ACS OMEGA

Open Access

http://pubs.acs.org/journal/acsodf

Article

Comparative Data Analysis of Virtual Screening Methodologies for Predicting Urease Inhibitory Activity

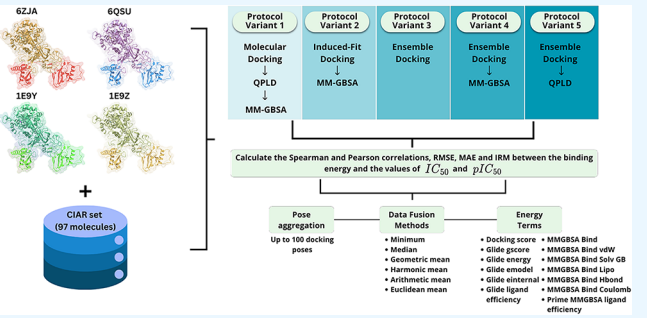
Elizabeth Valdés-Muñoz,[○] Gabriel J. Olguín-Orellana,[○] Sofía E. Ríos-Rozas, Melissa Alegría-Arcos, Natalia Morales, Vicente Rojas-Santander, Javier Farías-Abarca, Jonathan M. Palma, Erix W. Hernández-Rodríguez, Reynier Suardíaz, and Daniel Bustos*

Cite This: *ACS Omega* 2025, 10, 49641−49658

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Structure-based virtual screening (SBVS) is a fundamental approach in drug discovery, yet its predictive accuracy is highly dependent on methodological choices, scoring functions, and data processing strategies. This study systematically evaluates five protocol variants integrating molecular docking, induced-fit docking (IFD), quantum-polarized ligand docking (QPLD), ensemble docking (ED), and molecular mechanics/generalized Born surface area (MM-GBSA) in *Helicobacter pylori* urease employing four distinct crystallographic structures obtained from the protein data bank (PDB). We assess their predictive performance using statistical correlation metrics (Spearman and Pearson) and error-based measures (mean absolute error, root-mean-squared error, and inlier ratio metric). Additionally, we investigate the influence of data fusion techniques—minimum, median, arithmetic, geometric, harmonic, and Euclidean means—and varying numbers of docking poses (ranging from 1 to 100) on ligand ranking accuracy. Results indicate that MM-GBSA and ED consistently outperform other methods in compound ranking, although MM-GBSA exhibits higher errors in absolute binding energy predictions. While increasing the number of poses generally reduces predictive accuracy, the minimum fusion approach remains robust across all conditions. Comparisons between IC_{50} and pIC_{50} as experimental reference values reveal that pIC_{50} provides higher Pearson correlations, reinforcing its suitability for affinity prediction, while both metrics perform similarly in Spearman rankings. These findings refine SBVS workflows by optimizing scoring and pose aggregation strategies, highlighting the importance of method selection and data fusion techniques. The proposed framework enhances ligand prioritization in virtual screening campaigns and can be adapted to other therapeutic targets. Future research should explore adaptive scoring frameworks and machine-learning approaches to further improve the SBVS predictive reliability.



1. INTRODUCTION

Computer-aided drug design (CADD) encompasses *in silico* tools that integrate various theoretical disciplines, including cheminformatics, bioinformatics, and molecular modeling, to identify compounds with potential to modulate macromolecular functions.[1] These tools are broadly classified into structure-based virtual screening (SBVS), which requires knowledge of the target structure, and ligand-based virtual screening (LBVS), which relies on the properties of known active compounds.[2,3] While both approaches are widely used, each presents challenges: SBVS often produces false positives (FP), due to difficulties in accurately parametrizing ligand−receptor interactions, whereas LBVS encounters activity cliffs, where chemically similar molecules exhibit drastically different biological activities.[4,5] Overcoming these limitations is essential for improving the reliability of virtual screening in drug discovery.

The computationally more accessible methods employed in VS encompass various techniques, among which molecular docking, ensemble docking (ED), induced-fit docking (IFD), quantum parametrized ligand docking (QPLD), and molecular mechanics with generalized born and surface area (MM-GBSA) stand out. Molecular docking aims to predict the orientation and interaction of a ligand within the three-dimensional structure of a receptor protein.[6] ED extends this methodology by considering multiple conformations of the target protein, thereby improving prediction accuracy by offering a more realistic view of molecular interactions.[7] IFD

Received: May 13, 2025
Revised: September 17, 2025
Accepted: September 22, 2025
Published: October 14, 2025

accounts for the flexibility of both the ligand and some specific regions of the receptor, allowing side chains in the binding site to move or rearrange to better fit the ligand.[8,9] QPLD integrates quantum mechanics into docking to refine ligand atom charges, providing greater accuracy in predicting molecular interactions.[10] Lastly, MM-GBSA is an end-point method that focuses on calculating the free energy of binding in molecular interactions by decomposing it into enthalpic and entropic terms.[11]

Calibration of SBVS methodologies typically follows two approaches: comparing predicted binding energies with experimental data such as IC_{50} or K_i values and evaluating the consistency of predicted binding modes against cocrystallized ligand structures.[12] While root-mean-square deviation (RMSD) is widely used to assess the accuracy of binding mode predictions, its relevance diminishes when no cocrystallized ligand structurally similar to the screened compounds is available. In such scenarios, the ability to accurately predict binding energies becomes pivotal, as it provides a more direct and meaningful metric for identifying promising candidates for subsequent *in vitro* validation.[13] The accurate prediction of binding energies is crucial for reducing both FP and false negatives (FN) in VS. This arises when nonviable compounds are incorrectly predicted to bind strongly, leading to wasted resources and increased costs during experimental testing. Conversely, FN represents high-potential compounds erroneously discarded during screening, potentially missing opportunities for groundbreaking discoveries. Notably, a correct prediction of binding modes does not guarantee an accurate estimation of absolute or relative binding affinities. Consequently, SBVS protocols must be carefully calibrated to prioritize methods that not only predict binding modes accurately but also correlate strongly with experimental affinity data, ensuring a more reliable and cost-effective pathway for identifying effective inhibitors.[14]

To evaluate the relationship between predicted binding affinities and experimental values, various correlation measures and error metrics are employed to assess the predictive performance of SBVS methodologies. The Pearson correlation coefficient (r) is widely used to quantify the strength and direction of the linear relationship between two variables. It is most effective when a direct linear relationship is expected between predicted and experimental values, as often assumed in binding energy estimations.[15] However, Pearson's correlation is sensitive to outliers, which can distort its interpretability in data sets with high variability.[16] The Spearman rank correlation coefficient (ρ), in contrast, assesses the monotonic relationship between variables by comparing their ranked values, offering greater robustness to nonlinear trends and outliers. This makes it particularly valuable in scenarios where the relationship between predicted and experimental values deviates from linearity.[17] In addition to correlation coefficients, error metrics such as the mean absolute error (MAE) and root-mean-square error (RMSE) are frequently employed to quantify the deviation between predictions and experimental observations. MAE provides a straightforward measure of average predictive error, while RMSE, due to the squaring of errors, is more sensitive to larger deviations, highlighting significant deviations that may warrant further investigation.[18] These approaches collectively address critical limitations in VS, where an accurate ranking of molecules is essential for identifying promising candidates for experimental validation.

Selecting the most promising candidate molecules from molecular docking results requires effective aggregation strategies to interpret docking scores across multiple poses. One common approach is the top-scoring method, where molecule rank based on the pose with the most favorable predicted binding energy. While straightforward, this method can be influenced by outliers or scoring inconsistencies. To mitigate such issues, aggregate approaches that utilize information from many poses or all poses are often employed. For instance, calculating the median docking score provides a robust measure that minimizes the impact of extreme values, while various averaging techniques, such as arithmetic, geometric, or harmonic means, offer alternative perspectives on pose distributions.[19] The geometric mean, for example, is less affected by large variations and highlights consistent performance across poses,[20] whereas the harmonic mean emphasizes lower values, giving more weight to poses with stronger predicted binding affinities. Advanced methods, such as the Euclidean norm, treat docking scores as a multidimensional vector and calculate an aggregate that reflects the overall magnitude of the scores.[21] These approaches are particularly valuable when screening campaigns involve significant receptor flexibility, diverse ligand libraries, or scoring function variability, as they ensure that the ranking reflects a more comprehensive assessment of the binding potential. The choice of an aggregation strategy is pivotal, as it directly influences the identification of molecules with genuine binding potential for further experimental validation. The number of poses included in aggregation strategies is a critical yet often underexplored factor in VS. Considering too few poses risks missing alternative binding configurations that could provide important insights, while including too many poses may introduce noise, diluting the relevance of high-quality interactions. Striking the right balance is essential to ensuring that the selected aggregation method accurately reflects the true binding potential of a molecule. Despite its significance, a systematic evaluation of how pose quantity and quality influence aggregation performance and correlation with experimental data remains limited.[22]

Urease is a metalloenzyme that catalyzes the hydrolysis of urea into carbonic acid ($H_2CO_3^-$) and ammonia (NH_3^+). This enzymatic activity is essential for the survival of certain bacteria, fungi, and plants, but also presents significant agricultural and health challenges.[23] In agriculture, urease-mediated ureolysis contributes to nitrogen losses from fertilizers and animal manure, leading to environmental pollution through ammonia emissions and reduced crop efficiency. From a human health perspective, urease is a critical virulence factor in pathogens like *Helicobacter pylori* (Hp), enabling survival in acidic environments and promoting gastric diseases, including ulcers and cancers. Consequently, targeting urease with inhibitors has garnered substantial interest in both fields. The structural and functional characteristics of urease make it an excellent model for evaluating computational methodologies in drug discovery. Its well-characterized active site, featuring a highly conserved binuclear Ni^{2+} center coordinated by a hydroxide ion and a carbamylated lysine,[24] provides a robust framework for testing computational techniques. Moreover, the availability of high-resolution crystallographic structures further supports its use in SBVS studies. Urease's historical significance as the first enzyme discovered and crystallized highlights its pivotal role in biochemistry, offering a solid foundation for computational

https://doi.org/10.1021/acsomega.5c04457
ACS Omega 2025, 10, 49641−49658

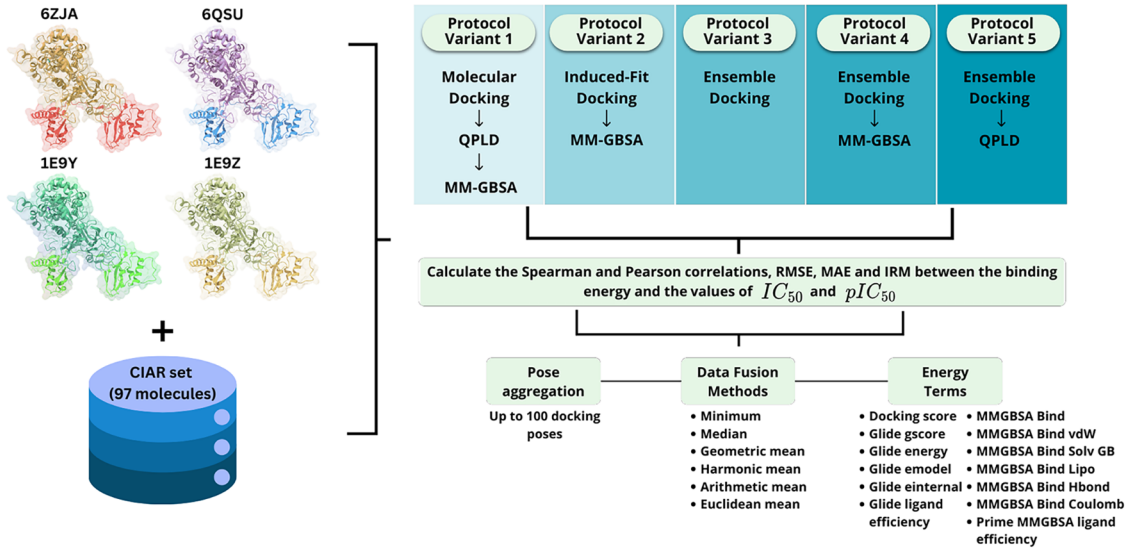


Figure 1. Schematic representation of the computational workflow used in this study.

analyses aimed at refining VS strategies. A wide range of chemical classes has been identified as urease inhibitors, among which coumarins have emerged as a prominent family due to their potent activity and structural diversity. Coumarins are naturally occurring compounds with notable pharmacological properties, including antimicrobial,[25,26] antitumor,[27] and anti-inflammatory[28] activities. These characteristics and their strong interaction with the urease active site underscore the relevance of employing coumarin derivatives as model compounds for validating and optimizing VS methodologies.

This study integrates multiple VS methodologies, including molecular docking, QPLD, ED, IFD, and MM-GBSA, into five protocol variants to systematically explore their predictive performance. To evaluate these methodologies, statistical correlation metrics (Spearman and Pearson) and error-based metrics (MAE, RMSE, and IRM) were employed to compare predicted binding energies with experimental (IC_{50} or pIC_{50}) values. Additionally, six data fusion methods (top-score or minimum, median, arithmetic mean, geometric mean, harmonic mean, and Euclidean mean) were applied across varying numbers of docking poses (data aggregation), ranging from one to one hundred, to assess their influence on ranking accuracy. Urease was selected as a model system, with a data set of coumarin-derived inhibitors termed CIAR (Coumarin Inhibitors Already Reported) and four urease crystallographic structures (6ZJA, 6QSU,[29,30] 1E9Y, and 1E9Z[30]) obtained from the Protein Data Bank (PDB,[31]). This work aims to refine SBVS workflows and enhance the identification of reliable candidates for experimental validation.

