Engineering and Production of Quality Viral Proteins in Prokaryotic and Eukaryotic Systems

Tesi doctoral

Mónica Martínez Alonso

2010

Departament de Genètica i de Microbiologia

Institut de Biotecnologia i de Biomedicina

Facultat de Biociències

Universitat Autònoma de Barcelona

PhD Programme in Biotechnology

Engineering and Production of Quality Viral Proteins in Prokaryotic and Eukaryotic Systems

Report presented by Mónica Martínez Alonso in order to complete the requirements to be granted the degree of Doctor of Philosophy in Biotechnology by the Autonomous University of Barcelona.

Mónica Martínez Alonso

Approval of the thesis directors,

Antonio Villaverde Corrales Neus Ferrer Miralles Rob Noad

A mis padres,

Contents

Contents

1. Introduction 7
1.1. Overview of the currently available protein production systems 11
1.2. *Escherichia coli* for recombinant protein production 14
1.2.1. Protein folding 18
1.2.2. Quality control in the bacterial cytoplasm 21
1.2.2.1. Chaperones 21
1.2.2.2. Proteases 26
1.2.3. Inclusion bodies 28
1.2.3.1. Morphology, composition and structure 29
1.2.3.2. Minimising inclusion body formation 32
1.2.3.3. Conformational quality of inclusion body proteins 33
1.3. The baculovirus-insect cell expression system 35
1.3.1. Overview of baculovirus biology 36
1.3.1.1. Baculovirus structure 36
1.3.1.2. Infection progress 37
1.3.2. Expression vectors 39
1.3.2.1. Transfer plasmids 40
1.3.2.2. Parental genomes 42
1.3.3. Insect hosts 47
1.3.3.1. Cell lines 47
1.3.3.2. Insect larvae 48
1.4. Model proteins 50
1.4.1. Green Fluorescent Protein 50
1.4.2. Foot-and-Mouth Disease Virus VP1 and VP2 capsid proteins 53
1.4.3. Human α-Galactosidase 54
1.5. Previous work 55
2. Objectives 57

Contents

3. Results 61
3.1. Article 1 63
3.2. Article 2 71
3.3. Article 3 77
3.4. Article 4 85
3.5. Article 5 93
4. Discussion 103
4.1. Independent control of protein yield and quality 106
4.2. Functional status of soluble protein 108
4.3. Bacterial folding modulators for eukaryotic systems 110
5. Conclusions 117
6. Annex I 121
7. Annex II 133
8. References 147
9. Acknowledgements 181

1. Introduction

Biotechnology is defined as the use of living organisms or biological substances to perform specific industrial or manufacturing processes, and as such, it has been known to mankind for a long time. Over 10,000 years ago, long before the term 'biotechnology' was even coined microorganisms were already used in fermentation processes to produce wine, beer or bread. Early farmers, even if unaware, also relied on biotechnology for crop improvement through careful seed selection to obtain higher yields or better taste.

More recently, the end of the 19th century experienced a significant improvement in health conditions in over-crowded industrial cities when large-scale sewage purification systems based on microbial activity were first introduced [1]. That was also the time when fermentation industry was born, as industrial processes were developed for the manufacture of chemicals such as acetone or butanol using bacteria [1]. Moreover, cowpox vaccines produced by Jenner in 1796 and the discovery of penicillin by Alexander Fleming in 1927 [2] and its further development in the 1940s are two examples of the early impact of biotechnology in the medical arena.

However, the development of biotechnology as we know it today would still need some major breakthroughs. The first came with the discovery of the double helix structure of DNA in 1953 by Watson and Crick [3], followed by the cracking of the genetic code by Marshall Nirenberg and Heinrich J. Matthaei in 1961 [4]. Soon after, in the early 1970s the discovery of new restriction enzymes by Paul Berg [5], combined with Herbert Boyer and Stanley Cohen's first genetic engineering of living organisms [6] gave way to recombinant DNA technology. The modern biotechnology era had just started.

Today, biotechnology is present in nearly all sectors of industry, with applications in major areas such as medicine, agriculture and crop production, and environment. Products derived from biotechnology have steadily increased over the years, and those commercially available today include antibiotics, antibodies, biofuels, fermented foods and beverages and recombinant proteins [1].

Proteins are the building blocks of life. No matter their origin, all proteins are assembled from a set of 20 amino acids linked together to form the linear chain that defines their primary structure. Being the most abundant macromolecules in living cells, they are also highly versatile. Their biological importance lies in the fact that proteins are the molecular tools required to carry out the functions encoded in the genome, with almost every event that takes place in a cell requiring action from one or several

proteins. Thus, they have important roles in cellular processes such as cell signalling, immune responses, cell adhesion or the cell cycle. Proteins also have structural roles and act as catalysts in many cell reactions [7]. This versatility translates in recombinant proteins (i.e. those derived from recombinant DNA) having applications in a wide variety of sectors, ranging from biopharmaceutical to enzyme and agricultural industries. Also, because they enter both industrial and therapeutic markets, recombinant proteins have a prominent position in the economical arena.

Although insulin was the first pharmaceutical produced as early as 1922 [8], the difficulty to obtain proteins from their natural sources in sufficient amounts for their study, characterisation and further use still represented a major roadblock. The availability of new restriction enzymes and recombinant DNA techniques, together with the parallel development of heterologous systems for recombinant protein production has resulted in an increasing number of commercially available biotechnological products, which has in turn boosted the biotechnological industry.

Some examples of the already marketed recombinant proteins include human insulin (which became the first *E. coli* produced biopharmaceutical approved by the FDA in 1982 [9]), growth hormone, Factor VIII or gamma interferon [10]. Enzymes are also marketed either for industrial use (amidase for the production of 6-aminopenicillanic acid, nitrile hydratase to produce acrylamide, amylases, proteases...) or to be used as therapeutic agents in the treatment of diseases like thromboses, cystic fibrosis, metabolic diseases or even cancer [1]. The production system must be carefully chosen to successfully obtain each of these proteins, as protein features and the processing abilities of the recombinant host will ultimately determine whether a protein can be obtained in a functional form.

1.1. Overview of the currently available protein production systems

The product to be obtained is the key element to be considered when choosing a production system. Depending on protein features such as size, origin or need for post-translational modifications, the available options will be narrowed down to the most convenient expression system. Production costs, time constraints and the yield and quality of the product must also be taken into account.

Prokaryotes are usually the first choice for protein production because of their fast growth and availability of easy-to-handle procedures. The many advantages of *Escherichia coli* make it the most widely used and best characterised microorganism. Cultivation is easy and essentially inexpensive. Recombinant gene expression is fast and high protein yields can be obtained in a cost-effective manner. Although recombinant proteins can be engineered for secretion to the periplasmic space, *E. coli* is often used for the production of cytoplasmic proteins. Despite the many advantages of this host, recombinant protein production in *Escherichia coli* has some drawbacks too. The two main obstacles encountered are proteolytic digestion by cell proteases [11] and accumulation of the protein in insoluble deposits, known as inclusion bodies (IBs) [12;13]. Both events are the result of the recombinant protein not being able to reach its native conformation. Although many strategies have been devised along the years to reduce inclusion body formation and promote the synthesis of soluble protein, protein deposition in inclusion bodies still represents a major bottleneck for protein production in this system. Moreover, eukaryotic proteins are often obtained as insoluble or inactive, due to the inability of the system to carry out complex post-translational modifications. However, the N-glycosylation system of *Campylobacter jejuni* has successfully been transferred to *E. coli*, rendering a strain capable of glycosylation [14].

Other bacteria can also be used as cell factories. *Bacillus* systems provide the advantage of stronger secretion compared to *E. coli*. Also, they have GRAS (Generally Recognised as Safe) status, which will eventually facilitate FDA approval of recombinant proteins obtained in this system. *Bacillus megaterium*, *B. subtilis*, *B. licheniformis* and *B. brevis* are often used for expression [1]. However, the production of many extracellular proteases by *B. subtilis* represents an important drawback.

Among the eukaryotic organisms, single-celled yeasts represent the simplest system. In common with *E. coli*, yeasts are also fast and cost-effective for protein production, offering high yields of the recombinant product and with the added advantage of being able to perform post-translational modifications. For this reason, many proteins which

fail to fold properly in *E. coli* or require post-translational modifications are produced in yeast. However, glycosylation patterns are different from higher eukaryotes [15]. The genetics of the system are well characterised, with the most common hosts being *Saccharomyces cerevisiae* and *Pichia pastoris*. Although approved biopharmaceuticals produced in yeast are derived exclusively from *S. cerevisiae* [15], *P. pastoris* is currently the most widely used yeast for heterologous protein expression due to its superior secretion characteristics [10].

Filamentous fungi provide complex post-translational modifications, which are then more similar to the mammalian version [10]. However, the system is not well characterised both genetically and physiologically, secretion yields are not competitive and proteases can hamper protein production [1;10].

Insect cells can perform post-translational modifications which are even more complex than those carried out in fungi. Being animal cells, cultivation is more difficult and expensive but they are still more resistant and easy to handle than mammalian systems. Their folding machinery is better suited for mammalian proteins, and thus soluble proteins of mammalian origin can be obtained [16]. Protein production is accomplished by infection of the insect cell host with a recombinant baculovirus encoding the target protein. Other advantages of this system include proper disulfide bond formation and high expression levels. The system is safe as baculovirus vectors have a restricted host range, infecting only insects but not vertebrates. Cells can be adapted to suspension cultures and chemically defined, serum-free media. Large proteins and also multi-protein complexes have been obtained, and simultaneous expression of multiple genes is also possible [17;18]. However, some shortcomings are also present. Proteins can sometimes be seen as intracellular aggregates [19;20], protease activity is high [10;21;22] and glycosylation patterns provided by insects still differ from mammalians, limiting protein half-life when administered to humans [23].

Mammalian cell lines are sometimes the only choice for expression of difficult proteins, especially heavily glycosylated ones. Expressed proteins are often soluble and active, and high yields are obtained. However, the system is expensive and process duration is long. Nevertheless, most of the approved therapeutic proteins have been obtained in hamster-derived cell lines, namely CHO (Chinese Hamster Ovary) and BHK (Baby Hamster Kidney) [15]. These cell lines can also be adapted to suspension cultures and defined serum-free media, which increases the biosafety of the recombinant

products. Although they are both recognised as safe regarding infectious and pathogenic agents [10], lack of contamination by viruses and DNA still needs to be proven [1].

Transgenic animals are also used to produce recombinant proteins in milk, egg white, blood, urine, seminal plasma and silk worm cocoons [1]. So far, milk has given the best results. Although production in milk is more cost-effective than in mammalian cell culture [1], safety concerns represent a great challenge because of possible transmission of infectious diseases (both viral and prion infections) and immunogenic responses [15].

Transgenic plants have also been used for production of recombinant proteins. The system presents many advantages, such as being cheap, highly productive, easy to scale up, and safe as it lacks human pathogens. Eukaryotic post-translational modifications are also available. However, disadvantages of transgenic plants include possible contamination with pesticides, herbicides and toxic plant metabolites [24], and the need to deal with the uncontrolled spread of the transgenic gene. Also, negative public perception of transgenic plants does not encourage their use as a promising system.

Besides recombinant protein production in prokaryotic or eukaryotic hosts, protein synthesis is also possible in cell-free expression systems, where transcription and translation reactions are carried out *in vitro*. This system is fast and simple, and an excellent alternative for proteins which are toxic for the host when produced *in vivo* [25].

Because they have been the two expression systems used in this study, both *E. coli* and the Baculovirus Expression System will be discussed in further detail.

1.2. *Escherichia coli* for recombinant protein production

Escherichia coli is the most widely used prokaryotic organism for expression of recombinant proteins [26]. Being one of the most studied microorganisms since early times, its genetics and physiology are well-known and this has facilitated the development of the wide set of molecular tools available today [15].

The use of *E. coli* as a host for protein production is relatively simple and inexpensive [27]. Added advantages include its short duplication time, growth to high cell densities, ease of cultivation and high yields of the recombinant product, which can accumulate up to around 30% of the total protein content of the cell [10;27;28]. Thus, it is not surprising that almost 30% of the recombinant proteins that are currently on the market are obtained in *E. coli* [15].

The basic requirement for protein production in *E. coli* is a strain that provides a suitable genetic background and harbours a compatible plasmid encoding the gene to be expressed [27]. The deep knowledge of the system provides flexibility and allows a better control of protein production. However, the choice of both strain and expression plasmid has to be carefully considered, as there are some key elements that need to be taken into account:

➢ *Host strain*

The most important feature to consider is the ability of the host strain to stably maintain the expression plasmid. Moreover, for some expression systems the host strain will also be required to provide relevant genetic elements (e.g., DE3 in the pET system).

Expression strains deficient in the main proteases have been developed with the aim of attaining a more efficient recovery of intact protein [29-31]. In this regard, there are currently many strains commercially available. BL21 is a non-pathogenic *E. coli* B strain deficient in ompT and Lon proteases. Novagen BLR is a recA$^-$ BL21 derivative, used to improve stability of plasmids with repetitive sequences. However, proteases are an important element of the protein quality control system, surveying conformational quality in cooperation with other folding assistants (*see section 1.2.2.2*). Therefore, although proteolysis is minimised in protease deficient mutants, this leads to the accumulation of the misfolded polypeptides in the form of inclusion bodies [32-34].

Strains for improved disulfide bond formation are also available. The genes for thioredoxin and glutathione reductases are disrupted in Novagen Origami (*trxB/gor*) and AD494 (*trxB*) strains, thus allowing disulfide bond formation in the cytoplasm of *E. coli*.

Other mutants can enhance soluble expression of difficult proteins (Avidis C41(DE3) and C43(DE3) strains) or allow for adjustable levels of protein expression (Novagen Tuner series). Rosetta and Rosetta-gami strains are also useful to alleviate use of codon bias (see below). A summary of *E. coli* strains commonly used for protein production is presented in Table 1.

Table 1. *E. coli* strains for recombinant protein production.

E. coli strain	Derived	Relevant features
AD494	K-12	Cytoplasmic disulfide bond formation enabled (trxB mutant)
BL21	B834	Deficient in lon and ompT proteases
BL21 trxB	BL21	Cytoplasmic disulfide bond formation enabled (trxB mutant) Deficient in lon and ompT proteases
BL21 CodonPlus-RIL	BL21	Deficient in lon and ompT proteases Overcome bias in codon usage (supplies AGG, AGA, AUA and CUA codons)
BL21 CodonPlus-RP	BL21	Deficient in lon and ompT proteases. Overcome bias in codon usage (supplies AGG, AGA and CCC codons)
BLR	BL21	Stabilizes tandem repeats (recA mutant) Deficient in lon and ompT proteases
B834	B strain	Met auxotroph; 35S-met labeling
C41	BL21	Mutant for expression of membrane proteins
C43	BL21	Double mutant for expression of membrane proteins
HMS174	K-12	Stabilizes tandem repeats (recA mutant) Rifampicin resistance
JM 83	K-12	Protein secretion to periplasm
Origami	K-12	Enhanced cytoplasmic disulfide bond formation (trxB/gor mutant)
Origami B	BL21	Enhanced cytoplasmic disulfide bond formation (trxB/gor mutant) Deficient in lon and ompT proteases
Rosetta	BL21	Deficient in lon and ompT proteases Overcome bias in codon usage (supplies AUA, AGG, AGA, CGG, CUA, CCC, and GGA codons)
Rosetta-gami	BL21	Enhanced cytoplasmic disulfide bond formation (trxB/gor mutant) Deficient in lon and ompT proteases Overcome bias in codon usage (supplies AUA, AGG, AGA, CGG, CUA, CCC, and GGA codons)

All strains are commercial, and most are also available as DE3 and DE3 pLysS strains.

Adapted from *Appl Microbiol Biotechnol*. 2006 Sep;72(2):211-22.

➢ *Plasmids for gene expression*

Plasmids are double-stranded circular DNA molecules that replicate independently of the host's chromosome.

Expression plasmids contain several genetic elements:

- The replicon, which contains the origin of replication that will in turn determine the plasmid copy number [35]. For multi-copy expression plasmids, ColE1 and p15A are the most common. Also, plasmid incompatibility groups must be taken into account when gene products are to be co-expressed from different plasmids. In that case, different replicon incompatibility groups will be required for plasmids to be compatible. In that regard, plasmids containing ColE1 and p15A are compatible, and thus are frequently combined for co-expression.

- Resistance markers, which confer a genetic trait that allows for artificial selection. Common resistance markers include ampicillin, kanamycin, chloramphenicol or tetracycline. Ampicillin resistance is obtained by expression of β-lactamase from the *bla* gene encoded in the plasmid. When secreted to the periplasm, the enzyme hydrolises the β-lactam ring. Kanamycin, chloramphenicol and tetracycline bind to the ribosomes, interfering with protein synthesis. Aminoglycoside phosphotransferases inactivate kanamycin in the periplasm, and resistance to chloramphenicol is provided by chloramphenicol acetyl transferase. Resistance to tetracycline can be conferred by several genes. However, *tetA* genes encoding a tetracycline efflux system or *tetM* and *tetQ,* encoding a protein that protects ribosomes from the inhibiting effects of tetracycline, are often used in molecular biology.

- Transcriptional promoters, which enable control of the gene expression levels in inducible systems. Ideally, promoters should be strong to provide high yields of the recombinant protein. It is also convenient that the inducer is cheap in order to minimise production costs. Promoter induction can be either thermal or chemical. Thermal induction will usually require a temperature upshift, whereas for chemical induction isopropyl-beta-D-thiogalactopyranoside (IPTG) is the most common molecule [36]. Minimising basal transcription is important, especially when the expression of target genes poses a cellular stress. This is achieved by the presence of a suitable repressor that will bind the promoter in absence of inducer.

- Translation initiation regions, which are necessary for ribosome binding to messenger RNA. Thus, these will include a ribosomal binding site (RBS) containing the

Shine-Dalgarno sequence located 7±2 nucleotides upstream the canonical AUG translation initiation codon used in efficient recombinant systems [37;38].

- Transcriptional terminators, which prevent transcription starting from irrelevant promoters or through the origin of replication. They are placed downstream of the sequence encoding the gene and stabilise mRNA by forming a stem loop at the three prime end [39].

- Translational terminators, which mediate translation termination usually by the stop codon UAA in *E. coli*. Efficiency can be increased by placing several stop codons together [40].

➢ *Stability of messenger RNA*

Gene expression levels mainly depend on four factors: efficiency of transcription, mRNA stability, frequency of translation and protein stability. Although transcription and translation have been thoroughly optimised in recombinant expression systems, mRNA stability is not often addressed. Therefore, gene expression is controlled by mRNA decay. Because of this, mRNA stability is an important factor in controlling gene expression levels as the expression rate depends directly on its stability, with the average half-life of mRNA in *E. coli* ranging from seconds to 20 minutes [41;42].

Messenger RNA is susceptible to degradation by cellular RNases, and protection depends on its folding, protection of ribosomes and polyadenylation, which in bacteria influences mRNA stability by promoting its decay. Thus, in poly(A)-deficient strains, mRNA is stabilised [43;44]. Moreover, commercially available mutant strains for the RNaseE gene (Invitrogen BL21 star) provide enhanced mRNA stability [45].

➢ *Bias in codon usage*

Because the genetic code is degenerate, most amino acids can be determined by more than one codon. Also, the preferred codons for each amino acid vary in different organisms and this can become a problem in recombinant expression systems.

Heterologous genes from viral origin, eukaryotes or archaeabacteria often contain high frequencies of codons which are rare in *E. coli* [46]. Because of the low availability of the tRNAs corresponding to rare codons, ribosomes are likely to stop at those

positions [47]. This leads to translational errors that can include amino acid substitutions, frameshifting or premature termination [48;49].

To overcome this bias, the recombinant gene sequence can be engineered so that rare codons are substituted by those which are optimal for the host system. Although this strategy can result in enhanced expression levels and reduced translational errors [50;51] it is also time-consuming, especially when considering biotechnological high-throughput applications. A faster alternative consists in co-transforming the host with a plasmid encoding the tRNAs corresponding to the problematic codons. Complementation plasmids and already transformed strains, such as Novagen Rosetta and Rosetta-gami, are commercially available for this purpose.

1.2.1. Protein folding

Protein folding is the process by which an unfolded polypeptide adopts its characteristic three-dimensional and functional structure. According to the fundamental principle of protein folding stated by Anfinsen in 1973, the folding of a protein is determined by its amino acid sequence, which contains all the information required for the protein to reach its native conformation [52]. The native conformation of a protein is usually the most thermodynamically stable, having the lowest Gibbs free energy. However, even if this means that thermodynamics is the driving force that guides protein folding it does not explain how most proteins reach their native conformation in a matter of seconds, as randomly exploring the billions of possible spatial conformations would take astronomical amounts of time. This view, which is known as the Levinthal paradox [53], assumes that the folding of every residue is independent from the rest. However, since folding is a cooperative process [54;55] every residue does not have to search for random conformation states, as their conformational freedom will be narrowed down by the folding of previous residues [56].

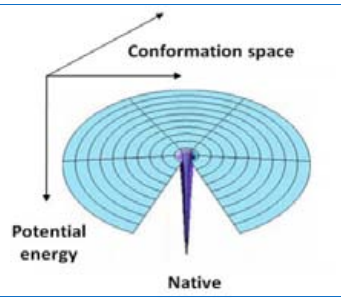

Figure 1. Model of the energy landscape for a polypeptide folding, according to Levinthal. Every residue folds independently from each other, so the time required for the protein to reach the native conformation is extremely large.

Adapted from *Nat Struct Biol.* 1997 Jan; **4**(1):10-19.

Levinthal also suggested that the stable conformation could have a higher energy if the lowest Gibbs energy was not kinetically accessible. Thus, different kinetic models have been proposed to solve the paradox.

The hydrophobic collapse model describes the initial stages of protein folding. Hydrophobic forces, which drive the collapse, arise from the repulsion between hydrophobic side chains of the protein and the hydrophilic water molecules of the environment. The collapse results in the protein being in the "molten globule" state, with hydrophobic side chains in the interior while the hydrophilic residues are on the surface. With a volume slightly larger than the native structure of the protein, the molten globule contains secondary structures but lacks a definite tertiary structure [57].

The nucleation theory proposes the existence of folding nuclei in the protein structure during the early stages of folding. The most recent view of this theory [58] proposes a mechanism in which weak nuclei are stabilised by long distance interactions.

Currently, the "new-view" in protein folding is illustrated by the folding funnel model proposed by Wolynes and co-workers [59]. This model, which requires a high cooperativity and is therefore very fast [57], describes both the thermodynamic and kinetic behaviour that unfolded polypeptides undergo to reach their native state, and is represented in terms of energy landscapes. Multiple pathways exist, and every single polypeptide can follow its own route. The number of possible conformations decreases towards the bottom of the funnel, and the folding is faster as the slope grows steeper [60]. For a protein which can only have two states, unfolded and native, a smooth mechanism is the simplest way of folding. A two-state folding reflects the existence of an energy barrier between unfolded and native states. When there is no energy barrier, this is called smooth folding [56] (Figure 2A), which is often seen when the viscosity of the solvent is the only limitation for protein folding [61]. Moreover, a protein can sometimes either fold by a two-state mechanism or adopt an intermediate conformation where unfolded and folded states coexist, which presents a kinetic trap (Figure 2B).

When cooperativity is not so high, distinct intermediates occur in the folding process, with local structures that can be different to those observed in the native structure for the same residues [57]. These structures may be locally favorable but unfavorable for the whole structure, which leads to kinetic traps determined by the presence of local energy barriers. This is represented by a rugged energy landscape (Figure 2C) which is often useful to picture the nucleation model, where local folding nuclei are formed prior to the molecule adopting its native conformation.

Sometimes polypeptides can fall into kinetic traps with a global free energy similar to that of the folded state. In this case, the deep kinetic trap results in the two conformers not being able to interconvert in a reasonable time scale, which may lead to misfolding and aggregation of the protein. The rough energy landscape corresponding to this scenario is depicted in Figure 2D.

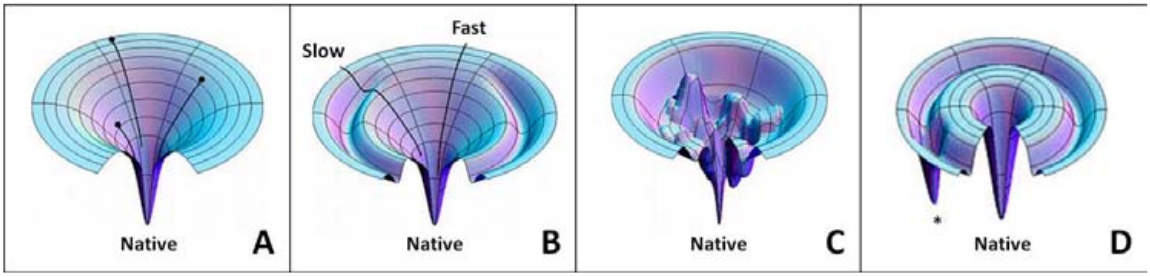

Figure 2. A) Smooth funnel for a protein following a two-step folding. **B)** Fast-folding process, in parallel with a slow-folding process involving a kinetic trap. **C)** Rugged energy landscape with kinetic traps and energy barriers for a multi-state folding protein. **D)** Rough energy landscape depicting a deep kinetic trap (*) easily accessible from unfolded conformations. Access to the global energy minimum will be very slow for trapped intermediates.

Adapted from *Nat Struct Biol.* 1997 Jan; **4**(1):10-19 (panels A-C) and *Proteins.* 1998 Jan; **30**(1):2-33 (panel D).

