Exploring Cryptic Pockets Formation in Targets of Pharmaceutical Interest with SWISH

Federico Comitani[†] and Francesco Luigi Gervasio[∗,†,‡]

Department of Chemistry, University College London, London WC1E 6BT, United Kingdom, and Institute of Structural and Molecular Biology, University College London, London WC1E 6BT, United Kingdom

E-mail: f.l.gervasio@ucl.ac.uk

Abstract

Cryptic (hidden) pockets are sites that are not visible on unliganded target proteins' structures and only become apparent when a ligand binds. They might provide a valid alternative to classical binding sites in otherwise "undruggable" targets, but their hidden nature makes it difficult to use standard structure-based or computer-aided drug discovery approaches. We recently developed a Hamiltonian replica exchange method (Sampling Water Interfaces through Scaled Hamiltonians or SWISH) that improves the sampling of hydrophobic cavities by scaling the interactions between water molecules and protein atoms. Here, we discuss further improvements to SWISH and its combination with fragment probe simulations. We tested the robustness and general applicability of the improved approach in a variety of pharmaceutically relevant targets. The chosen proteins: NPC2, p38α, LfrR, and hPNMT, represent a set of diversified and interesting targets harboring non-trivial cryptic binding sites. In all cases, the updated version of our algorithm efficiently explored the cryptic sites.

[∗]To whom correspondence should be addressed
[†]University College London
[‡]ISMB

1

Introduction

Large-scale sequencing projects of the human genome allowed a re-assessment of the cohort of protein targets that provide an opportunity for therapeutic intervention, showing clear potential to greatly expand their number.[1] However, many biologically validated targets are not amenable to classic substrate-competitive drug discovery strategies; thus significant efforts are now being focused toward alternative strategies, including the systematic identification and characterization of hidden (cryptic) allosteric pockets. Indeed, cryptic cavities represent an attractive alternative to substrate-competitive sites, as also shown in the case of K-RAS, an oncogene commonly found in human cancers. For many years the drug discovery community tried and failed to develop substrate-competitive K-RAS inhibitors, and many thought it to be undruggable. This all changed when a new cryptic pocket was serendipitously found.[2] What is more, it is generally thought that targeting such allosteric sites might improve specificity and selectivity.[3–5] However, their hidden nature makes them particularly difficult to identify. At variance to orthosteric pockets, which are generally manifest in the unbound protein structures, cryptic sites require conformational changes to occur in the protein prior to or in the presence of the ligand, whether by conformational selection, induced-fit or a mixed mechanism.[6–8] Some even require rearrangements of the interfacial interactions between multiple domains.[9] Structure-based and computer-aided drug discovery approaches, including virtual screening and free energy methods, generally assume a well-formed pocket and are thus ill-equipped to reveal cryptic sites.[10,11] Available experimental approaches such as fragment screening or tethering are laborious and not always effective.[12,13]

Most known cryptic pockets were found by chance, as it was the case for TEM1 β-lactamase, a relevant pharmacological target for antibiotic resistance, in which a previously unknown allosteric binding site was identified when small ligands were found wedged between two hydrophobic helices in a crystal.[14] Or indeed, in the case of K-RAS, in which the hidden site was unexpectedly revealed through tethering.[2]

Given the dearth of effective experimental and computational methods, the hunt for cryptic

pockets could greatly benefit from new modeling techniques. Very long Molecular Dynamics (MD) simulations, made possible by the increase in available computer power, proved successful in a few cases, especially when combined with Markov State Models.[4,15,16] Simulations with probe fragments, where the target protein is solvated in a mixture of water and small organic molecules with different physicochemical properties, have shown promise in identifying hydrophobic patches and inducing small conformational changes. Still, their success depends on the nature of the probes and the length of the simulations, and they usually fail to reveal deeply buried cryptic pockets.[17–23] Enhanced sampling simulations, with collective-variable based algorithms such as Metadynamics,[24,25] have been successfully used to improve the sampling of small and large conformational changes and characterize binding mechanisms.[26–33] However they require some *a priori* knowledge of the reaction coordinate involved in the pocket opening. Variables based on generic properties of cryptic pockets (i.e. their hydrophobicity) such as JEDI are being actively developed, but they still require prior knowledge of the cavity location.[34] At the other end of the spectrum are machine learning approaches such as Cryptosite, a hidden sites identification tool. They show promise, but are limited by the scarce availability of experimentally-determined cryptic pockets and lack the atomic-level insight that MD-based approaches can provide.[35] Databases such as the one compiled for Cryptosite also make evident the importance of an unambiguous definition of cryptic pockets, which reflects in the success of the algorithms used to hunt them. Some are based on comparing the accessible volume in the crystal structures of the bound and unbound forms. However, the trivial displacement of one or more side chains can lead to considerable fluctuations in such quantity, leading to false positives that cannot be classified as genuine cryptic sites. In spite of the numerous attempts, systematic and robust protocols to identify and characterize cryptic sites are still actively sought.

We recently developed a new Hamiltonian replica exchange-based method able to improve the exploration of hydrophobic patches and test their druggability. Sampling Water Interfaces through Scaled Hamiltonians (SWISH) is based on the scaling of non-bonded interactions be-

tween water molecules and apolar carbon and sulfur atoms of the solute. The scaling factor λ is gradually increased for different replicas within a Hamiltonian Replica-Exchange (HREX) framework, modulating the water-interface interactions and allowing highly hydrophobic, and possibly druggable, patches to open to the ligand. We initially tested SWISH on three relatively simple systems harboring cryptic sites: TEM1 β-lactamase, interleukin-2, which plays a role in the activation of T-cells and in graft tissue rejection, and a validated anti-cancer target, Polo-like kinase-1. Cryptic pockets on these targets were discovered by chance. SWISH, alone or combined with molecular probes, was able to correctly identify and characterize the three hidden sites and distinguish real binding interfaces from false positives.[36] However, the λ factor had to be carefully selected: high enough to open the cryptic pockets but not too high as it might unfold some target proteins. Here, we further improve the methodology by addressing the risk of unfolding and test it on a great variety of target types. To this end, we carefully inspected and analyzed known cryptic sites; this made clear the paucity of deeply buried and genuinely hidden cavities. Still, we were able to identify four well defined targets with experimentally validated cryptic pockets: NPC2, a small sterol transporter,[37] p38α, a mitogen-activated protein kinase,[38] LfrR, a bacterial efflux pump repressor,[39] and hPNMT, a catalyst for the synthesis of adrenaline.[40] Of these, three are also included in the Cryptosite database.[35] Our choice was guided by the necessity of testing our approach against a set of challenging and varied problems, to demonstrate its capabilities and find its limits.

Results

After an initial set up (see Methods), each of the chosen systems was subject to 0.1 μs of MD at constant temperature and volume, 0.5 μs of SWISH in aqueous solution (with TIP3P water) and 1 μs of SWISH in a mixed water/benzene solution. As shown in Ref. 36, benzene is one of the best probes for exploring cryptic hydrophobic cavities. Six HREX replicas

were employed with water-protein interaction scaled up to 1.35. To compare SWISH against simple mixed-solvent simulations, the opening of the pocket (as measured by the volume relative to the fully open state) and the presence of fragments in the cavity were examined against those obtained from 20 independent (not HREX) replicas of classical mixed-solvent MD with benzene, run for 0.3 μs each, for a total of more than 60 μs of simulations. Previously, we observed that using high λ scaling factors may lead to excessive deformation and partial unfolding of the proteins.[36] Still, a sufficiently high scaling factor is often crucial to assure a sufficient opening of the pockets under investigation. To keep the benefit of high scaling factors, while avoiding unfolding the proteins, here we introduce a flexible restraint based on the distance from the native contact map. The choice of a contact map was dictated by its effectiveness in keeping the structure folded while avoiding the excessive rigidity and the requirement of aligning the protein at each step typical of RMSD-based restraints. The restraint is only activated if and when the distance from the native crystallographic contact map reaches a pre-set maximum value (upper wall). The value for the wall was chosen to be slightly larger than the typical fluctuations observed in the unbiased MD runs, to allow for necessary movements.

NPC2

The Niemann-Pick type C disease is a fatal hereditary disorder characterized by an accumulation of cholesterol in lysosomes. Two genes NPC1 and NPC2 have been associated with this dysfunction; the two proteins they encode are responsible for the transport of free cholesterol (and other sterols) to different cellular regions.[37] Although structurally different, one is a large trans-membrane domain, while the other is a 130-amino acid glycoprotein, NPC1 and NPC2 work together in a single pathway, whose malfunctioning leads to a disruption of the normal intracellular cholesterol trafficking, impairing numerous functions such as the synthesis of membranes or sterol hormones.[41] The structure of NPC2 is quite simple, it is formed by a 7-strands β-sandwich fold, topologically identical to other members of the MD-2-related lipid recognition (ML) domain family, to which this protein belongs.

