nature structural & molecular biology

Genetic selection designed to stabilize proteins uncovers a chaperone called Spy

Shu Quan[1,2], Philipp Koldewey[1,2], Tim Tapley[1,2], Nadine Kirsch[1,2], Karen M Ruane[3], Jennifer Pfizenmaier[1,2], Rong Shi[3], Stephan Hofmann[1,2], Linda Foit[1,2], Guoping Ren[1,2], Ursula Jakob[2], Zhaohui Xu[4,5], Miroslaw Cygler[3,6] & James C A Bardwell[1,2]

To optimize the *in vivo* folding of proteins, we linked protein stability to antibiotic resistance, thereby forcing bacteria to effectively fold and stabilize proteins. When we challenged *Escherichia coli* to stabilize a very unstable periplasmic protein, it massively overproduced a periplasmic protein called Spy, which increases the steady-state levels of a set of unstable protein mutants up to 700-fold. *In vitro* studies demonstrate that the Spy protein is an effective ATP-independent chaperone that suppresses protein aggregation and aids protein refolding. Our strategy opens up new routes for chaperone discovery and the custom tailoring of the *in vivo* folding environment. Spy forms thin, apparently flexible cradle-shaped dimers. The structure of Spy is unlike that of any previously solved chaperone, making it the prototypical member of a new class of small chaperones that facilitate protein refolding in the absence of energy cofactors.

The folding of many proteins is assisted by molecular chaperones and other folding helpers in the cell[1]. Many chaperones act by inhibiting off-pathway events, such as aggregation, and serve a broad range of substrates in a stoichiometric manner. *In vitro* assays for chaperone activity are thus almost by necessity relatively insensitive, making these assays more useful in studying previously identified chaperones than in identifying new ones in crude cell lysates. Instead of being discovered directly for their capacity to assist *in vivo* protein folding processes, many chaperones were initially characterized because of their induction by stress conditions[2]. The lack of a sensitive and general *in vivo* assay for chaperone activity led us to wonder how complete the list of known chaperones is.

Previously, we developed a genetic system that directly links increased protein stability to increased antibiotic resistance[3] and thereby provides a selectable and quantitative *in vivo* measure of protein stability[3]. Our selection system makes use of a sandwich fusion between β-lactamase and unstable proteins, which effectively links the stability of the inserted proteins to the penicillin resistance of the host strain. We showed that β-lactamase tolerates the insertion of a well-folded protein in a surface loop (at position 197) and still retains enzymatic activity (**Fig. 1**). However, insertion of unstable proteins at this site negatively affects β-lactamase activity and decreases penicillin V resistance ($PenV^R$) *in vivo*[3]. This strategy allowed us to select for stabilized protein variants of a well-characterized protein, immunity protein 7 (Im7, also known as ImmE7)[3].

We now hypothesized that a similar strategy could even be used to select for host variants that alter the *in vivo* folding environment of individual proteins by enhancing protein folding or stability, either specifically or generally. Because our system functions in the periplasm, it may also provide an opportunity to explore protein folding in what has generally been considered to be a relatively chaperone-poor environment and, in doing so, uncover new chaperones. In selecting for host variants that stabilized a very unstable Im7 mutant, we discovered variants that massively overproduce a previously poorly characterized protein called Spy. We found Spy to be the founding member of a new class of chaperones that have a novel cradle-shaped structure and function as ATP-independent folding chaperones in the periplasm of *E. coli*.

RESULTS

Selection for optimized protein folding *in vivo*

We designed a dual selection system by inserting the same unstable test protein Im7-L53A I54A into the middle of β-lactamase, which is implicated in penicillin V resistance ($PenV^R$), and DsbA, which is implicated in cadmium resistance ($CdCl_2^R$)[4], as sandwich fusions (**Fig. 1**). Our overall experimental scheme is illustrated in **Figure 2**. The two resistance markers operate via very different mechanisms[4,5] and thus provide an independent measure of the stability of the inserted protein. In strains that coexpress both fusions, we reasoned that host mutations that simultaneously increase $PenV^R$ and $CdCl_2^R$ should positively affect the one thing that these two constructs apparently have in common, namely the stability of the inserted protein. Insertion of Im7 variants into DsbA after residue Thr99, a site that can tolerate protein insertions (**Fig. 1** and **Supplementary Fig. 1**),

[1]Howard Hughes Medical Institute, Chevy Chase, Maryland, USA. [2]Department of Molecular, Cellular and Developmental Biology, University of Michigan, Ann Arbor, Michigan, USA. [3]Department of Biochemistry, McGill University, Montreal, Quebec, Canada. [4]Department of Biological Chemistry, University of Michigan, Ann Arbor, Michigan, USA. [5]Life Sciences Institute, University of Michigan, Ann Arbor, Michigan, USA. [6]Biotechnology Research Institute, National Research Council, Montreal, Quebec, Canada. Correspondence should be addressed to J.C.A.B. (jbardwel@umich.edu).

Received 27 July 2010; accepted 15 December 2010; published online 13 February 2011; doi:10.1038/nsmb.2016

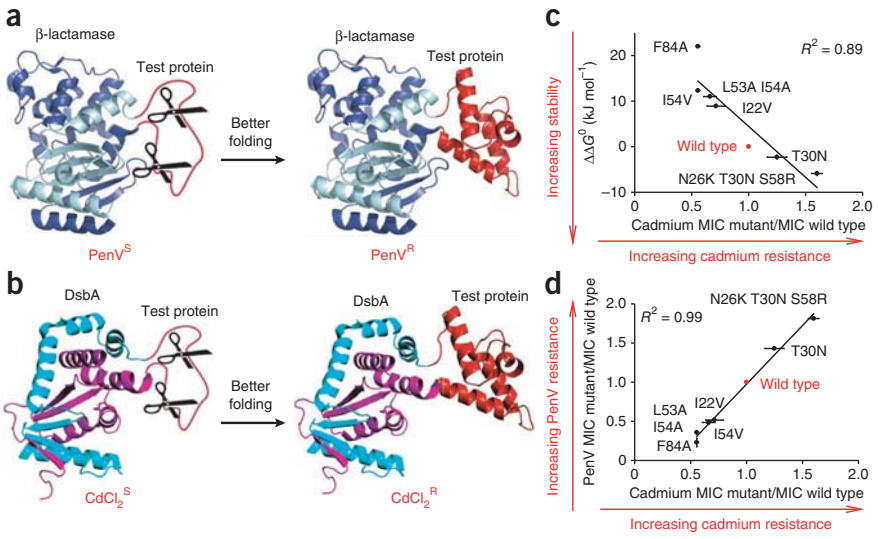

Figure 1 A dual fusion selection for enhancing *in vivo* protein stability. (**a**) Unstable test proteins inserted into β-lactamase are degraded by cellular proteases, producing penicillin-sensitive ($PenV^S$) strains. Improving the folding of the test proteins increases penicillin resistance ($PenV^R$)[3]. (**b**) Insertion of unstable test proteins into DsbA renders the strains sensitive to cadmium ($CdCl_2^S$). Improving the folding of the test proteins increases the strains' resistance toward cadmium ($CdCl_2^R$). (**c**) Thermodynamic stability of Im7 variants[3] (given as $\Delta\Delta G^0$ relative to wild-type Im7) correlates with the minimal inhibitory concentration (MIC) of cadmium for cells expressing the DsbA fusions to these variants. (**d**) Penicillin resistance of cells expressing the β-lactamase–Im7 fusions correlates with cadmium resistance when they also express the DsbA-Im7 fusion proteins with the same Im7 variants. Error bars indicate the s.d. of three independent measurements.

revealed an excellent correlation between $CdCl_2^R$ and Im7 variant stability (**Fig. 1c**). $CdCl_2^R$ also correlated well with $PenV^R$ when the same Im7 variants were inserted into β-lactamase (**Fig. 1d**). The highest resistance to both antimicrobials came from the most stable Im7 variants. These results suggested that host variants that stabilize Im7 should increase both $PenV^R$ and $CdCl_2^R$ and thus should be easily selected for on this basis.

