

Review

Open Access

## Sequence determinants of protein aggregation: tools to increase protein solubility

Salvador Ventura\*

Address: Institut de Biotecnologia i Biomedicina and Departament de Bioquímica i Biologia Molecular, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain

Email: Salvador Ventura\* - salvador.ventura@uab.es

\* Corresponding author

Published: 22 April 2005

Received: 17 March 2005

*Microbial Cell Factories* 2005, 4:11 doi:10.1186/1475-2859-4-11

Accepted: 22 April 2005

This article is available from: <http://www.microbialcellfactories.com/content/4/1/11>

© 2005 Ventura; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

*Escherichia coli* is one of the most widely used hosts for the production of recombinant proteins. However, very often the target protein accumulates into insoluble aggregates in a misfolded and biologically inactive form. Bacterial inclusion bodies are major bottlenecks in protein production and are hampering the development of top priority research areas such structural genomics. Inclusion body formation was formerly considered to occur via non-specific association of hydrophobic surfaces in folding intermediates. Increasing evidence, however, indicates that protein aggregation in bacteria resembles to the well-studied process of amyloid fibril formation. Both processes appear to rely on the formation of specific, sequence-dependent, intermolecular interactions driving the formation of structured protein aggregates. This similarity in the mechanisms of aggregation will probably allow applying anti-aggregational strategies already tested in the amyloid context to the less explored area of protein aggregation inside bacteria. Specifically, new sequence-based approaches appear as promising tools to tune protein aggregation in biotechnological processes.

### Review

#### Introduction

In the last decade, protein aggregation has moved beyond being a mostly ignored area of protein chemistry to become a key topic both in medical and biotechnological sciences [1]. The biological significance of protein deposition has been shown to be much higher than formerly thought. First, because the presence of insoluble protein deposits in human tissues correlates with the development of some debilitating human disorders of growing incidence such as Alzheimer's disease, Parkinson's disease, type II diabetes and the transmissible spongiform encephalopathies [2-4]. Second, because it has been shown that under cellular stress conditions, such as severe heat, massive protein misfolding exceeds the buff-

ering capacity of the folding quality machinery and results in the aggregation of proteins, which usually results in cell death [5,6]. Finally, the use of bacteria as factories for recombinant expression is limited by their intrinsic tendency to accumulate the target protein into inactive insoluble aggregates, called inclusion bodies (IBs). IBs are dense, amorphous protein deposits that can be found in both the cytoplasmic and periplasmic space of bacteria [7-11]. In fact, the formation of IBs is the main bottleneck in protein production, narrowing the spectrum of relevant polypeptides obtained by recombinant techniques and hampering the development of top priority research areas such as the *de novo* design of novel proteins, the rational modification of natural proteins or structural and functional genomics. The rising recognition of the crucial sig-

nificance of protein aggregation has resulted in a number of recent reviews [12-19]. This review focuses mainly on the role played by intrinsic polypeptide properties in protein aggregation.

One should distinguish between precipitates, in which proteins maintain the native folded conformation and aggregates, in which proteins adopt new non-native structures. The first type of self-assembly is generated during random precipitation of already native protein due to an environment promoted reduction of solubility in the polypeptide chain. Examples of these processes are salting out by ammonium sulfate or isoelectric precipitation. Reducing ionic force or shifting solution's pH results in immediate dissolution of these precipitates. The second type of macromolecular structures exhibits, without exception, an increase in  $\beta$ -sheet secondary structure content relative to the native conformation and very high concentrations of denaturants or detergents are needed to dissolve them into mainly unfolded polypeptide chains. We will focus our attention on these aggregates, which include amyloid fibrils, thermal aggregates and bacterial IBs. The progress made on the control of their aggregation propensities by means of primary sequence modulation is discussed.

#### **Protein aggregation is usually a specific process**

Protein aggregation has long been considered to be an unspecific process driven by the establishment of non-native contacts between proteins in totally or partially unfolded conformations to form a disordered precipitate. This idea was sustained in part by the diversity of morphologies of aggregates that were observed by techniques such as electron microscopy and atomic force microscopy [20]. This way, the typical amyloid aggregate is a long, straight and unbranched fibril with a diameter between 40 Å and 120 Å [21], whereas inclusion bodies appear as bigger globular electro-dense structures seen as refractile bodies under phase contrast microscopy usually with near 1 micron in diameter [22] and thermal aggregates are usually amorphous [23]. Recent work shows however that often aggregation is a much more specific event than previously expected at least in amyloid fibrils and bacterial IBs [24-27]. In fact, for many biotechnologically relevant proteins, isolation of the IBs is an efficient initial step in the purification process, since they contain usually more than 90% of recombinant protein [28], other proteins trapped in the aggregates are proteolytic fragments of the aggregating protein [29], other aggregation-prone polypeptides deposited by titration of chaperones during recombinant expression [30,31] or even contaminants from the purification process [11]. Similarly in Alzheimer and related neurodegenerative diseases *in vivo* amyloid plaques are composed primarily of the pathogenic aggregating protein rather than resulting from a widespread

recruitment of other amyloidogenic proteins, although proteins such as proteases or chaperones have been also found to co-localize in the amyloid deposits [32].

Amyloid fibrils are thought to form through self-assembly of protein monomers via a nucleation dependent pathway similar to the highly ordered process of protein crystallization [33]. This mechanism is also behind physiological ordered protein aggregation processes as viral coat assembly, microtubule formation or flagellum formation [33]. All these processes are characterized by an initial slow nucleation phase, in which the protein associates to form ordered oligomeric nucleus followed by a growth phase, in which the nucleus rapidly grows to form larger insoluble polymers. Addition of preformed protein nucleus during the lag time results in immediate polymerization. All these aggregation processes and in particular amyloid fibril formation are highly specific. This way, in the aggregation of  $\beta$ -amyloid protein, islet amyloid peptide, transthyretin, and prion protein the formation of amyloid fibrils is not seeded by preformed fibrils of similar amyloidogenic proteins [34-36]. Although it has been shown that some amyloid fibrils can accommodate up to 1% of a foreign peptide, indicating that some co-aggregation can occur [37], the efficiency of this event decreases rapidly as differences in protein sequence of co-aggregating proteins increases showing that specific protein-protein interactions are needed for amyloid fibril formation to occur [38].

