JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

COMMUNICATION

pubs.acs.org/JACS

Metastability of Native Proteins and the Phenomenon of Amyloid Formation

Andrew J. Baldwin,[†] Tuomas P. J. Knowles,[†] Gian Gaetano Tartaglia,[†] Anthony W. Fitzpatrick,[†] Glyn L. Devlin,[†] Sarah Lucy Shammas,[†] Christopher A. Waudby,[†] Maria F. Mossuto,[†] Sarah Meehan,[†] Sally L. Gras,[‡] John Christodoulou,[#] Spencer J. Anthony-Cahill,[§] Paul D. Barker,[†] Michele Vendruscolo,[†] and Christopher M. Dobson*[,†]

[†]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, U.K.

[‡]Department of Chemical and Biomolecular Engineering, University of Melbourne, Parkville, Victoria 3010, Australia

[#]Department of Structural and Molecular Biology, University College London, Darwin Building, Gower Street, London WC1E 6BT, U.K.

[§]Department of Chemistry, Western Washington University, Bellingham, Washington 98225-9150, United States

Supporting Information

ABSTRACT: An experimental determination of the thermodynamic stabilities of a series of amyloid fibrils reveals that this structural form is likely to be the most stable one that protein molecules can adopt even under physiological conditions. This result challenges the conventional assumption that functional forms of proteins correspond to the global minima in their free energy surfaces and suggests that living systems are conformationally as well as chemically metastable.

Although amyloid structures were initially identified in connection with a variety of pathological conditions such as Alzheimer's disease, Creutzfeldt-Jacob disease, and type II diabetes,[1−3] many other protein molecules that differ substantially in both amino acid sequence and length have recently been shown to be able to form amyloid fibrils *in vitro*.[2,4] As a result it has been proposed that the ability to form assemblies of this type is a generic and intrinsic property of polypeptide molecules.[4]

A detailed and quantitative knowledge of the factors that stabilize the structures of amyloid fibrils would not only provide great insight into the nature and evolution of biologically active protein molecules but also contribute very significantly to the targeted design of therapeutic agents for amyloid diseases, as well as hasten the development of nanomaterials with core components based on amyloid fibrils.[5] However, despite the great progress that has been made recently, particularly through the use of solid-state NMR spectroscopy[6,7] and cryo-electron microscopy,[8] a detailed and quantitative understanding of the relative stabilities of such species has remained elusive. In the present study, we address this issue by defining the difference in free energy between the monomeric and the fibrillar forms of a series of polypeptide systems (Table 1), ranging from short peptides (e.g., GNNQQNY, $TTR_{(105-115)}$) and model systems (e.g., the PI3-SH3 domain) to proteins or protein fragments associated with human diseases[2] (e.g., α-synuclein, insulin, lysozyme, the amyloid β peptide Aβ(1−40), β2-microglobulin), selected to represent a variety of different characteristics in sequence and structure of the monomeric state.

In order to gain insight into the nature of the free energy surfaces defining the different states of specific proteins, we performed direct *in vitro* measurements (Supporting Information S1) of the free energy difference, ΔG_{el}, between a protein molecule in solution and incorporated into a fibril. When studying fibrils formed from a given protein at the end of the growth reaction,[9] the concentration of soluble monomers has been found to tend toward a fixed steady-state value, M_S, for a wide range of different total protein concentrations, M_T.[10−13]

