nature biotechnology

Recombinant protein folding and misfolding in *Escherichia coli*

François Baneyx & Mirna Mujacic

The past 20 years have seen enormous progress in the understanding of the mechanisms used by the enteric bacterium *Escherichia coli* to promote protein folding, support protein translocation and handle protein misfolding. Insights from these studies have been exploited to tackle the problems of inclusion body formation, proteolytic degradation and disulfide bond generation that have long impeded the production of complex heterologous proteins in a properly folded and biologically active form. The application of this information to industrial processes, together with emerging strategies for creating designer folding modulators and performing glycosylation all but guarantee that *E. coli* will remain an important host for the production of both commodity and high value added proteins.

The enteric bacterium *Escherichia coli* is one of the most extensively used prokaryotic organisms for genetic manipulations and for the industrial production of proteins of therapeutic or commercial interest. Compared with other established and emerging expression systems[1], *E. coli* offers several advantages, including growth on inexpensive carbon sources, rapid biomass accumulation, amenability to high cell-density fermentations and simple process scale-up. Because of its long history as a model system, *E. coli* genetics are very well characterized and many tools have been developed for chromosome engineering and to facilitate gene cloning and expression. If heterologous proteins do not require complex post-translational modifications and are expressed in a soluble form, *E. coli* is usually first selected to obtain enough material for biochemical and/or structural studies and for the subsequent large-scale production of valuable gene products. It is, however, not uncommon that overexpressed recombinant proteins fail to reach a correct conformation and undergo proteolytic degradation or associate with each other to form insoluble aggregates of nonnative proteins known as inclusion bodies.

Over the past 20 years, there has been considerable progress in the fundamental understanding of the mechanisms used by *E. coli* to support *de novo* protein folding, manage stress-induced protein misfolding and decide whether misfolded polypeptides should be refolded or degraded. Here, we review this body of knowledge and how it has been exploited to promote the high-level production of heterologous proteins in a correct and bioactive conformation in the bacterial cytoplasm and periplasm.

Protein misfolding and inclusion body formation

In the crowded milieu of the *E. coli* cytoplasm where transcription and translation are tightly coupled and one protein chain is released from the ribosome every 35 seconds[2], an environment where macromolecule concentration can reach 300–400 mg/ml (ref. 3), protein folding is an extraordinary challenge. In general, small (<100 residues), single domain host proteins efficiently reach a native conformation owing to their fast folding kinetics, whereas large multidomain and overexpressed recombinant proteins often require the assistance of folding modulators. Folding helpers include molecular chaperones, which favor on-pathway folding by shielding interactive surfaces from each other and from the solvent, and folding catalysts that accelerate rate-limiting steps, such as the isomerization of peptidyl-prolyl bonds from an abnormal *cis* to a *trans* conformation and the formation and reshuffling of disulfide bonds.

For a heterologous protein, failure to rapidly reach a native conformation or to interact with folding modulators in a timely fashion has two possible consequences: partial or complete deposition into insoluble aggregates known as inclusion bodies or degradation. The likelihood of misfolding is increased by the routine use of strong promoters and high inducer concentrations that can lead to product yields exceeding 50% of the total cellular protein. Under such conditions, the rate of protein aggregation is often much greater than that of proper folding and folding modulators are rapidly titrated. A second factor contributing to inclusion body formation is the inability of bacteria to support all post-translational modifications that a protein requires to fold. For instance, the formation of intra- or intermolecular disulfides is not possible in the reducing cytoplasm of wild-type *E. coli*, which results in the aggregation of certain disulfide bond–rich proteins (e.g., Fab antibody fragments).

Inclusion bodies can accumulate in the cytoplasm or periplasm depending on whether or not a recombinant protein has been engineered for secretion. The target typically accounts for 80–95% of the inclusion body material and is contaminated by outer membrane proteins, ribosomal components and a small amount of phospholipids and nucleic acids that likely adsorb upon cell lysis[4]. Folding modulators (e.g., DnaK, GroEL and IbpA/B) are sometimes—but not always—associated with inclusion bodies[5,6]. Cytoplasmic inclusion

Departments of Chemical Engineering and Bioengineering, University of Washington, Box 351750, Seattle, Washington 98195, USA. Correspondence should be addressed to F.B. (baneyx@u.washington.edu).

Published online 4 November 2004; doi:10.1038/nbt1029

Table 1 Cytoplasmic chaperones

Family	Name	Cofactors	Function	Substrate specificity	ATP requirement
Hsp100 (AAA+)[a]	ClpB		Disaggregase	Segments enriched in aromatic and basic residues	+
Hsp90	HtpG		Possible folding/secretory chaperone	Unknown	+
Hsp70	DnaK	DnaJ, GrpE	Folding chaperone	Segments of four to five hydrophobic amino acids, enriched in leucine and flanked by basic residues	+
	HscA	HscB	Iron-sulfur cluster protein assembly	LPPVK motif in iron-sulfur cluster protein assembly IscU	+
	HscC	YbeV, YbeS	σ^{70} regulation	Unknown	+
Hsp60	GroEL	GroES	Folding chaperone	α/β folds enriched in hydrophobic and basic residues	+
Hsp33	Hsp33		Holding chaperone	Unknown	–
DJ-1 superfamily	Hsp31		Holding chaperone	Unknown	–[b]
Small Hsps	IbpA, IbpB		Holding chaperone	Unknown	–[c]
PPIase	TF		Holding chaperone, PPIase	Eight amino acid motif enriched in aromatic and basic residues	–
SecB	SecB		Secretory chaperone	Nine amino acid motif enriched in aromatic and basic residues	–

[a]AAA, ATPases associated with a variety of cellular activities. [b]ATP binding negatively regulates the chaperone activity of Hsp31 at high temperatures[23]. [c]ATP binding to certain small Hsps triggers conformational changes and substrate release[17].

bodies are porous ovoids or cylinders with maximum characteristic length and volume of 1 μm and 0.6 μm^3, respectively[6–8]. However, hemispherical inclusion bodies of 0.5-μm diameter have been observed in the periplasm[7]. In the cytoplasm, inclusion bodies grow from structured folding intermediates[9] at nearly constant rates and around nucleation cores that are mutually exclusive. Thus, multiple inclusions of different sizes may be present within a single cell[6]. Because inclusion bodies are resistant to proteolysis and contain large amounts of relatively pure material, their formation is often exploited for the production of proteins that are toxic, unstable or easy to refold. Finding optimal conditions for efficient refolding requires considerable optimization, but acceptable yields can usually be achieved using established strategies[10].

Cytoplasmic folding modulators

In *E. coli* and other systems, host protein misfolding is not uncommon. It may result from premature termination of translation, failure of a newly synthesized chain to reach a correct conformation or from loss of structure triggered by environmental stress. To cope with this situation, cells have evolved largely conserved mechanisms to favor proper *de novo* folding, refold partially folded proteins, dissolve aggregates and dispose of irretrievably damaged proteins.

Molecular chaperones, a ubiquitous class of folding modulators, play a central role in the conformational quality control of the proteome by interacting with, stabilizing and remodeling a wide range of nonnative polypeptides. Although constitutively expressed under balanced growth conditions, many chaperones are upregulated upon heat shock or other insults that increase cellular protein misfolding (including heterologous protein expression[11]) and are therefore classified as stress or heat shock proteins (Hsps). Mechanistically, molecular chaperones rely on the differential exposure of structured hydrophobic domains to the solvent to bind nonpolar segments that would normally be buried within the core of their substrates. Although there are subtle differences in the composition of client protein recognition sequences (and thus some degree of selectivity in substrate capture), the typical chaperone target is a short unstructured stretch of hydrophobic amino acids flanked by basic residues and lacking acidic residues (**Table 1**). The fact that such motifs are common explains why chaperones are so promiscuous.

Molecular chaperones can be divided into three functional subclasses based on their mechanism of action (**Fig. 1**). 'Folding' chaperones (e.g., DnaK and GroEL) rely on ATP-driven conformational changes to mediate the net refolding/unfolding of their substrates. 'Holding' chaperones (e.g., IbpB) maintain partially folded proteins on their surface to await availability of folding chaperones upon stress abatement. Finally, the 'disaggregating' chaperone ClpB promotes the solubilization of proteins that have become aggregated as a result of stress.

In the *E. coli* cytoplasm, *de novo* folding involves three chaperone systems: trigger factor (TF), DnaK-DnaJ-GrpE and GroEL-GroES (reviewed in refs. 12,13). TF, a three-domain protein that binds ribosomes with moderate affinity in the vicinity of the peptide exit site, is ideally positioned to interact with short nascent chains. Although the central domain of TF exhibits peptidyl-prolyl *cis*/*trans* isomerase (PPIase) activity, proline residues are not necessary for substrate capture[14] and native TF clients are primarily large (>60 kDa) multidomain proteins, which represent 10–20% of the *E. coli* proteome[15]. Longer nascent chains or newly synthesized proteins may alternatively be captured by DnaK, a chaperone whose substrate pool overlaps with that of TF[15]. DnaK is targeted to high-affinity sites by the cochaperone DnaJ, which activates tight substrate binding by triggering hydrolysis of DnaK-bound ATP. Substrate ejection is controlled by GrpE-catalyzed ADP/ATP exchange. Once released, a newly synthesized protein may reach a native conformation, undergo additional cycles of interactions with DnaK (and possibly TF) until it folds, or be transferred to the downstream GroEL-GroES system which handles about 10% of newly synthesized host proteins[16]. GroEL is an ≈800-kDa oligomer organized as two stacked homoheptameric rings, one of which is always bound by the cochaperone GroES[12]. GroEL substrates, which consist of structured but nonnative proteins up to 60-kDa in size, are bound by the free ring and allowed to fold at infinite dilution within the central chamber in a process controlled by reversible GroES capping and conformational changes orchestrated by ATP binding and hydrolysis[12].

