



Published in final edited form as:

Proteins. 2011 December ; 79(12): 3249–3259. doi:10.1002/prot.23194.

The pK_a Cooperative: A Collaborative Effort to Advance Structure-Based Calculations of pK_a values and Electrostatic Effects in Proteins

Jens E. Nielsen¹, M. R. Gunner^{2,*}, and E. Bertrand García-Moreno³

¹School of Biomolecular and Biomedical Science, Centre for Synthesis and Chemical Biology, UCD Conway Institute, University College Dublin, Belfield, Dublin 4, Ireland ²Department of Physics, City College of New York, New York, NY 10031 ³Department of Biophysics, Johns Hopkins University, Baltimore, MD 21218

Abstract

The pK_a Cooperative <http://www.pkacoop.org> was organized to advance development of accurate and useful computational methods for structure-based calculation of pK_a values and electrostatic energy in proteins. The Cooperative brings together laboratories with expertise and interest in theoretical, computational and experimental studies of protein electrostatics. To improve structure-based energy calculations it is necessary to better understand the physical character and molecular determinants of electrostatic effects. The Cooperative thus intends to foment experimental research into fundamental aspects of proteins that depend on electrostatic interactions. It will maintain a depository for experimental data useful for critical assessment of methods for structure-based electrostatics calculations. To help guide the development of computational methods the Cooperative will organize blind prediction exercises. As a first step, computational laboratories were invited to reproduce an unpublished set of experimental pK_a values of acidic and basic residues introduced in the interior of staphylococcal nuclease by site-directed mutagenesis. The pK_a values of these groups are unique and challenging to simulate owing to the large magnitude of their shifts relative to normal pK_a values in water. Many computational methods were tested in this 1st Blind Prediction Challenge and critical assessment exercise. A workshop was organized in the Telluride Science Research Center to assess objectively the performance of many computational methods tested on this one extensive dataset. This volume of *PROTEINS: Structure, Function, and Bioinformatics* introduces the pK_a Cooperative, presents reports submitted by participants in the blind prediction challenge, and highlights some of the problems in structure-based calculations identified during this exercise.

Keywords

proteins; pK_a values; electrostatics; pH; ionizable; staphylococcal nuclease; simulation

This volume of *Proteins: Structure, Function, and Bioinformatics* is dedicated to protein electrostatics and specifically, to a series of papers focused on the critical evaluation of computational methods for structure-based calculations of pK_a values of ionizable groups in proteins. These include papers that specifically analyze the pK_a s of a series of introduced buried changed residues in staphylococcal nuclease,^{1–12} those that carry out simulation on other systems^{13–18}, those that provide new experiments to challenge simulation,^{19–22} and a

*Corresponding Author: 212-650-5557; gunner@sci.ccnycunyu.edu.

technical overview.²³ The papers describe results from the 1st Blind Prediction Challenge for pK_a calculations organized by the pK_a Cooperative (<http://www.pkacoop.org>). The central goals of this Cooperative are to organize critical assessment exercises that benchmark the performance of current computational methods for structure-based calculation of pK_a values and related properties in proteins; to support the design of novel or improved computational methods for structure-based energy calculation; and to encourage experimental research to generate new and innovative data that challenge simulation methods and enhance fundamental insight into electrostatic effects in proteins. This paper describes the aims and organization of the pK_a Cooperative, the data used for its 1st Blind Prediction Challenge, and discusses some of the conclusions from this first critical assessment exercise.

pK_a Values Matter

The structure and function of proteins and nucleic acids cannot be understood fully without detailed understanding of contributions from electrostatic forces.^{24,25} Approximately 300 pK_a values had been measured in proteins and used as benchmarks values before 2009.^{26,27} Ionizable residues (Lys, Arg, His, Asp, Glu,) account for approximately 25% of all residues in the average globular protein.^{28,29} They determine or modulate many essential properties of proteins, including structure, function, stability, solubility, dynamics and inter-molecular interactions. The accurate prediction of pK_a values and electrostatic energies in proteins is therefore an important aspect of any structure-based energy calculation procedure.

The binding of protons (H^+) constitutes the smallest and least disruptive reaction that a protein can experience. If structure-based calculations cannot reproduce this, the simplest possible perturbation of a protein, attempts at calculating other, more complex processes, should be approached with great care. Because H^+ binding entails the creation or elimination of a formal charge, it is possible to examine the energetics of this process rigorously, starting from the physical principles of classical electrostatics and statistical thermodynamics.^{30–38} The electrostatic interactions between charged groups are among the strongest and most long-range interactions in biology. This is precisely the reason that ionizable groups can have significant influence on the structure, function, stability, and dynamics of proteins.

The properties of ionizable groups are influenced significantly by their surroundings. The pK_a values of acids and bases can be quite different when the ionizable moiety is surrounded by water or by protein, membrane or by other nonaqueous environments. For this reason, the pK_a values are very sensitive to changes in protein structure, to association with other proteins or with other macromolecules, including membrane bilayers, and even with small molecules. This sensitivity to the microenvironment is central to the essential role of H^+ transfer processes in biological energy transduction and that is what structure-based calculations attempt to reproduce quantitatively. This sensitivity is also why the calculation of pK_a values of ionizable groups in proteins is one of the best ways to test our understanding of the structural basis of the energetics and function of proteins.

Structure-based Simulations are Necessary

The experimental measurement of pK_a values can help in the interpretation of the structural basis of many properties of proteins. There are, however, limitations to the extent to which experimentation alone can be used to determine how the structure of a protein determines the observed pK_a values. Structure-based calculations that predict pK_a values through the application of physical models of electrostatic forces to high-resolution structures of proteins provide an essential quantitative link between structure and energetics, and therefore between structure and function. Validated or calibrated pK_a calculations also promise to be useful to predict pK_a values in the many situations where experimental measurements are

not possible. As with all simulations, the models underlying the methods for calculation of pK_a values determine the amount of physical insight that can be obtained from a calculation. The initial efforts of the pK_a Cooperative have been focused on attempts to reproduce the shifts in pK_a values of ionizable groups in proteins relative to their pK_a values in water. Electrostatic forces are assumed to be the major determinant of the pK_a value shifts in proteins, thus these simulations test the accuracy of calculation of electrostatic forces. The effects of pH on the stability of proteins, on their energetics, dynamics and interactions with other macromolecules and with small molecules, and the sensitivity of proteins to salts and to solutes that affect water properties, are all problems subsumed under the central problem of protonation thermodynamics. These are all areas considered by the pK_a Cooperative.

