ACS PHYSICAL CHEMISTRY Au
AN OPEN ACCESS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/physchemau
Article

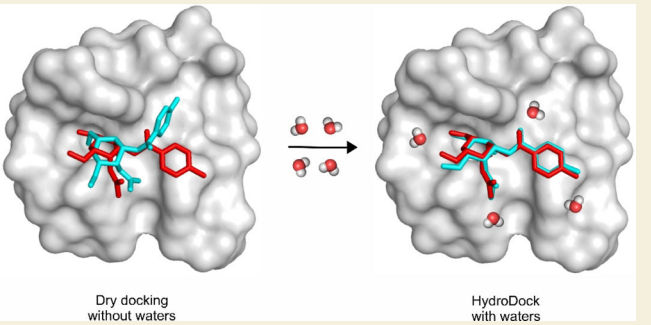
Robust Ligand Docking into Challenging Hydrated Binding Pockets

Bayartsetseg Bayarsaikhan, Csaba Hetényi, and Balázs Zoltán Zsidó*

Cite This: https://doi.org/10.1021/acsphyschemau.6c00072
Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Correctly positioning water molecules at the drug-target interface is a major challenge in structure-based drug design. Predicting how water structure reorganizes during drug docking into the target pocket is particularly difficult. Previously, a molecular dynamics-based protocol, HydroDock, was introduced and tested for docking small ligands into hydrated ion channels. Here, we extend the applicability of HydroDock to challenging targets with open binding pockets, populated by highly mobile water molecules that are the most troublesome for drug design. HydroDock improved docking accuracy, with increases of 28% (10%) in median structural (and ranking) performance. The comparison with other methods showed that HydroDock is a robust solution, even for problematic systems in which docking without explicit water molecules failed to find the correct drug-binding mode.



KEYWORDS: *docking, water molecule, refinement, molecular dynamics, small molecule*

1. INTRODUCTION

There have been various docking methods introduced for the accurate prediction of drug-target interactions.[1] However, fast docking methods often fail to capture the dynamic nature of binding pockets,[1] particularly when interfacial water molecules are involved.[2] These waters are not passive spectators—they mediate hydrogen bonding, stabilize conformations, and influence both ligand-binding affinity and specificity.[2−4] Yet their small size, high mobility, and dynamic hydrogen-bonding networks[5] make them difficult to model. Certain binding pockets, like the medically important narrow channels (,[6] Figure 1A) provide topological constraints that reduce the movement of water molecules. Therefore, structurally narrow channels are less challenging targets compared to open binding pockets. Open pockets (deep and shallow in Figure 1A) pose decreased topological constraint[4] to keep water molecules fixed in space. In wide-deep binding pockets, water molecules near the bulk water can freely leave the pocket. Only water molecules at the bottom of the deep binding pocket are kept by the surrounding topology. In wide-shallow pockets, the movement of water molecules is not limited by topology. The decreased topological constraint might lead to missing crucial hydrogen-bonding networks, which result in inaccurate energy calculations and hinder reliable ligand ranking.[2,7,8] Therefore, open binding pockets are challenging targets, especially when water molecules are involved.

Recent efforts to incorporate water molecules into docking calculations, whether as displaceable waters[11−18] or fully flexible waters,[19−21] have yielded mixed results, as summarized in Table S1. A recent review notes that deep learning-based tools cannot yet outperform physicochemical docking methods because they lack the physics relevant to docking,[22] especially for water molecules.[23]

To address the challenge of incorporating water molecules into docking for open binding pockets, we use HydroDock.[6] HydroDock requires only the dry target and ligand structures to generate hydrated complexes (Figure 1). HydroDock addresses the chicken-and-egg problem of hydrated docking by performing simultaneous dry docking and water prediction steps, followed by a short molecular dynamics (MD) refinement. Previously, the protocol achieved[6] excellent agreement between experimental structures and HydroDock-determined structures for narrow channels (Figure 1A). In this study, we aim to extend the applicability of HydroDock from docking in narrow channels to docking in proteins with challenging open binding pockets. As test systems (Table 1), we used medically important and structurally challenging anticancer drug targets, including Heat shock protein 90-alpha (HSP90),[24] Bruton's tyrosine kinase (BTK),[25] and cyclin-dependent kinase 2 (CDK2).[26] Antiinfective drug targets, such as the Type 1 fimbrin D-mannose-specific (FimH) adhesin[27] and the sialic acid-binding periplasmic protein (SiaP).[28] Together, a diverse set of 21 protein−small molecule ligand

Received: April 29, 2026
Revised: August 4, 2026
Accepted: August 4, 2026

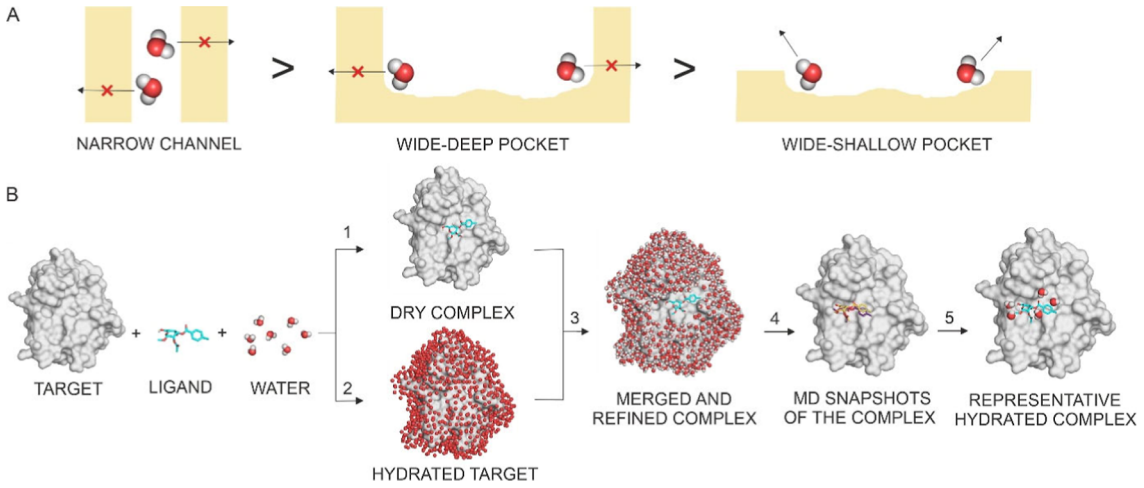


Figure 1. A) Comparison of a narrow channel, a wide-deep, and a wide-shallow binding pocket. Less-than signs indicate the decreasing order of topological constraint for the water molecule to leave into the bulk solvent. In the shallow pocket, water molecules can freely exchange with bulk waters in the absence of topological constraints. In the deep binding pocket and narrow channel, water molecules can not escape as quickly from the deeper pockets within the pocket. (**B**) Workflow of HydroDock protocol shown for System 3t1l. The target protein and ligand conformations are shown as surface and stick representations, respectively. The numbering of HydroDock steps follows the explanation in the main text. The surface water molecules produced by MobyWat[9,10] in the second step, are shown as red spheres; otherwise, stick representations are used for water molecules.

Table 1. Properties of the Protein–Ligand Systems Investigated in this Study

holo (PDB ID)	res[a] (Å)	apo (PDB ID)	res[a] (Å)	target name	IWC[b] (3.5)	ligand MW[c]	$pIC_{50}/pK_i/pK_D = -\log_{10}(M)$[d]	PS[e]	#rotatable bonds
1ikg [29]	1.9	3pte [30]	1.6	D-alanyl-D-alanine carboxypeptidase	5	374.390	-	72.18	11
1tou [31]	2	3q6l	1.4	fatty acid-binding protein, adipocyte	5	321.319	*6*[31]	65.97	3
2byi [24]	1.6	1yes [32]	2.2	heat shock protein 90-alpha	7	422.843	*6.34*[24]	69.12	5
2qcm [33]	1.67	2jgy	1.95	uridine 5′-monophosphate synthase	4	354.207	*3.35*[33]	45.33	6
2sim [34]	1.6	2sil [34]	1.6	sialidase	3	291.255	*4.42*[34]	70.36	5
2v4c [28]	1.7	2cey [35]	1.7	sialic acid-binding periplasmic protein	8	268.218	-	73.89	4
2xjx [36]	1.66	1yes [32]	2.2	heat shock protein 90-alpha	4	409.521	*9.15*	53.19	4
3t1l [37]	1.6	3zsl [38]	1.08	galectin-3	4	354.352	*4.26*	24.66	7
4j58 [39]	1.28	2bit [40]	1.71	peptidyl-prolyl cis–trans isomerase f, mitochondrial	5	276.334	*<3.30*[39]	46.27	4
4j59 [39]	1.92	2bit [40]	1.71	peptidyl-prolyl cis–trans isomerase f, mitochondrial	3	402.489	*5.85*[39]	62.4	5
4j5d [39]	1.32	2bit [40]	1.71	peptidyl-prolyl cis–trans isomerase f, mitochondrial	5	431.326	*5.96*[39]	66.14	5
4j5e [39]	0.99	2bit [40]	1.71	peptidyl-prolyl cis–trans isomerase f, mitochondrial	4	382.456	*4.94*[39]	69.23	6
4x50 [27]	2	5mca [41]	1.6	type 1 fimbrin D-mannose specific adhesin	3	332.348	*7.75*	62.82	4
5er1 [42]	2	4ape [43]	2.1	endothiapepsin	3	520.705	-	55.16	17
5fbe [44]	1.43	2xw9 [45]	1.2	kallikrein-7	5	465.422	*4.85*[44]	73.11	7
5fck [44]	1.86	2xw9 [45]	1.2	complement factor D	6	484.911	*7.52*[44]	41.94	6
5h9r [46]	1.58	3zsl [38]	1.08	galectin-3	3	503.499	-	38.31	13
5iyy [47]	0.9	1kex [48]	1.9	neuropilin-1	4	584.667	*4.77*	37.38	9
5nat [49]	1.17	2xw9 [45]	1.2	complement factor D	7	446.422	*4.70*[49]	73.45	4
5nkh [50]	1.29	5ek7	1.9	ephrin type-A receptor 2	4	561.098	-	66.32	5
6q4g [26]	0.98	1hcl [51]	1.8	cyclin-dependent kinase 2	4	269.259	-	77.39	3

