Shape fluctuations of random polyampholyte and intrinsically disordered protein sequences

Min-Kyung Chae, Nam-Kyung Lee, Youngkyun Jung, A. Johner

Shape fluctuations of random polyampholyte and intrinsically disordered protein sequences

Min-Kyung Chae,[†,‡] Nam-Kyung Lee,[*,†] Youngkyun Jung,[*,¶] and Albert Johner[*,§]

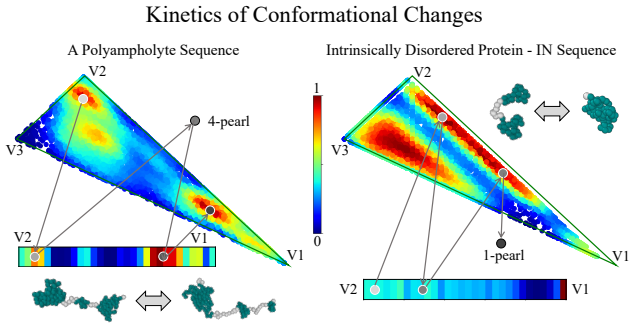
†*Department of Physics and Astronomy, Sejong University, Seoul 05006, Korea*

‡*National Institute for Mathematical Sciences, Daejeon 34047, South Korea*

¶*Supercomputing Center, Korea Institute of Science and Technology Information, Daejeon 34141, Korea*

§*Institut Charles Sadron CNRS-Unistra, Universite de Strasbourg, F-67000 Strasbourg, France*

E-mail: lee@sejong.ac.kr; yjung@kisti.re.kr; albert.johner@ics-cnrs.unistra.fr



Abstract

Charged polymers with quenched charge sequences adopt pearl-necklace structures due to the interplay between charge disorder and thermal fluctuations. Partially globular pearl necklace conformations from random (+/−) sequences are intrinsically heterogeneous and exhibit various structures characterized by different number of pearls.

Based on the molecular dynamics simulations of random model polyampholytes (PAs), we study dynamics of pearls and associated conformational heterogeneity in pearl structures. Fast nucleation/dissolution of small pearls controls the dynamics of the number of pearls, which is well described by first order kinetic equations. Most of individual sequences we considered have a rather stable number of large pearls. Only a few sequences allow for fast switchings in the number of large pearls following complex processes (trajectories in parameter space) going through states with different number of pearls. Processes are most complex in the tail of the switching time distribution. The specific sequence of two Intrinsically Disordered Proteins (IDPs) is studied along the same lines. Our study shows the resemblance in shape fluctuations between the IDP sequences and the fast-switching PA sequences.

Introduction

Water soluble polymers can be manipulated without organic solvent and are considered as environmental friendly. Most water soluble polymers are electrically charged and are called polyelectrolytes (PEs). Polymers comprising charged monomers inserted along their backbone can be obtained by chemical modification of an otherwise insoluble backbone. Alternatively a copolymer of neutral and charged monomers can be synthesized.[1] Polyelectrolytes are widely used in colloid science,[2] for membranes,[3] to obtain polyanion–polycation complexes (coacervates) as bulk materials[4–6] or at a solid/liquid interface.[7] Among charged polymers there is a special class called polyampholytes (PAs).[8] PAs carry charges of both signs and can form coacervates[9,10] Most proteins are PAs. A special class of proteins having no well-defined ground state and called Intrinsically Disordered Proteins (IDPs) share many common features with synthetic PAs.[11–15] The conformation of synthetic PAs (without hydrogen bonds) results from a balance between local attraction of opposite charges promoting collapse and global repulsion due to the net charge (if any) promoting overall extension. Both PAs and hydrophobic PEs, where the local attraction stems from effective short range attraction between monomers avoiding the water molecules, can be described by the pearl-necklace model introduced in the nineties.[16,17] In this model, collapsed pearls satisfying local attraction are separated by stretched strings satisfying global repulsion. It was recognized soon that a regular hydrophobic PEs is usually not accurately described by the ground-state but includes large fluctuations within the necklace affecting the size and number of pearls.[18] In the case of synthetic PAs where the sequence of charges is not controlled but random, the disorder along the sequence additionally contributes to the distribution of conformations for a set of sequences. The first simulation study taking ensemble averages over random PAs addressed the geometrical polymer size through the averaged square radius of gyration.[19] Scaling arguments illustrated by numerical simulations showed that blockiness of the sequence controls the overall PA structure.[20] In an early study for long PAs with all monomers charged, a variety of structures with complex topologies was found.[21] In our own studies of shorter partially charged PAs necklaces are rather regular and only rare double bridges were found.

In a recent simulation study[22] we showed that pearl-necklaces with asymmetric pearls are most often encountered with a set of preferred conformations counting one large pearl and a few small ones of comparable size. (See, Fig. 1). Necklaces with two large pearls are less numerous while necklaces with all beads similarly-sized are hardly found. The previous study was limited to chains counting $N = 202$ monomers, of which 68 monomers are charged with either sign at regular intervals.

In the present work we study the structural heterogeneity residing in random PA and IDP sequences. We classify randomly generated PA sequences according to their number of pearls and further analyze subclasses with similar asymmetry (number of large pearls). Most sequences essentially populate one subclass. Only a few spread over more subclasses. We then investigate the pearl dynamics in the necklace structures at equilibrium. This is at variance with the formation of pearls during the collapse transition triggered by a change in external conditions.[23,24] The fluctuation of the number of pearls is first studied qualitatively by tracking each sequence in the parameter space. While the number of small pearls rapidly fluctuates, the number of large pearls is much more stable. In a second step, the statistics of the number of pearls is described by a set of transition rates. The explanation for the slower fluctuation in the number of large pearls is established mainly based on a relatively fast fluctuating subset of sequences for which we have better statistics. Finally, our findings are discussed in a somewhat larger frame encompassing the well-defined sequences of IDPs which share many features of PAs. It turns out that IDP sequences behave similarly to the few PA sequences that undergo large shape fluctuations.

Simulation Description

In the molecular dynamics simulations, we consider IDP sequences and single PA chains with random charge correlations in a weakly poor solvent.

A PA is modeled as a charged bead-spring chain consisting of N monomers including $(N + 2)/3$ charged monomers. Every third monomer out of $N = 202$ monomers carry a charge ± 1 with an unbiased statistics satisfying global average of $\langle Q \rangle = 0$. As such, all end monomers carry charges. When all sites are positively charged, the corresponding net charge

$Q_{\max}$ is 68.[22] Each Q ensemble consists of those sequences satisfying the given net charge condition. For each net charge $Q = 16$, 20, and 24, data of 40 - 45 independent sequences are considered in a dilute solution at concentration $c = 2.02 \times 10^{-4}\,\sigma^{-3}$.

We also consider two IDPs in near Θ-solvent condition. IDP is modeled as a PA consisting of N monomers at concentration $c = 6.41 \times 10^{-3}\,\sigma^{-3}$. The charge sequence of monomers and spacing between charges are faithfully taken from the residues of the corresponding IDP. In particular, IN and ProTα are considered. IN consisting of 56 residues has the net charge of -4 where 16 monomers carry charges. ProTα consisting of 110 residues has the net charge of -43 and there are 63 charged monomers. The sequence informations are given in Table 1.

