THE PROTEIN SOCIETY **WILEY**

Cold spots are universal in protein–protein interactions

Sagara N.S. Gurusinghe | Ben Oppenheimer | Julia M. Shifman

Department of Biological Chemistry, The Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel

Correspondence
Julia M. Shifman, Department of Biological Chemistry, The Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel.
Email: jshifman@mail.huji.ac.il

Abstract

Proteins interact with each other through binding interfaces that differ greatly in size and physico-chemical properties. Within the binding interface, a few residues called hot spots contribute the majority of the binding free energy and are hence irreplaceable. In contrast, cold spots are occupied by suboptimal amino acids, providing possibility for affinity enhancement through mutations. In this study, we identify cold spots due to cavities and unfavorable charge interactions in multiple protein–protein interactions (PPIs). For our cold spot analysis, we first use a small affinity database of PPIs with known structures and affinities and then expand our search to nearly 4000 homo- and heterodimers in the Protein Data Bank (PDB). We observe that cold spots due to cavities are present in nearly all PPIs unrelated to their binding affinity, while unfavorable charge interactions are relatively rare. We also find that most cold spots are located in the periphery of the binding interface, with high-affinity complexes showing fewer centrally located colds spots than low-affinity complexes. A larger number of cold spots is also found in non-cognate interactions compared to their cognate counterparts. Furthermore, our analysis reveals that cold spots are more frequent in homo-dimeric complexes compared to heterocomplexes, likely due to symmetry constraints imposed on sequences of homodimers. Finally, we find that glycines, glutamates, and arginines are the most frequent amino acids appearing at cold spot positions. Our analysis emphasizes the importance of cold spot positions to protein evolution and facilitates protein engineering studies directed at enhancing binding affinity and specificity in a wide range of applications.

KEYWORDS

binding affinity, binding interfaces, cold spots of binding, hot spots of binding, protein–protein interactions

1 | INTRODUCTION

Protein–protein interactions (PPIs) play a key role in diverse biological functions including muscle contraction, signal transduction, cell metabolism, macromolecular

assembly, and other cellular processes.[1] Each PPI is characterized with a particular binding affinity (K_D) that is compatible with PPI functional role in the cell. As such, PPIs that are responsible for cellular life/death decisions, such as for example, toxin/antitoxin interactions, bind to each other with extremely high affinities.[2,3] PPIs that are involved in signaling usually possess medium binding affinities, allowing them to interact with their partners on the intermediate time scale. Transient and multi-specific interactions are characterized with weak binding affinities as quick dissociation of one protein from another is required for the correct functioning of such complexes.[4]

Proteins interact with each other through surface patches, termed binding interfaces, that are distinguishable from non-interface protein surface.[5,6] PPI binding interfaces differ greatly in size, geometry, and physico-chemical properties.[7–12] Molecular interactions across the binding interface, such as hydrophobic burial, hydrogen bonding, van der Waals, and electrostatic interactions largely determine PPI binding affinity. Multiple studies attempted to predict binding affinity from various structural features[13–21] and revealed that each single structural feature contributes weakly to PPI binding affinity. Only taking many features together, one can explain the more than 10 orders of magnitude span in K_D values observed in nature.[22]

Multiple works showed that various positions at the binding interface contribute differently to the binding free energy.[23–28] A small number of positions referred to as hot spots could contribute as much as three quarters of the binding free energy.[24] Hot spot positions are usually conserved among species,[29] located at the center of the binding interface and are most frequently occupied by large amino acids such as tryptophan, tyrosine, and arginine.[25] Furthermore, hot spots are often clustered forming hot regions,[30–32] within which mutations are coupled to each other.[33]

While hot spots have been extensively studied, little attention has been given to alternative, cold spot positions, that is, positions that are occupied by suboptimal amino acids and present imperfections in the binding interface design.[34] Yet, cold spot positions are important in protein evolution as they may support function of low-affinity and transient PPIs. Furthermore, they might play a crucial role in determining binding specificity, providing a way to discriminate against undesired binding partners. In addition, studies of cold spot positions could assist in designing experiments that aim to optimize protein-based therapeutics.

In our previous work, we introduced a concept of cold spots and examined a few PPI structures that contained cold spot positions.[35–38] From these examples we

suggested that cold spots in PPIs could occur via two distinct scenarios.[34] In the first scenario, a wild type amino acid at a cold spot position does not directly interact with the partner protein, resulting in a cavity. Upon mutation to a larger amino acid, new favorable intermolecular interactions are created, increasing PPI affinity. In the second scenario, an unfavorable interaction is present at the cold spot position in the wild-type PPI; removal of such an unfavorable interaction through mutation results in affinity enhancement, sometimes by several orders of magnitude. We further postulated that cold spot frequency and location might correlate with PPI binding affinity but have not substantiated our hypothesis with a high-throughput cold spot analysis.[34] Thus, cold spot frequency among various PPIs, their spatial distribution and structural context remained largely unknown.

To close this knowledge gap, we developed new software that recognizes cold spot positions from PPI structure without expensive $\Delta\Delta G_{bind}$ calculations and thus can be used for high-throughput PPI analysis. Using this software, we identify cold spots in thousands of PPIs, first using a small database of PPIs with known structures and affinities and then in a large database of hetero- and homo-dimeric complexes with known structures. We find that cold spots are common in PPIs with all binding affinities, more frequently located at the periphery of the binding interface and do not form clusters. Additionally, we find that non-cognate complexes contain a higher number of cold spots compared to their cognate counterparts. We also observe that cold spots are occupied most frequently by glycines, glutamates, and arginines, while only arginines are frequent at hot spot positions. We further explain how identification of cold spots could greatly assist in various strategies for drug design.

2 | RESULTS

To investigate cold spot frequency and distribution among various PPIs, we searched for imperfections in packing and in charge distribution in multiple binding interfaces. To identify imperfections in packing, we developed an algorithm that searches for cavities by placing random dots in the binding interface and identifying those dots that are not within van der Waals radius from any protein atom (Figure 1a). The dots that fell into empty spaces were clustered into separate cavities. Clusters of at least 10 dots were considered as cold spots due to imperfections in packing. Such cavities could be accessible or inaccessible to solvent yet all of them belonged to the PPI binding interface. In addition to identifying cavities, we developed an algorithm that searches for unfavorable interactions across the binding interface that
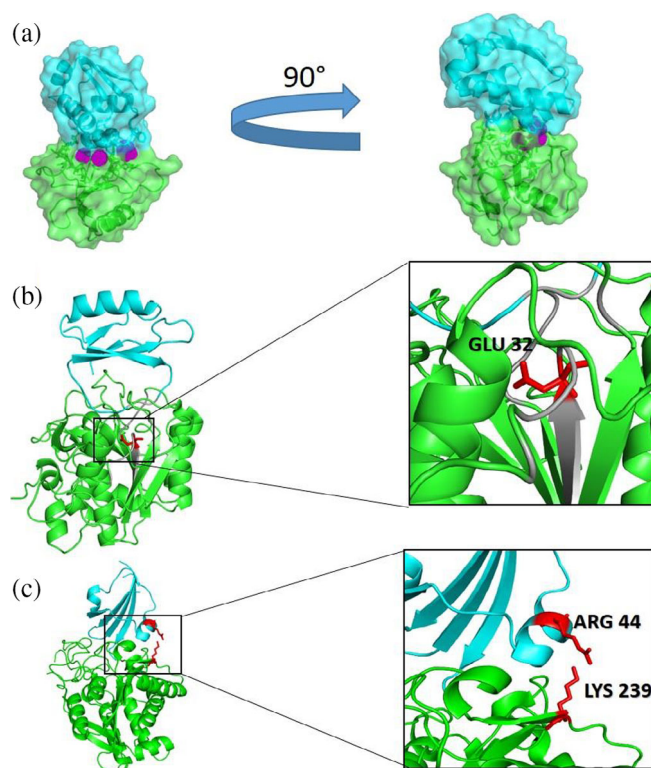
FIGURE 1 Three types of cold spots in protein complexes. (a) Cold spots due to cavities. A structure of a complex between Ribonuclease (green) and Barstar (cyan), showing cold spot cavities in magenta (PDB ID 1AY7). (b) Cold spots due to CH interactions. The left figure shows Subtilisin (green) bound to Chymotrypsin Inhibitor-2 (cyan) with a cold spot residue shown in red. Hydrophobic residues around the cold spot are shown in gray. The right figure zooms into the cold spot (PDB ID 2SNI). (c) Cold spots due to SC interactions. The left figure shows Carboxypeptidase A1 (green) bound to Metallocarboxypeptidase inhibitor (cyan) with cold spots shown in red (PDB ID 2ABZ). The right figure zooms into the cold spots, showing two positively charged residues in close contact with each other

belong to two common types: (1) charge-hydrophobic (CH) interactions where a charged residue is buried in the hydrophobic environment and (2) same charge (SC) interactions where two oppositely charged residues are coming in close proximity to each other. In CH interaction, substitution of the charged residue with a hydrophobic one usually results in binding affinity enhancement, giving rise to a single cold spot (Figure 1b). In SC interaction, either of the charged residues could be replaced to improve affinity, thus giving rise to two cold spots (Figure 1c).

