Review

Journal of the Brazilian Chemical Society

https://dx.doi.org/10.21577/0103-5053.20240168

J. Braz. Chem. Soc. **2024**, *35*, 10, e-20240168, 1-21

The Interplay between Protein Frustration and Hotspot Formation

Thamires Q. Froes[a] *and Marcelo S. Castilho*[*,b]

[a]*Centro de Pesquisa Gonçalo Moniz, Fiocruz, R. Waldemar Falcão, 121, Candeal, 40296-710 Salvador-BA, Brazil*

[b]*Programa de Pós-Graduação em Farmácia, Faculdade de Farmácia, Universidade Federal da Bahia (UFBA), Campus Universitário de Ondina, 40170-110 Salvador-BA, Brazil*

Hotspots situated at protein-protein interfaces, allosteric and orthosteric binding sites, traditionally play a crucial role in the initial phases of drug discovery, as they are useful to identify druggable targets and guide the optimization of fragments into lead compounds. Despite the high level of protein frustration, which can be thought of as the percentage of residues that cannot independently achieve minimum energy due to tridimensional restraints, observed at those locations, limited efforts have been made to investigate any potential connection between the localized frustration and hotspots in proteins. This review paper aims to partially address this knowledge gap by proposing that the origin of hotspots is, at least in part, due to localized frustration in proteins. While there is no hard evidence to support this hypothesis, the results obtained by integrating *in silico* tools predicting hot spot locations, with those calculating localized protein frustration, suggest that druggable hotspots are surrounded by residues involved in frustrated contacts. Additionally, the inclusion of long-range electrostatic terms in the protein frustration calculation enables the identification of an equal to a higher number of frustrated residues. These observations indirectly suggest that localized protein frustration around druggable hotspots could guide the development of more potent leads.

Keywords: hotspot, DRUGpy, propensity maps, fractional overlap, Debye-Hückel terms

1. Introduction

Professor Eliezer Barreiro and co-workers[1] advocated that medicinal chemists exploit their chemical intuition to integrate knowledge from diverse domains, including organic chemistry and biology, to elucidate the mechanisms of drug actions and devise novel chemical entities. To achieve this objective, both basic and applied research play crucial roles, either expediting hit identification, and lead optimization, or the pre-clinical studies that ensure the safety and efficacy of candidate drugs entering clinical trials.[2,3]

Both ligand-based and structure-based molecular modeling studies play their role in this pipeline, offering either a simplistic, physics-based description of the biological targets[4] or quantum-based insights into the electronic properties of ligands.[5] These studies also encompass mathematical models, linear or otherwise, guiding virtual screening campaigns, ADMET (absorption, distribution, metabolism, excretion, and toxicity) exclusion criteria, etc.[6-8] Integrating concepts and foundations from diverse fields often facilitates performing these tasks from a novel and more efficient perspective. In a contribution towards this goal, this paper explores how the well-established minimum frustration theory[9] relates to the hotspots found in orthosteric and allosteric binding sites. The text is organized as follows: a brief description of the frustration theory and its application to protein folding is provided, followed by an introduction to a quick and efficient tool for calculating frustration in proteins (Frustratometer webserver).[10] The reader is then informed about the high proportion of frustrated residues in binding sites, paving the way for the second section, which delves into the concept of hotspots, demonstrates how they can be identified, and introduces another tool widely used for predicting the location of these hotspots (FTMap webserver).[11-13] The focus of this section will be on

*e-mail: castilho@ufba.br
Editor handled this article: Carlos Maurício R. de Sant'Anna (Guest)

I dedicate this work to Prof. Eliezer J. Barreiro, who first ignited my passion for medicinal chemistry and will always serve as a shining example of dedication to advancing this field of knowledge, not only in Brazil but also worldwide.

druggable hotspots, where drug-like molecules are expected to bind with nanomolar affinity, but this section will also describe the fundamentals of the Fragment hotspot map server,[14,15] and how it can be exploited to gain further insight into the molecular interactions that are responsible for hotspot formation. The subsequent section compares and contrasts whether highly frustrated residues, as predicted by the Frustratometer webserver, surround druggable hotspots, as calculated with the DRUGpy plugin,[16] or the hotspots predicted with Fragment hotspot map.[14,15] Consequently, readers will gain insights into the correlation between hotspots and protein frustration on enzymes and selected examples. The concluding remarks will bring the readers' attention to the work that should be carried out to prove that protein frustration is indeed at the root of hotspots.

2. A Brief Overview of Protein Frustration

While high-resolution single structures are essential for understanding biological processes and protein interactions, Nussinov *et al*.[17] argue that a conformational view is necessary for an in-depth analysis of how proteins function at the conformational level and how they are regulated in the cell.[17] This argument is rooted in the fact that the sequence-structure-function dogma leads to some misconceptions. Firstly, it implies that the binding of a small molecule to an incompatible conformation of the protein always "pushes" it to change its shape. Instead, we should recognize that binding might also involve conformational selection from the available options at a given temperature, necessitating only minor changes for optimal steric and electrostatic complementarity.[18,19] To differentiate whether conformational changes occur before (conformational selection) or after (induced fit) ligand binding, or involve a mix of both,[20] it is recommended to use biophysical techniques that do not rely on ensemble-averaged data, such as single-molecule Förster resonance energy transfer.[21] Secondly, it fails to identify the protein' preferred activation pathway, as it requires the consideration not only of conformational states but also their occupancies.[22-24] Under physiological conditions, proteins are predominantly in their inactive conformations, and effector binding, whether covalent or not, can stabilize the active conformation, enabling the crossing of kinetic barriers necessary for biological activity.[25,26] Lastly, it struggles to adequately explain allosteric mechanisms [17,27,28]

The concept of the energy landscape is instrumental to overcoming such limitations since it describes biological molecules as dynamic ensembles that adapt to their surroundings.[29,30] Hence, it maps all the conformations a protein can adopt, with lower energy corresponding to a higher population. Then, minor differences in energies among conformations mean that a slight energy change can easily shift an inactive conformation to an active one. This transition translates into a biological signal, the intensity of which is directly proportional to the energy gap between the active and inactive conformations.[31]

There is substantial evidence suggesting that natural proteins navigate a minimally frustrated energy landscape towards their native state, as opposed to exploring various alternative conformations.[9,32-35] Frustration arises when a physical system cannot independently achieve minimum energy for each of its components.[36] This can result from geometric reasons or competition between interacting elements. For instance, the formation of a specific secondary structure in one region may prevent the formation of another in a distant part.

Localized frustration in proteins is typically found at binding sites[37] and regions associated with conformational changes required for allosteric regulation.[38,39] Proteins tolerate these frustrated regions due to their role in protein dynamics and function. Proteins perform chemical activities that impose restrictions conflicting with the requirements of self-organization.[40] For instance, complex formation required for the proteins' activity (e.g., protein-protein or protein-ligand complexes) may conflict with those stabilizing the folding of one of the partners. As highlighted by Ferreiro *et al*.,[36] the molecular origin of binding frustration is related to intramolecular contacts present in the unbound form that need to be disrupted before intermolecular contacts of the complex can form.

Analyzing protein frustration requires a reliable method to measure the degree of satisfaction of a general energy function. A quantum mechanics description is unnecessary for this type of system, as most forces involved in protein frustration can be adequately accounted for using classical force field descriptions, or even the simplified coarse-grained description.[31] In such an approach, molecular interactions may be modeled with forces averaged over the solvent environment, and each amino acid in the polymer may be described as a pseudoatom with specific encoded properties.[41-43] For instance, the associative memory, water-mediated, structure and energy model (AWSEM) force field employs a three-atom-*per*-residue representation with explicit hydrogen bonds,[44] but differs from Takada's model in its relatively simple hydrophobic interactions and context-dependent electrostatic interactions, even though both were originally developed for folding studies.[45] A full description of the AWSEM model is available as the supplementary material of paper published by Davtyan *et al*.[44] Briefly, the solvent-averaged free energy

function of the protein chain (V_{total}) is calculated as a sum of 6 terms (equation 1):

$$V_{total} = V_{backbone} + V_{contact} + V_{burial} + V_{HB} + V_{AM} + V_{DSB} \quad (1)$$

where: $V_{backbone}$ accounts for the protein chain's connectivity using harmonic potentials to maintain specific distances between atoms, typically between Cα, N, and C atoms of adjacent residues. It ensures the correct bond angles around the Cα atom to preserve the protein's secondary structure, models the energetics of the chain's dihedral angles to ensure that the torsional angles adopt physically reasonable values, enforces allowed conformations based on the Ramachandran plot, and applies repulsive potentials to avoid steric clashes; $V_{contact}$ is split into two terms that have a direct correlation with the frustrated contacts: (*i*) a pairwise additive potential that depends on the distance between the Cβ atoms of interacting residues, capturing the direct, non-bonded interactions between the atoms; (*ii*) a term that switches between water-mediated and protein-mediated interaction weights, capturing the role of water in mediating interactions between residues that are not in direct contact; V_{burial} term mimics the hydrophobic effect in protein folding, dictating whether the residue should be buried or exposed depending on its hydrophobicity; V_{HB} term captures the energetics of hydrogen bonds, ensuring physically realistic geometries for these interactions; V_{AM} is the associative memory term that incorporates information from known protein structures to bias the simulation towards native-like conformations; V_{DSB} accounts for the desolvation barrier, helping to simulate the difficulty of bringing residues near each other without forming direct contacts or hydrogen bonds.

Using coarse-grained descriptions of the system, sequence permutation is a straightforward way to probe the effect of topological frustration.[36] To analyze the existence of energetic conflicts in a folded protein, the energy of structural or sequence decoys is measured concerning the native state. For instance, the "frustration" index can be calculated by comparing the stability of native interactions, represented by the total energy of the protein in the native configuration ($E^{T,N}$) (equation 2) to a distribution of decoy interactions.

$$E^{T,N} = \sum_{k\#i}^{N} \left(E^{i;k}_{contact} + E^{i;k}_{water}\right) + E^{i}_{burial} \quad (2)$$

where the first two terms account for the interactions that residue (i) makes with residues (k), either in a direct contact ($E^{i;k}_{contact}$) or in a water-mediated interaction ($E^{i;k}_{water}$), while the third term describes the burial energy for residue (i) (E^{i}_{burial}).

This total energy is then compared to the average total energy of the decoy set $\langle E^{T,U}_{i,j} \rangle$, which can be generated by randomizing the identities of the interacting amino acids (i) and (j), while keeping all other interaction parameters at their native values.[46] From these two values, the mutational frustration index (F^{m}_{ij}) can be calculated as shown in equation 3:

$$F^{m}_{ij} = \frac{E^{T,N}_{i,j} - \left\langle E^{T,U}_{i,j'} \right\rangle}{\sqrt{\frac{1}{N\sum_{k=1}^{n}\left(E^{T,U}_{i',j} - \left\langle e^{T,U}_{i',j'} \right\rangle\right)2}}} \quad (3)$$

when the decoy set is generated by changing not only the residue identities but also their positions in space, the "configurational frustration" (F^{c}_{ij}) is calculated to describe the interaction energies of the same pairs of residues when they interact in a non-native ensemble of structures. While the mutational frustration index can be computed using only the coordinates of the folded native backbone configuration, computing the configurational frustration depends on generating non-native decoy configurations, such as those available in long molecular dynamics simulations.

The local frustration indexes (mutational and configurational) are represented as the Z-score of the free energy of parts of the native structure compared to the distribution of the energy of rearranged decoys.[37] If the energy of a native pair of interacting residues (i and j) falls at the most favorable end of the distribution, the interaction (contact) is considered minimally frustrated. Conversely, regions where permutation would lower the free energy of the system are labeled as having highly frustrated contacts (Figure 1), indicating that the original pair of residues (i and j) make contacts in the high-energy end of the distribution.[37] The localized frustration may result in direct frustrated contacts, which refer to interactions between amino acid residues that are close to each other in the protein structure, or water-mediated frustrated contacts, which account for interactions between amino acid residues that are not in direct contact but interact through one or more water molecules.

The Frustratometer webserver[10] computes these frustration indexes from the Protein Data Bank (PDB) structure and enables users to download the results as a PyMOL[48] session, which displays highly frustrated contacts between residues as red-solid lines, while minimally frustrated contacts are represented by green-solid lines. Furthermore, water-mediated highly frustrated contacts are depicted as dashed lines, following the same color scheme (highly frustrated: red, minimally frustrated: green), within the PyMOL session (Figure 1).

