

Approaching protein design with multisite λ dynamics: Accurate and scalable mutational folding free energies in T4 lysozyme

Ryan L. Hayes,¹ Jonah Z. Vilseck,¹ and Charles L. Brooks III^{1,2*}

¹Department of Chemistry, University of Michigan, Ann Arbor, Michigan, 48109

²Biophysics Program, University of Michigan, Ann Arbor, Michigan, 48109

Received 17 May 2018; Accepted 15 August 2018

DOI: 10.1002/pro.3500

Published online 00 Month 2018 proteinscience.org

Abstract: The estimation of changes in free energy upon mutation is central to the problem of protein design. Modern protein design methods have had remarkable success over a wide range of design targets, but are reaching their limits in ligand binding and enzyme design due to insufficient accuracy in mutational free energies. Alchemical free energy calculations have the potential to supplement modern design methods through more accurate molecular dynamics based prediction of free energy changes, but suffer from high computational cost. Multisite λ dynamics (MS λ D) is a particularly efficient and scalable free energy method with potential to explore combinatorially large sequence spaces inaccessible with other free energy methods. This work aims to quantify the accuracy of MS λ D and demonstrate its scalability. We apply MS λ D to the classic problem of calculating folding free energies in T4 lysozyme, a system with a wealth of experimental measurements. Single site mutants considering 32 mutations show remarkable agreement with experiment with a Pearson correlation of 0.914 and mean unsigned error of 1.19 kcal/mol. Multisite mutants in systems with up to five concurrent mutations spanning 240 different sequences show comparable agreement with experiment. These results demonstrate the promise of MS λ D in exploring large sequence spaces for protein design.

Keywords: protein folding; catalysis; binding; molecular dynamics

Additional Supporting Information may be found in the online version of this article.

Statement of Importance and Impact: Modern protein design algorithms have had remarkable success on a variety of targets; however, more accurate complementary methods are needed. Multisite λ dynamics (MS λ D) is a particularly efficient and scalable alchemical free energy method offering the potential for greater accuracy and rigor. Its accuracy and scalability in computing protein stability are demonstrated here on T4 lysozyme, and suggest MS λ D holds great promise in protein design.

Grant sponsor: NSF CHE 1506273; Grant sponsor: NIH GM037554 and GM103695.

*Correspondence to: Charles L. Brooks III, Biophysics Program, University of Michigan, Ann Arbor, Michigan 48109.

E-mail: brookscl@umich.edu

Introduction

The estimation of changes in free energy upon mutation is one of the central problems of protein design. Relevant free energies include the folding free energy, ligand binding free energy, catalytic transition state binding free energy, protein–protein association free energy, and others. These free energies can be used to design proteins with desired stability, ligand binding, catalytic, and dimerization properties, respectively. Estimation of the appropriate free energy can guide *in silico* evolution, through Monte Carlo sampling of sequence space at some selection temperature^{1,2} to optimize the property of interest.

Modern protein design algorithms, such as Rosetta^{3,4} and others,^{5–8} typically estimate free energies for a particular sequence and conformation using heuristic or empirical free energy estimates. Sequence and conformational space are explored simultaneously

through Monte Carlo sampling of rotamer libraries of various side chains. This general approach has had remarkable success in designing new sequences with desired structural properties,⁹ as well as moderate success in designing sequences with desired dimerization,¹⁰ ligand binding,¹¹ and catalytic properties.^{12,13}

In spite of this success, several unmet challenges remain in ligand binding¹⁴ and enzyme design that would benefit from improved predictions of free energy differences. Designed enzymes continue to show much lower activity than natural enzymes,^{15–17} but typically possess enough initial activity that their efficiency may be raised to moderate levels through directed evolution.^{12,13} Although directed evolution can improve catalytic efficiency, better design algorithms are highly desirable, because better initial enzyme designs also tend to possess superior activities after directed evolution.¹³ Current enzyme design methods suffer from a variety of limitations and approximations. Limitations include the inability to dock theozymes¹⁸ with more than about four catalytic residues into candidate protein scaffolds,¹⁶ which limits designed enzymes to simple reactions. Approximations include implicit solvent energy functions, lack of long range electrostatics, and the neglect of backbone flexibility during the design process.¹⁷ Due to these approximations, many enzyme active sites are no longer stable after redesign. In contrast, explicit solvent molecular dynamics (MD) simulations utilize more accurate energy functions and explicitly account for protein flexibility. Indeed, long MD simulations are frequently used in a diagnostic fashion to assess the kinetic stability of designed sequences before experimental testing.^{15,19}

Alchemical free energy methods can make rigorous estimates of thermodynamic stability using MD simulations. Free energy methods include free energy perturbation (FEP),²⁰ thermodynamic integration (TI),²¹ multisite λ dynamics (MS λ D),^{22–24} enveloping distribution sampling,^{25,26} and many others.^{27,28} These methods have most commonly been applied to compute solvation free energies of ligands,^{22,29,30} drug binding free energies,^{31,32} and pKa values of titratable groups in proteins.^{33–36} Studies using free energy methods to estimate changes in peptide and protein free energy upon side chain mutation are much rarer.^{37–41} Last year, two large-scale applications of free energy methods to protein folding and binding upon mutation appeared.^{42,43} Such studies could in principle be used to guide protein design, but in practice are far too slow to be of practical use in driving design. However, in one case, rigorous alchemical free energy calculations were used in post-design evaluation to assess thermodynamic stability and eliminate a less promising sequence.⁴¹

MS λ D is a unique free energy technique with several features that make it sufficiently efficient to drive protein design. First, MS λ D is highly efficient, and

requires one to two orders of magnitude less computation than other free energy methods to achieve comparable levels of precision.^{24,38} Second, MS λ D can be tuned to screen the most promising sequences from a larger set without wasting sampling on the remaining sequences.^{22,44} Third, MS λ D is highly scalable, allowing the comparison of hundreds of physical systems in a single simulation,⁴⁵ whereas conventional techniques like FEP and TI require independent simulations for each pairwise comparison of physical states. Scalability to combinatorial sequence spaces allows MS λ D to mitigate the high cost of MD simulations by searching much larger swaths of sequence space than FEP or Rosetta during each *in silico* evolutionary step. λ dynamics is over two decades old, but has matured substantially in recent years via generalization to multiple sites,²⁴ the use of implicit constraints to focus sampling on physical endpoint states,⁴⁶ enhanced sampling with biasing potential replica exchange (BP-REX),⁴⁷ adaptive landscape flattening (ALF) to remove alchemical barriers,⁴⁸ and the use of soft-core interactions.^{48,49}

In this study, we seek to assess the accuracy limits and demonstrate the scalability of MS λ D for use in future protein design projects. T4 lysozyme is an ideal test system, as changes in folding free energy have been measured for hundreds of mutants through the work of Brian Matthews and coworkers.⁵⁰ T4 lysozyme residue L133 was also the system used by Kollman in the first free energy simulations of protein mutation.^{37,51,52} We begin by computing single site mutational free energies at seven sites, A42, A98, L99, M102, M106, V149, and F153, which were chosen because they possessed multiple mutants and a large range of folding free energy changes. Predicted $\Delta\Delta G_{\text{Folding}}$ agree well with experiment (Pearson correlation of 0.914 and mean unsigned error of 1.19 kcal/mol). Next, we turned our attention to simultaneous mutations at multiple sites exploring up to 240 sequences concurrently. Agreement for these systems is of roughly equal quality (Pearson correlation of 0.860 and mean unsigned error of 1.12 kcal/mol). The ability of MS λ D to predict protein stability with high accuracy in these large combinatorial sequence spaces suggests this method holds promise for identifying the most favorable mutations in future protein design projects.

Results

Single site mutants

Mutation sites were chosen from and compared against the experimental data in Reference [50]. Single site mutants were chosen according to several criteria: first, stabilities at a particular site must have been measured at the same pH in the C54T/C97A background, a mutant with the one native disulfide bond between C54 and C97 removed. Second, at least three sequences (native and two mutants), excluding G and P for technical reasons, must have been

measured. Finally, the range of stabilities must be at least 3.5 kcal/mol. Seven sites, listed in Table I, met these criteria.