The Need for a Cooperative Effort

Attempts to use X-ray structures to calculate electrostatic effects in proteins have been ongoing for 40 years.^{31,39,40} During the last two decades it has become evident that the prediction of pK_a values in biological macromolecules is anything but straightforward.^{24,25,32–38} Approximately 300 pK_a values had been measured in proteins and used as benchmark values before 2009.^{26,27,41–45} Most of these are pK_a values of surface ionizable groups, whose properties are governed primarily by the dielectric response of water. The average measured perturbation in pK_a values for surface residues is 0.8 pK_a units.⁴³ Thus, early simulation methods struggled to match the data better than a simple assumption that pK_a values are equal to those of isolated residues in water (the NULL model).⁴⁶ Currently, pK_a calculations with many different computational models achieve RMSD values for surface residues of around 1 pK_a unit (at 25 °C this is equivalent to an error of 1.36 kcal/mol).²⁶ Thus, the errors are of similar magnitude to the protein-induced shifts in pK_a values – an unacceptable situation. Surprisingly, the accuracy of pK_a prediction algorithms has barely changed during the last decade, pointing towards stagnation in the field despite new methods and algorithms being published. Part of the problem has been the lack of useful data that could be used as a benchmark to challenge the calculations and to help identify problems in calculations that need to be corrected.

Despite progress, our ability to use structure to reproduce the effects of electrostatic interactions on proteins is very limited and insufficiently tested. Judging by the claims of the success and utility of structure-based energy calculation algorithms and methods, these problems are also not fully appreciated or acknowledged by some of the research groups active in the area of structure-based energy calculations. The same situation exists in the fields of computational protein/ligand docking and protein design. The pK_a Cooperative was organized primarily to help efforts by the computational community to develop improved methods for structure-based energy calculations, to generate more rigorous analyses of reported data, and to increase the awareness of the following issues:

1. Conclusions based solely on calculated/simulated data are likely to be misleading unless a rigorous error analysis is performed. As this is rarely done it is rarely possible to determine if a given structure-based calculation is sufficiently accurate to be useful for quantitative analysis of structure-function relationships in proteins.
2. Good agreement between experiment and calculation when using a small number of experimental data points may be coincidental. This is especially true if agreement is measured by a correlation coefficient, or if a data set is used that is not appropriate for gauging the performance of a computational method. Most data sets being used for evaluation of structure-based calculations are for water-exposed ionizable groups with very small pK_a shifts. These data are not useful to identify weaknesses and failures in calculations. One goal is to define a standard data set

that has been endorsed by the community, against which all computational methods should be tested to raise the bar for what constitutes a good agreement between experiment and simulation.

3. pK_a calculations are highly underdetermined since many parameters are needed to generate a pK_a value. Often only a few experimental pK_a values are available for any given protein. The validity of a calculation would be enhanced if many, related properties such as the pH dependence of protein stability, ionic strength effects and effects of mutations on electrostatic properties of proteins, were examined concurrently. Thus, the dimension of the comparison between calculated and measured data needs to be expanded. In addition, the sensitivity of the results to the parameters needs to be characterized better. Our lack of understanding of the parameters used also makes it difficult to compare the results and conclusions of different simulation techniques.
4. Conclusions from calculations need to be supported by rigorous calculation of thermodynamic quantities (pK_a values or ΔpK_a and the underlying Gibbs free energies) as well as their errors and sensitivity to parameters. Although empirical methods for structure-based pK_a calculations have made great strides, the use of these methods for electrostatics calculations without rigorous attempts to calculate thermodynamic parameters is of concern.

It is clear that understanding the structural and physical origins of electrostatic effects in proteins in detail remains extremely challenging and that it should be considered as a work in progress. Exaggerated claims of successful accounting of the structural origins of electrostatic effects should be viewed with skepticism until algorithms are subjected to more stringent testing than has been possible in the past. Given the importance of protein electrostatics to a large variety of problems in biology, further efforts in this area are warranted. The pK_a Cooperative was organized to foment and to steer further research and developments in protein electrostatics, and to raise the stringency with which structure-based calculations are evaluated and applied.

The pK_a Cooperative

The pK_a Cooperative (<http://www.pkacoop.org>) was established in response to the urgent need for reliable and useful methods for structure-based calculation of pK_a values and electrostatic energies of proteins, especially for methods that incorporate the underlying physics of electrostatics and the statistical thermodynamics of pH-dependent phenomena realistically. The ultimate goal of research in this area is to improve our understanding of the structural and physical origins of electrostatic effects and to be able to predict pK_a values and electrostatic energy from structure. It is important that structure-based calculations and predictions be accurate for the right physical reason so they allow us to elucidate the physics underlying all electrostatic effects in proteins.

In addition to encouraging more rigorous simulations, the pK_a Cooperative will try to act as a clearinghouse to organize, curate and distribute experimental data useful for testing structure-based electrostatics calculations. It will organize critical assessment exercises for blind and objective benchmarking of methods for structure-based calculations of pK_a values and other protein properties that are sensitive to electrostatic forces. In this role it will gather experimental data before publication and notify research groups interested in carrying out blind tests of their computational methods. It will also encourage new experimental investigation of electrostatic phenomena that can contribute the physical insight needed to guide the development of new computational methods. The Cooperative encourages close interactions between the participants; this should lead to collaboration between different

simulation techniques and to experimental studies being designed for the explicit purpose of testing improved algorithms.

The pK_a Cooperative was organized primarily by laboratories with expertise in structure-based calculations when a large set of experimental pK_a values for internal ionizable groups in proteins became available from the García-Moreno lab at Johns Hopkins University.⁴⁷⁻⁵⁰ This dataset, if used properly, had the promise to enable unprecedented, rigorous benchmarking of computational methods. It consists of pK_a values of Lys, Asp, Glu and Arg at 25 internal positions in staphylococcal nuclease and crystallographic structures of many of these proteins.⁴⁷⁻⁵⁰ In 2007, when the experimental analysis of the 100 proteins was nearing completion, it became clear it would be useful if the community was organized to predict the results as a challenge to individual methods. To enable critical assessment of computational methods in a blind manner, the data were withheld from publication, deposited with the pK_a Cooperative, and unavailable to the participants contributing to the 1st Blind Prediction Challenge. The group of laboratories, many of them represented in this volume of *Proteins*, agreed to submit pK_a predictions prior to a meeting in Telluride Science Research Center in the summer of 2009. Results of their calculations were submitted to the Cooperative prior to the workshop. The workshop was sobering and invigorating as the exercise succeeded in identifying strengths but also weaknesses in the computational algorithms. Only 30% of the experimental data were released at the workshop to allow groups to refine their algorithms subsequent to the workshop, and to resubmit predictions for blind assessment. The goal of this critical benchmarking exercise was to identify areas where improvements are needed to enhance the performance of these algorithms, to look for trends in the performance of algorithms based on different approaches, and to identify the key problems where experimental input was needed. Although the blind prediction challenge was useful for individual groups to discover shortcomings about their own methods, this exercise has been less successful in bridging different simulation methods.

