# Charge fluctuation effects on the shape of flexible polyampholytes with applications to Intrinsically disordered proteins

Himadri S. Samanta[1] and D. Thirumalai[1]

[1]*Department of Chemistry, University of Texas at Austin, TX 78712*

(Dated: April 15, 2018)

Polyampholytes (PAs) contain positively and negatively charged monomers that are distributed randomly along the polymer chain. The interaction between charges is assumed to be given by the Debye-Huckel potential. We show that the size of the PA is determined by an interplay between electrostatic interactions, giving rise to the polyelectrolyte (PE) effect due to net charge per monomer ($\sigma$), and an effective attractive PA interaction due to charge fluctuations, $\delta\sigma$. The interplay between these terms gives rise to non-monotonic dependence of the radius of gyration, $R_g$ on the inverse Debye length, $\kappa$ when PA effects are important ($\frac{\delta\sigma}{\sigma} > 1$). In the opposite limit, $R_g$ increases monotonically with increasing $\kappa$. The theory is used to calculate the changes in $R_g$ on $N$, the number of residues for a set of Intrinsically Disordered Proteins (IDPs). For a certain class of IDPs, with $N$ between 24 to 185, the size grows as $R_g \sim N^{0.6}$, which agrees with data from Small Angle X-ray Scattering (SAXS) experiments. The theory also describes quantitatively the $R_g$ measured using SAXS experiments for the wild type Tau proteins and fragments derived from it. In this case we find that $R_g \sim N^{0.5}$. The variations in the scaling between these two sets are a consequence of differing values of $\sigma$ and $\delta\sigma$, as well as conformational heterogeneity in the IDPs.

PACS numbers:

## I. INTRODUCTION

The shapes and dynamics of polyampholytes (PAs), which are polymers with monomers that carry both positive and negative charges, have been extensively studied [1–8]. Polyampholytes naturally occur in aqueous solution if the monomers contain acidic and basic groups. In this sense, all proteins are PAs in which charged residues are interspersed between hydrophobic and hydrophilic residues. Because of the simultaneous presence of positive and negative charges, the conformations of the PAs are determined by an interplay of electrostatic interactions, charge fluctuation effects (see below), as well as the stiffness of the backbone. In simple terms, we expect that repulsion between like-charges would stretch the chain whereas attraction would tend to make the polymer compact. Of course, in random PAs this balance is determined on an average performed over an ensemble of sequences. If the number, $N$, of monomers is large then the PA is predicted to adopt compact conformations if the polymer is overall neutral (the number of positive and negative charges nearly cancel). On the other hand if there is residual charge on the PA it is likely to be extended. It should be noted that there are differences in the behavior of the dependence of the radius of gyration, $R_g$, on $N$ depending on whether the the chain is globally neutral (plus and minus charges exactly cancel) or statistically neutral [9] (residual charge when averaged over a large number of sequences scales as $\sqrt{N}$ with $N \gg 1$). Thanks to several insightful theoretical studies [2–4, 10], the complex phase behavior of PAs as a function of salt concentration and temperature of have been elucidated.

More recently, there has been renewed interest in PAs in the biophysics community because many eukaryotic proteins contain an unusually large fraction of charged residues [11–13]. As a consequence the favorable hydrophobic interactions cannot overcome the residual electrostatic interactions. Consequently, this class of proteins do not adopt globular structures unless it is in complex with another partner protein. Polypeptide sequences with this characteristic are referred to as intrinsically disordered proteins or IDPs because they do not have stable ordered structures under physiological conditions. It is also the case that there are protein sequences in which only certain regions are disordered under nominal conditions. Because of the preponderance of such sequences and their roles in a variety of cellular functions and the potential role they play in diseases [13, 14], there is heightened interest in understanding their structural and dynamical properties [15–18]. The IDPs, whose backbone is relatively flexible (persistence length in the range (0.6 - 1.0) nm), are low complexity sequences containing a large fraction of charged residues and smaller fraction of hydrophobic residues compared to their counterparts that adopt well-defined structures in isolation. As a consequence water is likely to be a good or at best a $\theta$ solvent, which means that $R_g \approx N^{\nu}$ where $\nu$ is approximately 0.6 or 0.5. There are differences between IDPs and random PAs. (i) The sequences of IDPs are quenched, thus making it necessary to understand the conformations of a specific sequence. In other words, two sequences with identical charge composition could have drastically different structural characteristics. Of course, this could be the case for random PAs as well although this aspect has not been investigated as much. (ii) Unlike the case of PAs for which $N \gg 1$, which allows one to develop analytical and scaling type arguments using well-developed methods in polymer physics, typically studied IDPs have finite $N$, at best on the order of a few hundred residues. (iii) IDPs also contain uncharged amino acids,

which are not usually considered when treating PAs using theory and simulations. Despite these differences, concepts from polyelectrolytes (PEs) and PAs have been used to envision the conformations of IDPs using the difference between positive and negative charge ($\sigma$) and net charge as appropriate variables [19].

The importance of sequence effects on the $R_g$ of PAs was first illustrated in a key note by Srivastava and Muthukumar [3]. Using Monte Carlo simulations, with $N = 50$, they showed that there are substantial variations in $R_g$ in PAs (containing only charged monomers) for a globally neutral chain. This study showed that the location of charges (sequence specificity) plays a crucial role in determining the conformational properties. More recently, Firman and Ghosh [20] used Edwards model for charged polymers, encoding for the precise sequence in order to calculate $R_g$s for small $N$. Their theory successfully accounted for simulations of synthetic IDPs [21], containing only a mixture of positive and negative charged residues.

