



# Assessment of a Single Decoupling Alchemical Approach for the Calculation of the Absolute Binding Free Energies of Protein-Peptide Complexes

Denise Kilburg<sup>1,2</sup> and Emilio Gallicchio<sup>1,2,3\*</sup>

<sup>1</sup> Department of Chemistry, Brooklyn College, Brooklyn, NY, United States, <sup>2</sup> Ph.D. Program in Chemistry, The Graduate Center, City University of New York, New York, NY, United States, <sup>3</sup> Ph.D. Program in Biochemistry, The Graduate Center, City University of New York, New York, NY, United States

## OPEN ACCESS

### Edited by:

Emil Alexov,  
Clemson University, United States

### Reviewed by:

Nobuyuki Matubayasi,  
Osaka University, Japan  
Jana Shen,  
University of Maryland, Baltimore,  
United States

### \*Correspondence:

Emilio Gallicchio  
egallicchio@brooklyn.cuny.edu

### Specialty section:

This article was submitted to  
Molecular Recognition,  
a section of the journal  
Frontiers in Molecular Biosciences

**Received:** 05 January 2018

**Accepted:** 21 February 2018

**Published:** 08 March 2018

### Citation:

Kilburg D and Gallicchio E (2018)  
Assessment of a Single Decoupling  
Alchemical Approach for the  
Calculation of the Absolute Binding  
Free Energies of Protein-Peptide  
Complexes. *Front. Mol. Biosci.* 5:22.  
doi: 10.3389/fmolb.2018.00022

The computational modeling of peptide inhibitors to target protein-protein binding interfaces is growing in interest as these are often too large, too shallow, and too feature-less for conventional small molecule compounds. Here, we present a rare successful application of an alchemical binding free energy method for the calculation of converged absolute binding free energies of a series of protein-peptide complexes. Specifically, we report the binding free energies of a series of cyclic peptides derived from the LEDGF/p75 protein to the integrase receptor of the HIV1 virus. The simulations recapitulate the effect of mutations relative to the wild-type binding motif of LEDGF/p75, providing structural, energetic and dynamical interpretations of the observed trends. The equilibration and convergence of the calculations are carefully analyzed. Convergence is aided by the adoption of a single-decoupling alchemical approach with implicit solvation, which circumvents the convergence difficulties of conventional double-decoupling protocols. We hereby present the single-decoupling methodology and critically evaluate its advantages and limitations. We also discuss some of the challenges and potential pitfalls of binding free energy calculations for complex molecular systems which have generally limited their applicability to the quantitative study of protein-peptide binding equilibria.

**Keywords:** protein-peptide complexes, binding free energy, alchemical calculations, single-decoupling, implicit solvent, LEDGF/p75, HIV integrase

## 1. INTRODUCTION

Protein-protein interactions are pervasive in biological systems as they drive and regulate critical functions within the cell (Kastritis and Bonvin, 2013). Concomitantly, there is a strong interest in developing therapeutic drugs targeting protein-protein interactions (Higuero et al., 2009; Basse et al., 2013; Labb et al., 2013; Arkin et al., 2014). The rationale of using small-molecules or peptides to influence protein-protein binding rests on the observation that, even though they may involve large and complex molecular assemblies consisting of thousands of atoms, often protein-protein interactions are commanded by a relatively small number of interfacial amino acid residues that are part of short linear recognition motifs (London et al., 2010). Standard drug design approaches employed

for the development of small molecule inhibitors have been found generally insufficient compared to peptide constructs in disrupting protein-protein interactions. Peptides, being larger than most small molecule inhibitors, can target large and shallow binding sites with a small ratio of binding energy per atom (Kastritis and Bonvin, 2013), which are often categorized as “undruggable” by standard definition.

A large variety of approaches are available to model small molecule-protein interactions. These range from ligand-based cheminformatic and pharmacophore models (Lavecchia and Di Giovanni, 2013; Yan et al., 2016), to structure-based docking and scoring (Ewing et al., 2001; Gray et al., 2003; Verdonk et al., 2003; Friesner et al., 2004; Kozakov et al., 2006; Zhou et al., 2007; Perryman et al., 2014; Pierce et al., 2014) and physics-based atomistic models (Gilson et al., 1997; Wang et al., 2015; Bell et al., 2016; Ellis et al., 2016; Zuckerman and Chong, 2017). Which are the focus of this work. The ultimate goal of physics-based computational models of binding is the quantitative prediction of equilibrium constants of binding (or, equivalently, standard free energies of binding) from statistical mechanics principles and models of interatomic interactions (Chang et al., 2007; Gallicchio and Levy, 2011). Pathway-based models of binding measure the free energy changes along a thermodynamic path linking the unbound and bound states of the molecular complex. These take the form of physical pathways in which ligand and receptor are brought together along a spatial coordinate (Hénin et al., 2005; Gumbart et al., 2013; Jo et al., 2015; Lapelosa, 2017), as well as pathways in so-called alchemical space in which ligand-receptor interactions are progressively dialed-in (Chodera et al., 2011; Deng et al., 2017).

Models targeting protein-peptide binding are generally not as established (Kilburg and Gallicchio, 2016). Bioinformatic and knowledge-based approaches based, for example, on sequence and structure-based homology and phylogenetic profiles, are commonly employed to search for likely interacting protein pairs (Shoemaker and Panchenko, 2007; Zhang et al., 2012). The ability of predicting binding affinities of protein-peptide interactions from structural models is a crucial challenge that would enhance our understanding of biological regulatory systems and be highly beneficial for drug design and development. Even though estimating binding free energies from physical principles is a daunting computational problem, worthwhile attempts should be made to solve it. A main obstacle is that the size and flexibility of peptides make them difficult systems to model (You et al., 2016). Following the work of characterizing physical pathways approaches for this problem (Hénin et al., 2005; Gumbart et al., 2013; Jo et al., 2015; Kilburg and Gallicchio, 2016; Lapelosa, 2017), the purpose of this work is to assess the applicability of alchemical methods for the estimation of protein-peptide binding free energies.

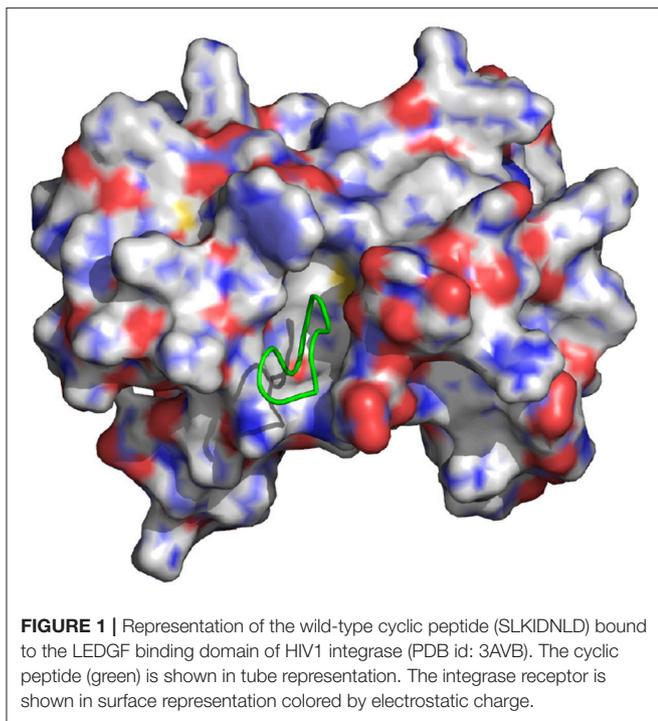
The double-decoupling method (Gilson et al., 1997), is the leading approach for the alchemical calculation of absolute binding free energies from first principles. It is based on a thermodynamic cycle involving the free energies of decoupling the ligand from the solution with and without the presence of the receptor. One of the difficulties of this approach is that the binding free energy, which in reversible binding processes is relatively small and weakly dependent on ligand size, is

obtained as the difference of two much larger values which grow with ligand size and ligand charge and whose statistical and systematic errors combine additively (Deng and Roux, 2009). For example, for a protein-peptide complex, the double-decoupling approach would require the challenging calculation of the large solvation free energy of the peptide within an uncertainty small in comparison to the binding free energy (ideally a fraction of a kilocalorie per mole). Furthermore, the free energy of decoupling from the receptor environment, which arguably involves much slower and complex reorganization processes than the decoupling from the solution, would have to be converged to the same degree (Gumbart et al., 2012).

To begin to address these challenges, in this work we assess the applicability of a single-decoupling alchemical method (Gallicchio et al., 2010) (referred to hereafter as “SDM” for short) to the calculation of protein-peptide binding free energies. As outlined below, the method employs a  $\lambda$ -dependent potential energy function which interpolates between the dissociated and associated states of the complex in solution, avoiding the intermediate gas-phase state required by the double-decoupling method. Because it computes the binding free energy directly, the method is expected to exhibit favorable convergence properties sufficient to handle challenging protein-peptide systems. Indeed the method has been already tested with relatively large ligands including some to the same receptor site (discussed below) which is the object of this work (Lapelosa et al., 2012; Gallicchio et al., 2014).

HIV-1 is a RNA retrovirus that must integrate a reverse transcribed copy of its RNA genome into host DNA to permanently infect a cell (Smith and Daniel, 2006). Essential viral catalytic enzymes include reverse transcriptase (RT) and integrase (IN). The former converts the viral RNA genome into double stranded cDNA. Once inside the nucleus, IN catalyzes the insertion of viral cDNA into the host chromosome (Cherepanov et al., 2005a). Integrase catalyzes this reaction in two ways: (1) it removes the 3' terminal GT nucleotides from both ends of the viral cDNA and (2) inserts the newly processed 3' termini into host chromosomal DNA (Cherepanov et al., 2005b). However, HIV1-IN itself cannot associate with chromatin and requires an endogenous transcriptional coactivator, p75, more commonly known as LEDGF (lens epithelium-derived growth factor) (Cherepanov et al., 2005b). In recent years, several small molecule inhibitors of the HIV1-IN interaction with LEDGF have been designed by focusing on interactions with selected residues of LEDGF, most notably Ile365, Asp366, and Leu368 (Cherepanov et al., 2005b; Rhodes et al., 2011). However, known mutations of HIV1-integrase have shown high resistance to these inhibitors, prompting further study into other potential interactions between LEDGF and integrase (Rhodes et al., 2011).

An extensive study by Tsiang et al. (2009) reported the binding affinities of native LEDGF to HIV1-IN as well as many LEDGF derived peptides, including both linear and cyclized motifs. One of the goals of their study was to find the smallest LEDGF-peptide sequence capable of inhibiting integrase. They found that small peptides are indeed capable of displacing LEDGF from HIV1-IN and concluded that cyclizing the peptides resulted in greater potency. Building on this, Rhodes et al. (2011) studied the interactions between HIV1-IN and multiple linear and



cyclic peptides derived from the SLKIDNLD (residues 362-369) binding motif of LEDGF (**Figure 1**). This motif is located on a loop region, between the  $\alpha 1$  and  $\alpha 2$  helices of LEDGF, which is considered the main interaction site (Cherepanov et al., 2005b). Rhodes et al. found that, in agreement with previous studies (Tsiang et al., 2009), the linear peptide, H-SLKIDNLD-OH, gave no activity up to 1 mM, while a cyclized version of the same sequence produced an  $IC_{50}$  value of 70.0  $\mu M$  using a strand transfer inhibition assay.

In this work we conduct alchemical single-decoupling calculations of the binding free energies of a set of LEDGF-derived peptides to the HIV1-IN receptor. The results recapitulate the effect of mutations relative to the wild-type binding motif of LEDGF, particularly those at positions that have been found to be critical for binding, and they provide structural, energetic, and dynamical interpretations of the observed binding affinities. The equilibration of the systems and the convergence of the results are carefully analyzed. The method is found to yield reliable binding free energies of cyclic peptides while it fails to produce a converged estimate for a linear peptide. We discuss the SDM methodology and the procedure of choosing appropriate simulation parameters. We also discuss some of the challenges and potential pitfalls of binding free energy calculations of complex molecular systems which have generally limited their applicability to the quantitative study of protein-peptide binding equilibria.

