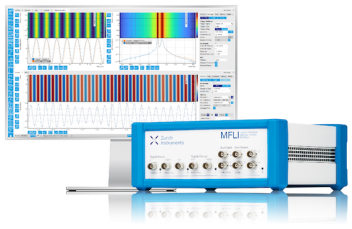
The Journal of Chemical Physics

RESEARCH ARTICLE | NOVEMBER 04 2024

Statistics of protein electrostatics

Taylor Colburn ; Setare Mostajabi Sarhangi; Dmitry V. Matyushov

J. Chem. Phys. 161, 175101 (2024)
https://doi.org/10.1063/5.0229619



The Journal of Chemical Physics ARTICLE pubs.aip.org/aip/jcp

Statistics of protein electrostatics

Cite as: J. Chem. Phys. 161, 175101 (2024); doi: 10.1063/5.0229619
Submitted: 22 July 2024 • Accepted: 18 October 2024 •
Published Online: 4 November 2024

Taylor Colburn,[1] Setare Mostajabi Sarhangi,[1] and Dmitry V. Matyushov[2,a)]

AFFILIATIONS

[1] Department of Physics, Arizona State University, P.O. Box 871504, Tempe, Arizona 85287-1504, USA
[2] School of Molecular Sciences and Department of Physics, Arizona State University, P.O. Box 871504, Tempe, Arizona 85287-1504, USA

[a)] Author to whom correspondence should be addressed: dmitrym@asu.edu

ABSTRACT

Molecular dynamics simulations of a small redox-active protein plastocyanin address two questions. (i) Do protein electrostatics equilibrate to the Gibbsian ensemble? (ii) Do the electrostatic potential and electric field inside proteins follow the Gaussian distribution? The statistics of electrostatic potential and electric field are probed by applying small charge and dipole perturbations to different sites within the protein. Nonergodic (non-Gibbsian) sampling is detectable through violations of exact statistical rules constraining the first and second statistical moments (fluctuation–dissipation relations) and the linear relation between free-energy surfaces of the collective coordinate representing the Hamiltonian electrostatic perturbation. We find weakly nonergodic statistics of the electrostatic potential (simulation time of 0.4–1.0 μs) and non-Gibbsian and non-Gaussian statistics of the electric field. A small dipolar perturbation of the protein results in structural instabilities of the protein–water interface and multi-modal distributions of the Hamiltonian energy gap. The variance of the electrostatic potential passes through a crossover at the glass transition temperature $T_{tr} \simeq 170$ K. The dipolar susceptibility, reflecting the variance of the electric field inside the protein, strongly increases, with lowering temperature, followed by a sharp drop at T_{tr}. The linear relation between free-energy surfaces can be directly tested by combining absorption and emission spectra of optical dyes. It was found that the statistics of the electrostatic potential perturbation are nearly Gibbsian/Gaussian, with little deviations from the prescribed statistical rules. On the contrary, the (nonergodic) statistics of dipolar perturbations are strongly non-Gibbsian/non-Gaussian due to structural instabilities of the protein hydration shell.

I. INTRODUCTION

This study explores the statistics of electrostatic fluctuations produced by the protein–water thermal bath inside proteins. Molecular dynamics (MD) simulations are performed to study general rules applicable to the statistics of two electrostatic observables: the electrostatic potential and the electric field inside a small globular protein plastocyanin (PC). This protein participates as an electron shuttle in the electron transport chains of plants,[1] delivering electrons from photosystem I to photosystem II. The distribution of the electrostatic potential at the PC's active site is a major factor[2] affecting the activation barrier for the reaction of electron transfer between PC and electron donor/acceptor in the membrane-bound protein complexes of two photosystems.[3,4] In contrast, the distribution of the electric field is a major factor defining the medium-induced shift and broadening of optical spectral lines of photoactive chromophores in proteins.[5] Both electrostatic observables can, therefore, be connected to laboratory observations. Keeping that in mind, we view them here as generic properties of proteins solvated in water and ask general questions about the statistics of these two collective electrostatic variables and the dependence of their statistical moments on temperature.

While the statistics of the electrostatic potential ϕ and the electric field E can be directly calculated from MD trajectories, we will follow the protocol of experimental interrogation of physical properties by observing the response of the protein–water system to a small perturbation, allowing one to use linear response theories.[6,7] In particular, we will introduce a change in the system Hamiltonian ΔH by either altering partial charges on tagged residues or by placing a dipole moment $\Delta\mu$ on a bond within the protein residue. The goal here is to establish the basic phenomenology of protein electrostatics by looking at the response of the thermal bath to such small multipolar perturbations.

Deviations of the system from equilibrium can be characterized by a set of collective coordinates establishing external perturbations. The reversible work required to bring the system out of equilibrium

defines the free energy surface (potential of mean force, PMF). The system is expected to reach PMF's minimum at equilibrium. Such a generally multidimensional free-energy surface can be approximated by a paraboloid near the equilibrium set of the chosen collective coordinates. This is depicted as a one-dimensional surface along a single collective coordinate shown in Fig. 1. The position of mechanical equilibrium is thus characterized by zero total force acting on the protein: $\mathbf{f}_{\rm tot} = 0$. The total force is a sum of a number of forces of different physical nature, of which electrostatic (el) and van der Waals (vdW) forces are the two main contributions. Neglecting other components, one can thus write $\mathbf{f}_{\rm el} = -\mathbf{f}_{\rm vdW}$.

While the total force reaches zero at mechanical equilibrium, the individual force components do not have to become zero. It is, in fact, well established that proteins possess high electric fields in their interior,[5,8] implying that electrical and vdW forces are separately out of equilibrium ($\mathbf{f}_{\rm el,vdW} \neq 0$) when the overall mechanical equilibrium is reached for the protein. This is indicated by the two shifted parabolas for electrostatic (el) and vdW free-energy surfaces shown in Fig. 1. The slopes for the electrostatic and vdW surfaces are nonzero at equilibrium: the slope of the electrostatic free-energy surface indicates a nonvanishing average electric field within the protein at equilibrium.

The total free energy surface can be approximated by a parabola near equilibrium, and one anticipates Gaussian statistics for a collective coordinate $u = u_{\rm el} + u_{\rm vdW}$ specified in the following. However, given that components $u_{\rm el,vdW}$ are not at mechanical equilibrium, Gaussian statistics do not have to apply to them. In other words, while the distribution of u samples the states near the equilibrium and is necessarily Gaussian, the components $u_{\rm el,vdW}$ sample the wings of the distribution and can potentially be non-Gaussian. Here, we focus solely on the electrostatic component to examine, by MD simulations, whether the electrostatic potential and the electric field follow the Gaussian statistics. This question is not easy to answer directly since sampling of distribution wings requires advanced sampling techniques.[9] We will, therefore, resort to analyzing relations between statistical moments of the Hamiltonian perturbation in the unperturbed and perturbed states required by the Gaussian statistics. We also ask the question of whether the second central moments of these perturbations follow the linear temperature law ($\propto T$), known as the Johnson–Nyquist law,[10] anticipated by the fluctuation–dissipation theorem (FDT).[11,12]

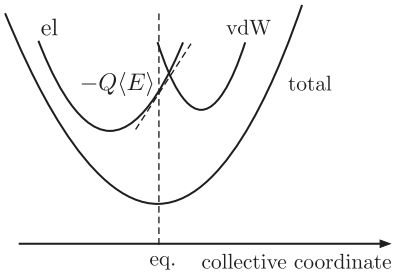


FIG. 1. Schematic drawing of the free-energy surface of the protein–water system vs an unspecified collective coordinate. The total free energy is at minimum at the equilibrium configuration. The electrostatic (el) and van der Waals (vdW) components of the free energy do not have to be at equilibrium and their corresponding minima can be shifted from the equilibrium configurations. The slope of the electrostatic potential energy quantifies the average electric force, $-Q\langle E\rangle$, in the protein (Q is the protein charge).