Residues involved in interactions with an average Z-score (indicative of neutral frustration) are excluded from the PyMOL session to enhance clarity. Additionally,

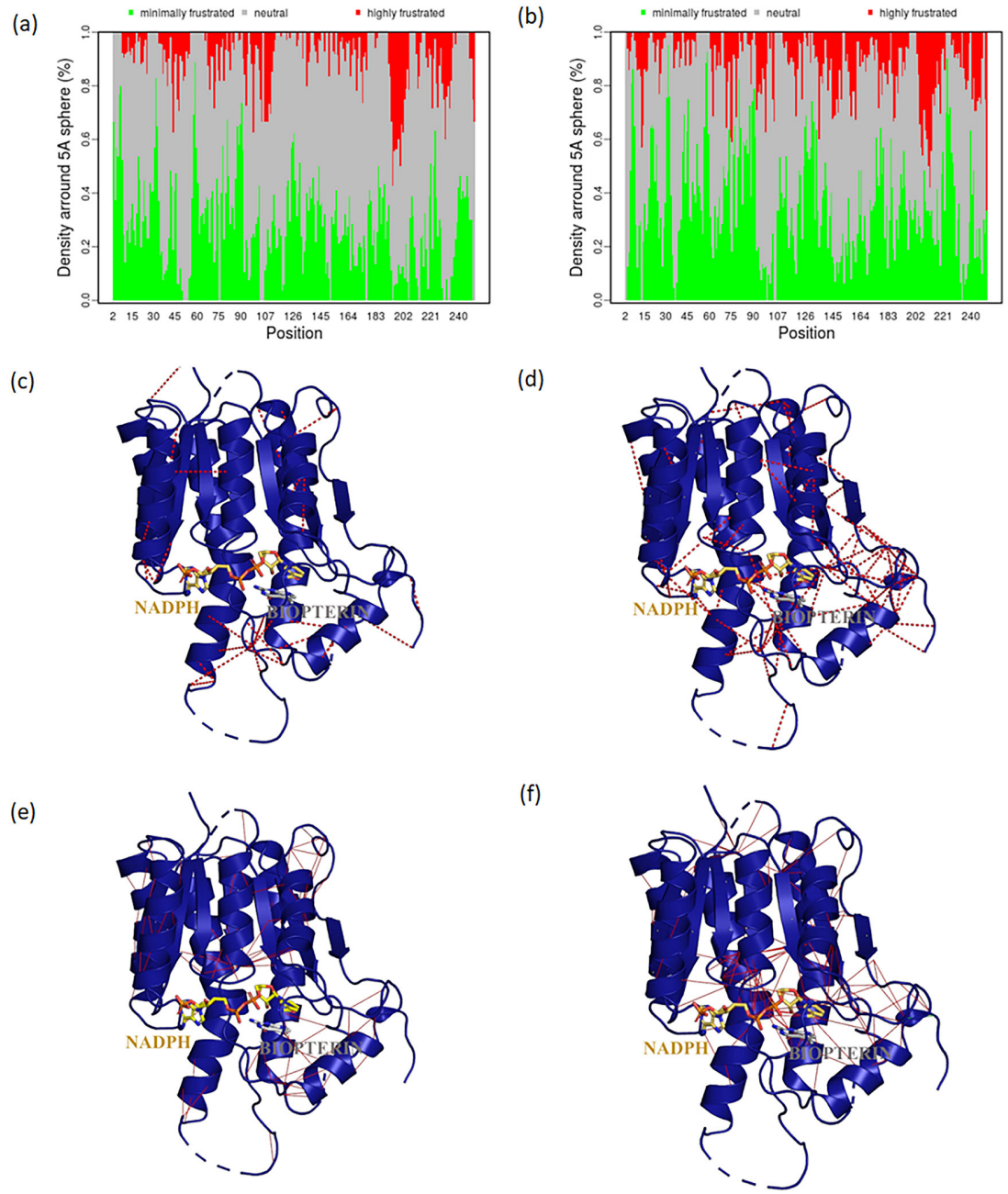

Figure 1. *Tb*PTR1 (PDB ID 4CLR)[47] localized frustration, calculated by the Frustratometer webserver,[10] is depicted on the left side for configurational frustration and on the right side for mutational frustration. (a,b) Density of frustrated residues (direct and water-mediated) within a 5 Å sphere radius. The localized frustration on the protein structure is highlighted by red dashed lines connecting residues engaged in highly frustrated water-mediated contacts (c,d) frustrated direct contacts, as solid red lines (e,f). The protein is presented in the cartoon, with the cofactor (nicotinamide adenine dinucleotide phosphate (NADPH)) and the substrate (biopterin) displayed in sticks. Frustration levels are represented as follows: minimally frustrated contacts in green, and highly frustrated contacts in red. Minimally frustrated contacts (green) are omitted in (c), (d), (e) and (f).

users have the option to visualize the results through three plots: a map displaying frustrated contacts for residues i and j, the counting of frustrated residues within a 5 Å radius of residue i, and the density (%) of frustrated residues within a 5 Å radius of residue i. In all plots, residues with neutral frustration are depicted in gray, but the user cannot distinguish between direct and water-mediated frustrated contacts in the plots (Figure 1).

In 2019, Chen *et al.*[49] introduced an all-atom model for computing localized protein frustration using the Rosetta energy function. For a matter of consistency, two types of Rosetta-based frustration indices were calculated: the atomistic "mutational-frustration", and the atomistic "configurational-frustration". The authors acknowledge that one of the biggest challenges of all-atoms descriptions is to build the ensemble of non-native conformations in a

reasonable amount of time, as the results reported in their paper took hundreds of core hours (see the supplementary material of Chen *et al*.).[49]

A comparison between the atomistic and AWSEM-frustratometer results shows a poor correlation, but some general trends are observed: (*i*) both methods agree that the vast majority of interactions have minimal frustration; (*ii*) there is a set of non-neutral residues identified by both methods; (*iii*) however, there is a significant amount of non-neutral frustrated residues identified only by one method or the other. Chen *et al*.[49] ascribe this discrepancy to different sets of parameters and cutoff values employed in each force field, without specifying which parameter or cutoff has the largest impact on the results. Similarly, the authors acknowledge that the results from both methods were not expected to match, without further justifying why. Rosetta-atomistic frustration indices are better suited for predicting sequence mutations that could improve the folding of a highly frustrated protein. On the other hand, the coarse-grained version available through the Frustratometer webserver remains the most adequate and cost-effective method to assess localized frustration in potential therapeutic targets. Alternatively, users can utilize an R package,[50] which also employs the coarse-grained version, to compare and contrast localized frustration in large protein datasets.

The original AWSEM forcefield accounts for local electrostatic interactions with the solvent but ignores long-range electrostatic interactions.[51] However, electrostatic interactions play an important role in the recognition between the protein and its ligand,[45,52] as the electrostatic interactions can introduce conformational changes when the partners come into close proximity. For that reason, Tsai *et al*.[51] added Debye-Huckel terms that describe long-range electrostatic interactions, into the AWSEM package.[53] Whether considering, or not, these long-range interactions improve the relationship between frustration and hotspot prediction will be addressed in the last part of "4.1. Hotspot glimpses into frustrated proteins" sub-section.

3. Hotspots on Proteins and their Importance to Drug Design

In the past couple of decades, drug development has predominantly focused on target-based approaches, relying primarily on *in vitro* assays.[54,55] According to this paradigm, macromolecules that play a crucial role in physiopathological processes and are considered druggable become suitable targets for drug development.[56] Therefore, identifying and characterizing the ligand's binding site are recognized as crucial steps for the success of drug development campaigns. The identification of binding sites can be accomplished through either experimental means[57] or computational strategies. Various *in silico* methods are available to identify cavities or pockets where molecules may bind, utilizing geometric and/or energetic criteria (Table 1).

Table 1. Software employed to binding site (hotspot) prediction

Name	Prediction criteria	Reference
CASTp	geometric	58,59
Pocketfinder	geometric	60
Q-SiteFinder	energetic	61
ConCavity	evolution (homology)	62
SiteMap	energetic	63
Ftsite (FTMap)	solvent mapping	12,13
Ligsite (fragment hotspot map)	atomic propensity weighted by buriedness	15

Typically, binding sites are situated in non-exposed, hydrophobic cavities that harbor regions, termed hotspots, that have a disproportional contribution to the ligand's binding energy with the macromolecule. The meaning of "hotspot" can vary across research fields.[64-66] In the context of protein science, hotspots refer to residues whose mutation to alanine alters the ligand's binding energy by at least 2 kcal mol^{-1}.[67] Identifying hotspots traditionally involves generating numerous single-point mutant proteins, purifying them, and assessing the mutation's impact on the binding energy. Mutagenesis studies cannot distinguish between residues critical for protein folding and those relevant for ligand binding,[67] which are the main interest of medicinal chemists.

For drug design purposes, it is noteworthy that hotspots coincide with sites where various organic solvent molecules bind,[68,69] even though the biophysical explanations for this phenomenon are not fully understood.[70] Despite experimental solvent-mapping provides unique insights into ligand binding by considering protein conformational flexibility,[71] it is not feasible for many proteins due to the harsh experimental conditions required (e.g., soaking the protein crystal in a dimethyl sulfoxide (DMSO) 50% solution for several hours). In response to these challenges, several *in silico* tools have been developed to predict hotspot locations on protein-protein interfaces, including DrugScorePPI,[72] ISIS,[73] PIIMS web server,[74] ECMIS,[75] WaterMap[76] and mixed solvent molecular dynamics.[77] Unless the solvation penalty is considered, *in silico* prediction of hotspots yields several false positives, along with the experimentally validated true positives.[55,56] FTMap

addresses this limitation by incorporating a desolvation energy term based on a continuum electrostatics model.[57,64] In essence, FTMap's energy potentials encompass van der Waals forces, electrostatics, a cavity term (to depict the contribution of hydrophobic enclosure), and a statistical pairwise potential (to represent other solvation effects). The latter two terms offer a simplified description of how the protein environment influences solvation. As a result, the FTMap server replicates the results obtained with experimental protein solvent mapping[78] by minimizing and clustering molecular probes (16 different organic solvents) on the protein surface. The probes typically gather around crevices and/or pockets (e.g., binding sites), and when their clusters overlap, they create 'consensus' sites (CSs), where multiple solvent molecules are likely to bind. Among them, those with at least 16 probes clusters are designated as hotspots[79] (Figure 2). The results from solvent mapping can be downloaded as a PyMOL session or as a PDB file, which includes the consensus sites along with the coordinates of the query protein.

The default parameters of FTMap exclude all water molecules and ligands present in the query. As a result, pockets where cofactors and prosthetic groups bind are likely to accumulate most of the consensus sites. If the user wishes to prevent this situation, a mask should be generated using the coordinates of the pocket to be excluded from mapping.

Based on the position, size, and distance of these consensus sites found in the protein crevices, the sites can be categorized by the hotspots they display as: druggable by drug-like compounds (D), not druggable due to the absence of hotspots (N), druggable only by a large chemotype (e.g., foldamer) (D_L), druggable only by macrocycles, peptide mimetics, and charged compounds (D_s), borderline druggable (micromolar affinity) only by a large chemotype (B_L) and borderline druggable (micromolar affinity) only by macrocycles, peptide mimetics, and charged compounds (B_s).[80] The accurate classification of the target using this approach relies on selecting the appropriate consensus sites, as different combinations

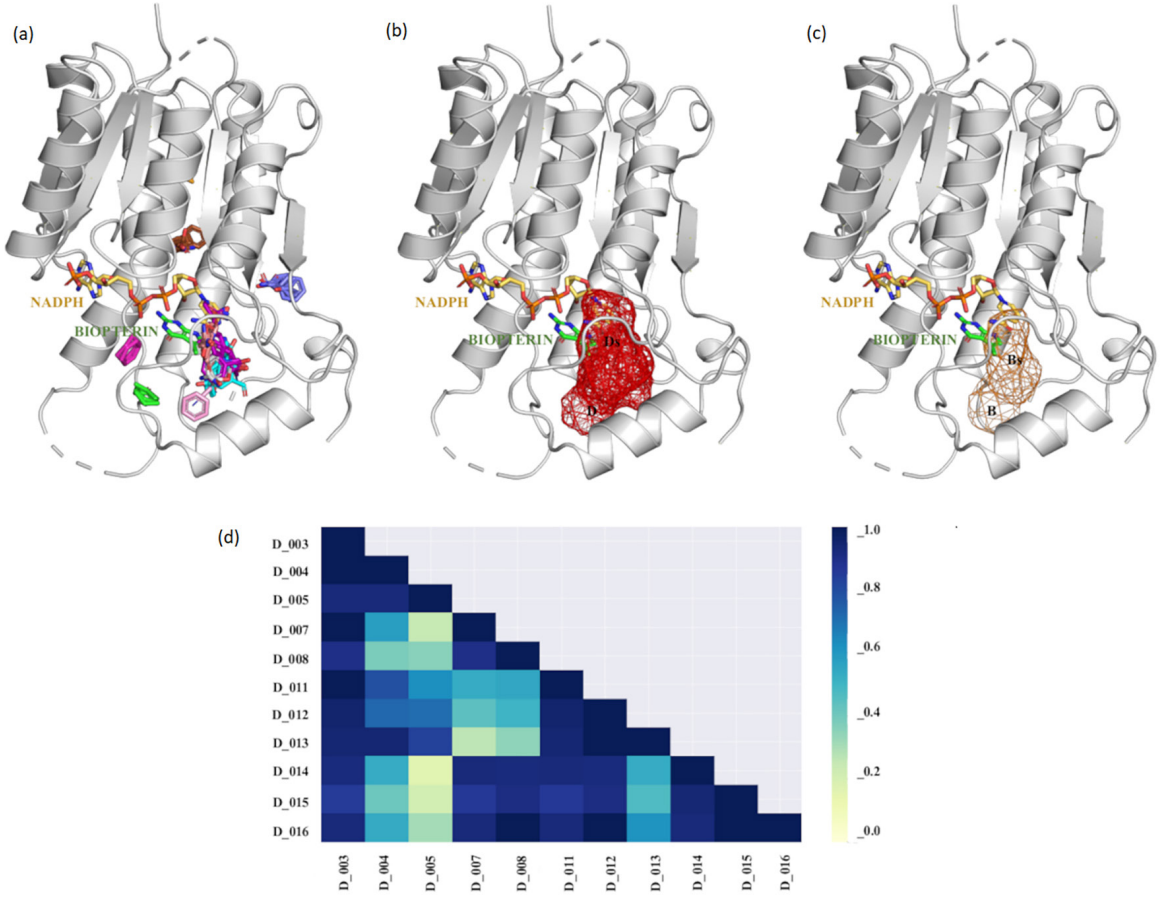

Figure 2. Solvent mapping of *Tb*PTR1 (PDB ID 4CLR).[47] (a) Consensus sites identified by FTMap when the cofactor binding site is occupied (protein mask = NADPH coordinates) and (b,c) hotspots calculated with the DRUGpy plugin using the output of FTMap (center: druggable (D and D_S); right: borderline-druggable (B and B_S)). The protein is depicted in cartoon, the molecular probe clusters are shown in sticks and ranked according to the number of probes in each cluster, and hotspots are shown as a mesh (red for D and D_S, orange for B and B_S); (d) heatmap showing the fractional overlap of druggable hotspots, ranging from zero (no overlap: light yellow) to 1.0 (complete overlap: dark blue).

of CSs yield the various types of hotspots employed to classify the targets druggability. Consequently, this task is biased by the experience of the medicinal chemist. To overcome this limitation, Teixeira *et al*.[81] devised a PyMOL plugin that automates the process and offers insights into the overlap among hotspots classified as D, Ds, B, and Bs, as well as between any of those hotspots and the ligand (Figure 2). First, the DRUGpy algorithm identifies all subsets of up to 3 CSs (default parameter) and calculates the center-to-center inter-distances, the maximum distance among the CSs, and the strength of the main CS, measured by the number of probe clusters in CS0, to assess if the ensemble of CSs forms a druggable (D or Ds) or a borderline-druggable (B or Bs) hotspot. Second, hotspots belonging to the druggable classes (D and Ds) are displayed as red meshes, while borderline-druggable (B and Bs) are depicted as orange meshes. Each hotspot is considered an object in PyMOL, allowing their superposition with each other or to another object (e.g., a ligand) to be calculated as a fractional overlap, ranging from zero (no overlap) to one (total superposition) (Figure 2d). The information provided by this plugin played a crucial role in bringing attention to underexplored features of the dihydroorotate dehydrogenase (DHODH) binding site,[82] as well as illuminating the optimal pathway to enhance the efficiency of ligands for a series of β-secretase enzyme 1 (BACE-1) inhibitors.[83]

The hotspots calculated with DRUGpy rely on solvent mapping carried out with FTMap, which considers the static information contained in the PDB structure used as the query. Thus, mapping an unliganded protein structure or a liganded one will provide different results. Accounting for side-chain flexibility does not significantly impact the solvent mapping result.[84] Consequently, PDB structures with minor shifts in their side chains generally provide similar hotspots. The results are not identical because minor modifications in the CSs' center-to-center and maximal distances may affect hotspot classification. One must be aware that even without a change in hotspot classification, its features (total number of probes, CSs center-to-center distance, maximal distance among CSs) might have slight changes, similar to those observed for side-chain movement.[84]

It has been shown that overall protein conformational flexibility must be considered for improved identification of binding sites, especially cryptic and allosteric ones, using the solvent mapping approach,[85] as CS location and strength change from one conformation to another. Unfortunately, there is no in-depth analysis of how protein conformation flexibility affects hotspot classification, requiring users to consider each case individually.

