Investigating the Role of Pharmacophore sites in Protein-Ligand Binding

MohammadSadeq Mottaqi[1]

[1]*Ph.D. Program in Biochemistry, The Graduate Center, The City University of New York, New York, NY 10016, USA*

Abstract

Protein-ligand binding is a fundamental process in biochemistry that underlies numerous biological functions and serves as a key target for drug discovery. This study investigates the role of pharmacophore sites in protein-ligand binding, leveraging computational tools and statistical methods to quantify and characterize the interactions mediated by different types of pharmacophore sites across a diverse dataset of protein-ligand complexes from the Protein Data Bank (PDB). By categorizing atoms as hydrogen bond donors, hydrogen bond acceptors, or hydrophobic regions, and identifying potential interaction partners based on distance criteria, the study quantified the interaction frequencies for each pharmacophore site. Statistical analysis using the Z-score method revealed outlier sites with significantly higher interaction frequencies ($p < 0.05$), highlighting key pharmacophoric features that drive ligand binding and stability. The quantitative interaction matrix generated serves as a valuable resource for predicting ligand binding behavior and informing the rational design of novel therapeutics with enhanced efficacy and specificity. The integration of quantitative and visual analyses offers a comprehensive understanding of the complex protein-ligand binding dynamics. This study advances the theoretical understanding of ligand-protein interactions and provides a targeted approach for drug design, with potential implications for the development of more effective and selective therapeutics.

Keywords: Rational drug design; Protein-ligand binding; Pharmacophore sites; Binding affinity

1. Introduction

The interaction between ligands and proteins is a fundamental aspect of biochemistry, underpinning numerous biological processes and serving as a key target for pharmaceutical interventions. The binding affinity between a drug molecule (ligand) and its target protein is a crucial factor in determining the efficacy and specificity of the drug. A high binding affinity indicates that the drug binds strongly to its target, potentially leading to greater therapeutic effects at lower doses. Conversely, a low binding affinity may result in reduced efficacy and increased off-target interactions, which can lead to adverse side effects. Therefore, understanding the factors that influence ligand-protein binding affinity is of paramount importance in drug discovery and development (Nero et al., 2018).

Central to the ligand-protein interaction are pharmacophore sites, which are specific structural features within ligands that facilitate their binding to protein receptors. Pharmacophore sites can be categorized into three main types: hydrogen bond donors, hydrogen bond acceptors, and hydrophobic (greasy) regions (Ferreira et al., 2015). The arrangement and chemical properties of these sites play a crucial role in determining the affinity and specificity of ligand-protein binding. By identifying and optimizing the pharmacophore sites within a ligand, researchers can design molecules with enhanced binding affinity and selectivity for their target proteins (Giordano et al., 2022).

Despite the recognized importance of pharmacophore sites in drug discovery, there remains a need for a more comprehensive and quantitative understanding of their role in ligand-protein interactions. Previous studies have largely focused on individual pharmacophore sites or specific ligand-protein complexes, providing valuable but limited insights. A broader, systematic analysis of pharmacophore site interactions across a diverse range of proteins and ligands would greatly enhance our understanding of the underlying principles governing ligand binding and inform the rational design of novel drug molecules.

Moreover, recent advancements in computational tools and the availability of high-resolution structural data from sources like the Protein Data Bank (PDB) have opened up new opportunities for in-depth exploration of pharmacophore site interactions. By leveraging these resources and applying rigorous statistical methods, we can identify key patterns and trends that may have been overlooked in previous studies.

In this study, we aim to investigate the role of pharmacophore sites in ligand-protein binding, with a focus on the implications for drug discovery. Our approach involves a systematic analysis of a diverse dataset of protein-ligand complexes, utilizing computational tools and statistical methods to quantify and characterize the interactions mediated by different types of pharmacophore sites. By identifying outlier sites with significantly higher interaction frequencies, we seek to pinpoint the key pharmacophoric features that drive ligand binding and stability. These insights can then be applied to the design of novel drug molecules with improved binding affinity and specificity for their target proteins. By elucidating the critical conserved pharmacophore sites and their interaction patterns, we aim to inform the rational design of novel therapeutics with enhanced efficacy and reduced side effects.