[a]Crystallographic resolution. [b]Interfacial Water Count, the number of water molecules within 3.5 Å of both target and ligand. [c]Molecular weight of the ligand. [d]IC_{50} values are shown in italic, and K_d values in bold text. [e]Percentage of the surface area of the ligands buried upon binding.

complexes with a broad molecular weight (268–585 Da) and experimental binding affinity (nM-mM, Table 1) range was assembled. The set of complexes encompasses ligands of various sizes, both lightly and tightly bound, that bind to challenging open binding pockets, and therefore, is adequate to extend and test HydroDock.[6]

2. RESULTS AND DISCUSSION

2.1. Structural Performance of Protein–Ligand Binding

The structural performance of the HydroDock protocol was evaluated using standard procedures, calculating the all-atom root-mean-square deviation (RMSD, Methods) of the ligand relative to its experimental binding mode. The initial dry-

https://doi.org/10.1021/acsphyschemau.6c00072
ACS Phys. Chem Au XXXX, XXX, XXX–XXX

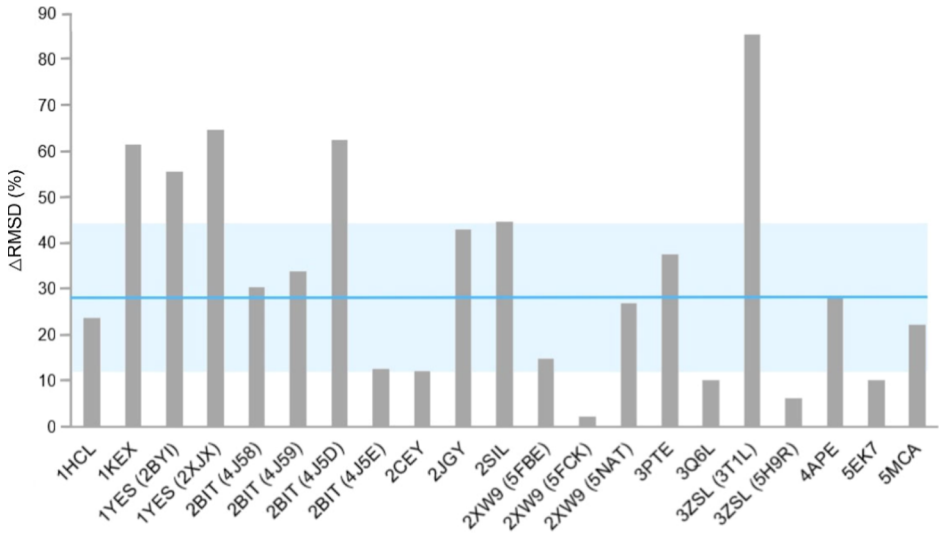


Figure 2. Summary of HydroDock performance across all test systems listed in Table 1 using apo target structures. The corresponding holo system is stated in brackets for clarity. Structural refinement results as indicated by ΔRMSD (%) values calculated between $RMSD_{start}$ and $RMSD_{best}$ values obtained by the HydroDock protocol (see also Table S2). The median of ΔRMSD is indicated by a horizontal blue line, and the corresponding MAD by a transparent blue background.

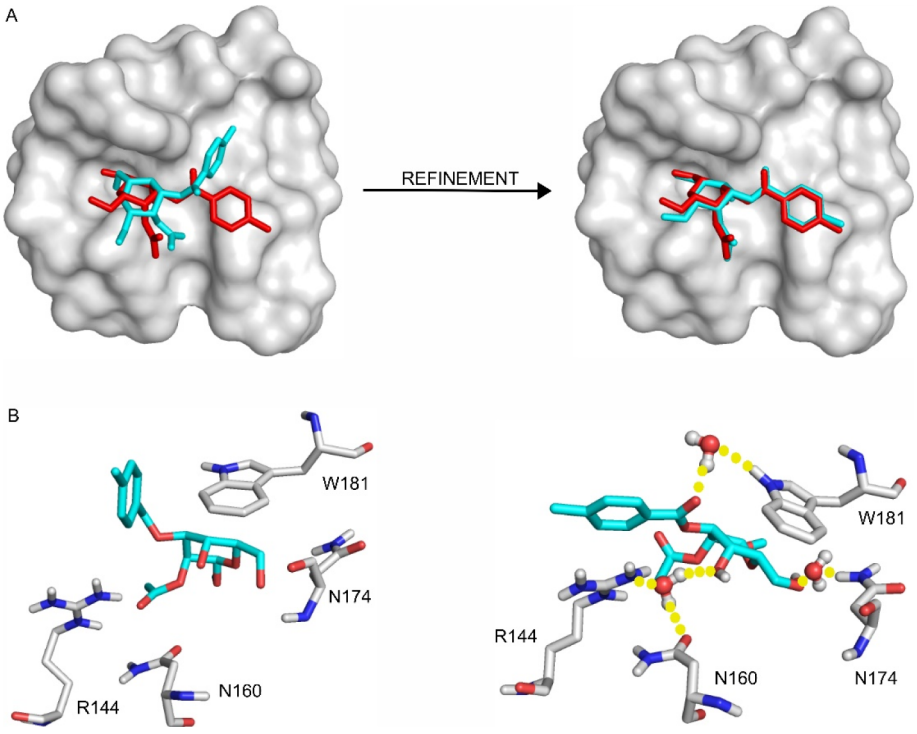


Figure 3. Comparison of the initial dry-docked (left) and the HydroDock-refined best binding mode (right) produced for human Galectin-3 in complex with methyl 2-*O*-acetyl-3-*O*-toluoyl-beta-D-talopyranoside (PDB ID: 3zsl) (holo 3t1l). The initial (left) and the best (right) binding modes are shown as cyan sticks. The crystallographic ligand-binding mode is shown as red sticks for reference. The target is shown as a gray surface (**A**) and as a gray all-atom stick representation (**B**) as a blow-up of the binding pocket, and interacting amino acids are labeled according to PDB: 3zsl (holo 3t1l). Water molecules are shown as red and white sticks. Important hydrogen bonds formed through water molecules are shown with yellow dotted lines. Water molecules are necessary for the experimentally observed target−ligand interactions, which are absent in the dry-docked binding mode.

docked binding mode ($RMSD_{start}$, Table S2) was compared with the best pose produced by HydroDock ($RMSD_{best}$, Methods). The improvement achieved by HydroDock is reported as a percentage relative to $RMSD_{start}$ (ΔRMSD%, Table S2, Figure 2, eqs 1 and 2).

$$\Delta RMSD = RMSD_{start} - RMSD_{best} \quad (1)$$

$$\Delta RMSD(\%) = \frac{\Delta RMSD \times 100}{RMSD_{best}} \quad (2)$$

Reporting ΔRMSD% may be more informative than ΔRMSD (also available in Table S2) for systems with different $RMSD_{start}$ values (as is the case for the systems in the diverse set of Table 1). In the main text, results for apo targets are

https://doi.org/10.1021/acsphyschemau.6c00072
ACS Phys. Chem Au XXXX, XXX, XXX−XXX

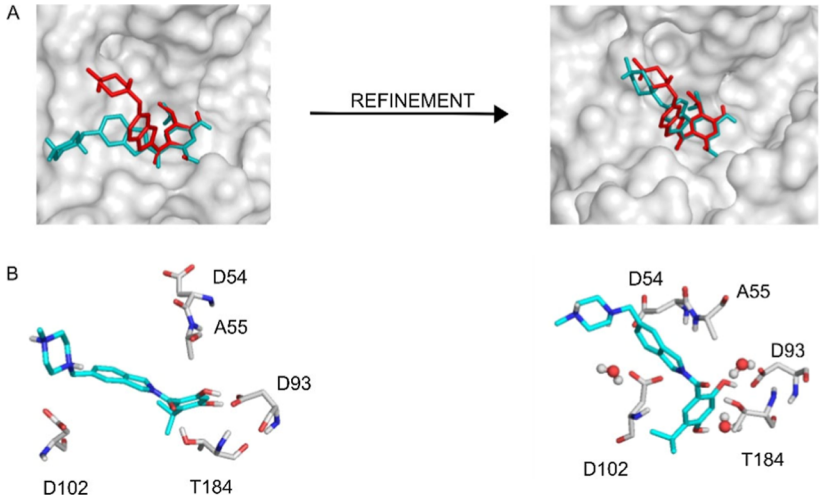


Figure 4. A) Comparison of the dry-docked (left) and the refined best binding mode produced by HydroDock (right) for HSP90 in complex with onalespib (PDB ID: 1yes) (holo 2xjx). The docked and refined binding modes are shown as teal all-atom representation sticks; the crystallographic ligand binding mode is shown in red sticks for reference. (**B**) A blow-up image of the binding pocket. Interacting amino acids are shown as gray all-atom representation sticks and labeled according to PDB: 1yes (holo 2xjx). In the dry-docked binding mode, the ligand interacts with D102, and in the refined binding mode, the interaction with D102 is replaced by a water molecule to push the ligand toward the experimental binding mode.

shown, and in the Supporting Information, values for holo systems are provided (Table S3).