However, unlike any of the other ML proteins, NPC2 lacks an apparent hydrophobic cavity to accommodate the ligands in its unbound (apo) state. The hidden cavity, which eluded early experimental attempts, was finally observed in 2007 when the protein was crystallized bound to a cholesterol-3-O-sulfate (C3S).[23,41] The comparison between apo and holo crystal structures (see Fig. 1 a right) shows a subtle rearrangement of the side chains within the hydrophobic core β strands, which move just enough to fit around the sterol. While other members of the ML family have a proper lid regulating the access to the site, in NPC2 this structure is substituted by the βD and $\beta E - \beta F$ loop lining the entrance (highlighted in Fig. 1 a left). Thus, it has been proposed that the pocket in the apo NPC2 is closed to compensate for the lack of the lid.[23,41]

NPC2 represents a good case for its intrinsic difficulty, a deeply buried pocket within the β-sandwich core and the expected specificity to sterol-like compounds. The cholesterol-3-O-sulfate wedges itself within the two β-sandwich faces, displacing the sidechains of Y100 on the $\beta E - \beta F$ loop and F66 on the βD strand, which act as a gate to the pocket through π-stacking in the apo structure. A few other minor structural changes, mostly affecting the side chains of Y36 and L94, are also observed in the protein core itself.[37,41]

Interestingly, previous attempts to open the pocket with probe simulations were mostly unsuccessful, as only resorcinol was able to transiently open the cavity.[23] All in all, this represents an excellent target to test and compare our new approach.

In Fig 2 a, violin plots with the distribution of the relative pocket volume as calculated with Epock[42] are shown for the different NPC2 simulations. The protein core is tight in the apo structure, reaching no more than 30% of its maximum opening during the 0.1 μs of plain MD in water. In particular, it can be observed how 0.3 μs of mixed-solvent simulations are not sufficient to open the cavity. In most cases, the pocket remains closed, with a couple of exception reaching up 40% of the open pocket volume. The most external area, close to the βD and the $\beta E - \beta F$ loop, where the hydrophilic head of sterols would normally reside, is easily accessible to benzene molecules in all simulations, but the probes can only bind su-

perficially under these circumstances. Analyzing the violin plots of simulations with SWISH alone and those with SWISH with benzene probes (Fig 2 a) it is clear that in this target the combination of SWISH with probes is the most effective in increasing the accessible volume. This is consistent with previous observations,[36] where the most successful cryptic site opening was obtained when SWISH was coupled with hydrophobic fragments. The combination of the two approaches works best than any of the two alone because SWISH facilitates the opening of the cryptic pockets, allowing the probes to bind and in turn stabilize the cavities themselves. In NPC2, higher accessible volumes correspond to conformations where the βD strand, left free to move, detaches from the main protein body. In the white replica, at $\lambda = 1.0$, the volume size mostly fluctuates around 10 and 20% in values reaching up to almost 60%. The two peaks correspond to closed conformations without and with a bound benzene in the β-sandwich respectively. The probe density difference (Fig. 3 a) further clarifies the situation. In this case, a fragment binding hotspot is found deep within the β-sandwich, far from more superficial areas sampled by the probes without SWISH. This is reflected also in the increase in frame occupancy when counting the number of full fragments within the sampled volume radius as shown in the bar plot of Fig. 3 right.

In agreement with the holo crystal structure, Y100 and F66 move out of the way during these simulations, granting access to ligands inside the pocket, while minor transient rearrangements are observed within the β-sandwich in the presence of benzene fragments, as in the case of Y36 and L94, whose side chains are facing the protein core. Benzene rings are tightly wedged within the β-sandwich, and trapped when the closed βD and $\beta E - \beta F$ loop conformation from in the white replica.

p38α

The second system we chose, p38α, belongs to the mitogen-activated serine-threonine protein kinase enzymes (MAPK) family. It is the prototype of four mammalian isoforms and is involved in regulating numerous cell cycle processes, apoptosis, and tumor suppression. Its activation is stimulated when cells are subject to stress, shock or inflammation signals,

7

and abnormal activity can lead to chronic inflammatory diseases and cancer.[39,43] Four different activation modes have been observed so far, one of which is caused by the interaction with phosphatidylinositol ether lipid analogs which leads both to autophosphorylation and activation of p38α. These lipids bind to a cryptic pocket far from both the ATP binding site and the DFG motif, and close to the MAPK inset (MKI), defined by $\alpha 1L14$ and $\alpha 2L14$ and the loop interconnecting them.[39,44] While most holo structures available in the PDB database present inhibitors bound to the orthosteric cellular ATP binding site, the structure crystallized by Xing and al. in 2009 has a diaryl pyrazole compound (I46) in the MKI site.[45] The binding region is not visible in the apo structure, as the αE-αF linker covers it entirely. At variance from the buried pocket observed in NP2, the MKI site is more superficial, but it requires backbone rearrangements and a flip of the W197 side-chain in the αE-αF loop (shown in Fig. 1 b), making it an interesting system to test the capabilities of our methodology. In this case, we observe higher variability in the success of probes MD in opening the pocket. In some cases, the benzene is able to displace the αE-αF loop and increase the volume up to its maximum capacity, with an almost flat distribution, but in others, the pocket is closed, maintaining the volume under 40% simulations (see Fig. 2 b left). The MD simulation in water alone is able to reach up to 60%, suggesting a cavity opening mechanism compatible with a conformational selection and an evident sampling problem in the probe simulations. On its own, SWISH is able to induce the necessary displacement of the MKI from the rest of the protein body and the consequent removal of the αE-αF loop from the cryptic site in a concerted motion. In Fig. 2 b center, we can observe how in the white replica, the volume opening can reach 80%. The combined approach is comparably successful (Fig. 2 b right). The frames distribution of the white replica in the simulations with SWISH and probes is characterized by two equiprobable volume values at $\sim$20% and $\sim$40%. In these situations, the αE-αF loop partially covers the binding site, respectively without or with benzene probes bound inside. Looking at the bar plot of Fig. 3 right, we can see how the cavity is much more accessible than in the case of NPC2, having at least one

8

fragment (sometimes more) within the volume radius for at least half of the simulations on average. The standard deviation is however quite large ($\sim 25\%$) in line with what observed in the volumes analysis. SWISH improves the average occupancy and thus the sampling of the pocket, but again some variability is observed. This was expected due to the high flexibility of the MKI area and suggests the need for longer simulations.

To allow the binding of benzene, the aromatic side chains of W197 and H199, pointing directly inside the pocket in the apo conformation, need to be displaced. While removing H199, partially exposed to solvent, allows for an easier access to the site, W197 needs to flip outward, in order to free enough space for the aromatic probes. The opening of the site and the removal of the histidine and tryptophan, also allows the M198 side-chain to transiently move internally and point toward the MKI helices. The advantage of scaling the hydrophobic interactions is markedly evident when analyzing the volumetric density difference against the $\lambda = 1.0$ SWISH replica, in Fig. 3 b. A well defined hot-spot of benzene occupancy is observed at the cryptic site, on the diaryl pyrazole binding coordinates. Previous works with molecular probes were also able to identify the region when using isopropanol but didn't observe a complete displacement of W197 and H199 as in this case.[23]

LfrR

The transcription factor LfrR is a member of the TetR family of transcriptional repressors and modulates the expression of the LfrA multidrug efflux pump in the *Mycobacterium smegmatis*. Efflux pumps regulate the egress of antibiotics and other compounds recognized as foreign from the bacterial cells; over-expression of these proteins, by the deletion of LfrR, has been associated with antibiotic resistance.[46] LfrA was first the first efflux pump to be described in mycobacteria, a family which includes *Mycobacterium tuberculosis*, responsible for tuberculosis, among its members.[38] Both apo and holo X-ray crystal structures are comprised of a ~ 350 residues asymmetric homodimer, with most of the differences located in the $\alpha 6$ and $\alpha 7$ helices (see Fig. 1 c left), close to the proflavine (PRL) binding site, which can be either helical or disordered. Each homodimer could in principle bind a molecule of

proflavine independently, but no evidence of double binding was observed experimentally;[38] it is unclear if this feature is dependent on the original asymmetry of the apo structure. Interestingly, the presence of the ligand can further affect the structure of these segments; the disordered $\alpha6$ loop needs to be removed from inside the binding pocket to make space for the ligand, hiding the site in the apo crystal, forcing also a rearrangement of $\alpha7$. Furthermore, these areas assume helical conformations in both monomers with the ligand bound. Their high structural variability is expected to modulate the DNA-binding properties of the repressor, through their proximity to $\alpha4$ and $\alpha1$.[23,38]

Comparing the available apo and holo crystals as in Fig. 1 c right, the helical content of $\alpha6$ seems to be most affected by the presence of the ligand in the site. Y106 and Y107 in the $\alpha6$-$\alpha7$ linker represent the biggest obstacle to the binding. With PRL, the loop is displaced outward and while Y106 engages in hydrogen bonds with the ligand, Y107 binds to the side-chains of the complementary monomer (e.g. G156).[38]

This system represents an interesting challenge as previous works with isopropanol probes were not able to displace the unraveled $\alpha6$ and $\alpha7$ helices occluding the pocket.[23]

In this case, SWISH with pure water seems to be insufficient to induce and stabilize the cryptic pocket, and no substantial increase of cavity volume from the MD in water is observed in the central panel of Fig. 2 c. The mixed-solvent MD simulations were sometimes successful. The accessible volume goes from 10% of its maximum opening (slightly more than what observe with MD in water) to up to 60-80% in a few cases. In stark contrast, SWISH combined with the benzene fragments is very effective in inducing the opening of the cavity. Indeed, all of the SWISH replicas in the presence of probes are able to sample a volume opening of up to 80%, with a peak in the density distribution at about 35%-40%, typical of configurations where the $\alpha6$-$\alpha7$ loop partially closes back around the bound benzene rings. A high sampling of the volume area both with or without SWISH is reflected in the fragments occupancy graph (Fig. 3 right), where a 15-20% improvement is still observed in SWISH replicas, where up to three fragments can enter the pocket. The cavity is free

of probes for less than 5% of the time and good consistency ($\sim$5% standard deviation) is observed on 0.1 μs blocks, against almost 20% of standard deviation in the mixed-solvent simulations without SWISH.

