Posted on 14 Dec 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.167101052.24746048/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Three types of hotspots- weak, moderate and strong in protein-protein and protein-peptide interactions

Kiran Kumar A.[1] and R.S. Rathore[1]

[1]Central University of South Bihar

December 14, 2022

Abstract

Protein-protein and protein-peptide interactions (PPIs and PPepI) belong to a similar category of interactions yet seemingly subtle differences exist among them. To characterize differences between protein-protein (PPI) and protein-peptide interactions (PPepI), we have focussed on two important class of residues- hotspot and anchor residues. Using implicit solvation-based free-energy calculations, a very large-scale alanine scanning has been performed on benchmarking dataset, consisting of over 5500 interface residues. The differences in the two categories are more pronounced, if the hotspot data is divided into three distinct types, namely - weak hotspots (having binding free energy loss upon Ala mutation, ΔΔG, 2-10 kcal/mol), moderate hotspots (ΔΔG, 10-20 kcal/mol) and strong hotspots (ΔΔG [?] 20 kcal/mol). The analysis suggests that for PPI, weak hotspots are predominantly populated by polar and hydrophobic residues. The distribution shifts towards charged and polar residues for moderate hotspot and charged residues (principally Arg) are overwhelmingly present in the strong hotspot. In contrast, in the PPepI dataset, the distribution shifts from predominantly hydrophobic and polar (in the weak type) to almost similar preference for polar, hydrophobic and charged residues (in moderate type) and finally the charged residue (Arg) and Trp are mostly occupied in the strong type. In anchor residue class of both categories, the preferred residues are Arg, Tyr and Leu, possesing bulky side chaing and which also strike a delicate balance between side chain flexibility and rigidity. The present knowledge should aid in effective design of biologics, when augmentation or disruption of PPI are substituted with the peptide-based mimics.

Three types of hotspots- weak, moderate and strong in protein-protein and protein-peptide interactions

Kiran Kumar A.[1] and R.S. Rathore*[1]

[1]Department of Bioinformatics, School of Earth, Biological and Environmental Sciences, Central University of South Bihar, Gaya, India 824236

* Correspondence: rsrathore@cusb.ac.in

Abstract

Protein-protein and protein-peptide interactions (PPIs and PPepI) belong to a similar category of interactions yet seemingly subtle differences exist among them. To characterize differences between protein-protein (PPI) and protein-peptide interactions (PPepI), we have focussed on two important class of residues- hotspot and anchor residues. Using implicit solvation-based free-energy calculations, a very large-scale alanine scanning has been performed on benchmarking dataset, consisting of over 5500 interface residues. The differences in the two categories are more pronounced, if the hotspot data is divided into three distinct types, namely - weak hotspots (having binding free energy loss upon Ala mutation, ΔΔG, 2-10 kcal/mol), moderate hotspots (ΔΔG, 10-20 kcal/mol) and strong hotspots (ΔΔG [?] 20 kcal/mol). The analysis suggests that for PPI, weak hotspots are predominantly populated by polar and hydrophobic residues. The distribution shifts towards

Posted on 14 Dec 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.167101052.24746048/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

charged and polar residues for moderate hotspot and charged residues (principally Arg) are overwhelmingly present in the strong hotspot. In contrast, in the PPepI dataset, the distribution shifts from predominantly hydrophobic and polar (in the weak type) to almost similar preference for polar, hydrophobic and charged residues (in moderate type) and finally the charged residue (Arg) and Trp are mostly occupied in the strong type. In anchor residue class of both categories, the preferred residues are Arg, Tyr and Leu, possesing bulky side chaing and which also strike a delicate balance between side chain flexibility and rigidity. The present knowledge should aid in effective design of biologics, when augmentation or disruption of PPI are substituted with the peptide-based mimics.

Keywords: - hotspot, anchor residue, protein-protein interaction, protein-peptide interaction, alanine scanning, free energy calculations

Introduction

Protein-protein interactions (PPI) and protein-peptide interactions (hereafter referred to as -PPepI) play a crucial role in many cell-signaling and metabolic processes. In recent years, they have become attractive targets for drug and biologics discovery [1-8]. To interfere or modulate PPI, protein-protein complex may be disrupted or interactions may be augmented by peptides or mimetics. Peptides have the advantage of larger shelf life, less prone to proteolysis, feasibility of their oral delivery and flexibility in optimization, screening and synthesis, similar to small molecule based drug. Thus, to augment or disrupt PPI by peptide-based therapeutic, it is worthwhile to understand the differences between PPI and PPepI.