Aggregation into IBs during recombinant protein expression has been usually thought to occur via non-specific association of hydrophobic patches on the surface of folding intermediates. However, the reduced number of IBs (usually one) formed during recombinant protein expression in bacteria suggested that they may be formed by the growth of a reduced number of "founder" aggregates in a nucleation-like mechanism. In this respect, the aggregation of the P22 coat protein has been extensively characterized [39-42] and it was demonstrated that when partially folded species of this protein were mixed *in vitro* with those of tailspike protein, no co-aggregation occurs, despite the fact that both form IBs when expressed individually in bacteria [43]. The folding intermediates for each protein preferred to self-associate indicating specificity in the *in vitro* aggregation process and suggesting that specific interactions may underlie IBs formation in the cell. Very recently, we have confirmed this extent by showing that the preformed IBs of an aggregation-prone  $\beta$ -galactosidase variant are able to act as effective aggregating cores for the aggregation of its soluble, partially folded counterpart in a dose-dependent manner [27]. Moreover, the aggregation process is highly specific as shown by the fact that preformed IBs promote deposition of homologous but not heterologous polypeptides. Both protein

sequence and conformation appear to play a role in the establishment of specific intermolecular contacts between aggregating polypeptide chains to form IBs, since aggregated  $\beta$ -galactosidase moiety in IBs do not recognize the properly folded tetrameric enzyme [27].

Inclusion bodies in mammalian cells, the so-called aggresomes, are far more complex structures than those in bacteria containing many proteins, including molecular chaperones, components of the ubiquitin-proteasome system, centrosomal material, and cytoskeletal proteins [44]. This suggested that co-aggregation of misfolded, damaged, or mutant proteins with normal cellular proteins could explain both the presence of multiple proteins in IBs and the toxicity associated with protein aggregation in many neurodegenerative diseases [45]. However, also in this complex system, protein aggregation into IBs exhibits exquisite specificity even among extremely hydrophobic substrates expressed at very high levels [46]. Thus, independent of the source, both amyloid fibril formation and IBs aggregation depend, at least partially, on the formation of specific protein-protein interactions between non-native species.

#### **Different polypeptides aggregate into similar structures**

The formation of amyloid fibrils was initially associated to a reduced number of proteins related to recognized pathological situations. Nevertheless, a growing number of globular proteins not related to disease can be induced to generate similar fibrils *in vitro*, albeit in some cases only in non-native conditions, leading to the suggestion that the ability to form amyloids is intrinsic to many or all polypeptides when their normal folding pathways are compromised [47-50]. This appears to be true for IBs as well since deposition in such structures has been reported in the recombinant expression of many, but not all, heterologous genes and in the high level expression of several endogenous genes [7,51,52].

No sequence or structural similarities are apparent between any of the proteins that display the ability to form amyloids. Prior to fibrillation, amyloidogenic polypeptides may be rich in  $\beta$ -sheet,  $\alpha$ -helix,  $\beta$ -helix, or combine  $\alpha$ -helices and  $\beta$ -sheets. They may be globular proteins with a stable unique conformation in the native state or belong to the class of natively unfolded proteins. Despite these differences, the fibrils formed by different polypeptides display many common properties including high content of  $\beta$ -sheet secondary structure forming a core cross- $\beta$  architecture in which continuous  $\beta$ -sheets are formed with  $\beta$ -strands running perpendicular to the long axis of the fibrils [53].

As in the case of amyloids, proteins incorporated in IBs are not related either structurally or sequentially and deposi-

tion during heterologous expression in bacteria has been reported for small, large, monomeric or multimeric proteins. The internal architecture of IBs has long thought by molecular biologists to be amorphous, despite the fact that several observations in the early 90's pointed to the presence of ordered structure in IBs [54-56]. The use of attenuated total reflectance FTIR in IBs formed by all- $\alpha$ , all- $\beta$  or  $\alpha$  +  $\beta$  showed that in all cases, even for all- $\beta$  proteins, significant new  $\beta$  structure, compared to that in the native conformation, was observed. Interestingly, the amount of secondary structure in the inclusion body varies from one protein to another, as does the amount of disordered structure. More recently, others and we have recapitulated these studies in previously unexplored protein systems, showing clearly that the intermolecular interactions leading to aggregation in IBs in the cell involve  $\beta$ -sheet like interactions [27,57]. Although the exact nature of the intermolecular interactions is unknown, and could be different in different IBs, the overall FTIR data suggest that the newly formed  $\beta$ -sheets in IBs are tightly packed with short hydrogen bonds providing them high stability. These features are reminiscent of those stabilizing the structure of amyloid fibrils [53]. In addition, Thioflavin-T and Congo red, two dyes used for the diagnostic of amyloid structures also bind to IBs, confirming thus certain resemblance in the internal organization of both kinds of aggregates [27]. Also, even if we still lack structural information on thermal aggregates purified directly from bacteria under stress conditions, it has been shown that *in vitro* heat denaturation leads to the formation of thermal aggregates that display the  $\beta$ -sheet signature as analyzed by FTIR [58] and are also able to bind amyloid dyes [59].

