Published in final edited form as:
J Chem Theory Comput. 2023 August 08; 19(15): 4863–4882. doi:10.1021/acs.jctc.3c00333.

Identifying and overcoming the sampling challenges in relative binding free energy calculations of a model protein:protein complex

Ivy Zhang[1,2], **Dominic A. Rufa**[1,3], **Iván Pulido**[1], **Michael M. Henry**[1], **Laura E. Rosen**[4], **Kevin Hauser**[4], **Sukrit Singh**[1,*], **John D. Chodera**[1,*]

[1]Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065;

[2]Tri-Institutional PhD Program in Computational Biology and Medicine, Weill Cornell Medical College, Cornell University, New York, NY 10065;

[3]Tri-Institutional PhD Program in Chemical Biology, Weill Cornell Medical College, Cornell University, New York, NY 10065;

[4]Vir Biotechnology, San Francisco, CA, USA

Abstract

Relative alchemical binding free energy calculations are routinely used in drug discovery projects to optimize the affinity of small molecules for their drug targets. Alchemical methods can also be used to estimate the impact of amino acid mutations on protein:protein binding affinities, but these calculations can involve sampling challenges due to the complex networks of protein and water interactions frequently present in protein:protein interfaces. We investigate these challenges by extending a GPU-accelerated open-source relative free energy calculation package (Perses) to predict the impact of amino acid mutations on protein:protein binding. Using the well-characterized model system barnase:barstar, we describe analyses for identifying and characterizing sampling problems in protein:protein relative free energy calculations. We find that mutations with sampling problems often involve charge-changes, and inadequate sampling can be attributed to slow degrees of freedom that are mutation-specific. We also explore the accuracy and efficiency of current state-of-the-art approaches—alchemical replica exchange and alchemical replica exchange with solute tempering—for overcoming relevant sampling problems. By employing sufficiently long simulations, we achieve accurate predictions (RMSE 1.61, 95% Cl: [1.12, 2.11] kcal/mol), with 86% of estimates within 1 kcal/mol of the experimentally-determined relative binding free energies and 100% of predictions correctly classifying the sign

[*]**For correspondence:** john.chodera@choderalab.org (JDC); sukrit.singh@choderalab.org (SS).
Author Contributions
Conceptualization: IZ, JDC; Data Curation: IZ; Formal Analysis: IZ; Funding Acquisition: JDC, LER; Investigation: IZ; Methodology: IZ, DAR, JDC, KH; Project Administration: JDC, LER, SS; Resources: JDC, LER; Software: IZ, DAR, IP, MMH; Supervision: JDC, LER, SS; Visualization: IZ; Writing - Original Draft: IZ; Writing - Review & Editing: IZ, JDC, SS, DAR, LER, KH, IP, MMH

Disclaimers
The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

of the changes in binding free energies. Ultimately, we provide a model workflow for applying protein mutation free energy calculations to protein:protein complexes, and importantly, catalog the sampling challenges associated with these types of alchemical transformations. Our free open-source package (Perses) is based on OpenMM and available at https://github.com/choderalab/perses.

INTRODUCTION

Predicting the impact of amino acid mutations on protein:protein binding has important applications

Protein:protein interactions (PPIs) underlie fundamental biological processes, such as transcriptional regulation (e.g., p53:MDM2 [1]), signal transduction (e.g., GRB2-EGFR [2]), and membrane fusion (e.g., SARS-CoV-2 RBD:ACE2 [3]). As protein:protein interactions are driven by binding events, changes in protein:protein binding affinity often have functional impact. Even a single amino acid substitution can significantly alter binding and function, which can give rise to disease [4], impact the fitness of a pathogen [5], or alter the activity of monoclonal antibody drugs [6]. Thus, quantifying the impact of an amino acid mutation on protein:protein binding is highly useful for predicting the functional implication of a mutation, as it can provide mechanistic understanding of disease-associated genetic variants [7, 8] and facilitate the design of biologic drugs such as monoclonal antibodies [9, 10].

Alchemical free energy calculations represent an accurate and generalizable approach for estimating mutational impact on PPIs

There are many experimental and computational approaches for quantifying the impact of a mutation on protein:protein binding. Experimental approaches, while highly accurate and generally considered “ground truth,” can be resource intensive, often requiring significant amounts of human labor time and costly reagents and instruments [11–14]. Computational methods, which can circumvent these resource challenges, serve as a complementary approach that can be combined with experimental methods to enable more efficient acquisition of high confidence data. Examples include computationally inexpensive methods (e.g., MM/PBSA and MM/GBSA [15, 16], machine learning (ML) models [17, 18], and Rosetta-based methods [19, 20]), as well as computationally expensive methods such as alchemical free energy calculations. Several studies have compared the tradeoffs between using computationally inexpensive methods and alchemical free energy calculations to predict the impact of a point mutation on binding [21–25]. While computationally inexpensive methods can be accurate for certain systems [21, 22], these methods often fail to account for key biophysical phenomena (i.e., conformational heterogeneity, explicit solvent interactions, multiple protonation states), and therefore perform worse on systems which require modeling of these properties [23–25]. Despite their increased computational cost, alchemical free energy calculations can account for these biophysical phenomena using rigorous statistical mechanics, so they tend to demonstrate better accuracy than the cheaper methods and are generalizable to any PPI with available structural data [24–27]. Ultimately, the optimal choice in approach will depend on the scientific goal and the computational resources available. However, given their accuracy and generalizability, as well as rapid advancements in graphics processing units (GPUs) that have made it feasible to carry out these calculations in reasonable timeframes [28–31], alchemical free energy calculations represent a highly promising approach for predicting mutational effect on protein:protein binding.

Relative alchemical binding free energy calculations aim to predict the impact of a mutation on the free energy of binding ($\Delta\Delta G_{binding}$)

While there are numerous types of alchemical free energy calculations [32], relative alchemical binding free energy (RBFE) calculations estimate the relative binding free energies ($\Delta\Delta G_{binding}$) between two chemically similar complexes, e.g., protein:protein complexes that differ by an amino acid mutation (Figure 1A). In protein mutation RBFE calculations [21, 23–25], the wild-type (WT) residue is transformed into the mutant residue through molecular dynamics simulations of alchemical (non-physical) intermediate states bridging the WT and mutant states (Figure 2A). This alchemical transformation is performed in two phases, complex and apo, which correspond to the mutating protein in the presence and absence of a protein binding partner, respectively (Figure 1A). The change in free energy associated with each phase (ΔG_{phase}) is estimated, and the difference in the ΔG_{phase}S gives an estimate of the impact of the mutation on the binding free energy, $\Delta\Delta G_{binding}$ (Figure 1A).

Achieving sufficient sampling of protein and water conformations is particularly challenging for RBFE calculations applied to protein:protein interactions

During the last couple of decades, RBFE calculations have become increasingly widely used in drug discovery projects for predicting the effects of small molecule modifications on protein:small molecule binding [31, 34–40]. In comparison to small molecule transformations, application of RBFE calculations to protein mutations has been relatively limited, though recent studies have demonstrated that these methods can accurately predict mutational impact on protein:small molecule binding [21–23, 41] and protein:protein binding [24, 25, 42–46] for a number of biologically-relevant complexes.

One reason for their lack of widespread use stems from the size of protein:protein complexes, which can frequently involve approximately double the number of atoms as are present in protein:small molecule complexes, making them more computationally expensive to simulate. However, the bigger hurdle has been the sampling challenges associated with alchemical transformations in protein:protein complexes [24, 47]. RBFE calculations require drawing decorrelated samples from the configurational probability distributions at each alchemical state [48], a nontrivial task for protein:protein complexes because the energy landscapes often contain many minima which can give rise to slow degrees of freedom. Slow degrees of freedom are more prevalent in protein:protein complexes because protein:protein interfaces are generally broader than protein:small molecule interfaces and typically involve complex protein and water interaction networks [4, 49]. Upon mutation, extensive reorganization of the mutating residue along with its closely-packed neighborhood of interfacial protein residues and waters may be required before one can draw decorrelated samples.

Pinpointing slow degrees of freedom can help address the sampling problems in protein:protein RBFE calculations, but existing approaches are not automated

Determining the slow degrees of freedom causing sampling problems in alchemical free energy calculations is useful because it may enable improved sampling via methods that accelerate known slow degrees of freedom (e.g., metadynamics [50, 51], umbrella sampling [52], adaptive biasing force [53]). Moreover, identifying slow degrees of freedom helps enumerate the common challenges associated with alchemical transformations and the limitations of existing methods, which will facilitate the improvement of existing methods or the design of new ones. However, pinpointing the slow degrees of freedom in protein:protein interfaces can be challenging because it typically involves careful manual inspection of simulation trajectories [33, 54, 55], a process which requires biophysical intuition and can be tedious even for experienced practitioners. Moreover, the manual inspection approach is not scalable as examining the trajectories for tens or hundreds of mutations would be extremely impractical.

Here, we investigate sampling problems associated with protein mutation relative free energy calculations using (1) terminally-blocked amino acids, a small and simple test system relatively free of interfacial complexities and (2) barnase:barstar, a well-studied protein:protein complex. We augment an existing open-source relative free energy calculation package (Perses [56], https://github.com/choderalab/perses) to carry out these calculations and describe experiments and automated analyses that identify likely causes of sampling problems. We find that sampling challenges are more likely to occur for charge-changing mutations and can be attributed to mutation-dependent slow degrees of freedom. We also compare the accuracy and efficiency of state-of-the-art enhanced sampling approaches-alchemical replica exchange (AREX) [31, 57, 58] and alchemical replica exchange with solute tempering (AREST) [59, 60]—for overcoming the sampling challenges. We find that given sufficient simulation time, our predictions are accurate with respect to experiment (RMSE 1.61, 95% Cl: [1.12, 2.11] kcal/mol), with 86% of predictions lying within 1 kcal/mol of experimental $\Delta\Delta G_{\text{binding}}$ s and 100% of predictions having the correct sign.

THEORY

We perform alchemical free energy calculations using two state-of-the-art enhanced sampling approaches: (1) alchemical replica exchange (AREX) [31, 57, 58], the current recommended approach based on best practices [27] and (2) alchemical replica exchange with solute tempering (AREST) [59, 60], a sampling scheme which builds upon AREX by increasing the temperature of a region around the mutating residue and has been shown to improve sampling over AREX for some transformations [60–62]. Here, we give a brief overview of the salient aspects of each method, as well as the general approach we take to alchemical free energy calculations for protein mutations. The alchemical approach is implemented in an open-source package (Perses [56], available at https://github.com/choderalab/perses). Complete simulation details can be found in the Detailed Methods.

Alchemical transformation

Alchemical free energy calculations aim to sample a set of alchemical states which are defined such that two endstates of interest are bridged by alchemical intermediate states with modified Hamiltonians. The Hamiltonians are modified such that the nonbonded interactions (and potentially valence terms) of the WT and mutant residues are gradually transformed between interacting and non-interacting. The first alchemical state, called the WT endstate, typically involves the WT residue fully interacting with its environment and the mutant residue completely non-interacting. The last alchemical state, known as the mutant endstate, involves the mutant residue fully interacting and the WT residue non-interacting. A series of intermediate alchemical states bridging the endstates is defined such that WT and mutant residues are partially interacting with their environments to varying extents. Taken together, these alchemical states form the alchemical transformation (Figure 2A). Note that amino acids with backbone cycles (such as proline) require a modified version of this approach to avoid the noninteracting residue influencing the conformational distribution of fully interacting residues [63], but the mutations in this study do not involve this type of amino acid.

To characterize an alchemical transformation, we need to specify both *which* interactions will be alchemically modified during the transformation and *how* we will modify them. We first identify the alchemical interactions by defining an atom mapping, which exploits the partial similarity in WT and mutant topologies by pairing up the WT and mutant atoms that will share coordinates. The atom mapping is then used to classify each atom into an “atom class”: “unique old” atoms are unmapped atoms that are only present in the WT residue, “unique new” atoms are unmapped atoms that are only present in the mutant residue, “core” atoms are mapped atoms that are shared between the WT and mutant residues (and include atoms in the residues immediately preceding and following the mutating residue), and “environment” atoms are mapped atoms that are shared between the topologies but lie outside of the core atoms. Interactions involving “unique old”, “unique new” and “core” atoms are considered alchemical interactions. Since increasing the number of alchemical interactions also increases the thermodynamic length (i.e., the distance between alchemical states) [64], an atom mapping should be defined to maximize the number of atoms mapped between the two residues, which minimizes thermodynamic length. An optimal mapping finds a balance between minimizing thermodynamic length and taking advantage of the built-in enhanced sidechain sampling that occurs when the unmapped atoms are non-interacting (i.e., the nonbonded interactions are scaled to zero, so the sidechains can more easily sample alternate rotameric states). To this end, we chose an atom mapping that maps all atoms between the WT and mutant residues up to and including the beta carbon (but not including beta hydrogens). Because we constrain bonds to hydrogen, we un-map any hydrogen atoms whose bond lengths would change between WT and mutant.

To specify how the energies should be modified during the alchemical transformation, we introduce an alchemical parameter $\lambda \in [0,1]$ into the potential energy function $U(x)$, forming the alchemical potential energy function $U(x; \lambda)$. The alchemical potential $U(x; \lambda)$ is typically evaluated at a different λ value for each alchemical state. For the WT endstate, $U(x; \lambda = 0)$ is identical to the unmodified WT potential with the addition of the standard valence terms (and not the nonbonded interactions) of the unique new atoms (sometimes called “dummy” atoms). For the mutant endstate, $U(x; \lambda = 1)$ is identical to the unmodified mutant potential, but with unique old atoms as noninteracting dummy atoms that only retain their valence terms. For the alchemical intermediate states $(\lambda \in (0,1))$, $U(x; \lambda)$ is a modified potential where interactions involving the WT and mutant residues are scaled to varying extents. The set of λ values sampled in the alchemical transformation is termed the “alchemical protocol”. To define the alchemical protocol for this study, we selected evenly spaced λ values from a simple linear function (Supplementary Figure 1A). The number of λ values used for each calculation was chosen such that the neighboring alchemical states have good phase space overlap (for more details, see Detailed Methods), which permits robust estimation of free energy differences between the fully-interacting WT and mutant endstates [65, 66].

Using the alchemical parameter λ, we define the potential energy functions for electrostatic and steric interactions. We compute the alchemically modified electrostatics interaction energy according to the Particle Mesh Ewald (PME) method [67] with linearly interpolated charges. For the direct space electrostatics contribution:

$$\begin{aligned} U_{\text{direct}}(r; \lambda) &= C\frac{q_i(\lambda)q_j(\lambda)}{r_{\text{eff}}(r, \lambda)} \cdot \mathsf{erfc}(\alpha r_{\text{eff}}(r, \lambda)) \\ q_i(\lambda) &= \chi_i^{\text{old}} q_i^{\text{old}}(\lambda) + \chi_i^{\text{new}} q_i^{\text{new}}(\lambda) + \chi_i^{\text{core}}\left(q_i^{\text{old}}(\lambda) + q_i^{\text{new}}(\lambda)\right) + \chi_i^{\text{env}} q_i^{\text{old}} \\ q_i^{\text{old}}(\lambda) &= (1 - \lambda) q_i^{\text{old}} \\ q_i^{\text{new}}(\lambda) &= \lambda q_i^{\text{new}} \end{aligned} \tag{1}$$

where α is an internal PME parameter (calculated based on the PME error tolerance and the cutoff distance, with dimension 1/length), C is the Coulomb constant (with dimension energy/length^2), $q_i(\lambda)$ and $q_j(\lambda)$ are the functions for computing the potentially alchemically modified charges of atoms i and j (with dimension of charge), and q_i^{old} and q_i^{new} are the charges of atom i in the old topology and new topology, respectively. χ_i^{old}, χ_i^{new}, χ_i^{core}, and χ_i^{env} are indicator functions denoting whether atom i belongs in the unique old, unique new, core, and environment atom classes, respectively.

$r_{\text{eff}}(r, \lambda)$ denotes the e*ffective interaction distance* used for computing the interaction energy, and depends on both the actual inter-particle separation r and the alchemical parameter λ. To avoid singularities in the computation of electrostatics (and sterics) energies, we use a softcore approach that involves “lifting” certain inter-atomic distances into the “4th dimension”, inspired by work of Pomès [68]:

$$\begin{aligned} r_{\text{eff}}(r, \lambda) &= \sqrt{r^2 + w(\lambda)^2} \\ w(\lambda) &= w_{\text{lifting}} \cdot \left(\chi_{ij}^{\text{old}} \lambda + \chi_{ij}^{\text{new}}(1 - \lambda)\right) \\ \chi_{ij}^{\text{old}} &= \begin{cases} 1 & \chi_i^{\text{old}} + \chi_j^{\text{old}} \geq 0 \\ 0 & else \end{cases} \\ \chi_{ij}^{\text{new}} &= \begin{cases} 1 & \chi_i^{\text{new}} + \chi_j^{\text{new}} \geq 0 \\ 0 & else \end{cases} \end{aligned}$$

(2)

where r is the distance between atoms i and $j, w(\lambda)$ is the function for computing the lifting distance, and w_{lifting} is the maximal lifting distance, which was selected to minimize the number of alchemical states needed to produce robust free energy estimates while maintaining good overlap among neighboring alchemical states. χ_{ij}^{old} is an indicator function that assumes the value of unity only when at least one of the atoms (i and j) belongs in the unique old atom class (and zero otherwise), and χ_{ij}^{new} is an indicator function indicating whether at least one of the atoms (i and j) belongs in the unique new atom class. For more details on this approach, see Detailed Methods.

