JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC

Article

Improving pK_a Predictions with Reparameterized Force Fields and Free Energy Calculations

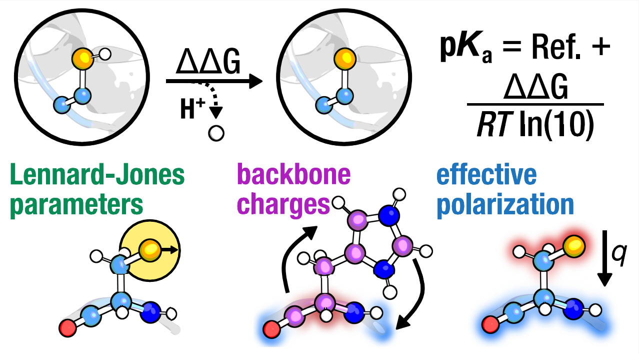
Carter J. Wilson, Vytautas Gapsys,* and Bert L. de Groot*

Cite This: *J. Chem. Theory Comput.* 2025, 21, 4095−4106

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Given the growing interest in designing targeted covalent inhibitors, methods for rapidly and accurately probing pK_as—and, by extension, the reactivities—of target cysteines are highly desirable. Complementary to cysteine, histidine is similarly relevant due to its frequent presence in protein active sites and its unique ability to exist in two tautomeric states. Here, we demonstrate that nonequilibrium free energy calculations can accurately determine the pK_a values of both residues, often outperforming conventional predictors. Importantly, we find that (1) increasing the van der Waals radius of cysteine's sulfur atom, (2) modifying the backbone charges of histidine, and (3) introducing effective polarization by downscaling the side chain partial charges of both residues can all significantly improve pK_a prediction accuracy. Using the modified CHARMM36m force field on the full dataset reduces the prediction error from 2.12 ± 0.27 pK to 1.28 ± 0.15 pK and increases the correlation with experiment from 0.25 ± 0.09 to 0.58 ± 0.08. Similarly, using the modified Amber14SB force field decreases the error from 3.21 ± 0.29 pK to 1.69 ± 0.23 pK and improves the correlation from 0.15 ± 0.10 to 0.36 ± 0.10.



■ INTRODUCTION

Cysteine is unique among the 20 proteogenic amino acids due to the large atomic radius of its sulfur atom and the relative weakness of the corresponding S−H bond. This imbues it with remarkable nucleophilicity, facilitating spontaneous reactions even under mild conditions.[1] The inherent nucleophilicity of a given cysteine is governed by its pK_a, a value that implies the favorability of the ionization state of the thiol. Solvent-exposed cysteines have values near 8,[2,3] while buried cysteines or those located in a unique protein microenvironment can range from 3 to 12.[4,5] Given their variable pK_a values and unique properties, cysteine residues play various functional roles in redox and nucleophilic catalysis,[6] metal binding,[7] environmental sensing,[8] and structural formation.[9,10]

In recent years, interest has grown in targeting cysteine residues with covalent inhibitors.[11] To overcome poor target selectivity and drug resistance, an electrophilic warhead moiety may be incorporated into a reversible submicromolar inhibitor to covalently bind a nucleophilic residue: this modification can dramatically increase therapeutic potency.[12,13] Members of this class of reactive molecules are commonly referred to as targeted covalent inhibitors (TCIs) and predicting their affinity and reversibility is particularly desirable.[14]

A key step in a TCI binding and reaction landscape is the deprotonation of the cysteine thiol and the formation of the nucleophilic thiolate. Experimental exchange-rate studies have shown that the equilibrium between the protonated and deprotonated states of solvent-exposed cysteine side chains is fast[15] (i.e., $10^{12}\cdot\ M^{-1}s^{-1}$) and that the protonation rate and pK_a are well correlated.[15,16] That is to say, the pK_a of a particular cysteine provides the relevant information about the energy required to form the nucleophilic thiolate and, by extension, the propensity for covalent modification.

Experimental methods for determining the pK_a value of a cysteine can involve kinetic assays, spectrophotometric titrations, or NMR spectroscopy; however, in a purely computational *in silico* screen of potential covalent modifiers, the ability to rapidly and accurately probe the reactivity of a target cysteine under various conditions is highly desirable. Theoretical approaches motivated by the thermodynamic cycle given in Figure 1, present a compelling alternative to experiment and can often be seamlessly integrated alongside existing computational free energy workflows.

Here, we consider a cysteine residue in the protein and a capped model peptide (i.e., ACE-Ala-Cys-Ala-NH$_2$) in both vacuum and water. We have the reference pK_a^o = 8.55 and as

Received: January 9, 2025
Revised: February 23, 2025
Accepted: February 25, 2025
Published: April 2, 2025

https://doi.org/10.1021/acs.jctc.5c00031

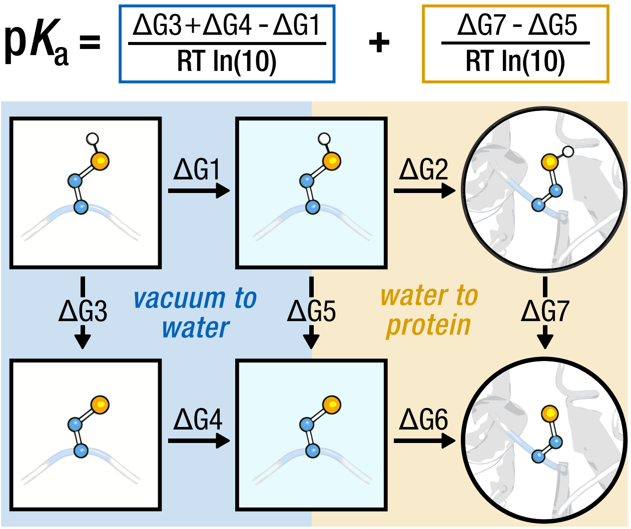

pKa = (ΔG3+ΔG4 - ΔG1) / (RT ln(10)) + (ΔG7 - ΔG5) / (RT ln(10))
ΔG1
ΔG2
ΔG3
vacuum to water
ΔG5
water to protein
ΔG7
ΔG4
ΔG6

Figure 1. Complete p$K_a$thermodynamic cycle. The horizontal arrows mark the transfer of a titratable residue between different environments: vacuum (left), water (middle), protein (right). The vertical arrows denote the free energy difference between the deprotonated and protonated form in a corresponding environment. For cysteine, the free energy associated with proton transfer from vacuum to water is known. By using this reference pK_a^o we need only consider the rightmost cycle: the free energy difference of a deprotonation event in water and in protein.

such can neglect the free energy of moving a (de)protonated residue from vacuum to water (Figure 1, left cycle). To determine the pK_a we then need only consider the free energy difference of a deprotonation event in water and in the protein (Figure 1, right cycle).

Recently we demonstrated that atomistic molecular dynamics simulations paired with nonequilibrium alchemical free energy calculations are capable of accurately resolving the pK_a values of aspartate, glutamate, and lysine.[17] Here, we assess the ability of our pmx-based, nonequilibrium switching (NES) approach to calculate the pK_a values of 40 cysteines and 22 histidines across a range of wildtype and mutant proteins. We compare our results to two conventional predictor methods and the previously reported performance of three MD-based approaches, replica-exchange thermodynamic integration, constant-pH molecular dynamics, and free energy perturbation. Taken as a whole, our results demonstrate that MD-approaches, including NES, provide pK_a prediction accuracy comparable to conventional predictors; however, we also find that this accuracy can be increased well above conventional methods by employing parameters that are refit to more accurately reproduce QM and experimental observables, i.e., (1) rescaling the vdW radii of cysteine sulfur thiolate and (2) altering the backbone charges of histidine in Amber14SB, and (3) charge-scaling CHARMM36m and Amber14SB. Using the triple-modified Amber14SB force field we achieve an average unsigned error of 2.37 ± 0.29 pK for cysteine and 0.50 ± 0.10 pK for histidine, whereas using charge-scaled CHARMM36m, we achieve errors of 1.61 ± 0.21 pK for cysteine and 0.71 ± 0.16 pK for histidine.

METHODOLOGY

Cysteine Analogues: QM Simulations. *Ab initio* molecular dynamics (AIMD) simulations were performed within the Born–Oppenheimer approximation; where the electronic structure of the system is solved using the Gaussian plane-wave (GPW) approach to DFT,[18,19] implemented in the QUICKSTEP[20] subroutine of CP2K.[21] We used the standard LIBXC library[22] for the exchange and correlation of the revPBE functional[23,24] and applied Grimme's DFT-D3 dispersion corrections with zero-damping.[25] We use the Goedecker–Teter–Hutter pseudopotentials[26,27] optimized for PBE to represent the core electrons and the TZV2P basis set. Simulations consisted of an initial 10 ps equilibration, followed by a production run in the NVT ensemble for another 200 ps. The temperature was maintained at 298 K by a massive Nose-Hoover chain thermostat with a time constant of 3 ps. We set an energy convergence threshold of 10^{-10} Ha and a convergence tolerance for the SCF cycle of 10^{-6} Ha. Because the *ab initio* simulations were to be performed in the NVT ensemble, achieving an accurate initial box volume was important. To this end, we performed classical molecular dynamics implemented in GROMACS with the OPC water model to produce an initial system configuration. OPC was chosen because it more accurately reproduces the relevant bulk properties of water compared to conventional 3-point models (i.e., TIP3P).[28] The result of these classical simulations was a final cubic box size of L = 15 Å, containing 109 water molecules and a methylthiolate molecule.

Cysteine Analogues: MD Simulations. Classical MD simulations were performed in three types of boxes: one identical to that used in AIMD simulations (i.e., L = 15 Å) and two larger boxes: L = 30 Å and L = 60 Å (Figure S1). We found no significant differences in the solvation structure when comparing the 15 and 30 Å boxes (Figure S3a). This observation was independent of the cutoff used for the simulation of the larger box (Figure S4).