Across the 21 systems in our test set, dry-docked $RMSD_{start}$ values ranged from 1.5 to 9.1 Å, indicating moderate to poor starting positions with a median of 4.1 Å (Table S2), hence an overall mis-docking situation. HydroDock improved the initial dry-docking results in all cases, as indicated by the ΔRMSD values in Table S2, achieving a median ΔRMSD improvement of 28%. Ten systems exceeded the median improvement, with 5 even exceeding the median absolute deviation (Figure 2). This demonstrates the protocol's robustness across a diverse set of protein−ligand complexes with challenging open binding pockets, which is attributable to the molecular dynamics simulations used. However, the extent of improvement varied by system (Figure 2), the largest gains were observed when the initial pose retained stabilizing interactions with the target, such as hydrogen bonds or π-stacking contacts, whereas moderate gains occurred when the pose was highly incorrect or constrained by deep, sterically restrictive binding pockets that limited ligand mobility. These trends indicate that HydroDock performs best when the initial docking pose captures key native interactions and the binding pocket geometry enables conformational accessibility.

Over 10% ΔRMSD improvement was observed for 19 systems, with the largest ΔRMSD% improvement observed for human Galectin-3 in complex with methyl 2-*O*-acetyl-3-*O*-toluoyl-beta-D-talopyranoside, namely the apo 3zsl (3t1l) system, achieving 85% improvement with the HydroDock protocol (Figure 3A). The initial dry-docked binding mode showed some experimental-like contacts with the target (Figure 3B). However, the toluoyl group of the ligand binds differently from the reference position, failing to form the cation−π stacking interaction with R144, which is experimentally identified as one of the key stabilizing contacts in the Galectin-3 binding pocket.[37] In the refined binding mode, the oxo group adjacent to the toluoyl group interacts with W181 through a water molecule, rotating the toluoyl group into the correct position. Furthermore, two hydroxyl groups of the ligand interact with hydrophilic amino acids, such as R144, N160, and N174, through water molecules, adjusting and stabilizing the ligand in a near-native structure ($RMSD_{best}$ = 0.45 Å) (Figure 3B). The MobyWat surface water success rate (SR, see Methods) for this system was 100% (Table S3) in the 3.5 Å binding pocket defined by the superimposed ligand from the holo counterpart 3t1l system, providing an appropriate hydration layer for refinement by the HydroDock protocol.

Another example of the role of the appropriately predicted hydration layer is the 1yes (2xjx) system. In the dry-docked binding mode, the isoindoline moiety of the ligand deviated markedly from the experimental reference ligand binding mode (Figure 4, $RMSD_{start}$ = 5 Å). Thus, the dry-docked binding mode failed to form contacts with D54 and A55, which are important amino acids in ligand binding to HSP90.[36] In the absence of void-filling water molecules, the isoindoline group of the dry-docked ligand interacts with D102 instead of D54 and A55, and only in the refined structure is this interaction replaced by a water molecule. Therefore, the isoindoline group is pushed toward D54 and A55 to form the experimentally observed binding interactions (Figure 4). As a result, the ligand reaches its correct binding conformation ($RMSD_{best}$ of 1.8 Å), achieving a ΔRMSD improvement of 64% during refinement. The MobyWat-predicted surface water molecules within the 3.5 Å binding pocket defined by the superimposed ligand from the holo counterpart 2xjx system (SR = 100%, Table S3) were necessary to accurately reproduce the experimental binding mode. Five other systems with significant improvements are also highlighted in Figures S1−S5.

The protocol was also applied to holo target structures (Table S3). The median ΔRMSD improvement was 17%. All systems improved (except one), and 14 systems showed at least a 10% improvement in ΔRMSD. The improvements for the apo targets are larger than those of the holo targets, explained by the higher starting RMSD values (4.09 Å for apo vs 2.92 Å for holo; median $RMSD_{start}$ in Tables S3 vs S4) due to the absence of the ligand-binding conformation in the apo target. Starting from a higher $RMSD_{start}$ (corresponding to worse ligand-binding modes on apo) is more challenging and requires the robust approach of HydroDock, which also handles ligand-induced target conformational changes during the final MD refinement step. Consequently, larger improve-

https://doi.org/10.1021/acsphyschemau.6c00072
ACS Phys. Chem Au XXXX, XXX, XXX−XXX

ments can be achieved for apo targets compared to holo targets.

2.2. Structural Performance of Predicting Water Molecules

The 21 holo experimental complex structures contained 97 interfacial water molecules, of which the HydroDock prediction step (using the molecular dynamics-based MobyWat program[9,10]) found 94% within 1.7 Å (Table S3), an excellent result according to the literature.[2,3] We predicted 389 water molecules, effectively filling the voids in the experimental structures and completing the hydration shell. This void-free hydration is essential for stabilizing docked complexes,[52,53] preventing artificial intramolecular contacts, and enabling realistic hydrogen-bond networks at the binding interface.[54] The strength of the prediction not only lies in finding experimental water molecules, because experiments do not provide a complete void-free hydration layer.[2] The combination of identifying the validated experimental water positions with high accuracy (Figure 5) and predicting additional waters to fill the complete binding pocket demonstrates the true strength of water molecule prediction.

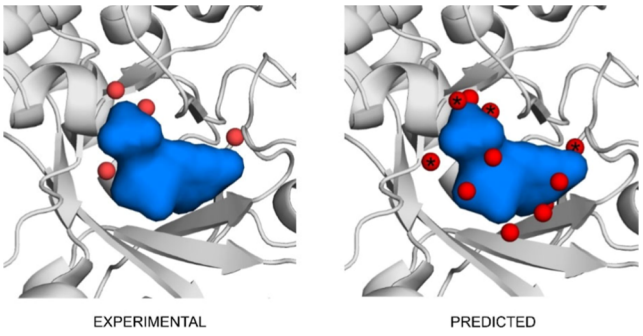


Figure 5. Experimental (left) and predicted (right) water molecules for system 2qcm (apo 2jgy). The target is shown as a gray cartoon, the ligand as a blue surface, and water molecules are red spheres. Asterisks indicate the predicted water molecules within 1.7 Å of the experimental water positions.

For example, in the case of the holo 2qcm system, there are 4 experimental water molecules in the binding pocket of the ligand (defined by a 3.5 Å radius around the experimental ligand, see Methods), MobyWat found all 4 of them, and predicted a total of 11 water molecules to fill the entire binding pocket (Figure 5). This example shows how water molecule prediction can find and complement the valid experimental water positions. The correct water positions are then refined by molecular dynamics to aid the ligand binding mode refinement with accurate hydrogen bond orientations. In the case of the holo 2qcm system, the 100% success rate of water positioning resulted in over 40% $\Delta RMSD_{best}$ improvement after refinement by HydroDock (Table S3).

As a basis for validation, crystallographic water molecules are used. However, it is worth noting that even the experimental determination of water molecules faces many limitations.[2] Especially, for larger systems, with many water molecules, the solvent is disordered, which increases noise, leading to difficult solvent assignation even in the first hydration shell,[55] where B-factor-based assignment of water molecules can be performed.[56] Furthermore, overfitting of electron density data improves resolution but can result in misleading identification of water sites.[57,58]

2.3. Ranking Performance

The initial dry-docked binding mode ($RMSD_{start}$, Table S2) was compared with the top-ranked pose produced by HydroDock ($RMSD_{top}$, Methods). The improvement achieved by HydroDock is reported as the percentage relative to $RMSD_{start}$ ($\Delta RMSD\%$, eqs 1 and 2, where $RMSD_{best}$ is substituted with $RMSD_{top}$).

Several ranking strategies were tested because the base HydroDock representative selection strategy did not yield sufficient improvement for open binding pockets.

The most commonly used strategy is to select the most energetically favorable binding mode. We evaluated several ranking methods based on target−ligand interaction energies (E_{inter}, see Methods), including their Lennard−Jones (LJ) and Coulomb components, calculated both with and without explicit water molecules. Several possibilities were tested, including treating water molecules as either part of the target or the ligand when calculating E_{inter} and its components. Among these, ranking by LJ energy with water molecules treated as part of the target (LJ tw-l) yielded the best performance, achieving a median $\Delta RMSD$ of 9.7% for the apo set (Table S5).

We then tested a hybrid strategy in which a representative binding mode (the closest structure to the average coordinates) was selected from the top 10% of LJ tw-l-ranked conformations. While this approach improved on the dry-docking results (median $\Delta RMSD$ of 6.6%, Table S6), its performance was notably lower than that of LJ tw-l ranking alone (9.7%).

To further refine the ranking, we applied frame elimination by discarding conformations with RMSD >5 Å or >10 Å relative to the first frame, thereby removing outliers such as dissociated poses or those drifting into nonrelevant subpockets. While this filtering alone had limited impact (Table S7), combining it with LJ tw-l energy ranking and representative selection yielded the greatest improvement (Figure 6). This resulted in a median 10% improvement in $\Delta RMSD$ for the apo set (Tables S4 and S7a). The median improvement was exceeded by 10 systems, including 4 that exceeded the median absolute deviation margin (Figure 6).

In addition to RMSD-based ranking, the frame with the most retained experimentally similar water-mediated contacts can be selected to demonstrate that the protocol not only improves geometric accuracy but also recovers chemically relevant interactions among water, the target, and the ligand. As an example, the MD trajectory of 3zsl(3t1l) is analyzed, and all 5 experimental water-mediated contacts are retained, and one additional is found after 14.4 ns of simulation (Figure S6).

2.4. Comparison with Other Methods

To answer the challenges described in the Introduction, relatively few methods have been developed to incorporate explicit water molecules into docking simulations (Table S1,[13,19−21,59]). The performance of three methods, AutoDock-GIST,[13] RosettaLigand,[21] and AutoDock Vina 1.2.0[60] (reference methods, Table 2) was compared to HydroDock.