Folding in the cellular environment presents an extra challenge. In the very crowded *E. coli* cytoplasm, transcription and translation are tightly coupled. With proteins being released from the ribosomes at a rate of one every 35 seconds [62], the cytoplasm becomes a very crowded space where macromolecule concentrations can reach 300-400 mg/mL [63]. Because of this, many proteins need assistance of folding modulators to reach their native conformation. This requirement is dramatically increased in the context of recombinant protein production, when the cell has an additional input of *de novo* synthesis. In fact, folding modulators are considered to be limiting in these conditions.

During folding, proteins can establish persistent non-native interactions that significantly affect their structure and biological functions. This is known as “misfolding” [64]. Misfolded and incompletely folded polypeptides expose hydrophobic stretches that would be hidden in the native conformation, which makes them prone to aggregation [65]. Failure of proteins to fold correctly, or to remain properly folded, gives rise to malfunctioning of living systems [66-68]. In humans, diseases related to incorrect protein folding, which prevents their normal function, include cystic fibrosis [66] and

some types of cancer [69]. Proteins with high tendency to misfold can form aggregates within cells or in the extracellular space, which can also be deposited in tissues such as brain, heart or spleen [67;68;70;71]. Disorders involving aggregate deposition in tissues include Alzheimer's and Parkinson's diseases, the spongiform encephalopathies and type II diabetes. Thus, living organisms have cellular factors responsible for avoiding aggregation by assisting in protein folding, such as molecular chaperones and folding catalysts [72;73]. In addition, proteases assist in protein quality control by degrading irreversibly damaged polypeptides which cannot be rescued by the action of chaperones.

1.2.2. Quality control in the bacterial cytoplasm

Surveillance of protein quality is accomplished by the coordinated action of chaperones and proteases, which act together to assist protein folding, prevent accumulation of misfolded polypeptides, remove protein from aggregates and degrade folding-reluctant species [74]. Thus, the system promotes solubility by minimising aggregation. Solubility, expressed as the relative amount of recombinant protein in the soluble cell fraction, is the parameter commonly used to evaluate the success of biotechnological processes regarding protein quality [75;76]. Although in *E. coli* quality control takes place both in the cytoplasm and the periplasm, this section will focus on the cytosolic branch of the quality control system.

1.2.2.1. Chaperones

The term "chaperone" was first used by Ron Laskey in 1978 to describe an activity associated to nucleoplasmin in *Xenopus* oocytes, which allowed the correct assembly of histones into nucleosomes [77]. Currently, the term chaperone includes a much wider set of more than 20 protein families which have a major role in the quality control of the proteome [74;78;79]. Although chaperones are constitutively expressed in physiological conditions, they become upregulated under stress situations. As thermal stress promotes an increase of chaperone levels in the cell, they have traditionally been named as heat shock proteins (Hsp) [80]. However, not all heat shock proteins are chaperones and *vice versa*. In *E. coli* this stress response is positively regulated at the transcriptional level by the product of the *rpoH* gene, the factor σ^{32}, which binds as an alternative σ

subunit to the RNA polymerase and targets it to the promoters of the heat shock genes [81;82].

Molecular chaperones constitute one of the better characterised groups of folding modulators, highly conserved in all kingdoms of life. These ubiquitous proteins play a central role in the conformational control of the proteome by helping other polypeptides reach their native conformation without affecting their folding rates or becoming part of their final structure. Chaperones bind hydrophobic patches of amino acids that would normally be buried within the core of the substrate protein, but have become exposed to the solvent because of their incorrect folding. The transient formation of chaperone-substrate complexes shields misfolded polypeptides from interacting with each other [83]. Chaperones normally target short unstructured stretches of hydrophobic amino acids which lack acidic residues and are flanked by basic ones. These motifs are extremely common, which explains why chaperones are so promiscuous [84].

Based on their mechanism of action, molecular chaperones can be divided into three functional subclasses:

- Folding chaperones, which drive the net refolding/unfolding of their bound substrates through ATP-mediated conformational changes. These chaperones promote the yield of correctly folded proteins without affecting their folding rates. Folding chaperones in the *E. coli* cytoplasm are the trigger factor (TF) [85] and the DnaK-DnaJ-GrpE and GroELS systems [86].

- Holding chaperones, which bind to partially folded proteins and stabilise them until folding chaperones become available, thus preventing them from aggregation [87-89]. In *E. coli*, the best characterised holding chaperones are IbpA and IbpB, which belong to the group of small Hsp family [90] and are commonly found within inclusion bodies [91]. Hsp31 is another cytoplasmic modulator in this group, which binds early unfolding intermediates under severe stress conditions and therefore prevents overloading of the DnaK-DnaJ-GrpE system [92]. Another holdase is Hsp33, a redox-regulated chaperone that deals with oxidative protein misfolding [93].

- Disaggregating chaperones, which promote protein removal from inclusion bodies and other aggregates formed under prolonged or severe stress conditions [84;94]. Solubilisation of protein aggregates occurs through ATP-driven conformational changes, and polypeptides are transferred to folding chaperones for refolding [83]. ClpB

is the best characterised disaggregase, and works together with DnaK and IbpAB chaperones assisting refolding and promoting the solubilisation of protein aggregates [95-97].

I. Trigger factor

Trigger factor is a three-domain cytosolic chaperone which associates to the large subunit of the ribosomes, close to the exit site, where it binds to nascent polypeptidic chains and thus stabilises them [84]. This chaperone also exhibits peptidyl-prolyl cis/trans isomerase activity (PPIase), although the presence of proline residues in its substrates is not required [98]. Unlike other chaperones, trigger factor is not an ATPase [99]. In addition, trigger factor is not a heat shock protein either. Indeed, it is induced upon cold shock and thus enhances cell viability at low temperatures [100].

Therefore, trigger factor aids in *de novo* protein folding by stabilising nascent chains or targeting them to other chaperones, like the DnaK-DnaJ-GrpE system with which it has been shown to cooperate [101].

II. The Hsp70 system: DnaK, DnaJ and GrpE

After being released from trigger factor, a newly synthesised polypeptide can either fold into its native conformation without any further help or require assistance of other chaperone sets. In this early stage of folding, polypeptides will expose unfolded segments. The major cytosolic chaperones involved in the recognition of this set of substrates are the Hsp70 system [102], that being highly conserved is present in all kingdoms of life. The bacterial member of the Hsp70 family is the chaperone DnaK, which acts together with its cofactor DnaJ (the Hsp40 homologue) [103] and a nucleotide exchange factor named GrpE [104;105]. Although all the three proteins are induced by heat shock, only DnaK has ATPase activity.

DnaK has a wide set of roles in the multichaperone network, such as folding newly synthesised polypeptides [73;106], mediating ATP-dependent unfolding, preventing aggregation, stabilising substrates for refolding by GroELS [107-113], solubilising protein aggregates in cooperation with ClpB and Ibps [88;107;114-118], participating in proteolysis [119;120] and protecting proteins against oxidative damages [121;122]. Moreover, it is also a negative regulator of the heat shock response acting in

cooperation with DnaJ, which binds the σ^{32} subunit of the RNA polymerase and targets it for degradation by the inner-membrane associated Ftsh protease [82].

DnaK is a monomeric protein with an N-terminal ATPase domain, a substrate binding site formed by two β-sheets and a C-terminal domain that interacts with partner proteins to modulate their function [123;124]. DnaK has two functional states depending on the phosphorylation state of the bound nucleotide. Affinity for substrates is low when DnaK is bound to ATP and high when bound to ADP [125-128]. DnaJ is a modular dimeric protein with at least four distinct domains. The J domain is a highly conserved motif which stimulates the ATPase activity of DnaK, converting it to the high affinity ADP-DnaK state [129]. DnaJ has chaperone activity itself and the C-terminal region seems to be the substrate binding site [99]. GrpE is a homodimer that binds to DnaK in a ratio of 2:1 [130;131]. It binds to the ATPase domain of DnaK causing the dissociation of ADP which determines the transition to the low affinity state. This results in release of the substrate from the chaperone [132;133].

During the functional cycle of the Hsp70 system the target polypeptide is first bound by DnaJ, which recognises hydrophobic stretches in its structure. The DnaJ-bound polypeptide is then transferred to DnaK, which is bound to ATP and thus in a low affinity state. Both DnaJ and the substrate stimulate the ATPase activity of DnaK, which hydrolyses ATP switching to the high affinity ADP-bound state. Thus, a stable ADP-DnaK substrate complex is formed. GrpE binding to DnaK stimulates nucleotide exchange and therefore ADP is dissociated, destabilising the interaction between DnaK and its substrate, which is then released. After completion of this cycle, the released polypeptide can fold to its native conformation, require more cycles in this system or be transferred to the GroELS chaperones. Proteins which have unfolded as a result of stress conditions can also be refolded by this system [84].

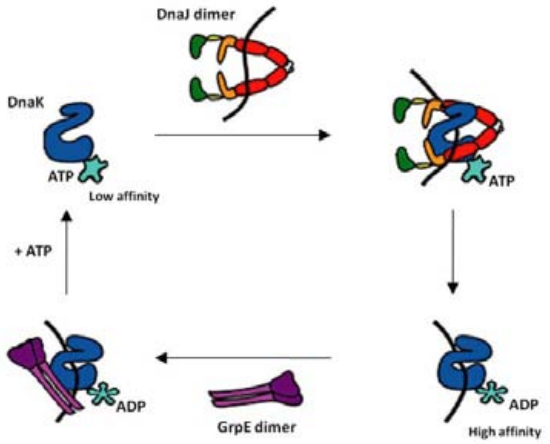

Figure 3. Functional cycle of the bacterial Hsp70 system.

Adapted from *Mol Microbiol.* 2007 Nov;66(4):840-57.

III. ClpB

Clp ATPases are members of the AAA family of proteins (ATPases Associated with a variety of cellular Activities) [134]. The highly conserved AAA module is the key feature of this family. Structurally, they are formed by subunits arranged in ring-shaped complexes [135-138].

ClpB is one of the main Clp ATPases in *E. coli*. This chaperone is a member of the Hsp100 family and is also induced upon heat shock [99]. ClpB acts by forming a ring-shaped hexameric structure and translocating its substrate protein through an axial channel [139]. It works as a "disaggregase" in cooperation with DnaK-DnaJ-GrpE, reverting aggregation [95;109;115;140]. It has an important role in quality control by removing protein from aggregates in cooperation with DnaK, reducing aggregate size and exposing hydrophobic surfaces [107;114]. Disaggregation is facilitated by the presence of small heat shock proteins within the aggregates [95], but complete renaturation of the partially unfolded substrates requires transfer from ClpB to DnaK [107;118].

IV. The Hsp60 system: GroEL and GroES

The GroEL-GroES system handles around 10% of newly synthesised proteins [141]. This is the only chaperone system of the *E. coli* cytoplasm essential for life under all growth conditions [142]. GroEL is a bacterial chaperonin of around 60 kDa which belongs to the Hsp60 family. Structurally, GroEL forms a large oligomer of approximately 800 kDa organised as two stacked homoheptameric rings, with its cochaperone GroES (a member of the Hsp10 family) always bound to one of the rings [73]. GroEL substrates are structured but non-native proteins up to 60 kDa in size [143]. The mechanism of this chaperone complex is well established *in vitro* [144-147]. In the substrate acceptor state of GroEL, GroES and seven ADP molecules are bound to the same ring. During the folding process, substrates are bound by the GroEL free ring. Then, ATP binding to the newly occupied ring mediates a conformational change [148] that renders GroEL able to bind GroES [73]. A second conformational change results in displacement of the substrate to a chamber defined by the GroEL ring and the GroES cap. This also results in GroES and ADP release from the opposite ring, as well as any previously encapsulated polypeptide. By this mechanism, partially folded polypeptides are allowed to fold at infinite dilution inside the GroEL cavity. Usually, more than one cycle of binding and release will be

required for a protein to fold into its native state [99]. Equally to the Hsp70 system, GroEL-GroES can also refold polypeptides which have become unfolded under stress conditions [84].

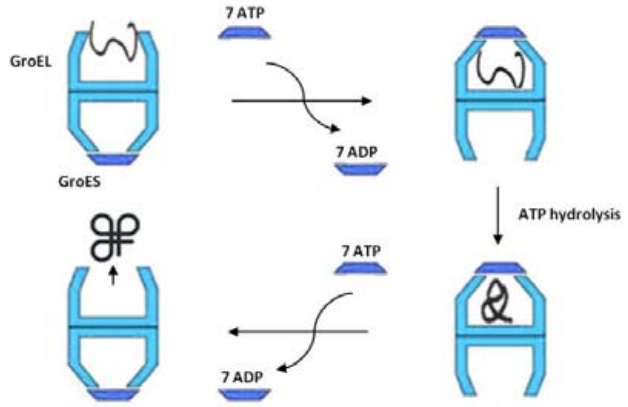

Figure 4. Functional cycle of the bacterial Hsp60 system.

Adapted from *Curr Biol.* 2005 Sep 6;15(17):R661-3.

V. Small heat shock proteins

Small heat shock proteins are ubiquitous and conserved proteins belonging to the group of the holding chaperones [90]. In *E. coli*, the best characterised are the Inclusion Body Proteins (Ibp) which receive their name because of their frequent association with inclusion bodies [91] and are usually found forming large oligomeric structures (80, 129). Bacterial IbpA and IbpB are two homologous proteins of 14 and 16 kDa respectively, encoded on a single operon [91]. Although IbpB is mainly soluble, it comigrates to the insoluble fraction when produced with the insoluble IbpA [149]. Their function is not well understood, but they seem to bind hydrophobic stretches of thermally unfolded polypepdides protecting them from aggregation until the stress disappears. Then, Ibp-bound polypeptides are transferred to DnaK or GroEL for refolding [149-152]. Recently, IbpA and IbpB have been shown to assist in the disaggregating and refolding activity of ClpB [95].

1.2.2.2. Proteases

Proteases have an important role in the control of protein quality, because by degrading misfolded polypeptides they guarantee that abnormal species do not accumulate in the cell, which in turn allows for amino acid recycling. Targets for

degradation include truncated polypeptides, kinetically trapped folding intermediates which are sensitive to proteolysis and partially folded protein species that after many folding attempts have still failed to reach their native conformation [84]. In the *E. coli* cytoplasm, Lon and ClpP are the two main proteases [30;153;154].

I.Lon

The homotetrameric serine protease Lon is formed by 87 kDa subunits with three functional domains. Substrate recognition and binding are associated to its N-terminus, while central and C-terminus domains are linked to ATPase and proteolytic activities, respectively [84]. Lon is responsible for bulk protein degradation[155;156], and it also has a regulatory function associated to proteolysis of proteins which are designed to be unstable (e.g. SulA).

II. ClpP

Together with Lon, the protease ClpP is believed to be responsible for the degradation of abnormal proteins [155]. Although it also intervenes in bulk degradation of folded and misfolded polypeptides, ClpP is specifically in charge of truncated proteins which have been tagged for degradation [157]. ClpP is structured as two stacked heptamers of 23 kDa subunits, and forms a complex with ClpA and ClpX, two members of the Hsp100 family of ATPases [158-160]. Only when complexed to ClpA and ClpX is the degrading system fully-competent, as ClpP alone can digest small peptides but not large ones or proteins [99]. ClpA and ClpX flank the rings of ClpP and act as molecular chaperones, unfolding proteins in an ATP-dependent manner and translocating them into ClpP central channel [161].

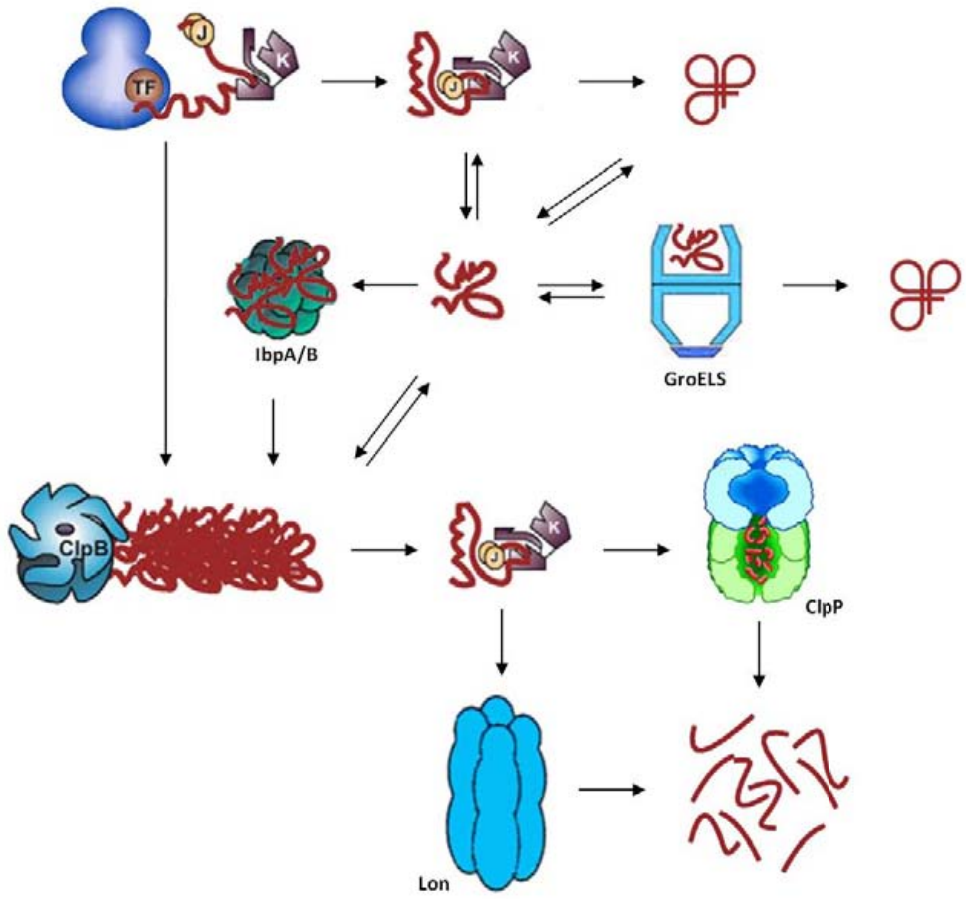

Figure 5. Conventional model of protein folding, aggregation and proteolysis in the cytoplasm of *E. coli*. Newly synthesised polypeptides can fold to their native state, aggregate or be proteolysed, in a process that is tightly regulated by the quality control system.

Adapted from *Nat Biotechnol.* 2004 Nov;22(11):1399-408.

1.2.3. Inclusion bodies

In 1975, Prouty and co-workers described for the first time the formation of amorphous proteinaceous granules in *E. coli* cells growing in the presence of canavanine [162]. These deposits contained abnormal cell proteins and were not surrounded by membranes. Although this was first thought to be an irrelevant cell response in non-physiological conditions, it turned out to be a common feature in recombinant cells used as factories for protein production [13] and protein deposition in the form of insoluble deposits, known as inclusion bodies, is still today a major roadblock in the recovery of soluble and functional recombinant proteins.

Under the non-physiological conditions induced by overexpression of recombinant proteins, the amount of available chaperones in producing cells becomes a limiting factor [62;163;164]. Intermolecular contacts of exposed hydrophobic stretches in the unfolded polypeptides are then favoured because of the high yields of recombinant protein and the limited availability of folding modulators. This situation results in deposition of folding intermediates [165], especially if they are resistant to proteolysis [166], leading to aggregation. Bacteria are well prepared genetically to respond to adverse natural conditions, such as mild protein denaturation under high temperatures [88;167]. However, despite the many cell responses triggered during recombinant protein production, no natural mechanism which favours protein folding has been found [168-175]. Even though, some heat-shock genes including chaperones and proteases are upregulated in response to recombinant stress [91;176-179], but still this response is not enough to prevent inclusion body formation.

From a biotechnological point of view, inclusion bodies have been regarded as a parameter to control in bacterial cell factories [180]. Because aggregation as IBs is not associated to particular protein sequences [181] predicting yield or solubility for a new protein production process becomes an obstacle. Therefore, recombinant protein production in bacteria remains a trial-and-error process.

1.2.3.1. Morphology, composition and structure

Inclusion bodies are insoluble protein deposits observed as cylindrical or ovoid refractile particles of up to 2 μm^3 under an optical microscope [182] and as electron-dense aggregates lacking a defined structure by transmission electron microscopy [183;184]. Usually, one or two inclusion bodies are formed per cell [185] and generally localise in the bacterial cytoplasm, although secreted proteins can also aggregate in the periplasmic space [186]. The surface topology of inclusion bodies can vary from rough to smooth [183], and they present a porous architecture [187] and high level of hydration which are in agreement with density data [188].

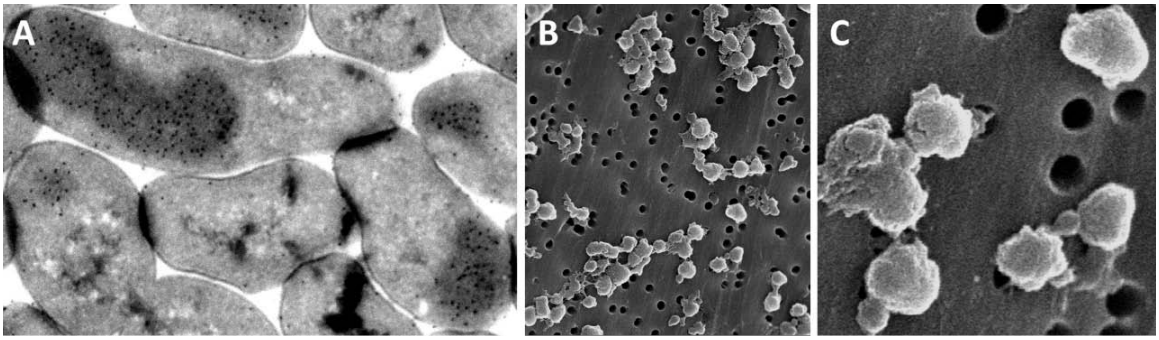

Figure 6. A) Transmission electron microscopy micrograph of an *Escherichia coli* strain producing inclusion bodies. **B)** and **C)** Scanning electron microscopy micrographs of purified inclusion bodies. (García-Fruitós *et al*, not published).

Generally, the major component of inclusion bodies is the target recombinant protein itself, which can account for 50 to 90% of the insoluble protein [189]. However, other cell components can be found associated to inclusion bodies, either adsorbed or entrapped in their structure. For instance, lipids, nucleic acids, lipopolysaccharides and outer membrane proteins can coprecipitate with inclusion bodies during sedimentation by centrifugation [183], although they are not integral components. Membrane proteins can be removed from inclusion bodies by detergent washing and other procedures that do not unfold proteins but solubilise membrane proteins [190;191]. Detergents, EDTA, and enzymes to degrade DNA or the bacterial cell wall are also used in washing procedures [13;192-194]. Truncated versions of the target protein and other plasmid-derived proteins (e.g. those conferring antibiotic resistance) can also be found within inclusion bodies [163;179;191;195-197].

Heat-shock proteins have also been found associated to inclusion bodies. DnaK is localised in the surface of inclusion bodies [184], and can be recovered during sucrose density centrifugation together with ClpB [198]. GroEL is also found in small amounts inside the aggregates, but absent from their surface [184]. In addition, inclusion body proteins IbpA and IbpB received their names after being described as IB components of unknown function [91].

Aggregation has long been regarded as an unspecific process driven by random interaction of exposed hydrophobic patches, resulting in aggregates with no specific internal molecular architecture. However, there is now an increasing body of evidence against this view [199-205], which pictures inclusion bodies as highly ordered structures. Fourier-Transform Infra-Red (FTIR) analysis reveals a characteristic formation of new β-

sheet structures [32;200;206;207] at expenses of α-helices [65;204], even in rich-β-sheet native proteins [203;208]. This newly formed β-sheet is non-native, creating a tightly packed extended intermolecular β-sheet conformation [65].

Remarkably, this enrichment in β-sheet structures is one of the features that inclusion bodies share with amyloid fibril formation [32;200;204;209] together with structural homogeneity [32;65;200;201;208], amyloid-tropic dye binding [200] and cytotoxicity linked to amyloid-like structures [206]. Moreover, for amyloid fibrils sequence determinants act as "hot spots" for aggregation, modulating the specific nucleation of amyloid proteins [210-213]. In the case of inclusion bodies, several observations support the high specifity of their formation process. Besides being essentially composed of the recombinant protein [182;209;214], their presence in reduced numbers [182] suggests their formation could be driven by the growth of a small number of founder aggregates acting as nucleation cores. This is supported by several observations. First, *in vitro* refolding studies of proteins in complex mixtures have shown specificity in polypeptide association during aggregation [215]. Second, folding intermediates of different IB-forming proteins tend to self-associate *in vitro* instead of coaggregating [199]. Third, coexpression of two proteins encoded in the same gene leads to the formation of two types of cytoplasmic aggregates, showing the selectivity of the process [191]. Furthermore, preformed inclusion bodies can act as seeding nuclei for aggregation of their soluble counterparts, but not of unrelated proteins, in a dose-dependent manner [200].

The increase in non-native β-sheet structures does not necessarily involve the full unfolding of the IB-embedded protein. Actually, native-like structure of soluble and inclusion body versions of several proteins has been shown to be highly similar. These include IL-2 [203], β-lactamase [216], *Pseudomonas fragi* lipase [201], human growth hormone and interferon-alpha-2b [202], recombinant *E. coli* β-galactosidase [209], and fluorescent proteins [208;217]. The presence of native-like structure in inclusion bodies seems to facilitate solubilisation of the embedded proteins. In this line, human granulocyte-colony stimulating factor (hGCSF) produced in *E. coli* at low temperatures forms "non classical" inclusion bodies which contain high amounts of correctly folded protein, enabling protein extraction from these IBs using non denaturing conditions and low concentrations of polar solvents [218].

