

Article

# Coupling of Charge Regulation and Conformational Equilibria in Linear Weak Polyelectrolytes: Treatment of Long-Range Interactions via Effective Short-Ranged and pH-Dependent Interaction Parameters

Pablo M. Blanco <sup>1,\*</sup> , Sergio Madurga <sup>1,\*</sup> , Francesc Mas <sup>1</sup>  and Josep L. Garcés <sup>2</sup>

<sup>1</sup> Physical Chemistry Unit, Department of Materials Science and Physical Chemistry & Research Institute of Theoretical and Computational Chemistry (IQTCUB) of Barcelona University (UB), 08028 Barcelona, Catalonia, Spain; fmas@ub.edu

<sup>2</sup> Department of Chemistry, Technical School of Agricultural Engineering & Agrotecnio of Lleida University (UdL), 25003 Lleida, Catalonia, Spain; jlgarces@quimica.udl.cat

\* Correspondence: pmblanco@ub.edu (P.M.B.); s.madurga@ub.edu (S.M.)

Received: 15 June 2018; Accepted: 23 July 2018; Published: 24 July 2018



**Abstract:** The classical Rotational Isomeric State (RIS) model, originally proposed by Flory, has been used to rationalize a wide range of physicochemical properties of neutral polymers. However, many weak polyelectrolytes of interest are able to regulate their charge depending on the conformational state of the bonds. Recently, it has been shown that the RIS model can be coupled with the Site Binding (SB) model, for which the ionizable sites can adopt two states: protonated or deprotonated. The resulting combined scheme, the SBRIS model, allows for analyzing ionization and conformational equilibria on the same foot. In the present work, this approach is extended to include pH-dependent electrostatic Long-Range (LR) interactions, ubiquitous in weak polyelectrolytes at moderate and low ionic strengths. With this aim, the original LR interactions are taken into account by defining effective Short-Range (SR) and pH-dependent parameters, such as effective microscopic protonation constants and rotational bond energies. The new parameters are systematically calculated using variational methods. The machinery of statistical mechanics for SR interactions, including the powerful and fast transfer matrix methods, can then be applied. The resulting technique, which we will refer to as the Local Effective Interaction Parameters (LEIP) method, is illustrated with a minimal model of a flexible linear polyelectrolyte containing only one type of rotating bond. LEIP reproduces very well the pH dependence of the degree of protonation and bond probabilities obtained by semi-grand canonical Monte Carlo simulations, where LR interactions are explicitly taken into account. The reduction in the computational time in several orders of magnitude suggests that the LEIP technique could be useful in a range of areas involving linear weak polyelectrolytes, allowing direct fitting of the relevant physical parameters to the experimental quantities.

**Keywords:** polyelectrolytes; charge regulation; long-range interactions; Debye–Hückel interactions; transfer matrix; Ising models; semi-grand canonical ensemble; Monte Carlo simulations; conformational equilibria; variational methods

## 1. Introduction

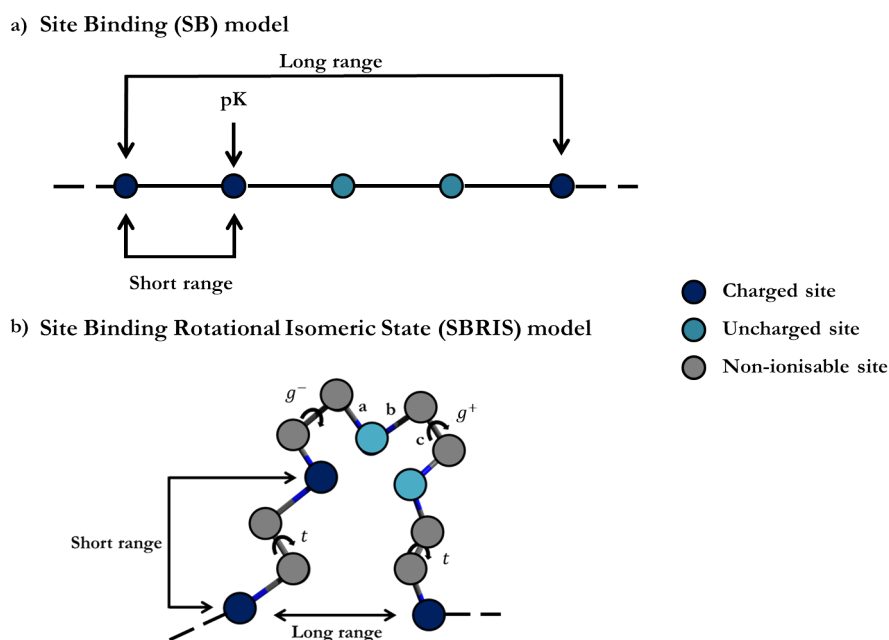
The ionization state of charged macromolecules in solution is regulated by the binding of small ions (protons, metal ions, etc.) present in the backward medium. In particular, acid-basic equilibria in weak polyelectrolytes represent the paradigmatic mechanism of charge regulation due to the ubiquitous presence of proton ions in aqueous solution. These processes are of paramount importance

to understand the physicochemical behavior of charged macromolecules in a wide range of situations. Just to mention a few examples, charge regulation plays a fundamental role in receptor–ligand equilibria in biochemical systems [1–4], supramolecular chemistry [5–7], the role of natural organic matter in geochemical cycle of metal ions [8], wastewater treatment [9], stability of colloidal systems [10], advanced coating in material science [11–13] or drug delivery [14]. Charge regulation can take place on rigid structures, such as surfaces or nano-particles [15], but in general polyelectrolytes are flexible and conformational and ionization degrees of freedom are strongly coupled. This fact can result in dramatic structural changes in the macromolecule. Classical examples are the helix–coil transitions of poly(peptides) [16], the swelling of poly(methacrylic) acid in a very narrow range of pH [17] or the strong influence of ionization in the folding of proteins [18]. More recently, the importance of the ionization configuration in the conformational properties of intrinsically disordered proteins, whose function–structure relationship still remains a controversial matter, has been recognized [19,20].

The understanding of the ionization processes has been mainly based on the so-called Site Binding (SB) model. In this approach, the ionization configuration of the macromolecule is defined as a set of sites which can be in two possible states, i.e., protonated or deprotonated, as outlined in Figure 1a. The free energy is then parametrized by site-specific microscopic protonation constants and interaction energies between sites. Triplet or higher-order interactions among sites can also be considered. Once the system is parametrized, the machinery of statistical mechanics can be used in order to quantify the relevant physical properties such as titration curves, site-specific binding probabilities, macroscopic protonation constants, microscopic protonation enthalpies, site–site binding correlation, etc. [5,15,21–25]. For systems with a small number of sites ( $N \leq 20$ ), the necessary thermal averages can be performed by direct enumeration, while, for a large number of sites, Monte Carlo (MC) simulations become necessary [26–37].

In the case of linear polyelectrolytes, the transfer matrix method can be used to compute the relevant thermal averages [15]. This powerful and elegant technique was originally designed to solve the classical Ising model of ferromagnets. It is based on the fact that the partition function of a system with  $N + 1$  sites can be related in a recursive way to the one of a system with  $N$  sites. The resulting recursive relationship can be expressed in terms of the transfer matrix, whose elements represent the contributions of the new site to the partition function, for a given state of the preceding site [38,39]. The technique is very versatile and can be generalized to systems composed with repetitive units (spins, bonds or binding sites), which can take in principle more than two states. When applied to the binding of ions to polyelectrolytes, the method can be adapted to include a wide range of phenomena such as triplet interactions between sites [21], chelate complexation of metal ions [23], proton binding to polyampholytes [40,41], protein–DNA binding [42], super-capacitor charging [43] or coupling between ionization and conformational degrees of freedom [44–46].

Probably the most productive application of transfer matrices was proposed by Flory in the context of the Rotational Isomeric State (RIS) model [47,48], aiming to compute conformational properties of neutral linear molecules. The RIS model relies on the observation that, although a particular bond can adopt in principle any rotation angle, only those of minimum energy (typically *trans*, *gauche+* and *gauche-*) are significantly populated. As a consequence, each bond can be regarded as a ‘unit’ of the system adopting three possible states. The corresponding partition function and the necessary thermal averages (bond probabilities, end-to-end distance, radius of gyration, etc.) can be calculated using a proper product of transfer matrices. In recent works [45,46], it has been shown that SB and RIS models can be combined in a unique scheme so that conformational and ionization equilibria can be analyzed on the same foot. It has been shown that all the matrixial expressions of RIS can be systematically extended to account for the ionization degrees of freedom. The resulting SBRIS model, outlined in Figure 1b, has been recently applied to the detailed characterization of the conformational and ionization properties of linear poly(ethylene)imine [46].



**Figure 1.** (a) Outline of the Site Binding (SB) model for a linear polyelectrolyte represented as a linear chain of ionizable sites. The ionization state of the macromolecule is characterized by a set of variables  $s = \{s_i\}$ ,  $i = 1, \dots, N$ , which can adopt two possible values:  $s_i = 1$  if the site  $i$  is protonated (dark blue circles) and  $s_i = 0$  if it is deprotonated (cyan circles); (b) sketch of a linear chain joining ionizable sites by means of rotating bonds as an example of the Site Binding-Rotational Isomeric State (SBRIS) model. Both ionization and conformation degrees of freedom are now taken into account. In the depicted chain, only the bonds type c are able to rotate, which can take three possible states of minimum energy, i.e., *trans*, *gauche+* and *gauche-*.

The main limitation of the transfer matrices used in SB, RIS and SBRIS models is that they can only deal with Short-Range (SR) interactions [49,50]. SR interactions are chemically specific and can produce important correlations between neighbouring sites and bonds. They cannot be modeled by simple continuous force fields (such as van der Waals or Debye–Hückel potentials) [51] but, in exchange, they can be easily implemented in a transfer matrix scheme. For polyelectrolytes, however, this is an important restriction, due to the Long-Range (LR) nature of coulombic interactions, which severely restricts the range of application of the transfer matrix approach. In practice, the possibility of neglecting LR coulombic interactions must be restricted to high ionic strengths, an important limitation specially for polyelectrolytes which become insoluble under such conditions [52–55].

In a recent paper [56], the SB model has been extended to include LR interactions by introducing a modified free energy involving Local Effective Interaction Parameters (LEIP), which account for the LR interactions in an effective way. The LR force field is thus replaced by a short-ranged effective one. The new local effective parameters, i.e., effective protonation free energies, effective pair interactions and so on, can be systematically calculated by using the Gibbs–Bogoliubov variational principle [39]. The resulting modified free energy converges very fast to the exact free energy. It was found that the correction to the site protonation  $pK$  (first order correction) is enough to obtain an excellent, exact from a practical point of view, agreement between theory and MC simulations. This previous study, however, was restricted to rigid molecules, and conformational degrees of freedom were not taken into account.

The main goal of the present work is to extend the LEIP method to account for the coupling between charge regulation and conformational equilibria involving LR interactions. In addition to allow much faster computations of ionization/conformational properties (computational times are reduced in orders of magnitude), the methodology here presented adds new physical insight in the

interplay of conformational and ionization degrees of freedom in polymeric structures. For instance, the energy of the *gauche* state of a bond will now depend on the pH and the ionic strength, even if such a bond does not hold any ionizable group. The use of the LEIP technique in the SB model is reviewed in Section 2. In Section 3, the technique is generalized in order to include conformational equilibria coupled to LR coulombic interactions represented by the Debye–Hückel potential. In Section 4, semi-grand canonical Monte Carlo simulations are introduced as a tool to test LEIP accuracy when applied to flexible polyelectrolytes. In Section 5, LEIP theoretical results are compared to MC simulations. The new ideas here introduced are illustrated with a minimal model of a flexible linear weak polyelectrolyte containing only one type of rotating bond.