Supporting these data, the density difference (see Fig. 3 c left) once again shows a substantial improvement of the cryptic cavity opening with SWISH and fragments. The limited size of benzene does not require a complete rearrangement of the α6-α7 loop during the binding but causes the side chains of its residues to reorient. In particular, Y107 points now toward the protein core and occasionally forms π-π interactions with the organic probes in the pocket. On the other hand, Y106 is still close to the original orientation observed in the apo crystal, but in a few cases attempts to flip underneath the unraveled loop. Completing this rotation would allow the side-chain to approach the complementary subunit, in line with the holo crystal conformation.

Fig. 3 c shows how the fragments density hotspots approach the PRL binding site on both sides. In particular, SWISH allows for the aromatic rings to bind between the α6-α7 linker and α4 and induce the rearrangement of α6, while previous probes simulations (even when performed with different probes) were unsuccessful in displacing the loop occluding the PRL binding cavity.[23] Given this result, we can conclude that the shift of α6 against α4, induced either by the natural substrate PRL or by a combination of SWISH with probes, is crucial for the opening of the pocket. This is further supported by the fractional occupation of PRL in the dimeric holo crystal. Indeed an elongated α6 is observed in the unbound monomer[38] which may impede its displacement. We also tested the effect of the constraints on hindering the movement and preventing the binding of fragments to the cavity. As there are two sites (one bound and one unbound) in the crystal structure, we tried to maintain the inherent asymmetry when defining the contact maps for the constraints: one cavity was let free to fluctuate and the other was kept comparatively fixed by the wall on the contact map. As expected, the results were strikingly different. While the benzene fragments bound abundantly (see Fig. 3 c) to the constraint-free monomer, no opening of the cavity is observed in

11

the more rigid subunit.

hPNMT

The human phenylethanolamine N-methyltransferase (hPNMT) acts as a catalyst for the synthesis in vivo of adrenaline (epinephrine) from noradrenaline. Given the proposed roles that adrenaline plays in the mammal central nervous systems, including in the regulation of blood pressure, body temperature, and respiration, as well as the modulation of the pituitary gland and hypothalamus functions, this protein has been subject to intense scrutiny in recent years.[40] hPNMT has been thus successfully co-crystallized with S-adenosyl-l-homocysteine (SAH) and a number of different inhibitors.[40,47,48]

Matin et al.[40] were among the first to produce a bound (holo) crystal structure of hPNMT in complex with both SAH and the inhibitor SK&F 29661 (PDB code: 1HNN). Their work was followed by numerous crystallization attempts with alternative inhibitors, occupying the same binding site to different degrees; these structures are isomorphous of 1HNN with the exception of local deformations and side chains rearrangements to accommodate the different ligands.[48] To our knowledge, however, no apo structure is available in the PDB database to date. These compounds bind to the orthosteric active site, and the absence of information on the apo structure is a clear complication. However, the conformational changes and the volume variability that has been observed when comparing the inhibitor-bound structures have been used as evidence for the cryptic nature of the cavity. Thus hPNMT has been used to test other algorithms for cryptic pockets such as JEDI.[34] From 1HNN to 2G8N, the pocket volume increases when the protein is in complex with a selective inhibitor (3-Hydroxymethyl-1,2,3,4-tetrahydroisoquinoline or F83), which requires the displacement of the K57 side chain, and a small movement of the $\alpha 3$ helix (see Fig. 1 d left).[47,49] The two ligands share an aromatic isoquinoline, which binds under $\alpha 2$ and the $\beta H1$-$\beta H2$ hairpin, while the displacement of K57 is induced by an extra chlorophenyl in F83 linked to the sulfonate group. While the 1HNN structure was chosen as the starting structure, the larger 2G8N, in complex with F83, was used as reference ligand to evaluate the pocket volume.

The protein core is composed of a central 7-fold mixed β sheet flanked by helices on both sides. The site is deeply buried, covered by a tight lid comprised of different motifs, $\alpha 1$-$\alpha 2$ and $\alpha 6$ helices in the N-terminus, the $\beta H1$-$\beta H2$ hairpin and the ω loop (highlighted in Fig. 1 d left).[40] Surprisingly, only a minor rearrangement in these motifs (1.4 Å of RMSD of difference on the Cαs) was observed after equilibrating the system when starting from the holo structure deprived of both ligands. The cavity is however occluded by the repositioning of a few side chains. K57, in the $\alpha 3$ helix, N39 and Y40 in the $\alpha 2$, and Y209, belonging to $\beta H1$, which move away from the cavity core to make space for the inhibitors in the holo state (in Fig. 4). It is important to notice that the elongated side chain of K57 reduces by steric occlusion the cavity also when shorter inhibitors are present, as in the case of 1HNN, while it is pushed aside by the chlorophenyl of F83 as observed in 2G8N. It has been hypothesized that the removal of the $\alpha 2$ $\alpha 3$ linker covering the cavity, culminating in P42 and P43, highly conserved among PMNTs (see Fig. 1 d right), may act as a gating mechanism to the ligand binding.[40] However, fluctuations in the position of N-terminal helix $\alpha 1$, which is close to the SAH binding site, but is not tightly attached to the protein body, were observed already during the 0.1 μs NVT simulation with water, without the need of rescaling the hydrophobic interactions. These fluctuations were dampened when the NVT simulation was repeated in the presence of SAH. We thus chose not to include this part in the definition of the contact map wall used in the simulations with fragments and SWISH, as its movement may help to displace $\alpha 2$ and the loops nearby, and open the cavity to the fragments, as an alternative to the first gating hypothesis proposed.

The elusive nature of the cavity found in hPNMT, as well as its structural features (it is deeply buried and movements of numerous secondary structures take part in the binding), make it a particularly difficult and interesting test case.

During the 0.1 μ s MD simulation in TIP3P water, the deeply buried cavity volume fluctuates around 15% of its maximum capacity (see Fig. 2 d). Adding benzene probes improves the situation slightly with an increase to 20% in most replicas and up to 60-70% in a few

exceptional cases. As for the p38α, SWISH alone is able to induce the opening of the pocket, by means of fluctuations in the $\alpha 1 - 3$ helices and the complementary ω and $\beta H1$-$\beta H2$ loops. A conformational selection mechanism may be at play here, but the absence of SAH, which has been shown to stabilize the helical regions, could also justify this behavior. When the scaling reaches $\lambda 1.35$, the ω-loop and the $\beta H1$-$\beta H2$ hairpin detaches from $\alpha 2$, which is not subject to constraints on the contact map. This leads to a full opening of the cavity and increase in solvent accessibility. The exchanges through the HREX algorithm, however, allow recovering normal conformation. Interestingly, this effect is dampened when benzene probes are added, as their presence inside the cavity keeps the motifs closer together at all lambdas. A pocket opening of no more than 60% of its maximum volume is sufficient to allow the probes to enter the binding site. Once inside, the lid closes the pocket again. This suggests that displacement in the ω-loop and the $\alpha 2$-$\alpha 3$ linker, rather than the $\alpha 2$-$\alpha 3$ loop may act as gating mechanism when inhibitors approach the protein.

A similar picture is observed in the bar plot of Fig. 3 right, where no more then one ring is observed within the monitored volume area for up to 20% of the time on average in the mixed-solvent MD runs, with an extremely wide standard deviation of 30%, showing that in some simulations the aromatic probes do not reach the pocket at all. This percentage improves to 40% when SWISH is activated and more consistently within 0.1 μs blocks, the standard deviation reaching only 20%.

Benzene fragments tend to occupy the isoquinoline binding area, as highlighted by occupancy hotspots in Fig. 3 d. On the other hand, no fragments are found wedged between $\alpha 2$ and $\alpha 3$, as for the F83 chlorophenyl. The area is well sampled by the probes during the SWISH simulations, but they do not seem to stay at lower λ values as the pocket closes. This shows the need for a sulfonate or other charged groups to displace the charged K57 side-chain.