2. MATERIALS AND METHODS

This study systematically evaluates the impact of SBVS methodologies, protocol variants, statistical metrics, data fusion, and pose aggregation strategies on predictive performance (Figure 1). Five protocol variants were designed, integrating sequential combinations of molecular docking, QPLD, ED, IFD, and MM-GBSA, where the output of one method serves as the input for the next. By leveraging the complementary strengths of these approaches, a comprehensive assessment of ligand−receptor interactions was achieved. To quantify the reliability and accuracy of predictions, statistical metrics were employed to compare predicted binding energies with experimental values. These include Spearman and Pearson correlations, as well as error metrics such as MAE, RMSE, and IRM, which measure deviations between predicted and observed inhibitory activity.

2.1. Preparation of the CIAR Data Set. The CIAR data set was constructed from a literature search using the PUBMED search engine with the query: "urease AND coumarin AND *H. pylori*". This data set consists of 141 molecules, of which 96 are competitive inhibitors of the urease enzyme, characterized by their IC_{50} values. The remaining 45 molecules were experimentally tested in the same published studies and reported as inactive, because they either exhibited no measurable inhibitory effect or had IC_{50} values above the detection limits under the experimental conditions employed. These molecules were published in[32−37] (see Supporting Files). It is important to note that, unlike classification-focused VS studies, the present work is based on regression analyses aimed at predicting continuous IC_{50} values for urease inhibitors. While decoy molecules are widely used in classification tasks to assess enrichment and model discrimination, incorporating decoys or experimentally inactive compounds into regression analyses would require assigning arbitrarily high IC_{50} values, which could introduce significant bias and distort correlation metrics. Therefore, the 45 compounds initially identified as inactive in our data set were not included in this study, as their incorporation would have necessitated hypothetical affinity values inconsistent with the regression objectives.

Subsequently, the data set was prepared in the Maestro−Schrödinger suite (Schrödinger Release 2021-1: Maestro, Schrödinger, LLC, New York, NY, 2021),[38] by assigning partial charges using the OPLS3e force field (Schrödinger Release 2021-1: OPLS3e Force Field, Schrödinger, LLC, New York, NY, 2021).[39] Additionally, ligand protonation states were determined at pH 7.4 using the PROPKA tool.[40]

2.2. Preparation of Urease Structures. Currently, there are four experimentally determined urease structures from *H. pylori* (*HpU*) deposited in the PDB.[31] These structures (PDB IDs: 6ZJA, 6QSU,[29] 1E9Y, and 1E9Z,[30] Table 1) served as starting points for all the target-dependent methodologies

https://doi.org/10.1021/acsomega.5c04457
ACS Omega 2025, 10, 49641−49658

Table 1. Structural Information on *HpU* Crystallographic Models Used in the Study

PDB code	resolution (Å)	method	ligand	area (Å^2)	volume (Å^3)
6ZJA	2.0	cryogenic electron microscopy	DJM	379 379	300 364
6QSU	2.4	cryogenic electron microscopy	BME	242 700	132 359
1E9Y	3.0	X-ray	HAE	276 773	171 188
1E9Z	3.0	X-ray		108 058	30 792

applied in this study. Briefly, The *HpU* topology consists of α and β subunits. The 6ZJA and 6QSU structures are assembled as dodecamers, i.e., 12 *HpU* units. In contrast, 1E9Y and 1E9Z are trimers (3 α and β units). Literature reports indicate that each of these units ($\alpha + \beta$) functions independently within the bacterial cell; therefore, to reduce the computational costs, only one functional unit was considered for each structure.

The preparation of these proteins was carried out in the Maestro−Schrödinger suite and included correcting partial charges and bond orders using the OPLS3e force field.[39] Hydrogen atoms were added, and the protonation states of the amino acids were defined by using the PROPKA program at pH 7.4. Finally, the energy minimization of hydrogen atoms was performed to relax the structures using the gradient descent algorithm. The Ni^{2+} ions necessary for enzymatic catalysis were retained in the structures. In the same way, three of these structures were solved in a complex with an inhibitor, as shown in Table 1, which were retained for the subsequent step. The volume and area of the catalytic pocket of *HpU* were computed using the CASTpFold server.[41]

2.3. Molecular Dynamics Simulations and Identification of Conformations. The four previously prepared urease structures were subjected to 2 μs-long simulations. These simulations included the cocrystallized ligand within the catalytic site and both Ni^{2+} ions. Previous unpublished simulations revealed that running MD simulations in the apo form of the protein leads to occlusion of the catalytic site, significantly reducing its volume and preventing docking programs from accurately identifying binding poses. All-atom MD simulations were performed in an explicit solvent model (SPC; Single Point Charge) under NPT ensemble conditions, maintaining a constant temperature and pressure (300 K and 1 atm). The Desmond simulation package within the Maestro−Schrödinger suite (Schrödinger Release 2021-1: Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2021) was utilized, with the OPLS3e force field to ensure accurate representation of molecular interactions.

To accurately maintain the coordination geometry of the catalytic Ni^{2+} ions during the MD simulations, positional restraints were applied to both metal ions. Additionally, for the three protein−ligand crystallographic complexes (urease bound to HAE, BME, and DJM), bond restraints were defined between the Ni^{2+} ions and their coordinating atoms in the bound ligand, as well as the histidines and carbamylated lysine in the active site. These constraints were generated using Maestro's standard metal-binding setup and retained throughout the 2 μs MD simulations to preserve the experimentally observed coordination geometry.

Upon completion, trajectory analyses were conducted using the Bio3D library[42] in *R* to calculate the RMSD of the catalytic residues and generate a dendrogram for each trajectory. Based on this analysis, 25 representative conformational states were selected per MD simulation, corresponding to the 25 clusters identified. These conformational states, along with the experimentally resolved structures, served as starting points for subsequent SBVS methods.

2.4. Application of Protocol Variants and Methods to Identify the Best Correlation with Experimental Values. Five protocol variants, illustrated in Figure 1, were designed to evaluate their correlation with experimental values. These variants combine different SBVS methods, including standard docking, IFD, MM-GBSA, QPLD, and ED.

Variants 1 and 2 were applied to the experimentally resolved structures (Table 2), leveraging the crystallographic data to provide a baseline for performance. Variants 3, 4, and 5, however, were applied to the 100 conformational states derived from MD simulations (25 conformations per system across four structures), offering a broader exploration of protein flexibility to better capture realistic ligand−receptor interactions.

Table 2. Variables are Analyzed in This Study

variable	description
protein	the *HpU* crystallographic structures used in the study (Table 1)
frame	the conformational snapshot (if applicable) from MD simulations. 0 in crystallographic structures.
ExpVariable	the experimental variable (IC_{50} or pIC_{50}).
DataFusion	the data fusion method used for docking scores (minimum, geometric, harmonic, arithmetic, Euclidean means, median, Table 3).
EnergyTerm	the energy-based metric analyzed (e.g., docking score, MM-GBSA energy, Table S1).
pose	the data aggregation considering the number of docking poses used in a data fusion method (ranging from 1 to 100).
Pearson and Spearman	correlation coefficients with experimental values.
MAE and RMSE	error-based metrics quantifying the deviation from experimental values.
IRM	the Inlier Ratio Metric, measuring predictive performance robustness.
N	the number of data point in each correlation.
method	the specific computational approach used (e.g., docking, MM-GBSA, QPLD, ED, IFD).
variant	the protocol variant applied, which determines the methodological pipeline (Figure 1).

A detailed description of each method's implementation within the context of this study is provided below, highlighting how each protocol addresses the challenges of accurately predicting binding affinities.

2.4.1. Standard Docking. Standard docking simulations were conducted using the Glide module (Schrödinger Release 2021-1: Glide, Schrödinger, LLC, New York, NY, 2021)[43] within the Maestro−Schrödinger suite, employing the Standard Precision (SP) scoring function. The CIAR data set compounds were docked into the urease catalytic site, using a docking grid composed of two overlapping components: a smaller, focused grid (15 Å × 15 Å × 15 Å) embedded within a larger grid (36 Å × 36 Å × 36 Å), as reported in ref 44. The grid center was defined by the centroid of key *HpU* catalytic site residues: H136, H138, KCX219, H221, H248, H274, C322, R338, and D362, and the two nickel ions. Docking simulations generated up to 10 poses per compound. In all docking experiments, no explicit metal−ligand restraints were applied. This decision was based on the absence of

https://doi.org/10.1021/acsomega.5c04457
ACS Omega 2025, 10, 49641−49658

experimentally resolved binding modes for these types of ligands and to avoid introducing structural bias into the docking poses. Instead, the focus was on the unbiased sampling of potential binding interactions, allowing the scoring functions to penalize poses that failed to establish relevant contacts with the active site. The same logic was applied to IFD, ED, and QPLD schemes.

2.4.2. MM-GBSA. MM-GBSA, unlike standard docking, does not predict binding sites or ligand interaction modes. Instead, it refines the protein−ligand complex geometry and provides a more accurate estimate of binding free energy by incorporating additional energetic terms not accounted for in standard docking. The method uses docking poses as the starting points for its calculations.

The protein−ligand-ion complexes were re-evaluated using MM-GBSA, implemented in the Maestro−Schrödinger suite via the Prime tool (Schrödinger Release 2021-1: Prime, Schrödinger, LLC, New York, NY, 2021). This method calculates the binding free energy (ΔG_{bind}) by decomposing it into key enthalpic contributions: (1) molecular mechanics energy terms (van der Waals and electrostatic interactions) and (2) solvation effects comprising polar and nonpolar components. This protocol applies local minimization to the protein−ligand complex, allowing flexible residues within an 8.0 Å radius of the ligand while constraining the remainder of the protein to maintain the global fold. The ΔG_{bind} corresponds to the minimized, partially constrained system.

2.4.3. Quantum-Polarized Ligand Docking. These calculations enable the reparameterization of the ligand partial charges within the binding site context, followed by a subsequent redocking of the ligand. Consequently, similar to MM-GBSA, this method requires previous docking results as its starting point (based on Section 2.4.1).

Partial charges were determined using Jaguar software (Schrödinger Release 2021-1: Jaguar, Schrödinger, LLC, New York, NY, 2021).[45] The quantum mechanical calculation was carried out with the B3LYP functional and the lacvp* basis set without applying any implicit solvation model. Partial charges computed from the QM step were transferred to the ligand by using the PULL stage of the QPLD workflow. Subsequently, redocking was performed in Glide XP mode, applying an RMSD cutoff of 0.5 Å, a maximum atomic displacement of 1.3 Å, and an LVDW scaling factor of 0.8. Initially, up to 77 poses per ligand were generated and subsequently filtered down to 10 final poses based on scoring criteria. Ligand charges obtained from the QM calculation were retained and explicitly incorporated during the redocking stage to ensure accurate electrostatic modeling.

2.4.4. Induced-Fit Docking. IFD is a molecular docking approach that accounts for receptor flexibility by allowing both the ligand and key residues in the binding site to adjust their conformations upon interaction. Unlike rigid-body and flexible docking, the IFD enables structural rearrangements, improving the prediction of binding modes and affinities. This method is particularly useful when docking ligands into highly flexible binding sites or when experimental structures may not fully capture induced conformational changes upon ligand binding.

IFD simulations were performed using the Maestro−Schrödinger suite, following a three-step protocol: (1) a standard docking was conducted using the OPLS3 force field and the SP scoring function, generating 20 poses per compound, (2) structural refinement was applied, allowing residues within 5.0 Å of the ligand to adopt alternative side-chain conformations to optimize interactions, and (3) redocking was performed, filtering poses based on an energy threshold of 30 kcal/mol from the most favorable pose per compound. The final output consisted of 10 poses per compound, which were used for further analysis.

2.4.5. Ensemble Docking. ED is an extension of standard molecular docking that accounts for receptor flexibility by considering multiple conformations of the target protein. Unlike conventional docking, which typically relies on a single rigid structure, ensemble docking leverages an ensemble of conformations—often derived from MD simulations or experimental structural variations. This approach improves docking accuracy by capturing the dynamic nature of the binding site, which may undergo conformational changes upon ligand binding. Compared to IFD, which refines protein side chains in response to ligand presence, ensemble docking explores global conformational diversity by sampling multiple pregenerated receptor states, offering a more comprehensive picture of potential binding interactions.

ED was performed by using 100 conformations obtained from long MD simulations of the four experimentally determined *HpU* structures (25 conformations per crystallographic structure). These receptor states were prepared using the same protocol as that described in Section 2.4.1 Standard Docking, ensuring consistency in grid generation and docking conditions. For each MD ensemble frame, docking was conducted using the Glide module in SP mode, generating 10 poses per compound per frame.