Proteins possess a wide range of relaxation times nearly continuously distributed between picosecond and millisecond time scales or even longer.[13–16] The broad tail in the distribution of relaxation times implies that equilibrium sampling is not always attained on a given observation time $\tau_{\rm obs}$ associated either with the length of the simulation trajectory in numerical simulations or with the resolution time of the spectrometer in laboratory experiments.[17,18] Nonequilibrium sampling implies that Gibbsian statistics (defined in the following) do not apply and one needs a measure of deviations from the anticipated equilibrium sampling.[19,20] These concerns are combined with the picture of nonequilibrium electrostatics in proteins (Fig. 1) to address the following questions. (i) Do Gibbsian statistics describe protein electrostatics? (ii) Are the statistics of electrostatic fluctuations Gaussian and do they follow the standard prescriptions of the FDT?[6,7,11]

The collective coordinate adopted in our study is the change of the electrostatic interaction energy between a tagged protein residue and the surrounding protein–water thermal bath. The collective variable is thus the variation of the system Hamiltonian ΔH. All motions of the medium are projected out on this collective coordinate to create a single-coordinate PMF, that is, the system free energy vs ΔH. The picture of the smooth parabola shown in Fig. 1 might not reflect the reality since it was proposed that the landscape of proteins can be rugged,[13,15,18,21,22] which many local minima of nearly equal free energy values (similar to fragile glasses in glass science[23]). This scenario seems unlikely given an enormous reduction of the entire protein–water phase space to a single collective coordinate. The PMFs we find are indeed mostly smooth and we only find some occasions of PMFs with two minima for the electric field, carrying some similarity with basins of attraction assigned to potential energy landscapes of glassformers.[24,25] When this happens, statistics of electrostatic fluctuations become distinctly non-Gibbsian and non-Gaussian.

In what follows, we describe MD simulations introducing small electrostatic perturbations to two residues and the active site of PC. The main goal is to formulate the general statistical rules for the statistics of the electrostatic potential and the electric field produced by the protein–water thermal bath at specific locations inside the protein. We study the distributions of the potential and field variables and test them on compliance with the static limit of the FDT.[26,27]

It is often believed that active sites of enzymes are evolutionally designed to provide special electrostatic properties, allowing lower activation barrier for enzymatic reactions.[28,29] Electron-transfer proteins, to which PC belongs, can also be viewed as enzymes lowering the activation barrier for redox reactions changing the oxidation state of the active site.[30] To address the difference between generic electrostatic properties of the protein interior and those of the active site, we compare the statistics of electrostatic fluctuations at two selected protein residues, which carry no biological function, with that at the active site of PC. The active site of PC is altered both by changing the oxidation state of the Cu ion and by placing a perturbation dipole on the Cu–S bond connecting the Cu ion with sulfur of methionine ligating it (see the supplementary material). We find that the statistics of electrostatic fluctuations essentially

follow the same general rules for all sites chosen here, suggesting that there is little special about the active site from the perspective of the protein electrostatics. The protein interior is characterized by generic electrostatic properties. They are related to non-vanishing and strongly inhomogeneous electric field inside the protein (Fig. 1) and structural instabilities of the hydration water toward small dipolar perturbations (see in the following).

II. STATISTICS OF ELECTROSTATIC FLUCTUATIONS

The goal of this study is to ascertain which level of theory is required to describe the statistics of protein electrostatic variables. The universally adopted starting point is the Gibbs ensemble[31] defined here by the following conditions: (i) Boltzmann weight $\exp[-\beta H]$, $\beta = (k_B T)^{-1}$ for states with different energy (Hamiltonian) of the system H, (ii) Gibbs postulate[32] assigning equal statistical weights to all states with the same energy H (principle of equal "*a priori*" probabilities), and (iii) the ergodicity assumption—implying the ability to sample all states of the phase space on the observation time t_{obs}. The last requirement, in practical terms, anticipates that t_{obs} is longer than the slowest relaxation time in the system τ_α usually assigned to α relaxation.[23] The latter assumption is potentially violated in short simulations of slowly relaxing proteins. We, however, identify yet another reason for nonergodicity in the following: the expansion of the available phase space allowed by structural instabilities of the protein–water interface.

The starting configuration of the system is the Hamiltonian of the wild-type protein H_0. Several variations of the electrostatic part of the Hamiltonian are introduced here to measure the response of the medium to electrostatic perturbations. First, partial charges are varied by $\Delta q_i = 0.3 q_i$ on a tagged amino acid (proline 16, PRO16) such that the system Hamiltonian changes by $\Delta H = H_\alpha - H_0 = \sum_j \Delta q_j \phi_j$, where ϕ_j are fluctuating site electrostatic potentials. The state $\alpha = 1$ (PRO16, Fig. 2) is obtained by applying this modification

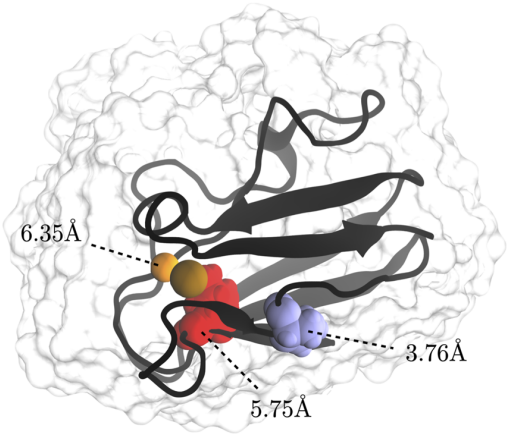


FIG. 2. Structure of PC (PDB: 1YLB) with three mutations sites used in this study: proline 16 (PRO16, iceblue), leucine 5 (LEU5, red), and the copper ion (orange) and sulfur atom (ochre) of the active site. The average distances between the mutation sites and the nearest water of the hydration shell are: 3.76 Å (C_α of PRO16), 6.35 Å (Cu ion), and 5.75 Å (C_α of LEU5).

(see the supplementary material for more details). The Hamiltonian of the altered state is $H_\alpha = H_0 + \Delta H$, and $\alpha = 1$. We use the dimensionless variable $x = \beta \Delta H$.

The second type of perturbation of the wildtype protein involves the introduction of the altered dipole moment at the carbonyl group C=O of the tagged amino acid. This perturbation introduces additional charges $\Delta q_d = \pm 0.255e$ to C and O atoms of the carbonyl group ($\alpha = 2$ for PRO16 and $\alpha = 4$ for leucine 5, LEU5) yielding an additional carbonyl dipole $\Delta\mu = \Delta q_d r_{CO}$, where r_{CO} is the fluctuating C=O bond length. The resulting average dipole moment is $\langle \Delta\mu \rangle \simeq 1.51$ D with the standard deviation of $\simeq 0.03$ D.

The interaction of the carbonyl dipole with the surrounding thermal bath can be described in the lowest multipolar order as the dipole-field interaction. We, therefore, define states $\alpha = 2, 3$, and 4 with the Hamiltonian $H_\alpha = H_0 + \Delta H = H_0 - \Delta\boldsymbol{\mu} \cdot \mathbf{E}$, where $\mathbf{E}$ is the electrostatic field from the rest of the protein and the water solvent at the tagged protein site. The assignment of the perturbation to the electric field was tested for consistency by direct calculations involving potentials ϕ_j at two atomic charge perturbations $\Delta q_{d,j}$.

The dipole moment perturbation is also applied to the active site of PC, where the dipole moment of the Cu–S bond connecting the Cu atom of the active site to the ligating sulfur atom of methionine is modified by placing $0.4e$ on the Cu atom and $-0.4e$ on the sulfur atom (state $\alpha = 3$). Similar to the case of potential alteration, we introduce the dimensionless variable $y = \beta \Delta H$ for the alteration of the residue carbonyl and active site dipoles.

The final energy perturbation introduces a single charge $\Delta q = 0.3 q^{(0)}$ at the Cu atom of the active site (state 5; $q^{(0)}$ is the partial atomic charge of Cu in the wildtype PC). This mutation alters the overall charge of the simulation cell and requires finite-size corrections to the Ewald sums calculations[33–35] (see the supplementary material). This perturbation is designed to probe the statistics of the electrostatic potential inside the protein and is described by the dimensionless variable $z = \beta \Delta H = \beta \Delta q \phi_{Cu}$, where ϕ_{Cu} is the electrostatic potential at the Cu atom.

