Supporting Information for: Approaching Protein Design with Multisite λ Dynamics: Accurate and Scalable Mutational Folding Free Energies in T4 Lysozyme

Ryan L. Hayes,[†] Jonah Z. Vilseck,[†] and Charles L. Brooks III[*,†,‡]

Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109, United States, and Biophysics Program, University of Michigan, Ann Arbor, Michigan 48109, United States

E-mail: brookscl@umich.edu

[*]To whom correspondence should be addressed
[†]Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109, United States
[‡]Biophysics Program, University of Michigan, Ann Arbor, Michigan 48109, United States

Detailed Explanation of Updated Techniques

The potential for MSλD is given by

$$V = V(x_0,x_0) + \sum_{s}^{M}\sum_{i}^{N_s} \lambda_{si}(V(x_0,x_{si}) + V(x_{si},x_{si})) + \sum_{i}^{M}\sum_{t=s+1}^{M}\sum_{i}^{N_s}\sum_{j}^{N_t} V(x_{si},x_{tj}) + V_{\mathrm{Bias}}(\{\lambda\}), \tag{S1}$$

where λ_{si} is the alchemical coordinate for mutation i at site s, M is the number of mutation sites, N_s is the number of mutants (including native) at a particular site s, $V(a,b)$ is the interaction energy between particles in groups a and b, where x_0 is the group of environment atoms common to all systems and x_{si} is the group of atoms unique to mutant i at site s, and $V_{\mathrm{Bias}}(\{\lambda\})$ is a multidimensional biasing potential to facilitate sampling in λ space.

The previous set of biasing potentials

$$V_{\mathrm{Fixed}} = \sum_{s}^{M}\sum_{i}^{N_s} \phi_{si}\lambda_{si} \tag{S2}$$

$$V_{\mathrm{Quad}} = \sum_{s}^{M}\sum_{i}^{N_s}\sum_{j>i}^{N_s} \psi_{si,sj}\lambda_{si}\lambda_{sj} \tag{S3}$$

$$V_{\mathrm{End}} = \sum_{s}^{M}\sum_{i}^{N_s}\sum_{j\neq i}^{N_s} \omega_{si,sj}\lambda_{sj}\lambda_{si}/(\lambda_{si}+\alpha), \tag{S4}$$

with $\alpha = 0.017$, were initially optimized to flatten alchemical free energy landscapes observed with hard-core interactions. An additional biasing potential of the form

$$V_{\mathrm{Skew}} = \sum_{s}^{M}\sum_{i}^{N_s}\sum_{j\neq i}^{N_s} \chi_{si,sj}\lambda_{sj}(1-\exp(-\lambda_{si}/\sigma)), \tag{S5}$$

with $\sigma = 0.18$, was found to give improved fits to free energy profiles obtained with soft-core interactions. The previous flattening algorithm was effective in optimizing parameters in Equations S2-S4, but had been carefully designed to handle a variety of special cases on an individual basis.

Consequently, it was not easily modified to accommodate additional biasing potentials such as Equation S5.

In the present work a more generalizable WHAM-based least-squares approach was used to learn biasing potential parameters to flatten the landscape. Flattening consists of several dozen rounds of very short simulations with iterative updates to the biasing potential parameters ϕ_{si}, $\psi_{si,sj}$, $\chi_{si,sj}$, and $\omega_{si,sj}$. During parameter optimization, recent trajectories are combined using WHAM,[1] and 1D free energy profiles are computed along λ_{si} as well as $\lambda_{si}/(\lambda_{si}+\lambda_{sj})$ such that $(\lambda_{si}+\lambda_{sj})>0.8$ as before, but with 400 bins from 0 to 1. An additional 2D profile on λ_{si} and λ_{sj} with 20 bins in each direction is also computed. The entropy of the implicit constraints is subtracted from the free energy profiles as before. A scoring function for proposed changes in parameters at each site s is constructed as

$$E=\sum_{p}^{\text{Profiles}}\sum_{b}^{\text{Bins}}\frac{w_{pb}}{2}(G_{pb}+\Delta G_{pb}-\bar{G}_p)^2+\sum_{i}^{\text{Biases}}\frac{k}{2{\gamma_0}^2}\Delta{\gamma_i}^2 \tag{S6}$$

$$\Delta G_{pb}=\sum_{i}^{\text{Biases}}\frac{\partial G_{pb}}{\partial\gamma_i}\Delta\gamma_i\,, \tag{S7}$$

and the scoring function is minimized at each site s independently. Here, γ_i runs over all biasing potential parameters in the set $\{\phi_{si},\psi_{si,sj},\chi_{si,sj},\omega_{si,sj}\}$. Regularization parameters are $k=1(\text{ kcal/mol})^2$, with γ_0 set to $\phi_0=2$ kcal/mol, $\psi_0=8$ kcal/mol, $\chi_0=2$ kcal/mol, $\omega_0=1$ kcal/mol depending on the type of parameter. For 1D λ_{si} profiles, the weight $w_{pb}=100$ for the last bin or $w_{pb}=1$ otherwise, to discourage sharp charges in the free energy near the $\lambda_{si}=1$ endpoint which subsequently result in undersampling or oversampling (trapping) at that endpoint. The weight $w_{pb}=2/(N_s-1)$ for all bins b of profiles p so that the total weight of all $N_s(N_s-1)/2$ transition profiles is constant relative to the weight of the N_s different λ_{si} profiles. The free energy terms G_{pb} and $\partial G_{pb}/\partial\gamma_i$ are calculated with WHAM from the states sampled during recent simulations. The first term of Equation S6 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 regularization term prevents overfitting between subsequent flattening runs. The scoring function in

Equation S6 is minimized with $\bar{G}_{pb}$ and $\Delta\gamma_i$ as free parameters using least squares. If any parameter at site s changes more than 1.5 times its respective limit of ϕ_0, ψ_0, χ_0, or ω_0, the changes in all parameters are scaled uniformly such that the largest change is 1.5 times ϕ_0, ψ_0, χ_0, or ω_0. This flattening algorithm can be easily generalized to other biasing potentials as needed, and converges comparably with the old algorithm using the more limited set of biasing potentials.

In the previous version of ALF,[2] an initial estimation of the fixed biases was performed, but was omitted in the present work because it added complexity without saving much time. As before, 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 followed by 1 ns of sampling. Previously, optimization for 100 ps and 1 ns runs included sampling from the last 10 and 6 cycles respectively in WHAM-based estimates of free energy profiles, but it was noted that convergence was more rapid using only the last 5 cycles as poorly sampled simulations could not misguide parameter estimation for as many cycles. 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 of ns equilibration followed by 5 ns of sampling combining data from the last 3 cycles was added to the protocol for all sites. For production, 5 independent simulations were run for 40 ns each.

It was observed that often some substituents stopped being sampled after the system had relaxed, therefore the first quarter of all production runs was discarded for 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 Table S2 for details.) A physical state of the system is achieved when one λ per site is equal to 1, and all the others are 0. The free energy of the physical states can be estimated by binning states for which all sites have one λ greater than $\lambda_c = 0.99$, Boltzmann inverting the populations, and subtracting the biasing potentials.[2] Results from different independent trials and replicas of the final production run were combined using WHAM,[1] free energies were computed from endpoint populations with a λ cutoff

of 0.99,[2,3] and uncertainties were computed from bootstrap analysis.[2]

Table S1: Optimal VB-REX Parameters

Sites	B[a]	N Replicas	Flat Replica[b]
1	1.6	3	2
3	1.0	4	2
4	0.8	4	2
5	0.6	6	3

[a] in (kcal/mol)
[b] The (one indexed) replica whose energy landscape is flattened

Inspired by BP-REX,[4] 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 (Table S1), where B is tuned to optimize exchange rates, while the number of replicas and the replica targeted by ALF are tuned to sample primarily physical endpoint states in the final replica. VB-REX simulations used it during both sampling and production, and production simulations were only sampled for 20 ns due to faster convergence. For the 5 mutation site system, even the small 1.8 kcal/mol barrier in the sixth replica relative to the third replica was adequate to increase the sampling of approximately physical states ($\lambda > 0.99$ at all sites) by about two orders of magnitude. Optimal VB-REX parameters elsewhere will depend on the number of sites and mutations.

Three different electrostatic models were considered in this work: 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. FSHIFT subtracts a constant from the electrostatic force so that it is zero at the cutoff while FSWITCH multiplies the electrostatic force by a function that decays smoothly from one to zero within a narrow range. The electrostatic potential is then obtained by integrating the electrostatic force. See equations 6 and 9 in Reference 5. FSWITCH gives a smaller perturbation to electrostatics than FSHIFT, but both neglect long range electrostatic effects, which are accounted for with PME.