This steady-state concentration is approached both when additional free monomer is added to pre-equilibrated samples and when residual free monomers are separated from the aggregates by ultracentrifugation.[10,11] By measuring the values of M_S and M_T at equilibrium (Figure 1), it is then possible in principle to determine experimentally the change in free energy, ΔG_{el}, associated with fibril elongation (Figure 2A). We adopted a strategy analogous to that employed for measuring the free energies of folding of the native states of proteins through their denaturation, although in contrast to the folding of many small proteins, amyloid growth is not a two-state process and is more accurately described as a linear polymerization process[14,15] (Supporting Information S2). In this strategy, therefore, the overall aggregation reaction is taken to consist of a sequence of reversible additions of monomers M to aggregates F of size i, governed by the equilibrium constant K such that $M + F_i \leftrightarrow F_{i+1}$ and $[F_{i+1}] = K[F_i][M]/c^0$, where c^0 is the standard concentration, 1 mol/L. In the presence of chemical denaturants, the free energy of elongation in the fibrillar state relative to the free energy of the soluble state can be written as $\Delta G_{el} = m[D] + \Delta G^0_{el}$, where m is a cooperativity coefficient[16] and ΔG^0_{el} is the free energy difference in the absence of denaturants. This linear polymerization picture yields the fraction of soluble protein M_S over the total protein concentration M_T as[11]

$$\frac{M_S}{M_T} = \frac{M_T K/c^0 + 1/2 - \sqrt{M_T K/c^0 + 1/4}}{M_T^2 K^2/(c^0)^2}$$

Received: February 25, 2011
Published: June 08, 2011

Table 1. Thermodynamic Stabilities of Amyloid Forms of the Peptides and Proteins Analyzed in This Work (See Supporting Information S6)[a]

system	abbreviation	no. of residues	ΔG^0_{el} (kJ mol^{-1})	$\Delta G^0_{el}/N$ (kJ mol^{-1})	pH
GNNQQNY	GNN	7	-13.7 ± 3.0	-2.0	7
$TTR_{(105-115)}$	TTR	11	-23.3 ± 0.5	-2.12	2
$TTR_{(105\text{-}115)}$RGD	TTRRGD	17	-23.2 ± 3.0	-1.36	2
Aβ(1−40)	AB(1−40)	40	-46.7^b	-1.17	7
Aβ(1−40)F20P	AB(1−40)F20P	40	-35.0^b	-0.88	7
bovine insulin	Ins2	51	-54.5 ± 1.4	-1.1	2
bovine insulin	Ins7	51	-37.0 ± 2.0	-0.73	7
bovine PI3-SH3	SH3	86	-38 ± 11	-0.44	2
human β2-microglobulin	B2M	119	-41.4^c	-0.34	7
human lysozyme	Lys2	130	-65.0 ± 5.5	-0.50	2
human lysozyme	Lys7	130	-40.0 ± 2.0	-0.31	7
human α-synuclein	Asyn	140	-33.0 ± 2.0	-0.24	7
human αB-crystallin	ABC	174	-28.5 ± 2.0	-0.16	7

[a] The experimentally determined elongation free energies of the fibrils measured here, or taken from the literature, are listed along with the number of residues in the polypeptide chain from which they were formed, together with the solution conditions where their stabilities were measured. The mean uncertainty in the free energies estimated both from multiple measurements and the scatter of the data about the best fit was 5%. [b] Ref 13. [c] Ref 15.

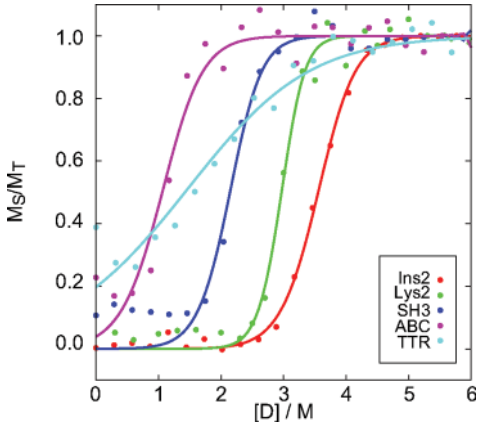

Figure 1. Variation of the fraction of soluble protein, M_S/M_T, in equilibrated fibril samples as a function of denaturant concentration. The corresponding fraction of insoluble aggregates is given by $(M_T-M_S)/M_T$. These data are fitted to the linear polymerization model[10,11] (lines) to obtain an estimate of ΔG^0_{el} (Table 1). The denaturant used was GdSCN in all cases apart from that of TTR, where high concentrations of GdHCl were sufficient to dissociate fully the fibrils.

where $K = \exp(-(\Delta G^0_{el} + m[D])/RT)$ is the equilibrium constant, R the gas constant, and T the temperature. Hence, the measurement of the population of free monomer M_S at equilibrium, in the presence of known concentrations of chemical denaturants, is a direct measure of the free energy of elongation of the fibrils ΔG_{el} (Figure 2A and Table 1).

