

Toward Reconciling the Standard Binding Free Energy of Lenacapavir to HIV-1 Capsid with Experiment: Thermodynamic Effects of Solvent Buffer and Ligand Reorganization

Published as part of *The Journal of Physical Chemistry B special issue "Molecular Simulation and Computational Chemistry: The Legacy of Peter A. Kollman"*.

Qinfang Sun, Emilio Gallicchio, Ronald Levy, and Nanjie Deng*

Cite This: <https://doi.org/10.1021/acs.jpcb.5c06714>

Read Online

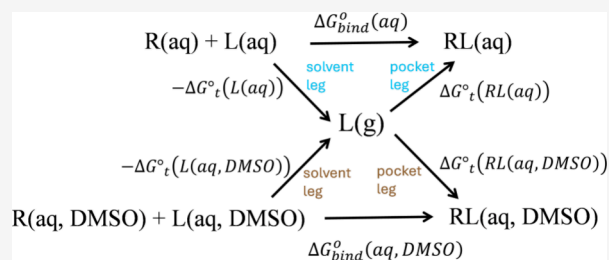
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: We report a large thermodynamic effect of solvent buffer on the standard binding free energy for a large hydrophobic ligand and show that a realistic comparison with the experimental binding affinity requires correctly accounting for the solvent reference state and ligand reorganization. We focus on lenacapavir (LEN; MW \approx 1 kDa), an HIV-1 capsid inhibitor with very low aqueous solubility. Using several absolute binding free energy (ABFE) protocols including double-decoupling method (DDM), potential-of-mean-force (PMF) approaches, and the Alchemical Transfer Method (ATM), we obtained standard binding free energy values in neat water of $\Delta G_{\text{bind}}^0 = -26.4$ to -30.0 kcal/mol for LEN binding to the HIV-1 capsid CA dimer, much stronger than the SPR-derived affinity measured in 5% DMSO buffer (≈ -13.4 kcal/mol). We analyze the discrepancy and identify two dominant contributors to the calculated overbinding. (i) Solvent reference state: A thermodynamic cycle analysis and solvation free energy calculations reveal that the 5% DMSO buffer stabilizes the free ligand by ~ -4 kcal/mol relative to neat water. Additionally, we propose that the enrichment of the hydrophobic cosolvent in the apo binding pocket makes it more energetically costly to displace DMSO upon ligand binding. (ii) Ligand reorganization: incomplete treatment of LEN's internal conformational reorganization in the ABFE protocols leads to overbinding; a DDM variant with explicit ligand reorganization reduces the overestimate by ~ 6 kcal/mol. Together, considerations of these effects significantly reduce the discrepancy between the ABFE calculations and experiments. Our results suggest that for large, hydrophobic ligands, quantitative agreement between ABFEs and experiments requires (a) reporting ΔG_{bind}^0 in the appropriate assay buffer (not simply in water) and (b) explicit treatment of ligand reorganization.



INTRODUCTION

Accurately predicting absolute binding free energies (ABFEs) is essential for understanding the thermodynamics that drive molecular recognition and for guiding rational drug design.^{1–7} Contemporary ABFE workflows can achieve mean unsigned errors of ~ 1 – 3 kcal/mol for compounds that satisfy Lipinski's Rule of Five (≤ 500 Da; modest flexibility)^{4,8–16} (although in some challenging cases protein reorganization can lead to errors of up to ~ 10 kcal/mol⁴). While most approved small-molecule drugs fall in this range of size and flexibility, a small but growing subset of approved drugs and clinical candidates extends to larger sizes, sometimes 600–1000 Da with ≈ 10 or more rotatable bonds.^{17,18} For such large and flexible ligands, achieving adequate sampling of ligand and protein reorganization and binding-site water exchange within practical simulation times can be very challenging. A further experimental complication is that, for hydrophobic ligands, very low aqueous solubility often necessitates cosolvents (e.g., a few percent DMSO, sometimes with a trace surfactant) to

prevent precipitation during affinity measurements; this changes the solvent reference state and can shift the observed binding equilibrium relative to neat water, an important consideration that is often overlooked in absolute binding free energy calculations.¹⁹

Here we focus on lenacapavir (LEN), an HIV-1 capsid inhibitor, to assess how current ABFE methods perform for a large hydrophobic ligand. LEN is a long-acting antiretroviral drug in HIV/AIDS treatment and prevention.²⁰ It targets the HIV-1 capsid protein (CA) and interferes with processes such as nuclear import, reverse transcription, and uncoating.^{20,21}

Received: September 27, 2025

Revised: January 29, 2026

Accepted: January 30, 2026

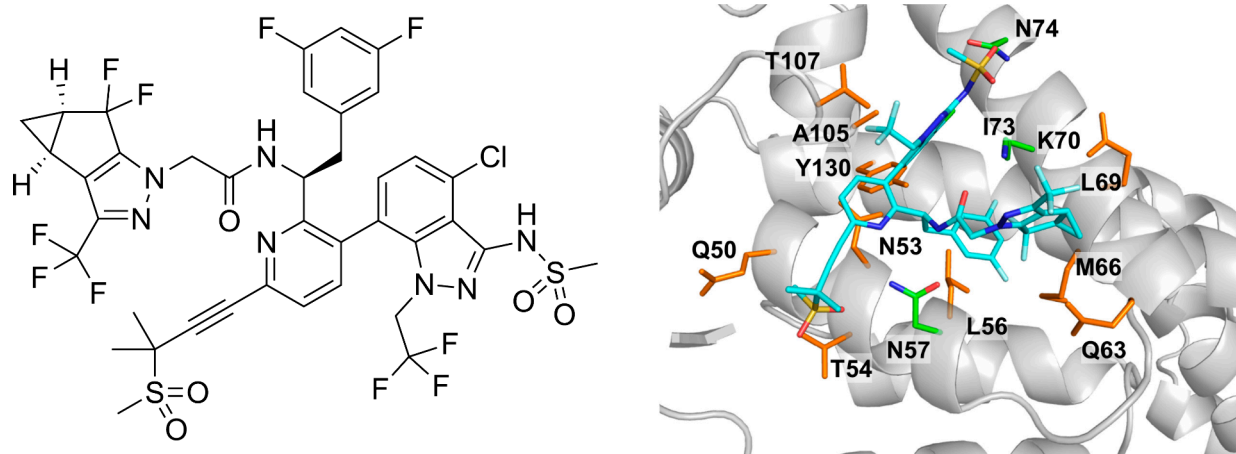


Figure 1. Left: chemical structures of lenacapavir. Right: X-ray structure of lenacapavir (blue) bound at the dimeric interface of two CA monomers.

With 968 Da molecular weight and 13 rotatable bonds, it ranks among the largest FDA-approved small-molecule drugs, in terms of both molecular weight and structural complexity (Figure 1). LEN has very low solubility in water, and its reported experimental binding affinity was measured by surface plasmon resonance (SPR) in 5% (v/v) DMSO buffer.²²

We applied several ABFE methods, implemented across different simulation packages, to compute the standard binding free energy of LEN binding to the HIV-1 capsid dimer in water. The ABFE methods used include the standard double decoupling method (DDM),^{23,24} the recently developed Alchemical Transfer Method (ATM),^{25–29} and a hybrid of the alchemical and potential of mean force method (Alchem-PMF).³⁰ To improve the sampling of water equilibration when decoupling a ligand of this size, we also test a fragment-wise, progressive double decoupling scheme

These ABFE protocols produced ΔG_{bind}^0 values much more favorable than the experimental estimate from SPR in 5% (v/v) DMSO. The overestimation persisted across methods, software platforms, structural models, and protein oligomeric forms. Performing the alchemical calculation in fragments (progressive/fragment-wise decoupling) or changing force-field parameters had only a modest impact on ΔG_{bind}^0 . Alternative ligand protonation states and protein reorganization also produced only small changes. We therefore examined two effects for the binding of this large hydrophobic ligand, one often treated inadequately and the other largely neglected in ABFE calculations: (i) the ligand-reorganization free energy cost and (ii) the assay solvent composition. As shown below, considering both effects helps reconcile the ABFE calculations with the experiment.

METHODS

System Preparation

Starting structures of HIV-1 capsid (CA) in complex with LEN were taken from the experimental holo (PDB: 6v2f) and apo (PDB: 4xfx) capsid hexamer. The dimer was the standard oligomeric form used in this work for all ABFE calculations and all methods (Schrödinger ABFEP, GROMACS DDM, ATM, Progressive-DDM, DDM with ligand reorganization, and Alchem-PMF). The hexameric form was investigated exclusively using the Schrödinger ABFEP to assess potential structural stability or allosteric effects on the calculated binding free energy. LEN was modeled in both the protonated and deprotonated forms. All systems were solvated in explicit water, neutralized, and brought to an experimental ionic strength (150 mM

NaCl). Unless otherwise noted, TIP3P water was used across the ABFE protocols. Equilibration comprised restrained minimization (protein heavy atoms and ligands), gradual heating to 300 K, and NPT relaxation. All simulations used particle-mesh Ewald electrostatics and standard long-range dispersion settings from the respective engines.

ABFEP from Schrödinger Inc

We employed the Absolute Binding Free Energy Perturbation (ABFEP) approach in the Schrödinger Suite (release 2024-2),⁴ which is based on the double decoupling scheme.²⁴ In this method, the calculation begins with the ligand in pure solvent, where the van der Waals and electrostatic interactions within the ligand as well as between the ligand and water molecules are gradually turned off, creating a “dummy” ligand molecule. This dummy ligand is then restrained to the protein binding site using the Boresch–Karplus cross-link restraints.²⁴ Subsequently, the van der Waals and electrostatic interactions within the ligand and between the ligand and the protein and surrounding water are gradually restored, after which the cross-link restraints are released.

The ABFEP runs were initiated from the holo and apo capsid hexamers. Apo-based complexes were generated by superimposing apo and holo structures, transferring the ligand to apo structures (Maestro, Schrödinger 2024-2), where no clashes were observed between the apo protein and Lenacapavir. The CA dimer was extracted from either the apo- or holo hexamer structures. The simulations employed the OPLS4 force field.³¹ The default Schrödinger ABFEP protocol was followed: complexes were solvated in an SPC water box (8 Å buffer for charged LEN, 5 Å for neutral LEN), neutralized with counterions, and supplemented with 0.15 M salt. An MD relaxation protocol was first performed, followed by 1 ns of Grand Canonical Monte Carlo³² μVT simulations at 300 K with protein backbone heavy-atom restraints. This step was necessary to identify the atoms for cross-link restraints. Subsequently, 15–30 ns ABFEP simulations per lambda window were carried out with 108 lambda windows for charged LEN and 68 lambda windows for neutral LEN. The replica exchange production simulation was carried out in the NPT ensemble with solute tempering (REST).³³ The free energy of the transformation was calculated using multistate Bennett’s acceptance ratio method (MBAR).³⁴ This workflow is automated within the FEP+ module of Schrödinger 2024-2.

