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

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

Received January 7, 2011
Revised February 7, 2011
Accepted February 10, 2011

Research Article

Global conformations of proteins as predicted from the modeling of their CZE mobility data

Estimations of protein global conformations in well-specified physicochemical microenvironments are obtained through global structural parameters defined from polypeptide-scale analyses. For this purpose protein electrophoretic mobility data must be interpreted through a physicochemical CZE model to obtain estimates of protein equivalent hydrodynamic radius, effective and total charge numbers, hydration, actual ionizing pK and pH-near molecule. The electrical permittivity of protein domain is also required. In this framework, the solvent drag on proteins is obtained via the characteristic friction power coefficient associated with the number of amino acid residues defining the global chain conformation in solution. Also, the packing dimension related to the spatial distribution of amino acid residues within the protein domain is evaluated and discussed. These scaling coefficients together with the effective and total charge number fractions of proteins provide relevant interpretations of protein global conformations mainly from collapsed globule to hybrid chain regimes. Also, protein transport properties may be estimated within this framework. In this regard, the central role played by the friction power coefficient in the evaluation of these properties is highlighted.

Keywords:
Chain packing dimension / Effective protein charge number fraction / Electrophoretic mobility modeling / Friction power coefficient / Protein global conformations
DOI 10.1002/elps.201100016

1 Introduction

At present CZE is a successful analytical technique for the separation of proteins, peptides and amino acids (see for instance [1–7] and the references therein). Also, the combination of CZE with MS and other complementary apparatuses are valuable tools for these purposes ([7–10] and the references therein). In this regard, one expects that this field will be increasing in the upcoming years with high benefits for long-term research programs involving both the characterizations of biomacromolecules and the elucidation of their biological functional properties. Within this framework, there are other CZE models for polypeptides, covering different aspects and phenomena with varied degrees of complexity. These model-types interpreting the effective electrophoretic mobility of polypeptides and amino acids provide additional physicochemical properties and average chain structural parameters helping one to visualize global conformations [11–28] and their possible destabilization paths ([29] and the references therein). In this regard, the present work shows specifically the relevance of extracting useful information from effective CZE mobility data. In fact, it is well recognized in the literature [27, 30–33] the need to provide global structural parameters of polypeptides (for instance, global shape descriptors and transport properties) apart from their local structural data usually obtained through involved experimental techniques determining the occurrence of specific secondary structural motifs (α-helixes, β-strands, reversed turns, random coils (RCs), etc.). Therefore, the definitions of protein global structural parameters must be related to the presence of possible secondary conformations, which are in turn a result of the interaction among amino acid residues within the solvent microenvironment, where these polyampholytic chains are immersed [27, 34–36].

From these physical aspects, it is also important to visualize how globular proteins follow folding mechanisms toward quite compact conformations [37], which are the preferred spatial distributions of amino acid residues of

Correspondence: Dr. Julio A. Deiber, INTEC, Güemes 3450, S3000GLN, Santa Fe, Argentina
E-mail: treoflu@santafe-conicet.gov.ar
Fax: +54-342-4550944

Abbreviations: **AAS,** amino acid sequence; **BCA,** bovine carbonic anhydrase; **CG,** collapsed globule; **HC,** hybrid chain; **HCA,** human carbonic anhydrase; **HZ,** hybrid zone; **PE,** polyelectrolyte; **PLLCEM,** perturbed Linderstrøm–Lang capillary electrophoresis model; **RC,** random coil; **STN,** Staphylococcal nuclease

these native chains in protocol solvents. Basically, the understanding of driving forces responsible for global chain conformations under different physicochemical conditions provides information to elucidate more basic questions [37, 38]. In this regard, it is also clear that polypeptides in solution are electrically charged hetero-chains (N amino acid residues randomly distributed in a chain building up a biomacromolecule) with interplays among different side groups involving hydrophobic, dispersion and electrostatic forces, within a hydrated microenvironment of weak bonded water molecules [27].

The present work is concerned with the estimation of some global structural parameters of proteins defined here and in [26], which then are used to study the global conformations of these analytes at different pHs, together with their possible destabilization paths when the pH varies. These results require the knowledge of basic physicochemical properties quantified via the modeling of the effective CZE mobility. Therefore, here these protein properties in well-defined solvents are obtained via the simple Perturbed Linderstrøm–Lang Capillary Electrophoresis Model (PLLCEM) [17, 26]. In this regard, it is crucial to place emphasis on the fact that the value of polypeptide electrophoretic mobility is a powerful information, allowing interpretations that concern to the high phenomenological coupling among chain–solvent friction, intra-molecular hydrodynamic interaction and electrostatic charge interactions of ionizing groups, within the whole hydrodynamic polypeptide domain. Although there are other conventional experimental techniques (for instance, measurements of sedimentation and diffusion coefficients and intrinsic viscosity) providing relevant data and results, one finds that their associated models [39] describe a rather weak coupling between these phenomena, thus limiting further characterizations of polypeptide chains mainly from the electrostatic point of view.

The present work is organized as follows. Section 2 describes briefly the main PLLCEM numerical outputs that are useful to calculate global structural parameters of the proteins considered here. These parameters are employed then in Section 3 to discuss polypeptide chain characterizations and conformations with specific case-studies involving three globular proteins analyzed previously via CZE: bovine carbonic anhydrase (BCA) [40], human carbonic anhydrase (HCA) [41]; Staphylococcal nuclease (STN) [42]. Also, estimations of transport properties such as the average friction, sedimentation and diffusion coefficients together with additional calculations are provided in Supporting Information of the present work. Needs for further research are also proposed in order to deepen the development of global structural parameters that are useful in estimating protein global conformations.

2 Theoretical considerations

The estimations of polypeptide physicochemical properties required here are carried out through data of the effective electrophoretic mobility μ_p of proteins for running CZE protocols with well-specified solvent bulk-pH, ionic strength I, temperature T, electrical permittivity ε and viscosity η. Also, protein molar mass M and amino acid sequence (AAS) of N amino acid residues with their protein data bank (PDB) inter-atomic distances r_{ij} are considered initial data. Therefore, the PLLCEM is used for these purposes [17, 26]. The description and development of this model for polypeptides in general may be found in details with the appropriate hypotheses and some limitations in [17, 18, 22, 23, 26, 27, 29] (see also Supporting Information of the present work).