Solvation free energy calculations of the charged methylthiolate molecule were performed in both a 30 and 60 Å box. Importantly, we found no change in the calculated free energy between the two box sizes (Figure S5a) which suggests the 30 Å box is sufficiently large to minimize the non-neutral simulation cell artifact associated with the calculation.

GROMACS 2023[29] was used to run all simulations. Simulations were carried out in the NVT ensemble with a constant temperature of 298 K, maintained using a Nosé–Hoover thermostat with 3 ps coupling time. In the 30 and 60 Å boxes, long-range electrostatic interactions were calculated using the Particle-Mesh Ewald method[30] with a real-space cutoff of 1.2 nm and grid spacing of 0.12 nm with CHARMM and a real-space cutoff of 1.0 nm and grid spacing of 0.125 nm with Amber. For CHARMM the Lennard-Jones interactions were force-switched off between 1.0 and 1.2 nm, while for Amber, a cutoff at 1.0 nm was used and a dispersion correction was applied to the energy and pressure.

In the 15 Å box, electrostatic interaction cutoffs were 0.7 nm with a 0.12 nm spacing with CHARMM and 0.7 nm with a 0.125 nm spacing with Amber. For CHARMM the Lennard-Jones interactions were force switched off between 0.5 and 0.7 nm, while for Amber they were cut off at 0.7 nm and a dispersion correction was applied to the energy and pressure.

We used the CHARMM TIP3P water model with the nonzero Lennard-Jones parameters on hydrogen atoms (i.e., mTIP3P)[31] and plain TIP3P[32] for the CHARMM and Amber systems, respectively. In the case of CHARMM we also assessed the impact of using the OPC water model on the solvation structure and measured free energies.

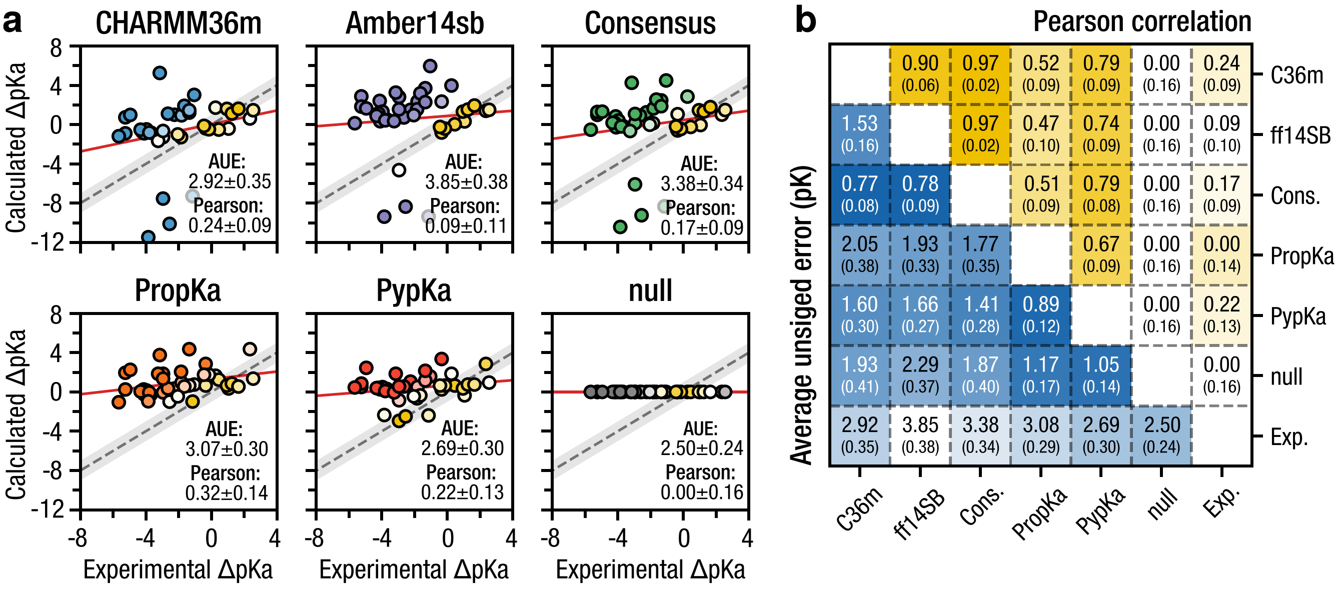


Figure 2. Overall performance: original force fields and predictors. (a) Correlation between the calculated and experimental cysteine pK_a values. Marker color indicates deviation from experiment where yellow indicates a minimum AUE from experiment. Regression lines are shown in red, and the gray error band represents a 1 pK unit deviation from experiment. Consensus is the average of CHARMM36m and Amber14SB. (b) Pearson correlations (upper right triangle) and AUEs (lower left triangle) between ΔpK_a estimates were calculated for each method over the entire data set. Comparison with experiment means that the bottom row and rightmost column correspond to the overall performance.

Production simulations were 50 ns and the first 10 ns of simulation were discarded as equilibration. From the remaining 40 ns: (1) radial distribution functions were computed; and (2) 200 nonequilibrium transitions of 20 ps were generated. Free energies were computed as described in the final paragraph of the following section.

GROMACS/pmx: Nonequilibrium Alchemy. pmx[33] was used for system setup, hybrid topology generation, and analysis. Initial protein structures were taken from the PDB database and mutations were introduced using pdbfixer. In total we consider 12 proteins and 40 cysteine residues and 22 histidines (Table S1). A double system in a single box setup was used, with a 3 nm distance between the protein and peptide (ACE-Ala-X-Ala-NH_2); this ensured a neutral box at every step of the alchemical transformation. To ensure that the protein and peptide did not interact, a single Cα in each molecule was positionally restrained. We used the CHARMM36m[34] (with mTIP3P[31]) and Amber14SB[35] (with TIP3P[32]) force fields, both having previously performed well for simulations involving nonequilibrium alchemical calculations.

GROMACS 2023 was used to run all simulations. For all systems, an initial minimization was performed using the steepest descent algorithm. A constant temperature corresponding to the reference experimental setup was maintained implicitly using the leapfrog stochastic dynamics integrator[36,37] with a friction constant of γ = 0.5 ps^{-1}. The pressure was maintained at 1 bar using the Parrinello−Rahman barostat[38] with a coupling time constant of 5 ps. The integration time step was set to 2 fs. Long-range electrostatic interactions were calculated using the Particle-Mesh Ewald method[30] with a real-space cutoff of 1.2 nm and grid spacing of 0.12 nm. Lennard-Jones interactions were force-switched off between 1.0 and 1.2 nm. Bonds to hydrogen atoms were constrained using the Parallel LINear Constraint Solver.[39]

Production simulations were 50 ns in length and run in quadruplicate. The first 10 ns of simulation was discarded as equilibration and from the remaining 40 ns, 400 nonequilibrium transitions of 200 ps were generated. Work values from the forward and backward transitions were collected using thermodynamic integration and these were used to estimate the corresponding free energy with Bennett's acceptance ratio[40] as a maximum likelihood estimator relying on the Crooks Fluctuation Theorem.[41] Bootstrapping was used to estimate the uncertainties of the free energy estimates, and these were propagated when calculating ΔΔG values.

As shown in Figure 1, we convert between the ΔΔG of deprotonation and the protein pK_a via:

$$\mathrm{p}K_a = \mathrm{p}K_a^{\circ} + \frac{\Delta\Delta G}{RT\ \ln(10)} \tag{1}$$

using the experimentally reported temperature and corresponding reference values for cysteine $\mathrm{p}K_a^{\circ} = 8.55$ and histidine $\mathrm{p}K_a^{\circ} = 6.54$.[3]

Conventional Predictors. Based on popularity and previously documented cysteine pK_a prediction performance[42] we considered only two methods: PropK_a (v3.4)[43] and PypK_a (v2.9.4).[44] PropK_a is an empirical predictor where the ΔG contributions are described by charge−charge, desolvation, and hydrogen-bonding interactions. Default settings were used when performing the calculation. PypK_a uses Monte Carlo simulations to probe the various side chain states and employs DelPhi[45] to resolve the PBE. Default settings were used, except for the salt concentration, which was set according to the experimental setup.

Recently, Molecular Operating Environment (MOE) was used to calculate the pK_a values of a large cysteine residue data set.[42] We compare NES with this method on the overlapping 20 residue data set.

