Biochimica et Biophysica Acta 1834 (2013) 2107–2115

Contents lists available at SciVerse ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbapap

Analysis of protein aggregation kinetics using short amino acid peptide tags

Monsur Alam Khan, Mohammad Monirul Islam [1], Yutaka Kuroda *

Department of Biotechnology and Life Science, Tokyo University of Agriculture and Technology, 2-24-16 Nakamachi, Koganei-shi, Tokyo 184-8588, Japan

ARTICLE INFO

Article history:
Received 18 April 2013
Received in revised form 15 June 2013
Accepted 18 June 2013
Available online 28 June 2013

Keywords:
Poly amino acid peptide tag
Protein solubility
Aggregation kinetics
Equilibration time

ABSTRACT

Understanding protein solubility, and consequently aggregation, is an important issue both from an academic and a biotechnological application viewpoints. Here we report the effects of 10 representative amino acids on the aggregation kinetics of proteins. The effects were determined by measuring the solubility of a simplified bovine pancreatic trypsin inhibitor (BPTI) variant, to which short artificial tags containing the amino acid of interest were added at its C-terminus. We determined the solubility of the tagged variants as a function of equilibration time (20 min to 48 h) and total protein concentration ranging from 0.10 mg/ml to 25.0 mg/ml. We observed, as anticipated, that proteins precipitated when the total protein concentration exceeded a critical value. However, when the total protein concentration was further increased, the apparent solubility reached a concentration above the critical value, and slowly decreased to a value under the critical concentration upon increasing the equilibration period. We rationalized these observations by identifying three different solubility values, the "transient solubility (*TS*)", the "aggregation initiation concentration (*AIC*)" and the "long-term solubility (*LS*)". *AIC* and *LS* are parameters determined essentially by the amino acid types composing the tags and could be considered as an amino acid's intrinsic property. On the other hand, *TS* is an apparent solubility that is measured after some (20 min in our case) equilibration time and is often considered as the "solubility" of the protein. Similar aggregation kinetic patterns were observed with natural proteins, indicating the generality of the observations made using our model protein.

1. Introduction

Protein aggregation is attracting much attention in physico-chemical studies [1]. Amyloidogenic protein aggregation, which is associated with several neurodegenerative diseases, is the focus of intensive research [2–5]. On the other hand, non-amyloidogenic or amorphous protein aggregation is much less characterized, but it is a major concern in biotechnology and pharmaceutical industries, especially in the production and storage of therapeutic proteins [6]. This is because the aggregation tendency (or low solubility) of therapeutic proteins, not only reduces their production and therapeutic efficacy, but may also increase the risk of immunogenetic reactions [7].

Though several aspects of amorphous and amyloidogenic protein aggregation might be related, few studies have specifically addressed the mechanisms of amorphous aggregation, which are barely understood. Amorphous protein aggregation can be either reversible or virtually irreversible if aggregates are formed by partially or fully unfolded or chemically altered proteins [8]. For example, insulin forms reversible aggregates [9], whereas, the monoclonal antibody, IgG2, aggregates irreversibly through chemical modification [10]. In addition, a protein might aggregate simultaneously through different mechanisms, as observed for an interleukin-1 receptor antagonist [11].

Except for general trends, the physico-chemical aspects of amorphous protein aggregation are not well understood [12,13]. Generally, hydrophobic proteins are aggregation prone [14] whereas proteins containing many charged residues on their surfaces are highly soluble [15,16]. Some studies also reported that large proteins are likely to be more aggregation prone than small ones [14], but this observation might simply reflect their tendency to partially unfold. Additionally, aggregation is thought to occur through a nucleation extension model similar to the helical polymerization model used to describe F-actin polymerization rather than a simple linear extension model [17]. However, the difficulty to isolate nuclei of amorphous aggregates and distinguish them unambiguously from small aggregates [18] has hampered the biophysical analysis of amorphous aggregation.

