Biophysical studies of protein solubility and amorphous aggregation by systematic mutational analysis and a helical polymerization model

Yutaka Kuroda¹

Received: 26 September 2017 / Accepted: 13 November 2017 / Published online: 4 January 2018
© International Union for Pure and Applied Biophysics (IUPAB) and Springer-Verlag GmbH Germany, part of Springer Nature 2017

Abstract

At concentrations above solubility, a protein aggregates, most often into amorphous aggregates, and loses its function. However, unlike amyloidogenic aggregates, which are β -sheeted fibrillar aggregates often related to neurodegenerative diseases, amorphous aggregates, where proteins aggregate/oligomerize without forming specific high-order structures, are rarely the focus of biophysical studies. Hence, protein solubility with respect to amorphous aggregation remains to be fully characterized from a biophysical viewpoint. Here, I briefly describe the structural nature of proteins in amorphous aggregates before discussing systematic mutational analyses that aim to rationalize the contribution of individual amino acids to the solubility of a protein. The discussion is expected to demonstrate that protein solubility, and, accordingly, amorphous aggregation, can be understood using thermodynamic and biophysical rationales similar to those used in the study of protein stability or, more recently, amyloidogenesis. Finally, I will argue that the mathematical formalism of the helical polymerization model (HPM) proposed by Oosawa, Kasai, and Asakura's group can be readily adapted to provide a thermodynamic description of a system containing amorphous aggregates and soluble particles. The HPM and HPM-derived models imply the presence of nuclei or seeds for amorphous aggregates, similar to those hypothesized in crystallogenesis and amyloidogenesis.

Keywords Solubility · Precipitation · Mutational analysis · Thermodynamics analysis · Helical polymerization model (HPM) · Amorphous aggregation

Introduction

Protein solubility issues are encountered in diverse areas of protein chemistry. Practically, solubility problems can be overcome by optimizing the sample's solution condition (Shiraki et al. 2002; Arakawa and Tsumoto 2003; Hamada and Shiraki 2007). Biophysical studies of protein solubility and, thus, amorphous aggregation performed using exhaustive mutational and thermodynamic analysis are still scarce (Iwura et al. 2014; Kamal et al. 2016). A potential reason for this limited interest may be that amorphous aggregates are rarely

related to diseases. Another reason may be the absence of high-order structures in amorphous aggregates, which is in sharp contrast with the regularly structured β -sheeted fibrils observed in amyloid aggregates.

Despite the fact that thermodynamic analysis does not require structural information *per se*, the structural nature of both the aggregated state and the soluble state can impact on the energetics of protein association and, consequently, the thermodynamics of amorphous aggregation as well as solubility (Fig. 1). A well-known example is the unfolded protein having an aggregation tendency much higher than the same protein in the folded state, where the hydrophobic residues are buried into the protein interior (Baldwin 2012). Thus, protein stability, which is a protein's internal property, strongly influences its solubility or aggregation propensity, which is a colloidal property.

Several factors may mislead the thermodynamic analysis of amorphous aggregation or solubility. An important and obvious factor is the slow kinetics of aggregation, which may lead to erroneous interpretations if the measurements are not performed under equilibrium conditions. Additionally,

---

This article is part of a special issue on 'Biomolecules to Bio-nanomachines - Fumio Arisaka 70th Birthday' edited by Damien Hall, Junichi Takagi and Haruki Nakamura.

✉ Yutaka Kuroda
ykuroda@cc.tuat.ac.jp

¹ Department of Biotechnology and Life Sciences, Graduate School of Engineering, Tokyo University of Agriculture and Technology, 2-24-16 Nakamachi, Koganei-shi, Tokyo 184-8588, Japan

Fig. 1 Conceptual diagram of protein aggregation. In this paper, we simplify our discussion by restricting it to the four states depicted within the dotted frame. Arrows a and b describe, respectively, the aggregation of natively folded and denatured (or unfolded) proteins. c describes the structural stability of proteins. In this review, we consider aggregation 1, 2 and 3 as “amorphous” aggregates. Diagram adapted with permission from Kato A. et al. [*Seibutsu Butsuri*, 48, 185–9 (2008), in Japanese; edited by the Biophysical Society of Japan]

amorphous aggregation can be reversible, but it is not always so, and one should obviously apply equilibrium thermodynamics only to systems that are reversible (Schön et al. 2017). Furthermore, as discussed above, we should keep in mind that the structural states of both the soluble and the aggregated state do strongly influence the thermodynamics of protein aggregation.