## 2. METHODS

### 2.1. System Preparation

The LEDGF binding site construct of HIV integrase was prepared as previously reported (Gallicchio et al., 2014). Water molecules,

bound ligands, and crystallization ions were removed and protein sidechain protonation states were assigned assuming pH 7 and absence of significant pKa shifts (Glu and Asp deprotonated and Lys and Arg protonated). A key residue in the LEDGF binding site, His171, was protonated at the  $N\delta$  position as previously investigated (Gallicchio et al., 2014). The cyclic SLKIDNLD peptide was adapted from the crystal structure of HIV integrase from which it was bound, 3AVB, as formerly reported (Rhodes et al., 2011). Four cyclic mutants (ALKIDNLD, ALKIDNMD, SLKINNLD, and SLKADNLD) in addition with a linear peptide H-SLKIDNLD-OH were constructed by modifying the cyclic SLKIDNLD peptide using the Maestro program (Schrödinger, LLC).

### 2.2. Single-Decoupling Binding Free Energy Protocol

Absolute binding free energies of the HIV-IN with the cyclic and linear H-SLKIDNLD-OH peptide complexes were computed with implicit solvation using the Single Decoupling (SDM) (Gallicchio et al., 2010) version of Double Decoupling Method (DDM) (Gilson et al., 1997). SDM was originally introduced as the Binding Energy Distribution Analysis Method (BEDAM) (Gallicchio et al., 2010) when it was based on the Weighted Analysis Histogram Method (WHAM) (Kumar et al., 1992). Recent applications of SDM have been based on multi-state binless free energy inference methods (Shirts and Chodera, 2008; Tan et al., 2012). Similarly to double-decoupling, in SDM the standard binding free energy  $\Delta G_b^\circ$  between a receptor R and the ligand L is expressed as:

$$\Delta G_b^\circ = -k_B T \ln C^\circ V_{\text{site}} + \Delta G_b \quad (1)$$

where  $C^\circ$  is the standard concentration of ligand molecules (1M or  $1,668 \text{ \AA}^{-3}$ ),  $V_{\text{site}}$  is the volume of the binding site (see below), and  $\Delta G_b$  is the excess binding free energy, defined as the free energy difference between the coupled state, in which the receptor and ligand are fully interacting, and the uncoupled state, in which the receptor and ligand are only interacting with the solvent and not with each other. In both the coupled and uncoupled states the ligand is sequestered in the binding site region as defined below.

Unlike the Double-Decoupling Method (DDM), which requires two alchemical calculations going through an intermediate “vacuum” state (Gilson et al., 1997; Boresch et al., 2003), SDM employs a  $\lambda$ /temperature-dependent reduced effective potential energy function which allows for a direct alchemical thermodynamic path between the uncoupled and coupled states of the complex, as follows:

$$U_\lambda(\mathbf{r}) = \beta[U_0(\mathbf{r}) + \lambda u(\mathbf{r})], \quad (2)$$

where  $\beta = 1/k_B T$ ,

$$U_0(\mathbf{r}) = U_R(\mathbf{r}_R) + U_L(\mathbf{r}_L) \quad (3)$$

is the effective potential energy function corresponding to the uncoupled state, and

$$u(\mathbf{r}_R, \mathbf{r}_L) = U(\mathbf{r}_R, \mathbf{r}_L) - U(\mathbf{r}_R) - U(\mathbf{r}_L) \quad (4)$$

is the binding energy function defined as the effective potential energy difference between the coupled and uncoupled states of the complex in conformation  $r = (r_r, r_L)$ , obtained by translating the ligand from the solvent and into the receptor without changing the internal coordinates of either molecule. It is straightforward to show that the  $\lambda$ -dependent potential energy function in Equation (2) linearly interpolates between the uncoupled and coupled states of the complex as  $\lambda$  is varied from 0 to 1.

In lieu of simulating each  $\lambda$  and/or temperature state independently and to improve conformational sampling efficiency, with SDM we utilize a Hamiltonian replica exchange (HREM)  $\lambda$ /temperature-hopping approach where  $\lambda$  and temperature values are swapped periodically according to a Monte Carlo procedure with acceptable exchanges adhering to the Metropolis algorithm (Pal et al., 2016).

To further improve convergence of the free energy at values around  $\lambda = 0$ , we apply in this work a “soft core” binding energy function:

$$u'(r) = \begin{cases} u_{\max} \tanh\left(\frac{u(r)}{u_{\max}}\right) & \text{if } u(r) > 0 \\ u(r) & \text{if } u(r) \leq 0 \end{cases} \quad (5)$$

where  $u_{\max}$  is some large positive value compared to the thermal energy. In this work we have set  $u_{\max} = 1000$  kcal/mol. The energy function modeled in Equation (5) replaces the binding energy function in Equation (4), when applicable, by limiting the maximum value of the binding energy for very small  $\lambda$  values while leaving the values of the favorable binding energies unchanged. The resulting data is subsequently processed using UWHAM analysis (Tan et al., 2012) to calculate the binding free energy of the complex.

The reorganization binding free energy is calculated by taking the difference between the binding free energy and the average interaction energy as:

$$\Delta G_{\text{reorg}}^{\circ} = \Delta G_b^{\circ} - \Delta E_b \quad (6)$$

Additionally, an advantage of calculating the binding free energy multiple temperatures is the availability of the conformational entropy of binding  $\Delta S_b^{\circ}$

$$\Delta S_b^{\circ} = -\frac{\partial}{\partial T} \Delta G_b^{\circ} \quad (7)$$

As we have only several binding free energies at discrete temperatures, the above differential was solved by using a linear least squares approximation.

$$f = \sum_i \left( \frac{b + mx_i - y_i}{\sigma_i} \right)^2 \quad (8)$$

$$\frac{\partial f}{\partial m} = \sum_i \left( \frac{(b + mx_i - y_i)x_i}{\sigma_i^2} \right) = 0 \quad (9)$$

and

$$\frac{\partial f}{\partial b} = \sum_i \left( \frac{(b + mx_i - y_i)}{\sigma_i^2} \right) = 0 \quad (10)$$

Solving these two simultaneous equations yields

$$\Delta S_b^{\circ} = \frac{\sum(1/\sigma_i^2) \sum(x_i y_i / \sigma_i^2) - \sum(x_i / \sigma_i^2) \sum(y_i / \sigma_i^2)}{\sum(1/\sigma_i^2) \sum(x_i^2 / \sigma_i^2) - (\sum x_i / \sigma_i^2)^2} \quad (11)$$

The error on the entropy was calculated using the errors on the individual points.

$$\frac{1}{\sigma^2} = \sum \left( \frac{(x_i - \langle x \rangle)^2}{\sigma_i^2} \right) \quad (12)$$

Finally, the binding reorganization energy  $\Delta E_{\text{reorg}}$  was computed as the residual of the reorganization free energy of binding after subtraction of the entropic component:

$$\Delta E_{\text{reorg}} = \Delta G_{\text{reorg}} + T \Delta S_b^{\circ} \quad (13)$$

The errors reported for:  $\Delta E_{\text{reorg}}$  and  $\Delta G_{\text{reorg}}$  were calculated using standard error propagation. For  $\Delta E_b$  we used the standard error of the mean. All statistical uncertainties, including error bars in plots, are reported as twice the standard deviation (96% confidence interval).

### 2.3. Computational Details

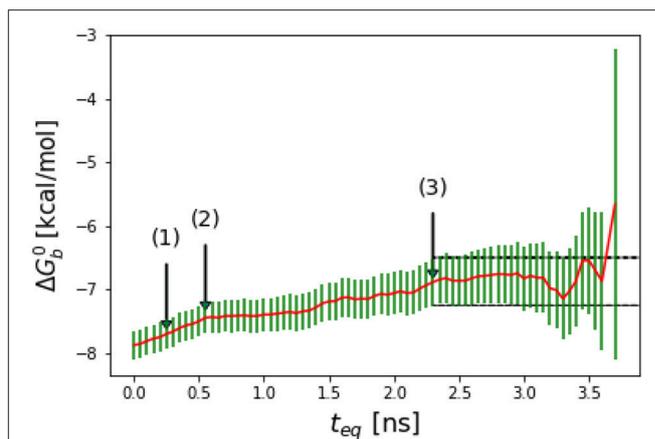
In this work we employ an effective potential energy function based on the AGBNP2 implicit solvent model (Gallicchio and Levy, 2004; Gallicchio et al., 2009), and the OPLS-AA force field (Jorgensen et al., 1996; Kaminski et al., 2001) for the covalent and non-covalent interactions. Parallel molecular dynamics simulations were performed with the IMPACT program (Banks et al., 2005). Replica exchange conformational sampling was conducted for all combinations of eight temperature spanning 300 to 379 K, and 26 intermediate  $\lambda$  steps at  $\lambda = 0.0, 0.002, 0.0048, 0.006, 0.008, 0.01, 0.015, 0.02, 0.0225, 0.025, 0.03, 0.0325, 0.035, 0.04, 0.07, 0.1, 0.25, 0.35, 0.45, 0.55, 0.65, 0.71, 0.78, 0.85, 0.92$ , and 1, for a total of 208 replicas. The binding site volume was defined as any conformation in which the peptide center of mass is within 6 Å of the center of mass of the  $C_{\alpha}$  atoms of residues 168-174 and 178 of chain A and residues 95-99, 102, 125, 128, 129, and 132 of chain B (residue and chain designations according to 3NFB crystal structure) of HIV1-IN. The peptide was sequestered within the binding site by means of a flat-bottom harmonic potential with a force constant of 3.0 kcal/mol Å<sup>2</sup> applied to atoms with distances greater than 6 Å. The volume of the binding site is calculated to be 904 Å<sup>3</sup> corresponding to  $k_B T \ln C^{\circ} V_{\text{site}} = 0.69$  kcal/mol. As in previous work (Gallicchio et al., 2014), the  $C_{\alpha}$  atoms of the residues of the CCD domain of HIV1-IN were restrained to the overall structure by means of spherical harmonic restraints with a force constant of 1 kcal/mol Å<sup>2</sup>, excluding the residues closest to the binding site (residues 167-178 of chain A and 95-102 and 124-132 of chain B) which were left unrestrained.

We simulated the complexes of HIV1-IN with six peptides: cyclic SLKIDNLD, linear H-SLKIDNKL-OH, cyclic ALKIDNLD, cyclic ALKIDNMD, cyclic SLKINNLD, and cyclic SLKADNLD. Multi-dimensional replica exchange calculations were performed for about 3.3 ns of molecular dynamics per replica, or 867 ns of average simulation time for each complex and approximately 5.2  $\mu$ s of simulation time in total. Binding energies were sampled with the frequency of 25 ps for a total of 35,000 binding energy samples per complex. All binding free energy calculations were conducted on Louisiana State University's XSEDE SuperMic supercomputer using the ASyncRE job distribution middleware (Gallicchio et al., 2015).

## 2.4. Error Analysis and Determination of Equilibration and Convergence

Statistical uncertainties of binding free energy estimates were calculated using UWHAM's built-in Fisher's analysis. The UWHAM formalism, based on mathematical distribution theory, uses the curvature of the likelihood function to calculate the variance of free energy estimates. A detailed derivation is provided by Tan et al. (2012). Fisher's analysis assumes uncorrelated data, we therefore assessed the statistical inefficiency of the binding energy time series data by computing uncertainties based on a standard  $1/\sqrt{N}$  error on progressively smaller, random sets of the data and then subsequently plotted those errors against the error provided by UWHAM that contained the same number of data points. From these comparisons, as well as the data obtained from running autocorrelation analysis, we have concluded that binding energies were collected with sufficiently small frequency so as to make statistical correlations negligible.

Most molecular dynamics simulations are initiated with structures that are atypical of equilibrium conformations. As binding energy calculations are sensitive to small perturbations in configuration, it is typical practice to remove an initial portion of the trajectory in which the system is approaching equilibration so as not to adulterate the equilibrium result (Klimovich et al., 2015). To determine the amount of initial data to eliminate, in this work we employ a method similar in the spirit of reverse cumulative averaging from Yang and Karplus (2003) and the autocorrelation analysis discussed recently by Chodera (2016). In this approach, we examine the time series of binding free energy estimates as a function of increasing equilibration time  $t_{eq}$  measured from the beginning of the simulation. Specifically, we define  $\Delta G_b^0[t_{eq}]$  as the binding free energy estimate obtained by discarding initial data up to simulation time  $t_{eq}$ . The sequence of binding free energy estimates so obtained is referred to as the reverse cumulative profile (Yang and Karplus, 2003). In these profiles (see Figure 2 as an example) binding free energy estimates to the right, at long equilibration times, are those least affected by bias introduced by unequilibrated data at early times. Conversely, estimates to the left, at short equilibration times, are expected to be the most biased. The decision of which equilibration time to pick is not obvious because the least biased estimates, since they correspond to the smallest binding energy samples, are also those with the largest statistical uncertainties.



