UAB

Universitat Autònoma
de Barcelona

Regulation of recombinant protein solubility and conformational quality in *Escherichia coli*

PhD Thesis

Elena Garcia i Fruitós

Facultat de Biociències

Departament de Genètica i de Microbiologia

Director: Dr. Antonio Villaverde i Corrales

2008

Regulation of recombinant protein solubility and conformational quality in *Escherichia coli*

Tesi doctoral

Elena Garcia Fruitós

2008
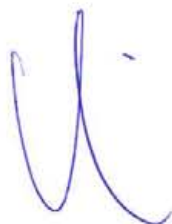
Departament de Genètica i de Microbiologia

Programa de doctorat en Biotecnologia

Regulation of recombinant protein solubility and conformational quality in *Escherichia coli*

Memòria presentada per Elena Garcia i Fruitós, per optar al grau de doctor, per la Universitat Autònoma de Barcelona.

Vist i plau del director de la tesi,

Dr. Antonio Villaverde Corrales

Elena Garcia i Fruitós

Bellaterra, Març de 2008

Als meus pares,

al Marcel i a l'avi Joan,

I.Contents

Contents

II.Introduction

All the processes that take place in a cell require one or more proteins, meaning that they are essential components of life.

Proteins are macromolecules consisting of amino acid units, all of them being constructed with combinations of only 20 amino acids. The primary structure of a protein molecule is determined by the sequence of amino acids connected by peptide bonds forming a polypeptide chain. Once the amino acid chain is synthesized, the protein folds by a physical process that might be eventually assisted by other proteins, reaching its characteristic three-dimensional structure that is the final, functional conformation. However, although it is known that all the proteins must properly fold into their correct native conformation to be functional, their final conformation cannot be predicted from their primary amino acid sequence, being protein folding mechanisms one of the most challenging problems in biology today.

Many proteins of relevant industrial or medical value are produced in low amounts in their natural sources. However, at the end of the seventies, the development of recombinant DNA technologies opened a new promising era for protein production in high amounts for both research and industrial applications. This had a tremendous impact, for example, in many areas of medicine as a tool to produce new drugs for the treatment of diseases and genetic disorders. Genetic engineering permits the introduction of the encoding genes of the protein of interest into recipient cells, where these genes are positioned downstream of regulable promoters in movable genetic elements, mainly plasmids. Under suitable conditions, these transgenic cells acting as protein production bio-factories would be expected to act as unlimited and inexpensive source of rare, highly valuable proteins not only for proteomics and structural functional genomics[1] but also for large-scale preparative purposes. The quality as well as the quantity of the produced recombinant protein is greatly influenced by the chosen biological cell system.

Bacteria have been the most commonly used organisms for protein production, specially the enterobacteria *Escherichia coli*, not only for the low cost of the used processes, but also for its fast growth. Generally, in *Escherichia coli*, the rather small host cell proteins can fold properly, adopting a native, biological active conformation. However, when producing heterologous proteins, specially those with eukaryotic or

viral origin, important obstacles appear during the protein production process: a) in most cases, the protein is produced in a non functional conformation; b) sometimes the formed product is toxic for the cell; c) the protein often results proteolytically degraded[2]; d) the product is accumulated as an insoluble, non-functional protein aggregates, known as inclusion bodies[3].

Therefore, even though the important advantages of the use of bacteria as a expression system, *Escherichia coli* also presents some drawbacks, such as its inability to carry most of the post-transcriptional modifications, often required for eukaryotic protein function, the lack of a secretion mechanism to release the protein to the medium, and the inability to create an oxidative environment to facilitate disulfide bond formation required to achieve the final, functional structure of some proteins. Therefore, this leads to the production of proteins which are not always suitable for immediate use. This means that, to date, many proteins have been excluded from the biotechnological and pharmaceutical market because they cannot be produced in high yields as soluble and active products.

To avoid protein folding problems encountered in bacteria under overexpression conditions, mainly secretion and post-transcriptional modifications, alternative host cells, such as yeast, filamentous fungi, mammalian or insect cells, have been explored. Nevertheless, an enormous number of deficiencies in these systems such as difficulty of genetic manipulation, low productivity and high costs, shows that these organisms are not ideal for this aim and that, even when bacteria show some obstacles in the production process and often this system has to be optimized for specific products, it is, in most of the cases, the best choice.

II.1 *Escherichia coli* as a cell factory for recombinant proteins

Two of the most important advantages of the use of the versatile, gram-negative bacteria *Escherichia coli*, as a cell factory for the production of proteins of therapeutic or commercial interest, are the possibility to produce high levels of recombinant proteins easily and through a low cost process. This microorganism has been extensively used since early recombinant DNA times and, hence, nowadays, the wide spectrum of cloning and gene expression vectors, mutant strains and mutagenic events

as well as the important knowledge of bacterial metabolism, allow genetic tailoring and the optimization of the culture conditions in each situation, permitting scale-up the process when necessary. For routine protein expression, *Escherichia coli* BL21, K12 and their derivatives are the most commonly used.

II.1.1 Key elements in recombinant protein production

There are several key elements regarding the strategy design for protein production of recombinant proteins in *Escherichia coli* such as the promoter system and the codon usage, among others.

II.1.1.1 Plasmid

A bacterial plasmid is a small circular dsDNA molecule, a species of nonessential extrachromosomal DNA that replicates autonomously as a stable component of the cell. The regulation of plasmids differs considerably from the regulation of chromosomal replication. However, the machinery involved in the replication of plasmids is similar to that of chromosomal replication. The plasmid copy number can vary between one to several hundred per cell, as well as its size, that can occur in a range from one to several hundred kilobases.

A number of central characteristics such as origin of replication (ori), antibiotic resistance marker, transcriptional promoters, translational initiation regions (TIRs) as well as transcriptional and translational terminators, are essential in the design of recombinant expression systems such as plasmids.

II.1.1.1.1 *Replicon*

The replicon contains the origin of replication which can support the autonomous replication of the plasmid. The origin of replication is essential not only for replication but also as a key element regulating the plasmid copy number. Most of the plasmids used in recombinant protein expression replicate by the ColE1 or the p15A replicons[4].

II.1.1.1.2 *Resistance markers*

An antibiotic resistance marker is a gene that confers resistance to antibiotics and it is a tool in molecular biology to confirm the uptake of genes (e.g., carried by plasmids) by

bacteria. The most common resistance markers in recombinant expression plasmids confer resistance to ampicillin, kanamycin, chloramphenicol or tetracycline. Ampicillin is a β-lactam antibiotic and the ampicillin resistance is accomplished by expression of β-lactamase from the *bla* gene. β-lactamase enzyme is secreted to the periplasm, where it catalyses the hydrolysis of the β-lactam ring. Kanamycin is inactivated by aminoglycoside trasnferases, chloramphenicol by the *cat* gene product, a chloramphenicol acetyl transferase, and a wide spectrum of genes confer resistance to tetracycline.

For the production of human therapeutic proteins, the use of other antibiotic resistances should be considered to avoid the potential of human allergic reactions.

II.1.1.1.3 Promoters

A promoter is a regulatory region of DNA located upstream of a gene (approximately 10 to 100 bp upstream of the ribosome-binding side), providing a control point for gene transcription. The promoter contains specific DNA sequences that are recognized by proteins known as transcription factors and these factors, at the same time, are recognized by RNA polymerase, the enzyme involved in the RNA synthesis of the gene coding region. Therefore, promoters represent critical elements to direct the level of gene transcription.

To choose a suitable promoter for recombinant protein expression, several criteria should be considered. First of all, recombinant protein expression plasmids require a strong transcriptional promoter to obtain high levels of protein production; the promoter should be capable to promote the production of protein in excess, desirably representing up to 10-30 % or more of the total cellular protein[5]. Secondly, the use of a promoter with a low level of basal expression is extremely important when the protein produced may be detrimental for the host cell. Furthermore, a non complete repression of the protein expression could cause plasmid instability, a slower cell growth rate and, in consequence, a decrease of the recombinant protein production[5]. Moreover, if possible, the induction of recombinant gene expression should be simple and cheap[6]. Thermal (e.g., λ p_L) and chemical (e.g., trp) induction of gene expression has been commonly used in large-scale protein production. In *Escherichia coli* lactose-

regulated promoters such as lac, tac and trc are very well-known[7]. The induction of these promoters could be achieved by adding the lactose analogous isopropyl-β-D-thiogalactopyranoside (IPTG), a powerful chemical inducer, specially used in basic research; however, in some cases it is not recommended due to its toxicity and cost when high levels are required[8]. Another widely used IPTG-inducible expression system is based on the T7 RNA polymerase[7,9]. An interesting inexpensive alternative for heterologous production in *Escherichia coli* is the L-arabinose inducible P_{BAD} promoter[7,10]. On the other hand, thermal induction is also effective, although it might not always be suitable because the heat-shock response is stimulated. For instance, the phage promoter p_L is regulated by the temperature-sensitive repressor CI857, being completely functional at 30°C, but inactivated at 42°C, therefore allowing gene expression[11-13]. Other types of promoters used, although not as frequently as the previously described, are cold-responsive promoters, to produce recombinant proteins at low temperature (avoiding in that way protein aggregation)[14] and pH-responsive promoters[15,16].

II.1.1.2 Stability of messenger RNA

Messenger RNA (mRNA) stability and the efficiency of mRNA translation determine gene expression levels and, consequently, protein production. The average half-life of mRNA in *Escherichia coli* at 37°C ranges from seconds to 20 minutes[17], depending on mRNA stability. RNA decay in cells is mediated by RNases such as endonucleases and 3´exonucleases.

Even though there are some nucleotidic sequences that improve mRNA stability, there is not a "universal stabiliser". Therefore, the best option for a given protein and host strain must be determined for each case.

II.1.1.3 Codon usage

It is also important to consider the codon usage of the target gene and of the cell host, because the lack of coincidence can affect both the quantity and the quality of the target protein. Amino acids are encoded by more than one codon and each organism shows a non-random usage of synonymous codons[18,19], being tRNA population in the

cytoplasm a reflection of the codon bias[20]. Usually, heterologous proteins are enriched with codons that are rare in *Escherichia coli* and this would be a problem to produce them efficiently in this microorganism. The main problems observed in recombinant protein overexpression in *Escherichia coli* due to the codon usage patterns are translational stalling at positions requiring incorporation of amino acids coupled to minor codon tRNAs[21], premature translation termination, translation frameshift, and amino acid misincorporation[22-25].

Although some studies show that modifications of culture conditions might solve this obstacle by shifting the codon usage bias[7], it has also been reported that, upon these modifications, most of the rare tRNAs remain unchanged[20]. In addition, two alternative strategies have been developed to minimize the effect of the preferential codon bias in *Escherichia coli*. The first approach relies on site-directed mutagenesis of the target sequence, without modifying the encoded protein product, to generate host-specific codons coexpression[26-29]. The second approach consists of rare tRNA coexpression from accompanying plasmids[17,30-32].

II.1.1.4 Strain

The genetic background is highly important for recombinant protein production. Nowadays, advantageous strains are available for a number of individual applications (section II.4).

II.2 Protein folding and misfolding

Proteins have multiple and indispensable roles in all the organisms, therefore being the most abundant molecules in biological systems other than water. In the *Escherichia coli* cytoplasm, for example, one protein chain is released from the ribosome every 35 seconds[33].

The most fundamental example of self-assembly process in living systems is protein folding, the process through which unfolded polypeptide chains convert into tightly folded compact structures with biological functions. The pioneering studies of protein folding were done by Anfinsen between the 1960s and 1970s, in which he described that the amino acid sequence of a protein encodes its functional three-dimensional structure[34]. Although nowadays, the underlying mechanism by which this complex process takes place is becoming better understood, by using physical and chemical techniques as well as computational methods, there is still much to do. In fact, understanding the mechanism by which proteins fold in detail is the first step on the path to resolve the molecular mechanism of conformational diseases such as Alzheimer, type II diabetes and Creutzfeldt-Jakob, among others, and to develop successfully protein-based medicines.

Although there are just 20 different types of amino acids, the number of sequences that can be generated through their combination is much greater than the number of atoms in the universe. However, natural proteins are just a small representation of these combinations, therefore becoming a selected group of molecules carefully chosen by the evolution. There seems to be a common mechanism for the folding of all proteins, irrespective of their sequence or native structure[35]. *In vivo*, proteins can fold whilst the nascent chain is still attached to the ribosome[36], either in the cytoplasm after release from the ribosome or in specific compartments such as endoplasmic reticulum (ER)[37]. Nevertheless, the important question is how the necessary information for proper folding is encoded in the amino acid sequence and how a protein can reach the unique native state in a finite time[38]. If we consider the total number of possible conformations that are accessible to a polypeptide, we realise that finding a particular structure in all the conformational space would take an

astronomical length of time, being many orders of magnitude greater than the real time required for proteins to fold. This paradox , known as Levinthal paradox[38], has recently been solved with the development of the so called "new view"[39], in which folding is described as a stochastic search of conformational space rather than as a series of mandatory structural transitions[40-43]. This view involves also the "energy landscape" concept, which describes the free energy of a given polypeptide chain as a function of its conformational properties (figure 1), being the native state of a protein that with the lowest free energy. In essence, the inherent fluctuations in the conformation of an incompletely folded polypeptide enable the contact even of residues located at very different positions in the amino acid sequence. Therefore, as correct (native-like) interactions are more stable than non-native ones, this search mechanism is able to find the structure with the lowest energy[43,44]. As the native state is

Unfolded states (10^{16})

Free energy

Number of native interactions

Transition state (10^3)

Number of residue contacts

Native structure (1)

Figure 1. Schematic energy landscape for protein folding (adapted from Sven Frokjaer and Daniel E. Otzen (2005) *Nature Reviews Drug Discovery* vol. 4(4): 298-306 and Christopher M. Dobson (2004) *Methods* vol. 34(1): 4-14).

approached, the conformational space accessible to the polypeptide chain is reduced[41]. The fundamental mechanism of protein folding involves the formation of a folding-nucleus of residues in the protein, around which the remainder structure condenses rapidly[45].

Small (< 100 residues), single-domain proteins do not require many partially folded intermediates to reach a native conformation, and usually only extreme conditions

unfold them[46]. In contrast, folding of large, multidomain proteins involve several intermediates prior to the formation of the completely folded native state. They usually fold in modules that finally interact forming the fully native structure[47-49] and, additionally, often require the assistance of folding modulators.

The term "misfolding" is used to describe the process that results in a protein acquiring a sufficient number of persistent non-native interactions to affect its overall architecture and/or its properties in a biologically significant manner[50]. Misfolded and incompletely folded molecules are susceptible to aggregate, due to the exposure of hydrophobic regions that are buried in the native state[51] (figure 2). To avoid aggregation, cells of living organisms have auxiliary factors, including folding catalysts that accelerate rate-limiting steps and molecular chaperones, that assist protein folding[37,52]. Moreover, the cell quality control mechanism targets for destruction any protein molecule that has not folded correctly (figure 2).

Figure 2. Conventional model of protein folding, aggregation and proteolysis. A chain newly synthesized on a ribosome may fold to a native state, can aggregate or can be proteolysed. In living systems, environmental conditions and the quality control system highly regulate the transition between the different states.

II.2.1 Misfolding, aggregation and derived diseases

Protein folding is a complex process that plays a key role within the cell; hence, any failure in this process will give rise to aggregation. Protein aggregation can not only cause major economical and technical problems in biotechnology and pharmaceutical industries, but also leads to conformational diseases such as cystic fibrosis, type II diabetes, prion diseases and amyloid diseases (table 1)[53,54]. Inclusion bodies are

aggregates formed in bacteria (section II.5), while aggresomes (microtubule-dependent inclusion bodies) and amyloid fibrils appear in eukaryotic cells.

The most dramatic examples of the occurrence of these structures are undoubtedly disease states[53]. Probably the best characterized conformational pathologies are neurodegenerative diseases involving the deposition of amyloid fibrils or plaques such as Alzheimer's disease, Parkinson's disease, Huntington's disease and the transmissible spongiform encephalopathies (prion diseases such as bovine spongiform encephalopathy, Mad Cow disease and Creutzfeldt-Jakob disease in humans). Depending on the disease, protein deposits can be formed in brain, in vital organs such as liver and spleen, or in skeletal tissue. Recently, it has been found that proteins without any connection with disease states can also form amyloid fibrils[55-63], meaning that this ability is a generic property of polypeptide chains[64]. Though the ability of proteins to form such fibrils may be common, the propensity to convert into this structure can vary for each polypeptide chain, depending on the composition and sequence[65,66]. Therefore, evolution has avoided, in general terms, sequences that tend to promote formation of amyloids.

Regardless of the amino acid sequence of a protein, amyloid fibrils are similar and composed of a repetitive assembly of β-sheets oriented perpendicularly to the fibril axis and linked by hydrogen bonds[67], being β-strand formation a powerful driving force for aggregation, in contrast to that of α-helices[68]. This β-structure can be observed by using techniques like IR spectroscopy, solid state NMR and circular dichroism (CD). Amyloid deposits are long, unbranched, fibrous and twisted structures of a few nanometers in diameter that show a characteristic "cross-beta" X-ray fiber diffraction pattern[67] as well as a characteristic optical behaviour, such as green birefringence, on binding certain dye molecules such as Congo red[69].

The first phase in amyloid formation seems to involve the formation of rather disordered, soluble oligomeric structures known as pre-fibrillar species. Then, these structures are transformed to species with more distinctive morphologies often called protofilaments or protofibrils. In the last few years, it has been described that pre-fibrillar aggregates of proteins associated with neurological diseases can be highly damaging to cells, in contrast to the mature fibrils being relatively benign[70,71].

Table 1. Protein-misfolding diseases.

Disease	Misfolded protein	Protein aggregates
Alzheimer's disease	β-peptide	Extracellular plaques
Parkinson's disease	Tau/α-synuclein	Intracellular plaques and Lewy bodies
Prion diseases	Prp^{SC}	Prionic plaques
α_1-antitrypsin deficiency	α_1-antitrypsin	Hepatocyte endoplasmic reticulum aggregates
Type II diabetes	Amylin (Islet Amyloid Polypeptide –IAPP-)	Pancreatic amyloid plaques
Marfan syndrome	Fibrillin-1	Microfibrillar aggregates
Retinitis pigmentosa	Rhodopsin	Endoplasmic reticulum aggregates
Amyotrophic lateral sclerosis	SOD1	Bunina bodies
Huntington's disease	Huntingtin containing polyglutamines	Nuclear and cytoplasmic aggregates
Cystic fibrosis	Cystic fibrosis transmembrane conductance regulator (CFTR)	Endoplasmic reticulum aggregates

II.3 Quality control machinery

Under stress conditions, such as high temperatures and recombinant protein overproduction, the protein quality control machinery is stimulated. This represents a natural cellular defense devoted to prevent protein misfolding and accumulation of aggregating proteins, based mainly on the activity of chaperones and proteases.

II.3.1 Chaperones

The term chaperone was first used to describe an activity associated with nucleoplasmin in *Xenopus oocytes*[72]. From that moment on, the term has been expanded to include more than 20 protein families with a central role in the conformational quality control of the proteome[73-75]. Specifically, molecular chaperones are a group of structurally diverse proteins highly conserved in all three kingdoms of life which form a complex network to assist proper protein folding, prevent their deposition and dissolve deposits of misfolded proteins[76-81]. Even though chaperones are constitutively expressed under physiological conditions, many of them are upregulated under stress conditions; this regulation is mainly due to the sigma factor σ^{32} encoded by *rpoH* gene[82]. Since chaperone abundance increases in cells upon thermal stress, these molecules have been traditionally named heat shock proteins (Hsps)[83]. However, despite this historical link, it should be clarified that not all chaperones are heat shock proteins and not all heat shock proteins are chaperones.

Molecular chaperones (table 2) can be divided into 3 functional subclasses based on their mechanism of action:

"Folding" chaperones

This group mediates the folding of their substrates in an ATP-dependent process. This chaperones increase the yield of properly folded proteins but not the rate of folding. In the *Escherichia coli* cytoplasm the three chaperone systems involved in this process are trigger factor (TF) (section II.3.1.1), DnaK-DnaJ-GrpE (section II.3.1.2) and GroEL-GroES (section II.3.1.3).

"Holding" chaperones
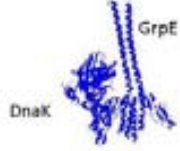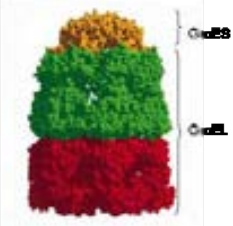
This second subclass of chaperones maintains proteins partially folded on its surface to await availability of folding chaperones upon stress situation, preventing polypeptides from aggregation[81,84,85]. The most extensively characterized bacterial holdases are IbpA and IbpB (section II.3.1.4), both belonging to the small Hsp family[86]. Hsp31 and Hsp33 are also classified as holdases; while Hsp31 binds early unfolding intermediates in times of severe stress, thereby preventing overloading of the DnaK-DnaJ-GrpE system[87,88], Hsp33 manages oxidative protein misfolding[89].

"Disaggregating" chaperones

Among this group, bacterial ClpB (section II.3.1.2.1) is the best characterized chaperone. It has a secondary role, assisting refolding and promoting the solubilisation of proteins that have become aggregated as a result of stress[90,91]. This chaperone acts in cooperation with DnaK and IbpAB chaperones[92-94].

Besides the functions mentioned above, it has also been described that chaperones can act as neurodegenerative repressors[78,95] and as modulators of fibril formation[80]. Moreover, chaperones work in cooperation with proteases (section II.3.2), another important component of the quality control machinery that mediates the degradation of proteins that cannot be properly folded[96].

Table 2. Elements of the *Escherichia coli* protein quality control machinery.

Escherichia coli member	Family	Cofactors	Function	ATP requirement	Structure
Trigger factor (TF)	PPIase		Holding, PPIase	No	(1)
DnaK	Hsp70	DnaJ GrpE	Folding, holding, disaggregation, regulation	Yes	(2)
ClpB	Hsp100		Disaggregation	Yes	(3)
GroEL	Hsp60	GroES	Folding	Yes	(4)
IbpA IbpB	Small heat shock proteins		Holding	No	(5)
ClpA	Hsp100	ClpP	Folding, degradation	Yes	(6)

ClpX	Hsp100	ClpP	Folding, degradation	Yes	
					(7)

(1) Deported from a RasMol representation according to the coordinates given by Lars Ferbitz *et al.* (2004) *Nature* vol. 431(7008): 590-6.

(2) Deported from a RasMol representation according to the coordinates given by Celia J. Harrison *et al.* (1997) *Science* vol. 276(5311): 431-5.

(3) Deported from a RasMol representation according to the coordinates given by Sukyeong Lee *et al.* (2003) *Cell* vol. 115(2): 229-40.

(4) Adapted from Zhaohui Xu *et al.* (1997) *Nature* vol. 388(6644): 741-50.

(5) Deported from a RasMol representation according to the coordinates given by Christopher K. Kennaway *et al.* (2005) *Journal of Biological Chemistry* vol. 280(39): 33419-25.

(6) Deported from a RasMol representation according to the coordinates given by Fusheng Guo *et al.* (2002) *Journal of Biological Chemistry* vol. 277(48): 46743-52.

(7) Deported from a RasMol representation according to the coordinates given by Dong Young Kim and Kyeong Kyu Kim (2003) *Journal of Biological Chemistry* vol. 278(50): 50664-70.

II.3.1.1 Trigger factor

The ribosome-associated trigger factor (TF) is a three-domain protein that binds to the large subunit of the ribosomes, in the vicinity of the peptide exit site, to interact with nascent polypeptides and protect them[97]. Trigger factor exhibits both peptidyl-prolyl *cis/trans* isomerase (PPIase) and chaperone activity[98-100]. Therefore, this molecular chaperone supports the *de novo* folding by binding to nascent chains (figure 3). Once the substrate is released, trigger factor can cycle back to the ribosome, waiting for the next substrate[97].

II.3.1.2 Hsp70 system: DnaK, DnaJ, GrpE

Hsp70 family is encoded in all living organisms' genomes, being one of the most conserved proteins in the evolution[101-103]. There are three Hsp70 (DnaK, HscA and HscC) in *Escherichia coli*, being DnaK the best characterized. DnaK is the most general chaperone and the centre of the multichaperone network (figure 3), having different roles: 1) mediates ATP-dependent unfolding, 2) prevents aggregation, 3) stabilises the substrates for refolding by GroELS (section II.3.1.3) [100-106], 4) participates in proteolysis[107,108], cooperating in some cases with Lon protease (section II.3.2.1), 5)

folds new synthesized polypeptides[37, 109], 6) solubilises protein aggregates in cooperation with ClpB (section II.3.1.2.1) and Ibps (section II.3.1.4)[81,104, 110-114], 7) protects proteins against oxidative damages,[115,116] and 8) negatively regulates the heat shock response[117] minimizing the expression of the heat shock σ^{32} regulon, which encodes the main chaperones and proteases, including DnaK itself[118-120].

DnaK has an N-terminal ATPase domain of 44 kDa, two β-sheets forming a substrate binding site and a C-terminal domain of 27kDa that can interact with partner proteins to modulate chaperone function[121,122]. DnaK partners are a J-domain protein (JDP) co-chaperone, belonging to the Hsp40 family, termed DnaJ[123] and a nucleotide exchange factor (NEF) named GrpE. When ATP is bound, DnaK binds the substrate through weak, hydrophobic interactions and hydrogen bonds[124]. Upon ATP hydrolysis, there is a conformational change that stabilise substrate binding[100]. In this process, the co-chaperone DnaJ has an important role accelerating the rate of ATP hydrolysis, while the co-chaperone GrpE accelerates the exchange of ADP with ATP, leading to the substrate ejection. The released polypeptide may reach a native conformation, undergo additional cycles in the chaperone system until it folds, or is transferred to GroEL-GroES (section II.3.1.3)[125]. The system formed by DnaK chaperone and DnaJ and GrpE co-chaperones is usually abbreviated KJE.

II.3.1.2.1 ClpB

ClpB is an ATP-dependent molecular chaperone, member of Hsp100 family. Specifically, ClpB is a "disagregase" that works in cooperation with DnaK-DnaJ-GrpE reverting aggregation[94,100,126,127] (figure 3). This molecular chaperone has an important role, in cooperation with DnaK, in dissolving protein aggregates, reducing the aggregate size and exposing hydrophobic surfaces[2,104,111,113]. However, the full recovery of renaturated proteins cannot be achieved until the partially unfolded substrate is transferred from ClpB to DnaK[81,104,111,128,129].

II.3.1.3 Hsp60 system: GroEL and GroES

GroEL is a bacterial chaperonine of approximately 60 kDa that belongs to Hsp60 family. This molecular chaperonine, essential for growth at all temperatures[130], prevents

aggregation [131], acting as the main folder element in the chaperone network[132]. GroEL is formed by two stacked homoheptameric rings which define a central cavity in which incompletely folded polypeptides up to around 60 kDa[133] can properly fold (figure 3). When ATP is bound a conformational change takes place[134], rendering GroEL competent to bind the 10 kDa accessory protein GroES[37]. The GroES-bound undergo a second conformational movement of GroEL, allowing the folding of the non-native polypeptide. If the protein has not reached the native state, a further round of binding and attempted folding follows.

II.3.1.4 Small heat shock proteins

The best defined small heat shock proteins (sHsps) in bacteria have been Inclusion Bodies Proteins (Ibps), proteins regularly associated to inclusion bodies[135] and commonly organised in large oligomeric structures[86,136]. There are two different types of Ibps encoded on a single-operon[135,137], IbpA and IbpB of 14 and 16 kDa size, respectively. Although IbpA is insoluble and IbpB is mainly soluble, IbpB comigrates to the insoluble fraction when produced with IbpA[138]. Even though Ibps function is not well understood, they seem to recognise hydrophobic patches in unfolded proteins, remaining bound to these polypeptides, protecting them from aggregation, until they are transferred to DnaK or GroEL for refolding[138-141]. Moreover, it has been recently described that IbpA and IbpB facilitate the disaggregation and refolding activity of ClpB[94] (figure 3).
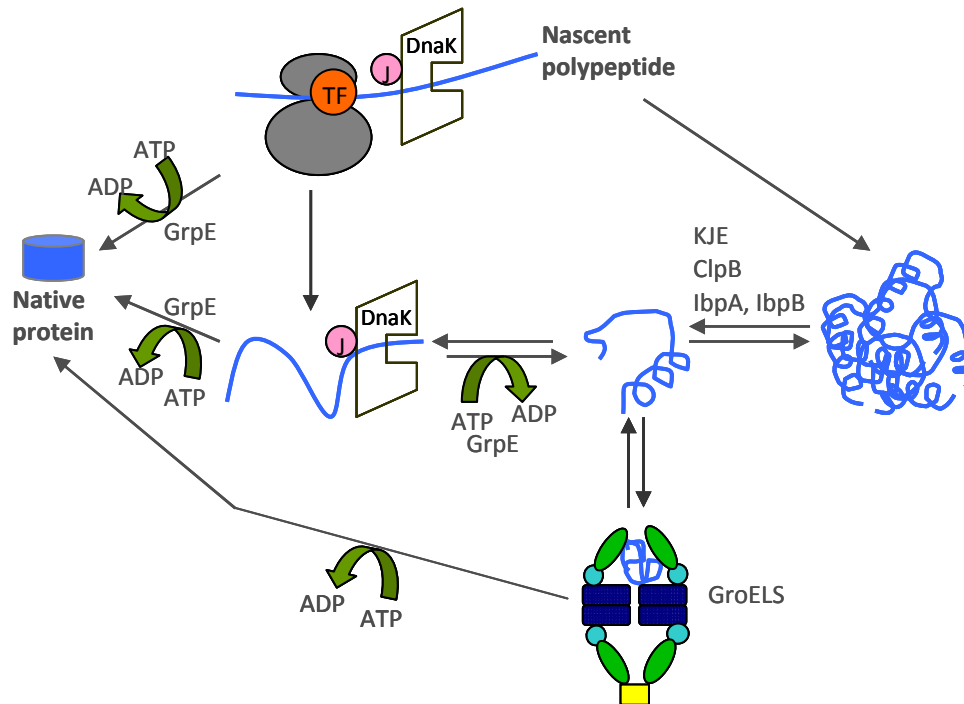
Figure 3. Trigger factor (TF) binds to nascent polypeptides. KJE (DnaK, DnaJ (J) and GrpE) and GroELS (GroEL and GroES) systems assist protein intermediates to reach their native form. The small heat shock proteins IbpA and IbpB, together with ClpB, cooperate with KJE system in the disaggregation process.