Here, I describe systematic mutational experiments that aim at providing a biophysical rationale for protein solubility with respect to amorphous aggregation. I show that, with careful consideration of key factors such as equilibrium or aggregation rates, one can use thermodynamic concepts to describe a protein's solubility or aggregation tendency. Finally, I discuss how the helical polymerization model (HPM) (Oosawa and Kasai 1962), initially proposed by Oosawa et al. to analyze the transition of G-actin to F-actin, can be readily adapted to provide a thermodynamic description of protein solubility and amorphous aggregation at the molecular level.

Structure of proteins in amorphous aggregates

It is a popular belief that amorphous aggregation occurs upon the unfolding of proteins. Here, I describe two recent NMR experiments showing that the association of natively folded or near-natively folded proteins can, indeed, form amorphous aggregates.

The first example is γ D-crystallin, whose aggregation is famous for provoking cataracts, a rare example of a disease caused by amorphous aggregates. It was believed that γ D-crystallin aggregates because the protein denatures (unfolds) upon aging, UV radiation, or sidechain oxidization. However,

solid-state NMR indicated that proteins aggregated under physiological conditions kept a well-ordered, native-like conformation, with no evidence of dynamic or static disorder (Boatz et al. 2017). To date, Boatz et al. mention that γ D-crystallin formed different aggregates at acidic pH.

In the second example, ^{15}N hetero single quantum correlation (HSQC) spectra showed that a low-solubility bovine pancreatic trypsin inhibitor (BPTI) variant, BPTI-22 (a BPTI variant containing 22 alanines), associated in a native-like structure (Kato et al. 2007). At a low concentration of 250 μM , BPTI-22 yielded a high-quality NMR spectrum (HSQC spectra) with sharp and well-dispersed peaks. On the other hand, at a concentration of 3.8 mM, the peaks broadened because of aggregation (Fig. 2) but appeared at the same chemical shift position as at 250 μM , strongly suggesting that the aggregated BPTI-22 retained a native-like structure.

These two examples show that amorphous aggregates do not result solely from the association of denatured proteins, as often imagined, but they can also result from the association of natively or near-natively folded proteins.

Hydrophobicity models and protein solubility

Hydrophobicity (and, thus, its counterpart, hydrophilicity) models are commonly used for rationalizing the solubility of proteins and other macromolecules (Kyte and Doolittle 1982). However, the original purpose of the hydrophobic/hydrophilic model was to describe an equilibrium between a molecule's (often amino acid) solubility in a polar (aqueous) and a non-polar environment (Nozaki and Tanford 1971). This model

Fig. 2 NMR spectrum (HSQC) of a low-solubility BPTI variant (BPTI-22) measured at 3.8 mM. The cross-peak shapes in the HSQC spectrum of BPTI-22 (3.8 mM) are broad, indicative of aggregation. However, the positions of several cross-peaks are the same as in the native spectrum, strongly suggesting that the overall BPTI fold remains intact. The assignments of cross-peaks with typical chemical shift positions are transferred from Kato et al. (2007), where the measurement and data analysis parameters are also given. The inset shows a photograph of the NMR sample

provides a good rationale for the structural stabilization of globular proteins by approximating its interior as non-polar and its exterior as polar (Kauzmann 1959; Tanford 1997), and, thus, for unfolded proteins having a higher aggregation tendency than natively folded proteins. On the other hand, the model is poorer at describing changes in protein solubility or amorphous aggregation tendency upon a single or a few mutations. This is especially true when the natively or near-natively folded proteins aggregate by keeping their hydrophobic residues buried in the protein's interior.

