



OPEN

# Large-scale application of free energy perturbation calculations for antibody design

Fangqiang Zhu<sup>1</sup>✉, Feliza A. Bourguet<sup>1</sup>, William F. D. Bennett<sup>1</sup>, Edmond Y. Lau<sup>1</sup>, Kathryn T. Arrildt<sup>1</sup>, Brent W. Segelke<sup>1</sup>, Adam T. Zemla<sup>2</sup>, Thomas A. Desautels<sup>3</sup> & Daniel M. Faissol<sup>3</sup>✉

Alchemical free energy perturbation (FEP) is a rigorous and powerful technique to calculate the free energy difference between distinct chemical systems. Here we report our implementation of automated large-scale FEP calculations, using the Amber software package, to facilitate antibody design and evaluation. In combination with Hamiltonian replica exchange, our FEP simulations aim to predict the effect of mutations on both the binding affinity and the structural stability. Importantly, we incorporate multiple strategies to faithfully estimate the statistical uncertainties in the FEP results. As a case study, we apply our protocols to systematically evaluate variants of the m396 antibody for their conformational stability and their binding affinity to the spike proteins of SARS-CoV-1 and SARS-CoV-2. By properly adjusting relevant parameters, the particle collapse problems in the FEP simulations are avoided. Furthermore, large statistical errors in a small fraction of the FEP calculations are effectively reduced by extending the sampling, such that acceptable statistical uncertainties are achieved for the vast majority of the cases with a modest total computational cost. Finally, our predicted conformational stability for the m396 variants is qualitatively consistent with the experimentally measured melting temperatures. Our work thus demonstrates the applicability of FEP in computational antibody design.

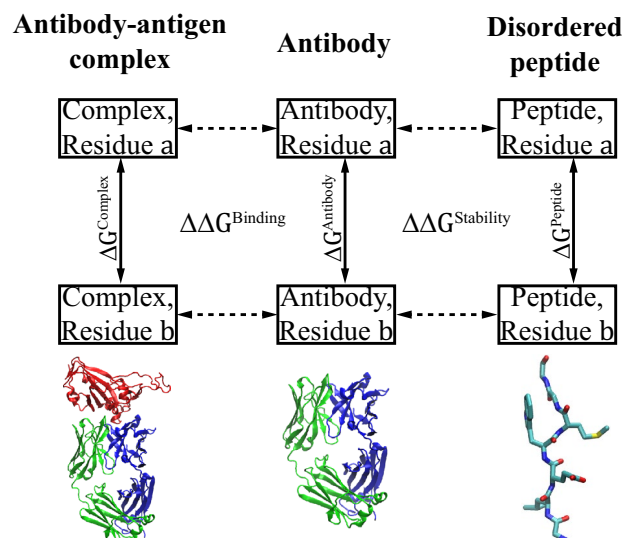
Alchemical free energy perturbation (FEP) simulation<sup>1</sup> is a rigorous physics-based method to calculate the free energy difference between distinct chemical systems. Due to recent technological advancement, FEP is now capable of accurately predicting relative binding affinities<sup>2–4</sup> and has found increasingly more applications in drug development.

Antibodies are proteins produced by the immune system to bind specific antigenic proteins and thereby neutralize the antigen. Therapeutically, antibodies can be designed and manufactured as effective medicines against the infection. Recently, FEP has been successfully applied to predict the relative binding affinity between antibody and antigen<sup>5,6</sup>, thus potentially facilitating antibody design<sup>7–9</sup> and optimization.

In our ongoing battle against the current COVID19 pandemic<sup>10</sup>, we aim to develop antibodies to neutralize SARS-CoV-2 and other coronaviruses. Toward this goal, we introduced FEP simulations as one of the computational tools for our multidisciplinary team. Specifically, the basic task of FEP is to predict the change in the binding affinity due to proposed mutations on the antibody. In our workflow, batches of mutations are routinely proposed for evaluation, and quick turnarounds are necessary for further decision making. Such requirements necessitate automated processing of many FEP calculations. In this situation, it is not feasible to manually examine all the simulations individually and identify potential problems therein. Therefore, a faithful and automated estimation of the uncertainties in such calculations is especially important, as it could provide the level of confidence for the FEP results when data from many different sources are integrated to inform the decision making. With these requirements in mind, we implemented automated protocols for performing FEP calculations with Hamiltonian replica exchange<sup>11</sup> using the Amber software package<sup>12</sup>, along with an uncertainty estimation that incorporates a number of factors in the analysis of the simulations.

Whereas the binding affinity to antigen is a critical component for the efficacy of an antibody, the structural stability is important for the developability<sup>13</sup>, such as manufacture, storage, and distribution of the antibody.

<sup>1</sup>Biosciences and Biotechnology Division, Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, USA. <sup>2</sup>Global Security Computing Division, Computing Directorate, Lawrence Livermore National Laboratory, Livermore, USA. <sup>3</sup>Computational Engineering Division, Engineering Directorate, Lawrence Livermore National Laboratory, Livermore, USA. ✉email: zhu13@llnl.gov; faissol1@llnl.gov



**Figure 1.** Schematics for the FEP calculations in this study. The free energy differences  $\Delta G^{\text{Complex}}$ ,  $\Delta G^{\text{Antibody}}$ , and  $\Delta G^{\text{Peptide}}$  between systems with residues *a* and *b* are obtained from the corresponding FEP simulations, and the results further determine (Eqs. (1) and (2)) the effect of mutation on the binding ( $\Delta\Delta G^{\text{Binding}}$ ) and the stability ( $\Delta\Delta G^{\text{Stability}}$ ) of the antibody. The illustrated protein structures are from the m396 antibody<sup>25</sup>, rendered by the VMD program<sup>43</sup> (version 1.9.3).

Therefore, computational stability evaluation<sup>14</sup> is desired in our antibody design. Recent studies demonstrated that FEP<sup>15–18</sup> and other computational approaches<sup>19–24</sup> could provide reasonable predictions for the relative stability of protein mutants. In this project, we also incorporated stability prediction by FEP calculations, such that our FEP protocols evaluate the effect of proposed mutation on both the binding affinity and the conformational stability of the antibody.

In this report, we take the m396 antibody<sup>25</sup> as a case study to demonstrate our FEP protocols. m396 is known to neutralize the coronavirus SARS-CoV-1 by binding to the receptor binding domain (RBD) of the viral spike protein<sup>25</sup>, but does not bind or neutralize SARS-CoV-2<sup>26</sup>. The RBD is a self-contained and stable domain, and isolated RBDs could independently bind m396 without other parts of the spike protein<sup>25</sup>. One of our objectives is to modify relevant residues of m396 such that the mutated antibody could bind the SARS-CoV-2 RBD. The focus of this article is only on the implementation of FEP calculations, and our much broader efforts in the antibody design will be reported in due course.

## Methods

In this study, the basic task of an FEP calculation is to evaluate two effects due to the mutation of a single residue on the antibody from type *a* to type *b*, as schematized in Fig. 1. The first effect, quantified by

$$\Delta\Delta G^{\text{Binding}} = \Delta G^{\text{Complex}} - \Delta G^{\text{Antibody}}, \quad (1)$$

describes the shift in equilibrium between the bound and unbound states. Here  $\Delta G^{\text{Complex}}$  is the difference in free energy between the bound systems with the concerned residue in type *a* and type *b*, respectively. Similarly,  $\Delta G^{\text{Antibody}}$  is the free energy difference between the unbound antibody systems of *a* and *b*. The difference between the two  $\Delta G$ s above,  $\Delta\Delta G^{\text{Binding}}$ , directly determines the change in the binding affinity due to the mutation. The second effect, quantified by

$$\Delta\Delta G^{\text{Stability}} = \Delta G^{\text{Antibody}} - \Delta G^{\text{Peptide}}, \quad (2)$$

describes the shift in equilibrium between the native (folded) and denatured (unfolded) conformations of the antibody. The structure of denatured antibody is unknown and is thus approximated by a highly simplified model of 7-residue peptide<sup>18</sup> here.

Because our FEP simulations are under constant volume, the results here correspond to the differences in the Helmholtz free energy (*A*). In contrast, the thermodynamics of constant-pressure systems should be described by the Gibbs free energy (*G*), with  $G = A + PV$ , where *P* and *V* are respectively the pressure and the volume. However, given that the contribution of the *PV* term to the  $\Delta\Delta G$  is typically small for biochemical systems, we consider  $\Delta\Delta A \approx \Delta\Delta G$  and use *A* and *G* interchangeably in this article.

In this section, we will first review the fundamental theories involved in FEP calculations. We will then describe our specific FEP simulations to evaluate the m396 antibody variants for their conformational stability and their binding to the SARS-CoV-1 and SARS-CoV-2 RBDs. We will also describe our experiments to measure the thermal stability of the antibodies.