The model used here is based on well-established physicochemical theories, and can also predict polypeptide chain properties within the classical and quite linear experimentation ranges of CZE runs used for these analytes, where some simplifying hypotheses apply [12, 16, 26]. In this regard, the PLLCEM provides the electrical state of a given analyte [23, 26] through total positive Z_+, total negative Z_-, effective $Z = |Z_+| - |Z_-|$ and total $Z_T = |Z_+| + |Z_-|$ charge numbers. Ionizing groups are confined in the total hydrodynamic volume $V_H = 4\pi a_H^3/3$, where the Stokes or equivalent hydrodynamic radius a_H is introduced. This volume is decomposed into the protein compact volume $V_c = M\nu_p/N_A = 4\pi a_c^3/3$ with equivalent compact radius a_c and the hydration volume V_w. Here, N_A is the Avogadro constant and ν_p is the average protein-specific volume calculated through $\nu_p = \sum_{i=1}^{N} M_i\nu_i/M$[26], where ν_i is the specific volume and M_i is the molar mass of each amino acid residue composing the AAS. An estimate of protein hydration δ (water mass/protein mass) is obtained from $\delta \approx [(a_H/a_c)^3 - 1](\nu_p/\nu_w)$[18], where ν_w is the specific volume of the protocol solvent. Also, the protein hydration number $H = \delta M/18$ (number of water molecules per chain) are obtained by summing each hydration number of ionizing, polar and non-polar groups at the local pH (see details in [22]). These last expressions establish one of the PLLCEM convergence criteria (see also [26] and Supporting Information). Additional water molecules designated H_d are considered when the protein is denatured in some degrees as explained and discussed in [26]. Also, this model provides the pK_i values of ionizing groups yielding a shift ΔpK_i in the reference pK_i^r reported in [43] as a result of the charge regulation phenomenon present in these charged macromolecules [17, 44, 45]. Further, the pH-microenvironment pH_i around the i-ionizing group is also accounted, which is a function of the mean field near-molecule pH, designated here pH^* [17]. The electrical permittivity ε' within protein domain [26] is estimated from the expression $\varepsilon' = \varepsilon\delta/(1+\delta) + \varepsilon_p/(1+\delta)$, which uses the protein hydration as the weighing parameter between protein ε_p [46] and solvent ε electrical permittivities. Another relevant property provided by the PLLCEM is the shape-orientation factor $\Omega = 6\pi\eta a_H/f$, which is the ratio $\Omega = \mu_p/\mu$ between the effective analyte mobility μ_p and the effective mobility of the equivalent sphere $\mu = eZf_H(\kappa a_H)/6\pi\eta a_H(1+\kappa a_H)$ [26]. In these expressions, e is the elementary charge, $f_H(\kappa a_H)$ is

Henry's function [47], f is the chain–solvent friction coefficient, $\kappa = \sqrt{2e^2 N_A I 10^3/\varepsilon k_B T}$ is the inverse of the particle screening length and k_B is the Boltzmann constant. It is clear that the shape-orientation factor Ω [18] associated with the particle shape and orientation density distribution function of main axes may be further modeled, for instance, through equations providing the effective mobility of cylindrical or spheroidal particles [18, 23, 26, 48, 49]. In this particular aspect, the protein friction in the electrophoretic movement is represented through a particle that preserves the hydrodynamic volume $V_H = V_c + V_w$ having also the right f value achieved through the characteristic particle shape. Within the PLLCEM framework, the basic characterization of proteins is obtained, as a first approximation, directly with parameters Ω and a_H [26] and the charge state defined through Z and Z_T, allowing then the application of an appropriate chain or particle model that is compatible with the average value of the protein–solvent friction coefficient (the average orientation of the main particle axis is included in Ω [23]). Here in particular, we express the solvent drag on a protein through the friction power coefficient defined below [27], to account intra-chain hydrodynamic interaction in the domain of the global chain conformation.

Therefore from the above framework, several protein global structural parameters may be studied (see also [26, 27]). The analysis of chain packing within the hydrodynamic domain defined through a_H requires the consideration of the average monomer radius $a_o = \sum_{i=1}^{N} a_i/N$[26] where different amino acid residues, each one having a radius $a_i = \{3v_i M_i/(4\pi N_A)\}^3$, are considered. Consequently, the chain packing dimension g_p is expressed [26],

$$g_p = \log N/\log(a_H/a_o) \leq 3 \quad (1)$$

Equation (1) indicates that for $g_p = 3$ the protein has the compact volume, with $\delta = 0$ and $a_H = a_c$. Thus in general, g_p carries in some degree condensed structural information for further analysis, mainly in relation to the spatial distributions of chain amino acid residues, like those resembling rod-, disk- and spherical-like particles, as the value of this parameter increases from around 1.7 to 2.8 [26].

The average protein–solvent friction coefficient is $f = 6\pi\eta a_o N^{g_f}$, where $g_f \leq 1$ is the corresponding friction power coefficient [27, 50]. The maximum friction is achieved for $g_f = 1$, when a free draining chain in creeping flow with negligible intra-chain hydrodynamic interaction is considered [27, 50]. Also $1/3 < g_f < 1$ when hydrodynamic interaction among chain units is present [27, 51, 52]. These limit values for both the neutral homopolymer-chains and the polypeptide heteropolymer-chains (proteins) are obtained through different mechanisms. In the former chains, they relate mainly to the effect of solvent quality and temperature, while in the later ones, they are due significantly to the electrical state of particles for a given solvent and temperature [27, 51–53]. From Ω and f, one readily obtains (see also Supporting Information),

$$g_f = \frac{1}{g_p} - \frac{\log\Omega}{\log N} \quad (2)$$

This equation indicates that for spherical particles, and to a good approximation for aspherical particles satisfying a random flight orientation distribution function of their major axes ($\Omega \approx 1$) [23, 26, 27] the power friction coefficient is $g_f \approx 1/g_p$, which are particular cases and the less frequent ones, as discussed in [27]. For ideal chains $g_f \approx 1/2$, while for globular polypeptides when fluctuation-attractive electrostatic interactions are relevant (or when the solvent is poor for neutral chains) $g_f \approx 1/3$ [27, 51, 52] yielding a Collapsed Globule (CG) conformation. On the other side, for self-avoiding random chains $g_f \approx 3/5$.