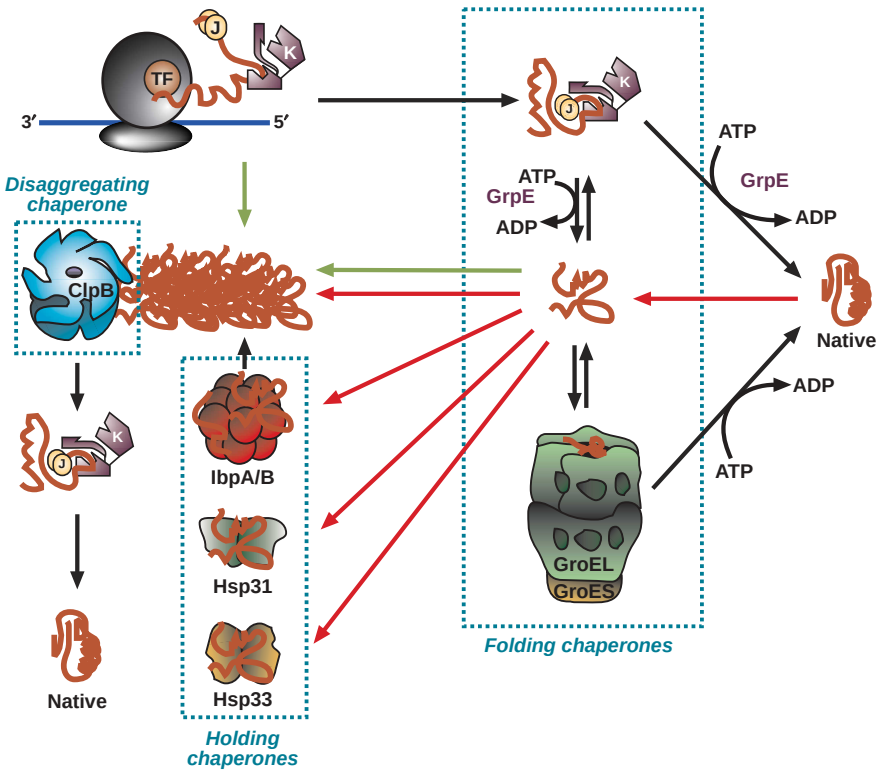

Figure 1 Chaperone-assisted protein folding in the cytoplasm of *E. coli*. Nascent polypeptides requiring the assistance of molecular chaperones first encounter TF or DnaK-DnaJ. Both chaperones engage solvent-exposed stretches of hydrophobic amino acids, shielding them from the solvent and each other. After undocking from TF- or GrpE-mediated release from DnaK, folding intermediate may reach a native conformation, cycle back to DnaK-DnaJ or be transferred to the central chamber of GroEL for folding at infinite dilution upon GroES capping. In times of stress (red arrows), thermolabile proteins unfold and aggregate. IbpB binds partially folded proteins on its surface to serve as a reservoir of unfolded intermediates until folding chaperones become available and intercalates within large aggregates. The holding chaperones Hsp33 and Hsp31 become important under oxidative and severe thermal stress, respectively. ClpB promotes the shearing and disaggregation of thermally unfolded host proteins and cooperates with DnaK-DnaJ-GrpE to reactivate them once stress has abated. Recombinant proteins that miss an early interaction with TF or DnaK/DnaJ, that undergo multiple cycles of abortive interactions with folding chaperones or titrate them out, accumulate in inclusion bodies (green arrows).

In addition to mediating proper *de novo* folding, DnaK and GroEL refold host proteins that become unfolded when cells experience environmental stress. They are assisted in this task by holding chaperones (holdases) that stabilize partially folded proteins without actively promoting their remodeling (**Fig. 1** and **Table 1**). The most extensively characterized holdases belong to the small Hsp family[17]. The bacterial representatives, IbpA and IbpB, are two homologous 16-kDa proteins encoded on a single operon[5]. IbpB forms large oligomers and relies on temperature-driven exposure of structured hydrophobic domains to capture unfolded intermediates of denaturation-prone proteins on its surface[18]. Once stress abates, IbpB-bound species are engaged by DnaK, and if necessary transferred to GroEL, for refolding[19].

Hsp33, which was identified on the basis of its thermal induction, is also classified as a holdase[20]. The main function of this redox-regulated chaperone is to manage oxidative protein misfolding[20]. Under balanced conditions, Hsp33 is a reduced monomer that coordinates a zinc atom via four conserved cysteines. When cells are exposed to reactive oxygen species—a situation that often accompanies heat shock—the cytoplasm becomes more oxidizing, Hsp33 monomers form intramolecular disulfide bonds, which trigger zinc release, and the protein adopts a dimeric conformation exhibiting chaperone activity[20]. The thioredoxin and glutaredoxin systems (see below) rapidly reduce Hsp33 disulfides in a process that does not cause substrate release but primes the chaperone for fast inactivation. Upon return to nonstress conditions, DnaK-DnaJ engage the bound substrate and refold it alone or with the help of GroEL-GroES[21].

Hsp31, a recently characterized cytoplasmic folding modulator[22,23] also functions as a holdase that binds early unfolding intermediates in times of severe stress, thereby preventing overloading of the DnaK-DnaJ-GrpE system[24]. The interface of this homodimer of 31-kDa units contains a ≈20-Å hydrophobic bowl proximal to flexible linker-loop regions that shield large nonpolar patches on either side of the bowl[25,26]. Temperature-induced motion of the linker-loop domains allows efficient capture of unfolding intermediates by uncovering high-affinity binding sites adjacent to the bowl[27]. The linker-loop region may also play a role in substrate ejection by returning to its original position upon stress abatement. Although Hsp31 is not an ATPase, its chaperone activity is negatively regulated by ATP binding at high temperature[23], possibly to coordinate substrate capture with the needs of the chaperone network.

Because folding and holding chaperones fail to abrogate protein aggregation under severe or prolonged stress conditions, *E. coli* possesses a third line of defense to manage the deleterious effects associated with misfolding: active aggregate solubilization. Disaggregation is performed by ClpB, a member of the Hsp100 family of ring-forming ATPases that also include ClpA, ClpX and ClpY, three proteins whose primary function is in proteolysis[28] (see below). The structure of *Thermus thermophilus* ClpB suggests that solubilization may rely on both a 'crowbar' action involving a long surface exposed coiled-coil domain, as well as net unfolding of the substrate by threading through the ≈16-Å central pore of the chaperone[29,30]. ClpB-mediated disaggregation is facilitated by intercalation of small Hsps within the aggregates[31] but renaturation requires transfer of partially folded substrate from ClpB to DnaK[32,33]. Interestingly, DnaK-DnaJ can solubilize small aggregates *in vitro*[34] and their interaction with large aggregates may be necessary for the initial steps of ClpB-driven disaggregation[35].

More poorly characterized folding modulators include HtpG, which may play a role in *de novo* folding and secretion, the specialized DnaK paralogs HscA and HscC (**Table 1**), and SlpA and SlyD, two homologous PPIases of unclear function.

Protein export

Proteins synthesized in the cytoplasm may remain in this compartment, integrate within the inner membrane or translocate to the periplasm[36] (**Fig. 2**). In *E. coli*, the vast majority of proteins destined for export are secreted by the Sec-dependent pathway and are