1.2.3.2. Minimising inclusion body formation

Inclusion body formation has affected the development of biotechnology, because even when inclusion bodies are a rich source of protein, the refolding processes required to recover the protein in a native form are complex and expensive [219]. For this reason, much effort has been made to minimise or prevent inclusion body formation, aiming to improve the yield of soluble protein.

Because recombinant protein can account up to around 30% of the total cell protein and this produces an enormous metabolic load on the *E. coli* expression machinery [28], many strategies have been devised to minimise aggregation, either based on a tight control of the cellular milieu or in favouring protein folding.

Besides the use of genetically engineered strains that favour production of soluble protein, (which has already been discussed in section 1.2) other factors can be considered to increase protein solubility. For instance, the composition of growth media affects the levels of soluble protein, and by optimising media composition it has been possible to reduce expression times, increase soluble fraction yield and enhance biological activity of human PDE-3A, PDE-5A and p38-α Map kinase enzymes [28;220]. Moreover, certain proteins can require the presence of specific cofactors in the growth media to fold properly, which can include metal ions (e.g., iron-sulphur) or polypeptide-cofactors (e.g., flavin-mononucleotide). Thus, addition of these factors to the growth media can improve both protein solubility and folding rates [221;222].

Another common strategy consists of lowering the growth temperature of the culture. Protein expression at temperatures below the optimal of 37 °C for *E. coli* growth usually leads to increased stability and correct folding because the hydrophobic interactions that determine inclusion body formation are temperature dependent [223;224]. This has resulted in a number of proteins being successfully expressed in a soluble form in *E. coli* [208;225;226]. Moreover, a number of chaperones show increased expression at low temperatures, which results in better protein quality under these conditions [227]. In addition, reduced degradation of recombinant protein has been observed within a temperature range of 15-23 °C due to poor activity of some of the heat shock proteases [228;229]. However, reduced yields and poor turnover of the recombinant protein are frequent disadvantages when using this strategy because low temperatures result in reduced transcription and translation rates.

Coproduction of folding modulators has been a widely used strategy aimed to overcome limited chaperone availability during recombinant protein expression, but the obtained results are controversial and inconsistent [83;230;231]. Some of the positive reports required coproduction of the major cytosolic chaperone systems (DnaK-DnaJ-GrpE or GroELS) to observe any increase in solubility [113;232-237] or even combinations of them, the most successful being KJE, ClpB and ELS [75]. Although the best results have been obtained when coexpressing several sets of folding modulators, determining the best set of chaperones for a certain target protein is still a trial and error process.

Another common approach consists of metabolic engineering through fusion protein technology, which usually leads to soluble expression [28]. "Tags" consist of proteins or peptides which are fused to the target protein and help to the proper folding of their fusion partners, thereby leading to enhanced solubility [238]. Tags are also convenient for affinity purification, and they can also be expression reporters or provide added advantages, such as protection from proteolysis. The successful use of small peptides (<30 amino acids) called SET tags [239] is promising because their small size may lead to less folding interference making the protein suitable for structural studies without needing to remove the tag, which sometimes results in loss of solubility. Nevertheless, if tags need to be removed, linking the target protein to its fusion partner through a protease-specific recognition sequence will provide an easy separation method by cleavage with the specific protease. For this purpose, TEV protease from tobacco etch virus is often used because of its high specificity and ease of production [240;241].

1.2.3.3. Conformational quality of inclusion body proteins

Ever since recombinant DNA technology was implemented, biotechnological processes have focused on maximising protein solubility [84] often disregarding conformational quality or assuming it to be linked to solubility [242]. However, an increasing number of studies report the existence of different conformational states of proteins trapped in inclusion bodies, many of them being at least partially active.

Back in 1989, Worrall and Goss reported specific activity in inclusion bodies formed by *E. coli* β-galactosidase [243]. Soon after, Tokatlidis and co-workers showed highly active inclusion bodies formed by *Clostridium thermocellum* endoglucanase D [244].

Later on, structural data presented by Oberg and co-workers described the existence of native-like secondary structure present in inclusion bodies [203].

More recently, data from our group showed that biological activity is also retained in fluorescent proteins, which remain highly fluorescent even when trapped in inclusion bodies [217]. Moreover, active inclusion bodies have also been found in the periplasm [245].

The presence of active polypeptides as structural components of inclusion bodies suggests that solubility and functionality are not necessarily linked. In fact, the presence of aggregates has also been reported in the soluble cell fraction [198]. On this background, we decided to further explore the scenario of recombinant protein production and test the coincidence of solubility and activity as indicators of conformational quality.

1.3. The baculovirus-insect cell expression system

Baculovirus-mediated expression of foreign genes emerged in the early 1980s as a promising system which seemed capable of providing both the high yields obtained in bacteria and the eukaryotic post-translational modifications provided by mammalian systems. Although these expectations turned out to be not completely realistic, important technological advances over the past 20 years have overcome the main drawbacks of the system, which is increasingly popular for recombinant protein production.

Baculoviruses are a large group of dsDNA viruses that infect arthropods, mainly insects. Their host range is very limited, and often restricted to just one species. However, *Autographa californica* multicapsid nucleopolyhedrosis virus (*Ac*MNPV) has a broader host range, being able to infect around 25 lepidopteran insects [246]. *Ac*MNPV is the most studied and exploited member of the Baculoviridae family, and was used to develop the first expression vectors [247;248]. Indeed, the backbone of most of the vectors available today is still based on its genome.

A key feature of baculoviruses enabled their development as vectors for recombinant protein production. Late in the infection cycle, progeny virions are coated with a protective matrix formed of a virus-encoded protein called polyhedrin, which is produced in very large amounts reaching up to 30-50% of the total cellular protein at the end of the baculovirus life cycle [246;249]. However, polyhedrin is not essential in cell culture, as it is not required for virus replication in cultured insect cells [250]. For this reason, it can be replaced by the gene of interest to obtain very high levels of the target protein. Indeed, this is one of the main advantages of the baculovirus system, with yields as high as≥ 100 mg of the target protein per litre of infected cells [246]. Moreover, in contrast to bacterial systems, the formation of inclusion bodies is rarely observed.

Eukaryotic protein processing capabilities are another important advantage of the baculovirus system. However, these pathways are not identical to those of higher eukaryotes, and also baculovirus infection can have an unfavourable effect on the processing functions of the infected host [251;252].

The baculovirus system is also a powerful tool to obtain multiprotein subunit complexes [253]. Production of virus-like particles which can be used as immunogens [254] is a clear example of its important applications.

Besides the baculovirus, the system has another essential component which is of course the host. Lepidopteran cell lines are the most frequent hosts, although alternatively an insect host can be used. In both cases, *Spodoptera frugiperda* and *Trichoplusia ni* are the most common hosts [249].

1.3.1. Overview of baculovirus biology

This section will focus on the main features of the virus structure and life cycle that will provide the basis for comprehending the principles of the baculovirus expression system.

1.3.1.1. Baculovirus structure

Baculoviridae is a diverse group of double-stranded circular DNA genomes [255], between 80-200 kbp long [256]. These viruses get their name from their rod-shape morphology (*baculum* meaning "stick" in Latin). Virus capsids are usually 40-50 nm in diameter and 200-400 nm in length [257]. For viruses carrying larger DNA genomes, as can be the case with recombinant viruses, the capsid length can extend to accommodate the insert [258]. Also, virions have polarity because the ends of the capsids are structurally different [258]. The two commonly used baculoviruses for expression vectors, *Autographa californica* multicapsid nucleopolyhedrovirus (*Ac*MNPV) and *Bombyx mori* nucleopolyhedrovirus (*Bm*NPV), both have genomes of approximately 130 kpb.

Nucleocapsids are synthesised in the nucleus of infected cells and acquire a membrane envelope either budding through the plasma membrane, forming the extracellular or *budded* virus, or within the cell nucleus. Nucleocapsids that are enveloped in the nucleus are also occluded within a crystalline protein matrix, forming the *occluded* virus. Viral *occlusion bodies* (also called *polyhedra* because of their shape) are formed in the nucleus as well, and consist of one or more enveloped nucleocapsids embedded in a crystalline protein matrix [259], which is polyhedrin in the case of nucleopolyhedroviruses (NPV). Depending on the number of nucleocapsids contained in the occlusion bodies, NPV can be divided into single (SNPV) or multiple (MNPV). Occlusion bodies also have an outer coat called *calyx*, which is thought to increase their stability [249].

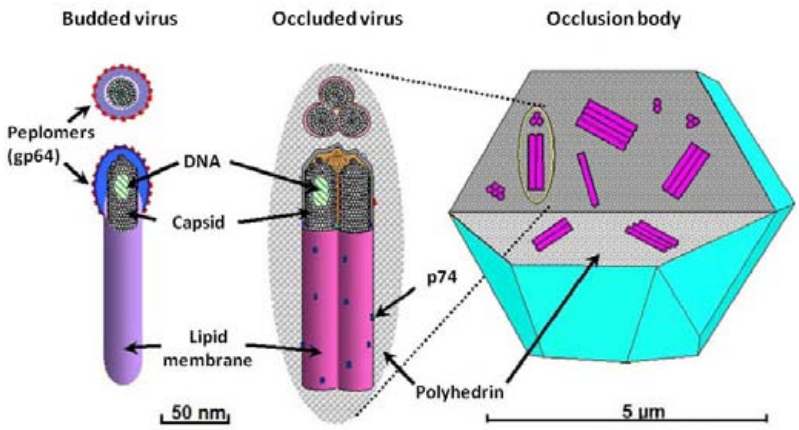

Figure 7. Structure of the different forms of multinucleopolyhedroviruses throughout their life cycle.

Adapted from Wikipedia.

Although nucleocapsids are thought to be identical in both budded and occluded viruses, their membranes are biochemically different. Budded viruses have projections in one end of their structure, called *peplomers*, that contain the glycoprotein gp64 which is absent in occluded viruses. Protein gp64 is involved in virus entry into cells by endocytosis during secondary infection [260], while enveloped viruses liberated from occlusion bodies enter cells by a different route [261]. Also, the O-glycosylated protein gp41 and protein p74 are present in occluded virus but not in the budded form.

A second type of occluded baculovirus exists in the baculovirus family. These are called the granulosis viruses (GV), and in contrast to NPV they have only a single virion embedded in a very small occlusion body. In this case, the matrix protein is granulin. Moreover, some baculoviruses do not synthesise an occluded form, and are consequently name *nonoccluded* baculoviruses.

1.3.1.2. Infection progress

Infection in the insect has two distinct phases. Primary infection is caused when larvae ingest polyhedra as contaminants of their food. Upon arrival to the insect midgut, polyhedra are dissolved in the alkaline environment and release the embedded virions [262], which enter midgut cells after fusing to the membrane of the microvilli [263]. This takes places during the early phase of infection, when cells are reprogrammed for virus replication.

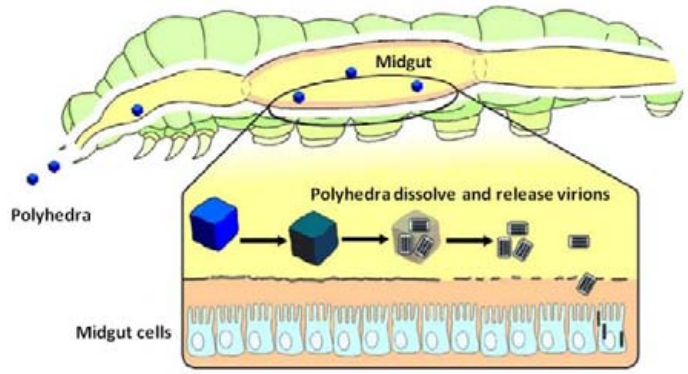

Figure 8. Baculovirus infection of an insect host.

Adapted from http://www.microbiologybytes.com /virology/kalmakoff/baculo/baculo.html

Nucleocapsids can then be transported to the nucleus, where they replicate, or to the basal side of the cells for rapid budding [264]. During the secondary phase of the infection both budded viruses and polyhedra are produced. The late phase of infection is characterised by extensive DNA replication and release of budded virus [249]. Released virions reach the hemocoel and are transported via the hemolymph to other tissues, causing a systemic infection [249].

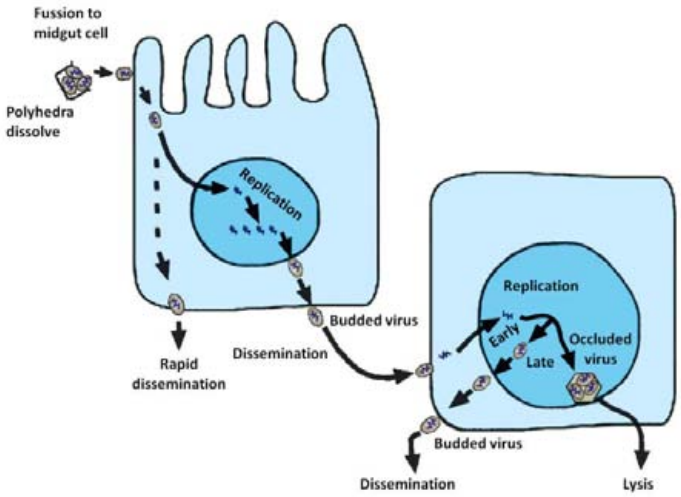

Figure 9. Phases of baculovirus infection.

Adapted from http://www.microbiologybytes.com /virology/kalmakoff/baculo/baculo.html

The very late phase of infection is characterised by hyperexpression of polyhedrin and P10 [263]. During this phase polyhedra accumulate in the nucleus and the production of budded virus is greatly reduced, if not terminated [265]. By the end of the infection larvae liquefy due to extensive cell lysis, in which P10 protein is involved [266;267]. The insect literally melts, becoming a sac of milky fluid containing polyhedra which are released to the environment upon cuticle breakage. Because polyhedra are relatively stable in the environment, they can reinitiate the infection cycle when consumed by a new host.

1.3.2. Expression vectors

The classic baculovirus expression vector consists of a recombinant baculovirus genome which contains a foreign nucleic acid sequence encoding the target protein under the control of a polyhedrin promoter. The heterologous gene is generally placed in the polyhedrin locus of the viral genome, replacing the wild-type polyhedrin. This recombinant baculovirus can be used to infect cultured insect cells or larvae, yielding high transcription levels during the very late phase of infection, which is usually translated to high levels of recombinant protein production.

Because baculovirus genomes are large, they usually contain one or more recognition sites for restriction endonucleases. By the time that these first baculovirus vectors were being developed no known restriction enzymes that lacked recognition sites in the genome had been described, so homologous recombination was the chosen method to insert the foreign genes into the baculovirus genome [247;248]. This method involved the construction of a "transfer" plasmid containing the heterologous gene flanked by baculoviral sequences homologous to the polyhedrin locus, which would then be cotransfected into cultured cells together with purified genomic DNA of wild-type *Ac*MNPV. However, the process was highly inefficient because a double crossover recombination was necessary to knock out the polyhedrin gene while knocking-in the gene encoding the target protein, so only about 0.1% recombinants were obtained [250]. Plaque assays were required to isolate the small amount of recombinant baculoviruses from the large parental background, and then visual screening for the polyhedron-negative phenotype allowed for identification of recombinant virus. However, this was a critical step constraining the use of the system, as identifying the recombinants could be a difficult task.

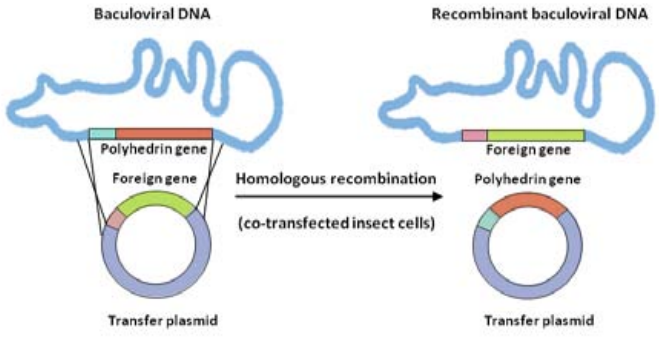

Figure 10. Baculovirus expression vector obtained by homologous recombination.

Adapted from *Methods Enzymol.* 2009;463:191-222.

To overcome these technical limitations, and also to improve the system in other ways, many modifications have been developed over the years, involving both the parental genomes and the transfer plasmids.

1.3.2.1. Transfer plasmids

Transfer plasmids are used to transfer the foreign gene into the viral genome by means of homologous recombination [249]. A typical transfer plasmid contains the gene of interest under control of a baculovirus promoter (which is often polyhedrin) and flanked by sufficient amount of viral DNA to allow recombination.

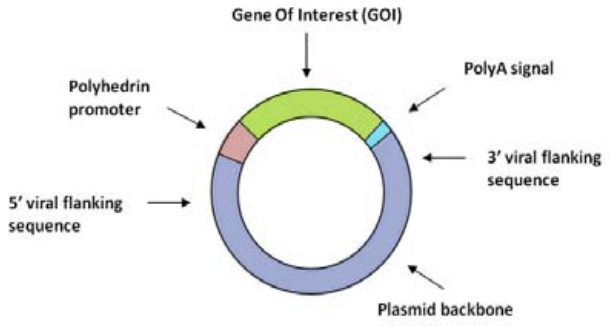

Figure 11. Baculovirus transfer plasmid. Adapted from *Methods Enzymol*. 2009;463:191-222.

Several factors must be considered when cloning the gene of interest into the transfer plasmid. First, it is important to use genes without introns because although low levels of splicing have been reported [268] strong protein expression has not been observed from spliced mRNAs. Also, the AUG context is important for initiation of translation [249]. The AUG contexts for several promoters are shown in Table 2.

Table 2. AUG contexts of highly expressed *Ac*MNPV proteins.

Gene	Transcription	AUG context	Reference
polh	Very late	CCUAUAAAU**AUG**CCGG	[269]
p10	Very late	UUUACAAUC**AUG**UCAA	[270]
p6.9	Late	AAUUUAAAC**AUG**GUUU	[271]
vp39	Late	GGCAACAAU**AUG**GCGC	[272]
Consensus		A Y**AUG** Y	

Adapted from *Baculovirus Expression Vectors. A Laboratory Manual*. Oxford University Press, 1994.

The choice of promoter is also important, as it will determine the production levels of the target protein. Polyhedrin is a common strong promoter which is equally efficient in either orientation with respect to the *Ac*MNPV genome [273]. Moreover, two polyhedrin promoters or two very late promoters (e.g., *polh* and *p10*) work at almost optimal efficiency when placed back-to-back to drive expression of two different genes [274-276]. The *p10* promoter has a similar strength to *polh*. Both are very late promoters with essential TAAG sequences at their transcriptional start point, and the region from around the initiation point to the ATG is sufficient to promote high transcription levels [249;275]. Although these promoters are very effective, a decline in the level of post-translational modifications at very late times post infection has been reported [277;278]. Thus, the use of late promoters such as vp39 or p6.9 may present an advantage when uniformity of post-translational modifications is important because the proteins will have additional hours to move through the endoplasmic reticulum and Golgi apparatus, in spite of the lower yields that will be obtained with the use of such promoters [249]. Early promoters such as *ie1* have also been used, and although these promoters drive lower levels of transcription they seem to promote higher quality products, being especially useful for secreted proteins [279-284].

Polyadenylation signals are also required for viral transcripts to be processed at their 3' end [249]. The polyadenylation signal for polyhedrin is located in the downstream *orf*1629 [285], and because this is an essential gene *polh* deletions in polyhedrin-based transfer plasmids do not extend into this *orf*. Therefore, the *polh* polyadenylation site is maintained in the viral genome, so it will not be necessary to include a polyadenylation signal in the transfer plasmid. For transfer plasmids supplying back-to-back promoters in the polyhedrin region, transcripts extending in opposite direction to the wild-type polyhedrin gene transcription are expected to terminate at a polyadenylation signal at the 3' end of the flanking *orf*603 [286;287]. For *p10*-based vectors, a polyadenylation signal is located downstream the stop codon, and this is usually included in the available transfer plasmids [249].

A bias in codon usage has been described for highly expressed baculovirus genes, such as polyhedrin or p10. However, heterologous genes with rare codons can be well expressed in the baculovirus system, as reported for *E. coli* β-galactosidase [247]. Nonetheless, the UAA codon is preferred for termination, as it is used for most AcMNPV genes [249].

The main objective of modifying transfer plasmids was to facilitate identification of recombinant baculovirus plaques by visual screening. For that purpose, marker genes such as *E. coli* β-galactosidase were introduced under the control of baculovirus promoters [288]. However, this could be a trap because the presence of the marker gene could indicate a single crossover homologous recombination, which produces recombinant baculoviruses containing the entire transfer plasmid and thus being genetically unstable. For this reason, further screening would be required to map the position of the foreign gene in the baculovirus genome and confirm that a double crossover recombination event had taken place.

A second modification of transfer plasmids was aimed at facilitating expression and purification of the recombinant protein. This included addition of sequences such as secretory signal peptides or purification tags, as well as replacing the polyhedrin promoter with alternate baculovirus promoters or multiple promoter elements that would allow coexpression of multiple recombinant proteins in the same cell during infection [289], as discussed.

1.3.2.2. Parental genomes

Modifications in the parental genomes have been addressed to solve technical problems related to isolation of recombinant viruses and to enhance the production of the target protein.

I. Enhancing recombination efficiency

The first important step towards increasing the efficiency of recombination was made with the development of a baculovirus with a linearisable DNA genome [290]. This was achieved by introducing a unique *Bsu*36I restriction site in the polyhedrin locus. Linearising the parental DNA prevented its replication, which reduced the number of parental virus after recombination. Homologous recombination was still possible, and indeed restored the ability of the baculovirus vector to replicate. This approach increased the efficiency of baculovirus vector production up to 10-20%.

The next improvement was the development of BakPAK6^{TM}, a recombinant baculovirus that could be gapped with *Bsu*36I deleting a portion of *orf*1629, which

encodes an essential phosphoprotein of the viral nucleocapsid [291], and that also included an *E. coli lacZ* gene which allowed for easy detection of recombinants as the white plaques on a blue background [292]. In this case, recombinant baculovirus production increased to about 95%. This was commercialised by ClonTech.

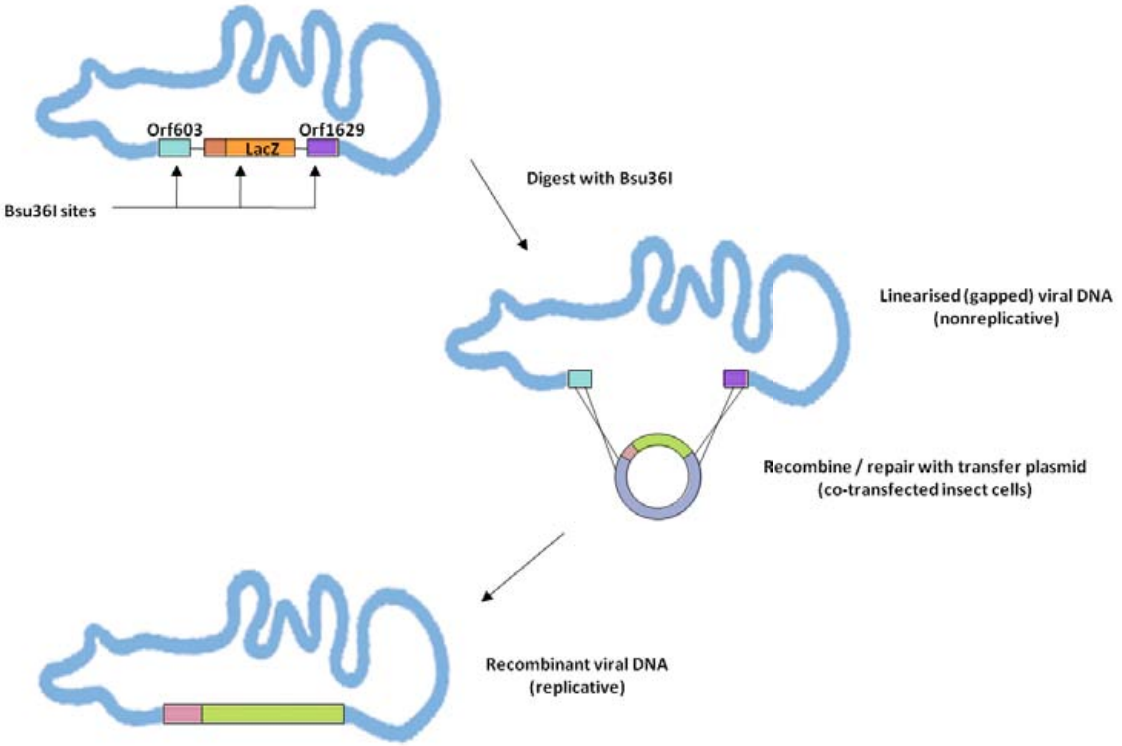

Figure 12. Baculovirus expression vector obtained by homologous recombination with a linearised/gapped parental viral genome.

Adapted from *Methods Enzymol*. 2009;463:191-222.

In parallel to linearisable genomes, another approach based on genetic transposition was developed [293]. Key to this method was the creation of a new *E. coli* strain that contained an autonomously replicating bacmid which included a copy of the entire baculovirus genome and a helper plasmid encoding a transposase. The bacmid contained an *E. coli lacZ* gene and a "mini-Att Tn7" site, an attachment site used during transposition. The transfer plamid contained the target gene flanked by the ends of Tn7, and thus could be transposed to the polyhedrin locus of the bacmid when transformed into the bacteria. The *lacZ* gene would be knocked out of the bacmid upon transposition, and the recombinants could be selected by standard blue-white screening. This system is

commercialised as Bac-to-Bac™ by Invitrogen, and provides a 100% efficiency of recombinant baculovirus production. However, recombinant viruses are genetically unstable upon passage in insect cells, seemingly because they retain the bacterial replicon [294].

