Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

J Am Chem Soc. Author manuscript; available in PMC 2015 June 25.

Published in final edited form as:

J Am Chem Soc. 2013 November 13; 135(45): 16968–16976. doi:10.1021/ja406995j.

Protein Dielectric Constants Determined from NMR Chemical Shift Perturbations

Predrag Kukic[†,‖], **Damien Farrell**[†], **Lawrence P. McIntosh**[‡], **Bertrand García-Moreno E.**[§], **Kristine Steen Jensen**[⊥], **Zigmantas Toleikis**[⊥], **Kaare Teilum**[⊥], and **Jens Erik Nielsen**[†,∇,*]

[†]School of Biomolecular and Biomedical Science, Centre for Synthesis and Chemical Biology, UCD Conway Institute, University College Dublin, Belfield, Dublin 4, Ireland [‡]Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver, BC Canada [§]Department of Biophysics, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218, USA [⊥]Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, 2200 Copenhagen N, Denmark

Abstract

Understanding the connection between protein structure and function requires a quantitative understanding of electrostatic effects. Structure-based electrostatics calculations are essential for this purpose, but their use have been limited by a long-standing discussion on which value to use for the dielectric constants (ε_{eff} and ε_p) required in Coulombic models and Poisson-Boltzmann models. The currently used values for ε_{eff} and ε_p are essentially empirical parameters calibrated against thermodynamic properties that are indirect measurements of protein electric fields. We determine optimal values for ε_{eff} and ε_p by measuring protein electric fields in solution using direct detection of NMR chemical shift perturbations (CSPs). We measured CSPs in fourteen proteins to get a broad and general characterization of electric fields. Coulomb's law reproduces the measured CSPs optimally with a protein dielectric constant (ε_{eff}) from 3 to 13, with an optimal value across all proteins of 6.5. However, when the water-protein interface is treated with finite difference Poisson-Boltzmann calculations, the optimal protein dielectric constant (ε_p) rangedsfrom 2-5 with an optimum of 3. It is striking how similar this value is to the dielectric constant of 2-4 measured for protein powders, and how different it is from the ε_p of 6-20 used in models based on the Poisson-Boltzmann equation when calculating thermodynamic parameters. Because the value of $\varepsilon_p = 3$ is obtained by analysis of NMR chemical shift perturbations instead of thermodynamic parameters such as pK_a values, it is likely to describe only the electric field and thus represent a more general, intrinsic, and transferable ε_p common to most folded proteins.

[*]**Corresponding Author**jens.nielsen@gmail.com.
[‖]**Present Address** Department of Chemistry, University of Cambridge, Lensfield Road, CB2 1EW, Cambridge, UK
[∇]**Present Address** Protein Design, Novozymes A/S, Brudelysvej 26, 2880 Bagsvaerd, Denmark

ASSOCIATED CONTENT

Supporting Information. Supporting Materials and Mehods; Supporting Results; Figure S1. Example of pH-dependent CSP on an amide group; Figure S2-S7. Structural distribution of CSPs in six proteins; Figure S8. Optimal values of ε_p and ε_{eff} using cumulative ^{15}N and $^{1}H^{N}$ CSPs measured in all proteins; Table S1. RMSD between experimental and calculated Δδef; Table S2. Sets of crystal structures used to ensembles of static structures. This material is available free of charge via the Internet at http://pubs.acs.org.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Introduction

Some of the most fundamental biochemical reactions, such as enzymatic catalysis[1], redox reactions[2], H^+ transfer[3], and ion homeostasis[4] are governed by electrostatic effects. To understand the structural and physical basis of such biological processes, it is necessary to know the magnitude and molecular determinants of electrostatic forces and energies in proteins. Owing to the difficulties inherent to the experimental measurement of electrostatic effects in proteins, we typically use structure-based calculations to estimate electrostatic energies in proteins. These theoretical calculations are an essential tool for dissecting structure-function relationships and properties of biomolecules, but they are notoriously sensitive to the input structure[5,6] and to the parameters used such as the dielectric constants, and the charge-radius force field[7]. In particular, the value of the dielectric constants in these calculations remains highly contentious. What is clear is that the accuracy and utility of computational methods for structure-based electrostatics calculations is limited by our inability to describe dielectric effects quantitatively. Here we present data suggesting what the optimal value of the protein dielectric constant is when calculating electric fields with a Poisson-Boltzmann model framework and when using a simple Coulombic model.

Most structure-based calculations of electrostatic fields treat some part of the protein-water system as a dielectric continuum whose polarizability is described implicitly by a dielectric constant. To maximize the ability of a theoretical model to reproduce dielectric properties of proteins, the parameters that it employs (i.e. the charge-radius force field and its dielectric constant) are usually calibrated against benchmarks consisting of thermodynamic parameters for simpler systems, such as solvation free energies of ions in different polar solvents[8], changes in stability induced by changes in pH, in ionic strength, or by mutations[9], pK_a values[10], peptide acidity constants[11] and redox potentials[12]. The problem is that although these thermodynamic parameters do reflect the magnitude of the electrostatic potential, they represent a convolution of many other factors as well. The dielectric constants obtained by calibration against thermodynamic data are therefore model dependent and experiment-dependent. They are empirical parameters calibrated to reproduce experimental benchmarks and to account implicitly for any physical factors that are not treated explicitly in the models[13]. In the present paper we focus on measuring protein electric fields via NMR spectroscopy and on using these experimental measurements to guide electrostatic field calculations. To this end, we analyze the measured electric field-dependent chemical shifts to extract the corresponding dielectric constants that reproduce them most accurately when using the Poisson-Boltzmann equation or Coulomb's law.

Spectroscopic observables such as Stark shifts and NMR chemical shifts offer a more direct measure of the magnitude and direction of electric fields in biomolecules than thermodynamic parameters such as pK_a values. In proteins, chemical shifts of ^1H, ^{15}N and ^{13}C nuclei measured with NMR spectroscopy represent a particularly rich source of information about electric fields. The relationship between the electric field, **E,** and the chemical shift reported by a nucleus, δ_{ef}, was first identified by Pople[14] and later formulated by Buckingham[15] in what is known as Buckingham's equation:

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$\delta_{ef} = \boldsymbol{A}_{\gamma}\mathbf{E}_{y} + \boldsymbol{B}_{\gamma,\gamma}\mathbf{E}_{\gamma}^{2} + \boldsymbol{A}_{\gamma\delta}\frac{\partial \mathbf{E}_{\gamma}}{\partial \mathbf{r}_{\delta}} \quad (1)$$

$\mathbf{E}_{\gamma}$ in Buckingham's equation is the component of the uniform electric field at the nucleus along the coordinate γ, and $\partial\mathbf{E}_{\gamma}/\partial\mathbf{r}_{\delta}$ represents its gradient along the coordinate δ (See fig. S1b). A_{γ}, $B_{\gamma,\gamma}$ and $A_{\gamma\delta}$ are well-defined nucleus-dependent response properties that describe the sensitivity of the chemical shift δ_{ef} to the applied uniform electric field and to its gradient. This sensitivity is particularly significant in $^{1}H^{N}$ and ^{15}N nuclei involved in highly polarizable covalent bonds, such as the backbone N–H and C–N bonds. Therefore, chemical shifts reported by these nuclei can be used as high resolution probes of through-space electrostatic effects in proteins.

