FlexVASP-S: Flexible Steric-Aware 3D Geometric Framework for Explaining Protein Binding Specificity

Yangying Liu[1] **Houliang Zhou**[1] **Lifang He**[1] **Brian Y. Chen**[1*]

[1]Lehigh University

byc@lehigh.edu

Abstract

The design of selective ligands requires the discovery of binding site properties that distinguish targets from similar off-targets. From a steric perspective, this problem is exacerbated by the intrinsic flexibility of proteins, which can hide variations in shape that could be exploited for selective design. To address this problem, we propose FlexVASP-S, the first interpretable 3D geometic learning framework that identifies regions of hindrance in binding cavities that select for different ligands. FlexVASP-S identifies these regions despite noisy conformational variations by using flexible structural representations and explainable AI techniques. We tested FlexVASP-S on two superfamilies of proteins that have subfamilies with different ligand preferences, producing conformational variations of every protein with molecular dynamics simulations. Even though substantial motion occurs, FlexVASP-S identified regions that accommodate ligands preferred by one subfamily and not by others, recapitulating established experimental findings. It also accurately classified cavities with similar binding preferences. The code is publicly shared at https://github.com/LehighInfolab/FlexVASP-S.

1 Introduction

Steric hindrance is a major design concern for small-molecule ligands that selectively bind protein targets. At protein binding sites, where shape complementarity is most critical, assessing the influence of steric hindrance is challenged by conformational flexibility [1, 2, 3, 4, 5, 6]. Backbone and sidechain motions can conceal the boundaries of a cavity that a ligand might occupy, and they can obscure differences in cavities that drive different binding preferences [7, 8, 9], even if the ligand itself induces no special fit. In such cases, finding differentially impinged regions inside the binding sites of subtly different proteins would yield insights that are useful for designing ligand selectivity.

To address this challenge, we propose FlexVASP-S, the first tool for automatically explaining differences in steric hindrance that control ligand binding preferences. FlexVASP-S uses voxelized probabilistic representations of binding cavities called "frequent regions" [9], which describe subregions of cavities that are frequently solvent accessible, based on sampled timesteps from molecular dynamics (MD) simulations. The result is a representation of cavity shape that both reflects the motion seen in the simulation and augments the dataset. The frequent regions are passed to a novel 3D-CNN to classify protein subfamilies, and to Grad-CAM++ [10], an interpretation technique that identifies voxels that are salient for classification. This framework enables us to test the novel hypothesis that salient voxels actually delineate regions of differential steric hindrance that cause proteins to select different ligands. In our results, FlexVASP-S detected subtle variations in cavity shape that cause differences in ligand preferences that were observed in established experimental findings. These findings point to applications in refining small molecule design for selective therapy.

2 Methods

2.1 Representing Binding Sites with Frequent Regions

Dataset overview and conformational sampling. To test how FlexVASP-S can distinguish cavities with different ligand preferences, we selected the Enolase and the Serine Protease superfamilies, which
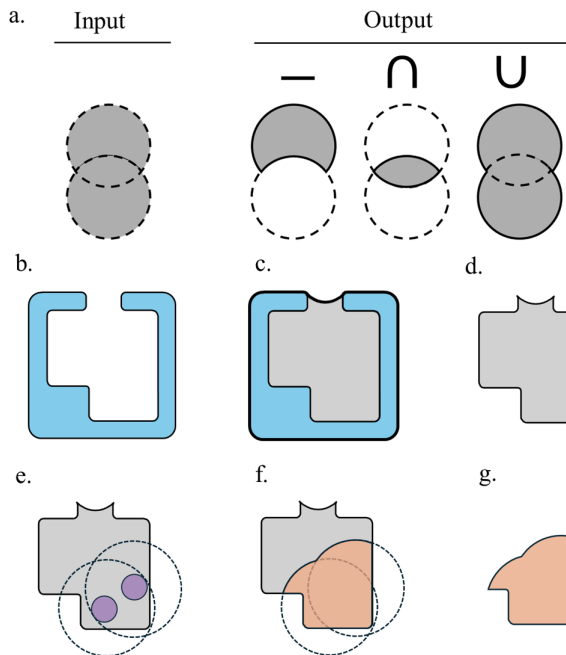
Figure 1: Representing Binding sites with three dimensional solids

(a) We manipulate geometric solids using CSG (Constructive Solid Geometry) operations, which take input solids (gray, dotted outline) and compute unions ($\cup$), intersections ($\cap$) and differences ($-$) as outputs (gray, solid outline).

(b) We begin with a protein structure (blue) defined by its molecular surface (solid outline).

(c) An envelope surface (thick outline) is computed using the standard rolling probe algorithm[11].

(d) The binding region is the CSG difference: envelope surface $-$ molecular surface (grey).

(e) The binding neighborhood is the union of 5Å radius spheres centered on ligand atoms (dotted).

(f) Binding site = Binding region ($\cap$) Binding neighborhood (orange).