We first tested our program for cold spot identification on a small affinity database of 133 PPIs with available structures in both bound and unbound states and measured binding affinities ranging from 10^{-5} to 10^{-14} M.[39] Our analysis of the affinity database showed
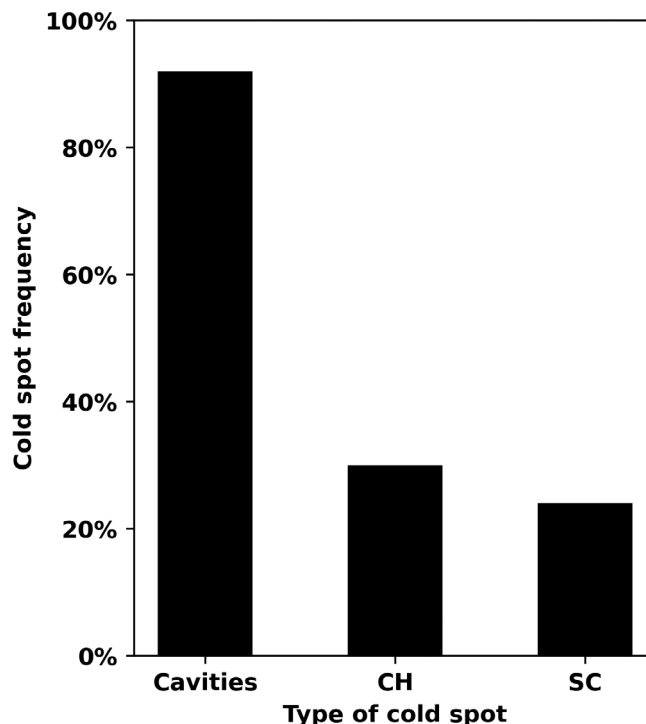

FIGURE 2 Frequency of cold spot occurrence. Frequency of the three types of cold spots in the affinity database. Complexes that contain at least one cold spot due to cavity, CH, and SC interactions were considered for the count

that out of 133 complexes, 123 (92%) contained cold spots due to imperfections in packing, demonstrating that cavities are very frequent in PPIs. In comparison, only 40 complexes (30%) contained cold spots due to buried CH interactions, and only 33 complexes (25%) due to SC interactions, indicating that unfavorable charge interactions are notable but less frequent than cavities in PPIs (Figure 2). The number of cold spots due to CH and SC interactions did not exceed 5 per complex. In comparison, a maximum of eight cold spots per complex due to cavities were observed (see Supporting Information).

Using the same database, we next divided the database into two groups of similar size, high-affinity complexes ($K_D < 10^{-8}$ M) and low-affinity complexes ($K_D > 10^{-8}$ M) and examined whether high-affinity complexes contained fewer cold spots compared to low-affinity complexes. Interestingly, no correlation was observed between complex K_D and the number of cold spots due to cavities, CH interactions, SC interactions, and due to all the three scenarios (Figure S1). However, we observed that the total number of cold spots was correlated with the difference in the accessible surface area upon binding, ΔASA (Figure S2A), suggesting that larger interfaces are likely to have more non-optimized positions. Further analysis showed that among different types of colds spots, only cold spots due to cavities show strong
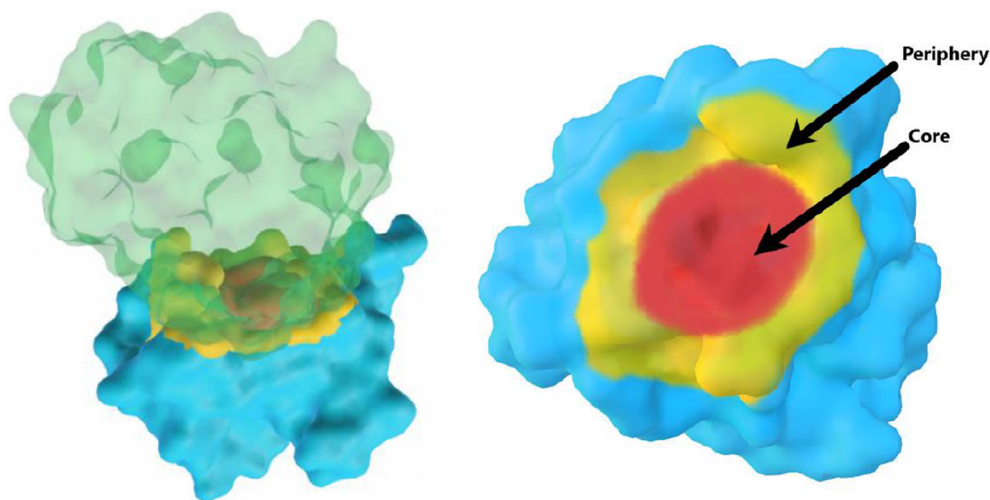
FIGURE 3 Binding interface of a PPI divided into core and periphery Left: Side view of the complex. Right: top view of one of the proteins in the complex, showing the periphery (yellow) and the core (red) of the binding interface. For each atom in the binding interface, a Periphery Index (PI) was calculated by dividing atom-to-surface distance by the maximum atom-to-surface distance for the binding interface (see Section 6 for details). Atoms with PI higher than 0.68 were assigned to the core of the binding interface and atoms with PI less than 0.68 were assigned to the periphery of the interface

correlation with ΔASA (Figure S2b–d). To investigate whether PPIs with certain functions contain a higher number of cold spots, we made use of the functional categorization in the affinity database[39] (see Section 6 for details) and analyzed cold spots in antigen–antibody complexes, G protein complexes, enzyme-containing complexes, and other receptor-containing complexes. Our data show no significant difference in cold spot number among these functional classes of PPIs, except for slightly higher number of cold spots due to CH interactions in enzyme-containing complexes (Figure S3). Thus, all functional classes could contain PPIs with different level of evolutionary optimality. Yet, enzyme-containing complexes might exhibit less optimized interfaces more frequently, as burial of a charged residue might be a prerequisite for catalysis.

While we did not observe direct correlation between interaction K_D and the number of cold spots, we postulated that location of a cold spot might be important in determining free energy of binding, with more central cold spots being more deleterious to binding compared to peripheral cold spots. To check this hypothesis, we divided binding interfaces in two areas, interface core and interface periphery according to their distance from the center of the binding interface and each area containing approximately the same number of atoms (Figure 3). Subsequently, each cold spot was assigned to either interface core or interface periphery (see Section 6 for details). Our analysis shows that 62% of the cold spots were located at the periphery of the interface. Similar results were observed for different types of cold spots where
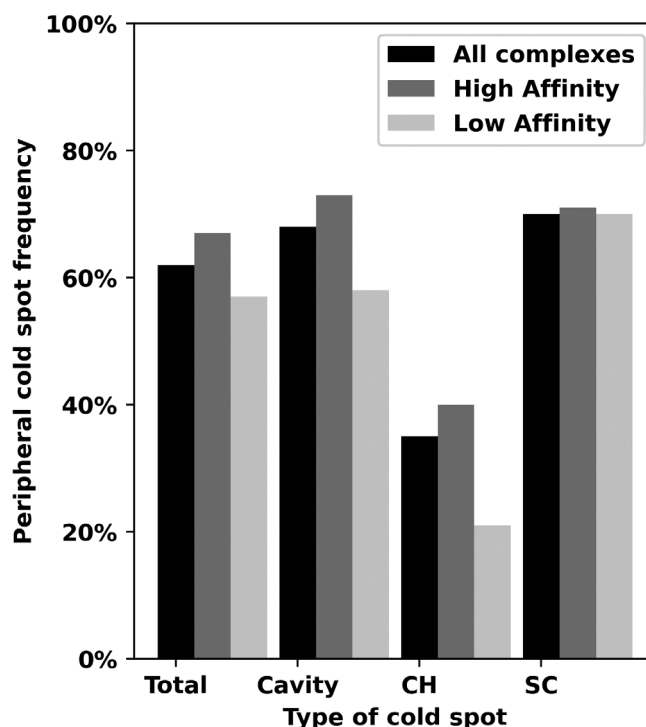
FIGURE 4 Frequency of peripheral cold spots in the affinity database. Frequency of peripheral cold spots due to cavities, CH and SC interactions in all complexes (black), high-affinity complexes (dark gray) and low-affinity complexes (light gray). Cold spots were assigned as peripheral if their PI was <0.68