From the PLLCEM one can also evaluate the protein electrical state through additional global structural properties like the effective $\Delta\sigma = |Z|/N$ and the total $\sigma = Z_T/N$ charge number fractions. The importance of these global properties is that for chains in dilute solution they satisfy simple scaling relationships [27, 51, 53–55] suggesting four possible global chain conformations for polyampholytes in general (proteins in this work). Therefore, the following ranges of coordinates $\Delta\sigma$ and σ delimiting different chain conformational regimes may be obtained [27, 54, 55] (see the schematic divisions of regimes illustrated in Fig. 1 for more details). The RC regime implies $\Delta\sigma < 1/\sqrt{u}N^{(1-g_f/2)}$ and $\sigma < 1/uN^{(1-g_f)}$, which may destabilize into the polyelectrolyte (PE) regime when $\Delta\sigma > 1/\sqrt{u}N^{(1-g_f/2)}$. Alternatively, the RC regime destabilizes forming a CG due to fluctuating-attractive electrostatic forces when $\sigma > 1/uN^{(1-g_f)}$. Another relevant chain destabilization is the transition from CG regime to the hybrid chain (HC) regime [27], where $\sqrt{\sigma/N} < \Delta\sigma < u^x\sigma^y$ for $\sigma > 1/uN^{(1-g_f)}$, with $x = 1/2(1-g_f)$ and $y = (1-g_f/2)/(1-g_f)$. This regime is also designated "necklace" in the literature (mainly for polyampholytic homochains) because its basic conformation is composed by hybrid zones (HZs) connected by strings, each one involving

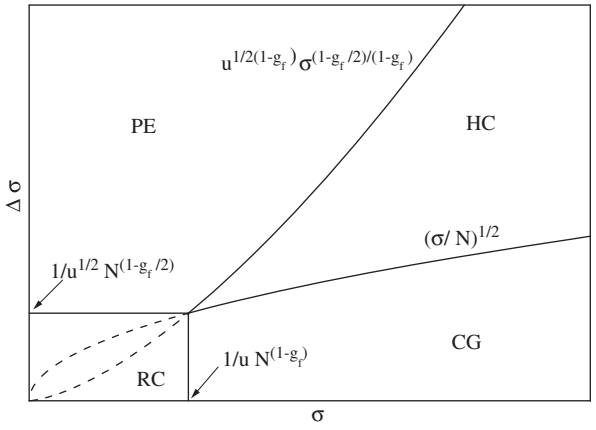

Figure 1. Scheme of the $\Delta\sigma$–σ plot, relating effective $\Delta\sigma$ and total σ charge number fractions, obtained with the polypeptide electrophoretic mobility and the PLLCEM. Full lines delimit the RC, PE, HC and CG regimes. Scaling relationships involving the ratio $u = l_B/L$ between Bjerrum length l_B and characteristic length L associated with chain flexibility, friction power coefficient g_f and total number of amino acid residues N define the transition curve from one regime another.

different number of blobs (see discussion below) which are units formed with a given number of monomers (amino acid residues in the present work). Finally a transition from HC to PE regimes is found when $\Delta\sigma > u^x\sigma^y$ and $\sigma > 1/uN^{(1-g_f)}$. These transitions may be also described in the opposite directions by reversing the inequities placed above (see Fig. 1). It is interesting to observe that the transition from CG to HC involving the line $\Delta\sigma \approx \sqrt{\sigma/N}$ (or equivalently $|Z| \approx \sqrt{Z_T}$) may be explained from a mechanism similar to Rayleigh's instability of a charged liquid drop (see, for instance [51]) where the particle surface energy is exceeded by the repulsion energy generated via unbalanced electrostatic charges. This transition is relevant when proteins are under consideration, as discussed below. In the above expressions, parameter $u = l_B/L$ [51, 53, 56] compares Bjerrum length $l_B = e^2/4\pi\varepsilon' k_B T$ involving the electrical permittivity of protein domain ε' [26] with the characteristic chain length L (typically for proteins $3.8 \le L \le 2a_o$, where 3.8 Å is the distance between consecutive C_α in the backbone chain). In this regard, taking into account the approximate number of amino acid residues (monomers) per blob $g_b \approx (u\sigma)^{-2}$ [51] and the relatively high values of $u = l_B/L$ found for proteins ($u \approx 5$), the blob size is close to the characteristic chain length L for typical total charge fractions of these analytes (thus $g_b \approx 1$). This specific result for natural polypeptides is important for visualizing HZs as partial chain conformations with a tendency to form secondary structures and clusters of them, which involve direct interactions of several amino acid residues. Thus for natural polypeptide, the presence of random chain blobs as occurring in synthetic polyampholytes with smaller characteristic chain length L, in principle, would not be much expected.

It is also clear that global conformations may present different characteristics within each chain regime analyzed above. Chain regimes are described by rather fuzzy limiting lines involving the transitions between them. Limiting lines are placed via basic scale analyses (Fig. 1), where pre-factors are approximate and the powers provide the relevant physical meaning [27, 51, 52]. In this framework additional estimations may be carried out specifically for the HC regime [51]. In fact, for polypeptides, an HC has basically N_a average number of amino acid residues of average radius a_o per HZ. These HZs are interconnected by strings each one having N_b average number of amino acid residues. These numbers keep a fine electrostatic balance between effective $\Delta\sigma$ and total σ charge number fractions [27, 51]. Therefore, for the purposes of the present work, the important relationships valid only for the HC regime are the following:

$$N_a \approx \sigma/\Delta\sigma^2 \quad (3)$$

$$N_b \approx \sqrt{\sigma}/\Delta\sigma \equiv \sqrt{N_a} \quad (4)$$

which are estimated from the basic scaling laws [51]. Although Eqs. (3) and (4) are established within the approximate framework of scale analyses giving values of N_a and N_b that are not necessarily integer numbers, they are still useful quantifiers to describe and discuss the relative tendency of chains to generate HZs in the HC regime comprised between the two transition limits involving the HC→CG and HC→PE destabilizations (Fig. 1), as the chain electrostatic state varies with pH. In this regard, one possible arrangement is that the number of HZs in the whole chain designated as n here gives $(n+1)$ strings alternating with HZs. The other two possibilities are n and $(n-1)$ strings for HZ–string and HZ–HZ terminating chain arrangements, respectively. Based on these results the percent fraction of total HZs in the whole HC may be defined as $\Lambda = 100nN_a/N$, which is an indication of the relative chain tendency to form secondary conformations and clusters of them, as a first approximation, within the framework of scaling global structural parameters. In Section 3, the case of string–string terminal ends giving the lower Λ is used for simplicity only; thus from a simple balance of amino acid residues of the basic AAS one readily gets $n \approx (N - N_b)/(N_a + N_b)$.