Despite the fact that the different types of aggregates share similar characteristics, they are obviously not identical and exhibit a series of distinctive features. First, most amyloid fibrils are SDS-insoluble, whereas SDS can usually dissolve IBs. This observation is in agreement with the higher extent of  $\beta$ -sheet content of amyloids relative to that in IBs, in which the presence of some native or disordered structure can be still detected [27,60]. As a result amyloids would display more and stronger intermolecular non-covalent interactions that would provide them with higher order and stability in front of denaturation, while sharing similar overall connectivity between polypeptide chains than this present in IBs. Also, the regulation of amyloid and bacterial aggregates formation *in vivo* appears to be somehow different. In this sense, it has been demonstrated that in yeast the formation of amyloids by the Sup 35 prion is highly dependent on the presence of the Hsp 104 chaperone [61]. In contrast, the role of the bacterial Hsp 104 homologue, ClpB, in the regulation of inclusion body formation in *E. coli* is more controversial, some studies indicating that, as in the case of Hsp

104, it binds preferentially to the hydrophobic surface of aggregated protein [62], while others suggesting only a moderate role in the process of aggregation, which is mainly controlled by the chaperones DnaK and GroEL [63]. Interestingly, the bacterial chaperone GroEL is able to modulate both *in vitro* [64] and *in vivo* in mammalian cells [65] the aggregation of proteins involved in amyloid pathologies, suggesting that in spite of the constraints imposed by the different cellular contexts some similitude may still exist between the mechanisms of bacterial and eukaryotic protein aggregation.

Regardless of the existence of some structural or functional differences between the aggregates formed in bacteria and those in eukaryotic cells, in both cases there is an inherent tendency to kidnap misfolded protein in the interior of such supra-molecular structures. It is suggested that this is a mechanism evolved to reduce the potential toxicity of partially folded monomers or small oligomers, which by exposing large hydrophobic surfaces could interact inappropriately with a wide range of cellular components, hampering this way cell function [66]. In these sense, specific aggregation could be a conserved strategy playing a cellular protective role.

#### **Sequence modulates protein aggregation**

One of the major unanswered questions of protein aggregation is the specificity with which primary sequence determines both the aggregation propensity and the specific details of the aggregated structure. The hypothesis that the ability of proteins to form ordered aggregates is a general property of the polypeptide chain rather than to be limited to a restricted set of proteins [2] seems reasonable, especially if the main driving force for aggregation is the formation of an inter-backbone hydrogen-bonded network leading to the above described  $\beta$ -sheets structures, since all polypeptides regardless of sequence share the polypeptide backbone. In this regard, IBs and amyloid formation abilities has not been associated a priori to particular protein sequences, being this fact, an additional obstacle to predict the yield of a given protein in a new production process or its cellular toxicity. However, in recent times it is coming clear that sequence modulates aggregation, giving a chance to control the unwanted protein deposition phenomena.

A first indication that sequence controls deposition comes from the observation that not all regions of a polypeptide are equally important for determining the aggregation propensities. This way, we have proved recently that very short specific amino acid stretches can act as facilitators or inhibitors in the incorporation of globular proteins into amyloid fibrils [67]. These relevant regions are usually known as aggregation "hot spots". Aggregation-prone regions are blocked in the native state of globular proteins

because their side chains are usually hidden in the interior of the protein hydrophobic core or already involved in the establishment of the network of native contacts that stabilizes a protein. This is the reason why globular proteins rarely aggregate from their native states. Destabilization usually results in an increased population of partially folded molecules and is well established as a triggering factor in disorders associated with the deposition of proteins that are globular in their normal functional states [68].

Accordingly, peptides and proteins involved in the most prevalent human neurodegenerative diseases are mostly unstructured within the cell [3]. In these disorders, protein deposition does not require the unfolding of a globular native conformation and occurs by direct self-assembly of the unstructured polypeptide chains, in which aggregation-prone, usually hydrophobic, regions are already exposed to solvent. The presence of aggregation "hot spots" have been already described in the peptides and proteins underlying Alzheimer's, Creutzfeldt-Jakob disease, or some systemic amyloidogenic disorders [69-71]. Independent of the native conformation and stability of the protein, the high level of expression during recombinant production results in a large number of polypeptides emerging from the ribosome in at least partially unfolded conformations which usually associate to form IBs. Even if not yet proved, it is thinkable that the presence of aggregation prone sequences in these conformers will influence at least partially the equilibrium between aggregated and folded protein during recombinant expression. Interestingly, it is observed that proteins assembled into amyloid *in vitro* usually render insoluble during recombinant protein expression. For example, this happens for proteins involved in disease such as A $\beta$ 42 amyloid peptide,  $\beta$ -2-microglobulin, mammalian prions and human islet amyloid polypeptide [72-75].

The study of the effects of mutations on the formation of amyloid fibrils and IBs also point to the role of sequence as an aggregation controller. Two types of mutations should be distinguished according to their ability to destabilize or not significantly the native state of the protein. As stated before, destabilizing mutations favour aggregation by originating an ensemble of partially unfolded conformations allowing this way the establishment of inter-molecular interactions. In addition, it has been shown that punctual mutations can also facilitate aggregation without affecting the native state stability when they promote the conversion of already unfolded or partially folded polypeptides into oligomeric forms that further aggregate to form insoluble species. In these cases, protein aggregation has been found to be tuned by mutations that change the polarity, the secondary structure propensity or the net charge of the polypeptide. In general, increases in

