

# Understanding Cryptic Pocket Formation in Protein Targets by Enhanced Sampling Simulations

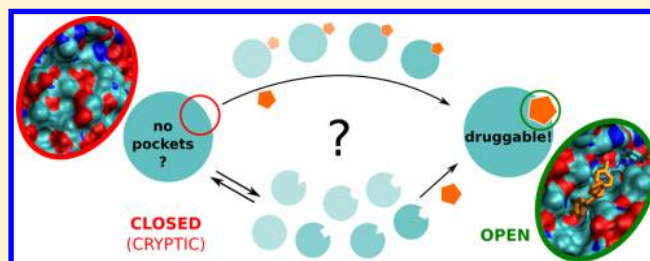
Vladimiras Oleinikovas,<sup>†</sup> Giorgio Saladino,<sup>†</sup> Benjamin P. Cossins,<sup>§</sup> and Francesco L. Gervasio<sup>\*,‡</sup>

<sup>†</sup>Department of Chemistry, and <sup>‡</sup>Institute of Structural and Molecular Biology, University College London, London WC1E 6BT, United Kingdom

<sup>§</sup>UCB Pharma, Slough SL1 3WE, United Kingdom

## Supporting Information

**ABSTRACT:** Cryptic pockets, that is, sites on protein targets that only become apparent when drugs bind, provide a promising alternative to classical binding sites for drug development. Here, we investigate the nature and dynamical properties of cryptic sites in four pharmacologically relevant targets, while comparing the efficacy of various simulation-based approaches in discovering them. We find that the studied cryptic sites do not correspond to local minima in the computed conformational free energy landscape of the unliganded proteins. They thus promptly close in all of the molecular dynamics simulations performed, irrespective of the force-field used. Temperature-based enhanced sampling approaches, such as Parallel Tempering, do not improve the situation, as the entropic term does not help in the opening of the sites. The use of fragment probes helps, as in long simulations occasionally it leads to the opening and binding to the cryptic sites. Our observed mechanism of cryptic site formation is suggestive of an interplay between two classical mechanisms: induced-fit and conformational selection. Employing this insight, we developed a novel Hamiltonian Replica Exchange-based method “SWISH” (Sampling Water Interfaces through Scaled Hamiltonians), which combined with probes resulted in a promising general approach for cryptic site discovery. We also addressed the issue of “false-positives” and propose a simple approach to distinguish them from druggable cryptic pockets. Our simulations, whose cumulative sampling time was more than 200  $\mu$ s, help in clarifying the molecular mechanism of pocket formation, providing a solid basis for the choice of an efficient computational method.



## INTRODUCTION

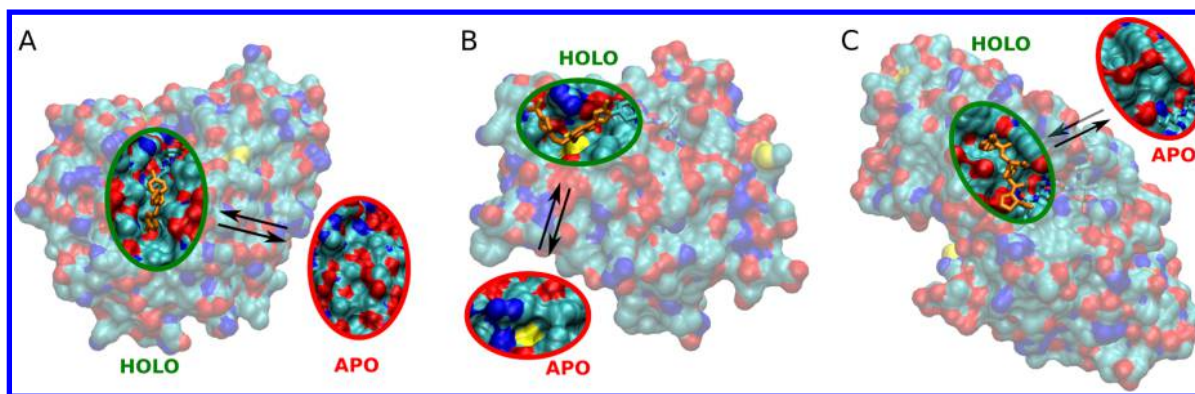
Drug discovery is an increasingly expensive endeavor as many therapeutic discovery projects fail in clinical trials.<sup>1–5</sup> Late-stage failures are especially costly, and lack of efficacy of lead compounds is often cited as the main reason for them. This problem might be addressed by using validated targets. In the postgenomic era, an increasing number of validated targets with a well-defined role in disease are arising from large-scale genomic sequencing complemented by proteome analysis.<sup>6</sup> However, not all validated targets are therapeutically tractable (or “druggable”).<sup>7–9</sup> Many attractive targets involve protein–protein interactions (PPI) or other allosteric regions lacking apparent binding sites and for a long time were considered undruggable. This notion is now changing as an increasing number of proteins previously considered intractable have been successfully targeted with small molecules, greatly extending the range of usable targets.<sup>10,11</sup> What is more, many validated targets have ligandable pockets, which are not obvious from structural information on the unbound protein and only become clear when a drug is bound. These “cryptic” or “hidden” pockets offer an attractive opportunity for therapy, as it was recently shown in the case of K-RAS, the most commonly mutated oncogene in human cancers. For 30 years,

researchers had tried and failed to develop K-RAS inhibitors, to the point it was thought to be undruggable. Yet very recently, a new cryptic pocket was found by tethered compounds and successfully targeted.<sup>12</sup> The molecular mechanism by which cryptic sites are formed is not clear. Although ligands seem to be necessary in their opening, it is still debated whether they do it by an induced-fit, conformational selection, or a “mixed” mechanism.<sup>13–15</sup> Because of these uncertainties and their hidden nature, cryptic pockets are difficult to identify by both experimental and computational methods.