We compute the PME reciprocal space and self-energy contributions using the default energy functions in OpenMM [29], but with linearly interpolated charges, where interpolation was performed in the same manner as was done for the direct space.

We compute the alchemically modified sterics interaction energy as a standard Lennard-Jones 12–6 potential [69–71] with linearly interpolated σ and ϵ and "lifted" interaction distances to create a softcore potential:

$$\begin{aligned} U_{\text{sterics}}(r;\lambda) &= 4\epsilon_{ij}\left(\lambda\right)x\left(x-1.0\right);\, x=\left(\frac{\sigma_{ij}(\lambda)}{r_{\text{eff}}(r,\lambda)}\right)^{6} \\ \sigma_{ij}(\lambda) &= \frac{\sigma_i(\lambda)+\sigma_j(\lambda)}{2};\, \sigma_i\left(\lambda\right) = \chi_i^{\text{old}}\sigma_i^{\text{old}} + \chi_i^{\text{new}}\sigma_i^{\text{new}} + \chi_i^{\text{core}}\left(\left(1-\lambda\right)\sigma_i^{\text{old}} + \lambda\sigma_i^{\text{new}}\right) + \chi_i^{\text{env}}\sigma_i^{\text{old}} \\ \epsilon_{ij}(\lambda) &= \sqrt{\epsilon_i(\lambda)\cdot\epsilon_j(\lambda)};\, \epsilon_i\left(\lambda\right) = \chi_i^{\text{old}}\epsilon_i^{\text{old}}\left(\lambda\right) + \chi_i^{\text{new}}\epsilon_i^{\text{new}}\left(\lambda\right) + \chi_i^{\text{core}}\left(\epsilon_i^{\text{old}}\left(\lambda\right) + \epsilon_i^{\text{new}}\left(\lambda\right)\right) + \chi_i^{\text{env}}\epsilon_i^{\text{old}}\left(\lambda\right) \\ \epsilon_i^{\text{old}}\left(\lambda\right) &= \left(1-\lambda\right)\epsilon_i^{\text{old}};\, \epsilon_i^{\text{new}}\left(\lambda\right) = \lambda\epsilon_i^{\text{new}} \end{aligned} \tag{3}$$

Here, $\sigma_{ij}(\lambda)$ is the function for computing the potentially alchemically modified distance at which the interaction energy crosses zero for atoms i and j . σ_i^{old} and σ_i^{new} are the distances at which the energy equals zero for atom i in the old topology and new topology, respectively. $\epsilon_{ij}(\lambda)$ is the function for computing the potentially alchemically modified interaction strength for atoms i and j. ϵ_i^{old} and ϵ_i^{new} are the interaction strengths for atom i in the old topology and new topology, respectively.

For charge-changing mutations, we ensure the system remains electrostatically neutral by transforming a water molecule in the WT system into a sodium or chloride ion in the mutant system. The ΔΔGs for charge-changing mutations in terminally-blocked amino acids are internally consistent, indicating that in the absence of sampling problems, our counterion scheme enables robust estimation of free energies (Figure 3A). Further details on this implementation can be found in Detailed Methods.

We estimate free energy differences using the Multistate Bennett Acceptance Ratio (MBAR), which is an asymptotically unbiased estimator that, in the large sample limit, often has lower variance compared to other commonly used estimators [66].

Alchemical replica exchange (AREX)

Alchemical free energy calculations must sample from a chain of alchemical intermediate states bridging the two endstates of interest (Figure 2A). Because the introduction or deletion of bulky residues can often frustrate sampling within alchemical states in which these residues are almost fully interacting, alchemical free energy calculations often use replica exchange simulations to help reduce correlation times and over-come sampling challenges. Replica exchange enhances sampling by allowing each replica to visit multiple alchemical states, including those states which may help more rapidly decorrelate slow degrees of freedom because of their modified Hamiltonians [31,57,58]. Here, we refer to this approach as alchemical replica exchange (AREX).

AREX can be thought of as a Markov Chain Monte Carlo (MCMC) algorithm that aims to generate equilibrium samples from a family of K probability densities corresponding to the K alchemical states:

$$x_k \sim p(x_k \mid s_k) \propto \exp[-u_{s_k}(x)] k = 1, \ldots, K \tag{4}$$

where x_k is a configuration drawn from state k, s_k is the k^{th} state, and $u_{s_k}(x)$ is the potential energy of sample x at state s_k.

To generate equilibrium samples, AREX utilizes weakly-coupled replicas (copies of the system of interest), where the number of replicas is typically equal to the number of alchemical states K. AREX employs a Gibbs sampling framework where in each iteration n, the positions $X_n = \{x_k\}_{k=1}^{K}$ of all K replicas are first updated with molecular dynamics simulations, yielding X_{n+1}, and then the permutation set of alchemical state indices $S_n = \{s_k\}_{k=1}^{K}$ associated with the corresponding positions are updated based on the updated positions X_{n+1}, yielding S_{n+1}:

$$\begin{aligned} X_{n+1} &\sim P(X_{n+1} \mid X_n, S_n) \\ S_{n+1} &\sim P(S_{n+1} \mid X_{n+1}) \end{aligned} \tag{5}$$

In sufficiently long simulations, the resulting samples (X_n, S_n) are distributed with respect to the joint probability density $P(X_n, S_n)$ such that

$$P\left(X, S\right) \equiv \prod_{k=1}^{K} p(x_k \mid s_k) \tag{6}$$

The algorithm updates the state indices by exchanging the alchemical state labels for pairs of replicas according to a Metropolis criterion that compares the energies of the two replicas considered for swapping. The replica swap acceptance rate will depend on how well the alchemical states overlap, i.e., the thermodynamic length between states (Figure 2B). Numerous methods can be used to update the states S, including attempting exchanges only between replicas visiting neighboring thermodynamic states (where state overlap is highest). Here, we use a simple strategy that attempts to draw an independent permutation S_{n+1} given configuration X_{n+1} by attempting many swaps of pairs of alchemical state indices, which has been shown to enhance mixing and reduce correlation times [72].

AREX involves running many cycles of molecular dynamics followed by exchange attempts with the goal of ensuring all replicas ultimately perform a random walk through all alchemical states (Figure 2B). If the exchanges are accepted at a sufficiently high rate over the course of the simulation, we expect to observe improved sampling because configurational correlation times associated with the alchemical region are likely decreased for alchemical states with partially interacting residues.

Alchemical replica exchange with solute tempering (AREST)

AREST is AREX with an added layer of sophistication that aims to enhance sampling to a greater extent than AREX. AREST involves running AREX with a REST (replica exchange solute tempering [59]) region, a user-defined set of atoms for which the effective temperature is increased in alchemical intermediate states (Figure 2C–D). Therefore, in AREST, the alchemical states do not solely differ by the extent to which the WT and mutant residues are interacting with their environment, they also differ by the effective temperature of the REST region. Although introducing differences in effective temperature will increase the thermodynamic length between alchemical states (Figure 2C), the goal is to decrease the correlation time of the slowest degrees of freedom sufficiently to compensate for the decrease in state overlap, yielding more decorrelated samples in the same amount of total simulation time.

To incorporate REST into AREX, we classify each bond, angle, torsion, and nonbonded interaction as “REST” (all atoms in the interaction are part of the REST region), “inter” (at least one atom is part of the REST region and at least one atom is not), or “non-REST” (none of the atoms are part of the REST region) based on an initial conformation. Each interaction energy is multiplied by a scale factor depending on the REST class. Therefore, total potential energy is defined as:

$$\begin{gathered} u_{\text{total}}(\lambda) = \alpha(\lambda, T_{\max}, T_0) \cdot u_{\text{rest}}(\lambda) + \sqrt{\alpha(\lambda, T_{\max}, T_0)} \cdot u_{\text{inter}}(\lambda) + u_{\text{nonrest}}(\lambda) \\ \alpha(\lambda, T_{\max}, T_0) \propto T_0 / T_{\max} \end{gathered} \tag{7}$$

where T_0 is the temperature of the desired distribution and $T_{\max}$ is the user-selected maximum effective temperature. The function we use to define the REST scale factor, $\alpha(\lambda, T_{\max}, T_0)$, is shown in Supplementary Figure 1B. Note that when $\lambda = 0$ or $1, \alpha(\lambda, T_{\max}, T_0) = 1$ to ensure the endstates are unscaled.

METHODS AND SYSTEMS

For complete details on system setup and simulation parameters, see Detailed Methods.

Barnase:barstar

Our investigation primarily focuses on the bacterial protein:protein complex barnase:barstar. The interaction of barnase, an extracellular ribonuclease, with its intracellular inhibitor, barstar, regulates RNA degradation in bacterial cells with a binding free energy of −19 kcal/mol [73]. Solvated barnase:barstar simulation models contain only ~ 41,000 atoms (including hydrogens and solvent), making it a computationally tractable system for studying sampling challenges in high-affinity protein:protein interfaces (Figure 1C, see Detailed Methods for system preparation details). Barnase:barstar has been well-studied both computationally [25, 47, 74] and experimentally [73, 75–77]). The barnase:barstar mutations considered in this work come from Schreiber et al. [73], who used stopped-flow measurements to derive experimental relative binding free energies ($\Delta\Delta G_{binding}$s) for 14 single amino acid substitutions across 13 residue positions in either barnase or barstar. The $\Delta\Delta G_{binding}$s for this set of mutations span an unusually large dynamic range (7.8 kcal/mol) with a statistical error of 0.1 kcal/mol, and involves a diverse set of amino acids, making it particularly useful for assessing quantitative predictive models. All mutations occur within or are in close proximity to the barnase:barstar interface, which is a complex network of interactions dominated by electrostatic interactions and coordinated by buried waters (Figure 1C) [73]. Therefore, the mutations tend to disrupt numerous interfacial interactions, potentially requiring significant conformational and water reorganization to achieve equilibrium, which may give rise to sampling challenges.

Terminally-blocked amino acids

As a control to the complexity of barnase:barstar, we also study terminally-blocked amino acids, which lack the complex interaction networks of barnase:barstar and have relatively few solute degrees of freedom. Specifically, we introduce mutations in small, solvated amino acids in two different environments: either terminally-blocked with ACE and NME caps at the N- and C-termini, respectively (ACE-X-NME), or terminally blocked with ALA residues with natural zwitterionic termini (ALA-X-ALA) (Figure 1D). The terminally-blocked mutation set consists of the same amino acid mutations as in the barnase:barstar mutation set, but contains only 10 total mutations (instead of 14 for barnase:barstar) as some of the barnase:barstar mutations involve the same amino acid transformation at different residue positions. By introducing the same mutations into the terminally-blocked amino acids, we separate the sampling challenges present in the barnase:barstar interface from the common challenges associated with alchemical free energy calculations.

To obtain relative free energies ($\Delta\Delta G$s), we estimate the free energy differences (ΔGs) for two phases. For barnase:barstar, we are interested in the $\Delta\Delta G_{binding}$, so the two simulation phases are complex and apo (Figure 1A). For terminally-blocked amino acids, there is no notion of binding, so the two phases are: ACE-X-NME and ALA-X-ALA (Figure 1B).

RESULTS

In the following sections, we investigate the sampling problems associated with applying relative free energy calculations to predict the impact of mutations in a model protein:protein complex. We establish an open-source workflow which consists of: (1) identifying mutations that are potentially plagued by sampling problems, (2) determining the slow degrees of freedom responsible for poor sampling, and (3) exploring state-of-the-art approaches for improving sampling. We first apply our workflow to a simple test system, terminally-blocked amino acids. We then focus the rest of the work on sampling challenges at the complex protein:protein interface of barnase:barstar, benchmarking to experimentally-determined binding free energies. While this study mainly focuses on analyzing the sampling problems in one protein:protein complex, past studies have performed similar types of analyses on other protein:ligand and protein:protein complexes [24, 54], so we expect our approach to be generalizable to other systems.

1 Mutations at protein:protein interfaces can be challenging for alchemical replica exchange relative free energy calculations, likely due to inadequate sampling in complex phase simulations

1.1 The relative free energy differences (ΔΔGs) for terminally-blocked amino acid mutations are internally consistent, well converged, and relatively absent of sampling problems—We first establish that running alchemical replica exchange (AREX) with our alchemical approach (e.g., alchemical protocol, atom mapping, softcore and counterion approaches, etc.) is free of sampling and convergence issues when the mutation is not located in the context of a complex network of protein interactions. We estimated the ΔΔGs of 10 terminally blocked amino acid mutations between two environments: ACE-X-NME and ALA-X-ALA, where X is an amino acid. For each mutation, we ran simulations in both the forward (A→B) and reverse (B→A) directions, where A corresponds to the amino acid in the WT barnase:barstar crystal structure (PDB ID: 1BRS). A mutation was considered internally consistent if the ΔΔG for the forward mutation (A→B) was within statistical error of the −ΔΔG for the reverse mutation (B→A). We found that with 5 ns/replica AREX simulations, all of the mutations are internally consistent and the forward ΔΔGs match the negative of the reverse ΔΔGs with high accuracy (root mean square error (RMSE): 0.21, 95% confidence interval (CI): [0.12, 0.28] kcal/mol, Figure 3A).

We next confirm that our calculations lack replica mixing bottlenecks and convergence issues. We first checked that there are no replica mixing bottlenecks for any of the mutations, indicating that the alchemical states are spaced such that they have reasonable overlap (Supplementary Figure 2). We next determined the extent to which the free energy difference of mutating WT→Mutant in one phase (ΔG) is converged because a converged ΔG indicates that the simulation has likely sampled all relevant degrees of freedom sufficiently. We assessed convergence by monitoring the changes in ΔG as a function of simulation time, which we call a “ΔG time series”. A ΔG time series was considered converged if, within the last five nanoseconds, it appeared flat with a close-to-zero slope (0 ± 0.1 kcal/mol/ns) (Supplementary Figure 3A). A ΔG time series was considered not converged if the magnitude of the slope of the last 5 ns was not within statistical uncertainty of 0 kcal/mol/ns. We found that for all 10 mutations in both phases and in both the forward and reverse directions, the slope of the ΔG time series is within statistical uncertainty of 0 kcal/mol/ns (Figure 3C–D, Supplementary Figure 3B–C), suggesting that the calculations are converged and relatively free of sampling problems.

Finally, we verify that 5 ns/replica AREX simulations thoroughly sample the slowest degrees of freedom for terminally blocked amino acid mutations [78]. We monitored the ϕ and ψ angles for the ACE-X-NME phase of two representative mutations with significant sampling problems in barnase:barstar, A2T (ALA to THR at residue 2) and R2A (ARG to ALA at residue 2) (see Section 2). If the ϕ and ψ degrees of freedom are thoroughly sampled, the time series should rapidly decorrelate. We quantify the extent to which each time series is hindered by slow correlation times by estimating its statistical inefficiency, $g = 2\tau + 1$, which is proportional to the autocorrelation of the time series τ [79, 80]. Since the sampling interval for the time series is 0.1 ns, if the statistical inefficiency (g) is close to 0.1 ns, the samples are completely decorrelated, and the larger the value of g, the more correlated the samples are [80]. We observed that for both representative mutations, ACE-X-NME phase simulations thoroughly sample both angles with g close to 0.1 ns (Supplementary Figure 4), providing further support that the terminally-blocked amino acid calculations are converged and have minimal sampling problems.

These results demonstrate that in the absence of significant sampling problems, running AREX with our alchemical approach can provide reliable estimates of relative free energy differences (ΔΔGs), indicating that we can use this approach to explore the challenges associated with applying RBFE calculations to interfacial residues in the barnase:barstar protein:protein complex.

1.2 Several barnase:barstar mutation predictions show poor accuracy due to slow convergence of the complex phase free energy difference ($\Delta G_{complex}$), suggesting the presence of sampling problems—

We assess the performance of AREX on predicting barnase:barstar relative binding free energies ($\Delta\Delta G_{binding}$s). We ran 10 ns/replica AREX simulations for the 14 mutations in the barnase:barstar mutation set in both the forward (i.e., mutations start from crystal structure residue) and reverse directions, resulting in a total of 28 $\Delta\Delta G_{binding}$ predictions. We first compared the predicted versus experimental $\Delta\Delta G_{binding}$s and considered a mutation to be significantly discrepant if the 95% Cls of its predicted and experimental $\Delta\Delta G_{binding}$s were not within 1 kcal/mol of each other. We observed relatively poor agreement (RMSE: 2.49, 95% Cl: [1.32, 3.74] kcal/mol) with 7% (2/28) of the predictions having the wrong sign and 21% (6/28) of the predictions considered significantly discrepant (Figure 6B, Supplementary Figure 6A). Moreover, when we compared the forward and negative of the reverse $\Delta\Delta G_{binding}$s for each mutation, we found that 21% (3/14) of mutations have poor internal consistency (i.e., the forward and negative reverse $\Delta\Delta G_{binding}$s are not within statistical error of each other) (Figure 6A or Figure 3B). We refer to the following mutations, all with poor accuracy with respect to experiment, as significantly discrepant mutations: A42T, R87A, D35A, H102A, A29Y, Q83R (Supplementary Figure 6A). A subset of these mutations (A42T, R87A, Q83R) also has poor internal consistency (Figure 6A, Supplementary Table 1).