When the chain is in a solvent including salt, one should expect Coulombic screening effects for both the fluctuation-attractive electrostatic forces promoting the CG conformation and the electrostatic repulsion forces approaching the chain toward the PE regime. Thus, critical values $\sigma \approx 1/uN^{(1-g_f)}$ and $\Delta\sigma \approx 1/\sqrt{u}N^{(1-g_f/2)}$ for the two transitions of the RC regime become higher along the curve $|Z| = \sqrt{Z_T}$, as the solvent ionic strength I increases around the θ-solvent conditions [51, 53]. It should be observed that the experimental protocols used here have solvents rather low in salt.

To visualize possible destabilization paths of proteins with pH variations, as a first approximation we decided to include the reference titration curve in the characteristic plot $\Delta\sigma$–σ (see for instance Fig. 2) which involves the raw effective electrical charge number Z^r of proteins defined as $Z^r = \sum_{i=1}^{N} Z_i^r$, where Z_i^r is the raw charge number of each i-ionizing group expressed simply $Z_i^r = \pm 1/\{1+10^{\mp(pK_i^r - pH)}\}$ as obtained also from the outputs of the PLLCEM; here

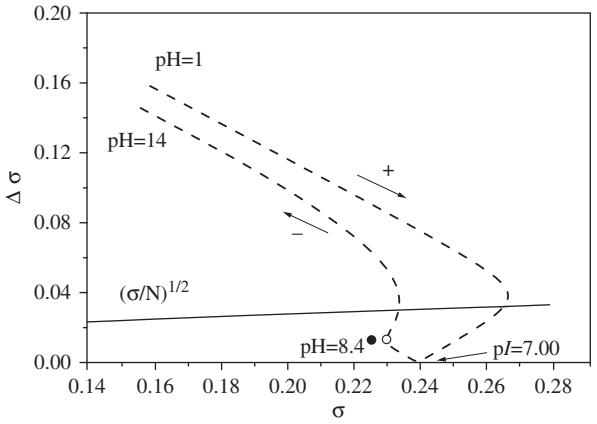

Figure 2. $\Delta\sigma$–σ plot for BCA at protocol pH 8.4, 10 mM and 25°C of the CZE run. The dashed line indicates the raw titration curve of BCA when the pH is changed from 1 to 14 passing through the $pI \approx 7.00$, located at $\sigma \approx 0.240$ and $\Delta\sigma = 0$. Full line refers to the transition HC↔CG. Symbols (○) and (●) indicate coordinates (σ, $\Delta\sigma$) for the raw and charge-regulated electrostatic states of the protein, respectively.

signs are defined according to basic or acid properties of ionizing groups [17]. For these specific calculation purposes only, one considers $pK_i^r \approx pK_i$, and $pH_i \approx pH$, as a first approximation, by neglecting rather small effects of the charge regulation phenomenon (see [17] and its addendum and [22] for a detailed discussion on these expressions).

Finally, the protein transport properties like the friction coefficient $f = 6\pi\eta a_o N^{g_f}$, the sedimentation coefficient $s = M(1 - \nu_p/\nu_w)/\{6\pi\eta a_o N^{g_f} N_A\}$ and the diffusion coefficient $D = k_B T/\{6\pi\eta a_o N^{g_f}\}$ may be evaluated with numerical values of Ω and a_H, at the running protocol conditions [27]. These expressions show the central role played by the power friction coefficient g_f to evaluate s and D, once it has been determined from the output data provided by the PLLCEM and the use of Eq. (2).

3 Results and discussion

Electrophoretic mobility values of the three proteins studied here were obtained from [40–42]. Their associated protocols for the CZE runs together with the corresponding protein AAS were described and analyzed in [17] mainly by considering the presence of metal ion Zn^{+2} in BCA and HCA, affecting directly Z and Z_T.

The basic physicochemical properties, global structural parameters and conformation types obtained from the PLLCEM are fully reported in Supporting Information (Tables SI-1 to SI-7). Additional helpful material within this framework may be also found in [26] for proteins and [27] for peptides.

3.1 Analysis of protein basic regimes in the Δσ–σ plot

For the purpose of visualizing the relevance of the Δσ–σ plot (Fig. 1) to study protein global conformations via CZE, it is appropriate, for instance, to analyze the BCA as illustrated in Fig. 2. This figure is obtained from the protein electrophoretic mobility value and the theoretical framework analyzed in Section 2. It also includes the raw titration curve of this protein when the pH is changed from 1 to 14 passing through the pI located in the σ-axis. This curve is indicated with a dashed line to observe approximately the possible global conformational states that the protein may take as the pH varies. For this protein, the positive branch of the raw titration curve is above the negative one in the HC regime. The numerical results obtained from the PLLCEM at the protocol pH 8.4 places however this protein within the CG regime (full dot) with a rather small ''charge regulation'' effect in the protein electrostatic state, as indicated by the proximity of the coordinates (σ, Δσ) for both the raw (open dot) and the regulated (full dot) electrostatic states. In this plot, the pI is of course located in the abscissa, for $\sigma \approx 0.24$ and $\Delta\sigma = 0$. A different situation is illustrated in Fig. 3 for the HCA indicating in general that each protein may have its Δσ–σ plot as ''fingerprint''. In particular the positive and