Many available cryptic pocket structures, such as those studied here, were discovered serendipitously. They helped to understand the nature of these sites and their potential use as therapeutic targets. Available experimental approaches rely on expensive procedures such as large-scale fragment screening,<sup>16</sup> site-directed tethering,<sup>17</sup> or the use of antibodies.<sup>18</sup> As cryptic pockets are mostly not evolved to bind small molecules and likely have fewer possible binders as compared to catalytic sites, the discovery effort has to be more focused.<sup>19</sup> Simulations and modeling can play a significant role in the systematic search for

Received: May 26, 2016

Published: October 11, 2016



**Figure 1.** Crystal structures of TEM1 (A), IL2 (B), and PLK1 (C) complexes showing the cryptic sites bound to ligands, and their corresponding regions in *apo* structures, in which the pockets are unidentifiable. Protein surfaces are colored by element: cyan, red, blue, and yellow for carbon, oxygen, nitrogen, and sulfur, respectively. Cryptic binding sites in *holo* structures are circled in green with corresponding surface patch of *apo* structure displayed in red ovals. Ligands are shown in sticks with fragments that bind to cryptic sites highlighted in orange. Alternative comparison between the *apo* and *holo* structures is shown in Figure S1.

cryptic sites. Ad hoc machine learning approaches are starting to emerge,<sup>20</sup> but they rely on experimentally known cryptic pockets, whose number is still limited. Many early attempts to identify cryptic pockets by atomistic simulations were limited by the short time-scale accessible,<sup>21,22</sup> sampling effectively side-chain rearrangements, but less successful in predicting more significant conformational changes.<sup>23</sup> Simulations using small organic fragments as probes<sup>24–27</sup> and mixed-solvent<sup>28</sup> have been quite successful in identifying binding hot-spots. For instance, in the case of the eukaryotic translation initiation factor 4E protein, a cryptic pocket that opens only transiently during MD simulations with water is stabilized by a water–benzene mixture.<sup>28</sup> Yet they are also limited by the time-scale problem and cannot usually find cryptic sites that take longer to open.<sup>29</sup> The ever-increasing computational resources allowed cryptic pocket predictions based on much longer simulations complemented by Markov state models.<sup>30</sup> Still, even long simulations, lasting hundreds of microseconds, fall short when dealing with transient, high-in-energy pockets.

Enhanced sampling techniques have been successfully employed to accelerate the exploration of the relevant conformational space in complex biological systems<sup>31–33</sup> and were successful in localizing cryptic pockets.<sup>34,35</sup> Unfortunately, many of these methods, as Umbrella Sampling,<sup>36</sup> Metadynamics,<sup>37,38</sup> or Transition Path Sampling,<sup>39</sup> rely on the use of Collective Variables (CVs) or a path approximating the reaction coordinate, and require prior knowledge on the location and features of the opening pockets. As this knowledge is mostly unavailable for cryptic binding sites, significant efforts have been devoted to design general CVs, such as JEDI,<sup>40</sup> which sample the opening of hydrophobic cavities. These CVs, however, still require prior knowledge on the location of the binding sites. Methods based on multiple replica of the system such as Parallel Tempering (PT)<sup>41</sup> or Hamiltonian Replica Exchange (HREX)<sup>42</sup> offer an attractive alternative to CV-based enhanced sampling approaches, but thus far have not been systematically employed for this purpose.

The aim of this work is both to devise a robust and generally applicable simulation-based approach and to better understand the driving forces for cryptic pocket formation.<sup>15,43</sup> We compare long MD simulations with both Parallel Tempering and a novel HREX-based approach “SWISH” (Sampling Water Interfaces through Scaled Hamiltonians) in combination with fragment-based simulations, on four systems. Three are of