MS λ D simulations were run on each of the seven single site mutants featuring three to nine side chains per simulation, each scaled by its own λ variable, according to the procedure outlined in Methods. Briefly, the landscape was flattened using an updated ALF framework to optimize sampling, and then five independent trial simulations were run for each site for either 40 ns without variable bias replica exchange (VB-REX, see Methods) or 20 ns with VB-REX. The entire procedure was run once with force switching (FSWITCH) electrostatics⁵³ and once with particle mesh Ewald (PME) electrostatics^{36,54,55} to explore the influence of long range electrostatics on the free energy changes.

The overall agreement with experiment is shown in Figure 1. FSWITCH and PME yielded comparable accuracy with Pearson correlations of 0.914/0.893, mean unsigned error (MUE) of 1.19/1.10 kcal/mol, and root mean squared error (RMSE) of 1.39/1.50 kcal/mol, respectively.

In most cases, mutations within an individual site had substantially higher correlations and lower RMSE than when all the sites were combined (Table II). Three sites, A42, A98, and M102, proved difficult to sample in preliminary simulations, consequently VB-REX was employed to sample these sites more aggressively. These sites proved difficult for two reasons. M102 involved a charge change, which is known to cause serious artifacts in free energy calculations,⁵⁶ via mutation to a buried lysine. A42 and A98 both involved mutations from small to large residues in the protein core, which in general may require relaxation of slow degrees of freedom to accommodate the extra steric bulk.

To investigate whether relaxation of slow degrees of freedom limited convergence in these particular mutations, crystal structures for each of the seven sites^{57–65} were visually inspected to identify distances that correlate with mutation for use as structural reaction coordinates. Distances averaged over all frames where a particular substituent was on ($\lambda > 0.99$) agreed fairly well with the crystal structures, though slight thermal expansion of about 0.1 Å was seen in simulations and distances for S and T were

Table I. Single Site Mutants

Site	Mutations	pH
A42	FILSV	3.0
A98	CFILMSVW	3.0
L99	AFIMV	3.0
M102	AK	5.4
M106	KL	3.0
V149	CIMST	5.4 ^a
F153	AILMV	3.0

^a V149T was measured at a different pH (3.0) from other mutations at this site

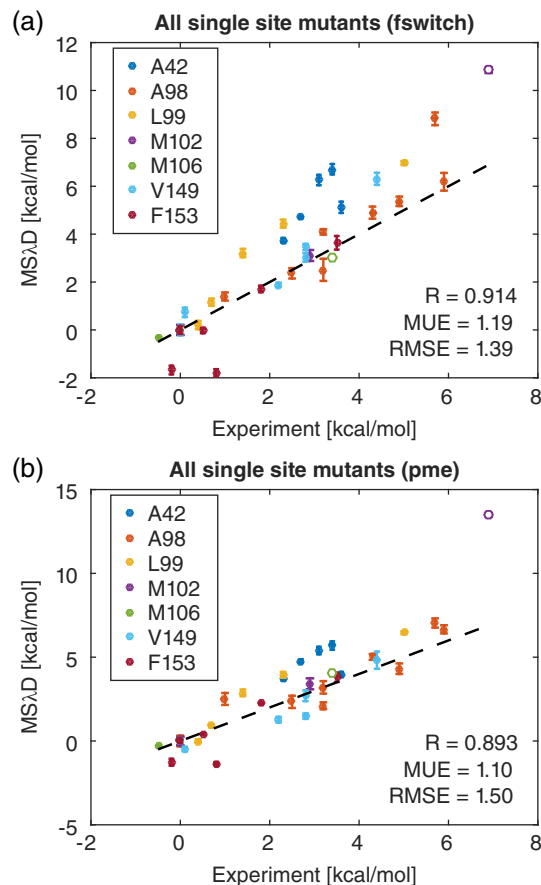


Figure 1. Comparison between predicted MS λ D and experimental folding free energies at all seven mutation sites using (a) FSWITCH and (b) PME electrostatics. Open circles denote charged mutants M102K and M106K. Dashed lines denote $y = x$, and free energies are plotted relative to wild type with C54T/C97A.

overestimated by 0.5 Å (Supporting Information Table S3). Crystal structure analysis also suggests a possible explanation for the large error in M102K:

Table II. Site Specific Results

Site	RMSE ^a	R
FSWITCH		
A42	1.12	0.952
A98	1.01	0.938
L99	0.97	0.968
M102	1.82	0.989
M106	0.25	1.000
V149	0.71	0.960
F153	0.98	0.884
All	1.39	0.914
PME		
A42	0.92	0.920
A98	0.82	0.926
L99	0.82	0.970
M102	2.99	0.983
M106	0.29	0.999
V149	0.59	0.944
F153	0.91	0.897
All	1.50	0.893

^a Centered RMSE in kcal/mol: $\left(\frac{(\Delta\Delta G_{\text{MS}\lambda\text{D}} - \Delta\Delta G_{\text{Exp}})^2}{n} \right)^{1/2}$

although M102K remains buried, S_δ in M106 reorients to directly contact M102K. Interaction with a polarizable sulfur atom likely ameliorates the cost of charge burial but is neglected in our fixed charge models.

As all reaction coordinates identified in crystals were demonstrated to correlate with λ in our simulations (Supporting Information Table S3), the autocorrelation time in these reaction coordinates gives a lower bound for time scale of relaxations that are necessary to obtain converged estimates of free energy. Autocorrelation times are shown in Figure 2 and show a clear split between more rapidly relaxing sites (L99, M106, and V149) and more slowly relaxing sites (A42, A98, M102, and F153). Notably, three out of four of the more slowly relaxing sites were sites that required VB-REX for efficient sampling. These results also suggest that A98 sampling could use further improvement due to its long autocorrelation time.

Multisite mutants

Encouraged by our findings from the single site mutation studies, we turned our attention to multisite mutants. Three systems comprising 3, 4, and 5 mutation sites were selected from methionine scan and core redesign studies (Table III). The three site, four site, and five site systems comprised 8 ($2 \times 2 \times 2$), 24 ($2 \times 2 \times 3 \times 2$), and 240 ($3 \times 5 \times 4 \times 2 \times 2$) sequences with all combinatorial permutations of the individual mutations in Table III. Of these 8, 24, and 240 sequences, only 6, 14, and 9 have been experimentally measured at the same pH,⁵⁰ (see

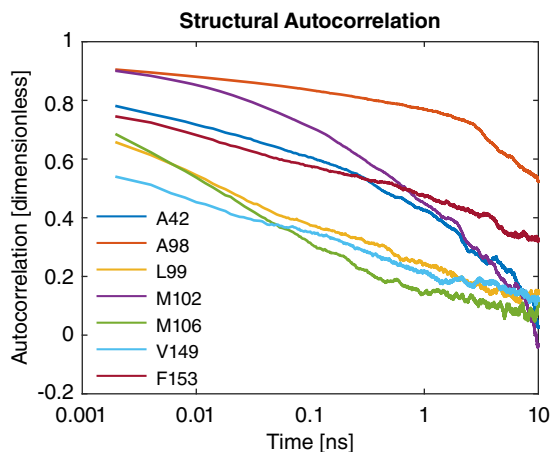


Figure 2. Autocorrelation in structural reaction coordinates at all seven sites (see Supporting Information for reaction coordinate definitions). Sites divide into three groups: L99, M106, and V149 decay rapidly; A42, M102, and F153 decay more slowly; and A98 decays very slowly. This division roughly parallels the sites requiring VB-REX and the sites not requiring VB-REX. At sites A42 and M102 two distances were chosen to capture the variety of structural changes observed. The autocorrelation in the distance more strongly correlated with λ is plotted; the other autocorrelation decays more rapidly in both cases.

Table III. Multisite Mutants

Site	Mutations
3 Site pH = 5.4	
I17	M
I27	M
L33	M
4 Site pH = 3.0	
L99	F
M102	L
V111	FI
F153	L
5 Site pH = 3.0	
L121	AI
A129	LMVW
L133	AMV
V149	I
F153	L

Supporting Information Tables S5–S9 for identities of these sequences), so only results for these sequences are compared with experiment.

Interactions between side chains at different sites are scaled by the product of their λ variables, so mutating side chains at two sites only interact when they are both on, which allows MS λ D to explicitly account for coupling between sites. VB-REX was applied to all systems to increase sampling near the alchemical endpoints, otherwise most endpoints were not sampled using a λ cutoff of 0.99 in the five site system.

