From Water Networks to Binding Affinities: Resolving Solvation Dynamics for Accurate Protein-Ligand Predictions

Marharyta Blazhynska,[†,‡] Narjes Ansari,[¶] Louis Lagardère,[*,†] and Jean-Philip Piquemal[*,†]

†*Sorbonne Université, Laboratoire de Chimie Théorique, UMR 7616 CNRS, 75252 Paris Cedex 05, France*

‡*Present adress: Qubit Pharmaceuticals, Advanced Research Department, 75014 Paris, France*

¶*Qubit Pharmaceuticals, Advanced Research Department, 75014 Paris, France*

E-mail: louis.lagardere@sorbonne-universite.fr; jean-philip.piquemal@sorbonne-universite.fr

Abstract

Water molecules play a critical role in mediating protein–ligand interactions by forming bridging hydrogen bonds and contributing to ligand solvation. However, their intricate behavior, such as frequent exchange with bulk solvent or persistent stabilization in the binding site, makes the accurate binding free-energy estimation via molecular dynamics-based approaches challenging. Particularly, inadequate sampling of water reorganization might not only bias computed affinities but also obscure key interactions, making adequate rehydration of the binding site violated upon calculations. To address this, we employ the polarizable AMOEBA force field together with Lambda-ABF-OPES, an integrated enhanced-sampling framework, which combines λ-dynamics, multiple-walker adaptive biasing force, and exploratory version of on-the-fly probability enhanced sampling technique. Such approach enables efficient rehydration of the binding site and allows robust sampling of water exchange and reorganization events without explicitly including any water-related collective variable. Applied to five water-containing protein–ligand complexes with diverse ligand types and binding-site environments, the approach yields binding affinities efficiently and in good agreement with experimental data, demonstrating that Lambda-ABF-OPES captures dynamic water networks and provides robust and reproducible absolute binding free-energy estimation towards chemical accuracy.

Introduction

The structural and dynamic properties of water are fundamental to governing molecular recognition and the specificity of protein–ligand binding.[1–4] At the binding interface, water molecules can bridge protein and ligand interactions, potentially modulating both their affinity and selectivity. Structural studies reveal that some crystallographically leveraged interfacial waters may form stable hydrogen-bonding bridges that reinforce recognition, whereas others exchange rapidly with bulk solvent, dynamically contributing to the specificity and strength of binding.[2,3,5,6] Such behavior presents substantial complexity for the accurate evaluation of absolute binding free energies, particularly due to the difficulties in adequately sampling the associated solvation process.[5,7–10]

To account for hydration effects, several Monte Carlo (MC)-based water placement and rehydration methods have been introduced in the literature.[11–14] While powerful for hydration mapping, such methods face important limitations for rigorous binding free-energy calculations. In particular, MC moves are generally independent of protein and ligand conformational dynamics, preventing a realistic description of the coupled reorganization of solvent, binding pose, and side chains. Moreover, their stochastic sampling can leave kinetically relevant water exchange pathways undersampled, leading to incomplete rehydration and potentially biased estimates of binding affinities.[13]

A widely employed molecular-dynamics (MD) based approach for estimating binding affinities is the molecular mechanics/Poisson-Boltzmann or generalized Boltzmann surface area (MM-PBSA/GBSA) strategies, which combine molecular mechanics (MM) energies, solvation free energies, and entropic contributions derived from MD simulations.[15,16] Herein, the solvation component approximates the free energy associated with transferring the isolated ligand, the protein, and their complex from vacuum to solvent. Despite its computational efficiency and broad applicability, MM-PBSA/GBSA tend to underestimate water-mediated contributions and neglect dynamic rearrangements of interfacial waters and protein side chains, leading to limited accuracy for complexes with substantial solvent displacement or conformational change.[17–24]

Theory-driven alternatives to the discussed end-state MD and MC–based strategies employ

enhanced-sampling algorithms that explore the free-energy landscape along carefully chosen collective variables (CVs) governing the slow degrees of freedom in protein–ligand complexes.[25] These methods aim to overcome sampling bottlenecks and provide a more rigorous statistical-mechanical foundation for binding free-energy estimation. Prominent representatives include geometry-,[23,26] metadynamics-,[27–29] and alchemical-based (i.e, free-energy perturbation (FEP)) schemes.[23,30–38] Depending on the chosen formalism, the ligand is either physically separated from the binding site or reversibly decoupled/annihilated through a series of intermediate states. In both cases, the vacated pocket becomes accessible to solvent molecules, and water infiltration can alter critical protein–ligand interactions or hinder the reformation of bridging waters during ligand reassociation. An accurate treatment of this rehydration process is therefore essential, as uncontrolled water behavior can compromise both the structural integrity of the binding interface and the reliability of computed thermodynamics. Faithful sampling of pocket rehydration remains particularly challenging for buried or semi-buried ligands, and has been shown to be crucial for achieving quantitative agreement with experiment in alchemical simulations.[9,39,40]

Among the methodologies developed to explicitly account for solvent behavior, implementation of water-based CVs have recently emerged as a powerful extension of enhanced-sampling methods.[41–43] These approaches incorporate water occupancy, density, and hydrogen-bonding patterns as CVs to systematically capture the coupled dynamics of water, ligand, and protein. By treating key water molecules as active participants in conformational transitions, such strategy enable a realistic description of hydration effects and water-mediated rearrangements that remain largely inaccessible to traditional MD or MC sampling. While highly accurate, water-based CV approaches are computationally demanding, and their success relies on the careful design and validation of system-specific water descriptors.

To address these challenges, we probed Lambda-ABF-OPES,[44,45] an alchemical-based methodology building upon λ-dynamics.[46,47] In this framework, the alchemical parameter, λ, which modulates van der Waals and electrostatic interactions between the ligand and its environment in both bound and unbound states, is treated as a continuous dynamical variable rather than being fixed at

discrete intervals. By integrating λ-dynamics with multiple-walker adaptive biasing force (ABF) and the exploratory version of the on-the-fly probability enhanced sampling (OPES), known as OPES-Explore or OPES$_e$[48], the approach facilitates efficient exploration of relevant configurational space. Specifically, multiple-walker ABF enables independent simulations to exchange biasing information, while OPES$_e$ constructs a bias potential adaptively from the instantaneous estimate of the free-energy surface, avoiding overfilling, reducing noise relative to Gaussian deposition schemes, and accelerating convergence compared to well-tempered metadynamics, which relies on pre-defined Gaussian kernels that can lead to slower adaptation to the underlying free-energy landscape.[49–51] Together, these components enhance sampling along both λ and orthogonal degrees of freedom, effectively overcoming barriers associated with ligand rearrangement, solvation effects, and protein conformational changes.

Additionally, efficient sampling depends on the careful selection of a set of CVs to confine the accessible phase space when the ligand is partially or fully decoupled, thereby improving convergence.[52,53] To this end, the Lambda-ABF-OPES scheme employs a single distance-to-bound configuration (DBC) CV[45,54], defined as the root-mean-square deviation (RMSD) of the ligand relative to a moving reference of the protein binding site. A flat-bottom harmonic restraint applied to the DBC introduces an adaptive boundary, restraining the ligand position, orientation, and conformation while still allowing relevant conformational changes. Such strategy ensures comprehensive sampling of essential binding-site configurations while avoiding less informative states, thereby enhancing both efficiency and reliability of free energy calculations.

It should be noted that the accuracy of absolute binding free-energy estimates also critically depends on the choice of force field. Classical fixed-charge force fields often fail to capture the subtleties of polarization and water-mediated effects, whereas polarizable models[55–57] offer a more realistic description of solvation and protein–ligand interactions.[58,59] In particular, polarizable water models allow dynamic adjustment of dipole moments, yielding hydration structures and water occupancies that are consistent with experiment.[60–63] Combining Lambda-ABF-OPES with the polarizable AMOEBA force field[64] allows for an accurate modeling of polarization effects

and water-mediated interactions across five ligand–protein complexes encompassing diverse ligand chemotypes (cyclic and aliphatic), binding affinities (moderate to strong), and pocket topologies (buried and semi-buried) (Table 1). These complexes span bromodomains, heat shock proteins, and a major urinary protein, covering a range of binding environments where water-mediated and polarization effects are expected to be critical.[9,13,65] The resulting binding affinities exhibit good agreement with experimental data (Table 2), highlighting that Lambda-ABF-OPES provides accurate, reproducible, and computationally efficient free-energy estimation, while explicitly resolving rehydration dynamics essential to solvent-mediated binding.

Table 1. Overview of Studied Protein-Ligand Complexes and Binding Environments

PDB id	Protein	Ligand	Resolution (Å)	Complex type	Total atoms
5I1Q[66]	Second bromodomain of Transcription initiation factor 1 (TAF1(2))	3-[6-(but-3-en-1-yl)-7-oxo-6,7-dihydro-1H-pyrrolo[2,3-c]pyridin-4-yl]-N,N-dimethylbenzamide	1.50	semi-buried	61,679
1I05[65]	Major urinary protein 1	6-hydroxy-6-methyl-heptan-3-one	2.0	buried	51,743
2XJG[67]	Heat shock protein HSP90-α	(1,3-dihydro-isoindol-2-yl)-(2-hydroxy-5-isopropyl-4-methoxy-phenyl)-methanone	2.25	semi-buried	37,806
2XAB[67]	Heat shock protein HSP90-α	(1,3-dihydro-isoindol-2-yl)-(2,4-dihydroxy-5-isopropyl-phenyl)-methanone	1.90	semi-buried	57,296
5I40[66]	Bromo - domain containing protein 9	6-methyl-1,6-dihydro-7H-pyrrolo[2,3-c]pyridin-7-one	1.04	semi-buried	37,935

Methods

Molecular Dynamics Protocol

Each protein–ligand complex was first subjected to CHARMM-GUI platform[68,69] to generate a fully solvated system with neutralizing counterions and 0.15 M NaCl. The resulting PDB file was then split into separate protein, ligand, water, and ion components, each converted to Tinker XYZ format using the *pdbxyz* program from the Tinker suite.[70,71] Ligand coordinates were converted to SDF format and processed with Poltype[72,73] to generate the corresponding force field parameter set, including multipoles and polarizabilities. The complete system was subsequently reassembled using the *xyzedit* program [70,71] to produce the final Tinker-HP[74] input configuration.

Each studied protein-ligand complex underwent an energy minimization using a limited-memory Broyden–Fletcher-Goldfarb–Shanno (L-BFGS) algorithm. The minimization was carried out in two steps: first, the complex was minimized without the inclusion of the electrostatic and polarizable terms to optimize the initial structure and alleviate steric clashes. In the second step, the minimization was repeated with the addition of these terms to ensuring an accurate account for polarizable effects. Afterwards, the complexes were adjusted to a 1-ns pre-equilibration in NVT ensemble to stabilize temperature fluctuations of solvent while maintaining a fixed volume. MD production simulations in the NPT ensemble (300 K, 1 atm) were then conducted using the BAOAB-RESPA multiple-step integrator[75], implemented within the GPU-accelerated Tinker-HP software package[76,77]. The hydrogen mass repartitioning[78] was applied to enable a 3-fs integration time step. The simulations were conducted with an Ewald cutoff of 7 Å and a van der Waals cutoff of 12 Å, balancing computational efficiency and accuracy in long-range interactions.

Design of the Distance-to-Bound-Configuration Collective Variable

An accurate definition of the DBC CV is essential for restraining the ligand within the binding site while allowing for its relevant conformational flexibility.[34,79] The DBC requires two atom subsets: (i) rigid ligand atoms used to describe the ligand relative orientation and conformation to

the binding site, and (ii) a set of protein anchor atoms that provide a moving reference frame. To identify an optimal ligand–anchor pair, we employed a systematic procedure based on conventional MD trajectory analysis. Such CVs can be defined, monitored and biased through the use of the Colvars[52] library linked with the molecular dynamics engine.

We define the anchor region as the set of C_α atoms located within 6 Å of the ligand in the reference structure. These atoms are spatially close enough to represent the local binding-site environment while minimizing contributions from distal protein motions. From this initial pool, only atoms exhibiting minimal fluctuations during conventional MD simulations are retained. Similarly, for the ligand, all heavy atoms (i.e., non-hydrogen atoms) are initially considered, and only those maintaining structural stability are selected. Specifically, atomic stability is quantified by computing the root mean square fluctuation (RMSF) over the trajectory. Atoms with RMSF values below a user-defined cutoff (typically 0.6–1.0 Å) are selected. This ensures that only conformationally rigid atoms (i.e., least affected by thermal motion) contribute to the DBC calculation. For highly rigid complexes, a fixed RMSF cutoff of 0.6 Å is usually sufficient, whereas more flexible systems may require a higher threshold.

The distribution of DBC values across the trajectory reflects the range of native fluctuations between the ligand and anchor atoms. The upper restraint limit (DBC_{max}) is defined as the 95th percentile of this distribution, accommodating natural motion without artificially biasing the system. The selected ligand–anchor pair is then subjected to a flat-bottom harmonic restraint with a force constant of 100 $kcal \cdot mol^{-1} \cdot Å^{-2}$[54]. Reference coordinates are taken from the last frame of the conventional MD trajectory to ensure the DBC CV is defined relative to an equilibrated configuration. This selective procedure not only provides a physically meaningful definition of the DBC CV but also offers an implicit check on complex dynamics. For instance, if a stringent DBC (e.g., 0.7 Å) results in very few anchor atoms being selected, this may indicate substantial flexibility in the protein near the binding site, potentially impacting the reliability of the restraint.

