De novo determination of near-surface electrostatic potentials by NMR

Binhan Yu[a], Channing C. Pletka[a], B. Montgomery Pettitt[a], and Junji Iwahara[a,1]

[a]Department of Biochemistry and Molecular Biology, Sealy Center for Structural Biology and Molecular Biophysics, University of Texas Medical Branch, Galveston, TX 77555

Edited by Lewis E. Kay, University of Toronto, Toronto, Ontario, Canada, and approved May 2, 2021 (received for review March 8, 2021)

Electrostatic potentials computed from three-dimensional structures of biomolecules by solving the Poisson–Boltzmann equation are widely used in molecular biophysics, structural biology, and medicinal chemistry. Despite the approximate nature of the Poisson–Boltzmann theory, validation of the computed electrostatic potentials around biological macromolecules is rare and methodologically limited. Here, we present a unique and powerful NMR method that allows for straightforward and extensive comparison with electrostatic models for biomolecules and their complexes. This method utilizes paramagnetic relaxation enhancement arising from analogous cationic and anionic cosolutes whose spatial distributions around biological macromolecules reflect electrostatic potentials. We demonstrate that this NMR method enables de novo determination of near-surface electrostatic potentials for individual protein residues without using any structural information. We applied the method to ubiquitin and the Antp homeodomain–DNA complex. The experimental data agreed well with predictions from the Poisson–Boltzmann theory. Thus, our experimental results clearly support the validity of the theory for these systems. However, our experimental study also illuminates certain weaknesses of the Poisson–Boltzmann theory. For example, we found that the theory predicts stronger dependence of near-surface electrostatic potentials on ionic strength than observed in the experiments. Our data also suggest that conformational flexibility or structural uncertainties may cause large errors in theoretical predictions of electrostatic potentials, particularly for highly charged systems. This NMR-based method permits extensive assessment of near-surface electrostatic potentials for various regions around biological macromolecules and thereby may facilitate improvement of the computational approaches for electrostatic potentials.

electrostatics | nuclear magnetic resonance | proteins | ions | DNA

Due to the fundamental importance of electrostatic interactions in chemistry and biology, electrostatic potentials are invaluable information for the understanding of molecular recognition, enzymatic catalysis, and other functions of proteins and nucleic acids (1–4). Quantification of electrostatics is also important for successful protein engineering (5) and structure-based drug design (6). Computational approaches based on the Poisson–Boltzmann theory are commonly used to calculate electrostatic potentials from three-dimensional (3D) molecular structures (1, 7). Owing to available software such as Adaptive Poisson-Boltzmann Solver (APBS) (8, 9) and DelPhi (10, 11), computation of the electrostatic potentials around biomolecules has gained widespread popularity in the fields of molecular biophysics, structural biology, and medicinal chemistry.

However, the computed electrostatic potentials may not necessarily be accurate even if the 3D structures are precisely and accurately determined. Importantly, the Poisson–Boltzmann theory is approximate with known limitations. The electrostatic models based on this theory are valid under assumptions, which simplify the calculations (12). The lack of consideration of correlations between ions can diminish accuracy in calculations of electrostatic potentials for systems at high ionic strength (13). Due to the assumption of a dielectric continuum, the electrostatic potentials predicted with the Poisson–Boltzmann theory may be inaccurate for zones near the first hydration layer. Electrostatic potentials predicted for regions near highly charged molecular surfaces may also be inaccurate due to the assumption of linear dielectric response. Nonetheless, the Poisson–Boltzmann theory can accurately predict electrostatic interactions at longer range (7). The extent of validity for such electrostatic potentials near molecular surfaces remains to be addressed more rigorously through experiments.

Despite the need, experimental validation of computed electrostatic potentials is rather rare and methodologically limited for biological macromolecules. The validity of electrostatic models has been examined using pK_a data on titratable side-chain moieties (14–16), redox potentials of redox-active groups (17, 18), and electron–electron double resonance (19). Among them, pK_a data have been most commonly used for the validation, but even fundamentally incorrect electrostatic models can reproduce pK_a data (20). Electrostatic fields can be experimentally determined by vibrational spectroscopy, for example, for nitrile groups that are conjugated to cysteine thiol moieties of proteins (21, 22). However, the approaches utilizing vibrational spectroscopy or electron–electron double resonance provide only limited information about the extrinsically introduced probes, which may perturb native systems.

In this paper, we present a unique and powerful method for de novo determination of near-surface electrostatic potentials for

Significance

Electrostatic potentials are important information for our understanding of biomolecular functions as well as for protein engineering and drug design. Typically, electrostatic potentials around biomolecules are computed from three-dimensional structures. However, the theory behind this computation is approximate with some limitations. Validation of the electrostatic theory through experiments is therefore essential yet remains inadequate. This paper presents a unique spectroscopic method to experimentally determine electrostatic potentials near the molecular surfaces without using any structural information. This method allows for examination of theoretical electrostatic potentials for many different regions of biomolecules. Structure-independent determination of electrostatic potentials for various biomolecular systems may lead to improvement of theoretical models, which would ultimately have broad impacts on basic research and pharmaceutical development.

Author contributions: J.I. designed research; B.Y. and C.C.P. performed research; B.Y. and J.I. analyzed data; B.M.P. provided guidance on electrostatic theory; and B.Y. and J.I. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

See online for related content such as Commentaries.

[1]To whom correspondence may be addressed. Email: j.iwahara@utmb.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2104020118/-/DCSupplemental.

Published June 14, 2021.

Downloaded by guest on June 15, 2021

many protein residues, regardless of their side-chain types, and without using any chemical modifications. In this method, data of NMR paramagnetic relaxation enhancement (PRE) arising from analogous charged paramagnetic cosolutes are analyzed for ^{1}H nuclear magnetizations of proteins (Fig. 1). The PRE data reflect the electrostatic biases in spatial distributions of charged paramagnetic cosolutes and permit the determination of near-surface electrostatic potentials around proteins without using any structural information. The de novo determination of near-surface electrostatic potentials can greatly facilitate the examination of theoretical models for electrostatics of biological macromolecules.