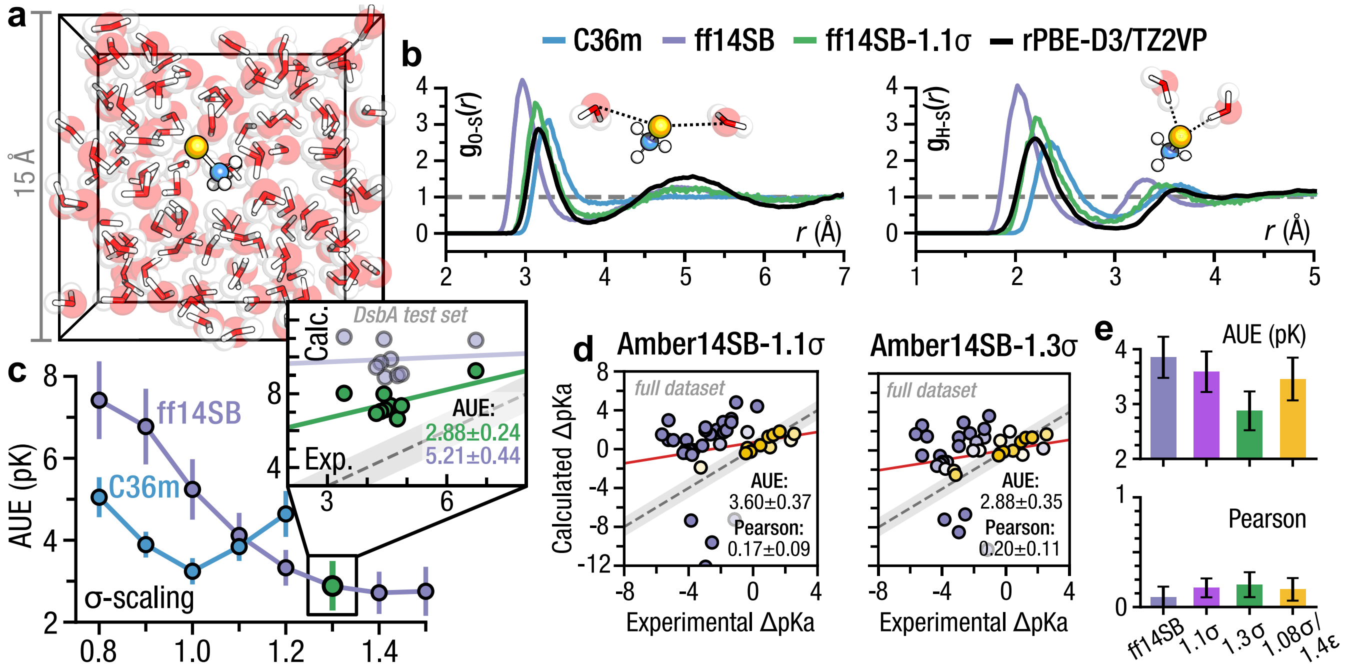


Figure 3. Thiolate reparameterization: Amber14SB-1.3σ. (a) *Ab initio* simulation setup with methylthiolate and ≈100 water molecules. (b) Radial distribution function between the methylthiolate sulfur and water oxygen (left) and methylthiolate sulfur and water hydrogen (right). The *ab initio* distribution is shown as a black line. (c) pK_a prediction performance for CHARMM36m and Amber14SB on the DsbA test set as a function of σ-value scaling. Amber14SB-1.3σ performance is marked in green. (c, inset) Correlation between the calculated and experimental pK_a values on the DsbA test set, with regression lines indicated. The gray error band represents a 1 pK unit deviation from experiment. (d) Correlation between the calculated and experimental cysteine pK_a values. Marker color indicates deviation from experiment where yellow indicates a minimum AUE from experiment. Regression lines are shown in red, and the gray error band represents a 1 pK unit deviation from experiment. (e) Prediction performance across the full data set, comparing the scaled Amber14SB force fields.

RESULTS

Overall Performance: Original Force Fields. Double free energy differences ($\Delta\Delta G$) were calculated for a set of 40 residues, allowing us to robustly evaluate performance on a large data set.

Figure 2 summarizes the main findings: our NES approach performs comparably to *in silico* predictors, with CHARMM36m yielding an average unsigned error (AUE) of 2.92 ± 0.35 pK as compared to 3.09 ± 0.29 pK and 2.71 ± 0.29 pK for PropK_a and PypK_a, respectively. This performance was also reflected in the Pearson correlation, which was 0.24 ± 0.09 with CHARMM36m compared to 0.31 ± 0.14 and 0.22 ± 0.12 with PropK_a and PypK_a, respectively. On the 20 residue subset evaluated by MOE,[42] MOE exhibited an improved accuracy of 1.79 ± 0.24 pK as compared to 2.49 ± 0.46 pK and 2.95 ± 0.45 pK for CHARMM36m and Amber14SB, respectively. With respect to accuracy, no method significantly exceeds a null predictor, which assumes $\Delta pK_a = 0$.

Previous work has illustrated that an accurate determination of the pK_a may require accounting for residue coupling.[46] With respect to cysteine residues, often found at enzyme active sites, the relevance of coupling is expected to become even more pronounced. Elsewhere, we introduced a coupling formalism that improved pK_a prediction accuracy;[46] here, we apply this approach to 12 cysteine residues found near other titratable groups. Consistent with previous work, the pK_a values of coupled residues were predicted with lower accuracy, when the coupling was not explicitly accounted for. Accounting for coupling, however, could in part remedy this (Figure S2). The observed improvement was less pronounced for Amber14SB (AUE without coupling: 3.85 ± 0.38 pK, AUE with coupling: 3.53 ± 0.37 pK) compared to CHARMM36m (AUE without coupling: 2.92 ± 0.35 pK, AUE with coupling: 2.28 ± 0.28 pK). On the 20 residue subset evaluated with MOE, accounting for coupling was even more pronounced for CHARMM36m, shifting the accuracy from 2.49 ± 0.46 pK to 1.53 ± 0.29 pK.

Previous work has assessed the ability of different MD-based approaches to predict cysteine pK_a values[42,47] we can compare our performance on the overlapping data sets. For 18 cysteine residues, Awoonor-Williams and Rowley,[47] found a thermodynamic integration, replica-exchange scheme with the CHARMM36 force field gave an AUE of 1.67 ± 0.40 pK (compared to 1.64 ± 0.40 pK with CHARMM36m and NES). More recently, Awoonor-Williams and coworkers[42] found that on 25 residues, a Monte Carlo, constant-pH approach paired with CHARMM36m gave an AUE of 2.42 ± 0.36 pK (compared to 1.70 ± 0.34 pK with CHARMM36m and NES).

Consistent with previous work, CHARMM36m performed significantly better than plain Amber14SB (Figure 2); this large discrepancy led us to investigate the underlying parameterization differences.

Thiolate Reparameterization: Amber14SB-1.3σ. In the Amber family of force fields, both the thiol and thiolate sulfur atoms share the same atom type, while the partial charge assignments between the two residues differ. Thiolate sulfur has a more diffuse electron density and a larger ionic radius, characteristics that will be reflected in the Lennard-Jones parameters, particularly the σ-value.

Simulations of methylthiolate using both classical molecular dynamics with Amber14SB and CHARMM36m, as well as *ab initio* molecular dynamics, revealed substantially different

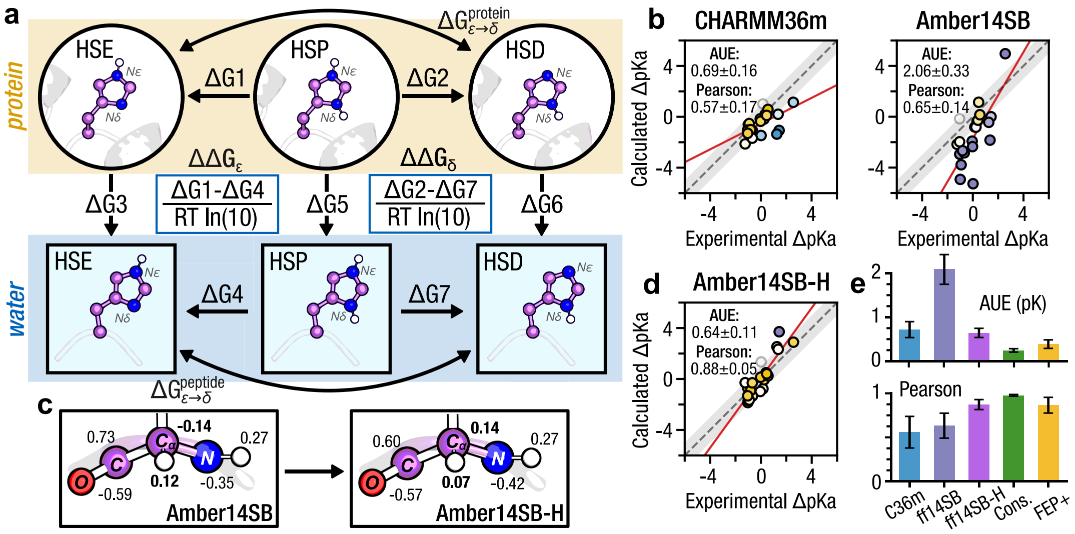


Figure 4. Histidine partial charges: Amber14SB-H. (a) Thermodynamic cycle corresponding to the pK_a of histidine. (b) Correlation between the calculated and experimental histidine pK_a values. Marker color indicates deviation from experiment where yellow indicates a minimum AUE from experiment. Regression lines are indicated in red, and the gray error band represents a 1 pK unit deviation from experiment. (c) Comparison of backbone partial charges between Amber14SB and Amber14SB-H. (d) Correlation between calculated and experimental histidine pK_a values for the modified Amber14SB-H. Error bands and marker color scheme match panel b. (e) Comparative prediction performance across the considered force fields. Consensus is the average of CHARMM36m and Amber14SB-H.

hydration structures (Figure 3a,b). Specifically, Amber14SB methylthiolate exhibited a radial distribution function (RDF) peak backshift of 0.5 nm compared to AIMD, suggesting a potentially erroneous hydration structure (Figure 3b). We note that while the revPBE-D3 functional has previously been shown to well reproduce the solvation structures of water[48] and anions like chloride[49,50] as a pure GGA functional it has a tendency to overdelocalize electrons, which may shift the first peak position of $g(r)$ to a larger distance. Importantly, the position of the first peak in our O−S RDF, i.e., $r \approx 3.16$ Å is consistent with that observed in two separate AIMD studies of methylthiolate solvation which calculated values of $r \approx 3.20$ Å and $r \approx 3.10$ Å, employing higher-level hybrid and range-separated functionals.[47,51]

Given the significant discrepancy in prediction performance between CHARMM36m and Amber14SB, as well as the notable differences in RDFs, we rescaled the Lennard-Jones σ value to improve agreement with the AIMD RDF and potentially improve pK_a prediction performance.