GROMACS DDM

A GROMACS implementation of DDM¹⁰ (Figure 5) was run using Amber ff99SB-ILDN³⁵ for protein and GAFF2³⁶ for LEN with AM1-BCC³⁷ partial charges. Electrostatics were decoupled first over 11 linear Coulomb lambda windows ($\lambda = 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0$), followed by van der Waals decoupling with 17 LJ lambda windows ($\lambda = 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.94, 0.985, 1.0$) using soft-core potential. Each

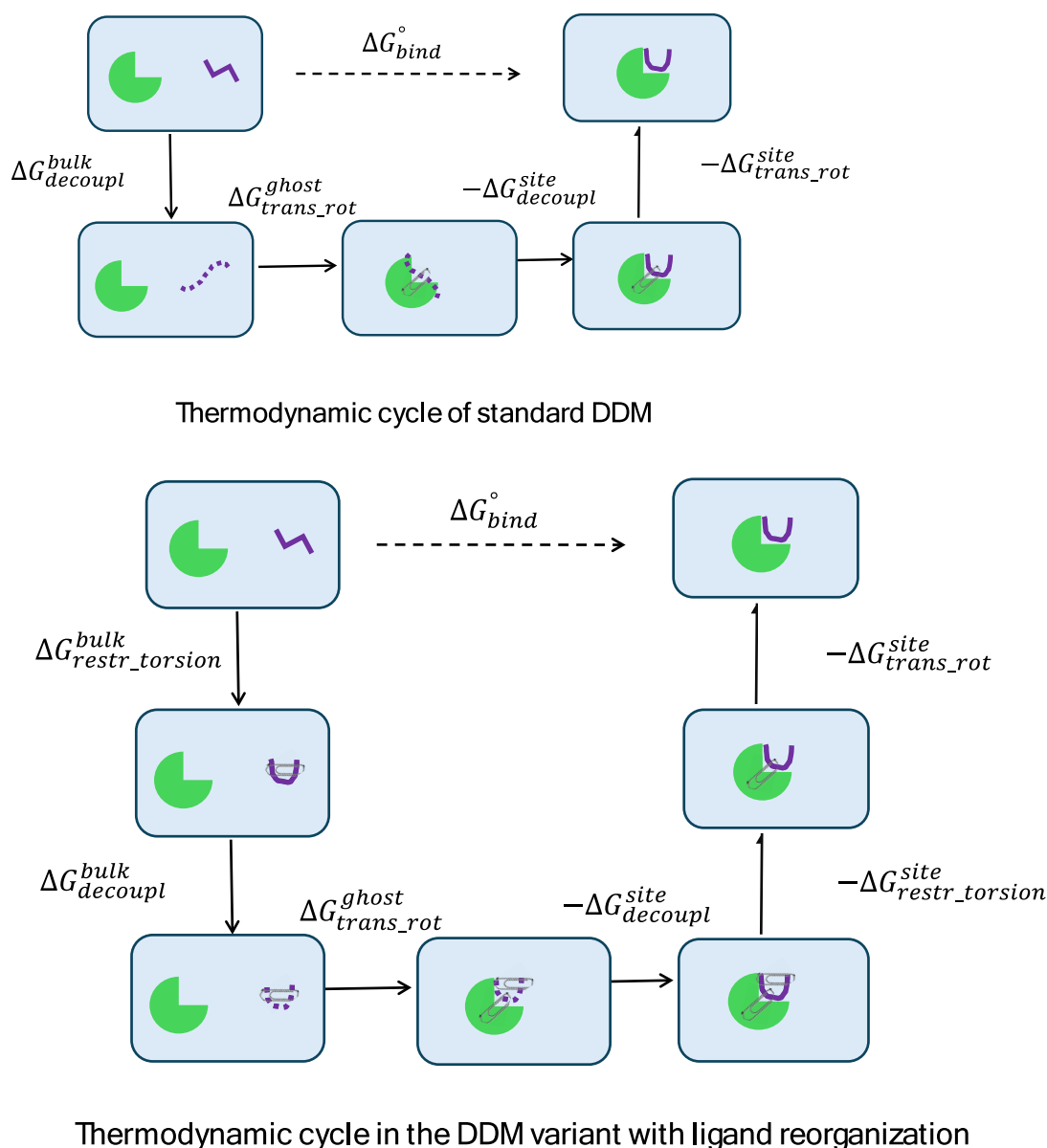


Figure 2. Thermodynamic cycles of (top) standard DDM and (bottom) the DDM with ligand reorganization.

window was equilibrated for 5 ns, followed by 70 ns of simulation to monitor convergence. The last 35 ns were used to compute ΔG_{bind}^0 . The same lambda schedules were used in both the solvent and complex decoupling legs. We have examined the conformational ensembles of the decoupled ligand in both the solvent leg and the complex leg by comparing the sampled distributions of selected dihedral angles: see [Supporting Information Figure S7](#). As expected, these distributions are very similar between the two decoupling legs.

The Boresch–Karplus translation/rotation restraint²⁴ of the ligand was applied in the complex leg using 17 bonded lambda windows ($\lambda = 0.0, 0.01, 0.025, 0.05, 0.075, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0$). At each window $dH/d\lambda$ data were collected from 2 ns simulation and the free energy of application of the Boresch–Karplus restraints was estimated through multistate Bennett’s acceptance ratio method (MBAR).³⁴ The free energy of releasing the same set of ligand translational/rotational restraints in bulk solution at 1 M standard state was calculated analytically.^{24,38}

DDM with Explicit Ligand Reorganization

To address LEN’s many internal torsion degrees of freedom and substantial internal flexibility, we employed a ligand restrain-and-

release variant of DDM, following earlier treatment of conformational free energy schemes.^{15,39}

In standard DDM, i.e.

$$\Delta G_{bind}^0 = -\Delta G_{trans_rot}^{site} - \Delta G_{decoupl}^{site} + \Delta G_{trans_rot}^{ghost} + \Delta G_{decoupl}^{bulk} \quad (1)$$

the accurate calculation of the two decoupling free energies $\Delta G_{decoupl}^{site}$ and $\Delta G_{decoupl}^{bulk}$ which have the largest magnitudes among all the terms in the DDM cycle, is quite challenging because during decoupling the ligand samples an increasingly larger conformational space as the ligand–environment interactions are slowly turned off.

In the DDM variant with ligand reorganization,

$$\Delta G_{bind}^0 = -\Delta G_{trans_rot}^{site} - \Delta G_{restr_torsion}^{site} - \Delta G_{decoupl}^{site} + \Delta G_{trans_rot}^{ghost} + \Delta G_{decoupl}^{bulk} + \Delta G_{restr_torsion}^{bulk} \quad (2)$$

harmonic torsional restraints were applied to ligand rotatable bonds in both the complex and solution environments before decoupling, resulting in a more restricted conformational search space which reduces the difficulty of sampling; see [Figure 2](#). By avoiding the need to sample the large conformational change during the decoupling simulations in the bulk solution and in the complex, the DDM variant

facilitates more accurate calculations of $\Delta G_{\text{decouple}}^{\text{site}}$ and $\Delta G_{\text{decouple}}^{\text{bulk}}$ the two most crucial terms in DDM and improves the ABFE estimate.

To calculate the free energy of applying/releasing the torsional restraints in the bound and unbound ligands (i.e., $\Delta G_{\text{restr_torsion}}^{\text{site}}$ and $\Delta G_{\text{restr_torsion}}^{\text{bulk}}$), harmonic torsional restraints were applied to all the rotatable bonds of LEN using 17 restraint lambda windows ($\lambda = 0.0, 0.01, 0.025, 0.05, 0.075, 0.1, 0.15, 0.2, 0.25, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0$). While 2 ns of the $dH/d\lambda$ data per lambda were sufficient for converging the bound state leg, more than 50 ns per lambda were needed to converge the free energy of the unbound state leg to within 0.6 kcal/mol. All other setups are the same as the standard GROMACS DDM described above.

Progressive DDM (Schrödinger RBFEP, OPLS4)

In this DDM variant designed to overcoming any bottleneck from sampling the solvent exchange during the decoupling of large ligand,⁴⁰ LEN is divided into a core (L) and three functional groups (A, B, C). The RBFEP module was used to calculate relative binding free energies in a stepwise manner, progressing from $L^*ABC \rightarrow L$, as illustrated in Figure S4 (Supporting Information). The absolute binding free energy of the core was then determined using the Schrödinger ABFEP module, following the methodology described above, which includes the use of the Boresch–Karplus restraints on the ligand translation and rotation.

Force Field Builder³¹ (Schrödinger Suite, release 2024-2) was used to parametrize any missing torsions in the LEN-related ligands studied here. The default RBFEP protocol, as implemented in the FEP+ module,^{20,41} was employed: 24 lambda windows and an 8 Å SPC water buffer were used for charged ligands, whereas 16 lambda windows and a 5 Å SPC water buffer were applied for neutral ligands. Following a relaxation stage, 20 ns production simulations were performed in the NPT ensemble with REST,³³ and relative free energies were computed via multistate Bennett's acceptance ratio method (MBAR).^{34,42} To evaluate the consistency of the FEP+ predictions, cycle closure correction⁴³ was applied, and the resulting ΔG^{pred} values were converted to ΔG^{pred} using the absolute binding free energy of the core ligand obtained from the ABFEP protocol.