To characterize differences between PPI and PPepI, in the present work we have focused on two important class of residues, namely, hotspot and anchor residues. At the interface of molecular complex, certain residues play a crucial role in governing such interactions. These residues contributing most to the binding of molecular complexes are termed as hotspot residues. As a rule of thumb, a hotspot residue is defined, if its energy contribution to the binding of complex is more than 2 kcal/mol [9-10]. The binding affinity and specificity of the PPI and PPepI are in essence governed by such energetically important hotspot residues. Mutations at hotspots have been shown to cause dissociation of complexes or disruption in the signaling processes [11]. Additionally, anchoring residue also play specific role in the initial stage of molecular recognition. They avoid kinetically costly structural rearrangements in the binding pathway allowing for a relatively smooth recognition process [12]. Anchor residues are hotspot residues that bury the highest amounts of solvent accessible surface area ([?] 100 $Å^2$) upon binding [12-13].

To quantitatively probe the energetic contributions of residue to the overall binding of molecular complexes, earlier experimental alanine scanning mutagenesis were attempted [9]. This involves experimentally mutating each residue at the interface to alanine and then measuring the effect of the mutation on binding to the partner protein or peptide. Extending the previous definition of hotspot more specifically with the Ala-scanning method, a residue is considered a hotspot residue, if its mutation to alanine gives rise to loss in binding affinity [?] 2 kcal/mol. Early on, limited data were generated using experiment mutagenesis for many diverse complexes [14-22], and few databases on experimentally determined hotspot residues in protein-protein complexes such as AB-bind [14], ASEdb [15], ATLAS [16], BID [17], DACUM [18], dbAMEPNI [19], dbMPIKT [20], PROXiMATE [21] and SKEMPI 2.0 [22] have also been developed. The identification of hotspot residues by experimental techniques is a costly and time consuming job, which has led to the development of a series of computational algorithms to predict hotspots at the protein-protein interfaces [1, 23-39]. Serrano's group have developed FoldX based on an empirical potential for rapid evaluation of effect of mutations in proteins and nucleic acids [23,24]. Flex_ddG method in Rosetta uses a combination of sophisticated Monte Carlo sampling, minimization, and specialized force fields [25-27]. BeAtMuSiC is a tool, which employ statistical potentials adapted to a coarse-grained representation of protein structures [28] and mCSM (mutation Cutoff Scanning Matrix) is based on graph-based structural signatures, which encodes distance patterns between atoms to represent protein residue environments [29]. Various other methods and their applications have also been discussed in the literature [30-36, 44-45].

In the present work, we have used Bioluminate Residue scanning approach implemented in Schrodinger [37-

Posted on 14 Dec 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.167101052.24746048/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

38]. The method uses implicit solvation based MM-GBSA (Molecular Mechanics with Generalized Born and Surface Area) with the OPLS2005 force field, VSGB2.0 solvent model and rotamer search algorithms from Prime [37-38]. To understand the differences in the nature of hotspots in protein-protein and protein-peptide complexes using benchmarking data sets, a comprehensive Ala scanning, comprising of over 5500 mutation analyses were performed.

Materials and Methods

2.1 Benchmarking dataset

Benchmarking dataset of protein-protein and protein-peptide complexes were taken from Vrevenet al. [39] and Hauser et al. [40], respectively. These datasets were earlier used for benchmarking studies of docking programs. Protein-protein benchmarking dataset is composed of non-redundant high-quality structures (resolution $<$ 3.25 A). Protein-peptide complexes were composed of peptides with 3 to 12 residues (resolution $<$ 2.0 A, R-free $<$ 0.3). The dataset consists of 134 protein-protein and 50 protein-peptide complexes. The PDB IDs are provided in Table 1.

2.2 Protein preparation

PPI and PPepI dataset entries were imported in Schrodinger Maestro [38] and prepared as discussed here. Each of the entries in both the benchmarking dataset were prepared using a Protein Preparation Wizard utility available in Maestro [38]. For each PDB, protein hydrogens were stereochemically added and appropriate ionization states (at pH 7.0) for the acidic and basic amino acid residues were maintained. Missing side-chain atoms were added using a rotamer library of Xiang and Honig [41] and missing backbone atoms were fixed using a loop modeling procedure implemented in Prime [38]. Subsequently, the structure optimization, which utilizes maximization of hydrogen bonding, was carried out to i) identify the most probable positions for freely rotating –OH and –SH hydrogens, ii) to assign the correct protonation states of charged residues, iii) tautomers of His, and iv) Chi 'flip' of Asn, Gln and His residues. Finally, to relieve any steric clashes within atoms, the structures were energy minimized using OPLS2005 force field until the heavy-atom displacement converged to root mean square deviation (RMSD) of 0.30 A. Many entries were manually examined and processed, where it was not possible to prepare using the default option.

