Electrostatics Controls the Formation of Amyloid Superstructures in Protein Aggregation

Vito Foderà,[1,*] Alessio Zaccone,[1,2,*] Marco Lattuada,[3,†] and Athene M. Donald[1]

[1]*Sector of Biological and Soft Systems, Department of Physics, Cavendish Laboratory, University of Cambridge, JJ Thomson Avenue, Cambridge CB3 0HE, United Kingdom*

[2]*Theory of Condensed Matter, Department of Physics, Cavendish Laboratory, University of Cambridge, JJ Thomson Avenue, Cambridge CB3 0HE, United Kingdom*

[3]*ETH Institute for Chemical and Bioengineering, HCI F135, Wolfgang Pauli Strasse 10, 8093 Zurich, Switzerland*

(Received 5 May 2013; published 5 September 2013)

The possibility for proteins to aggregate in different superstructures, i.e. large-scale polymorphism, has been widely observed, but an understanding of the physicochemical mechanisms behind it is still out of reach. Here we present a theoretical model for the description of a generic aggregate formed from an ensemble of charged proteins. The model predicts the formation of multifractal structures with the geometry of the growth determined by the electrostatic interactions between single proteins. The model predictions are successfully verified in comparison with experimental curves for aggregate growth allowing us to reveal the mechanism of formation of such complex structures. The model is general and is able to predict aggregate morphologies occurring both *in vivo* and *in vitro*. Our findings provide a framework where the physical interactions between single proteins, the aggregate morphology, and the growth kinetics are connected into a single model in agreement with the experimental data.

DOI: 10.1103/PhysRevLett.111.108105 PACS numbers: 87.14.em, 87.15.bk, 87.15.km, 87.15.nr

Understanding the connection between growth mechanisms and morphology is a central problem for modelling self-assembling biological systems [1]. This basic topic in condensed matter and biophysics was already emphasized by the far-seeing work of D'Arcy Thompson at the beginning of the last century, focusing on the need for quantitatively describing the specific physical interactions leading to different structural arrangements [2]. Protein aggregation is a central area in current biophysics research mainly because of its connection to neurodegenerative diseases [3]. An increasing interest has recently been addressed towards understanding the occurrence of pronounced microscopic polymorphism in the formation of aggregates of amyloid origin, i.e. fibrils [4–6]. Moreover, both *in vivo* and *in vitro*, amyloid aggregates may generally conserve their basic structural arrangement of cross β sheet, yet exhibit significantly different packing into three-dimensional superstructures.

Under destabilizing conditions and sufficiently high protein concentrations, a number of model proteins have been shown to aggregate into different forms [7–9], mainly depending on the pH of the solution [7]. Close to the isoelectric point (pI) of the protein (i.e. no net charge on the protein), compact spherical aggregates with radius up to 1 μm (particulates) are detected [10]. On the other hand at low pH (i.e. high charge on the protein), elongated amyloid fibrils [11] occur together with amyloid spherulites [12]. Spherulites [with radii up to hundreds of μm, Fig. S1 in the Supplemental Material [13]] are thought to be composed of a central and compact part (precursor) surrounded by a low-density outer part (shell) [14]. They are rich in β structures [12], show a positive labelling when bound to amyloid sensitive dyes [12], and they may have a role in human amyloid pathologies [15]. However, even though the occurrence of such a variety of morphologies is widely observed, the connection between the macroscopic final morphology and details of the growth kinetics is still out of reach. This leads also to the central and still unexplored problem of linking the aggregation kinetics curves with the microstructural details of the growing aggregate. Several models based on nucleation assume generic mechanisms for protein assembly and disassembly but do not allow for a proper connection between the large-scale morphology of the aggregates and the interactions and phenomena happening at shorter time and length scales.

Here we bridge the gap between the kinetic description of the overall process, the predictions of superstructures, and the physics of interprotein interactions. By means of a statistical mechanical model, we provide direct evidences that electrostatic interactions between single protein molecules determine the final amyloid superstructure through a multifractal growth process. By comparison with experimental data, we prove that our coarse-grained model quantitatively predicts both the overall kinetics and the large-scale morphology for spherulite-forming systems. Importantly, the proposed framework is general and can be used to recover the amyloid fibril morphology and the occurrence of particulates in the limit of high protein charge and uncharged proteins, respectively.