## 2. Simultaneous Treatment of Short- and Long-Range Interactions in Rigid Molecules

The ionization state of a macromolecule with  $N$  ionizable sites can be characterized by a set of variables  $s = \{s_i\}$ ,  $i = 1, \dots, N$ , which can adopt two possible values:  $s_i = 1$  if the site  $i$  is protonated, and  $s_i = 0$  if it is deprotonated. The corresponding reduced free energy can be expressed in terms of the variables  $s_i$  by means of the so-called cluster expansion [24]

$$\frac{H(s)}{\ln 10} = \sum \mu_i s_i + \sum_{i>j} \phi_{ij} s_i s_j + \sum_{i>j>k} \tau_{ijk} s_i s_j s_k + \dots, \quad (1)$$

where  $\mu_i = \text{pH} - \text{p}K_i = -\log(K_i a_H)$  is the reduced chemical potential, which depends on the proton activity,  $a_H$ , and the protonation  $\text{p}K$ -value of the ionizable site  $i$ ,  $\text{p}K_i$ ;  $\phi_{ij}$  represents the interaction energy of the sites  $i$  and  $j$ ;  $\tau_{ijk}$  accounts for possible triplet interactions among sites  $i$ ,  $j$  and  $k$ , and so on. The term “reduced” refers to the fact that the chemical potential incorporates both the pH and the protonation  $\text{p}K$ , which simplifies the subsequent expressions. The interaction (or cluster) parameters are expressed in thermal units, i.e.,  $\beta = 1/k_B T = 1$ , and divided by a factor  $\ln 10$  in order to be compared in the pH scale. Note that the conformation degrees of freedom are omitted in Equation (1), so that the interaction parameters should be understood as proper averages over the conformational states. The mathematical form of these averages is not trivial and expressions for them are given in [45]. Throughout this work, we will assume that a site is charged when it is protonated, i.e., we are dealing with poly-cations. However, the subsequent arguments are also applicable to poly-anions with a suitable change in the protonation variables [15]. The expansion of the free energy (1) usually converges very fast to the exact free energy, and, for most of the cases, the inclusion of triplet interactions is enough to accurately reproduce the measurable quantities, such as the degree of ionization of the individual sites [22]. These can be obtained from  $H(s)$  by means of the semi-grand canonical partition function

$$\Xi = \sum_s e^{-H(s)}. \quad (2)$$

The average degree of protonation of a particular site  $i$  is related to  $\Xi$  as

$$\theta_i = \langle s_i \rangle = -\frac{\partial \log \Xi}{\partial \mu_i} = \frac{1}{\ln 10} \frac{\partial \Omega}{\partial \mu_i}, \quad (3)$$

where  $\Omega = -\ln \Xi$  is the thermodynamic potential associated with the semi-grand canonical ensemble. The average number of bound protons is given by

$$v = \left\langle \sum_i s_i \right\rangle = \frac{\partial \Omega}{\partial \ln a_H}. \quad (4)$$

The correlation of the protonation degrees of two sites  $i$  and  $j$ , a quantity which will be used later, can be expressed as

$$h_{ij} = \langle s_i s_j \rangle = -\frac{1}{(\ln 10)^2} \frac{\partial \Omega}{\partial \mu_i \partial \mu_j}. \quad (5)$$

As can be seen, the quantification of all the relevant physical quantities relies in the accurate determination of the partition function  $\Xi$ . If the number of sites is small ( $N \leq 20$ ),  $\Xi$  can be evaluated by direct enumeration of all the possible ionization states. Otherwise, Monte Carlo (MC) simulations must be performed. In some cases, however, methods borrowed from Statistical Mechanics can be used. Among them, probably the most elegant one is the transfer matrix method, consisting of relating the partition function for a system with  $N + 1$  sites with that with  $N$  sites in a recursive way. This method was firstly used in the exact solution of the Ising model of ferromagnets [38,39]. The link between both partition functions is the transfer matrix whose elements are the Boltzmann factors corresponding to the increase in the reduced free energy. For instance, for the linear polyelectrolyte sketched in Figure 1a and assuming only nearest neighbour interactions, the partition function can be expressed as [47]

$$\Xi = \mathbf{q} \mathbf{T}^N \mathbf{p}^T, \tag{6}$$

where  $\mathbf{T}$  is the transfer matrix

$$\mathbf{T} = \begin{pmatrix} 1 & z \\ 1 & zu \end{pmatrix}. \tag{7}$$

$z = Ka_H$  represents the reduced activity and  $\epsilon = -\log u$  is the interaction free energy between neighbouring sites.  $\mathbf{q} = (1, 0)$  and  $\mathbf{p} = (1, 1)$  are the initiating and terminating vectors. This would be the simplest use of the transfer matrix.

The main limitation of the transfer matrix methods is that they can be only used when Long-Range (LR) interactions are neglected, since the size of the transfer matrices grows exponentially with the range of the interactions [50]. This is an important limitation of the method when dealing with polyelectrolytes, since it can be only used at high enough ionic strengths, for which the screening is enough to avoid the LR interactions. In a recent paper [56], we introduced a method which allows for including the LR interactions in a very accurate way. In this approach, the full free energy Equation (1) is replaced by a new one involving only Short-Range (SR) interaction parameters, accounting for the LR interactions in an effective way. The resulting formalism deals with both SR and LR interactions simultaneously. The method can be used for any kind of molecular or surface geometry, but it is restricted to rigid structures, so that the conformational degrees of freedom are not explicitly taken into account. Since the main goal of this work is to extend this formalism to flexible molecules and polyelectrolytes, we briefly outline the basic ideas of the method. The details of the derivations are given in Reference [56]. Although the following arguments can be readily generalized to the general form of the free energy Equation (1), let us consider the simplest case of a rigid linear chain with identical sites, such as the one sketched in Figure 1a. For this system,  $\mu_1 = \mu_2 = \dots = \mu$  and triplet interactions are omitted, i.e.,  $\tau_{ijk} = 0$ . The reduced free energy  $H$  can be split into two contributions  $H = H_0(x) + \Delta H(x)$  such as

$$\begin{aligned} \frac{H_0}{\ln 10} &= (\mu - x) \sum_i s_i + \epsilon \sum_i s_i s_{i+1}, \\ \frac{\Delta H}{\ln 10} &= \sum_{j>i+1} \phi_{ij} s_i s_j + x \sum_i s_i, \end{aligned} \tag{8}$$

where  $x$  is a parameter to be determined. Note that  $H_0$  corresponds to a reduced free energy containing only nearest neighbour interactions of energy  $\epsilon = \phi_{i,i+1}$ , which can be exactly solved by using the transfer matrix (7). Now, we can use the Gibbs–Bogoliubov variational principle [39,57]

$$\Omega \leq \tilde{\Omega} = \Omega_0(x) + \langle \Delta H(x) \rangle_0 \tag{8}$$

to determine the optimal value of  $x$ , where  $\Omega_0(x) = -\ln \Xi_0$  and  $\langle \dots \rangle_0$  represent the free energy and

the thermal average corresponding to  $H_0$ , respectively. Minimizing  $\tilde{\Omega}$  with respect to  $x$ , it is found that  $x$  fulfills the equation [56]

$$x = \frac{d\varphi_0/dx}{dv_0/dx} = \frac{d\varphi_0}{dv_0}, \quad (9)$$

where

$$\varphi_0 = \left\langle \sum_{j>i+1} \phi_{ij} s_i s_j \right\rangle_0 = \sum_{j>i+1} \phi_{ij} h_{ij}^0 \quad (10)$$

is the LR energy averaged over the unperturbed free energy  $H_0$ , whose correlation function  $h_{ij}^0$ , can be exactly evaluated using (5). If the optimal value for  $x$  is used in the computations, the variational principle (8) implies that all the thermal averages (degree of protonation, correlation functions, etc.) can be obtained replacing the average  $\langle \dots \rangle$  by  $\langle \dots \rangle_0$ , which can be exactly determined since only SR interactions are involved. Equation (9) provides a transparent physical interpretation of  $x$ : it is the average change in the LR interaction energy when a new proton is bound to the molecule at a given pH-value. As expected,  $x$  vanishes in the absence of LR interactions and the nearest neighbour interaction model becomes exact. Therefore,  $x$  can be interpreted as the necessary correction to the reduced chemical potential  $\mu$  in order to account for the LR interactions but in a local effective way. By the definition of  $\mu = \text{pH} - \text{pK}$ ,  $x$  can also be understood as the correction to the site pK-value, so that  $\text{pK}^{\text{eff}} = \text{pK} - x$  is the effective pK-value, and it represents the extra energetic cost of the site protonation due to the presence of LR interactions. We will refer to this procedure as the Local Effective Interaction Parameters (LEIP) method. It is important to highlight that LEIP, unlike other approaches involving some mean-field approximation (such as the Bragg–Williams approximation in Ising models), includes the correlations via Equation (10), although in an approximate way. This approximation, however, results in being extremely accurate, as can be observed in Figure 2a, where the titration curves corresponding to a rigid linear chain with identical sites are depicted. The chosen parameters are  $\text{pK} = 9$  and  $\epsilon = 1.5$ . In this model, the LR interactions between distant sites are described by the Debye–Hückel potential

$$\phi_{ij} = \frac{1}{\ln 10} \frac{\ell_B e^{-\kappa d_{ij}}}{d_{ij}} ; \quad j > i + 1, \quad (11)$$

where  $\ell_B \simeq 0.7$  nm is the Bjerrum length in water at 298 K,  $d_{ij}$  is the distance between the sites  $i$  and  $j$ , and  $\kappa^{-1}$  (nm) =  $0.304/\sqrt{I(\text{M})}$  is the Debye length at the ionic strength  $I$ . For a rigid linear chain, as the one shown in Figure 1,  $d_{ij} = |j - i|b$ , where  $b$  is the separation between consecutive protonating sites. We have plotted the titration curves obtained by Monte Carlo (MC) simulations in the semi-grand canonical ensemble, i.e., at constant pH (blue circles), together with the ones calculated using Equations (9) and (10) (continuous lines) for all the range of ionic strengths and  $b = 0.2$  nm. Surprisingly, simulated and calculated curves overlap, so that, for this model, the LEIP solution can be regarded as exact from the practical point of view. The computational cost of LEIP methods is many orders of magnitude lower than that required in MC simulations, allowing the fitting of parameters to experimental titration curves. The correction to the pK,  $x$ , shown in Figure 2b, increases in lowering the pH (i.e., increasing the charge), and in decreasing the ionic strength (lower electrostatic screening), since the energetic cost to protonate a site increases with the macromolecular charge and with the intensity of the LR interactions.