Discussion

The variety of systems and cryptic cavities presented here, differing in sizes, accessibility level and conformational changes upon binding, would normally require different computational approaches specifically tailored to reveal each cavity. For instance, small conformational changes stabilized by a conformational selection mechanism might be ideally suited to mixed solvent MD (as in the case of p38α). While larger scale conformational changes can be effectively explored by CV-based enhanced sampling algorithms. However, when little is known about the cryptic binding site and its opening mechanism, a general and robust approach is needed. Here we have shown how SWISH combined with mixed solvent simulations can quickly test the presence of druggable hydrophobic sites in a variety of systems even when no previous knowledge of the pocket position and opening mechanism is available.

For example, benzene probes were found wedged deep in the β strands core of the NPC2 cholesterol transport, an area suitable to accommodate the hydrophobic end of sterols. The conformational change requires the opening of rimming loops and the removal of Y100 and F66 side chains, followed by the displacement of a few residues internal to the β structure, leading to a modest increase of volume. Reducing hydrophobicity against the solvent allows for an increase in the sampling of regions deeper in the protein core when compared to standard mixed-solvent MD runs. This methodology fares well when compared with simple fragment simulations; binding to the MKI site of p38α was observed in previous fragments MD simulations with isopropanol but the flip of W197 and H199 was incomplete.[23] SWISH combined with probes allows not only the displacement of the αE-αF loop covering the site but also of W197 and H199. Similarly, in the case of LfrR efflux pump repressor, the displacement of the $\alpha 6$-$\alpha 7$ linker is a necessary step to allow the binding of the probes. When in its extended conformation, the loop occludes the PRL binding site and has proven problematic to deal with simple fragments MD simulations. Our method was able to induce the displacement of this lid and allow the binding of the probes. Movements in the $\alpha 6$ helix promoted by the hydrophobic scaling may play a crucial role in the cavity emergence. This

hypothesis is supported by the absence of binding in one of the two subunits, where the relative position of $\alpha6$ to $\alpha4$ is constrained. Finally, the method was tested on a more complex case, characterized by a deeply buried pocket and limited experimental information. To work with hPNMT we had first to recover a plausible apo structure, as all crystals in the PDB database are bound to inhibitors. Small aromatic probes were able to find the inhibitors binding site, which required the opening of the ω-loop and the linker between the $\alpha2$ and $\alpha3$ helices and conformational changes in a number of side-chains occupying the cavity. Marked differences in the pocket volume behavior are observed when comparing the plots of Fig. 2. In NPC2 an induced-fit mechanism might be at play, with a slight increase in volume ratio when the fragments approach the site, at the other end of the spectrum, p38α pocket fluctuates widely, reaching up to 60% of its maximum volume during the MD simulations alone, suggesting that conformational selection plays a role to a certain degree in this particular protein. Finally, the pocket opening mechanism observed in LfrR and hPNMT seems to be compatible with a mixed mechanism.

The method performs consistently well on four significantly different cases, some of which were already successfully investigated with carefully selected probes, while others were still lacking a full characterization. Interestingly, short SWISH with probes simulations (0.1 μs) with 6 replicas perform better than 20 independent probes simulations each lasting 0.3μs; in LfrR, the difference is remarkable. One can thus easily appreciate the speed and efficiency of our proposed approach.

The method is very suitable to explore and quickly identify hidden ligandable cavities in the target proteins. The use of walls on contact map differences from the native fold allows to increase the scaling, which may be necessary for some of the most buried pockets, while maintaining the fold of the protein intact. On the downside, however, they require some knowledge of the system as either the approximate area of binding needs to be excluded from the contact map restraints, or the value of the wall has to be sufficiently large to allow for a complete conformational change upon binding. This was the case of LfrR, where a less

careful choice of restraints may prevent the opening of the cavity and fragment binding. At the same time, LfrR multidomain structure is highly sensitive to the hydrophobic interactions scaling and without any wall on the contact map, the helical parts of the protein tend to drift from the main body even at relatively low lambda values. A possible approach for such cases is to start with an initial general exploration of hydrophobic patches without restraints and low scaling and, once possible candidate areas have been identified, to follow the protocol described here to refine the search and fully investigate the cavities. Alternatives to the contact map, such as residue pair coordination or covariance-based techniques, could also in principle also be used. The trade-off among complexity, specificity, and cost should be however properly evaluated when making this choice. In our simulations, we observed that the contact map is an effective choice and more efficient than the commonly used RMSD. The position density difference of the probes in simulations with and without SWISH (shown in Fig. 3 left) provides a useful indication of the presence of real cryptic pockets to be confirmed by a follow-up druggability analysis, and should distinguish them from false positives. The issue of false positives has been also addressed in Ref. 36, where SWISH was applied to ubiquitin, which does not harbor any hidden cavity. As expected, no major density hotspots were observed in such control system. Similarly, other than the validated target cavities, no relevant hotspots within the protein folds were found in any of the four systems here described.

Interestingly, our results were robust against excessive lambda factors, allowing for a certain degree of flexibility in this choice. When local unfolding was observed in the cryptic site area, the replica exchange algorithm allowed to recover secondary structure motifs at lower scaling factors.

It is surprising to see how a single simple fragment, benzene, works effectively with SWISH in identifying a deeply buried binding site such as in the case of NPC2, where classical MD simulations gave mixed results depending on the probe used. This bodes well for a more targeted approach, using a varied choice of probes with different physicochemical properties,

and scaling replicas tuned to the cavity depth and hydrophobicity, when characterizing in detail a system and testing the effective druggability of its cavities. A good example where this might help is hPNTM, where the benzene probes were able to correctly identify the binding site of isoquinoline, but were easily blocked by the charged lysine.

Conclusion

We presented here a general and reliable protocol to determine the presence of hidden druggable cavities based on a refined version of our enhanced sampling method SWISH and benzene probes. We tested it on four different and challenging targets, characterized by experimentally known pockets difficult to approach with other computational strategies. The results confirm that the method is indeed able in all cases to correctly identify the position of the sites and sample the conformational changes regardless of the binding mechanism while maintaining the correct fold of the protein. We here improved on the original SWISH-based approach, further proving it as a robust and general tool to reveal druggable cryptic cavities.

Methods

With the exception of hPNMT, the cryptic sites here investigated were selected from a more comprehensive database recently redacted by Cimermancic et al., CryptoSite.[35] The database provides Protein Data Bank[50] codes for both apo and holo structures, in particular 1NEP[37] and 1HKA[41] for apo and holo of NPC2 respectively, 2WGB and 2V57[38] for LfrR, and 4E5B[39] and 3HL7[45] for p38α. In the case of hPNMT, no known apo experimental structure was available at the time of this study. As reported in more detail the text, we decided thus to start from the holo crystal structure 1HNN published by Martin et al. in 2001,[40] while 2G8N[47] was used as a comparison to identify position and shape of the pocket volume. Any exogenous compound was removed from the PDB file and the structure was

subject to a standard setup like the other systems. A 0.1 μs-long MD simulation was run to test the pocket stability. During this relaxation, a slight collapse of the inhibitor binding pocket volume was observed, as detailed in the text, and it was taken as apo reference for the cavity volume comparison (see Fig. 4).

Once an adequate apo structure was identified for each system, missing residues were added where needed with SwissModel[51] according to the respective UniProt sequences. Standard protonation and rotameric states were chosen for these models, which were then embedded in 10 Å buffer of TIP3P water molecules and counterions to ensure charge neutrality. Possible atomic clashes and inconsistencies in the structure were removed through 5000 steps of geometrical optimization with steepest descent and 500 steps of conjugated gradient, followed by 200 ps of gradual heating to 300 K and 10 ns of NPT equilibration. During the equilibration stage, the Parrinello-Rahman[52] barostat maintained a target pressure of 1 bar isotropically, while a velocity rescale thermostat[53] was applied throughout all the simulations. The same minimization and equilibration protocol were applied to the mixed-solvent counterparts, which were prepared starting from the equilibrated MD structure. In this case, $\sim$1 Mol of benzene molecules were added to the solvent. The ligands were parametrized with Amber Generalized Force Field (GAFF)[54] with an inter-ligand repulsion term added to the non-bonded interactions to prevent clustering and other artifacts and maintain a diffusive behavior. Charges were calculated with Gaussian09[55] at the Hartree-Fock level with a 6-31G(d) basis set. All the MD simulations were run with Gromacs 5.1.4[56] and the Amber FF14SB[57] force field, in the NVT ensemble. The PME-Switch algorithm was used for electrostatic interactions with a cut-off of 1 nm, while a single cut-off of 1.2 nm was used for Van der Waals interactions. The HREX implementation of Plumed 2[58] plugin for Gromacs was chosen to run SWISH. Exchanges between replicas were attempted every 2ps.

Epock[42] was used to monitor the accessible volume of the cryptic pockets, while the rest of the analysis was carried out with the help of MDTraj[59] library for Python 2.7. Pymol 1.8[60] was used to obtain images of the structures and volume maps were produced with VMD

1.9.2.[61] These were obtained as the difference in the average density of the benzene $C\alpha$s between the molecular probes simulations with and without SWISH.