The electroneutrality is imposed by counterions compensating the net charge on the PA backbone with no added salt. Counterion condensation effect does not play a role here.[22]

The excluded volume interactions are modeled by the truncated-shifted Lennard-Jones(LJ) potential: $U_{\rm LJ}(r) = 4\epsilon_{\rm LJ}[(\sigma/r)^{12} - (\sigma/r)^6 - (\sigma/r_c)^{12} + (\sigma/r_c)^6]$ for $r < r_c$ and 0 elsewhere. Here $\epsilon_{\rm LJ}$ represents the strength of the LJ potential and r denotes the center-to-center distance between two interacting particles. In simulations of random PAs, the value of the interaction parameter is set to $\epsilon_{\rm LJ} = 0.6\,k_{\rm B}T$ and the cutoff distance r_c is set to 2.5σ for monomer-monomer interactions so that polymers are in moderately poor solvent condition. For monomer–counterion and counterion–counterion interactions, we set the interaction parameter $\epsilon_{\rm LJ} = 1.0\,k_{\rm B}T$ and the cut off distance $2^{1/6}\sigma$, which leads to the interactions purely repulsive. In simulations of IDP sequences, the value of interaction parameter is set to $\epsilon_{\rm LJ} = 0.4\,k_{\rm B}T$ for monomer–monomer interactions so that IDPs are in near Θ-solvent condition. Other parameters remain the same as the PA model.

Two charged particles interact via the Coulomb potential $U_{\rm C}(r) = k_{\rm B}T l_{\rm B} z_i z_j / r_{ij}$, where z_i and z_j are the charge valence of particle i and j, respectively. The strength of the electrostatic interactions is determined by the Bjerrum length and it is set to $l_{\rm B} = 3\ \sigma$ and $1.5\ \sigma$ for PA and IDP, respectively. The long-range electrostatic interactions are calculated by the particle-particle-particle-mesh (PPPM) method implemented in LAMMPS software package.[25]

The chain connectivity is ensured by the finite extension nonlinear elastic (FENE) potential between two consecutive beads[22] and the equilibrium bond length is 1.0σ.

In order to explore the phase space, we integrated Newton's equations of motion using the velocity Verlet algorithm with an integration time step $\delta t = 0.01\tau$, where LJ time $\tau = \sigma(m/\epsilon_{\rm LJ})^{1/2}$ is the characteristic time scale with bead mass $m = 1$. A Langevin thermostat with the damping constant $1.0\tau^{-1}$ was used to keep the system at the fixed temperature $T = 1.0\ \epsilon_{\rm LJ}/k_{\rm B}$.

We first performed 6×10^6 integration steps (which is equivalent to $6 \times 10^6\ \delta t = 6 \times 10^6 \times 0.01\,\tau = 6 \times 10^4 \tau$) in order for the mean square radius of gyration of the chain to relax to its equilibrium values. After equilibration, we ran additional $4 \times 10^7 (= 4 \times 10^5 \tau)$ integration steps and collected data every $10^4 (= 10\,\tau)$ time steps.

To identify pearl-necklace structures, we need to determine whether a given monomer belongs to a globular pearl or to a string. We first apply the algorithm suggested by Liao *et al.*[26] We make lists of monomers belonging to globules such that the number of neighboring monomers within the sphere of cutoff radius $R_{\rm cut} = 2.3\,\sigma$ is larger than $N_{cut} = 8$. Other monomers are considered as parts of strings. The second step introduces a merge process for clusters if their closest distance is smaller than $r_{min} = 1.5\sigma$. These procedures are repeated until the lists no longer change. In the figures below, monomers belonging to globules are shown in green and the other ones in white. The adopted criterion works well for distinguishing monomers belonging to a globular phase.

Results

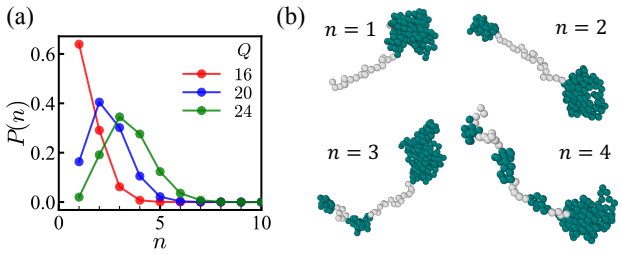


Figure 1: (a) The probability distributions of n-pearl states $P(n)$ for ensembles of various net charges $Q = 16$, 20, and 24. (b) Snapshots show typical structures of pearl-necklaces with 1- 4 pearls from sequences of $Q = 20$. Monomers belonging to the globules are shown in green.

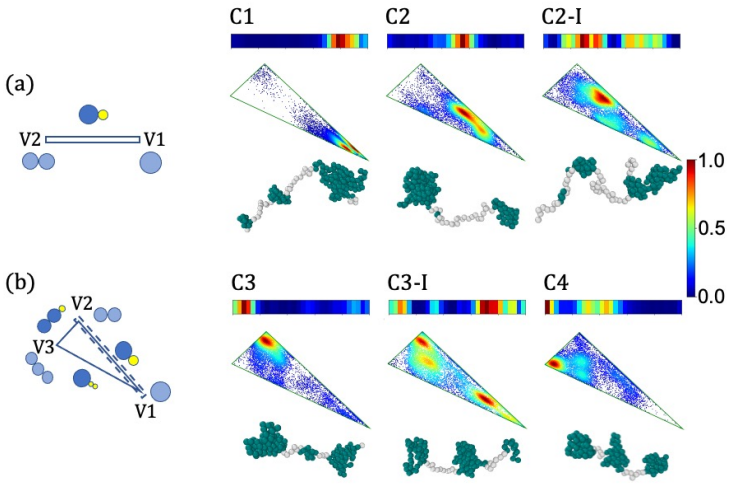


Figure 2: (left) The segmental diagram of 2-pearl state (a) and triangular diagram of 3-pearl state (b). Vertex Vn_l represents the number of large pearls n_l. (right) Classification of PA sequences. 2-pearl/3-pearl structures are shown for representative single sequences of each class. The shapes are classified based on their 3-pearl structures: (C1) one large pearl, (C2) two uneven large pearls, (C3) two comparable large pearls, and (C4) three large pearls. C2-I and C3-I are variations from C2 and C3 types, respectively and the secondary structures are found in addition to the most populated structures of C2 and C3. Sequences are given in Table 1. The net charge is $Q = 20$ except for C4 where $Q = 24$. The color bar indicates the density of states in [0,1] scale. Each snapshot is representative of the most visited regions for the studied sequence.