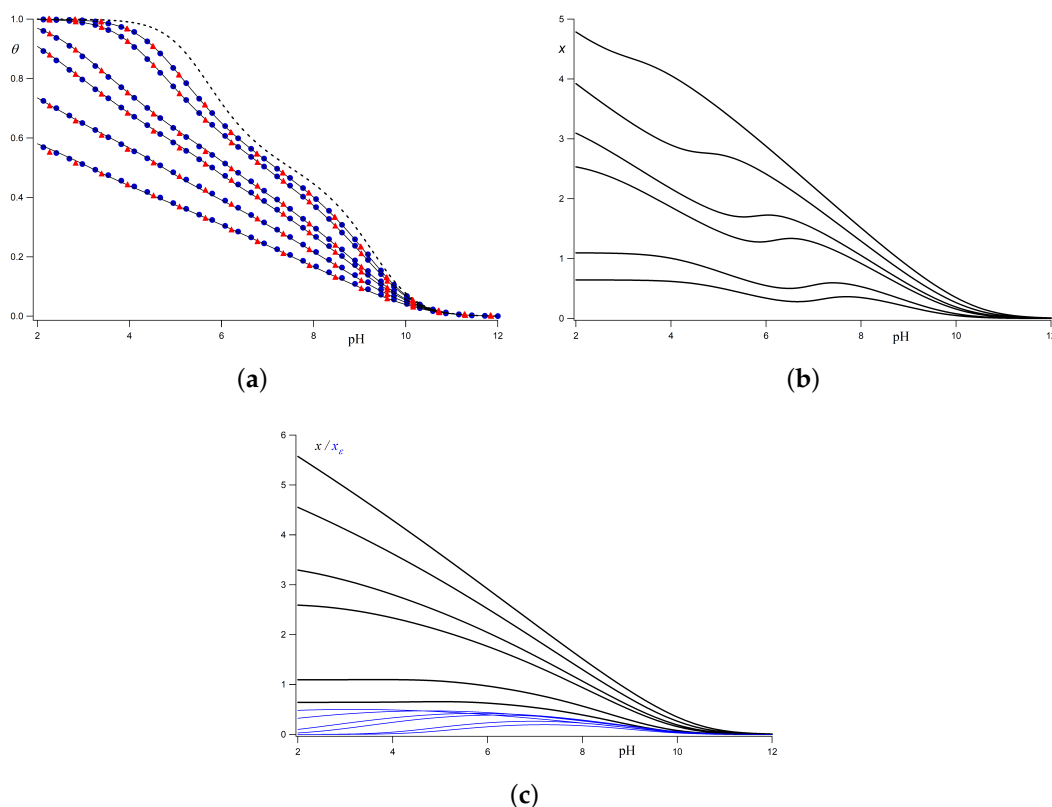
Another advantage of the LEIP method is that it can be systematically improved by selectively correcting other cluster parameters. For instance, one could decide to correct, not only the pK-value

( $pK \rightarrow pK - x$ ), but also the nearest neighbour interaction energy ( $\epsilon \rightarrow \epsilon + x_\epsilon$ ). Proceeding in the same way, it can be shown that  $x$  and  $x_\epsilon$  fulfill the nonlinear system of equations [56]

$$\mathbf{J} \begin{pmatrix} x \\ x_\epsilon \end{pmatrix} = \begin{pmatrix} \left( \frac{\partial \varphi_0}{\partial x} \right)_{x_\epsilon} \\ \left( \frac{\partial \varphi_0}{\partial x_\epsilon} \right)_x \end{pmatrix}; \mathbf{J} = \begin{pmatrix} \left( \frac{\partial v_0}{\partial x} \right)_{x_\epsilon} & \left( \frac{\partial D_0}{\partial x} \right)_{x_\epsilon} \\ \left( \frac{\partial v_0}{\partial x_\epsilon} \right)_x & \left( \frac{\partial D_0}{\partial x_\epsilon} \right)_x \end{pmatrix}, \quad (12)$$

where  $v_0(x, x_\epsilon)$  and  $D_0 = \langle s_i s_{i+1} \rangle_0 = h_{12}^0(x, x_\epsilon)$  represent the average number of protons and the average number of nearest neighbour interactions, respectively, which can be exactly calculated using  $H_0$ . Solving Equation (12), the correction to the  $pK$  and  $\epsilon$  are obtained as functions of the  $pH$ . The physical meaning of  $x$  and  $x_\epsilon$  becomes transparent if Equation (12) are rewritten in terms of  $v_0$  and  $D_0$  as independent variables. After some elementary algebra,  $x$  and  $x_\epsilon$  adopt the much simpler form

$$x = \left( \frac{\partial \varphi_0}{\partial v_0} \right)_{D_0}; \quad x_\epsilon = \left( \frac{\partial \varphi_0}{\partial D_0} \right)_{v_0}. \quad (13)$$



**Figure 2.** (a) Titration curves corresponding to a rigid linear chain with interacting ionizable sites separated by a distance  $b = 0.2$  nm obtained using Monte Carlo (MC) simulations (blue circles), Local Effective Interaction Parameters (LEIP) method correcting only the  $pK$ -value (continuous line) and LEIP method correcting the  $pK$ -value and the nearest neighbour interaction energy  $\epsilon$  (red triangles). The chosen parameters are  $pK = 9$  and  $\epsilon = 1.5$ . The Long-Range (LR) interactions are calculated using the Debye–Hückel potential. The dashed line represents the titration curve in the absence of LR interactions. Note that the correction to the  $pK$ -value is enough to reproduce almost exactly the MC simulations and no significant improvement is obtained in correcting  $\epsilon$ ; (b) correction  $x$  to the  $pK$ -value using the LEIP method; (c) corrections  $x$  (black lines) and  $x_\epsilon$  (blue lines) to the  $pK$ -value and the nearest neighbour interaction energy  $\epsilon$ , respectively. In all the figures, from top to bottom, the ionic strengths are 1 M, 0.5 M, 0.1 M, 0.05 M, 0.01 M and 0.001 M.

Equation (13) states that  $x$  represents the increase in  $\varphi_0$  for a constant number of interactions  $D_0$ , while  $x_\epsilon$  can be interpreted as the change in  $\varphi_0$  in creating a nearest neighbour interaction, keeping constant the number of bound protons  $\nu_0$ . Intuitively, one can guess that  $x_\epsilon$  is much smaller than  $x$ , so that the correction to the  $pK$  is enough to reproduce almost exactly the exact free energy, generating physical properties almost indistinguishable from the MC simulations. In Figure 2a, the titration curves have been recalculated using the correction to  $\epsilon$ . As expected, no significant improvement is obtained.  $x$  and  $x_\epsilon$  as functions of the pH are shown in Figure 2c, where it is clearly observed that  $x_\epsilon$  is much lower than  $x$  for all the ionic strengths. Note that the wavy behaviour of  $x$  in Figure 2b is no longer present in Figure 2c, and seems to be replaced by the contribution  $x_\epsilon$ . Using the same procedure, corrections to higher order interactions, such as triplet or next-nearest neighbour interactions, can be calculated until the desired accuracy is obtained, and expressions of type (13) can be generalized in a straightforward manner. The same treatment leads to very good results for heterogeneous polyelectrolytes and polyampholytes, by correcting the  $pK$ -values of the different kind of sites ( $pK_i \rightarrow pK_i - x_i$ ) [56].

### 3. Coupling of Ionization and Conformational Equilibria

For a linear macromolecule composed by  $M$  bonds, a particular conformational state is denoted by a set of variables  $c = \{c_\alpha\}$ ,  $j = 1, \dots, M$ . The variables  $c_\alpha$  can adopt several values corresponding to the rotational angles of the bond  $\alpha$ . The possible states of the bonds are usually chosen as those of minimum energy, three in the simplest situation: *trans*, *gauche+* and *gauche-*. The selection of a finite number of rotational states instead of working with the full continuous rotational potential greatly simplifies the statistical mechanics treatment, and constitutes the basis of the Rotational Isomeric State (RIS) model, mainly developed by Flory [47]. In the case of linear polymers, the transfer method can be used to determine the conformational partition function  $\Xi_{\text{rot}}$ , which can be expressed as

$$\Xi_{\text{rot}} = \mathbf{q} \mathbf{U}_1 \mathbf{U}_2 \dots \mathbf{U}_{M-1} \mathbf{U}_M \mathbf{p}^T, \quad (14)$$

where  $\mathbf{U}_\alpha$  is the transfer matrix corresponding to the bond  $\alpha$ . For a symmetric chain, for which the states *gauche+* and *gauche-* have the same energy and identical bonds, the transfer matrices are of the form

$$\mathbf{U} = \begin{pmatrix} 1 & \sigma & \sigma \\ 1 & \sigma\psi & \sigma\omega \\ 1 & \sigma\omega & \sigma\psi \end{pmatrix}, \quad (15)$$

where  $\sigma$ ,  $\psi$  and  $\omega$  are the Boltzmann factors associated with the conformational energies of the bonds:  $-k_B T \ln \sigma$  is the free energy of the *gauche* states while  $\psi$  ( $\omega$ ) are related to the interaction energies between two consecutive *gauche* states of different (same) orientation.  $\psi$  and  $\omega$  equate one if the rotation of the bonds is independent.  $\mathbf{q} = (1, 0, 0)$  and  $\mathbf{p} = (1, 1, 1)$  are the initiating and terminating vectors. As in the SB model, the necessary thermal averages can be obtained by performing proper derivatives of the partition function. For instance, the average number of bonds in the *gauche* state is given by [47]

$$g = \frac{\partial \ln \Xi_{\text{rot}}}{\partial \ln \sigma}. \quad (16)$$

The RIS model can be generalized in order to take into account the protonation degrees of freedom. If the macromolecule is in a protonation state  $s$ , the pair  $(s, c)$  defines a *roto-microstate* with reduced free energy

$$F(s, c) = F_{\text{rot}}(c) + F_p(s, c). \quad (17)$$

$F_{\text{rot}}(c)$  is the free energy corresponding to the fully deprotonated state of each conformer, while  $F_p(s, c)$  represents the reduced free energy due to the protonation process, which, for a given conformation, can be expressed as

$$\frac{\beta F_p(s, c)}{\ln 10} = \sum_i \mu_i(c) s_i + \sum_{j>i} \phi_{ij}(c) s_i s_j, \tag{18}$$

where triplet interactions have been neglected. Note that the cluster parameters now depend on the conformational state  $c$ . The reduced free energy Equations (17) and (18) combines the RIS and the SB model and defines the SBRIS model, which allows for studying conformational and ionization properties on the same foot. In recent publications, the SBRIS model has been used to explain conformational transitions in weak linear polyelectrolytes [45] and in the characterization of ionization/conformational properties of linear poly(ethylenimine) [46]. The probability of a specified roto-microstate is given by

$$p(s, c) = \frac{e^{-\beta F(s, c)}}{\Xi_{\text{SBRIS}}}, \tag{19}$$

where the SBRIS partition function  $\Xi_{\text{SBRIS}}$  is defined as

$$\Xi_{\text{SBRIS}} = \sum_{s, c} e^{-\beta F(s, c)}. \tag{20}$$

The SBRIS partition function can alternatively be expressed in the fashion

$$\Xi_{\text{SBRIS}} = \sum_s \Xi_{\text{rot}}(s), \tag{21}$$

where  $\Xi_{\text{rot}}(s)$  denotes the rotational partition function for the macromolecule in a ‘frozen’ binding configuration  $s = \{s_1, s_2, \dots, s_N\}$ .  $\Xi_{\text{rot}}(s)$  can then be calculated as a RIS partition function as in Equation (14), but now decorating the transfer matrices with the suitable binding parameters. The sum over the protonation states can be performed by using proper matricial methods described elsewhere [46]. They are outlined as supplementary information and here we just comment the final results. The SBRIS partition function is obtained by replacing the conformational RIS transfer matrices  $\mathbf{U}$  (Equation (15)) for suitable super-matrices. The rule is that, if a bond is holding at its ends two ionization groups,  $\mathbf{U}$  must be replaced by

$$\mathbf{U} \rightarrow \mathbf{B} = \begin{pmatrix} \mathbf{U} & \mathbf{Uz} \\ \mathbf{U} & \mathbf{Uuz} \end{pmatrix}, \tag{22}$$

where  $\mathbf{u}$  is a diagonal matrix containing the Boltzmann factors corresponding to the short-range interactions

$$\mathbf{u} = \begin{pmatrix} u_t & 0 & 0 \\ 0 & u_g & 0 \\ 0 & 0 & u_g \end{pmatrix}. \tag{23}$$

In matrix (23),  $-k_B T \ln u_t$  and  $-k_B T \ln u_g$  represent the short-range interaction energy between two protonated sites separated by a bond in *trans* and *gauche* conformation, respectively. For the bonds which do not hold ionization sites, the substitution is

$$\mathbf{U} \rightarrow \mathbf{B} = \begin{pmatrix} \mathbf{U} & \mathbf{0} \\ \mathbf{0} & \mathbf{U} \end{pmatrix}. \tag{24}$$

The resulting SBRIS partition function reads

$$\Xi_{\text{SBRIS}} = \mathbf{r} \mathbf{B}_1 \mathbf{B}_2 \dots \mathbf{B}_{M-1} \mathbf{B}_M \mathbf{t}^T, \quad (25)$$

where  $\mathbf{r} = (\mathbf{q} \mathbf{q})$  and  $\mathbf{t} = (\mathbf{p} \mathbf{p})$  are the initiating and terminating vectors, respectively. The average number of bound protons and the bond state probabilities can be again obtained by proper derivatives of Equation (25). It can also be shown that matricial expressions for other physical quantities derived in the context of the RIS model can also be generalized to ionizable molecules by performing suitable substitutions by super-matrices [45]. For instance, there are available matricial expressions for the average square distance between two sites of the chain. These expressions are used in this work to estimate the average distance between charged sites and the corresponding LR interaction energy.