2. Results

2.1. Data Overview and Categorization

The dataset used in this study, obtained from the Protein Data Bank (PDB), consisted of 8 proteins, each with a query file containing pharmacophore sites and several corresponding ligands. The total number of ligands varied for each protein. Table 1 summarizes the dataset statistics. Atoms within the proteins' pharmacophore query files were categorized as 'Acceptor', 'Donor', or 'Grease' based on the labels provided in the file. This categorization laid the foundation for subsequent analysis steps tailored to each type of pharmacophoric feature.

Protein	Number of ligands
Aldehyde dehydrogenase 1A3 (ALDR)	122
Urokinase plasminogen activator surface receptor (UROK)	97
Proto-oncogene tyrosine-protein kinase (SRC)	72
RAC-alpha serine/threonine-protein kinase (AKT1)	71
Hepatocyte growth factor receptor (MET)	69
Peroxisome proliferator-activated receptor alpha (PPARA)	53
Mobile colistin resistance protein (MCR)	27
Wee1 protein kinase (WEE1)	20

* **Table 1.** Summary statistics of the datasets used throughout the study

2.2. Identification and Quantification of Potential Interactions

Using PyMOL (Schrödinger LLC, 2015) and Python scripting, potential interaction partners were identified for each pharmacophoric category based on their chemical properties and spatial configuration. The criteria for selecting donor, acceptor, and hydrophobic atoms were as follows:

Donor atoms: Atoms that could act as hydrogen bond donors were identified based on their ability to donate a hydrogen atom. These typically included atoms like nitrogen (N) or oxygen (O) that had attached

hydrogen atoms. For example, in amino groups (-NH2), the nitrogen atom with its bonded hydrogen atoms was considered a potential hydrogen bond donor.

Acceptor atoms: Atoms that could serve as hydrogen bond acceptors were identified based on their ability to accept a hydrogen bond. These generally involved electronegative atoms like oxygen (O) or nitrogen (N) that had lone pairs of electrons available for bonding. For instance, in carbonyl groups (C=O), the oxygen atom with its lone pairs was considered a potential hydrogen bond acceptor.

Hydrophobic atoms: Atoms contributing to hydrophobic interactions were identified based on their non-polar character and lack of polar bonding capabilities. These typically included carbon atoms (C) in hydrophobic chains or rings not connected to any nitrogen, oxygen, or fluorine atoms. For example, the carbon atoms in an alkyl chain or a benzene ring were considered as potential hydrophobic interaction sites. By defining these criteria, the study aimed to provide a clear and consistent approach for identifying potential interaction partners within the ligands. This allowed for a systematic analysis of the role of pharmacophores in ligand-protein binding.

The quantification of potential interactions was performed by calculating the distances between atoms. For hydrogen bonds, donor-acceptor distances within 3.2 Å were considered potential interactions. This distance cutoff was chosen based on the typical range of hydrogen bond lengths observed in biological systems. Hydrophobic interactions were assessed based on distances less than 4 Å between non-polar groups, as this range encompasses the van der Waals radii of the interacting atoms and allows for effective hydrophobic contacts.

The number of potential interaction partners within these distance thresholds was recorded for each atom, representing the count of potential hydrogen bonds or hydrophobic contacts. This quantitative approach provided a measure of the interaction frequencies for each pharmacophore site, enabling the identification of sites with significantly higher interaction potential compared to others within the same category. Supplementary table 1 includes the total interaction frequencies for all the proteins in the dataset.

2.3. Statistical Analysis of Interaction Frequencies

To identify outlier pharmacophore sites with significantly higher interaction frequencies compared to other sites within the same category, the Z-score method was applied. This non-parametric statistical approach was chosen due to the skewed distribution of interaction counts, which is common in biological data. The Z-score, also known as the standard score, measures how many standard deviations an observation or data point is from the mean of the dataset. The analysis revealed several pharmacophore sites that exhibited significantly higher interaction frequencies ($p < 0.05$) compared to their peers. These outliers were found across two pharmacophoric categories: acceptors and grease atoms.

