Manuscript (with track changes)

Symmetry-Breaking Transitions in the Early Steps of Proteins self-assembly.

Carmelo La Rosa[a], Marcello Condorelli[a], Giuseppe Compagnini[a], Fabio Lolicato[a,b,c], Danilo Milardi[d], Trang Do[e], Mikko Karttunen[f], Martina Pannuzzo[g], Ayyalusamy Ramamoorthy[h], Franca Fraternali[i], Francesca Collu[i], Human Rezaei[l], Birgit Strodel[k]and Antonio Raudino[a,]*.

[a]Department of Chemical Sciences, University of Catania, Viale A. Doria 6, I-95125, Catania, Italy.

[b]Department of Physics, University of Helsinki, P.O. Box 64, FI-00014, Helsinki, Finland.

[c]Heidelberg University Biochemistry Center, Heidelberg, Germany

[d]CNR, Istituto di Biostrutture e Bioimmagini, Unità Organizzativa e di Supporto di Catania, Via P. Gaifami 18, I-95126, Catania, Italy.

[e]Department of Chemistry & Waterloo Institute for Nanotechnology, University of Waterloo, 200 University Avenue West, Waterloo, Ontario, Canada N2L 3G1.

[f]Department of Chemistry and Department of Applied Mathematics, Western University, 1151 Richmond Street, London, Ontario N6A 5B7, Canada

[g]Istituto Italiano di Tecnologia (IIT), Via Morego 30, 16163 Genoa, Italy

[h]Department of Chemistry and Department of Biophysics, University of Michigan, Ann Arbor, Michigan 48109-1055, USA.

[i]Randall Division of Cell and Molecular Biophysics, New Hunt's House King's College London, Guy's Campus SE1 1UL London UK.

[l]Virologie et Immunologie Moléculaires, Institut National de la Recherche Agronomique, F-78352 Jouy-en-Josas, France.

[k]Institute of Complex Systems, Structural Biochemistry (ICS-6), ForschungszentrumJülich, 52425 Jülich, Germany.

Abstract

Protein misfolding and assembly are complex, intertwined processes resulting in the development of a heterogeneous population of aggregates closely related to many chronic pathological conditions including Type 2 Diabetes Mellitus and Alzheimer's Disease. To address this issue, here we develop a theoretical model in the general framework of linear stability analysis. According to this model, self-assemblies of peptides with pronounced conformational flexibility may become, under particular conditions, unstable and spontaneously evolve toward an alternating array of partially ordered and disordered monomers. The predictions of the theory were verified by atomistic Molecular Dynamics (MD) simulations of Islet Amyloid Polypeptide (IAPP) used as a paradigm of aggregation-prone polypeptides (proteins). Simulations of dimeric, tetrameric and hexameric human-IAPP self-assemblies at physiological electrolyte concentration reveal an alternating distribution of the smallest domains (of the order of the peptide mean length) formed by partially ordered (mainly β-strands) and disordered (turns and coil) arrays. Periodicity disappears upon weakening of the inter-peptide binding, a result in line with the predictions of the theory. To further probe the general validity of our hypothesis, we extended the simulations to other peptides, the Aβ(1-40) amyloid peptide, and the ovine prion peptide as well as to other proteins (SOD1 dimer) that do not belong to the broad class of intrinsically disordered proteins. In all cases, the oligomeric aggregates show an alternate distribution of partially ordered and disordered monomers. We also carried out Surface Enhanced Raman Scattering (SERS) measurements of hIAPP as an experimental validation of both the theory and in silico simulations.

Keywords: Intrinsically Disordered Proteins, Symmetry-Breaking, Molecular Dynamics, Analytical Model, Oligomers.

Introduction

In spontaneous symmetry-breaking, a system possessing a symmetry at its high temperature phase loses this symmetry upon cooling in the absence of external fields below the phase transition temperature. Ferromagnetism is probably the most well-known example: net magnetization arises as the local spins become ordered upon cooling below the Curie temperature (Chaikin and Lubensky 1995). Spontaneous symmetry breaking is important in a plethora of phenomena such as single-double bond alternation in Chemistry, molecular chirality, the formation of liquid-crystalline phases, charge and spin-density waves, superconductivity, the onset of polarity in cells and so on (Grason and Santangelo 2006; Yannoni and Clarke 1983; Chaikin and Lubensky 1995; Fuß *et al.* 2000; Gorman and Chakrabartty 2001; Peierls 2001; Nguyen *et al.* 2005, 2016; Tamashiro and Schiessel 2006; Sitkiewicz *et al.* 2014; Klinger *et al.* 2014; vandenAkker *et al*. 2015).

In this work, we try to extend the above concepts of spontaneous symmetry breaking to solid-state peptide (protein) crystals and to oligomeric self-aggregate protein in solution. It is well-known that in the crystal lattice the unit cell of most proteins contains just a single protein, while the unit cell of some protein systems may contain two or more *unequal* proteins (Weber and Steitz 1987; Deng *et al.* 1993) (see, as an example, the SOD1 structure reported in fig S1 of Supplementary Information (SI)).

This behavior is often observed also in the liquid phase, where proteins can stick together forming oligomers by spontaneous self-assembly. In this case, aggregates may arrange themselves as an array of identical monomeric units or, alternatively, they may exhibit periodic patters. For instance, dimeric proteins such as wt (Deng *et al.* 1993)and apo (Banci et al. 2012) superoxide dismutase (SOD1) and CRP-CAMP (Weber and Steitz 1987) exert their biological activity as *asymmetric* dimers.

We would like to mention that the possibility of heterogeneous conformational distribution along a protein aggregate has been previously suggested in literature (Gorman and Chakrabartty 2001; Sitkiewicz *et al*. 2014; Klinger *et al.* 2014; vandenAkker *et al.* 2015) and scattered observations have been reported in MD studies of Intrinsically Disordered Proteins (IDPs) (Reddy *et al.* 2010; Dupuis *et al*. 2011; Barz *et al.* 2014; Huy *et al.* 2016; Zhang *et al.* 2016; Nguyen *et al.* 2016; Das *et al.* 2017).

In this paper we address this issue by building up a simplified theoretical model (both in 1-D and 3-D) where protein aggregates are represented as an array of two different conformers: a spheroidal and quite disordered conformer alternating with a more ordered and elongated structure. Notably, the elongated shape enables the unit to be packed more tightly than the spheroidal one. The theoretical model developed here predicts that, under particular conditions, the system becomes unstable and

produce an alternating (patterned) array of ordered and disordered conformers when: a) their internal energy difference is small; b) the monomer-monomer interaction is large; c) there is a change in the geometry on going from ordered to disordered conformers.

The best candidates for observing the foreseen effects is the broad family IDPs because they fulfill all the above requirements. Furthermore, IDPs misfolding and aggregation into toxic oligomers or amyloid fibrils has attracted increasing interest in the last decades since it is associated with a large number of diseases (such as Alzheimer's disease (AD) and type 2 diabetes (T2D)). Particular attention was paid to oligomers to define the molecular determinants that may steer amyloid vs amorphous aggregation(La Rosa *et al*. 2016; Scollo *et al.* 2018): in fact, poorly structured, small sized oligomers may be more toxic than large, mature fibrils (Chiti and Dobson 2006; Riek and Eisenberg 2016; Sciacca *et al*. 2018). This is being proven by the development of new toxicity inhibitors that interact with oligomers instead of the already studied beta-breakers(Sciacca et al. 2017; Savelieff *et al*. 2019). A hallmark of AD is the aggregation of amyloid-β (Lorenzo *et al.* 1994) (Aβ); and human amylin (Höppener et al. 2000) (IAPP), a 37-residue polypeptide, is the major constituent of the pancreatic amyloid deposits found in patients with T2D. Both Aβ and IAPP belong to the IDPs family.

To verify the general validity of the model, we performed Molecular Dynamics (MD) simulations of di-, tetra- and hexameric assemblies of hIAPP in pure water and in a 0.1 M electrolyte solution, a condition in which the intermolecular repulsion is weaker. Accurate experimental tests of our model are not currently available since present instrumentation do not support nanometer and millisecond resolution. However, Surface Enhanced Raman Scattering (SERS) measurements are able to detect the average evolution of helix, sheet and random coil secondary structures as a function of time. These measurements confirm the heterogeneous nature of hIAPP aggregates. We extended MD simulations to other IDPs (e.g. Aβ peptide) and amyloidogenic globular proteins (e.g., ovine prion peptide) to probe the general validity of the proposed model. Finally, we performed accurate (in the microsecond scale) MD simulations of Super Oxide Dismutase dimer, a well-known protein that does not belong to the IDPs class. Similarly, to IDPs, SOD self-aggregates and exerts its enzymatic activity as an asymmetric tight-bound dimer.

Experimental Methods

hIAPP monomerization. hIAPP with a purity >99% was purchased from Bachem (Bubendorf, Switzerland) and 1,1,1,3,3,3-hexa-fluoro-2-propanol (HFIP) with a purity of 99% w from Sigma-Aldrich. In order to prevent preformed aggregates, hIAPP was initially dissolved in HFIP at a concentration of 1 mg/ml and then lyophilized overnight. The lyophilized powder was dissolved in dimethyl sulfoxide (DMSO) to obtain a stock solution with a final concentration of 100 μM. Each stock solution of hIAPP was used immediately after preparation.

Metal colloids and SERS. Colloidal metal nanoparticles (NPs) were prepared by pulsed laser ablation in liquid using the second harmonic (532 nm) output of a Nd:YAG laser operating at 10 Hz and at a fluence of 2 J/cm^2. A metal target (Ag) submerged in 5 ml of Millipore water was used as a substrate for ablation(Compagnini *et al.* 2007). The ablation process lasted 10 minutes. The concentration of the colloidal dispersion was estimated to be at around $1 \cdot 10^{-5}$M, assuming an average molar extinction coefficient (Messina *et al.* 2012; Fazio *et al.* 2013) of $2.5 \cdot 10^4$ M^{-1}cm^{-1}.The size (20 nm) of metal NPs was estimated to be 35 nm by AFM. SERS analysis was performed with 532 nm laser irradiation in the backscattering mode using a Witec Alpha 300 RS instrument. All the spectra were acquired in water containing 10^{-3} M of NaCl (high salt concentration, NPs self-aggregate into large size assemblies so decreasing the SERS effect) with a concentration of 0.25 μM of hIAPP.

In SERS, the intensity of the Raman spectrum is greatly enhanced by metal (typically Ag or Au) particles or nanostructures. This can increase the intensity of the Raman spectrum up to a factor of 10^{10} - 10^{11} (see, e.g., Le Ru, *et al.* 2007)). Proteins' secondary structures can be determined by SERS by following the characteristic bands associated with the CONH (amide) group. The so-called amide I and III are the most sensitive bands. Typically, an α-helix shows a band in the range of 1640-1658 cm^{-1} (amide I) and 1264-1272 cm^{-1} (amide III). β-sheet bands are in the range 1665-1680 cm^{-1} (amide I) and 1227-1242 cm^{-1} (amide III) while random coil lies in the range 1660-1665 cm^{-1} (amide I) and 1230-1240 cm^{-1} (amide III) (Rygula *et al.* 2013). These wavenumbers are red shifted by about 20-30 cm^{-1} in a SERS measurement[4]. The amide I band is the most sensitive to secondary structure (Kurouski *et al.* 2015). The amide III band gives information about the oligomerization state of amyloidogenic proteins. For instance, Aβ(1-42) oligomers show a red shift of 27 cm^{-1} with respect to monomers (Wang *et al.* 2013).

Simulation methods and setup

Amyloidogenic proteins simulations were carried out using Gromacs software (Hess *et al*. 2008). In all of the cases, the systems were first energy minimized and pre-equilibrated by using the steepest descents algorithm and successive 1 ns simulations in the NVT and NPT ensembles. With the exception of the 20-mer Aβ (1-40) system, all systems were simulated in explicit water. The details of the implicit solvent simulations are given in Sec. 3.2.3

The common details for all explicit solvent simulations are the following: the GROMOS 54A7 force field (Schmid *et al.* 2011) was used for the hIAPP and Aβ systems, and the GROMOS 53A6 (Oostenbrink *et al.* 2004) for the ovine prion system. SPC (simple point charge) water model (Berendsen *et al.* 1981) was employed. The time step was set to 2 fs and the temperature was kept constant at 300 K using the V-rescale algorithm (Bussi *et al.* 2007) with the time constant set to 0.1 ps. Periodic boundary conditions were applied and the Parrinello-Rahman algorithm (Parrinello and Rahman 1981) was applied for isotropic pressure coupling (1 bar) with the exception of the prion simulation that used the Berendsen algorithms (Berendsen *et al.* 1984*)*. The particle-mesh Ewald (PME) algorithm (Darden *et al.* 1993) was used for electrostatic interactions and overall charge neutrality was preserved by adding counterions when necessary. A solution of 0.1 M NaCl was used in all explicit solvent systems. All simulations were repeated at least three times. For secondary structure analysis DSSP (Kabsch and Sander 1983) and STRIDE (Frishman and Argos 1995) were used. To monitor equilibration, RMSD (the root mean square deviation) was used. The rest of the details of each of the systems are provided below.

Simulation details of the hIAPP systems. The NMR structure of hIAPP bound to sodium dodecylsufate (SDS) micelles (pdb ID: 2KB8 (Patil *et al.* 2009); the first structure in the PDB file) was used as the initial structure for the hIAPP monomer. Monomer, dimer, tetramer and hexamer systems were simulated. Details of the individual systems are listed next.

hIAPP monomers. The hIAPP monomer was placed in a cubic solvent box of linear length 7.7 nm. Each of the three independent simulations lasted 400 ns.

hIAPP dimers. Final independent configurations from the hIAPP monomer simulations were used to construct a dimer. Initially, the monomers were placed 4.7 nm from each other. The dimer was then simulated for 500 ns (for each independent simulation run) at 300 K. The simulation box was of the same size as above.

hIAPP tetramers. The last frames from independent dimer simulations were used to construct a tetramer. A larger simulation box (16 x 10 x 10 nm^3) was used. Initially, the two dimers were placed 5 nm away from each other to eliminate unwanted self-interactions. Each of the three independent

simulations was run for a total of 1050 ns using a heating cycle to eliminate possible kinetic traps as follows: 1) 300 ns at 300 K. 2) Followed by 20 ns at 400 K, and 3) finally 730 ns at 300 K.

hIAPP hexamers. Final configurations from tetramer simulations and two monomers from the monomer simulations were used to set up the three independent simulations. Initially, the components were placed about 4 nm away from each other. The same simulation box size as for the tetramer systems was used. The hexamer systems were simulated for a total of 1020 ns each run using a heating cycle as follows: 1) 300 ns at 300 K, 2) 20 ns at 400 K, 3) 250 ns at 300 K, 4) 50 ns at 400 K, 5) 100 ns at 300 K, 6) 200 ns at 400 K, and 7) 100 ns at 300 K.

hIAPP hexamers in pure water. Using the final configurations obtained after 1020 ns of simulation from the previous section, salt was removed (the counterions were left to ensure overall charge neutrality). The hexamer in pure water was then simulated for 500 ns at 300 K.

Aβ (1-40) systems. To investigate the generality of the partially ordered-disordered periodic arrays, we extended our investigation to three additional systems.

hIAPP:Aβ (1-40) 1:1. Monomer structures for hIAPP and Aβ (1-40) were taken from above. Each of the three independent systems was simulated for 500 ns at 300 K.

Implicit solvent simulations. In addition to Aβ (1-40) in explicit solvent, 20-mer Aβ (1-40) systems using the implicit solvent generalized Born surface area (GBSA (Qiu *et al.* 1997)) method were simulated. These simulations were performed in the NVT ensemble using the Langevin thermostat (Grest and Kremer 1986). The OPLS/AA force field (Kaminski *et al.* 2001) was used for the peptides and no NaCl was added. As with other systems, three independent simulations were performed each simulation being 500 ns.

Ovine prion systems. Prion (Pappalardo *et al.* 2007) aggregation was simulated by taking 18 molecules of H2H3 from OvPrP (PDB ID 1UW3, residues from C182 to C217) (Haire *et al.* 2004) in a β-rich conformation, obtained from a prior MD simulation (Chakroun *et al.* 2010) in water. The 18 prion monomers were placed in random orientation on a grid in a rectangular simulation box of dimensions 14 nm x 14 nm x 10 nm with 1.4 nm spacing and 63,668 water molecules. Each of the three individual trajectories was 200 ns. In the prion systems, Cys182 and Cys217 were connected with a disulphide bridge (Haire *et al.* 2004).

Superoxide dismutase. For atomistic molecular dynamics simulations of Superoxide dismutase (SOD) we used the SOD dimeric crystallographic structure [PDB id: 1SPD (Deng *et al.* 1993)] the

PDB structure of SOD was solvated CHARMM-GUI (Lee *et al.* 2016). A 150 mM concentration of KCl was added to mimic physiological condition.

The CHARMM36m force field (Huang *et al.* 2017) was employed to describe the protein structure, salt ions and TIP3 water model.

The system was first energy minimized using the steepest descend algorithm followed by an equilibration step under NpT conditions for 10 ns. In this stage, all protein atoms were restrained with a harmonic potential with a force constant of 1000 kJ mol^{-1} nm^{-2}. The temperature was kept constant at 310 K using the Nose-Hoover thermostat (Nosé and Klein 1983; Evans and Holian 1985) with a time constant of 1.0 ps. Isotropic pressure coupling scheme was applied using the Berendsen algorithm (Berendsen *et al.* 1984) with a time constant set to 5.0 ps. The Verlet scheme, with a cut-off distance of 1.2 nm is set to search the short-range neighbours every 20 steps. Particle mesh Ewald method (Darden *et al.* 1993) was used to handle the electrostatic interactions. The cut-off length of 1.2 nm was used for both electrostatic and van der Waals interactions. To constrain the hydrogen bonds, the LINCS algorithm (Hess *et al.* 1997) was employed and periodic boundary conditions were applied in all directions. For the production MD runs, we removed all the restraints applied to the proteins and used the and Parrinello-Rahman barostat (Bussi *et al.* 2007) instead of the Berendsen's algorithm (Berendsen *et al.* 1984). See **Table 1** for simulations details.

All simulations were carried out using an integration time step of 2 fs using the GROMACS 2018 simulation package (Abraham *et al.* 2015).

Table 1. Simulations Details.

# Protein	# Water	# K +	# Cl -	Time (ns)	Temperature (K)
2	49001	153	139	1200	310 – 372*
2	49001	153	139	200	310

* The temperature has been increased linearly from 310 to 372 K during the first 500 ns. The temperature was then kept constant at 372 K until the system reached 1200 ns.

Theoretical Methods

The famous Peierls instability (Peierls 2001) shares significant resemblances (but there are also significant differences) with the systems investigated here. In the essence, the Peierls Theorem

(Peierls 2001) states that a system consisting of a one-dimensional crystal with constant lattice spacing and electrons becomes unstable and develops a periodic modulation of electron density if there is any coupling between the electrons and the lattice. This is due to a competition between the electronic and elastic energies. An application is provided by poly-acetylene (Yannoni and Clarke 1983; Fuß *et al.* 2000) (…-CH=CH-CH=CH-…) and chemically related molecules: Experimental (Yannoni and Clarke 1983) and theoretical (Fuß *et al.* 2000) data confirm the onset of less symmetric structures, where alternating array of electron-rich short (1.36 Å) and electron-poor long (1.44 Å) bonds is preserved. Analogously, shape instabilities arise from the competition between electrostatic repulsion and surface tension: increasing the size of a droplet beyond a critical size may lead to capillary instabilities that eventually breaks larger droplets into smaller ones (Tamashiro and Schiessel 2006). For instance, uniformly charged tubules made up of identical surfactant molecules can become unstable evolving toward less symmetric structures such as undulated cylinders or arrays of juxtaposed spheroidal micelles (pearling instability (Grason and Santangelo 2006; Nguyen *et al.* 2005)). Similar effects were also observed in lyotropic lamellar systems which show alternation in the lamellar repeat distance (Porcar *et al.* 2000; Harries *et al.* 2006; Del Favero *et al.* 2009).

We start by developing an idealized model to investigate if a linear aggregate of interacting peptides (proteins) that exist in different conformational states may exhibit a transition from a *homogeneous* to a space-modulated structure.

Free energy

We first consider infinitely long one-dimensional (1-D) pre-fibrillar aggregates along the *z*-axis. The opposite cases of a dimeric structure and the extension to the 3-D infinite aggregates will be discussed separately. Proteins exhibit conformational flexibility that, for the sake of simplicity, is restricted to two interchanging states alone: Φ and Ψ. The ordered Ψ arrangement comprises α-helices and β-strands, while the disordered Φ conformation includes coils and turns. A description of protein structure and function in terms of pure Φ and Ψ states is a gross simplification. In the real world, proteins can be described by a weighted combination of Φ and Ψ states. Usually, the ordered conformations prevail at room temperature (and in absence of denaturating chemicals), while, in the case of IDPs, the relative abundance of ordered and disordered domains within the same molecule is comparable over a wide range of experimental conditions.

The total free energy of proteins array, G, can be decomposed into three main contributions: a) a term related to the entropy of mixing ($T \cdot S_{MIXING}$) which favors an identical of Φ and Ψ states, b)

the self-energy (U_{SELF}) which measures the stability of a given conformation by specific intramolecular bonds and solvent interactions, and c) the energy of interaction (U_{INT}), which expresses the conformation-dependent interaction between nearest-neighbor proteins. Combining the three contributions gives the total free energy

$$G = -TS_{MIXING} + U_{SELF} + U_{INT}, \qquad (1)$$

Next, we develop functional forms for the individual terms.

Let φ_n be the local fraction of Φ conformations of a protein at site *n* along the 1-D array, and 1-φ_n the fraction of Ψ conformations. Standard mean-field expression for Φ and Ψ mixing entropy S_{MIX} reads

$$-TS_{MIX} \approx k_B T \sum_n^N [\varphi_n \log \varphi_n + (1-\varphi_n)\log(1-\varphi_n)], \qquad (2)$$

where k_B is the Boltzmann constant, T the absolute temperature and the sum spans over the N molecules of the aggregate.

The simplest expression for the self-energy of the array is

$$U_{SELF} = \sum_n^N [\varphi_n g_\Phi + (1-\varphi_n) g_\Psi)], \qquad (3)$$

Where g_Φ and g_Ψ are the internal energies of the Φ and Ψ conformations, respectively. Notice that g_Φ and g_Ψ are strongly affected by the interactions of the protein with its environment.

In the nearest-neighbor approximation, the energy of interaction for a 1-D array of self-aggregated proteins can be written as

$$U_{INT} = \tfrac{1}{2}\sum_n^N u_{INT}\left(\left|z_n - z_{n\pm1}\right|, \varphi_n\right), \qquad (4)$$

where u_{INT} is the potential acting between nearest-neighbor proteins at positions z_n and z_{n+1}. We define it as

$$u_{INT}\left(\left|z_n - z_{n\pm1}\right|, \varphi_n\right) \approx \tfrac{1}{2}[-Pe^{-\gamma_P(\varphi_n)\cdot(z_n - z_{n\pm1})} + Qe^{-\gamma_Q(\varphi_n)\cdot(z_n - z_{n\pm1})}],$$

where the positive constants *P* and *Q* measure the strengths of the attractive and repulsive interactions, respectively, while $\gamma_P(\varphi_n)$ and $\gamma_Q(\varphi_n)$ (with $\gamma_Q > \gamma_P$) measure their decay lengths. The parameters P, Q, $\gamma_P(\varphi_n)$ and $\gamma_Q(\varphi_n)$ can be related to experimentally accessible quantities as shown later on. Since conformational arrangements strongly change upon aggregation (Gsponer and Vendruscolo 2006), also the decay profile of the intermolecular protein-protein interactions must change. These variations are mainly associated to the proteins changing from an ordered to a disordered conformation. In the following we assume that the disordered Φ configuration is more swollen than

the ordered Ψ arrangement where the elongate shape enables a tighter packing (in SI we report some conformation-related gyration radii of IDPs).