**FIGURE 2** | Representative reverse cumulative convergence plot for the binding free energy for the complex with peptide Mutant 2 (S362A/L368M) highlighting the different methods for choosing equilibration time. (1) represents the equilibration time,  $t_{eq}$ , returned by the quantitative protocol we adopted in this work as described in the text, (2) represents the  $t_{eq}$  chosen qualitatively based on the first observed inflection point of the plot and (3) represents the smallest  $t_{eq}$  corresponding to a free energy estimate statistically indistinguishable from those at longer equilibration times. The resultant  $\Delta G_b^0(t_{eq})$  are  $-7.7 \pm 0.2$ ,  $-7.5 \pm 0.2$ , and  $-7.1 \pm 0.4$  kcal/mol with the three methods, respectively. The three methods yield statistically equivalent results in this case.

A suitable equilibration time can be chosen qualitatively as, for example, the smallest equilibration time that gives a free energy estimate statistically indistinguishable from those at longer equilibration times (see Figure 2) (Yang and Karplus, 2003).

In this work we have explored the quantitative protocol for the choice of the equilibration time recently proposed for averages of correlated time series (Chodera, 2016). When applied to a generic time series  $x_i = x(t_i)$ , this method consists of picking the equilibration time so as to maximize the effective number,  $n_{eff}$ , of statistically independent samples of the time series (Allen and Tildesley, 1993) which remains after removing the first  $n_{eq}$  data points collected prior to the chosen equilibration time (see Figure 3 for an example). The number of statistically independent samples is given by

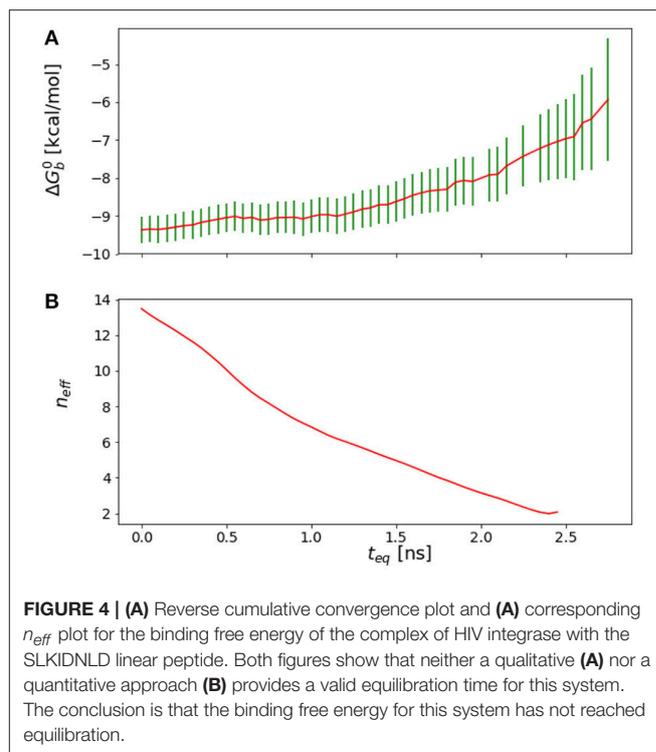
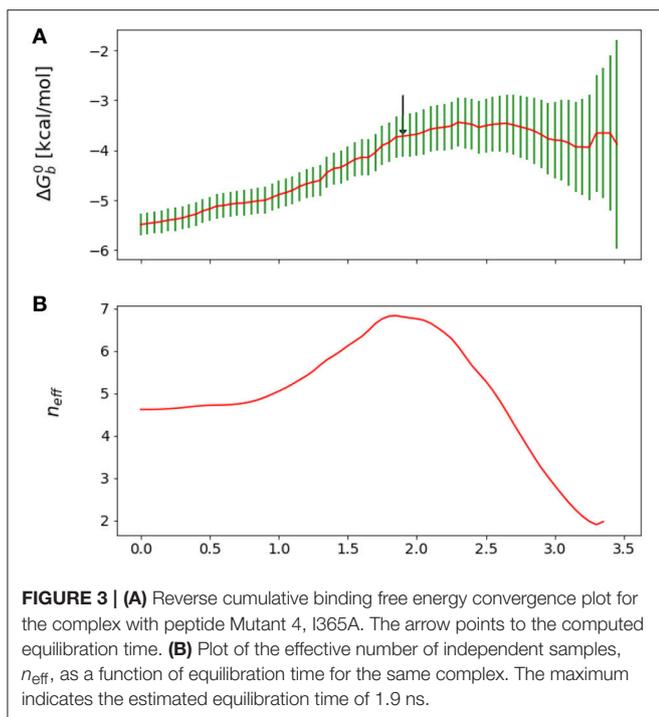
$$n_{eff} = \frac{n - n_{eq}}{g(n_{eq})}, \quad (14)$$

where  $n$  is the total number of points in the time series,  $n - n_{eq}$  is the number of points remaining in the time series after removing the first  $n_{eq}$  data points, and

$$g(n_{eq}) = 1 + 2\tau_{eq} \quad (15)$$

is the statistical inefficiency of the time series, where  $\tau_{eq}$  is the correlation length given by:

$$\tau_{eq} = \sum_{t=1}^{t_{max}-1} \left(1 - \frac{t}{t_{max}}\right) C_t \quad (16)$$



where the autocorrelation function,  $C_t$  is defined as:

$$C_t = \frac{\langle x_n x_{n+t} \rangle - \langle x_n \rangle^2}{\langle x_n^2 \rangle - \langle x_n \rangle^2} \quad (17)$$

It was found that this approach incorporates the trade-off, discussed above, between minimizing the variance while at the same time minimizing the bias in the estimate. The variance obviously decreases as the number of samples included in the estimate increases, that is as  $n_{\text{eq}}$  decreases. In contrast, Chodera observed that the correlation length, and, consequently, the statistical inefficiency, grow as atypical samples at the beginning of the simulation are included (Chodera, 2016), presumably because these introduce an overall drift in the time series.

We obtained reasonable estimates of equilibration times by applying the approach above to the reverse cumulative profile of the binding free energy  $\Delta G_b^0[t_{\text{eq}}]$ . We rationalize this observation by noting that, while the reverse cumulative profile is not a time series of an observable,  $n_{\text{eff}}$  still captures the trade-off between maximization of the number of samples and minimization of the statistical inefficiency, which is larger at small equilibration times due to the overall drift of the reverse cumulative profile in that region (see Figure 3). Specifically, we considered the family of reverse cumulative free energy profiles  $\Delta G_b^0[t > t_{\text{eq}}]$ , that is those derived by the full reverse cumulative profile after removal of the values for  $t < t_{\text{eq}}$ . For each profile we performed autocorrelation analysis to compute the corresponding correlation length  $\tau(t_{\text{eq}})$ , the statistical inefficiency  $g(t_{\text{eq}}) = 1 + 2\tau(t_{\text{eq}})$  and effective number of statistically independent samples  $n_{\text{eff}} \propto (t_{\text{max}} - t_{\text{eq}})/g(t_{\text{eq}})$ . We then selected the optimal equilibration time as that one that maximizes  $n_{\text{eff}}$ . As illustrated in Figure 3, in the

case of mutant 4 (I365A), for example, this procedure returns an equilibration time of 1.9 ns per replica.

In general, we observe that the outlined procedure returns small equilibration times for systems that would be considered to have equilibrated quickly based on qualitative arguments, statistical uncertainties, and the shape of the reverse cumulative profiles. Conversely, the procedure returned relatively long equilibration times (see Results) for systems that qualitatively appear to equilibrate slowly, for example those that display persistent drift of the reverse cumulative profile for short equilibration times (Figures 3, 4). In the case of a linear peptide, considered from multiple perspective to have failed to achieve equilibration, this procedure correctly indicated the lack of a reasonable equilibration time as the  $n_{\text{eff}}$  curve did not display a clear maximum as shown in Figure 4.

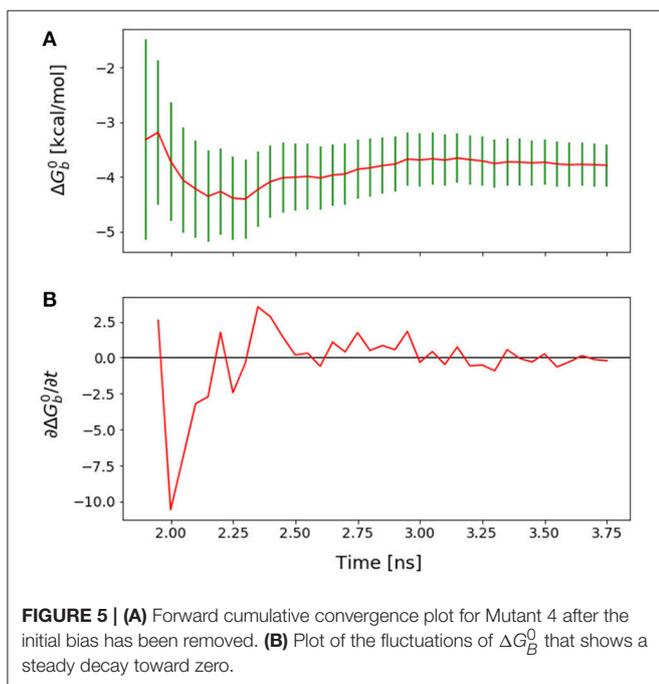
Having determined the equilibration time, we then assessed convergence of the free energy calculations by examining the conventional forward cumulative profiles (see Figure 5 and Figure S1 for an example) obtained considering only binding energy samples collected at times following the equilibration time until some maximum time  $t$ , which is varied. Systems that are considered converged display a forward cumulative profile with monotonically decreasing uncertainties, ideally random fluctuations, or, more often, small downward or upward drifts contained within computed uncertainties. To better assess convergence, it has also been helpful to consider the quantity  $(\delta \Delta G_b^0 / \delta t)$ , which measures the differential effect of adding samples contained between  $t$  and  $t + \Delta t$ . As shown for example in Figure 5, in converged calculations this function progressively decays to zero with smaller and smaller fluctuations with increasing simulation time.

### 3. RESULTS

#### 3.1. Thermodynamic Parameters

Computed thermodynamic parameters for the binding of five cyclic LEDGF-derived peptides considered in this work are listed in **Table 1**. The complexes with wildtype (SLKIDNLD), mutant 1 (ALKIDNLD), and double-mutant 2 (ALKIDNMD) have been studied by Rhodes et al. (2011) who obtained  $IC_{50}$  values and crystal structures. Mutants 3 (SLKINNLN) and 4 (SLKADNLD) have been included in this study to probe the thermodynamic and structural consequences of mutations in residues known to be critical for successful binding (Tsiang et al., 2009). We have been unsuccessful at computing the binding free energy of the linear wild type peptide (see below), which showed weak or no activity in experimental tests (Tsiang et al., 2009; Rhodes et al., 2011).

The computed binding free energies for wildtype and mutant 1 cyclic peptides are qualitatively consistent with  $\mu\text{M}$ -range  $IC_{50}$  values (Rhodes et al., 2011) measured for these peptides ( $85.2 \pm 24.5 \mu\text{M}$  and  $39.7 \pm 7.1 \mu\text{M}$ , respectively). Also in agreement with inhibition experiments, mutation of the S366 serine residue to alanine causes minor changes in binding affinity. The double mutant S366A/L368M (M2 in **Table 1**), while undergoing a greater conformational change relative to wild type and mutant 1 (**Figures 7, 8**), is predicted to have similar binding affinity to the wild type and the single mutant (**Figure 6B**). Experimental values for this double-mutant are contradictory. Rhodes et al. (2011) report no detectable inhibition activity for this peptide, while they were able to obtain a crystal structure of it bound to HIV integrase (PDB id 3AVJ) (Rhodes et al., 2011). Taken together, this evidence suggests that, while mutant 2 (M2) is capable of binding to HIV integrase, somehow it can not inhibit it effectively.