It should be noted that although water molecules were monitored throughout the binding free energy calculations, they were not included in the DBC selection due to their high mobility and

dynamic exchange between the bulk and the binding site, which could introduce excessive fluctuations, reducing the robustness of the CV. Besides, their inclusion could also obscure the conformational changes of the protein-ligand complex, making it more challenging to obtain a well-converged free-energy profile. To enhance sampling efficiency and improve convergence of the free energy profiles, we employed an enhanced sampling strategy, the details of which are discussed in the next subsection.

Lambda-ABF-OPES Enhanced Sampling for Binding Free Energy

Accurate binding free energy calculations require efficient sampling of the alchemical dynamic variable, λ, which governs the decoupling of the ligand and the accompanying reorganization of solvent and protein. To address this, we employed Lambda-ABF-OPES, a hybrid scheme that combines the multiple-walker ABF algorithm [49,80] with OPES$_e$, an adaptive enhanced-sampling technique favoring rapid exploration of relevant configurational space.[48,81,82] In practice, the alchemical variable is treated as an extended CV via the Colvars library[52] coupled with Tinker-HP. Four independent walkers were run in parallel for each electrostatic and van der Waals free-energy contributions in bound and unbound states. To ensure statistical robustness, the entire simulation set was repeated three times, thereby improving the reliability of the resulting free energy estimates.

The smooth evolution of λ is stabilized using Langevin dynamics. The ABF component accelerates convergence by distributing the bias across multiple independent walkers that share mean-force information, producing accurate force estimates with reduced statistical noise. Additionally, OPESe constructs an analytical bias from a kernel density estimate of the sampled distribution. Using a well-tempered target, OPESe flattens free-energy barriers while preserving the overall landscape.[27,81,83,84] An important advantage of including OPES$_e$ is its ability to favor controlled exploration of the λ space and thus capture transition states that emerge during alchemical transformations, including solvent rearrangements or protein conformational changes, while avoiding kinetic trapping.

Acting synergistically, OPES$_e$ drives broad exploration of λ, whereas the multiple-walker ABF refines the free-energy profile through replicated local force estimation. Together, these components form a robust and computationally efficient framework for sampling complex alchemical transformations and estimating the associated free-energy differences. Further technical details on Lambda-ABF-OPES are provided in Refs. 44,45. Building on this sampling framework, we next describe the thermodynamic cycle and restraint strategy used to compute standard binding free energies, including the application of DBC restraints and the decomposition of van der Waals and electrostatic contributions along the alchemical pathway.

Thermodynamic Framework and Restraint Strategy

The thermodynamic cycle used to compute the standard binding free energy, $\Delta G^{\circ}_{\text{bind}}$, follows three restrained legs connecting the bound and unbound states (Fig. 1):

- In the bound complex (step 1), a DBC flat-bottom harmonic restraint is applied to prevent ligand drift during decoupling in the solution. This restraint is active only during alchemical van der Waals and electrostatic transformations in the site. The free energy associated with this process is $\Delta G^{*}_{\text{site}}$. In these transformations, a soft-core potential[85,86] was employed for the van der Waals interactions to avoid singularities and atom clashes during decoupling. The pairwise 14-7 buffered potential was empirically fine-tuned to:

$$V_{\text{sc}}(r_{ij};\lambda) = \lambda^{\beta}\,\epsilon_{ij}\,t_1 t_2$$

$$t_1 = \frac{(1+\delta)^{\,n-m}}{\alpha(1-\lambda)^2 + (\rho+\delta)^{\,n-m}}, \quad t_2 = \frac{1+\gamma}{\alpha(1-\lambda)^2 + \rho^m + \gamma} - 2$$

$$\rho = \frac{r_{ij}}{r_{\text{min},ij}}, \quad n = 14,\ m = 7,\ \delta = 0.07,\ \gamma = 0.12,\ \beta = 3,\ \alpha = 0.2,$$

where r_{ij} is the interatomic distance, ϵ_{ij} is the is the well-depth, λ is the alchemical coupling parameter, ρ is the interatomic distance normalized by the minimum energy distance, $r_{\min,ij}$. The exponent, β, of 3 was found to effectively control the steepness of the soft-core repulsion to prevent atom overlap, while the scaling factor, α, of 0.2 ensures smooth interpolation of the potential during decoupling.

- The unbound ligand in solution is then decoupled from its condensed phase to the gas phase under no restraints, evaluating both van der Waals and electrostatic contributions (step 2). $\Delta G^*_{\mathrm{bulk}}$ captures the complete removal of solute–solvent interactions for the unbound state.

- The DBC flat-bottom restraint was designed to exert negligible influence in the fully coupled (bound) state, yet it substantially reduces the configurational entropy of the fully decoupled ligand in the gas phase,[54]. To quantify this penalty, ΔG_{DBC} was evaluated in the gas phase using thermodynamic integration (TI) (step 3).

Three independent simulations were performed to assess variance of the free energy contribution, with the coupling parameter varied from 0 (no DBC applied) to 1 (fully applied DBC restraint) (see Supplementary Information (SI) for more details). In these simulations, harmonic flat-bottom restraints with a force constant of 100 kcal·mol^{-1}·Å^{-2} were applied to both the DBC and the ligand center-of-mass (COM) displacement CVs. The DBC cutoff is set based on the natural fluctuations of the ligand in the bound state. The COM radius is typically increased by 1 Å relative to the DBC to ensure that translational motion is not artificially more restrictive than the configurational definition of binding. When the prescribed DBC width is already $\geq$ 1 Å, however, this additional margin is unnecessary. The restraint limit already permits sufficient flexibility, so the COM upper wall is applied directly without adding the extra offset.

The free-energy cost of this volumetric confinement, $\Delta G^\circ_{\mathrm{V}}$ is then calculated analytically by treating the ligand COM as a point particle restricted to a sphere of radius, R. This correction term accounts for the translational entropy loss of the gas-phase ligand due to COM confinement and ensures proper normalization to the bulk molecular density. Considering that the volume per

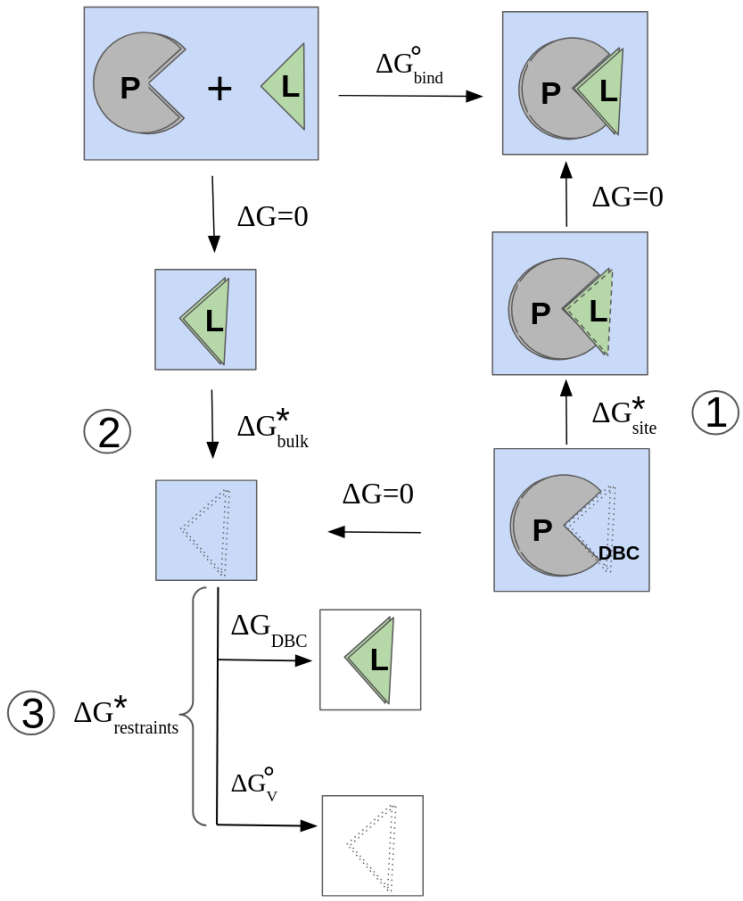

Figure 1. Thermodynamic cycle used to compute the standard binding free energy via alchemical transformations. Step 1: A DBC flat-bottom harmonic restraint is applied to the ligand in the bound complex to prevent drift during van der Waals and electrostatic decoupling in the binding site (ΔG^*_{site}). Step 2: The ligand is fully decoupled in solution under no restraints, capturing the free energy of removing solute–solvent interactions (ΔG^*_{bulk}). Step 3: The free energy cost associated with applying the DBC restraint in the gas phase is computed, including the configurational and translational entropy penalties (ΔG^*_{gas}).

molecule in the bulk is V, the analytical form becomes:

$$\Delta G^\circ_{\text{V}} = -\frac{1}{\beta} \ln \left(\frac{V_R}{V} \right),$$

where V is the system volume, $1/\beta = k_{\text{B}}T$ is the Boltzmann constant times temperature, and $V_R = \frac{4}{3}\pi r_R^3$ is the accessible spherical volume under the COM restraint.

In total, the standard binding free energy is obtained as:

$$\Delta G^{\circ}_{\mathrm{bind}} = \Delta G^{*}_{\mathrm{site}} - \Delta G^{*}_{\mathrm{bulk}} + \Delta G^{*}_{\mathrm{gas}}, \text{where}$$

$$\Delta G^{*}_{\mathrm{gas}} = \Delta G_{\mathrm{DBC}} + \Delta G^{\circ}_{\mathrm{V}},$$

$\Delta G^{*}_{\mathrm{site}}$ and $\Delta G^{*}_{\mathrm{bulk}}$ consist of two additive terms each raising from the van der Waals and electrostatic contributions.

Results and Discussions

The decomposed free-energy contributions, along with the comparison between calculated and experimental binding affinities for all complexes, are summarized in Table 2. In this section, we focus on two representative protein–ligand complexes, chosen to illustrate distinct mechanisms of water-mediated interactions within the binding site. These examples highlight the nuanced interplay between water mobility and binding affinity: (i) a dynamic regime in which water molecules are continuously exchanged, yet critical water-bridging interactions are consistently maintained (5I1Q complex), and (ii) a system where water molecules remain largely stable, forming persistent bridging interactions (1I05 complex).

All necessary starting configurations and input files, including the DBC collective variables used in the condensed-phase simulations (conventional MD and binding free energy calculations), are provided in the accompanying ZIP archive in SI. The remaining setup instructions for Colvars configuration in unbound and gas calculations are provided in SI and in the online git-hub repository of Lambda-ABF-OPES.[87] A brief discussion of the remaining complexes is presented at the end of this section, with additional binding affinity and water mobility details also available in SI.

Table 2. Decomposed Contributions to Standard Binding Affinities for Protein–Ligand Complexes.

PDB id	ΔG^*_{site}		ΔG^*_{bulk}		ΔG^*_{gas}		ΔG°_{bind}	ΔG_{exp}	$t^{replica}_{tot}$, ns[a]
	ΔG^{site}_{ELEC}	ΔG^{site}_{VdW}	ΔG^{bulk}_{ELEC}	ΔG^{bulk}_{VdW}	ΔG_{DBC}	ΔG°_V			
5I1Q[66]	-56.3 ± 0.9	25.4 ± 0.2	-48.0 ± 0.1	37.9 ± 0.1	8.5 ± 0.0	2.4	-9.9 ± 0.3	-10.0 (TR-FRET)[66]	15
1I05[65]	-24.8 ± 0.1	3.9 ± 0.2	-29.9 ± 0.1	21.0 ± 0.1	2.0 ± 0.1	2.5	-6.3 ± 0.0	-6.0 (ITC)[88]	22
2XJG*[67]	-57.7 ± 0.0	20.1 ± 0.1	-48.9 ± 0.0	36.6 ± 0.4	10.6 ± 0.2	2.5	-12.2 ± 0.1	-10.7 (ITC)[67]	25
2XAB*[67]	-85.0 ± 0.3	43.0 ± 0.9	-51.9 ± 0.1	32.4 ± 0.1	7.4 ± 0.6	3.1	-12.0 ± 0.4	-12.6 (ITC)[67]	16
5I40[66]	-33.5 ± 0.1	2.9 ± 0.2	-28.7 ± 0.0	9.6 ± 0.1	2.1 ± 0.1	2.4	-7.0 ± 0.1	-6.7 (TR-FRET)[66]	23

All ΔG values are in kcal·mol^{-1}. $t^{replica}_{tot}$ denotes the total simulation time required for convergence in solvated simulations. The superscript [a] indicates an additional 20 ns used for ΔG_{DBC} calculations in the gas phase. *For 2XJG and 2XAB, values correspond to the models used in the main text (truncated and full, respectively); comparative data for alternative models are reported in Table S1.

5I1Q: Semi-buried Ligand Binding

Conventional MD Analysis

The second bromodomain of TAF1 (TAF1(2)) is a reader domain that recognizes acetylated lysines on histones and other proteins, thereby playing a critical role in transcriptional regulation. Its hydrophobic binding cavity accommodates asparagine and tyrosine residues that contribute to recognition through direct and water-mediated hydrogen bonds. While this function is essential for normal cellular processes, dysregulation of TAF1(2) has been implicated in neurodegeneration, cancer, and cardiovascular disorders.[89–91] To inhibit TAF1(2), small-molecule ligands target its binding pocket, either by mimicking acetyl-lysine interactions or by disrupting critical hydrogen-bonding and hydrophobic contacts.[66,92,93] Among these, pyrrolopyridone derivatives have emerged as promising scaffolds due to their capacity to create both direct and water-mediated interactions within the bromodomain site.