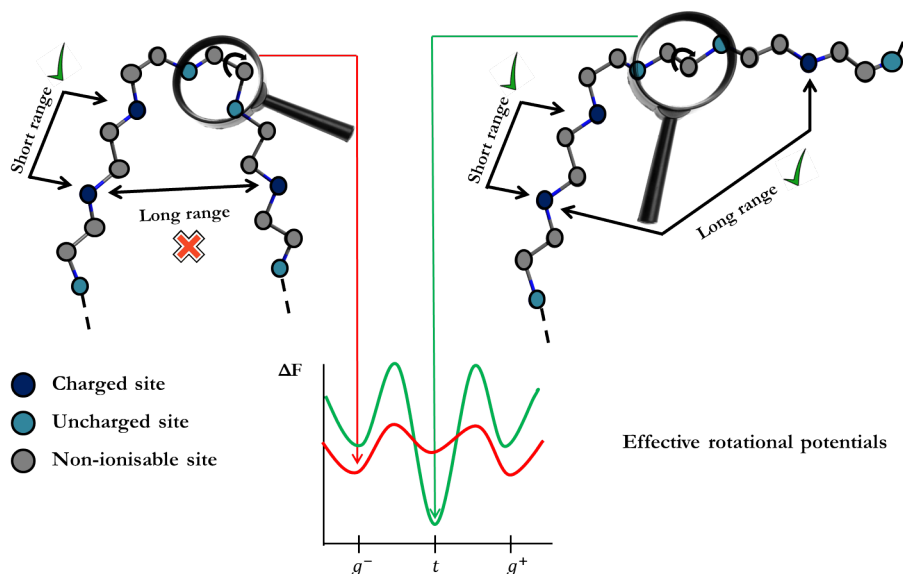
As commented on in the preceding section, transfer matrix methods can only be applied if only SR interactions are taken into account. In this work, we propose to use the LEIP technique to include the LR interactions via local parameters, as done in the case of rigid molecules. Now, however, not only the ionization parameters, such as  $pK \rightarrow pK - x$ , but also the conformational parameters must be corrected as outlined in Figure 3. In the simplest case, with only one kind of rotating bonds, the substitution  $p\sigma \rightarrow p\sigma + x_\sigma$  where  $p\sigma = -\log \sigma$  will be necessary. The treatment is almost identical to the one used for rigid molecules. Now, the “unperturbed” free energy is  $\Omega_0 = -\ln \Xi_{\text{SBRIS}}(x, x_\sigma)$ . It can be easily shown that the corrections  $x$  for the  $pK$  and  $x_\sigma$  for  $p\sigma$  fulfill equivalent equations to (12)

$$\mathbf{J} \begin{pmatrix} x \\ x_\sigma \end{pmatrix} = \begin{pmatrix} \left( \frac{\partial \varphi_0}{\partial x} \right)_{x_\sigma} \\ \left( \frac{\partial \varphi_0}{\partial x_\sigma} \right)_x \end{pmatrix}; \quad \mathbf{J} = \begin{pmatrix} \left( \frac{\partial v_0}{\partial x} \right)_{x_\sigma} & \left( \frac{\partial g_0}{\partial x} \right)_{x_\sigma} \\ \left( \frac{\partial v_0}{\partial x_\sigma} \right)_x & \left( \frac{\partial g_0}{\partial x_\sigma} \right)_x \end{pmatrix}, \quad (26)$$

where  $v_0(x, x_\sigma)$  represents the average number of bound protons (Equation (4)) and  $g_0(x, x_\sigma)$  the average number of bonds in the *gauche* state (Equation (16)), calculated using the unperturbed free energy. If we use  $v_0$  and  $g_0$  as independent variables, instead of  $x$  and  $x_\sigma$ , Equation (26) can be rewritten in a similar fashion as Equation (13)

$$x = \left( \frac{\partial \varphi_0}{\partial v_0} \right)_{g_0}; \quad x_\sigma = \left( \frac{\partial \varphi_0}{\partial g_0} \right)_{v_0}, \quad (27)$$

which essentially tell us that  $x$  represents the average change in  $\varphi_0$  when a new proton is bound (keeping constant the number of bonds in *gauche*) while  $x_\sigma$  is the the average change in  $\varphi_0$  when a bond is brought to its *gauche* state (keeping constant the number of bound protons). Note that the LEIP method always leads to expressions for the interaction corrections of the same type of Equations (13) and (27).



**Figure 3.** Outline of the Local Effective Interaction Parameters (LEIP) method. LEIP accounts for the Short-Range (SR) interactions exactly, but LR interactions are replaced by effective SR free energies. In the resulting scheme, only SR interactions are present, which considerably simplify the theoretical treatment. The bonds “feel” the presence of the LR interactions in an effective way, which leads to an apparent pH-dependent rotational energy. Other parameters, such as the protonation p*K*-values, also become pH-dependent due to the presence of LR interactions.

As in the previous section, LR interactions are described by the Debye–Hückel potential, although the method could in principle be applied to other kind of interactions such as van der Waals interactions. Moreover, by including convenient “hard core” terms in the interaction potentials, the excluded volume effect could in principle be taken into account. The study of this effect, however, is not trivial and it is out of the scope of this work. Unlike rigid molecules, for flexible molecules, the average LR interaction energy  $\varphi_0$  for the unperturbed free energy can only be approximately calculated. In this work, as a first approximation, we assume that

$$\varphi_0 = \left\langle \sum_{ij} \phi(d_{ij}) s_i s_j \right\rangle_0 \simeq \sum_{ij} \phi\left(\sqrt{\langle d_{ij}^2 \rangle_0}\right) \langle s_i s_j \rangle_0. \quad (28)$$

This approximation could in principle be improved by using higher order moments of  $d_{ij}$ . Matricial expressions for  $\langle d_{ij}^2 \rangle_0$  and higher moments were derived by Flory and Jernigan [47,58] for neutral chains. Here, these expressions are modified in order to account for the protonation degrees of freedom. An outline of the derivations is provided as supplementary information.

#### 4. Monte Carlo Simulations

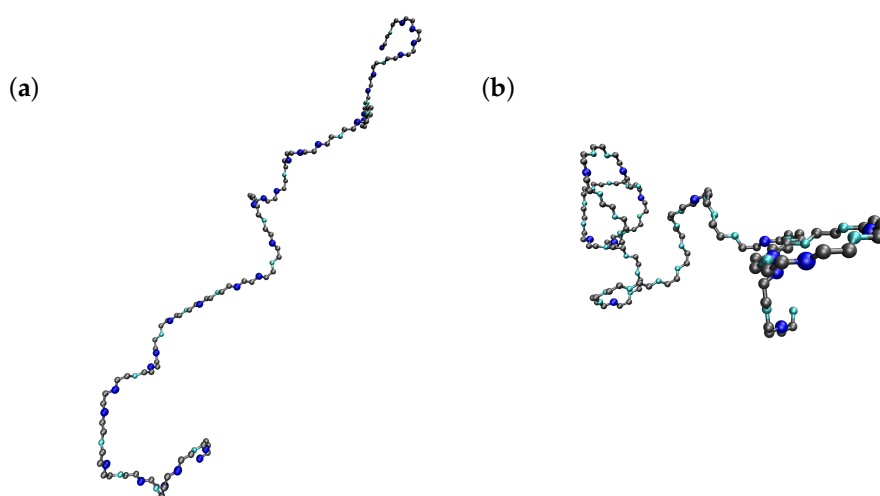
In order to estimate the accuracy of the LEIP method when applied to flexible polyelectrolytes, we compare the theoretical values with those resulting from MC simulations. Two main MC techniques have been previously proposed: the Reaction Ensemble approach, for which the pH is a calculated quantity [59,60], and the constant pH method, corresponding to the semi-grand canonical ensemble [33,61–63]. Since the control variable in the LEIP method is the pH-value, as indicated by the reduced free energies in Equations (1) and (17), the constant pH method has been chosen here. In previous studies about polyelectrolyte ionization properties, both Reaction Ensemble and constant pH methods have been coupled to Molecular Dynamics schemes in order to deal with explicit

ions. In this work, free protons, co- and counter-ions are not explicit in the simulations and the screening effects are taken into account via the Debye length parameter,  $\kappa^{-1}$ . The MC code generalises the one previously used in the computation of conformational and ionization properties of linear poly(ethylenimine) [46]. The polyelectrolyte is modeled as a linear chain with rigid bond lengths and angles. Bonds can adopt one of the three states of minimum energy (*trans*, *gauche+* or *gauche-*). Each change of a bond state implies a 120° rotation of its dihedral angle and the recalculation of distances among the sites situated before and after the rotating bond. The linear chain is composed of interacting nodes which can correspond to inert or protonating groups. In Figure 4, two snapshots of Monte Carlo simulations at ionic strength 0.001 M and two pH-values (four in Figure 4a and eight in Figure 4b) are presented. As in Figures 1 and 3, the ionizable sites are depicted in blue (dark blue if they are protonated and cyan otherwise). It can be observed that a decrease in the pH-value promotes the elongation of the chain, and the consequent reduction of the electrostatic repulsion, by increasing the number of bonds in the *trans* conformations.