We base our model for aggregate formation on the calculation of the free energy of a spherical cluster with radius R,

$$\Delta F = -\frac{qU}{2} + \frac{3\gamma}{R}\frac{4\pi R^3}{3} + \frac{(Nn_c e)^2}{8\pi\varepsilon\varepsilon_0 R} - kT\log(X_N/N), \quad (1)$$

where q is the total number of contacts between pairs of molecules in the aggregate, U is the energy per contact, γ is the surface tension [16], R is the radius of the cluster, N is the number of the molecules in the cluster, n_c is the effective number of charges on a single particle, e is the elementary charge, ε and ε_0 are the relative dielectric constant and the permittivity in vacuum, and X_N is the volume fraction of molecules in the cluster. The first term is related to the binding between single molecules in a cluster and the second one is the correction factor due to missing contacts at the surface [16]. The first two terms in Eq. (1) correspond to the well-known terms in the Weizsacker formula [17]. The third term represents the electrostatic (Born) energy required to move a charged protein from infinity to the charged aggregate and is consistent with the ion charging formula previously reported [16,18]. This term is related to the overall electrostatic repulsion between the molecules [16,18] and, due to the complexity of the electrostatic interaction between two proteins, n_c cannot be referred only to the absolute number of charges on the protein but depends on several other factors [19–23]. As a consequence, the electrostatic term in Eq. (1) represents an effective term including all the possible contributions. Finally, the last term is the entropic contribution arising from the loss of translational degrees of freedom when particles are bound to the cluster [24].

Equation (1) can be rewritten (see the Supplemental Material [13]) for the case of fractal growth to obtain the free energy of the cluster as a function of R, the fractal dimension d_f, the radius of the single protein a, the number of particle nearest-neighbors Z, the effective number of interactions f, and a binding energy $n_E\ kT$,

$$\Delta F = -\frac{Z}{2}\left(\frac{R}{a}\right)^{d_f} n_E kT + \frac{f n_E kT d_f}{a^{(d_f-1)}} R^{(d_f-1)} + \frac{n_c^2 e^2}{8\pi\varepsilon\varepsilon_0 a^{2d_f}} R^{(2d_f-1)} - kT \log\frac{a^3}{R^3} \qquad (2)$$

Equation (2) represents the free energy for an aggregate growing with a generic d_f. We evaluate Eq. (2) for a spherical growing aggregate with $d_f = 3$ and for single (globular) proteins of radius $a = 2$ nm at different effective number of charges n_c (see the Supplemental Material [13]). We assume a binding energy of $10kT$ which is compatible with a $d_f \leq 3$ growth [16]. Importantly, in the specific case of aggregating proteins, the first term of the right-hand side of Eq. (2) refers to the binding energy between already destabilized and aggregation-prone molecules. Figure 1 shows the free energy profile as a function of the cluster radius. For each n_c value, the free energy shows an initial constant value followed by a minimum and a steep and indefinite increase towards positive values. Varying n_c in Eq. (2) changes the value of the minimum and its position shows a well-defined exponential decay as n_c increases (inset).

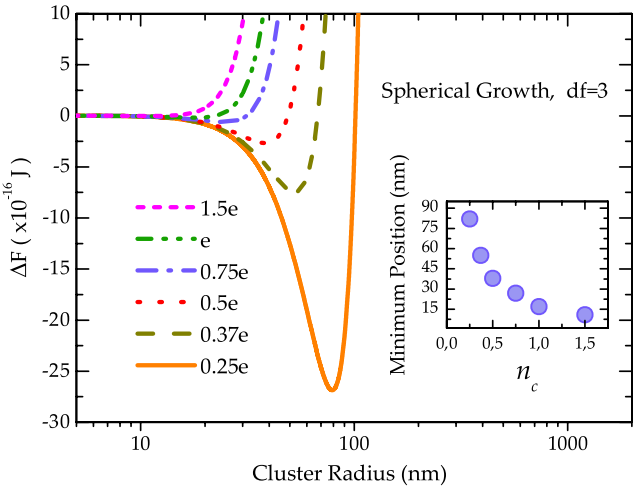

FIG. 1 (color online). Free energy calculated by means of Eq. (2) for $d_f = 3$ considering different values of effective charge on a single particle $n_c e$. Inset: energy minimum position as a function of n_c.

