Chapter 15
Systems-Level Analysis of Protein Quality in Inclusion Body-Forming *Escherichia coli* Cells

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

Contents

Abstract Recombinant proteins produced in *Escherichia coli* often aggregate as amorphous masses of insoluble material known as inclusion bodies. Being quite homogeneous in their composition, inclusion bodies display amyloid-like properties such as sequence-dependent protein-protein interactions, seeding-driven deposition of their components and β-sheet intermolecular architecture. However, inclusion bodies formed by different proteins and enzymes also show important extents of native-like secondary structure and include significant proportions of properly folded, functional protein, which makes them suitable to be used in catalytic processes. Inclusion bodies are formed as a result of the incapability of the quality

A. Villaverde (✉)
Institut de Biotecnologia i de Biomedicina, Universitat Autònoma de Barcelona, Bellaterra, Barcelona, 08193 Spain
e-mail: antoni.villaverde@uab.es

S.Y. Lee (ed.), *Systems Biology and Biotechnology of Escherichia coli*,
DOI 10.1007/978-1-4020-9394-4_15,

control cell system to cope with the non physiological amounts of misfolding-prone proteins produced upon recombinant gene expression. Multiple cellular proteins involved in the quality control, namely chaperones and proteases, participate in their formation and co-ordinately determine the amount of aggregated protein, the size of aggregates and the main structural and functional properties of the embedded polypeptides, such as their inner molecular organization.

15.1 Recombinant Protein Production: An Historical Overview

The discovery of restriction enzymes in the 70s offered one of the most powerful tools in molecular biology, dramatically fuelling the progress of recombinant DNA technologies. Before the systematic use of restriction enzymes, genetic manipulation was restricted to poorly controlled genetic modifications such as those caused in bacterial genomes by bacteriophages and plasmids. Restriction enzymes permitted the isolation and cloning of genes and their regulated expression in heterologous cell hosts, such as bacteria. This allowed the production of polypeptides that, being of interest for scientific, pharmaceutical or industrial purposes, occurred in low amounts in their natural sources and therefore, were difficult to obtain. This simple gene-cloning-and-expression strategy offered a solid methodological background on which the modern biotechnology fully developed. The use of cells (mainly microbial) as biological systems for the regulated production of recombinant proteins (and also of natural substances of biotechnological interest) originated the "Cell Factory" concept. This notion, underlying any man-driven, cell-mediated production process in single cells, refers to the engineering of the cell's biosynthetic machinery and the supporting genetic programme for applied purposes.

In early DNA recombinant times, it was believed that recombinant protein production in microbial cells would be the source of any relevant protein of pharmacological interest with high added value (such as immunogens, hormones, enzymes and complex molecular assemblies such as virus-like particles) as well as enzymes of straightforward industrial applicability (such as lipases, glycosidases, proteases, etc.). Therefore, the implementation of recombinant DNA technologies was predicted to result into a dramatic positive impact in biotechnology and biomedicine, expanding the spectrum of protein products available in the market. However, those expectations were rapidly frustrated since generally the quality of recombinant proteins produced in bacteria was not comparable to that of those obtained from natural sources, and therefore those recombinant proteins were not suitable for use. Essentially, the major bottlenecks encountered during recombinant protein production are proteolytic digestion by cell proteases (Enfors 1992) and aggregation as insoluble protein deposits known as inclusion bodies (IBs) (Georgiou and Valax 1996, Marston 1986). Human insulin was among the first proteins for which the accumulation in morphologically discrete aggregates in the bacterial cytoplasm was shown

(Paul et al. 1983, Williams et al. 1982), which delayed its further development as a pharmaceutical in human therapy.

The majority of proteins deposited as inclusion bodies are produced in non-functional conformation, in particular if they are of eukaryotic origin and contain disulfide bonds and, thus, require solubilisation and refolding for generation of the biologically active version of the protein (Clark 2001, Fahnert et al. 2004, Jungbauer and Kaar 2007, Middelberg 2002, Vallejo and Rinas 2004). Although inclusion body formation in general leads to additional down-stream steps during protein production and purification, inclusion body based production processes involving solubilization and refolding are economically viable options for many biopharmaceuticals. For example, human insulin, nowadays produced as recombinant protein in tons per year quantities, is produced using two major routes (Walsh 2005). One route involves the production of proinsulin in form of inclusion bodies using *E. coli* as expression host with subsequent solubilization and refolding procedures. The other route involves the utilization of yeast-based expression systems leading to the secretion of a soluble proinsulin into the culture supernatant. Both routes are economically viable.

Both proteolysis and IB formation result from the unability of many recombinant proteins to reach their native conformation in recombinant cells, and the efforts addressed to minimize them have only resulted partially successful. Therefore, the number of recombinant proteins that have entered the biotechnological market represents only a very minor fraction of those that have ever been produced in heterologous cells. At least partially, the incomplete exploitation of the recombinant DNA technologies for protein production can be attributed to a limited understanding of the cell physiology under the non-physiological cellular conditions associated to recombinant gene expression.

15.1.1 Protein Production in Escherichia coli

Since the first experiments of gene cloning and expression, the gram-negative bacterium *Escherichia coli* has been universally used as a convenient host for protein production. Despite other heterologous hosts have been progressively incorporated into production processes (namely gram-positive bacteria, yeast, insect cells, mammalian cells, filamentous fungi and others) (Gasser et al. 2008), *E. coli* is still a main cell factory for protein production, essentially because of its high growth rate, the high cell densities reached in fed-batch cultures, the relatively inexpensive growth media, the deep knowledge of its genetics and the availability of diverse genetic tools, such as plasmids, transposons and viruses acting on this species. Although recombinant proteins can be obtained in the cell periplasm if fused to secretion peptides, *E. coli* is mostly used to produce proteins in the cytoplasm, which need to be recovered from cell extracts after cell disruption.

However, many recombinant proteins produced in *E. coli* are unable to reach their native conformation, especially if they have eukaryotic origins or require posttranslational modifications such as disulfide bridge formation for their folding,

and are rapidly degraded by cell proteases. Among the set of *E. coli* proteases, the ATP-dependent proteases Lon and ClpP are responsible for the degradation of most of the recombinant polypeptides (Maurizi 1992). Although the use of protease-deficient mutants as cell hosts has been explored as a method to enhance the stability of recombinant proteins (Baneyx and Georgiou 1991, Gottesman et al. 1997, Tomoyasu et al. 2001b), the issue appears progressively more complex as the physiology of *in vivo* protein folding is better understood. In this regard, cell proteases are an important arm of the quality control system, in which they act in cooperation with folding assistant proteins to survey conformational quality (Fig. 15.1). Therefore, minimizing proteolysis in protease deficient mutants leads to the accumulation of misfolded protein species as IBs (Garcia-Fruitos et al. 2007a, Rosen et al. 2002, Vera et al. 2005). During the growth of non recombinant *E. coli* cells at 37 °C, protein aggregation affects only a background fraction of cell proteins (Gonzalez-Montalban et al. 2006), being rather irrelevant from a quantitative point of view. However, both cell growth at high temperatures and the production of recombinant proteins cause aggregation of cell proteins or recombinant species, respectively, and trigger the expression of heat-shock genes (many of which encode chaperones and proteases

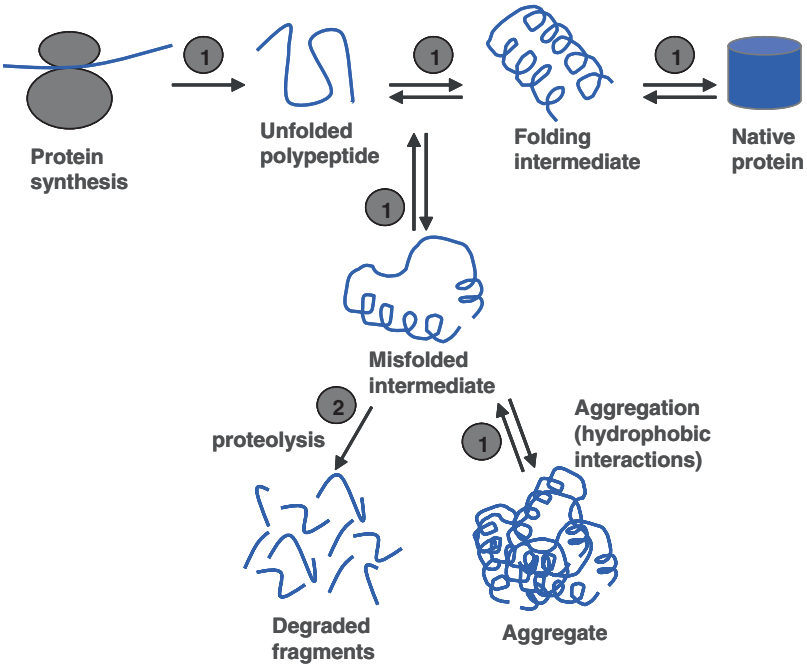

Fig. 15.1 Conventional model of protein folding, aggregation and proteolysis. A chain newly synthesized on a ribosome may fold to a native state, can aggregate or can be proteolysed. In living systems, environmental conditions and the quality control system highly regulate the transition between the different states. (1) Chaperones assist protein intermediates and misfolded proteins to reach their native state. (2) Proteases proteolyse misfolded proteins that have failed to reach a native conformation

for quality control). In recombinant cells, significant fractions of the recombinant protein are often found as IBs. This indicates that the quality control is inefficient in the processing of non-physiological amounts of heterologous proteins, what results in production processes rendering insoluble and biologically unuseful material. Recombinant protein misfolding and aggregation is one of the major concerns when facing *in vivo* protein production processes.

15.2 Molecular Basis of Protein Folding

Proteins have multiple and critical roles in all organisms, being the most abundant molecules in biological systems other than water. Protein folding is the process through which unfolded, nascent polypeptide chains convert into tightly folded compact structures with biological functions. Pioneering studies on protein folding by Anfinsen showed that the amino acid sequence of a protein encodes its functional three-dimensional structure (Anfinsen 1973). The underlying mechanism by which this complex process takes place is becoming progressively understood, because of the development of both physicochemical techniques and computational methods. In fact, understanding protein folding is not only relevant for biotechnological purposes but also to solve the molecular mechanisms responsible for conformational diseases such as Alzheimer, type II diabetes and Creutzfeldt-Jakob, among others.

Apparently, there is a common mechanism for the folding of the enormous spectrum of proteins in nature, irrespective of their native structure or amino acid sequence (Snow et al. 2002), in which the necessary information to reach a unique native state in a finite time is defined (Karplus 1997). Among the total number of possible conformations that a polypeptide could reach, finding a particular structure would take a length of time many orders of magnitude greater than the real time required for proteins to fold. This inconsistency, known as Levinthal paradox (Karplus 1997), has been solved with the development of the so called "new view" (Yon 2001), in which folding is described as a stochastic search of conformational space rather than as a series of mandatory structural transitions (Baldwin 1994, Dill and Chan 1997, Matagne and Dobson 1998, Wolynes et al. 1995). In essence, the inherent fluctuations in the conformation of an incompletely folded polypeptide enable the contact even of residues located at very different positions in the amino acid sequence. Therefore, as correct (native-like) interactions are more stable than non-native ones, this search mechanism is able to find the structure with the lowest energy (Baldwin 1994, Dinner et al. 2000). As the native state is approached, the conformational space accessible to the polypeptide chain is reduced (Wolynes et al. 1995). The fundamental mechanism of protein folding involves the formation of a folding-nucleus of residues in the protein, around which the remainder structure rapidly condenses (Otzen and Fersht 1998).

Small, single-domain proteins do not require many partially folded intermediates to reach a native conformation, and usually only extreme conditions unfold them (Jackson 1998). In contrast, folding of large, multidomain proteins involve several intermediates prior to the formation of the completely folded native state. They usu-

ally fold in modules that finally interact to conform the fully native structure (Khan et al. 2003, Panchenko et al. 1996, Vendruscolo et al. 2003) but often require the assistance of folding modulators, namely isomerases and foldases. The requirement of such cell elements dramatically increases in the context of recombinant protein production, in which the host cell receives an extremely dramatically high input of *de novo* synthesized polypeptides. In fact, chaperones are considered limiting factors in recombinant cells.