**Free energy difference between two systems.** Consider two systems  $i$  and  $j$ , both consisting of an identical set of atoms but with different energy functions  $E_i(\vec{X})$  and  $E_j(\vec{X})$ , respectively, where  $\vec{X}$  represents the microstate of the system, i.e., the positions and velocities of all the atoms. Under constant temperature ( $T$ ) and volume, the difference in the Helmholtz free energy between the two systems is given by:

$$A_j - A_i = -k_B T \ln \frac{\int \exp[-E_j(\vec{X})/k_B T] d\vec{X}}{\int \exp[-E_i(\vec{X})/k_B T] d\vec{X}}, \quad (3)$$

where  $k_B$  is the Boltzmann constant.

The free energy difference above can be rewritten in a more practical and useful form:

$$A_j - A_i = -k_B T \ln \left\langle \exp \left[ -\frac{E_j(\vec{X}) - E_i(\vec{X})}{k_B T} \right] \right\rangle_i, \quad (4)$$

in which  $\langle \rangle_i$  represents the expected value in the equilibrium ensemble of system  $i$ . This expected value can be estimated as the average of  $\exp\left(-\frac{\Delta E_{ij}}{k_B T}\right)$ , where  $\Delta E_{ij}(\vec{X}) \equiv E_j(\vec{X}) - E_i(\vec{X})$ , with  $\vec{X}$  taken from an equilibrium sampling of system  $i$  using, e.g., molecular dynamics (MD) simulations. This method<sup>27</sup> is termed exponential averaging (EXP). Alternatively, if one performs equilibrium sampling for system  $j$  instead, the free energy can also be obtained in a similar way:

$$A_j - A_i = k_B T \ln \left\langle \exp \left[ -\frac{E_i(\vec{X}) - E_j(\vec{X})}{k_B T} \right] \right\rangle_j, \quad (5)$$

with the expected value above estimated as the average of  $\exp\left(-\frac{\Delta E_{ji}}{k_B T}\right)$ , where  $\Delta E_{ji} = E_i(\vec{X}) - E_j(\vec{X}) = -\Delta E_{ij}$ .

If equilibrium sampling is available for both systems  $i$  and  $j$ , the Bennett acceptance ratio (BAR) method<sup>28</sup> can also provide the free energy difference  $\Delta A \equiv A_j - A_i$ , with better statistical accuracy. The BAR method is based on the following relation:

$$\left\langle \frac{1}{1 + \exp\left[\frac{\Delta E_{ij}(\vec{X}) - \Delta A}{k_B T}\right]} \right\rangle_i = \left\langle \frac{1}{1 + \exp\left[\frac{\Delta E_{ji}(\vec{X}) + \Delta A}{k_B T}\right]} \right\rangle_j. \quad (6)$$

Again, the expected values above can be estimated as the corresponding averages, with  $\Delta E_{ij}(\vec{X})$  and  $\Delta E_{ji}(\vec{X})$  obtained from the sampled microstates in systems  $i$  and  $j$ , respectively. Equation (6) thus contains only one unknown,  $\Delta A$ , which can be numerically solved. If the sample sizes for systems  $i$  and  $j$  are different, Eq. (6) could be slightly modified into a more general form that minimizes the statistical error<sup>28</sup>. In this study, since the two involved systems always have identical sample size, Eq. (6) is the optimal form to calculate the  $\Delta A$ .

**Sampling with intermediate windows.** Although the free energy methods in “Free energy difference between two systems” are correct and rigorous, they are only practically useful if the two systems are sufficiently similar such that their equilibrium distributions of the microstates  $\vec{X}$  are largely overlapped. If the two systems are substantially different, direct applications of Eqs. (4)–(6) will be practically difficult to achieve reliable results. In such cases, one could introduce intermediate systems (each conventionally called a “window”), calculate the free energy difference between adjacent systems using the methods in “Free energy difference between two systems”, and then add these increments to obtain the free energy difference between the two end systems.

Suppose we aim to calculate the free energy difference between two dissimilar systems,  $a$  and  $b$ , with the same set of atoms. Typically, intermediate systems can be constructed by introducing a parameter  $\lambda$ , in the range of  $[0, 1]$ , into the energy function  $E(\vec{X}; \lambda)$ , such that  $E(\vec{X}; \lambda = 0) = E_a(\vec{X})$  and  $E(\vec{X}; \lambda = 1) = E_b(\vec{X})$ . Now the system free energy  $A(\lambda)$  is also a function of  $\lambda$ , and the target free energy difference becomes  $A_b - A_a = A(\lambda = 1) - A(\lambda = 0)$ . Intermediate systems between  $a$  and  $b$  can thus be defined by a series of  $\lambda$  values:  $0 = \lambda_1 < \lambda_2 < \dots < \lambda_N = 1$ , such that systems  $i$  and  $i + 1$  are sufficiently similar. With all the incremental changes  $A(\lambda_{i+1}) - A(\lambda_i)$  calculated, their sum then corresponds to  $A_b - A_a$ .

As another alternative to calculate the free energy difference, thermodynamic integration (TI) is based on the following relation:

$$\frac{d}{d\lambda} A(\lambda) = \left\langle \frac{\partial E(\vec{X}; \lambda)}{\partial \lambda} \right\rangle_\lambda, \quad (7)$$

in which  $\langle \rangle_\lambda$  represents the expected value in the system with the energy function  $E(\vec{X}; \lambda)$ . Therefore,

$$A_b - A_a = \int_0^1 \left\langle \frac{\partial E(\vec{X}; \lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda. \quad (8)$$

The simulation setup for TI is identical to that for FEP, both entailing equilibrium sampling in systems  $E(\vec{X}; \lambda_i)$  for  $i=1,2,\dots,N$ . To apply TI, we estimate  $f_i \equiv \left\langle \frac{\partial E(\vec{X}; \lambda)}{\partial \lambda} \right\rangle_{\lambda_i}$  by averaging the  $\frac{\partial E(\vec{X}; \lambda)}{\partial \lambda}$  value over the sampled microstates for each system  $i$ . Then a numerical integration of Eq. (8) using the trapezoidal rule will provide the free energy difference:

$$A_b - A_a \approx \sum_{i=1}^{N-1} \frac{1}{2} (f_i + f_{i+1}) (\lambda_{i+1} - \lambda_i). \quad (9)$$

**Energy function in alchemical system.** To obtain the free energy difference between two systems ( $a$  and  $b$ ) that do not share an identical atom set, an “alchemical” system can be constructed. All the atoms in  $a$  and  $b$  are present in the alchemical system. In Amber<sup>12</sup>, the microstate of an alchemical system is described as  $\vec{X} = (\vec{X}_0, \vec{X}_a^{\text{SC}}, \vec{X}_b^{\text{SC}})$ . Here  $\vec{X}_0$  represents the common atoms in both systems. In addition, two separate sets of coordinates,  $\vec{X}_a^{\text{SC}}$  and  $\vec{X}_b^{\text{SC}}$ , represent the atoms unique to systems  $a$  and  $b$ , respectively, which are called “soft core” (SC) atoms.

The system energy is the sum of kinetic energy and potential energy. Kinetic energy is independent of  $\lambda$  and thus does not contribute to the energy difference  $\Delta E$ . Therefore, we only need to consider the potential energy  $V$  in the calculations. In Amber<sup>12</sup>, the potential energy for the alchemical system has the following form:

$$V(\vec{X}; \lambda) = (1 - \lambda) \cdot V_a(\vec{X}; \lambda) + \lambda \cdot V_b(\vec{X}; \lambda). \quad (10)$$

In the potential  $V_a$  above, the atoms in  $\vec{X}_b^{\text{SC}}$  have no nonbond interaction with  $\vec{X}_0$  or  $\vec{X}_a^{\text{SC}}$  and are thus decoupled from the rest of the system. Similarly, in  $V_b$ , the  $\vec{X}_a^{\text{SC}}$  atoms are decoupled from the rest of the system. Furthermore, the dependence of  $V_a$  on  $(\vec{X}_0, \vec{X}_a^{\text{SC}})$  and the dependence of  $V_b$  on  $(\vec{X}_0, \vec{X}_b^{\text{SC}})$  conform to the standard potentials in the MD force field for systems  $a$  and  $b$ , respectively, except that the nonbond interactions involving the SC atoms are modified as described below, to avoid the “end-point catastrophes”.