Stark shifts have already been shown to be useful as probes of electric fields in a few proteins and in an enzyme active site[16-21]. Chemical shift perturbations (CSP) measured with NMR spectroscopy hold promise of being equally useful to this end. CSPs can be induced by changing the charge state of a protein either by removing or inserting charged residues[16], by binding charged ligands, by changing the ionic strength of the solution, by incorporation NMR sensitive nuclei[22] or by changing pH[17,19].

The CSP reported by a nucleus, $\Delta\delta_{tot}$, can be affected primarily by three factors. It can be caused directly by changes in the electric field, $\Delta\delta_{ef}$, by changes in magnetic effects, $\Delta\delta_{m}$, or by conformational changes, $\Delta\delta_{s}$,:

$$\Delta\delta_{tot} = \Delta\delta_{ef} + \Delta\delta_{m} + \Delta\delta_{s} \quad (2)$$

To obtain as clean a measurement of the electric field effect as possible it is important to reduce the contribution of non-electrostatic effects to the measured $\Delta\delta_{tot}$. For this reason, only proteins that do not experience significant pH-induced magnetic and conformational changes were studied[16,17,19]. In these proteins, $\Delta\delta_{tot}$ reflects predominantly the $\Delta\delta_{ef}$ component, whose magnitude and direction is predictable (Eq. 1). The two other components, $\Delta\delta_{m}$ and $\Delta\delta_{s}$, make minor contributions. The first of these, $\Delta\delta_{m}$, is well-defined[23,24] and often negligible except near aromatic rings that are structurally perturbed over the course of a titration. The second, $\Delta\delta_{s}$, results from any titration-induced changes in structure, such as changes in torsion angles, bond lengths, bond angles or hydrogen bonding geometry close to a reporter nucleus[25]. The direction and magnitude of this component is expected to show a complex dependence on the structural environment of a given reporter nucleus[25]. Therefore, for the proteins studied here, $\Delta\delta_{tot}$ represents a measurement of the electric field, $\Delta\delta_{ef}$, with an added amount of noise proportional to $(\Delta\delta_{s}+\Delta\delta_{m})$, with a direction that is uncorrelated with the direction of $\Delta\delta_{ef}$. Any one measurement of the electric field at a single nucleus can include an error from conformational changes or magnetic effects. However, through simultaneous measurements performed for all nuclei in a single or several proteins, this error cancels out and hence reliable values of $\Delta\delta_{ef}$ can be obtained.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Previous measurements of electric fields with Stark shifts or NMR chemical shifts have been limited to a few individual proteins[16-19,26] or to a single location inside an enzyme active site[20,21]. To characterize the propagation of electric field in proteins in a more systematic manner, we used chemical shift perturbations measured with NMR spectroscopy in nine proteins, in four variants of these proteins, and in one enzyme-substrate intermediate. This large dataset allowed a self-consistent and systematic description of electric field propagation inside proteins, which in turn allowed an unprecedented determination of the value of the dielectric constant of proteins.

Materials and Methods

ACBP and hGRX

^{13}C, ^{15}N labeled bovine ACBP and hGRX were prepared as described previously[27,28]. A 1mM sample of the proteins was prepared in 100 mM KCl, 1 mM DSS, 0.02% NaN_3 and 10% D_2O. NMR spectra of ACBP (hGRX) were recorded in the pH interval from 3.3 (2.7) to 11.0 (8.9), in steps of approximately 0.25 by addition of NaOH. All spectra at each pH value were referenced to DSS. The pK_a values for His residues in ACBP were determined from the ^{1}H-^{13}C HSQC, and for Asp/Glu from the ^{13}CO and H^N spectrum[29]. The pK_a values for Cys residues in hGRX were determined from 2D ^{1}H-^{13}C HSQC of the aliphatic side-chain resonances.

Prediction of CSPs

The $\Delta\delta_{ef}$ was calculated as a difference in chemical shielding of the nucleus upon the (de)protonation of an ionizable group using Buckingham's equation (Eq. 1). Nuclear shielding polarizabilities, A_γ, for ^{15}N and $^{1}H^N$ nuclei were used from[19]. Although values for the second term in Buckingham's formula, $B_{\gamma,\gamma}$, have not been reported for ^{15}N and $^{1}H^N$ in the literature, their negligible contribution to $\Delta\delta_{ef}$ was assumed based on data reported for ^{19}F and ^{13}C nuclei[22,30]. The third term in Buckingham's formula, the gradient contribution, was calculated according to[31]. The electric field, $\mathbf{E}_\gamma$, at the position of backbone ^{15}N and $^{1}H^N$ nuclei was calculated with the PBE implemented in the APBS program (single Debye-Hückel boundary condition; final grid resolution of 0.25 Å/grid point)[32] and with Coulomb's law. The values of ε_p and ε_{eff} were taken from the range [1, 30] with a step of 0.5. 20 ns MD simulations were generated as described in detail in SI.

MD simulations

MD simulations of 20 ns length were performed for each of the six protein systems using the GROMACS package with the Amber ff99SB force field[33] and the TIP3P water model[34]. A real space cut-off of 10.0 Å on non-bonded interactions, periodic boundary conditions and particle-mesh Ewald electrostatics were used throughout the simulation. Prior to MD simulations, the systems were neutralized to a total charge of zero by adding Cl^- and Na^+ ions, energy minimized and equilibriated in the NVT ensemble over 200 ps. The temperature was controlled with a Berendsen coupling[35] to a water bath at 300 K using a temperature-coupling constant of 0.1 ps. Snapshot structures were taken every 10 ps starting at 5 ns of the simulation time and each snapshot was used in CSP predictions (Eq. 1).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Results

Protein selection

Some of the pH-dependent ^{15}N-^{1}H HSQC spectra used in this analysis were recorded previously and were generously made available by other researchers. Included in this set were: (I) bovine beta-lactoglobulin (βLG)[36], (II) B1 domain of protein G (GB1)[37], (III) plastocyanin from *A. variabilis* (A.v. Pc)[19], (IV) plastocyanin from *Phormidium laminosum* (P.l. Pc)[38-40], (V) xylanase from *Bacillus circulans* (B.c. xylanase)[41], (VI) the catalytic domain of β-(1,4)-glycosidase Cex from *Cellulomonas fimi* (CexCD)[42], and (VII) a highly stable form of staphylococcal nuclease (SNase) known as Δ+PHS[43]. The pH-dependent ^{15}N-^{1}H HSQC spectra for two additional proteins were measured specifically for this study: (I) bovine acyl-coenzyme A binding protein (ACBP), and (II) human glutaredoxin 1 (hGRX). Spectra previously recorded for four variants of SNase with substitutions D21N, L38D, L38E and L38K, and for one long-lived enzyme-substrate intermediate of CexCD covalently modified by 2,4-dinitrophenyl 2-deoxy-2-fluoro-β-cellobioside (2FCb-CexCD)[44] were also analyzed. The complete data set consisted of 1861 ^{15}N and 1861 $^{1}H^{N}$ chemical shifts of backbone amides measured as a function of pH.