Exploring the matrix of rescaled σ- and ϵ-values within the interval [0.5, 1.5] with 0.1 spacing, we found that a σ-value of 1.1 well reproduced the oxygen−sulfur RDF from the AIMD trajectories (Figure 3b). Expanding the grid to include more values yielded the same conclusion. Because adjusting the ϵ-value led to only marginal enhancements (Figure S3b), we refrained from unnecessarily fitting both parameters.

We performed a similar analysis to determine an optimal value for predicting the solvation free energy of methylthiolate.[52] Consistent with the RDF analysis, we observed that modifications to σ yielded more significant improvements than changes to ϵ (Figure S5b); however, achieving an accurate solvation-free energy required a σ-scaling of 1.3 (Figure S5a).

With the primary goal of improving pK_a prediction performance we probed the pK_a values of wild type DsbA and seven mutants. Because changes in ϵ had a limited effect on solvent structure or solvation free energy, we decided to only scan σ-values on the interval [0.8, 1.5]. We observed a sigmoidal improvement in accuracy that was saturated for σ = 1.3 with an AUE of 2.88 ± 0.24 pK (Figure 3c) and a Pearson correlation of 0.31 ± 0.60. This accuracy was significantly increased from unscaled Amber14SB which had an AUE of 5.21 ± 0.44 pK and a correlation of 0.00 ± 0.51 on the DsbA test set.

Using Amber14SB-1.3σ on the full data set gave an AUE of 2.88 ± 0.35 pK and a Pearson correlation of 0.20 ± 0.11 (Figure 3d): markedly improved from the performance of plain Amber14SB, but still worse than the accuracy previously reported for aspartate, glutamate, and lysine. Using the σ-value that maximized agreement with the *ab initio* determined solvation structure (i.e., σ = 1.1), yielded an AUE of 3.60 ± 0.47 pK and a Pearson correlation of 0.17 ± 0.09 (Figure 3d). Previous efforts to reparameterize the thiolate parameters for cysteine targeted both the σ- and ϵ-values.[51] Optimizing against the AIMD solvation structure of methylthiolate, the researchers found a scaling factor of 1.08 for σ and 1.40 for ϵ provided the best agreement. Using these parameters we observed a statistically significant AUE improvement over the original Amber14SB of 0.38 pK (Figure 3e, Figure S7) which is comparable to the ≈0.5 pK improvement reported using the same parameters with Amber99SB on a different data set.[42] As discussed above, scaling σ by 1.1, which best matched our AIMD solvation structure, resulted in a nonsignificant AUE improvement of 0.25 pK, while further scaling to 1.3 did yield a significant improvement of 0.97 pK (Figure 3e).

We performed an identical analysis for CHARMM36m, which indicated that although a larger σ value (i.e., $\sigma \approx 1.15$) was required to achieve an accurate experimental solvation free energy (Figure S5a), the default LJ parameters could effectively reproduce the RDF data (Figure 3b) and maximize pK_a prediction accuracy on the DsbA test set (Figure 3c); this observation led us to leave the σ parameter untouched. Using the OPC water model (rather than mTIP3P) resulted in an identical position of the first solvation shell (Figure S6a) and suggested the same scaling factor was required to reproduce the experimental solvation free energy (Figure S6b).

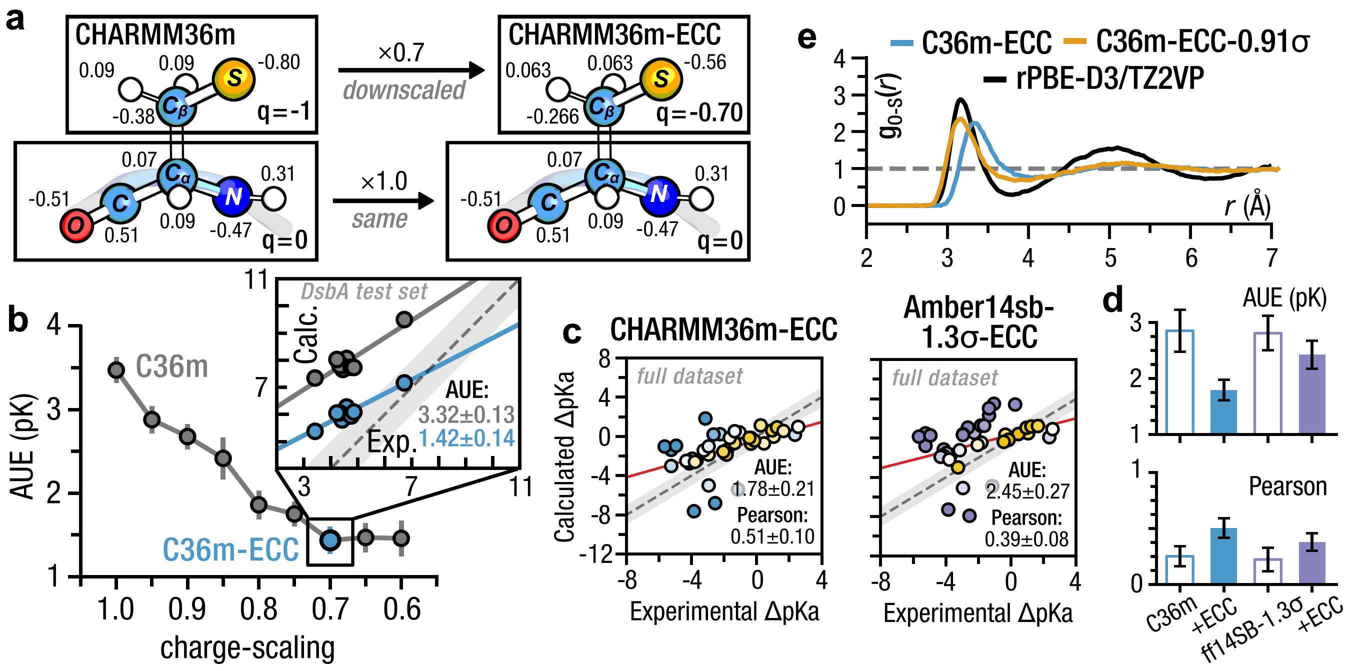


Figure 5. Effective polarization: CHARMM36m-ECC. (a) Charge scaling scheme where the side chain unit charge is scaled while the backbone remains fixed. (b) AUE on the DsbA test set as a function of scaling factor. The value of 0.70 for which the error saturates is marked in blue. (b, inset) Correlation between the calculated and experimental DsbA test set pK_a values. Regression lines are shown, and the gray error band represents a 1 pK unit deviation from experiment. (c) Correlation between the calculated and experimental cysteine pK_a values. Marker color indicates deviation from experiment where yellow indicates a minimum AUE from experiment. Regression lines are indicated in red, and the gray error band represents a 1 pK unit deviation from experiment. (d) Prediction performance comparison between the unscaled (empty bars) and charge-scaled (filled bars) force fields. (e) Solvation structure of charge-scaled methylthiolate with and without σ-scaling on the sulfur atom.

In short, the default Amber14SB cysteine thiolate parameters are erroneous and increasing σ or both σ and ϵ is required to better reproduce QM and experimental observables. Nevertheless, even with this modification, CHARMM36m still exceeds the accuracy of Amber14SB−1.3σ.

Histidine Partial Charges: Amber14SB-H. In the course of our coupling analysis, we found that histidine pK_a values are predicted significantly higher with Amber14SB than with CHARMM36m (Figure 4b). We previously observed lower accuracy for the prediction of lysine pK_as with the Amber14SB force field: this was traced to the partial charge difference of the backbone between the charged and uncharged lysine species.[17] We hypothesized that this may also play a role for histidine, as here too the backbone partial charges differ between the doubly (denoted HSP) and singly protonated histidine residues (denoted HSD and HSE). To further investigate, we computed 22 histidine pK_a values that were taken from a full data set previously probed using equilibrium free energy calculations.[53] These calculations were performed using both plain Amber14SB and a modified version, here called Amber14SB-H, where the partial charges of the protonated histidine backbone are those previously reported by Best et al (Figure 4c).[54]

To account for the fact that two neutral histidine tautomers can exist with the proton present on $N\delta$ (HSD) or $N\epsilon$ (HSE), we perform two sets of free energy calculations: HSP → HSD and HSP → HSE, which yield two relative free energies of deprotonation: $\Delta\Delta G_\delta$ and $\Delta\Delta G_\epsilon$. Taking their difference gives the relative free energy of tautomer interconversion:

$$\Delta\Delta G_\delta - \Delta\Delta G_\epsilon = -(\Delta G^{\text{prot}}_{\epsilon\to\delta} - \Delta G^{\text{pep}}_{\epsilon\to\delta}) = -\Delta\Delta G_{\epsilon\to\delta} \tag{2}$$

which we can combine with the absolute free energy of tautomer conversion—determined from the experimental microscopic pK_a values as $\Delta G^{\text{pep}}_{\epsilon\to\delta} \approx 2.2$ kJ/mol[55]—to get the overall relative free energy of deprotonation (Figure 4a):

$$\Delta\Delta G = \Delta\Delta G_\epsilon - \frac{1}{\beta}\ln(1 + e^{-\beta(\Delta G_{\epsilon\to\delta} + \Delta\Delta G_\epsilon - \Delta\Delta G_\delta)}) \tag{3}$$

from which we can determine the pK_a via eq 1.