hydrophobicity and  $\beta$ -sheet propensity result in increased aggregation whereas an increase in the overall net charge decreases this tendency [24,76,77]. There are a good number of protein systems in which it has been shown that point mutations may dramatically affect the amount of aggregate formation; these include the P22 tailspike protein, single-chain antibodies, interferon- $\gamma$ , colicin A, Che Y, immunoglobulin domains, and interleukin-1 $\beta$  for IBs formation [43,78-83] and SH3-domains, acylphosphatase, amylin, prion peptides,  $\alpha$ -synuclein, amyloid- $\beta$  peptide and tau for amyloid formation [25,67,84-88]. Notably, mutant proteins with reduced *in vitro* amyloid propensity are expressed usually in *E. coli* as more soluble proteins than the natural occurring ones [89], whereas providing a previously *in vitro* soluble protein increased amyloid propensity results in accumulation as IBs during recombinant expression [90,91]. Moreover, when amyloid proteins have been designed de novo, all proteins displaying amyloid properties *in vitro* accumulated *in vivo* as bacterial IBs [92], but the rational introduction of point mutations that convert these aggregation-prone proteins into monomeric  $\beta$ -sheet forms allowed their expression in bacteria in soluble forms [93]. These observations strongly suggest that both aggregation phenomena are related and depend in last term on tendency to self-aggregate associated to individual protein sequences. This way, it appears that the study of bacterial models may contribute significantly in the future to the understanding of protein misfolding and aggregation, since they are fast, simple and biologically relevant experimental systems. Conversely, it is thinkable that the application of successful anti-depositional strategies derived from the numerous studies dealing with amyloid fibril formation to the less explored area of protein aggregation within the cell may provide clues to optimize biotechnological protein production. In this regard, simple sequence-based computational approaches have been developed very recently which permit to predict with reasonable accuracy the aggregation propensity of polypeptides [94-97]. In particular, TANGO a statistical mechanics algorithm based on the physico-chemical principles of  $\beta$ -sheet formation, extended by the assumption that the core regions of an aggregate are fully buried, accurately predicts the aggregation propensity of a data set of more than 200 different peptides [95,96]. Without doubt, these new algorithms born in the sinus of the amyloid area are going to be very useful tools for the rational modification of polypeptides for biotechnological applications, opening the door to a fully automated, sequence-based design strategies to improve the solubility of proteins of industrial interest.

#### **Perspectives: Towards rational design of protein solubility**

There is an increasing need for the efficient production of genetically engineered proteins as a result of the success of the genome sequencing projects. From the different host

that may be used to produce this large set of proteins, bacteria, mainly *E. coli*, still appears as the default option, particularly when the biological activity of the protein does not depend on post-translational modifications. *E. coli* is fast and inexpensive to culture, easy to handle and manipulate genetically and usually renders high levels of recombinant products. However, expression of recombinant proteins in *E. coli* often results in the accumulation of the protein product in inactive IBs in the cell. The recovery of bioactive proteins from IBs is a complex process. Still, IBs formation is such a frequent phenomena in protein production that a large number of in house and commercial protocols and solutions have been developed in order to obtain pure, active and soluble protein from IBs [17,98]. Nevertheless, the purification of protein from IBs usually requires the optimization of refolding conditions for each individual target, the recovery yields are usually poor and one should be sure that the refolding procedure does not affect the integrity and activity of the recovered protein. In addition, purification of over-expressed soluble proteins is faster and cheaper than obtaining it in a pure form from IBs, especially at large scale. Overall, optimizing the levels of soluble protein is nowadays a more attractive strategy to increase pure and active protein yield than recovering highly expressed protein in aggregated form [99].

The observation that natural proteins are usually soluble in their biological environments may help to maximize soluble expression levels in recombinant approaches. This way, nature has provided proteins with a reasonable conformational stability in the native state, in which most of the hydrophobic residues, amide and carboxyl groups and aggregation-prone sequence stretches are buried or involved in intra-molecular interactions. This appears as a very successful strategy used to avoid aggregation, since few proteins are able to aggregate from its stable native conformation. Along with this, over-stabilized proteins of thermophilic organisms are usually expressed in soluble forms during recombinant protein production [100-102] and a positive correlation between thermostability and solubility has been recently reported [103]. In addition, the analysis of protein databases has shown that highly aggregating sequences are less frequent in proteins than innocuous amino acid combinations and that, if present; they are surrounded by amino acids that disrupt their aggregating capability [94]. These evidences support the suggestion that natural protein sequences have evolved in part to code for structural characteristics other than those included in the native fold, such as avoidance of aggregation. According to this, protein aggregation results from a failure of the natural protective strategies under special circumstances, such as recombinant protein expression.

Using rational design to engineer target proteins in order to emulate and reinforce natural anti-aggregation mecha-

nisms, taking advantage of the above mentioned computational methods to predict aggregation, appears thus as a reasonable approach to overwhelm protein deposition and optimizing the levels of soluble protein in biotechnological processes. Few, but successful experimental steps have been taken already in this direction. First, improving thermodynamic stability by rational mutation has been shown to render more soluble heterologous protein versions [104]. Second, it has been proven that decreasing the intrinsic propensity to aggregate of the partially unfolded state of an aggregation-prone protein by modulating the net polypeptide charge and introduction of electrostatic repulsions also results in increased solubility [105]. Finally, the analysis, identification and disruption by mutation of sequential "hot spots" of aggregation has allowed the recovering from the *E. coli* supernatant of previously aggregated polypeptides [67,93,106].

## Conclusion

The raising interest to understand the mechanisms underlying protein aggregation in the cell has crystallized in a good number of recent relevant studies in an area whose biological significance is coming of central importance in biotechnology. The scenario emerging from these efforts is especially encouraging because one can foresee a future in which rational design of protein solubility based on natural laws will allow to tune aggregation, permitting to over-pass the main bottleneck in high throughput expression projects.

## Acknowledgements

The author is recipient of a "Ramón y Cajal" contract awarded by the MCYT-Spain and co-financed by the Universitat Autònoma de Barcelona (UAB) and founded by PNL2004-40 from UAB.