Here, we develop a theory to investigate the effects of charge fluctuations on the shapes of random PAs. In our model there is a probability, $p_+$, ($p_-$) that a monomer at location $s$ is positively (negatively) charged. The fluctuations arise because the normalized charge distribution is taken to be stochastic given by,

$$P[\sigma(s)] = p_+\delta[\sigma(s)-1]+p_-\delta[\sigma(s)+1]+(1-p_+-p_-)\delta[\sigma(s)]. \tag{1}$$

The charges are measured in units of $e^-$. Because some monomers do not carry a charge (like in IDPs) $(p_+ + p_-) \neq 1$. The mean $\langle\sigma(s)\rangle$ gives the net charge, $p_+ - p_-$, and the expression for the square of the charge fluctuations is, $\langle\delta\sigma^2(s)\rangle = p_+ + p_- - (p_+ - p_-)^2$. We show that due to $\langle\delta\sigma^2(s)\rangle$ the $R_g$ is altered substantially, and could even induce a coil-globule transition even when the total charge on the PA is not globally neutral. Because of the opposing behavior of polyelectrolyte (PA) ($\sigma \neq 0$) and PA effects arising from charge fluctuations ($\langle\delta\sigma^2(s) \neq 0$), the dependence of $R_g$ on the Debye screening length is non-monotonic. The phase diagram in the $[\langle\sigma(s)\rangle,\langle\delta\sigma(s)\rangle]$ plane is rich. We also apply the theory to calculate $R_g$ of specific IDP sequences. Remarkably, the theory reproduces quantitatively the $R_g$ values for the wild type Tau protein and various fragments obtained from the wild type Tau, which have been measured by Small Angle X-ray Scattering (SAXS) experiments [22]. In Tau, and other IDPs, charge fluctuations arise because of conformational heterogeneity, which we demonstrate explicitly elsewhere [23]. From now on we drop the angular brackets in both $\langle\sigma\rangle$ and $\langle\delta\sigma\rangle$.

## II. THEORY

We begin by considering the Edwards Hamiltonian for a polymer chain:

$$\mathcal{H} = \frac{3k_BT}{2a_0^2}\int_0^N \left(\frac{\partial\vec{r}}{\partial s}\right)^2 ds + k_BTV(\vec{r}(s)), \tag{2}$$

where $\vec{r}(s)$ is the position of the monomer $s$, $a_0$ is the monomer size, $N$ is the number of monomers. The first term in the Eq.(2) accounts for chain connectivity, and the second term represents the sum of excluded volume interactions, electrostatic interactions, and effects of charge fluctuations (see below) due to the random values of charges in different positions in the ensemble of sequences. The expression for $V(\vec{r}(s))$ is,

$$\begin{aligned} V(\vec{r}(s)) &= \frac{v_0}{(2\pi a_0^2)^{3/2}}\sum_{s,s'=0}^{N}\exp[-\frac{(\vec{r}(s)-\vec{r}(s'))^2}{2a_0^2}] \\ &+ l_B\int_0^N\int_0^N ds\ ds'\ \sigma(s)\sigma(s')\frac{e^{-\kappa|\vec{r}(s)-\vec{r}(s')|}}{\mid \vec{r}(s)-\vec{r}(s')\mid} \\ &= V_0 + V_1(\mid \vec{r}(s)-\vec{r}(s')\mid). \end{aligned} \tag{3}$$

The first term in Eq.(3) accounts for the non-specific two body excluded volume interactions. It differs insignificantly from the usual $\delta$ function potential used in the standard Edwards model. Of course, when $a_0$ is small compared to $R_g$, the precise form of this term is irrelevant, as long as it is short-ranged. In a good solvent ($v_0 > 0$), the polymer chain swells with $R_g \sim a_0N^\nu$ ($\nu \approx 0.6$), where as in a poor solvent ($v_0 < 0$), the size of the polymer is $R_g \sim a_0N^\nu$ ($\nu \approx 1/3$). Here, we consider a PA in a good solvent ($v_0 > 0$).

From Eq.(3) one may obtain an effective interaction term between charges on the PA chain. By following the theory developed previously [24], we use the Hubbard-Stratonovich transformation to decouple the product of charges $\sigma(s)\sigma(s')$ in Eq.3. The partition function may be written as,

$$\begin{aligned} Z = \mathcal{N}^{-1} \int d[\psi(\vec{r})]\exp\Big[-\frac{1}{2}\int d\vec{r}d\vec{r'}\psi(\vec{r}) \\ V_1^{-1}(\mid \vec{r}(s)-\vec{r}(s')\mid)\psi(\vec{r'})\Big] Z_\psi \end{aligned} \tag{4}$$

where, $Z_\psi = \int d[\vec{r}]\exp\left[-V_0 - i\int ds\sigma(s)\psi(\vec{r}(s))\right]$, and $\mathcal{N} = \int d[\psi(\vec{r})]\exp[-\frac{1}{2}\int d\vec{r}d\vec{r'}\psi(\vec{r})V_1^{-1}(\mid \vec{r}(s)-\vec{r}(s')\mid)\psi(\vec{r'})]$. If we assume that the charge distribution (Eq. 1) is annealed it suffices to average $Z_\psi$ over the sequence of charges. With assumption that the charges $\sigma(s)$ at distant sites are not correlated, one may write down the partition function averaged over sequence of charges to second order in $\psi$, [24] as,