We next demonstrate that all mutations with significant discrepancy have sufficiently overlapping alchemical states and some have slow $\Delta G_{complex}$ convergence. To assess state overlap, we checked for sufficient replica mixing in both phases of simulation and found that the replicas mix well, indicating that the $\Delta\Delta G_{binding}$ discrepancies are not a result of poor overlap of alchemical states (Supplementary Figure 5). We next determined whether the free energy difference of mutating WT→Mutant in each phase (ΔG) has converged by checking whether the slope of the last 5 ns of the ΔG time series (i.e., ΔG as a function of simulation time) is within statistical uncertainty of zero (0 ± 0.1 kcal/mol/ns). We found that 67% (4/6) of the significantly discrepant mutations (A42T, R87A, H102A, and Q83R) have $\Delta G_{complex}$s with slow convergence, suggesting that the corresponding simulations may contain significant sampling problems (Figure 3E–G). For the remaining 33% (2/6) of significantly discrepant mutations (D35A and A29Y), the $\Delta G_{complex}$s do converge within 10 ns (Figure 3G), which indicates that they may have minimal sampling problems, though it is possible that the slowest degrees of freedom in these simulations have correlation times longer than 10 ns and therefore have not yet been sampled.

Finally, we show that sampling problems often occur in complex phase simulations, especially for charge-changing mutations. We extended the convergence analysis to all 28 barnase:barstar mutations and observed that most of the mutations (27/28) have ΔG_{apo}s that converge within 10 ns/replica (Figure 3H). However, 25% (7/28) of mutations have $\Delta G_{complex}$s that do not converge, some of which are mutations that have predicted $\Delta\Delta G_{complex}$s close to experiment (R83Q, A39D, A35D) and should therefore be considered problematic mutations (Figure 3G). More importantly, this analysis indicates that convergence may be more difficult to achieve in the complex phase simulations, likely because of difficulties in sampling. Furthermore, out of the seven mutations with poor $\Delta G_{complex}$ convergence, six of the mutations are charge-changing, suggesting that sampling may be more challenging for charge-changing mutations (Figure 3G).

In summary, we identified several barnase:barstar mutations with predicted $\Delta\Delta G_{complex}$s that exhibit poor accuracy and described an approach for identifying mutations that potentially have significant sampling challenges. We found that 32% (9/28) of the mutations have discrepant $\Delta\Delta G_{complex}$s or slow $\Delta G_{complex}$s convergence with 10 ns/replica AREX simulations. Moreover, while sampling problems are absent for terminally-blocked amino acid mutations, they are likely present in the complex phase for several barnase:barstar mutations, most of which involve charge-changes. In the next section, we attempt to identify the slow degrees of freedom causing sampling challenges.

2 Poor complex phase sampling can occur due to mutation-dependent slow protein or water degrees of freedom

We choose two significantly discrepant mutations for deeper analysis of potential sampling challenges, A42T and R87A, each of which also has poor internal consistency (Supplementary Figure 6A, Figure 3B). These mutations encompass distinct types of transformations: A42T is a reverse mutation that involves a neutral, small to medium amino acid change (ALA to THR) and R87A is a forward mutation that involves a charge-changing, large to small transformation (ARG to ALA). Both mutations have slowly converging complex phase free energy differences ($\Delta G_{complex}$s) which are likely a result of sampling problems (Figure 3F). In this section, we confirm the presence of sampling problems in A42T and R87A complex phase simulations and identify slow degrees of freedom likely causing poor sampling.

2.1 Sampling challenges can be caused by hindered protein conformational dynamics—

We first hypothesize that slow $\Delta G_{complex}$ convergence can be attributed to poor sampling of slow protein backbone or side chain motion in the A42T and R87A complex phase simulations. To test this hypothesis, we ran 10 ns/replica complex phase AREX simulations where we imposed restraints on the heavy-atom coordinates to eliminate protein motion as a source of slow degrees of freedom, and compared the $\Delta G_{complex}$ time series with and without restraints for each mutation. Because these restraints significantly reduce protein motion, if the slow $\Delta G_{complex}$ convergence is caused by slow protein motions that are insufficiently sampled, the restraints should eliminate the sampling problem and the $\Delta G_{complex}$ time series should converge immediately. The A42T $\Delta G_{complex}$ time series with restraints converges rapidly within 10 ns, lacking the downward trend that is present in the $\Delta G_{complex}$ time series without restraints and indicating that the A42T complex phase simulation has a protein sampling problem (Figure 4A). However, for R87A, the $\Delta G_{complex}$ time series with restraints is within error of the time series without restraints, exhibiting the same downward trend as the unrestrained time series, which suggests the lack of convergence in the R87A $\Delta G_{complex}$ is not solely caused by protein sampling problems (Figure 4D). Although this analysis can help determine the presence (or absence) of protein sampling problems, it does not identify the specific slow degrees of freedom that are likely causing sampling problems.

2.2 Poor sampling can specifically be attributed to individual sidechain torsions, interfacial contacts, or nearby waters—

We next determine the specific degrees of freedom which may be responsible for slow $\Delta G_{complex}$ convergence by identifying conformational degrees of freedom that are tightly coupled to $\partial U/\partial\lambda$ [54]. We are particularly interested in mutations with slowly-varying $\partial U/\partial\lambda$ (i.e., highly correlated and with large statistical inefficiency, g), because slowly-varying $\partial U/\partial\lambda$s indicate slow $\Delta G_{complex}$ convergence. We monitored hundreds of protein and water degrees of freedom near the protein:protein interface over time because we observed that slow convergence is more common for $\Delta G_{complex}$ than ΔG_{apo} (Figure 3G–H). Specifically, we monitored the following degrees of freedom: backbone and rotameric torsion transitions near the interface, residue contacts within or between binding partners, and waters near the alchemical residue. We then computed the Pearson correlation coefficient (PCC) between $\partial U/\partial\lambda$ and each degree of freedom, averaging over all replicas. For mutations with slowly-varying $\partial U/\partial\lambda$, the most highly coupled (largest magnitude PCC) degree of freedom is likely implicated in slow $\Delta G_{complex}$ convergence. We emphasize that the slow degrees of freedom discussed hereafter are relatively slow (in the context of the degrees of freedom we analyzed and in the timescales of our simulations), and that they are not necessarily the globally slowest degrees of freedom for each alchemical transformation.

For A42T, which has slowly-varying $\partial U/\partial\lambda$ (g = 6.4 ns, Figure 5), the degrees of freedom with the largest magnitude PCCs are the χ_1 angle of T42 (PCC: −0.63, 95% CI: [−0.68, −0.56], Figure 5) and the distance between interface barstar residues T42 and E76 (PCC: 0.61, 95% Cl: [0.53, 0.66], Figure 5). The time series of a typical replica show that both degrees of freedom are highly correlated with $\partial U/\partial\lambda$ and slowly sample two metastable states during the 50 ns replica trajectory (Figure 4B–C). The relatively slow sampling of sidechain rotamer and interface contact metastable states (correlation time: 9.2 ns for T42 χ_1 and 9.5 ns for T42-E76, Figure 4B–C) likely explains the slow convergence of the A42T $\Delta G_{complex}$ time series in Figure 3F. Water sampling does not seem to play a significant role in causing poor $\Delta G_{complex}$ convergence for A42T, as the waters near T42 are only weakly correlated to $\partial U/\partial\lambda$ (PCC: 0.30, 95% CI: [0.21, 0.37], Figure 5).

For R87A, which also has slowly-varying $\partial U/\partial\lambda$ ($g = 32.1$ ns, Figure 5), the degree of freedom with the largest magnitude PCC is the number of waters near A87 (PCC −0.74, 95% CI: [−0.78, −0.68], Figure 5). The correlation is also particularly high for the distance between interface residues R87 (barnase) and D39 (barstar) (PCC: −0.73, 95% CI: [−0.76, −0.69], Figure 5). The time series of a representative replica shows that both degrees of freedom are highly correlated with $\partial U/\partial\lambda$ (Figure 4E–F). Both degrees of freedom also have long correlation times (9.6 ns for R87-D39 and 14.3 ns for neighboring waters) and slow equilibration times (evidenced by the upward trend in both degree of freedom time series), which suggests the slow convergence of R87A $\Delta G_{complex}$ is likely explained by slowness in R87-D39 and nearby waters (Figure 4E–F).

We next attempted to identify general trends in the slow degrees of freedom across all barnase:barstar mutations and found that there is no common degree of freedom (or category of degrees of freedom) that is implicated in all complex phase sampling problems (Figure 5). We observed that backbone torsions consistently show PCCs less than 0.5 in magnitude (which is smaller than the PCCs of the other degrees of freedom), indicating that backbone torsions are unlikely to be the primary cause of sampling problems. The other four categories—sidechain torsions, intra interface contacts, inter interface contacts, and neighboring waters—each have many high correlation values (magnitude of PCC greater than 0.5), but no single category explains the majority of sampling problems. Therefore, the slowest degrees of freedom are highly variable depending on the mutation (Figure 5).

Finally, we observed that for complex phase simulations, mutations involving charge-changes show slower convergence than charge-preserving transformations. 83% (10/12) of neutral mutations have $\partial U/\partial\lambda$ time series with $g < 1$, whereas 100% (16/16) of charge-changing mutations have $g > 1$ (Figure 5). We emphasize that the slow convergence of charge-changing mutations predominantly occurs in the complex (and not apo) phase (Figure 3G–H), indicating that introduction of a counterion to accommodate charge-changes does not significantly contribute to slow convergence. Instead, the sampling difficulties for charge-changing mutations likely emerge as a result of the strong network of electrostatic interactions at the barnase:barstar interface (Figure 1C).

One limitation of this work is that we studied only one protein:protein complex, and it is possible that other types of sampling problems are present in other protein:protein complexes. From our focused experiments, we cannot extrapolate how common the barnase:barstar sampling issues are for other protein:protein complexes, though it seems likely that the issues observed here are sufficiently fundamental in origin to be present in other complexes. It is worth remarking that the uniquely strong electrostatic nature of the barnase:barstar interface may exacerbate sampling challenges compared to other PPIs with less electrostatically-driven binding. The barnase:barstar interface involves 14 hydrogen bonds, more than the average protein:protein complex [75]. Of the 14 hydrogen bonds, most involve at least one charged residue, which is also atypical for protein:protein complexes [75]. Further work will be necessary to determine the extent to which the sampling problems observed in barnase:barstar are similar to those in other protein:protein systems and identify other mechanisms by which sampling problems could manifest.

Another caveat of this work is that the degrees of freedom explored in this analysis are not exhaustive; other, more complex collective variables (e.g., identified by time-lagged independent component analysis (TICA) [81, 82]) may correlate with $\partial U / \partial \lambda$ even more highly than those explored here. Nevertheless, our scan of simple degrees of freedom reveals specific slow degrees of freedom (sidechain torsions, interfacial contacts, or nearby waters) likely implicated in slow $\Delta G_{complex}$ convergence. Moreover, we found that the degrees of freedom causing poor sampling are highly dependent on the mutation. This analysis serves as an example approach for diagnosing sampling problems in other protein:protein complexes. In the next section, we explore approaches for ameliorating the sampling challenges.

3 Given sufficient simulation time, AREX and AREST can provide converged and accurate $\Delta\Delta G_{binding}$ predictions

We explore two potential solutions for overcoming the observed sampling challenges: (1) running much longer simulations with the same sampling strategy (AREX) with the goal of exceeding the relevant slow correlation times to enable convergence, and (2) using an enhanced sampling strategy that aims to reduce the correlation times to shorter timescales. For (2), we consider the addition of solute tempering to alchemical replica exchange (AREST). We explore the extent to which each approach improves convergence for the complex phase simulations of all barnase:barstar mutations, with a special focus on A42T and R87A.

3.1 Significantly longer (50 ns/replica) complex phase AREX simulations yield improved $\Delta G_{complex}$ convergence and adequate sampling of slow conformational degrees of freedom—We first demonstrate that running longer (50 ns/replica) complex phase AREX simulations improves barnase:barstar $\Delta\Delta G_{binding}$ predictions. We found that the accuracy of the predictions improved: the RMSE decreased from 2.49 (95% Cl: [1.32, 3.74]) kcal/mol with 10 ns/replica AREX to 1.61 (95% Cl: [1.12, 2.11]) kcal/mol with 50 ns/replica AREX (Figure 6D). Moreover, with 50 ns/replica AREX simulations, 86% (24/28) of predictions are close to experiment and all mutations have the correct sign (Figure 6D, Supplementary Figure 6C). We also found that the internal consistency improved: the RMSE decreased from 3.07 (95% Cl: [0.89, 4.76]) kcal/mol for 10 ns/replica AREX to 0.89 (95% Cl: [0.25, 1.43]) kcal/mol for 50 ns/replica AREX

(Figure 6C). Finally, we found that with 50 ns/replica AREX simulations, the convergence of $\Delta G_{\text{complex}}$ improved significantly, such that 100% (28/28) of mutations have converged (Supplementary Figure 11B). We then confirmed that the improved convergence for A42T and R87A is a result of more thorough sampling of the likely slowest degrees of freedom associated with each mutation. Examination of representative time series shows that between 10–50 ns, the slow degrees of freedom are sampled more comprehensively than with only 10 ns (Figure 4B, F).

We next show that the poor accuracy with respect to experiment for mutations with significantly discrepant $\Delta\Delta G_{\text{binding}}$s (even after 50 ns) are likely due to errors in force field parameters or extreme sampling problems. Despite the improved predictions obtained from running longer AREX, 14% (4/28) of the mutations still demonstrate significantly poor accuracy: D35A, A35D, Q83R, and A29Y (Figure 6D). Common reasons for discrepant $\Delta\Delta G_{\text{binding}}$ predictions include insufficient protein or water sampling, errors in force field parameters, and failure to model multiple protonation states [33]. We found that A29Y has a significantly discrepant $\Delta\Delta G_{\text{binding}}$ (and relatively poor internal consistency) because the mutant tyrosine residue does not sample the relevant energetically favorable orientations that enable it to contribute favorably to the barnase:barstar interface (details in Supplementary Information B).

We next investigated the causes of discrepancy for the remaining significantly discrepant mutations (D35A, A35D, and Q83R), all of which pass the internal consistency check with sufficient simulation time (100 ns/replica for Q83R and 50 ns/replica for the other two mutations, see Figure 6C and Supplementary Table 1). We first assessed whether the discrepancies are a result of failing to account for all relevant protonation states and found that protonation states are not the cause of these discrepancies (see Supplementary Information C). Given the absence of protonation state problems, the discrepancies are likely due to inaccurate force field parameters or insufficient sampling (of a slow degree of freedom with a correlation time longer than 50 ns). However, it is worth noting that with sufficient simulation time, the sign is correct for each of these discrepant $\Delta\Delta G_{\text{binding}}$s, indicating that the estimates are still useful in characterizing whether a mutation is energetically favorable or unfavorable (Figure 6D).

Despite the exceptions described above, we emphasize that running longer AREX improved our barnase:barstar predictions (RMSE is 1.61, 95% CI: [1.12, 2.11] kcal/mol, Figure 6D), indicating that for several mutations, sampling was insufficient with 10 ns/replica AREX but sufficient with 50 ns/replica. Moreover, 50 ns/replica may not be necessary depending on the desired accuracy, e.g., to achieve an RMSE of less than 2 kcal/mol for barnase:barstar predictions, ~20 ns/replica AREX simulations should be sufficient (Figure 7C).

3.2 AREST convergence is comparable to that of AREX for most mutations

—We next demonstrate that running 50 ns/replica AREST simulations also yields improved barnase:barstar $\Delta\Delta G_{\text{binding}}$ with respect to 10 ns/replica AREX simulations (for a comparison to 10 ns/replica AREST simulations, see Supplementary Information A.7). We ran 50 ns/replica AREST (with radius = 0.5 nm and $T_{\text{max}} = 600\ \text{K}$, see Supplementary Information D for details on REST parameter selection) for the complex phase of all barnase:barstar mutations and observed sufficient replica mixing for all mutations (Supplementary Figure 12). We observed improvement in the accuracy with respect to experiment; the RMSE decreased from 2.49 (95% Cl: [1.32, 3.74]) kcal/mol for 10 ns/ replica AREX to 1.65 (95% Cl: [1.23, 2.04]) kcal/mol for 50 ns/replica AREST (Figure 6F, Supplementary Figure 6D). We also found that the internal consistency significantly improved with the RMSE decreasing from 3.07 (95% Cl: [0.89, 4.76]) kcal/mol for 10 ns/replica AREX to 0.53 (95% Cl: [0.33, 0.71]) kcal/mol for 50 ns/replica AREST (Figure 6E). Finally, we also observed that with 50 ns/replica AREST simulations, 100% (28/28) of the complex free energy differences ($\Delta G_{complex}$s) converge (Supplementary Figure 11C).