2.5. Data Analysis. All method outputs were exported as CSV files for further analysis. Data processing and statistical analyses were conducted using R within the RStudio environment, employing multiple libraries to evaluate the performance of SBVS methodologies and to compare predicted binding affinities with experimental IC_{50} values. Key libraries used included dplyr, ggplot2, tidyr, gridExtra, readxl, openxlsx, stringr, and networkD3. The analyses encompassed statistical correlation assessments, error metrics calculations, and ranking evaluations based on different Data Fusion methods.

To ensure reproducibility and facilitate adoption by other researchers, all data analysis workflows were implemented as *R*-Markdown reports, which are provided as Supporting Information and deposited in GitHub. These scripts are easily adapted to analyze different data sets, allowing researchers to apply the same methodological framework to their own VS benchmarks.

2.5.1. Unified Data Set Structure. To systematically evaluate the impact of different methodological choices, all processed data were integrated into a single structured data set containing the key variables listed in Table 2.

2.5.2. Evaluation of Energy-Based Parameters. In addition to evaluating the total binding energy, we analyzed specific energetic components to determine whether certain energy terms correlate more strongly with experimental values. The binding free energy was decomposed into individual components, such as electrostatic, van der Waals, hydrophobic, and solvation terms (Table S1). These components were tested against IC_{50} and pIC_{50} values to assess whether an alternative ranking criterion could enhance the VS reliability. If a specific energy parameter exhibits a stronger correlation with experimental values, it could be prioritized over traditional docking scores in molecular ranking workflows.

https://doi.org/10.1021/acsomega.5c04457
ACS Omega 2025, 10, 49641−49658

2.5.3. Data Fusion and Pose Aggregation. To enhance ranking accuracy and account for multiple docking poses per ligand, six data fusion methods were implemented (Table 3).

Table 3. Formulas Employed in This Study[a]

	metric	formula	R code
data fusion metric	minimum value	$\min(CE_1, CE_2 \ldots CE_n)$	
	arithmetic mean	$\frac{1}{n}\sum_{i=1}^{n} CE_i$	
	geometric mean	$\sqrt[n]{\prod_{i=1}^{n} CE_i}$	
	harmonic mean	$\left(\frac{\sum_{i=1}^{n} CE_i^{-1}}{n}\right)^{-1}$	
	euclidean mean	$\sqrt{\frac{\sum_{i=1}^{n} CE_i^{2}}{n}}$	
	median	$\text{median}(CE_1, CE_2 \ldots CE_n)$	
correlation and error	Spearman correlation	$1 - \frac{6\sum d_i^2}{n(n^2-1)}$	
	MAE	$\frac{1}{n}\sum_{i=1}^{n} \lvert CE_i - IC_{50,i}\rvert$	
	RMSE	$\sqrt{\frac{1}{n}\sum_{i=1}^{n} (CE_i - IC_{50,i})^2}$	
	IRM	$\frac{N_{\text{inliers}}}{N_{\text{total}}}$	

[a]CE: Computed energy.

Top-score (Minimum): Selects the *N*th best-scoring pose per ligand.

Median: Reduces the impact of outliers by selecting the central value.

Arithmetic mean: Averages of the scores across poses.

Geometric mean: Dampens extreme values while retaining overall trends.

Harmonic mean: Assign greater weight to stronger binding scores (lower values).

Euclidean Mean: Treats docking scores as multidimensional vectors, calculating an overall magnitude.

Each Data fusion method was applied incrementally from one pose to one M pose to assess the impact of pose aggregation on predictive performance. This systematic evaluation aimed to refine consensus scoring methodologies and improve molecular ranking strategies in VS campaigns.

2.5.4. IC_{50} vs pIC_{50}: Comparison of Experimental Variables. To determine whether IC_{50} or pIC_{50} serves as a reliable experimental reference for correlation analysis, we directly compared their Spearman and Pearson correlations with computational predictions. The difference between their correlation values was calculated as follows

$$\text{Spearman} = |\text{Spearman } IC_{50}| - |\text{Spearmanp } IC_{50}| \quad (1)$$

$$\text{Pearson} = |\text{Pearson } IC_{50}| - |\text{Pearsonp } IC_{50}| \quad (2)$$

This calculation not only identified differences but also assessed their directionality: whether IC_{50} or pIC_{50} yields stronger correlations with computed binding scores. A negative difference indicates that pIC_{50} outperforms IC_{50}, whereas a positive difference suggests that IC_{50} is superior. Additionally, to statistically validate whether pIC_{50} significantly improves or decreases the correlation coefficients over IC_{50}, paired *t* tests were performed by comparing the median Pearson and Spearman correlation values obtained for each combination of protein and data fusion method. These analyses were conducted both globally across all data fusion strategies and separately for each fusion method.

2.5.5. Interaction and Metal Coordination Analysis. To characterize the binding interactions between ligands and protein residues, we analyzed the .csv output files generated by the Glide docking module, which reports per-pose interaction annotations. Each interaction is classified in the Type field based on internal geometric criteria, involving both distance and angular thresholds specific to each interaction class. Hydrogen bonds are defined by a donor−acceptor distance between 2.5 and 3.5 Å and an angle between the donor, hydrogen, and acceptor atoms approaching 180°, ensuring linearity and optimal orbital overlap. Hydrophobic contacts are identified based on nonpolar atom proximity, within a 3.5−5.0 Å range, and are not subject to angular constraints. $\pi-\pi$ stacking interactions include both π-face (parallel) and π-edge (T-shaped) geometries. In π-face interactions, the centroid−centroid distance between aromatic rings falls between 3.3 and 4.0 Å, with ring planes in parallel alignment. In π-edge interactions, an ideal ∼90° angle between ring planes is expected, and the centroid-to-plane distance is also under 5.0 Å. $\pi-$Cation interactions are classified when a positively charged group (typically a quaternary ammonium or protonated amine) is located within 4.0−6.0 Å of the center of an aromatic ring. These interactions are directional but do not require strict angular thresholds. Salt bridges involve electrostatic interactions between oppositely charged groups such as a carboxylate and a protonated amine. Glide classifies these when the distance between the charged atoms is ≤4.0 Å, often around 3.0−3.7 Å. Halogen bonds occur when a halogen atom (acting as an electrophilic donor) interacts with a nucleophilic atom (oxygen, nitrogen, or sulfur). These interactions are highly directional, with ideal X···A distances (X = halogen, A = acceptor) between 2.8 and 3.5 Å and angles C−X···A close to 180°, consistent with the presence of a σ-hole. Metal coordination interactions are assigned when a ligand atom approaches a metal center (e.g., Ni^{2+}) within a distance consistent with coordination geometry, under 3.4 Å. While angle constraints vary depending on the specific metal and its preferred geometry (e.g., octahedral, square planar), Glide primarily evaluates distance-based criteria for scoring and pose ranking.

For each protein structure and its corresponding 25-frame ensemble, we conducted a systematic analysis of ligand−residue and ligand−metal interactions using a two-step computational pipeline. First, all annotated interactions were extracted from the Glide-generated. csv files using a custom Python script. This script parsed the Type and RecResidue fields to classify interactions (e.g., hydrogen bonds, $\pi-\pi$ stacking, metal coordination, salt bridges) and associate them with their corresponding residues. For each docking pose in each frame, interaction types were grouped by residue and counted, resulting in residue-wise frequency tables per frame. In the second step, a custom R script aggregated the interaction frequencies across all frames for each protein and applied a $\log_{10}$ transformation ($\log_{10}(1 + \text{count})$) to facilitate

https://doi.org/10.1021/acsomega.5c04457
ACS Omega 2025, 10, 49641−49658

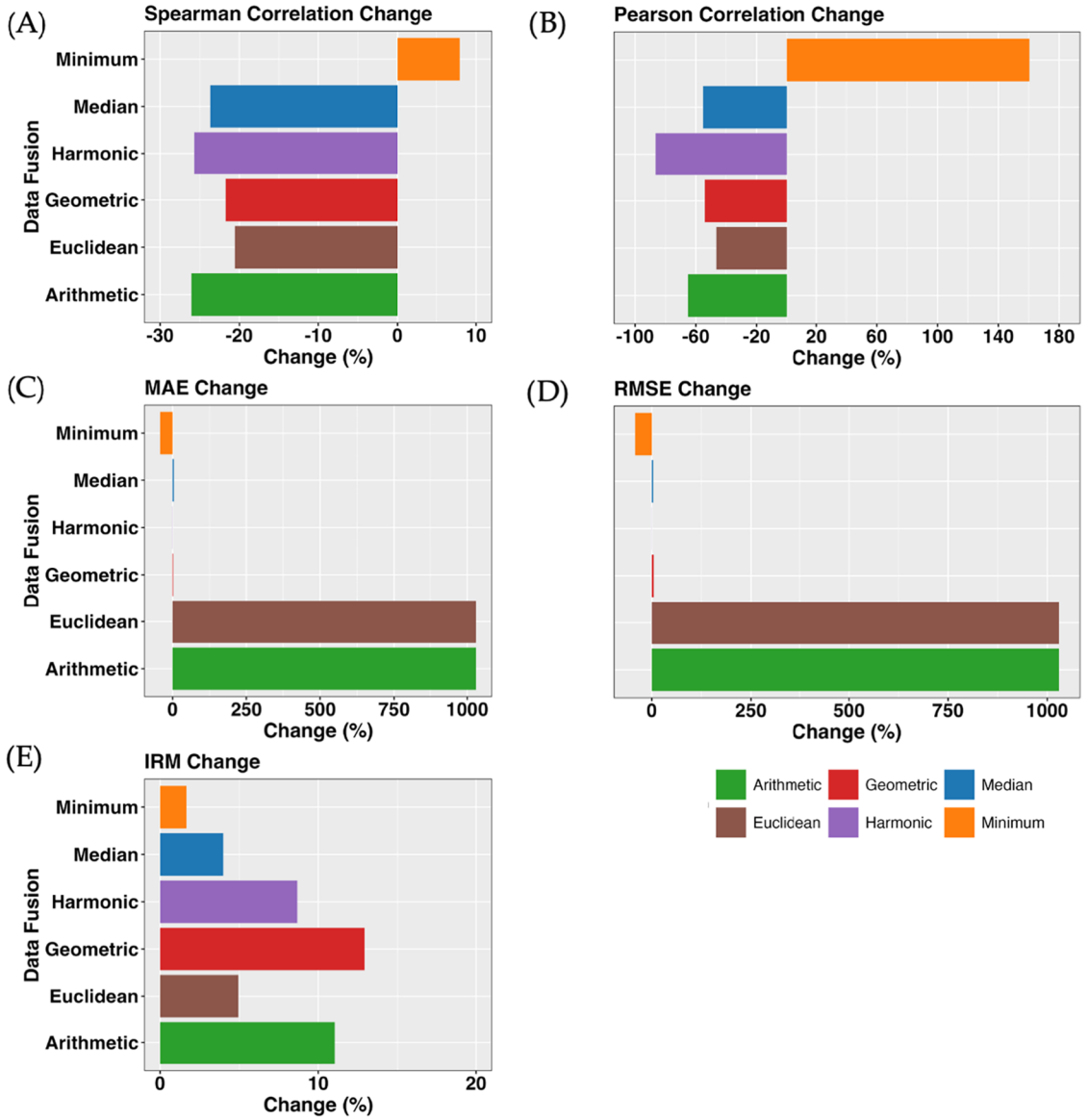


Figure 2. Impact of increasing the number of poses on different performance metrics across data fusion methods. The percentage change in five key metrics (Spearman correlation, Pearson correlation, MAE, RMSE, and IRM) is shown for different data fusion methods when increasing the number of poses from 10 to 100. (A) Spearman correlation change, indicating the variation in rank-order relationships between predicted and experimental values. (B) Pearson correlation change, reflecting the alteration in linear correlation strength. (C) MAE change, quantifying the absolute error deviation between predictions and experimental values. (D) RMSE change, highlighting error sensitivity to large deviations. (E) IRM change, illustrating the proportion of predictions that remain within the predefined acceptable threshold. Bars indicate the relative percentage changes for each data fusion method.

comparative visualization and mitigate the effects of outliers. The resulting data were used to construct heatmaps representing interaction frequencies per residue. To improve the interpretability of the heatmaps while maintaining a focus on biologically relevant contacts, we applied a residue filtering strategy based on interaction frequency. Specifically, we retained the top 12 interacting entities (residues or Ni^{2+} ions) per protein, and the interaction count corresponding to the 12th-ranked entry was used as a cutoff threshold. As this approach was applied independently to each structure, the resulting thresholds varied depending on the total interaction density observed: 10 000 for 1E9Y, 150 for 1E9Z, 7000 for 6QSU, and 11 000 for 6ZJA. The comparatively lower cutoff in 1E9Z reflects the reduced number of ligand−residue contacts observed across its ensemble, likely due to structural features such as reduced binding site accessibility.

In addition to frequency-based analyses, we computed the minimum distance between each ligand pose and the two Ni^{2+} ions for all complexes generated by standard and ensemble docking. For each docking pose, we calculated the shortest distance between any non-hydrogen atom of the ligand and each metal ion. This allowed us to assess whether energetically favorable poses also exhibited close proximity to the metal center, indicative of potential coordination-based interactions. Color-coded panels also highlight pose rank and conformational frame, where applicable.