Specifically, between an atom in  $\vec{X}_0$  and another atom in  $\vec{X}_a^{\text{SC}}$ , the van der Waals (VDW) term in  $V_a$  has the modified form:

$$u_a^{\text{VDW}}(r_{ij}; \lambda) = 4\epsilon \left\{ \frac{1}{[\alpha\lambda + (r_{ij}/\sigma)^6]^2} - \frac{1}{\alpha\lambda + (r_{ij}/\sigma)^6} \right\}, \quad (11a)$$

where  $r_{ij}$  is the distance between the two atoms and  $\alpha$  is a user-defined parameter. When  $\alpha = 0$  or  $\lambda = 0$ ,  $u_a^{\text{VDW}}$  recovers the standard Lennard–Jones potential with the parameters  $\sigma$  and  $\epsilon$ . Similarly, between an atom in  $\vec{X}_0$  and another atom in  $\vec{X}_b^{\text{SC}}$ , the VDW term in  $V_b$  is:

$$u_b^{\text{VDW}}(r_{ij}; \lambda) = 4\epsilon \left\{ \frac{1}{[\alpha(1 - \lambda) + (r_{ij}/\sigma)^6]^2} - \frac{1}{\alpha(1 - \lambda) + (r_{ij}/\sigma)^6} \right\}. \quad (11b)$$

In addition, the electrostatic terms involving the SC atoms are modified as:

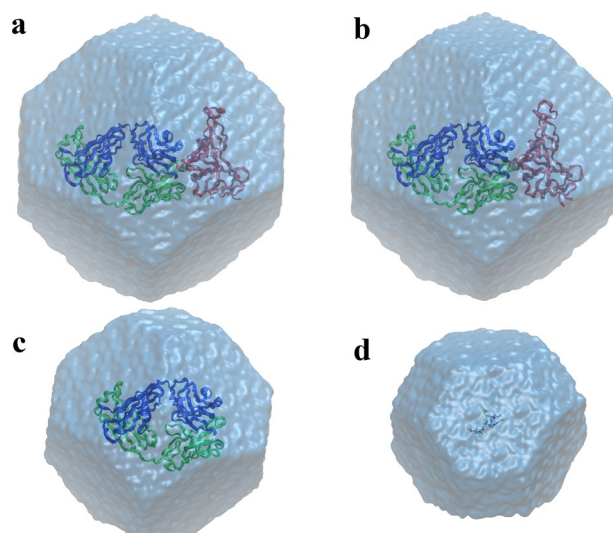
$$u_a^{\text{EL}}(r_{ij}; \lambda) = \frac{q_i q_j}{4\pi\epsilon_0 \sqrt{\beta\lambda + r_{ij}^2}}, \quad (12a)$$

$$u_b^{\text{EL}}(r_{ij}; \lambda) = \frac{q_i q_j}{4\pi\epsilon_0 \sqrt{\beta(1 - \lambda) + r_{ij}^2}}, \quad (12b)$$

where  $q_i$  and  $q_j$  are the charges, and  $\beta$  is a user-defined parameter. When  $\beta = 0$ ,  $u_a^{\text{EL}}$  and  $u_b^{\text{EL}}$  recover the standard Coulomb potential.

With the designs above, when  $\lambda$  is 0 or 1, the  $V(\vec{X}; \lambda)$  in Eq. (10) becomes the potential energy of system  $a$  or  $b$ , respectively. The systems with the intermediate  $\lambda$  values thus establish an alchemical path between  $a$  and  $b$ .

**Alchemical system for mutation of protein residues.** Our task here is to evaluate the mutation of a single residue in the protein from type  $a$  to type  $b$ , excluding proline (Pro) and any charge-reversing mutation (i.e., between a positively and a negatively charged types). The SC atoms (i.e., in  $\vec{X}_a^{\text{SC}}$  or  $\vec{X}_b^{\text{SC}}$ ) in our setup



**Figure 2.** Four simulation systems in this study. **(a)** Complex of the m396 Fab and the SARS-CoV-1 RBD. **(b)** Complex of the m396 Fab and the SARS-CoV-2 RBD. **(c)** m396 Fab. **(d)** A 7-residue peptide representing the local structure of denatured antibody. Images were rendered by VMD<sup>43</sup> (version 1.9.3).

consist of the side chain atoms of the residue other than the  $C_{\beta}$ , except when  $a$  or  $b$  is a Gly, in which case the SC includes the hydrogen atoms bonded to the  $C_{\alpha}$  as well as the side chain of the non-Gly residue.

For charge-changing mutations, i.e., between a charged type and a neutral one, a co-alchemical ion is introduced to render the system electrically neutral for both end states. Specifically, the co-alchemical ion is a  $\text{Na}^+$  if the charged residue is negative, or a  $\text{Cl}^-$  if the residue is positively charged. In the end state corresponding to the neutral protein residue, the co-alchemical ion is assigned a zero charge, thus ensuring that the systems at both ends have identical total charge. The co-alchemical ion was initially placed in the bulk solution. In the simulations, we applied positional restraints (with a spring constant of  $1 \text{ kcal/mol/\text{\AA}^2}$ ) on the co-alchemical ion as well as a  $C_{\alpha}$  atom at the center of the protein, to ensure that the co-alchemical ion is always in the bulk region and has no close encounter with the protein.

**Preparation and equilibration of wildtype m396 systems.** In this study, FEP simulations were performed on four systems: (1) m396 bound to SARS-CoV-1 RBD; (2) m396 bound to SARS-CoV-2 RBD; (3) m396 alone; (4) 7-residue peptides taken from m396 or its variants, as models for denatured antibody. We describe the setup for each system (Fig. 2) below.

The protein complex of an m396 antigen-binding fragment (Fab) bound to the SARS-CoV-1 RBD was taken from the crystal structure<sup>25</sup> 2DD8. The m396 Fab consists of a heavy chain and a light chain, folded into a variable domain and a constant domain. The RBD consists of residues 321–512 in the spike protein of SARS-CoV-1. The RBD and the Fab have 4 and 5 disulfide bonds, respectively. The glycan molecule attached to Asn330 of the RBD in the crystal structure<sup>25</sup> was not included in our model. Standard protonation states at pH 7 were assigned to all the residues. In particular, all His residues were neutral, with the proton on the  $\epsilon$  nitrogen (i.e., the “HIE” type in Amber<sup>12</sup>). The protein was solvated by 40,814 water molecules. In addition, 109  $\text{Na}^+$  ions and 113  $\text{Cl}^-$  ions were included to make the system electrically neutral and to establish a bulk NaCl solution of  $\sim 150 \text{ mM}$ . The simulation system (Fig. 2a) consists of a total of 131,986 atoms. Periodic boundary conditions with a truncated octahedron unit cell were imposed in all the simulations.

The structure of the SARS-CoV-2 RBD was taken from the crystal structure<sup>29</sup> 7BZ5, which has the best resolution among all the currently available structures. We superimposed this RBD against the SARS-CoV-1 RBD<sup>25</sup> in 2DD8 to obtain a complex structure of the SARS-CoV-2 RBD and the m396 Fab. Specifically, the superimposition was based on the alignment of the  $C_{\alpha}$  atoms in residues 336–456 and 491–516 of the SARS-CoV-2 RBD to the  $C_{\alpha}$  atoms in residues 323–443 and 477–502 of the SARS-CoV-1 RBD. One missing residue in the crystal structure<sup>29</sup>, His519 of the RBD, was added back to the model based on another crystal structure<sup>30</sup>, 6M0J, of the complete SARS-CoV-2 RBD. Our simulation system (Fig. 2b) consists of residues 334–528 of the SARS-CoV-2 RBD, the m396 Fab, 41,248 water molecules, 110  $\text{Na}^+$  ions and 115  $\text{Cl}^-$  ions, with a total of 133,320 atoms. The structure (in pdb file) of this system after equilibration is provided in Supplementary Information.

The unbound antibody system was also built from the 2DD8 crystal structure<sup>25</sup>. The simulation system (Fig. 2c) consists of the m396 Fab, 22,777 water molecules, 60  $\text{Na}^+$  ions and 62  $\text{Cl}^-$  ions, with a total of 74,789 atoms. Crystal structures revealed that the conformations of the m396 variable domain are quite similar in the RBD-bound and the unbound states<sup>25</sup>. In addition, all the systems were subject to MD equilibration to allow the relaxation of the protein structure.

Each peptide system consists of the residue to be mutated as well as its 3 preceding and 3 succeeding residues in the antibody sequence, thus a total of 7 residues. The N- and C-termini of the peptide are capped by the

[–C(=O)–CH<sub>3</sub>] and the [–NH–CH<sub>3</sub>] groups, respectively. Similar to other systems, the peptide was placed in a bulk solution of ~150 mM NaCl. Figure 2d shows one example of the peptide simulation system.

For each system, we first fixed all the protein atoms and performed 500 steps of conjugate-gradient minimization for the water and ions, followed by an MD equilibration of 2 ns. We then relaxed the protein and performed a minimization of 500 steps and an equilibration of 20 ns for the entire system. All the simulations in this stage were under a constant pressure of 1 atm. We then calculated the average of the system volume in the last 10 ns of the simulation trajectory and fixed the system volume at this average. All the subsequent FEP simulations were under constant volume. The length of the truncated octahedron unit cell (see Fig. 2) is ~120 Å for the RBD/Fab complex systems and 99.1 Å for the antibody system.

**FEP simulations for m396.** In our protocols, the number of  $\lambda$  windows is  $N=24$  for charge-changing mutations and  $N=12$  for others. The  $\lambda$  values for the windows are evenly spaced between 0 and 1, including the two ends. For a mutation from the wildtype  $a$  to a new type  $b$ , the input for the calculation is the equilibrated wildtype system described in “Preparation and equilibration of wildtype m396 systems”. For charge-changing mutations, we also introduce a co-alchemical ion (see “Alchemical system for mutation of protein residues”) in the simulation system as follows. Depending on whether the wildtype  $a$  is charged or not, the co-alchemical ion is placed by taking an existing ion or by replacing a water molecule. For each ion or water molecule in the system, we calculate its shortest distance to any protein atom or their periodic images, and the ion/water with the longest distance is converted into the co-alchemical ion.