Equilibrium Statistics

We obtained configurations from ensembles of $Q = 16$, 20 and 24, with a time resolution of 10 τ. The structures are analyzed according to the pearl-necklace model. For each Q ensemble, the probability distribution of n-pearl states $P(n)$ is very dispersed and the most common number of pearls increases with Q, (Fig. 1). A necklace contains n_l large pearls and n_s small pearls. For a given number of pearls $n = n_l + n_s$, the pearl sizes usually vary widely but not randomly so. The global statistics is dominated by configurations with one large pearl and much smaller pearls of similar size. (Fig. 1 (b)). Configurations where one small pearl is replaced by a somewhat larger one are also common. Other configurations are considerably less abundant, less so for large net charge Q. Each sequence exhibits thermal fluctuations markedly smaller than the thermal fluctuation of the ensemble of quenched sequences with the exception of a few rare, but interesting, sequences. In order to characterize the PA structure, we classify sequences into subclasses, according to their configurational

fluctuations, here mainly in the two and three pearl states (see, Fig. 2). The two and the three pearl states lead to similar conclusions. Almost all sequences sampled can be clearly attributed to one of the four classes: C1, C2, C3 and C4 shown in Fig. 2. For the ensemble of $Q = 20$, only about 6% of sequences spread a lot in configuration space (e.g. C2-I, C3-I) and are subject to large thermal fluctuations between very different shapes occurring with similar probability. It turned out that frequent changes of the number of pearls mostly involves fluctuation of small pearls for each of the four sequence classes.

For a given number of pearls, we adopt the simplex representation introduced earlier in Ref.[22] as shown in Fig. 2. First the n-pearls are ordered by decreasing size ($i = 1, 2, 3, ...n$) and the mass distribution among pearls N_i ($\Sigma_i N_i = N_p$) is characterized by the set of coordinates $x_i = N_i/N_p - 1/n$ measuring the difference between the even share fraction $1/n$ and the actual share fraction taken by pearl i. By construction the sum of all x_i vanishes ($\Sigma x_i = 0$) and we discard the smallest one. The ranking imposes $n-1$ extra linear constraints among the x_i. Also, the smallest pearl must have a positive mass, which adds another linear constraint. For an n-pearl structure, the size-space is a $(n-1)$-simplex. For two pearls, the size-space reduces to the segment $x_1 \in [0, 0.5)$ (Fig. 2). For three pearls the parameter space x_1, x_2 reduces to a triangle (Fig. 2) and the three vertices, V1,V2, and V3, correspond to one large pearl, two and three equal size large pearls, respectively. In the vicinity of a vertex, inside the triangle, the distribution is complemented by small pearls. The sides of the triangle correspond to two equal large pearls, two equal small pearls, the 2-pearl limit, respectively (Fig. 1 (c) and Fig. 2). Vertices are labeled as Vn_l with the number of large pearls n_l. The line connecting V1 and V2 in 3-pearl diagram corresponds to the 2-pearl state. In Ref.,[22] it is suggested that the energetically most favorable configurations comprise one large pearl and that the distribution spreads with increasing net charge Q. Many examples of the segmental (two pearls) and triangular (three pearls) representation will be shown (Figs. 2, 4, 5, 6, 7).

Dynamics of the number of peals n in the necklace.

The survival probability $\tilde{P}_n(t)$ of the n-pearl structure has been evaluated from the simulations for various number of pearls including the single globule state. The half time $t_{1/2,n}$,

Table 1: Sequence informations of selected sequences of PA and IDP. The net charge of PA sequences is $Q = 20$ except for C4 where $Q = 24$. Positive/negative charges are indicated as 1 and -1. Neutral charges are represented as zeros only for IDP sequences which have irregular spacings between charges. Some large blocks (≥ 6) are indicated in red.

	sequences
C1	-11-1-11-11-11-1 1-111-11-1111 11-1-111111-1 1111-1 11-1-1-1 1-11-11-11-1-11 11-11111111 -111-11111
C2	111-11-111-1-1 11-11111111 -111-1-11-1-111 -11-1111-1-1-1-1 111-1111-111 11111-11-111 1-111-1-11-1
C2-I	11-1-111111-1 11-11111-1-1-1 11-11-1-1-1111 -1-1111-1-1111 1-1111-11111 -1-111-111111 1 -11-111-1-1
C3	-1111111-1-1-1 11-111111-1-1 1-1-111-1-1111 -11-11111111 11-1-111-1-1-11 1111-1-1-111-1 111-11-11 1
C3-I	-111111-1111 1-1-1-111-11-11 -11111-11-1-11 111-11-11111 1-1-1-111-1-1-1-1 11-11111111 11111 -1-1 -1
C4	11111-1-1-11-1111-111-111111 1-111-1-11-11-111111-1-11-111111 111-11111-1-11-11-1-11-1-11111
IN (N=56)	00-100-1100-1 -10-11000001 0000-100000 0001-100000 -1100010-100 0000-10
ProT$\alpha(N = 110)$	00-1000-1000 -10001-101-11 1-100-1-10-100 1-100000000 -1-100-10-10-10 -10-1-1-1-1-1-100 -1-1-1-1-1-1-1-10-1 0-1-1-1-10-1-1-1-1 -10-10000110 0-1-1-1-1-1-1-10-1 011010-1-1-1-1

the time required for the initial population $\tilde{P}_n(t = 0)$ to be reduced by half, was measured for each Q ensemble (see, Table 2 (a)). As shown in Fig. 1, $Q = 20$ offers good statistics for all values of n considered, but $Q = 16$ and Q =24 less so. When small number of pearls are considered (n =1 and 2), Q =16 and 20 show similar half times but the corresponding values of $Q = 24$ are smaller. For $n = 3$ and 4, half times of $Q = 20$ and $Q = 24$ are similar. We also measure the characteristic times t_d of the extended quasi-exponential tail in the survival probability: $\tilde{P}_n(t) \sim e^{-t/t_d}$. The inverse values of the measured relaxation times t_d^{-1} are shown in Table 2 (b). The single globule state shows markedly longer relaxation time than necklaces. The relaxation time is shorter for a higher number of pearls and bridges. Nonetheless the associated frequency is not just proportional to the number of bridges (pearls). Under given time resolution of 10τ, many switching events in $Q = 24$ ensemble remain unresolved in the simulation and this makes the estimates for relaxation times and half-life times inaccurate. Due to the occurrence of small pearls which can form/dissolve easily on pre-existing strings or tails the half times and relaxation times of a given number of pearls is indeed expected to be short, as observed. We also plot the instantaneous values of the inverse relaxation time ${t_{d,n}}^{-1}(t) = \frac{1}{\tilde{P}_n(t)} \frac{d\tilde{P}_n(t)}{dt}$ in Table 2 (c). The number of switching events at short times are significantly larger than the stationary values for small n. The rate saturates to the stationary value, t_d^{-1}.

Table 2: (a) Half times $t_{1/2}$ and (b) inverse relaxation time ${t_d}^{-1}$ of n-pearl states for each Q ensemble. Having the time resolution of 10 τ, the half time falls $\lesssim 20\tau$ in units of LJ time τ. The values are extracted by (linear) extrapolation. (c) Instantaneous values of transition rate ${t_{d,n}}^{-1}(t)$ leaving from n-pearl state as a function of time. Symbols represent net charge of PAs, $Q = 16$ (red $\circ$), 20 (blue $\square$), and 24 (green $\triangle$).