Compared to plain Amber14SB, using Amber14SB-H significantly reduced the AUE from 2.06 ± 0.33 pK to 0.64 ± 0.11 pK and increased the Pearson correlation from 0.65 ± 0.14 to 0.88 ± 0.05 (Figure 4d,e). Unlike previously observed for aspartate, glutamate, and lysine, we found a consensus estimate for CHARMM36m and Amber14SB-H resulted in an predictor that exceeded the performance of either method alone (i.e., AUE: 0.24 ± 0.04 pK, Pearson correlation: 0.98 ± 0.01). This level of accuracy exceeded that achieved using FEP+ (i.e., 0.39 pK) on the same 22 pK_a data set (Figure 4e).[53]

Our results suggest that NES can resolve histidine pK_a values as accurately as FEP+ and further supports our previous suggestion that free energy calculations, in particular pK_a calculations, with Amber14SB should employ the more recent, Best et al. partial charges.[54] We note that this partial charge suggestion may also apply to Amber19SB, which utilizes the same backbone charges as Amber14SB.

Effective Polarization: CHARMM36m-ECC. Traditional MM force fields do not explicitly account for electronic polarizability. While the Drude[56] and AMOEBA[57] force fields explicitly introduce this missing electronic polarization, it can also be introduced implicitly. The electronic continuum correction (ECC) models the simulated system as a collection of point charges embedded in a medium.[58,59] This medium has a dielectric constant of ≈2, corresponding to the high-frequency dielectric of most condensed phase environments.[59] Applying this as a screening factor into the Coulomb equation effectively scales charges by $\frac{1}{\sqrt{2}} \approx 0.7$. In condensed-phase calculations, like the ones considered here, the ECC approximation is reasonable and has been shown to improve thermodynamic and kinetic observables across diverse biomolecular systems.[60−62] Given that sulfur is significantly more polarizable than oxygen and nitrogen, we hypothesized that the notably poorer pK_a prediction performance for cysteine—compared to glutamate, aspartate, histidine, and lysine—stems from inaccurately modeled electrostatic interactions between the cysteine thiolate and its protein-residue neighborhood. By reintroducing the missing polarization implicitly, we aimed to refine the representation of this local environment and, in turn, improve the accuracy of our free energy calculations.

We rescaled all full unit charges (i.e., charged side chains and ions) (Figure 5a) on the interval [0.60, 1.00] with a 0.05 increment in CHARMM36m and recomputed the pK_a values of the DsbA test set. CHARMM36m was chosen because of its higher accuracy in predicting cysteine pK_a and because recent charge scaling efforts have successfully employed this force field.[61]

We observed the accuracy saturated to an AUE of 1.42 ± 0.14 pK for a scaling of 0.70 (Figure 5b), very close to 0.75, which is the value often used within ECC scaling frameworks (e.g., prosEECo75). Applying this 0.70 scaled CHARMM36m force field on the entire cysteine data set reduced the AUE from 2.92 ± 0.35 pK to 1.78 ± 0.21 pK and increased the correlation with experiment from 0.24 ± 0.09 to 0.51 ± 0.09 (Figure 5c,d). Accounting for residue coupling improved the accuracy of CHARMM36m-ECC even further, shifting the overall AUE from 1.78 ± 0.21 pK to 1.61 ± 0.21 pK (Figure S2). Compared to plain CHARMM36m, CHARMM36m-ECC did not significantly improve the already strong histidine pK_a prediction performance (i.e., 0.69 ± 0.16 pK vs 0.71 ± 0.16 pK). Charge-scaling did also not completely resolve the significant pK_a underestimation observed for YopH tyrosine phosphatase (PDB: 1YPT) (Figure S8a). While we could exactly reproduce the relative effects of two nearby mutations (Figure S8c), the absolute pK_a values were downshifted by ≈4 pK units (Figure S8b).

Having scaled down the charge of cysteine, we have also increased the effective radius of the side chain atoms, in particular sulfur. Comparing the solvation structure of charge-scaled methylthiolate to the AIMD simulations, we found a slight increase in the position of the first RDF peak (Figure S9a). Scaling the sulfur σ by 0.91 maximized overlap between with the MD and AIMD RDF curves (Figure 5e, Figure S9b).

As a cross check we also computed solvation free energies of charge- and σ-scaled methylthiolate. Within the ECC framework, absolute free energies cannot be compared directly with experiment but must be adjusted to account for the scaling (see SI methods). After adjusting the values, we found that similar to unscaled CHARMM36m, a slightly larger σ scaling is required to reproduce the experimental solvation free energy (Figure S9b).

Applying the charge-scaled *and* σ-scaled CHARMM36m force field on the 1A2L test set showed no significant improvement (Figure S9c); we did not probe the entire data set with this doubly modified force field.

As an alternative to scaling all unit charges, we charge-scaled only the probed cysteine and balanced the missing negative charge by scaling the ions in solution. Computing the pK_a values on the DsbA test set revealed a similar improvement trend as observed for global charge scaling, but nevertheless yielded a slightly poorer accuracy at a 0.7 scaling (Figure S11). This difference is quite small and would seem to suggest that charge-scaled interactions of the probed cysteine itself and not that of other charged species is the major determinant of improved accuracy. In certain highly charged contexts (i.e., enzyme active sites), the accuracy improvement from charge-scaling other nearby residues is likely to play a larger role.

As a complete alternative to charge-scaling we also considered the more traditional reparameterization approach of redistributing the charge on the side chain. We found that altering the proportion of charge on the Cβ carbon and sulfur did not improve pK_a prediction accuracy on the DsbA test set (Figure S10).

Given the success with CHARMM36m, we also investigated charge-scaling with Amber14SB and Amber14SB−1.3σ. Amber presents difficulties because unlike CHARMM the side chain does not carry a full integer charge and cannot be simply scaled. Instead we linearly interpolate between the protonated and deprotonated cysteine to get a charge-scaled residue. Charge scaling Amber14SB improved prediction accuracy on the DsbA test set but failed to meaningfully saturate on the interval [1.00, 0.60], while charge-scaling Amber14SB-1.3σ improved prediction accuracy which was maximized for 0.80 and appeared to degrade for further scaling (Figure S11). Probing Amber14SB-1.3σ with 0.80 charge scaling on the entire data set improved the accuracy from an AUE of 2.88 ± 0.38 pK to 2.45 ± 0.27 pK (Figure 5c,d); this improvement was smaller than that observed for CHARMM36m. Accounting for coupling also pushed the accuracy slightly higher to 2.37 ± 0.29 pK. Compared to plain Amber14SB-H, charge-scaling did improve histidine pK_a prediction accuracy decreasing the average unsigned error from 0.64 ± 0.11 pK to 0.50 ± 0.10 pK.

To assess structural effects of charge-scaling on the simulated ensembles (e.g., unfolding) we analyzed the fluctuation profiles of the end-state ensembles. To compare the unscaled and scaled force fields we calculated the absolute difference between the residue-wise RMSF profile: $|\text{RMSF}_A - \text{RMSF}_B|/\text{RMSF}_B$ and took the average. This measure varied between proteins but was roughly ≈0.15 Å (Figure S12), which was comparable to the difference observed between unscaled Amber14SB and CHARMM36m. While a comprehensive validation of charge-scaled force fields against their unscaled counterparts is beyond the scope of this work, our observations—along with previous work demonstrating the stability of charge-scaled force fields in larger systems over longer time scales[62]—we suggest that charge-scaling is unlikely to deleteriously destabilize folded protein systems.

Taken as a whole, our results suggest charge scaling is a viable, force field independent method for improving or maintaining the already strong performance of nonpolarizable force fields in pK_a-related, free energy calculations.

Overall Performance: Modified Force Fields. Figure 6 summarizes the main findings: on the full dataset our NES approach with the modified CHARMM36m force field and residue coupling accounted for, significantly exceeds the

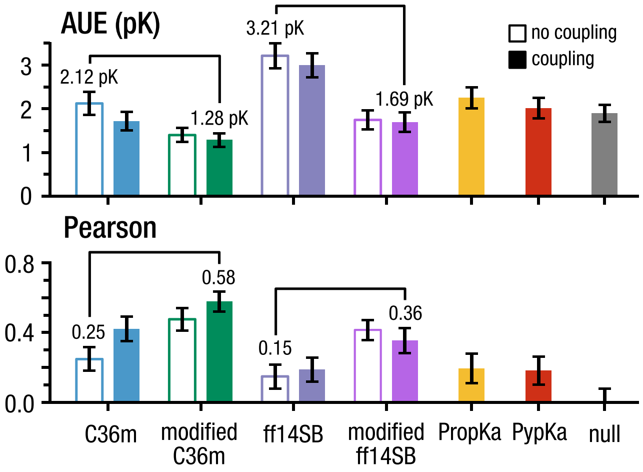


Figure 6. Overall performance: modified force fields. Combined cysteine and histidinepK_a prediction performance comparison between unmodified and modified CHARMM36m and Amber14SB force fields, with and without coupling accounted for, PropKa, PypKa, and a null model.

performance of plain CHARMM36m reducing the average unsigned error from 2.12 ± 0.27 pK to 1.28 ± 0.15 pK and increasing the correlation with experiment from 0.25 ± 0.09 to 0.58 ± 0.08. For Amber14SB, our force field modifications decrease the error from 3.21 ± 0.29 pK to 1.69 ± 0.23 pK and increase the correlation from 0.15 ± 0.10 to 0.36 ± 0.10. On the same dataset PropK_a and PypK_a gave errors of 2.25 ± 0.24 pK and 2.02 ± 0.23 pK, and correlations of 0.19 ± 0.12 and 0.18 ± 0.12, respectively. The average unsigned error of a null model was 1.89 ± 0.19 pK. Considering cysteine predictions within a certain tolerance, CHARMM36m-ECC correctly predicts 39 ± 8% of residues within 1 pK and 73 ± 7% within 2 pK, compared to 22 ± 6% and 44 ± 8% with PypK_a and 15 ± 6% and 34 ± 7% with PropK_a. We also note that CHARMM36m-ECC exceeds the AUE of a null model by 0.72 ± 0.21 pK or 0.91 ± 0.27 pK depending on whether coupling is accounted for.