$$\begin{aligned} < Z_\psi >_{seq} &= \int \mathcal{D}[\vec{r}]\exp\{-i\sigma\int\psi(\vec{r})c(\vec{r})d\vec{r} \\ &- \frac{1}{2}(\delta\sigma)^2\int[\psi^2(\vec{r}) - <\psi^2(\vec{r})>_\psi]c(\vec{r})d\vec{r} \end{aligned} \tag{5}$$

where the average value of the charge on the chain, $\sigma =< \sigma(s) >= p_+ - p_-$, the charge fluctuation, $(\delta\sigma)^2 =< \sigma^2(s) - < \sigma(s) >^2 >= p_+ + p_- - (p_+ - p_-)^2$ and the local monomer density, $c(\vec{r}) = \int ds \delta(\vec{r}(s) - \vec{r})$. The term involving $(\delta\sigma)^2$, arising from the charge fluctuations, gives rise to the so called PA effect, which is manifested as an effective attractive interaction of the screened Coulomb potential. Using Eq.(4) and Eq.(5), we perform the needed integration over $\psi(\vec{r})$ to obtain the following expression for the effective two body interaction term between charges on the PA,

$$\begin{aligned} \mathcal{V}(\vec{r}(s)) &= \frac{v}{(2\pi a_0^2)^{3/2}} \sum_{s,s'=0}^{N} \exp[-\frac{(\vec{r}(s) - \vec{r}(s'))^2}{2a_0^2}] \quad (6) \\ &+ \sigma^2 l_B \int\int ds\, ds' \, \frac{e^{-\kappa|\vec{r}(s)-\vec{r}(s')|}}{\mid \vec{r}(s) - \vec{r}(s') \mid} \\ &- \frac{1}{2}(\delta\sigma)^4 l_B^2 \sum_{\{s,s'\}} \frac{e^{-2\kappa|\vec{r}(s)-\vec{r}(s')|}}{\mid \vec{r}(s) - \vec{r}(s') \mid^2}. \end{aligned}$$

We neglect the three body interactions in the effective Hamiltonian in Eq.(6), which would be important if the PA were in a poor solvent. Note that the coefficient of the charge fluctuation term differs from that in an important previous study [2]. We believe that this difference arises because we averaged the partition function over the distribution of charges using Eq. 1, which was not explicitly performed by Higgs and Joanny [2].

The screened Coulomb potential, the second term in Eq.(6), accounts for the interactions between charges separated by a distance $\mid \vec{r}(s) - \vec{r}(s') \mid$. The strength of the unscreened electrostatic interactions is characterized by the Bjerrum length $l_B = e^2/\epsilon k_B T$. The Debye screening length, $\kappa^{-1}$ determines the range of the electrostatic interactions. By changing the value of $\kappa$, and hence the range of charge interactions, the PA chain could undergo a coil-to-globule transition. The value of $\kappa$ may be the changed by decreasing or increasing the salt concentration. The dimensionless parameter, $\sigma$, determines the net charge per residue on the polyelectrolyte chain. For a particular sequence, fraction $p = p_+ + p_-$ of the monomers are charged with the charge on each monomer being $\pm e$. Therefore, the net charge per monomer is $\sigma =\mid p_+ - p_- \mid$. The third term in Eq.(6) is the attractive interaction term that is proportional to charge fluctuations $(\delta\sigma)$. The PA affect arises due to the interaction between charge and dipoles formed between sequence of positive and negative charges. The charge-dipole interaction term decays as $\mid \vec{r}(s) - \vec{r}(s') \mid^2$ and it is effectively screened (with a screening length $1/2\kappa$) due to the presence of other dipoles. In the absence of the third term the Hamiltonian would describe a polyelectrolyte, whose phases as function of temperature and $\kappa$ have been previous described using the methods used here [25].

In order to obtain $R_g$, we adopt the Edwards-Singh (ES) type variational calculation [26], which has been extensively used in the polymer literature [2, 25, 27–29]. More recently, the method was used to study sequence dependence of collapse of polypeptide chains [30] and polyelectrolytes [20] with application to a special class of synthetic IDPs. In developing the theory, we assume that interactions between charges exist only between specific monomers, described by the second and third term in Eq.(6). The sum is over the set of specific contacts between pairs $\{s_i, s_j\}$. We use the contact maps of IDP, generated in coarse-grained simulations of IDPs [23], in order to assign the specific interactions.

The ES method is a variational type (referred to as the uniform expansion method) calculation that represents the exact Hamiltonian by a Gaussian chain with an effective monomer size, which is determined as follows. Consider a virtual chain without excluded volume interactions, whose radius of gyration $\langle R_g^2 \rangle = Na^2/6$ [26], described by the Hamiltonian,

$$\mathcal{H}_v = \frac{3k_BT}{2a^2} \int_0^N \left(\frac{\partial \vec{r}}{\partial s}\right)^2 ds. \quad (7)$$

The monomer size in the trial Hamiltonian is $a$. We split the deviation $\mathcal{W}$ between the virtual chain Hamiltonian and the real Hamiltonian as,

$$\mathcal{H} - \mathcal{H}_v = k_BT\mathcal{W} = k_BT(\mathcal{W}_1 + \mathcal{W}_2), \quad (8)$$

where

$$\mathcal{W}_1 = \frac{3}{2}\left(\frac{1}{a_0^2} - \frac{1}{a^2}\right) \int_0^N \left(\frac{\partial \vec{r}}{\partial s}\right)^2 ds, \; \mathcal{W}_2 = \mathcal{V}(\vec{r}(s)). \quad (9)$$

The radius of gyration is $R_g^2 = \frac{1}{N}\int_0^N \langle \vec{r}^2(s)\rangle ds$, with the average being, $\langle \vec{r}^2(s)\rangle = \frac{\int r^2 e^{-\mathcal{H}_v/k_BT} e^{-\mathcal{W}} \delta\vec{r}}{\int e^{-\mathcal{H}_v/k_BT} e^{-\mathcal{W}} \delta\vec{r}} = \frac{\langle \vec{r}^2(s) e^{-\mathcal{W}}\rangle_v}{\langle e^{-\mathcal{W}}\rangle_v}$, where, $\langle \cdots \rangle_v$ denotes the average over $\mathcal{H}_v$.