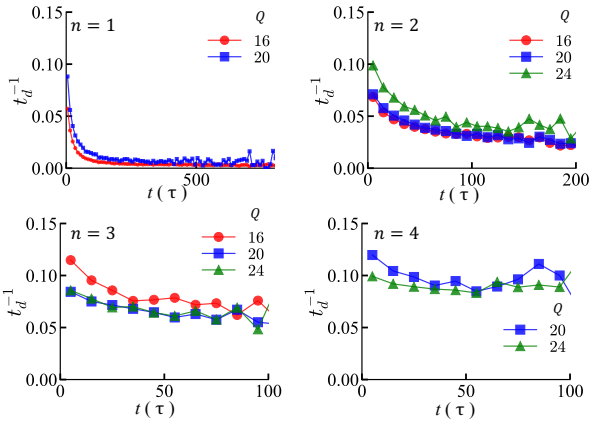
(a)

$t_{1/2,n}$	n =1	$n = 2$	$n = 3$	$n = 4$
Q =16	13.2	9.8	6.9	5.7
$Q = 20$	13.9	9.5	8.4	6.7
$Q = 24$	6.3	7.6	8.2	7.5

(b)

${t_d}^{-1}$	n =1	$n = 2$	$n = 3$	$n = 4$
Q =16	0.003	0.032	0.077	0.133
$Q = 20$	0.007	0.024	0.076	0.105
$Q = 24$	0.004	0.043	0.070	0.090

(c)



One way to characterize the dynamics of pearl number distribution is through the transition rates. Rates between n-pearl state and $(n \pm 1)$-pearl state can be studied from the simulation data. Transition rates are approximated as constants. To put this approximation to a test we try to reproduce the evolution of the pearl populations starting out from an out of equilibrium distribution.

We construct master equations for the populations of n-pearl states, P_1, P_2, P_3..., with a fixed number of pearls smaller than the cut off n_{max}. The n_{max} is 6, 9, and 11 for $Q =$ 16, 20 and 24 ensembles, respectively. Populations with a higher number of pearls ($n > n_{max}$) are negligible. In table 3, we show ω_{m_2,m_1} values for $m_1, m_2 \leq 6$. The full data is

available as SI. The transition rates ω_{m_2,m_1} are obtained from the number of switching events detected in the simulation at equilibrium. If the time resolution used in the detection is small enough, transitions between non-adjacent states must vanish. This yields a first accuracy test. Encouraged by the data, we assume a constant transition rate over the whole time (See the discussion of the time dependence of the apparent $\omega(t) = t_d^{-1}(t)$, Table 2).

We count the total number of events Ω_{m_2,m_1} switching from the m_1-pearl state to the m_2-pearl state for each Q ensemble. The number of switching between m_1 and m_2 is (almost) symmetric, confirming the detailed balance at equilibrium (at least for the retained populations), (See, SI). The events are counted for $3.5 \times 10^5 \tau$ ($Q = 16$ and 20) and $3.0 \times 10^5 \tau$ ($Q = 24$) with time resolution of 10τ. The mean transition rate ω_{m_2,m_1} computed from the total number of events as $\omega_{m_2,m_1} = \tilde{\Omega}_{m_2,m_1}/P_{m_1}$ where $\tilde{\Omega}$ is the number of events per unit time averaged by the number of the sequences. The values of ω_{m_2,m_1} are listed in Table 3.

The master equations describing the pearl number dynamics with constant switching rates can be written:

$$\begin{aligned}
\frac{dP_1(t)}{dt} &= -\omega_{21}P_1(t) + \omega_{12}P_2(t) \\
\frac{dP_2(t)}{dt} &= \omega_{21}P_1(t) - (\omega_{12} + \omega_{32})P_2(t) + \omega_{23}P_3(t) \\
\frac{dP_n(t)}{dt} &= \omega_{n,n-1}P_{n-1}(t) - (\omega_{n-1,n} + \omega_{n+1,n})P_n(t) + \omega_{n,n+1}P_{n+1}(t) \\
\frac{dP_{n_{max}}(t)}{dt} &= \omega_{n_{max},n_{max}-1}P_{n_{max}-1}(t) - \omega_{n_{max}-1,n_{max}}P_{n_{max}}(t). \qquad (1)
\end{aligned}$$

To construct the master equation Eqs. 1, we consider switchings $\{\omega_{m_2,m_1}\}$ satisfying $m_2 = m_1 \pm 1$. Unresolved switching events $m_2 = m_1 \pm 2, 3$ turn out to be rare for $Q = 16$ less so for $Q = 20$ but not at all for $Q = 24$ where the determination of the rates $m_2 = m_1 \pm 1$ for the master equation is expected to be inaccurate.

The Fig. 3 shows time evolution of the n-pearl populations $P_n(t)$ for Q =16 and 20. The evolution of the populations $P_n(t)$ with a chosen initial condition (e.g. 100% in n =1) is directly extracted from the simulation data using the ensembles satisfying the initial condition. This evolution is compared to the solution of the corresponding master equation based on the simple, constant, transitions rates.

Table 3: The mean transition rate ω_{m_2,m_1} for (a) $Q = 16$ (b) $Q = 20$ (c) Q =24. The index $m_1(m_2)$ corresponds to row (column). The switching rates used for the master equations (Eq. 1) are highlighted by colors.

		m_2					
(a) $Q = 16$		1	2	3	4	5	6
m_1	1	-	0.012	0.001	0.000	0.000	0.000
	2	0.026	-	0.010	0.001	0.000	0.000
	3	0.010	0.039	-	0.004	0.000	0.000
	4	0.004	0.032	0.036	-	0.002	0.000
	5	0.002	0.020	0.038	0.021	-	0.000
	6	0.000	0.001	0.003	0.003	0.002	-

		m_2					
(b) $Q = 20$		1	2	3	4	5	6
m_1	1	-	0.032	0.007	0.001	0.000	0.000
	2	0.013	-	0.027	0.005	0.001	0.000
	3	0.004	0.036	-	0.015	0.003	0.000
	4	0.001	0.020	0.043	-	0.007	0.001
	5	0.001	0.011	0.035	0.035	-	0.003
	6	0.000	0.006	0.025	0.037	0.022	-

		m_2					
(c) $Q = 24$		1	2	3	4	5	6
m_1	1	-	0.439	0.219	0.061	0.010	0.001
	2	0.045	-	0.374	0.147	0.031	0.004
	3	0.013	0.207	-	0.266	0.081	0.015
	4	0.004	0.101	0.334	-	0.161	0.041
	5	0.002	0.048	0.227	0.361	-	0.090
	6	0.001	0.025	0.145	0.316	0.305	-

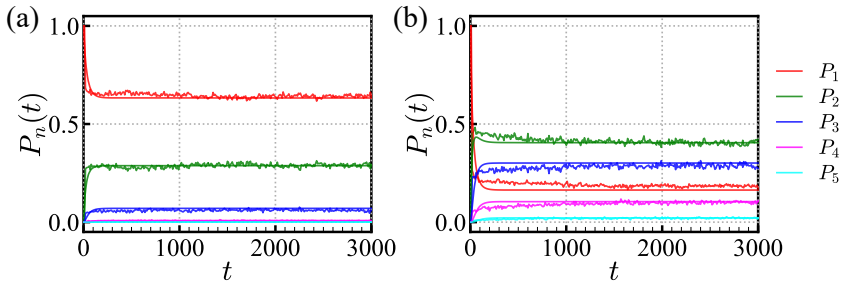


Figure 3: Evolutions of n-pearl state populations $P_n(t)$ from the ensembles for (a) Q =16 and (b) $Q = 20$. We set the initial configuration with only $n = 1$ so that $P_1(t = 0) = 1$. Master equation results using values of w_{m_2,m_1} in Table 3 are shown as solid lines. Each population converges to the equilibrium statistics of n-pearls shown in Fig. 1 (a).