Consistent with experimental findings (Tsiang et al., 2009; Rhodes et al., 2011), mutation of the critical D366 residue (M3 in **Table 1**) resulted in a significant reduction of binding affinity. This residue is known to form a strong hydrogen bonding interaction motif with the backbone atoms of His171 and Glu170 of HIV integrase (**Figure 8**), which is also found in many small-molecule inhibitors (Gallicchio et al., 2014). The loss of this interaction is observed to lead to major disruption of the peptide-receptor binding interface and loss of binding affinity. Analogously, mutation of I365 to alanine (M4 in **Table 1**) causes disruption of favorable hydrophobic packing peptide-receptor interactions and loss of binding affinity, although to a smaller degree than the D366N mutation.

**Table 1** also reports the decomposition of the computed binding free energies into average binding energies and reorganization free energies, and the further decomposition of the latter into reorganization energy and conformational entropy of binding. Average binding energies ( $\Delta E_b$ , third column in **Table 1**) reflect the strength of direct interatomic interactions between the peptide and the protein receptor. These are relatively constant around  $-58 \text{ kcal/mol}$  across the peptide set, with the exception of mutant 3, for which  $\Delta E_b$  is about 20 kcal/mol less favorable than the others. This behavior confirms that the loss of binding affinity for mutant 3 is mainly due to the loss of key hydrogen bonding interactions between D366 and a histidine residue in the binding pocket of the receptor (**Figure 8**). The magnitude of the interaction energy loss (20 kcal/mol) is also consistent with the loss of two strong hydrogen bonds and the corresponding favorable electrostatic interactions.

In contrast, the weakening of the binding of mutant 4 (M4 in **Table 1**), in which a bulky isoleucine residue is replaced by alanine, displays a completely different thermodynamic signature. The average binding energy for this peptide complex is of similar magnitude than the higher affinity peptides, and it is actually predicted to be greater than that of mutant 1 even though their predicted binding constants differ by at least two orders of magnitude. The weaker binding of mutant 4 is predicted to be mainly due to a large unfavorable reorganization free energy (fourth column in **Table 1**), which, in turn, is caused by unusually large intramolecular energy strain ( $\Delta E_{\text{reorg}}$ ) only partially compensated by a smaller entropic penalty ( $T\Delta S_b^0$ ). As further discussed below, this behavior is consistent with the observed structural response of the peptide and receptor attempting to compensate for the loss of favorable packing interactions (**Figure 8D**). In contrast, the thermodynamic parameters collected in **Table 1** indicate that the loss of hydrogen bonding interactions in mutant 3 results, effectively, in the unbinding of the peptide with little or no tendency for conformational reorganization aimed at regaining binding as shown in **Figure 8C** (see below for a discussion of the observed structural changes).

#### 3.2. Structural Considerations

The equilibration times of the complexes are found to be highly correlated with the extent of their conformational reorganization to go from the starting structure (which in all cases was the crystal structure of the wildtype peptide) to

the equilibrated conformational ensemble. Thus, as expected, the binding free energy of the wildtype peptide converged relatively quickly (**Figure 9B**). Conversely, the I365A mutation caused extensive conformational rearrangements which required a longer equilibration (**Figure 6D**).

Consistent with the crystal structure (3AVB), residues Ser362, Lys364, Ile365, Asp366, and Asn367 of the cyclic WT peptide form the primary intermolecular interactions with the receptor (**Figure 7**). The carboxylate group of Asp366 of the peptide forms the strong hydrogen bonds with the NH backbone groups of Glu170 and His171 of the receptor which are the hallmark of the LEDGF-HIV IN complex (Cherepanov et al., 2005a) as well as those with peptido-mimetic synthetic inhibitors (Peat et al., 2014). The carboxylate group of Asp366 is further anchored

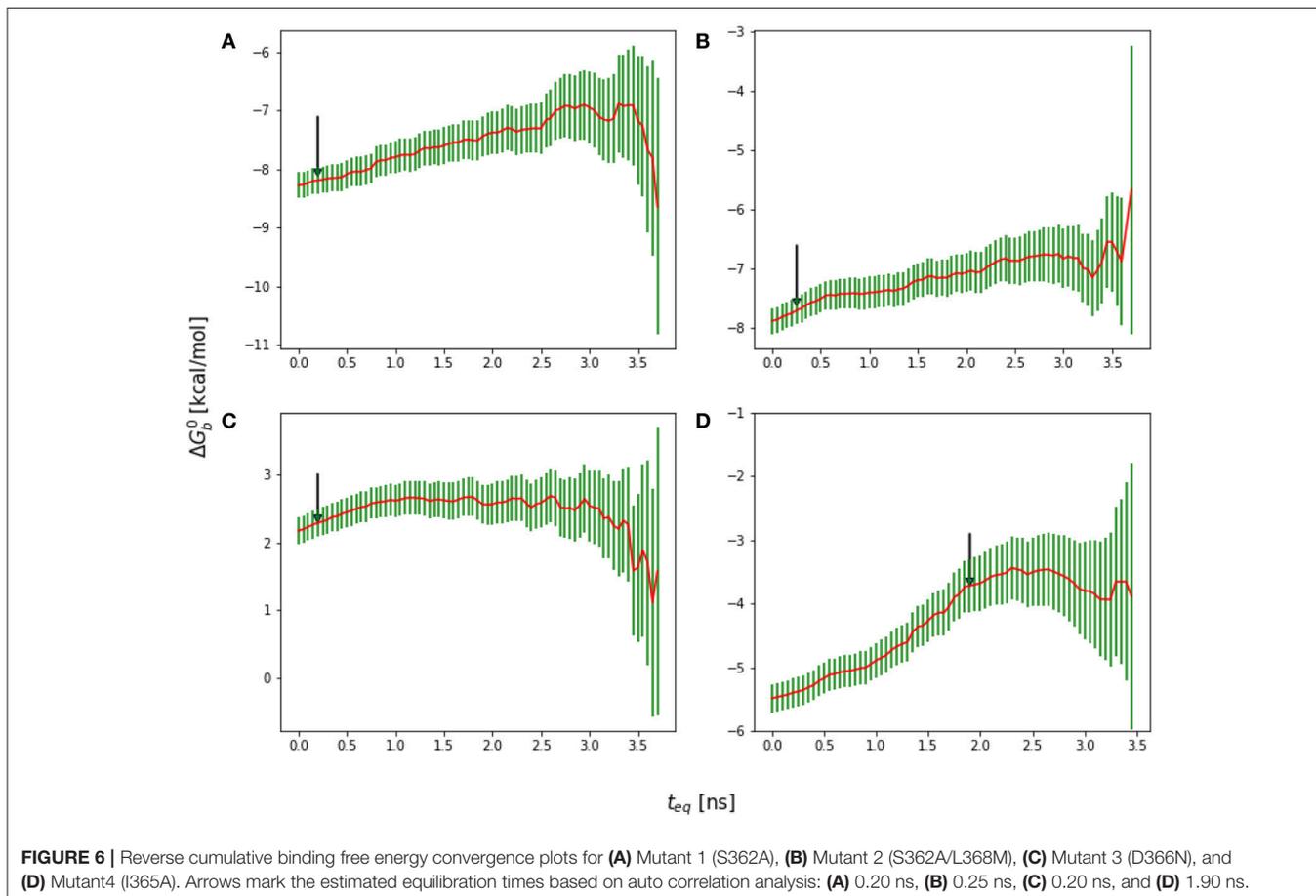
by a hydrogen bond with the hydroxyl group of Thr174 and with a salt bridge with the protonated nitrogen atom of His171. Both Ser362 and Lys364 of the peptide forms strong interactions with the sidechain of Glu170. Ile365 forms backbone-backbone interactions with Gln168 and Thr174. The sidechain of Ile365 is additionally nestled into an hydrophobic pocket of the receptor lined by residues Ala128, Trp131, Trp132, and Met178. The solvent-exposed residues of the peptide (Leu363, Leu368, and Asp369), do not interact consistently with the receptor although they make occasional intramolecular interactions with each other.

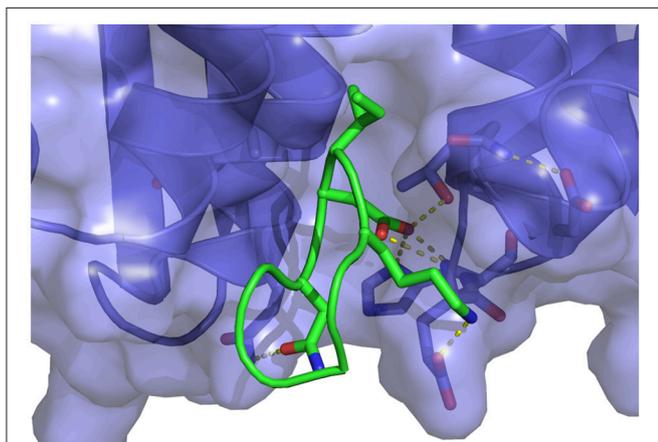
A serine to alanine mutation of residue 362 (M1 mutant) causes a slight, but statistically significant decrease in the predicted binding affinity (**Table 1**). Auto-correlation analysis of

**TABLE 1** | Computed thermodynamic parameters for the binding between HIV1-integrase and a series of cyclic peptides derived from the LEDGF protein.

Peptide	$\Delta G_b^{\circ a}$	$\Delta E_b^a$	$\Delta G_{reorg}^a$	$\Delta E_{reorg}^a$	$T\Delta S_b^{\circ a}$
SLKIDNLD (WT)	$-8.6 \pm 0.4$	$-59.3 \pm 0.5$	$50.6 \pm 0.5$	$23.6 \pm 1.9$	$27.0 \pm 1.8$
ALKIDNLD (M1)	$-8.1 \pm 0.2$	$-56.9 \pm 0.4$	$48.8 \pm 0.4$	$22.1 \pm 0.4$	$26.7 \pm 0.1$
ALKIDNMD (M2)	$-7.6 \pm 0.2$	$-59.5 \pm 0.5$	$51.9 \pm 0.5$	$24.6 \pm 0.8$	$27.3 \pm 0.6$
SLKINNLD (M3)	$2.7 \pm 0.2$	$-36.1 \pm 0.6$	$38.3 \pm 0.6$	$12.4 \pm 0.6$	$25.9 \pm 0.6$
SLKADNLD (M4)	$-3.7 \pm 0.2$	$-57.6 \pm 0.5$	$53.9 \pm 0.5$	$33.2 \pm 0.6$	$20.7 \pm 0.4$

<sup>a</sup> In kcal/mol.





**FIGURE 7** | Representative structure extracted from the simulation of the wildtype cyclic peptide (WT in **Table 1**) bound to the LEDGF-binding domain of HIV integrase. The following intermolecular interactions are highlighted: Lys364-Glu170, Asp366-Thr174, His171, and Asn367-Gln95. Also shown in the intramolecular interaction between Asp168 and Gln169 of the peptide and the key bidentated interaction between Asp366 and the backbone nitrogen atoms of Glu170 and His171.

the reverse cumulative binding free energy profiles (**Figure 6A**) revealed only a small equilibration bias of about 1 kcal/mol, consistent with the relatively minor conformational changes relative to the wildtype peptide (**Figure 8A**). The most noticeable structural difference is the shift of the interaction between Glu170 and Ser362, now absent, to Asn 367 and the concomitant weakening of the Lys364-Glu170 salt bridge which now occurs less often. These subtle changes are the cause of the weakening of the binding energy and the slight less favorable binding free energy relative to the wildtype (**Table 1**).

The double mutation S362A:L368M (M2 in **Table 1**) resulted in no significant change of binding free energy relative to the S362A single-mutant. However, the methionine sidechain at position 368 now affords additional polar interactions with Thr125 and it induces the recruitment of Asp369 (previously completely solvent exposed) in the same interaction (**Figure 8B**). As a result, the peptide shifts overall position toward the 124-131  $\alpha$ -helix of the receptor.

The D366N mutation resulted, essentially, in the dissociation of the complex (**Figure 8C**), consistent with the unfavorable computed binding free energy (**Table 1**, M3 mutant) and previous experiments (Tsiang et al., 2009). The peptide was kept in proximity of the pocket only by the applied tether (see Methods). The conformations of the receptor and the peptide rapidly reorganized to screen the unmet positive charge of His171 while the Asn366 sidechain shifted away from the Glu170 backbone. The rapid equilibration of the peptide, freed from strong intermolecular interactions, is the cause of the smaller equilibration time observed (**Figure 6C**) for the M3 mutant relative to the other mutants.