In order to allow the proteins within an aggregate to have different conformations, we introduce a modulation of the decay lengths $\gamma_i(\varphi_n)$ (with i =P or Q) linked to the local conformational population (probability) φ_n. To the lowest order

$$\gamma_i(\varphi_n) \approx \gamma_i(\bar{\varphi}) + \frac{\partial \gamma_i}{\partial \bar{\varphi}} \cdot (\varphi_n - \bar{\varphi}), \qquad (5)$$

where $\bar{\varphi}$ is the averaged conformational population for a homogeneous equally-spaced array. The equation for $\gamma_i(\varphi_n)$ allows for the length scales of intermolecular interactions to be modulated around their average value depending on the parameters $\bar{\gamma}_i^* \equiv \frac{\partial \gamma_i}{\partial \bar{\varphi}}$ that measure the response of γ_i to variations of the conformational population $\bar{\varphi}$.

First, the free energy, Eq.(1), is minimized for a homogeneous array $\varphi_n = \bar{\varphi}$ and $|z_n - z_{n+1}|$=$\bar{a}$, where $\bar{a}$ is the protein-protein mean distance in the homogeneous state. This procedure, shown in SI (Eq.(17Sa)), yields

$$g_\Phi - g_\Psi + k_B T \log \frac{\bar{\varphi}}{1-\bar{\varphi}} + |\alpha| \Lambda\, \bar{a} = 0, \qquad (6)$$

where $\Lambda \equiv \frac{\bar{\gamma}_Q \bar{\gamma}_P^* - \bar{\gamma}_P \bar{\gamma}_Q^*}{\bar{\gamma}_Q - \bar{\gamma}_P}$ and $-|\alpha|$ is the protein-protein adhesion energy (a thorough discussion will be made after Eq.(13)). Solution to Eq.(6) yields an expression for the averaged conformational population $\bar{\varphi}$ in an equally-spaced 1-D lattice.

Stability criterion

The next step is the calculation of the fluctuations around $\bar{a}$ and $\bar{\varphi}$. This is most conveniently done at the continuum limit by replacing the discrete variables φ_n and $z_n - z_{n\pm1}$ by their continuum analogues.

$$z_n - z_{n\pm1} \approx \bar{a} \pm \frac{\partial \eta(z)}{\partial z}\bar{a} + \frac{1}{2}\frac{\partial^2 \eta(z)}{\partial z^2}\bar{a}^2 + \ldots \;, \quad \varphi_n = \varphi(z) \qquad (7)$$

where $\bar{a}$ is the average distance between two nearby proteins. Then, inserting Eq.(7) in Eq.(1), expanding it in power series for $\varphi(z) - \bar{\varphi}$, $\partial\eta(z)/\partial z$ and $\partial^2\eta(z)/\partial z^2$ and retaining terms up to second order gives (see SI, Eq.(17Sb))

$$G \approx G_o + \bar{a}^{-1} \int_{-\ell}^{+\ell} \left[\tfrac{1}{2} A_1 \left(\frac{\partial^2 \eta}{\partial z^2} \right)^2 + \tfrac{1}{2} A_2 \left(\frac{\partial \eta}{\partial z} \right)^2 + \tfrac{1}{2} A_3 (\varphi - \bar{\varphi})^2 + A_4 \frac{\partial^2 \eta}{\partial z^2} (\varphi - \bar{\varphi}) \right] dz, \quad (8)$$

where G_o is the total energy without fluctuations, $2\ell = N\bar{a}$ is the length of an aggregate of N proteins, and $A_1 = \frac{1}{4}|\alpha|\bar{a}^4 \bar{\gamma}_P \bar{\gamma}_Q$, $A_2 = |\alpha|\bar{a}^2 \bar{\gamma}_P \bar{\gamma}_Q$, $A_3 = k_B T[\bar{\varphi}(1-\bar{\varphi})]^{-1} - |\alpha|\bar{a}^2 \frac{(\bar{\gamma}_Q^*)^2 \bar{\gamma}_P - (\bar{\gamma}_P^*)^2 \bar{\gamma}_Q}{\bar{\gamma}_Q - \bar{\gamma}_P}$ and $A_4 = \frac{1}{2}|\alpha|\bar{a}^3 \bar{\gamma}_P \bar{\gamma}_Q \left(\frac{\bar{\gamma}_Q^* - \bar{\gamma}_P^*}{\bar{\gamma}_Q - \bar{\gamma}_P} \right)^2$. With Eq.(8), by denoting the monomer mass by m and introducing a kinetic energy term as $K = \frac{\bar{a}^{-1}}{2} \int_{-\ell}^{+\ell} \left[m \left(\frac{\partial \eta}{\partial t} \right)^2 \mathrm{dz} \right]$, we can now construct a Lagrangian for the system, $\mathcal{L} = K - U$. Let $I \equiv \int_0^t \mathcal{L} dt$, the motion is such that the variation of I is zero (see, e.g., (Goldstein *et al.* 2001). This procedure leads to the Euler-Lagrange equations shown in eq.(18S). By using the explicit form of $\mathcal{L}$ derived above, eventually we obtain a simple system of two linear partial differential equations

$$-\rho\bar{a} \frac{\partial^2 \eta}{\partial t^2} = A_1 \frac{\partial^4 \eta}{\partial z^4} - A_2 \frac{\partial^2 \eta}{\partial z^2} + A_4 \frac{\partial^2 \varphi}{\partial z^2}, \qquad (9a)$$

$$0 = A_3(\phi - \bar{\phi}) + A_4 \frac{\partial^2 \eta}{\partial z^2}, \qquad (9b)$$

where $\rho \equiv m/\bar{a}$ is the linear density of the protein array. Assuming symmetry across zero, use of Fourier series expansions: $\eta = \sum_q \eta_q(t) e^{iqz} + \text{c.c.}$ and: $\phi - \bar{\phi} = \sum_q \phi_q(t) e^{iqz} + \text{c.c.}$ (c.c. = Conjugated Complex), yields from Eqs.(9)

$$\frac{d^2 \eta_q(t)}{dt^2} = \Theta^2(q) \eta_q(t) \qquad \phi_q(t) = q^2 \frac{A_4}{A_3} \eta_q(t), \qquad (10)$$

where: $\Theta^2(q) \equiv \frac{1}{\rho\bar{a}} \left(\left(A_1 - \frac{A_2^2}{A_3} \right) q^4 + A_2 q^2 \right)$ depends on the wave number q. It can be noted that when $\Theta(q)>0$ (stable region), the solutions to Eq.(10) are: $\eta_q \propto \exp(i|\Theta(q)|t)$ and $\varphi_q \propto \exp(i|\Theta(q)|t)$, that is, both protein-protein distance and the protein conformational population oscillate near the equilibrium values $\bar{a}$ and $\bar{\varphi}$. The amplitudes of those oscillations are usually small and are calculated by adding to Eq.(10) a Fourier series representation of the thermal noise. The resulting Langevin's equation can be solved by well-known procedures (Risken 1984). On the contrary, when $\Theta(q)<0$ (unstable region), the solutions are $\eta_q \propto \exp(+|\Theta(q)|t)$ and $\varphi_q \propto \exp(+|\Theta(q)|t)$. In other words, the system becomes unstable and even the smallest fluctuations applied at $t = 0$ grow exponentially in time. The growth of the protein distance η and that of the

conformational population φ suggests that, beyond some critical values, the homogeneous system is no longer stable against infinitesimal fluctuations and breaks down into *a patched structure*. This behavior is akin to spinodal decomposition (Onuki 2002).

It is worth mentioning that not all fluctuations in the unstable region grow at the same rate: those with the largest $\Theta(q)$ will grow the fastest. The q value which maximizes $\Theta(q)$, q^* say, is obtained from the condition: $\frac{\partial\Theta(q)}{\partial q} = 0$. The final dimensions of the protein domains (defined as the protein clusters inside the patterned aggregate having similar conformational population) will keep memory of this dynamic processes and would be of order q^*. Using the explicit expressions for $\Theta(q)$ (Eq.(10)) and A_i (Eq.(8)) and using $\frac{\partial\Theta(q)}{\partial q} = 0$, we find that the fastest growth rate of the patterned structures arises when

$$q^*\bar{a} = \left(-2\frac{\frac{k_BT}{\bar{\varphi}(1-\bar{\varphi})}-|\alpha|\bar{a}^2 f_1}{\frac{k_BT}{\bar{\varphi}(1-\bar{\varphi})}-|\alpha|\bar{a}^2(f_1+f_2)}\right)^{1/2} \qquad (11)$$

where: $f_1 = \frac{(\bar{\gamma}_Q^*)^2\bar{\gamma}_P-(\bar{\gamma}_P^*)^2\bar{\gamma}_Q}{\bar{\gamma}_Q-\bar{\gamma}_P}$ and $f_2 = \bar{\gamma}_P\bar{\gamma}_Q\left(\frac{\bar{\gamma}_Q^*-\bar{\gamma}_P^*}{\bar{\gamma}_Q-\bar{\gamma}_P}\right)^2$. Equation (11) is the main result of the theory, predicting stability and wavelength (size) of the patterned structures. Indeed, only if the right-hand side of Eq.(11) is real, stable patterned structures may exist. This condition is satisfied only if in Eq.(11) the following inequalities are fulfilled: $|\alpha|\bar{a}^2 f_1 < \frac{k_BT}{\bar{\varphi}(1-\bar{\varphi})} < |\alpha|\bar{a}^2(f_1+f_2)$. Near the lower boundary, $\frac{k_BT}{\bar{\varphi}(1-\bar{\varphi})} \approx |\alpha|\bar{a}^2(f_1+f_2)$, the patterned structures exhibit the highest $q^* \to q_{max}$ (i.e., a close alternation of ordered and disordered proteins). On the contrary, on approaching the upper boundary, $\frac{k_BT}{\bar{\varphi}(1-\bar{\varphi})} \approx |\alpha|\bar{a}^2 f_1$, we find: $q^* \to 0$, that is, the patterned phase contains very large domains. To calculate the boundaries in a phase diagram, we need the concentrations of ordered and disordered conformations, $\bar{\varphi}$ and $(1-\bar{\varphi})$, given by Eq.(6) and depend on the same parameters that affect the protein-protein interaction. Whence, combing Eqs.(6) and (11) the boundaries between homogeneous and patched aggregates are obtained solving the algebraic equations

$$\frac{\left(1+\exp(-\frac{\Delta G_{UNF}(|\alpha|)}{k_BT})\right)^2}{\exp(-\frac{\Delta G_{UNF}(|\alpha|)}{k_BT})} = \frac{|\alpha|\bar{a}^2}{k_BT}(f_1+f_2), \qquad \frac{\left(1+\exp(-\frac{\Delta G_{UNF}(|\alpha|)}{k_BT})\right)^2}{\exp(-\frac{\Delta G_{UNF}(|\alpha|)}{k_BT})} = \frac{|\alpha|\bar{a}^2}{k_BT}f_1 \qquad (12a)$$

where:

$$\Delta G_{UNF}(|\alpha|) \equiv g_\Phi - g_\Psi + |\alpha|\Lambda\bar{a} = \Delta G^o_{UNF} + |\alpha|\Lambda\bar{a} \qquad (12b)$$

is the unfolding energy of the protein self-aggregate. The energy difference $g_\Phi - g_\Psi$ can be identified with the unfolding free energy of the *isolated* protein, ΔG^o_{UNF}. On the contrary, no direct measurements of the self-adhesion energy $|\alpha|$are available, while the experimental free energy variation upon the assembly of N identical proteins reads: $\Delta G_{ADH} = \frac{1}{N}(U(N) - N \cdot U(N=1))$. This formula accounts for the reorganization energy of the proteins self-energy upon the formation of a *N-mer* and it is valid in the limit *N>>1* (negligible end effects). Exploiting Eqs.(1)-(5), we derived in SI (Eq.27S) a compact expression for ΔG_{ADH} (per protein molecule)as a function of $|\alpha|$ for a 1-D aggregate

$$|\Delta G_{ADH}| \approx |\alpha|\left(1 - \frac{\exp\left(-\frac{\Delta G^o_{UNF}}{k_B T}\right)}{\frac{1}{2}\left(1+\exp\left(-\frac{\Delta G^o_{UNF}}{k_B T}\right)\right)^2}\frac{|\alpha|}{k_B T}\Lambda^2\bar{a}^2 + O\left(\left(\frac{|\alpha|}{k_B T}\Lambda^2\bar{a}^2\right)^2\right)\right) \qquad (13)$$

showing how the self-adhesion energy $|\alpha|$is modified by the reorganization effects (the second term in the right hand side of Eq.(13)).

By eliminating$|\alpha|$ from the system of equations (12a,b) and (13), we may calculate the ΔG_{ADH} vs. ΔG^0_{UNF} phase diagram. Specifically, if a point defined by the two accessible control parameters ΔG_{ADH} and ΔG^0_{UNF} falls inside the unstable region, large fluctuations in conformational population and intermolecular distances begin to develop within the aggregate, eventually leading to patterned structures. On the contrary, inside the stable region, the protein aggregate remains homogeneous.

So, while the requirements for the formation of patterned structures can be predicted with a good degree of accuracy and the final morphology can be somehow inferred by Eq. (11), what is lost in our linear stability analysis is the final conformational composition of the mature protein aggregates. This goal can be reached only by including higher order terms in the series expansion of the energy functional (Eq. 8). The mathematics, however, becomes very involved (non-linear soliton-like partial differential equations (Drazin and Johnson 1989)) and the number of parameters increases, this task is beyond the aims of the present study and will be addressed in the next section by MD simulations. Accordingly, one expects to observe, under particular conditions ($\Theta(q)<0$), an alternating ordered array of small domains richer in Φ and Ψ arrangements, respectively as shown in Fig.1. We have extended the theory to 3-D aggregates, details are reported in SI. In the limit of isotropic interactions, the final formulas remain unchanged, apart from a renormalization of the wavenumber q. The calculated final arrangements for the 1-D and 3-D cases are pictorially shown in Fig.1.

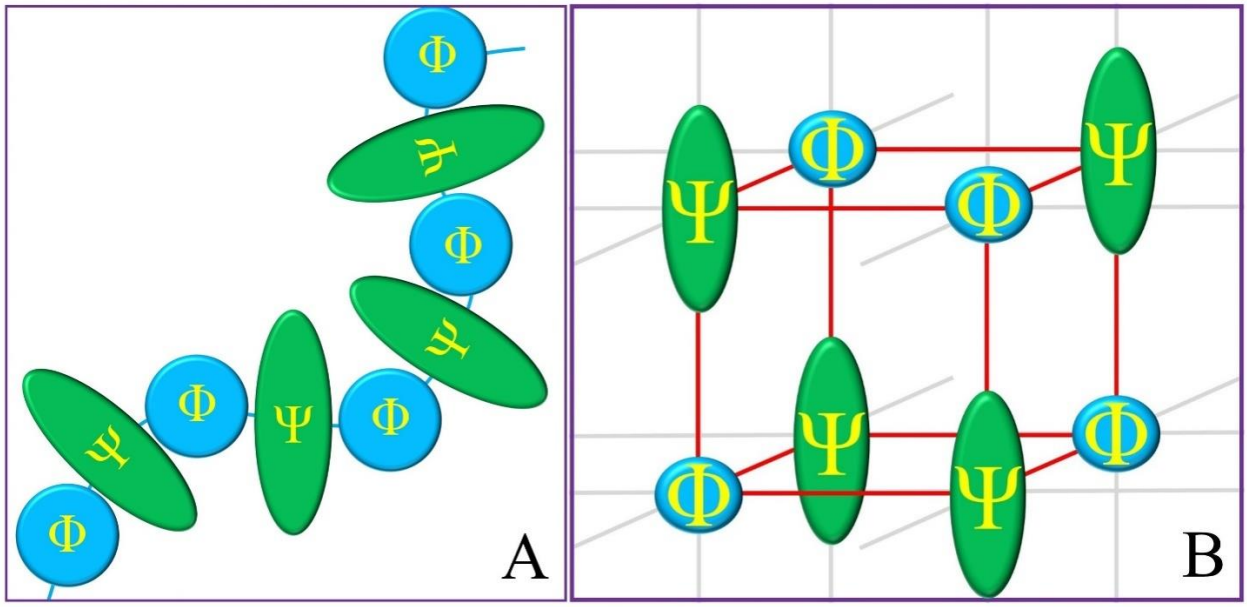

Figure 1. The most likely configurations of two-state proteins above a critical threshold (see the text for the definition). For the sake of clarity, proteins have been depicted by the pure Φ and Ψ states (fully disordered and fully ordered configurations, respectively). In our model, proteins are described by a weighted combination of Φ and Ψ states. Panel A: 1-D arrangement. Panel B: 3-D arrangement.

Then, we numerically solved the system of equations (12a,b) and (13) in order to obtain a qualitative phase diagram reported in Fig. 2. The drawing shows the boundaries among stable and unstable regions as a function of two control parameters. We selected as parameters two experimental quantities: the (attractive) protein-protein interaction free energy, ΔG_{ADH} against the unfolding free energy ΔG_{UNF}. If a point identified by the above defined control parameters falls inside the unstable region, large fluctuations in conformational population and intermolecular distances begin to develop, eventually leading to a patterned structure. Conversely, inside the stable region the protein aggregate remains homogeneous. Figures have been calculated for selected values of the parameters $\bar{\gamma}_i$, $\bar{\gamma}_i^*$ and $\bar{a}$. The phase diagrams in Fig. 2 clearly shows two distinct regions, homogeneous and patched, respectively.

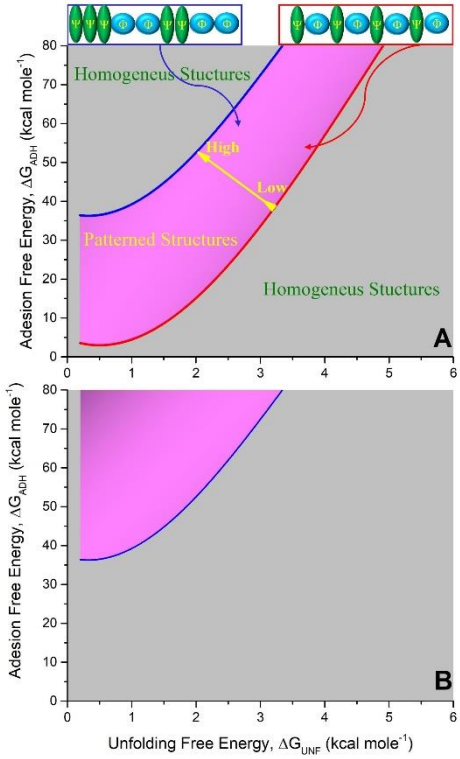

Figure 2. Phase diagram showing the stability regions (homogeneous and patterned structures) for a 1-D array of proteins with internal conformational flexibility. The protein-protein self-adhesion energy, ΔG_{ADH}, is plotted against the unfolding energy ΔG^0_{UNF} of the isolated monomer. ***Panel A***: strong geometrical variations upon protein conformational transition, ***panel B***: weak geometrical deformations (20% of those of panel A). *Large* and *small* labels the size of the patterned structures, pictorially sketched at the top of the figure.

Lastly, we repeated the calculations in the simpler case of dimeric protein aggregates. Results, reported below and in SI, predict a qualitatively similar behavior as that of infinite arrays.

Main qualitative conclusions from the model

Let us summarize the main features of the theoretical model before discussing the MD simulations and SERS experiments. An equally spaced 1-D array made of protein repeat units that may assume two different geometrical arrangements becomes unstable when: i) the internal energies of Φ (disordered) and Ψ (ordered) arrangements are slightly different (of order of $1k_BT \approx 0.5\ kcalmole^{-1}$ at room temperature); ii) there are large differences in the geometries of the Φ and Ψ arrangements, and iii) there are strong interactions between the proteins (of order of $10\ k_BT$). The phase diagram for a 1-D array of proteins is shown in Fig. 2. By comparison, it is evident that strong

protein-protein interactions stabilize the patterned structures, provided a critical strength of the interactions (of order of a few k_BT) has been reached. Thus, the formation of periodic structures is unlikely, but not unrealistic. For variation of the intermolecular forces decay length upon Φ to Ψ protein arrangement as small as 10%, and for internal energy difference between Φ and Ψ of order of a few k_BT and proteins radii within the experimental range, patterned 1-D structures emerge, provided the protein-protein interaction energy is of order of 10 k_BT. Obviously, other parameters, like the protein-protein mean distance $\bar{a}$, modify the detailed shape of the curves, so an "universal" phase diagram cannot be drawn even using more advanced models. Nevertheless, our simple theory unveils a close relationship between patterns geometry, self-adhesion energy and unfolding energy, a relation that would be obscured in a more detailed picture.

Once the system becomes unstable, fluctuations grow bringing the protein array toward less symmetric configurations. The growth of local heterogeneities depends on the wave number: patterns with the largest wave number q grow fast near the lower boundary of the instability region (see Fig. 2), while large patterns (i.e. $q \to 0$) appear on approaching the upper boundary. Thus, it is conceivable that the most likely final organization is given by an alternating array of mostly ordered and mostly disordered proteins as shown schematically in Fig. 1.

Results and Discussion

A comparison of the experimental data with model predictions

The qualitative model developed above was tested for some relevant classes of proteins with known assembly and folding free energies. The energies of aggregation, ΔG_{ADH}, and the energy difference between Φ and Ψ internal states in the monomeric state, ΔG^{o}_{UNF}, were estimated from calorimetric and chemical denaturation measurements. Fig. 3 show phase diagram of patterned and homogenous of some amyloidogenic and non-amyloidogenic proteins. The free energy of self-adhesion, ΔG_{ADH} (a negative quantity because of the favourable protein-protein interactions in self-assembled proteins) was calculated from the experimental dissociation energy: $\Delta G_{ADH} = -\Delta G_{DISS}$.