Results

De Novo Determination of Near-Surface Electrostatic Potentials. To determine the near-surface electrostatic potentials, we used two derivatives of 2,2,5,5-tetramethylpyrrolidine-N-oxyl nitroxide (PROXYL): namely, amino-methyl-PROXYL and carboxy-PROXYL (Fig. 1). These paramagnetic compounds are analogous but differ in charge at neutral pH: amino-methyl-PROXYL is cationic ($+1e$) whereas carboxy-PROXYL is anionic ($-1e$). For PRE arising from paramagnetic cosolute molecules, the PRE rate for a transverse ^{1}H magnetization (Γ_2) is given as follows (23):

$$\Gamma_2 = 4\pi\xi c_P \tau_c \int_0^\infty r^{-4} \exp\left[-\frac{U(\boldsymbol{r})}{k_B T}\right] dr, \qquad [1]$$

where c_P is the concentration of the paramagnetic cosolute; τ_c is the correlation time for the dipole–dipole interaction between the macromolecular ^{1}H nucleus and the unpaired electron of the cosolute; r is the distance between the unpaired electron and the ^{1}H nucleus; $U(\boldsymbol{r})$ is the potential of mean force for the cosolute as a function of the position vector $\boldsymbol{r}$ for the unpaired electron

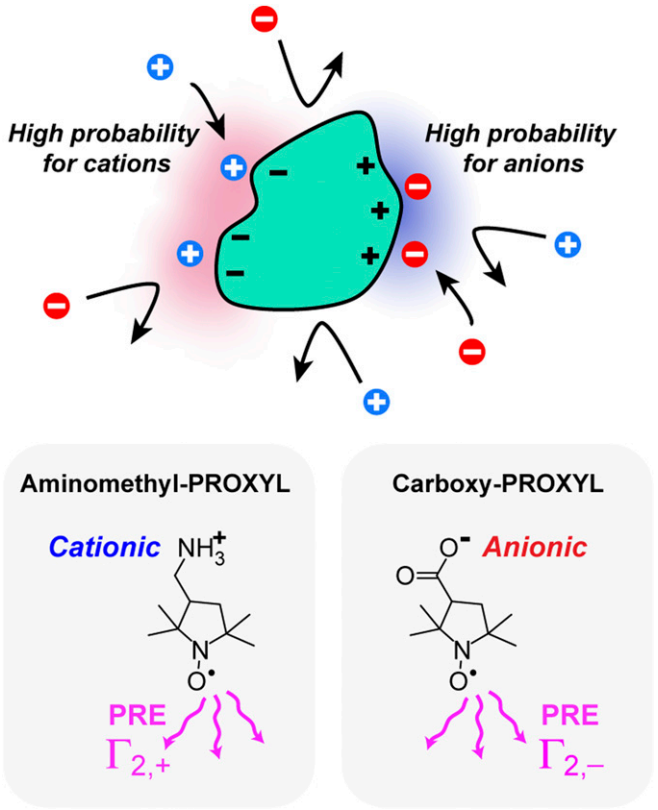

Fig. 1. NMR PRE arising from cationic amino-methyl-PROXYL or anionic carboxy-PROXYL reflects their spatial distribution bias due to near-surface electrostatic potentials around a biological macromolecule.

with respect to the ^{1}H nucleus; k_B is the Boltzmann constant; T is temperature; and ξ is a parameter reflecting other spin properties such as the gyromagnetic ratios. More details about Eq. **1** and the relevant parameters are provided in *SI Appendix*. The integral in Eq. **1** is conveniently discretized on a 3D lattice grid surrounding the macromolecule:

$$\Gamma_2 = \xi c_P \tau_c \sum_i v r_i^{-6} \exp\left[-\frac{U_i}{k_B T}\right], \qquad [2]$$

where i is the index of a grid point, and v is the volume of the voxel each grid point represents. Eq. **2** is similar to the empirical equation used for the so-called Otting–LeMaster approach (24–26) but differs in that each r_i^{-6} term involves a Boltzmann factor representing the bias in the spatial distribution of the cosolute molecule due to the effective potential U_i. When comparing PREs arising from the cationic and anionic PROXYL derivatives at the same concentration, the ratio of Γ_2 rates will be as follows:

$$\Gamma_{2,+} \Big/ \Gamma_{2,-} = \sum_i r_i^{-6} \exp\left[-\frac{U_{+,i}}{k_B T}\right] \Big/ \sum_i r_i^{-6} \exp\left[-\frac{U_{-,i}}{k_B T}\right], \qquad [3]$$

where the annotations "+" and "−" are for the cationic and anionic PROXYL derivatives, respectively. The cancellation of τ_c in Eq. **3** requires identical diffusional properties for the cationic and anionic PROXYL derivatives, which we confirmed through NMR-based diffusion measurements (*SI Appendix*, Fig. S1). Due to the strong distance dependence imposed by r_i^{-6}, each Γ_2 rate is dominated by terms for a near-surface zone close to the observed ^{1}H nucleus. We referred to it as the effective near-surface (ENS) zone. As shown in Fig. 2, the ENS zone for each observed protein ^{1}H nucleus is relatively narrow in the total space surrounding a macromolecule.

Here, for the ENS zone of each observed ^{1}H atom, we define a potential as follows:

$$\phi_{ENS} = -\frac{k_B T}{2e} \ln\left(\Gamma_{2,+} / \Gamma_{2,-}\right), \qquad [4]$$

in which e is the elementary charge. The physical meaning of the ϕ_{ENS} potential becomes apparent by considering a hypothetical case in which the potentials U_+ and U_- are uniform within the ENS zone. Under this condition, Eq. **3** is reduced to the following:

$$\Gamma_{2,+} \Big/ \Gamma_{2,-} = \exp\left[-\frac{\Delta U_{ENS}}{k_B T}\right], \qquad [5]$$

where $\Delta U_{ENS} = U_+ - U_-$. Tetramethyl nitroxide compounds such as PROXYL or TEMPOL are known to preferentially interact with hydrophobic patches on protein surface (23, 26, 27). Because the hydrophobic moiety of PROXYL is identical for the two PROXYL derivatives, it is likely that the hydrophobic contribution in each U term is approximately cancelled in ΔU_{ENS}. Considering that the major difference between the cationic and anionic PROXYL derivatives is the charge ($+1e$ versus $-1e$), it is reasonable to assume that the difference ΔU_{ENS} is dominated by the electrostatic interaction and given by the following:

$$\Delta U_{ENS} = 2e\phi, \qquad [6]$$

where ϕ is the electrostatic potential within the ENS zone. From Eqs. **4**–**6**, it is obvious that $\phi_{ENS} = \phi$ in this hypothetical case with a uniform electrostatic potential within the ENS zone. In general cases with nonuniform electrostatic potentials, ϕ_{ENS} represents an effective electrostatic potential for the ENS zone. As demonstrated in the next subsection, ϕ_{ENS} is close to the average of the electrostatic