We next show that while the two methods predict similar $\Delta\Delta G_{binding}$s for each mutation (Supplementary Figure 14), AREST converges more efficiently than AREX for two mutations with sampling problems, but not for the rest of the barnase:barstar mutations. We monitored the discrepancy in predicted $\Delta\Delta G_{binding}$ with respect to experiment as a function of time and compared the discrepancy time series for 50 ns/replica AREST versus that from 50 ns/replica AREX simulations. We analyzed the discrepancies in $\Delta\Delta G_{binding}$s for each of the seven mutations identified as potentially containing sampling problems due to slow $\Delta G_{complex}$ convergence with 10 ns/replica AREX: A42T, R87A, R83Q, Q83R, H102A, A35D, and A39D (Figure 3G). For A42T, the AREST discrepancy flattens out (to a close-to-zero discrepancy) more quickly than that of AREX, indicating that for A42T, AREST converges with less simulation time than AREX (Figure 7A). Similarly, for R87A, the AREST discrepancy starts to flatten out around 10 ns, while the AREX discrepancy doesn’t start to flatten out until ~40 ns, demonstrating that for R87A, AREST converges faster than AREX (Figure 7D). We next investigated why AREST yields faster convergence by comparing AREX and AREST sampling of the likely slowest degrees of freedom (T42 χ_1 angle for A42T and number of waters near A87 for R87A) in representative time series. We found that AREST more thoroughly samples these degrees of freedom and the statistical inefficiencies of the AREST time series are smaller than those of AREX, indicating that the faster convergence of AREST is due to reduction of relevant correlation times (Figure 7B, E).

Importantly, we found that for the remaining 71% (5/7) of mutations potentially containing sampling problems, the discrepancy in $\Delta\Delta G_{binding}$ does not converge to zero significantly faster for AREST than AREX (Supplementary Figure 15). Finally, to assess convergence across all mutations, we monitored the root mean square error (RMSE) and mean unsigned error (MUE) over time and observed that for both RMSE and MUE, the AREX and AREST time series are within error of each other (Figure 7C, F). Therefore, although AREST shows faster convergence than AREX for A42T and R87A, AREST convergence is comparable to that of AREX when comparing the two sampling strategies over all barnase:barstar mutations.

DISCUSSION

Widespread application of RBFE calculations to protein:protein complexes is primarily limited by the simulation time required to achieve reliable estimates

For some mutations, running RBFE calculations long enough to achieve converged, accurate, and reliable predictions can be computationally expensive, depending on the simulation time required and the computing resources available. For example, to achieve highly accurate RBFE predictions (RMSE ~1.6 kcal/mol) for barnase:barstar, the most challenging mutations (i.e., charge-changing mutations with sampling challenges) require 50 ns/replica for the complex phase and 10 ns/replica for the apo phase. This amounts to ~220 graphics processing unit (GPU) hours per mutation on an NVIDIA A100 graphics card—at the cost of roughly $920 per mutation on an equivalent instance on Amazon Web Services (AWS) (Supplementary Information A.12). However, we emphasize that we obtained converged and accurate $\Delta\Delta G_{binding}$ estimates for most of the mutations with 10 ns/replica AREX (Section 1.2), indicating that most mutations would not require such computationally expensive simulations (and instead would cost ~62 GPU hours and $260 per mutation on AWS). Taken together, our results demonstrate that given current best practices sampling strategies and state-of-the-art computing resources, the primary limiting factor in applying RBFE calculations to protein:protein complexes is the computational cost associated with achieving sufficient sampling for a small subset of mutations.

Given that similar types of sampling problems are also challenging for small molecule transformations [54, 55, 83], finding ways to reduce computational cost for alchemical transformations with difficult sampling problems will be highly useful for the development of alchemical free energy calculations in general. One straightforward approach for reducing computational cost involves waiting for improvements in hardware. GPU performance has rapidly improved over the last decade and will continue to improve in the coming years [84]. There are also particularly exciting developments in the realm of cheaper parallelization through the introduction of wider GPUs that enable a single GPU to be partitioned into multiple instances (e.g., NVIDIA’s Multi-Instance GPU feature).

Improvement of AREX and AREST simulation parameters may reduce the simulation time required for converged ΔΔG estimates for mutations with sampling challenges

Beyond anticipating advancements in hardware, a promising avenue for decreasing computational cost involves further optimizing the AREX and AREST simulation parameters used in this study. For both AREX and AREST, we chose the same number of alchemical intermediate states for all neutral mutations and a different, larger number of states for all charge-changing mutations. Additionally, we defined the alchemical and REST scaling protocols for each state according to simple, piecewise linear functions. Moreover, for AREST, we chose the REST parameters (radius and T_{max}) by exploring a small set of extreme REST parameters (Supplementary Information D).

Although we confirmed that our AREX and AREST parameter choices do not result in any replica mixing bottlenecks (Supplementary Figures 2, 5, 12), there are likely alternative protocol parameters which could provide more efficient $\Delta\Delta G_{binding}$ convergence. Ideally, each mutation would have optimized protocol parameters that provides a converged and accurate $\Delta\Delta G_{binding}$ estimate in the minimal amount of simulation time. However, because the search space for each of the parameters is large, brute-force optimization is unfeasible and even exploration of extreme values for each parameter for each mutation would be quite computationally expensive. Therefore, future work could involve development of methods for mutation-specific parameter optimization. Furthermore, there are also opportunities for optimizing mutation-independent protocol parameters, such as the integrator timestep [85], alchemical functional form [86–89], and Particle Mesh Ewald error tolerance [67] which may reduce simulation time.

Adaptation of other enhanced sampling methods for use in alchemical free energy calculations may also decrease the simulation time required to sufficiently sample difficult transformations

There are many existing methods for enhancing sampling in molecular dynamics simulations [90], many of which accelerate sampling of known slow degrees of freedom in a targeted manner [50–53, 91–96]. Some existing enhanced sampling methods also identify the slow degrees of freedom (as an intermediate step) [97, 98], but they do not necessarily identify the slow degrees of freedom that are most highly coupled to the alchemical coordinate (i.e. $\partial U / \partial \lambda$), which are responsible for slow convergence of RBFEs. Future work could involve incorporating existing enhanced sampling methods into alchemical free energy calculations to further improve sampling and convergence, as has been demonstrated for simple test systems [99]. Furthermore, when adapting methods that identify slow degrees of freedom, it will be important to account for coupling to the alchemical coordinate.

CONCLUSIONS

In this work, we explored the sampling challenges associated with applying relative binding free energy (RBFE) calculations to estimate the impact of protein mutations in a model protein:protein complex (barnase:barstar). We found that sampling problems are absent when the mutation is not located in the context of a complex network of protein interactions (i.e. in terminally-blocked amino acids), but are present in the complex phase for several barnase:barstar mutations, yielding slow convergence of $\Delta G_{complex}$s. Moreover, most of the mutations with complex phase sampling and convergence problems involve charge-changes. Furthermore, we attributed the barnase:barstar complex phase sampling problems to specific slow degrees of freedom (individual sidechain torsions, interfacial contacts, and nearby waters) which are highly dependent on the mutation. Finally, we found that given sufficient simulation time (50 ns/replica), both AREX and AREST can address most of the aforementioned sampling problems, with both methods demonstrating comparable convergence for most mutations.

Ultimately, our analyses and findings provide a model framework for diagnosing and mitigating sampling problems in other protein:protein complexes. By facilitating deep investigation of these sampling challenges in an open-source manner, our study lays the groundwork for the development of better methods for improving sampling in protein:protein RBFE calculations and free energy calculations in general.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to acknowledge Gyorgy Snell (ORCID: 0000-0003-1475-659X) for helpful input during project initiation. We also thank Peter Eastman (0000-0002-9566-9684) and Levi Naden (0000-0002-3692-5027) for helpful feedback on software implementation (Perses [56], https://github.com/choderalab/perses and OpenMMTools, https://github.com/choderalab/openmmtools). We thank OpenEye Scientific Software and Schrodinger for providing us with free academic licenses. We are grateful to Hannah Bruce MacDonald (0000-0002-5562-6866), Hannah Baumann (0000-0002-1736-7744), Bert de Groot (0000-0003-3570-3534), the reviewers, and the editors for their constructive feedback.

Funding

IZ acknowledges support from Vir Biotechnology Inc, a Molecular Sciences Software Institute Seed Fellowship, and the Tri-Institutional Program in Computational Biology and Medicine. JDC acknowledges support from NIH grant P30 CA008748, NIH grant R01 GM121505, and the Sloan Kettering Institute. DAR acknowledges support from a Molecular Sciences Software Institute Seed Fellowship and the Tri-Institutional Program in Chemical Biology. SS is a Damon Runyon Quantitative Biology Fellow supported by the Damon Runyon Cancer Research Foundation (DRQ-14-22) and acknowledges support from NIH grant R01 GM121505.

Disclosures

JDC is a current member of the Scientific Advisory Board of OpenEye Scientific Software, Redesign Science, Ventus Therapeutics, and Interline Therapeutics, and has equity interests in Redesign Science and Interline Therapeutics. The Chodera laboratory receives or has received funding from multiple sources, including the National Institutes of Health, the National Science Foundation, the Parker Institute for Cancer Immunotherapy, Relay Therapeutics, Entasis Therapeutics, Silicon Therapeutics, EMD Serono (Merck KGaA), AstraZeneca, Vir Biotechnology, Bayer, XtalPi, Interline Therapeutics, the Molecular Sciences Software Institute, the Starr Cancer Consortium, the Open Force Field Consortium, Cycle for Survival, a Louis V. Gerstner Young Investigator Award, and the Sloan Kettering Institute. A complete funding history for the Chodera lab can be found at http://choderalab.org/funding.

KH and LER are employees of and may hold shares in Vir Biotechnology Inc.

References

[1]. Momand J, Zambetti GP, Olson DC, George D, and Levine AJ. The mdm-2 oncogene product forms a complex with the p53 protein and inhibits p53-mediated transactivation. Cell, 69(7):1237–1245, June 1992. [PubMed: 1535557]

[2]. Lowenstein EJ, Daly RJ, Batzer AG, Li W, Margolis B, Lammers R, Ullrich A, Skolnik EY, Bar-Sagi D, and Schlessinger J. The SH2 and SH3 domain-containing protein GRB2 links receptor tyrosine kinases to ras signaling. Cell, 70(3):431–442, August 1992. [PubMed: 1322798]

[3]. Zhou Peng, Yang Xing-Lou, Wang Xian-Guang, Hu Ben, Zhang Lei, Zhang Wei, Si Hao-Rui, Zhu Yan, Li Bei, Huang Chao-Lin, Chen Hui-Dong, Chen Jing, Luo Yun, Guo Hua, Jiang Ren-Di, Liu Mei-Qin, Chen Ying, Shen Xu-Rui, Wang Xi, Zheng Xiao-Shuang, Zhao Kai, Chen Quan-Jiao, Deng Fei, Liu Lin-Lin, Yan Bing, Zhan Fa-Xian, Wang Yan-Yi, Xiao Geng-Fu, and Shi Zheng-Li. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature, 579(7798):270–273, March 2020. [PubMed: 32015507]

[4]. Yates Christopher M and Sternberg Michael J E. The effects of non-synonymous single nucleotide polymorphisms (nsSNPs) on protein-protein interactions. J. Mol. Biol, 425(21):3949–3963, November 2013. [PubMed: 23867278]

[5]. Liu Yang, Liu Jianying, Plante Kenneth S, Plante Jessica A, Xie Xuping, Zhang Xianwen, Ku Zhiqiang, An Zhiqiang, Scharton Dionna, Schindewolf Craig, Widen Steven G, Menachery Vineet D, Shi Pei-Yong, and Weaver Scott C. The N501Y spike substitution enhances SARS-CoV-2 infection and transmission. Nature, 602(7896):294–299, February 2022. [PubMed: 34818667]

[6]. Focosi Daniele, McConnell Scott, Casadevall Arturo, Cappello Emiliano, Valdiserra Giulia, and Tuccori Marco. Monoclonal antibody therapies against SARS-CoV-2. Lancet Infect. Dis, 22(11):e311–e326, November 2022. [PubMed: 35803289]

[7]. Thomson Emma C, Rosen Laura E, Shepherd James G, Spreafico Roberto, da Silva Filipe Ana, Wojce-chowskyj Jason A, Davis Chris, Piccoli Luca, Pascall David J, Dillen Josh, Lytras Spyros, Czudnochowski Nadine, Shah Rajiv, Meury Marcel, Jesudason Natasha, De Marco Anna, Li Kathy, Bassi Jessica, O'Toole Aine, Pinto Dora, Colquhoun Rachel M, Culap Katja, Jackson Ben, Zatta Fabrizia, Rambaut Andrew, Jaconi Stefano, Sreenu Vattipally B, Nix Jay, Zhang Ivy, Jarrett Ruth F, Glass William G, Beltramello Martina, Nomikou Kyriaki, Pizzuto Matteo, Tong Lily, Cameroni Elisa-betta, Croll Tristan I, Johnson Natasha, Di Iulio Julia, Wickenhagen Arthur, Ceschi Alessandro, Harbison Aoife M, Mair Daniel, Ferrari Paolo, Smollett Katherine, Sallusto Federica, Carmichael Stephen, Garzoni Christian, Nichols Jenna, Galli Massimo, Hughes Joseph, Riva Agostino, Ho Antonia, Schiuma Marco, Semple Malcolm G, Openshaw Peter J M, Fadda Elisa, Baillie J Kenneth, Chodera John D, ISARIC4C Investigators, COVID-19 Genomics UK (COG-UK) Consortium, Rihn Suzannah J, Lycett Samantha J, Virgin Herbert W, Telenti Amalio, Corti Davide, Robertson David L, and Snell Gyorgy. Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. Cell, 184(5):1171–1187.e20, March 2021. [PubMed: 33621484]

[8]. Mellor Paul, Marshall Jeremy D S, Ruan Xuan, Whitecross Dielle E, Ross Rebecca L, Knowles Margaret A, Moore Stanley A, and Anderson Deborah H. Patient-derived mutations within the n-terminal domains of p85α impact PTEN or rab5 binding and regulation. Sci. Rep, 8(1):7108, May 2018. [PubMed: 29740032]

[9]. Lippow Shaun M, Wittrup K Dane, and Tidor Bruce. Computational design of antibody-affinity improvement beyond in vivo maturation. Nat. Biotechnol, 25(10):1171–1176, October 2007. [PubMed: 17891135]

[10]. Clark Louis A, Ann Boriack-Sjodin P, Eldredge John, Fitch Christopher, Friedman Bethany, Hanf Karl J M, Jarpe Matthew, Liparoto Stefano F, Li You, Lugovskoy Alexey, Miller Stephan, Rushe Mia, Sherman Woody, Simon Kenneth, and Van Vlijmen Herman. Affinity enhancement of an in vivo matured therapeutic antibody using structure-based computational design. Protein Sci, 15(5):949–960, May 2006. [PubMed: 16597831]

[11]. Velazquez-Campoy Adrian, Leavitt Stephanie A, and Freire Ernesto. Characterization of Protein-Protein interactions by isothermal titration calorimetry. In Fu Haian, editor, Protein-Protein Interactions: Methods and Applications, pages 35–54. Humana Press, Totowa, NJ, 2004.

[12]. Fowler Douglas M and Fields Stanley. Deep mutational scanning: a new style of protein science. Nat. Methods, 11 (8):801–807, August 2014. [PubMed: 25075907]

[13]. Madeira Alexandra, Vikeved Elisabet, Nilsson Anna, Sjögren Benita, Andrén Per E, and Svenningsson Per. Identification of protein-protein interactions by surface plasmon resonance followed by mass spectrometry. Curr. Protoc. Protein Sci, Chapter 19:Unit 19.21, August 2011.