Data from the simulation and master equation compare favorably for $Q = 16$ and $Q = 20$. (Fig. 3) The master equation with the *constant* transition rates captures the long time relaxation correctly but misses the short time evolution ($t \lesssim 10\tau$). .

Shape fluctuations and the relaxation of large pearls

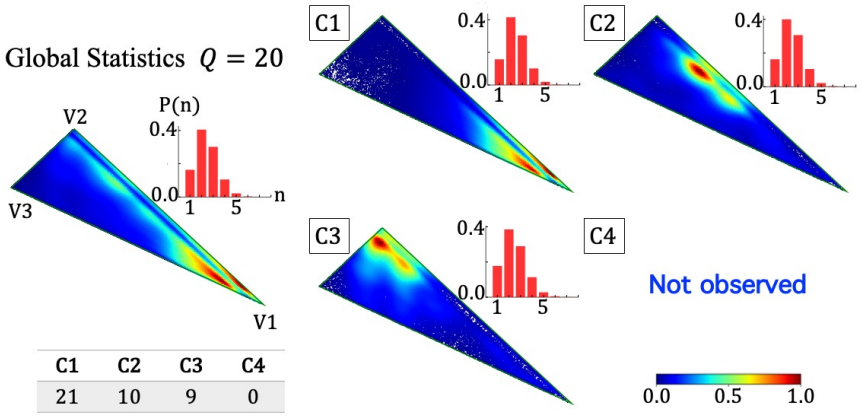


Figure 4: 2-simplex $\{x_1, x_2\}$ (3-pearl state) representations for each PA type in $Q = 20$ ensemble. The color code indicates the density of visited states on a scale of 0 to 1. Accumulated probability distributions of n-pearl state $P(n)$ are shown as red histograms for each group. The sequences are classified into 3 different types depending on the pattern of visited points in 2-simplex. Each type takes 52.5%, 25% and 22.5%. The global ensemble of 40 sequences (left panel) has the most visited state closer to the vertex of $V1$ where a single large pearl dominates. The majority are of C1 type of which population density is similar to global statistics. C1 type and C3 type have the most visited state near $V1$ and $V2$, respectively. Most of conformations of C2 type is along a line connecting vertex $V1$ and $V2$. No C4 type pattern is found.

To illustrate the structural heterogeneity we mainly refer to the 3-pearl states and distinguish several types of sequences according to the patterns of the triangular diagram. What is surprising is the high heterogeneity of the distribution of configurations inside the triangle where most of the area is very depleted. Relaxation dynamics of the pearl number n is dominated by the fluctuation of small pearls. We are interested in the large-scale fluctuations that accompany changes in the number of large pearls n_l and in the types of sequences in which these changes easily occur. A large shape fluctuation would be, for example, between those conformations involving one large pearl and those involving two large pearls (e.g. changing corner from $V1$ to $V2$ or vice versa in the triangular representation). This

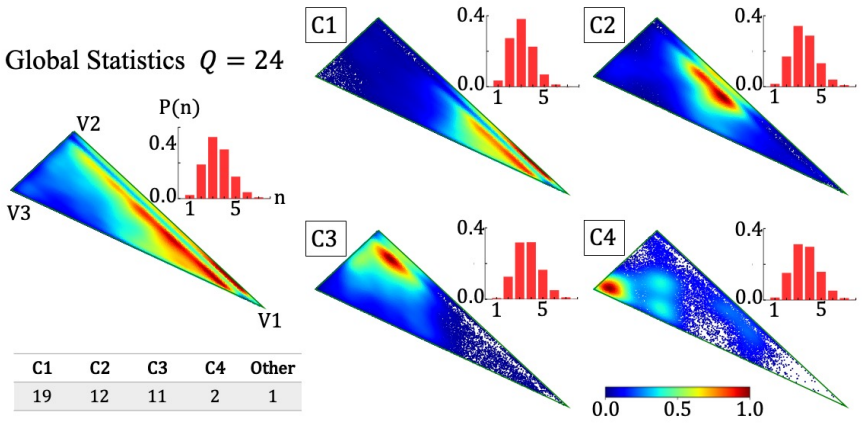


Figure 5: 2-simplex $\{x_1, x_2\}$ (3-pearl state) representations for each PA type in $Q = 24$ ensemble. The color code indicates the density of visited states on a scale of 0 to 1. Accumulated probability distributions of n-pearl state $P(n)$ are shown as red histograms for each group. 44 sequences are classified into 4 different types depending on the pattern of visited points in 2-simplex. Each ensemble takes 43.2%, 27.3%, 25% and 4.5% The global ensemble of 44 sequences has the most visited state closer to the vertex of $V1$ where a single large pearl dominates but spreads more as compared to $Q = 20$. Due to the large net charge, C4 type where conformations consisting of pearls of alike-size develop. C1 type, C3 type and C4 type have the most visited state near $V1$,$V2$ and $V3$, respectively. Most of conformations of C2 type are along a line connecting vertex $V1$ and $V2$.

fluctuation has many paths inside the triangle with a continuous conversion. These paths go through very rare configurations and hence experience a substantial energy barrier in all cases investigated. The necklace rather changes number of pearls (say from 3 to 2 and 1) then undergoes part of the conversion in those states before coming back to 3 pearls in a new configuration. Such cycles are repeated and the representative point of the PA shifts progressively in the triangular diagram. The change in n_l mainly occurs during visits to the 2-pearl state. These dynamical aspects are detailed below.

Sequences of type C1 are dominant at small Q (The rare 3-pearl states for $Q = 16$ mostly belong to C1 type.) For large Q, other types C2, C3, C4 develop. For $Q = 20$, we find C1, C2, C3 (Fig. 4). The pearl number distributions are similar ($P_2 > P_3 > P_1 > P_4$) for all classes. Transition rates (stationary values) are $\omega_{2,3} \approx 0.035$ - 0.039 (global average 0.038) and $\omega_{4,3} \approx 0.021$ - 0.022 (global average 0.021). For all classes, the rate $\omega_{2,3}$ is larger than $\omega_{4,3}$.

