Maria V. Piaggio[1]
Marta B. Peirotti[2]
Julio A. Deiber[2]

[1]Cátedra de Bioquímica Básica de Macromoléculas, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral (UNL), Santa Fe, Argentina
█Check █
[2]Instituto de Desarrollo Tecnológico para la Industria Química (INTEC), UNL, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Santa Fe, Argentina

Effect of background electrolyte on the estimation of protein hydrodynamic radius and net charge through capillary zone electrophoresis

Two physicochemical models are proposed for the estimation of both hydrodynamic radius and net charge of a protein when the capillary zone electrophoretic mobility at a given protocol, the set of p*K* of charged amino acids, and basic data from Protein Data Bank are available. These models also provide a rationale to interpret appropriately the effects of solvent properties on protein hydrodynamic radius and net charge. To illustrate the numerical predictions of these models, experimental data of electrophoretic mobility available in the literature for well-defined protocols are used. Five proteins are considered: lysozyme, staphylococcal nuclease, human carbonic anhydrase, bovine carbonic anhydrase, and human serum albumin. Numerical predictions of protein net charges through these models compare well with the results reported in the literature, including those found asymptotically through protein charge ladder techniques. Model calculations indicate that the hydrodynamic radius is sensitive to changes of the protein net charge and hence it cannot be assumed constant in general. Also, several limitations associated with models for estimating protein net charge and hydrodynamic radius from protein structure, amino acid sequence, and experimental electrophoretic mobility are provided and discussed. These conclusions also show clear requirements for further research.

Keywords: Electrophoretic mobility / Hydrodynamic radius / Protein net charge / Solvent–protein interactions solvent properties DOI 10.1002/elps.200500161

1 Introduction

In the last 15 years, capillary zone electrophoresis (CZE) has been intensively applied to the separation and characterization of many small and large electrically charged molecules. At present, among the different kinematical units suspended in the solvent, peptides and proteins are the ones most frequently found in the literature, mainly due to the challenges offered by the proteomic project. In this context, an important study becomes relevant, namely, the one specifically associated with the appropriate formulation of the background electrolyte (BGE) for a given CZE run (see, *e.g.*, [1–4]). One of the basic concerns in this sense is how pH, ionic strength *I*, temperature *T*, solvent viscosity η, and electrical permittivity ε may be controlled as relevant properties of the BGE to get proteins at an appropriate state of electrical charge, which in turn may be defined for different purposes like: (A) Minimize the effect of electrostatic interaction of charged groups of residual amino acids in the protein, to avoid significant pK_i shifts (i = 1, 2, . . ., N indicates different charged groups at different positions in the protein), thus leading to a more predictable electropherogram and rational protocol formulation from a theoretical point of view. The evaluation and prediction of have ΔpK_i gained much attention [5–14] in order to understand better protein roles in biological systems. (B) Analyze available electrokinetic and electrostatic theories, providing solutions of analyte mobility with direct application to the interpretation of protein CZE results, expressed through electropherogram outputs. These specific graphs, being rich in physicochemical information of the analytes separated, are also very sensitive to the formulation of the running solvent, *i.e.*, the BGE. Nevertheless, this information must be extracted from the electropherogram carefully because the mobility μ of each analyte separated is proportional to the ratio between the net electrical charge and the hydrodynamic friction coefficient *f*. Here, *e* is the unit charge and *Z*s the protein net charge number. The values of numerator and

Correspondence: Dr. Julio A. Deiber, Instituto de Desarrollo Tecnológico para la Industria Química (INTEC), Güemes 3450, S3000GLN Santa Fe, Argentina
E-mail: treoflu@ceride.gov.ar
Fax: +54-342-455-0944

Abbreviations: AAS, amino acid sequence; **LLCEM**, Linderstrøm-Lang capillary electrophoresis model; **LLM**, Linderstrøm-Lang model; **LRCLM**, linear regression charge ladder model; **PBMCM**, Poisson–Boltzmann–Monte Carlo model; **PDB**, Protein Data Bank; **PLLCEM**, perturbed Linderstrøm-Lang capillary electrophoresis model

denominator in this ratio are difficult to separate; at least complementary experiments are carried out apart from CZE. An alternative is to introduce further theoretical developments to model these terms. Thus, it is clear that a hydrodynamic radius a_H is necessary to define properly the friction coefficient in terms of the most elementary form, *viz.*, a spherical particle, when continuum hypothesis is invoked. Other particle forms may be used with the need of additional protein structural information and this approach has been found useful [15–17].

Out of the CZE framework, the estimation of the net charge of proteins in a given solvent has been studied by using protein structural details from both X-ray crystallographic and NMR data. At present, computational models and solution algorithms are available (see, *e.g.*, [6–9], and citations therein), which are based on basic principles of electrostatics, electrokinetics, and statistical thermodynamics. Throughout this work, we refer to these models as "detailed models" in general. Valuable results have been achieved in this rigorous track, with very costly and time-consuming computational runs. Nevertheless, these models are still difficult to be applied in the most practical situations, like in CE. For instance, including side chain flexibility for the electrostatic interactions in the calculations of protein titration models through Poisson–Boltzmann equation and Monte Carlo sampling procedure [7] (designated here as PBMCM (Poisson–Boltzmann–Monte Carlo model)) requires averages of charge group protonations over the accessible states of the protein. These states are of the order of $(2N_c)^N$, where N is the number of titration sites and N_c is the number of alternative conformations, including the basic one pertaining to the X-ray crystallographic conformation. These computations require the application of finite difference numerical techniques, with different degrees of grid refinements, and frequently the resulting problem leads to situations impractical computationally, when the lowest thermodynamic free energy of the macromolecule must be found to achieve the correct result (see also [7, 18], and citations therein). On the other hand, these models have also been providing useful results thus allowing the formulation of simpler ones. The predictions of the PBMCM are used to compare results provided by simple models that keep a direct relationship with CZE experimental data [19].

Simple models are useful by presenting a valuable physical insight of the difficult calculations carried out through detailed models when the net charge of proteins in a given solvent is required. Therefore, from the analysis above, it is clear that the search of simple models, although empirical in some aspects, also becomes necessary as expressed more recently in the literature (see, *e.g.*, [15, 17, 19–29]).

More specifically, for the estimation *via* CZE of eZ and a_H of a given protein in a BGE with defined values of pH, I, T, η, and ε, one has to fix the level of detail of the macromolecular structure on which a mathematical model is constructed. In addition, this model may also have different purposes as described by Reijenga and Kenndler [30]. In general, the requirements of the analyte structure and model of course depend on the type of scientific problem and application. For instance, structural details needed for studying the protein folding problem, or the evaluation of ΔpK_i of charged groups in the protein, may be different from those required in the formulation of an efficient CZE separation of a mixture of these analytes. In this respect, we present here an analysis of CZE models for charged proteins, where the basic data used concerning analyte macromolecular structure are the amino acid sequence (AAS) and the values of the geometric distances r_{ij} between pairs of charged groups, designated i and j among the N possible pairs in a protein. These two types of information are readily available in the Protein Data Bank (PDB). Here, different atom conformations from those reported in the PDB are not considered because the fine protein structure cannot be treated without entering into more complex calculations through detailed models.

Therefore, in this work we propose and discuss two simple physicochemical models to estimate each terms of the ratio eZ/f of proteins, when the experimental value of their electrophoretic mobilities are available for a given CZE protocol, and also compare results with the data reported in the literature. The analysis and evaluation of the charge regulation phenomenon, also designated proton-binding cooperativity [19, 23, 31], must be included and hence modeled with some details. We expect that the results obtained may be of utility for those carrying out frequent CZE runs in research and development laboratories mainly to optimize and propose the appropriate protocols of protein separations. This target has been considered from different points of view, as one may infer through more recent published works [16, 20], which present practical advantages, calculation difficulties, as well as estimations of the protein hydrodynamic radius, usually assumed a constant. Therefore, these aspects are analyzed and discussed below.

It is worth observing that one of the proposal found in the literature to simply estimate eZ and a_H of a protein consists in using protein charge ladders with CZE (see, *e.g.*, [17, 19, 20–29]). This method introduces changes in the protein by converting charged groups into electrically neutral ones, usually by modification of the Lys ε-NH_3^+ and/or carboxyl groups. Therefore, in CZE these modified proteins separate into different peaks or rungs, which may then be used for further calculations like eZ and a

constant value of a_H for the first peaks of the protein charge ladder. It is observed that this method presents practical disadvantages such as the need to modify chemically the protein and the estimation of a_H from the slope of the line defined by the electrophoretic mobility of the first converted fractions *versus* $n\Delta Z$. In this expression, n is the number of charged groups converted to neutral ones, independent of positions in the protein, and ΔZ is the charge change from two consecutive numbers of conversions. Throughout our work, this model is designated as LRCLM (linear regression charge ladder model) based on the use of linear regression analysis of mobilities of the protein charge ladder. In this model the approximation $|\Delta Z| \cong 1$ is usually introduced [20]. In this sense, Sharma *et al.* [19] and Allison *et al.* [17] suggested an interesting correction based on a previous work [12] to avoid this assumption, because in most of the cases studied, this change is not necessarily unity.