Mutational analyses are expected to help rationalizing the impact of amino acid's physico-chemical properties on protein aggregation tendency [19,20]. In one such example, the solubility of Ribonuclease Sa was examined by systematically mutating Thr76, which is located

Abbreviations: BPTI, bovine pancreatic trypsin inhibitor; BPTI-19, simplified single-disulfide-bonded BPTI variant containing 19 alanines; Protein aggregation, in vitro amorphous protein aggregation; *AIC*, aggregation initiation concentration; *TS*, transient solubility; *LS*, long-term solubility; *R*, correlation coefficient

* Corresponding author. Tel./fax: +81 42 388 7794.
E-mail address: ykuroda@cc.tuat.ac.jp (Y. Kuroda).
[1] Present address: Department of Biochemistry and Molecular Biology, University of Chittagong, Chittagong-4331, Bangladesh.

1570-9639/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.bbapap.2013.06.013

on its molecular surface, to all of the 20 natural amino acids [21]. In another example, Ankyrin repeat protein's solubility was analyzed by substituting solvent exposed hydrophobic Leu with positively charged Arg, eventually yielding highly soluble variants [22]. However, insights from mutational analysis remain relatively limited, as the effects of the local physico-chemical environment on solubility are difficult to disentangle from genuine amino acid solubility properties [23].

Aggregation kinetics is a determining factor for processing proteins in the biotechnological and pharmaceutical industries [24], but it is often overlooked in biochemical and biophysical studies, partly because aggregation can be very slow to occur. Although some studies have stressed its importance in rationalizing protein solubility and aggregation [25], few mutational analyses of aggregation kinetics have been reported [26]. In another example, the addition of an acidic tail of synuclein (ATS) fused to hGH, G-CSF and leptin derivatives significantly reduced the aggregation rates of the host protein [27]. However, a systematic approach for rationalizing protein aggregation kinetics remains to be developed.

Recently, we used a simplified BPTI variant [28] as a model protein and reported short poly-amino acid peptide tags attached to its termini that can manipulate protein solubility in an essentially context-independent manner [29], and we used them to measure the contribution of individual amino acids on protein solubility for ten different types of amino acids [30]. In this work, we expanded this strategy to analyze how the amino acid types contribute to the kinetics of protein aggregation. The major finding of the present study is that, for each amino acid, at least two solubility parameters (which we coined the *AIC* and the *LS*) and presumably one or more aggregation rates (that we did not determine here) would be necessary for rationalizing protein solubility from its amino acid component as a function of time.

2. Materials and methods

2.1. Mutant design, expression and purification

Solubility patterns were assessed with BPTI variants tagged with poly-amino acids tags for ten different amino acid types. The poly-amino acid tagged variants were named according to the number and types of amino acids included. For example, C5S denotes a BPTI variant with five Ser residues added at the C-Terminus. Two Gly were added as a spacer between BPTI and the poly-amino acid tag.

All variants were prepared from a pMMHa vector encoding the BPTI-19 variant [28], and by introducing a DNA sequence encoding the poly-amino acid tag, to the C-terminus of BPTI-19 using QuikChange site directed mutagenesis (Stratagene). The sequences were confirmed by DNA sequencing (ABI PRISM 3130xl Genetic Analyzer). Expression of all poly-amino acid tagged variants were performed in an *Escherichia coli* JM109(DE3) pLysS cell line and purified by pH-precipitation, and reverse phase HPLC as previously described [19]. Protein identities were confirmed by MALDI mass spectroscopy (AB SCIEX TOF/TOF™ 5800), and the purified proteins were preserved at −30 °C as lyophilized powder.

2.2. Protein solubility measurement

We measured protein solubility at 25 °C, pH 4.7 (50 mM acetate buffer) and pH 7.7 (50 mM Tris–HCl buffer) by using the amorphous precipitation method with 1.3 M ammonium sulfate (1.7 M for Lysozyme) [30,31]. Ammonium sulfate was necessary to obtain reliable and reproducible data as we discussed in our previous paper [30]. Protein solubility was determined as the maximum protein concentration in the supernatant of protein solutions, as described previously [21,29]. Sample solutions were equilibrated for different times (20 min, 6 h, 12 h, 24 h and 48 h). After equilibration, the samples were centrifuged at 20,000 rpm for 20 min at 25 °C and the protein concentration in the supernatant was determined by measuring the absorbance at 280 nm using a Nano Drop 2000C spectrometer (Thermo Scientific).