The term "misfolding" is used to describe the process that results in a protein acquiring a sufficient number of persistent non-native interactions to affect its overall architecture and/or its properties in a biologically significant manner (Dobson 2004). Misfolded and incompletely folded molecules are susceptible to aggregate, due to the exposure of hydrophobic regions that are buried in the native state (Fink 1998) (Fig. 15.1). To avoid aggregation, cells of living organisms have auxiliary factors, including folding catalysts that accelerate rate-limiting steps, and molecular chaperones that assist protein folding (Gething and Sambrook 1992, Hartl and Hayer-Hartl 2002). Moreover, such cell quality control mechanism targets for proteolytic destruction any protein molecule that has not folded correctly (Fig. 15.1). Protein misfolding in recombinant bacteria and other microbial cell factories is a major concern in Biotechnology, as misfolding not only results in protein degradation and/or aggregation but also in a global conformational stress status that triggers a set of cell responses.

15.3 The *Escherichia coli* Quality Control System

The protein quality control machinery is mainly based on the activity of chaperones and proteases that co-ordinately act assisting protein folding, preventing accumulation of misfolded species, removing protein from aggregates and degrading folding-reluctant species (Bukau et al. 2006). Therefore, this system's coordinated activity promotes protein solubility by minimizing the amount of aggregated species. In the biotechnological context, solubility is the parameter commonly used to evaluate the quality of a recombinant protein in a production process (de Marco et al. 2007, Schultz et al. 2006), and it is given as the amount of recombinant protein present in the soluble cell fraction relative to the total amount of recombinant protein occurring in the cell (usually expressed as percentage). In *E. coli*, this system is composed by periplasmic and cytoplasmic arms, which control polypeptides secreted and retained in the cytoplasm respectively. Periplasmic quality control has been extensively reviewed elsewhere (Miot and Betton 2004) and the next sections will mainly focus on the cytoplasmic regulators of protein folding and quality.

15.3.1 Chaperones and Proteases

The term chaperone was first used to describe an activity associated with nucleoplasmin in *Xenopus* oocytes (Laskey et al. 1978). Since then, the term has been expanded to include more than 20 protein families with a central role in the

conformational quality control of the proteome (Bukau et al. 2006, Ellis 1987, Young et al. 2004). Specifically, molecular chaperones are a group of structurally diverse proteins highly conserved in all kingdoms of life which form a complex network to assist proper protein folding, prevent their deposition and dissolve deposits of misfolded proteins (Kazemi-Esfarjani and Benzer 2000, Krobitsch and Lindquist 2000, Mogk et al. 1999, Muchowski et al. 2000, Warrick et al. 1999). Even though chaperones are constitutively expressed under physiological conditions, many of them are upregulated under conformational stress conditions. In *E. coli*, such regulation is mainly controlled by the sigma factor σ^{32}, encoded by *rpoH* gene (Straus et al. 1987). Since chaperone abundance increases in cells upon thermal stress, these molecules have been traditionally named heat shock proteins (Hsps) (Lemaux et al. 1978), although not all chaperones are heat shock proteins and not all heat shock proteins are chaperones. Molecular chaperones can be divided into three functional subclasses based on their mechanism of action:

"Folding" chaperones mediate the folding of their substrates in an ATP-dependent process. These cell molecules increase the yield of properly folded proteins but not the folding rate. In the *E. coli* cytoplasm the three chaperone systems involved in this process are trigger factor (TF) (Bukau et al. 2000), DnaK-DnaJ-GrpE and GroEL-GroES (Grantcharova et al. 2001).

"Holding" chaperones maintain proteins partially folded on their surface to await availability of folding chaperones upon stress conditions, preventing polypeptides from aggregation (Ehrnsperger et al. 1997, Mogk et al. 1999, Veinger et al. 1998). The most extensively characterized bacterial holdases are IbpA and IbpB, both belonging to the small Hsp family (Narberhaus 2002) and commonly found within IBs (Allen et al. 1992) with a suspected role in its physiological disintegration (Lethanh et al. 2005). Hsp31 (Sastry et al. 2002) and Hsp33 (Graf and Jakob 2002) are also classified as holdases; while Hsp31 binds early unfolding intermediates in times of severe stress, thereby preventing overload of the DnaK-DnaJ-GrpE system (Malki et al. 2003, Mujacic et al. 2004), Hsp33 manages oxidative protein misfolding (Graf and Jakob 2002).

Finally, "disaggregating" chaperones promote protein removal from IBs and other aggregates (Rinas et al. 2007). Among them, ClpB is the best characterized. It has a secondary role, assisting refolding and promoting the solubilisation of proteins that have become aggregated as a result of stress (Ben-Zvi and Goloubinoff 2001, Schirmer et al. 1996). This chaperone acts in cooperation with DnaK and IbpAB chaperones (Mogk et al. 2003, Schlieker et al. 2004, Thomas and Baneyx 2000).

Moreover, at least several chaperones, including DnaK, ClpA and ClpX, work in cooperation with proteases (Garcia-Fruitos et al. 2007b, Hoskins et al. 1998, Matouschek 2003).

15.3.1.1 Trigger Factor

The ribosome-associated trigger factor (TF) is a three-domain protein that binds to the large subunit of the ribosomes, in the vicinity of the peptide exit site, to interact with nascent polypeptides and protect them (Bukau et al. 2000). Trigger factor

exhibits both peptidyl-prolyl *cis/trans* isomerase (PPIase) and chaperone activity (Hoffmann and Rinas 2004, Huang et al. 2000, Nishihara et al. 2000). Therefore, this molecular chaperone supports the *de novo* folding by binding to nascent chains. Once the substrate is released, trigger factor can cycle back to the ribosome, waiting for the next substrate molecule (Bukau et al. 2000).

15.3.1.2 The Hsp70 System: DnaK, DnaJ, GrpE

Hsp70 family proteins are encoded in all living organisms' genomes, being one of the most conserved family in the evolution (Gupta and Singh 1994, Hunt and Morimoto 1985, Lindquist and Craig 1988). In *E. coli*, there are three Hsp70 proteins (namely DnaK, HscA and HscC), being DnaK the best characterized. DnaK is a key element of the multichaperone network, having different recognized roles: (1) it mediates ATP-dependent unfolding, (2) prevents aggregation, (3) stabilises the substrates for refolding by GroELS (Goloubinoff et al. 1999, Gupta and Singh 1994, Hoffmann and Rinas 2004, Hunt and Morimoto 1985, Lindquist and Craig 1988, Nishihara et al. 1998, Thomas and Baneyx 1996a), (4) participates in proteolysis (Bukau 1993, Yura and Nakahigashi 1999), cooperating in some cases with Lon protease, (5) folds newly synthesized polypeptides (Hartl and Hayer-Hartl 2002, Teter et al. 1999), (6) solubilises protein aggregates in cooperation with ClpB and Ibps (Ben-Zvi and Goloubinoff 2001, Carrio and Villaverde 2002, Glover and Tkach 2001, Goloubinoff et al. 1999, Mogk and Bukau 2004, Mogk et al. 1999, Zolkiewski 1999), (7) protects proteins against oxidative damages (Echave et al. 2002, Fredriksson et al. 2005) and (8) negatively regulates the heat shock response (Nagai et al. 1994) minimizing the expression of the heat shock σ^{32} regulon, which encodes the main chaperones and proteases, including DnaK itself (Morita et al. 2000, Tomoyasu et al. 2001a, Tomoyasu et al. 1998).

DnaK has an N-terminal ATPase domain of 44 kDa, two β-sheets forming a substrate binding site and a C-terminal domain of 27 kDa that can interact with partner proteins to modulate chaperone function (Genevaux et al. 2007, Genevaux et al. 2001). DnaK partners are a J-domain protein (JDP) co-chaperone, belonging to the Hsp40 family, termed DnaJ (Hennessy et al. 2005) and a nucleotide exchange factor (NEF) named GrpE. When ATP is bound, DnaK binds the substrate through weak, hydrophobic interactions and hydrogen bonds (Zhu et al. 1996). Upon ATP hydrolysis, there is a conformational change that stabilises substrate binding (Hoffmann and Rinas 2004). In this process, the co-chaperone DnaJ has an important role accelerating the ATP hydrolysis rate, while the co-chaperone GrpE accelerates the exchange of ADP with ATP, leading to substrate ejection. The released polypeptide may reach a native conformation, undergo additional cycles in the chaperone system until it folds, or be transferred to GroEL-GroES (Ewalt et al. 1997). The system formed by DnaK chaperone and DnaJ and GrpE co-chaperones is usually abbreviated as KJE.

15.3.1.3 ClpB

ClpB is an ATP-dependent molecular chaperone, member of Hsp100 family. Specifically, ClpB is a "disagregase" that works in cooperation with DnaK-DnaJ-GrpE reverting aggregation (Carrio and Villaverde 2001, Hoffmann and Rinas 2004,

Mogk et al. 2003, Parsell et al. 1994). This molecular chaperone has an important role, in cooperation with DnaK, in dissolving protein aggregates, reducing the aggregate size and exposing hydrophobic surfaces (Ben-Zvi and Goloubinoff 2001, Goloubinoff et al. 1999, Zolkiewski 1999). However, the full recovery of the native state cannot be achieved until the partially unfolded substrate is transferred from ClpB to DnaK (Glover and Lindquist 1998, Goloubinoff et al. 1999, Mogk et al. 1999, Motohashi et al. 1999, Zolkiewski 1999).

15.3.1.4 Hsp60 System: GroEL and GroES

GroEL is a bacterial chaperonin of approximately 60 kDa that belongs to the Hsp60 family. This molecular chaperone, essential for growth at all temperatures (Fayet et al. 1989), prevents aggregation (Kedzierska et al. 1999), acting as the main folder element in the chaperone network (Grantcharova et al. 2001). GroEL is formed by two stacked homoheptameric rings which define a central cavity in which incompletely folded polypeptides up to around 60 kDa (Sakikawa et al. 1999) can properly fold. When ATP is bound, a conformational change takes place (Ranson et al. 2001) rendering GroEL competent to bind the 10 kDa accessory protein GroES (Hartl and Hayer-Hartl 2002). The GroES-bound GroEL protein undergoes a second conformational modification, allowing the folding of the non-native polypeptide. If the protein has not reached the native state, a further round of binding and attempted folding follows.

15.3.1.5 Small Heat Shock Proteins

The best defined small heat shock proteins (sHsps) in bacteria have been Inclusion Body Proteins (Ibps), which are regularly associated to inclusion bodies (Allen et al. 1992) and commonly organized in large oligomeric structures (Haslbeck 2002, Narberhaus 2002). There are two different types of Ibps encoded on a single-operon (Allen et al. 1992, Chuang et al. 1993), IbpA and IbpB of 14 and 16 kDa size, respectively. Although IbpA is insoluble and IbpB is mainly soluble, IbpB comigrates to the insoluble fraction when produced with IbpA (Kuczynska-Wisnik et al. 2002). Even though Ibps function is not well understood, they seem to recognize hydrophobic patches in unfolded proteins, remaining bound to these polypeptides and protecting them from aggregation until they are transferred to DnaK or GroEL for refolding (Kitagawa et al. 2002, Kuczynska-Wisnik et al. 2002, Shearstone and Baneyx 1999, Thomas and Baneyx 1998). Moreover, it has been recently described that IbpA and IbpB facilitate the disaggregation and refolding activity of ClpB (Mogk et al. 2003).

15.3.1.6 Proteases

Proteolysis of misfolded proteins that have failed to reach a native conformation plays a crucial role in the quality control system, preventing the aggregation of abnormal polypeptides as well as allowing the amino acid recycling within the cell. The main proteases of *E. coli* cytoplasm are ClpP and Lon (Gottesman et al. 1997,

Maurizi 1992, Wickner et al. 1999). These heat-shock ATP-dependent proteases recognize hydrophobic surfaces, as chaperones do (Wickner et al. 1999). Moreover, these cell proteases degrade not only unprotected, misfolded polypeptides localized in the soluble cell fraction (Carrio et al. 1999, Maurizi 1992), but also those found embedded in protein aggregates (Corchero et al. 1997, Vera et al. 2005).

Lon is a tetrameric serine protease of 87 kDa subunits containing three functional domains. Its N-terminus is involved in substrate recognition and binding, its central domain is responsible for ATPase activity and its C-terminus domain has proteolytic activity. In addition to being responsible for bulk protein degradation (Missiakas et al. 1996, Tomoyasu et al. 2001b), Lon also exerts a regulatory function by degrading a class of proteins that are designed to be unstable.