2.3 Hotspot identification using residue scanning

The Ala scan was performed on the prepared library of protein-protein and protein-peptide complexes using the Residue Scanning module in Bioluminate [38]. Ala mutation was carried out, one residue at a time, for all the interface residues. The method calculates relative binding affinity values (G_{bind} or simply G; also referred as Affinity in the reference [38]) and stability ($[?][?]G_{stability}$) between the mutant and the wild type protein using implicit solvation based MM-GBSA and using the thermodynamic cycle approach. The details of thermodynamic cycle, which allows one to calculate the net G_{bind} and $[?][?]G_{stability}$, taking the advantage of state function nature of the free energy is described in literature [37,42]. For interface residues, default 4A distance between residues in either of the protein chains was used [38].

The $\Delta[?]G_{bind}$ is the change in binding affinity between binding partners upon point mutation. One of the difference (Δ) is between the bound and unbound state of binding partners using MM-GBSA and the another difference (Δ) is between wild type and the mutant. The positive value indicates loss in binding affinity upon mutation to Ala and negative value indicates gain in affinity. The calculations were carried out using the Schrödinger Prime MM-GBSA, which uses an implicit (continuum) solvation model [43]. $\Delta G_{stability}$ or ΔStability (solvated) is the change in the stability of the protein upon mutation, also calculated using the Prime energy function. The stability was defined as the difference in free energy between the folded state of molecule and the corresponding unfolded state were estimated from Gly-X-Gly tripeptide, where X is the residue under consideration [38]. A negative value refers that the mutant is more stable. Residue involving covalent linkage (having $\Delta[?]G_{bind}>$ 80 kcal/mol) were discarded. Finally, all the hotspot residues with $\Delta[?]G_{bind}$ in 2-80 kcal/mol range were considered for analysis. All the results were analysed using Maestro visualizer, Bioluminate utilities and Microsoft Excel.

Posted on 14 Dec 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.167101052.24746048/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

2.4 Data clustering & Multiple linear regression analysis

Data clustering was carried out using k-means clustering in Weka platform using Elbow method. Multiple Linear Regression (MLR) analysis was carried out using *lm* library in R to decipher correlation between outcome variable, $\Delta\Delta$G and all its components that contribute to total free energy calculations as predictor variable assuming that they follow approximately linear relationship.

2.5 Anchor residues

Anchor residues and their ΔSASA (solvent accessible surface area) were calculated for the both PPI and PepPI benchmark datasets from ANCHOR webserver [13]. ΔSASA is the change in the solvent accessible surface area for the side-chain upon binding. The program also estimates the contribution of side-chain to the binding free energy. ΔSASA and binding free energy were calculated for all interface residues. Anchor residues were identified that burry the solvent accessible surface area, ΔSASA [?] 100 Å^2 upon binding.

Results and Discussion

Protein-protein and protein-peptide benchmark dataset having 134 and 50 dimer complexes, respectively, were used for calculations and analysis. To identify the hotspots, all interface residues of both partner chains in the complex were mutated to alanine and relative binding affinity ([?][?]G) values were calculated upon mutation of all these 184 complexes. Ala scan was carried out using Bioluminate residue scanning module [37-38]. The positive value of [?][?]G indicates loss in binding affinity upon mutation and suggest quantitatively the importance of the particular residue. A negative value refers to the gain in binding affinity indicating that the mutant binds better.

3.1 Frequency distribution of hotspot and anchor residues

In all, calculations were made for a total of 5774 interface residues. To our knowledge, this is one of the comprehensive study on hotspot residues in PPI and PPepI. Out of 5774 residues, 3732 residues amounting to 64.6% of the total dataset belong to hotspot categories having [?][?]G [?] 2 kcal/mol. This is suggestive of the fact that nature has remarkably optimized a great majority (˜65%) of the interface residues in protein-protein complexes during evolution. This finding is also in contrast to earlier notion that complex interface comprises only few hot spot residues either isolated or in clusters [51]. The frequency distribution is given in Table 2. Fig. 1 illustrates the histogram of hotspots in PPI and PPepI categories. Altogether both charged and polar residues contribute about 60% of hotspots. In the histogram, the frequency curves drawn for both categories are in perfect sync with a very minor difference at Gln, indicating the similar overall tendency observed in PPI and PPepI. Arg, Tyr, Leu, Lys and Gln, are the preferred hotspot residues at the PPI interfaces with Arg alone accounting for over 10% in the frequency distribution. Met, His, Trp, Gly and Cys are the least preferred hotspot residues with Cys presence is mere 0.1 %. In contrast, Tyr, Leu, Arg and Ile are the most favoured hotspot residues in PPepI category and Cys, Gly, Gln, Met and Trp are the least preferred ones. Examining the trend, the PPI dataset is characterized by the dominance of charged and polar residues followed by hydrophobic residues, whereas PPepI dataset, the polar and hydrophobic followed by charged residues predominantly occupy the frequency distribution. The fact that negatively charged residues are not the ones among preferred hotspot residues suggest that the electrostatic complementarity is not a predominant factor in PPI and PPepI as well.