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.[6,7] Using PME in MSλD simulations is not trivial, and we follow the approach of Jana Shen:[8] the short range piece is scaled according to Equation S1, while in the long range piece, the charges of an atom in substituent *si* are scaled by λ_{si}, which means the interaction between two atoms is scaled by the product of their λ's, even if the atoms are in the same site or even the same substituent, (which would normally result in a scaling of 0 or λ, respectively). PME exclusion terms were scaled by λ_{si} for exclusions between a side chain and the environment, $\lambda_{si}\lambda_{tj}$ for exclusions between sites, 0 for exclusions between different side chains at the same site, and λ_{si}^2 for exclusions within the same side chain. While this treatment alters the alchemical pathway between endpoints, the energy of the endpoints themselves is not affected, so the results will be unaffected. Alchemical simulations involving a net change in charge can result in serious artifacts.[9–11] While corrections for these effects have been devised for PME,[12] they were not applied in the present work since only two of the considered mutants involved a charge change. Application of these corrections is deferred until it can be tested more systematically.

For FSWITCH simulations, the switching radius was 10 Å, the cutoff radius was 12 Å, and the neighbor list radius was 15 Å. For PME simulations, the grid spacing was 1 Å, the interpolation order was 6, κ was 0.32 $Å^{-1}$, the switching radius (for van der Waals interactions) was 9 Å, the cutoff radius was 10 Å, and the neighbor list radius was 12 Å.

Convergence of Multisite Results

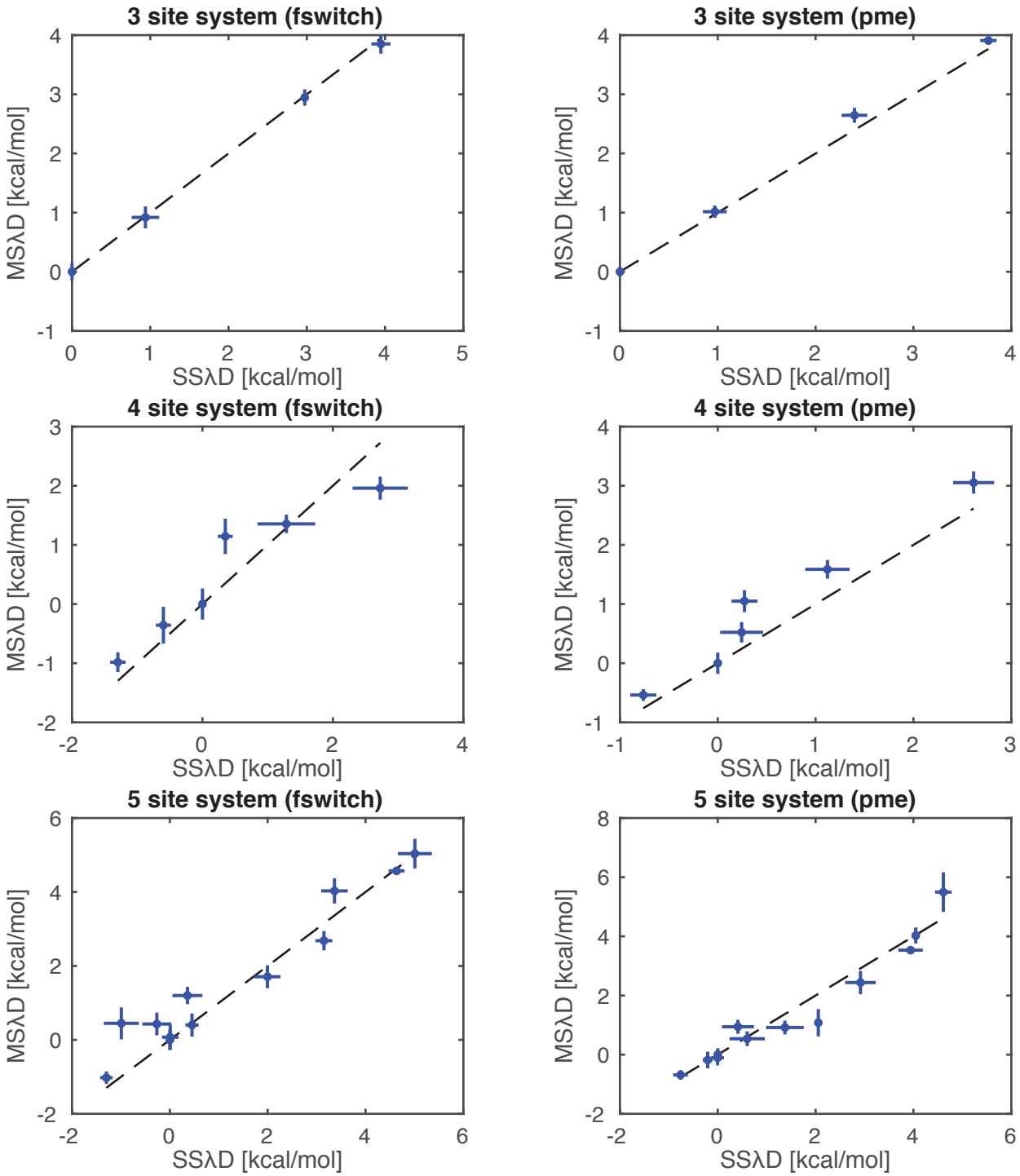

Figure S1: Comparison of SSλD and MSλD results in the multisite systems. Dashed lines denote $y = x$. Agreement in the 3 site system is good, suggesting sampling is well converged. Agreement is poorer in the 4 site and 5 site systems, suggesting sampling or convergence issues in these systems. RMS differences are roughly 0.5 kcal/mol or less, which is substantially smaller than the deviations from experiment, suggesting that sampling is a small but non-negligible source of error.

Flattening with Production Runs

Sometimes landscape flattening parameters derived from 5 ns simulations did not allow effective sampling during the last three quarters of the production simulations. This occurs either because insufficient sampling was obtained during the 5 ns flattening runs, possibly due to trapping in one or more iterations, or because the system is not fully relaxed by the end of 5 ns. This was only observed in the protein side of the thermodynamic cycle ($\Delta G_{\text{Folded}}(S_1 \rightarrow S_2)$); the pentapeptide mimic of the unfolded ensemble converged rapidly and flattened efficiently (and did not require VB-REX for the A42, A98, and M102).

There were fewer slowly converging sites in the multisite systems. The 3 site and 4 site systems converged well with their initial biases. The 5 site system failed to sample a few sequences due to the large sequence space and possible cooperative effects between mutations, but reoptimizing the biases seemed unlikely to improve matters, so production was not used to reoptimize the biases for any of the multisite systems.

When rerunning production several times due to poor sampling, it is possible to artificially improve results by stopping after a production run that agrees well with experiment, but this inflates agreement and is not possible in a prospective study. To avoid this pitfall, only $\Delta G_{\text{Folded}}(S_1 \rightarrow S_2)$ and its statistical uncertainty were computed after a particular round of production. $\Delta\Delta G_{\text{Folding}}(S_1 \rightarrow S_2)$ was not computed, nor was it compared to experiment until after the biases were deemed adequately converged.

Four measures of convergence were used for the biasing potential parameters. The statistical uncertainty of $\Delta G_{\text{Folded}}(S_1 \rightarrow S_2)$ was computed from the simulation. If the statistical uncertainty was above 0.5 kcal/mol, this indicated poor convergence. Since the results are especially sensitive to the free energy of the reference sequence (the native sequence), the biases were judged unconverged if the uncertainty of the reference sequence was above 0.3 kcal/mol. For the small number of independent trials (5) is is possible to get serendipitously obtain roughly the same free energy results from all trials even when the biases are poor. Therefore the changes in the fixed biases $\Delta\phi_{si}$ based on the current production run were also compared. If the range between the maximum

and minimum $\Delta\phi_{si}$ was greater than roughly $3kT$, this indicates the biases are unconverged, as the ratio in sampling will be roughly 1:20 for this difference. Furthermore, the lower the minimum of $\Delta\phi_{si}$, ($\Delta\phi_{s0}$ is always subtracted from $\Delta\phi_{si}$ so that the fixed bias of the reference sequence ϕ_{s0} remains 0), the more difficult it will be to sample the reference sequence, so we sought to keep this above $-2kT$. With these limits in mind, some sites are more difficult to sample, and one must concede the sampling is not going to improve upon further bias optimization, while other sites sample very easily, and one more round of production could substantially improve statistical uncertainty even though the criteria are met, therefore bias optimization still requires some human oversight. Reasoning for decisions to reoptimize biases is provided in the notes column of Table S2. All production runs started on iteration 64, after 50 rounds of 100 ps flattening, 10 rounds of 1 ns flattening, and 3 rounds of 5 ns flattening.