Identification of useful chemical shift perturbations (CSPs)

The titration of ionizable groups in a protein is reflected in the chemical shifts of backbone amides. To extract $\Delta\delta_{tot}$ values that originate from the titration of a single ionizable group, pH-dependent chemical shifts reported by a ^{15}N or $^{1}H^{N}$ nucleus were fitted using an automated F-test based algorithm[45,46]. Only $\Delta\delta_{tot}$ values with a magnitude above the uncertainty level for the specific nucleus (0.1 ppm for ^{15}N and 0.03 ppm for $^{1}H^{N}$ (ref. 47)) were considered. Each extracted $\Delta\delta_{tot}$, denoted “observed CSP”, was assigned as arising from the (de)protonation of a particular ionizable group if its apparent pK_a matched the pK_a of the ionizable group (Figure S1 and Supporting Results in the Supporting Information). The spatial distribution of nuclei reporting $\Delta\delta_{tot}$ values associated with a specific ionization event provides information about the distribution of the electric field emanating from a single charge. Instances where it was clear that an ionization event did not induce a CSP in a reporter nucleus were also recorded. These are termed ‘absent CSPs’ and they are important, as these are cases where the contributions by a charge to the electric field dissipated before they could induce a CSP in the reporter nucleus.

The majority of the ionizable groups in the nine proteins that were investigated are solvent-exposed. Their pK_a values are not perturbed significantly relative to the normal pK_a values of isolated amino acids in water. Therefore, the ionizable groups of the same type of residues (Asp/Glu, His, Lys) tend to have similar pK_a values in the proteins. This similarity of pK_a values precludes unambiguous assignment of the majority of the $\Delta\delta_{tot}$ values in the experimental data to any one specific ionizable group. For this reason, these groups were not included in the analysis. The only groups used in the analysis of electric fields were the groups where the CSP could be associated with the change in charge of a specific residue. On this basis, 9 ionizable groups were identified whose apparent pK_a values could be identified unambiguously in the experimental CSPs (details given in SI). These groups are: His14 (pK_a=7.1) and His30 (pK_a=6.7) in ACBP; His61 (pK_a=7.1) and His92 (pK_a=5.1) in

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

A.v. Pc; His 92 (pK_a=5.1) in P.l. Pc; Glu44 (pK_a=4.0) and Glu94 (pK_a=4.0) in hGRX, His80 (pK_a=7.9) in CexCD and Asp21 (pK_a=6.5) in SNase. The final dataset used to analyze electric fields included the effects of nine ionizable groups on a total of 786 ^{15}N and 786 $^{1}H^{N}$ nuclei.

Statistics on the number of observed and absent CSPs associated with each ionizable group are presented in Table 1. The spatial distribution of CSPs is illustrated for ACBP in Figure 1 and in Figures S2-S7 for all remaining proteins. The number of observed CSPs originating from His80 in CexCD and Asp21 in SNase was further validated by requiring the absence of their observed CSPs in the spectra of the 2FCb-CexCD intermediate or in the D21N variant of SNase. The nuclei inside the protein core do not report CSPs originating from the solvent-exposed His14 in ACBP and Glu94 and Glu44 in hGRX. Indeed, only a handful of nuclei along the protein surface were sensitive to the electric field changes originating from the ionization of these groups. Only 9 ^{15}N (and 9 $^{1}H^{N}$) CSPs were associated with the titration of His14 in ACBP, whereas the corresponding numbers associated with the titration of Glu94 and Glu44 in hGRX were 10 ^{15}N (and 8 $^{1}H^{N}$) and 14 ^{15}N (and 11 $^{1}H^{N}$), respectively. In contrast, the CSPs associated with the ionization of partially or fully buried groups, such as His30 in ACBP, His92 and His61 in A.v. Pc, His92 in P.l. Pc, His80 in CexCD and Asp21 in SNase, are reported by nuclei positioned both deep in the core or on the protein surface. The number of CSPs originating from the titration of these groups were: His30 in ACBP (33 ^{15}N and 31 $^{1}H^{N}$), His92 (36 ^{15}N and 34 $^{1}H^{N}$) and His61 (30 ^{15}N and 26 $^{1}H^{N}$) in A.v. Pc, His92 (38 ^{15}N and 32 $^{1}H^{N}$) in P.l. Pc, His80 (57 ^{15}N and 46 $^{1}H^{N}$) in CexCD and Asp21 (33 ^{15}N and 31 $^{1}H^{N}$) in SNase.

Differences in the number of the observed CSPs originating from the titration of solvent exposed and buried ionizable groups can be rationalized in terms of expected general properties of electric fields in proteins: electric fields originated from solvent exposed side chains are screened by bulk water, which has a high dielectric response. In contrast, electric fields inside the protein can propagate over larger distances because inside the protein the dielectric response is more muted than in water. Similar arguments can be applied to CSPs caused by titration-induced conformational changes ($\Delta\delta_s$ in Eq. 2): the titration of a buried ionizable group is more likely to induce a large conformational perturbation. Hence the ionization of a buried group can affect a much larger number of nuclei than the ionization of solvent exposed groups, which are less likely to induce conformational reorganization.

Corroboration that CSPs reflect through-space electric field effects

To analyze CSPs rigorously in terms of propagation of electric fields originating from the ionizable moieties, it is important to ascertain that the assigned $\Delta\delta_{tot}$ reflect primarily the propagation of through-space electrostatic effects. To test this we compared the direction of the observed $\Delta\delta_{tot}$ (downfield/upfield shift with increased pH) with the corresponding direction of the electric field-induced CSPs, $\Delta\delta_{ef}$, calculated from the crystal structures using Buckingham's equation and Coulomb's law. Whereas the local structural contribution to the CSP, $\Delta\delta_s$, is expected to result in either upfield or downfield shifts depending on the structural environment of a given nucleus, the direction of the electric field-induced CSP, $\Delta\delta_{ef}$, is always determined by the position of the titrating site relative to the reporter nucleus

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(Eq. 1). Therefore, a high correlation between the directions predicted from Buckingham's equation and the direction of the observed $\Delta\delta_{tot}$ would indicate that electrostatic effects make the dominant contribution to the observed $\Delta\delta_{tot}$.

A strong correlation was found between the directions of the observed $\Delta\delta_{tot}$ and the calculated $\Delta\delta_{ef}$ values; 226 of 260 (87%) observed ^{15}N $\Delta\delta_{tot}$ and 189 of 228 (83%) observed $^{1}H^{N}$ $\Delta\delta_{tot}$ displayed shifts in the direction predicted by Buckingham's equation. These results show convincingly that the electric field makes the dominant contribution to the observed $\Delta\delta_{tot}$. Nevertheless, to further minimize artifacts of non-electrostatic effects in the electric field analysis, we excluded 34 of the ^{15}N and 39 of the $^{1}H^{N}$ nuclei for which an agreement in CSPs direction was not obtained.

Dielectric constants obtained by analysis of electric fields with static structures

The ability of the Poisson-Boltzmann equation (PBE) and of Coulomb's law to reproduce experimental results was tested using both single static structures (see Figure 2 for an overview) as well as ensembles of structures. When a Coulomb interaction measured experimentally is analyzed with a primitive continuum model based on Coulomb's law, the dielectric parameter that reproduces the experimental quantity is referred to as the effective dielectric constant, ε_{eff} [13]. More sophisticated models based on numerical solutions of the Poisson-Boltzmann equation describe dielectric effects of the protein with a parameter that we call ε_p. CSP values were used to identify values of ε_{eff} and ε_p that most accurately reproduce the experimental data.