cavities and SC interactions occurred in the interface periphery in 68% and 70% cases, respectively (Figure 4). In contrast, only 37% of cold spots due to CH interactions

were located in the periphery of the binding interface. We next compared cold spot occurrence in the core and in the periphery for high- and low-affinity PPIs. We found that 67% of the cold spots were located in the periphery for high-affinity complexes while only 57% for low affinity complexes (p value 3.93e−05). Decrease in frequency of peripheral cold spots with decreased affinity was also observed for cavities and CH interactions when analyzing them separately. Here, 73% and 40% of cold spots were located in the periphery for high-affinity complexes while only 58% and 21% for low-affinity complexes for cavities and CH interactions, respectively (p-values of .001 and .013, respectively). In contrast, no difference was found between high- and low-affinity complexes for SC interactions that occurred predominantly at the periphery of the interface in both groups.

In order to check whether cold spots cluster together, we calculated pairwise distances between two cold spots in the same complex (see Section 6 for details). We observed that only 5% of cold spots were found within 6 Å from each other, while 65% of cold spots were located more than 12 Å apart (Figure 5). Similar results were observed when analyzing separately pairwise distances for the three different classes of cold spots (Figure S4).
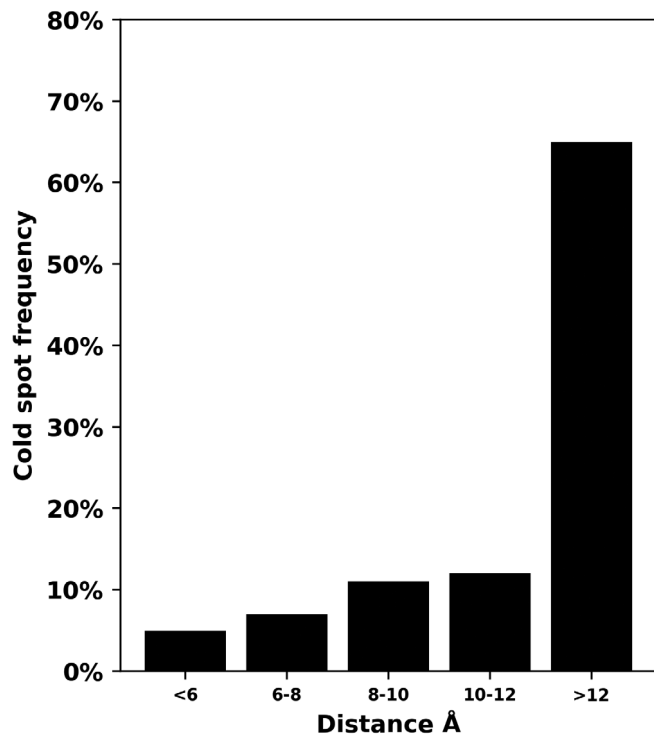

FIGURE 5 Histogram of pairwise distances between cold spots. Histogram includes cold spots due to all three scenarios. Cα–Cα distances were analyzed for cold spots due to CH and SC interactions. Center–center distances were analyzed for cold spots due to cavities

Thus, the majority of cold spots do not cluster together but spread randomly throughout the protein interface.

In summary, the above results demonstrate that cold spots are rather frequent in PPIs independent of their binding affinity, they do not form clusters and show different spatial distribution in high- and low-affinity complexes, with higher affinity complexes exhibiting fewer centrally located colds spots compared to low-affinity complexes.

3 | COLD SPOTS IN COGNATE AND NON-COGNATE INTERACTIONS

While our results showed no correlation between the number of cold spots and PPI affinity, we wondered whether this result was due to a relatively small number of cold spots per PPI. We thought that the number of cold spots could be used to distinguish between structurally similar cognate complexes that have evolved for high-affinity interactions and non-cognate complexes that bind each other only due to structural similarity. Since no database of cognate and non-cognate complexes exists, we analyzed a few exemplary cases of such PPIs.[40–47] First, we looked at interactions between basic pancreatic trypsin inhibitor (BPTI) interacting with serine proteases such as trypsin, chymotrypsin, and mesotrypsin (PDB ID 2PTC, 1CBW, and 2R9P, respectively). No cold spots were discovered in the cognate BPTI/trypsin complex that has been highly optimized for binding (Figure 6a). In contrast, the non-cognate complex of Chymotrypsin/BPTI contained two cold spots, one due to cavity near position 15 and the other one in the periphery of the interface (Figure 6b). The non-cognate Mesotrypsin/BPTI complex contained four cold spots due to SC interactions (Figure 6c).

Second, we looked at pairs of colicin/immunity complexes that exhibit ultra-high affinity for cognate interactions and several orders of magnitude lower affinity for non-cognate interactions. We analyzed the structure of colicin E9 DNase domain with its cognate immunity protein IM9 (PDB ID 1EMV) and the structure of the corresponding non-cognate PPI between E9 DNase and IM2 (PDB ID 2WPT). In the cognate complex, we did not observe any cavities or CH/ SC interactions, which can act as cold spots. But in the non-cognate complex, we observed two distinct cavities and a CH interaction that could act as cold spots (Figure 7). Examination of 15 pairs of cognate and non-cognate complexes containing various PPIs revealed that 12 non-cognate complexes contained a higher number of cold spots compared to the corresponding cognate complexes, while the remaining three contained the same number
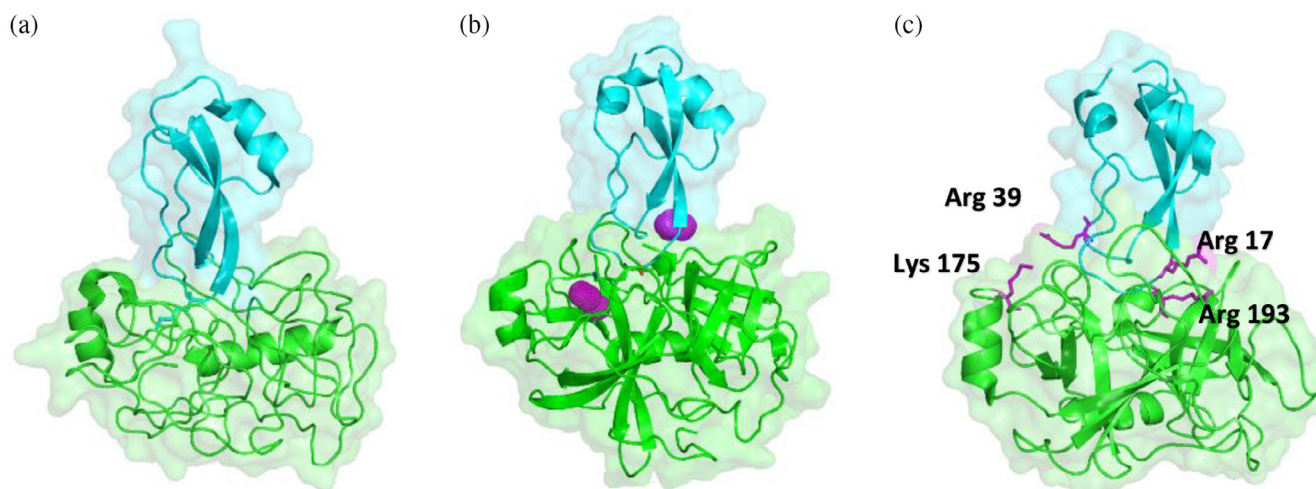
FIGURE 6 Comparison of cold spots in cognate and non-cognate complexes of BPTI-serine protease complex structures. (a) The structure of cognate complex between BPTI (cyan) and bovine trypsin (green) (PDB ID 2PTC) showing no cold spots. (b) The structure of the non-cognate complex between BPTI (cyan) and bovine alpha-chymotrypsin (green) (PDB ID 1CBW). Two cold spots due to cavities are shown in magenta. (c) The structure of the non-cognate complex between BPTI (cyan) and human mesotrypsin (green) (PDB ID 2R9P). Four cold spots due to SC interactions are shown in magenta
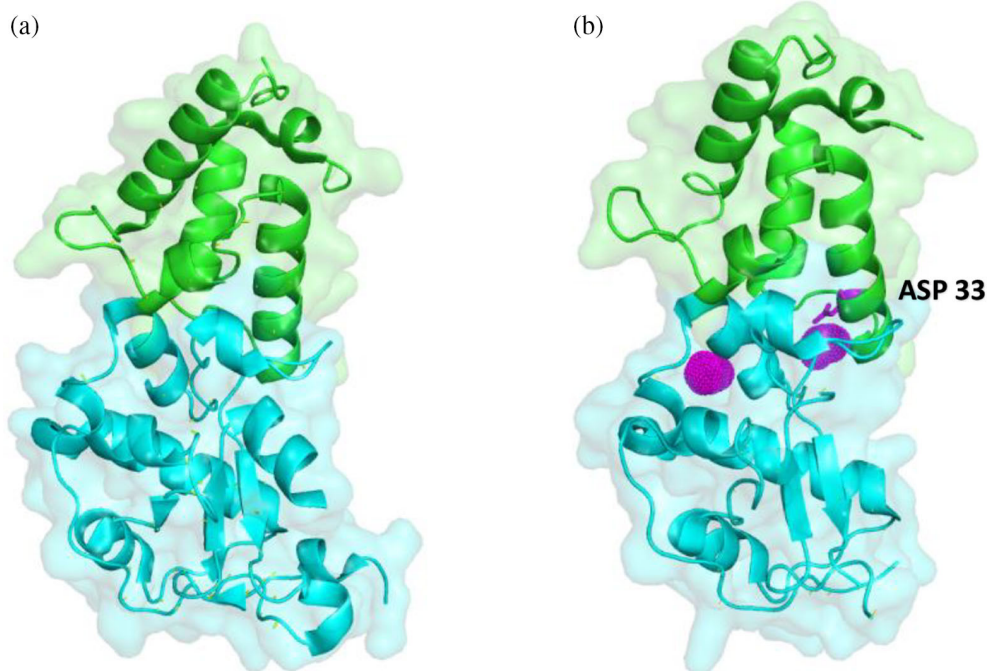