ATM

The Alchemical Transfer Method (ATM)^{29,28,44} implements ABFE and RBFE protocols formally equivalent to DDM and FEP+, except that the alchemical transformation is achieved by a direct coordinate transformation rather than a series of decoupling or mutating steps of the ligand atoms by modifications of the force field parameters. In the ATM-ABFE protocol, for example, the ligand coordinates are displaced by a fixed amount from the binding site region to a corresponding area in the solvent bulk. In the ATM-RBFE protocol, the two ligands are displaced in opposite directions to switch places in the binding site and the bulk solvent bulk. Because the displacement involves the whole ligand, the ATM-ABFE protocol cannot handle ligands as big as LEN. Hence, in this application, we adopted a progressive ABFE unbinding protocol similar to the progressive DDM approach described above. We used a sequence of nine RBFE steps starting with LEN (lenacapavir in Figure S2, Supporting Information) and ending with fragment I9. Fragment I9 was then transferred to solution using the ATM-ABFE protocol (Figure S2). The sum of the free energies of these ten steps is the excess binding free energy of LEN with a negative sign. For each step, we employed the standard ATM setup workflow, alchemical potential, and λ -schedule (https://github.com/Gallicchio-Lab/AToM-OpenMM/blob/master/examples/RBFE/cdk2/scripts/asyncnc_template.cnt). The center of the binding site was defined as the centroid of the C-alpha atoms of residues 37, 38, 41, 135, 169, 172, 173, 179, 182, 183, 270, 273, 274, 276, 277, 283, 286, 287, 289, 290, 293, 294, 325, 326, 327, and 350 of capsid (PDB ID: 6VKV). The center of LEN was defined as the atom labeled "1" in Figure S2. To define the binding site region, the distance between the center of the binding site and the center of the ligand was restrained with a tolerance of 2 Å based on the observed fluctuations of the position of LEN during a 20 ns preparatory MD simulation. The alignment atoms employed during the ATM-RBFE

steps are labeled "1", "2", and "3" in Figure S2. The length of the ATM displacement vector was set to 43 Å.

Alchem-PMF (Physical Pathway with Alchemical Softening)

ABFE was also computed with Alchem-PMF,³⁰ which combines a physical extraction pathway (PMF) with alchemical softening/recoupling of sterically hindering groups. In brief, the standard PMF defines a pulling coordinate $r = |r(P1-L1)|$ between a receptor anchor atom (P1) and a ligand anchor atom (L1), together with five angular restraints ($\theta, \varphi, \Theta, \Phi, \Psi$) that fix ligand orientation and the linear pulling direction relative to the receptor; umbrella sampling along r yields the one-dimensional PMF $w(r)$ under these orientational restraints. The ABFE is assembled from (i) the reversible work to apply orientational restraints in the bound complex; (ii) the reversible work of moving the restrained ligand from the binding site to a bulk location along r (from the PMF $w(r)$); and (iii) the free energy of analytic release of restraints to the 1 M standard state in bulk.

For enclosed sites or ligands with buried substituents, direct pulling can encounter high, poorly sampled steric barriers and impede water exchange in the vacated pocket. Alchem-PMF circumvents this by alchemically removing (or softening) nonbonded interactions for the minimal set of sterically obstructing atoms (typically a localized ligand moiety and, if needed, nearby receptor side chains) before the PMF pulling, then restoring those interactions after the ligand reaches bulk. The resulting thermodynamic cycle adds two alchemical terms to the standard PMF expression: a bound-state decoupling free energy ($\Delta G_{\text{decouple_bound}}$) and a bulk-state recoupling free energy ($\Delta G_{\text{decouple_bulk}}$). Thirty umbrella windows spanned bound to bulk with a single radial force constant; each window was equilibrated before production and analyzed by WHAM. In the case of the LEN-CA complex, the ligand atoms that are deeply buried in the binding pocket (Figure S3) are alchemically decoupled before pulling.

Calculation of Solvation Free Energy of LEN in DMSO Buffers

The solvation free energies of LEN in DMSO/water buffers were performed using GROMACS at two DMSO concentrations: 5% and 10% v/v. Both LEN and DMSO were parametrized using the General AMBER Force Field GAFF2³⁶ with the AM1-BCC³⁷ partial charge model. The LEN was solvated in a box of mixed solvents containing DMSO/water. The mixed solvent box was prepared using the `gmx insert-molecules` command in GROMACS. To achieve 5% and 10% v/v DMSO concentrations, 20 and 40 DMSO molecules were added, respectively, to a box of 1450 TIP3P water molecules at random positions.

To calculate the solvation free energy, the LEN was first decoupled electrostatically from the solvent by using 11 Coulomb lambda windows ($\lambda_{\text{Coul}} = 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0$), and then the van der Waals solute–solvent interactions were turned off by using 17 LJ lambda windows ($\lambda_{\text{LJ}} = 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.94, 0.985, 1.0$). Soft-core potential was applied during the latter stage to avoid singularities. In each λ window the system was equilibrated for 10 ns and followed by 20 ns of production run to collect $\left\langle \frac{dH}{d\lambda} \right\rangle$ in each lambda windows. The results of $\left\langle \frac{dH}{d\lambda} \right\rangle$ were then used to compute the solvation free energy using thermodynamic integration (TI).

RESULTS

Table 1 shows the ABFE of Lenacapavir with the HIV-1 capsid dimer calculated using different methods. The protocols span both alchemical and physical pathway approaches:

- Schrödinger ABFEP and GROMACS DDM follow the standard double-decoupling method (DDM).^{23,24} The DDM with ligand reorganization uses the same GROMACS DDM protocol but adds ligand torsion-

Table 1. Absolute Binding Free Energy of Calculated by Using Different ABFE Methods

Method (Code, force field)	ΔG_{bind}^0 (kcal/mol)
ABFEP (Schrödinger Inc., OPLS4) ⁴	-26.4 ± 1.1
DDM (GROMACS, Amber ff99SB-ILDN/GAFF2) ¹⁰	-28.7 ± 0.5
DDM with ligand reorganization (GROMACS, Amber ff99SB-ILDN/GAFF2) ¹⁰	-22.5 ± 1.0
Alchem-PMF (GROMACS, Amber ff99SB-ILDN/GAFF2) ³⁰	-27.5 ± 0.3
ATM ABFE Protocol (OpenMM, Amber ff14SB/GAFF2) ^{29,44,45}	-27.2 ± 1.1
Progressive DDM (Schrödinger, OPLS4) ⁴	-30.1 ± 1.2
Experiment (SPR, in 5% DMSO) ²²	-13.4

restrain/release legs to better account for the ligand reorganization cost.

- Alchem-PMF combines a physical pathway PMF extraction with alchemical decoupling to aid removal of a bulky ligand from a buried pocket.³⁰
- ATM^{25,27,28} and Progressive DDM⁴⁰ apply fragment-based strategies that transfer or alchemically decouple ligand fragments in stages, which facilitates water equilibration during solute cavity annihilation and recreation.

The error bars in Table 1 are estimated from variations in ΔG_{bind}^0 obtained from multiple independent runs, each initialized using either different LEN-CA dimer starting structures selected from the 6 LENs in the CA hexamer structure (for ABFEP, DDM/GROMACS, and Progressive DDM) or from different thermalized starting structures with different initial velocities (for Alchem-PMF and ATM).

Apparent convergence of the calculated ΔG_{bind}^0 was assessed by computing ΔG_{bind}^0 on nonoverlapping time blocks (see the convergence plots Figure S5 in the Supporting Information). For ABFEP, Alchem-PMF, and Progressive DDM the block-wise estimates plateau within ~15–30 ns (for ABFEP and Progressive DDM, this refers to the simulation time per λ window; for Alchem-PMF, to simulation time per umbrella sampling window), with late-time variation ≤ 0.5 –0.8 kcal/mol. For ATM-ABFE the plateau is reached between 35 and 40 ns. Both GROMACS DDM and DDM with ligand reorganization showed a small late-time drift of ~0.9–1.0 kcal/mol between 45 and 80 ns (Figure S5).

Several observations follow from Table 1.

First, all methods significantly overestimate the experimental binding free energy of -13.4 kcal/mol, predicting values that are 9–16 kcal/mol more favorable. DDM with ligand reorganization produced a smaller but still substantial overestimation of ~9 kcal/mol. This suggests that inadequate treatment of ligand reorganization is one source of the discrepancy; we return to this in a later section.

Second, consistent results are obtained for the five methods, ABFEP (Schrödinger), DDM (GROMACS), Alchem-PMF (GROMACS), ATM ABFE Protocol (OpenMM) and Progressive DDM (implemented using ABFEP and FEP+ from Schrödinger) across different thermodynamic pathways, software package, and force fields. The two standard DDM implementations, Schrödinger ABFEP and GROMACS DDM which use different force field and sampling protocols (e.g., use of REST2 or not), produce similar standard binding free energies (-26 to -29 kcal/mol). The results from the hybrid Alchem-PMF (-27.5 kcal/mol) and the ATM ABFE Protocol

(-27.2 kcal/mol) fall within the same range as the standard DDM, indicating that switching between a purely alchemical pathway, a hybrid physical-plus-alchemical scheme, and an alchemical-transfer pathway does not affect the final affinity estimate for this system.

Third, both fragment-based protocols, ATM-ABFEP and Progressive DDM, gave ΔG_{bind}^0 values in a narrow range from -27 to -30 kcal/mol, closely matching the results from the other protocols. Like ATM-ABFEP, Progressive DDM was designed to improve water equilibration during cavity creation, yet it gave a slightly more favorable value (~3.7 kcal/mol) than regular ABFEP that used the same OPLS4 force field. This indicates that, for this system, insufficient sampling of ligand-cavity annihilation/creation is unlikely to explain the much larger ~13 kcal/mol overestimation relative to the SPR measurement.

Because the overestimation is large across very different ABFE engines and pathways, we next examined common sources of ABFE error to see whether any of them could plausibly account for the large gap of 9–16 kcal/mol between the computation and the experiments.

CONVENTIONAL SUSPECTS DO NOT EXPLAIN THE OVERESTIMATION IN ΔG_{BIND}^0

Ligand Protonation State

Lenacapavir's sulfonamide nitrogen can be either protonated or deprotonated near physiological pH. A relative binding free energy (FEP+ from Schrödinger Inc.) calculation between the two protonation states gave $\Delta G_{bind}^{LEN \rightarrow LEN^-} = -1.1$ kcal/mol impact on affinity. This minor shift confirms that the protonation state is not a significant contributor to the large overestimation.

Protein Oligomeric form

Because mature HIV-1 capsid is primarily composed of CA hexamers, we compared ABFEP results for the dimer and hexamer. The two receptors yielded nearly identical affinities (-27.4 ± 1.1 vs -27.1 ± 1.0 kcal/mol), differing by <0.5 kcal/mol. Thus, receptor oligomeric state can be ruled out as a source of systematic error.