ClpP is a protein organized as two stacked heptamers of 23 kDa each. Their substrates are folded, misfolded or incompletely synthesized proteins that are targeted for degradation. This protease forms a complex with two members of the Hsp100 family of ATPases (ClpA and ClpX) (Hoskins et al. 1998, Levchenko et al. 1995, Wickner et al. 1994) to form a fully-competent degrading machinery. ClpA and ClpX, which are flanking the rings of ClpP, act as molecular chaperones, unfolding proteins in an ATP-dependent manner and translocating substrates into the ClpP central channel (Matouschek 2003).

15.4 Composition of Inclusion Bodies

In general, the major component of IBs is the recombinant protein itself that can reach up to around 95% of the deposited protein material (Villaverde and Carrio 2003). However, in addition to the target protein other plasmid or host cell derived proteins or other cell components coprecipitate during IB recovery, adsorb to IBs or can get even entrapped *in vivo* during IB construction. For instance, lipids, DNA and outer membrane proteins are not integral IB components but coprecipitate after mechanical cell breakage with the aggregates during sedimentation by centrifugation (Bowden et al. 1991). The outer membrane proteins, for example, are also found in the particulate fraction after cell breakage prior to induction of recombinant protein synthesis and in cells not producing recombinant proteins (Hart et al. 1990, Rinas and Bailey 1992, Rinas and Bailey 1993, Rinas et al. 1993, Schmidt et al. 1999). These outer membrane proteins can be removed from IB preparations by detergent washing and other procedures that do not unfold proteins but solubilise membrane proteins (Estapé and Rinas 1996, Hart et al. 1990). Other non-integral macromolecular host cell contaminants of crude IB preparations, e.g. nucleic acids, phospholipids, and lipopolysaccharides are also removed by washing procedures using buffers composed of detergents, EDTA as well as cell wall- and DNA-degrading enzymes (Harris et al. 1986, Hartley and Kane 1988, Marston 1986, Marston and Hartley 1990, Marston et al. 1984, Schoemaker et al. 1985, Sugrue et al. 1990). In addition to the outer membrane protein OmpA, which constitutes the major portion of contaminating proteins in crude IB preparations (Hart et al. 1990, Rinas and

Bailey 1992, Rinas et al. 1993), other host cell or plasmid-encoded proteins also coprecipitate after mechanical cell breakage of IB-containing cells but also in corresponding control cells not producing the recombinant protein. Examples include the other outer membrane proteins OmpF and OmpC (Hart et al. 1990, Rinas and Bailey 1992, Rinas et al. 1993), other membrane proteins such as the flavoprotein subunit of succinate dehydrogenase SdhA, and ribosomal subunit proteins L7/L12 (Rinas and Bailey 1992). Moreover, the plasmid-encoded cI857 repressor, a thermolabile protein used for controlling temperature-inducible lambda promoter based expression systems, has been found in the insoluble cell fraction of IB-containing cells but also in respective control cells suggesting that its aggregation is not related to target protein production (Rinas et al. 2007).

However, there are also other proteins specifically associated with the aggregated fraction of inclusion body producing cells which are not found in the aggregated fraction of respective control cells. For instance, truncated versions of the recombinant target protein, other plasmid-encoded proteins e.g. those conferring resistance to antibiotics, and defined host cell proteins have been found entrapped within bacterial IBs (Hart et al. 1990, Jurgen et al. 2000, Neubauer et al. 2007, Rinas and Bailey 1992, Rinas and Bailey 1993, Rinas et al. 1993, Wagner et al. 2007). In particular, putative DnaK substrates such as the elongation factor Tu (ET-Tu) and the metabolic enzymes dihydrolipoamide dehydrogenase (LpdA), tryptophanase (TnaA), and D-tagatose-1,6-bisphosphate aldolase (GatY) have been identified only in the aggregated fraction of inclusion body producing cells (Rinas et al. 2007). GatY, in particular, a notoriously insoluble protein depending on GroEL (Chapman et al. 2006, Kerner et al. 2005) and DnaK for proper folding (Mogk et al. 1999), has also been found in other inclusion body preparations (Josef Lengeler and Peter Neubauer, personal communication). In some cases entrappment of precursors of membrane and periplasmic proteins into cytoplasmic IBs has been reported (Rinas and Bailey 1993, Wagner et al. 2007)

The most prominent host cell derived protein contaminants of IBs were identified as members of the heat-shock protein family (Allen et al. 1992). As their function was completely unknown at that time these IB contaminants were named inclusion body proteins (IbpA and IbpB). Since then, their presence within bacterial IBs has been further reported (Lethanh et al. 2005, Wagner et al. 2007) but also their absence in IB preparations has been noted (Rinas et al. 2007). Today, their function is still not completely understood. *In vitro* studies on thermal aggregates indicate that both together efficiently stabilize thermally aggregated proteins in a disaggregation competent state and allow more effective reactivation through the disaggregating chaperones ClpB and DnaK (Kuczynska-Wisnik et al. 2002, Laskowska et al. 2004, Lewandowska et al. 2007, Matuszewska et al. 2005, Veinger et al. 1998). Moreover, *in vivo* studies revealed that the presence of IbpA and IbpB renders the aggregated polypeptides in a conformationally more native state, with higher enzymatic activity compared to IBs produced in *ibpAB* deletion strains (Kuczynska-Wisnik et al. 2004). Other members of the heat shock protein family, namely the chaperones DnaK and GroEL, have also been found associated with IBs (Carrio and Villaverde 2002). DnaK is localized preferentially on the surface of inclusion bodies (Carrio

and Villaverde 2005) and, together with ClpB, is recovered with low density protein aggregates during sucrose density centrifugation (Schrodel and de Marco 2005). The presence of DnaK has also been verified in other inclusion body preparations (Rinas et al. 2007, Wagner et al. 2007). GroEL, on the other hand, is homogeneously distributed in the cytosol, absent from the IB surface, but found in minor amounts also inside the aggregates (Carrio and Villaverde 2005). During *in vitro* recovery by sucrose density centrifugation, GroEL is recovered together with IbpB with high density protein aggregates (Schrodel and de Marco 2005). However, other researchers report absence or at most very small quantities of GroEL in IB preparations (Bowden et al. 1991, Carrio and Villaverde 2002, Rinas and Bailey 1992, Rinas et al. 1993, Rinas et al. 2007, Wagner et al. 2007).

It has been long debated if intracellular protein aggregation as IBs is a specific process between identical protein chains or is a process where different protein chains interact with each other forming mixed aggregates. *In vitro* mixed refolding studies using the P22 tailspike and coat proteins revealed that the two proteins did not coaggregate with each other but only with themselves, suggesting that aggregation is caused by specific interactions among protein chains (Speed et al. 1996). Moreover, *in vitro* seeding of pure soluble protein solutions with purified IBs revealed that aggregation of the soluble protein was only induced when seeding occurred with IBs composed of the same protein but not with IBs composed of unrelated proteins (Carrio et al. 2005). Also, recent experiments using Fluorescence Resonance Energy Transfer (FRET) indicated that coproduction of two different aggregation prone proteins in *E. coli* does not lead to mixed intermolecular interactions between the different protein chains (Morell et al. 2008). In this line, *in vivo* studies using a human cell line (HEK293) revealed that coexpression of two unrelated aggregation prone proteins did not lead to coaggregation but to deposition in separate aggregates in the same cell, suggesting strong specificity of protein aggregation (Rajan et al. 2001). On the other hand, kanamycin phosphotransferase, a plasmid encoded protein conferring resistance to kanamycin, can only be solubilized under conditions that also solubilize the plasmid-encoded target protein, bovine growth hormone, strongly suggesting that both proteins are tightly associated within IBs (Schoner et al. 1985). However, tight association of proteins within bacterial IBs does not necessarily imply interactions between unrelated peptide chains but could simply indicate colocalization of small protein aggregates within inclusion bodies. This is not unexpected having in mind that the cellular environment is a very crowded space (Ellis 2001), with protein concentrations in the cytoplasm in the order of 200 g/L (Neidhardt and Umbarger 1996) and concentrations of all macromolecules together reaching more than 340 g/L (Zimmerman and Minton 1993, Zimmerman and Trach 1991). Moreover, protein diffusion experiments suggest that in solutions containing proteins at concentrations comparable to those found in biological fluid media, the diffusive transport of larger proteins and aggregates may be slower than in dilute solution by several orders of magnitude (Muramatsu and Minton 1988). By using very strong expression systems, induction leads to almost exclusive synthesis of the target protein (Schoner et al. 1985). For example, temperature-induction of recombinant protein synthesis can increase total

protein synthesis rates four fold (with 60% of protein synthesis dedicated to the synthesis of the target protein) but leading only to 10% target protein accumulation (Hoffmann and Rinas 2000, Hoffmann and Rinas 2001). Thus, during a limited period of time very high protein synthesis rates can occur in protein producer cells which can explain entrappment of normally soluble host cell or plasmid derived proteins into bacterial inclusion bodies. In this line, high level expression of an aggregation prone target protein can also lead to entrappment of another aggregation prone plasmid-encoded protein into inclusion bodies although this other plasmid-encoded protein is produced at lower rates during target protein synthesis compared to respective control conditions without target protein overexpression (Neubauer et al. 2007, Rinas and Bailey 1993). Thus, non-target but aggregation prone proteins might directly aggregate where they are synthesized due to diffusive transport limitations during high level target proteins synthesis thereby leading to inclusion bodies of mixed micro aggregates.

15.5 Structural Properties of Bacterial Inclusion Bodies

IBs are protein aggregates with spherical or ovoid shapes, formed either in the cytoplasm or the periplasm, and that are observed as refractile particles (usually one or two per cell) by optical microscopy (Bowden et al. 1991, Carrio et al. 2005) and as electrodense masses by transmission electron microscopy (Bowden et al. 1991). Soluble polypeptides can be extracted *in vitro* from IBs by denaturation and refolding sequential procedures (Rudolph and Lilie 1996, Vallejo and Rinas 2004), that permit to obtain soluble protein species through protein-tailored protocols. Interestingly, the arrest of protein synthesis in recombinant bacteria promotes the fast disintegration of IBs (Carrio et al. 1999) proving that they result from an unbalanced equilibrium between protein deposition and cell mediated protein removal, in which both chaperones and proteases are involved. This fact is also being considered when designing *in vivo* protein recovery protocols (de Marco et al. 2007).

The inner molecular organization of bacterial IBs has been a matter of deep scientific discussion. To date, some spectroscopic techniques have been developed or fitted to the analysis of bacterial IBs, namely Circular Dichroism (Chiti et al. 1998, Lewandowska et al. 2007, Plakoutsi et al. 2005, Umetsu et al. 2004, Umetsu et al. 2005), Raman Spectroscopy (Przybycien et al. 1994), Dynamic Light Scattering (Grudzielanek et al. 2007, Plakoutsi et al. 2005), and Nuclear Magnetic Resonance (Umetsu et al. 2004). However, Fourier-Transform Infrared Spectroscopy (FTIR) has proven to be the most useful and powerful tool for this purpose (Ami et al. 2005, Jevsevar et al. 2005), especially Attenuated Total Reflection-FTIR (ATR-FTIR) (Gonzalez-Montalban et al. 2006, Gonzalez-Montalban et al. 2007b, Vera et al. 2007). FTIR, in contrast to other optical spectroscopic methods, resolves measurements which are essentially unaffected by light scattering on residual protein-lipid interactions or contaminant membrane fragments. For this reason, this technique was originally developed for the study of structural characterization of membrane

or lipid-associated proteins (Chapman and Haris 1989, Surewicz and Mantsch 1988, Surewicz et al. 1988).

Infrared spectroscopy is a form of vibrational spectroscopy which reports directly on the secondary structure of the proteins. For this purpose, the major areas of interest in the spectra are Amide I and Amide II bands. Amide I band arises predominantly (about 80%) from the C = O stretching vibration of the amide functional group which absorbs basically in the 1600–1700cm^{-1} region. The Amide II band arises from N-H bending and C-N stretching vibrations which absorb in the 1500–1600cm^{-1} region. However, structural studies on protein aggregation are usually based on evaluations of Amide-I-band contour, since only this band is a sensitive marker of secondary structure, being the analysis of Amide II band, in general, less relevant.