Amino acid's solubility parameters

The solubility of individual amino acids have been measured, independently from hydrophobic scales, in the early 1970s and standard values for amino acid solubility have been compiled (Lide (ed) 2009). However, such solubility measurements are subject to artifacts because they were carried out at extremely high peptide (or amino acid) concentrations, which can cause jellification, poor phase separation, sticky fluids, pH shifts, etc. (Kramer et al. 2012).

Two recent studies have measured the contribution of amino acids to the relative solubility change of a model protein using a systematic mutational analysis. In one of them, the solubility of ribonuclease Sa was examined by systematically

mutating Thr76, which is located on its molecular surface, to all of the 20 natural amino acids (Trevino et al. 2007). This mutational analysis is probably the first attempt to determine the contribution of amino acids to protein solubility in a systematic way.

In the second example, we determined the amino acid's contribution to a protein's solubility by measuring the solubility of a model protein, a simplified BPTI variant (Kuroda and Kim 2000), to which we fused 5-residue peptide tags composed of a single amino acid type (Kato et al. 2007; Islam et al. 2012). BPTI's solubility was defined as the fraction of proteins remaining in the supernatant after a 20-min centrifugation at 5 °C and in the presence of 1.3 M ammonium chloride, and this was carried out for ten representative amino acids (acidic, basic, etc.). As anticipated, positively charged residues significantly increased the solubility of the protein (despite the presence of ammonium chloride), at both pH 4.7 and 8.7, whereas the hydrophobic 5-Ile tag markedly reduced the solubility of BPTI. 5-Asp and Glu tags barely affected BPTI's solubility at pH 4.7, but increased solubility eight- to ten-fold at pH 8.7, attributable to the side chains ionization. Noteworthy, such pH dependence is not described in hydrophobicity scales. Overall, these two studies suggest that systematic mutational analysis measuring relative solubility changes could yield a solubility propensity scale, which might provide the basis for calculating the relative solubility of a polypeptide from its amino acid sequence alone.

Let us describe a recent molecular dynamics (MD) study that we carried out in order to decipher the molecular mechanisms of peptide and protein's solubility, and where we used multi-peptide systems containing 27 to 54 tetra-peptides and 3×10^4 water molecules (Kuroda et al. 2016). The clustering of tetra-peptides made of a single amino acid type representing all natural amino acids, except Gly and Cys, was monitored during 100-ns simulations. The results (fractions of tetra-peptides forming clusters) were, overall, in line with experimental amino acid's solubility as well as hydrophobicity scales (Table 1), and also demonstrated a concentration dependency for Ala₄ peptides. Additionally, the calculations indicated that the solubility of peptides containing charged amino acids originated from repulsive Coulomb energies, whereas "hydrophobic" amino acids, lacking repulsive electrostatic interaction, clustered predominantly through van der Waals interactions and accessorially through main-chain H-bonds. Finally, statistical analysis indicated that small clusters containing 2–4 peptides formed first and grew or merged into larger clusters, suggesting the existence of nuclei in amorphous aggregation (Fig. 3). Altogether, this benchmark type study was surprising in that the solubility of amino acids is fairly well reproduced using standard MD simulation methods, without a need to introduce artificial attractive or repulsive forces (Table 1).