For large Q, like $Q = 24$, all C1, C2, C3, C4 types are developed (Fig. 5). The probabilities

to find n-pearl states P_n are ($P_3 > P_4 > P_2 > P_5 > P_1$) for global ensemble of 45 sequences. We find for C1 type $P_3 > P_2 > P_4 > P_5 > P_1$, for C2 and C3 types, $P_3 \sim P_4 > P_2 > P_5 > P_1$, for C4 type $P_3 > P_4 > P_2 \sim P_5 > P_6 > P_1 \sim P_7$. Considering the stationary values of $\omega_{2,3}$, the global average is $\omega_{2,3} \approx 0.028$. C1 type has the largest switching rate $\omega_{2,3} \approx 0.032$ and other types have values between 0.024 – 0.026. Considering the stationary values of $\omega_{4,3}$, the global average is $\omega_{4,3} \approx 0.032$ and C3 has the largest switching rate $\omega_{4,3}^{C3} \approx 0.039$ but values for other types remain close 0.031 – 0.035. For $Q = 24$ the transition rates are at best indicative as is also reflected by the fact that the stationary value of the rates is smaller than the equilibrium rate shown in Table. 3. The two large pearls are asymmetric in type C2 and symmetric in C3 type. Conformations of C4 type do have less well-defined pearls (fat strings).

We trace the switchings and shape fluctuations by recording the shapes and by following the trajectories in simplex and between simplexes. Some of these trajectories are shown in Fig. 6 for $Q = 24$. When a small pearl separates from a large pearl or merges with a large pearl, switching n accompanies the size change of the large pearl. This type of splitting is encountered for C2, shown in Fig. 6 (b). Switching of n can occur by fluctuation of small pearl in the string without interaction of the large pearls. This type of switching is common for all C1, C2, C3 and dominant in C3 (See, Fig. 6). C1, C2 shapes have high probability to have a tail. Both termini of C3 types are occupied by a globular cluster/pearl. In all cases where the configurations are rather localized in the triangular representation, strong shape fluctuations affecting the number of large pearls n_l are not observed in the (relatively) short time trajectories shown in Fig. 6.

Most of the individual sequences preferentially populate a certain area in the triangle (and in the segmental 2-pearl representation). There are few sequences with very scrambled triangular diagrams where two or more distant regions of the diagram are well populated. Three examples with $Q = 20$ are presented in Fig. 7. These sequences can easily switch the number of large pearls, n_l. We analyze the dynamics of these sequences. Some are hybrids between two of the previous types. Examples are (a) within C2 (b) C2-hy-C1, (c) C3-hy-C1, which we labeled in Fig. 7. as C2, C2-I and C3-I, respectively. In these sequences fluctuations affecting the large pearls in the 3-pearl state are easily realized. They imply

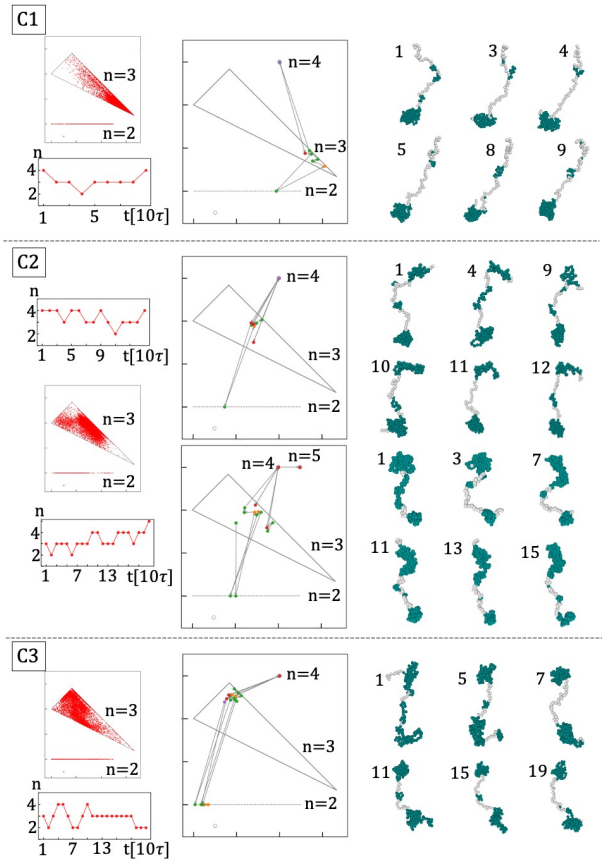


Figure 6: Typical shape fluctuations of sequences belong to C1, C2, C3 types ($Q = 24$). Accumulated visits (left panels) and short time trajectories (right panels) are shown on 1,2,3-simplex representation. 1-pearl state is a point, 2-pearl state as a 1-simplex $x_1 \in [0, 0.5)$, 3-pearl state is triangular 2-simplex $\{x_1, x_2\}$. The states with larger number of pearls ($n \geq 4$) are shown as a point above the 2-simplex for convenience. $n(t)$ shows the change in the number of pearls over a short period of time, and snapshots over that times are shown for each type of sequences. The unit time in graphs is 10τ. A different color for each point indicates different value of n from the previous time.

multiple switchings in the number of pearls, more so for longer times involving the uniglobule state. The 2-pearl simplex, which is also scrambled, plays an important role. Large pearls are deeply affected at times as short as $\sim 100\tau$ with an average relaxation time of $\lesssim 1000\tau$. The average switching time is apparently longer for transitions between more distant regions despite of large variances.

To extend the analogy between PA and IDP, we construct the simplex diagrams for two IDPs sequences that we have considered in the context of translocation previously: IN and ProTα (Fig. 8). We are interested in the sequences here, more/different efforts are needed

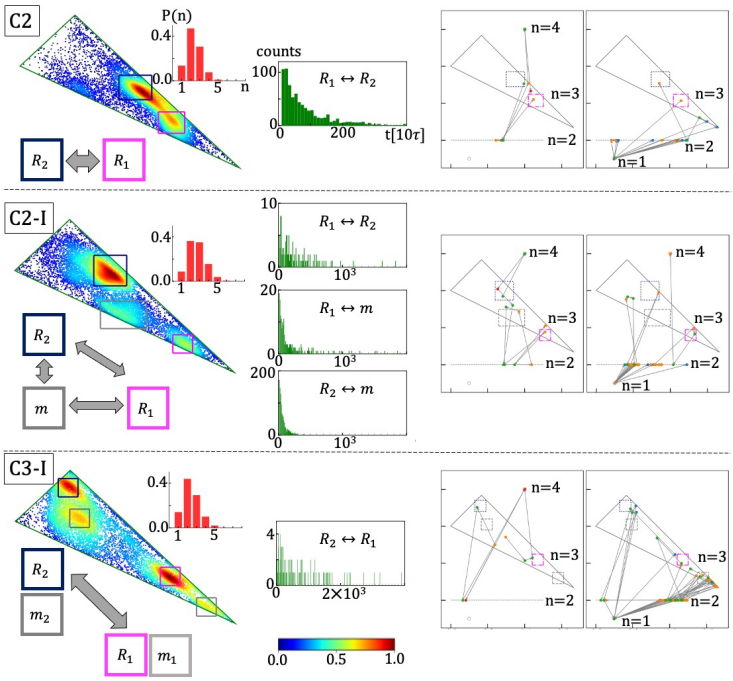


Figure 7: Trajectories of switching events between distant conformations in (a) C2 and in hybrids (b) C2-I and (c) C3-I. The left panels show the probability distribution $P(n)$ of n-pearl states for the given sequence, along with a triangular diagram representing the density of visits. The middle panels show the distributions of switching times between marked regions in 2-simplexes for C2, C2-I and C3-I. The average switching times are 126τ, 287τ and 738τ with rms deviations of 309τ, 360τ, 835τ for C2, C2-I and C3-I, respectively. In C2-I, switching between R_1 and m, R_2 and m have average times of 242τ and 97τ The right panels show typical short ($< 100\tau$) and long ($\sim 1000\tau$) trajectories on simplex representations. The trajectories start at the time leaving from one of the marked box (R2) and end at the time entering the marked magenta box (R1). 1-pearl state is a point, 2-pearl state as a 1-simplex $x_1 \in [0, 0.5)$, 3-pearl state is triangular 2-simplex $\{x_1, x_2\}$. The states with larger number of pearls ($n \geq 4$) are shown as a point above the 2-simplex for convenience.

to describe IDPs in vivo, including chemically realistic residues.