In the MC simulations, the free energy of the system is divided into SR and LR terms

$$F(s, c) = F_{\text{SR}} + F_{\text{LR}} + \sum_i (\text{pH} - \text{p}K_i) s_i. \quad (29)$$

The SR term is computed using SBRIS free energy (Equation (17)) which involves the energies present in the transfer matrices ( $\sigma$ ,  $\psi$ ,  $\omega$ ,  $u_t$  and  $u_g$ ), while the LR contribution is calculated using the Debye–Hückel potential (Equation (11)). If  $F_{\text{LR}}$  is set to zero, the obtained results coincide, within the numerical error, with those obtained using the transfer matrix method. This was one of the tests used to check the reliability of the Monte Carlo code. A Metropolis algorithm [15,27] is used to generate roto-microstates at constant pH in a chain with 50 ionizable sites (i.e., 148 nodes or 147 bonds). In each new MC configuration, the polyelectrolyte can change either (i) the conformational state of a rotating bond or (ii) the ionization state of a binding site, with trial probabilities of 0.999 and 0.001, respectively. These trial probabilities allow us to obtain a good equilibration of the conformational structure for each ionization state and the system does not become trapped in local minima. The probability to accept a new configuration is obtained by computing the free energy difference ( $\Delta F(s, c)$ ) between trial and actual conformations. When the state of the bond  $\alpha$  is changed, the following free energy differences must be calculated: (i) the conformational energy of bond  $\alpha$  and its interaction with bonds  $\alpha \pm 1$  (corresponding to the parameters  $\sigma$ ,  $\psi$  and  $\omega$ ); (ii) the electrostatic SR interaction between the two sites bound to  $\alpha$  when they are charged (corresponding to  $u_t$  and  $u_g$ , which depend on the new conformation of  $\alpha$ ); and (iii) the change in the LR Debye–Hückel interaction among sites before and after  $\alpha$ , which involves the recalculation of the distances between the charged sites. On the other hand, a change in the ionization state of a site  $s_i$  implies to recalculate: (i) the reduced chemical potential of the site  $i$  by an amount  $\Delta\mu_i = (\text{pH} - \text{p}K_i)\Delta n$ , where  $\Delta n = \pm 1$  is the variation in the number of protons; (ii) the SR repulsive interaction between  $s_i$  and  $s_{i\pm 1}$ ; and (iii) the LR Debye–Hückel interactions between the trial protonating site and the rest of ionized sites. Once  $\Delta F(s, c)$  is computed, the new configuration is always accepted if  $\Delta F(s, c) < 0$  and accepted with a probability  $\exp(-\beta\Delta F(s, c))$  if  $\Delta F(s, c) > 0$ . The values presented are the average over eight different MC simulations. Each MC simulation has been equilibrated in the first  $5 \times 10^7$  configurations and the thermal averages have been computed in the following  $4.5 \times 10^8$  realizations. The simulations were performed using a parallel code developed in C++ on a 126 CPU cluster. For each pH and ionic strength (one point of the curves), typical jobs were run using 8 CPUs during 1 to 2 h.

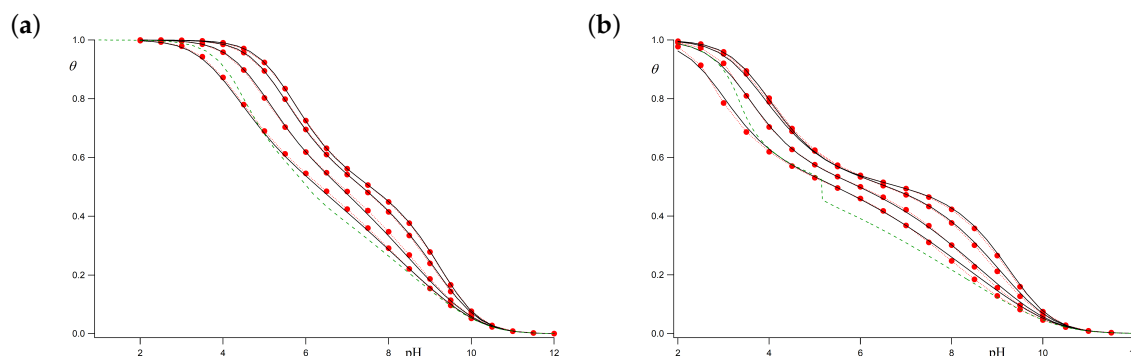


**Figure 4.** Two snapshots of Monte Carlo simulations with  $pK = 9$ ,  $\epsilon_t = 1$ ,  $u_g = 0$ ,  $\sigma = 10$  and pH = 4 (a) and pH = 8 (b). Note that elongated conformations are promoted at low pH as a consequence of polyelectrolyte global charge increase.

## 5. Results and Discussion

As a model system, we use the linear polyelectrolyte outlined in Figure 1b, with protonating sites situated every three chain positions. Only c bonds are allowed to rotate and they can adopt the three states of minimum energy, i.e., *trans*, *gauche+* or *gauche-*. The conformation of c bonds determines the intensity of the SR interactions between neighbouring protonated sites. The rest of bonds (a and b in the figure) are forced to be in the *trans* state. The energy of the *gauche* state of the c bonds is denoted as  $p\sigma = -\log \sigma$ . The c bonds are assumed to rotate independently when the macromolecule is uncharged ( $\omega = \psi = 1$  in Equation (15)). The protonating sites are considered to be identical with the same protonation pK-value. The interactions between protonated sites are characterized by the energies  $\epsilon_t = -\log u_t$  (when the bond c is in *trans* state) and  $\epsilon_g = -\log u_g$  (when the bond c is in *gauche* state). In the computations, the used values of the bond length and the bond angle are 0.2 nm and  $120^\circ$ , respectively. This model can be regarded as a minimal model of a flexible polyelectrolyte, with only four energetic parameters involved ( $\sigma$ ,  $\epsilon_t$ ,  $\epsilon_g$  and pK), and it is here used to illustrate the application of the LEIP method to the analysis of the interplay of conformational and protonation degrees of freedom.

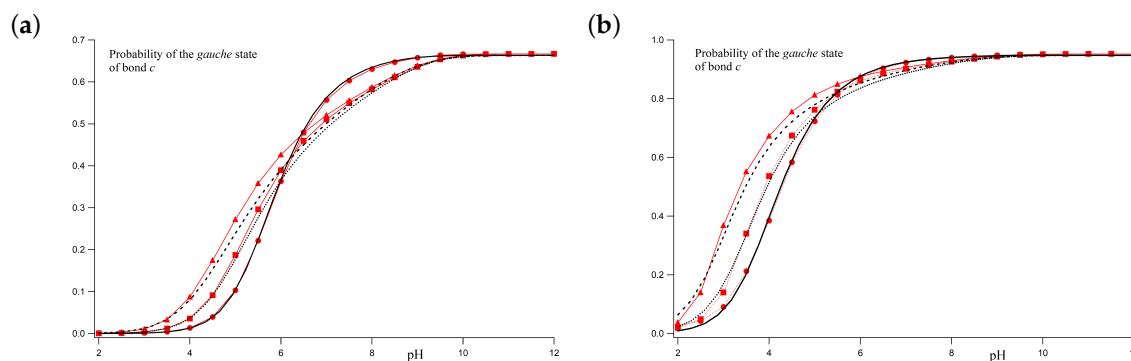
Let us firstly consider the case for which c bonds can freely rotate when the adjacent sites are deprotonated (i.e.,  $\sigma = 1$ ). When both sites are charged, however, the very strong SR repulsion hinders the *gauche* conformation, so that we take  $u_g = 0$  (i.e.,  $\epsilon_g \rightarrow \infty$ ). The resulting titration curves are shown in Figure 5a for ionic strengths ranging from 1 M to 0.001 M. The chosen parameters are pK = 9 and  $\epsilon_t = 1$ . The black continuous lines represent the average protonation degree  $\theta$  calculated using the LEIP method correcting both the pK-value ( $pK \rightarrow pK - x$ ) and the conformational energy of c bonds ( $p\sigma \rightarrow p\sigma + x_\sigma$ ), while the red circles represent the results of the MC simulations. It is observed that the LEIP method reproduces very accurately the MC simulations for all the range of pH-values and ionic strengths. The dashed line depicts the values provided by LEIP for  $I = 0.001$  M if only the pK-value is corrected, while  $\sigma$  remains constant. Although relative good prediction of the MC simulations is obtained, the quality of the titration curve clearly improves if  $\sigma$  is corrected. This means that the rotational energy of the c bonds is affected by LR interactions even if their pendant sites are not charged, as a result of the tendency of the chain to separate the rest of charged groups. Actually, the system behaves as if c bonds “feel” the LR interactions in an effective way. This effect is more remarkable in the subsequent case.



**Figure 5.** Titrations curve for the model polyelectrolyte depicted in Figure 2 with  $pK = 9$ ,  $\epsilon_t = 1$ ,  $u_g = 0$  (i.e.,  $\epsilon_g \rightarrow \infty$ ) and  $\sigma = 1$  (a);  $\sigma = 10$  (b). The chosen ionic strengths are, from top to bottom, 1 M, 0.1 M, 0.01 M and 0.001 M. Black continuous lines represent calculations using the LEIP method in which two effective short-range parameters ( $pK$  and  $p\sigma$ ) has been corrected. Red circles depict the MC values. The green dashed line corresponds to the LEIP values for  $I = 0.001$  M when only the  $pK$ -value is corrected, while  $p\sigma$  is kept constant.

Let us check the accuracy of LEIP method when the *gauche* states of the *c* bonds are favored, for instance, because of the existence of hydrogen bond, which means that  $\sigma > 1$ . When the adjacent sites are both protonated, on the contrary, the electrostatic repulsion is so strong that the *gauche* states are forbidden ( $u_g = 0$ ). In this case, the conformational propensity changes when the ionization state of the sites change. Figure 5b compares the titration curves obtained using LEIP correcting  $pK$  and  $p\sigma$  and MC simulations for  $pK = 9$ ,  $\epsilon_t = 1$  and  $\sigma = 10$ . As can be observed, LEIP method and MC simulations yield to almost identical titration curves. In this case, however, the correction of  $p\sigma$  becomes compulsory. If only  $pK$  is corrected, the titration curve obtained by LEIP at 0.001 M (green dashed line) exhibits a phase transition-like behaviour at  $pH \approx 5$ . This is an artifact resulting from the impossibility to explain the complex interplay of charge regulation and conformational transition without taking into account the influence of LR interactions in the effective energy of the *gauche* state.

For the two cases commented above, the *gauche* state probabilities versus the pH are shown in Figure 6a,b. Markers correspond to MC simulations while black lines represent the theoretical values at ionic strengths 1 M, 0.01 M and 0.001 M, for  $pK = 9$ ,  $\epsilon_t = 1$  and  $u_g = 0$ . Figure 6a corresponds to the case with  $\sigma = 1$ . A good correspondence between simulated and theoretical profiles is obtained for all the ionic strengths. Since at low pH-values, the polyelectrolyte is almost fully protonated, the *gauche* state probability tends to zero because of the high electrostatic repulsion between the nearest charged sites in the *gauche* position ( $u_g = 0$ ). On the other hand, at high pH-values, the macromolecule is completely uncharged and the *c* bonds are freely to rotate. As a result, the probability of the two *gauche* conformers tends to  $2/3$ . For  $I = 1$  M, the LR interactions can be neglected and total correspondence between simulated and calculated values is found. At higher ionic strengths (0.01 M and 0.001 M), for which the Debye–Hückel potential is not screened enough, some small differences arise. However, still now, good agreement between the LEIP method and MC simulations is observed.