Downloaded by guest on June 15, 2021

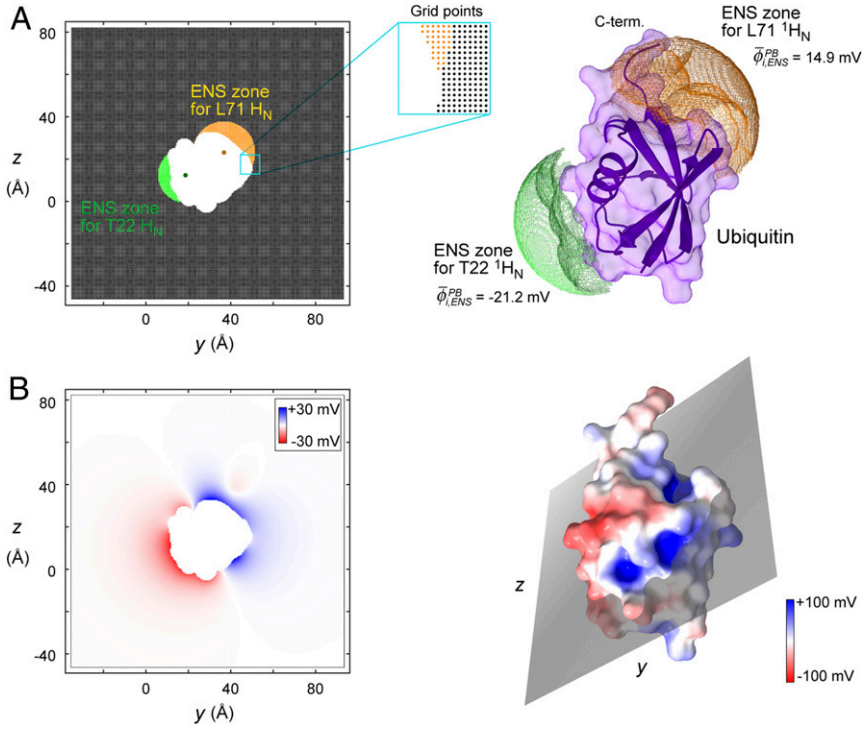

Fig. 2. Electrostatic potentials and ENS zones for PROXYL-induced PRE for the T22 and L71 ^1H_N nuclei of ubiquitin (PDB: 1UBQ). (*A*) The ENS zones for the ^1H_N nuclei of T22 and L71. (*Left*) Two-dimensional (2D) plane containing the H_N atoms of T22 and L71 (shown in dark green and dark orange, respectively). Possible positions of the paramagnetic center of PROXYL derivatives around a ubiquitin molecule are shown in lattice representations (with a spacing of 0.5 Å). The grid points within a zone making 68% contribution to $\sum r_i^{-6}$, where r_i is the distance between the ^1H nucleus and a grid point i in the exterior space, are colored in green (T22) or orange (L71). These zones are also shown in the 3D structure. (*B*) Exterior electrostatic potentials at the grid points on the 2D plane shown in *A*. Electrostatic potentials at 30 mM ionic strength were computed using the APBS software. The location of the 2D plane is also indicated by a smaller rectangle in a 3D surface electrostatic potential map. The average values of electrostatic potentials at grid points within the ENS zones ($\overline{\phi}^{PB}_{i,ENS}$) were −21.2 mV for T22 ^1H_N and 14.9 mV for L71 ^1H_N. The corresponding ϕ^{PB}_{ENS} potentials calculated using all grid points (both inside and outside the ENS zones) along with Eqs. **3** and **4** were −20.1 mV for T22 ^1H_N and 13.5 mV for L71 ^1H_N. The threshold (68%) for the ENS zones was chosen as a value at which $\overline{\phi}^{PB}_{i,ENS}$ becomes close to ϕ^{PB}_{ENS} for the majority of ^1H_N nuclei (see *Results*). Although the physical meaning of $\overline{\phi}^{PB}_{i,ENS}$ is clearer than that of ϕ^{PB}_{ENS}, we use ϕ^{PB}_{ENS} potentials to compare with experimental data because ϕ^{PB}_{ENS} does not depend on any arbitrary threshold.

potential within the ENS zone. Importantly, Eq. **4** enables determination of the ENS electrostatic potentials ϕ_{ENS} from $\Gamma_{2,+}$ and $\Gamma_{2,-}$ data without using any structural information. Experimental ϕ_{ENS} data allow us to examine theoretical electrostatic models through direct comparison with theoretical predictions using Eqs. **3** and **4**.

Comparison with the Poisson–Boltzmann Equation-Based Electrostatic Potentials. We applied this method to ubiquitin, a 76-residue protein containing 11 positively charged and 11 negatively charged side chains. We measured PRE $\Gamma_{2,+}$ and $\Gamma_{2,-}$ rates at 25 °C for backbone ^1H_N nuclei of ^{15}N ubiquitin in solutions containing 10 mM amino-methyl- or carboxy-PROXYL at pH 7.5 and the

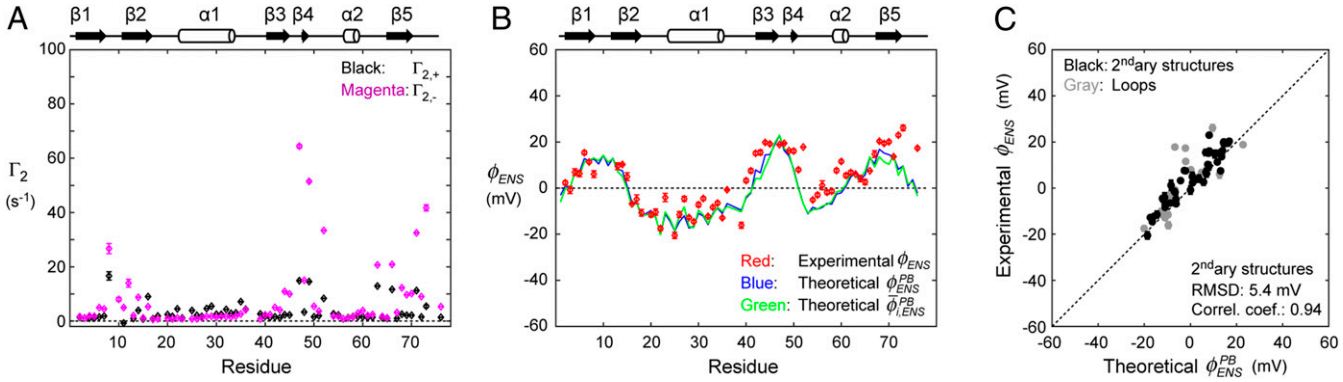

Fig. 3. ENS electrostatic potentials ϕ_{ENS} determined from NMR PRE data for backbone H_N atoms of ubiquitin at pH 7.5 and the ionic strength of 30 mM. (*A*) ^1H PRE rates Γ_2 measured with the cationic and anionic PROXYL derivatives (10 mM). (*B*) Experimental ENS electrostatic potentials ϕ_{ENS} (red) determined from the PRE data for individual residues. Theoretical ENS electrostatic potentials ϕ^{PB}_{ENS} (blue) and $\overline{\phi}^{PB}_{i,ENS}$ (green) predicted from the 1.8-Å resolution crystal structure (PDB: 1UBQ) are also plotted. (*C*) Correlation plots for the experimental ϕ_{ENS} data and the theoretical ϕ^{PB}_{ENS} data. Data points for the secondary structure regions are shown in black and those for loop regions are shown in gray. The corresponding data for ubiquitin at the ionic strength of 130 mM are shown in *SI Appendix*.

