DOI: 10.1002/cbic.200900144

Position-Dependent Electrostatic Protection against Protein Aggregation

Alexander K. Buell,[a] Gian Gaetano Tartaglia,[b] Neil R. Birkett,[b] Christopher A. Waudby,[b] Michele Vendruscolo,[b] Xavier Salvatella,[b] Mark E. Welland,*[a] Christopher M. Dobson,*[b] and Tuomas P. J. Knowles*[a]

Proteins can in general perform their functions only by remaining soluble in their biological environment, and much effort has therefore been devoted to understanding the chemical and physical properties of amino acid sequences that promote their solubility.[1–3] The failure of proteins to remain in a soluble state is increasingly recognised as a central problem in contexts ranging from biotechnology to human disease.[4–7] A relationship between the kinetic stability of polypeptide chains in solution and their fundamental physicochemical properties has been established through the finding that global charge and hydrophobicity are closely linked with the rates at which given peptides and proteins aggregate to form amyloid structures.[8,9] It has also been found that specific patterns of amino acids form "aggregation-prone" regions that play a key role in determining the aggregation process.[10–15] Despite their potentially detrimental effects to biological functionality through facilitation of protein aggregation, such aggregation-prone regions are nevertheless often naturally present in proteins because they essentially encode for the formation of β sheets when these structures are required for folding. Thus, in order to disfavour uncontrolled aggregation and promote folding, it has been proposed that evolution has inserted "gatekeeper" residues[16,17] in the sequence in close proximity to the aggregation-prone regions. Such gatekeeper residues are electrostatically charged and prevent the ordered association of aggregation-prone regions of different molecules.[10,12,16,18–20]

Gatekeeper residues have been observed in various systems, but the magnitude and specificity of their effects have remained elusive. Herein we therefore focus on a quantitative characterisation of the gatekeeping effect through a combination of biosensor measurements, atomic force microscopy, and quantitative sequence-based design approaches.[11,12] To investigate the influence of the position of charges on aggregation rates, we generated chemically conservative mutations within the sequence of the SH3 domain of bovine phosphatidylinositol 3′-kinase (PI3K-SH3), an 84-residue non-disease-related amyloidogenic protein[21] that has proved to be a powerful system for the study of the mechanism of protein aggregation.[22] We show that this approach enables quantitative characterisation of the extreme sensitivity that the kinetics of aggregation has toward electrostatic effects. We find that specific mutations that modify the charge of strategically positioned gatekeeper residues can lead to dramatic changes in the aggregation rates. Such changes are several orders of magnitude greater than if the same mutations were incorporated into less sensitive regions. These results demonstrate the importance of the contributions that quantitative chemical approaches can provide to the understanding of complex biomolecular systems, and provide insight into rational strategies for stabilising soluble forms of proteins.

To introduce a supplementary charge under the acidic denaturing conditions at which PI3K-SH3 amyloid fibril formation is most readily observed,[23] the strategy adopted was to change individual glutamic acid residues into lysine residues. For the converse effect, subtraction of a charge, the strategy was to change a lysine residue to a glutamine residue. These mutations were chosen because they have, apart from the change in charge, only a small effect on the overall physicochemical properties of the sequence, in particular, its hydrophobicity and secondary structure propensity. As the global effects of changes in charge on protein aggregation are relatively well established,[8,12] we focus here on elucidating the dependence of the effect of charge on its position in the sequence of the protein.

The positions of the mutations chosen for this study were selected by screening all possible amino acid substitutions using an algorithm[12] that computes an aggregation propensity profile for a protein sequence by taking into account physicochemical parameters such as charge, hydrophobicity, and patterns of varying polarity. Values of this propensity (Z_{agg}) at individual residues can then be defined, and changes resulting from specific mutations can be predicted. Mutation of an amino acid not only alters this Z_{agg} value at the particular position of the mutation but also in the adjacent regions. Summation of the values of the aggregation profile above a threshold value yields an overall Z_{agg} value for a given sequence.[12] In order to probe the effect of the position of the charge within the sequence, we chose two glutamic acid residues in close proximity with each other; these residues are individually predicted to lead to a significant decrease in aggregation rate after being substituted by a residue that is charged at this pH. In addition, we generated a variant in which the charge was decreased, located close to the N terminus of the protein, a region where adding and removing a charge is predicted by the Z_{agg} scores to have only a small effect.