[14]. Kim Moonil, Park Kyoungsook, Jeong Eun-Ju, Shin Yong-Beom, and Chung Bong Hyun. Surface plasmon resonance imaging analysis of protein-protein interactions using on-chip-expressed capture protein. Anal. Biochem, 351(2):298–304, April 2006. [PubMed: 16510110]

[15]. Chen Fu, Liu Hui, Sun Huiyong, Pan Peichen, Li Youyong, Li Dan, and Hou Tingjun. Assessing the performance of the MM/PBSA and MM/GBSA methods. 6. capability to predict protein-protein binding free energies and re-rank binding poses generated by protein-protein docking. Phys. Chem. Chem. Phys, 18(32):22129–22139, August 2016. [PubMed: 27444142]

[16]. Genheden Samuel and Ryde UIf. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opin. Drug Discov, 10(5):449–461, May 2015. [PubMed: 25835573]

[17]. Rodrigues Carlos H M, Pires Douglas E V, and Ascher David B. mmCSM-PPI: predicting the effects of multiple point mutations on protein-protein interactions. Nucleic Acids Res, 49(W1):W417–W424, July 2021. [PubMed: 33893812]

[18]. Liu Ye, Yeung William S B, Chiu Philip C N, and Cao Dandan. Computational approaches for predicting variant impact: An overview from resources, principles to applications. Front. Genet, 13:981005, September 2022. [PubMed: 36246661]

[19]. Kortemme Tanja and Baker David. A simple physical model for binding energy hot spots in protein-protein complexes. Proc. Natl. Acad. Sci. U. S. A, 99(22):14116–14121, October 2002. [PubMed: 12381794]

[20]. Barlow Kyle A, Conchúir Shane Ó, Thompson Samuel, Suresh Pooja, Lucas James E, Heinonen Markus, and Kortemme Tanja. Flex ddg: Rosetta Ensemble-Based estimation of changes in Protein-Protein binding affinity upon mutation. J. Phys. Chem. B, 122(21):5389–5399, May 2018 [PubMed: 29401388]

[21]. Aldeghi Matteo, Gapsys Vytautas, and de Groot Bert L. Accurate estimation of ligand binding affinity changes upon protein mutation. ACS Cent Sci, 4(12):1708–1718, December 2018. [PubMed: 30648154]

[22]. Aldeghi Matteo, Gapsys Vytautas, and de Groot Bert L. Predicting kinase inhibitor resistance: Physics-Based and Data-Driven approaches. ACS Cent Sci, 5(8):1468–1474, August 2019. [PubMed: 31482130]

[23]. Hauser Kevin, Negron Christopher, Albanese Steven K, Ray Soumya, Steinbrecher Thomas, Abel Robert, Chodera John D, and Wang Lingle. Predicting resistance of clinical abl mutations to targeted kinase inhibitors using alchemical free-energy calculations. Commun Biol, 1:70, June 2018. [PubMed: 30159405]

[24]. Clark Anthony J, Gindin Tatyana, Zhang Baoshan, Wang Lingle, Abel Robert, Murret Colleen S, Xu Fang, Bao Amy, Lu Nina J, Zhou Tongqing, Kwong Peter D, Shapiro Lawrence, Honig Barry, and Friesner Richard A. Free energy perturbation calculation of relative binding free energy between broadly neutralizing antibodies and the gp120 glycoprotein of HIV-1. J. Mol. Biol, 429(7):930–947, April 2017. [PubMed: 27908641]

[25]. Clark Anthony J, Negron Christopher, Hauser Kevin, Sun Mengzhen, Wang Lingle, Abel Robert, and Friesner Richard A. Relative binding affinity prediction of Charge-Changing sequence mutations with FEP in Protein-Protein interfaces. J. Mol. Biol, 431(7):1481–1493, March 2019. [PubMed: 30776430]

[26]. Zwanzig Robert W. High-Temperature equation of state by a perturbation method. i. nonpolar gases. J. Chem. Phys, 22(8):1420–1426, August 1954.

[27]. Mey Antonia S J S, Allen Bryce K, Bruce Macdonald Hannah E, Chodera John D, Hahn David F, Kuhn Maximilian, Michel Julien, Mobley David L, Naden Levi N, Prasad Samarjeet, Rizzi Andrea, Scheen Jenke, Shirts Michael R, Tresadern Gary, and Xu Huafeng. Best practices for alchemical free energy calculations [article v1.0]. Living J Comput Mol Sci, 2(1), 2020

[28]. Kutzner Carsten, Kniep Christian, Cherian Austin, Nordstrom Ludvig, Grubmüller Helmut, de Groot Bert L, and Gapsys Vytautas. GROMACS in the cloud: A global supercomputer to speed up alchemical drug design. J. Chem. Inf. Model, 62(7):1691–1711, April 2022. [PubMed: 35353508]

[29]. Eastman Peter, Swails Jason, Chodera John D, McGibbon Robert T, Zhao Yutong, Beauchamp Kyle A, Wang Lee-Ping, Simmonett Andrew C, Harrigan Matthew P, Stern Chaya D, Wiewiora Rafal P, Brooks Bernard R, and Pande Vijay S. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. PLoS Comput. Biol, 13(7):e1005659, July 2017 [PubMed: 28746339]

[30]. He Xibing, Liu Shuhan, Lee Tai-Sung, Ji Beihong, Man Viet H, York Darrin M, and Wang Junmei. Fast, accurate, and reliable protocols for routine calculations of Protein-Ligand binding affinities in drug design projects using AMBER GPU-TI with ff14SB/GAFF. ACS Omega, 5(9):4611–4619, March 2020. [PubMed: 32175507]

[31]. Wang Lingle, Wu Yujie, Deng Yuqing, Kim Byungchan, Pierce Levi, Krilov Goran, Lupyan Dmitry, Robinson Shaughnessy, Dahlgren Markus K, Greenwood Jeremy, Romero Donna L, Masse Craig, Knight Jennifer L, Steinbrecher Thomas, Beuming Thijs, Damm Wolfgang, Harder Ed, Sherman Woody, Brewer Mark, Wester Ron, Murcko Mark, Frye Leah, Farid Ramy, Lin Teng, Mobley David L, Jorgensen William L, Berne Bruce J, Friesner Richard A, and Abel Robert. 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(7):2695–2703, February 2015 [PubMed: 25625324]

[32]. Hahn David F, Bayly Christopher I, Bruce Macdonald Hannah E, Chodera John D, Mey Antonia SJ S, Mobley David L, Benito Laura Perez, Schindler Christina E M, Tresadern Gary, and Warren

Gregory L. Best practices for constructing, preparing, and evaluating protein-ligand binding affinity benchmarks [article v0.1]. Living J Comput Mol Sci, 4(1), August 2022.

[33]. Cournia Zoe, Allen Bryce, and Sherman Woody. Relative binding free energy calculations in drug discovery: Recent advances and practical considerations. J. Chem. Inf. Model, 57(12):2911–2937, December 2017. [PubMed: 29243483]

[34]. Kuhn Bernd, Tichý Michal, Wang Lingle, Robinson Shaughnessy, Martin Rainer E, Kuglstatter Andreas, Benz Jörg, Giroud Maude, Schirmeister Tanja, Abel Robert, Diederich François, and Hert Jérôme. Prospective evaluation of free energy calculations for the prioritization of cathepsin L inhibitors. J. Med. Chem, 60(6):2485–2497, March 2017. [PubMed: 28287264]

[35]. Meier Katharina, Bluck Joseph P, and Christ Clara D. Use of free energy methods in the drug discovery industry. In Free Energy Methods in Drug Discovery: Current State and Future Directions, volume 1397 of ACS Symposium Series, pages 39–66. American Chemical Society, November 2021.

[36]. Ross Gregory, Lu Chao, Scarabelli Guido, Albanese Steven, Houang Evelyne, Abel Robert, Harder Edward, and Wang Lingle. The maximal and current accuracy of rigorous protein-ligand binding free energy calculations. ChemRxiv, October 2022.

[37]. Sherborne Bradley, Shanmugasundaram Veerabahu, Cheng Alan C, Christ Clara D, DesJarlais Renee L, Duca Jose S, Lewis Richard A, Loughney Deborah A, Manas Eric S, McGaughey Georgia B, Peishoff Catherine E, and van Vlijmen Herman. Collaborating to improve the use of free-energy and other quantitative methods in drug discovery. J. Comput. Aided Mol. Des, 30(12):1139–1141, December 2016. [PubMed: 28013427]

[38]. Schindler Christina E M, Baumann Hannah, Blum Andreas, Böse Dietrich, Buchstaller Hans-Peter, Burgdorf Lars, Cappel Daniel, Chekler Eugene, Czodrowski Paul, Dorsch Dieter, Eguida Merveille KI, Follows Bruce, Fuchß Thomas, Grädler Ulrich, Gunera Jakub, Johnson Theresa, Lebrun Catherine Jorand, Karra Srinivasa, Klein Markus, Knehans Tim, Koetzner Lisa, Krier Mireille, Leiendecker Matthias, Leuthner Birgitta, Li Liwei, Mochalkin Igor, Musil Djordje, Neagu Constantin, Rippmann Friedrich, Schiemann Kai, Schulz Robert, Steinbrecher Thomas, Tanzer Eva-Maria, Lopez Andrea Unzue, Follis Ariele Viacava, Wegener Ansgar, and Kuhn Daniel. Large-Scale assessment of binding free energy calculations in active drug discovery projects. J. Chem. Inf. Model, 60(11):5457–5474, November 2020. [PubMed: 32813975]

[39]. Wang Lingle, Chambers Jennifer, and Abel Robert. Protein-Ligand binding free energy calculations with FEP. Methods Mol. Biol, 2022:201–232, 2019.

[40]. Abel Robert, Wang Lingle, Harder Edward D, Berne BJ, and Friesner Richard A. Advancing drug discovery through enhanced free energy calculations. Acc. Chem. Res, 50(7):1625–1632, July 2017. [PubMed: 28677954]

[41]. Bastys Tomas, Gapsys Vytautas, Doncheva Nadezhda T, Kaiser Rolf, de Groot Bert L, and Kalinina Olga V. Consistent prediction of mutation effect on drug binding in HIV-1 protease using alchemical calculations. J. Chem. Theory Comput, 14(7):3397–3408, July 2018. [PubMed: 29847122]

[42]. Park Hwangseo and Jeon Young Ho. Free energy perturbation approach for the rational engineering of the antibody for human hepatitis B virus. J. Mol. Graph. Model, 29(5):643–649, February 2011. [PubMed: 21159534]

[43]. Xia Zhen, Huynh Tien, Kang Seung-Gu, and Zhou Ruhong. Free-energy simulations reveal that both hydrophobic and polar interactions are important for influenza hemagglutinin antibody binding. Biophys. J, 102(6):1453–1461, March 2012. [PubMed: 22455929]

[44]. Zhou Ruhong, Das Payel, and Royyuru Ajay K. Single mutation induced H3N2 hemagglutinin antibody neutralization: a free energy perturbation study. J. Phys. Chem. B, 112(49):15813–15820, December 2008. [PubMed: 19367871]

[45]. Das Payel, Li Jingyuan, Royyuru Ajay K, and Zhou Ruhong. Free energy simulations reveal a double mutant avian H5N1 virus hemagglutinin with altered receptor binding specificity. J. Comput. Chem, 30(11):1654–1663, August 2009. [PubMed: 19399777]

[46]. La Serra Maria Antonietta, Vidossich Pietro, Acquistapace Isabella, Ganesan Anand K, and De Vivo Marco. Alchemical free energy calculations to investigate Protein-Protein interactions: the case of the CDC42/PAK1 complex. J. Chem. Inf. Model, 62(12):3023–3033, June 2022. [PubMed: 35679463]

[47]. Patel Dharmeshkumar, Patel Jagdish Suresh, and Ytreberg F Marty. Implementing and assessing an alchemical method for calculating Protein-Protein binding free energy. J. Chem. Theory Comput, 17(4):2457–2464, April 2021. [PubMed: 33709712]

[48]. Shirts Michael R, Mobley David L, and Chodera John D. Chapter 4 Alchemical free energy calculations: Ready for prime time? In Spellmeyer DC and Wheeler R, editors, Annual Reports in Computational Chemistry, volume 3, pages 41–59. Elsevier, January 2007.

[49]. Larsen TA, Olson AJ, and Goodsell DS. Morphology of protein-protein interfaces. Structure, 6(4):421–427, April 1998. [PubMed: 9562553]

[50]. Laio Alessandro and Parrinello Michele. Escaping free-energy minima. Proc. Natl. Acad. Sci. U. S. A, 99(20):12562–12566, October 2002 [PubMed: 12271136]

[51]. Sutto Ludovico, Marsili Simone, and Gervasio Francesco Luigi. New advances in metadynamics. Wiley Interdiscip. Rev. Comput. Mol. Sci, 2(5):771–779, September 2012.

[52]. Torrie GM and Valleau JP. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. J. Comput. Phys, 23(2):187–199, February 1977.

[53]. Darve Eric, Rodríguez-Gómez David, and Pohorille Andrew. Adaptive biasing force method for scalar and vector free energy calculations. J. Chem. Phys, 128(14):144120, April 2008. [PubMed: 18412436]

[54]. Baumann Hannah M, Gapsys Vytautas, de Groot Bert L, and Mobley David L. Challenges encountered applying equilibrium and nonequilibrium binding free energy calculations. J. Phys. Chem. B, 125(17):4241–4261, May 2021. [PubMed: 33905257]

[55]. Mobley David L. Let's get honest about sampling. J. Comput. Aided Mol. Des, 26(1):93–95, January 2012. [PubMed: 22113833]

[56]. Rufa Dominic A, Zhang Ivy, Bruce Macdonald Hannah E, Grinaway Patrick B, Pulido Iván, Henry Mike M, Rodríguez-Guerra Jaime, Wittmann Matt, Albanese Steven K, Glass William G, Silveira Ana, Schaller David, Naden Levi N, and Chodera John D. Perses, June 2022.

[57]. Sugita Yuji, Kitao Akio, and Okamoto Yuko. Multidimensional replica-exchange method for free-energy calculations. J. Chem. Phys, 113(15):6042–6051, October 2000.

[58]. Woods Christopher J, Essex Jonathan W, and King Michael A. The development of Replica-Exchange-Based Free-Energy methods. J. Phys. Chem. B, 107(49):13703–13710, December 2003.

[59]. Wang Lingle, Friesner Richard A, and Berne BJ. Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2). J. Phys. Chem. B, 115(30):9431–9438, August 2011. [PubMed: 21714551]

[60]. Wang Lingle, Berne BJ, and Friesner Richard A. On achieving high accuracy and reliability in the calculation of relative protein-ligand binding affinities. Proc. Natl. Acad. Sci. U. S. A, 109(6):1937–1942, February 2012. [PubMed: 22308365]

[61]. Wang Lingle, Deng Yuqing, Knight Jennifer L, Wu Yujie, Kim Byungchan, Sherman Woody, Shelley John C, Lin Teng, and Abel Robert. Modeling local structural rearrangements using FEP/REST: Application to relative binding affinity predictions of CDK2 inhibitors. J. Chem. Theory Comput, 9(2):1282–1293, February 2013. [PubMed: 26588769]

[62]. Cole Daniel J, Tirado-Rives Julian, and Jorgensen William L. Enhanced monte carlo sampling through replica exchange with solute tempering. J. Chem. Theory Comput, 10(2):565–571, February 2014. [PubMed: 24803853]

[63]. Liu Shuai, Wang Lingle, and Mobley David L. Is ring breaking feasible in relative binding free energy calculations? J. Chem. Inf. Model, 55(4):727–735, April 2015. [PubMed: 25835054]

[64]. Crooks Gavin E. Measuring thermodynamic length. Phys. Rev. Lett, 99(10):100602, September 2007. [PubMed: 17930381]

[65]. Bennett Charles H. Efficient estimation of free energy differences from monte carlo data. J. Comput. Phys, 22(2):245–268, October 1976.

[66]. Shirts Michael R and Chodera John D. Statistically optimal analysis of samples from multiple equilibrium states. J. Chem. Phys, 129(12):124105, September 2008. [PubMed: 19045004]

[67]. Essmann Ulrich, Perera Lalith, Berkowitz Max L, Darden Tom, Lee Hsing, and Pedersen Lee G. A smooth particle mesh ewald method. J. Chem. Phys, 103(19):8577–8593, November 1995.

[68]. Pomès Régis, Eisenmesser Elan, Post Carol B, and Roux Benoît. Calculating excess chemical potentials using dynamic simulations in the fourth dimension. J. Chem. Phys, 111(8):3387–3395, August 1999.

[69]. Jones JE. On the determination of molecular fields.—i. from the variation of the viscosity of a gas with temperature. Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, 106(738):441–462, 1924.

[70]. Jones JE. On the determination of molecular fields. —ii. from the equation of state of a gas. Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, 106(738):463–477, 1924.

[71]. Lennard-Jones JE. Cohesion. Proc. Phys. Soc. London, 43(5):461, September 1931.

[72]. Chodera John D and Shirts Michael R. Replica exchange and expanded ensemble simulations as gibbs sampling: simple improvements for enhanced mixing. J. Chem. Phys, 135(19):194110, November 2011. [PubMed: 22112069]

[73]. Schreiber G and Fersht AR. Energetics of protein-protein interactions: analysis of the barnase-barstar interface by single mutations and double mutant cycles. J. Mol. Biol, 248(2):478–486, April 1995. [PubMed: 7739054]

[74]. Wang Jinan and Miao Yinglong. Protein-Protein Interaction-Gaussian accelerated molecular dynamics (PPI-GaMD): Characterization of protein binding thermodynamics and kinetics. J. Chem. Theory Comput, 18(3):1275–1285, March 2022. [PubMed: 35099970]

[75]. Buckle AM, Schreiber G, and Fersht AR. Protein-protein recognition: crystal structural analysis of a barnase-barstar complex at 2.0-Å resolution. Biochemistry, 33(30):8878–8889, August 1994. [PubMed: 8043575]

[76]. Schreiber Gideon and Fersht Alan R. Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering. Biochemistry, 32(19):5145–5150, May 1993. [PubMed: 8494892]

[77]. Schreiber G, Buckle AM, and Fersht AR. Stability and function: two constraints in the evolution of barstar and other proteins. Structure, 2(10):945–951, October 1994. [PubMed: 7866746]

[78]. Ramachandran GN, Ramakrishnan C, and Sasisekharan V. Stereochemistry of polypeptide chain configurations. J. Mol. Biol, 7:95–99, July 1963 [PubMed: 13990617]

[79]. Chodera John D. A simple method for automated equilibration detection in molecular simulations. J. Chem. Theory Comput, 12(4):1799–1805, April 2016. [PubMed: 26771390]

[80]. Chodera John D, Swope William C, Pitera Jed W, Seok Chaok, and Dill Ken A. Use of the weighted histogram analysis method for the analysis of simulated and parallel tempering simulations. J. Chem. Theory Comput, 3(1):26–41, January 2007. [PubMed: 26627148]

[81]. Pérez-Hernández Guillermo, Paul Fabian, Giorgino Toni, De Fabritiis Gianni, and Noé Frank. Identification of slow molecular order parameters for markov model construction. J. Chem. Phys, 139(1):015102, July 2013. [PubMed: 23822324]

[82]. Schwantes Christian R and Pande Vijay S. Improvements in markov state model construction reveal many Non-Native interactions in the folding of NTL9. J. Chem. Theory Comput, 9(4):2000–2009, April 2013. [PubMed: 23750122]

[83]. Lim Nathan M, Wang Lingle, Abel Robert, and Mobley David L. Sensitivity in binding free energies due to protein reorganization. J. Chem. Theory Comput, 12(9):4620–4631, September 2016. [PubMed: 27462935]

[84]. Mims Christopher. Huang's law is the new moore's law, and explains why nvidia wants arm. https://www.wsj.com/articles/huangs-law-is-the-new-moores-law-and-explains-why-nvidia-wants-arm-11600488001, September 2020. Accessed: 2023-2-14.