The data in Fig. 1 predict a growth of a spherical aggregate with the size controlled by n_c. After reaching the energy minimum, the aggregate can no longer evolve with the same geometric features. In Eq. (2) information about the structure is encompassed by the fractal dimension d_f. Calculations of the energy profile have been performed at 5 different d_f values. When the d_f is decreased, the free energy minimum turns out to be shifted towards higher values of radius. This means that, after reaching the first minimum ($d_f = 3$), the aggregating system can explore new minima of its free energy only if the morphology of the growth changes, i.e. if d_f decreases. Since this change happens continuously, this leads to a multifractal profile for the free energy [Fig. 2(a)], with more compact objects energetically favorable in the early stages of the growth (see Fig. S2 in the Supplemental Material [13]). The system will follow a pathway of energy minimization [solid line in Fig. 2(a)] leading to an aggregate with a compact central structure with $d_f = 3$ (hereafter called precursor) and an outer part with a decreasing fractal dimension as a function of the radius (hereafter called shell). This has some similarities to the well-known Rayleigh electrospray effect, where a spherical drop overcomes the overcharging by spraying the liquid outwards [25,26]. The model proposed here allows one to calculate how the multifractal profile evolves during the aggregate growth at different values of n_c. After the precursor formation, d_f shows a decrease for $n_c = 0.5$ and $n_c = 1$ [Fig. 2(b)]. Interestingly, decreasing the n_c value down to 0.001 leads to an aggregate growing with $d_f = 3$ for tens of μm [Fig. 2(b)] before a significant decrease in d_f can take place.

Extrapolating the d_f vs R relationship from Fig. 2(b) also allows us to estimate the change in density during the

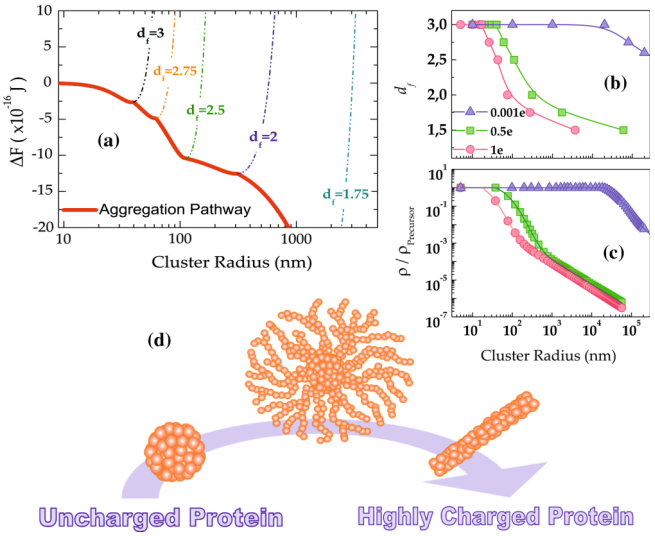

FIG. 2 (color online). (a) Free energy profile for a growing cluster as a function of the radius calculated at 5 different fractal dimensions by means of Eq. (2) ($n_c e = 0.5e$). The solid line indicates the most energetic favorable pathway for the aggregate growth. (b) Fractal dimension and (c) density of the aggregate normalized by the precursor density during the cluster growth: $n_c e = 0.001e$ (triangles), $n_c e = 0.5e$ (squares), and $n_c e = 1e$ (circles). (d) Illustrative sketch of the superstructures: from particulates (in the limit of $n_c = 0$), to amyloid fibrils ($n_c > 1$) (Fig. S3 in the Supplemental Material [13]). For $0 < n_c < 1$ precursor-shell growth is predicted (i.e. spherulites).