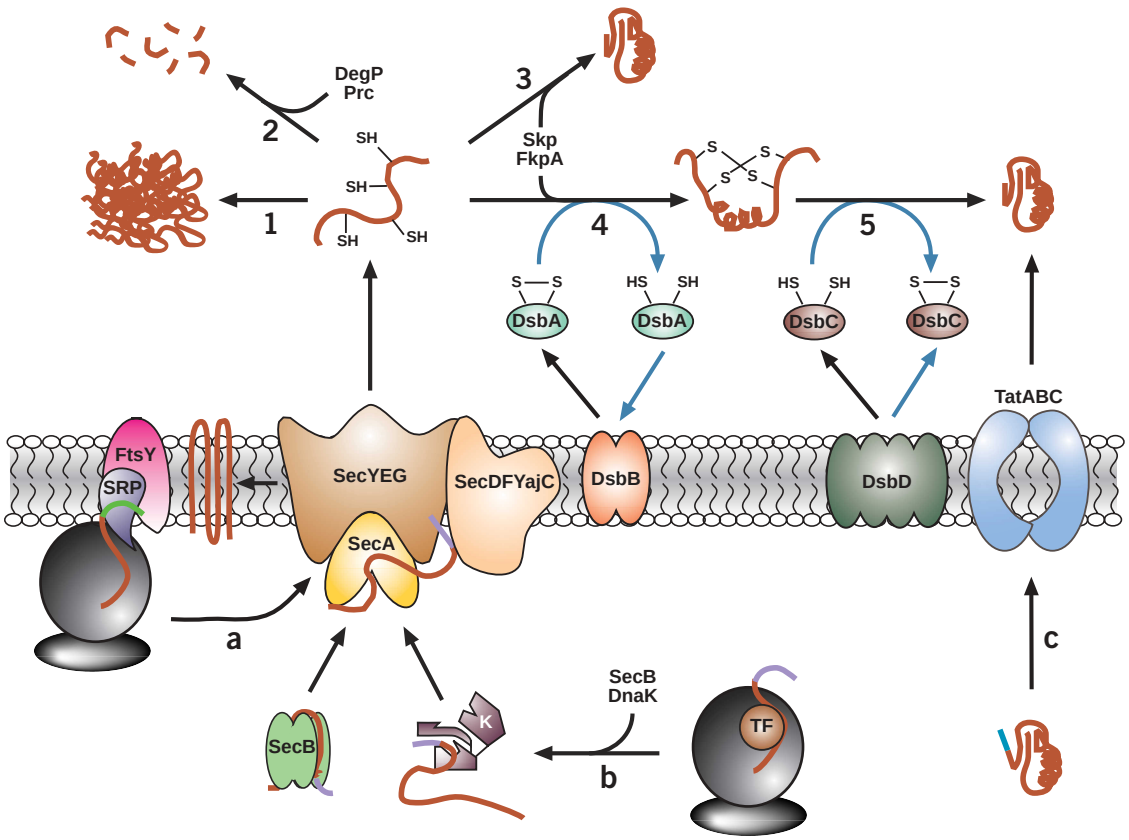

Figure 2 Export and periplasmic folding pathways. Proteins destined for export can be translocated across the inner membrane in three different fashions. (**a**) Preproteins with highly nonpolar signal sequences (green) or transmembrane segments of inner membrane proteins are recognized by SRP which, along with TF, scans nascent chains. SRP-dependent export involves delivery of the ribosome-nascent chain complex to FtsY and subsequent translocation through the SecYEG-SecDFYajC translocon. (**b**) The vast majority of preproteins have less hydrophobic signal sequence (lavender) and undergo Sec-dependent export. TF associates with the nascent polypeptide, halting cotranslational folding. As the chain grows, TF dissociates and the polypeptide is transferred to SecB or DnaK that maintain it in an extended conformation. Delivery to SecA and ATP-dependent translocation through SecYEG completes the process. (**c**) Preproteins with signal sequences containing the twin-arginine motif (cyan) are exported via the Tat-dependent pathway in a folded form. After cleavage of the signal sequence, partially folded periplasmic proteins may aggregate (1), undergo proteolysis (2) or reach a native conformation, possibly with the assistance of folding modulators (3). Cysteine pairs in proteins containing disulfide bonds are oxidized by DsbA (4) whereas incorrect disulfides are isomerized by DsbC (5). These oxidoreductases are reactivated by DsbB and DsbD, respectively. Black arrows show products obtained after each step, whereas blue arrows represent electron flow.

synthesized with an amino-terminal signal sequence, 20–30 amino acids in length, that consists of a hydrophobic core followed by a proteolytic cleavage site. Efficient export of the resulting preproteins requires targeting to the membrane-associated translocation apparatus in an extended conformation. The homotetrameric secretory chaperone SecB maintains large (>200 residues) preproteins in an export-competent form by using two 70-Å-long hydrophobic channels running along its sides[37]. Although generic chaperone such as DnaK and GroEL can also perform this duty[38], SecB has the advantage of containing an acidic 'top' region that allows it to dock and transfer the protein cargo to the peripheral membrane protein SecA[37]. Through cycles of ATP-hydrolysis, SecA drives itself and the preprotein into the pore formed by the integral membrane proteins SecYEG. Translocation to the periplasm is dependent upon the proton motive force and facilitated by the SecDFYajC complex (**Fig. 2**, path a). In the process, the signal sequence is removed by the membrane-associated Lep or Lsp signal peptidases (the latter being specific for glyceride-modified prolipoproteins).

A subset of proteins is exported via the signal recognition particle (SRP)-dependent pathway[39] (**Fig. 2**, path b). Bacterial SRP, composed of a 48-kDa GTPase termed Ffh and the 114 nt-long 4.5S RNA, can bind either to the signal sequence of certain secretory proteins (provided that it is highly hydrophobic) or to transmembrane segments of inner membrane proteins as they emerge from the ribosome. The SRP-bound ribosome nascent chain complex (RNC) is then targeted to the membrane-bound receptor FtsY. Upon GTP hydrolysis by SRP and its receptor, the SRP-RNC-FtsY complex dissociates and the RNC is transferred to SecYEG for cotranslational translocation in a process that involves SecA[40].

Because of its ability to bind nascent proteins, TF also plays a role in both Sec- and SRP-dependent protein secretion. In the former pathway, TF seems to sequester nascent chains for relatively long periods of times (perhaps until half of a typical 30-kDa protein has been translated), thereby allowing larger, unstructured proteins to be efficiently engaged by SecB[41]. In the case of SRP-dependent export, TF and SRP, which share a common attachment site (the L23 ribosomal protein), both sample nascent chains exiting from the ribosome. The appearance of a highly hydrophobic signal sequence or transmembrane segment leads to high-affinity SRP binding, and subsequent interaction with FtsY results in TF ejection and initiation of cotranslational translocation[41].

Table 2 Periplasmic folding modulators

Classification	Protein	Substrates
Generic chaperones	Skp (OmpH)	Outer membrane proteins and misfolded periplasmic proteins
	FkpA	Broad substrate range
Specialized chaperones	SurA	Outer membrane proteins
	LolA	Outer membrane lipoproteins
	PapD (and its family)	Proteins involved in P pili biosynthesis
	FimC	Proteins involved in type 1 pili biosynthesis
PPIases	SurA	Outer membrane β-barrel proteins
	PpiD	Outer membrane β-barrel proteins
	FkpA	Broad substrate range
	PpiA (RotA)	Unknown
Proteins involved in disulfide bond formation	DsbA	Reduced cell-envelope proteins
	DsbB	Reduced DsbA
	DsbC	Proteins with nonnative disulfides
	DsbG	Proteins with nonnative disulfides
	DsbD	Oxidized DsbC, DsbG and CcmG
	DsbE (CcmG)	Cytochrome *c* biogenesis
	CcmH	Cytochrome *c* biogenesis

Whereas both Sec- and SRP-dependent pathways handle preproteins that have not yet reached a native conformation, the twin-arginine (Tat)-dependent secretion pathway exclusively deals with folded or partially folded proteins. Proteins exported via the Tat pathway are produced with a signal sequence that contains a conserved—but not absolutely required[42]—twin arginine motif and most natural substrates are redox cofactor-binding proteins necessary for anaerobic respiration. Four integral membrane proteins, TatA, TatB, TatC and TatE, make up the Tat-export machinery (**Fig. 2**, path c). Although, the precise mechanism of Tat-dependent transport remains controversial[43], it has been postulated that the TatBC complex recognizes substrate proteins and delivers them to TatA, which forms a transport channel capable of accommodating substrates with diameters of up to 70 Å[44]. TatE seems to be interchangeable with TatA[45].

Periplasmic folding modulators

The periplasm contains a single bona fide chaperone termed Skp that captures unfolded proteins as they emerge from the Sec translocation apparatus (**Fig. 2**) and whose primary function is to assist the folding and membrane insertion of outer membrane proteins[46,47]. In the Skp homotrimer, α-helical tentacles extending from a β-barrel body define a central cavity that can accommodate nonnative substrates or folding modules up to ≈20-kDa in size[48]. Consistent with the absence of a periplasmic ATP pool, Skp chaperone activity is ATP independent[48] and this chaperone is likely a holdase.

Other periplasmic folding modulators include the PPIases SurA, FkpA, PpiA and PpiD (**Table 2**). Among these, FkpA has the most generic folding activity[49] and combines PPIase and chaperone functions[50]. FkpA is a V-shaped homodimer with N-terminal segments responsible for dimerization and chaperone activity and C-terminal PPIase domains[51]. FkpA is believed to cradle partially folded substrates within the hydrophobic cleft formed at the dimerization interface, allowing the flexible C-terminal domains easy access to prolyl bonds requiring isomerization.

SurA, which contains two parvulin-like PPIase domains, relies on its chaperone activity–rather than PPIase activity–to support the maturation of trimeric outer membrane proteins[52]. It resembles an asymmetric dumbbell and contains a deep cleft within its core module that may be responsible for substrate binding[53]. The fact that SurA preferentially recognizes an *Ar*-X-*Ar* motif (where *Ar* is an aromatic and X any residue) that is common in outer membrane proteins but infrequent in other polypeptides may explain its substrate specificity[54].

