction of ligand-binding sites of proteins by molecular docking calculation for a random ligand library: Prediction of Ligand-Binding Sites. *Protein Sci.* **2011**, *20*, 95–106.

(4) Frembgen-Kesner, T.; Elcock, A. H. Computational sampling of a cryptic drug binding site in a protein receptor: explicit solvent molecular dynamics and inhibitor docking to p38 MAP kinase. *J. Mol. Biol.* **2006**, *359*, 202–214.

(5) Cimermancic, P.; Weinkam, P.; Rettenmaier, T. J.; Bichmann, L.; Keedy, D. A.; Woldeyes, R. A.; Schneidman-Duhovny, D.; Demerdash, O. N.; Mitchell, J. C.; Wells, J. A.; Fraser, J. S.; Sali, A. CryptoSite: Expanding the druggable proteome by characterization and prediction of cryptic binding sites. *J. Mol. Biol.* **2016**, *428*, 709–719.

(6) Škrhák, V.; Riedlova, K.; Novotný, M.; Hoksza, D. Cryptic binding site prediction with protein language models. 2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). 2023; pp 2883–2888.

(7) Meller, A.; Ward, M.; Borowsky, J.; Kshirsagar, M.; Lotthammer, J. M.; Oviedo, F.; Ferres, J. L.; Bowman, G. R. Predicting locations of cryptic pockets from single protein structures using the PocketMiner graph neural network. *Nat. Commun.* **2023**, *14*, 1177.

(8) Kuzmanic, A.; Bowman, G. R.; Juarez-Jimenez, J.; Michel, J.; Gervasio, F. L. Investigating cryptic binding sites by molecular dynamics simulations. *Acc. Chem. Res.* **2020**, *53*, 654–661.

(9) Iida, S.; Nakamura, H. K.; Mashimo, T.; Fukunishi, Y. Structural fluctuations of aromatic residues in an APO-form reveal cryptic binding sites: Implications for fragment-based drug design. *J. Phys. Chem. B* **2020**, *124*, 9977–9986.

(10) Meller, A.; Bhakat, S.; Solieva, S.; Bowman, G. R. Accelerating cryptic pocket discovery using AlphaFold. *J. Chem. Theory Comput.* **2023**, *19*, 4355–4363.

(11) Yanagisawa, K.; Moriwaki, Y.; Terada, T.; Shimizu, K. EXPRORER: Rational cosolvent set construction method for cosolvent molecular dynamics using large-scale computation. *J. Chem. Inf. Model.* **2021**, *61*, 2744–2753.

(12) Smith, R. D.; Carlson, H. A. Identification of cryptic binding sites using MixMD with standard and accelerated molecular dynamics. *J. Chem. Inf. Model.* **2021**, *61*, 1287–1299.

(13) Prangé, T.; Schiltz, M.; Pernot, L.; Colloc'h, N.; Longhi, S.; Bourguet, W.; Fourme, R. Exploring hydrophobic sites in proteins with xenon or krypton. *Proteins* **1998**, *30*, 61–73.

(14) Roose, B. W.; Zemerov, S. D.; Dmochowski, I. J. Xenon-protein interactions: Characterization by X-ray crystallography and hyper-CEST NMR. *Methods Enzymol.* **2018**, *602*, 249–272.

(15) Wiebelhaus, N.; Singh, N.; Zhang, P.; Craig, S. L.; Beratan, D. N.; Fitzgerald, M. C. Discovery of the xenon-protein interactome using large-scale measurements of protein folding and stability. *J. Am. Chem. Soc.* **2022**, *144*, 3925–3938.

(16) Spaggiari, S.; Kepp, O.; Rello-Varona, S.; Chaba, K.; Adjemian, S.; Pype, J.; Galluzzi, L.; Lemaire, M.; Kroemer, G. Antiapoptotic activity of argon and xenon. *Cell Cycle* **2013**, *12*, 2636–2642.

(17) Rutkai, G.; Thol, M.; Span, R.; Vrabec, J. How well does the Lennard-Jones potential represent the thermodynamic properties of noble gases? *Mol. Phys.* **2017**, *115*, 1104–1121.

(18) GROMACS 2022.4 manual; nonbonded-interactions. `https://manual.gromacs.org/2022.4/reference-manual/functions/nonbonded-interactions.html`, Accessed: 2023-12-9.