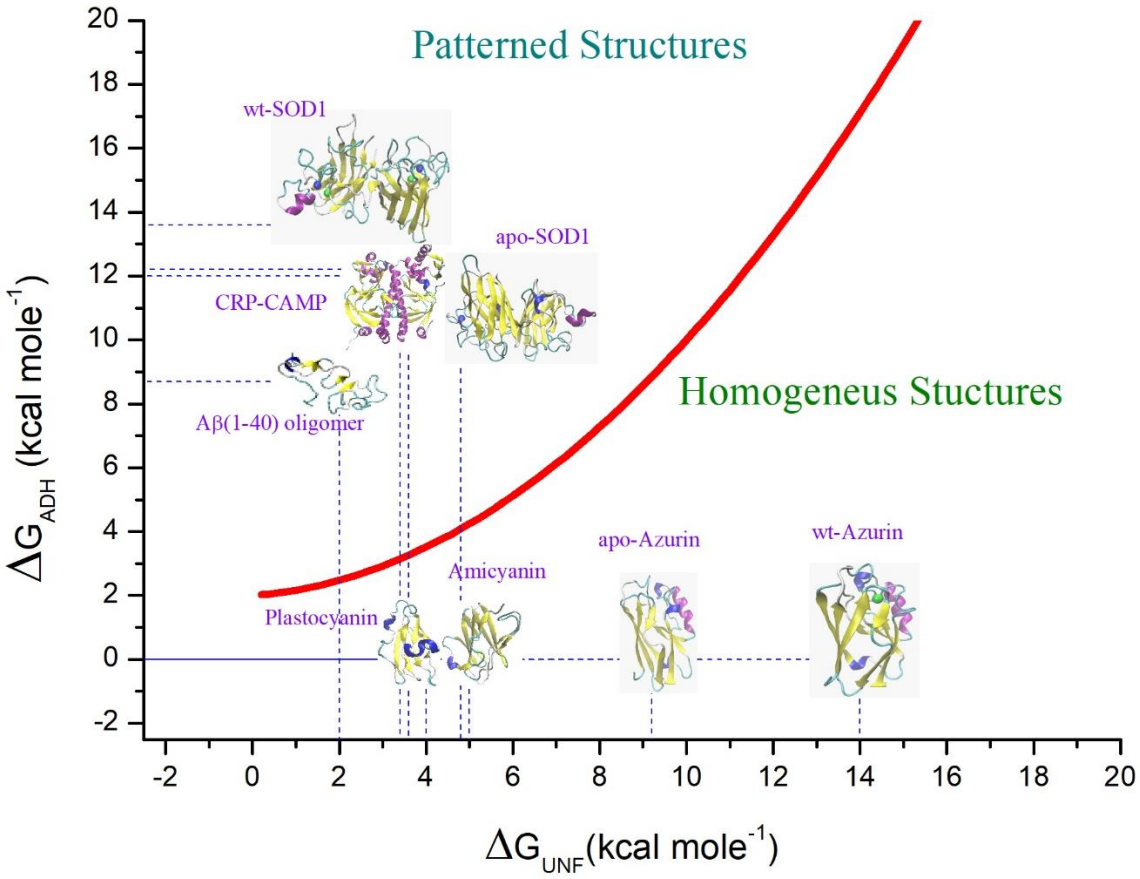

Figure. 3. Phase diagram (ΔG_{ADH} of self-adhesion versus ΔG^{o}_{UNF} of unfolding) of some proteins. The plot includes amyloidogenic, (Aβ(1-40) (Iljina *et al. 2016)*) and globular proteins in aqueous solutions (wt-SOD1 (Khare *et al.* 2004; Stathopulos *et al*. 2006), apo-SOD1 (Svensson *et al.* 2006, 2010), plastocyanin (Milardi *et al.* 1998), amicyanin (La Rosa *et al*. 2002), wt-azurin (La Rosa *et al.* 1995), apo-azurin (Pappalardo *et al.* 2008) and CRP-CAMP (Cheng *et al.* 1993)). Red line represents the qualitative boundary between patterned and homogenous aggregate structure. The *ΔG* values are reported in table S1 of SI.

Superoxide dismutase (SOD1) is a thermodynamically well characterized globular dimeric metal-protein. SOD1 undergoes thermal or chemical denaturation following a two-step pathway: in a first step the dimeric form dissociates into two monomers and then unfolding occurs with a Gibbs free energy change of 13.5 kcal mole^{-1} (dissociation) and 3.5 kcal mole monomer^{-1} (unfolding) respectively. This datum suggests that SOD1 lies in the patterned structures side of the phase diagram 3. In order to verify this conjecture, accurate MD simulations of SOD1 were performed. MD data are in agreement with model expectations, showing that the two subunits have a different number of amino acid in ordered conformations (sum of α-helix and β-sheet) as reported in Fig.S1. This difference decreases at increasing temperatures during unfolding process (Fig. S2). The dimer asymmetry was found also in the crystal structure (pdb code 1SPD (Deng *et al.* 1993)). In fact, the two monomers show 56% and 34% of amino acids with ordered secondary structures, respectively. On the contrary, globular proteins, such as blue-copper proteins, that do not show any tendency to

form multi-meric aggregates, lie in the homogeneous side of Fig. 3 because of the negligible self-adhesion energy and the large unfolding energy. Moreover, even though the crystallographic unit cell of azurin contains four proteins, no differences in the conformational population were detected. For additional information, Table S1 reports the Gibbs free energy of some patterned and homogeneous proteins.

Molecular dynamics test of the theoretical model

IAPP oligomers formation. During the simulation, hIAPP monomer (over 500 ns at 300 K in 0.1 M NaCl solution) undergoes a conformational transition toward a structure exhibiting an antiparallel cross β-sheet (residues 24-28 and 15-20, yellow color in Figure S3) with a connecting turn (cyan) encompassing residues 21-23 according to literature data (Milardi *et al.* 2008; Reddy *et al.* 2010; Sciacca *et al.* 2018). The secondary structure evolution of dimer and a snapshot of the final structure are shown in Fig. 5 (contact map is shown in Fig. S4). The first monomer retained its antiparallel β-conformation for 140±20 ns. Then it became partially ordered, with shorter β-sheets and a small but noticeable α-helical portion at one end. The second monomer preserved its antiparallel β-conformation for 80 ns, then quickly unfolded and became disordered with hints of short discontinued helices and a β-bridge. Interestingly, the arrangement of a partially ordered monomer and a largely disordered partner was found to maintain stable intermolecular contacts in the individual simulations. To best describe the dimer secondary structures, a statistical analysis over 500 ns of three independent runs was performed. Dimeric complexes were clustered based on their RMSD values using the Gromos algorithm (Daura et al. 1999) and a RMSD cut-off for two structures to be neighbors within 0.6 nm. This procedure yields 3 most populated clusters C1 (25%, rmsd=0.44 nm), C2 (24%, rmsd=0.38 nm) and C3 (21%, rmsd=0.51 nm). Other small clusters where found, the most rapresentative one being populated up to 5%. For each cluster, the most representative structure and the centroid of the structures in a given cluster, are represented in Figure 5.

These observations partially disagree with the implicit solvent simulations of Dupuis *et al.* (Dupuis *et al.* 2011) who found still an asymmetric composition of the dimer, but with a different conformational distribution. Our simulations do, however, agree with those of Derreumaux *et al.* (Tarus *et al.* 2015) who investigated the Aβ (1-40) dimers by atomistic simulations and found a limited percentage of helix and sheet conformations using both explicit solvent simulations as well implicit solvent methods for free energy analysis (Tarus *et al.* 2015; Nguyen *et al.* 2016) .

An interesting consequence of the order-disorder alternation is that the radii of gyration of the single peptides within the aggregate are not identical. This is clearly seen in Fig. 4 which shows them for two hIAPP monomers within a dimeric aggregate. Such an effect is difficult to detect in single snapshots from the aggregate, but becomes unambiguous after averaging over long times (see Fig. 5). Similar results were observed also in hexamers confirming that the heterogeneous nature of the peptides aggregates persists even after a long equilibration time.

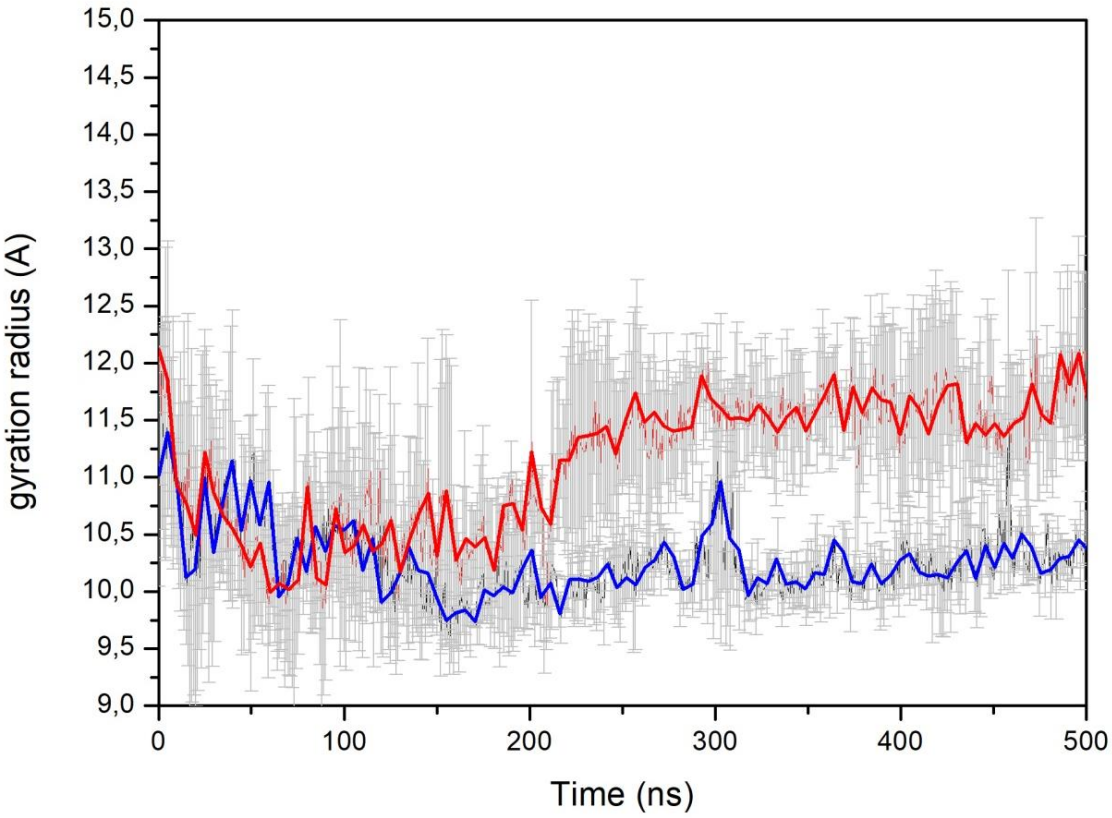

Figure 4. Radius of gyration of hIAPP dimer. Simulation of the early stage time evolution of a dimeric aggregate consisting of two *identical* monomers brought at constant distance at time t = 0 ns. Notice: 1) the formation of an asymmetric dimer made up of two juxtaposed unequally sized monomers at late times; 2) the intense noise due to the continuous conformational rearrangement obscures the dimer asymmetry unless a statistical analysis has been performed.

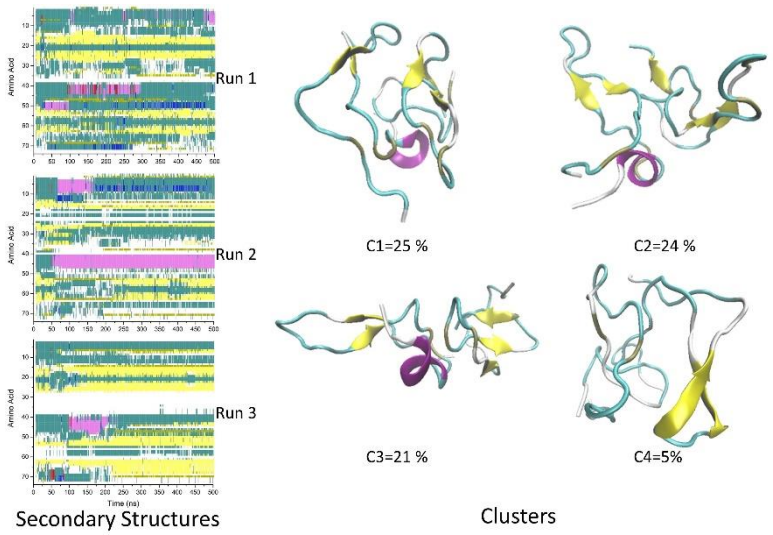

Figure 5. Left: DSSP plot of the secondary structure evolution of a hIAPP dimer over 500 ns of three independent runs. Right: Cluster analysis of dimeric assembly of hIAPP in 0.1 M electrolyte solution. Notice the coexistence of partially ordered (mainly β-sheets) and partially disordered domains in the dimeric aggregate. Color code: white-coil; light yellow-β-sheet; dark yellow-isolated bridge; pink-α-helix; blue-3-10 helix; red-π-helix; green-turn.

The tetramer resembles the dimer, i.e., the tetramer shows the partially ordered-disordered sequence (Figs. S5-S7).

The peptides self-assembled into a hexameric aggregate after about 50 ns. The aggregate exhibits the characteristic partially order-disorder arrangement as observed for dimers and tetramers. The average percentage of different secondary structures in the last 60 ns of the 1020 ns simulation (Fig. S8) shows random coil is the dominant conformation in all molecules while helical structures are rarely sampled. Interestingly, Fig. S8 also shows that the molecules with more β-strand tend to have less turns (molecules 2, 4, and 6) while molecules with less β-strand tend to have more of them (molecules 1, 3, and 5). Unlike with dimers and tetramers, the alternating ordered-disordered correlation between the different molecules in hexamers is less evident. Nonetheless, the strand-turn complementary shows an interesting partially ordered-disordered feature *within* each of the peptides in the hexamer aggregate (see fig.6D). The time evolution of the secondary structures is shown in Fig. S9. Contact map and a snapshot of the final hexameric hIAPP assembly are also reported (Fig. S10, S11). The radius of gyration of the hexamer hIAPP is reported in Fig S12.

Our MD results are summarized in Fig. 6 which clearly shows the order-disorder alternation in the conformational population of hIAPP monomers, dimers, tetramers and hexamers.

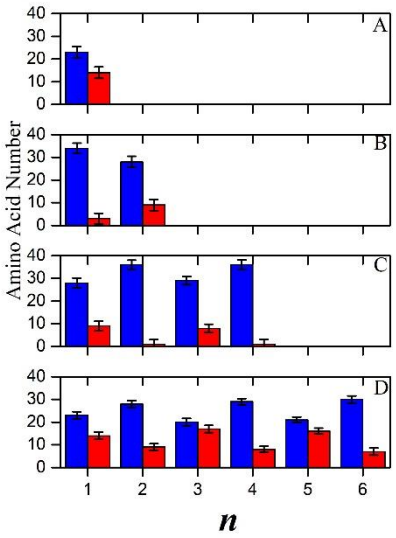

Figure 6. Fractions of ordered structures and their indeterminations, calculated by VMD and averaged over ten frames of the last 50 ns of simulation, along an array of hIAPP peptide suspended in a 0.1 M electrolyte solution. *n* labels the position of a generic *n-th* peptide within the linear aggregate. Blue bars: disordered secondary structures considered as a sum of coils and turns. Red bars: partially ordered secondary structures as a sum of α-helices and β-strands. A = monomer, B = dimer, C = tetramer, D = hexamer.

To further investigate the validity of the model, we performed MD simulations of the hexamer without salt (only counterions were present). The theory predicts the disappearance of the modulated structures for weakly adhesive assemblies (Fig. 2). When no added salt is present, the electrostatic repulsion is stronger due to lack of screening. Strong repulsion weakens the stability of the assembly and destroys the modulated phases. Results are summarized in Fig. S13 where we report the fraction of ordered (blue) and disordered (red) arrangements in a hIAPP hexamer aggregate in pure water.

The comparison with the corresponding system embedded in a 0.1 M electrolyte solution (Fig. 6D) is striking: virtually, no modulated phases exist in pure water and the distribution, apart from position n=3, resembles that of a monomer (Fig. 6D and S2). Interestingly, the ordered structures of hIAPP in pure water are mainly helices, while in 0.1 M electrolyte solution the most abundant component is the β-strand arrangement (Fig. S3). The strong sensitivity of the conformational landscape to electrolytes in amyloid Aβ-peptides has been recently shown by Smith and Cruz (Smith and Cruz 2013) using MD simulations. In addition, in another previous study using pure water no alternation was observed in aggregates of the closely related Aβ(1-42) fragment (Masman *et al*. 2009).

*Aβ (1-40) large oligomers formation.*Lastly, the analysis of the trajectories reported by Strodel *et al.* for a ribbon-like 20-mer aggregate of Aβ(1-40) (Barz *et al.* 2014) yields an alternate distribution along both the short and the long-axes of the ribbon (see Fig. 7).

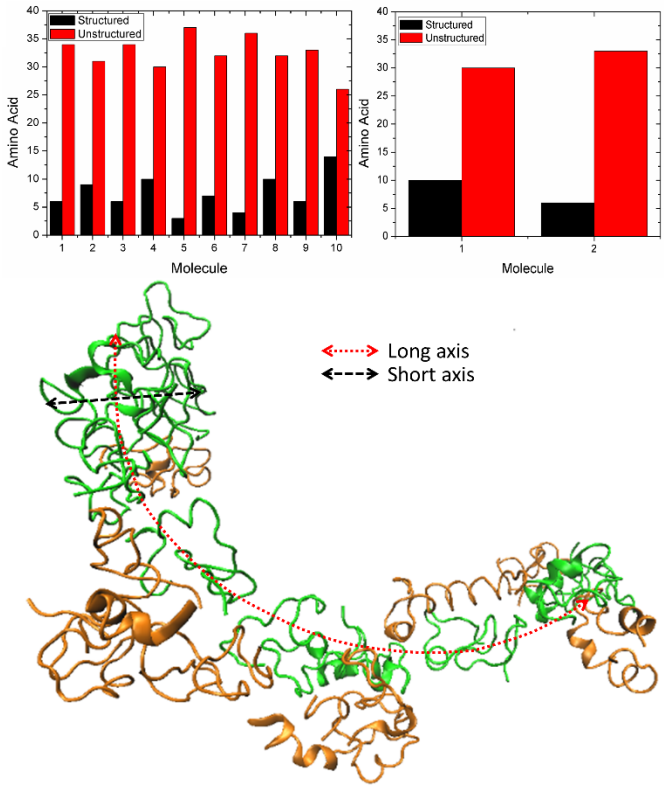

Figure 7. Top left histogram: fraction of ordered and unordered structures averaged over ten frames of the last 50 ns of simulation counted over the long axis (red axis of cartoon representation). Top right histogram: fraction of ordered and unordered structures evaluated over the short axis (black axis of cartoon representation). Below: cartoon representation of semi-toroidal aggregate formed by 20 Aβ(1-40) molecules after 500 ns of implicit solvent simulation.

Also, ovine prion protein and mixture of 1:1 Molar ratio of IAPP and Aβ(1-40) show the same behavior as reported in SI (see Fig.S15 and S16).

Surface-Enhanced Raman Scattering (SERS) test of the theoretical model

SERS is a technique able to detect the secondary structures of amyloidogenic peptides and proteins at very low concentrations. Although spatial and temporal resolutions are not comparable to MD, at low concentration SERS is able to reveal, with a good precision, the secondary structure of amyloidogenic proteins (D'Urso *et al.* 2018). Here we use SERS to determine inhomogeneous

secondary structures within small aggregates of IAPPs. SERS spectra of 10 nM hIAPP solution containing 1 μM of silver Nano Particles (NPs) were recorded at different times. Representative spectra of the most informative spectral regions are shown in Figs. S17-S19. As time proceeds, proteins self-aggregate from monomers to oligomers, and eventually to proto-fibrils and fibrils. The aggregation kinetics of these assemblies is well known and can be followed by conventional light scattering (Nag *et al.* 2011) and fluorescence lifetime correlation spectroscopy (Paredes *et al.* 2012) measurements.

We can obtain quantitative information by comparing the relative intensities of the various bands inside each spectrum. Since all signals assigned to α-helix, β-strand and random coil are well resolved, in Table 2 we report the results of such an analysis in the case of the β-sheet to random coil intensity ratio, considered at different times after mixing hIAPP and silver NPs. The time evolution of the amide III band confirms that at low incubation time (0-15 min.) the peptide exists as a monomeric species in agreement with Dynamic Light Scattering measurements (Nag *et al.* 2011) performed at the same monomer concentration. After 40 minutes of incubation, the sharp amide III band broadens and the three peaks assigned to helix, sheet and random coil are rather convoluted (Wang *et al.* 2013). Lastly, at very long incubation times (> 360 min) the formation of proto-fibrils is observed (Fändrich 2012).

Our SERS results unambiguously confirm that the distribution of the peptide conformational populations becomes *broader* upon monomer to dimer self-aggregation (Table 2), while retaining a comparable heterogeneity upon increasing size of the aggregates. Although SERS data cannot straightforwardly prove the order-disorder alternation in the aggregates, they are consistent because:

a) significant changes in conformational populations are observed on passing from monomers to ~~a~~ dimers and, spectroscopically, larger oligomers behave as a collection of ordered-disordered dimers as predicted by the theory and demonstrated by MD;

b) our MD simulations cover microsecond range, SERS data run over minutes. Therefore, the simulated alternate order-disordered arrangement is not a transient state but it persists over much longer times than those typical of MD simulations (10^{-6} s).

The data in Table 2 highlight an overall agreement between experiments and simulations.

Table 2. Secondary structures evaluated as the ratio (β-sheet)/(random coil) from SERS measurements recorded at different times *(minutes)*. At each monomer concentration aggregates modify their own size as time goes by, as reported in the literature (Sumner Makin and Serpell 2004; Nguyen *et al.* 2016). In parenthesis we report the (β-sheet)/(random coil) ratio taken from MD data obtained for the different aggregates. Simulation and SERS kinetics have different time scales. Here, molecular dynamics are used to calculate the secondary structure of each aggregation state. The error over SERS measurements was calculated by considering I and III amide band of three independent spectra. [a)] Average of (β-sheet)/(random coil) ratio calculated by MD over monomer (0.48) and dimer (0.75). [b)] Average of (β-sheet)/(random coil) ratio calculated over dimers (0.75), tetramers (0.80) and hexamers (0.73).

Time (minutes)	β-sheet/coil ratio from SERS spectra	Aggregation Status and (β-sheet)/(coil) ratio from MD data
0	0.5±0.2	Monomers (0.48±0.10)
15	1.0±0.3	Mostly Dimers (0.85±0.26[a])
40	1.1±0.2	Small Oligomers (0.76±0.46[b])
120	1.0±0.3	Large Oligomers
240	1.2±0.3	Proto-fibrils + large oligomers
360	2.0±0.3	Proto-fibrils + large Oligomers

Our main results are a theoretical model predicting an instability that leads to periodic distribution of partially ordered and disordered peptide structures in aggregates, and the confirmation of the theoretical predictions using MD simulations and SERS. The main findings from the MD simulations are summarized in Figs. 6 and 7 which show the percentages of the secondary structures in a hIAPP hexamer, and the partially ordered (α-helix and β-strand) and disordered (turn and coil) structures in all hIAPP systems, respectively. More details on conformational arrangements are given in Fig. S14 for ovine prion and in Fig. S15 for equimolar mixtures of hIAPP and Aβ(1-40).

High sensitivity SERS spectroscopy was used to provide information about the β-sheet to random coil ratio in hIAPP aggregates upon oligomerization. The results are consistent with the appearance of modulated structures. Since the theory is based on the hypothesis of conformational symmetry breaking, change in the ratio between ordered and disordered structures observed upon passing from monomer to aggregates is a clear indication of the conformational changes occurring upon aggregation. Moreover, as predicted by the theory and MD data, this ratio remains constant upon increasing the size of the oligomeric structures.