FIGURE 7 Comparison of cold spots in cognate and non-cognate complexes of Colicin DNase/Immunity proteins. The left figure shows the cognate complex of colicin DNase E9 (cyan) with Immunity protein 9 (green) (PDB ID 1EMV) with no cold spots observed. The right figure shows the non-cognate complex of colicin DNase E9 (cyan) with immunity protein 2 (green) (PDB ID 2WPT). Two cold spots due to cavities and one cold spot due to CH interaction is shown in magenta

of cold spots (Table 1, *p*-value of 0.0027). This result agrees with our general understanding that binding interfaces of cognate complexes exhibit higher evolutionary optimality relative to non-cognate interfaces. Note that the appearance of even one additional cold spot could decrease binding affinity of this PPI by several orders of magnitude.[34]
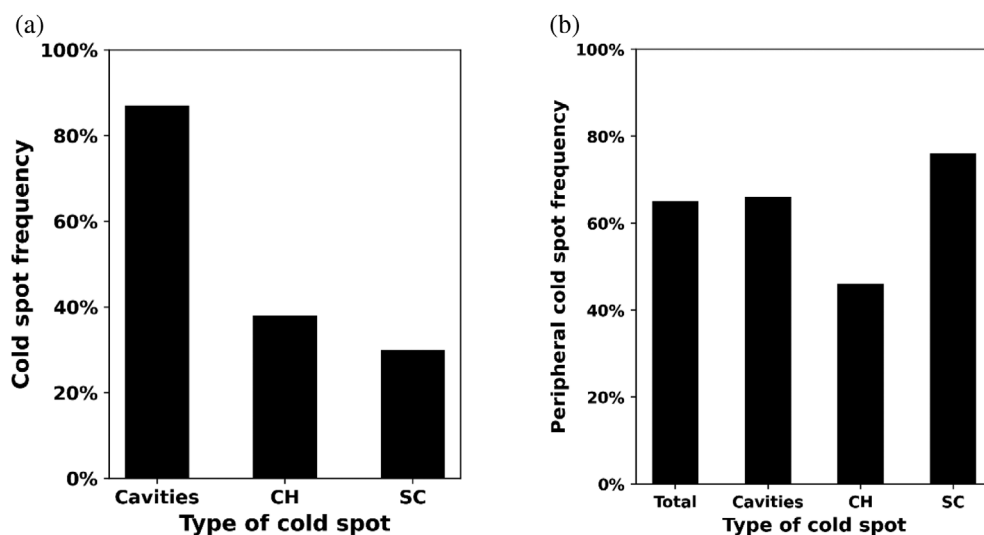
4 | COLD-SPOT IDENTIFICATION ON THE PDB SCALE

To further extend our cold spot analysis, we identified cold spots in all available nonredundant protein–protein complex structures in the PDB. To avoid inaccuracies in cold spot identification, we limited our dataset to

complexes with structures solved to high resolution, without hetero atoms in the binding interface, and containing binding partners that are larger than 40 residues. Our structural dataset thus included a total of 3,826 protein complexes in which 1,507 structures belonged to heterodimeric complexes and 2,319 structures—to homodimeric complexes. Our analysis showed that *87%* of PPIs in our structural dataset contained at least one cavity that can act as cold spot (Figure 8a). In comparison, *38%* of PPIs contained cold spots due to CH interactions and *30%* due to SC interactions, similarly to the results

observed for the small affinity database. Among identified cold spots, 66%, 46%, and 76% lied in the periphery of the binding interface for cavities, CH, and SC interactions, respectively with 65% of the total number of cold spots located in the periphery of the interface (Figure 8b). Further analysis showed that on average, a PPI contained 4.6 ± 3.73 cold spots per complex with 8.3% of PPIs exhibiting no cold spots and some containing as many as 18 cold spots (Figure 9).

We next explored whether there is any difference between colds spot occurrence in hetero- and homo-complexes. Our analysis showed that *85%* of heterocomplexes and *88%* of homocomplexes contained at least one cavity which can act as cold spot (Figure 10). Cold spots due to SC interactions were present in *24%* of hetero-complexes and *34%* of homo-complexes. Finally, cold spots due to CH interactions were present in *35%* of heterocomplexes and *39%* of homocomplexes. Thus, in all three scenarios, homodimers contained a higher number of cold spots compared to heterocomplexes (*p*-values of .011, 7.63e−21, and 4.81e−17 for cold spots due to cavities, SC interactions and CH interactions, respectively).