We searched for host mutations that enable the proper folding of Im7-L53A I54A, a very unstable Im7 variant[6]. Using ethyl methanesulfonate (EMS), we randomly mutagenized strain SQ1306, which contains sandwich fusions between this unstable Im7 variant and both DsbA and β-lactamase. We selected for variants with enhanced $PenV^R$ and then screened those for enhanced $CdCl_2^R$.

Plate screens revealed that about 13% (35 of 263) of the strains that had gained $PenV^R$ had simultaneously acquired resistance to $CdCl_2$. Ten independently isolated mutant strains that were resistant to both PenV and $CdCl_2$ (EMS1–EMS10, **Supplementary Table 1a**) were selected for further analysis. Eight of these contained substantially elevated levels of the β-lactamase–Im7 fusion protein, as determined by quantitative western blots (**Supplementary Fig. 2**). Because EMS4 and EMS9 contained the highest amounts of the sandwich fusion protein and showed resistance to high concentrations of both PenV and cadmium, they were chosen for further analysis. To confirm these host variants had improved the folding and consequently increased the protein levels of the Im7 protein itself, we transformed EMS4 and EMS9 with plasmids encoding Im7 and three different destabilized Im7 mutants in the absence of the fusion. Both EMS4 and EMS9 strain backgrounds accumulated large amounts of Im7 proteins in the periplasm, in contrast to the parental $PenV^S$ and $CdCl_2^S$ strain SQ1306 (**Fig. 3a**). Quantification using an Agilent Bioanalyzer showed that Im7 proteins made up 7–10% of the periplasmic content in EMS4 and EMS9, a 34- to 92-fold increase over the wild-type strain (**Fig. 3a**, **Table 1** and **Supplementary Table 1b**). The Im7 proteins were extracted from the periplasm in the absence of detergent, indicating that they accumulate in a soluble form.

Variants massively overexpress the periplasmic protein Spy

In examining the periplasmic extracts of the Im7 overexpressing strains, it was impossible to ignore the strong induction of a 15.9-kDa protein, which was identified by mass spectrometry analysis as Spy (spheroplast protein Y)[7]. The Spy protein accounted for up to 48% of total periplasmic content in eight of the ten tested $PenV^R$ or $CdCl_2^R$ EMS mutant strains (**Fig. 3**, **Supplementary Fig. 2** and **Supplementary Table 1b**), suggesting that Spy overexpression might be involved in the stabilization of our Im7 variants.

Spy expression is under the control of the Bae and Cpx periplasmic stress response systems[7–9], which are induced by a variety of stress conditions known to cause protein unfolding and aggregation[2,10]. To identify the mutation(s) that caused the massive upregulation of Spy expression in EMS4 and EMS9, we sequenced the *spy* gene, its regulators *cpxARP* and *baeRS*, and a number of other candidate genes picked because they encode periplasmic chaperones, proteases,

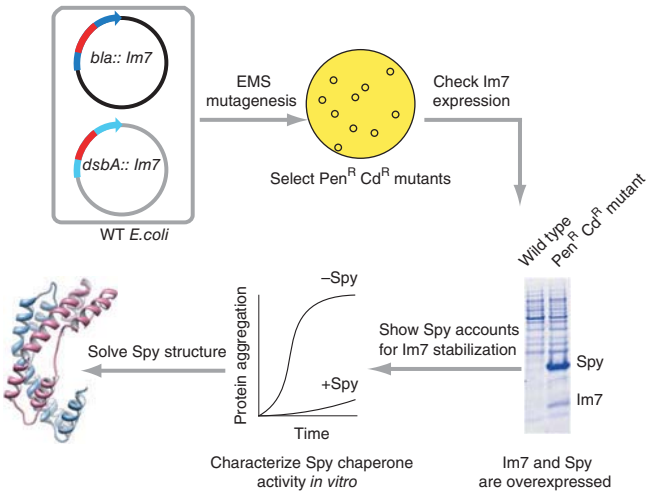

Figure 2 Overall experimental scheme. Fusion constructs that link the stability of the poorly folded Im7 protein to $PenV^R$ (*bla::Im7*) and to $CdCl_2^R$ (*dsbA::Im7*) were introduced into the same *E. coli* strain. Following mutagenesis by EMS treatment, mutants that simultaneously enhanced both $PenV^R$ and $CdCl_2^R$ were selected, and levels of Im7 in these strains were measured. The mutants massively increased levels of Im7 as well as a host protein called Spy. Genetic experiments showed that increased Spy levels are necessary and sufficient to increase Im7 levels. The Spy protein was examined for molecular chaperone activity *in vitro* and found to be highly effective as a chaperone in preventing protein aggregation and aiding refolding. Spy was then crystallized, its structure was solved and mutants were made based on its structure to explore the interaction of Spy with substrates.

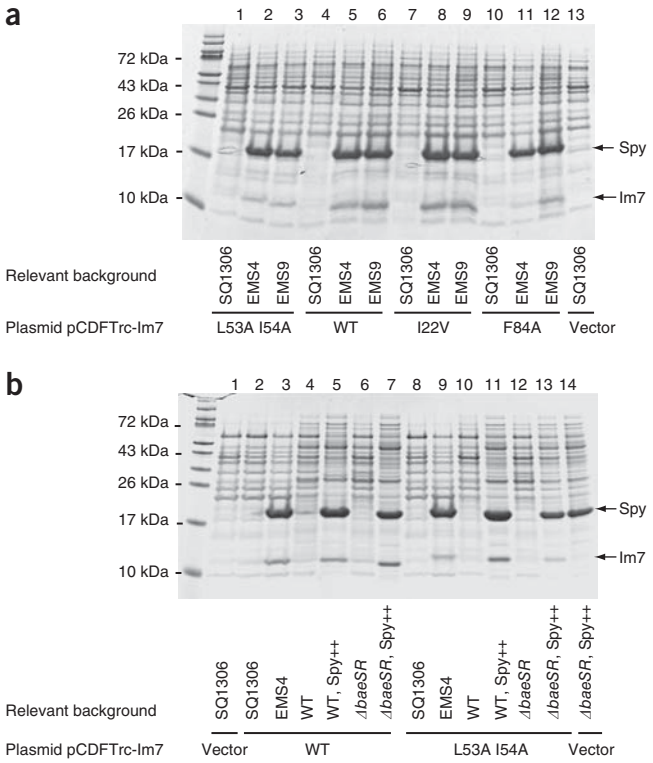

Figure 3 Im7 and Spy are abundant in the periplasm of EMS strains. (**a**) The PenV^R and $CdCl_2{}^R$ mutants EMS4 and EMS9 and the parental SQ1306 strain were transformed with plasmids encoding wild-type Im7 (WT) or the destabilized variants L53A I54A, I22V and F84A[3,6]. Periplasmic extracts were prepared and analyzed by SDS-PAGE. (**b**) Spy overexpression is sufficient to enhance Im7 levels in both wild-type and *ΔbaeSR* backgrounds to those seen in EMS4. Plasmid-encoded Im7 and its destabilized variants were expressed in SQ765 wild-type (WT) or *ΔbaeSR* backgrounds, with or without the coexpression of plasmid-encoded Spy (designated as Spy++). Aggregated or insoluble proteins were not expected to be extracted using our periplasmic extraction procedure and therefore was not detected in these experiments.