II.3.2 Proteases

Proteolysis of misfolded proteins that have failed to reach a native conformation plays a crucial role in the quality control system, preventing the aggregation of abnormal polypeptides as well as allowing the amino acid recycling within the cell. The main proteases of *Escherichia coli* cytoplasm are ClpP and Lon. These heat-shock ATP-depedent proteases recognize hydrophobic surfaces, as chaperones do[142]. Moreover, these cell proteases degrade not only unprotected, misfolded polypeptides localized in the soluble cell fraction[143], but also those found embedded in protein aggregates[144].

II.3.2.1 Lon

Lon is a tetrameric serine protease of 87 kDa subunits containing three functional domains. Its N-terminus is involved in substrate recognition and binding, its central domain is responsible for ATPase activity and its C-terminus domain has proteolytic activity. In addition to being responsible for bulk protein degradation[145,146], Lon also

exerts a regulatory function by degrading a class of proteins that are designed to be unstable.

II.3.2.2 ClpP

ClpP protease is a protein organised as two stacked heptamers of 23 kDa units. Their substrates are folded, misfolded or incompletely synthesized proteins that are targeted for degradation. This protease forms a complex with two members of the Hsp100 family of ATPases (ClpA and ClpX)[147-149] to form a fully-competent degrading machinery. ClpA and ClpX, which are flanking the rings of ClpP, act as molecular chaperones, unfolding proteins in an ATP-dependent manner and translocating substrates into ClpP central channel[150].

II.4 Improving solubility of recombinant proteins

Minimization or prevention of inclusion body formation is an attractive alternative to obtain a high degree of accumulation of soluble product in the bacterial cell. Nevertheless, protein stability and solubility cannot be predicted in advance and the used strategies have not shown the same degree of success for different polypeptides[98,151].

The main strategies used to minimise inclusion bodies formation are those included below (figure 4):

Escherichia coli cytoplasm

- Cultivation at reduced temperatures
- Weak promoters or partial induction
- Optimization of media composition
- Suitable cultivation strategy
- Mutant strains
- Coproduction of folding modulators
- Protein engineering

Inclusion body

Increased amounts of soluble protein
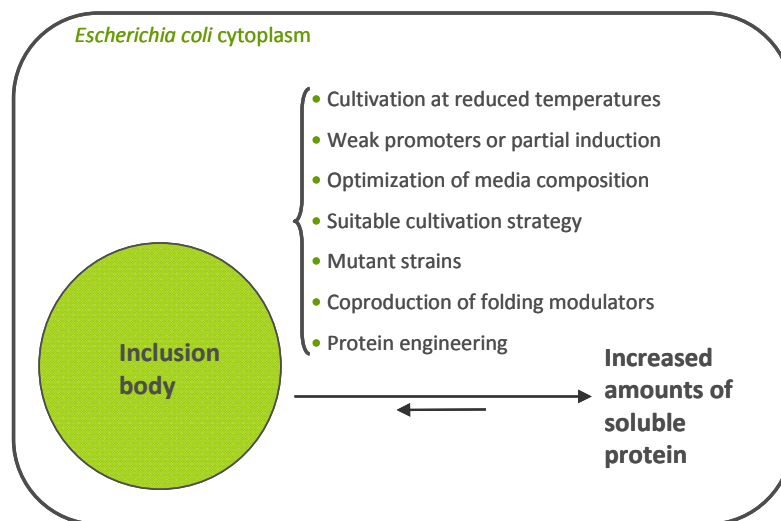
Figure 4. Strategies used to reduce inclusion body formation.

Protein production at low temperatures

A well-known procedure to limit inclusion body formation consists of cultivation at reduced temperatures[152,153]. A wide number of proteins have been successfully produced in a soluble form by following this strategy[154-158]. The use of low temperatures has the advantage of reducing the hydrophobic interactions that contribute to protein misfolding and aggregation[159,160]. Moreover, some heat shock proteases induced during recombinant protein production are poorly active at low temperatures, being protein degradation substantially reduced[157,161]. Altogether, this accounts for the expression of more stable and prone to properly fold proteins at temperatures below the optimal of 37°C for *Escherichia coli* growth[162]. Nevertheless, drawback are also present in the use of this approach; as low temperatures lead to

reduced replication, transcription and translation rates[163], this results in the obtention of reduced yields of the recombinant protein.

Different strategies like a system based on cold-inducible promoters have been developed to optimize the expression of heterologous proteins at low temperatures. The most widely used cold-inducible promoter is *cspA*, which is highly active at low temperature and is well repressed at and above 37°C[164]. Moreover, another system based on the coexpression of chaperones from the psychrophilic bacterium that allows protein expression at 4°C has been recently developed[165].

Weak promoters

Solubility can be improved by using weaker promoters or by using strong promoters under conditions of partial induction[166]. For promoters based on *lac*-derived control elements (e.g., the trc promoter), isopropyl-D-thiolgalactopiranoside (IPTG) concentrations below 100 μM are suitable for partial induction[160]. A weaker promoter or a partially induced promoter leads to a reduction of recombinant protein concentration which favors folding. However, under these conditions, bacterial growth is slower, thus resulting in decreased amount of biomass.

Modification of media composition and cultivation strategies

The composition of growth media can also help minimizing inclusion body formation. By optimising media composition, reduced expression times, increased soluble fraction yield and enhanced biological activity of enzymes have been achieved[162]. Moreover, the folding of certain proteins might require the presence of a specific cofactor, such as metal ions or polypeptide cofactors, in the growth media. Therefore, we can also increase both protein solubility and folding rates after adding such cofactors[167-170].

On the other hand, the election of a suitable cultivation strategy can also help optimising soluble protein production. While in batch cultivation there is a limited control of the growth, fed-batch cultivation permits the regulation of several factors, allowing a real time optimisation of growth conditions[171].

Escherichia coli genetically modified strains

The genetic background is extremely important for recombinant protein expression and, hence, *Escherichia coli* mutant strains have contributed significantly to the soluble expression of recombinant proteins. As recombinant gene products are commonly sensitive to proteolysis, one of the most commonly used strains is BL21(DE3)[17], a strain that can grow in minimal media and that is deficient in ompT and Lon proteases. BL21 derivative mutants have also been extensively used; these derivatives include: BLR, a *recA*⁻ strain, specially useful for stabilisation of target plasmids containing repetitive sequences; origami strains (*trxB*⁻/*gor*⁻), used for production of proteins containing disulfide bonds; Rosetta strains, chosen for the overexpression of a rare tRNA expression vector and C41(DE3) and C43(DE3) strains, frequently utilized for the production of membrane proteins[7,172].

Coexpression of folding modulators

One of the most extensively used approaches to improve the yields of soluble proteins in the *Escherichia coli* cytoplasm involves the coproduction of folding modulators such as molecular chaperones. The chaperone machinery is saturated by the massive protein overproduction, giving rise to inclusion body formation; thus, the production of molecular chaperones can improve protein folding and solubilisation. This strategy was first proven in 1989 with the overexpression of the chaperone system *groELS*, which clearly increased the solubility of Rubisco enzyme produced in *Escherichia coli*[173]. Since then, there have been many successful attempts to produce recombinant proteins in a soluble form by coexpression of chaperones[98,174-177]. However, the selection of the suitable chaperone(s) for a determined target protein is still a trial-and-error approach and so far the best results have been obtained by the production of several sets of folding modulators[178].

In some cases chaperone overproduction is beneficial not only regarding higher solubility, but also regarding the total concentration of the target protein[179,180]. There are, however, restrictions that limit routine application of this approach[181] and there are many, and often unpublished studies, in which this strategy fails to improve protein solubility.

Fusion tags

A different strategy used for maximizing protein solubility consists of protein engineering through directed mutation or gene fusion[162,182-187]. The use of affinity tags (proteins or peptides that are fused to the protein of interest) is not only exploited in recombinant protein purification, but also to improve protein yield, to prevent proteolysis and to increase solubility *in vivo*[5,17,188]. Among the most potent solubility enhancing tags, we can find *Escherichia coli* maltose binding protein (MBP) and *Escherichia coli* N-utilizing substance A (NusA), this last one also showing high expression levels[172]. In addition, the use of small, synthetic peptide tags called SET (solubility enhancing tag) has also been successful for some proteins[189], becoming a new promising approach because of the small size of these tags (< 30 amino acids), which may lead to less folding interference making all proteins suitable for structural studies without the need of removing the tag.

However, this technique also has some disadvantages, such as the eventual need to remove the tag and the question of whether the protein of interest remains in its native state and biologically functional once the tag has been removed. Moreover, it should be added that the engineering approach has to be adapted for each particular polypeptide.

II.5 Inclusion bodies

The formation of amorphous proteinaceous granules in *Escherichia coli* was first described in cells growing in presence of the amino acid analog canavanine[190]. These granules were deposits of abnormal cell proteins and they were not surrounded by any defined superficial layer. This observation was thought to be irrelevant. However, after the implementation of DNA recombinant technology it was seen that, under overexpression conditions, these granules were the rule rather than the exception[191].

Inclusion bodies are frequently observed during the production of heterologous proteins in transformed microorganisms[191] and, nowadays, are still one of the main bottlenecks in the use of bacteria as cell factories for foreign protein production, being an important obstacle for their obtention in a soluble form. Moreover, inclusion body formation has not been associated to particular protein sequences, the expression system or the host cells, to date not being possible to predict precisely the aggregation propensity of a new protein[192]. This unpredictability defines a trial-and-error landscape of recombinant protein production in bacteria. However, in recent times, it is becoming more evident that protein sequence modulates aggregation, being the aggregation propensity determined by some particular regions of the polypeptide known as aggregation "hot spots"[193,194]. In this regard, in the last years, different computational approaches have been developed to predict the aggregation propensity of polypeptides[195-199], becoming useful tools with biotechnological applications.

From another point of view, these aggregates are considered an interesting and convenient model to investigate the molecular basis of protein folding and aggregation. Understanding the aggregation process not only is important for the production of functional proteins, but also to solve the molecular mechanism of a growing number of conformational diseases that have become the object of intense research and that are not well characterized.

II.5.1 Structure and morphology

Inclusion bodies are *in vivo*-formed aggregates found in the insoluble cell fraction and not surrounded by any superficial layer. These aggregates have been observed in both prokaryotic and eukaryotic cells, although they are especially common in bacteria producing mammalian or viral proteins. Normally, there is one inclusion body per cell, eventually two, and they are porous particles with a high level of hydratation[200]. They have been usually seen as refractile particles with a size between 0.4 μm and 2 μm by optical microscopy (figure 5A)[201,202] and as electron-dense aggregates lacking defined structure by transmission electron microscopy (figure 5B)[202]. Even though these aggregates are commonly found in the bacterial cytoplasm, usually in a polar situation, secreted protein aggregates are frequently localized in the bacterial periplasmic space[203]. Once purified, these particles are generally ovoid or cylindrical (figure 5C and 5D), with differently smoothed surfaces[201,202,204]. The Inclusion body density is higher than that of many other cellular components; therefore, they can be easily separated by high-speed centrifugation after cell disruption[205].

Figure 5. A) Phase contrast optical micrograph of *Escherichia coli* strain overproducing VP1LAC protein; the arrows indicate inclusion bodies (Elena García-Fruitós *et al.* not published). B) Transmission electronic microscopy micrograph of *Escherichia coli* strain overproducing VP1LAC protein; the arrow indicates an inclusion body (Maria del Mar Carrió *et al.* 2005). C) Scanning electron microscopy micrographs of purified inclusion bodies of VP1GFP protein (Elena García-Fruitós *et al.* not published). D) Scanning electron microscopy micrographs of purified inclusion bodies of VP1GFP protein (Elena García-Fruitós *et al.* not published).

Although inclusion bodies have in general been described as highly stable aggregates resistant to proteases, this view has changed in the last ten years. Specifically, through scanning electron microscopy, it has been demonstrated that untreated inclusion

bodies and those treated with trypsin show a completely different aspect. While non treated inclusion bodies exhibit a rod-shaped morphology with a smooth surface, those treated with trypsin show fragmentation lines that sub-particulate the aggregate, proving that inclusion bodies are not only sensitive to proteolytic attack but also that this proteolysis is not surface restricted[204].

Even though the recombinant protein is the main inclusion body component[182,201,206], its percentage in these aggregates is variable[143]. Besides recombinant protein, minor amounts of other cell proteins, components of the translational apparatus, lipids as well as DNA and RNA fragments, that have unespecifically precipitated, can also be found in variable concentration among inclusion body components[206,207]. Additionally, folding-assistant proteins, such as small heat shock proteins IbpA and IbpB, have been identified as regular components of inclusion bodies[135]; moreover, the presence of GroEL and DnaK chaperones in inclusion bodies has also been reported[208]. This last fact suggests that the presence of components of the quality control system could be linked to cell-mediated protein processing.

In addition, aggregation as amyloid fibrils and as inclusion bodies shows coincident traits such as a predominant β-sheet architecture and the ability to bind amyloid-tropic dyes (table 3).

Table 3. Functional and structural coincident traits of inclusion bodies resembling those of amyloid fibrils.

Feature	Inclusion bodies
High purity of the aggregates	Carrió et al. (1998) *FEMS Microbiol. Lett..* vol. 169(1): 9-15
Intermolecular, cross β-sheet organisation	Przybycien et al. (1994) *Protein Eng* vol. 7(1):131-6, Carrió et al. (2005) *J. Mol. Biol.* vol. 347(5):1025-37
Amyloid-tropic dye binding	Carrió et al. (2005) *J. Mol. Biol.* vol. 347(5):1025-37
Sequence specific aggregation	Speed et al. (1996) *Nature Biotech* vol. 14(10):1283-7, Carrió et al. (2005) *J. Mol. Biol.* vol. 347(5):1025-37
Chaperone-modulated aggregation	Thomas et al. (1997) *Appl Biochem Biotechnol.*vol. 66(3):197-238, Carrió et al. (2003) *FEBS Letters* vol. 537(1-3):215-21
Seeding-driven aggregation	Carrió et al. (2005) *J. Mol. Biol.* vol. 347(5):1025-37
Chaperones localized in the aggregates	Carrió et al. (2005) *J. Bacteriol* vol. 187(10):3599-601
Aggregation hot spots	Tenidis K et al. (2000) *J. Mol. Biol.* Vol. 295(4):1055-71
Protection from cytotoxicity	González-Montalbán et al. (2005) *J. Biotechnol* vol. 118(4):406-12

II.5.2 Formation

Inclusion body formation takes place under high rate protein synthesis, specifically, when misfolded or partially folded proteins with exposed hydrophobic surfaces cross-interact, and this drives to the deposition of these folding intermediates[209].

To better understand inclusion body formation, first of all, we need to consider that overproduction itself is enough to increase nascent polypeptide chain concentration. Also, under these high level production conditions, chaperones and proteases (section II.3) can become limiting factors, allowing misfolded proteins to escape from the quality control machinery. Therefore, these events, as well as the inability of bacteria to support all post-translational modifications that some proteins need to properly fold or to complete their folding process[210], induce the formation of these inactive aggregates called inclusion bodies. In addition, the sequence of the polypeptide itself can determine its fractioning between the soluble and insoluble cell fractions.

It was generally believed that inclusion bodies were inert, unorganised and proteolytic resistant aggregates that undergo a parsimonious accumulation growth by product accumulation. On the contrary, recent studies have shown that these aggregates are highly dynamic, plastic, selforganised proteinaceous structures with an important biotechnological potential[143]. To better understand inclusion body dynamism, it should be considered that their formation might be the result of two processes, namely the continuous incorporation of precipitated protein around nucleation cores and the release of solubilised and refolded proteins. Consequently, there is an equilibrium between protein income and outcome that, under an actively protein overproducing situation, is unbalanced towards protein aggregation driving to inclusion body growth. However, the situation is inverted in absence of *de novo* protein synthesis or in ageing cultures[114,143], where inclusion bodies are disintegrated in few hours, increasing the soluble protein and its biological activity. The fact that inclusion bodies are not a dead-end process in the quality control network of the cell, but transient reservoirs of misfolded polypeptide chains, can have a significant impact on protein recovery when using bacteria as a cell factory.

It has also been demonstrated that inclusion bodies are not only dynamic structures that regularly grow through the productive phase of the culture but also that they aggregate by a nucleation-like mechanism forming cores that are mutually exclusive[201,211]. Additionally, a significant number of observations suggest that aggregation occurs by a highly specific seeding process rather than by nonspecific coaggregation[51,211-214].

The increase of cytoplasmic inclusion bodies volume upon high-level expression may span in some cases the entire diameter of an *Escherichia coli* cell.

II.5.3 Inclusion bodies as a source of soluble proteins

In a production context, inclusion bodies represent a major obstacle, its formation being difficult to prevent. However, from an opposite point of view, the formation of inclusion bodies can be advantageous, being a convenient source of relatively pure polypeptides that can be recovered, denatured and refolded to obtain functional, soluble protein. When big amounts of resolubilised protein have to be obtained, inclusion body formation can be favoured by using high temperature culture conditions[215], protease-deficient strains[216], or through the fusion of aggregation-prone domains[217].

II.5.3.1 *In vivo* recovery of inclusion body proteins

As mentioned above, after arresting protein synthesis both the amount and the activity of the soluble protein increase, concomitantly with a progressive volumetric reduction of the inclusion bodies[126]. Thus, inclusion bodies can be solubilised *in vivo* after protein synthesis arrest and at least a fraction of the solubilised protein might reach its native conformation with fully biological activity. When protein production is arrested, availability of molecular chaperones in the cell increases, facilitating inclusion body protein release and folding[178]. The specific profile of this process could vary between proteins[204].

II.5.3.2 *In vitro* recovery of inclusion body proteins

In vitro protein refolding from purified and solubilised inclusion bodies is a method of choice for functional protein recovery in a high number of cases[218], especially when the yield of soluble protein is poor. In these cases, inclusion body protein recovery is better than that obtained by purifying the soluble fraction[215]. Although *in vitro* protein refolding is not an easy procedure[219] and not always a reasonable solution, the careful optimization and the continuous developing and improvement of this strategy done in the last decades has allowed the efficient refolding of even complex proteins at an industrial scale[218]. Therefore, the conversion of insoluble protein into soluble and correctly folded products has become an interesting solution in a variety of situations, avoiding the use of mammalian cells or other expression systems. The general strategy used for protein recovery includes 3 basic steps: 1) isolation and washing of inclusion bodies (table 4 and section II.5.3.2.1), 2) solubilisation of aggregated protein (table 4 and section II.5.3.2.2) and 3) refolding of solubilised proteins, the last one being the most critical step (table 4 and section II.5.3.2.4).

There is a wide number of protocols that can be used for this purpose and if the right method is chosen high yields of recombinant protein with a native conformation can be obtained. It has been described that after a global optimisation of the whole protocol, the yield of soluble, biologically active protein that can be recovered usually range between 10 and 50 %; however, in some cases higher recoveries have been obtained. Nevertheless, time-consuming efforts are required to optimize the protocols usually case-by-case, most of the times conducing to irregular results that are not always profitable for the industry market[220].

II.5.3.2.1 *Inclusion body purification*

The first step to finally obtain soluble, active protein from inclusion bodies is the purification of these aggregates. Inclusion bodies are generally easily separated from the cell debris, because of its high density, by centrifugating cells extracts obtained by using ultrasonication or lysozyme treatment of bacterial cells for small volumes, French press for medium volumes and high pressure homogenization for large volumes[221,222]. Cell lysis should be maximally effective to prevent cosedimentation of cell debris.

However, even after an optimal cell lysis, some impurities such as lipids, DNA and outer membrane proteins will remain coprecipitated with inclusion bodies[223]. Therefore, several washing steps should be applied to remove these contaminants and improve the homogeneity of inclusion body components. These washing steps can include DNase treatment to remove DNA, low concentrations of chaotropic agents such as urea or guanidinium chloride (Gdn-HCl) and detergents like Triton X-100 or SDS to remove membrane proteins as well as other contaminants. It has to be taken into account that urea and Gdn-HCl must be used carefully, because in some cases they can completely disintegrate inclusion bodies[218,221].

After cell disruption, an alternative method that can be used to isolate inclusion bodies is sucrose gradient centrifugation[224,225]. This method, though technically more complex, can offer a more precise separation and, consequently, very pure inclusion body recovery from *Escherichia coli* lysates. Using an appropriate protocol for inclusion body isolation and purification, 95 % pure preparations can be obtained.

II.5.3.2.2 Protein solubilisation

After inclusion body isolation, these aggregates have to be solubilised to obtain unfolded proteins that in a final step (section II.5.3.2.4) would be refolded into an active conformation. Protein denaturalization is usually achieved using high concentrations (6-8 M) of strong denaturants and chaotropic reagents, such as urea or Gdn-HCl [221,226], both of them promoting the complete unfolding of the protein and the disruption of intermolecular interactions[227,228]. Gdn-HCl is in general preferable to urea, because, although more expensive, it is a stronger chaotrop that can even denature inclusion body proteins resistant to urea solubilisation. Additionally, urea solubilisation is pH-dependent and, moreover, solutions can contain isocyanate which can lead to the carbamylation of the free amino groups of the polypeptides[229,230]. In consequence, when urea is the method of choice, agents containing free amino acid groups should be added in the solubilisation solution.

Alternatively to Gdn-HCl or urea, the cationic surfactant cetyltrimethylammonium chloride (CTAC)[231], SDS detergent or N-acetyltrimethylammonium bromide (CTAB)[232] can be used for inclusion body solubilisation. On the other hand, it has also been

described that extreme pH, irrespective of the presence of denaturants, can also favour protein solubilisation[233]; however, this approach is not suitable for any protein because acid cleavage and irreversible chemical modifications like deamination or alkaline desulfuration of cysteine residues can occur.

Proteins containing cysteines also need the presence of reducing agents in the solubilisation buffer to avoid inter- and intramolecular disulfide bonds formed by air oxidation, keeping the cysteines in a reduced state. Specifically, low molecular weight thiols reagents such as β-mercaptoethanol, 1,4-dithiotreitol (DTT) or 1,4-dithioerythritol (DTE), in combination with chaotrops, are used to reduce disulfide bonds by thiol-disulfide exchange. Chelating agents like ethylene diamine tetraacetic acid (EDTA) or ethylene glycol tetraacetic acid (EGTA) are frequently added to the solubilisation buffer to avoid metal-catalysed air oxidation of cysteines. Once the solubilisation step is finalised, all the reducing substances have to be completely removed, because they could interfere the final refolding step (section II.5.3.2.4).

Finally, it should be mentioned that proteases that could coaggregate in inclusion bodies should be removed to avoid proteolytic degradation during solubilisation[57,207].

II.5.3.2.3 Purification of solubilised inclusion body proteins

Although a rigorous purification of the solubilised protein is not a prerequisite for efficient protein refolding and proteins normally can be directly renatured after their release, in some cases[234-237] a purification step is required before any refolding attempt. This is recommended when the protein of interest represents less than 2-5 %[238] of the total cell protein or less than 60 % of the total inclusion body protein[236].

II.5.3.2.4 Protein refolding

After solubilisation, protein refolding is necessary to finally generate soluble and biologically active proteins. This renaturation step, which can require from seconds to days, is usually the most delicate and, although the optimization of the inclusion body purification, washing and solubilisation is important, the whole efficiency of the process is determined by this final step.

Renaturation of denatured polypeptides to recover biologically active forms is accomplished by using *in vitro* folding techniques by which the excess of denaturants and reducing agents are removed, creating a favorable environment that allows the spontaneous protein folding[219,226]. However, *in vitro* refolding frequently leads to a poor recovery of properly refolded protein and this is mainly due to the spontaneous formation of aggregates in this step. Specifically, upon *in vitro* protein folding, refolding, a first-order reaction involving intramolecular interactions, and aggregation, a second or higher-order reaction involving intermolecular interactions, are continuously competing, being the second one in advantage under high protein concentrations[159,221,227]. The loss of the protein secondary structure in denatured proteins leads to the exposure of hydrophobic patches that can interact and tend to aggregate[239]. In this regard, low initial concentrations of denatured proteins reduce aggregation[227,240,241]. In addition, there are some small molecules such as L-arginine (0,4-1 M)[242], specific cofactors like Zn^{2+} or Ca^{2+}, glycerol, sucrose, polyethylene glycol, denaturant agents at low concentration like urea (1-2M) and Gdn-HCl (0,5-1,5 M) and detergents such as Chaps, SDS, CTAB and Triton X-100 that are useful to reduce the aggregation propensity, improving the solubility and stability of the unfolded protein and thus the final efficiency of the renaturation[226,228]. All these substances are easy to remove, except detergents whose removal requires a special treatment after protein refolding.

The more frequently used techniques for the recovery of native proteins are dilution, dyalisis, diafiltration, size exclusion chromatography and immobilisation onto a solid support[228]. Again the optimum process cannot be deduced from the protein molecular properties and it has to be studied case by case.

Direct dilution

Dilution of the solubilisation buffer (where unfolded proteins are concentrated) in a renaturation buffer, is the simplest procedure to allow protein refolding. Frequently, renaturation has to be performed at low concetrations (1-10 µg/ml) to favour protein refolding instead of aggregation[221,226]. However, this might become a problem in large-scale volumes, because this dilution method needs huge refolding vessels, large amounts of buffers and additional concentration steps after renaturation, altogether

meaning a high production cost. To solve this problem, a dilution method variant called pulse renaturation, consisting in the continuous or by-pulse addition of denatured protein in the refolding buffer, has been developed[240,243,244]. Therefore, large amounts of protein can be refolded in the same buffer vessel, recovering high final concentrations of properly folded protein.

Dyalisis and diafiltration

Dyalisis[245] and diafiltration[246] are buffer-exchange methods by which the denaturant is removed gradually, what prompts to recover also gradually a native environment in which the polypeptides can properly fold. The major drawbacks of these methods are the presence of aggregates in the buffer and the non-specific protein binding to the membrane, both reducing the efficiency of the process. However, despite these problems, after optimising conditions, high refolding yields can be obtained[221].

Size exclusion chromatography

The denaturant can be removed at the same time that proteins are refolded by using size exclusion chromatography (SEC)[247,248 249]. The denaturation solution is injected through a column where the proteins are trapped and after the refolding buffer addition, proteins refold and are eluted, recovering a high concentration of protein compared to the dilution method. With this method, protein-protein interactions are minimised due to the protein immobilisation onto a solid support and, in consequence, the aggregation is substantially reduced. A variation of this methodology consists of an equilibration of the SEC column with a gradient of denaturants[241].

Although the yield of refolded protein obtained by SEC is comparable to that recovered by using direct dilution method, SEC offers an important advantage over direct dilution refolding: the refolded proteins, due to the fractioning by size, are free from contaminants, aggregates and also unfolded proteins[228].

Solid support immobilisation

In this case, a stable protein-matrix complex is formed through non-specific or specific interactions and, as in the chromatographic approach, the aggregation is minimised because the proteins are spatially isolated[228]. Once the proteins have been refolded, they are detached from the matrix. Some examples of this method are: unfolded proteins that bind to ion exchange resins[250-253], proteins with a histidine tag which bind

to immobilised metal ions[250,254,255] and proteins fused to a glutathione S-transferase fragment which binds to an anion exchange matrix[256].

Additionally to the above-mentioned techniques, it has to be considered that the efficiency of *in vitro* folding processes can be affected by physical and chemical parameters such as pH, ionic strength, pressure, temperature and buffer components[221]. Moreover, protein engineering and the addition of chaperones can eventually enhance the yield of *in vitro* refolded protein from inclusion bodies.

Proteins containing disulfide-bonds in their native state need special requirements after an *in vitro* folding process[226]. Even though the number of possible disulfide bonds formed increases dramatically with the number of cysteines present in a protein, the correct disulfide bond formation is biased by the free energy gained upon the acquisition of the correct native conformation. To obtain a successful product, an appropriate oxidative environment has to be established to reach the complete renaturation of these disulfide-bonded proteins. Initially, the cysteine oxidation was performed with molecular oxygen or air, in a reaction catalysed by Cu^{2+} ions[257,258]. However, despite the low cost of this procedure the efficiency is very low. Nowadays, the oxidation can also be achieved with low molecular weight thiols in reduced and oxidized forms ("oxido-shuffling" reagents), such as reduced and oxidized gluthation (GSH, GSSG), cysteine/cystine, cysteamine/cystamine, di-β-hydroxyethil and disulfide/2-mercaptoethanol at a concentration of 5-15 mM. Apart from the above-mentioned methods, S-sulfonation is used to improve refolding of proteins containing disulfide bonds[218,259], since this reagent prevents the aggregation along the refolding process.