The number of systems included in the reference methods ranged from 6 to 23; in the present study, 21 systems (Table 2) were used with HydroDock. The median $RMSD_{start}$ value of the HydroDock systems was at least two times higher than that of the reference studies (Figure 7A, Table S8), indicating that the diverse systems of the present study were overall more challenging hydration-wise than the reference ones.

https://doi.org/10.1021/acsphyschemau.6c00072
ACS Phys. Chem Au XXXX, XXX, XXX−XXX

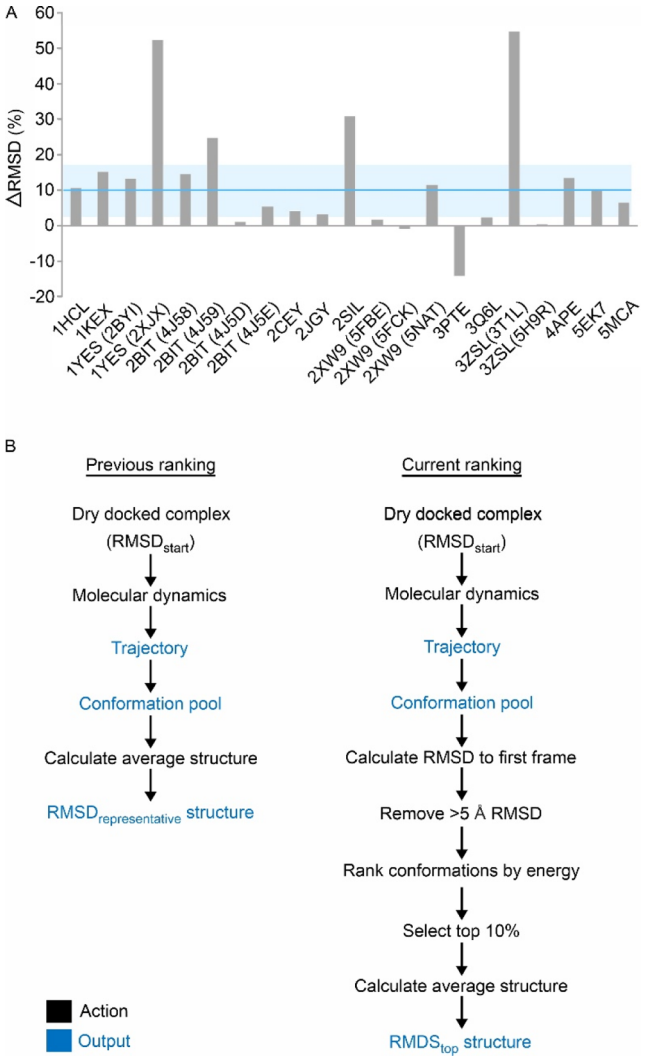


Figure 6. **A**) Ranking accuracy as indicated by ΔRMSD (%) calculated between RMSD$_{start}$ and RMSD$_{top}$ values selected by the top-performing ranking strategy in this study (see also Table S4). The median of ΔRMSD is indicated by a horizontal blue line, and the corresponding MAD by a transparent blue background. (**B**) Comparison of the previous HydroDock and current final ranking strategies. RMSD$_{start}$ is the starting structure of the refinement, and RMSD$_{top}$ is the final structure. The difference between the two is expressed as ΔRMSD.

From its definition, a high RMSD$_{start}$ indicates that the starting position of the ligand is far from the experimental binding mode. Therefore, improving the ligand poses of a high RMSD$_{start}$ shows the robustness of HydroDock.

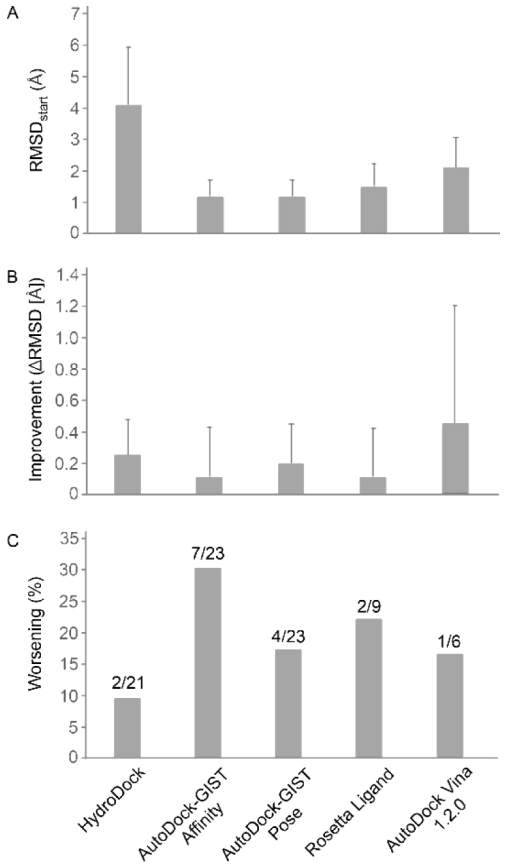


Figure 7. Comparison of the performance of HydroDock with other methods that use water molecules to improve dry docking accuracy. (**A**) The RMSD of the dry-docked starting structure for each method before the inclusion of water molecules. (**B**) The absolute improvement in RMSD after the inclusion of water molecules compared to the starting structure. (**C**) The ratio of worsened cases, where the inclusion of water molecules leads to an increased (worsened) RMSD$_{top}$ compared to the starting dry-docked (RMSD$_{start}$) position.

HydroDock showed the best absolute improvement in ranking (ΔRMSD, eq 1) after including water molecules in docking (Figure 7B), compared to the reference methods, except for AutoDock Vina. However, in the case of AutoDock Vina, the median absolute deviation of the reported ΔRMSD values for the 6 systems[60] is significantly larger than the median itself, and therefore, it cannot be used for a valid statistical comparison with the other methods.

Including water molecules not only improves docking accuracy but, in some cases, it can worsen the RMSD

Table 2. Docking Methods and Protocols Investigated

name	description	count of test systems
HydroDock[6] (with an updated ranking)	MD-based prediction of hydration structure, dry docking with AutoDock, and an additional MD refinement of the ligand pose in the hydrated binding pocket	21
AutoDock-GIST Affinity & AutoDock-GIST Pose[13]	MD-based prediction and GIST[a]-based selection of displaced and conserved water molecules, and docking with AutoDock to the hydrated binding pocket	23
RosettaLigand[21]	water molecules are handled as part of the target or the ligand during docking, and the scoring involves an implicit solvent model	9
AutoDock Vina 1.2.0[60]	water molecules are not involved in docking, but are used for the calculation of an entropic gain upon desolvation of the ligand	6

[a]Grid Inhomogeneous Solvation Theory.

https://doi.org/10.1021/acsphyschemau.6c00072
ACS Phys. Chem Au XXXX, XXX, XXX−XXX

(negative ΔRMSD). If the number of worsened cases is high relative to the total number of systems, the method is not really robust. This ratio (worsened/total cases) across all systems was the lowest for HydroDock (Figure 7C), also highlighting its robustness. The robustness is likely due to the final MD refinement step,[61] which is not present in the reference methods. Notably, HydroDock also applies MD to generate water positions (Section 2.2), achieving high success rates and contributing to its robustness. It is worth to mention that our preliminary results show the MD simulations can be simplified to reduce computational demand by using a single water droplet with a tiny 5- or 10 Å water layer around the protein instead of a simulation box filled with an order of magnitude larger number of waters (see Supplementary Method, Table S9). Further investigations on this "droplet simplification" will be presented in a forthcoming study.

2.5. Case Study

The case of Cyclophilin D inhibitor antivirals was selected to highlight how HydroDock captures chemically meaningful hydration effects. Cyclophilin D has a hydrophobic binding pocket with amino acids F60, M61, F113, and L122. This hydrophobic pocket (amino acids in sticks, Figure 8A) was targeted by a series of ligand modifications (4j58, 4j59, 4j5d, and 4j5e[39]) at the pyrrolidine terminus (asterisk in Figure 8A). The common part of these ligands (Figure 8A) aligns well with one another, and the differences in their antiviral activity arise from the modification of their terminal pyrrolidine ring located in the center of the hydrophobic pocket. In the HydroDock $RMSD_{best}$ structure for the base compound (4j58, Figure 8B), there are no water molecules in the hydrophobic pocket, and the urea oxygen of the molecule does not bind to R55, an important amino acid for antiviral activity (4j58, Figure 8B). Consequently, this compound has no antiviral activity (low pEC_{50}, Table 1). In the HydroDock $RMSD_{best}$ structure for 4j5e, a compound with improved activity (Table 1), there is still no connection with R55, but the water molecules around the methoxy-phenyl moiety arrange in a clathrate-like structure (Figure 8C). To optimize the water-mediated interactions, 11 water molecules of the clathrate contribute to the hydrophobic effect[62−64] sticking the substituted phenyl ring of the ligand and L122 together. 4j59 (Figure 8D) and 4j5d (Figure 8E) are equally strong inhibitors (Table 1). In both cases, R55 binds to the urea oxygen of the ligands, and the hydrophobic moieties of the ligands are also stabilized by smaller clusters of water in a clathrate-like assembly.