Considering general determinants of accuracy we observed—consistent with previous work[17]—that the pK_a itself is a reasonable predictor of accuracy: performance degraded as a function of pK_a (Figure S13a). Decreasing pK_a also correlated with solvent exposure i.e., buried residues tended to have lower pK_a values (Figure S13b) and, by extension, solvent accessibility correlated with the prediction error (Figure S13c). In this data set, buried residues are also more frequently involved in coupling (Figure S13c, triangle markers). As noted earlier, coupling is an important determinant of accuracy, with poorer prediction performance observed for coupled residues compared to uncoupled ones. As also noted earlier, this discrepancy can be remedied by explicitly accounting for coupling, which restores the accuracy to a level comparable to that observed for uncoupled residues (Figure S2).

Taken together, this final comparison highlights that NES pK_a prediction accuracy can be significantly increased by accounting for charge scaling and residue coupling. These enhancements push the accuracy well above unscaled force fields and conventional predictor methods.

DISCUSSION

Here, we assess the ability of nonequilibrium switching (NES) free energy calculations to resolve the pK_a values of 40 cysteine and 22 histidine residues across 10 wildtype and mutant proteins. Given the widespread use of free energy calculations in lead optimization and the growing interest in designing targeted covalent inhibitors, *in silico* methods for determining the deprotonation free energy of specific cysteines in the presence or absence of bound molecules is highly desirable.

Our results highlight three force field modifications that can improve pK_a prediction accuracy for cysteine and histidine: (1) increasing the vdW radius of the deprotonated cysteine sulfur in Amber14SB; (2) altering the backbone partial charges of doubly protonated histidine in Amber14SB; and (3) charge scaling all unit charges in CHARMM36m and Amber14SB. Our investigation is not intended to provide definitive parameters for either force field or an absolute strategy for improving relative free energy calculations, particularly pK_a prediction, instead, we aim to highlight potential avenues for further investigation and development.

On the full data set of 40 cysteines and 22 histidines, we found the strongest performing force field, CHARMM36m-ECC, to exhibit an AUE of 1.61 ± 0.21 pK for cysteine; this accuracy exceeds conventional predictors and a null model. While increasing the vdW of sulfur and charge-scaling both improved the performance of Amber14SB in predicting cysteine pK_a values, the final accuracy of 2.36 ± 0.29 pK remains lower than that of CHARMM36m, suggesting further reparameterization of the residue would be required.

In the case of histidine, we found the accuracy could be significantly improved by taking a consensus of the Amber14SB-H and CHARMM36m charge-scaled force fields which yielded an AUE of 0.24 ± 0.04 pK and a Pearson correlation of 0.98 ± 0.01. Even standing alone, Amber14SB-H with charge-scaling attained an accuracy of 0.50 ± 0.10 pK and correlation of 0.85 ± 0.06, while CHARMM36m with charge-scaling gave an accuracy of 0.71 ± 0.16 pK and correlation of 0.56 ± 0.15.

We note that while NES paired with the charged-scaled CHARMM36m force field represents the strongest predictor reported here, our results suggest inherent limitations of conventional force fields in accurately capturing the local electrostatic environment of the cysteine thiolate. The deprotonated sulfur is significantly more polarizable than oxygen or nitrogen, making pK_a predictions particularly sensitive to local electrostatics, hydrogen bonding, and screening effects within the protein. These factors are not fully accounted for and likely explain why prediction accuracy is poorer compared to the less polarizable amino acids i.e., aspartate and lysine. In light of this, alternative approaches may be necessary to improve physics-based cysteine pK_a predictions. One potential avenue is the use of explicitly polarizable force fields[56,57] which could offer a more accurate description of electrostatics compared to the implicitly polarizable force fields employed here. Another potential direction is the integration of MM/ML end-state corrections.[63,64] As machine learning potentials become more capable of reliably modeling charged species at longer ranges, they could be used to correct the solvent and protein branches of the thermodynamic cycle in Figure 1, ultimately leading to more accurate pK_a estimates.

In summary we find MD-approaches, including NES, can resolve the pK_a values of cysteine and histidine residues with accuracy that exceeds conventional methods; however, this requires modification to the underlying MD force fields. The largest accuracy improvement we observed was for charge scaling the CHARMM36m force field, a result that will likely extend to other force fields and could remedy the poorer accuracy previously observed for predicting the effect of charge-changing mutations on protein thermostability[65] and binding affinity.[66] More work will help determine the consequences of charge-scaling; however, this work and the recent work of others[61,62] seems to suggest that charge-scaling may be a general method to enhance the accuracy of nonpolarizable MM force fields with only minimal and predictable costs.

■ ASSOCIATED CONTENT

Data Availability Statement

Calculated pK_a values and modified force field files are available at https://github.com/deGrootLab/pka_reparam_2025.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.5c00031.

> A derivation of the ECC solvation free energy relationship, supplementary figures, and supplemental tables, which contain references to PDB structures[67−86] and experimental pK_a values[4,5,87−102] (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Vytautas Gapsys − *Computational Biomolecular Dynamics Group, Max Planck Institute for Multidisciplinary Sciences, Göttingen 37077, Germany; Computational Chemistry, Janssen Research & Development, Janssen Pharmaceutica N. V., Beerse B-2340, Belgium;* orcid.org/0000-0002-6761-7780; Email: vgapsys@gwdg.de

Bert L. de Groot − *Computational Biomolecular Dynamics Group, Max Planck Institute for Multidisciplinary Sciences, Göttingen 37077, Germany;* Email: bgroot@gwdg.de

Author

Carter J. Wilson − *Computational Biomolecular Dynamics Group, Max Planck Institute for Multidisciplinary Sciences, Göttingen 37077, Germany;* orcid.org/0000-0002-8992-6269

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.5c00031

Funding

Open access funded by Max Planck Society.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

B.d.G. acknowledges support from BioExcel CoE (www.bioexcel.eu), a project funded by the European Union contract HORIZON-EUROHPC-JU-2021-COE-01-02. The authors thank Chenggong Hui for helpful discussions on charge scaling.

■ REFERENCES

(1) Maurais, A. J.; Weerapana, E. Reactive-cysteine profiling for drug discovery. *Curr. Opin. Chem. Biol.* **2019**, *50*, 29−36.

(2) Cannan, R. K.; Knight, B. C. J. G. Dissociation Constants of Cystine, Cysteine, Thioglycollic Acid and α-Thiolactic Acid. *Biochem. J.* **1927**, *21*, 1384−1390.

(3) Thurlkill, R. L.; Grimsley, G. R.; Scholtz, J. M.; Pace, C. N. pK values of the ionizable groups of proteins. *Protein Sci.* **2006**, *15*, 1214−1218.

(4) Pinitglang, S.; Watts, A. B.; Patel, M.; Reid, J. D.; Noble, M. A.; Gul, S.; Bokth, A.; Naeem, A.; Patel, H.; Thomas, E. W.; Sreedharan, S. K.; Verma, C.; Brocklehurst, K. A Classical Enzyme Active Center Motif Lacks Catalytic Competence until Modulated Electrostatically. *Biochemistry* **1997**, *36*, 9968−9982.

(5) Tolbert, B. S.; Tajc, S. G.; Webb, H.; Snyder, J.; Nielsen, J. E.; Miller, B. L.; Basavappa, R. The Active Site Cysteine of Ubiquitin-Conjugating Enzymes Has a Significantly Elevated pKa: Functional Implications. *Biochemistry* **2005**, *44*, 16385−16391.

(6) Giles, N. M.; Giles, G. I.; Jacob, C. Multiple roles of cysteine in biocatalysis. *Biochem. Biophys. Res. Commun.* **2003**, *300*, 1−4.

(7) Giles, N. M.; Watts, A. B.; Giles, G. I.; Fry, F. H.; Littlechild, J. A.; Jacob, C. Metal and Redox Modulation of Cysteine Protein Function. *Chem. Biol.* **2003**, *10*, 677−693.

(8) Klomsiri, C.; Karplus, P. A.; Poole, L. B. Cysteine-Based Redox Switches in Enzymes. *Antioxid. Redox Signal* **2011**, *14*, 1065−1077.

(9) Bechtel, T. J.; Weerapana, E. From structure to redox: The diverse functional roles of disulfides and implications in disease. *Proteomics* **2017**, *17*, 1600391.

(10) Pace, N. J.; Weerapana, E. Diverse Functional Roles of Reactive Cysteines. *ACS Chem. Biol.* **2013**, *8*, 283−296.

(11) Boike, L.; Henning, N. J.; Nomura, D. K. Advances in covalent drug discovery. *Nat. Rev. Drug Discovery* **2022**, *21*, 881−898.

(12) Tuley, A.; Fast, W. The Taxonomy of Covalent Inhibitors. *Biochemistry* **2018**, *57*, 3326−3337.

(13) Zhang, T.; Hatcher, J. M.; Teng, M.; Gray, N. S.; Kostic, M. Recent Advances in Selective and Irreversible Covalent Ligand Development and Validation. *Cell Chem. Biol.* **2019**, *26*, 1486−1500.

(14) Sutanto, F.; Konstantinidou, M.; Dömling, A. Covalent inhibitors: A rational approach to drug discovery. *RSC Med. Chem.* **2020**, *11*, 876−884.

(15) Wallerstein, J.; Weininger, U.; Khan, M. A. I.; Linse, S.; Akke, M. Site-Specific Protonation Kinetics of Acidic Side Chains in Proteins Determined by pH-Dependent Carboxyl 13C NMR Relaxation. *J. Am. Chem. Soc.* **2015**, *137*, 3093−3101.