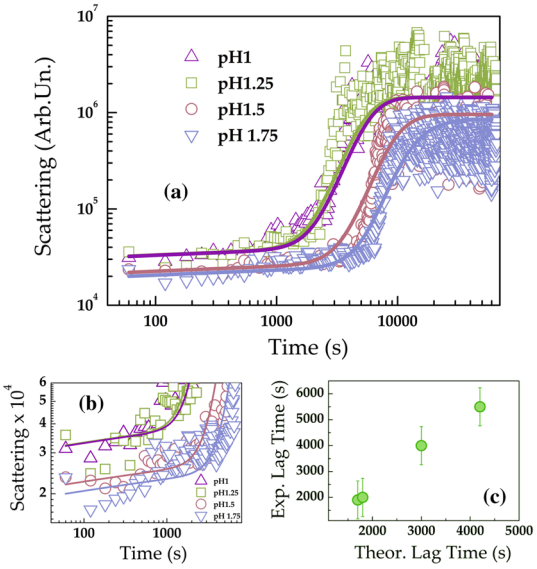

FIG. 3 (color online). (a) Static light scattering intensity as a function of time during insulin spherulite formation at different pH values for a bovine insulin sample 4 mg/ml, 25 mM NaCl during incubation at 60 °C for 24 h. Solid lines represent simulated curves according to the theoretical model. (b) Zoom on the early stages of the process. (c) Comparison between experimental and simulated lag times of the process. Error bars represent absolute deviations observed on three replicates.

aggregate growth compared to the precursor [Fig. 2(c); see the Supplemental Material [13]]. After the formation of the precursor, a decrease of the density is predicted for the data at $n_c = 0.5$ and $n_c = 1$, i.e. spherulite formation. For the data at $n_c = 0.001$, a significant decrease of the density is expected only when the aggregate reaches a radius >20 μm [triangles in Fig. 2(c)]. In the limit of $n_c = 0$, d_f is constant and equal to 3 for the entire growth, i.e. particulates. Conversely, for $n_c > 2$, the growth basically proceeds with $1 < d_f < 2$ from the early stages, i.e. elongated fibrillar structures (Fig. S3 in the Supplemental Material [13]). These predictions are sketched in Fig. 2(d) and are in agreement with what is experimentally observed [10–12,27].

Now the question is if we can quantitatively describe the temporal course of experimentally observed aggregate growth by our microscopic model. We consider static light scattering data for spherulite growth in samples of bovine insulin during incubation at 60 °C and at different pH values in the range 1–1.75 [28]. Decreasing the pH in this range would mainly increase the positive charge on the protein [29].

Aggregation kinetics (symbols) in Fig. 3(a) show the well-known sigmoidal profile, with an initial lag time that decreases as the pH is lowered. A closer view to the lag time shows an increase in the signal already in the very early stages of the process [Fig. 3(b)]. After that, an abrupt increase in the growth rate of aggregates characterizes the temporal profile before reaching a plateau. To date, only generic and qualitative explanations are suggested for the early increase of the signal [30,31]. In order to compare the predictions of our theoretical framework with experimental data we consider the master kinetic (population balance) equations for the aggregation process

$$\frac{dC_k}{dt} = \frac{1}{2}\sum_{i+j=k} K_{ij} C_i C_j - C_k \sum_{i=1}^{\infty} K_{ik} C_i - K_k^B C_k + \sum_{i=k+1}^{\infty} K_{ik}^B C_i, \quad (3)$$

where C_i is the concentration of aggregates with mass i (i.e., made of i protein molecules), and K_{ij} is the kernel determining the rate of aggregation between two aggregates, one with mass i and the other with mass j, explicitly including both the van der Waals contribution and a repulsive electrostatic interaction (see the Supplemental Material [13]). The last two terms in Eq. (3) account for thermal breakup of a cluster of size k and generation of a k cluster by breakup of a cluster of size $k + i$. For systems in which the thermal breakup is not relevant the last two terms are negligible. This is actually the case of our system ($\sim 10kT$; see Sec. 3 in the Supplemental Material [13]).