Aggregation as IBs has been long thought to be an unspecific process driven by the random interactions of hydrophobic patches, thus rendering protein aggregates with no specific internal molecular architecture. However, more recently, evidences against this view have been rapidly increasing (Mozell et al. 2008, Wang et al. 2008, Ami et al. 2005, Ami et al. 2006, Carrio et al. 2005, Oberg et al. 1994, Przybycien et al. 1994, Umetsu et al. 2004), picturing IBs as highly ordered structures. As FTIR analysis reveals, IBs seem to build up through a constant type of intermolecular protein interactions, resulting in a molecular architecture characterized by the formation of new β-sheet structures (Ami et al. 2003, Carrio et al. 2005, Garcia-Fruitos et al. 2007b, Gonzalez-Montalban et al. 2007b, Umetsu et al. 2005) at expenses of α-helical structures (Fink 1998, Przybycien et al. 1994), even common to rich-β-sheet native proteins (Oberg et al. 1994, Vera et al. 2007). In cases where the aggregation-prone protein is an all-α-protein, as it happens with interleukin-4(IL-4) (Umetsu et al. 2004), IBs are characterized by a sharp increment in β-sheet content and by an almost undetectable α-helical moieties signal. These structural data suggest that the new formed β-sheet structures may be interacting in a different way from native β-sheet conformation, probably by a network of hydrogen bonds between different chains creating a tightly packed extended, intermolecular β-sheet conformation (Fink 1998). Altogether, these observations seem to point out that the interactions leading to IB formation and the molecular reorganization that aggregated proteins undergo within the deposit are not likely to be unspecific interactions.

Table 15.1 Common structural features of amyloid fibrils and inclusion bodies

Structural characteristics	References
Structural homogeneity	(Ami et al. 2005, Carrio et al. 2005, Fink 1998, Garcia-Fruitos et al. 2007b, Vera et al. 2007)
Intermolecular, cross β-sheet organization or enrichement of β-structures	(Carrio et al. 2005, Garcia-Fruitos et al. 2007b, Gonzalez-Montalban et al. 2006, Przybycien et al. 1994)
Amyloid-tropic dye binding	(Carrio et al. 2005)
Cytotoxicity linked to amyloid-like structures	(Gonzalez-Montalban et al. 2007b)

Interestingly, all these secondary structural features greatly resemble to those that have been proven to characterize amyloid fibril formation (Table 15.1). In the case of amyloid fibrils, sequence determinants acting as aggregating "hot-spots" seem to modulate the specific nucleation of amyloid proteins (Ivanova et al. 2004, Ventura 2005, Ventura and Villaverde 2006, Ventura et al. 2004). In fact, in recent years it has been shown that IBs formation is a highly specific process since this kind of protein aggregates are essentially formed by the recombinant protein (Carrio et al. 1998, Garcia-Fruitos et al. 2005a, Gonzalez-Montalban et al. 2006) and organized in a very homogeneous architecture (Ami et al. 2005, Fink 1998). Furthermore, preformed IBs can seed specifically misfolded counterparts promoting the deposition of homologous but not heterologous domains (Carrio et al. 2005). As in the case of amyloid fibrils, whose formation seems to be preceded by the formation of intermediate amyloid-like species linked to cellular toxicity (Bucciantini et al. 2002, 2004, Stefani and Dobson 2003), IB structure is reported to be deleterious for mammalian cells in a structural-dependent manner (Gonzalez-Montalban et al. 2007b).

Intriguingly, the increase in the newly formed, non-native β-sheet content does not necessarily involve the full unfolding of the protein sequestered in IBs. In fact, there is a significant number of reports indicating the occurrence of native-like secondary structure of IB polypeptides (Table 15.2).

The native structure of the soluble IL-2 and its IB counterpart is almost identical, with only packing the degree and the nature of molecular β-sheet interaction being the main differences (Oberg et al. 1994). In fact, the little variations in FTIR signals seem to reflect subtle rather than significant changes in the secondary structure. TEM β-lactamase IBs seem to retain about 60% of the native secondary structure of the soluble protein (Georgiou et al. 1994). A particular example is represented by recombinant hyperthermophilic archaeon proteins. At least 3 proteins (namely PH0979, PH0628 and PH1830), when embedded in IBs, maintained some degree of a native-like rigid secondary structure (Umetsu et al. 2004). These observations were also made for IBs formed by *Pseudomonas fragi* lipase, human growth hormone and interferon-alpha-2b (Ami et al. 2005, Ami et al. 2006). VP1LAC, a re-

Table 15.2 Coincidence of native-like structure and amyloid-like aggregation pattern in inclusion bodies

IB-forming protein	Amyloid-like structure	Native-like structure	References
K97V 1L-1β	Yes	Yes	(Oberg et al. 1994)
hDHFR	Yes	Yes	(Garcia-Fruitos et al. 2005b)
VP1GFP	Yes	Yes	(Garcia-Fruitos et al. 2005b)
Aβ42(F19D)-BFP	Yes	Yes	(Garcia-Fruitos et al. 2005b)
hG-CSF	Yes	Yes	(Jevsevar et al. 2005)
LPF	Yes	Yes	(Ami et al. 2005)
h-GH	Yes	Yes	(Ami et al. 2006)
IFN-alpha-2b	Yes	Yes	(Ami et al. 2006)
S65T GFP	Yes	Yes	(Vera et al. 2007)
VP1LAC	Yes	Yes	(Gonzalez-Montalban et al. 2006)

combinant *E. coli* β-galactosidase, retained a great amount of native-like structure when forming IBs in a mutant strain lacking a fully functional chaperone GroEL (Gonzalez-Montalban et al. 2006). This trait is also common to recombinant fluorescent proteins. The structural analysis by ATR-FTIR and fluorescence measurements showed that green fluorescent protein (GFP) (Vera et al. 2007), VP1GFP (a GFP fused to a foot-and-mouth disease virus capsid protein) and Aβ42(F19D)-BFP (an amyloid peptide fused to blue fluorescent protein (BFP)) (Garcia-Fruitos et al. 2005b) aggregated as IBs maintain native-like structure. This feature seems to permit an easier solubilization of the embedded protein. In this regard, L-arginine can easily disaggregate GFP (Tsumoto et al. 2003) and β2 microglobulin (Umetsu et al. 2005) from IBs due to the retention of native-like structure of the embedded polypeptides. Human granulocyte-colony stimulating factor (hG-CSF) produced in *E. coli* at low temperatures enables the formation of "non-classical" IBs, which contain high amounts of correctly folded hG-CSF. HG-CSF can be readily extracted from these "non-classical" IBs by nondenaturing conditions and low concentrations of polar solvents (Jevsevar et al. 2005).

15.6 Strategies to Minimize Inclusion Body Formation

In general, the refolding processes required to recover the protein in a native form are complex, expensive and not always convenient from an industrial point of view (Vallejo and Rinas 2004). For this reason, much effort has been invested to minimize IB formation during the production process itself, aiming to improve the yield of soluble protein species. Recombinant protein can account up to around 30% of the total cell protein, producing an enormous metabolic load on the *E. coli* biosynthetic machinery (Sahdev et al. 2008). Thus, as summarized below, some of the strategies devised to minimize aggregation are based on a tight control of the *E. coli* cellular milieu, while others are addressed to favour protein folding by either physicochemical or biological approaches.

15.6.1 Media Composition

The folding of certain proteins requires the presence of specific cofactors in the growth media, such as metal ions (e.g., iron-sulphur) or polypeptide-cofactors (e.g., flavin-mononucleotide). By adding these factors to the growth media, both protein solubility and folding rates can be enhanced (Apiyo and Wittung-Stafshede 2002, Bruser et al. 2003). The composition of growth media also affects the levels of soluble protein. By optimizing media composition, reduced expression times, increased soluble fraction yield and enhanced biological activity of enzymes have been achieved. These modifications have been recently reviewed (Sahdev et al. 2008).

15.6.2 Protein Production at Low Temperatures

A number of proteins have been successfully produced in a soluble form in *E. coli* by lowering the growth temperature of the culture (Chesshyre and Hipkiss 1989, Niiranen et al. 2007, Schein and Noteborn 1988, Vera et al. 2007). As the hydrophobic interactions that determine IB formation are temperature-dependent, protein production at temperatures below the optimal of 37 °C for *E. coli* growth usually leads to increased stability and correct folding (Sahdev et al. 2008). Moreover, the increased production of a number of chaperones also accounts for the better protein quality obtained at lower growth rates (Ferrer et al. 2003). In addition, some of the heat shock proteases induced during recombinant protein production are poorly active at low temperatures. This accounts for the reduced degradation of recombinant protein observed within a temperature range of 15–23 °C (Hunke and Betton 2003, Spiess et al. 1999).

However, disadvantages are also present in the use of this strategy, as low temperatures lead to reduced transcription and translation rates, which results in low yields and poor turnover of the recombinant protein.

15.6.3 Genetic Modification of Producing Escherichia coli Strains

Genetic background largely affects recombinant protein production. Ideally, host strains should be deficient in the most harmful proteases, confer a stable maintenance of the expression plasmid and be compatible with the expression system chosen by providing the genetic elements required (e.g., DE3 strain for the pET expression system) (Sorensen and Mortensen 2005a).

E. coli BL21 (Novagen, USA) is one among the most common hosts. The nonpathogenic *E. coli* B strains can grow in minimal media and are deficient in *ompT* and Lon proteases, providing increased protein stability. The most important BL21 derivatives include:

- *BLR* RecA$^-$ for stabilization of target plasmids containing repetitive sequences.
- *trxB/gor mutants* for enhancement of cytoplasmic disulfide bond formation (Novagen Origami and AD494 strains).
- *Rosetta-gami strains* for overcoming codon bias associated problems through the overexpression of a rare tRNA expression vector, in addition to the *trxB*/gor mutation described above.
- *lac ZY* deletion mutants for uniform and adjustable protein expression in all the cells (Novagen Tuner series).
- Origami-B strains derived from a *lac ZY* mutant of BL21, also including *trxB* and *gor* mutations and OmpT and Lon deficiencies of *BL21*.
- Avidis C1(DE3) and C43(DE3) strains, for soluble expression of IB prone and membrane proteins.

15.6.4 Co-production of Folding Modulators

Molecular chaperones important for the control of protein quality are believed to be limiting in bacterial cell factories. Therefore, co-production strategies have been widely tested to overcome limitations due to IB formation during recombinant protein expression, but to date the results obtained are in general controversial and inconsistent (Baldwin 1986, Baneyx 2004, Thomas et al. 1997). Some successful examples of improved solubility by coproduction of some of the major cytosolic chaperones (namely the DnaK-DnaJ-GrpE system or the GroEL-ES complex) are human ORP150, human lysozyme, p50csk protein tyrosine kinase, phosphomannose isomerase, endostatin, transglutaminase and fusion protein PreS2-S′-β-galactosidase (Amrein et al. 1995, Dale et al. 1994, de Marco et al. 2000, Nishihara et al. 2000, Proudfoot et al. 1996, Thomas and Baneyx 1996a, Thomas and Baneyx 1996b, Yokoyama et al. 1998). However, although a trial and error approach is still needed to determine the best set of chaperones for a determined target protein, so far the best results have been obtained by coexpression of several sets of folding modulators.

Recently, a systematic analysis of the combined power of the major cytosolic chaperone systems of *E. coli* (KJE, ELS, ClpB and IbpAB) was performed (de Marco et al. 2007). Of the 50 proteins tested, the solubility of around 50% of them was improved by chaperone co-overproduction, being KJE, ClpB and ELS the most successful combination. The study also suggested an enhancement of the native state acquisition due to chaperone overproduction.

Optimization of the procedure was done by allowing chaperone-assisted folding in absence of protein synthesis, which was blocked by either inducer withdrawal or chloramphenicol addition. Solubility yields increased in comparison to the one-step procedure, with some of the proteins requiring the two-step procedure for any solubilisation. Coproduction of IbpAB also improved solubility, even being the only combination that solubilised some of the proteins tested.

15.6.5 Fusion Tags

A different strategy consists of fusion protein technology, in which a solubility "tag" is fused to the target protein (Sahdev et al. 2008). Tags are proteins or peptides that upon fusion, help to the proper folding of their fusion partners and lead to enhanced solubility of the protein (Esposito and Chatterjee 2006). Some tags can also be used for affinity purification, and provide advantages such as protection from proteolysis or being expression reporters (GFP). When solubility tags do not double as affinity tags, they may be combined with another hexahistidine (His6) tag, this way allowing for purification. The use of small peptide tags called SET tags has also been successful for some proteins (Zhang et al. 2004). The small size of these tags (< 30 amino acids) may lead to less folding interferences, making the protein suitable for structural studies without the need of removing the tag.