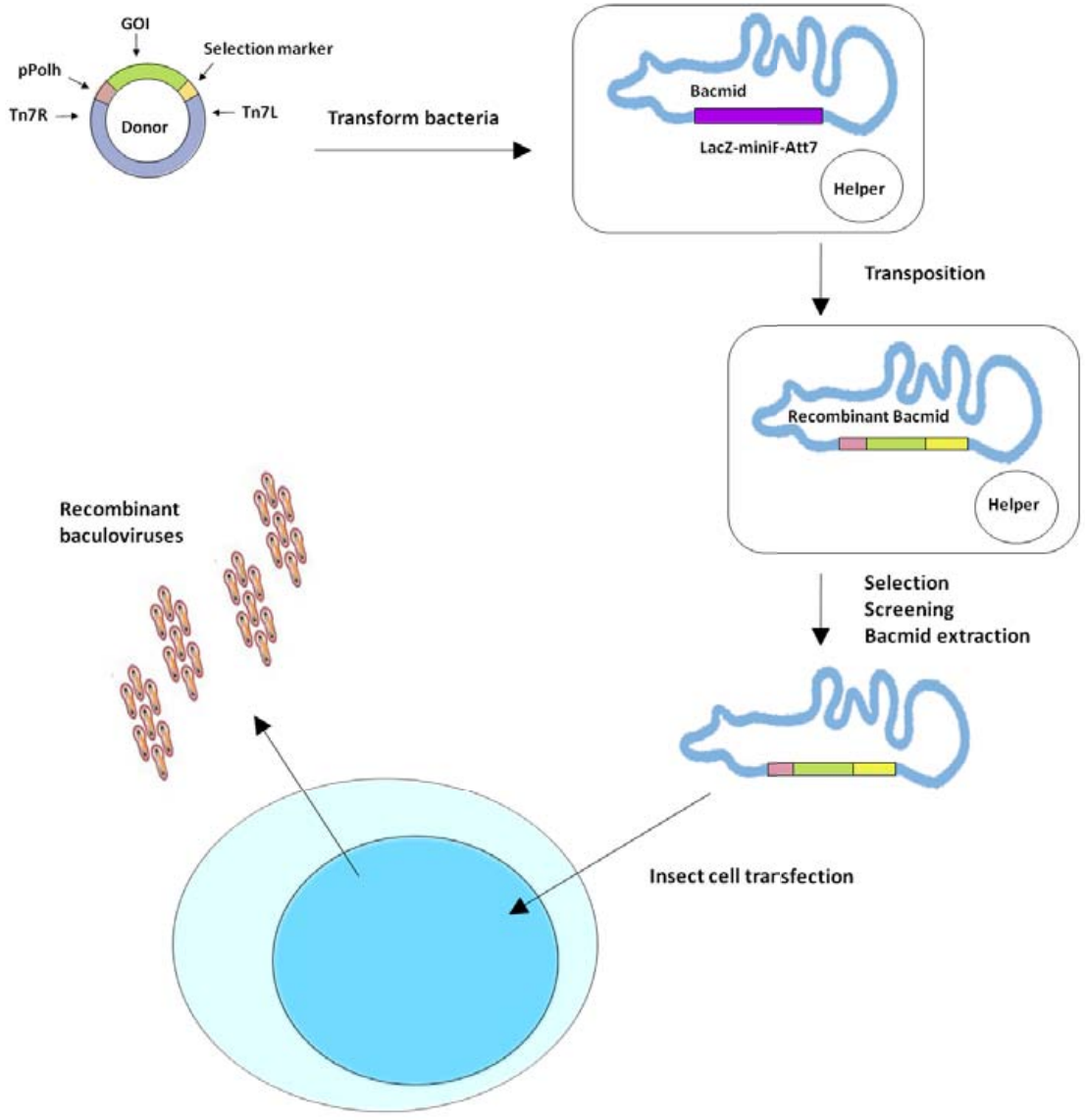

Figure 13. Baculovirus expression vector obtained by transposition.

Adapted from *Methods Enzymol*. 2009;463:191-222.

Recently, a new method consisting of cross-hybridising the linearisable baculoviral DNA and bacmid strategies has been developed [295]. This approach relies on a bacmid that contains a recombinant baculoviral genome with a bacterial replicon in the

polyhedrin locus and a deletion in the *orf*1629 gene. The bacmid can replicate in *E. coli* but not in insect cells, so it can easily be produced in *E. coli* and used to cotransfect insect cells together with the transfer plasmid. Homologous recombination restores the *orf*1629 deletion, knocking-in the gene of interest and at the same time knocking out the bacterial replicon. This is marketed under the name of *flashBAC*™ by Oxford Expression Techonologies, and yields very high levels of recombinant baculovirus production. However, despite the bacmid not being able to replicate in insect cells, progeny derived from the defective parental viral genome can be obtained by genetic complementation when the *orf*1629 product is provided in *trans* by the recombinant virus. For this reason, plaque assay is still recommended to purify the recombinant virus.

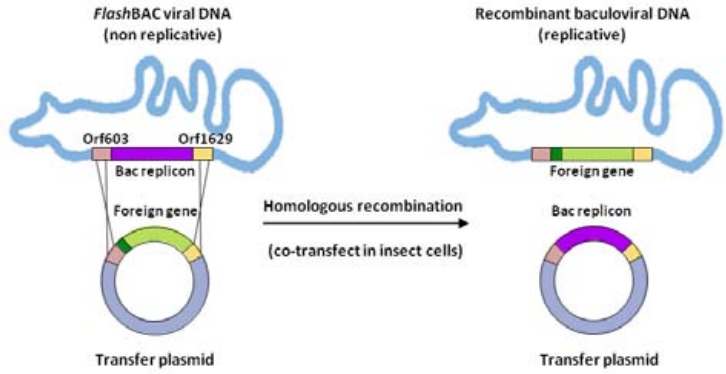

Figure 14. Baculovirus expression vector obtained using the *flash*BAC method.

Adapted from *Methods Enzymol*. 2009;463:191-222.

Baculovirus vectors can also be produced *in vitro* by site-specific recombination [246]. A prelinearised virus genome contains an *E. coli lacZ* gene and a herpes simplex virus thymidine kinase gene flanked by site-specific recombination sites from bacteriophage lambda (attR1 and attR2) replacing the polyhedrin coding sequence. The target gene is encoded in an entry plasmid, flanked by recombination sites attL1 and attL2. Genome and plasmid are mixed *in vitro* in presence of a purified recombinase to obtain the recombinant baculovirus. The mixture is transfected into insect cells, which are cultured in presence of gancyclovir to select against replication of parental viral DNA. This is commercialised by Invitrogen as BaculoDirect™.

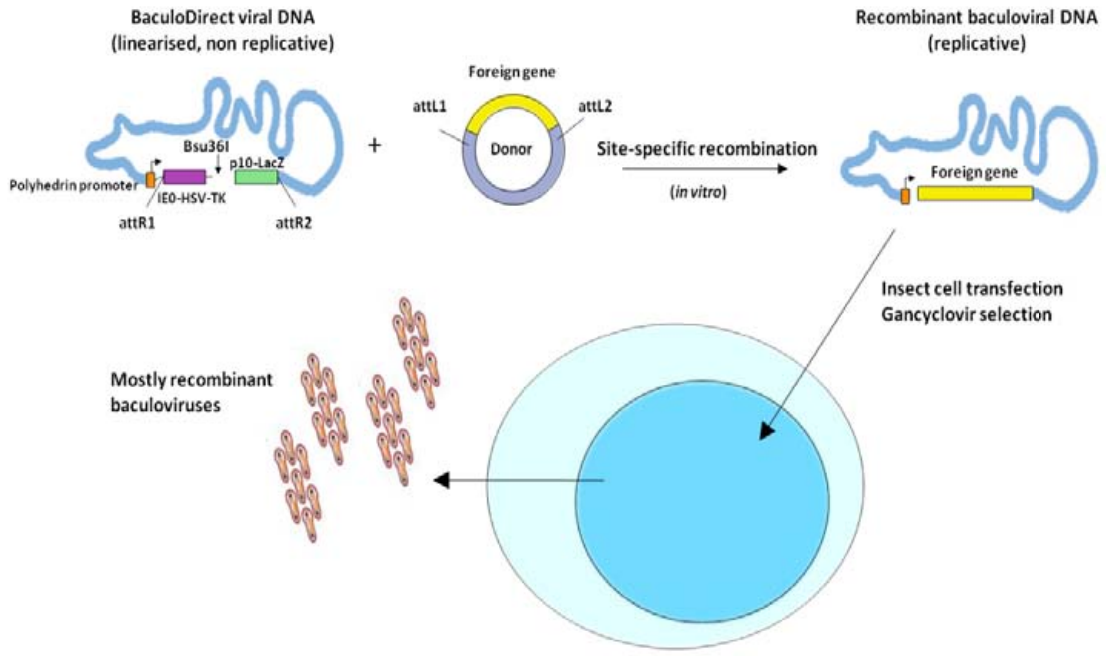

Figure 15. Baculovirus expression vector obtained using BaculoDirect.

Adapted from *Methods Enzymol*. 2009;463:191-222.

II. Improving protein production

The general approach used to improve protein production in the baculovirus system consists of deleting nonessential genes that are thought to interfere with heterologous protein production or to degrade the target protein. However, addition of new heterologous genes to the baculovirus genome has also been reported.

Chitinase [296] and cathepsin-like protease [297] have been deleted in several commercial vectors, and although the impact of these deletions is not totally clear less degradation of foreign glycoproteins has been shown [298]. Chitinase is a resident endoplasmic reticulum protein [299] thought to interfere with protein secretion by saturation of the host translocation machinery [298]. Thus, deletion of the chitinase gene is expected to increase the yields of secreted proteins.

Parental baculovirus DNA lacking a functional *p10* gene is also available commercially under the name of DiamondBac^{TM}. As *p10* is involved in cell lysis [266], it is expected that infected cells will retain higher viabilities throughout the course of infection. Moreover, in DiamondBac the *p10* gene has been replaced by a protein disulfide

isomerase (PDI), a chaperone that drives disulfide bridge formation, thus increasing solubility and secretion of the target protein [300].

Other baculovirus vectors encoding heterologous protein processing enzymes have also been described. Polydnavirus *vankyrin* gene under control of the p10 promoter [301] has been found to prolong the viability of *Sf9* cells infected with baculovirus, which can thereby enhance the production of the target protein. In this line, heterologous glycosyltransferases [302;303] or enzymes involved in CMP-sialic acid biosynthesis [304] under the control of baculovirus *ie1* promoters have been used to expand the processing capabilities of the baculovirus system.

Production of multi-subunit complexes has also been addressed. Although transfer plasmids allow coproduction of several proteins, the number of genes that can be inserted in the plasmid is limited by its size. Moreover, the use of repeated sequences such as promoters or terminators can result in recombination events [294;305]. As a solution, a new system allowing each protein in the complex to be expressed from different loci has been developed [18]. Because single gene insertions are distributed along the genome, these problems are overcome. Furthermore, the system is based on lambda red recombination [306], which allows fast generation of recombinants in *E. coli*.

1.3.3. Insect hosts

Insect hosts constitute the second half of the baculovirus system. Lepidopteran insects are hosts for many viruses from the *Baculoviridae* family, including *Ac*MNPV. Although cell lines are the most frequent choice at laboratory scale, insect larvae provide an interesting alternative to cell culture scale-up for producing large amounts of recombinant protein, with the added advantage of reduced production costs.

1.3.3.1. Cell lines

The first established lepidopteran cell lines were described by Grace in 1962 [307], and so far over 250 insect cell lines have been described [308]. Two of the most common cells used with *Ac*MNPV vectors are *Sf9* and *Sf21* cell lines, both originated from IPLB-SF-21 cells derived from pupal ovarian tissue from the fall armyworm *Spodoptera frugiperda* [309]. The other common cell line originated from adult ovarian cells of the

cabbage looper *Trichoplusia ni*, which was originally described as BTI Tn 5B-1 [310;311] and is now marketed by Invitrogen as High Five™.

These cell lines can grow in adherent and suspension cultures, and thus can be easily scaled-up in shake flasks, spinner flasks, or bioreactors to obtain large amounts of recombinant proteins [312;313]. Moreover, *Sf9* and *Sf21* cells are also routinely used to plaque purify and quantify recombinant baculovirus vectors.

Insect cell cultures grow at an optimal temperature of 28 °C. Since the cells are loosely adherent neither trypsin nor EDTA is required for subculture. Also, CO_2 incubators are not necessary because insect cell culture media are buffered with phosphate instead of carbonate. Moreover, cells can grow both in media supplemented with serum or in serum-free media, both of which are commercially available.

Currently, transgenic insect cell lines are already in the market. One of the most important modifications has been the introduction of constitutively expressed mammalian genes involved in post-translational processing, with the aim of obtaining partially humanised glycosylations [314-318]. A *Sf9*-derived insect cell line with an extended N-glycosylation pathway is commercialised by Invitrogen under the name of MIMIC™ [317].

Another transgenic *Sf9* cell line derivative contains a polydnavirus *vankyrin* gene expressed constitutively under the control of an immediate-early baculovirus promoter, which enhances the life span of the cells [301]. Three vankyrin-enhanced *Sf9* derivatives are already marketed by ParaTechs.

1.3.3.2. Insect larvae

Although proteins produced in cell culture are easier to purify and usually have more uniform post-translational modifications than those obtained in insect larvae as a result of only one cell type being involved in protein synthesis, the main drawback of scaling-up protein production in cell culture is the cost. As culture media are expensive, the use of large volumes may become prohibitive. Moreover, bioreactors will often be required to handle large culture volumes, which will add to the cost of the production process [249].

Larvae offer the advantage of being cheaper to maintain because they do not require growth media or sterile conditions [319]. However, protein production in larvae will

require feeding and handling living insects. Moreover, protein purification may become more difficult due to the presence of insect parts as contaminating products. Nevertheless, although yields of recombinant proteins produced in insect larvae can be reduced due to protein aggregation [320], larvae can still be regarded as natural bioreactors for recombinant protein production.

In addition to *Spodoptera frugiperda* and *Trichoplusia ni* larvae being used as hosts, *Bombyx mori* larvae are also commonly used, mainly due to the inability of growing large culture volumes of *Bombyx mori* cells [249].

1.4. Model proteins

Several proteins have been used in this study as reporters of protein aggregation and conformational quality. Our main model protein has been a chimeric fusion protein between the aggregation-prone VP1 capsid protein of foot-and-mouth disease virus (FMDV) and the green fluorescent protein (GFP). Nonetheless, for some specific experiments VP1 has also been used without the fluorescence reporter, along with FMDV VP2 and human α-galactosidase.

1.4.1. Green Fluorescent Protein

In 1962, Shimomura reported the existence of a green fluorescent protein in the jellyfish *Aequorea victoria* [321]. In nature, this protein absorbs the blue bioluminescence of its partner protein, the calcium activated aequorin, and converts it to the greenish glow observed in living animals [322]. Although green fluorescent proteins exist in other organisms [323] and recently GFPs from *Renilla mullerei*, *Renilla reniformis* and *Ptilosarcus gurneyi* have been cloned and patented [324], the *Aequorea* GFP gene was the first to be cloned [325] and expressed in heterologous systems [326;327] and thus is the most widely used today. GFP is still fluorescent without the need of jellyfish-specific enzymes or cofactors; therefore, the gene contains all the necessary information for the correct formation of the chromophore.

Wild-type GFP consists of a single chain of 238 amino acid residues which is highly stable and resistant to proteolysis and has two absorption maxima at about 395 and 475 nm, with excitation at the major peak of 395 nm yielding an emission maximum at 508 nm [328]. The structure of GFP is an 11-stranded β-barrel wrapped around a single α-helix where the chromophore attaches, remaining buried in the centre of the cylinder, which has been called a β-can [328;329]. The barrel forms an almost perfect cylinder which is 42 Å long and 24 Å in diameter [328].

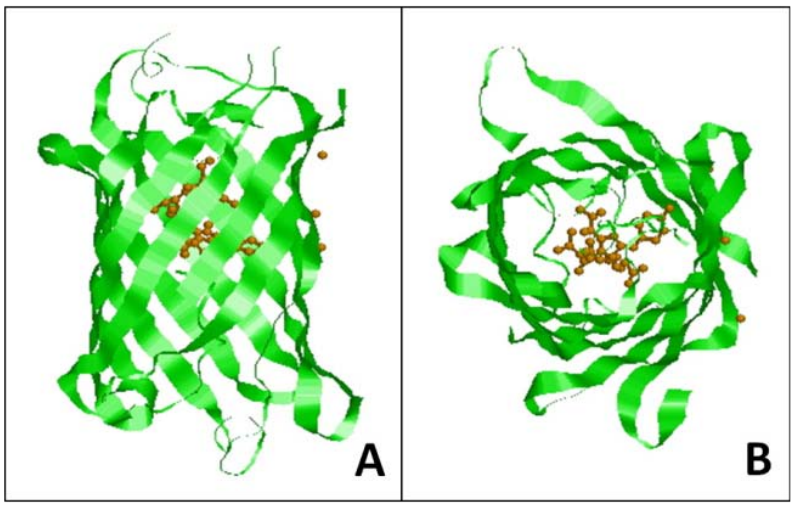

Figure 16. *Aequorea victoria* GFP tridimensional structure. **A)** Frontal view. **B)** Axial view. The chromophore is coloured in orange. (Images exported from a Rasmol representation, PDB file 1EMA).

The chromophore is a *p*-hydroxybenzylideneimidazolinone [325;330] formed from residues 65–67, which are Ser-Tyr-Gly in wild-type GFP. The currently accepted mechanism for chromophore formation is shown in Figure 17. GFP folds nearly into its native conformation before the imidazolinone is formed by a nucleophilic attack of the amide of Gly67 on the carbonyl of Ser65, followed by dehydration. Then, molecular oxygen dehydrogenates the α-β bond of Tyr66 conjugating its aromatic ring with the imidazolinone [331-333]. Since O_2 is required [327;331], GFP is probably not functional in obligate anaerobes.

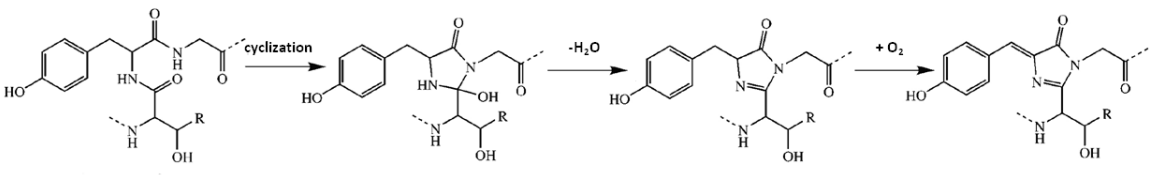

Figure 17. Proposed mechanism for chromophore formation.

Adapted from *Annu Rev Biochem*. 1998;67:509-44.

Although wild-type GFP folds efficiently at room temperature or below, it tends to misfold at higher temperatures. However, this temperature sensitivity is restricted to the folding process, and after GFP has matured correctly at a low temperature it is stable and fluorescent up to 65 °C. In any case, GFP has been optimised for expression at 37 °C.

For that purpose, the most often used mutations are F64L and V163A, which improve the folding efficiency but not GFP brightness [334].

Other mutants have been developed to enhance the properties of GFP. One of the most common consists of a replacement of Ser65 by Thr, or S65T [335]. This mutation changes the excitation spectra of GFP to a single peak that is also shifted to 490 nm, which renders the protein more compatible with standard optical filter sets. Mutations rendering altered emission spectra have also been explored, and some of the most representative are listed in Table 3. Most of these GFP derivatives are resistant to photobleaching [332;336], probably because the fluorophore is well protected.

Table 3. Summary of the most important GFP derivatives.

Class	Protein	Excitation (nm)	Emission (nm)	Oligomerisation
Far-red	mPlum	590	649	Monomer
Red	mCherry	587	610	Monomer
	tdTomato	554	581	Tandem dimer
	mStrawberry	574	596	Monomer
	J-Red	584	610	Dimer
	DsRed-monomer	556	586	Monomer
Orange	mOrange	548	562	Monomer
	mKO	548	559	Monomer
Yellow-green	mCitrine	516	529	Monomer
	Venus	515	528	Weak dimer
	YPet	517	530	Weak dimer
	EYFP	514	527	Weak dimer
Green	Emerald	487	509	Weak dimer
	EGFP	488	507	Weak dimer
Cyan	CyPet	435	477	Weak dimer
	mCFPm	433	475	Monomer
	Cerulean	433	475	Weak dimer
UV-excitable green	T-Sapphire	399	511	Weak dimer

Adapted from *Nat Methods*. 2005 Dec;2(12):905-9.

Some of the applications of GFP include its broad use in cellular biology as a fusion tag to monitor localisation and fate of host proteins in cells [337-340], and also as a cell marker or reporter of gene expression *in vivo* [326]. Moreover, GFP can also be used as an active indicator for protease action [341], or pH [342] and calcium [343] sensitivity.

1.4.2. Foot-and-Mouth Disease Virus VP1 and VP2 capsid proteins

The virion of foot-and-mouth disease virus contains 60 copies each of the four structural proteins forming the capsid. Three of these proteins, VP1, VP2 and VP3, are wedge-shaped, eight-stranded β-sandwiches partially exposed to the surface while VP4 and the N termini of VP1 and VP3 are located at the capsid interior [344].

The surface of the particle is fairly smooth with a major protruding element in VP1, which is called the G-H loop. This loop is highly flexible and comprises about 20 residues around positions 140-160 [345]. The G-H loop also contains a highly conserved Arg-Gly-Asp (RGD) triplet that interacts with integrin receptors in the cell surface, and the major antigenic site [346-348]. For serotype C, different overlapping epitopes have been mapped in the G-H loop [348]. The highly exposed C terminus of VP1 has also been related to both the antigenic and receptor binding properties of the virus [349;350].

Both VP1 and VP2 proteins used in this study are from serotype C (isolate C-S8c1) of the virus.

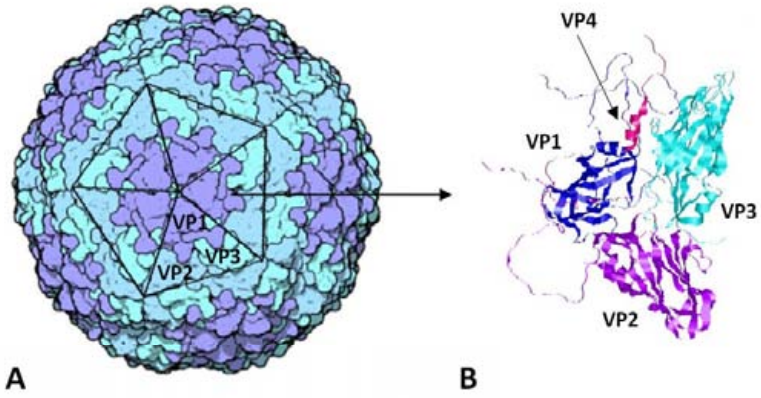

Figure 18. A) FMDV virion structure. The capsid is composed of 12 pentamers with 5 protomers each (adapted from illustration by David S. Goodsell of The Scripps Research Institute) and **B)** Ribbon representation of the capsid proteins forming a protomer. (Image exported from a Rasmol representation, PDB file 1FMD).

1.4.3. Human α-Galactosidase

The enzyme α-galactosidase is responsible of removing galactose from glycosylated macromolecules in the lysosomes [351]. The absence of functional α-galactosidase in humans results in a condition known as Fabry disease, a lysosomal storage disease caused by the accumulation of the enzyme's substrates in the tissues. In addition, some α-galactosidase mutations have been associated to enzyme aggregation in Fabry condition [352].

Human α-galactosidase is a homodimeric glycoprotein with each monomer consisting of two domains, a $(\beta/\alpha)_8$ barrel that contains the active site and a C-terminal domain formed of eight antiparallel β strands on two sheets in a β sandwich [353].

Figure 19. Human α-galactosidase tridimensional structure. (Image exported from a Rasmol representation, PDB file 1R46).

1.5. Previous work

The objectives of this thesis were defined after a first study that involved coproduction of folding modulators with the aim of determining their impact on both protein solubility and conformational quality (see Annex I).

Background to that study was the controversy of results obtained after coproduction of folding modulators, which although expected to increase solubility often resulted in inconsistent results that were attributed to particular requirements for folding modulators or to specific features of particular proteins. Moreover, reports of active inclusion bodies were becoming increasingly frequent. Thus, we decided to explore whether solubility was a good indicator of protein quality, or instead biological activity would be a more suitable parameter.

For that purpose, we coproduced a recombinant GFP together with the DnaKJ chaperone pair and analysed the impact of the folding modulators both on the fractioning and conformational quality of the reporter protein. The experimental approach consisted of determining protein and fluorescence levels in both soluble and insoluble protein fractions that had been produced with or without the chaperone pair and under different gene expression conditions modulated by the inducer concentration. Our results indicated a different impact of the chaperones on protein solubility and quality, as while solubility of the protein was only affected by its own yield, DnaK promoted a quality enhancement at low production levels, which was impaired by a chaperone excess that also resulted in proteolysis of the recombinant protein. Moreover, soluble and insoluble protein populations displayed a coincident quality profile, and the variability observed for the soluble fraction was associated to the existence of oligomers in that population.

Thus, the results of this work prompted us to disregard solubility as a good indicator of protein quality, since these parameters were divergently controlled by DnaK.

2. Objectives

The aim of the study was to explore the patent divergence in the control of protein solubility and conformational quality observed in bacterial cells actively producing recombinant proteins and test whether this principle is also true for eukaryotic systems.