Table 1 Averaged mean cluster size, solubility, and hydrophobicity scales

Amino acid residues	Parameters						
	MCS ^a	Individual amino acid solubility ^b (g/kg)	Hydropathy ^c	Hydrophilicity ^d (kcal/mol)	Hydrophobicity ^e (kcal/mol)	Contribution to protein solubility ^f (mg/mL)	Average number of monomers ^a
Ile	25.1	34.2	4.5	0.2	2.5	0.1	0.6
Leu	26.2	23.8	3.8	0.1	1.8	na	0.3
Val	22.6	88.0	4.2	0.4	1.5	na	1.2
Gln	24.4	42.0	− 3.5	11.8	− 0.3	4.7	1.2
Asn	19.0	25.1	− 3.5	12.1	− 0.2	15.4	3.9
Thr	12.0	90.6	− 0.7	7.3	0.4	na	5.4
Ser	6.2	250.0	− 0.8	7.5	− 0.3	4.6	9.0
Ala	3.4	166.9	1.8	0.5	0.5	na	13.7
Pro	2.6	1625	− 1.6	na	− 3.3	2.0	12.9
Asp	1.1	5.4	− 3.5	13.3	− 7.4	14.2	25.4
Glu	1.0	8.6	− 3.5	12.6	− 9.9	11.0	26.5
Arg	1.0	182.6	− 4.5	22.3	− 11.2	8.9	24.6
Lys	1.1	5.8	− 3.9	11.9	− 4.2	17.0	25.4
Phe	23.1	27.9	2.8	3.2	2.5	na	0.4
Tyr	21.8	0.5	− 1.3	8.5	2.3	na	1.4
Trp	24.8	13.2	− 0.9	8.3	3.4	na	0.7
His	25.1	43.5	− 3.2	12.6	0.5	0.9	0.5
Met	24.8	56.0	1.9	3.9	1.3	na	0.7

^a MCS values are the mean cluster sizes. MSCs and the number of monomers averaged over the last 10 ns of simulation are reported in Kuroda et al. (2016).

^b Lide ed, (2009)

^c Kyte and Doolittle (1982)

^d Wolfenden et al. (1981)

^e Nozaki and Tanford (1971)

^f Islam et al. (2012)

Time dependence of solubility and aggregation rates

Aggregation rates are an essential, but often overlooked, factor in biochemical and biophysical studies of amorphous

aggregation. Despite a few earlier reports (Nohara et al. 1999), a systematic approach for rationalizing protein's amorphous aggregation kinetics has yet to be developed. We recently expanded our aforementioned work by Islam et al. (2012) in order to analyze the contribution of

Fig. 3 All atom molecular dynamics simulations of amino acid solubility. The calculations were performed for more than 18 systems, each containing 27 tetra-peptides and about 30,000 water molecules. The tetra-peptides were made of a single amino acid type for all natural peptides, except Gly and Cys. Snapshots at 1, 5, 10, and 50 ns for tetra-peptides made of Ile, Asn, and Arg are shown. The peptides are shown using van der Waals representation. Water molecules are not shown. Details are given in Kuroda et al. (2016)

individual amino acids to the kinetics of protein aggregation (Khan et al. 2013). The major finding of the study was that, for each amino acid, in addition to a parameter describing its aggregation rate, at least two solubility parameters would be necessary for rationalizing the time dependence of protein solubility (Fig. 4). We coined these two parameters, the aggregation initiation concentration (*AIC*) and the long-term solubility (*LTS*).

Early in our study, we noticed that, besides the type of amino acids fused, BPTI's solubility depended on the equilibration time. For example, the amount of protein in the supernatant of the His-tagged variant, BPTI-C5H, reduced sharply from 10.5 mg/mL to 2.5 mg/mL with the extension of equilibration time from 20 min to 48 h, which was obviously indicative of a slow aggregation rate (Fig. 4a). In addition, we observed that the protein concentration in the supernatant or “apparent solubility” reduced faster for solutions

with a high total protein concentration, which was highly unexpected. For example, BPTI-C5H aggregated faster at a total protein concentration of 20.0 mg/mL compared to that of 15.0 mg/mL (Fig. 4a).

Thus, protein concentration in the supernatant of a saturated sample, measured after 20 min, did not persist, and we termed this solubility “transient solubility (*TS*)”. *TS* is not a fundamental parameter in defining a protein's solubility. Rather, it may be erroneously taken for a protein's solubility if one measures the amount of protein in the supernatant before the system reaches equilibrium. *TS* decreased with time only when the total protein concentration was higher than the *AIC*, and reached the *LTS* after 48 h, at which time we assumed that equilibrium was reached. With respect to the nucleation-extension or helical polymerization model (see next section), the *AIC* would correspond to the minimum concentration at which an aggregation nucleus can form, whereas the *LTS* (or actually the final equilibrium solubility) is the minimum concentration at which growth of the nucleus to large aggregates occurs. In many ways, both the *LTS* and the *AIC* could be considered as the “solubility” of a protein, but we believe that the *LTS* is close to the final equilibrium solubility and is an important parameter. Similarly, the *AIC* is also an important parameter because the protein concentration in the supernatant remains constant over the long term if the total protein concentration is lower than the *AIC*, where no nucleus can form.