We first examined the binding properties of butenyl-pyrrolo-pyridinone-dimethylbenzamide (Fig. 2A). The crystal structure reported by Crawford et al.[66] reveals a water-mediated hydrogen-bonding network that bridges the ligand and conserved residues within the binding site. Hydrophobic stabilization is provided by the dimethylbenzamide moiety, while the butenyl side chain contributes conformational adaptability that may optimize ligand accommodation within the pocket.

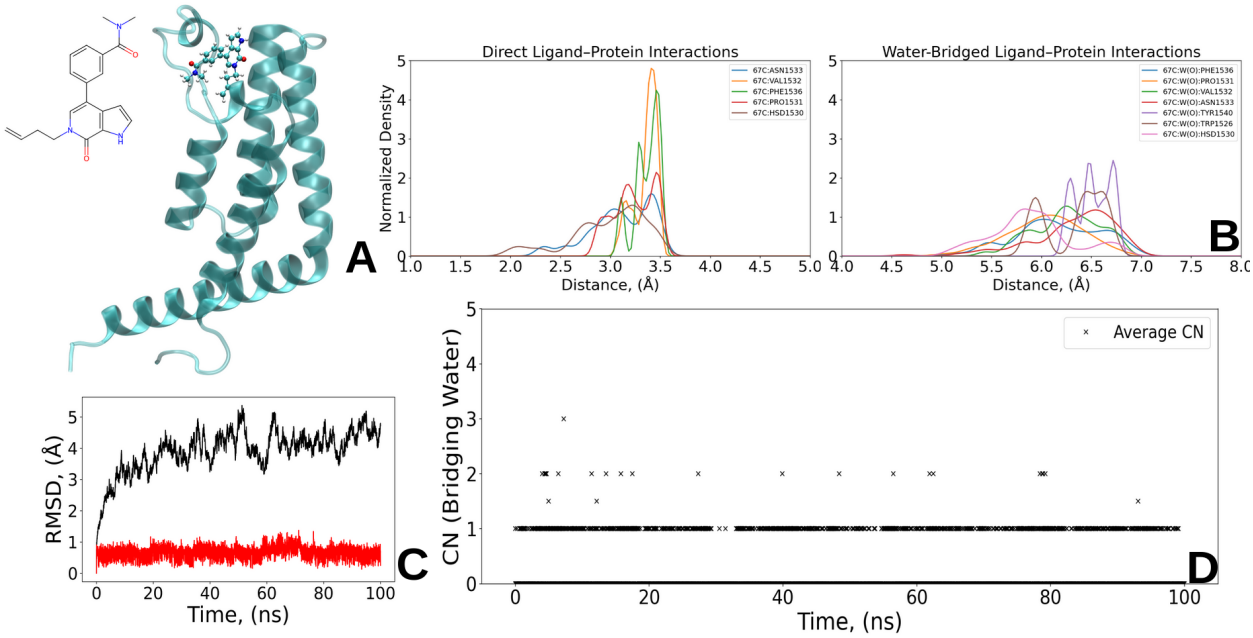

Figure 2. (A) Structure of the second bromodomain of TAF1(2) in complex with a semi-buried pyrrolopyridone derivative. (B) Density distributions of ligand–protein interactions: direct hydrogen bonds (left) and water-mediated O-bridging hydrogen bonds within the binding site (right). (C) RMSD evolution of protein C_α atoms (black) and ligand heavy atoms (red) over the 100-ns conventional MD simulation. (D) Average coordination number (CN) of O-bridging water molecules involved in mediating protein–ligand interactions.

To elucidate the dynamic behavior of these interactions, we performed a 100-ns conventional MD simulation and analyzed the hydrogen-bond patterns, solvation structure, and water coordination dynamics (Fig. 2B–D).

The density distributions show that oxygen atoms of the ligand and protein engage in both direct and water-mediated hydrogen bonding, with intervening water oxygen atoms transiently bridging ligand–protein oxygen pairs to form the O-bridging motif (i.e., where oxygen atoms of both protein and ligand are involved; see Fig. 2B and Fig. S1). Additional N/O-bridging motifs, in which either the protein or ligand nitrogen atoms participate in water-mediated links, further enrich the hydration network (Fig. S2–S4 in the SI). For example, residues ASN1533, HSD1530, VAL1532, and PRO1531 form direct interactions involving protein nitrogen–ligand nitrogen, protein nitrogen–ligand oxygen, protein oxygen–ligand nitrogen, and protein oxygen–ligand oxygen pairs, while also supporting water-mediated bridging between these atoms. Likewise, PHE1536

participates through protein oxygen–ligand nitrogen interactions, and TYR1540 through protein oxygen–ligand nitrogen contacts, both reinforced by water bridges. To simplify the presentation of our results of conventional MD, only the O-bridging patterns are discussed in the main text.

The RMSD profiles in Fig. 2C reveal that the protein backbone undergoes some fluctuations, stabilizing around 4–5 Å, mostly due to flexible winged helix domain that responses to the solvation. While the ligand remains highly stable (RMSD $\approx$ 1 Å), suggesting a well-defined binding mode within the pocket. The coordination number (CN) analysis (Fig. 2D) reveals a transient increase of approximately two O-bridging water molecules over time, reflecting the continuous exchange and dynamic reorganization of solvent molecules within the binding pocket. To extend this observation, we computed individual CNs for each water-mediated ligand–protein pair (see Fig. S1–S4 for N/O bridging patterns) and exhibited intermittent fluctuations indicative of rapid water rearrangements rather than static occupancy.

Providing a broader perspective, an overall CN analysis of water molecules within 4 Å of the ligand, together with the corresponding radial distribution function (RDF) comparison (Fig. 5A), captures the general hydration dynamics. The RDF between ligand and water oxygen atoms shows a pronounced first peak at $\sim$3.3 Å, consistent with strong hydrogen bonding, followed by a secondary rise that reflects the presence of the structured solvation shell around the semi-buried ligand. Together, these findings underscore the highly dynamic yet organized nature of water within the TAF1(2) binding pocket and align with previous observations by Aldeghi et al.[94], who demonstrated that structured water molecules in bromodomain binding sites can be displaced depending on the bound ligand while still contributing to complex stability and adaptability.

Previous studies suggested that water displacement in bromodomain complexes may be ligand-induced, with ligand binding and conformational dynamics perturbing the local pocket and triggering reorganization of hydration networks.[9,66,94] Motivated by these findings, we sought to investigate whether a similar mechanism operates in the TAF1(2)–butenyl-pyrrolo-pyridinone-dimethyl benzamide complex. Specifically, we focused on the two crystallographic water sites, W1 and W2, which are positioned to mediate potential protein-ligand interactions.

As a preliminary assessment, we examined whether global changes in the ligand dipole moment correlated with the CN of bridging water molecules. No meaningful correlation was observed (Fig. S5, SI), suggesting that overall hydration changes around the ligand are not directly coupled to its global polarization. This lack of a global signal motivated a more site-specific analysis of hydration dynamics at the W1 and W2 positions, where discrete water exchange events were evident in the trajectory (Fig. 2B, left).

For each site, we tracked oxygen atoms of all water molecules within 2 Å of the crystallographic positions of W1 and W2 throughout the simulation. The resulting occupancy profiles (Fig. 3B, D) revealed distinct behaviors: W1 was transiently occupied by water molecule ID 1831 for 90 ns, after which it remained unoccupied for the final 10 ns. By contrast, W2 remained continuously hydrated, but the identity of the occupying water molecules alternated, with some molecules exhibiting short residence times (e.g., IDs 40952 and 49019) and others persisting longer. These observations are in line with our CN evaluation of total water bridging molecules (Fig. 2E).

We then extracted instantaneous dipole moments of ligand oxygen atoms directly interacting with these waters and constructed binary occupancy signals (occupied = 1, vacant = 0) based on site-specific exchange events. To quantify the temporal coupling between ligand polarization and hydration dynamics, we computed lagged Pearson correlation coefficients over 4 ns windows. Positive correlation peaks at finite lag indicate that ligand dipole changes precede water exchange, whereas negative correlations imply that dipole fluctuations follow hydration events. This analysis revealed heterogeneous coupling across the two sites. At W2, modest positive correlations at short positive lags suggest that local ligand dipole fluctuations precede and may facilitate water exchange, consistent with ligand-driven perturbation of the hydration shell (Fig. 3D). In contrast, the W1 site exhibited a pronounced negative correlation centered near zero lag, indicating that ligand dipole changes respond to, rather than induce, hydration events (Fig. 3E).

Taken together, these results highlight the heterogeneous nature of hydration–ligand coupling within the TAF1(2) binding pocket: while the W2 site demonstrates features of ligand-induced water displacement, the W1 site behaves as a passive sensor of solvent fluctuations. These obser-

vations underscore the importance of site-specific analyses for understanding the dynamic interplay between ligands and structured water molecules in bromodomain recognition.

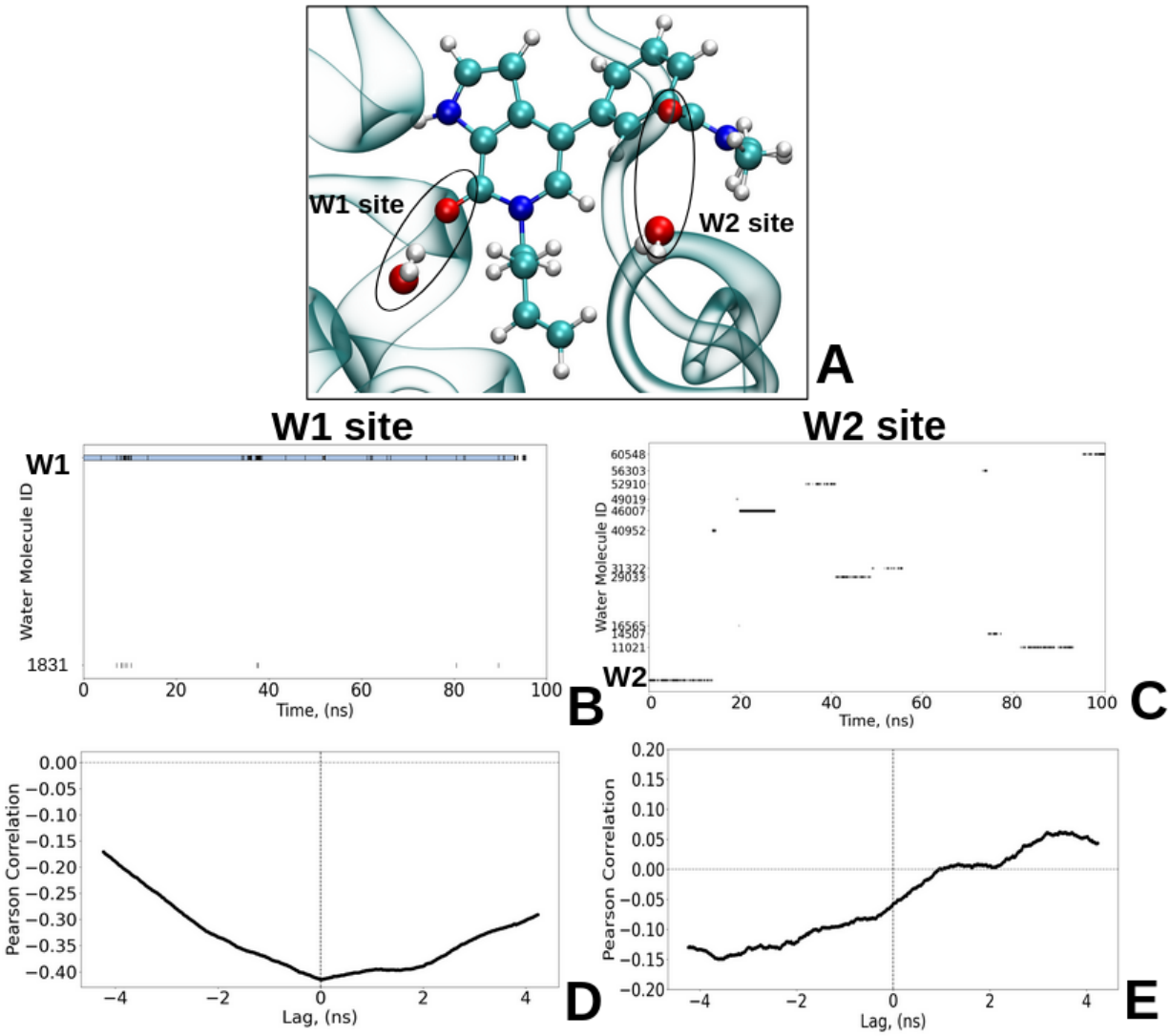

Figure 3. (A) Structural overview of the TAF1(2) binding pocket showing crystallographic water molecules W1 and W2 (red spheres) that bridge ligand oxygen atoms. Ligand–hydration coupling regions are circled in black. Time-resolved occupancy traces for W1 (B) and W2 (C), illustrating transient water replacement and exchange events. Lagged Pearson correlation between ligand oxygen-dipole fluctuations and water-site occupancy for W1 (D) and W2 (E), indicating site-specific coupling between ligand polarization and hydration dynamics.

Binding Affinity Calculation

To evaluate the binding affinity of the TAF1(2)–pyrrolopyridone compound, we determined the optimal set of protein and ligand atoms for the DBC restraint following the procedure described above. For this complex, the DBC_{max} value was 0.8 Å with the selected protein and ligand atoms used in the calculation shown in Fig. 4A.