We probed the rates of fibril growth using a combination of nanogram mass change measurements by quartz crystal microbalance (QCM) sensors and determination of the increase in fibril length by atomic force microscopy (AFM). The QCM has

[a] A. K. Buell, Prof. M. E. Welland, Dr. T. P. J. Knowles
Nanoscience Centre, University of Cambridge
J. J. Thomson Avenue, CB3 0FF, Cambridge (UK)
Fax: (+44) 1223-760304
E-mail: mew10@cam.ac.uk
tpjk2@cam.ac.uk

[b] Dr. G. G. Tartaglia, Dr. N. R. Birkett, C. A. Waudby, Dr. M. Vendruscolo, Dr. X. Salvatella, Prof. C. M. Dobson
Department of Chemistry, University of Cambridge
Lensfield Road, CB2 1EW, Cambridge (UK)
Fax: (+44) 1223-336362
E-mail: cmd44@cam.ac.uk

recently been shown to be a very precise measurement technique for the determination of amyloid fibril elongation rates.[24] For fast-aggregating mutants (WT, K16Q), the growth rates were determined from QCM mass shifts (Figure 1 E) and for slow-growing mutants (E52K, E61K) we found that direct AFM measurements of the fibril length increase (Figure 2) over time intervals of days resulted in convenient probe of their elongation rate. We have previously shown that these two methods yield rates in good agreement,[24] and their combination in the present work allows us to probe aggregation reactions with rates spanning more than four orders of magnitude. A crucial aspect of the present work is that we have developed a method to accurately control the covalent attachment of the fibrils to the gold surface of the QCM crystal by a cystamine spacer (Figure 1 A and B, ref. [25]); see the Experimental Section. The level of activation of the fibrils with disulfide groups was adjusted to approximately 5–10% as demonstrated by mass spectrometry (Figure 1 B) in order to ensure sufficient attachment but minimal interference with the seeding ability of the fibrils. The polymerization reactions for all mutants were initiated by seed fibrils grown from the wild-type protein. Cross-seeding experiments with K16Q seeds (data not shown) yielded similar ratios of the elongation rates as for the data obtained by using wild-type seeds ($r_{K16Q:K16Q}/r_{K16Q:WT} = 2.6 \pm 0.5$ versus $r_{WT:K16Q}/r_{WT:WT} = 2.0 \pm 0.3$, for which the indices represent $r_{seed:monomer}$); this observation shows that the effects on the aggregation rates of single point mutations primarily stem from the physicochemical changes in the sequence and that differences induced by the seeding process remain small. In agreement with this result, the AFM images (Figure 2) confirm that the WT seed fibrils grow with identical morphologies whether the growth solution contains the wild-type or mutant proteins. In accord with previous studies,[8,12] we observe a marked correlation between the overall net charge of the protein and the absolute fibril elongation rate, with increased charge leading to a decreased growth rate. This result is attributable to an increased electrostatic repulsion between a fibril and its monomeric precursors.

Examination of the specific values of the rates shows that in agreement with the predicted Z_{agg} profile, the mutation of E to K at positions 52 and 61 generated significantly different (more than two orders of magnitude) elongation rates, with