Results are shown in Figure 3 and Table IV. The four site system had a narrower range of stabilities than the other two systems and consequently had a lower Pearson correlation with experiment; however, the four site system also had a lower RMSE with experiment, demonstrating the ability of MS λ D to quantify the small effect of these mutations. Overall, Pearson correlations were 0.860/0.845, the RMSEs were 1.28/1.11 kcal/mol, and the MUEs were 1.12/1.02 kcal/mol when FSWITCH and PME respectively were compared with experiment. The agreement with experiment of these multisite mutational results are strikingly comparable with the single site results, suggesting errors do not increase rapidly with the number of mutations.

In order to determine how much error came from the force field and how much came from sampling, each of the sites was mutated individually in single site λ dynamics simulations (SS λ D) run with VB-REX. These point mutations were then compared against the same point mutations in the MS λ D ensemble (Fig. 3(c)) and Supporting Information (Fig. S1). SS λ D and MS λ D agreed closely in the three site system (Supporting Information Fig. S1). Differences for the four site and five site systems were slightly outside of statistical uncertainty, suggesting there are reproducible differences that we attribute to lack of sampling in the MS λ D simulations. These differences are substantially smaller than the deviation from experiment, suggesting most of

the deviation is due to force field errors or shared approximations of the unfolded ensemble.

It should be noted that of the 240 sequences in the five site system, either five or two sequences were not sampled and 23 or 13 sequences were insufficiently sampled across all independent trials to estimate statistical significance in the FSWITCH or PME simulations, respectively (Supporting Information Tables S7–S9). The sequences exhibiting these characteristics were all less stable than the average of the other 210 sequences, typically by 2–3 kcal/mol. This outcome is likely due to cooperative effects between pairs of conflicting mutations. Since such sequences are particularly unstable, they are of less interest in protein design, and inability to sample them within the context of an MS λ D design process is only a minor concern.

Discussion

Comparison with other methods

Due to the importance of changes in protein stability upon mutation, a wide variety of techniques exist to predict $\Delta\Delta G_{\text{Folding}}(S_1 \rightarrow S_2)$. Techniques range from fast physics-based, knowledge-based, and machine learning approaches^{3,66,67}, to implicit solvent methods^{68,69}, to rigorous alchemical free energy calculations in explicit solvent.^{37–43}

In one study comparing a variety of fast prediction methods including Rosetta,³ FoldX,⁶⁷ CC/PBSA⁶⁸ and others, MUEs ranged from 1.00 to 1.68 kcal/mol and Pearson correlations ranged from 0.26 to 0.59.⁶⁶ While some of these methods may appear to be competitive with MS λ D in terms of MUE, this is because these methods are designed to minimize error and thus make conservative estimates, while most errors by MS λ D are overestimates. Consequently, the Pearson correlations, which normalize out overprediction and underprediction, reveal MS λ D clearly gives superior predictions. It is noteworthy that the worst performing algorithm was Rosetta, which suggests current design algorithms are still able to design highly stable foldable proteins despite poor prediction of free energies because they capture the underlying physics of hydrophobic burial and side chain packing. Consequently, MS λ D may enable even more accurate design of proteins, because in addition to these effects, it also captures backbone flexibility and ligand binding energetics, which are poorly modeled by current design algorithms.^{14,17,70,71}

It is also instructive to compare MS λ D directly to other alchemical free energy methods. Most alchemical free energy studies of protein mutations have been small^{37–41}; however, two large scale tests of FEP for calculating folding free energies⁴² and protein–protein binding free energies⁴³ recently appeared. The binding free energy study achieved excellent accuracy (0.68 kcal/mol RMSE, Pearson correlation

0.70) due to the small data range, aggressive sampling (100 ns per window in some cases), neglect of charge changing mutations, and ability to ignore the unfolded ensemble.^{43,*}

The FEP folding free energy study allows a more direct comparison with our results because both studies must grapple with the unfolded ensemble. In a survey of 712 mutations in 10 proteins, that study achieved a Pearson correlation of 0.63, an RMSE of 2.06 kcal/mol,[†] and an MUE of 1.58 kcal/mol.⁴² MS λ D achieved superior results with Pearson correlations of 0.914/0.893, RMSEs of 1.39/1.50 kcal/mol, and MUEs of 1.15/1.07 kcal/mol for FSWITCH/PME, but the differences in the mutation sets call for a more nuanced comparison. First, 25% of the mutations in the FEP study involved charge changes, while only 6.3% of mutations in the present study did. Removing these mutants gives FEP a correlation coefficient of 0.69, an RMSE of 1.84 kcal/mol,[‡] and an MUE of 1.38 kcal/mol. MS λ D results improve to correlation coefficients of 0.901/0.896, RMSEs of 1.30/1.09 kcal/mol, and MUEs of 1.12/0.93 kcal/mol for FSWITCH/PME. Another noteworthy difference is in the relative number of large to small mutations (decreasing number of heavy atoms), which may simply leave a void, versus small to large mutations (increasing number of heavy atoms), which may require slow relaxation of the protein core to relieve a steric clash. The FEP dataset is comprised of mostly large to small mutations derived from alanine and glycine scans; on average, mutations involved loss of three heavy atoms, and only 7.6% mutations involved an equally sized or larger side chain. In the MS λ D dataset, 68.8% of the mutations were to an equally sized or larger side chain. The small to large mutants were harder to sample and gave larger errors, suggesting the MS λ D dataset is a more challenging dataset. Removing the large to small and equal-sized mutants from the MS λ D dataset gives roughly the same agreement with experiment as removing the charge change mutants, and does not result in further improvements because only 10 mutants remain, and of these, half are to F153,

*In studies of protein binding the two relevant ensembles are the bound and unbound states, both of which can be sampled starting from well defined structures, whereas in protein folding the two ensembles are folded and unfolded. Due to the conformational heterogeneity and slow relaxation in the true unfolded ensemble, it is typically approximated using a short peptide. The approximations in the unfolded ensemble seem to be one of the primary causes of the difference in accuracy between References 42,43.

†Reference 42 reported an RMSE of 2.27 kcal/mol, but appears to use RMSE relative to native, rather than centered RMSE. Given that a mean signed error of 0.95 kcal/mol was observed, the centered RMSE would be 2.06 kcal/mol.

‡Reference 42 reported an RMSE of 2.07 kcal/mol. If the mean signed error remained 0.95 kcal/mol, the centered RMSE would be 1.84 kcal/mol.

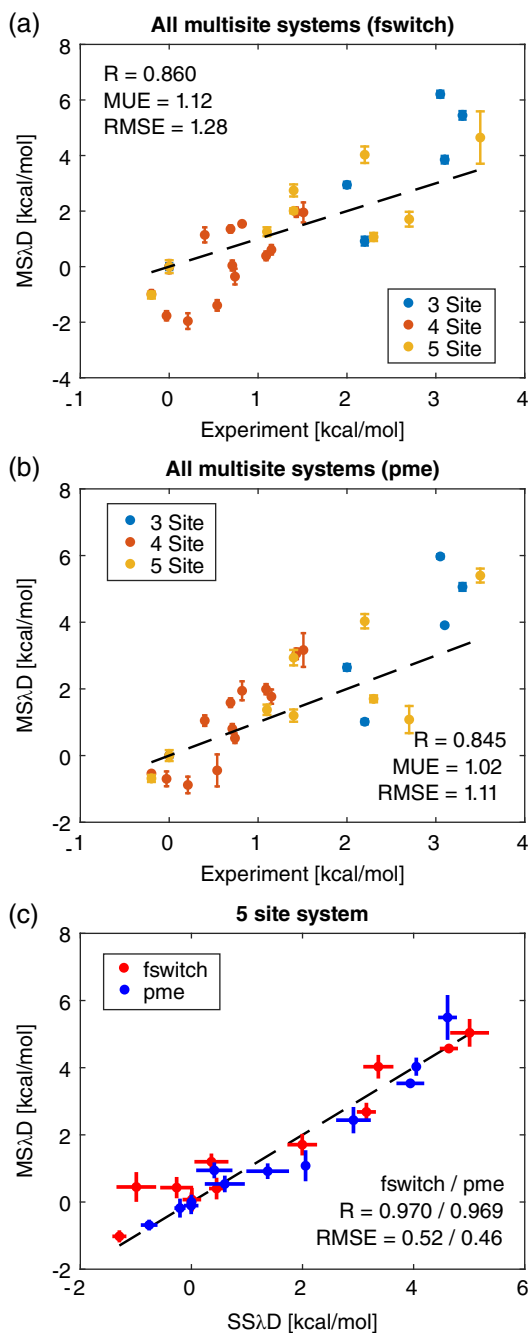


Figure 3. Accuracy of MSΔD in multisite systems. Comparison of multisite mutant results with experiment using (a) FSWITCH and (b) PME electrostatics. (c) Comparison of 5 site results from MSΔD dynamics with more converged results run on a single site only (SSΔD). Comparison with SSΔD is only affected by sampling errors, comparison with experiment is affected by sampling and force field errors. Dashed lines denote $y = x$, and free energies are plotted relative to wild type with C54T/C97A.

where the force field may not fully capture aromatic interactions.