(19) Tang, K. T.; Toennies, J. P. New combining rules for well parameters and shapes of the van der Waals potential of mixed rare gas systems. *Z. Phys. D At. Mol. Clust.* **1986**, *1*, 91–101.

(20) Hickman, A. B.; Klein, D. C.; Dyda, F. Melatonin biosynthesis: the structure of serotonin N-acetyltransferase at 2.5 A resolution suggests a catalytic mechanism. *Mol. Cell* **1999**, *3*, 23–32.

(21) Wolf, E.; De Angelis, J.; Khalil, E. M.; Cole, P. A.; Burley, S. K. X-ray crystallographic studies of serotonin N-acetyltransferase catalysis and inhibition. *J. Mol. Biol.* **2002**, *317*, 215–224.

(22) Qin, B. Y.; Bewley, M. C.; Creamer, L. K.; Baker, E. N.; Jameson, G. B. Functional implications of structural differences between variants A and B of bovine beta-lactoglobulin. *Protein Sci.* **1999**, *8*, 75–83.

(23) Kontopidis, G.; Holt, C.; Sawyer, L. The ligand-binding site of bovine beta-lactoglobulin: evidence for a function? *J. Mol. Biol.* **2002**, *318*, 1043–1055.

(24) Shewchuk, L. M.; Hassell, A. M.; Ellis, B.; Holmes, W. D.; Davis, R.; Horne, E. L.; Kadwell, S. H.; McKee, D. D.; Moore, J. T. Structure of the Tie2 RTK domain: self-inhibition by the nucleotide binding loop, activation loop, and C-terminal tail. *Structure* **2000**, *8*, 1105–1113.

(25) Hodous, B. L. et al. Synthesis, structural analysis, and SAR studies of triazine derivatives as potent, selective Tie-2 inhibitors. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 2886–2889.

(26) Antonysamy, S. S. et al. Fragment-based discovery of hepatitis C virus NS5b RNA polymerase inhibitors. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 2990–2995.

(27) Di Marco, S.; Volpari, C.; Tomei, L.; Altamura, S.; Harper, S.; Narjes, F.; Koch, U.; Rowley, M.; De Francesco, R.; Migliaccio, G.; Carfí, A. Interdomain communication in hepatitis C virus polymerase abolished by small molecule inhibitors bound to a novel allosteric site. *J. Biol. Chem.* **2005**, *280*, 29765–29770.

(28) Vijayalakshmi, J.; Padmanabhan, K. P.; Mann, K. G.; Tulinsky, A. The isomorphous structures of prethrombin2, hirugen-, and PPACK-thrombin: changes accompanying activation and exosite binding to thrombin. *Protein Sci.* **1994**, *3*, 2254–2271.

(29) Eigenbrot, C.; Kirchhofer, D.; Dennis, M. S.; Santell, L.; Lazarus, R. A.; Stamos, J.; Ultsch, M. H. The factor VII zymogen structure reveals reregistration of beta strands during activation. *Structure* **2001**, *9*, 627–636.

(30) Kadono, S. et al. Structure-based design of P3 moieties in the peptide mimetic factor VIIa inhibitor. *Biochem. Biophys. Res. Commun.* **2005**, *327*, 589–596.

(31) Reuter, K.; Sanderbrand, S.; Jomaa, H.; Wiesner, J.; Steinbrecher, I.; Beck, E.; Hintz, M.; Klebe, G.; Stubbs, M. T. Crystal structure of 1-deoxy-D-xylulose-5-phosphate reductoisomerase, a crucial enzyme in the non-mevalonate pathway of isoprenoid biosynthesis. *J. Biol. Chem.* **2002**, *277*, 5378–5384.

(32) Yajima, S.; Hara, K.; Iino, D.; Sasaki, Y.; Kuzuyama, T.; Ohsawa, K.; Seto, H. Structure of 1-deoxy-D-xylulose 5-phosphate reductoisomerase in a quaternary complex with a magnesium ion, NADPH and the antimalarial drug fosmidomycin. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2007**, *63*, 466–470.

(33) Kim, K.-H.; Bong, Y.-J.; Park, J. K.; Shin, K.-J.; Hwang, K. Y.; Kim, E. E. Structural basis for glutamate racemase inhibition. *J. Mol. Biol.* **2007**, *372*, 434–443.