Table 15.3 Commonly used tags for solubility enhancement

Tag	Protein	Solubility enhancement	Affinity purification
MBP	Maltose-binding protein	Yes	Yes
GST	Glutathione-S-transferase	Yes	Yes
Trx	Thioredoxin	Yes	No
NusA	N-Utilization substance	Yes	No
SUMO	Small ubiquitin-modifier	Yes	No
SET	Solubility-enhancing tag (synthetic)	Yes	No
DsbC	Disulfide bond C	Yes	No
Skp	Seventeen kilodalton protein	Yes	No
T7PK	Phage T7 protein kinase	Yes	No
GB1	Protein G B1 domain	Yes	No
ZZ	Protein A IgG ZZ repeat domain	Yes	No
His6	Hexahistidine tag	No	Yes
FLAG	FLAG tag peptide	No	Yes
BAP	Biotin acceptor peptide	No	Yes
Strep-II	Streptavidin-binding peptide	No	Yes
CBP	Calmodulin-binding peptide	No	Yes

Table adapted from reference (Esposito and Chatterjee 2006)

This technique poses some technical disadvantages, such as the need for tag removal and the question of whether the protein of interest remains in its native state and active once the tag has been removed. Nevertheless, if the target protein is linked to its fusion partner through a protease-specific recognition sequence, this will allow for an easy separation of the purified recombinant protein by cleavage with the specific protease. Because of its high specificity and ease of production, one of the most commonly used proteases is TEV, from tobacco etch virus (Kapust et al. 2002, Kapust et al. 2001).

Some commonly used tags, either for solubility enhancement or combined affinity purification, are listed in Table 15.3.

15.7 Conformational Quality and Biological Activity of Recombinant Proteins in Inclusion Bodies

Although it has been historically believed that proteins deposited as IBs were devoid of any biological activity, independent studies of unrelated aggregating enzymes and fluorescent proteins have demonstrated that IBs are enzymatically active or fluorescent respectively (de Groot and Ventura 2006, Garcia-Fruitos et al. 2005a, Garcia-Fruitos et al. 2005b, Kuczynska-Wisnik et al. 2004, Tokatlidis et al. 1991, Vera et al. 2007, Worrall and Goss 1989). The functional protein species do not occur in the IB interface but in the core of the aggregates, indicating that active polypeptides are not mere contaminants from the soluble cell fraction but true structural components (Garcia-Fruitos et al. 2007a). This is in agreement with the observation of native-like secondary structure in IBs as discussed above, and indicates that solubility and biological activity are not linked parameters. Therefore, aggregation of recombi-

nant proteins as IBs does not split the population of recombinant polypeptides into functional and non functional (Gonzalez-Montalban et al. 2007a), and aggregation determinants must then be defined stretches instead of large protein segments, and not necessarily linked to active sites or fluorophors. Probably, aggregation patches coexist in a single polypeptide molecule, with properly folded regions and conformational quality of protein embedded in IBs depending on how fast the aggregation occurs after protein synthesis (de Groot and Ventura 2006, Waldo et al. 1999). On the other hand, the occurrence in recombinant cells of "soluble aggregates", namely protein deposits present in the soluble fraction (Schrodel and de Marco 2005, Sorensen and Mortensen 2005b, Ventura and Villaverde 2006) is another indicator that solubility is not matching conformational quality, and strongly suggests that there is a wide spectrum of protein conformations in both soluble and insoluble cell fractions (Ventura and Villaverde 2006). The fact that the biological activity of both soluble and insoluble recombinant protein versions is favoured or impaired in

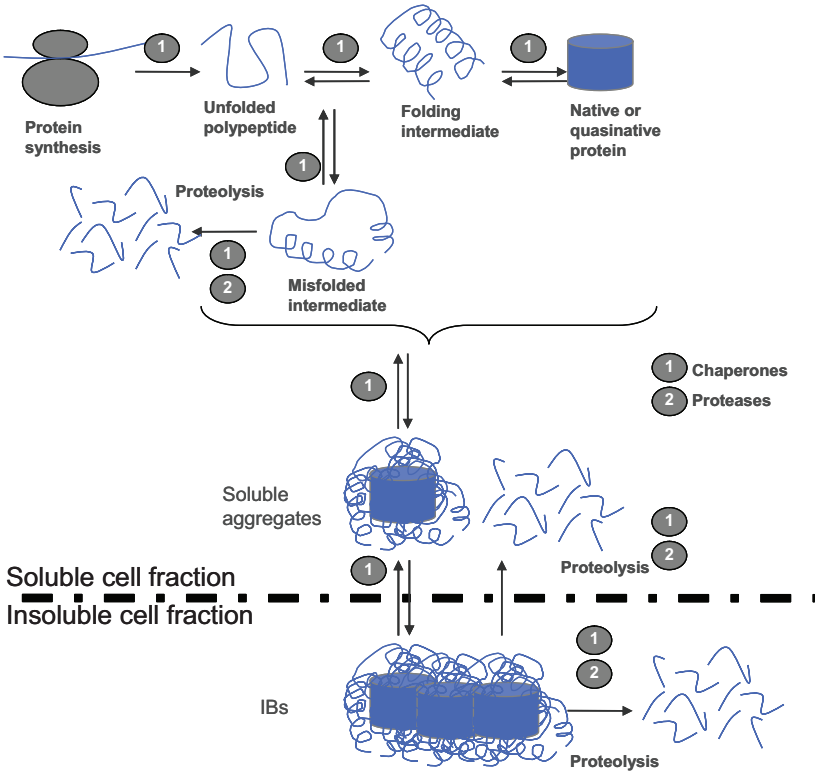

Fig. 15.2 Novel model of protein folding, aggregation and proteolysis in the *E. coli* cytoplasm. Several conformational versions of newly synthesized polypeptides, including those reaching native or native-like forms, can interact to form soluble aggregates, the putative precursors of inclusion bodies. Both soluble aggregates and inclusion bodies are then expected to be heterogeneous regarding protein folding status. The formation of insoluble inclusion bodies is highly favoured at high concentrations of recombinant protein. Chaperones (1) regulate aggregation and disaggregation but also protease (2)-mediated digestion of both soluble and insoluble protein versions

Table 15.4 Inclusion bodies used as biocatalysers

Inclusion bodies as biocatalysers	References
β-Galactosidase	Garcia-Fruitos et al. 2007a
Polyphosphate kinase	Nahalka et al. 2006
D-amino acid oxidase fusion protein	Nahalka and Nidetzky 2007
Maltodextrin phosphorylase fusion protein	Nahalka 2008
Sialic acid aldolase fusion protein	Nahalka et al. 2008

parallel by experimental conditions such as growth temperature (Vera et al. 2007) or availability of chaperones (Martinez-Alonso et al. 2007), indicates that IBs are not excluded from quality control but fully integrated in the cell processing of aberrant proteins (Fig. 15.2).

Regarding practical issues, functional IBs (such as those formed by enzymes) have been proposed as useful catalysts in bioprocesses without the need of protein removal and *in vitro* refolding (Garcia-Fruitos et al. 2007a, Garcia-Fruitos et al. 2005b). This principle has recently been proven with a diversity of aggregating recombinant enzymes such as D-amino oxidase from *Trigonopis variabilis* (Nahalka and Nidetzky 2007), polyphosphate kinase (Nahalka et al. 2006), maltodextrin phosphorylase from *Pyrococcus furious* (Nahalka 2008) and sialic acid aldolase (Nahalka et al. 2008), and opens new and challenging possibilities in the biotechnological market of recombinant proteins (Table 15.4).

15.8 Complex Systems Control of Protein Quality, Aggregation and IB Formation

Interestingly, the dramatic impact that different mutations in chaperone and protease genes have on IB disintegration (Carrio and Villaverde 2003, Vera et al. 2005) indicates that many components of the cell quality coordinately regulate the biology of these aggregates. In this context, a recent study shows that the total or partial intactivation of different genes of the *E. coli* quality control apparatus (including *dnaK, groEL, groES, clpA, clpP* and *lon*) results, as expected, in less solubility, but, surprisingly, in much more functional proteins in both soluble and insoluble populations (Garcia-Fruitos et al. 2007b). In particular, a deficiency in the chaperone DnaK, which is essentially found on the IBs surface in wild type recombinant cells (Carrio and Villaverde 2005), promotes the accumulation of high amounts of highly fluorescent GFP in IBs. This and other intriguing recent findings, such as the DnaK-inhibited activation and folding of β-galactosidase within IBs (Gonzalez-Montalban et al. 2008), the negative effect of DnaK on GFP folding and fluorophore activation (Garcia-Fruitos et al. 2007b, Martinez-Alonso et al. 2007), the DnaK-mediated stimulation of Lon- and Clp-mediated recombinant protein degradation (Garcia-Fruitos et al. 2007b) and the impact of DnaK on the partitioning of recombinant proteins into soluble and insoluble cell fractions (Garcia-Fruitos et al. 2005a, Gonzalez-Montalban et al. 2006) show that the quality control in general and

the particular role of DnaK as a chaperone might have been largely misunderstood, specially regarding IB-forming recombinant cells.

By using GFP as a reporter recombinant protein, it has been determined that, in wild type cells, proteolysis acts on aggregation-prone but functional (or suitable to be activated) polypeptides. However, in cells deficient in chaperones such as ClpB, DnaK or GroEL or proteases such as Lon and ClpP, protein stability significantly increases. It seems that Lon and ClpP, in cooperation with DnaK, ClpB and others, proteolyse polypeptides on the IB surface (Garcia-Fruitos et al. 2007b), probably associated to their release during the continuous *in vivo* IB reconstruction (Carbonell and Villaverde 2002, Carrio et al. 1999, Carrio and Villaverde 2001, Carrio and Villaverde 2002, Corchero et al. 1997, Cubarsi et al. 2005). On the contrary, IbpA and IbpB play an antagosistic role, protecting recombinant proteins from proteolysis (Garcia-Fruitos et al. 2007b, Han et al. 2004). Interestingly, the combination of all these events and in particular, the unexpected role of DnaK in promoting proteolytic digestion of functional protein species and impairing *in situ* IB protein folding, results in a negative correlation between solubility and biological activity (and therefore, conformational quality) of recombinant proteins (Martínez-Alonso et al. 2008, and (Garcia-Fruitos et al. 2007b)). These observations point out that solubility is not a parameter representative of protein quality, since in recombinant cells conformational quality and solubility show a divergent genetic control. At least under recombinant protein production conditions, the bacterial quality control system tends to promote solubility at expenses of conformational quality, what could partially explain the inconsistent results found under coexpression of particular chaperones as discussed above. Also, the soluble and insoluble fractions, as virtual cell compartments, do not have much biological sense regarding protein quality and activity (Fig. 15.2).

To sum up, IBs, rather than being mere molecular "dust-balls" of the protein folding pipeline, are transient but highly dynamic protein reservoirs, fully integrated in the protein quality system, and whose formation and maintenance implies the complex activities of multigenetic networks. From a functional side, IB formation involves a tight cell control of protein folding and proteolytic stability.

Acknowledgments The authors appreciate the finantial support received for their research in recombinant protein production and aggregation, through grants BIO2005-23732-E and BIO2007-61194 MEC, Spain, and 2005SGR-00956 AGAUR, Spain.