On a larger scale (microns), alternating patterns have also been recently observed in Tip-Enhanced Raman Spectroscopy (TERS) measurements on hIAPP fibrils where an inhomogeneous configuration of secondary structures was found along the fibril surface. In addition, a recent ion mobility mass spectrometry study reported the coexistence of compact and extended structures in an ensemble of low-order Aβ peptide oligomers. Particularly interesting is a recent study using advanced scanning microscopy (vandenAkker *et al.* 2015) and vibrational sum-frequency generation spectroscopy on Aβ fibrils (vandenAkker *et al.* 2011). Other observations have been reported in Refs. (Gorman and Chakrabartty 2001; Klinger *et al.* 2014; vandenAkker *et al.* 2015). These works provide further support to our hypothesis.

Conclusion

The simple theoretical model we have developed assumes a coupling between internal arrangement and intermolecular interactions: for a given intermolecular distance, the peptide (protein) internal conformations may change in order to minimize the total free energy. In turn, the new conformations influence the strengths of the interactions among self-assembled peptides (proteins). When the energy difference between ordered and disordered structures become similar and the interactions are appreciable, spatially modulated patterns may spontaneously emerge. This kind of phenomena are widespread in condensed matter and are often referred to as "pre-transitional" effects. They are observed, for instance, in lipid bilayers, where a strongly corrugated phase (the Ripple phase) appears on approaching the gel to liquid-crystalline phase transition (Akabori and Nagle 2015). According to our model, IDPs are the best candidates for observing an alternate distribution of conformations as confirmed by accurate MD simulations on different self-assemblies. Even non-IDP systems such as wt and apo SOD1 and CRP-CAMP form dimers with a distinct ordered-disordered distribution. Although simulations were performed in different laboratories by using different force fields and water descriptors, data converge toward the concept of symmetry breaking.

The results also qualitatively explain why, in general, amyloid aggregates escape a detailed structural characterization by X-ray analysis (Sumner Makin and Serpell 2004). Provided the partially ordered-disordered arrangement persists also in larger aggregates and at longer times, one may infer that the diffraction patterns of an alternate crystalline-amorphous array of scattering units yields well-resolved Bragg peaks to which a broad amorphous *halo* is superimposed. A thorough discussion of X-ray diffraction for such a complex system is beyond the aims of the present paper. However, the particular peptide arrangement observed for small hIAPP assemblies might contribute to explaining

the coarse resolution and the complex diffraction patterns of mature hIAPP fibrils (Masman et al. 2009).

To the best of our knowledge, this is the first systematic study focusing on the spontaneous periodic distortion in supramolecular peptide and protein aggregates, the possibility of heterogeneous conformational distribution along an aggregate has been previously suggested in literature (Gorman and Chakrabartty 2001; Sitkiewicz *et al.* 2014; Klinger *et al.* 2014; vandenAkker *et al.* 2015) and scattered observations have been reported in MD studies (Reddy *et al*. 2010; Dupuis *et al.* 2011; Barz *et al.* 2014; Huy *et al.* 2016; Zhang *et al.* 2016; Nguyen *et al.* 2016; Das *et al.* 2017). We proved by MD simulations that the order-disorder alternation in IDPs oligomers is a dynamic process that emerges only after statistical averaging. Such a result is consistent with the low energies required for the onset of the alternation in respect to the homogeneous state. We propose that the occurrence of small patterns of alternate ordered-disordered patterns may provide a novel rationale to explain the molecular origin of fibril polymorphism as well as the lack of short-range molecular order in mature fibrils.

Acknowledgment Discussions with Profs. E. Del Favero and L. Cantù (University of Milan, Italy) are gratefully acknowledged. MK would like to thank Natural Sciences and Engineering Council (NSERC) of Canada for financial support. Computational resources were providing by Compute Canada. DM thanks the Italian MIUR program PRIN 20157WZM8A for financial support. AR, GC and CLR have been financed by the Progetto di Dipartimento 2017-2020. FC thanks the Swiss National Science Foundation for funding. Raman measurements were performed at Bio-Nanotech Research and Innovation Tower (BRIT), University of Catania, Italy.

References

Abraham MJ, Murtola T, Schulz R, et al (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1–2:19–25. https://doi.org/10.1016/j.softx.2015.06.001

Akabori K, Nagle JF (2015) Structure of the DMPC lipid bilayer ripple phase. Soft Matter 11:918–926. https://doi.org/10.1039/C4SM02335H

Banci L, Bertini I, Blaževitš O, et al (2012) Interaction of Cisplatin with Human Superoxide Dismutase. J Am Chem Soc 134:7009–7014. https://doi.org/10.1021/ja211591n

Barz B, O. Olubiyi O, Strodel B (2014) Early amyloid β-protein aggregation precedes conformational change. Chemical Communications 50:5373–5375. https://doi.org/10.1039/C3CC48704K

Berendsen HJC, Postma JPM, Gunsteren WF van, Hermans J (1981) Interaction Models for Water in Relation to Protein Hydration. In: Intermolecular Forces. Springer, Dordrecht, pp 331–342

Berendsen HJC, Postma JPM, van Gunsteren WF, et al (1984) Molecular dynamics with coupling to an external bath. The Journal of Chemical Physics 81:3684–3690. https://doi.org/10.1063/1.448118

Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. J Chem Phys 126:014101. https://doi.org/10.1063/1.2408420

Chaikin PM, Lubensky TC (1995) Principles of Condensed Matter Physics. Cambridge University Press

Chakroun N, Prigent S, Dreiss CA, et al (2010) The oligomerization properties of prion protein are restricted to the H2H3 domain. FASEB J 24:3222–3231. https://doi.org/10.1096/fj.09-153924

Cheng X, Gonzalez ML, Lee JC (1993) Energetics of intersubunit and intrasubunit interactions of Escherichia coli adenosine cyclic 3',5'-phosphate receptor protein. Biochemistry 32:8130–8139. https://doi.org/10.1021/bi00083a011

Chiti F, Dobson CM (2006) Protein misfolding, functional amyloid, and human disease. Annu Rev Biochem 75:333–366. https://doi.org/10.1146/annurev.biochem.75.101304.123901

Compagnini G, Messina E, Puglisi O, Nicolosi V (2007) Laser synthesis of Au/Ag colloidal nano-alloys: Optical properties, structure and composition. Applied Surface Science 254:1007–1011. https://doi.org/10.1016/j.apsusc.2007.07.177

Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. The Journal of Chemical Physics 98:10089–10092. https://doi.org/10.1063/1.464397

Das P, Chacko AR, Belfort G (2017) Alzheimer's Protective Cross-Interaction between Wild-Type and A2T Variants Alters Aβ42 Dimer Structure. ACS Chem Neurosci 8:606–618. https://doi.org/10.1021/acschemneuro.6b00357

Daura X, Gademann K, Jaun B, et al (1999) Peptide Folding: When Simulation Meets Experiment. Angewandte Chemie International Edition 38:236–240. https://doi.org/10.1002/(SICI)1521-3773(19990115)38:1/2<236::AID-ANIE236>3.0.CO;2-M

Del Favero E, Raudino A, Brocca P, et al (2009) Lamellar stacking split by in-membrane clustering of bulky glycolipids. Langmuir 25:4190–4197

Deng HX, Hentati A, Tainer JA, et al (1993) Amyotrophic lateral sclerosis and structural defects in Cu,Zn superoxide dismutase. Science 261:1047–1051

Drazin PG, Johnson RS (1989) Solitons: An Introduction, 2 edition. Cambridge University Press, Cambridge England ; New York

Dupuis NF, Wu C, Shea J-E, Bowers MT (2011) The Amyloid Formation Mechanism in Human IAPP: Dimers have β-strand Monomer-Monomer Interfaces. J Am Chem Soc 133:7240–7243. https://doi.org/10.1021/ja1081537

D'Urso L, Condorelli M, Puglisi O, et al (2018) Detection and characterization at nM concentration of oligomers formed by hIAPP, Aβ(1–40) and their equimolar mixture using SERS and MD simulations. Phys Chem Chem Phys 20:20588–20596. https://doi.org/10.1039/C7CP08552D

Evans DJ, Holian BL (1985) The Nose–Hoover thermostat. J Chem Phys 83:4069–4074. https://doi.org/10.1063/1.449071

Fändrich M (2012) Oligomeric intermediates in amyloid formation: structure determination and mechanisms of toxicity. J Mol Biol 421:427–440. https://doi.org/10.1016/j.jmb.2012.01.006

Fazio E, Cacciola A, Mezzasalma AM, et al (2013) Modelling of the optical absorption spectra of PLAL prepared ZnO colloids. Journal of Quantitative Spectroscopy and Radiative Transfer 124:86–93. https://doi.org/10.1016/j.jqsrt.2013.02.028

Frishman D, Argos P (1995) Knowledge-based protein secondary structure assignment. Proteins 23:566–579. https://doi.org/10.1002/prot.340230412

Fuß W, Haas Y, Zilberg S (2000) Twin states and conical intersections in linear polyenes. Chemical Physics 259:273–295. https://doi.org/10.1016/S0301-0104(00)00200-7

Goldstein H, Jr CPP, Safko JL (2001) Classical Mechanics, 3 edition. Pearson, San Francisco, NJ

Gorman PM, Chakrabartty A (2001) Alzheimer beta-amyloid peptides: structures of amyloid fibrils and alternate aggregation products. Biopolymers 60:381–394. https://doi.org/10.1002/1097-0282(2001)60:5<381::AID-BIP10173>3.0.CO;2-U

Grason GM, Santangelo CD Undulated cylinders of charged diblock copolymers. Eur Phys J E 20:335–346. https://doi.org/10.1140/epje/i2006-10025-1

Grest GS, Kremer K (1986) Molecular dynamics simulation for polymers in the presence of a heat bath. Phys Rev A 33:3628–3631. https://doi.org/10.1103/PhysRevA.33.3628

Gsponer J, Vendruscolo M (2006) Theoretical approaches to protein aggregation. Protein Pept Lett 13:287–293

Haire LF, Whyte SM, Vasisht N, et al (2004) The crystal structure of the globular domain of sheep prion protein. J Mol Biol 336:1175–1183. https://doi.org/10.1016/j.jmb.2003.12.059

Harries D, Podgornik R, Parsegian VA, et al (2006) Ion induced lamellar-lamellar phase transition in charged surfactant systems. J Chem Phys 124:224702. https://doi.org/10.1063/1.2198534

Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: A linear constraint solver for molecular simulations. Journal of Computational Chemistry 18:1463–1472. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H

Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J Chem Theory Comput 4:435–447. https://doi.org/10.1021/ct700301q

Höppener JWM, Ahrén B, Lips CJM (2000) Islet Amyloid and Type 2 Diabetes Mellitus. New England Journal of Medicine 343:411–419. https://doi.org/10.1056/NEJM200008103430607

Huang J, Rauscher S, Nawrocki G, et al (2017) CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. Nat Methods 14:71–73. https://doi.org/10.1038/nmeth.4067

Huy PDQ, Vuong QV, La Penna G, et al (2016) Impact of Cu(II) Binding on Structures and Dynamics of Aβ42 Monomer and Dimer: Molecular Dynamics Study. ACS Chem Neurosci 7:1348–1363. https://doi.org/10.1021/acschemneuro.6b00109

Iljina M, Garcia GA, Dear AJ, et al (2016) Quantitative analysis of co-oligomer formation by amyloid-beta peptide isoforms. Scientific Reports 6:. https://doi.org/10.1038/srep28658

Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577–2637. https://doi.org/10.1002/bip.360221211

Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. J Phys Chem B 105:6474–6487. https://doi.org/10.1021/jp003919d

Khare SD, Caplow M, Dokholyan NV (2004) The rate and equilibrium constants for a multistep reaction sequence for the aggregation of superoxide dismutase in amyotrophic lateral sclerosis. PNAS 101:15094–15099. https://doi.org/10.1073/pnas.0406650101

Klinger AL, Kiselar J, Ilchenko S, et al (2014) A synchrotron-based hydroxyl radical footprinting analysis of amyloid fibrils and prefibrillar intermediates with residue-specific resolution. Biochemistry 53:7724–7734. https://doi.org/10.1021/bi5010409

Kurouski D, Duyne RPV, Lednev IK (2015) Exploring the structure and formation mechanism of amyloid fibrils by Raman spectroscopy: a review. Analyst 140:4967–4980. https://doi.org/10.1039/C5AN00342C

La Rosa C, Milardi D, Grasso D, et al (1995) Thermodynamics of the thermal unfolding of azurin. Journal of Physical Chemistry 99:14864–14864

La Rosa C, Milardi D, Grasso DM, et al (2002) A model for the thermal unfolding of amicyanin. European Biophysics Journal 30:559–570

La Rosa C, Scalisi S, Lolicato F, et al (2016) Lipid-assisted protein transport: A diffusion-reaction model supported by kinetic experiments and molecular dynamics simulations. J Chem Phys 144:184901. https://doi.org/10.1063/1.4948323

Le Ru EC, Blackie E, Meyer M, Etchegoin PG (2007) Surface Enhanced Raman Scattering Enhancement Factors: A Comprehensive Study. J Phys Chem C 111:13794–13803. https://doi.org/10.1021/jp0687908

Lee J, Cheng X, Swails JM, et al (2016) CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. J Chem Theory Comput 12:405–413. https://doi.org/10.1021/acs.jctc.5b00935

Lorenzo A, Razzaboni B, Weir GC, Yankner BA (1994) Pancreatic islet cell toxicity of amylin associated with type-2 diabetes mellitus. Nature 368:756–760. https://doi.org/10.1038/368756a0

Masman MF, Eisel ULM, Csizmadia IG, et al (2009) In Silico Study of Full-Length Amyloid β 1−42 Tri- and Penta-Oligomers in Solution. J Phys Chem B 113:11710–11719. https://doi.org/10.1021/jp901057w

Messina E, D'Urso L, Fazio E, et al (2012) Tuning the structural and optical properties of gold/silver nano-alloys prepared by laser ablation in liquids for optical limiting, ultra-sensitive spectroscopy, and optical trapping. Journal of Quantitative Spectroscopy and Radiative Transfer 113:2490–2498. https://doi.org/10.1016/j.jqsrt.2012.06.023

Milardi D, La Rosa C, Grasso D, et al (1998) Thermodynamics and kinetics of the thermal unfolding of plastocyanin. Eur Biophys J 27:273–282. https://doi.org/10.1007/s002490050134

Milardi D, Pappalardo M, Pannuzzo M, et al (2008) The role of the Cys2-Cys7 disulfide bridge in the early steps of Islet amyloid polypeptide aggregation: A molecular dynamics study. Chemical Physics Letters 463:396–399. https://doi.org/10.1016/j.cplett.2008.07.110

Nag S, Sarkar B, Bandyopadhyay A, et al (2011) Nature of the Amyloid-β Monomer and the Monomer-Oligomer Equilibrium. J Biol Chem 286:13827–13833. https://doi.org/10.1074/jbc.M110.199885

Nguyen PH, Sterpone F, Pouplana R, et al (2016) Dimerization Mechanism of Alzheimer Aβ40 Peptides: The High Content of Intrapeptide-Stabilized Conformations in A2V and A2T Heterozygous Dimers Retards Amyloid Fibril Formation. J Phys Chem B 120:12111–12126. https://doi.org/10.1021/acs.jpcb.6b10722

Nguyen TT, Gopal A, Lee KYC, Witten TA (2005) Surface charge relaxation and the pearling instability of charged surfactant tubes. Phys Rev E 72:051930. https://doi.org/10.1103/PhysRevE.72.051930

Nosé S, Klein ML (1983) Constant pressure molecular dynamics for molecular systems. Molecular Physics 50:1055–1076. https://doi.org/10.1080/00268978300102851

Onuki A (2002) Phase Transition Dynamics, Cambridge University Press, 2009. In: Cambridge Core

Oostenbrink C, Villa A, Mark AE, van Gunsteren WF (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. J Comput Chem 25:1656–1676. https://doi.org/10.1002/jcc.20090

Pappalardo M, Milardi D, Grasso D, La Rosa C (2007) Steered molecular dynamics studies reveal different unfolding pathways of prions from mammalian and non-mammalian species. New J Chem 31:901–905. https://doi.org/10.1039/B700764G

Pappalardo M, Sciacca M, Milardi D, et al (2008) Thermodynamics of azurin folding: The role of copper ion. Journal of Thermal Analysis and Calorimetry 93:575–581

Paredes JM, Casares S, Ruedas-Rama MJ, et al (2012) Early Amyloidogenic Oligomerization Studied through Fluorescence Lifetime Correlation Spectroscopy. Int J Mol Sci 13:9400–9418. https://doi.org/10.3390/ijms13089400

Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. Journal of Applied Physics 52:7182–7190. https://doi.org/10.1063/1.328693

Patil SM, Xu S, Sheftic SR, Alexandrescu AT (2009) Dynamic α-Helix Structure of Micelle-bound Human Amylin. J Biol Chem 284:11982–11991. https://doi.org/10.1074/jbc.M809085200

Peierls RE (2001) Quantum Theory of Solids - Oxford University Press, USA, 2001, Chapter 5.3

Porcar L, Marignan J, Ligoure C, Gulik-Krzywicki T (2000) Effect of a Nonadsorbing Polymer on the Stability of a Two-Solvent Lamellar Phase: Experimental and Theoretical Study of Critical Points of Lamellar/Lamellar Phase Separations. Langmuir 16:2581–2594. https://doi.org/10.1021/la991193a

Qiu D, Shenkin PS, Hollinger FP, Still WC (1997) The GB/SA Continuum Model for Solvation. A Fast Analytical Method for the Calculation of Approximate Born Radii. J Phys Chem A 101:3005–3014. https://doi.org/10.1021/jp961992r

Reddy AS, Wang L, Singh S, et al (2010) Stable and metastable states of human amylin in solution. Biophys J 99:2208–2216. https://doi.org/10.1016/j.bpj.2010.07.014

Riek R, Eisenberg DS (2016) The activities of amyloids from a structural perspective. Nature 539:227–235. https://doi.org/10.1038/nature20416

Risken H (1984) The Fokker-Planck Equation: Methods of Solution and Applications. Springer-Verlag, Berlin Heidelberg

Rygula A, Majzner K, Marzec KM, et al (2013) Raman spectroscopy of proteins: a review. J Raman Spectrosc 44:1061–1076. https://doi.org/10.1002/jrs.4335

Savelieff MG, Nam G, Kang J, et al (2019) Development of Multifunctional Molecules as Potential Therapeutic Candidates for Alzheimer's Disease, Parkinson's Disease, and Amyotrophic Lateral Sclerosis in the Last Decade. Chem Rev 119:1221–1322. https://doi.org/10.1021/acs.chemrev.8b00138

Schmid N, Eichenberger AP, Choutko A, et al (2011) Definition and testing of the GROMOS force-field versions 54A7 and 54B7. Eur Biophys J 40:843–856. https://doi.org/10.1007/s00249-011-0700-9

Sciacca MFM, Tempra C, Scollo F, et al (2018) Amyloid growth and membrane damage: Current themes and emerging perspectives from theory and experiments on Aβ and hIAPP. Biochim Biophys Acta Biomembr. https://doi.org/10.1016/j.bbamem.2018.02.022

Sciacca MicheleFM, Romanucci V, Zarrelli A, et al (2017) Inhibition of Aβ Amyloid Growth and Toxicity by Silybins: The Crucial Role of Stereochemistry. ACS Chem Neurosci 8:1767–1778. https://doi.org/10.1021/acschemneuro.7b00110

Scollo F, Tempra C, Lolicato F, et al (2018) Phospholipids Critical Micellar Concentrations Trigger Different Mechanisms of Intrinsically Disordered Proteins Interaction with Model Membranes. J Phys Chem Lett 9:5125–5129. https://doi.org/10.1021/acs.jpclett.8b02241

Sitkiewicz E, Kłoniecki M, Poznański J, et al (2014) Factors influencing compact-extended structure equilibrium in oligomers of aβ1-40 peptide--an ion mobility mass spectrometry study. J Mol Biol 426:2871–2885. https://doi.org/10.1016/j.jmb.2014.05.015

Smith MD, Cruz L (2013) Effect of Ionic Aqueous Environments on the Structure and Dynamics of the Aβ21–30 Fragment: A Molecular-Dynamics Study. J Phys Chem B 117:6614–6624. https://doi.org/10.1021/jp312653h

Stathopulos PB, Rumfeldt JAO, Karbassi F, et al (2006) Calorimetric Analysis of Thermodynamic Stability and Aggregation for Apo and Holo Amyotrophic Lateral Sclerosis-associated Gly-93 Mutants of Superoxide Dismutase. Journal of Biological Chemistry 281:6184–6193. https://doi.org/10.1074/jbc.M509496200

Sumner Makin O, Serpell LC (2004) Structural Characterisation of Islet Amyloid Polypeptide Fibrils. Journal of Molecular Biology 335:1279–1288. https://doi.org/10.1016/j.jmb.2003.11.048

Svensson A-KE, Bilsel O, Kayatekin C, et al (2010) Metal-free ALS variants of dimeric human Cu,Zn-superoxide dismutase have enhanced populations of monomeric species. PLoS ONE 5:e10064. https://doi.org/10.1371/journal.pone.0010064

Svensson A-KE, Bilsel O, Kondrashkina E, et al (2006) Mapping the folding free energy surface for metal-free human Cu,Zn superoxide dismutase. J Mol Biol 364:1084–1102. https://doi.org/10.1016/j.jmb.2006.09.005

Tamashiro MN, Schiessel H (2006) Rayleigh instability of charged aggregates: Role of the dimensionality, ionic strength, and dielectric contrast. Phys Rev E 74:021412. https://doi.org/10.1103/PhysRevE.74.021412

Tarus B, Tran TT, Nasica-Labouze J, et al (2015) Structures of the Alzheimer's Wild-Type Aβ1-40 Dimer from Atomistic Simulations. J Phys Chem B 119:10478–10487. https://doi.org/10.1021/acs.jpcb.5b05593

vandenAkker CC, Deckert-Gaudig T, Schleeger M, et al (2015) Nanoscale Heterogeneity of the Molecular Structure of Individual hIAPP Amyloid Fibrils Revealed with Tip-Enhanced Raman Spectroscopy. Small 11:4131–4139. https://doi.org/10.1002/smll.201500562

vandenAkker CC, Engel MFM, Velikov KP, et al (2011) Morphology and Persistence Length of Amyloid Fibrils Are Correlated to Peptide Molecular Structure. J Am Chem Soc 133:18030–18033. https://doi.org/10.1021/ja206513r

Wang Q, Wang Y, Lu HP (2013) Revealing the secondary structural changes of amyloid β peptide by probing the spectral fingerprint characters. J Raman Spectrosc 44:670–674. https://doi.org/10.1002/jrs.4253

Weber IT, Steitz TA (1987) Structure of a complex of catabolite gene activator protein and cyclic AMP refined at 2.5 A resolution. J Mol Biol 198:311–326

Yannoni CS, Clarke TC (1983) Molecular Geometry of cis- and trans-Polyacetylene by Nutation NMR Spectroscopy. Phys Rev Lett 51:1191–1193. https://doi.org/10.1103/PhysRevLett.51.1191

Zhang Y, Hashemi M, Lv Z, Lyubchenko YL (2016) Self-assembly of the full-length amyloid Aβ42 protein in dimers. Nanoscale 8:18928–18937. https://doi.org/10.1039/C6NR06850B

Manuscript (Clean)

Symmetry-Breaking Transitions in the Early Steps of Proteins self-assembly.