Refolding followed by purification is generally preferable as some of the high molecular weight aggregates can be co-purified along with contaminants in a single step.

Table 4. General strategy for *in vitro* recovery of inclusion body proteins.

	Procedure	Comments

Step 1
 	
Isolation and washing of inclusion bodies
	•Cell disruption:
	  -Ultrasonication or lysozyme treatment (small volumes).
	  -French press (medium volumes).
	  -High pressure homogenization (large volumes).
	•Low speed centrifugation.
	•Removal of contaminants with DNase treatment, low concentrations of chaotropes (urea or Gdn-HCl) and detergents (Triton X-100 or SDS).
		*Lysozyme-EDTA treatment before cell homogenization facilitates cell disruption.
		*Optional: sucrose gradient to improve inclusion body purity.
		*The final preparations have usually around 95 % of purity.

Step 2 **Solubilisation of aggregated protein**	•High concentration (6-8 M) of strong denaturants and chaotropic agents (urea or Gdn-HCl). Alternatively, to urea or Gdn-HCl, cationic surfactants (CTAC, SDS or CTAB) can be used.	*Gdn-HCl is preferable to urea, because urea can carbamylate amino groups and its action is pH-dependent. *Proteins containing cysteines require reducing agents (β-mercaptoethanol, DTT or DTE). Additionally, the use of chelating agents (EDTA or EGTA) avoid oxidation of cysteines.

Step 3
 	
Refolding of solubilised protein
	•Remove the excess of denaturants and reducing agents with one of the following methods:
	  -Dilution.
	  -Dialysis or diafiltration.
	  -Chromatography.
	  -Immobilization onto a solid support.
		*Critical step that can require from seconds to days.
		*The presence of contaminants can reduce the refolding yield.
		*For proteins containing disulfide bonds, refolding buffers should be complemented with redox agents.

II.6 Conformational quality of inclusion body proteins

Pioneering studies by Worrall and coworkers[317] and by Tokatlidis and colaborators[318] suggested, for the first time, that inclusion bodies formed in *Escherichia coli* exhibit biological activity. In this context, some years later, Oberg and coworkers[260] mentioned the occurrence of proteins with native-like structures embedded in inclusion bodies. Therefore, as these works seemed to suggest the presence of protein with activity and native structure localized in inclusion bodies, we decided to further explore this phenomenon in detail to better characterize the conformational quality of proteins embedded in inclusion bodies formed in *Escherichia coli*.

II.7 Model proteins

In this section, the model proteins used in this study are shown in detail. The specific properties that make these proteins suitable as conformational quality reporters will be discussed.

II.7.1 β-galactosidase

Escherichia coli β-galactosidase (β-D-galactohydrolase, E.C.3.2.1.23.) is the product of the *lac*Z operon and it belongs to the so-called "4/7 superfamily" of glycosyl hydrolases[261,262]. This enzyme hydrolyses lactose into glucose and galactose, allowing *Escherichia coli* to grow on lactose as carbon source.

II.7.1.1 Structure

The 1023 amino acid sequence of the *Escherichia coli* β-galactosidase monomer was determined in 1970[263] and its tridimensional structure shows that it is predominantly composed by α-helix[264]. β-galactosidase is a tetramer of four identical subunits of 116.353 Da each[264] (figure 6), although in some situations 8, 16 or even more monomers can interact and aggregate[265]. These subunits interact by non-covalent unions[264]. Each subunit consists of five domains, the third of which comprises an important part of the active site[266]. It has been described that the enzyme has two contact surfaces, the long interface and the activating interface, being this last one where the active site is localized.

Figure 6. *Escherichia coli* β-galactosidase tridimensional structure (deported from a RasMol representation according to the coordinates given by Raymond H. Jacobson *et al.* (1994) *Nature* vol. 369(6483): 761-6).

II.7.1.2 Enzymatic activity

β-galactosidase enzyme has two catalytic activities, always generating products with the same stereochemistry as the starting substrate. First, it hydrolyzes the β1,4 bond of the disaccharide lactose, generating glucose and galactose, two compounds that the cell can use as a carbon source[267]. Secondly, it also catalyses the isomerisation of lactose (galactosil-β-D-(1,4)-galactopyranose) to allolactose (galactosil-β-D-(1,6)-glucopyranose), which is the natural inducer of the *Escherichia coli lac* operon[268].

The residues that form the active site are mainly from domain 3, although loops from the first, second and fifth domain also contribute[266]. The amino acids that play a key role during the catalysis have been identified by using mutagenesis techniques: Glu-461[269], Tyr 503[270], Glu-537[271], His-540[272] and Gly-794, this last one being the residue that allows the enzyme-substrate union[273]. Each subunit also has a Mg^{2+} or Mn^{2+} binding-site. Although not indispensable, these cofactors can increase significantly the β-galactosidase activity[274]. Additionally, Na^+ and K^+ can also improve the hydrolase activity of this enzyme. The optimal conditions for β-galactosidase activity are 35^oC and pH=7.2, although the enzyme is stable between pH=6 and 8[275].

Interestingly, β-galactosidase can be split in two non functional peptides, N-terminal (α-fragment) and C-terminal (ω-fragment). Both fragments can spontaneously reassemble into a functional enzyme in a process called complementation. Specifically, α-complementation[276] is observed in specific laboratory strains in which the small α-fragment is encoded by a plasmid, while the large ω-fragment is encoded in *trans* by the bacterial chromosome. On the other hand, ω-complementation is the process in which the fragment encoded in the bacterial genome, carrying a deletion in the carboxy-terminus, interacts with the ω-peptide.

II.7.1.3 β-galactosidase applications

β-galactosidase has a rich history in science and has become an invaluable tool for molecular biology and biochemistry[277]. The possibility of using lactose-analogous substrates that upon hydrolysis give colored products has allowed the use of simple, fast and easy colorimetric assays, being the product quantifiable by

espectrophotometry[278]. The most commonly used substrates with chromogenic properties are orto-nitrophenil-β-D-galactopyranoside (ONPG), a compound that once hydrolyzed gives a yellow product, chlorophenol galactopyranoside (CPRG), that gives a red product and 5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside (X-gal), that becomes blue once hydrolyzed. Therefore, due to the easy detection of these products, β-galactosidase has been widely used as a reporter to better understand many biological processes.

Another important characteristic of this enzyme is that the fusion of other proteins at carboxyl or amino termini does not interfere with its enzymatic activity[279,280], being modified β-galactosidases a good model to study fusion proteins. Specifically, the amino terminus of the enzyme has been described as very tolerant to deletions and insertions of high molecular mass domains[276].

To sum up, β-galactosidase is a useful tool for improving recombinant protein production in *Escherichia coli*[281], improving protein stability and minimizing its toxicity[282]. However, in certain cases, the protein is proteolysed in the cell, once the final product is obtained. Although in many of the cases fusion proteins can be used in the chimeric form[283-285], in some situations the presence of β-galactosidase can be undesirable and the homologous domain can be removed by either chemical or enzymatic reactions[286].

II.7.2 Green Fluorescent Protein

Green Fluorescent Proteins (GFP) exist in a variety of marine organisms such as hydrozoa (*Aequorea, Obelia* and *Phialidium*) and anthozoa (*Renilla*). However, *Aequorea victoria GFP* gene is the only one that has been cloned and well characterized.

The GFP of the Pacific Northwest jellyfish *Aequorea victoria* is an accessory protein of the chemioluminiscent calcium binding aequorin protein, and a greenish bioluminescence is emitted upon Ca^{2+} stimulation[287]. The GFP gene contains all the necessary information for its proper folding and neither jellyfish-specific enzyme nor substrates or cofactors are needed to reach its native and functional conformation[288-290].

II.7.2.1 Structure

GFP was first crystallized in 1974, but the crystal structure was solved in 1996[291,292]. Wild-type GFP crystallizes as a monomer[293]. However, *Renilla* GFP is a dimer, only dissociated under denaturing conditions[294].

GFP is a highly stable, proteolytic-resistant single chain protein of 238 amino acid residues and a molecular mass of approximately 28 kDa[291]. This protein is an eleven-stranded β-barrel with a α-helix running up to the axis of the cylinder (figure 7). The β-barrel, a 24 Å-diameter cylinder with a length of 42 Å[292], is the so-called β-can. The chromophore, responsible for its fluorescence and known as "light in the can", is attached to the α-helix and positioned near the geometric center of the cylinder (figure 7). The GFP chromophore is formed by three adjacent residues Serine 65-dehydroTyrosine 66-Glycine 67[295,296], and its protection is so complete that classical fluorescence quenching agents have almost no effect on GFP fluorescence[297].

It has to be considered that almost all the primary sequence is used to build the β-barrel and the axial helix, so it is difficult to reduce the protein size with large deletions. Specifically, the use of deletion analysis has been defined that the amino acids 7-229 are the minimal domain in GFP required for keeping the fluorescence[298].

Figure 7. *Aequorea victoria* Green fluorescent protein tridimensional structure (deported from a RasMol representation according to the coordinates given by Mats Ormö *et al.* (1996) *Science* vol. 273(5280): 1392-5). β-barrel structure in grey and chromophore in green.

II.7.2.2 Chromophore formation

The chromophore (fluorophore) formation is divided in several steps (figure 8). First of all, GFP folds into a nearly native conformation. Then, there is a nucleophilic attack of the amino group of Gly67 onto the carbonyl group of Ser65 followed by dehydratation to achieve the imidazolidinone ring formation. Finally, the C^{alpha}-C^{beta} bond of Tyr66 is

oxidized and only at this stage can the chromophore absorbe blue light and emit fluorescence[299].

It is important to note that GFP cannot become fluorescent in obligate anaerobes, because O_2 is needed for the complete maturation of the protein. Once GFP is mature, O_2 has no further effects[300]. Therefore, the oxidation seems to be the slowest step in GFP maturation.
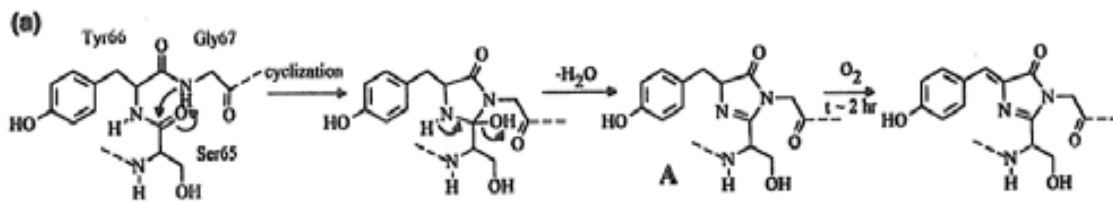
Figure 8. GFP chromophore formation (adapated from Andrew B. Cubitt *et al.* (1995) *Trends in Biochemical Sciences* vol. 20(11): 448-55).

II.7.2.3 Absorbance and fluorescence properties

There is a wide number of GFP variants that has been designed to improve the folding properties, accelerate the chromophore formation, change the emission and excitation wavelengths and/or improve fluorescence emission. These variants are classified in seven groups, based on the chromophore residues[301]. One of these GFP variants is S65T-GFP, a mutant particularly useful in biology in which Serine 65 has been substituted by a Threonine[302]. While wild-type GFP has a major excitation peak at 395 nm and a minor peak at 475 nm, S65T-GFP, due to the amino acid substitution, it has just an excitation peak at 490 nm. Therefore, S65T-GFP mutant combines simple excitation and emission spectra peaking at 490 and 509 nm respectively with sixfold greater brightness than wild-type, a lower sensitivity versus photobleaching (photochemical destruction of the fluorophore) and a fourfold faster chromophore oxidation than the wild-type GFP.

Folding of wild-type GFP is sensitive to temperature and therefore reaching the native structure is inefficient at high temperatures. However it is interesting to point out that, although the temperature restricts protein folding, once the protein has matured properly at low temperatures it is stable and fluorescent at temperatures up to 65°C.

II.7.2.4 GFP applications

β-galactosidase, firefly luciferase and bacterial luciferase have been widely used as reporters in molecular and cell biology. However, all these enzymes require an exogenously added substrate, a fact that can limit their applications in biology. Besides, in the last twenty years, GFP has attracted a lot of attention in biochemistry, cell biology and molecular biology, because it is easy to produce in a broad variety of organisms, including bacteria, fungi, yeasts, plants, insects and nematodes, and because its detection only requires irradiation.

Some of the most important applications of GFP are: a) reporter of gene expression, b) fusion tag to monitor protein dynamics, c) a tool to study protein-protein interactions and even for studying chaperonin-mediated folding, d) pH sensor and e) detection of protease activity[299,303,304]. Specifically, for fusion protein constructions, gene encoding GFP is fused in frame, in carboxy or amino terminii[299], with the other gene of interest, to finally produce it in the organism of interest.

Moreover, there are several GFP-variants such as blue fluorescent protein (BFP), yellow fluorescent protein (YFP) or cyan fluorescent protein (CFP), all of them are obtained through punctual mutations in the chromophore or near the chromophore of GFP.

As a result of the variety of applications, several variants from the original wild type green fluorescent protein (wtGFP) have been developed. Several of these variants have different excitation and emission spectra than wtGFP (table 5). For maximal sensitivity when quantifying these GFPs, it is important to use the most appropriate filter set.

Table 5. Excitation and emission spectra of GFP and GFP variants.

GFP-variant	Excitation maximum wavelength (nm)	Emission maximum wavelength (nm)
wtGFP	395	509
S65T-GFP	490	509
BFP	383	447
CFP	439	476
YFP	514	527

II.7.3 VP1 protein

VP1 is a capsid protein of foot-and-mouth disease virus (FMDV) (figure 9). In particular, different regions of VP1 have been produced as recombinant proteins[305] or reproduced in synthetic peptides[306], as vaccine components in immunization trials. This protein has the two most important antigenic overlapping sites of the virus. In serotype C, the antigenic site is called site A, which is located between residues 138 and 150 within the G-H loop[307]. Site A contains several B-cells epitopes[308-310] including an Arg- Gly-Asp (RGD) tripeptide, used for the virus to bind to cellular receptor[311].
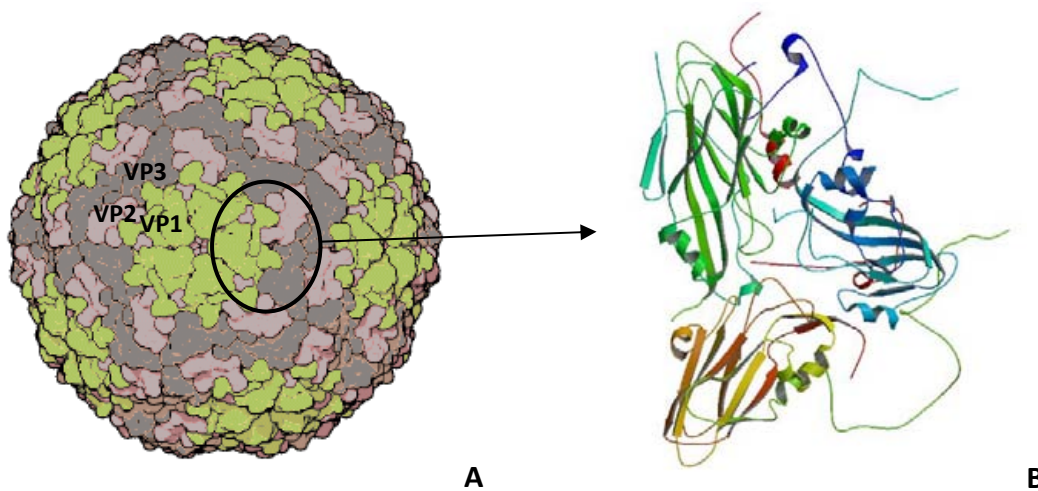
Figure 9. A) Foot and mouth disease virus (FMDV) particle (adapted from David S. Goodsell (2001) The Scripps Research Institute). B) Ribbon protein diagram of FMDV VP1 –blue-, VP2 –green- and VP3 –yellow- (deported from a RasMol representation according to the coordinates given by Elizabeth E. Fry *et al.* (1993) *EMBO Journal* vol. 18(3): 543-54).

II.7.4 VP1LAC and VP1GFP hybrid proteins

One of the most successful applications of β-galactosidase and GFP has been as reporter partners, being GFP specially attracting because its localisation can be easily determined by fluorescence microscopy[312]. On the other hand, foot-and-mouth disease virus (FMDV) VP1 protein has high propensity to aggregate, becoming a very useful tool to study recombinant protein aggregation[143,201]. Hence, in our work, we have used VP1 fused to β-galactosidase and to GFP as a model to study viral protein production in *Escherichia coli*. VP1LAC plasmid encodes the complete VP1 protein (23 kDa) joined to the amino termini of β-galactosidase (116 kDa), a stable fusion protein that forms inclusion bodies[216] (figure 10). The other model protein used, VP1GFP, is a

construct where VP1 has been fused to GFP (28 kDa) (figure 10). Both recombinant proteins, VP1LAC and VP1GFP, preserve the enzymatic activity of the β-galactosidase and the fluorescence of the GFP, respectively.

Figure 10. Schematic representation of VP1GFP and VP1LAC fusion proteins.
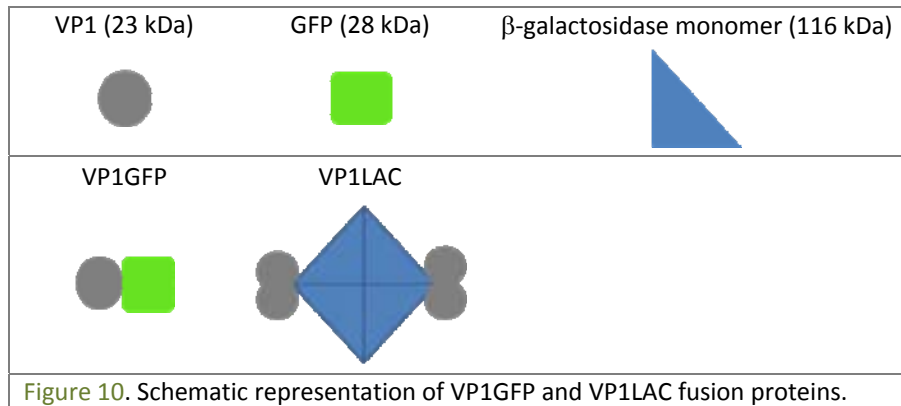
III.Objectives

The purpose of this study is the detailed characterization of the main determining factors involved in the regulation of the conformational quality and solubility of heterologous proteins produced in the *Escherichia coli* cytoplasm. To reach this aim we have adressed the following specific issues:

1. The study of the role of DnaK chaperone concerning both protein activity and stability in *Escherichia coli* cells overproducing recombinant proteins.

2. The comparison of different culture growth conditions to improve protein solubility, stability and biological activity.

3. The exploration of the conformational quality of inclusion body-embedded proteins by:

    3.1. Characterization of inclusion body-linked biological activity.

    3.2. Analysis of the secondary structure of different proteins forming inclusion bodies.

    3.3. Comparison of inclusion body proteins and their soluble versions regarding conformational quality.

4. The evaluation of possible inclusion body applications in the biotechnological industry.

5. The study of the biological role of the bacterial quality control system surveilling the conformational quality of both soluble and inclusion body protein versions.

IV.Results

IV.1 Paper 1

Folding of a misfolding-prone β-galactosidase in absence of DnaK

Elena García-Fruitós, M. Mar Carrió, Anna Arís, Antonio Villaverde

In *Escherichia coli*, DnaK and their cochaperones DnaJ and GrpE have valuable functions in processing folding reluctant protein chains. Even though in cells devoid of functional DnaK the quantity of aggregated polypeptides is higher than in wild type cells, there is still a significant amount of recombinant protein in the soluble cell fraction. In this work, we have explored and compared the status of the soluble VP1LAC protein in both wild type and *dnak* mutant, through the analysis of its enzymatic activity and stability. Specifically, we have analysed VP1LAC gene expression at both lag and late exponential growth phases. This study shows that soluble VP1LAC produced in DnaK$^-$ cells is less stable and less active than when produced in wild type DnaK$^+$ cells. Additionally, the obtained results revealed that a late induction strategy in batch cultures enhances the half-life and the specific activity of soluble VP1LAC, irrespectively of DnaK. Hence, it seems that the minimisation of recombinant protein biosynthesis in the late exponential phase of culture growth favours proper protein folding.

Folding of a Misfolding-Prone β-Galactosidase in Absence of DnaK

Elena García-Fruitós, M. Mar Carrió, Anna Arís, Antonio Villaverde

Institut de Biotecnologia i de Biomedicina and Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona, Bellaterra, 08193 Barcelona, Spain; telephone: 34 935812148; fax: 34 935812011; e-mail: antoni.villaverde@uab.es

Abstract: In absence of chaperone DnaK, bacterially produced misfolding-prone proteins aggregate into large inclusion bodies, but still a significant part of these polypeptides remains in the soluble cell fraction. The functional analysis of the model β-galactosidase fusion protein VP1LAC produced in DnaK⁻ cells has revealed that the soluble version exhibits important folding defects and that it is less stable and less active than when produced in wild-type DnaK⁺ cells. In addition, we have observed that the induction of gene expression at the very late exponential phase enhances twofold the stability of VP1LAC, a fact that in DnaK⁻ background results in a dramatic increase of its specific activity up to phenotypically detectable levels. These results indicate that the chaperone DnaK is critical for the folding of misfolding-prone proteins and also that the soluble form reached in its absence by a fraction of polypeptides is not necessarily supportive of biological activity. In the case of *E. coli* β-galactosidase, the catalytic activity requires assembling into tetramers and the fine organization of the activating interfaces holding the active sites, what might not be properly reached in absence of DnaK.
© 2005 Wiley Periodicals, Inc.

Keywords: aggregation; chaperones; DnaK; heat-shock; protein folding; recombinant protein

INTRODUCTION

Protein folding of around 5%–10% of cell polypeptides is assisted by components of the cellular quality control system, and this percentage increases up to more than 30% under conformational stress conditions such as high growth temperature or overproduction of foreign polypeptides (Feldman and Frydman, 2000; Fink, 1999). In *E. coli*, among the cell elements devoted to assist protein folding, DnaK and their co-chaperones DnaJ and GrpE, and GroEL and its co-chaperone GroES, have main and sequential functions in processing folding reluctant protein chains (Goloubinoff et al., 1999). On the other side, small heat-shock proteins bind to and protect misfolded proteins from aggregation (Shearstone and Baneyx, 1999; Thomas and Baneyx, 1998). Protein aggregation, as either loose aggregates or inclusion bodies, occurs by failure of folding and holding activities, a common situation when chaperones are under-titrated. Being elements from the protein quality control system, cell proteases degrade unprotected, misfolded polypeptides in the soluble cell fraction (Carrió et al., 1999) but also when associated to protein aggregates (Corchero et al., 1997). On the other hand, ClpB, assisted by DnaK and small heat-shock proteins solubilize aggregates in vitro (Lee and Vierling, 2000; Mogk et al., 2003a,b; Weibezahn et al., 2004), and the ClpB-DnaK pair might also be involved in the physiological dissolution of bacterial inclusion bodies when the substrate load for chaperones declines (Carrió and Villaverde, 2001). DnaK is also, a negative modulator of the heat-shock system and minimizes expression of the σ^{32} regulon that encodes the main cytoplasmatic chaperones and proteases including DnaK itself (Morita et al., 2000; Tomoyasu et al., 1998, 2001).

In cells devoid of a functional DnaK, overproduction of foreign proteins results into large inclusion bodies but still an important fraction of recombinant polypeptides are found in the soluble cell fraction (Carrió and Villaverde, 2003). By using an engineered β-galactosidase, we have here explored the status of this protein population through the analysis of their stability and enzymatic activity. Interestingly, the absence of DnaK prevents proper folding and even soluble polypeptide chains are dramatically less stable and active than their counterparts produced in wild type DnaK⁺ cells. This indicates that the folding activities of DnaK are critical for the production of a misfolding-prone β-galactosidase fusion protein and that cannot be complemented by other elements of the heat-shock response.

MATERIALS AND METHODS

Strains and Plasmids

The *Escherichia coli* strains used in this study were the Lac⁻ streptomycin resistant MC4100 (araD139 Δ(argF-lac) U169 rpsL150 relA1 flbB5301 deoC1 ptsF25 rbsR)

(Sambrook et al., 1989) and its tetracycline and streptomycin resistant *dnaK756* (DnaK⁻) derivative JGT20 encoding a non-functional chaperone (Thomas and Baneyx, 1998). pJVP1LAC, a pJCO46 derivative (Corchero and Villaverde, 1998), encodes an amino-terminal β-galactosidase fusion protein carrying the VP1 capsid protein of foot-and-mouth disease virus that dramatically reduces the solubility of the whole fusion resulting in its aggregation as inclusion bodies (Corchero et al., 1996). The expression of the gene fusion is under the control of the strong lambda promoters p_R and p_L and regulated by a plasmid-encoded CI^{ts} repressor.

Culture Conditions

LB medium (Sambrook et al., 1989) with the required antibiotics at a final concentration of 100 μg/mL each was used for both plate and liquid cultures. X-Gal was added to plates at 40 μg/mL. Liquid cultures were grown in 500 mL flasks in a shaking water bath at 28°C and 200 rpm until the desired optic density. Then, cultures were shifted to 42°C to induce recombinant gene expression for 3 h. After that, samples taken at 0, 1, 2, and 3 h were processed for enzymatic activity and immunodetection of the fusion protein.

Determination of the Enzymatic Activity

The analysis of total β-galactosidase activity was performed on selected samples through a variant of Miller's protocol (Miller, 1972) that uses chloroform instead toluene as permeabilizing agent, and the activity values are presented as corrected by the OD of the culture. Further details can be found elsewhere (Feliu et al., 1998). To determine β-galactosidase activity in either soluble and insoluble cell fractions, 2.5 mL culture samples were disrupted by sonication and centrifuged for 15 min at 14,000 rpm. The soluble fraction was directly used for the analysis, and the pellet was resuspended in 100 μL of Z-buffer before the activity analysis. Thermal inactivation experiments were performed by incubating the soluble cell fractions at 47°C. Results are presented as the average of three independent experiments.

Western Blot

Cells were concentrated up to a predefined OD$_{550}$, ice-jacketed and disrupted by ultrasonication as described (Feliu et al., 1998). These samples were centrifuged at 14,000 rpm for 15 min at 4°C to separate soluble and insoluble cell fractions. Inclusion bodies were further purified as described when necessary (Carrió et al., 2000). Samples were boiled in loading buffer (Laemmli, 1970) for 10 min, submitted to PAGE and analyzed by Western Blot as described (Carrió and Villaverde, 2003). Bands were developed by using a rabbit serum against either β-galactosidase or GroEL and quantified within the linear range with the Quantity One (BioRad, Hercules, CA) software (using commercial proteins as standards) after high resolution scanning. Results are presented as the average of three independent experiments.

RESULTS

Activity of a β-galactosidase Fusion Protein Produced in Absence of DnaK Depends on the Specific Gene Expression Conditions

In recombinant *E. coli*, the functional deficiency in the chaperone DnaK results in the dramatic accumulation of misfolding-prone proteins and in a low yield of the soluble version when compared with wild-type cells (Carrió and Villaverde, 2003). We evaluated the biological activity of the soluble protein as resulting from a DnaK-independent folding process by an in situ analytical system that permits the easy monitoring of β-galactosidase enzymatic activity. *E. coli* MC4100 (wild-type)/pJVP1LAC and JGT20 (DnaK⁻)/pJVP1LAC cells were plated on X-gal LB dishes at 42°C to promote expression of the temperature-regulated *VP1LAC* gene. Wild-type colonies showed a blue precipitate that was slightly lighter when plates were incubated at 28°C before triggering *VP1LAC* gene expression (Fig. 1, top). This was probably caused by reduced productivity when induction of gene expression was triggered on old cultures.

As expected, colonies of DnaK⁻ cells were white, indicative a low amount (if any) of functional VP1LAC (Fig. 1, left). However, when cell-containing plates were incubated at 28°C before their transfer to 42°C, DnaK⁻ colonies exhibited intense blue color being then indistinguishable from those of the wild-type (Fig. 1, right). Therefore, we decided to further explore the apparent variability of VP1LAC solubility observed in absence of DnaK.
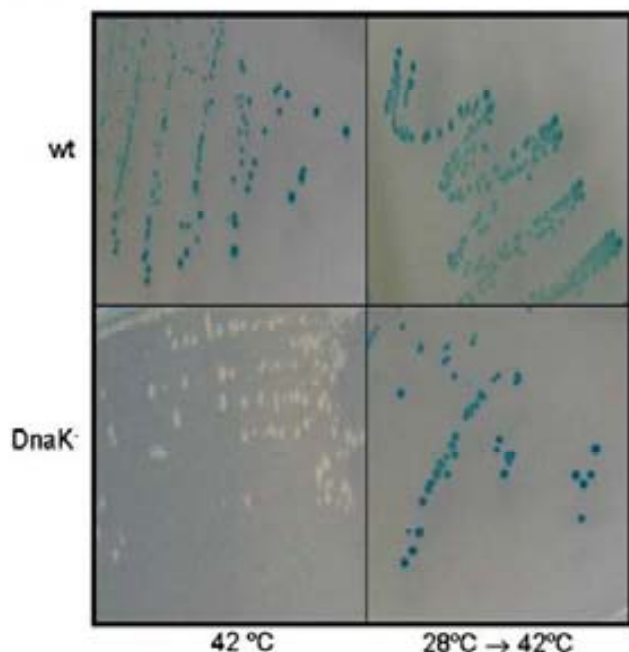
Figure 1. Accumulation of the X-gal blue product in colonies of the wild-type MC4100 strain and its DnaK⁻ derivative JGT20, both containing the plasmid pJVP1LAC. Cells were cultured on LB plates plus the required antibiotics and the lactose analog for 36 h, either always at 42°C (left) or at 28°C for 24 h and then shifted at 42 for 12 additional hours (right). [Color figure can be seen in the online version of this article, available at www.interscience.wiley.com.]