It is worth mentioning that FTMap provides a list of hydrogen-bonded and nonbonded contacts between the residues and the probes which can hint at crucial molecular interaction within the binding site.[86] However, the output of this server provides no information on the type of interactions (hydrophobic, H-donor or H-acceptor) that are responsible for maintaining a fragment-sized molecule inside the hotspot. Kozakova *et al*.[87] have demonstrated that the overlap of fragments with the main hot spot (largest probe cluster), quantified as the fractional overlap (FO), indicates whether the fragment binding mode is conserved upon fragment-to-lead development. Their assertion is based on the understanding that a significant portion of the protein-ligand binding free energy arises from interactions within the main hot spot. According to these authors, a fragment with an FO greater than 0.8 is expected to maintain its binding profile when its molecular weight is increased, as long as its pharmacophore groups are not modified. They note, however, that this condition is not necessary for retaining the binding mode, as there are cases where a fragment's binding profile is conserved even when the FO is below the specified threshold. For example, this may occur when the fragment exhibits high FO values to secondary, but still strong, consensus sites (i.e., > 13 probe clusters).

The lack of information regarding the molecular interaction responsible for preserving the fragment orientation within the hotspot is, to some extent, explained by the simplistic potentials employed for calculating their binding energy to the protein within a reasonable timeframe. According to Radoux *et al*.,[15] the Fragment Hotspot Map server[14] addresses this issue. Hence, the fundamentals of hotspot prediction with this server are summarized below: initially, atomic propensities, derived from a library of intermolecular interactions called Isostar, are extracted to indicate how many times more likely than random it is for an atomic probe (aromatic CH (hydrophobic), uncharged NH (donor), and carbonyl oxygen (acceptor)) to be located at a grid point on the protein surface, based on data from the Cambridge Structural Database.[88] Among the sampled grid points, those situated in cavities of the protein, as identified by the LIGSITE[15,89] algorithm, are also assigned a buriedness score ranging from zero (completely solvent-exposed) to seven (completely buried), and those scoring at least five are retained. To highlight the location of the hotspots, the remaining grid points are sampled with molecular probes (toluene, aniline, and phenol), to exclude false positives (i.e., fake binding sites), and weighted by their buriedness score (Figure 3).

The output of the Fragment Hotspot Maps server, in the form of a PyMOL session containing a grid file, can

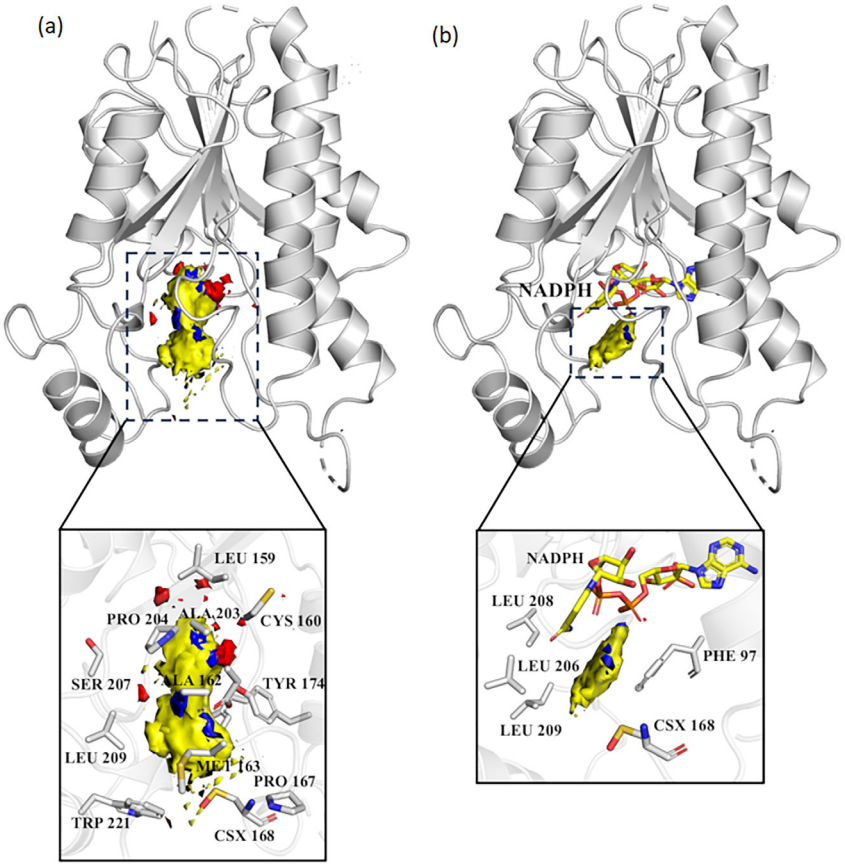

Figure 3. Hotspots of *Tb*PTR1 (PDB ID: 4CLR)[47] according to Fragment Hotspot Maps. (a) The calculation of atomic propensities was carried out with no ligand or water molecules in the input file; (b) the molecule of NADPH was considered for the calculation of atomic propensities. The hydrophobic map is depicted in yellow, H-bond donor in blue and H-bond acceptor in red. All maps are contoured at level 17. Residues surrounding (5 Å radius) the probe propensity maps are highlighted in the insets. CSX refers to a modified cysteine residue.

be examined at various contour levels, which lack specific physical meanings; instead, they indicate how closely a given interaction (hydrophobic, H-donor, or H-acceptor) mimics hotspot-yielding environment. A contour of 0 displays every grid point that has been sampled by the molecular probes, maps with contour levels of at least 10 highlight the protein's binding sites, while those with levels of at least 14 correspond to hotspots. Maps with contour levels higher than 17 emphasize molecular interactions that anchor the fragment in its position, within a strong hotspot. Consequently, a high-scoring polar interaction (e.g., H-donor, or H-acceptor maps scoring > 17) is anticipated to be situated in a deeply buried and hydrophobic environment.

4. Where Do the Hotspots Come From?

As demonstrated in the preceding section, hotspots arise from the protein environment surrounding them, characterized by features such as a concave topology and a mixed composition of hydrophobic and polar groups, described as a "mosaic-like" pattern by Kozakov *et al.*[12] These structural characteristics favor the binding of small molecules, including solvents and fragments. It has been emphasized that upon binding, fragments displace constrained water molecules,[90] including those unhappy waters situated in hydration sites with a positive ΔG (Gibbs free energy) compared to bulk water,[76] or those involved in strong polar interactions with the protein. The latter situation is only feasible if the ligand can replace a geometrically strained water-protein hydrogen bond, resulting in a significant entropy gain with minimal enthalpy loss. This water-centric hypothesis is one of the putative underlying reasons for the existence of hotspots. While there is not yet direct evidence for an alternative frustration-centric hypothesis, this section will build upon previous works that have already utilized either protein hotspots or their frustration to demonstrate how the complementary analysis provided by the other approach can lead to similar outcomes and conclusions. Although a validation study is beyond the scope of this review, we believe the reader will be convinced of the compelling evidence connecting frustration and hotspots in proteins.

4.1. Hotspot glimpses into frustrated proteins

Evolutionary constraints, which prioritize function, often come into conflict with those optimizing folding, resulting in local protein frustration.[37,91] To gain experimental insights linking frustration to the folding pathway and protein function, Di Silvio *et al*.[92] analyzed proteins with a common structural signature composed of 5 or 6 β-strands and 2 or 3 α-helices,[93] known as the PDZ domain, named after the proteins in which it was first observed,[94] focusing on those with two experimentally accessible sequential transition states (TS1 and TS2).[95] The first state is exemplified by protein tyrosine phosphatase-bas-like (PDZ2, PDB ID 1GM1),[96] with its active site located between helix a2 and strand b2. An insertion of five residues (VLFDK) at the beginning of the b2-b3 loop mimics the second state (PDZ2a, PDB ID 1OZI),[97] which lacks catalytic activity.[97] Comparing the frustration patterns between these two structures (native *versus* denatured-like) and the complexed PDZ2 structure (PDB ID 1VJ6)[98] (Figure 4) reveals an increased number of highly frustrated water-mediated contacts near the VLFDK-loop. In contrast, the frustration in the complex and free PDZ2 conformations is quite similar. Instead of attributing the lack of activity in PDZ2a to the increased frustration around this region, the authors of the original paper emphasize that the frustration patterns in the native and denatured-like conformations differ and they exploit kinetic data from 13 PDZ2a mutants (including the one that produced the 1OZI X-ray structure) to argue that differences in the early stages of folding explain why TS2 cannot adopt the correct (active) conformation.

Aiming to shed additional light on why the insertion of the VLFDK loop renders the protein inactive, we took a different approach, focusing on the mutational frustration within the binding site of denatured-like (PDB ID 1OZI) and native-like (PDB ID 1GM1) conformations, along with the position and shape of hotspots within them, as calculated with DRUGpy.

From a structural perspective, the additional loop impacts the orientation of the α2 helix and reduces the pocket's flexibility to accommodate ligands.[97] From the hotspots' perspective, the native-like conformation (PDB ID 1OZI) has a druggable hotspot in the active site. Its size, represented by the maximal distance (MD) among the CSs, decreases in the denatured-like conformation (PDB ID 1GM1 D1, MD = 10.3 Å *versus* PDB ID 1OZI D1, MD = 9.7 Å), being classified as "Ds" for "druggable small". The frustration pattern around this hotspot (within a 5 Å radius) shows that highly frustrated contacts in both native- and denatured-like structures are largely the same. While this result is disappointing, it illustrates that combining data from Frustratometer and FTMap does not always yield positive results.

One might argue that the shape of the main hotspot (D1) could explain why the denatured-like conformation lacks activity. To test this hypothesis, one would need to analyze hotspots in several proteins with the PDZ domain and compare their biological activities, what is beyond the scope of our work. A simpler approach involves using the FTMove server[99] to survey proteins with at least 90% sequence identity to a query that has the PDZ domain (e.g., 1GM1), and then conducting hotspot analysis on these proteins. This approach may miss some proteins with the PDZ domain but minimizes confounding effects due to sequence dissimilarities affecting hotspot location and size. Using this conservative approach, we identified 8 additional unique PDBs that have the PDZ domain (Table 2).

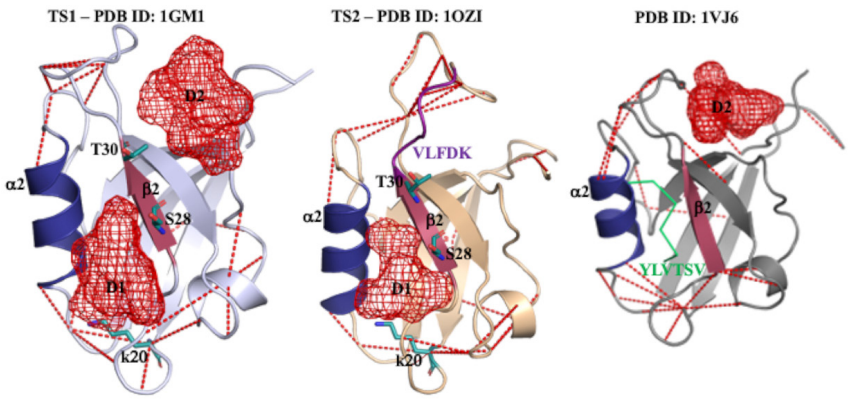

Figure 4. Frustration pattern and hotspot analysis of PDZ2as (PDB ID: 1OZI) and PDZ2 in free (PDB ID 1GM1) and complexed (PDB ID 1VJ6) forms. Mutational frustration of the protein was computed using the Frustratometer 2.0 server,[10] and hotspot analysis was performed with the DRUGpy plugin,[16] utilizing the output from the FTMap server.[11-13] The protein structure is represented in cartoon form, with helix α2 colored in density-blue and strand β2 in raspberry. The VLFDK is highlighted in purple, and the ligand of PDZ2 is presented in a green ribbon, while the druggable hotspots are depicted as red meshes. Water-mediated frustrated contacts are indicated by dashed red lines, while residues within 5 Å from the main hotspot (D1), making water-mediated frustrated contacts, are shown as sticks.