Protein Reorganization

When receptor reorganization induced by ligand binding is not adequately sampled within the finite trajectories used in the ABFE, the calculated ΔG_{bind}^0 can be biased. In holo-initiated DDM runs, insufficient sampling of the apo-like conformations in the fully decoupled λ window tends to overestimate affinity; in apo-initiated runs, insufficient sampling of holo-like conformations in the fully coupled λ window tends to underestimate affinity. Accordingly, the difference between holo- and apo-initiated ABFEs ($\Delta \Delta G_{bind}^0$) provides a practical upper-bound diagnostic for the free energy cost of receptor reorganization.^{46–49,4} To this end, we have computed ΔG_{bind}^0 starting from both holo and apo capsid models in dimer and hexamer contexts. Complexes based on the apo structures were prepared by superimposing apo and holo capsids and transferring the bound LEN and deleting the holo protein from the complex. Across all setups, ΔG_{bind}^0 values vary between -25.7 and -26.4 kcal/mol, with a ≤ 1 kcal/mol difference between holo- and apo-initiated calculations (holo slightly more favorable for both dimer and hexamer cases). If we believe that $\Delta \Delta G_{bind}^0$ computed from the experimental holo and apo structures reflects the protein reorganization cost, then

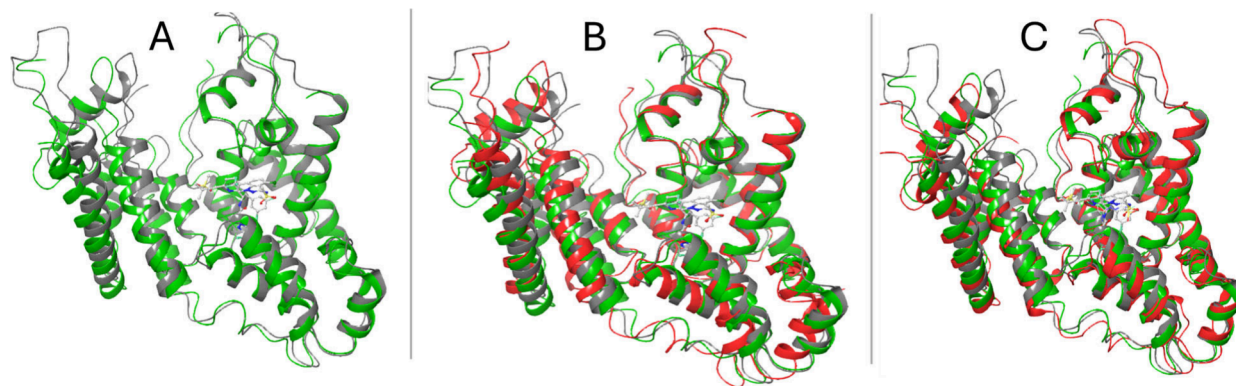


Figure 3. (A) Experimental dimer structures from PDB 4XFX (apo, gray) and 6V2F (holo, green). (B) Experimental dimer structures from PDB 4XFX (apo, gray) and 6V2F (holo, green), along with the capsid dimer conformation in the decoupled state initiated from the experimental holo complex (red). (C) Experimental dimer structures from PDB 4XFX (apo, gray) and 6V2F (holo, green), together with the capsid dimer conformation in the MD-relaxed complex initiated from the apo structure with docked LEN (red).

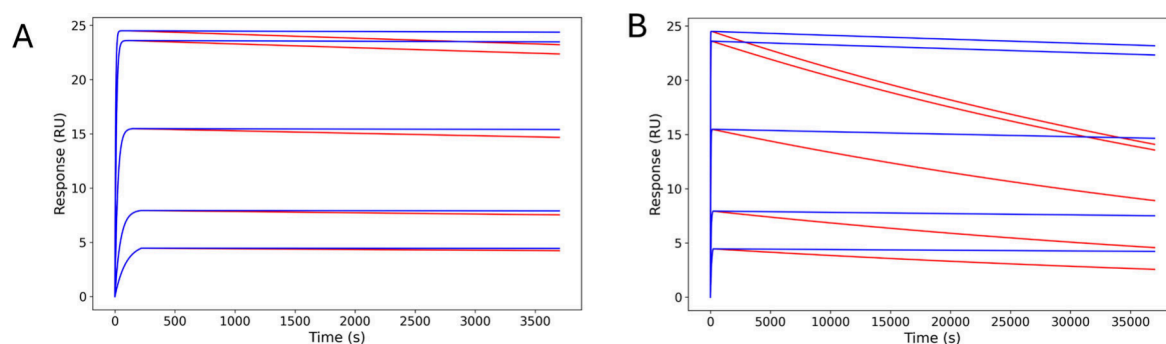


Figure 4. A: The binding fit curves using the originally reported k_{on} and k_{off} (1.5×10^{-5}) are shown in orange, while the fit curves using the same k_{on} but a 10-fold smaller k_{off} (1.5×10^{-6}) is shown in blue. B: The time scale here is 10 times longer than that in panel A.

this small shift suggests that protein conformational reorganization does not contribute substantially to the ~ 9 – 16 kcal/mol gap between computed ABFEs and the SPR measurement.

When we superimposed capsid dimers from the holo and apo experimental structures on the N-terminal domain (NTD) of the monomer that makes the most extensive contact with LEN, we observed distal allosteric differences, particularly in the NTD of the other monomer (Figure 3A). In our DDM runs, the decoupled complex-leg end state, which should correspond to the apo protein if well converged, relaxed to an intermediate between the holo- and apo structures, especially in regions exhibiting allostery (Figure 3B). Moreover, in MD-relaxed apo+LEN starting structures, the allosteric pattern did not fully match that of the holo state (Figure 3C).

Taken together, these observations indicate that distal allosteric reorganization is not fully sampled and may contribute modestly to the difference between calculated ΔG_{bind}^0 and the SPR result, even though the small calculated $\Delta \Delta G_{\text{bind}}^0$ between the results obtained using the apo and holo structures indicates that reorganization is unlikely to be a dominant source of the ~ 9 – 16 kcal/mol discrepancy. This suggests that HIV-1 capsid differs from more flexible targets (e.g., kinases), where protein reorganization penalties was shown to reach ~ 5 kcal/mol and significantly impact binding free energy calculations.⁴⁹

No Fitting Artifacts in the SPR Kinetics

To test whether the long dissociation tail in the SPR sensorgrams (Figure 1A of ref 6) could reflect an overestimated k_{off} (and thus an artificially weak affinity), we

regenerated association/dissociation curves using the reported global-fit parameters (Figure S1; Equations S1–S4) and then overlaid dissociation traces computed with a 10-fold smaller k_{off} while keeping k_{on} fixed. The reduced k_{off} curves decay markedly more slowly and no longer match the measured traces (Figure 4A); when the time axis is extended by 10 \times , the mismatch is even more pronounced (Figure 4B). Thus, the original global fit does not appear to overestimate the k_{off} .

Taken together, these control calculations show that none of the factors considered explain the ~ 9 – 16 kcal/mol gap between computation and experiment. We now focus on two effects: ligand reorganization free energy cost for which DDM with ligand reorganization provides a better account (Table 1) and the effects from the experimental solvent (5% DMSO) used in the SPR affinity measurements.

THE DDM VARIANT WITH LIGAND REORGANIZATION REDUCES THE LEN OVERBINDING

Applying the modified DDM variant with the ligand torsion restrain/release cycle gave a standard binding free energy of -22.5 ± 1.0 kcal/mol, i.e., ≈ 6 kcal weaker than $\Delta G_{\text{bind}}^0 = -28.7 \pm 0.5$ kcal/mol from the standard DDM using the same GROMACS DDM setup (Table 1). LEN is a large molecule with 13 rotatable bonds which makes the sampling of the ligand degrees of freedom challenging for standard ABFE protocols employing tens of nanoseconds of simulation time to capture the ligand reorganization free energy. As shown in Figure 5, LEN undergoes a very large reorganization between

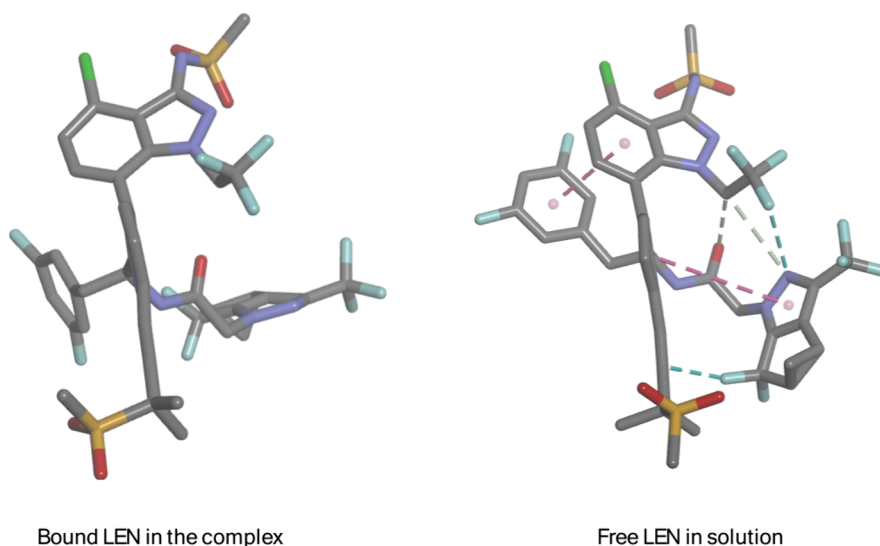


Figure 5. Left: the experimental structure of the bound LEN in the complex. Right: a representative MD structure of the free LEN in solution. Red dashed lines indicate pi-pi stacking interactions. Green lines indicate VDW interactions.