A classical model used to estimate protein net charge is the Linderstrøm-Lang model (LLM), presented and discussed elsewhere [19, 23]. When the underlying hypothesis is accepted, the LLM has the limitations that a_H is unknown. One alternative is to use a value for the hydrodynamic radius from the LRCLM, assuming a constant for the first steps of the protein charge ladder [19].

By considering the practical aspects of the models mentioned above, a convenient procedure is to solve the complete mathematical problem associated with the basic equations proposed in these works, having as experimental data only the value of the electrophoretic mobility of the native protein, without the need of carrying out, for instance, CZE of the protein charge ladder. This model is designated here as LLCEM (Linderstrøm-Lang CE model) based on the use of LLM and only one experimental data of CZE. In general, one observes that changes of eZ affects a_H, even to the precision required and discussed in previous works. This aspect is analyzed and illustrated below.

Another proposal to solve the central problem studied in this work is to estimate a_H from the determination of the sedimentation coefficient or the diffusion coefficient [15]. This method needs the evaluation of the function $a_H(\mathrm{pH}, I, T, \eta, \varepsilon)$, considering the values that the independent variables involved may take. From the framework proposing the modeling of the ratio eZ/f, at present it is clear that classical methods (see a relevant revision of these methods in [16]) provide the model-independent protein net charge without invoking the AAS in general.

Most of the simple models evaluating eZ and eventually a_H *via* CZE have as a reference point the LLM, which in turn can give an approximation of eZ only. We show that the LLCEM proposed here provides both eZ and a_H with unique experimental data μ as long as the mathematical problem is solved completely: this means without an estimation of a constant a_H. In fact, the LLM carries out a correction of the wild protein charge eZ^o calculated directly from the knowledge of both the pK_i^o values (i = 1, 2, . . ., N) and the number n_i of each type of the N charged groups in the protein at the pH near particle. This will be another interesting conclusion in our work in the sense that to our best knowledge the LLM has not been completely solved (it is meant in terms of the LLCEM) in most of the circumstances that it was applied. This aspect is further analyzed below.

Therefore, the present work is organized as follows. In Section 2 we present the LLCEM proposed and solved numerically in this work, which asymptotically reduces to the LLM when a_H is assumed constant and estimated through other sources. Also in Section 2, the LLCEM is reconsidered to include electrostatic interaction between pairs of charged groups through charge perturbations defined as the difference between wild and regulated charge of each charged group. This model is designated as perturbed Linderstrøm-Lang CE model (PLLCEM) due to the charge perturbations introduced. For the formulation of the PLLCEM several assumptions of previous works are considered [12, 17, 19]. Emphasis is placed on the mathematical strategy to solve the resulting well-posed problems associated with both LLCEM and PLLCEM, by identifying an equal number of equations and unknowns. Also in Section 2 the basic steps of the numerical algorithm are described including convergence criteria placed within the iterative process. Section 3 presents numerical results for typical proteins studied in the literature, thus facilitating comparisons of our calculations with other model predictions. Finally, several limitations associated with the models for estimating protein net charge and hydrodynamic radius from protein structure, AAS, and experimental electrophoretic mobility are discussed and provided in Section 4. These conclusions also set clear requirements for further research of this subject.

2 Materials and methods

2.1 Modeling the effective charge and hydrodynamic radius of proteins through CZE█Check █

The models studied here consider globular proteins as having at least a tertiary structure, which is, in principle, assumed equivalent to a sphere of hydrodynamic radius a_H with $f = 6\pi\eta a_H$. Although this hypothesis may be readily eliminated to include other particle forms [15, 17],

we postpone this discussion in favor of a clear and simple presentation of the models. In this section, PLLCEM is described first because this model reduces to LLCEM when electrostatic interactions between pairs of charged groups are not considered.

2.2 PLLCEM

The protein has a net charge eZ for a given CZE protocol, expressed as

$$eZ = \sum_{i=1}^{N} eZ_i \tag{1}$$

where eZ is the charge of the i-charged group in the protein. The value of Z_i is calculated through

$$Z_i = \pm \frac{1}{1 + 10^{\mp(pK_i^o - pH_i)}} \tag{2}$$

where signs (+) and (−) are defined according to basic or acidic properties of charged groups. Also pH_i is the pH estimated near the i-charged group, generating this description for one possible mathematical approach describing the proton-binding cooperativity phenomenon (also see below for the definition of ΔpK_i). Equations 1 and 2 are quite difficult to solve because the physicochemical details around charged groups within the protein domain require quantitations. In this sense, it is appropriate to visualize that detailed models describe the proton-binding cooperativity phenomenon through free energy changes associated with (a) the solvation of charged groups, (b) the electrostatic interaction between protein charged groups and solvent ions (Poisson–Boltzmann equation for an atomic radius of around 1.4 Å is typically used, where an electrical permittivity different from that of the BGE is estimated), and (c) the electrostatic interaction of charged groups with both other charged groups and dipoles present in the protein. On the other hand, in simple models, one introduces the hydrodynamic radius a_H to evaluate the mean field free energy change associated with (b), while those pertaining to (a) and (c) are treated as perturbations of charged groups defined in Eq. 5. These free energy changes are associated with electrical potentials, the sum of which determines the Boltzmann distribution of proton near the generic i-charged group. Thus the pH_i can be expressed through three terms when dipole effects are neglected, as follows (also see the Appendix):

$$pH_i = pH^* + \frac{e^2}{\ln(10)k_BT}\left(\frac{\Delta Z_i}{4\pi r_i^o \varepsilon}\left(\frac{\varepsilon}{\varepsilon'} - 1\right) + \sum_{\substack{j=1 \\ j\neq i}}^{N} \frac{\Delta Z_j}{4\pi\varepsilon'} \frac{\exp(-\kappa' \cdot r_{ij})}{r_{ij}}\right) \tag{3}$$

where k_B is the Boltzmann constant and ε' is the electrical permittivity within the domain of the protein (this aspect is further explained below). Since the inverse of the screening length [32] is $\kappa = \sqrt{2e^2 I N_A 10^3/\varepsilon k_B T}$, where N_A is the Avogadro number, one also gets $\kappa' = \kappa\sqrt{\varepsilon'/\varepsilon}$. In Eq. 3, is the pH^* near the protein (assumed as a particle) and can be evaluated by using the mean field approximation introduced together with the inclusion of the protein hydrodynamic radius. Thus, the protein particle is assumed to have a surface potential provided by the Poisson–Boltzmann equation for a solvent having properties pH, I, T, η, and ε specified. In order to generate a simple problem, Henry solution of the particle electrophoretic mobility is used where the Debye–Hückel approximation is invoked. Nevertheless, this approximation does not necessarily limit the basic framework of the model as discussed below. Consequently, one can readily express (also see the Appendix)

$$pH^* = pH + \frac{e^2 Z}{\ln(10)k_BT4\pi\varepsilon a_H(1 + \kappa a_H)} \tag{4}$$

Equation 4 reflects the electrostatic and hydrodynamic interactions between protein and BGE, and inevitably the introduction of a_H generates an assumption not found in detailed models. Thus, here we are dealing with the well-known Henry particle and its associated hypothesis (see, *e.g.*, [32]).

The second term on the right-hand side of Eq. 3 is associated with the free energy perturbation when charged groups of amino acids in solution (defining the protein wild charge) are transferred to a cavity in the protein [5]; here atom radius r_i^o is considered through Born equation [33]. The third term involves the effect on the protonation state of i-charged group caused by the electrostatic interactions with perturbed charges ΔZ_j of the other charged groups j = 1, 2, . . ., (N-1) in the protein. To evaluate this term, data of the distances r_{ij} between pairs of charged atoms are obtained from PDB. Thus, from this source one obtains the coordinates (generic position vector r_i) of each charged atom referred to a coordinate system. The distances between pairs of charged atoms may be readily calculated as $r_{ij} = \sqrt{(r_i - r_j)\cdot(r_i - r_j)}$. These distances belong to native proteins obtained from X-ray crystallography. When ions are binding the protein, one also needs to know which are the charged groups coordinated in order to estimate additional distances involving these ions (see examples given below concerning Zn^{2+} in 1CA2 and 1V9E). For the purpose of the present model, point charges are only considered. Therefore, the expression

$$\Delta Z_j = Z_j - Z_j^o \tag{5}$$

is required in Eq. 3, where Z_j^o is the wild charge of site *j*. Thus, the electrostatic interaction between pairs of charged groups appears in the model as a perturbation of the mean field solution associated with a particle of radius a_H.

The other equation relevant in the model is the expression of the electrophoretic mobility μ of the protein in the BGE, obtained experimentally and expressed (for high particle potentials, Poisson–Boltzmann equation with boundary conditions is required) as follows:

$$\mu = \frac{eZ}{6\pi\eta a_H(1 + \kappa a_H)} f(\kappa a_H) \tag{6}$$

where $f(\kappa a_H)$ is Henry function expressed here through an explicit fitting equation [34]. Equations 1–6 compose the PLLCEM of this work (Eqs. 2, 3 and 5 are generic for any *i* and they provide 3*N* equations), where the set $\{a_H, Z, Z_i, pH^*, pH_i, \Delta Z_i\}$ of $3(N + 1)$ unknowns may be obtained through a numerical algorithm (Section 2.4), once the experimental value of μ is provided under a well-defined CZE protocol. It is precisely in this aspect that the choice of pH, *I*, *T*, η, and ε becomes relevant in the CZE characterization of proteins.