The I365A mutant (M4 in **Table 1**) displayed the most significant and lengthy equilibration of the binding free energy among the peptides studied (**Figure 6D**). Unlike the other

mutants, the most significant structural rearrangement occurred for the receptor. The void in the hydrophobic binding pocket left from the isoleucine to alanine mutation slowly collapsed around the alanine sidechain (**Figure 8D**). Concomitantly, the peptide moves deeper into the binding cavity causing the solvent-exposed residues (Leu363, Leu368, and Asp369) to make increased interactions with the receptor while slightly decreasing interactions between Asp366 and Glu170/His171. These rearrangements recover most of the interaction energy of binding (**Table 1**, third column) at the expense of higher conformational strain (fifth column), resulting in a binding free energy 5 kcal/mol less favorable than wildtype in agreement with experimental evidence (Cherepanov et al., 2005b; Tsiang et al., 2009).

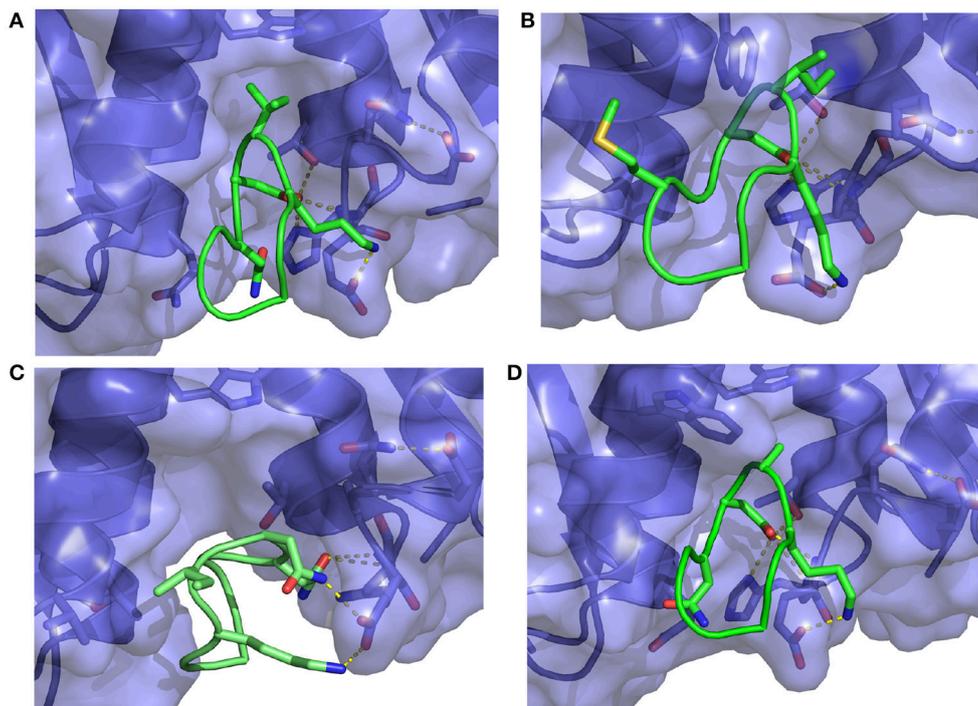
### 3.3. The Potential Pitfalls of Protein-Peptide Binding

Alchemical binding free energy calculations require careful planning and assessment of the results (Klimovich et al., 2015). This is particularly so for protein-peptide complexes due to their large conformational variability. Here we discuss the challenges we encountered in applying our calculation protocols, tuned for small molecule binding, to peptides.

The first challenge we encountered turned out to be related to our initial choice of the  $\lambda$  schedule, which was based on the one that had been used with success on large ligands for the same receptor site (Gallicchio et al., 2014). With this  $\lambda$  schedule we collected a substantial amount of binding free energy data that, when displayed using a traditional forward cumulative plots showed no particular disturbance (**Figure 9B**) beyond a persistent drift signaling slow convergence. It was only when we analyzed the reverse cumulative plot (**Figure 9A**) that we discovered that the slow convergence was a symptom of a serious numerical artifact.

The reverse cumulative plot (**Figure 9A**) shows an abrupt shift in the binding free energy corresponding to an equilibration time of approximately 1 ns. Subsequent examination of the error bounds returned by the UWHAM (Tan et al., 2012) procedure, revealed that without the equilibration data bias introduced by the binding energy data collected at times prior to 1 ns, the uncertainty in the computed binding free energy was so large (on the order of 1000 kcal/mol) as to make the results meaningless.

It was eventually established that the failure in obtaining a reasonable binding free energy estimate was due in this case to the choice of the  $\lambda$  schedule, which lacked sufficient intermediate  $\lambda$  values between 0.1 and 0.4 to ensure sufficient ensemble “overlap” of adjacent alchemical states (Klimovich et al., 2015). Specifically, after some equilibration, for  $\lambda < 0.1$  all of the binding energy values clustered around the maximum value  $u_{\max}$  [see Equation (5)], which were not observed in any of the other  $\lambda$  states. As a result, the UWHAM multi-state maximum likelihood function (Tan et al., 2012) lacked a well-defined maximum corresponding to the binding free energy profile. This problem had not been apparent when unequilibrated data at the beginning of the simulation, which contained a small amount of binding energy data bridging alchemical states across  $\lambda = 0.1$ ,



**FIGURE 8** | Representative structures extracted from the simulations of **(A)** the S362A mutant cyclic peptide (M1 in **Table 1**) and **(B)** the S362A:L368M double mutant cyclic peptide (M2), **(C)** the D366N cyclic peptide (M3), and **(D)** the I365A mutant cyclic peptide (M4) with the HIV integrase LEDGF-binding domain. All interactions shown were previously described in **Figure 7**. Notice in **(C)** that the D366N mutation causes severe distortion of the peptide as well as the receptor relative to the bound conformation of the complex.

was included in the estimate, resulting in a forward cumulative convergence plot not clearly out of the ordinary.

As a result of this assessment, we modified the  $\lambda$  schedule to insert intermediate  $\lambda$  states between 0.1 and 0.4 more likely to sample both large and smaller binding energy values. This resulted in binding free energy estimates with reasonable error bounds when analyzed with both forward and reverse cumulative plots (**Figures 9C,D**).

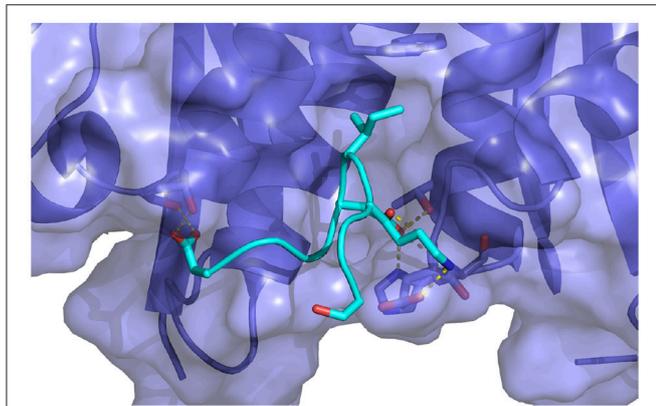
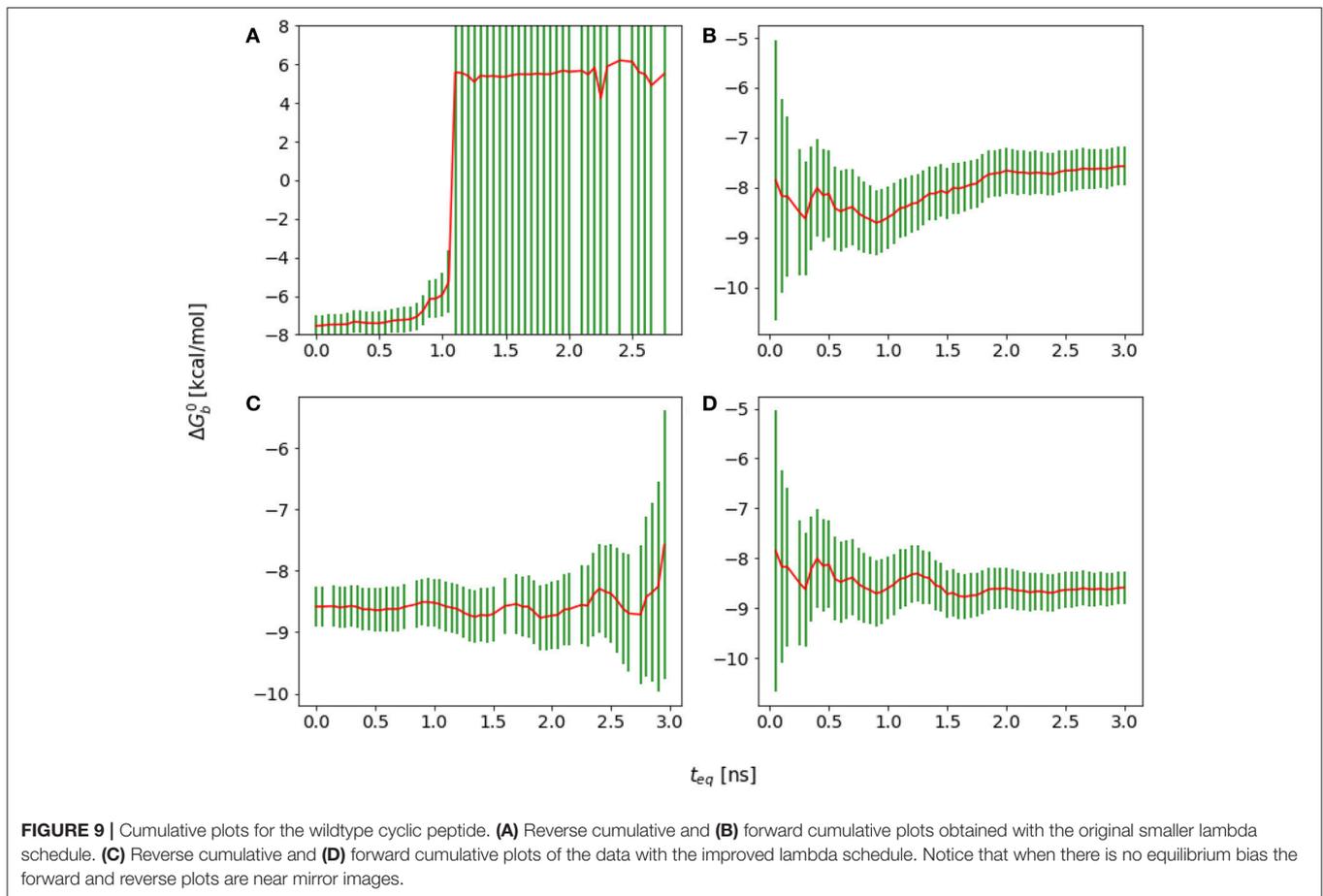
This experience highlights the need for careful examination of equilibration artifacts and error bounds when assessing the results of alchemical binding free energy calculations. Clearly, in this case, the choices of simulation length and  $\lambda$  schedule tuned for small-molecule binding were not suitable for the modeling of the binding of the larger peptides studied here. The optimized  $\lambda$  schedule addressed the requirement of a suitable, unbroken, thermodynamic path between the coupled and uncoupled states of the complex and enabled the assessment of equilibration bias in the binding free energy estimates. However it did not guarantee convergence of the binding free energy estimates when these are affected by challenges of a different nature, as the following example illustrates.

One of aims in this work had been to study the thermodynamic factors that favor the binding of the cyclic wild-type peptide over the linear one, H-SKIDNLD-OH (Tsiang et al., 2009; Rhodes et al., 2011) (**Figure 10**). However, we were unable to obtain a converged binding free energy for the

linear form of the wild-type peptide (H-SKIDNLD-OH). As shown in **Figure 4**, the reverse cumulative plot for the complex with this peptide does not show a plateau, indicating that the system was still in the equilibration phase up until the longest simulation time considered. Lack of convergence is confirmed by the forward cumulative plot including all of the binding energy samples (**Figure 11A**).