stress response regulators or multidrug exporters—*surA, skp*, *fkpA*, *prc*, *degP*, *ptr*, *ompP*, *ompT*, *rcsABCDF* and *mdtABCD*—along with their upstream regulator sequences. We discovered that both EMS4 and EMS9 contained mutations in *baeS* but not in any of the other sequenced genes or regulatory sequences (**Supplementary Table 1c**). We then sequenced the *baeS* gene in all other independently isolated PenV^R or $CdCl_2{}^R$ strains (EMS1–EMS10) and determined that it was mutated in all strains except EMS7, a strain that failed to overproduce Spy or Im7. A total of five different mutations were found in *baeS* (**Supplementary Table 1c**).

BaeS is a putative histidine kinase that, together with the proposed response regulator BaeR, makes up the two-component BaeSR envelope stress response regulation system[2,7–9]; this system has previously been shown to regulate *spy* and a few other periplasmic stress genes. We picked one *baeS* mutation, *baeS-R416C*, for further analysis. We established that the *baeS-R416C* mutation was necessary and sufficient for enhanced Im7 expression (**Supplementary Fig. 3**). Transcriptional analysis of EMS4 revealed not only a massive upregulation of *spy* mRNA but also a substantial induction of other known downstream targets of BaeSR (**Supplementary Table 1d**), suggesting that the *baeS-R416C* mutation caused the constitutive activation of the BaeSR envelope stress response.

Spy overexpression is sufficient to enhance Im7 expression

To determine if Spy overproduction alone is responsible for the enhanced levels of the unstable Im7 protein or whether other BaeSR-regulated proteins are involved as well, we overproduced Spy from the pTrc promoter to levels similar to those seen in the EMS4 mutant, in a *baeSR* null background. We found that for the various destabilized Im7 mutants, overexpression of Spy, even in the absence of a functional BaeSR system, led to soluble Im7 levels that were very similar to those observed in EMS4 by SDS-PAGE analysis (**Fig. 3b**). Quantification of Im7 and Spy levels showed that upon Spy overexpression, Im7 levels increased 100- to 700-fold (**Table 1**). The overexpression of downstream targets of BaeSR other than *spy* apparently does not contribute to the observed PenV resistant phenotype, as their individual deletions had no effect on PenV resistance (**Supplementary Fig. 3**). Based on these results, we concluded that Spy overproduction is necessary and sufficient to increase the levels of soluble periplasmic Im7. Strains deficient in the protease DegP showed increased levels of Im7 (**Supplementary Fig. 4a**), so it appears that Spy is acting at least in part to protect Im7 from proteolysis.

Spy has chaperone activity *in vitro*

Although sequences homologous to Spy are present in a wide variety of enterobacteria, protobacteria and some cyanobacteria (**Supplementary Fig. 5**), very little was previously known about Spy function[7]. Deletion of *spy* was reported to cause slight induction of *degP* and *rpoH*, two genes under the control of σ^E, the stress-response σ-factor involved in outer membrane protein biogenesis, leading to the suggestion that Spy may also be involved in this process[11]. Using quantitative reverse transcription polymerase chain reaction (qRT-PCR), we were unable to detect substantial induction of these or other periplasmic stress-regulated genes upon deletion of *spy* in our strain background (**Supplementary Table 1d**), suggesting that *spy* deletion does not cause substantial defects in membrane integrity.

Our finding that Spy overexpression leads to the accumulation of an otherwise highly unstable protein instead suggested that Spy might

Table 1 Spy induction increases Im7 levels substantially

Strains compared	Relevant host genotypes[a] compared	Im7 variant	Fold increase in Im7 level
SQ1406/SQ1414	EMS4/WT	L53A I54A	66 ± 28
SQ1410/SQ1414	EMS9/WT	L53A I54A	92 ± 44
SQ1805/SQ1809	*spy*$^{++}$/WT	L53A I54A	280 ± 145
SQ1826/SQ1830	*ΔbaeSR, spy*$^{++}$/*ΔbaeSR*	L53A I54A	684 ± 162
SQ1405/SQ1413	EMS4/WT	WT	34 ± 12
SQ1409/SQ1413	EMS9/WT	WT	43 ± 24
SQ1804/SQ1808	*spy*$^{++}$/WT	WT	165 ± 93
SQ1825/SQ1829	*ΔbaeSR, spy*$^{++}$/*ΔbaeSR*	WT	211 ± 83
SQ1407/SQ1415	EMS4/WT	I22V	51 ± 9
SQ1411/SQ1415	EMS9/WT	I22V	48 ± 23
SQ1806/SQ1810	*spy*$^{++}$/WT	I22V	103 ± 14
SQ1827/SQ1831	*ΔbaeSR, spy*$^{++}$/*ΔbaeSR*	I22V	470 ± 176
SQ1408/SQ1416	EMS4/WT	F84A	43 ± 18
SQ1412/SQ1416	EMS9/WT	F84A	67 ± 16
SQ1807/SQ1811	*spy*$^{++}$/WT	F84A	694 ± 243
SQ1828/SQ1832	*ΔbaeSR, spy*$^{++}$/*ΔbaeSR*	F84A	589 ± 102

Data indicate extent of Im7 overproduction in various strains (see **Supplementary Table 1a** for details of strains). Fold increases in Im7 levels were quantified with an Agilent Bioanalyzer 2100 by direct comparison between periplasmic extracts prepared from various pairs of strains (see Online Methods). Values are the average of measurements from three biological samples ± s.d. WT, wild type.
[a]*spy*$^{++}$ designates cells that overexpress Spy from pTrc-spy.