The microscopic rates can be calculated based on a conventional diffusion-limited aggregation scheme (see the Supplemental Material [13]) and they fully account for four basic interactions, which have all been computed

using the Derjaguin-Landau-Verwey-Overbeek theory: (1) monomer-monomer, (2) oligomer-oligomer, (3) shell-monomer, and (4) shell-oligomer interaction, where the cluster in the precursor regime ($d_f = 3$) is treated as a dielectric sphere, while, for clusters in the multifractal regime, the reactivity of a monomer is considered (see the Supplemental Material [13]). The geometry of the growth is also taken into account by implementing the d_f evolution predicted by the theory into the master equation (see the Supplemental Material [13]). Together with the above hypotheses, pairs of values of n_c and precursor radius, as obtained from the model (Table S1 [13]) have been used to simulate curves with different electrostatic properties. The scattering profiles are then obtained from structure factors calculated by the Fisher-Burford equation (see Fig. S5 and the Supplemental Material [13]). Simulations [solid lines in Figs. 3(a) and 3(b)] are able to predict both the initial slow increase in the light scattering curves and the rapid growth of signal before reaching the plateau. The agreement can also be seen by considering the experimental lag times versus the theoretical prediction [Fig. 3(c)].

We can now go back to our original question: can we relate the temporal curve with details of the large-scale morphology of the growing aggregate? Potential curves for the interactions between all the species can be estimated [Fig. 4(a), Eq. 2.10 in the Supplemental Material [13]] together with changes in oligomer and precursor populations as a function of time [Fig. 4(b)]. Moreover, in Fig. 4(b) the scattering curve (dashed line) is also shown

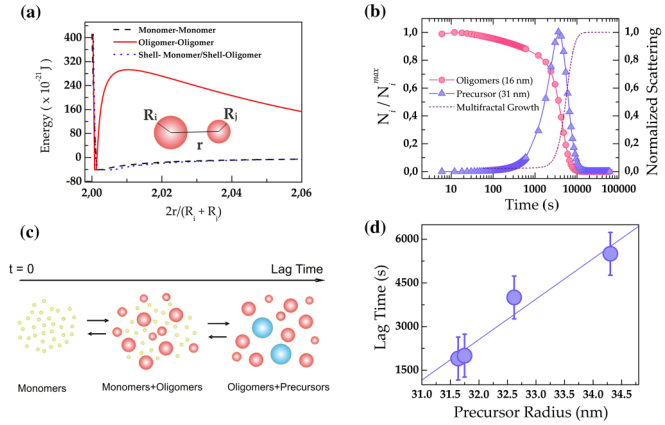

FIG. 4 (color online). (a) Potential energy for monomer-monomer (dashed line), oligomer-oligomer (close to the critical size of the precursor, solid line), shell-monomer, and shell-oligomer interactions (dotted line). With the word "oligomer" we refer to an aggregate with a number of units higher than 1 (monomer) and lower than the number of molecules in the precursor. (b) Oligomer and precursor population during the aggregation process as predicted by the model for a sample at pH 1. The scattering curve is shown to visualize the activation of the multifractal growth. (c) Sketch of the mechanism during the lag time: from monomers to oligomers and precursors. (d) Experimental lag time as a function of the simulated precursor size.

to visualize the profile of the multifractal growth. All these data are for the kinetics at $p\mathrm{H} = 1$; analogous trends were obtained at other pH values. The association of individual proteins proceeds without any significant energy barrier [dashed line in Fig. 4(a)], so that a rapid formation of oligomers with an average radius of ~16 nm [circles in Fig. 4(b)] takes place in the early stages of the process. This also explains the initial increase observed in the scattering curves [Fig. 3(b)] and it is in agreement with previously reported experimental data [32,33]. This is further verified by the size distributions obtained by dynamic light scattering experiments (see Fig. S6 in the Supplemental Material [13]). Afterwards, oligomers associate [decrease of the oligomer fraction, circles in Fig. 4(b)] until they reach a critical radius leading to a specific potential barrier for oligomer-oligomer interaction up to 288×10^{-21} J ($\sim 70kT$). This barrier makes further association between oligomers with critical radius extremely unlikely [solid line in Fig. 4(a)]. This critical size defines the radius of the precursor, the number of which increases until the end of the lag phase [~ 4000 s, triangles in Fig. 4(b)]; after that, shell growth is dominant and takes place through association between precursors with smaller oligomers and/or residual monomers. This shell growth can proceed without any significant barrier [dotted line in Fig. 4(a)], leading to the consumption of the precursor population [4000–10 000 s; triangles in Fig. 4(b)] and the formation of the multifractal structure [increase in the scattering curve, dashed line in Fig. 4(b)]. These quantitative results are sketched in Fig. 4(c).