Analysis of molecular dynamics trajectories revealed that the lack of convergence was due to the slow equilibration of the relative populations of two competing conformational states of the complex (**Figure 11B**). The first state corresponds to a compact conformation of the linear peptide tethered by an intramolecular salt bridge similar to the conformation of the cyclic peptide. The other, more extended, conformation resulted from the replacement of the intramolecular salt bridge with intermolecular salt bridges between the termini of the peptide and residues lying on the rim of the protein receptor pocket (**Figure 10**). This extended bound conformation, whose formation was found to be reproducible across independent runs, is strongly favored by binding energy (approximately  $-120$  kcal/mol, compared with binding energies of half this magnitude for cyclic peptides and for the compact conformation of the linear peptides—see **Table 1** and **Figure 11**—and strongly disfavored by reorganization free energy.

Because of the trade-off between energetic and reorganization/entropic opposing thermodynamic driving forces, the compact



**FIGURE 10** | Key interactions observed in the conformational ensemble at low binding energies (around  $-120$  kcal/mol) of the complex of HIV integrase with the linear peptide. Notable differences relative to the complexes with the cyclic peptides include the participation of Asp369 with  $\alpha_2$  helix residues of HIV integrase.

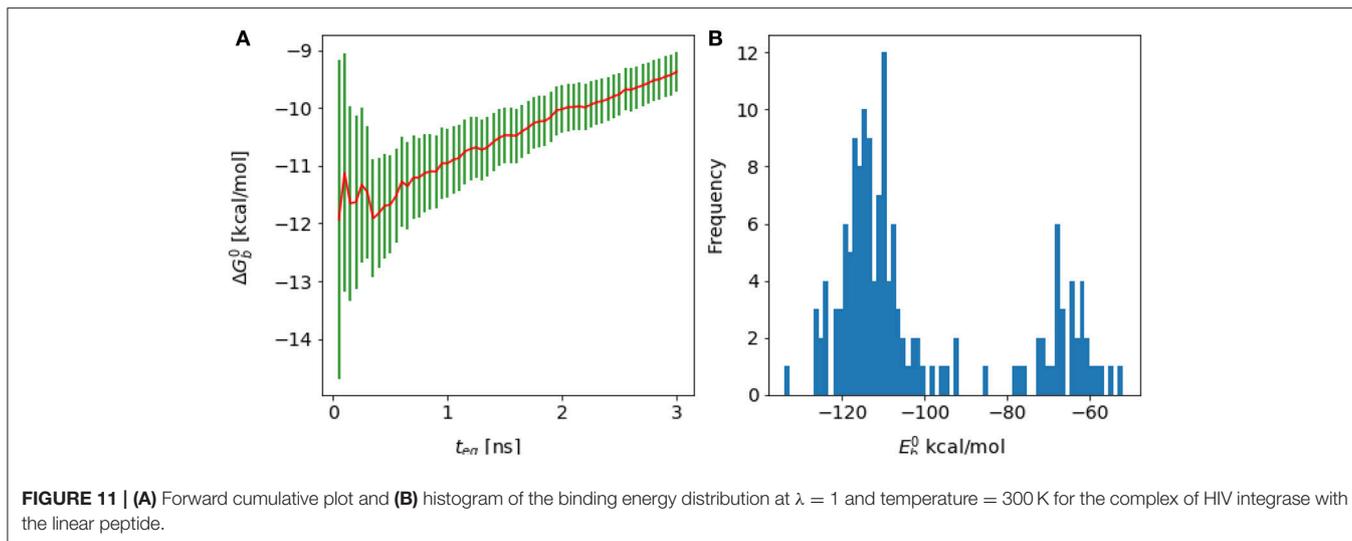
and extended binding modes were found to coexist at  $\lambda = 1$ , producing a characteristic bimodal binding energy distribution (**Figure 11**). The binding free energy estimate is found to be very sensitive to the relative population of these two binding modes,

which, however, could not be determined with certainty due to lack of interconversion events. As a result, the binding free energy exhibits a slow drift as a function of simulation time with no identifiable convergence point (**Figure 4**).

Cases, such as this, of slow convergence due to slow intramolecular conformational equilibration can not be easily addressed by conventional conformational sampling acceleration methods based, for example, on Hamiltonian replica exchange on the alchemical variable as done here. Some success has been gained with biasing potentials targeting specific degrees of freedom (Ensing et al., 2006; Kim et al., 2010; Procacci et al., 2013; Cavalli et al., 2014), such as dihedral angles (Wang et al., 2012; Lindert et al., 2013). We have not attempted, in this work, to obtain a converged estimate of the relative population of the two binding modes of the linear peptide using these methods. This is partly because of the awareness that implicit solvation may not be an optimal model for conformational equilibria involving salt bridges (Okur et al., 2008; Gallicchio et al., 2009), even if these could be sampled thoroughly.

## 4. DISCUSSION

Despite decades of intense medical research, HIV infections continue to be a major widespread problem for worldwide health



with 37 million people living with AIDS and 21 million on antiretroviral therapy (World Health Organization, 2017). The integrase enzyme of the HIV virus is one of the main targets of ongoing anti-viral treatments. Raltegravir, for example, is an inhibitor of integrase's strand transfer reaction approved for clinical use in 2008. It remains one of the most used integrase inhibitors in conjunction with other viral inhibitors in highly active antiretroviral treatment courses (Summa et al., 2008). Even though they are able to significantly reduce viral load, a significant downside of strand transfer integrase inhibitors is that HIV becomes readily resistant to their effects (Murray et al., 2007; Sarkis et al., 2008). There has been therefore great interest in identifying alternative inhibitory pathways for HIV integrase.

The interaction between integrase and LEDGF/p75 is considered one of the most promising untapped targets. LEDGF/p75 is an endogenous cofactor that is essential to viral replication. It facilitates the integration of viral DNA by transporting the viral protein into the cell nucleus and inserting the viral DNA into host DNA. LEDGF/p75 is also believed to play a role in the protection of integrase from degradation by the host cell (Smith and Daniel, 2006). Because it targets a host protein, the LEDGF binding domain of HIV integrase is believed to be less susceptible to the insurgence of resistance mutations. There have been efforts to identify peptidomimetic compounds capable of disrupting the interaction between HIV integrase and LEDGF/p75 (Gallicchio et al., 2014; Peat et al., 2014).

Cherepanov et al. (2005a) reported in 2005 the crystal structure of the complex between HIV integrase and the integrase binding domain of LEDGF (PDB code: 2B4J). The structure revealed key residues responsible for mediating the interaction. Ile365 of LEDGF/p75 is found into the hydrophobic cavity of the integrase binding site that consists of Leu102, Ala128, Ala129, Trp131, Trp132, Thr174, and Met178. Ile365 of LEDGF also formed a hydrogen bond between its backbone amide and the backbone carbonyl group of Gln168 of integrase. Asp366 of LEDGF makes two-pair hydrogen bonds with Glu170 and His171 of integrase. In addition, a salt-bridge is found between Lys364

of LEDGF and Glu170 of integrase. Mutations of these critical residues weaken the LEDGF/integrase interaction (Tsiang et al., 2009).

Based on these earlier studies, Rhodes et al. (2011) investigated the structures of HIV integrase bound to a series of cyclic peptides including the wild-type sequence SLKIDNLD, corresponding to residues 362-369 of LEDGF/p75, as well as single and double mutants. Rhodes et al. obtained crystal structures for 13 of the complexes they investigated (PDB codes: 3AV9, and 3AVA through 3AVN) (Rhodes et al., 2011). While we did not use the obtained crystal structures as templates for our S362A and S362A:L368M mutants, our resulting structures matched well with those reported by Rhodes et al. (2011). The structure for the wild-type cyclic peptide matched most of the intermolecular interactions found in the structure of the complex with LEDGF/p75.

We simulated the SLKIDNLD, ALKIDNLD, and ALKIDNMD cyclic peptides investigated by Rhodes et al. (2011) to compare directly with their findings. We also simulated two other cyclic peptides with known fatal mutations, D366N and I365A, to compare with earlier experiments (Cherepanov et al., 2005a,b; Tsiang et al., 2009). The results of our calculations recapitulate all of the experimental findings and identify the specific interactions responsible for the binding trends (Figures 7, 8, and Table S1). We were able to confirm that three residues of the peptides, corresponding to Lys364, Ile365, and Asp367 of LEDGF, are responsible for the most critical protein-peptide interactions. Conversely, the residue pairs Leu363-Leu368 and Ser362-Asn367 form the most stable intramolecular interactions within the peptides.

In this study, we used an alchemical approach to compute the binding free energies for the complexes of HIV integrase with one linear peptide and five cyclic peptides derived from LEDGF. A significant portion of the effort was devoted to error and convergence analysis (Shirts, 2013; Klimovich et al., 2015). In normal circumstances, visual inspection of the binding data time series could be sufficient to get a good estimate of

equilibration time and convergence. In more complex systems, as in this case, variations of binding free energy estimates over the course of the simulation are slow and gradual and require careful consideration. The onset of equilibration is particularly slight, but not less important, when the starting conformation is close to the equilibrium conformation. To estimate equilibration times, that is the amount of data to neglect at the beginning of the simulation, we employed techniques inspired by previous work (Yang and Karplus, 2003; Chodera, 2016). We tested various approaches and concluded that the process described above based on reverse cumulative averaging gave the most reliable and consistent results. We established that, while failing for the linear peptide, the method yielded robust binding free energy estimates for a series of cyclic peptides.

This work represents a rare successful application of an alchemical binding free energy method to the calculation of converged absolute binding energies of protein-peptide complexes (Kilburg and Gallicchio, 2016). This has been possible by employing multi-dimensional replica exchange conformational sampling and by adopting an implicit model of solvation which speeds up equilibration and convergence by removing the fluctuations of the solvent. Statistical errors are further reduced by designing a direct alchemical path from the unbound to bound states of the complex. This is in contrast to double decoupling strategies (Gilson et al., 1997), necessary with explicit solvation, which require going through an intermediate decoupled “vacuum” state of the ligand. Because the free energy changes and the corresponding errors leading to and from the intermediate decoupled state are generally large, double decoupling estimates are affected by slow convergence (Deng and Roux, 2009), especially for large and charged ligands such as these (Gumbart et al., 2012).

Potential of mean force methods (Hénin et al., 2005; Woo and Roux, 2005; Comer et al., 2014; Sandberg et al., 2015; Casanovas et al., 2017; Deng et al., 2017) have been applied to protein-peptide binding (Hénin et al., 2005; Gumbart et al., 2013; Jo et al., 2015; Lapelosa, 2017). These circumvent the difficulties of double decoupling alchemical methods by following the direct physical path of the ligand in and out of the receptor. Nevertheless, slow conformational reorganization remains a serious bottleneck in these calculations (Gan and Roux, 2009). In some cases convergence has been achieved only by imposing stiff conformational restraints, which introduce significant systematic bias unless rigorous and time-consuming treatments are applied (Gumbart et al., 2013). With the exception of weak positional restraints on portions of the receptor backbone distant from the peptide binding site and the wide flat-bottom center of mass tether required by alchemical theory (Gilson et al., 1997), in the present simulations the receptor and the peptides were free to explore a wide variety of conformations so as to properly respond to the applied mutations (Figure 8).

The implicit description of the solution environment, while critical to achieving converged results, is likely less accurate than the more established and tested explicit solvent descriptions. Implicit solvation models, based on the rigorous concept of

the solvent potential of mean force (Roux and Simonson, 1999), are not intrinsically less accurate than other models. However, because they are asked to model a free energy rather than a potential energy, they are more difficult to design and parametrize than explicit models of solvation (Chakavorty et al., 2016). Some implicit solvent implementations, in particular, are known to overemphasize the formation of salt bridges between protein residues (Okur et al., 2008; Wickstrom et al., 2015).

Here we adopt the AGBNP2 implicit solvent model (Gallicchio et al., 2009) which incorporates short-ranged terms and other features designed specifically to tune intermolecular interactions and, in particular, the occurrence of protein salt bridges as seen in experimental NMR structures and explicit solvent simulations. Despite these steps, it is possible that the long-lived intermolecular electrostatic interactions between the termini of the linear peptide and the protein receptor site and the opposing intramolecular interactions between the termini, which prevented convergence of the binding free energy in that case, are artifacts of the AGBNP2 implicit solvent model and the specific choice of fixed protonation state (Ellis et al., 2016; Harris et al., 2017). The model may overestimate the strength of salt bridges or overestimate the free energy barrier separating them from solvent-separated states. The good quantitative agreement between calculated and experimental binding affinities for the cyclic peptides, which also form salt bridges with the protein receptor, suggests that the latter circumstance is more likely. Thus, it appears that the failure for the linear peptide is caused by the slow interconversion rates between competing ion-paired conformations which prevents the reliable estimation of their relative populations and, ultimately, of their relative contribution to the binding free energy.