For that purpose, we set the following objectives:

1. Explore conditions that can enhance simultaneously solubility and conformational quality of recombinant proteins.

2. Study the conformational quality of soluble recombinant proteins by determining:
    a. Biological activity
    b. Extent of native-like structure

3. Construct vectors that allow production of our model proteins in a eukaryotic system.

4. Confirm whether conditions enhancing solubility and conformational quality in bacterial systems are also valid in a eukaryotic system.

5. Determine the effect of the major bacterial folding modulator, the chaperone DnaK, on protein quality and solubility in an environment that does not support its associated proteolytic activity.

3. Results

3.1. Article 1

Yield, solubility and conformational quality of soluble proteins are not simultaneously favored in recombinant *Escherichia coli*.

Mónica Martínez-Alonso, Elena García-Fruitós, Antonio Villaverde.

Biotechnology and Bioengineering, Vol. 101, No 6, 1353-8 (December 2008).

In this work we pursued the first objective of the study. Since solubility and quality are divergently controlled we explored whether it was possible to engineer them independently to simultaneously enhance both parameters.

The experimental design consisted of a two-step approach that combined genetic and process engineering. Since our purpose was to improve solubility of recombinant proteins, we chose a genetic background where our model protein was obtained mainly as highly functional inclusion bodies. Solubility was enhanced by appropriately adjusting growth temperature and gene expression rate. However, conditions promoting high protein yields resulted in poor conformational quality of the recombinant protein. Thus, since high yields of soluble and active protein cannot be gained simultaneously, the requirement of either solubility or functionality for a determined protein must be clearly established before designing its production process.

COMMUNICATION TO THE EDITOR

BIOTECHNOLOGY AND BIOENGINEERING

Yield, Solubility and Conformational Quality of Soluble Proteins Are Not Simultaneously Favored in Recombinant *Escherichia coli*

Mónica Martínez-Alonso, Elena García-Fruitós, Antonio Villaverde

Institute for Biotechnology and Biomedicine, Department of Genetics and Microbiology, Autonomous University of Barcelona, and CIBER-BBN Network in Bioengineering, Biomaterials and Nanomedicine, Bellaterra, 08193 Barcelona, Spain; telephone: 34-935812148; fax: 34-935812011; e-mail: avillaverde@servet.uab.es

Received 10 February 2008; revision received 29 April 2008; accepted 20 May 2008

Published online 3 June 2008 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/bit.21996

ABSTRACT: Many enzymes or fluorescent proteins produced in *Escherichia coli* are enzymatically active or fluorescent respectively when deposited as inclusion bodies. The occurrence of insoluble but functional protein species with native-like secondary structure indicates that solubility and conformational quality of recombinant proteins are not coincident parameters, and suggests that both properties can be engineered independently. We have here proven this principle by producing elevated yields of a highly fluorescent but insoluble green fluorescent protein (GFP) in a DnaK$^-$ background, and further enhancing its solubility through adjusting the growth temperature and *GFP* gene expression rate. The success of such a two-step approach confirms the independent control of solubility and conformational quality, advocates for new routes towards high quality protein production and intriguingly, proves that high protein yields dramatically compromise the conformational quality of soluble versions.

Biotechnol. Bioeng. 2008;101: 1353–1358.

KEYWORDS: protein folding; protein quality; solubility; IBs

Introduction

Very often, the bacterial production of recombinant proteins results in the formation of insoluble protein aggregates known as inclusion bodies (IBs) (Villaverde and Carrio, 2003). Improving solubility has been a main goal in protein production, and a spectrum of genetic, process engineering and physicochemical approaches have been explored with relative degree of success (Sorensen and Mortensen, 2005b). In particular, the co-production of appropriate sets of chaperones along with a misfolding-prone protein results in enhanced solubility ratios (de Marco et al., 2007) although, at least in some cases, in clearly lower protein yield and stability (Garcia-Fruitos et al., 2007).

Recently, by using Fourier transform infrared (FTIR) spectroscopy procedures (Ami et al., 2005, 2006; Oberg et al., 1994), it is being recognized that IBs contain important extents of properly folded, functional polypeptides (Ventura and Villaverde, 2006), and that protein aggregation in recombinant *Escherichia coli* does not necessarily imply loss of biological activity, neither in the cytoplasm (Garcia-Fruitos et al., 2005b) nor in the periplasm (Arie et al., 2006). The occurrence of functional protein versions in IBs seems to be inversely dependent on the aggregation rate (de Groot and Ventura, 2006). Interestingly, specific activity or fluorescence emission of recombinant enzymes and fluorescent proteins respectively is similar when comparing soluble and IB versions (Garcia-Fruitos et al., 2005a; Gonzalez-Montalban et al., 2006; Martinez-Alonso et al., 2007). This might result from a combination of functional protein species forming IBs and the occurrence of soluble aggregates (de Marco and Schroedel, 2005) that might contain, at different proportions, misfolded and non-functional proteins. Therefore, enhancing the solubility of a recombinant protein, irrespective of the used procedure, does not necessarily enhance the yield of functional versions (Gonzalez-Montalban et al., 2007).

Green fluorescent protein (GFP) fusions are excellent models to monitor conformational quality, as the proper conformation for fluorescence emission is reached during the last folding steps (maturation) of GFP (Herberhold et al., 2003; Scheyhing et al., 2002; Zhang et al., 2006). In a recent study, we have observed that *E. coli* mutant cells deficient in different chaperones or proteases were more fluorescent than wild type cells when producing an aggregation-prone GFP (mGFP) (Garcia-Fruitos et al., 2007), stressing the fact

Correspondence to: A. Villaverde

Contract grant sponsor: MEC, AGAUR

Contract grant numbers: BIO2007-61194; 2005SGR-00956

that conformational quality and solubility are not completely matching protein properties. In this context, we wondered if both parameters might be modulated by selectable conditions to enhance the yield of soluble but also biologically efficient protein. By using a misfolding-prone GFP variant we show here that the total cellular amount of functional protein can be dramatically enhanced by producing it in a DnaK$^-$ background, although it occurs as large IBs. In absence of DnaK, solubility of such functional polypeptides can be stimulated by appropriately adjusting growth temperature and gene expression rate. The success of such a combined, two-step (genetic and process) approach proves that solubility and conformational quality can be independently engineered, offering new strategies to optimize recombinant protein production processes. The results presented here indicate, however, that high protein yield dramatically compromises the conformational quality of the soluble product versions, being both parameters mutually exclusive. Therefore, recombinant production processes should be designed on the basis of the preferential outcome regarding yield and functionality.

Materials and Methods

Strains and Plasmids

E. coli pseudo wild type strain MC4100 (*araD139 Δ(argF-lac) U169 rpsL150 relA1 flbB5301 deoC1 ptsF25 rbsR*) (Sambrook et al., 1989) and its derivatives JGT3 (Δ*clpB::kan*), JGT4 (*clpA::kan*), JGT6 (*zjd::Tn10 groES30*), JGT17 (Δ*ibp::kan*), JGT19 (*clpP::cat*), JGT20 (*dnak756 thr::Tn10*) (Thomas and Baneyx, 1996), BB4564 (*groEL140 zjd::Tn10 zje::ΩSpc^r/Str^r*) (Ziemienowicz et al., 1993) and BB2395 (Δ*lon146::miniTn10*) (Tomoyasu et al., 2001) were used in this work. All these strains were transformed with plasmid pTVP1GFP (Garcia-Fruitos et al., 2007), which was used to drive the expression of a GFP fusion protein (mGFP) containing the aggregation-prone VP1 capsid protein of the foot-and-mouth disease virus. The chimerical *VP1GFP* gene is under the control of the ITPG-inducible *trc* promoter.

Culture and Gene Expression Conditions

Bacterial strains were cultured at 37°C and 250 rpm in shake flasks, in Luria–Bertani (LB) rich medium with 100 μg/mL ampicillin, up to an OD_{550} of 0.4. Then, the expression of the recombinant gene was triggered by addition of IPTG at three different final concentrations (namely 0.01, 0.1, or 1 mM) generally used to trigger recombinant gene expression (Lin et al., 2007; Wang et al., 2007). Aliquots of the culture were submitted then at different growth temperatures (16, 22, 27, 32, 37, or 42°C), again in the range used for the growth of protein-producing recombinant bacteria (Sambrook et al., 1989; Villaverde et al., 1993). Samples for analysis were taken when the culture reached an OD_{550} around 3. All experiments were performed in triplicate.

Protein Analysis

Samples of bacterial cultures (15 mL) were centrifuged (for 15 min at 15,000*g*) to harvest cells, and pellets were resuspended in 2 mL of phosphate-buffered saline (PBS) with one tablet of Protease Inhibitor Cocktail (Roche, ref. 1836170) per 10 mL of buffer. For analysis of the soluble fraction, 1-mL aliquots of the resuspended cells were ice-jacketed and sonicated for a minimum of 5 min at 50 W under 0.5 s cycles, or longer when required for total disruption of the cells (Feliu et al., 1998). After centrifugation for 15 min at 15,000*g*, the supernatant, corresponding to the soluble fraction, was mixed with denaturing buffer (Laemmli, 1970) at appropriate ratios for further Western Blot analysis.

The remaining 1-mL aliquots were used to purify IBs by repeated washing with detergent as described (Carrio et al., 2000) and resuspended in denaturing buffer. Samples were boiled for 20 min, and appropriate volumes were loaded onto denaturing gels for Western Blot analysis. mGFP was immunodetected using a rabbit polyclonal antibody against GFP (Santa Cruz Biotechnology, Inc., Santa Cruz, CA). Blots were scanned at high resolution and bands quantified using Quantity One software from Bio-Rad (Hercules, CA), using different amounts of commercial GFP as standards. Determinations were always done in triplicate and within the linear range, and they were used to calculate the specific activity values.

Fluorescence Determination

Soluble cell fraction samples were appropriately diluted in PBS and their fluorescence measured without any further treatment. IBs were purified as described above, and resuspended in PBS for fluorescence analysis. Determinations were carried out using a Cary Eclipse Fluorescence Spectrophotometer (Variant, Inc., Palo Alto, CA) and under continuous stirring. Excitation wavelength was 450 nm, and measures were taken at 510 nm. All experiments were performed in triplicate. The obtained data, combined with mGFP protein amounts determined by immunoanalysis, were used to calculate the specific fluorescence emission of both soluble and mGFP IBs.

Results

E. coli Genetic Background and Yield of Active Protein

We explored here the fluorescence distribution between soluble and insoluble cell fraction in several *E. coli* mutants, deficient in quality control functions, to select one with higher total fluorescence per cell. For further engineering

attempts, and to explore up to what extent solubility and functionality can be modulated, we were interested in strains with the fluorescent protein population being mainly insoluble. As observed (Table I), DnaK$^-$, ClpB$^-$, Lon$^-$ and ClpP$^-$ mutants produced significantly higher fluorescence emission than wild type cells. Interestingly, in all these cases, most of the fluorescent GFP accumulated as IBs, a fact that has been associated to a strong inhibition of DnaK$^-$surveyed proteolysis of functional protein species (Garcia-Fruitos et al., 2007). Interestingly, among these highly fluorescent mutants, DnaK$^-$ cells showed the lowest ratio between soluble and insoluble fluorescence (0.8 vs. 8.0 in the wild type). Hence, we decided to use this mutant to explore if conventional methods to enhance solubility could promote a more favorable distribution of functional protein between IBs and the soluble cell fraction, thus enhancing the occurrence of both soluble and fluorescent GFP.

Impact of Temperature and Gene Expression Rates on Protein Solubility and Conformational Quality

Therefore, we analyzed the fluorescence emission in DnaK$^-$ cells producing mGFP at different temperatures, from 16 to 42°C. As observed (Fig. 1A), both the total fluorescence per biomass and the particular fraction of emission associated with IBs increased with temperature, showing a sudden upshift between 27 and 32°C. However, the fluorescence associated with soluble protein only slightly decreased at the same temperature range, proving a positive effect of temperature on the absolute yield of functional and insoluble (but not soluble) protein. In this context, the ratio between soluble and insoluble fluorescence significantly increased at low temperatures, reaching 4.2 at 16°C (and dropping to 0.6 at 42°C). At 27°C or below, the prevalence of soluble fluorescent protein was then more than fourfold higher than at 32°C or higher temperatures.

With regard to protein production at each growth condition, we observed that while the amount of soluble mGFP showed a slight peak at 27°C, amounts of both total and insoluble mGFP increased with temperature (Fig. 1B). This resulted in a strong dependence of solubility (from 19.5% to 54%) on temperature. Altogether, these data suggested important differences in the temperature-mediated evolution of protein quality, depending on the soluble–insoluble protein status. In agreement (Fig. 1C), the specific emission of insoluble mGFP was poorly affected by temperature, although a minimum was observed at 27°C. However, the conformational quality of total mGFP increased with decreasing temperatures in an exponential pattern, what was essentially accounted for by the soluble fraction, since the specific fluorescence of aggregated mGFP was unaffected by temperature. At 27°C then, the soluble fluorescence was slightly higher than at other temperatures (Fig. 1A), probably because quality and solubility were both favored and yield was still high when compared to that obtained at lower temperatures (Fig. 1B).

We used then this intermediate growth temperature to analyze the effects of the IPTG concentration on solubility and protein quality in the range of doses commonly used for recombinant gene expression. As observed (Fig. 2A), total fluorescence per biomass was significantly lower at 0.01 mM than at the other tested concentrations (namely 0.1 and 1 mM), that produced very similar values. However, protein yield was strongly dependent on IPTG concentration (Fig. 2B). When combined with fluorescence data, these results suggest dramatic effects of IPTG on protein quality. This was indeed confirmed when determining the specific fluorescence of produced mGFP as distributed among different fractions (Fig. 2C). Medium IPTG values (0.1 mM) resulted in higher quality protein than that obtained at 1 mM. This fact accounts for the similar fluorescence per cell observed at these two IPTG doses (Fig. 1A) even when the higher protein yield was obtained at 1 mM IPTG (Fig. 2B).

Discussion

Solubility has been universally considered as the best indicator of recombinant protein quality. Therefore, gaining solubility is a main goal in protein production processes, and numerous strategies have been tested in this regard (Sorensen and Mortensen, 2005b). Many of them are based

Table I. Fluorescence emission observed in the soluble and IB fractions in mGFP-producing cells.

Phenotype	Total fraction		Soluble fraction		Inclusion bodies		Ratio of soluble/IB fluorescence[b]	Solubility (%)[c]
	units/OD[a]	%	units/OD	%	units/OD	%		
wt (MC4100)	405.6 ± 13.1	100	331.0 ± 71.6	100	40.9 ± 29.3	100	8.0	40.8 ± 11.1
DnaK$^-$	547.3 ± 77.9	134.9 ± 19.2	200.6 ± 33.7	60.6 ± 10.2	235.8 ± 42.5	575.6 ± 103.8	0.8	20.7 ± 3.4
GroEL140	342.9 ± 35.3	84.5 ± 8.7	245.2 ± 23.8	74.1 ± 7.2	47.1 ± 7.5	114.9 ± 18.4	5.2	30.2 ± 5.4
ClpB$^-$	515.6 ± 24.4	127.1 ± 6.0	310.7 ± 24.9	93.8 ± 7.5	231.8 ± 25.3	565.8 ± 61.9	1.3	31.2 ± 11.8
ClpA$^-$	543.1 ± 45.5	133.9 ± 11.2	46.9 ± 7.1	14.1 ± 2.1	66.2 ± 18.6	161.6 ± 45.5	0.7	44.2 ± 9.9
GroES$^-$	440.2 ± 5.3	108.5 ± 1.3	379.2 ± 15.9	114.5 ± 4.8	105.1 ± 13.0	256.7 ± 31.8	3.6	27.1 ± 7.8
IbpAB$^-$	303.2 ± 16.1	74.7 ± 4.0	261.3 ± 45.7	78.9 ± 13.8	62.9 ± 9.3	153.6 ± 22.8	4.1	30.1 ± 9.8
ClpP$^-$	522.4 ± 31.2	128.7 ± 7.7	431.0 ± 38.0	130.2 ± 11.4	237.8 ± 29.2	580.3 ± 71.4	1.8	25.0 ± 8.7
Lon$^-$	686.4 ± 40.9	169.2 ± 10.1	400.7 ± 21.8	121.0 ± 6.5	261.9 ± 12.5	639.4 ± 30.5	1.5	14.6 ± 3.5

[a]Total fluorescence data in this strain set (columns 1 and 2) have been obtained and shown in a previous study (Garcia-Fruitos et al., 2007).
[b]Quotient between soluble and insoluble fluorescence in a given sample.
[c]Amount of soluble mGFP relative to its total amount in the cell.

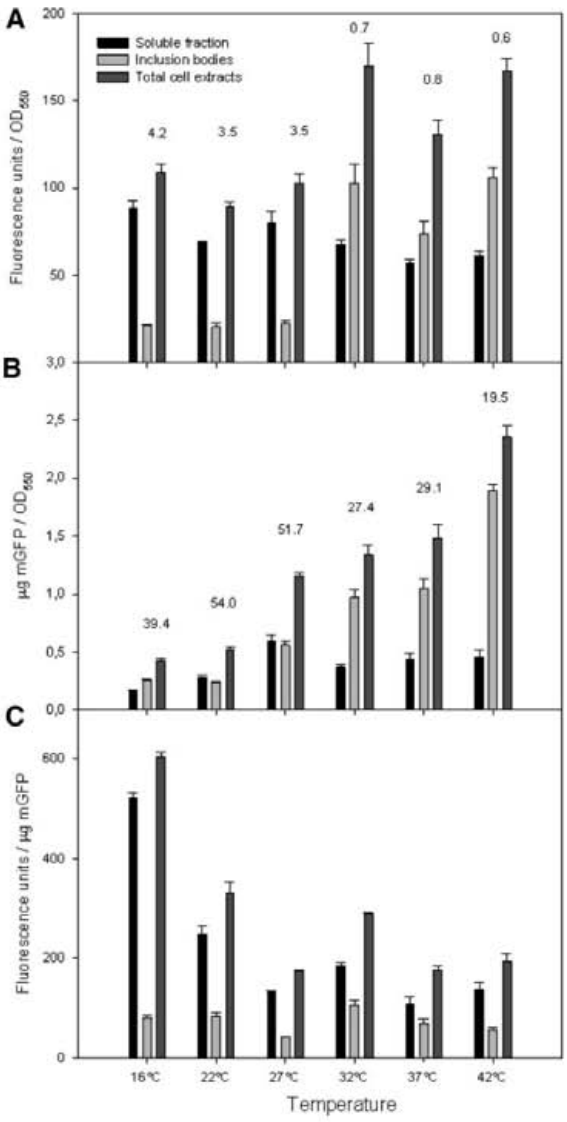

Figure 1. Effect of growth temperature on the fluorescence per cell biomass (**A**), yield (**B**) and specific fluorescence of mGFP (**C**). The ratio of soluble and insoluble fluorescence (panel A) and mGFP solubility (panel B) are also indicated for each temperature.

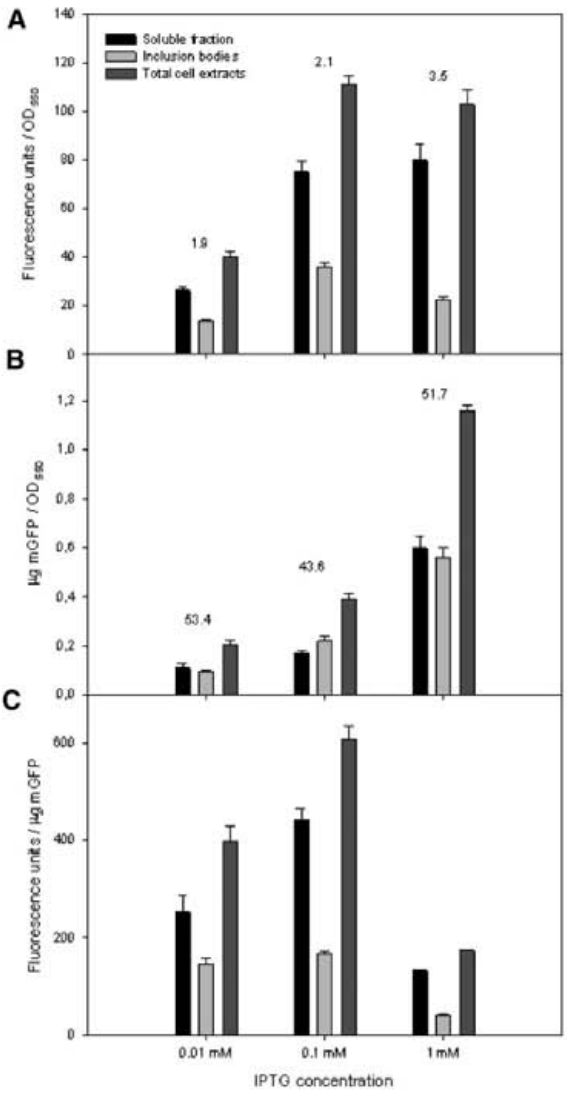

Figure 2. Effect of IPTG concentration on the fluorescence per cell biomass (**A**), yield (**B**) and specific fluorescence of mGFP (**C**). The ratio of soluble and insoluble fluorescence (panel A) and mGFP solubility (panel B) are also indicated for each dose.

on the production of chaperones along with the target protein, since they are believed to be limiting for recombinant protein folding. The selection of appropriate combinations of chaperones has resulted in higher solubility values (de Marco et al., 2007; Nishihara et al., 1998), usually expressed as the percentage of soluble over total protein. However, a detailed analysis of published data suggests that at least in some cases, increasing solubility through chaperone co-production would reduce the final protein yield. This concept has been clearly shown by the co-production of the DnaK–DnaJ pair, which dramatically reduces the proteolytic stability and yield of an IB-forming GFP (Garcia-Fruitos et al., 2007). In fact, a comprehensive genetic analysis of protein production in *E. coli* has recently indicated that cell mutations increasing solubility minimize the conformational quality of the soluble protein (Garcia-Fruitos et al., 2007). This fact, and other findings relevant to functionality of soluble and insoluble polypeptides in

recombinant bacteria, clearly prove that solubility and conformational quality are non-matching (and potentially divergent) protein properties (Gonzalez-Montalban et al., 2007).

For industrial processes requiring functional products, the production of highly active polypeptides (irrespective of their solubility) would be more appealing than high percentages (but poor yields) of soluble and moderately active polypeptides. However, solubility is obviously required for applications such as crystallographic determination or in vivo protein delivery for therapeutic purposes among others. In this work, we have explored how solubility of highly functional proteins (produced in a convenient DnaK$^-$ background) can be successfully manipulated through process engineering by manipulating growth temperature and gene expression rate. Interestingly, DnaK$^-$ cells producing an engineered GFP were up to 2.5-fold more fluorescent than wild type bacteria producing the same protein, despite the high aggregation level observed in the mutant. Furthermore, the total GFP yield in absence of DnaK was 1.6-fold higher than that obtained in wild type cells (Garcia-Fruitos et al., 2007), being the DnaK$^-$ background an appealing source of highly functional but aggregated protein that might be solubilized by a convenient choice of process parameters. In this regard, knocking *dnaK* gene has only a slight influence on growth rate at medium growth temperatures ($\mu = 0.932\ h^{-1}$ versus $\mu = 1.074\ h^{-1}$ in the wild type, at 37°C and in presence of 1 mM IPTG; not shown), what makes the mutant perfectly suitable to be used in bioprocesses.

Temperature, in the physiological ranges between 16 and 42°C, has a positive impact on the total yield of mGFP. This is exclusively accounted for by an increase in the amount of aggregated protein since the yield of the soluble version is only slightly affected (Fig. 1B). The total fluorescence per cell undergoes an up-shift above 27°C, but again it is accounted for exclusively by the insoluble cell fraction (Fig. 1A). Finally, the conformational quality of soluble mGFP is dramatically and progressively impaired by temperature while specific fluorescence of IB mGFP remains nearly constant (Fig. 1C). The influence of IPTG concentration is more modest regarding the variation range of the studied parameters, which follow a less progressive pattern than the one defined by temperature. However, the divergent evolution of yield (and total fluorescence) and the functional quality of the soluble protein version is also evident (Fig. 2). Importantly, by combining the appropriate temperature (27°C) and IPTG dose (0.1 mM), the distribution of fluorescence between soluble and insoluble shifted from 0.8 (Table I and Fig. 1A) to 2.1 (Fig. 2A). Of course, better distribution values can be reached at 1 mM IPTG (3.5), but at the expense of protein quality measured by specific fluorescence (Fig. 2C).