The improved accuracy of MSΔD relative to conventional FEP may be attributed to a variety of causes. First, while both studies are large systematic studies, the FEP study tackled more than an order of magnitude more mutations. Consequently, the FEP

study merely identified poorly converged mutants and noted that they degraded the quality of the results. In contrast, due to its smaller scope, the present study could identify poorly converged mutants and focus sampling on them with VB-REX and reoptimization of biasing potential parameters. Second, the accuracy of both methods is limited by the force field used, and it is possible that the CHARMM36 force field used here is more accurate than OPLS3, though at least one study suggests that the two force fields to have comparable accuracy.⁴³ Finally, MSΔD is more computationally efficient; one study cites a factor of 20–30⁴⁵; see Supporting Information for further discussion. Here, MSΔD requires 0.22 simulations per mutant (32 mutants in 7 simulations), while FEP requires 12–16 simulations per mutant (55–73 times more simulations per mutant).⁸ This allows MSΔD to use longer 20–40 ns simulations with multiple independent trials, while FEP was limited to a single trial of 5 ns per simulation per window. As in Methods, substantial differences were noted between the biasing potentials obtained during the 5 ns flattening runs and the optimal biases in MSΔD production, suggesting the system was still relaxing after 5 ns. Auto-correlations in structural reaction coordinates confirm this for some sites (Fig. 2). Consequently, the FEP simulations likely did not fully relax to equilibrium.

Looking forward

The present study allows a comparison of FSWITCH and PME electrostatics. In the systems considered here, FSWITCH and PME achieved comparable agreement with experiment. While the differences are small, FSWITCH performed better for single site mutants until charge mutations (which PME was not corrected for) were removed. For multisite mutants, PME performed better in terms of RMSE. Moving forward it will be worthwhile to monitor the relative performance of the two methods, as each has advantages. PME is more computationally efficient than FSWITCH because shorter cutoffs may be used. PME appears to perform worse if uncorrected for charge changes,⁵⁶ but likely stabilizes the rest of the protein structure more effectively.

Currently, MSΔD does not treat all amino acid mutations equally. In particular, charge changes are handled less well than charge conserving mutations, and mutations to or from G and P have been avoided due to technical challenges. Corrections for charge changes with PME electrostatics have been derived,⁵⁶

[§]The use of replica exchange complicates the matter. Of the seven MSΔD simulations, four used a single replica and three used three replicas for a total of 13 replicas for 32 mutants. The 12–16 FEP simulations per mutant are coupled by replica exchange, so they are technically a single simulation, but they are just as expensive as 12–16 independent simulations, though replica exchange improves the quality of the results.

Table IV. Multisite Results

System	SS λ D ^a		Experiment ^b	
	RMSE ^c	R	RMSE ^d	R
FSWITCH				
3 site	0.04	1.000	1.43	0.841
4 site	0.47	0.946	0.98	0.726
5 site	0.52	0.970	1.03	0.806
All			1.28	0.860
PME				
3 site	0.10	0.999	1.29	0.856
4 site	0.24	0.981	0.87	0.904
5 site	0.46	0.969	1.15	0.790
All			1.11	0.845

^a Errors due to only sampling^b Errors due to force field, pentapeptide approximation of unfolded ensemble, and sampling^c Centered RMSE in kcal/mol: $((\Delta\Delta G_{\text{MS}\lambda\text{D}} - \Delta\Delta G_{\text{SS}\lambda\text{D}})^2) - (\Delta\Delta G_{\text{MS}\lambda\text{D}} - \Delta\Delta G_{\text{SS}\lambda\text{D}})^2)^{1/2}$ ^d Centered RMSE in kcal/mol: $((\Delta\Delta G_{\text{MS}\lambda\text{D}} - \Delta\Delta G_{\text{Exp}})^2) - (\Delta\Delta G_{\text{MS}\lambda\text{D}} - \Delta\Delta G_{\text{Exp}})^2)^{1/2}$

but the recent FEP folding study demonstrated that large errors remain even when these corrections are applied.⁴² Such errors have been attributed to sampling,⁴² but some error would remain even with perfect sampling due to the fixed protonation state of mutating and surrounding residues. MS λ D is ideally poised to consider such protonation changes with the use of constant pH MD simulations, which utilize the same MS λ D formalism.^{33–36} Glycine mutations, which were considered in the FEP studies,^{42,43} require the ability to change the charge on the C $_{\alpha}$ atom and to scale CMAP interactions^{72,73} by λ , which are not currently implemented in the CHARMM molecular dynamics package. Proline mutations are even more challenging, requiring topology changes via the opening of a ring.⁷⁴ Such mutations could be allowed through the use of soft bonds^{75,76} and other advances. Overcoming these limitations with charged, G, and P mutants represents an important direction for future MS λ D development.

The present study aimed to assess the level of accuracy attainable with MS λ D, not necessarily to obtain the desired free energy changes with the

Table V. Comparison of Alchemical Methods

Method	Mutants	R	RMSE	MUE
All Point Mutants				
FEP (Reference [42])	712	0.63	2.06 ^a	1.58
MS λ D with FSWITCH	32	0.914	1.39	1.15
MS λ D with PME	32	0.893	1.50	1.07
Charge Conserving Point Mutants				
FEP (Reference [42])	534	0.69	1.84 ^b	1.38
MS λ D with FSWITCH	30	0.901	1.30	1.12
MS λ D with PME	30	0.896	1.09	0.93

^a The originally cited RMSE of 2.27 was uncentered^b The originally cited RMSE of 2.07 was uncentered

greatest efficiency. In particular, full production was run with five independent trials, and then discarded if the biasing potentials were determined to be suboptimal. Running fewer production runs to see if biases were converged before launching all five trials^{||} or combining subsequent production runs together with the weighted histogram analysis method (WHAM)⁷⁷ would likely allow comparable precision with even less computational expense and should be considered in the future.

Sampling is a constant concern, and it is worth noting the sampling difficulties encountered with this approach. While MS λ D gives good results for even the nine sequence A98 system, the very long structural autocorrelation time in this system (Fig. 2) reveals sampling difficulties that are obscured by VB-REX. Individual replicas are either trapped in a state transitioning slowly between W and F or in a state rapidly transitioning between the remaining seven side chains. As both states are included in the full set of replicas, reasonable free energy estimates can still be obtained. However, advanced sampling techniques beyond VB-REX, such as orthogonal space random walk (OSRW)²⁷ or replica exchange with solute tempering (REST)⁷⁸ may allow more facile transitions between all mutants and yield superior results. In addition, biasing potentials changed between the 5 ns flattening simulations and the full length production simulations, which suggests that the system does not fully relax in 5 ns. It is likely the system is still not fully relaxed after 20–40 ns, so longer production simulations might allow further relaxation, which could improve agreement with experiment.

Prospects of MS λ D for protein design

The FEP folding study offered discouraging prospects for computational protein design, suggesting that favorable mutations could only be enriched by a factor of three above random, or possibly more if aggressive filtering were used.⁴² Such a strategy is not accurate enough to generate new stable sequences and thus requires constant experimental testing, and is of use primarily for designing enriched libraries for directed evolution. While useful, this may represent a step backward from the current state of the art in enzyme design, because it is perturbative and requires an existing enzyme with initial activity as a starting point, whereas programs like Rosetta are generative and can propose an enzyme with a new function as a starting point for directed evolution.^{12,13,79}

We believe the present study offers more potential for two reasons. First, errors are smaller, suggesting MS λ D can enrich favorable mutations at a greater rate. Second, errors for multisite mutants are comparable with errors for single site mutants, so

^{||}Reference 43 optimized computational expense in a similar fashion

errors appear not to compound significantly. This suggests that while the force field and experiment may disagree slightly, they capture the same physical effects, and do not diverge rapidly as one moves through sequence space. Therefore, MS λ D offers the ability to accurately search combinatorially large sequence spaces, which is computationally intractable with FEP. Consequently, by searching larger regions of sequence space, MS λ D possesses more potential for finding favorable cooperative mutations that can be aggressively screened.