Specialized conformational sampling techniques based on collective variables (Zheng et al., 2008; Di Leva et al., 2014) and constant pH molecular dynamics (Mongan et al., 2004; Ellis and Shen, 2015) could be applied here to treat these specific slow degrees of freedom related to salt-bridge formation.

While successful in this circumstance, the methodology we employed here for the estimation of protein-peptide binding is not without hurdles. The choice of  $\lambda$  schedule as well as the temperature ladder of the multi-dimensional replica exchange conformational sampling turned out to be critical for the overall realization of convergence. The larger the ligand the more dense the  $\lambda$  and temperature schedules have to be to form enough critical overlap between alchemical states. Careful monitoring of equilibration bias and analysis of convergence proved essential to ensure the robustness and reproducibility of the results in these difficult systems. Furthermore, due to the larger size of the systems, the large number of replicas, and the slow conformational reorganization as compared to small-molecule inhibitors, these calculations are quite expensive and time consuming. We are currently working to port our SDM implementation to GPUs and parallel processors to leverage their high computational power (Zhang et al., 2017).

These hurdles, however, should not be seen as a deterrent. Free energy modeling of protein-peptide binding is an exciting, emerging field with great potential in biomedical research and drug discovery. We have shown here that our

single-decoupling strategy, while not perfect, can be used to calculate converged binding free energies for protein-peptide complexes. Advancements in models, algorithms and computer hardware are progressing steadily and making calculations more efficient and practical.

## 5. CONCLUSIONS

In this work we evaluate a single-decoupling alchemical method (SDM) combined with implicit solvation and advanced conformational sampling strategies toward the calculation of the binding free energies of protein-peptide complexes. We report converged binding free energy estimates of a set of cyclic LEDGF-derived peptides to the HIV1-IN receptor. The binding free energy estimates recapitulate the observed effect of mutations relative to the wild-type binding motif. We show that careful error analysis and monitoring of equilibration and convergence is essential to ensure the reliability of the results. For example, the analysis detected the failure of the method for a linear peptide due to conformational trapping. In overall, the results of this work confirm that, together with advanced conformational sampling strategies, accurate solvation models, careful quality control, and careful

convergence analysis, the single-decoupling alchemical strategy employed here is a viable approach to the quantitative calculation of the binding free energies of protein-peptide complexes.

## AUTHOR CONTRIBUTIONS

DK and EG jointly designed the research and wrote the paper. DK performed the calculations and the numerical analysis.

## ACKNOWLEDGMENTS

The calculations conducted as part of this work have been conducted on the SuperMIC cluster at Louisiana State University supported by a generous computer time allocation (TG-MCB150001) by the NSF XSEDE program.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolf.2018.00022/full#supplementary-material>

## REFERENCES

- Allen, M. P., and Tildesley, D. J. (1993). *Computer Simulation of Liquids*. New York, NY: Oxford University Press.
- Arkin, M. R., Tang, Y., and Wells, J. A. (2014). Small-molecule inhibitors of protein-protein interactions: progressing toward the reality. *Chem. Biol.* 21, 1102–1114. doi: 10.1016/j.chembiol.2014.09.001
- Banks, J. L., Beard, J. S., Cao, Y., Cho, A. E., Damm, W., Farid, R., et al. (2005). Integrated modeling program, applied chemical theory (IMPACT). *J. Comp. Chem.* 26, 1752–1780. doi: 10.1002/jcc.20292
- Basse, M. J., Betzi, S., Bourgeois, R., Bouzidi, S., Chetrit, B., Hamon, V., et al. (2013). 2P2Idb: a structural database dedicated to orthosteric modulation of protein-protein interactions. *Nucl. Acids Res.* 41, D824–D827. doi: 10.1093/nar/gks1002
- Bell, D. R., Qi, R., Jing, Z., Xiang, J. Y., Mejias, C., Schnieders, M. J., et al. (2016). Calculating binding free energies of host-guest systems using the AMOEBA polarizable force field. *Phys. Chem. Chem. Phys.* 18, 30261–30269. doi: 10.1039/C6CP02509A
- Boresch, S., Tettinger, F., Leitgeb, M., and Karplus M. (2003). Absolute binding free energies: a quantitative approach for their calculation. *J. Phys. Chem. B* 107, 9535–9551. doi: 10.1021/jp0217839
- Casasnovas, R., Limongelli, V., Tiwary, P., Carloni, P., and Parrinello, M. (2017). Unbinding kinetics of a p38 map kinase type II inhibitor from metadynamics simulations. *J. Am. Chem. Soc.* 139, 4780–4788. doi: 10.1021/jacs.6b12950
- Cavalli, A., Spitaleri, A., Saladino, G., and Gervasio, F. L. (2014). Investigating drug-target association and dissociation mechanisms using metadynamics-based algorithms. *Acc. Chem. Res.* 48, 277–285. doi: 10.1021/ar500356n
- Chakavorty, A., Li, L., and Alexov, E. (2016). Electrostatic component of binding energy: interpreting predictions from poisson-boltzmann equation and modeling protocols. *J. Comp. Chem.* 37, 2495–2507. doi: 10.1002/jcc.24475
- Chang, C. A., Chen, W., and Gilson, M. K. (2007). Ligand configurational entropy and protein binding. *Proc. Natl. Acad. Sci. U.S.A.* 104, 1534–1539. doi: 10.1073/pnas.0610494104
- Cherepanov, P., Ambrosio, A. L. B., Rahman, S., Ellenberger, T., and Engelman, A. (2005a). Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75. *Proc. Natl. Acad. Sci. U.S.A.* 102, 17308–17313. doi: 10.1073/pnas.0506924102
- Cherepanov, P., Sun, Z.-Y., Rahman, S., Maertens, G., Wagner, G., and Engelman, A. (2005b). Solution structure of the HIV-1 integrase-binding domain in LEDGF/p75. *Nature, Struct. Mol. Biol.* 12, 526–532. doi: 10.1038/nsmb937
- Chodera, J. D., Mobley, D. L., Shirts, M. R., Dixon, R. W., Branson, K., and Pande, V. S. (2011). Alchemical free energy methods for drug discovery: progress and challenges. *Curr. Opin. Struct. Biol.* 21, 150–160. doi: 10.1016/j.sbi.2011.01.011
- Chodera, J. D. (2016). A simple method for automated equilibration detection in molecular simulations. *J. Chem. Theory Comput.* 12, 1799–1805. doi: 10.1021/acs.jctc.5b00784
- Comer, J., Gumbart, J. C., Hénin, J., Lelievre, T., Pohorille, A., and Chipot, C. (2014). The adaptive biasing force method: everything you always wanted to know but were afraid to ask. *J. Phys. Chem. B* 119, 1129–1151. doi: 10.1021/jp506633n
- Deng, Y., and Roux, B. (2009). Computations of standard binding free energies with molecular dynamics simulations. *J. Phys. Chem. B*, 113, 2234–2246. doi: 10.1021/jp807701h
- Deng, N., Wickstrom, L., Cieplak, P., Lin, C., and Yang, D. (2017). Resolving the ligand binding specificity in C-MYC G-Quadruplex DNA: absolute binding free energy calculations and SPR experiment. *J. Phys. Chem. B* 121, 10484–10497. doi: 10.1021/acs.jpcc.7b09406
- Di Leva, F. S., Novellino, E., Cavalli, A., Parrinello, M., and Limongelli, V. (2014). Mechanistic insight into ligand binding to G-quadruplex DNA. *Nucl. Acids Res.* 42, 5447–5455. doi: 10.1093/nar/gku247
- Ellis, C. R., and Shen, J. (2015). pH-dependent population shift regulates BACE1 activity and inhibition. *J. Am. Chem. Soc.* 137, 9543–9546. doi: 10.1021/jacs.5b05891
- Ellis, C. R., Tsai, C.-C., Hou, X., and Shen, J. (2016). Constant pH molecular dynamics reveals pH-modulated binding of two small-molecule BACE1 inhibitors. *J. Phys. Chem. Lett.* 7, 944–949. doi: 10.1021/acs.jpclett.6b00137
- Ensing, B., De Vivo, M., Liu, Z., Moore, P., and Klein, M. L. (2006). Metadynamics as a tool for exploring free energy landscapes of chemical reactions. *Acc. Chem. Res.* 39, 72–81. doi: 10.1021/ar040198i
- Ewing, T. J., Makino, S., Skillman, A. G., and Kuntz, I. D. (2001). DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J. Comp. Aid. Mol. Des.* 15, 411–428. doi: 10.1023/A:1011115820450
- Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., et al. (2004). Glide: a new approach for rapid, accurate docking and scoring.