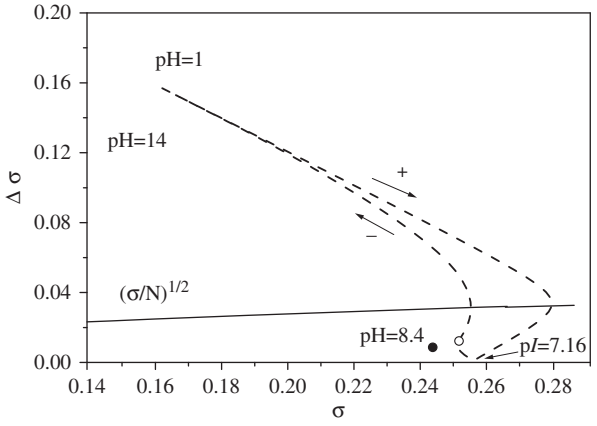

Figure 3. Δσ–σ plot for HCA at protocol pH 8.4, 10 mM and 25°C of the CZE run. The dashed line indicates the raw titration curve of HCA when the pH is changed from 1 to 14 passing through the p$I \approx 7.16$ located at $\sigma \approx 0.258$ and $\Delta\sigma = 0$. Full line refers to the transition HC↔CG. Symbols (○) and (●) indicate coordinates (σ, Δσ) for the raw and charge-regulated electrostatic states of the protein, respectively.

negative curve branches of the HCA start and overlap each other at the extreme pH values in the HC regime. The HCA is also placed in the CG regime at the protocol pH 8.4 showing a higher charge regulation effect at this pH than that found for the BCA. Proteins HCA and BCA show different σ values at the pIs as observed in Figs. 2 and 3. Other case of interest in this context of the discussion is the STN for pH 8.9 (Fig. 4) as a basic state of reference, which is rather on the transition CG→HC described with the curve $|Z| = \sqrt{Z_T}$, or equivalently with $\Delta\sigma = \sqrt{\sigma/N}$, designated here the Kantor–Kardar curve. Consistently with the discussion above, Table SI-2 provides information concerning the global conformations of the three proteins. It was thus found quantitatively in this table that BCA and HCA satisfy $|Z| < \sqrt{Z_T}$ indicating clearly that, under the protocol conditions of the CZE runs, they are in the CG regime, while the STN is near the destabilization curve involving the transition CG→HC (see also Figs. 2–4). Furthermore, Figs. 4–8 illustrate the evolution of the STN from the transition CG→HC to the quite fully denatured state in the transition HC→PE of the Δσ–σ plot (Fig. 8), as the pH is decreased from 8.9 to 2.8 following the CZE protocols used. These results are discussed below in conjunction with global structural parameters.

3.2 Analysis of protein basic states via the packing dimension coefficient g_p

It is important to point out that the three proteins placed below the Kantor–Kardar curve present $g_p \approx 2.8$ (Table SI-2) corresponding to global conformations with a rather 3-D spatial distribution of amino acids residues, as expected and indicated in Section 2. In fact, it was found in [27] that for peptides, g_p may take lower values ($1.7 < g_p < 2.4$, from

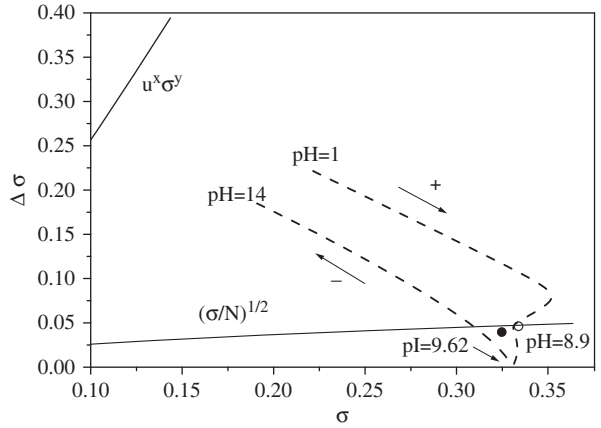

Figure 4. Δσ–σ plot for STN at protocol pH 8.9, 14 mM and 25°C of the CZE run. The dashed line indicates the raw titration curve of STN when the pH is changed from 1 to 14 passing through the pI ≈ 9.62 located at σ ≈ 0.33 and Δσ = 0. Upper and lower full lines refer to transitions PE ↔ HC and HC ↔ CG, respectively. Symbols (○) and (●) indicate coordinates (σ, Δσ) for the raw and charge-regulated electrostatic states of the protein, respectively.

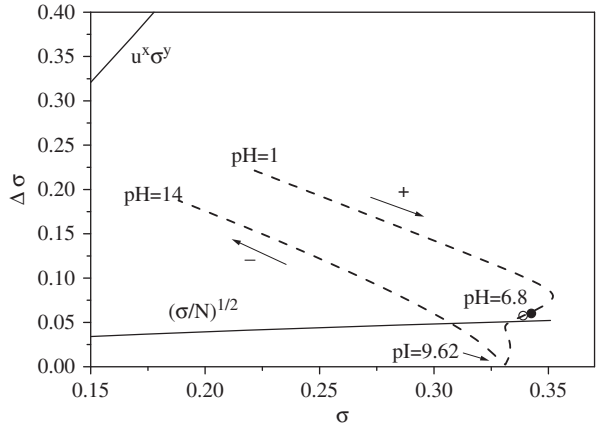

Figure 5. Δσ–σ plot for STN at protocol pH 6.8, 26 mM and 25°C of the CZE run. Other data are the same as in Fig. 4.

rather linear to near planar distributions of amino acid residues) thus allowing one to visualize the importance of this global structural parameter. In this regard, the STN has been studied here for a sequence of pHs to analyze the values taken by the structural parameter g_p. Thus, Table SI-5 shows that the STN at pH 8.9 is in the CG regime (Fig. 4) with $g_p \approx 2.82$, while at pH 6.8 it is in the HC regime (Fig. 5) with $g_p \approx 2.73$. These two global conformations belong to the native state and to the destabilization onset of the STN, respectively, with the asymptotic values of g_p expected (from higher to lower values), yielding thus a rather compact 3-D spatial distribution of amino acid residues, as indicated above. Nevertheless, for pHs 5.7 and 4.1 the STN is partially denatured in the HC regime (Figs. 6 and 7) and calculations also provide consistently that $g_p \approx 2.53$ and 2.35, respectively (Table SI-5). These results show that the STN evolves from 3-D spatial conformations at pHs 8.9 and 6.8 toward rather