Enzymatic Activity of VP1LAC Increases in DnaK⁻ Cells Under *Late* Gene-Expression Conditions

The pattern of VP1LAC yield and activity suggested by data from Figure 1 was then tested in liquid cultures by triggering *VPLAC* gene expression at both lag and late exponential phases (*early* and *late* strategies respectively, Fig. 2A). In the DnaK⁻ background, the *late* induction strategy promoted higher levels of β-galactosidase activity than the *early* induction (Fig. 2B), the observed tendency being coincident with the pattern shown in Figure 1. To evaluate if this could be determined by variable productivity, we analyzed the total amount of VP1LAC 3 h after induction of gene expression in both strains. As observed, the yield of VP1LAC decayed in wild-type cells when gene expression was promoted in old cultures (Fig. 3A). Although higher variability was observed when comparing different experiments in absence of DnaK (Fig. 3A), in most of them VP1LAC amounts were also lower in old cultures. While the enzymatic activity in wild-type cells decreased concomitantly, in the DnaK⁻ background it surprisingly increased inversely to the reduction of the VP1LAC amount (Fig. 3B). On the other hand, the fraction of soluble VP1LAC shifted from $13.9 \pm 0.5\%$ in young cultures to $29.3 \pm 0.9\%$ in old cultures (not shown), what resulted in a similar value of soluble VP1LAC around 4,500 ng/mL/OD units (an increase in the relative solubility was also observed in the wild-type from $11.8 \pm 1.6\%$ to $80.7 \pm 7.5\%$, not shown). All this data indicated that at the average specific activity of VP1LAC in the DnaK⁻ background was higher after *late* gene expression.

GroEL Levels and VP1LAC Proteolytic Digestion Do Not Account for the Dissimilar Enzymatic Activity in DnaK⁻ Cells

Being DnaK a negative regulator of the σ^{32} regulon (Gamer et al., 1996), its absence results in the constitutive production of main cytoplasmatic heat-shock proteins including chaperones

and proteases ((Petersson et al., 2004) and references therein). However, the results presented above suggested that DnaK-independent folding could be improved by *late* induction, and this fact might involve the activity of other chaperones. To explore putative variations in the level of other heat-shock proteins that could account for more efficient VP1LAC folding, GroEL amounts were determined in the DnaK⁻ background 3 h after induction of VP1LAC gene expression induced according to either *early* or *late* strategies. Also, GroEL was determined before induction to analyze possible fluctuations of its cellular concentration during plain culture growth. However, the concentration of this chaperone was constant in all the tested conditions (Fig. 4A), indicating that heat-shock gene expression levels were not affected by the cell status at different phases of culture growth.

The higher activity observed in late-induced DnaK⁻ cultures might be also accounted by an inhibition of the proteolytic activity on misfolded VP1LAC polypeptides in DnaK⁻ cells resulting in more opportunities for proper folding. Alternatively, a more efficient digestion of misfolded and inactive polypeptides versus folded and active polypeptides could improve the global quality of the soluble protein and consequently enhance the resulting specific activity. Although such variations in the proteolytic activity were not expected, we analyzed the digestion patterns of VP1LAC in both *early* and *late* expression conditions regarding the occurrence of specific degradation intermediates and their presence relative to that of the full-length form, without finding significant qualitative or quantitative differences that could support the mentioned hypothesis (Fig. 4B).

Stability of the Soluble VP1LAC Is Lower in Absence Than in Presence of DnaK, and it Is Enhanced Under *Late* Expression Conditions

The increase of the enzymatic activity observed in absence of DnaK was further explored by considering the insoluble
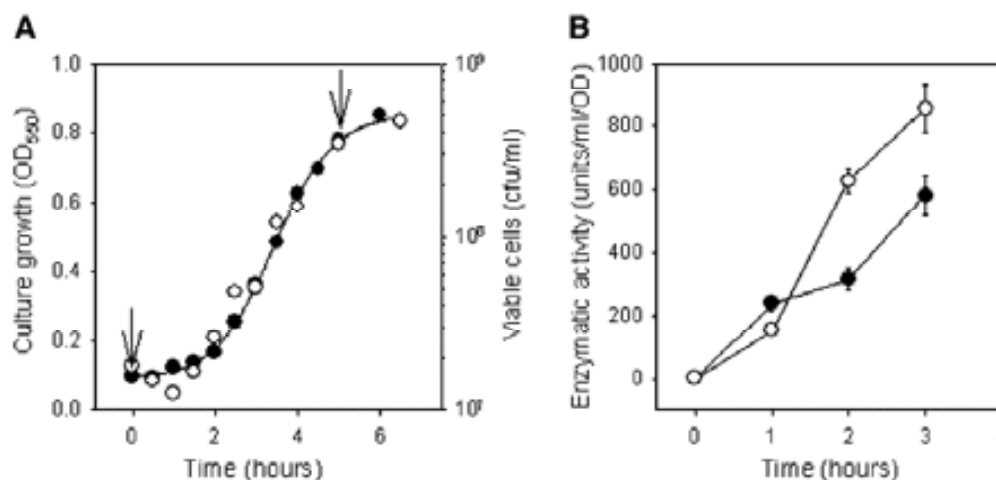
Figure 2. A: Growth of JGT20/pJVP1LAC cells at 28°C in shaker flask, measured by both optical density (black dots, left axis) and viable cell counts (white dots, right axis). Arrows indicate the time points selected to trigger *VP1LAC* gene expression, either at early, lag phase ($OD_{550} = 0.115$) or at late, exponential phase ($OD_{550} = 0.800$). These gene induction strategies were named *early* and *late* respectively for convenience. B: Evolution of β-galactosidase activity in JGT20/pJVP1LAC after induction of *VP1LAC* gene expression, triggered at either *early* (black dots) or *late* (white dots) growth phases.
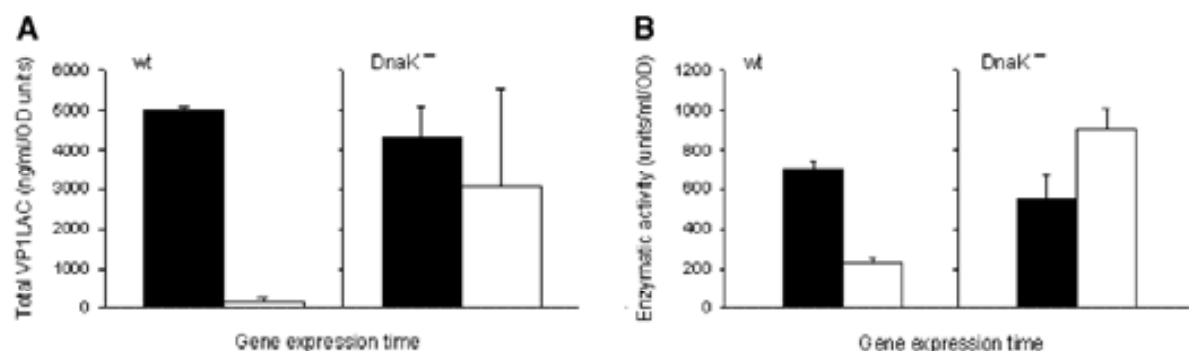
Figure 3. Total amount of VP1LAC observed by Western blot (A) and β-galactosidase activity (B) in cell extracts of MC4100/pJVP1LAC and JGT20/pJVP1LAC cultures. Data were collected 3 h after induction of gene expression triggered at either *early* (black bars) or *late* (white bars) growth phases. Differences between activity data from the two induction strategies are significant (*P* < 0.001).

cell fraction. It is know that a part of the polypeptides aggregated as inclusion bodies can still exhibit biological activities (Carrió and Villaverde, 2003; Tokatlidis et al., 1991; Worrall and Goss, 1989) and this could contribute to the total activity observed in the cultures (Figs. 1 and 3). The specific activity of VP1LAC was determined in both genetic backgrounds and gene expression conditions. As observed (Table I), the specific activity of the soluble VP1LAC produced in presence of DnaK was between 9- and 15-fold higher than when produced in the mutant, proving that DnaK is critical for the proper folding of this protein. On the other hand, the contribution of insoluble VP1LAC in the global activity was important, but it could not account by itself the global variation of the enzyme activity. Interestingly, we observed that the *late* induction strategy augmented the specific activity of the soluble form of VP1LAC in both strains.

Both the different protein activity in the tested strains as well as the observed increase mediated by *late* gene expression could be therefore dependent on the stability of VP1LAC. To check this possibility, we studied the thermal stability of the soluble fusion protein under the relevant culture conditions. The half-life of VP1LAC in wild-type

cells was twofold higher than in the *dnaK* mutant (Fig. 5), indicating that the absence of DnaK could result in deficient folding even of proteins that finally occurs in the soluble cell fraction. However, in both strains the half-life of the soluble fraction was doubled under *late* induction conditions, being coincident with the variations of the specific activity shown in Table I.

DISCUSSION

The bacterial production of foreign recombinant proteins, especially when driven at high rates, usually results in the formation of bacterial aggregates known as inclusion bodies. The requirement for more folding-assistant proteins than those physiological available results in the deposition of misfolded forms of the protein. Among the complex chaperone network surveying the protein conformational quality, DnaK (along with their co-chaperones DnaJ and GrpE) drives folding intermediates into folding attempts and deliver folding reluctant forms to the subsequent GroEL–GroES folding apparatus (Goloubinoff et al., 1999; Nishihara et al., 1998; Thomas and Baneyx, 1996). Also, in cooperation with ClpB and small heat-shock proteins, DnaK refolds polypeptide chains from thermal aggregates (Mogk et al., 1999, 2003a,b; Mogk and Bukau, 2004) and it might be a main ruler of the physiological disintegration of inclusion bodies (Carrió and Villaverde, 2002; Schlieker et al., 2002). In absence of functional DnaK, the total yield of recombinant protein and the relative amount of aggregated polypeptides are higher than in wild-type cells, and inclusion bodies concomitantly up-sized probably because of failures in both folding and refolding activities (Carrió and Villaverde, 2003;
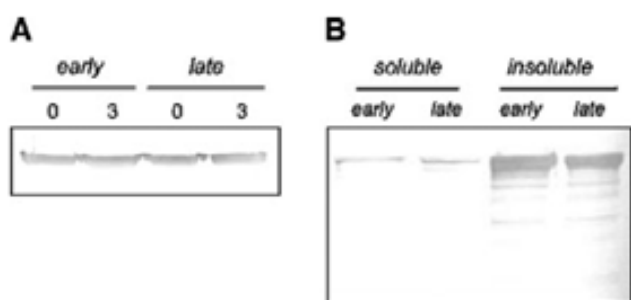
Figure 4. A: Immunodetection of GroEL in JGT20/pJVP1LAC cells growing at 28°C at OD$_{550}$ = 0.115 (**lane 1**) and at OD$_{550}$ = 0.800 (**lane 3**). Three hours after induction of *VP1LAC* expression, triggered at these OD values, GroEL was immunodetected in **lanes 2** and **4**, respectively. The same biomass was loaded in every lane for comparison. The slight differences in GroEL amount were not significant. B: VP1LAC immunodetected in both soluble and insoluble cell fractions 3 h after induction of gene expression, according to *early* and *late* strategies.

Table I. Specific activity (in units/µg) of VP1LAC in different cell fractions, strains, and culture conditions.

	WT		DnaK⁻	
	Soluble	Insoluble	Soluble	Insoluble
Early	1771.5 ± 179.4	338.5 ± 9.4	189.8 ± 31.5	153.7 ± 26.9
Late	5415.5 ± 502.1	1039.8 ± 32.1	365.8 ± 26.4	119.9 ± 14.4
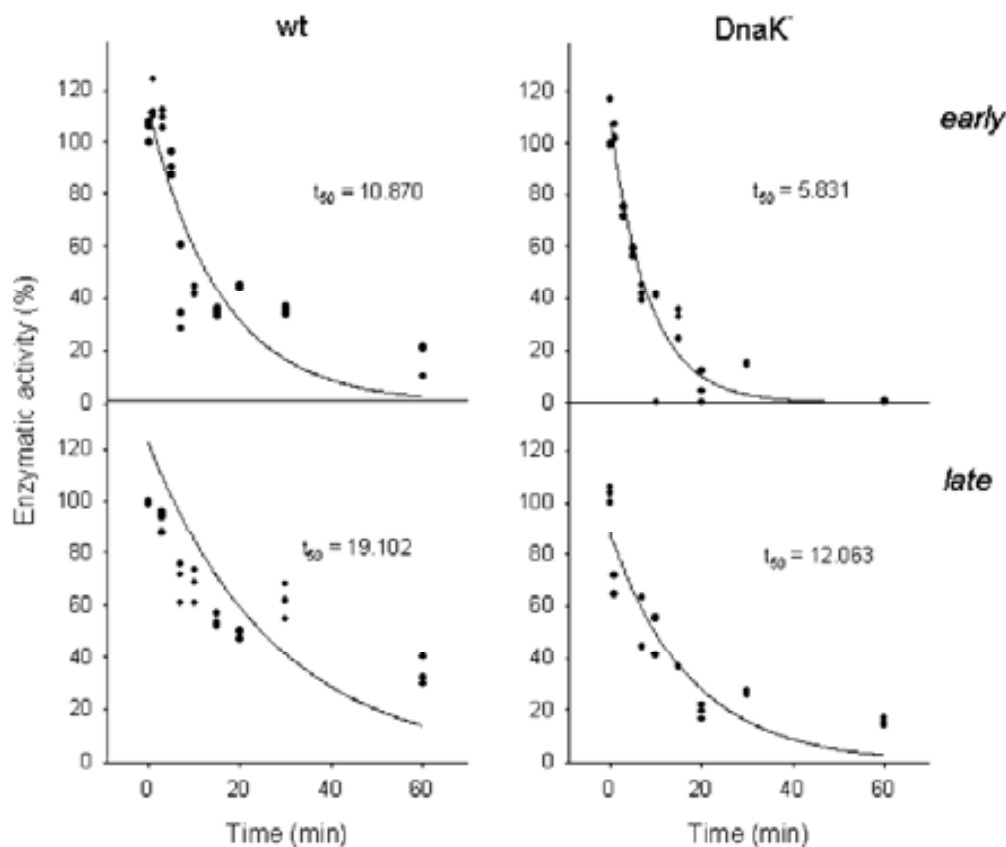
Figure 5. Thermal inactivation of VP1LAC in the soluble cell fraction of either MC4100 or JGT20, 3 h after the induction of gene expression according to *early* and *late* strategies. The time (in min) at which 50% of the activity is lost is indicated as t_{50}. Parameters of the adjusted curves were significant in all the cases ($P < 0.001$).

Petersson et al., 2004). These observations indicate that DnaK is crucial for the conformational surveillance and that DnaK activities cannot be completely complemented by other chaperones.

However, a still significant fraction of soluble recombinant protein is observed when recombinant protein production is promoted in a DnaK⁻ background (Carrió and Villaverde, 2003), in absolute values being comparable to that found in wild-type cells. We have explored the status of this protein fraction by analyzing the denaturation kinetics of the model β-galactosidase fusion produced in the strain JGT20. At the early exponential phase of the culture thermal stability of VP1LAC is about twofold lower than when produced in wild-type cells (Fig. 5). This fact results in a specific activity of the soluble VP1LAC about 10 times lower than that produced in presence of DnaK (Table I) and in a clear Lac⁻ phenotype of JT20/pJVP1LAC (Fig. 1). This observation indicates that only a minor portion of the soluble VP1LAC polypeptides reaches its functional form in absence of active DnaK, this chaperone being critical to drive VP1LAC into a conformational stable and active form even when the rest of cytoplasmic chaperones are constitutively expressed. Since *E. coli* β-galactosidase is only active in the tetrameric form and the active sites are generated at the activating interfaces by contacts between facing monomers (Jacobson et al.,

1994), the DnaK folding activity might be involved in promoting the conformational disposition of contact surfaces required for enzyme activation. Interestingly, and in the line of this suggestion, DnaK is necessary for α-complementation of the enzyme (Lopes Ferreira and Alix, 2002), what requires a stable interaction between the complementing fragments.

When the production of VP1LAC is triggered within a defined time period at the beginning of the stationary phase (Fig. 2), its stability is enhanced about two times irrespectively of DnaK (Fig. 5). In this phase of culture growth, biosynthesis of recombinant proteins is dramatically minimized (Vila et al., 1997) resulting in a lower yield of VP1LAC (Fig. 3). A slow rate of recombinant protein synthesis is generally recognized as promoter of proper protein folding (Friehs and Reardon, 1993; Murby et al., 1996; Schein, 1991; Thomas et al., 1997), and because of the lower substrate load for cell chaperones their titration by misfolded protein chains resulting minimized. This stabilization is naturally reflected by an increase in the specific activity of soluble VP1LAC (Table I) that although being moderate could be biologically relevant for the DnaK⁻ strain in which the folding machinery is operating close to the optimal conditions. In this last case, this fact augmented the enzymatic activity of the culture despite the lower yield of VP1LAC (Figs. 2 and 3) up to the reversion of the Lac⁻ phenotype (Fig. 1). However, even

under the low production conditions provided by the *late* induction strategy the acting chaperones were not able to drive the population of soluble VP1LAC chains into a proper native-like conformation (note its specific activity being still 15-fold lower than that VP1LAC in wild-type cells, Table I).

The bacterial chaperone network is at least partially redundant. Among other indications, overproduction of GroEL–GroES reverses the lethal phenotype of cells lacking both the trigger factor and DnaK (Vorderwulbecke et al., 2004). This fact proves that despite due size constrictions, the sets of substrates recognized by DnaK and GroEL are only partially matching (Deuerling et al., 2003; Ewalt et al., 1997; Houry. 2001; Mogk et al., 1999), folding activities of DnaK can be complemented by GroELS. In this regard, the chaperone GroEL interacts with *E. coli* β-galactosidase (Ayling and Baneyx, 1996; Boels et al., 1999), and although the large size of the enzyme monomer (116 kDa) prevents it from encapsulation by GroEL–GroES ring complexes, restricted to 55–57 kDa (Ewalt et al., 1997; Sakikawa et al., 1999), this chaperone is still active in folding larger substrates (Chaudhuri et al., 2001; Song et al., 2003). Since the residual folding of VP1LAC in DnaK⁻ is improved by low production rates, it must be then modulated by GroEL or other cell factors that become more available at low production rates. However, the results presented here indicate that the folding-assistant proteins alternative to DnaK are only moderately efficient in directing the folding of a misfolding prone β-galactosidase.

References

Ayling A. Baneyx F. 1996. Influence of the GroE molecular chaperone machine on the in vitro refolding of *Escherichia coli* beta-galactosidase. Protein Sci 5(3):478–487.

Boels K, Carrió MM, Arís A, Corchero JL, Villaverde A. 1999. Distinct chaperone affinity to folding variants of homologous recombinant proteins. Biotechnol Lett 21:531–536.

Carrió MM, Corchero JL, Villaverde A. 1999. Proteolytic digestion of bacterial inclusion body proteins during dynamic transition between soluble and insoluble forms. Biochim Biophys Acta 1434(1):170–176.

Carrió MM, Cubarsi R, Villaverde A. 2000. Fine architecture of bacterial inclusion bodies. FEBS Lett 471(1):7–11.

Carrió MM, Villaverde A. 2001. Protein aggregation as bacterial inclusion bodies is reversible. FEBS Lett 489(1):29–33.

Carrió MM, Villaverde A. 2002. Construction and deconstruction of bacterial inclusion bodies. J Biotechnol 96(1):3–12.

Carrió MM, Villaverde A. 2003. Role of molecular chaperones in inclusion body formation. FEBS Lett 537(1–3):215–221.

Chaudhuri TK, Farr GW, Fenton WA, Rospert S, Horwich AL. 2001. GroEL/GroES-mediated folding of a protein too large to be encapsulated. Cell 107(2):235–246.

Corchero JL, Cubarsi R, Enfors S, Villaverde A. 1997. Limited in vivo proteolysis of aggregated proteins. Biochim Biophys Res Commun 237(2):325–330.

Corchero JL, Viaplana E, Benito A, Villaverde A. 1996. The position of the heterologous domain can influence the solubility and proteolysis of beta-galactosidase fusion proteins in *E. coli*. J Biotechnol 48(3): 191–200.

Corchero JL, Villaverde A. 1998. Plasmid maintenance in *Escherichia coli* recombinant cultures is dramatically, steadily, and specifically influenced by features of the encoded proteins. Biotechnol Bioeng 58(6):625–632.

Deuerling E, Patzelt H, Vorderwulbecke S, Rauch T, Kramer G, Schaffitzel E, Mogk A, Schulze-Specking A, Langen H, Bukau B. 2003. Trigger Factor and DnaK possess overlapping substrate pools and binding specificities. Mol Microbiol 47(5):1317–1328.

Ewalt KL, Hendrick JP, Houry WA, Hartl FU. 1997. In vivo observation of polypeptide flux through the bacterial chaperonin system. Cell 90(3):491–500.

Feldman DE, Frydman J. 2000. Protein folding in vivo: The importance of molecular chaperones. Curr Opin Struct Biol 10(1):26–33.

Feliu JX, Cubarsi R, Villaverde A. 1998. Optimized release of recombinant proteins by ultrasonication of *E. coli* cells. Biotechnol Bioeng 58(5): 536–540.

Fink AL. 1999. Chaperone-mediated protein folding. Physiol Rev 79(2): 425–449.

Friehs K, Reardon KF. 1993. Parameters influencing the productivity of recombinant *E. coli* cultivations. Adv Biochem Eng Biotechnol 48:53–77.

Gamer J, Multhaup G, Tomoyasu T, McCarty JS, Rudiger S, Schonfeld HJ, Schirra C, Bujard H, Bukau B. 1996. A cycle of binding and release of the DnaK, DnaJ and GrpE chaperones regulates activity of the *Escherichia coli* heat shock transcription factor sigma32. EMBO J 15(3):607–617.

Goloubinoff P, Mogk A, Zvi AP, Tomoyasu T, Bukau B. 1999. Sequential mechanism of solubilization and refolding of stable protein aggregates by a bichaperone network. Proc Natl Acad Sci USA 96(24):13732–13737.

Houry WA. 2001. Mechanism of substrate recognition by the chaperonin GroEL. Biochem Cell Biol 79(5):569–577.

Jacobson RH, Zhang XJ, DuBose RF, Matthews BW. 1994. Three-dimensional structure of beta-galactosidase from *E. coli*. Nature 369(6483):761–766.

Laemmli UK. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227(259):680–685.

Lee GJ, Vierling E. 2000. A small heat shock protein cooperates with heat shock protein 70 systems to reactivate a heat-denatured protein. Plant Physiol 122(1):189–198.

Lopes Ferreira N, Alix JH. 2002. The DnaK chaperone is necessary for alpha-complementation of beta-galactosidase in *Escherichia coli*. J Bacteriol 184(24):7047–7054.

Miller JH. 1972. Experiments in molecular genetics. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Mogk A, Bukau B. 2004. Molecular chaperones: Structure of a protein disaggregase. Curr Biol 14(2):R78–R80.

Mogk A, Deuerling E, Vorderwulbecke S, Vierling E, Bukau B. 2003a. Small heat shock proteins, ClpB and the DnaK system form a functional triade in reversing protein aggregation. Mol Microbiol 50(2):585–595.

Mogk A, Schlieker C, Friedrich KL, Schonfeld HJ, Vierling E, Bukau B. 2003b. Refolding of substrates bound to small Hsps relies on a disaggregation reaction mediated most efficiently by ClpB/DnaK. J Biol Chem 278(33):31033–31042.

Mogk A, Tomoyasu T, Goloubinoff P, Rudiger S, Roder D, Langen H, Bukau B. 1999. Identification of thermolabile *Escherichia coli* proteins: Prevention and reversion of aggregation by DnaK and ClpB. EMBO J 18(24):6934–6949.

Morita MT, Kanemori M, Yanagi H, Yura T. 2000. Dynamic interplay between antagonistic pathways controlling the sigma 32 level in Escherichia coli. Proc Natl Acad Sci USA 97(11):5860–5865.

Murby M, Uhlen M, Stahl S. 1996. Upstream strategies to minimize proteolytic degradation upon recombinant production in *Escherichia coli*. Protein Expr Purif 7(2):129–136.

Nishihara K, Kanemori M, Kitagawa M, Yanagi H, Yura T. 1998. Chaperone coexpression plasmids: Differential and synergistic roles of DnaK-

DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in *Escherichia coli*. Appl Environ Microbiol 64(5):1694–1699.

Petersson L, Carrió MM, Vera A, Villaverde A. 2004. The impact of dnaKJ overexpression on recombinant protein solubility results from antagonistic effects on the control of protein quality. Biotechnol Lett 26(7):595–601.

Sakikawa C, Taguchi H, Makino Y, Yoshida M. 1999. On the maximum size of proteins to stay and fold in the cavity of GroEL underneath GroES. J Biol Chem 274(30):21251–21256.

Sambrook J, Fritsch E, Maniatis T. 1989. Molecular cloning, a laboratory manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Schein CH. 1991. Optimizing protein folding to the native state in bacteria. Curr Opin Biotechnol 2(5):746–750.

Schlieker C, Bukau B, Mogk A. 2002. Prevention and reversion of protein aggregation by molecular chaperones in the *E. coli* cytosol: Implications for their applicability in biotechnology. J Biotechnol 96(1):13–21.

Shearstone JR, Baneyx F. 1999. Biochemical characterization of the small heat shock protein IbpB from *Escherichia coli*. J Biol Chem 274(15):9937–9945.

Song JL, Li J, Huang YS, Chuang DT. 2003. Encapsulation of an 86-kDa assembly intermediate inside the cavities of GroEL and its single-ring variant SR1 by GroES. J Biol Chem 278(4):2515–2521.

Thomas JG, Ayling A, Baneyx F. 1997. Molecular chaperones, folding catalysts, and the recovery of active recombinant proteins from *E. coli*. To fold or to refold. Appl Biochem Biotechnol 66(3):197–238.

Thomas JG, Baneyx F. 1996. Protein folding in the cytoplasm of *Escherichia coli*: Requirements for the DnaK-DnaJ-GrpE and GroEL-GroES molecular chaperone machines. Mol Microbiol 21(6):1185–1196.

Thomas JG, Baneyx F. 1998. Roles of the *Escherichia coli* small heat shock proteins IbpA and IbpB in thermal stress management: Comparison with ClpA, ClpB, and HtpG In vivo. J Bacteriol 180(19):5165–5172.

Tokatlidis K, Dhurjati P, Millet J, Beguin P, Aubert JP. 1991. High activity of inclusion bodies formed in *Escherichia coli* overproducing Clostridium thermocellum endoglucanase D. FEBS Lett 282(1):205–208.

Tomoyasu T, Mogk A, Langen H, Goloubinoff P, Bukau B. 2001. Genetic dissection of the roles of chaperones and proteases in protein folding and degradation in the *Escherichia coli* cytosol. Mol Microbiol 40(2):397–413.

Tomoyasu T, Ogura T, Tatsuta T, Bukau B. 1998. Levels of DnaK and DnaJ provide tight control of heat shock gene expression and protein repair in *Escherichia coli*. Mol Microbiol 30(3):567–581.

Vila P, Corchero JL, Cubarsi R, Villaverde A. 1997. Enhanced fitness of recombinant protein synthesis within the stationary phase of *E. coli* batch cultures. Biotechnol Lett 19(3):225–228.

Vorderwulbecke S, Kramer G, Merz F, Kurz TA, Rauch T, Zachmann-Brand B, Bukau B, Deuerling E. 2004. Low temperature or GroEL/ES overproduction permits growth of *Escherichia coli* cells lacking trigger factor and DnaK. FEBS Lett 559(1–3):181–187.

Weibezahn J, Bukau B, Mogk A. 2004. Unscrambling an egg: Protein disaggregation by AAA+ proteins. Microb Cell Fact 3(1):1.

Worrall DM, Goss NH. 1989. The formation of biologically active beta-galactosidase inclusion bodies in *Escherichia coli*. Aust J Biotechnol 3(1):28–32.

IV.2 Paper 2

Aggregation as bacterial inclusion bodies does not imply inactivation of

enzymes and fluorescent proteins

Elena García-Fruitós, Núria González-Montalbán, Montse Morell, Andrea Vera, Rosa

María Ferraz, Anna Arís, Salvador Ventura and Antonio Villaverde

Microbial Cell Factories, Vol. 4, September 12, 2005

The aim of this study was to analyse in more detail not only the conformational quality of the soluble protein, but also that of the protein aggregated as inclusion bodies. For that purpose, we quantitatively explored the biological activity of the soluble and aggregated recombinant proteins in *Escherichia coli* MC4100 strain overexpressing the human dihydrofolate reductase (hDHFR), an engineered *Escherichia coli* β-galactosidase (VP1LAC), a green fluorescent protein fused to VP1 protein (VP1GFP) and a blue fluorescent protein fused to Aβ-amyloid peptide (Aβ42(F19D)-BFP).

