JOURNAL OF PROTEOMICS 118 (2015) 21–38

Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jprot

A feature analysis of lower solubility proteins in three eukaryotic systems☆

Razvan F. Albu[a,1], Gerard T. Chan[a,1], Mang Zhu[a,1], Eric T.C. Wong[a], Farnaz Taghizadeh[b], Xiaoke Hu[b], Arya E. Mehran[b], James D. Johnson[b], Jörg Gsponer[a], Thibault Mayor[a,*]

[a]Department of Biochemistry and Molecular Biology and Centre for High-Throughput Biology, University of British Columbia, Vancouver, BC, V6T 1Z4 Canada

[b]Department of Cellular and Physiological Sciences, University of British Columbia, Vancouver, BC, V6T 1Z3 Canada

ARTICLE INFO

Available online 25 October 2014

Keywords:
Proteomics
Mass spectrometry
Computational biology
Proteostasis
Low solubility proteins
Functional aggregation

ABSTRACT

Because misfolded and damaged proteins can form potentially harmful aggregates, all living organisms have evolved a wide variety of quality control mechanisms. However, the timely clearance of aggregation-prone species may not always be achieved, potentially leading to the accumulation of low solubility proteins. At the same time, promiscuity, which can be a driving force for aggregation, is also important to the functionality of certain proteins which have a large number of interaction partners. Considerable efforts have been made towards characterizing why some proteins appear to be more aggregation-prone than others. In this study, we analyze the features of proteins which precipitate following centrifugation in unstressed yeast cells, human SH-SY5Y cells and mouse brain tissue. By normalizing for protein abundance, we devised an approach whereby lower solubility proteins are reliably identified. Our findings indicate that these tend to be longer, low abundance proteins, which contain fewer hydrophobic amino acids. Furthermore, low solubility proteins also contain more low complexity and disordered regions. Overall, we observed an increase in features that link low solubility proteins to functional aggregates. Our results indicate that lower solubility proteins from three biologically distinct model systems share several common traits, shedding light on potentially universal solubility determinants.

Biological significance

We set up a novel approach to identify lower solubility proteins in unstressed cells by comparing precipitated proteins with those that remain soluble after centrifugation. By analyzing three eukaryotic model systems in parallel, we were able to identify traits which cross the species barrier, as well as species-specific characteristics. Notably, our analyses revealed a number of primary and secondary structural features that set apart lower solubility proteins, a number of which connected them to a greater potential for promiscuity. This article is part of a Special Issue entitled: Protein dynamics in health and disease. Guest Editors: Pierre Thibault and Anne-Claude Gingras.

☆ This article is part of a Special Issue entitled: Protein dynamics in health and disease. Guest Editors: Pierre Thibault and Anne-Claude Gingras.

* *Corresponding author.*
E-mail address: mayor@mail.ubc.ca (T. Mayor).

[1] Equal contribution.

http://dx.doi.org/10.1016/j.jprot.2014.10.011
1874-3919/

1. Introduction

The intracellular space is extremely crowded, containing a vast number of proteins that populate the cytosol and cellular compartments. This is potentially problematic, since proteins, by their nature, have a high propensity for aggregation, especially if damaged or improperly folded. The proper assembly of a natively folded protein is not trivial, and even after the native conformation has been attained, unfolding and aggregation can still occur through instability incurred by translational errors or mutations, or be caused by a wide variety of possible stressors, such as high temperature, high pressure and low pH [1–3]. Protein aggregation can be harmful to the cell, both through the potential loss-of-function due to the lack of available functional protein [4,5], as well as through toxic gain-of-function, such as the formation of β-sheet-rich aggregates [6]. To combat the accumulation of misfolded and aggregation-prone species, a complex protein quality control network is in place to help mediate their refolding or, failing that, their proteolysis. For instance, the ubiquitin proteasome system plays a major role in clearing misfolded proteins to which poly-ubiquitin chains are covalently attached [7–9]. If the ubiquitin proteasome system fails, or is overwhelmed (for example, as a result of chemical inhibition of the proteasome), misfolded proteins have been shown to accumulate in the cytosol, sometimes with devastating effects [10–13]. To compensate, cells often attempt to sequester these proteins to specific quality control compartments [14], such as aggresomes [15,16], JUxtaNuclear Quality control (JUNQ), Insoluble PrOtein Deposit (IPOD) compartments [17] or Q-bodies [18]. The cells may then attempt to clear these structures either by macroautophagy [16,19], or by asymmetrically partitioning them during cell division [20–23].