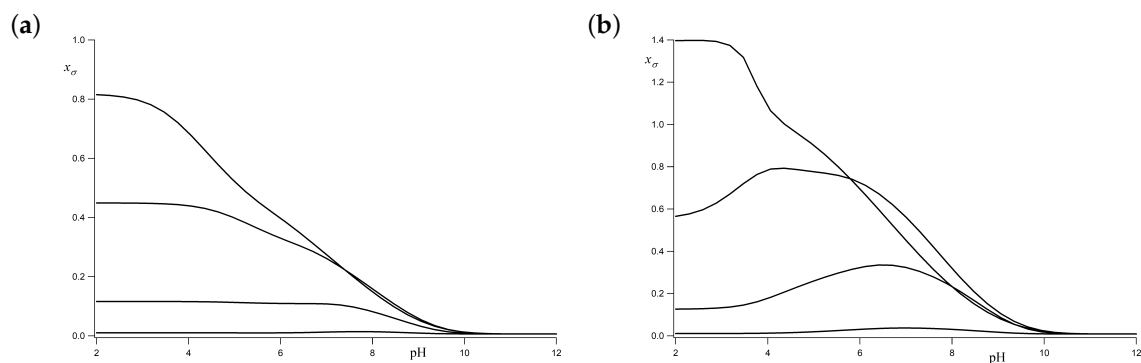


**Figure 6.** *Gauche* state probabilities of bond *c* versus the pH-value computed by means of MC simulations (red markers) and using the LEIP method (black lines) for ionic strengths: 1 M (circles and continuous line), 0.01 M (squares and dotted line) and 0.001 M (triangles and dashed lines). The chosen parameters are  $pK = 9$ ,  $\epsilon_t = 1$ ,  $u_g = 0$  and  $\sigma = 1$  (a);  $\sigma = 10$  (b).

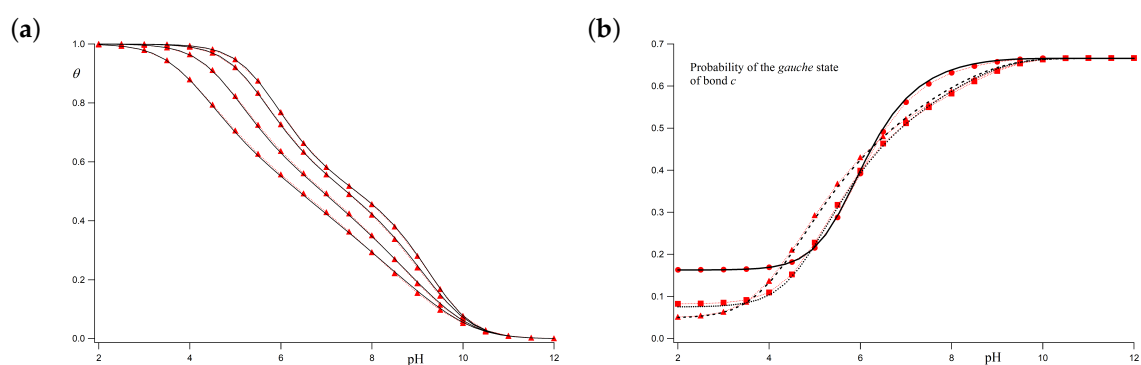
Figure 6b corresponds to the case with  $\sigma = 10$ . Simulated and theoretical probabilities are also in good agreement. Again, the *gauche* state probability tends to zero for low pH-values, while, at high pH-values, the population of the *gauche* conformer is  $2\sigma / (2\sigma + 1) = 0.95$ . A continuous transition from *gauche* to *trans* conformations as the pH decreases is observed. This transition is sharper than in the previous case. Note that, from the LEIP point of view, this transition occurs because of a double effect. On the one hand, there is the charging process, so that two adjacent sites tend to minimize the repulsion when the bonds adopt the *trans* conformation. This effect is present even when LR interactions are not present. On the other hand, the effective *gauche* state energy is increasing due to the average effect of the LR interactions ( $p\sigma \rightarrow p\sigma + x_\sigma$ ). Both effects are important to correctly reproduce the MC simulations. Otherwise, the lack of flexibility in the conformational energy leads to the spurious phase transition observed in Figure 5b (green dashed line).

Figure 7 shows the LEIP method correction  $x_\sigma$  to the bond conformational energy ( $p\sigma \rightarrow p\sigma + x_\sigma$ ) for  $\sigma = 1$  (Figure 7a) and  $\sigma = 10$  (Figure 7b). In both cases, it is observed that  $x_\sigma$  tends to zero at high pH-values, since the molecule is uncharged and no LR interactions are present, so no correction is necessary. As a general tendency,  $x_\sigma$  tends to increase as the pH decreases due to the charging process and the corresponding increase in the LR interactions. This effect is larger at low values of the ionic strength since the Debye–Hückel potential is less screened. For the case  $\sigma = 10$ , a wavy behaviour is observed for ionic strengths 0.1 M and 0.01 M and  $x_\sigma$  exhibits a smooth maximum at  $pH \simeq 4$ , which coincides with the pH regime where the *trans* to *gauche* transition is sharper. This fact could be due to correlations between the rotation of neighbouring bonds or because part of the correction is effectively included in the  $pK$ -correction  $x$ . Further analysis would probably be necessary in order to clarify this point.

In the two cases discussed above, we have taken  $u_g = 0$ , which means that bonds between two protonated sites cannot be in the *gauche* state. Let us now relax this condition and take a finite value for  $u_g$ , so that the electrostatic interaction between two charged sites in *gauche* is not forbidden but only penalized. LEIP predictions (black lines) and MC simulations (red markers) are plotted in Figure 8. Figure 8a shows the computed titration curves with  $\epsilon_g = 2$  at ionic strengths ranging, from top to bottom, from 1 M to 0.001 M. Excellent agreement between the theoretical predictions and simulations is obtained for all the ionic strengths, so that the relaxation of the condition  $u_g = 0$  does not seem to affect the accuracy of the LEIP approach. *Gauche* state probabilities versus pH at three ionic strengths are depicted in Figure 8b: 1 M (circles and continuous line), 0.01 M (squares and dotted line) and 0.001 M (triangles and dashed line). As expected, even at low pH-values, some bonds can remain in the *gauche* state due to the finite value of  $u_g$ . Despite the complexity of the obtained profiles for this case, LEIP is able to accurately reproduce the MC simulations.



**Figure 7.** Correction to the *gauche* state energy  $x_\sigma$  versus the pH for  $pK = 9$ ,  $\epsilon_t = 1$ ,  $u_g = 0$  and  $\sigma = 1$  (a);  $\sigma = 10$  (b). From the bottom to the top, the ionic strengths are 1 M, 0.1 M, 0.01 M and 0.001 M.  $x_\sigma$  represents the average effective energy “felt” by c bonds as a result of the LR interactions.



**Figure 8.** Titration curves (a) and *gauche* state probabilities (b) obtained using the LEIP method (black lines) and MC simulations (red markers). The chosen parameters are  $pK = 9$ ,  $\epsilon_t = 1$ ,  $\epsilon_g = 2$  and  $\sigma = 1$ . (a) the ionic strength are, from top to bottom, 1 M, 0.1 M, 0.01 M and 0.001 M; (b) three different ionic strengths are shown:  $I = 1$  M (circles and continuous line), 0.01 M (squares and dotted line) and 0.001 M (triangles and dashed line).

## 6. Conclusions

The ionization and conformational properties of polyelectrolytes are determined by a combination of Short-Range (SR) and Long-Range (LR) interactions between bonds and ionizable sites. In particular, electrostatic LR interactions can only be neglected at high enough ionic strengths, which is an important limitation for many macromolecular systems of interest. The present work explores the possibility of defining local, short-ranged, interaction parameters which are corrected to account for the LR interactions in an effective way. The new parameters are systematically calculated by using variational methods and equations for them are provided. The resulting approach, the Local Effective Interaction Parameters (LEIP) method, was firstly developed to study the binding properties of rigid polyelectrolytes. In this paper, these ideas are extended to flexible polyelectrolytes, for which conformational and ionization equilibria (charge regulation) are strongly coupled. With this aim, LEIP is combined with the Site Binding Rotational Isomeric State (SBRIS) model in order to deal simultaneously with conformational and protonation degrees of freedom for the full range of ionic strengths.

The LEIP method is illustrated by using a model of a linear symmetric polyelectrolyte containing protonating sites situated regularly along the polymer backbone. The charged sites interact by means of the Debye–Hückel potential, which accounts for the electrostatic screening in an average way, while excluded volume effects are neglected. The bonds linking the ionizable sites can be in three possible states, i.e., *trans*, *gauche+* and *gauche-*. This model with only four relevant parameters

(protonation  $pK$ -value, *gauche* state energy and SR electrostatic interactions between neighbouring sites through bonds in *trans* and *gauche* states) can be regarded as a minimal model of a flexible polyelectrolyte where conformational and binding equilibria are strongly coupled. The LEIP method is applied to correct both the protonation  $pK$ -values and the *gauche* state energy. As a result, local pH dependent rotational potentials are obtained. The correction to the *gauche* energy represents the contribution of the LR interactions in rotating a bond to its *gauche* state.

The degree of protonation and the *gauche* state probabilities obtained using the LEIP method are compared with those computed using semi-grand canonical Monte Carlo (MC) simulations. In all of the studied cases, the agreement between LEIP and MC simulations is excellent. The computational cost, however, is orders of magnitude lower in the LEIP method. This fact allows using LEIP to directly fit parameters to experimental information. The LEIP method could also represent a complementary tool to the study of other aspects of the polyelectrolyte physical chemistry, such as the dependence of the molecular size on the pH, the influence of excluded volume interactions, the presence of attractive hydrophobic interactions or the competitive binding of metal ions. The clarification of these points, which have not been the subject of the present study, would be desirable in order to extend the applicability of the LEIP method. We think that the ideas presented here could be useful in the design of pH-dependent force fields based on experimental ionization and conformational properties.

**Supplementary Materials:** General equations for SBRIS partition function and mean square distance between sites, necessary in the computations of this work, are provided as supplementary information (can be available online at <http://www.mdpi.com/2073-4360/10/08/811/s1>). The transfer matrices corresponding to the model polyelectrolyte used here are also reported.

**Author Contributions:** Conceptualization, F.M. and J.L.G.; Methodology, F.M. and J.L.G.; Software, P.M.B. and S.M.; Validation, P.M.B. and S.M.; Investigation, P.M.B. and J.L.G.; Writing—Original Draft Preparation, P.M.B., S.M., F.M. and J.L.G.; Writing—Review and Editing, P.M.B., S.M., F.M. and J.L.G.; Visualization, P.M.B. and J.L.G.

**Funding:** We acknowledge the financial support from Generalitat de Catalunya (Grants 2017SGR1033, 2017SGR1329 and XrQTC). Josep Lluís Garcés also acknowledges the Spanish Ministry of Science and Innovation (project CTM2016-78798-C2-1-P). Sergio Madurga and Francesc Mas acknowledge the funding of the EU project 8SEWP-HORIZON 2020 (692146). Pablo M. Blanco also acknowledges the financial support from Grant (FI-2017) of Generalitat de Catalunya.