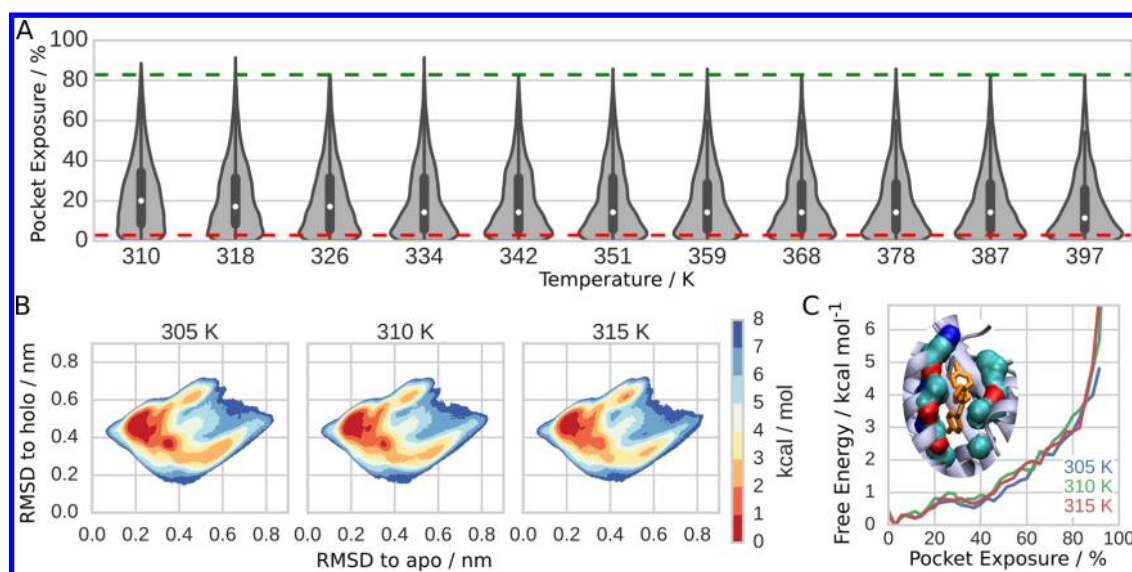
pharmaceutical interest and harbor experimentally validated cryptic pockets: TEM1  $\beta$ -lactamase,<sup>44</sup> interleukin-2 (IL2),<sup>45</sup> and Polo-like kinase-1 (PLK1)<sup>46</sup> (see Figure 1).  $\beta$ -Lactamase is responsible for widespread antibiotic resistance due to its ability to hydrolyze  $\beta$ -lactam antibiotics. Its cryptic binding site, which is not observed in the *apo* crystal structure,<sup>47</sup> has been discovered serendipitously<sup>44</sup> when crystals revealed two small organic molecules inserted at a predominantly hydrophobic interface between  $\alpha$ -helices H11 and H12, 1.6 nm away from the active site. IL2 plays a key role in the activation of T-cells and in the rejection of tissue grafts.<sup>10</sup> Its cryptic binding site coincides with its partner protein binding surface. The discovery of this cryptic site led to one of the earliest examples of successful small-molecule inhibitor against PPIs. The site elongates across the natural PPI interface, of which a significant part is apolar, and has been explored with increasingly longer ligands, revealing a surprisingly large pocket.<sup>45</sup> PLK1 is a validated anticancer target in which phosphorylation-dependent PPIs play a crucial role in regulating cell mitosis. Its cryptic site was discovered in a complex with a synthetic phosphopeptide,<sup>46</sup> revealing a binding cavity that is not exposed when bound to natural peptides.<sup>48</sup> The fourth system, Ubiquitin, is not known to harbor cryptic pockets. It has been chosen to act as a control for “false positives” as its structure has been characterized by crystallography and its flexibility and weak superficial binding sites by NMR.<sup>49,50</sup>

## RESULTS AND DISCUSSION

We begin with testing whether or not the cryptic pockets are populated in the absence of the ligands allowing their sampling by long unbiased MD simulations.<sup>22,30</sup> We run several long MD simulations of TEM1, IL2, and PLK1 with different force-fields for 1–5  $\mu$ s each. However, they fail to reveal an appreciable opening of the cryptic sites when starting from the *apo* crystal structures (see Supporting Information and Figure S2).

### Cryptic Pockets Are Unstable without Ligands and Higher Temperature Does Not Improve Their Sampling.

The Parallel Tempering<sup>41</sup> approach does not rely on the choice of system-dependent CVs and significantly enhances the sampling through excursions at higher temperatures; it is thus a seemingly suitable method for our purpose. However, to our initial surprise, increasing the temperature does not help exploring the cryptic sites for any of our chosen systems (see



**Figure 2.** TEM1 cryptic site stability without a ligand. (A) Violin plots of pocket exposure distribution at different temperatures during unbiased Parallel Tempering simulation. Reference values for *apo* and *holo* crystal structures are indicated with red and green dashed lines, respectively. For clarity, only every third replica is shown, to see full range of replicas for TEM1, IL2, and PLK1 systems in Figure S3. (B) Two-dimensional free energy surfaces at different equilibrium temperatures as a function of RMSD relative to *apo* and *holo* crystal structures. (C) Reweighted free energy as a function of pocket exposure with an image of pocket lining atoms used to estimate this variable visualized as a surface and ligand shown in orange sticks. All of the above results show instability of the cryptic pocket conformation.

Figure 2A and Figure S3). Throughout the 500 ns-long multiple replica simulations, the known cryptic sites remained mostly unidentifiable with an average exposure (see Supporting Information) of 23%, 8%, and 13% for TEM1, IL2, and PLK1, respectively. To better understand this, we investigated the TEM1 cryptic site stability and its thermodynamic balance.

Starting equilibrium simulations from the open (*holo*-like) conformation with ligands removed, we repeatedly observe a prompt pocket closure irrespective of the force-field used. Indeed, the results obtained with amber03<sup>51</sup> and amber14SB<sup>52</sup> are in qualitative agreement with those obtained with amber99SB\*ildn;<sup>53–55</sup> while with charmm22\*<sup>56,57</sup> we observe local unfolding (see Supporting Information and Figure S4).

The thermodynamic balance of the binding site formation was further studied with Parallel Tempering Metadynamics (PTMetaD) simulations.<sup>58</sup> We estimated the free energy surface (FES) describing the interconversion between *apo*- and *holo*-like structures of TEM1 in the absence of the ligands. As CVs we used root mean squared deviation (RMSD) from the open and closed crystal structures and coupled the simulations to a range of temperatures (see Methods). The reconstructed FES shows that, as expected, the dominant state of  $\beta$ -lactamase is akin to the crystallographic *apo* structure in which the cryptic site is closed (Figure 2B). However, it also confirmed that the *holo* structure does not correspond to a stable minimum, and pocket opening is thus rarely observed. After reweighting the FES<sup>59</sup> as a function of the pocket exposure (see Figure 2C), it is clear that the open conformation is  $\sim 3$  kcal/mol higher in energy, disfavoring the formation of the cryptic site. Figure 2C also reiterates the lack of a minimum and, as a result, a real transition state for cryptic site opening. In turn, this explains the lack of any significant pocket opening in the long equilibrium simulations. While fully open states are much higher in energy, the calculated free energy profile is in agreement with previous suggestions that equilibrium fluctuations can sample partially open states.<sup>30</sup> As the sampling of the open states might be force-field dependent, we repeated the