## References

- Smith A: **Protein misfolding**. *Nature* 2003, **426**:883.
- Dobson CM: **Protein folding and misfolding**. *Nature* 2003, **426**:884-890.
- Rochet JC, Lansbury PT Jr: **Amyloid fibrillogenesis: themes and variations**. *Curr Opin Struct Biol* 2000, **10**:60-68.
- Cohen FE, Kelly JW: **Therapeutic approaches to protein-misfolding diseases**. *Nature* 2003, **426**:905-909.
- Morimoto R, Tissieres A, Georgopoulos C: **The Biology of Heat Shock Proteins and Molecular Chaperones**. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY; 1994.
- Weibezahn J, Tessarz P, Schlieker C, Zahn R, Maglica Z, Lee S, Zentgraf H, Weber-Ban EU, Dougan DA, Tsai FT, Mogk A, Bukau B: **Thermotolerance requires refolding of aggregated proteins by substrate translocation through the central pore of ClpB**. *Cell* 2004, **119**:653-65.
- Marston FA: **The purification of eukaryotic polypeptides synthesized in *Escherichia coli***. *Biochem J* 1986, **40**:1-12.a2.
- Georgiou G, Telford JN, Shuler ML, Wilson DB: **Localization of inclusion bodies in *Escherichia coli* overproducing beta-lactamase or alkaline phosphatase**. *Appl Environ Microbiol* 1986, **52**:1157-61.
- Georgiou G, Valax P: **Expression of correctly folded proteins in *Escherichia coli***. *Curr Opin Biotechnol* 1996, **7**:190-197.
- Baneyx F: **Recombinant protein expression in *Escherichia coli***. *Curr Opin Biotechnol* 1999, **10**:411-421.
- Georgiou G, Valax P: **Isolating inclusion bodies from bacteria**. *Methods Enzymol* 1999, **309**:48-58.
- Stefani M: **Protein misfolding and aggregation: new examples in medicine and biology of the dark side of the protein world**. *Biochim Biophys Acta* 2004, **1739**:5-25.
- Ross CA, Poirier MA: **Protein aggregation and neurodegenerative disease**. *Nat Med* 2004, **10**:S10-7.
- Dobson CM: **Principles of protein folding, misfolding and aggregation**. *Semin Cell Dev Biol* 2004, **15**:3-16.
- Ohnishi S, Takano K: **Amyloid fibrils from the viewpoint of protein folding**. *Cell Mol Life Sci* 2004, **61**:511-24.
- Lee HG, Petersen RB, Zhu X, Honda K, Aliev G, Smith MA, Perry G: **Will preventing protein aggregates live up to its promise as prophylaxis against neurodegenerative diseases?** *Brain Pathol* 2003, **13**:630-8.
- Cabrita LD, Bottomley SP: **Protein expression and refolding – A practical guide to getting the most out of inclusion bodies**. *Biotechnol Annu Rev* 2004, **10**:31-50.
- Fahnert B, Lilie H, Neubauer P: **Inclusion bodies: formation and utilisation**. *Adv Biochem Eng Biotechnol* 2004, **89**:93-142.
- Villaverde A, Carrio MM: **Protein aggregation in recombinant bacteria: biological role of inclusion bodies**. *Biotechnol Lett* 2003, **25**:1385-95.
- Dobson CM: **Protein-misfolding diseases: Getting out of shape**. *Nature* 2002, **418**:729-730.
- Ventura S, Lacroix E, Serrano L: **Insights into the origin of the tendency of the P13-SH3 domain to form amyloid fibrils**. *J Mol Biol* 2002, **322**:1147-58.
- Williams DC, Van Frank RM, Muth WL, Burnett JP: **Cytoplasmic inclusion bodies in *Escherichia coli* producing biosynthetic human insulin proteins**. *Science* 1982, **215**:687-9.
- Kundu B, Guptasarma P: **Use of a hydrophobic dye to indirectly probe the structural organization and conformational plasticity of molecules in amorphous aggregates of carbonic anhydrase**. *Biochem Biophys Res Commun* 2002, **293**:572-7.
- Chiti F, Calamai M, Taddei N, Stefani M, Ramponi G, Dobson CM: **Studies of the aggregation of mutant proteins in vitro provide insights into the genetics of amyloid diseases**. *Proc Natl Acad Sci USA* 2002, **99**:16419-16426.
- Chiti F, Stefani M, Taddei N, Ramponi G, Dobson CM: **Rationalization of the effects of mutations on peptide and protein aggregation rates**. *Nature* 2003, **424**:805-8.
- Chiti F, Taddei N, Baroni F, Capanni C, Stefani M, Ramponi G, Dobson CM: **Kinetic partitioning of protein folding and aggregation**. *Nat Struct Biol* 2002, **9**:137-43.
- Carrio M, Gonzalez-Montalban N, Vera A, Villaverde A, Ventura S: **Amyloid-like properties of bacterial inclusion bodies**. *J Mol Biol* 2005, **347**:1025-1037.
- Clark ED: **Protein refolding for industrial processes**. *Curr Opin Biotechnol* 2001, **12**:202-7.
- Corchero JL, Viaplana E, Benito A, Villaverde A: **The position of the heterologous domain can influence the solubility and proteolysis of b-galactosidase fusion proteins**. *J Biotechnol* 1996, **48**:191-200.
- Rinas U, Bailey JE: **Protein compositional analysis of inclusion bodies produced in recombinant *Escherichia coli***. *Appl Microbiol Biotechnol* 1992, **37**:609-614.
- Hoffmann F, Rinas U: **Roles of heat-shock chaperones in the production of recombinant proteins in *Escherichia coli***. *Advan Biochem Eng Biotechnol* 2004, **89**:143-61.
- De la Torre JC: **Is Alzheimer's disease a neurodegenerative or a vascular disorder? Data, dogma, and dialectics**. *Lancet Neurol* 2004, **3**:184-90.
- Harper JD, Lieber CM, Lansbury PT Jr: **Atomic force microscopic imaging of seeded fibril formation and fibril branching by the Alzheimer's disease amyloid-beta protein**. *Chem Biol* 1997, **4**:951-9.
- Jarrett JT, Lansbury PT Jr: **Amyloid fibril formation requires a chemically discriminating nucleation event: studies of an amyloidogenic sequence from the bacterial protein OsmB**. *Biochemistry* 1992, **31**:12345-52.
- Kocisko DA, Priola SA, Raymond GJ, Chesebro B, Lansbury PT Jr, Caughey B: **Species specificity in the cell-free conversion of prion protein to protease-resistant forms: a model for the scrapie species barrier**. *Proc Natl Acad Sci U S A* 1995, **92**:3923-7.
- Saraiva MJ, Birken S, Costa PP, Goodman DS: **Amyloid fibril protein in familial amyloidotic polyneuropathy, Portuguese**