Despite the battery of quality control strategies designed to clear away or limit the harm caused by damaged and misfolded species, protein aggregation remains the hallmark of more than 40 diseases in humans alone [6,24]. While diseases such as amyloidosis can affect a wide range of tissues, or, indeed, have systemic effects [6], neurodegenerative diseases in particular have been the focus of considerable research efforts, owing to the devastating nature of the symptoms associated with protein aggregation occurring within neural tissue. Though the aggregates found in most neurodegenerative diseases are typically characterized by one major component, such as α-synuclein in the case of Parkinson's disease (PD) [25,26] and amyloid-beta fibrils in Alzheimer's disease (AD) [27,28], more recent proteomics studies have revealed a highly diverse composition of the disease-related inclusions [29,30]. Perhaps unsurprisingly, aggregates from a wide variety of diseases tend to contain ubiquitin [31,32], indicating that the proteins which form them have been flagged for degradation, but have somehow escaped cellular quality control pathways.

With the exception of spongiform encephalopathies, which are caused by prions [33], neurodegenerative diseases are not thought to be transmissible between individuals. However, recent evidence has shown that neurodegenerative disease-associated proteins can behave in a manner very similar to prions, in that small amounts of disease-associated aggregates can nucleate the aggregation of endogenous proteins, and can also spread to other unaffected cells [34], as seen for α-synuclein [35,36], tau [37–39] and huntingtin [40]. It has been postulated that any protein can, in theory, form aggregates [41,42], though some seem to have a higher propensity to do so. This predisposition is correlated to inherent properties, such as high hydrophobicity, low charge and the propensity to form β-sheets [43], which define particularly aggregation-prone sequences of amino acids [44–46]. For instance, a stretch of six residues from an amylogenic domain has been shown to induce aggregation when introduced into the non-amylogenic protein α-spectrin [47].

Several model systems have been employed to the study of protein aggregation in the context of disease [29,30] and in connection to prions and amyloid-like aggregates [4,48], aging [49,50], heat shock [51–53] or when the cells are exposed to proteasome inhibitors [54]. However, subjecting the cells to stress would naturally favor the discovery of proteins particular to those experimental conditions. By contrast, we were interested in studying the landscape of protein aggregation in unstressed cells. We identified lower solubility proteins by quantitative mass spectrometry and characterized them by computational methods in three model systems: yeast cells, human neuronal tissue culture cells and mouse brain tissue. Our results revealed a number of traits that set apart lower solubility proteins in all three eukaryotic systems, and which potentially connect them to the formation of functional aggregates or macro-molecular assemblies.

2. Materials and methods

2.1. Chemical reagents

All chemicals used in this project were of analytical grade or higher, and were supplied by Sigma-Aldrich, unless otherwise specified. Tris-base, NaCl and acetonitrile (HPLC grade) were supplied by Fisher. Protease inhibitor cocktail (PIC), PMSF and Endoproteinase LysC were supplied by Roche. Trypsin (sequencing grade) was purchased from Promega. $(NH_4)HCO_2$ (ammonium formate) used with high pH reversed phase offline fractionation (see below) was obtained by mixing ammonium hydroxide (final concentration—0.7% (vol/vol)) with formic acid (final concentration—0.45% (vol/vol)). Both chemicals were purchased from Fisher. SILAC amino acids were supplied by Cambridge Isotopes. For formaldehyde labeling, light formaldehyde was supplied by Sigma, deuterated formaldehyde by Cambridge Isotopes. Cyanoborohydride was purchased from Acros Organics.

2.2. Biological sample preparation

The SILAC yeast strains (YTM1173) are in the S288C background (MATa, his3Δ1, leu2Δ0, ura3Δ0, MET15, arg4Δ::KanMX6, lys2Δ0) and were grown in YNB media (with 2% (wt/vol) dextrose and 0.002% (wt/vol) adenine, uracil, tryptophan, histidine, methionine and leucine) supplemented with either light (Lys0, Arg0) or heavy (Lys4, Arg6) amino acids, or, for triplex analysis, light (Lys0, Arg0), medium (Lys4, Arg6) and heavy (Lys8, Arg10) (to a working concentration of 30 mg/L for Lys and 20 mg/L for Arg) for at least 7 generations (>99% labeling efficiency) at 25 °C. Cultures were grown to mid log