Assuming that the deviation $\mathcal{W}$ is small, we can calculate the average to first order in $\mathcal{W}$. The result is, $\langle \vec{r}^2(s)\rangle \approx \frac{\langle \vec{r}^2(s)(1-\mathcal{W})\rangle_v}{\langle (1-\mathcal{W})\rangle_v} \approx \langle \vec{r}^2(s)(1-\mathcal{W})\rangle_v \langle (1+\mathcal{W})\rangle_v$ and the radius of gyration becomes,

$$\begin{aligned} < R_g^2 > &= \frac{1}{N}\int_0^N \langle \vec{r}^2(s)\rangle ds \quad (10) \\ &= \frac{1}{N}\int_0^N [\langle \vec{r}^2(s)\rangle_v + \langle \vec{r}^2(s)\rangle_v \langle \mathcal{W}\rangle_v - \langle \vec{r}^2(s)\mathcal{W}\rangle_v] ds. \end{aligned}$$

If we choose the effective monomer size $a$ in $\mathcal{H}_v$, such that the first order correction (second and third terms in the right hand side of Eq.(10)) vanishes, then the size of the chain is, $\langle R_g^2 \rangle = Na^2/6$. This is an estimate of the exact $\langle R_g^2 \rangle$, and is only an approximation as we have neglected $\mathcal{W}^2$ and higher powers of $\mathcal{W}$. Thus, in the ES theory, we determine $a$ using Eq. (10),

$$\frac{1}{N}\int_0^N [\langle \vec{r}^2(s)\rangle_v \langle \mathcal{W}\rangle_v - \langle \vec{r}^2(s)\mathcal{W}\rangle_v] ds = 0. \quad (11)$$

The equation above leads to a self-consistent equation for $a$, and is given by [26]:

$$\frac{1}{a_0^2} - \frac{1}{a^2} = \frac{\frac{1}{N}\int_0^N [\langle \vec{r}^2(s)\rangle_v \langle \mathcal{V}\rangle_v - \langle \vec{r}^2(s)\mathcal{V}\rangle_v] ds}{\frac{a^2}{N}\int_0^N ds\ \langle \vec{r}^2(s)\rangle_v}. \quad (12)$$

By calculating the averages in the Fourier space ($\vec{r}_n = \frac{1}{N}\int_1^N \cos\left(\frac{\pi n s}{N}\right)\vec{r}(s)ds$; $\vec{r}(s) = 2\sum_{n=1}^{N}\cos\left(\frac{\pi n s}{N}\right)\vec{r}_n$; $R_g^2 = 2\sum_n \langle|\vec{r}_n^{\,2}|\rangle$), we obtain

$$\begin{aligned}\frac{1}{a_0^2} - \frac{1}{a^2} &= \frac{4Na_0^3}{9\pi\sum\frac{1}{n^2}}\sum_{s,s'=0}^{N}\frac{C_1^{ss'}}{(\frac{a^2N}{3\pi^2}C_2^{ss'}+a_0^2)^{\frac{5}{2}}} \\ &+\frac{4N\sigma^2 l_B}{9\pi^3\sum\frac{1}{n^2}}\sum_{\{s,s'\}}C_1^{ss'}\left(\frac{\pi^{1/2}(1-\frac{2\kappa^2a^2NC_2^{ss'}}{3\pi^2})}{4(\frac{2a^2N}{3\pi^2}C_2^{ss'})^{3/2}}+\frac{\pi\kappa^3}{2}e^{(\frac{a^2\kappa^2NC_2^{ss'}}{3\pi^2})}\mathrm{erfc}[\kappa\sqrt{\frac{2a^2N}{3\pi^2}C_2^{ss'}}]\right) \\ &-\frac{4N(\delta\sigma)^4 l_B^2}{9\pi^3\sum\frac{1}{n^2}}\sum_{\{s,s'\}}C_1^{ss'}\int_0^\infty dq\ q^3\left(\pi-\arctan\left(\frac{2\kappa}{q}\right)\right)\exp\left(-q^2\frac{2a^2N}{3\pi^2}C_2^{ss'}\right)\end{aligned} \quad (13)$$

where, $C_1^{ss'} = \sum_{n=1}^{N}\frac{1-\cos[n\pi(s-s')/N]}{n^4}$ and $C_2^{ss'} = \sum_{n=1}^{N}\frac{1-\cos[n\pi(s-s')/N]}{n^2}$. In obtaining Eq. 13 we have used $v_0 = \frac{4\pi a_0^3}{3}$ in Eq. 3