A plot of the experimentally measured fibril energies per residue as a function of residue number (Figure 2B) reveals that the thermodynamic stability of the fibrillar relative to the soluble state depends primarily on sequence-independent characteristics of the polypeptide chains involved. This finding is in marked contrast to the free energies of native folded states relative to unfolded chains (Figure 2A), which are highly dependent on their specific amino acid sequences. Furthermore, we found that the fibril free energies can be described well by simple subextensive power laws of the form

$$\frac{G^0_{el}}{N} = \varepsilon_0 + \varepsilon_1 N^{\gamma}$$

with the exponent $-1 < \gamma < 0$. Such behavior fundamentally reflects the fact that the connectivity of the chain does not allow an independent search of the optimal interactions within the fibril for each amino acid. For example, shorter sequences will exhibit less packing constraints than longer ones; indeed, fibrils formed from short sequences such as GNNQQNY and $TTR_{(105-115)}$, where the entire chain can be incorporated into the cross-β core structure as each molecule forms a single strand, can be considered to be optimally stable in thermodynamic terms, yielding a free energy gain relative to the soluble state of about -2 kJ mol^{-1} per residue. But our experimental data show that geometric and topological frustration inherent in the packing of a longer polypeptide chain into the fibrillar state leads to a decrease of this value to about -0.5 kJ mol^{-1} per residue for fibrils formed from polypeptide sequences with 100 residues. These findings can be formalized by simple scaling arguments as mentioned above (Supporting Information S3), which provide a practical formulation of this characteristic frustration that, as in spin glasses,[17] leads to the emergence of relatively simple subextensive power law relations as discussed above. The phenomenon of frustration is also likely to be the origin of the observation that fibrils formed from longer protein sequences have on average a lower Young's modulus[5] than those formed from shorter peptides,[18] indicating that the local intermolecular interactions per unit length within the fibril core decrease with increasing sequence length.

The largest deviation from the trend shown in Figure 2B is for fibrils of lysozyme formed at pH 2.0, for which the predicted free energy of elongation ΔG^0_{el} is about 30% lower than the value measured experimentally. Such variations from the stability predicted solely on the basis of the scaling arguments can be used to estimate the magnitude by which we can expect specific factors associated with sequence and with experimental conditions to affect core stability, including, as is the case here for