(16) Chen, J.; Yadav, N. N.; Stait-Gardner, T.; Gupta, A.; Price, W. S.; Zheng, G. Thiol-water proton exchange of glutathione, cysteine, and N-acetylcysteine: Implications for CEST MRI. *NMR Biomed.* **2020**, 33, No. e4188.

(17) Wilson, C. J.; Karttunen, M.; de Groot, B. L.; Gapsys, V. Accurately Predicting Protein pKa Values Using Nonequilibrium Alchemy. *J. Chem. Theory Comput.* **2023**, *19*, 7833−7845.

(18) Lippert, G.; Hutter, J.; Parrinello, M. A hybrid Gaussian and plane wave density functional scheme. *Mol. Phys.* **1997**, *92*, 477−487.

(19) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133−A1138.

(20) VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.* **2005**, *167*, 103−128.

(21) Kühne, T. D.; Iannuzzi, M.; Del Ben, M.; Rybkin, V. V.; Seewald, P.; Stein, F.; Laino, T.; Khaliullin, R. Z.; Schütt, O.; Schiffmann, F.; et al. CP2K: An electronic structure and molecular dynamics software package - Quickstep: Efficient and accurate electronic structure calculations. *J. Chem. Phys.* **2020**, *152*, 194103.

(22) Marques, M. A.; Oliveira, M. J.; Burnus, T. Libxc: A library of exchange and correlation functionals for density functional theory. *Comput. Phys. Commun.* **2012**, *183*, 2272−2281.

(23) Zhang, Y.; Yang, W. Comment on "Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1998**, *80*, 890−890.

(24) Goerigk, L.; Grimme, S. A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. *Phys. Chem. Chem. Phys.* **2011**, *13*, 6670.

(25) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

(26) Goedecker, S.; Teter, M.; Hutter, J. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B* **1996**, *54*, 1703−1710.

(27) Hartwigsen, C.; Goedecker, S.; Hutter, J. Relativistic separable dual-space Gaussian pseudopotentials from H to Rn. *Phys. Rev. B* **1998**, *58*, 3641−3662.

(28) Izadi, S.; Anandakrishnan, R.; Onufriev, A. V. Building Water Models: A Different Approach. *J. Phys. Chem. Lett.* **2014**, *5*, 3863−3871.

(29) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1−2*, 19−25.

(30) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: AnN·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

(31) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.

(32) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.

(33) Gapsys, V.; Michielssens, S.; Seeliger, D.; de Groot, B. L. pmx: Automated protein structure and topology generation for alchemical perturbations. *J. Comput. Chem.* **2015**, *36*, 348−354.

(34) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71−73.

(35) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696−3713.

(36) Van Gunsteren, W. F.; Berendsen, H. J. C. A Leap-frog Algorithm for Stochastic Dynamics. *Mol. Simul.* **1988**, *1*, 173−185.

(37) Goga, N.; Rzepiela, A. J.; de Vries, A. H.; Marrink, S. J.; Berendsen, H. J. C. Efficient Algorithms for Langevin and DPD Dynamics. *J. Chem. Theory Comput.* **2012**, *8*, 3637−3649.

(38) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.

(39) Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 116−122.

(40) Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976**, *22*, 245−268.

(41) Crooks, G. E. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E* **1999**, *60*, 2721−2726.

(42) Awoonor-Williams, E.; Golosov, A. A.; Hornak, V. Benchmarking In Silico Tools for Cysteine pKa Prediction. *J. Chem. Inf. Model.* **2023**, *63*, 2170−2180.

(43) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKaPredictions. *J. Chem. Theory Comput.* **2011**, *7*, 525−537.

(44) Reis, P. B. P. S.; Vila-Viçosa, D.; Rocchia, W.; Machuqueiro, M. PypKa: A Flexible Python Module for Poisson-Boltzmann-Based pKa Calculations. *J. Chem. Inf. Model.* **2020**, *60*, 4442−4448.

(45) Klapper, I.; Hagstrom, R.; Fine, R.; Sharp, K.; Honig, B. Focusing of electric fields in the active site of Cu-Zn superoxide dismutase: Effects of ionic strength and amino-acid modification. *Proteins* **1986**, *1*, 47−59.

(46) Wilson, C. J.; de Groot, B. L.; Gapsys, V. Resolving coupled pH titrations using alchemical free energy calculations. *J. Comput. Chem.* **2024**, *45*, 1444−1455.

(47) Awoonor-Williams, E.; Rowley, C. N. Evaluation of Methods for the Calculation of the pKa of Cysteine Residues in Proteins. *J. Chem. Theory Comput.* **2016**, *12*, 4662−4673.

(48) Ruiz Pestana, L.; Mardirossian, N.; Head-Gordon, M.; Head-Gordon, T. Ab initio molecular dynamics simulations of liquid water using high quality meta-GGA functionals. *Chem. Sci.* **2017**, *8*, 3554−3565.

(49) Zhou, K.; Qian, C.; Liu, Y. Quantifying the Structure of Water and Hydrated Monovalent Ions by Density Functional Theory-Based Molecular Dynamics. *J. Phys. Chem. B* **2022**, *126*, 10471−10480.

(50) DelloStritto, M.; Xu, J.; Wu, X.; Klein, M. L. Aqueous solvation of the chloride ion revisited with density functional theory: Impact of correlation and exchange approximations. *Phys. Chem. Chem. Phys.* **2020**, *22*, 10666−10675.

(51) Pedron, F. N.; Messias, A.; Zeida, A.; Roitberg, A. E.; Estrin, D. A. Novel Lennard-Jones Parameters for Cysteine and Selenocysteine in the AMBER Force Field. *J. Chem. Inf. Model.* **2023**, *63*, 595−604.

(52) Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. *J. Phys. Chem.* **1994**, *98*, 1978−1988.

(53) Coskun, D.; Chen, W.; Clark, A. J.; Lu, C.; Harder, E. D.; Wang, L.; Friesner, R. A.; Miller, E. B. Reliable and Accurate Prediction of Single-Residue pKa Values through Free Energy Perturbation Calculations. *J. Chem. Theory Comput.* **2022**, *18*, 7193−7204.

(54) Best, R. B.; de Sancho, D.; Mittal, J. Residue-Specific *α*-Helix Propensities from Molecular Simulation. *Biophys. J.* **2012**, *102*, 1462−1467.

(55) Tanokura, M. 1H-NMR study on the tautomerism of the imidazole ring of histidine residues. *Biochim. Biophys. Acta* **1983**, *742*, 576−585.

(56) Lemkul, J. A.; Huang, J.; Roux, B.; MacKerell, A. D. An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications. *Chem. Rev.* **2016**, *116*, 4983−5013.

(57) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B* **2010**, *114*, 2549−2564.

(58) Leontyev, I. V.; Stuchebrukhov, A. A. Electronic continuum model for molecular dynamics simulations. *J. Chem. Phys.* **2009**, *130*, 085102.

(59) Leontyev, I. V.; Stuchebrukhov, A. A. Electronic Continuum Model for Molecular Dynamics Simulations of Biological Molecules. *J. Chem. Theory Comput.* **2010**, *6*, 1498−1508.

(60) Tolmachev, D. A.; Boyko, O. S.; Lukasheva, N. V.; Martinez-Seara, H.; Karttunen, M. Overbinding and Qualitative and Quantitative Changes Caused by Simple Na+ and K+ Ions in Polyelectrolyte Simulations: Comparison of Force Fields with and without NBFIX and ECC Corrections. *J. Chem. Theory Comput.* **2020**, *16*, 677−687.

(61) Nencini, R.; Tempra, C.; Biriukov, D.; Riopedre-Fernandez, M.; Cruces Chamorro, V.; Polák, J.; Mason, P. E.; Ondo, D.; Heyda, J.; Ollila, O. H. S.; Jungwirth, P.; Javanainen, M.; Martinez-Seara, H. Effective Inclusion of Electronic Polarization Improves the Description of Electrostatic Interactions: The prosECCo75 Biomolecular Force Field. *J. Chem. Theory Comput.* **2024**, *20*, 7546−7559.

(62) Hui, C.; de Vries, R.; Kopec, W.; de Groot, B. L.Effective Polarization in Potassium Channel Simulations: Ion Conductance, Occupancy, Voltage Response, and Selectivity*bioRxiv***2024**

(63) Rufa, D. A.; Bruce Macdonald, H. E.; Fass, J.; Wieder, M.; Grinaway, P. B.; Roitberg, A. E.; Isayev, O.; Chodera, J. D.Towards chemical accuracy for alchemical free energy calculations with hybrid physics-based machine learning/molecular mechanics potentials*bioRxiv***2020**

(64) Devereux, C.; Smith, J. S.; Huddleston, K. K.; Barros, K.; Zubatyuk, R.; Isayev, O.; Roitberg, A. E. Extending the Applicability of the ANI Deep Learning Molecular Potential to Sulfur and Halogens. *J. Chem. Theory Comput.* **2020**, *16*, 4192−4202.

(65) 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.* **2016**, *55*, 7364−7368.

(66) Sampson, J. M.; Cannon, D. A.; Duan, J.; Epstein, J. C.; Sergeeva, A. P.; Katsamba, P. S.; Mannepalli, S. M.; Bahna, F. A.; Adihou, H.; Guéret, S. M.; Gopalakrishnan, R.; Geschwindner, S.; Rees, D. G.; Sigurdardottir, A.; Wilkinson, T.; Dodd, R. B.; De Maria, L.; Mobarec, J. C.; Shapiro, L.; Honig, B.; Buchanan, A.; Friesner, R. A.; Wang, L. Robust Prediction of Relative Binding Energies for Protein-Protein Complex Mutations Using Free Energy Perturbation Calculations. *J. Mol. Biol.* **2024**, *436*, 168640.