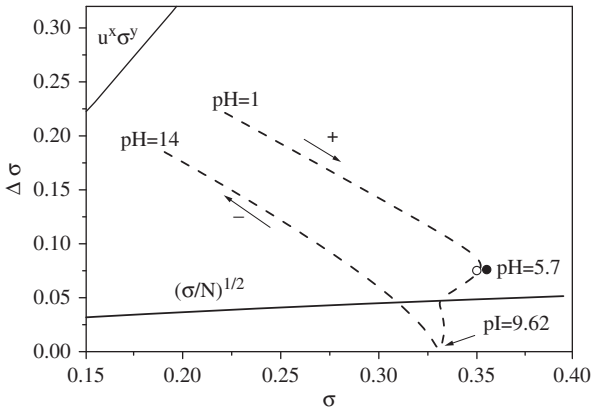

Figure 6. Δσ–σ plot for STN at protocol pH 5.7, 36 mM and 25°C of the CZE run. Other data are the same as in Fig. 4.

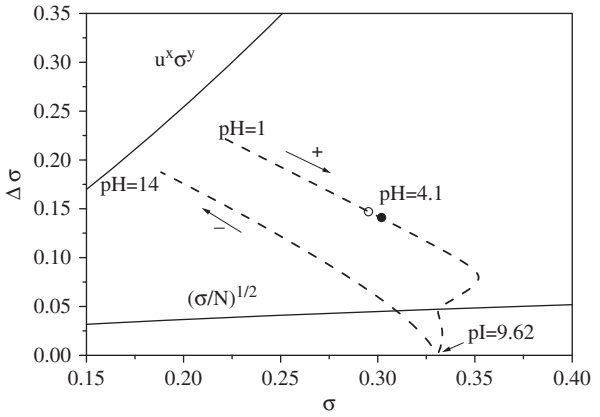

Figure 7. Δσ–σ plot for STN at protocol pH 4.1, 55 mM and 25°C of the CZE run. Other data are the same as in Fig. 4.

planar global conformations at pHs 5.7 and 4.1 as a consequence of the lower values of g_p obtained and depicted in Table SI-5. Furthermore, at pH 2.8, this table indicates that the STN is in the transition between HC and PE regimes (Fig. 8). Table SI-5 also shows that $g_p \approx 1.80$ under this low pH, which is associated with the onset of a rather linear distribution of amino acids residues, consistently with the expected PE regime conformations for the near fully denatured states of globular proteins.

3.3 Analysis of proteins in the HC regime

From the consideration of the HC regime of the STN, several basic concepts presented in Section 2 may be considered with an appropriate physical interpretation. First from basic expressions described in this section, one observes that as the HC conformation approaches the CG, the approximate limits $n \to 1$, $N_a \to N$ and $N_b \to 0$ must be satisfied. On the other hand, the HC conformation near the transition HC → PE presents a relatively high number of small-size HZs serving as ''nucleation points'' for the formation of secondary structure. Thus at this limit, one

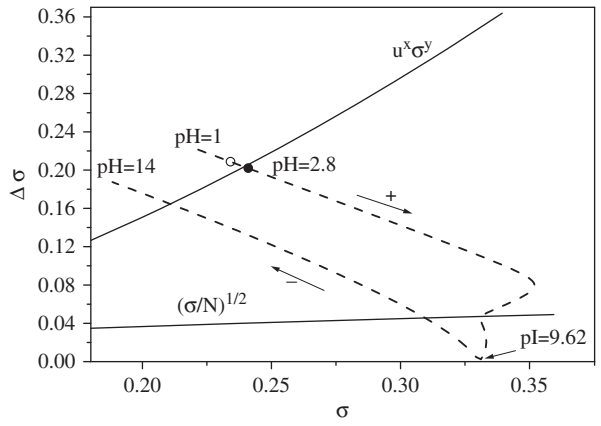

Figure 8. Δσ–σ plot for STN at protocol pH 2.8, 5.5 mM and 25°C of the CZE run. Other data are the same as in Fig. 4.

obtains $n \to 0$, $N_a \to 0$ and $N_b \to N$ as expected for the PE regime (see also Table SI-6). These results implies that after consecutive changes of physicochemical microenvironments (changes in pH) from the transition PE → HC toward the transition HC → CG (see the number decreasing sequence of Figs. 7 to 5), there is a process generating the growing in size of HZs (Table SI-6) with the reduction of HZ number. In fact, average string size and number of amino acid residues per string accompany this evolution by satisfying the balance of total amino acid residues. Since Table SI-6 presents values of n, N_a, N_b and Λ for the STN in the HC regime, it is then clear that global conformations of this protein in the HC regime may be analyzed through Eqs. (3) and (4). Thus results indicate that on the average this protein is composed of HZs with around $N_a \approx 5.9$, 15.2, 61.4 and 94.8 amino acid residues, for pH 2.8, 4.1, 5.7 and 6.8, respectively, while the corresponding strings have $N_b \approx 2.4$, 3.9, 7.8 and 9.7, which are feasible and consistent values less than N, obtained from the protein electrical state at the well-defined pH* and protocol pH (see Table SI-5). These results are also visualized through Λ indicating the tendency of the polypeptide chain to present HZs in the HC regime. Thus, Λ values are around 69.6, 77.5, 84 and 84.8, respectively, taking into account that the last state is approaching the curve $|Z| \approx \sqrt{Z_T}$, as the pH value progresses to the pI. Since this protein is partially denatured at the HC regime, the values of n are consistent with the condition deduced above indicating that the number of HZs kept by the STN in this regime must be higher as the transition HC → PE is approached. From these analyses it is clear that N_a, N_b and Λ belonging to the HC regime are relevant global structural parameters describing denaturing states of globular proteins, perhaps helping one to understand the existence of intermediate states from the native to the fully unfolded protein (see also the so called ''all-to none transitions'' discussed elsewhere [37]). In fact, in Table SI-6, one also observes clearly that intermediate protein states between PE and CG regimes may involve different HZ numbers, i.e. $n = 17.7$, 7.6, 2.0 and 1.3 for pH = 2.8, 4.1, 5.7 and 6.8, respectively.