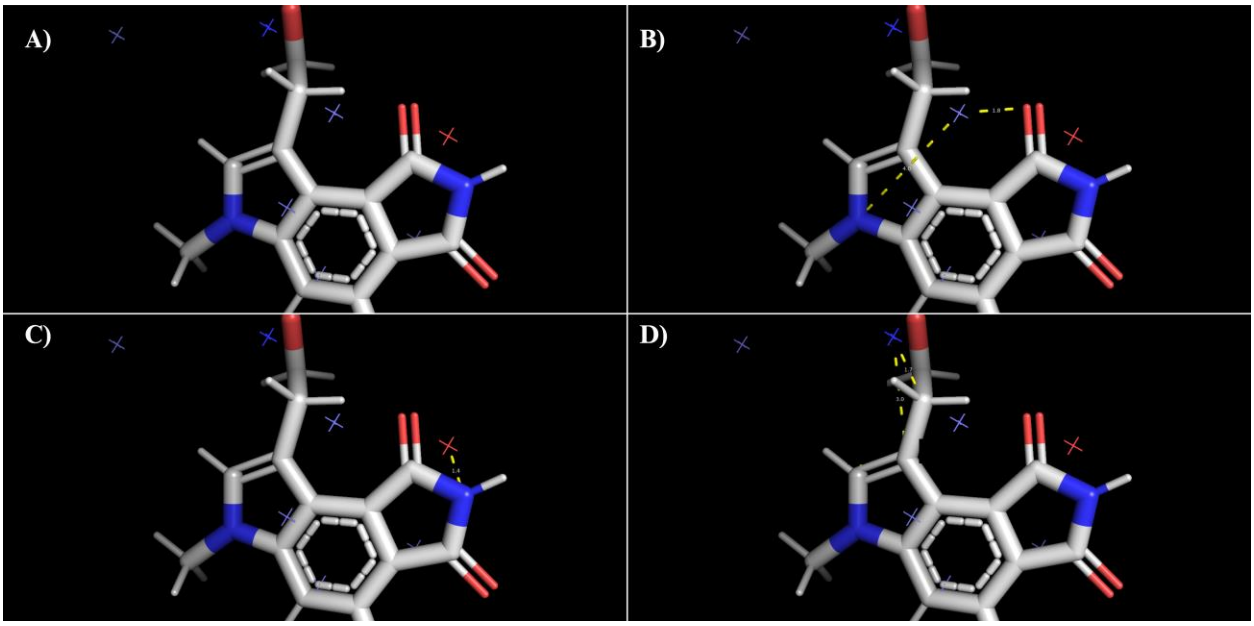
Notably, a centrally located donor pharmacophore site demonstrated an exceptionally high number of potential hydrogen bonds, suggesting its crucial role in ligand recognition and binding, and stabilizing ligand-protein complexes. Among the grease atoms, specific pharmacophore sites were identified as key contributors to hydrophobic interactions, further emphasizing their significance in the binding process. Table 2 presents the results of the statistical test.

Protein	Pharmacophore	Site's Label
Urokinase plasminogen activator surface receptor (UROK)	Atom 3	Grease
Hepatocyte growth factor receptor (MET)	Atom 4	Acceptor
	Atom 7	Grease
	Atom 8	Grease
Peroxisome proliferator-activated receptor alpha (PPARA)	Atom 7	Grease
	Atom 8	Grease
Mobile colistin resistance protein (MCR)	Atom 4	Grease
Wee1 protein kinase (WEE1)	Atom 5	Grease

* **Table 2.** Outlier pharmacophore sites identified by the Z-score method. The table lists the site names, their corresponding labels (Grease or Acceptor), and their classification as outliers based on the statistical analysis of interaction frequencies.