Our protocols consist of a forward calculation and a backward calculation. In the forward calculation,  $\lambda_1 = 0$  corresponds to the residue type  $a$  and  $\lambda_N = 1$  corresponds to the type  $b$ . We first carried out a “mutating” simulation, which starts at  $\lambda_1$  and equilibrates the system at each  $\lambda_i$  for 0.5 ns before changing  $\lambda$  to the next value  $\lambda_{i+1}$ . In this way, the residue type was gradually transformed from type  $a$  to type  $b$  in the mutating simulation. Snapshots in this simulation were used to initiate the sampling simulations below.

In the production runs, all the  $\lambda$  windows were sampled simultaneously, each by an individual simulation called a replica. Furthermore, Hamiltonian replica exchange<sup>11</sup> was employed to allow neighboring windows to exchange their replicas. The criterion for swapping windows  $i$  and  $i + 1$  is based on the change in the total energy due to the exchange:  $\Delta E = [E(\vec{X}_{i+1}; \lambda_i) + E(\vec{X}_i; \lambda_{i+1})] - [E(\vec{X}_i; \lambda_i) + E(\vec{X}_{i+1}; \lambda_{i+1})]$ , with the swapping probability given by<sup>11</sup>  $\min[\exp(-\Delta E/k_B T), 1]$ . The simulations were run for 5 ns per window unless noted otherwise, with the exchanges attempted every 0.4 ps. The last 4 ns of the simulation trajectories were used for free energy calculations.

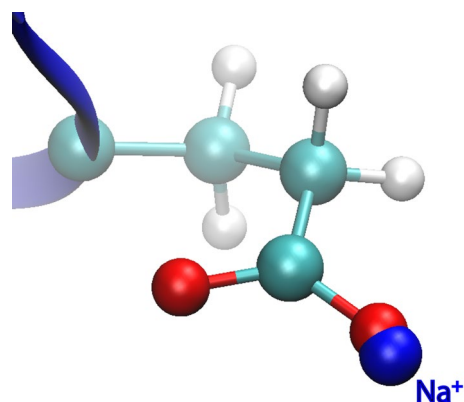
As described earlier, the mutating simulation ended with the residue in type  $b$ . We further performed an equilibration of this mutant system for 10 ns. Starting from the equilibrated mutant, we carried out a backward calculation with the residue mutated from  $b$  to  $a$ . Similar to the forward calculation above, the backward calculation involves a mutating simulation and subsequent sampling simulations with Hamiltonian replica exchange. The comparison of the forward and backward calculations serves as an important indicator for the convergence of the sampling, as will be elaborated in “Analysis and error estimation”.

**Common simulation protocols.** We adopted the Amber ff14SB force field<sup>31</sup> for the proteins and the TIP3P model<sup>32</sup> for the water molecules. All simulations were run using a time step of 2 fs. All bond lengths involving hydrogen atoms were constrained using the SHAKE algorithm. Nonbond interactions were calculated with a cutoff distance of 8 Å. Full electrostatics was calculated using the Particle Mesh Ewald (PME) method<sup>33</sup>. A constant temperature of 300 K was maintained in all the simulations using the Langevin dynamics method with a collision frequency of 5 ps<sup>-1</sup>. We used the Amber software package<sup>12</sup> to run the simulations on a computer cluster of IBM Power9 nodes equipped with NVIDIA TeslaV100 (Volta) GPUs. For technical reasons, we adopted an earlier version<sup>12</sup> (Amber18) of the simulation engine (pmemd.cuda) in this study, which worked with better stability than Amber20 on our particular hardware and encountered fewer random crashes for the FEP simulations.

**Analysis and error estimation.** At every 0.4 ps, when exchanges of the replicas were attempted, the Amber program<sup>12</sup> output the potential energies  $V(\vec{X}_i; \lambda_i)$  and  $V(\vec{X}_{i\pm 1}; \lambda_i)$  for each window. These output data allowed us to calculate the  $\Delta E_{ij}$  and  $\Delta E_{ji}$  values in Eqs. (4)–(6). Among all the mentioned techniques (“Free energy difference between two systems” and “Sampling with intermediate windows”), we adopted the BAR method (Eq. (6)) to calculate the free energy differences between adjacent windows. We then added these incremental differences to obtain the free energy difference  $\Delta A_{ab} \equiv A_b - A_a$  between the two end states.

However, it is not straightforward to directly derive the statistical errors in the BAR results. Therefore, we combined a few strategies to estimate the uncertainties. One strategy is to obtain the uncertainty from some of the alternative methods (see “Free energy difference between two systems” and “Sampling with intermediate windows”) that are easier for estimating the statistical errors therein. In addition, given that all those methods would give the same ideal  $\Delta A_{ab}$  value if the sampling were infinite, the differences between their results could reflect to a lesser extent the deviation from the ideal  $\Delta A_{ab}$  due to the finite sampling. Furthermore, a more important strategy is to estimate the uncertainty by comparing different simulations. The specifics of our error estimation are described below.

We first estimated the statistical errors in the EXP method (Eq. (4)), where the free energy difference  $\Delta A_{ij}$  is obtained from the average of  $\exp\left(-\frac{\Delta E_{ij}}{k_B T}\right)$  in window  $i$ . The variance of this average can be estimated by block averaging<sup>34</sup> on the  $\exp\left(-\frac{\Delta E_{ij}}{k_B T}\right)$  trajectory, and the variance in  $\Delta A_{ij}$  can be further obtained. The sum of all the variances for  $\Delta A_{ij}$  between adjacent windows was then taken as the variance of the end-to-end  $\Delta A_{ab}$ , and the



**Figure 3.** A snapshot from an FEP simulation with a Trp residue mutated to a Glu residue, in the window of  $\lambda = 0.65$ . The side chain of the Glu is shown, with an oxygen atom in very close contact to a  $\text{Na}^+$  ion from the bulk solution, thus exhibiting a particle collapse problem. Image was rendered by VMD<sup>43</sup> (version 1.9.3).

square root of this variance was the standard error for the calculated  $\Delta A_{ab}$ . Furthermore, if the deviation between the  $\Delta A_{ab}$  values calculated by the EXP (Eq. (4)) and the BAR (Eq. (6)) methods exceeds the standard error above, we would take this deviation as the new estimate for the standard error.

As the  $\Delta A_{ij}$  may also be calculated by the EXP method based on the sampling of window  $j$  (Eq. (5)), we followed the steps above to obtain another standard error. We then took the larger of the two standard errors as the statistical error estimated from the EXP method.

In addition,  $\Delta A_{ab}$  can be calculated by the TI method (Eqs. (8) and (9)), which offers a straightforward uncertainty estimation. In particular, the variance of each  $f_i$  may be estimated by block averaging<sup>34</sup>, and the statistical error of  $\Delta A_{ab}$  can be further obtained. We took the larger of the error estimates from the EXP and the TI methods as the standard error for the FEP sampling.

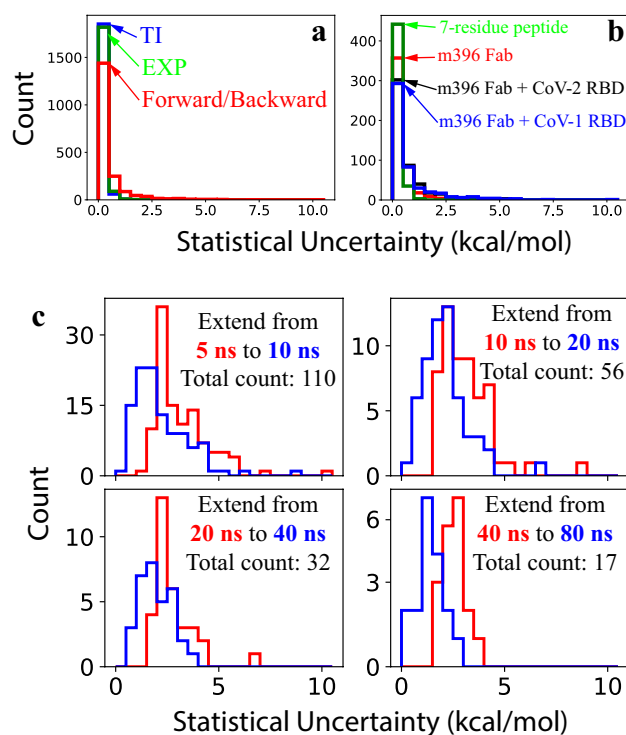
For the forward and backward FEP calculations, following the procedure above, we obtained the  $\Delta A_{ab}^{\text{forward}}$  and  $\Delta A_{ab}^{\text{backward}}$  along with their variances  $\text{var}(\Delta A_{ab}^{\text{forward}})$  and  $\text{var}(\Delta A_{ab}^{\text{backward}})$ , respectively. We always took the average  $(\Delta A_{ab}^{\text{forward}} + \Delta A_{ab}^{\text{backward}})/2$  as our final result for  $\Delta A_{ab}$ . Accordingly, the standard error for the  $\Delta A_{ab}$  would be  $\sqrt{\text{var}(\Delta A_{ab}^{\text{forward}}) + \text{var}(\Delta A_{ab}^{\text{backward}})}/2$ . However, this estimated error will be overwritten by  $|\Delta A_{ab}^{\text{forward}} - \Delta A_{ab}^{\text{backward}}|/2$  if the latter is larger. In such cases, the large deviation between the results from the forward and backward calculations will determine our statistical uncertainty.