SWISH protocol. A common methodological protocol was applied to study and compare the efficacy of SWISH on the four different systems. 20 independent replicas of the water and benzene mixed-solvent boxes were simulated for 0.3 μs each, while 0.1 μs of plain MD with only TIP3P waters was run as a control system. SWISH was then applied first in the absence of benzene, for 0.5 μs with 6 HREX replicas and a scaling factor λ between 1 and 1.35. Similarly, 6 HREX replicas with probes were run for 1.0 μs each, with the same parameters.

Restraints were applied to all proteins, to prevent the general unfolding at higher λ values. A contact map was applied to monitor the distances between pair of key representative atoms belonging secondary structures of the protein according to the Timescapes[62] definition. The pairs were chosen looking at the most consistent contacts between equilibrated apo and holo structures, so to exclude any region that underwent natural conformational changes, and were limited to ~100-200 in number to ease the computational burden. The movements of these atoms were restrained to abide by average and fluctuations observed during a simple MD simulation with a spring constant of 3000 kJ/(mol·nm). An area of ~6-8 Å around the alleged cryptic pocket (as defined by the center of mass of the ligand in the holo structure) was left free of these constraints, allowing for local rearrangements of secondary structures upon ligand binding. In the case of LfrR, extra constraints on long-range domain-domain interactions were applied to maintain the fold.

Violin plots were produced monitoring the accessible volume centered at the center of mass of the ligand residue neighbors within ~6-8 Å and with radius chosen upon visual inspection to best fit the situation. Occupancy bars were calculated counting the probes falling in full within the same radius. As an exception, the resulting C3S center of mass is relatively superficial and the volume center was thus positioned deeper in the NPC2 β-sandwich, to

assure the exclusion of superficial interfaces.

Acknowledgement

References

(1) Hopkins, A. L.; Groom, C. R. The druggable genome. *Nat. Rev. Drug Discovery* **2002**, *1*, 727–730.

(2) 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* **2013**, *503*, 548–551.

(3) Wenthur, C. J.; Gentry, P. R.; Mathews, T. P.; Lindsley, C. W. Drugs for allosteric sites on receptors. *Annu. Rev. Pharmacol. Toxicol.* **2014**, *54*, 165–84.

(4) Johnson, D. K.; Karanicolas, J. Selectivity by Small-Molecule Inhibitors of Protein Interactions Can Be Driven by Protein Surface Fluctuations. *PLOS Comput. Biol.* **2015**, *11*, e1004081.

(5) Cossins, B.; Lawson, A. Small Molecule Targeting of ProteinâĂŞProtein Interactions through Allosteric Modulation of Dynamics. *Molecules* **2015**, *20*, 16435–16445.

(6) Monod, J.; Wyman, J.; Changeux, J.-P. On the nature of allosteric transitions: A plausible model. *J. Mol. Biol.* **1965**, *12*, 88–118.

(7) Csermely, P.; Palotai, R.; Nussinov, R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem. Sci.* **2010**, *35*, 539–46.

(8) Beglov, D.; Hall, D. R.; Wakefield, A. E.; Luo, L.; Allen, K. N.; Kozakov, D.; Whitty, A.; Vajda, S. Exploring the structural origins of cryptic sites on proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E3416–E3425.

(9) Wells, J. A.; McClendon, C. L. Reaching for high-hanging fruit in drug discovery at proteinâĂŞprotein interfaces. *Nature* **2007**, *450*, 1001–1009.

(10) Campbell, S. J.; Gold, N. D.; Jackson, R. M.; Westhead, D. R. Ligand binding: functional site location, similarity and docking. *Curr. Opin. Struct. Biol.* **2003**, *13*, 389–95.

(11) Nisius, B.; Sha, F.; Gohlke, H. Structure-based computational analysis of protein binding sites for function and druggability prediction. *J. Biotechnol.* **2012**, *159*, 123–134.

(12) Ludlow, R. F.; Verdonk, M. L.; Saini, H. K.; Tickle, I. J.; Jhoti, H. Detection of secondary binding sites in proteins using fragment screening. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 15910–5.

(13) Erlanson, D. A.; Braisted, A. C.; Raphael, D. R.; Randal, M.; Stroud, R. M.; Gordon, E. M.; Wells, J. A. Site-directed ligand discovery. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 9367–72.

(14) Horn, J. R.; Shoichet, B. K. Allosteric Inhibition Through Core Disruption. *J. Mol. Biol.* **2004**, *336*, 1283–1291.

(15) 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.* **2012**, *109*, 11681–11686.

(16) Hart, K. M.; Moeder, K. E.; Ho, C. M. W.; Zimmerman, M. I.; Frederick, T. E.; Bowman, G. R. Designing small molecules to target cryptic pockets yields both positive and negative allosteric modulators. *PLOS ONE* **2017**, *12*, e0178678.

(17) Guvench, O.; MacKerell, A. D. Computational Fragment-Based Binding Site Identification by Ligand Competitive Saturation. *PLOS Comput. Biol.* **2009**, *5*, e1000435.

(18) Seco, J.; Luque, F. J.; Barril, X. Binding Site Detection and Druggability Index from First Principles. *J. Med. Chem.* **2009**, *52*, 2363–2371.

(19) Schmidtke, P.; Barril, X. Understanding and Predicting Druggability. A High-Throughput Method for Detection of Drug Binding Sites. *J. Med. Chem.* **2010**, *53*, 5858–5867.

(20) Lexa, K. W.; Carlson, H. A. Full Protein Flexibility Is Essential for Proper Hot-Spot Mapping. *J. Am. Chem. Soc.* **2011**, *133*, 200–202.

(21) Bakan, A.; Nevins, N.; Lakdawala, A. S.; Bahar, I. Druggability Assessment of Allosteric Proteins by Dynamics Simulations in the Presence of Probe Molecules. *J. Chem. Theory Comput.* **2012**, *8*, 2435–2447.

(22) Tan, Y. S.; Śledź, P.; Lang, S.; Stubbs, C. J.; Spring, D. R.; Abell, C.; Best, R. B. Using Ligand-Mapping Simulations to Design a Ligand Selectively Targeting a Cryptic Surface Pocket of Polo-Like Kinase 1. *Angew. Chem., Int. Ed.* **2012**, *51*, 10078–10081.

(23) Kimura, S. R.; Hu, H. P.; Ruvinsky, A. M.; Sherman, W.; Favia, A. D. Deciphering Cryptic Binding Sites on Proteins by Mixed-Solvent Molecular Dynamics. *J. Chem. Inf. Model.* **2017**, *57*, 1388–1401.

(24) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562–6.

(25) Laio, A.; Gervasio, F. L. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Rep. Prog. Phys.* **2008**, *71*, 126601.

(26) Sutto, L.; Gervasio, F. L. Effects of oncogenic mutations on the conformational free-energy landscape of EGFR kinase. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 10616–21.

(27) Papaleo, E.; Sutto, L.; Gervasio, F. L.; Lindorff-Larsen, K. Conformational Changes and Free Energies in a Proline Isomerase. *J. Chem. Theory Comput.* **2014**, *10*, 4169–4174.

(28) Bunney, T. D.; Wan, S.; Thiyagarajan, N.; Sutto, L.; Williams, S. V.; Ashford, P.; Koss, H.; Knowles, M. A.; Gervasio, F. L.; Coveney, P. V.; Katan, M. The Effect of Mutations on Drug Sensitivity and Kinase Activity of Fibroblast Growth Factor Receptors: A Combined Experimental and Theoretical Study. *EBioMedicine* **2015**, *2*, 194–204.

(29) Comitani, F.; Melis, C.; Molteni, C. Elucidating ligand binding and channel gating mechanisms in pentameric ligand-gated ion channels by atomistic simulations. *Biochem. Soc. Trans.* **2015**, *43*, 151–6.

(30) Lovera, S.; Morando, M.; Pucheta-Martinez, E.; Martinez-Torrecuadrada, J. L.; Saladino, G.; Gervasio, F. L. Towards a Molecular Understanding of the Link between Imatinib Resistance and Kinase Conformational Dynamics. *PLOS Comput. Biol.* **2015**, *11*, e1004578.

(31) Marino, K. A.; Sutto, L.; Gervasio, F. L. The Effect of a Widespread Cancer-Causing Mutation on the Inactive to Active Dynamics of the B-Raf Kinase. *J. Am. Chem. Soc.* **2015**, *137*, 5280–5283.

(32) Comitani, F.; Limongelli, V.; Molteni, C. The Free Energy Landscape of GABA Binding

to a Pentameric Ligand-Gated Ion Channel and Its Disruption by Mutations. *J. Chem. Theory Comput.* **2016**, *12*, 3398–3406.

(33) Kuzmanic, A.; Sutto, L.; Saladino, G.; Nebreda, A. R.; Gervasio, F. L.; Orozco, M. Changes in the free-energy landscape of p38α MAP kinase through its canonical activation and binding events as studied by enhanced molecular dynamics simulations. *eLife* **2017**, *6*.