Table S2: Bias Reoptimization with Production Runs

FSWITCH						
System	Iteration	Ref Unc[a]	Max Unc[a]	Min $\Delta\phi_{si}$	Range $\Delta\phi_{si}$	Notes
V149	64	0.223	0.223	-0.78	1.44	Converged
M102	64	0.113	0.227	-1.83	1.83	Min $\Delta\phi_{si}$ bad
M102	65	0.222	0.241	-0.67	0.92	Converged
L99	64	0.130	0.170	-1.02	1.46	Converged
A98	64	0.571	0.712	-2.76	2.76	All bad
A98	65	0.229	0.734	-0.69	2.68	Reference sequence is better
A98	66	0.436	0.564	-1.62	2.65	All bad again
A98	67	0.186	0.465	-0.31	1.89	Converged: Range $\Delta\phi_{si}$ bad
A42	64	0.085	0.598	0.00	1.95	Range $\Delta\phi_{si}$ bad
A42	65	0.155	0.220	0.00	0.69	Converged
M106	64	0.047	0.047	-0.75	0.75	Converged
F153	64	0.166	0.271	-0.25	0.31	Converged
PME						
System	Iteration	Ref Unc[a]	Max Unc[a]	Min $\Delta\phi_{si}$	Range $\Delta\phi_{si}$	Notes
V149	64	0.164	0.521	-1.23	1.44	Min $\Delta\phi_{si}$ and Max unc bad
V149	65	0.146	0.512	-0.43	0.90	Converged: Max unc bad
M102	64	nan	nan	-2.66	2.66	All bad
M102	65	nan	nan	-1.33	1.93	All bad
M102	66	0.877	1.083	-2.09	2.23	Improving
M102	67	0.273	0.336	-0.85	0.85	Converged
L99	64	0.061	0.195	-0.99	0.99	Converged
A98	64	0.165	0.315	-0.81	1.98	Range $\Delta\phi_{si}$ bad
A98	65	0.653	0.670	-0.95	2.84	Range $\Delta\phi_{si}$ worse
A98	66	0.328	0.372	-0.93	1.72	Converged: Ref unc bad
A42	64	0.369	0.369	-1.02	1.54	Ref unc bad
A42	65	0.323	0.323	-1.05	1.05	Ref unc bad
A42	66	0.283	0.283	-0.44	0.58	Converged
M106	64	0.076	0.106	-0.23	0.23	Converged
F153	64	0.126	0.355	-0.23	0.81	Max unc could improve
F153	65	0.123	0.273	-0.62	0.68	Converged

[a]Unc is uncertainty

Computational Efficiency

A direct estimation of the relative computational efficiencies of MSλD and FEP is difficult, either in terms simulation time or CPU time. The central issue is that one can run either method for more or less time, so a fair comparison requires simulations that obtain the same level of statistical precision on the same systems. Given the substantially lower efficiency of FEP, running the simulations required for such a comparison is undesirable due to computational expense. Indeed, in another large scale MSλD study, FEP controls on a small fraction of the systems studied with MSλD required far more computing resources than the entire MSλD study.[13] That study concluded MSλD was roughly 20-30 times more efficient in simulation time than FEP.

As a large scale study of FEP folding free energies, Reference 14 is the closest comparison available, but direct comparison is not possible for two reasons. First, that study only cites their accuracy relative to experiment, and makes no estimate of precision. Second, that study examined mostly large to small mutations which converge more rapidly than the small to large mutations in the present study. A related study, Reference 15 found that obtaining accurate predictions for small to large mutations with tryptophan required 100 ns per window, or 20 times more sampling than Reference 14 used. With these caveats, Reference 14 took 60 ns of simulation time (12 windows of 5 ns each) for neutral mutations and 80 ns of simulation time (16 windows of 5 ns each) for charge changing mutations, requiring an average of 65 ns per mutation.

In the present study, we made little attempt to optimize computational efficiency by minimizing the amount of time spent in landscape flattening. Instead of trying to obtain comparable precision with FEP in less time, we used a comparable amount of time to standard FEP studies and obtained more accurate results with higher precision. We traded away the computational advantage of MSλD to run longer simulations, which improved the precision and accuracy of our results, and to run multiple independent trials, which enabled us to more robustly estimate statistical precision. A set of 5 independent trials without VB-REX takes a total of 200 ns (40 ns each), and with VB-REX takes 300 ns (20 ns each times three replicas). If these production runs were poorly sampled, they could only be used to further optimize the biases, resulting in a large waste of simulation time

that could have been saved with a more careful procedure. The total simulation time including flattening for the single site mutants was 117 ns per mutation (3746 ns total) for FSWITCH simulations and 148 ns per mutation (4746 ns total) for PME simulations. Without flattening, both methods took 53 ns per mutation (1700 ns total). It is worth noting that these times would be divided by five if we had run no independent trials (Reference 14 ran no independent trials), that the majority of the computational expense was focused on the three sites that were more difficult to converge: A42, A98, and M102, and that the multisite simulations ran for substantially less time per mutation. It should be emphasized that while Reference 14 and the present study used comparable amounts of simulation time per mutation, this is not an indication of comparable computational efficiency, as the present study obtained high accuracy results with high precision, while Reference 14 obtained lower accuracy results of unknown precision.

Finally, the conversion from simulation time to CPU time is rather meaningless. Both FEP and MSλD add very little computational overhead to standard MD simulations, so comparisons are dominated by the efficiency of the molecular dynamics engine and the hardware used. Reference 14 simulations were run on a GTX780 GPU cluster while the present simulations were run on a faster GTX980 cluster. The present simulations were run with the CHARMM molecular dynamics engine, which has an efficient GPU implementation, but is still only half as fast as OpenMM. Simulations on our cluster with our code require 1.3 hours per ns on a single GPU for a 72 Å box in the folded ensemble, but an industrial application of MSλD would ideally involve more efficient code that effectively scales to multiple GPUs and runs on the most recent hardware.

Structural Analysis

In order to quantify structural relaxation within our simulations, crystal structures for each of the 7 sites [16–24] were visually inspected to identify reaction coordinates that correlate with mutation. Reaction coordinates were chosen from crystal structures rather than from the trajectories themselves because thermal vibrations tend to obscure slower motions. If the reaction coordinates identified in crystals are subsequently seen to correlate with λ in our simulations, then the autocorrelation time in these reaction coordinates indicates the speed of structural relaxations that are necessary to obtain converged estimates of free energy.

Site A42 is missing crystal structures for three of the six mutations considered, suggesting structural disruption by some of the more aggressive mutations may be larger than what is observed in the remaining three structures. Even the rather conservative mutation A42V displaces residues S36-S44 rather substantially by wedging more steric bulk into the interface between these residues and the rest of the protein. Because this motion is rather large and collective, two distances were chosen as possible reaction coordinates: A42C_α-T34C_α and A42C_α-Y25C_α.

Site A98 was also missing structures for three of the mutants and exhibited large structural variations. A98M showed especially large RMSD (1.10 Å) because of a change in the opening of the cleft between the N terminal and C terminal lobes. Fitting only the C terminal lobe (residues 1-10 and 70-162) gave a somewhat lower RMSD of 0.58 Å. Increasing the size of the mutated side chain seemed to force the C_α atom of A98 further out of the cavity the side chain occupied; this was well captured by the A98C_α-T152C_α distance.

Site AL99 showed surprisingly little structural variation. The largest and smallest mutants, L99A and L99F showed slight changes that would be expected for protein collapse around a smaller side chain or protein displacement to to a larger side chain. For L99A, the side chain of Y88 moves closer by 0.2 Å. For L99F, V103 and V111 move closer by about 0.3 Å, and the side chain of L84 moves away by 0.5 Å. Though nothing moved very much, the distance L99C_α-L84C_γ was selected for giving the largest variation.

Although there are several moving parts in the M102 mutations, it is surprising how little

structural disruption there is for the buried charge in M102K. M102 is pulled more deeply into the protein for both M102A and M102K, and F153 moves to meet it, so M102C_α-F153C_ζ was chosen as one of the distances. While M102K does not become solvent exposed, it results in the flip of several dihedrals in M106. These result in substantial changes to the M102C_α-M106C_ε distance, so this was chosen as the second reaction coordinate. The net effect of these dihedral flips in M106 is to reposition M106S_δ directly against the amine of M102K. While M102K is still buried, this interaction with a soft, polarizable sulfur atom may partially compensate for charge burial. The lack of polarization in our forcefield as well as the generally poor treatment of S and P atoms in fixed charge forcefields may partially account for our very large error in M102K.