In order to reliably determine protein aggregation kinetics, it was important to respect constant equilibration time and total protein concentration. We thus prepared separate solutions (50 mM buffer, 2.6 M ammonium sulfate in 50 mM buffer and a protein stock solution

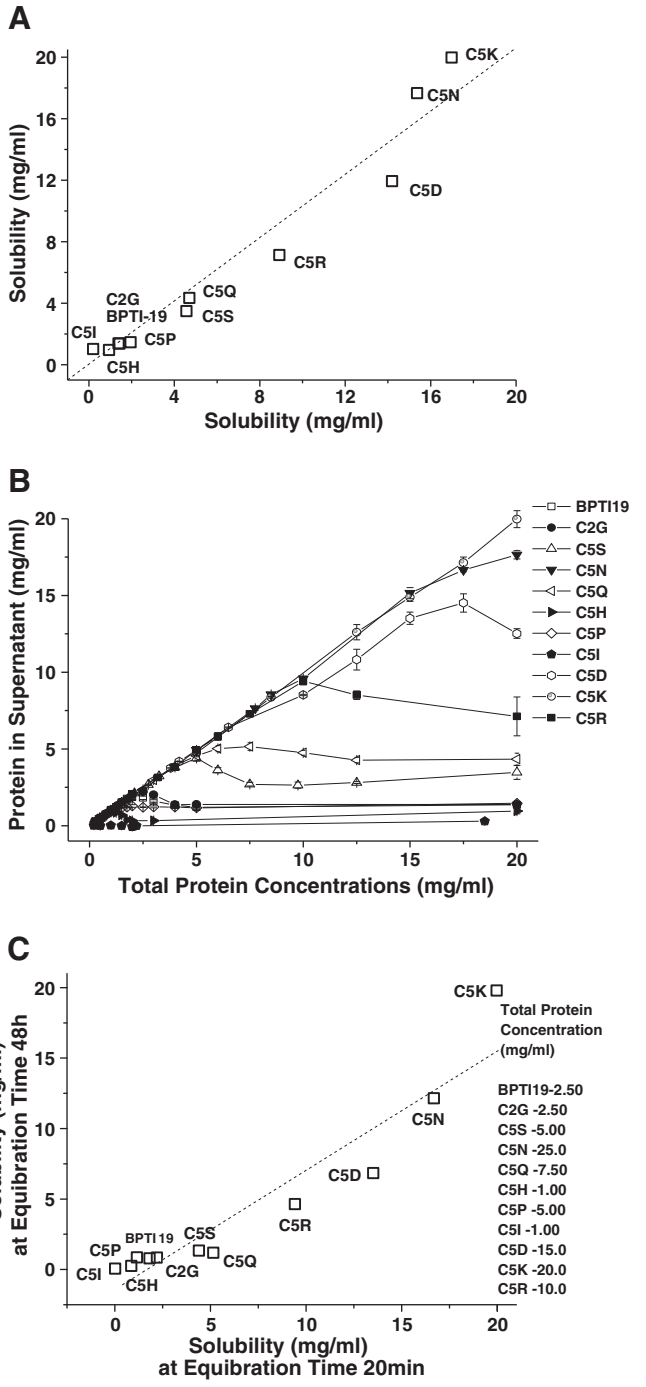

Fig. 1. Solubility values of the poly-amino acid tagged BPTI variants under different conditions at pH 7.7, (A) comparison with previously published values, X-axis represents the reference value, [30], and the Y-axis represents values in the present study, (B) solubility values with different total protein concentration. For figures (A) and (B), the equilibration period was 20 min. (C) Solubility values for different equilibration time (20 min vs. 48 h). The total protein concentrations are indicated on the right side of the figure.