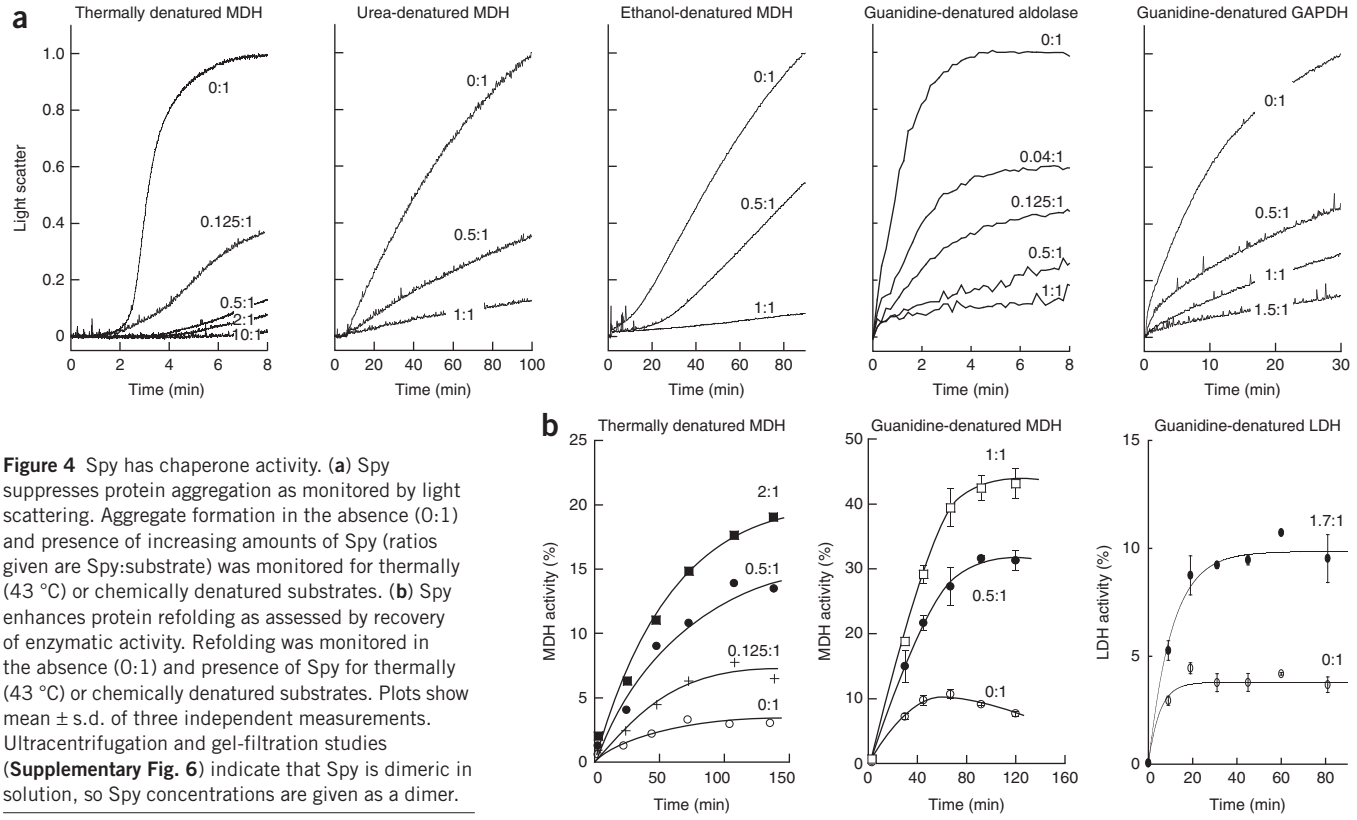

Figure 4 Spy has chaperone activity. (**a**) Spy suppresses protein aggregation as monitored by light scattering. Aggregate formation in the absence (0:1) and presence of increasing amounts of Spy (ratios given are Spy:substrate) was monitored for thermally (43 °C) or chemically denatured substrates. (**b**) Spy enhances protein refolding as assessed by recovery of enzymatic activity. Refolding was monitored in the absence (0:1) and presence of Spy for thermally (43 °C) or chemically denatured substrates. Plots show mean ± s.d. of three independent measurements. Ultracentrifugation and gel-filtration studies (**Supplementary Fig. 6**) indicate that Spy is dimeric in solution, so Spy concentrations are given as a dimer.

function as a chaperone that facilitates protein folding in the bacterial periplasm. To assess its chaperone activity, we purified Spy and analyzed its influence on the aggregation of a number of substrate proteins *in vitro* (**Fig. 4a**). We first tested the effect of Spy on the aggregation of thermally denatured malate dehydrogenase (MDH) and found that addition of increasing amounts of Spy substantially reduced protein aggregation. Even substoichiometric quantities of Spy effectively inhibited the aggregation process, suggesting that Spy is a highly efficient chaperone. Analysis of the effects of Spy on urea-denatured MDH revealed similar results and showed that Spy effectively prevents MDH from aggregating. We found that Spy, which is strongly induced by the protein denaturant ethanol[9], protects MDH from ethanol-mediated aggregation. We also observed a strong suppression of aggregation of chemically denatured aldolase and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) by Spy. Given that heat, urea and ethanol use different mechanisms to unfold proteins, we concluded that Spy must have a general affinity for a wide range of different protein-unfolding intermediates. Combined, our results strongly suggested that we had identified a new, general chaperone in the periplasm of *E. coli*.

Most known ATP-dependent chaperones, such as the DnaK and GroEL systems, function as folding chaperones[12]. They use cycles of ATP binding and hydrolysis to regulate substrate binding and release, thus facilitating protein folding[12]. In contrast, most known ATP-independent chaperones function as holding chaperones, which

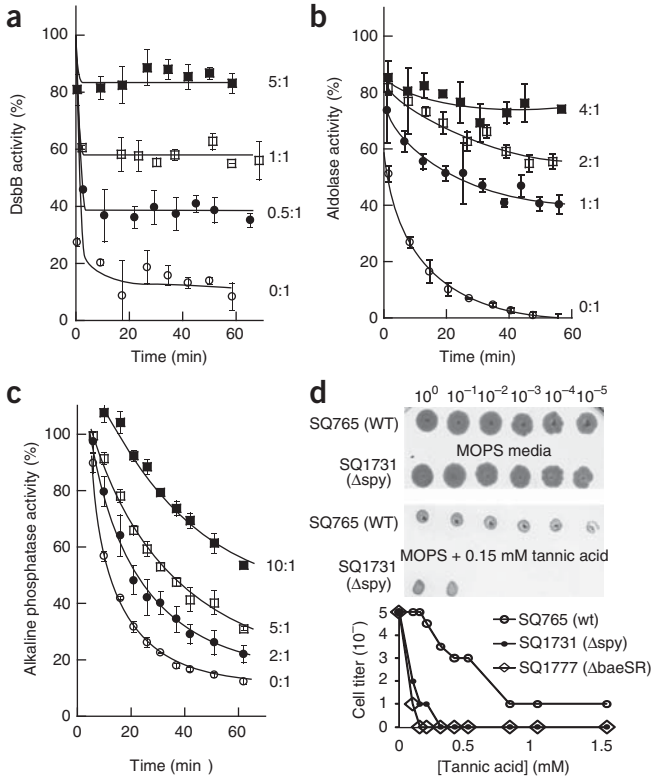

Figure 5 Spy protects DsbB, aldolase and alkaline phosphatase from tannic acid–induced activity loss. Spy concentrations are given as a dimer. Plots show mean ± s.d. for three independent measurements. (**a**) Enzymatic activity of *E. coli* DsbB (0.5 μM) incubated in 100 μM tannic acid in the absence or presence of increasing amounts of Spy (ratios given are Spy:substrate). (**b**) Enzymatic activity of rabbit muscle aldolase (0.5 μM) incubated in 16 μM tannic acid in the absence or presence of increasing amounts of Spy. (**c**) Enzymatic activity of *E. coli* alkaline phosphatase (AP) (1 μM) incubated in 500 μM tannic acid in the absence or presence of increasing amounts of Spy. (**d**) *spy* and *baeSR* deletion strains are tannin sensitive.