In PPI dataset, 249 anchor residues were recognized (37 weak, 118 moderate and 87 strong hotspots types; Supplementary. Table 1). In PPepI category, 92 anchor residues were identified (12 weak, 46 moderate and 34 strong types; Supplementary. Table 2). Anchor residues mostly occur for PPI dataset in moderate and strong types. The anchor residues comprise of 8.3% of hot spot residues and about 5.3% of total residues investigated in PPI and PPepI category. Anchor residues demonstrate similar trend with Arg, Leu, Tyr Gln, Phe as the most preferred anchor residue for PPI. For PPepI hydrophobic residue predominantly occupy at the interface - Leu, Ile, Val, Phe, Tyr & Arg (Fig.1).

3.2 Three types of hotspots: weak, moderate and strong hotspots

Posted on 14 Dec 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.167101052.24746048/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

The difference in preference of hotspot residues between PPI and PPepI dataset are not very much evident with the overall frequency distribution (Fig. 1). Therefore, we tried to cluster the dataset obtained from residue scanning. The data clustering suggested the possible clusters may be 3 to 5. Upon manual examination and observing the trends, it was considered reasonable to divide the hotspot residues into approximately three different types. The difference of hotspot residues was found to be most pronounced in the following three approximately different [?][?]G ranges, we refer them as weak hotspots (loss in [?][?]G in 2-10 kcal/mol range), moderate hotspots ([?][?]G in 10-20 kcal/mol range) and strong hotspot ([?][?]G >20 kcal/mol).

Out of 3732 hotspots, a great majority of 68.7% (2565) belong to weak hotspot type. For PPI dataset, Gln, Leu Tyr are the most preferred. This is followed by Asn, Val, Lys, Glu, Ser and Pro, which also have substantial presence at the PPI interface. In contrast, in PPepI, Leu and Tyr are the most preferred hotspot residues with Leu having an overwhelming contribution in the distribution. Val, Thr, Pro and Ile also possess large frequencies in the distribution (Fig. 2). Thus, among weak hotspot type, in PPI, the high occurrence is observed for polar residues followed by hydrophobic residues and minor fraction of charged residues are also present. On the other hand, in PPepI data, hydrophobic residues are more preferred as compared to polar residues. Somewhat similar trend was observed for anchor residue in PPI category was observed, even though there are very few data observed in weak type. Frequency distribution for Gln is the highest followed by Asn and Lys. In PPepI, the paucity of data precluded us for any reliable predictions.

The data for moderate type ([?][?]G in 10-20 kcal/mol) is shown in Fig 3. About 25.4% of data (949) belong to moderate hotspot type. In contrast to the weak type, Arg is overwhelmingly present (˜18%) followed by Tyr (˜12%) even though Lys and Leu also possess sizable frequencies (about 10%) in the distribution. Thus, the distribution in PPI category is dominated by charged and polar residues and minor fraction of hydrophobic residues are also present. In contrast, the distribution of PPepI data is dominated by substantial presence of polar (Tyr), hydrophobic (Leu, Ile) and charged (Arg) residues. Among the anchor residues in PPI, Leu is dominant followed by Arg, Tyr and Gln. However, in PPepI, highest frequencies were observed for only hydrophobic residues Leu, Ile, Val and Phe.

Out of 3732 hotspots, only 5.8% (218) belong to the strong hotspot type. The strong hotspot type is completely dominated by Arg residue being the single most dominant residue in PPI occupying frequency of ˜42%. For PPepI category, Arg followed by Trp are the dominant residues, occupying frequencies of ˜26% and ˜20%, respectively (Fig. 4). Again for anchor residues, similar trend was observed in PPI with Arg predominantly present. For PPepI, Arg and Trp are preferred residues. Other than Arg, the bulky hydrophobic side chain of Trp also serves as suitable candidate for anchor residue in PPepI category.

Thus, going from the weak to the strong hotspot types, the PPI and PPepI categories tend to close the gap. In the weak type, differences are prominent with polar residues followed by hydrophobic and minor fraction of charged residues in PPI; hydrophobic followed by polar residues in PPepI category. Moving towards the moderate category, the nature of interactions shift towards the polar side in PPI with dominance of charged and polar residues. Hotspot nature in PPepI categories is represented by all three types of residues – polar, hydrophobic and charged. Finally, in the strong type, only Arg dominate the distribution in PPI, and in PPepI Arg as well as Trp are overwhelmingly present (Table 3).