In regards to the PLLCEM, one observes that there is no need to estimate a_H from any additional source. In fact, the protein hydrodynamic radius is a direct solution of the model. This conclusion introduces a difference from the previous works where the modeling has been carried out with an estimation of a constant a_H. This parameter, of course, is inevitable when data of CZE must be considered.

After the model is solved for a given protein in a specific solvent, it is of interest to estimate the ΔpK_i values, having as reference state the pK_i^o values of the charged amino acids alone in solution, which defines the wild charge eZ^o of a protein. Therefore,

$$\Delta pK_i = pK_i^o - pK_i = \frac{e\zeta_i}{\ln(10)k_B T} \tag{7}$$

where ζ is the electrical potential of the *i*-charged group, expressed as

$$\zeta_i = \frac{eZ}{4\pi\varepsilon a_H(1 + \kappa a_H)} + \frac{e\Delta Z_i}{4\pi\varepsilon r_i^o}\left(\frac{\varepsilon}{\varepsilon'} - 1\right) + \sum_{\substack{j=1 \\ j\neq i}}^{N} \frac{e\Delta Z_j}{4\pi\varepsilon'} \frac{\exp(-\kappa' r_{ij})}{r_{ij}} \tag{8}$$

consistently with Eqs. 3 and 4 (also see the Appendix).

From Eqs. 3 and 7 one shows that $pK_i^o - pH_i \equiv pK_i - pH$. In addition, $\Delta pK_i = pK_i^o - pK_i = pH_i - pH = \Delta pH_i$. This last expression indicates that the ΔpK_i shifts of the protein charged groups are a direct consequence of ΔpH_i shifts, which is a difference between the pH near the charged group in the protein and the pH of the solvent, *i.e.*, the BGE.

An aspect to point out here before describing the LLCEM is the inclusion of particle shape in these models. This effect may be accounted for simply by introducing additional structural information of the protein [16, 17]. Thus, the particle shape must be defined *a priori* by specifying, for instance, the ratio of major to minor axes for prolate or oblate spheroids [35]. Then, one dimension is expressed in terms of the other, and the basic form of Henry solution is readily retained. Also, a numerical simulation by using the boundary element method may be carried out [17], which is a significantly more complex procedure placing emphasis mainly on the hydrodynamic friction exerted by the solvent on irregularly shaped particles. This hydrodynamic aspect is becoming a relevant way to explore more detailed protein shapes consistently with their hydration state.

2.3 LLCEM

The LLCEM is asymptotically obtained by neglecting the terms containing $\Delta Z_j = Z_j - Z_j^o$ associated with the perturbation of charged groups in the PLLCEM. Therefore, Eq. 2 reduces to

$$Z_i = \pm \frac{n_i}{1 + 10^{\mp(pK_i^o - pH^*)}} \tag{9}$$

Equations 1, 4, 6 and 9 compose the LLCEM of this work (Eq. 9 is generic for any *i* and provides *N* equations) where the set $\{a_H, Z, Z_i, pH^*\}$ of $(N + 3)$ unknowns may be obtained through an appropriate numerical algorithm (Section 2.4), once the experimental value of μ is provided under a well-defined CZE protocol.

From Eq. 9 it is clear that the LLCEM cannot distinguish positions of charged groups in the protein. In fact, from Eqs. 3 and 7, one readily finds that $pK_i^o - pH^* = pK_i - pH$ and $\Delta pK_i = pK_i^o - pK_i = pH^* - pH = \Delta pH$, indicating that the LLCEM predicts the same ΔpK_i for a given *i*-charged group, independent of the position of this group in the protein AAS, because in this model ΔpH is the same for all charged groups. Despite this limitation, it is worth observing that electrostatic interaction effects are usually small for protein with a low net charge. Even more, one can use a solvent pH in the CZE run that minimizes the effect of ΔpK_i on the values of both and a_H.

The relevant aspect concerning LLCEM is that the values eZ and a_H of a given protein in a well-defined BGE can be obtained by measuring only the protein

mobility μ. The results thus generated are a good approximation to the expected values, as it is illustrated in Section 3.

2.4 Numerical procedure

We describe the algorithm of PLLCEM only, because that pertaining to LLCEM is similar and simpler. Thus, Eqs. 1–6 are solved through a numerical procedure written in FORTRAN**█Provide supplier█** that is composed by the following principle steps. (1) Input of physical constants, protocol data involving solvent properties pH, I, T, η, and ε, reference values of pK_i^o (also see below for a discussion concerning this aspect) and an estimation of the electrical permittivity of the protein–solvent domain ε' (a value between $5\varepsilon_o$ and $80\varepsilon_o$, where ε_o is electrical permittivity of free space; also see below for a discussion on this aspect of the lysozyme protein). (2) Input of protein AAS from the PDB where the alphanumeric information is converted into two digits: one indicating the type of amino acid and the other referring to the position in the AAS for their use in the program logic. (3) Input of structural properties from the PDB involving the set $\{r_{ij}\}$ of X-ray crystallographic distance between pairs of the N-charged sites, taking as loci the corresponding charged atom. (4) Input of the protein electrophoretic mobility for the specific BGE. (5) With these inputs the algorithm searches the numbers of each type of charged group, and hence an estimation of the wild protein charge $eZ^o = \sum_{i=1}^{N} eZ_i^o$ is obtained at the solvent pH with the set $\{pK_i^o\}$. Of course, this result is quite different from that found experimentally. (6) An estimation of the hydrodynamic radius designated as a_H^o is introduced to initialize the program. This value is obtained from the average distance to the protein center of the outer amino acids located in a shell width of around 7 Å (around a length of two peptide bonds). (7) Iteration starts with initializations $pH_i = pH$ and $Z_i = Z_i^o$ and the evaluation of Eqs. 1–5. Then with the value of eZ, at each iteration, Eq. 6 is solved to find root a_H through Newton method. This procedure requires satisfying the experimental value μ within a relative error of 10^{-6}. After solving for a_H, at each iteration k, an additional convergence criterion is imposed; thus the following equation must be satisfied to stop the iteration process:

$$\frac{\left|a_H^k - a_H^{k-1}\right|}{a_H^k} \leq 10^{-6} \tag{10}$$

(8) Once the value of a_H is obtained, which is a rigorous solution of the mathematical problem placed by the model, one can proceed with other calculations by using Eqs. 7, 8 and basic outputs. Throughout this work, numerical results are reported with round-off error of 1/100 and the unit Å is used for the hydrodynamic radius as the most indicative for the discussion of results. Before ending this section, it is relevant to indicate that the evaluation of a_H, at each iteration in the above numerical algorithm, requires a very low relative error due to the shape of $f(\kappa a_H)$ when κa_H takes intermediate values. This situation is typical for proteins used in these studies.

Finally, in this context of analysis, the limits of validity when applying the Debye–Hückel approximation and Henry solution in the above models are explored with the help of numerical solutions of Poisson–Boltzmann problem carried out by O'Brien and White [36]. Thus, one may evaluate ordinate $Y = \frac{3}{2}\frac{e\mu\eta}{\varepsilon k_B T}$, abscise $X = \frac{e\zeta}{k_B T}$, and parameter $P = \kappa a_H$ with ζ and a_H obtained from PLLCEM and LLCEM, the experimental data μ, and those values pertaining to the CZE protocol, in order to know if the calculated point is close to the corresponding curve of the mobility generalized plots reported elsewhere (see also [32]). It is simple to conclude in general that for $Y \approx X$ and both Y and X less than around 2, the model results are compatible with the hypothesis under consideration for $P < 3$. Nevertheless, it is interesting to point out that the framework of these models has no limitations in this sense. In fact, although the closure of the mathematical problem is easily visualized when classical hypotheses are invoked, out of the range of their validity, the Poisson–Boltzmann equation must be included with appropriate boundary conditions. Thus, the resulting models would be more time consuming from the computational point of view but still very feasible to be solved.

3 Results and discussion

To illustrate the numerical predictions of LLCEM and PLLCEM, experimental data of the electrophoretic mobility available in the literature for well-defined protocols are used. These data involve four proteins: lysozyme (2LYZ), staphylococcal nuclease (1STN), human carbonic anhydrase (1CA2), and bovine carbonic anhydrase (1V9E).

First 2LYZ was studied with LLCEM by using the experimental data $\mu = 1.67 \times 10^{-8}\ m^2/s \times V$ extracted from the electropherogram reported in [19] where the CZE protocol was: pH = 8.4, I = 8 mM, and T = 25°C in aqueous solvent. The set of pK_i^o used is $\{pK^o_{NH_2} = 7.5;\ pK^o_{Arg} = 12.5;\ pK^o_{Asp} = 4.0;\ pK^o_{Glu} = 4.4;\ pK^o_{His} = 6.3;\ pK^o_{Lys} = 10.5;\ pK^o_{Tyr} = 9.6;\ pK^o_{COOH} = 3.8\}$ [19]. Also the number of charged groups is $\{n_{Arg} = 11;\ n_{Asp} = 7;\ n_{Glu} = 2;$

$n_{His} = 1$; $n_{Lys} = 6$; $n_{Tyr} = 3$}. The numerical results obtained were a_H = 22.22 Å and Z = 6.34. This formally shows that the solution of the LLCEM provides both the protein electrical charge and the hydrodynamic radius as numerical outputs. Also, it was found that ζ = 32 mV and pH* = 8.93, indicating that [H^+] is lower around the protein as expected.