The detailed contributions to the standard binding free energy are summarized in Table 2. Both van der Waals and electrostatic terms exhibit large magnitudes in the site and bulk calculations (see Fig.4 (B-E)), however, it is their differences that determine the net binding affinity. The van der Waals contribution favors binding by -12.5 kcal·mol^{-1} (i.e., $37.9 - 25.4$), whereas the electrostatic term provides an additional stabilization of -8.3 kcal·mol^{-1}. These results demonstrate that both van der Waals and electrostatic interactions contribute favorably to binding, with van der Waals interactions providing the dominant stabilization. After including the standard-state volume and the DBC restraint corrections, the resulting standard binding free energy is of -9.9 kcal·mol^{-1} that is in good agreement with experimental value (ΔG_{exp} $=-10.0$ kcal·mol^{-1}, Table 2)[66]. It should be noted that this level of accuracy was achieved within a relatively short simulation time required to reach the convergence, highlighting the efficiency of our approach (see Fig. S6-S8).

The critical role of dynamic binding pocket resolvation sampling in this complex was explicitly highlighted by Ge et al.[9], who noted that extensive ligand motion within the binding site can hinder the insertion or rearrangement of water molecules. Such behavior poses a particular challenge for binding affinity calculations, as it can lead to poor convergence and unreliable free-energy estimates. These observations underscore the importance of accurately modeling ligand-induced water displacement to achieve robust and reproducible results. To evaluate how effectively our protocol samples rehydration dynamics, we estimated generalized CNs of water molecules within 4 Å across all three replicas (Fig. 5B-D), integrating data from four walkers per replica and compared to the same function along the conventional MD (Fig. 5A). The resulting CNs were mapped to the alchemical coupling parameter, λ, allowing estimation of the mean and standard deviation of CN as a function of λ. Across both electrostatic and van der Waals transformations, the

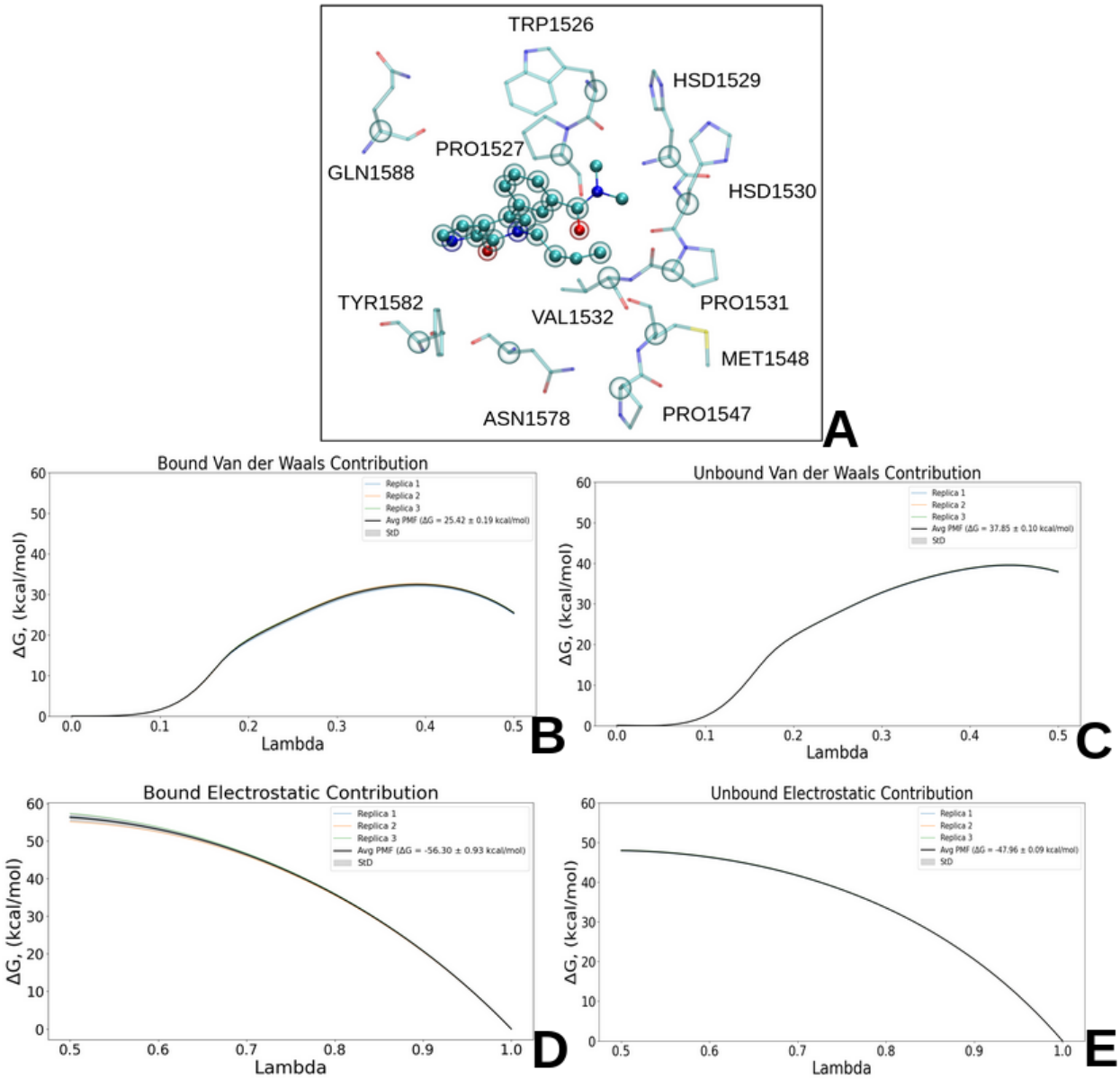

Figure 4. (A) Selection of atoms for the DBC CV shown in transparent VDW representation: C_α atoms of the binding site (anchor group) and selected ligand atoms used for binding affinity calculations. Hydrogen atoms are omitted for clarity. Plots of the PMF contributions for the bound van der Waals (B) and electrostatic (D) interactions, and unbound van der Waals (C) and electrostatic (E) interactions across three replicas of four walkers each.

global average CN remains approximately 22 water molecules, closely matching the values obtained from conventional simulations of the same complex. It should be noted, the van der Waals transformation reveals several additional bridging sites compared to electrostatic one that emerge primarily at low λ values (0.0–0.2), when electrostatic interactions are largely turned off and the

ligand–protein contacts begin to weaken. Notably, the highest solvent contribution (Fig. 5C) occurs within the same λ range, reflecting transient rehydration of the binding pocket as the ligand progressively decouples and solvent molecules occupy the vacated interaction space. Thereafter, the generalized CN converges to a plateau corresponding to the equilibrium hydration level of the bound state, underscoring the robustness of our protocol in capturing binding-site solvation across the full spectrum of interaction-scaling conditions within the Lambda-ABF-OPES framework.

Although the RDFs differ in absolute g(r) values, the characteristic peak at $\sim$3.3 Å (Fig. 5A) is consistently reproduced within the electrostatic component (Fig. 5B). This behavior arises because, in the λ-range of 0.5–1.0 the electrostatic interactions are scaled, yet the van der Waals forces remain fully active, preserving the packing effects that stabilize water molecules near the binding site. Consequently, the hydration shell persists, albeit with reduced intensity. In contrast, during the van der Waals transformation, where dispersion and repulsion terms are scaled from λ = 0.0 (no van der Waals interactions) to 0.5 (fully active), the RDFs lose their pronounced first-shell structure, consistent with weakened packing and a more diffuse solvation environment.

To assess the effectiveness of the rehydration process at recovering key bridging interactions, we analyzed the individual coordination numbers (CNs) for each O-bridging pattern (Fig. S9–S10). Both bound-state transformations cover the essential set of the protein bridging sites with the highest densities (i.e., TYR1540, ASN1533, PRO1531, and HSD1530), previously observed in conventional MD simulations (Fig. 2B). Such close correspondence demonstrates that the rehydration protocol effectively restores the characteristic network of bridging waters observed under equilibrium conditions. Moreover, N/O-bridging motifs follow the expected trends for both transformations (Fig. S10–S16), yet bridging between a protein nitrogen and a ligand oxygen remains flat at low λ values (0.0–0.2) during the van der Waals transformation (Fig. S11). This behavior is consistent with the corresponding water-bridged density distribution (Fig. S2-S4), indicating that such bridging is less geometrically favorable, therefore, contributes little to the hydration sampling of the bridged sites.

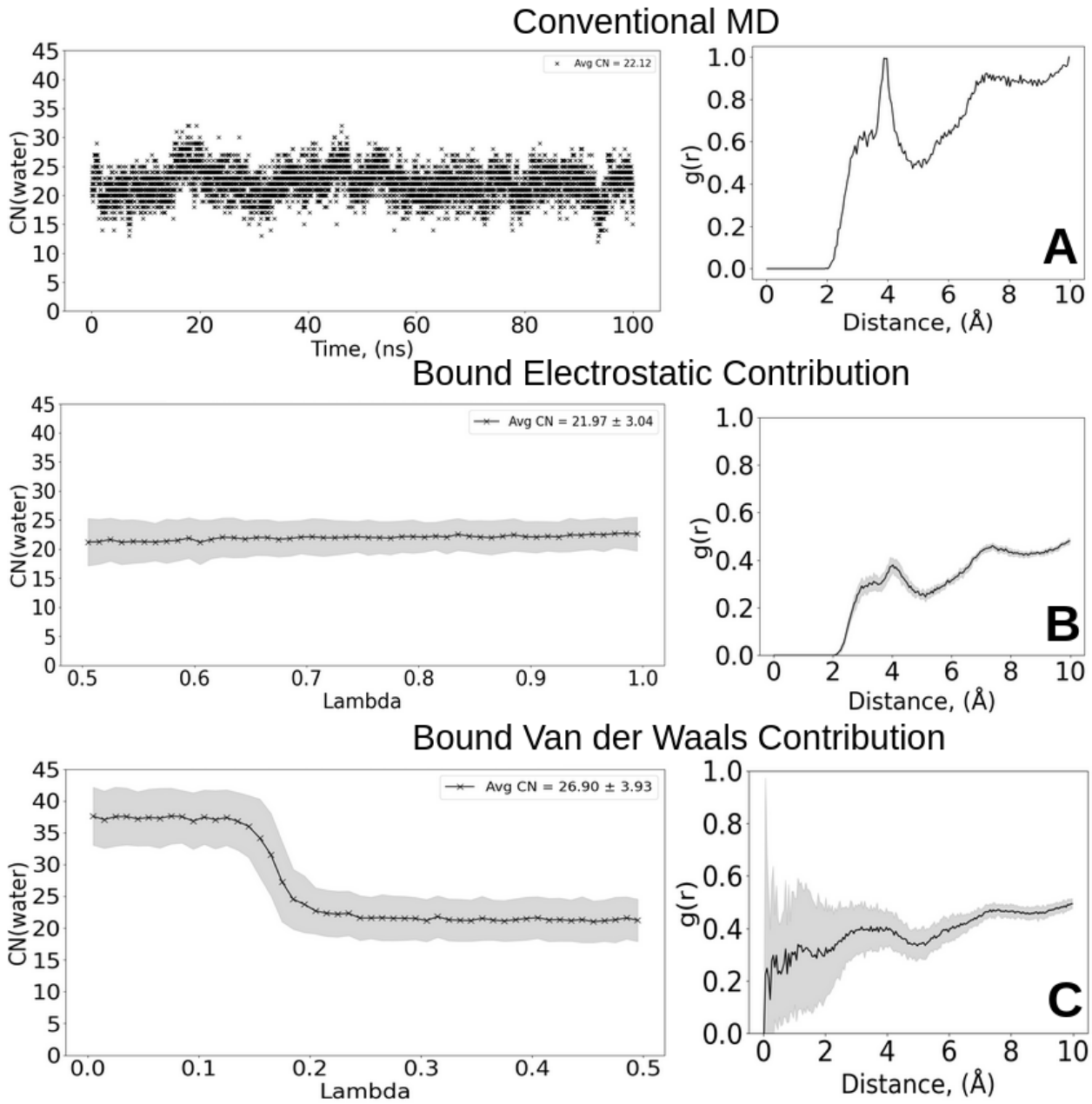

Figure 5. Generalized coordination number (CN) of water molecules within 4 Å of the ligand and corresponding radial distribution functions (RDFs) of water and ligand oxygen atoms. Panels show (A) conventional MD, (B) electrostatic, and (C) van der Waals binding-affinity contributions. (B, C) CNs are averaged over all walkers and replicas (shaded areas indicate standard deviation), collected at $\delta\lambda = 0.01$. RDFs illustrate changes in the ligand solvation shell along the respective λ-scaling transformations.

1I05: Buried Ligand Binding

Conventional MD

Major urinary protein I (MUP-I) is a member of the lipocalin family, abundantly present in male mouse urine, where it functions as a pheromone-binding protein.[65,95,96] Structurally, MUP-I adopts the canonical lipocalin fold: an eight-stranded antiparallel β-barrel capped by a short α-helix, enclosing a deeply buried hydrophobic cavity (Fig. 6(A, B)). Its natural ligand, 6-hydroxy-6-methyl-3-heptanone, is a small ketone bearing a polar hydroxyl group central to pheromone signaling.[65,97] Isothermal titration calorimetry (ITC) measurements reported by Timm et al.[65] indicate that ligand binding is predominantly enthalpy-driven ($\Delta H = -13$ kcal·mol^{-1}), consistent with the extensive noncovalent interactions observed in crystallography, but partially offset by an entropic penalty ($-T\Delta S = +7$ kcal·mol^{-1}).