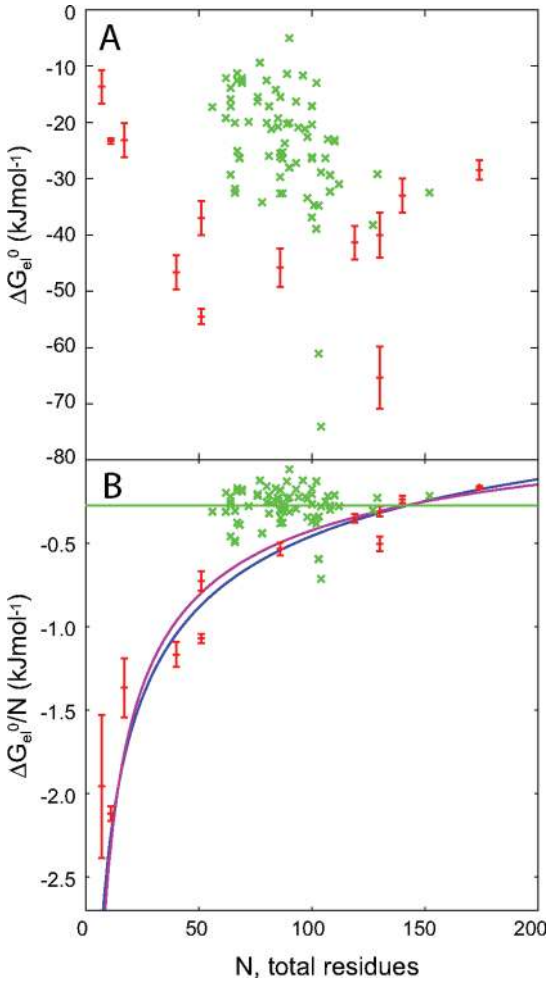

Figure 2. (A) Experimentally determined elongation free energies ΔG_{el}^0 as a function of the number of residues in the protein chain N (red points). (B) Experimentally determined elongation free energies ΔG_{el}^0 per residue as a function of residue number N. These data are fitted with $\gamma = -1/2$ (purple line, see Supporting Information S3) and with $\gamma = -1/3$ (green line, see Supporting Information S3). For comparison, the free energy per residue of the native state is shown for a set of 76 globular proteins (green, see Supporting Information S5). Model parameters for the fibrillar state obtained from fitting to the experimental data are $\varepsilon_0 = 0.4 \pm 0.1$ and $\varepsilon_1 = -8.9 \pm 0.8$ for $\gamma = -1/2$, and $\varepsilon_0 = 1.1 \pm 0.1$ and $\varepsilon_1 = -7.5 \pm 0.6$ for $\gamma = -1/3$.

lysozyme, the presence of disulfide bonds and a multidomain architecture of the native state. This situation contrasts with trends observed in the stability of the native states of globular proteins against unfolding, where, for example, we find that the mean free energy per residue for the native states of 76 representative small proteins is -0.27 kJ mol^{-1} (Figure 2A, green), but the correlation coefficient between the free energy per residue and the number of residues is only 0.2.

This result demonstrates that, in contrast to the situation found here for amyloid fibrils, there is no simple relationship between the number of residues in natively folded globular proteins and their thermodynamic stabilities. This situation is likely to be a result of the fact that native structures have been selected by evolutionary pressure to have properties that optimize biological function and are therefore not determined simply by the polymeric nature of polypeptide chains. Importantly, the energy difference between the soluble and the fibrillar states of a polypeptide chain, in the majority of cases studied here, exceeds the energy difference between the unfolded and natively folded states (Figure 2). An interesting feature revealed through our scaling analysis is, however, that there is a sequence length of approximately 100 residues at which amyloid fibrils are predicted to have their maximal stability relative to the corresponding soluble state. Indeed, the condition

$$\frac{dG_{tot}}{dN}|_{N_0} = 0$$

with

$$G_{tot}(N) = \varepsilon_0 N + \varepsilon_1 N^{\gamma + 1}$$

enables the optimal length to be predicted (Supporting Information S4), i.e.,

$$N_0 = \left[\frac{-\varepsilon_0}{\gamma\varepsilon_1 + \varepsilon_1}\right]^{1/\gamma}$$

which is 123 for $\gamma = -1/2$ and 94 for $\gamma = -1/3$. This result suggests an explanation for the observation that, although the average length of proteins in the human genome is over 400 residues,[19,20] the peptides and proteins that are commonly involved in disorders related to amyloid formation are all significantly shorter, such as the IAPP peptide in Type II diabetes (37 residues), the Aβ peptide in Alzheimer's disease (42 residues), and α-synuclein in Parkinson's disease (140 residues). Our results show that the energy gained from the transition of a longer sequence from its soluble state to an amyloid configuration would be substantially less than for these shorter disease-associated sequences, leading to the interesting possibility that a reduced propensity to form intractable amyloid deposits could be one of the driving forces for the evolution of large protein molecules.

The results that we have presented here have profound implications on our understanding of the thermodynamics and kinetics of protein molecules. When the concentration M_S of a protein exceeds a critical value[9,12−15] given by

$$M_S^{max} = \exp(\Delta G_{el}^0/k_B T)$$

the native state is not thermodynamically stable and a protein can in principle lower its overall free energy through amyloid formation, in the same manner in which other types of molecules that exceed their solubility limit have a tendency to form insoluble amorphous or crystalline structures.