Although it has been generally believed that inclusion body proteins are biologically inactive, our results revealed that there is an unexpectedly high biological activity associated to these aggregates. Moreover, the obtained results also indicated that native-like structure in bacterial inclusion bodies coexists with an amyloid, molecular β-sheet organisation of the embedded proteins. Therefore, these facts indicate that aggregation as inclusion bodies does not split protein population into active and inactive fractions, being solubility an inappropriate indicator of conformational quality. All these data suggest that active inclusion bodies formed by enzymes could be potentially used as efficient industrial catalysts for bioprocesses, becoming an attractive alternative to *in vitro* protein refolding procedures.

Microbial Cell Factories

BioMed Central

Open Access

Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins

Elena García-Fruitós[1,2], Nuria González-Montalbán[1,2], Montse Morell[1], Andrea Vera[1,2], Rosa María Ferraz[1,2], Anna Arís[1,2], Salvador Ventura[1,3] and Antonio Villaverde*[1,2]

Address: [1]Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Bellaterra, 08193 Barcelona, Spain, [2]Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona, Bellaterra, 08193 Barcelona, Spain and [3]Departament de Bioloquímica i de Biologia Molecular, Universitat Autònoma de Barcelona, Bellaterra, 08193 Barcelona, Spain

Email: Elena García-Fruitós - Elena.Garcia.Fruitos@uab.es; Nuria González-Montalbán - nuria.gonzalez.montalban@uab.es; Montse Morell - mmorell@bioinf.uab.es; Andrea Vera - andrea.vera@uab.es; Rosa María Ferraz - RosaMaria.Ferraz@uab.es; Anna Arís - anna.aris@uab.es; Salvador Ventura - salvador.ventura@uab.es; Antonio Villaverde* - avillaverde@servet.uab.es

* Corresponding author

Abstract

Background: Many enzymes of industrial interest are not in the market since they are bio-produced as bacterial inclusion bodies, believed to be biologically inert aggregates of insoluble protein.

Results: By using two structurally and functionally different model enzymes and two fluorescent proteins we show that physiological aggregation in bacteria might only result in a moderate loss of biological activity and that inclusion bodies can be used in reaction mixtures for efficient catalysis.

Conclusion: This observation offers promising possibilities for the exploration of inclusion bodies as catalysts for industrial purposes, without any previous protein-refolding step.

Background

Protein misfolding is a common event during bacterial over-expression of recombinant genes [1]. The aggregation of insoluble polypeptide chains as inclusion bodies has seriously restricted the spectrum of proteins marketed by the biotechnology industry. Being widespreadly believed that inclusion body proteins are biologically inactive and therefore useless in bioprocesses, many aggregation-prone products have been disregarded for commercialisation. Protein solubility can be tailored by either process [2] or protein [3] engineering, although most efforts have been addressed to minimize inclusion body formation by co-production of folding modulators

[4], or to refold purified inclusion body proteins by chemical denaturation followed by refolding procedures [5]. Both strategies need to be adapted to particular protein species and they render largely variable results regarding the final soluble protein yield.

Interestingly, independent reports have noted enzymatic activity associated to inclusion bodies formed by recombinant enzymes [6-8], but the extent of these side-observations has been never quantified and its biological and biotechnological relevance remained unexplored. In this work, we have quantitatively explored the biological

Table 1: Enzymatic activity or fluorescence of inclusion bodies produced in *E. coli*

Construct name	Reference	Functional protein	Fraction of inclusion body protein (range, %) [a]	Aggregating domain or protein (all in the N-terminal position)	Specific activity or emission [b] (enzymatic units/mg or fluorescence units/mg)		Activity of the inclusion body fraction relative to that of soluble protein (%) [c]
					Soluble protein	Inclusion bodies	
VP1LAC	This work and [9]	*E. coli* β-galactosidase	35.6–45.9	FMDV VP1 capsid protein	698.3 ± 153.0	1162.5 ± 256.0	166.4
hDHFR	[25]	Human dihydrofolate reductase	28.4–36.8	none	$8.0 \ 10^{-2} \pm 2.6 \ 10^{-2}$	$4.7 \ 10^{-3} \pm 0.9 \ 10^{-3}$	5.9
VP1GFP	This work	Green fluorescent protein	82.5–88.4	FMDV VP1 capsid protein	359.5 ± 66.0	70.4 ± 10.1	19.5
Aβ42(F19D)-BFP	[26]	Blue fluorescent protein	61.4–65.3	Aβ42(F19D)	118.1 ± 10.2	36.3 ± 2.2	30.7

[a] The percentage of protein found in inclusion bodies relative to the total intracellular amount of recombinant protein. Values were determined from different samples taken at 3 and 5 h after triggering recombinant gene expression.
[b] These values were determined in samples taken between 3 and 5 h after triggering recombinant gene expression.
[c] Specific activity or fluorescence emission of inclusion bodies relative to the values determined for the soluble counterpart fraction. Protein amounts were determined by Western blot analysis as described and enzymatic assays performed by conventional procedures. Excitation wavelengths were 450 nm for VP1GFP and 360 nm for Aβ42(F19D)-BFP.

activity of inclusion body recombinant proteins and their potential use for bioprocesses in the aggregated form.

Results

To determine the occurrence of active protein in inclusion bodies we analysed those formed upon overproduction of the wild-type human dihydrofolate reductase (hDHFR) and an engineered *E. coli* β-galactosidase fused to the aggregation-prone foot-and-mouth disease virus (FMDV) VP1 capsid protein (VP1LAC). In addition, we explored fluorescence emission of green and blue fluorescent proteins (GFP and BFP respectively) fused to different aggregating polypeptides, namely the FMDV VP1 and a point mutant of the human Aβ-amyloid peptide (Aβ(F19D)), by comparing specific fluorescence emission of protein in the soluble cell fraction and purified inclusion bodies. Upon overproduction, all these proteins form cytoplasmic inclusion bodies in *E. coli*, the fraction of the aggregated protein ranging between 28 and 88 % of the total recombinant production (Table 1). Surprisingly, both enzymatic activity and specific fluorescence of inclusion body proteins were unexpectedly high (Table 1), ranging from 6 to 166 % of that of their counterparts occurring in the soluble cell fraction. This fact indicates that protein inactivation mediated by *in vivo* aggregation is only moderate. In addition, it is shown that protein packaging as bacterial inclusion bodies into inter-molecular β-sheet architecture (characterized by the presence of a peak

around 1620 cm⁻¹ that dominates the FTIR spectrum in the amide I region) [9,10] in these model proteins (Figure 1) is compatible with the functionality of enzyme active sites and fluorophores. In this context, VP1GFP and Aβ42(F19D)-BFP inclusion bodies are noticeably fluorescent inside the producing cells (Figure 2).

We wondered if active inclusion bodies could be then used in suspension as efficient catalysts for bioprocesses. If so, the straightforward use of these particles, that in addition are easily removable from the reaction mixture once the reaction is completed by low speed centrifugation, would be a convenient alternative to in vitro protein refolding before use, a complex procedure for which efficiencies are highly variable but in general low [5]. The enzymatic activity of soluble and inclusion body versions of both VP1LAC and hDHFR was then monitored in reaction mixtures. As observed (Figure 3A and 3B), inclusion body-embedded enzymes performed very efficiently as catalysts of enzymatic reactions. Substrate hydrolysis mediated by the insoluble form of VP1LAC was significantly faster than that mediated by the same amount of the soluble version (Figure 3A), while substrate processing by hDHFR was slower when driven from inclusion bodies but still important (Figure 3B). These observations are nicely compatible with the specific activities displayed by both versions of these proteins (Table 1).
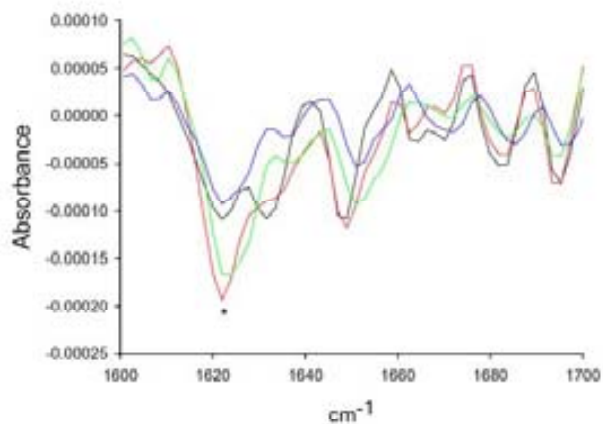
Figure 1
FTIR spectra of inclusion bodies formed by either VP1LAC (black), hDHFR (green), VP1GFP (red) or Aβ42(F19D)-BFP (blue) in the amide I region [9]. The asterisk labels the peak indicative of extended inter-molecular β-sheet structures in bacterial inclusion bodies.
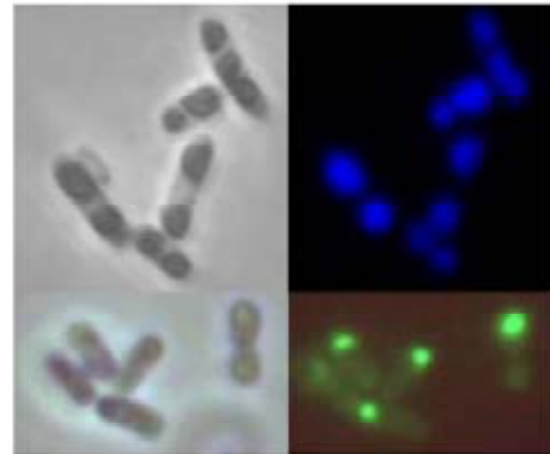
Figure 2
Optical micrographs of Aβ42(F19D)-BFP (top) and VP1GFP (bottom) inclusion bodies by phase contrast (left) and fluorescent microscopy (right).

Discussion

The quantitative similarity between protein activity in the soluble cell fraction and that of the aggregated forms of both enzymes and fluorescent proteins (Table 1) demonstrates that physiological aggregation as inclusion bodies does not necessarily split protein population into active and inactive fractions. Probably, protein solubility (observed as the occurrence in the soluble cell fraction) does not necessarily indicate the acquisition of a correctly folded and thus active structure. In this context, soluble micro-aggregates have been described [11] and recently characterized in detail [12]. The non complete coincidence between solubility and folding has been previously indicated by exhaustive mutational analysis of model proteins [13], showing that the genetic determinants of protein aggregation and misfolding are not coincident. In this way, natively unfolded proteins are unstructured but soluble [14]. Therefore, determinations of GFP-fusions solubility by using fluorescence as reporter [15] could have eventually been indicative of folding-misfolding extend rather than solubility-insolubility, since inclusion bodies formed by GFP fusions can be highly fluorescent (Figure 2). Furthermore, solubility does not appear to be an all-or-nothing attribute and polypeptide chains might exhibit a continuum of folding states in both soluble and insoluble cell fractions, between which they are dynamically transferred with the assistance of cellular folding modulators [16]. In this context, the occurrence and evolution of 'soluble' aggregates in bacteria (namely misfolded species occurring in the soluble cell fraction and presumably inactive) [12] could explain the variable specific activity observed in the soluble cell fraction of bacteria producing recombinant β-galactosidases [17].

Inversely, our results prove a major occurrence of native or native-like protein in inclusion bodies. In fact, deposition as inclusion bodies might even result in the enrichment of active species as suggested by the specific activity (166 % of that found in the soluble cell fraction; Table 1) and catalytic properties (Figure 3A) of VP1LAC inclusion bodies. This observation can be then again indirectly indicative of the presence of enzymatically inactive protein in the soluble cell fraction, since protein deposition is not expected to favour a correct folding.

Finally, although the existence of native-like structure in bacterial inclusion body proteins has been previously reported [18], here we demonstrate that this is not anecdotic but probably the architectonic nature of these kind of aggregates, as inclusion bodies formed by four structurally different proteins all display significantly high biological activity. Interestingly, the active and properly folded polypeptides in inclusion bodies coexist with a molecular β-sheet organization also manifest in all cases, although the extent of β-sheet structure and its coincidence with the biological activity of the aggregates cannot be quantitatively evaluated. Since is highly improbable that enzyme active sites involved in the intermolecular β-sheet struc-
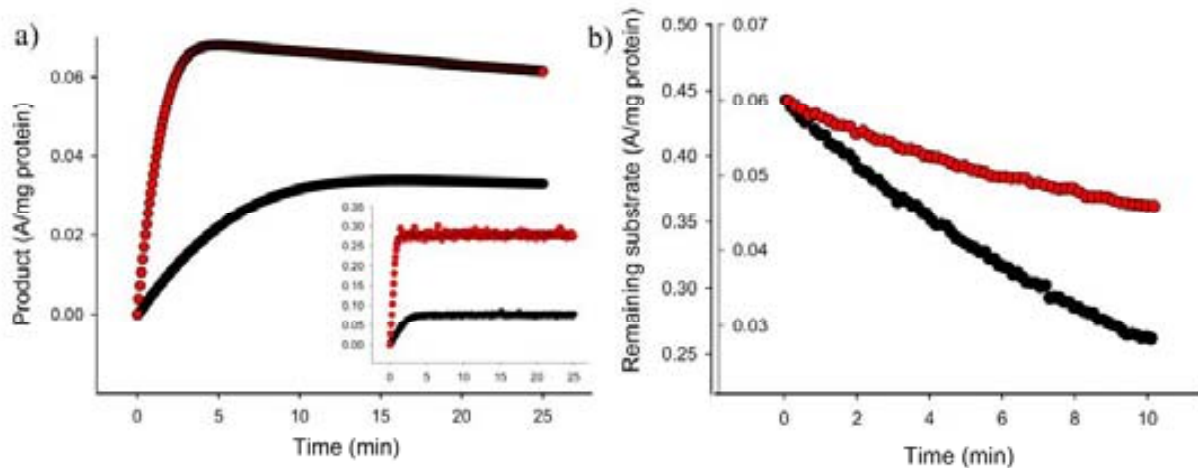
Figure 3
A) Product formed by soluble (black symbols) or inclusion body (red symbols) VP1LAC through ONPG hydrolysis as determined at 414 nm. Very coincident results have been obtained by using CPRG as alternative substrate (see the small panel), whose hydrolysis product was determined at 540 nm. B) Conversion of NADPH into NADP+ associated to tetrahydrofolate formation mediated by soluble (black symbols, left scale) and inclusion body (red symbols, right scale) hDHFR. Absorbance was determined at 340 nm.

ture could be themselves active, we suggest that enzymatic activity or fluorescence are supported by properly folded molecules or molecule segments. Aggregation, observed as protein deposition driven by intermolecular interactions between solvent-exposed hydrophobic patches [9] would not necessarily disturb the conformation of all protein domains, and the active site would be still functional if misfolded, aggregation-prone regions are located in a distant site of the polypeptide chain. Alternatively, properly folded and active molecules could coexist with β-sheet-enriched (inactive) versions of the same species, and both situations could in fact take place simultaneously in single aggregate units. Further structural and functional analysis would hopefully solve this issue.

From an applied point of view, inclusion bodies, being formed by sequence-specific interaction between homologous protein patches result in highly pure protein microparticles [9]. Since they are also porous and highly hydrated [19], efficient substrate diffusion would probably occur for most of the (or at least many) biotechnologically relevant aggregated enzymes, thus opening the possibility for a new industrial market of enzymatically active inclusion bodies.

Conclusion
Results presented here prove that aggregation of recombinant proteins as bacterial inclusion bodies does not nec-

essarily inactivate them, despite the enriched intermolecular β-sheet structure observed in those formed by the tested model proteins. The extent of protein activity varies depending on the specific protein, but even the lowest functional values observed are still high enough to consider the use of inclusion body enzymes in bioprocesses, without any previous refolding step. The eventual incorporation of inclusion bodies in industrial catalysis could represent an important conceptual shift in the biotechnology market.

Methods
Strain, plasmids and culture conditions
E. coli MC4100 [20] was used for all the experiments. Plasmids encoding hDHFR and Aβ42(F19D)-BFP have been previously described and appropriate references can be found in Table 1. Briefly, in the Aβ42(F19D)-BFP vector (6.7 Kb) the DNA sequence encoding the 42-mer Alzheimer's amyloid peptide, (bearing a Phe[19]→Asp mutation to reduce its *in vivo* aggregation rate), is fused upstream of the BFP gene and under the control of the T7 promoter, in a pET-28 based vector. In the product, the two protein sequences were separated by 12-mer linker stretch to provide flexibility to the fusion protein and limit steric constraints between domains. pTVP1LAC was constructed by moving the *SalI-NcoI* VP1LAC fusion-encoding DNA segment (3.5 Kb) from pJVP1LAC (8.5 Kb) to the cloning vector pTRC99A [20]. The resulting pTVP1LAC construct

(7.7 Kb) was used to direct the production of VP1LAC. The *lacZ* gene was further replaced there by an appropriate GFP-encoding DNA segment (0.7 Kb) through digestion with *EcoRI* and *BamHI*, rendering pTVP1GFP (5.5 Kb). All the production processes were performed in shaker-flask cultures growing at 37°C in LB rich medium [20] plus 100 μg/ml ampicillin for plasmid maintenance, and recombinant gene expression was induced when the OD_{550} reached 0.4, by adding 1 mM IPTG. Cell samples were taken at 3 and 5 h after induction of gene expression.

Analysis of enzymatic activity
Culture samples of 2.5 ml were jacketed in ice, disrupted by sonication for 5 min at 50 W under 0.5 s cycles [21] and centrifuged at 4°C for 15 min at 15000 g. The supernatant was directly used for the analysis as the soluble cell fraction. Inclusion bodies were purified by a detergent-washing protocol as described [19] and used in suspension for activity analysis. β-Galactosidase activity of both soluble cell fraction and inclusion bodies of VP1LAC was determined in microplates as described [7,22] under continuous stirring at 250 rpm. Kinetic analysis of VP1LAC enzymatic activity was monitored in 120 μl reaction mixtures with either 2 mM ONPG (pH 8.4) or 2 mM CPRG (pH 7.0). The hDHFR activity was determined by incubating 50 μl of the protein sample and 850 μl of the appropriate assay buffer (0.1 M K_3PO_4 pH 7.4, 1 mM DTT, 0.5 M KCl, 1 mM EDTA and 20 mM ascorbic acid) for 10 minutes at room temperature. Then, 50 μl of 2 mM 7,8-dihidrofolate and 50 μl of 2 mM NADPH were added and hDHFR activity was recorded every 15 seconds during 4 minutes at 340 nm. Protein concentration in all the assays was adjusted between 2 and 3 μg/ml.

Fluorescence (at 510 nm for GFP and 460 nm for BFP) was recorded in a Perkin-Elmer 650-40 fluorescence spectrophotometer by using excitation wavelengths of 450 nm and 360 nm for GPF and BFP respectively. Fluorescence was measured in 1 ml samples using dilutions when necessary. Both enzymatic activities and fluorescence were determined in triplicate.

Quantitative protein analysis
Samples of bacterial cultures (10 ml) were low-speed centrifuged (15 min at 12000 g) to harvest the cells. For protein quantification in soluble cell fractions, samples were resuspended in 400 μl of Z buffer without β-mercaptoethanol [23] with one tablet of protease inhibitor cocktail (Roche, ref. 1 836 170) per 10 ml buffer. Such mixtures, once jacketed in ice, were sonicated for 5 min (or longer when required to achieve a complete disruption) at 50 W under 0.5 s cycles as described [21], and centrifuged for 15 min at 12000 g. The supernatant was mixed with denaturing buffer at appropriate ratios [24]. For the determination of inclusion body protein, these structures were

purified by repeated detergent washing as described [19] and resuspended in denaturing buffer [24]. After boiling for 20 min, appropriate sample volumes were loaded onto denaturing gels. For Western blot, polyclonal antibodies specific for each protein were used as previously described [17]. Dried blots were scanned at high resolution and bands quantified by using the Quantity One software from Bio Rad, by using appropriate protein dilutions of known concentration as controls. Determinations were always done within the linear range and they were used to calculate the specific activity values.

Conformational analysis by FTIR spectroscopy
For FTIR spectroscopy analysis, purified inclusion bodies were dried for two hours in a Seepd-Vac system before analysis to reduce water interference in the infrared spectra. The FTIR spectrum of the dry samples was analysed directly in a Bruker Tensor FTIR spectrometer. All processing procedures were carried out so as to optimise the quality of the spectrum in the amide I region, between 1600 cm^{-1} and 1700 cm^{-1}. Second derivatives of the amide I band spectra were used to determine the frequencies at which the different spectral components were located. A general description of FTIR procedures can be found elsewhere [9,10].

Abbreviations
BFP blue fluorescent protein

CPRG phenol red β-D-galactopyranoside

FMDV foot-and-mouth disease virus

FTIR fourier transform infrared

GFP green fluorescent protein

HDHFR human dihydropholate reductase

IPTG isopropyl-β-D-thiogalactopyranoside

ONPG ortho-nitrophenyl β-D-galactopyranoside

Authors' contributions
EGF performed most of the experiments and prepared the final data and figures. NGM, A. Vera and AA analysed protein amounts by Western blot, RMF performed enzyme kinetics, MM performed part of optical microscopy analysis and SV part of FTIR analysis and data interpretation. A. Villaverde directed the work and prepared the manuscript.

Acknowledgements

References

1. Baneyx F, Mujacic M: **Recombinant protein folding and misfolding in Escherichia coli.** *Nat Biotechnol* 2004, **22**:1399-1408.
2. Strandberg L, Enfors SO: **Factors influencing inclusion body formation in the production of a fused protein in Escherichia coli.** *Appl Environ Microbiol* 1991, **57**:1669-1674.
3. Rinas U, Tsai LB, Lyons D, Fox GM, Stearns G, Fieschko J, Fenton D, Bailey JE: **Cysteine to serine substitutions in basic fibroblast growth factor: effect on inclusion body formation and proteolytic susceptibility during in vitro refolding.** *Biotechnology (N Y)* 1992, **10**:435-440.
4. Baneyx F, Palumbo JL: **Improving heterologous protein folding via molecular chaperone and foldase co-expression.** *Methods Mol Biol* 2003, **205**:171-197.
5. Vallejo LF, Rinas U: **Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins.** *Microb Cell Fact* 2004, **3**:11.
6. Tokatlidis K, Dhurjati P, Millet J, Beguin P, Aubert JP: **High activity of inclusion bodies formed in Escherichia coli overproducing Clostridium thermocellum endoglucanase D.** *FEBS Lett* 1991, **282**:205-208.
7. Garcia-Fruitos E, Carrio MM, Aris A, Villaverde A: **Folding of a misfolding-prone beta-galactosidase in absence of DnaK.** *Biotechnol Bioeng* 2005, **90**:869-875.
8. Worrall DM, Goss NH: **The formation of biologically active beta-galactosidase inclusion bodies in Escherichia coli.** *Aust J Biotechnol* 1989, **3**:28-32.
9. 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.
10. Ami D, Natalello A, Gatti-Lafranconi P, Lotti M, Doglia SM: **Kinetics of inclusion body formation studied in intact cells by FT-IR spectroscopy.** *FEBS Lett* 2005, **579**:3433-3436.
11. Sorensen HP, Mortensen KK: **Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli.** *Microb Cell Fact* 2005, **4**:1.
12. de Marco A, Schroedel A: **Characterization of the aggregates formed during recombinant protein expression in bacteria.** *BMC Biochem* 2005, **6**:10.
13. 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**:137-143.
14. Wright PE, Dyson HJ: **Intrinsically unstructured proteins: reassessing the protein structure-function paradigm.** *J Mol Biol* 1999, **293**:321-331.
15. Waldo GS, Standish BM, Berendzen J, Terwilliger TC: **Rapid protein-folding assay using green fluorescent protein.** *Nat Biotechnol* 1999, **17**:691-695.
16. Carrio MM, Villaverde A: **Protein aggregation as bacterial inclusion bodies is reversible.** *FEBS Lett* 2001, **489**:29-33.
17. Cazorla D, Feliu JX, Villaverde A: **Variable specific activity of Escherichia coli beta-galactosidase in bacterial cells.** *Biotechnol Bioeng* 2001, **72**:255-260.
18. Oberg K, Chrunyk BA, Wetzel R, Fink AL: **Nativelike secondary structure in interleukin-1 beta inclusion bodies by attenuated total reflectance FTIR.** *Biochemistry* 1994, **33**:2628-2634.
19. Carrio MM, Cubarsi R, Villaverde A: **Fine architecture of bacterial inclusion bodies.** *FEBS Lett* 2000, **471**:7-11.
20. Sambrook J, Fritsch E, Maniatis T: *Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY* 1989.
21. Feliu JX, Cubarsi R, Villaverde A: **Optimized release of recombinant proteins by ultrasonication of E. coli cells.** *Biotechnol Bioeng* 1998, **58**:536-540.
22. Ferraz RM, Aris A, Villaverde A: **Profiling the allosteric response of an engineered beta-galactosidase to its effector, anti-HIV antibody.** *Biochem Biophys Res Commun* 2004, **314**:854-860.
23. Miller JH: *Experiments in Molecular Genetics Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY* 1972.
24. Laemmli UK: **Cleavage of structural proteins during the assembly of the head of bacteriophage T4.** *Nature* 1970, **227**:680-685.
25. Davies JF, Delcamp TJ, Prendergast NJ, Ashford VA, Freisheim JH, Kraut J: **Crystal structures of recombinant human dihydrofolate reductase complexed with folate and 5-deazafolate.** *Biochemistry* 1990, **29**:9467-9479.
26. Sánchez de Groot N, Avilés FX, Vendrell J, Ventura S: **Mutagenesis of the central hydrophobic cluster in Ab42 Alzheimer's peptide. Simple rules to predict the aggregation propensities of polypeptides.** submitted 2005.

IV.3 Paper 3

Localization of functional polypeptides in bacterial inclusion bodies

Elena García-Fruitós, Anna Arís and Antonio Villaverde

Applied and Environmental Microbiology, Vol. 73 No. 1, January, 2007

In the previous work, we determined that the specific activity of recombinant proteins found in both soluble and insoluble cell fractions is not dramatically different, suggesting the potential use of purified inclusion bodies as functional particles in catalytic bioprocesses. To better characterise this potential, we explored in detail the molecular organisation of these aggregates.

By using VP1GFP as model protein, we identified the localization of fluorescence by image analysis of VP1GFP inclusion bodies. Specifically, we proved that the fluorescence distribution in inclusion bodies formed at different temperatures is not homogeneous in all the aggregate but is concentrated in the core, being the surface layer poor in functional fluorescent protein (see also annex I). Additionally, this result allows us to state that the occurrence of properly folded and functional protein in inclusion bodies is not a mere contaminant unspecifically attached to the aggregates' surface, but a natural component of inclusion bodies.

On the other hand, we also analysed both the activity and the amount of protein that remains associated to VP1LAC inclusion bodies in a reaction mixture in presence of a β-galactosidase substrate. Upon resuspension, a significant fraction of functional protein is immediately released to the solvent, indicating that active forms of VP1LAC might be exposed to the inclusion body-solvent interface due to the highly porous and hydrated nature of these aggregates. When comparing this behaviour in the presence or absence of its substrate, we determined that, interestingly, soluble functional protein release in the reaction mixtures might be enhanced in the presence of its specific substrate. Protein release from VP1GFP inclusion bodies is, instead, very low.

Localization of Functional Polypeptides in Bacterial Inclusion Bodies[▽]

Elena García-Fruitós, Anna Arís, and Antonio Villaverde*

*Institut de Biotecnologia i de Biomedicina and Departament de Genètica i de Microbiologia,
Universitat Autònoma de Barcelona, Bellaterra, 08193 Barcelona, Spain*

Bacterial inclusion bodies, while showing intriguing amyloid-like features, such as a β-sheet-based intermolecular organization, binding to amyloid-tropic dyes, and origin in a sequence-selective deposition process, hold an important amount of native-like secondary structure and significant amounts of functional polypeptides. The aggregation mechanics supporting the occurrence of both misfolded and properly folded protein is controversial. Single polypeptide chains might contain both misfolded stretches driving aggregation and properly folded protein domains that, if embracing the active site, would account for the biological activities displayed by inclusion bodies. Alternatively, soluble, functional polypeptides could be surface adsorbed by interactions weaker than those driving the formation of the intermolecular β-sheet architecture. To explore whether the fraction of properly folded active protein is a natural component or rather a mere contaminant of these aggregates, we have explored their localization by image analysis of inclusion bodies formed by green fluorescent protein. Since the fluorescence distribution is not homogeneous and the core of inclusion bodies is particularly rich in active protein forms, such protein species cannot be passively trapped components and their occurrence might be linked to the reconstruction dynamics steadily endured in vivo by such bacterial aggregates. Intriguingly, even functional protein species in inclusion bodies are not excluded from the interface with the solvent, probably because of the porous structure of these particular protein aggregates.

Procedures for in vitro protein refolding are under continuous development (22, 32, 35), since many proteins of industrial or pharmacological interest are produced in recombinant microorganisms, especially bacteria, as insoluble aggregates called inclusion bodies (IBs) (38). Recent insights into the structure and physiology of bacterial IBs have revealed that at least a significant fraction of the embedded protein occurs in a properly folded native-like form (36) and that for aggregates formed by enzymes, this fact is reflected by the occurrence of enzymatic activity associated with these particles (16, 34, 39). While for hormones or other drugs to be used in vivo, in vitro solubilization of IBs and refolding of IB proteins would still be required to allow their proper use (27), enzymes to be used in bioprocesses could be employed straight after production, skipping any refolding step. This is particularly appealing since the specific activity found in IB enzymes, although variable when comparing different protein species, is not dramatically different from that exhibited by the soluble counterparts (15, 16), and on the other hand, refolding procedures render yields of active protein that are usually far from 100%.