(g) Output binding site (orange).

each have three subfamilies with different ligand preferences that are created in part by differences in steric hindrance. To eliminate redundancy, one member of any pair of proteins with sequence identity greater than 90% was removed, resulting in 16 Enolase and 19 Serine Protease structures, but the Enolases and the Serine Proteases have imbalanced subfamilies. Mandelate Racemases contain fewer proteins than the other Enolases, and Elastases have fewer proteins than other Serine Protease subfamilies. To mitigate subfamily imbalance, we upsample minority subfamilies by sampling them more frequently and downsample majority subfamilies by randomly discarding examples, so that all classes contribute more evenly during training. This level of manual curation and mechanism annotation is not yet scalable to dozens of superfamilies, so we focused on two systems with detailed experimental characterization to ensure high-quality labels and interpretable case studies. For each protein, we used GROMACS [12] to perform a 100-ns all-atom explicit solvent MD simulation to produce conformational samples. Datasets and simulations are detailed in Section A.2. For each protein, 1000 were selected uniformly and aligned to the original Protein DataBank (PDB) structure, which was first aligned to a pivot protein for the entire superfamily. The serine proteases were pivoted on Bovine chymotrypsin (PDB: 8GCH) and the enolase superfamily was pivoted on mandelate racemase from *Pseudomonas putida* (PDB: 1MDR), which were selected because they have co-crystallized ligands. Our simulations include only the protein and nearby waters, so conditions causing induced fit are not explicitly represented. Finally, pClay [13] was used to compute constructive solid geometry (CSG) operations (unions, intersections and differences) to generate a volumetric representation for the binding site (summarized in Fig. 1).

Generating frequent regions. Frequent regions were developed for FAVA [9] in earlier work, but only used for classification and analyzing amino acids. Here, and in Section A.4, we paraphrase the method for generating frequent regions. A frequent region is defined as a spatial zone that is solvent-accessible in more than k out of N conformational samples of a binding site. By setting k, the user excludes, from the shape represented, the rare gaps or clefts that occur at frequencies less than k/N as being conformational noise. Here, our simulations provide $N = 1000$ conformational samples (details are described in Appendix A.4), and we tested k in the range {15, 25, 50, 70, 100, 150, 200 } as a hyperparameter.

Frequent region voxelization. To represent the frequent regions as a vector, we embed them into a consistent lattice, discretizing it into voxels of side length $\Delta = 0.5\,\text{Å}$, yielding voxel tensors (e.g., $\mathcal{T}_{\text{Enolase}} \in \mathbb{R}^{28 \times 34 \times 30}$, $\mathcal{T}_{\text{Serprot}} \in \mathbb{R}^{34 \times 38 \times 42}$). Using constructive solid geometry (CSG) to compute the intersection between each voxel and the frequent region, each voxel value v_{ijk} is computed as the fraction of frequent region volume contained in the voxel (details in A.5). This calculation forms a tensor $\mathcal{T} = [v_{ijk}] \in \mathbb{R}^{N_x \times N_y \times N_z}$ which serves as input to the our 3D CNN.
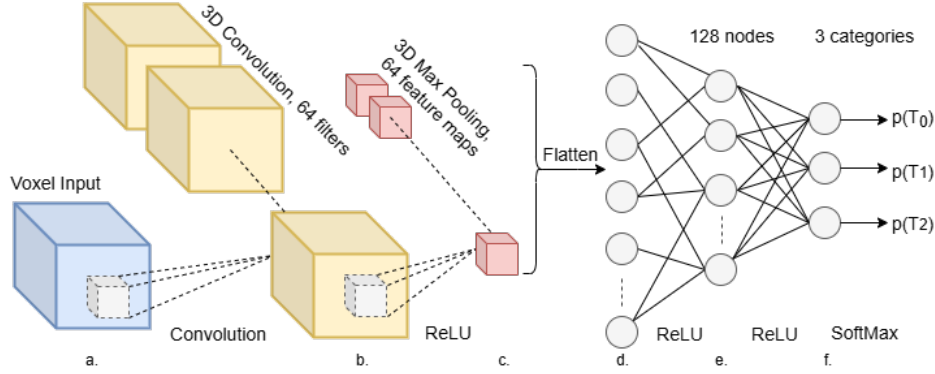
Figure 2: 3D-CNN Architecture. **(a)** Voxel inputs with each of the three dimensions (blue). **(b)** 3D convolutional layers with 64 filters. A ReLU activation function is used in both layers. **(c)** 3D max pooling layer producing outputs of size (x/2, y/2, z/2) with ReLU activation function. **(d)** Flattening layer with ReLU activation function. **(e)** Fully connected layer reducing layer d to 128 notes, ReLU activation layer. **(f)** Output layer with 3 nodes corresponding to the number of classes with softmax activation function, to produce probabilities of classification

2.2 A 3D Geometric CNN for Classifying Frequent Regions

The input is a single-channel 3D tensor $x \in \mathbb{R}^{1 \times N_x \times N_y \times N_z}$, corresponding to the voxelized representation of a protein cavity ($\mathcal{T}_{\text{Enolase}} \in \mathbb{R}^{28 \times 34 \times 30}$, $\mathcal{T}_{\text{Serprot}} \in \mathbb{R}^{34 \times 38 \times 42}$). This architecture is shown in Fig. 2. First, a 3D convolutional layer with 64 filters of size $5 \times 5 \times 5$ (stride = 1, padding = 2) is applied, followed by the rectified linear unit (ReLU) activation function: $h_1 = \text{ReLU}(\text{Conv3D}(x))$. The output is then downsampled by a 3D max-pooling layer with kernel size $2 \times 2 \times 2$ and stride $2 : h_2 = \text{MaxPool3D}(h_1)$. After pooling, the feature maps are flattened into a one-dimensional vector. For the enolase superfamily, the flattened dimension is $64 \times 14 \times 17 \times 15$, while for the serine protease superfamily it is $64 \times 17 \times 19 \times 21$. This flattened vector is then projected into a 128-dimensional latent space through a fully connected layer: $h_3 = \text{ReLU}(W_1 h_2 + b_1)$, $h_3 \in \mathbb{R}^{128}$. To prevent overfitting, a dropout layer with dropout probability $p = 0.7$ is applied: $h_4 = \text{Dropout}(h_3)$. Finally, the latent representation is passed through a second fully connected layer that maps to the target number of protein subfamilies: $\hat{y} = W_2 h_4 + b_2$, $\hat{y} \in \mathbb{R}^{\text{num\_classes}}$. On our data, it was always three.