- type. **Definition of molecular abnormality in transthyretin (prealbumin).** *J Clin Invest* 1984, **74**:104-19.
37. MacPhee CE, Dobson CM: **Formation of mixed fibrils demonstrates the generic nature and potential utility of amyloid nanostructures.** *J Am Chem Soc* 2000, **122**:12707-12713.
  38. Krebs MR, Morozova-Roche LA, Daniel K, Robinson CV, Dobson CM: **Observation of sequence specificity in the seeding of protein amyloid fibrils.** *Protein Sci* 2004, **13**:1933-8.
  39. Gordon CL, King J: **Temperature-sensitive mutations in the phage P22 coat protein which interfere with polypeptide chain folding.** *J Biol Chem* 1993, **268**:9358-68.
  40. Teschke CM, King J: **Folding of the phage P22 coat protein in vitro.** *Biochemistry* 1993, **32**:10839-47.
  41. Speed MA, Wang DI, King J: **Multimeric intermediates in the pathway to the aggregated inclusion body state for P22 tail-spike polypeptide chains.** *Protein Sci* 1995, **4**:900-8.
  42. Speed MA, Morshead T, Wang DI, King J: **Conformation of P22 tailspike folding and aggregation intermediates probed by monoclonal antibodies.** *Protein Sci* 1997, **6**:99-108.
  43. Speed MA, Wang DI, King J: **Specific aggregation of partially folded polypeptide chains: the molecular basis of inclusion body composition.** *Nat Biotechnol* 1996, **14**:1283-7.
  44. Kopito RR: **Aggresomes, inclusion bodies and protein aggregation.** *Trends Cell Biol* 2000, **10**:524-30.
  45. Steffan JS, Kazantsev A, Spasic-Boskovic O, Greenwald M, Zhu YZ, Gohler H, Wanker EE, Bates GP, Housman DE, Thompson LM: **The Huntington's disease protein interacts with p53 and CREB-binding protein and represses transcription.** *Proc Natl Acad Sci USA* 2000, **97**:6763-6768.
  46. Rajan RS, Illing ME, Bence NF, Kopito RR: **Specificity in intracellular protein aggregation and inclusion body formation.** *Proc Natl Acad Sci USA* 2001, **98**:13060-5.
  47. Pallares I, Vendrell J, Aviles FX, Ventura S: **Amyloid fibril formation by a partially structured intermediate state of alpha-chymotrypsin.** *J Mol Biol* 2004, **342**:321-31.
  48. Litvinovich SV, Brew SA, Aota S, Akiyama SK, Haudenschild C, Ingham KC: **Formation of amyloid-like fibrils by self-association of a partially unfolded fibronectin type III module.** *J Mol Biol* 1998, **280**:245-58.
  49. Villegas V, Zurdo J, Filimonov VV, Aviles FX, Dobson CM, Serrano L: **Protein engineering as a strategy to avoid formation of amyloid fibrils.** *Protein Sci* 2000, **9**:1700-1708.
  50. Fandrich M, Fletcher MA, Dobson CM: **Amyloid fibrils from muscle myoglobin.** *Nature* 2001, **410**:165-166.
  51. Schein CH: **Optimizing protein folding to the native state in bacteria.** *Curr Opin Biotechnol* 1991, **2**:746-50.
  52. Carrio MM, Villaverde A: **Construction and deconstruction of bacterial inclusion bodies.** *J Biotechnol* 2002, **96**:3-12.
  53. Wetzel R: **Ideas of order for amyloid fibril structure.** *Structure* 2002, **10**:1031-6.
  54. Oberg K, Chrnyk BA, Wetzel R, Fink AL: **Nativelike secondary structure in interleukin-1-beta inclusion bodies by attenuated total reflectance FT-IR.** *Biochemistry* 1994, **33**:2628-2634.
  55. Georgiou G, Valax P, Ostermeier M, Horowitz PM: **Folding and aggregation of TEM  $\beta$ -lactamase: analogies with the formation of inclusion bodies in *Escherichia coli*.** *Protein Sci* 1994, **3**:1953-1960.
  56. Przybycien TM, Dunn JP, Valax P, Georgiou G: **Secondary structure characterization of beta-lactamase inclusion bodies.** *Protein Eng* 1994, **7**:131-136.
  57. Ami D, Bonocchi L, Cali S, Orsini G, Tonon G, Doglia SM: **FT-IR study of heterologous protein expression in recombinant *Escherichia coli* strains.** *Biochim Biophys Acta* 2003, **1624**:6-10.
  58. Ismail AA, Mantsch HH, Wong PT: **Aggregation of chymotrypsinogen: portrait by infrared spectroscopy.** *Biochim Biophys Acta* 1992, **1121**:183-8.
  59. Goloubinoff P, Mogk A, Zvi AP, Tomoyasu T, Bukau B: **Sequential mechanism of solubilization and refolding of stable protein aggregates by a chaperone network.** *Proc Natl Acad Sci U S A* 1999, **96**:13732-7.
  60. Umetsu M, Tsumoto K, Ashish K, Nitta S, Tanaka Y, Adschiri T, Kumagai I: **Structural characteristics and refolding of in vivo aggregated hyperthermophilic archaeon proteins.** *FEBS Lett* 2004, **557**:49-56.