Carmelo La Rosa[a], Marcello Condorelli[a], Giuseppe Compagnini[a], Fabio Lolicato[a,b,c], Danilo Milardi[d], Trang Do[e], Mikko Karttunen[f], Martina Pannuzzo[g], Ayyalusamy Ramamoorthy[h], Franca Fraternali[i], Francesca Collu[i], Human Rezaei[l], Birgit Strodel[k]and Antonio Raudino[a,]*.

[a]Department of Chemical Sciences, University of Catania, Viale A. Doria 6, I-95125, Catania, Italy.

[b]Department of Physics, University of Helsinki, P.O. Box 64, FI-00014, Helsinki, Finland.

[c]Heidelberg University Biochemistry Center, Heidelberg, Germany

[d]CNR, Istituto di Biostrutture e Bioimmagini, Unità Organizzativa e di Supporto di Catania, Via P. Gaifami 18, I-95126, Catania, Italy.

[e]Department of Chemistry & Waterloo Institute for Nanotechnology, University of Waterloo, 200 University Avenue West, Waterloo, Ontario, Canada N2L 3G1.

[f]Department of Chemistry and Department of Applied Mathematics, Western University, 1151 Richmond Street, London, Ontario N6A 5B7, Canada

[g]Istituto Italiano di Tecnologia (IIT), Via Morego 30, 16163 Genoa, Italy

[h]Department of Chemistry and Department of Biophysics, University of Michigan, Ann Arbor, Michigan 48109-1055, USA.

[i]Randall Division of Cell and Molecular Biophysics, New Hunt's House King's College London, Guy's Campus SE1 1UL London UK.

[l]Virologie et Immunologie Moléculaires, Institut National de la Recherche Agronomique, F-78352 Jouy-en-Josas, France.

[k]Institute of Complex Systems, Structural Biochemistry (ICS-6), ForschungszentrumJülich, 52425 Jülich, Germany.

Abstract

Protein misfolding and assembly are complex, intertwined processes resulting in the development of a heterogeneous population of aggregates closely related to many chronic pathological conditions including Type 2 Diabetes Mellitus and Alzheimer's Disease. To address this issue, here we develop a theoretical model in the general framework of linear stability analysis. According to this model, self-assemblies of peptides with pronounced conformational flexibility may become, under particular conditions, unstable and spontaneously evolve toward an alternating array of partially ordered and disordered monomers. The predictions of the theory were verified by atomistic Molecular Dynamics (MD) simulations of Islet Amyloid Polypeptide (IAPP) used as a paradigm of aggregation-prone polypeptides (proteins). Simulations of dimeric, tetrameric and hexameric human-IAPP self-assemblies at physiological electrolyte concentration reveal an alternating distribution of the smallest domains (of the order of the peptide mean length) formed by partially ordered (mainly β-strands) and disordered (turns and coil) arrays. Periodicity disappears upon weakening of the inter-peptide binding, a result in line with the predictions of the theory. To further probe the general validity of our hypothesis, we extended the simulations to other peptides, the Aβ(1-40) amyloid peptide, and the ovine prion peptide as well as to other proteins (SOD1 dimer) that do not belong to the broad class of intrinsically disordered proteins. In all cases, the oligomeric aggregates show an alternate distribution of partially ordered and disordered monomers. We also carried out Surface Enhanced Raman Scattering (SERS) measurements of hIAPP as an experimental validation of both the theory and in silico simulations.

Keywords: Intrinsically Disordered Proteins, Symmetry-Breaking, Molecular Dynamics, Analytical Model, Oligomers.

Introduction

In spontaneous symmetry-breaking, a system possessing a symmetry at its high temperature phase loses this symmetry upon cooling in the absence of external fields below the phase transition temperature. Ferromagnetism is probably the most well-known example: net magnetization arises as the local spins become ordered upon cooling below the Curie temperature (Chaikin and Lubensky 1995). Spontaneous symmetry breaking is important in a plethora of phenomena such as single-double bond alternation in Chemistry, molecular chirality, the formation of liquid-crystalline phases, charge and spin-density waves, superconductivity, the onset of polarity in cells and so on (Grason and Santangelo 2006; Yannoni and Clarke 1983; Chaikin and Lubensky 1995; Fuß *et al.* 2000; Gorman and Chakrabartty 2001; Peierls 2001; Nguyen *et al.* 2005, 2016; Tamashiro and Schiessel 2006; Sitkiewicz *et al.* 2014; Klinger *et al.* 2014; vandenAkker *et al*. 2015).

In this work, we try to extend the above concepts of spontaneous symmetry breaking to solid-state peptide (protein) crystals and to oligomeric self-aggregate protein in solution. It is well-known that in the crystal lattice the unit cell of most proteins contains just a single protein, while the unit cell of some protein systems may contain two or more *unequal* proteins (Weber and Steitz 1987; Deng *et al.* 1993) (see, as an example, the SOD1 structure reported in fig S1 of Supplementary Information (SI)).

This behavior is often observed also in the liquid phase, where proteins can stick together forming oligomers by spontaneous self-assembly. In this case, aggregates may arrange themselves as an array of identical monomeric units or, alternatively, they may exhibit periodic patters. For instance, dimeric proteins such as wt (Deng *et al.* 1993)and apo (Banci et al. 2012) superoxide dismutase (SOD1) and CRP-CAMP (Weber and Steitz 1987) exert their biological activity as *asymmetric* dimers.

We would like to mention that the possibility of heterogeneous conformational distribution along a protein aggregate has been previously suggested in literature (Gorman and Chakrabartty 2001; Sitkiewicz *et al*. 2014; Klinger *et al.* 2014; vandenAkker *et al.* 2015) and scattered observations have been reported in MD studies of Intrinsically Disordered Proteins (IDPs) (Reddy *et al.* 2010; Dupuis *et al*. 2011; Barz *et al.* 2014; Huy *et al.* 2016; Zhang *et al.* 2016; Nguyen *et al.* 2016; Das *et al.* 2017).

In this paper we address this issue by building up a simplified theoretical model (both in 1-D and 3-D) where protein aggregates are represented as an array of two different conformers: a spheroidal and quite disordered conformer alternating with a more ordered and elongated structure. Notably, the elongated shape enables the unit to be packed more tightly than the spheroidal one. The theoretical model developed here predicts that, under particular conditions, the system becomes unstable and

produce an alternating (patterned) array of ordered and disordered conformers when: a) their internal energy difference is small; b) the monomer-monomer interaction is large; c) there is a change in the geometry on going from ordered to disordered conformers.

The best candidates for observing the foreseen effects is the broad family IDPs because they fulfill all the above requirements. Furthermore, IDPs misfolding and aggregation into toxic oligomers or amyloid fibrils has attracted increasing interest in the last decades since it is associated with a large number of diseases (such as Alzheimer's disease (AD) and type 2 diabetes (T2D)). Particular attention was paid to oligomers to define the molecular determinants that may steer amyloid vs amorphous aggregation(La Rosa *et al*. 2016; Scollo *et al.* 2018): in fact, poorly structured, small sized oligomers may be more toxic than large, mature fibrils (Chiti and Dobson 2006; Riek and Eisenberg 2016; Sciacca *et al*. 2018). This is being proven by the development of new toxicity inhibitors that interact with oligomers instead of the already studied beta-breakers(Sciacca et al. 2017; Savelieff *et al*. 2019). A hallmark of AD is the aggregation of amyloid-β (Lorenzo *et al.* 1994) (Aβ); and human amylin (Höppener et al. 2000) (IAPP), a 37-residue polypeptide, is the major constituent of the pancreatic amyloid deposits found in patients with T2D. Both Aβ and IAPP belong to the IDPs family.

To verify the general validity of the model, we performed Molecular Dynamics (MD) simulations of di-, tetra- and hexameric assemblies of hIAPP in pure water and in a 0.1 M electrolyte solution, a condition in which the intermolecular repulsion is weaker. Accurate experimental tests of our model are not currently available since present instrumentation do not support nanometer and millisecond resolution. However, Surface Enhanced Raman Scattering (SERS) measurements are able to detect the average evolution of helix, sheet and random coil secondary structures as a function of time. These measurements confirm the heterogeneous nature of hIAPP aggregates. We extended MD simulations to other IDPs (e.g. Aβ peptide) and amyloidogenic globular proteins (e.g., ovine prion peptide) to probe the general validity of the proposed model. Finally, we performed accurate (in the microsecond scale) MD simulations of Super Oxide Dismutase dimer, a well-known protein that does not belong to the IDPs class. Similarly, to IDPs, SOD self-aggregates and exerts its enzymatic activity as an asymmetric tight-bound dimer.

Experimental Methods

hIAPP monomerization. hIAPP with a purity >99% was purchased from Bachem (Bubendorf, Switzerland) and 1,1,1,3,3,3-hexa-fluoro-2-propanol (HFIP) with a purity of 99% w from Sigma-Aldrich. In order to prevent preformed aggregates, hIAPP was initially dissolved in HFIP at a concentration of 1 mg/ml and then lyophilized overnight. The lyophilized powder was dissolved in dimethyl sulfoxide (DMSO) to obtain a stock solution with a final concentration of 100 μM. Each stock solution of hIAPP was used immediately after preparation.

Metal colloids and SERS. Colloidal metal nanoparticles (NPs) were prepared by pulsed laser ablation in liquid using the second harmonic (532 nm) output of a Nd:YAG laser operating at 10 Hz and at a fluence of 2 J/cm^2. A metal target (Ag) submerged in 5 ml of Millipore water was used as a substrate for ablation(Compagnini *et al.* 2007). The ablation process lasted 10 minutes. The concentration of the colloidal dispersion was estimated to be at around $1 \cdot 10^{-5}$M, assuming an average molar extinction coefficient (Messina *et al.* 2012; Fazio *et al.* 2013) of $2.5 \cdot 10^4$ M^{-1}cm^{-1}.The size (20 nm) of metal NPs was estimated to be 35 nm by AFM. SERS analysis was performed with 532 nm laser irradiation in the backscattering mode using a Witec Alpha 300 RS instrument. All the spectra were acquired in water containing 10^{-3} M of NaCl (high salt concentration, NPs self-aggregate into large size assemblies so decreasing the SERS effect) with a concentration of 0.25 μM of hIAPP.

In SERS, the intensity of the Raman spectrum is greatly enhanced by metal (typically Ag or Au) particles or nanostructures. This can increase the intensity of the Raman spectrum up to a factor of 10^{10} - 10^{11} (see, e.g., Le Ru, *et al.* 2007)). Proteins' secondary structures can be determined by SERS by following the characteristic bands associated with the CONH (amide) group. The so-called amide I and III are the most sensitive bands. Typically, an α-helix shows a band in the range of 1640-1658 cm^{-1} (amide I) and 1264-1272 cm^{-1} (amide III). β-sheet bands are in the range 1665-1680 cm^{-1} (amide I) and 1227-1242 cm^{-1} (amide III) while random coil lies in the range 1660-1665 cm^{-1} (amide I) and 1230-1240 cm^{-1} (amide III) (Rygula *et al.* 2013). These wavenumbers are red shifted by about 20-30 cm^{-1} in a SERS measurement[4]. The amide I band is the most sensitive to secondary structure (Kurouski *et al.* 2015). The amide III band gives information about the oligomerization state of amyloidogenic proteins. For instance, Aβ(1-42) oligomers show a red shift of 27 cm^{-1} with respect to monomers (Wang *et al.* 2013).

Simulation methods and setup

Amyloidogenic proteins simulations were carried out using Gromacs software (Hess *et al*. 2008). In all of the cases, the systems were first energy minimized and pre-equilibrated by using the steepest descents algorithm and successive 1 ns simulations in the NVT and NPT ensembles. With the exception of the 20-mer Aβ (1-40) system, all systems were simulated in explicit water. The details of the implicit solvent simulations are given in Sec. 3.2.3

The common details for all explicit solvent simulations are the following: the GROMOS 54A7 force field (Schmid *et al.* 2011) was used for the hIAPP and Aβ systems, and the GROMOS 53A6 (Oostenbrink *et al.* 2004) for the ovine prion system. SPC (simple point charge) water model (Berendsen *et al.* 1981) was employed. The time step was set to 2 fs and the temperature was kept constant at 300 K using the V-rescale algorithm (Bussi *et al.* 2007) with the time constant set to 0.1 ps. Periodic boundary conditions were applied and the Parrinello-Rahman algorithm (Parrinello and Rahman 1981) was applied for isotropic pressure coupling (1 bar) with the exception of the prion simulation that used the Berendsen algorithms (Berendsen *et al.* 1984*)*. The particle-mesh Ewald (PME) algorithm (Darden *et al.* 1993) was used for electrostatic interactions and overall charge neutrality was preserved by adding counterions when necessary. A solution of 0.1 M NaCl was used in all explicit solvent systems. All simulations were repeated at least three times. For secondary structure analysis DSSP (Kabsch and Sander 1983) and STRIDE (Frishman and Argos 1995) were used. To monitor equilibration, RMSD (the root mean square deviation) was used. The rest of the details of each of the systems are provided below.

Simulation details of the hIAPP systems. The NMR structure of hIAPP bound to sodium dodecylsufate (SDS) micelles (pdb ID: 2KB8 (Patil *et al.* 2009); the first structure in the PDB file) was used as the initial structure for the hIAPP monomer. Monomer, dimer, tetramer and hexamer systems were simulated. Details of the individual systems are listed next.

hIAPP monomers. The hIAPP monomer was placed in a cubic solvent box of linear length 7.7 nm. Each of the three independent simulations lasted 400 ns.

hIAPP dimers. Final independent configurations from the hIAPP monomer simulations were used to construct a dimer. Initially, the monomers were placed 4.7 nm from each other. The dimer was then simulated for 500 ns (for each independent simulation run) at 300 K. The simulation box was of the same size as above.

hIAPP tetramers. The last frames from independent dimer simulations were used to construct a tetramer. A larger simulation box (16 x 10 x 10 nm^3) was used. Initially, the two dimers were placed 5 nm away from each other to eliminate unwanted self-interactions. Each of the three independent

simulations was run for a total of 1050 ns using a heating cycle to eliminate possible kinetic traps as follows: 1) 300 ns at 300 K. 2) Followed by 20 ns at 400 K, and 3) finally 730 ns at 300 K.

hIAPP hexamers. Final configurations from tetramer simulations and two monomers from the monomer simulations were used to set up the three independent simulations. Initially, the components were placed about 4 nm away from each other. The same simulation box size as for the tetramer systems was used. The hexamer systems were simulated for a total of 1020 ns each run using a heating cycle as follows: 1) 300 ns at 300 K, 2) 20 ns at 400 K, 3) 250 ns at 300 K, 4) 50 ns at 400 K, 5) 100 ns at 300 K, 6) 200 ns at 400 K, and 7) 100 ns at 300 K.

hIAPP hexamers in pure water. Using the final configurations obtained after 1020 ns of simulation from the previous section, salt was removed (the counterions were left to ensure overall charge neutrality). The hexamer in pure water was then simulated for 500 ns at 300 K.

Aβ (1-40) systems. To investigate the generality of the partially ordered-disordered periodic arrays, we extended our investigation to three additional systems.

hIAPP:Aβ (1-40) 1:1. Monomer structures for hIAPP and Aβ (1-40) were taken from above. Each of the three independent systems was simulated for 500 ns at 300 K.

Implicit solvent simulations. In addition to Aβ (1-40) in explicit solvent, 20-mer Aβ (1-40) systems using the implicit solvent generalized Born surface area (GBSA (Qiu *et al.* 1997)) method were simulated. These simulations were performed in the NVT ensemble using the Langevin thermostat (Grest and Kremer 1986). The OPLS/AA force field (Kaminski *et al.* 2001) was used for the peptides and no NaCl was added. As with other systems, three independent simulations were performed each simulation being 500 ns.

Ovine prion systems. Prion (Pappalardo *et al.* 2007) aggregation was simulated by taking 18 molecules of H2H3 from OvPrP (PDB ID 1UW3, residues from C182 to C217) (Haire *et al.* 2004) in a β-rich conformation, obtained from a prior MD simulation (Chakroun *et al.* 2010) in water. The 18 prion monomers were placed in random orientation on a grid in a rectangular simulation box of dimensions 14 nm x 14 nm x 10 nm with 1.4 nm spacing and 63,668 water molecules. Each of the three individual trajectories was 200 ns. In the prion systems, Cys182 and Cys217 were connected with a disulphide bridge (Haire *et al.* 2004).

Superoxide dismutase. For atomistic molecular dynamics simulations of Superoxide dismutase (SOD) we used the SOD dimeric crystallographic structure [PDB id: 1SPD (Deng *et al.* 1993)] the

PDB structure of SOD was solvated CHARMM-GUI (Lee *et al.* 2016). A 150 mM concentration of KCl was added to mimic physiological condition.

The CHARMM36m force field (Huang *et al.* 2017) was employed to describe the protein structure, salt ions and TIP3 water model.

The system was first energy minimized using the steepest descend algorithm followed by an equilibration step under NpT conditions for 10 ns. In this stage, all protein atoms were restrained with a harmonic potential with a force constant of 1000 kJ mol^{-1} nm^{-2}. The temperature was kept constant at 310 K using the Nose-Hoover thermostat (Nosé and Klein 1983; Evans and Holian 1985) with a time constant of 1.0 ps. Isotropic pressure coupling scheme was applied using the Berendsen algorithm (Berendsen *et al.* 1984) with a time constant set to 5.0 ps. The Verlet scheme, with a cut-off distance of 1.2 nm is set to search the short-range neighbours every 20 steps. Particle mesh Ewald method (Darden *et al.* 1993) was used to handle the electrostatic interactions. The cut-off length of 1.2 nm was used for both electrostatic and van der Waals interactions. To constrain the hydrogen bonds, the LINCS algorithm (Hess *et al.* 1997) was employed and periodic boundary conditions were applied in all directions. For the production MD runs, we removed all the restraints applied to the proteins and used the and Parrinello-Rahman barostat (Bussi *et al.* 2007) instead of the Berendsen's algorithm (Berendsen *et al.* 1984). See **Table 1** for simulations details.

All simulations were carried out using an integration time step of 2 fs using the GROMACS 2018 simulation package (Abraham *et al.* 2015).

Table 1. Simulations Details.

# Protein	# Water	# K +	# Cl -	Time (ns)	Temperature (K)
2	49001	153	139	1200	310 – 372*
2	49001	153	139	200	310

* The temperature has been increased linearly from 310 to 372 K during the first 500 ns. The temperature was then kept constant at 372 K until the system reached 1200 ns.

Theoretical Methods

The famous Peierls instability (Peierls 2001) shares significant resemblances (but there are also significant differences) with the systems investigated here. In the essence, the Peierls Theorem

(Peierls 2001) states that a system consisting of a one-dimensional crystal with constant lattice spacing and electrons becomes unstable and develops a periodic modulation of electron density if there is any coupling between the electrons and the lattice. This is due to a competition between the electronic and elastic energies. An application is provided by poly-acetylene (Yannoni and Clarke 1983; Fuß *et al.* 2000) (…-CH=CH-CH=CH-…) and chemically related molecules: Experimental (Yannoni and Clarke 1983) and theoretical (Fuß *et al.* 2000) data confirm the onset of less symmetric structures, where alternating array of electron-rich short (1.36 Å) and electron-poor long (1.44 Å) bonds is preserved. Analogously, shape instabilities arise from the competition between electrostatic repulsion and surface tension: increasing the size of a droplet beyond a critical size may lead to capillary instabilities that eventually breaks larger droplets into smaller ones (Tamashiro and Schiessel 2006). For instance, uniformly charged tubules made up of identical surfactant molecules can become unstable evolving toward less symmetric structures such as undulated cylinders or arrays of juxtaposed spheroidal micelles (pearling instability (Grason and Santangelo 2006; Nguyen *et al.* 2005)). Similar effects were also observed in lyotropic lamellar systems which show alternation in the lamellar repeat distance (Porcar *et al.* 2000; Harries *et al.* 2006; Del Favero *et al.* 2009).

We start by developing an idealized model to investigate if a linear aggregate of interacting peptides (proteins) that exist in different conformational states may exhibit a transition from a *homogeneous* to a space-modulated structure.

Free energy

We first consider infinitely long one-dimensional (1-D) pre-fibrillar aggregates along the *z*-axis. The opposite cases of a dimeric structure and the extension to the 3-D infinite aggregates will be discussed separately. Proteins exhibit conformational flexibility that, for the sake of simplicity, is restricted to two interchanging states alone: Φ and Ψ. The ordered Ψ arrangement comprises α-helices and β-strands, while the disordered Φ conformation includes coils and turns. A description of protein structure and function in terms of pure Φ and Ψ states is a gross simplification. In the real world, proteins can be described by a weighted combination of Φ and Ψ states. Usually, the ordered conformations prevail at room temperature (and in absence of denaturating chemicals), while, in the case of IDPs, the relative abundance of ordered and disordered domains within the same molecule is comparable over a wide range of experimental conditions.

The total free energy of proteins array, G, can be decomposed into three main contributions: a) a term related to the entropy of mixing ($T \cdot S_{MIXING}$) which favors an identical of Φ and Ψ states, b)

the self-energy (U_{SELF}) which measures the stability of a given conformation by specific intramolecular bonds and solvent interactions, and c) the energy of interaction (U_{INT}), which expresses the conformation-dependent interaction between nearest-neighbor proteins. Combining the three contributions gives the total free energy

$$G = -TS_{MIXING} + U_{SELF} + U_{INT}, \qquad (1)$$

Next, we develop functional forms for the individual terms.

Let φ_n be the local fraction of Φ conformations of a protein at site *n* along the 1-D array, and 1-φ_n the fraction of Ψ conformations. Standard mean-field expression for Φ and Ψ mixing entropy S_{MIX} reads

$$-TS_{MIX} \approx k_B T \sum_n^N [\varphi_n \log \varphi_n + (1-\varphi_n)\log(1-\varphi_n)], \qquad (2)$$

where k_B is the Boltzmann constant, T the absolute temperature and the sum spans over the N molecules of the aggregate.

The simplest expression for the self-energy of the array is

$$U_{SELF} = \sum_n^N [\varphi_n g_\Phi + (1-\varphi_n) g_\Psi)], \qquad (3)$$

Where g_Φ and g_Ψ are the internal energies of the Φ and Ψ conformations, respectively. Notice that g_Φ and g_Ψ are strongly affected by the interactions of the protein with its environment.

In the nearest-neighbor approximation, the energy of interaction for a 1-D array of self-aggregated proteins can be written as

$$U_{INT} = \tfrac{1}{2}\sum_n^N u_{INT}\left(\left|z_n - z_{n\pm1}\right|, \varphi_n\right), \qquad (4)$$

where u_{INT} is the potential acting between nearest-neighbor proteins at positions z_n and z_{n+1}. We define it as

$$u_{INT}\left(\left|z_n - z_{n\pm1}\right|, \varphi_n\right) \approx \tfrac{1}{2}[-Pe^{-\gamma_P(\varphi_n)\cdot(z_n - z_{n\pm1})} + Qe^{-\gamma_Q(\varphi_n)\cdot(z_n - z_{n\pm1})}],$$

where the positive constants *P* and *Q* measure the strengths of the attractive and repulsive interactions, respectively, while $\gamma_P(\varphi_n)$ and $\gamma_Q(\varphi_n)$ (with $\gamma_Q > \gamma_P$) measure their decay lengths. The parameters *P*, *Q*, $\gamma_P(\varphi_n)$ and $\gamma_Q(\varphi_n)$ can be related to experimentally accessible quantities as shown later on. Since conformational arrangements strongly change upon aggregation (Gsponer and Vendruscolo 2006), also the decay profile of the intermolecular protein-protein interactions must change. These variations are mainly associated to the proteins changing from an ordered to a disordered conformation. In the following we assume that the disordered Φ configuration is more swollen than

the ordered Ψ arrangement where the elongate shape enables a tighter packing (in SI we report some conformation-related gyration radii of IDPs).