2.3 Interpretable 3D Geometric Learning Architecture

We integrated Grad-CAM++ [10] into the framework to interpret the salience of individual voxels for classification. Grad-CAM++ generates class-specific saliency maps through a weighted combination of positive partial derivatives of the output score for a class c with respect to the feature maps of the last convolutional layer (details are described in Appendix A.6), these weights identify the contribution of different voxels for classification, delineating the geometric subregion within the frequent region that is most salient for distinguishing proteins by subfamilies, and thus their binding specificity. By comparing these salient regions across different subfamilies, we identify amino acids intersecting with these salient regions as key amino acids that confer binding specificity through steric hindrance mechanisms, this process is shown in Fig. 3. Given a voxel's spatial location (x, y, z) for a particular class c, we use Grad-CAM++ to generate a class-specific saliency map L^c as: $L^c_{x,y,z} = \sum_k w^c_k A^k_{x,y,z}$.

3 Results

3.1 Generated Explanations are Consistent with Established Findings

We investigated whether residues from one protein that intersect with salient voxels of a protein from another subfamily contribute to differences in binding preferences between the two subfamilies, according to established experimental findings. We interpret the overlaps as potential determinants of subfamily-specific binding: When a residue from one protein occupies voxels that are salient for another subfamily, we interpret that residue as sterically hindering a region that must remain frequently solvent accessible in the other subfamily. Cross-subfamily overlaps between residues and salient voxels thus provide a steric explanation for specificity.
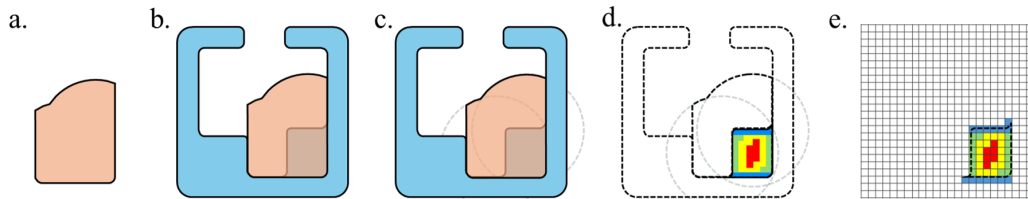
Figure 3: Saliency map localization. **(a)** Frequent region of one protein. **(b)** Superposing the frequent region onto a protein conformation from another subfamily reveals that a portion of the binding site is sterically blocked at the corresponding location in that protein. **(c)** Superposing union of binding ligand atoms spheres onto the protein. **(d)** Grad-CAM++ identifies the saliency region (rainbow-colored area, where different colors represent varying levels of saliency) within the binding site, which coincides with the blocked region of the protein in (c), indicating that binding specificity is driven by steric hindrance. The amino acids forming the blocked region alter binding specificity to the protein by creating steric constraints. **(e)** Translation of salient regions into the weighted voxel representation. Red, yellow, green and blue squares are voxels with decreasing salience.

In the serine proteases, Val216 in elastase blocks a binding site region that remains solvent accessible in trypsin (Fig.4). This difference helps explain why the two subfamilies exhibit different ligand preferences. When FlexVASP-S is applied to trypsin (1ane), the most salient voxels form a region inside the binding pocket that is used to accommodate ligands. Superposing this salient region onto an elastase structure (1b0e) shows that it overlaps with Val216, indicating that this residue physically occupies the space that is accessible in trypsin. Consequently, elastase cannot frequently adopt the same open cavity shape as trypsin in this region, and Val216 acts as a key residue that contributes to binding specificity through steric hindrance. This interpretation is consistent with experimental observations reported by Kunori [14]. Moreover, the binding affinity difference between trypsin-ligand interaction and elastase-ligand interaction also support our result. Experimental binding free energies for variants of OMTKY3, a serine protease inhibitor, show that the same P1 residue can bind both a trypsin-like protease (SGPB) and elastase (PPE), but with much weaker affinity in elastase [15]. For example, the P1 Lys variant of OMTKY3 binds SGPB with $\Delta G_{\text{bind}} \approx -11.3$, kcal/mol, while it binds PPE with $\Delta G_{\text{bind}} \approx -6.3$, kcal/mol, a ~ 5, kcal/mol penalty corresponding to an estimated 10^3–10^4-fold loss in affinity.

In the enolase superfamily, Glu211 of enolase (1ebh) is intersected with salient voxel region of protein (1bkh) from Muconate Lactonizing Enzyme (MLE), which also showed Glu211 has steric hindrance effects and contribute to enolase's binding specificity. This is supported by [16], Glu211's carboxylate side chain is positioned specifically to interact with the 3-OH group of 2-phosphoglycerate (2-PGA). This positioning not only facilitates hydrogen bonding but also physically constrains the substrate, it prevents other molecules or incorrectly oriented substrates from binding effectively. In other words, the side chain of Glu211 distinguishes enolase from MLE because it contributes a steric barrier that would invade the binding cavity of MLE, interfering with its function and specificity.
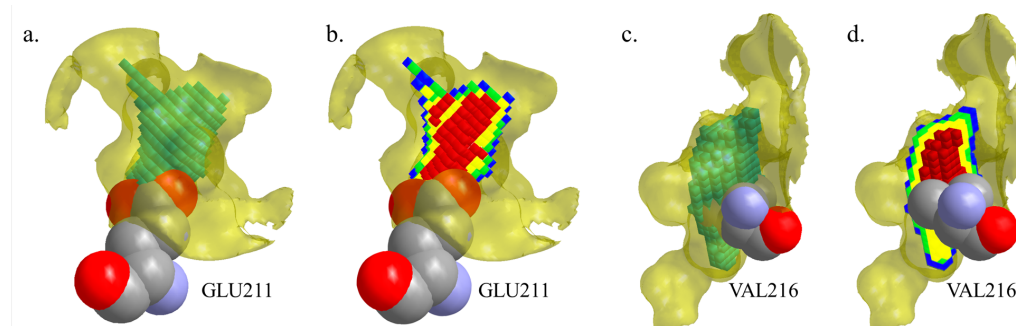