M106 causes motion of M102C_α by up to 0.7 Å, however this motion is perpendicular to the distance between M106 and M102, and thus poorly captured by a distance. M106L pulls the M106C_α 0.6 Å deeper into the protein, so M106C_α-W138C_α was chosen as the reaction coordinate.

V149 mutations are relatively simple. Smaller side chains pull the C_α atom into the pocket formerly occupied by the side chain and larger side chains push it out. The V149C_α-N101C_α distance captures this effect well.

F153 is similarly simple. Mutation away from the native F cause decreases in steric bulk, and all pull the C_α atom into the vacated pocket by about 0.6 Å. F153C_α-L99C_α was chosen to measure this effect.

The mean distance in the simulations for a particular mutation was calculated by averaging over all frames for which λ of that mutation was greater than 0.99. Overall, the mean distance observed in simulations agreed rather closely with the distance observed in the crystal structure (Table S3). The distances were typically slightly larger in simulations by 0.1 to 0.2 Å due to thermal vibrations. S and T mutants typically occupied more volume than expected from crystal structures, resulting in 0.5 Åincreases in the relevant distance. All chosen distances exhibited correlation with λ for at least some of the side chains during the simulations.

It is noteworthy that three of the four sites displaying slower autocorrelation decay (A42, A98,

M102, and M153) were sites identified as requiring enhanced sampling with VB-REX. Furthermore, while VB-REX should accelerate structural relaxation in these systems, their relaxation was still slower than the remaining systems. The autocorrelation times further underline the fact that while VB-REX was likely sufficient for A42 and M102, more aggressive sampling techniques are needed to sample A98 well, due to its very slow autocorrelation decay.

Table S3: Single Site Structural Analysis

AA	XTAL	XTAL BB RMSD	XTAL Distance	Mean Sim. Dist.	Corr. (λ & Dist.)
A42: (A42C$_\alpha$-T34C$_\alpha$ / A42C$_\alpha$-Y25C$_\alpha$)					
A	1l63	0.00	5.91 / 6.22	5.93 / 6.77	-0.16 / -0.22
F	—-	—-	—- / —-	6.22 / 7.02	-0.11 / -0.14
I	—-	—-	—- / —-	6.63 / 7.71	0.20 / 0.25
L	—-	—-	—- / —-	6.42 / 7.79	0.06 / 0.28
S	206l	0.12	5.85 / 6.43	6.11 / 6.85	-0.07 / -0.15
V	1qtb	0.24	6.46 / 7.00	6.48 / 7.30	0.17 / 0.12
A98: (A98C$_\alpha$-T152C$_\alpha$)					
A	1l63	0.00	6.32	6.52	-0.08
C	1qsb	0.23	7.10	7.10	0.06
F	—-	—-	—-	8.57	0.18
I	—-	—-	—-	7.48	0.03
L	1qs5	0.32	7.70	7.99	0.20
M	1qth	1.10	7.46	7.49	0.05
S	125l	0.11	6.51	7.05	-0.34
V	1qs9	0.17	7.16	7.31	-0.06
W	—-	—-	—-	8.36	0.15
L99: (L99C$_\alpha$-L84C$_\gamma$)					
L	1l63	0.00	8.25	8.36	-0.13
A	1l90	0.10	8.30	8.50	-0.06
F	1l91	0.14	8.94	9.07	0.54
I	1l92	0.10	8.29	8.06	-0.31
M	1l93	0.10	8.52	8.67	-0.02
V	1l94	0.12	8.07	8.12	-0.29
M102: (M102C$_\alpha$-F153C$_\zeta$ / M102C$_\alpha$-M106C$_\varepsilon$)					
M	1l63	0.00	6.76 / 4.22	6.98 / 5.11	0.24 / -0.36
A	222l	0.18	6.24 / 5.01	6.40 / 4.76	-0.15 / -0.43
K	1l54	0.15	6.37 / 7.14	6.67 / 7.51	-0.08 / 0.64
M106: (M106C$_\alpha$-W138C$_\alpha$)					
M	1l63	0.00	8.71	8.79	0.14
K	231l	0.20	8.75	8.39	-0.17
L	234l	0.14	8.38	8.61	0.03
V149: (V149C$_\alpha$-N101C$_\alpha$)					
V	1l63	0.00	7.64	7.78	-0.19
C	1g07	0.12	7.66	7.95	-0.05
I	1g0q	0.16	7.90	7.95	-0.06
M	1cv6	0.15	8.19	8.35	0.31
S	1g06	0.15	7.46	8.00	0.02
T	126l	0.12	7.51	8.03	0.03
F153: (F153C$_\alpha$-L99C$_\alpha$)					
F	1l63	0.00	8.45	8.62	0.40
A	1l85	0.22	7.77	7.46	-0.36
I	1l86	0.25	8.06	8.09	0.13
L	1l87	0.15	8.15	7.50	-0.22
M	1l88	0.17	7.94	7.97	0.03
V	1l95	0.28	8.01	8.14	0.13

Individual Results

Table S4: Single Site Folding Free Energies

AA	Experiment	FSWITCH	PME
		A42	
A	0.00	0.00 ± 0.17	0.00 ± 0.28
F	3.60	5.12 ± 0.23	4.00 ± 0.13
I	3.10	6.26 ± 0.22	5.40 ± 0.23
L	3.40	6.71 ± 0.22	5.70 ± 0.25
S	2.30	3.74 ± 0.10	3.73 ± 0.13
V	2.70	4.70 ± 0.05	4.78 ± 0.10
		A98	
A	0.00	0.00 ± 0.19	0.00 ± 0.33
C	1.00	1.40 ± 0.17	2.51 ± 0.35
F	5.90	6.19 ± 0.37	6.66 ± 0.24
I	4.90	5.37 ± 0.20	4.31 ± 0.32
L	4.30	4.90 ± 0.25	5.01 ± 0.17
M	3.20	2.51 ± 0.46	2.10 ± 0.21
S	2.50	2.36 ± 0.22	2.34 ± 0.37
V	3.20	4.09 ± 0.11	3.20 ± 0.37
W	5.70	8.82 ± 0.26	7.04 ± 0.28
		L99	
L	0.00	0.00 ± 0.14	0.00 ± 0.06
A	5.00	6.97 ± 0.08	6.46 ± 0.09
F	0.40	0.20 ± 0.17	-0.00 ± 0.11
I	1.40	3.21 ± 0.17	2.88 ± 0.20
M	0.70	1.16 ± 0.15	0.89 ± 0.08
V	2.30	4.44 ± 0.17	3.95 ± 0.17
		M102	
M	0.00	0.00 ± 0.21	0.00 ± 0.26
A	2.90	3.11 ± 0.23	3.41 ± 0.33
K	6.90	10.85 ± 0.12	13.49 ± 0.12
		M106	
M	0.00	0.00 ± 0.06	0.00 ± 0.08
K	3.40	2.98 ± 0.04	4.09 ± 0.11
L	-0.50	-0.31 ± 0.03	-0.32 ± 0.04
		V149	
V	0.00	0.00 ± 0.23	0.00 ± 0.15
C	2.20	1.89 ± 0.09	1.27 ± 0.17
I	0.10	0.73 ± 0.20	-0.47 ± 0.12
M	2.80	3.03 ± 0.18	1.48 ± 0.16
S	4.40	6.31 ± 0.25	4.82 ± 0.51
T	2.80	3.46 ± 0.13	2.70 ± 0.32
		F153	
F	0.00	0.00 ± 0.17	0.00 ± 0.13
A	3.50	3.64 ± 0.28	3.80 ± 0.27
I	0.50	-0.03 ± 0.11	0.42 ± 0.04
L	-0.20	-1.68 ± 0.18	-1.26 ± 0.20
M	0.80	-1.79 ± 0.15	-1.35 ± 0.09
V	1.80	1.71 ± 0.14	2.24 ± 0.05

Table S5: 3 Site System Folding Free Energies

AA	Experiment	FSWITCH	PME
IIL	0.00	0.00 ± 0.12	0.00 ± 0.05
IIM	2.00	2.94 ± 0.11	2.65 ± 0.10
IML	3.10	3.85 ± 0.14	3.91 ± 0.04
IMM	3.05	6.21 ± 0.12	5.97 ± 0.07
MIL	2.20	0.92 ± 0.16	1.01 ± 0.08
MIM		2.41 ± 0.14	2.36 ± 0.12
MML		4.53 ± 0.16	4.40 ± 0.09
MMM	3.30	5.45 ± 0.15	5.06 ± 0.12