The *AIC* and *LTS* are also valid for describing the aggregation of natural proteins, as shown for lysozyme, though it took longer to reach equilibrium (Fig. 4b). For a solution with a total protein concentration of 15.0 mg/mL, the lysozyme concentration was 6.9 mg/mL after 48 h and, finally, 5.8 mg/mL after 7 days of equilibration in 1.7 M ammonium sulfate and pH 4.7, which can be considered to have reached equilibrium. On the other hand, samples prepared at 6.5 mg/mL, which was the *AIC* under this condition, did not aggregate.

Furthermore, the slow rate of aggregation could provide a rationale for a peculiar observation that lysozyme's solubility was higher when the total concentration was higher (Trevino et al. 2008). We speculate that this bizarre observation might originate from the slow speed of precipitation and the fact that lysozyme's “solubility” was measured before the system had reached equilibrium. A similar situation is visible in Fig. 4b, where lysozyme dissolved at 15 mg/mL has an “apparent solubility” (or *TS*) of 8.3 mg/mL, as measured after a 60-min incubation. However, the apparent solubility is only 7.3 mg/mL when the total protein concentration is 10.0 mg/mL. This effect eventually disappeared at longer incubation time as the system moved toward equilibrium.

Fig. 4 Time dependence of the aggregation process. The concentrations of BPTI-C5H (a) and lysozyme (b) in the supernatant measured after centrifugation at $20,000 \times g$ for 20 min are shown on the vertical axis. BPTI-C5H samples were incubated at 25 °C for 20 min, 60 min, 6 h, 12 h, 24 h, and 48 h before centrifugation (a). Lysozyme samples were incubated for 20 min, 6 h, 12 h, 1 day, 2 days, and 7 days before centrifugation (b). The other experimental details are given in Khan et al. (2013). The total protein concentration is indicated on the horizontal axis. The aggregation initiation concentration (*AIC*) and long-term solubility (*LTS*) for lysozyme are indicated by arrows

The helical polymerization model applied to protein's amorphous aggregation

The HPM was initially developed to describe the formation of F-actin (Oosawa and Kasai 1962). The HPM provides a theoretical framework for systems that undergo an abrupt transition from a monomeric state to an oligomeric or colloidal state (Oosawa and Kasai 1962). Besides actin (Arisaka et al. 1975), HPM-derived models have been applied to various systems, such as tobacco mosaic virus (Klug and Durham 1972) and tubulin (Sutherland and Sturtevant 1976).

To date, thermodynamic models do not require structural information as discussed above. Before describing how the HPM formalism applies to protein amorphous aggregation, let us first discuss the more popular linear model. In the linear model (Fig. 5a), the equilibrium constant (Sutherland and Sturtevant 1976) for adding a single monomer to the aggregate is k_a . According to the linear model, the concentration (M_i) of aggregates containing i molecules is written in terms of the free monomer concentration in the supernatant m as:

$$M_i = k_a^{(i-1)} m^i \quad (1)$$

The amount of protein molecules (m_{Tot}) is then given by:

$$m_{\text{Tot}} = \sum_{i=1}^{\infty} i M_i = \sum_{i=1}^{\infty} i k_a^{(i-1)} m^i = \frac{m}{(1 - k_a m)^2} \quad (2)$$

The major problem in applying Eq. 2 to experimental observations is that the number of monomers in the supernatant rises gradually as m_{Tot} , the total number of protein molecules, increases to a limit value at infinite total concentration (Fig. 5a). Equation 2 could provide an explanation for solubility increasing along with total protein concentration (Trevino et al. 2008), but such an observation is anecdotal and counter-intuitive, and it is reasonably assigned to slow aggregation rates as discussed above. Thus, Eq. 2 does not yield a sharp transition with a concentration above which newly added molecules associate, i.e., a solubility value, which contradicts experimental observation.