Data Used in the 1st Blind Prediction Challenge

The set of pK_a values that were used for this first large scale exercise organized by the Cooperative was unique as they represent 100 acidic and basic residues that were introduced into the interior of one protein using site-directed mutagenesis. The pK_a values were determined by the García-Moreno group for variants of staphylococcal nuclease (SNase) with Lys, Asp, Arg or Glu at 25 internal locations.⁴⁷⁻⁵⁰ The variant proteins were designed to position a single ionized group in the interior of SNase to measure the effect of desolvating the ionizable group and to evaluate plausible compensation from newly formed favorable interactions or from structural reorganization. This yielded experimental data about highly perturbed pK_a values for a large number of residues at different positions in the protein, and provided a unique dataset for blind predictions. The pK_a measurements were done either by linkage analysis of the pH dependence of stability measured with equilibrium denaturation measurements at different pH values and/or with NMR spectroscopy or direct potentiometric methods. The pK_a values of some internal groups in the set are shifted by almost 6 pK_a units relative to pK_a values of ionizable groups in water. At the time of the blind prediction exercise the majority of the pK_a values had not been released. Crystal structures of many of these variants were released and made available through the Protein Data bank in advance of publication.

The pK_a values of internal ionizable groups are very different from those on the surface of proteins. The latter are governed primarily by the dielectric properties of water and by the high flexibility of the protein-water interface. For this reason they tend to be very similar to the normal pK_a values of ionizable groups in water and can be reproduced with moderate success with a variety of different methods, including ones that may have empirical or

unphysical assumptions and parameters. Prior to the emergence of the SNase data set the community had been working with very few data points, which coupled to the use of many parameters in the calculations, made it particularly difficult to assess the validity of the representation of protein relaxation coupled to protonation changes. For example, the question of the meaning of the dielectric constant of a protein, and of the relevance of this question proper, had been impossible to examine in detail prior to the development of the data set of 100 SNase variants.

The pK_a values of internal residues engineered through site directed mutagenesis by the Hopkins group are highly anomalous and very different from the normal values of ionizable groups in water.⁴⁷⁻⁵⁰ Larger perturbations in pK_a values give a greater range for the targets and are much more challenging to simulate. Also, if methods cannot reproduce pK_a values for ionizable groups in proteins with accuracy higher than 1 pK_a unit, it is important that the experimental pK_a shifts used to evaluate computational methods be much larger than 1 pK_a unit. From the magnitude of the shifts in the pK_a values it was clear from the outset that this unique set of experimental data would allow unprecedented and stringent benchmarking of computational methods for structure-based calculations, and that it would help identify weaknesses and strengths in existing methods. The pK_a values of internal groups are exquisitely sensitive to local and global structural details; therefore they are also useful to test the ability of algorithms to reproduce contributions from a variety of structural factors that can affect the pK_a of these groups.

The shifts in pK_a values of the internal ionizable groups in SNase are always in the direction that promotes the neutral form of the ionizable groups. This suggests that they are primarily determined by the dehydration (desolvation) experienced by the ionizable groups when they are buried in the interior of the protein. The desolvation appears to be poorly counterbalanced by significant compensating factors to stabilize ionization. These internal ionizable groups test the ability of an algorithm to simulate structural relaxation of proteins coupled to the ionization of the internal groups. As these structural changes can involve a range of time and length-scales, their effect on the pK_a values of the internal ionizable groups proved to be challenging to calculate. The degree to which the pK_a values of the artificial internal ionizable groups engineered in SNase are comparable to naturally occurring internal ionizable groups has emerged as an important question that remains to be established experimentally. It is clear that many naturally occurring ionizable groups exist in microenvironments that have evolved to ensure that these internal groups are charged, whereas others can have highly shifted pK_a values comparable to those found in SNase.^{51,52}

Overview of Computational Methods Tested in the 1st Blind Prediction Challenge

One problem at the heart of pK_a calculations is that the ionization of acidic and basic residues in a protein can be coupled so the change in all protonation state must be calculated simultaneously. Many simulation techniques use Monte Carlo (MC) methods to generate the Boltzmann distribution of microstates as a function of pK_a .⁵³ The energies underlying the simulations can come from empirical force fields, from molecular mechanics (MM) force fields or from Density Functional Theory in quantum mechanics (QM) or QM/MM simulations. As there was scant QM analysis of the SNase variants this latter approach will not be considered here although this is an area that should receive considerable attention in the future. The key long-range electrostatic energies are determined from semi-empirical methods,^{2,5,6,13} from Poisson-Boltzmann (PB)^{3,4,7,9,11,12} or Generalized Born (GB) continuum methods,^{8,10} semi-microscopic Protein Dipole-Langevin Dipole (PLDL) methods¹⁸ or simulations with explicit waters.^{1,16} The calculations must take into account the motions of the protein that accompany the protonation/deprotonation events. These are

incorporated by empirical screening functions,^{2,5,6,13} by adding a discontinuous dielectric constant for water and protein,^{3,4,7,9,11,12} or by explicit motions ranging from sampling of polar proton positions,^{7,9,11,12} of side chain rotamers³ to full degrees of freedom in a molecular dynamics (MD) simulation.^{1,16} Accounting for the heterogeneous response of proteins is generally considered the chief difficulty in modeling pK_a values in proteins. Alexov, Mehler et al provided an overview of the various methods used here.²³

Among the methods tested in the 1st Blind Prediction Challenge in 2009 were: (1) calculations based on empirical methods (Mehler,⁶ Jensen,¹³ and Olsson⁵ labs); (2) calculations based on continuum electrostatics methods (Knapp,⁴ Gunner,³ Alexov,¹¹ Song,⁷ Word/Nicholls¹² and Warwicker⁹ labs); (3) calculations based on constant pH/MD (Shen,⁸ Williams/McCammon,¹⁰ Brooks¹ and Baptista¹⁶ labs), (4) calculations based on other microscopic or semi-microscopic methods (Cui (QM/MM)⁵⁴ and Warshel (PLDL)⁵⁵ had published on the SNase system prior to the 1st Blind Prediction Challenge). Most of the groups that worked on the SNase data had already simulated much of the same available dataset of largely surface residues. All found that the buried residues presented here were considerably more challenging to model and that they offer the opportunity for meaningful benchmarking.