The PLLCEM was also used to study 2LYZ by considering parameter ε' and the electrostatic interaction between pairs of charged groups. The AAS and the set of distances {r_{ij}} were obtained from the PDB. Since the value of ε' depends on the protein hydration and compaction in relation to the formulation of the BGE, its direct prediction is quite difficult. The choice at present is to use an estimated value as indicated above. Throughout this work, the value $\varepsilon' = 20\varepsilon_o$ was considered in most of the calculations as an approximation, based on the capability of the model to yield results comparable to those reported with the PBMCM (also see below). We found that a_H = 22.26 Å, Z = 6.87, and ζ = 32 mV. Thus, the effect of the electrostatic interaction of charged groups becomes evident mainly on the net protein charge, while the particle potential does not change within the round-off error considered here. The values of ΔpK_i obtained did not exceed 0.63. This value corresponds to Asp at position 66. It is then clear that the protein charge difference between the two situations analyzed here is around 7.7%. Also the evaluation of dimensionless coordinates yields $Y \approx X \approx 1.2$ and P = 0.65, indicating that the Debye–Hückel approximation and Henry solution are valid. In what follows, we do not comment any more on these coordinates as long as the quantitative requirements for validity are fulfilled.

Table 1 presents a comparison of net charge number and hydrodynamic radius for 2LYZ obtained through different models under the protocol specified above. It is worth pointing out from this table that the Z obtained with the PLLCEM is close to that reported in [19] with the PBMCM. The radius obtained from PLLCEM, LLCEM, and RLCLM are almost the same, while that reported for LLM is an average value imposed to the calculations, which resulted from an average of the radii measured using charge ladders in 8, 33, and 108 mM solutions. This last aspect is pointed out in Section 1 in relation to the need of finding a_H as a root of LLM.

Table 1. Predictions of net charge number and hydrodynamic radius of 2LYZ through PLLCEM and LLCEM and comparison with values of other models

Model	PLLCEM	LLCEM	PBMCM [19]	LLM [19]	LRCLM [19]
Z	6.87	6.34	6.90	6.30	7.30
a_H, Å	22.26	22.22	a)	b)	22.35

a) Value not required.
b) Hydrodynamic radius is imposed at 21 Å [19].

A crucial study consisted in testing the hypothesis of previous works that considered the hydrodynamic radius constant for different steps of 2LYZ charge ladder. Figure 1 shows that values of a_H calculated through both PLLCEM and LLCEM are near constant for the three first steps only. These results show that there is a functionality between a_H and eZ, within the theoretical framework of electrostatic and electrokinetics. Thus, as the protein loses positive charges along the charge ladder of six acylations by changing the net charge number from 6.87 to 0.94, a_H decreases as indicated in Fig. 1 mainly due to the decrease of charge repulsions (also see below for a discussion concerning the negligible effect on a_H of acetic anhydride as acetylating agent). This figure also depicts that values of obtained from both models as a function of n are close.

It is relevant to point out here that the number of possible acetylation sites in the protein N_{AAS} for a given value of (the number of charged groups converted) is $N_{AAS} = n_a!/(n_a - n)!n!$, where n_a refers to the total number of conversions through the protein charge ladder procedure. For instance, for n = 3 of the six Lys in the 2LYZ yields N_{ASS} = 20. Therefore, one can calculate the average parameters of these possibilities corresponding to a given n. Nevertheless, calculations with simple models show in general very little dependence of results on N_{ASS}.

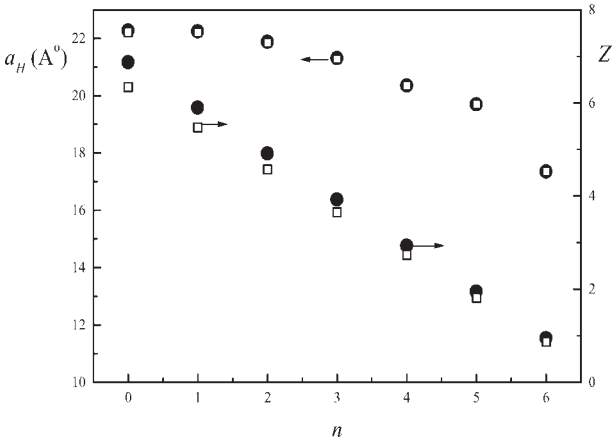

Figure 1. Hydrodynamic radius a_H and net charge number Z of 2LYZ as a function of the number n acetylated Lys ε-amino groups. Symbols (●) and (□) refer to numerical predictions of PLLCEM and LLCEM, respectively.

Further analysis of 2LYZ indicates that while LLCEM and LLM provide uniform ΔpK_i (this means that ΔpK_i are the same for all the charged groups in the protein), the PLLCEM makes a clear distinction of ΔpK_i because r_{ij} data are included in the model (see paragraph below Eq. 8 and the Appendix, where appropriate mathematical expressions are deduced). We found through PLLCEM, for instance, that the amino-terminal pK_{NH_2} of 2LYZ takes a value 7.47 for the unmodified protein, almost in the track of the experimental value reported as 7.5 in [19] for a higher ionic strength (33 mM) than that studied here (8 mM). This rough comparison is carried out with this difference in I because the electrophoretic mobility of 2LYZ is not available in [19] for the higher value of ionic strength to perform exact calculations. For the peracetylated protein (all Lys residues acetylated), we obtained $pK_{NH_2} \approx 7.74$ when $\varepsilon' \approx 10$ to approximate again the data reported experimentally as 7.8 in [19], indicating also a change of the solvent–protein environment. This result, in addition, shows the need of a good estimate of ε', free from any fitting process as described previously, and hence this is a clear limitation of the PLLCEM at the present state of its formulation. It is also relevant to point out that this limitation is intrinsic also to the well-known PBMCM (detailed model) where the specification of ε' around charged groups within the protein domain is an open question yet [7–9, 18], and further researches will be required to step forward in this difficult point. On the other hand, when the LLCEM is used in this comparison, pK_{NH_2} is around 6.97 for the unmodified protein and 7.40 for the peracetylated protein, which are approximations with uniform ΔpK_i, as stated above.

Another case study carried out in this work involved 1CA2. The protocol data are pH = 8.4, I = 10 mM, and T = 25°C. We calculated and used the values of μ for the native protein and its 20 acetylations provided by the CZE of the protein charge ladder reported in the literature [20]. The number of charged groups calculated from PDB is $\{n_{Arg} = 7;\ n_{Asp} = 19;\ n_{Cys} = 1;\ n_{Glu} = 13;\ n_{His} = 12;\ n_{Lys} = 23;\ n_{Tyr} = 8\}$. In addition, the terminal amino group is considered preacetylated and hence not reactive, and the Zn^{2+}–OH ion is coordinated by three His residues having a $pK^{o}_{Zn^{2+}-OH} = 7.0$ [37]. The set of pK^{o}_i used here is $\{pK^{o}_{Arg} = 12.0;\ pK^{o}_{Asp} = 4.0;\ pK^{o}_{Cys} = 9.3;\ pK^{o}_{Glu} = 4.4;\ pK^{o}_{His} = 6.5;\ pK^{o}_{Lys} = 10.7;\ pK^{o}_{Tyr} = 10.2;\ pK^{o}_{COOH} = 4.9\}$ [38]. In addition, the AAS and the set of distances $\{r_{ij}\}$ were obtained from the PDB. The values of Z and a_H of native 1CA2 calculated with PLLCEM and LLCEM compare well with those found by other authors as reported in Table 2. One concludes that calculations from the first steps of the protein charge ladder provide results close to those obtained directly through PLLCEM and LLCEM.

Table 2. Predictions of net charge number and hydrodynamic radius of 1CA2 through PLLCEM and LLCEM and comparison with LRCLM

Models	PLLCEM	LLCEM	LRCLM [20]
Z	−2.13	−2.06	−2.30
a_H, Å	26.47	26.33	27.00

In addition, Fig. 2 shows a_H as a function of the number of acetylations n. In this figure, one observes that for the first seven acetylations the hydrodynamic radius is fairly constant. However for $n > 7$ the assumption of constant a_H is not acceptable. What is relevant to point out here is that PLLCEM and LLCEM provide both eZ and a_H without the need of carrying out a protein charge ladder, and that the results thus obtained compare well with those found with LRCLM, which are valid only for the first acetylations steps.