Abbreviations

IBs:	inclusion bodies
TF:	trigger factor
KJE:	DnaK-DnaJ-GrpE system
Hsp:	heat shock protein
sHsps:	small heat shock proteins
Ibps:	inclusion bodies proteins
FTIR:	Fourier-transform infrared spectroscopy
ATR-FTIR:	Attenuated Total Reflection-FTIR
GFP:	Green fluorescent protein
BFP:	Blue fluorescent protein
hG-CSF:	Human granulocyte-colony stimulating factor
PPIase:	peptidyl-prolyl *cis/trans* isomerase
NEF:	nucleotide exchange factor

References

Allen SP, Polazzi JO, Gierse JK et al. (1992) Two novel heat shock genes encoding proteins produced in response to heterologous protein expression in *Escherichia coli*. J Bacteriol 174(21):6938–47

Ami D, Bonecchi L, Cali S et al. (2003) FT-IR study of heterologous protein expression in recombinant *Escherichia coli* strains. Biochim Biophys Acta 1624(1–3):6–10

Ami D, Natalello A, Gatti-Lafranconi P et al. (2005) Kinetics of inclusion body formation studied in intact cells by FT-IR spectroscopy. FEBS Lett 579(16):3433–6

Ami D, Natalello A, Taylor G et al. (2006) Structural analysis of protein inclusion bodies by Fourier transform infrared microspectroscopy. Biochim Biophys Acta 1764(4):793–9

Amrein KE, Takacs B, Stieger M et al. (1995) Purification and characterization of recombinant human p50csk protein-tyrosine kinase from an *Escherichia coli* expression system overproducing the bacterial chaperones GroES and GroEL. Proc Natl Acad Sci USA 92(4):1048–52

Anfinsen CB (1973) Principles that govern the folding of protein chains. Science 181(96):223–30

Apiyo D, Wittung-Stafshede P (2002) Presence of the cofactor speeds up folding of *Desulfovibrio desulfuricans* flavodoxin. Protein Sci 11(5):1129–35

Baldwin RL (1986) Temperature dependence of the hydrophobic interaction in protein folding. Proc Natl Acad Sci USA 83(21):8069–72

Baldwin RL (1994) Protein folding. Matching speed and stability. Nature 369(6477):183–4

Baneyx F (2004) Keeping up with protein folding. Microb Cell Fact 3(1):6

Baneyx F, Georgiou G (1991) Construction and characterization of *Escherichia coli* strains deficient in multiple secreted proteases: protease III degrades high-molecular-weight substrates *in vivo*. J Bacteriol 173(8):2696–703

Ben-Zvi AP, Goloubinoff P (2001) Review: mechanisms of disaggregation and refolding of stable protein aggregates by molecular chaperones. J Struct Biol 135(2):84–93

Bowden GA, Paredes AM, Georgiou G (1991) Structure and morphology of protein inclusion bodies in *Escherichia coli*. Biotechnology (N Y) 9(8):725–30

Bruser T, Yano T, Brune DC et al. (2003) Membrane targeting of a folded and cofactor-containing protein. Eur J Biochem 270(6):1211–21

Bucciantini M, Calloni G, Chiti F et al. (2004) Prefibrillar amyloid protein aggregates share common features of cytotoxicity. J Biol Chem 279(30):31374–82

Bucciantini M, Giannoni E, Chiti F et al. (2002) Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. Nature 416(6880):507–11

Bukau B (1993) Regulation of the *Escherichia coli* heat-shock response. Mol Microbiol 9(4): 671–80

Bukau B, Deuerling E, Pfund C et al. (2000) Getting newly synthesized proteins into shape. Cell 101(2):119–22

Bukau B, Weissman J, Horwich A (2006) Molecular chaperones and protein quality control. Cell 125(3):443–51

Carbonell X, Villaverde A (2002) Protein aggregated into bacterial inclusion bodies does not result in protection from proteolytic digestion. Biotechnol Lett 24(23):1939–1944

Carrio M, Gonzalez-Montalban N, Vera A et al. (2005) Amyloid-like properties of bacterial inclusion bodies. J Mol Biol 347(5):1025–37

Carrio MM, Corchero JL, Villaverde A (1998) Dynamics of *in vivo* protein aggregation: building inclusion bodies in recombinant bacteria. FEMS Microbiol Lett 169(1):9–15

Carrio MM, Corchero JL, Villaverde A (1999) Proteolytic digestion of bacterial inclusion body proteins during dynamic transition between soluble and insoluble forms. Biochim Biophys Acta 1434(1):170–6

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

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

Carrio MM, Villaverde A (2003) Role of molecular chaperones in inclusion body formation. FEBS Lett 537(1–3):215–21

Carrio MM, Villaverde A (2005) Localization of chaperones DnaK and GroEL in bacterial inclusion bodies. J Bacteriol 187(10):3599–601

Chapman D, Haris PI (1989) Biomembrane structures. Fourier transform infrared spectroscopy and biomembrane technology. Biochem Soc Trans 17(6):951–3

Chapman E, Farr GW, Usaite R et al. (2006) Global aggregation of newly translated proteins in an *Escherichia coli* strain deficient of the chaperonin GroEL. Proc Natl Acad Sci USA 103(43):15800–5

Chesshyre JA, Hipkiss AR (1989) Low temperatures stabilize interferon α-2 against proteolysis in *Methylophilus methylotrophus* and *Escherichia coli*. Appl Microbiol Biotechnol 31(2): 158–162

Chiti F, Taddei N, van Nuland NA et al. (1998) Structural characterization of the transition state for folding of muscle acylphosphatase. J Mol Biol 283(4):893–903

Chuang SE, Burland V, Plunkett G, 3rd et al. (1993) Sequence analysis of four new heat-shock genes constituting the *hslTS/ibpAB* and *hslVU* operons in *Escherichia coli*. Gene 134(1):1–6

Clark ED (2001) Protein refolding for industrial processes. Curr Opin Biotechnol 12(2):202–7

Corchero JL, Cubarsi R, Enfors S et al. (1997) Limited *in vivo* proteolysis of aggregated proteins. Biochem Biophys Res Commun 237(2):325–30

Cubarsi R, Carrio MM, Villaverde A (2005) A mathematical approach to molecular organization and proteolytic disintegration of bacterial inclusion bodies. Math Med Biol 22(3):209–26

Dale GE, Schonfeld HJ, Langen H et al. (1994) Increased solubility of trimethoprim-resistant type S1 DHFR from *Staphylococcus aureus* in *Escherichia coli* cells overproducing the chaperonins GroEL and GroES. Protein Eng 7(7):925–31

de Groot NS, Ventura S (2006) Protein activity in bacterial inclusion bodies correlates with predicted aggregation rates. J Biotechnol 125(1):110–3

de Marco A, Deuerling E, Mogk A et al. (2007) Chaperone-based procedure to increase yields of soluble recombinant proteins produced in *E. coli*. BMC Biotechnol 7:32

de Marco A, Volrath S, Bruyere T et al. (2000) Recombinant maize protoporphyrinogen IX oxidase expressed in *Escherichia coli* forms complexes with GroEL and DnaK chaperones. Protein Expr Purif 20(1):81–6

Dill KA, Chan HS (1997) From Levinthal to pathways to funnels. Nat Struct Biol 4(1):10–9

Dinner AR, Sali A, Smith LJ et al. (2000) Understanding protein folding via free-energy surfaces from theory and experiment. Trends Biochem Sci 25(7):331–9
Dobson CM (2004) Principles of protein folding, misfolding and aggregation. Semin Cell Dev Biol 15(1):3–16
Echave P, Esparza-Ceron MA, Cabiscol E et al. (2002) DnaK dependence of mutant ethanol oxidoreductases evolved for aerobic function and protective role of the chaperone against protein oxidative damage in *Escherichia coli*. Proc Natl Acad Sci USA 99(7):4626–31
Ehrnsperger M, Graber S, Gaestel M et al. (1997) Binding of non-native protein to Hsp25 during heat shock creates a reservoir of folding intermediates for reactivation. Embo J 16(2):221–9
Ellis J (1987) Proteins as molecular chaperones. Nature 328(6129):378–9
Ellis RJ (2001) Macromolecular crowding: obvious but underappreciated. Trends Biochem Sci 26(10):597–604
Enfors SO (1992) Control of *in vivo* proteolysis in the production of recombinant proteins. Trends Biotechnol 10(9):310–5
Esposito D, Chatterjee DK (2006) Enhancement of soluble protein expression through the use of fusion tags. Curr Opin Biotechnol 17(4):353–8
Estapé D, Rinas U (1996) Optimized procedures for purification and solubilization of basic fibroblast growth factor inclusion bodies. Biotechnol. Tech 10(7):481–484
Ewalt KL, Hendrick JP, Houry WA et al. (1997) *In vivo* observation of polypeptide flux through the bacterial chaperonin system. Cell 90(3):491–500
Fahnert B, Lilie H, Neubauer P (2004) Inclusion bodies: formation and utilisation. Adv Biochem Eng Biotechnol 89:93–142
Fayet O, Ziegelhoffer T, Georgopoulos C (1989) The *groES* and *groEL* heat shock gene products of *Escherichia coli* are essential for bacterial growth at all temperatures. J Bacteriol 171(3): 1379–85
Ferrer M, Chernikova TN, Yakimov MM et al. (2003) Chaperonins govern growth of *Escherichia coli* at low temperatures. Nat Biotechnol 21(11):1266–7
Fink AL (1998) Protein aggregation: folding aggregates, inclusion bodies and amyloid. Fold Des 3(1):R9–23
Fredriksson A, Ballesteros M, Dukan S et al. (2005) Defense against protein carbonylation by DnaK/DnaJ and proteases of the heat shock regulon. J Bacteriol 187(12):4207–13
Garcia-Fruitos E, Aris A, Villaverde A (2007a) Localization of functional polypeptides in bacterial inclusion bodies. Appl Environ Microbiol 73(1):289–94
Garcia-Fruitos E, Carrio MM, Aris A et al. (2005a) Folding of a misfolding-prone beta-galactosidase in absence of DnaK. Biotechnol Bioeng 90(7):869–75
Garcia-Fruitos E, Gonzalez-Montalban N, Morell M et al. (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 et al. (2007b) Divergent genetic control of protein solubility and conformational quality in *Escherichia coli*. J Mol Biol 374(1):195–205
Gasser B, Saloheimo M, Rinas U et al. (2008) Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. Microb Cell Fact 7:11
Genevaux P, Georgopoulos C, Kelley WL (2007) The Hsp70 chaperone machines of *Escherichia coli*: a paradigm for the repartition of chaperone functions. Mol Microbiol 66(4):840–57
Genevaux P, Schwager F, Georgopoulos C et al. (2001) The *djlA* gene acts synergistically with *dnaJ* in promoting *Escherichia coli* growth. J Bacteriol 183(19):5747–50
Georgiou G, Valax P (1996) Expression of correctly folded proteins in *Escherichia coli*. Curr Opin Biotechnol 7(2):190–7
Georgiou G, Valax P, Ostermeier M et al. (1994) Folding and aggregation of TEM beta-lactamase: analogies with the formation of inclusion bodies in *Escherichia coli*. Protein Sci 3(11):1953–60
Gething MJ, Sambrook J (1992) Protein folding in the cell. Nature 355(6355):33–45
Glover JR, Lindquist S (1998) Hsp104, Hsp70, and Hsp40: a novel chaperone system that rescues previously aggregated proteins. Cell 94(1):73–82

Glover JR, Tkach JM (2001) Crowbars and ratchets: hsp100 chaperones as tools in reversing protein aggregation. Biochem Cell Biol 79(5):557–68

Goloubinoff P, Mogk A, Zvi AP et al. (1999) Sequential mechanism of solubilization and refolding of stable protein aggregates by a bichaperone network. Proc Natl Acad Sci USA 96(24): 13732–7

Gonzalez-Montalban N, Garcia-Fruitos E, Ventura S et al. (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 (2007a) Recombinant protein solubility – does more mean better? Nat Biotechnol 25(7):718–20

Gonzalez-Montalban N, Natalello A, Garcia-Fruitos E et al. (2008) In situ protein folding and activation in bacterial inclusion bodies. Biotechnol Bioeng 100(4):797–802

Gonzalez-Montalban N, Villaverde A, Aris A (2007b) Amyloid-linked cellular toxicity triggered by bacterial inclusion bodies. Biochem Biophys Res Commun 355(3):637–42

Gottesman S, Wickner S, Maurizi MR (1997) Protein quality control: triage by chaperones and proteases. Genes Dev 11(7):815–23

Graf PC, Jakob U (2002) Redox-regulated molecular chaperones. Cell Mol Life Sci 59(10): 1624–31

Grantcharova V, Alm EJ, Baker D et al. (2001) Mechanisms of protein folding. Curr Opin Struct Biol 11(1):70–82

Grudzielanek S, Velkova A, Shukla A et al. (2007) Cytotoxicity of insulin within its self-assembly and amyloidogenic pathways. J Mol Biol 370(2):372–84

Gupta RS, Singh B (1994) Phylogenetic analysis of 70 kD heat shock protein sequences suggests a chimeric origin for the eukaryotic cell nucleus. Curr Biol 4(12):1104–14

Han MJ, Park SJ, Park TJ et al. (2004) Roles and applications of small heat shock proteins in the production of recombinant proteins in *Escherichia coli*. Biotechnol Bioeng 88(4): 426–36

Harris TJ, Patel T, Marston FA et al. (1986) Cloning of cDNA coding for human tissue-type plasminogen activator and its expression in *Escherichia coli*. Mol Biol Med 3(3):279–92

Hart RA, Rinas U, Bailey JE (1990) Protein composition of Vitreoscilla hemoglobin inclusion bodies produced in *Escherichia coli*. J Biol Chem 265(21):12728–33

Hartl FU, Hayer-Hartl M (2002) Molecular chaperones in the cytosol: from nascent chain to folded protein. Science 295(5561):1852–8

Hartley DL, Kane JF (1988) Properties of inclusion bodies from recombinant *Escherichia coli*. Biochem Soc Trans 16(2):101–2