Downloaded by guest on June 15, 2021

ionic strength of 30 mM (Fig. 3*A* and *SI Appendix*, Table S1). We identified 63 ^1H_N nuclei, which exhibited statistically significant PRE rates and $\Gamma_{2,+}/\Gamma_{2,-}$ ratios, using the criteria described in the *Materials and Methods* section. For each of these ^1H_N nuclei, we determined the ENS electrostatic potential ϕ_{ENS} using Eq. **4** and the measured $\Gamma_{2,+}$ and $\Gamma_{2,-}$ rates. Fig. 3*B* shows the ENS electrostatic potentials ϕ_{ENS} (red) determined from $\Gamma_{2,+}/\Gamma_{2,-}$ data for individual residues of ubiquitin at the ionic strength of 30 mM. The range of the ϕ_{ENS} potentials determined for ubiquitin was from −21 to +26 mV.

Using the APBS software and a 1.8-Å resolution crystal structure of ubiquitin (Protein Data Bank [PDB]: 1UBQ), we also performed the Poisson–Boltzmann theory-based calculations of electrostatic potentials. From the electrostatic potentials ϕ_i computed at all grid points in the exterior space, we predicted the ϕ_{ENS} potential (ϕ^{PB}_{ENS}) for each ^1H_N nucleus using Eqs. **3** and **4** along with $U_{+,i} = e\phi_i$ and $U_{-,i} = -e\phi_i$. The accessibility of each voxel was assessed using van der Waals radii of macromolecular atoms and an empirically determined radius of 3.5 Å for the paramagnetic center of the PROXYL derivatives (*SI Appendix*, Fig. S3). We also calculated the average ($\overline{\phi}^{PB}_{i,ENS}$) of the electrostatic potentials at grid points only within a proximal zone making a 68% contribution to $\sum r^{-6}$ for each ^1H_N nucleus. At this threshold (68%), the average potentials $\overline{\phi}^{PB}_{i,ENS}$ within these zones were in excellent agreement with the ϕ^{PB}_{ENS} potentials calculated for all grid points (Fig. 3*B*). The experimental ϕ_{ENS} data agreed remarkably well with the theoretical data, particularly for secondary structure regions. A correlation plot comparing the experimental ϕ_{ENS} and theoretical ϕ^{PB}_{ENS} potentials is shown in Fig. 3*C*. The rmsd between them was 5.4 mV, and the correlation coefficient was 0.94 for secondary structure regions of ubiquitin at the ionic strength of 30 mM. Experimental ϕ_{ENS} data were also consistent with $\overline{\phi}^{PB}_{i,ENS}$, supporting the notion that the ϕ_{ENS} potential represents an effective electrostatic potential in the ENS zone. These results clearly indicate that the Poisson–Boltzmann theory predicts the near-surface electrostatic potentials for this system well and that our current experimental approach allows for straightforward comparison with electrostatic models.

Impact of Conformational Variations. The majority of large differences between the experimental and theoretical ϕ_{ENS} data for ubiquitin were found in loop regions. To examine the extent to which conformational variations affect the prediction of near-surface potentials, we conducted the same analysis of the ensemble of 10 NMR structures determined by Bax and coworkers (PDB: 1D3Z) (28). The ϕ^{PB}_{ENS} potentials calculated for the NMR structures were largely consistent with those calculated for the crystal structure, and larger variations were observed for loop regions (*SI Appendix*, Fig. S4*A*). For further examination of the impact of conformational variations, we also generated 100 structures with varied conformations. The backbone atoms of the original crystal structure were kept fixed while the side-chain conformations were varied through high temperature dynamics under the influence of the conformational database potentials (29) using the Xplor-NIH software (30). The structural variations caused changes in ϕ^{PB}_{ENS} potentials (*SI Appendix*, Fig. S4*B*). For the secondary structures, the mean of SDs of ϕ^{PB}_{ENS} potentials for the 100 structures were 4.1 mV, which is comparable to the rmsd (5.4 mV) between the ϕ^{PB}_{ENS} data for 1UBQ and the experimental ϕ_{ENS} data. The SDs of ϕ^{PB}_{ENS} due to the structural variations were larger for loop regions, which also exhibited the large differences between the experimental and theoretical ϕ_{ENS} data. These results collectively suggest that the larger discrepancy for the residues in the loop regions of ubiquitin is mainly attributed to the conformational flexibility and/or structural uncertainties, which represents an inherent limitation of structure-based electrostatic potential predictions.

Examination at a Higher Ionic Strength. Typical Poisson–Boltzmann and Debye–Hückel models do not take ionic correlation effects into consideration and are known to overestimate the dependence of the activity coefficients on the ionic strength (I) when $I > \sim$100 mM (31). To examine the impacts of an increase in I on the experimental and theoretical near-surface electrostatic potential data, we conducted the same experiments for ubiquitin at I = 130 mM. The PRE $\Gamma_{2,+}$ and $\Gamma_{2,-}$ rates at 25 °C are shown in *SI Appendix*, Fig. S5*A* and Table S2. For many of the residues, the magnitudes of PRE were smaller at I = 130 mM, implying weaker electrostatic interactions between the protein and the charged cosolutes at the higher ionic strength. Due to the decreased magnitude of PRE and the lower sensitivity of the cryogenic NMR probe at higher ionic strength, the number of ^1H_N nuclei that met the criteria for ϕ_{ENS} determination was smaller at I = 130 mM (50 ^1H_N nuclei) than at I = 30 mM (63 ^1H_N nuclei). Fig. 4*A* compares experimental ϕ_{ENS} data for the secondary structure regions of ubiquitin at I = 30 and 130 mM. There was a clear trend that the magnitudes of ϕ_{ENS} potentials at I = 130 mM were smaller than those at I = 30 mM. The linear regression for the correlation plot gave a slope of 0.74 ± 0.07 (with the uncertainty estimated at a confidence interval [CI] of 95%). Fig. 4*B* shows the corresponding ϕ^{PB}_{ENS} data predicted by the Poisson–Boltzmann theory. The slope for the linear regression for the correlation between ϕ^{PB}_{ENS} data at I = 30 and 130 mM was 0.64 ± 0.01, showing that the Poisson–Boltzmann theory predicts a stronger dependence on ionic strength. Nonetheless, the agreement between the experimental ϕ_{ENS} and theoretical ϕ^{PB}_{ENS} data at I = 130 mM (with an rmsd of 5.1 mV and a correlation coefficient of 0.89 for the secondary structure regions; reference *SI Appendix*, Fig. S5 *B* and *C*) was as good as at I = 30 mM. Thus, the Poisson–Boltzmann theory predicts the ENS electrostatic potentials at the ionic strengths of 30 and 130 mM almost equally well, though it appears to predict a slightly stronger dependence on I in the current range.