**Protein thermal shift experiments.** We performed experiments to measure the thermal stability for a group of selected m396 variants, which can be used to validate some of our FEP predictions. Humanized recombinant antibodies were purchased from Genscript. To determine melting temperatures, each antibody was subjected to protein thermal shift assay according to the manufacturer (ThermoFisher). Briefly, each antibody was diluted to 0.1 mg/ml in PTS buffer plus 1X PTS dye. Thermal shift was performed on an ABI 7500 Fast Real-Time PCR machine (ThermoFisher) according to the manufacturer recommendation of increasing the temperature from 25 to 99 °C at 1 °C per minute. Thermal profiles were analyzed using Protein Thermal Shift Software (ThermoFisher, version 1.4, <https://www.thermoFisher.com/order/catalog/product/4466038>) using multi-peak analysis.

## Results

As described in “Methods”, we implemented automated protocols to evaluate antibody mutations using FEP and Hamiltonian replica exchange<sup>11</sup> along with a systematic uncertainty estimation. In this section, we describe the application of the protocols in our effort of optimizing the m396 antibody. In our definition, a positive  $\Delta \Delta G^{\text{Binding}}$  (see Eq. (1)) means a weaker binding affinity compared to the wildtype, and a positive  $\Delta \Delta G^{\text{Stability}}$  (see Eq. (2)) means a poorer conformational stability for the antibody mutant.

**Particle collapse problem.** Some of our preliminary calculations involving the Asp or Glu residues exhibited abnormally large statistical errors, thus indicating severe sampling problems. Visual inspection of the corresponding simulation trajectories revealed a “particle collapse problem”<sup>4</sup>, with an SC oxygen atom in the carboxylic group of Asp or Glu making abnormally close contact with a  $\text{Na}^+$  ion from the bulk solution, as illustrated in Fig. 3. This behavior could be explained by Eqs. (11) and (12) for the interaction between the two overlapping atoms. In these equations, the default Amber parameters are  $\alpha = 0.5$ ,  $\beta = 12 \text{ \AA}^2$ , and the force field parameters for the pair of oxygen and  $\text{Na}^+$  atoms are  $q_i = -0.8188 e$ ,  $q_j = 1 e$ , and  $\epsilon = 0.1355074 \text{ kcal/mol}$ . With these parameters, a complete overlap of the two atoms (i.e.,  $r_{ij} = 0$ ) in the system with  $\lambda = 0.5$  corresponds to a VDW energy (Eq. (11)) of 6.5 kcal/mol and an electrostatic energy (Eq. (12)) of  $-111.2 \text{ kcal/mol}$ , or a net interaction energy (Eq. (10)) of  $-52.3 \text{ kcal/mol}$ . In the default setting, therefore, a collapse of the two atoms into each other is energetically favorable<sup>35</sup>.



**Figure 4.** Distributions of the statistical uncertainties in the  $\Delta G$  calculations. (a) Histograms for the statistical errors estimated by three different methods, i.e., EXP, TI, and difference between the forward and backward simulation results, as described in “Analysis and error estimation”. (b) Histograms of the statistical errors in the original calculations, with 5 ns per window, for the four systems. (c) Comparisons of the uncertainty histograms after extending some of the sampling simulations from 5 to 10 ns, from 10 to 20 ns, from 20 to 40 ns, and from 40 to 80 ns.

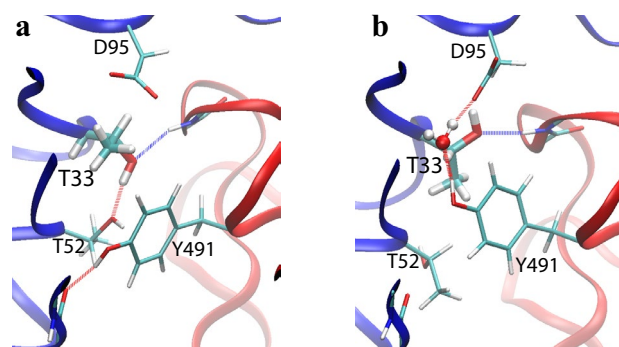
In principle, the unphysical behaviors in the intermediate  $\lambda$  windows should not affect the correctness of the calculated free energy difference between the two ends. In this case, however, ion association and dissociation have very slow kinetics, and the presence or absence of the ion has a significant effect on the energies. Consequently, it is difficult to fully equilibrate some of the intermediate  $\lambda$  windows, thus resulting in very large statistical errors.

In the SC potentials (Eqs. (11) and (12)), the parameter  $\alpha$  controls the softness of the particle core, and  $\beta$  determines the limit of the electrostatic energy at short distances. To solve the particle collapse problem, therefore, one may decrease  $\alpha$  to make the VDW energy more repulsive and thus make the atoms “harder”, or increase  $\beta$  to reduce the attractiveness of the electrostatic energy. In this study, we reduced  $\alpha$  from the default value of 0.5 to 0.3 for all the FEP simulations. The new  $\alpha$  value is also within the examined range in a previous systematic evaluation<sup>35</sup>. In addition, mutations between Asp and Glu appeared to be more vulnerable to the problem, and we thus further increased the  $\beta$  parameter from the default of 12 Å<sup>2</sup> to 20 Å<sup>2</sup> and increased the number of windows from 12 to 24 for such cases. After the adjustment of the SC parameters, the particle collapse problem no longer appeared to be significant in our simulations.

**Evaluating single mutations.** We applied our automated FEP protocols to exhaustively evaluate single mutations for 27 residues located on the m396 binding interface to the RBD. Each of the wildtype residues was mutated to all the other types except Pro and charge-reversing ones. In total, this task involved 480 single mutations from the wildtype m396. For each mutation, we calculated the  $\Delta G$  values for the complex systems of m396 bound to SARS-CoV-1 and SARS-CoV-2 RBDs, the isolated m396 system, and the 7-residue peptide system (Fig. 2).

We also performed uncertainty estimation for each  $\Delta G$  calculation. As explained in “Analysis and error estimation”, our statistical error was essentially taken as the largest among three estimates based on the EXP method, the TI method, and the comparison of forward and backward simulations, respectively. Figure 4a shows the distributions of the errors estimated by the three individual methods for all the  $\Delta G$  calculations. Whereas the EXP and TI methods provided similar statistical errors overall, the difference between results from the forward and backward FEP simulations revealed substantially larger errors in many cases. Indeed, most of the large estimated uncertainties here were due to the deviation between the forward and backward results, thus confirming that our method of comparing simulations starting from distinct initial coordinates would offer a more stringent error estimation than methods based on single simulations do<sup>36</sup>.





**Figure 5.** Snapshots of two alternative side chain conformations for the residues around Thr33 in the binding interface between the m396 heavy chain (blue) and the SARS-CoV-1 RBD (red). H-bonds between the residues are shown. In panel (b), a water molecule is also involved in the H-bond network. A proper equilibration between these alternative conformations would require long simulation times, thus resulting in slow convergence of the FEP calculations. Images were rendered by VMD<sup>43</sup> (version 1.9.3).

Figure 4b displays the distributions of the estimated standard errors for the  $\Delta G$  calculations in the four systems. Although the sampling simulations (5 ns per window) were relatively short, most of our calculations nonetheless had satisfactory statistical errors. Overall, the statistical errors here appear to be negatively correlated to the structural complexity of the system, with the simplest 7-residue peptide system having the smallest uncertainty and the antigen–antibody complex systems having the largest uncertainties.

For the  $\Delta\Delta G^{\text{Binding}}$  and  $\Delta\Delta G^{\text{Stability}}$  calculations in this study, we aimed to achieve a statistical accuracy of 2 kcal/mol or better. For the cases in which the statistical error exceeded 2 kcal/mol, therefore, we extended the sampling simulations from 5 to 10 ns per window. Subsequently, for the cases with the errors still above 2 kcal/mol in the 10-ns sampling, we further extended the simulations to 20 ns per window. Similarly, we performed extension runs for the subset of simulations with large errors to 40 ns and then to 80 ns. As shown in Fig. 4c, each extension clearly reduced the overall statistical errors in the  $\Delta G$  calculation and reduced the number of cases with unacceptable uncertainty. After all the extensions (up to 80 ns per window), the vast majority of our calculated  $\Delta\Delta G^{\text{Binding}}$  and  $\Delta\Delta G^{\text{Stability}}$  values had statistical errors below 2 kcal/mol. Calculation results for all the single mutations are provided in Supplementary Information.