From Eq.(13), we can calculate the effective monomer size $a$, and hence the chain size $< R_g^2 >= \frac{a^2N}{6}$. However, without having to solve Eq.(13) numerically, we can define the $\theta$-like point, which signals the onset of a potential transition from a coil to globule state in the PA. At the $\theta$-point, the repulsive terms exactly balance the PA term. Since at the $\theta$-point, the PA behaves as a Gaussian chain, with $a = a_0$, we substitute this value for $a$ in Eq.13 to determine the it. Thus, from Eq.(13), the critical charge fluctuation value, at which the PA term equals the excluded volume and PE terms is,

$$\begin{aligned}(\delta\sigma_\theta^2)^2 &= \Bigg[\frac{4Na_0^3}{9\pi\sum\frac{1}{n^2}}\sum_{s,s'=0}^{N}\frac{C_1^{ss'}}{(\frac{a^2N}{3\pi^2}C_2^{ss'}+a_0^2)^{\frac{5}{2}}} \\ &+\frac{4N\sigma^2 l_B}{9\pi^3\sum\frac{1}{n^2}}\sum_{\{s,s'\}}C_1^{ss'}\left(\frac{\pi^{1/2}(1-\frac{2\kappa^2a^2NC_2^{ss'}}{3\pi^2})}{4(\frac{2a^2N}{3\pi^2}C_2^{ss'})^{3/2}}+\frac{\pi\kappa^3}{2}e^{(\frac{a^2\kappa^2NC_2^{ss'}}{3\pi^2})}\mathrm{erfc}[\kappa\sqrt{\frac{2a^2N}{3\pi^2}C_2^{ss'}}]\right)\Bigg] / \\ &\left[\frac{4Nl_B^2}{9\pi^3\sum\frac{1}{n^2}}\sum_{\{s,s'\}}C_1^{ss'}\int_0^\infty dq\ q^3\left(\pi-\mathrm{Arctan}\left(\frac{2\kappa}{q}\right)\right)\exp\left(-q^2\frac{2a^2N}{3\pi^2}C_2^{ss'}\right)\right]\end{aligned} \quad (14)$$

The numerator in Eq.(14) is a consequence of the repulsion containing excluded volume interactions and polyelectrolyte term. The denominator encodes the PA effect, determining the extent to which the size of the polymer changes due to charge fluctuations. Using Eq.(14) we obtain the dependence of $\delta\sigma_\theta$ on $N$. Scaling $n$ by $N$, it can be sshown that $C_1^{ss'} \sim \frac{1}{N^2}$ and $C_2^{ss'} \sim 1$. From these result, we obtain, $\delta\sigma_\theta \sim \sqrt{N}$. The implication is that for $N \gg 1$, charge fluctuations have to be extremely large to drive coil to globule transition unless the PA is globally neutral. Because even for statistically neutral PA the PE term would not be irrelevant, we surmise that a genuine coil to globule transition transition may not be easily realizable in long PAs, which is in accord with the results in a previous study [9]. By implication our theory suggests that a maximally compact IDPs would be difficult to obtain for generic IDP sequences if the fractions of + and - charged residues is on the order of (0.4 - 0.5). Of course, to establish the various conformations IDPs or PAs adopt as the nominal parameters ($\sigma$, $\delta\sigma$, and $\kappa$) are varied, will require performing detailed calculations as was previously done for polyelectrolytes [31].

## III. RESULTS:

From Eq. 6 it is easy to show that that the size of the PA should be determined by $\frac{\delta\sigma}{\sigma}$, which can be written as $\sqrt{(1-\sigma_A)/\sigma}$ where $\sigma_A = \frac{(p_+ - p_-)^2}{\sigma}$, which in the IDP literature literature is referred to as the charge asymmetry parameter. Fig. (1), displaying the dependence of the radius of gyration for a PA chain, with randomly distributed charges, on the screening length, shows that the $R_g$ changes non-monotonically as $\kappa$ increases when $\frac{\delta\sigma}{\sigma} = 2.4$. In this charge-fluctuation dominated regime, the behavior in Fig. (1) can be explained by noting that at small values of $\kappa$, $R_g$ increases due to the PE term until $\kappa l_b = 0.13$. Interestingly, such an increase has been observed in two IDPs (N-terminal domain of HIV-1 integrase and human prothymosin $\alpha$) at low ionic concentration [32]. For the PA whose $R_g$ is shown in Fig. (1), at $\kappa l_B = 0.13$ the attractive and repulsive terms balance each other. Upon further increase in $\kappa$ the decrease in $R_g$ (but the chain is not a globule) is due to the dominance of the PA term. In the opposite limit when $\frac{\delta\sigma}{\sigma} = 0.1$, both the dimensions of the chain are dominated by the PE term and $R_g$ increases with $\kappa$. We expect that at sufficiently large values of $\kappa l_B$ the consequences of PE and PA effects are negligible, and hence, $R_g$ would have the value expected for a Flory random coil ($\nu = 0.6$).

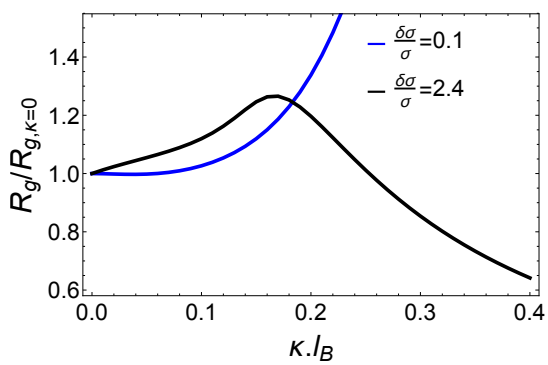


FIG. 1: Non-monotonic increase in the radius of gyration, $R_g$, with increasing value of the inverse screening length $\kappa$ for a PA chain (N=150).