Apart from the obvious potential of enzyme IBs as catalyzers, the occurrence of properly folded, active enzymes poses intriguing structural questions. The conformational background sustaining the IB molecular structure lies on an extended, intermolecular β-sheet architecture (6) that coexists with various amounts of a population of native-like, correctly folded polypeptides (1–3, 26). Such a β-sheet pattern is progressively lost at the expense of native-like structure when the temperature at which IBs are formed decreases (19, 28), indicating the existence of several categories of protein aggregates in regard to their molecular organization and even global morphology (13). However, the process that selects in vivo the protein species to be deposited with regard to its conformational status, and especially the way in which both properly folded and β-sheet-rich species coexist, remains unexplored. It has been suggested that in single polypeptides, specific domains with a misfolded status could act as aggregating elements (and organize through intermolecular β-sheet interactions), while others might remain properly folded (and fully functional if containing the active site) (36). On the other hand, it is known that during IB isolation, contaminating cell proteins get attached to the IB surface (17), which is probably "sticky" from the exposure of hydrophobic patches (6). Therefore, active polypeptides also eventually could be surface trapped if they abound in the soluble cell fraction of the producing cell. In that case, the enzymatic activity displayed by IBs would lie on contaminant protein species rather than on structural elements. The experimental approaches to solving this issue are not obvious, since it is not possible to distinguish in situ conformational states within such nanoparticles.

To gain insights into the molecular organization of functional IBs, we have taken alternative approaches, namely, exploring the localization of the biological activity in actively catalyzing enzyme-based IBs and generating fluorescence emission maps of green fluorescent protein (GFP)-based IBs. Intriguingly, although an important part of the active protein species is easily released from catalyzing IBs, the core of such aggregates (but not the surface layers) is rich in functional protein. These results are discussed in the context of the porous structure of these protein aggregates, also considering the highly dynamic protein deposition and release processes that drive the in vivo building of bacterial IBs.

---

* Corresponding author. Mailing address: Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Bellaterra, 08193 Barcelona, Spain. Phone: 34 935812148. Fax: 34 935812011. E-mail: avillaverde@servet.uab.es.

MATERIALS AND METHODS

Strain, plasmids, and culture conditions. *Escherichia coli* MC4100 (29) was used for all the experiments. Plasmids pTVP1GFP and pTVP1LAC (16) encode engineered versions of GFP and β-galactosidase, respectively, both carrying the VP1 capsid protein of foot-and-mouth disease virus fused at the amino termini. This viral protein dramatically reduces the solubility of the whole fusions, resulting in aggregation of fusion proteins as IBs. All of the production processes were performed with shaker flask cultures growing at 37°C in LB rich medium (29) plus 100 μg/ml ampicillin for plasmid maintenance, and recombinant gene expression was induced when the optical density at 550 nm reached 0.4 by adding 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG). Cell samples were taken at 3 h after induction of gene expression. For the comparative analysis of IBs formed at different temperatures, samples were taken from IPTG-treated cultures at an optical density between 2.9 and 3.1, irrespective of the time taken for growth (3 h at 37°C but longer at lower temperatures).

IB-mediated catalysis conditions and determination of enzymatic activity. IBs were purified by a detergent-washing protocol as described previously (8), resuspended in phosphate-buffered saline (PBS), and diluted either 5 or 50 times for VP1GFP or VP1LAC, respectively. For the analysis of VP1LAC, two aliquots of each sample were prepared and kept at 37°C in agitation. In one of them, 5 ml of 6 mM o-nitrophenyl-β-D-galactopyranoside (ONPG) (in PBS) was added, while the other was used as an internal control. To monitor the ONPG hydrolysis reaction, samples were taken every 5 min for 1 h and the enzymatic activity of VP1LAC was determined in 120-μl reaction mixtures in microplates, as described previously (14, 15). Also, to determine the localization of the enzymatic activity in such a reaction mixture, at different times of the catalysis process, 1-ml samples were taken at three different times (t_0, just before the ONPG addition), t_1, and t_2 (2 min and 30 min after ONPG addition, respectively). These samples were centrifuged at a low speed (for 5 min at 15,000 × g), and the supernatant was used for the analysis of the soluble fraction of the enzymatic reaction (associated with protein released from IBs), while the resulting pellet, resuspended in 1 ml PBS, was used for the analysis of the protein still associated with IBs. VP1GFP IBs prepared as described were incubated at 37°C, and samples taken at different times were centrifuged at 15,000 × g for 5 min. Fluorescence of both soluble and insoluble fractions was determined in a Cary Eclipse fluorescence spectrophometer (Variant).

To analyze the enzymatic activity in these fractions and finally the specific activity, a second substrate, rendering red products upon hydrolysis by β-galactosidase (chlorophenol red β-D-galactopyranoside), was used to avoid the yellow background linked to the ONPG products already present in the samples. These assays were performed with 120-μl reaction mixtures in microplates with 6 mM chlorophenol red β-D-galactopyranoside, as described previously (14), for 16 h. The enzymatic activity was calculated by measuring the slope of the linear part of each graph plotted against the reaction time. All determinations were done with at least three independent experiments.

Quantitative protein analysis. For protein quantification, supernatants and IB fractions were boiled for 15 or 25 min, respectively. Appropriate sample volumes were loaded onto denaturing gels for immunodetection. For Western blotting, polyclonal antibodies specific for β-galactosidase was used as previously described (12). Bands were quantified by means of the Quantity One software from Bio-Rad, using appropriate protein dilutions of known concentrations as controls. All of these analyses were done in at least three independent experiments. Protein amounts were finally employed to determine specific activities of the distinct samples.

Confocal microscopy analysis. For image analysis, samples of VP1GFP-producing cells 3 h after IPTG addition were fixed with 0.1% formaldehyde and stored at 4°C until observed. Photographs were taken by using a Leica TCS SP2 AOBS confocal microscope (excitation wavelength at 488 nm and emission wavelength at 500 to 600 nm; optical lens magnification, 63×; 1,024 by 1,024 pixels; zooms between 4 and 8). For the analysis of the resulting images, we used the Adobe Photoshop software and two different lookup tables (tables of cross-references linking index numbers to output values) with coincident results. By this, we determined the colors and intensity values with which a particular image is to be displayed, producing color maps in which each pixel's value is treated as an index number instead of a definite color. The particular lookup table displayed in the Results section was "Metamorph."

RESULTS

In vivo distribution of active IB polypeptides. In a previous study, we analyzed in situ β-galactosidase protein material in
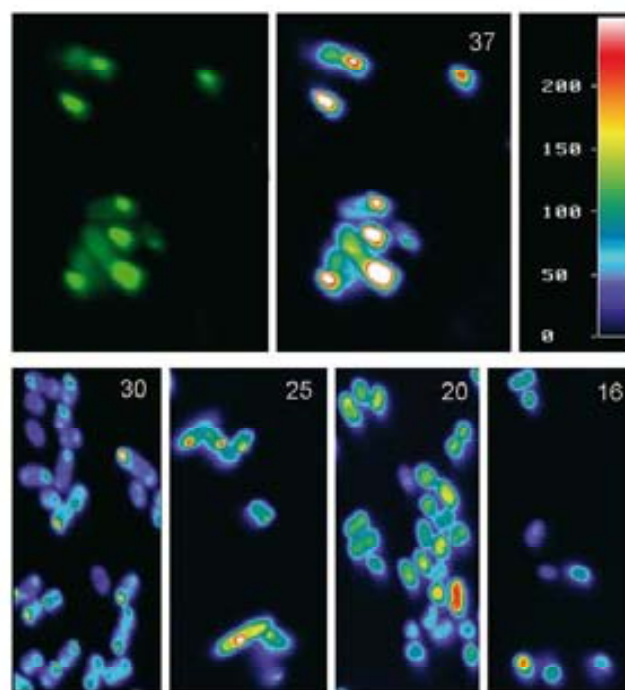

FIG. 1. Top. Fluorescence microscopy (left) and Metamorph image analysis (right) of VP1GFP IBs formed at 37°C. The color scale is depicted at the right. Bottom. Metamorph image analysis of IBs formed at different growth temperatures (indicated by numbers at the upper right corners).

sections of IB-bearing cells by immunodetection without noting any specific distribution of the enzyme in the aggregates (see Fig. 1 in reference 11). Since most of the well-formed, aged VP1LAC IBs are composed by VP1LAC (up to more than 90% of the IB protein material; see Fig. 4 in reference 7), this fact indicates that the density of IBs is rather homogeneous. Although specific IB protein density has not been directly investigated for other proteins and production conditions, the common aggregation mechanics (5), secondary structure pattern (2, 3, 6, 16, 19), and architectural data independently obtained from different IBs (5, 8) do not point out the homogenous distribution of IB polypeptides as being a particular, protein-restricted feature. Since it was not technically possible to map in situ the occurrence of active VP1LAC in VP1LAC IBs, we instead analyzed the fluorescence distribution in VP1GPF IBs to identify the localization of functional protein and any possible unbalanced distribution of fluorescent protein material. Like VP1LAC IBs, the aggregates formed by VP1GFP are highly active, and they are fluorescent (16). In this regard, confocal analysis of VP1GFP-producing cells through 0.04-μm virtual sections (note that IB diameter occurs between around 0.5 and 1 μm [7, 8, 11, 16]) rendered intriguing images in which there was a clear gradation in the emission intensity from low (external layer) to high (the IB core) (Fig. 1, top). The same pattern was consistently observed at suboptimal growth temperatures, namely, 30, 25, 20, and 16°C, known to favor both protein solubility (33, 37) and conformational quality of IB polypeptides (19, 37). Even at 16°C, when refrac-
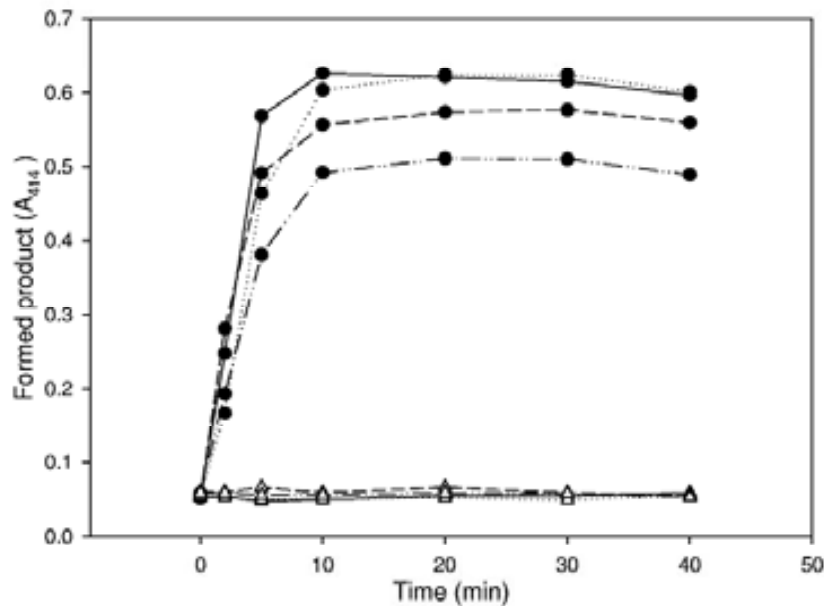
FIG. 2. Formation of the ONPG hydrolysis product as mediated by VP1LAC IBs (filled symbols) in four independent experiments. As a control, the absorbance of IB samples in the absence of ONPG is also shown (empty symbols).

tile IBs are hardly formed (37), the concentric fluorescence pattern was observed in some individual cells.

As indicated above, the obtained emission maps cannot be accounted for by any strong radial distribution of protein density. Therefore, functional, properly folded polypeptides are specifically found at the core of the aggregates, while their surface layer is poor in functional protein (note that the differences in the fluorescence emissions between such protein populations are at least twofold [Fig. 1]). This fact could be due to the dynamics of the in vivo IB building process that results from an unbalanced equilibrium between protein deposition and removal (9, 10). Disaggregating chaperones, namely DnaK, ClpB, and small heat shock proteins, act cooperatively on misfolded polypeptides at the aggregate interface (23, 25, 30, 31). Since, as derived from IB structural analysis (1, 2, 3, 26, 28), protein aggregation is not a highly selective process that involves functional polypeptides, a more selective removal of misfolded proteins at the IBs' surfaces (as suggested (30, 31) could enrich the nucleus with native-like, active species. This possibility is compatible with the gain of conformational homogeneity observed during the volumetric growth of IB (8), since the ratio between core and surface material increases with IB volume.

Distribution and release of active polypeptides in catalyzing IBs. In a previous work (16), we suggested that enzyme-based IBs, since they contain functional proteins, could be useful catalyzers in enzymatic processes, and in fact, both β-galactosidase and human dihydrofolate reductase efficiently processed their respective substrates as embedded in IBs. To better understand how the reaction is performed by IBs in the context of the activity distribution seen in Fig. 1, we determined the occurrence of β-galactosidase enzymatic activity during substrate hydrolysis mediated by VP1LAC IBs. In the presence of an enzyme substrate (ONPG), resuspended VP1LAC IBs cat-

alyzed the product formation kinetics with a very conventional profile (Fig. 2). Since IBs are highly porous and hydrated structures (5, 8), substrate diffusion to the core would not be unexpected. However, to explore to what extent such an enzymatic process was directed by enzyme molecules associated with or released from IBs, we determined the enzymatic activities in the insoluble and soluble fractions of the reaction mixture at different times of the process, as well as the enzyme present in each fraction. Intriguingly, a significant part of the enzymatic activity (between 7 and 8%) was found in the soluble fraction upon IB resuspension in the reaction buffer before substrate addition (Fig. 3A, time zero). Note that this occurred after IB isolation by a procedure that involves repeated detergent washing steps (8). Since the protein amount in the soluble fraction was very low (not shown; lower than 0.0002%), such a protein fraction must exhibit a specific activity higher than average for the aggregates and would not be linked to surface polypeptides. The immediate release of functional protein was also observed for VP1GFP IBs to an extent very similar to that for VP1LAC (5.0%, Fig. 3C). This fact suggested that fluorescent VP1GFP polypeptides, since they are not located at the IB surface layer, might be not completely excluded from the interface with the solvent because of the highly porous and hydrated IB architecture (5, 8). A similar situation could take place with VP1LAC IBs if their functional architecture is comparable to that of VP1GFP IBs. Interestingly, during substrate hydrolysis, an increasing fraction of the enzymatic activity is found not to be linked to IBs (Fig. 3A), and at 30 min, it essentially represents the total activity in the reaction mixture. In the absence of a substrate, VP1LAC IBs incubated under the same conditions also split the activity in soluble and insoluble fractions but to much lesser extent (up to around 50% after a 30-min incubation) (Fig. 3B). Despite the fact that this substrate-mediated modulation of the activity fractioning was
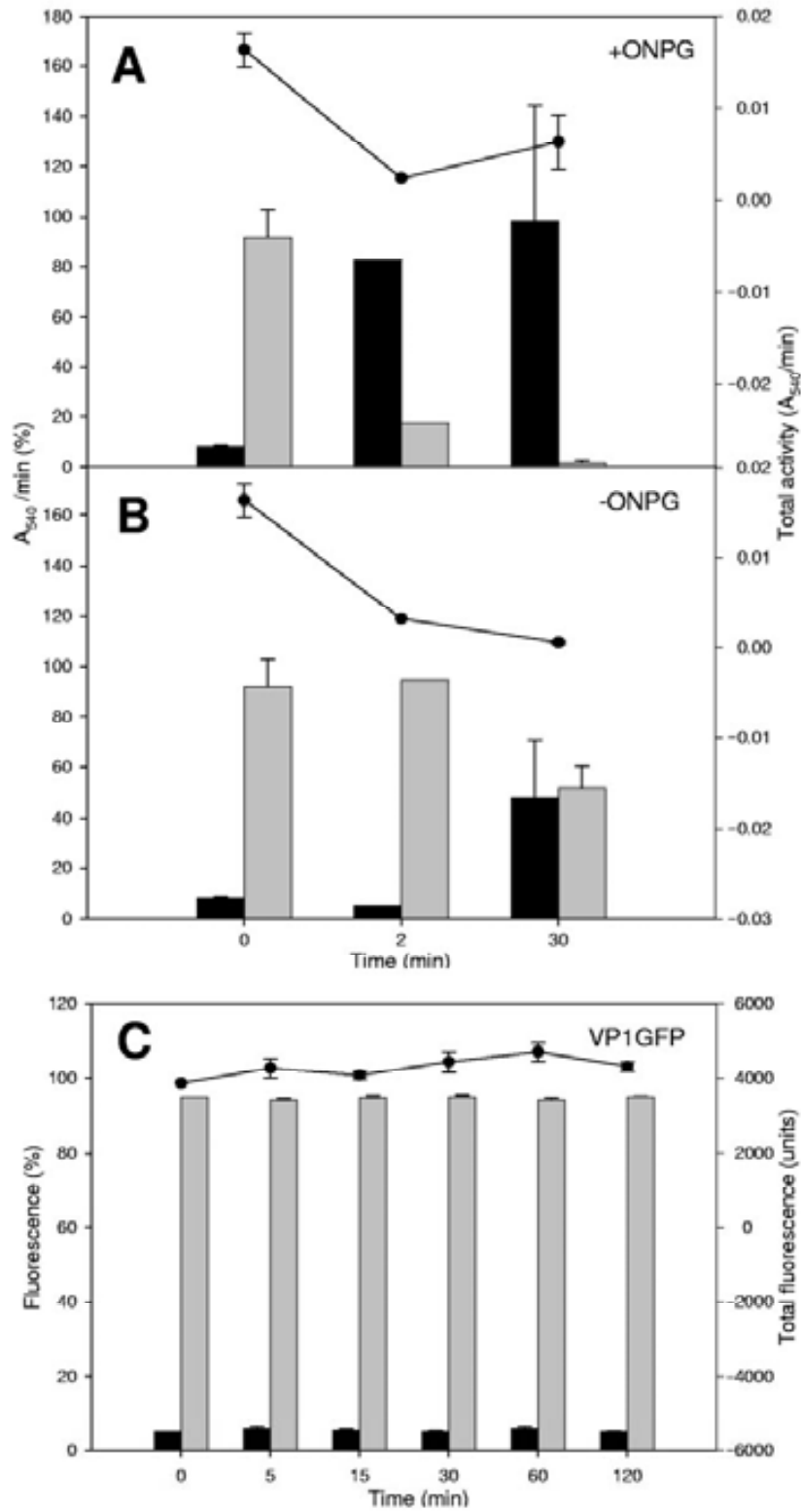
FIG. 3. Relative biological activity (measured by either absorbance or fluorescence; left scale) in suspended VP1LAC (A, B) or VP1GFP (C) IBs. Soluble (black bars) and insoluble (gray bars) fractions were analyzed. The total activity in absolute values is shown as dots (top, right scale).

FUNCTIONAL MAPPING OF INCLUSION BODIES   293

clear in every individual experiment, we obtained only fairly significant differences ($P = 0.057$), probably because of the high variability found between experiments. It must be noted, however, that the total enzymatic activity decreased by more than sixfold in VP1LAC IBs when the substrate was absent (Fig. 3B) but only moderately in actively catalyzing reaction mixtures (with the substrate) (Fig. 3A). Also, 30 min after substrate addition, the specific activity of soluble VP1LAC was estimated to be 10-fold higher than the average for the remaining IB protein species (not shown). On the other hand, the fluorescence of VP1GFP remains associated to IBs for a long time (Fig. 3C), apart from the small fraction that remained constant after being immediately released.

DISCUSSION

Despite the structural similarities recently recognized between IBs and amyloids (6), bacterial aggregates are formed by an unbalanced, highly dynamic equilibrium between protein deposition and removal (10, 38), which implies a continuous reconstruction through exchange of polypeptides between the soluble and insoluble cell fractions (36). The occurrence of native-like structure in IB protein (1, 2, 3, 6, 26, 28) indicates that the aggregation process is not highly selective, involving polypeptides that at least to a significant extent are properly folded. As a consequence, instead of being inert structures, IBs, when formed by proteins with measurable biological activity, result in active nanoparticles with potential applications in catalytic bioprocesses (16, 34, 39). The extent of active (properly folded) protein in IBs is variable depending on the specific polypeptide (16), the environmental conditions under which IBs have been formed (such as temperature or the curve growth phase) (15, 19, 28), and the genetic background of the producer strain (15, 18). This conformational variability can be better understood in the context of a continuum of forms that aggregates formed in bacteria can adopt, including loose aggregates occurring in the soluble cell fraction, amyloid-like fibers, aggregates in the insoluble cell fraction, and conventional, refractile IBs (13). In true IBs, the coexistence of both active and inactive polypeptides has generated intriguing discussions about how such protein species could coexist and in particular if single polypeptides could exhibit both properly folded domains, accounting for the native-like structure observed in infrared spectroscopy analysis (1, 2, 3, 26) (and conferring biological activity if embracing the active site), and misfolded protein stretches, responsible for the intermolecular beta-sheet organization supporting the IB architecture (6, 28).

We have proved in this study that the localization of fluorescence emission in GFP-containing IBs is not homogeneous in all of the aggregate body but is concentrated in its core (Fig. 1). Although the approach used does not allow determination of the extent of misfolded portions of VP1GFP, which does not affect the fluorophore performance, this observation indicates that the active protein (with global proper folding) is not limited to the aggregate surface, which could have been eventually accounted for by in vivo sequence-specific association of soluble and functional polypeptides from the soluble cell fraction to the IB's surface. The fluorescence distribution pattern is consistent when observing IBs formed at different temperatures below 37°C, known to minimize aggregation (33) but

enhance the conformational quality of the embedded protein (37). Therefore, the occurrence of functional protein is not an artifact from a weakly stringent purification process, but such active forms are a structural, natural component of IBs. Interestingly, a significant fraction of functional protein is immediately released to the solvent upon resuspension, indicating that despite their nuclear localization, active forms might be exposed to the IB-solvent interface. This can be accounted for by the highly porous architecture and hydrated nature of the IB (5, 8), which must be also supportive of substrate diffusion in IB-mediated catalysis (16) (Fig. 2). Obviously we cannot completely discard some an extent of spontaneous refolding of surface-attached inactive protein, but the progressive loss of activity in catalyzing VP1LAC IBs prompts us to favor the hypothesis of active protein release. In this context, the appearance of soluble functional protein in the reaction mixtures might be enhanced in catalyzing IBs, while it does not occur in VP1GFP IBs and occurs only moderately in VP1LAC IBs in the absence of the enzyme substrate (Fig. 3). Although the variability of the obtained data prevented robust significant support of this hypothesis, it is likely that the catalytic process itself would induce subtle conformational modifications in the active, aggregated polypeptides, promoting their release. Despite the fact that molecular chaperones are tightly associated to IBs (4, 11, 20, 21), more research is needed to know whether such protein release in vitro is modulated by such cell proteins or rather is mechanical process.

On the other hand, the core localization of functional protein could be due to different selectivities of aggregating and disaggregating polypeptides in vivo regarding the conformational status. If, as suggested, the chaperone (or protease)-mediated release from aggregates is surface restricted and specifically targeted to misfolded species (24, 30, 31) (while aggregation seem to be less selective regarding the folding status), the unbalanced equilibrium favoring the in vivo volumetric growth of IBs would progressively enrich the aggregates' core with properly folded polypeptides.

REFERENCES

1. Ami, D., L. Bonecchi, S. Cali, G. Orsini, G. Tonon, and S. M. Doglia. 2003. FT-IR study of heterologous protein expression in recombinant *Escherichia coli* strains. Biochim. Biophys. Acta 1624:6–10.
2. Ami, D., A. Natalello, P. Gatti-Lafranconi, M. Lotti, and S. M. Doglia. 2005. Kinetics of inclusion body formation studied in intact cells by FT-IR spectroscopy. FEBS Lett. 579:3433–3436.
3. Ami, D., A. Natalello, G. Taylor, G. Tonon, and D. S. Maria. 2006. Structural analysis of protein inclusion bodies by Fourier transform infrared microspectroscopy. Biochim. Biophys. Acta 1764:793–799.
4. Boels, K., M. M. Carrió, A. Aris, J. L. Corchero, and A. Villaverde. 1999. Distinct chaperone affinity to folding variants of homologous recombinant proteins. Biotechnol. Lett. 21:531–536.
5. Bowden, G. A., A. M. Paredes, and G. Georgiou. 1991. Structure and morphology of protein inclusion bodies in *Escherichia coli*. Biotechnology (N. Y.) 9:725–730.
6. Carrio, M., N. Gonzalez-Montalban, A. Vera, A. Villaverde, and S. Ventura. 2005. Amyloid-like properties of bacterial inclusion bodies. J. Mol. Biol. 347:1025–1037.
7. Carrio, M. M., J. L. Corchero, and A. Villaverde. 1998. Dynamics of in vivo protein aggregation: building inclusion bodies in recombinant bacteria. FEMS Microbiol. Lett. 169:9–15.

8. Carrio, M. M., R. Cubarsi, and A. Villaverde. 2000. Fine architecture of bacterial inclusion bodies. FEBS Lett. 471:7–11.
9. Carrio, M. M., and A. Villaverde. 2001. Protein aggregation as bacterial inclusion bodies is reversible. FEBS Lett. 489:29–33.
10. Carrio, M. M., and A. Villaverde. 2002. Construction and deconstruction of bacterial inclusion bodies. J. Biotechnol. 96:3–12.
11. Carrio, M. M., and A. Villaverde. 2005. Localization of chaperones DnaK and GroEL in bacterial inclusion bodies. J. Bacteriol. 187:3599–3601.
12. Cazorla, D., J. X. Feliu, and A. Villaverde. 2001. Variable specific activity of Escherichia coli beta-galactosidase in bacterial cells. Biotechnol. Bioeng. 72:255–260.
13. de Marco, A., and A. Schroedel. 2005. Characterization of the aggregates formed during recombinant protein expression in bacteria. BMC Biochem. 6:10.
14. Ferraz, R. M., A. Aris, and A. Villaverde. 2004. Profiling the allosteric response of an engineered beta-galactosidase to its effector, anti-HIV antibody. Biochem. Biophys. Res. Commun. 314:854–860.
15. Garcia-Fruitos, E., M. M. Carrio, A. Aris, and A. Villaverde. 2005. Folding of a misfolding-prone beta-galactosidase in absence of DnaK. Biotechnol. Bioeng. 90:869–875.
16. Garcia-Fruitos, E., N. Gonzalez-Montalban, M. Morell, A. Vera, R. M. Ferraz, A. Aris, S. Ventura, and A. Villaverde. 2005. Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. Microb. Cell Fact. 4:27.
17. Georgiou, G., and P. Valax. 1999. Isolating inclusion bodies from bacteria. Methods Enzymol. 309:48–58.
18. Gonzalez-Montalban, N., E. Garcia-Fruitos, S. Ventura, A. Aris, and A. Villaverde. 2006. The chaperone DnaK controls the fractioning of functional protein between soluble and insoluble cell fractions in inclusion body-forming cells. Microb. Cell Fact. 5:26.
19. Jevsevar, S., V. Gaberc-Porekar, I. Fonda, B. Podobnik, J. Grdadolnik, and V. Menart. 2005. Production of nonclassical inclusion bodies from which correctly folded protein can be extracted. Biotechnol. Prog. 21:632–639.
20. Jurgen, B., H. Y. Lin, S. Riemschneider, C. Scharf, P. Neubauer, R. Schmid, M. Hecker, and T. Schweder. 2000. Monitoring of genes that respond to overproduction of an insoluble recombinant protein in Escherichia coli glucose-limited fed-batch fermentations. Biotechnol. Bioeng. 70:217–224.
21. Lethanh, H., P. Neubauer, and F. Hoffmann. 2005. The small heat-shock proteins IbpA and IbpB reduce the stress load of recombinant Escherichia coli and delay degradation of inclusion bodies. Microb. Cell Fact. 4:6.
22. Li, M., Z. G. Su, and J. C. Janson. 2004. In vitro protein refolding by chromatographic procedures. Protein Expr. Purif. 33:1–10.
23. Mogk, A., E. Deuerling, S. Vorderwulbecke, E. Vierling, and B. Bukau. 2003. Small heat shock proteins, ClpB and the DnaK system form a functional triade in reversing protein aggregation. Mol. Microbiol. 50:585–595.
24. Mogk, A., D. Dougan, J. Weibezahn, C. Schlieker, K. Turgay, and B. Bukau.