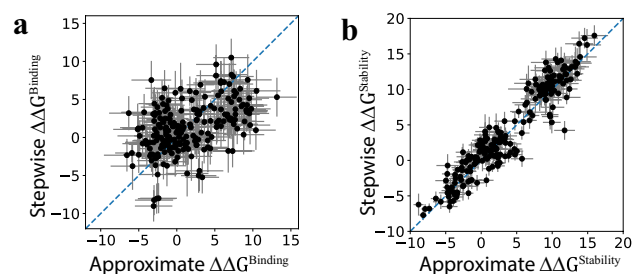
Several  $\Delta G$  calculations still had relatively large statistical errors after the extensive sampling (80 ns per window), mainly due to the presence of alternative conformations of the protein residues. Figure 5 shows such an example taken from the FEP simulations for the complex of m396 and SARS-CoV-1 RBD, with the wildtype Thr33 in the m396 heavy chain mutated to an Asp. In the simulations, Thr33 and nearby residues exhibited multiple conformations, two of which are shown in the figure. The conformations differ in the sidechain rotamer states of Thr33 and Thr51 as well as the local H-bond network. Furthermore, spontaneous transitions between these conformations were rare in the simulations, and proper equilibration of the conformations was thus difficult and would require very long simulation times. Consequently, the statistical uncertainty here was relatively high, as manifested by the deviation between the results from the forward and backward calculations. Such conformational variability is a common cause of slow convergence and large uncertainty in the FEP simulations. One potential solution to the problem would be to run a separate FEP calculation for each of the conformational states (e.g., each rotamer). However, to properly combine the results and obtain the overall  $\Delta G$  between the two end states, we also need to know the equilibrium probabilities of the alternative conformations in at least one end state. If spontaneous transitions between these conformations are too rare, other enhanced sampling techniques<sup>36</sup> could be employed to calculate such probabilities.

**Evaluating multi-point mutants.** Antibody design typically involves mutations of multiple residues from the wildtype template, and the capability of FEP to evaluate multi-point mutants is thus desirable. Given that our protocols take one mutation at a time, we adopted a multi-step approach to handle the multi-point mutants. For example, the free energy shift for a double-mutant ( $a, b$ ) relative to the wildtype (denoted as 0 below) is equal to the sum of two stepwise shifts:

$$\Delta\Delta G_{0\rightarrow a,b} = \Delta\Delta G_{0\rightarrow a} + \Delta\Delta G_{a\rightarrow a,b} = \Delta\Delta G_{0\rightarrow b} + \Delta\Delta G_{b\rightarrow a,b}, \quad (13)$$

where the first step is between the wildtype and an intermediate single-point mutant, and the second step is between the single-point mutant and the target double-point mutant. With this approach, we evaluated a given set of 221 mutants of m396, each consisting of mutations on two to twelve residues. Here we do not discuss how the mutants were proposed, and instead only focus on our FEP evaluation of these provided antibodies for their relative conformational stability and binding affinity to the SARS-CoV-2 RBD.

In principle, each  $n$ -point mutant should take  $n$  steps to evaluate. However, because different mutants may share common single mutations, a proper scheduling could maximize the number of shared intermediates and thereby reduce the total number of the steps to evaluate all the given mutants. In this study, we adopted a greedy heuristic below to minimize the total computational cost. First, similar to the Hamming distance, we defined the distance between two mutants as the number of residues that are different between the two. For example, the



**Figure 6.** Comparisons of the  $\Delta\Delta G$  values for 221 multi-point mutants obtained from two methods, as described in the text. The approximate  $\Delta\Delta G$  for the mutant is the sum of the  $\Delta\Delta G$  values for the corresponding single mutations with respect to the wildtype. The stepwise  $\Delta\Delta G$  for the mutant is calculated by adding the incremental  $\Delta\Delta G$  values when the wildtype is transformed to the mutant one residue at a time. The  $\Delta\Delta G$  is shown for both the binding to the SARS-CoV-2 RBD (a) and the stability of the mutant (b). All the units are in kcal/mol.

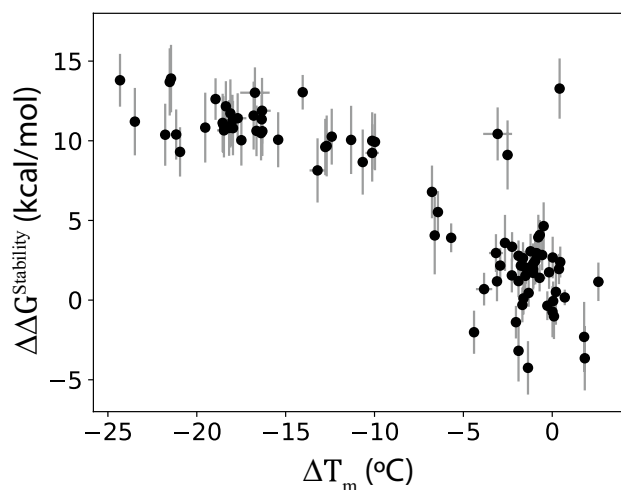
distance between the wildtype and an  $n$ -point mutant is  $n$ . We then created an “evaluation set” of mutants and defined the distance from any mutant  $x$  to the evaluation set as the shortest distance from  $x$  to any member of the set. The sum of the distances from the 221 target mutants to the evaluation set was denoted as  $L$ . Initially, the evaluation set contained the wildtype only, giving rise to  $L=1,410$  for our case here. Next, given that we already evaluated 480 single mutations (see “Evaluating single mutations”), we added the corresponding 480 single-point mutants to the evaluation set, which reduced  $L$  to 1,189. Subsequently, we added new mutants to the evaluation set one at a time. At each step, the new mutant must be derived from an existing member in the evaluation set by mutating one more residue. Among all the eligible new mutants that can be added for the current step, we chose the one that minimizes the  $L$  for the resulting new evaluation set. Such steps were repeated until  $L$  was reduced to zero, when all the targets were in the evaluation set. With this strategy, it took 554 steps to reduce  $L$  from 1,189 to 0, with each step evaluated by an individual FEP calculation. The  $\Delta\Delta G$  for any target mutant with respect to the wildtype was then obtained by summing up the  $\Delta\Delta G$  values from the individual steps.

Alternatively, by assuming additivity<sup>37</sup>, the  $\Delta\Delta G$  for a given multi-point mutant can be approximated by adding the  $\Delta\Delta G$  values for the individual single mutations from the wildtype. For example, the  $\Delta\Delta G$  for the double-mutant in Eq. (13) can be approximated as  $\Delta\Delta G_{0\rightarrow a,b} \approx \Delta\Delta G_{0\rightarrow a} + \Delta\Delta G_{0\rightarrow b}$ . With the  $\Delta\Delta G$  values available for all the single mutations, this approximation allows a direct estimate of the  $\Delta\Delta G$  for any given multi-point mutant without running additional simulations.

As shown in Fig. 6, we examined the additivity by comparing the approximate  $\Delta\Delta G$  values above to the more rigorous  $\Delta\Delta G$  obtained from the stepwise FEP calculations described earlier. Overall, for  $\Delta\Delta G^{\text{Stability}}$  (Fig. 6b), the two methods agree reasonably well, thus suggesting that the  $\Delta\Delta G^{\text{Stability}}$  values for the single mutations are roughly additive and can be used to approximately evaluate the conformational stability of a multi-point mutant. In the meantime, we caution that the stability calculations here were based on highly simplified models for the denatured state, and experimental measurements are still required to confirm our finding for the additivity of  $\Delta\Delta G^{\text{Stability}}$ . In contrast, the agreement is much poorer for the  $\Delta\Delta G^{\text{Binding}}$  (Fig. 6a), thus indicating that these residues at the binding interface may not act completely independently and that the  $\Delta\Delta G^{\text{Binding}}$  values for the single mutations may not be quantitatively predictive for their collective effects on the binding affinity.

We also experimentally measured the melting temperature ( $T_m$ ) for some of the m396 mutants above (see “Protein thermal shift experiments”). Most of the mutants exhibited two peaks in the thermal shift curve<sup>38</sup>, in which cases we took the first peak as the  $T_m$ . We performed four measurements for each mutant and took the mean and the standard error as the reported  $T_m$  and its uncertainty, respectively. All the computational results and the  $T_m$  data for the m396 multi-point mutants are provided in Supplementary Information. Figure 7 shows the change ( $\Delta T_m$ , relative to the wildtype) in the experimental melting temperature vs. the FEP-predicted  $\Delta\Delta G^{\text{Stability}}$  for each tested variant. Although simplified theories<sup>24,39</sup> predicted a linear relation between  $\Delta\Delta G^{\text{Stability}}$  and  $\Delta T_m$  for small changes (i.e., when the two are close to zero), their relation over large ranges has no known quantitative form and would be system specific. Therefore, we did not attempt to fit the data points in Fig. 7 into a mathematical (such as linear) function. Rather, some qualitative classification would be more practical for this dataset. Indeed, most of the mutants with small  $\Delta\Delta G^{\text{Stability}}$  also have small  $\Delta T_m$ , thus consistently indicating similar stabilities relative to the wildtype. Moreover, with a few exceptions, the mutants with large positive  $\Delta\Delta G^{\text{Stability}}$  tend to have much lower  $T_m$  values, thus indicating significantly reduced stability.