The question is then whether living systems, ever or even in general, operate under conditions of such metastability under normal circumstances. In order to answer this question, we have obtained values from the literature for the physiological concentrations of as many as the proteins considered in this study as we could find (Figure 3). Remarkably, these values are generally found to exceed the limiting concentrations determined in this study, in some cases by a substantial margin.

Thus, for example, the physiological concentration of lysozyme in plasma[22] typically exceeds that corresponding to the critical value (Table 1) by a factor of over 50. The results that we describe in this paper indicate, therefore, that the stability relative to amyloid formation of the native states, at least for small proteins, is not a

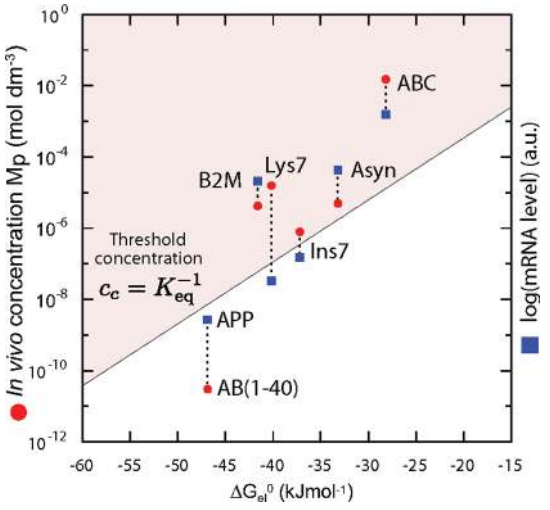

Figure 3. Experimentally measured *in vivo* concentrations of a range of proteins studied here (M_P, filled red circles) are compared to the limiting concentrations (M_S^{max}, black line); the red shaded area indicates the region in which the *in vivo* concentration exceeds its limiting value and conversion to the amyloid state is thermodynamically favored. The estimates of protein concentrations obtained from mRNA expression levels (blue squares) are also shown for comparison (values from ref 21).

consequence of thermodynamic factors but of kinetic ones, which are associated with the aggregation process *in vivo* and which act to reduce the possibility of uncontrolled amyloid formation. Such a situation is not, however, unique in biological systems, as the covalent structures of the biomolecules themselves are metastable relative to their chemical components.[23]

We conclude, therefore, that although the native folds of proteins are almost universally stable under physiological conditions with respect to unfolding, the native states of either folded or natively unfolded proteins may be only metastable with respect to amyloid formation, because of the high kinetic barriers associated with the *in vivo* self-assembly of polypeptide chains. As the loss of solubility resulting from amyloid formation is generally highly detrimental to the biochemical processes that take place in the cell as a result of a loss of normal function or a gain of toxic function,[1−3] living organisms have evolved a range of regulatory and quality control mechanisms with the aim of preventing aggregation over their normal life spans.[1,3,4] Such mechanisms act to raise still further the intrinsic kinetic barriers to aggregation, thus adding further protection. This protection may be particularly strong in higher organisms, as the average thermodynamic stability of amyloid fibrils should decrease with the increasing average protein length found from prokaryotes to eukaryotes (see Figure 2). In some circumstances, however, for example as a result of mutations, stress, and aging, aggregation either of individual proteins or at the proteome level does occur on biologically relevant time scales.[24,25] These factors cause protein molecules to overcome the kinetic barriers that maintain them in their soluble states, enabling them to convert into thermodynamically more stable aggregates, at least some of which are linked to highly debilitating and currently incurable pathogenic conditions.[2]

■ ASSOCIATED CONTENT

Supporting Information. Methods S1−S6. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
cmd44@cam.ac.uk

■ ACKNOWLEDGMENT

We thank Wilhelm Huck, Maarten Biesheuvel, David Chandler, and David Bowler for helpful discussions. This work was supported by the EPSRC, BBSRC, EMBO (A.J.B.), St John's College, Cambridge (T.P.J.K.), the Australian Academy of Science (S.L.G.), the Wellcome Trust (C.M.D.), and the Leverhulme Trust (M.V., C.M.D.).