Finally, we examined amino acid composition at cold spot positions. First, we calculated the frequency of each amino acid that participates in CH and SC interactions (Figure 11a,b). We observed that negatively charged amino acids are found more frequently in CH interactions compared to positively charged amino acids with glutamate occupying the largest fraction of cold spots. Similarly, in cold spots due to SC interactions, we also observed that negatively charged amino acids are more frequent compared to positively charge amino acids, but arginine appears with the highest frequency while lysine appears least. Second, we calculated frequency of amino acids, appearing next to a cavity that has been assigned as a cold spot. Since there are several residues

TABLE 1 Comparison of number of cold spots in cognate and non-cognate complexes

Cognate complexes				Non-cognate complexes			
PDB	CH	SC	Cavity	PDB	CH	SC	Cavity
2PTC	0	0	0	1CBW	0	0	2
1EMV	0	0	0	2WPT	1	0	2
3L3T	0	0	0	1CA0	0	0	2
2PTC	0	0	0	2TGP	0	0	0
2PTC	0	0	0	2R9P	0	4	0
3BZD	0	0	1	2AQ3	0	0	2
1BRS	2	0	2	1AY7	2	0	3
2VIR	0	0	1	2VIS	0	0	1
2JB0	1	0	3	3GKL	1	3	1
3L3T	0	0	0	5NX1	0	0	2
1EFN	0	0	1	1AVZ	0	1	0
3L3T	0	0	0	1BRC	0	0	4
1P2C	1	0	1	1MLC	3	0	0
2PTC	0	0	0	3U1J	0	0	2
3L3T	0	0	0	4U30	0	0	1

FIGURE 8 Frequency of cold spot occurrence and cold spot location in the structural dataset. (a) Frequency of the three types of cold spots. Complexes that contain at least one cold spot due to cavity, CH, and SC interactions were considered for the count. (b) Frequency of peripheral cold spots showing the total number of cold spots, cold spots due to cavities, CH, and SC interactions. Cold spots were assigned as peripheral if their PI was <0.68
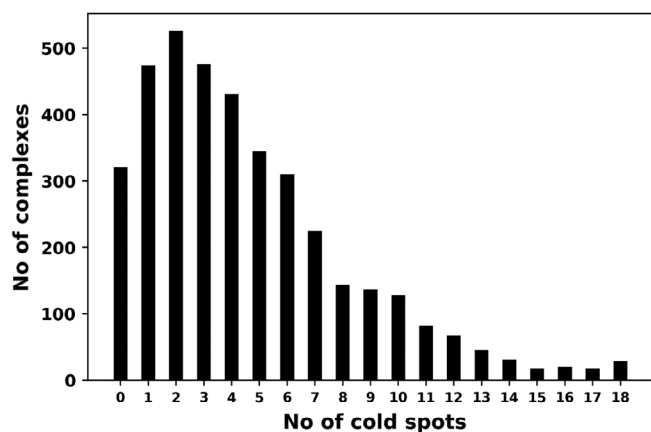
FIGURE 9 Histogram of the total number of cold spots per PPI in the structural database
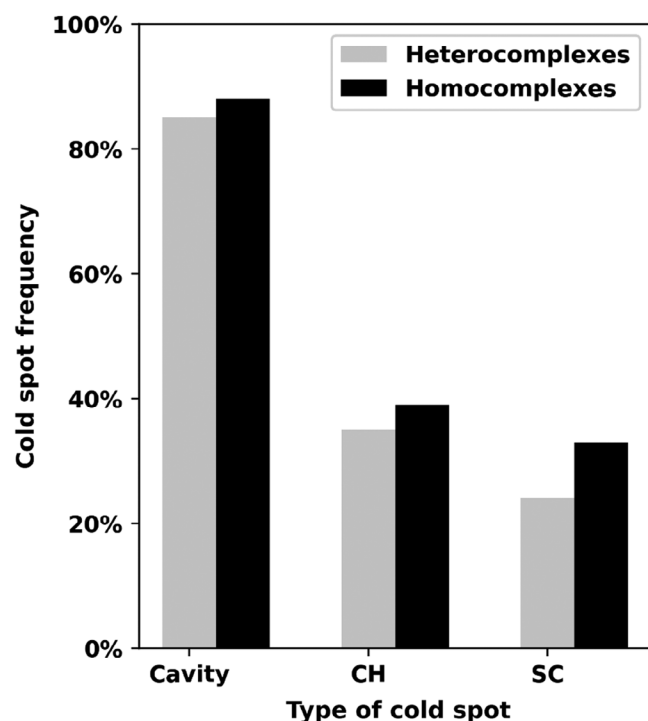

FIGURE 10 Frequency of cold spots in hetero- and homo-complexes. Percentage of complexes that contain at least one cold spot due to cavity, CH, and SC interactions in hetero-complexes (gray) and homo-complexes (black)

surrounding a cavity, we selected the residue with the closest Cα to the center of the cavity and with a side chain protruding toward the center of the cavity as such a residue could be mutated to eliminate the cavity and enhance affinity. Not surprisingly, we observed that small amino acids such as Gly, Ala, and Ser appear most commonly at cold spots due to cavities constituting 24%, 10%, and 9% of cold spots cavities, respectively (Figure 11c).

5 | DISCUSSION

Previous studies identified that various positions in the binding interface contribute differently to binding free energy with cold spot positions having potential to bring affinity enhancement upon mutations. This study examined the frequency and the spatial distribution of cold spots resulting from three types of unfavorable interactions: cavities, SC, and CH interactions. We found that cold spots are rather frequent in PPIs with an average PPI containing four cold spots. Especially frequent are cold spots due to cavities, which are present in ∼90% of all PPIs. Such positions are usually occupied by small amino acids in the wild-type complex; mutating them to larger amino acids could eliminate the cavity, enhancing binding affinity. The frequent occurrence of cold spot cavities in PPIs implies that the affinity of most PPIs could be easily improved through protein engineering and design. In fact, this strategy has been already pursued by a number of studies where affinity of a PPI was increased by filling up a cavity at the binding interface.[36,38,48]

Large-scale analysis of PPI structures showed that most cold spots are located at the periphery of the binding interface. This finding is even more prominent in high-affinity complexes that might be more evolutionarily optimized compared to low-affinity complexes. The preference for cold spots to be located in the interface periphery agrees with our general understanding of binding interface architecture, with binding hot-spots frequently occurring in the binding interface center.[25] The predominantly peripheral location of cold spots might be the reason why cold spots do not occur in clusters unlike binding hot spots.[25] This geometry of hot-spot/cold spot distribution is important for protein evolution, where low-affinity complexes could be first created with central hot spot residues that dock the complex in a particular orientation, surrounded by cold spot positions. Evolution could then work to introduce affinity enhancing mutations and to eliminate a large number of cold spots, reaching nearly optimal interactions with minimal number of cold spots in some complexes.[49] Yet, some cold spots do occur in the core of the interface. For example, in a complex between α-chymotrypsin and BPTI, cold spot position was experimentally observed at the very central binding interface position 15, where multiple mutations to hydrophobic amino acids lead to affinity improvement.[49]

Previous studies have shown that three amino acids, tryptophan, arginine, and tyrosine have been found most frequently at hot spot positions.[25] Predominance of these amino acids at hot spots could be explained by their large hydrophobic surface areas and an ability to participate in
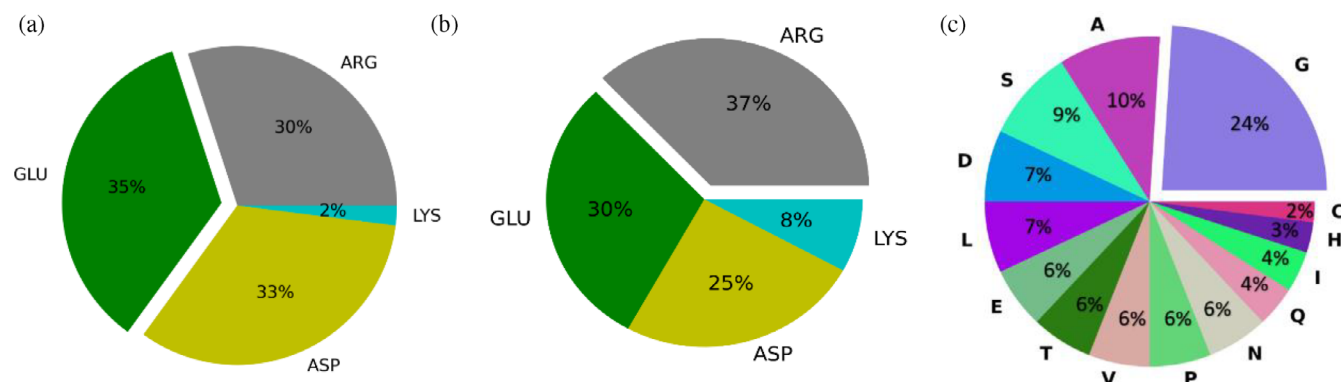
FIGURE 11   Amino acid composition of cold spots. Cold spots due to CH interactions (a), SC interactions (b), and cavities (c)

hydrogen bond interactions. In comparison, the majority of cold spots are formed by a glutamic acid and an arginine at CH and SC cold spots, respectively. The high frequency of arginine in SC cold spots is not surprising since arginine is found with high frequency in protein–protein interfaces[50,51] and its long side chain increases its probability of interacting with another arginine or lysine across the binding interface (Figure 1c). The preference of glutamate in CH interactions might be due to its high frequency in protein interfaces.[52,53] Additional explanation might be relatively high occurrence of cation/pi interactions in protein structures.[54] Such interactions were excluded from our count of CH cold spot positions, thus making negatively charged amino acids appear more frequently as cold spots. On the other hand, lysine is found least frequently among all charged amino acids in both SC and CH interactions. In addition, lysine appears rarely in hot-spots.[25] This finding could be explained by a generally low propensity of lysines to be found in protein–protein interfaces.[52] Not surprisingly, we also observed that small amino acids such as glycine and alanine are found most frequently in cold spots due to cavities, while these amino acids do not occupy hot spot positions.