Examination for a Highly Charged System: The Antp Homeodomain–DNA Complex. As mentioned in the introduction, it has been suggested that the Poisson–Boltzmann theory may inaccurately predict electrostatic potentials for highly charged systems (12). Previous studies showed that the Poisson–Boltzmann equation-based approaches can accurately predict the total number of ions accumulated around highly charged proteins (32) and nucleic acids (33–35).

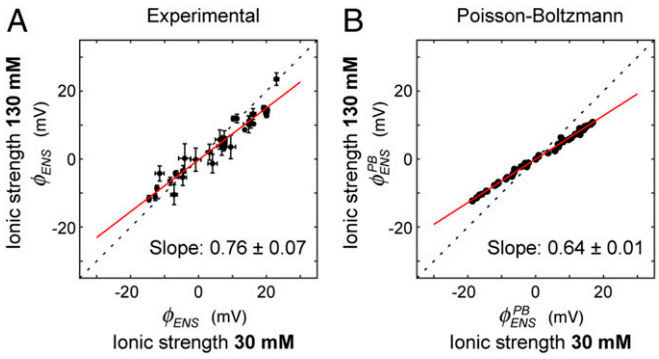

Fig. 4. Comparison of ENS electrostatic potentials for backbone H_N atoms in the secondary structures of ubiquitin at ion strengths of 30 and 130 mM. (*A*) Comparison of experimental ϕ_{ENS} data. (*B*) Comparison of computational ϕ^{PB}_{ENS} data predicted by the Poisson–Boltzmann theory. The red lines represent linear regression. The uncertainties in the slopes were estimated at a CI of 95%. Diagonals are indicated in broken lines as a guide to the eye.

Downloaded by guest on June 15, 2021

However, the total number of accumulated ions is more relevant to long-range electrostatic potentials. The validity of the Poisson–Boltzmann theory for near-surface zones remains to be examined for highly charged systems. For such examination, we applied our method to the complex of the Antp homeodomain and a 15-base pair (bp) DNA duplex. The total charge of this protein–DNA complex is −16*e*. Using the complex of ^{15}N-labeled protein and unlabeled DNA, we measured the ^{1}H PRE $\Gamma_{2,+}$ rates for amino-methyl-PROXYL and $\Gamma_{2,-}$ rates for carboxy-PROXYL (Fig. 5*A*; see also *SI Appendix*, Table S3) and determined the ϕ_{ENS} potentials for $^{1}H_N$ atoms in the complex at 25 °C (Fig. 5*B*). The determined ϕ_{ENS} potentials ranged from −54 to −8 mV. Unlike ubiquitin's case, backbone $^{1}H_N$ nuclei of the Antp homeodomain–DNA complex exhibited only negative ϕ_{ENS} potentials, reflecting the strong influence from the large negative charge of DNA.

We also computed ϕ_{ENS}^{PB} and $\overline{\phi}_{i,ENS}^{PB}$ using the electrostatic potentials calculated with the Poisson–Boltzmann equation for a 2.4-Å resolution crystal structure (PDB: 9ANT). Fig. 5*B* shows the experimental ϕ_{ENS} data (red) and the theoretical ϕ_{ENS}^{PB} (blue) and $\overline{\phi}_{i,ENS}^{PB}$ (green) data for the Antp homeodomain–DNA complex. For loop regions, differences between experimental ϕ_{ENS} and computational ϕ_{ENS}^{PB} data for this complex were remarkably larger than those for ubiquitin. Conformational flexibility or structural uncertainties may influence ϕ_{ENS}^{PB} predictions more strongly for highly charged systems. This trend may arise partly from the significant packing and orientational correlation of PROXYL-based ions, which are not well represented in Poisson–Boltzmann theory. Nonetheless, for the secondary structure regions of the Antp homeodomain bound to DNA, the rmsd between experimental and theoretical ϕ_{ENS} data were 5.0 mV, which is even lower than the corresponding rmsd for ubiquitin at 30 mM, and the correlation coefficient was 0.84. Thus, the Poisson–Boltzmann theory appears to produce reasonable predictions, at least for the structurally well-defined regions of this highly charged system.

Discussion

Our NMR-based method has several advantages over previous experimental approaches for validation of theoretical models of protein electrostatics. The most remarkable advantage is that the NMR method can directly provide the near-surface electrostatic potentials for many protein residues simultaneously regardless of residue types. Previous experimental approaches provide only limited information on particular types of groups. Although validations based on pK_a data have been common (14–16), the pK_a-based method does not provide near-surface electrostatic potentials. Experimental pK_a determination requires fitting to experimental data at various pH values and is impacted by secondary effects (e.g., conformational changes and deprotonation/protonation of nearby residues due to the variation in pH) (36). In our paramagnetic NMR method, experiments are performed at a single pH and do not require any fitting procedures. Although electron paramagnetic resonance or vibrational spectroscopic methods are able to provide electrostatic potentials or fields (19, 21, 22), information from each experiment with these methods is limited to a single extrinsic probe that is conjugated to a protein and may perturb its native molecular properties. Our paramagnetic NMR-based method provides electrostatic potentials for many residues in native molecules and does not require any protein modification. Although we measured the ϕ_{ENS} potentials only for $^{1}H_N$ nuclei of ^{15}N-labeled proteins in our current study, the same method can in principle be applied to ^{1}H nuclei of CH/CH$_2$/CH$_3$ groups in ^{13}C/^{15}N-labeled proteins as well. This method is relatively easy in sample preparation, experimental procedures, and data analysis but allows for de novo determination of near-surface electrostatic potentials for a large number of protein residues without using any structural information.

In conclusion, we have developed a unique and powerful experimental method that permits straightforward examination of theoretical models for biomolecular electrostatics. Experimental ϕ_{ENS} data can serve for benchmarks for testing electrostatic models. Comparison of the paramagnetic NMR-based data and theoretical data are straightforward and greatly facilitate validation of theoretical models. Our current work illuminates the effectiveness and weakness of the Poisson–Boltzmann theory. Further experiments with the de novo method can produce large data sets of near-surface electrostatic potentials for various regions of biomolecules under various conditions. Although some discrepancy may arise from structural uncertainties, comparison of the experimental and computational electrostatic potentials will promote the improvement of theoretical models on protein electrostatics. Such improvement of electrostatic models would have broad impacts on molecular biophysics, structural biology, protein engineering, and drug development.