One of the features that distinguish the periplasm from the cytoplasm is its oxidizing environment. Indeed, in wild-type *E. coli*, stably disulfide-bonded proteins are only found in the cell envelope where disulfide formation and isomerization is catalyzed by a set of thiol-disulfide oxidoreductases known as the Dsb proteins[55,56]. DsbA, a soluble periplasmic protein containing a C-P-H-G active site embedded in a thioredoxin-like fold uses its highly reactive Cys30 to promote disulfide transfer to substrate proteins by the formation of mixed disulfide species (**Fig. 2**, path 4). It is kept in an active oxidized state by DsbB, an inner membrane protein exposing two loops to the periplasm, each containing two cysteines. DsbA recycling involves initial attack of the Cys104-Cys130 DsbB disulfide and coordinated involvement of all four DsbB cysteines before release of oxidized DsbA[56]. If incorrect disulfide bonds form in proteins containing more than two cysteines, the disulfide bond isomerase DsbC comes to the rescue (**Fig. 2**, path 5). This soluble V-shaped homodimer is structurally similar to Skp with N-terminal dimerization domains and C-terminal thioredoxin folds containing C-G-Y-C active sites[57]. DsbC is thought to capture folding intermediates within the uncharged cleft formed by its dimerization interface and to use its reduced Cys98 to attack disulfides in substrate proteins, thereby catalyzing isomerization in a process involving mixed disulfide intermediates. DsbC is maintained in a reduced state by the inner membrane protein DsbD at the expense of NADH oxidation in the cytoplasm[55,56]. DsbA, DsbC and DsbG all exhibit chaperone activity, presumably because a partially folded structure is needed to allow efficient disulfide formation and isomerization in substrate proteins.

Proteolysis

The degradation of misfolded proteins by host proteases guarantees that abnormal polypeptides do not accumulate within the cell and allows amino acid recycling. Targets for degradation include prematurely terminated polypeptides, proteolytically vulnerable folding intermediates that are kinetically trapped off-pathway, and partially folded proteins that have failed to reach a native conformation after multiple cycles of interactions with folding modulators.

In the cytoplasm, proteolytic degradation is initiated by five ATP-dependent heat shock proteases (Lon, ClpYQ/HslUV, ClpAP, ClpXP and FtsH) and completed by peptidases that hydrolyze sequences 2–5 residues in length. These proteases consist of a remodeling component or domain that binds substrate proteins and couples ATP hydrolysis to unfolding and transfer of the polypeptide to an associated protease domain or proteolytic component.

Lon is a tetrameric serine protease of 87-kDa subunits containing three functional domains. Its N-terminus is involved in substrate recognition and binding whereas its central and C-terminus domain

are responsible for ATPase and proteolytic activities, respectively. In addition to being responsible for bulk protein degradation[58,59], Lon also exerts a regulatory function by degrading a class of proteins that are designed to be unstable (e.g., SulA).

ClpA (84 kDa), ClpX (46 kDa) and ClpY (49 kDa) assemble into hexameric ATPase rings that recognize and actively unfold proteins destined for degradation or remodeling by threading them through a central channel in a process fueled by ATP-driven conformational changes[60]. Both ClpA and ClpX associate with the same serine proteolytic component ClpP (a protein organized as two stacked heptamers of 23-kDa subunits) via 'ClpP loops' containing a [L/I/V]-G-[F/L] tripeptide absent in ClpB[61]. Because ClpP presents two identical faces for ClpA/ClpX interactions, complexes consisting of one or two ClpA or ClpX rings bound to one ClpP double-ring as well as heteromeric ClpA:ClpP:ClpX complexes can all form *in vitro*[62]. It has been suggested that ClpA:ClpP:ClpA particles may be best suited to carry out the remodeling of a subset of substrate proteins that are released without degradation[62]. ClpY specifically binds to either or both ends of ClpQ, a proteolytic component organized as two stacked hexameric rings of 19-kDa subunits, to form a structure that resembles the eukaryotic 26S proteaseome[63].

Together with Lon, ClpYQ is thought to be primarily responsible for the degradation of abnormal proteins[58]. Although also involved in bulk proteolysis, ClpXP, and to a lesser extent ClpAP, specifically degrade prematurely terminated proteins that have been modified by attachment of an SsrA tag (AANDENYALAA) at their C terminus[64]. Shuttling of tagged substrates to ClpX requires binding of the adaptor protein SspB to the SsrA tag and its subsequent interaction with the N terminus of ClpX[65]. ClpAP does not appear to require such an usher protein. However, an adaptor protein termed ClpS that associate with the N terminus of ClpA redirects ClpAP protease activity from soluble proteins to aggregated species[66].

FtsH (HflB), the only ATP-dependent cytoplasmic protease associated with the inner membrane, is organized as an hexamer of 71-kDa subunits that associates with dimers or hexamers of the HflK-HflC inner membrane proteins to form 1-MDa complexes[67]. FtsH relies on its cytoplasmic metalloprotease active site to degrade both membrane-embedded and soluble cytosolic proteins including the heat shock sigma factor σ^{32} and SsrA-tagged proteins. The role of HflKC in FtsH function is poorly understood.

Accumulation of misfolded proteins also occurs in the cell envelope owing to temperature increase, oxidative stress or improper formation of disulfide bonds. The primary housekeeping periplasmic protease is DegP a hexamer formed by staggered association of trimeric rings[68]. The proteolytic sites of DegP are located within an inner cavity bounded by mobile side walls formed by PDZ domains. PDZ regions control access to the protease chamber and are likely involved in substrate binding[69]. At low temperatures, DegP switches function from protease to chaperone[70] but the physiological relevance of this activity remains unclear. A second generic periplasmic protease, Prc (Tsp), cleaves proteins with nonpolar C-termini, membrane proteins and unfolded polypeptides with broad primary sequence specificity[71]. Prc contains a single PDZ domain that has been implicated in substrate binding[72].

Additional cell envelope proteases include DegS, DegQ, Protease III and OmpT. DegS is an inner membrane homotrimer that senses protein misfolding by binding Y-X-F tripeptides (where X is any amino acid) exposed at the C-terminus of immature outer membrane porins. Motion of the PDZ domains results in the cleavage of RseA, a transmembrane protein that sequesters the extracytoplasmic factor σ^{24}, thereby allowing high-level production of proteins involved in combating protein misfolding in the cell envelope[73]. The more poorly characterized serine protease DegQ hydrolyzes denatured substrates at discrete V/I-X locations[74], whereas protease III (the *ptr* gene product) is involved in the degradation of both protein fragments and larger abnormal proteins[75]. OmpT is an outer membrane protein specific for paired basic residues, which is organized in a vase-like structure with a serine active site that faces the growth milieu[76]. Because OmpT readily adsorbs to inclusion bodies during cell lysis and remains active under highly denaturing conditions[77], *ompT* null strains should always be used if the refolding route is chosen.

Cytoplasmic folding pathways engineering

Many eukaryotic proteins of therapeutic or commercial interest possess complex tertiary and quaternary structures and often require the formation of multiple disulfide bonds and other post-translational modifications to reach a native, biologically active conformation. Producing these proteins in *E. coli* can be challenging because the cellular environment, folding machinery and conformational quality control checkpoints of prokaryotes are quite different from those of eukaryotes. Not surprisingly, inclusion body formation and proteolytic degradation are commonly observed upon heterologous protein overexpression in *E. coli*.

A traditional approach to alleviate these problems involves reducing the synthesis rate of the target gene product to promote proper folding. This can be achieved by using weaker promoters or by decreasing the concentration of gratuitous inducer. For promoters based on *lac*-derived control elements (e.g., the *tac* or *trc* promoters), isopropyl D-thiogalactopyranoside (IPTG) concentrations below 100 μM are suitable for partial induction. However, because the P_{BAD} promoter operates in an 'all or none' fashion in wild-type cells, graded induction by subsaturating arabinose concentrations is only possible in strains that have been engineered to constitutively transport arabinose[78,79].

An alternative strategy is to decrease the temperature at which the recombinant protein accumulates. The use of low temperatures has the combined advantages of slowing down transcription and translation rates and of reducing the strength of hydrophobic interactions that contribute to protein misfolding. The drawback of this approach (as is the case with low inducer concentrations) is a reduction in productivity. Because traditional promoter systems exhibit reduced efficiency below 15°C, cold-inducible promoters, such as that of the major *E. coli* cold-shock gene *cspA*, are best suited for driving transcription at very low temperatures[80]. *cspA*-driven transcription is also useful for the expression of proteolytically sensitive and membrane-associated gene products[81] and companion strains that relieve promoter repression after prolonged incubation at low temperature are available[82]. Second generation *cspA*-based expression vectors have recently been described[83].

One of the most extensively used approaches to improve the yields of soluble proteins in the *E. coli* cytoplasm involves coexpression of molecular chaperones implicated in *de novo* protein folding. The beneficial effects of an increase in the intracellular concentration of TF, DnaK-DnaJ (with or without GrpE) and GroEL-GroES is well documented and a number of plasmids compatible with the routinely used ColE1-derived expression vectors are available[84]. DnaK-DnaJ or TF overexpression is suitable to increase the solubility of proteins requiring the assistance of chaperones in the early stages of their folding pathway[85]. For folding intermediates that rapidly transit through the TF/DnaK or require help at later folding stages, GroEL-GroES coexpression may be most beneficial. Technically, the GroEL-GroES encapsulation mechanism should limit the usefulness of this system to proteins smaller than ≈60 kDa. Nevertheless, larger proteins may also

benefit from GroEL-GroES coexpression[86], presumably because GroES-independent stabilization of partially folded domains by GroEL facilitates correct folding of the remainder of the chain[87]. If aggregation-prone intermediates are formed at both early and late stages of the folding pathway, coordinated expression of DnaK-DnaJ (or TF) and GroEL-GroES may be required to maximize recovery of the target protein in a soluble form[85,88].