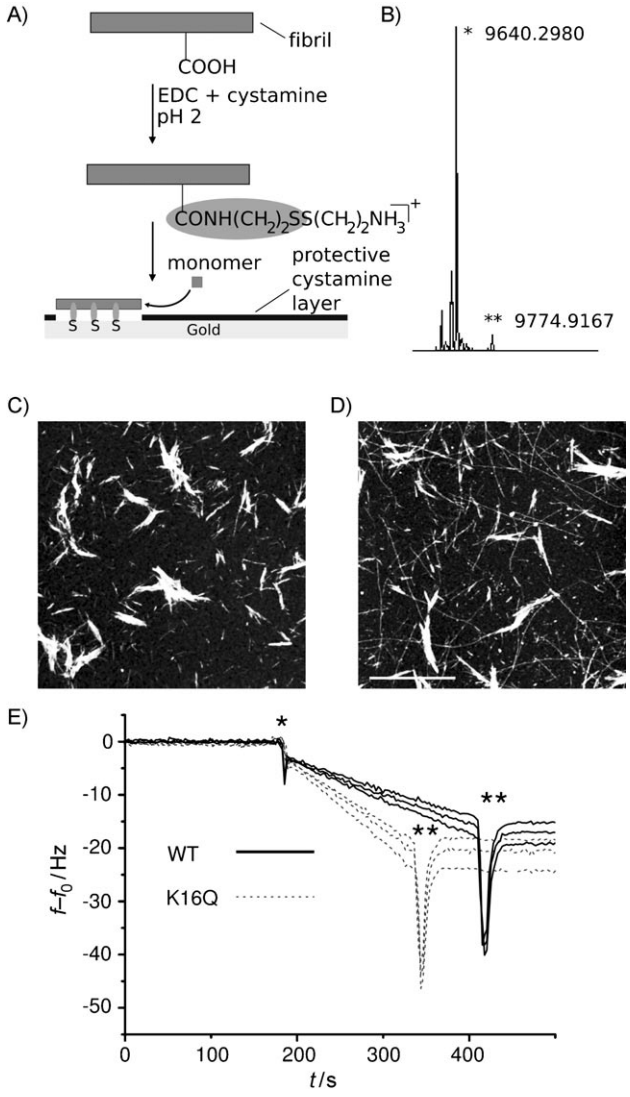

Figure 1. A) Schematic illustration of the surface attachment strategy and subsequent fibril elongation. To verify that the required level of modification was performed, fibrils were functionalized and then dissociated into composing monomers by a pH change from 2 to 11. B) The subsequent mass spectrum shows functionalized (**) and nonderivatized (*) PI3K-SH3 molecules. AFM images of: C) fibrils bound to gold substrates, and D) fibrils exposed to monomeric PI3K-SH3 protein for 2 h (scale bar: 1 μm). E) QCM frequency shifts probing mass loading from fibril growth for WT SH3 protein and the K16Q mutant under otherwise identical conditions described in the text. Injection of protein solution is at the time indicated by (*), and the injection of protein-free 10 mM HCl corresponds to (**). The three traces correspond to three frequency overtones (top to bottom: $n = 7, 5, 3$) measured in the experiments.

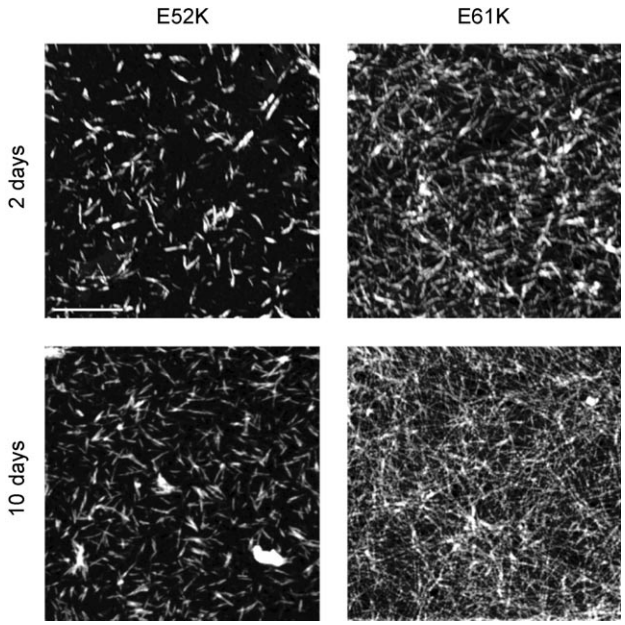

Figure 2. AFM images of WT fibrils bound to gold substrates and incubated in solutions of slow-growing monomeric mutant PI3K-SH3 proteins under the conditions detailed in the text for the times indicated; scale bar: 1 μm.

position 52 being more sensitive to the additional charge than position 61 (Figure 3). In addition, the K16Q mutant, with a decreased net charge in a region predicted from Z_{agg} calculations

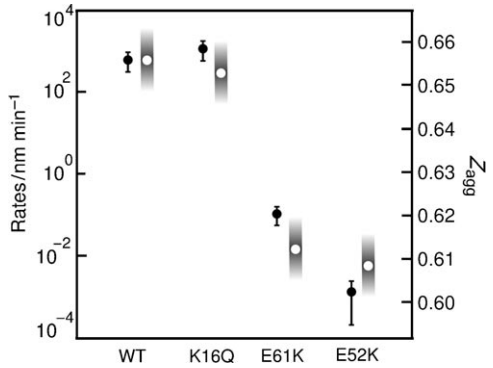

Figure 3. Measured elongation rates (black points, left scale) and the corresponding calculated Z_{agg} values (white circles with error ranges indicated as gradients, right scale).