All simulations, minimizations, and equilibrations were performed with NAMD.[36] For the initial system, conjugate gradient descent minimization was performed for 250 000 steps. Then, NPT equilibration was performed for 10 ns using Langevin dynamics with the following parameters: a piston period of 100 fs, a piston decay time of 50 fs, a piston target pressure of 1.013 bar, and constant temperature targeted at 300 K. NVT production simulations followed up using the same parameters as in NPT, but with the constant pressure controls removed. The particle-mesh Ewald technique with a cutoff distance of 12.0 Å was used to treat the long-range electrostatics.[9] All the simulations were generated using 2 fs time steps. The production runs were done with the simulation box containing 11 091 TIP3P water molecules and no ions neutralizing the total charge to avoid electrostatic contributions from the electrolyte. The relevant corrections[35] for the uniform charge background in calculations of electrostatics with Ewald sums are discussed in the supplementary material.

A. Gibbsian/Gaussian statistics

A number of general statistical rules apply to the perturbation variable $u = x, y, z$, imposing constraints on possible forms of the probability density $P_\alpha(u)$ and the corresponding free energy

function (PMF) $F_\alpha(u)$. Assuming the Gibbsian distribution of the variable u, one can derive[37–39] a general relation between the probability densities $P_0(u)$ in the initial (wildtype) state and probability densities in the perturbed states $P_\alpha(u)$,

$$\ln P_\alpha(u) = \ln P_0(u) - u + \ln(Q_0/Q_\alpha), \tag{1}$$

where Q_0 and Q_α are corresponding partition functions. If the free-energy surfaces $f_\alpha(u) = \beta F_\alpha(u) = -\ln P_\alpha(u)$ are associated with the probability densities, one obtains a linear relation between $f_\alpha(u)$,

$$f_\alpha(u) = f_0(u) + u - \Delta f_\alpha. \tag{2}$$

Here, $\Delta f_\alpha = -\ln(Q_\alpha/Q_0) = f_\alpha - f_0$ is the difference of thermodynamic free energies (scaled with β) in states α and 0. This relation is limited only by the assumption of the Gibbsian statistics,[31] as defined above. It provides the difference of thermodynamic free energies in two states Δf_α [in contrast to PMFs $f_\alpha(u)$] from the intercept of the straight line of $f_\alpha(u) - f_0(u)$ vs u.

If the number of particles involved in fluctuations of u is large, the central limit theorem prescribes $P_\alpha(u)$ to approach the Gaussian distribution, resulting in parabolas for $f_\alpha(u)$,

$$f_\alpha(u) = f_\alpha + \frac{(u - \langle u\rangle_\alpha)^2}{2\sigma^2}. \tag{3}$$

The variances of two Gaussian distributions,

$$\sigma^2 = \langle(\delta u)^2\rangle_0 = \langle(\delta u)^2\rangle_\alpha \tag{4}$$

must be equal if the linear relation (2) is satisfied.

Combining Gibbsian and Gaussian statistics imposes very significant constraints on the values of the first, $\langle u\rangle_\alpha$, and second, σ_α^2, statistical moments. The difference of first statistical moments connects to the variance in the static limit of the FDT,[26]

$$\Delta = \langle u\rangle_0 - \langle u\rangle_\alpha = \langle(\delta u)^2\rangle. \tag{5}$$

Here, based on Eq. (4), we put $\Delta_\alpha = \Delta_0 = \Delta$. One additionally finds the connection between the free energy difference and the mean of the first moments: $\Delta f_\alpha = u_m$, where $u_m = (\langle u\rangle_0 + \langle u\rangle_\alpha)/2$ at $\alpha > 0$.

Equal results are obtained when a sufficiently small perturbation u is chosen to allow the use of the perturbation theory (supplementary material). When the first-order perturbation theory is applied to the Hamiltonians $\Delta H_\alpha = H_\alpha - H_0$, one arrives at the results equivalent to assuming Gaussian distributions $P_\alpha(u)$. Deviations of the first and second moments from Eqs. (4) and (5) signal the breakdown of the Gaussian approximation. The free-energy functions $f_\alpha(u)$ must be non-parabolic in this case. Even in this case, the Gibbsian statistics can still be maintained if the linear relation in Eq. (1) is satisfied.[40]

Deviations from Eq. (5) signal either the appearance of the non-Gaussian statistics or the violation of the Gibbsian statistics discussed in the following. Both scenarios can be quantified in terms of the ratio of an effective temperature $T_{\rm eff}$ and the kinetic temperature T according to the following relation:[27]

$$\frac{T_{\rm eff}}{T} = \frac{\langle(\delta u)^2\rangle_{\max}}{\Delta}. \tag{6}$$

TABLE I. Statistics of $u = x, y, z$ obtained by mutating the wildtype state 0 (T = 300 K).

States	$T_{\rm eff}/T$	Δ	σ_0^2	σ_α^2	κ^{-1}
0,1 (x)[a]	1.26	0.58	0.69	0.72	1.22
0,2 (y)[b]	2.85	2.17	1.08	6.20	
0,3 (y)	1.68	7.08	1.47	11.86	
0,4 (y)	1.08	3.71	4.00	2.60	1.0
0,5 (z)[c]	1.50	0.80	1.17	1.20	1.02

[a]States involved in the calculations and the variable x, y, or z specifying the alteration of the system Hamiltonian.
[b]The distribution of the perturbed state is bimodal and the linear relation [Eq. (2)] does not hold.
[c]Finite size corrections are applied to account for altering charge of the simulation cell when the charge mutation is introduced (see the supplementary material). One finds $T_{\rm eff}/T = 0.98$ without corrections, in a good agreement with κ^{-1} calculated from uncorrected free-energy surfaces.

Here, in anticipation of the breakdown of the Gaussian statistics and the violation of Eq. (4), $\langle(\delta u)^2\rangle_{\max}$ specifies the maximum variance between σ_0^2 and σ_α^2. We show in the following that $\sigma_\alpha^2/\sigma_0^2$ can be as high as 7–8 (Table I) and merely taking the mean of two variances does not provide a meaningful metric of the non-Gaussian statistics.

B. Nonergodic sampling

Given that the Gibbsian statistics is the only assumption used in deriving Eqs. (1) and (2), the violation of the linear relation between the PMFs is an indication of non-Gibbsian statistics. Our simulations presented in the following indeed show that the linear relation between $f_\alpha(u)$ [Eq. (2)] is often violated and, phenomenologically, can be replaced with an alternative linear relation involving a slope lower than unity ($\kappa \le 1$),

$$f_\alpha(u) = f_0(u) + \kappa u - \Delta f_\alpha^{\rm neq}. \tag{7}$$

Here, given the nonequilibrium sampling, we have also specified a nonequilibrium offset $\Delta f_\alpha^{\rm neq}$. It is important to emphasize that $f_\alpha(u)$ in Eq. (7) are given in terms of time averages, which do not coincide with ensemble averages when ergodicity is broken. Therefore, the functions $f_\alpha(u)$ in Eq. (7) depend on the observation time and do not carry the universal meaning of the corresponding Gibbsian ensemble averages in Eq. (2). We do not introduce separate notations for the former since only time averages are allowed to us in numerical simulations.

The violation of the Gibbsian statistics leading to $T_{\rm eff} > T$ has been found in MD simulations of a number of redox active proteins,[27,39] with the separation between $T_{\rm eff}$ and T increasing for large membrane-bound proteins of electron-transport chains of bacterial photosynthesis and respiration.[41,42] One anticipates that longer sampling is required for larger proteins, with the resulting stronger violation of the equality between $T_{\rm eff}$ and T. We have chosen a small redox-active protein here with the hope to achieve full equilibration on the simulation time scale of 0.4–1.0 μs to study the statistics of fully equilibrated electrostatic variables. However, even for a small protein, there are noticeable deviations between $T_{\rm eff}$ and T attributed to the violation of equilibrium sampling and to non-Gibbsian statistics (Table I).