The strength of MS λ D lies in its ability to consider large combinatorial sequence spaces. Modern protein design algorithms function through a Monte Carlo annealing search of sequence space, in which tens to hundreds of thousands of point mutations are proposed sequentially⁸⁰ with dozens to hundreds of alternative designs attempted in parallel.^{9,11} The improved accuracy of alchemical free energy approaches comes at a high computational cost: multisite simulations with MS λ D required 54 h for flattening and 26 h for production, which limits alchemical simulations to several dozen steps in sequence space. The point mutations considered by FEP are inadequate to optimize a protein sequence in such a small number of steps; however, MS λ D enables much larger jumps in sequence space which may allow sequence optimization in a small number of steps.

The ability of MS λ D to consider concurrent mutations is critical for efficiently navigating the topological complexities of sequence space^{81,82} such as fitness valleys. A simple example of a fitness valley is a case where a small and a large residue are in contact in a well-packed core. Even if switching the residue sizes is favorable, the pathways to do so lie through a fitness valley because either an unfavorable void or steric clash is introduced before the second compensating mutation can occur. Current design algorithms can cross fitness valleys of several $kT_{\text{Selection}}$ through the sheer number of Monte Carlo steps. In contrast, alchemical methods like FEP that only propose point mutations cannot cross fitness valleys in the small number of Monte Carlo steps available. MS λ D can overcome this limitation by considering concurrent mutations, which allow movement directly to the other side of the fitness valley in a single Monte Carlo step.

Conclusions

Multisite λ dynamics is a highly efficient and scalable alchemical free energy method that allows high accuracy estimation of changes in folding free energy for both single site and multisite mutants. The efficiency enables longer simulations that appear to relax more closely to equilibrium, while scalability allows the exploration of large combinatorial sequence spaces. This study represents a first step towards protein design with MS λ D by quantifying the accuracy of

MS λ D and demonstrating its scalability to large sequence spaces.

Additional work is required to extend MS λ D to poorly treated charged residues and technically excluded G and P residues. Extension of MS λ D to charged residues will be aided by the identical formalism underlying MS λ D and constant pH MD. These advances will enable protein design with the full palette of amino acids.

The strength of MS λ D lies in its ability to consider large combinatorial sequence spaces, and find cooperative mutations within them that enable efficient navigation of the protein sequence landscape. This scalability is essential due to the high cost of alchemical free energy calculations. The accuracy, efficiency, and combinatorial scalability of MS λ D suggest it will become a powerful tool for protein design in the future.

Methods

Multisite λ dynamics

Like most free energy methods, MS λ D relies on a thermodynamic scheme such as the one shown in Figure 4 to calculate free energy differences. As free energy is a state function, sequences S_1 and S_2 can be compared by calculating $\Delta\Delta G_{\text{Folding}}(S_1 \rightarrow S_2)$ in one of the two ways:

$$\begin{aligned} \Delta\Delta G_{\text{Folding}}(S_1 \rightarrow S_2) &= \Delta G_{\text{Folding}}(S_2) - \Delta G_{\text{Folding}}(S_1) \\ &= \Delta G_{\text{Folded}}(S_1 \rightarrow S_2) - \Delta G_{\text{Unfolded}}(S_1 \rightarrow S_2) \end{aligned} \quad (1)$$

While the first expression converges slowly because many folding transitions are required for convergence, the second expression converges rapidly and is readily obtained by alchemically transforming from S_1 to S_2 in both the folded and unfolded ensembles.

Alchemical free energy methods accomplish this transformation through a coupling parameter λ , which tunes the potential energy function from sequence S_1 at $\lambda = 0$ through nonphysical intermediates to sequence S_2 at $\lambda = 1$. In traditional alchemical methods, several MD simulations are run at closely

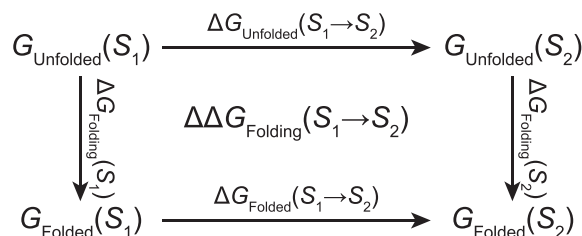


Figure 4. Thermodynamic scheme that enables the calculation of $\Delta\Delta G_{\text{Folding}}(S_1 \rightarrow S_2)$, the change in folding free energy upon mutation from sequence S_1 to sequence S_2 . The folded ensemble is modeled starting from the crystal structure, the unfolded ensemble is modeled using a capped pentapeptide.

spaced fixed values of λ , and free energy differences are then estimated using FEP,²⁰ TI,²¹ or multistate Bennett acceptance ratio (MBAR).⁸³

In MS λ D, λ is not fixed, but is rather a dynamic variable with a pseudomass and velocity that is propagated along with the other coordinates in the equations of motion. Consequently, MS λ D only requires two simulations to estimate $\Delta\Delta G_{\text{Folding}}(S_1 \rightarrow S_2)$, one for the folded ensemble and one for the unfolded ensemble. Furthermore, while traditional free energy methods are limited to pairwise comparisons between sequences, MS λ D can evaluate multiple mutations at the same site or even multiple sites using a multidimensional λ space.²⁴

Updated techniques

Recently, landscape flattening and soft-core interactions have been shown to give marked improvements in MS λ D predictions of $\Delta\Delta G$.⁴⁸ Those methods are used here with several important updates, including a new biasing potential that better flattens soft-core landscapes and a more general flattening algorithm. A generalization of BP-REX⁴⁷ called VB-REX is introduced to improve sampling on multisite systems. Finally, we extend the treatment of electrostatics from FSWITCH to PME following the work of Shen.³⁶ See Supporting Information for full details.

The previous set of biasing potentials were designed to flatten alchemical free energy landscapes observed with hard-core interactions. An additional biasing potential of the form

$$V_{\text{Skew}} = \sum_s^M \sum_i^{N_s} \sum_{j \neq i}^{N_s} \chi_{si,sj} \lambda_{sj} (1 - \exp(-\lambda_{si}/\sigma)), \quad (2)$$

with $\sigma = 0.18$, was found to give improved fits to free energy profiles obtained with soft-core interactions. The previous flattening algorithm handled a variety of special cases on an individual basis, and could not be easily modified to accommodate additional biasing potentials such as Equation 2.

In the present work a more generalizable approach was used to learn biasing potential parameters to flatten the landscape. Free energy profiles are computed from several recent sampling iterations using WHAM,⁷⁷ the entropy of the implicit constraints is subtracted,⁴⁶ and changes in the biasing potential parameters $\Delta\varphi_{si}$, $\Delta\psi_{si, sj}$, $\Delta\chi_{si, sj}$, and $\Delta\omega_{si, sj}$ are computed for the next round of sampling. A scoring function for proposed changes in parameters at each site s is constructed as

$$E = \sum_p^{\text{Profiles Bins}} \sum_b \frac{w_{pb}}{2} (G_{pb} + \Delta G_{pb} - \bar{G}_p)^2 + \sum_i^{\text{Biases}} \frac{k}{2\gamma_0^2} \Delta\gamma_i^2 \quad (3)$$

$$\Delta G_{pb} = \sum_i^{\text{Biases}} \frac{\partial G_{pb}}{\partial \gamma_i} \Delta\gamma_i, \quad (4)$$

where γ_i runs over all biasing potential parameters in the set $\{\varphi_{si}, \psi_{si, sj}, \chi_{si, sj}, \omega_{si, sj}\}$. The first term of Equation 3 assesses landscape flatness and predicts the effects of changes in biasing parameters on landscape flatness through the linear approximation of ΔG_{pb} , and the second term prevents overfitting. The scoring function is minimized at each site s independently using least squares with \bar{G}_{pb} and $\Delta\gamma_i$ as free parameters. This flattening algorithm can be easily generalized to other biasing potentials.