Table 2 Data collection, phasing and refinement statistics for Spy

	SeMet Data
Data collection	
Space group	$P6_2$
Cell dimensions	
a, *b*, *c* (Å)	69.0, 69.0, 124.3
α, β, γ (°)	90, 90, 120
Wavelength	0.9792
Resolution (Å)[a]	50-2.6 (2.66-2.6)
R_{sym}[a]	0.116 (0.575)
$I / \sigma I$[a]	37.1 (2.9)
Completeness (%)[a]	97.8 (83.1)
Redundancy[a]	9.3 (5.2)
Refinement	
Resolution (Å)	2.6
No. reflections	10,139
R_{work} / R_{free}	0.243/0.281
No. atoms	
Protein	1,512
Ion	21
Water	9
B-factors	
Protein	81
Ion	63
Water	83
R.m.s. deviations	
Bond lengths (Å)	0.014
Bond angles (°)	1.32

There are two independent molecules in the asymmetric unit, each containing 11 methionine residues. The molecules are related by a noncrystallographic twofold symmetry and are very similar. No NCS restraints were applied during refinement and the two molecules superimpose with r.m.s. deviation of 0.3 Å for 96 Cα atoms and 1.1 Å for all atoms.
[a]Values in parenthesis correspond to the highest-resolution outer shell.

prevent protein aggregation but usually lack the ability to support protein folding[13]. To test whether Spy can support protein folding, even though it is localized to the ATP-devoid environment of the bacterial periplasm, we analyzed its influence on the refolding yield of chemically and thermally unfolded proteins. We found that Spy substantially increased the refolding yield of a number of substrates (**Fig. 4b**). Because the assays were done in the absence of any cofactors, these results strongly suggest that Spy has intrinsic protein folding capacity, providing an excellent explanation of how Spy overexpression is sufficient to substantially increase the amount of folded Im7 protein *in vivo*.

Spy protects proteins from tannin inactivation

qRT-PCR-based measurements indicated that induction of *spy* mRNA is increased nearly 500-fold in response to tannin treatment[14]. Induction of the gene for IbpB, an *E. coli* small heat-shock protein homolog also involved in inhibiting protein aggregation, is increased 48-fold by tannins, making it second only to *spy* in induction by tannins[14]. Tannins, which have long been used to tan leather, are the fourth most abundant component of vascular plant tissue and are synthesized by plants as a protection against bacterial and fungal infections[15]. Tannins are thought to be responsible for the astringent taste of many human food substances including red wine, strong tea and unripe fruit[16]. Some forage crops contain up to 25% tannins by dry weight. *E. coli* found in the gut of herbivores is thus exposed to high concentrations of tannins, driving it to develop tannin resistance[17]. Tannins can have human disease–related antimicrobial effects[18]. The tannins present in cranberry juice, for instance, act as potent inhibitors of the attachment of pathogenic *E. coli* to the uroepithelium and are thought to explain the effectiveness of cranberry juice in preventing urinary tract infections[19]. Only small quantities of tannins are required to aggregate proteins, a feature that may be responsible for their antimicrobial activity[20]. Their astringent taste in foods may be due to the precipitation of mouth proteins[21]. The astringent taste of strong tea is often reduced by addition of milk, which drives co-precipitation of the tannins with the disordered protein casein that is present in milk[21]. Tannic acid has also been reported to have potent anti-amyloidogenic activity[22–24].

We tested the effects of Spy on tannin-mediated inactivation of the *E. coli* membrane protein DsbB (**Fig. 5a**), aldolase (**Fig. 5b**) and the *E. coli* periplasmic protein alkaline phosphatase (**Fig. 5c**). Spy protected all three proteins from tannic acid–induced activity loss. Consistent with this *in vitro* result, we discovered that *spy* null mutants are highly sensitive to tannins (**Fig. 5d**) and show decreased alkaline phosphatase activity *in vivo* (**Supplementary Fig. 2**). *baeSR* mutants have been reported to be tannin sensitive[14], and we found that most, but not all, of the tannin sensitivity could be attributed to induction of Spy (**Fig. 5d**). The antimicrobial action of tannins can substantially alter the microbial content of the gut; indeed, the relatively high tannin resistance of enterobacteria, which we show here is mediated at least in part by Spy, appears to be responsible for allowing the population of fecal enterobacteria to increase up to 19-fold in rats fed a high-tannin diet[17]. We conclude that the induction of Spy is likely to be involved in protecting cells from tannin-induced protein aggregation and inactivation *in vitro* and *in vivo*.

Spy may not have a regulatory role like its homolog CpxP

The Spy protein is 29% identical to CpxP, an inhibitory component of the CpxRA regulatory system[25]. CpxP binds to the periplasmic domain of CpxA, inhibiting its autokinase activity. The presence of unfolded proteins causes the release of CpxP, thereby activating the Cpx response[25,26]. These observations suggest that CpxP might be acting as one of the very few known periplasmic chaperones, targeting itself and its unfolded protein cargo to the protease DegP for degradation[27,28]. We found that CpxP has weak chaperone activity *in vitro* and CpxP overproduction causes the accumulation of Im7 in otherwise wild-type strains (**Supplementary Fig. 4**). To assess whether Spy might be involved in regulating BaeSR or the other periplasmic stress-response systems, we carried out qRT-PCR in strains either lacking or overexpressing *spy* (**Supplementary Table 1d**). We found no notable influence on the expression of Bae-, Cpx-, σ^E-, σ^{32}-, Rcs- or Psp-regulated genes, suggesting that Spy, unlike CpxP, may not be playing a major regulatory role. Instead, our results strongly suggest that Spy functions directly as a molecular chaperone.

Spy has a novel chaperone fold—an α-helical cradle

To gain insights into the mechanism of Spy's chaperone action, we crystallized and determined the three-dimensional structure of histidine-tagged Spy (**Table 2**). The crystal structure shows that Spy molecules associate into tightly bound dimers (**Fig. 6**). Size exclusion chromatography and analytical ultracentrifugation (**Supplementary Fig. 6**) confirmed this oligomerization state and revealed that Spy is also dimeric in solution. Each Spy monomer consists of four α-helices (α1–α4). The 28 N-terminal residues and 14 C-terminal residues are disordered in the crystal. Helices α1, α2 and α3 fold into a hairpin, with α1 and α2 forming one arm and α3 the other arm (**Fig. 6a** and **Supplementary Fig. 6**). Helix α3 has nine turns and is bent in