Table S6: 4 Site System Folding Free Energies

AA	Experiment	FSWITCH	PME
LMVF	0.00	0.00 ± 0.23	0.00 ± 0.15
LMVL	-0.20	-0.98 ± 0.14	-0.54 ± 0.07
LMFF	1.43	1.96 ± 0.17	3.05 ± 0.16
LMFL	1.09	0.39 ± 0.16	1.99 ± 0.15
LMIF	0.69	1.35 ± 0.13	1.59 ± 0.13
LMIL		0.06 ± 0.15	0.80 ± 0.06
LLVF	0.74	-0.36 ± 0.28	0.52 ± 0.14
LLVL		-0.93 ± 0.20	0.22 ± 0.17
LLFF	1.51	1.95 ± 0.36	3.17 ± 0.51
LLFL		0.62 ± 0.18	1.73 ± 0.30
LLIF		1.04 ± 0.15	1.63 ± 0.23
LLIL		-0.01 ± 0.12	0.85 ± 0.15
FMVF	0.40	1.14 ± 0.27	1.05 ± 0.16
FMVL	-0.03	-1.77 ± 0.17	-0.70 ± 0.22
FMFF		1.41 ± 0.16	2.45 ± 0.15
FMFL		-0.74 ± 0.15	0.45 ± 0.27
FMIF	0.82	1.54 ± 0.10	1.94 ± 0.28
FMIL		-0.95 ± 0.12	0.33 ± 0.29
FLVF	0.71	0.04 ± 0.18	0.80 ± 0.14
FLVL	0.21	-1.96 ± 0.28	-0.88 ± 0.25
FLFF		1.73 ± 0.39	3.12 ± 0.09
FLFL		-0.91 ± 0.08	0.25 ± 0.38
FLIF	1.15	0.61 ± 0.17	1.77 ± 0.21
FLIL	0.54	-1.40 ± 0.19	-0.45 ± 0.48

Table S7: 5 Site System Folding Free Energies - Part 1

AA	Experiment	FSWITCH	PME
LALVF	0.00	0.00 ± 0.23	0.00 ± 0.16
LALVL	-0.20	-1.02 ± 0.12	-0.69 ± 0.11
LALIF		0.40 ± 0.27	-0.18 ± 0.23
LALIL		-0.45 ± 0.14	-0.79 ± 0.13
LAAVF		5.04 ± 0.36	5.50 ± 0.61
LAAVL		3.72 ± 0.17	3.44 ± 0.35
LAAIF		5.14 ± 0.70	3.96 ± 0.44
LAAIL		3.86 ± 0.21	3.07 ± 0.31
LAMVF		0.07 ± 0.29	-0.11 ± 0.20
LAMVL		-0.90 ± 0.18	-0.07 ± 0.14
LAMIF		0.32 ± 0.18	-0.29 ± 0.17
LAMIL		-0.44 ± 0.14	-0.36 ± 0.10
LAVVF		$4.57 \pm nan^a$	$3.53 \pm nan^a$
LAVVL		2.37 ± 0.16	2.58 ± 0.17
LAVIF		5.41 ± 0.34	3.17 ± 0.12
LAVIL		2.90 ± 0.17	2.44 ± 0.33
LLLVF		0.45 ± 0.39	0.94 ± 0.18
LLLVL		0.55 ± 0.13	0.81 ± 0.26
LLLIF		0.97 ± 0.20	0.43 ± 0.17
LLLIL		1.15 ± 0.15	0.92 ± 0.27
LLAVF		4.91 ± 0.29	$5.79 \pm nan^a$
LLAVL		5.31 ± 0.40	4.77 ± 0.36
LLAIF		$5.16 \pm nan^a$	4.81 ± 0.31
LLAIL		$4.39 \pm nan^a$	4.24 ± 0.42
LLMVF		1.08 ± 0.46	0.14 ± 0.32
LLMVL		0.37 ± 0.26	0.89 ± 0.14
LLMIF		0.33 ± 0.22	0.37 ± 0.23
LLMIL		0.44 ± 0.23	0.98 ± 0.22
LLVVF		3.75 ± 0.69	3.89 ± 0.30
LLVVL		3.83 ± 0.21	$3.77 \pm nan^a$
LLVIF		3.23 ± 0.43	2.69 ± 0.18
LLVIL		4.02 ± 0.48	2.81 ± 0.30
LMLVF		1.20 ± 0.19	0.92 ± 0.18
LMLVL		-0.09 ± 0.16	-0.19 ± 0.23
LMLIF		1.32 ± 0.14	0.75 ± 0.18
LMLIL		0.40 ± 0.18	-0.20 ± 0.20
LMAVF		3.95 ± 0.39	2.97 ± 0.46
LMAVL		3.69 ± 0.23	3.22 ± 0.17
LMAIF		4.51 ± 0.47	3.01 ± 0.19
LMAIL		4.11 ± 0.22	3.29 ± 0.21
LMMVF		0.79 ± 0.27	0.39 ± 0.21
LMMVL		-0.84 ± 0.13	-0.08 ± 0.17
LMMIF		0.73 ± 0.17	0.49 ± 0.19
LMMIL		-0.55 ± 0.21	-0.35 ± 0.21
LMVVF		3.53 ± 0.20	2.61 ± 0.25
LMVVL		3.00 ± 0.20	2.54 ± 0.23
LMVIF		3.29 ± 0.43	2.45 ± 0.26
LMVIL		2.99 ± 0.17	2.39 ± 0.35
LVLVF		0.43 ± 0.26	0.54 ± 0.19
LVLVL		0.01 ± 0.14	0.12 ± 0.08
LVLIF		1.10 ± 0.22	0.07 ± 0.21
LVLIL		0.47 ± 0.10	0.30 ± 0.07
LVAVF		$8.04 \pm nan^a$	5.16 ± 0.58
LVAVL		4.87 ± 0.28	4.21 ± 0.19
LVAIF		$inf^b \pm nan^a$	3.97 ± 0.48
LVAIL		4.87 ± 0.72	4.21 ± 0.52
LVMVF		0.26 ± 0.35	0.45 ± 0.30
LVMVL		-0.47 ± 0.15	0.12 ± 0.14
LVMIF		1.06 ± 0.20	0.62 ± 0.23
LVMIL		0.02 ± 0.13	0.34 ± 0.10
LVVVF		5.60 ± 0.38	4.06 ± 0.70
LVVVL		3.03 ± 0.19	2.93 ± 0.18
LVVIF		$5.28 \pm nan^a$	3.26 ± 0.18
LVVIL		3.69 ± 0.28	3.19 ± 0.18
LWLVF	2.20	4.03 ± 0.30	4.03 ± 0.21
LWLVL		4.42 ± 0.58	3.27 ± 0.26
LWLIF		4.38 ± 0.58	2.84 ± 0.53
LWLIL		6.09 ± 0.83	3.34 ± 0.41
LWAVF		4.09 ± 0.27	3.92 ± 0.24
LWAVL		4.20 ± 0.25	4.11 ± 0.26
LWAIF		4.41 ± 0.20	3.72 ± 0.10
LWAIL		4.41 ± 0.28	3.90 ± 0.14
LWMVF		2.07 ± 0.16	1.47 ± 0.22
LWMVL		1.95 ± 0.11	2.06 ± 0.15
LWMIF		1.88 ± 0.33	1.30 ± 0.15
LWMIL		1.99 ± 0.16	1.54 ± 0.33
LWVVF		4.73 ± 0.15	3.64 ± 0.18
LWVVL		4.67 ± 0.15	3.97 ± 0.22
LWVIF		4.38 ± 0.24	3.22 ± 0.11
LWVIL		4.90 ± 0.18	3.59 ± 0.14