From the above discussion it is clear that by following the steps along the path of pH-changes in the direction toward the pI of the STN, there is a clear indication that secondary conformations tend to be completed, accompanied simultaneously by the growth of HZs (Table SI-6). Thus at this stage, closer to the Kantor–Kardar curve, an HZ may be also composed of quite complex organizations between several secondary structure (these HZs are called pears beads in synthetic polyampholytes) which are also connected by strings. This process continues until the Kantor–Kardar curve is found at around pH 6.8 for the STN, where the HC gets the transition to CG consistently with $n \to 1$, $N_a \to N$ and $N_b \to 0$, as one expects from the scale analysis in Section 2. Thus in general, the CG of a polypeptide may be visualized in this framework as formed by complex structures of secondary conformations and strings in a quite compact state, evolving from the HC regime. The results here indicate that in these case-studies of the three globular proteins, the RC regime as starting point is less possible (see, for instance [57], and also [58] describing counter-cases belonging to ''native unfolded proteins'', which are targets for future investigations). Certainly, in the CG regime there exist also many possible chain organizations that would depend on coordinates (σ, Δσ) values. Thus additional global structural parameters may be formulated in this regime, and at present, we have this physical aspect as a subject to be elucidated.

Taking into account that blobs at both the HC and CG regimes are not much promoted in polypeptides due to the fact $g_b \approx 1$, it is also clear that in particular the HC has HZs with either small ''nucleation zones'' or secondary conformations (alternatively structured organizations of them), all connected by unstructured RCs or PE strings (strings composed of blobs would not be expected as indicated above). These strings may be also minimized with the growing of HZs before entering the CG regime, as one would expect in general for native functional globular proteins.

To conclude this specific analysis, from the five CZE runs studied here for the STN, this protein presents three states that are partially denatured structures in the HC regime (see also Tables SI-4 to 7). In fact, Figs. 4–8 show that by following the approximate raw titration paths of this protein indicated as dashed line in the Δσ–σ plot, global conformations start from the CG regime at pH 8.9, then the HC regime is obtained mainly for pH 6.8, 5.7 and 4.1, finally reaching the transition curve between the HC and PE regimes, above which the STN may be fully denatured for pH < 2.8 in a rather linear spatial distribution of amino acid residues (see also Table SI-5).

3.4 Analysis of protein basic states via the friction power coefficient g_f

The three proteins (BCA, HCA and STN) at the CG regime have rather compact conformations and Table SI-2 shows consistently that $g_f < 1/2$ (at θ-condition $g_f = 1/2$) indicating

the presence of significant ''hydrodynamic friction shielding'' around each amino acid residue (low free draining chains). Furthermore, from this table one observes that the power friction coefficients for the BCA and HCA may be a little higher than 1/3 due to the fact their protocol pH 8.4 is quite distinct from the corresponding protein p*I*s. On the other hand, the STN at pH 8.9 and $pH^* \approx 9.18$ is closer to the p*I* = 9.6 (Tables SI-1 and SI-2) giving g_f around 1/3, despite it is also located near the transition curve $|Z| \approx \sqrt{Z_T}$, where proteins begin to destabilize. This value would be expected for the fully CG near the p*I*. In general around this curve one should find the onset of globular protein denaturation, starting with the molten globule state, taking also into account that pre-factors of scaling laws are approximate values. Interesting is to observe in Table SI-5, that increasing the pH from 2.8 to 8.9 the values of g_f for the STN varies from 0.57 to 0.27 by crossing the θ-point at around pH 3.45. Thus, along the pH-evolution process, this protein changes from a state of quite free draining chain, where intra-chain hydrodynamic interaction is still important, to a state of shielded hydrodynamic friction of amino acid residues.

Following Section 2, the transport properties of these proteins are presented in Tables SI-3 and SI-7 to highlight the interplay among friction power coefficient, shape orientation factor and equivalent hydrodynamic radius as required in the determination of the diffusion and sedimentation coefficients of proteins [26]. These relationships indicate clearly that, apart from evaluating the hydrodynamic volume swept by the protein along its movement in the CZE run, values of g_f showing the degree of chain free draining are required to be able to evaluate first the average chain friction coefficient and then the diffusion coefficient as a relevant protein global structural parameter.

Finally, taking into account the above discussion, it is also clear that several approximations associated with the definitions of global structural parameters should be further investigated by taking as case-studies other proteins presenting more complex AAS with disulfide bonds, which place additional constraints in the statistics of chain electrostatic interactions. In this regard, although the average physicochemical properties provided by the PLLCEM may still apply, as a first approximation, the definition of additional global structural parameters may be required allowing one to interpret other microstructural aspects not considered in Eqs. (1)–(4).

4 Concluding remarks

The present work demonstrated that the electrophoretic mobility of proteins is a powerful information allowing interpretations that concern the high phenomenological coupling among chain–solvent friction, intra-molecular hydrodynamic interaction and electrostatic charge interactions of ionizing groups, within the whole hydrodynamic polypeptide domain. In addition, the modeling of a protein electrophoretic mobility to estimate protein basic physicochemical and hydrodynamic properties allows one the analysis of relevant global structural parameters describing protein global conformations and regimes, which may be valuable information to study protein function properties at different pHs. Emphasis is placed on the fact that the Δσ–σ plot is derived from the protein CZE mobility data and its appropriate physicochemical and hydrodynamic modeling. The construction of the Δσ–σ plot in order to study protein global conformations via CZE is illustrated and also highlighted as electrostatic ''fingerprint'' of polypeptides.

The authors thank the financial aid received from Universidad Nacional del Litoral, Santa Fe, Argentina (CAI+D 2009) and CONICET (PIP 112-200801-01106).

The authors have declared no conflict of interest.

5 References

[1] Coufal, P., Zuska, J., van de Goor, T., Smith, V., Gaš, B., *Electrophoresis* 2003, *24*, 671–677.

[2] Včelaková, K., Zusková, I., Kenndler, E., Gaš, B., *Electrophoresis* 2004, *25*, 309–317.

[3] Kašička, V., *Electrophoresis* 2006, *27*, 142–175.

[4] Kašička, V., *Electrophoresis* 2008, *29*, 179–206.