Analysis with Coulomb's law—Analysis of CSPs with Coulomb's law yields an optimal value of ε_{eff} of 6.5 for the ^{15}N datasets, with an RMSD of 0.28 ppm between calculated and experimentally observed CSPs. For the $^{1}H^{N}$ datasets, the optimal ε_{eff} is 7.0 with an RMSD of 0.07 ppm; however ε_{eff} values from 6 to 15.0 (Table 2 and Table S1) yield RMSD values that are witin 0.1 ppm. This wide range of ε_{eff} values thus predicts the observed CSPs almost as accurately as the optimal ε_{eff} value. If we instead calculate the fraction of CSPs predicted correctly we find that ε_{eff} values between 6 and 15 do best for the observed CSPs, whereas the absent CSPs are predicted accurately only with ε_{eff} values above 15. This is illustrated for ACBP in Figure 3, and for all other proteins in Figures S3-S8. Figure 3B displays the RMSD between calculated and observed $\Delta\delta_{ef}$ (solid line), and calculated and absent CSPs (dashed line). The fraction of reporter nuclei in which the observed $\Delta\delta_{ef}$ (solid line) and the absent CSPs (dashed line) predicted within the experimental error are displayed in Figure 3C. In all proteins it was observed that optimal ε_{eff} in terms of RMSD, predicts the existence of CSPs in 20% of ^{15}N and $^{1}H^{N}$ nuclei for which CSPs were not observed in the NMR spectra. This highlights the inability of Coulomb's law to correctly model the effect of solvent screening in a protein-water system.

Analysis with a finite difference Poisson-Boltzmann model—The optimal value of ε_p obtained by analysis of both ^{15}N and $^{1}H^{N}$ datasets in all proteins is 3.0. (Tables 2 and S1). With this value the RMSD between the observed and calculated $\Delta\delta_{ef}$s is 0.25 ppm for ^{15}N data and 0.06 ppm for $^{1}H^{N}$ data. The analysis with the PBE methodology yielded the following observations: (1) The magnitudes of the observed $\Delta\delta_{ef}$ are predicted in the PBE

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

model most successfully when a low value of ε_p is used regardless of whether performance was evaluated with RMSD or with fractional statistics. (2) The optimal value of $\varepsilon_p = 3$ predicted both the presence and the absence of the CSPs in the reporter nuclei. (3) The optimal values of ε_p obtained by analysis of both observed and absent CSPs confirmed the ability of the PBE method to account for the screening of electric fields with the two-dielectric system (protein, ε_p; solvent, ε_s=80). (4) The range of ε_p values obtained by analysis of nine different ionizable groups in six different proteins is unexpected low, and falls in the range 2 to 5, which is much narrower than the range obtained in calculations with Coulomb's law (Table 2). (5) The value of $\varepsilon_p = 3$ contrasts sharply with the values of ε_p in the range 6 to 20 commonly used in structure-based pK_a calculations with PB models to maximize agreement between experimental and calculated thermodynamic parameters such as pK_a values and redox potentials[48]. The value of value of $\varepsilon_p = 3$ is strikingly similar to the value of the protein dielectric constant measured with dry proteins in experiments with parallel plate capacitors.

The difference between the magnitude of the protein dielectric constant resolved by analysis with the one dielectric model in Coulomb's law and with the two-dielectric calculations with PB electrostatics suggest that screening by bulk water plays an important role in attenuating the propagation of electric fields inside proteins. This effect cannot be accounted for correctly with Coulomb's law because the role of bulk water is subsumed into the single dielectric constant used in the analysis, but the effect is reproduced quite well in finite difference Poisson-Boltzmann calculations. This shows that screening by bulk solvent cannot be neglected in theoretical models for the calculation of electrostatic effects inside proteins.

Analysis of electric fields with conformational ensembles

NMR chemical shifts reflect ensemble-averaged chemical properties of a nucleus. For this reason, it is necessary to move beyond analysis based on calculations with single static structures and to examine the extent to which the conformational diversity present in the ensemble in solution could reproduce the structural diversity probed by the NMR spectroscopy experiment in solution. To this end ensembles were generated from crystal structures and from molecular dynamics (MD) trajectories.

Ensemble of static crystal structures—An ensemble of static structures was generated using structures available in the Protein Data Bank (Table S2). These different structures were obtained from different crystallization conditions, with variants, or in the presence of ligands. When the number of existing crystal structures for any one protein of interest was less than 5, models solved by NMR spectroscopy were included in the ensemble. The total number of structures used to create the ensemble of each protein (average C_α-RMSD to the representative static structure shown in parenthesis) was: 5 (1.5 Å) for ACBP, 4 (0.8 Å) for A.v. Pc, 8 (0.5 Å) for P.l. Pc, 14 (0.3 Å) for CexCD and 36 (0.9 Å) for SNase.

Use of these ensembles of static structures improved quantitative agreement between the experimentally observed and calculated CSPs (Table S1, Figures 3 and S3-S8).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Improvement in RMSD between the observed and calculated $\Delta\delta_{ef}$ was on average 0.02 ppm (0.01 ppm) for the PBE model and the ^{15}N ($^{1}H^{N}$) chemical shift dataset. The corresponding improvement for calculations with Coulomb's law was 0.04 ppm (0.02 ppm). The largest improvement was for ACBP, 0.03 ppm for both ^{15}N and $^{1}H^{N}$ datasets, whose structures show the largest diversity, with an average RMSD of 1.46 Å. These results, similar to those reported by Best *et al.*[49], highlight the importance of accurate accounting of protein dynamics in making comparisons with ensemble-averaged experimental data. Although the improvement is quite modest in our case, the results suggest that even a modest number of structures determined under different conditions can improve the agreement with NMR chemical shift experiments although the improvement is quite modest in our case.

Ensemble from MD trajectory—Ensembles of MD structures were generated from 20 ns trajectories in which the ionizable groups of interest were kept in the neutral state. The average RMSD between the generated conformations and the initial static structures were: 2.18 Å for ACBP, 2.43 Å for A.v. Pc, 3.61 Å for P.l. Pc, 1.74 Å for hGRX, 1.52 Å for CexCD, and 1.46 Å for SNase.

Use of ensembles of MD-generated proteins did not improve quantitative agreement between experimentally measured and calculated secondary titrations with the PBE model or with Coulomb's law (Table S1, Figures 3 and S2-S8). The RMSD between observed and calculated CSPs increased on average by 0.01 ppm (0.01 ppm) using the PBE model and ^{15}N (or $^{1}H^{N}$) chemical shift dataset, and 0.01 ppm (0.01 ppm) using Coulomb's model. The inability of the MD-generated conformational ensembles to reproduce ensemble-averaged experimental data may originate with differences in time scales between the NMR experiment and the relatively short MD simulations, or with the inherent inability of MD simulations to sample relevant conformations far from the crystal structure. Whereas chemical shifts are averaged by dynamic processes on the microsecond time-scale, the MD trajectories were much shorter and consequently not able to account for the true ensemble of the protein in solution, as reflected in the NMR spectroscopy data.

Microscopic modeling

The response of permanent and induced dipoles to the gain or loss of charge is not homogenous throughout a protein. In continuum electrostatic models this reorganization is supposed to be accounted for implicitly with the protein dielectric constant. The agreement between measured and calculated data can be greatly affected by how this response is handled in the calculations. To examine this question, the reorganization of permanent dipoles was modeled explicitly by generating two separate ensembles of MD conformations. In one of the ensembles the ionizable group of interest was treated as being charged and in the other one it was assigned the neutral state. Using the ensembles in both states, $\Delta\delta_{ef}$ values were calculated using Coulomb's law and Buckingham's formula and then compared against the experimental results. This is a crude approach to microscopic modeling of structural reorganization in response to ionization events.