More intriguingly, the data presented here indicate that yield, solubility and conformational quality of soluble proteins cannot be favored simultaneously in recombinant *E. coli*. This fact must be seriously considered in protein production processes, since the production strategy should be clearly targeted to protein yield, solubility or product quality. In this regard, many of the non-coincident reports regarding the success of given strategies for improved protein production (Baneyx and Palumbo, 2003; Baneyx and Mujacic, 2004; de Marco et al., 2000, 2007; de Marco and De Marco, 2004; Sorensen and Mortensen, 2005a; Schultz et al., 2006) and the unpredictability and product-dependence of the chaperone co-production approach (de Marco, 2007; de Marco et al., 2007) are probably accounted for (at least in many cases) by the different parameters through which process success is measured, namely solubility, yield, or functionality. While evidences that enhancing solubility does not imply better protein quality are now stronger (Gonzalez-Montalban et al., 2007), the results presented here furthermore indicate that conditions promoting high protein yield and high soluble yield are clearly adverse for conformational quality (Fig. 3). In this context, the distribution of fluorescence between soluble and insoluble cell fractions (Fig. 3A) and the specific fluorescence of soluble mGFP (Fig. 3B) are negatively

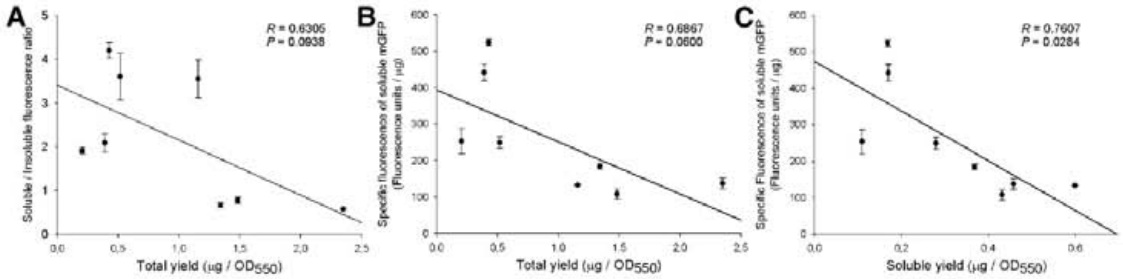

Figure 3. Influence of total (**A** and **B**) and soluble (**C**) mGFP yield on soluble/insoluble fluorescence ratio (A) and specific fluorescence of soluble mGFP (B and C). All the conditions shown in Figure 1 and Figure 2 were used in this analysis. In all cases, the data set also fitted to exponential decay, single, two parameter equations (not shown) with important extents of statistic significance (A, $P = 0.1209$; B, $P = 0.0444$; C, $P = 0.0292$).

affected by the total production of mGFP. Likewise, the lower the yield of soluble mGFP, the higher its conformational quality is (Fig. 3C), strongly supporting the concept that gaining yield and quality cannot be reached simultaneously.

This work has been supported by grants BIO2007-61194 (MEC) and 2005SGR-00956. MMA and EGF are recipients of predoctoral fellowships from MEC, Spain. The authors thank Salvador Bartolomé for helpful technical assistance.

References

Ami D, Natalello A, Gatti-Lafranconi P, Lotti M, Doglia SM. 2005. Kinetics of inclusion body formation studied in intact cells by FT-IR spectroscopy. FEBS Lett 579:3433–3436.

Ami D, Natalello A, Taylor G, Tonon G, Maria DS. 2006. Structural analysis of protein inclusion bodies by Fourier transform infrared microspectroscopy. Biochim Biophys Acta 1764:793–799.

Arie JP, Miot M, Sassoon N, Betton JM. 2006. Formation of active inclusion bodies in the periplasm of *Escherichia coli*. Mol Microbiol 62:427–437.

Baneyx F, Mujacic M. 2004. Recombinant protein folding and misfolding in *Escherichia coli*. Nat Biotechnol 22:1399–1408.

Baneyx F, Palumbo JL. 2003. Improving heterologous protein folding via molecular chaperone and foldase co-expression. Methods Mol Biol 205:171–197.

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

de Groot NS, Ventura S. 2006. Protein activity in bacterial inclusion bodies correlates with predicted aggregation rates. J Biotechnol 125:110–113.

de Marco A, Schroedel A. 2005. Characterization of the aggregates formed during recombinant protein expression in bacteria. BMC Biochem 6:10.

de Marco A, Volrath S, Bruyere T, Law M, Fonne-Pfister R. 2000. Recombinant maize protoporphyrinogen IX oxidase expressed in *Escherichia coli* forms complexes with GroEL and DnaK chaperones. Protein Exp Purif 20:81–86.

de Marco A. 2007. Protocol for preparing proteins with improved solubility by co-expressing with molecular chaperones in *Escherichia coli*. Nat Protoc 2:2632–2639.

de Marco A, De Marco V. 2004. Bacteria co-transformed with recombinant proteins and chaperones cloned in independent plasmids are suitable for expression tuning. J Biotechnol 109:45–52.

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.

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

Garcia-Fruitos E, Carrio MM, Aris A, Villaverde A. 2005a. Folding of a misfolding-prone beta-galactosidase in absence of DnaK. Biotechnol Bioeng 90:869–875.

Garcia-Fruitos E, Gonzalez-Montalban N, Morell M, Vera A, Ferraz RM, Aris A, Ventura S, Villaverde A. 2005b. Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. Microb Cell Fact 4:27.

Garcia-Fruitos E, Martinez-Alonso M, Gonzalez-Montalban N, Valli M, Mattanovich D, Villaverde A. 2007. Divergent genetic control of protein solubility and conformational quality in *Escherichia coli*. J Mol Biol 374:195–205.

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.

Gonzalez-Montalban N, Garcia-Fruitos E, Villaverde A. 2007. Recombinant protein solubility—Does more mean better? Nat Biotechnol 25:718–720.

Herberhold H, Marchal S, Lange R, Scheyhing CH, Vogel RF, 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.

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

Lin GZ, Lian YJ, Ryu JH, Sung MK, Park JS, Park HJ, Park BK, Shin JS, Lee MS, Cheon CI. 2007. Expression and purification of His-tagged flavonol synthase of *Camellia sinensis* from *Escherichia coli*. Protein Exp Purif 55:287–292.

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 Lett 273:187–195.

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 64:1694–1699.

Oberg K, Chrunyk BA, Wetzel R, Fink AL. 1994. Nativelike secondary structure in interleukin-1 beta inclusion bodies by attenuated total reflectance FTIR. Biochemistry 33:2628–2634.

Sambrook J, Fritsch E, Maniatis T. 1989. Molecular Cloning, A Laboratory Manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Scheyhing CH, Meersman F, Ehrmann MA, Heremans K, Vogel RF. 2002. Temperature-pressure stability of green fluorescent protein: A Fourier transform infrared spectroscopy study. Biopolymers 65:244–253.

Schultz T, Martinez L, de MA , 2006. The evaluation of the factors that cause aggregation during recombinant expression in *E. coli* is simplified by the employment of an aggregation-sensitive reporter. Microb Cell Fact 5:28.

Sorensen HP, Mortensen KK. 2005a. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. J Biotechnol 115: 113–128.

Sorensen HP, Mortensen KK. 2005b. Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. Microb Cell Fact 4:1.

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 21:1185–1196.

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.

Ventura S, Villaverde A. 2006. Protein quality in bacterial inclusion bodies. Trends Biotechnol 24:179–185.

Villaverde A, Benito A, Viaplana E, Cubarsi R. 1993. Fine regulation of cI857-controlled gene expression in continuous culture of recombinant *Escherichia coli* by temperature. Appl Environ Microbiol 59:3485–3487.

Villaverde A, Carrio MM. 2003. Protein aggregation in recombinant bacteria: Biological role of inclusion bodies. Biotechnol Lett 25: 1385–1395.

Wang J, Tan H, Zhao ZK. 2007. Over-expression, purification, and characterization of recombinant NAD-malic enzyme from *Escherichia coli* K12. Protein Exp Purif 53:97–103.

Zhang L, Patel HN, Lappe JW, Wachter RM. 2006. Reaction progress of chromophore biogenesis in green fluorescent protein. J Am Chem Soc 128:4766–4772.

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.

3.2. Article 2

The functional quality of soluble recombinant polypeptides produced in *Escherichia coli* is defined by a wide conformational spectrum.

Mónica Martínez-Alonso, Nuria González-Montalbán, Elena García-Fruitós and Antonio Villaverde.

Applied and Environmental Microbiology, Vol. 74, No 23, 7431-3 (December 2008).

The finding of oligomeric versions in the soluble fraction of our recombinant protein prompted us to further characterise this protein population in terms of functional quality and molecular organisation.

To that end, we analysed the distribution of the soluble fraction of an aggregation-prone recombinant GFP along a sucrose density gradient. The protein widely dispersed along the gradient, indicating the presence of differently sized species within the soluble population. Furthermore, the fluorescence profile did not match the protein distribution, indicative of a variable functional status in the soluble fraction. Further purification of one of the protein species observed in the gradient still resulted in a heterogeneous population of microaggregates, as observed by transmission electron microscopy. These soluble aggregates were also heterogeneous regarding their secondary structure, as evidenced by the presence of both non native and native-like conformations. However, the prevalence of native-like structures accounted for the higher functionality of the soluble protein compared to inclusion bodies. Being structurally more homogeneous than their soluble counterparts, IBs can be regarded as a narrow subpopulation among the total recombinant protein species. Therefore, the observed protein quality can be regarded as a statistical average of all the existing protein species.

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Dec. 2008, p. 7431–7433 Vol. 74, No. 23
0099-2240/08/$08.00+0 doi:10.1128/AEM.01446-08

The Functional Quality of Soluble Recombinant Polypeptides Produced in *Escherichia coli* Is Defined by a Wide Conformational Spectrum▽

Mónica Martínez-Alonso, Nuria González-Montalbán, Elena García-Fruitós, and Antonio Villaverde*

Institute for Biotechnology and Biomedicine, Department of Genetics and Microbiology, Autonomous University of Barcelona, and CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Spain, Bellaterra, 08193 Barcelona, Spain

Received 27 June 2008/Accepted 24 September 2008

We have observed that a soluble recombinant green fluorescent protein produced in *Escherichia coli* occurs in a wide conformational spectrum. This results in differently fluorescent protein fractions in which morphologically diverse soluble aggregates abound. Therefore, the functional quality of soluble versions of aggregation-prone recombinant proteins is defined statistically rather than by the prevalence of a canonical native structure.

The quality of recombinant proteins produced in bacteria and other host cells represents a major matter of concern in biological protein production and determines the potential use of target proteins for functional applications or structural analysis (2, 17). In contrast to what was formerly believed, the straightforward measurement of the soluble protein yield or the ratio between the soluble and total protein yields (usually given as a percentage of solubility) is not a useful indicator of quality. Properly folded and functional polypeptides often aggregate as inclusion bodies; therefore, an important fraction of functional protein species occurs in the insoluble cell fraction (7, 9, 18). Thus, the specific biological activity rather than the presence of the protein species in the soluble cell fraction reveals the conformational quality of the product and, therefore, its biotechnological potential. In this regard, an increasing number of structural analyses reveal the coexistence, in the embedded protein species, of a cross-β-sheet-based, amyloid-like organization (3) and also that of a native secondary structure (5). In inclusion bodies formed by enzymes, the associated enzymatic activity is sufficient for efficient in situ substrate processing. The proposal of biologically active inclusion bodies being usable as catalyzers (10) has resulted in the incorporation of diverse enzymes, in the form of inclusion bodies (including β-galactosidase, D-amino acid oxidase, maltodextrin phosphorylase, sialic acid aldolase, and polyphosphate kinase) (6, 11–15), into different types of enzymatic processes.

The molecular organization and quality of the soluble protein population, which is generally believed to adopt the native, functional conformation, have been studied much less. Therefore, in this conventional view, soluble proteins are expected to show a rather narrow conformational spectrum and to be highly functional. However, the recent finding that the solubility and conformational quality in recombinant bacteria are divergently controlled (8) seriously challenges such an assumption. Moreover, several independent observations clearly argue against a model picturing the soluble protein fraction as fully functional and structurally homogeneous. First, the specific activity of a recombinant β-galactosidase aggregated as inclusion bodies is, under defined production conditions, higher than that of its soluble counterpart (7). This strongly suggests that the activity of soluble protein species represents an average of the numbers of coexisting active and inactive protein forms. In the same context, reducing the growth temperature of recombinant *Escherichia coli* cells from 37°C to 16°C results in a significant increase of the specific emission of a recombinant green fluorescent protein (GFP) (19), indicating that at 37°C an important fraction of soluble species have not maturated to an optimal conformation for fluorescence emission. Finally, the finding of the so-called soluble aggregates, which are fibril-like structures in *E. coli* cells overproducing GFP variants (4), indicates that even soluble species can display an amyloidal organization, such as that found in inclusion bodies, and that GFP in these soluble aggregates may eventually fold into a form very different from the native conformation. Aggregation of soluble versions of other structurally diverse proteins, including β-galactosidase (1) and different maltose-binding fusion proteins (16), has also been reported. The functional properties of such soluble aggregates remain unexplored. Therefore, the meaning of "solubility" in both structural and functional terms is as yet essentially obscure.

To clarify the folding scenery of the soluble protein population in recombinant bacteria and the biological significance of solubility, we have examined (through fluorescence emission) the conformational quality in the soluble fraction of recombinant *E. coli* MC4100 cells producing a model GFP (mGFP) by conventional culture and induction of gene expression procedures (8). mGFP is a GFP fused to the VP1 capsid protein of foot-and-mouth disease virus that drives protein aggregation immediately upon production in *E. coli*. Consequently, most of the fusion protein is found in the form of inclusion bodies and only up to 45% of the total mGFP yield occurs in the soluble cell fraction (8).

Cell pellets were obtained by centrifugation of bacterial cultures producing mGFP at 6,000 × *g* for 30 min at 4°C, frozen overnight at −80°C, and resuspended in lysis buffer (50 mM Tris-HCl [pH 8], 100 mM NaCl, 1 mM EDTA [pH 8]). Ice-

* Corresponding author. Mailing address: Institute for Biotechnology and Biomedicine, Autonomous University of Barcelona, Bellaterra, 08193 Barcelona, Spain. Phone: 34 935813086. Fax: 34 935812011. E-mail: avillaverde@servet.uab.es.

▽ Published ahead of print on 3 October 2008.

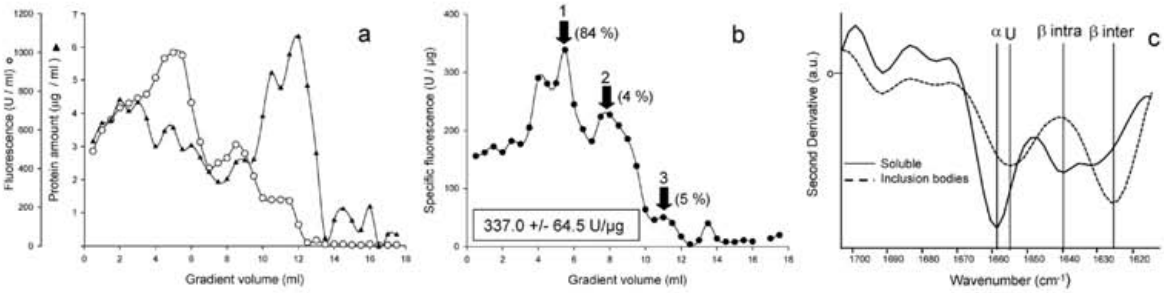

FIG. 1. (a) Distribution of the soluble version of the aggregation-prone mGFP (triangles) and fluorescence emission (circles) along a sucrose gradient. (b) Specific fluorescence emission of soluble GFP populations. Arrows indicate discrete peaks of highly fluorescent protein subfractions (with GFP purity values given as percentages), and the inset shows values for the average specific fluorescence of soluble GFP in the whole soluble cell fraction. (c) FTIR spectra (normalized at the tyrosine peak by use of GRAMS/AI spectroscopic software, version 7) of isolated GFP inclusion bodies (dashed line) and the purified GFP population (from peak 1) (continuous line). Vertical lines indicate the relevant peaks corresponding to α-helix (α), unfolded stretches (U), native intramolecular β-sheet (β intra), and intermolecular cross β-sheet (β inter).

jacketed samples were sonicated for 35 min at 50 W in cycles of 0.5 s. The soluble cell fraction was separated from cell debris and inclusion bodies by centrifugation at 15,000 × *g* for 15 min at 4°C and was submitted to density gradient ultracentrifugation (ranging from 0 to 80% sucrose) at 92,444 × *g* for 16 h at 4°C. Recombinant GFP widely dispersed along the gradient, showing a profile that was not coincident with that of the total fluorescence (Fig. 1a). In contrast to what would be expected for a monodispersal conformational model, such a lack of matching indicates a variable functional status within the soluble protein population. This functional heterogeneity was further confirmed by Western blot quantitative analysis of GFP as described previously (8) to determine the specific fluorescence, which peaked as three independent protein fractions (Fig. 1b). The GFP population enclosed in peak number 1 was further purified by size exclusion chromatography using fast protein liquid chromatography equipment, resulting in a highly (84%) pure and fluorescent population. Therefore, according to the conventional concept of soluble protein characteristics, the protein species found therein would be conformationally homogeneous and functional, having reached the native conformation. However, when we analyzed this particular population by transmission electron microscopy we observed a spectrum of soluble aggregates ranging from particulate material to fibril-like structures, with both categories demonstrating immunoreactivity with anti-GFP antibodies (Fig. 2a) as inclusion bodies themselves (Fig. 2b). Moreover, the Fourier transform infrared spectroscopy (FTIR) pattern of these protein species was compared with that of GFP inclusion bodies. The insoluble aggregates were observed to be rich in intermolecular β-sheets, with a minor occurrence of native-like α-helix and/or unfolded stretches (Fig. 1c). However, the soluble GFP did show a wider

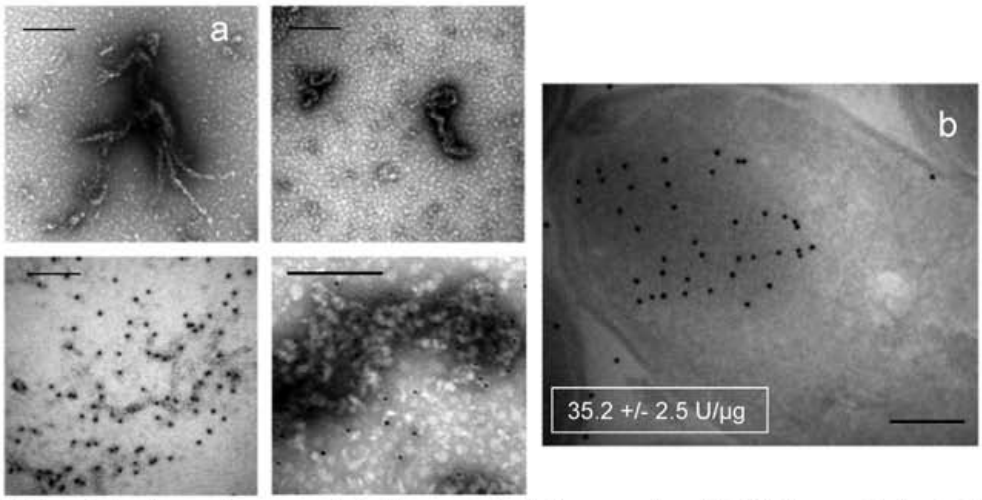

FIG. 2. (a) Transmission electron microscopy analysis of fast protein liquid chromatography-purified GFP from peak 1, showing fibril-like and particulate microaggregates (top). At the bottom, the results of immunodetection of GFP on the same samples are shown. (b) Immunolabeling of GFP in inclusion bodies, on cryosections of GFP-producing *E. coli* cells. In the inset figures, the specific fluorescence values of GFP determined on isolated inclusion bodies are indicated. Bars, 0.2 μm.

set of conformational types in which α-helix and intramolecular native-like β-sheets were prominent. The significant occurrence of these protein forms, indicative of native conformation, accounts for the higher specific emission level found in the soluble protein species within peak 1 (and also peak 2; Fig. 1b) compared with that of inclusion bodies (shown in the inset of Fig. 2b). In addition, such native-like folding patterns in soluble material were accompanied by unfolded and intermolecular β-sheet species. The smoother peaks in the FTIR plot of soluble protein (resulting from the contribution of several bands between 1,650 and 1,630 cm^{-1}) compared to those of the more defined inclusion body FTIR spectra and the higher number of structural patterns again stressed the high structural heterogeneity of soluble protein species. Polypeptides aggregated as inclusion bodies (Fig. 2b) would then represent a particular and narrow subpopulation among the total recombinant protein species which, upon clustering through intermolecular interactions, deposit as insoluble material. Still being conformationally diverse, they would be significantly more homogeneous regarding their folding state than their soluble counterparts.

In summary, the soluble fraction of a recombinant GFP produced in *E. coli* is composed of diverse protein populations with distinct specific fluorescence characteristics, with the most pure and fluorescent subfraction (peak 1) still being formed by a spectrum of protein forms and microaggregates. All those protein species, and the less fluorescent variants of peaks 2 and 3, generate, as a set, an average specific emission that defines the quality of the recombinant protein in the soluble cell fraction (note the average specific fluorescence of soluble GFP given in the inset of Fig. 1b). The protein "quality" of model mGFP, and presumably of other recombinant polypeptides, could then be statistically observed as the relative abundance of the most active protein species, which would be closer to the canonical native conformation. Adjusting either the production conditions or the genetic composition of the cell quality control system (either by knocking down or by overexpressing chaperone or protease genes) would alter protein quality by unbalancing the prevalence of active (native or native-like) and inactive (misfolded) protein species in the cell context.

We appreciate the financial support to our research on recombinant protein production through grants BIO2007-61194 and BIO2005-23732-E (MEC) and grant 2005SGR-00956 (AGAUR). We are also grateful for the financial support provided by the CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN, promoted by ISCIII), Spain. M.M.-A., N.G.-M., and E.G.-F. are recipients of fellowships from MEC, Spain.

We thank Salvador Bartolomé and Alejandro Sànchez Chardi for helpful technical assistance.

REFERENCES

1. **Aris, A., and A. Villaverde.** 2000. Molecular organization of protein-DNA complexes for cell-targeted DNA delivery. Biochem. Biophys. Res. Commun. **278:**455–461.
2. **Baneyx, F., and M. Mujacic.** 2004. Recombinant protein folding and misfolding in *Escherichia coli*. Nat. Biotechnol. **22:**1399–1408.
3. **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.
4. **de Marco, A., and A. Schroedel.** 2005. Characterization of the aggregates formed during recombinant protein expression in bacteria. BMC Biochem. **6:**10.
5. **Doglia, S. M., D. Ami, A. Natalello, P. Gatti-Lafranconi, and M. Lotti.** 2008. Fourier transform infrared spectroscopy analysis of the conformational quality of recombinant proteins within inclusion bodies. Biotechnol. J. **3:**193–201.
6. **Garcia-Fruitos, E., A. Aris, and A. Villaverde.** 2007. Localization of functional polypeptides in bacterial inclusion bodies. Appl. Environ. Microbiol. **73:**289–294.
7. **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.
8. **Garcia-Fruitos, E., M. Martinez-Alonso, N. Gonzalez-Montalban, M. Valli, D. Mattanovich, and A. Villaverde.** 2007. Divergent genetic control of protein solubility and conformational quality in *Escherichia coli*. J. Mol. Biol. **374:**195–205.
9. **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.
10. **Gonzalez-Montalban, N., E. Garcia-Fruitos, and A. Villaverde.** 2007. Recombinant protein solubility—does more mean better? Nat. Biotechnol. **25:**718–720.
11. **Nahálka, J.** 2008. Physiological aggregation of maltodextrin phosphorylase from *Pyrococcus furiosus* and its application in a process of batch starch degradation to α-D-glucose-1-phosphate. J. Ind. Microbiol. Biotechnol. **35:**219–223.
12. **Nahalka, J., I. Dib, and B. Nidetzky.** 2008. Encapsulation of *Trigonopsis variabilis* D-amino acid oxidase and fast comparison of the operational stabilities of free and immobilized preparations of the enzyme. Biotechnol. Bioeng. **99:**251–260.
13. **Nahálka, J., P. Gemeiner, M. Bucko, and P. G. Wang.** 2006. Bioenergy beads: a tool for regeneration of ATP/NTP in biocatalytic synthesis. Artif. Cells Blood Substit. Immobil. Biotechnol. **34:**515–521.
14. **Nahálka, J., A. Vikartovska, and E. Hrabarova.** 2008. A crosslinked inclusion body process for sialic acid synthesis. J. Biotechnol. **134:**146–153.
15. **Navrátil, M., P. Gemeiner, J. Klein, E. Sturdík, A. Malovíková, J. Nahálka, A. Vikartovská, Z. Dömény, and D. Smogrovicová.** 2002. Properties of hydrogel materials used for entrapment of microbial cells in production of fermented beverages. Artif. Cells Blood Substit. Immobil. Biotechnol. **30:**199–218.
16. **Sachdev, D., and J. M. Chirgwin.** 1999. Properties of soluble fusions between mammalian aspartic proteinases and bacterial maltose-binding protein. J. Protein Chem. **18:**127–136.
17. **Sørensen, H. P., and K. K. Mortensen.** 2005. Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. Microb. Cell Fact. **4:**1.
18. **Ventura, S., and A. Villaverde.** 2006. Protein quality in bacterial inclusion bodies. Trends Biotechnol. **24:**179–185.
19. **Vera, A., N. Gonzalez-Montalban, A. Aris, and A. Villaverde.** 2007. The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. Biotechnol. Bioeng. **96:**1101–1106.

3.3. Article 3

Learning about protein solubility from bacterial inclusion bodies.

Mónica Martínez-Alonso, Nuria González-Montalbán, Elena García-Fruitós and Antonio Villaverde.

Microbial Cell Factories, Vol. 8, 4-8 (January 2009).

In this work we summarised all the recent findings that support a new concept of protein quality, which can no longer be linked to solubility.

Although formerly believed to be insoluble deposits of inactive protein, inclusion bodies are actually rich in functional protein species with native secondary structure. This opens avenues for a straightforward application of enzymatic inclusion bodies as catalysers for industrial bioprocesses. Moreover, easy extraction of active polypeptides has been achieved without the need of complex refolding procedures after adequate engineering of IB protein quality, which may translate in enhanced *in vitro* release of functional protein.

Soluble protein can no longer be thought of as a homogeneous population of protein species either, since the existence of soluble aggregates with variable functional conformations prompts to consider recombinant proteins as a “continuum of forms” rather than the classic soluble and insoluble cell fractions.

In addition, protein production should be targeted to yield, quality or solubility of the recombinant product, as these parameters are under a divergent control and cannot be enhanced at the same time.