[85]. Fass Josh, Sivak David A, Crooks Gavin E, Beauchamp Kyle A, Leimkuhler Benedict, and Chodera John D. Quantifying Configuration-Sampling error in langevin simulations of complex molecular systems. Entropy, 20(5), May 2018.

[86]. Lee Tai-Sung, Lin Zhixiong, Allen Bryce K, Lin Charles, Radak Brian K, Tao Yujun, Tsai Hsu-Chun, Sherman Woody, and York Darrin M. Improved alchemical free energy calculations with optimized smoothstep softcore potentials. J. Chem. Theory Comput, 16(9):5512–5525, September 2020. [PubMed: 32672455]

[87]. Oshima Hiraku and Sugita Yuji. Modified hamiltonian in FEP calculations for reducing the computational cost of electrostatic interactions. J. Chem. Inf. Model, 62(11):2846–2856, June 2022. [PubMed: 35639709]

[88]. Ge Yunhui, Hahn David F, and Mobley David L. A benchmark of electrostatic method performance in relative binding free energy calculations. J. Chem. Inf. Model, 61(3):1048–1052, March 2021. [PubMed: 33686853]

[89]. Gapsys Vytautas, Seeliger Daniel, and de Groot Bert L. New Soft-Core potential function for molecular dynamics based alchemical free energy calculations. J. Chem. Theory Comput, 8(7):2373–2382, July 2012. [PubMed: 26588970]

[90]. Yang Yi Isaac, Shao Qiang, Zhang Jun, Yang Lijiang, and Gao Yi Qin. Enhanced sampling in molecular dynamics. J. Chem. Phys, 151(7):070902, August 2019. [PubMed: 31438687]

[91]. Evans Rhys, Hovan Ladislav, Tribello Gareth A, Cossins Benjamin P, Estarellas Carolina, and Gervasio Francesco L. Combining machine learning and enhanced sampling techniques for efficient and accurate calculation of absolute binding free energies. J. Chem. Theory Comput, 16(7):4641–4654, July 2020 [PubMed: 32427471]

[92]. Suruzhon Miroslav, Bodnarchuk Michael S, Ciancetta Antonella, Wall Ian D, and Essex Jonathan W. Enhancing ligand and protein sampling using sequential monte carlo. J. Chem. Theory Comput, 18(6):3894–3910, June 2022. [PubMed: 35588256]

[93]. Bussi Giovanni, Gervasio Francesco Luigi, Laio Alessandro, and Parrinello Michele. Free-energy landscape for beta hairpin folding from combined parallel tempering and metadynamics. J. Am. Chem. Soc, 128(41):13435–13441, October 2006. [PubMed: 17031956]

[94]. Zeller Fabian and Zacharias Martin. Adaptive biasing combined with hamiltonian replica exchange to improve umbrella sampling free energy simulations. J. Chem. Theory Comput, 10(2):703–710, February 2014. [PubMed: 26580047]

[95]. Limongelli Vittorio, Bonomi Massimiliano, and Parrinello Michele. Funnel metadynamics as accurate binding free-energy method. Proc. Natl. Acad. Sci. U. S. A, 110(16):6358–6363, April 2013. [PubMed: 23553839]

[96]. Zheng Lianqing, Chen Mengen, and Yang Wei. Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems. Proc. Natl. Acad. Sci. U. S. A, 105(51):20227–20232, December 2008. [PubMed: 19075242]

[97]. Tiwary Pratyush and Berne BJ. Spectral gap optimization of order parameters for sampling complex molecular systems. Proc. Natl. Acad. Sci. U. S. A, 113(11):2839–2844, March 2016. [PubMed: 26929365]

[98]. Ribeiro João Marcelo Lamim, Bravo Pablo, Wang Yihang, and Tiwary Pratyush. Reweighted autoencoded variational bayes for enhanced sampling (RAVE). J. Chem. Phys, 149(7):072301, August 2018. [PubMed: 30134694]

[99]. Hsu Wei-Tse, Piomponi Valerio, Merz Pascal T, Bussi Giovanni, and Shirts Michael R. Adding alchemical variables to metadynamics to enhance sampling in free energy calculations. arXiv [cond-mat.stat-mech], June 2022.

[100]. McGibbon Robert T., Beauchamp Kyle A., Harrigan Matthew P., Klein Christoph, Swails Jason M., Hernández Carlos X., Schwantes Christian R., Wang Lee-Ping, Lane Thomas J., and Pande Vijay S.. Mdtraj: A modern open library for the analysis of molecular dynamics trajectories. Biophysical Journal, 109(8):1528–1532, 2015. doi: 10.1016/j.bpj.2015.08.015 [PubMed: 26488642]

[101]. Hunter JD. Matplotlib: A 2d graphics environment. Computing in Science & Engineering, 9(3):90–95, 2007. doi: 10.1109/MCSE.2007.55

[102]. The PyMOL molecular graphics system, version 2.5.1. Schrödinger, LLC, New York, NY.

[103]. Case DA, Aktulga HM, Belfon K, Ben-Shalom IY, Brozell SR, Cerutti DS, Cheatham TE III, Cisneros GA, Cruzeiro VWD, Darden TA, Duke RE, Giambasu G, Gilson MK, Gohlke H, Goetz AW, Harris R, Izadi S, Izmailov SA, Jin C, Kasavajhala K, Kaymak MC, King E, Kovalenko A, Kurtzman T, Lee TS, LeGrand S, Li P, Lin C, Liu J, Luchko T, Luo R, Machado M, Man V, Manathunga M, Merz KM, Miao Y, Mikhailovskii O, Monard G, Nguyen H, O'Hearn KA, Onufriev A, Pan F, Pantano S, Qi R, Rahnamoun A, Roe DR, Roitberg A, Sagui C, Schott-Verdugo S, Shen J, Simmerling CL, Skrynnikov NR, Smith J, Swails J, Walker RC, Wang

J, Wei H, Wolf RM, Wu X, Xue Y, York DM, Zhao S, and Kollman PA. Amber 2021. University of California, San Francisco, 2021.

[104]. Schrödinger Releases 2021–2 and 2021–3. Maestro. Schrödinger, LLC, New York, NY, 2021.

[105]. Ramachandran S and Udgaonkar JB. Stabilization of barstar by chemical modification of the buried cysteines. Biochemistry, 35(26):8776–8785, July 1996. [PubMed: 8679642]

[106]. Jorgensen William L, Chandrasekhar Jayaraman, Madura Jeffry D, Impey Roger W, and Klein Michael L. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys, 79(2):926–935, July 1983.

[107]. Li Pengfei, Song Lin Frank, and Merz Kenneth M Jr. Systematic parameterization of monovalent ions employing the nonbonded model. J. Chem. Theory Comput, 11(4):1645–1657, April 2015. [PubMed: 26574374]

[108]. Maier James A, Martinez Carmenza, Kasavajhala Koushik, Wickstrom Lauren, Hauser Kevin E, and Simmerling Carlos. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. J. Chem. Theory Comput, 11(8):3696–3713, August 2015. [PubMed: 26574453]

[109]. Hauser Kevin, Essuman Bernard, He Yiqing, Coutsias Evangelos, Garcia-Diaz Miguel, and Simmerling Carlos. A human transcription factor in search mode. Nucleic Acids Res, 44(1):63–74, January 2016. [PubMed: 26673724]

[110]. Leimkuhler Benedict and Matthews Charles. Efficient molecular dynamics using geodesic integration and solvent-solute splitting. Proc. Math. Phys. Eng. Sci, 472(2189):20160138, May 2016. [PubMed: 27279779]

[111]. Leimkuhler Benedict and Matthews Charles. Robust and efficient configurational molecular sampling via langevin dynamics. J. Chem. Phys, 138(17):174102, May 2013 [PubMed: 23656109]

[112]. Darden Tom, York Darrin, and Pedersen Lee. Particle mesh ewald: An N.log(N) method for ewald sums in large systems. J. Chem. Phys, 98(12):10089–10092, June 1993.

[113]. Shirts Michael R, Mobley David L, Chodera John D, and Pande Vijay S. Accurate and efficient corrections for missing dispersion interactions in molecular simulations. J. Phys. Chem. B, 111(45):13052–13063, November 2007. [PubMed: 17949030]

[114]. Beutler Thomas C, Mark Alan E, van Schaik René C, Gerber Paul R, and van Gunsteren Wilfred F. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. Chem. Phys. Lett, 222(6):529–539, June 1994

[115]. Wennberg Christian L, Murtola Teemu, Páll Szilárd, Abraham Mark J, Hess Berk, and Lindahl Erik. Direct-Space corrections enable fast and accurate Lorentz-Berthelot combination rule Lennard-Jones lattice summation. J. Chem. Theory Comput, 11(12):5737–5746, December 2015. [PubMed: 26587968]

[116]. Virtanen Pauli, Gommers Ralf, Oliphant Travis E., Haberland Matt, Reddy Tyler, Cournapeau David, Burovski Evgeni, Peterson Pearu, Weckesser Warren, Bright Jonathan, van der Walt Stéfan J., Brett Matthew, Wilson Joshua, Millman K. Jarrod, Mayorov Nikolay, Nelson Andrew R. J., Jones Eric, Kern Robert, Larson Eric, Carey CJ, Polat IIlhan, Feng Yu, Moore Eric W., VanderPlas Jake, Laxalde Denis, Perktold Josef, Cimrman Robert, Henriksen Ian, Quintero EA, Harris Charles R., Archibald Anne M., Ribeiro Antônio H., Pedregosa Fabian, van Mulbregt Paul, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2. [PubMed: 32015543]

[117]. Mongan John, Case David A, and McCammon J Andrew. Constant ph molecular dynamics in generalized born implicit solvent. J. Comput. Chem, 25(16):2038–2048, December 2004 [PubMed: 15481090]

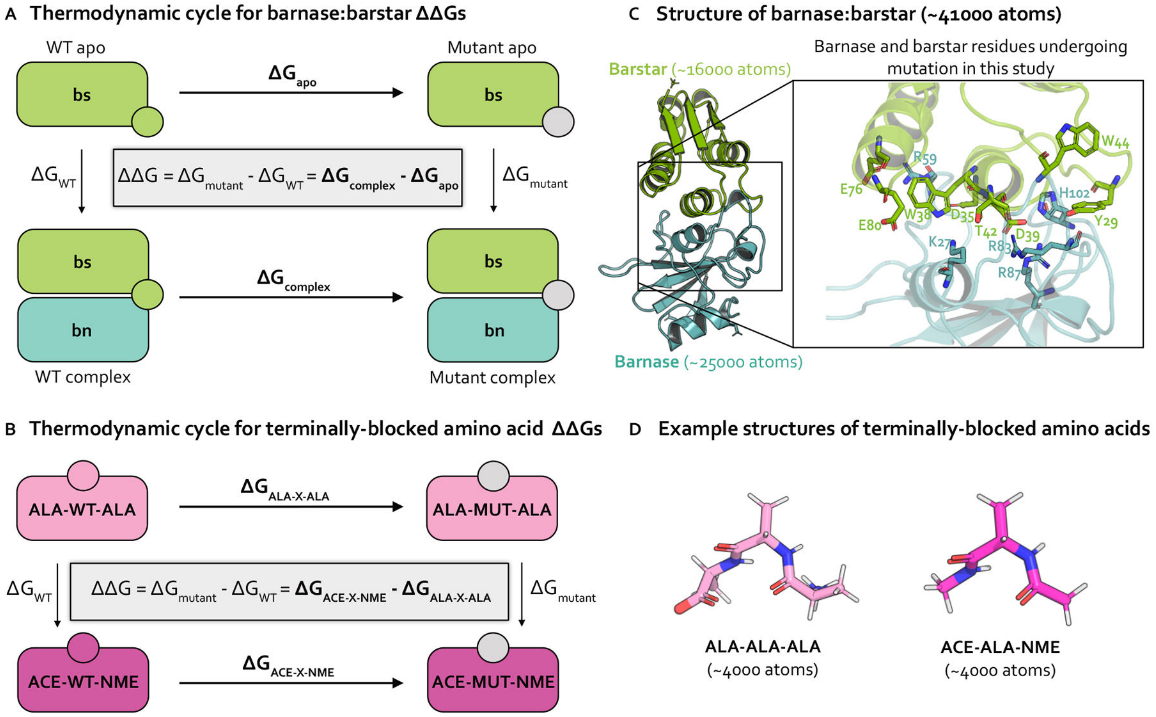
A Thermodynamic cycle for barnase:barstar ΔΔGs
WT apo
Mutant apo
ΔG$_{apo}$
bs
bs
ΔG$_{WT}$
ΔΔG = ΔG$_{mutant}$ - ΔG$_{WT}$ = ΔG$_{complex}$ - ΔG$_{apo}$
ΔG$_{mutant}$
bs
bn
ΔG$_{complex}$
bs
bn
WT complex
Mutant complex
C Structure of barnase:barstar (~41000 atoms)
Barstar (~16000 atoms)
Barnase and barstar residues undergoing mutation in this study
R59
E76
E80
W38
D35
T42
D39
K27
R83
R87
H102
W44
Y29
Barnase (~25000 atoms)
B Thermodynamic cycle for terminally-blocked amino acid ΔΔGs
ALA-WT-ALA
ΔG$_{ALA\text{-}X\text{-}ALA}$
ALA-MUT-ALA
ΔG$_{WT}$
ΔΔG = ΔG$_{mutant}$ - ΔG$_{WT}$ = ΔG$_{ACE\text{-}X\text{-}NME}$ - ΔG$_{ALA\text{-}X\text{-}ALA}$
ΔG$_{mutant}$
ACE-WT-NME
ΔG$_{ACE\text{-}X\text{-}NME}$
ACE-MUT-NME
D Example structures of terminally-blocked amino acids
ALA-ALA-ALA
(~4000 atoms)
ACE-ALA-NME
(~4000 atoms)

Figure 1. Relative free energy calculations predict the impact of single point mutations using thermodynamic cycles that each involve transformations in two environments.

(A) Thermodynamic cycle representing how the relative binding free energy ($\Delta\Delta G_{binding}$) can be computed for a protein mutation in the barnase:barstar complex. By cycle closure, the ΔΔG equation shown inside the thermodynamic cycle can be recovered. In practice, it is easier to compute the horizontal legs (ΔG_{apo} and $\Delta G_{complex}$, shown in bold) [33], which involve transforming a WT residue (green circle) into a mutant residue (gray circle). The free energy differences for each phase (apo and complex) are subtracted to compute the $\Delta\Delta G_{binding}$ **(B)** Thermodynamic cycle representing how the relative free energy (ΔΔG) can be computed for a protein mutation between two phases of terminally-blocked amino acids. The horizontal legs ($\Delta G_{ALA\text{-}X\text{-}ALA}$ and $\Delta G_{ACE\text{-}X\text{-}NME}$), shown in bold) are simulated, which involve transforming a WT residue (magenta or pink circle) into a mutant residue (gray circle). The free energy differences for each phase (ACE-X-NME and ALA-X-ALA) are subtracted to compute the ΔΔG. **(C)** Structural model of barnase:barstar (PDB ID: 1BRS) with barstar shown in green and barnase shown in blue. Barstar and barnase contain ~ 16000 and ~ 25000 atoms, respectively (including hydrogens and solvent). Zoomed-in view of the barnase:barstar interface shows the 13 residues undergoing mutation in this study (all of which are interfacial) as sticks. Nitrogen atoms shown in blue and oxygen atoms are shown in red. **(D)** Example structural models of terminally-blocked amino acids: ALA-X-ALA and ACE-X-NME (where X is ALA) shown in pink and magenta, respectively. Each terminally-blocked amino acid contains ~ 4000 atoms (including hydrogens and solvent). Nitrogen atoms are depicted in blue, oxygen atoms in red, and hydrogen atoms in white.