3.3 Uniqueness of Arg, Tyr and Leu makes them favourite residues for interaction

Dividing the hotspots into a three different types has revealed the subtle characteristic patterns of hotspots. Arg is the clear choice among the strong types of hotspots and anchor residues in either of the protein-protein and protein-peptide interfaces. Why Arg is the highly preferred hotspot and anchor residue for interactions? The overwhelming presence of Arg residue as hotspot at the interacting interface stems from a multitude of facts [10]. The guanidinium group of Arg is one of the weakliest hydrated cations due to charge delocalization. This makes the Arg side chain easier to bury [46] as compared to other hydrophobic residues [47]. The anchor residue calculations suggest the predominant occurrence of Arg with ΔSASA (unbound – bound) area ranging from 101-205 $Å^2$, quantitatively indicating that it has the ability to substantially bury its surface. Buried Arg side chain has several advantages as compared to other residues. It is charged,

Posted on 14 Dec 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.167101052.24746048/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

extensively hydrogen-bonded (can donate up to 5 hydrogen bonds), having high pK, high flexibility to interact, and also possess the ability to interact by stacking with other planar side chain groups in proteins [46, 48, 49]. Thus among 5-charges residues, Arg was selected in nature to mediate protein-protein and protein-peptide interactions.

Similarly, examining the overall trend (Fig. 1 and Table 2), Tyr and Leu are highly preferred ones among polar and hydrophobic residues. These amino acids possess a striking balance of flexibility, rigidity and steric bulk. The amino acids having sufficient steric bulk and flexibility is required to generate a structurally plastic regions enabling the binding interface to mold itself to optimize complementarity. Additionally, the physicochemical properties of tyrosine i.e. amphipathic and bulky side chain, which is capable of forming nonpolar, hydrogen-bonding, cation-π and π-π stacking interactions, making it one of the most effective polar residues for mediating molecular recognition [50].

3.4 Multiple factors govern the interaction of hotspot residues

To examine the predominant forces governing the interaction for the hotspot residues, it would be worthwhile to correlate [?][?]G values of hotspot residues with respect to changes in various types of interactions upon mutation. Multiple linear regression analysis was carried out to examine the correlation between [?][?]G values with energy contributions from Coulomb, lipophilic, hydrogen bonding, van der Waal's interactions, packing desolvation, entropy and surface complementarity**.** Strong correlation exists (in the range of 0.9 -1, with p-value close to 0) between the binding affinity [?][?]G and its components ([?][?]G of Coulomb, lipophilic, hydrogen bonding, van der Waal's interactions and desolvation). Notably, these interactions contribute in equal proportion in the making of hotspot residues and no single factor predominantly governs over other in the hotspot characteristics across all three weak, moderate and strong types. As peptide has binding characteristic similar to protein-ligand interactions, some degree of correlation (˜0.1) was also observed between [?][?]G and its component [?][?]G (surface complementarity) in PPepI.

CONCLUSION

Peptides and peptidomimetics are straightforward alternatives to protein-based biologics due to multiple advantages of larger shelf life, feasibility of oral delivery, flexibility of optimization, screening and synthesis. The knowledge of subtle differences between protein-protein and protein-peptide interactions should aid in the effective design of peptide-based biologics. In the present study, we have focussed on two important class of residues, namely, hotspot and anchor residues, to characterize differences between protein-protein and protein-peptide interactions.

Using implicit solvation-based free energy calculations, alanine scanning has been extensively performed on benchmarking datasets and hotspot and anchor residues were identified, which has revealed many interesting findings. The presence of sizable population (about 65%) of hotspot residues at the interface of the complex suggest that nature has remarkably optimized a great majority of the interface residues responsible for protein-protein interactions during evolution. It turned out that the differences in the two categories – PPI and PPepI are readily apparent, once we group the hotspot data into three distinct types, namely - weak hotspots (having binding free energy loss upon Ala mutation, ΔΔG in 2-10 kcal/mol range), moderate hotspots (ΔΔG, 10-20 kcal/mol) and strong hotspots (ΔΔG, 20 kcal/mol and higher). Correlation studies using MLR suggest that calculated free energy of binding of hotspot directly correlate with coulomb, lipophilic, hydrogen bonding, van der Waal interactions and desolvation penalty and no specific preference of any of the factor(s) over other was observed across all three types of hotspots. The analysis suggests that for PPI the preference is charged and polar followed by hydrophobic residues while for PPepI it is polar and hydrophobic followed by charged residues. In PPI, weak hotspots are predominantly populated by polar and hydrophobic residues. The distribution shifts towards charged and polar residues for moderate type, and charged residue (Arg) is overwhelmingly present in the strong type. In contrast, in the protein-peptide dataset, the distribution shifts from predominantly hydrophobic & polar (in the weak type) to more or less similar preference for polar, hydrophobic and charged residues and finally the charged residue (Arg) and Trp are mostly occupied in the strong type. Similar trend has been observed for anchor residues in both

Posted on 14 Dec 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.167101052.24746048/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

categories. The present work is an attempt to characterize and distinguish PPI and PPepI, focussing on two important class of residues. Further work is required to facilitate the discovery of new generations of peptide and peptidomimetic modulators. which can be utilized in the effective design of biologics.