Microbial Cell Factories

BioMed Central

Commentary

Open Access

Learning about protein solubility from bacterial inclusion bodies

Mónica Martínez-Alonso[1,2], Nuria González-Montalbán[1,2], Elena García-Fruitós[1,2] and Antonio Villaverde*[1,2]

Address: [1]Institute for Biotechnology and Biomedicine and Department of Genetics and Microbiology, Autonomous University of Barcelona, Barcelona, Spain and [2]CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Spain

Email: Mónica Martínez-Alonso - monica.martinez.alonso@uab.cat; Nuria González-Montalbán - Nuria.Gonzalez.Montalban@uab.cat; Elena García-Fruitós - Elena.Garcia.Fruitos@uab.es; Antonio Villaverde* - avillaverde@servet.uab.es

* Corresponding author

Published: 8 January 2009

Microbial Cell Factories 2009, **8**:4 doi:10.1186/1475-2859-8-4

Received: 19 December 2008
Accepted: 8 January 2009

This article is available from: http://www.microbialcellfactories.com/content/8/1/4

Abstract

The progressive solving of the conformation of aggregated proteins and the conceptual understanding of the biology of inclusion bodies in recombinant bacteria is providing exciting insights on protein folding and quality. Interestingly, newest data also show an unexpected functional and structural complexity of soluble recombinant protein species and picture the whole bacterial cell factory scenario as more intricate than formerly believed.

Commentary

The conformational quality of soluble recombinant proteins is an emerging matter of concern, especially when the obtained products are to be used for functional or interactomic analyses [1]. In the context of recombinant protein production, the general believing that soluble protein species are properly folded and fully functional in contrast to the misfolded and inactive protein versions trapped in insoluble inclusion bodies [2], cannot be longer supported by current research data. The dropping of independent references to inclusion bodies as entities formed by functional protein species with native secondary structure is progressively increasing, and the structural and functional diversity of the model proteins used in these studies [3-13] does leave little room to speculate about this fact as being an artefact or a peculiarity of a limited number of protein species. Recent reviews in this area have presented properly folded proteins as natural components of inclusion bodies [10,14], indirectly compromising the paradigm of recombinant protein solubility as equivalent to protein conformational quality [15].

Indeed, the occurrence of functional proteins as important components of bacterial aggregates prompts to reconsider the conformational quality of protein species occurring in the soluble cell fraction of inclusion body-forming cells, that might be lower than expected. Several indirect observations are also in this line; (i) the functional quality of recombinant proteins in *E. coli* is affected in parallel by physical parameters such as temperature (high temperature impairs protein activity in both soluble and insoluble cell fractions) [16] and physiological conditions such as the availability of chaperones (a molar excess of DnaKJ inactivates both soluble and insoluble recombinant proteins) [17]; (ii) *in vivo* disintegration of inclusion bodies is strongly dependent on proteolytic degradation [18-21] for which DnaK is required [20], indicating a tight surveillance of the quality control system over aggregated protein species; (iii) inclusion body-forming proteins can complete their folding process once embedded in these aggregates [22]; (iv) the soluble versions of recombinant proteins can occur as soluble aggregates [23,24]; (v) the functional quality (measured for a model

enzyme as its specific activity and fluorescent proteins by specific emission) of soluble protein versions can be lower than that of the inclusion body counterparts [3], and be eventually improved by reducing the growth temperature of recombinant cells from 37 to 16°C [16]. This indicates that at 37°C, an important fraction of soluble protein species are inactive, suggesting that they have not reached their native conformation. This has been very recently explored by sub-fractioning the soluble population of an inclusion body-forming recombinant GFP and their subsequent functional analysis. Indeed, there is a large functional diversity within the soluble protein population (accompanied by an extremely high abundance of soluble aggregates, either globular or fibrilar) [24], that prompts to observe the specific fluorescence of the soluble protein version as an average rather than a canonical value defined by a single type of molecular species.

In this scenario, recombinant proteins in producing cells can be seen as adopting "a continuum of forms" [23] expanding from soluble to insoluble cell fractions, and inclusion bodies as insoluble "clusters" of protein species [19]. Therefore, soluble versions of a given protein would not necessarily show better conformational quality than the aggregated counterparts, although the average biological activity (specific activity for enzymes or specific fluorescence for fluorescent proteins) is in general higher in the soluble cell fraction [3,24]. Interestingly, the specific enzymatic activities (or fluorescence emission) of soluble and insoluble protein versions tend to adopt similar values under specific conditions such as in DnaK knockout mutants [25,26]. Therefore, the soluble and insoluble "virtual" cell fractions in bacteria [14] are now regarded as more virtual than ever, as the main feature distinguishing soluble and inclusion body protein species might be the dispersed-clustered status rather than the biological activity.

From a practical point of view, these emerging concepts about protein aggregation in recombinant bacteria have remarkable implications. First, inclusion bodies formed by enzymes can be straightforward used as catalysers in industry-relevant enzymatic reactions skipping any previous *in vitro* refolding protocols [5-7]. Second, the quality of inclusion body proteins can be dramatically enhanced by producing them at suboptimal temperatures. This should not only permit the production of inclusion bodies with improved catalyzing properties but it also might favour the controlled *in vitro* release of functional proteins from these aggregates. In this regard, the recovery of functional proteins from inclusion bodies has been a largely used strategy when a desired protein species showed a high aggregation tendency. Such an approach implies separation of inclusion bodies, efficient protein unfolding under extreme denaturation conditions and further refolding through complex (and often unsuccessful) step strategies to be optimized for any particular protein species [27]. However, in the last years, an increasing piece of evidence points out that inclusion bodies with high content of native-like structure could be easily solubilised in non-denaturing conditions avoiding strong denaturation and refolding steps. A set of non related proteins, namely GFP [28], archaeon proteins, cytokines, immunoglobulin-folded proteins [29] and β-2-microglobulin [30], have been successfully extracted from inclusion bodies without the need of denaturing conditions, basically using as solubilising agents L-arginine and GdnHCl at non-denaturing concentrations [28,29]. Also in this line, Menart and co-workers observed that functional proteins could be easily extracted from inclusion bodies using non denaturing mild detergents and polar solvents, provided that the cells would have been cultured under suboptimal temperatures [12]. Such inclusion bodies, being a straightforward source of soluble proteins, were named "non-classical" because of their unexpected high content of functional, extractable species. Although sufficient data has been now accumulated to infer that in general, inclusion bodies are non-classical by nature (regarding the unlink between solubility and activity) [15], this interesting approach would potentially permit to skip complex refolding procedures by engineering the quality of inclusion body proteins during the production process. In very recent papers, Peternel and co-workers reported not only the successful extraction of functional polypeptides from inclusion bodies but also the fact that, in some cases, the biological activity of these inclusion body-solubilised proteins was comparable or even higher that than found in the soluble fraction. For instance, human granulocyte-stimulating factor (hG-CSF), GFP and lymphotoxin α (LT-α) extracted from inclusion bodies represented around the 98%, 40% and 25%, respectively, of the total biological activity and fluorescence emission in the recombinant protein producing-cells [31,32]. Again, the different structural and biological properties of the proteins for which this principle has been proved indicate that the extractability of functional proteins from inclusion bodies is not a particular issue, although its applicability at large scale needs to be further evaluated. On the other side, as an additional strategy, the specific activity of inclusion body proteins can be successfully enhanced by down-regulating the levels of recombinant gene expression [33,34].

Finally, since early recombinant DNA times, when the formation of inclusion bodies was noticed as a general undesirable event [35], enhancing protein solubility has been compulsory pursued through diverse approaches. The need for soluble proteins for many research, industrial and pharmaceutical applications has pushed microbiologists, biochemists and chemical engineers to modify cell, protein and process conditions (using protease-deficient

cells, chaperone co-production, removing hydrophobic regions, fusion of solubility tags, minimizing the growth rate or using weak gene expression induction conditions among others), in an attempt to favour the occurrence of the target protein in the soluble cell fraction [36-41]. However, solubility is often observed as an academic parameter, namely the quotient (in %) between soluble and total protein and therefore with a questionable practical value. Interestingly, it is very rare to find in the literature measures of solubility simultaneous to determinations of protein yield or functional quality, when attempting a novel strategy to minimize inclusion body formation, such as for instance, the co-production of chaperones along with the recombinant protein species. In this regard, enhancing the levels of trigger factor and GroELS increases the solubility of a recombinant lysozyme that shows a specific activity lower than in absence of additional chaperones [42]. Other chaperone sets have been observed to promote solubility of target proteins [40,43,44] without a detailed analysis of protein quality and activity or by determining specific activity referring it to cell extracts or total (recombinant or not) protein [45,46]. Also, there are clear indications that the solubility enhancement under such conditions might eventually be associated to an increase of soluble aggregates [47]. Interestingly, lower protein yields obtained during chaperone co-production result in higher enzymatic activity in cell extracts and enhanced solubility, as observed by cyclodextrin glycosyltransferase [48] and mouse endostatin and human lysozyme respectively [42].

Furthermore, fine analyses of solubility in combination with other more useful parameters such as yield of soluble polypeptide or the biological activity reveal intriguing physiological events. For instance, co-production of the DnaKJ chaperone pair along with a target recombinant protein indeed favours solubility but at expenses of protein quality and yield [20]. In fact, enhancing the intracellular levels of DnaK, alone or within distinct chaperone sets (a common strategy to increase solubility) [40], dramatically diminishes protein stability through the stimulation of Lon- and ClpP-dependent proteolysis of inclusion body polypeptides [20]. In this regard, both yield and quality of a model recombinant GFP and other unrelated proteins are largely enhanced in DnaK$^-$ mutants [20,25,26], in which the solubility percent value is, as expected, lower than in wilt type hosts. More intriguingly, plotting solubility percent data versus protein yield of functional quality renders extremely good but negative correlations, under different genetic backgrounds [20] or production conditions [49]. Preliminary data about non bacterial protein production systems from our group obtained by M. Martínez Alonso (not shown) indicate that such a negative correlation between yield (or quality) and solubility could be a general issue. Therefore, when designing a protein production process the most pertinent strategy should be chosen depending on what parameter (yield, quality or solubility) is the most relevant to the final use of the protein. Eventually, recombinant protein solubility could be merely dependent on the intracellular concentration of the recombinant protein itself, what would ultimately fit with the enhanced solubility observed at low growth rates, low temperatures and weak doses of gene expression inductor [36,41].

There are still exciting issues regarding bacterial inclusion bodies that deserve full scientific attention, such as the solving of the inner molecular organization that allows the occurrence of proper folded species within a general amyloid-like aggregate pattern [50,51]. Also, the sequence-dependent nature of protein aggregation [52-54] is still poorly known from a mechanistic point of view. From the biotechnological side, it is widely accepted that production of aggregation-prone protein triggers cell responses to conformational stress [55-58], irrespective of the host used as cell factory [59]. If such set of physiological responses cannot be efficiently controlled, enhancing protein solubility without renouncing to protein quality might be then a mirage. Surfing the complex network of cell activities that regulate protein aggregation (for instance, through rational metabolic engineering) could be a choice strategy to approach the production of soluble and high quality recombinant proteins. For such a more gentle use of cell factories, a deeper comprehension of the recombinant cell physiology and quality control system is urgently needed.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MMA, NGM and EGF have equally contributed to this work.

Acknowledgements

We appreciate the financial support received for the design and production of recombinant proteins for biomedical applications from MEC (PETRI 95-0947.OP.02, BIO2005-23732-E, BIO2007-61194), AGAUR (2005SGR-00956), CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Spain, and from the European Science Foundation, which is also funded by the European Commission, Contract no. ERAS-CT-2003-980409 of the Sixth Framework Programme.

References

1. de Marco A: **Minimal information: an urgent need to assess the functional reliability of recombinant proteins used in biological experiments.** *Microb Cell Fact* 2008, **7**:20.
2. Baneyx F, Mujacic M: **Recombinant protein folding and misfolding in Escherichia coli.** *Nat Biotechnol* 2004, **22**:1399-1408.
3. Garcia-Fruitos E, Gonzalez-Montalban N, Morell M, Vera A, Ferraz RM, Aris A, *et al.*: **Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins.** *Microb Cell Fact* 2005, **4**:27.

4. Arie JP, Miot M, Sassoon N, Betton JM: **Formation of active inclusion bodies in the periplasm of Escherichia coli.** *Mol Microbiol* 2006, **62:**427-437.
5. Nahalka J, Vikartovska A, Hrabarova E: **A crosslinked inclusion body process for sialic acid synthesis.** *J Biotechnol* 2008, **134:**146-153.
6. Nahalka J: **Physiological aggregation of maltodextrin phosphorylase from Pyrococcus furiosus and its application in a process of batch starch degradation to alpha-D: -glucose-1-phosphate.** *J Ind Microbiol Biotechnol* 2008, **35:**219-223.
7. Nahalka J, Dib I, Nidetzky B: **Encapsulation of Trigonopsis variabilis D-amino acid oxidase and fast comparison of the operational stabilities of free and immobilized preparations of the enzyme.** *Biotechnol Bioeng* 2008, **99:**251-260.
8. Nahalka J, Nidetzky B: **Fusion to a pull-down domain: a novel approach of producing Trigonopsis variabilisD-amino acid oxidase as insoluble enzyme aggregates.** *Biotechnol Bioeng* 2007, **97:**454-461.
9. Nahalka J, Gemeiner P, Bucko M, Wang PG: **Bioenergy beads: a tool for regeneration of ATP/NTP in biocatalytic synthesis.** *Artif Cells Blood Substit Immobil Biotechnol* 2006, **34:**515-521.
10. Doglia SM, Ami D, Natalello A, Gatti-Lafranconi P, Lotti M: **Fourier transform infrared spectroscopy analysis of the conformational quality of recombinant proteins within inclusion bodies.** *Biotechnol J* 2008, **3:**193-201.
11. Garcia-Fruitos E, Aris A, Villaverde A: **Localization of functional polypeptides in bacterial inclusion bodies.** *Appl Environ Microbiol* 2007, **73:**289-294.
12. Jevsevar S, Gaberc-Porekar V, Fonda I, Podobnik B, Grdadolnik J, Menart V: **Production of nonclassical inclusion bodies from which correctly folded protein can be extracted.** *Biotechnol Prog* 2005, **21:**632-639.
13. 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.
14. Ventura S, Villaverde A: **Protein quality in bacterial inclusion bodies.** *Trends Biotechnol* 2006, **24:**179-185.
15. Gonzalez-Montalban N, Garcia-Fruitos E, Villaverde A: **Recombinant protein solubility-does more mean better?** *Nat Biotechnol* 2007, **25:**718-720.
16. Vera A, Gonzalez-Montalban N, Aris A, Villaverde A: **The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures.** *Biotechnol Bioeng* 2007, **96:**1101-1106.
17. Martinez-Alonso M, Vera A, Villaverde A: **Role of the chaperone DnaK in protein solubility and conformational quality in inclusion body-forming Escherichia coli cells.** *FEMS Microbiol Lett* 2007, **273:**187-195.
18. Vera A, Aris A, Carrio M, Gonzalez-Montalban N, Villaverde A: **Lon and ClpP proteases participate in the physiological disintegration of bacterial inclusion bodies.** *J Biotechnol* 2005, **119:**163-171.
19. Rinas U, Hoffmann F, Betiku E, Estape D, Marten S: **Inclusion body anatomy and functioning of chaperone-mediated in vivo inclusion body disassembly during high-level recombinant protein production in Escherichia coli.** *J Biotechnol* 2007, **127:**244-257.
20. Garcia-Fruitos E, Martinez-Alonso M, Gonzalez-Montalban N, Valli M, Mattanovich D, Villaverde A: **Divergent Genetic Control of Protein Solubility and Conformational Quality in Escherichia coli.** *J Mol Biol* 2007, **374:**195-205.
21. Alcala P, Feliu JX, Aris A, Villaverde A: **Efficient accommodation of recombinant, foot-and-mouth disease virus RGD peptides to cell-surface integrins.** *Biochem Biophys Res Commun* 2001, **285:**201-206.
22. Gonzalez-Montalban N, Natalello A, Garcia-Fruitos E, Villaverde A, Doglia SM: **In situ protein folding and activation in bacterial inclusion bodies.** *Biotechnol Bioeng* 2008, **100:**797-802.
23. de Marco A, Schroedel A: **Characterization of the aggregates formed during recombinant protein expression in bacteria.** *BMC Biochem* 2005, **6:**10.
24. Martinez-Alonso M, Gonzalez-Montalban N, Garcia-Fruitos E, Villaverde A: **The functional quality of soluble recombinant polypeptides produced in Escherichia coli is defined by a wide conformational spectrum.** *Appl Environ Microbiol* 2008.
25. 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.
26. Gonzalez-Montalban N, Garcia-Fruitos E, Ventura S, Aris A, Villaverde A: **The chaperone DnaK controls the fractioning of functional protein between soluble and insoluble cell fractions in inclusion body-forming cells.** *Microb Cell Fact* 2006, **5:**26.
27. Vallejo LF, Rinas U: **Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins.** *Microb Cell Fact* 2004, **3:**11.
28. Tsumoto K, Umetsu M, Kumagai I, Ejima D, Arakawa T: **Solubilization of active green fluorescent protein from insoluble particles by guanidine and arginine.** *Biochem Biophys Res Commun* 2003, **312:**1383-1386.
29. Tsumoto K, Umetsu M, Kumagai I, Ejima D, Philo JS, Arakawa T: **Role of arginine in protein refolding, solubilization, and purification.** *Biotechnol Prog* 2004, **20:**1301-1308.
30. Umetsu M, Tsumoto K, Nitta S, Adschiri T, Ejima D, Arakawa T, *et al.*: **Nondenaturing solubilization of beta2 microglobulin from inclusion bodies by L-arginine.** *Biochem Biophys Res Commun* 2005, **328:**189-197.
31. Peternel S, Grdadolnik J, Gaberc-Porekar V, Komel R: **Engineering inclusion bodies for non denaturing extraction of functional proteins.** *Microb Cell Fact* 2008, **7:**34.
32. Peternel S, Jevsevar S, Bele M, Gaberc-Porekar V, Menart V: **New properties of inclusion bodies with implications for biotechnology.** *Biotechnol Appl Biochem* 2008, **49:**239-246.
33. Jung KH: **Enhanced enzyme activities of inclusion bodies of recombinant beta-galactosidase via the addition of inducer analog after L-arabinose induction in the araBAD promoter system of Escherichia coli.** *J Microbiol Biotechnol* 2008, **18:**434-442.
34. Jung KH, Yeon JH, Moon SK, Choi JH: **Methyl alpha-D-glucopyranoside enhances the enzymatic activity of recombinant beta-galactosidase inclusion bodies in the araBAD promoter system of Escherichia coli.** *J Ind Microbiol Biotechnol* 2008, **35:**695-701.
35. Marston FA: **The purification of eukaryotic polypeptides synthesized in Escherichia coli.** *Biochem J* 1986, **240:**1-12.
36. Sorensen HP, Mortensen KK: **Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli.** *Microb Cell Fact* 2005, **4:**1.
37. Butt TR, Edavettal SC, Hall JP, Mattern MR: **SUMO fusion technology for difficult-to-express proteins.** *Protein Expr Purif* 2005, **43:**1-9.
38. Mansell TJ, Fisher AC, DeLisa MP: **Engineering the protein folding landscape in gram-negative bacteria.** *Curr Protein Pept Sci* 2008, **9:**138-149.
39. Wall JG, Pluckthun A: **Effects of overexpressing folding modulators on the in vivo folding of heterologous proteins in Escherichia coli.** *Curr Opin Biotechnol* 1995, **6:**507-516.
40. de Marco A, Deuerling E, Mogk A, Tomoyasu T, Bukau B: **Chaperone-based procedure to increase yields of soluble recombinant proteins produced in E. coli.** *BMC Biotechnol* 2007, **7:**32.
41. Sorensen HP, Mortensen KK: **Advanced genetic strategies for recombinant protein expression in Escherichia coli.** *J Biotechnol* 2005, **115:**113-128.
42. Nishihara K, Kanemori M, Yanagi H, Yura T: **Overexpression of trigger factor prevents aggregation of recombinant proteins in Escherichia coli.** *Appl Environ Microbiol* 2000, **66:**884-889.
43. Song L, Yuan HJ, Coffey L, Doran J, Wang MX, Qian S, *et al.*: **Efficient expression in E. coli of an enantioselective nitrile hydratase from Rhodococcus erythropolis.** *Biotechnol Lett* 2008, **30:**755-762.
44. de Marco A, De M V: **Bacteria co-transformed with recombinant proteins and chaperones cloned in independent plasmids are suitable for expression tuning.** *J Biotechnol* 2004, **109:**45-52.
45. Lee DH, Kim MD, Lee WH, Kweon DH, Seo JH: **Consortium of fold-catalyzing proteins increases soluble expression of cyclohexanone monooxygenase in recombinant Escherichia coli.** *Appl Microbiol Biotechnol* 2004, **63:**549-552.
46. Liu D, Schmid RD, Rusnak M: **Functional expression of Candida antarctica lipase B in the Escherichia coli cytoplasm – a screening system for a frequently used biocatalyst.** *Appl Microbiol Biotechnol* 2006, **72:**1024-1032.

47. Haacke A, Fendrich G, Ramage P, Geiser M: **Chaperone overexpression in Escherichia coli: Apparent increased yields of soluble recombinant protein kinases are due mainly to soluble aggregates.** *Protein Expr Purif* 2008.
48. Kim SG, Kweon DH, Lee DH, Park YC, Seo JH: **Coexpression of folding accessory proteins for production of active cyclodextrin glycosyltransferase of Bacillus macerans in recombinant Escherichia coli.** *Protein Expr Purif* 2005, **41**:426-432.
49. Martinez-Alonso M, Garcia-Fruitos E, Villaverde A: **Yield, solubility and conformational quality of soluble proteins are not simultaneously favored in recombinant Escherichia coli.** *Biotechnol Bioeng* 2008, **101**:1353-1358.
50. 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.
51. Wang L, Maji SK, Sawaya MR, Eisenberg D, Riek R: **Bacterial inclusion bodies contain amyloid-like structure.** *PLoS Biol* 2008, **6**:e195.
52. Speed MA, Wang DI, King J: **Specific aggregation of partially folded polypeptide chains: the molecular basis of inclusion body composition.** *Nat Biotechnol* 1996, **14**:1283-1287.
53. Morell M, Bravo R, Espargaro A, Sisquella X, Aviles FX, Fernandez-Busquets X, *et al.*: **Inclusion bodies: specificity in their aggregation process and amyloid-like structure.** *Biochim Biophys Acta* 2008, **1783**:1815-1825.
54. Rajan RS, Illing ME, Bence NF, Kopito RR: **Specificity in intracellular protein aggregation and inclusion body formation.** *Proc Natl Acad Sci USA* 2001, **98**:13060-13065.
55. Jurgen B, Lin HY, Riemschneider S, Scharf C, Neubauer P, Schmid R, *et al.*: **Monitoring of genes that respond to overproduction of an insoluble recombinant protein in Escherichia coli glucose-limited fed-batch fermentations.** *Biotechnol Bioeng* 2000, **70**:217-224.
56. Goff SA, Goldberg AL: **Production of abnormal proteins in E. coli stimulates transcription of lon and other heat shock genes.** *Cell* 1985, **41**:587-595.
57. Allen SP, Polazzi JO, Gierse JK, Easton AM: **Two novel heat shock genes encoding proteins produced in response to heterologous protein expression in Escherichia coli.** *J Bacteriol* 1992, **174**:6938-6947.
58. Lesley SA, Graziano J, Cho CY, Knuth MW, Klock HE: **Gene expression response to misfolded protein as a screen for soluble recombinant protein.** *Protein Eng* 2002, **15**:153-160.
59. Gasser B, Saloheimo M, Rinas U, Dragosits M, Rodriguez-Carmona E, Baumann K, *et al.*: **Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview.** *Microb Cell Fact* 2008, **7**:11.

3.4. Article 4

Rehosting of bacterial chaperones for high-quality protein production.

Mónica Martínez-Alonso, Verónica Toledo-Rubio, Rob Noad, Ugutz Unzueta, Neus Ferrer-Miralles, Polly Roy, Antonio Villaverde.

Applied and Environmental Microbiology, Vol. 75, No 24, 7850-4 (December 2009).

Although coproduction of folding modulators has been a widely tested strategy to improve soluble protein production in bacteria, the results obtained have been controversial. Promising sets of folding modulators usually include the chaperone pair DnaKJ. However, besides their folding activity DnaKJ also act as proteolytic enhancers in cooperation with bacterial proteases, which may account, at least partially, for some of the negative results obtained when coproducing folding modulators in an attempt to increase protein solubility. Since the DnaKJ pair has been widely conserved in evolution, we envisaged rehosting of this chaperone set as a way to uncouple their valuable foldase activity from the associated proteolysis observed in bacterial systems. Using again our recombinant GFP as a model protein, we constructed baculovirus vectors that allowed its production either alone or together with the chaperone pair upon infection of insect cells. Deposition of the target protein in insoluble but fluorescent clusters was in agreement with solubility and quality not being coincident parameters. When we coproduced the chaperone pair and evaluated solubility and conformational quality of the reporter protein, we observed enhanced yield and biological activity. Also, stability was increased compared to when the protein was produced in *E. coli*, indicative of no DnaK-mediated proteolysis. However, in agreement with our observations for bacterial systems, yield and quality of the recombinant protein could not be increased in parallel in this eukaryotic system either. Positive effects of this set of bacterial folding modulators were also observed for the production of three other different proteins.