IN sequence is mostly found in uniglobular state ($\sim 72\%$) and about 25% of populations are found in 2-pearl state. The 2-pearl state segmental simplex representation contains two populated regions: a broad region (R2) closer to symmetric side, V2, and the other very narrow region (R1) with large asymmetry V1, which is reminiscent of uniglobular state. The less abundant 3-pearl structures ($\sim 3\%$) are scrambled between C2-C3 and C4 types. Similar to some PA sequences (C2-I or C3-I) analyzed in Fig. 7, in Fig. 8 (a), we show switching time distribution of IN between uniglobular states and 2-pearl states. The average switching time is $\sim 11 \pm 3\tau$. In order to switch the number of large pearls within 1-simplex, PEs often

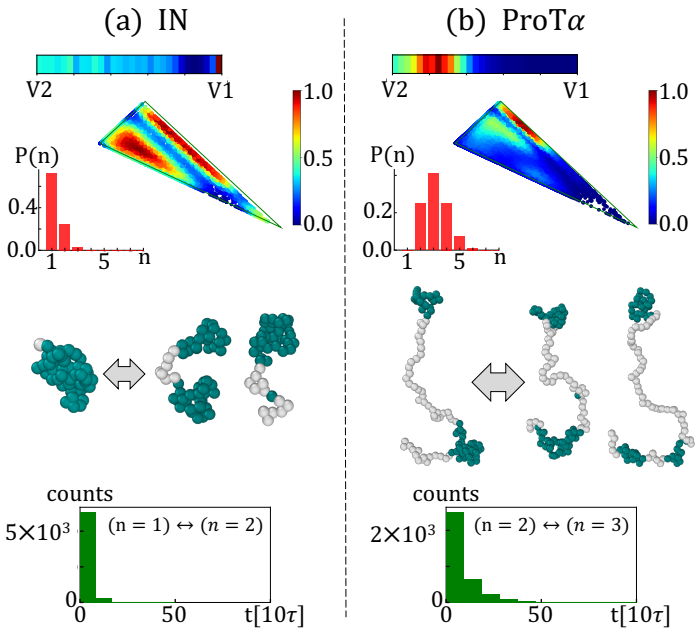


Figure 8: Structural heterogeneity analysis of IDP: (a) IN and (b) ProTα. The 2-pearl and 3-pearl simplex representation and the pearl number distributions $P(n)$. The bottom panels (green histogram) are switching time distributions between 1- and 2- pearl states of IN and between 2- and 3- pearl states of ProTα. Typical conformations are shown where pearls are identified as green colors.

go through uniglobular state or less often 3-pearl state.

For ProTα, the most abundant state is 3-pearl state $\sim 41\%$ (Fig. 8 (b)). 2- and 4-pearl states are also common ($\sim 25\%$ each). As expected from the large central negatively charged block (see, Table 1), 2-pearl states are predominantly in symmetric state, yet spread in the segmental 2-pearl representation. The 3-pearl diagram of ProTα resembles that of IN with the C3 type region (area closer to V2) being most populated (yet less extended towards C2). But, at variance with IN, there is no subdominant C4 type region (area closer to V3). The subdominant region remains almost of C3 type implying that at least one of the pearls is significantly smaller than others. Thus, the number of large pearls n_l remains at 2 for the most of times and 3- and 4-pearl states are short-lived. The short-time dynamics of switching between 2- and 3-pearl states with typical switching time $\sim 15 \pm 10\tau$ are governed by small pearl fluctuations.

Conclusions

The global statistics of the structure of random PAs shows a wide dispersion in the number of pearls even for a given net charge. It also shows dispersion in the pearl-size distribution for a fixed number of pearls as can be seen for three pearls from the triangular representations. In the same time most of the individual sequences have a qualitatively well defined pearl-size distribution for a given number of pearls and preferentially populate a certain area in the triangle, which does not necessarily reflect the global distribution over sequences. Few sequences explore wider areas of the triangle.

The pearl-necklaces are rather versatile with a fast fluctuating number of pearls. This fluctuations mainly affect small pearls forming and dissolving. The number of large pearls bearing more than their share $(1/n)$ of the monomers is much more stable. In the simulation the relaxation time associated with the number of small pearls is only of order a few 10 τ. The relaxation time of the number of large pearl is typically longer than the time reached by our simulations.

To further illustrate the relaxation of large pearls we consider in detail the case of medium charged sequences ($Q = 20$) in the predominant two and three pearl necklace states. The few sequences with the characteristics of dispersed triangular diagram (i.e. C2-I, C3-I) rather easily switch the number of large pearls n_l with mean relaxation times of several 100 τ. However the process usually includes (many) visits to necklace-states with different number of pearls n. Most of the changes in n_l occurs typically in $n = 2$ state and also during conversions in n. The switching of n_l in the two pearl state ($n = 2$) occasionally go through the uni-globule state ($n = 1$). Closer inspection of these sequences shows good dispersion of the charge along some long subsequence(s) not involving any large blocks, which is in some analogy with IDPs. Large blocks of the majority sign can exist away from these subsequences. To expand on this analogy, we construct simplex diagrams for two IDPs, IN and ProTα. In the spirit of our PA model we considered the two IDP sequences with the charges regularly spaced along the sequence and obtained dispersed 2-pearl and 3-pearl diagrams (results not shown). We also reached similar conclusions by considering sequences reflecting the actual charge spacing set in the same solvent quality for all residues.

What is striking for IN is the very scrambled 2-pearl diagram that contains a very localized densely populated asymmetric state and a wide spread more symmetric medium dense region. This offers the possibility of fast large scale fluctuations between symmetric and asymmetric necklaces. The 2- and 3-pearl diagrams of ProTα indicate that ProTα exists as rather symmetric 2-pearl structures, or as 3-pearl structures containing one significantly smaller pearl than the two other pearls. In accordance with these fluctuation statistics, both IDPs considered here quickly switch the number of large beads like the minority hybrid sequences discussed for PAs. While IDPs share generic properties of PAs such as the behavior in salty solution,[11] the statistics and dynamics of their fluctuations follows those of a tiny minority of hybrid PA sequences. IDPs are rather peculiar PAs. Our analysis suggests that synthetic polymers mimicking IDPs which can be synthesized[27] show conformational plasticity like IDPs and could be potentially used as multitaskers.