As in the previous version of ALF,⁴⁸ 50 cycles of optimization were run with 100 ps of equilibration and 100 ps of sampling, followed by 10 cycles of optimization with 100 ps of equilibration and 1 ns of sampling. We observed that biasing potentials at the end of the 1 ns cycles were still suboptimal for some of the more slowly converging sites (e.g. A42, A98, and M102), so an additional phase with three cycles of 2 ns of equilibration and 5 ns of sampling were added to the protocol for all sites. For production, five independent simulations were run for 40 ns each, or 20 ns when using VB-REX. It was observed often that some substituents stopped being sampled after the system had relaxed, therefore the first quarter of all production runs was discarded as equilibration. If sampling was poor during the remainder of production, landscape flattening parameters were reestimated and an additional one to three rounds of production were run, until all substituents were being sampled efficiently during the remainder of production. (See Supporting Information Table S2 for details.)

Inspired by BP-REX,⁴⁷ a new form of biasing potential replica exchange on the variable biases (VB-REX) was implemented to encourage interconversion of side chains in low replicas, and endpoint sampling in high replicas. This was achieved by changing the biasing potential parameters between neighboring replicas by $\Delta\psi_{si, sj} = 2B$ and $\Delta\omega_{si, sj} = 0.5B$ for all si - sj combinations, in order to change the barrier height by B . Parameters vary with the number of sites, where B is tuned to optimize exchange rates, while the number of replicas and the replica flattened by ALF are tuned to sample primarily physical endpoint states in the final replica. For the five mutation site system, even the small 1.8 kcal/mol barrier in the final replica was adequate to increase the sampling of approximately physical states ($\lambda > 0.99$ at all sites) by about two orders of magnitude.

Three different electrostatic models were considered in this work: force shifting (FSHIFT), FSWITCH, and PME. In preliminary simulations, FSHIFT showed poorer correlation with experiment and roughly 0.5 kcal/mol greater RMSE, so only FSWITCH and PME results are reported. FSWITCH multiplies the electrostatic force by a function that decays smoothly from one to zero within a narrow range, but neglects long range electrostatics.⁵³ PME splits the electrostatic

potential into a short range piece accurately calculated with cutoffs and a smooth long range piece accurately calculated in Fourier space.^{54,55} Using PME in MS λ D simulations is not trivial, and we follow the approach of Shen.³⁶ Alchemical simulations involving a net change in charge can result in significant artifacts.^{84–86} While corrections for these effects have been devised for PME,⁵⁶ they were not applied in the present work as only two of the considered mutants involved a charge change. Application of these corrections is deferred until it can be tested more systematically.

Simulation details

Simulations began from the T4 lysozyme structure with the disulfide C54T/C97A mutated out (PDB ID: 1L63).⁸⁷ The unfolded ensemble was modeled through the use of short capped pentapeptides starting from their conformation in the crystal structure. The pentapeptide N-terminus was capped with an acetyl group ($-\text{COCH}_3$) and the C-terminus was capped with an amide group ($-\text{NH}_2$). The unfolded ensemble for multisite systems was treated as an independent sum of the pentapeptide systems for each individual mutation site. Folded lysozyme and unfolded pentapeptides were solvated with TIP3P⁸⁸ in 72 Å and 40 Å cubic boxes respectively giving margins of at least 10 Å on each side. Residues were protonated according to pKa predictions by PropKa⁸⁹ at pH values corresponding to experimental measurements, either 3.0 or 5.4. At a pH of 5.4, all K, R, and H were positive, and all D and E were negative. At a pH of 3.0, only D47, E62, D89, D92, and D159 remained negative. NaCl ions were added to a molality of 100 millimolar, and the system was neutralized by switching an appropriate number of Na^+ to Cl^- .

Simulations were run with the CHARMM molecular dynamics engine^{90,91} using the DOMDEC module⁹² for GPU. Folded simulations achieved roughly 18 ns/day and unfolded pentapeptide simulations achieved 67 ns/day with each replica using four OpenMP threads on an Intel Xeon E5-2650 v3 CPU and an Nvidia GeForce GTX 980 GPU. Simulations used the CHARMM36 force field.^{73,93} Mutating residues were modeled using the patch facility of CHARMM. Every mutation site had a single C_α atom and side chains for each mutation were attached starting at the C_β atom with appropriate bonds, angles, dihedrals, impropers, and 1–4 interactions. Each side chain is assigned to a different block in the CHARMM block module. The nonbonded potential terms of each side chain are scaled by the appropriate λ (or product of λ s for site-site interactions) and do not interact with other side chains at the same site. Soft-core nonbonded interactions were used as described in Reference [48]. Most bonded interactions are not scaled by λ in order to maintain side chain geometry when it is in the off state. Historically, dihedrals have not been scaled in MS λ D, but scaling

dihedrals that include side chain atoms was necessary here, otherwise the side chain $C_\beta-C_\alpha-\text{NH}-\text{C}$ and $C_\beta-C_\alpha-\text{C}-\text{O}$ dihedrals (with dihedral amplitudes of 1.8 and 1.4 kcal/mol) would be double counted by the number of side chains at a site. As these dihedrals involve the same rotatable bonds as protein backbone φ and ψ angles, respectively, such double counting would have adverse consequences on the protein backbone conformations. CMAP interactions^{72,73} do not include side chain atoms, so only one CMAP interaction is present per residue, and scaling it is unnecessary.

Acknowledgment

We gratefully acknowledge funding from the NIH (GM037554 and GM103695) and the NSF (CHE 1506273). We thank Michael Crowley for providing a version of CHARMM which was subsequently modified to make DOMDEC parallelization and replica exchange compatible. We thank Efrosini Artikis for help with PropKa. This work is dedicated to the memory of Peter Kollman and Antti-Pekka Hynninen. Peter was the first to apply alchemical free energy methods to T4 lysozyme mutations. Antti-Pekka wrote most of the GPU code in CHARMM, which enabled this project to be completed in a timely manner.

References

1. Pande V, Grosberg A, Tanaka T (1997) Statistical mechanics of simple models of protein folding and design. *Biophys J* 73:3192–3210.
2. Morcos F, Schafer NP, Cheng RR, Onuchic JN, Wolynes PG (2014) Coevolutionary information, protein folding landscapes, and the thermodynamics of natural selection. *Proc Nat Acad Sci U S A* 111:12408–12413.
3. Rohl CA, Strauss CE, Misura KM, Baker D (2004) Protein structure prediction using Rosetta. *Methods Enzymol* 383:66–93.
4. Leaver-Fay A et al. In: Johnson ML, Brand L, Eds, 2011. *Computer Methods, Part C*. Volume 487. *Methods in Enzymology*: Academic Press; p. 545–574.
5. Dahiyat BI, Mayo SL (1997) De novo protein design: fully automated sequence selection. *Science* 278:82–87.
6. Bolon DN, Mayo SL (2001) Enzyme-like proteins by computational design. *Proc Nat Acad Sci U S A* 98: 14274–14279.
7. Lassila JK, Privett HK, Allen BD, Mayo SL (2006) Combinatorial Methods for Small-Molecule Placement in Computational Enzyme Design. *Proc Nat Acad Sci U S A* 103:16710–16715.
8. Druart K, Palmai Z, Omarjee E, Simonson T (2016) Protein:ligand binding free energies: a stringent test for computational protein design. *J Comput Chem* 37: 404–415.
9. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D (2003) Design of a novel globular protein fold with atomic-level accuracy. *Science* 21: 1364–1368.
10. Karanicolas J, Corn JE, Chen I, Joachimiak LA, Dym O, Peck SH, Albeck S, Unger T, Hu W, Liu G, Delbecq S, Montelione GT, Spiegel CP, Liu DR, Baker D (2011) A