To cross-check this analysis, CZE results for 1V9E were considered. The protocol data are the same as those of 1CA2. We calculated and used the values of μ for the

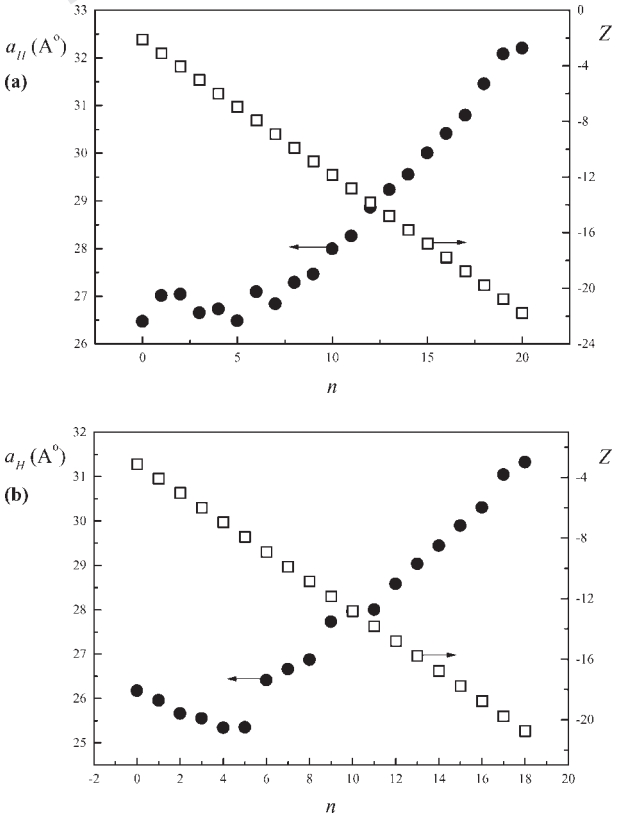

Figure 2. PLLCEM predictions of hydrodynamic radius a_H and net charge number Z of 1CA2 (a) and 1V9E (b) as a function of the number n of acetylated Lys ε-amino groups.

native protein and its 18 acetylations provided by the CZE of the protein charge ladder reported in the literature [21]. The number of charged groups calculated from PDB is $\{n_{Arg} = 9;\ n_{Asp} = 19;\ n_{Glu} = 11;\ n_{His} = 11;\ n_{Lys} = 18;\ n_{Tyr} = 7\}$. In addition, the terminal amino group is pre-acetylated and the Zn^{2+}–OH ion is coordinated by three His residues having $pK^{o}_{Zn^{2+}-OH} = 7.0$ [21]. The set of pK^{o}_{i} used is $\{pK^{o}_{Arg} = 12.5;\ pK^{o}_{Asp} = 3.5;\ pK^{o}_{Glu} = 4.5;\ pK^{o}_{His} = 6.2;\ pK^{o}_{Lys} = 10.3;\ pK^{o}_{Tyr} = 10.3;\ pK^{o}_{COOH} = 3.2\}$ [21]. In addition, the AAS and the set of distances $\{r_{ij}\}$ were obtained from the PDB. The values of of native 1V9E calculated with PLLCEM and LLCEM compare fairly well with those found by these authors as reported in Table 3.

Table 3. Predictions of net charge number and hydrodynamic radius of 1V9E through PLLCEM and LLCEM and comparison with values of other models

Models	PLLCEM	LLCEM	LRCLM [21]	LLM [21]
Z	−3.12	−3.06	−3.40	−3.00
a_H, Å	26.17	25.99	20.50	[a]

a) Hydrodynamic radius is imposed at 20.5 Å [21, 25].

It is also interesting to point out here that the difference in the net charge predictions between PLLCEM and LRCLM is approximately 10%, which is an overestimation of the latter model reported in [21]. Nevertheless, comparison of the hydrodynamic radius prediction shows a significant difference for this case, having taken into account that Gitlin *et al.* [21] used a value of 20.5 Å, which was calculated with an average value of the partial specific volume of proteins (0.72 mL/g); thus, electrokinetic effects at the level of a Henry particle are neglected. It is clear that this value of a_H is not the root that results from LLCEM as indicated in Table 3. Nevertheless, Gitlin *et al.* [21] also used a_H = 25 Å in the LLM yielding a value of, which is closer to the predictions of PLLCEM and LLCEM.

For 1V9E, a_H as a function of the number of acetylations n is fairly constant for around the first four acetylations only. However, for $n > 4$ the assumption of constant a_H is not appropriate. Thus, the same conclusions as those mentioned above for the 1CA2 may be obtained here.

An estimate of the value of ΔZ for two consecutive acylation in the 1V9E can be obtained simply from the LLCEM. For the first five rungs, ΔZ takes values between −0.94 and −0.95; thus, for practical calculations ΔZ may be considered approximately constant for this protein, although different from −1.

It is relevant to point out that in Fig. 2, the values of a_H reported for both 1CA2 and 1V9E cover a high number of acylations (20 and 18, respectively). For high values of acylations, the coordinate P does not exceed 1.06, while Y and X remain low, indicating that the linearized Poisson–Boltzmann approximation is still valid and the parametric effect of P on the results is not observed.

Before considering a different protein, it is appropriate to analyze the effect of the molecular weight of acylating agents used to produce the synthesized charge ladders, of course, in the context of the models studied in this work. For this purpose the electrophoretic mobilities provided by Colton *et al.* [28] for this type of analysis are used with LLCEM and PLLCEM. In these calculations two agents are considered: acetic anhydride (low molecular weight) and *N*-hydroxysuccinimidyl cholate (high molecular weight of 389 Da). We found in coincidence with these authors that as the number of acylation n increases in 1V9E, there exists an increasing difference in hydrodynamic radius between the two types of acylations, indicating that the high molecular weight agent provides an additional radius change, apart from that observed due to other physicochemical effects associated with protein net charge. Thus, when 5, 10, and 13 acylations are included, for instance, in 1V9E, the LLCEM results indicate that the differences in hydrodynamic radius at each acylation are 1.53, 1.95, and 2.64 Å, respectively, while the net charge number remains almost unchanged (changes are less than around 0.16). Nevertheless, when one to three acylations are considered for example, no difference is observed when these two different agents are used. Similar results are found with the PLLCEM. In this sense, it should be pointed out that throughout this work, we have used mobility experimental data of proteins modified with the low molecular weight acylating agent only, to avoid the effect of protein chemical change on the predictions of a_H.

Another study carried out in this work involved 1STN, where changes of solvent pH and ionic strength I were included according to CZE data available in the literature [39, 40] for T = 25 C. The set of pK^{o}_{i} used here is $\{pK^{o}_{NH_2} = 7.4;\ pK^{o}_{Arg} = 12.0;\ pK^{o}_{Asp} = 4.0;\ pK^{o}_{Glu} = 4.4;\ pK^{o}_{His} = 6.5;\ pK^{o}_{Lys} = 10.7;\ pK^{o}_{Tyr} = 10.2;\ pK^{o}_{COOH} = 4.9\}$ [38]. The number of charged groups calculated from PDB is $\{n_{Arg} = 5;\ n_{Asp} = 8;\ n_{Glu} = 12;\ n_{His} = 4;\ n_{Lys} = 23;\ n_{Tyr} = 7\}$. In addition, the AAS and the set of distances $\{r_{ij}\}$ were obtained from the PDB. Table 4 presents numerical results obtained with PLLCEM and LLCEM and also those values reported in [15, 39] for comparison. This table shows that the values of Z from these models are closer to the predictions reported in [15], while the values of a_H obtained are rather different

Table 4. Predictions of net charge number and hydrodynamic radius of 1STN through PLLCEM and LLCEM and comparison with values reported by other authors

pH	I, mM	PLLCEM		LLCEM		Kálmán *et al.* [39]		Winzor *et al*. [15]	
		Z_i	a_H, Å	Z_i	a_H, Å	Z_i	a_H, Å	Z_i	a_H, Å
2.8	5.5	31.02	53.71	29.30	52.34	33.0	50.1	a)	a)
4.1	55.0	22.58	29.53	17.78	28.28	20.6	26.8	19.6	27.0
5.7	36.0	11.43	23.95	10.69	23.52	14.8	25.2	13.2	21.0
6.8	25.5	8.61	21.00	8.0	20.85	12.9	25.1	9.5	21.0
8.9	14.0	5.66	22.16	5.29	21.48	7.49	24.6	5.6	21.0
9.5	6.5	3.75	24.63	2.74	20.94	5.40	30.5	a)	a)

a) Values not reported.

from those reported in [15, 39]. Nevertheless, in the former reference the same hydrodynamic radius for three different pH was used, and the values of Z and a_H reported in [39] are estimated from CZE of protein mutations, a procedure that may be considered as a one-step protein charge ladder. In this sense, deviations to the expected values may be found because for a better prediction a charge ladder with additional steps is needed. Therefore, the advantage of PLLCEM and LLCEM on other simple models becomes clear, providing as solution both Z and a_H.