Summary of Results of the 1st Blind Prediction Challenge

In comparing the individual methods, it was rare for a method to predict the direction of the shift in pK_a relative to the solution value incorrectly. However, the methods that used Monte Carlo sampling with restricted conformational searching tended to exaggerate the stabilization of the neutral forms of the internal ionizable groups. In contrast, the methods that used MD based sampling tended to exaggerate the stabilization of the ionized introduced residues.

The most successful methods belong to no particular class of algorithms, and among the best methods we thus find PB-based methods using Monte Carlo sampling, MD-based methods and a rule-based method, with very little separating the performance of these algorithms. In terms of physical understanding it is disappointing that a simple rule-based method performs as well as the more physics-based methods (MD and MC-PB); Carstensen *et al.*¹³ put forward a possible explanation for why the empirical methods work well elsewhere in this volume. In summary, the pK_a calculation methods still tend to have a significant ad hoc component, which is disappointing given the large amount of effort put into developing and testing protein electrostatics algorithms and energy functions over the last 2–3 decades. The blind challenge did help identify several areas where progress is needed:

1. For algorithms for calculation of pK_a values (and any structure-based energy calculation algorithm) the dependence of results on the parameters, input structure and general setup must be examined and reported. A common failure is to omit the effect of using different input structures.⁵⁶ The analysis of convergence through sufficient sampling becomes particularly important in methods that rely on MD simulations.^{1,16} Clear guidelines for sensitivity analysis to describe the dependence of results on input parameters has been lacking.^{5,12}
2. Metrics for rigorous comparison of calculated and experimental data, with control over sample size, number of parameters in the models, size of experimental pK_a shifts, etc, are needed. The use of F-tests and RMSD values is a good starting point but more needs to be done in this area.⁵⁷
3. Finally, the field should move beyond qualitative descriptions used to characterize the performance of pK_a calculation algorithms. Adjectives such as 'good', 'above average', 'highly significant' etc. need to be replaced with hard numbers that

describe accurately the performance of each model. To this end it is essential to continue to organize blind prediction exercises, to ensure that algorithms are tested rigorously.

Brief Overview of Some Simulation Methods Tested

Continuum Electrostatics

Approximately half the submissions use Monte Carlo sampling of ionization states in a protein with the electrostatic energies obtained via the Poisson-Boltzmann equation. These include the work of Alexov,¹¹ Gunner,³ Knapp,⁴ Song,⁷ Warwicker,⁹ and Word¹² These methods are fast compared to those that depend on MD techniques, and they use more easily separable parameters thus enabling test of the sensitivity to specific input variables. By design, these methods account for many of the motions coupled to ionization events by mean field methods encapsulated in the dielectric constant. As the range of explicitly determined interactions is much smaller than in MD-based approaches it is also easier to distinguish specific interactions that affect a pK_a value. As all of these methods tend to over stabilize the neutral form of the internal ionizable group, often leading to shifts of several pH units in the calculated pK_a values. Flexibility was increased implicitly by increasing the dielectric constant,³ by improving the search for cavities that would allow for greater and implicit water penetration,⁴ and by smoothing the dielectric surface.¹² Conformational degrees of freedom were explored more explicitly by devices ranging from side chain rotamer sampling³ to the use of MD¹¹ and Rosetta⁷ to generate multiple protein structures. One other general conclusion about these methods is that the input parameters may still not be fully optimized. Thus, changing charge sets and different Lennard Jones parameters were found to improve the agreement between calculated and measured pK_a values.^{3,12} These types of calculations tend to depend significantly on the input structures. Calculations with x-ray structures of proteins were usually better than those performed with *in silico* models of the variants of SNase with internal ionizable groups.

Constant pH MD Simulations

Calculations based on constant pH/MD were submitted by Shen,⁸ Williams/McCammon,¹⁰ Brooks¹ and Baptista.¹⁶ These simulations generally achieved a closer match between measured and calculated pK_a values than did continuum electrostatic methods with Monte-Carlo sampling, especially when a low dielectric constant was used in the continuum calculations. In contrast to what was observed in continuum electrostatics calculations, the MD-based approaches tended to over-stabilize the ionized form of the internal ionizable group, showing that flexibility might be exaggerated in these calculations. As the conformation of the whole protein must come to equilibrium with each charge state of all residues these methods all find that the changes in protonation can lead to structural changes that converge slowly. In addition, the strength of interactions quite distant from the ionizing group can affect the pK_a and these appear to be exaggerated in some of these calculations. Thus, the possibility that the parameters overestimate the strength of somewhat distant pairwise interactions between ionizable groups was mentioned as possible cause of poor performance on several residues in several of the CPHMD papers.

Empirical Methods

Calculations based on empirical methods were contributed by (Mehler,⁶ Jensen,¹³ and Olsson.⁵ These methods rely on parameterized analytical energy functions, with the distribution of protonation states generated by Monte Carlo sampling. These methods are very fast, often taking mere seconds to perform a full protein titration. These programs need training sets to generate parameters; the SNase variants represented a set of residues that were novel and thus not well modeled. However, with modest tuning of the parameters these

programs continue to provide fairly good match to data. This is the group of simulations where the changes made to better simulate the SNase data are least likely to be transferable to other situations.

Areas where improvement might be possible

A variety of methods found that explicit handling of structural relaxation coupled to the ionization of internal groups greatly improved the results. This could be achieved by increasing the protein dielectric constant, allowing for water penetration, increasing the sampling of structures optimized around ionized or neutral residue of interest, or running full MD simulations. Although all these changes were used to improve the agreement between experimental and calculated pK_a values, it is not known if the structures resulting from MD, Rosetta or other simulation methods reproduce real conformational changes. The question remains whether the improvement of calculations by manipulating the protein dielectric constant in continuum electrostatic methods means we better understand the dielectric response of the protein to ionization events, or that a high dielectric constant minimizes incorrect interaction energies. Likewise the changes in molecular structure found by MD-based methods may indicate how the structure of the protein changes in response to the ionization of an internal group, or it may simply reflect the properties of local minima achieved with the parameters used. This lack of knowledge arises because all pK_a calculations are highly underdetermined, with many parameters and atomic positions collectively leading to the treatment of a few protonation equilibria in a given protein. For example, Nielsen et al¹³ considered the possibility that completely random structural perturbations that produced somewhat less densely packed protein structures could lead to improved agreement between calculated and measured pK_a values of internal groups.