2.4. Visual Representation of Key Pharmacophore Sites

The visual representation of pharmacophore sites within the protein-ligand binding pockets is crucial for understanding the spatial arrangement and interaction dynamics. Figure 1 illustrates the pharmacophore sites and their interactions for the urokinase plasminogen activator surface receptor (UROK) and one of its ligands. Panel A displays the pharmacophore sites, while panels B, C, and D highlight the interactions of donor, acceptor, and hydrophobic (greasy) sites, respectively. These visualizations corroborate our quantitative findings, providing a depiction of how different pharmacophore sites contribute to the ligand binding process.



* **Figure 1.** Visualization of pharmacophore sites and their interactions for UROK; (A) Pharmacophores for UROK and one of its ligands. (B) Pharmacophore donor sites and their potential interactions. (C) Pharmacophore acceptor sites and their interactions. (D) Hydrophobic (greasy) site and its interactions. The yellow dashed lines in panels B and D represent potential hydrogen bonds and hydrophobic interactions, respectively.

3. Methods

3.1. Data Collection

The dataset for this study was meticulously selected from the Protein Data Bank (PDB) harmonized by Ligand Expo database (Feng et al., 2004), focusing on complexes characterized by clear definitions of ligand-protein interactions where pharmacophore sites with significant contribution in ligand binding mechanisms are specified. The dataset contained files for 8 proteins, each having one query file containing pharmacophore sites and several ligands varied for each protein. Table 1 summarizes the dataset statistics.

3.2. Computational Tools and Software

The analysis leveraged PyMOL (version 2.5) (Schrödinger LLC, 2015), a molecular visualization system, to inspect and manipulate the structures of protein-ligand complexes. Additionally, statistical analysis was conducted using Python (version 3.11), particularly utilizing libraries such as Pandas and Numpy for data manipulation and SciPy for statistical testing.

3.3. Analysis Procedure

The analysis involved a two-phase approach. The first phase, conducted in PyMOL, focused on the quantification of the number of bonds and the creation of a data frame. The second phase, carried out in Python, aimed to identify outlier pharmacophore sites based on their interaction frequency.

3.3.1. Categorization of Atoms

Each atom in the dataset was categorized based on labels provided in the 'Label' column of the data file. These labels indicated whether an atom functioned as an 'Acceptor', 'Donor', or was part of a 'Grease' group, which typically indicates involvement in hydrophobic interactions. This categorical information was crucial for directing subsequent analysis steps specific to each type of pharmacophoric feature. Acceptors were targeted for analysis concerning their ability to form hydrogen bonds with donors. Donors were analyzed based on their potential to engage acceptors in hydrogen bonds. Grease atoms were evaluated for their role in hydrophobic interactions.

3.3.2. Identification of Interaction Partners

For each pharmacophoric category, potential interaction partners were identified using structural criteria established through molecular visualization and analysis tools.

Donor and acceptor atoms identification: Within the ligand-protein complexes, atoms that could act as hydrogen bond donors or acceptors were identified based on their chemical properties and spatial configuration. Donors typically included atoms like nitrogen or oxygen that had attached hydrogen atoms, indicating their ability to donate a hydrogen bond. Acceptors generally involved electronegative atoms like oxygen or nitrogen capable of receiving a hydrogen bond.

Hydrophobic interaction partners: Atoms contributing to hydrophobic interactions were identified based on their non-polar character and lack of polar bonding capabilities. These typically included carbon atoms in hydrophobic chains or rings which are not connected to any nitrogen, oxygen, or fluor atoms

3.3.3. Measurement of Potential Interactions

Once potential donors, acceptors, and hydrophobic partners were identified, the next step was to quantify the interaction opportunities for each categorized atom. The primary method for determining the existence of a potential interaction involved calculating the distance between atoms using PyMOL and Python scripting. For hydrogen bonds, the distance between the donor hydrogen and the acceptor atom needed to be within a typical hydrogen bond length (< 3.2 Å). For hydrophobic interactions, distances were assessed to determine if non-polar groups were within effective proximity (< 4 Å) conducive to van der Waals interactions. For each atom, the number of potential interaction partners within the specified distance threshold was counted. This count represented the number of potential hydrogen bonds for donors and acceptors, or potential hydrophobic contacts for grease atoms.