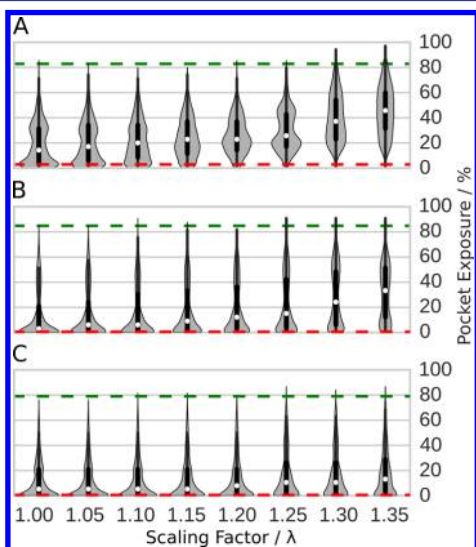
PTMetaD simulation using the same amber03 force-field of ref 30. Although the cryptic pocket is still unstable, the protein spends significantly more time in a partially open conformation (see Figure S5). Not only the main minimum is shifted from  $\sim 0$  to 20% partial exposure, but also the energy penalty to achieve an exposure above 80% is 1 kcal/mol less than that of the amber99SB\*-ILDN force-field.

As expected from the PT simulations, the free energy difference between *apo* and *holo* is hardly affected by the change in temperature (see Figure 2B,C). The estimated entropic contribution to the conformational change indeed was found to be very small and negative for both force-fields (see Figure S6), explaining why pocket exploration is not helped by higher temperatures. However, we cannot exclude alternative roles of entropy in other systems.

**SWISH Approach Efficiently Explores the Cryptic Pockets.** Our above-discussed results are suggestive of a major role of the ligands in opening and stabilizing the cryptic sites. As the druggable pockets are known to be apolar/hydrophobic,<sup>8,40,60,61</sup> we sought an enhanced sampling approach that would replicate the effect of a small ligand by changing the protein–water interactions. Inspired by simulations with probes<sup>27</sup> and studies about the water–protein interaction effect on protein folding and stability,<sup>62,63</sup> we developed the HREX-based technique SWISH (Sampling Water Interfaces through Scaled Hamiltonians). By progressively scaling the nonbonded interactions of solvent molecules with protein atoms, we shift the water properties toward more ligand-like behavior to increase cryptic site opening. After a few trials, we found that the best results are obtained by applying the scaling to apolar protein atoms only: carbon and sulfur (see Supporting Information). The higher is the value of scaling factor  $\lambda$ , the stronger is water affinity to apolar protein surface patches. Our preliminary tests have also shown that the range of  $\lambda$  needs to be chosen sensibly to avoid the unfolding of the proteins. We settled for a range of  $\lambda$  that keeps the fluctuations

of the  $\alpha$ -carbon RMSD similar to those observed in the Parallel Tempering simulations.

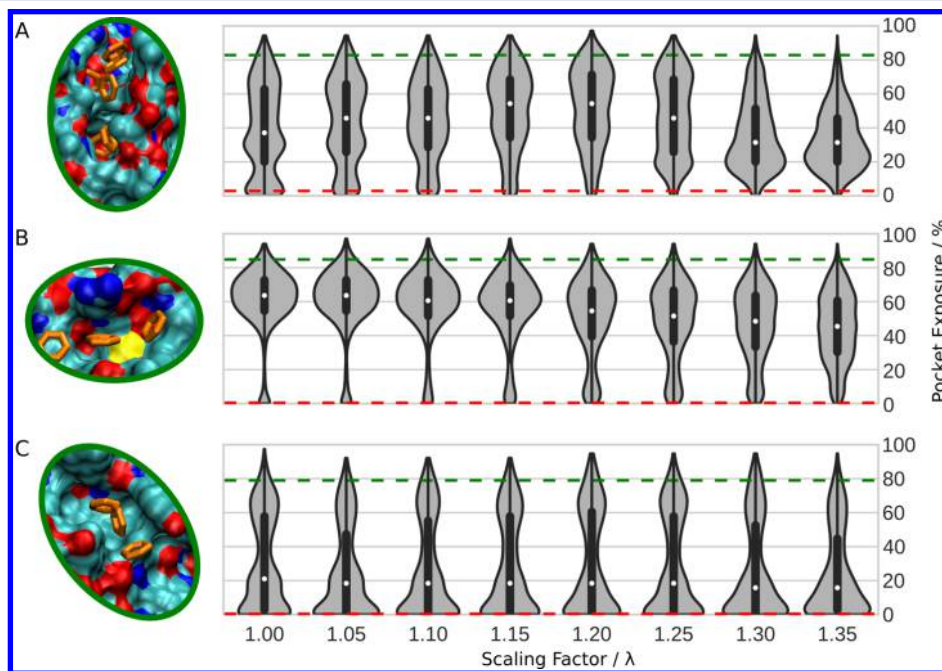
We performed 500 ns long SWISH simulations for each target system. In all three cases, we found that as  $\lambda$  increases, the known druggable sites become more exposed (Figure 3).



**Figure 3.** Violin plots show pocket exposure distribution at different scaling factor  $\lambda$  during unbiased ligand-free SWISH simulations of (A) TEM1, (B) IL2, and (C) PLK1 model systems. In each case, reference values for corresponding *apo* and *holo* crystal structures are indicated with red and green dashed lines, respectively. All three systems show enhanced populations for states with exposed cryptic sites as  $\lambda$  is increased.

This positive result suggests that the approach can be effective in cryptic binding site exploration even without ligands. As expected, the pocket population in replicas with  $\lambda = 1.0$ , corresponding to the standard Hamiltonian, is low, in agreement with the previous results of the other enhanced sampling approaches.