Data Analysis

One of the simple elements of analysis that is dealt with in the papers of Word/Nicholls¹² and Olsson⁵ in particular is the treatment of pK_a values that are beyond the limits of calculation or of measurement. These are dealt with rather inconsistently through the volume. Most analysis simply treated the error as zero if the simulation is lower/higher than the published pK_a limit for an acid/base. Calculated pK_a values are moved to be at the limit of simulation, often pH 0–14, rather than pushing the simulation to higher or lower values. The paper of Word and Nicholls¹² provides some better guidance for future benchmark tests. In addition, Olsson⁵ commented on the degrees of difficulty of different test cases, judging the larger the pK_a shift from solution the harder the pK_a should be to calculate. This is an important when comparing submissions that analyzed different sub-sets of the full dataset. Improving our understanding of the underlying statistics of 20 laboratories analyzing some or all of a 100 residues dataset will allow us to better realize when one method represent a breakthrough.

In Praise of Blind Predictions

The utility of blind predictions lies in the objectivity gained in testing a given method without the possibility of subconscious or instinctual tweaking of parameters to improve agreement between calculation and experiment. Blind predictions provide a true measure of the state of development of a particular approach and helps identify areas where improvements are necessary. At a certain point a community focused on the development of computational methods for prediction and simulation becomes confident of its success. That is precisely when computational methods should be subjected to blind prediction challenges. When post-dictions are attempted instead of predictions, the temptation to carry out calculations to the point where the simulated answer is close to the experimental one is very strong. In blind prediction exercises each participating lab must make decisions about their

best simulation practices and live with the outcome, afterwards using failures to improve their practices. In some areas, coarse-grained approximations contribute considerable insight into difficult problems, and in these cases a blind prediction challenge would be superfluous. In the case of protein electrostatics the field has matured to the point where methods for structure-based calculations have to be challenged and benchmarked to demonstrate that they can reproduce details of pK_a values and of the factors that determine them. That is the level of detail at which they have to operate.

Blind predictions are loaded with practical complications, and it speaks highly of this community that the 1st Blind Prediction Challenge leading to the Telluride Workshop took place in a spirit of deep curiosity and friendly cooperation. In the 1st Blind Prediction Challenge we carefully avoided creating a climate in which some labs were recognized as winners and others as losers. To the extent that the field of protein moves forward as a result of this exercise, all are winners. The time frame of the type of challenge that was organized can also be very problematic for laboratories that do not have sufficient staff to attempt to solve problems or complete calculations in a rigid timescale. There is always the danger that blind prediction challenges and critical assessments can cripple a field, if they drive methodological changes away from physically based methods towards empirical tweaks that lead to improved agreement between calculated and measured parameters, with an attendant loss of physical understanding. The pK_a Cooperative attempts to balance the positive aspects of forcing laboratories to evaluate their simulations and calculations critically and stringently, while biasing towards improved physical understanding of the underlying physics of the problem of interest. These results of this 1st Blind Prediction Challenge were very sobering but also highly stimulating.

Limitations of the Experimental Data Used in the 1st Blind Prediction Challenge

The internal ionizable groups in SNase were buried in mostly hydrophobic microenvironments in the interior of the protein. Even in the case of some of the introduced Lys residues (e.g. Lys-62, Lys-36, Lys-66) that are close to the backbone polar atoms of residues 19–22, which might be expected to stabilize the charged form of the Lys residues, these potential interactions were not identified by the simulations as being important. According to the calculations the set of SNase variants used in the challenge appear to provide more insight into desolvation energies than about the influence of hydrogen bonds and Coulomb interactions as determinants of pK_a values of internal groups –although this implied computational finding remains to be demonstrated experimentally. Although pairwise Coulomb interactions must be present, the loss of solvation energy was generally viewed as the primary perturbation of the pK_a of these internal residues in SNase. This segregation of solvation and pairwise Coulomb interactions in this collection of data can be seen as a strength of this set of pK_a values as it provides more direct insight into desolvation energies. On the other hand, it does not allow detailed exploration of the balance between desolvation and Coulomb forces. In proteins evolved to contain internal ionizable groups the calculations suggest that desolvation energies often appear to be compensated by Coulomb interactions.^{28,29} This balance is generally proposed anytime a buried acid or base has a pK_a that is not much shifted from the value in solution. However, this assertion cannot be cleanly established experimentally. The balancing of desolvation energies that destabilize and the Coulomb interactions that stabilize the charged state reduce shifts in pK_a values and can lead to a mitigation of errors. More experimental data are needed in which Coulomb effects experienced by internal ionizable groups can be examined rigorously. The ongoing engineering of variants of SNase with internal ionizable pairs or with ionizable groups in more polar cages will be useful to test the ability to simulate pair-wise charge-charge and charge-dipole interactions accurately.

New Challenges Ahead

In the 1st Blind Prediction Challenge the emphasis was strictly on the ability of computational methods to reproduce the pK_a values of internal ionizable groups in SNase. Future challenges for simulation of pK_a values of internal groups in proteins can become more stringent if attention is given to the following areas:

1. *Coupling between ionizable groups.* Faced with the demanding challenge represented by the large shifts in pK_a values measured for 100 internal groups by the García-Moreno Lab, the participants in the critical assessment exercise focused primarily on reproducing the pK_a values of the internal introduced groups, without paying attention to the pK_a of the surface residues, especially the His, Asp and Glu residues in SNase whose pK_a values have been reported previously.^{58–60} Owing to the long-range character of electrostatic forces, the ionization of different groups on the protein mutually influence each other. Coupling is most likely when internal ionizable groups are involved because their interactions with surface ionizable groups occur primarily through the protein, where depending on dielectric relaxation properties, electrostatic effects could be strong even between distant groups.⁶¹ The coupling between the pK_a values of internal groups and surface residues represents additional information to be simulated and analyzed. As experimental measurements of coupling between internal and surface residues become available, these will represent important constraints that pK_a calculations will have to satisfy. Thus, in the future it will be useful for the ionization state of both surface and internal ionizable groups of the protein to be reported for all calculations.
2. *Conformational Reorganization Coupled to the Ionization of Internal Groups.* The range of structural and dynamic responses to the ionization of internal groups in SNase or in other proteins is not known. Preliminary data suggest that the ionization of internal groups can trigger local conformational reorganization.^{62–67} This is consistent with the high dielectric constants required in continuum electrostatic methods to come close to reproducing the pK_a values of the internal groups. There is a critical need for improved experimental measurements of the range of conformational reorganization that can be promoted by the ionization of internal groups. Again, when this level of experimental detail of the effects of ionization on protein structure and dynamics becomes known it will constitute important constraints for calculations.

The calculations have to take into account the motions that accompany the protonation/deprotonation reactions. If conformational reorganization is found to be more prevalent than currently recognized, it may be that simulation techniques will have to depend more on use of Monte Carlo methods to generate the Boltzmann distribution of microstates as a function of pH to achieve convergence. Or it may be that advanced MD methods will be required to sample sufficient conformational space. These are areas where progress will be difficult and where experimental data are needed to guide further developments.