The HPM overcomes this conundrum by introducing two equilibrium constants. It considers that the first molecules associate with a low association constant k_0 . Once the aggregate reaches a certain size (j), it assumes that each newly added molecule associates with a novel association constant k_b . For amorphous aggregation, it is reasonable to assume that $k_b < k_0$ because the interactions are non-specific, and the surface to which a molecule can bind is larger for an aggregate than for a monomeric molecule.

Fig. 5 Protein solubility simulated using the Oosawa–Kasai polymerization model. The total amount of protein (m_{Tot}) is shown on the horizontal axis and the concentration of monomeric proteins (m) in the supernatant is shown on the vertical axis. The curves in **a** and **b** were calculated by using a linear polymerization model with an association constant (k_a) of 0.2 (1/5) mM^{-1} . In **b**, the curves calculated for k_a with 1/6 mM^{-1} and 1/4 mM^{-1} are shown with dashed-dotted and dashed lines, respectively. The curves in **c** were calculated by using a helical polymerization model (HPM) with an association $k_b = 0.25$ (1/4) mM^{-1} and $k_0 = 0.01$ (1/100) mM^{-1} and assuming that monomers remain in the supernatant. The value of k_0 had no visible influence on the curves within a range of 0.1 M^{-1} –0.01 M^{-1} . Similarly, j had a minor influence on the curve's appearance, as shown here for $j = 2$ (dashed line), 10 (dashed-dotted line), and 25 (solid line). The curves essentially overlapped for j above 25

Thus, according to the HPM, the concentration M_i of aggregates containing i molecules is expressed in terms of the free monomer concentration (m) as:

$$M_i = ik_0^{(i-1)} m^i \text{ for } i < j, \text{ and } M_i = ik_0^{(j-1)} k_b^{(i-j+1)} m^i \text{ for } i \geq j \quad (3)$$

The total protein concentration m_{Tot} is:

$$m_{\text{Tot}} = \sum_{i=1}^{j-1} i k_0^{i-1} m^i + \sum_{i=j}^{\infty} i k_0^{j-1} k_b^{(i-j+1)} m^i \quad (4)$$

This value is valid at equilibrium and does not take into account aggregation rates. k_0 and k_b would roughly correspond, to the *AIC* and the *LTS*, respectively (Khan et al. 2013). Let us examine how Eq. 4 describes the free monomer concentration m as a function of total protein molecules m_{Tot} using the three parameters, k_0 , k_b , and j (Fig. 5b). First, k_b is experimentally determined by measuring the critical concentration. k_0 is assumed to be smaller than k_b but it cannot be determined precisely as the curve is strongly dependent on k_0/k_b . However, by computing m as a function of m_{Tot} , for $k_0/k_b = 0.1 \sim 10^{-4}$, one sees that k_0 has no experimentally detectable effect on the amount of monomers (data not shown). Similarly, j , which is the size of aggregates at which the association constant changes from k_0 to k_b , has little influence if it is larger than 10-mers (Fig. 5c). Thus, in equilibrium, protein amorphous aggregation is well described using the HPM with essentially one variable, k_b , the others having no or little experimentally detectable effect on the curve's appearance.

A direct consequence of the HPM and HPM-derived models is the existence of a seed or a nucleus in the formation of amorphous aggregates (Xue 2015). Both our MD simulation and observation of an *AIC* are in line with the existence of a nucleus in amorphous aggregation. However, we note that the nucleus is a “thermodynamic state” which are unlikely to contain a specific structure. Moreover, the number of molecules in the seed can vary within the fitting error of j , which is large for amorphous aggregation ($j \sim 10$).

Scope and limitation of the HPM for describing amorphous aggregation

The HPM provides a good description for amorphous aggregation. However, we note that, unlike previous systems described using the HPM, there are most probably no specific structures in amorphous aggregates (including in the seeds), and the parameters discussed in the previous section are obviously averaged values and their structural interpretation should be carried out judiciously.