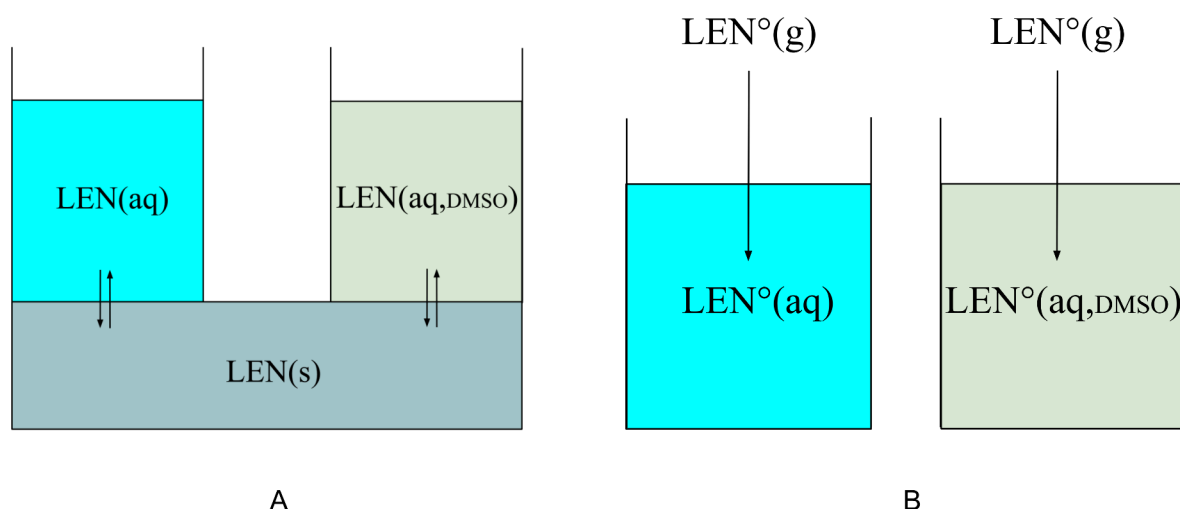


Figure 6. A. Diagram that illustrates the derivation of eq 4. Solid LEN ($\text{LEN}(s)$) is at bottom and saturated LEN solutions in pure water (left) and a mixture of 5% DMSO and water (right) are in equilibrium with the solid. The chemical potential of LEN is the same in these three environments. B. Diagrams illustrating the transfer free energy of LEN from the gas phase to a pure water solution (left) and a 5% DMSO/water mixture (right) at their standard states denoted by a superscript circle.

its structure in solution and that in the bound complex. While the free LEN in solution is stabilized by multiple nonbonded intramolecular interactions, such favorable intramolecular interactions are completely absent in the bound LEN (Figure 5), which is stabilized by favorable interactions with the protein (not shown in Figure 5).

By restraining ligand torsions in the bound, fully coupled state and releasing them in a separate step, when the ligand is fully coupled in solution (see Figure 2 and eq 2), the DDM variant better accounted for such extensive ligand reorganization in ABFE calculation. In contrast, the limited simulation times per λ window used in the standard ABFE protocols were insufficient to fully explore the unbound conformational ensemble of LEN and hence underestimated the reorganization free energy penalty. This conclusion is supported by the convergence analysis (see Figure S5, Supporting Information) generally indicating a slow upward drift of the ABFE binding free energy estimates toward less negative values as the

simulations with the standard protocols are extended to longer times.

REMAINING ERROR OF ~ 9 KCAL/MOL: IMPACT OF SOLVENT REFERENCE STATES

Even with the improvement in accounting for the ligand reorganization effect using the DDM variant of Figure 2, $\Delta G_{bin}^{4\sigma}$ is still about -9 kcal/mol too favorable compared with that obtained from experiment, i.e., -22.5 kcal/mol vs -13.4 kcal/mol (Table 1). In this section, we provide quantitative evidence and a thermodynamic framework to show that this remaining error is likely attributable to the different solvent conditions used in the SPR measurements (5% v/v DMSO buffer) and ABFE calculations (neat water). LEN is practically insoluble in water (solubility $S \approx 0.31$ $\mu\text{g}/\text{mL}$) but highly soluble in DMSO ($S \approx 100$ to 195 mg/mL). The addition of less polar DMSO to the solution has two principal effects on the binding affinity of LEN for the protein pocket. First, the

presence of DMSO increases the solubility of free LEN in the buffer solution, which as shown below leads to a weakening of binding by a factor equivalent to the free energy cost of transferring LEN from the buffer solution to neat water. Second, it is reasonable to expect that the preferential absorption (enrichment) of DMSO by the nonpolar protein binding pocket in the buffer solution increases the free energy cost of displacing the DMSO molecules by the LEN relative to that in water. Since absolute binding free energy is determined by the difference between the transfer free energy of the ligand into the binding pocket and that into solution (Figure 5), each of the above effects contributes to a weaker standard binding free energy of the LEN with the capsid protein, as shown below in a detailed semiquantitative thermodynamic analysis.

Estimate of the Difference of Insertion Free Energy of LEN in Pure Water and DMSO/Water Solution Using Experimental Solubilities and Solvation Free Energy Calculations

Consider saturated solutions of LEN in pure water (LEN(aq)) and in a 5% DMSO water mixture (LEN(aq,DMSO)) in equilibrium with solid LEN (LEN(s)) (Figure 6A). By definition, the concentration of LEN in each solution is the corresponding solubility. Assuming ideal solutions, the chemical potential $\mu(\text{L(aq)})$ of LEN can be written as

$$\mu(\text{L(aq)}) = \mu^\circ(\text{L(aq)}) + k_B T \ln \frac{S(\text{L(aq)})}{C^\circ} \quad (3)$$

where $\mu^\circ(\text{L(aq)})$ is the standard chemical potential and $S(\text{L(aq)})$ is the solubility of LEN in water. The standard state is an ideal solution at $C^\circ = 1$ M concentration. The expression of the chemical potential of LEN in the DMSO/water mixture is written analogously to eq 3, i.e.

$$\begin{aligned} \mu(\text{L(aq, DMSO)}) &= \mu^\circ(\text{L(aq, DMSO)}) \\ &+ k_B T \ln \frac{S(\text{L(aq, DMSO)})}{C^\circ} \end{aligned} \quad (3')$$

Because the chemical potential of LEN in each saturated solution is equal to the chemical potential of solid LEN, the chemical potentials of LEN in the two saturated solutions in Figure 6A are the same. Equating the two expressions eq 3 and eq 3', we arrive at the following expression for the difference of standard chemical potentials of LEN in the DMSO/water mixture and in pure water in terms of the respective solubilities:

$$\begin{aligned} \Delta\mu^\circ &= \mu^\circ(\text{L(aq, DMSO)}) - \mu^\circ(\text{L(aq)}) \\ &= -k_B T \ln \frac{S(\text{L(aq, DMSO)})}{S(\text{L(aq)})} \end{aligned} \quad (4)$$

The aqueous solubility of LEN has been reported as $S(\text{L(aq)}) = 0.31 \mu\text{g/mL}$ at room temperature.⁵⁰ Using the reported $320 \mu\text{M}$ LEN in 4.3% DMSO buffer²¹ as a proxy for solubility, i.e., $\approx 10^3$ -fold over water, eq 4 implies $\Delta\mu^\circ = -k_B T \ln \frac{S(\text{L(aq,DMSO)})}{S(\text{L(aq)})} = -4.1 \text{ kcal/mol}$ at 300 K.

Now consider the standard molar transfer free energy of LEN from the gas phase to water, $\Delta G^\circ_t(\text{L(aq)}) = \mu^\circ(\text{L(aq)}) - \mu^\circ(\text{L(g)})$, and to the DMSO/water mixture (Figure 6B). The difference between the transfer free energies is

$$\begin{aligned} \Delta\Delta G^\circ_t(\text{L(aq)} \rightarrow \text{L(aq, DMSO)}) &= \Delta G^\circ_t(\text{L(aq, DMSO)}) \\ &- \Delta G^\circ_t(\text{L(aq)}) = \mu^\circ(\text{L(aq, DMSO)}) - \mu^\circ(\text{L(aq)}) \end{aligned} \quad (5)$$

which is exactly eq 4. Hence this analysis shows that the transfer free energy of LEN from the gas phase to the DMSO/water solution is predicted to be -4.1 kcal/mol more favorable than the transfer to water.

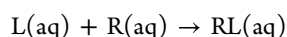
In addition to estimating the $\Delta\Delta G^\circ_t(\text{L(aq)} \rightarrow \text{L(aq,DMSO)})$ using experimental solubilities, we also calculated this quantity using free energy simulations by alchemically annihilating a ligand in DMSO/water and pure water and taking the difference of the free energy changes. Using GROMACS with GAFF parameters for DMSO and TIP3P for water, we found that the calculated solvation free energy shifts relative to water are

$$\begin{aligned} \Delta\Delta G^\circ_t(\text{L(aq)} \rightarrow \text{L(aq, DMSO)}) &= \Delta G^\circ_t(\text{L(aq, DMSO)}) \\ &- \Delta G^\circ_t(\text{L(aq)}) = -3.35 \pm 1.03 \text{ kcal/mol} \end{aligned}$$

at 5% DMSO and $-4.43 \pm 0.87 \text{ kcal/mol}$ at 10% DMSO. These results are consistent with the estimate from eq 4 using experimental solubilities at 5% DMSO ($\approx -4.1 \text{ kcal/mol}$) and show the expected concentration dependence (stronger stabilization at 10% than at 5%). The general agreement between the calculation and the solubility-based thermodynamic estimates provides quantitative evidence that the solvation free energy of LEN in DMSO/water is substantially more favorable than that in pure water, which as shown below contributes to the weaker binding affinity in the DMSO/water buffer.

Effect of DMSO on the Binding Affinity of LEN for the Receptor

The standard molar binding free energy $\Delta G^\circ_{\text{bind}}$ of a ligand L to a receptor R to form the complex RL in water



is the difference between the standard chemical potentials of the complex and the free receptor and ligand

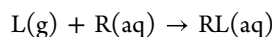
$$\Delta G^\circ_{\text{bind}}(\text{aq}) = \mu^\circ(\text{RL(aq)}) - \mu^\circ(\text{R(aq)}) - \mu^\circ(\text{L(aq)}) \quad (6)$$

Motivated by the DDM thermodynamic cycle (Figure 5), $\Delta G^\circ_{\text{bind}}$ can be expressed in terms of the free energies of transfer of the ligand to water and to the receptor binding site in water by adding and subtracting the chemical potential of the ligand in vacuum $\mu^\circ(\text{L(g)})$

$$\begin{aligned} \Delta G^\circ_{\text{bind}}(\text{aq}) &= [\mu^\circ(\text{RL(aq)}) - \mu^\circ(\text{R(aq)}) - \mu^\circ(\text{L(g)})] \\ &- [\mu^\circ(\text{L(aq)}) - \mu^\circ(\text{L(g)})] \end{aligned} \quad (7)$$

$$= \Delta G^\circ_t(\text{RL(aq)}) - \Delta G^\circ_t(\text{L(aq)}) \quad (8)$$

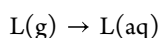
where $\Delta G^\circ_t(\text{RL(aq)})$ is the free energy of transfer of the ligand from the gas phase to the receptor binding site in water,



i.e.,

$$\Delta G^\circ_t(\text{RL(aq)}) = \mu^\circ(\text{RL(aq)}) - \mu^\circ(\text{R(aq)}) - \mu^\circ(\text{L(g)})$$

and $\Delta G^\circ_t(\text{L(aq)})$ is the free energy of transfer of the ligand from the gas phase to water:



with

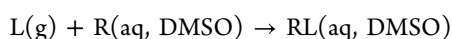
$$\Delta G_t^\circ(L(aq)) = \mu^\circ(L(aq)) - \mu^\circ(L(g))$$