Together, these analyses provide a spatial and energetic overview of ligand binding preferences, revealing that many top-ranking poses exhibit favorable interactions with catalytic

https://doi.org/10.1021/acsomega.5c04457
ACS Omega 2025, 10, 49641−49658

(A)

DataFusion	N	Mean	SD	Q1	Median	Q3	Outliers ≥ Q3	Outliers ≥ 0.9	Outliers = 1.0
Euclidean	14790	0.097	0.257	-0.063	0.128	0.264	3666	138	84
Geometric	14790	0.096	0.253	-0.057	0.123	0.255	3669	137	80
Median	14790	0.093	0.257	-0.064	0.126	0.257	3665	146	83
Arithmetic	14790	0.093	0.258	-0.065	0.122	0.258	3675	148	82
Harmonic	14790	0.089	0.251	-0.064	0.114	0.247	3671	125	75
Minimum	16052	0.076	0.247	-0.071	0.096	0.232	3989	151	138

(B)

Protein	DataFusion	N	Mean	SD	Q1	Median	Q3	Outliers ≥ Q3	Outliers ≥ 0.9	Outliers = 1.0
1E9Y	Arithmetic	4734	0.079	0.253	-0.143	0.082	0.286	1180	0	0
1E9Y	Euclidean	4734	0.087	0.250	-0.126	0.100	0.293	1184	0	0
1E9Y	Geometric	4734	0.089	0.245	-0.108	0.106	0.281	1183	0	0
1E9Y	Harmonic	4734	0.081	0.244	-0.115	0.092	0.273	1180	0	0
1E9Y	Median	4734	0.079	0.251	-0.138	0.094	0.286	1180	0	0
1E9Y	Minimum	5260	0.063	0.223	-0.112	0.059	0.239	1309	0	0
1E9Z	Arithmetic	1146	0.115	0.585	-0.377	0.194	0.524	287	148	82
1E9Z	Euclidean	1146	0.118	0.588	-0.376	0.200	0.546	286	138	84
1E9Z	Geometric	1146	0.117	0.575	-0.312	0.194	0.529	285	137	80
1E9Z	Harmonic	1146	0.103	0.576	-0.312	0.174	0.512	287	125	75
1E9Z	Median	1146	0.114	0.582	-0.321	0.198	0.543	286	146	83
1E9Z	Minimum	892	0.075	0.664	-0.429	0.174	0.556	222	151	138
6QSU	Arithmetic	4302	0.082	0.171	-0.033	0.087	0.212	1076	0	0
6QSU	Euclidean	4302	0.084	0.174	-0.030	0.092	0.219	1072	0	0
6QSU	Geometric	4302	0.082	0.171	-0.038	0.085	0.207	1071	0	0
6QSU	Harmonic	4302	0.076	0.169	-0.040	0.078	0.197	1073	0	0
6QSU	Median	4302	0.084	0.174	-0.034	0.089	0.220	1063	0	0
6QSU	Minimum	4780	0.063	0.184	-0.058	0.070	0.194	1191	0	0
6ZJA	Arithmetic	4608	0.111	0.186	-0.003	0.164	0.252	1144	0	0
6ZJA	Euclidean	4608	0.115	0.184	0.002	0.163	0.254	1150	0	0
6ZJA	Geometric	4608	0.110	0.182	-0.008	0.159	0.250	1141	0	0
6ZJA	Harmonic	4608	0.105	0.179	-0.018	0.154	0.243	1146	0	0
6ZJA	Median	4608	0.112	0.186	-0.005	0.163	0.250	1142	0	0
6ZJA	Minimum	5120	0.101	0.175	-0.017	0.139	0.233	1277	0	0

Figure 3. Analysis of data fusion performance across proteins. (A) Summary statistics of Spearman correlation values across different Data Fusion methods, considering all proteins collectively. The table reports the mean, standard deviation (SD), quartiles (Q1, median, and Q3), and number of data points (N) for each Data Fusion approach. The last three columns highlight the presence of outliers, including values exceeding the third quartile (Q3), those above the 0.9 threshold, and cases in which the correlation is exactly 1.0. (B) Breakdown of Spearman correlation statistics per protein, illustrating the variability in Data Fusion performance across different systems. Each row corresponds to a specific Protein-Data Fusion combination, with color intensity indicating the relative magnitude.

residues or approach the metal center, while others adopt plausible alternative binding modes, including flap-loop occlusion.

3. RESULTS AND DISCUSSIONS

3.1. Is Considering More Poses Better? Understanding the impact of increasing the number of poses on the predictive performance is critical in molecular modeling and docking studies. Given the ligand conformation variability and scoring function uncertainty, determining whether additional poses enhance or hinder prediction accuracy is essential. To address this, we evaluated the rate of change in correlation and error metrics when considering 10 poses versus 100 poses across different data fusion methods.

We specifically analyzed Spearman and Pearson correlations, which assess the ranking and linear relationship between predicted and experimental values, respectively. Additionally, to quantify potential increases in prediction error, we examined the MAE, RMSE, and IRM. These error-based metrics provide insight into the reliability of predictions when more poses are included. To evaluate the impact of increasing the number of poses from 10 to 100, we computed the percentage change for each metric using the following formula

https://doi.org/10.1021/acsomega.5c04457
ACS Omega 2025, 10, 49641−49658

$$\text{change} = \frac{\text{metric}_{100\text{poses}} - \text{metric}_{10\text{poses}}}{|\text{metric}_{10\text{poses}}|} \times 100 \quad (3)$$

This normalization expresses the variation in each metric as a percentage relative to its initial value at 10 poses, making it easier to compare across different data fusion methods.

The Spearman correlation (Figure 2A), which evaluates the monotonic relationship between experimental and predicted values, followed a pattern similar to that of Pearson. All fusion methods except the minimum showed a decline with an increasing number of poses. The largest decrease was observed in the Median method (−23.67%), followed by harmonic (−25.69%) and arithmetic (−26.06%). The minimum fusion approach was the only exception, showing a modest improvement of +7.95%, suggesting that selecting the lowest-energy pose per ligand preserves rank-order relationships better than averaging across multiple poses. The Pearson correlation (Figure 2B), which measures the linear relationship between the predicted and experimental values, exhibited an even more pronounced decline. The Harmonic mean suffered the largest drop (−86.53%), followed by arithmetic (−65.18%) and geometric (−54.09%). Interestingly, the minimum method again demonstrated a substantial improvement (+160.51%), further reinforcing the idea that selecting the lowest-energy pose leads to better correlation with experimental values at 100 poses.

These findings suggest that increasing the number of poses generally introduces noise rather than enhances predictive power. In Spearman correlation, beyond a certain threshold, additional poses do not necessarily preserve the rank-order relationships. The minimum method, however, appears to benefit from a larger sampling, likely because it prioritizes the most energetically favorable conformation, filtering out less meaningful poses that could introduce inconsistencies. While correlation metrics provide insight into the consistency of predictions, error-based measures allow us to assess their absolute accuracy. To determine whether considering more poses improves or worsens precision. MAE (Figure 2C) measures the absolute deviation between predicted and experimental values. The Arithmetic and Euclidean fusion methods exhibited the largest increases in MAE (+1029.65%) with more poses, followed by median (+4.05%) and geometric (+1.95%), indicating a significant decline in prediction accuracy. Conversely, the minimum method showed a remarkable decrease (−42.05%), suggesting that its approach of selecting a single best pose reduces the absolute prediction error. RMSE (Figure 2D), which is more sensitive to large deviations than MAE, mirrored the same trend. The Arithmetic and Euclidean fusion methods suffered the largest increases in RMSE (+1029.65%), whereas the minimum method exhibited the largest improvement (−42.05%), reinforcing its superiority in reducing extreme errors when considering more poses. IRM (Figure 2E) evaluates the proportion of predictions that fall within an acceptable error threshold. Unlike the other metrics, all fusion methods exhibited an increase, suggesting that more poses slightly improve the likelihood of predictions falling within the accepted range. The geometric method had the highest IRM increase (+12.92%), followed by arithmetic (+11.04%) and harmonic (+8.67%), whereas the minimum method showed a more modest increase (+1.65%).

3.2. How Do Data Fusion Methods Behave Across Different Proteins? A preliminary exploratory analysis of the data set reveals a lack of strong Spearman correlations across all Data Fusion methods when considering the combined data from four proteins. This trend persists across all evaluated energy terms, pose aggregation methods (up to 10 poses), computational methods, and protocol variants. Figure 3A summarizes these findings by presenting statistical descriptors of Spearman correlation across data fusion methods without differentiating between proteins.

The overall mean Spearman correlation across all data fusion methods is 0.092, with a median value of 0.126. The third quartile (Q3), representing the upper 25% of the correlations, does not exceed 0.264, reinforcing the weak overall correlations. In terms of sample sizes (N), each data fusion method typically includes approximately 14 790 samples, except for the Minimum approach, which has a slightly larger data set (16 052 samples).

Despite the low overall correlations, Figure 3A highlights the presence of high-correlation outliers in the last three columns. The number of instances where the Spearman correlation surpasses the third quartile (Outliers ≥ Q3) is relatively uniform across most data fusion methods, with values around 3665−3675. However, the Minimum method shows a slightly higher count (3989 instances), suggesting that this method might capture favorable correlations in a small subset of cases. When considering extreme values (outliers ≥ 0.9), the overall frequency remains low, ranging between 125 and 151 occurrences, again with the Minimum method exhibiting the highest count (151 instances). Cases where Spearman correlation reaches exactly 1.0 are rarer, with a maximum count of 138 cases in the minimum approach, while all other methods remain below 84 occurrences. These findings suggest that although strong correlations are rare, the minimum fusion method appears to be the only approach consistently capturing a meaningful subset of high-correlation cases.

While Figure 3A aggregates correlations across all proteins, Figure 3B provides a per-protein breakdown, revealing substantial variability across individual data sets. The first major observation is that some proteins (e.g., 1E9Z) exhibit noticeably higher correlation variability, as evidenced by their standard deviations (SDs) reaching 0.664 for the minimum fusion method, whereas other proteins remain below 0.258 across different methods. This suggests that for certain proteins, the fusion approach significantly influences correlation variability.

The sample sizes (N) also vary by protein, with some data sets being significantly smaller. For example, 1E9Z has only 1146 samples per method, while others, such as 6QSU and 6ZJA, have 4302 and 4608 samples, respectively. These differences suggest that data set size may influence the stability of correlation estimates. Interestingly, the Minimum method again shows the highest number of extreme outliers (Spearman ≥ 0.9 or = 1.0), particularly in 1E9Z, reinforcing the trend seen in Figure 3A.

Interestingly, despite its lower N, 1E9Z exhibits the highest number of Spearman correlations above 0.9, setting it apart from the other proteins. This might suggest an apparently superior predictive power for this protein when using certain Data Fusion methods. However, this finding should be interpreted with caution. As highlighted in Figure S1 and Table 2, the restricted binding site volume in 1E9Z may artificially limit ligand flexibility, reducing the range of docking scores and inflating correlations. Thus, while high Spearman correlations for 1E9Z may appear promising, they could be a