The generality of the theory allows us to predict the dependence of the size of an IDP. In Fig.(2) we plot the $\kappa$ dependence of radius of gyration of a Tau protein fragment (K25Tau ($N$=185)). To perform the calculations, we used the contact map generated in simulations based on the Self-Organized Polymer (SOP)-IDP model, which captures accurately the measured structure factors for a variety of IDPs [23]. Using the input from simulations, which accounts for heterogeneity of the conformational ensembles of IDP, we find that the $R_g$ of K25Tau changes non-monotonically with increasing value of $\kappa$ (Fig. (2)). The size of K25Tau protein increases with $\kappa$ until it reaches a maximum at $\kappa l_B \approx 0.28$ (ion density for a monovalent salt is $\approx$ 30.5 mM), where the protein behaves like a polymer in a $\theta$-solvent. With further increase in $\kappa$, $R_g$ decreases just as for the random PA chain (Fig. 1). From these results, we conclude that charge fluctuations are substantial in K25Tau.

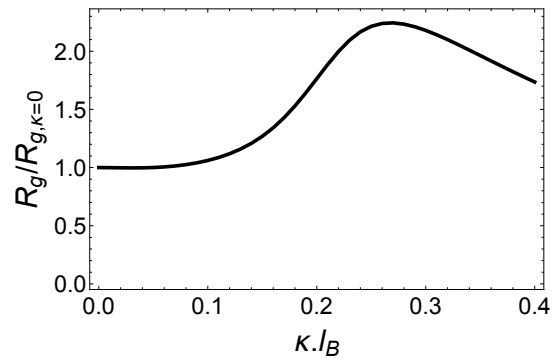


FIG. 2: The size of the chain ($R_g$) changes non-monotonically with increasing value of $\kappa$ for K25Tau (N=185) protein. The value of $\delta\sigma$ is 0.9.

The 3D plot in fig.(3) shows the phase diagram for various parameter values of net charge per monomer ($\sigma$) and charge fluctuations described by $\delta\sigma$. For small values of $\sigma$, the change in $R_g$ is small over a range of values of $\delta\sigma$. For a large value of net charge, say $\sigma = 0.5$, there is a significant change in $R_g$ at a particular value of $\delta\sigma$. The value of $\delta\sigma_\theta$ increases with $\sigma$ for a PA chain. If we further increase $\sigma$ (e.g. $\sigma = 0.8$), $R_g$ of the chain remains unchanged over a range of $\delta\sigma$, indicating that PE effects dominate.

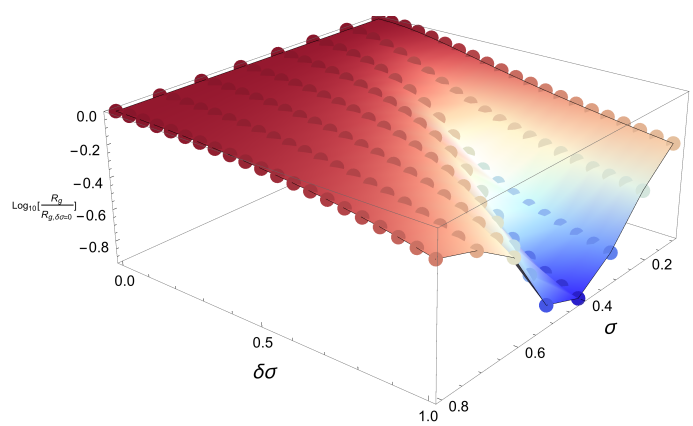


FIG. 3: Phase diagram for different values of net charge $\sigma$ and the charge fluctuation parameter $\delta\sigma$. The chain size decreases monotonically as $\delta\sigma$ increases. The parameters for the plot: $N = 150$ and $\kappa = 0.4nm^{-1}$.

In order to calculate $R_g$ for several IDPs (see Fig. 4), with $N$ ranging from 24 (HISTATIN5) to 185 (An16) we used the average contact maps from simulations [23], which restricts the summation in Eq.6 to specific sites on the IDP. The dependence of $R_g$ on the chain length for

a set of IDPs (listed in the caption to Fig. 4) is shown in Fig. 4. The theory shows that $R_g \sim N^{0.6}$, implying that these IDPs behave as self avoiding polymers, similar to the results in the simulations for PAs [9]. The scaling in Fig. 2 has a weaker $N$ dependence than predicted by the renormalization group argument ($R_g \sim N$) for long PAs [33]. The $N$ dependence in Fig. 4 has higher power than the result in [2] ($R_g \sim N^{1/3}$). It appears that for values of $\sigma$ observed in this set of IDPs the random coil behavior is the apt description.

As an additional application to a class of related IDPs, we show in Fig.(5) the chain length dependence of the radius of gyration of wild type Tau and a number of fragments. The $R_g$ values for these sequences have been determined by SAXS experiments [22]. We find that $R_g \sim N^{1/2}$, which differs from the random coil scaling for the set of IDPs whose $R_g$ values are plotted in Fig.(4). Simulations based on the SOP-IDP model leads to $N^{0.5}$ for the set of IDP sequences listed in Fig. 5 [23].