Aggregation in the denatured state is readily incorporated in the HPM by considering:

This aggregation scheme yields $\tilde{k}_0 = k_0 k_{\text{denaturation}}$, where $\tilde{k}_0$ is the new association constant and $k_{\text{denaturation}}$ is the equilibrium constant between the native and denatured state of the monomer. Similarly, one could presumably include the time dependence of protein aggregation using an HPM-derived formalism as previously shown by Oosawa and Asakura (1975) or more sophisticated models (Hall et al. 2015).

We also note that the present model holds at low protein concentration, and exactly only for infinitely dilute systems, which is somewhat contradictory in the case of amorphous aggregation, as it necessarily occurs at fairly high concentration. One calculates that, at 1 mM (about 6.5 mg/mL), a BPTI molecule is surrounded by a sphere of solvent molecules with a radius approximately equal to six times its own radius. The solution is, thus, dense and one can expect that molecular crowding may start to influence the thermodynamics at concentrations where solubility experiments are usually carried out (Zimmerman and Minton 1993; Hall and Minton 2002).

Conclusion

The main point of our discussion was to present thermodynamic and mutational analyses of protein solubility with respect to amorphous aggregation. It appears that a thermodynamic description of protein amorphous aggregation is indeed reasonable within the mathematical framework provided by the helical polymerization model (HPM). Both the HPM and molecular dynamics (MD) simulation suggest the existence of a nucleus or seed during the formation of aggregates, which is in line with the experimental observation of an aggregation initiation concentration (*AIC*) and a long-term solubility (*LTS*) in both bovine pancreatic trypsin inhibitor (BPTI) variants and lysozyme. Further biophysical investigation of protein solubility and amorphous aggregation are worthy of being carried out, as molecular crowding and aggregation kinetics remain to be taken fully into account.

Acknowledgements This article is dedicated to the 70th birthday anniversary of Professor Fumio Arisaka. I thank laboratory members, especially Drs. Atsushi Kato, Mohammad M. Islam, and Alam M. Khan for discussion and experimental data, and Ms. Patricia McGahan for the English grammar editing. This work was supported by a Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for Scientific Research (KAKENHI-21300110).

Compliance with ethical standards

Conflict of interest Yutaka Kuroda declares that he has no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by the author.