Haslbeck M (2002) sHsps and their role in the chaperone network. Cell Mol Life Sci 59(10): 1649–57

Hennessy F, Nicoll WS, Zimmermann R et al. (2005) Not all J domains are created equal: implications for the specificity of Hsp40-Hsp70 interactions. Protein Sci 14(7):1697–709

Hoffmann F, Rinas U (2000) Kinetics of heat-shock response and inclusion body formation during temperature-induced production of basic fibroblast growth factor in high-cell-density cultures of recombinant *Escherichia coli*. Biotechnol Prog 16(6):1000–7

Hoffmann F, Rinas U (2001) On-line estimation of the metabolic burden resulting from the synthesis of plasmid-encoded and heat-shock proteins by monitoring respiratory energy generation. Biotechnol Bioeng 76(4):333–40

Hoffmann F, Rinas U (2004) Roles of heat-shock chaperones in the production of recombinant proteins in *Escherichia coli*. Adv Biochem Eng Biotechnol 89:143–61

Hoskins JR, Pak M, Maurizi MR et al. (1998) The role of the ClpA chaperone in proteolysis by ClpAP. Proc Natl Acad Sci USA 95(21):12135–40

Huang GC, Li ZY, Zhou JM et al. (2000) Assisted folding of D-glyceraldehyde-3-phosphate dehydrogenase by trigger factor. Protein Sci 9(6):1254–61

Hunke S, Betton JM (2003) Temperature effect on inclusion body formation and stress response in the periplasm of *Escherichia coli*. Mol Microbiol 50(5):1579–89

Hunt C, Morimoto RI (1985) Conserved features of eukaryotic hsp70 genes revealed by comparison with the nucleotide sequence of human hsp70. Proc Natl Acad Sci USA 82(19): 6455–9

Ivanova MI, Sawaya MR, Gingery M et al. (2004) An amyloid-forming segment of beta2-microglobulin suggests a molecular model for the fibril. Proc Natl Acad Sci USA 101(29):10584–9

Jackson SE (1998) How do small single-domain proteins fold? Fold Des 3(4):R81–91

Jevsevar S, Gaberc-Porekar V, Fonda I et al. (2005) Production of nonclassical inclusion bodies from which correctly folded protein can be extracted. Biotechnol Prog 21(2):632–9

Jungbauer A, Kaar W (2007) Current status of technical protein refolding. J Biotechnol 128(3):587–96

Jurgen B, Lin HY, Riemschneider S et al. (2000) Monitoring of genes that respond to overproduction of an insoluble recombinant protein in *Escherichia coli* glucose-limited fed-batch fermentations. Biotechnol Bioeng 70(2):217–24

Kapust RB, Tozser J, Copeland TD et al. (2002) The P1′ specificity of tobacco etch virus protease. Biochem Biophys Res Commun 294(5):949–55

Kapust RB, Tozser J, Fox JD et al. (2001) Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. Protein Eng 14(12): 993–1000

Karplus M (1997) The Levinthal paradox: yesterday and today. Fold Des 2(4):S69–75

Kazemi-Esfarjani P, Benzer S (2000) Genetic suppression of polyglutamine toxicity in Drosophila. Science 287(5459):1837–40

Kedzierska S, Staniszewska M, Wegrzyn A et al. (1999) The role of DnaK/DnaJ and GroEL/GroES systems in the removal of endogenous proteins aggregated by heat-shock from *Escherichia coli* cells. FEBS Lett 446(2–3):331–7

Kerner MJ, Naylor DJ, Ishihama Y et al. (2005) Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. Cell 122(2):209–20

Khan F, Chuang JI, Gianni S et al. (2003) The kinetic pathway of folding of barnase. J Mol Biol 333(1):169–86

Kitagawa M, Miyakawa M, Matsumura Y et al. (2002) *Escherichia coli* small heat shock proteins, IbpA and IbpB, protect enzymes from inactivation by heat and oxidants. Eur J Biochem 269(12):2907–17

Krobitsch S, Lindquist S (2000) Aggregation of huntingtin in yeast varies with the length of the polyglutamine expansion and the expression of chaperone proteins. Proc Natl Acad Sci USA 97(4):1589–94

Kuczynska-Wisnik D, Kedzierska S, Matuszewska E et al. (2002) The *Escherichia coli* small heat-shock proteins IbpA and IbpB prevent the aggregation of endogenous proteins denatured *in vivo* during extreme heat shock. Microbiology 148(Pt 6):1757–65

Kuczynska-Wisnik D, Zurawa-Janicka D, Narkiewicz J et al. (2004) *Escherichia coli* small heat shock proteins IbpA/B enhance activity of enzymes sequestered in inclusion bodies. Acta Biochim Pol 51(4):925–31

Laskey RA, Honda BM, Mills AD et al. (1978) Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA. Nature 275(5679):416–20

Laskowska E, Bohdanowicz J, Kuczynska-Wisnik D et al. (2004) Aggregation of heat-shock-denatured, endogenous proteins and distribution of the IbpA/B and Fda marker-proteins in *Escherichia coli* WT and grpE280 cells. Microbiology 150(Pt 1):247–59

Lemaux PG, Herendeen SL, Bloch PL et al. (1978) Transient rates of synthesis of individual polypeptides in *E. coli* following temperature shifts. Cell 13(3):427–34

Lethanh H, Neubauer P, Hoffmann F (2005) The small heat-shock proteins IbpA and IbpB reduce the stress load of recombinant *Escherichia coli* and delay degradation of inclusion bodies. Microb Cell Fact 4(1):6

Levchenko I, Luo L, Baker TA (1995) Disassembly of the Mu transposase tetramer by the ClpX chaperone. Genes Dev 9(19):2399–408

Lewandowska A, Matuszewska M, Liberek K (2007) Conformational properties of aggregated polypeptides determine ClpB-dependence in the disaggregation process. J Mol Biol 371(3):800–11

Lindquist S, Craig EA (1988) The heat-shock proteins. Annu Rev Genet 22:631–77

Malki A, Kern R, Abdallah J et al. (2003) Characterization of the *Escherichia coli* YedU protein as a molecular chaperone. Biochem Biophys Res Commun 301(2):430–6

Marston FA (1986) The purification of eukaryotic polypeptides synthesized in *Escherichia coli*. Biochem J 240(1):1–12

Marston FA, Hartley DL (1990) Solubilization of protein aggregates. Methods Enzymol 182: 264–76

Marston FAO, Lowe PA, Doel MT et al. (1984) Purification of Calf Prochymosin(Prorennin) Synthesized in *Escherichia coli*. Bio/Technology 2(9):800–804

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(2):187–95

Martinez-Alonso M, González-Montalbán N, Garcia-Fruitós E, Villaverde A. (2008) The functional quality of soluble recombinant polypeptides produced in *Escherichia coil* is defined by a wide conformational spectrum. Appl Environ Microbiol 74(23):7431–3

Matagne A, Dobson CM (1998) The folding process of hen lysozyme: a perspective from the 'new view'. Cell Mol Life Sci 54(4):363–71

Matouschek A (2003) Protein unfolding–an important process *in vivo*? Curr Opin Struct Biol 13(1):98–109

Matuszewska M, Kuczynska-Wisnik D, Laskowska E et al. (2005) The small heat shock protein IbpA of *Escherichia coli* cooperates with IbpB in stabilization of thermally aggregated proteins in a disaggregation competent state. J Biol Chem 280(13):12292–8

Maurizi MR (1992) Proteases and protein degradation in *Escherichia coli*. Experientia 48(2): 178–201

Middelberg AP (2002) Preparative protein refolding. Trends Biotechnol 20(10):437–43

Miot M, Betton JM (2004) Protein quality control in the bacterial periplasm. Microb Cell Fact 3(1):4

Missiakas D, Schwager F, Betton JM et al. (1996) Identification and characterization of HsIV HsIU (ClpQ ClpY) proteins involved in overall proteolysis of misfolded proteins in *Escherichia coli*. Embo J 15(24):6899–909

Mogk A, Bukau B (2004) Molecular chaperones: structure of a protein disaggregase. Curr Biol 14(2):R78–80

Mogk A, Deuerling E, Vorderwulbecke S et al. (2003) Small heat shock proteins, ClpB and the DnaK system form a functional triade in reversing protein aggregation. Mol Microbiol 50(2):585–95

Mogk A, Tomoyasu T, Goloubinoff P et al. (1999) Identification of thermolabile *Escherichia coli* proteins: prevention and reversion of aggregation by DnaK and ClpB. Embo J 18(24): 6934–49

Morell M, Bravo R, Espargaro A et al. (2008): Inclusion bodies: Specificity in their aggregation process and amyloid-like structure. Biochim Biophys Acta 1783(10):1815–25

Morita MT, Kanemori M, Yanagi H et al. (2000) Dynamic interplay between antagonistic pathways controlling the sigma 32 level in *Escherichia coli*. Proc Natl Acad Sci U S A 97(11):5860–5

Motohashi K, Watanabe Y, Yohda M et al. (1999) Heat-inactivated proteins are rescued by the DnaK.J-GrpE set and ClpB chaperones. Proc Natl Acad Sci USA 96(13):7184–9

Muchowski PJ, Schaffar G, Sittler A et al. (2000) Hsp70 and hsp40 chaperones can inhibit self-assembly of polyglutamine proteins into amyloid-like fibrils. Proc Natl Acad Sci USA 97(14):7841–6

Mujacic M, Bader MW, Baneyx F (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(3):849–59

Muramatsu N, Minton AP (1988) Tracer diffusion of globular proteins in concentrated protein solutions. Proc Natl Acad Sci USA 85(9):2984–8

Nagai H, Yuzawa H, Kanemori M et al. (1994) A distinct segment of the sigma 32 polypeptide is involved in DnaK-mediated negative control of the heat shock response in *Escherichia coli*. Proc Natl Acad Sci USA 91(22):10280–4

Nahalka J (2008) 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 35(4):219–23

Nahalka J, Gemeiner P, Bucko M et al. (2006) Bioenergy beads: a tool for regeneration of ATP/NTP in biocatalytic synthesis. Artif Cells Blood Substit Immobil Biotechnol 34(5):515–21

Nahalka J, Nidetzky B (2007) Fusion to a pull-down domain: a novel approach of producing Trigonopsis variabilisD-amino acid oxidase as insoluble enzyme aggregates. Biotechnol Bioeng 97(3):454–61

Nahalka J, Vikartovska A, Hrabarova E (2008) A crosslinked inclusion body process for sialic acid synthesis. J Biotechnol 134(1–2):146–53

Narberhaus F (2002) Alpha-crystallin-type heat shock proteins: socializing minichaperones in the context of a multichaperone network. Microbiol Mol Biol Rev 66(1):64–93; table of contents

Neidhardt FC, Umbarger HE (1996) Chemical composition of *Escherichia coli*. In: (ed) *Escherichia coli* and *Salmonella*, American Society for Microbiology Press, Washington, D.C.