3. *Accuracy of structures.* A limited number of structures of SNase variants with internal ionizable groups were available when the first blind predictions were submitted to the pK_a Cooperative. Several dozen structures have now been released by the García-Moreno lab through the Protein Data Bank. The participants in the blind prediction challenge who simulated structures can now go back and examine the accuracy of their simulated structures to determine how the pK_a values calculated in real structures and in simulated structure compare. This represents an important first step to gauge the ability of various approaches to simulate the

relaxation that can occur as internal groups change their ionization state. Similarly, the simulations that depend on MD trajectories to sample different states of the protein can compare ensembles computed with MD simulations with ensembles obtained with NMR spectroscopy. The expected consequences of ionization of internal groups on structure and dynamics can thus be predicted and tested. In the future some structures can be released prior to publication to focus simulations on the analysis of pK_a s while other can be held back to challenge the community to better understand the coupling between the mutations and the structural changes.

4. *Water penetration:* Water penetration into the hydrophobic interior of the protein was observed in some of the earliest structures with internal Asp and Glu at position 66 in SNase.⁶⁵ Internal water molecules have been found in many of the variants of SNase.⁶⁵ The role of explicit water penetration was ignored in most of the calculations that were performed for the blind critical assessment. MD has been shown to be useful to reproduce the patterns of hydration that are observed in crystal structures obtained under cryogenic conditions.^{68–70} The ability of macroscopic methods to reproduce the presence of internal water molecules implicitly through the dielectric constant and the radii used to describe the water inaccessible surface, explored here by Knapp,⁴ will have to be tested extensively. The presence of internal water molecules is not an artifact of the SNase variants; many of the internal ionizable groups found naturally in proteins, for example in proteins involved in H^+ transport, are found in association with internal water molecules.⁷¹

Future Role for the pK_a Cooperative

The 1st Blind Challenge showed that the calculation of pK_a values of internal groups in proteins is extremely challenging. It remains to be established if the calculations that succeeded in reproducing the pK_a values of some of the internal groups do so for the right physical reasons. The Challenge showed that by increasing protein flexibility and tuning available parameters it was possible to improve the agreement between simulated and measured pK_a values of internal acidic and basic residues in SNase. This leaves us with a number of questions for future study: Will these empirical modifications be validated in future simulations? For methods with implicit solvent, are optimized parameters that improved agreement between calculated and measured pK_a values in SNase transferable to other proteins? Can the changes in structures coupled to the ionization of internal groups as predicted with MD simulations be validated with NMR spectroscopy experiments? Can comparison of measured and calculated data with small molecules be used to validate changes in parameters for charges and radii used in various methods? Can improvements in calculations achieved by comparison of the data with SNase achieve equal success with other proteins? Are calculations with small proteins transferable to calculations with large proteins, and are calculations with globular proteins transferable to calculations with membrane proteins?

The pK_a Cooperative hopes to play a role channeling and organizing efforts in the development of structure-based computational methods for calculation of electrostatic effects. If it is successful, the Cooperative will help the community avoid unnecessary repetition of effort, it will foment the development of truly novel approaches, it will help expose deficiencies in algorithms for structure-based calculations, and stimulate novel experiments to gain the physical insight needed to guide the development of new computational approaches. The Cooperative will enhance synergy and interactions between groups with expertise in experimental, computational or theoretical approaches, and it will facilitate the organization and maintenance of sets of data that should become the standard used to benchmark or to train a wide variety of computational methods in this field.

Acknowledgments

The pK_a Cooperative is a collaborative effort involving a large group of laboratories (<http://www.pkacoop.org>) many of which are contributors to this volume of *Proteins*. Credit for the success of the 1st Blind Prediction Challenge organized by the Cooperative goes to the laboratories that submitted blind predictions, to the ones that have contributed experimental data to the Cooperative, and to laboratories that have played a central role in administration of the Cooperative and handling of data submissions. We would like to thank Nana Nesbitt and the Telluride Science Research Center for providing a wonderful venue for the meetings of the Cooperative. MRG gratefully acknowledges NSF MCB-1022208 and the infrastructure support of NIH 5G12 RR03060. BGME gratefully acknowledges support from NSF MCB-0743422, NIH GM-061597 and NIH GM-073838. JN was supported by the following grants from Science Foundation Ireland: President of Ireland Young Researcher award; Grant number: 04/Y11/M537; Research Frontiers award; Grant number: 08/RFP/BIC1140

References

1. Arthur E, Yesselman J, Brooks C. Predicting Extreme pKa Shifts in Staphylococcal Nuclease Mutants with Constant pH Molecular Dynamics. *Proteins*. 2011 n/a-n/a.
2. Czodrowski P. Blind, one-eyed, or eagle-eyed? pKa calculations during blind predictions with staphylococcal nuclease. *Proteins*. 2011 n/a-n/a.
3. Gunner MR, Zhu X, Klein MC. MCCE analysis of the pKas of introduced buried acids and bases in staphylococcal nuclease. *Proteins*. 2011 n/a-n/a.
4. Meyer T, Kieseritzky G, Knapp E-W. Electrostatic pKa computations in proteins: Role of internal cavities. *Proteins*. 2011 n/a-n/a.
5. Olsson MHM. Protein electrostatics and pKa blind predictions contribution from empirical predictions of internal ionizable residues. *Proteins*. 2011 n/a-n/a.
6. Shan J, Mehler EL. Calculation of pKa in proteins with the microenvironment modulated-screened coulomb potential. *Proteins*. 2011 n/a-n/a.
7. Song Y. Exploring conformational changes coupled to ionization states using a hybrid Rosetta-MCCE protocol. *Proteins*. 2011 n/a-n/a.
8. Wallace JA, Wang Y, Shi C, Pastoor KJ, Nguyen B-L, Xia K, Shen JK. Toward accurate prediction of pKa values for internal protein residues: The importance of conformational relaxation and desolvation energy. *Proteins*. 2011 n/a-n/a.
9. Warwicker J. pKa predictions with a coupled finite difference Poisson–Boltzmann and Debye–Hückel method. *Proteins*. 2011 n/a-n/a.
10. Williams SL, Blachly PG, McCammon JA. Measuring the successes and deficiencies of constant pH molecular dynamics: A blind prediction study. *Proteins*. 2011 n/a-n/a.
11. Witham S, Talley K, Wang L, Zhang Z, Sarkar S, Gao D, Yang W, Alexov E. Developing hybrid approaches to predict pKa values of ionizable groups. *Proteins*. 2011
12. Word JM, Nicholls A. Application of the Gaussian dielectric boundary in Zap to the prediction of protein pKa values. *Proteins*. 2011 n/a-n/a.
13. Carstensen T, Farrell D, Huang Y, Baker NA, Nielsen JE. On the development of protein pKa calculation algorithms. *Proteins*. 2011 n/a-n/a.
14. Couch V, Stuchebrukhov A. Histidine in continuum electrostatics protonation state calculations. *Proteins*. 2011 n/a-n/a.
15. Itoh S, Damjanovic A, Brooks BR. pH Replica-Exchange Method based on discrete protonation states. *Proteins*. 2011
16. Machuqueiro M, Baptista AM. Is the prediction of pKa values by constant-pH molecular dynamics being hindered by inherited problems? *Proteins*. 2011 n/a-n/a.
17. Polydorides S, Amara N, Aubard C, Plateau P, Simonson T, Archontis G. Computational protein design with a generalized born solvent model: Application to asparaginyl-tRNA synthetase. *Proteins*. 2011 n/a-n/a.
18. Warshel A, Dryga A. Simulating electrostatic energies in proteins: Perspectives and some recent studies of pKas, redox, and other crucial functional properties. *Proteins*. 2011 n/a-n/a.