Nevertheless, there are many—and often unpublished—studies in which coexpression of folding modulators fails to improve recombinant protein solubility. The underlying mechanisms are unclear but may be related to the need for timely interactions with specific folding modulators or to the substrate folding pathway itself. Indeed, it was recently reported that binding of TF and DnaK to nascent firefly luciferase chains redirects the folding of this protein from an efficient eukaryotic cotranslational mode to a slower post-translational pathway that is accompanied by aggregation[89]. Consequently, certain multidomain eukaryotic proteins that have evolved to take advantage of cotranslational folding may not benefit from chaperone overproduction. It should finally be noted that chaperone overexpression may also reduce the overall yield of recombinant proteins[90] (defined as the sum of soluble and insoluble fractions), possibly by transiently stabilizing off-pathway intermediates that are subsequently degraded by host proteases. In such cases, an increase in the yields of soluble and bioactive product may be achieved in strains bearing mutations in chaperone systems[91].

Stable disulfide bonds do not form in the cytoplasm of *E. coli* owing to their rapid reduction by the combined action of thioredoxins and glutaredoxins[92]. Enzymes from both pathways share a thioredoxin fold and a C-X-X-C active site. In their reduced form, the thioredoxins TrxA and TrxC attack disulfides in substrate proteins and leave them reduced while becoming oxidized in the process. Thioredoxin reductase (TrxB) recycles oxidized TrxA/C by reducing active site disulfides in an NADPH-dependent manner. GrxA, GrxB and GrxC perform a similar disulfide bond–reductase function but belong to the glutaredoxin pathway. They are kept in a reduced state through the action of tripeptide glutathione (the product of the *gshA* and *gshB* genes), which is in turn reduced by glutathione reductase (Gor).

Identification of the members of thioredoxin and glutaredoxin pathways and subsequent elucidation of their roles in disulfide bond reduction has made it possible to manipulate the *E. coli* cytoplasm to rationally promote disulfide bond formation in heterologous proteins. Production of oxidized proteins in the cytoplasm was first demonstrated in *trxB* mutants[93] and later shown to be due to a reversal of function of TrxA and TrxC from reductases to oxidases, owing to their accumulation in a disulfide-bonded form in the absence of TrxB[94]. Later work showed that the cytoplasm could be made even more reducing by incubating *trxB* cells at low temperatures[90,95] or by combining *trxB* and *gshA* or *gor* null mutations[96]. Although aerobic growth of the double mutants was impeded[96], suppressor strains exhibiting good growth characteristics were isolated and shown to be suitable for enhancing disulfide bond formation in heterologous proteins[97,98]. The yields of properly disulfide-bonded proteins in *trxB gor* suppressor cells can be further increased by coexpressing folding modulators including TF, GroEL-GroES and variants of Skp and DsbC that remain cytoplasmic due to the removal of their signal sequences[97–99]. In the last case, the chaperone activity of DsbC, rather than its disulfide isomerase activity, is determining in enhancing folding[98].

Despite the usefulness of the strategies described in the above paragraphs, recombinant gene products are commonly sensitive to proteolysis. In such cases, strains lacking ATP-dependent cytoplasmic proteases may be useful expression hosts, particularly if a single protease is responsible for degradation and if production is carried out at the laboratory scale. However, this approach is not without drawbacks. For example, thermosensitive *ftsH* mutants exhibit poor growth characteristics and *E. coli* K-12 mutants lacking Lon are filamentous and unsuitable for high-density fermentations. The fact that the *E. coli* B strain BL21 retains good growth characteristics while lacking both *lon* and *ompT* explains the popularity of this host. It is, however, not known if the cells compensate for these deficiencies by upregulating the concentration of other proteases.

Export pathways engineering

Recombinant proteins can be targeted to the periplasmic space in a Sec-dependent fashion by fusing naturally occurring signal sequences (e.g., those of PelB, OmpA, MalE or PhoA) to their N terminus. Periplasmic expression has a number of advantages over cytoplasmic production. First, an authentic N terminus can be obtained after removal of the signal sequence by leader peptidases. Second, the periplasm is conducive to disulfide bond formation because of the presence of the Dsb machinery. Third, there are fewer proteases in the periplasm compared to the cytoplasm and many have specific substrates. Finally, because the periplasm contains fewer proteins and because its content can be selectively released by osmotic shock or other strategies[100,101] purification of the target protein is facilitated.

One of the difficulties associated with heterologous protein secretion is inefficient export, which manifests itself by the degradation or aggregation of preproteins in the cytoplasm, and, in the cases of highly hydrophobic or integral membrane proteins, by membrane jamming which is associated with toxicity and eventual cell death. The use of low temperatures[81] or the cooverexpression of chaperones involved in secretion (e.g., SecB, DnaK-DnaJ and GroEL-GroES) may alleviate these problems. However, the benefits of the latter approach are highly dependent on the signal-sequence/mature-protein combination[102].

E. coli mutants originally selected for their ability to support the export of proteins with defective signal sequences offer an alternative route to promote secretion. One such allele, *prlA4*, encodes a defective version of SecY that leads to enhanced translocation rates, increased affinity of SecA for the SecYEG translocon, reduced reliance of Sec transport on the proton motive force and export of preproteins with folded domains[103,104]. The observation that inactivation of TF (encoded by *tig*) accelerates protein export and reduces the dependency of preproteins on secretory factors such as SecB[105,106] suggests that Δ*tig* strains will also be useful to enhance heterologous protein secretion. Finally, an artificial increase in signal sequence hydrophobicity may redirect translocation from the Sec- to the SRP-dependent pathway and concomitantly eliminate toxicity effects associated with membrane jamming by tightly coupling secretion with translation[105].

Owing to its relatively recent characterization, the potential of the bacterial Tat-dependent export pathway for heterologous protein secretion has not yet been fully explored. However, successful translocation of several heterologous proteins (including single chain and Fab antibody fragments[107]) suggest that Tat-dependent secretion will be a valuable tool for the secretion of heterologous proteins that assume a folded or a partially folded form before reaching the Sec machinery. Already, it has been shown that coexpression of the phage shock protein PspA improves the secretory capacity of the Tat system[108].

Engineering of periplasmic folding pathways

As they emerge in an unfolded or partially folded form on the periplasmic side of the inner membrane, heterologous proteins also confront the task of reaching a native conformation. As previously mentioned, the two cell envelope chaperones exhibiting the widest

substrate specificity are Skp and the PPIase FkpA (**Table 2**). Several studies have shown that coexpression of these folding helpers enhances heterologous protein folding and reduces degradation and periplasmic inclusion body formation[49,50,109–111]. Similarly, coover-production of DsbA or DsbC alone or in combination with DsbB and DsbD can facilitate the folding of proteins containing complex patterns of disulfide bonds[112–116]. It should finally be noted that, because the outer membrane is permeable to small solutes (<600 Da), an alternative way to reduce aggregation in the periplasm is to supplement the growth medium with small nonmetabolizable sugars, such as sucrose and rafinose[117,118]. By equilibrating within the periplasm, these sugars directly affect folding pathways, presumably by increasing protein chemical potentials via preferential exclusion effects[117].

The increased understanding of the specificity and mode of action of cell-envelope proteases suggests that a few simple precautions may go a long way in alleviating the problem of periplasmic degradation. First, because DegP is the major housekeeping protease and recognizes commonly occurring paired hydrophobic residues[69], *degP* null hosts should be routinely used. Second, proteins containing nonpolar C-terminal sequence should be expressed in *prc* (or perhaps *degS*) mutants, or as N-terminal fusions to carrier proteins. Finally, if proteolysis remains a problem, strains containing multiple mutations in the *degP*, *ompT*, *ptr* and *prc* genes may prove valuable. Nevertheless, the tradeoff associated with their reduced growth rates should be carefully considered[75,119].

The road ahead

The growing understanding of the principles that govern protein folding and misfolding in *E. coli*, the availability of sophisticated tools for chromosome engineering and rapid progress in PCR and directed evolution methodologies, have opened many new avenues of research. Already, it is obvious that one is not limited to traditional systems because the chaperone activity either of periplasmic folding helpers (e.g., Skp and DsbC[98,99]) or of oxidoreductases (e.g., TrxA[120,121]) can be co-opted to improve the folding of cytoplasmic recombinant proteins. Progress in the study of the structure-function relationship of folding modulators will undoubtedly allow rational improvements and/or coordination of chaperone and catalytic activities to promote more efficient heterologous protein folding.

The 'irrational' route also holds enormous potential. In a landmark study, Weissman and coworkers[122] showed that successive rounds of *in vivo* screening and DNA shuffling can be used to evolve GroEL variants exhibiting greatly enhanced ability to fold green fluorescent protein. Similar strategies could be applied to other chaperone-substrate pairs to isolate designer folding modulators dedicated to the efficient folding of specific, high-value recombinant proteins. That ClpB possesses disaggregation activity raises the possibility that engineered or evolved variants could solubilize cytoplasmic inclusion bodies and release component proteins in a conformation committed to folding, thereby enhancing the overall yields of native species during fermentation.