Clustering these structures with the three already discussed above, using their CSs as descriptors (Figure 5) shows that the denatured-like conformation (PDB ID 1OZI) clusters with the PDZ2 domain from human phosphatase HPTP1E (PDB ID 3PDZ), the *cis*-form of a photo-switchable PDZ domain crosslinked with an azobenzene derivative (PDB ID 2M0Z), the second PDZ domain from human PTP1E in complex with RA-GEF2 peptide (PDB ID 3LNY), and the second PDZ domain from human PTP1E (PDB ID 3LNX). The native-like conformation (PDB ID 1GM1) clusters with the PDZ2 domain from human phosphatase HPTP1E complexed with a peptide (PDB ID 1D5G), the apo form of a PDZ2 domain from hPTP1E (PDB ID 7QCX), the PDZ2 domain from hPTP1E complexed with the RA-GEF2 peptide (PDB ID 7QCY), the *trans*-form of a photo-switchable PDZ domain crosslinked with an azobenzene derivative (PDB ID 2M10), and PDZ2 from PTP-BL in complex with the C-terminal ligand from the APC protein (PDB ID 1VJ6). According to this analysis, PDBs in the denatured-like cluster have more similar CSs to each other than to those in the native-like cluster. If this similarity analysis is extended to the hotspots calculated from the CSs, proteins in the denatured-like cluster would be expected to have Ds hotspots in the pocket between helix α2 and strand β2, whereas proteins in the native-like cluster should have D hotspots in the equivalent pocket.

While all proteins within the denatured-like cluster, except the second PDZ domain from human PTP1E in complex with RA-GEF2 peptide (PDB ID 3LNY), have the expected Ds hotspots, the second PDZ domain from human PTP1E (PDB ID 3LNX) and the PDZ2 domain from human phosphatase HPTP1E (PDB ID 3PDZ) also have druggable hotspots at that location (Figure 5). This finding invalidates the initial hypothesis and highlights the need for further studies to understand when protein frustration and hotspot analysis will provide complementary insights, or not.

The catalytic activity of enzymes depends on the spatial orientation of residues involved in catalysis, even if they conflict with the overall fold.[46] To investigate how these conflicts impact the biological function of enzymes, Freiberger *et al*.[105] calculated the local frustration patterns for all protein enzymes from the Catalytic Site Atlas.[106] They observed that highly frustrated interactions follow the same pattern as the ones found in the binding or allosteric sites of globular proteins.[37,39,107] The density of minimally frustrated contacts around the catalytic residues, up to 3.5 Å, is below the overall density observed elsewhere in the enzymes, while the density of highly frustrated interactions is above average, regardless of the catalytic residue burial, or the catalytic mechanism involved. The primary contributor to this observation seems to be the identity of the residues surrounding the catalytic sites, as captured in the mutational frustration index, rather than their conformational state (configurational frustration).[31,108] Consequently, we incorporated a comparison of hotspots into the characterization of protein frustration in pairs of enzymes sharing the same fold (CATH classification),[109] but catalyzing different reactions, which has already been conducted by Freiberger *et al*.[105] (Figure 6).

Accordingly, both cyclomaltodextrin glucanotransferase (PDB ID 1CDG)[110] and glucan 1,4-α-maltohydrolase (PDB ID 1QHO)[111] exhibit a TIM barrel fold and possess the catalytic residues D-D-E-H. As inferred from the analysis conducted by Freiberger *et al*.,[105] the examination

Table 2. X-ray structures sharing at least 90% sequence identity to 1GM1, identified with FTMove,[99] which had their *in silico* solvent mapping assessed and hotspots calculated within DRUGpy. PDB IDs are shaded according to the hierarchical cluster analysis shown in Figure 5

PDB ID	UNIPROT	Number of CSs (CS0)[a]	Main hotspots class[b]	Reference
1D5G	Q12923	10 (17)	**D** and D_s	100
1GM1	Q64512	09 (19)	**D** and D_s	96
1OZI	Q64512	11 (21)	$\mathbf{D_s}$	97
1VJ6[c]	Q64512	07 (21)	$\mathbf{D_s}$	98
2M0Z	Q12923	10 (22)	$\mathbf{D_s}$	101
2M10	Q12923	10 (18)	**D** and D_s	101
3LNX	Q12923	09 (25)	**D** and D_s	102
3LNY	Q12923	09 (15)	**B** and B_s	102
3PDZ	Q12923	10 (21)	**D** and D_s	103
7QCX	Q12923	09 (18)	$\mathbf{B_s}$	104
7QCY	Q12923	09 (26)	**D** and B_s	104

[a]Number of consensus sites (CSs) calculated with FTMap and the strength of the main CS (CS0), measured by the number of probe clusters; [b]when more than one hotspot class is predicted in the active site, pocket located between helix α2 and strand β2, only the first one, highlighted in bold, is considered, since the other class can be viewed as a subset of the first; [c]the ligand of 1VJ6 was ignored during the solvent mapping in FTMove, thus the result shown here is different from the one depicted in Figure 4.

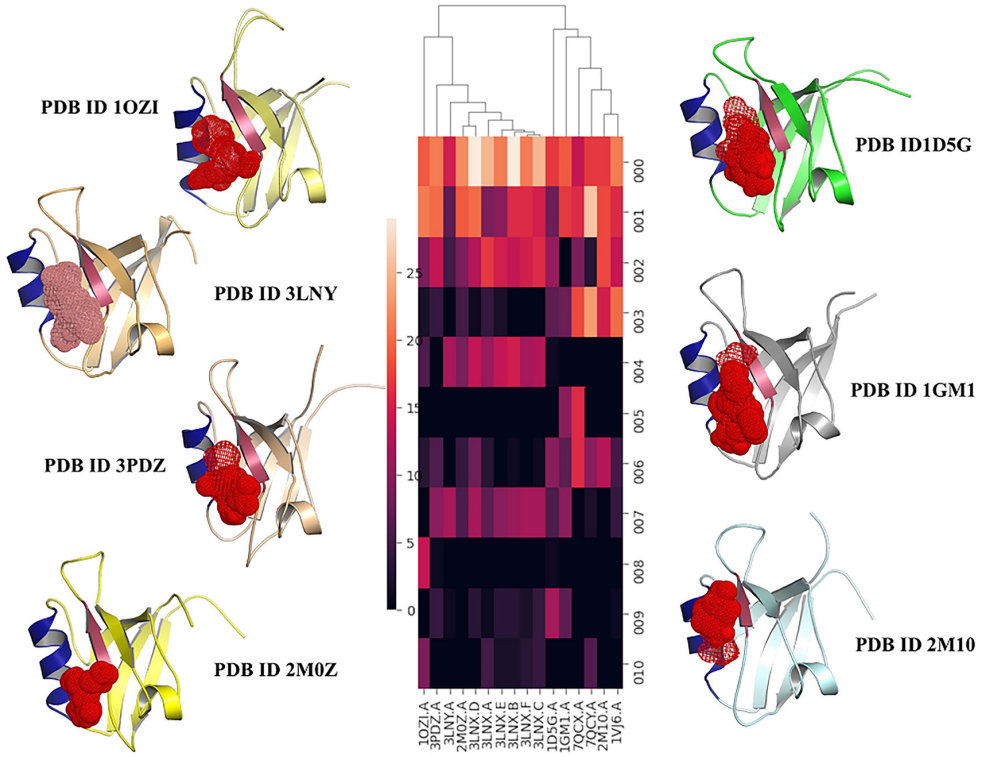

Figure 5. Hotspot analysis of representative X-ray structures sharing the PDZ domain, clustered according to their CSs using the FTMove server. (Left) The main hotspot from representative X-ray structures belonging to the same cluster as the denatured-like conformation (PDB ID 1OZI); (center) heatmap calculated by the FTMove server, colored according to the number of probe clusters in each CS, showing how the X-ray structures with the PDZ domain can be clustered into two families (as depicted in the dendrogram) based on their CSs. The CS number is represented on the right-hand side of the heatmap, while the number of probe clusters is on the left-hand side; (right) the main hotspot from representative X-ray structures belonging to the same cluster as the native-like conformation (PDB ID 1GM1). The protein structure is represented in cartoon, with helix α2 colored in density-blue and strand β2 in raspberry.

of frustrated contacts (mutational frustration) within a 5 Å radius from the main hotspots in 1CDG (druggable) and 1QHO (borderline-druggable) reveals a substantial proportion of residues making direct frustrated interactions in both enzymes: 1QHO has 16 out of 43 residues (37.2%) involved in frustrated molecular interactions, while 1CDG has 14 out of 39 residues (36%).

Moreover, the residues participating in frustrated interactions in both structures (1QHO: N13, F21, W101, D127, D190, R226, Y230, H233, D256, W258, F259; 1CDG: W101, A137, F183, R227, D229, A230, K232, P235, D257, D328) display similar features in equivalent positions. This result can be explained by the fact that the druggable hotspot from 1CDG partially overlaps with the borderline-druggable hotspot from 1QHO, whose classification changed due to a reduced number of probe clusters in the main CS (CS0). Besides the difference in hotspot strength, there is no clear distinction between them that would explain the preference for one enzyme's substrate over the other.

The lack of distinction between the hotspot features is expected since FTMap was not designed to identify the specific interactions responsible for the existence of the hotspot or for locking a ligand into a specific orientation within the binding site. As this type of information can be obtained through the Fragment Hotspot Map server (FHM),[14,15] comparing the predicted hotspots from this server might provide insights into the connection between the interactions underlying the hotspot and the localized frustration around the catalytic cleft. In contrast to FTMap, FHM predicts three hotspots in each structure, one of them located in the catalytic pocket, and one of them coincides with the location of a borderline-druggable hotspot in both enzymes (Figure S1, Supplementary Information section), hydrophobic interactions (indicated by the yellow surface contoured at a 17 level) are present in the same region as the druggable hotspot in 1CDG. On the other hand, only a small polar surface (acceptor) is found in the equivalent position in 1QHO.

The imidazole glycerol phosphate synthase subunit HisF (PDB ID 2A0N)[112] and the thiamine-phosphate diphosphorylase (PDB ID 2TPS)[113] share the same fold, with their catalytic sites located in the same region of the protein. However, the percentage of water-mediated highly frustrated residues around the hotspots (within a 4 Å radius) found in the pocket containing the catalytic residues of each

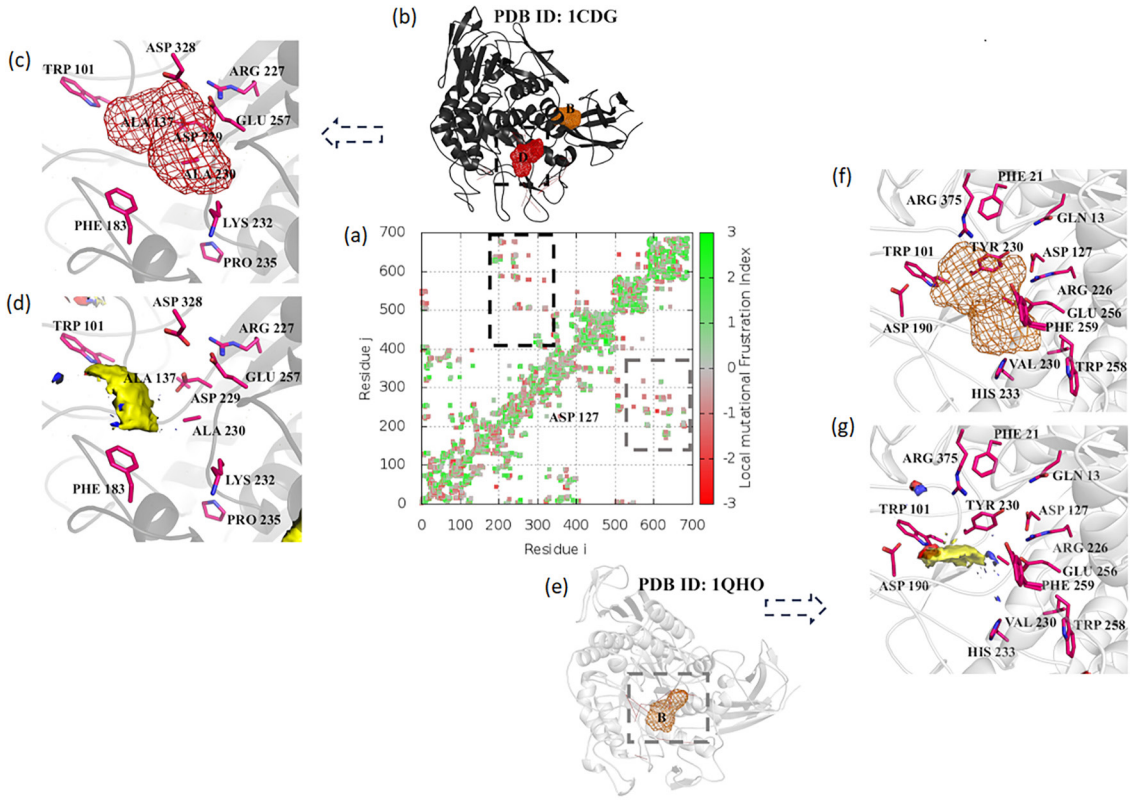

Figure 6. Protein frustration and hotspot analysis for a pair of proteins sharing the same fold (PDB IDs 1CDG and 1QHO) that catalyze different biochemical reactions. (a) Contact map of frustrated residues: green - minimally frustrated, red - highly frustrated. The upper left quadrant depicts the data for 1CDG, while the lower right quadrant shows the data for 1QHO. Residues from the catalytic cleft surrounding the main hotspot are boxed; (b) water-mediated frustrated interactions (dashed red lines) and the main hotspot (druggable shown in red mesh) for 1CDG; (c) zoomed view of frustrated residues (sticks) within 5 Å from the main hotspots in 1CDG; (d) propensity maps calculated with the FHM server for hydrophobic (yellow), donor (blue), and acceptor (red) interactions, contoured at level 17, for the 1CDG active site; (e) water-mediated frustrated interactions (dashed red lines) and the main hotspot (borderline-druggable shown in orange mesh) for 1QHO; (f) zoomed view of frustrated residues (sticks) within 5 Å from the main hotspots in 1QHO; (g) propensity maps calculated with the FHM server for hydrophobic (yellow), donor (blue), and acceptor (red) interactions, contoured at level 14 (transparent color) and 17 (solid color), for the 1QHO active site.

enzyme differs significantly: 8 out of 36 residues (23.5%) for the imidazole glycerol phosphate synthase subunit HisF, compared to 6 out of 15 residues (40%) for thiamine-phosphate diphosphorylase (Figure S2, Supplementary Information section). Another distinction is that according to the results provided by DRUGpy, the imidazole glycerol phosphate synthase subunit HisF is classified as a druggable target, whereas thiamine-phosphate diphosphorylase exhibits only borderline-druggable hotspots within the active site. A secondary borderline-druggable hotspot is predicted in this structure, close to the protein N-terminal portion, but no highly frustrated residues around this putative pocket is found. Moreover, the FHM server suggests this region does not have hotspots, probably due to its high solvent exposure. Taken together, these data suggest the secondary hotspot should not be considered for drug development purposes. Last but not least, the analysis of the propensity maps calculated with FHM suggests that polar interactions play a much more significant role in the hotspot of the imidazole glycerol phosphate synthase subunit HisF than in thiamine-phosphate diphosphorylase, as predicted by their hydrophobic surfaces contoured at level 14. This type of information is instrumental in understanding which modifications should be made in putative inhibitors of each enzyme to improve their affinity.