In order to allow the proteins within an aggregate to have different conformations, we introduce a modulation of the decay lengths $\gamma_i(\varphi_n)$ (with i =P or Q) linked to the local conformational population (probability) φ_n. To the lowest order

$$\gamma_i(\varphi_n) \approx \gamma_i(\bar{\varphi}) + \frac{\partial \gamma_i}{\partial \bar{\varphi}} \cdot (\varphi_n - \bar{\varphi}), \qquad (5)$$

where $\bar{\varphi}$ is the averaged conformational population for a homogeneous equally-spaced array. The equation for $\gamma_i(\varphi_n)$ allows for the length scales of intermolecular interactions to be modulated around their average value depending on the parameters $\bar{\gamma}_i^* \equiv \frac{\partial \gamma_i}{\partial \bar{\varphi}}$ that measure the response of γ_i to variations of the conformational population $\bar{\varphi}$.

First, the free energy, Eq.(1), is minimized for a homogeneous array $\varphi_n = \bar{\varphi}$ and $|z_n - z_{n+1}|$=$\bar{a}$, where $\bar{a}$ is the protein-protein mean distance in the homogeneous state. This procedure, shown in SI (Eq.(17Sa)), yields

$$g_\Phi - g_\Psi + k_B T \log \frac{\bar{\varphi}}{1-\bar{\varphi}} + |\alpha| \Lambda \, \bar{a} = 0, \qquad (6)$$

where $\Lambda \equiv \frac{\bar{\gamma}_Q \bar{\gamma}_P^* - \bar{\gamma}_P \bar{\gamma}_Q^*}{\bar{\gamma}_Q - \bar{\gamma}_P}$ and $-|\alpha|$ is the protein-protein adhesion energy (a thorough discussion will be made after Eq.(13)). Solution to Eq.(6) yields an expression for the averaged conformational population $\bar{\varphi}$ in an equally-spaced 1-D lattice.

Stability criterion

The next step is the calculation of the fluctuations around $\bar{a}$ and $\bar{\varphi}$. This is most conveniently done at the continuum limit by replacing the discrete variables φ_n and $z_n - z_{n\pm1}$ by their continuum analogues:

$$z_n - z_{n\pm1} \approx \bar{a} \pm \frac{\partial \eta(z)}{\partial z} \bar{a} + \frac{1}{2} \frac{\partial^2 \eta(z)}{\partial z^2} \bar{a}^2 + \ldots \;, \quad \varphi_n = \varphi(z) \qquad (7)$$

where $\bar{a}$ is the average distance between two nearby proteins. Then, inserting Eq.(7) in Eq.(1), expanding it in power series for $\varphi(z) - \bar{\varphi}$, $\partial\eta(z)/\partial z$ and $\partial^2\eta(z)/\partial z^2$ and retaining terms up to second order gives (see SI, Eq.(17Sb))

$$G \approx G_o + \bar{a}^{-1} \int_{-\ell}^{+\ell} \left[\tfrac{1}{2} A_1 \left(\frac{\partial^2 \eta}{\partial z^2} \right)^2 + \tfrac{1}{2} A_2 \left(\frac{\partial \eta}{\partial z} \right)^2 + \tfrac{1}{2} A_3 (\varphi - \bar{\varphi})^2 + A_4 \frac{\partial^2 \eta}{\partial z^2} (\varphi - \bar{\varphi}) \right] dz, \quad (8)$$

where G_o is the total energy without fluctuations, $2\ell = N\bar{a}$ is the length of an aggregate of N proteins, and $A_1 = \frac{1}{4}|\alpha|\bar{a}^4 \bar{\gamma}_P \bar{\gamma}_Q$, $A_2 = |\alpha|\bar{a}^2 \bar{\gamma}_P \bar{\gamma}_Q$, $A_3 = k_B T[\bar{\varphi}(1-\bar{\varphi})]^{-1} - |\alpha|\bar{a}^2 \frac{(\bar{\gamma}_Q^*)^2 \bar{\gamma}_P - (\bar{\gamma}_P^*)^2 \bar{\gamma}_Q}{\bar{\gamma}_Q - \bar{\gamma}_P}$ and $A_4 = \frac{1}{2}|\alpha|\bar{a}^3 \bar{\gamma}_P \bar{\gamma}_Q \left(\frac{\bar{\gamma}_Q^* - \bar{\gamma}_P^*}{\bar{\gamma}_Q - \bar{\gamma}_P} \right)^2$. With Eq.(8), by denoting the monomer mass by m and introducing a kinetic energy term as $K = \frac{\bar{a}^{-1}}{2} \int_{-\ell}^{+\ell} \left[m \left(\frac{\partial \eta}{\partial t} \right)^2 \mathrm{dz} \right]$, we can now construct a Lagrangian for the system, $\mathcal{L} = K - U$. Let $I \equiv \int_0^t \mathcal{L} dt$, the motion is such that the variation of *I* is zero (see, e.g., (Goldstein *et al.* 2001). This procedure leads to the Euler-Lagrange equations shown in eq.(18S). By using the explicit form of $\mathcal{L}$ derived above, eventually we obtain a simple system of two linear partial differential equations

$$-\rho\bar{a} \frac{\partial^2 \eta}{\partial t^2} = A_1 \frac{\partial^4 \eta}{\partial z^4} - A_2 \frac{\partial^2 \eta}{\partial z^2} + A_4 \frac{\partial^2 \varphi}{\partial z^2}, \qquad (9a)$$

$$0 = A_3(\phi - \bar{\phi}) + A_4 \frac{\partial^2 \eta}{\partial z^2}, \qquad (9b)$$

where $\rho \equiv m/\bar{a}$ is the linear density of the protein array. Assuming symmetry across zero, use of Fourier series expansions: $\eta = \sum_q \eta_q(t) e^{iqz} + \text{c.c.}$ and: $\phi - \bar{\phi} = \sum_q \phi_q(t) e^{iqz} + \text{c.c.}$ (c.c. = Conjugated Complex), yields from Eqs.(9)

$$\frac{d^2 \eta_q(t)}{dt^2} = \Theta^2(q) \eta_q(t) \qquad \phi_q(t) = q^2 \frac{A_4}{A_3} \eta_q(t), \qquad (10)$$

where: $\Theta^2(q) \equiv \frac{1}{\rho\bar{a}} \left(\left(A_1 - \frac{A_2^2}{A_3} \right) q^4 + A_2 q^2 \right)$ depends on the wave number *q*. It can be noted that when $\Theta(q)>0$ (stable region), the solutions to Eq.(10) are: $\eta_q \propto \exp(i|\Theta(q)|t)$ and $\varphi_q \propto \exp(i|\Theta(q)|t)$, that is, both protein-protein distance and the protein conformational population oscillate near the equilibrium values $\bar{a}$ and $\bar{\varphi}$. The amplitudes of those oscillations are usually small and are calculated by adding to Eq.(10) a Fourier series representation of the thermal noise. The resulting Langevin's equation can be solved by well-known procedures (Risken 1984). On the contrary, when $\Theta(q)<0$ (unstable region), the solutions are $\eta_q \propto \exp(+|\Theta(q)|t)$ and $\varphi_q \propto \exp(+|\Theta(q)|t)$. In other words, the system becomes unstable and even the smallest fluctuations applied at $t = 0$ grow exponentially in time. The growth of the protein distance η and that of the

conformational population φ suggests that, beyond some critical values, the homogeneous system is no longer stable against infinitesimal fluctuations and breaks down into *a patched structure*. This behavior is akin to spinodal decomposition (Onuki 2002).

It is worth mentioning that not all fluctuations in the unstable region grow at the same rate: those with the largest $\Theta(q)$ will grow the fastest. The q value which maximizes $\Theta(q)$, q^* say, is obtained from the condition: $\frac{\partial\Theta(q)}{\partial q} = 0$. The final dimensions of the protein domains (defined as the protein clusters inside the patterned aggregate having similar conformational population) will keep memory of this dynamic processes and would be of order q^*. Using the explicit expressions for $\Theta(q)$ (Eq.(10)) and A_{i} (Eq.(8)) and using $\frac{\partial\Theta(q)}{\partial q} = 0$, we find that the fastest growth rate of the patterned structures arises when

$$q^*\bar{a} = \left(-2\frac{\frac{k_BT}{\bar{\varphi}(1-\bar{\varphi})} - |\alpha|\bar{a}^2 f_1}{\frac{k_BT}{\bar{\varphi}(1-\bar{\varphi})} - |\alpha|\bar{a}^2(f_1+f_2)}\right)^{1/2} \qquad (11)$$

where: $f_1 = \frac{(\bar{\gamma}_Q^*)^2\bar{\gamma}_P - (\bar{\gamma}_P^*)^2\bar{\gamma}_Q}{\bar{\gamma}_Q - \bar{\gamma}_P}$ and $f_2 = \bar{\gamma}_P\bar{\gamma}_Q\left(\frac{\bar{\gamma}_Q^* - \bar{\gamma}_P^*}{\bar{\gamma}_Q - \bar{\gamma}_P}\right)^2$. Equation (11) is the main result of the theory, predicting stability and wavelength (size) of the patterned structures. Indeed, only if the right-hand side of Eq.(11) is real, stable patterned structures may exist. This condition is satisfied only if in Eq.(11) the following inequalities are fulfilled: $|\alpha|\bar{a}^2 f_1 < \frac{k_BT}{\bar{\varphi}(1-\bar{\varphi})} < |\alpha|\bar{a}^2(f_1+f_2)$. Near the lower boundary, $\frac{k_BT}{\bar{\varphi}(1-\bar{\varphi})} \approx |\alpha|\bar{a}^2(f_1+f_2)$, the patterned structures exhibit the highest $q^* \to q_{max}$ (i.e., a close alternation of ordered and disordered proteins). On the contrary, on approaching the upper boundary, $\frac{k_BT}{\bar{\varphi}(1-\bar{\varphi})} \approx |\alpha|\bar{a}^2 f_1$, we find: $q^* \to 0$, that is, the patterned phase contains very large domains. To calculate the boundaries in a phase diagram, we need the concentrations of ordered and disordered conformations, $\bar{\varphi}$ and $(1-\bar{\varphi})$, given by Eq.(6) and depend on the same parameters that affect the protein-protein interaction. Whence, combing Eqs.(6) and (11) the boundaries between homogeneous and patched aggregates are obtained solving the algebraic equations

$$\frac{\left(1+\exp(-\frac{\Delta G_{UNF}(|\alpha|)}{k_BT})\right)^2}{\exp(-\frac{\Delta G_{UNF}(|\alpha|)}{k_BT})} = \frac{|\alpha|\bar{a}^2}{k_BT}(f_1+f_2), \qquad \frac{\left(1+\exp(-\frac{\Delta G_{UNF}(|\alpha|)}{k_BT})\right)^2}{\exp(-\frac{\Delta G_{UNF}(|\alpha|)}{k_BT})} = \frac{|\alpha|\bar{a}^2}{k_BT}f_1 \qquad (12a)$$

where:

$$\Delta G_{UNF}(|\alpha|) \equiv g_\Phi - g_\Psi + |\alpha|\Lambda\bar{a} = \Delta G^o_{UNF} + |\alpha|\Lambda\bar{a} \qquad (12b)$$

is the unfolding energy of the protein self-aggregate. The energy difference $g_\Phi - g_\Psi$ can be identified with the unfolding free energy of the *isolated* protein, ΔG^o_{UNF}. On the contrary, no direct measurements of the self-adhesion energy $|\alpha|$are available, while the experimental free energy variation upon the assembly of *N* identical proteins reads: $\Delta G_{ADH} = \frac{1}{N}(U(N) - N \cdot U(N=1))$. This formula accounts for the reorganization energy of the proteins self-energy upon the formation of a *N-mer* and it is valid in the limit *N>>1* (negligible end effects). Exploiting Eqs.(1)-(5), we derived in SI (Eq.27S) a compact expression for ΔG_{ADH} (per protein molecule)as a function of $|\alpha|$ for a 1-D aggregate

$$|\Delta G_{ADH}| \approx |\alpha| \left(1 - \frac{\exp\left(-\frac{\Delta G^o_{UNF}}{k_B T}\right)}{\frac{1}{2}\left(1+\exp\left(-\frac{\Delta G^o_{UNF}}{k_B T}\right)\right)^2} \frac{|\alpha|}{k_B T} \Lambda^2 \bar{a}^2 + O((\frac{|\alpha|}{k_B T} \Lambda^2 \bar{a}^2)^2) \right) \qquad (13)$$

showing how the self-adhesion energy $|\alpha|$is modified by the reorganization effects (the second term in the right hand side of Eq.(13)).

By eliminating$|\alpha|$ from the system of equations (12a,b) and (13), we may calculate the ΔG_{ADH} vs. ΔG^0_{UNF} phase diagram. Specifically, if a point defined by the two accessible control parameters ΔG_{ADH} and ΔG^0_{UNF} falls inside the unstable region, large fluctuations in conformational population and intermolecular distances begin to develop within the aggregate, eventually leading to patterned structures. On the contrary, inside the stable region, the protein aggregate remains homogeneous.

So, while the requirements for the formation of patterned structures can be predicted with a good degree of accuracy and the final morphology can be somehow inferred by Eq. (11), what is lost in our linear stability analysis is the final conformational composition of the mature protein aggregates. This goal can be reached only by including higher order terms in the series expansion of the energy functional (Eq. 8). The mathematics, however, becomes very involved (non-linear soliton-like partial differential equations (Drazin and Johnson 1989)) and the number of parameters increases, this task is beyond the aims of the present study and will be addressed in the next section by MD simulations. Accordingly, one expects to observe, under particular conditions ($\Theta(q)<0$), an alternating ordered array of small domains richer in Φ and Ψ arrangements, respectively as shown in Fig.1. We have extended the theory to 3-D aggregates, details are reported in SI. In the limit of isotropic interactions, the final formulas remain unchanged, apart from a renormalization of the wavenumber *q*. The calculated final arrangements for the 1-D and 3-D cases are pictorially shown in Fig.1.

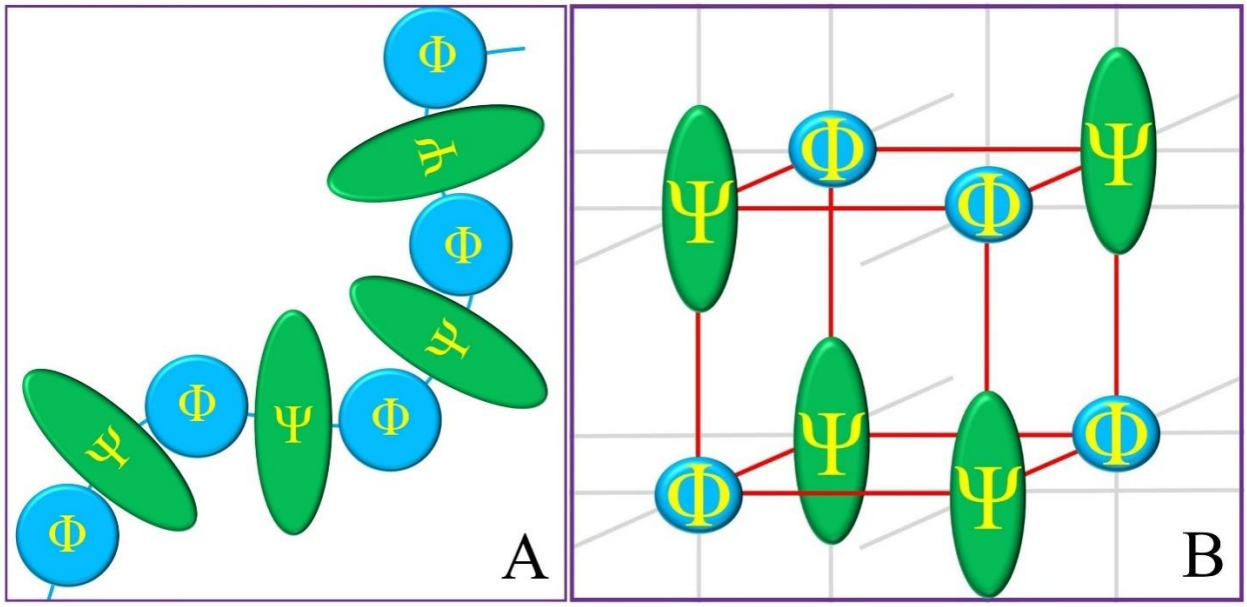

Figure 1. The most likely configurations of two-state proteins above a critical threshold (see the text for the definition). For the sake of clarity, proteins have been depicted by the pure Φ and Ψ states (fully disordered and fully ordered configurations, respectively). In our model, proteins are described by a weighted combination of Φ and Ψ states. Panel A: 1-D arrangement. Panel B: 3-D arrangement.

Then, we numerically solved the system of equations (12a,b) and (13) in order to obtain a qualitative phase diagram reported in Fig. 2. The drawing shows the boundaries among stable and unstable regions as a function of two control parameters. We selected as parameters two experimental quantities: the (attractive) protein-protein interaction free energy, ΔG_{ADH} against the unfolding free energy ΔG_{UNF}. If a point identified by the above defined control parameters falls inside the unstable region, large fluctuations in conformational population and intermolecular distances begin to develop, eventually leading to a patterned structure. Conversely, inside the stable region the protein aggregate remains homogeneous. Figures have been calculated for selected values of the parameters$\bar{\gamma}_i$, $\bar{\gamma}_i^*$and $\bar{a}$. The phase diagrams in Fig. 2 clearly shows two distinct regions, homogeneous and patched, respectively.

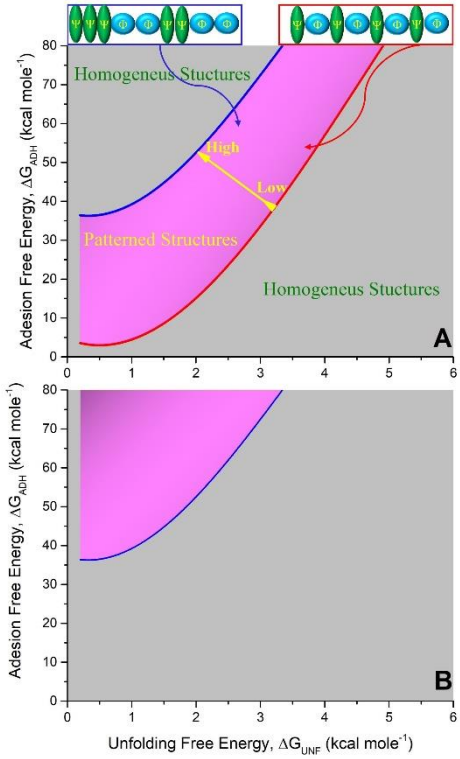

Figure 2. Phase diagram showing the stability regions (homogeneous and patterned structures) for a 1-D array of proteins with internal conformational flexibility. The protein-protein self-adhesion energy, ΔG_{ADH}, is plotted against the unfolding energy ΔG^0_{UNF} of the isolated monomer. ***Panel A***: strong geometrical variations upon protein conformational transition, ***panel B***: weak geometrical deformations (20% of those of panel A). *Large* and *small* labels the size of the patterned structures, pictorially sketched at the top of the figure.

Lastly, we repeated the calculations in the simpler case of dimeric protein aggregates. Results, reported below and in SI, predict a qualitatively similar behavior as that of infinite arrays.

Main qualitative conclusions from the model

Let us summarize the main features of the theoretical model before discussing the MD simulations and SERS experiments. An equally spaced 1-D array made of protein repeat units that may assume two different geometrical arrangements becomes unstable when: i) the internal energies of Φ (disordered) and Ψ (ordered) arrangements are slightly different (of order of $1k_BT \approx 0.5\ kcalmole^{-1}$ at room temperature); ii) there are large differences in the geometries of the Φ and Ψ arrangements, and iii) there are strong interactions between the proteins (of order of $10\ k_BT$). The phase diagram for a 1-D array of proteins is shown in Fig. 2. By comparison, it is evident that strong

protein-protein interactions stabilize the patterned structures, provided a critical strength of the interactions (of order of a few k_BT) has been reached. Thus, the formation of periodic structures is unlikely, but not unrealistic. For variation of the intermolecular forces decay length upon Φ to Ψ protein arrangement as small as 10%, and for internal energy difference between Φ and Ψ of order of a few k_BT and proteins radii within the experimental range, patterned 1-D structures emerge, provided the protein-protein interaction energy is of order of 10 k_BT. Obviously, other parameters, like the protein-protein mean distance $\bar{a}$, modify the detailed shape of the curves, so an "universal" phase diagram cannot be drawn even using more advanced models. Nevertheless, our simple theory unveils a close relationship between patterns geometry, self-adhesion energy and unfolding energy, a relation that would be obscured in a more detailed picture.

Once the system becomes unstable, fluctuations grow bringing the protein array toward less symmetric configurations. The growth of local heterogeneities depends on the wave number: patterns with the largest wave number q grow fast near the lower boundary of the instability region (see Fig. 2), while large patterns (i.e. $q \to 0$) appear on approaching the upper boundary. Thus, it is conceivable that the most likely final organization is given by an alternating array of mostly ordered and mostly disordered proteins as shown schematically in Fig. 1.

Results and Discussion

A comparison of the experimental data with model predictions

The qualitative model developed above was tested for some relevant classes of proteins with known assembly and folding free energies. The energies of aggregation, ΔG_{ADH}, and the energy difference between Φ and Ψ internal states in the monomeric state, ΔG^{o}_{UNF}, were estimated from calorimetric and chemical denaturation measurements. Fig. 3 show phase diagram of patterned and homogenous of some amyloidogenic and non-amyloidogenic proteins. The free energy of self-adhesion, ΔG_{ADH} (a negative quantity because of the favourable protein-protein interactions in self-assembled proteins) was calculated from the experimental dissociation energy: $\Delta G_{ADH} = -\Delta G_{DISS}$.