Acknowledgement Kiran Kumar thanks UGC, India for PhD fellowship.

REFERENCES

1. Cukuroglu E, Engin HB, Gursoy A, Keskin O. Hot spots in protein-protein interfaces: towards drug discovery. *Progress in Biophysics and Molecular Biology* . 2014; 116(2-3): 165-173. doi:10.1016/j.pbiomolbio.2014.06.003

2. Jubb H, Blundell TL, Ascher DB. Flexibility and small pockets at protein-protein interfaces: New insights into druggability.*Progress in Biophysics and Molecular Biology* . 2015; 119(1): 2-9. doi:10.1016/j.pbiomolbio.2015.01.009

3. Junaid M, Li CD, Shah M, Khan A, Guo H, Wei DQ. Extraction of molecular features for the drug discovery targeting protein-protein interaction of Helicobacter pylori CagA and tumor suppressor protein ASSP2. *Proteins: Structure, Function and Bioinformatics* . 2019; 87(10): 837-849. doi:10.1002/prot.25748

4. Lanzarotti E, Defelipe LA, Marti MA, Turjanski AG. Aromatic clusters in protein-protein and protein-drug complexes. *Journal of Cheminformatics* . 2020; 12(1): 1-9. doi:10.1186/s13321-020-00437-4

5. Rosell M, Fernández-Recio J. Hot-spot analysis for drug discovery targeting protein-protein interactions. *Expert Opinion on Drug Discovery* . 2018; 13(4): 327-338. doi:10.1080/17460441.2018.1430763

6. Sarvagalla S, Cheung CHA, Tsai JY, Hsieh HP, Coumar MS. Disruption of protein-protein interactions: Hot spot detection, structure-based virtual screening and: In vitro testing for the anti-cancer drug target-survivin. *RSC Advances* . 2016; 6(38): 31947-31959. doi:10.1039/c5ra22927h

7. Srivastava M, Suri C, Singh M, Mathur R, Asthana S. Molecular dynamics simulation reveals the possible druggable hotspots of USP7.*Oncotarget* . 2018; 9(76): 34289-34305. doi:10.18632/oncotarget.26136

8. Wang, L., Wang, N., Zhang, W., Cheng X, Yan Z, Shao G, Wang X, Wang R, Fu C. Therapeutic peptides: current applications and future directions. *Sig Transduct Target Ther.* 2022. 7, 48. https://doi.org/10.1038/s41392-022-00904-4.

9. Clackson T, Wells JA. A hot spot of binding energy in a hormone-receptor interface. *Science* . 1995; 267(5196): 383-386. doi:10.1126/science.7529940

10. Bogan AA, Thorn KS. Anatomy of hot spots in protein interfaces.*J Mol Biol* . 1998; 280(1): 1-9. doi:10.1006/jmbi.1998.1843

11. Csizmok V, Follis AV, Kriwacki RW, Forman-Kay JD. Dynamic protein interaction networks and new structural paradigms in signaling.*Chem Rev* . 2016; 116(11): 6424-6462. doi:10.1021/acs.chemrev.5b00548

12. Rajamani D, Thiel S, Vajda S, Camacho CJ. Anchor residues in protein–protein interactions. *Proc Natl Acad Sci USA* . 2004; 101(31): 11287. doi:10.1073/pnas.0401942101

13. Meireles LMC, Dömling AS, Camacho CJ. ANCHOR: a web server and database for analysis of protein-protein interaction binding pockets for drug discovery. *Nucleic Acids Res* . 2010; 38(Web Server issue): W407-411. doi:10.1093/nar/gkq502

14. Sirin S, Apgar JR, Bennett EM, Keating AE. AB-Bind: Antibody binding mutational database for computational affinity predictions.*Protein Science* . 2016; 25(2): 393-409. doi:10.1002/pro.2829

15. Thorn KS, Bogan AA. ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions.*Bioinformatics* . 2001; 17(3): 284-285. doi:10.1093/bioinformatics/17.3.284

Posted on 14 Dec 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.167101052.24746048/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

16. Borrman T, Cimons J, Cosiano M, et al. ATLAS: A database linking binding affinities with structures for wild-type and mutant TCR-pMHC complexes. *Proteins* . 2017; 85(5): 908-916. doi:10.1002/prot.25260

17. Fischer TB, Arunachalam KV, Bailey D, et al. Binding interface database (BID): a compilation of amino acid hot spots in protein interfaces. *Bioinformatics* . 2003; 19(11): 1453–1454. https://doi.org/10.1093/bioinformatics/btg163

18. Geng C, Vangone A, Bonvin AMJJ. Exploring the interplay between experimental methods and the performance of predictors of binding affinity change upon mutations in protein complexes. *Protein Eng Des Sel* . 2016; 29(8): 291-299. doi:10.1093/protein/gzw020