- de novo protein binding pair by computational design and directed evolution. *Mol Cell* 42:250–260.
11. Tinberg CE, Khare SD, Dou J, Doyle L, Nelson JW, Schena A, Jankowski W, Kalodimos CG, Johnsson K, Stoddard BL, Baker D (2013) Computational design of ligand-binding proteins with high affinity and selectivity. *Nature* 501:212–216.
 12. Blomberg R, Kries H, Pinkas DM, Mittl PRE, Grutter MG, Privett HK, Mayo SL, Hilvert D (2013) Precision is essential for efficient catalysis in an evolved Kemp eliminase. *Nature* 503:418–421.
 13. Khersonsky O, Kiss G, Rothlisberger D, Dym O, Albeck S, Houk KN, Baker D, Tawfik DS (2012) Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed Kemp eliminase KE59. *Proc Natl Acad Sci U S A* 109: 10358–10363.
 14. Morin A, Kaufmann KW, Fortenberry C, Harp JM, Mizoue LS, Meiler J (2011) Computational design of an endo-1,4- β -xylanase ligand binding site. *Protein Eng Des Sel* 24:503–516.
 15. Kiss G, Çelebi-Ölçüm N, Moretti R, Baker D, Houk KN (2013) Computational enzyme design. *Angew Chem Int Ed* 52:5700–5725.
 16. Baker D (2010) An exciting but challenging road ahead for computational enzyme design. *Protein Sci* 19: 1817–1819.
 17. Kries H, Blomberg R, Hilvert D (2013) De novo enzymes by computational design. *Curr Opin Chem Biol* 17: 221–228.
 18. Tantillo DJ, Jiangang C, Houk KN (1998) Theozymes and compuzymes: theoretical models for biological catalysis. *Curr Opin Chem Biol* 2:743–750.
 19. Osuna S, Jiménez-Osés G, Noey EL, Houk KN (2015) Molecular dynamics explorations of active site structure in designed and evolved enzymes. *Acc Chem Res* 48: 1080–1089.
 20. Zwanzig RW (1954) High-temperature equation of state by a perturbation method. I. nonpolar gases. *J Chem Phys* 22:1420–1426.
 21. Straatsma TP, Berendsen HJC (1988) Free energy of ionic hydration: analysis of a thermodynamic integration technique to evaluate free energy differences by molecular dynamics simulations. *J Chem Phys* 89: 5876–5886.
 22. Kong X, Brooks CL III (1996) λ -dynamics: a new approach to free energy calculations. *J Chem Phys* 105: 2414–2423.
 23. Knight JL, Brooks CL III (2009) λ -dynamics free energy simulation methods. *J Comput Chem* 30:1692–1700.
 24. Knight JL, Brooks CL III (2011) Multisite λ dynamics for simulated structure-activity relationship studies. *J Chem Theory Comput* 7:2728–2739.
 25. Christ CD, van Gunsteren WF (2007) Enveloping distribution sampling: a method to calculate free energy differences from a single simulation. *J Chem Phys* 126: 184110.
 26. Riniker S, Christ CD, Hansen N, Mark AE, Nair PC, van Gunsteren WF (2011) Comparison of enveloping distribution sampling and thermodynamic integration to calculate binding free energies of phenylethanolamine N-methyltransferase inhibitors. *J Chem Phys* 135: 024105.
 27. Zheng L, Chen M, Yang W (2008) Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems. *Proc Natl Acad Sci U S A* 105: 20227–20232.
 28. Ding X, Vilseck JZ, Hayes RL, Brooks CL III (2017) Gibbs sampler-based λ -dynamics and Rao-Blackwell estimator for alchemical free energy calculation. *J Chem Theory Comput* 13:2501–2510.
 29. Guthrie JP (2009) A blind challenge for computational solvation free energies: introduction and overview. *J Phys Chem B* 113:4501–4507.
 30. Geballe MT, Skillman AG, Nicholls A, Guthrie JP, Taylor PJ (2010) The SAMPL2 blind prediction challenge: introduction and overview. *J Comput Aided Mol Des* 24:259–279.
 31. Steinbrecher TB, Dahlgren M, Cappel D, Lin T, Wang L, Krilov G, Abel R, Friesner R, Sherman W (2015) Accurate binding free energy predictions in fragment optimization. *J Chem Inf Model* 55:2411–2420.
 32. Wang L, Wu Y, Deng Y, Kim B, Pierce L, Krilov G, Lupyan D, Robinson S, Dahlgren MK, Greenwood J, Romero DL, Masse C, Knight JL, Steinbrecher T, Beuming T, Damm W, Harder E, Sherman W, Brewer M, Wester R, Murcko M, Frye L, Farid R, Lin T, Mobley DL, Jorgensen WL, Berne BJ, Friesner RA, Abel R (2015) Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J Am Chem Soc* 137:2695–2703.
 33. Donnini S, Tegeler F, Groenhof G, Grubmüller H (2011) Constant pH molecular dynamics in explicit solvent with λ -dynamics. *J Chem Theory Comput* 7:1962–1978.
 34. Wallace JA, Shen JK (2012) Charge-leveling and proper treatment of long-range electrostatics in all-atom molecular dynamics at constant pH. *J Chem Phys* 137: 184105.
 35. Goh GB, Hulbert BS, Zhou H, Brooks CL III (2014) Constant pH molecular dynamics of proteins in explicit solvent with proton tautomerism. *Prot Struct Func Bioinform* 82:1319–1331.
 36. Huang Y, Chen W, Wallace JA, Shen J (2016) All-atom continuous constant pH molecular dynamics with particle mesh Ewald and titratable water. *J Chem Theory Comput* 12:5411–5421.
 37. Pitera JW, Kollman PA (2000) Exhaustive mutagenesis in silico: multicoordinate free energy calculations on proteins and peptides. *Prot Struct Func Bioinform* 41: 385–397.
 38. Guo Z, Durkin J, Fischmann T, Ingram R, Prongay A, Zhang R, Madison V (2003) Application of the λ -dynamics method to evaluate the relative binding free energies of inhibitors to HCV protease. *J Med Chem* 46: 5360–5364.
 39. Rizzo RC, Wang D-P, Tirado-Rives J, Jorgensen WL (2000) Validation of a model for the complex of HIV-1 reverse transcriptase with sustiva through computation of resistance profiles. *J Am Chem Soc* 122:12898–12900.
 40. Guo Z, Prongay A, Tong X, Fischmann T, Bogen S, Velazquez F, Venkatraman S, Njoroge FG, Madison V (2006) Computational study of the effects of mutations A156T, D168V, and D168Q on the binding of HCV protease inhibitors. *J Chem Theory Comput* 2:1657–1663.
 41. Simonson T, Ye-Lehmann S, Palmi Z, Amara N, Wydau-Dematteis S, Bigan E, Druart K, Moch C, Plateau P (2016) Redesigning the stereospecificity of tyrosyl-trna synthetase. *Prot Struct Func Bioinform* 84: 240–253.
 42. Steinbrecher T, Zhu C, Wang L, Abel R, Negron C, Pearlman D, Feyfant E, Duan J, Sherman W (2017) Predicting the effect of amino acid single-point mutations on protein stability: large-scale validation of MD-based relative free energy calculations. *J Mol Biol* 429: 948–963.
 43. Clark AJ, Gindin T, Zhang B, Wang L, Abel R, Murret CS, Xu F, Bao A, Lu NJ, Zhou T, Kwong PD,