As highlighted earlier, globular proteins typically display a low degree of frustration in their tertiary structure, with a significant portion of highly frustrated residues and hotspots predominantly located in their orthosteric binding site. Nonetheless, it has been demonstrated that specific regions within many proteins, including binding interfaces, continually sample varied conformations.[114,115] The ability to explore these new conformational states has proven crucial for their interaction with different partners.[116] An analysis of local frustration in protein complexes, where the formation is accompanied by a transition from disordered to ordered forms upon binding,[18] indicates lower local frustration in the final complexes compared to the unliganded disordered forms of the proteins, although frustration is not always eliminated. These findings suggest

that binding-induced folding often results in suboptimal interactions at the binding interface and in the structured part of the protein.

Freiberg *et al.*[18] demonstrated the differential binding of fuzzy regions in the translation initiation factor 2 subunit gamma (eIF2g) through the analysis of protein structures 3CW2[117] and 3I1F.[118] As reported in their work, the crystallographic conformation of the fuzzy region (residues 39-47) is stabilized by intramolecular interactions between Glu39-Thr46 and Glu40-Gly44. In contrast, the 3I1F structure shows stabilization due to charge-charge interactions between Glu39 and Arg43. Additionally, the Gly44 main chain in the latter forms a hydrogen bond with a Lys42 side-chain, while in the 3CW2 structure, Lys42 interacts with Asp283 of the structured domain. These distinct interaction profiles reflect different strategies for addressing localized frustration in each eIF2g conformation (Figure 7). DRUGpy-calculated hotspots indicate that both conformations only exhibit borderline-druggable hotspots. However, the shape and location of the hotspots are specific to each conformation. In 3CW2, all predicted hotspots are within 4 Å of the fuzzy region. Conversely, the dissimilar conformation of this region in 3I1F makes the cavity smaller than in 3CW2. As a consequence, hotspots are distributed between the fuzzy region and another pocket on the opposite side of the protein surface. The authors also discussed an example of fuzzy binding involving mitogen-activated protein kinase 10, based on the PDB structures 3V6R[119] and 4H3B.[120] Neither DRUGpy nor the FHM server predicts hotspots within 5 Å of that region. However, the raw output of FTMap indicates a consensus site with 10 probes near the fuzzy region (residues 369-382), and FHM server identifies the same region as a putative binding site (propensity maps with level > 10) (data not shown). The absence of hotspots in this region is not unexpected, given its location in a solvent-exposed and convex area.

Through the analysis of frustration in proteins capable of adopting numerous conformational states, commonly referred to as "intrinsically disordered" (IDP), Gianni *et al.*[116] proposed a unified model that links the frustration concept, accounting for conformational variations in proteins, with fuzziness, capturing the functional heterogeneity within protein interactions,[121] to elucidate how proteins exploit these features for fine-tuned, context-dependent regulation through binding to different partners. They conclude that frustration and fuzziness

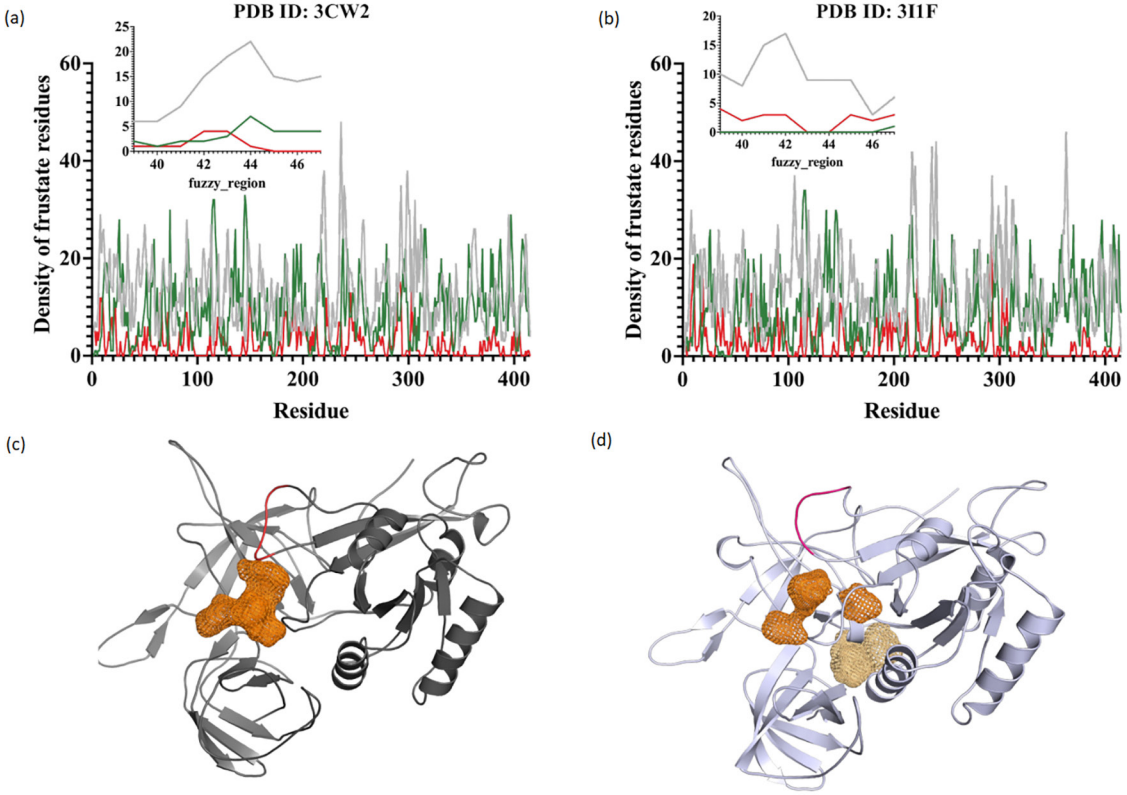

Figure 7. Hotspot analysis of frustrated fuzzy regions in the translation initiation factor 2 subunit gamma (eIF2g) (PDB IDs 3CW2 (a,c) and 3I1F (b,d)). (a,b) Density of neutral (gray), highly (red), and minimally (green) frustrated residues, as calculated by the mutational frustration index within the Frustratometer webserver.[10] The plot insets highlight the different patterns of frustration in the fuzzy region (residues E39-I47) for each crystallographic structure; (c,d) borderline-druggable (B) hotspots (orange) predicted with DRUGpy, within a 5 Å radius from the fuzzy region in eIF2g structures (colored in red). A secondary cluster of B hotspots is predicted to exist on the opposite face of 3I1F (light orange).

emerge from common physical principles of the energy landscape, i.e., its ruggedness, and can be considered as two sides of the same coin. Since the binding heterogeneity first observed in IDPs is much more common than initially expected,[122] it is important to make readers aware that "a fuzzy complex" exhibits structural heterogeneity that is maintained upon protein-protein interactions.[123] This heterogeneity reflects the structural and dynamic continuum between the unliganded and liganded conformations, and a more comprehensive description of fuzziness and its impact on binding can be found in the works of Fuxreiter.[121,124] While fuzziness and hotspots may not exhibit the same feature, it is evident that the divergent conformations adopted by IDPs significantly influence the strength and localization of hotspots.

A similar observation was made for *Pseudomonas aeruginosa* phosphomannomutase/phosphoglucomutase (PMM/PGM), where the catalytic mechanism relies on the movement of its C-terminal domain.[125] Although FTMap's top 3 consensus sites are located in the active site for three out of four representative PMM/PGM structures investigated, additional hot spots at domain-domain interfaces and hinge regions were identified, depending on the conformation the protein adopts (open, half-open, and closed conformers). In that study, hot spots close to each other or partially overlapping were combined. Based on the findings reported for this PMM/PGM, the authors argue that computational solvent mapping proves to be a more sensitive approach for detecting changes in binding sites and interdomain crevices than a simple RMSD (root mean square deviation) comparison among the structures. While substantial differences between the hot spot structures were observed upon large conformational shifts, the impact of side-chain movements is minor and can be accounted for using FTflex,[84] which employs a rotamer library to determine alternative conformations for residues within a 5 Å radius from the selected hotspot(s). The server automatically selects the set of rotamers that maximizes the number of probe clusters within each consensus site (CS), thus highlighting the "optimal" (maximally opened binding site) conformation for binding a ligand.

Whenever there is an "induced fit" or "conformational selection", electrostatic interactions play a role as interacting partners come into close proximity.[126,127] Therefore, Tsai *et al*.[51] investigated the role of electrostatic interactions in shaping both folding funnels and binding landscapes by introducing long-range electrostatic interactions (Debye-Hückel terms) into the AWSEM package.[51] Although no improvement was achieved in monomeric structure prediction, they demonstrated that the addition of long-range electrostatics does influence the binding mechanism, especially when a charged partner is involved. In some cases, long-range electrostatic interactions provide charge-charge stabilization, while in others, they cause frustration in the landscape that negatively impacts protein dimer formation (protein-protein binding). Since the paper did not describe any protein-small molecule complexes, a comparison of protein frustration, considering or not the Debye-Hückel terms, and hotspot locations is presented for the selected proteins described herein (Figure 8).

The dataset considered here (Table 3) suggests there is no single frustration index (mutational or configurational) that consistently exhibits a strong correlation with hotspots in the proteins investigated here. For 2TPS and 2A0N, the percentage of frustrated residues surrounding the main hotspot is higher when mutational frustration is taken

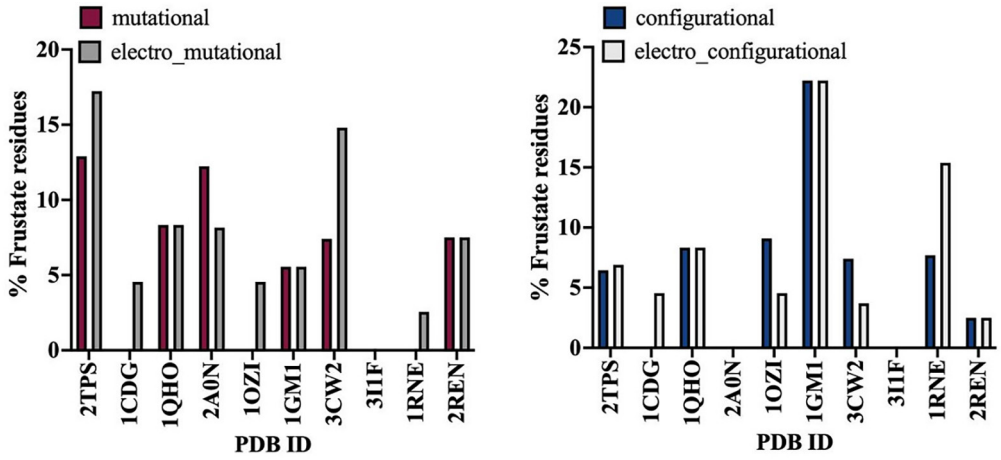

Figure 8. Percentage of highly frustrated residues (top 10%) within 5 Å of the main hotspot for unliganded PDZ2 (PDB ID 1GM1), PDZ2as (PDB ID 1OZI), cyclomaltodextrin glucanotransferase (PDB ID 1CDG), glucan 1,4-α-maltohydrolase (PDB ID 1QHO), imidazole glycerol phosphate synthase subunit HisF (PDB ID 2A0N), thiamine-phosphate diphosphorylase (PDB ID 2TPS) and the translation initiation factor 2 subunit gamma (eIF2g) (PDB IDs 3CW2 and 3I1F), and Renin (PDB Ids 2REN, 1RNE) in the presence or absence of long-range electrostatic interactions (mutational: left-hand side; configurational: right-hand side).

into consideration, whereas for the second PDZ domain of PTP-BL (PDB ID 1GM1), the proportion of frustrated residues appears to be much larger when configurational frustration is considered. Despite this shortcoming, the results presented in Figure 8 suggest that incorporating long-range electrostatic terms into the frustration calculation shows similar (PDB IDs1QHO, 1GM1, and 2REN)[129] or improved (PDB IDs 2TPS, 1CDG, 1OZI, 3CW2, and 1RNA)[130] correlation for all proteins, except for the imidazole glycerol phosphate synthase subunit hisF (PDB ID 2A0N) when mutational frustration is considered.

Table 3. Dataset of protein binding small molecules employed to compare protein frustration, considering or not the Debye-Hückel terms

PDB ID	UNIPROT	Number of CSs (CS0)[a]	Main hotspots class[b]	Reference
1GM1[c]	Q64512	9 (19)	D	96
1OZI	Q64512	11 (21)	D_s	97
1CDG	P43379	11 (18)	D	110
1QHO	P19531	13 (15)	B	111
2A0N	Q9X0C6	9 (20)	D	112
2TPS	P39594	15 (17)	D	113
3CW2[c]	Q980A5	14 (15)	B	117
3I1F	Q980A5	12 (17	D	118
2REN[c]	P00797	9 (21)	D	128
1RNE	P00797	9 (25)	D	129

[a]Number of consensus sites (CSs) calculated with FTMap and the strength of the main CS (CS0), measured by the number of probe clusters; [b]when overlapping hotspots from the same class were predicted in the active site, all of them were employed to calculate the 5 Å radius mentioned in Figure 8; [c]unliganded X-ray structures.