**Fragments Bind to the Cryptic Sites with a Mixed Mechanism.** With the further support for the hypothesis of ligand-induced cryptic pocket formation, we chose to perform a series of equilibrium simulations with probe molecules. We selected a set of six hydrophobic ring molecules: benzene, pyridine, imidazole, indole, pyrimidine, and tetrahydropyran, each of which are within the top 15 most common ring fragments in the FDA approved drugs,<sup>64</sup> benzene being the most common. We solvated the TEM1 *apo* structure with pre-equilibrated water boxes with  $\sim 1$  M probe concentration. To prevent phase separation, we used interligand repulsive potentials similar to those previously used in the literature.<sup>24</sup> For each probe, we ran 32 independent replicas of 100 ns in length. We observe several trajectories in which the fragments bind to the TEM1 cryptic pocket (see Figure S7), shifting the equilibrium to significant exposure. Closer inspection of the resulting structures from successful trajectories shows a wide open cryptic site bound by a few molecules intercalated between the helices 11 and 12 akin to the *holo* crystal structure. In the presence of ligands, the open conformations clearly correspond to an energy minimum. However, there is a great variability between the individual trajectories. In many cases, the ligands were unable to discover the site. Moreover, they tend to interact with unspecific sites on the surface. Extending the simulations up to a microsecond still proved to be insufficient to successfully explore the cryptic pocket (see Figure S67H,I). It is possible that, due to the energy penalty of



**Figure 4.** Violin plots show pocket exposure distribution at different scaling factor  $\lambda$  during unbiased SWISH simulations with benzene fragments of model systems: (A) TEM1, (B) IL2, and (C) PLK1. In each case, reference values for corresponding *apo* and *holo* crystal structures are indicated with red and green dashed lines, respectively. In the presence of probes, we observe a significant cryptic site population even at the “neutral” ( $\lambda = 1.0$ ) replicas. Structural trajectory snapshots of exposed cryptic sites bound by benzene probes, corresponding to maximum pocket exposure at  $\lambda = 1.0$ , are shown on the left-hand side for each system. Benzene is visualized as orange sticks and receptor surface in the same way as in Figure 1.

site opening, the fragments are not always able to access the hydrophobic site and hence induce the full site opening, hindering the convergence. In addition, a slow exchange rate of the bound fragments together with unspecific interactions also contribute to poor reproducibility.

From a mechanistic point of view, we observe first a slight opening of a channel (e.g., from the active site or through the helices) that allows the permeation of a fragment to the hydrophobic area underneath the two helices. Once a hydrophobic fragment resides there, one of the hydrophobic residues might come loose and expose a larger hydrophobic patch to which an increasing number of fragments bind. Finally, the cavity fully opens with multiple fragments bound to it (see Figure S8). This is suggestive of a mixed binding mechanism for the fragments whereby both conformational selection and induced-fit play a role (see analysis and discussion in Supporting Information). To verify the nature of the binding mechanism with more drug-like ligands, we repeated the simulations with CBT [*N,N*-bis(4-chlorobenzyl)-1*H*-1,2,3,4-tetraazol-5-amine], which was shown experimentally to bind to the cryptic pocket. Albeit we were not able to observe the binding of CBT during a standard MD simulations of average length, in agreement with the observations of ref 65, CBT successfully found the cryptic pocket when SWISH was used (see below).

Finally, we note that the chemical nature of the probe is important. We found that the site opening decreased going from benzene and pyrimidine to pyridine, imidazole, and indole, and was almost negligible in the case of tetrahydropyran (see Figure S7).

**Combining SWISH with Fragments Provides the Most Efficient Approach to Cryptic Pocket Detection.** We repeated the SWISH simulations in the presence of benzene and pyrimidine fragments for all systems (see Figure 4 and Figure S9). In all of the cases, we see a significant and consistent opening of the cryptic site. The increased effectiveness of the SWISH+fragment approach might be explained by the fact that, while SWISH sampling helps exposing the cryptic sites, those are further opened and stabilized by the fragments. The combined method is thus very general as it helps sampling systems with induced-fit, conformational selection and “mixed” binding mechanisms.<sup>15</sup> What is more, due to the scaling of protein–water interactions, at high  $\lambda$  values the ligands are out-competed by water. This behavior accelerates the sampling by allowing the fragments to quickly leave unspecific interaction sites and more efficiently explore the protein surface (Figure S9). We note that this also reduces the average pocket exposure at higher  $\lambda$  values as water molecules are less efficient at inducing the exposure than larger fragments. However, the fractional opening appears to be sufficient for the fragments to enter and further induce the sites once lower  $\lambda$  allows displacing the scaled waters. What typically happens is that at higher lambda the opening and closing of the pockets is faster, helping the binding of the fragments to cryptic pockets, but also the unbinding is faster, leading to a lower residence time. When a ligand bound to a replica at higher lambda exchanges with a replica at lower lambda values, the residence time of the fragments is higher, and thus the open states experience a longer lifetime. Thus, by combining the pocket opening power of the scaled replicas and the binding stability of the unbiased replica, the fragments access the pocket and stay bound. However, higher lambdas help the sampling up to a point; when lambda is too high, the protein starts to unfold

resulting in the exploration of mostly irrelevant unfolded states (see the Supporting Information).