- Shapiro L, Honig B, Friesner RA (2017) Free energy perturbation calculation of relative binding free energy between broadly neutralizing antibodies and the gp120 glycoprotein of HIV-1. *J Mol Biol* 429:930–947.
44. Guo Z, Brooks CL III, Kong X (1998) Efficient and flexible algorithm for free energy calculations using the λ -dynamics approach. *J Phys Chem B* 102:2032–2036.
 45. Vilseck JZ, Armacost KA, Hayes RL, Goh GB, Brooks CL III (2018) Predicting binding free energies in a large combinatorial chemical space using multisite λ dynamics. *J Phys Chem Lett* 9:3328–3332.
 46. Knight JL, Brooks CL III (2011) Applying efficient implicit nongeometric constraints in alchemical free energy simulations. *J Comput Chem* 32:3423–3432.
 47. Armacost KA, Goh GB, Brooks CL III (2015) Biasing potential replica exchange multisite λ -dynamics for efficient free energy calculations. *J Chem Theory Comput* 11:1267–1277.
 48. Hayes RL, Armacost KA, Vilseck JZ, Brooks CL III (2017) Adaptive landscape flattening accelerates sampling of alchemical space in multisite λ dynamics. *J Phys Chem B* 121:3626–3635.
 49. Zacharias M, Straatsma TP, McCammon JA (1994) Separation-shifted scaling, a new scaling method for Lennard-Jones interactions in thermodynamic integration. *J Chem Phys* 100:9025–9031.
 50. Baase WA, Liu L, Tronrud DE, Matthews BW (2010) Lessons from the lysozyme of phage T4. *Protein Sci* 19:631–641.
 51. Veenstra DL, Kollman PA (1997) Modeling protein stability: a theoretical analysis of the stability of T4 lysozyme mutants. *Protein Eng Des Sel* 10:789807.
 52. Wang L, Veenstra DL, Radmer RJ, Kollman PA (1998) Can one predict protein stability? An attempt to do so for residue 133 of T4 lysozyme using a combination of free energy derivatives, PROFEC, and free energy perturbation methods. *Prot Struct Func Bioinform* 32:438–458.
 53. Steinbach PJ, Brooks BR (1994) New spherical-cutoff methods for long-range forces in macromolecular simulation. *J Comput Chem* 15:667–683.
 54. Darden T, York D, Pedersen L (1993) Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems. *J Chem Phys* 98:10089–10092.
 55. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) A smooth particle mesh Ewald method. *J Chem Phys* 103:8577–8593.
 56. Rocklin GJ, Mobley DL, Dill KA, Hünenberger PH (2013) Calculating the binding free energies of charged species based on explicit-solvent simulations employing lattice-sum methods: an accurate correction scheme for electrostatic finite-size effects. *J Chem Phys* 139:184103.
 57. Nicholson H, Anderson DE, Pin SD, Matthews BW (1991) Analysis of the interaction between charged side chains and the α -helix dipole using designed thermostable mutants of phage T4 lysozyme. *Biochemistry* 30:9816–9828.
 58. Blaber M, Lindstrom JD, Gassner N, Xu J, Heinz DW, Matthews BW (1993) Energetic cost and structural consequences of burying a hydroxyl group within the core of a protein determined from Ala \rightarrow Ser and Val \rightarrow Thr substitutions in T4 lysozyme. *Biochemistry* 32:11363–11373.
 59. Liu R, Baase WA, Matthews BW (2000) The introduction of strain and its effects on the structure and stability of T4 lysozyme. *J Mol Biol* 295:127–145.
 60. Eriksson AE, Baase WA, Matthews BW (1993) Similar hydrophobic replacements of Leu99 and Phe153 within the core of T4 lysozyme have different structural and thermodynamic consequences. *J Mol Biol* 229:747–769.
 61. Baldwin E, Baase WA, Zhang X, Feher V, Matthews BW (1998) Generation of ligand binding sites in T4 Lysozyme by deficiency-creating substitutions. *J Mol Biol* 277:467–485.
 62. Dao-Pin S, Anderson DE, Baase WA, Dahlquist FW, Matthews BW (1991) Structural and thermodynamic consequences of burying a charged residue within the hydrophobic core of T4 lysozyme. *Biochemistry* 30:11521–11529.
 63. Lipscomb LA, Gassner NC, Snow SD, Eldridge AM, Baase WA, Drew DL, Matthews BW (1998) Context-dependent protein stabilization by methionine-to-leucine substitution shown in T4 lysozyme. *Protein Sci* 7:765–773.
 64. Xu J, Baase WA, Quillin ML, Baldwin EP, Matthews BW (2001) Structural and thermodynamic analysis of the binding of solvent at internal sites in T4 lysozyme. *Protein Sci* 10:1067–1078.
 65. Gassner NC, Baase WA, Lindstrom JD, Lu J, Dahlquist FW, Matthews BW (1999) Methionine and alanine substitutions show that the formation of wild-type-like structure in the carboxy-terminal domain of T4 lysozyme is a rate-limiting step in folding. *Biochemistry* 38:14451–14460.
 66. Potapov V, Cohen M, Schreiber G (2009) Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details. *Protein Eng Des Sel* 22:553–560.
 67. Guerois R, Nielsen JE, Serrano L (2002) Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J Mol Biol* 320:369–387.
 68. Benedix A, Becker CM, de Groot BL, Caffisch A, Bockmann RA (2009) Predicting free energy changes using structural ensembles. *Nat Methods* 6:3–4.
 69. Zhang Z, Wang L, Gao Y, Zhang J, Zhenirovskyy M, Alexov E (2012) Predicting folding free energy changes upon single point mutations. *Bioinformatics* 28:664–671.
 70. Allison B, Combs S, DeLuca S, Lemmon G, Mizoue L, Meiler J (2014) Computational design of protein-small molecule interfaces. *J Struct Biol* 185:193202.
 71. Morin A, Meiler J, Mizoue LS (2011) Computational design of protein-ligand interfaces: potential in therapeutic development. *Trends Biotechnol* 29:159–166.
 72. MacKerell AD Jr, Feig M, Brooks CL III (2004) Improved treatment of the protein backbone in empirical force fields. *J Am Chem Soc* 126:698699.
 73. Best RB, Mittal J, Feig M, MacKerell AD Jr (2012) Inclusion of many-body effects in the additive CHARMM protein CMAP potential results in enhanced cooperativity of α -helix and β -hairpin formation. *Biophys J* 103:1045–1051.
 74. Liu S, Wang L, Mobley DL (2015) Is ring breaking feasible in relative binding free energy calculations? *J Chem Inf Model* 55:727–735.
 75. Wang L, Deng Y, Wu Y, Kim B, LeBard DN, Wandschneider D, Beachy M, Friesner RA, Abel R (2017) Accurate modeling of scaffold hopping transformations in drug discovery. *J Chem Theory Comput* 13:42–54.
 76. Yu HS, Deng Y, Wu Y, Sindhikara D, Rask AR, Kimura T, Abel R, Wang L (2017) Accurate and reliable prediction of the binding affinities of macrocycles to their protein targets. *J Chem Theory Comput* 13:6290–6300.
 77. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. I. the method. *J Comput Chem* 13:1011–1021.

78. Wang L, Friesner RA, Berne BJ (2011) Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2). *J Phys Chem B* 115:9431–9438.
79. Privett HK, Kiss G, Lee TM, Blomberg R, Chica RA, Thomas LM, Hilvert D, Houk KN, Mayo SL (2012) Iterative approach to computational enzyme design. *Proc Natl Acad Sci U S A* 109:3790–3795.
80. Jin W, Kambara O, Sasakawa H, Tamura A, Takada S (2003) De novo design of foldable proteins with smooth folding funnel: automated negative design and experimental verification. *Structure* 11:581–590.
81. Weinreich DM, Lan Y, Wylie CS, Heckendorn RB (2013) Should evolutionary geneticists worry about higher-order epistasis? *Curr Opin Genet Dev* 23:700–707.
82. Otwinowski J, Plotkin JB (2014) Inferring fitness landscapes by regression produces biased estimates of epistasis. *Proc Natl Acad Sci U S A* 111:E2301–E2309.
83. Shirts MR, Chodera JD (2008) Statistically optimal analysis of samples from multiple equilibrium states. *J Chem Phys* 129:124105.
84. Hummer G, Pratt LR, Garcia AE (1996) Free energy of ionic hydration. *J Phys Chem* 100:1206–1215.
85. Hunenberger PH, McCammon JA (1999) Ewald artifacts in computer simulations of ionic solvation and ion-ion interaction: a continuum electrostatics study. *J Chem Phys* 110:1856–1872.
86. Hub JS, de Groot BL, Grubmüller H, Groenhof G (2014) Quantifying artifacts in Ewald simulations of inhomogeneous systems with a net charge. *J Chem Theory Comput* 10:381–390.
87. Nicholson H, Anderson DE, Pin SD, Matthews BW (1991) Analysis of the interaction between charged side chains and the α -helix dipole using designed thermostable mutants of phage T4 lysozyme. *Biochemistry* 30:9816–9828.
88. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79:926–935.
89. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH (2011) PROPKA3: consistent treatment of internal and surface residues in empirical pK_a predictions. *J Chem Theory Comput* 7:525–537.
90. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* 4:187–217.
91. Brooks BR, Brooks CL III, Mackerell AD Jr, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M (2009) CHARMM: the biomolecular simulation program. *J Comput Chem* 30:1545–1614.
92. Hynninen A-P, Crowley MF (2014) New faster CHARMM molecular dynamics engine. *J Comput Chem* 35:406–413.
93. Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M, MacKerell AD Jr (2012) Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and side-chain χ_1 and χ_2 dihedral angles. *J Chem Theory Comput* 8:3257–3273.