Materials and Methods

Proteins, DNA, and Other Materials. ^{15}N-labeled ubiquitin was purchased from Sigma-Isotec. ^{15}N-labeled Antp homeodomain bearing the C39S mutation was expressed in *Escherichia coli* and purified as described by Zandarashvili et al. (37) The unlabeled 15-bp DNA duplex with the sequence of CTCTAATGGCTTTCT (bold, the Antp recognition site) was prepared as described (37). The paramagnetic compounds, amino-methyl-PROXYL (Chemical Abstracts Service [CAS] number: 54606-49-4) and carboxy-PROXYL (CAS

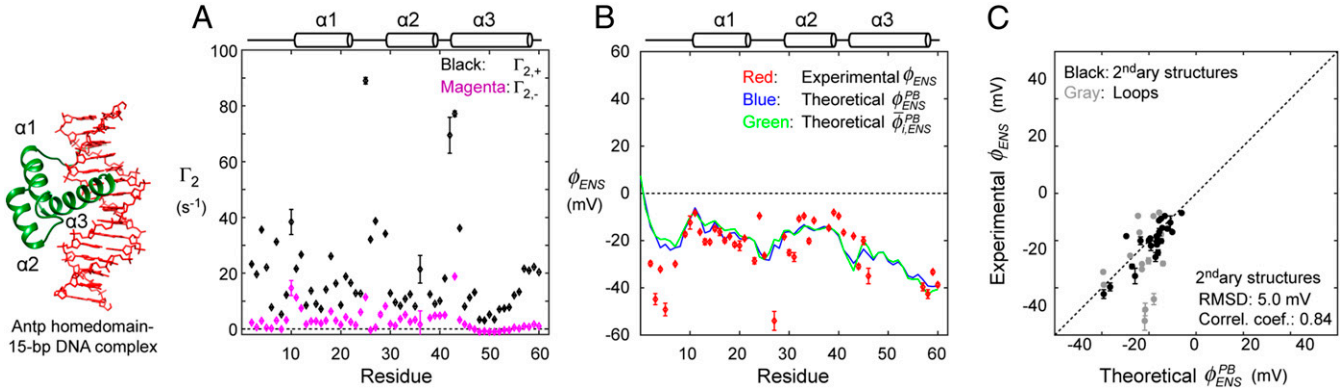

Fig. 5. Near-surface electrostatic potentials determined from ^{1}H PRE rates Γ_2 for protein backbone H_N atoms of the Antp homeodomain–DNA complex at pH 7.5 and the ionic strength of 30 mM. The data in panels *A*, *B*, and *C* are presented in the same manner as shown in Fig. 3. For straightforward comparison of the two systems, the axis scale is set identical for this figure and Fig. 3. Note that for the entire system, the near-surface electrostatic potentials are negative due to a strongly negatively charged DNA duplex.

Downloaded by guest on June 15, 2021

number: 2154-68-9), were purchased from Sigma-Aldrich. Other reagents were also purchased from Sigma-Aldrich.

Quantifications of PROXL Derivatives. For the de novo determination of near-surface electrostatic potentials, it is crucial to precisely measure the concentrations of amino-methyl-PROXYL and carboxy-PROXYL in the PRE experiments. Since pure amino-methyl-PROXYL is a highly viscous gel-like substance, it is impractical to quantify this compound by weighing. To resolve this problem, an NMR-based approach was used. In this approach, the PROXYL nitroxide (paramagnetic) sampled from a stock solution was completely reduced to hydroxylamine (diamagnetic) by adding an excess amount of ascorbic acid (38) and then a one-dimensional ^{1}H NMR spectrum was recorded in the presence of 10 mM dimethyl sulfoxide (DMSO) using a single scan (without any dummy scan) to excite ^{1}H nuclei at the Boltzmann equilibrium. The integrals of methyl ^{1}H NMR signals from reduced PROXYL and those from 10 mM DMSO as the internal standard were used to determine the concentration of the PROXYL derivative (an example is shown in *SI Appendix*, Fig. S2). Based on the concentrations of the PROXYL derivatives measured through this approach, the stock solutions of 20 mM carboxy- or amino-methyl-PROXYL in the NMR buffer were prepared.

Samples for NMR Experiments. NMR experiments were performed for samples containing 0.4 mM ^{15}N-labeled proteins. The NMR samples of ^{15}N ubiquitin at the ionic strength of 30 mM were 520-μL solutions containing 0.4 mM protein, 20 mM acetate, 28 mM Tris, 10 mM amino-methyl- or carboxy-PROXYL, 10 mM DMSO, and 5% D_2O at pH 7.5. For the ubiquitin samples at the ionic strength of 130 mM, the buffer also contained 100 mM KCl. The NMR samples of the Antp homeodomain–DNA complex contained 0.4 mM ^{15}N Antp homeodomain, 0.51 mM 15-bp DNA, 20 mM acetate, 28 mM Tris, 10 mM amino-methyl- or carboxy-PROXYL, 10 mM DMSO, and 5% D_2O at pH 7.5. Solutions containing no paramagnetic substance were prepared through buffer equilibration using an Amicon Ultra-4 centrifugal filter with a molecular weight cutoff at 3 kDa (Millipore EMD) and then were concentrated to ∼0.8 mM. The diamagnetic control samples containing no paramagnetic substance were prepared by diluting the high-concentration solution with the same buffer containing no PROXYL derivatives. The paramagnetic samples were prepared by mixing the high-concentration solution with the stock solution of 20 mM carboxy- or amino-methyl-PROXYL in the same buffer to the final concentration of 10 mM at the final step. This procedure ensures the exact final concentration of amino-methyl- or carboxy-PROXYL. It should be noted that buffer equilibration using a buffer containing a charged PROXYL derivative would cause a significant deviation of its final concentration from the original concentration due to ion accumulation or exclusion by proteins and DNA (32, 39). Each NMR sample was 0.5 mL and was sealed in a 5-mm tube (Norell).

NMR Experiments. The PRE rates Γ_2 for ^{1}H transverse magnetizations were measured for protein backbone ^1H_N nuclei of ^{15}N-labeled ubiquitin and the complex of ^{15}N-labeled Antp homeodomain and unlabeled 15-bp DNA. The two time-point approach (40) with a 10-ms difference was used to measure PRE Γ_2 rates. Uncertainties in Γ_2 rates were estimated through the error propagation method, as previously described (40). For each molecular system, the PRE experiments were conducted with three samples: one without any paramagnetic cosolute (diamagnetic) and the others with 10 mM amino-methyl-PROXYL (cationic) or carboxy-PROXYL (anionic). All NMR experiments were performed at 25 °C using a Bruker Avance III spectrometer equipped with a QCI cryogenic probe operated at the ^{1}H frequency of 600 MHz. The spectra were processed with the NMRPipe software (41). Analyses of the NMR spectra and quantification of signal intensities for PRE measurements were carried out using the NMRFAM-SPARKY software (42). For ubiquitin, the Biological Magnetic Resonance Bank entry 17769 was used for resonance assignment information. For the Antp homeodomain–DNA complex, the resonances assigned by Nguyen et al. (43) was used. Resonance assignment for each system was confirmed using 3D ^{15}N-edited nuclear Overhauser effect spectroscopy heteronuclear single-quantum coherence (NOESY-HSQC) spectra.