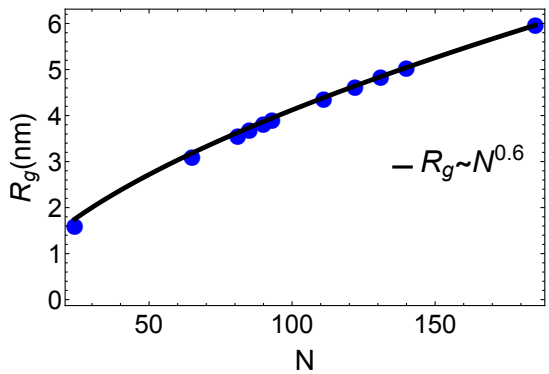


FIG. 4: The size $R_g$ increases with $N$ as $R_g \sim N^{0.6}$ for specific contacts (ACTR (N=65), An16 (N=185), aSyn (N=140), ERMnTAD (N=122), HISTATIN5 (N=24), hNHE1 (N=131), NUP153 (N=81), p53 (N=93), ProtA (N=111), SH4UDsrc (N=85) and Sic1 (N=90)). The scaling is same even when the restriction to specific interactions in Eq. 6 is removed. The parameters for the plot are $\sigma = 0.01$, $\delta\sigma = 0.54$, and $\kappa l_B = 0.28$.

## IV. CONCLUSIONS:

We developed a theory to quantitatively predict the effect of charge fluctuations ($\delta\sigma$) on the size of flexible PAs that is in a good solvent (excluded volume interactions are positive) as a function of the inverse Debye length and the net charge per monomer on the chain ($\sigma$). Interestingly, when charge fluctuations are non-negligible ($\frac{\delta\sigma}{\sigma}$ is greater than unity), the radius of gyration increases non-monotonically as $\kappa$ increases. When $\frac{\delta\sigma}{\sigma}$ is less than unity, $R_g$ increases with increasing $\kappa$. The generality of the theory allows us to predict $R_g$ for a number of IDPs. For a certain class of IDPs, we find the usual scaling of $R_g \sim N^{\nu}$ with $\nu = 0.6$, which coincides with the behavior expected for Flory random coils. Remarkably, our theory

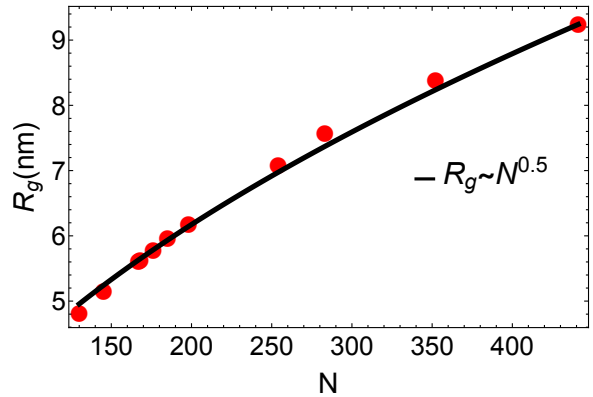


FIG. 5: The size $R_g$ increases with $N$ as $R_g \sim N^{0.5}$ for wild type Tau and fragments constructed from it. The various IDPs are (K18Tau ($N$=130), K17Tau ($N$=145), K27Tau ($N$=167), K10Tau ($N$=168), K32Tau ($N$=198), K44Tau ($N$=283), K23Tau ($N$=254), K25Tau ($N$=185), hTau23 ($N$=352), hTau40 ($N$=441) and K16Tau ($N$=176)). The parameters used to generate the plot are $\sigma = 0.01$, $\delta\sigma = 0.54$ and $\kappa l_B = 0.28$.

gives accurate estimates of the size of the Tau protein, and various fragments derived from it. This class of IDPs behaves as an ideal chain. The differences in the two scaling behavior between these IDPs can be rationalized in terms of the interplay between charge fluctuations and net charge per monomer.

What could be the origin of charge fluctuations in an IDP in which $\sigma$ (more precisely the precise sequence) is fixed? Even with $\sigma$ fixed, the ensemble of conformations that a typical IDP samples is heterogeneous, as implied by the $R_g$ scaling in Figs. 4 and 5. In sampling a large number of conformations the spatial distances between charged residues could vary greatly. Therefore, the effective charge of each conformation is different. In some of the conformations, positively and negatively charged residues would be , whereas in others they would be spatially well-separated. This gives rise to conformation-dependent effective attraction, which is quantified in our theory in terms of the average quantity $\delta\sigma$. Of course, the effective value of $\delta\sigma$ cannot be computed for a quenched charged sequence for a specific IDP without a suitable simulation. Therefore, it is difficult to construct phase diagrams of IDPs solely in terms of $\sigma$ or the differences between the number of positively and negatively charged residues. Construction of phase diagrams requires use of physical order parameters (here we used $R_g$), which necessarily involves quantitatively characterizing the conformations of the IDPs.

**Acknowledgements:** We are indebted to Upayan Baul and Debayan Chakraborty for providing the simulation results and for useful discussions. We thank Prof. D. Svergun for providing us SAXS data on the Tau protein. This work was supported by the National Foundation (CHE 16-36424) and the Collie-Welch Foundation (F-0019).

[1] S. F. Edwards, P. R. King, and P. Pincus. Phase changes in polyampholytes. *Ferroelectrics*, 30(1):3–6, 1980.

[2] P. Higgs and J.-F. Joanny. Theory of polyampholyte solutions. *J. Chem. Phys.*, 94:1543, 1991.

[3] D. Srivastava and M. Muthukumar. Sequence dependence of conformations of polyampholytes. *Macromolecules*, 29:2324–2326, 1996.

[4] J.L. Barrat and J.-F. Joanny Joanny. *Theory of Polyelectrolyte Solutions*, pages 1–66. Wiley-Blackwell, 2007.

[5] I. Borukhov, D. Andelman, and H. Orland. Random polyelectrolytes and polyampholytes in solution. *Eur. Phys. J. B - Condensed Matter and Complex Systems*, 5(4):869–880, Nov 1998.