In the case of a similar analysis carried out for configurational mutation, the inclusion of Debbie-Hückel terms yields comparable results for 2TPS, 1GM1, 1QHO, and 2REN, whereas 1OZI and 3CW2 have fewer frustrated residues than when these terms are not considered.

4.2. Uncharted frustration around hotspots in proteins

Brenke *et al*.[78] reported the *in silico* solvent mapping study on Renin, a well-known drug target for controlling increased blood pressure. They mapped both the ligand-free Renin structure (PDB ID 2REN) and the structures co-crystallized with aliskiren (PDB ID 2V0Z)[131] or a peptidomimetic inhibitor (PDB ID 1RNE). According to their results, the strength of the main binding site (CS0) is highest for the apo structure (22 probe clusters), followed by the aliskiren-bound structure (18 probe clusters) and the peptide-bound structure (14 probe clusters). Using the current version of the FTMap server (accessed in December 2023),[11] with standard parameters, one finds a somewhat different order (Apo structure 25 probe clusters, peptide-bound structure 21 probe clusters, and aliskiren-bound structure 19 probe clusters), although the locations of the binding sites remain largely unchanged. As the coordinates of the original consensus sites are not available, the ones calculated with the current version of the FTMap server were considered to calculate the number of frustrated residues around the main consensus site (CS0), as well as the druggable hotspots in each of these structures. The peptidomimetic-bound structure displays a lower proportion of highly frustrated residues (mutational frustration) around both the druggable hotspots (8 out of 39, 20.5%) and the main consensus site (4 out of 17, 24%) than the unliganded structure (9 out of 35, 25.7% for the druggable hotspots, and 2 out of 17, 12% for CS0) or the aliskiren-bound structure (12 out of 50, 24% for the druggable hotspots and 3 out of 16, 19% for CS0) (Figure 9, upper panel).

According to Brenke *et al*.,[78] the fact that the subsite where aliskiren, the first orally available drug targeting Renin,[132] displays the third-largest consensus site (fourth-largest probe cluster according to the solvent mapping carried out in December 2023), suggests this pocket exists before any ligand binds to it. The mutational frustration pattern around this subpocket changes with the presence of the ligand: in both cases analyzed, ligand binding decreases the number of residues engaged in frustrated interactions compared to the protein in the apo form (Figures 9g-9i). This example highlights the anticipated reduction in frustration (protein stabilization) due to ligand binding.

Furthermore, the authors highlight the largest, albeit biologically irrelevant, difference among the three structures, as revealed by FTMap, is a consensus site located between the S2 and S4 subsites that is present in both the apo-structure and the peptidomimetic-bound protein, but absent in the aliskiren-bound structure. The analysis of druggable hotspots suggests yet another difference: the number of consensus site combinations that yield druggable hotspots: 5 in the ligand-free Renin, 10 in the aliskiren-bound, and 16 in the peptidomimetic-bound structure (Figure S3, Supplementary Information section). This result indicates that upon ligand interaction, the binding site becomes even more druggable. This finding is in good agreement with the assertion that solvent mapping sheds light on subtle but crucial changes in the binding site.[125] Last but not least, the druggable hotspot in the peptide-bound Renin structure does not entirely encompass the S3 subpocket. This suggests that while a strong consensus site (CS) is identified in this subpocket when analyzing the APO structure with FTMap, ligand binding results in a conformational selection that displays

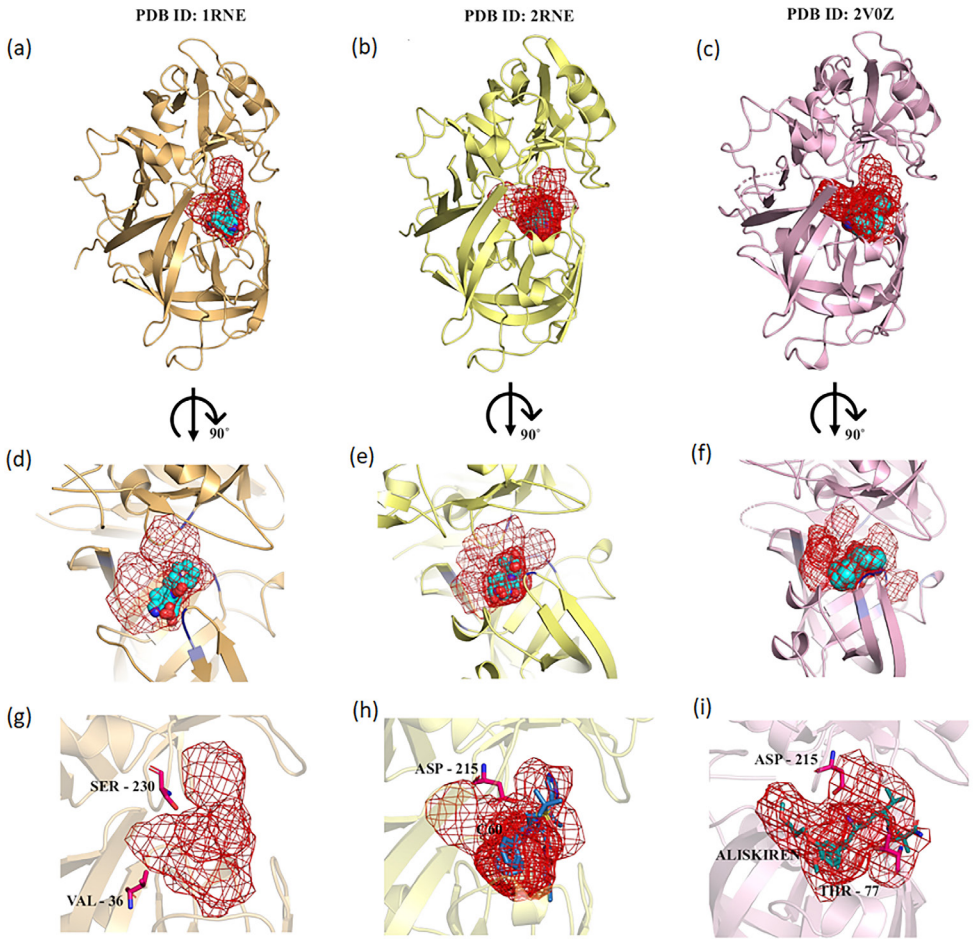

Figure 9. Mutational frustration around druggable hotspots in Renin. (Upper panel) druggable hotspots (red mesh) predicted from the output of the FTMap server (accessed in December 2023)[11] and the main consensus site (CS0 spheres) found in Renin (a) peptide-bound structure, (b) apo structure, (c) aliskiren-bound structure; (middle panel) (d-f) zoomed view of residues within a 5 Å radius from CS0 (dark-blue) or the druggable hotspots (light-blue) making highly frustrated interactions (mutational frustration); (lower panel) mapping residues that make highly frustrated interactions within the S3 subpocket shows that ligand binding ((h) peptidomimetic inhibitor, (i) aliskiren) changes the mutational frustration pattern of the apo form (g) within this region.

a somewhat different distribution of CSs, particularly in the S3 subpocket.

Previously, we conducted *in silico* solvent mapping on representative structures of DHODH from *Trypanosoma cruzi* (*Tc*DHODH) to capture the effect of conformational flexibility on hotspots.[82] This enzyme is responsible for the oxidation of dihydroorotate (DHO) to orotate (ORO) and acts as a soluble fumarate reductase, playing a crucial role in the parasite's *de novo* pyrimidine biosynthesis.[133]

The calculation of borderline-druggable and druggable hotspots suggests that when the catalytic loop (residues 129-137) is closed, there are no hotspots within the active site, as observed in the representative structure (PDB ID 2E6A). In the partially open loop conformation, druggable hotspots form, as seen in the DHODH complex with W76 (PDB ID 3W76), which turn into borderline-druggable if the loop adopts a wide-open conformation (PDB ID 3W72).[82]

The analysis of localized frustration in these three X-ray structures (Figure 10) reveals that overall water-mediated frustrated contacts are lowest in the DHODH-orotate complex (PDB ID 2E6A) and highest in the complex DHODH-W72 (PDB ID 3W72) (Figure 10, upper panel). The same trend is observed if only direct-frustrated contacts are considered (PDB ID 2E6A: 111 highly frustrated contacts, PDB ID 3W76: 168 highly frustrated contacts, and PDB ID 3W72: 179 highly frustrated contacts). Focusing on the catalytic loop, the number of residues making highly direct-frustrated contacts in the DHODH-orotate complex is twice the number found in the DHODH-W76 complex (4 *versus* 2 residues), which might explain the presence of druggable hotspots in that structure. The number of residues, in the catalytic loop, making highly frustrated direct contacts in the complex DHODH-W72 (PDB ID 3W72) is equal to one observed in the DHODH-orotate complex, possibly contributing to the "borderline-druggable" classification of the hotspots in that structure.

In our previous work,[82] the mentioned hotspots were calculated after excluding the FMN binding site from the solvent mapping step. This restraint was not applied when using the FHM server, as evidenced by a hydrophobic yellow surface over FMN (see Figure 8 in Froes *et al.*).[82] When this region is excluded from the FHM analysis, propensity maps contoured at level 17 do not show significant changes. However, when contoured at level 14, it becomes clear that in the closed-loop conformation (PDB ID 2E6A), it is easier to design molecules to bind the pocket "behind" the catalytic loop, than to target the active site. The propensity maps within this pocket are closer to residues not making highly-frustrated direct contacts. A similar trend is observed for the wide-open conformation. Although, propensity maps and hotstpot classification (druggable *versus* borderline-druggable), seem to be affected by localized frustration in their vicinity, it is too early to state if this is a general rule or to speculate on its implications for hit-to-lead optimization.

5. Concluding Remarks

The presence or absence of hotspots, their locations, as well as the percentage and localization of frustrated residues in each protein, are due to the 3D structure of the protein, which is ultimately encoded in its amino acid sequence. Therefore, these features are intertwined, and the preceding section reveals that residual protein frustration around (druggable) hotspots is significantly higher compared to the entire binding site or other regions on the protein surface. This indirectly suggests localized protein frustration might serve as the "driving force" for hotspot formation at specific pockets and crevices, thus being useful for the development of more potent ligands against selected therapeutic targets. However, to establish a frustration-centric hypothesis explaining the origin of hotspots, validation studies are necessary on proteins with fully characterized localized frustration and hotspots. This requires benchmarks where both experimental and computational analyses have been applied. Consequently, constructing such datasets is crucial for advancing our understanding of the intricate relationship between protein frustration and the emergence of hotspots. Achieving this may involve protein engineering guided by energy landscape theory,[134] followed by the experimental solvent mapping pipeline described by Agarwal *et al.*[135] While this goal is not yet attainable, a reasonable starting

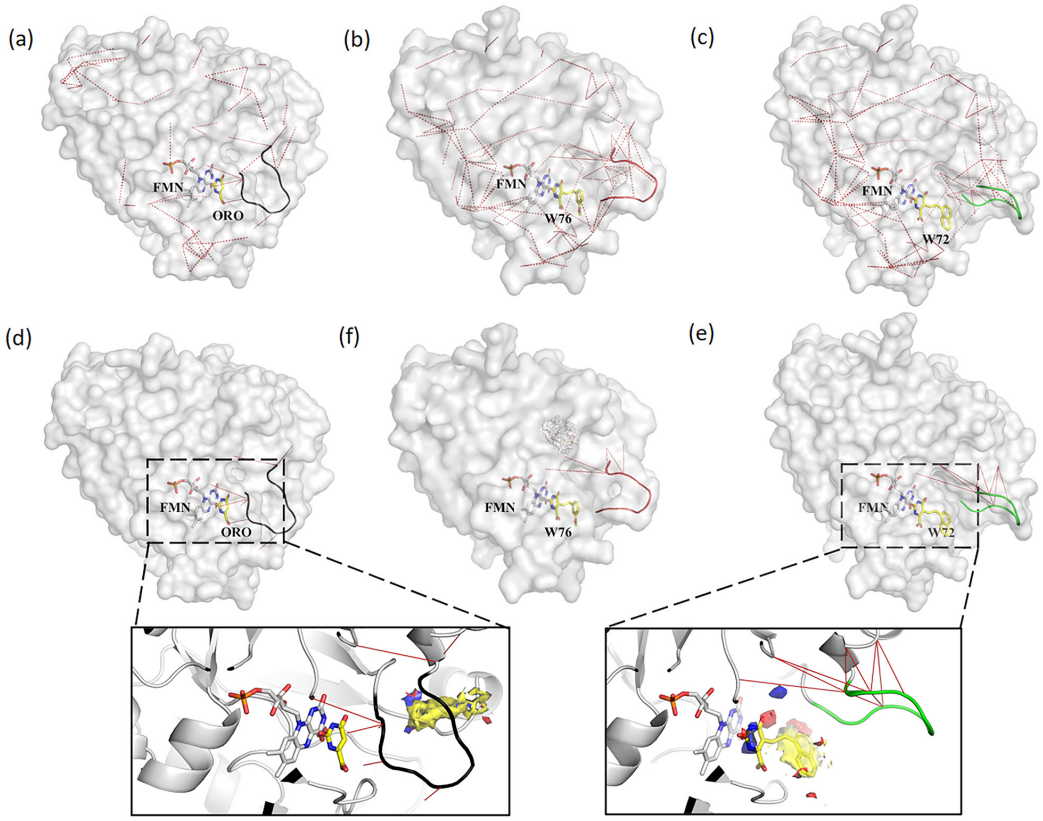

Figure 10. Localized protein frustration in representative crystallographic structures of TcDHOD (left: closed loop, center: open-loop, right: wide-open), as suggested by Froes *et al.*[82] (a-c) Direct (red solid-lines) and water mediated (red-dashed lines) highly frustrated contacts; (d-f) direct frustrated contacts in the flexible catalytic loop (black: closed, red: open, green: wide open). The inset highlights the probe propensity maps, calculated with FHM, when the cofactor (FMN) binding site is masked. The color scheme and contour level are the same as in previous figures.

point would be to explore the potential connection between reduced frustration and the absence of hotspots in less-druggable protein targets.[136]

On the other hand, it is equally crucial to investigate how well coarse-grained forcefields, such as AWSEM (utilized by the Frustratometer engine), can predict the location of hotspots. A more resource-intensive approach would be to employ the Charmm force field (used by FTMAP) to calculate all-atom frustration indexes. Improvement on either front would signify a better understanding of the physical parameters underlying hotspot origin. Consequently, this would be instrumental in supporting or refuting the frustration-centric origin of hotspots.