Finally, it has recently become apparent that *E. coli* can be manipulated to achieve post-translational modifications that have long been considered to be beyond its reach. For instance, engineering of the *Campylobacter jejuni* glycosylation pathway into *E. coli* has allowed (non-eukaryotic) N-linked glycosylation of the model protein AcrA[123]. More recently, Schultz and coworkers[124] evolved an orthogonal synthetase-tRNA pair suitable for the insertion of an exogenously added glycosylated amino acid (N-acetylglucosamine-serine) in response to amber codons, and obtained good yields of recombinant myoglobin containing N-acetylglucosamine-serine. Because N-acetylglucosamine serves as a substrate for the synthesis of more complex carbohydrates, further strain engineering to coexpress glycosyltransferases may provide an economically viable route for the production of therapeutic glycoproteins. Clearly, progress on the above issues and unexpected new discoveries all but guarantee the future of *E. coli* as an expression host.

ACKNOWLEDGMENTS

This work was supported by National Science Fund award BES-0097430.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/naturebiotechnology/

1. Baneyx, F. Protein expression technologies: current status and future trends (Horizon Biosciences, Norfolk, 2004).
2. Lorimer, G.H. A quantitative assessment of the role of the chaperonin proteins in protein folding *in vivo*. *FASEB J.* **10**, 5–9 (1996).
3. Ellis, R.J. & Minton, A.P. Join the crowd. *Nature* **425**, 27–28 (2003).
4. Valax, P. & Georgiou, G. Molecular characterization of β-lactamase inclusion bodies produced in *Escherichia coli*. 1. Composition. *Biotechnol. Prog.* **9**, 539–547 (1993).
5. Allen, S.P., Polazzi, J.O., Gierse, J.K. & Easton, A.M. Two novel heat shock genes encoding proteins produced in response to heterologous protein expression in *Escherichia coli*. *J. Bacteriol.* **174**, 6938–6947 (1992).
6. Carrió, M.M., Corchero, J.L. & Villaverde, A. Dynamics of *in vivo* protein aggregation: building inclusion bodies in recombinant bacteria. *FEMS Microbiol. Lett.* **169**, 9–15 (1998).
7. Bowden, G.A., Paredes, A.M. & Georgiou, G. Structure and morphology of inclusion bodies in *Escherichia coli*. *Bio/Technology* **9**, 725–730 (1991).
8. Taylor, G., Hoare, M., Gray, D.R. & Martson, F.A.O. Size and density of inclusion bodies. *Bio/Technology* **4**, 553–557 (1986).
9. Oberg, K., Chrunyk, B.A., Wetzel, R. & Fink, A.L. Nativelike secondary structure in interleukin 1-β inclusion bodies by attenuated total reflectance FTIR. *Biochemistry* **33**, 2628–2634 (1994).
10. Tsumoto, K., Ejima, D., Kumagai, I. & Arakawa, T. Practical considerations in refolding proteins from inclusion bodies. *Protein Expr. Purif.* **28**, 1–8 (2003).
11. Hoffmann, F. & Rinas, U. 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**, 1000–1007 (2000).
12. Hartl, F.U. & Hayer-Hartl, M. Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* **295**, 1852–1858 (2002).
13. Mujacic, M. & Baneyx, F. in Protein expression technologies: current status and future trends. (ed. Baneyx, F.) 85–148 (Horizon Biosciences, Norfolk, UK, 2004).
14. Patzelt, H. *et al.* Binding specificity of *Escherichia coli* trigger factor. *Proc. Natl. Acad. Sci. USA* **98**, 14244–14249 (2001).
15. Deuerling, E. *et al.* Trigger factor and DnaK possess overlapping substrate pools and binding specificities. *Mol. Microbiol.* **47**, 1317–1328 (2003).
16. Ewalt, K.L., Hendrick, J.P., Houry, W.A. & Hartl, F.U. *In vivo* observation of polypeptide flux through the bacterial chaperonin system. *Cell* **90**, 491–500 (1997).
17. Narberhaus, F. Alpha-crystallin-type heat shock proteins: socializing minichaperones in the context of a multichaperone network. *Microbiol. Mol. Biol. Rev.* **66**, 64–93 (2002).
18. Shearstone, J.R. & Baneyx, F. Biochemical characterization of the small heat shock protein IbpB from *Escherichia coli*. *J. Biol. Chem.* **274**, 9937–9945 (1999).
19. Veinger, L., Diamant, S., Buchner, J. & Goloubinoff, P. The small heat-shock protein IbpB from *Escherichia coli* stabilizes stress-denatured proteins for subsequent refolding by a multichaperone network. *J. Biol. Chem.* **273**, 11032–11037 (1998).
20. Graf, P.C. & Jakob, U. Redox-regulated molecular chaperones. *Cell. Mol. Life Sci.* **59**, 1624–1631 (2002).
21. Hoffmann, J.H., Linke, K., Graf, P.C., Lilie, H. & Jakob, U. Identification of a redox-regulated chaperone network. *EMBO J.* **23**, 160–168 (2004).
22. Malki, A., Kern, R., Abdallah, J. & Richarme, G. Characterization of the *Escherichia coli* YedU protein as a molecular chaperone. *Biochem. Biophys. Res. Commun.* **301**, 430–436 (2003).
23. Sastry, M.S.R., Korotkov, K., Brodsky, Y. & Baneyx, F. Hsp31, the *Escherichia coli yedU* gene product, is a molecular chaperone whose activity is inhibited by ATP at high temperatures. *J. Biol. Chem.* **277**, 46026–46034 (2002).
24. Mujacic, M., Bader, M.W. & Baneyx, F. *Escherichia coli* Hsp31 functions as a holding chaperone that cooperates with the DnaK-DnaJ-GrpE system in the management of protein misfolding under severe thermal stress conditions. *Mol. Microbiol.* **51**, 849–859 (2004).
25. Quigley, P.M., Korotkov, K., Baneyx, F. & Hol, W.G.J. The 1.6-Å crystal structure of the class of chaperones represented by *Escherichia coli* Hsp31 reveals a putative catalytic triad. *Proc. Natl. Acad. Sci. USA* **100**, 3137–3142 (2003).
26. Quigley, P.M., Korotkov, K., Baneyx, F. & Hol, W.G.J. A new native EcHsp31 crystal structure suggests key role of structural flexibility for chaperone function. *Protein Sci.* **13**, 269–277 (2004).
27. Sastry, M.S.R., Quigley, P.M., Hol, W.G.J. & Baneyx, F. The linker-loop region of *E. coli* chaperone Hsp31 functions as a thermal gate that modulates high affinity sub-