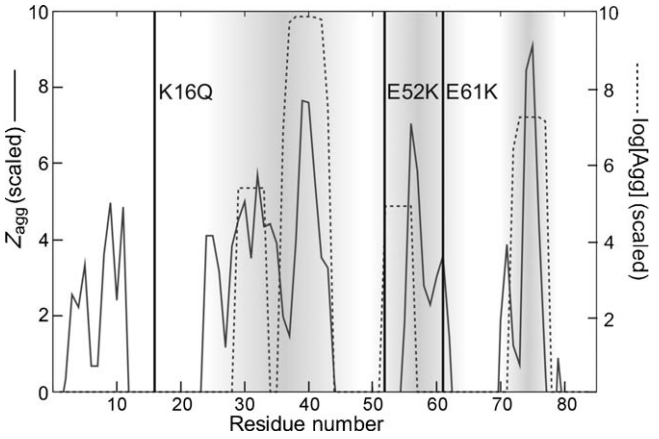

Figure 4. Aggregation propensity of the SH3 polypeptide calculated using the Zyggregator[12] (solid line) and Tango[10] (dashed line) algorithms. The vertical lines denote the position of the chemically conservative mutations analysed here, and the shaded zones highlight aggregation-prone regions.

to have low sensitivity to changes in charge, shows an elongation rate of a similar order of magnitude to that of the wild-type. The high-resolution nature of QCM measurements, however, where the growth of the same ensemble of fibrils can repeatedly be probed with different precursor proteins, allowed for the reproducible determination of a small but measurable difference between the elongation rates of WT and K16Q of approximately a factor of two (Figure 1 E), a value consistent with the previously proposed overall connection between decreased global charge and increased elongation rate.[8,12]

The present results demonstrate the high sensitivity of fibrous protein aggregation to the local chemical environment, even for otherwise chemically equivalent mutations. Insight into the origin of the differential response to charge addition can be gained by examining the aggregation propensity along the polypeptide chain as shown in Figure 4. The positions with high susceptibility to charge modification are found at the edges of aggregation-prone regions, where they are able to interfere most efficiently with the aberrant self-assembly of proteins driven by such regions. On the other hand, in zones that are not in the vicinity of aggregation-prone segments, changes in the charge state of the polypeptide chain only have a small effect on the aggregation rates of the proteins. We note that addition of charge within an aggregation-prone zone itself is not possible without modifying the other chemical characteristics of the sequence such as hydrophobicity.

In summary, we have shown that quantitative measurements over five orders of magnitude of the elongation rates of amyloid fibrils can be achieved by the use of QCM sensors in combination with scanning probe microscopy. Application of this technique in a rational mutational study of the amyloid assembly process of an SH3 protein has enabled us to predict and confirm the location of specific gatekeeper charges that strongly modulate and control protein aggregation depending on the local chemical environment provided by an amino acid sequence, thus shedding light in a uniquely detailed way on the aggregation mechanism at the single-residue scale, and opening up possibilities to stabilise electrostatically[26,27] soluble states of proteins with minimally invasive changes to the polypeptide sequence.

Experimental Section

WT-PI3K SH3 was expressed as described.[13] To induce fibril formation, the protein was incubated at a concentration of 1 mM in aqueous HCl (pH 2.0) at room temperature for 4–5 days; subsequently a new monomer solution was seeded with the preformed fibrils to maximize the fibrillar content. In a second step, fibril solution (100 μL, 1 mM) at pH 2.0 was mixed with a solution of cystamine dihydrochloride (100 μL, 1 M; Sigma) and EDC (*N*-(3-dimethylaminopropyl)-*N'*-ethylcarbodiimide hydrochloride, 1 mg; Sigma). The volume was then increased to 2 mL with HCl, pH 2.0. The solution was kept at room temperature for 30 min and then centrifuged for 2 h at 8900 *g*. The supernatant was removed, and the remaining pellet was re-dissolved in 2 mL HCl, pH 2.0. This purification procedure was repeated four times in order to minimize excess cystamine, which could compete with the fibrils for surface attachment. The solution of the cystamine-activated fibrils was sonicated with a probe sonicator for 30 s (3 s pulses, 3 s off) to decrease the length of the structures. Then 100 μL of the solution was deposited onto each of the QCM chips for 60 min in an atmosphere of 100% humidity. The surface was carefully rinsed with aqueous HCl (pH 2.0), and the sensor was inserted into the microbalance flow cell. After acquisition of a stable baseline, the device was ready for the kinetic measurements. The PI3K-SH3 mutants were designed by standard mutagenesis methodology (primers custom synthesized by Operon, plasmids sequenced by Geneservice) and were expressed and linked to the surface in an identical way to the WT.