https://doi.org/10.1021/acsomega.5c04457
ACS Omega 2025, 10, 49641−49658

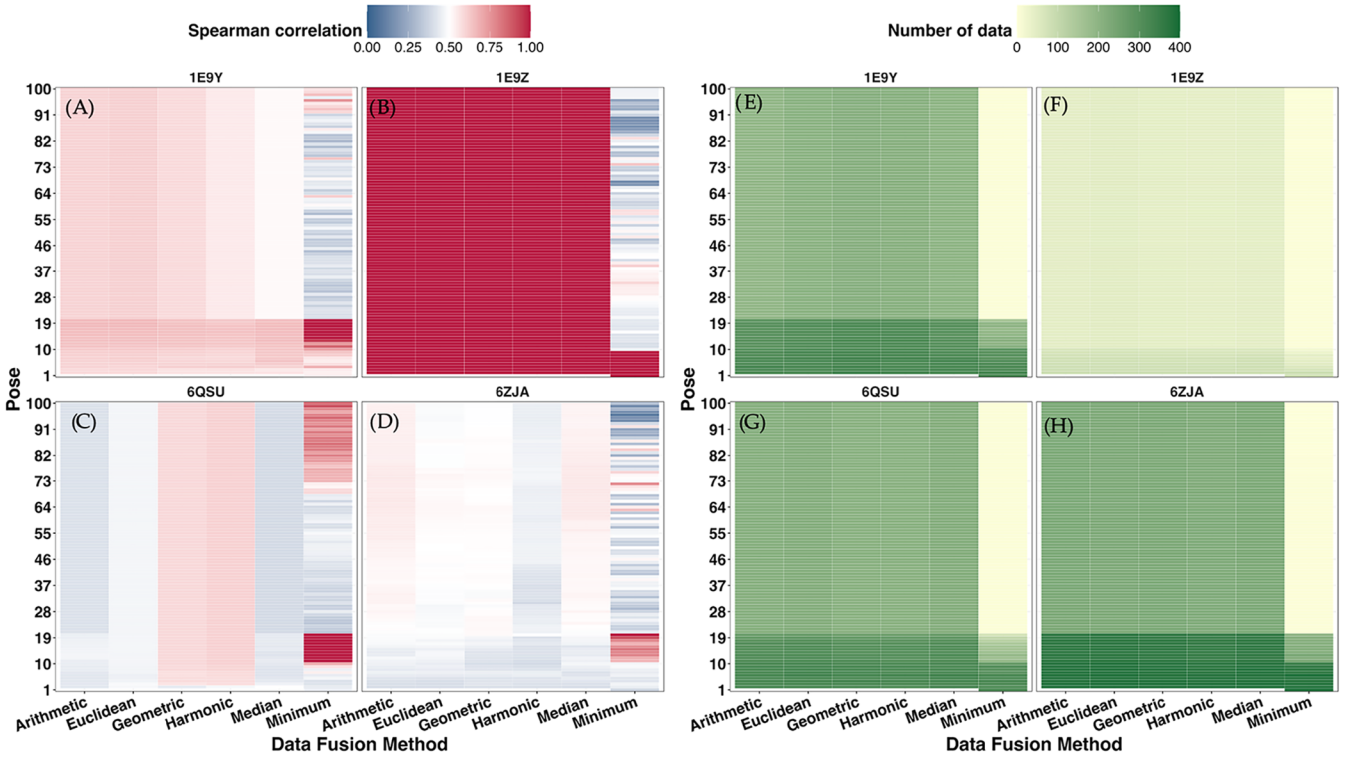


Figure 4. Impact of data fusion and pose aggregation on Spearman correlations and data count across proteins. (A−D) Heatmaps depicting the maximum Spearman correlation per pose for each data fusion method across the four proteins analyzed (1E9Y, 1E9Z, 6QSU, and 6ZJA). The color gradient from blue to red denotes increasing Spearman correlation values, with red indicating the strongest correlations. (E−H) Corresponding heatmaps showing the number of data points (*N*) used for each correlation calculation, where darker green indicates a higher number of data points.

consequence of reduced variability rather than a true improvement in the prediction accuracy.

3.3. Pose Aggregation and Data Fusion: Identifying Optimal Strategies Across Proteins. The results presented in Figure 4 reinforce the findings from previous analyses (Sections 3.1 and 3.2) regarding the impact of pose aggregation and data fusion strategies in the context of IC_{50}-based Spearman correlations. This analysis specifically focuses on positive Spearman correlations, as previous evaluations confirmed the absence of negative correlations. Figure 4 is structured into two main parts: (A−D) illustrate the correlation strength across poses, data fusion methods, and proteins, while (E−H) provide insights into the number of data points used for correlation calculations under the same conditions.

A key observation from Figure 4A−D is that 1E9Z achieves the highest Spearman correlation across multiple poses and data fusion methods. This is particularly evident in Figure 4B, where nearly all data fusion methods attain strong correlations (>0.75) across a broad range of poses. However, Figure 4F highlights that 1E9Z has the lowest number of data points, raising concerns about the robustness of these correlations. The Minimum fusion method, in particular, shows strong correlations at low pose counts, but its performance deteriorates as more poses are included. This suggests that 1E9Z's high correlations may be artificially inflated due to its smaller data set, limiting the generalizability of these findings.

Unlike 1E9Z, 1E9Y appears to be the most balanced protein, achieving moderate to high Spearman correlations across most data fusion methods (Figure 4A). The data set size for this protein is also considerably larger (Figure 4E), suggesting that its performance is not biased by data scarcity. However, a notable trend in 1E9Y is the gradual decline in correlation as the number of poses increases, suggesting that excessive pose aggregation introduces noise rather than enhances predictive power. The minimum fusion method follows an unusual pattern in this protein, displaying strong correlations at both high pose counts (≥90 poses) and low pose counts (≤20 poses) but a weaker performance at intermediate values. This suggests that Minimum fusion is highly sensitive to the number of poses used, resulting in inconsistent performance.

In contrast to 1E9Y and 1E9Z, proteins 6QSU and 6ZJA demonstrate the weakest correlation performance overall (Figure 4C,D). In 6QSU, only the Geometric and Harmonic fusion methods achieve appreciable correlations, while other fusion strategies struggle to surpass 0.5 Spearman correlation. The data count for 6QSU is relatively high (Figure 4G), suggesting that poor correlation performance is likely an intrinsic characteristic of this system rather than a consequence of limited data. Similarly, 6ZJA exhibits a widespread lack of strong correlations across all data fusion methods, reaffirming that this protein may be inherently difficult for predictive modeling.

Across all four proteins, the Minimum fusion method presents a highly variable performance. It achieves the highest correlations in 1E9Z but performs inconsistently in 1E9Y and 6QSU, showing strong correlations at extreme pose values but weak correlations in intermediate ranges. This irregularity may stem from Minimum fusion's reliance on the lowest-energy pose, which can represent either a biologically meaningful conformation or a statistical artifact due to pose selection bias. Importantly, Minimum fusion uses fewer data points compared

https://doi.org/10.1021/acsomega.5c04457
ACS Omega 2025, 10, 49641−49658

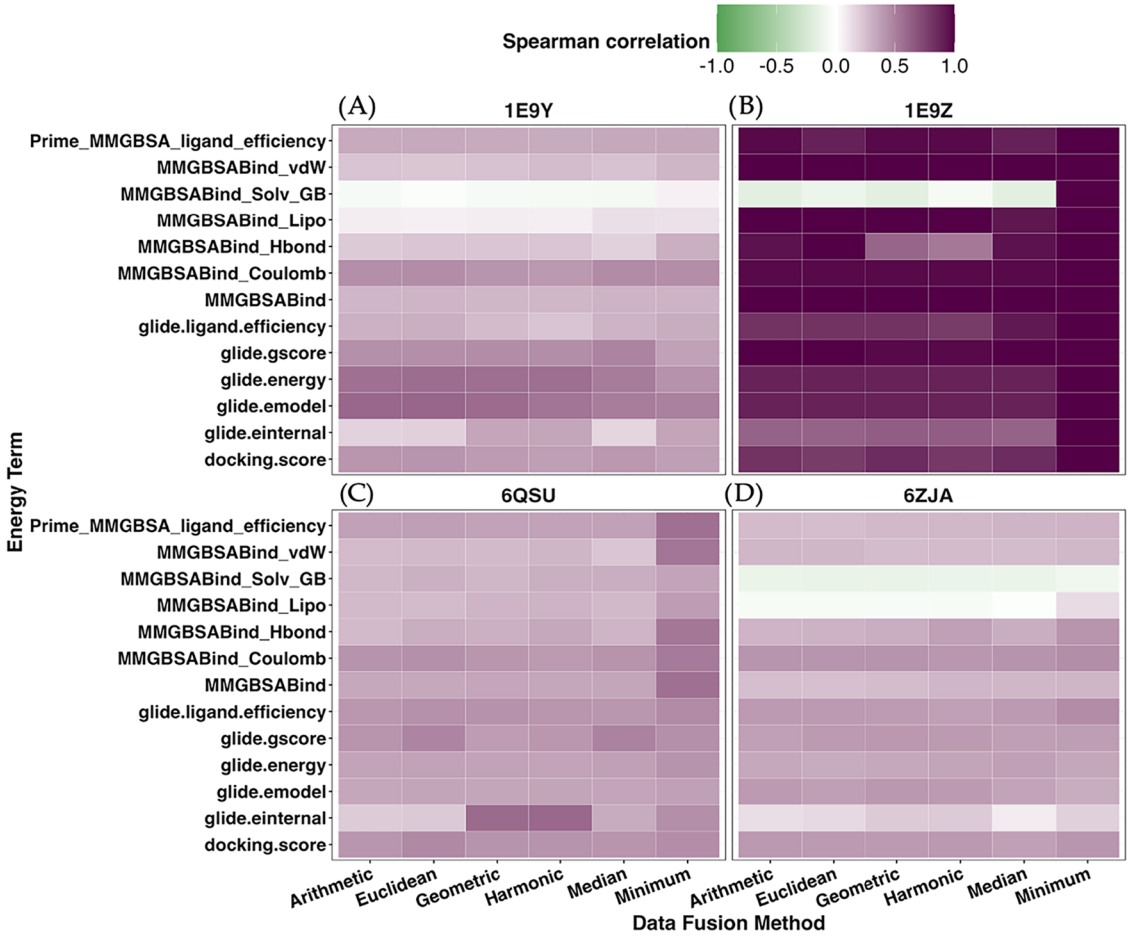


Figure 5. Impact of individual ET on Spearman correlation across proteins and data fusion methods. The heatmap presents the highest Spearman correlation values obtained for each ET (*y*-axis) across different Data Fusion Methods (*x*-axis) for each protein (A−D).

to other fusion methods (Figure 4E−H), meaning its results should be interpreted with caution.

The heatmaps in Figure 4A−D further underscore the importance of pose aggregation strategies. In 1E9Y, most data fusion methods maintain a stable correlation up to 20 poses, but performance declines beyond this threshold. This suggests that including additional poses beyond a certain point introduces greater variability rather than enhancing correlation strength. In 6QSU, the Geometric and Harmonic methods exhibit moderate performance, reinforcing the previous finding that these fusion approaches are less sensitive to pose aggregation bias.

3.4. Evaluation of Individual Energy Terms. Understanding the role of individual energy terms (ETs) in achieving high Spearman correlations is crucial for refining molecular docking and scoring methodologies. Figure 5 systematically evaluates the impact of each ET across the four studied proteins and data fusion strategies.

As expected, 1E9Z consistently exhibits the highest Spearman correlations across multiple ETs, reinforcing the previous findings that suggest a favorable structural context for strong correlations in this system. In contrast, 6ZJA and 6QSU display lower correlation values, highlighting potential challenges in capturing binding trends in these proteins. 1E9Y represents an intermediate case, where certain ETs achieve notable performance, albeit with greater variability across fusion methods.

A key question in docking studies is whether docking score, the default metric widely employed in the literature, provides robust predictive performance. The results indicate that while docking.score maintains moderate correlations (∼0.5) across proteins; it does not consistently outperform alternative docking-related ETs such as glide.gscore and glide.energy, whose correlations vary depending on the protein and fusion method. Notably, no single docking-based ET consistently outperforms across all proteins or data fusion methods, suggesting that relying on a single scoring function may be suboptimal for generalizable predictions.

Among the MM-GBSA energy terms (from Prime_MMGBSA_ligand_efficiency to MMGBSA_Bind), the Coulombic energy emerges as a key descriptor. This term exhibits strong correlations across all four proteins and all data fusion strategies, suggesting that electrostatic contributions to the binding energy play a pivotal role in capturing meaningful experimental trends. This finding highlights the importance of considering electrostatic interactions beyond traditional docking scores, particularly in systems in which charge complementarity is critical.

Conversely, Solv_GB and Lipo exhibit the weakest correlations among all analyzed terms, except in 6QSU, where they show a slightly better performance. The poor performance of Solv_GB (generalized Born solvation energy) and Lipo (lipophilic interactions) suggests that these components may not independently capture ligand-binding