Our model suggests that the difference in the lag time of the kinetics in Fig. 3(a) is basically related to the radius of the precursor. When the precursor is smaller, the time necessary to reach the critical radius for the shell growth is reduced [Fig. 4(d)]. It is worth noting that for a number of amyloidogenic systems, the abrupt growth in the aggregate sizes might find an elegant explanation in terms of secondary nucleation processes [33–36]. However, for systems mainly forming spherulites [27], the speeding up of the process is basically dictated by the change in growth geometry from a compact sphere (precursor) to an increasingly less compact geometry (shell) [12]. The latter is regulated by the minimization of the free energy of clustering (which is dominated by the electrostatic contribution). Importantly, our mathematical framework can recover the classical nucleation theory in the limit of weak attraction (see Sec. 3 in the Supplemental Material [13]).

In summary, using a combination of theoretical arguments, quantitative experiments, and simulations, we show that multifractal patterns arise in protein aggregation reactions due to the interplay of a random multiplicative process (growth) that evolves under the constraint of following a path of minimal free energy, the latter being dominated by electrostatics. Our approach naturally explains the occurrence of a range of protein aggregate structures

observed *in vivo* and *in vitro* controlled by electrostatic interactions. An adequate knowledge on how interprotein interactions are related to both the overall aggregation kinetics and the aggregate morphology is nowadays crucial [37]. Our framework provides the possibility to connect these three aspects, offering a new tool to single out, rationalize, and control the mechanisms behind protein aggregation phenomena. Furthermore, in view of the absence of restrictive assumptions in the proposed model, Eq. (2) could be, in principle, used to describe generic systems of charged particles undergoing random multiplicative and branching processes, e.g. dielectric breakdown of insulators [38–40].

We thank Mike Smith (University of Nottingham) for help in designing the experiments. We thank James Sharp and Clive Roberts (University of Nottingham) for useful discussions. V. F. thanks Lorenzo Di Michele (University of Cambridge) for useful discussions. V. F. thanks Bente Vestergaard (University of Copenhagen) for the inspiring discussions on polymorphism and superstructures in protein aggregation. Funding from the EPSRC (EP/H004939/1), the Swiss National Science Foundation (Grants No. PBEZP2-131153 and No. 200020-126487/1) is gratefully acknowledged. A. Z. acknowledges support from the Ernest Oppenheimer Fellowship at Cambridge.

*To whom correspondence should be addressed.
vf234@cam.ac.uk
az302@cam.ac.uk
†Present address: Adolphe Merkle Institute, University of Fribourg, Route de l'ancienne Papeterie, CP 209, CH-1723 Marly 1, Switzerland.