[5] Dolník, V., *Electrophoresis* 2008, *29*, 143–156.

[6] Kašička, V., *Electrophoresis* 2010, *31*, 122–146.

[7] El Rassi, Z., *Electrophoresis* 2010, *31*, 174–191.

[8] Stutz, H., *Electrophoresis* 2005, *26*, 1254–1290.

[9] Herrero, M., Ibáñez, E., Cifuentes, A., *Electrophoresis* 2008, *29*, 2148–2160.

[10] Desiderio, C., Iavarone, F., Rossetti, D., Messana, I., Castagnola, M., *J. Sep. Sci.* 2010, *33*, 2385–2393.

[11] Cifuentes, A., Poppe, H., *Electrophoresis* 1995, *16*, 516–524.

[12] Allison, S. A., *Macromolecules* 1996, *29*, 7391–7401.

[13] Adamson, N. J., Reynolds, E. C., *J. Chromatogr. B* 1997, *699*, 133–147.

[14] Cifuentes, A., Poppe, H., *Electrophoresis* 1997, *18*, 2362–2376.

[15] Castagnola, M., Rossetti, D. V., Corda, M., Pellegrini, M., Misiti, F., Olianas, A., Giardina, B., Messana, I., *Electrophoresis* 1998, *19*, 2273–2277.

[16] Allison, S. A., *Biophys. Chem.* 2001, *93*, 197–213.

[17] Piaggio, M. V., Peirotti, M. B., Deiber, J. A., *Electrophoresis* 2005, *26*, 3232–3246.

[18] Piaggio, M. V., Peirotti, M. B., Deiber, J. A., *Electrophoresis* 2006, *27*, 4631–4647.

[19] Xin, Y., Mitchell, H., Cameron, H., Allison, S. A., *J. Phys. Chem. B* 2006, *110*, 1038–1045.

[20] Benavente, F., Balaguer, E., Barbosa, J., Sanz-Nebot, V., *J. Chromatogr. A* 2006, *1117*, 94–102.

[21] Šolínová, V., Kašička, V., Sázelová, P., Barth, T., Mikšík, I., *J. Chromatogr. A* 2007, *1155*, 146–153.

[22] Piaggio, M. V., Peirotti, M. B., Deiber, J. A., *Electrophoresis* 2007, *28*, 3658–3673.

[23] Peirotti, M. B., Piaggio, M. V., Deiber, J. A., *J. Sep. Sci.* 2008, *31*, 548–554.

[24] Pei, H., Xin, Y., Allison, S. A., *J. Sep. Sci.* 2008, *31*, 555–564.

[25] Pei, H., Allison, S. A., *J. Chromatogr. A.* 2009, *1216*, 1908–1916.

[26] Piaggio, M. V., Peirotti, M. B., Deiber, J. A., *Electrophoresis* 2009, *30*, 2328–2336.

[27] Piaggio, M. V., Peirotti, M. B., Deiber, J. A., *J. Sep. Sci.* 2010, *33*, 2423–2429.

[28] Allison, S. A., Pei, H., Allen, M., Brown, J., Chang-II, K., Zhen, Y., *J. Sep. Sci.* 2010, *33*, 2439–2446.

[29] Piaggio, M. V., Peirotti, M. B., Deiber, J. A., *Electrophoresis* 2007, *28*, 2223–2234.

[30] Chow, P. Y., Fasman, G. D., *Biochemistry* 1974, *13*, 211–222.

[31] Arteca, G. A., *Can. J. Chem.* 1995, *73*, 241–248.

[32] Xiong, H., Buckwalter, B. L., Shieh, H.-M., Hecht, M. H., *Proc. Natl. Acad. Sci. USA* 1995, *92*, 6349–6353.

[33] Liu, H., Yao, X., Xue, C., Zhang, R., Liu, M., Hu, Z., Fan, B., *Anal. Chim. Acta* 2005, *542*, 249–259.

[34] Zhong, L., Johnson, C., *Proc. Natl. Acad. Sci. USA* 1992, *89*, 4462–4465.

[35] Šolínová, V., Kašička, V., Koval, D., Hlaváček, J., *Electrophoresis* 2004, *25*, 2299–2308.

[36] Plasson, R., Vayaboury, W., Giani, O., Cottet, H., *Electrophoresis* 2007, *28*, 3617–3624.

[37] Creighton, T. E., (Ed.), *Protein Folding*, W. H. Freeman and Company, New York 1992.

[38] Chen, Y., Ding, F., Nie, H., Serohijos, A., Sharma, S., Wilcox, K., Yin, S., Dokholyan, N., *Arch. Biochem. Biophys.* 2008, *469*, 4–19.

[39] Tanford, Ch., *Physical Chemistry of Macromolecules*, Wiley, USA 1961.

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

[41] 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, 2004, pp. 189–216.

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

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

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

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

[46] Bashford, D., *Front. Biosci.* 2004, *9*, 1082–1099.

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

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

[49] Kim, J. Y., Ahn, S. H., Kang, S. T., Yoon, B. J., *J. Colloid Interface Sci.* 2006, *299*, 486–492.

[50] Flory, P. J., *Principles of Polymer Chemistry*, 9th Edn, Cornell University Press, London 1975.

[51] Dobrynin, A. V., Colby, R. H., Rubinstein, M., *J. Polym. Sci. Part B Polym. Phys.* 2004, *42*, 3515–3538.

[52] Rubinstein, M., Colby, R. H., *Polymer Physics*, 1st Edn, Oxford University Press, Oxford 2004.

[53] Dobrynin, A. V., Rubinstein, M., Joanny, J.-F., *Macromolecules* 1997, *30*, 4332–4341.

[54] Kantor, Y., Kardar, M., *Phys. Rev. E* 1995, *51*, 1299–1312.

[55] Kantor, Y., Kardar, M., *Phys. Rev. E* 1995, *52*, 835–846.

[56] Higgs, P. G., Joanny, J.-F., *J. Chem. Phys.* 1991, *94*, 1543–1553.

[57] Shortle, D., *FASEB J.* 1996, *10*, 27–34.

[58] Uversky, V., Dunker, A. K., *Biochim. Biophys. Acta* 2010, *1804*, 1231–1264.