■ REFERENCES

(1) Dobson, C. M. *Nature* **2003**, *426*, 884–890.
(2) Chiti, F.; Dobson, C. M. *Annu. Rev. Biochem.* **2006**, *75*, 333–366.
(3) Haass, C.; Selkoe, D. J. *Nature Rev. Mol. Cell Biol.* **2007**, *8*, 101–112.
(4) Dobson, C. M. *Trends Biochem. Sci.* **1999**, *24*, 329–332.
(5) Knowles, T. P.; Fitzpatrick, A. W.; Meehan, S.; Mott, H. R.; Vendruscolo, M.; Dobson, C. M.; Welland, M. E. *Science* **2007**, *318*, 1900–1903.
(6) Tycko, R. Q. *Rev. Biophys.* **2006**, *39*, 1–55.
(7) Wasmer, C.; Lange, A.; Van Melckebeke, H.; Siemer, A. B.; Riek, R.; Meier, B. H. *Science* **2008**, *319*, 1523–1526.
(8) Sachse, C.; Fandrich, M.; Grigorieff, N. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 7462–7466.
(9) Kodali, R.; Wetzel, R. *Curr. Opin. Struct. Biol.* **2007**, *17*, 48–57.
(10) Kardos, J.; Yamamoto, K.; Hasegawa, K.; Naiki, H.; Goto, Y. *J. Biol. Chem.* **2004**, *279*, 55308–55314.
(11) O'Nuallain, B.; Shivaprasad, S.; Kheterpal, I.; Wetzel, R. *Biochemistry* **2005**, *44*, 12709–12718.
(12) Carulla, N.; Caddy, G. L.; Hall, D. R.; Zurdo, J.; Gairi, M.; Feliz, M.; Giralt, E.; Robinson, C. V.; Dobson, C. M. *Nature* **2005**, *436*, 554–558.
(13) Williams, A. D.; Shivaprasad, S.; Wetzel, R. *J. Mol. Biol.* **2006**, *357*, 1283–1294.
(14) Oosawa, F.; Kasai, M. *J. Mol. Biol.* **1962**, *4*, 10ff.
(15) Narimoto, T.; Sakurai, K.; Okamoto, A.; Chatani, E.; Hoshino, M.; Hasegawa, K.; Naiki, H.; Goto, Y. *FEBS Lett.* **2004**, *576*, 313–319.
(16) Fersht, A. R. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*; W. H. Freeman: New York, 1999.
(17) Derrida, B. *Phys. Rev. B* **1981**, *24*, 2613–2626.
(18) Kol, N.; Adler-Abramovich, L.; Barlam, D.; Shneck, R. Z.; Gazit, E.; Rousso, I. *Nano Lett.* **2005**, *5*, 1343–1346.
(19) Lander, E. S; et al. *Nature* **2001**, *409*, 860–921.
(20) Wang, D. Y.; Hsieh, M.; Li, W. H. *Mol. Biol. Evol.* **2005**, *22*, 142–147.
(21) Tartaglia, G. G.; Pechmann, S.; Dobson, C. M.; Vendruscolo, M. *Trends Biochem. Sci.* **2007**, *32*, 204–206.
(22) Morozov, V. I.; Pryatkin, S. A.; Kalinski, M. I.; Rogozkin, V. A. *Eur. J. Appl. Physiol.* **2003**, *89*, 257–262.
(23) Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. *Molecular Biology of the Cell*, 4th ed.; Garland Science Publishing: New York, 2002.
(24) David, D. C.; Ollikainen, N.; Trinidad, J. C.; Cary, M. P.; Burlingame, A. L.; Kenyon, C. *PLoS Biol.* **2010**, e1000450.
(25) Narayanaswamy, R.; Levy, M.; Tsechansky, M.; Stovall, G. M.; O'Connell, J. D.; Mirrielees, J.; Ellington, A. D.; Marcotte, E. M. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 10147–10152.