https://doi.org/10.1021/acsomega.5c04457
ACS Omega 2025, 10, 49641−49658

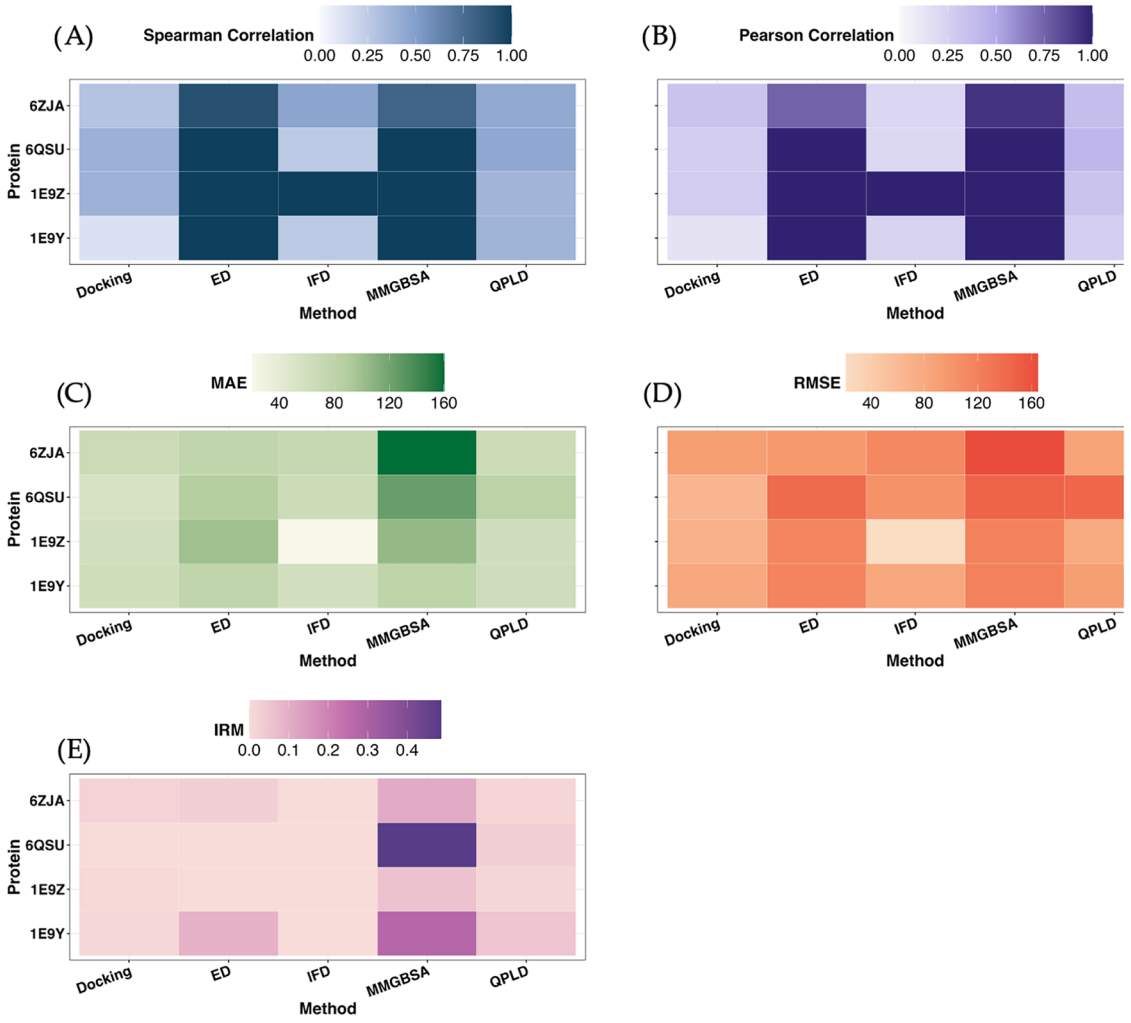


Figure 6. Comparative analysis of VS methodologies using five evaluation metrics: (A) Spearman correlation, (B) Pearson correlation, (C) MAE, (D) RMSE, and (E) IRM. The analysis was performed considering up to 20 poses per compound and applying the Minimum data fusion approach. Higher correlation values (A, B) indicate better agreement between predicted and experimental IC_{50} values, while lower error values (C, D) reflect greater predictive accuracy. Higher values in (E) indicate a greater proportion of correctly ranked active compounds within the top-scoring predictions, reflecting the method's ability to enrich true positives in virtual screening.

trends effectively. However, improved performance in 6QSU, indicates that hydrophobic effects might play a more significant role in this particular system.

3.5. Methods Comparison in VS: Performance Across Metrics. Figure 6 evaluates the performance of different computational methods, considering up to 20 poses per compound and using the Minimum data fusion approach, as supported by previous analyses.

Panels 6A and 6B illustrate the Spearman and Pearson correlation coefficients, respectively, providing a direct measure of each method's alignment with experimental IC_{50} values. ED and MM-GBSA consistently achieve the highest correlation across all four proteins, suggesting that they are the most reliable for ranking ligands by binding affinity. IFD, while displaying strong correlation values in 1E9Z, does not maintain a robust performance across the other proteins, likely due to its higher computational cost and increased sensitivity to structural variations. Notably, the case of 1E9Z requires careful interpretation, as previously discussed, since this protein's binding site remains structurally closed, reducing the number of generated poses compared to the other proteins, which may artificially inflate correlation values.

Error-based metrics, presented in panels 6C and 6D, further contextualize these results by evaluating the absolute and squared deviations from experimental IC_{50} values. MM-GBSA exhibits the highest MAE and RMSE values across all proteins, indicating that while it ranks compounds effectively, its absolute binding affinity predictions are less accurate. ED also shows relatively high MAE and RMSE values, though to a lesser extent than MM-GBSA, suggesting that despite its strong ranking performance, this method may introduce systematic error. Conversely, IFD and QPLD present lower overall error values, but given their lower correlation scores (particularly for QPLD), these methods may provide a better approximation of IC_{50} magnitudes at the cost of weaker ranking capabilities.

Panel 6E presents the IRM, which evaluates how well a method ranks the top-performing compounds relative to the entire data set. MM-GBSA stands out as the method with the highest IRM values across all proteins, indicating that it prioritizes active compounds more effectively than the other methodologies. ED follows closely, reinforcing its utility in ranking high-affinity ligands, while IFD and QPLD show moderate IRM values, further enhancing their limited generalizability across different protein targets.

https://doi.org/10.1021/acsomega.5c04457
ACS Omega 2025, 10, 49641−49658

Taken together, these findings suggest that ED and MM-GBSA are the most effective methodologies for VS, balancing a strong correlation with experimental data and reasonable ranking accuracy, though MM-GBSA suffers from higher prediction errors. IFD demonstrates competitive performance in 1E9Z but lacks consistency across other proteins, while QPLD and standard docking methods do not emerge as top-performing approaches in any of the evaluated metrics. The observed trade-offs among ranking ability, correlation strength, and absolute error underscore the importance of careful method selection based on the specific objectives of a VS campaign. Future work could focus on optimizing hybrid approaches, leveraging the superior ranking performance of MM-GBSA and ED while mitigating their error-prone estimations of binding affinities.

3.6. Impact of Protocol Variants on VS Performance.

Figure S2 presents a comparative analysis of five VS variants across four target proteins, using five evaluation metrics: Spearman correlation (A), Pearson correlation (B), MAE (C), RMSE (D), and IRM (E). This comparison offers insights into the relative performance of each protocol and highlights how different docking and scoring strategies impact predictive accuracy. Notably, variants 3, 4, and 5 are based on ED, whereas variants 1 and 2 follow alternative methodological pathways, as detailed in the methodological flowchart (Figure 1).

Spearman (A) and Pearson (B) correlations indicate that Variants 3 and 4 consistently outperform other protocols, reinforcing the importance of ED as an effective VS strategy. However, their relative ranking depends on the correlation metric: Variant 3 shows superior Spearman correlation across most proteins, while Variant 4 excels in the Pearson correlation. This divergence suggests that Variant 3 better preserves ranking relationships between predicted and experimental values, whereas Variant 4 more accurately models absolute binding energy values. Variant 2 also exhibits strong correlations but is notably weaker than Variants 3 and 4, particularly because its high correlation is observed in only two of the four proteins. This highlights a limitation in its generalizability across different targets. Conversely, Variant 5 consistently exhibits the lowest correlation values, indicating that its approach, combining ED and QPLD, does not significantly enhance predictive power. Additionally, the gray squares in Variant 5 for 1E9Z indicate missing data, which stem from the failure of QPLD to generate binding poses for this protein due to its highly enclosed binding site. This reinforces previous observations that 1E9Z's closed binding site restricted pose generation across multiple methodologies, impacting its evaluation in both method and variant comparisons.

The MAE (C) and RMSE (D) heatmaps reveal that variant 4 has the highest error values, closely followed by variant 3. This suggests that although variant 4 achieves high Pearson correlations, its predictions deviate more from the experimental IC_{50} values in absolute terms. This trade-off suggests that while MM-GBSA improves correlation, it also increases absolute prediction errors, consistent with Figure 6, where MM-GBSA exhibited the highest error rates. Interestingly, variants 2 and 5 show the lowest MAE values, suggesting greater numerical accuracy. However, since these variants also have the lowest correlation values, this indicates that while their predictions are numerically close to experimental IC_{50}s, they fail to capture the correct rank ordering of compounds. This underscores a critical trade-off: a method with a lower MAE/RMSE may not be the best choice if it fails to rank compounds in virtual screening.

IRM (E) highlights that variant 4 achieves the highest IRM values across all proteins, while variants 3 and 5 have the lowest. This suggests that variant 4 not only produces high correlation values but also maintains strong classification ability, which is crucial for practical VS applications. Notably, the IRM ranking aligns well with the previous method-based analysis (Figure 6), where MM-GBSA (dominant in variant 4) showed the highest IRM values. This suggests that although MM-GBSA introduces greater error (higher MAE/RMSE), it enhances classification robustness, making it a viable option when prioritizing hit identification over absolute binding energy accuracy.

3.7. Frame Dependence Across Variants 3, 4, and 5.

Figure S3 presents an in-depth analysis of the Spearman correlation as a function of frame selection across variants 3, 4, and 5 in the four proteins, considering only the "docking.score" or "MMGBSABind_Coulomb" energy terms. These terms were selected based on the findings from Figure 5 (Section 2.4), where these two energy terms emerged as the most reliable and consistently predictive across multiple proteins and data fusion methods. The heatmap illustrates how correlation values fluctuate across frames, emphasizing the impact of frame selection on the predictive performance. The results show that while some proteins exhibit relatively stable correlations, others display a strong frame-dependent variation, with specific frames showing markedly high or low correlation values.

Among the four proteins, 1E9Y exhibits the highest stability and consistently high Spearman correlation values across all variants. Most frames in this protein show strong correlations, suggesting that frame selection has a minimal impact on predictive performance. However, localized variations in the correlation strength indicate some degree of instability in certain frames.

Conversely, 1E9Z exhibits an extreme polarization of correlation values, while most frames show very strong correlations, while others display negative correlations. This pronounced dichotomy suggests that while certain frames yield highly predictive results, others introduce artifacts or noise, potentially distorting the analysis. Additionally, due to its structural constraints and highly enclosed binding site, 1E9Z failed to generate poses in any frame for variant 5, resulting in the empty white space in the heatmap. A similar pattern appears in specific frames of the other three proteins, where missing values indicate the absence of generated poses. This further underscores the role of structural constraints in modulating docking pose availability and its impact on correlation values.

The 6QSU protein exhibits a heterogeneous correlation pattern, characterized by substantial variability across frames and variants. While some frames achieve high Spearman correlations, the overall trend suggests a higher sensitivity to frame selection compared to those of 1E9Y and 1E9Z. This variability implies that frame selection in 6QSU may play a crucial role in optimizing predictive performance. Conversely, 6ZJA displays moderate correlation values across most frames without a clear dominance of highly correlated or uncorrelated regions. This suggests that 6ZJA is less responsive to frame selection than the other proteins, making it a less predictable system in terms of correlation stability.

https://doi.org/10.1021/acsomega.5c04457
ACS Omega 2025, 10, 49641−49658

A key finding across all four proteins is that variant 4 consistently achieves the highest correlations, outperforming variants 3 and 5. This trend is particularly evident in 1E9Y, 1E9Z, and 6QSU, where variant 4 consistently outperforms the other variants across most frames. However, in 6ZJA, this pattern does not hold, as variant 4 fails to exhibit the strongest correlations, suggesting that this protein behaves differently from the others. The underlying reasons for this discrepancy require further investigation, potentially linked to differences in ligand binding modes or the flexibility of the active site in 6ZJA.

A comprehensive analysis of frame selection reveals important nuances in how structural sampling impacts predictive performance. It is crucial to distinguish this concept from the discussion in Section 3.1 regarding pose sampling within a single docking run. Whereas increasing the number of poses beyond a certain point introduces noise by accumulating redundant or suboptimal ligand conformations, selecting diverse receptor frames from MD simulations can be beneficial, provided that only the most informative frames are included. Our results show that considering all frames together (All_Frames) does not necessarily yield higher Spearman correlations compared to selecting a single representative frame. While ensemble approaches theoretically leverage structural diversity, indiscriminately incorporating all frames may introduce noise or dilute meaningful correlations rather than enhance the predictive performance. These findings underscore the need for a refined selection process that identifies an optimal subset of frames, potentially offering a more effective strategy for ensemble-based VS.

Future research should explore methodologies that dynamically select the most informative frames using statistical or machine-learning-driven approaches, potentially improving the robustness and accuracy of ensemble docking strategies. It is important to note that the 1E9Z system behaves as an outlier in our analyses. Its highly enclosed binding site limits ligand flexibility and reduces the diversity of the docking poses, which may artificially elevate correlation values. Although 1E9Z demonstrated strong predictive performance for some protocols, these results should not be generalized. Therefore, our overall conclusions regarding the effectiveness of different VS methodologies are primarily based on trends observed across all four proteins rather than on the exceptional performance of this particular system.

3.8. Comparison of IC_{50} and pIC_{50} as Experimental Parameters for Correlation Analysis. Figure S4 presents a comparative evaluation of IC_{50} and pIC_{50} across different data fusion methods, assessing their impact on the Spearman and Pearson correlations for each protein. Differences were initially quantified by calculating the absolute difference in correlation coefficients, which allowed visualization of trends and directionality. Negative values indicated higher correlations for pIC_{50}, while positive values suggested that IC_{50} performed better. Across all proteins, Pearson correlations were consistently higher for pIC_{50}, supporting the idea that the log transformation reduces variability and improves linear relationships with predicted binding energies. In contrast, differences in Spearman correlations appeared to be small in magnitude, suggesting that rank-order relationships were less affected by the transformation.