The greatly increased effectiveness of the SWISH+fragment approach might be explained by the fact that, while SWISH sampling helps expose the cryptic sites, those are further opened and stabilized by the fragments. What is more, due to the scaling of protein–water interactions, at high  $\lambda$  values the ligands are out-competed by water. This behavior accelerates the sampling by allowing the fragments to quickly leave unspecific interaction sites and more efficiently explore the protein surface (Figure S8). We note that this also reduces the average pocket exposure at higher  $\lambda$  values as water molecules are less efficient at inducing the exposure than larger fragments. However, the fractional opening appears to be sufficient for the fragments to enter and further induce the sites once lower  $\lambda$  allows displacing the scaled waters. What typically happens is that at higher  $\lambda$  the opening and closing of the pockets is faster, helping the binding of the fragments to cryptic pockets, but also the unbinding is faster, leading to a lower residence time. When a ligand bound to a replica at higher  $\lambda$  exchanges with a replica at lower  $\lambda$  values, the residence time of the fragments is higher, and thus the open states experience a longer lifetime. Thus, by combining the pocket opening power of the scaled replicas and the binding stability of the unbiased replica, the fragments can access the pocket and stay bound. The combined method is very general as it helps sampling systems with induced-fit, conformational selection and “mixed” binding mechanisms.<sup>15</sup> As compared to plain MD with fragments or mixed solvent, it is much more efficient in exploring conformational changes with high activation barriers. Unlike CV-based approaches,<sup>34,35</sup> it does not need previous knowledge of the binding site.

We also performed SWISH simulations of TEM1 with the validated allosteric ligand CBT. At higher  $\lambda$ 's, one CBT molecule was able to bind to the cryptic pocket in  $\approx 100$  ns. After the first CBT was bound, a second molecule quickly entered the site, and the two explored orientations similar to those revealed by X-ray crystallography (see Figure S10). This result is also suggestive of a mixed binding mechanism for CBT, as the presence of one molecule in the cavity clearly helps the binding of additional ones.

## ■ DISTINGUISHING REAL CRYPTIC POCKETS FROM FALSE POSITIVES

An important question that needs to be addressed is how do we distinguish “real” cryptic sites from false positives? In general, one cannot be certain whether a novel cavity found by simulations and previously unreported by experiments is a false positive or a yet undiscovered cryptic site. However, this does not mean that the question cannot be addressed, especially if the aim is to distinguish between weakly binding superficial sites and buried cryptic sites. Take, for instance, the case of TEM1. Apart from the main cryptic site to which CBT binds, the other two minor cryptic sites have been proposed to be druggable.<sup>30</sup> We want to distinguish these sites (the main and possibly the two minor ones) from other more or less superficial binding “hotspots”. Both long repeated MD simulations ( $32 \times 1.1 \mu\text{s}$ ) and SWISH simulations with fragments were able to identify the superficial sites (see Figure S11). When we plot the isosurfaces of the volumetric maps showing the time-averaged occupation density, the hotspots corresponding to the two superficial sites are evident. This reflects the fact that the ligands spend more time in contact with those sites. At difference with the superficial sites,

however, only the SWISH simulations were able to systematically explore the main cryptic site between the helices 11 and 12, irrespective of the use of benzene or pyrimidine. Thus, the occupation of the main site in the long MD is much lower than that obtained by SWISH. Subtracting the ligand density obtained from the long MD simulations from that obtained with SWISH+fragments clearly reveals a hotspot corresponding to the buried cryptic site (Figure S11). We repeated the analysis on IL2 and PLK1 with similar results. On the basis of these results, we propose that buried cryptic sites can be detected by looking at the fragment occupation density differences between SWISH and standard MD simulations. This approach also provides a quick visual way to assess different hotspots.

Still, an important question is whether or not we can distinguish true pockets from false positive. To test whether or not the approach we propose also works in a case where no experimental cryptic pockets are known, we have tested it on Ubiquitin, whose structure and dynamics as well as a number of weak superficial binding sites have been experimentally characterized.<sup>49,50</sup> First, by computing the time-averaged fragment occupation, we observe with both long MD simulations and SWISH two weak superficial hotspots, which are interestingly located close to experimentally reported weak binding sites (Figure S11).<sup>50</sup> Yet when we compute the difference in the binding density, both hotspots disappear, consistent with the absence of observed buried cryptic binding pockets.

## CONCLUSIONS

We investigated the molecular mechanisms of cryptic pocket opening and compared the efficiency of different simulation-based approaches in finding them. Long standard MD simulations, free-energy calculations, and extensive Parallel Tempering simulations revealed the energy penalty of cryptic site opening resulting in a spontaneous closure in the absence of a ligand. Although long equilibrium simulations can sample rare fluctuations leading to the opening of the sites, the incomplete opening and short lifetime make their detection difficult. Also, the partially open states are quite different from the states observed in ligand-bound crystal structures and might prove impractical in virtual screening. The use of different force-fields confirms the qualitative picture; while some of them might underestimate the energy penalty to form the pockets, all agree on their lack of stability.

Our results indicate a role of induced-fit effects in cryptic site opening, and show a mixed mechanism in some cases, as the binding of benzene and the more drug-like ligand CBT to TEM1. We have demonstrated that our proposed SWISH method is able to induce the opening of the known cryptic binding sites in TEM1, IL2, and PLK1 systems by scaling protein–water interactions. Combining the SWISH method with suitable probe molecules proved to be the most efficient and reliable approach. At difference with CV-based enhanced sampling algorithms, it does not require prior knowledge of the site or the definition of optimal CVs to discover the cryptic sites. What is more, the difference in time-averaged fragment densities from SWISH and long MD simulations is able to distinguish them from nonspecific superficial binding hot-spots. To the best of our knowledge, this is the first successful example in which enhanced sampling techniques are combined with fragment molecule simulations to explore cryptic pockets. We also note that the SWISH approach might be applied to a

wide range of protein sampling problems including protein folding.