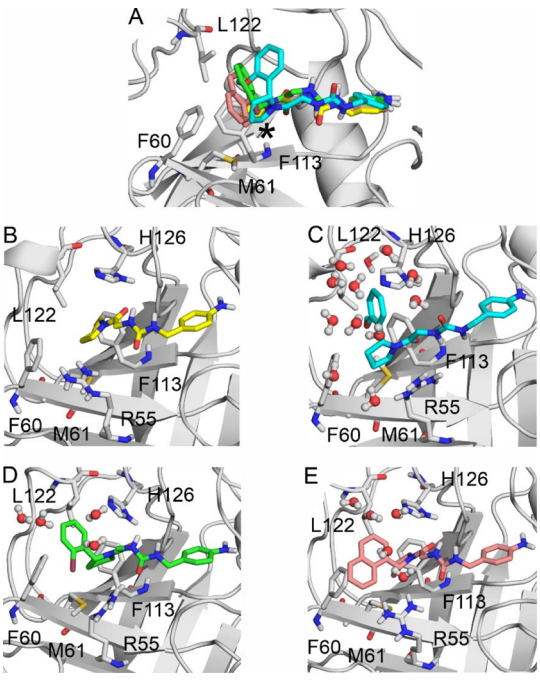


Figure 8. A case study of cyclophilin D inhibitors and the chemically meaningful hydration effects captured by HydroDock. (**A**) The binding modes of the substituted inhibitor molecules superimposed from their respective PDB structures (4j58 in yellow; 4j5e in cyan; 4j5d in green; 4j59 in pink). The amino acids of the hydrophobic binding pocket of cyclophilin D are shown in sticks, the terminal pyrrolidine ring of the inhibitors is marked with an asterisk. Systems 4j58 (**B**), 4j5e (**C**), 4j5d (**D**), and 4j59 (**E**) are shown with HydroDock-predicted water networks. The protein is shown as gray cartoon, the ligands and interacting amino acids as colored and gray all atom representation sticks, respectively. Water molecules are shown as red and white sticks and balls. Amino acids are labeled according to 4j58.[39]

These water molecules contributing to the hydrophobic effect do not mediate bridging interactions but instead "push" the two hydrophobic interaction partners toward each other by forming a common clathrate-like structure around them. Importantly, such a complex effect can be handled only with the explicit treatment of water molecules. Because hydrophobic binding pockets are present in many drug targets, MD-based methods with explicit water molecules (such as HydroDock) are key to accounting for complex water-mediated effects in drug design.

3. CONCLUSIONS

To extend and test the applicability of HydroDock to targets with challenging open binding pockets, a diverse test set of 21 pharmacologically important drug-target complexes of various hydration patterns was assembled. HydroDock resulted in a remarkable increase in structural (28%) and ranking (10%) performance compared to structures (mis)docked without water molecules. The structural improvement exceeded the median for half of the test set. HydroDock proved to be robust, as evidenced by the small number of cases that worsened. The robustness of the protocol can be attributed to the molecular dynamics step, which can improve even significantly deviated starting structures. It was also discussed how accurately predicted water molecules aided the improvement of ligand-binding poses via improved interactions with the target. A comparison with state-of-the-art methods showed that HydroDock achieves better results and greater robustness, even when starting from more challenging mis-docked structures.

4. METHODS

4.1. Selection of Test Systems

In the present study, structural files of protein-small molecule complexes were systematically mined from the Protein Data Bank (PDB,[65]). The filtering was based on the following criteria.

1. Structures determined by X-ray crystallography with a resolution of ≤2.5 Å Only one ligand is bound at the binding pocket
2. Unbound (apo) structures of the target protein are available

https://doi.org/10.1021/acsphyschemau.6c00072
ACS Phys. Chem Au XXXX, XXX, XXX−XXX

3. At least three interfacial (IF) water molecules within 3.5 Å
4. Noncovalently bound ligands
5. Drug-like and structurally diverse ligands
6. Ligands without peculiar ring systems that cannot be regenerated by redocking
7. No missing heavy atoms in ligands
8. No atoms with alternative locations in ligands

Further filtering was performed to remove complexes with sugar-containing ligands, lipid ligands, and targets containing cofactors or uncommon elements (e.g., Be, B, Si) near their binding pockets, or any nonstandard amino acid residues in direct contact with the ligands. Complexes whose apo target structures did not meet the general selection criteria were also excluded. Additionally, systems with missing residues or heavy atoms in the binding pocket region or in the ligands were excluded. The filtering resulted in a high quality diverse test set of 21 complexes (Table 1).

4.2. Ligand Preparation

Ligand structures were extracted directly from the structural files of their corresponding complexes in PyMol and subsequently processed for use in AutoDock 4.2.[66] Gasteiger−Marsili partial charges[67] were assigned to the ligand structures based on their protonation states at neutral pH 7.4 using Open Babel,[68] a software package, and were saved in MOL2 format. The resulting ligand structures were subjected to steepest descent optimization, followed by conjugate gradient minimization with a maximum of 100,000 steps and the MMFF94 force field.[69] In conjugate gradient minimization, the convergence threshold was set to 10^{-7} kcal mol^{-1} $Å^{-1}$. Furthermore, semiempirical quantum mechanical optimization was carried out with MOPAC[70] using PM7 parametrization,[71] with the upper limit for the gradient norm set to 0.001 kcal mol^{-1} $Å^{-1}$. The CHARMM-GUI[72] was used to assign bond parameters and atom types for the ligand topology using the general Amber force field,[73] which is designed to be compatible with the existing Amber force fields, including AMBER99SB-ILDN,[74] which we used to prepare the target topology.

4.3. Target Preparation

Both apo and holo forms of the proteins were prepared for docking. The atomic coordinates of the holo and apo target proteins (Table 1) were downloaded from the PDB. A clean target structure was generated by removing all nonprotein components, including ligands and all solvent molecules, except for important metal ions in the binding pocket. Swiss-Model[75] was used to detect and reconstruct missing residues and/or atoms in the target structure. The error-free, full-length protein sequence was obtained from UniProt.[76] PyMol[77] was used to add capping groups and to align the apo targets with their holo counterparts.

To prepare a molecular dynamics simulation in GROMACS,[78] a minimization was performed. The structure was placed in a dodecahedral box with a 1 nm cutoff between the solute and the box. The box was filled with explicit TIP3P water molecules,[79] and counterions (sodium or chloride ions) were added to neutralize the system using the gmx solvate program in GROMACS.[78] The simulation box was subjected to steepest descent (sd) optimization, with convergence thresholds set to 1000 kJ mol^{-1} nm^{-1}. Next, conjugate gradient (cg) optimization was performed, with convergence thresholds set to 10 kJ mol^{-1} nm^{-1}. In both steps, solute heavy atoms were position-restrained with a force constant of 1000 kJ mol^{-1} nm^{-1}. All simulations were performed with the AMBER99SB-ILDN force field[74] using the GROMACS software package.[78] After ligand and target preparation, the dry target and its respective ligands were used as starting points for HydroDock (see next Section).

4.4. HydroDock Protocol

4.4.1. Dry Docking (Step 1). Docking was performed for both the apo and holo forms of the proteins. AutoDock Tools[66] was used to add Gasteiger-Marsili partial charges[67] to the prepared target structures and convert them to PDBQT format. The target was treated as a rigid body, while the ligand was set to be fully flexible. Prior to docking, AutoGrid4 was used to generate the grid, with the grid center set to the average coordinates of the experimental ligand structure. The grid box size, defined as the number of grid points (N_{pts}), was calculated using eq 4.

$$N_{pts} = \frac{(L_{max} + T)}{G} \tag{4}$$

where L_{max} is the maximum length of the ligand, measured between the farthest two heavy atoms in the ligand; T is a distance tolerance (10 Å) between the edge of the docking box and the ligand; and G is the grid spacing, set to 0.375 Å. No explicit water molecules were adopted from the PDB structures.

Docking was performed using the Lamarckian Genetic Algorithm,[80] with 100 docking runs, a maximum of 25 million evaluations, and a population size of 150. The resulting 100 ligand-binding modes were structurally clustered (with a tolerance of 2 Å) and ranked by their AutoDock 4.2 binding free energy values. The ligand structure with the lowest calculated free energy of binding was selected, and neighboring docked ligand structures within 2 Å were grouped into the same rank. A new rank was then opened with an unused structure of the lowest calculated free energy of binding from the remaining structures, and the process was repeated until all ligand structures were collected into ranks. The rank-1 binding pose is used to calculate $RMSD_{start}$.

4.4.2. Building the Water Structures of the Target−Ligand Interface (Step 2). To prepare the protein for a MobyWat[9,10] water prediction, first a 1 ns-MD simulation is performed. For the MD simulation GROMACS package[78] is used with the AMBER99SB-ILDN force field.[73] The MD simulation box was filled with explicit TIP3P water molecules.[79] Long-range electrostatics were handled using the PME summation,[81] employing a Fourier spacing of 0.12 nm and a grid spacing of 1 Å. van der Waals interactions were truncated at an 11 Å cutoff. The solute and solvent were coupled separately to a reference temperature of 300 K using the velocity rescale algorithm[82] with a time constant of 0.1 ps. Pressure coupling used the Parrinello−Rahman algorithm[83,84] with a reference pressure of 1 bar, a time constant of 0.5 ps, and a compressibility of 4.5×10^{-5} bar^{-1}. Position restraints were applied to all heavy atoms of the target with a force constant of 1000 kJ mol^{-1} nm^{-2}.

The resulting trajectory (md.trr) of the TIP3P solvated system, containing atomic coordinates for all frames, was converted to a portable xtc format (system.xtc) using the trjconv utility in GROMACS. Using the resulting MD trajectory file, the surface water structure of the target was calculated in MobyWat's prediction mode with its all-inclusive,

https://doi.org/10.1021/acsphyschemau.6c00072
ACS Phys. Chem Au XXXX, XXX, XXX−XXX

identity-based prediction algorithm (IDa). MobyWat predicts waters from the trajectory of the solvated MD system, starting from solvent waters in the frames, resulting in a predicted layer of water molecules. During the calculation, a maximum distance from the target (dmax) of 5 Å was used, with clustering (ctol) and prediction (ptol) tolerances of 1.5 and 2.5 Å, respectively. The resulting hydrated target structure was then placed in a common coordinate system with the target containing the $RMSD_{start}$ ligand binding pose, using Pymol.[77] Water molecules conflicting with the ligand structure were removed using the editing mode of MobyWat at a minimum distance limit (dmin) of 1.75 Å.

4.4.3. Refinement (Step 3). The hydrated complex structures were subjected to a five-step robust refinement to optimize the orientation of hydrogen atoms on the predicted water molecules, facilitating water network formation. Energy minimization was performed as described in the Target Preparation Section.