For this protein case study, we also evaluated the effect of solvent pH on Z_i of some charged groups with low, intermediate, and high values of pK_i^o. In fact, by using the mobility of 1STN at different pH, Figs. 3–5 are constructed, which show the evolution of the net charge number of a selected group in the protein as pH is changed from 2.8 to 9.5. In these figures, Z_i^o and Z_i of i-group calculated with both LLCEM and PLLCEM are reported. For instance, in Fig. 3, Asp-40 with $pK_{Asp}^o = 4.0$ shows that the net charge number of this group is quite the same as its wild value, except for pH = 4. $\pm$ 2, where ΔpK_{Asp-40} becomes significant in the calculations. Figure 4 illustrates this aspect for His-121 where differences in the net charge number estimations are found around a pH close to $pK_{His}^o = 6.5$. Figure 5 presents another case when the pH of BGE is rather high affecting, for instance Lys-6 with $pK_{Lys}^o = 10.7$. With these results we formally prove that the effect on the protein net charge of both electrostatic interaction between pairs of charged groups and pH variation near molecule are mainly relevant for those groups having a pK_i^o value near the pH of the solvent [19, 28]. Therefore, one can formulate a BGE by choosing the pH with the constraint that its value must be far away from (or close to) the pK_i^o of a specific charged group to avoid (or not) a dependence on the magnitude of the ΔpK_i in order to get an appropriate protein separation in CZE.

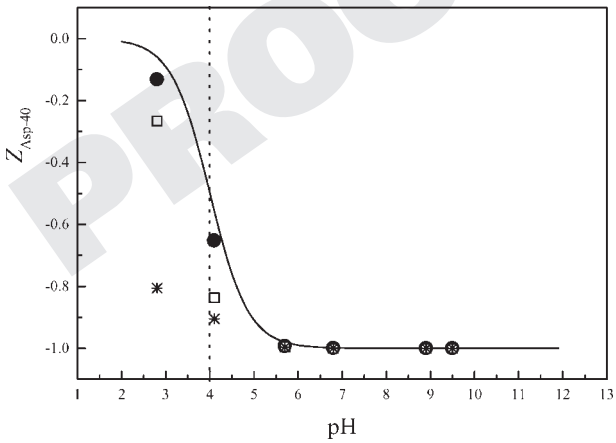

Figure 3. Net charge number Z_{Asp-40} of 1STN protein as a function of pH (see Table 4 for values of ionic strength). Full line refers to the wild charge prediction without charge regulation phenomena. Symbols (●), (□), and (*) indicate numerical predictions of the PLLCEM, LLCEM, and LLM, respectively. Vertical dot line is placed at pH equal to pK_{Asp}^o = 4. LLM uses a constant hydrodynamic radius of 21 Å.

In addition, in Figs. 3–5, calculations carried out with LLM are also reported. In this model, the hydrodynamic radius is an input and a fixed value as explained above, and hence differs conceptually from both the PLLCEM and LLCEM. In this sense, by using as estimate, for instance, $a_H = 21$ Å (this value is appropriate for a pH around 6.8; see Table 4), the predictions of Z_{Asp-40} (Fig. 3) and $Z_{His-121}$ (Fig. 4) with LLM provide values that deviate from those obtained with both PLLCEM and LLCEM. This situation is very critical for pH = 4.1 and 2.8 in relation to Z_{Asp-40} where the 1STN becomes denatured (acidic "swollen" state for pH = 4.1 and random coil for pH = 2.8). Thus, at these extreme pH a fixed radius for LLM and fixed distances r_{ij} for PLLCEM are not any more

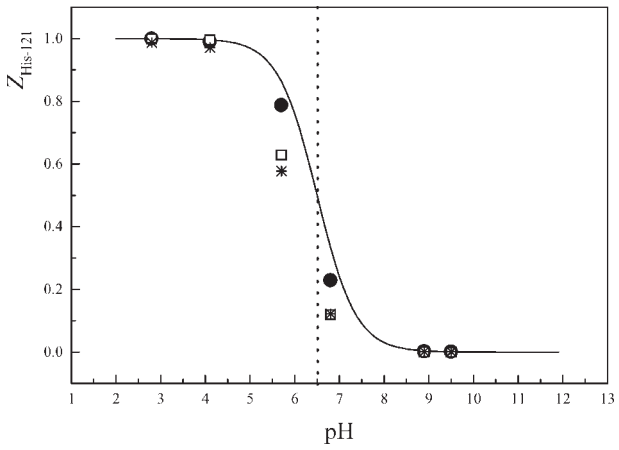

Figure 4. Net charge number $Z_{His-121}$ of 1STN protein as a function of pH (see Table 4 for values of ionic strength). Full line refers to the wild charge prediction without charge regulation phenomena. Symbols (●), (□), and (*) indicate numerical predictions of the PLLCEM, LLCEM, and LLM, respectively. Vertical dot line is placed at pH equal to pK^{o}_{His} = 6.5. LLM uses a constant hydrodynamic radius of 21 Å.

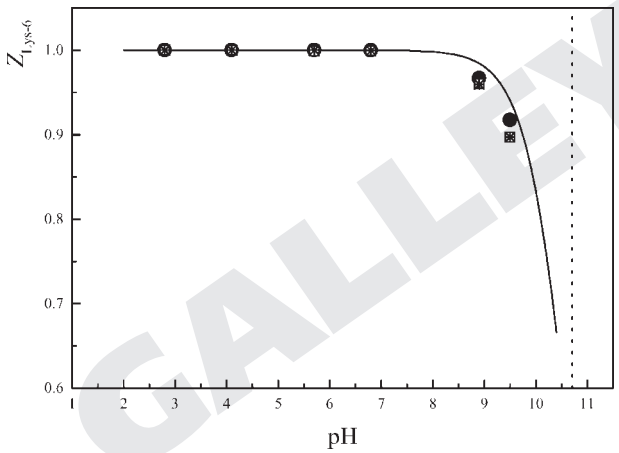

Figure 5. Net charge number Z_{Lys-6} of 1STN protein as a function of pH (see Table 4 for values of ionic strength). Full line refers to the wild charge prediction without charge regulation phenomena. Symbols (●), (□), and (*) indicate numerical predictions of PLLCEM, LLCEM, and LLM, respectively. Vertical dot line is placed at pH equal to pK^{o}_{Lys} = 10.7. LLM uses a constant hydrodynamic radius of 21 Å.

appropriate. In fact, at these extreme physicochemical conditions the LLCEM may be giving the best answer to the problem, because this model does not include these types of inputs. Of course the unexpected predictions of the LLM may be a consequence of imposing a constant radius on the calculations, as mentioned above in the presentation of these models. Also for this analysis, the above discussion concerning Tables 1 and 3 are appropriate, where it is evident that LLM, as it has been used in the literature most of the time, is not fully solved to predict the value of a_H for different pH and *I*.

A study when the ionic strength of the BGE is varied at a fixed pH is also of interest. From Eqs. 3 and 4, one infers that as long as *I*, and hence κ', are kept high enough small ΔpK_i should be expected. For the 1V9E, a study of the electrophoretic mobility at constant pH = 8.4 and varying ionic strengths was carried out by Carbeck *et al.* [24]. Therefore, the experimental data reported in [24] are used here to analyze the sensitivity of *I* in the predictions of LLCEM (this method does not require crystallographic distances, and hence, is simpler in order to make these calculations). Table 5 shows the results obtained when the native protein is studied. Thus, a_H decreases and *Z* increases with increments of *I* (the second parameter, however, does very weakly). Although these results are in part consistent with the physical view that one expects a smaller electrical double layer around the particle, it is also important to place emphasis on the fact that the numerical solution obtained is the result of solving a strong and coupled nonlinear mathematical problem, even in the case of the LLCEM. In fact, the effect of *I* is observed mainly through parameter κ placed in the denominator and in the function $f(\kappa a_H)$ both pertaining to Eq. 6, which in turn inputs smaller mobilities for this study in the numerical iterating process described in Section 2.4. Furthermore, when electrostatic effects between pairs of charged groups are considered through the PLLCEM, this effect is also validated. Other calculations carried out for different acetylations show similar trends with the variation of *I*.

Table 5. Prediction of net charge number and hydrodynamic radius of 1V9E through LLCEM, for the same protocol as that used in Table 3, with different values of ionic strength *I* (various concentrations of Li_2SO_4 in the electrophoresis buffer)

$[Li_2SO_4]$, mM	*Z*	a_H, Å	ΔpK_i
0	−3.06	25.99	−0.20
10	−3.09	24.13	−0.15
30	−3.12	23.19	−0.12
50	−3.13	22.23	−0.11

It is also worth pointing out that different effects may be achieved when aqueous solvent electrical permittivity is modified by adding, for instance, an organic solvent to the BGE [41], although in this case the problem requires to have a new set of pK^{o}_{i} referred to the thermodynamic conditions of the charged group dissociations [42, 43].