Figure 4: Salient regions of Muconate Lactonizing Enzyme, and enolase binding cavities. The binding cavities of Muconate Lactonizing Enzyme (pdb: 1bkh, panels **a,b**) and Rattus rattus trypsin (pdb: 1ane, panels **c,d**), are shown in transparent yellow. The most salient 600 voxels identified by FlexVASP-S are shown as teal cubes in **(a)** and **(c)**. In **(b)** and **(d)**, the gradient of red, yellow, green and blue cubes illustrates four groups of 150 cubes with decreasing salience. The salient region of Muconate Lactonizing Enzyme (1bkh) is intersected with amino acid Glu211 from enolase (1ebh) (panels, **a,b**), and salient region of trypsin (1ane) intersected with amino acid Val216 from Elastase (1b0e), these intersection suggested that these amino acids protrude into binding cavities and side chains of these amino acids occupy part of the spatial region that accommodates ligand, so their presence restricts the available region within binding cavity, imposing steric hinderance that influences the binding specificity by preventing non-complementary ligands from accessing the cavity.

		PCA+LR		FlexVASP-S	
		Accuracy	F1	Accuracy	F1
Enolase	**Overall**	60.62 ± 37.02	33.35 ± 23.82	$\mathbf{83.82 \pm 37.33}$	$\mathbf{83.99 \pm 37.38}$
	Mandelate Racemase	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
	Muconate Lactonizing Enzyme	51.24 ± 32.44	28.28 ± 15.55	$\mathbf{98.77 \pm 3.00}$	$\mathbf{99.30 \pm 1.70}$
	Enolase	84.44 ± 19.60	46.41 ± 21.99	$\mathbf{100.00 \pm 0.00}$	$\mathbf{100.00 \pm 0.00}$
Serine Protease	**Overall**	63.64 ± 41.18	47.60 ± 39.98	$\mathbf{86.25 \pm 34.03}$	$\mathbf{86.25 \pm 34.03}$
	Trypsin	86.17 ± 23.84	68.16 ± 35.09	$\mathbf{98.00 \pm 6.32}$	$\mathbf{98.00 \pm 6.32}$
	Chymotrypsin	34.36 ± 39.76	15.45 ± 18.97	$\mathbf{50.00 \pm 57.73}$	$\mathbf{50.00 \pm 57.73}$
	Elastase	0.00 ± 0.00	0.00 ± 0.00	$\mathbf{100.00 \pm 0.00}$	$\mathbf{100.00 \pm 0.00}$

Table 1: Performance of FlexVASP-S and PCA+LR on enolases and serine proteases. Accuracy and F1 scores are reported as average $\pm$ standard deviation. Scores from the better performing method are bolded.

3.2 Binding site clsssification performance

While the primary purpose of FlexVasp-S is to explain the relevance of different regions of potential hindrance, our approach employs a classifier to generate the explanations. The classifier, which considers many simulated timesteps of the binding site, is distinct from existing classifiers, to our knowledge. For this reason, we are not able to compare the classifier against existing methods. Nonetheless, a comparison is still possible by using principal components analysis (PCA) to generate features as input to a linear regression (LR) model. We refer to this baseline model as PCA + LR.

We evaluated FlexVASP-S and PCA+LR for classifying subfamily membership in the Enolase and Serine Protease families (Table 1). FlexVASP-S consistently outperformed the traditional PCA+LR model, achieving an accuracy of 83.82% and an F1 score of 83.99% on the Enolase dataset, as well as an accuracy of 83.98% and an F1 score of 84.90% on the Serine Proteases dataset. In contrast, the PCA+LR model attained only 60.62% accuracy and 33.36% F1 on Enolase, and 60.6% accuracy with a 72.59% F1 score on Serine Proteases. Notably, FlexVASP-S reached perfect or near-perfect classification for Enolase and Muconate Lactonizing Enzyme subfamilies but failed to distinguish Mandelate Racemase, and classified Elastase and Chymotrypsin, where PCA+LR failed. These results highlight FlexVASP-S's ability to capture subtle structural determinants of binding specificity that linear methods miss.

Some performance limitations were observed, especially in the Enolases. Mandelate Racemase has a prediction accuracy of 0, primarily due to two factors. First, the sample size for this subfamily is limited, with only 3 proteins with a total of 1800 frequent regions. In comparison, the Muconate Lactonizing Enzyme subfamily has 6 proteins and 3600 frequent regions, and the Enolase subfamily has 10 proteins and 6000 frequent regions. Second, an induced-fit effect also impacts prediction performance. The binding cavity of Mandelate Racemase has significant conformational differences between the ligand-free (apo) state and the ligand-bound state, and the model does not consider these ligand-induced rearrangements [17, 18]. As a result, simulation fails to accurately capture the true cavity shape during substrate binding, leading to low prediction accuracy.