With this more microscopic approach it was more difficult to reproduce the experimental data with structure-based calculations. The RMSD for ^{15}N or $^{1}H^{N}$ data increased by 0.12

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

ppm or 0.02 ppm, respectively, relative to the calculations with single static structures. A more detailed microscopic approach was not pursued, as this is an open-ended problem in sampling that probably reflects differences in the conformational reorganization sampled in solution and *in silico*. The calculations that were performed were detailed enough to reveal how challenging it will be to reproduce the propagation of electric fields inside the anisotropic and heterogeneous interior of proteins using detailed atomistic methods devoid of any implicit parameters, such as dielectric constants, to account for attenuation of electric fields by polarization of the protein.

Discussion

Structure-based computation of electrostatic effects in proteins is challenging because quantitative treatment of the dielectric response of proteins is not straightforward. Most models treat at least some parts of the protein-water system as a dielectric continuum and use dielectric constants to capture implicitly the effects of polarization. The problem is that the magnitude of the dielectric constants used in the different models is not transferable because these are simply *ad hoc* empirical parameters without any clear physical meaning.

The dielectric constant measured in parallel plate capacitor experiments using dry protein powders (ε_{powder}) is 2 in the high-frequency limit, and closer to 4 in the equilibrium regime and in the presence of a few water molecules per protein[50]. In contrast, ε_p is usually assigned values of 6 to 20 in PBE calculations to maximize agreement between experimental and calculated thermodynamic data (e.g. pK_a values, redox potentials, pH dependence of ΔG)[10,13,48,51-58]. The difference between ε_{powder} = 2 to 4 and ε_p = 6 to 20 is due mainly to two factors: 1) in dehydrated proteins all dynamic relaxation processes are suppressed and we therefore have a muted dielectric response, and 2) ε_p is found by maximizing the agreement with experimental thermodynamic properties that are influenced by many factors in addition to the electric field. The ε_p = 6 to 20 value are high because they account implicitly for all forms of dynamic, structural reorganization that are usually not treated explicitly in a computational model. In the case of thermodynamic data describing ionization events, the local structural reorganization is likely significant as indicated by the high dielectric constants needed to reproduce the experimental values. The correct explicit modeling of these reorganization events present a formidable challenge for quantitative structure-based energy calculations.

The experimental value of ε_p = 3 as determined here by analysis of CSPs, is significant because this dielectric constant likely represents a value closer to a general ε_p (i.e. the dielectric constant within a PBE framework that describes only the electric field) of proteins in solution. This value was obtained by analysis of the behavior of several proteins, with different folds, and with different structural and functional properties. This suggests that ε_p = 3 represent a general, intrinsic and probably transferable property of proteins with a size and dynamic characteristics similar to the set of proteins studied here. The similarity with the values of ε_p = 2 to 4 measured with dry protein powders suggests that it reflects primarily the polarization of electrons and contributions from fluctuations of permanent dipoles although this agreement could be coincidental.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

The difference between the value of ε_p measured in this study using the propagation of electric fields and the value of ε_p obtained through reverse pK_a calculations, reflects differences in the goals and the physical processes monitored in these two different types of experiments. The goal of a reverse pK_a calculation is to identify the value ε_p that reproduces an experimental pK_a value. In such an analysis, anything that affects the pK_a and is not treated explicitly in the model becomes absorbed in the dielectric constant. Because ionizable side chains (the reporters of pK_a values) are located in solvent-exposed regions that can undergo considerable structural reorganization, the ε_p values they report reflect the substantial structural heterogeneity and reorganization experienced by ionizable side chains when they titrate. ε_p values obtained by analysis of pK_a values therefore reflect dynamic properties of proteins and implicitly describe non-electrostatic effects that cause significant perturbation of pK_a values. In contrast, the sensors of the charge-dipole interactions examined in this study are the non-titratable backbone amides. Unlike side chains, the mobility of backbone C-N and N-H bonds is much lower and therefore less sensitive to ionization-induced structural changes. The backbone amides themselves do not titrate and it is therefore possible to measure an ε_p value in the absence of nearby structural reorganization - such a measurement is never possible in a reverse pK_a calculation. In fact, the data used in this analysis purposely excluded cases where conformational reorganization could contribute to the CSP, thus the value of $\varepsilon_p = 3$ obtained by analysis of electric fields reflected in CSPs is likely to reflect more directly an intrinsic ε_p of proteins, devoid of contributions from structural reorganization, water penetration, or any other factor that can affect the magnitude of electrostatic effects in proteins. These results suggest that the most accurate method for calculation of electrostatic effects in proteins (with PBE solvers) should use $\varepsilon_p = 3$ to reproduce implicitly the contributions from electronic polarization and high frequency dipolar fluctuations of the protein matter, with all other dynamic contributions treated explicitly. The latter if of course very challenging because it is difficult to sample correctly all the different microstates that contribute to thermodynamic properties of proteins in solution. It is important to note that the value of ε_p found here is dependent on the scaling factors used in Buckingham's equation, and although all evidence points to the values used here being quite accurate (see supplementary material), it is nevertheless possible that future findings could change this and thus affect our conclusions on the value of ε_p.

We used the exquisite sensitivity of ^{15}N and $^{1}H^{N}$ chemical shifts to the magnitude and direction of electric fields along the highly polarizable C-N and N-H bonds to assess the magnitude of electric fields in 14 proteins. The magnitudes of the measured electric field emanating from a single charge ranged from 1.2 MV/cm to 20.0 MV/cm, with a resolution of 1.2 MV/cm. The lower limit, and hence the sensitivity of the measurements, was determined by the uncertainty of the extracted $\Delta\delta_{ef}$ from the CSPs (0.1 ppm for ^{15}N and 0.03 ppm for $^{1}H^{N}$ nuclei[47]); the upper limit was detected for Leu36 in *A.v. Pc* (−1.66 ppm). To illustrate the significance of such large variations in the measured electric fields emanating from a single charge, consider how a single charge could affect the rate of a catalytic reaction. If a charge were able to produce electric fields in this range, parallel to a dipolar transition state separating unit charges over a distance of 1Å, the resulting increase on the reaction rate would range from 1.5 (1.2 MV/cm) to more than 2000 fold (20.0 MV/cm).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Conclusions

To reproduce electric fields inside proteins accurately with continuum electrostatics models it is necessary to treat the dielectric properties of proteins accurately. The CSP has exquisite information about the protein dielectric constant that has not been tapped before. Our goal is to determine the range of the dielectric constants, ε_p in Poisson-Boltzmann electrostatics and ε_{eff} in Coulomb's law, that best reproduce the experimental CSPs. The optimal value of ε_p in the PBE was of particular interest because the value of this quantity and its physical meaning has been the most contentious topic in theoretical calculations of electrostatic interactions in the last three decades[18,59].

The results obtained in this study have two important conclusions. (1) A low value of $\varepsilon_p = 3$ and a narrow range of values of ε_p from 2 to 5 is optimal for reproducing electric fields in the proteins studied. Because these ε_p values were obtained by analysis of NMR chemical shift perturbations and not by analysis of thermodynamic variables such as pK_a values, which are contaminated by non-electrostatic factors, they likely represent the best estimates of the most general and intrinsic value of the ε_p of proteins. (2) The differences between results from analysis with PB electrostatics and with a simple Coulomb formalism underscore the importance of screening effects of bulk water, which attenuates the propagation of the electric field inside proteins. This screening effect should always be accounted for in the theoretical models. It is reproduced well with a simple PBE-based model that treats the dielectric response of water and of protein with separate dielectric constants, but it is underestimated by a simple Coulomb expression as shown by the significant over-prediction of CSPs.