Determination of Near-Surface Electrostatic Potentials. ENS electrostatic potentials ϕ_{ENS} for individual backbone ^1H_N nuclei were determined from $\Gamma_{2,+}$ and $\Gamma_{2,-}$ data using Eq. **4**. The uncertainties in ϕ_{ENS} were estimated through error propagation (44) using the following:

$$\sigma_\phi = \frac{k_B T}{2e}\sqrt{\left(\sigma_+/\Gamma_{2,+}\right)^2 + \left(\sigma_-/\Gamma_{2,-}\right)^2}, \quad [7]$$

where σ_+ and σ_- are uncertainties in $\Gamma_{2,+}$ and $\Gamma_{2,-}$, respectively. To use only statistically significant PRE rates and $\Gamma_{2,+}/\Gamma_{2,-}$ ratios, we chose ^{1}H nuclei, which satisfied the following three criteria: $\Gamma_{2,+} > 3\sigma_+$, $\Gamma_{2,-} > 3\sigma_-$, and $\Gamma_{2,+}/\Gamma_{2,-} > 3(\Gamma_{2,+}/\Gamma_{2,-})\sqrt{(\sigma_+/\Gamma_{2,+})^2 + (\sigma_-/\Gamma_{2,-})^2}$. The calculations were performed using the MATLAB software (MathWorks).

Poisson–Boltzmann Equation–Based Calculations of Electrostatic Potentials. Theoretical electrostatic potentials were calculated with the APBS software (8, 9) by numerically solving the nonlinear Poisson–Boltzmann equation. Full-atom models for ubiquitin and the Antp–DNA complex were constructed from the crystal structures (PDB: 1UBQ and 9ANT) using the Xplor-NIH software (30). In this step, hydrogen and other atoms absent in the crystal structures but present in the actual molecules were added to the structure models. The PDB2PQR tool (45) was used to assign charges and radii to individual atoms. The AMBER ff99 force-field parameter set and the protonation states at pH 7.5 predicted with the PROPKA approach (46) were used. The grid used for the APBS calculations was 128 × 128 × 128 Å^3 for ubiquitin. A larger space of 160 × 160 × 160 Å^3 was used for the Antp homeodomain–DNA complex. The grid spacing was 0.5 Å in each case. The monovalent ion concentration was set to 30 or 130 mM, depending on the actual experimental ionic strength used. The solvent van der Waals radius and the ionic radius were set to 1.4 and 2.0 Å, respectively, for calculations of the electrostatic potential at each grid point. The dielectric constant of the solvent was set to 78.54, and the dielectric constant for the macromolecular interior (ε_i) was set to 2, which has been used for both proteins and DNA (33, 47). Other values (ε_i = 4, 10, 20) were also tested, which impacted ϕ^{PB}_{ENS} with an rmsd ≤ 1.1 mV (*SI Appendix*, Fig. S6). For each grid point, the accessibility of the paramagnetic moiety (i.e., nitroxide) was determined using a radius of 3.5 Å, which empirically gave the best agreement between experimental and computational ϕ_{ENS} data on amino-methyl-PROXYL and carboxy-PROXYL (*SI Appendix*, Fig. S3). The ENS zone for each ^{1}H nucleus was defined as the grid points that are located within the distance range making a 68% contribution to $\sum r^{-6}$. The $\overline{\phi}^{PB}_{i,ENS}$ potential was calculated as the average of electrostatic potentials for the grid points within the ENS zone. The ENS electrostatic potentials (ϕ^{PB}_{ENS}) were computed with Eqs. **3** and **4** using all grid points, including those inside and outside the ENS zone, along with $U_{+,i} = e\phi_i$ and $U_{-,i} = -e\phi_i$ for individual grid points. When Eq. **3** was used for calculating ϕ^{PB}_{ENS}, the accessibility factor a_i (0 for inaccessible grids; 1 for accessible grids) was multiplied to each r_i^{-6} term. The calculations of $\overline{\phi}^{PB}_{i,ENS}$ and ϕ^{PB}_{ENS} from the APBS outputs were conducted using the MATLAB software. Structures in Fig. 2 were drawn using ChimeraX (48).

Data Availability. All study data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. This work was supported by Grants R35-GM130326 (to J.I.) from the NIH and CHE-1709310 (to B.M.P.) from the NSF. We thank Dr. Tianzhi Wang for maintenance of the NMR instruments.

1. B. Honig, A. Nicholls, Classical electrostatics in biology and chemistry. *Science* **268**, 1144–1149 (1995).
2. A. Warshel *et al*., Electrostatic basis for enzyme catalysis. *Chem. Rev.* **106**, 3210–3235 (2006).
3. B. Yu, B. M. Pettitt, J. Iwahara, Dynamics of ionic interactions at protein–nucleic acid interfaces. *Acc. Chem. Res.* **53**, 1802–1810 (2020).
4. H. X. Zhou, X. Pang, Electrostatic interactions in protein structure, folding, binding, and condensation. *Chem. Rev.* **118**, 1691–1741 (2018).
5. C. L. Vizcarra, S. L. Mayo, Electrostatics in computational protein design. *Curr. Opin. Chem. Biol.* **9**, 622–626 (2005).
6. A. C. Anderson, The process of structure-based drug design. *Chem. Biol.* **10**, 787–797 (2003).
7. F. Fogolari, A. Brigo, H. Molinari, The Poisson-Boltzmann equation for biomolecular electrostatics: A tool for structural biology. *J. Mol. Recognit.* **15**, 377–392 (2002).
8. N. A. Baker, D. Sept, S. Joseph, M. J. Holst, J. A. McCammon, Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10037–10041 (2001).
9. E. Jurrus *et al*., Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* **27**, 112–128 (2018).
10. M. K. Gilson, K. A. Sharp, B. H. Honig, Calculating the electrostatic potential of molecules in solution: Method and error assessment. *J. Comput. Chem.* **9**, 327–335 (1988).
11. L. Li *et al*., DelPhi: A comprehensive suite for DelPhi software and associated resources. *BMC Biophys.* **5**, 9 (2012).
12. P. Grochowski, J. Trylska, Continuum molecular electrostatics, salt effects, and counterion binding–A review of the Poisson-Boltzmann theory and its modifications. *Biopolymers* **89**, 93–113 (2008).