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Dec. 2009, p. 7850–7854 Vol. 75, No. 24
0099-2240/09/$12.00 doi:10.1128/AEM.01532-09

Rehosting of Bacterial Chaperones for High-Quality Protein Production▽

Mónica Martínez-Alonso,[1,2] Verónica Toledo-Rubio,[1,2] Rob Noad,[3]† Ugutz Unzueta,[1,2] Neus Ferrer-Miralles,[1,2] Polly Roy,[3] and Antonio Villaverde[1,2]*

Institute for Biotechnology and Biomedicine and Department of Genetics and Microbiology, Universitat Autònoma de Barcelona, Bellaterra, 08193 Barcelona, Spain[1]; CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Bellaterra, 08193 Barcelona, Spain[2]; and Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, United Kingdom[3]

Received 30 June 2009/Accepted 2 October 2009

Coproduction of DnaK/DnaJ in *Escherichia coli* enhances solubility but promotes proteolytic degradation of their substrates, minimizing the yield of unstable polypeptides. Higher eukaryotes have orthologs of DnaK/DnaJ but lack the linked bacterial proteolytic system. By coexpression of DnaK and DnaJ in insect cells with inherently misfolding-prone recombinant proteins, we demonstrate simultaneous improvement of soluble protein yield and quality and proteolytic stability. Thus, undesired side effects of bacterial folding modulators can be avoided by appropriate rehosting in heterologous cell expression systems.

The production of recombinant proteins is an essential instrument in biotechnology and biomedicine, but it has not been fully optimized in all the cell systems commonly used as factories. An important fraction of recombinant proteins fail to reach their native conformation, triggering diverse cell stress responses, and are often deposited as insoluble aggregates (usually referred to as inclusion bodies) or degraded by cellular proteases (11). Furthermore, the quality of soluble recombinant proteins is frequently compromised by the occurrence of soluble aggregates (5) and, in general, by their conformational heterogeneity (7, 17). In the widely used bacterium *Escherichia coli*, many strategies have been explored to enhance recombinant protein solubility. Among them, the coproduction of folding modulators, believed to be limiting in cells actively producing recombinant proteins, has attracted special attention (14). However, the efficacy of such an approach has been highly controversial. While some authors have reported enhancement of protein solubility by coproducing specific chaperones or chaperone sets, many others have found poor improvement or even impairment of protein solubility and yield. In particular, the chaperone DnaK and its cochaperone DnaJ have been very frequently incorporated as folding modulators, being present in essentially all the promising chaperone sets (6, 8). *E. coli* DnaK is the major cytosolic chaperone that exhibits folding and disaggregase activities (Table 1). Moreover, it is a negative regulator of the heat shock response in cooperation with DnaJ by promoting the conformation-dependent FtsH-mediated proteolytic inactivation of the stress-activated RNA polymerase subunit σ^{32} (Table 1) (22). This subunit controls the expression of the heat shock genes whose products cope with conformational stress, thus increasing cell survival. DnaK and DnaJ also deliver misfolded or conformationally abnormal proteins (including recombinant proteins) to the Lon and ClpP proteases, resulting in reduced protein yields (13, 23, 24). Misfolding-prone green fluorescent protein (GFP) fusions synthesized in DnaK knockouts (10) rendered higher protein yields but reduced recombinant protein solubility compared to those in wild-type cells. In contrast, the overexpression of *dnaK* and *dnaJ* genes in *E. coli* enhanced the proportion of soluble recombinant proteins by stimulating Lon- and ClpP-mediated proteolysis of aggregated proteins, reducing overall protein yields (10). Very recently, the molecular basis of the DnaKJ-mediated proteolytic enhancement has been solved by fine dissection of the interaction between DnaKJ and σ^{32}. The binding of DnaK and DnaJ to distinct sites of the transcription factor promotes conformational modifications that expose a unique target site for the inactivating protease FtsH (22). Such conformational effect of the substrate seems to be the mechanistic platform of both foldase and disaggregase activities exhibited by DnaK/DnaJ on their substrates, including RepA and unfolded proteins (22). In recombinant *E. coli* cells, DnaK/DnaJ could mediate conformational perturbations of misfolded proteins at the surface of inclusion bodies, where DnaK localizes (4), exposing them to the stress proteases Lon and ClpP (10) and promoting digestion during refolding attempts. Such a dual role of DnaK/DnaJ in stimulating protein folding but also enhancing degradation of protease-sensitive protein species could be the cause of the controversial data obtained from their use as folding modulators and the reason for the only transient rise of soluble protein species refolded in vivo from bacterial inclusion bodies (3).

For the improved production of recombinant proteins, it would then be desirable to keep the DnaK/DnaJ foldase activity but eliminate the enhanced proteolysis indirectly promoted by these chaperones. Since DnaK/DnaJ are members of the highly conserved Hsp70 family, we anticipated that their foldase activity would be retained in organisms other than *E.*

* Corresponding author. Mailing address: Institute for Biotechnology and Biomedicine, Universitat Autònoma de Barcelona, Bellaterra, 08193 Barcelona, Spain. Phone: 34 935813086. Fax: 34 935812011. E-mail: antoni.villaverde@uab.es.
† Present address: Department of Pathology and Infectious Diseases, Royal Veterinary College, Hawkshead Lane, Hatfield, Hertfordshire AL9 7TA, United Kingdom.
▽ Published ahead of print on 9 October 2009.

TABLE 1. Biological activities of the *E. coli* chaperone DnaK

Activity	Cochaperone(s)	Substrate(s)	Reference(s)	Evolutionary relationship(s)
Foldase	DnaJ, GrpE	Unfolded/misfolded proteins	15	Hsp70 family member
Holding chaperone	Hsp31	Unfolded/misfolded proteins	21	Presumed Hsp70 family member
Disaggregase	ClpB, IbpAB	Protein aggregates	20, 25	Species specificity of Hsp104/Hsp70 and ClpB/DnaK cooperativity; a disaggregation activity has not yet been identified in mammalian cells
Negative regulator of the heat shock response (proteolytic enhancer)	DnaJ, FtsH	Transcription factor σ^{32}	22	Not reported
Proteolytic enhancer	DnaJ, Lon, ClpP	Abnormal proteins	13	Not reported
Proteolytic enhancer	DnaJ, Lon, ClpP	Recombinant proteins	10	Not reported

coli, but that was not the case for the proteolytic stimulation of recombinant proteins, which is highly dependent on the *E. coli* proteases Lon and ClpP (Table 1) (10). Under the assumption that insect cell proteases would not recognize the target sites exposed by the activity of DnaKJ (no Lon and ClpP orthologs have so far been identified in insects), we rehosted the *E. coli* DnaKJ pair for coproduction along with recombinant protein by using the baculovirus expression system. For these studies, we used a proteolytically unstable GFP (mGFP), already characterized for recombinant protein quality analysis in bacteria (10), and generated two transfer vectors designed to express mGFP either alone or together with the bacterial chaperones DnaK and DnaJ, using the vector pAcAB4 (2). mGFP and *dnaJ* genes were placed under the control of the p10 promoter, and *dnaK* was placed under the control of the polyhedrin promoter. Each transfer vector was cotransfected with Bsu36I-linearized viral DNA BAC10:KO1629 (27) into *Spodoptera frugiperda* Sf9 cells to obtain recombinant baculoviruses. Individual clones were plaque purified and further amplified, and the titers of the clones were determined before they were analyzed for mGFP expression. In the absence of chaperone coproduction, mGFP gene expression resulted in mild cytoplasmic fluorescence within 24 h postinfection (hpi), with punctate distribution indicative of inclusion body formation (Fig. 1A). Coexpression of mGFP along with DnaK and DnaJ rendered much higher and homogeneously distributed fluorescence. The degree of fluorescence was largely stable at least up to around 72 h and was reproducible at all the tested multiplicities of infection (MOIs) (ranging from 0.1 to 10 [Fig. 1B]). The enhancement of fluorescence in infected cells coexpressing *dnaK* and *dnaJ* was further confirmed by flow cytometry (Fig. 2A). These results were in marked contrast to those of similar experiments with *E. coli*, where coexpression of DnaK and DnaJ with mGFP resulted in much lower fluorescence levels than in the control cells (10).

Also, the coproduction of DnaK and DnaJ resulted in a dramatic, sixfold increase of the total amount of mGFP, indicating the absence of proteolysis. Being more important in the insoluble cell fraction, this still represented an almost twofold yield enhancement for the soluble version of the

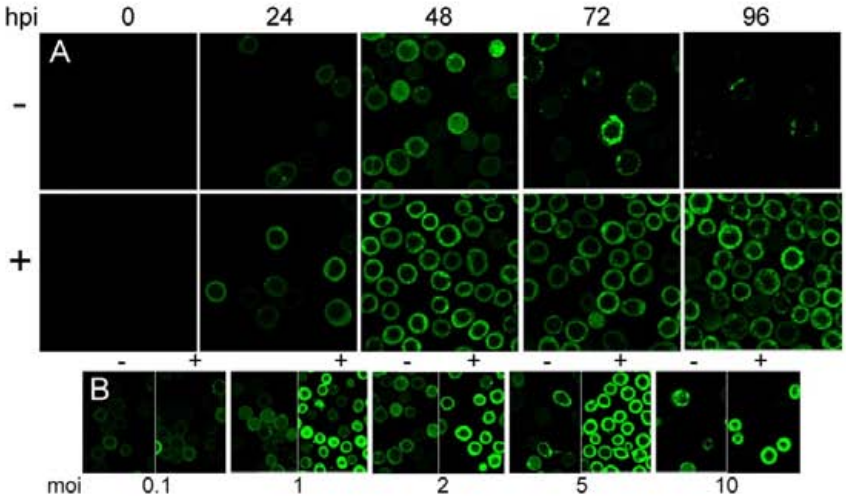

FIG. 1. Confocal microscopy images of baculovirus-infected Sf9 cells taken in a Leica TCS SP2 AOBS microscope. Batches of 1×10^6 cells were seeded on glass-bottom dishes, supplemented with 5% fetal calf serum, and infected at MOIs of 0.1, 1, 2, 5, and 10 with recombinant baculoviruses expressing mGFP in the presence (+) or absence (−) of DnaK and DnaJ. The time course experiment for a set MOI of 2 is shown in panel A, while cultured cells 48 h after infection at different MOIs are presented in panel B. The settings used to capture the images in panel B were maintained for direct comparison of the images.

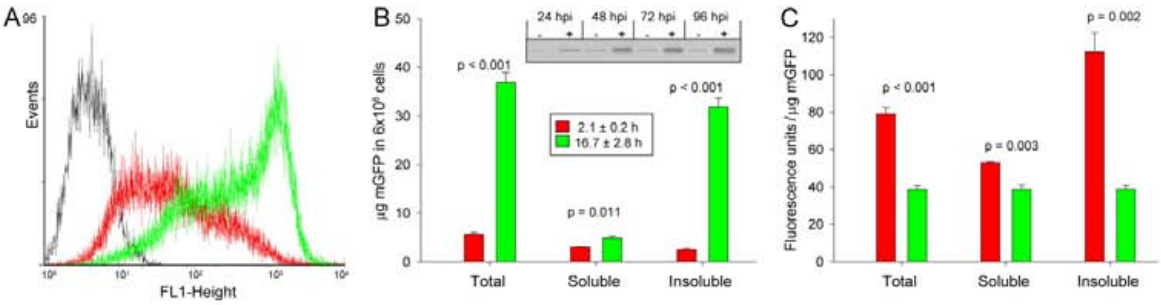

FIG. 2. (A) Flow cytometry analysis of baculovirus-infected Sf9 cells producing mGFP in the absence (red plot) or presence (green plot) of bacterial chaperones DnaK and DnaJ. Sf9 cell cultures were set up at a density of 2×10^6 cells/ml, supplemented with 5% fetal calf serum, and infected at an MOI of 2. Cells were harvested at 72 hpi and rinsed with cold phosphate-buffered saline (PBS). Flow cytometry analyses were performed with intact cells resuspended in PBS on a FACSCalibur system (Becton Dickinson). The fluorescence emission in the FL1 channel was analyzed using WinMDI 2.9 software. Uninfected cells (black plot) were used as a negative control, and measurements were recorded for three independent replicas. (B) Protein amounts in total, soluble, and insoluble cell fractions in the absence (red bars) or presence (green bars) of bacterial chaperones DnaK and DnaJ were estimated by Western blot analysis after disruption of cells harvested at 72 hpi. Sf9 cells were disrupted in lysis buffer (extraction buffer [50 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM EDTA] containing 1% Tergitol NP-9 [Sigma] and Roche's protease inhibitor cocktail [catalog no. 11836170001]) on ice for 30 min and Dounce homogenized. Cell lysates were treated with DNase (25 µg/ml) and $MgSO_4$ (10 mM), and soluble and insoluble cell fractions were separated by centrifugation at 9,500 × *g* for 10 min. The insoluble cell pellet was washed with extraction buffer containing 0.5% Triton X-100 and resuspended in extraction buffer. The inset box shows the estimated half-life for mGFP in the absence (red bars) or presence (green bars) of DnaKJ after protein synthesis arrest at 24 hpi by the addition of cycloheximide at a final concentration of 100 µg/ml. Western blot analysis of a time course experiment showing mGFP production in Sf9 cells at different hpi is depicted on the inset graph. Material from the same number of cells was loaded into the gels for protein determination. (C) Specific fluorescence of mGFP produced in the absence (red bars) or presence (green bars) of bacterial chaperones DnaK and DnaJ at 72 hpi. Fluorescence emissions of lysates and soluble and insoluble cell fractions were measured in triplicate, with no further treatment, using a Cary Eclipse fluorescence spectrophotometer (Varian, Inc.). Fluorescence data were combined with protein amounts to obtain the specific fluorescence of mGFP, defined as fluorescence units per µg of mGFP. The significance of the differences between data values determined in the absence and in the presence of DnaK/DnaJ is indicated through *P* values from an analysis of variance test in panels B and C.

protein (Fig. 2B). Consistent with previous studies, correctly folded polypeptides were deposited in inclusion bodies (9), confirming that solubility is not a good indicator of a successful protein production process (12). When the stability of mGFP produced alone or in the presence of *E. coli* DnaK and DnaJ was analyzed, mGFP's half-life increased from 2.1 ± 0.2 h to 16.7 ± 2.8 h (Fig. 2B, inset), indicative of an increased resistance to proteolytic degradation. This is the opposite of what has been observed in bacterial cells, where the half-life of mGFP was reduced from 5.9 ± 0.5 h to 1.9 ± 0.3 h in the presence of DnaKJ (10). This result confirms not only the lack of DnaKJ-mediated mGFP proteolysis in insect cells but also that the activity of these chaperones as folding mediators is able to protect from degradation by host cell proteases, probably by fully completing the folding process. In the absence of DnaK and DnaJ coexpression, mGFP demonstrated a clear heterogeneity in specific fluorescence when soluble and insoluble versions were compared (Fig. 2C). This is in marked contrast to what was observed when the chaperones were coexpressed, where specific fluorescence emissions of soluble and insoluble mGFPs were remarkably similar. On the other hand, the overall mGFP specific fluorescence in the presence of DnaK/DnaJ was lower than in the absence of these chaperones. This is consistent with recent observations reporting that higher protein yield in a production process necessarily results in a decrease of conformational quality (10, 16, 18).

To determine whether the positive effect of DnaK and DnaJ in assisting protein folding could be extended to the expression of other recombinant proteins in insect cells, we also tested coexpression of these proteins with foot-and-mouth disease virus (FMDV) VP1 and VP2 capsid proteins and human alpha-galactosidase A. Although potential gene dosage effects due to coinfection of the virus expressing the chaperones with that expressing the recombinant protein prevented us from a fine comparative analysis of total protein amounts, unlike in our studies with mGFP, we were able to assess changes in solubility that correlated with recombinant protein expression. In general, total protein amounts were comparable when produced alone or with DnaK/DnaJ, indicating the absence of important DnaK-induced proteolysis (Fig. 3A to C). For two of the three proteins (VP1 and alpha-galactosidase), the amounts of soluble protein were significantly enhanced, reaching up to more than twofold for the enzyme, in which the solubility also improved by more than 250% (Fig. 3B). For the third protein (FMDV VP2), there was no increase in the yield of soluble protein but in the protein quality. VP2 has a tendency to spontaneously form unwanted oligomers. When coexpressed with DnaK/DnaJ, formation of these aggregates was completely cleared (Fig. 3C). The quality of soluble alpha-galactosidase, which is protease sensitive, was also enhanced, with degradation largely minimized, indicative of DnaK/DnaJ-promoted protein stabilization (Fig. 3B). At least in mGFP, chaperone coproduction did not affect the efficiency of recovery from cell extracts in further purification processes (not shown), and we do not have any experimental data suggesting that this could happen with other proteins.

In contrast to the widespread use of folding modulators in bacteria (14), very few attempts have been made to improve cytoplasmic protein production in insect cells with the aid of

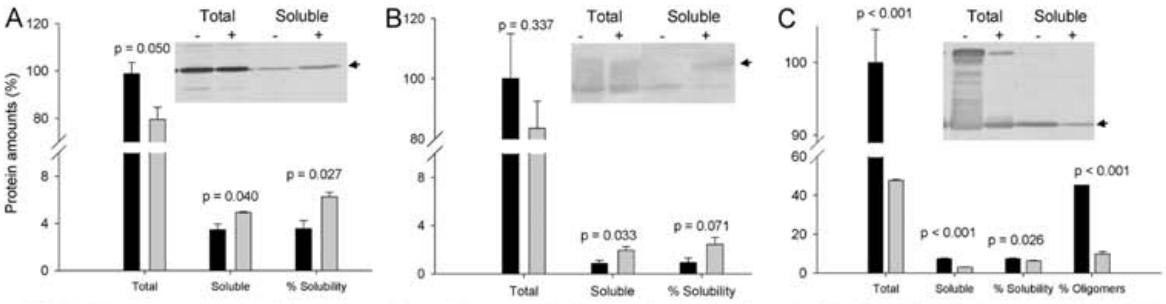

FIG. 3. Total amounts and soluble amounts of recombinant proteins in Sf9 cells coinfected at an MOI of 2 with baculoviruses encoding FMDV VP1 (A), human alpha-galactosidase (B), and FMDV VP2 (C) proteins (black bars), compared with protein data obtained during coinfection with a DnaK/DnaJ-encoding baculovirus, also at an MOI of 2 (gray bars). Data were obtained in triplicate at 72 hpi, and they were compared to the total protein amounts observed in the absence of chaperones. Percentages of solubility and VP2 oligomer occurrence are also included. Arrows indicate the positions of the full-length recombinant proteins. The significance of the differences between data values determined in the absence and in the presence of DnaK/DnaJ is indicated through *P* values from an analysis of variance test.

chaperones. In particular, the human versions of DnaK and DnaJ, namely, Hsp70 and Hsp40, respectively, have been coproduced along with target proteins (1, 19, 26), with all cases reporting slight increases in solubility, poor gain of yield, if any, and no references to protein stability and quality. Interestingly, gain of solubility has also been generally described for bacteria when the *dnaK* gene is coexpressed (6). Our data demonstrate that bacterial DnaK and DnaJ do have a significant effect on protein solubility in insect cells, suggesting that these chaperones may target a broader subset of misfolded protein substrates in insect cells than human Hsp70 and Hsp40. This highlights the relevance of function selection in multifunctional folding modulators for use in heterologous hosts. The results presented here strongly support chaperone rehosting as a new concept for high-quality recombinant protein production in insect cells that permits separation of the undesirable effects observed in *E. coli* from the valuable foldase activity. Coexpression of DnaK and DnaJ in insect cells dramatically enhances protein yield, proteolytic stability, protein solubility, and global biological activity, with unusually mild negative effects on protein quality.

We appreciate financial support through grants BIO2007-61194 and EUI2008-03610 (MICINN) to A.V. and grant BB/C504735/1 (BBSRC) to P.R. We also acknowledge the support of the CIBER de Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Spain. M.M.-A. is the recipient of a predoctoral fellowship from MEC, Spain. A.V. received an ICREA ACADEMIA award (Catalonia, Spain).

We thank Francisco Cortés, from Servei de Cultius Cel·lulars (UAB), for routine maintenance of insect cell cultures and both Servei de Microscòpia and Servei de Citometria (UAB) for helpful technical assistance.

REFERENCES

1. **Ailor, E., and M. J. Betenbaugh.** 1998. Overexpression of a cytosolic chaperone to improve solubility and secretion of a recombinant IgG protein in insect cells. Biotechnol. Bioeng. **58:**196–203.
2. **Belyaev, A. S., and P. Roy.** 1993. Development of baculovirus triple and quadruple expression vectors: co-expression of three or four bluetongue virus proteins and the synthesis of bluetongue virus-like particles in insect cells. Nucleic Acids Res. **21:**1219–1223.
3. **Carrio, M. M., and A. Villaverde.** 2001. Protein aggregation as bacterial inclusion bodies is reversible. FEBS Lett. **489:**29–33.
4. **Carrio, M. M., and A. Villaverde.** 2005. Localization of chaperones DnaK and GroEL in bacterial inclusion bodies. J. Bacteriol. **187:**3599–3601.
5. **de Marco, A., and A. Schroedel.** 2005. Characterization of the aggregates formed during recombinant protein expression in bacteria. BMC Biochem. **6:**10.
6. **de Marco, A.** 2007. Protocol for preparing proteins with improved solubility by co-expressing with molecular chaperones in *Escherichia coli.* Nat. Protoc. **2:**2632–2639.
7. **de Marco, A.** 2008. Minimal information: an urgent need to assess the functional reliability of recombinant proteins used in biological experiments. Microb. Cell Fact. **7:**20.
8. **de Marco, A., E. Deuerling, A. Mogk, T. Tomoyasu, and B. Bukau.** 2007. Chaperone-based procedure to increase yields of soluble recombinant proteins produced in *E. coli.* BMC Biotechnol. **7:**32.
9. **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.
10. **Garcia-Fruitos, E., M. Martinez-Alonso, N. Gonzalez-Montalban, M. Valli, D. Mattanovich, and A. Villaverde.** 2007. Divergent genetic control of protein solubility and conformational quality in *Escherichia coli.* J. Mol. Biol. **374:**195–205.
11. **Gasser, B., M. Saloheimo, U. Rinas, M. Dragosits, E. Rodriguez-Carmona, K. Baumann, M. Giuliani, E. Parrilli, P. Branduardi, C. Lang, D. Porro, P. Ferrer, M. L. Tutino, D. Mattanovich, and A. Villaverde.** 2008. Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. Microb. Cell Fact. **7:**11.
12. **Gonzalez-Montalban, N., E. Garcia-Fruitos, and A. Villaverde.** 2007. Recombinant protein solubility—does more mean better? Nat. Biotechnol. **25:**718–720.
13. **Jubete, Y., M. R. Maurizi, and S. Gottesman.** 1996. Role of the heat shock protein DnaJ in the lon-dependent degradation of naturally unstable proteins. J. Biol. Chem. **271:**30798–30803.
14. **Kolaj, O., S. Spada, S. Robin, and J. G. Wall.** 2009. Use of folding modulators to improve heterologous protein production in *Escherichia coli.* Microb. Cell Fact. **8:**9.
15. **Langer, T., C. Lu, H. Echols, J. Flanagan, M. K. Hayer, and F. U. Hartl.** 1992. Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding. Nature **356:**683–689.
16. **Martinez-Alonso, M., E. Garcia-Fruitos, and A. Villaverde.** 2008. Yield, solubility and conformational quality of soluble proteins are not simultaneously favored in recombinant *Escherichia coli.* Biotechnol. Bioeng. **101:**1353–1358.
17. **Martínez-Alonso, M., N. González-Montalbán, E. García-Fruitós, and A. Villaverde.** 2008. The functional quality of soluble recombinant polypeptides produced in *Escherichia coli* is defined by a wide conformational spectrum. Appl. Environ. Microbiol. **74:**7431–7433.
18. **Martinez-Alonso, M., N. Gonzalez-Montalban, E. Garcia-Fruitos, and A. Villaverde.** 2009. Learning about protein solubility from bacterial inclusion bodies. Microb. Cell Fact. **8:**4.
19. **Martinez-Torrecuadrada, J. L., S. Romero, A. Nunez, P. Alfonso, M. Sanchez-Cespedes, and J. I. Casal.** 2005. An efficient expression system for the production of functionally active human LKB1. J. Biotechnol. **115:**23–34.
20. **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.

21. **Mujacic, M., M. W. Bader, and F. Baneyx.** 2004. *Escherichia coli* Hsp31 functions as a holding chaperone that cooperates with the DnaK-DnaJ-GrpE system in the management of protein misfolding under severe stress conditions. Mol. Microbiol. **51:**849–859.
22. **Rodriguez, F., F. Arsene-Ploetze, W. Rist, S. Rudiger, J. Schneider-Mergener, M. P. Mayer, and B. Bukau.** 2008. Molecular basis for regulation of the heat shock transcription factor σ^{32} by the DnaK and DnaJ chaperones. Mol. Cell **32:**347–358.
23. **Sherman, M. Y., and A. L. Goldberg.** 1996. Involvement of molecular chaperones in intracellular protein breakdown. EXS **77:**57–78.
24. **Sherman, M. Y., and A. L. Goldberg.** 1992. Involvement of the chaperonin dnaK in the rapid degradation of a mutant protein in *Escherichia coli*. EMBO J. **11:**71–77.
25. **Weibezahn, J., B. Bukau, and A. Mogk.** 2004. Unscrambling an egg: protein disaggregation by AAA+ proteins. Microb. Cell Fact. **3:**1.
26. **Yokoyama, N., M. Hirata, K. Ohtsuka, Y. Nishiyama, K. Fujii, M. Fujita, K. Kuzushima, T. Kiyono, and T. Tsurumi.** 2000. Co-expression of human chaperone Hsp70 and Hsdj or Hsp40 co-factor increases solubility of over-expressed target proteins in insect cells. Biochim. Biophys. Acta **1493:**119–124.
27. **Zhao, Y., D. A. Chapman, and I. M. Jones.** 2003. Improving baculovirus recombination. Nucleic Acids Res. **31:**E6.