Of the observables in NMR spectroscopy, only the chemical shift is sensitive to long-range electric field effects[60]. Therefore, the optimal values of ε_p and ε_{eff} obtained from the analysis of CSPs cannot be corroborated independently with data from other high-resolution spectroscopic observable, such as J-coupling, residual dipolar coupling, order parameters or nuclear Overhauser effect restraints. The analysis presented here, however, was performed separately for different proteins and using experimental data reported by different nuclei types, ^{15}N and $^{1}H^{N}$ nuclei, known to report different sensitivities to the different factors that determine pH-dependent CSPs (e.g. $\Delta\delta_m$, $\Delta\delta_s$ and $\Delta\delta_{ef}$ in Eq. 2)[37]. In all protein systems studied the results obtained for optimal ε_p are correlated and therefore validate its narrow range in proteins in general. The almost identical, self-consistent results for optimal values of ε_p and ε_{eff} obtained by analysis of data reported by two nuclei types, ^{15}N and $^{1}H^{N}$, further support the validity of this analysis and its conclusions. These results represent yet another application of the NMR chemical shift; the most readily and accurately measured NMR observable that should lead to improved electrostatic calculation methodology and understanding of protein properties and functions in general. They constitute the first general, systematic description of general properties of electric field propagation probed by chemical shifts in proteins.

The observables obtained from NMR spectroscopy experiments are rich in information useful to validate the ability of computational models to account for the ensemble properties of proteins. We therefore probed the ability of ensembles, defined in two different ways, to

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

reproduce the diversity probed by the NMR chemical shift in solution. Ensembles consisting of multiple crystal structures of the same protein (or similar proteins), gave a slightly improved prediction of the measured CSPs. In contrast, conformational averaging generated from MD trajectories was unable to capture the dynamics reflected in experimental data acquired on the NMR time scale. These results suggest that standard MD simulations should be used with caution for this purpose.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENT

This work was supported by an SFI PIYR award (04/YI1/M537) to J.E.N., an SFI Research Frontiers Programme award (08/RFP/BIC1140) to J.E.N., a PhD scholarship from The Irish Research Council for Science Engineering and Technology (IRCSET) Graduate Research Education Programme (GREP) to P.K, a grant from the Danish Natural Science Research Council (272-060251) to K.T and grant GM073838 from the National Institutes of Health USA to B.G-M. E. NMR instrument support (L.P.M.) was provided by the Canadian Institutes for Health Research, the Canadian Foundation for Innovation, the British Columbia Knowledge Development Fund, the UBC Blusson Fund, and the Michael Smith Foundation for Health Research. The authors are grateful for experimental data provided by Peter B. Crowley, Jens Jørgen Led, Kazumasa Sakurai and Yuji Goto.

REFERENCES

1. Warshel A, Florian J. Proc. Natl. Acad. Sci. USA. 1998; 95:5950–5955. [PubMed: 9600897]
2. Varadarajan R, Zewert TE, Gray HB, Boxer SG. Science. 1989; 243:69–72. [PubMed: 2563171]
3. Luecke H, Richter HT, Lanyi JK. Science. 1998; 280:1934–1937. [PubMed: 9632391]
4. Doyle D, Cabral JM, Pfuetzner R, Kuo A, Gulbis J, Cohen S, Chait B, MacKinnon R. Science. 1998; 280:69–77. [PubMed: 9525859]
5. Nielsen JE, Andersen KV, Honig B, Hooft RW, Klebe G, Vriend G, Wade RC. Protein Eng. 1999; 12:657–662. [PubMed: 10469826]
6. Nielsen JE, McCammon JA. Protein Sci. 2003; 12:313–326. [PubMed: 12538895]
7. Carstensen T, Farrell D, Huang Y, Baker NA, Nielsen JE. Proteins. 2011; 79:3287–3298. [PubMed: 21744393]
8. Warshel A, Russell SQ. Rev. Biophys. 1984; 17:283–422.
9. Isom D, Castaneda CA, Cannon BR, Velu P, Garcia-Moreno B. Proc. Natl. Acad. Sci. USA. 2010; 107:16096–16100. [PubMed: 20798341]
10. Tynan-Connolly BM, Nielsen JE. Protein Sci. 2007; 16:239–249. [PubMed: 17189477]
11. Hernandez G, Anderson JS, LeMaster DM. Biochemistry. 2009; 48:6482–6494. [PubMed: 19507827]
12. Churg A, Warshel A. Biochemistry. 1986; 25:1675–1681. [PubMed: 3011070]
13. Schutz CN, Warshel A. Proteins. 2001; 44:400–417. [PubMed: 11484218]
14. Schneider W, Bernstein H, Pople J. J. Chem. Phys. 1958; 28:601–607.
15. Buckingham AD. Can. J. Chem. 1960; 38:300–307.
16. LeMaster DM, Minnich M, Parsons PJ, Anderson JS, Hernández G. J. Inorg. Biochem. 2006; 100:1410–1412. [PubMed: 16712938]
17. Boyd J, Domene C, Redfield C, Ferraro MB, Lazzeretti P. J. Am. Chem. Soc. 2003; 125:9556–9557. [PubMed: 12903999]
18. Kukic P, Farrell D, Søndergaard CR, Bjarndóttir U, Bradley J, Pollastri G, Nielsen JE. Proteins. 2010; 78:971–984. [PubMed: 19894279]
19. Hass MA, Jensen MR, Led JJ. Proteins. 2008; 72:333–343. [PubMed: 18214953]
20. Webb LJ, Boxer SG. Biochemistry. 2008; 47:1588–1598. [PubMed: 18205401]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