Given the additivity described earlier, the  $\Delta\Delta G^{\text{Stability}}$  of a multi-point mutant can be approximated by the sum of contributions from individual single mutations. This analysis allowed us to identify the mutation of Trp47 in the heavy chain to Asp as the major contributor for the large predicted  $\Delta\Delta G^{\text{Stability}}$ . Indeed, all the antibodies with this W47D mutation were predicted a highly positive  $\Delta\Delta G^{\text{Stability}}$ . However, although most of these mutants do have significantly reduced  $T_m$  values as expected, there are three clear outliers (at the upper-right corner of Fig. 7) with their  $T_m$  not very different from that of the wildtype. One possible reason for the discrepancy might be that those mutants could establish a somewhat differently folded conformation which is more probable but failed to be sampled by our MD simulations. In addition, the simplification in our models of denatured conformation could always be a source of inaccuracy. Further experiments on relevant variants are thus required to elucidate



**Figure 7.** The predicted  $\Delta\Delta G^{\text{Stability}}$  for the m396 multi-point mutants from the FEP simulations vs. the shift of their experimentally measured melting temperature relative to the wildtype m396. Some of the horizontal error bars are smaller than the size of the data marker and are thus invisible.

the underlying cause. Nonetheless, our calculations correctly identified the potentially destabilizing mutation of W47D, thus providing helpful information for the antibody design.

## Discussion

In this article, we report our implementation of automated FEP calculations to facilitate antibody design. Our protocols enable systematic large-scale predictions, such as scanning the residues in the binding interface and exhaustively evaluating all possible single mutations for these residues. A critical component in our protocols is a faithful estimation of the statistical uncertainty in the results. The uncertainty would determine the level of confidence when the individual results are considered in the decision making. Furthermore, abnormally high uncertainty could indicate pathological cases, such as the particle collapse problem identified in this study. Such alarms are especially important in the automated large-scale calculations, where manual inspections on every individual case become unfeasible. In addition, a faithful uncertainty estimation could identify the simulations that need to be improved. In this study, e.g., whereas most FEP simulations were relatively short, we extended only those simulations with large statistical errors to much longer sampling times.

Due to the stochastic nature of the sampling, no method guarantees correct uncertainty estimation for all cases<sup>36</sup>. For example, although most of the extended samplings reported in this study reduced the estimated uncertainty, the error estimates for a small fraction of the simulations became even larger after the extension. Therefore, it is entirely likely that among the simulations that were not extended, the uncertainty of some may have been underestimated and would have been revealed in extension runs. As another example, in our evaluation of the multi-point mutants (“Evaluating multi-point mutants”), each mutant is connected to the wildtype by a single path of intermediates, whereas establishing multiple paths could improve the uncertainty estimation by exploiting the thermodynamic consistency<sup>2</sup>. In general, longer and more simulations should always make the results more reliable. In practical applications, however, the increased computational cost associated with more samplings has to be considered as well. In this study, by allocating the computational resource to the cases with the most needs, most of our calculations appeared to achieve acceptable statistical uncertainties at a modest aggregated simulation time. Our approach thus strikes a good balance between the statistical accuracy and the computational cost.

An important strategy in our uncertainty estimation is to compare the results from two groups of FEP simulations starting from different initial coordinates, namely, a forward calculation from residue *a* to *b* and a backward calculation from *b* to *a*. In contrast to estimating errors from single simulations alone, this approach could often capture statistical uncertainties (see Fig. 4a) caused by slow equilibrations<sup>36</sup>, such as the case shown in Fig. 5 with multiple alternative conformations of the protein residues.

In the context of rapid large-scale screening here, we considered it acceptable if the statistical uncertainty is below 2 kcal/mol. In the meantime, we note that with sufficient sampling, FEP can often achieve a higher accuracy, e.g., 1 kcal/mol or even better. If an improved statistical accuracy is needed, however, we may always further extend the FEP sampling to reduce the uncertainty, as similarly done in this study. Furthermore, as experimental binding affinities are not available yet at this stage, we do not have a direct validation for our predicted  $\Delta\Delta G^{\text{Binding}}$  values and do not know their actual errors.

In addition to the relative binding affinity, our protocols also predict the relative conformational stability for the antibody mutant. Protein stability describes the equilibrium between its native (folded) and denatured (unfolded) conformations. Given that the denatured structure is not known, one needs to resort to highly simplified models to approximate the local environment of the mutated residue in the denatured conformation. In many FEP works, e.g., the denatured state was represented by just a single amino acid (i.e., the residue under

the mutation) in bulk solution. Recent studies<sup>18</sup> demonstrated that the accuracy of the predicted stability can be improved if a few flanking residues are also included in the model for the denatured state, as adopted in this study. Whereas our peptide system remains a crude model for the denatured protein, our predicted free energy shift for the measured m396 mutants appears to be consistent with the experimental melting temperature for the majority of the test cases. Furthermore, predictions of both the affinity and the stability by our FEP protocols would allow one to jointly consider these factors in the early stage of antibody design, thus helping develop antibodies with good manufacturability properties that will become important for the clinical applications at a later stage.

As FEP has been widely adopted in many applications, there are multiple variations in its implementation. In the following, we will briefly discuss the tradeoffs for some of the alternative implementations and suggest potential improvement for future applications.

First, as the case for all MD-based methods, the accuracy of FEP heavily relies on the faithfulness of the simulation system, especially the quality of the protein structures. In addition, the protonation states of the protein residues may also have a significant impact on the simulation. In this study, we adopted the simplest approach of assigning standard protonation state for all the residues. More careful treatment of the protonation states, such as by PROPKA<sup>40</sup> or other programs, should in principle improve the accuracy of our FEP calculations.

As mentioned earlier, longer and more simulations should always improve the statistical uncertainty but will increase the computational cost. Therefore, alternative implementations should be compared based on the same computational cost determined by the total simulation length. For example, one could repeat the same sampling simulation multiple times with different initial velocities<sup>41</sup>, or alternatively extend the sampling multiple times longer. Here we would argue that a longer simulation is often preferred over multiple short runs. The essence of equilibration is to “forget” the initial microstate during the simulation, such that the statistics become completely independent of the particular starting coordinates<sup>36</sup>. In this aspect, a long simulation should more likely forget the initial condition than multiple short simulations do, especially when involving a relaxation with nontrivial first passage times to some more probable structures. As demonstrated in this study, the discrepancy between our forward and backward simulations became increasingly smaller as the simulations were extended longer, and it is doubtful that the same could be achieved by repeatedly running those simulations but with a fraction of the sampling time for each run. On the other hand, when equilibration is fast and not of a concern, running multiple shorter simulations has the advantage of parallel execution and thus possibly a faster turnaround.

A related issue concerns the number and the placement of the  $\lambda$  windows in FEP. In this study, we adopted a simple recipe<sup>5,6</sup> of 12 or 24 windows depending on whether the mutation is charge-changing or not. However, we found that with the same number of windows, calculations involving bulky residues (such as Trp) normally have substantially higher statistical uncertainties than others, thus suggesting that it would be beneficial if the number of  $\lambda$  windows is determined according to the residue type, with more windows allocated to larger residues. Furthermore, with Hamiltonian replica exchange implemented, the exchange rate between adjacent windows could conveniently indicate whether the windows have sufficient overlap. Indeed, the accuracy of the calculated  $\Delta A$  (see “Free energy difference between two systems” and “Sampling with intermediate windows”) between two neighboring windows appears to be correlated to their exchange rate. Moreover, we observed that the exchange rates for the windows near the ends (i.e., when  $\lambda$  is near 0 or 1) tend to be substantially lower than those in the midrange, especially after our adjustment of the  $\alpha$  and  $\beta$  parameters (“Particle collapse problem”). Therefore, instead of uniformly placing the  $\lambda$  windows as in this study, the sampling efficiency may be improved by allocating denser windows near the ends and sparser windows in the midrange.

Our FEP implementation in this study adopted the one-step transformation where the electrostatic and VDW potentials vary simultaneously (Eqs. (11) and (12)) with  $\lambda$ . In contrast, multi-step transformation is an alternative scheme in which the SC atoms are decharged prior to the transformation of their VDW interactions<sup>35,41</sup>. Although technically the one-step transformation is slightly simpler, the multi-step transformation has the advantage of eliminating the particle collapse problem encountered here.