(67) Elliott, P. R.; Pei, X. Y.; Dafforn, T. R.; Lomas, D. A. Topography of a 2.0 Å structure of α1-antitrypsin reveals targets for rational drug design to prevent conformational disease. *Protein Sci.* **2000**, *9*, 1274−1281.

(68) Perkins, A.; Nelson, K. J.; Williams, J. R.; Parsonage, D.; Poole, L. B.; Karplus, P. A. The Sensitive Balance between the Fully Folded and Locally Unfolded Conformations of a Model Peroxiredoxin. *Biochemistry* **2013**, *52*, 8708−8721.

(69) Jia, Z.; Hasnain, S.; Hirama, T.; Lee, X.; Mort, J. S.; To, R.; Huber, C. P. Crystal Structures of Recombinant Rat Cathepsin B and a Cathepsin B-Inhibitor Complex. *J. Biol. Chem.* **1995**, *270*, 5527−5533.

(70) Wilson, M. A.; Collins, J. L.; Hod, Y.; Ringe, D.; Petsko, G. A. The 1.1-Å resolution crystal structure of DJ-1, the protein mutated in autosomal recessive early onset Parkinson's disease. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 9256−9261.

(71) Shen, Y.; Tang, L.; Zhou, H.; Lin, Z. Structure of human muscle creatine kinase. *Acta Crystallogr., Sect. D: Biol.* **2001**, *57*, 1196−1200.

(72) Lim, J. C.; Gruschus, J. M.; Ghesquière, B.; Kim, G.; Piszczek, G.; Tjandra, N.; Levine, R. L. Characterization and Solution Structure of Mouse Myristoylated Methionine Sulfoxide Reductase A. *J. Biol. Chem.* **2012**, *287*, 25589−25595.

(73) Daniels, D. S. Active and alkylated human AGT structures: A novel zinc site, inhibitor and extrahelical base binding. *EMBO J.* **2000**, *19*, 1719−1730.

(74) Pickersgill, R. W.; Rizkallah, P.; Harris, G. W.; Goodenough, P. W. Determination of the structure of papaya protease omega. *Acta Crystallogr. Sect. B: Struct. Sci.* **1991**, *47*, 766−771.

(75) Pickersgill, R. W.; Harris, G. W.; Garman, E. Structure of monoclinic papain at 1.60 Å resolution. *Acta Crystallogr. Sect. B: Struct. Sci.* **1992**, *48*, 59−67.

(76) Weichsel, A.; Gasdaska, J. R.; Powis, G.; Montfort, W. R. Crystal structures of reduced, oxidized, and mutated human thioredoxins: Evidence for a regulatory homodimer. *Structure* **1996**, *4*, 735−751.

(77) Barford, D.; Flint, A. J.; Tonks, N. K. Crystal Structure of Human Protein Tyrosine Phosphatase 1B. *Science* **1994**, *263*, 1397−1404.

(78) Miura, T.; Klaus, W.; Ross, A.; Güntert, P.; Senn, H. *J. Biomol. NMR* **2002**, *22*, 89−92.

(79) VanDemark, A. P.; Hofmann, R. M.; Tsui, C.; Pickart, C. M.; Wolberger, C. Molecular Insights into Polyubiquitin Chain Assembly. *Cell* **2001**, *105*, 711−720.

(80) Lin, Y.; Hwang, W. C.; Basavappa, R. Structural and Functional Analysis of the Human Mitotic-specific Ubiquitin-conjugating Enzyme, UbcH10. *J. Biol. Chem.* **2002**, *277*, 21913−21921.

(81) Stuckey, J. A.; Schubert, H. L.; Fauman, E. B.; Zhang, Z.-Y.; Dixon, J. E.; Saper, M. A. Crystal structure of Yersinia protein tyrosine phosphatase at 2.5 Å and the complex with tungstate. *Nature* **1994**, *370*, 571−575.

(82) Quillin, M. L.; Arduini, R. M.; Olson, J. S.; Phillips, G. N. High-Resolution Crystal Structures of Distal Histidine Mutants of Sperm Whale Myoglobin. *J. Mol. Biol.* **1993**, *234*, 140−155.

(83) Guddat, L. W.; Bardwell, J. C.; Martin, J. L. Crystal structures of reduced and oxidized DsbA: Investigation of domain motion and thiolate stabilization. *Structure* **1998**, *6*, 757−767.

(84) Jeng, M.-F.; Campbell, A.; Begley, T.; Holmgren, A.; Case, D. A.; Wright, P. E.; Dyson, H. High-resolution solution structures of oxidized and reduced *Escherichia coli* thioredoxin. *Structure* **1994**, *2*, 853−868.

(85) Alphey, M. S.; Gabrielsen, M.; Micossi, E.; Leonard, G. A.; McSweeney, S. M.; Ravelli, R. B.; Tetaud, E.; Fairlamb, A. H.; Bond, C. S.; Hunter, W. N. Tryparedoxins from Crithidia fasciculata and Trypanosoma brucei. *J. Biol. Chem.* **2003**, *278*, 25919−25925.

(86) Crow, A.; Acheson, R. M.; Le Brun, N. E.; Oubrie, A. Structural Basis of Redox-coupled Protein Substrate Selection by the Cytochrome c Biosynthesis Protein ResA. *J. Biol. Chem.* **2004**, *279*, 23654−23660.

(87) Griffiths, S. W.; King, J.; Cooney, C. L. The Reactivity and Oxidation Pathway of Cysteine 232 in Recombinant Human α1-Antitrypsin. *J. Biol. Chem.* **2002**, *277*, 25486−25492.

(88) Nelson, K. J.; Parsonage, D.; Hall, A.; Karplus, P. A.; Poole, L. B. Cysteine pKa Values for the Bacterial Peroxiredoxin AhpC. *Biochemistry* **2008**, *47*, 12860−12868.

(89) Hasnain, S.; Hirama, T.; Tam, A.; Mort, J. Characterization of recombinant rat cathepsin B and nonglycosylated mutants expressed in yeast. New insights into the pH dependence of cathepsin B-catalyzed hydrolyses. *J. Biol. Chem.* **1992**, *267*, 4713−4721.

(90) Witt, A. C.; Lakshminarasimhan, M.; Remington, B. C.; Hasim, S.; Pozharski, E.; Wilson, M. A. Cysteine pKa Depression by a Protonated Glutamic Acid in Human DJ-1. *Biochemistry* **2008**, *47*, 7430−7440.

(91) Wang, P.-F.; McLeish, M. J.; Kneen, M. M.; Lee, G.; Kenyon, G. L. An Unusually Low pKafor Cys282 in the Active Site of Human Muscle Creatine Kinase. *Biochemistry* **2001**, *40*, 11698−11705.

(92) Lim, J. C.; Gruschus, J. M.; Kim, G.; Berlett, B. S.; Tjandra, N.; Levine, R. L. A Low pK Cysteine at the Active Site of Mouse Methionine Sulfoxide Reductase A. *J. Biol. Chem.* **2012**, *287*, 25596−25601.

(93) Guengerich, F. P.; Fang, Q.; Liu, L.; Hachey, D. L.; Pegg, A. E. O6-Alkylguanine-DNA Alkyltransferase: Low pKa and High Reactivity of Cysteine 145. *Biochemistry* **2003**, *42*, 10965−10970.

(94) Forman-Kay, J. D.; Clore, G. M.; Gronenborn, A. M. Relationship between electrostatics and redox function in human thioredoxin: Characterization of pH titration shifts using two-dimensional homo- and heteronuclear NMR. *Biochemistry* **1992**, *31*, 3442−3452.

(95) Lohse, D. L.; Denu, J. M.; Santoro, N.; Dixon, J. E. Roles of Aspartic Acid-181 and Serine-222 in Intermediate Formation and Hydrolysis of the Mammalian Protein-Tyrosine-Phosphatase PTP1. *Biochemistry* **1997**, *36*, 4568−4575.

(96) Zhang, Z. Y.; Dixon, J. E. Active site labeling of the Yersinia protein tyrosine phosphatase: The determination of the pKa of the active site cysteine and the function of the conserved histidine 402. *Biochemistry* **1993**, *32*, 9340−9345.

(97) Jensen, K. S.; Pedersen, J. T.; Winther, J. R.; Teilum, K. The pKaValue and Accessibility of Cysteine Residues Are Key Determinants for Protein Substrate Discrimination by Glutaredoxin. *Biochemistry* **2014**, *53*, 2533−2540.

(98) Miranda, J. L. Position-dependent interactions between cysteine residues and the helix dipole. *Protein Sci.* **2003**, *12*, 73−81.

(99) Grauschopf, U.; Winther, J. R.; Korber, P.; Zander, T.; Dallinger, P.; Bardwell, J. C. Why is DsbA such an oxidizing disulfide catalyst? *Cell* **1995**, *83*, 947−955.

(100) Chivers, P. T.; Prehoda, K. E.; Volkman, B. F.; Kim, B.-M.; Markley, J. L.; Raines, R. T. Microscopic pKa Values of *Escherichia coli* Thioredoxin. *Biochemistry* **1997**, *36*, 14985−14991.

(101) Manta, B.; Möller, M. N.; Bonilla, M.; Deambrosi, M.; Grunberg, K.; Bellanda, M.; Comini, M. A.; Ferrer-Sueta, G. Kinetic studies reveal a key role of a redox-active glutaredoxin in the evolution of the thiol-redox metabolism of trypanosomatid parasites. *J. Biol. Chem.* **2019**, *294*, 3235−3248.

(102) Lewin, A.; Crow, A.; Oubrie, A.; Le Brun, N. E. Molecular Basis for Specificity of the Extracytoplasmic Thioredoxin ResA. *J. Biol. Chem.* **2006**, *281*, 35467−35477.