Downloaded by guest on June 15, 2021

13. R. R. Netz, H. Orland, Beyond Poisson-Boltzmann: Fluctuation effects and correlation functions. *Eur. Phys. J. E* **1**, 203–214 (2000).
14. J. E. Nielsen, M. R. Gunner, B. E. García-Moreno, The pK_a cooperative: A collaborative effort to advance structure-based calculations of pKa values and electrostatic effects in proteins. *Proteins* **79**, 3249–3259 (2011).
15. C. N. Schutz, A. Warshel, What are the dielectric "constants" of proteins and how to validate electrostatic models? *Proteins* **44**, 400–417 (2001).
16. A. Warshel, A. Dryga, Simulating electrostatic energies in proteins: Perspectives and some recent studies of pKas, redox, and other crucial functional properties. *Proteins* **79**, 3469–3484 (2011).
17. A. K. Churg, A. Warshel, Control of the redox potential of cytochrome c and microscopic dielectric effects in proteins. *Biochemistry* **25**, 1675–1681 (1986).
18. J. Lu, W. R. Kobertz, C. Deutsch, Mapping the electrostatic potential within the ribosomal exit tunnel. *J. Mol. Biol.* **371**, 1378–1391 (2007).
19. Y. K. Shin, W. L. Hubbell, Determination of electrostatic potentials at biological interfaces using electron-electron double resonance. *Biophys. J.* **61**, 1443–1453 (1992).
20. A. Warshel, A. Papazyan, Electrostatic effects in macromolecules: Fundamental concepts and practical modeling. *Curr. Opin. Struct. Biol.* **8**, 211–217 (1998).
21. A. Chattopadhyay, S. G. Boxer, Vibrational Stark effect spectroscopy. *J. Am. Chem. Soc.* **117**, 1449–1450 (1995).
22. A. T. Fafarman *et al*., Quantitative, directional measurement of electric field heterogeneity in the active site of ketosteroid isomerase. *Proc. Natl. Acad. Sci. U.S.A.* **109**, E299–E308 (2012).
23. Y. Okuno, A. Szabo, G. M. Clore, Quantitative interpretation of solvent paramagnetic relaxation for probing protein-cosolute interactions. *J. Am. Chem. Soc.* **142**, 8281–8290 (2020).
24. G. M. Clore, J. Iwahara, Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chem. Rev.* **109**, 4108–4139 (2009).
25. G. Hernández, C. L. Teng, R. G. Bryant, D. M. LeMaster, O_2 penetration and proton burial depth in proteins: Applicability to fold family recognition. *J. Am. Chem. Soc.* **124**, 4463–4472 (2002).
26. G. Pintacuda, G. Otting, Identification of protein surfaces by NMR measurements with a pramagnetic Gd(III) chelate. *J. Am. Chem. Soc.* **124**, 372–373 (2002).
27. A. M. Petros, L. Mueller, K. D. Kopple, NMR identification of protein surfaces using paramagnetic probes. *Biochemistry* **29**, 10041–10048 (1990).
28. G. Cornilescu, J. L. Marquardt, M. Ottiger, A. Bax, Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *J. Am. Chem. Soc.* **120**, 6836–6837 (1998).
29. G. M. Clore, J. Kuszewski, χ_1 rotamer populations and angles of mobile surface side chains are accurately predicted by a torsion angle database potential of mean force. *J. Am. Chem. Soc.* **124**, 2866–2867 (2002).
30. C. D. Schwieters, J. J. Kuszewski, G. M. Clore, Using Xplor-NIH for NMR molecular structure determination. *Prog. NMR Spect.* **48**, 47–62 (2006).
31. J. N. Chazalviel, *Coulomb Screening by Mobile Charges: Application to Materials Science, Chemistry, and Biology* (Springer, New York, 1999).
32. B. Yu, C. C. Pletka, J. Iwahara, Quantifying and visualizing weak interactions between anions and proteins. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2015879118 (2021).
33. Y. Bai *et al*., Quantitative and comprehensive decomposition of the ion atmosphere around nucleic acids. *J. Am. Chem. Soc.* **129**, 14981–14988 (2007).
34. M. Gebala, S. Bonilla, N. Bisaria, D. Herschlag, Does cation size affect occupancy and electrostatic screening of the nucleic acid ion atmosphere? *J. Am. Chem. Soc.* **138**, 10925–10934 (2016).
35. M. Gebala, D. Herschlag, Quantitative studies of an RNA duplex electrostatics by ion counting. *Biophys. J.* **117**, 1116–1124 (2019).
36. M. A. Hass, F. A. Mulder, Contemporary NMR studies of protein electrostatics. *Annu. Rev. Biophys.* **44**, 53–75 (2015).
37. L. Zandarashvili *et al*., Entropic enhancement of protein-DNA affinity by oxygen-to-sulfur substitution in DNA phosphate. *Biophys. J.* **109**, 1026–1037 (2015).
38. C. M. Paleos, P. Dais, Ready reduction of some nitroxide free radicals with ascorbic acid. *J. Chem. Soc. Chem. Commun.* **1977**, 345–346 (1977).
39. C. C. Pletka, R. Nepravishta, J. Iwahara, Detecting counterion dynamics in DNA-protein association. *Angew. Chem. Int. Ed. Engl.* **59**, 1465–1468 (2020).
40. J. Iwahara, C. Tang, G. Marius Clore, Practical aspects of ^{1}H transverse paramagnetic relaxation enhancement measurements on macromolecules. *J. Magn. Reson.* **184**, 185–195 (2007).
41. F. Delaglio *et al*., NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
42. W. Lee, M. Tonelli, J. L. Markley, NMRFAM-SPARKY: Enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327 (2015).
43. D. Nguyen, Z. A. Hoffpauir, J. Iwahara, Internal motions of basic side chains of the Antennapedia homeodomain in the free and DNA-bound states. *Biochemistry* **56**, 5866–5869 (2017).
44. P. R. Bevington, D. K. Robinson, *Data Reduction and Error Analysis for the Physical Sciences* (McGraw-Hill, New York, USA, ed. 3, 2003).
45. T. J. Dolinsky *et al*., PDB2PQR: Expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* **35**, W522–W525 (2007).
46. M. H. Olsson, C. R. Søndergaard, M. Rostkowski, J. H. Jensen, PROPKA3: Consistent treatment of internal and surface residues in empirical pK_a predictions. *J. Chem. Theory Comput.* **7**, 525–537 (2011).
47. L. Li, C. Li, Z. Zhang, E. Alexov, On the dielectric "constant" of proteins: Smooth dielectric function for macromolecular modeling and its implementation in DelPhi. *J. Chem. Theory Comput.* **9**, 2126–2136 (2013).
48. E. F. Pettersen *et al*., UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).

Downloaded by guest on June 15, 2021