(34) Cuchillo, R.; Pinto-Gil, K.; Michel, J. A Collective Variable for the Rapid Exploration of Protein Druggability. *J. Chem. Theory Comput.* **2015**, *11*, 1292–1307.

(35) Cimermancic, P.; Weinkam, P.; Rettenmaier, T. J.; Bichmann, L.; Keedy, D. A.; Wold-eyes, 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.

(36) 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.* **2016**, *138*, 14257–14263.

(37) 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*, 2512–7.

(38) 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*, 7531–7.

(39) Tzarum, N.; Eisenberg-Domovich, Y.; Gills, J. J.; Dennis, P. A.; Livnah, O. Lipid Molecules Induce p38α Activation via a Novel Molecular Switch. *J. Mol. Biol.* **2012**, *424*, 339–353.

(40) Martin, J. L.; Begun, J.; McLeish, M. J.; Caine, J. M.; Grunewald, G. L. Getting the Adrenaline Going. *Structure* **2001**, *9*, 977–985.

(41) Xu, S.; Benoff, B.; Liou, H.-L.; Lobel, P.; Stock, A. M. Structural basis of sterol binding by NPC2, a lysosomal protein deficient in Niemann-Pick type C2 disease. *J. Biol. Chem.* **2007**, *282*, 23525–31.

(42) Laurent, B.; Chavent, M.; Cragnolini, T.; Dahl, A. C. E.; Pasquali, S.; Derreumaux, P.; Sansom, M. S.; Baaden, M. Epock: rapid analysis of protein pocket dynamics. *Bioinformatics* **2015**, *31*, 1478–1480.

(43) Eckert, R. L.; Efimova, T.; Balasubramanian, S.; Crish, J. F.; Bone, F.; Dashti, S. p38 Mitogen-activated protein kinases on the body surface–a function for p38 delta. *J. Investig. Dermatol.* **2003**, *120*, 823–8.

(44) Diskin, R.; Engelberg, D.; Livnah, O. A Novel Lipid Binding Site Formed by the MAP Kinase Insert in p38α. *J. Mol. Biol.* **2008**, *375*, 70–79.

(45) Xing, L. et al. Structural Bioinformatics-Based Prediction of Exceptional Selectivity of p38 MAP Kinase Inhibitor PH-797804. *Biochemistry* **2009**, *48*, 6402–6411.

(46) Buroni, S.; Manina, G.; Guglierame, P.; Pasca, M. R.; Riccardi, G.; De Rossi, E. LfrR Is a Repressor That Regulates Expression of the Efflux Pump LfrA in Mycobacterium smegmatis. *Antimicrob. Agents Chemother.* **2006**, *50*, 4044–4052.

(47) Grunewald, G. L.; Seim, M. R.; Regier, R. C.; Martin, J. L.; Gee, C. L.; Drinkwater, N.; Criscione, K. R. Comparison of the binding of 3-fluoromethyl-7-sulfonyl-1,2,3,4- tetrahydroisoquinolines with their isosteric sulfonamides to the active site of phenylethanolamine N-methyltransferase. *J. Med. Chem.* **2006**, *49*, 5424–5433.

(48) Gee, C. L.; Drinkwater, N.; Tyndall, J. D.; Grunewald, G. L.; Wu, Q.; McLeish, M. J.;

Martin, J. L. Enzyme adaptation to inhibitor binding: A cryptic binding site in phenylethanolamine N-methyltransferase. *J. Med. Chem.* **2007**, *50*, 4845–4853.

(49) Grunewald, G. L.; Dahanukar, V. H.; Caldwell, T. M.; Criscione, K. R. Examination of the role of the acidic hydrogen in imparting selectivity of 7-(aminosulfonyl)-1,2,3,4-tetrahydroisoquinoline (SK and F 29661) toward inhibition of phenylethanolamine N-methyltransferase vs the α2-adrenoceptor. *J. Med. Chem.* **1997**, *40*, 3997–4005.

(50) Berman, H.; Henrick, K.; Nakamura, H. Announcing the worldwide Protein Data Bank. âĂŐNat. *Struct. Mol. Biol.* **2003**, *10*, 980–980.

(51) Bordoli, L.; Kiefer, F.; Arnold, K.; Benkert, P.; Battey, J.; Schwede, T. Protein structure homology modeling using SWISS-MODEL workspace. *Nat. Protoc.* **2009**, *4*, 1–13.

(52) Parrinello, M. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182.

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

(54) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(55) Frisch, M. J. et al. Gaussian09 Revision D.01, Gaussian Inc. Wallingford CT. 2010.

(56) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19–25.

(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) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185*, 604–613.

(59) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **2015**, *109*, 1528–32.

(60) DeLano, W. Pymol: An open-source molecular graphics tool. *CCP4 Newsletter On Protein Crystallography* **2002**,

(61) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.