The double-system/single-box setup<sup>42</sup> is an alternative simulation setup for FEP, in which two entities (such as one antibody and one peptide) are placed far apart from each other in the same simulation system. Then in the FEP simulations, the alchemical transformation from  $a$  to  $b$  in one entity occurs simultaneously with the transformation from  $b$  to  $a$  in the other entity. Despite a much larger system size and higher atom count for the simulations, this setup could obtain the  $\Delta\Delta G$  in a single FEP calculation, whereas in the traditional setup the  $\Delta\Delta G$  is the difference between two  $\Delta G$ s (Eqs. (1) and (2)), each from a separate FEP calculation. For charge-changing mutations, the double-system/single-box setup has a unique advantage of simultaneously satisfying the charge neutrality in both ends, thus making the co-alchemical ion unnecessary. In this study, we still adopted the traditional setup, partly due to its technical simplicity. In addition, as mutation on the isolated antibody is required for determining multiple  $\Delta\Delta G$ s here, it is more convenient to obtain the  $\Delta G^{\text{Antibody}}$  (Eqs. (1) and (2)) in just one FEP calculation.

In this article, we mainly focus on the FEP technique, using m396 as a case study. As mentioned in “Introduction”, our FEP implementation is a component in a much broader multidisciplinary effort of antibody design. Our project involves several other antibodies, a diverse set of computational approaches for proposing and evaluating antibody mutants, and extensive experiments. The other components of the project will be presented in forthcoming reports.

In summary, given the capability of FEP to rigorously predict relative free energies, our work here could help transform this powerful technique into a reliable and efficient tool that can be routinely applied to aid antibody design and to elucidate protein–protein interactions.

## Data availability

Computational and experimental data are provided in Supplementary Information. Relevant code (scripts) may be requested from the corresponding author.

Received: 17 December 2021; Accepted: 7 June 2022

Published online: 21 July 2022

## References

- Beveridge, D. L. & DiCapua, F. M. Free energy via molecular simulation: Applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Biophys. Chem.* **18**, 431–492 (1989).
- Wang, L. *et al.* Modeling local structural rearrangements using FEP/REST: Application to relative binding affinity predictions of CDK2 inhibitors. *J. Chem. Theory Comput.* **9**, 1282–1293 (2013).
- Abel, R. *et al.* Advancing drug discovery through enhanced free energy calculations. *Acc. Chem. Res.* **50**, 1625–1632 (2017).
- Lee, T. S. *et al.* Alchemical binding free energy calculations in AMBER20: Advances and best practices for drug discovery. *J. Chem. Inf. Model.* **60**, 5595–5623 (2020).
- Clark, A. J. *et al.* 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 (2017).
- Clark, A. J. *et al.* Relative binding affinity prediction of charge-changing sequence mutations with FEP in protein–protein interfaces. *J. Mol. Biol.* **431**, 1481–1493 (2019).
- Yamashita, T. Toward rational antibody design: Recent advancements in molecular dynamics simulations. *Int. Immunol.* **30**, 133–140 (2018).
- Sormanni, P., Aprile, F. A. & Vendruscolo, M. Third generation antibody discovery methods: In silico rational design. *Chem. Soc. Rev.* **47**, 9137–9157 (2018).
- Norman, R. A. *et al.* Computational approaches to therapeutic antibody design: Established methods and emerging trends. *Brief Bioinform.* **21**, 1549–1567 (2020).
- Lau, E. Y. *et al.* Discovery of small-molecule inhibitors of SARS-CoV-2 proteins using a computational and experimental pipeline. *Front. Mol. Biosci.* **8**, 678701 (2021).
- Fukunishi, H., Watanabe, O. & Takada, S. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *J. Chem. Phys.* **116**, 9058–9067 (2002).
- Case, D. A. *et al.* AMBER (University of California, 2018).
- Jain, T. *et al.* Biophysical properties of the clinical-stage antibody landscape. *Proc. Natl. Acad. Sci. USA* **114**, 944–949 (2017).
- Kuroda, D. & Tsumoto, K. Engineering stability, viscosity, and immunogenicity of antibodies by computational design. *J. Pharm. Sci.* **109**, 1631–1651 (2020).
- Gapsys, V., Michielssens, S., Seeliger, D. & de Groot, B. L. Accurate and rigorous prediction of the changes in protein free energies in a large-scale mutation scan. *Angew. Chem. Int. Ed. Engl.* **55**, 7364–7368 (2016).
- Steinbrecher, T. *et al.* 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 (2017).
- Ford, M. C. & Babaoglu, K. Examining the feasibility of using free energy perturbation (FEP+) in predicting protein stability. *J. Chem. Inf. Model.* **57**, 1276–1285 (2017).
- Duan, J., Lupyan, D. & Wang, L. Improving the accuracy of protein thermostability predictions for single point mutations. *Biophys. J.* **119**, 115–127 (2020).
- Olson, M. A. *et al.* Can template-based protein models guide the design of sequence fitness for enhanced thermal stability of single domain antibodies?. *Protein Eng. Des. Sel.* **28**, 395–402 (2015).
- Zeiske, T., Stafford, K. A. & Palmer, A. G. 3rd. Thermostability of enzymes from molecular dynamics simulations. *J. Chem. Theory Comput.* **12**, 2489–2492 (2016).
- Goldenzweig, A. & Fleishman, S. J. Principles of protein stability and their application in computational design. *Annu. Rev. Biochem.* **87**, 105–129 (2018).
- Pucci, F., Bernaerts, K. V., Kwasigroch, J. M. & Rooman, M. Quantification of biases in predictions of protein stability changes upon mutations. *Bioinformatics* **34**, 3659–3665 (2018).
- Bekker, G. J., Ma, B. & Kamiya, N. Thermal stability of single-domain antibodies estimated by molecular dynamics simulations. *Protein Sci.* **28**, 429–438 (2019).
- Huang, P. *et al.* Evaluating protein engineering thermostability prediction tools using an independently generated dataset. *ACS Omega* **5**, 6487–6493 (2020).
- Prabakaran, P. *et al.* Structure of severe acute respiratory syndrome coronavirus receptor-binding domain complexed with neutralizing antibody. *J. Biol. Chem.* **281**, 15829–15836 (2006).
- Wrapp, D. *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260–1263 (2020).
- Zwanzig, R. W. High-temperature equation of state by a perturbation method. 1. Nonpolar gases. *J. Chem. Phys.* **22**, 1420–1426 (1954).
- Bennett, C. H. Efficient estimation of free-energy differences from monte-carlo data. *J. Comput. Phys.* **22**, 245–268 (1976).
- Wu, Y. *et al.* A noncompeting pair of human neutralizing antibodies block COVID-19 virus binding to its receptor ACE2. *Science* **368**, 1274–1278 (2020).
- Lan, J. *et al.* Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 215–220 (2020).
- Maier, J. A. *et al.* ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).
- Jorgensen, W. L. *et al.* Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
- Darden, T., York, D. & Pedersen, L. Particle mesh Ewald—an N.Log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
- Flyvbjerg, H. & Petersen, H. G. Error-estimates on averages of correlated data. *J. Chem. Phys.* **91**, 461–466 (1989).
- Steinbrecher, T., Joung, I. & Case, D. A. Soft-core potentials in thermodynamic integration: Comparing one- and two-step transformations. *J. Comput. Chem.* **32**, 3253–3263 (2011).
- Meshkin, H. & Zhu, F. Toward convergence in free energy calculations for protein conformational changes: A case study on the thin gate of Mhp1 transporter. *J. Chem. Theory Comput.* **17**, 6583–6596 (2021).
- Jemimah, S. & Gromiha, M. M. Exploring additivity effects of double mutations on the binding affinity of protein–protein complexes. *Proteins* **86**, 536–547 (2018).
- King, A. C. *et al.* High-throughput measurement, correlation analysis, and machine-learning predictions for pH and thermal stabilities of Pfizer-generated antibodies. *Protein Sci.* **20**, 1546–1557 (2011).
- Becktel, W. J. & Schellman, J. A. Protein stability curves. *Biopolymers* **26**, 1859–1877 (1987).
- Olsson, M. H., Sondergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: Consistent treatment of internal and surface residues in empirical pKa predictions. *J. Chem. Theory Comput.* **7**, 525–537 (2011).
- Jespers, W. *et al.* QresFEP: An automated protocol for free energy calculations of protein mutations in Q. *J. Chem. Theory Comput.* **15**, 5461–5473 (2019).
- Aldeghi, M., de Groot, B. L. & Gapsys, V. Accurate calculation of free energy changes upon amino acid mutation. In *Computational Methods in Protein Evolution* Vol. 1851 (ed. Sikosek, T.) (Humana Press, 2018).
- Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).

## Acknowledgements

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory (LLNL) under contract DE-AC52-07NA27344 and was supported by Joint Program Executive Office (JPEO) agreement number 11647302, Defense Advanced Research Projects Agency (DARPA) agreement number HR0011154580, and Laboratory Directed Research and Development programs (20-ERD-032 and 20-ERD-064) at LLNL. We would like to thank C. Earnhart, A. Jenkins, S. Hopkins, S. Sundaram and J. Brase for helpful discussions and support. The computational resources were provided by Livermore Computing at LLNL.

## Author contributions

F.Z. performed the computational works in this report. F.A.B. performed the experiments. F.Z. wrote the manuscript with help from F.A.B, E.Y.L., K.T.A., T.A.D., W.F.D.B., B.W.S., A.T.Z., and D.M.F.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-14443-z>.

**Correspondence** and requests for materials should be addressed to F.Z. or D.M.F.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022