**Conflicts of Interest:** The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

## Abbreviations

The following abbreviations are used in this manuscript:

LEIP	Local Effective Interaction Parameters
LR	Long-Range
MC	Monte Carlo
RIS	Rotational Isomeric State
SB	Site Binding
SBRIS	Site Binding Rotational Isomeric State
SR	Short-Range

## References

1. Wyman, J.; Gill, S.J. *Binding and Linkage*; University Science Books: Mill Valley, CA, USA, 1990.
2. Warshel, A.; Sharma, P.K.; Kato, M.; Parson, W.W. Modeling electrostatic effects in proteins. *BBA Proteins Proteom.* **2006**, *1764*, 1647–1676. [[CrossRef](#)] [[PubMed](#)]
3. Lund, M.; Jönsson, B. Charge regulation in biomolecular solution. *Q. Rev. Biophys.* **2013**, *46*, 265–281. [[CrossRef](#)] [[PubMed](#)]
4. Lipfert, J.; Doniach, S.; Das, R.; Herschlag, D. Understanding Nucleic Acid–Ion Interactions. *Annu. Rev. Biochem.* **2014**, *83*, 813–841. [[CrossRef](#)] [[PubMed](#)]
5. Borkovec, M.; Koper, G.J.M.; Piguët, C. Ion binding to polyelectrolytes. *Curr. Opin. Colloid Interface Sci.* **2006**, *11*, 280–289. [[CrossRef](#)]
6. Hamacek, J.; Borkovec, M.; Piguët, C. Simple thermodynamics for unravelling sophisticated self-assembly processes. *Dalton Trans.* **2006**, *12*, 1473–1490. [[CrossRef](#)] [[PubMed](#)]
7. Li, Y.; Zhao, T.; Wang, C.; Lin, Z.; Huang, G.; Sumer, B.D.; Gao, J. Molecular basis of cooperativity in pH-triggered supramolecular self-assembly. *Nat. Commun.* **2016**, *7*, 13214. [[CrossRef](#)] [[PubMed](#)]
8. Stumm, W.; Morgan, J.J. Aquatic Chemistry. In *Chemical Equilibria and Rates in Natural Waters*, 3rd ed.; Wiley: New York, NY, USA, 1996.
9. Rivas, L.; Pereira, E.D.; Moreno-Villoslada, I. Water-soluble polymer–metal ion interactions. *Prog. Polym. Sci.* **2003**, *28*, 173–208. [[CrossRef](#)]
10. Trefalt, G.; Behrens, S.H.; Borkovec, M. Charge Regulation in the Electrical Double Layer: Ion Adsorption and Surface Interactions. *Langmuir* **2016**, *32*, 380–400. [[CrossRef](#)] [[PubMed](#)]
11. Dobrynin, A.V. Theory and simulations of charged polymers: From solution properties to polymeric nanomaterials. *Curr. Opin. Colloid Interface Sci.* **2008**, *13*, 376–388. [[CrossRef](#)]
12. Chapel, J.P.; Berret, J.F. Versatile electrostatic assembly of nanoparticles and polyelectrolytes: Coating, clustering and layer-by-layer processes. *Curr. Opin. Colloid Interface Sci.* **2012**, *17*, 97–105. [[CrossRef](#)]
13. Carnal, F.; Clavier, A.; Stoll, S. Polypeptide-nanoparticle interactions and corona formation investigated by monte carlo simulations. *Polymers* **2016**, *8*, 203. [[CrossRef](#)]
14. Hartig, S.M.; Greene, R.R.; Dikov, M.M.; Prokop, A.; Davidson, J.M. Multifunctional nanoparticulate polyelectrolyte complexes. *Pharm. Res.* **2007**, *24*, 2353–2369. [[CrossRef](#)] [[PubMed](#)]
15. Borkovec, M.; Jönsson, B.; Koper, G.J.M.J. *Surface and Colloid Science*; Matijevic, E., Ed.; Plenum Press: New York, NY, USA, 2001; Volume 16.
16. Olander, D.S.; Holtzer, A. The Stability of the Polyglutamic Acid alpha Helix. *J. Am. Chem. Soc.* **1968**, *90*, 4549–4560. [[CrossRef](#)] [[PubMed](#)]
17. Uyaver, S.; Seidel, C. First-order conformational transition of annealed polyelectrolytes in a poor solvent. *Europhys. Lett.* **2003**, *64*, 536–542.
18. Whitten, S.T.; García-Moreno, B.; Hilser, V.J. Local conformational fluctuations can modulate the coupling between proton binding and global structural transitions in proteins. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 4282–4287. [[CrossRef](#)] [[PubMed](#)]
19. Das, R.K.; Pappu, R.V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 13392–13397. [[CrossRef](#)] [[PubMed](#)]
20. Das, R.K.; Ruff, K.M.; Pappu, R.V. Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **2015**, *32*, 102–112. [[CrossRef](#)] [[PubMed](#)]
21. Borkovec, M.; Koper, G.J.M. Ising models of polyprotic acids and bases. *J. Phys. Chem.* **1994**, *98*, 6038–6045. [[CrossRef](#)]
22. Borkovec, M.; Koper, G.J.M. A cluster expansion method for the complete resolution of microscopic ionization equilibria from NMR titrations. *Anal. Chem.* **2000**, *72*, 3272–3279. [[CrossRef](#)] [[PubMed](#)]
23. Koper, G.J.M.; Borkovec, M. Binding of metal ions to polyelectrolytes and their oligomeric counterparts: An application of a generalized potts model. *J. Phys. Chem. B* **2001**, *105*, 6666–6674. [[CrossRef](#)]
24. Koper, G.J.M.; Borkovec, M. Proton binding by linear, branched, and hyperbranched polyelectrolytes. *Polymer* **2010**, *51*, 5649–5662. [[CrossRef](#)]
25. Borkovec, M.; Cakara, D.; Koper, G.J.M. Resolution of microscopic protonation enthalpies of polyprotic molecules by means of cluster expansions. *J. Phys. Chem. B* **2012**, *116*, 4300–4309. [[CrossRef](#)] [[PubMed](#)]

26. Cannas, S.A. One-dimensional Ising model with long-range interactions: A renormalization-group treatment. *Phys. Rev. B* **1995**, *52*, 3034–3037. [[CrossRef](#)]
27. Ullner, M.; Jönsson, B. Monte Carlo study of titrating polyelectrolytes in the presence of salt. *Macromolecules* **1996**, *29*, 6645–6655. [[CrossRef](#)]
28. Nishio, T. Monte Carlo studies on potentiometric titration of poly(glutamic acid). *Biophys. Chem.* **1998**, *71*, 173–184. [[CrossRef](#)]
29. Zarzycki, P. Monte Carlo study of the topographic effects on the proton binding at the energetically heterogeneous metal oxide/electrolyte interface. *Langmuir* **2006**, *22*, 11234–11240. [[CrossRef](#)] [[PubMed](#)]
30. Zarzycki, P. Effective adsorption energy distribution function as a new mean-field characteristic of surface heterogeneity in adsorption systems with lateral interactions. *J. Colloid Interface Sci.* **2007**, *311*, 622–627. [[CrossRef](#)] [[PubMed](#)]
31. Ulrich, S.; Seijo, M.; Stoll, S. A Monte Carlo study of weak polyampholytes: Stiffness and primary structure influences on titration curves and chain conformations. *J. Phys. Chem. B* **2007**, *111*, 8459–8467. [[CrossRef](#)] [[PubMed](#)]
32. Companys, E.; Garcés, J.L.; Salvador, J.; Galceran, J.; Puy, J.; Mas, F. Electrostatic and specific binding to macromolecular ligands. A general analytical expression for the Donnan volume. *Colloid Surf. A* **2007**, *306*, 2–13. [[CrossRef](#)]
33. Madurga, S.; Garcés, J.L.; Companys, E.; Rey-Castro, C.; Salvador, J.; Galceran, J.; Puy, J.; Mas, F. Ion binding to polyelectrolytes: Monte Carlo simulations versus classical mean field theories. *Theor. Chem. Acc.* **2009**, *123*, 127–135. [[CrossRef](#)]
34. Garcés, J.L.; Rey-Castro, C.; David, C.; Madurga, S.; Mas, F.; Pastor, I.; Puy, J. Model-independent link between the macroscopic and microscopic descriptions of multidentate macromolecular binding: Relationship between stepwise, intrinsic, and microscopic equilibrium constants. *J. Phys. Chem. B* **2009**, *113*, 15145–15155. [[CrossRef](#)] [[PubMed](#)]
35. Carnal, F.; Ulrich, S.; Stoll, S. Influence of explicit ions on titration curves and conformations of flexible polyelectrolytes: A Monte Carlo study. *Macromolecules* **2010**, *43*, 2544–2553. [[CrossRef](#)]
36. Carnal, F.; Stoll, S. Chain stiffness, salt valency, and concentration influences on titration curves of polyelectrolytes: Monte Carlo simulations. *J. Chem. Phys.* **2011**, *134*, 044909. [[CrossRef](#)] [[PubMed](#)]
37. Aleksandrov, A.; Polydorides, S.; Archontis, G.; Simonson, T. Predicting the Acid/Base Behavior of Proteins: A Constant-pH Monte Carlo Approach with Generalized Born Solvent. *J. Phys. Chem. B* **2010**, *114*, 10634–10648. [[CrossRef](#)] [[PubMed](#)]
38. Baxter, R.J. *Exactly Solved Models in Statistical Mechanics*; Academic Press: London, UK, 1982.
39. Chandler, D. *Introduction to Modern Statistical Mechanics*; Oxford University Press: Oxford, UK, 1987.
40. Koper, G.J.M.; Borkovec, M. Ising models of polyprotic acids and bases. *Langmuir* **1994**, *10*, 2863–2865.
41. Koper, G.J.M.; Borkovec, M. Exact affinity distributions for linear polyampholytes and polyelectrolytes. *J. Chem. Phys.* **1996**, *104*, 4204. [[CrossRef](#)]
42. Vladimir, B.T. General transfer matrix formalism to calculate DNA-protein-drug binding in gene regulation: Application to OR operator of phage  $\lambda$ . *Nucleic Acids Res.* **2007**, *35*, 3519–3867.
43. Lee, A.A.; Kondrat, S.; Kornyshev, A.A. Single-file charge storage in conducting nanopores. *Phys. Rev. Lett.* **2014**, *113*, 048701. [[CrossRef](#)] [[PubMed](#)]
44. Poland, D.; Scheraga, H.A. *Theory of Helix-Coil Transitions in Biopolymers*; Academic Press: New York, NY, USA, 1970.
45. Garcés, J.L.; Koper, G.J.M.; Borkovec, M. Ion binding to polyelectrolytes. *J. Phys. Chem. B* **2006**, *110*, 10937–10950. [[CrossRef](#)] [[PubMed](#)]
46. Garcés, J.L.; Madurga, S.; Borkovec, M. Coupling of conformational and ionization equilibria in linear poly(ethylenimine): A study based on the site binding/rotational isomeric state (SBRIS) model. *Phys. Chem. Chem. Phys.* **2014**, *16*, 4626–4638. [[CrossRef](#)] [[PubMed](#)]
47. Flory, P.L. *Statistical Mechanics of Chain Molecules*; John Wiley: New York, NY, USA, 1969.
48. Flory, P.L. Foundations of Rotational Isomeric State Theory and General Methods for Generating Configurational Averages. *Macromolecules* **1974**, *7*, 381–392. [[CrossRef](#)]
49. Vila, J. Statistical mechanics in homopolypeptides. I. Theoretical approach for titration process in  $\alpha$  helix. *J. Chem Phys.* **1986**, *84*, 6421–6425. [[CrossRef](#)]