The non-unity slope [Eq. (7)] also leads to the violation of the FDT equality between the first and second moments [Eq. (5)] and thus yields $T_{\rm eff} \neq T$. It still requires the equality between two Gaussian variances [Eq. (4)] when the Gaussian statistics hold. In that case, $f_\alpha(u)$ in Eq. (7) are parabolic and one obtains

$$\kappa = T/T_{\rm eff}, \quad \Delta f_\alpha^{\rm neq} = \Delta f_\alpha(T/T_{\rm eff}). \tag{8}$$

The slope of the linear relation in Eq. (7) is a direct measure of the effective temperature provided the distributions $P_\alpha(u)$ are Gaussian, which is often the case with long-range electrostatic fluctuations involving many particles (central limit theorem). An illuminating example from optical spectroscopy of proteins is discussed in the following. The deviation of the nonequilibrium offset $\Delta f_\alpha^{\rm neq}$ from the equilibrium free energy difference Δf_α can be observed in nonequilibrium injection of charge carriers into molecular junctions when the measured nonequilibrium resonance peak deviates from the equilibrium reduction potential.[42,43]

C. Statistics of ΔH

Table I presents the results of MD simulations for electrostatic perturbations produced at two protein residues and at the active site of PC in TIP3P water, as described in more detail in the supplementary material. As mentioned above, states $\alpha = 1, 2$ specify electrostatic perturbations applied to the PRO16 residue close to the surface of PC, while state 3 is obtained by placing a dipole on the Cu–S bond of the active site. State 4 is produced by placing a probe dipole to LEU5, and state 5 is obtained by increasing the partial charge of the Cu ion in the active site (Fig. 1).

The $0 \rightarrow 1$ perturbation does not involve changing the residue charge and can be viewed as a mixed probe involving a multipolar expansion starting from the dipole moment. The variances of the variable x are nearly equal in two states [Eq. (4)], but the slope of the linear relation is less than unity supporting Eqs. (7) and (8). In the case of potential perturbation $(0 \rightarrow 5)$, $T_{\rm eff} > T$ comes from applying the finite-size corrections[35] (see the supplementary material), which discriminate between Δ and σ^2: the statistics directly from simulations are Gibbsian and Eq. (5) is satisfied. It is currently unclear how to apply finite-size corrections to the free-energy surfaces, which are calculated directly from MD trajectories.

States targeting the electric potential largely fall in line with the expectations of linear response. Increases in the calculated effective temperatures span 26%–50% of the bath temperature when accounting for finite-size effects. State 1 shows a modest deviation from linear response, with the variance approximately conserved between states. While the active site showed the greatest response here, it was not by a large margin. This result runs counter to the idea that the active site has been specially crafted to possess unique electrostatic properties.

D. Electric field statistics

The statistics of the electric field probed by the perturbation y are both non-Gibbsian and strongly non-Gaussian. It appears that even small dipolar perturbations to various sites within the protein produce strongly nonlinear responses of the protein–water thermal bath. Changing the carbonyl dipole at PRO16 $(0 \rightarrow 2)$ results in a distinctly bimodal distribution of the hydration water in the perturbed state, thus leading to a bimodal free-energy surface $f_2(y)$ [Fig. 3(a)]. In contrast, placing the probe dipole to the Cu–S bond of the active site $(0 \rightarrow 3)$ makes the protein distribution both shifted and strongly broadened. It is likely that the broad basin in state 3 is a superposition of closely separated minima. In both cases, one finds $\sigma_\alpha > \sigma_0$ and non-Gaussian $f_\alpha(y)$ for $\alpha > 0$ [Figs. 3(a) and 3(b)]. The perturbation of LEU5 $(0 \rightarrow 4)$ results in less dramatic consequences because this residue is more separated from the hydration shell (Figs. 1 and S1). Moving from $0 \rightarrow 2$ to $0 \rightarrow 3$ to $0 \rightarrow 4$ shows a decrease in the non-Gaussian character of the electric field response. One can, therefore, conclude that it is the hydration layer of the protein that is responsible for the non-Gibbsian/non-Gaussian statistics of the electric field.

It is also clear that the linear relation between unperturbed and perturbed free-energy surfaces [Eq. (2)] cannot be satisfied if a one-well PMF is transformed to a two-well PMF by perturbation [Fig. 3(a)]. The induction of a structural transition by a perturbation is thus a signature of non-Gibbsian statistics of the perturbing collective variable.

The non-Gaussian statistics of the coordinate y result in Δ falling between two distinct variances for all dipolar perturbations (Table I). The Q-model distribution, derived to describe statistics of perturbations in electron-transfer reactions and optical spectroscopy,[40] satisfies this constraint. This generic distribution function describes transitions in systems with altering force constants of the harmonic medium between two states or, in other words, allowing different variances of the perturbation variable u in the original and perturbed states. The Q-model was derived in the Gibbs ensemble and thus follows the standard linear relation [Eq. (2)]. Consequently, the distributions in two states are skewed, but not bimodal. Figures 3(c) and 3(d) show $P_\alpha(u)$ produced with the Q-model based on the statistical moments provided by MD (supplementary material). It captures the general features for $0 \rightarrow 3, 4$ perturbations but, being a Gibbsian model, does not capture the bimodal form of $f_2(y)$ shown in Fig. 3(a) (Fig. S2).

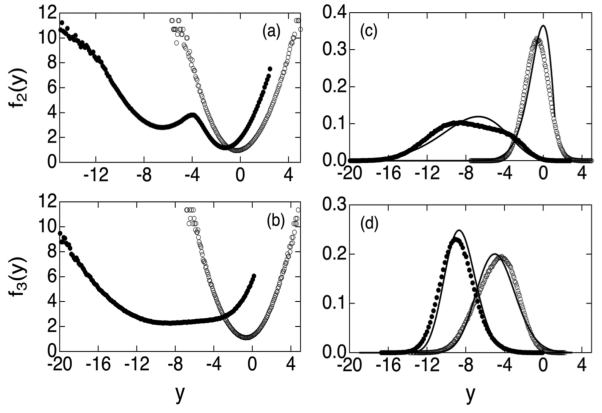


FIG. 3. $f_\alpha(y) = -\ln P_\alpha(y)$ for $\alpha = 0$ (open points) and $\alpha = 2$ (filled points) (a) and $\alpha = 0, 3$ (b). Distributions $P_\alpha(y)$ in state 3 (c) and state 4 (d). The points indicate MD calculations, and the solid lines refer to the Q-model[40] calculations based on σ_α^2, Δ, and y_m from MD simulations.

TABLE II. Separation of the total variance in the protein, water, and cross components. In addition, $T_{\rm eff}/T$ are shown for the protein (p) and water (w) components of the thermal bath.

States	$(T_{\rm eff}/T)_p$	σ_p^2	Δ_p	$(T_{\rm eff}/T)_w$	σ_w^2	Δ_w	σ^2	$2\sigma_{pw}$
0 (x)	1.64	0.43	0.26	1.65	0.52	0.32	0.69	−0.25
1 (x)		0.39			0.52		0.72	−0.19
0 (y)	3.15	0.89	0.28	3.95	0.97	1.89	1.08	−0.78
2 (y)		0.85			7.47		6.20	−2.13
0 (y)	1.74	1.35	4.24	1.02	1.46	2.84	1.47	−1.34
3 (y)		7.39			2.90		11.86	1.57
0 (y)	1.21	4.36	3.62	6.63	0.62	0.09	4.00	−0.98
4 (y)		3.12			0.39		2.60	−0.91
0 (z)[a]	98.0[b]	1.77	0.018	2.11	2.00	0.948	0.984	−2.53
5 (z)		1.25			1.56		1.012	−1.80

[a]No finite-size corrections applied to the protein and water components.
[b]The distributions are distinctly non-Gaussian, producing a very low shift Δ between the two states.

E. Protein and water components

All the perturbations considered here can be separated into corresponding protein (p) and water (w) components: $u = \beta\Delta H = p + w$. The overall variance of u is a composite quantity,

$$\sigma^2 = \sigma_p^2 + \sigma_w^2 + 2\sigma_{pw}, \tag{9}$$

including the self-components and a cross term $\sigma_{pw} = \langle \delta p \delta w \rangle$. The cross term is nearly universally negative (Table II), physically reflecting dielectric screening of the protein charges by hydration water.