21. Suydam IT, Snow CD, Pande VS, Boxer SG. Science. 2006; 313:200–204. [PubMed: 16840693]
22. Pearson J, Oldfield E, Lee F, Warshel A. J. Am. Chem. Soc. 1993; 115:6851–6861.
23. Perkins SJ, Dower SK, Gettins P, Wain-Hobson S, Dwek RA. Biochem. J. 1977; 165:223–225. [PubMed: 921745]
24. Williamson MP, Asakura T. J. Magn. Reson. B. 1993; 101:63–71.
25. de Dios A, Laws D, Oldfield E. J. Am. Chem. Soc. 1994; 116:7784–7786.
26. Anderson JS, Hernandez G, LeMaster DM. Biophys. Chem. 2009; 141:124–130. [PubMed: 19200635]
27. Thomsen J, Kragelund B, Teilum K, Knudsen J, Poulsen F. J. Mol. Biol. 2002; 318:805–814. [PubMed: 12054824]
28. Jensen K, Winther J, Teilum K. J. Am. Chem. Soc. 2011; 133:3034–3042. [PubMed: 21323311]
29. Tollinger M, Forman-Kay J, Kay L. J. Am. Chem. Soc. 2002; 124:5714–5717. [PubMed: 12010044]
30. Augspurger J, Pearson J, Oldfield E, Dykstra C, Park KD, Schwartz D. J. Magn. Reson. 1992; 100:342–357.
31. Batchelor J. J. Am. Chem. Soc. 1975; 97:3410–3415.
32. Baker NA, Sept D, Holst MJ, McCammon JA. IBM J. Res. Dev. 2001; 45:427–438.
33. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Proteins. 2006; 65:712–725. [PubMed: 16981200]
34. Mahoney M, Jorgensen W. J. Chem. Phys. 2000; 112:8910–8922.
35. Berendsen H, Postma J, van Gunsteren W, Di Nola A, Haak J. J. Chem. Phys. 1984; 81:3684–3690.
36. Sakurai K, Goto Y. Proc. Natl. Acad. Sci. USA. 2007; 104:15346–15351. [PubMed: 17878316]
37. Tomlinson JH, Green VL, Baker PJ, Willamson MP. Proteins. 2010; 78:3000–3016. [PubMed: 20715051]
38. Crowley PB, Otting G, Schlarb-Ridley, Canters BG, Ubbink M. J. Am. Chem. Soc. 2001; 123:10444–10453. [PubMed: 11673974]
39. Crowley P, Matias P, Khan A, Roessle M, Svergun D. Chem. Eur. J. 2009; 15:12672–12680. [PubMed: 19834935]
40. Crowley P, Matias P, Mi H, Firbank S, Banfield M, Dennison C. Biochemistry. 2008; 47:6583–6589. [PubMed: 18479147]
41. McIntosh LP, Naito D, Baturin SJ, Okon M, Joshi M, Nielsen JE. J. Biomol. NMR. 2011; 51:5–19. [PubMed: 21947911]
42. Schubert M, Poon D, Wicki J, Tarling C, Kwan E, Nielsen J, Withers S, McIntosh L. Biochemistry. 2007; 46:7383–7395. [PubMed: 17547373]
43. Castañeda C, Fitch C, Majumdar A, Khangulov V, Schlessman J, García-Moreno B. Proteins. 2009; 77:570–588. [PubMed: 19533744]
44. White A, Tull D, Johns K, Withers S, Rose D. Nat. Struct. Biol. 1996; 3:149–154. [PubMed: 8564541]
45. Farrell D, Sa-Miranda E, Webb H, Georgi N, Nielsen JE. Proteins. 2010; 78:843–857. [PubMed: 19899070]
46. Webb H, Tynan-Connolly BM, Lee GM, Farrell D, O'Meara F, Søndergaard CR, Teilum K, Hewage C, McIntosh LP, Nielsen JE. Proteins. 2011; 79:685–702. [PubMed: 21287606]
47. Bartik K, Redfield C, Dobson CM. Biophys. J. 1994; 66:1180–1184. [PubMed: 8038389]
48. Antosiewicz J, McCammon J, Gilson M. J. Mol. Biol. 1994; 238:415–436. [PubMed: 8176733]
49. Best R, Lindorff-Larsen K, DePristo M, Vendruscolo M. Proc. Natl. Acad. Sci. USA. 2006; 103:10901–10906. [PubMed: 16829580]
50. Bone S, Pethig R. J. Mol. Biol. 1985; 181:323–326. [PubMed: 2984434]
51. Nielsen JE, Vriend G. Proteins. 2001; 43:403–412. [PubMed: 11340657]
52. Demchuk E, Wade RC. J. Phys. Chem. 1996; 100:17373–17387.
53. Karshikoff A. Protein Eng. 1995; 8:243–248. [PubMed: 7479686]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

54. Yang AS, Gunner MR, Sampogna R, Sharp K, Honig B. Proteins. 1993; 15:252–265. [PubMed: 7681210]
55. Alexov EG, Gunner MR. Biophys. J. 1997; 72:2075–2093. [PubMed: 9129810]
56. Bashford D, Karplus M. Biochemistry. 1990; 29:10219–10225. [PubMed: 2271649]
57. Gordon JC, Myers JB, Folta T, Shoja V, Heath LS, Onufriev A. Nucleic Acids Res. 2005; 33:W368–371. [PubMed: 15980491]
58. Warwicker J. Protein Sci. 1999; 8:418–425. [PubMed: 10048335]
59. Warshel A, Sharma P, Kato M, Parson W. Biochim. Biophys. Acta. 2006; 1764:1647–1676. [PubMed: 17049320]
60. Osapay K, Case D. J. Am. Chem. Soc. 1991; 113:9436–9444.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

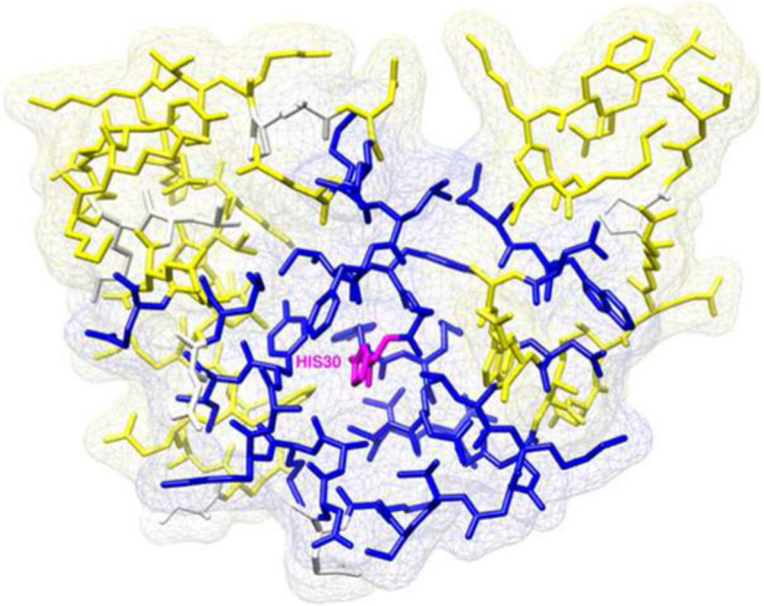

Figure 1.

Electric field propagation in ACBP caused by the protonation of His30 (magenta) and measured at backbone ^{15}N: (blue) residues with identified CSPs above 0.1ppm; (yellow) residues with CSPs bellow 0.1ppm; (white) unassigned resonances and proline residues.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

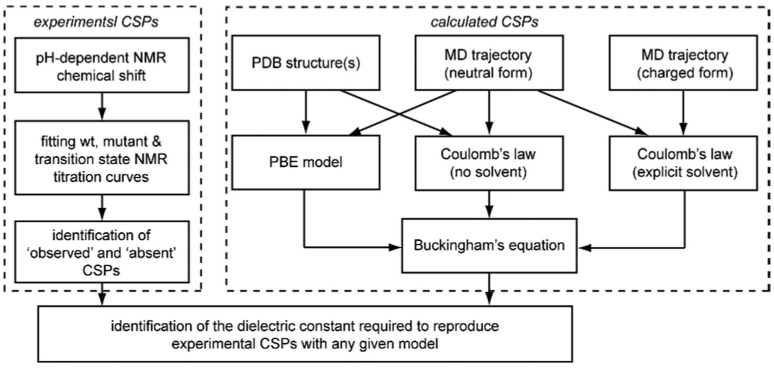

Figure 2.