With the purpose of illustrating our calculations with a protein of molecular weight higher than those of the proteins analyzed above, we carried out an experimental and theoretical study of the human serum albumin (1AO6). The protocol data were pH = 9.8, I = 110 mM, and T = 25°C. The CZE test yielded $\mu = 1.57 \times 10^{-8}$ m^2/s × V. The set of pK_i^o used is {p$K^o_{NH_2}$ = 9.27; pK^o_{Arg} = 12.48; pK^o_{Asp} = 3.65; pK^o_{Glu} = 4.25; pK^o_{His} = 6.; pK^o_{Lys} = 10.53; pK^o_{Tyr} = 10.07; pK^o_{COOH} = 2.35; pK^o_{Cys} = 8.18; } [44]. The number of charged groups calculated from PDB is {n_{Arg} = 24; n_{Asp} = 36; n_{Cys} = 35; n_{Glu} = 62; n_{His} = 16; n_{Lys} = 59; n_{Tyr} = 18}. In this case it is considered that 34 Cys are forming disulfide bridges (possible binding ions are neglected for this illustration). The LLCEM generated the following numerical results: ζ = –30 mV, a_H = 30.74 Å, Z = –22.22, and pH* = 9.28, while the PLLCEM with $\varepsilon' \approx 70\varepsilon_o$ provided: ζ = –30 mV, a_H = 31.73 Å, Z = –24.14, and pH* = 9.29. These values of pH* indicate, of course, that there is a higher proton concentration near the protein. One also observes that the net charge number of this protein is relatively high in relation to the proteins analyzed above. Further, the value of the electrical permittivity of the solvent–protein domain is relatively high for a net charge estimation of the protein with approximately the same value as that predicted with LLCEM. For this protein, $Y \approx 1.3$, $X \approx 1.2$, and P = 3.44. Thus, despite Y and X coordinates being rather close and still low, a relatively high value of P is found and parametric curves Y *versus* X start to separate one another. At this level of charge number and protein size, the classical hypothesis introduced in PLLCEM could be starting to fail, and the full Poisson–Boltzmann equation should be solved.

Finally, it is also important to point out here that, in general, pK_i^o are reported in the literature with an experimental precision of around 1/10 depending on the technique used for their measurements. Also these values may differ from one source of information to another (see, *e.g.*, [38, 44]). This experimental error is around the same value as the net charge variation of proteins found in mutations or in each step of the charge ladder procedure. This experimental error also affects the numerical results provided by the PLLCEM and LLCEM. In this sense, the quality of the data inputs in computational programs is relevant. At present, there exists a difficulty to find pK_i^o data for different BGE that can be used reliably, as reported by Vcelákova *et al.* [43]. Consequently, one of the emphases to be placed in future research should be on improving the experimental determinations of pK_i^o in different solvents at a well-specified ionic strength [43].

In order to have a visualization of how experimental errors associated with pK_i^o may affect the calculations provided by PLLCEM, we have imposed arbitrarily errors to the pK_i^o used with 2LYZ. Thus, one can assume by turns that the pK_i^o of a specific charged group contains an error of $\pm$1 and $\pm$0.1 also. The results are reported in Table 6, where one observes that a precision of 1/10 would be acceptable for the model predictions, even for those charged groups that have a pK_i^o close to the pH of the running protocol. On the other hand, an error $\pm$1 would be detrimental for some proteins, like it is illustrated in Table 6 for 2LYZ. Therefore, one concludes that the calculations obtained from PLLCEM provide an important guide on how to select the pH in a protocol for separating a mixture of proteins.

Table 6. Prediction of net charge number and hydrodynamic radius of 2LYZ through PLLCEM for the same protocol as that used in Table 1, when an error $\pm$1 and $\pm$0.1 U of pK_i^o is introduced in a protein charged group (one *per* time)

pK_i^o	a'_H, Å	Z'	a'_H–a_H	Z'–Z
pK^o_{Asp} ±1	22.26	6.872	0.000	0.001
pK^o_{Glu} ±1	22.26	6.872	0.000	0.001
pK^o_{His} ±1	22.33	6.918	0.07	0.047
pK^o_{Lys} ±1	18.36	6.863	−3.90	−0.008
pK^o_{Arg} ±1	22.21	6.845	−0.05	−0.026
pK^o_{Tyr} ±1	16.89	6.882	−5.37	0.011
pK^o_{Asp} ±0.1	22.26	6.871	0.000	0.000
pK^o_{Glu} ±0.1	22.26	6.871	0.000	0.000
pK^o_{His} ±0.1	22.26	6.872	0.000	0.001
pK^o_{Lys} ±0.1	22.16	6.850	−0.100	−0.021
pK^o_{Arg} ±0.1	22.26	6.872	0.000	0.001
pK^o_{Tyr} ±0.1	21.94	6.879	−0.320	0.008

The values of a'_H and Z' reported pertain to calculations with each pK_i^o changed by the error indicated, while a_H and Z are the values of the reference run of Table 1. Maximum differences are reported only.

Following the discussion of results, it is appropriate to point out here the limitations found in simple models like PLLCEM, LLCEM, LRCEM, and LLM for their direct application to more practical problems. These limitations may be summarized as follows:

(1) In the use of the PLLCEM a critical requirement is the determination of ε', which is quite difficult to estimate *a priori* for a given protein without having available data, for instance, from the PBMCM. Also a possible estimation of ε' may be obtained by achieving with the PLLCEM the same net charge as that reported through other classical methods. In addition, the LLCEM may provide an estimate in this sense, which does not consider the internal protein environment in calculations. The strength of the

PLLCEM, however, is that an estimation of the electrostatic interactions between charged groups within the protein is provided and values of ΔpK_i may be distinguished among charged groups, once ε' is available.

(2) Simple models under study faces the difficulty that the set $\{pK_i^o\}$ is required and the error inherent to experimental evaluations associated with the dissociation constants may introduce a relevant parametric sensitivity in the models (mainly when errors could exceed approximately 1/10 U of pK_i^o), giving changes in results of the same order of magnitude as those of ΔpK_i predicted. In this sense, an error no greater than 1/10 seems acceptable, for example, for 2LYZ.

(3) Calculations with PLLCEM and LLCEM indicate that the hypothesis of constant a_H and $|\Delta Z| \cong 1$ to evaluate $n\Delta Z$ in the LRCLM is an approximation for the first converted protein fractions in the electrophoresis of protein charge ladders.

(4) The predictions of eZ from LLM may be close to that found from LLCEM as long as the value a_H imposed in the former model is also close to the numerical value obtained with the latter one. Both models, of course, neglect the difference between pH_i and pH* giving the same ΔpK_i for all charged groups. The LLM may give inappropriate predictions concerning ΔpK_i as a consequence of imposing a rather constant radius in the calculations, as mentioned above (Figs. 3–5) in the presentation of these models.

(5) When the protein goes through significant changes from its native structure, the crystallographic distances r_{ij} may not be available in general and the PLLCEM cannot be useful in these cases, where results obtained with r_{ij} of native protein involves an additional approximation. Here one would prefer to use LLCEM, where the counting of n_i and associated ions are considered *a priori* only.

(6) Of course, those methods more empirical and quite model-independent in nature providing the protein net charge, like those proposed in Refs. [15, 16], must be used as reference points to overcome several uncertainty concerning the simple models described here. Results obtained with PBMCM are also useful in this sense, but they are scarce and not easily available.

From the above discussion it is clear that further research is required in the development of these types of models before they may be used in direct practical applications. Nevertheless, it is also relevant to indicate that the conceptual visualization of different phenomena participating in the determination of the net protein charge and hydrodynamic radius in different solvents becomes quite evident through the simple models studied here.

As long as native proteins are studied and an estimation of ε' is available, PLLCEM would be suggested to use in practice, which provides results closer to the PBMCM as shown for 2LYZ above. On the other hand, when the protein goes through significant changes from its native structure, one would prefer to use the LLCEM where the counting of n_i and other binding ions are considered while the crystallographic distances r_{ij} are ignored. Of course, methods quite model-independent in nature, like those proposed in [16], are to be used, mainly when protein structure and possible binding ions are not well known. In general, the precision of each method is a subject still to debate in the literature.

4 Concluding remarks

PLLCEM and LLCEM yield an estimation of both hydrodynamic radius and net charge of a protein as outputs of numerical solutions, when the CZE mobility at a given protocol, the set of pK_i^o of charged amino acids, and basic data from PDB (including binding ions) are provided. Numerical predictions of protein net charge through these models compare well with previous results reported in the literature, including those of the detailed PBMCM. These model predictions are also good when the net charge and hydrodynamic radius obtained are compared with the asymptotic responses of the simple LRCRM, which may provide an approximately constant hydrodynamic radius of the protein for the first steps of the protein charge ladder. Through our calculations, it is found that the hydrodynamic radius is sensitive to changes of the protein net charge for 2LYZ, 1CA2, 1V9E, and 1STN studied in this work.

As long as simple models are required for practical situations of CZE, it is important to consider the interplay between protein net charge and hydrodynamic radius in relation to the physicochemical properties of the solvent used like pH, ionic strength, electrical permittivity, and viscosity. In addition, a crucial aspect is the appropriate estimation of the electrical permittivity within the protein–solvent domain, which remains a topic of research for both detailed and simple models. In a wider context, PLLCEM and LLCEM are complementary tools of classical model-independent methods, in order to get and estimation of ΔpK_i associated with charged groups in the protein–solvent environment.

The authors wish to thank the financial aid received from SEPCYT-FONCyT (PICT 09–09752), UNL (CAI+D 12/I301), and CONICET (PIP 02554), Argentina.

Received February 25, 2005

5 References

[1] Kenndler, E., *Chem. Anal. Ser.* 1998, *146*, 25–76.

[2] Gaš, B., Coufal, P., Jaroš, M., Muzikář, J., Jelínek, I., *J. Chromatogr. A* 2001, *905*, 269–279.