Table II also presents $(T_{\rm eff}/T)_{p,w}$ from Eq. (6) for the protein and water components of the total Hamiltonian perturbation. The effective temperatures of components typically exceed those calculated for the total perturbation (Table I). This does not imply enhanced nonergodic sampling of component perturbations since $(T_{\rm eff}/T)_{p,w} > 1$ is allowed by standard statistical-mechanical arguments. For instance, perturbation theory suggests that the component variances can be obtained from the corresponding Δ_a, $a = p, w$ and the cross-correlation term[44] (supplementary material),

$$\sigma_a^2 = \Delta_a - \sigma_{pw}. \tag{10}$$

With a negative σ_{pw}, this equation leads to $\sigma_a^2 > \Delta_a$, as reported by the simulations. However, a quantitative agreement between Eq. (10) and simulations is poor. Nevertheless, it shows that coupling between protein and water is responsible for the breakdown of the standard relation between the first and second statistical moments: $\sigma_a^2 > \Delta_a$ and $(T_{\rm eff}/T)_{p,w} > 1$ do not violate the standard rules of Gibbsian statistics.

III. TEMPERATURE DEPENDENCE

All electrostatic perturbations introduced at three sites of PC are designed to satisfy the linear response approximation and to avoid the nonlinear electrostatic response. The linear response approximation[6,7] also prescribes the variance σ_α^2 to satisfy the Johnson–Nyquist temperature dependence,[10] predicting that thermal noise of a macroscopic statistical variable scales linearly with temperature,

$$(k_{\rm B}T\sigma_\alpha)^2 = 2k_{\rm B}T\lambda_\alpha. \tag{11}$$

The parameter λ_α here is known as the reorganization energy in theories of electron transfer when electrostatic perturbation is experimentally achieved by delivering an electron to a localized site in a half (involving an electrode) or a full (involving a donor and an acceptor moieties) redox reaction.[45] The electrostatic perturbations used here introduce either fractional charge Δq or a dipole moment $\Delta\mu$ to protein sites. When linear response is satisfied, λ_α scales linearly with either Δq^2 or $\Delta\mu^2$ and our results can be applied to practical situations of redox reaction or spectroscopy by rescaling Δq to $-1e$ and $\Delta\mu$ to the alteration of the chromophore dipole moment with optical excitation.

The reorganization energies $\lambda_\alpha(T)$ are expected to be constants according to the Johnson–Nyquist equation [Eq. (11)]. Therefore, we cast our MD data in the form of $\lambda_\alpha(T)$ to quantify potential deviations from this standard expectation due to either structural transitions or ergodicity breaking (glass transition) in the protein–water thermal bath. We, therefore, use Eq. (11) to split λ_α into the protein and water components and the cross term,

$$\lambda_\alpha = \lambda_\alpha^p + \lambda_\alpha^w + 2\lambda_\alpha^{pw}. \tag{12}$$

We find notable violations of $\lambda_\alpha(T) = Const$, and the character of these violations is affected by the choice of the electrostatic property, that is either the electrostatic potential or the electric field.

Figures 4(a) and 4(b) shows the reorganization energies $\lambda_\alpha(T)$ for two states $\alpha = 0, 1$ and the corresponding water components $\lambda_\alpha^w(T)$. One finds a crossover, at around 150–170 K, from constant values of λ at high temperatures, in accord with Eq. (11), to approximately linearly decreasing functions. Previous MD simulations of PC with the simulation time of ≃ 10 ns found the crossover at higher temperatures,[44] ≃ 200–220 K, close in magnitude to the temperature of dynamical transition[46] T_d reported by quasielastic neutron scattering with the observation time of ≃ 0.1–1 ns. The drop in T_d at the longer observation time ∼ 0.4–1.0 μs in the present

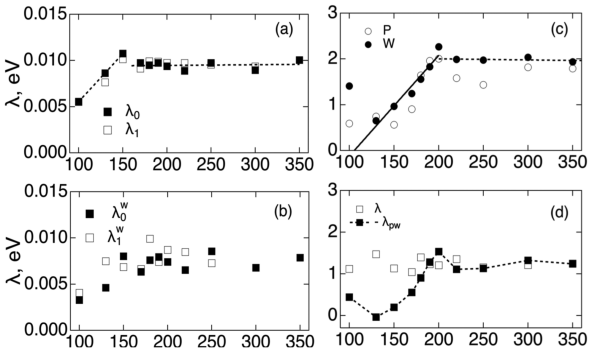


FIG. 4. (a) $\lambda_\alpha(T)$, where $\alpha = 0, 1$. (b) λ_α^w, where $\alpha = 0, 1$ (0,1 perturbation). (c) $\lambda_0^{p,w}(T)$ when the perturbation is achieved by adding the charge $\Delta q = -1e$ to the Cu atom of PC. (d) λ_0 and $-\lambda_0^{pw}$ with the perturbation as in panel (c).

simulations is consistent with the general interpretation of the dynamical transition as the point at which a major relaxation process, with the relaxation time $\tau(T)$, crosses the observation time: $\tau_{\rm obs} \simeq \tau(T_d)$.[47–50] A general outcome of this perspective is lowering of T_d for longer sampling times, as we find in our present simulations. The kink temperature, 150–170 K, found here comes close to the glass transition temperature of protein hydration water, $\simeq$ 170 K,[48] signaled by an abrupt change in the expansion coefficient of water in hydrated lysozyme powders. This connection to observations suggests that the kink shown in Figs. 4(a) and 4(b) can be attributed to hydration water.

A similar analysis is presented in Figs. 4(c) and 4(d), where $\lambda_0(T)$ is calculated for state 0, assuming that the perturbation involves adding the charge $\Delta q = -1e$ to the Cu atom of the PC's active site. This perturbation corresponds to a half redox reaction when an electron is transferred to PC to change its oxidation state. The results presented for state 5 in Table I are produced with $\Delta q = 0.1125e$ and can be connected to an electrochemical experiment by rescaling with Δq^{-2}, as is shown in Figs. 4(c) and 4(d).

The temperature dependences of the water and protein components of λ_0 shown in Fig. 4(c) resemble the results shown in Fig. 4(a): the component λ shows a kink at a slightly higher temperature. However, in contrast to Fig. 4(b), the total $\lambda_0(T)$ is nearly flat. This distinction must require the protein–water correlation term λ_0^{pw} to show a sharp change below the crossover temperature [Fig. 4(d)]. Interestingly, we find that the total reorganization energy and the cross terms nearly coincide in magnitude at high temperatures, but are opposite in sign, $\lambda_0 \simeq -\lambda_0^{pw}$. The total reorganization energy at high temperatures becomes

$$\lambda_0 \simeq \frac{1}{3}\left(\lambda_0^p + \lambda_0^w\right). \tag{13}$$

The cross term λ_α^{pw} physically reflects the screening of protein charges by interfacial water. A positive ionized residue at the protein surface induces a negative surface charge from the water dipoles oriented by the local electric field, leading to a negative sign of the cross term. A substantial drop in $|\lambda_0^{pw}|$ at lower temperatures and the ensuing breakdown of Eq. (13) signal a strong change in the ability of interfacial water to screen protein's charged residues. The individual components $\lambda_0^{p,w}$ drop at low temperatures, but so does the magnitude of mutual screening.

The Johnson–Nyquist temperature law predicts a linear scaling with temperature for the variance of a macroscopic variable. There are established violations of this law, such as the temperature scaling of the macroscopic dipole moment of a polar liquid, which is typically a decaying function of temperature.[51] Following the protocol established in Fig. 4 for the change in the system Hamiltonian induced by altering dipole moment, we define the parameter λ_α^E based on the variance of the electric field $\mathbf{E}$ at a given site within the protein calculated here for $\alpha = 0$,

$$\lambda_\alpha^E = \frac{1}{2}\beta\langle(\delta\mathbf{E})^2\rangle_\alpha. \tag{14}$$

The variance of the electric field at the tagged site can be connected, in isotropic media, to the dipolar susceptibility of the medium[52,53] $\chi_{d,\alpha} = \lambda_\alpha^E/3$. This susceptibility defines the linear response of the medium to a change in the site dipole and is often given by using the Onsager formula[52,54] in dielectric models of the medium response.