Overview of the approach for using CSPs for benchmarking computational methods for prediction of electric fields.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

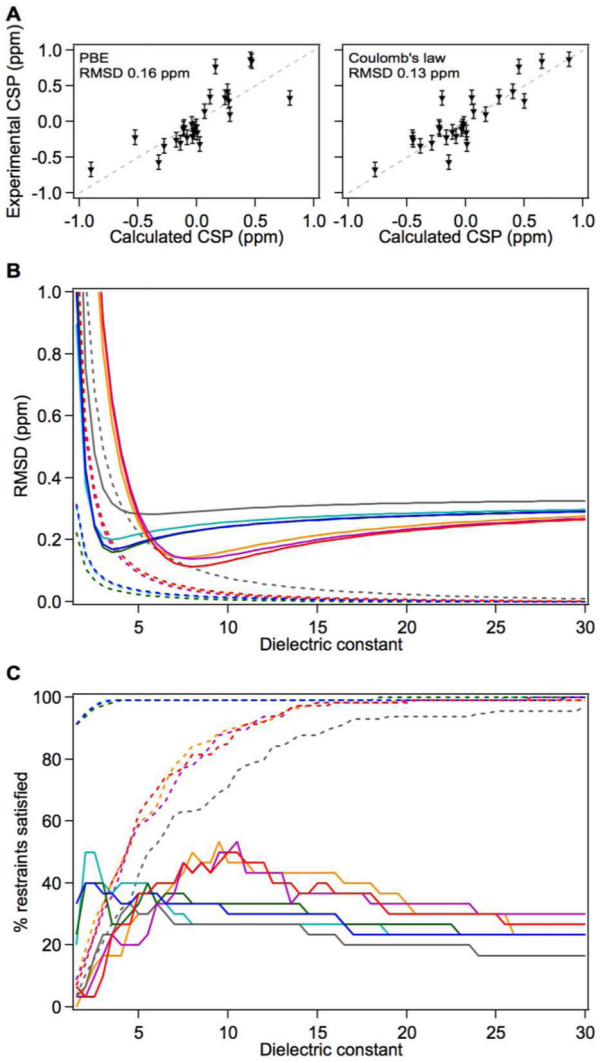

Figure 3.

Optimal values of ε_p and ε_{eff} using ^{15}N CSPs measured in ACBP. **(A)** Correlation between experimentally measured ^{15}N $\Delta\delta_{ef}$ emanating from the His30 charge and calculated ^{15}N $\Delta\delta_{ef}$ obtained from the static structure PDB ID: 1HB6 and with optimal dielectric constant values that minimize RMSD (ε_p=3.5; ε_{eff}=7.5). **(B)** Average RMSD between measured and predicted ^{15}N $\Delta\delta_{ef}$ in ACBP as a function of dielectric constant. Solid lines depict the RMSD of observed $\Delta\delta_{ef}$ and dashed lines depict the RMSD of absent $\Delta\delta_{ef}$. Color scheme: calculations with static structure with PBE (cyan) or with Coulomb's law (orange), calculations with an ensemble of static structures using PBE (blue) or Coulomb's law (red), calculations with an ensemble generated with MD simulations and PBE (green) or Coulomb's law (magenta), microscopic approach (grey). **(C)** Fraction of ^{15}N reporter nuclei whose predicted $\Delta\delta_{ef}$ are within the nucleus-specific experimental errors of the observed $\Delta\delta_{ef}$ (solid line) and absent $\Delta\delta_{ef}$ (dashed line). Color scheme is identical to (B).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Dataset used to measure the dielectric constant of proteins.

Protein (pH range)	Titr.group	pKa	side chain SASA (Å^2)	# of identified CSPs ^{15}N ($^1H^N$) $\Delta\delta_{ef}$	# of absent CSPs ^{15}N ($^1H^N$) $\Delta\delta_{ef}$
ACBP (pH 3.0-8.9)	His14	7.1	36	9 (9)	70 (66)
	His30	6.7	6	33 (31)	46 (43)
A.v. Pc (pH 4.0-8.4)	His92	5.1	7	36 (34)	56 (58)
	His61	7.1	14	30 (26)	62 (66)
P.l. Pc (pH 4.2-8.3)	His92	5.1	9	38 (32)	58 (63)
hGRX (pH 2.8-11.1)	Glu44	4.0	19	10 (8)	80 (80)
	Glu94	4.0	24	14 (11)	80 (80)
CexCD (pH 4.1-9.3)	His80	7.9	2	57 (46)	213 (232)
SNase (pH 1.4-11.1)	Asp21	6.5	2	33 (31)	72 (87)
Total				**260 (228)**	**737 (775)**

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Optimal values of the protein dielectric constant obtained with Poisson-Boltzmann electrostatics (ε_p) and with Coulomb's law (ε_{eff}) by analysis of CSPs reported by ^{15}N or $^{1}H^{N}$ (given in brackets) nuclei using a single crystal structure, an ensemble of crystal structures, or an ensemble generated with MD simulations.

Protein/group	ε_p from PBE			ε_{eff} from Coulomb's law			
	static structure	ensemble of stat.struct.	MD simulations	static structure	ensemble of stat.struct.	MD without water	full MD
ACBP	**3.5(3.5)**	**3.5(3.0)**	**3.5(4.0)**	**7.5(8.0)**	**8.0(9.0)**	**8.0(9.0)**	**5.5(8.5)**
Hisl4	2.0(2.0)	3.0(3.5)	3.5(6.0)	7.0(17.0)	10.0(20)	9.5(21.0)	4.0(60)
His30	3.5(3.5)	4.0(3.0)	3.5(3.0)	7.5(7.5)	7.0(7.5)	7.5(7.5)	6.0 (8.0)
A.v. Pc	**3.0(3.0)**	**2.5(3.0)**	**2.0(2.5)**	**5.5(4.5)**	**5.5(5.5)**	**5.5(7.5)**	**11.0(31.5)**
His6l	5.0(5.0)	4.0(9.5)	2.5(7.5)	8.0(16.5)	8.0(18.0)	9.5(20.0)	60(41.5)
His92	2.0(2.5)	2.0(2.0)	2.0(2.5)	4.0(4.0)	4.0(4.0)	4.0(5.0)	5.5(17.5)
P.l. Pc	**2.0(2.5)**	**2.0(2.5)**	**2.5(3.5)**	**4.0(4.5)**	**4.0(4.5)**	**4.0(5.0)**	**60(60)**
His92	2.0(2.5)	2.0(2.5)	2.5(3.5)	4.0(4.5)	4.0(4.5)	4.0(5.0)	60(60)
hGRX	**2.0(2.0)**		**2.0(2.0)**	**11.5(5.5)**		**11.5(4.5)**	**60(4.0)**
Glu44	2.0(2.0)		1.5(2.0)	14.5(3.5)		16.0(4.0)	60(3.5)
Glu94	2.0(3.5)		2.0(3.0)	11.0(10.0)		9.5(9.5)	5.5(5.0)
CexCD	**5.0(5.0)**	**5.0(5.0)**	**4.5(10.0)**	**7.5(7.0)**	**7.5(7.0)**	**8.0(9.0)**	**4.5(60)**
His80	5.0(5.0)	5.0(5.0)	4.5(10.0)	7.5(7.0)	7.5(7.0)	8.0(9.0)	4.5(60)
SNase	**2.5(3.5)**	**2.5(3.5)**	**2.0(2.5)**	**6.0(10.0)**	**6.0(9.0)**	**7.0(9.0)**	**46.5(51)**
Asp21	2.5(3.5)	2.5(3.5)	2.0(2.5)	6.0(10.0)	6.0(9.0)	7.0(9.0)	46.5(51)
ALL	**3.0(3.0)**	**3.0(3.5)**	**3.0(4.0)**	**7.0(6.5)**	**6.0(7.0)**	**7.0(7.0)**	**20.0(28.5)**