[3] Jaroš, M., Vceláková, K., Zusková, I., Gaš, B., *Electrophoresis* 2002, *23*, 2667–2677.

[4] Berli, C. L. A., Piaggio, M. V., Deiber, J. A., *Electrophoresis* 2003, *24*, 1587–1595.

[5] Demchuk, E., Wade, R. C., *J. Phys. Chem.* 1996, *100*, 17373–17387.

[6] Antosiewicz, J., McCammon, J. A., Gilson, M. K., *Biochemistry* 1996, *35*, 7819–7833.

[7] Beroza, P., Case, D. A., *J. Phys. Chem*. 1996, *100*, 20156–20163.

[8] Beroza, P., Fredkin, D. R., *J. Comp. Chem*. 1996, *17*, 1229–1243.

[9] Beroza, P., Fredkin, D. R., Okamura, M. Y., Feher, G., *Proc. Natl. Acad. Sci. USA* 1991, *88*, 5804–5808.

[10] Kuehner, D. E., Engmann, J., Fergg, F., Wernick, M., Blanch, H. W., Prausnitz, J. M., *J. Phys. Chem. B* 1999, *103*, 1368–1374.

[11] Kao, Y., Fitch, C. A., Vhattacharya, S., Sarkisian, C. J., Lecomte, J. T. J., García-Moreno, E. B., *Biophys. J*. 2000, *79*, 1637–1654.

[12] Lee, K. K., Fitch, C. A., García-Moreno, E. B., *Protein Sci.* 2002, *11*, 1004–1016.

[13] Lee, K. K., Fitch, C. A., Lecomte, J. T. J, García-Moreno, E. B., *Biochemistry* 2002, *41*, 5656–5667.

[14] Fitch, C. A., Karp, D. A., Lee, K. K., Stites, W. E., Lattman, E. E., García-Moreno, E. B., *Biophys. J.* 2002, *82*, 3289–3304.

[15] Winzor, D. J., Jones, S., Harding, S., *Anal. Biochem.* 2004, *333*, 225–229.

[16] Winzor, D. J., *Anal. Biochem.* 2004, *325,* 1–20.

[17] Allison, S. A., Carbeck, J. D., Chen, C., Burkes, F., *J. Phys. Chem. B* 2004, *108*, 4516–4524.

[18] You, T. J., Bashford, D., *Biophys. J*. 1995, *69*, 1721–1733.

[19] Sharma, U., Negin, R. S., Carbeck, J. D., *J. Phys. Chem. B* 2003, *107*, 4653–4666.

[20] Sharma, U., Carbeck, J. D., in: Strege, M. A., Lagu, A. L. (Eds.), *Capillary Electrophoresis of Proteins and Peptides. Methods in Molecular Biology*, Humana Press, Totowa, New Jersey 2004, pp. 189–216.

[21] Gitlin, I., Mayer, M., Whitesides, G. M., *J. Phys. Chem. B* 2003, *107*, 1466–1472.

[22] Carbeck, J. D., Negin, R. S., *J. Am. Chem. Soc*. 2001, *123*, 1252–1253.

[23] Menon, M. K., Zydney, A. L., *Anal. Chem*. 2000, *72*, 5714–5717.

[24] Carbeck, J. D., Colton, I. J., Anderson, J. R., Deutch, J. M., Whitesides, G. M., *J. Am. Chem. Soc*. 1999, *121*, 10671–10679.

[25] Menon, M. K., Zydney, A. L., *Anal. Chem*. 1998, *70*, 1581–1584.

[26] Córdova, E., Gao, J., Whitesides, G. M., *Anal. Chem.* 1997, *69*, 1370–1379.

[27] Gao, J., Whitesides, G. M., *Anal. Chem*. 1997, *69*, 575–580.

[28] Colton, I. J., Anderson, J. R., Gao, J., Chapman, R. G., Isaacs, L., Whitesides, G. M., *J. Am. Chem. Soc*. 1997, *119*, 12701–12709.

[29] Gao, J., Gomez, F. A., Härter, R., Whitesides, G. M., *Proc. Natl. Acad. Sci. USA* 1994, *91*, 12027–12030.

[30] Reijenga, J. C., Kenndler, E., *J. Chromatogr. A* 1994, *659*, 403–415.

[31] Linderstrøm-Lang, K. C., *Trav. Lab. Carlsberg* 1924, *15*, 1–29.

[32] Russel, W. B., Saville, D. A., Schowalter, W. R., *Colloidal Dispersions*, Cambridge University Press, Cambridge, UK 1989, pp. 96–108.

[33] Atkins, P. W., *Physical Chemistry*, Oxford University Press, Oxford 1995, pp. 316–317.

[34] Piaggio, M. V., Deiber, J. A., *Lat. Am. Appl. Res.* 2003, *33*, 261–268.

[35] Yoon, B. J., Kim, S., *J. Colloid Interface Sci.* 1989, *128*, 275–288.

[36] O'Brien, R. W., White, L. R., *J. Chem. Soc. Faraday Trans.* 1978, *74*, 1607–1626.

[37] Caravella, J. A., Carbeck, J. D., Duffy, D. C., Whitesides, G. M., Tidor, B., *J. Am. Chem. Soc*. 1999, *121*, 4340–4347.

[38] Creighton, T. E., *Proteins: Structures and Molecular Properties*, W. H. Freeman and Company, New York 1993, p. 6.

[39] Kálmán, F., Ma, S., Fox, R. O., Horváth, C., *J. Chromatogr. A* 1995, *705*, 135–154.

[40] Kálmán, F., Ma, S., Hodel, A., Fox, R. O., Horváth, C., *Electrophoresis* 1995, *16*, 595–603.

[41] Porras, S. P., Riekkola, M.-L., Kenndler, E., *Electrophoresis* 2003, *24*, 1485–1498.

[42] Beckers, J. L., Ackermans, M. T., Boček, P., *Electrophoresis* 2003, *24*, 1544–1552.

[43] Vceláková, K., Zusková, I., Kenndler, E., Gas, B., *Electrophoresis* 2004, *25*, 309–317.

[44] Nelson, D. L., Cox, M. M., *Lehninger Principios de Bioquímica*, Ediciones Omega S. A., Barcelona 2001, p. 118.

[45] Tanford, Ch., *Physical Chemistry of Macromolecules*, John Wiley & Sons, USA 1961, pp. 457–488.

6 Appendix

By following Boltzmann distribution of ions, the proton concentration $[H_i^+]$ around the protein *i*-charged group is expressed as

$$[H_i^+] = [H^+] \exp\left(-\frac{e\zeta_i}{k_B T}\right) \tag{A-1}$$

In Eq. A-1, $[H^+]$ is the molar concentration of protons in the BGE. Then, Eq. A-1 may be rewritten as follows:

$$pH_i = pH + \frac{1}{\ln(10)}\frac{e\zeta_i}{k_B T} \tag{A-2}$$

where ζ_i is the electrical potential of *i*-charged group expressed as

$$\zeta_i = \zeta + \Delta\zeta_i^B + \sum_{\substack{j=1 \\ j\neq i}}^{N} \Delta\zeta_{ij} \tag{A-3}$$

This equation involves (a) the particle potential (mean field approximation) [32]

$$\zeta = \frac{eZ}{4\pi\varepsilon a_H(1 + \kappa a_H)} \tag{A-4}$$

(b) the *i*-electrostatic potential perturbation $\Delta\zeta_i^B$, which is associated with the transfer of *i*-charged group in solution into a cavity in the protein [33], expressed as

$$\Delta\zeta_i^B = \frac{e\Delta Z_i}{4\pi r_i^o \varepsilon}\left(\frac{\varepsilon}{\varepsilon'} - 1\right) \quad \text{(A-5)}$$

and (c) the *i*-electrostatic potential perturbation $\Delta\zeta_{ij}$ due to the presence of *j*-charged group at a distance r_{ij}, expressed [12, 17, 19, 45] as

$$\Delta\zeta_{ij} = \frac{e\Delta Z_j}{4\pi\varepsilon'}\frac{\exp(-\kappa' r_{ij})}{r_{ij}} \quad \text{(A-6)}$$

From Eqs. A-1–A-6 one readily obtains Eq. 3.

In addition, the pH near the protein particle is derived from

$$[H^+]^* = [H]\exp\left(-\frac{e\zeta}{k_B T}\right) \quad \text{(A-7)}$$

where $[H^+]^*$ is the molar proton concentration near particle, and hence,

$$pH^* = pH + \frac{1}{\ln(10)}\frac{e\zeta}{k_B T} \quad \text{(A-8)}$$

From Eqs. A-4 and A-8, one readily obtains Eq. 4.

Finally, since $K_i^o = K_i \exp\left(-\frac{e\zeta_i}{k_B T}\right)$ where pK_i is the effective dissociation constant of *i*-charged group in the protein–solvent environment, one also finds

$$pK_i^o = pK_i + \frac{1}{\ln(10)}\frac{e\zeta_i}{k_B T} \quad \text{(A-9)}$$

This equation combined with Eq. A-2 yields the expression for ΔpK_i used in the calculations carried out with PLLCEM.