Ideally, several improvements could be made, but no qualitative change in our results is expected. Performing sufficient runs with explicit solvents to sample disordered polyampholytes is out of reach. Presence of explicit solvents would provide more realistic Zimm-type rather than Rouse-type dynamics. The fast and slow dynamic processes described above are entailed by the complex energy landscape and must remain true when explicit solvents are present. We did not vary the solvent quality in a systematic way. PA were simulated in weakly poor solvent condition which is the most common situation for synthetic PA. IDP were considered in marginal solvent (very close to the theta point). Data for IDP in weakly poor solvent (similar to that used for synthetic PA) can be found in the SI for comparison. In the case of poorer solvent, pearls are better defined with sharper interfaces. The simplex representation retains its overall structure but the highly populated islands are slightly more diffuse for the marginal solvent. The categories of slow and fast switching sequences are not affected.

Supporting Information: A. The equilibrium transition rate, B. Total number of switching events between different numbers of pearl states, C. Dynamics of ProTα structure in two different solvent conditions.

Acknowledgment Our work is supported by National Research Foundation grants provided by Korean government (2020R1A2B5B01002041 (NKL) and 2021R1F1A1064098 (YJ))

and the National Supercomputing Center (KSC-2022-CHA-0003).

References

(1) Corpart, J.; Candau, F. Formulation and polymerization of microemulsions containing a mixture of cationic and anionic monomers. *Colloid Polym. Sci.* **1993**, *271*, 1055–1067.

(2) Spalla, O. Nanoparticle interactions with polymers and polyelectrolytes. *Current Opinion in Colloid & Interface Science* **2002**, *7*, 179–185.

(3) Durmaz, E. N.; Sahin, S.; Virga, E.; de Beer, S.; de Smet, L. C. P. M.; de Vos, W. M. Polyelectrolytes as Building Blocks for Next-Generation Membranes with Advanced Functionalities. *ACS Appl. Polym. Mater.* **2021**, *3*, 4347–4374.

(4) Hariri, H. H.; Schlenoff, J. B. Saloplastic Macroporous Polyelectrolyte Complexes: Cartilage Mimics. *Macromolecules* **2010**, *43*, 8656–8663.

(5) Shamoun, R.; Hariri, H.; Ghostine, R.; Schlenoff, J. Thermal Transformations in Extruded Saloplastic Polyelectrolyte Complexes. *Macromolecules* **2012**, *45*, 9759–9767.

(6) Diddens, D.; Baschnagel, J.; Johner, A. Microscopic Structure of Compacted Polyelectrolyte Complexes : Insights from Molecular Dynamics Simulations. *ACS Macro Letters* **2019**, *8*, 123–127.

(7) Petrila, L.-M.; Bucatariu, F.; Mihai, M.; Teodosiu, C. Polyelectrolyte Multilayers: An Overview on Fabrication, Properties, and Biomedical and Environmental Applications. *Materials* **2021**, *14*, 4152.

(8) Lee, N. K.; Jung, Y.; Johner, A.; Joanny, J. F. Globular Polyampholytes: Structure and Translocation. *Macromolecules* **2021**, *54*, 2394–2411.

(9) Danielsen, S. P.; McCarty, J.; Shea, J.-E.; Delaney, K. T.; Fredrickson, G. H. Small ion effects on self-coacervation phenomena in block polyampholytes. *J. Chem. Phys.*, (2019) **2019**, *151*, 034904.

(10) Danielsen, S. P.; McCarty, J.; Shea, J.-E.; Delaney, K. T.; Fredrickson, G. H. Molecular design of self–coacervation phenomena in block polyampholytes. *Proc. Natl. Acad. Sci.* **2019**, *116*, 8224–8232.

(11) Müller-Späth, S.; Soranno, A.; Hirschfeld, V.; Hofmann, H.; Rüegger, S.; Reymond, L.; Nettels, D.; Schuler, B. Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 14609–14614.

(12) Samanta, H. S.; Chakraborty, D.; Thirumalai, D. Charge fluctuation effects on the shape of flexible polyampholytes with applications to intrinsically disordered proteins. *J. Chem. Phys.* **2018**, *149*, 163323.

(13) Bianchi, G.; Longhi, S.; Grandori, R.; Brocca, S. Relevance of Electrostatic Charges in Compactness, Aggregation, and Phase Separation of Intrinsically Disordered Proteins. *Int. J. Mol. Sci.* **2020**, *21*, 6208.

(14) Uversky, V. N. Intrinsically disordered proteins from A to Z. *Intl. J. Biochem. & Cell Biol.* **2011**, *43*, 1090–1103.

(15) McCarty, J.; Delaney, K. T.; Danielsen, S. P.; Fredrickson, G. H.; Shea, J.-E. Complete phase diagram for liquid–liquid phase separation of intrinsically disordered proteins. *J. Phys. Chem. Lett.* **2019**, *10*, 1644–1652.

(16) Dobrynin, A. V.; Rubinstein, M.; Obukhov, S. Cascade of transitions of polyelectrolytes in poor solvents. *Macromolecules* **1996**, *29*, 2974–2979.

(17) Kantor, Y.; Kardar, M. Excess charge in polyampholytes. *EPL (Europhys. Lett.)* **1994**, *27*, 643.

(18) Chae, M.-K.; Lee, N.-K.; Jung, Y.; Johner, A.; Joanny, J.-F. Structure of a hydrophobic polyelectrolyte chain with a random sequence. *Macromolecules* **2022**, *55*, 6275–6285

(19) Yamakov, V. and Milchev, A. and Limbach, H. J. and Dünweg, B. and Everaers, R. Conformations of random polyampholytes *Phys. Rev. Lett.* **2000**, *85*, 4305.

(20) Rumyantsev, A. M. and Johner, A. and de Pablo, J. J. Sequence Blockiness Controls the Structure of Polyampholyte Necklaces *ACS Macro Letters* **2021**, *1*, 1048–1054.

(21) Lee, N. and Obukhov, S. Multiple conformations in polyampholytes *Eur. Phys. J. B* **1998**, *1*, 371–376.

(22) Chae, M.-K.; Lee, N.-K.; Jung, Y.; Johner, A.; Joanny, J.-F. Partially Globular Conformations from Random Charge Sequences. *ACS Macro Letters* **2022**, *11*, 382–386.

(23) Lee, N.-K.; Thirumalai, D. Dynamics of collapse of flexible polyampholytes. *J. Chem. Phys.* **2000**, *113*, 5126–5129.

(24) Lee, N.-K.; Thirumalai, D. Dynamics of collapse of flexible polyelectrolytes in poor solvents. *Macromolecules* **2001**, *34*, 3446–3457.

(25) Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Comp. Phys.* **1995**, *117*, 1–19.

(26) Liao, Q.; Dobrynin, A. V.; Rubinstein, M. Counterion-correlation-induced attraction and necklace formation in polyelectrolyte solutions: Theory and simulations. *Macromolecules* **2006**, *39*, 1920–1938.

(27) Lutz, J.-F. Defining the Field of Sequence-Controlled Polymers. *Macromol. Rapid Com.* **2017**, *38*, 1700582.