4 Discussion

We have demonstrated FlexVASP-S, the first tool for automatically explaining differences in steric hindrance that control ligand binding preferences, even in the presence of molecular motion. We found that FlexVASP-S reliably identifies steric differences in binding cavities that underlie small-ligand specificity, and these computed differences align with experimentally established findings on steric hindrance-realted amino acids. These results indicate that FlexVASP-S can provide steric-hindrance–aware interpretations of binding cavities, rather than serving solely as a model optimized for maximal predictive accuracy.

Our examination of FlexVASP-S faced some practical challenges. First, FlexVASP-S relies on apo MD simulations. In such cases, the simulations may fail to capture the true cavity shape relevant to substrate binding, visible from induced fit, thereby degrading predictive performance. Second, evaluating FlexVASP-S requires data with many specific properties, including subfamilies with diverse binding preferences but nonredundant subfamily members, high quality crystal structures, and carefully elucidated specificity mechanisms. Longer simulations and more data sets would present practical hurdles but also an opportunity to better understand how machine learning systems can be used to explain the detailed mechanisms that govern selective binding. For the future, our findings here suggest that FlexVASP-S can be a framework for automatically interpreting binding specificity from structure and a starting point for computationally guiding the structure-based design of selective ligands.

Acknowledgements

References

[1] Maria I Zavodszky and Leslie A Kuhn. Side-chain flexibility in protein–ligand binding: the minimal rotation hypothesis. *Protein Science*, 14(4):1104–1114, 2005.

[2] Zhichao Miao and Yang Cao. Quantifying side-chain conformational variations in protein structure. *Scientific reports*, 6(1):37024, 2016.

[3] Brian G Godshall, Yisheng Tang, Wenjie Yang, and Brian Y Chen. An aggregate analysis of many predicted structures to reduce errors in protein structure comparison caused by conformational flexibility. *BMC structural biology*, 13(Suppl 1):S10, 2013.

[4] Maxim Shatsky, Ruth Nussinov, and Haim J Wolfson. Flexible protein alignment and hinge detection. *Proteins: Structure, Function, and Bioinformatics*, 48(2):242–256, 2002.

[5] Dima Kozakov, Laurie E Grove, David R Hall, Tanggis Bohnuud, Scott E Mottarella, Lingqi Luo, Bing Xia, Dmitri Beglov, and Sandor Vajda. The ftmap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nature protocols*, 10(5):733–755, 2015.

[6] Zhanwen Li, Lukasz Jaroszewski, Mallika Iyer, Mayya Sedova, and Adam Godzik. Fatcat 2.0: towards a better understanding of the structural diversity of proteins. *Nucleic acids research*, 48(W1):W60–W64, 2020.

[7] Daria Timonina, Yana Sharapova, Vytas Švedas, and Dmitry Suplatov. Bioinformatic analysis of subfamily-specific regions in 3d-structures of homologs to study functional diversity and conformational plasticity in protein superfamilies. *Computational and Structural Biotechnology Journal*, 19:1302–1311, 2021.

[8] Maxim Totrov. Ligand binding site superposition and comparison based on atomic property fields: identification of distant homologues, convergent evolution and pdb-wide clustering of binding sites. *BMC Bioinformatics*, 12(Suppl 1):S35, 2011.

[9] Ziyi Guo, Trevor Kuhlengel, Steven Stinson, Seth Blumenthal, Brian Y Chen, and Soutir Bandyopadhyay. A flexible volumetric comparison of protein cavities can reveal patterns in ligand binding specificity. In *Proceedings of the 5th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics*, pages 445–454, 2014.

[10] Aditya Chattopadhay, Anirban Sarkar, Prantik Howlader, and Vineeth N. Balasubramanian. Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 839–847. IEEE, 2018.

[11] Michael L Connolly. Solvent-accessible surfaces of proteins and nucleic acids. *Science*, 221(4612):709–713, 1983.

[12] Berk Hess, Carsten Kutzner, David Van Der Spoel, and Erik Lindahl. Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation*, 4(3):435–447, 2008.

[13] Georgi D Georgiev, Kevin F Dodd, and Brian Y Chen. Precise parallel volumetric comparison of molecular surfaces and electrostatic isopotentials. *Algorithms for Molecular Biology*, 15(1):11, 2020.

[14] Yuichi Kunori. Rodent alpha-chymases are elastase-like proteases. *Journal of Allergy and Clinical Immunology*, 111(2):S183, 2003.

[15] Tor Arne Heim Andberg. Protein-protein binding affinities calculated using the lie method. Master's thesis, Universitetet i Tromsø, 2011.

[16] Todd M Larsen, Joseph E Wedekind, Ivan Rayment, and George H Reed. A carboxylate oxygen of the substrate bridges the magnesium ions at the active site of enolase: structure of the yeast enzyme complexed with the equilibrium mixture of 2-phosphoglycerate and phosphoenolpyruvate at 1.8 å resolution. *Biochemistry*, 35(14):4349–4358, 1996.

[17] Daniel Seeliger and Bert L de Groot. Conformational transitions upon ligand binding: Holo structure prediction from apo conformations. *Biophysical Journal*, 98(3):428a, 2010.

[18] Adam D Lietzan, Mitesh Nagar, Elise A Pellmann, Jennifer R Bourque, Stephen L Bearne, and Martin St. Maurice. Structure of mandelate racemase with bound intermediate analogues benzohydroxamate and cupferron. *Biochemistry*, 51(6):1160–1170, 2012.

[19] Brian Y. Chen and Barry Honig. Vasp: a volumetric analysis of surface properties yields insights into protein-ligand binding specificity. *PLoS Computational Biology*, 6(8):e1000881, 2010.

[20] Irina Kufareva and Ruben Abagyan. Methods of protein structure comparison. In *Homology modeling: Methods and protocols*, pages 231–257. Springer, 2012.