Acknowledgements

This work was supported by the UK BBSRC, EPSRC, IRC in Nanotechnology, and the Wellcome and Leverhulme Trusts. A.K.B. is a recipient of the Sélection Internationale scholarship from ENS

Paris, and T.P.J.K. holds a Junior Research Fellowship at St. John's College, Cambridge.

Keywords: amyloid · atomic force microscopy · biosensors · gatekeeper residues · protein aggregation

[1] D. L. Wilkinson, R. G. Harrison, *Biotechnology* **1991**, *9*, 443–448.
[2] C.-S. Goh, N. Lan, S. M. Douglas, B. Wu, N. Echols, A. Smith, D. Milburn, G. T. Montelione, H. Zhao, M. Gerstein, *J. Mol. Biol.* **2004**, *336*, 115–130.
[3] P. Smialowski, A. J. Martin-Galiano, A. Mikolajka, T. Girschick, T. A. Holak, D. Frishman, *Bioinformatics* **2007**, *23*, 2536–2542.
[4] D. J. Selkoe, *Nature* **2003**, *426*, 900–904.
[5] C. M. Dobson, *Nature* **2003**, *426*, 884–890.
[6] S. B. Fowler, S. Poon, R. Muff, F. Chiti, C. M. Dobson, J. Zurdo, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 10105–10110.
[7] F. Chiti, C. M. Dobson, *Annu. Rev. Biochem.* **2006**, *75*, 333–366.
[8] F. Chiti, M. Stefani, N. Taddei, G. Ramponi, C. M. Dobson, *Nature* **2003**, *424*, 805–808.
[9] K. F. DuBay, A. P. Pawar, F. Chiti, J. Zurdo, C. M. Dobson, M. Vendruscolo, *J. Mol. Biol.* **2004**, *341*, 1317–1326.
[10] A.-M. Fernandez-Escamilla, F. Rousseau, J. Schymkowitz, L. Serrano, *Nat. Biotechnol.* **2004**, *22*, 1302–1306.
[11] A. P. Pawar, K. F. Dubay, J. Zurdo, F. Chiti, M. Vendruscolo, C. M. Dobson, *J. Mol. Biol.* **2005**, *350*, 379–392.
[12] G. G. Tartaglia, A. P. Pawar, S. Campioni, C. M. Dobson, F. Chiti, M. Vendruscolo, *J. Mol. Biol.* **2008**, *380*, 425–436.
[13] F. Rousseau, L. Serrano, J. W. H. Schymkowitz, *J. Mol. Biol.* **2006**, *355*, 1037–1047.
[14] H. Xiong, B. L. Buckwalter, H. M. Shieh, M. H. Hecht, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 6349–6353.
[15] E. Gazit, *FASEB J.* **2002**, *16*, 77–83.
[16] D. E. Otzen, M. Oliveberg, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 11746–11751.
[17] S. Matysiak, C. Clementi, *J. Mol. Biol.* **2006**, *363*, 297–308.
[18] M. Lopez de La Paz, L. Serrano, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 87–92.
[19] E. Monsellier, F. Chiti, *EMBO Rep.* **2007**, *8*, 737–742.
[20] E. Monsellier, M. Ramazzotti, N. Taddei, F. Chiti, *PLoS Comput. Biol.* **2008**, *4*, e1000199.
[21] J. I. Guijarro, M. Sunde, J. A. Jones, I. D. Campbell, C. M. Dobson, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 4224–4228.
[22] N. Carulla, G. L. Caddy, D. R. Hall, J. Zurdo, M. Gair, M. Feliz, E. Giralt, C. V. Robinson, C. M. Dobson, *Nature* **2005**, *436*, 554–558.
[23] J. Zurdo, J. I. Guijarro, J. L. Jimnez, H. R. Saibil, C. M. Dobson, *J. Mol. Biol.* **2001**, *311*, 325–340.
[24] T. P. J. Knowles, W. Shu, G. L. Devlin, S. Meehan, S. Auer, C. M. Dobson, M. E. Welland, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 10016–10021.
[25] C. M. Lin, K. A. Mihal, R. J. Krueger, *Biochim. Biophys. Acta Protein Struct. Mol. Enzymol.* **1990**, *1038*, 382–385.
[26] M. S. Lawrence, K. J. Phillips, D. R. Liu, *J. Am. Chem. Soc.* **2007**, *129*, 10110–10112.
[27] M. Vendruscolo, C. M. Dobson, *Nature* **2007**, *449*, 555.

Received: March 13, 2009
Published online on May 4, 2009