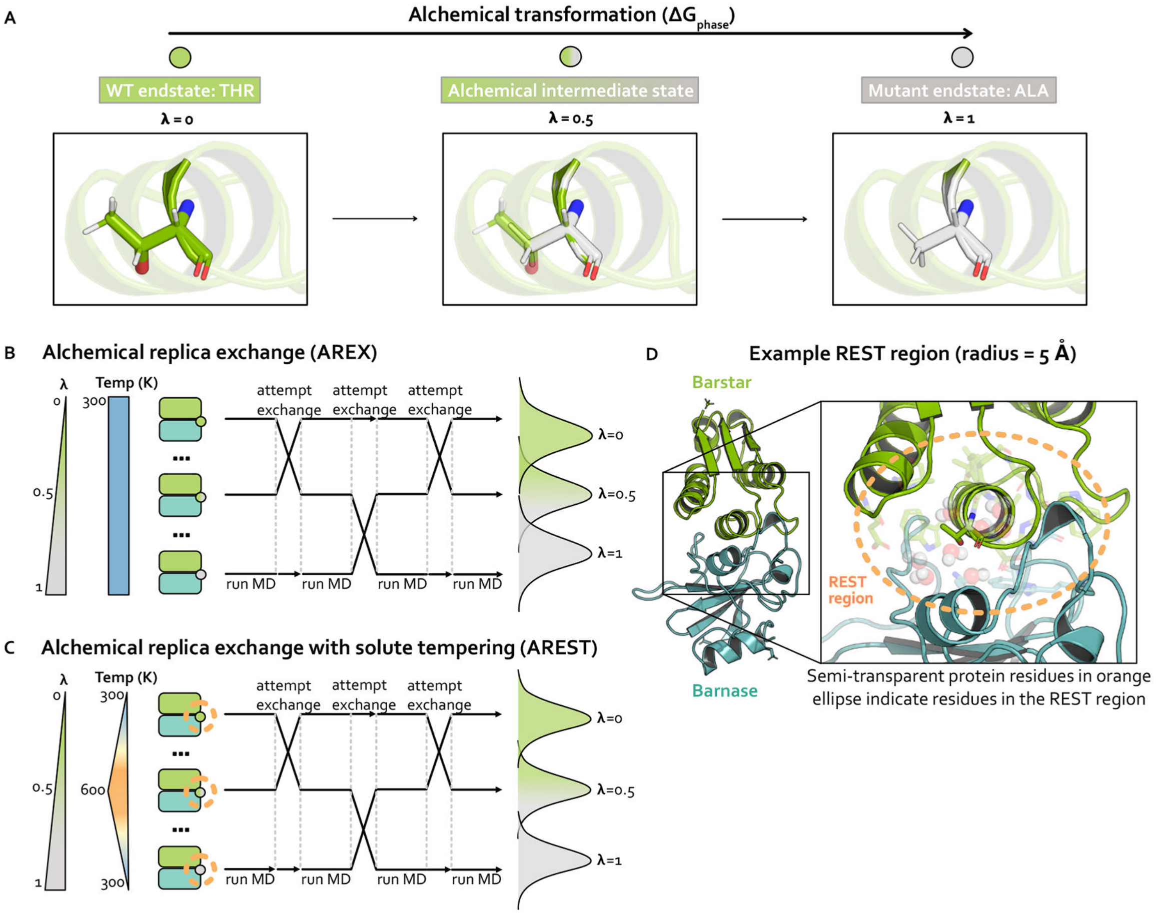

Figure 2. Strategies for sampling an alchemical transformation: Alchemical replica exchange (AREX) and alchemical replica exchange with solute tempering (AREST). AREST modifies AREX by introducing local heating around the alchemical region at intermediate alchemical states.

(A) Schematic representing an alchemical transformation (with one alchemical intermediate state) for one simulation phase. The WT ($\lambda = 0$, green) endstate contains a fully-interacting threonine residue and the mutant ($\lambda = 1$, gray) endstate contains a fully-interacting alanine residue. The alchemical intermediate ($\lambda = 0.5$, green-gray gradient) state contains partially interacting threonine and alanine residues. Nitrogen atoms are shown in blue and oxygen atoms are shown in red. **(B)** Schematic representing alchemical replica exchange (AREX), sometimes called Hamiltonian replica exchange among alchemical states, which utilizes multiple replicas (in this schematic, three replicas) to explore alchemical states that bridge the WT ($\lambda = 0$, green circle) and mutant ($\lambda = 1$, gray circle) fully interacting states. The temperature remains constant at 300 K for all alchemical states. Representative configurational distributions for each alchemical state are shown (on the right) to be overlapping for neighboring states, which is a requirement for accurate ΔΔG estimates. **(C)** Schematic representing alchemical replica exchange with solute tempering (AREST), which elevates the effective temperature for a small region (i.e., the REST region) to further enhance sampling. The REST region is shown as an orange, dashed circle. The effective temperature of the REST region reaches a maximum at 600 K at $\lambda = 0.5$. Representative configurational distributions for each alchemical state are shown (on the right) with less overlap than in AREX (panel B), because increasing the effective temperature usually causes increased thermodynamic length. **(D)** Structural model of barnase:barstar with barstar shown in green and barnase shown in blue. Zoomed-in view highlights an example REST region (orange, dashed circle) for residue T42 in barstar. T42 and neighboring residues (within 5 Å of T42) are shown as sticks and neighboring waters (also within 5 Å of T42) are shown as spheres.

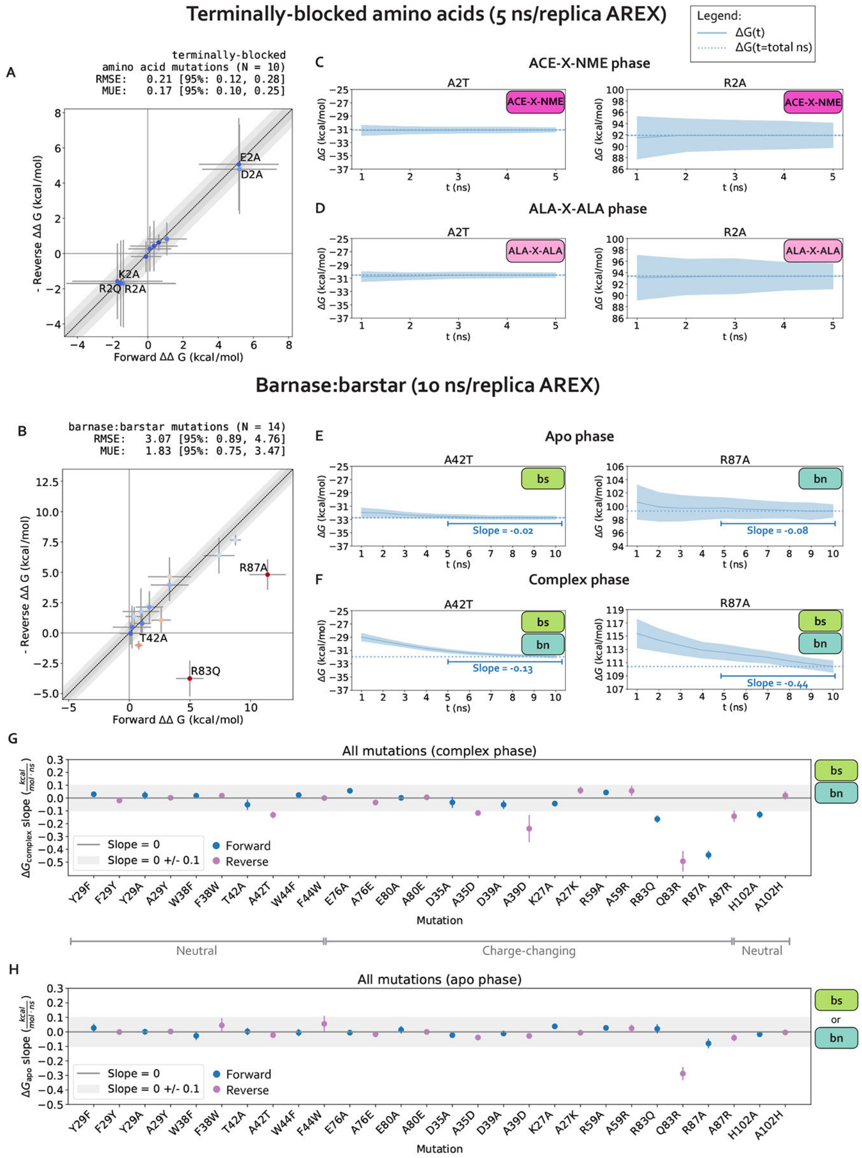

Figure 3. The relative free energy difference (ΔΔG) predictions for small terminally-blocked amino acid mutations are internally consistent and show good convergence, but several of the predictions for interfacial barnase:barstar mutations show poor internal consistency due to slow convergence of complex phase free energy differences ($\Delta G_{complex}$S).

(A) (Negative of the) Reverse versus forward ΔΔGs for each terminally-blocked amino acid mutation computed using alchemical replica exchange (AREX) simulations (number of states = 12 and 24 for neutral and charge mutations, respectively, and simulation time =5 ns/replica for each phase). Data points are labeled if the mutation involves a charge-change (to emphasize that our counterion introduction scheme works well in the absence of sampling problems). The y = x (black dotted) line represents zero discrepancy between forward and (negative of the) reverse ΔΔGs, the dark gray shaded region represents 0.5 kcal/mol discrepancy, and the light gray region represents 1 kcal/mol discrepancy. Data points are colored by how far they are from zero discrepancy (dark blue and red indicate close to and far from zero, respectively). Error bars represent two standard deviations and were computed by bootstrapping the decorrelated reduced potential matrices 200 times. Root mean square error (RMSE) and mean unsigned error (MUE) are shown with 95% confidence intervals obtained from bootstrapping the ΔΔGs 1000 times. **(B)** (Negative of the) Reverse versus forward ΔΔGs for each barnase:barstar mutation computed using alchemical replica exchange (AREX) simulations (number of states = 24 and 36 for neutral and charge mutations, respectively and simulation time = 10 ns/replica for each phase). Data points are labeled if the forward and (negative of the) reverse $\Delta\Delta G_{binding}$s are not within statistical error of each other (i.e., neither the forward nor the negative reverse $\Delta\Delta G_{binding}$ is within 1 kcal/mol of the 95% Cl for the other $\Delta\Delta G_{binding}$). For more details on the plot and error bars, refer to the caption for panel A. **(C)** Free energy difference (ΔG) time series for representative mutations A2T (left) and R2A (right) in the ACE-X-NME phase. Alchemical replica exchange simulations were performed with number of states = 12 and 24 for A2T and R2A, respectively and the simulation time was 5 ns/replica. Dashed line indicates the ΔG at t = 5 ns. Shaded region represents ± two standard deviations, which were computed by bootstrapping the decorrelated reduced potential matrices 200 times. **(D)** Same as (B), but for the ALA-X-ALA phase, instead of the ACE-X-NME phase. **(E)** Free energy difference (ΔG) time series for the apo phase of representative mutations with sampling problems: A42T (left) and R87A (right). Alchemical replica exchange simulations were performed with number of states = 24 and 36 for A42T and R87A, respectively and the simulation time was 10 ns/replica. Dashed line indicates the ΔG at t = 10 ns. For details on the error bars, refer to the caption for panel C. **(F)** Same as (E), but for the complex phase instead of the apo phase. **(G)** Slopes of the last 5 ns of the $\Delta G_{complex}$ time series for each barnase:barstar mutation are shown as blue (forward mutations) and purple (reverse mutations) circles. $\Delta G_{complex}$ time series were generated from complex phase AREX simulations (number of states = 24 and 36 for neutral and charge mutations, respectively, and simulation time = 10 ns/replica). Error bars represent 2 standard deviations and were computed using the SciPy `linregress` function. Slopes within error of the shaded gray region (0 ± 0.1 kcal/mol/ns) are close to zero and are therefore considered “flat.” **(H)** Same as (G), but for apo phase barnase or barstar mutations instead of complex phase barnase:barstar mutations.

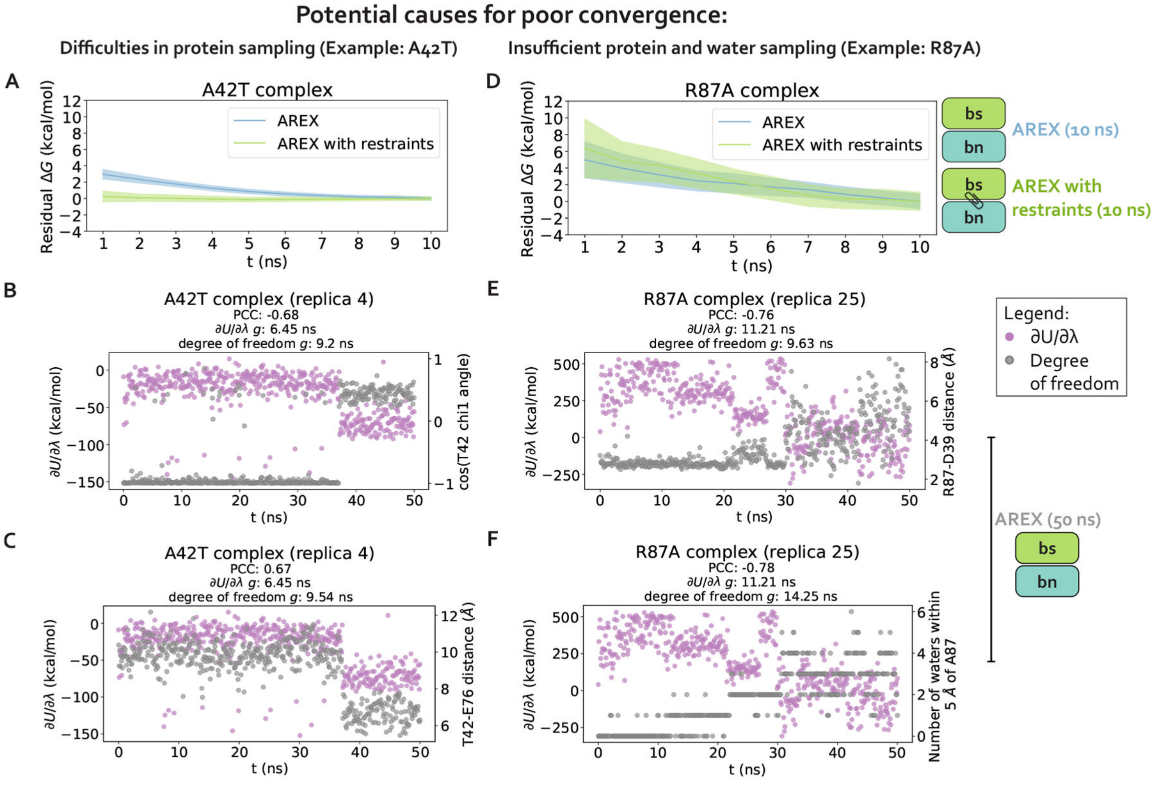

Figure 4. Complex phase convergence problems can arise due to insufficient sampling of protein and water degrees of freedom, e.g., a sidechain rotamer and intra-barstar contact for A42T and an inter-chain contact and neighboring waters for R87A.

(A) Residual complex phase free energy difference (ΔG) time series for AREX simulations of A42T (number of states = 24 and simulation time =10 ns/replica), where the residual ΔG is computed as $\Delta G(t) - \Delta G(t = 10\text{ns})$. Blue curve represents the time series for the AREX simulation without restraints and green curve represents the time series for the AREX simulation with heavy atoms restraints (force constant = 50 kcal/moIÅ^2). Shaded regions represent ± two standard deviations, which were computed by bootstrapping the decorrelated reduced potential matrices 200 times. **(B)** Time series for $\partial U / \partial \lambda$ (left y-axis, purple) and χ_1 angle for residue T42 (right y-axis, gray) for a representative replica (replica 4) of the A42T complex phase AREX simulation (number of states = 24, simulation time = 50 ns/replica). PCC indicates Pearson correlation coefficient and g indicates statistical inefficiency, which is proportional to the correlation time. $g = 0.1\text{ns}$ indicates very thorough sampling (because the sampling interval is 0.1 ns) and large values of g indicate poor sampling. **(C)** Time series for $\partial U / \partial \lambda$ (left y-axis, purple) and T42-E76 distance (right y-axis, gray) for a representative replica (replica 4) of the A42T complex phase AREX simulation (number of states = 24, simulation time = 50 ns/replica). **(D)** Same as (A), but for R87A instead of A42T (number of states = 36 and simulation time = 10 ns/replica) and using a force constant of 75 kcal/molÅ^2 instead of 50 kcal/molÅ^2. **(E)** Time series for $\partial U / \partial \lambda$ (left y-axis, purple) and R87-D39 distance (right y-axis, gray) for a representative replica (replica 25) of the R87A complex phase AREX simulation (number of states = 36, simulation time = 50 ns/replica). **(F)** Time series for $\partial U / \partial \lambda$ (left y-axis, purple) and number of waters within 5 Å of A87 (right y-axis, gray) for a representative replica (replica 25) of the R87A complex phase AREX simulation (number of states = 36, simulation time =50 ns/replica).