Our analysis showed that cold spots occur more frequently in homocomplexes compared to heterocomplexes. The most likely reason for this finding is that binding interfaces of homocomplexes are symmetrical with every single residue appearing twice in the binding interface. It is more difficult to optimize all intermolecular interactions when a constraint of symmetry is imposed, thus leaving some positions "stuck" in suboptimal interactions. No symmetry constraint has been imposed in heterocomplexes, allowing them to evolve more optimal interfaces. In fact, it has been shown that the density of intermolecular hydrogen bonds per interface residue is higher for heterocomplexes compared to homocomplexes.[55] In addition, several studies reported

that homodimers exhibit larger binding interfaces compared to heterodimers,[50,53,55–57] adding another possible explanation why homocomplexes contain a larger number of cold spots. The relatively small difference between occurrence of cold spots in the two groups could be due to additional constraints in heterodimeric PPIs such as constraints for solubility of individual proteins and/or constraints due to interactions with multiple binding partners.[58]

In this study, we did not observe a direct correlation between the number of cold spots and binding affinity. This is likely due to the fact that binding interfaces of PPIs are highly different in shape and sizes, making them difficult to compare. Furthermore, other types of interactions that were not examined in this study are important in determining PPI binding affinity.[59] However, cold spots were found more frequently in non-cognate complexes compared to cognate complexes. Non-cognate complexes usually exhibit same interaction mode and a very similar binding interface area compared to their cognate counterparts. Hence, most of the intermolecular interactions are preserved from cognate to non-cognate complex. Yet, several residues that differ could create unfavorable interactions in non-cognate complexes, resulting in cold spots due to cavities, CH, or SC interactions. In some instances, a hot-spot in a cognate complex could be converted into a cold-spot in the non-cognate complex.[49] Even one such cold spot could weaken affinity of this PPI by several orders of magnitude, especially if located in the core of the binding interface, where energetic contribution to affinity is generally high.[35,60]

Our high-throughput cold spot analysis thus explains some principles of PPI evolution. In addition, cold spot identification could be useful in designing drugs that target PPIs. For example, identifying cavities at binding interfaces could allow scientists to target such cavities with small molecules that serve as molecular glue. Stabilization of PPIs that act against disease or result in

degradation of a particular disease-associated protein is an attractive therapeutic strategy. One example of a PPI stabilizer is a drug Tafamidis that stabilizes transthyrectin dimeric interface and is used to treat transthyrectin amyloidosis.[61] Another example is forskolin that stabilizes interaction between the C1 and C2 subunits of adenylyl cyclase and is used to treat hypertension and respiratory disorders.[62] A variety of other natural products and synthetic compounds that stabilize PPIs have been already discovered and are being pursued as leads for drug development.[63,64] In addition to targeting cavities at binding interfaces, one can utilize unfavorable CH and SC interactions when designing protein-based therapeutics.[38,65] Such an approach is especially useful when engineering drugs from natural protein effectors that bind to their targets with suboptimal affinities; affinity of such PPIs could be enhanced by mutating cold spot residues to more optimal choices and hence enhancing PPI affinity and specificity.[66]

In conclusion, we show that cold spots are universal in PPIs, presenting imperfections of binding interface design. Such imperfections could be corrected through multiple choices of mutations, giving rise to affinity enhancement. The presence of cold spots in the majority of PPIs is needed to keep PPI affinity in the desired range that would support the PPI function in the cell. Some cold-spots might arise due to inability to satisfy all constraints in multispecific and homodimeric interactions. Other cold spots might arise from the requirements of folding or protein function in the unbound form. Cold spot positions could serve as great sites for enhancing binding affinity or specificity toward one species or one particular homolog. Thus, computational identification of cold spots would greatly assist the efforts of engineering high-affinity binders for various biotechnical applications.

6 | METHODS

6.1 | Database preparation

6.1.1 | Affinity dataset

Affinity dataset was built from the Kastritis database.[39] The database contains 144 nonredundant protein–protein complexes that have high-resolution structures available for both bound and unbound states and K_D measurements for each PPI. From this database, we excluded 11 PDB files that (a) include heteroatoms within 2 Å from the binding interface, (b) Lack the N-terminal tail close to the binding interface, (c) contain several binding interfaces close to each other complicating independent analysis of each interface. We thus were left with 133 PDB files in our affinity dataset. The database was also divided into two groups: high-affinity complexes ($K_D < 10^{-8}$ M) and low-affinity complexes ($K_D > 10^{-8}$).

6.1.2 | Structural dataset

Structural database was built from the whole Protein Data Bank (PDB) by extracting nonredundant x-ray structures of protein–protein complexes. From the structural database we excluded PPI structures that (1) were solved to resolution of worse than 2.5 Å; (2) Included heteroatoms within 2 Å of the binding interface; (3) contained short peptides of 40 residues or less as one of the chains; and (4) contained missing atoms in the binding interface region. The dataset was divided into heterodimeric complexes that were extracted directly from the PDB and homodimeric complexes that were downloaded from the database of annotated biological assemblies that distinguishes biological interfaces from crystal contact interfaces.[67] For our analysis, we downloaded homodimeric complexes with 15% or less error probability for biological assembly. In complexes where there is more than one biological assembly, we used the complex with the minimum QSbio error probability for the specific PDB ID.[67] After all exclusions, our structural database contained 1,507 hetero complexes and 2,319 homo complexes.

6.1.3 | Binding interface identification

Hydrogens were added to all PDB files with the MolProbity software[68] with asparagines, histidines, and glutamines allowed to flip. For each PPI, we first identified atoms that belong to the binding interface. The binding interface atoms were defined as all atoms on one chain that were within 4 Å from the second chain in the complex.

6.1.4 | Cavity identification

Cavities were defined as empty spaces within the binding interface. Random dots were placed within the binding interface of the protein complex and the dots, which were placed in an empty space and can accommodate a sphere of more than 1.4 Å, were retained to distinguish cavities. To cluster the dots and define different cavities, we applied the DBSCAN algorithm.[69] In this algorithm, the epsilon value (maximum distance between two points) was set to 2 and minimum point value was set to 10 dots per cluster, which corresponds to a volume higher than a water molecule.

6.1.5 | Defining unfavorable CH interactions

For each PPI, we first identified all charged buried atoms (4 Å from the surface of the protein) in the binding interface. We considered the charged oxygen atoms in the carboxylic side chains of aspartic acid and glutamic acid as negatively charged atoms while the charged nitrogen atoms in the amine side chains of lysine and arginine as positively charged atoms. For each buried charged atom, we searched for hydrophobic atoms belonging to a different residue within 4.5 Å distance from the charge. Such an interaction was defined as a charge-hydrophobic unfavorable interaction unless it belonged to one of the four exceptions listed below:

1. The charged atom participated in a favorable hydrogen bond. When the identified charged atom formed hydrogen bond(s) with other neighboring atoms and the total energy of the hydrogen bond was lower than −2 kcal/mol, the charged atom was excluded from the unfavorable interaction count. Hydrogen bond energy was calculated according to[70] with hydrogen bond equilibrium distance of 2.8 Å and a well depth of 8 kcal/mol.
2. The charged atom participated in a favorable pi–cation interaction.
   We defined pi–cation interaction when the charged atom was within a distance of 6 Å from the center of the aromatic ring and if it was located at an angle of 45° or less from the normal of the plane of the aromatic ring.[54,71]
3. The charged atom was within 3.5 Å from an atom with the opposite charge.
4. The charged atom participated in a favorable anion–aromatic interaction.
   We defined an anion–aromatic interaction if the charged atom of the negatively charged side chain is situated within 4 Å from the ring edge of an aromatic group.[72]

The charged residue that exhibited unfavorable CH interaction(s) was considered a cold spot residue.