[a] Insufficient sampling to estimate uncertainty

[b] Sequence not sampled

Table S8: 5 Site System Folding Free Energies - Part 2

AA	Experiment	FSWITCH	PME
AALVF	2.70	1.71 ± 0.27	1.08 ± 0.41
AALVL		2.54 ± 0.19	2.35 ± 0.19
AALIF		2.48 ± 0.19	1.60 ± 0.47
AALIL		3.05 ± 0.18	2.35 ± 0.24
AAAVF		$6.08 \pm nan^a$	$4.22 \pm nan^a$
AAAVL		$7.54 \pm nan^a$	4.94 ± 0.45
AAAIF		$inf^b \pm nan^a$	$4.41 \pm nan^a$
AAAIL		$inf^b \pm nan^a$	5.14 ± 0.81
AAMVF		1.76 ± 0.56	1.47 ± 0.50
AAMVL		3.65 ± 0.22	2.85 ± 0.47
AAMIF		3.14 ± 0.37	1.56 ± 0.64
AAMIL		3.13 ± 0.65	3.12 ± 0.22
AAVVF		4.34 ± 0.24	3.83 ± 0.69
AAVVL		5.53 ± 0.47	$4.83 \pm nan^a$
AAVIF		$5.98 \pm nan^a$	3.73 ± 0.78
AAVIL		6.59 ± 0.62	$4.46 \pm nan^a$
ALLVF		0.65 ± 0.20	1.57 ± 0.18
ALLVL		0.36 ± 0.13	0.93 ± 0.13
ALLIF		1.72 ± 0.21	1.52 ± 0.20
ALLIL		0.89 ± 0.17	1.06 ± 0.20
ALAVF		6.11 ± 0.57	$7.08 \pm nan^a$
ALAVL		4.72 ± 0.51	4.51 ± 0.23
ALAIF		$inf^b \pm nan^a$	5.35 ± 0.17
ALAIL		5.09 ± 0.64	4.34 ± 0.23
ALMVF		0.71 ± 0.23	1.58 ± 0.21
ALMVL		0.09 ± 0.21	1.33 ± 0.15
ALMIF		1.61 ± 0.31	1.62 ± 0.27
ALMIL		0.77 ± 0.41	1.34 ± 0.12
ALVVF		4.95 ± 0.44	5.68 ± 0.42
ALVVL		3.33 ± 0.10	4.07 ± 0.31
ALVIF		5.17 ± 0.22	4.01 ± 0.15
ALVIL		3.81 ± 0.50	3.96 ± 0.28
AMLVF		1.38 ± 0.22	1.58 ± 0.20
AMLVL	1.10	1.26 ± 0.16	1.37 ± 0.15
AMLIF	1.40	2.01 ± 0.09	1.20 ± 0.18
AMLIL		1.85 ± 0.17	1.50 ± 0.14
AMAVF		5.04 ± 0.55	4.36 ± 0.48
AMAVL		5.35 ± 0.14	4.32 ± 0.18
AMAIF		5.82 ± 0.33	4.15 ± 0.51
AMAIL		4.57 ± 0.73	4.13 ± 0.35
AMMVF		0.69 ± 0.20	1.70 ± 0.17
AMMVL		0.77 ± 0.10	1.10 ± 0.17
AMMIF		1.46 ± 0.18	1.53 ± 0.14
AMMIL		0.81 ± 0.20	1.34 ± 0.25
AMVVF		4.04 ± 0.20	4.14 ± 0.19
AMVVL		4.15 ± 0.21	3.63 ± 0.28
AMVIF		4.84 ± 0.25	3.56 ± 0.25
AMVIL		4.08 ± 0.19	3.44 ± 0.72
AVLVF		2.00 ± 0.12	2.04 ± 0.13
AVLVL		1.40 ± 0.10	1.97 ± 0.10
AVLIF		2.48 ± 0.38	2.47 ± 0.12
AVLIL		2.00 ± 0.15	1.97 ± 0.07
AVAVF		$5.96 \pm nan^a$	4.77 ± 1.16
AVAVL	3.50	4.65 ± 0.94	5.40 ± 0.21
AVAIF		$inf^b \pm nan^a$	5.93 ± 0.36
AVAIL		6.69 ± 0.71	5.42 ± 0.54
AVMVF		1.29 ± 0.32	1.91 ± 0.24
AVMVL	2.30	1.07 ± 0.15	1.70 ± 0.10
AVMIF		2.40 ± 0.39	1.89 ± 0.11
AVMIL		1.78 ± 0.21	1.98 ± 0.11
AVVVF		4.65 ± 0.37	4.38 ± 0.32
AVVVL		4.46 ± 0.10	4.42 ± 0.26
AVVIF		5.92 ± 0.63	5.23 ± 0.45
AVVIL		4.85 ± 0.28	4.70 ± 0.30
AWLVF		4.38 ± 0.24	4.34 ± 0.21
AWLVL		$5.83 \pm nan^a$	4.87 ± 0.31
AWLIF		5.31 ± 0.28	4.11 ± 0.17
AWLIL		$6.29 \pm nan^a$	5.37 ± 0.30
AWAVF		4.63 ± 1.04	6.11 ± 0.15
AWAVL		6.30 ± 1.03	7.51 ± 0.55
AWAIF		6.04 ± 0.74	6.18 ± 0.20
AWAIL		6.37 ± 0.39	6.67 ± 0.62
AWMVF		2.26 ± 0.09	2.27 ± 0.44
AWMVL		4.42 ± 0.21	5.18 ± 0.42
AWMIF		2.51 ± 0.17	2.12 ± 0.36
AWMIL		4.25 ± 0.55	4.22 ± 0.61
AWVVF		5.35 ± 0.18	4.77 ± 0.35
AWVVL		7.00 ± 0.25	6.30 ± 0.44
AWVIF		5.73 ± 0.16	4.93 ± 0.21
AWVIL		7.34 ± 0.40	6.10 ± 0.89

[a] Insufficient sampling to estimate uncertainty
[b] Sequence not sampled

Table S9: 5 Site System Folding Free Energies - Part 3

AA	Experiment	FSWITCH	PME
IALVF		2.68 ± 0.22	2.44 ± 0.34
IALVL		1.32 ± 0.13	1.68 ± 0.15
IALIF		3.71 ± 0.34	2.83 ± 0.25
IALIL		2.01 ± 0.21	1.80 ± 0.16
IAAVF		$7.65 \pm nan^a$	$inf^b \pm nan^a$
IAAVL		6.03 ± 0.35	4.96 ± 0.44
IAAIF		$9.11 \pm nan^a$	6.35 ± 0.25
IAAIL		6.48 ± 0.54	4.88 ± 0.27
IAMVF		3.00 ± 0.27	3.06 ± 0.27
IAMVL		1.64 ± 0.30	2.74 ± 0.16
IAMIF		2.86 ± 0.90	2.06 ± 0.28
IAMIL		1.96 ± 0.23	2.24 ± 0.21
IAVVF		5.84 ± 0.39	$4.80 \pm nan^a$
IAVVL		4.88 ± 0.35	3.12 ± 0.78
IAVIF		$6.08 \pm nan^a$	4.41 ± 0.50
IAVIL		5.51 ± 0.27	3.22 ± 0.59
ILLVF		1.69 ± 0.21	2.19 ± 0.17
ILLVL		0.84 ± 0.12	1.82 ± 0.18
ILLIF		2.17 ± 0.35	2.12 ± 0.21
ILLIL		1.53 ± 0.14	1.76 ± 0.17
ILAVF		6.04 ± 0.67	5.93 ± 0.27
ILAVL		4.27 ± 0.29	5.29 ± 0.29
ILAIF		5.19 ± 0.76	5.04 ± 0.34
ILAIL		5.15 ± 0.41	5.25 ± 0.38
ILMVF		1.10 ± 0.20	2.34 ± 0.14
ILMVL		0.23 ± 0.30	1.79 ± 0.29
ILMIF		1.35 ± 0.33	1.82 ± 0.25
ILMIL		1.06 ± 0.29	2.12 ± 0.18
ILVVF		3.93 ± 0.24	4.44 ± 0.17
ILVVL		3.13 ± 0.30	3.62 ± 0.19
ILVIF		3.58 ± 0.42	3.99 ± 0.21
ILVIL		3.96 ± 0.24	3.57 ± 0.27
IMLVF		2.65 ± 0.21	3.41 ± 0.28
IMLVL		1.18 ± 0.17	1.34 ± 0.08
IMLIF		2.45 ± 0.58	2.66 ± 0.16
IMLIL		1.79 ± 0.18	1.47 ± 0.14
IMAVF		$6.29 \pm nan^a$	5.85 ± 0.26
IMAVL		4.53 ± 0.37	4.57 ± 0.30
IMAIF		$6.75 \pm nan^a$	5.26 ± 0.21
IMAIL		6.12 ± 0.24	4.78 ± 0.17
IMMVF		1.65 ± 0.17	2.47 ± 0.21
IMMVL		0.50 ± 0.17	1.43 ± 0.12
IMMIF		2.11 ± 0.20	2.00 ± 0.13
IMMIL		1.05 ± 0.31	1.39 ± 0.14
IMVVF		4.49 ± 0.32	4.71 ± 0.35
IMVVL		3.94 ± 0.15	3.51 ± 0.38
IMVIF		5.47 ± 0.41	4.66 ± 0.17
IMVIL		4.70 ± 0.35	3.49 ± 0.32
IVLVF		2.62 ± 0.45	2.79 ± 0.19
IVLVL		1.15 ± 0.10	1.80 ± 0.09
IVLIF		4.13 ± 0.27	2.54 ± 0.10
IVLIL		1.82 ± 0.16	1.81 ± 0.12
IVAVF		$5.94 \pm nan^a$	$inf^b \pm nan^a$
IVAVL		5.16 ± 0.35	5.33 ± 0.19
IVAIF		$8.16 \pm nan^a$	$6.64 \pm nan^a$
IVAIL		6.42 ± 0.53	5.36 ± 0.21
IVMVF		1.79 ± 0.57	2.58 ± 0.15
IVMVL		0.68 ± 0.14	1.75 ± 0.16
IVMIF		3.06 ± 0.37	2.52 ± 0.20
IVMIL		1.55 ± 0.17	1.85 ± 0.17
IVVVF		4.80 ± 0.45	5.40 ± 0.31
IVVVL		3.88 ± 0.42	4.62 ± 0.24
IVVIF		6.41 ± 0.32	4.93 ± 0.18
IVVIL		4.79 ± 0.17	4.00 ± 0.27
IWLVF		5.32 ± 0.34	4.76 ± 0.60
IWLVL		4.67 ± 0.84	$5.80 \pm nan^a$
IWLIF		5.63 ± 0.32	4.60 ± 0.37
IWLIL		6.41 ± 0.56	5.86 ± 0.63
IWAVF		5.83 ± 0.23	5.71 ± 0.16
IWAVL		5.89 ± 0.21	6.57 ± 0.12
IWAIF		5.38 ± 0.41	5.32 ± 0.15
IWAIL		6.26 ± 0.27	5.76 ± 0.20
IWMVF	1.40	2.74 ± 0.22	2.94 ± 0.23
IWMVL		3.09 ± 0.16	4.31 ± 0.18
IWMIF		3.19 ± 0.17	2.79 ± 0.20
IWMIL		3.59 ± 0.28	4.04 ± 0.17
IWVVF		5.67 ± 0.22	5.13 ± 0.22
IWVVL		5.73 ± 0.20	5.68 ± 0.21
IWVIF		5.82 ± 0.21	5.00 ± 0.17
IWVIL		6.28 ± 0.16	5.37 ± 0.20