(62) Wriggers, W.; Stafford, K. A.; Shan, Y.; Piana, S.; Maragakis, P.; Lindorff-Larsen, K.; Miller, P. J.; Gullingsrud, J.; Rendleman, C. A.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. Automated Event Detection and Activity Monitoring in Long Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2009**, *5*, 2595–2605.
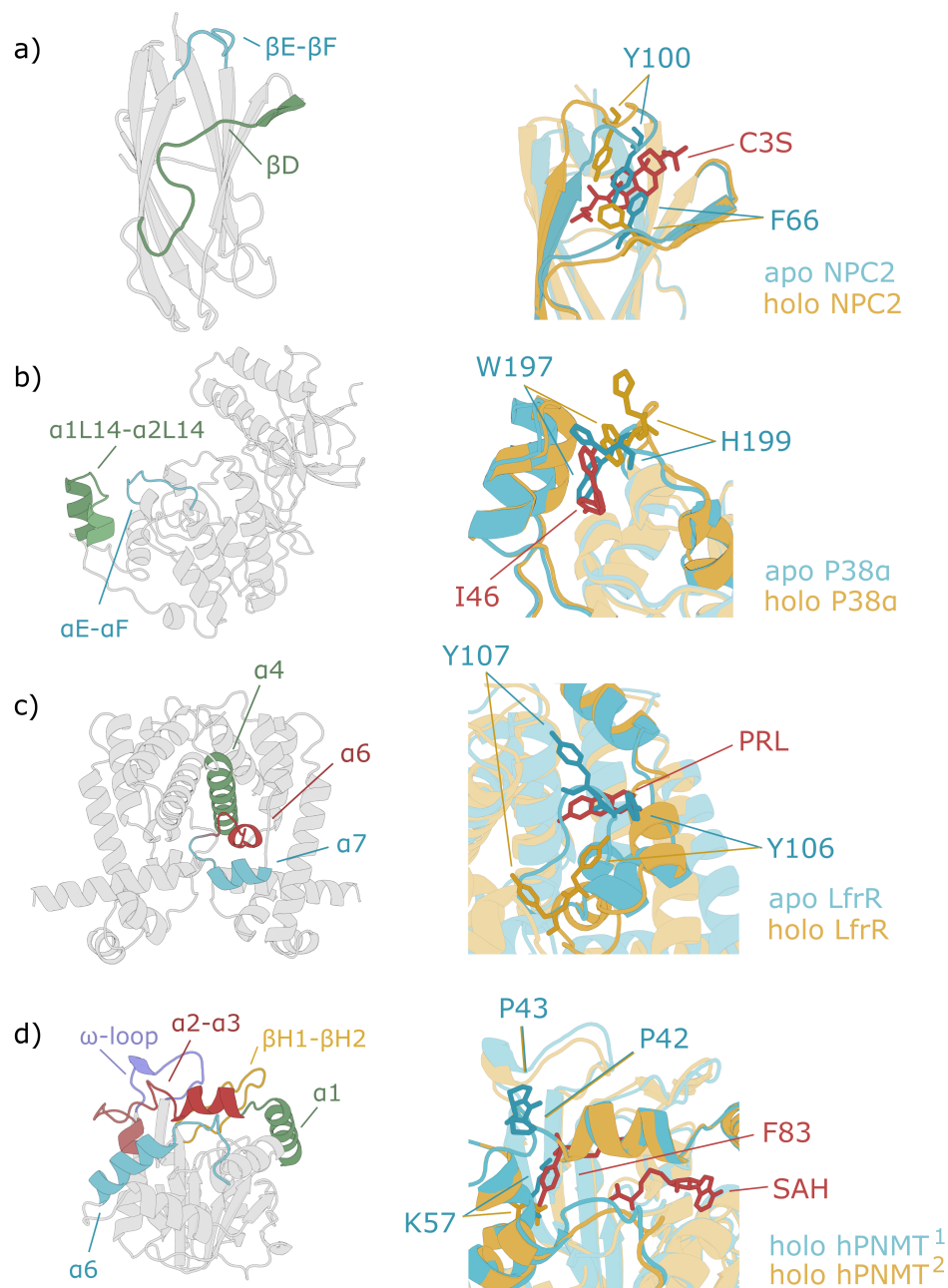
Graphical TOC Entry
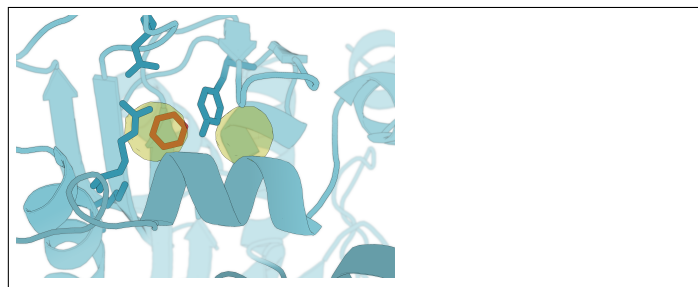
Figure 1: The four system here studied: NPC2 (a), p38α (b), LfrR (c) and hPNMT (d). On the left, Secondary structures relevant to the conformational changes observed during the emergence of cryptic pockets. On the right, comparison between apo (in cyan) and holo (in yellow) structures as cartoons (NPC2, PDB ID: 1NEP[37] and 1HKA;[41] LfrR, PDB ID: 2WGB and 2V57;[38] p38α, PDB ID: 4E5B[39] and 3HL7[45]). In the case of hPNMT the two different holo structures used (1, PDB ID: 1HNN[40] and 2, PDB ID: 2G8N[47]) are shown instead. Residues relevant to the apo to holo transition are shown as licorice. The ligands are superimposed as red licorice.
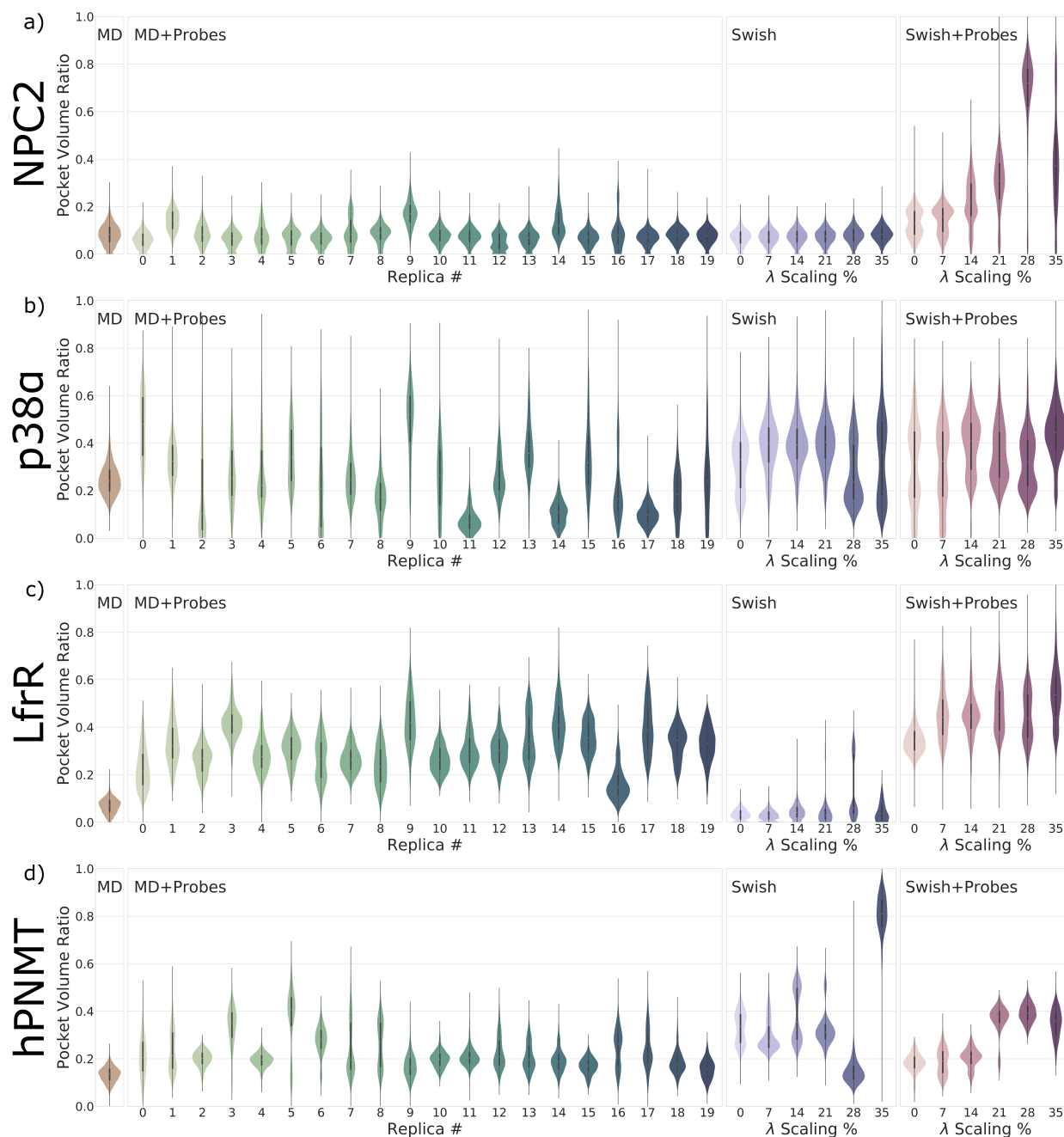
Figure 2: Violin plots representing the volume ratio for the studied cavities during the different simulations. From top to bottom: NPC2 (a), p38α (b), LfrR (c) and hPNMT (d). From left to right: classical MD (in yellow), twenty repetitions of mixed-solvent MD with benzene (in shades of green), SWISH in water (in shades of blue), and SWISH with mixed-solvent (in shades of red).
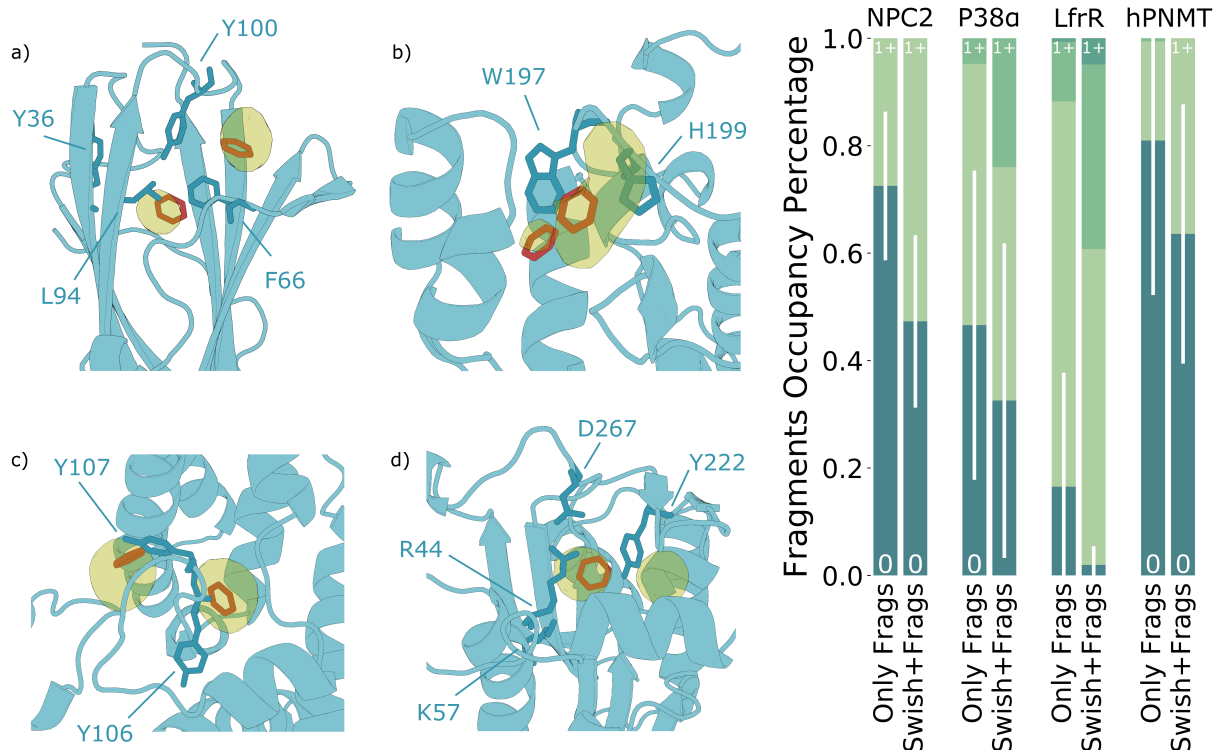