## METHODS

The protein structures were retrieved from the Protein Data Bank<sup>66</sup> (PDB entries 1JWP, 1PZO, 1M47, 1PY2 (chain D), 3FVH, 3P37 (chain A), and 2LJ5). Missing residues were added using the software Modeller,<sup>67</sup> according to the respective Uniprot sequences. The structural figures were made using VMD.<sup>68</sup>

Molecular Dynamics was performed with Gromacs 4.6.7.<sup>69</sup> Unless otherwise specified, all simulations were run at 300 K NVT ensemble with temperature coupled using V-rescale algorithm,<sup>70</sup> using the amber99SB\*-ILDN<sup>53–55</sup> and explicit solvent TIP3P<sup>71</sup> force-fields. PTMetaD<sup>58</sup> was run using the PLUMED<sup>72</sup> for 500 ns using binding site RMSD relative to *apo* and *holo* structures over the temperature range of 305–315 K using 9 replicas. Unbiased Parallel Tempering simulations (310–400 K; 32 replicas) and SWISH simulations ( $\lambda \in [1.0; 1.35]$ , 8 replicas, HREX<sup>73</sup>) were each performed for 500 ns. Ligands were parametrized using GAFF<sup>74</sup> with additional repulsive interligand potential. To monitor pocket exposure, we analyzed snapshots using fpocket.<sup>75</sup> For further details, see the Supporting Information.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b05425.

Supplementary results and discussion with accompanied figures and tables referred to in the main text, along with a more explicit account of methodology (PDF)

## AUTHOR INFORMATION

### Corresponding Author

\*f.l.gervasio@ucl.ac.uk

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

Support for this work was provided by BBSRC CASE Studentship (V.O.). G.S. and F.L.G. acknowledge EPSRC [grant no EP/M013898/1] for financial support. HecBioSim [EPSRC grant no EP/L000253/1] and PRACE are acknowledged for computer time.

## REFERENCES

- (1) Kola, I.; Landis, J. *Nat. Rev. Drug Discovery* **2004**, *3*, 711–716.
- (2) Avorn, J. *N. Engl. J. Med.* **2015**, *372*, 1877–1879.
- (3) Paul, S. M.; Mytelka, D. S.; Dunwiddie, C. T.; Persinger, C. C.; Munos, B. H.; Lindborg, S. R.; Schacht, A. L. *Nat. Rev. Drug Discovery* **2010**, *9*, 203–214.
- (4) Arrowsmith, J. *Nat. Rev. Drug Discovery* **2012**, *11*, 17–18.
- (5) Scannell, J. W.; Blanckley, A.; Boldon, H.; Warrington, B. *Nat. Rev. Drug Discovery* **2012**, *11*, 191–200.
- (6) Brown, D.; Superti-Furga, G. *Drug Discovery Today* **2003**, *8*, 1067–1077.
- (7) Hajduk, P. J.; Huth, J. R.; Fesik, S. W. *J. Med. Chem.* **2005**, *48*, 2518–2525.
- (8) Cheng, A. C.; Coleman, R. G.; Smyth, K. T.; Cao, Q.; Soulard, P.; Caffrey, D. R.; Salzberg, A. C.; Huang, E. S. *Nat. Biotechnol.* **2007**, *25*, 71–75.
- (9) Surade, S.; Blundell, T. L. *Chem. Biol.* **2012**, *19*, 42–50.
- (10) Wells, J. A.; McClendon, C. L. *Nature* **2007**, *450*, 1001–1009.
- (11) Vigil, D.; Cherfils, J.; Rossman, K. L.; Der, C. J. *Nat. Rev. Cancer* **2010**, *10*, 842–857.