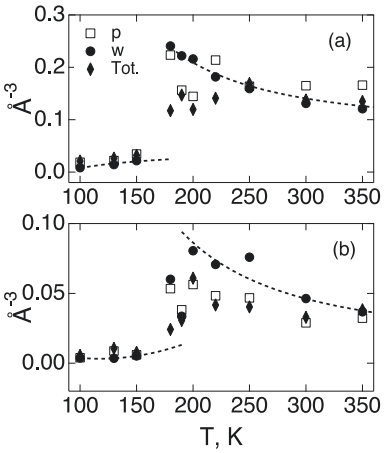


FIG. 5. Temperature dependence of $\lambda_0^E = \beta\langle(\delta E^2)^2\rangle_0/2$ produced by protein and water at the carbonyl dipole of PRO16 in state 0 (a) and at the Cu–S dipole of the active site (b). Filled diamonds indicate the overall λ_0^E values. The high-temperature data for the water component are fitted to the $a + b/T^2$ functionality (dashed line). The lower-temperature portion of the water data is fitted to an $a + b/T$ function to guide the eye.

The interaction of the protein–water electric field with a probe dipole is more short-ranged than the charge–potential interaction. One might expect some qualitative differences in the temperature dependence $\lambda_0^E(T)$ compared to $\lambda_0(T)$ and, given a shorter range of interactions, more pronounced deviations from the macroscopic Johnson–Nyquist temperature law. Indeed, $\lambda_0^E(T)$ (Fig. 5) is very different from $\lambda_0(T)$ shown in Fig. 4: instead of the kink shown in Fig. 4, we find a discontinuous transition at $T_{\rm tr} \simeq 170$ K. Figure 5 shows the results for $\lambda_0^E(T)$ for PRO16 and the active site of PC, while the results for the LEU5 dipolar perturbation are shown in Fig. S3.

In contrast to Fig. 4(c), a pronounced drop at $T_{\rm tr}$ is seen for both the protein and water components and for the total λ_0^E. The high-temperature branch of λ_0^E is notably increasing with temperature, approximately following the $a + b/T^2$ functionality (dashed lines shown in Fig. 5), in contrast to a plateau high-temperature region shown for λ_0 in Fig. 4. Overall, the transition to the lower-temperature branch is accompanied with a strong drop in the field fluctuations.

A. Non-Gaussian parameter

The appearance of crossovers in the temperature dependencies of second moments of either the Hamiltonian energy gap or the electric field raises the question of their physical origin. Such crossovers can generally be attributed to either protein–water structural transitions or to dynamical transitions (ergodicity breaking) when the time scale of fluctuations, increasing with lowering temperature, crosses the observation time $\tau_{\rm obs}$ (length of the trajectory

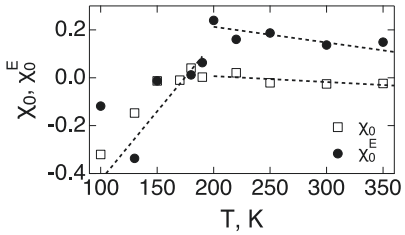


FIG. 6. Binder parameters $\chi_0(T)$ [$u = z$, Eq. (15)] and $\chi_0^E(T)$ [Eq. (16)] for the electric field at the active site of PC. The dashed lines are linear fits drawn to guide the eye.

in our case). To distinguish between the two scenarios, we calculated the Binder non-Gaussian parameter,[55,56]

$$\chi_\alpha = 1 - \frac{\langle(\delta u)^4\rangle_\alpha}{3\langle(\delta u)^2\rangle_\alpha^2}, \tag{15}$$

where $\delta u = u - \langle u\rangle$ and α specifies the equilibrium state used for statistical averages.

The Binder parameter is expected to show a narrow dip for a structural transition when two alternative states of the system become nearly equally accessible in a narrow range of temperatures around T_{tr}.[56] In contrast, we generally find a transition between high- and low-temperature branches at $T_{tr} \simeq 170$ K. For instance, Fig. 6 shows χ_0 for $u = z$. It is nearly zero at $T > 200$ K, followed by substantial negative values at lower temperatures. This crossover signals an onset of non-Gaussian statistics of z at low temperatures.

The analog of the Binder parameter for a vector variable has to account for the average over random vector orientations with the following result for the electric field **E**:

$$\chi_\alpha^E = 1 - \frac{3\langle(\delta\mathbf{E})^4\rangle_\alpha}{5\langle(\delta\mathbf{E})^2\rangle_\alpha^2}. \tag{16}$$

One obtains $\chi_\alpha^E = 0$ for $\delta\mathbf{E}$, satisfying the Maxwell distribution. Figure 6 shows χ_α^E for the electric field at the active site of PC. A similar trend is found for the electric field at the PRO16 site of the protein (Fig. S4). As for the scalar Binder parameter (Fig. 6), there is a change in the temperature dependence of χ_α^E, but no narrow dip expected for a structural transition. We, therefore, conclude that crossovers observed here have to be related to ergodicity breaking (glass transition) at low temperatures caused by the growing medium relaxation time reaching the level of the observation (simulation) time. A strong sensitivity of the hydration shell to small dipolar perturbations introduced to the protein (Fig. 3) suggests that a structural transition might be present at a lower temperature, but is pre-empted by the glass transition at $\simeq 170$ K.

IV. DISCUSSION

This study asks two major questions. (i) Do protein electrostatics equilibrate to the Gibbsian ensemble on the simulation time of 0.4–1.0 μs? (ii) Do the electrostatic potential and electric field inside proteins follow the Gaussian distribution? The first question is driven by the recognition of a broad distribution of relaxation processes in proteins, including time scales in the millisecond range, well beyond the typical length of numerical simulations. If the relaxing modes characterized by such long relaxation times affect protein electrostatics, the ensuing nonergodic sampling violates statistical–mechanical relations between the first and second statistical moments [Eq. (5)]. One anticipates that nonergodic sampling is pertinent to large proteins[20,57] and membrane-bound protein complexes[41] encountered in electron-transport chains of photosynthesis and respiration. Assuming that relaxation is slower for larger proteins, we have chosen here a small redox-active protein plastocyanin to achieve equilibrium sampling on currently accessible simulation times.

The violation of Eq. (5) (which connects the first and second statistical moments of the Hamiltonian perturbation ΔH) can be attributed to either nonergodic sampling or to the non-Gaussian statistics of ΔH. The latter also requires different variances σ_α^2 in the perturbed and unperturbed states [Eq. (4)]. In contrast, the violation of the linear relation between the free energy functions $f_\alpha(u)$ [Eq. (2)] is a direct signature of nonergodic sampling and a non-Gibbsian ensemble. Such violations are not only a matter of curiosity. It was suggested[27,39] that nonergodic sampling carries physiological significance since it lowers activation barriers for reactions involving redox-active proteins by the ratio $T/T_{\rm eff}$ [cf. to Eq. (8)],

$$\Delta F^\dagger = \Delta F^\dagger_{\rm erg}(T/T_{\rm eff}). \tag{17}$$

Here, $\Delta F^\dagger_{\rm erg}$ is the activation barrier when the conditions of ergodic (Gibbsian) sampling are satisfied and $\Delta F^\dagger$ is the activation barrier with nonergodic sampling. The ratio $T_{\rm eff}/T$ is expressed either in terms of the slope κ of the modified linear energy gap law [Eq. (7)] or through the ratio between the second and first central moments of the energy gap collective coordinate $u = \beta\Delta H$ [Eq. (6)]. The second relation connects the effective temperature to experimental observables.

Both parameters σ_α [Eq. (4)] and Δ [Eq. (5)] carry spectroscopic meaning: $k_BT\Delta$ corresponds to the medium-induced Stokes shift, and $k_BT\sigma_\alpha$ specifies inhomogeneous broadening in optical spectroscopy. When light photon is absorbed, the Hamiltonian perturbation is the energy of the light photon $hc\tilde{\nu}$ ($\tilde{\nu}$ is the wavenumber and c is the speed of light in vacuum) and the reduced perturbation coordinate becomes $x = \beta hc\tilde{\nu}$. We use absorption and emission lines of mStrawberry fluorescent protein[58] to illustrate the spectral analysis.