Neubauer A, Soini J, Bollok M et al. (2007) Fermentation process for tetrameric human collagen prolyl 4-hydroxylase in *Escherichia coli*: improvement by gene optimisation of the PDI/beta subunit and repeated addition of the inducer anhydrotetracycline. J Biotechnol 128(2): 308–21

Niiranen L, Espelid S, Karlsen CR et al. (2007) Comparative expression study to increase the solubility of cold adapted *Vibrio* proteins in *Escherichia coli*. Protein Expr Purif 52(1):210–8

Nishihara K, Kanemori M, Kitagawa M et al. (1998) Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in *Escherichia coli*. Appl Environ Microbiol 64(5):1694–9

Nishihara K, Kanemori M, Yanagi H et al. (2000) Overexpression of trigger factor prevents aggregation of recombinant proteins in *Escherichia coli*. Appl Environ Microbiol 66(3):884–9

Oberg K, Chrunyk BA, Wetzel R et al. (1994) Nativelike secondary structure in interleukin-1 beta inclusion bodies by attenuated total reflectance FTIR. Biochemistry 33(9):2628–34

Otzen DE, Fersht AR (1998) Folding of circular and permuted chymotrypsin inhibitor 2: retention of the folding nucleus. Biochemistry 37(22):8139–46

Panchenko AR, Luthey-Schulten Z, Wolynes PG (1996) Foldons, protein structural modules, and exons. Proc Natl Acad Sci USA 93(5):2008–13

Parsell DA, Kowal AS, Singer MA et al. (1994) Protein disaggregation mediated by heat-shock protein Hsp104. Nature 372(6505):475–8

Paul DC, Van Frank RM, Muth WL et al. (1983) Immunocytochemical demonstration of human proinsulin chimeric polypeptide within cytoplasmic inclusion bodies of *Escherichia coli*. Eur J Cell Biol 31(2):171–4

Plakoutsi G, Bemporad F, Calamai M et al. (2005) Evidence for a mechanism of amyloid formation involving molecular reorganisation within native-like precursor aggregates. J Mol Biol 351(4):910–22

Proudfoot AE, Goffin L, Payton MA et al. (1996) *In vivo* and *in vitro* folding of a recombinant metalloenzyme, phosphomannose isomerase. Biochem J 318 (Pt 2):437–42

Przybycien TM, Dunn JP, Valax P et al. (1994) Secondary structure characterization of beta-lactamase inclusion bodies. Protein Eng 7(1):131–6

Rajan RS, Illing ME, Bence NF et al. (2001) Specificity in intracellular protein aggregation and inclusion body formation. Proc Natl Acad Sci USA 98(23):13060–5

Ranson NA, Farr GW, Roseman AM et al. (2001) ATP-bound states of GroEL captured by cryo-electron microscopy. Cell 107(7):869–79

Rinas U, Bailey JE (1992) Protein compositional analysis of inclusion bodies produced in recombinant *Escherichia coli*. Appl Microbiol Biotechnol 37(5):609–14

Rinas U, Bailey JE (1993) Overexpression of bacterial hemoglobin causes incorporation of prebeta-lactamase into cytoplasmic inclusion bodies. Appl Environ Microbiol 59(2):561–6

Rinas U, Boone TC, Bailey JE (1993) Characterization of inclusion bodies in recombinant *Escherichia coli* producing high levels of porcine somatotropin. J Biotechnol 28(2–3):313–20

Rinas U, Hoffmann F, Betiku E et al. (2007) Inclusion body anatomy and functioning of chaperone-mediated *in vivo* inclusion body disassembly during high-level recombinant protein production in *Escherichia coli*. J Biotechnol 127(2):244–57

Rosen R, Biran D, Gur E et al. (2002) Protein aggregation in *Escherichia coli*: role of proteases. FEMS Microbiol Lett 207(1):9–12

Rudolph R, Lilie H (1996) *In vitro* folding of inclusion body proteins. Faseb J 10(1):49–56

Sahdev S, Khattar SK, Saini KS (2008) Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. Mol Cell Biochem 307(1–2):249–64

Sakikawa C, Taguchi H, Makino Y et al. (1999) On the maximum size of proteins to stay and fold in the cavity of GroEL underneath GroES. J Biol Chem 274(30):21251–6

Sastry MS, Korotkov K, Brodsky Y et al. (2002) Hsp31, the *Escherichia coli yedU* gene product, is a molecular chaperone whose activity is inhibited by ATP at high temperatures. J Biol Chem 277(48):46026–34

Schein CH, Noteborn (1988) Formation of soluble recombinant proteins in *Escherichia coli* is favored by lower growth temperature. Bio/Technology 6(3):291–294

Schirmer EC, Glover JR, Singer MA et al. (1996) HSP100/Clp proteins: a common mechanism explains diverse functions. Trends Biochem Sci 21(8):289–96

Schlieker C, Tews I, Bukau B et al. (2004) Solubilization of aggregated proteins by ClpB/DnaK relies on the continuous extraction of unfolded polypeptides. FEBS Lett 578(3):351–6

Schmidt M, Babu KR, Khanna N et al. (1999) Temperature-induced production of recombinant human insulin in high-cell density cultures of recombinant *Escherichia coli*. J Biotechnol 68(1):71–83

Schoemaker JM, Brasnett AH, Marston FA (1985) Examination of calf prochymosin accumulation in *Escherichia coli*: disulphide linkages are a structural component of prochymosin-containing inclusion bodies. Embo J 4(3):775–80

Schoner RG, Ellis LF, Schoner BE (1985) Isolation and purification of protein granules from *Escherichia coli* cells overproducing bovine growth hormone. Bio/Technology 3:151–154

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

Schultz T, Martinez L, de Marco A (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

Shearstone JR, Baneyx F (1999) Biochemical characterization of the small heat shock protein IbpB from *Escherichia coli*. J Biol Chem 274(15):9937–45

Snow CD, Nguyen H, Pande VS et al. (2002) Absolute comparison of simulated and experimental protein-folding dynamics. Nature 420(6911):102–6

Sorensen HP, Mortensen KK (2005a) Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. J Biotechnol 115(2):113–28

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

Speed MA, Wang DI, King J (1996) Specific aggregation of partially folded polypeptide chains: the molecular basis of inclusion body composition. Nat Biotechnol 14(10):1283–7

Spiess C, Beil A, Ehrmann M (1999) A temperature-dependent switch from chaperone to protease in a widely conserved heat shock protein. Cell 97(3):339–47

Stefani M, Dobson CM (2003) Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. J Mol Med 81(11):678–99

Straus DB, Walter WA, Gross CA (1987) The heat shock response of *E. coli* is regulated by changes in the concentration of sigma 32. Nature 329(6137):348–51

Sugrue R, Marston FA, Lowe PA et al. (1990) Denaturation studies on natural and recombinant bovine prochymosin (prorennin). Biochem J 271(2):541–7

Surewicz WK, Mantsch HH (1988) New insight into protein secondary structure from resolution-enhanced infrared spectra. Biochim Biophys Acta 952(2):115–30

Surewicz WK, Stepanik TM, Szabo AG et al. (1988) Lipid-induced changes in the secondary structure of snake venom cardiotoxins. J Biol Chem 263(2):786–90

Teter SA, Houry WA, Ang D et al. (1999) Polypeptide flux through bacterial Hsp70: DnaK cooperates with trigger factor in chaperoning nascent chains. Cell 97(6):755–65

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

Thomas JG, Baneyx F (1996a) Protein folding in the cytoplasm of *Escherichia coli*: requirements for the DnaK-DnaJ-GrpE and GroEL-GroES molecular chaperone machines. Mol Microbiol 21(6):1185–96

Thomas JG, Baneyx F (1996b) Protein misfolding and inclusion body formation in recombinant *Escherichia coli* cells overexpressing Heat-shock proteins. J Biol Chem 271(19):11141–7

Thomas JG, Baneyx F (1998) Roles of the *Escherichia coli* small heat shock proteins IbpA and IbpB in thermal stress management: comparison with ClpA, ClpB, and HtpG *In vivo*. J Bacteriol 180(19):5165–72

Thomas JG, Baneyx F (2000) ClpB and HtpG facilitate de novo protein folding in stressed *Escherichia coli* cells. Mol Microbiol 36(6):1360–70

Tokatlidis K, Dhurjati P, Millet J et al. (1991) High activity of inclusion bodies formed in *Escherichia coli* overproducing *Clostridium thermocellum* endoglucanase D. FEBS Lett 282(1):205–8

Tomoyasu T, Arsene F, Ogura T et al. (2001a) The C terminus of sigma(32) is not essential for degradation by FtsH. J Bacteriol 183(20):5911–7

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

Tomoyasu T, Ogura T, Tatsuta T et al. (1998) Levels of DnaK and DnaJ provide tight control of heat shock gene expression and protein repair in *Escherichia coli*. Mol Microbiol 30(3):567–81

Tsumoto K, Umetsu M, Kumagai I et al. (2003) Solubilization of active green fluorescent protein from insoluble particles by guanidine and arginine. Biochem Biophys Res Commun 312(4):1383–6

Umetsu M, Tsumoto K, Ashish K et al. (2004) Structural characteristics and refolding of *in vivo* aggregated hyperthermophilic archaeon proteins. FEBS Lett 557(1–3):49–56

Umetsu M, Tsumoto K, Nitta S et al. (2005) Nondenaturing solubilization of beta2 microglobulin from inclusion bodies by L-arginine. Biochem Biophys Res Commun 328(1):189–97

Vallejo LF, Rinas U (2004) Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins. Microb Cell Fact 3(1):11

Veinger L, Diamant S, Buchner J et al. (1998) The small heat-shock protein IbpB from *Escherichia coli* stabilizes stress-denatured proteins for subsequent refolding by a multichaperone network. J Biol Chem 273(18):11032–7

Vendruscolo M, Paci E, Dobson CM et al. (2003) Rare fluctuations of native proteins sampled by equilibrium hydrogen exchange. J Am Chem Soc 125(51):15686–7

Ventura S (2005) Sequence determinants of protein aggregation: tools to increase protein solubility. Microb Cell Fact 4(1):11

Ventura S, Villaverde A (2006) Protein quality in bacterial inclusion bodies. Trends Biotechnol 24(4):179–85

Ventura S, Zurdo J, Narayanan S et al. (2004) Short amino acid stretches can mediate amyloid formation in globular proteins: the Src homology 3 (SH3) case. Proc Natl Acad Sci USA 101(19):7258–63

Vera A, Aris A, Carrio M et al. (2005) Lon and ClpP proteases participate in the physiological disintegration of bacterial inclusion bodies. J Biotechnol 119(2):163–71
Vera A, Gonzalez-Montalban N, Aris A et al. (2007) The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. Biotechnol Bioeng 96(6):1101–6
Villaverde A, Carrio MM (2003) Protein aggregation in recombinant bacteria: biological role of inclusion bodies. Biotechnol Lett 25(17):1385–95
Wagner S, Baars L, Ytterberg AJ et al. (2007) Consequences of membrane protein overexpression in *Escherichia coli*. Mol Cell Proteomics 6(9):1527–50
Waldo GS, Standish BM, Berendzen J et al. (1999) Rapid protein-folding assay using green fluorescent protein. Nat Biotechnol 17(7):691–5
Walsh G (2005) Therapeutic insulins and their large-scale manufacture. Appl Microbiol Biotechnol 67(2):151–9
Wang L, Maji SK, Sawaya MR et al. (2008): Bacterial inclusion bodies contain amyloid-like strucure. PLoS Biol 6(8):e195
Warrick JM, Chan HY, Gray-Board GL et al. (1999) Suppression of polyglutamine-mediated neurodegeneration in *Drosophila* by the molecular chaperone HSP70. Nat Genet 23(4):425–8
Wickner S, Gottesman S, Skowyra D et al. (1994) A molecular chaperone, ClpA, functions like DnaK and DnaJ. Proc Natl Acad Sci USA 91(25):12218–22
Wickner S, Maurizi MR, Gottesman S (1999) Posttranslational quality control: folding, refolding, and degrading proteins. Science 286(5446):1888–93
Williams DC, Van Frank RM, Muth WL et al. (1982) Cytoplasmic inclusion bodies in *Escherichia coli* producing biosynthetic human insulin proteins. Science 215(4533):687–9
Wolynes PG, Onuchic JN, Thirumalai D (1995) Navigating the folding routes. Science 267(5204):1619–20
Worrall DM, Goss NH (1989) The formation of biologically active beta-galactosidase inclusion bodies in *Escherichia coli*. Aust J Biotechnol 3(1):28–32
Yokoyama K, Kikuchi Y, Yasueda H (1998) Overproduction of DnaJ in *Escherichia coli* improves *in vivo* solubility of the recombinant fish-derived transglutaminase. Biosci Biotechnol Biochem 62(6):1205–10
Yon JM (2001) Protein folding: a perspective for biology, medicine and biotechnology. Braz J Med Biol Res 34(4):419–35
Young JC, Agashe VR, Siegers K et al. (2004) Pathways of chaperone-mediated protein folding in the cytosol. Nat Rev Mol Cell Biol 5(10):781–91
Yura T, Nakahigashi K (1999) Regulation of the heat-shock response. Curr Opin Microbiol 2(2):153–8
Zhang YB, Howitt J, McCorkle S et al. (2004) Protein aggregation during overexpression limited by peptide extensions with large net negative charge. Protein Expr Purif 36(2):207–16
Zhu X, Zhao X, Burkholder WF et al. (1996) Structural analysis of substrate binding by the molecular chaperone DnaK. Science 272(5268):1606–14
Zimmerman SB, Minton AP (1993) Macromolecular crowding: biochemical, biophysical, and physiological consequences. Annu Rev Biophys Biomol Struct 22:27–65
Zimmerman SB, Trach SO (1991) Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*. J Mol Biol 222(3):599–620
Zolkiewski M (1999) ClpB cooperates with DnaK, DnaJ, and GrpE in suppressing protein aggregation. A novel multi-chaperone system from *Escherichia coli*. J Biol Chem 274(40):28083–6