- (12) Ostrem, J. M.; Peters, U.; Sos, M. L.; Wells, J. A.; Shokat, K. M. *Nature* **2013**, *503*, 548–551.
- (13) Koshland, D. E. *Proc. Natl. Acad. Sci. U. S. A.* **1958**, *44*, 98–104.
- (14) Monod, J.; Wyman, J.; Changeux, J.-P. *J. Mol. Biol.* **1965**, *12*, 88–118.
- (15) Csermely, P.; Palotai, R.; Nussinov, R. *Trends Biochem. Sci.* **2010**, *35*, 539–546.
- (16) Ludlow, R. F.; Verdonk, M. L.; Saini, H. K.; Tickle, I. J.; Jhoti, H. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *2015*, 201518946.
- (17) Erlanson, D. A.; Braisted, A. C.; Raphael, D. R.; Randal, M.; Stroud, R. M.; Gordon, E. M.; Wells, J. A. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 9367–9372.
- (18) Lawson, A. D. G. *Nat. Rev. Drug Discovery* **2012**, *11*, 519–525.
- (19) Cossins, B.; Lawson, A. *Molecules* **2015**, *20*, 16435–16445.
- (20) 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. *J. Mol. Biol.* **2016**, *428*, 709–719.
- (21) Schames, J. R.; Henchman, R. H.; Siegel, J. S.; Sottriffer, C. A.; Ni, H.; McCammon, J. A. *J. Med. Chem.* **2004**, *47*, 1879–1881.
- (22) Eyrisch, S.; Helms, V. *J. Med. Chem.* **2007**, *50*, 3457–3464.
- (23) Frembgen-Kesner, T.; Elcock, A. H. *J. Mol. Biol.* **2006**, *359*, 202–214.
- (24) Guvench, O.; MacKerell, A. D. *PLoS Comput. Biol.* **2009**, *5*, e1000435.
- (25) Lexa, K. W.; Carlson, H. A. *J. Am. Chem. Soc.* **2011**, *133*, 1–8.
- (26) Seco, J.; Luque, F. J.; Barril, X. *J. Med. Chem.* **2009**, *52*, 2363–2371.
- (27) Tan, Y. S.; Ślędz, P.; Lang, S.; Stubbs, C. J.; Spring, D. R.; Abell, C.; Best, R. B. *Angew. Chem., Int. Ed.* **2012**, *51*, 10078–10081.
- (28) Lama, D.; Brown, C. J.; Lane, D. P.; Verma, C. S. *Biochemistry* **2015**, *54*, 6535–6544.
- (29) Hall, D. R.; Enyedy, I. J. *Future Med. Chem.* **2015**, *7*, 337–353.
- (30) Bowman, G. R.; Geissler, P. L. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 11681–11686.
- (31) Bernardi, R. C.; Melo, M. C.; Schulten, K. *Biochim. Biophys. Acta, Gen. Subj.* **2015**, *1850*, 872–877.
- (32) Cavalli, A.; Spitaleri, A.; Saladino, G.; Gervasio, F. L. *Acc. Chem. Res.* **2015**, *48*, 277–285.
- (33) Masetti, M.; Cavalli, A.; Recanatini, M.; Gervasio, F. L. *J. Phys. Chem. B* **2009**, *113*, 4807–4816.
- (34) Pietrucci, F.; Vargiu, A. V.; Kranjc, A. *Sci. Rep.* **2015**, *5*, 18555.
- (35) Herbert, C.; et al. *Cancer Cell* **2013**, *23*, 489–501.
- (36) Torrie, G.; Valleau, J. *J. Comput. Phys.* **1977**, *23*, 187–199.
- (37) Laio, a.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562–12566.
- (38) Laio, A.; Gervasio, F. L. *Rep. Prog. Phys.* **2008**, *71*, 126601.
- (39) Dellago, C.; Bolhuis, P. G.; Csajka, F. S.; Chandler, D. *J. Chem. Phys.* **1998**, *108*, 1964.
- (40) Cuchillo, R.; Pinto-Gil, K.; Michel, J. *J. Chem. Theory Comput.* **2015**, *11*, 1292–1307.
- (41) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
- (42) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **2000**, *329*, 261–270.
- (43) D'Abramo, M.; Rabal, O.; Oyarzabal, J.; Gervasio, F. L. *Angew. Chem.* **2012**, *124*, 666–670.
- (44) Horn, J. R.; Shoichet, B. K. *J. Mol. Biol.* **2004**, *336*, 1283–1291.
- (45) Thanos, C. D.; Randal, M.; Wells, J. A. *J. Am. Chem. Soc.* **2003**, *125*, 15280–15281.
- (46) Ślędz, P.; Stubbs, C. J.; Lang, S.; Yang, Y.-Q.; McKenzie, G. J.; Venkitaraman, A. R.; Hyvönen, M.; Abell, C. *Angew. Chem., Int. Ed.* **2011**, *50*, 4003–4006.
- (47) Wang, X.; Minasov, G.; Shoichet, B. K. *J. Mol. Biol.* **2002**, *320*, 85–95.
- (48) Yun, S.-M.; et al. *Nat. Struct. Mol. Biol.* **2011**, *18*, 516–516.
- (49) Lange, O. F.; Lakomek, N.-A.; Fares, C.; Schroder, G. F.; Walter, K. F. A.; Becker, S.; Meiler, J.; Grubmuller, H.; Griesinger, C.; De Groot, B. L. *Science* **2008**, *320*, 1471–1475.
- (50) Chung, W. K.; Freed, A. S.; Holstein, M. A.; McCallum, S. A.; Cramer, S. M. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 16811–16816.
- (51) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.
- (52) Case, D. A.; et al. *AMBER 2015*; University of California: San Francisco, CA, 2015.
- (53) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct., Funct., Genet.* **2006**, *65*, 712–725.
- (54) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins: Struct., Funct., Genet.* **2010**, *1950*–1958.
- (55) Best, R. B.; Hummer, G. *J. Phys. Chem. B* **2009**, *113*, 9004–9015.
- (56) MacKerell, A. D.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (57) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. *Biophys. J.* **2011**, *100*, L47–L49.
- (58) Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2006**, *128*, 13435–13441.
- (59) Tiwary, P.; Parrinello, M. *J. Phys. Chem. B* **2015**, *119*, 736–742.
- (60) Schmidtke, P.; Barril, X. *J. Med. Chem.* **2010**, *53*, 5858–5867.
- (61) Morelli, X.; Bourgeois, R.; Roche, P. *Curr. Opin. Chem. Biol.* **2011**, *15*, 475–481.
- (62) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. *J. Phys. Chem. B* **2015**, *119*, 5113–5123.
- (63) Best, R. B.; Zheng, W.; Mittal, J. *J. Chem. Theory Comput.* **2014**, *10*, 5113–5124.
- (64) Taylor, R. D.; MacCoss, M.; Lawson, A. D. G. *J. Med. Chem.* **2014**, *57*, 5845–5859.
- (65) Bowman, G. R.; Bolin, E. R.; Hart, K. M.; Maguire, B. C.; Marqusee, S. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 201417811.
- (66) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *112*, 535–542.
- (67) Šali, A.; Blundell, T. L. *J. Mol. Biol.* **1993**, *234*, 779–815.
- (68) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (69) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; Van Der Spoel, D.; Hess, B.; Lindahl, E. *Bioinformatics* **2013**, *29*, 845–854.
- (70) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.
- (71) Jorgensen, W. L. *J. Am. Chem. Soc.* **1981**, *103*, 335–340.
- (72) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. *Comput. Phys. Commun.* **2009**, *180*, 1961–1972.
- (73) Bussi, G. *Mol. Phys.* **2014**, *112*, 1–6.
- (74) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (75) Le Guilloux, V.; Schmidtke, P.; Tuffery, P. *BMC Bioinf.* **2009**, *10*, 168.