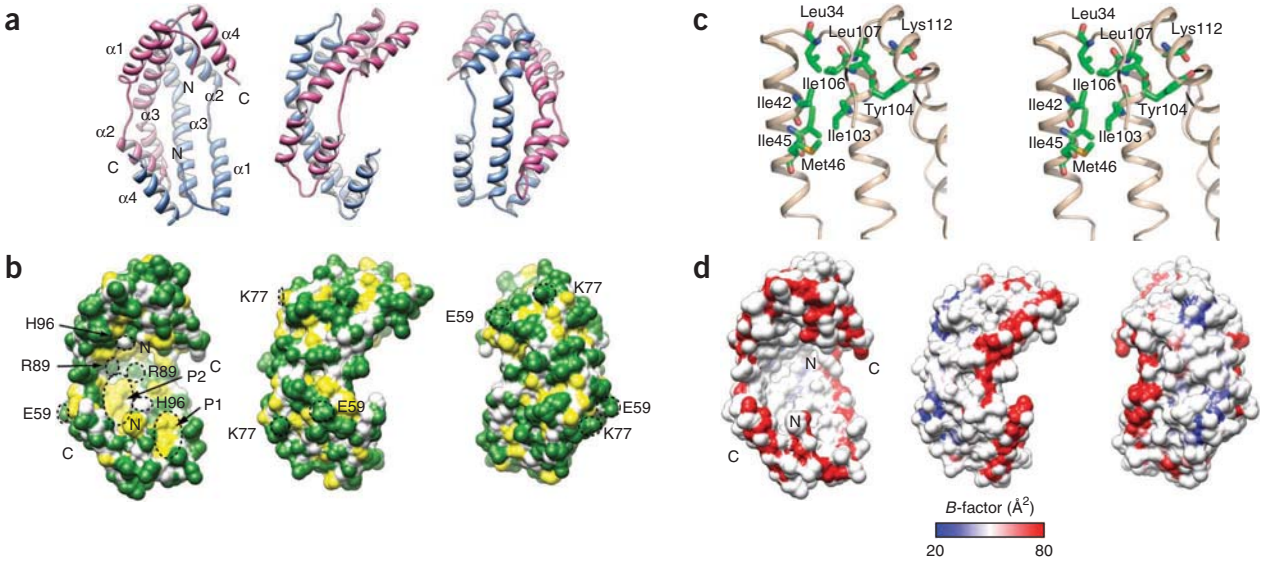

Figure 6 Crystal structure of the Spy dimer shown in three orientations rotated by 90° along the vertical axis. For ease of comparison, the orientations shown in panels **a**, **b** and **d** are identical. (**a**) Ribbon drawing shows an all–α-helical structure. One subunit is colored light blue and the other is colored magenta. The N and C termini and the secondary structural elements of the molecule are labeled. (**b**) Surface properties of Spy, colored as: backbone atoms, white; polar and charged side chain atoms, green; hydrophobic side chain atoms, yellow. Two predominantly hydrophobic patches in the concave surface are indicated as P1 (composed of Leu34, Ile42, Met46 and Ile103) and P2 (composed of Pro56, Met64, Ile68, Met85, Met93 and Met97). The residues labeled with fluorescent probes (for experiments in **Fig. 7**) are circled with black dashed lines. (**c**) Stereoview showing the cluster of hydrophobic residues at the tip of the cradle marked as P1 in **b**. These residues are well conserved among homologous sequences (**Supplementary Fig. 5**). (**d**) Structural flexibility of Spy dimer. The molecular surface representing Spy backbone atoms is colored based on the average *B*-factors for each residue. Note that the rim lining the concave surface has higher *B*-factors, indicating greater structural flexibility; in particular, the N and C termini are highly mobile.

the middle by ~30° because of partial unwinding at Met85–Glu86. The N-terminal half of α3 is parallel to α2, whereas its C-terminal half is parallel to α1. Helix α4 runs antiparallel to α3 and is inclined to it by ~45°. The first six ordered N-terminal residues, Phe29–Leu34, assume an extended conformation and follow along helix α4. Following the submission of our work, the crystal structure of Spy was determined as part of a high-throughput effort[29]; as expected, this structure is very similar to ours. The Spy dimer is formed through the antiparallel coiled-coiled interaction. The shape of the dimer is rather unusual, with one surface highly concave and the other convex, reminiscent of a cradle. The bottom of this cradle is formed by helices α3, the sides by the connection between helices α1 and α2 and the tips by helices α1 and α4. The cradle is extremely thin in cross-section; its average thickness is 9.2 Å (**Supplementary Fig. 6**). The contacts between the two monomers are extensive, burying a surface of ~1,850 Å^2 per monomer upon dimerization and suggesting high dimeric stability. Although the concave surface has an overall positive charge[29], it is also lined with a number of conserved apolar side chains, localized in two clusters that are exposed as hydrophobic patches (**Fig. 6b**,**c**).

Substrate interaction with Spy

To investigate the interaction between Spy and protein substrates, we labeled Spy with two different environmentally sensitive probes, acrylodan and 4-(*N*-(iodoacetoxy)ethyl-*N*-methyl)amino-7-nitrobenz-2-oxa-1,3-diazole (IANBD). Probe attachment sites were made by substituting cysteines at six different positions; these included residues exposed to both the concave (R89C, H96C) and convex sides of the cradle (E59C, K77C), as well as residues within the structurally disordered N and C termini (H24C and A128C) (**Fig. 6b**). We then determined the influence of substrate addition on the fluorescence of the probe-labeled Spy variants. Changes in the microenvironment near the probes caused by substrate addition could reflect either direct binding to the labeled region of the probe or structural rearrangements within Spy caused by substrate binding.

The Spy variants retained a substantial ability to inhibit the aggregation of aldolase, showing that they are at least partially active (**Supplementary Table 1e** and **Supplementary Fig. 7a**). For the interaction experiments, we used casein as a Spy substrate, because, unlike most of our other Spy substrates, it is intrinsically disordered but soluble; Im7-L53A I54A was also chosen because it is an excellent Spy substrate *in vivo*. Im7-L53A I54A *in vitro* is trapped as a partially unfolded but soluble intermediate[3]. Analytical gel filtration revealed the presence of an apparently stable complex between Spy and casein (**Fig. 7a**). Equimolar addition of casein or Im7 substantially decreased Spy-mediated refolding of MDH (**Fig. 7b**,**c**). We interpreted these data as likely due to direct competition for the same binding site on Spy. We then measured changes in the fluorescence emission spectra of the various labeled Spy variants upon addition of equimolar quantities of casein or Im7-L53A I54A, thereby monitoring potential environmental changes in the vicinity of the labeled residues upon substrate binding. The fluorescence of acrylodan attached via H24C, E59C, R89C, H96C and A128C substantially increased and blue-shifted with casein addition, suggesting that the region near these residues becomes more hydrophobic in the presence of casein (**Fig. 7d** and **Supplementary Fig. 7b**). Acrylodan-labeled H24C and A128C Spy variants also showed fluorescence increases and slight blue shifts upon Im7-L53A I54A addition (**Fig. 7e** and **Supplementary Fig. 7c**). IANBD is generally not as sensitive as acrylodan in its ability to reflect changes in the fluorophore environment. Nevertheless, IANBD-labeled H24C, H96C and A128C Spy variants showed substantially decreased fluorescence upon casein binding (**Supplementary Fig. 7d**). Because the fluorescence of IANBD is quenched in hydrophobic environments[30], these results suggest that these residues become more hydrophobic in character with casein binding. In contrast, the fluorescence of acrylodan attached