Analogously, we can write the equation for the binding free energy for the DMSO/water mixture as

$$\begin{aligned} \Delta G_{bind}^\circ(aq, DMSO) &= [\mu^\circ(RL(aq, DMSO)) - \mu^\circ(R(aq, DMSO)) \\ &\quad - \mu^\circ(L(g))] - [\mu^\circ(L(aq, DMSO)) - \mu^\circ(L(g))] \end{aligned} \quad (7')$$

$$= \Delta G_t^\circ(RL(aq, DMSO)) - \Delta G_t^\circ(L(aq, DMSO)) \quad (8')$$

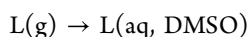
where $\Delta G_t^\circ(RL(aq, DMSO))$ is the free energy of transfer of the ligand from the gas phase to the receptor binding site in DMSO/water mixture:



with

$$\begin{aligned} \Delta G_t^\circ(RL(aq, DMSO)) &= \mu^\circ(RL(aq, DMSO)) - \mu^\circ(R(aq, DMSO)) \\ &\quad - \mu^\circ(L(g)) \end{aligned}$$

And $\Delta G_t^\circ(L(aq, DMSO))$ is the free energy of transfer of the ligand from the gas phase to DMSO/water mixture:



with

$$\Delta G_t^\circ(L(aq, DMSO)) = \mu^\circ(L(aq, DMSO)) - \mu^\circ(L(g))$$

Taking the difference between eq 8' and eq 8, we arrive at an expression for the difference between the standard binding free energy in the DMSO/water mixture and pure water:

$$\begin{aligned} \Delta \Delta G_{bind}^\circ(aq \rightarrow (aq, DMSO)) &= \Delta G_{bind}^\circ(aq, DMSO) - \Delta G_{bind}^\circ(aq) \end{aligned} \quad (9)$$

$$\begin{aligned} &= \Delta \Delta G_t^\circ(RL(aq) \rightarrow RL(aq, DMSO)) \\ &\quad - \Delta \Delta G_t^\circ(L(aq) \rightarrow L(aq, DMSO)) \end{aligned} \quad (10)$$

Equation 10 is the central equation for understanding the effects of adding DMSO on the binding affinity. The first term in eq 10, $\Delta \Delta G_t^\circ(RL(aq) \rightarrow RL(aq, DMSO))$, i.e., the “pocket leg” contribution in the thermodynamic cycle in Figure 7, is the difference between the free energy of transferring a gas-phase ligand into the protein pocket in the DMSO/water mixture and that in pure water, i.e.

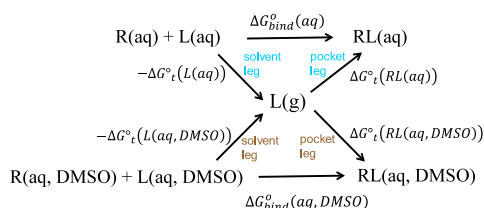


Figure 7. Thermodynamic cycle that connects ABFEs in an aqueous solution and DMSO/water mixture.

$$\begin{aligned} \Delta \Delta G_t^\circ(RL(aq) \rightarrow RL(aq, DMSO)) &= \Delta G_t^\circ(RL(aq, DMSO)) - \Delta G_t^\circ(RL(aq)) \end{aligned} \quad (10-1)$$

$$\begin{aligned} &= [\mu^\circ(RL(aq, DMSO)) - \mu^\circ(R(aq, DMSO)) \\ &\quad - \mu^\circ(L(g))] - [\mu^\circ(RL(aq)) - \mu^\circ(R(aq)) - \mu^\circ(L(g))] \end{aligned} \quad (10-2)$$

The second term in eq 10, $\Delta \Delta G_t^\circ(L(aq) \rightarrow L(aq, DMSO))$, i.e., the “solvent leg” contribution in Figure 7, is the difference between the free energy of transferring a gas-phase ligand into the DMSO/water and that into pure water,

$$\begin{aligned} \Delta \Delta G_t^\circ(L(aq) \rightarrow L(aq, DMSO)) &= [\mu^\circ(L(aq, DMSO)) - \mu^\circ(L(g))] - [\mu^\circ(L(aq)) \\ &\quad - \mu^\circ(L(g))] \end{aligned} \quad (10-3)$$

We have already shown that, in the case of LEN, the second term (“solvent-leg term”) in eq 10 $\Delta \Delta G_t^\circ(L(aq) \rightarrow L(aq, DMSO))$ is negative (≈ -4.1 kcal/mol using the experimental solubilities, or ≈ -3.35 kcal/mol using calculated solvation free energy difference; see eqs 4 and 5 and the discussions given there), which according to eq 10 contributes to making the binding free energy in DMSO about 3–4 kcal/mol less favorable than in pure water.

We next consider the first term (“pocket-leg term”) in eq 10, $\Delta \Delta G_t^\circ(RL(aq) \rightarrow RL(aq, DMSO))$. We propose that for a hydrophobic receptor pocket this term also likely makes the binding free energy in DMSO less favorable than in water. A thermodynamic-cycle analysis, presented in the Supporting Information, suggests that this pocket-leg contribution to $\Delta \Delta G_{bind}^\circ(aq \rightarrow (aq, DMSO))$ can be decomposed into a cavity-formation term, associated with displacing DMSO preferentially bound in the hydrophobic pocket, and a ligand–pocket interaction term that largely cancels between water and DMSO buffer. We hypothesize that the cavity term dominates and is positive, thereby disfavoring binding in the DMSO buffer relative to water and reinforcing the trend inferred from the solvent-leg contribution (see Supporting Information, eqs S5–S8 and Figure S6).

Our analysis on the plausible mechanisms by which DMSO affects binding affinity is consistent with previous reports that the addition of DMSO weakens apparent binding relative to DMSO-free aqueous buffer, even though those studies did not resolve the underlying physical origin. Senac et al.¹⁹ examined the effect of adding DMSO to the binding affinity in host–guest complexes of 1-adamantane carboxylic acid (ADA) with β - and γ -cyclodextrin and found that 5% DMSO reduced the affinity of ADA for β -cyclodextrin by ~ 2.1 kcal/mol, whereas the effect on the looser γ -cyclodextrin complex was smaller (~ 0.7 kcal/mol), indicating that DMSO weakens tight, hydrophobic cavities more strongly. Cubrilovic and Zenobi⁵¹ compared three protein–ligand complexes in 0–8% DMSO and observed that the largest affinity loss (≈ 10.5 -fold or 1.4 kcal/mol shift in ΔG_{bind}°) occurred for carbonic anhydrase binding chlorothiazide, which is the least water-soluble ligand in their set, while the more water-soluble ligands (Pefabloc and NAG3) showed smaller decreases. This trend of solubility dependence is consistent with one of the mechanisms we propose here: when a ligand is poorly soluble in neat water, adding a few percent of DMSO stabilizes the free ligand in the bulk solvent and shifts the binding equilibrium (see eq 10).

LEN is an extreme case with a molar mass 3–5 times the ligands in those studies, and the LEN binding pocket is also significantly more extensive. In addition, the SPR buffer employed in the measurement of LEN-CA binding affinity also contains 0.05% Tween-20 surfactant which further increases the LEN solubility, so it is reasonable to expect a much larger buffer-induced binding affinity decrease relative to neat water.

DISCUSSION

This work evaluates the performance of current absolute binding free energy (ABFE) protocols on a large hydrophobic ligand (lenacapavir, LEN) and substantially accounts for an initially very large discrepancy with the experiment through two physically distinct effects: ligand reorganization and the solvent reference state. Across diverse thermodynamic pathways and software code (standard DDM, Alchem-PMF, ATM, progressive DDM), the calculated values in neat water converge to $\Delta G^{\circ}_{bind} \approx -26.4$ to -30.0 kcal/mol, while SPR in 5% (v/v) DMSO reports ≈ -13.4 kcal·mol⁻¹ (Table 1). Control calculations indicate that changes in ligand protonation, receptor oligomeric form, and force field each shift ΔG°_{bind} by only ~ 1 kcal/mol, and SPR fitting artifacts were ruled out by explicit sensitivity analyses of k_{off} (Figure 4). These observations focus our attention on: (i) the ligand reorganization, since LEN undergoes extensive structural rearrangement upon binding that is challenging to sample in standard ABFE calculations due to its many torsional degrees of freedom, and (ii) the effect of buffer composition, as the experiments were carried out in a mixed solvent while the simulations used neat water.

Ligand reorganization indeed turns out to be an important contributor to the overbinding of the LEN. With 13 rotatable bonds, the LEN has substantial internal flexibility. A DDM variant that explicitly restrains and releases ligand torsions, following earlier work such as BAT2,¹⁵ reduces the overbinding by ~ 6 kcal/mol relative to standard DDM by avoiding the need to sample the large ligand conformational change during decoupling (Figure 2).

On the other hand, several ABFE protocols take a different approach by applying enhanced sampling during ligand decoupling while keeping the ligand conformation unrestrained. For example, Schrödinger's ABFEP⁴ and the recently developed protocol by Wu et al.¹⁴ employ REST or Hamiltonian replica exchange in λ space to accelerate exploration of the ligand's conformational ensemble across all λ windows, instead of applying torsional restraints. In principle, such enhanced-sampling schemes should improve convergence, but in practice their impact is system dependent: enlarging the accessible conformational space can also increase correlation times and, in some systems, even slow convergence.⁵² For LEN-CA, the restraint/release DDM variant yields a better estimate of ΔG°_{bind} than the current implementation of Schrödinger ABFEP, which relies on enhanced sampling to flatten the unrestrained ligand's conformational landscape (Table 1). However, it is difficult to draw firm conclusions from a single large ligand, particularly because the two protocols also use different force fields. It would therefore be informative in future work to compare these restrained versus enhanced-sampling strategies more systematically, using a common force field and a set of ligands spanning a range of sizes and flexibilities, to assess their relative ability to capture ligand-reorganization penalties.