The total absorption and emission lines shown in Fig. 7 are superpositions of individual inhomogeneously broadened vibronic lines with corresponding Franck–Condon weights.[59] The decomposition of the band shapes (supplementary material) allows one to extract two individual 0-0 Gaussian lines for absorption and emission,

$$P_\alpha(x) \propto \exp\left[-\frac{(x - x_\alpha)^2}{2\sigma_\alpha^2}\right], \tag{18}$$

where α = abs, em. The relative shift of their maxima is the medium-induced Stokes shift $k_BT\Delta$ [Fig. 7(a)]. Normalizing the functions $P_\alpha(x)$ and taking the logarithm of their ratio produces the difference of the corresponding PMFs and allows one to test the linear

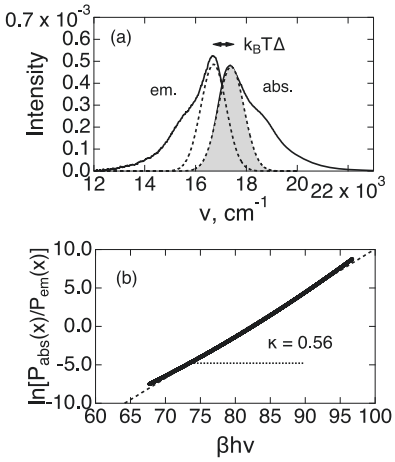


FIG. 7. (a) Normalized absorption (abs.) and emission (em.) spectra of mStrawberry fluorescent protein.[58] The band envelopes are decomposed in vibronic lines, and the dashed lines show the 0-0 transitions for abs. (shaded) and em. (b) $\ln[P_{abs}(x)/P_{em}(x)]$ vs $x = \beta h\nu = \beta hc\bar{\nu}$ (tick line). The dashed line refers to a linear fit with the slope $\kappa = 0.56$.

relation in Eqs. (1) and (2). This test applied to mStrawberry fluorescent protein[58] indeed leads to a nearly straight line with the slope $\kappa \simeq 0.56$ [Fig. 7(b)]. One, therefore, obtains $T_{\text{eff}}/T \simeq 1.8$. A practical difficulty of applying this analysis is to find a proper chromophore with well-resolved absorption and emission lines, but the algorithm itself provides a direct test of the linear relation between the ground and photoexited PMFs.

A surprising outcome of our simulations is a strong response of the protein–water thermal bath to a relatively small, $\Delta\mu \simeq 1.5$ D, change in a site dipole, which is consistent with or below the typical dipole changes for chromophores in optical transitions (e.g., $\Delta\mu \simeq 6.5$ D for green fluorescent proteins[60]). We find that altering the site dipole results in a bimodal distribution of the energy gap coordinate or to a broad basin most likely arising from a superposition of multiple minima (Fig. 3). The response and broadening are reduced for the residue LEU5 in the protein core compared to PRO16 close to the protein surface (Fig. S1). A strong nonlinear response of the hydration water is, therefore, the origin of the non-Gibbsian [violation of Eq. (2)] and non-Gaussian [violation of Eq. (4)] statistics.

A high susceptibility of the hydration shell of PC toward a small dipolar perturbation is indicative of the proximity of the protein–water system to the point of structural instability. This situation is reminiscent of divergent susceptibilities of bulk materials near phase transitions.[61] A small dipolar perturbation induces a structural re-arrangement of the hydration layer. This outcome seems to be general for interfacial water. A small dipolar probe placed near water interfacing a hard-sphere/Lennard-Jones cavity produces a structural transition similar to that observed here for the protein–water interface.[62] In the former case, one finds a bimodal shape of the electric field distribution. In addition, the dipolar susceptibility of interfacial water passes through a sharp spike as the magnitude of the probe dipole increases. Experimental data also report a number of structural instabilities of interfacial water, often leading to the release of O–H dangling bonds, when the charge of the substrate is altered.[63–66] All these results indicate that water in the interface can adopt alternative structures and can fluctuate between them when the barrier separating the corresponding minima is altered by the solute–solvent electrostatics. Surface protein residues are a source of such an external perturbation, making the hydration shell split into nano-domains and carry properties of relaxor ferroelectrics.[67]

A reason for nonergodic sampling of protein electrostatics cited above is the separation of time scales, i.e., the inability to sample the system's phase space on the observation time shorter than a subset of slow relaxation times. However, structural instabilities and the appearance of multi-minima PMFs upon an electrostatic perturbation offer a new mechanism for nonergodic sampling of solvated biomolecules. The phase transition in bulk materials is a prime example of statistical nonergodicity, given that only the configuration space describing a given thermodynamic phase is accessible on the experimental observation time.[68] The appearance of new structural states of the protein–water system, unavailable in the unperturbed configuration, constitutes an expansion of the available configuration space that breaks the ergodicity assumption of the Gibbs ensemble. The situation at hand is shown in Fig. 8.

Figure 8 shows the expansion of the phase space available to the ensemble average when the protein–water system develops a structural instability upon perturbation. The phase space effectively available to the system in the one-minimum unperturbed state is within the energy k_BT near the bottom of the potential well. When the potential well develops two minima in the perturbed state, the available phase space significantly expands. Such an expansion of the phase space accessible to sampling is not anticipated in the derivation of the linear relation between the unperturbed and perturbed PMFs [Eq. (2)]. The phase space expansion is analogous to altering a set of constraints defining the ensemble,[69] and it results in violations of the standard statistical relations based on the preservation of constraints. The key to this picture is a close proximity of the

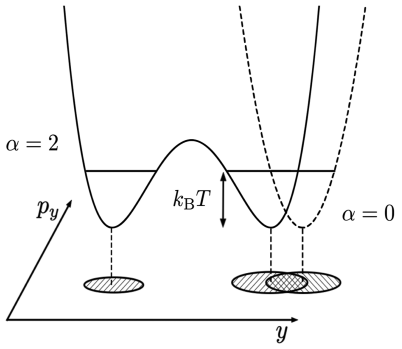


FIG. 8. Phase space (y, p_y) of one minimum in the unperturbed state $\alpha = 0$ and the combined phase space of two minima in the perturbed state $\alpha = 2$. The phase space available to sampling is substantially enhanced due to structural instability leading to a bimodal PMF. The effective phase space available for sampling is within the range k_BT in the PMF well.

protein–water interface to points of structural instability. Such instabilities do not have to be global and can involve rearrangements of clusters of waters around specific parts of the protein surface or of some surface residues.

V. CONCLUSIONS

To summarize, we found nearly Gibbsian/Gaussian statistics of the electrostatic potential perturbation, in line with the corresponding fluctuation–dissipation theorem. Dipolar perturbations, however, disturb the protein–water interface and their statistics are strongly non-Gibbsian/non-Gaussian. While there are no non-Gibbsian models to account for the new statistical rules, the ergodic (Gibbsian) Q-model accounts for the basic features of the non-Gaussian statistics in response to dipolar perturbations. The dipolar susceptibility characterizing the electric field variance [Eq. (14)] shows a strong increase with cooling, interrupted by a dynamical glass transition at ≃ 170 K.