[21] Ambrish Roy and Yang Zhang. Recognizing protein-ligand binding sites by global structural alignment and local geometry refinement. *Structure*, 20(6):987–997, 2012.

[22] Mohammad A. Siddiq, Georg K. A. Hochberg, and Joseph W. Thornton. Evolution of protein specificity: insights from ancestral protein reconstruction. *Current Opinion in Structural Biology*, 47:113–122, 2017.

[23] Antonio Rausell, François Juan, Joaquín Pazos, and Alfonso Valencia. Protein interactions and ligand binding: from protein subfamilies to functional specificity. *Proceedings of the National Academy of Sciences*, 107(5):1995–2000, 2010.

[24] Abdullah Kahraman, Richard J. Morris, Roman A. Laskowski, and Janet M. Thornton. Shape variation in protein binding pockets and their ligands. *Journal of Molecular Biology*, 368(1):283–301, 2007.

[25] Lizbeth Hedstrom. An overview of serine proteases. *Current Protocols in Protein Science*, 26(1):21–10, 2001.

[26] Lizbeth Hedstrom. Serine protease mechanism and specificity. *Chemical Reviews*, 102(12):4501–4524, 2002.

[27] John A. Gerlt, Patricia C. Babbitt, and Ivan Rayment. Divergent evolution in the enolase superfamily: the interplay of mechanism and specificity. *Archives of Biochemistry and Biophysics*, 433(1):59–70, 2005.

[28] Margaret E. Glasner, John A. Gerlt, and Patricia C. Babbitt. Evolution of enzyme superfamilies. *Current Opinion in Chemical Biology*, 10(5):492–497, 2006.

[29] A. D. Buckingham and B. D. Utting. Intermolecular forces. *Annual Review of Physical Chemistry*, 21(1):287–316, 1970.

[30] Berk Hess. P-lincs: A parallel linear constraint solver for molecular simulation. *Journal of Chemical Theory and Computation*, 4(1):116–122, 2008.

A   Appendix

A.1   Background

Protein families formed through divergent evolution usually share common structural characteristics, but homologous proteins within a superfamily also develop distinct ligand-binding preferences, leading to the emergence of different subfamilies [7, 19]. Such specificity is mainly associated with local structural variations at binding sites rather than global differences [7, 8]. Traditional structure comparison methods, which rely on sequence or backbone similarity often overlook subtle variations such as side-chain flexibility that influence binding specificity [1, 2, 3, 4]. Moreover, these methods typically describe structural similarity or differences with a single numerical score like root mean square deviation (RMSD). While RMSD is useful for classifying structures, it is less effective in capturing and representing subtle differences between closely related proteins. For example, in some cases, the RMSD between two conformations of the same protein can be as high as that between two entirely unrelated structures [6]. Proteins can show similar RMSD values despite substantial global changes induced by substrate binding or complex formation [6, 20].

In most protein–ligand interactions, a few residues in binding pocket have major influences on partner recognition and binding affinity [21, 22]. Understanding how these residues confer specificity through steric, electrostatic, or hydrogen-bonding mechanisms is critical for explaining functional diversity, disease-related mutations, and guiding protein redesign [21, 23, 24, 22]. This understanding must also be developed in the presence of protein flexibility.

Identifying the exact mechanisms behind specific binding remains a challenging problem that often requires expert knowledge. Specifically, it requires experts to visually check protein-ligand complexes and their binding cavities to identify the mechanism through which amino acids act. This process is time-consuming, it requires a deep understanding of molecular interactions, and can be subjective depending on the expert's experience. If software could automatically unveil specificity mechanisms, it might accelerate the drug design process. Unfortunately, there is no algorithm capable of automatically explaining the mechanisms of residues that drive protein–ligand binding specificity.

Moreover, identifying a reliable and distinguishable region that confers binding specificity to proteins requires consideration of structural conformational flexibility. A single fixed conformation of protein might not be representative of regions that control specificity. Some binding sites are only visible in specific conformations [5]. When the proteins studied are closely related, conformational flexibility in side chains or the backbone can obscure subtle similarities or differences within their ligand-binding cavities that are critical for controlling specificity [9].

To address the challenge of explaining binding specificity, this paper presents FlexVASP-S, a steric-aware 3D geometric framework that isolates steric hindrance by focusing solely on spatial constraints from amino acid side chains. FlexVASP-S leverages voxelized "frequent regions" derived from ensembles of flexible conformations and employs a 3D-CNN with Grad-CAM++ to both classify protein subfamilies and highlight specificity-related regions. The framework (1) detects subtle cavity differences that determine binding specificity, (2) provides mechanistic interpretability without manual expert inspection, and (3) identifies key regions by integrating multiple conformations rather than relying on a single static structure.

Superfamily	Subfamily	PDB
Serine Proteases	Trypsins	2f91, 1fn8, 2eek, 1h4w, 1bzx, 1aq7, 1ane, 1aks, 1trn, 1a0j
	Chymotrypsins	1eq9, 4cha, 1kdq, 8gch
	Elastases	1elt, 1b0e
Enolase	Enolases	1e9i, 1pdy, 1iyx, 1ebh, 1te6, 1w6t, 2xsx, 2pa6, 3otr, 3qn3
	Mandelate Racemases	2ox4, 1mdr, 2og9
	Muconate Lactonizing Enzyme	2pgw, 1bkh, 3dgb, 3i4k, 2zad, 1jpm

Table A.1: PDB codes of selected families used in this study.