Figure 3: Left: fragment density difference maps between mixed-solvent simulations with and without SWISH obtained for (a) NPC2, (b) p38α, (c) LfrR and (d) hPNMT. In cyan are the relevant residues in an exemplary configuration, while in red are the benzene fragments in proximity to the pocket observed in the same snapshot. Superimposed in yellow are the volumetric maps. Right: average fragments occupancy of the cryptic pockets for simulations without (left) or with (right) SWISH. In dark cyan is the percentage with no probes within the area, while frames with at least one or more probes are shown in shades of green. White bars represent the standard deviation measured on the twenty mixed-solvent MD replicas and on 0.1 μs long blocks of the SWISH and probes simulations, for the frames in absence of fragments.
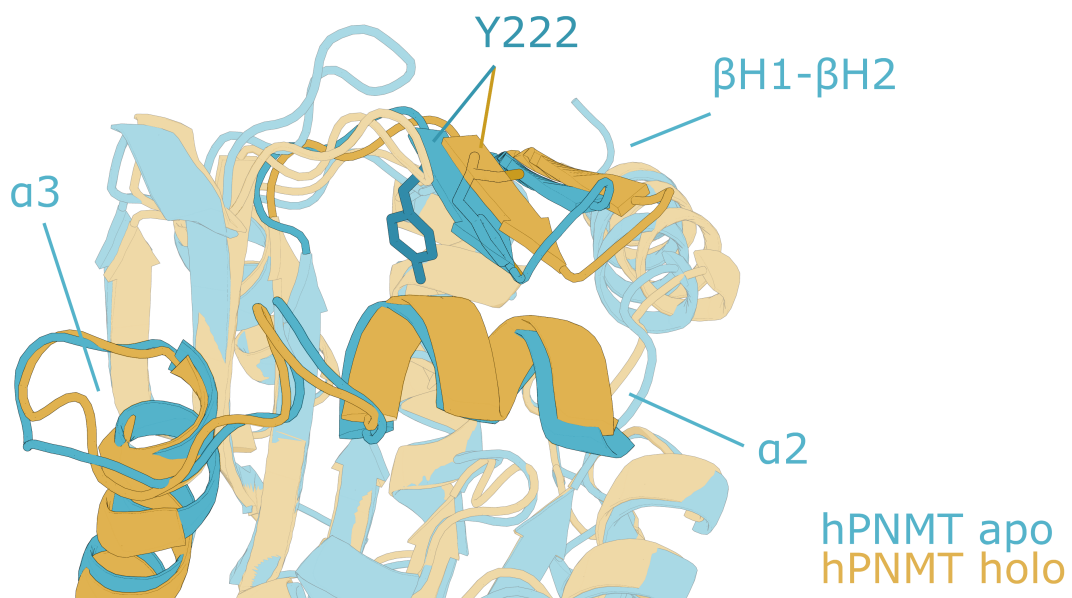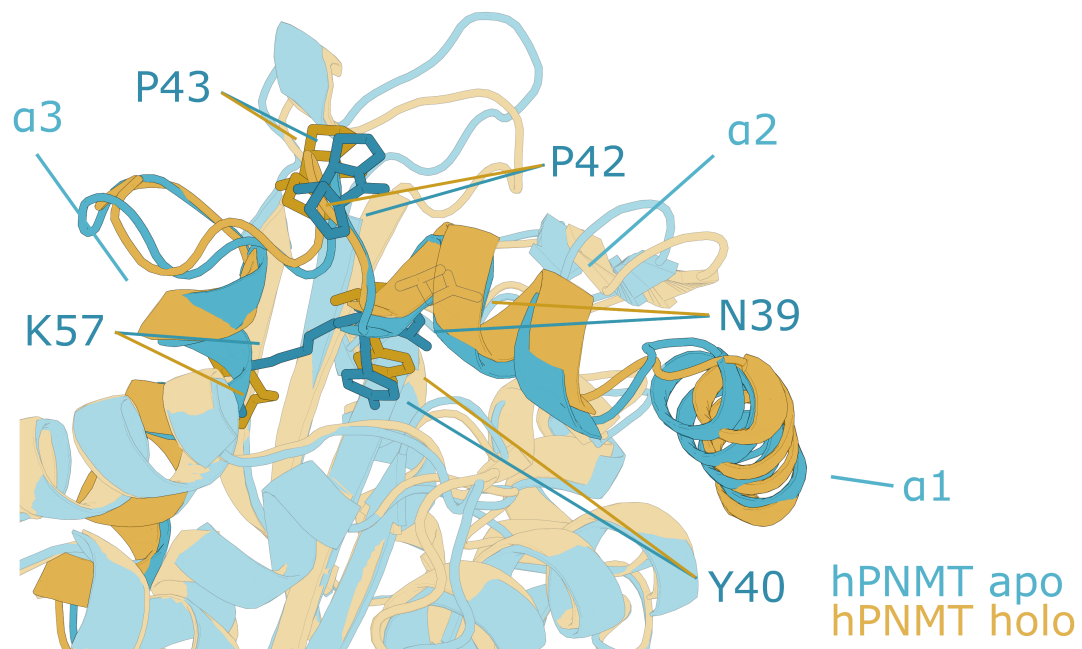
Figure 4: Comparison between the apo structure of hPNMT, obtained after a 10ns NPT equilibration, and the holo structure (PDB ID: 2G8N[47]), here superimposed as cartoon in cyan and yellow respectively. Residues relevant to the apo to holo transition are shown as licorice. The $\alpha 1$ to $\alpha 3$ motifs and the $\beta H1$-$\beta H2$ hairpin are highlighted against the rest of the protein.
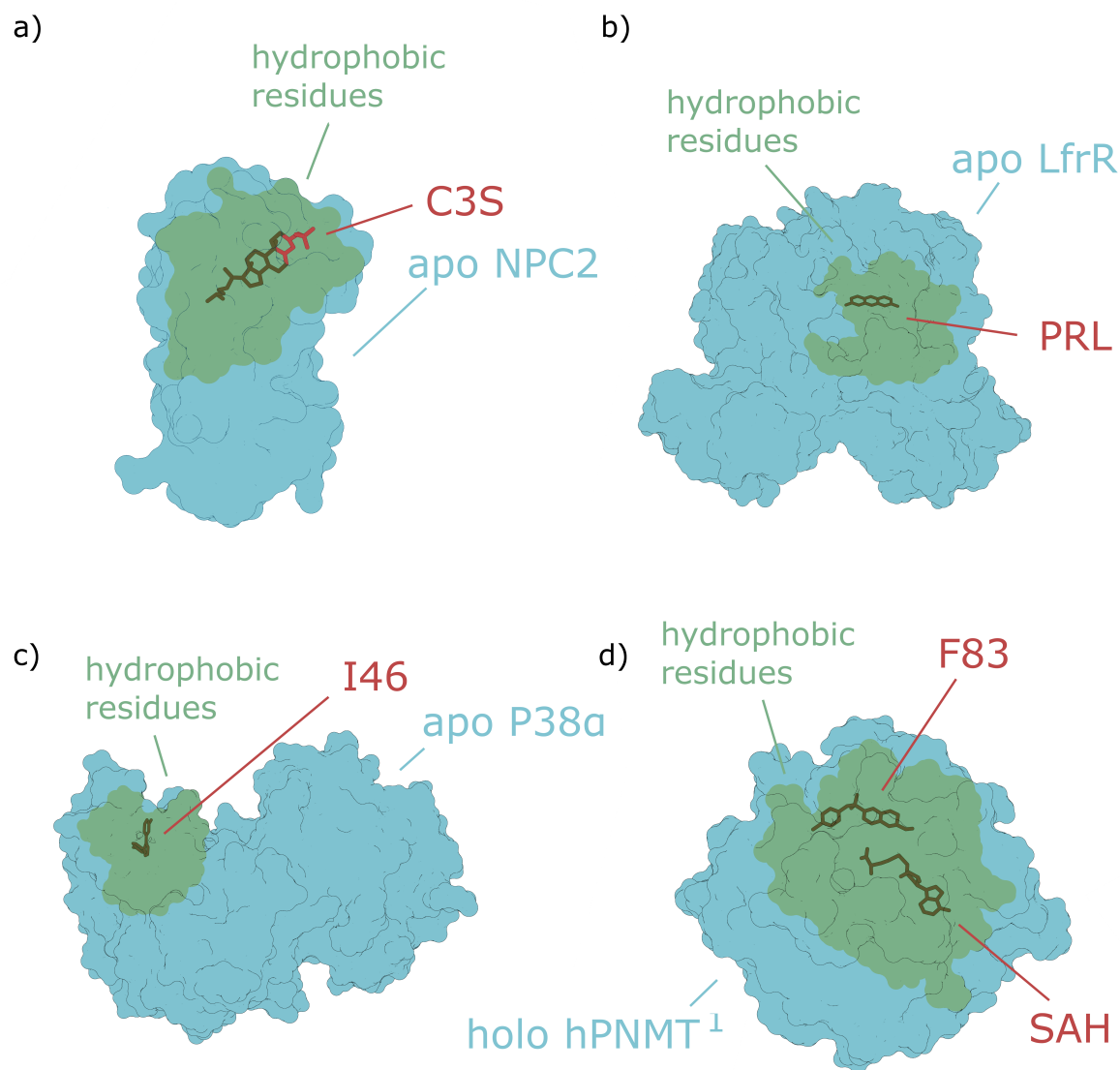
Figure 5: Surface representation of the apo structures (in cyan) of (a) NPC2, (b) p38α, (c) LfrR and (d) hPNMT, and their respective ligands (in red licorice) superimposed. In green are the surface patches of hydrophobic residues within 10 Å from the ligand.