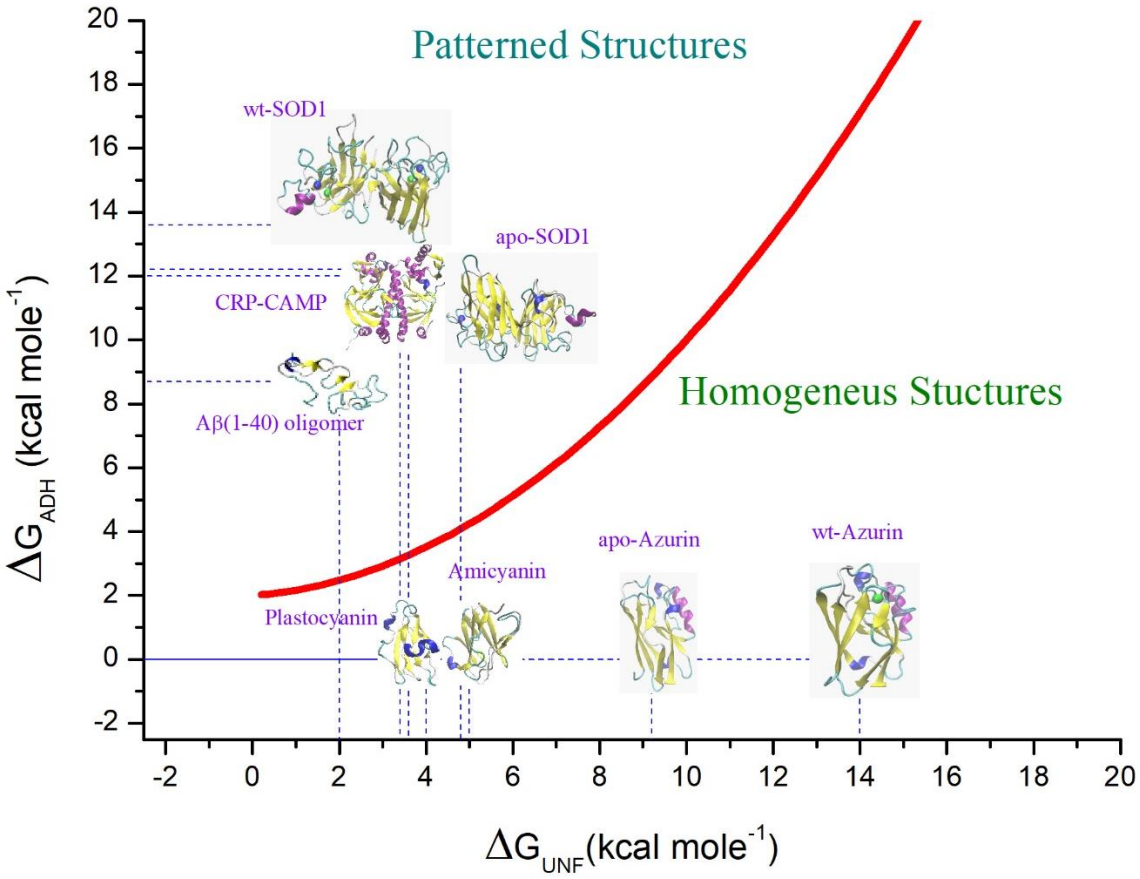

Figure. 3. Phase diagram (ΔG_{ADH} of self-adhesion versus ΔG^{o}_{UNF} of unfolding) of some proteins. The plot includes amyloidogenic, (Aβ(1-40) (Iljina *et al. 2016)*) and globular proteins in aqueous solutions (wt-SOD1 (Khare *et al.* 2004; Stathopulos *et al*. 2006), apo-SOD1 (Svensson *et al.* 2006, 2010), plastocyanin (Milardi *et al.* 1998), amicyanin (La Rosa *et al*. 2002), wt-azurin (La Rosa *et al.* 1995), apo-azurin (Pappalardo *et al.* 2008) and CRP-CAMP (Cheng *et al.* 1993)). Red line represents the qualitative boundary between patterned and homogenous aggregate structure. The *ΔG* values are reported in table S1 of SI.

Superoxide dismutase (SOD1) is a thermodynamically well characterized globular dimeric metal-protein. SOD1 undergoes thermal or chemical denaturation following a two-step pathway: in a first step the dimeric form dissociates into two monomers and then unfolding occurs with a Gibbs free energy change of 13.5 kcal mole^{-1} (dissociation) and 3.5 kcal mole monomer^{-1} (unfolding) respectively. This datum suggests that SOD1 lies in the patterned structures side of the phase diagram 3. In order to verify this conjecture, accurate MD simulations of SOD1 were performed. MD data are in agreement with model expectations, showing that the two subunits have a different number of amino acid in ordered conformations (sum of α-helix and β-sheet) as reported in Fig.S1. This difference decreases at increasing temperatures during unfolding process (Fig. S2). The dimer asymmetry was found also in the crystal structure (pdb code 1SPD (Deng *et al.* 1993)). In fact, the two monomers show 56% and 34% of amino acids with ordered secondary structures, respectively. On the contrary, globular proteins, such as blue-copper proteins, that do not show any tendency to

form multi-meric aggregates, lie in the homogeneous side of Fig. 3 because of the negligible self-adhesion energy and the large unfolding energy. Moreover, even though the crystallographic unit cell of azurin contains four proteins, no differences in the conformational population were detected. For additional information, Table S1 reports the Gibbs free energy of some patterned and homogeneous proteins.

Molecular dynamics test of the theoretical model

IAPP oligomers formation. During the simulation, hIAPP monomer (over 500 ns at 300 K in 0.1 M NaCl solution) undergoes a conformational transition toward a structure exhibiting an antiparallel cross β-sheet (residues 24-28 and 15-20, yellow color in Figure S3) with a connecting turn (cyan) encompassing residues 21-23 according to literature data (Milardi *et al.* 2008; Reddy *et al.* 2010; Sciacca *et al.* 2018). The secondary structure evolution of dimer and a snapshot of the final structure are shown in Fig. 5 (contact map is shown in Fig. S4). The first monomer retained its antiparallel β-conformation for 140±20 ns. Then it became partially ordered, with shorter β-sheets and a small but noticeable α-helical portion at one end. The second monomer preserved its antiparallel β-conformation for 80 ns, then quickly unfolded and became disordered with hints of short discontinued helices and a β-bridge. Interestingly, the arrangement of a partially ordered monomer and a largely disordered partner was found to maintain stable intermolecular contacts in the individual simulations. To best describe the dimer secondary structures, a statistical analysis over 500 ns of three independent runs was performed. Dimeric complexes were clustered based on their RMSD values using the Gromos algorithm (Daura et al. 1999) and a RMSD cut-off for two structures to be neighbors within 0.6 nm. This procedure yields 3 most populated clusters C1 (25%, rmsd=0.44 nm), C2 (24%, rmsd=0.38 nm) and C3 (21%, rmsd=0.51 nm). Other small clusters where found, the most rapresentative one being populated up to 5%. For each cluster, the most representative structure and the centroid of the structures in a given cluster, are represented in Figure 5.

These observations partially disagree with the implicit solvent simulations of Dupuis *et al.* (Dupuis *et al.* 2011) who found still an asymmetric composition of the dimer, but with a different conformational distribution. Our simulations do, however, agree with those of Derreumaux *et al.* (Tarus *et al.* 2015) who investigated the Aβ (1-40) dimers by atomistic simulations and found a limited percentage of helix and sheet conformations using both explicit solvent simulations as well implicit solvent methods for free energy analysis (Tarus *et al.* 2015; Nguyen *et al.* 2016) .

An interesting consequence of the order-disorder alternation is that the radii of gyration of the single peptides within the aggregate are not identical. This is clearly seen in Fig. 4 which shows them for two hIAPP monomers within a dimeric aggregate. Such an effect is difficult to detect in single snapshots from the aggregate, but becomes unambiguous after averaging over long times (see Fig. 5). Similar results were observed also in hexamers confirming that the heterogeneous nature of the peptides aggregates persists even after a long equilibration time.

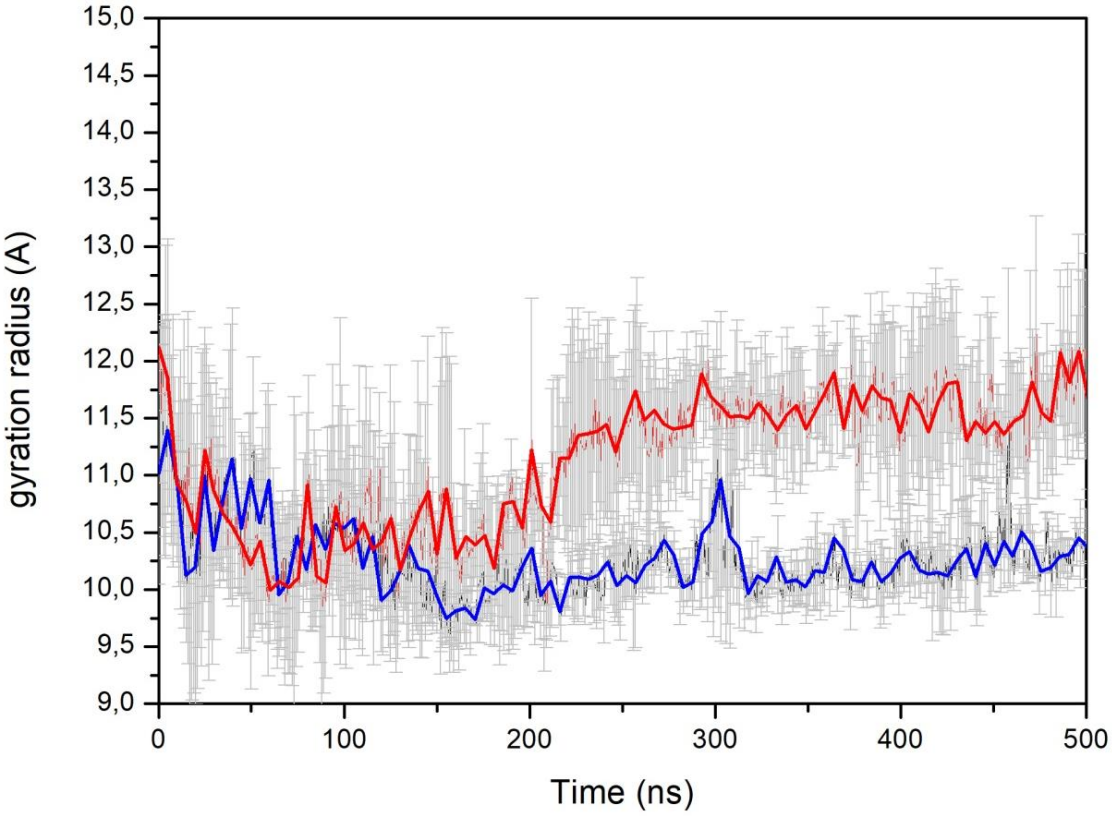

Figure 4. Radius of gyration of hIAPP dimer. Simulation of the early stage time evolution of a dimeric aggregate consisting of two *identical* monomers brought at constant distance at time t = 0 ns. Notice: 1) the formation of an asymmetric dimer made up of two juxtaposed unequally sized monomers at late times; 2) the intense noise due to the continuous conformational rearrangement obscures the dimer asymmetry unless a statistical analysis has been performed.

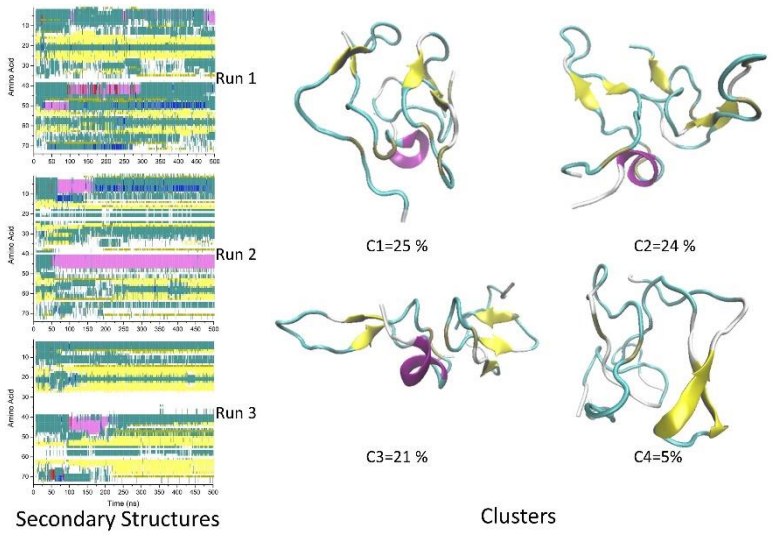

Figure 5. Left: DSSP plot of the secondary structure evolution of a hIAPP dimer over 500 ns of three independent runs. Right: Cluster analysis of dimeric assembly of hIAPP in 0.1 M electrolyte solution. Notice the coexistence of partially ordered (mainly β-sheets) and partially disordered domains in the dimeric aggregate. Color code: white-coil; light yellow-β-sheet; dark yellow-isolated bridge; pink-α-helix; blue-3-10 helix; red-π-helix; green-turn.

The tetramer resembles the dimer, i.e., the tetramer shows the partially ordered-disordered sequence (Figs. S5-S7).

The peptides self-assembled into a hexameric aggregate after about 50 ns. The aggregate exhibits the characteristic partially order-disorder arrangement as observed for dimers and tetramers. The average percentage of different secondary structures in the last 60 ns of the 1020 ns simulation (Fig. S8) shows random coil is the dominant conformation in all molecules while helical structures are rarely sampled. Interestingly, Fig. S8 also shows that the molecules with more β-strand tend to have less turns (molecules 2, 4, and 6) while molecules with less β-strand tend to have more of them (molecules 1, 3, and 5). Unlike with dimers and tetramers, the alternating ordered-disordered correlation between the different molecules in hexamers is less evident. Nonetheless, the strand-turn complementary shows an interesting partially ordered-disordered feature *within* each of the peptides in the hexamer aggregate (see fig.6D). The time evolution of the secondary structures is shown in Fig. S9. Contact map and a snapshot of the final hexameric hIAPP assembly are also reported (Fig. S10, S11). The radius of gyration of the hexamer hIAPP is reported in Fig S12.

Our MD results are summarized in Fig. 6 which clearly shows the order-disorder alternation in the conformational population of hIAPP monomers, dimers, tetramers and hexamers.

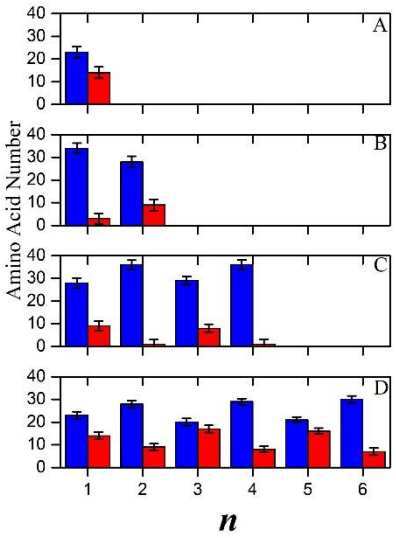

Figure 6. Fractions of ordered structures and their indeterminations, calculated by VMD and averaged over ten frames of the last 50 ns of simulation, along an array of hIAPP peptide suspended in a 0.1 M electrolyte solution. *n* labels the position of a generic *n-th* peptide within the linear aggregate. Blue bars: disordered secondary structures considered as a sum of coils and turns. Red bars: partially ordered secondary structures as a sum of α-helices and β-strands. A = monomer, B = dimer, C = tetramer, D = hexamer.

To further investigate the validity of the model, we performed MD simulations of the hexamer without salt (only counterions were present). The theory predicts the disappearance of the modulated structures for weakly adhesive assemblies (Fig. 2). When no added salt is present, the electrostatic repulsion is stronger due to lack of screening. Strong repulsion weakens the stability of the assembly and destroys the modulated phases. Results are summarized in Fig. S13 where we report the fraction of ordered (blue) and disordered (red) arrangements in a hIAPP hexamer aggregate in pure water.

The comparison with the corresponding system embedded in a 0.1 M electrolyte solution (Fig. 6D) is striking: virtually, no modulated phases exist in pure water and the distribution, apart from position n=3, resembles that of a monomer (Fig. 6D and S2). Interestingly, the ordered structures of hIAPP in pure water are mainly helices, while in 0.1 M electrolyte solution the most abundant component is the β-strand arrangement (Fig. S3). The strong sensitivity of the conformational landscape to electrolytes in amyloid Aβ-peptides has been recently shown by Smith and Cruz (Smith and Cruz 2013) using MD simulations. In addition, in another previous study using pure water no alternation was observed in aggregates of the closely related Aβ(1-42) fragment (Masman *et al*. 2009).

*Aβ (1-40) large oligomers formation.*Lastly, the analysis of the trajectories reported by Strodel *et al.* for a ribbon-like 20-mer aggregate of Aβ(1-40) (Barz *et al.* 2014) yields an alternate distribution along both the short and the long-axes of the ribbon (see Fig. 7).

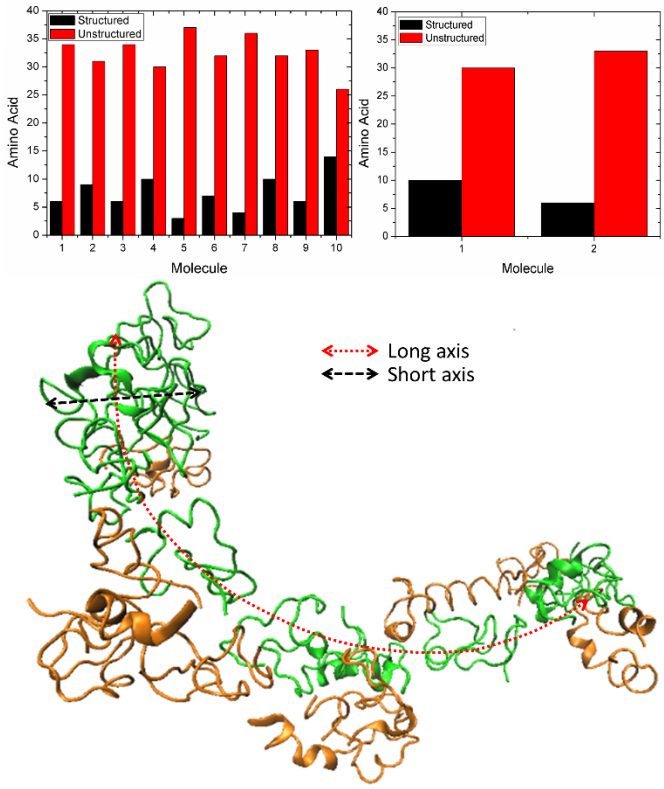

Figure 7. Top left histogram: fraction of ordered and unordered structures averaged over ten frames of the last 50 ns of simulation counted over the long axis (red axis of cartoon representation). Top right histogram: fraction of ordered and unordered structures evaluated over the short axis (black axis of cartoon representation). Below: cartoon representation of semi-toroidal aggregate formed by 20 Aβ(1-40) molecules after 500 ns of implicit solvent simulation.

Also, ovine prion protein and mixture of 1:1 Molar ratio of IAPP and Aβ(1-40) show the same behavior as reported in SI (see Fig.S15 and S16).

Surface-Enhanced Raman Scattering (SERS) test of the theoretical model

SERS is a technique able to detect the secondary structures of amyloidogenic peptides and proteins at very low concentrations. Although spatial and temporal resolutions are not comparable to MD, at low concentration SERS is able to reveal, with a good precision, the secondary structure of amyloidogenic proteins (D'Urso *et al.* 2018). Here we use SERS to determine inhomogeneous

secondary structures within small aggregates of IAPPs. SERS spectra of 10 nM hIAPP solution containing 1 μM of silver Nano Particles (NPs) were recorded at different times. Representative spectra of the most informative spectral regions are shown in Figs. S17-S19. As time proceeds, proteins self-aggregate from monomers to oligomers, and eventually to proto-fibrils and fibrils. The aggregation kinetics of these assemblies is well known and can be followed by conventional light scattering (Nag *et al.* 2011) and fluorescence lifetime correlation spectroscopy (Paredes *et al.* 2012) measurements.

We can obtain quantitative information by comparing the relative intensities of the various bands inside each spectrum. Since all signals assigned to α-helix, β-strand and random coil are well resolved, in Table 2 we report the results of such an analysis in the case of the β-sheet to random coil intensity ratio, considered at different times after mixing hIAPP and silver NPs. The time evolution of the amide III band confirms that at low incubation time (0-15 min.) the peptide exists as a monomeric species in agreement with Dynamic Light Scattering measurements (Nag *et al.* 2011) performed at the same monomer concentration. After 40 minutes of incubation, the sharp amide III band broadens and the three peaks assigned to helix, sheet and random coil are rather convoluted (Wang *et al.* 2013). Lastly, at very long incubation times (> 360 min) the formation of proto-fibrils is observed (Fändrich 2012).

Our SERS results unambiguously confirm that the distribution of the peptide conformational populations becomes *broader* upon monomer to dimer self-aggregation (Table 2), while retaining a comparable heterogeneity upon increasing size of the aggregates. Although SERS data cannot straightforwardly prove the order-disorder alternation in the aggregates, they are consistent because:

a) significant changes in conformational populations are observed on passing from monomers to ~~a~~ dimers and, spectroscopically, larger oligomers behave as a collection of ordered-disordered dimers as predicted by the theory and demonstrated by MD;

b) our MD simulations cover microsecond range, SERS data run over minutes. Therefore, the simulated alternate order-disordered arrangement is not a transient state but it persists over much longer times than those typical of MD simulations (10^{-6} s).

The data in Table 2 highlight an overall agreement between experiments and simulations.

Table 2. Secondary structures evaluated as the ratio (β-sheet)/(random coil) from SERS measurements recorded at different times *(minutes)*. At each monomer concentration aggregates modify their own size as time goes by, as reported in the literature (Sumner Makin and Serpell 2004; Nguyen *et al.* 2016). In parenthesis we report the (β-sheet)/(random coil) ratio taken from MD data obtained for the different aggregates. Simulation and SERS kinetics have different time scales. Here, molecular dynamics are used to calculate the secondary structure of each aggregation state. The error over SERS measurements was calculated by considering I and III amide band of three independent spectra. [a)] Average of (β-sheet)/(random coil) ratio calculated by MD over monomer (0.48) and dimer (0.75). [b)] Average of (β-sheet)/(random coil) ratio calculated over dimers (0.75), tetramers (0.80) and hexamers (0.73).

Time (minutes)	β-sheet/coil ratio from SERS spectra	Aggregation Status and (β-sheet)/(coil) ratio from MD data
0	0.5±0.2	Monomers (0.48±0.10)
15	1.0±0.3	Mostly Dimers (0.85±0.26[a])
40	1.1±0.2	Small Oligomers (0.76±0.46[b])
120	1.0±0.3	Large Oligomers
240	1.2±0.3	Proto-fibrils + large oligomers
360	2.0±0.3	Proto-fibrils + large Oligomers

Our main results are a theoretical model predicting an instability that leads to periodic distribution of partially ordered and disordered peptide structures in aggregates, and the confirmation of the theoretical predictions using MD simulations and SERS. The main findings from the MD simulations are summarized in Figs. 6 and 7 which show the percentages of the secondary structures in a hIAPP hexamer, and the partially ordered (α-helix and β-strand) and disordered (turn and coil) structures in all hIAPP systems, respectively. More details on conformational arrangements are given in Fig. S14 for ovine prion and in Fig. S15 for equimolar mixtures of hIAPP and Aβ(1-40).

High sensitivity SERS spectroscopy was used to provide information about the β-sheet to random coil ratio in hIAPP aggregates upon oligomerization. The results are consistent with the appearance of modulated structures. Since the theory is based on the hypothesis of conformational symmetry breaking, change in the ratio between ordered and disordered structures observed upon passing from monomer to aggregates is a clear indication of the conformational changes occurring upon aggregation. Moreover, as predicted by the theory and MD data, this ratio remains constant upon increasing the size of the oligomeric structures.