strate binding at elevated temperatures. *Proc. Natl. Acad. Sci. USA* **101**, 8587–8592 (2004).
28. Weibezahn, J., Bukau, B. & Mogk, A. Unscrambling an egg: protein disaggregation by AAA+ proteins. *Microb. Cell Fact.* **3**, 1 (2004).
29. Lee, S. *et al.* The structure of ClpB. A molecular chaperone that rescues proteins from an aggregated state. *Cell* **115**, 229–240 (2003).
30. Schlieker, C. *et al.* Substrate recognition by the AAA+ chaperone ClpB. *Nat. Struct. Mol. Biol.* **11**, 607–615 (2004).
31. Mogk, A., Deuerling, E., Vorderwulbecke, S., Vierling, E. & Bukau, B. Small heat shock proteins, ClpB and the DnaK system form a functional triade in reversing protein aggregation. *Mol. Microbiol.* **50**, 585–595 (2003).
32. Goloubinoff, P., Mogk, A., Ben Zvi, A.P., Tomoyasu, T. & Bukau, B. Sequential mechanism of solubilization and refolding of stable protein aggregates by a bichaperone network. *Proc. Natl. Acad. Sci. USA* **96**, 13732–13737 (1999).
33. Zolkiewski, M. ClpB cooperates with DnaK, DnaJ and GrpE in suppressing protein aggregation. *J. Biol. Chem.* **274**, 28083–28086 (1999).
34. Diamant, S., Ben-Zvi, A.P., Bukau, B. & Goloubinoff, P. Size-dependent disaggregation of stable protein aggregates by the DnaK chaperone machinery. *J. Biol. Chem.* **275**, 21107–21113 (2000).
35. Zietkiewicz, S., Krzewska, J. & Liberek, K. Successive and synergistic action of the Hsp70 and Hsp100 chaperones in protein disaggregation. *J. Biol. Chem.* (in the press) (2004).
36. de Keyzer, J., van der Does, C. & Driessen, A.J. The bacterial translocase: a dynamic protein channel complex. *Cell. Mol. Life Sci.* **60**, 2034–2052 (2003).
37. Xu, Z., Knafels, J.D. & Yoshino, K. Crystal structure of the bacterial protein export chaperone SecB. *Nat. Struct. Biol.* **7**, 1172–1177 (2000).
38. Altman, E., Kumamoto, C.A. & Emr, S.D. Heat-shock proteins can substitute for SecB function during protein export in *Escherichia coli*. *EMBO J.* **10**, 239–245 (1991).
39. Harms, N., Luirink, J. & Oudega, B. in *Molecular chaperones in the cell* (ed. Lund, P.) 35–60 (Oxford University Press, New York, 2001).
40. Driessen, A.J.M., Manting, E.H. & van der Does, C. The structural basis of protein targeting and translocation in bacteria. *Nat. Struct. Biol.* **8**, 492–498 (2001).
41. Buskiewicz, I. *et al.* Trigger factor binds to ribosome-signal recognition particle (SRP) complexes and is excluded by binding of the SRP receptor. *Proc. Natl. Acad. Sci. USA* **101**, 7902–7906 (2004).
42. DeLisa, M.P., Samuelson, P., Palmer, T. & Georgiou, G. Genetic analysis of the twin arginine translocator secretion pathway in bacteria. *J. Biol. Chem.* **277**, 29825–29831 (2002).
43. Brüser, T. & Sanders, C. An alternative model of the twin arginine translocation system. *Mircobiol. Res.* **158**, 7–17 (2003).
44. Palmer, T. & Berks, B.C. Moving folded protein across the bacterial cell membrane. *Microbiology* **149**, 547–556 (2003).
45. Sargent, F. *et al.* Overlapping functions of components of a bacterial Sec-independent protein export pathway. *EMBO J.* **17**, 3640–3650 (1998).
46. Chen, R. & Henning, U. A periplasmic protein (Skp) of *Escherichia coli* selectively binds a class of outer membrane proteins. *Mol. Microbiol.* **19**, 1287–1294 (1996).
47. Schafer, U., Beck, K. & Muller, M. Skp, a molecular chaperone of gram-negative bacteria is required for the formation of soluble periplasmic intermediates of outer membrane proteins. *J. Biol. Chem.* **274**, 24567–24574 (1999).
48. Walton, T.A. & Sousa, M.C. Crystal structure of Skp, a prefoldin-like chaperone that protects soluble and membrane proteins from aggregation. *Mol. Cell* **15**, 367–374 (2004).
49. Missiakas, D., Betton, J.M. & Raina, S. New components of protein folding in extracytoplasmic compartments of *Escherichia coli* SurA, FkpA and Skp/OmpH. *Mol. Microbiol.* **21**, 871–884 (1996).
50. Arie, J.P., Sassoon, N. & Betton, J.M. Chaperone function of FkpA, a heat shock prolyl isomerase, in the periplasm of *Escherichia coli*. *Mol. Microbiol.* **39**, 199–210 (2001).
51. Saul, F.A. *et al.* Structural and functional studies of FkpA from *Escherichia coli*, a *cis/trans* peptidyl-prolyl isomerase with chaperone activity. *J. Mol. Biol.* **335**, 595–608 (2004).
52. Behrens, S., Maier, R., de Cock, H., Schmid, F.X. & Gross, C.A. The SurA periplasmic PPIase lacking its parvulin domains functions *in vivo* and has chaperone activity. *EMBO J.* **20**, 285–294 (2001).
53. Bitto, E. & McKay, D.B. Crystallographic structure of SurA, a molecular chaperone that facilitates folding of outer membrane porins. *Structure* **10**, 1489–1498 (2002).
54. Bitto, E. & McKay, D.B. The periplasmic molecular chaperone protein SurA binds a peptide motif that is characteristic of integral outer membrane proteins. *J. Biol. Chem.* **278**, 49316–49322 (2003).
55. Hiniker, A. & Bardwell, J.C.A. Disulfide bond isomerization in prokaryotes. *Biochemistry* **42**, 1179–1185 (2003).
56. Kadokura, H., Katzen, F. & Beckwith, J. Protein disulfide bond formation in prokaryotes. *Annu. Rev. Biochem.* **72**, 111–135 (2003).
57. McCarthy, A.A. *et al.* Crystal structure of the protein disulfide bond isomerase, DsbC, from *Escherichia coli*. *Nat. Struct. Biol.* **7**, 196–199 (2000).
58. Missiakas, D., Schwager, F., Betton, J.M., Georgopoulos, C. & Raina, S. Identification and characterization of HslV HslU (ClpQ ClpY) proteins involved in overall proteolysis of misfolded proteins in *Escherichia coli*. *EMBO J.* **15**, 6899–6909 (1996).
59. Tomoyasu, T., Mogk, A., Langen, H., Goloubinoff, P. & Bukau, B. Genetic dissection of the roles of chaperones and proteases in protein folding and degradation in the *Escherichia coli* cytosol. *Mol. Microbiol.* **40**, 397–413 (2001).
60. Matouschek, A. Protein unfolding—an important process *in vivo*? *Curr. Opin. Struct. Biol.* **13**, 98–109 (2003).
61. Kim, Y.-I. *et al.* Molecular determinants of complex formation between Clp/Hsp100 ATPases and the ClpP peptidase. *Nat. Struct. Biol.* **8**, 230–233 (2001).
62. Ortega, J., Lee, H.S., Maurizi, M.R. & Steven, A.C. ClpA and ClpX ATPases bind simultaneously to opposite ends of ClpP peptidase to form active hybrid complexes. *J. Struct. Biol.* **146**, 217–226 (2004).
63. Bochtler, M. *et al.* The structure of HslU and the ATP-dependent protease HslU-HslV. *Nature* **403**, 800–805 (2000).
64. Gottesman, S., Roche, E., Zhou, Y.N. & Sauer, R.T. The ClpXP and ClpAP proteases degrade proteins with carboxyl-terminal peptide tails added by the SsrA-tagging system. *Genes Dev.* **12**, 1338–1347 (1998).
65. Dougan, D.A., Weber-Ban, E. & Bukau, B. Targeted delivery of an ssrA-tagged substrate by the protein SspB to it cognate AAA+ protein ClpX. *Mol. Cell* **12**, 373–380 (2003).
66. Dougan, D.A., Reid, B.G., Horwich, A.L. & Bukau, B. ClpS, a substrate modulator of the ClpAP machine. *Mol. Cell* **9**, 673–683 (2002).
67. Saikawa, N. & Akiyama, Y. & Ito, K. FtsH exists as an exceptionally large complex containing HflKC in the plasma membrane of *Escherichia coli*. *J. Struct. Biol.* **146**, 123–129 (2004).
68. Krojer, T., Garrido-Franco, M., Huber, R., Ehrmann, M. & Clausen, T. Crystal structure of DegP (HtrA) reveals a new protease-chaperone machine. *Nature* **416**, 455–459 (2002).
69. Jones, C.H. *et al. Escherichia coli* DegP protease cleaves between paired hydrophobic residues in a natural substrate: the PapA pilin. *J. Bacteriol.* **184**, 5762–5771 (2002).
70. Spiess, C., Beil, A. & Ehrmann, M. A temperature-dependent switch from chaperone to protease in a widely conserved heat shock protein. *Cell* **97**, 339–347 (1999).
71. Keiler, K.C. *et al.* C-terminal specific protein degradation: activity and substrate specificity of the Tsp protease. *Protein Sci.* **4**, 1507–1515 (1995).
72. Spiers, A. *et al.* PDZ domains facilitate binding of high temperature requirement protease A (HtrA) and tail-specific protease (Tsp) to heterologous substrates through recognition of the small stable RNA A (ssrA)-encoded peptide. *J. Biol. Chem.* **277**, 39443–39449 (2002).
73. Walsh, N.P., Alba, B.M., Bose, B., Gross, C.A. & Sauer, R.T. OMP peptide signals initiate the envelope-stress response by activating DegS protease via relief of inhibition mediated by Its PDZ domain. *Cell* **113**, 61–71 (2003).
74. Waller, P.R. & Sauer, R.T. Characterization of *degQ* and *degS*, *Escherichia coli* genes encoding homologs of the DegP protease. *J. Bacteriol.* **178**, 1146–1153 (1996).
75. Baneyx, F. & Georgiou, G. Construction and characterization of *Escherichia coli* strains deficient in multiple secreted proteases: protease III degrades high molecular weight substrates *in vivo*. *J. Bacteriol.* **173**, 2696–2703 (1991).
76. Vandeputte-Rutten, L. *et al.* Crystal structure of the outer membrane protease OmpT from *Escherichia coli* suggests a novel catalytic site. *EMBO J.* **20**, 5033–5039 (2001).
77. White, C.B., Chen, Q., Kenyon, G.L. & Babbitt, P.C. A novel activity of OmpT. Proteolysis under extreme denaturing conditions. *J. Biol. Chem.* **270**, 12990–12994 (1995).
78. Khlebnikov, A., Datsenko, K.A., Skaug, T., Wanner, B.L. & Keasling, J.D. Homogeneous expression of the P_{BAD} promoter in *Escherichia coli* by constitutive expression of the low-affinity high-capacity AraE transporter. *Microbiology* **147**, 3241–3247 (2001).
79. Morgan-Kiss, R.M., Wadler, C. & Cronan, J.E.J. Long-term and homogeneous regulation of the *Escherichia coli araBAD* promoter by use of a lactose transporter of relaxed specificity. *Proc. Natl. Acad. Sci. USA* **99**, 7373–7377 (2002).
80. Vasina, J.A. & Baneyx, F. Expression of aggregation-prone proteins at low temperatures: a comparative study of the *E. coli cspA* and *tac* promoter systems. *Protein Expr. Purif.* **9**, 211–218 (1997).
81. Mujacic, M., Cooper, K.W. & Baneyx, F. Cold-inducible cloning vectors for low-temperature protein expression in *Escherichia coli*: application to the production of a toxic and proteolytically sensitive fusion protein. *Gene* **238**, 325–332 (1999).
82. Vasina, J.A., Peterson, M.S. & Baneyx, F. Scale-up and optimization of the low-temperature inducible *cspA* promoter system. *Biotechnol. Prog.* **14**, 714–721 (1998).
83. Qing, G. *et al.* Cold-shock induced high-yield protein production in *Escherichia coli*. *Nat. Biotechnol.* **22**, 877–882 (2004).
84. Baneyx, F. & Palumbo, J.L. Improving heterologous protein folding via molecular chaperone and foldase co-expression. *Methods Mol. Biol.* **205**, 171–197 (2003).
85. Nishihara, K., Kanemori, M., Yanagi, H. & Yura, T. Overexpression of trigger factor prevents aggregation of recombinant proteins in *Escherichia coli*. *Appl. Environ. Microbiol.* **66**, 884–889 (2000).
86. Roman, L.J. *et al.* High-level expression of functional rat neuronal nitric oxide synthase in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **92**, 8428–8432 (1995).
87. Ayling, A. & Baneyx, F. Influence of the GroE molecular chaperone machine on the *in vitro* folding of *Escherichia coli* β-galactosidase. *Protein Sci.* **5**, 478–487 (1996).
88. Nishihara, K., Kanemori, M., Kitagawa, M., Yanaga, H. & Yura, T. Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of Japanese cedar pollen, Cryj2, in *Escherichia coli*. *Appl. Environ. Microbiol.* **64**, 1694–1699 (1998).
89. Agashe, V.R. *et al.* Function of trigger factor and DnaK in multidomain protein folding: increase in yield at the expense of folding speed. *Cell* **117**, 199–209 (2004).
90. Schneider, E.L., Thomas, J.G., Bassuk, J.A., Sage, E.H. & Baneyx, F. Manipulating the aggregation and oxidation of human SPARC in the cytoplasm of *Escherichia coli*.