19. Liu L, Xiong Y, Gao H, Wei D-Q, Mitchell JC, Zhu X. dbAMEPNI: a database of alanine mutagenic effects for protein–nucleic acid interactions. *Database (Oxford)* . 2018; 2018. doi:10.1093/database/bay034

20. Liu Q, Chen P, Wang B, Zhang J, Li J. dbMPIKT: a database of kinetic and thermodynamic mutant protein interactions. *BMC Bioinformatics* . 2018; 19(1): 455. doi:10.1186/s12859-018-2493-7

21. Jemimah S, Yugandhar K, Michael Gromiha M. PROXiMATE: a database of mutant protein-protein complex thermodynamics and kinetics.*Bioinformatics* . 2017; 33(17): 2787-2788. doi:10.1093/bioinformatics/btx312

22. Jankauskaite J, Jiménez-García B, Dapkunas J, Fernández-Recio J, Moal IH. SKEMPI 2.0: an updated benchmark of changes in protein-protein binding energy, kinetics and thermodynamics upon mutation.*Bioinformatics* . 2019; 35(3): 462-469. doi:10.1093/bioinformatics/bty635

23. Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J Mol Biol* . 2002; 320(2): 369-387. doi:10.1016/S0022-2836(02)00442-4

24. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. *Nucleic Acids Research* . 2005; 33(2): W382–W388. https://doi.org/10.1093/nar/gki387

25. Kortemme T, Baker D. A simple physical model for binding energy hot spots in protein–protein complexes. *Proc Natl Acad Sci USA* . 2002; 99(22): 14116-14121. doi:10.1073/pnas.202485799

26. Kortemme T, Kim DE, Baker D. Computational alanine scanning of protein-protein interfaces. *Sci STKE* . 2004; 2004(219): pl2. doi:10.1126/stke.2192004pl2

27. Barlow KA, Ó Conchúir S, Thompson S, Suresh P, Lucas JE, Heinonen M, Kortemme T. Flex ddG: Rosetta Ensemble-Based Estimation of Changes in Protein–Protein Binding Affinity upon Mutation. *J. Phys. Chem. B.* 2018; 122, 5389–5299. 10.1021/acs.jpcb.7b11367.

28. Dehouck Y.; Kwasigroch J. M.; Rooman M.; Gilis D. BeAtMuSiC: prediction of changes in protein–protein binding affinity on mutations.*Nucleic Acids Res* . 2013; 41, W333–W339. 10.1093/nar/gkt450.

29. Pires D. E. V.; Ascher D. B.; Blundell T. L. mCSM: predicting the effects of mutations in proteins using graph-based signatures.*Bioinformatics* 2014; 30, 335–342. 10.1093/bioinformatics/btt691.

30. Witvliet DK, Strokach A, Giraldo-Forero AF, Teyra J, Colak R, Kim PM. ELASPIC web-server: proteome-wide structure-based prediction of mutation effects on protein stability and binding affinity.*Bioinformatics* . 2016; 32(10): 1589-1591. doi:10.1093/bioinformatics/btw031

31. Geng C, Vangone A, Folkers GE, Xue LC, Bonvin AMJJ. iSEE: Interface structure, evolution, and energy-based machine learning predictor of binding affinity changes upon mutations. *Proteins* . 2019; 87(2): 110-119. doi:10.1002/prot.25630

32. Li M, Simonetti FL, Goncearenco A, Panchenko AR. MutaBind estimates and interprets the effects of sequence variants on protein–protein interactions. *Nucleic Acids Res* . 2016; 44(Web Server issue): W494-W501. doi:10.1093/nar/gkw374

Posted on 14 Dec 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.167101052.24746048/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.

33. Benedix A, Becker CM, de Groot BL, Caflisch A, Bockmann RA. Predicting free energy changes using structural ensembles. *Nature Methods* . 2009; 6(1): 3-4. doi:10.1038/nmeth0109-3

34. Brender JR, Zhang Y. Predicting the effect of mutations on protein-protein binding interactions through structure-based interface profiles. *PLOS Computational Biology* . 2015; 11(10): e1004494. doi:10.1371/journal.pcbi.1004494

35. Petukh M, Dai L, Alexov E. SAAMBE: Webserver to predict the charge of binding free energy caused by amino acids mutations. *Int J Mol Sci* . 2016; 17(4). doi:10.3390/ijms17040547

36. Rosell M, Fernandez-Recio J. Hot-spot analysis for drug discovery targeting protein-protein interactions. *Expert Opin Drug Discov* . 2018;13(4):327-338. doi:10.1080/17460441.2018.1430763

37. Beard H, Cholleti A, Pearlman D, Sherman W, Loving KA. Applying physics-based scoring to calculate free energies of binding for single amino acid mutations in protein-protein complexes. *Plos One* . 2013; 8(12): e82849. doi:10.1371/journal.pone.0082849

38. Schrodinger Release 2019–3, Maestro, Protein Preparation Wizard, Prime, MM-GBSA, Schrodinger, LLC, New York, NY, 2020.