On a larger scale (microns), alternating patterns have also been recently observed in Tip-Enhanced Raman Spectroscopy (TERS) measurements on hIAPP fibrils where an inhomogeneous configuration of secondary structures was found along the fibril surface. In addition, a recent ion mobility mass spectrometry study reported the coexistence of compact and extended structures in an ensemble of low-order Aβ peptide oligomers. Particularly interesting is a recent study using advanced scanning microscopy (vandenAkker *et al.* 2015) and vibrational sum-frequency generation spectroscopy on Aβ fibrils (vandenAkker *et al.* 2011). Other observations have been reported in Refs. (Gorman and Chakrabartty 2001; Klinger *et al.* 2014; vandenAkker *et al.* 2015). These works provide further support to our hypothesis.

Conclusion

The simple theoretical model we have developed assumes a coupling between internal arrangement and intermolecular interactions: for a given intermolecular distance, the peptide (protein) internal conformations may change in order to minimize the total free energy. In turn, the new conformations influence the strengths of the interactions among self-assembled peptides (proteins). When the energy difference between ordered and disordered structures become similar and the interactions are appreciable, spatially modulated patterns may spontaneously emerge. This kind of phenomena are widespread in condensed matter and are often referred to as "pre-transitional" effects. They are observed, for instance, in lipid bilayers, where a strongly corrugated phase (the Ripple phase) appears on approaching the gel to liquid-crystalline phase transition (Akabori and Nagle 2015). According to our model, IDPs are the best candidates for observing an alternate distribution of conformations as confirmed by accurate MD simulations on different self-assemblies. Even non-IDP systems such as wt and apo SOD1 and CRP-CAMP form dimers with a distinct ordered-disordered distribution. Although simulations were performed in different laboratories by using different force fields and water descriptors, data converge toward the concept of symmetry breaking.

The results also qualitatively explain why, in general, amyloid aggregates escape a detailed structural characterization by X-ray analysis (Sumner Makin and Serpell 2004). Provided the partially ordered-disordered arrangement persists also in larger aggregates and at longer times, one may infer that the diffraction patterns of an alternate crystalline-amorphous array of scattering units yields well-resolved Bragg peaks to which a broad amorphous *halo* is superimposed. A thorough discussion of X-ray diffraction for such a complex system is beyond the aims of the present paper. However, the particular peptide arrangement observed for small hIAPP assemblies might contribute to explaining

the coarse resolution and the complex diffraction patterns of mature hIAPP fibrils (Masman et al. 2009).

To the best of our knowledge, this is the first systematic study focusing on the spontaneous periodic distortion in supramolecular peptide and protein aggregates, the possibility of heterogeneous conformational distribution along an aggregate has been previously suggested in literature (Gorman and Chakrabartty 2001; Sitkiewicz *et al.* 2014; Klinger *et al.* 2014; vandenAkker *et al.* 2015) and scattered observations have been reported in MD studies (Reddy *et al*. 2010; Dupuis *et al.* 2011; Barz *et al.* 2014; Huy *et al.* 2016; Zhang *et al.* 2016; Nguyen *et al.* 2016; Das *et al.* 2017). We proved by MD simulations that the order-disorder alternation in IDPs oligomers is a dynamic process that emerges only after statistical averaging. Such a result is consistent with the low energies required for the onset of the alternation in respect to the homogeneous state. We propose that the occurrence of small patterns of alternate ordered-disordered patterns may provide a novel rationale to explain the molecular origin of fibril polymorphism as well as the lack of short-range molecular order in mature fibrils.

Acknowledgment Discussions with Profs. E. Del Favero and L. Cantù (University of Milan, Italy) are gratefully acknowledged. MK would like to thank Natural Sciences and Engineering Council (NSERC) of Canada for financial support. Computational resources were providing by Compute Canada. DM thanks the Italian MIUR program PRIN 20157WZM8A for financial support. AR, GC and CLR have been financed by the Progetto di Dipartimento 2017-2020. FC thanks the Swiss National Science Foundation for funding. Raman measurements were performed at Bio-Nanotech Research and Innovation Tower (BRIT), University of Catania, Italy.

References

Abraham MJ, Murtola T, Schulz R, et al (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1–2:19–25. https://doi.org/10.1016/j.softx.2015.06.001

Akabori K, Nagle JF (2015) Structure of the DMPC lipid bilayer ripple phase. Soft Matter 11:918–926. https://doi.org/10.1039/C4SM02335H

Banci L, Bertini I, Blaževitš O, et al (2012) Interaction of Cisplatin with Human Superoxide Dismutase. J Am Chem Soc 134:7009–7014. https://doi.org/10.1021/ja211591n

Barz B, O. Olubiyi O, Strodel B (2014) Early amyloid β-protein aggregation precedes conformational change. Chemical Communications 50:5373–5375. https://doi.org/10.1039/C3CC48704K

Berendsen HJC, Postma JPM, Gunsteren WF van, Hermans J (1981) Interaction Models for Water in Relation to Protein Hydration. In: Intermolecular Forces. Springer, Dordrecht, pp 331–342

Berendsen HJC, Postma JPM, van Gunsteren WF, et al (1984) Molecular dynamics with coupling to an external bath. The Journal of Chemical Physics 81:3684–3690. https://doi.org/10.1063/1.448118

Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. J Chem Phys 126:014101. https://doi.org/10.1063/1.2408420

Chaikin PM, Lubensky TC (1995) Principles of Condensed Matter Physics. Cambridge University Press

Chakroun N, Prigent S, Dreiss CA, et al (2010) The oligomerization properties of prion protein are restricted to the H2H3 domain. FASEB J 24:3222–3231. https://doi.org/10.1096/fj.09-153924

Cheng X, Gonzalez ML, Lee JC (1993) Energetics of intersubunit and intrasubunit interactions of Escherichia coli adenosine cyclic 3',5'-phosphate receptor protein. Biochemistry 32:8130–8139. https://doi.org/10.1021/bi00083a011

Chiti F, Dobson CM (2006) Protein misfolding, functional amyloid, and human disease. Annu Rev Biochem 75:333–366. https://doi.org/10.1146/annurev.biochem.75.101304.123901

Compagnini G, Messina E, Puglisi O, Nicolosi V (2007) Laser synthesis of Au/Ag colloidal nano-alloys: Optical properties, structure and composition. Applied Surface Science 254:1007–1011. https://doi.org/10.1016/j.apsusc.2007.07.177

Darden T, York D, Pedersen L (1993) Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. The Journal of Chemical Physics 98:10089–10092. https://doi.org/10.1063/1.464397

Das P, Chacko AR, Belfort G (2017) Alzheimer's Protective Cross-Interaction between Wild-Type and A2T Variants Alters Aβ42 Dimer Structure. ACS Chem Neurosci 8:606–618. https://doi.org/10.1021/acschemneuro.6b00357

Daura X, Gademann K, Jaun B, et al (1999) Peptide Folding: When Simulation Meets Experiment. Angewandte Chemie International Edition 38:236–240. https://doi.org/10.1002/(SICI)1521-3773(19990115)38:1/2<236::AID-ANIE236>3.0.CO;2-M

Del Favero E, Raudino A, Brocca P, et al (2009) Lamellar stacking split by in-membrane clustering of bulky glycolipids. Langmuir 25:4190–4197

Deng HX, Hentati A, Tainer JA, et al (1993) Amyotrophic lateral sclerosis and structural defects in Cu,Zn superoxide dismutase. Science 261:1047–1051

Drazin PG, Johnson RS (1989) Solitons: An Introduction, 2 edition. Cambridge University Press, Cambridge England ; New York

Dupuis NF, Wu C, Shea J-E, Bowers MT (2011) The Amyloid Formation Mechanism in Human IAPP: Dimers have β-strand Monomer-Monomer Interfaces. J Am Chem Soc 133:7240–7243. https://doi.org/10.1021/ja1081537

D'Urso L, Condorelli M, Puglisi O, et al (2018) Detection and characterization at nM concentration of oligomers formed by hIAPP, Aβ(1–40) and their equimolar mixture using SERS and MD simulations. Phys Chem Chem Phys 20:20588–20596. https://doi.org/10.1039/C7CP08552D

Evans DJ, Holian BL (1985) The Nose–Hoover thermostat. J Chem Phys 83:4069–4074. https://doi.org/10.1063/1.449071

Fändrich M (2012) Oligomeric intermediates in amyloid formation: structure determination and mechanisms of toxicity. J Mol Biol 421:427–440. https://doi.org/10.1016/j.jmb.2012.01.006

Fazio E, Cacciola A, Mezzasalma AM, et al (2013) Modelling of the optical absorption spectra of PLAL prepared ZnO colloids. Journal of Quantitative Spectroscopy and Radiative Transfer 124:86–93. https://doi.org/10.1016/j.jqsrt.2013.02.028

Frishman D, Argos P (1995) Knowledge-based protein secondary structure assignment. Proteins 23:566–579. https://doi.org/10.1002/prot.340230412

Fuß W, Haas Y, Zilberg S (2000) Twin states and conical intersections in linear polyenes. Chemical Physics 259:273–295. https://doi.org/10.1016/S0301-0104(00)00200-7

Goldstein H, Jr CPP, Safko JL (2001) Classical Mechanics, 3 edition. Pearson, San Francisco, NJ

Gorman PM, Chakrabartty A (2001) Alzheimer beta-amyloid peptides: structures of amyloid fibrils and alternate aggregation products. Biopolymers 60:381–394. https://doi.org/10.1002/1097-0282(2001)60:5<381::AID-BIP10173>3.0.CO;2-U

Grason GM, Santangelo CD Undulated cylinders of charged diblock copolymers. Eur Phys J E 20:335–346. https://doi.org/10.1140/epje/i2006-10025-1

Grest GS, Kremer K (1986) Molecular dynamics simulation for polymers in the presence of a heat bath. Phys Rev A 33:3628–3631. https://doi.org/10.1103/PhysRevA.33.3628

Gsponer J, Vendruscolo M (2006) Theoretical approaches to protein aggregation. Protein Pept Lett 13:287–293

Haire LF, Whyte SM, Vasisht N, et al (2004) The crystal structure of the globular domain of sheep prion protein. J Mol Biol 336:1175–1183. https://doi.org/10.1016/j.jmb.2003.12.059

Harries D, Podgornik R, Parsegian VA, et al (2006) Ion induced lamellar-lamellar phase transition in charged surfactant systems. J Chem Phys 124:224702. https://doi.org/10.1063/1.2198534

Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: A linear constraint solver for molecular simulations. Journal of Computational Chemistry 18:1463–1472. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H

Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J Chem Theory Comput 4:435–447. https://doi.org/10.1021/ct700301q

Höppener JWM, Ahrén B, Lips CJM (2000) Islet Amyloid and Type 2 Diabetes Mellitus. New England Journal of Medicine 343:411–419. https://doi.org/10.1056/NEJM200008103430607

Huang J, Rauscher S, Nawrocki G, et al (2017) CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. Nat Methods 14:71–73. https://doi.org/10.1038/nmeth.4067

Huy PDQ, Vuong QV, La Penna G, et al (2016) Impact of Cu(II) Binding on Structures and Dynamics of Aβ42 Monomer and Dimer: Molecular Dynamics Study. ACS Chem Neurosci 7:1348–1363. https://doi.org/10.1021/acschemneuro.6b00109

Iljina M, Garcia GA, Dear AJ, et al (2016) Quantitative analysis of co-oligomer formation by amyloid-beta peptide isoforms. Scientific Reports 6:. https://doi.org/10.1038/srep28658

Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577–2637. https://doi.org/10.1002/bip.360221211

Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. J Phys Chem B 105:6474–6487. https://doi.org/10.1021/jp003919d

Khare SD, Caplow M, Dokholyan NV (2004) The rate and equilibrium constants for a multistep reaction sequence for the aggregation of superoxide dismutase in amyotrophic lateral sclerosis. PNAS 101:15094–15099. https://doi.org/10.1073/pnas.0406650101

Klinger AL, Kiselar J, Ilchenko S, et al (2014) A synchrotron-based hydroxyl radical footprinting analysis of amyloid fibrils and prefibrillar intermediates with residue-specific resolution. Biochemistry 53:7724–7734. https://doi.org/10.1021/bi5010409

Kurouski D, Duyne RPV, Lednev IK (2015) Exploring the structure and formation mechanism of amyloid fibrils by Raman spectroscopy: a review. Analyst 140:4967–4980. https://doi.org/10.1039/C5AN00342C

La Rosa C, Milardi D, Grasso D, et al (1995) Thermodynamics of the thermal unfolding of azurin. Journal of Physical Chemistry 99:14864–14864

La Rosa C, Milardi D, Grasso DM, et al (2002) A model for the thermal unfolding of amicyanin. European Biophysics Journal 30:559–570

La Rosa C, Scalisi S, Lolicato F, et al (2016) Lipid-assisted protein transport: A diffusion-reaction model supported by kinetic experiments and molecular dynamics simulations. J Chem Phys 144:184901. https://doi.org/10.1063/1.4948323

Le Ru EC, Blackie E, Meyer M, Etchegoin PG (2007) Surface Enhanced Raman Scattering Enhancement Factors: A Comprehensive Study. J Phys Chem C 111:13794–13803. https://doi.org/10.1021/jp0687908

Lee J, Cheng X, Swails JM, et al (2016) CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. J Chem Theory Comput 12:405–413. https://doi.org/10.1021/acs.jctc.5b00935

Lorenzo A, Razzaboni B, Weir GC, Yankner BA (1994) Pancreatic islet cell toxicity of amylin associated with type-2 diabetes mellitus. Nature 368:756–760. https://doi.org/10.1038/368756a0

Masman MF, Eisel ULM, Csizmadia IG, et al (2009) In Silico Study of Full-Length Amyloid β 1−42 Tri- and Penta-Oligomers in Solution. J Phys Chem B 113:11710–11719. https://doi.org/10.1021/jp901057w

Messina E, D'Urso L, Fazio E, et al (2012) Tuning the structural and optical properties of gold/silver nano-alloys prepared by laser ablation in liquids for optical limiting, ultra-sensitive spectroscopy, and optical trapping. Journal of Quantitative Spectroscopy and Radiative Transfer 113:2490–2498. https://doi.org/10.1016/j.jqsrt.2012.06.023

Milardi D, La Rosa C, Grasso D, et al (1998) Thermodynamics and kinetics of the thermal unfolding of plastocyanin. Eur Biophys J 27:273–282. https://doi.org/10.1007/s002490050134

Milardi D, Pappalardo M, Pannuzzo M, et al (2008) The role of the Cys2-Cys7 disulfide bridge in the early steps of Islet amyloid polypeptide aggregation: A molecular dynamics study. Chemical Physics Letters 463:396–399. https://doi.org/10.1016/j.cplett.2008.07.110

Nag S, Sarkar B, Bandyopadhyay A, et al (2011) Nature of the Amyloid-β Monomer and the Monomer-Oligomer Equilibrium. J Biol Chem 286:13827–13833. https://doi.org/10.1074/jbc.M110.199885

Nguyen PH, Sterpone F, Pouplana R, et al (2016) Dimerization Mechanism of Alzheimer Aβ40 Peptides: The High Content of Intrapeptide-Stabilized Conformations in A2V and A2T Heterozygous Dimers Retards Amyloid Fibril Formation. J Phys Chem B 120:12111–12126. https://doi.org/10.1021/acs.jpcb.6b10722

Nguyen TT, Gopal A, Lee KYC, Witten TA (2005) Surface charge relaxation and the pearling instability of charged surfactant tubes. Phys Rev E 72:051930. https://doi.org/10.1103/PhysRevE.72.051930

Nosé S, Klein ML (1983) Constant pressure molecular dynamics for molecular systems. Molecular Physics 50:1055–1076. https://doi.org/10.1080/00268978300102851

Onuki A (2002) Phase Transition Dynamics, Cambridge University Press, 2009. In: Cambridge Core

Oostenbrink C, Villa A, Mark AE, van Gunsteren WF (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. J Comput Chem 25:1656–1676. https://doi.org/10.1002/jcc.20090

Pappalardo M, Milardi D, Grasso D, La Rosa C (2007) Steered molecular dynamics studies reveal different unfolding pathways of prions from mammalian and non-mammalian species. New J Chem 31:901–905. https://doi.org/10.1039/B700764G

Pappalardo M, Sciacca M, Milardi D, et al (2008) Thermodynamics of azurin folding: The role of copper ion. Journal of Thermal Analysis and Calorimetry 93:575–581

Paredes JM, Casares S, Ruedas-Rama MJ, et al (2012) Early Amyloidogenic Oligomerization Studied through Fluorescence Lifetime Correlation Spectroscopy. Int J Mol Sci 13:9400–9418. https://doi.org/10.3390/ijms13089400

Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. Journal of Applied Physics 52:7182–7190. https://doi.org/10.1063/1.328693

Patil SM, Xu S, Sheftic SR, Alexandrescu AT (2009) Dynamic α-Helix Structure of Micelle-bound Human Amylin. J Biol Chem 284:11982–11991. https://doi.org/10.1074/jbc.M809085200

Peierls RE (2001) Quantum Theory of Solids - Oxford University Press, USA, 2001, Chapter 5.3

Porcar L, Marignan J, Ligoure C, Gulik-Krzywicki T (2000) Effect of a Nonadsorbing Polymer on the Stability of a Two-Solvent Lamellar Phase: Experimental and Theoretical Study of Critical Points of Lamellar/Lamellar Phase Separations. Langmuir 16:2581–2594. https://doi.org/10.1021/la991193a

Qiu D, Shenkin PS, Hollinger FP, Still WC (1997) The GB/SA Continuum Model for Solvation. A Fast Analytical Method for the Calculation of Approximate Born Radii. J Phys Chem A 101:3005–3014. https://doi.org/10.1021/jp961992r

Reddy AS, Wang L, Singh S, et al (2010) Stable and metastable states of human amylin in solution. Biophys J 99:2208–2216. https://doi.org/10.1016/j.bpj.2010.07.014

Riek R, Eisenberg DS (2016) The activities of amyloids from a structural perspective. Nature 539:227–235. https://doi.org/10.1038/nature20416

Risken H (1984) The Fokker-Planck Equation: Methods of Solution and Applications. Springer-Verlag, Berlin Heidelberg

Rygula A, Majzner K, Marzec KM, et al (2013) Raman spectroscopy of proteins: a review. J Raman Spectrosc 44:1061–1076. https://doi.org/10.1002/jrs.4335

Savelieff MG, Nam G, Kang J, et al (2019) Development of Multifunctional Molecules as Potential Therapeutic Candidates for Alzheimer's Disease, Parkinson's Disease, and Amyotrophic Lateral Sclerosis in the Last Decade. Chem Rev 119:1221–1322. https://doi.org/10.1021/acs.chemrev.8b00138

Schmid N, Eichenberger AP, Choutko A, et al (2011) Definition and testing of the GROMOS force-field versions 54A7 and 54B7. Eur Biophys J 40:843–856. https://doi.org/10.1007/s00249-011-0700-9

Sciacca MFM, Tempra C, Scollo F, et al (2018) Amyloid growth and membrane damage: Current themes and emerging perspectives from theory and experiments on Aβ and hIAPP. Biochim Biophys Acta Biomembr. https://doi.org/10.1016/j.bbamem.2018.02.022

Sciacca MicheleFM, Romanucci V, Zarrelli A, et al (2017) Inhibition of Aβ Amyloid Growth and Toxicity by Silybins: The Crucial Role of Stereochemistry. ACS Chem Neurosci 8:1767–1778. https://doi.org/10.1021/acschemneuro.7b00110

Scollo F, Tempra C, Lolicato F, et al (2018) Phospholipids Critical Micellar Concentrations Trigger Different Mechanisms of Intrinsically Disordered Proteins Interaction with Model Membranes. J Phys Chem Lett 9:5125–5129. https://doi.org/10.1021/acs.jpclett.8b02241

Sitkiewicz E, Kłoniecki M, Poznański J, et al (2014) Factors influencing compact-extended structure equilibrium in oligomers of aβ1-40 peptide--an ion mobility mass spectrometry study. J Mol Biol 426:2871–2885. https://doi.org/10.1016/j.jmb.2014.05.015

Smith MD, Cruz L (2013) Effect of Ionic Aqueous Environments on the Structure and Dynamics of the Aβ21–30 Fragment: A Molecular-Dynamics Study. J Phys Chem B 117:6614–6624. https://doi.org/10.1021/jp312653h

Stathopulos PB, Rumfeldt JAO, Karbassi F, et al (2006) Calorimetric Analysis of Thermodynamic Stability and Aggregation for Apo and Holo Amyotrophic Lateral Sclerosis-associated Gly-93 Mutants of Superoxide Dismutase. Journal of Biological Chemistry 281:6184–6193. https://doi.org/10.1074/jbc.M509496200

Sumner Makin O, Serpell LC (2004) Structural Characterisation of Islet Amyloid Polypeptide Fibrils. Journal of Molecular Biology 335:1279–1288. https://doi.org/10.1016/j.jmb.2003.11.048

Svensson A-KE, Bilsel O, Kayatekin C, et al (2010) Metal-free ALS variants of dimeric human Cu,Zn-superoxide dismutase have enhanced populations of monomeric species. PLoS ONE 5:e10064. https://doi.org/10.1371/journal.pone.0010064

Svensson A-KE, Bilsel O, Kondrashkina E, et al (2006) Mapping the folding free energy surface for metal-free human Cu,Zn superoxide dismutase. J Mol Biol 364:1084–1102. https://doi.org/10.1016/j.jmb.2006.09.005

Tamashiro MN, Schiessel H (2006) Rayleigh instability of charged aggregates: Role of the dimensionality, ionic strength, and dielectric contrast. Phys Rev E 74:021412. https://doi.org/10.1103/PhysRevE.74.021412

Tarus B, Tran TT, Nasica-Labouze J, et al (2015) Structures of the Alzheimer's Wild-Type Aβ1-40 Dimer from Atomistic Simulations. J Phys Chem B 119:10478–10487. https://doi.org/10.1021/acs.jpcb.5b05593

vandenAkker CC, Deckert-Gaudig T, Schleeger M, et al (2015) Nanoscale Heterogeneity of the Molecular Structure of Individual hIAPP Amyloid Fibrils Revealed with Tip-Enhanced Raman Spectroscopy. Small 11:4131–4139. https://doi.org/10.1002/smll.201500562

vandenAkker CC, Engel MFM, Velikov KP, et al (2011) Morphology and Persistence Length of Amyloid Fibrils Are Correlated to Peptide Molecular Structure. J Am Chem Soc 133:18030–18033. https://doi.org/10.1021/ja206513r

Wang Q, Wang Y, Lu HP (2013) Revealing the secondary structural changes of amyloid β peptide by probing the spectral fingerprint characters. J Raman Spectrosc 44:670–674. https://doi.org/10.1002/jrs.4253

Weber IT, Steitz TA (1987) Structure of a complex of catabolite gene activator protein and cyclic AMP refined at 2.5 A resolution. J Mol Biol 198:311–326

Yannoni CS, Clarke TC (1983) Molecular Geometry of cis- and trans-Polyacetylene by Nutation NMR Spectroscopy. Phys Rev Lett 51:1191–1193. https://doi.org/10.1103/PhysRevLett.51.1191

Zhang Y, Hashemi M, Lv Z, Lyubchenko YL (2016) Self-assembly of the full-length amyloid Aβ42 protein in dimers. Nanoscale 8:18928–18937. https://doi.org/10.1039/C6NR06850B