[1] T. Savin, N. A. Kurpios, A. E. Shyer, P. Florescu, H. Liang, L. Mahadevan, and C. J. Tabin, Nature (London) **476**, 57 (2011).
[2] D. W. Thompson, *On Growth and Form* (Cambridge University Press, Cambridge, England, 1917).
[3] F. Chiti, P. Webster, N. Taddei, A. Clark, M. Stefani, G. Ramponi, and C. M. Dobson, Proc. Natl. Acad. Sci. U.S.A. **96**, 3590 (1999).
[4] A. Loksztejn and W. Dzwolak, J. Mol. Biol. **395**, 643 (2010).
[5] D. Thirumalai, G. Reddy, and J. E. Straub, Acc. Chem. Res. **45**, 83 (2012).
[6] M. D. Griffin, M. L. Mok, L. M Wilson, C. L. Pham, L. J. Waddington, M. A. Perugini, and G. J. Howlett, J. Mol. Biol. **375**, 240 (2008).
[7] V. Vetri, M. D'Amico, V. Foderà, M. Leone, A. Ponzoni, G. Sberveglieri, and V. Militello, Archive Biochem. Biophys. **508**, 13 (2011).
[8] F. Massi, D. Klimov, D. Thirumalai, and J. E. Straub, Protein Sci. **11**, 1639 (2002).
[9] V. Foderà, S. Pagliara, O. Otto, U. F. Keyser, and A. M. Donald, J. Phys. Chem. Lett. **3**, 2803 (2012).
[10] M. R. H. Krebs, G. L. Devlin, and A. M. Donald, Biophys. J. **92**, 1336 (2007).
[11] L. Nielsen, R. Khurana, A. Coats, S. Frokjaer, J. Brange, S. Vyas, V. N. Uversky, and A. L. Fink, Biochemistry **40**, 6036 (2001).
[12] M. R. Krebs, C. E. Macphee, A. F. Miller, I. E. Dunlop, C. M. Dobson, and A. M. Donald, Proc. Natl. Acad. Sci. U.S.A. **101**, 14420 (2004).
[13] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.111.108105 for supplemental theoretical details and experimental data.
[14] M. R. Krebs, E. H. Bromley, S. S. Rogers, and A. M. Donald, Biophys. J. **88**, 2013 (2005).
[15] E. House, K. Jones, and C. Exley, J. Alzheimer's Disease **25**, 43 (2011).
[16] J. Groenewold and W. K. Kegel, J. Phys. Chem. B **105**, 11702 (2001).
[17] N. I. Lebvoka, *Advances in Polymer Science* (Springer, New York, 2012).
[18] J. Israelachvilli, *Intermolecular and Surface Forces* (Academic Press, New York, 1985).
[19] R. Piazza, Curr. Opin. Colloid Interface Sci. **8**, 515 (2004).
[20] D. N. Petsev and P. G. Vekilov, Phys. Rev. Lett. **84**, 1339 (2000).
[21] D. Chandler, Nature (London) **437**, 640 (2005).
[22] R. Piazza, Curr. Opin. Colloid Interface Sci. **5**, 38 (2000).
[23] G. N Patargias, S. A. Harris, and J. H. Harding, J. Chem. Phys. **132**, 235103 (2010).
[24] J. A. L. Jones, *Soft Condensed Matter* (Oxford University Press, Oxford, 2002); A. Zaccone and E. M. Terentjev, Phys. Rev. Lett. **108**, 038302 (2012).
[25] L. Rayleigh, Philos. Mag. **14**, 184 (1882).
[26] A Gomez and K. Tang, Phys. Fluids **6**, 404 (1994).
[27] V. Foderà and A. M. Donald, Eur. Phys. J. E **33**, 273 (2010).
[28] M. I. Smith, V. Foderà, J. S. Sharp, C. J. Roberts, and A. M. Donald, Colloids Surf. B **89**, 216 (2012).
[29] J. Haas, E. Vöhringer-Martinez, A. Bögeholdm, D. Matthesm, U. Hensenm, A. Pelahm, B. Abel, and H. Grubmüller, Chembiochem **10**, 1816 (2009).
[30] M. Manno, E. F. Craparo, A. Podestà, D. Bulone, R. Carrotta, V. Martorana, G. Tiana, and P. L. San Biagio, J. Mol. Biol. **366**, 258 (2007).
[31] F. Librizzi, V. Foderà, V. Vetri, C. Lo Presti, and M. Leone, Eur. Biophys. J. **36**, 711 (2007).
[32] S. Grudzielanek, V. Smirnovas, and R. Winter, J. Mol. Biol. **356**, 497 (2006).
[33] V. Foderà, S. Cataldo, F. Librizzi, B. Pignataro, P. Spiccia, and M. Leone, J. Phys. Chem. B **113**, 10830 (2009).
[34] T. P. Knowles, C. A. Waudby, G. L. Devlin, S. I. Cohen, A. Aguzzi, M. Vendruscolo, E. M. Terentjev, M. E. Welland, and C. M. Dobson, Science **326**, 1533 (2009).
[35] R. Jansen, W. Dzwolak, and R. Winter, Biophys. J. **88**, 1344 (2005).
[36] V. Foderà, M. van de Weert, and B. Vestergaard, Soft Matter **6**, 4413 (2010).
[37] L. L. Blazer and R. R. Neubig, Neuropsychopharmacology **34**, 126 (2009).
[38] H. E. Stanley and P. Meakin, Nature (London) **335**, 405 (1988).
[39] L. Niemeyer, L. Pietronero, and H. J. Wiesmann, Phys. Rev. Lett. **52**, 1033 (1984).
[40] L. Pietronero and H. J. Wiesmann, J. Stat. Phys. **36**, 909 (1984).