References

- Arakawa T, Tsumoto K (2003) The effects of arginine on refolding of aggregated proteins: not facilitate refolding, but suppress aggregation. *Biochem Biophys Res Commun* 304(1):148–152
- Arisaka F, Noda H, Maruyama K (1975) Kinetic analysis of the polymerization process of actin. *Biochim Biophys Acta* 400(2):263–274
- Baldwin RL (2012) Gas–liquid transfer data used to analyze hydrophobic hydration and find the nature of the Kauzmann–Tanford hydrophobic factor. *Proc Natl Acad Sci U S A* 109(19):7310–7313
- Boatz JC, Whitley MJ, Li M, Gronenborn AM, van der Wel PCA (2017) Cataract-associated P23T gammaD-crystallin retains a native-like fold in amorphous-looking aggregates formed at physiological pH. *Nat Commun* 8:15137
- Hall D, Minton AP (2002) Effects of inert volume-excluding macromolecules on protein fiber formation. I. Equilibrium models. *Biophys Chem* 98(1–2):93–104
- Hall D, Kardos J, Edskes H, Carver JA, Goto Y (2015) A multi-pathway perspective on protein aggregation: implications for control of the rate and extent of amyloid formation. *FEBS Lett* 589(6):672–679
- Hamada H, Shiraki K (2007) L-argininamide improves the refolding more effectively than L-arginine. *J Biotechnol* 130(2):153–160
- Islam MM, Khan MA, Kuroda Y (2012) Analysis of amino acid contributions to protein solubility using short peptide tags fused to a simplified BPTI variant. *Biochim Biophys Acta* 1824(10):1144–1150
- Iwura T, Fukuda J, Yamazaki K, Arisaka F (2014) Conformational stability, reversibility and heat-induced aggregation of alpha-1-acid glycoprotein. *J Biochem* 156(6):345–352
- Kamal MZ, Kumar V, Satyamurthi K, Das KK, Rao NM (2016) Mutational probing of protein aggregates to design aggregation-resistant proteins. *FEBS Open Biol* 6(2):126–134
- Kato A, Maki K, Ebina T, Kuwajima K, Soda K, Kuroda Y (2007) Mutational analysis of protein solubility enhancement using short peptide tags. *Biopolymers* 85(1):12–18
- Kauzmann W (1959) Some factors in the interpretation of protein denaturation. *Adv Protein Chem* 14:1–63
- Khan MA, Islam MM, Kuroda Y (2013) Analysis of protein aggregation kinetics using short amino acid peptide tags. *Biochim Biophys Acta* 1834(10):2107–2115
- Klug A, Durham ACH (1972) The disk of TMV protein and its relation to the helical and other modes of aggregation. *Cold Spring Harb Symp Quant Biol* 36:449–460
- Kramer RM, Shende VR, Motl N, Pace CN, Scholtz JM (2012) Toward a molecular understanding of protein solubility: increased negative surface charge correlates with increased solubility. *Biophys J* 102(8):1907–1915
- Kuroda Y, Kim PS (2000) Folding of bovine pancreatic trypsin inhibitor (BPTI) variants in which almost half the residues are alanine. *J Mol Biol* 298(3):493–501
- Kuroda Y, Suenaga A, Sato Y, Kosuda S, Taiji M (2016) All-atom molecular dynamics analysis of multi-peptide systems reproduces peptide solubility in line with experimental observations. *Sci Rep* 6:19479
- Kyte J, Doolittle RF (1982) A simple method for displaying the hydrophobic character of a protein. *J Mol Biol* 157(1):105–132
- Lide DR (ed) (2009) CRC handbook of chemistry and physics. 90th Edition, Section 7–1.
- Nohara D, Mizutani A, Sakai T (1999) Kinetic study on thermal denaturation of hen egg-white lysozyme involving precipitation. *J Biosci Bioeng* 87(2):199–205
- Nozaki Y, Tanford C (1971) The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions. Establishment of a hydrophobicity scale. *J Biol Chem* 246(7):2211–2217
- Oosawa F, Asakura S (1975) Thermodynamics of the polymerization of protein. Academic Press, New York
- Oosawa F, Kasai M (1962) A theory of linear and helical aggregations of macromolecules. *J Mol Biol* 4:10–21
- Schön A, Clarkson BR, Jaime M, Freire E (2017) Temperature stability of proteins: analysis of irreversible denaturation using isothermal calorimetry. *Proteins* 85(11):2009–2016
- Shiraki K, Kudou M, Fujiwara S, Imanaka T, Takagi M (2002) Biophysical effect of amino acids on the prevention of protein aggregation. *J Biochem* 132(4):591–595
- Sutherland JW, Sturtevant JM (1976) Calorimetric studies of the in vitro polymerization of brain tubulin. *Proc Natl Acad Sci U S A* 73(10):3565–3569
- Tanford C (1997) How protein chemists learned about the hydrophobic factor. *Protein Sci* 6(6):1358–1366
- Trevino SR, Scholtz JM, Pace CN (2007) Amino acid contribution to protein solubility: Asp, Glu, and Ser contribute more favorably than the other hydrophilic amino acids in RNase Sa. *J Mol Biol* 366(2):449–460
- Trevino SR, Scholtz JM, Pace CN (2008) Measuring and increasing protein solubility. *J Pharm Sci* 97(10):4155–4166
- Wolfenden R, Andersson L, Cullis PM, Southgate CCB (1981) Affinities of amino acid side chains for solvent water. *Biochemistry* 20:849–855
- Xue WF (2015) Nucleation: the birth of a new protein phase. *Biophys J* 109(10):1999–2000
- Zimmerman SB, Minton AP (1993) Macromolecular crowding: biochemical, biophysical, and physiological consequences. *Annu Rev Biophys Biomol Struct* 22:27–65