3.3.4. Recording Results

Results from these measurements were compiled systematically. Each atom's detail, its category, the count of potential interactions (hydrogen bonds or hydrophobic contacts), and the atom's specific partners were recorded for further statistical analysis.

3.4. Statistical Methods

Due to the non-parametric nature of the data, with skewed distributions expected in interaction counts, the Z-score method was used to identify outlier pharmacophore sites. This method is particularly suited for detecting data points that significantly deviate from the mean, making it useful for identifying pharmacophore sites with unusually high interaction frequencies compared to other sites within the same category.

The Z-score, also known as the standard score, measures how many standard deviations an observation or data point is from the mean of the dataset. It is calculated by subtracting the population mean from an individual raw score and then dividing the difference by the population standard deviation. Mathematically, it can be represented as:

$$z = (X - \mu) / \sigma$$

Where Z is the Z-score, X is the individual data point (interaction frequency), μ is the population mean, and σ is the population standard deviation.

In the context of this study, the Z-score was used to compare the interaction frequency of each pharmacophore site to the mean interaction frequency within its respective category (donor, acceptor, or grease). The magnitude of the Z-score represents the distance between the individual data point and the mean in terms of standard deviations. Pharmacophore sites with Z-scores greater than a certain threshold ($Z > 1.5$) were considered outliers, indicating that their interaction frequencies were significantly higher than the majority of the sites within the same category.

4. Discussion

The present study provides a comprehensive investigation into the role of pharmacophore sites in protein-ligand binding, offering valuable insights that have both theoretical and practical implications. By leveraging computational tools and statistical methods, we have quantified and characterized the interactions mediated by different types of pharmacophore sites across a diverse dataset of protein-ligand complexes. The identification of outlier sites with significantly higher interaction frequencies highlights the key pharmacophoric features that drive ligand binding and stability.

From a theoretical perspective, our findings contribute to a deeper understanding of the fundamental principles governing ligand-protein interactions. The quantitative interaction matrix generated provides a framework for predicting ligand binding behavior based on the arrangement and chemical properties of pharmacophore sites. This aligns with previous studies that have emphasized the importance of considering the three-dimensional structure and chemical features of ligands in understanding their binding affinity and specificity (Kairys et al., 2024; Xu et al., 2017). Our results further extend this concept by demonstrating the critical role of specific pharmacophore sites in driving these interactions.

The identification of outlier pharmacophore sites with exceptionally high interaction frequencies has significant implications for drug discovery and development. By pinpointing the key pharmacophoric features that contribute to strong and stable ligand binding, our study provides a targeted approach for designing novel drug molecules with improved efficacy and reduced side effects. This is consistent with the findings of recent studies that have successfully applied pharmacophore-based approaches to identify potent and selective inhibitors of various therapeutic targets (Correia et al., 2022). Our results suggest that focusing on the optimization of specific pharmacophore sites, rather than the ligand as a whole, could lead to more efficient and effective drug design strategies.

The visual representations of the key pharmacophore sites within the protein binding pockets provide valuable insights into the spatial arrangement and interaction networks that facilitate ligand binding. These visualizations complement the quantitative findings and offer a more intuitive understanding of the complex dynamics at play.

The limitations of this study are as follows; (1) While our dataset included a diverse range of protein-ligand complexes, it is possible that the findings may not be generalizable to all protein families or ligand classes. Future studies could expand the dataset to include a broader range of proteins and ligands, allowing for a more comprehensive assessment of the role of pharmacophore sites across different biological systems. Additionally, (2) the use of more advanced statistical methods, such as machine learning algorithms, could potentially uncover more complex patterns and relationships between pharmacophore sites and ligand binding behavior.