To assess the conformational stability of the MUP-I–ligand complex, we performed MD-based structural analysis. The Cα RMSD plateaued at $\sim$3.0 Å, indicating that the lipocalin fold remained structurally stable (Fig. 6C). In contrast, the ligand heavy-atom RMSD fluctuated between 0.5–1.5 Å, reflecting local conformational flexibility within the binding pocket, as the ligand remained bound while exploring several distinct internal rotational degrees of freedom. To determine which internal motions are permitted upon binding and whether the protein selectively stabilizes particular ligand conformers, we performed a torsional analysis of the ligand backbone and the hydroxyl–ketone dihedral. Backbone dihedrals predominantly adapted trans conformations ($\sim$80%, Fig. 6D), whereas the hydroxyl–ketone dihedral favored mostly a gauche$-$ conformation ($\sim$87%), with minor population of gauche$+$ ($\sim$10%), and trans ($\sim$3%) (Fig. 6F).

Compared to the TAF1(2)-related complex (i.e., 5I1Q) discussed above, where continuous water displacement was observed during the simulation, the MUP-I binding site exhibited markedly different behavior. Here, the ligand is deeply buried within the cavity, and the two water molecules identified in the crystal structure remained stably coordinated throughout the conventional MD trajectory. To quantify this persistent hydration, we applied a generalized CN analysis by evaluating

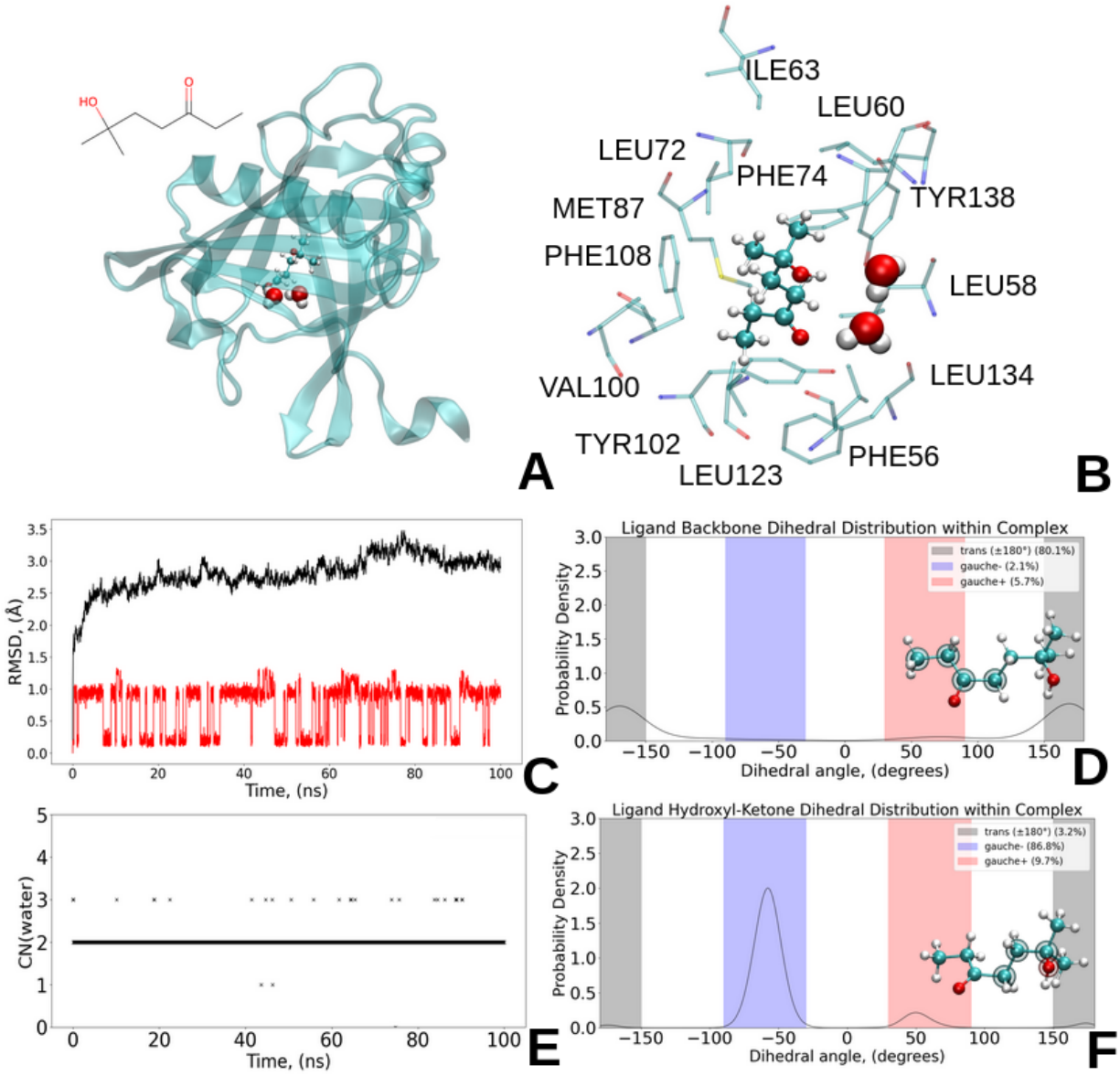

Figure 6. (A) Structure of MUP-I in complex with 6-hydroxy-6-methyl-3-heptanone. (B) Representative binding pose within the MUP-I lipocalin β-barrel cavity. (C) RMSD of protein Cα atoms (black) and ligand heavy atoms (red) over 100 ns. (E) Coordination number (CN) of water molecules interacting with the ligand at distances less than 4 Å from the ligand, showing an average of two bound waters. (D) Probability distribution of ligand backbone dihedral angles. (F) Probability distribution of the ligand hydroxyl–ketone dihedral, showing populations of trans, gauche+, and gauche− conformations. Atoms analyzed are shown in transparent van der Waals representation.

water occupancy within 4 Å of the ligand (Fig. 6E). The CN profile confirmed stable interactions involving an average of two water molecules, in agreement with crystallographic observations.

These water molecules, together with surrounding residues, define the key stabilizing interactions in the complex, where the ligand is primarily anchored via van der Waals contacts with largely nonpolar cavity residues (see Fig. S17–S18 for decomposed hydrogen-bond density functions)[65]. Polar contributions arise from TYR102, which caps the entrance and shields the binding site from solvent, and from the two buried water molecules, which form O-bridged hydrogen bonds with TYR138 and the backbone carbonyls of PHE56 and LEU58, completing the interaction network (Fig. 6B).

To probe the dynamics of these O-bridged network, we analyzed dipole–dipole correlations among the oxygen atoms of the two buried waters (W1 and W2) and oxygen atoms of their interacting partners (Fig. 7A-F). Pearson correlation coefficients were used to quantify the degree of coordinated dipole fluctuations, with values close to 1 indicating highly correlated motion and stable hydrogen bonding. W1 showed the strongest correlations, particularly with TYR138 ($r \approx 0.99$) and LEU58 ($r \approx 0.93$), while both residues were less correlated with W2 ($r \approx 0.90$) (Fig. 7C,E). Although PHE56 forms a direct hydrogen bond, its correlation with W2 remained modest Fig. 7F). The ligand oxygen atoms exhibited cooperative correlations with both water molecules ($r \approx 0.96$) Fig. 7D), and the two water oxygen atoms were themselves strongly correlated ($r \approx 0.94$), together forming a tightly coordinated hydrogen-bonding network Fig. 7B).

Notably, although W2 engages less strongly with cavity residues, its strong correlation with W1 suggests that the stabilization of W1 may promote the persistence of W2 within the binding site. To further characterize the nature of this coupling, we examined the dipole-moment distributions of W1, W2, and bulk water (Fig. S17) and revealed that W1 displayed the highest polarization ($\mu = 2.8$ D, $\sigma = 0.2$), slightly exceeding that of bulk water[59] ($\mu = 2.7$ D, $\sigma = 0.2$), whereas W2 was less polarized ($\mu = 2.5$ D, $\sigma = 0.2$). The enhanced polarization and strong residue correlations of W1 indicate that it experiences substantial electrostatic stabilization from the protein-ligand environment. By contrast, W2 low dipole moment suggests weaker direct electrostatic support. Therefore, its retention in the cavity more likely arises from geometric confinement and cooperative coupling to W1 and the ligand, rather than strong, individual stabilization by surrounding residues.

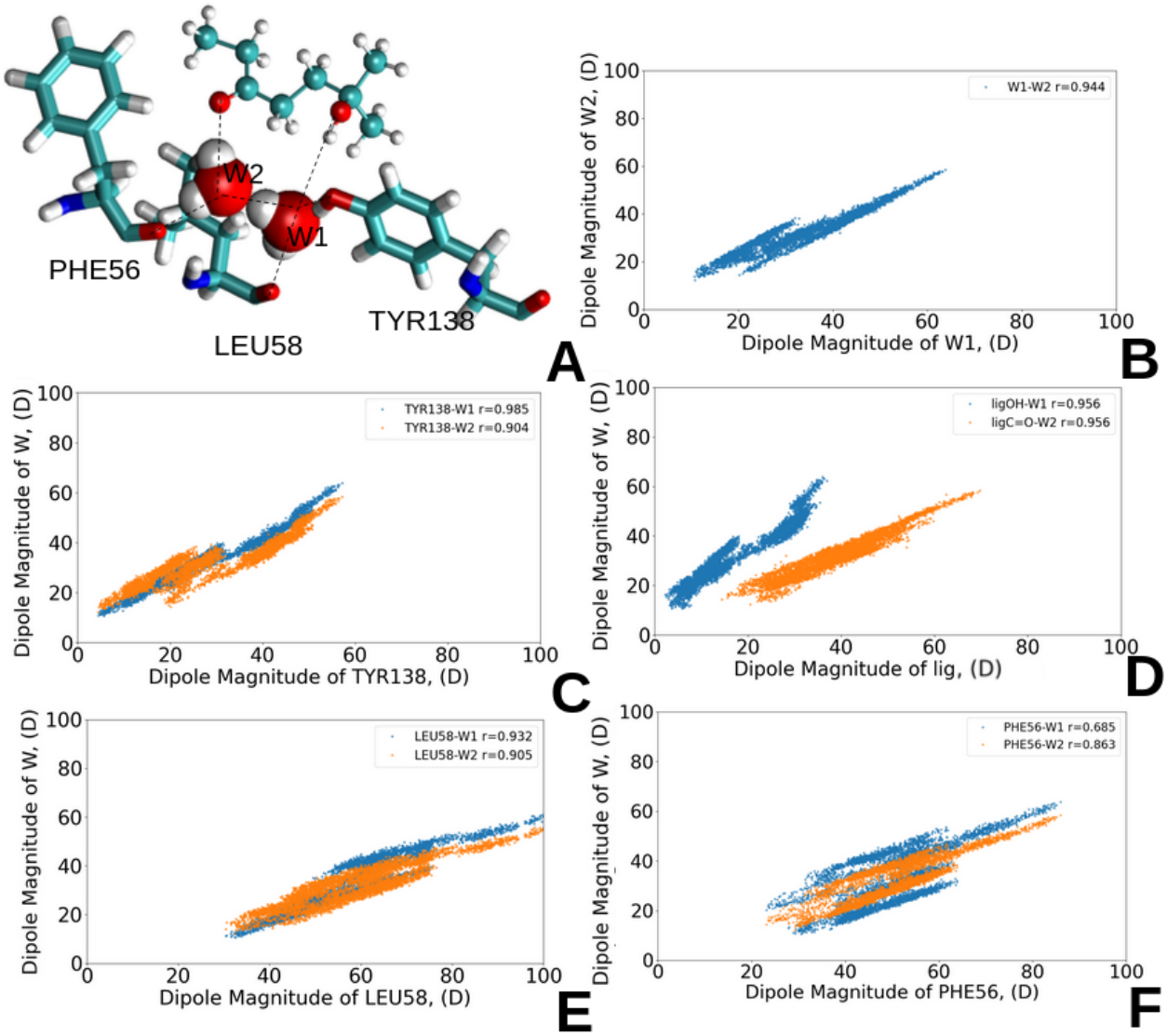

Figure 7. (A) Schematic of the water-mediated O-bridging hydrogen-bond network. Correlation of dipole magnitudes between individual oxygen atoms of buried water molecules (W1, W2) (B), oxygen atom of TYR138 with W1 and W2 (C), ligand hydroxyl and carbonyl oxygen atomss (ligOH and ligC=O, respectively) with W1 and W2 (D), oxygen atom of LEU58 with W1 and W2 (E), and oxygen atom of PHE56 with W1 and W2 (F). Pearson correlation coefficients,r, indicate the strength of dynamic coupling of dipole fluctuations.

Binding Affinity Calculation

The studied 1I05 complex exemplifies the challenges common to recognition complexes: despite relatively low binding affinity (Table 2), it undergoes substantial conformational rearrangements, where flexibility tends to reduce affinity but enhance specificity, making adequate sampling of

transition states particularly demanding [98–100]. Previous investigations by Fu et al.[36] highlighted these difficulties, showing that incomplete exploration of orientational states during the alchemical annihilation stage prevented convergence of free energy estimates. To mitigate this, slow-growth strategy [101] with carefully tuned restraints were employed to stabilize the binding mode, yet additional complications arose from two buried crystallographic water molecules. Maintaining the experimentally observed hydrogen-bonding network required artificial restraints on these molecules during alchemical transformations, a strategy that, while effective, remains ad hoc and complex-specific, limiting its broader applicability.