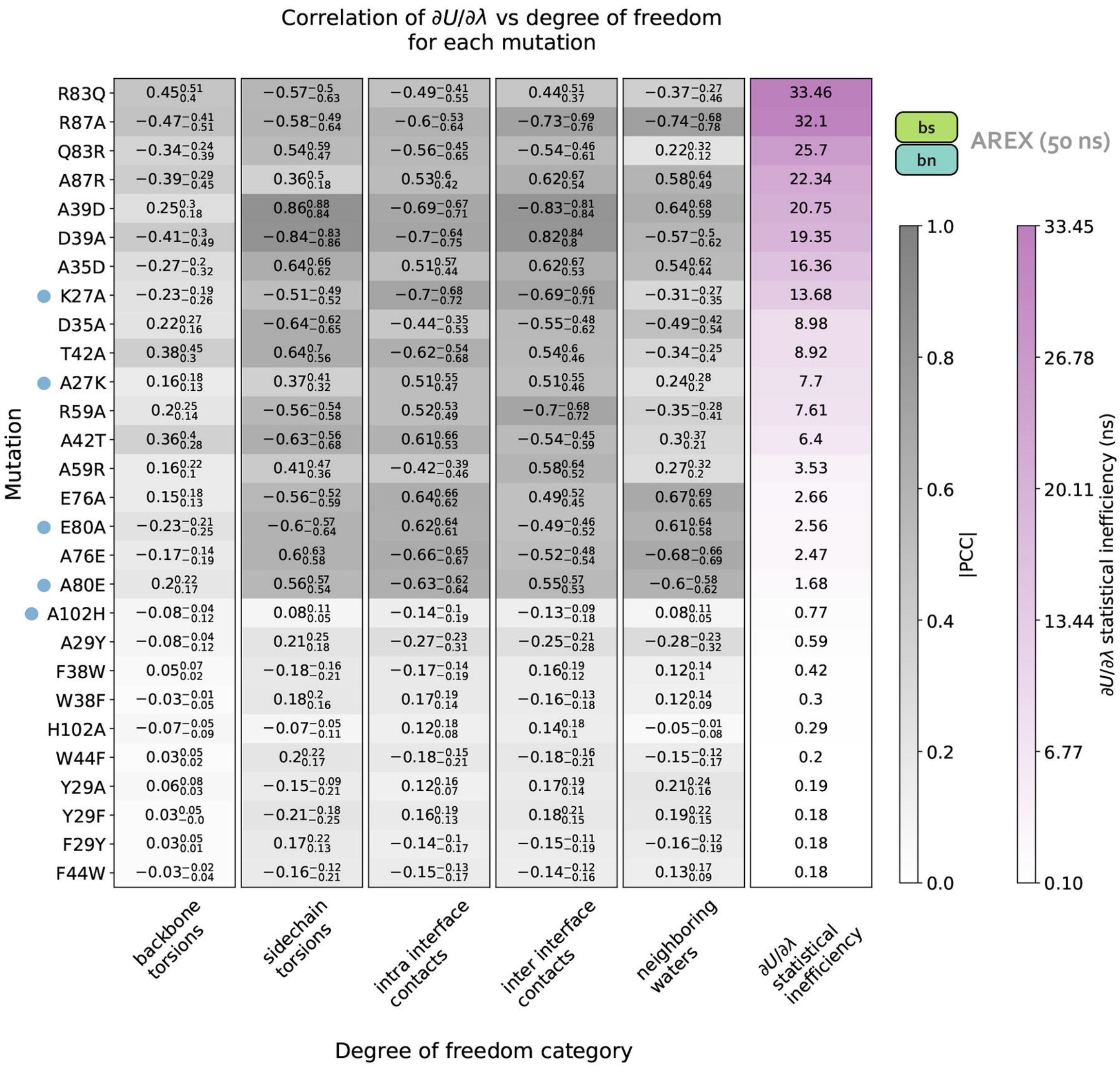

Figure 5. Charge-changing mutations demonstrate worse complex phase sampling than neutral mutations and the slowest degrees of freedom responsible for poor sampling are highly variable depending on the mutation.

Data in this plot was generated from 50 ns/replica complex phase AREX simulations. Each row of the heatmap corresponds to a mutation and each of the first five columns corresponds to a degree of freedom category: backbone torsions, sidechain torsions, intra-interface contacts, inter-interface contacts, and neighboring waters. Each category contains a set of degrees of freedom, i.e., the backbone torsions category contains the ϕ and ψ angles for all interface residues, sidechain torsions contains the χ_1, χ_2, χ_3, and χ_4 angles for all interface residues (if the angle is present for the residue), intra-interface contacts contains pairs of interface residues that are within the same chain, inter-interface contacts contains pairs of interface residues that span different chains, and neighboring waters involves monitoring the number of waters within 5 Å of the mutating residue. Each heatmap value (in the first five columns) is the maximum of (the absolute value of) the Pearson correlation coefficients (PCCs) between $\partial U/\partial\lambda$ and each of the degrees of freedom in the corresponding category for the corresponding mutation. For example, the top left value of the heatmap indicates that for R83Q, the backbone torsion with maximum correlation to $\partial U / \partial \lambda$ has a PCC of 0.45. The background colors for the PCC values are different shades of gray, with darker grays indicating values closer to 1. The subscript and superscript values associated with each PCC represent the 95% confidence interval. Each heatmap value in the last column corresponds to the statistical inefficiency of $\partial U / \partial \lambda$ across all replica trajectories for the corresponding mutation. Statistical inefficiency is proportional to the correlation time, where a value of 0.1 ns indicates very thorough sampling (because the sampling interval is 0.1 ns) and large values indicate poor sampling. Statistical inefficiency values are colored different shades of purple, with darker colors indicating larger values. The rows of the heatmap are ordered from highest to lowest by the statistical inefficiency of $\partial U / \partial \lambda$ across all replicas. Blue dots indicate mutations for which the degree of freedom with the largest magnitude PCC is relatively far from the mutating residue. See Detailed Methods for more information about this analysis.

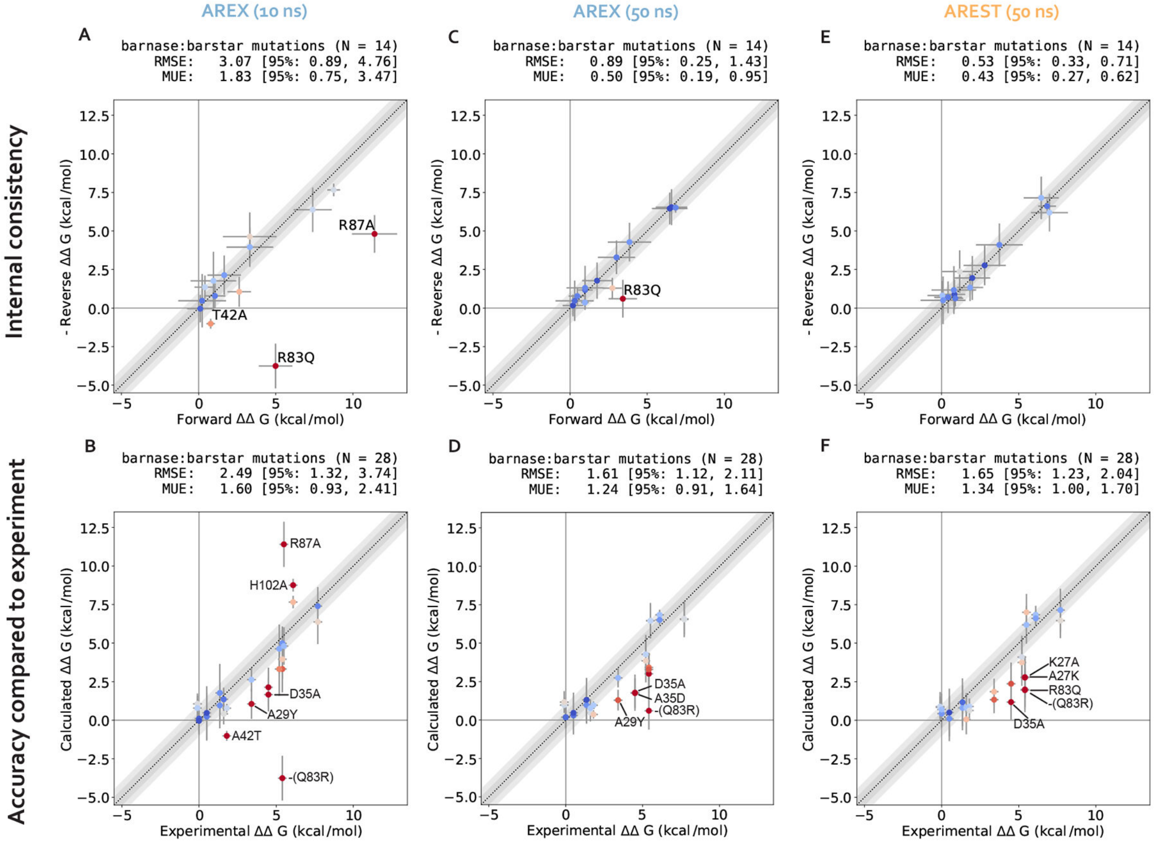

Figure 6. Running long (50 ns/replica) simulations of alchemical replica exchange (AREX) and alchemical replica exchange with solute tempering (AREST) yields improved $\Delta\Delta G_{binding}$ predictions with respect to 10 ns/replica AREX simulations.

(A) (Negative of the) Reverse versus forward $\Delta\Delta G_{binding}$s for each barnase:barstar mutation computed from AREX simulations (number of states = 24 and 36 for neutral and charge mutations, respectively and simulation time = 10 ns/replica for each phase). The y = x (black dotted) line represents zero discrepancy between forward and (negative of the) reverse $\Delta\Delta G_{binding}$s, the dark gray shaded region represents 0.5 kcal/mol discrepancy, and the light gray region represents 1 kcal/mol discrepancy. Data points are colored by how far they are from zero discrepancy (dark blue and red indicate close to and far from zero, respectively). Data points are labeled if the forward and (negative of the) reverse $\Delta\Delta G_{binding}$s are not within statistical error of each other (i.e., neither the forward nor the negative reverse $\Delta\Delta G_{binding}$ is within 1 kcal/mol of the 95% Cl for the other $\Delta\Delta G_{binding}$). Error bars represent two standard deviations and were computed by bootstrapping the decorrelated reduced potential matrices 200 times. Root mean square error (RMSE) and mean unsigned error (MUE) are shown with 95% confidence intervals obtained from bootstrapping the data 1000 times. **(B)** Calculated versus experimental $\Delta\Delta G_{binding}$s for each barnase:barstar mutation computed from AREX simulations (number of states = 24 and 36 for neutral and charge mutations, respectively and simulation time = 10 ns/replica for each phase). The y = x (black dotted) line represents zero discrepancy between calculated and experimental $\Delta\Delta G_{binding}$s, the dark gray shaded region represents 0.5 kcal/mol discrepancy, and the light gray region represents 1 kcal/mol discrepancy. Data points are labeled if the 95% Cls of the calculated and experimental $\Delta\Delta G_{binding}$s are not within 1 kcal/mol of each other. For more details on the plot and error bars, refer to the caption for panel A. **(C)** Same as (A), but using 50 ns/replica AREX simulations for the complex phase and 10 ns/replica AREX simulations for the apo phase instead of 10 ns/replica AREX simulations for both phases. **(D)** Same as (B), but using 50 ns/replica AREX simulations for the complex phase and 10 ns/replica AREX simulations for the apo phase instead of 10 ns/replica AREX simulations for both phases. **(E)** Same as (A), but using 50 ns/replica AREST simulations for the complex phase and 10 ns/replica AREX simulations for the apo phase instead of 10 ns/replica AREX simulations for both phases. **(F)** Same as (B), but using 50 ns/replica AREST simulations for the complex phase and 10 ns/replica AREX simulations for the apo phase instead of 10 ns/replica AREX simulations for both phases.

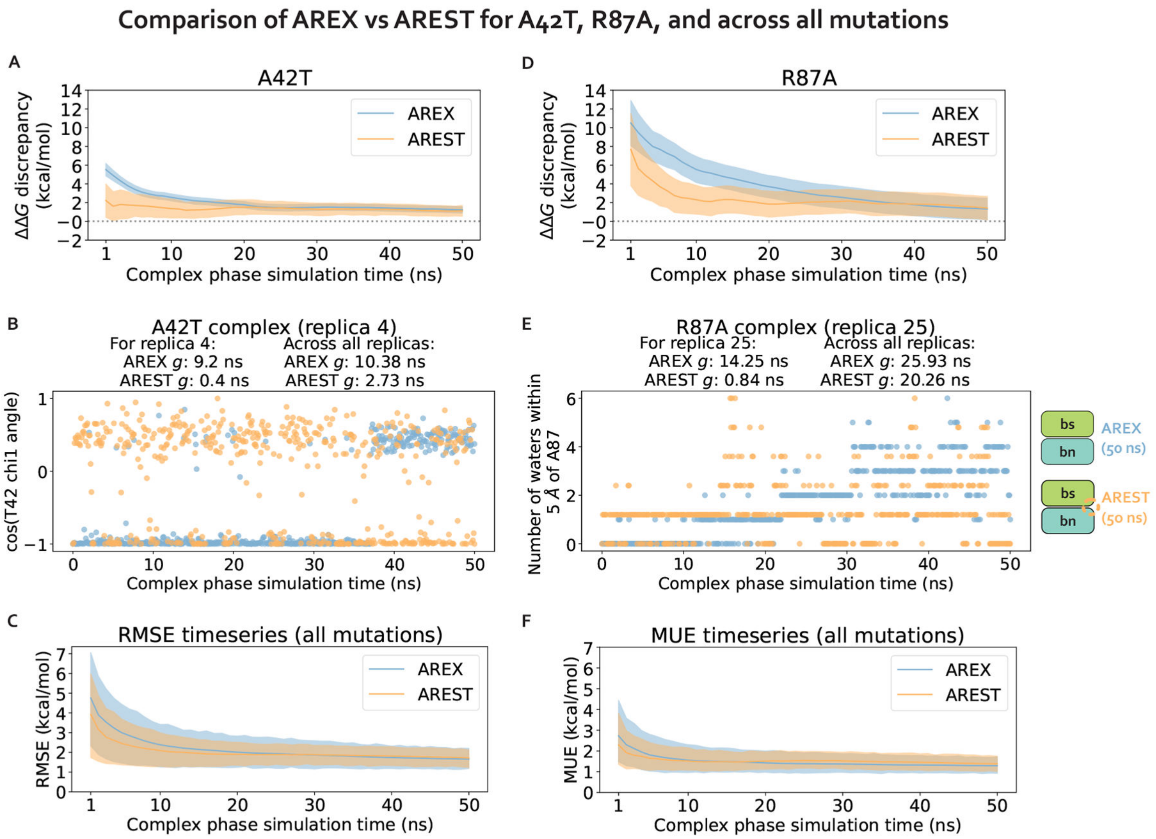

Figure 7. Alchemical replica exchange with solute tempering (AREST) and alchemical replica exchange (AREX) demonstrate comparable convergence for most barnase:barstar mutations.
(A) $\Delta\Delta G_{binding}$ discrepancy (with respect to experiment) time series for A42T. The discrepancy was computed as $\Delta G_{complex} - \Delta G_{apo} - \Delta\Delta G_{experiment}$, where $\Delta G_{complex}$ corresponds to the (AREX or AREST) complex phase ΔG at a particular time point, ΔG_{apo} corresponds to the apo phase ΔG computed from a 10 ns/replica AREX simulation, and $\Delta\Delta G_{experiment}$ is the experimental value from Schreiber et al [73]. AREX time series shown in blue and AREST time series (with radius = 0.5 nm, T_{max} = 600 K) shown in orange. Number of states is 24 for both AREX and AREST. Shaded regions represent ± two standard deviations, computed by bootstrapping the decorrelated reduced potential matrices 200 times. Gray dashed line indicates $\Delta\Delta G_{binding}$ discrepancy = 0. **(B)** Time series of the χ_1 angle for residue T42 for a representative replica (replica 4) of the A42T complex phase AREX simulation (blue) and AREST simulation (orange) (number of states = 24, simulation time = 50 ns/replica). g indicates statistical inefficiency, which is proportional to the correlation time. $g = 0.1$ ns indicates very thorough sampling (because the sampling interval is 0.1 ns) and large values of g indicate poor sampling. **(C)** Time series of the root mean square error (RMSE) (with respect to experiment) for the $\Delta\Delta G_{binding}$s of all barnase:barstar mutations. The $\Delta\Delta G_{binding}$s used to compute the RMSE at each time point were computed as $\Delta G_{complex} - \Delta G_{apo}$for each mutation, where $\Delta G_{complex}$ corresponds to the (AREX or AREST) complex phase ΔG at a particular time point and ΔG_{apo} corresponds to the apo phase ΔG computed from a 10 ns/replica AREX simulation. AREX time series shown in blue and AREST time series (with radius = 0.5 nm, T_{max} = 600K) shown in orange.

Number of states is 24 for neutral mutations and 36 for charge-changing mutations. Shaded regions represent ± two standard deviations, computed by bootstrapping 1000 times. **(D)** Same as (A), but for R87A instead of A42T. Number of states is 36 for both AREX and AREST. (E) Time series of the number of waters within 5 Å of residue A87 for representative replica (replica 25) of the R87A complex phase AREX simulation (blue) and AREST simulation (orange) (number of states = 36, simulation time =50 ns/replica). (F) Same as (C) but for mean unsigned error (MUE) instead of RMSE.