The energy-minimized structure was subjected to a 100 ps NPT MD simulation. Long-range electrostatics were treated with the PME summation,[81] using a Fourier spacing of 0.12 nm and a grid spacing of 1 Å. van der Waals interactions were truncated at 11 Å. The solute and solvent were coupled separately to a reference temperature of 300 K using the velocity rescale algorithm[82] with a time constant of 0.1 ps. Pressure coupling used the Parrinello−Rahman algorithm[83,84] with a reference pressure of 1 bar, a time constant of 0.5 ps, and a compressibility of 4.5×10^{-5} bar^{-1}. Position restraints were applied only to the backbone Cα atoms of the target, with a force constant of 1000 kJ mol^{-1} nm^{-2}. The second round of sd and cg optimizations was performed with the same settings as the first round, but with position restraints applied solely to the backbone Cα atoms. All simulations were performed with the AMBER99SB-ILDN force field[74] using the GROMACS software package.[78]

4.4.4. Generating MD Snapshots of the Target−Ligand Complex (Step 4). Finally, a 50 ns NPT MD simulation was performed on the hydrated and equilibrated complex structures using the same settings described in Step 3. Only the target Cα atoms were position-restrained, while the amino acid side chains, the ligand, and the solvent were allowed to move. The resulting trajectory file was converted to a portable xtc format using the trjconv utility of GROMACS. This conversion also accounts for periodic boundary conditions, ensuring that the solute remains intact and centered within the simulation box. First, each frame was aligned to the experimental target structure using Cα atoms via the confrms utility. Complex snapshots were extracted as individual PDB files at 0.1 ns intervals (conformation pool) and subjected to further analysis. RMSD values were calculated between each ligand binding mode extracted from the conformation pool and the corresponding reference binding mode, as described in the Evaluation metrics section. The ligand binding mode with the lowest RMSD among all extracted snapshots ($RMSD_{best}$) was used to assess the structural performance of HydroDock, while the RMSD of the top-ranked binding mode ($RMSD_{top}$) was determined as described in the Ranking section.

4.4.5. Selection of Representative Binding Mode (Step 5). An average ligand coordinate was calculated from all ligand binding modes in the conformation pool generated in Step 4 using a shell script. RMSD values were then calculated between the average ligand coordinate and each ligand binding mode in the pool. The ligand binding mode with the lowest RMSD relative to the average structure was selected as the representative binding mode. The RMSD between this representative binding mode and the experimental reference ligand binding mode is referred to as $RMSD_{representative}$.

4.5. Ranking

All structures in the conformation pool (Step 4) were ranked by their target−ligand intermolecular interaction energies (E_{inter}, eq 5) and by the components separately (i.e., Lennard−Jones and Coulomb terms). All energy calculations were performed on the trajectory file (system.xtc from Step 4) with and without water molecules using Pantherg, an in-house program. The Coulomb term ($E_{Coulomb}$) was calculated using the distance-dependent dielectric function of Mehler and Solmajer with Amber2012 force field parameters.[85]

$$E_{inter} = E_{LJ} + E_{Coulomb} = \sum_{i,j}^{N_T N_L} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon_r r_{ij}} \right)$$

$$A_{ij} = \varepsilon_{ij} R_{ij}^{12}$$

$$B_{ij} = 2\varepsilon_{ij} R_{ij}^{6}$$

$$R_{ij} = R_i + R_j$$

$$\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j}$$

$$\varepsilon_r = A + \frac{B}{1 + k e^{-\lambda Br}} \quad (5)$$

where N_T and N_L are the number of target and ligand atoms, respectively; r_{ij} is the actual distance between the ith (ligand) and jth (target) atoms; q is the partial charge of an atom; ε_0 is the permittivity of vacuum; ε_r, distance−dependent relative permittivity; ε_{ij} is the potential well depth at equilibrium; R_{ij} is the internuclear distance at equilibrium; $B = \varepsilon_0 - A$ in which ε_0 is the dielectric constant of water at 25 °C; A, λ and k are constants.[86]

The RMSD value calculated between the top-ranked binding mode and the experimental reference ligand binding mode was referred to as $RMSD_{top}$.

4.6. Evaluation Metrics

RMSD values were measured between the heavy atoms of the predicted (C in eq 5) and the experimental reference (R) ligand binding mode according to eq 6.

$$RMSD = \sqrt{\frac{1}{NH} \sum_{i=1}^{NH} |C_i - R_i|^2} \quad (6)$$

where NH is the number of heavy atoms in the ligand, R is the space vector of the ith heavy atom of the experimental reference ligand structure in the respective Protein Databank (PDB) coordinate files, and C is the space vector of the ith heavy atom of the predicted ligand structure.

4.6.1. Improvement of RMSD. The improvement of starting structures upon the HydroDock protocol was expressed by ΔRMSD. ΔRMSD is measured as the difference between $RMSD_{start}$ (the RMSD calculated for the starting ligand conformation obtained from dry docking), and $RMSD_x$, a solution produced by the HydroDock protocol, which is $RMSD_{best}$ in the case of structural performance and $RMSD_{representative}$ or $RMSD_{top}$ in the case of ranking perform-

https://doi.org/10.1021/acsphyschemau.6c00072
ACS Phys. Chem Au XXXX, XXX, XXX−XXX

ance. The improvement is also expressed as a percentage (ΔRMSD (%)) relative to $RMSD_{start}$.

4.6.2. Success Rate. The Validation mode of MobyWat was used for success rate (SR) calculation, which quantifies the match between predicted and experimental (reference) water positions. For interface water prediction the binding site is defined by the 3.5 Å radius of the ligand. The identification of a reference water position was considered successful when the predicted water position was within 1.7 and 1.5 Å from the reference position (Table S3). If the distance criterion is met, then the predicted and reference water positions are considered as matching pairs. The number of matches is divided by the number or reference water positions and multiplied by 100 to get the SR as a percentage.

■ ASSOCIATED CONTENT

Data Availability Statement

Data files are available at https://zenodo.org/records/18919800. The compressed data file contains the in- and output files of HydroDock protocol for both the holo and apo systems, the parameter files and the programs necessary to produce them, and a Table of Contents file with detailed folder descriptions. The programs necessary for producing the results, MobyWat (http://mobywat.com/) and GROMACS (https://www.gromacs.org/), are open source programs. The in-house programs that were used to calculate RMSD and interaction energy values are provided in the data depository.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsphyschemau.6c00072.

> Additional computational details, and methods, including description of docking methods; per system breakdown of structural and ranking performance, success rates for water prediction; comparison with reference methods; illustrations of dry and refined binding modes; water-mediated interaction recovery metrics; droplet molecular dynamics simulation (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Balázs Zoltán Zsidó − *Pharmacoinformatics Unit, Department of Pharmacology and Pharmacotherapy, Medical School, University of Pécs, 7624 Pécs, Hungary;* orcid.org/0000-0001-9101-9644; Email: zsido.balazs@pte.hu

Authors

Bayartsetseg Bayarsaikhan − *Pharmacoinformatics Unit, Department of Pharmacology and Pharmacotherapy, Medical School, University of Pécs, 7624 Pécs, Hungary;* Present Address: Bayartsetseg Bayarsaikhan Institute of Mathematics and Digital Technology, Mongolian Academy of Sciences, Peace Avenue 54B, 13,330 Ulaanbaatar, Mongolia

Csaba Hetényi − *Pharmacoinformatics Unit, Department of Pharmacology and Pharmacotherapy, Medical School, University of Pécs, 7624 Pécs, Hungary;* orcid.org/0000-0002-8013-971X

Complete contact information is available at: https://pubs.acs.org/10.1021/acsphyschemau.6c00072

Author Contributions

B.B.: Investigation, Writing—original draft. **C.H.**: Conceptualization, Supervision, Writing—review and editing. **B.Z.Z.**: Conceptualization, Supervision, Methodology, Writing—review and editing.

Funding

This paper was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences (BO/00189/24/5). The work was supported by the National Research, Development and Innovation Office (PharmaLab, RRF-2.3.1-21-2022-00015).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge the Digital Government Development and Project Management Ltd. for awarding us access to the Komondor HPC facility based in Hungary.

■ REFERENCES

(1) Luo, J.; Wei, W.; Waldispühl, J.; Moitessier, N. Challenges and Current Status of Computational Methods for Docking Small Molecules to Nucleic Acids. *Eur. J. Med. Chem.* **2019**, *168*, 414−425.

(2) Zsidó, B. Z.; Bayarsaikhan, B.; Börzsei, R.; Szél, V.; Mohos, V.; Hetényi, C. The Advances and Limitations of the Determination and Applications of Water Structure in Molecular Engineering. *Int. J. Mol. Sci.* **2023**, *24*, 11784.

(3) Zsidó, B. Z.; Hetényi, C. The Role of Water in Ligand Binding. *Curr. Opin. Struct. Biol.* **2021**, *67*, 1−8.

(4) Zsidó, B. Z.; Hetényi, C. Water in Drug Design: Pitfalls and Good Practices. *Expert Opin. Drug Discovery* **2025**, *20*, 745−764.

(5) Roos, G.; Vanpoucke, D. E. P.; Blossey, R.; Lensink, M. F.; Murray, J. S. Quantifying Water Hydrogen Bonding from the Surface Electrostatic Potential at Varying Iso-Density Contours. *J. Chem. Phys.* **2025**, *163*, 114112.

(6) Zsidó, B. Z.; Börzsei, R.; Szél, V.; Hetényi, C. Determination of Ligand Binding Modes in Hydrated Viral Ion Channels to Foster Drug Design and Repositioning. *J. Chem. Inf. Model.* **2021**, *61*, 4011−4022.

(7) Hu, X.; Maffucci, I.; Contini, A. Advances in the Treatment of Explicit Water Molecules in Docking and Binding Free Energy Calculations. *Curr. Med. Chem.* **2020**, *26*, 7598−7622.

(8) de Beer, S.; Vermeulen, N.; Oostenbrink, C. The Role of Water Molecules in Computational Drug Design. *Curr. Top. Med. Chem.* **2010**, *10*, 55−66.