19. Cymes GD, Grosman C. Estimating the pK_a values of basic and acidic side chains in ion channels using electrophysiological recordings: A robust approach to an elusive problem. *Proteins*. 2011 n/a-n/a.
20. Ensign DL, Webb LJ. Factors determining electrostatic fields in molecular dynamics simulations of the ras/effector interface. *Proteins*. 2011 n/a-n/a.
21. Loladze VV, Makhatadze G. Analysis of Electrostatic Interactions in the Denatured State Ensemble of the N-terminal Domain of L9 Under Native Conditions. *Proteins*. 2011
22. Meng W, Raleigh D. Analysis of Electrostatic Interactions in the Denatured State Ensemble of the N-terminal Domain of L9 Under Native Conditions. *Proteins*. 2011 n/a-n/a.
23. Alexov E, Mehler E, Baker N, Baptista A, Huang Y, Milletti F, Nielsen J, Farrell D, Carstensen T, Olsson M, Shen J, Warwicker J, Williams S, Word M. Progress in the prediction of pK_a values in proteins. *Proteins*. 2011 n/a-n/a.
24. Decoursey TE. Voltage-gated proton channels and other proton transfer pathways. *Physiol Rev*. 2003; 83:475–579. [PubMed: 12663866]
25. Garcia-Moreno EB, Fitch CA. Structural interpretation of pH and salt-dependent processes in proteins with computational methods. *Methods Enzymol*. 2004; 380:20–51. [PubMed: 15051331]
26. Stanton CL, Houk KN. Benchmarking pK_a prediction methods for residues in proteins. *J Chem Theory Comput*. 2008; 4:951–966.
27. Song Y, Mao J, Gunner MR. MCCE2: Improving Protein pK_a Calculations with Extensive Side Chain Rotamer Sampling. *J Comp Chem*. 2009; 30:2231–2247. [PubMed: 19274707]
28. Spassov VZ, Ladenstein R, Karshikoff AD. Optimization of the electrostatic interactions between ionized groups and peptide dipoles in proteins. *Protein Sci*. 1997; 6:1190–1196. [PubMed: 9194179]
29. Kim J, Mao J, Gunner MR. Are acidic and basic groups in buried proteins predicted to be ionized? *J Mol Biol*. 2005; 348:1283–1298. [PubMed: 15854661]
30. Parsegian A. Energy of an ion crossing a low dielectric membrane: solutions to four relevant electrostatic problems. *Nature*. 1969; 221:844–846. [PubMed: 5765058]
31. Kassner RJ. Effects of nonpolar environments on the redox potentials of heme complexes. *Proc Natl Acad Sci USA*. 1972; 69:2263–2267. [PubMed: 4506096]
32. Warshel A, Russell ST. Calculations of electrostatic interactions in biological systems and in solutions. *Q Rev Biophys*. 1984; 17:283–422. [PubMed: 6098916]
33. Honig B, Nicholls A. Classical electrostatics in biology and chemistry. *Science*. 1995; 268:1144–1149. [PubMed: 7761829]
34. Bashford D. Macroscopic electrostatic models for protonation states in proteins. *Front Biosci*. 2004; 9:1082–1099. [PubMed: 14977531]
35. Baker NA. Improving implicit solvent simulations: a Poisson-centric view. *Curr Opin Struct Biol*. 2005; 15:137–143. [PubMed: 15837170]
36. Gunner MR, Mao J, Song Y, Kim J. Factors influencing energetics of electron and proton transfers in proteins. What can be learned from calculations. *Biochim Biophys Acta*. 2006; 1757:942–968. [PubMed: 16905113]
37. Chen J, Brooks CL 3rd, Khandogin J. Recent advances in implicit solvent-based methods for biomolecular simulations. *Curr Opin Struct Biol*. 2008; 18:140–148. [PubMed: 18304802]
38. Wallace JA, Shen JK. Predicting pK_a values with continuous constant pH molecular dynamics. *Methods Enzymol*. 2009; 466:455–475. [PubMed: 21609872]
39. Tanford C, Roxby R. Interpretation of protein titration curves. Application to lysozyme. *Biochemistry*. 1972; 11:2192. [PubMed: 5027621]
40. Matthew JB, Gurd FR, Garcia-Moreno B, Flanagan MA, March KL, Shire SJ. pH-dependent processes in proteins. *CRC Crit Rev Biochem*. 1985; 18:91–197. [PubMed: 3899508]
41. Forsyth WR, Antosiewicz JM, Robertson AD. Empirical relationships between protein structure and carboxyl pK_a values in proteins. *Proteins: Struct Funct Genet*. 2002; 48:388–403. [PubMed: 12112705]
42. Edgcomb SP, Murphy KP. Variability in the pK_a of histidine side-chains correlates with burial within proteins. *Proteins: Struct Funct Genet*. 2002; 49:1–6. [PubMed: 12211010]