A.2 Serine protease and Enolase superfamily Dataset

Serine protease superfamilies can exhibit endo- or exopeptidase activity, with some members involved in protein digestion. For example, trypsins, chymotrypsins, and elastases are endoproteases that cleave polypeptides into shorter chains. The specificity of these three enzymes, however, is determined by their substrate rather than structural or mechanistic features [25]. For trypsins, chymotrypsins, and elastases, substrate recognition is primarily restricted to the S1 site, whereas other serine proteases can recognize substrates beyond S1. Each enzyme exhibits distinct substrate preferences, leading to their binding specificity: chymotrypsin favors hydrophobic P1 residues, forming a large hydrophobic pocket; trypsin prefers positively charged P1 residues due to its negatively charged S1 site; elastase favors small aliphatic P1 residues, as its S1 site is smaller than that of chymotrypsin or trypsin [26].

Proteins in the enolase superfamily share a common partial reaction in which they abstract the α-proton from a carboxylate substrate, generating an enolate anion intermediate stabilized by a crucial Mg^{2+} ion. The specific active site structure then directs these intermediates toward different end products [27, 28]. The shared active site resides within a modified $(\beta/\alpha)_8$ -barrel, with conserved residues arranged around it. A second domain caps the active site, conferring substrate specificity for some enolase enzymes. While metal-coordinating residues are highly conserved, the identity and position of the general base that abstracts the proton are subfamily-specific [28].

Based on the two superfamilies above, we included 16 Enolase structures and 19 Serine Protease structures, but the resulting Enolase and Serine Protease datasets are imbalanced across subfamilies. Because we only consider two protein superfamilies, the limited sample size is unavoidable and may reduce the representativeness of the training data. Moreover, the dataset is necessarily small because the number of available structures for these subfamilies in the RCSB is itself limited. Many additional PDB entries are redundant with our current set, and therefore cannot provide meaningful new training examples.

A.3 Molecular dynamics simulation

To generate conformational samples for each protein in our dataset, we performed molecular dynamics (MD) simulations using GROMACS 2021.4 with CUDA 11.6 support. Protein structures were obtained from the PDB database, and all water molecules were removed prior to simulation.

Each protein was centered in a cubic water box with a 1.0 nm buffer from the box edge. We used the GROMOS96 54a7 force field and the SPC\E water model [29], with periodic boundary conditions applied throughout. Sodium and potassium ions were added at low concentrations (<0.1% salinity) to neutralize the system. Energy minimization was performed using the steepest descent algorithm, followed by a 1 ns NPT equilibration in four 250 ps steps. Position restraints were gradually reduced from 1000 to 0 kJ/(mol·nm), and the backbone restraints were released for the production run. Pressure and temperature were maintained at 1 bar (Parrinello-Rahman) and 300 K (Nosé-Hoover), respectively. Bond constraints were handled using the P-LINCS algorithm [30], and electrostatics were computed with PME [12]. The production MD simulation ran for 100 ns with a 1 fs timestep using OpenMPI. Afterward, trajectories were converted to PDB format with water removed and aligned to the initial structure. A total of 1000 conformational snapshots were selected from the trajectory for further analysis.

A.4 Generating Frequent Regions

Representing all frequent regions that describe exactly k out of N samples would require computing all $\binom{N}{k}$ possible combinations. That effort would be impractical for most values of k, so we take 600 random samples from the space of $\binom{N}{k}$ combinations. Overall, the approach, following Fig. A.1, starts by generating the binding site for every conformational sample (Fig. A.1c).

Next, we randomly select k conformations from $T_i \subseteq \{1, 2, \ldots, N\}$ to get an spatial intersection $O_i = \cap_{i=1}^k T_i$. The frequent region is the union of all intersections across the 600 iterations: $\bigcup_{i=1}^{600} O_i$ (Fig. A.1). The process is of course affected by the random selection, so we repeat this random process 500 times, creating 500 varying frequent regions describing each binding site for downstream tasks. These sampling quantities were selected in our preliminary work to be large enough to similar samples of frequent regions, for reproducibility.
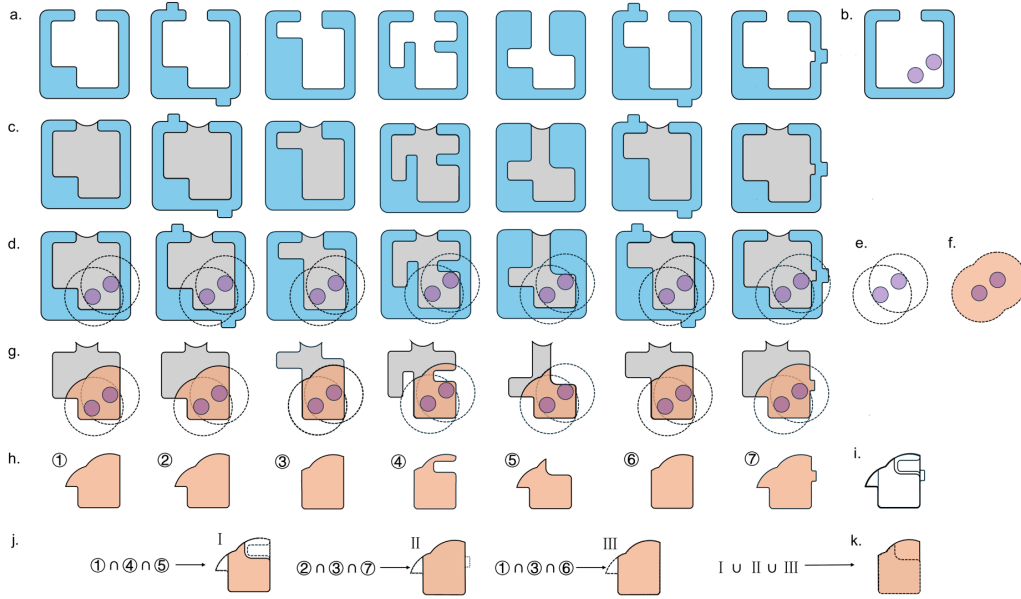
Figure A.1: Extracting frequent regions. (a) Sampled conformations $T_0, T_1, \ldots, T_N$ of protein T. (b) The pivot structure t' with ligand atoms (purple). (c) Binding cavities of each conformational sample (grey). (d) Structural alignment of each conformational sample to t' allows the ligand to mark the ligand binding site in each conformation. (e) Spheres are defined at 5Å radius around ligand atoms. (f) The union of ligand-centered spheres, for binding site computations. (g) Generating the binding site (orange) of each conformation. (h) The binding site of each protein cavity. (i) All binding sites, superposed. (j) Rare intersections of k randomly selected binding sites (I), and common intersectionss of k (for simplicity of illustration, we set k = 3 in this example) randomly selected sites (II, III). (k) The CSG union of the intersections of I, II, III defines the frequent region. 500 of these are generated for each protein in our data set.