2004. Broad yet high substrate specificity: the challenge of AAA+ proteins. J. Struct. Biol. 146:90–98.
25. Mogk, A., C. Schlieker, K. L. Friedrich, H. J. Schonfeld, E. Vierling, and B. Bukau. 2003. Refolding of substrates bound to small Hsps relies on a disaggregation reaction mediated most efficiently by ClpB/DnaK. J. Biol. Chem. 278:31033–31042.
26. Oberg, K., B. A. Chrunyk, R. Wetzel, and A. L. Fink. 1994. Nativelike secondary structure in interleukin-1 beta inclusion bodies by attenuated total reflectance FTIR. Biochemistry 33:2628–2634.
27. Panda, A. K. 2003. Bioprocessing of therapeutic proteins from the inclusion bodies of Escherichia coli. Adv. Biochem. Eng. Biotechnol. 85:43–93.
28. Przybycien, T. M., J. P. Dunn, P. Valax, and G. Georgiou. 1994. Secondary structure characterization of beta-lactamase inclusion bodies. Protein Eng. 7:131–136.
29. Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
30. Schlieker, C., I. Tews, B. Bukau, and A. Mogk. 2004. Solubilization of aggregated proteins by ClpB/DnaK relies on the continuous extraction of unfolded polypeptides. FEBS Lett. 578:351–356.
31. Schlieker, C., J. Weibezahn, H. Patzelt, P. Tessarz, C. Strub, K. Zeth, A. Erbse, J. Schneider-Mergener, J. W. Chin, P. G. Schultz, B. Bukau, and A. Mogk. 2004. Substrate recognition by the AAA+ chaperone ClpB. Nat. Struct. Mol. Biol. 11:607–615.
32. Singh, S. M., and A. K. Panda. 2005. Solubilization and refolding of bacterial inclusion body proteins. J. Biosci. Bioeng. 99:303–310.
33. Strandberg, L., and S. O. Enfors. 1991. Factors influencing inclusion body formation in the production of a fused protein in Escherichia coli. Appl. Environ. Microbiol. 57:1669–1674.
34. Tokatlidis, K., P. Dhurjati, J. Millet, P. Beguin, and J. P. Aubert. 1991. High activity of inclusion bodies formed in Escherichia coli overproducing Clostridium thermocellum endoglucanase D. FEBS Lett. 282:205–208.
35. Vallejo, L. F., and U. Rinas. 2004. Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins. Microb. Cell Fact. 3:11.
36. Ventura, S., and A. Villaverde. 2006. Protein quality in bacterial inclusion bodies. Trends Biotechnol. 24:179–185.
37. Vera, A., N. Gonzalez-Montalban, A. Aris, and A. Villaverde. 29 September 2006, posting date. The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. Biotechnol. Bioeng. doi: 10.1002/bit.21218.
38. Villaverde, A., and M. M. Carrio. 2003. Protein aggregation in recombinant bacteria: biological role of inclusion bodies. Biotechnol. Lett. 25:1385–1395.
39. Worrall, D. M., and N. H. Goss. 1989. The formation of biologically active beta-galactosidase inclusion bodies in Escherichia coli. Aust. J. Biotechnol. 3:28–32.

IV.4 Paper 4

Divergent genetic control of protein solubility and conformational quality

in *Escherichia coli*

Elena García-Fruitós, Mónica Martínez-Alonso, Núria González-Montalbán, Minoska

Valli, Diethard Mattanovich and Antonio Villaverde

Our previous reports indicated that inclusion bodies contain not only important extents of native-like secondary structure, but also detectable levels of biological activity. These observations prove that, contrary to what has been generally believed, protein solubility is not matching conformational quality. However, it was not known how these two parameters are regulated by the cell. Therefore, we decided to investigate the role of the evolutionarily conserved quality control apparatus (composed essentially by chaperones and proteases) regulating the solubility and the conformational quality of aggregation-prone recombinant proteins. To perform this analysis we used *Escherichia coli* MC4100 wild-type strain and a set of mutants (ClpB⁻, ClpA⁻, GroES, IbpAB⁻, DnaK⁻, ClpP⁻, GroEL, Lon⁻) deficient in different cytosolic chaperones and proteases critical for the cytoplasmic quality control. Given that the specific emission of green fluorescent protein (GFP) has been successfully used as a marker of conformational quality, the model protein chosen to perform this study was VP1GFP.

We envisaged a decrease in solubility in the absence of relevant chaperones or proteases. However, we also noticed an increase in the fluorescence emission per cell, especially in DnaK⁻ cells. To further study this fact, we determined VP1GFP half-life in all the mutants, observing that proteolysis was clearly inhibited in all the cases except in IbpAB⁻ cells. Specifically, VP1GFP stability in DnaK⁻ and ClpB⁻ was really close to that found in the absence of the proteases involved in the degradation (Lon, ClpP or ClpP ATPase subunit ClpA), proving that both DnaK and ClpB are positively mediating proteolysis, preferentially that of insoluble but functional polypeptides.

Summarising, all these data make evident that the bacterial quality control system promotes protein solubility at the expense of conformational quality (resulting in biological activity), through an overcommitted proteolysis. On the other hand, we could also conclude that cell-mediated conformational control acts irrespectively of protein solubility.

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Divergent Genetic Control of Protein Solubility and Conformational Quality in *Escherichia coli*

Elena García-Fruitós[1], Mónica Martínez-Alonso[1] Nuria Gonzàlez-Montalbán[1], Minoska Valli[2,3] Diethard Mattanovich[2,3] and Antonio Villaverde[1]*

[1]*Institute for Biotechnology and Biomedicine, Department of Genetics and Microbiology and CIBER-BBN Networking Centre on Bioengineering Biomaterials and Nanomedicine Autonomous University of Barcelona, Bellaterra 08193 Barcelona, Spain*

[2]*University of Natural Resources and Applied Life Sciences Vienna, Department of Biotechnology, Institute of Applied Microbiology, Vienna Austria*

[3]*School of Bioengineering University of Applied Sciences FH-Campus Vienna, Austria*

In bacteria, protein overproduction results in the formation of inclusion bodies, sized protein aggregates showing amyloid-like properties such as seeding-driven formation, amyloid-tropic dye binding, intermolecular β-sheet architecture and cytotoxicity on mammalian cells. During protein deposition, exposed hydrophobic patches force intermolecular clustering and aggregation but these aggregation determinants coexist with properly folded stretches, exhibiting native-like secondary structure. Several reports indicate that inclusion bodies formed by different enzymes or fluorescent proteins show detectable biological activity. By using an engineered green fluorescent protein as reporter we have examined how the cell quality control distributes such active but misfolded protein species between the soluble and insoluble cell fractions and how aggregation determinants act in cells deficient in quality control functions. Most of the tested genetic deficiencies in different cytosolic chaperones and proteases (affecting DnaK, GroEL, GroES, ClpB, ClpP and Lon at different extents) resulted in much less soluble but unexpectedly more fluorescent polypeptides. The enrichment of aggregates with fluorescent species results from a dramatic inhibition of ClpP and Lon-mediated, DnaK-surveyed green fluorescent protein degradation, and it does not perturb the amyloid-like architecture of inclusion bodies. Therefore, the *Escherichia coli* quality control system promotes protein solubility instead of conformational quality through an overcommitted proteolysis of aggregation-prone polypeptides, irrespective of their global conformational status and biological properties.

© 2007 Elsevier Ltd. All rights reserved.

*Keywords:* protein folding; protein solubility; DnaK; *E. coli*; quality control

Introduction

In bacteria, overproduction of different proteins results in their deposition as amyloid-like aggregates called inclusion bodies.[1] Inclusion body formation is a sequence-specific, seeding-driven aggregation process[1,2] in which misfolded protein species, folding intermediates and presumably oligomeric species,

cross-react in amorphous protein deposits rich in a pleated β-sheet architecture.[3] Such aggregates are stained with amyloid-tropic dies, such as thioflavin-T and Congo red,[1] and when exposed to mammalian cells show a toxicity profile indistinguishable from that caused by mammalian amyloids.[4]

However, bacterial inclusion bodies also contain important extents of native-like secondary structure.[5–7] Therefore, when they are formed by enzymes or fluorescent proteins they show catalytic properties or are highly fluorescent.[8] Therefore, in inclusion body-forming proteins aggregation determinants and properly folded regions coexist in single polypeptides. This is obvious when observing the core of green fluorescent protein (GFP) inclusion bodies rich in fluorescent species.[9] The occurrence of

---

*Corresponding author. E-mail address: avillaverde@servet.uab.es.

Abbreviations used: GFP, green fluorescent protein; mGFP, misfolding-prone GFP fusion protein; FTIR, Fourier transform infrared.
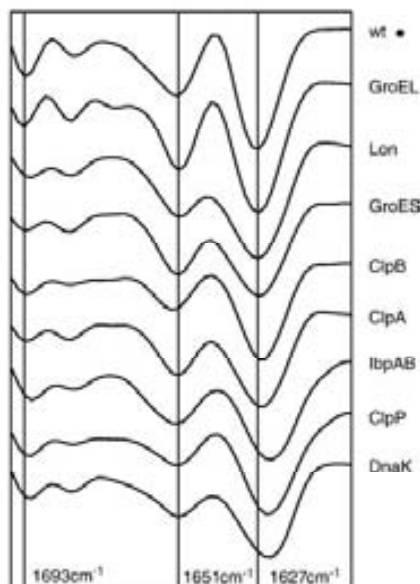
Figure 1. Second derivative spectra in the amide I region of inclusion bodies formed in either the parental strain or the derivative mutants. Vertical lines correspond to the observed extended β-sheet (1627 cm^{-1}), α-helix and/or unfolded structure (1651 cm^{-1}) and antiparallel and/or loop structure peaks (1693 cm^{-1}). Labels at the right indicate the protein whose function is totally or partially lost, and the wild-type strain is highlighted by a dot for reference, here and in the other Figures.

functional protein in inclusion bodies[3] indicates that, contrary to what has been generally believed, protein solubility and conformational quality are not coincident properties.[10] In fact, the soluble fraction of inclusion body-producing cells contains "soluble aggregates", observed, among others, in glutathion transferase,[11] β-galactosidase[12] and maltose-bind-

ing protein.[13,14] Such oligomeric protein species might be eventually less functional than the equivalent "true" soluble species. The variability in the specific activities observed in different soluble enzymes when produced under different environmental conditions strongly supports this presumption.[7,15–18] How the amyloidal deposition of conformationally hybrid, misfolded but functional proteins is regulated by the cell is not known.

In *Escherichia coli*, an evolutionarily conserved quality control apparatus (essentially composed of chaperones and proteases) is believed to survey the conformational status of the cell protein pool.[19–21] The bacterial protein DnaK, which is analogous to the eukaryotic Hsp70, regulates the whole set of quality control genes by modulating the levels of active σ 32 heat-shock transcription factor.[22–24] Being itself a chaperone, DnaK cooperates with DnaJ and GrpE for protein holding and folding,[25–30] and in association with ClpB and the small heat shock proteins IbpA and IbpB, removes individual polypeptides from aggregates for eventual refolding and activation.[31,32,63] Conventionally, solubility and conformational quality are considered to ne unequivocally connected, enhancing the cellular levels of DnaK and other chaperones through plasmid-encoded expressible genes that should result in higher solubility and biological activity.[33,34] However, such experimental approaches, generally improving solubility,[35] have often yielded inconsistent results.[36,37]

In the light of recent observations indicating that conformational quality and solubility are distinguishable protein properties,[10] it is not clear how these two parameters are genetically regulated by the cell and how misfolded but active polypeptides are selected from the soluble cell fraction for deposition. Here, we present evidence that the bacterial quality control system acts coordinately to promote solubility at the expense of conformational

Table 1. Fluorescence emission per cell biomass observed in mGFP-producing cells

| Phenotype | Fluorescence per cell biomass | | | | | | Function of the missing protein | Relevant references |
| | Conventional fluorometry | | | | Flow cytometry | | | |
	Units/OD	%^{a}	Units/ng mGFP	%^{a}	Mean FL1 units/cell	%^{a}		
wt (MC4100)^b	405.6±13.1	100	1027.4±89.2	100	176.5±10.1	100		
DnaK⁻	547.3±77.8	134.9±19.1	1506.9±89.2	146.6±8.6	443.2±15.8	250.9±8.9	Heat-shock system regulator, foldase, disagregase	22,27–29, 31,57,63,73
GroEL140^c	342.8±35.3	84.529±8.7	1062.9±119.8	103.4±11.6	237.5±2.6	134.5±1.4	Foldase subunit	27,28,30
ClpB⁻	515.6±24.3	127.10±6.0	1628.6±78.7	158.5±7.6	290.2±7.8	164.3±4.4	Disagregase	31,57,63
ClpA⁻	543.0±45.4	133.87±11.2	1756.5±401.1	170.9±39.0	284.2±9.4	160.9±5.3	ATPase, foldase	43,74–77
GroES⁻	440.2±5.2	108.52±1.3	1408.4±62.7	137.0±8.3	218.6±11.7	123.7±5.4	Foldase subunit	27,28,30
IbpAB⁻	303.2±16.1	74.74±3.9	1036.2±123.5	100.8±12.0	150.1±2.4	85.0±1.4	Disaggregase	31,57,63,67,78
ClpP⁻	522.3±31.1	128.77±7.6	1821.5±144.8	177.2±14.0	377.4±2.5	213.7±1.1	Protease	43,74–77,79
Lon⁻	686.4±40.9	169.22±10.0	2168.0±463.9	211.0±45.1	295.3±15.4	168.4±5.1	Protease	80–84

The columns display, from left to right, fluorescence per total biomass (OD), fluorescence per specific GFP protein, and fluorescence per single cell.
^{a} Fluorescence values in % are referred to those observed in MC4100.
^{b} MC4100 is used as wild-type cells regarding the protein quality control system.
^{c} Protein GroEL140 is partially active at 37 °C, allowing cell growth at this temperature (although not at 43 °C).
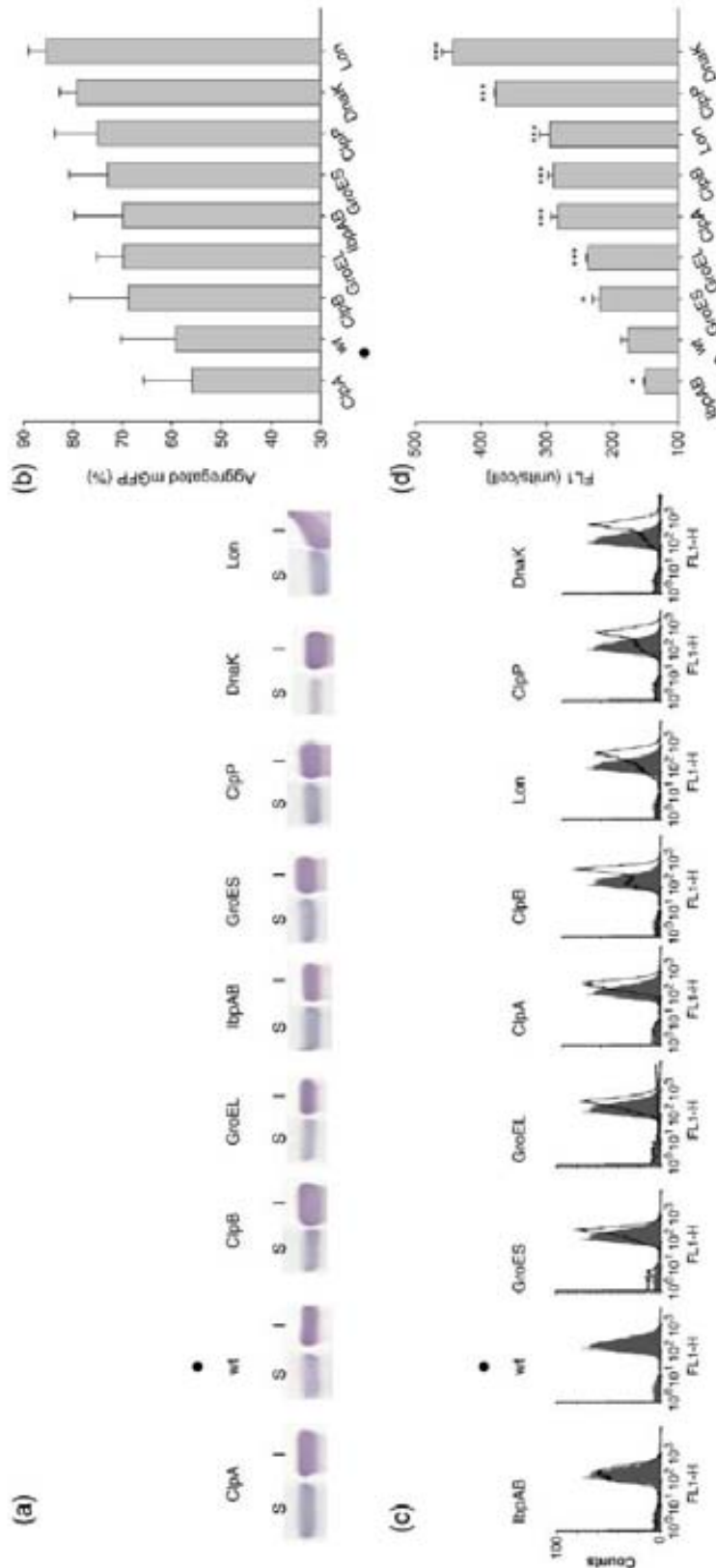
Figure 2.. (a) Immunodetection of mGFP in the soluble (S) and insoluble (I) cell fractions; (b) extent of aggregation (percentage of aggregated mGFP with respect to total mGFP) in different mutant strains. (c) Flow cytometry analysis of mGFP-producing cells (black lines), showing the wild-type strain for reference (filled symbols); (d) fluorescence emission/cell in all these strains. Values significantly different from those of wild-type cells are indicated by the p parameter obtained in an ANOVA test as follows: ***$p<0.01$ <**$p<0.05$ <*$p<0.1$. Data are ordered numerically following an increasing pattern, here and in the other Figures.

quality, and that high levels of solubility are achieved by a powerful proteolysis of target proteins in the bacterial cytosol under a complex regulation. At least Lon and ClpP proteases, directed by DnaK, degrade aggregation-prone but functional protein species by an over-committed activity that, while minimizing aggregation, dramatically reduces the cellular amounts of functional versions of misfolded protein species.

Results

Fluorescence emission of GFP requires the presence of the chromophore attached to the inner α-helix,[38] and the proper folding of the β-barrel and the complete maturation of the whole polypeptide through late folding steps that eliminate the residual water and the resulting quenching.[38–40] Therefore, the specific emission of GFP has been used successfully as a marker of the conformational quality of GFP-containing misfolding-prone proteins.[8,20,41,42] To explore the eventual divergent control of solubility and conformational quality we have analysed both parameters in a reporter, misfolding-prone mGFP fusion protein when produced in a set of mutants devoid of functional chaperones and proteases critical for the cytoplasmic quality control. In all these strains, as well as in the parental wild-type MC4100, the reporter mGFP formed cytoplasmic inclusion bodies with a consistent pattern of secondary structure characteristic of amyloid deposits (Figure 1). Considering the functions disrupted in these strains (Table 1), we envisaged a dramatic decrease of both mGFP solubility and fluorescence.

In this context, the proportion of insoluble mGFP increased in mutant cells showing that, as expected, total or partial inactivation of different arms of the quality control results in more protein aggregation (Figure 2(a) and (b)). A progressive downshift in the β-sheet peaks in all the mutants indicated mild variations in the molecular organization of the aggregates, tending to a more compact cross β-sheet architecture when comparing with the wild-type (Figure 1). However the fluorescence emission per cell was also higher in the absence of relevant chaperones (except IbpAB) or proteases than in wild-type cells (Figure 2(c) and (d)). This unexpected observation was confirmed by conventional fluorimetry (Table 1). It was especially surprising that cells lacking a functional DnaK, which strongly enhanced aggregation, were also highly fluorescent (Figure 2(c) and (d)). In such a DnaK⁻ background, fluorescence returned to wild-type levels when producing supplementary amounts of DnaK from a DnaK-encoding plasmid (Figure 3).

When checking the soluble cell fraction and purified inclusion bodies separately, we observed that both soluble and aggregated mGFP were more fluorescent in most of the tested mutants than in the wild-type cells (Figure 4). The higher level of specific fluorescence occurred in genetic backgrounds in which the aggregation of the reporter protein was
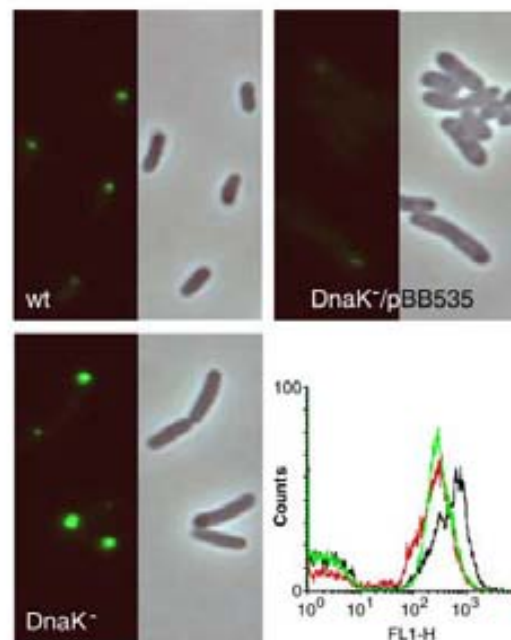
Figure 3. Fluorescence microscopy images of wild-type, DnaK⁻ and DnaK⁻ cells overproducing DnaK (encoded in pBB535). Flow cytometry of these strains is comparatively shown in red, black and green, respectively.

more intense (compare with Figure 2). To check if the enhanced emission of the mGFP produced in these cases could be due to less saturation of the folding machinery and, consequently, to more opportunities for proper folding, we determined the intracellular concentration of mGFP. However, much more mGFP was observed in most of the mutant strains (Figure 5(a)).

We hypothesized that protein stability could be the major contributor to the different GFP pools observed in the mutant set. While a lower degradation rate of misfolding-prone species would be expected in protease-deficient cells (in either Lon⁻ or ClpP⁻ cells), such a phenotypic trait would not be evident in the other genetic backgrounds tested. Therefore, degradation of mGFP was monitored in the absence of protein synthesis and its half-life was determined in several independent experiments. Intriguingly, proteolysis of mGFP was clearly inhibited in all the tested mutants except in IbpAB⁻ cells (Figure 5(b)). Furthermore, the stability of mGFP in the absence of DnaK or ClpB was close to that observed in the absence of Lon, ClpP, or the ClpP ATPase subunit ClpA,[43] proving that both DnaK and ClpB are positive regulators of mGFP degradation. Again, and in agreement with complementation data shown in Figure 3, the transformation of DnaK⁻ cells with an expressible, DnaK-encoding plasmid accelerated the degradation of mGFP to wild-type rates. Therefore, chaperone-modulated digestion of misfolding-prone but functional polypeptides appears as the
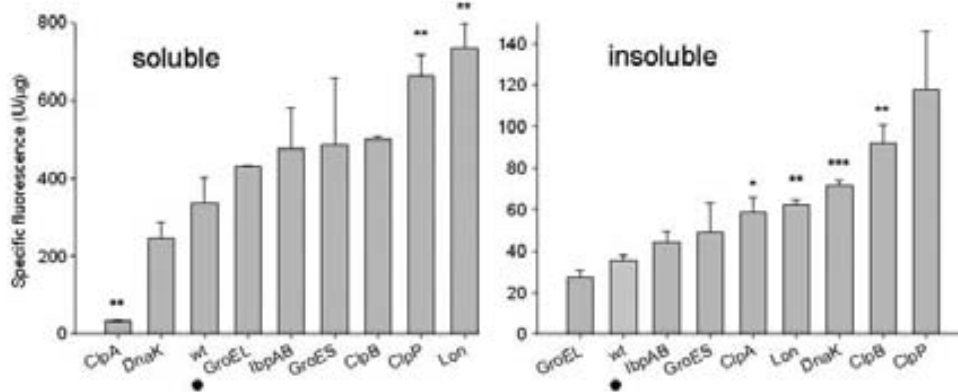
Figure 4. Specific fluorescence of mGFP in the soluble and insoluble cell fractions.

main determinant of the relatively small amounts of fluorescent mGFP in wild-type cells.

Discussion

The functional meaning of protein aggregation is still controversial. Intriguingly, many examples of bacterial inclusion bodies formed by highly functional enzymatic forms have been reported, including those formed by β-galactosidase,[44] endoglucanase-D,[45] β-lactamase,[46] rHtr1 aliase[46] and dihydrofolate reductase,[8] and by fusion proteins containing β-galactosidase,[8] β-lactamase and alkaline phosphatase. In a similar way, GFP,[20,47] GFP fusion versions[8,20] and BFP[8,42] form highly fluorescent inclusion bodies. This indicates that solubility is not matching conformational quality, and the connection between the occurrence of aggregation determinants and the average folding status of the protein is still far from being solved. Numerous pieces of evidence point to defined protein patches as drivers of the aggregation process,[48] which can occur without necessarily

compromising the biological activity of the protein deposited.[8,49] The progressive characterization of individual genes expressed during the *E. coli* heat-shock response, and the identification of the functions displayed by the encoded proteins, pictured the existence in *E. coli* and in higher organisms, of a complex protein quality system that mediates the elimination of folding-reluctant, potentially cytotoxic species. The quality control is executed through the combined action of chaperones and proteases and is especially relevant under conformational stresses such as those caused by high temperatures or high rates of protein synthesis. However, as discussed below, the physiological mechanics of the quality control, at the system level, might have been largely misunderstood.

Recombinant bacteria are one of the best models to investigate protein quality control, since many of the foreign proteins over-produced therein tend to form a special type of amyloid aggregates known as inclusion bodies.[1,50] It has been generally believed that aggregation is promoted by a high substrate load of the quality control system. Therefore, the
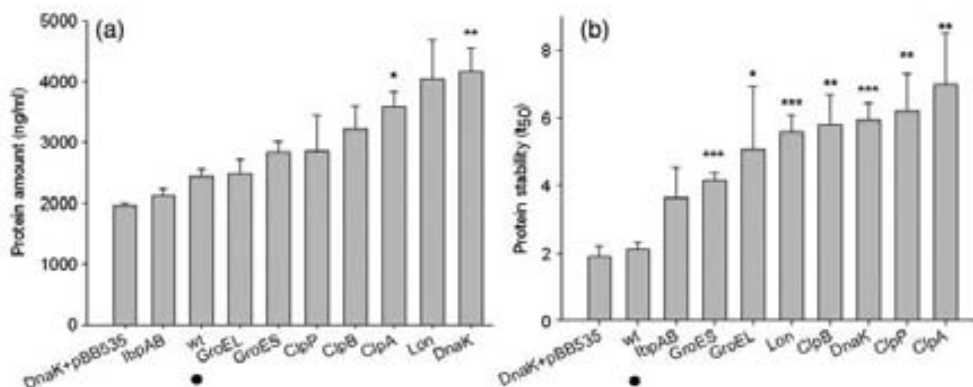
Figure 5. (a) Total intracellular amounts and (b) proteolytic stability of mGFP in the different producing strains. Values significantly different from those of wild-type cells are indicated by the *p* parameter obtained in an ANOVA test as follows: ***$p < 0.01 <$ **$p < 0.05 <$ *$p < 0.1$. Labels indicate the protein whose function is lost, and the wild-type strain is highlighted by an asterisk. The presence of pBB535 is indicated. Total mGFP amounts and half-life correlated within the limits of statistical significance ($p = 0.052$).

insufficient availability of folding modulators and proteases would fail in preventing the accumulation of misfolded species. Therefore, the additional supply of chaperones from encoding plasmids has been explored repeatedly as a way to gain solubility in production processes,[33,37] and solubility has been taken as universal indicator of conformational quality. In this context, functional deficiencies in the main chaperones such as DnaK or GroEL were expected to result in lower conformational quality (therefore less functional recombinant protein) than when the quality control system is completely active.

When analyzing the fluorescence of different mutant strains (deficient in GroEL, ClpA, ClpB, ClpP, Lon and DnaK activities) producing a misfolding-prone version of GFP, we surprisingly noted more fluorescence emission per cell than in the wild-type (Figure 2(d)). This was due to a generally higher specific emission of both soluble and insoluble mGFP populations (Figure 4). However, solubility in these genetic backgrounds was not concomitantly high. On the contrary, the aggregation of the reporter GFP was enhanced (Figure 2(b)) and the Fourier transform infrared (FTIR) analysis of the inclusion bodies showed a band down-shift from 1627 cm^{-1} to around 1623 cm^{-1} (Figure 1), indicative of a more condensed, amyloid-like cross molecular β-sheet architecture. In particular, when observing aggregation in the absence of DnaK, the high level of fluorescence exhibited by the aggregates returned to wild-type levels when complemented by a *dnaK*-carrying expression vector (Figure 3). In summary, the disruption of different arms of the *E. coli* cytosolic quality control resulted, as expected, in less solubility but, contrary to what is generally presumed, in much more functional proteins in both soluble and insoluble populations (Figure 4). Note that, for instance, the big aggregates formed in the absence of DnaK were extremely fluorescent (Figure 3).

Since the cellular amounts of mGFP were also enhanced in most of the strains tested (Figure 5(a)), and this fact correlated with proteolytic stability (Figure 5(b)), it was reasonable to infer that in cells with a fully functional quality control system, proteolysis acts largely on aggregation-prone but functional (or suitable to be activated) polypeptides. Such over-committed activity must remove insoluble protein species thus increasing solubility in relative terms, but resulting in less cell fluorescence. This is illustrated by the negative role of DnaK in protein stability. In the absence of this chaperone, GFP half-life increases dramatically to levels comparable with those observed in the absence of Lon (Figure 5(b)), one of the main cytosolic proteases degrading recombinant proteins.[51] However, the expression of plasmid-harboured copies of *dnaK* reduced fluorescence (Figure 3), minimized accumulation of mGFP by lowering its half-life to wild-type levels (Figure 5(a) and (b)), and reduced aggregation to around 30 % (not shown).

While Lon and ClpP (and its ATPase subunit ClpA) are known to be involved in the degradation of most recombinant proteins in *E. coli*,[52–54] the roles

of DnaK and ClpB (and other chaperones identified here) in positively mediating proteolysis are less apparent. In this respect, it could be considered that greater stability could be an indirect consequence of the enhanced aggregation observed in the absence of these chaperones. Although it is generally believed that aggregation as inclusion bodies prevents proteolysis,[36] *in situ* or release-associated digestion of inclusion body-forming proteins has been described.[51,55,56] In fact, the absence of main cell proteases results in oversized aggregates,[51] indicating that, on the contrary, inclusion body formation is regulated strongly by proteolysis. In addition, the absence of a negative correlation between solubility and the half-life of mGFP ($p = 0.5907$, data not shown) is further evidence that aggregation as inclusion bodies is not the cause of proteolytic resistance but *vice versa*.