(34) Arkin, M. R.; Randal, M.; DeLano, W. L.; Hyde, J.; Luong, T. N.; Oslob, J. D.; Raphael, D. R.; Taylor, L.; Wang, J.; McDowell, R. S.; Wells, J. A.; Braisted, A. C.

Binding of small molecules to an adaptive protein-protein interface. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 1603–1608.

(35) Thanos, C. D.; Randal, M.; Wells, J. A. Potent small-molecule binding to a dynamic hot spot on IL-2. *J. Am. Chem. Soc.* **2003**, *125*, 15280–15281.

(36) Muchmore, S. W.; Sattler, M.; Liang, H.; Meadows, R. P.; Harlan, J. E.; Yoon, H. S.; Nettesheim, D.; Chang, B. S.; Thompson, C. B.; Wong, S. L.; Ng, S. L.; Fesik, S. W. X-ray and NMR structure of human Bcl-xL, an inhibitor of programmed cell death. *Nature* **1996**, *381*, 335–341.

(37) Lee, E. F.; Czabotar, P. E.; Smith, B. J.; Deshayes, K.; Zobel, K.; Colman, P. M.; Fairlie, W. D. Crystal structure of ABT-737 complexed with Bcl-xL: implications for selectivity of antagonists of the Bcl-2 family. *Cell Death Differ.* **2007**, *14*, 1711–1713.

(38) Friedland, N.; Liou, H.-L.; Lobel, P.; Stock, A. M. Structure of a cholesterol-binding protein deficient in Niemann-Pick type C2 disease. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 2512–2517.

(39) Zhang, H.; Astrof, N. S.; Liu, J.-H.; Wang, J.-H.; Shimaoka, M. Crystal structure of isoflurane bound to integrin LFA-1 supports a unified mechanism of volatile anesthetic action in the immune and central nervous systems. *FASEB J.* **2009**, *23*, 2735–2740.

(40) Guckian, K. M.; Lin, E. Y.-S.; Silvian, L.; Friedman, J. E.; Chin, D.; Scott, D. M. Design and synthesis of a series of meta aniline-based LFA-1 ICAM inhibitors. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5249–5251.

(41) Yang, H.; Yang, M.; Ding, Y.; Liu, Y.; Lou, Z.; Zhou, Z.; Sun, L.; Mo, L.; Ye, S.; Pang, H.; Gao, G. F.; Anand, K.; Bartlam, M.; Hilgenfeld, R.; Rao, Z. The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 13190–13195.

(42) Lu, I.-L.; Mahindroo, N.; Liang, P.-H.; Peng, Y.-H.; Kuo, C.-J.; Tsai, K.-C.; Hsieh, H.-P.; Chao, Y.-S.; Wu, S.-Y. Structure-based drug design and structural biology study of novel nonpeptide inhibitors of severe acute respiratory syndrome coronavirus main protease. *J. Med. Chem.* **2006**, *49*, 5154–5161.

(43) Patel, S.; Vuillard, L.; Cleasby, A.; Murray, C. W.; Yon, J. Apo and inhibitor complex structures of BACE (beta-secretase). *J. Mol. Biol.* **2004**, *343*, 407–416.

(44) Björklund, C.; Oscarson, S.; Benkestock, K.; Borkakoti, N.; Jansson, K.; Lindberg, J.; Vrang, L.; Hallberg, A.; Rosenquist, A.; Samuelsson, B. Design and synthesis of potent and selective BACE-1 inhibitors. *J. Med. Chem.* **2010**, *53*, 1458–1464.

(45) Bellinzoni, M.; Buroni, S.; Schaeffer, F.; Riccardi, G.; De Rossi, E.; Alzari, P. M. Structural plasticity and distinct drug-binding modes of LfrR, a mycobacterial efflux pump regulator. *J. Bacteriol.* **2009**, *191*, 7531–7537.

(46) Amaral, M.; Kokh, D. B.; Bomke, J.; Wegener, A.; Buchstaller, H. P.; Eggenweiler, H. M.; Matias, P.; Sirrenberg, C.; Wade, R. C.; Frech, M. Protein conformational flexibility modulates kinetics and thermodynamics of drug binding. *Nat. Commun.* **2017**, *8*.