6.1.6 | Defining unfavorable SC interactions

We next identified charged atoms in the binding interface that were situated within 4.5 Å from another charge of the same sign belonging to a different amino acid. Similar to CH interactions, we considered the charged oxygen atoms in the carboxylic side chains of aspartic acid and glutamic acid as negatively charged amino acids while the charged nitrogen atoms in the amine side chains of lysine and arginine as positively charged atoms. Such an interaction was considered to be unfavorable with two exceptions stated below:

1. The charged residue participates in favorable hydrogen bonds. Similar to CH interactions, when the identified charged atom formed hydrogen bond(s) with other neighboring atoms with a total energy lower than −2 kcal/mol, the charged atom was excluded from SC interaction count. Out of the two charged residues that form SC interactions, if only one charged residue participates in a favorable hydrogen bond, only that residue is excluded from the cold spot count.
2. The charged residues are found within 3.5 Å from atoms with an opposite charge. If the SC interaction(s) was unfavorable, both residues that contain the charge atoms were considered as cold spots.

6.1.7 | Dividing binding interface into core and periphery

For each atom in the binding interface, a periphery index (PI) was calculated. For this purpose, we first calculated atom-to-surface distance, that is, the distance between this atom and the closest atom on the surface of the protein–protein complex that is not part of the interface. To calculate PI for a particular atom, its atom-to-surface distance was divided by the maximum atom-to-surface distance for the binding interface. Atoms with PI higher than 0.68 were assigned to the core of the binding interface and atoms with PI less than 0.68 were assigned to the periphery of the interface, such that both groups would contain approximately the same number of atoms. Each cold spot due to CH or SC interaction was then assigned with a PI according to the location of the charged atom belonging to the cold spot residue (Figure 3). To classify colds spots due to cavities, the distance between the center of each cavity and the closest atom on the surface of the protein that is not a part of the interface was calculated. To calculate PI, the distance between the center of each cavity and the closest atom on the surface was divided by the maximum atom-to-surface distance for the binding interface. Cavities with PI smaller than 0.68 were classified as peripheral while cavities with PI higher than 0.68 were defined as core (Figure 3).

6.1.8 | Cold spots cluster analysis

We calculated pairwise distances between cold spots to analyze whether they are clustered in the same region. In these calculations, we used the position of the Cα atom of

the cold spots due to CH and SC interactions and the position of the center of the cavity—for cold spots due to cavities. Pairwise distances between cold spots were analyzed for all types of cold spots together and for each type of cold spots separately.

6.1.9 | Dividing the dataset into different functional classes

Affinity database was divided into three main classes of complexes: antigen–antibody, enzyme-containing, and other complexes in which sub classes were introduced into the latter two. The other class was divided into three subgroups as G proteins, receptors containing, and miscellaneous. We made use of this categorization and divided our dataset into four classes: antigen–antibody, G proteins, enzyme-containing complexes, and other receptor containing complexes.

6.1.10 | Calculating p values

We calculated p values using one-tailed unpaired T test for Figures 4 and 10. For the figure S3, we used the two-tailed unpaired T test.

AUTHOR CONTRIBUTIONS

Sagara N. S. Gurusinghe: Formal analysis (lead); investigation (lead); methodology (equal); software (lead); validation (lead); visualization (lead); writing – original draft (supporting); writing – review and editing (supporting). **Ben Oppenheimer:** Investigation (supporting); methodology (equal); software (equal); validation (supporting); writing – review and editing (supporting). **Julia M. Shifman:** Conceptualization (lead); funding acquisition (lead); project administration (lead); supervision (lead); writing – original draft (lead); writing – review and editing (lead).

ACKNOWLEDGMENTS

DATA AVAILABILITY STATEMENT
All data are provided in Supporting Information.

ORCID
Julia M. Shifman https://orcid.org/0000-0001-6678-6497

REFERENCES
1. Braun P, Gingras AC. History of protein-protein interactions: From egg-white to complex networks. Proteomics. 2012;12: 1478–1498.
2. Navolotskaya EV, Sadovnikov VB, Zinchenko DV, Vladimirov VI, Zolotarev YA. Interaction of cholera toxin B subunit with rat intestinal epithelial cells. Russ J Bioorganic Chem. 2018;44:403–407.
3. Badran AH, Guzov VM, Huai Q, et al. Continuous evolution of Bacillus thuringiensis toxins overcomes insect resistance. Nature. 2016;533:58–63.
4. Cohen A, Rosenthal E, Shifman JM. Analysis of structural features contributing to weak affinities of ubiquitin/protein interactions. J Mol Biol. 2017;429:3353–3362.
5. Tonddast-Navaei S, Skolnick J. Are protein-protein interfaces special regions on a protein's surface? J Chem Phys. 2015;143: 243149.
6. Ozgur S, Basquin J, Kamenska A, Filipowicz W, Standart N, Conti E. Structure of a human 4E-T/DDX6/CNOT1 complex reveals the different interplay of DDX6-binding proteins with the CCR4-NOT complex. Cell Rep. 2015;13:703–711.
7. Chothia C, Janin J. Principles of protein-protein recognition. Nature. 1975;256:705–708.
8. Miller S, Lesk AM, Janin J, Chothia C. The accessible surface area and stability of oligomeric proteins. Nature. 1988;328:834–836.
9. Flock T, Hauser AS, Lund N, Gloriam DE, Balaji S, Babu MM. Selectivity determinants of GPCR-G-protein binding. Nature. 2017;545:317–322.
10. Keskin O, Tuncbag N, Gursoy A. Predicting protein-protein interactions from the molecular to the proteome level. Chem Rev. 2016;116:4884–4909.
11. Guo F, Li SC, Wei Z, Zhu D, Shen C, Wang L. Structural neighboring property for identifying protein-protein binding sites. BMC Syst Biol. 2015;9:1–9.
12. Gromiha MM, Yugandhar K, Jemimah S. Protein–protein interactions: Scoring schemes and binding affinity. Curr Opin Struct Biol. 2017;44:31–38.
13. Kumar H, Shifman JM. Predicting the consequences of mutations. Protein interactions: Computational methods, analysis and applications. Singapore: World Scientific Publishing Co., 2020; p. 145–165.
14. Zhou P, Wang C, Tian F, Ren Y, Yang C, Huang J. Biomacromolecular quantitative structure-activity relationship (BioQSAR): A proof-of-concept study on the modeling, prediction and interpretation of protein-protein binding affinity. J Comput Aided Mol Des. 2013;27:67–78.
15. Horton N, Lewis M. Calculation of the free energy of association for protein complexes. Protein Sci. 1992;1:169–181.
16. Dehouck Y, Kwasigroch JM, Rooman M, Gilis D. BeAtMuSiC: Prediction of changes in protein-protein binding affinity on mutations. Nucleic Acids Res. 2013;41:333–339.
17. Benedix A, Becker CM, de Groot BL, Caflisch A, Böckmann RA. Predicting free energy changes using structural ensembles. Nat Methods. 2009;6:3–4.
18. Vangone A, Bonvin AMJJ. Contacts-based prediction of binding affinity in protein–protein complexes. Elife. 2015;4:e07454.
19. Raucci R, Laine E, Carbone A. Local interaction signal analysis predicts protein-protein binding affinity. Structure. 2018;26: 905–915.

20. Pires DEV, Ascher DB. mCSM-AB: A web server for predicting antibody–antigen affinity changes upon mutation with graph-based signatures. Nucleic Acids Res. 2016;44:W469–W473.

21. Rodrigues CHM, Myung Y, Pires DEV, Ascher DB. MCSM-PPI2: Predicting the effects of mutations on protein-protein interactions. Nucleic Acids Res. 2019;47:W338–W344.

22. Erijman A, Rosenthal E, Shifman JM. How structure defines affinity in protein-protein interactions. PLoS One. 2014;9: e110085.

23. Clackson T, Wells JA. A hot spot of binding energy in a hormone-receptor interface. Science (80-). 1995;267:383–386.

24. David A, Razali R, Wass MN, Sternberg MJE. Protein-protein interaction sites are hot spots for disease-associated nonsynonymous SNPs. Hum Mutat. 2012;33:359–363.

25. Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces. J Mol Biol. 1998;280:1–9.

26. Rosell M, Fernández-Recio J. Hot-spot analysis for drug discovery targeting protein-protein interactions. Expert Opin Drug Discov. 2018;13:327–338.