A.5 Voxelization

After the "frequent regions" within the protein structures are determined, they are converted into a grid-based representation for 3D Convolutional Neural Network (3D-CNN) analysis. Generating a bounding box that is consistent across all proteins and can accommodate all proteins from the same superfamilies within it. Each bounding box is discretized into cubes (voxels) of side length:

$$\Delta = 0.5 \,\text{Å} \tag{1}$$

So, the number of voxels along each axis is:

$$N_x = \left\lceil \frac{L_x}{\Delta} \right\rceil, \quad N_y = \left\lceil \frac{L_y}{\Delta} \right\rceil, \quad N_z = \left\lceil \frac{L_z}{\Delta} \right\rceil$$

Thus, the voxelized tensor shape is:

$$\mathcal{T} \in \mathbb{R}^{N_x \times N_y \times N_z} \tag{2}$$

To accommodate a whole number of these cubes, the bounding box is adjusted with slight padding as necessary. The voxelized tensor shape of two superfamilies are:

$$\mathcal{T}_{\text{Enolase}} \in \mathbb{R}^{28 \times 34 \times 30} \tag{3}$$

$$\mathcal{T}_{\text{Serprot}} \in \mathbb{R}^{34 \times 38 \times 42} \tag{4}$$

By performing constructive solid geometry (CSG) operations with the cubes, we compute the portion of the cavity volume that lies within each voxel. Formally, let $V_{\text{cavity}} \subset \mathbb{R}^3$ denote the cavity region,

and let C_{ijk} denote the cubic region corresponding to voxel (i, j, k). The voxel occupancy value is defined as:

$$v_{ijk} = \frac{\text{Vol}(V_{\text{cavity}} \cap C_{ijk})}{\text{Vol}(C_{ijk})} \in [0, 1] \tag{5}$$

Thus, the resulting tensor of voxel volumes is:

$$\mathcal{T} = [v_{ijk}] \in \mathbb{R}^{N_x \times N_y \times N_z} \tag{6}$$

This tensor $\mathcal{T}$ is then used as the input for training or classification with the 3D-CNN.

A.6 Grad-CAM++

A weighted combination of the positive partial derivatives of the last convolutional layer feature maps with respect to a specific class score as weights are used to generate a visual explanation for the corresponding class label. Here, we try to find the most important regions and amino acids that contribute to the classification. Given a voxel's spatial location (x, y, z) for a particular class c, we use Grad-CAM++ to generate a class-specific saliency map L^c as:

$$L^c_{x,y,z} = \sum_k w^c_k A^k_{x,y,z} \tag{7}$$

where $A^k_{x,y,z}$ is the kth feature map in the last convolutional layer of 3D-CNN, and w^c_k is the corresponding weight defined as follows:

$$w^c_k = \sum_x \sum_y \sum_z \alpha^{kc}_{x,y,z} \cdot relu\left(\frac{\partial Y^c}{\partial A^k_{z,y,z}}\right) \tag{8}$$

$$\alpha^{kc}_{x,y,z} = \frac{\dfrac{\partial^2 Y^c}{\left(\partial A^k_{x,y,z}\right)^2}}{2\dfrac{\partial^2 Y^c}{\left(\partial A^k_{x,y,z}\right)^2} + \sum_p \sum_q \sum_r A^k_{p,q,r}\left\{\dfrac{\partial^3 Y^c}{\left(\partial A^k_{x,y,z}\right)^3}\right\}} \tag{9}$$

where $Y^c = \exp(S^c)$ is the class score, and S^c is the penultimate layer score of class c. $\frac{\partial Y^c}{\partial A^k_{x,y,z}}$, $\frac{\partial^2 Y^c}{\left(\partial A^k_{x,y,z}\right)^2}$, and $\frac{\partial^3 Y^c}{\left(\partial A^k_{x,y,z}\right)^3}$ are the first-, second-, and third-order gradients w.r.t. $A^k_{x,y,z}$.

A.7 Experimental Design

For each protein in the training family, 1,000 conformational samples are generated, from which 500 frequent region samples are extracted. To train the 3D-CNN model, all frequent region samples of one protein are set aside as an evaluation set, while the remaining dataset undergoes 5-fold cross-validation to optimize model parameters and prevent overfitting. The best-performing model weights from the five folds are then used to predict the subfamily category of the evaluation samples. This procedure is repeated iteratively for each protein, ensuring that evaluation data remain independent from model training. Additionally, the number of conformational samples k used to generate each frequent region influences classification performance, and an optimal k is determined by testing a range of values (15, 25, 50, 70, 100, 150, 200) based on prior study[9].