(47) Brough, P. A. et al. Combining hit identification strategies: fragment-based and in silico approaches to orally active 2-aminothieno[2,3-d]pyrimidine inhibitors of the Hsp90 molecular chaperone. *J. Med. Chem.* **2009**, *52*, 4794–4809.

(48) modules/pymol/util.py at 03d7a7fcf0bd95cd93d710a1268dbace2ed77765 · schrodinger/pymol-open-source.

(49) Campello, R. J. G. B.; Moulavi, D.; Sander, J. *Advances in Knowledge Discovery and Data Mining*; Lecture notes in computer science; Springer Berlin Heidelberg: Berlin, Heidelberg, 2013; pp 160–172.

(50) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **2011**, *32*, 2319–2327.

(51) Gowers, R.; Linke, M.; Barnoud, J.; Reddy, T.; Melo, M.; Seyler, S.; Domański, J.; Dotson, D.; Buchoux, S.; Kenney, I.; Beckstein, O. MDAnalysis: A python package for the rapid analysis of molecular dynamics simulations. Proceedings of the Python in Science Conference. 2016; pp 98–105.

(52) Shirts, M. R.; Klein, C.; Swails, J. M.; Yin, J.; Gilson, M. K.; Mobley, D. L.; Case, D. A.; Zhong, E. D. Lessons learned from comparing molecular dynamics engines on the SAMPL5 dataset. *J. Comput. Aided Mol. Des.* **2017**, *31*, 147–161.

(53) Páll, S.; Hess, B. A flexible algorithm for calculating pair interactions on SIMD architectures. *arXiv [physics.comp-ph]* **2013**,

(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.* **1997**, *18*, 1463–1472.

(55) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

(56) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(57) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.

(58) 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.* **1983**, *79*, 926–935.

(59) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

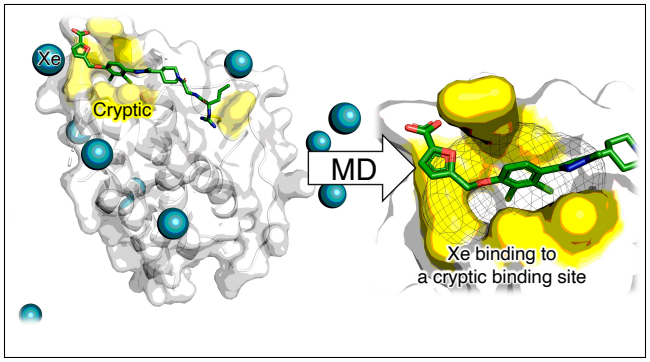
(60) Bauer, P.; Hess, B.; Lindahl, E. GROMACS 2022.4 Manual. 2022.

(61) Sousa da Silva, A. W.; Vranken, W. F. ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Res. Notes* **2012**, *5*, 367.

(62) Roose, B. W.; Zemerov, S. D.; Wang, Y.; Kasimova, M. A.; Carnevale, V.; Dmochowski, I. J. A structural basis for 129 Xe hyper-CEST signal in TEM-1 β-lactamase. *Chemphyschem* **2019**, *20*, 260–267.

(63) Vithani, N.; Zhang, S.; Thompson, J. P.; Patel, L. A.; Demidov, A.; Xia, J.; Balaeff, A.; Mentes, A.; Arnautova, Y. A.; Kohlmann, A.; Lawson, J. D.; Nicholls, A.; Skillman, A. G.; LeBard, D. N. Exploration of cryptic pockets using enhanced sampling along normal modes: A case study of KRAS G12D. *J. Chem. Inf. Model.* **2024**, *64*, 8258–8273.

(64) Škrhák, V.; Novotný, M.; Feidakis, C. P.; Krivák, R.; Hoksza, D. CryptoBench: cryptic protein-ligand binding sites dataset and benchmark. *Bioinformatics* **2024**, *41*.

(65) Mizukoshi, Y.; Takeuchi, K.; Tokunaga, Y.; Matsuo, H.; Imai, M.; Fujisaki, M.; Kamoshida, H.; Takizawa, T.; Hanzawa, H.; Shimada, I. Targeting the cryptic sites: NMR-based strategy to improve protein druggability by controlling the conformational equilibrium. *Sci. Adv.* **2020**, *6*.

TOC Graphic