(9) Jeszenoi, N.; Bálint, M.; Horváth, I.; Van Der Spoel, D.; Hetényi, C. Exploration of Interfacial Hydration Networks of Target-Ligand Complexes. *J. Chem. Inf. Model.* **2016**, *56*, 148−158.

(10) Jeszenoi, N.; Horváth, I.; Bálint, M.; Van Der Spoel, D.; Hetényi, C. Mobility-Based Prediction of Hydration Structures of Protein Surfaces. *Bioinformatics* **2015**, *31*, 1959−1965.

(11) Huang, N.; Shoichet, B. K. Exploiting Ordered Waters in Molecular Docking. *J. Med. Chem.* **2008**, *51*, 4862.

(12) Sun, H.; Zhao, L.; Peng, S.; Huang, N. Incorporating Replacement Free Energy of Binding-Site Waters in Molecular Docking. *Proteins* **2014**, *82*, 1765−1776.

(13) Uehara, S.; Tanaka, S. AutoDock-GIST: Incorporating Thermodynamics of Active-Site Water into Scoring Function for Accurate Protein-Ligand Docking. *Molecules* **2016**, *21*, 1604.

(14) Forli, S.; Olson, A. J. A Force Field with Discrete Displaceable Waters and Desolvation Entropy for Hydrated Ligand Docking. *J. Med. Chem.* **2012**, *55*, 623−638.

(15) Lie, M. A.; Thomsen, R.; Pedersen, C. N. S.; Schiøtt, B.; Christensen, M. H. Molecular Docking with Ligand Attached Water Molecules. *J. Chem. Inf. Model.* **2011**, *51*, 909−917.

(16) Verdonk, M. L.; Chessari, G.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Nissink, J. W. M.; Taylor, R. D.; Taylor, R. Modeling Water Molecules in Protein-Ligand Docking Using GOLD. *J. Med. Chem.* **2005**, *48*, 6504−6515.
(17) The Particle Concept: Placing Discrete Water Molecules during Protein-Ligand Docking Predictions - PubMed Available online: https://pubmed.ncbi.nlm.nih.gov/10336380/(accessed on 3 April 2026).
(18) Therrien, E.; Weill, N.; Tomberg, A.; Corbeil, C. R.; Lee, D.; Moitessier, N. Docking Ligands into Flexible and Solvated Macromolecules. 7. Impact of Protein Flexibility and Water Molecules on Docking-Based Virtual Screening Accuracy. *J. Chem. Inf. Model.* **2014**, *54*, 3198−3210.
(19) Murphy, R. B.; Repasky, M. P.; Greenwood, J. R.; Tubert-Brohman, I.; Jerome, S.; Annabhimoju, R.; Boyles, N. A.; Schmitz, C. D.; Abel, R.; Farid, R.; et al. WScore: A Flexible and Accurate Treatment of Explicit Water Molecules in Ligand-Receptor Docking. *J. Med. Chem.* **2016**, *59*, 4364−4384.
(20) Mahmoud, A. H.; Masters, M. R.; Yang, Y.; Lill, M. A. Elucidating the Multiple Roles of Hydration for Accurate Protein-Ligand Binding Prediction via Deep Learning. *Commun. Chem.* **2020**, *3*, 19.
(21) Lemmon, G.; Meiler, J. Towards Ligand Docking Including Explicit Interface Water Molecules. *PLoS One* **2013**, *8*, No. e67536.
(22) Buttenschoen, M.; Morris, G. M.; Deane, C. M. PoseBusters: AI-Based Docking Methods Fail to Generate Physically Valid Poses or Generalise to Novel Sequences. *Chem. Sci.* **2024**, *15*, 3130−3139.
(23) Dahlström, K. M.; Salminen, T. A. Apprehensions and Emerging Solutions in ML-Based Protein Structure Prediction. *Curr. Opin. Struct. Biol.* **2024**, *86*, 102819.
(24) Brough, P. A.; Barril, X.; Beswick, M.; Dymock, B. W.; Drysdale, M. J.; Wright, L.; Grant, K.; Massey, A.; Surgenor, A.; Workman, P. 3-(5-Chloro-2,4-Dihydroxyphenyl)-Pyrazole-4-Carboxamides as Inhibitors of the Hsp90 Molecular Chaperone. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 5197−5201.
(25) Shi, Q.; Tebben, A.; Dyckman, A. J.; Li, H.; Liu, C.; Lin, J.; Spergel, S.; Burke, J. R.; McIntyre, K. W.; Olini, G. C.; et al. Purine Derivatives as Potent Bruton's Tyrosine Kinase (BTK) Inhibitors for Autoimmune Diseases. *Bioorg. Med. Chem. Lett.* **2014**, *24*, 2206−2211.
(26) Wood, D. J.; Lopez-Fernandez, J. D.; Knight, L. E.; Al-Khawaldeh, I.; Gai, C.; Lin, S.; Martin, M. P.; Miller, D. C.; Cano, C.; Endicott, J. A.; et al. FragLites-Minimal, Halogenated Fragments Displaying Pharmacophore Doublets. An Efficient Approach to Druggability Assessment and Hit Generation. *J. Med. Chem.* **2019**, *62*, 3741−3752.
(27) Fiege, B.; Rabbani, S.; Preston, R. C.; Jakob, R. P.; Zihlmann, P.; Schwardt, O.; Jiang, X.; Maier, T.; Ernst, B. The Tyrosine Gate of the Bacterial Lectin FimH: A Conformational Analysis by NMR Spectroscopy and X-Ray Crystallography. *ChemBioChem* **2015**, *16*, 1235−1246.
(28) Darby, J. F.; Hopkins, A. P.; Shimizu, S.; Roberts, S. M.; Brannigan, J. A.; Turkenburg, J. P.; Thomas, G. H.; Hubbard, R. E.; Fischer, M. Water Networks Can. Determine the Affinity of Ligand Binding to Proteins. *J. Am. Chem. Soc.* **2019**, *141*, 15818−15826.
(29) McDonough, M. A.; Anderson, J. W.; Silvaggi, N. R.; Pratt, R. F.; Knox, J. R.; Kelly, J. A. Structures of Two Kinetic Intermediates Reveal Species Specificity of Penicillin-Binding Proteins. *J. Mol. Biol.* **2002**, *322*, 111−122.
(30) Kelly, J. A.; Kuzin, A. P. The Refined Crystallographic Structure of ADD-Peptidase Penicillin-Target Enzyme at 1.6 Å Resolution. *J. Mol. Biol.* **1995**, *254*, 223−236.
(31) Ringom, R.; Axen, E.; Uppenberg, J.; Lundbäck, T.; Rondahl, L.; Barf, T. Substituted Benzylamino-6-(Trifluoromethyl)Pyrimidin-4(1H)-Ones: A Novel Class of Selective Human A-FABP Inhibitors. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 4449−4452.
(32) Stebbins, C. E.; Russo, A. A.; Schneider, C.; Rosen, N.; Hartl, F. U.; Pavletich, N. P. Crystal Structure of an Hsp90-Geldanamycin Complex: Targeting of a Protein Chaperone by an Antitumor Agent. *Cell* **1997**, *89*, 239−250.
(33) Wittmann, J. G.; Heinrich, D.; Gasow, K.; Frey, A.; Diederichsen, U.; Rudolph, M. G. Structures of the Human Orotidine-5′-Monophosphate Decarboxylase Support a Covalent Mechanism and Provide a Framework for Drug Design. *Structure* **2008**, *16*, 82−92.
(34) Crennell, S. J.; Garman, E. F.; Philippon, C.; Vasella, A.; Laver, W. G.; Vimr, E. R.; Taylor, G. L. The Structures OfSalmonella TyphimuriumLT2 Neuraminidase and Its Complexes with Three Inhibitors at High Resolution. *J. Mol. Biol.* **1996**, *259*, 264−280.
(35) Müller, A.; Severi, E.; Mulligan, C.; Watts, A. G.; Kelly, D. J.; Wilson, K. S.; Wilkinson, A. J.; Thomas, G. H. Conservation of Structure and Mechanism in Primary and Secondary Transporters Exemplified by SiaP, a Sialic Acid Binding Virulence Factor from Haemophilus Influenzae. *J. Biol. Chem.* **2006**, *281*, 22212−22222.
(36) Woodhead, A. J.; Angove, H.; Carr, M. G.; Chessari, G.; Congreve, M.; Coyle, J. E.; Cosme, J.; Graham, B.; Day, P. J.; Downham, R.; et al. Discovery of (2,4-Dihydroxy-5-Isopropylphenyl)-[5-(4-Methylpiperazin-1-Ylmethyl)-1,3-Dihydroisoindol-2-Yl]-Methanone (AT13387), a Novel Inhibitor of the Molecular Chaperone Hsp90 by Fragment Based Drug Design. *J. Med. Chem.* **2010**, *53*, 5956−5969.
(37) Collins, P. M.; Öberg, C. T.; Leffler, H.; Nilsson, U. J.; Blanchard, H. Taloside Inhibitors of Galectin-1 and Galectin-3. *Chem. Biol. Drug Des.* **2012**, *79*, 339−346.
(38) Saraboji, K.; Hakansson, M.; Genheden, S.; Diehl, C.; Qvist, J.; Weininger, U.; Nilsson, U. J.; Leffler, H.; Ryde, U.; Akke, M.; et al. The Carbohydrate-Binding Site in Galectin-3 Is Preorganized To Recognize a Sugarlike Framework of Oxygens: Ultra-High-Resolution Structures and Water Dynamics. *Biochemistry* **2012**, *51*, 296−306.
(39) Ahmed-Belkacem, A.; Colliandre, L.; Ahnou, N.; Nevers, Q.; Gelin, M.; Bessin, Y.; Brillet, R.; Cala, O.; Douguet, D.; Bourguet, W.; et al. Fragment-Based Discovery of a New Family of Non-Peptidic Small-Molecule Cyclophilin Inhibitors with Potent Antiviral Activities. *Nature Communications* **2016**, *7*, 12777.
(40) Schlatter, D.; Thoma, R.; Küng, E.; Stihle, M.; Müller, F.; Borroni, E.; Cesura, A.; Hennig, M. Crystal Engineering Yields Crystals of Cyclophilin D Diffracting to 1.7 Å Resolution. urn:issn:0907−4449 **2005**, *61*, 513−519..
(41) Rabbani, S.; Fiege, B.; Eris, D.; Silbermann, M.; Jakob, R. P.; Navarra, G.; Maier, T.; Ernst, B. Conformational Switch of the Bacterial Adhesin FimH in the Absence of the Regulatory Domain: Engineering a Minimalistic Allosteric System. *J. Biol. Chem.* **2018**, *293*, 1835−1849.
(42) ER1: A Rational Approach to the Design of Antihypertensives. X-Ray Studies of Complexes between Aspartic Proteinases and Aminoalcohol Renin Inhibitors Available online: https://www.ncbi.nlm.nih.gov/Structure/pdb/5ER1 (accessed on 3 April 2026).
(43) Pearl, L.; Blundell, T. The Active Site of Aspartic Proteinases. *FEBS Lett.* **1984**, *174*, 96−101.
(44) Maibaum, J.; Liao, S. M.; Vulpetti, A.; Ostermann, N.; Randl, S.; Rüdisser, S.; Lorthiois, E.; Erbel, P.; Kinzel, B.; Kolb, F. A.; et al. Small-Molecule Factor D Inhibitors Targeting the Alternative Complement Pathway. *Nat. Chem. Biol.* **2016**, *12*, 1105−1110.
(45) Forneris, F.; Ricklin, D.; Wu, J.; Tzekou, A.; Wallace, R. S.; Lambris, J. D.; Gros, P. Structures of C3b in Complex with Factors B and D Give Insight into Complement Convertase Formation. *Science (1979).* **2010**, *330*, 1816−1820.
(46) Hsieh, T. J.; Lin, H. Y.; Tu, Z.; Lin, T. C.; Wu, S. C.; Tseng, Y. Y.; Liu, F. T.; Hsu, S. T. D.; Lin, C. H. Dual Thio-Digalactoside-Binding Modes of Human Galectins as the Structural Basis for the Design of Potent and Selective Inhibitors. *Sci. Rep.* **2016**, *6*, 29457.
(47) Mota, F.; Fotinou, C.; Rana, R. R.; Chan, A. W. E.; Yelland, T.; Arooz, M. T.; O'Leary, A. P.; Hutton, J.; Frankel, P.; Zachary, I.; et al. Architecture and Hydration of the Arginine-Binding Site of Neuropilin-1. *FEBS J.* **2018**, *285*, 1290−1304.