1. method and assessment of docking accuracy. *J. Med. Chem.* 47, 1739–1749. doi: 10.1021/jm0306430
- Gallicchio, E., and Levy, R. M. (2004). AGBNP: an analytic implicit solvent model suitable for molecular dynamics simulations and high-resolution modeling. *J. Comput. Chem.* 25, 479–499. doi: 10.1002/jcc.10400
- Gallicchio, E., and Levy, R. M. (2011). Recent theoretical and computational advances for modeling protein-ligand binding affinities. *Adv. Prot. Chem. Struct. Biol.* 85, 27–80. doi: 10.1016/B978-0-12-386485-7.00002-8
- Gallicchio, E., Deng, N., He, P., Perryman, A. L., Santiago, D. N., Forli, S., et al. (2014). Virtual screening of integrase inhibitors by large scale binding free energy calculations: the SAMPL4 challenge. *J. Comp. Aided Mol. Des.* 28, 475–490. doi: 10.1007/s10822-014-9711-9
- Gallicchio, E., Lapelosa, M., and Levy, R. M. (2010). Binding energy distribution analysis method (BEDAM) for estimation of protein-ligand binding affinities. *J. Chem. Theory Comput.* 6, 2961–2977. doi: 10.1021/ct1002913
- Gallicchio, E., Paris, K., and Levy, R. M. (2009). The AGBNP2 implicit solvation model. *J. Chem. Theory Comput.* 5, 2544–2564. doi: 10.1021/ct900234u
- Gallicchio, E., Xia, J., Flynn, W. F., Zhang, B., Samlalsingh, S., Mentas, A., et al. (2015). Asynchronous replica exchange software for grid and heterogeneous computing. *Comp. Phys. Comm.* 196, 236–246. doi: 10.1016/j.cpc.2015.06.010
- Gan, W., and Roux, B. (2009). Binding specificity of SH2 domains: insight from free energy simulations. *Proteins: Struct. Fun. Bioinf.* 74, 996–1007. doi: 10.1002/prot.22209
- Gilson, M. K., Given, J. A., Bush, B. L., and McCammon, J. A. (1997). The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys. J.* 72, 1047–1069. doi: 10.1016/S0006-3495(97)78756-3
- Gray, J. J., Moughon, S., Wang, C., Schueler-Furman, O., Kuhlman, B., Rohl, C. A., et al. (2003). Protein-protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. *J. Mol. Biol.* 331, 281–299. doi: 10.1016/S0022-2836(03)00670-3
- Gumbart, J. C., Roux, B., and Chipot, C. (2012). Standard binding free energies from computer simulations: what is the best strategy? *J. Chem. Theory Comput.* 9, 794–802. doi: 10.1021/ct3008099
- Gumbart, J. C., Roux, B., and Chipot, C. (2013). Efficient determination of protein-protein standard binding free energies from first principles. *J. Chem. Theory Comput.* 9, 3789–3798. doi: 10.1021/ct400273t
- Harris, R. C., Tsai, C.-C., Ellis, C. R., and Shen, J. (2017). Proton-coupled conformational allostery modulates the inhibitor selectivity for  $\beta$ -secretase. *J. Chem. Phys. Lett.* 8, 4832–4837. doi: 10.1021/acs.jpcl.7b02309
- Hénin, J., Pohorille, A., and Chipot, C. (2005). Insights into the recognition and association of transmembrane  $\alpha$ -helices. the free energy of  $\alpha$ -helix dimerization in glycoporphin A. *J. Am. Chem. Soc.* 127, 8478–8484. doi: 10.1021/ja050581y
- Higuero, A. P., Schreyer, A., Bickerton, G. R., Pitt, W. R., Groom, C. R., and Blundell, T. L. (2009). Atomic interactions and profile of small molecules disrupting protein-protein interfaces: the TIMBAL database. *Chem. Biol. Drug Des.* 74, 457–467. doi: 10.1111/j.1747-0285.2009.00889.x
- Jo, S., Jiang, W., and Roux, B. (2015). Quantifying protein-protein binding energy and entropy using molecular dynamics simulations. *Biophys. J.* 108:41a. doi: 10.1016/j.bpj.2014.11.251
- Jorgensen, W. L., Maxwell, D. S., and Tirado-Rives, J. (1996). Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* 118, 11225–11236. doi: 10.1021/ja9621760
- Kaminski, G. A., Friesner, R. A., Tirado-Rives, J., and Jorgensen, W. L. (2001). Evaluation and reparameterization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* 105, 6474–6487. doi: 10.1021/jp003919d
- Kastritis, P. L., and Bonvin, A. M. (2013). On the binding affinity of macromolecular interactions: daring to ask why proteins interact. *J. R. Soc. Interface* 10:20120835. doi: 10.1098/rsif.2012.0835
- Kilburg, D., and Gallicchio, E. (2016). Recent advances in computational models for the study of protein-peptide interactions. *Adv. Prot. Chem. Struct. Biol.* 105, 27–57. doi: 10.1016/bs.apcsb.2016.06.002
- Kim, J., Keyes, T., and Straub, J. E. (2010). Generalized replica exchange method. *J. Chem. Phys.* 132:224107. doi: 10.1063/1.3432176
- Klimovich, P. V., Shirts, M. R., and Mobley, D. L. (2015). Guidelines for the analysis of free energy calculations. *J. Comp. Aid. Mol. Des.* 29, 397–411. doi: 10.1007/s10822-015-9840-9
- Kozakov, D., Brenke, R., Comeau, S. R., and Vajda, S. (2006). PIPER: an FFT-based protein docking program with pairwise potentials. *Proteins: Struct. Fun. Bioinf.* 65, 392–406. doi: 10.1002/prot.21117
- Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A., and Rosenberg, J. M. (1992). The weighted histogram analysis method for free-energy calculations on biomolecules. I. the method. *J. Comput. Chem.* 13, 1011–1021. doi: 10.1002/jcc.540130812
- Labb, C. M., Laconde, G., Kuenemann, M. A., Villoutreix, B. O., and Sperandio, O. (2013). iPPI-DB: a manually curated and interactive database of small non-peptide inhibitors of protein-protein interactions. *Drug. Discov. Today* 18, 958–968. doi: 10.1016/j.drudis.2013.05.003
- Lapelosa, M., Gallicchio, E., and Levy, R. M. (2012). Conformational transitions and convergence of absolute binding free energy calculations. *J. Chem. Theory Comput.* 8, 47–60. doi: 10.1021/ct200684b
- Lapelosa, M. (2017). Free energy of binding and mechanism of interaction for the MEEVD-TPR2A peptide-protein complex. *J. Chem. Theory Comput.* 13, 4514–4523. doi: 10.1021/acs.jctc.7b00105
- Lavecchia, A., and Di Giovanni, C. (2013). Virtual screening strategies in drug discovery: a critical review. *Curr. Med. Chem.* 20, 2839–2860. doi: 10.2174/09298673113209990001
- Lindert, S., Bucher, D., Eastman, P., Pande, V., and Andrew McCammon, J. (2013). Accelerated molecular dynamics simulations with the amoeba polarizable force field on graphics processing units. *J. Chem. Theory Comput.* 9, 4684–4691. doi: 10.1021/ct400514p
- London, N., Movshovitz-Attias, D., and Schueler-Furman, O. (2010). The structural basis of peptide-protein binding strategies. *Structure* 18, 188–199. doi: 10.1016/j.str.2009.11.012
- Mongan, J., Case, D. A., and McCammon, J. A. (2004). Constant pH molecular dynamics in generalized Born implicit solvent. *J. Comput. Chem.* 25, 2038–2048. doi: 10.1002/jcc.20139
- Murray, J. M., Emery, S., Kelleher, A. D., Law, M., Chen, J., Hazuda, D. J., et al. (2007). Antiretroviral therapy with the integrase inhibitor Raltegravir alters decay kinetics of HIV, significantly reducing the second phase. *Aids* 21, 2315–2321. doi: 10.1097/QAD.0b013e3282f12377
- Okur, A., Wickstrom, L., and Simmerling, C. (2008). Evaluation of salt bridge structure and energetics in peptides using explicit, implicit, and hybrid solvation models. *J. Chem. Theory Comput.* 4, 488–498. doi: 10.1021/ct7002308
- Pal, R. K., Haider, K., Kaur, D., Flynn, W., Xia, J., Levy, R. M., et al. (2016). A combined treatment of hydration and dynamical effects for the modeling of host-guest binding thermodynamics: the SAMPL5 blinded challenge. *J. Comp. Aided Mol. Des.* 31, 29–44. doi: 10.1007/s10822-016-9956-6
- Peat, T. S., Dolezal, O., Newman, J., Mobley, D., and Deadman, J. J. (2014). Interrogating HIV integrase for compounds that bind—a SAMPL challenge. *J. Comp. Aided Mol. Des.* 28, 347–362. doi: 10.1007/s10822-014-9721-7
- Perryman, A. L., Santiago, D. N., Forli, S., Martins, D. S., and Olson, A. J. (2014). Virtual screening with autodock vina and the common pharmacophore engine of a low diversity library of fragments and hits against the three allosteric sites of HIV Integrase: participation in the SAMPL4 protein-ligand binding challenge. *J. Comp. Aided Mol. Des.* 28, 1–13. doi: 10.1007/s10822-014-9709-3
- Pierce, B. G., Wiehe, K., Hwang, H., Kim, B. H., Vreven, T., and Weng, Z. (2014). ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics* 30, 1771–1773. doi: 10.1093/bioinformatics/btu097
- Procacci, P., Bizzarri, M., and Marsili, S. (2013). Energy-driven undocking (EDU-HREM) in solute tempering replica exchange simulations. *J. Chem. Theory Comput.* 10, 439–450. doi: 10.1021/ct400809n
- Rhodes, D. I., Peat, T. S., Vandegraaff, N., Jeevarajah, D., Rhodes, J., Peat, D. I., et al. (2011). Crystal structures of novel allosteric peptide inhibitors of HIV Integrase identify new interactions at the LEDGF binding site. *ChemBioChem* 12, 2311–2315. doi: 10.1002/cbic.201100350
- Roux, B., and Simonson, T. (1999). Implicit solvent models. *Biophys. Chem.* 78, 1–20. doi: 10.1016/S0301-4622(98)00226-9
- Sandberg, R. B., Banchelli, M., Guardiani, C., Menichetti, S., Caminati, G., and Procacci, P. (2015). Efficient nonequilibrium method for binding free energy calculations in molecular dynamics simulations. *J. Chem. Theory Comput.* 11, 423–435. doi: 10.1021/ct500964e

- Sarkis, C., Philippe, S., Mallet, J., and Serguera, C. (2008). Non-integrating lentiviral vectors. *Curr. Gene. Ther.* 8, 430–437. doi: 10.2174/156652308786848012
- Shirts, M. R., and Chodera, J. D. (2008). Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* 129:124105. doi: 10.1063/1.2978177
- Shirts, M. R. and Mobley, D. L. (2013). An introduction to best practices in free energy calculations. *Methods Mol. Biol.* 924, 271–311. doi: 10.1007/978-1-62703-017-5\_11
- Shoemaker, B. A., and Panchenko, A. R. (2007). Deciphering protein–protein interactions. part ii. computational methods to predict protein and domain interaction partners. *PLoS Comput. Biol.* 3:e43. doi: 10.1371/journal.pcbi.0030043
- Smith, J. A., and Daniel, R. (2006). Following the path of the virus: the exploitation of host DNA repair mechanisms by retroviruses. *ACS Chem. Biol.* 1, 217–226. doi: 10.1021/cb600131q
- Summa, V., Petrocchi, A., Bonelli, F., Crescenzi, B., Donghi, M., Ferrara, M., et al. (2008). Discovery of Raltegravir, a potent, selective orally bioavailable HIV-Integrase inhibitor for the treatment of HIV-AIDS infection. *J. Med. Chem.* 51, 5843–5855. doi: 10.1021/jm800245z
- Tan, Z., Gallicchio, E., Lapelosa, M., and Levy, R. M. (2012). Theory of binless multi-state free energy estimation with applications to protein–ligand binding. *J. Chem. Phys.* 136:144102. doi: 10.1063/1.3701175
- Tsiang, M., Jones, G. S., Hung, M., Mukund, S., Han, B., Liu, X., et al. (2009). Affinities between the binding partners of the HIV-1 integrase dimer–lens epithelium–derived growth factor (IN dimer–LEDGF) complex. *J. Biol. Chem.* 284, 33580–33599. doi: 10.1074/jbc.M109.040121
- Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., and Taylor, R. D. (2003). Improved protein–ligand docking using GOLD. *Proteins: Struct. Fun. Bioinf.* 52, 609–623. doi: 10.1002/prot.10465
- Wang, I., Berne, B. J., and Friesner, R. A. (2012). On achieving high accuracy and reliability in the calculation of relative protein–ligand binding affinities. *Proc. Natl. Acad. Sci. U.S.A.* 109, 1937–1942. doi: 10.1073/pnas.1114017109
- Wang, L., Wu, Y., Deng, Y., Kim, B., Pierce, L., Krilov, G., et al. (2015). Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J. Am. Chem. Soc.* 137, 2695–2703. doi: 10.1021/ja512751q
- Wickstrom, L., Deng, N., He, P., Montes, A., Nguyen, C., Gilson, M. K., et al. (2015). Parameterization of an effective potential for protein–ligand binding from host–guest affinity data. *J. Mol. Recognition* 29, 10–21. doi: 10.1002/jmr.2489
- Woo, H.-J., and Roux, B. (2005). Calculation of absolute protein–ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. U.S.A.* 102, 6825–6830. doi: 10.1073/pnas.0409005102
- World Health Organization (2017). *10 facts on HIV/AIDS*. Available online at: <http://www.who.int/features/factfiles/hiv/en/>.
- Yan, X., Liao, C., Liu, Z., Hagler, A. T., Gu, Q., and Xu, J. (2016). Chemical structure similarity search for ligand-based virtual screening: methods and computational resources. *Curr. Drug Targets* 17, 1580–1585. doi: 10.2174/1389450116666151102095555
- Yang, W., Bitetti-Putzer, R., and Karplus, M. (2003). Free energy simulations: use of the reverse cumulative averaging to determine the equilibrated region and the time required for convergence. *J. Chem. Phys.* 120, 2618–2628. doi: 10.1063/1.1638996
- You, W., Huang, Y. M., Kizhake, S., Natarajan, A., and Chang, C. E. (2016). Characterization of promiscuous binding of phosphor ligands to breast-cancer-gene 1 (BRCA1) C-terminal (BRCT): molecular dynamics, free energy, entropy and inhibitor design. *PLoS Comp. Biol.* 12:e1005057. doi: 10.1371/journal.pcbi.1005057
- Zhang, B., Kilburg, D., Eastman, P., Pande, V. S., and Gallicchio, E. (2017). Efficient gaussian density formulation of volume and surface areas of macromolecules on graphical processing units. *J. Comp. Chem.* 38, 740–752. doi: 10.1002/jcc.24745
- Zhang, Q. C., Petrey, D., Deng, L., Qiang, L., Shi, Y., Thu, C. A., et al. (2012). Structure-based prediction of protein–protein interactions on a genome-wide scale. *Nature* 490, 556–560. doi: 10.1038/nature11503
- Zheng, L., Chen, M., and Yang, W. (2008). Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems. *Proc. Natl. Acad. Sci. U.S.A.* 105, 20227–20232. doi: 10.1073/pnas.0810631106
- Zhou, Z., Felts, A. K., Friesner, R. A., and Levy, R. M. (2007). Comparative performance of several flexible docking programs and scoring functions: enrichment studies for a diverse set of pharmaceutically relevant targets. *J. Chem. Inf. Model.* 47, 1599–1608. doi: 10.1021/ci7000346
- Zuckerman, D. M., and Chong, L. T. (2017). Weighted ensemble simulation: review of methodology, applications, and software. *Ann. Rev. Biophys.* 46, 43–57. doi: 10.1146/annurev-biophys-070816-033834

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Kilburg and Gallicchio. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.