39. Vreven T, Moal IH, Vangone A, et al. Updates to the Integrated Protein-Protein Interaction Benchmarks: Docking Benchmark Version 5 and Affinity Benchmark Version 2. *J Mol Biol* . 2015;427(19):3031-3041. doi:10.1016/j.jmb.2015.07.016

40. Hauser AS, Windshugel B. LEADS-PEP: A Benchmark Data Set for Assessment of Peptide Docking Performance. *J ChemInf Model* . 2016;56(1):188-200. doi:10.1021/acs.jcim.5b00234

41 . Xiang Z, Honig B. Extending the accuracy limits of prediction for side-chain conformations. *J Mol Biol* . 2001;311(2):421-430. doi:10.1006/jmbi.2001.4865

42. Reddy MR, Reddy CR, Rathore RS, et al. Free energy calculations to estimate ligand-binding affinities in structure-based drug design.*Curr Pharm Des* . 2014; 20(20): 3323-3337. doi:10.2174/13816128113199990604

43. Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities .*Expert Opin Drug Discov* . 2015;10(5):449-461. doi:10.1517/17460441.2015.1032936

44.Sirin S, Pearlman DA, Sherman W. Physics-based enzyme design: predicting binding affinity and catalytic activity. *Proteins* . 2014;82(12):3397-3409. doi:10.1002/prot.24694

45. Negron C, Pearlman DA, Del Angel G. Predicting mutations deleterious to function in beta-lactamase TEM1 using MM-GBSA. *PLoS ONE* . 2019;14(3):e0214015. doi:10.1371/journal.pone.0214015

46. Mason PE, Neilson GW, Dempsey CE, Barnes AC, Cruickshank JM. The hydration structure of guanidinium and thiocyanate ions: implications for protein stability in aqueous solution. *Proc Natl Acad Sci USA.* 2003; 100(8): 4557-61. doi: 10.1073/pnas.0735920100.

47. Samanta U, Bahadur RP, Chakrabarti P. Quantifying the accessible surface area of protein residues in their local environment.*Protein Eng.* 2002; 15(8): 659-67. doi: 10.1093/protein/15.8.659.

48. Pace CN, Grimsley GR, Scholtz JM. Protein ionizable groups: pK values and their contribution to protein stability and solubility.*J Biol Chem.* 2009; 284(20): 13285-9. doi: 10.1074/jbc.R800080200.

49. Chakrabarti P, Bhattacharyya R. Geometry of nonbonded interactions involving planar groups in proteins. *Prog Biophys Mol Biol.* 2007; 95(1-3): 83-137. doi: 10.1016/j.pbiomolbio.2007.03.016.

50. Koide S, Sidhu SS. The importance of being tyrosine: lessons in molecular recognition from minimalist synthetic binding proteins. ACS Chem Biol. 2009; 4(5): 325-34. doi: 10.1021/cb800314v.

51. 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(5): 1281-94. doi:

10.1016/j.jmb.2004.10.077.

Hosted file

Figure1.docx available at https://authorea.com/users/566224/articles/612975-three-types-of-hotspots-weak-moderate-and-strong-in-protein-protein-and-protein-peptide-interactions

Hosted file

Figure2.docx available at https://authorea.com/users/566224/articles/612975-three-types-of-hotspots-weak-moderate-and-strong-in-protein-protein-and-protein-peptide-interactions

Hosted file

Figure3.docx available at https://authorea.com/users/566224/articles/612975-three-types-of-hotspots-weak-moderate-and-strong-in-protein-protein-and-protein-peptide-interactions

Hosted file

Figure4.docx available at https://authorea.com/users/566224/articles/612975-three-types-of-hotspots-weak-moderate-and-strong-in-protein-protein-and-protein-peptide-interactions

Hosted file

Table1.docx available at https://authorea.com/users/566224/articles/612975-three-types-of-hotspots-weak-moderate-and-strong-in-protein-protein-and-protein-peptide-interactions

Hosted file

Table2.docx available at https://authorea.com/users/566224/articles/612975-three-types-of-hotspots-weak-moderate-and-strong-in-protein-protein-and-protein-peptide-interactions

Hosted file

Table3.docx available at https://authorea.com/users/566224/articles/612975-three-types-of-hotspots-weak-moderate-and-strong-in-protein-protein-and-protein-peptide-interactions

Posted on 14 Dec 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/au.167101052.24746048/v1 — This a preprint and has not been peer reviewed. Data may be preliminary.