Supplementary Information

Supplementary information is available free of charge at http://jbcs.sbq.org.br, as PDF file.

Thamires Q. Froes *has a degree in Pharmacy and master's degree from UFBA. PhD in Biotechnology from UEFS. She was a postdoc at the Protein Crystallography Laboratory of Ribeirão Preto (FCFRP) for 4 years and is currently a postdoc at the Fiocruz (BA). She has experience in the medicinal chemistry, structure-based drug design and biological assays.*

Marcelo S. Castilho *has a degree in Pharmacy from Universidade Estadual Paulista Júlio de Mesquita Filho (1997), a master's degree from the Chemical Institute of Universidade de São Paulo, Ribeirão Preto (2000) and a PhD in Physics from Universidade de São Paulo. He held several positions in the Medicinal Chemistry and Biological Chemistry division of the Brazilian Chemical Society and received the vanguard researcher award at the 10th edition of BrazMedChem (2022). Presently he is a full professor of Medicinal Chemistry at the Universidade Federal da Bahia. He has large experience with biological assays and molecular modeling.*

References

1. Pedreira, J. G. B.; Franco, L. S.; Barreiro, E. J.; *Curr. Top. Med. Chem.* **2019**, *19*, 1679. [Crossref]
2. Dambach, D. M.; Misner, D.; Brock, M.; Fullerton, A.; Proctor, W.; Maher, J.; Lee, D.; Ford, K.; Diaz, D.; *Chem. Res. Toxicol.* **2016**, *29*, 452. [Crossref]
3. Seyhan, A. A.; *Transl. Med. Commun.* **2019**, *4*, 18. [Crossref]
4. Brown, S. P.; Muchmore, S. W.; *J. Med. Chem.* **2009**, *52*, 3159. [Crossref]
5. Arodola, O. A.; Soliman, M. E. S.; *Drug Des., Dev. Ther.* **2017**, *11*, 2551. [Crossref]
6. Ferreira, L. L. G.; Andricopulo, A. D.; *Drug Discovery Today* **2019**, *24*, 1157. [Crossref]
7. Neves, B. J.; Braga, R. C.; Alves, V. M.; Lima, M. N. N.; Cassiano, G. C.; Muratov, E. N.; Costa, F. T. M.; Andrade, C. H.; *PLoS Comput. Biol.* **2020**, *16*, e1007025. [Crossref]
8. Neves, B. J.; Braga, R. C.; Melo-Filho, C. C.; Moreira-Filho, J. T.; Muratov, E. N.; Andrade, C. H.; *Front. Pharmacol.* **2018**, *9*, 1275. [Crossref]
9. Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G.; *Annu. Rev. Phys. Chem.* **1997**, *48*, 545. [Crossref]
10. Frustratometer Server, http://frustratometer.qb.fcen.uba.ar/, accessed in August 2024.
11. FTMap Computational Solvent Mapping, https://ftmap.bu.edu/login.php, accessed in August 2024.
12. Kozakov, D.; Grove, L. E.; Hall, D. R.; Bohnuud, T.; Mottarella, S. E.; Luo, L.; Xia, B.; Beglov, D.; Vajda, S.; *Nat. Protoc.* **2015**, *10*, 733. [Crossref]
13. Ngan, C.-H.; Hall, D. R.; Zerbe, B.; Grove, L. E.; Kozakov, D.; Vajda, S.; *Bioinformatics* **2012**, *28*, 286. [Crossref]
14. Fragment Hotspot Maps, https://fragment-hotspot-maps.ccdc.cam.ac.uk/, accessed in August 2024.
15. Radoux, C. J.; Olsson, T. S. G.; Pitt, W. R.; Groom, C. R.; Blundell, T. L.; *J. Med. Chem.* **2016**, *59*, 4314. [Crossref]
16. DRUGPy, https://pypi.org/project/DRUGpy/, accessed in August 2024.
17. Nussinov, R.; Liu, Y.; Zhang, W.; Jang, H.; *RSC Chem. Biol.* **2023**, *4*, 850. [Crossref]
18. Freiberger, M. I.; Wolynes, P. G.; Ferreiro, D. U.; Fuxreiter, M.; *J. Phys. Chem. B* **2021**, *125*, 2513. [Crossref]
19. Karlsson, E.; Andersson, E.; Dogan, J.; Gianni, S.; Jemth, P.; Camilloni, C.; *J. Biol. Chem.* **2019**, *294*, 1230. [Crossref]
20. Morando, M. A.; Saladino, G.; D'Amelio, N.; Pucheta-Martinez, E.; Lovera, S.; Lelli, M.; López-Méndez, B.; Marenchino, M.; Campos-Olivas, R.; Gervasio, F. L.; *Sci. Rep.* **2016**, *6*, 24439. [Crossref]
21. Maus, H.; Hinze, G.; Hammerschmidt, S. J.; Schirmeister, T.; Basché, T.; *ChemistryEurope* **2023**, *1*, e202300060. [Crossref]
22. Draper-Joyce, C.; Furness, S. G. B.; *ACS Pharmacol. Transl. Sci.* **2019**, *2*, 285. [Crossref]
23. Nicolaï, A.; Delarue, P.; Senet, P.; *Sci. Rep.* **2015**, *5*, 18128. [Crossref]
24. Teruel, N.; Mailhot, O.; Najmanovich, R. J.; *PLoS Comput. Biol.* **2021**, *17*, e1009286. [Crossref]
25. Wodak, S. J.; Paci, E.; Dokholyan, N. V.; Berezovsky, I. N.; Horovitz, A.; Li, J.; Hilser, V. J.; Bahar, I.; Karanicolas, J.; Stock, G.; Hamm, P.; Stote, R. H.; Eberhardt, J.; Chebaro, Y.;

Dejaegere, A.; Cecchini, M.; Changeux, J.-P.; Bolhuis, P. G.; Vreede, J.; Faccioli, P.; Orioli, S.; Ravasio, R.; Yan, L.; Brito, C.; Wyart, M.; Gkeka, P.; Rivalta, I.; Palermo, G.; McCammon, J. A.; Panecka-Hofman, J.; Wade, R. C.; Di Pizio, A.; Niv, M. Y.; Nussinov, R.; Tsai, C.-J.; Jang, H.; Padhorny, D.; Kozakov, D.; McLeish, T.; *Structure* **2019**, *27*, P566. [Crossref]
26. Sosnick, T. R.; *Prot. Sci.* **2008**, *17*, 1308. [Crossref]
27. Nussinov, R.; Tsai, C.-J.; *Cell* **2013**, *153*, 293. [Crossref]
28. Maloney, R. C.; Zhang, M.; Liu, Y.; Jang, H.; Nussinov, R.; *Cell. Mol. Life Sci.* **2022**, *79*, 281. [Crossref]
29. Ding, J.; Lee, Y.-T.; Bhandari, Y.; Schwieters, C. D.; Fan, L.; Yu, P.; Tarosov, S. G.; Stagno, J. R.; Ma, B.; Nussinov, R.; Rein, A.; Zhang, J.; Wang, Y.-X.; *Nat. Commun.* **2023**, *14*, 714. [Crossref]
30. Lu, S.; Jang, H.; Muratcioglu, S.; Gursoy, A.; Keskin, O.; Nussinov, R.; Zhang, J.; *Chem. Rev.* **2016**, *116*, 6607. [Crossref]
31. Clarke, J.; Cota, E.; Fowler, S. B.; Hamill, S. J.; *Structure* **1999**, *7*, P1145. [Crossref]
32. Onuchic, J. N.; Wolynes, P. G.; *Curr. Opin. Struct. Biol.* **2004**, *14*, 70. [Crossref]
33. Wolynes, P. G.; *Philos. Trans. R. Soc., A* **2005**, *363*, 453. [Crossref]
34. Bryngelson, J. D.; Onuchic, J. N.; Socci, N. D.; Wolynes, P. G.; *Proteins: Struct., Funct., Bioinf.* **1995**, *21*, 167. [Crossref]
35. Bryngelson, J. D.; Wolynes, P. G.; *Proc. Natl. Acad. Sci. U S. A.* **1987**, *84*, 7524. [Crossref]
36. Ferreiro, D. U.; Komives, E. A.; Wolynes, P. G.; *Curr. Opin. Struct. Biol.* **2018**, *48*, 68. [Crossref]
37. Ferreiro, D. U.; Hegler, J. A.; Komives, E. A.; Wolynes, P. G.; *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 19819. [Crossref]
38. Miyashita, O.; Onuchic, J. N.; Wolynes, P. G.; *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 12570. [Crossref]
39. Ferreiro, D. U.; Hegler, J. A.; Komives, E. A.; Wolynes, P. G.; *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 3499. [Crossref]
40. Lubchenko, V.; *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 10635. [Crossref]
41. Gopi, S.; Singh, A.; Suresh, S.; Paul, S.; Ranu, S.; Naganathan, A. N.; *Phys. Chem. Chem. Phys.* **2017**, *19*, 20891. [Crossref]
42. Capelli, R.; Paissoni, C.; Sormanni, P.; Tiana, G.; *J. Chem. Phys.* **2014**, *140*, 195101. [Crossref]
43. Schafer, N. P.; Kim, B. L.; Zheng, W.; Wolynes, P. G.; *Isr. J. Chem.* **2014**, *54*, 1311. [Crossref]
44. Davtyan, A.; Schafer, N. P.; Zheng, W.; Clementi, C.; Wolynes, P. G.; Papoian, G. A.; *J. Phys. Chem. B* **2012**, *116*, 8494. [Crossref]
45. Truong, H. H.; Kim, B. L.; Schafer, N. P.; Wolynes, P. G.; *J. Chem. Phys.* **2013**, *139*, 121908. [Crossref]
46. Ferreiro, D. U.; Komives, E. A.; Wolynes, P. G.; *Q. Rev. Biophys.* **2014**, *47*, 285. [Crossref]
47. Khalaf, A. I.; Huggan, J. K.; Suckling, C. J.; Gibson, C. L.; Stewart, K.; Giordani, F.; Barrett, M. P.; Wong, P. E.; Barrack, K. L.; Hunter, W. N.; *J. Med. Chem.* **2014**, *57*, 6479. [Crossref]
48. DeLano, W. L.; *The PyMOL Molecular Graphics System*, version 3.0; Schrödinger, LLC., 2024.
49. Chen, J.; Schafer, N. P.; Wolynes, P. G.; Clementi, C.; *J. Phys. Chem. B* **2019**, *123*, 4497. [Crossref]
50. Rausch, A. O.; Freiberger, M. I.; Leonetti, C. O.; Luna, D. M.; Radusky, L. G.; Wolynes, P. G.; Ferreiro, D. U.; Parra, R. G.; *Bioinformatics* **2021**, *37*, 3038. [Crossref]
51. Tsai, M.-Y.; Zheng, W.; Balamurugan, D.; Schafer, N. P.; Kim, B. L.; Cheung, M. S.; Wolynes, P. G.; *Prot. Sci.* **2016**, *25*, 255. [Crossref]
52. Bitencourt-Ferreira, G.; Veit-Acosta, M.; de Azevedo Jr., W. F. In *Docking Screens for Drug Discovery*; de Azevedo Jr., W. F., ed.; Humana: New York, 2019, p. 67. [Crossref]
53. Debye, P.; Hückel, E.; *Phys. Z.* **1923**, *24*, 185. [Link] accessed in August 2024
54. Zheng, W.; Thorne, N.; McKew, J. C.; *Drug Discovery Today* **2013**, *18*, 1067. [Crossref]
55. Vasaikar, S.; Bhatia, P.; Bhatia, P. G.; Yaiw, K. C.; *Biomedicines* **2016**, *4*, 27. [Crossref]
56. Hopkins, A. L.; Groom, C. R.; *Nat. Rev. Drug Discovery* **2002**, *1*, 727. [Crossref]
57. Hajduk, P. J.; Huth, J. R.; Fesik, S. W.; *J. Med. Chem.* **2005**, *48*, 2518. [Crossref]
58. Binkowski ,T. A.; Naghibzadeh, S.; Liang, J.; *Nucleic Acids Res.* **2003**, *31*, 3352. [Crossref]
59. Tian, W.; Chen, C.; Lei, X.; Zhao, J.; Liang, J.; *Nucleic Acids Res.* **2018**, *46*, W363. [Crossref]
60. An, J.; Totrov, M.; Abagyan, R.; *Mol. Cell. Proteomics* **2005**, *4*, 752. [Crossref]
61. Laurie, A. T. R.; Jackson, R. M.; *Bioinformatics* **2005**, *21*, 1908. [Crossref]
62. Capra, J. A.; Laskowski, R. A.; Thornton, J. M.; Singh, M.; Funkhouser, T. A.; *PLoS Comput. Biol.* **2009**, *5*, e1000585. [Crossref]
63. Halgren, T.; *Chem. Biol. Drug Des.* **2007**, *69*, 146. [Crossref]
64. Myers, N.; Mittermeler, R. A.; Mittermeler, C. G.; da Fonseca, G. A. B.; Kent, J.; *Nature* **2000**, *403*, 853. [Crossref]
65. Grob, N. M.; Remarcik, C.; Rössler, S. L.; Wong, J. Y. K.; Wang, J. C. K.; Tao, J.; Smith, C. L.; Loas, A.; Buchwald, S. L.; Eaton, D. L.; López, M. P.; Pentelute, B. L.; *ACS Chem. Biol.* **2024**, *19*, 101. [Crossref]
66. Madsen, A. V.; Mejias-Gomez, O.; Pedersen, L. E.; Morth, J. P.; Kristensen, P.; Jenkins, T. P.; Goletz, S.; *Comput. Struct. Biotechnol. J.* **2024**, *23*, 199. [Crossref]
67. Morrow, J. K.; Zhang, S.; *Curr. Pharm. Des.* **2012**, *18*, 1255. [Crossref]
68. Hall, D. R.; Ngan, C. H.; Zerbe, B. S.; Kozakov, D.; Vajda, S.; *J. Chem. Inf. Model.* **2012**, *52*, 199. [Crossref]
69. Mattos, C.; Ringe, D.; *Nat. Biotechnol.* **1996**, *14*, 595. [Crossref]
70. Ngan, C. H.; Bohnuud, T.; Mottarella, S. E.; Beglov, D.; Villar, E. A.; Hall, D. R.; Kozakov, D.; Vajda, S.; *Nucleic Acids Res.* **2012**, *40*, W271. [Crossref]