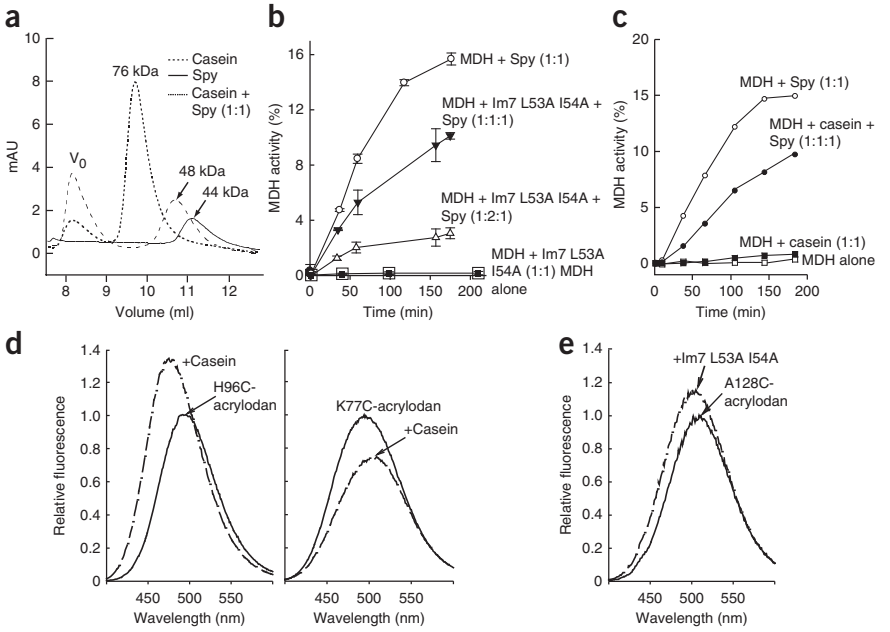

Figure 7 Spy binds the disordered model substrate protein casein and the *in vivo* substrate protein Im7-L53A I54A. (**a**) Analytical gel filtration of Spy, casein or a 1:1 mixture of Spy and casein. The molecular weight of a dimer of Spy is 31 kDa, but it elutes as a 44-kDa molecule, consistent with the elongated form of the dimer seen in the crystal structure. The molecular weight of casein is 23–26 kDa. (**b**) Competition assay between urea-denatured Im7-L53A I54A and MDH for Spy in MDH refolding. Refolding of chemically denatured MDH was monitored by recovery of enzymatic activity plotted as a fraction of the activity of native (that is, nondenatured) MDH. Note that the curves of MDH alone (open squares) and MDH + Im7-L53A I54A (closed squares) overlap precisely. (**c**) Competition assay between casein and MDH for Spy binding. Refolding of chemically denatured MDH was monitored as in **b**. (**d**) Normalized fluorescence emission spectra of acrylodan-labeled Spy mutants H96C and K77C in the absence or presence of an equimolar amount of casein. Note that the fluorescence change upon casein addition for Spy K77-acrylodan is opposite that from Spy H96C-acrylodan. (**e**) Normalized fluorescence emission spectra of acrylodan-labeled Spy A128C in the absence or presence of an equimolar amount of urea-denatured Im7-L53A I54A.

via K77C is decreased with casein addition (**Fig. 7d**), indicating that the region near this residue probably becomes more hydrophilic with substrate binding. That nearly all our Spy mutants show substantial changes in their environment upon substrate addition suggests that substrate binding occurs over large regions of Spy (including both the concave and convex sides of the cradle). Alternatively, major structural rearrangements or higher-order oligomerization reactions could occur within Spy upon substrate binding, or both.

DISCUSSION

It may seem unusual that a chaperone as effective as Spy has not been studied in an organism as well-characterized as *E. coli*. However, because chaperones are usually effective only in stoichiometric quantities, chaperone assays are generally not sensitive enough to enable their purification from crude lysates by activity. Instead, chaperones have often been first identified because their genes are induced in protein unfolding conditions[2]. Our approach of linking folding to selectable markers opens up the possibility of directing evolution to alter the *in vivo* folding environment to specifically enhance the folding of a given unstable protein and provides a new route for chaperone discovery.

Spy has a unique cradle shape that is unlike that of any other chaperone whose structure is known. Spy lacks a globular core and would therefore be expected to have higher flexibility than the average globular protein. The molecule averages 9.2 Å in thickness, less than the 12-Å diameter of a single α-helix. Its thinness places a disproportionate number of side chains on the protein surface. The highest backbone temperature factors are observed in surfaces and bumps extending from the concave side of the cradle, particularly in the connectors between helices α1 and α2, which form the sides of the cradle, suggesting the possibility for bending and twisting of the molecule (**Fig. 6d**). Spy's shape combined with its apparent flexibility leads us to propose a model for Spy action that involves the shielding of aggregation-sensitive regions on substrate proteins, which are revealed upon partial unfolding by interaction with Spy. Spy's apparently highly flexible nature may allow the chaperone to accommodate a variety of partially unfolded protein substrates. The extremely high flexibility of another periplasmic chaperone, HdeA, for instance, appears to allow it to bind numerous substrate proteins and prevent aggregation[31].

Addition of 4 mM tannic acid results in the massive induction of Spy so that it comprises ~25% of periplasmic proteins (**Supplementary Fig. 2d**), similar to the induction seen in our *baeS* constitutive EMS strains (**Supplementary Table 1b**). Of *E. coli*'s ~4,280 genes, *spy* is one of those most strongly induced by butanol[32,33], a protein unfolding agent[34], leading to Spy comprising ~20% of periplasmic proteins (**Supplementary Fig. 2d**). Ethanol, another well-known protein denaturant and potent inducer of the heat-shock response[35], also strongly induces Spy so that it makes up ~5% of periplasmic proteins[9] (**Supplementary Fig. 2d**). Spy is also strongly induced by other conditions that induce protein unfolding, such as overproduction of misfolded PapG and NlpE[7]. Strong induction of Spy is thus physiological, not something that only occurs in our BaeS constitutive mutants. That Spy is induced so strongly by protein unfolding or precipitation agents that it becomes up to 25% of the periplasmic contents implies a pressing need for it to respond to unfolding conditions. Other chaperones also attain high percentage abundances after stresses that unfold proteins. GroEL, for instance, becomes 12% of the total cellular protein during growth at 46 °C (ref. 36), but the abundance of Spy is more extreme, particularly when calculated on a molar basis. The total concentration of protein in cells is ~300 g l^{-1} (ref. 37), making the molar concentration of the 31-kDa Spy dimer in the periplasm up to 2.4 mM. GroEL functions as a 790-kDa 14-mer, so its molar concentration in the cell when 12% of total cellular contents is only about 0.05 mM, or about 50-fold less abundant on a molar basis than the physiological levels of induced Spy. Even though Im7 can attain ~10% of the cell protein upon Im7 overproduction, Spy is even more abundant; upon Spy overproduction, approximately two Spy dimers are present in the periplasm for every Im7 monomer.

The extreme abundance of Spy leads us to propose a speculative model for Spy action in which it protects proteins *in vivo* by binding to aggregation-prone regions exposed on the surface of unstable proteins, coating these regions (or possibly even the entire unstable protein)

with a thin layer that inhibits proteolysis and/or aggregation. Upon Spy overproduction, the levels of unstable Im7 mutants go from barely detectable to up to 10% of periplasmic extracts (**Supplementary Table 1d**), likely because Spy is very effective in inhibiting *in vivo* proteolysis and/or aggregation. Changes in the fluorescence of environmentally sensitive probes labeled with acrylodan or IANBD are consistent with either Spy binding to substrates via large parts of the molecule or major rearrangements occurring in the Spy structure upon substrate binding, or both. Definition of the precise substrate binding site, the stoichiometry of its interaction with substrates and its precise mechanism of action, including the possible involvement of co-chaperones, awaits future experimentation.

Although suppression of protein aggregation does not require energy, release and refolding of bound substrate proteins usually does[38]. ATP is absent in the periplasm and in our assays was not required for the refolding function of Spy, indicating that the chaperone activity of Spy is ATP-independent. Spy thus appears to be one of the very few chaperones that actively support protein refolding in the absence of any obvious energy source, suggesting that Spy uses a mechanism to control substrate binding and release that is different from that of previously characterized chaperones. How Spy enables refolding in the absence of energy cofactors is a provocative question for future research. At this point, we can only speculate that the apparently highly dynamic nature of the Spy structure may permit structural fluctuations that not only allow it to mold to various proteins but also enable it to release them.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/nsmb/.