(48) Lee, C. C.; Kreusch, A.; McMullan, D.; Ng, K.; Spraggon, G. Crystal Structure of the Human Neuropilin-1 B1 Domain. *Structure* **2003**, *11*, 99−108.
(49) Lorthiois, E.; Anderson, K.; Vulpetti, A.; Rogel, O.; Cumin, F.; Ostermann, N.; Steinbacher, S.; Mac Sweeney, A.; Delgado, O.; Liao, S. M.; et al. Discovery of Highly Potent and Selective Small-Molecule Reversible Factor D Inhibitors Demonstrating Alternative Complement Pathway Inhibition in Vivo. *J. Med. Chem.* **2017**, *60*, 5717−5735.
(50) Heinzlmeir, S.; Lohse, J.; Treiber, T.; Kudlinzki, D.; Linhard, V.; Gande, S. L.; Sreeramulu, S.; Saxena, K.; Liu, X.; Wilhelm, M.; et al. Chemoproteomics-Aided Medicinal Chemistry for the Discovery of EPHA2 Inhibitors. *ChemMedChem* **2017**, *12*, 999−1011.
(51) Schulze-Gahmen, U.; De Bondt, H. L.; Kim, S. H. High-Resolution Crystal Structures of Human Cyclin-Dependent Kinase 2 with and without ATP: Bound Waters and Natural Ligand as Guides for Inhibitor Design†. *J. Med. Chem.* **1996**, *39*, 4540−4546.
(52) Glen, R. C.; Cole, J. C.; Stewart, J. J. P. Prediction of Enzyme Inhibition (IC50) Using a Combination of Protein-Ligand Docking and Semiempirical Quantum Mechanics. *J. Mol. Model.* **2025**, *31*, 209.
(53) Li, Y.; Zhang, Y.; Qi, M.; Yi, L.; Yu, Y.; Yang, W.; Hu, Y. Solubility Behavior of Diosmetin in Pure and Binary Solvents: Thermodynamic Evaluation and Molecular-Level Interpretation. *Fluid Phase Equilib.* **2026**, *602*, 114611.
(54) Mieres-Perez, J.; Almeida-Hernandez, Y.; Sander, W.; Sanchez-Garcia, E. A Computational Perspective to Intermolecular Interactions and the Role of the Solvent on Regulating Protein Properties. *Chem. Rev.* **2025**, *125*, 7023−7056.
(55) Biedermannová, L.; Schneider, B. Hydration of Proteins and Nucleic Acids: Advances in Experiment and Theory. A Review. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2016**, *1860*, 1821−1835.
(56) Mattos, C.; Ringe, D. Solvent Structure. In *International Tables for Crystallography*; International Union of Crystallography: Chester, England, 2006; pp 623−647.
(57) Ladbury, J. E. Just Add Water! The Effect of Water on the Specificity of Protein-Ligand Binding Sites and Its Potential Application to Drug Design. *Chem. Biol.* **1996**, *3*, 973−980.
(58) Horváth, I.; Jeszenői; Norbert, M. B. Gábor Paragi and Csaba Hetényi A Fragmenting Protocol with Explicit Hydration for Calculation of Binding Enthalpies of Target-Ligand Complexes at a Quantum Mechanical Level. *Int. J. Mol. Sci.* **2019**, *20*.
(59) Lu, J.; Hou, X.; Wang, C.; Zhang, Y. Incorporating Explicit Water Molecules and Ligand Conformation Stability in Machine-Learning Scoring Functions. *J. Chem. Inf. Model.* **2019**, *59*, 4540−4549.
(60) Eberhardt, J.; Santos-Martins, D.; Tillack, A. F.; Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J. Chem. Inf. Model.* **2021**, *61*, 3891−3898.
(61) Kolokouris, D.; Kalenderoglou, I. E.; Kolocouris, A. Inside and Out of the Pore: Comparing Interactions and Molecular Dynamics of Influenza A M2 Viroporin Complexes in Standard Lipid Bilayers. *J. Chem. Inf. Model.* **2021**, *61*, 5550−5568.
(62) Chandler, D. Interfaces and the Driving Force of Hydrophobic Assembly. *Nature* **2005**, *437*, 640−647.
(63) Southall, N. T.; Dill, K. A.; Haymet, A. D. J. A View of the Hydrophobic Effect. *J. Phys. Chem. B* **2002**, *106*, 521−533.
(64) Cheng, Y.-K.; Rossky, P. J. Surface Topography Dependence of Biomolecular Hydrophobic Hydration. *Nature* **1998**, *392*, 696−699.
(65) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(66) Morris, G. M.; Ruth, H.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **2009**, *30*, 2785−2791.
(67) Gasteiger, J.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity-a Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219−3228.
(68) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *Chem. inf.* **2011**, *3*, 33.
(69) Halgren, T. Merck Molecular Force Field. V. Extension of MMFF94 Using Experimental Data, Additional Computational Data, and Empirical Rules. *J. Comput. Chem.* **1996**.
(70) Stewart, J. J. P. *MOPAC2016*: Stewart Computational Chemistry: Colorado Springs, CO 2016
(71) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods VI: More Modifications to the NDDO Approximations and Re-Optimization of Parameters. *J. Mol. Model.* **2013**, *19*, 1−32.
(72) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859−1865.
(73) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(74) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber Ff99SB Protein Force Field. *Proteins* **2010**, *78*, 1950−1958.
(75) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; De Beer, T. A. P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Res.* **2018**, *46*, W296−W303.
(76) Bateman, A.; Martin, M. J.; Orchard, S.; Magrane, M.; Adesina, A.; Ahmad, S.; Bowler-Barnett, E. H.; Bye-A-Jee, H.; Carpentier, D.; Denny, P.; et al. UniProt: The Universal Protein Knowledgebase in 2025. *Nucleic Acids Res.* **2025**, *53*, D609−D617.
(77) DeLano, W.L. *The PyMOL Molecular Graphics System* Schrödinger, LLC 2021
(78) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26*, 1701−1718.
(79) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(80) Morris, G.; Goodsell, D.; Halliday, R. S.; Huey, R.; Hart, W.; Belew, R.; Olson, A. Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *J. Comput. Chem.* **1998**.
(81) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(82) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
(83) Parrinello, M.; Rahman, A. Crystal Structure and Pair Potentials: A Molecular-Dynamics Study. *Phys. Rev. Lett.* **1980**, *45*, 1196.
(84) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(85) Wang, J.; Cieplak, P.; Li, J.; Cai, Q.; Hsieh, M. J.; Luo, R.; Duan, Y. Development of Polarizable Models for Molecular Mechanical Calculations. 4. van Der Waals Parametrization. *J. Phys. Chem. B* **2012**, *116*, 7088−7101.
(86) Mehler, E. L.; Solmajer, T. Electrostatic Effects in Proteins: Comparison of Dielectric and Charge Models. *Protein Eng.* **1991**, *4*, 903−910.

https://doi.org/10.1021/acsphyschemau.6c00072
ACS Phys. Chem Au XXXX, XXX, XXX−XXX