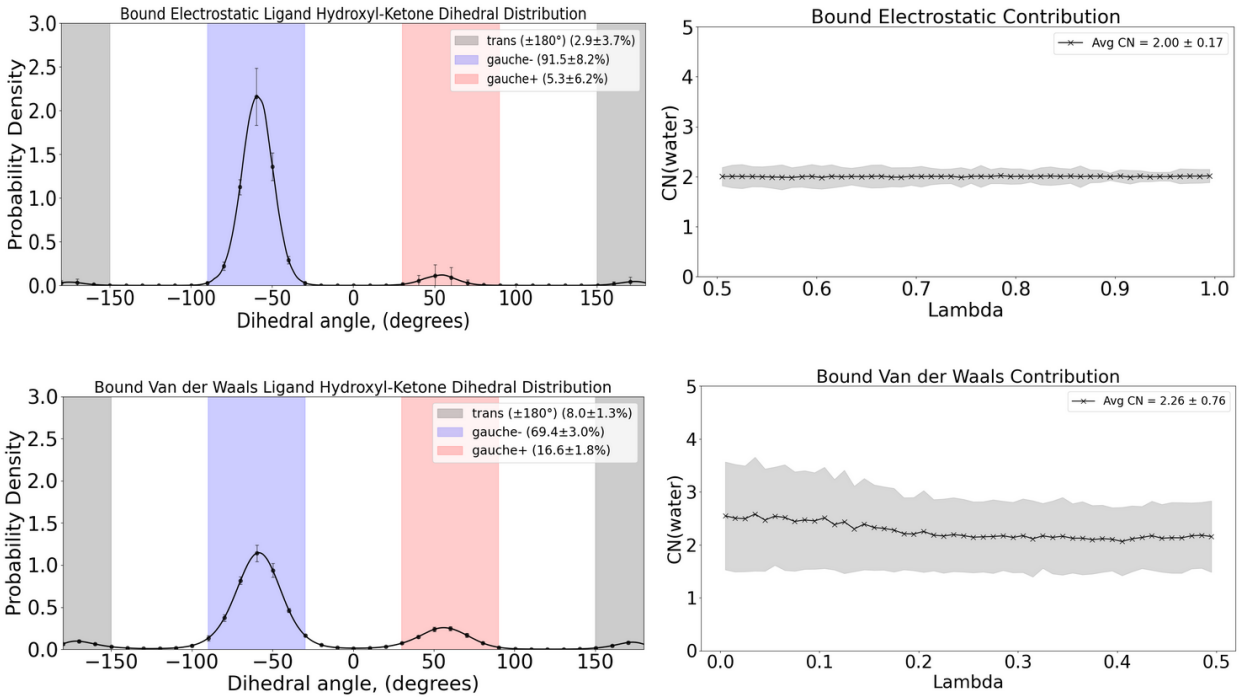

Figure 8. Probability distribution of the ligand hydroxyl–ketone dihedral with population of trans, gauche+, and gauche− conformations (left). Generalized coordination number (CN) of bridging water molecules within 4 Å of the ligand, shown across their respective λ-ranges collected at $\delta\lambda$ of 0.01 and averaged over all walkers (right). Shaded gray areas indicate the standard deviation. The data collected for the bound electrostatic and van der Waals binding-affinity contributions.

In the present study, the Lambda-ABF-OPES framework addresses both of these challenges without requiring complex-specific modifications. By enhancing sampling along DBC coordinates during alchemical decoupling, the approach allows the ligand to surmount internal rotational barriers that would otherwise confine it to metastable basins. Analysis of the dihedral probability dis-

tribution substantiates this point: while backbone torsions remain centered on the trans conformer under both electrostatic and van der Waals decoupling (Fig. S20-S21), the ketone–hydroxyl torsion displays distinct behavior (Fig. 8). Under van der Waals decoupling, the trans state is further stabilized and both gauche− and gauche+ conformers are populated, whereas under electrostatic decoupling the distribution mirrors equilibrium sampling (Fig. 6), with gauche− predominating. Additionally, the hydration environment of the binding pocket is likewise captured without artificial restraints. Notably, the exchange and rehydration of buried waters occur spontaneously, with average CNs of 2.3 ± 0.8 (van der Waals) and 2.0 ± 0.2 (electrostatics) (Fig. 8).

The resulting binding free energy ($\Delta G^{\circ}_{\mathrm{bind}} = -6.3 \pm 0.0$ kcal·mol^{-1}) is in close agreement with experimental affinity data ($\Delta G_{\mathrm{exp}} = -6.0$ kcal·mol^{-1}, Table 2), underscoring both the accuracy and robustness of Lambda-ABF-OPES for flexible, buried ligands such as 6-hydroxy-6-methyl-3-heptanone in MUP-I. More broadly, this case study illustrates the framework ability to extend alchemical free energy calculations to other complexes where ligand flexibility and buried hydration have previously hindered convergence.

Relative to prior BFEE2 simulations on the same complex[36], which required approximately 300 ns of total sampling and employed positional restraints to treat buried waters as part of the protein, the present Lambda-ABF-OPES calculations reached convergence within accumulated simulation time of only 42 ns (i.e., 22ns in the condensed phase and 20 ns in gas, Table 2). In contrast to the earlier BFEE2 protocol, water exchange and hydrogen-bond reformation occurred spontaneously during the simulation, eliminating the need for artificial restraints. This outcome demonstrates not only the superior efficiency of Lambda-ABF-OPES but also its broader applicability to complexes in which conventional alchemical methods struggle to reconcile ligand flexibility with pocket stability.

Remaining Complexes

5I40

In addition to the TAF1(2)–ligand complex (PDB id 5I1Q), we examined the BRD9–ligand complex (PDB id 5I40)[66] to explore differences in water-mediated binding environments and ligand specificity between these bromodomain subtypes. Unlike TAF1(2), BRD9 exhibits variations in binding-cavity size, shape, and flexibility, as well as amino acid substitutions that influence ligand access and hydration networks. The ligand, 6-methyl-1,6-dihydro-7H-pyrrolo[2,3-c]pyridin-7-one, acts as an acetyl-lysine mimetic, forming two-point hydrogen bonds with ASN100 via the pyridone carbonyl and pyrrole nitrogen, while its aromatic core engages in π–π stacking with the TYR106 gatekeeper.[66]

Throughout the 100-ns conventional MD simulation, the overall geometry of the protein–ligand complex was preserved. In contrast, the hydration sites, including those bridging ASN100 and TYR106, exhibited pronounced dynamism. Density distribution plots reveal a broad spatial distribution of bridging water positions (Fig. S22–S24), indicating that the waters are inherently mobile rather than localized. A conserved water molecule has also been reported to participate in a tetra-coordinated hydrogen-bonding network involving ASN95, TYR57, and MET65 based on static structural analysis.[102] However, while the crystallographic configuration highlights potential hydration motifs, it does not necessarily reflect their persistence under dynamic conditions and was not maintained in our conventional MD simulations.

Regarding binding-affinity calculations, the corresponding PMFs were fully converged (Fig. S25–S27), and the total CN of water molecules within 4 Å of the ligand agrees well with the CN profiles obtained from van der Waals and electrostatic alchemical transformations (Fig. S28), further supporting active solvent exchange and a dynamically adaptive hydration shell. Such dynamic rehydration is effectively captured by our binding-affinity calculations, highlighting the ability of the Lambda-ABF-OPES framework to model ligand-induced water rearrangements in flexible binding environments.

Hsp90-related complexes

The N-terminal domain of Hsp90 is an ATP-binding chaperone module essential for the maturation and stability of many oncogenic proteins. Its binding pocket is characterized by a polar subsite flanked by a hydrophobic cavity. Two studied HSP90–ligand complexes, 2XAB and 2XJG, ligands of which differ by a single chemical modification at the capping group of the resorcinol pharmacophore: the hydroxyl group in the lead compound (2XAB) is replaced with a methoxy-substituent in 2XJG.[67] Despite this minimal change, the crystallographic data underlying the two complexes slightly differ in resolution, which might impact model construction and subsequent simulation. 2XAB was refined at 1.9 Å, enabling a slightly higher confidence in sidechain orientations and ordered solvent placement compared to 2XJG (e.g., 2.25 Å, Table 1).

Initially, unresolved histidine-rich segments in disordered regions and in the binding site were built for both 2XAB and 2XJG and assigned as neutral δ-tautomers (HSD), consistent with their local environment and potential hydrogen-bond network. The simulations using these full models (i.e., with disordered regions) revealed that inclusion of the disordered segments had very different effects in the two complexes. For 2XAB, the loops had a minor impact on the computed binding free energy providing a good agreement with the experimental data ($\Delta G^{\circ}_{\mathrm{bind}} = -12.0 \pm 0.4$ kcal·mol^{-1} vs. $\Delta G_{\mathrm{exp}} = -12.6$ kcal·mol^{-1}, Table 2), indicating that the full model is stable and can be used for accurate binding affinity evaluation. In contrast, for 2XJG, the disordered loops were highly mobile and introduced large-amplitude fluctuations that dominated the conformational ensemble, artificially destabilizing the bound state by $\sim$16 kcal·mol^{-1} (Table S1). This high intrinsic mobility makes the full 2XJG model unsuitable for reliable binding free energy calculations. Based on these observations, truncated models (omitting the disordered segments) were then generated for both complexes and tested. Here, histidine protonation within the binding site was also carefully assigned, as misassignment can alter local electrostatic potentials, reorient ordered waters, and change hydrogen-bond networks, routinely affecting computed ΔG by several $k_B T$. Protonation states were determined by visual inspection of potential inter- and intramolecular hydrogen-bonding networks: two neutral ϵ-tautomers HSE77, HSE189, HSE210, and one neutral

δ-tautomer HSD154. The results showed that the truncated 2XJG model restored stability and gave physically meaningful binding energetics ($\Delta G^{\circ}_{\mathrm{bind}} = -12.2 \pm 0.1$ kcal·mol^{-1} vs. $\Delta G_{\mathrm{exp}} = -10.7$ kcal·mol^{-1}, Table 2), while for 2XAB, the truncation had a minor impact on the computed binding free energy ($\sim$1.8 kcal·mol^{-1} less favorable), meaning that the full model including the disordered loops remains preferable for 2XAB to accurately capture the binding environment. All subsequent analyses were therefore performed using the truncated model for 2XJG and the full model for 2XAB, ensuring accurate representation of binding energetics without artefactual penalties.

To further characterize binding interactions, we analyzed density distributions of both direct protein–ligand hydrogen bonds and water-mediated interactions (Fig. S29–S32 for 2XAB and Fig. S37–S40 for 2XJG). For 2XAB, the distributions are generally more pronounced and numerous than in 2XJG, reflecting the solvent accessibility of the ligand exposed hydroxyl group. Several direct interaction patterns are shared between the two complexes, including contacts with PHE138 and MET98, and many direct hydrogen bonds are additionally stabilized by bridging solvent molecules. In both complexes, extensive displacement of pocket waters was observed, highlighting the dynamic nature of solvent participation and its contribution to the binding landscape. This observation aligns with Ge et al.[9], who reported that crystallographic waters in these complexes are highly dynamic and require extensive sampling to converge. Specifically, they found that the 2XJG complex is particularly difficult to equilibrate, requiring approximately 145 million force evaluations in conventional MD simulations. Their study also employed BLUES[103], which combines nonequilibrium candidate Monte Carlo (NCMC) with classical MD, and grand canonical Monte Carlo (GCMC)[104–107], which allows insertion and deletion of solvent molecules during equilibration. Yet, even with these strategies, convergence of the local solvent network still required millions of force evaluations, highlighting the slow-exchanging sites and labile environment in 2XJG, whose behavior resists to straightforward explanation.

We carefully evaluated solvent effects in both complexes, and rehydration was effectively captured, consistent with the calculated binding affinities. Comparative plots of the coordination number of water molecules within 4 Å of the ligand during conventional MD, as well as as a function

of the alchemical coupling parameter λ, are provided (Fig. S33–S34 for 2XAB and Fig. S41–S42 for 2XJG). Notably, in the bound state, the electrostatic and van der Waals contributions closely resemble the coordination number observed in standard MD, further supporting the robustness of our methodology. Convergence of the obtained PMFs is shown in Fig. S35–S36 and Fig. S43–S44 for 2XAB and 2XJG, respectively.

Conclusion and Perspectives

The present study demonstrates that resolving the dynamic role of water is both essential and achievable in binding affinity calculations. By combining the polarizable AMOEBA force field with the Lambda-ABF-OPES framework, which unifies enhanced sampling with efficient restraint schemes, our approach evaluates binding affinities towards chemical accuracy in complexes where water molecules either remain stably bound or dynamically exchange within the binding site. In this context, water-bridged complexes can be classified into two general categories: those with persistent water molecules and those characterized by rapid water exchange. In both scenarios, Lambda-ABF-OPES captures the relevant hydration dynamics and provides reliable estimates of protein-ligand binding affinities.

Our methodology proved effective across both buried and semi-buried water-mediated complexes. Importantly, Lambda-ABF-OPES does not rely on ad hoc restraints or complex-specific modifications, offering a generalizable framework for quantitatively reliable modeling of biomolecular recognition processes in which solvent dynamics play a decisive role. Moreover, despite employing fully polarizable models, the total simulation time required for convergence was only on the order of tens of nanoseconds per complex, ensuring both good agreement with experimental data and computational efficiency for rigorous binding affinity evaluation. These features also open the possibility of coupling the present framework with fast neural network foundation models[108–110] to further accelerate accurate free energy predictions.