27. Garcia-Seisdedos H, Empereur-Mot C, Elad N, Levy ED. Proteins evolve on the edge of supramolecular self-assembly. Nature. 2017;548:244–247.

28. Cukuroglu E, Engin HB, Gursoy A, Keskin O. Hot spots in protein-protein interfaces: Towards drug discovery. Prog Biophys Mol Biol. 2014;116:165–173.

29. Vakser IA, Matar OG, Lam CF. A systematic study of low-resolution recognition in protein-protein complexes. Proc Natl Acad Sci U S A. 1999;96:8477–8482.

30. Chakrabarti P, Janin J. Dissecting protein-protein recognition sites. Proteins: Struct Funct Genet. 2002;47:334–343.

31. Liu Q, Li J. Protein binding hot spots and the residue-residue pairing preference: A water exclusion perspective. BMC Bioinformatics. 2010;11:244.

32. Keskin O, Ma B, Nussinov R. Hot regions in protein-protein interactions: The organization and contribution of structurally conserved hot spot residues. J Mol Biol. 2005;345:1281–1294.

33. Kenneth Morrow J, Zhang S. Computational prediction of protein hot spot residues. Curr Pharm Des. 2012;18:1255–1265.

34. Shirian J, Sharabi O, Shifman JM. Cold spots in protein binding. Trends Biochem Sci. 2016;41:739–745.

35. Meenan NAG, Sharma A, Fleishman SJ, et al. The structural and energetic basis for high selectivity in a high-affinity protein-protein interaction. Proc Natl Acad Sci U S A. 2010; 107:10080–10085.

36. Diskin R, Scheid JF, Marcovecchio PM, et al. Increasing the potency and breadth of an HIV antibody by using structure-based rational design. Science. 2011;334:1289–1293.

37. Leung I, Dekel A, Shifman JM, Sidhu SS. Saturation scanning of ubiquitin variants reveals a common hot spot for binding to USP2 and USP21. Proc Natl Acad Sci U S A. 2016;113:8705–8710.

38. Aizner Y, Sharabi O, Shirian J, et al. Mapping of the binding landscape for a picomolar protein-protein complex through computation and experiment. Structure. 2014;22:636–645.

39. Kastritis PL, Moal IH, Hwang H, et al. A structure-based benchmark for protein-protein binding affinity. Protein Sci. 2011;20:482–491.

40. Hartley RW. Directed mutagenesis and Barnase-Barstar recognition. Biochemistry. 1993;32:5978–5984.

41. Li W, Hamill SJ, Hemmings AM, Moore GR, James R, Kleanthous C. Dual recognition and the role of specificity-determining residues in colicin E9 DNase-immunity protein interactions. Biochemistry. 1998;37:11771–11779.

42. Vincent JP, Lazdunski M. Trypsin-pancreatic trypsin inhibitor association. Dynamics of the interaction and role of disulfide bridges. Biochemistry. 1972;11:2967–2977.

43. Bode W. The transition of bovine trypsinogen to a trypsin-like state upon strong ligand binding. II. The binding of the pancreatic trypsin inhibitor and of isoleucine-valine and of sequentially related peptides to trypsinogen and to p-guanidinobenzoate-trypsinoge. J Mol Biol. 1979;127:357–374.

44. Fleury D, Wharton SA, Skehel JJ, Knossow M, Bizebard T. Antigen distortion allows influenza virus to escape neutralization. Nat Struct Biol. 1998;5:119–124.

45. Cauerhff A, Goldbaum FA, Braden BC. Structural mechanism for affinity maturation of an anti-lysozyme antibody. Proc Natl Acad Sci U S A. 2004;101:3539–3544.

46. Arold S, O'Brien R, Franken P, et al. RT loop flexibility enhances the specificity of Src family SH3 domains for HIV-1 Nef. Biochemistry. 1998;37:14683–14691.

47. Cho S, Swaminathan CP, Bonsor DA, et al. Assessing energetic contributions to binding from a disordered region in a protein-protein interaction. Biochemistry. 2010;49:9256–9268.

48. Sammond DW, Eletr ZM, Purbeck C, Kimple RJ, Siderovski DP, Kuhlman B. Structure-based protocol for identifying mutations that enhance protein-protein binding affinities. J Mol Biol. 2007;371:1392–1404.

49. Heyne M, Shirian J, Cohen I, et al. Climbing up and down binding landscapes through deep mutational scanning of three homologous protein-protein complexes. J Am Chem Soc. 2021; 143:17261–17275.

50. Bahadur RP, Chakrabarti P, Rodier F, Janin J. Dissecting subunit interfaces in homodimeric proteins. Proteins: Struct Funct Genet. 2003;53:708–719.

51. Talavera D, Robertson DL, Lovell SC. Characterization of protein-protein interaction interfaces from a single species. PLoS One. 2011;6:e21053.

52. Levy ED. A simple definition of structural regions in proteins and its use in analyzing Interface evolution. J Mol Biol. 2010; 403:660–670.

53. Yan C, Wu F, Jernigan RL, Dobbs D, Honavar V. Characterization of protein-protein interfaces. Protein J. 2008;27:59–70.

54. Dougherty DA. The cation-π interaction. Acc Chem Res. 2013; 46:885–893.

55. Zhanhua C, Gan JG-K, Lei L, Sakharkar MK, Kangueane P. Protein subunit interfaces: Heterodimers versus homodimers. Bioinformation. 2005;1:28–39.

56. Jones S, Thornton JM. Principles of protein-protein interactions. Proc Natl Acad Sci U S A. 1996;93:13–20.

57. Caffrey DR. Are protein-protein interfaces more conserved in sequence than the rest of the protein surface? Protein Sci. 2004; 13:190–202.

58. Fromer M, Shifman JM. Tradeoff between stability and multi-specificity in the design of promiscuous proteins. PLoS Comput Biol. 2009;5:e1000627.

59. Yugandhar K, Gromiha MM. Protein-protein binding affinity prediction from amino acid sequence. Bioinformatics. 2014;30: 3583–3589.

60. Li L, Zhao B, Cui Z, Gan J, Sakharkar MK, Kangueane P. Identification of hot spot residues at protein-protein interface. Bioinformation. 2006;1:121–126.

61. Lorente L, Botta M, Ottmann C. Small-molecule modulators of protein-protein interactions: Focus on 14-3-3 PPIs. Protein–protein interaction regulators. London: The Royal Society of Chemistry, 2021, pp. 249–279.

62. Simpson J, Pálvölgyi A, Antoni FA. Direct stimulation of adenylyl cyclase 9 by the fungicide imidazole miconazole. Naunyn Schmiedebergs Arch Pharmacol. 2019;392:497–504.

63. Chang C, Lin SM, Satange R, et al. Targeting protein-protein interaction interfaces in COVID-19 drug discovery. Comput Struct Biotechnol J. 2021;19:2246–2255.

64. Soini L, Leysen S, Davis J, Ottmann C. Molecular glues to stabilise protein–protein interactions. Curr Opin Chem Biol. 2022; 69:102169.

65. Sharabi O, Shirian J, Grossman M, Lebendiker M, Sagi I, Shifman J. Affinity- and specificity-enhancing mutations are frequent in multispecific interactions between TIMP2 and MMPs. PLoS One. 2014;9:e93712.

66. Bonadio A, Shifman JM. Computational design and experimental optimization of protein binders with prospects for biomedical applications. Protein Eng Des Sel. 2021;34:1–7.

67. Dey S, Ritchie DW, Levy ED. PDB-wide identification of biological assemblies from conserved quaternary structure geometry. Nat Methods. 2018;15:67–72.

68. Davis IW, Leaver-Fay A, Chen VB, et al. MolProbity: All-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res. 2007;35:W375–W383.

69. Ester M, Kriegel H-P, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining; 1996; 226–231.

70. Gordon DB, Marshall SA, Mayo SL. Energy functions for protein design. Curr Opin Struct Biol. 1999;9:509–513.

71. Gallivan JP, Dougherty DA. Cation-π interactions in structural biology. Proc Natl Acad Sci U S A. 1999;96:9459–9464.

72. Thomas KA, Smith GM, Thomas TB, Feldmann RJ. Electronic distributions within protein phenylalanine aromatic rings are reflected by the three-dimensional oxygen atom environments. Proc Natl Acad Sci U S A. 1982;79:4843–4847.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.