43. Davies MN, Toseland CP, Moss DS, Flower DR. Benchmarking pK_a prediction. *BMC Biochem.* 2006; 7:1–12. [PubMed: 16401342]
44. Grimsley GR, Scholtz JM, Pace CN. A summary of the measured pK values of the ionizable groups in folded proteins. *Protein Sci.* 2009; 18:247–251. [PubMed: 19177368]
45. Webb H, Tynan-Connolly BM, Lee GM, Farrell D, O'Meara F, Sondergaard CR, Teilum K, Hewage C, McIntosh LP, Nielsen JE. Remeasuring HEWL pK_a values by NMR spectroscopy: Methods, analysis, accuracy, and implications for theoretical pK_a calculations. *Proteins.* 2011; 79:685–702. [PubMed: 21287606]
46. Antosiewicz J, McCammon JA, Gilson MK. The determinants of pK_as in proteins. *Biochemistry.* 1996; 35:7819–7833. [PubMed: 8672483]
47. Isom DG, Cannon BR, Castaneda CA, Robinson A, Garcia-Moreno B. High tolerance for ionizable residues in the hydrophobic interior of proteins. *Proc Natl Acad Sci U S A.* 2008; 105:17784–17788. [PubMed: 19004768]
48. Isom DG, Castaneda CA, Cannon BR, Velu PD, Garcia-Moreno EB. Charges in the hydrophobic interior of proteins. *Proc Natl Acad Sci U S A.* 2010; 107:16096–16100. [PubMed: 20798341]
49. Isom DG, Castaneda CA, Cannon BR, Garcia-Moreno B. Large shifts in pK_a values of lysine residues buried inside a protein. *Proc Natl Acad Sci U S A.* 2011; 108:5260–5265. [PubMed: 21389271]
50. Harms MJ, Schlessman JL, Sue GRBE, B.E. G-M. Arginine residues at internal positions in a protein are always charged. *Proc Natl Acad Sci U S A.* 2011 Submitted.
51. Balashov SP. Protonation reactions and their coupling in bacteriorhodopsin. *Biochim Biophys Acta.* 2000; 1460:75–94. [PubMed: 10984592]
52. Brzezinski P. Redox-driven membrane-bound proton pumps. *Trends Biochem Sci.* 2004; 29:380–387. [PubMed: 15236746]
53. Beroza P, Fredkin DR, Okamura MY, Feher G. Protonation of interacting residues in a protein by a Monte Carlo method: application to Lysozyme and the photosynthetic reaction center of *Rhodobacter sphaeroides*. *Proc Natl Acad Sci USA.* 1991; 88:5804–5808. [PubMed: 2062860]
54. Ghosh N, Cui Q. pK_a of residue 66 in Staphylococcal nuclease. I. Insights from QM/MM simulations with conventional sampling. *J Phys Chem B.* 2008; 112:8387–8397. [PubMed: 18540669]
55. Kato M, Warshel A. Using a charging coordinate in studies of ionization induced partial unfolding. *J Phys Chem B.* 2006; 110:11566–11570. [PubMed: 16771433]
56. Nielsen JE, McCammon JA. On the evaluation and optimization of protein X-ray structures for pK_a calculations. *Prot Sci.* 2003; 12:313–326.
57. Nicholls A. What do we know?: simple statistical techniques that help. *Methods Mol Biol.* 2011; 672:531–581. [PubMed: 20838984]
58. Lee KK, Fitch CA, Lecomte JT, Garcia-Moreno EB. Electrostatic effects in highly charged proteins: salt sensitivity of pK_a values of histidines in staphylococcal nuclease. *Biochemistry.* 2002; 41:5656–5667. [PubMed: 11969427]
59. Baran KL, Chimenti MS, Schlessman JL, Fitch CA, Herbst KJ, Garcia-Moreno BE. Electrostatic effects in a network of polar and ionizable groups in staphylococcal nuclease. *J Mol Biol.* 2008; 379:1045–1062. [PubMed: 18499123]
60. Castaneda CA, Fitch CA, Majumdar A, Khangulov V, Schlessman JL, Garcia-Moreno BE. Molecular determinants of the pK_a values of Asp and Glu residues in staphylococcal nuclease. *Proteins.* 2009; 77:570–588. [PubMed: 19533744]
61. Pey AL, Rodriguez-Larrea D, Gavira JA, Garcia-Moreno B, Sanchez-Ruiz JM. Modulation of buried ionizable groups in proteins with engineered surface charge. *J Am Chem Soc.* 2010; 132:1218–1219. [PubMed: 20055447]
62. Damjanovic A, Schlessman JL, Fitch CA, Garcia AE, Garcia-Moreno EB. Role of flexibility and polarity as determinants of the hydration of internal cavities and pockets in proteins. *Biophys J.* 2007; 93:2791–2804. [PubMed: 17604315]
63. Damjanovic A, Wu X, Garcia-Moreno EB, Brooks BR. Backbone relaxation coupled to the ionization of internal groups in proteins: a self-guided Langevin dynamics study. *Biophys J.* 2008; 95:4091–4101. [PubMed: 18641078]

64. Karp DA, Stahley MR, Garcia-Moreno B. Conformational consequences of ionization of Lys, Asp, and Glu buried at position 66 in staphylococcal nuclease. *Biochemistry*. 2010; 49:4138–4146. [PubMed: 20329780]
65. Chimenti MS, Khangulov VS, Robindon AC, Heroux A, Majumdar A, Schlessman JL, B.E. García-Moreno. Structural reorganization triggered by the introduction of charge in the hydrophobic interior of a protein: survey of 25 internal Lys residues. *Structure*. 2011 Submitted.
66. Damjanovic A, Brooks BR, Garcia-Moreno B. Conformational relaxation and water penetration coupled to ionization of internal groups in proteins. *J Phys Chem A*. 2011; 115:4042–4053. [PubMed: 21428436]
67. Dwyer JJ, Gittis AG, Karp DA, Lattman EE, Spencer DS, Stites WE, Garcia-Moreno EB. High apparent dielectric constants in the interior of a protein reflect water penetration. *Biophys J*. 2000; 79:1610–1620. [PubMed: 10969021]
68. Damjanovic A, Garcia-Moreno B, Lattman EE, Garcia AE. Molecular dynamics study of water penetration in staphylococcal nuclease. *Proteins*. 2005; 60:433–449. [PubMed: 15971206]
69. Schlessman JL, Abe C, Gittis A, Karp DA, Dolan MA, Garcia-Moreno EB. Crystallographic study of hydration of an internal cavity in engineered proteins with buried polar or ionizable groups. *Biophys J*. 2008; 94:3208–3216. [PubMed: 18178652]
70. Damjanovic A, Miller BT, Wenaus TJ, Maksimovic P, B. E G-M, Brooks BR. Study of the Coupling between Conformation and Water Content in the Interior of a Protein. *J Chem Inf Model*. 2008; 48:2021–2029. [PubMed: 18834189]
71. Luecke H, Schobert B, Richter HT, Cartailler JP, Lanyi JK. Structure of bacteriorhodopsin at 1.55 Å resolution. *J Mol Biol*. 1999; 291:899–911. [PubMed: 10452895]