Being MD-based, our approach also provides mechanistic insights into binding, revealing how

both direct and water-mediated interactions stabilize ligands within their binding pockets. This capability enables a deeper understanding of the physical principles governing molecular recognition and energetics. Nonetheless, the accuracy of such calculations remains inherently sensitive to the quality of the underlying structural models. Incompletely resolved or highly flexible regions, particularly disordered loops can introduce artificial fluctuations and bias free-energy estimates if not treated carefully. For such systems, alternative structural hypotheses or ensemble-based modeling should be considered before quantitative analysis. Additionally, highly flexible or surface-exposed ligands, such as peptides or other proteins, may still pose challenges due to extensive solvent exposure during alchemical transformations. Addressing such cases will be the focus of future works. Integrating the presented framework into high-throughput pipelines could further accelerate drug discovery, where accurate and efficient ranking of candidate ligands is critical.

Acknowledgments

This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 810367), project EMC2 (JPP). We thank the Grand Équipement de Calcul Intensif (GENCI), Institut du Développement et des Ressources en Informatique (IDRIS), and Centre Informatique de l'Enseignement Supérieur (CINES), France, for their support in term of computational ressources through grant no. A0180716167 and A0150712052.

Competing Interests

LL and JPP are shareholders and co-founders of Qubit Pharmaceuticals. The remaining authors declare no other competing interests.

References

(1) Ball, P. Water as an active constituent in cell biology. *Chem. Rev.* **2008**, *108*, 74–108.

(2) Rudling, A.; Orro, A.; Carlsson, J. Prediction of Ordered Water Molecules in Protein Binding Sites from Molecular Dynamics Simulations: The Impact of Ligand Binding on Hydration Networks. *J. Chem. Inf. Model.* **2018**, *58*, 350–361.

(3) Maurer, M.; Oostenbrink, C. Water in protein hydration and ligand recognition. *J. Mol. Recognit.* **2019**, *32*, e2810.

(4) Tari, L. W. *Methods Mol. Biol.*; 2012; Vol. 819; pp 259–270.

(5) Lu, Y.; Wang, R.; Yang, C. Y.; Wang, S. Analysis of ligand-bound water molecules in high-resolution crystal structures of protein-ligand complexes. *J. Chem. Inf. Model.* **2007**, *47*, 668–675.

(6) de Beer, S. B. A.; Vermeulen, N. P. E.; Oostenbrink, C. The role of water molecules in computational drug design. *Curr. Top. Med. Chem.* **2010**, *10*, 257–270.

(7) Li, Z.; Lazaridis, T. Water at biomolecular binding interfaces. *Phys. Chem. Chem. Phys.* **2007**, *9*, 573–581.

(8) 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, PMID: 31518131.

(9) Ge, Y.; Wych, D. C.; Samways, M. L.; Wall, M. E.; Essex, J. W.; Mobley, D. L. Enhancing Sampling of Water Rehydration on Ligand Binding: A Comparison of Techniques. *J. Chem. Theory Comput.* **2022**, *18*, 1359–1381.

(10) Setiadi, J.; Biedermann, F.; Nau, W. M.; Gilson, M. K. Thermodynamics of Water Displacement from Binding Sites and its Contributions to Supramolecular and Biomolecular Affinity. *Angew. Chem. Int. Ed.* **2025**, *64*, e202505713.

(11) Bodnarchuk, M. S.; Viner, R.; Michel, J.; Essex, J. W. Strategies to Calculate Water Binding Free Energies in Protein–Ligand Complexes. *J. Chem. Inf. Model.* **2014**, *54*, 1623–1633.

(12) Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. Prediction of the Water Content in Protein Binding Sites. *J. Phys. Chem. B* **2009**, *113*, 13337–13346.

(13) Ge, Y.; Melling, O. J.; Dong, W.; Essex, J. W.; Mobley, D. L. Enhancing sampling of water rehydration upon ligand binding using variants of grand canonical Monte Carlo. *J. Comput. Aided Mol. Des.* **2022**, *36*, 767–779.

(14) Deng, J.; Cui, Q. Efficient Sampling of Cavity Hydration in Proteins with Nonequilibrium Grand Canonical Monte Carlo and Polarizable Force Fields. *J. Chem. Theory Comput.* **2024**, *20*, 1897–1911.

(15) Srinivasan, J.; Cheatham, T. E.; Cieplak, P.; Kollman, P. A.; Case, D. A. Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate-DNA Helices. *J. Am. Chem. Soc.* **1998**, *120*, 9401–9409.

(16) Gohlke H., C. D. Converging Free Energy Estimates: MM-PB(GB)SA Studies on the Protein-Protein Complex Ras-Raf. *J Comput. Chem.* **2004**, *25(2)*, 238–50.

(17) Massova, I.; Kollman, P. A. Combined Molecular Mechanical and Continuum Solvent Approach (MM-PBSA/GBSA) to Predict Ligand Binding. *Perspect. drug discov. des.* **2000**, *18*, 113–135.

(18) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000**, *33*, 889–897.

(19) Homeyer, N.; Gohlke, H. Free Energy Calculations by the Molecular Mechanics Poisson-Boltzmann Surface Area Method. *Mol. Inform.* **2012**, *31*, 114–22.

(20) Baron, R.; Setny, P.; McCammon, J. A. Water in Cavity-Ligand Recognition. *J. Am. Chem. Soc.* **2010**, *132*, 12091–12097.

(21) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities. *Expert Opin. Drug Discov.* **2015**, *10*, 449–461.

(22) Swanson, J. M. J.; Henchman, R. H.; McCammon, J. A. Revisiting Free Energy Calculations: a Theoretical Connection to MM/PBSA and Direct Calculation of the Association Free Energy. *Biophys. J.* **2004**, *86*, 67–74.

(23) Gumbart, J. C.; Roux, B.; Chipot, C. Standard Binding Free Energies from Computer Simulations: What Is the Best Strategy? *J. Chem. Theory Comput.* **2013**, *9*, 794–802.

(24) Blazhynska, M.; Goulard Coderc de Lacam, E.; Chen, H.; Roux, B.; Chipot, C. Hazardous shortcuts in standard binding-free energy calculations. *J. Phys. Chem. Lett.* **2022**, *13*, 6250–6258.

(25) Chipot, C. Free Energy Methods for the Description of Molecular Processes. *Annu. Rev. Biophys.* **2023**, *52*, 113–138.

(26) Fu, H.; Chen, H.; Blazhynska, M.; Goulard Coderc de Lacam, E.; Szczepaniak, F.; Pavlova, A.; Shao, X.; Gumbart, J. C.; Dehez, F.; Roux, B.; Cai, W.; Chipot, C. Accurate Determination of Protein:Ligand Standard Binding Free Energies from Molecular Dynamics Simulations. *Nat. Protoc.* **2022**, *17*, 1114–1141.

(27) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.

(28) Limongelli, V.; Bonomi, M.; Parrinello, M. Funnel Metadynamics as Accurate Binding Free-Energy Method. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 6358–6363.

(29) Raniolo, S.; Limongelli, V. Ligand Binding Free-Energy Calculations with Funnel Metadynamics. *Nat. Protoc.* **2020**, *15*, 2837–2866.

(30) Zwanzig, R. W. High-Temperature Expansion for the Free Energy of a System Interacting with a Heat Bath. *J. Chem. Phys.* **1954**, *22*, 1420–1426.

(31) Landau, L. D.; Lifshitz, E. M. *Course of Theoretical Physics, Vol. 5: Statistical Physics*; Pergamon Press: London, 1980; Translated from the Russian by Peierls, E. and Peierls, R. F., pp. 893–895.

(32) Mobley, D. L.; Chodera, J. D.; Dill, K. A. Confine-and-release method: obtaining correct binding free energies in the presence of protein conformational change. *J. Chem. Theory Comput.* **2007**, *3*, 1231–1235.

(33) Ross, G. A.; Lu, C.; Scarabelli, G.; Albanese, S. K.; Houang, E.; Abel, R.; Harder, E. D.; Wang, L. The maximal and current accuracy of rigorous protein-ligand binding free energy calculations. *Commun. Chem.* **2023**, *6*, 222.

(34) Ebrahimi, M.; Hénin, J. Symmetry-adapted restraints for binding free energy calculations. *J. Chem. Theory Comput.* **2022**, *18*, 2494–2502.

(35) Fu, H.; Chen, H.; Cai, W.; Shao, X.; Chipot, C. BFEE2: automated, streamlined, and accurate absolute binding free-energy calculations. *J. Chem. Inf. Model.* **2021**, *61*, 2116–2123.

(36) Fu, H.; et al. Accurate determination of protein:Ligand standard binding free energies from molecular dynamics simulations. *Nat. Protoc.* **2022**, *17*, 1114–1141.

(37) Liu, R.; et al. Accelerating and automating the free energy perturbation absolute binding free energy calculation with the RED-E function. *J. Chem. Inf. Model.* **2023**, *63*, 7755–7767.

(38) Bian, H.; Shao, X.; Chipot, C.; Cai, W.; Fu, H. A formally exact method for high-throughput absolute binding-free-energy calculations. *Nat. Comput. Sci.* **2025**, *5*, 621–626.

(39) Wahl, J.; Smieško, M. Assessing the Predictive Power of Relative Binding Free Energy Calculations for Test Cases Involving Displacement of Binding Site Water Molecules. *J. Chem. Inf. Model.* **2019**, *59*, 754–765.

(40) Jiang, W.; Hodoscek, M.; Roux, B. Computation of absolute hydration and binding free energy with free energy perturbation distributed replica-exchange molecular dynamics. *J. Chem. Theory Comput.* **2009**, *5*, 2583–2588.

(41) Rizzi, V.; Bonati, L.; Ansari, N.; Parrinello, M. The role of water in host-guest interaction. *Nat. Commun.* **2021**, *12*, 93.

(42) Ansari, N.; Rizzi, V.; Parrinello, M. Water regulates the residence time of Benzamidine in Trypsin. *Nat. Commun.* **2022**, *13*, 5438.

(43) Schulze, M.; Khakhula, T.; Piasentin, N.; Aureli, S.; Rizzi, V.; Gervasio, F. L. All You Need Is Water: Converging Ligand Binding Simulations with Hydration Collective Variables. *bioRxiv* **2025**, Preprint.

(44) Lagardère, L.; Maurin, L.; Adjoua, O.; El Hage, K.; Monmarché, P.; Piquemal, J.-P.; Hénin, J. Lambda-ABF: Simplified, portable, accurate, and cost-effective alchemical free-energy computation. *J. Chem. Theory Comput.* **2024**, *20*, 4481–4498.

(45) Ansari, N.; Jing, Z. F.; Gagelin, A.; Hédin, F.; Aviat, F.; Hénin, J.; Piquemal, J.-P.; Lagardère, L. Lambda-ABF-OPES: Faster Convergence with High Accuracy in Alchemical Free Energy Calculations. *J. Phys. Chem. Lett.* **2025**, *16*, 4626–4634.

(46) Kong, X.; Brooks, C. L. λ-dynamics: A new approach to free energy calculations. *J. Chem. Phys.* **1996**, *105*, 2414–2423.

(47) Knight, J. L.; Brooks, C. L. I. Lambda-dynamics free energy simulation methods. *J. Comput. Chem.* **2009**, *30*, 1692–1700.

(48) Invernizzi, M.; Parrinello, M. Rethinking Metadynamics: From bias potentials to probability Distributions. *J. Phys. Chem. Lett.* **2020**, *11*, 2731–2736.

(49) Minoukadeh, K.; Chipot, C.; Lelièvre, T. Potential of Mean Force Calculations: A Multiple-Walker Adaptive Biasing Force Approach. *J. Chem. Theory Comput.* **2010**, *6*, 1008–1017.

(50) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100*, 020603.

(51) Barducci, A.; Bonomi, M.; Parrinello, M. Linking well-tempered metadynamics simulations with experiments. *Biophys. J.* **2010**, *98*, L44–L46.

(52) Fiorin, G.; Klein, M. L.; Hénin, J. Using collective variables to drive molecular dynamics simulations. *Mol. Phys.* **2013**, *111*, 3345–3362.

(53) Hénin, J.; Fiorin, G.; Chipot, C.; Klein, M. L. Exploring multidimensional free energy landscapes using time-dependent biases on collective variables. *J. Chem. Theory Comput.* **2010**, *6*, 35–47.

(54) Salari, R.; Joseph, T.; Lohia, R.; Hénin, J.; Brannigan, G. A Streamlined, general approach for computing ligand binding free energies and its application to GPCR-bound cholesterol. *J. Chem. Theory Comput.* **2018**, *14*, 6560–6573.

(55) Shi, Y.; Ren, P.; Schnieders, M.; Piquemal, J.-P. *Reviews in Computational Chemistry Volume 28*; John Wiley & Sons, Ltd, 2015; Chapter 2, pp 51–86.

(56) Melcr, J.; Piquemal, J.-P. Accurate Biomolecular Simulations Account for Electronic Polarization. *Front. mol. biosci.* **2019**, *6*, 143.

(57) Jing, Z.; Liu, C.; Cheng, S. Y.; Qi, R.; Walker, B. D.; Piquemal, J.-P.; Ren, P. Polarizable Force Fields for Biomolecular Simulations: Recent Advances and Applications. *Annual Review of Biophysics* **2019**, *48*, 371–394.

(58) de Courcy, B.; Piquemal, J.-P.; Garbay, C.; Gresh, N. Polarizable Water Molecules in Ligand-Macromolecule Recognition. Impact on the Relative Affinities of Competing Pyrrolopyrimidine Inhibitors for FAK Kinase. *Journal of the American Chemical Society* **2010**, *132*, 3312–3320, PMID: 20178314.

(59) Klesse, G.; Rao, S.; Tucker, S. J.; Sansom, M. S. Induced Polarization in Molecular Dynamics Simulations of the 5-HT3 Receptor Channel. *J. Am. Chem. Soc.* **2020**, *142*, 9415–9427.