Even with this improvement from the DDM with ligand reorganization, an ~ 9 kcal/mol gap to the SPR value remains, indicating that ligand reorganization alone cannot reconcile computation and experiment. Our thermodynamic analysis and solvation free energy calculations suggest that the solvent reference state provides the dominant missing term. SPR was executed in a 5% DMSO buffer. In such a buffer, two effects can systematically weaken binding relative to pure water: (1) bulk stabilization of free LEN in the DMSO/water buffer (higher solubility, lower chemical potential)^{53,54} and (2) the proposed cosolvent occupancy of the hydrophobic pocket, which likely raises the work to form the complex by displacing DMSO. A buffer-transfer analysis using experimental solubilities yields a bulk shift of about +4.1 kcal/mol at 300 K (eqs 4 and 5), consistent with the expected direction. We also calculated the solvation free energy of LEN in the 5% DMSO buffer and that in pure water using standard alchemical decoupling simulations, which gave $\Delta \Delta G^{\circ}_i(L(aq) \rightarrow L(aq,DMSO)) \approx -3.35$ kcal/mol, confirming the thermodynamic analysis based on experimental solubilities. The binding pocket contribution is also hypothesized to disfavor binding in the DMSO buffer, because DMSO likely enriches hydrophobic pockets and is expected to be more costly to displace on binding than bound water. Together, the quantitative estimate of the bulk stabilization and qualitatively hypothesized pocket effects help rationalize the residual ~ 9 kcal/mol overbinding after DDM with ligand reorganization and explain why water-reference ABFEs are not directly comparable to K_d measured in DMSO-containing buffers, and fully buffer-matched ABFE calculations would be required for rigorous validation.

Protein reorganization appears to be secondary for this system. Holo- versus apo-initiated calculations in both dimeric and hexameric capsid contexts differ by ~ 1 kcal/mol, suggesting limited net protein reorganization cost along the simulated pathway (Figure 3). Some distal allosteric shifts are still visible upon decoupling; they may contribute to the remaining discrepancy.

Lastly, the possibility that force field limitations also contribute to the remaining discrepancy still exists. As seen from Figure 5, the unbound ligand in solution features extensive intramolecular π - π interactions, whereas such interactions are not involved in the intermolecular recognition between the LEN and binding pocket (structure not shown). It was reported that classical force fields tend to underestimate the strengths of π - π interactions (Sherrill et al.⁵⁵). It is therefore conceivable that our ABFE simulations using the standard force fields underestimated the thermodynamic stability of the unbound state, thereby shifting the binding equilibrium to favor the bound state and contributing to the remaining overestimation.

IMPLICATIONS FOR LIGAND SCREENING IN DRUG DISCOVERY

Although we find that adding only 5% DMSO can shift the apparent binding free energy of LEN by many kilocalories per mole, we do not conclude that the widespread use of DMSO-containing buffers in high-throughput screening is inherently problematic. In most ligand screening and SAR campaigns, the main goal is to obtain *relative* affinities or activities across a series of chemically related compounds under the same assay conditions. In that context, a substantial part of the DMSO effect (bulk solubility enhancement plus preferential pocket adsorption) is expected to be similar across analogs within a

congeneric series, so that much of the cosolvent contribution cancels when considering binding free energy differences among ligands. Our work highlighted that cosolvents can strongly perturb *absolute* binding free energies, particularly for very hydrophobic ligands like LEN, and that this must be taken into account when one aims at quantitative comparison between computation and experiment.

In addition, we do not regard 5% DMSO as a realistic model of the physiological environment experienced by a drug molecule. The intracellular milieu is a crowded, compositionally complex mixture of proteins, nucleic acids, lipids, metabolites, and other components and does not contain DMSO at appreciable concentrations. Instead, DMSO is introduced into *in vitro* assays primarily as a technical cosolvent to increase the solubility and enable reliable measurements. Our results do not argue that DMSO is more “physiological” than water but rather use DMSO as a concrete example to demonstrate that any cosolvent or molecular crowder can have a large thermodynamic impact on binding and that the effective *in vivo* recognition landscape can differ substantially from that inferred from simple aqueous or DMSO-based *in vitro* conditions.

LIMITATIONS AND FUTURE WORK

In this study, we estimated the buffer effect semiquantitatively using solubility-based thermodynamic arguments, solvation free energy calculations, and qualitative pocket-solvation reasoning. A rigorous treatment mimicking experimental buffer conditions requires running ABFE calculations in a DMSO/water buffer for a system with a large, topologically complex binding site: the LEN pocket comprises three connected subpockets, two of which are deeply buried. Mixed solvents can exhibit slow exchange between pocket and bulk (DMSO has a larger excluded volume than water), so achieving convergence will likely require advanced sampling techniques^{56,57} through insertion–deletion of water and DMSO cosolvent targeting pocket solvent to accelerate equilibration. In addition, the presence of surfactants (e.g., 0.05% Tween-20 in the SPR buffer) above its critical micelle concentration (CMC) further complicates the bulk reference state through the potential micelle-assisted solvation of LEN. A rigorous treatment of these multicomponent equilibria is challenging and beyond the scope of the present work. Such binding free energy calculations that mimic the experimental buffer conditions could be an important direction for future studies.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpcb.5c06714>.

Procedure for SPR curve fitting, figures of sample response curve vs time, sequence of RBEF (relative binding free energy) steps used in the ATM calculation of the ABFE for LEN, buried ligand atoms of LEN that are alchemically decoupled from the environment in Alchem-PMF for computing the ABFE of LEN, sequence of stepwise alchemical removal of functional groups from LEN in the progressive DDM calculation of the ABFE of LEN, convergence in the ΔG_{bind}° calculated using different ABFE methods employed in this study, thermodynamic cycle analysis of the sign of the first term in eq 10, $\Delta\Delta G_{i}^{\circ}(\text{RL}(\text{aq})\rightarrow\text{RL}(\text{aq,DMSO}))$, i.e., the

“pocket-leg” contribution to binding free energy difference in the two buffers, conformational ensembles of the decoupled ligand in both the solvent leg and the complex leg in the standard GROMACS DDM simulation: comparisons of the sampled distributions of selected dihedral angles (PDF)

AUTHOR INFORMATION

Corresponding Author

Nanjie Deng – Department of Chemistry and Physical Sciences, Pace University, New York, New York 10038, United States; orcid.org/0000-0001-8805-3526; Email: nanjie.deng@gmail.com

Authors

Qinfang Sun – Center for Biophysics and Computational Biology and Department of Chemistry, Temple University, Philadelphia, Pennsylvania 19122, United States

Emilio Gallicchio – Department of Chemistry, Brooklyn College, the City University of New York, Brooklyn, New York 11210, United States; orcid.org/0000-0002-2606-4913

Ronald Levy – Center for Biophysics and Computational Biology and Department of Chemistry, Temple University, Philadelphia, Pennsylvania 19122, United States; orcid.org/0000-0001-8696-5177

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jpcb.5c06714>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Drs. Chung Wong, Bin Zhang and R. S. K. Vijayan for helpful discussions. The calculations were run on the ACCESS allocation resource TGMCB100145 and a shared computing cluster at Temple University supported by National Institutes of Health S10 OD020095. This work was supported in part by the National Institutes of Health, grants NIH U54AI170855 subcontract and NIH R01 AI178849. E.G. acknowledges support from the National Institute of General Medical Sciences (NIH 1R15GM151708). The authors are pleased to submit this article to the special issue of *The Journal of Physical Chemistry* in memory of Peter Kollman. Peter was one of the first to use molecular dynamics simulations in the 1980s to study protein–ligand binding, and an early developer of free energy simulations for that purpose. R.L. remembers with fondness many discussions with Peter Kollman about the subject, some of which took place over delicious dinners in Paris.

REFERENCES

- (1) Kollman, P. A.; et al. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000**, *33*, 889–97.
- (2) Kollman, P. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417.
- (3) Gallicchio, E.; Levy, R. M. Recent Theoretical and Computational Advances for Modeling Protein-Ligand Binding Affinities. *Adv. Protein Chem. Struct. Biol.* **2011**, *85*, 27–80.
- (4) Chen, W.; et al. Enhancing Hit Discovery in Virtual Screening through Absolute Protein-Ligand Binding Free-Energy Calculations. *J. Chem. Inf. Model.* **2023**, *63*, 3171–3185.