Nat. Biotechnol. **15**, 581–585 (1997).
91. Carrió, M.M. & Villaverde, A. Role of molecular chaperones in inclusion body formation. *FEBS Lett.* **537**, 215–221 (2003).
92. Ritz, D. & Beckwith, J. Roles of thiol-redox pathways in bacteria. *Annu. Rev. Microbiol.* **55**, 21–48 (2001).
93. Derman, A.I., Prinz, W.A., Belin, D. & Beckwith, J. Mutations that allow disulfide bond formation in the cytoplasm of *Escherichia coli*. *Science* **262**, 1744–1747 (1993).
94. Stewart, E.J., Åslund, F. & Beckwith, J. Disulfide bond formation in the *Escherichia coli* cytoplasm: an *in vivo* role reversal for the thioredoxins. *EMBO J.* **17**, 5543–5550 (1998).
95. Derman, A.I. & Beckwith, J. *Escherichia coli* alkaline phosphatase localized to the cytoplasm acquires enzymatic activity in cells whose growth has been suspended: a caution for gene fusion studies. *J. Bacteriol.* **177**, 3764–3770 (1995).
96. Prinz, W.A., Åslund, F., Holmgren, A. & Beckwith, J. The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the *Escherichia coli* cytoplasm. *J. Biol. Chem.* **272**, 15661–15667 (1997).
97. Bessette, P.H., Åslund, F., Beckwith, J. & Georgiou, G. Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm. *Proc. Natl. Acad. Sci. USA* **96**, 13703–13708 (1999).
98. Jurado, P., Ritz, D., Beckwith, J., de Lorenzo, V. & Fernandez, L.A. Production of functional single-chain Fv antibodies in the cytoplasm of *Escherichia coli*. *J. Mol. Biol.* **320**, 1–10 (2002).
99. Levy, R., Weiss, R., Chen, G., Iverson, B.L. & Georgiou, G. Production of correctly folded Fab antibody fragments in the cytoplasm of *Escherichia coli trxB gor* mutants via the coexpression of molecular chaperones. *Protein Expr. Purif.* **23**, 338–347 (2001).
100. Robbens, J., Raeymaekers, A., Steidler, L., Fiers, W. & Remaut, E. Production of soluble and active recombinant murine interleukin-2 in *Escherichia coli*: high level expression, Kil-induced release and purification. *Protein Expr. Purif.* **6**, 481–486 (1995).
101. Wan, E. & Baneyx, F. TolAIII co-overexpression facilitates the recovery of periplasmic recombinant proteins into the growth medium of *Escherichia coli*. *Protein Expr. Purif.* **14**, 13–22 (1998).
102. Bergès, H., Joseph-Liauzun, E. & Fayet, O. Combined effects of the signal sequence and the major chaperone proteins on the export of human cytokines in *Escherichia coli*. *Appl. Environ. Microbiol.* **62**, 55–60 (1996).
103. Nouwen, N., Kruijff, B. & Tommassen, J. *prlA* suppressors in *Escherichia coli* relieve the proton electrochemical gradient dependency of translocation of wild-type precursors. *Proc. Natl. Acad. Sci. USA* **93**, 5953–5957 (1996).
104. van der Wolk, J.P.W. *et al. PrlA4* prevents the rejection of signal sequence defective preproteins by stabilizing the SecA-SecY interaction during the initiation of translocation. *EMBO J.* **17**, 3631–3639 (1998).
105. Bowers, C.W., Lau, F. & Silhavy, T.J. Secretion of LamB-LacZ by the signal recognition particle pathway of *Escherichia coli*. *J. Bacteriol.* **185**, 5697–5705 (2003).
106. Lee, H.C. & Bernstein, H.D. Trigger factor retards protein export in *Escherichia coli*. *J. Biol. Chem.* **277**, 43527–43535 (2002).
107. DeLisa, M.P., Tullman, D. & Georgiou, G. Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway. *Proc. Natl. Acad. Sci. USA* **100**, 6115–6120 (2003).
108. DeLisa, M.P., Lee, P., Palmer, T. & Georgiou, G. Phage shock protein PspA of *Escherichia coli* relieves saturation of protein export via the Tat pathway. *J. Bacteriol.* **186**, 366–373 (2004).
109. Bothmann, H. & Plückthun, A. Selection for a periplasmic factor improving phage display and functional periplasmic expression. *Nat. Biotechnol.* **16**, 376–380 (1998).
110. Bothmann, H. & Plückthun, A. The periplasmic *Escherichia coli* peptidylprolyl *cis,trans*-isomerase FkpA.I. Increased functional expression of antibody fragments with and without *cis*-prolines. *J. Biol. Chem.* **275**, 17100–17105 (2000).
111. Hayhurst, A. & Harris, W.J. *Escherichia coli* Skp chaperone coexpression improves solubility and phage display of single-chain antibody fragments. *Protein Expr. Purif.* **15**, 336–343 (1999).
112. Jeong, K.J. & Lee, S.Y. Secretory production of human leptin in *Escherichia coli*. *Biotechnol. Bioeng.* **67**, 398–407 (2000).
113. Kurokawa, Y., Yanagi, H. & Yura, T. Overexpression of protein disulfide isomerase DsbC stabilizes multiple-disulfide-bonded recombinant protein produced and transported to the periplasm in *Escherichia coli*. *Appl. Environ. Microbiol.* **66**, 3960–3965 (2000).
114. Kurokawa, Y., Yanagi, H. & Yura, T. Overproduction of bacterial protein disulfide isomerase (DsbC) and its modulator (DsbD) markedly enhances periplasmic production of human nerve growth factor in *Escherichia coli*. *J. Biol. Chem.* **276**, 14393–14399 (2001).
115. Qiu, J., Swartz, J.R. & Georgiou, G. Expression of active human tissue-type plasminogen activator in *Escherichia coli*. *Appl. Environ. Microbiol.* **64**, 4891–4896 (1998).
116. Wulfing, C. & Rappuoli, R. Efficient production of heat-labile enterotoxin mutant proteins by overexpression of *dsbA* in a *degP*-deficient *Escherichia coli* strain. *Arch. Microbiol.* **167**, 280–283 (1997).
117. Bowden, G.A. & Georgiou, G. The effect of sugars on β-lactamase aggregation in *Escherichia coli*. *Biotechnol. Prog.* **4**, 97–101 (1988).
118. Sawyer, J.R., Schlom, J. & Kashmiri, S.V.S. The effects of induction conditions on production of a soluble anti-tumor SFv in *Escherichia coli*. *Prot. Eng.* **7**, 1401–1406 (1994).
119. Meerman, H.J. & Georgiou, G. Construction and characterization of a set of *E. coli* strains deficient in all known loci affecting the proteolytic stability of secreted recombinant proteins. *Bio/Technology* **12**, 1107–1110 (1994).
120. Kern, R., Malki, A., Holmgren, A. & Richarme, G. Chaperone properties of *Escherichia coli* thioredoxin and thioredoxin reductase. *Biochem. J.* **371**, 965–972 (2003).
121. Yasukawa, T. *et al.* Increase of solubility of foreign proteins in *Escherichia coli* by coproduction of the bacterial thioredoxin. *J. Biol. Chem.* **270**, 25328–25331 (1995).
122. Wang, J.D., Herman, C., Tipton, K.A., Gross, C.A. & Weissman, J.S. Directed evolution of substrate-optimized GroEL/S chaperonins. *Cell* **111**, 1027–1039 (2002).
123. Wacker, M. *et al.* N-linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science* **298**, 1790–1793 (2002).
124. Zhang, Z. *et al.* A new strategy for the synthesis of glycoproteins. *Science* **303**, 371–373 (2004).