  61. Parsell DA, Kowal AS, Singer MA, Lindquist S: **Protein disaggregation mediated by heat-shock protein Hsp104.** *Nature* 1994, **372**:475-8.
  62. Ben-Zvi AP, Goloubinoff P: **Mechanisms of disaggregation and refolding of stable protein aggregates by molecular chaperones.** *J Struct Biol* 2001, **135**:84-93.
  63. Carrio MM, Villaverde A: **Role of molecular chaperones in inclusion body formation.** *FEBS Lett* 2003, **537**:215-21.
  64. Gozu M, Lee YH, Ohhashi Y, Hoshino M, Naiki H, Goto Y: **Conformational dynamics of beta(2)-microglobulin analyzed by reduction and reoxidation of the disulfide bond.** *J Biochem (Tokyo)* 2003, **33**:731-6.
  65. Carmichael J, Chatellier J, Woolfson A, Milstein C, Fersht AR, Rubinsztein DC: **Bacterial and yeast chaperones reduce both aggregate formation and cell death in mammalian cell models of Huntington's disease.** *Proc Natl Acad Sci U S A* 2000, **97**:9701-5.
  66. Bucciantini M, Giannoni E, Chiti F, Baroni F, Formigli L, Zurdo J, Taddei N, Ramponi G, Dobson CM, Stefani M: **Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases.** *Nature* 2002, **416**:507-11.
  67. Ventura S, Zurdo J, Narayanan S, Parreno M, Mangués R, Reif B, Chiti F, Giannoni E, Dobson CM, Aviles FX, Serrano L: **Short amino acid stretches can mediate amyloid formation in globular proteins: the Src homology 3 (SH3) case.** *Proc Natl Acad Sci U S A* 2004, **101**:7258-63.
  68. Kelly JW: **The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways.** *Curr Opin Struct Biol* 1998, **8**:101-106.
  69. Tjernberg LO, Näslund J, Lindqvist F, Johansson J, Karlström AR, Thyberg J, Terenius L, Nordstedt C: **Arrest of beta-amyloid fibril formation by a pentapeptide ligand.** *J Biol Chem* 1996, **271**:8545-85.
  70. Patino M, Liu JJ, Glover JR, Lindquist S: **Support for the prion hypothesis for inheritance of a phenotypic trait in yeast.** *Science* 1996, **273**:622-626.
  71. Tenidis K, Waldner M, Bernhagen J, Fischle W, Bergmann M, Weber M, Merkle ML, Voelter W, Brunner H, Kapurniotou A: **Identification of a penta- and hexapeptide of islet amyloid polypeptide (IAPP) with amyloidogenic and cytotoxic properties.** *J Mol Biol* 2000, **295**:1055-1071.
  72. Wurth C, Guimard NK, Hecht MH: **Mutations that reduce aggregation of the Alzheimer's A $\beta$ 42 peptide: an unbiased search for the sequence determinants of A $\beta$  amyloidogenesis.** *J Mol Biol* 2002, **319**:1279-90.
  73. McParland VJ, Kad NM, Kalverda AP, Brown A, Kirwin-Jones P, Hunter MG, Sunde M, Radford SE: **Partially unfolded states of beta-2-microglobulin and amyloid formation in vitro.** *Biochemistry* 2000, **39**:8735-8746.
  74. Legname G, Baskakov IV, Nguyen HO, Riesner D, Cohen FE, DeArmond SJ, Prusiner SB: **Synthetic mammalian prions.** *Science* 2004, **305**:673-676.
  75. Lopes DH, Colin C, Degaki TL, de Sousa AC, Vieira MN, Sebollela A, Martinez AM, Bloch C Jr, Ferreira ST, Sogayar MC: **Amyloidogenicity and cytotoxicity of recombinant mature human islet amyloid polypeptide (rIAPP).** *J Biol Chem* 2005, **279**:42803-42810.
  76. Fink AL: **Protein aggregation: folding aggregates, inclusion bodies and amyloid.** *Fold Des* 1998, **3**:9-23.
  77. de la Paz ML, Goldie K, Zurdo J, Lacroix E, Dobson CM, Hoenger A, Serrano L: **De novo designed peptide-based amyloid fibrils.** *Proc Natl Acad Sci USA* 2002, **99**:16052-16057.
  78. Izard J, Parker MW, Chartier M, Duché D, Baty D: **A single amino acid substitution can restore the solubility of aggregated colicin A mutants in *Escherichia coli*.** *Protein Eng* 1994, **7**:1495-1500.
  79. Wetzel R, Perry LJ, Veilleux C: **Mutations in human interferon gamma affecting inclusion body formation identified by a general immunochemical screen.** *Biotechnology (NY)* 1991, **9**:731-737.
  80. Krueger JK, Stock J, Schutt CE: **Evidence that the methylesterase of bacterial chemotaxis may be a serine hydrolase.** *Biochim Biophys Acta* 1992, **1119**:322-326.
  81. Wetzel R, Chrnyk BA: **Inclusion body formation by interleukin-1b depends on the thermal sensitivity of a folding intermediate.** *FEBS Lett* 1994, **350**:245-248.
  82. Nieba L, Honegger A, Krebber C, Pluckthun A: **Disrupting the hydrophobic patches at the antibody variable/constant domain interface: improved in vivo folding and physical char-**