[a] Insufficient sampling to estimate uncertainty

[b] Sequence not sampled

Table S10: 3 Site System Folding Free Energies - SSλD

AA	Experiment	FSWITCH	PME
IIL	0.00	0.00 ± 0.00	0.00 ± 0.00
MIL	2.20	0.94 ± 0.15	0.97 ± 0.10
IML	3.10	3.95 ± 0.10	3.77 ± 0.07
IIM	2.00	2.97 ± 0.02	2.40 ± 0.12

Table S11: 4 Site System Folding Free Energies - SSλD

AA	Experiment	FSWITCH	PME
LMVF	0.00	0.00 ± 0.00	0.00 ± 0.00
FMVF	0.40	0.35 ± 0.09	0.27 ± 0.12
LLVF	0.74	-0.60 ± 0.09	0.24 ± 0.20
LMFF	1.43	2.73 ± 0.40	2.62 ± 0.19
LMIF	0.69	1.29 ± 0.42	1.12 ± 0.21
LMVL	-0.20	-1.30 ± 0.09	-0.76 ± 0.12

Table S12: 5 Site System Folding Free Energies - SSλD

AA	Experiment	FSWITCH	PME
LALVF	0.00	0.00 ± 0.00	0.00 ± 0.00
AALVF	2.70	2.00 ± 0.23	$2.06 \pm nan$[a]
IALVF		3.15 ± 0.14	2.92 ± 0.28
LLLVF		-0.99 ± 0.32	0.41 ± 0.29
LMLVF		0.36 ± 0.27	1.37 ± 0.35
LVLVF		-0.27 ± 0.26	0.60 ± 0.33
LWLVF	2.20	3.37 ± 0.24	$4.05 \pm nan$[a]
LAAVF		5.01 ± 0.32	4.61 ± 0.14
LAMVF		0.01 ± 0.13	-0.00 ± 0.10
LAVVF		4.63 ± 0.13	3.94 ± 0.22
LALIF		0.45 ± 0.10	-0.21 ± 0.07
LALVL	-0.20	-1.30 ± 0.09	-0.76 ± 0.12

[a] Insufficient sampling to estimate uncertainty

References

(1) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *Journal of Computational Chemistry* **1992**, *13*, 1011–1021.

(2) Hayes, R. L.; Armacost, K. A.; Vilseck, J. Z.; Brooks, C. L., III *Journal of Physical Chemistry B* **2017**, *121*, 3626–3635.

(3) Knight, J. L.; Brooks, C. L., III *Journal of Computational Chemistry* **2009**, *30*, 1692–1700.

(4) Armacost, K. A.; Goh, G. B.; Brooks, C. L., III *Journal of Chemical Theory and Computation* **2015**, *11*, 1267–1277.

(5) Steinbach, P. J.; Brooks, B. R. *Journal of Computational Chemistry* **1994**, *15*, 667–683.

(6) Darden, T.; York, D.; Pedersen, L. *Journal of Chemical Physics* **1993**, *98*, 10089–10092.

(7) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *Journal of Chemical Physics* **1995**, *103*, 8577–8593.

(8) Huang, Y.; Chen, W.; Wallace, J. A.; Shen, J. *Journal of Chemical Theory and Computation* **2016**, *12*, 5411–5421.

(9) Hummer, G.; Pratt, L. R.; García, A. E. *Journal of Physical Chemistry* **1996**, *100*, 1206–1215.

(10) Hünenberger, P. H.; McCammon, J. A. *Journal of Chemical Physics* **1999**, *110*, 1856–1872.

(11) Hub, J. S.; de Groot, B. L.; Grubmüller, H.; Groenhof, G. *Journal of Chemical Theory and Computation* **2014**, *10*, 381–390.

(12) Rocklin, G. J.; Mobley, D. L.; Dill, K. A.; Hünenberger, P. H. *Journal of Chemical Physics* **2013**, *139*, 184103.

(13) Vilseck, J. Z.; Armacost, K. A.; Hayes, R. L.; Goh, G. B.; Brooks, C. L., III *Journal of Physical Chemistry Letters* **2018**, *9*, 3328–3332.

(14) Steinbrecher, T.; Zhu, C.; Wang, L.; Abel, R.; Negron, C.; Pearlman, D.; Feyfant, E.; Duan, J.; Sherman, W. *Journal of Molecular Biology* **2017**, *429*, 948–963.

(15) Clark, A. J.; Gindin, T.; Zhang, B.; Wang, L.; Abel, R.; Murret, C. S.; Xu, F.; Bao, A.; Lu, N. J.; Zhou, T.; Kwong, P. D.; Shapiro, L.; Honig, B.; Friesner, R. A. *Journal of Molecular Biology* **2017**, *429*, 930–947.

(16) Nicholson, H.; Anderson, D. E.; Pin, S. D.; Matthews, B. W. *Biochemistry* **1991**, *30*, 9816–9828.

(17) Blaber, M.; Lindstrom, J. D.; Gassner, N.; Xu, J.; Heinz, D. W.; Matthews, B. W. *Biochemistry* **1993**, *32*, 11363–11373.

(18) Liu, R.; Baase, W. A.; Matthews, B. W. *Journal of Molecular Biology* **2000**, *295*, 127–145.

(19) Eriksson, A. E.; Baase, W. A.; Matthews, B. W. *Journal of Molecular Biology* **1993**, *229*, 747–769.

(20) Baldwin, E.; Baase, W. A.; jun Zhang, X.; Feher, V.; Matthews, B. W. *Journal of Molecular Biology* **1998**, *277*, 467–485.

(21) Dao-Pin, S.; Anderson, D. E.; Baase, W. A.; Dahlquist, F. W.; Matthews, B. W. *Biochemistry* **1991**, *30*, 11521–11529.

(22) Lipscomb, L. A.; Gassner, N. C.; Snow, S. D.; Eldridge, A. M.; Baase, W. A.; Drew, D. L.; Matthews, B. W. *Protein Science* **1998**, *7*, 765–773.

(23) Xu, J.; Baase, W. A.; Quillin, M. L.; Baldwin, E. P.; Matthews, B. W. *Protein Science* **2001**, *10*, 1067–1078.

(24) Gassner, N. C.; Baase, W. A.; Lindstrom, J. D.; Lu, J.; Dahlquist, F. W.; Matthews, B. W. *Biochemistry* **1999**, *38*, 14451–14460.