To statistically validate these observations, we performed paired t tests comparing median correlation coefficients obtained for IC_{50} and pIC_{50} across all proteins and data fusion methods (Table S2). The results confirmed that pIC_{50} yields significantly higher Pearson correlations globally (mean difference = 0.391, t = 11.98, $p < 2.29 \times 10^{-11}$) and across all individual data fusion methods, although the magnitude of improvement varied. Notably, the Minimum fusion method exhibited the largest mean difference (0.500) for Pearson correlations.

Interestingly, while differences in Spearman correlations appeared small in absolute terms in Figure S4, the paired t tests revealed these differences to be statistically significant, albeit of smaller magnitude than Pearson. Globally, Spearman correlations were significantly higher for pIC_{50} (mean difference = 0.714, t = 12.14, $p < 1.77 \times 10^{-11}$), with consistent significance across all fusion methods. These findings indicate that log transformation of IC_{50} to pIC_{50} not only improves linear correlations but also subtly enhances rank-based associations, supporting the use of pIC_{50} as the preferred experimental parameter in virtual screening workflows.

3.9. Selection of Optimal Protocol Configurations for VS Performance. As a final step in the methodological evaluation, the most frequently occurring parameter combinations associated with high correlation performance were identified (Table S3). Specifically, we quantified the frequency of unique tuples (data fusion, energyterm, method, variant) that achieved Spearman correlation values above 0.75 when IC_{50} was used as the experimental variable. This analysis aimed to establish practical guidelines for selecting the most robust parameter configurations in VS workflows.

The results indicate that MM-GBSA-based scoring functions dominate among the highest-performing tuples, with MMGBSABind_Coulomb appearing in multiple configurations. The Arithmetic and Euclidean data fusion methods are among the most frequently observed across different energy terms, suggesting that these aggregation strategies enhance consistency in ranking active compounds. Additionally, variant 4 emerges as the most recurrent, supporting previous findings that ED followed by MM-GBSA rescoring provides superior correlation with experimental IC_{50} values. Interestingly, variant 2 also appears frequently, supporting the idea that induced-fit docking (IFD) followed by MM-GBSA is a viable alternative for cases in which ED is computationally prohibitive. However, while variant 4 exhibits a strong performance, it also shows variability in the choice of energy terms, particularly in the inclusion of MMGBSABind_Hbond and MMGBSABind_Lipo, which may introduce method-dependent fluctuations in ranking accuracy. The frequency analysis further highlights that geometric and median data fusion approaches contribute to a high correlation performance, particularly when paired with Coulomb-based MM-GBSA calculations. These findings further support the robustness of Coulombic interactions as key determinants in the binding affinity estimation.

3.10. Analysis of Protein−Ligand and Metal Interactions. In addition to scoring evaluations, all docking poses were analyzed to compute residue−ligand and metal−ligand interactions across the four protein structures and their 25-frame ensembles. Heatmaps of interaction frequencies (Figures S5−S8) were generated to identify both amino acid residues and Ni^{2+} ions most frequently engaged by the ligands. Our analyses revealed that several coumarin derivatives exhibited direct interactions with the binuclear Ni^{2+} center, suggesting possible coordination-driven binding modes similar to those observed in the crystal structures of known urease inhibitors. However, numerous compounds also engaged in alternative

https://doi.org/10.1021/acsomega.5c04457
ACS Omega 2025, 10, 49641−49658

binding patterns, involving key residues such as His322, His274, Asp362, and Ala169, all of which are known from crystallographic data to play crucial roles in urease catalysis, metal coordination, and inhibitor binding. Given the absence of crystallographic data for coumarins bound to urease and the possibility of alternative inhibition mechanisms, such as allosteric effects or flap-loop stabilization, we intentionally did not discard docking poses that did not directly interact with key residues or with the nickel ions, in order to avoid bias and to preserve the blind, exploratory nature of our VS. Notably, our analyses confirm that many of the observed ligand−protein contacts align with residues previously identified as critical in urease (PDB IDs 6ZJA, 1E9Y), supporting the biological plausibility of the docking predictions. To complement the interaction frequency analysis, we calculated the distances between the heavy atoms of each ligand pose and the two Ni^{2+} ions in the active site across all docking simulations. This analysis was performed for both the four crystallographic structures and the 100-conformation ensemble docking experiments. In these plots (Figure S9A−H), we observed that in most cases, particularly for 1E9Y, 6QSU, and 6ZJA, poses with more favorable docking energies tend to be positioned closer to the nickel ions, often within 5 Å, indicating potential coordination-driven interactions. Even in the more enclosed binding site of 1E9Z, the lowest-energy poses were those closer to the Ni^{2+} center, albeit with higher overall distances due to the closed conformation of the active site. These findings support the notion that nickel coordination may contribute significantly to the binding affinity of many coumarin derivatives. However, given the absence of experimental crystallographic data for coumarins bound to urease and the potential for alternative inhibition mechanisms, we intentionally did not discard poses solely on the basis of metal coordination distances. Docking algorithms inherently penalize poses that do not favorably engage the catalytic site, thus reducing the likelihood of retaining irrelevant distant poses.

3.11. Limitations of the Study. While our study provides a comprehensive evaluation of VS methodologies for urease inhibition, several limitations must be acknowledged. First, although we applied metal coordination restraints (via distance-based bonds) during the molecular dynamics simulations of PDB structures containing coresolved ligands (e.g., DJM, HAE, BME), such constraints were intentionally not applied during the docking of coumarin derivatives. This decision was guided by the absence of crystallographic or cryo-EM data defining a conserved binding mode for coumarins in urease and was made to preserve the blind and exploratory nature of our VS approach. Nevertheless, postdocking interaction analyses (Figures S5−S8) and distance-based evaluations (Figure S9) revealed that most top-scoring ligands do interact with key residues, including His322, His274, and Asp362, or approach the Ni^{2+} ions at coordination-compatible distances (<5 Å), even without explicit restraints. This suggests that the docking algorithm captures relevant binding features without artificially biasing pose selection.

Second, the OPLS3e force field used in all simulations, while broadly applicable, has documented limitations in modeling transition metal coordination. This may affect the energetic ranking of some ligands, particularly those interacting directly with the binuclear Ni^{2+} center. Our distance-energy plots (Figure S9) revealed that although many energetically favorable ligands are close to the Ni^{2+} ions, the scoring function does not always reflect a monotonic relationship between proximity and binding energy, hinting at limitations in the force field's parametrization.

Third, we did not filter docking poses based on their ability to coordinate the metal ions. While such filtering could prioritize plausible metallobinding interactions, it would introduce a mechanistic bias that might exclude ligands acting through alternative mechanisms, such as flap-loop stabilization or peripheral allosteric binding. Our analyses indicate that several coumarin derivatives interact with the flap-loop region, a flexible segment distal to the metal center, supporting the possibility of nonclassical inhibitory pathways.

Finally, our work adopted a "blind" VS philosophy, reflecting real-world applications in which binding modes of novel scaffolds are unknown. While this promotes unbiased chemical space exploration, it may overlook inhibitory mechanisms that rely on precise metal coordination. Future studies incorporating metallo-specific docking protocols, quantum mechanics/molecular mechanics (QM/MM) approaches, or experimental validations will be valuable to further refine the predictive accuracy of such pipelines.

4. CONCLUSIONS

This study systematically evaluated the predictive performance of various SBVS methods, protocol variants, statistical metrics, and data fusion approaches to improve the ligand ranking accuracy. By integrating multiple SBVS techniques and analyzing their predictive capabilities, we identified key trends that refine computational workflows for drug discovery, particularly in urease inhibition, but also extensible to other targets. Notably, while the 1E9Z system yielded high correlation metrics, we caution that its enclosed active site renders it an outlier. Thus, our conclusions are not solely based on this system but reflect consistent findings across all of the studied proteins.

One of the main findings was that increasing the number of docking poses from 10 to 100 generally reduced the predictive performance across most data fusion methods. Both Pearson and Spearman correlations declined, while error metrics such as MAE and RMSE increased, suggesting that excessive pose aggregation introduces noise rather than enhancing accuracy. The notable exception was the Minimum fusion method, which demonstrated improvements across all metrics.

When comparing SBVS methodologies, MM-GBSA and ED emerged as the most effective in ranking compounds, consistently achieving higher Spearman and Pearson correlations. However, MM-GBSA, while excelling in ranking performance, exhibited higher absolute errors (MAE and RMSE), suggesting that although it effectively orders compounds by binding affinity, its direct estimation of free energy values requires careful interpretation. IFD showed competitive results, particularly in 1E9Z, but lacked a broad applicability across all protein targets. In contrast, QPLD and standard docking methods were outperformed by these more advanced approaches.

Another key aspect of the analysis was the comparison between IC_{50} and pIC_{50} as experimental reference values. The results showed that pIC_{50} consistently achieved better Pearson correlations, reinforcing that log transformation enhances the linearity between the predicted and experimental values. However, Spearman correlations remained largely unaffected by this transformation, suggesting that for ranking purposes, IC_{50} and pIC_{50} yield comparable results. These findings

https://doi.org/10.1021/acsomega.5c04457
ACS Omega 2025, 10, 49641−49658

indicate that when absolute accuracy is a priority, pIC_{50} should be favored, but for rank-based predictions, either metric is equally suitable.

Furthermore, the study identified high-performing parameter combinations for which Spearman correlations exceeded 0.75. Among these, MM-GBSA-based scoring functions, particularly MMGBSA_dGBind_Coulomb, were the most frequently observed. Additionally, data fusion strategies such as Arithmetic and Euclidean means were consistently linked to strong predictive performance, emphasizing their effectiveness in molecular ranking.

The results of this study have direct implications for refining SBVS workflows by optimizing pose aggregation strategies, improving data fusion methods, and selecting the most reliable energy terms for molecule ranking. The findings suggest that hybrid approaches, combining the ranking strength of MM-GBSA with the lower error rates observed in other methods, such as IFD, could further enhance predictive accuracy. Moreover, integrating machine-learning techniques to dynamically select the best parameters for data fusion and pose selection could improve the prediction reliability. While this study focused on urease inhibitors, the methodologies presented are applicable to other protein targets in drug discovery, particularly for enzymes with flexible active sites.

Despite these advances, several limitations should be considered. Structural constraints in certain urease variants, such as 1E9Z, may have influenced correlation values, limiting the generalizability of findings to more flexible proteins. Additionally, the computational cost of advanced methodologies such as IFD and MM-GBSA remains a challenge for large-scale VS campaigns.

Overall, this study advances SBVS strategies by providing a systematic evaluation of VS methodologies, data fusion techniques, and ranking metrics. By refining computational strategies and identifying key parameters that enhance predictive accuracy, these findings pave the way for more effective workflows. Future research should focus on expanding these methodologies to other targets and incorporating adaptive scoring frameworks that further enhance the reliability and efficiency in CADD.

■ ASSOCIATED CONTENT

Data Availability Statement

Data supporting the findings of this study are available within the electronic Supporting Information. A R-language pipeline was deposited in GitHub link: https://github.com/DanielBustosG/Comparative-Data-Analysis-of-Virtual-Screening-Methodologies-for-Predicting-Urease-Inhibitory-Activi.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.5c04457.

Additional figures (Figures S1−S9) and tables (Tables S1−S3) that provide detailed results and complementary analyses; these include: comparative cavity volume estimations across MD-derived frames (Figure S1); performance evaluation of virtual screening protocols across different statistical metrics (Figures S2−S4); frequency heatmaps of protein−ligand and metal−ligand interactions across ensemble docking runs for all studied urease structures (Figures S5−S8); distance−energy plots evaluating pose proximity to catalytic Ni^{2+} ions (Figure S9); descriptions and values of all energy terms used in docking and MM-GBSA analyses (Table S1); statistical comparisons between IC_{50} and pIC_{50} using multiple correlation metrics (Table S2); and frequency analysis of high-performance energy term combinations across fusion methods (Table S3); additionally, the coumarin-derived molecules used in the study, including their 2D structures, standardized titles, and reported experimental pIC_{50} values from the literature (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Daniel Bustos − *Laboratory of Bioinformatics and Computational Chemistry, Department of Translational Medicine, Faculty of Medicine, Universidad Católica del Maule, Talca 3480094, Chile;* orcid.org/0000-0002-2136-2305; Email: dbustos@ucm.cl

Authors

Elizabeth Valdés-Muñoz − *Doctorate in Translational Biotechnology, Center for Biotechnology of Natural Resources, Universidad Catolica del Maule, Talca 3480094, Chile; Laboratory of Bioinformatics and Computational Chemistry, Department of Translational Medicine, Faculty of Medicine, Universidad Católica del Maule, Talca 3480094, Chile;* orcid.org/0009-0007-3111-7729

Gabriel J. Olguín-Orellana − *Departamento de Farmacología, Facultad de Ciencias Biológicas, Universidad de Concepción, Concepción 4030000, Chile;* orcid.org/0000-0003-1239-5962

Sofía E. Ríos-Rozas − *Laboratory of Bioinformatics and Computational Chemistry, Department of Translational Medicine, Faculty of Medicine, Universidad Católica del Maule, Talca 3480094, Chile;* orcid.org/0009-0003-4907-3964

Melissa Alegría-Arcos − *Data Science Research Center, Faculty of Engineering and Business, Universidad de las Américas, Santiago 7500000, Chile;* orcid.org/0000-0002-9372-9153