- acterization of an engineered scFv fragment. *Protein Eng* 1997, **10**:435-444.
83. Chrunyk BA, Wetzel R: **Breakdown in the relationship between thermal and thermodynamic stability in an interleukin-1 beta point mutant modified in a surface loop.** *Protein Eng* 1993, **6**:733-738.
  84. Azriel R, Gazit E: **Analysis of the minimal amyloid-forming fragment of the islet amyloid polypeptide. An experimental support for the key role of the phenylalanine residue in amyloid formation.** *J Biol Chem* 2001, **276**:34156-34161.
  85. Salmons M, Malesani P, De Gioia L, Gorla S, Bruschi M, Molinari A, Della Vedova F, Pedrotti B, Marrari MA, Awan T, Bugiani O, Forloni G, Tagliavini F: **Molecular determinants of the physicochemical properties of a critical prion protein region comprising residues 106-126.** *Biochem J* 1999, **342**:207-214.
  86. Conway KA, Lee SJ, Rochet JC, Ding TT, Williamson RE, Lansbury PT Jr: **Acceleration of oligomerization, not fibrillization, is a shared property of both alpha-synuclein mutations linked to early-onset Parkinson's disease: Implications for pathogenesis and therapy.** *Proc Natl Acad Sci USA* 2000, **97**:571-576.
  87. Esler WP, Stimson ER, Ghilardi JR, Lu YA, Felix AM, Vinters HV, Mantyh PW, Lee JP, Maggio JE: **Point substitution in the central hydrophobic cluster of a human beta-amyloid congener disrupts peptide folding and abolishes plaque competence.** *Biochemistry* 1996, **35**:13914-13921.
  88. Barghorn S, Zheng-Fischhofer Q, Ackmann M, Biernat J, von Bergen M, Mandelkow EM, Mandelkow E: **Structure, microtubule interactions, and paired helical filament aggregation by tau mutants of frontotemporal dementias.** *Biochemistry* 2000, **39**:11714-11721.
  89. Wigley WC, Stidham RD, Smith NM, Hunt JF, Thomas PJ: **Protein solubility and folding monitored in vivo by structural complementation of a genetic marker protein.** *Nature Biotechnol* 2001, **19**:131-136.
  90. Sirangelo I, Malmo C, Casillo M, Mezzogiorno A, Papa M, Irace G: **Tryptophanyl substitutions in apomyoglobin determine protein aggregation and amyloid-like fibril formation at physiological pH.** *J Biol Chem* 2002, **277**:45887-45891.
  91. Hammarstrom P, Sekijima Y, White JT, Wiseman RL, Lim A, Costello CE, Altland K, Garzuly F, Budka H, Kelly JW: **D18G transthyretin is monomeric, aggregation prone, and not detectable in plasma and cerebrospinal fluid: a prescription for central nervous system amyloidosis?** *Biochemistry* 2003, **42**:6656-6663.
  92. West MW, Wang W, Patterson J, Mancias JD, Beasley JR, Hecht MH: **De novo amyloid proteins from designed combinatorial libraries.** *Proc Natl Acad Sci USA* 1999, **96**:11211-11216.
  93. Wang W, Hecht MH: **Rationally designed mutations convert de novo amyloid-like fibrils into soluble monomeric beta-sheet proteins.** *Proc Natl Acad Sci USA* 2002, **99**:2760-2765.
  94. de la Paz ML, Serrano L: **Sequence determinants of amyloid fibril formation.** *Proc Natl Acad Sci U S A* 2004, **101**:87-92.
  95. Fernandez-Escamilla AM, Rousseau F, Schymkowitz J, Serrano L: **Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins.** *Nat Biotechnol* 2004, **22**:1302-6.
  96. Linding R, Schymkowitz J, Rousseau F, Diella F, Serrano L: **A comparative study of the relationship between protein structure and beta-aggregation in globular and intrinsically disordered proteins.** *J Mol Biol* 2004, **342**:345-53.
  97. DuBay KF, Pawar AP, Chiti F, Zurdo J, Dobson CM, Vendruscolo M: **Prediction of the absolute aggregation rates of amyloidogenic polypeptide chains.** *J Mol Biol* 2004, **341**:1317-26.
  98. Vallejo LF, Rinas U: **Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins.** *Microb Cell Fact* 2004, **3**:11.
  99. Sorensen HP, Mortensen KK: **Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli.** *Microb Cell Fact* 2005, **4**:1.
  100. Maloney AP, Callan SM, Murray PG, Tuohy MG: **Mitochondrial malate dehydrogenase from the thermophilic, filamentous fungus Talaromyces emersonii.** *Eur J Biochem* 2004, **271**:3115-26.
  101. Bertoldo C, Armbrecht M, Becker F, Schafer T, Antranikian G, Liebl W: **Cloning, sequencing, and characterization of a heat- and alkali-stable type I pullulanase from Anaerobranca gottschalkii.** *Appl Environ Microbiol* 2004, **70**:3407-16.
  102. Fang TY, Hung XG, Shih TY, Tseng WC: **Characterization of the trehalosyl dextrin-forming enzyme from the thermophilic archaeon Sulfolobus solfataricus ATCC 35092.** *Extremophiles* 2004, **8**:335-43.
  103. Idicula-Thomas S, Balaji PV: **Understanding the relationship between the primary structure of proteins and its propensity to be soluble on overexpression in Escherichia coli.** *Protein Sci* 2005, **14**:582-92.
  104. Yan G, Cheng S, Zhao G, Wu S, Liu Y, Sun W: **A single residual replacement improves the folding and stability of recombinant cassava hydroxynitrile lyase in E. coli.** *Biotechnol Lett* 2003, **25**:1041-7.
  105. Calloni G, Zoffoli S, Stefani M, Dobson CM, Chiti F: **Investigating the effects of mutations on protein aggregation in the cell.** *J Biol Chem* 2005, **280**:10607-13.
  106. Fan D, Li Q, Korando L, Jerome WG, Wang J: **A monomeric human apolipoprotein E carboxyl-terminal domain.** *Biochemistry* 2004, **43**:5055-64.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