71. Mattos, C.; Bellamacina, C. R.; Peisach, E.; Pereira, A.; Vitkup, D.; Petsko, G. A.; Ringe, D.; *J. Mol. Biol.* **2006**, *357*, 1471. [Crossref]
72. Krüger, D. M.; Gohlke, H.; *Nucleic Acids Res.* **2010**, *38*, W480. [Crossref]
73. Ofran, Y.; Rost, B.; *PLoS Comput. Biol.* **2007**, *3*, e119. [Crossref]
74. Wu, F. X.; Yang, J.-F.; Mei, L.-C.; Wang, F.; Hao, G.-F.; Yang, G.-F.; *J. Chem. Inf. Model.* **2021**, *61*, 14. [Crossref]
75. Shingate, P.; Manoharan, M.; Sukhwal, A.; Sowdhamini, R.; *BMC Bioinf.* **2014**, *15*, 303. [Crossref]
76. Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A.; *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 808. [Crossref]
77. Ghanakota, P.; Carlson, H. A.; *J. Med. Chem.* **2016**, *59*, 10383. [Crossref]
78. Brenke, R.; Kozakov, D.; Chuang, G.-Y.; Beglov, D.; Hall, D.; Landon, M. R.; Mattos, C.; Vajda, S.; *Bioinformatics* **2009**, *25*, 621. [Crossref]
79. Hall, D. R.; Kozakov, D.; Whitty, A.; Vajda, S.; *Trends Pharmacol. Sci.* **2015**, *36*, 724. [Crossref]
80. Kozakov, D.; Hall, D. R.; Napoleon, R. L.; Yueh, C.; Whitty, A.; Vajda, S.; *J. Med. Chem.* **2015**, *58*, 9063. [Crossref]
81. Teixeira, O.; Lacerda, P.; Froes, T. Q.; Nonato, M. C.; Castilho, M. S.; *J. Comput.-Aided Mol. Des.* **2021**, *35*, 871. [Crossref]
82. Froes, T. Q.; Zapata, L. C. C.; Akamine, J. S.; Castilho, M. S.; Nonato, M. C.; *Curr. Top. Med. Chem.* **2021**, *21*, 2134. [Crossref]
83. Froes, T. Q.; Barbosa, D. B.; do Bomfim, M. R.; Leite, F. H. A.; Castilho, M. S. In *Computational Modeling of Drugs Against Alzheimer's Disease*; Roy, K., ed.; Humana: New York, 2023, p. 99. [Crossref]
84. Grove, L. E.; Hall, D. R.; Beglov, D.; Vajda, S.; Kozakov, D.; *Bioinformatics* **2013**, *29*, 1218 [Crossref]; FTFlex - Flexible Protein Mapping, https://ftflex.bu.edu/, accessed in August 2024.
85. Egbert, M.; Jones, G.; Collins, M. R.; Kozakov, D.; Vajda, S.; *J. Mol. Biol.* **2022**, *434*, 167587. [Crossref]
86. Hall, D. R.; Kozakov, D.; Vajda, S. In *Computational Drug Discovery and Design*; Baron, R., ed.; Humana: Totowa, 2012, p. 13. [Crossref]
87. Kozakova, D.; Hall, D. R.; Jehle, S.; Luo, L.; Ochiana, S. O.; Jones, E. V.; Pollastri, M.; Allen, K. N.; Whitty, A.; Vajda, S.; *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, E2585. [Crossref]
88. The Cambridge Crystallographic Data Centre (CCDC), https://www.ccdc.cam.ac.uk/, accessed in August 2024.
89. Hendlich, M.; Rippmann, F.; Barnickel, G.; *J. Mol. Graph. Model.* **1997**, *15*, 359. [Crossref]
90. Ichihara, O.; Shimada, Y.; Yoshidome, D.; *ChemMedChem* **2014**, *9*, 2708. [Crossref]
91. Jäger, M.; Zhang, Y.; Bieschke, J.; Nguyen, H.; Dendle, M.; Bowman, M. E.; Noel, J. P.; Gruebele, M.; Kelly, J. W.; *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 10648. [Crossref]
92. Di Silvio, E.; Brunori, M.; Gianni, S.; *Angew. Chem., Int. Ed.* **2015**, *54*, 10867. [Crossref]
93. Fanning, A. S.; Anderson, J. M.; *Curr. Biol.* **1996**, *6*, 1385. [Crossref]
94. De Vries, L.; Lou, X.; Zhao, G.; Zheng, B.; Farquhar, M. G.; *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 12340. [Crossref]
95. Calosci, N.; Chi, C. N.; Richter, B.; Camilloni, C.; Engström, Å.; Eklund, L.; Travaglini-Allocatelli, C.; Gianni, S.; Vendruscolo, M.; Jemth, P.; *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 19241. [Crossref]
96. Walma, T.; Spronk, C. A. E. M.; Tessari, M.; Aelen, J.; Schepens, J.; Hendriks, W.; Vuister, G. W.; *J. Mol. Biol.* **2002**, *316*, 1101. [Crossref]
97. Walma, T.; Aelen, J.; Nabuurs, S. B.; Oostendorp, M.; van den Berk, L.; Hendriks, W.; Vuister, G. W.; *Structure* **2004**, *12*, 11. [Crossref]
98. Gianni, S.; Walma, T.; Arcovito, A.; Calosci, N.; Bellelli, A.; Engström, Å.; Travaglini-Allocatelli, C.; Brunori, M.; Jemth, P.; Vuister, G. W.; *Structure* **2006**, *14*, 1801. [Crossref]
99. Welcome to FTMove, https://ftmove.bu.edu/, accessed in August 2024.
100. Kozlov, G.; Banville, D.; Gehring, K.; Ekiel, I.; *J. Mol. Biol.* **2002**, *320*, 813. [Crossref]
101. Buchli, B.; Waldauer, S. A.; Walser, R.; Donten, M. L.; Pfister, R.; Blöchliger, N.; Steiner, S.; Caflisch, A.; Zerbe, O.; Hamm, P.; *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 11725. [Crossref]
102. Zhang, J.; Sapienza, P. J.; Ke, H.; Chang, A.; Hengel, S. R.; Wang, H.; Phillips Jr., G. N.; Lee, A. L.; *Biochemistry* **2010**, *49*, 9280. [Crossref]
103. Kozlov, G.; Gehring, K.; Ekiel, I.; *Biochemistry* **2000**, *39*, 2572. [Crossref]
104. Ashkinadze, D.; Kadavath, H.; Pokharna, A.; Chi, C. N.; Friedmann, M.; Strotz, D.; Kumari, P.; Minges, M.; Cadalbert, R.; Königl, S.; Güntert, P.; Vögeli, B.; Riek, R.; *Nat. Commun.* **2022**, *13*, 6232. [Crossref]
105. Freiberger, M. I.; Guzovsky, A. B.; Wolynes, P. G.; Parra, R. G.; Ferreiro, D. U.; *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116*, 4037. [Crossref]
106. Furnham, N.; Holliday, G. L.; de Beer, T. A. P.; Jacobsen, J. O. B.; Pearson, W. R.; Thornton, J. M.; *Nucleic Acids Res.* **2014**, *42*, D485. [Crossref]
107. Li, W.; Wolynes, P. G.; Takada, S.; *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 3504. [Crossref]
108. Gianni, S.; Jemth, P.; *IUBMB Life* **2014**, *66*, 449. [Crossref]
109. Orengo, C. A.; Michie, A. D.; Jones, S.; Jones, D. T.; Swindells, M. B.; Thornton, J. M.; *Structure* **1997**, *5*, 1093. [Crossref]
110. Lawson, C. L.; van Montfort, R.; Strokopytov, B.; Rozeboom, H. J.; Kalk, K. H.; de Vries, G. E.; Penninga, D.; Dijkhuizen, L.; Dijkstra, B. W.; *J. Mol. Biol.* **1994**, *236*, 590. [Crossref]
111. Dauter, Z.; Dauter, M.; Brzozowski, A. M.; Christensen, S.; Borchert, T. V.; Beier, L.; Wilson, K. S.; Davies, G. J.;

Biochemistry **1999**, *38*, 8385. [Crossref]
112. Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB), *2A0N*, https://www.rcsb.org/structure/2A0N, accessed in August 2024.
113. Chiu, H.-J.; Reddick, J. J.; Begley, T. P.; Ealick, S. E.; *Biochemistry* **1999**, *38*, 6460. [Crossref]
114. Miskei, M.; Antal, C.; Fuxreiter, M.; *Nucleic Acids Res.* **2017**, *45*, D228. [Crossref]
115. Delaforge, E.; Kragelj, J.; Tengo, L.; Palencia, A.; Milles, S.; Bouvignies, G.; Salvi, N.; Blackledge, M.; Jensen, M. R.; *J. Am. Chem. Soc.* **2018**, *140*, 1148. [Crossref]
116. Gianni, S.; Freiberger, M. I.; Jemth, P.; Ferreiro, D. U.; Wolynes, P. G.; Fuxreiter, M.; *Acc. Chem. Res.* **2021**, *54*, 1251. [Crossref]
117. Stolboushkina, E.; Nikonov, S.; Nikulin, A.; Bläsi, U.; Manstein, D. J.; Fedorov, R.; Garber, M.; Nikonov, O.; *J. Mol. Biol.* **2008**, *382*, 680. [Crossref]
118. Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB), *3I1F*, https://www.rcsb.org/structure/3I1F, accessed in August 2024.
119. Zhang, T.; Inesta-Vaquera, F.; Niepel, M.; Zhang, J.; Ficarro, S. B.; Machleidt, T.; Xie, T.; Marto, J. A.; Kim, N.; Sim, T.; Laughlin, J. D.; Park, H.; LoGrasso, P. V.; Patricelli, M.; Nomanbhoy, T. K.; Sorger, P. K.; Alessi, D. R.; Gray, N. S.; *Chem. Biol.* **2012**, *19*, 140. [Crossref]
120. Laughlin, J. D.; Nwachukwu, J. C.; Figuera-Losada, M.; Cherry, L.; Nettles, K. W.; LoGrasso, P. V.; *Structure* **2012**, *20*, 2174. [Crossref]
121. Fuxreiter, M.; *J. Mol. Biol.* **2018**, *430*, 2278. [Crossref]
122. Hatos, A.; Monzon, A. M.; Tosatto, S. C. E.; Piovesan, D.; Fuxreiter, M.; *Nucleic Acids Res.* **2022**, *50*, D509. [Crossref]
123. Tompa, P.; Fuxreiter, M.; *Trends Biochem. Sci.* **2008**, *33*, 2. [Crossref]
124. Fuxreiter, M.; *Biochem. Soc. Trans.* **2020**, *48*, 2557. [Crossref]
125. Chuang, G.-Y.; Mehra-Chaudhary, R.; Ngan, C.-H.; Zerbe, B. S.; Kozakov, D.; Vajda, S.; Beamer, L. J.; *Prot. Sci.* **2010**, *19*, 1662. [Crossref]
126. Schreiber, G.; Fersht, A. R.; *Nat. Struct. Biol.* **1996**, *3*, 427. [Crossref]
127. Wang, Q.; Zhang, P.; Hoffman, L.; Tripathi, S.; Homouz, D.; Liu, Y.; Waxham, M. N.; Cheung, M. S.; *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 20545. [Crossref]
128. Sielecki, A. R.; Hayakawa, K.; Fujinaga, M.; Murphy, M. E. P.; Fraser, M.; Muir, A. K.; Carilli, C. T.; Lewicki, J. A.; Baxter, J. D.; James, M. N. G.; *Science* **1989**, *243*, 1346. [Crossref]
129. Rahuel, J.; Priestle, J. P.; Grütter, M. G.; *J. Struct. Biol.* **1991**, *107*, 227. [Crossref]
130. Dock-Bregeon, A. C.; Chevrier, B.; Podjarny, A.; Johnson, J.; de Bear, J. S.; Gough, G. R.; Gilham, P. T.; Moras, D.; *J. Mol. Biol.* **1989**, *209*, 459. [Crossref]
131. Rahuel, J.; Rasetti, V.; Maibaum, J.; Rüeger, H.; Göschke, R.; Cohen, N.-C.; Stutz, S.; Cumin, F.; Fuhrer, W.; Wood, J.; Grütter, M. G.; *Cell Chem. Biol.* **2000**, *7*, 493. [Crossref]
132. Jensen, C.; Herold, P.; Brunner, H. R.; *Nat. Rev. Drug Discovery* **2008**, *7*, 399. [Crossref]
133. Takashima, E.; Inaoka, D. K.; Osanai, A.; Nara, T.; Odaka, M.; Aoki, T.; Inaka, K.; Harada, S.; Kita, K.; *Mol. Biochem. Parasitol.* **2002**, *122*, 189. [Crossref]
134. Malagrinò, F.; Diop, A.; Pagano, L.; Nardella, C.; Toto, A.; Gianni, S.; *Curr. Opin. Struct. Biol.* **2022**, *72*, 153. [Crossref]
135. Agarwal, S.; Smith, M.; De La Rosa, I.; Verba, K. A.; Swartz, P.; Segura-Totten, M.; Mattos, C.; *Acta Crystallogr., Sect. D: Struct. Biol.* **2020**, *76*, 1001. [Crossref]
136. Krasowski, A.; Muthas, D.; Sarkar, A.; Schmitt, S.; Brenk, R.; *J. Chem. Inf. Model.* **2011**, *51*, 2829. [Crossref]

Submitted: February 17, 2024
Published online: August 20, 2024