SUPPLEMENTARY MATERIAL

The supplementary material encompasses the details of the simulation protocol, analysis of the results, and derivation of equations listed in the text. In addition, details of applying the Q-model and the band-shape analysis of optical lines are provided.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation (Grant No. CHE-2154465). The supercomputer time was provided through Extreme Science and Engineering Discovery Environment (XSEDE) allocation MCB080071 and through ASU's Research Computing.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Taylor Colburn: Data curation (equal); Formal analysis (equal). **Setare Mostajabi Sarhangi**: Data curation (equal); Formal analysis (equal). **Dmitry V. Matyushov**: Conceptualization (equal); Funding acquisition (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1]E. I. Solomon, R. K. Szilagyi, S. D. George, and L. Basumallick, Chem. Rev. **104**, 419 (2004).
[2]D. N. LeBard and D. V. Matyushov, J. Phys. Chem. B **112**, 5218 (2008).
[3]D. G. Nicholls and S. J. Ferguson, *Bioenergetics4* (Academic Press, Amsterdam, 2013).
[4]R. E. Blankenship, *Molecular Mechanisms of Photosynthesis* (Blackwell Science, Williston, VT, 2003).
[5]S. G. Boxer, J. Phys. Chem. B **113**, 2972 (2009).
[6]R. Kubo, *Lectures in Theoretical Physics* (Interscience Publishers, Inc., New York, 1959), Vol. 1, p. 120.
[7]J.-P. Hansen and I. R. McDonald, in *Theory of Simple Liquids*, 4th ed. (Academic Press, Amsterdam, 2013).
[8]S. D. Fried and S. G. Boxer, Annu. Rev. Biochem. **86**, 387 (2017).
[9]M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon Press, Oxford, 1996).
[10]R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics, Vol. I: Mainly Mechanics, Radiation, and Heat* (Addison-Wesley, Reading, 1963).
[11]R. Kubo, Rep. Prog. Phys. **29**, 255 (1966).
[12]U. M. B. Marconi, A. Puglisi, L. Rondoni, and A. Vulpiani, Phys. Rep. **461**, 111 (2008).
[13]H. Frauenfelder, "The physics of proteins," in *An Introduction to Biological Physics and Molecular Biophysics*, edited by S. S. Chan and W. S. Chan (Springer, New York, 2010).
[14]J. R. Lewandowski, M. E. Halse, M. Blackledge, and L. Emsley, Science **348**, 578 (2015).
[15]E. Dorbath, A. Gulzar, and G. Stock, J. Chem. Phys. **160**, 074103 (2024).
[16]G. Haran and H. Mazal, J. Chem. Phys. **153**, 130902 (2020).
[17]K. Lyu, H. Chen, J. Gao, J. Jin, H. Shi, D. K. Schwartz, and D. Wang, Biomacromolecules **23**, 4709 (2022).
[18]A. N. Hassani, L. Haris, M. Appel, T. Seydel, A. M. Stadler, and G. R. Kneller, J. Chem. Phys. **159**, 141102 (2023).
[19]X. Hu, L. Hong, M. Dean Smith, T. Neusius, X. Cheng, and J. C. Smith, Nat. Phys. **12**, 171 (2015).
[20]J. Li, J. Xie, A. Godec, K. R. Weninger, C. Liu, J. C. Smith, and L. Hong, Chem. Sci. **13**, 9668 (2022).
[21]J. D. Bryngelson and P. G. Wolynes, Proc. Natl. Acad. Sci. U. S. A. **84**, 7524 (1987).
[22]B. Hille, Annu. Rev. Biophys. **51**, 1 (2022).
[23]C. A. Angell, Science **267**, 1924 (1995).
[24]F. H. Stillinger, Science **267**, 1935 (1995).
[25]S. Büchner and A. Heuer, Phys. Rev. E **60**, 6507 (1999).
[26]A. Crisanti and F. Ritort, J. Phys. A: Math. Gen. **36**, R181 (2003).
[27]D. V. Matyushov, J. Mol. Liq. **266**, 361 (2018).
[28]A. Warshel, P. K. Sharma, M. Kato, Y. Xiang, H. Liu, and M. H. M. Olsson, Chem. Rev. **106**, 3210 (2006).
[29]S. D. Fried, S. Bagchi, and S. G. Boxer, Science **346**, 1510 (2014).
[30]H. B. Gray and J. R. Winkler, Q. Rev. Biophys. **36**, 341 (2003).
[31]L. D. Landau and E. M. Lifshits, *Statistical Physics* (Pergamon Press, New York, 1980).
[32]D. A. McQuarrie, *Statistical Mechanics* (University Science Books, Sausalito, CA, 2000).
[33]G. Hummer, L. R. Pratt, and A. E. García, J. Chem. Phys. **107**, 9275 (1997).
[34]R. Ayala and M. Sprik, J. Phys. Chem. B **112**, 257 (2008).
[35]D. V. Matyushov, J. Chem. Phys. **155**, 114110 (2021).
[36]J. C. Phillips, D. J. Hardy, J. D. C. Maia, J. E. Stone, J. V. Ribeiro, R. C. Bernardi, R. Buch, G. Fiorin, J. Hénin, W. Jiang, R. McGreevy, M. C. R. Melo, B. K. Radak, R. D. Skeel, A. Singharoy, Y. Wang, B. Roux, A. Aksimentiev, Z. Luthey-Schulten, L. V. Kalé, K. Schulten, C. Chipot, and E. Tajkhorshid, J. Chem. Phys. **153**, 044130 (2020).
[37]C. H. Bennett, J. Comput. Phys. **22**, 245 (1976).
[38]M. Tachiya, J. Phys. Chem. **93**, 7050 (1989).
[39]D. V. Matyushov, J. Phys.: Condens. Matter **27**, 473001 (2015).
[40]D. V. Matyushov and G. A. Voth, J. Chem. Phys. **113**, 5413 (2000).
[41]D. R. Martin and D. V. Matyushov, Sci. Rep. **7**, 5495 (2017).
[42]D. V. Matyushov, Phys. Chem. Chem. Phys. **25**, 7589 (2023).
[43]B. Zhang, W. Song, J. Brown, R. Nemanich, and S. Lindsay, J. Am. Chem. Soc. **142**, 6432 (2020).
[44]D. N. LeBard and D. V. Matyushov, Phys. Rev. E **78**, 061901 (2008).

[45]R. A. Marcus and N. Sutin, Biochim. Biophys. Acta, Rev. Bioenerg. **811**, 265 (1985).
[46]W. Doster, Biochim. Biophys. Acta, Proteins Proteomics **1804**, 3 (2010).
[47]S. Khodadadi, S. Pawlus, J. H. Roh, V. Garcia Sakai, E. Mamontov, and A. P. Sokolov, J. Chem. Phys. **128**, 195106 (2008).
[48]W. Doster, S. Busch, A. M. Gaspar, M.-S. Appavou, J. Wuttke, and H. Scheer, Phys. Rev. Lett. **104**, 098101 (2010).
[49]S. Magazù, F. Migliardo, and A. Benedetto, J. Phys. Chem. B **115**, 7736 (2011).
[50]S. Seyedi and D. V. Matyushov, Soft Matter **13**, 8188 (2017).
[51]D. V. Matyushov and R. Richert, J. Chem. Phys. **144**, 041102 (2016).
[52]G. van der Zwan and J. T. Hynes, J. Phys. Chem. **89**, 4181 (1985).
[53]T. Samanta and D. V. Matyushov, Phys. Rev. Res. **3**, 023025 (2021).
[54]L. Onsager, J. Am. Chem. Soc. **58**, 1486 (1936).
[55]A. Rahman, Phys. Rev. **136**, A405 (1964).
[56]K. Binder and D. W. Heermann, *Monte Carlo Simulation in Statistical Mechanics* (Springer-Verlag, Berlin, 1992).
[57]S. Krishnan, A. Aksimentiev, S. Lindsay, and D. Matyushov, ACS Phys. Chem. Au **3**, 444 (2023).
[58]M. Drobizhev, N. S. Makarov, S. E. Tillo, T. E. Hughes, and A. Rebane, J. Phys. Chem. B **116**, 1736 (2012).
[59]M. Bixon and J. Jortner, Adv. Chem. Phys. **106**, 35 (1999).
[60]M. Drobizhev, P. R. Callis, R. Nifosì, G. Wicks, C. Stoltzfus, L. Barnett, T. E. Hughes, P. Sullivan, and A. Rebane, Sci. Rep. **5**, 13223 (2015).
[61]H. E. Stanley, *Introduction to Phase Transitions and Critical Phenomena* (Oxford University Press, New York, 1987).
[62]A. D. Friesen and D. V. Matyushov, J. Chem. Phys. **135**, 104501 (2011).
[63]Y. R. Shen and V. Ostroverkhov, Chem. Rev. **106**, 1140 (2006).
[64]J.-J. Velasco-Velez, C. H. Wu, T. A. Pascal, L. F. Wan, J. Guo, D. Prendergast, and M. Salmeron, Science **346**, 831 (2014).
[65]C.-Y. Li, J.-B. Le, Y.-H. Wang, S. Chen, Z.-L. Yang, J.-F. Li, J. Cheng, and Z.-Q. Tian, Nat. Mater. **18**, 697 (2019).
[66]G. Gonella, E. H. G. Backus, Y. Nagata, D. J. Bonthuis, P. Loche, A. Schlaich, R. R. Netz, A. Kühnle, I. T. McCrum, M. T. M. Koper, M. Wolf, B. Winter, G. Meijer, R. K. Campen, and M. Bonn, Nat. Rev. Chem. **5**, 466 (2021).
[67]D. R. Martin and D. V. Matyushov, J. Phys. Chem. Lett. **6**, 407 (2015).
[68]R. G. Palmer, Adv. Phys. **31**, 669 (1982).
[69]H. Reiss, *Methods of Thermodynamics* (Dover Publications, Inc., Mineola, 1996).