Natalia Morales − *Laboratory of Bioinformatics and Computational Chemistry, Department of Translational Medicine, Faculty of Medicine and Doctorate in Engineering, Faculty of Engineering, Universidad Católica del Maule, Talca 3480094, Chile*

Vicente Rojas-Santander − *Laboratory of Bioinformatics and Computational Chemistry, Department of Translational Medicine, Faculty of Medicine, Universidad Católica del Maule, Talca 3480094, Chile;* orcid.org/0009-0005-8791-9515

Javier Farías-Abarca − *Laboratory of Bioinformatics and Computational Chemistry, Department of Translational Medicine, Faculty of Medicine, Universidad Católica del Maule, Talca 3480094, Chile;* orcid.org/0009-0006-1828-4656

Jonathan M. Palma − *Faculty of Engineering, Universidad de Talca, Curicó 3460000, Chile;* orcid.org/0000-0002-3924-1907

Erix W. Hernández-Rodríguez − *Laboratory of Bioinformatics and Computational Chemistry, Department of Translational Medicine, Faculty of Medicine, Universidad Católica del Maule, Talca 3480094, Chile;* orcid.org/0000-0002-9231-7552

https://doi.org/10.1021/acsomega.5c04457
ACS Omega 2025, 10, 49641−49658

Reynier Suardíaz − *Department of Physical Chemistry, Faculty of Chemical Sciences, Complutense University of Madrid, 28040 Madrid, Spain;* orcid.org/0000-0002-1035-9020

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.5c04457

Author Contributions

[O]E.V.-M. and G.J.O.-O. contributed equally to this work. E.V.-M., G.J.O.-O., and S.E.R.-R. have computed the calculations. M.A. and N.M.R. have helped with coding the analysis. V.R.-S. and J.F.-A. have supported data organization, curation, and visualization. J.P.-O., E.H.-R., and R.S. have contributed to the computer equipment, review, and discussion of the results. D.B. has conceptualized the idea, overseen all tasks, and coded the notebooks. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding

D.B. thanks ANID FONDECYT de Iniciación 11220444.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

J.P. thanks ANID FONDECYT Regular 1241305. E.W.H.-R. thanks ANID FONDECYT de Iniciación en Investigación #11230033. R.S. would like to thank the Royal Society of Chemistry for research funding (Grants R19-3409, R20-6912, and R21-6448709305) and MCIN/AEI/10.13039/501100011033 (Grants PID2020-113147GA-I00 and PID2021-122839NB-I00).

■ REFERENCES

(1) Niazi, S. K.; Mariam, Z. Computer-Aided Drug Design and Drug Discovery: A Prospective Analysis. *Pharmaceuticals* **2024**, *17*, 22.
(2) Rasul, H. O.; Ghafour, D. D.; Aziz, B. K.; Hassan, B. A.; Rashid, T. A.; Kivrak, A. Decoding Drug Discovery: Exploring A-to-Z In Silico Methods for Beginners. *Appl. Biochem. Biotechnol.* **2025**, *197*, 1453.
(3) Vemula, D.; Jayasurya, P.; Sushmitha, V.; Kumar, Y. N.; Bhandari, V. CADD, AI and ML in drug discovery: A comprehensive review. *Eur. J. Pharm. Sci.* **2023**, *181*, 106324.
(4) Oliveira, T.; Silva, M.; Maia, E.; Silva, A.; Taranto, A. Virtual Screening Algorithms in Drug Discovery: A Review Focused on Machine and Deep Learning Methods. *Drugs Drug Candidates* **2023**, *2* (2), 311−334. May
(5) Cruz-Monteagudo, M.; Medina-Franco, J. L.; Pérez-Castillo, Y.; Nicolotti, O.; Cordeiro, M. N.; Borges, F. Activity cliffs in drug discovery: Dr Jekyll or Mr Hyde? *Drug Discovery Today* **2014**, *19*, 1069−1080, DOI: 10.1016/j.drudis.2014.02.003.
(6) Huang, S. Y.; Zou, X. Advances and challenges in Protein-ligand docking. *Int. J. Mol. Sci.* **2010**, *11*, 3016.
(7) Amaro, R. E.; Baudry, J.; Chodera, J.; et al. Ensemble Docking in Drug Discovery. *Biophys. J.* **2018**, *114* (10), 2271−2278.
(8) Nabuurs, S. B.; Wagener, M.; De Vlieg, J. A flexible approach to induced fit docking. *J. Med. Chem.* **2007**, *50* (26), 6507−6518.
(9) Sherman, W.; Beard, H. S.; Farid, R. Use of an induced fit receptor structure in virtual screening. *Chem. Biol. Drug Des* **2006**, *67* (1), 83−84.
(10) Kalyaanamoorthy, S.; Chen, Y. P. P. Quantum polarized ligand docking investigation to understand the significance of protonation states in histone deacetylase inhibitors. *J. Mol. Graphics Modell.* **2013**, *44*, 44−53.
(11) Tuccinardi, T. What is the current value of MM/PBSA and MM/GBSA methods in drug discovery? *Expert Opin. Drug Discovery* **2021**, *16*, 1233−1237, DOI: 10.1080/17460441.2021.1942836.
(12) Zhu, H.; Zhang, Y.; Li, W.; Huang, N. A Comprehensive Survey of Prospective Structure-Based Virtual Screening for Early Drug Discovery in the Past Fifteen Years. *Int. J. Mol. Sci.* **2022**, *23* (24), 15961. Dec.
(13) Michino, M.; Abola, E.; Brooks, C. L.; et al. Community-wide assessment of GPCR structure modelling and ligand docking: GPCR Dock 2008. *Nat. Rev. Drug Discovery* **2009**, *8* (6), 455−463 *8:6*. May
(14) Boyles, F.; Deane, C. M.; Morris, G. M. Learning from the ligand: using ligand-based features to improve binding affinity prediction. *Bioinformatics* **2020**, *36* (3), 758−764. Feb.
(15) Brown, B. P.; Mendenhall, J.; Geanes, A. R.; Meiler, J. "General Purpose Structure-Based Drug Discovery Neural Network Score Functions with Human-Interpretable Pharmacophore Maps. *J. Chem. Inf. Model.* **2021**, *61* (2), 603−620. Feb.
(16) Pernet, C. R.; Wilcox, R.; Rousselet, G. A. "Robust correlation analyses: False positive and power validation using a new open source matlab toolbox. *Front. Psychol.* **2013**, *3* (JAN), 39803. Jan.
(17) Schober, P.; Schwarte, L. A. Correlation Coefficients: Appropriate Use and Interpretation. *Anesth. Analg.* **2018**, *126* (5), 1763−1768. May
(18) Chai, T.; Draxler, R. R. "Root mean square error (RMSE) or mean absolute error (MAE)? *Geosci. Model Dev. Discuss.* **2014**, *7*, 1525−1534.
(19) Luo, Q.; Zhao, L.; Hu, J.; Jin, H.; Liu, Z.; Zhang, L. The scoring bias in reverse docking and the score normalization strategy to improve success rate of target fishing. *PLoS One* **2017**, *12* (2), No. e0171433. Feb.
(20) Wang, K.; Lyu, N.; Diao, H.; et al. GM-DockZn: a geometry matching-based docking algorithm for zinc proteins. *Bioinformatics* **2020**, *36* (13), 4004−4011. Jul.
(21) Cao, Y.; Shen, Y. Bayesian Active Learning for Optimization and Uncertainty Quantification in Protein Docking. *J. Chem. Theory Comput.* **2020**, *16* (8), 5334−5347. Aug.
(22) Ashtawy, H. M.; Mahapatra, N. R. "Task-Specific Scoring Functions for Predicting Ligand Binding Poses and Affinity and for Screening Enrichment. *J. Chem. Inf. Model.* **2018**, *58* (1), 119−133. Jan.
(23) Svane, S.; Sigurdarson, J. J.; Finkenwirth, F.; Eitinger, T.; Karring, H. Inhibition of urease activity by different compounds provides insight into the modulation and association of bacterial nickel import and ureolysis. *Sci. Rep.* **2020**, *10* (1), 8503 DOI: 10.1038/s41598-020-65107-9.
(24) Krajewska, B. Functional, catalytic and kinetic properties: A review. *J. Mol. Catal. B:Enzym.* **2009**, *59* (1−3), 9−21.
(25) Musicki, B.; Periers, A. M.; Laurin, P.; et al. Improved antibacterial activities of coumarin antibiotics bearing 5′,5′-dialkylnoviose: biological activity of RU79115. *Bioorg. Med. Chem. Lett.* **2000**, *10* (15), 1695−1699.
(26) Al-Haiza, M. A.; Mostafa, M. S.; El-Kady, M. Y. Synthesis and Biological Evaluation of Some New Coumarin Derivatives. *Molecules* **2003**, *8* (2), 275−286. *Molecules*, Feb.
(27) Vianna, D. R.; Hamerski, L.; Figueiró, F.; et al. Selective cytotoxicity and apoptosis induction in glioma cell lines by 5-oxygenated-6,7-methylenedioxycoumarins from Pterocaulon species. *Eur. J. Med. Chem.* **2012**, *57*, 268−274. Nov.
(28) Küpeli, E.; Tosun, A.; Yesilada, E. Anti-inflammatory and antinociceptive activities of Seseli L. species (Apiaceae) growing in Turkey. *J. Ethnopharmacol.* **2006**, *104* (3), 310−314. Apr.
(29) Cunha, E. S.; Chen, X.; Sanz-Gaitero, M.; Mills, D. J.; Luecke, H. Cryo-EM structure of Helicobacter pylori urease with an inhibitor in the active site at 2.0 Å resolution. *Nat. Commun.* **2021**, *12* (1), 230.
(30) Ha, N. C.; Oh, S. T.; Sung, J. Y.; Cha, K. A.; Lee, M. H.; Oh, B. H. Supramolecular assembly and acid resistance of Helicobacter pylori urease. *Nat. Struct. Biol.* **2001**, *8* (6), 505−509.
(31) Berman, H. M.; et al. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235−242. Jan.
(32) Rashid, U.; Rahim, F.; Taha, M.; et al. Synthesis of 2-acylated and sulfonated 4-hydroxycoumarins: In vitro urease inhibition and molecular docking studies. *Bioorg. Chem.* **2016**, *66*, 111−116.

(33) Salar, U.; Nizamani, A.; Arshad, F.; et al. Bis-coumarins; non-cytotoxic selective urease inhibitors and antiglycation agents. *Bioorg. Chem.* **2019**, *91*, No. 103170.
(34) Alomari, M.; Taha, M.; Imran, S.; et al. Design, synthesis, in vitro evaluation, molecular docking and ADME properties studies of hybrid bis-coumarin with thiadiazole as a new inhibitor of Urease. *Bioorg. Chem.* **2019**, *92*, No. 103235. August
(35) Naz, F.; Latif, M.; Salar, U.; et al. 4-Oxycoumarinyl linked acetohydrazide Schiff bases as potent urease inhibitors. *Bioorg. Chem.* **2020**, *105*, No. 104365.
(36) Khan, I.; Khan, A.; Ahsan Halim, S.; et al. Exploring biological efficacy of coumarin clubbed thiazolo[3,2−b][1,2,4]triazoles as efficient inhibitors of urease: A biochemical and in silico approach. *Int. J. Biol. Macromol.* **2020**, *142*, 345−354.
(37) Khan, K. M.; Iqbal, S.; Lodhi, M. A.; et al. Biscoumarin: New class of urease inhibitors; Economical synthesis and activity. *Bioorg. Med. Chem.* **2004**, *12* (8), 1963−1968.
(38) Maestro, S. *Schrödinger Release 2021−1*; LLC: New York, NY, 2020.
(39) Roos, K.; Wu, C.; Damm, W.; et al. OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. *J. Chem. Theory Comput.* **2019**, *15* (3), 1863−1874.
(40) Olsson, M. H. M.; SØndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent treatment of internal and surface residues in empirical pKa predictions. *J. Chem. Theory Comput* **2011**, *7* (2), 525−537.
(41) Tian, W.; Chen, C.; Lei, X.; Zhao, J.; Liang, J. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res.* **2018**, *46* (W1), W363−W367. Jul.
(42) Grant, B. J.; Skjærven, L.; Yao, X. Q. The Bio3D packages for structural bioinformatics. *Protein Sci.* **2021**, *30* (1), 20−30. Jan.
(43) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47* (7), 1750−1759.
(44) Valenzuela-Hormazabal, P.; Sepúlveda, R. V.; Alegría-Arcos, M.; et al. Unveiling Novel Urease Inhibitors for Helicobacter pylori: A Multi-Methodological Approach from Virtual Screening and ADME to Molecular Dynamics Simulations. *Int. J. Mol. Sci.* **2024**, *25* (4), 1968.
(45) Bochevarov, A. D.; Harder, E.; Hughes, T. F.; et al. Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences. *Int. J. Quantum Chem.* **2013**, *113* (18), 2110−2142.

https://doi.org/10.1021/acsomega.5c04457
ACS Omega 2025, 10, 49641−49658