50. Reed, C.E.; Reed, W.F. Monte Carlo study of titration of linear polyelectrolytes. *J. Chem. Phys.* **1992**, *96*, 1609–1620. [[CrossRef](#)]
51. Ullner, M. Comments on the Scaling Behavior of Flexible Polyelectrolytes within the Debye-Huckel Approximation. *J. Phys. Chem. B* **2003**, *7*, 8097–8110. [[CrossRef](#)]
52. Alfrey, T.; Morawetz, H. Amphoteric Polyelectrolytes. I. 2-Vinylpyridine-Methacrylic Acid Copolymers *J. Am. Chem. Soc.* **1952**, *74*, 436–438.
53. Mazur, J.; Silberberg, A.; Katchalsky, A. Potentiometric behavior of polyampholytes. *J. Polym. Sci.* **1959**, *35*, 43–70. [[CrossRef](#)]
54. David, C.; Companys, E.; Galceran, J.; Garcés, J.L.; Mas, F.; Rey-Castro, C.; Salvador, J.; Puy, J. Competitive ion complexation to polyelectrolytes: Determination of the stepwise stability constants. The  $\text{Ca}^{2+}/\text{H}^{+}$ /polyacrylate system. *J. Phys. Chem. B* **2007**, *111*, 10421–10430. [[CrossRef](#)] [[PubMed](#)]
55. David, C.; Companys, E.; Galceran, J.; Garcés, J.L.; Mas, F.; Rey-Castro, C.; Salvador, J.; Puy, J. Competitive  $\text{Cd}^{2+}/\text{H}^{+}$  complexation to polyacrylic acid described by the stepwise and intrinsic stability constants. *J. Phys. Chem. B* **2008**, *112*, 10092–10100. [[CrossRef](#)] [[PubMed](#)]
56. Garcés, J.L.; Madurga, S.; Rey-Castro, C.; Mas, F. Dealing with long-range interactions in the determination of polyelectrolyte ionization properties. Extension of the transfer matrix formalism to the full range of ionic strengths. *J. Polym. Sci. Part B* **2017**, *55*, 275–284. [[CrossRef](#)]
57. Burak, Y.; Netz, R.R. Charge Regulation of Interacting Weak Polyelectrolytes. *J. Phys. Chem. B* **2004**, *108*, 4840–4849. [[CrossRef](#)]
58. Flory, P.J.; Jernigan, R.L. Second and fourth moments of chain molecules. *J. Chem. Phys.* **1965**, *42*, 3509–3519. [[CrossRef](#)]
59. Landsgesell, J.; Holm, C.; Smiatek, J. Wang-Landau Reaction Ensemble Method: Simulation of Weak Polyelectrolytes and General Acid-Base Reactions. *J. Chem. Theory Comput.* **2016**, *13*, 852–862. [[CrossRef](#)] [[PubMed](#)]
60. Landsgesell, J.; Holm, C.; Smiatek, J. Simulation of weak polyelectrolytes: A comparison between the constant pH and the reaction ensemble method. *Eur. Phys. J. Spec. Top.* **2017**, *226*, 725–736. [[CrossRef](#)]
61. Baptista, A.M.; Teixeira, V.H.; Soares, C.M. Constant-pH molecular dynamics using stochastic titration. *J. Chem. Phys.* **2002**, *117*, 4184–4200. [[CrossRef](#)]
62. Mongan, J.; Case, D.A.; McCammon, J.A. Constant pH Molecular Dynamics in Generalized Born Implicit Solvent. *J. Comput. Chem.* **2004**, *25*, 2038–2048.
63. Madurga, S.; Rey-Castro, C.; Pastor, I.; Vilaseca, E.; David, C.; Garcés, J.L.; Puy, J.; Mas, F. A semi-grand canonical Monte Carlo simulation model for ion binding to ionizable surfaces: Proton binding of carboxylated latex particles as a case study. *J. Chem. Phys.* **2011**, *135*, 1–11. [[CrossRef](#)] [[PubMed](#)]



Supplementary information to “Coupling of charge regulation and conformational equilibria in linear weak polyelectrolytes: treatment of long range interactions *via* effective short-ranged and pH-dependent interaction parameters”

July 23, 2018

## 1 Matricial expression for the SBRIS partition function for a linear polyelectrolyte

Let us express the SBRIS partition function as

$$\Xi_{\text{SBRIS}} = \sum_s \Xi_{\text{rot}}(s) \quad (1)$$

where  $\Xi_{\text{rot}}(s)$  is the partition function for the molecule in a specific binding state  $s = \{s_1, s_2, \dots, s_N\}$ . For simplicity, let us suppose that all the  $N-1$  bonds hold two protonating sites at their ends.

The partition function for every 'frozen' binding configuration can then be expressed as a the RIS partition function but decorating the transfer matrices with suitable binding parameters.  $\Xi_{\text{rot}}(s)$  adopts the form [1, 2, 3]

$$\Xi_{\text{rot}}(s) = \mathbf{p} \prod_{i=1}^{N-1} \mathbf{U}_i^{s_i s_{i+1}} \mathbf{q}^T \quad (2)$$

In the simplest case  $\mathbf{U}_i^{00} = \mathbf{U}_i^{10}$ ;  $\mathbf{U}_i^{01} = z \cdot \mathbf{U}_i$  and  $\mathbf{U}_i^{11} = z \cdot \mathbf{U}_i \mathbf{u}_i$  where  $\mathbf{U}_i$  are the transfer matrices typical of the RIS model for a symmetric polymer

$$\mathbf{U}_i = \begin{pmatrix} 1 & \sigma & \sigma \\ 1 & \sigma\psi & \sigma\omega \\ 1 & \sigma\omega & \sigma\psi \end{pmatrix}_i \quad (3)$$

where  $z$  is the reduced activity of the site and  $\mathbf{u}$  is a diagonal matrix containing the Boltzmann factors corresponding to the short range interactions between

charged sites

$$\mathbf{u}_i = \begin{pmatrix} u_t & 0 & 0 \\ 0 & u_g & 0 \\ 0 & 0 & u_g \end{pmatrix}_i \quad (4)$$

$-k_B T \ln u_t$  and  $-k_B T \ln u_g$  represent the short range interaction energy between two sites separated by a bond in *trans* and *gauche* conformation, respectively. The next step is to calculate the sum in Eqn. (1), which can be done using the identity

$$\sum_s \left( \prod_{i=1}^{N-1} \mathbf{U}^{s_i s_{i+1}} \right) = (\mathbf{E} \mathbf{E}) \prod_{i=1}^{N-1} \begin{pmatrix} \mathbf{U}_i^{00} & \mathbf{U}_i^{01} \\ \mathbf{U}_i^{10} & \mathbf{U}_i^{11} \end{pmatrix} \begin{pmatrix} \mathbf{E} \\ \mathbf{E} \end{pmatrix} \quad (5)$$

where  $\mathbf{E}$  is the  $3 \times 3$  identity matrix. Combining Eqns. (5), (1) and (2), we obtain the SBRIS partition function.

$$\Xi_{\text{SBRIS}} = \mathbf{s} \prod_{i=1}^{N-1} \begin{pmatrix} \mathbf{U}_i^{00} & \mathbf{U}_i^{01} \\ \mathbf{U}_i^{10} & \mathbf{U}_i^{11} \end{pmatrix} \mathbf{t}^T \quad (6)$$

where  $\mathbf{s} = (\mathbf{p} \mathbf{p})$  and  $\mathbf{t} = (\mathbf{q} \mathbf{q})$ . Note that the SBRIS partition function is obtained from the RIS partition function by replacing

$$\mathbf{U} \rightarrow \begin{pmatrix} \mathbf{U} & \mathbf{U}z \\ \mathbf{U} & \mathbf{U}z \end{pmatrix}; \quad \mathbf{p} \rightarrow (\mathbf{p} \mathbf{p}); \quad \mathbf{q} \rightarrow (\mathbf{q} \mathbf{q}) \quad (7)$$

It can be easily shown that if some of the matrices in Eqn. (2) does not depend on any index  $s_i$ , the proper substitution is

$$\mathbf{U} \rightarrow \begin{pmatrix} \mathbf{U} & \mathbf{0} \\ \mathbf{0} & \mathbf{U} \end{pmatrix} \quad (8)$$

which can be necessary for the transfer matrices corresponding to bonds which do not hold any ionizable site.

## 2 Calculation of the mean square distance between two nodes of the chain

The matrix summation trick used above can also be applied to calculate other physical quantities which become pH-dependent in the SBRIS approach. Let  $f(c, s)$  be any quantity and  $\langle f(c, s) \rangle$  its SBRIS thermal average. Then

$$\langle f(s, c) \rangle = \sum_{s,c} f(s, c) \frac{e^{-\beta F(s,c)}}{\Xi_{\text{SBRIS}}} = \sum_s \pi(s) \left( \sum_c \frac{f(s,c) e^{-\beta F(s,c)}}{\Xi_{\text{rot}}(s)} \right) = \sum_s \pi(s) \langle f(c, s) \rangle_c \quad (9)$$

where

$$\pi(s) = \frac{\Xi_{\text{rot}}(s)}{\Xi_{\text{SBRIS}}} \quad (10)$$

represents the probability of a protonation state  $s$  and  $\langle f(c, s) \rangle_c$  is the average for that fixed protonation state. Note now that the quantity between brackets in Eqn. (9) represents a typical RIS average which can be calculated using the (conveniently decorated) transfer matrices  $\mathbf{U}_i^{s_i s_{i+1}}$ . In particular, matricial expressions for the mean square distance between two nodes,  $k$  and  $l$ , of a linear chain are available [1, 4]. For a fixed ionization state, the result is

$$\langle d_{kl}^2 \rangle (s) = \frac{2}{\Xi_{\text{rot}}(s)} \mathbf{p} \prod_{i=1}^k \mathbf{U}_i^{s_i s_{i+1}} [\mathbf{E} \mathbf{0} \mathbf{0} \mathbf{0} \mathbf{0}] \prod_{r=k+1}^l \mathbf{G}_r^{s_r s_{r+1}} \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{E} \end{bmatrix} \prod_{j=l+1}^{N-1} \mathbf{U}_j^{s_j s_{j+1}} \mathbf{q} \quad (11)$$

where the matrices  $\mathbf{G}$  are proper super-matrices which can be expressed in terms of the transfer matrices  $\mathbf{U}_i^{s_i s_{i+1}}$ , the translation matrices and the bond vectors. The details and derivations are given in the Chapter 4 of the classical Flory's book [1] (see Eqn. 35). Introducing (11) in (9) and using the summation (5),  $\langle d_{kl}^2 \rangle$  is calculated as a function of the pH.

### 3 Minimal symmetric polyelectrolyte model

The model used in this work to illustrate the LEIP method consists of a polyelectrolyte like the one in Fig. 2b. In this molecule the ionizable sites are separated by three bonds **a**, **b** and **c**. The structure is very similar to the one of poly(oxyethylene) and the corresponding transfer matrices are given can be found in [1, 3]

$$\mathbf{U}_a = \begin{pmatrix} 1 & \sigma_a & \sigma_a \\ 1 & \sigma_a \psi & \sigma_a \omega \\ 1 & \sigma_a \omega & \sigma_a \psi \end{pmatrix}; \quad \mathbf{U}_b = \begin{pmatrix} 1 & \sigma_a & \sigma_a \\ 1 & \sigma_a \psi' & \sigma_a \omega' \\ 1 & \sigma_a \omega' & \sigma_a \psi' \end{pmatrix}; \quad \mathbf{U}_c = \begin{pmatrix} 1 & \sigma & \sigma \\ 1 & \sigma \psi & \sigma \omega \\ 1 & \sigma \omega & \sigma \psi \end{pmatrix}$$

The conformational model here used can be obtained by taking  $\sigma_a = 0$ , which assures that the **a** and **b** bonds are always in the *trans* state, and  $\omega = \psi = \omega' = \psi' = 1$ , which means that the bonds of the deprotonated molecule are independent.

### References

- [1] Flory, P. L. *Statistical Mechanics of Chain Molecules*; John Wiley: New York, **1969**
- [2] Garcés J. L.; Koper G. J. M.; Borkovec M. *J. Phys. Chem. B* **2006**, *110*, 10937-10950
- [3] Garcés J. L.; Madurga S.; Borkovec M. *Phys. Chem. Chem. Phys.*, **2014**, *16*, 4626-4638

[4] Flory P. J.; Jernigan R. L. *J. Chem. Phys.* **1965**, 42, 3509-3519