[6] A. V. Dobrynin and M. Rubinstein. Flory theory of a polyampholyte chain. *J. Phys. II France*, 5(5):677–695, 1995.

[7] A. V. Dobrynin, R. H. Colby, and M. Rubinstein. Polyampholytes. *Journal of Polymer Science Part B: Polymer Physics*, 42(19):3513–3538, 2004.

[8] N. Lee and D. Thirumalai. Dynamics of collapse of flexible polyampholytes. *J. Chem. Phys.*, 113(13):5126–5129, 2000.

[9] V. Yamakov, A. Milchev, H. J. Limbach, B. Dünweg, and R. Everaers. Conformations of random polyampholytes. *Phys. Rev. Lett.*, 85:4305, 2000.

[10] AM Gutin and EI Shakhnovich. Effect of a net charge on the conformation of polyampholytes. *Phys. Rev. E*, 50(5):R3322–R3325, 1994.

[11] P. E. Wright and H. J. Dyson. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biology*, 16:18–29, 12 2015.

[12] M. van der Lee, R.and Buljan, B. Lang, R. J. Weatheritt, G. W. Daughdrill, A. K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D. T. Jones, P. M. Kim, R. W. Kriwacki, C. J. Oldfield, R. V. Pappu, P. Tompa, V. N. Uversky, P. E. Wright, and M. M. Babu. Classification of Intrinsically Disordered Regions and Proteins. *Chem. Rev.* , 114(13):6589–6631, 2014.

[13] C. J. Oldfield and A. K. Dunker. Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions. In Kornberg, RD, editor, *Ann. Rev. Biochem.*, volume 83 of *Annual Review of Biochemistry*, pages 553–584. 2014.

[14] RI Dima and D Thirumalai. Proteins associated with diseases show enhanced sequence correlation between charged residues. *Bioinformatics*, 20(15):2345–2354, 2004.

[15] Rahul K. Das, Kiersten M. Ruff, and Rohit V. Pappu. Relating sequence encoded information to form and function of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.*, 32:102–112, 2015.

[16] W. Zheng, A. Borgia, K. Buholzer, A. Grishaev, B. Schuler, and R. B. Best. Probing the Action of Chemical Denaturant on an Intrinsically Disordered Protein by Simulation and Experiment. *J. Am. Chem. Soc.*, 138(36):11702–11713, 2016.

[17] B. Schuler, A. Soranno, H. Hofmann, and D. Nettels. Single-Molecule FRET Spectroscopy and the Polymer Physics of Unfolded and Intrinsically Disordered Proteins. *Ann. Rev. Biophys.*, 45:207–231, 2016.

[18] Z. A. Levine and J.-E. Shea. Simulations disordered proteins and systems with conformational heterogeneity. *Curr. Opin. Struct. Biol.*, 43:95–103, 2017.

[19] V. N. Uversky. Natively unfolded proteins: A point where biology waits for physics. *Protein Sci.*, 11:739–756, 2002.

[20] T. Firman and K. Ghosh. Sequence charge decoration dictates coil-globule transition in intrinsically disordered proteins. *J. Chem. Phys.*, 148:123305, 2018.

[21] R. K. Das and R. V. Pappu. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci.*, 110:13392–13397, 2013.

[22] E. Mylonas, A. Hascher, P. Bernado, M. Blackledge, E. Mandelkow, and D. I. Svergun. Domain conformation of tau protein studied by solution small-angle X-ray scattering. *Biochemistry*, 47(39):10345–10353, 2008.

[23] U. Baul, D. Chakraborty, M. L. Mugnai, and D. Thirumalai. A sequence-speci c two bead per residue model (sop-idp) for intrinsically disordered proteins. *Unpublished*, 2018.

[24] B.-Y. Ha and D. Thirumalai. Persistence length of intrinsically stiff polyampholyte chains. *J. Phys. II France*, 7:887–902, 1997.

[25] B.-Y. Ha and D. Thirumalai. Conformations of a polyelectrolyte chain. *Phys. Rev. A*, 46:R3012–R3015, Sep 1992.

[26] S. F. Edwards and P. Singh. Size of a polymer molecule in solution. part 1.—excluded volume problem. *J. Chem. Soc., Faraday Trans. 2*, 75:1001–1019, 1979.

[27] M. Muthukumar and S. F. Edwards. Extrapolation formulas for polymer solution properties. *The Journal of Chemical Physics*, 76(5):2720–2730, 1982.

[28] M Muthukumar. Adsorption of a polyelectrolyte chain to a charged surface . *J. Chem. Phys.*, 86(12):7230–7235, 1987.

[29] B.-Y. Ha and D. Thirumalai. Persistence length of flexible polyelectrolyte chains. *The Journal of Chemical Physics*, 110(15):7533–7541, 1999.

[30] H. S. Samanta, P. I. Zhuravlev, M. Hinczewski, N. Hori, S. Chakrabarti, and D. Thirumalai. Protein collapse is encoded in the folded state architecture. *Soft Matter*, 13:3622–3638, 2017.

[31] N Lee and D Thirumalai. Dynamics of collapse of flexible polyelectrolytes in poor solvents. *Macromolecules*, 34(10):3446–3457, 2001.

[32] S. Mueller-Spaeth, A. Soranno, V. Hirschfeld, H. Hofmann, S. Rueegger, L. Reymond, D. Nettels, and B. Schuler. Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proc. Natl. Acad. Sci.*, 107(33):14609–14614, 2010.

[33] Y. Kantor and M. Kardar. Polymers with random self-interactions. *Europhys. Lett.*, 14:421, 1991.