- (5) Wang, C.; Nguyen, P. H.; Pham, K.; Huynh, D.; Le, T. B.; Wang, H.; Ren, P.; Luo, R. Calculating Protein-Ligand Binding Affinities with Mmpbsa: Method and Error Analysis. *J. Comput. Chem.* **2016**, *37*, 2436–46.
- (6) Minh, D. D. L.; Cooper, D. A.; Xie, B.; Shi, L. Ligand Reorganization for End-Point Binding Free Energy Calculations: Identifying Preferred Poses of Fentanyl in the Mu Opioid Receptor. *J. Chem. Theory Comput.* **2025**, *21*, 943–950.
- (7) Fu, H.; Shao, X.; Cai, W.; Chipot, C. Taming Rugged Free Energy Landscapes Using an Average Force. *Acc. Chem. Res.* **2019**, *52*, 3254–3264.
- (8) Deng, N. J.; Zhang, P.; Cieplak, P.; Lai, L. Elucidating the Energetics of Entropically Driven Protein-Ligand Association: Calculations of Absolute Binding Free Energy and Entropy. *J. Phys. Chem. B* **2011**, *115*, 11902–10.
- (9) Aldeghi, M.; Heifetz, A.; Bodkin, M. J.; Knapp, S.; Biggin, P. C. Accurate Calculation of the Absolute Free Energy of Binding for Drug Molecules. *Chem. Sci.* **2016**, *7*, 207–218.
- (10) Deng, N.; Cui, D.; Zhang, B. W.; Xia, J.; Cruz, J.; Levy, R. Comparing Alchemical and Physical Pathway Methods for Computing the Absolute Binding Free Energy of Charged Ligands. *Phys. Chem. Chem. Phys.* **2018**, *20*, 17081–17092.
- (11) Karrenbrock, M.; Borsatto, A.; Rizzi, V.; Lukauskis, D.; Aureli, S.; Luigi Gervasio, F. Absolute Binding Free Energies with Oneopes. *J. Phys. Chem. Lett.* **2024**, *15*, 9871–9880.
- (12) Khalak, Y.; Tresadern, G.; Aldeghi, M.; Baumann, H. M.; Mobley, D. L.; de Groot, B. L.; Gapsys, V. Alchemical Absolute Protein-Ligand Binding Free Energies for Drug Design. *Chem. Sci.* **2021**, *12*, 13958–13971.
- (13) 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.
- (14) Wu, Z.; Konig, G.; Boresch, S.; Cossins, B. P. Optimizing Absolute Binding Free Energy Calculations for Production Usage. *J. Chem. Theory Comput.* **2025**, *21*, 8330–8340.
- (15) Heinzelmann, G.; Huggins, D. J.; Gilson, M. K. Bat2: An Open-Source Tool for Flexible, Automated, and Low Cost Absolute Binding Free Energy Calculations. *J. Chem. Theory Comput.* **2024**, *20*, 6518–6530.
- (16) Clark, F.; Robb, G. R.; Cole, D. J.; Michel, J. Automated Adaptive Absolute Binding Free Energy Calculations. *J. Chem. Theory Comput.* **2024**, *20*, 7806–7828.
- (17) Egbert, M.; Whitty, A.; Keseru, G. M.; Vajda, S. Why Some Targets Benefit from Beyond Rule of Five Drugs. *J. Med. Chem.* **2019**, *62*, 10005–10025.
- (18) Doak, B. C.; Over, B.; Giordanetto, F.; Kihlberg, J. Oral Druggable Space Beyond the Rule of 5: Insights from Drugs and Clinical Candidates. *Chem. Biol.* **2014**, *21*, 1115–42.
- (19) Senac, C.; Desgranges, S.; Contino-Pepin, C.; Urbach, W.; Fuchs, P. F. J.; Taulier, N. Effect of Dimethyl Sulfoxide on the Binding of 1-Adamantane Carboxylic Acid to Beta- and Gamma-Cyclodextrins. *ACS Omega* **2018**, *3*, 1014–1021.
- (20) Link, J. O.; et al. Clinical Targeting of Hiv Capsid Protein with a Long-Acting Small Molecule. *Nature* **2020**, *584*, 614–618.
- (21) Bester, S. M.; et al. Structural and Mechanistic Bases for a Potent Hiv-1 Capsid Inhibitor. *Science* **2020**, *370*, 360–364.
- (22) Bester, S. M.; Adu-Ampratwum, D.; Annamalai, A. S.; Wei, G.; Briganti, L.; Murphy, B. C.; Haney, R.; Fuchs, J. R.; Kvaratskhelia, M. Structural and Mechanistic Bases of Viral Resistance to Hiv-1 Capsid Inhibitor Lenacapavir. *mBio* **2022**, *13*, No. e0180422.
- (23) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The Statistical-Thermodynamic Basis for Computation of Binding Affinities: A Critical Review. *Biophys. J.* **1997**, *72*, 1047–69.
- (24) Boresch, S.; Tettering, F.; Leitgeb, M.; Karplus, M. Absolute Binding Free Energies: A Quantitative Approach for Their Calculation. *J. Phys. Chem. B* **2003**, *107*, 9535–9551.
- (25) Azimi, S.; Gallicchio, E. Potential Distribution Theory of Alchemical Transfer. *J. Chem. Phys.* **2025**, *162*, 054106.
- (26) Gallicchio, E. Relative Binding Free Energy Estimation of Congeneric Ligands and Macromolecular Mutants with the Alchemical Transfer Method with Coordinate Swapping. *J. Chem. Inf. Model.* **2025**, *65*, 3706–3714.
- (27) Khuttan, S.; Azimi, S.; Wu, J. Z.; Dick, S.; Wu, C.; Xu, H.; Gallicchio, E. Taming Multiple Binding Poses in Alchemical Binding Free Energy Prediction: The Beta-Cyclodextrin Host-Guest Sampl9 Blinded Challenge. *Phys. Chem. Chem. Phys.* **2023**, *25*, 24364–24376.
- (28) Azimi, S.; Wu, J. Z.; Khuttan, S.; Kurtzman, T.; Deng, N.; Gallicchio, E. Application of the Alchemical Transfer and Potential of Mean Force Methods to the Sampl8 Host-Guest Blinded Challenge. *J. Comput. Aided Mol. Des.* **2022**, *36*, 63–76.
- (29) Wu, J. Z.; Azimi, S.; Khuttan, S.; Deng, N.; Gallicchio, E. Alchemical Transfer Approach to Absolute Binding Free Energy Estimation. *J. Chem. Theory Comput.* **2021**, *17*, 3309–3319.
- (30) Cruz, J.; Wickstrom, L.; Yang, D.; Gallicchio, E.; Deng, N. Combining Alchemical Transformation with a Physical Pathway to Accelerate Absolute Binding Free Energy Calculations of Charged Ligands to Enclosed Binding Sites. *J. Chem. Theory Comput.* **2020**, *16*, 2803–2813.
- (31) Lu, C.; et al. Opls4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space. *J. Chem. Theory Comput.* **2021**, *17*, 4291–4300.
- (32) Ross, G. A.; Russell, E.; Deng, Y.; Lu, C.; Harder, E. D.; Abel, R.; Wang, L. Enhancing Water Sampling in Free Energy Calculations with Grand Canonical Monte Carlo. *J. Chem. Theory Comput.* **2020**, *16*, 6061–6076.
- (33) Wang, L.; Friesner, R. A.; Berne, B. J. Correction to “Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (Rest2)”. *J. Phys. Chem. B* **2011**, *115*, 11305.
- (34) Shirts, M. R.; Chodera, J. D. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *J. Chem. Phys.* **2008**, *129*, 124105.
- (35) 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*, 1950–8.
- (36) 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*, 1157–74.
- (37) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. Am1-Bcc Model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132–146.
- (38) Boresch, S. On Analytical Corrections for Restraints in Absolute Binding Free Energy Calculations. *J. Chem. Inf. Model.* **2024**, *64*, 3605–3609.
- (39) 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.
- (40) Liu, R.; Yao, Y.; Huang, W.; Zhong, Y.; Luo, H. B.; Li, Z. Divide-and-Conquer Abfe: Improving Free Energy Calculations by Enhancing Water Sampling. *J. Chem. Theory Comput.* **2025**, *21*, 3712–3725.
- (41) Li, M.; et al. Mechanisms of Hiv-1 Integrase Resistance to Dolutegravir and Potent Inhibition of Drug-Resistant Variants. *Sci. Adv.* **2023**, *9*, No. eadg5953.
- (42) Bennett, C. H. Efficient Estimation of Free Energy Differences from Monte Carlo Data. *J. Comput. Phys.* **1976**, *22*, 245–268.
- (43) Wang, L.; Deng, Y.; Knight, J. L.; Wu, Y.; Kim, B.; Sherman, W.; Shelley, J. C.; Lin, T.; Abel, R. Modeling Local Structural Rearrangements Using Fep/Rest: Application to Relative Binding Affinity Predictions of Cdk2 Inhibitors. *J. Chem. Theory Comput.* **2013**, *9*, 1282–93.
- (44) Azimi, S.; Khuttan, S.; Wu, J. Z.; Pal, R. K.; Gallicchio, E. Relative Binding Free Energy Calculations for Ligands with Diverse Scaffolds with the Alchemical Transfer Method. *J. Chem. Inf. Model.* **2022**, *62*, 309–323.

(45) Azimi, S.; Wu, J. Z.; Khuttan, S.; Kurtzman, T.; Deng, N.; Gallicchio, E. Application of the Alchemical Transfer and Potential of Mean Force Methods to the Sampl8 Host-Guest Blinded Challenge. *Journal of Computer-Aided Molecular Design* **2022**, *36*, 63–76.

(46) Gallicchio, E.; Deng, N.; He, P.; Wickstrom, L.; Perryman, A. L.; Santiago, D. N.; Forli, S.; Olson, A. J.; Levy, R. M. Virtual Screening of Integrase Inhibitors by Large Scale Binding Free Energy Calculations: The Sampl4 Challenge. *J. Comput. Aided Mol. Des* **2014**, *28*, 475–90.

(47) Gizzio, J.; Thakur, A.; Haldane, A.; Post, C. B.; Levy, R. M. Evolutionary Sequence and Structural Basis for the Distinct Conformational Landscapes of Tyr and Ser/Thr Kinases. *Nat. Commun.* **2024**, *15*, 6545.

(48) Gizzio, J.; Thakur, A.; Haldane, A.; Levy, R. M. Evolutionary Divergence in the Conformational Landscapes of Tyrosine Vs Serine/Threonine Kinases. *Elife* **2022**, *11*, e83368.

(49) Fajer, M.; Borrelli, K.; Abel, R.; Wang, L. Quantitatively Accounting for Protein Reorganization in Computer-Aided Drug Design. *J. Chem. Theory Comput* **2023**, *19*, 3080–3090.

(50) https://www.gilead.com/en-ca/-/media/gilead-canada/pdfs/science/sunlenca_pm_en.pdf.

(51) Cubrilovic, D.; Zenobi, R. Influence of Dimethylsulfoxide on Protein-Ligand Binding Affinities. *Anal. Chem.* **2013**, *85*, 2724–30.

(52) Khuttan, S.; Gallicchio, E. What to Make of Zero: Resolving the Statistical Noise from Conformational Reorganization in Alchemical Binding Free Energy Estimates with Metadynamics Sampling. *J. Chem. Theory Comput* **2024**, *20*, 1489–1501.

(53) Hong, R. S.; et al. Free Energy Perturbation Approach for Accurate Crystalline Aqueous Solubility Predictions. *J. Med. Chem.* **2023**, *66*, 15883–15893.

(54) Mondal, S.; et al. A Free Energy Perturbation Approach to Estimate the Intrinsic Solubilities of Drug-Like Small Molecules. *ChemRxiv* **2019**, DOI: 10.26434/chemrxiv.10263077.v1.

(55) Sherrill, C. D.; Sumpter, B. G.; Sinnokrot, M. O.; Marshall, M. S.; Hohenstein, E. G.; Walker, R. C.; Gould, I. R. Assessment of Standard Force Field Models against High-Quality Ab Initio Potential Curves for Prototypes of Pi-Pi, Ch/Pi, and Sh/Pi Interactions. *J. Comput. Chem.* **2009**, *30*, 2187–93.

(56) Deng, Y.; Roux, B. Computation of Binding Free Energy with Molecular Dynamics and Grand Canonical Monte Carlo Simulations. *J. Chem. Phys.* **2008**, *128*, 115103.

(57) Ben-Shalom, I. Y.; Lin, C.; Kurtzman, T.; Walker, R. C.; Gilson, M. K. Simulating Water Exchange to Buried Binding Sites. *J. Chem. Theory Comput* **2019**, *15*, 2684–2691.



CAS BIOFINDER DISCOVERY PLATFORM™

**PRECISION DATA
FOR FASTER
DRUG
DISCOVERY**

CAS BioFinder helps you identify
targets, biomarkers, and pathways

Unlock insights

CAS
A Division of the
American Chemical Society