The University of Michigan

Department of Chemistry
Biophysics Program
2006 Chemistry
930 N. University Avenue
Ann Arbor, Michigan 48109-1055

Charles L. Brooks III, Ph.D.
Cyrus Levinthal Distinguished
Professor of Chemistry and Biophysics
Warner-Lambert/Parke-Davis
Professor of Chemistry
Professor of Biophysics

Phone 734/647-6682
Fax 734/6471604
E-mail brookscl@umich.edu

August 6, 2018

Professor Ron Levy
Department of Chemistry
Temple University
Philadelphia, PA

Dear Ron,

Please find our revised manuscript entitled “Approaching Protein Design with Multisite λ Dynamics: Accurate and Scalable Mutational Folding Free Energies in T4 Lysozyme” (ID PRO-18-0114), co-authored by RL Hayes, JZ Vilseck and CL Brooks III.

We have revised the manuscript per the reviewer comments and provide a detailed response to the reviewers below. We hope that out paper is now acceptable for publication in Protein Science.

Sincerely,

Charles L. Brooks III
Cyrus Levinthal Distinguished University Professor
of Chemistry and Biophysics
Warner-Lambert/Parke-Davis Professor
of Chemistry
Professor of Biophysics
Professor of Chemistry
Chair of Biophysics

Charles L. Brooks III
Cyrus Levinthal Distinguished University Professor
of Chemistry and Biophysics 2
Warner-Lambert/Parke-Davis Professor of Chemistry
Professor of Biophysics

Referee(s)' Comments to Author:

Reviewer 1

Comments to the Author

This is a high quality paper that is easy to referee. Lambda-dynamics is extended to the protein design problem of computing stabilities for a large library of lysozyme sequences, which differ at a few positions. Lambda-dynamics has improved over the years in important ways and it provides here accurate folding free energy differences for 240 lysozyme sequences using a single MD simulation. This is at the lower limit of bona fide protein design problems but the method has the obvious potential to scale further. The method opens new possibilities for protein design, providing high-accuracy, medium-throughput simulations that can be applied to a problem eg after a first, very high-throughput pass has been done with a less accurate design tool. Overall, the methodology is very original and important. The quality of the data and the discussion are high. The presentation is clear and complete. The paper is very well-suited to Protein Science.

Thanks for a favorable review. Your explanation of potential use cases is quite insightful.

Reviewer 2

Comments to the Author

This manuscript reports on the performance of a computational method for computing free energy differences known as multisite λ dynamics. Application is to predicting changes in protein stability upon mutation of side chains in T4 lysozyme, for which there is a wealth of data. A major advantage of MSλD is the need for only simulating the ends states and that the free energy can be computed for multiple sequences simultaneously. This allows the search of a much larger sequence space than possible with FEP, which is restricted to one site at a time and several intermediate simulations between the end states are needed. The multisite results are encouraging and a nice general discussion of the value of MSλD for protein design is presented. On the other hand, the Methods and Supplementary sections are highly technical and directed to someone already intimately familiar with MSλD. Nevertheless, because the methodology is directed toward protein design, and the results indicate accurate free energy differences can be obtained, the paper should be of interest to the general audience of Protein Science if the following points can be addressed in a revised ms.

A. The multiple site calculations are of most interest but some additional explanation is needed for how the calculation is done and what is the outcome. The results in Fig 2 are a major part of the work being reported, so that clarifying the points below would help the reader understand these results.

1. The sentence on p. 8 “The 3 site, 4 site and 5 site systems comprised 8, 24, and 240 sequences with 6, 14, and 9 experimental measurements at the same pH” needs to be explained. The 8, 24 and 240 is some combination of the mutants listed in table 3 but it’s not clear how the number of sequences is determined. And, how is this very large number of sequences (8+24+240) connected to the much smaller number of experimental measurements (6+14+9)? The basis for these numbers needs to be spelled out a bit more.

We added more details to the introduction of these numbers on page 8, and note that we only compare the experimentally measured sequences with experiment.

2. What are the experimental data plotted in Figure 2 in the case of multisite mutations? The table in ref 49 with experimental data shows single mutation free energy changes. Double mutants are typically not the sum of single mutants, so what is compared for multisite FE

Charles L. Brooks III
Cyrus Levinthal Distinguished University Professor
of Chemistry and Biophysics
Warner-Lambert/Parke-Davis Professor of Chemistry
Professor of Biophysics

differences? Are the MSλD results somehow using FE values for only single sites relative to WT? Further, the figure is labeled "All multisite mutants" yet the description quoted above in #1 gives 8+24+240 sequences, which clearly is not the number of values plotted in figure 2. So what is meant by "all multisite mutants"?

While most of the mutations in ref 49 are point mutants, several dozen are multisite mutants. We have modified the text to mention we only compare with the experimentally measured sequences. We also changed the titles of the figures from "all multisite mutants" to "all multisite systems".

3. How are the sites treated in a multisite MSλD run? It is my understanding that all targeted sites are scaled in a multidimensional λ space in a single run. If one residue has two or more substitutions (e.g. V111 in Table 3), are all of the amino acid types mutated in one run? How are the interactions handled between residues of different sites? For example, do mutants at residue 99 interact with mutants at residue 102 in the 4 site system?

We modified the single site section to mention all mutations at a site are present in the same simulation, and added the sentence "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 MSλD to explicitly account for coupling between sites," to the multisite section to address this confusion. So the answer to your question is yes, whichever side chain is on (with nonzero λ) at residue 99 will interact with whichever side chain is on at site 102.

4. What is set to zero in Figure 1 and 2? Presumably it is WT but it is not specified in the caption.

We added the phrase "free energies are plotted relative to wild type with C54T/C97A" to both captions.

5. How is convergence assessed for the MSλD results? Is there some measure of in terms of the evolution of λ?

Convergence is typically assessed through the statistical uncertainty in the 5 duplicate runs. Table S2 in the Supporting Information shows the lack of convergence that can occur if optimal biasing potentials are not used. Convergence is also touched on in Figure 3c since SSλD simulations should be more converged than MSλD simulations because fewer sequences need to be sampled. Temporal convergence is beyond the scope of this article, but we ran 20 to 40 ns because this is in the ballpark of what is conventional with MSλD (10 to 30 ns in recent publications).

B. The point that MSλD is more efficient than FEP is made repeatedly and argued in a descriptive manner with statements about the number of simulations for FEP vs MSλD, etc. The efficiency of MSλD makes sense, but could a more quantitative comparison be given, such as a rough estimate of the relative CPU time required for a large scale study using FEP compared to using MSλD?

No, a more quantitative comparison cannot be given without expending a lot of computer time. In another study of drug binding, we ran FEP on a small subset, roughly 7%, of the ligands studied with MSλD, but the FEP simulations took more time than the MSλD simulations. In that study a difference of a factor of 20-30 was estimated. The central issue is that one has to obtain free energy estimates of comparable precision on the same system, as some systems (e.g. small to large vs. large to small mutations in proteins) converge at different rates. We ran our simulations for roughly the same amount of time as

Charles L. Brooks III
Cyrus Levinthal Distinguished University Professor
of Chemistry and Biophysics 4
Warner-Lambert/Parke-Davis Professor of Chemistry
Professor of Biophysics

standard FEP simulations, but obtained far better precision and accuracy. Running long enough FEP simulations to obtain that level of precision would be intractable. We added a long discussion to the SI and mention it in the main text.

C. The authors should make some kind of comparison between the mutant structures from MSλD with the crystal structures. In particular, there are structural changes noted in ref 49 associated with the L99A cavity. Such a comparison would be useful, particularly given that the authors speculate that relaxation of the structure influences their results.

This was the most difficult and most fruitful suggestion. We have added 3 pages to the SI and two paragraphs and a figure to the main text discussing structural relaxation in the context of solved structures. Ironically, the structural relaxations for L99 are quite rapid, but this is not the case for several of the other sites.

Some minor points on the written presentation are the following:
1. P. 5: “C54T/C97A background” should be defined

Done

2. Footnote to Table 2 is a squared quantity but the values are labeled root mean squared

Thanks, we have corrected the equation in tables 2 and 4

3. P. 10 last paragraph: “5/2 sequences” and “23/13 sequences.” What is the slash?

The slash denotes either FSWITCH or PME simulations. We have reworded the sentence to clarify.

4. P. 15 The first paragraph of section Prospects of MSλD for Protein Design is not clear. Suggest rewriting.

We have rewritten this section to clarify the difference between state-of-the-art protein design methods which are generative, and the FEP approach which is currently only accurate enough to perturb existing sequences.

5. P. 21-22 bottom. One sentence states only side chain dihedrals are scaled, yet there is discussion about phi and psi angles also being scaled. Please clarify.

Some dihedrals including C_β involve the same rotatable bond as the ϕ and ψ angles. We have clarified this point in the text.

http://brooks.chem.lsa.umich.edu *brookscl@umich.edu*