Another avenue for future research is the integration of our findings with experimental studies to validate the identified key pharmacophore sites and their predicted interactions. This could involve targeted mutagenesis of specific pharmacophore sites to assess their impact on ligand binding affinity and specificity, or the synthesis of novel ligands designed to optimize interactions with the identified key sites. Such experimental validation would strengthen the practical applications of our findings and provide a more robust foundation for rational drug design.

5. Conclusion

This study provides a robust and comprehensive analysis of the role of pharmacophore sites in ligand-protein binding, offering critical insights that enhance our understanding of drug discovery mechanisms. By utilizing advanced computational tools and statistical methods, we systematically characterized the interactions mediated by different types of pharmacophore sites across a diverse dataset of protein-ligand complexes. Our identification of key pharmacophore sites with significantly higher interaction frequencies underscores their pivotal role in driving ligand binding and stability, highlighting specific features that can be targeted to optimize drug efficacy and specificity.

From a theoretical standpoint, our findings advance the foundational knowledge of ligand-protein interactions, emphasizing the importance of spatial arrangement and chemical properties in determining binding affinity. The quantitative interaction matrix developed in this study serves as a valuable predictive framework for ligand binding behavior, supporting the rational design of novel therapeutics. Practically, the insights gained from this study have implications for the pharmaceutical industry. By pinpointing specific pharmacophore sites that contribute to strong and stable ligand binding, we offer a targeted approach for drug design, potentially leading to the development of more effective and selective therapeutics. The integration of quantitative data with visual representations of binding interactions enhances our understanding of the complex dynamics at play, providing a comprehensive toolset for structure-based drug design. Future research can build on these findings by expanding the dataset and incorporating experimental validation, further bridging the gap between computational predictions and real-world applications in drug discovery.

Availability of data and materials

The datasets used in this study can be downloaded from this link. The code used for the analysis is available on GitHub at: https://github.com/mmottaqii/PyMol_conserved_pharmacophore_sites.

References

Correia, L. C., Ferreira, J. V, de Lima, H. B., Silva, G. M., da Silva, C. H. T. P., de Molfetta, F. A., & Hage-Melim, L. I. S. (2022). Pharmacophore-based virtual screening from phytocannabinoids as antagonist r-CB1. *Journal of Molecular Modeling*, *28*(9), 258. https://doi.org/10.1007/s00894-022-05219-3

Feng, Z., Chen, L., Maddula, H., Akcan, O., Oughtred, R., Berman, H. M., & Westbrook, J. (2004). Ligand Depot: a data warehouse for ligands bound to macromolecules. *Bioinformatics*, *20*(13), 2153–2155. https://doi.org/10.1093/bioinformatics/bth214

Ferreira, L. G., Dos Santos, R. N., Oliva, G., & Andricopulo, A. D. (2015). Molecular Docking and Structure-Based Drug Design Strategies. *Molecules*, *20*(7), 13384–13421. https://doi.org/10.3390/molecules200713384

Giordano, D., Biancaniello, C., Argenio, M. A., & Facchiano, A. (2022). Drug Design by Pharmacophore and Virtual Screening Approach. *Pharmaceuticals*, *15*(5). https://doi.org/10.3390/ph15050646

Kairys, V., Baranauskiene, L., Kazlauskiene, M., Zubrienė, A., Petrauskas, V., Matulis, D., & Kazlauskas, E. (2024). Recent advances in computational and experimental protein-ligand affinity determination techniques. *Expert Opinion on Drug Discovery*, *19*(6), 649–670. https://doi.org/10.1080/17460441.2024.2349169

Nero, T. L., Parker, M. W., & Morton, C. J. (2018). Protein structure and computational drug discovery. *Biochemical Society Transactions*, *46*(5), 1367–1379. https://doi.org/10.1042/BST20180202

Schrödinger LLC. (2015). *The PyMOL Molecular Graphics System, Version~1.8*.

Xu, X., Yan, C., & Zou, X. (2017). Improving binding mode and binding affinity predictions of docking by ligand-based search of protein conformations: evaluation in D3R grand challenge 2015. *Journal of Computer-Aided Molecular Design*, *31*(8), 689–699. https://doi.org/10.1007/s10822-017-0038-1