(60) Jaffrelot Inizan, T.; Célerse, F.; Adjoua, O.; El Ahdab, D.; Jolly, L.-H.; Liu, C.; Ren, P.; Montes, M.; Lagarde, N.; Lagardère, L.; Monmarché, P.; Piquemal, J.-P. High-resolution mining of the SARS-CoV-2 main protease conformational space: supercomputer-driven unsupervised adaptive sampling. *Chem. Sci.* **2021**, *12*, 4889–4907.

(61) El Ahdab, D.; Lagardère, L.; Inizan, T. J.; Célerse, F.; Liu, C.; Adjoua, O.; Jolly, L.-H.; Gresh, N.; Hobaika, Z.; Ren, P.; Maroun, R. G.; Piquemal, J.-P. Interfacial Water Many-Body Effects Drive Structural Dynamics and Allosteric Interactions in SARS-CoV-2 Main Protease Dimerization Interface. *J. Phys. Chem. Lett.* **2021**, *12*, 6218–6226, PMID: 34196568.

(62) El Khoury, L. et al. Computationally driven discovery of SARS-CoV-2 Mpro inhibitors: from design to experimental validation. *Chem. Sci.* **2022**, *13*, 3674–3687.

(63) Blazhynska, M.; Lagardère, L.; Liu, C.; Adjoua, O.; Ren, P.; Piquemal, J.-P. Water–glycan interactions drive the SARS-CoV-2 spike dynamics: insights into glycan-gate control and camouflage mechanisms. *Chem. Sci.* **2024**, *15*, 14177–14187.

(64) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B* **2010**, *114*, 2549–2564.

(65) Timm, D. E.; Baker, L.; Mueller, H.; Zidek, L.; Novotny, M. V. Structural basis of pheromone binding to mouse major urinary protein (MUP-I). *Protein Science* **2001**, *10*, 997–1004.

(66) Crawford, T. D. et al. Diving into the Water: Inducible Binding Conformations for BRD4, TAF1(2), BRD9, and CECR2 Bromodomains. *J. Med. Chem.* **2016**, *59*, 5391–5402.

(67) Woodhead, A. J. 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.

(68) 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.

(69) Kong, L.; Park, S. J.; Im, W. CHARMM-GUI PDB Reader and Manipulator: Covalent Ligand Modeling and Simulation. *J. Mol. Biol.* **2024**, *436*, 168554.

(70) Rackers, J. A.; Wang, Z.; Lu, C.; Laury, M. L.; Lagardère, L.; Schnieders, M. J.; Piquemal, J.-P.; Ren, P.; Ponder, J. W. Tinker 8: Software Tools for Molecular Design. *J. Chem. Theory Comput.* **2018**, *14*, 5273–5289.

(71) Wang, Z. Tinker9: Next Generation of Tinker with GPU Support. https://github.com/TinkerTools/tinker9, 2021; Washington University in St. Louis.

(72) Wu, J. C.; Chattree, G.; Ren, P. Automation of AMOEBA polarizable force field parameterization for small molecules. *Theor. Chem. Acc.* **2012**, *131*, 1138.

(73) Walker, B.; Liu, C.; Wait, E.; Ren, P. Automation of AMOEBA polarizable force field for small molecules: Poltype 2. *J. Comput. Chem.* **2022**, *43*, 1530.

(74) Lagardère, L.; Jolly, L.-H.; Lipparini, F.; Aviat, F.; Stamm, B.; Jing, Z. F.; Harger, M.; Torabifard, H.; Cisneros, G. A.; Schnieders, M. J.; Gresh, N.; Maday, Y.; Ren, P.; Ponder, J. W.; Piquemal, J.-P. Tinker-HP: a Massively Parallel Molecular Dynamics Package for Multiscale Simulations of Large Complex Systems with Advanced Polarizable Force Fields. *Chem. Sci.* **2018**, *9*, 956–972, Open Access.

(75) Lagardère, L.; Aviat, F.; Piquemal, J.-P. Pushing the Limits of Multiple-Time-Step Strate-

gies for Polarizable Point Dipole Molecular Dynamics. *The Journal of Physical Chemistry Letters* **2019**, *10*, 2593–2599.

(76) Lagardère, L.; Jolly, L.-H.; Lipparini, F.; Aviat, F.; Stamm, B.; Jing, Z. F.; Harger, M.; Torabifard, H.; Cisneros, G. A.; Schnieders, M. J.; Gresh, N.; Maday, Y.; Ren, P. Y.; Ponder, J. W.; Piquemal, J.-P. Tinker-HP: a massively parallel molecular dynamics package for multiscale simulations of large complex systems with advanced point dipole polarizable force fields. *Chem. Sci.* **2018**, *9*, 956–972.

(77) Adjoua, O.; Lagardère, L.; Jolly, L.-H.; Durocher, A.; Very, T.; Dupays, I.; Wang, Z.; Inizan, T. J.; Célerse, F.; Ren, P.; Ponder, J. W.; Piquemal, J.-P. Tinker-HP: Accelerating Molecular Dynamics Simulations of Large Complex Systems with Advanced Point Dipole Polarizable Force Fields Using GPUs and Multi-GPU Systems. *J. Chem. Theory Comput.* **2021**, *17*, 2034–2053.

(78) Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J. Chem. Theory Comput.* **2015**, *11*, 1864–1874.

(79) Santiago-McRae, E.; Ebrahimi, M.; Sandberg, J. W.; Brannigan, G.; Hénin, J. Computing Absolute Binding Affinities by Streamlined Alchemical Free Energy Perturbation (SAFEP) [Article v1.0]. *Living J. Comput. Mol. Sci.* **2023**, *5*, 2067, Tutorial article; includes example calculations using NAMD and VMD.

(80) Comer, J.; Phillips, J.; Schulten, K.; Chipot, C. Multiple-walker strategies for free-energy calculations in NAMD: Shared adaptive biasing force and walker selection rules. *J. Chem. Theory Comput.* **2014**, *10*, 5276–5285.

(81) Invernizzi, M.; Parrinello, M. Exploration vs Convergence Speed in Adaptive-Bias Enhanced Sampling. *J. Chem. Theory Comput.* **2022**, *18*, 3988–3996.

(82) Ray, D.; Ansari, N.; Rizzi, V.; Invernizzi, M.; Parrinello, M. Rare Event Kinetics from Adaptive Bias Enhanced Sampling. *J. Chem. Theory Comput.* **2022**, *18*, 6500–6509.

(83) Branduardi, D.; Bussi, G.; Parrinello, M. Metadynamics with Adaptive Gaussians. *J. Chem. Theory Comput.* **2012**, *8*, 2247–2254.

(84) Dama, J. F.; Rotskoff, G.; Parrinello, M.; Voth, G. A. Transition-Tempered Metadynamics: Robust, Convergent Metadynamics via On-the-Fly Transition Barrier Estimation. *J. Chem. Theory Comput.* **2014**, *10*, 3626–3633.

(85) Schnieders, M. J.; Baltrusaitis, J.; Shi, Y.; Chattree, G.; Zheng, L.; Yang, W.; Ren, P. The Structure, Thermodynamics, and Solubility of Organic Crystals from Simulation with a Polarizable Force Field. *J. Chem. Theory Comput.* **2012**, *8*, 1721–1736, PMID: 22582032.

(86) Halgren, T. A. The representation of van der Waals (vdW) interactions in molecular mechanics force fields: potential form, combination rules, and vdW parameters. *Journal of the American Chemical Society* **1992**, *114*, 7827–7843.

(87) Ansari, N. ABFE-L-ABF-OPES. https://github.com/ansarinarjes/ABFE-L-ABF-OPES, 2024; Accessed: 2025-10-16.

(88) Sharrow, S. D.; Vaughn, J. L.; Zídek, L.; Novotny, M. V.; Stone, M. J. Pheromone binding by polymorphic mouse major urinary proteins. *Protein Sci.* **2002**, *11*, 2247–2256.

(89) Sanchez, R.; Zhou, M. M. The role of human bromodomains in chromatin biology and gene transcription. *Curr. Opin. Drug Discov. Devel.* **2009**, *12*, 659–665.

(90) Filippakopoulos, P.; Picaud, S.; Mangos, M.; Keates, T.; Lambert, J.-P.; Barsyte-Lovejoy, D.; Felletar, I.; Volkmer, R.; Müller, S.; Pawson, T.; Gingras, A.-C.; Arrowsmith, C. H.; Knapp, S. Histone Recognition and Large-Scale Structural Analysis of the Human Bromodomain Family. *Cell* **2012**, *149*, 214–231.

(91) Crombie, E. M.; Cleverley, K.; Timmers, H. T. M.; Fisher, E. M. C. The roles of TAF1 in neuroscience and beyond. *R. Soc. Open Sci.* **2024**, *11*, 240790.

(92) Romero, F. A.; Taylor, A. M.; Crawford, T. D.; Tsui, V.; Côté, A.; Magnuson, S. Disrupting Acetyl-Lysine Recognition: Progress in the Development of Bromodomain Inhibitors. *J. Med. Chem.* **2016**, *59*, 1271–1298, PMID: 26572217.

(93) Li, Y.; Zhao, J.; Gutgesell, L. M.; Shen, Z.; Ratia, K.; Dye, K.; Dubrovskyi, O.; Zhao, H.; Huang, F.; Tonetti, D. A.; Thatcher, G. R. J.; Xiong, R. Novel Pyrrolopyridone Bromodomain and Extra-Terminal Motif (BET) Inhibitors Effective in Endocrine-Resistant ER+ Breast Cancer with Acquired Resistance to Fulvestrant and Palbociclib. *J. Med. Chem.* **2020**, *63*, 7186–7210, PMID: 32453591.

(94) Aldeghi, M.; Ross, G. A.; Bodkin, M. J.; et al. Large-scale analysis of water stability in bromodomain binding pockets with grand canonical Monte Carlo. *Commun. Chem.* **2018**, *1*, 19.

(95) Flower, D. R. The lipocalin protein family: Structure and function. *Biochem. J.* **1996**, *318*, 1–14.

(96) Zhou, Y.; Rui, L. Major urinary protein regulation of chemical communication and nutrient metabolism. *Vitam. Horm.* **2010**, *83*, 151–163.

(97) Novotny, M. V.; Jemiolo, B.; Wiesler, D.; Ma, W.; Harvey, S.; Xu, F.; Xie, T. M.; Carmack, M. A unique urinary constituent, 6-hydroxy-6-methyl-3-heptanone, is a pheromone that accelerates puberty in female mice. *Chem. Biol.* **1999**, *6*, 377–383.

(98) Dunker, A. K.; Garner, E.; Guilliot, S.; Romero, P.; Albrecht, K.; et al. Protein disorder and the evolution of molecular recognition: theory, predictions and observations. Pacific Symposium on Biocomputing. 1998; pp 473–484.

(99) Zhou, H. X. Intrinsic disorder: signaling via highly specific but short-lived association. *Trends Biochem. Sci.* **2012**, *37*, 43–48.

(100) Chu, X.; Wang, J. Specificity and Affinity Quantification of Flexible Recognition from Underlying Energy Landscape Topography. *PLoS Comput. Biol.* **2014**, *10*, e1003782.

(101) Chipot, C.; Kollman, P. A.; Pearlman, D. A. Alternative Approaches to Potential of Mean Force Calculations: Free Energy Perturbation versus Thermodynamic Integration. Case Study of Some Representative Nonpolar Interactions. *J. Comput. Chem.* **1996**, *17*, 1112–1131.

(102) Amadasi, A.; Surface, A.; Spyrakis, F.; Cozzini, P.; Mozzarelli, A.; Kellogg, G. E. Robust classification of "relevant" water molecules in putative protein binding sites. *J. Med. Chem.* **2008**, *51*, 1063–1067.

(103) Nilmeier, J. P.; Crooks, G. E.; Minh, D. D. L.; Chodera, J. D. Nonequilibrium Candidate Monte Carlo Is an Efficient Tool for Equilibrium Simulation. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, E1009–E1018.

(104) Adams, D. Chemical Potential of Hard-Sphere Fluids by Monte Carlo Methods. *Mol. Phys.* **1974**, *28*, 1241–1252.

(105) Adams, D. Grand Canonical Ensemble Monte Carlo for a Lennard-Jones Fluid. *Mol. Phys.* **1975**, *29*, 307–311.

(106) Mezei, M. A Cavity-Biased (T, V, μ) Monte Carlo Method for the Computer Simulation of Fluids. *Mol. Phys.* **1980**, *40*, 901–906.

(107) Mezei, M. Grand-Canonical Ensemble Monte Carlo Study of Dense Liquid: Lennard-Jones, Soft Spheres and Water. *Mol. Phys.* **1987**, *61*, 565–582.

(108) Plé, T.; Adjoua, O.; Benali, A.; Posenitskiy, E.; Villot, C.; Lagardère, L.; Piquemal, J.-P. A

Foundation Model for Accurate Atomistic Simulations in Drug Design. 2025; 10.26434/chemrxiv-2025-f1hgn-v3.

(109) Benali, A. et al. Pushing the Accuracy Limit of Foundation Neural Network Models with Quantum Monte Carlo Forces and Path Integrals. 2025; https://arxiv.org/abs/2504.07948.

(110) Cattin, C.; Plé, T.; Adjoua, O.; Gouraud, N.; Lagardère, L.; Piquemal, J.-P. Accelerating Molecular Dynamics Simulations with Foundation Neural Network Models using Multiple Time-Step and Distillation. 2025; https://arxiv.org/abs/2510.06562.