Both DnaK and ClpB act as a disaggregase complex[31,57,58] that remove aggregated polypeptides for eventual refolding. In inclusion body-forming cells, DnaK accumulates on the inclusion body surfaces,[59] where disaggregation is expected to occur. When either DnaK or ClpB are absent, the spontaneous disintegration of inclusion bodies is largely delayed,[60] while the co-production of both chaperones favors aggregate dissolution.[61] On the other hand, physiological removal of inclusion body proteins is temporarily associated with protein cleavage, and the mathematical modelling of inclusion body dissolution repeatedly indicated that protein digestion is intimately associated with protein removal.[55,62] How DnaK and ClpB could mechanistically promote Lon- or ClpP-mediated digestion during protein extraction from aggregates is not clear and requires further investigation. Interestingly, the small heat shock proteins IbpA and IbpB, which act in cooperation with DnaK and ClpB in protein refolding,[57,63] do not seem to be positively involved in proteolysis, as their absence does not result in stabilization of mGFP (Figure 5). On the contrary, these small chaperones display protease-protective roles on misfolded proteins,[64] which could antagonistically regulate *the in situ* proteolysis promoted by DnaK-ClpB.

Intriguingly, the divergence between solubility and conformational quality recently proposed in the context of recombinant protein production is clearly illustrated by the negative relationship between the relative occurrence of soluble species and their specific emission (Figure 6(a)).[10] This fact supports the suggestion that insoluble but functional polypeptides are preferred targets for DnaK-controlled proteolysis, and indicated that biological activity is not compromised by the occurrence of aggregation determinants. In this respect, it has been shown recently that factors favouring protein folding, such as low temperature,[20] or appropriate levels of DnaK,[65] activate soluble and inclusion body protein versions simultaneously. We determine here that the specific fluorescence of soluble and insoluble GFP correlate well when these parameters are compared pair-wise in different strains, although the level of
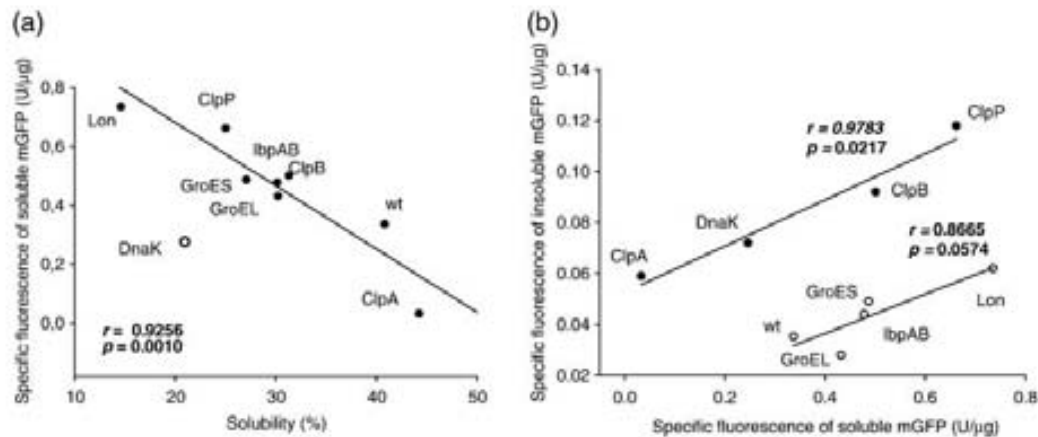
Figure 6. (a) Pair-wise comparison of solubility and the specific emission of soluble mGFP in different producing strains. The data from DnaK⁻ cells has been excluded from the analysis but the coordinate is indicated by a white circle. When including this value, the correlation is still significant ($r = 0.698$; $p = 0.036$). (b) Pair-wise comparison of the specific fluorescence of soluble and insoluble mGFP in all the different producing strains. Correlation parameters are shown in each panel. The slopes are 0.0906 and 0.0767 for the upper and lower lines, respectively.

fluorescence of aggregated protein is higher in the set lacking DnaK, ClpA, ClpB or ClpP than in others (Figure 6(b)). This fact indicates that these four proteins participate in the discrimination of functional species between soluble and insoluble fractions. It also supports the hypothesis that inclusion bodies are not excluded from conformational control, which acts irrespective of protein solubility. Interestingly, the slopes of both correlation lines are numerically very similar (Figure 6(b)), indicating again a high level of regularity in the genetic control of the conformational quality of soluble and aggregated species. The similarity between the specific activity of soluble and inclusion body versions of a recombinant β-galactosidase (structurally and functionally different from GFP) in the absence of DnaK,[7,15] indicates that the genetics of quality control described here is not limited to the model protein used in the present study.

Therefore, solubility, resulting from the combination of the protein folding process and an over-committed, highly selective proteolysis should be reconsidered as a universal indicator of conformational quality.

Materials and Methods

Strains and plasmids

The *Escherichia coli* strains used in this work were MC4100 (*araD139 Δ(argF-lac) U169 rpsL150 relA1 flbB5301 deoC1 ptsF25 rbsR*),[66] used as a wild-type, and its derivatives JGT3 (*ΔclpB::kan*), JGT4 (*clpA::kan*), JGT6 (*zjd::Tn10 groES30*), JGT17 (*Δibp::kan*), JGT19 (*clpP::cat*), JGT20 (*dnak756 thr::Tn10*),[67] BB4564 (*groEL140 zjd::Tn10 zje::ΩSpcʳ/Strʳ*),[20,30,68] and BB2395 (*Δlon146::miniTn10*).[53] These strains were transformed with pTVP1GFP,[8] encoding a misfolding-prone fusion version of GFP (mGFP), in which the VP1 capsid protein of foot-and-mouth disease virus is joined at the amino terminus and produced under

the control of the *trc* promoter. The chaperone-encoding plasmid pBB535 (Specʳ), has been described.[53] Briefly, it drives the expression of the *dnaK* and *dnaJ* genes under the control of the IPTG-inducible P$_{A1/lac-O1}$ promoter and it is fully compatible with pTVP1GFP.

Culture conditions

All the production processes were performed in shake-flask cultures growing at 37 °C and 250 rpm in LB rich medium,[66] plus 100 μg/ml of ampicillin and eventually 100 μg/ml of spectinomycin for plasmid maintenance. Expression of VP1GFP gene was induced when the absorbance at 550 nm reached 0.4, by adding IPTG to 1 mM. Cell samples were taken at 2 h after induction of gene expression for further analysis. For the *in vivo* determination of proteolytic stability, protein synthesis was arrested by adding chloramphenicol to 200 μg/ml, and the cultures were further incubated at 37 °C and sampled at different times. Data were obtained from three or more independent experiments.

Flow cytometry

Flow cytometric analyses were performed after sample fixation with 0.1 % (v/v) formaldehyde overnight at 4 °C, on a FACSCalibur system (Becton Dickinson) with 488 nm excitation from a 15 mW air-cooled argon-ion laser. Fluorescence emission was acquired through a 530/30 nm bandpass filter (FL1) in logarithmic mode. For every sample, the fluorescence emission in channel FL1 was analyzed with a FACSCalibur and converted in a spreadsheet by the WinMDI 2.8 software. During the elaboration of the data, we calculated the geometric mean for every sample.

Determination of total and specific fluorescence

Culture samples of 1 ml were jacketed in ice, disrupted by sonication for 4 min at 50 W under 0.5 s cycles as described,[69] and centrifuged at 4 °C for 15 min at 15,000g. The supernatant was used directly for the analysis as the soluble cell fraction. Inclusion bodies were purified by a

detergent-washing protocol as described,[70] and used in suspension for activity analysis. Fluorescence at 510 nm was recorded with a Cary Eclipse fluorescence spectrophotometer (Variant) by using 450 nm as the excitation wavelength. Fluorescence was measured, in triplicate, in 1 ml samples using dilutions when necessary, and corrected by either biomass or amounts of mGFP protein determined as described below.

Quantitative protein analysis

Samples of bacterial cultures (10 ml) were centrifuged at low speed (15 min at 15,000g) to harvest the cells and resuspended in 2 ml of PBS with one tablet of Protease Inhibitor Cocktail (Roche, ref. 1 836 170) per 10 ml of buffer. These samples, once jacketed in ice, were sonicated for 5 min (or longer when required to achieve a complete disruption) at 50 W under 0.5 s cycles as described,[69] and centrifuged for 15 min at 15,000g. The supernatant was mixed with denaturing buffer at the appropriate ratios.[71] For the determination of inclusion body protein, these aggregates were purified by repeated washing with detergent as described,[70] and resuspended in denaturing buffer.[71] For the analysis of proteolytic stability, protein synthesis was arrested by adding chloramphenicol at 200 μg/ml 2 h after of induction of gene expression, and the cultures were further incubated at 37 °C. Samples taken at different times were then centrifuged for 15 min at 15,000g and pellets were resuspended in denaturing buffer.[71] After boiling for 15 min (for the soluble fraction) or 25 min (for inclusion bodies), appropriate sample volumes were loaded onto denaturing gels. For Western blot, a polyclonal serum directed against GFP was used. Dried blots were scanned at high resolution and bands were quantified by using the Quantity One software from Bio Rad, with the appropriate protein dilutions of known concentration as controls. Slight bands compatible partially with proteolysed mGFP and with full-length GFP were observed in some cases, representing always <4% of the total immunoreactive material and were included in all the analysis. Determinations were done within a linear range and they were used to calculate the specific activity values. We performed a nested ANOVA to analyse the different samples.

FTIR spectroscopy

Purified inclusion bodies were dried for 2 h in a Speed-Vac system before analysis to reduce water interference in the infrared spectra. The FTIR spectrum of the dry samples was analysed directly into a Bruker Tensor FTIR spectrometer. All processing procedures were carried out so as to optimize the quality of the spectrum in the amide I region, between 1600 cm^{-1} and 1700 cm^{-1}. Second derivatives of the amide I band spectra were used to determine the frequencies at which the different spectral components were located. A general description of FTIR procedures can be found elsewhere.[72]

Acknowledgements

References

1. Carrio, M., Gonzalez-Montalban, N., Vera, A., Villaverde, A. & Ventura, S. (2005). Amyloid-like properties of bacterial inclusion bodies. *J. Mol. Biol.* **347**, 1025–1037.
2. Speed, M. A., Wang, D. I. & King, J. (1996). Specific aggregation of partially folded polypeptide chains: the molecular basis of inclusion body composition. *Nature Biotechnol.* **14**, 1283–1287.
3. Ventura, S. & Villaverde, A. (2006). Protein quality in bacterial inclusion bodies. *Trends Biotechnol.* **24**, 179–185.
4. Gonzalez-Montalban, N., Villaverde, A. & Aris, A. (2007). Amyloid-linked cellular toxicity triggered by bacterial inclusion bodies. *Biochem. Biophys. Res. Commun.* **355**, 637–642.
5. Oberg, K., Chrunyk, B. A., Wetzel, R. & Fink, A. L. (1994). Nativelike secondary structure in interleukin-1 beta inclusion bodies by attenuated total reflectance FTIR. *Biochemistry*, **33**, 2628–2634.
6. Ami, D., Natalello, A., Taylor, G., Tonon, G. & Maria, D. S. (2006). Structural analysis of protein inclusion bodies by Fourier transform infrared microspectroscopy. *Biochim. Biophys. Acta*, **1764**, 793–799.
7. Gonzalez-Montalban, N., Garcia-Fruitos, E., Ventura, S., Aris, A. & Villaverde, A. (2006). The chaperone DnaK controls the fractioning of functional protein between soluble and insoluble cell fractions in inclusion body-forming cells. *Microb. Cell Fact.* **5**, 26.
8. Garcia-Fruitos, E., Gonzalez-Montalban, N., Morell, M., Vera, A., Ferraz, R. M., Aris, A. et al. (2005). Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. *Microb. Cell Fact.* **4**, 27.
9. Garcia-Fruitos, E., Aris, A. & Villaverde, A. (2007). Localization of functional polypeptides in bacterial inclusion bodies. *Appl. Environ. Microbiol.* **73**, 289–294.
10. Gonzalez-Montalban, N., Garcia-Fruitos, E. & Villaverde, A. (2007). Recombinant protein solubility-does more mean better? *Nature Biotechnol.* **25**, 718–720.
11. Stegemann, J., Ventzki, R., Schrodel, A. & de Marco, A. (2005). Comparative analysis of protein aggregates by blue native electrophoresis and subsequent sodium dodecyl sulfate-polyacrylamide gel electrophoresis in a three-dimensional geometry gel. *Proteomics*, **5**, 2002–2009.
12. Aris, A. & Villaverde, A. (2000). Molecular organization of protein-DNA complexes for cell-targeted DNA delivery. *Biochem. Biophys. Res. Commun.* **278**, 455–461.
13. Sachdev, D. & Chirgwin, J. M. (1999). Properties of soluble fusions between mammalian aspartic proteinases and bacterial maltose-binding protein. *J. Protein Chem.* **18**, 127–136.
14. Nomine, Y., Ristriani, T., Laurent, C., Lefevre, J. F., Weiss, E. & Trave, G. (2001). Formation of soluble inclusion bodies by hpv e6 oncoprotein fused to

maltose-binding protein. *Protein Expr. Purif.* **23**, 22–32.

15. Garcia-Fruitos, E., Carrio, M. M., Aris, A. & Villaverde, A. (2005). Folding of a misfolding-prone beta-galactosidase in absence of DnaK. *Biotechnol. Bioeng.* **90**, 869–875.

16. Vasina, J. A. & Baneyx, F. (1996). Recombinant protein expression at low temperatures under the transcriptional control of the major *Escherichia coli* cold shock promoter cspA. *Appl. Environ. Microbiol.* **62**, 1444–1447.

17. Leandro, P., Lechner, M. C., Tavares dA, I. & Konecki, D. (2001). Glycerol increases the yield and activity of human phenylalanine hydroxylase mutant enzymes produced in a prokaryotic expression system. *Mol. Genet. Metab.* **73**, 173–178.

18. Ferrer, M., Chernikova, T. N., Timmis, K. N. & Golyshin, P. N. (2004). Expression of a temperature-sensitive esterase in a novel chaperone-based *Escherichia coli* strain. *Appl. Environ. Microbiol.* **70**, 4499–4504.

19. Bukau, B., Weissman, J. & Horwich, A. (2006). Molecular chaperones and protein quality control. *Cell*, **125**, 443–451.

20. Vera, A., Gonzalez-Montalban, N., Aris, A. & Villaverde, A. (2007). The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. *Biotechnol. Bioeng.* **96**, 1101–1106.

21. Miot, M. & Betton, J. M. (2004). Protein quality control in the bacterial periplasm. *Microb. Cell Fact.* **3**, 4.

22. Liberek, K. & Georgopoulos, C. (1993). Autoregulation of the *Escherichia coli* heat shock response by the DnaK and DnaJ heat shock proteins. *Proc. Natl Acad. Sci. USA*, **90**, 11019–11023.

23. Tomoyasu, T., Ogura, T., Tatsuta, T. & Bukau, B. (1998). Levels of DnaK and DnaJ provide tight control of heat shock gene expression and protein repair in *Escherichia coli*. *Mol. Microbiol.* **30**, 567–581.

24. Gamer, J., Multhaup, G., Tomoyasu, T., McCarty, J. S., Rudiger, S., Schonfeld, H. J. *et al.* (1996). A cycle of binding and release of the DnaK, DnaJ and GrpE chaperones regulates activity of the *Escherichia coli* heat shock transcription factor sigma32. *EMBO J.* **15**, 607–617.

25. Kedzierska, S., Staniszewska, M., Wegrzyn, A. & Taylor, A. (1999). The role of DnaK/DnaJ and GroEL/GroES systems in the removal of endogenous proteins aggregated by heat-shock from *Escherichia coli* cells. *FEBS Letters*, **446**, 331–337.

26. Jubete, Y., Maurizi, M. R. & Gottesman, S. (1996). Role of the heat shock protein DnaJ in the lon-dependent degradation of naturally unstable proteins. *J. Biol. Chem.* **271**, 30798–30803.

27. Gragerov, A., Nudler, E., Komissarova, N., Gaitanaris, G. A., Gottesman, M. E. & Nikiforov, V. (1992). Cooperation of GroEL/GroES and DnaK/DnaJ heat shock proteins in preventing protein misfolding in *Escherichia coli*. *Proc. Natl Acad. Sci. USA*, **89**, 10341–10344.

28. Langer, T., Lu, C., Echols, H., Flanagan, J., Hayer, M. K. & Hartl, F. U. (1992). Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding. *Nature*, **356**, 683–689.

29. Thomas, J. G. & Baneyx, F. (1996). Protein folding in the cytoplasm of *Escherichia coli*: requirements for the DnaK-DnaJ-GrpE and GroEL-GroES molecular chaperone machines. *Mol. Microbiol.* **21**, 1185–1196.

30. Ziemienowicz, A., Skowyra, D., Zeilstra-Ryalls, J., Fayet, O., Georgopoulos, C. & Zylicz, M. (1993). Both the *Escherichia coli* chaperone systems, GroEL/GroES and DnaK/DnaJ/GrpE, can reactivate heat-treated RNA polymerase. Different mechanisms for the same activity. *J. Biol. Chem.* **268**, 25425–25431.

31. Schlieker, C., Tews, I., Bukau, B. & Mogk, A. (2004). Solubilization of aggregated proteins by ClpB/DnaK relies on the continuous extraction of unfolded polypeptides. *FEBS Letters*, **578**, 351–356.

32. Thomas, J. G. & Baneyx, F. (2000). ClpB and HtpG facilitate de novo protein folding in stressed *Escherichia coli* cells. *Mol. Microbiol.* **36**, 1360–1370.

33. Sorensen, H. P. & Mortensen, K. K. (2005). Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microb. Cell Fact.* **4**, 1.

34. Sorensen, H. P. & Mortensen, K. K. (2005). Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *J. Biotechnol.* **115**, 113–128.

35. de Marco, A., Deuerling, E., Mogk, A., Tomoyasu, T. & Bukau, B. (2007). Chaperone-based procedure to increase yields of soluble recombinant proteins produced in *E. coli*. *BMC. Biotechnol.* **7**, 32.

36. Baneyx, F. & Mujacic, M. (2004). Recombinant protein folding and misfolding in *Escherichia coli*. *Nature Biotechnol.* **22**, 1399–1408.

37. Baneyx, F. & Palumbo, J. L. (2003). Improving heterologous protein folding via molecular chaperone and foldase co-expression. *Methods Mol. Biol.* **205**, 171–197.

38. Herberhold, H., Marchal, S., Lange, R., Scheyhing, C. H., Vogel, R. F. & Winter, R. (2003). Characterization of the pressure-induced intermediate and unfolded state of red-shifted green fluorescent protein: a static and kinetic FTIR, UV/VIS and fluorescence spectroscopy study. *J. Mol Biol.* **330**, 1153–1164.

39. Zhang, L., Patel, H. N., Lappe, J. W. & Wachter, R. M. (2006). Reaction progress of chromophore biogenesis in green fluorescent protein. *J. Am. Chem. Soc.* **128**, 4766–4772.

40. Scheyhing, C. H., Meersman, F., Ehrmann, M. A., Heremans, K. & Vogel, R. F. (2002). Temperature-pressure stability of green fluorescent protein: a Fourier transform infrared spectroscopy study. *Biopolymers*, **65**, 244–253.

41. de Groot, N. S. & Ventura, S. (2006). Effect of temperature on protein quality in bacterial inclusion bodies. *FEBS Letters*, **580**, 6471–6476.

42. de Groot, N. S. & Ventura, S. (2006). Protein activity in bacterial inclusion bodies correlates with predicted aggregation rates. *J. Biotechnol.* **125**, 110–113.

43. Hoskins, J. R., Pak, M., Maurizi, M. R. & Wickner, S. (1998). The role of the ClpA chaperone in proteolysis by ClpAP. *Proc. Natl Acad. Sci. USA*, **95**, 12135–12140.

44. Worrall, D. M. & Goss, N. H. (1989). The formation of biologically active beta-galactosidase inclusion bodies in *Escherichia coli*. *Aust. J Biotechnol.* **3**, 28–32.

45. Tokatlidis, K., Dhurjati, P., Millet, J., Beguin, P. & Aubert, J. P. (1991). High activity of inclusion bodies formed in *Escherichia coli* overproducing *Clostridium thermocellum* endoglucanase D. *FEBS Letters*, **282**, 205–208.

46. Kuczynska-Wisnik, D., Zurawa-Janicka, D., Narkiewicz, J., Kwiatkowska, J., Lipinska, B. & Laskowska, E. (2004). *Escherichia coli* small heat shock proteins IbpA/B enhance activity of enzymes sequestered in inclusion bodies. *Acta Biochim. Pol.* **51**, 925–931.

47. Tsumoto, K., Umetsu, M., Kumagai, I., Ejima, D. & Arakawa, T. (2003). Solubilization of active green fluorescent protein from insoluble particles by

guanidine and arginine. *Biochem. Biophys. Res. Commun.* **312**, 1383–1386.

48. Ventura, S. (2005). Sequence determinants of protein aggregation: tools to increase protein solubility. *Microb. Cell Fact.* **4**, 11.

49. Arie, J. P., Miot, M., Sassoon, N. & Betton, J. M. (2006). Formation of active inclusion bodies in the periplasm of *Escherichia coli*. *Mol. Microbiol.* **62**, 427–437.

50. Carrio, M. M. & Villaverde, A. (2002). Construction and deconstruction of bacterial inclusion bodies. *J. Biotechnol.* **96**, 3–12.

51. Vera, A., Aris, A., Carrio, M., Gonzalez-Montalban, N. & Villaverde, A. (2005). Lon and ClpP proteases participate in the physiological disintegration of bacterial inclusion bodies. *J. Biotechnol.* **119**, 163–171.

52. Maurizi, M. R. (1992). Proteases and protein degradation in *Escherichia coli*. *Experientia*, **48**, 178–201.

53. Tomoyasu, T., Mogk, A., Langen, H., Goloubinoff, P. & Bukau, B. (2001). Genetic dissection of the roles of chaperones and proteases in protein folding and degradation in the *Escherichia coli* cytosol. *Mol. Microbiol.* **40**, 397–413.

54. Rozkov, A. & Enfors, S. O. (2004). Analysis and control of proteolysis of recombinant proteins in *Escherichia coli*. *Advan. Biochem. Eng. Biotechnol.* **89**, 163–195.

55. Cubarsi, R., Carrio, M. M. & Villaverde, A. (2001). In situ proteolytic digestion of inclusion body polypeptides occurs as a cascade process. *Biochem. Biophys. Res. Commun.* **282**, 436–441.

56. Carbonell, X. & Villaverde, A. (2002). Protein aggregated into bacterial inclusion bodies does not result in protection from proteolytic digestion. *Biotechnology Letters*, **24**, 1939–1944.

57. Mogk, A., Schlieker, C., Friedrich, K. L., Schonfeld, H. J., Vierling, E. & Bukau, B. (2003). Refolding of substrates bound to small Hsps relies on a disaggregation reaction mediated most efficiently by ClpB/DnaK. *J. Biol. Chem.* **278**, 31033–31042.

58. Weibezahn, J., Schlieker, C., Tessarz, P., Mogk, A. & Bukau, B. (2005). Novel insights into the mechanism of chaperone-assisted protein disaggregation. *Biol. Chem.* **386**, 739–744.

59. Carrio, M. M. & Villaverde, A. (2005). Localization of chaperones DnaK and GroEL in bacterial inclusion bodies. *J. Bacteriol.* **187**, 3599–3601.

60. Carrio, M. M. & Villaverde, A. (2003). Role of molecular chaperones in inclusion body formation. *FEBS Letters*, **537**, 215–221.

61. Rinas, U., Hoffmann, F., Betiku, E., Estape, D. & Marten, S. (2007). Inclusion body anatomy and functioning of chaperone-mediated in vivo inclusion body disassembly during high-level recombinant protein production in *Escherichia coli*. *J. Biotechnol.* **127**, 244–257.

62. Corchero, J. L., Cubarsi, R., Enfors, S. & Villaverde, A. (1997). Limited in vivo proteolysis of aggregated proteins. *Biochem. Biophys. Res. Commun.* **237**, 325–330.

63. Mogk, A., Deuerling, E., Vorderwulbecke, S., Vierling, E. & Bukau, B. (2003). Small heat shock proteins, ClpB and the DnaK system form a functional triade in reversing protein aggregation. *Mol. Microbiol.* **50**, 585–595.

64. Han, M. J., Lee, J. W. & Lee, S. Y. (2005). Enhanced proteome profiling by inhibiting proteolysis with small heat shock proteins. *J. Proteome. Res.* **4**, 2429–2434.

65. Martinez-Alonso, M., Vera, A. & Villaverde, A. (2007). Role of the chaperone DnaK in protein solubility and conformational quality in inclusion body-forming *Escherichia coli* cells. *FEMS Microbiol. Letters*, **273**, 187–195.

66. Sambrook, J., Fritsch, E. & Maniatis, T. (1989). *Molecular Cloning, A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

67. Thomas, J. G. & Baneyx, F. (1998). Roles of the *Escherichia coli* small heat shock proteins IbpA and IbpB in thermal stress management: comparison with ClpA, ClpB, and HtpG In vivo. *J. Bacteriol.* **180**, 5165–5172.

68. Mogk, A., Tomoyasu, T., Goloubinoff, P., Rudiger, S., Roder, D., Langen, H. & Bukau, B. (1999). Identification of thermolabile *Escherichia coli* proteins: prevention and reversion of aggregation by DnaK and ClpB. *EMBO J.* **18**, 6934–6949.

69. Feliu, J. X., Cubarsi, R. & Villaverde, A. (1998). Optimized release of recombinant proteins by ultra-sonication of *E. coli* cells. *Biotechnol. Bioeng.* **58**, 536–540.

70. Carrio, M. M., Cubarsi, R. & Villaverde, A. (2000). Fine architecture of bacterial inclusion bodies. *FEBS Letters*, **471**, 7–11.

71. Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680–685.

72. Gonzalez-Montalban, N., Garcia-Fruitos, E., Ventura, S., Aris, A. & Villaverde, A. (2006). The chaperone DnaK controls the fractioning of functional protein between soluble and insoluble cell fractions in inclusion body-forming cells. *Microb. Cell Fact.* **5**, 26.

73. Liberek, K., Galitski, T. P., Zylicz, M. & Georgopoulos, C. (1992). The DnaK chaperone modulates the heat shock response of *Escherichia coli* by binding to the sigma 32 transcription factor. *Proc. Natl Acad. Sci. USA*, **89**, 3516–3520.

74. Gottesman, S., Clark, W. P. & Maurizi, M. R. (1990). The ATP-dependent Clp protease of *Escherichia coli*. Sequence of clpA and identification of a Clp-specific substrate. *J. Biol. Chem.* **265**, 7886–7893.

75. Wickner, S., Gottesman, S., Skowyra, D., Hoskins, J., McKenney, K. & Maurizi, M. R. (1994). A molecular chaperone, ClpA, functions like DnaK and DnaJ. *Proc. Natl Acad. Sci. USA*, **91**, 12218–12222.

76. Singh, S. K., Guo, F. & Maurizi, M. R. (1999). ClpA and ClpP remain associated during multiple rounds of ATP-dependent protein degradation by ClpAP protease. *Biochemistry*, **38**, 14906–14915.

77. Hoskins, J. R., Singh, S. K., Maurizi, M. R. & Wickner, S. (2000). Protein binding and unfolding by the chaperone ClpA and degradation by the protease ClpAP. *Proc. Natl Acad. Sci. USA*, **97**, 8892–8897.

78. Allen, S. P., Polazzi, J. O., Gierse, J. K. & Easton, A. M. (1992). Two novel heat shock genes encoding proteins produced in response to heterologous protein expression in *Escherichia coli*. *J. Bacteriol.* **174**, 6938–6947.

79. Thompson, M. W., Singh, S. K. & Maurizi, M. R. (1994). Processive degradation of proteins by the ATP-dependent Clp protease from *Escherichia coli*. Requirement for the multiple array of active sites in ClpP but not ATP hydrolysis. *J. Biol. Chem.* **269**, 18209–18215.

80. Goff, S. A. & Goldberg, A. L. (1985). Production of abnormal proteins in *E. coli* stimulates transcription of lon and other heat shock genes. *Cell*, **41**, 587–595.

81. Maurizi, M. R. (1987). Degradation in vitro of bacteriophage lambda N protein by Lon protease from *Escherichia coli*. *J. Biol. Chem.* **262**, 2696–2703.

82. Van Melderen, L., Thi, M. H., Lecchi, P., Gottesman, S., Couturier, M. & Maurizi, M. R. (1996). ATP-dependent degradation of CcdA by Lon protease. Effects of secondary structure and heterologous subunit interactions. *J. Biol. Chem.* **271**, 27730–27738.

83. Gonzalez, M., Frank, E. G., Levine, A. S. & Woodgate, R. (1998). Lon-mediated proteolysis of the *Escherichia coli* UmuD mutagenesis protein: in vitro degradation and identification of residues required for proteolysis. *Genes Dev.* **12**, 3889–3899.

84. Smith, C. K., Baker, T. A. & Sauer, R. T. (1999). Lon and Clp family proteases and chaperones share homologous substrate-recognition domains. *Proc. Natl Acad. Sci. USA*, **96**, 6678–6682.