Accession codes. Protein Data Bank: the coordinates and structure factors have been deposited with accession code 3O39.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

We thank G. Morgan and S. Radford for communicating results before publication and for the Im7-L53A I54A protein, T. Franzmann for conducting the ultracentrifugation experiments, D. Reichmann, C. Cremers and A. Malik for advice, M. Lei and Y. Chen for providing vector and reagents for Spy purification and C. Munger for initial purification and crystallization of Spy. Protein identification was done by the Protein Structure Facility at the University of Michigan. Diffraction data were collected at the Canadian Macromolecular Crystallography Facility-1 beamline at the Canadian Light Source. We would like to thank S. Labiuk for data collection. The Howard Hughes Medical Institute funded this work (to J.C.A.B). M.C. acknowledges financial support from Canadian Institutes of Health Research grant GSP-48370.

AUTHOR CONTRIBUTIONS

J.C.A.B. designed the study and wrote the manuscript, with contributions from M.C. and S.Q. S.Q., P.K., T.T., N.K., K.M.R., R.S., J.P., S.H. and G.R. conducted the experiments and collected and analyzed the data. J.C.A.B., Z.X. and M.C. further analyzed the data. L.F. and U.J. provided technical support and conceptual advice.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/nsmb/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Kerner, M.J. *et al.* Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. *Cell* **122**, 209–220 (2005).
2. Duguay, A.R. & Silhavy, T.J. Quality control in the bacterial periplasm. *Biochim. Biophys. Acta* **1694**, 121–134 (2004).
3. Foit, L. *et al.* Optimizing protein stability *in vivo*. *Mol. Cell* **36**, 861–871 (2009).
4. Stafford, S.J., Humphreys, D.P. & Lund, P.A. Mutations in *dsbA* and *dsbB*, but not *dsbC*, lead to an enhanced sensitivity of *Escherichia coli* to Hg^{2+} and Cd^{2+}. *FEMS Microbiol. Lett.* **174**, 179–184 (1999).
5. De Pascale, G. & Wright, G.D. Antibiotic resistance by enzyme inactivation: from mechanisms to solutions. *ChemBioChem* **11**, 1325–1334 (2010).
6. Spence, G.R., Capaldi, A.P. & Radford, S.E. Trapping the on-pathway folding intermediate of Im7 at equilibrium. *J. Mol. Biol.* **341**, 215–226 (2004).
7. Raffa, R.G. & Raivio, T.L. A third envelope stress signal transduction pathway in *Escherichia coli*. *Mol. Microbiol.* **45**, 1599–1611 (2002).
8. MacRitchie, D.M., Buelow, D.R., Price, N.L. & Raivio, T.L. Two-component signaling and gram negative envelope stress response systems. *Adv. Exp. Med. Biol.* **631**, 80–110 (2008).
9. Bury-Moné, S. *et al.* Global analysis of extracytoplasmic stress signaling in *Escherichia coli*. *PLoS Genet.* **5**, e1000651 (2009).
10. Price, N.L. & Raivio, T.L. Characterization of the Cpx regulon in *Escherichia coli* strain MC4100. *J. Bacteriol.* **191**, 1798–1815 (2009).
11. Raivio, T.L., Laird, M.W., Joly, J.C. & Silhavy, T.J. Tethering of CpxP to the inner membrane prevents spheroplast induction of the cpx envelope stress response. *Mol. Microbiol.* **37**, 1186–1197 (2000).
12. Hartl, F.U. & Hayer-Hartl, M. Converging concepts of protein folding *in vitro* and *in vivo*. *Nat. Struct. Mol. Biol.* **16**, 574–581 (2009).
13. Winter, J. & Jakob, U. Beyond transcription—new mechanisms for the regulation of molecular chaperones. *Crit. Rev. Biochem. Mol. Biol.* **39**, 297–317 (2004).
14. Zoetendal, E.G., Smith, A.H., Sundset, M.A. & Mackie, R.I. The BaeSR two-component regulatory system mediates resistance to condensed tannins in *Escherichia coli*. *Appl. Environ. Microbiol.* **74**, 535–539 (2008).
15. Scalbert, A. Antimicrobial properties of tannins. *Phytochemistry* **30**, 3875–3883 (1991).
16. Hernes, P.J. *et al.* Tannin diagenesis in mangrove leaves from a tropical estuary: A novel molecular approach. *Geochim. Cosmochim. Acta* **65**, 3109–3122 (2001).
17. Smith, A.H. & Mackie, R.I. Effect of condensed tannins on bacterial diversity and metabolic activity in the rat gastrointestinal tract. *Appl. Environ. Microbiol.* **70**, 1104–1115 (2004).
18. Serrano, J., Puupponen-Pimia, R., Dauer, A., Aura, A.M. & Saura-Calixto, F. Tannins: current knowledge of food sources, intake, bioavailability and biological effects. *Mol. Nutr. Food Res.* **53**, S310–S329 (2009).
19. Howell, A.B., Vorsa, N., Marderosian, A.D. & Foo, L.Y. Inhibition of the adherence of P-fimbriated *Escherichia coli* to uroepithelial-cell surfaces by proanthocyanidin extracts from cranberries. *N. Engl. J. Med.* **339**, 1085–1086 (1998).
20. Zanchi, D. *et al.* Tannin-assisted aggregation of natively unfolded proteins. *Europhys. Lett.* **82**, 58001 (2008).
21. Hofmann, T. *et al.* Protein binding and astringent taste of a polymeric procyanidin, 1,2,3,4,6-penta-O-galloyl-beta-D-glucopyranose, castalagin, and grandinin. *J. Agric. Food Chem.* **54**, 9503–9509 (2006).
22. Ono, K., Hasegawa, K., Naiki, H. & Yamada, M. Anti-amyloidogenic activity of tannic acid and its activity to destabilize Alzheimer's β-amyloid fibrils *in vitro*. *Biochim. Biophys. Acta* **1690**, 193–202 (2004).
23. Kim, W. *et al.* A high-throughput screen for compounds that inhibit aggregation of the Alzheimer's peptide. *ACS Chem. Biol.* **1**, 461–469 (2006).
24. Lee, L.L., Ha, H., Chang, Y.T. & DeLisa, M.P. Discovery of amyloid-β aggregation inhibitors using an engineered assay for intracellular protein folding and solubility. *Protein Sci.* **18**, 277–286 (2009).
25. Raivio, T.L., Popkin, D.L. & Silhavy, T.J. The Cpx envelope stress response is controlled by amplification and feedback inhibition. *J. Bacteriol.* **181**, 5263–5272 (1999).
26. Danese, P.N. & Silhavy, T.J. CpxP, a stress-combative member of the Cpx regulon. *J. Bacteriol.* **180**, 831–839 (1998).
27. DiGiuseppe, P.A. & Silhavy, T.J. Signal detection and target gene induction by the CpxRA two-component system. *J. Bacteriol.* **185**, 2432–2440 (2003).
28. Isaac, D.D., Pinkner, J.S., Hultgren, S.J. & Silhavy, T.J. The extracytoplasmic adaptor protein CpxP is degraded with substrate by DegP. *Proc. Natl. Acad. Sci. USA* **102**, 17775–17779 (2005).
29. Kwon, E., Kim, D.Y., Gross, C.A., Gross, J.D. & Kim, K.K. The crystal structure *Escherichia coli* Spy. *Protein Sci.* **19**, 2252–2259 (2010).
30. Powl, A.M., Wright, J.N., East, J.M. & Lee, A.G. Identification of the hydrophobic thickness of a membrane protein using fluorescence spectroscopy: Studies with the mechanosensitive channel MscL. *Biochemistry* **44**, 5713–5721 (2005).
31. Tapley, T.L. *et al.* Structural plasticity of an acid-activated chaperone allows promiscuous substrate binding. *Proc. Natl. Acad. Sci. USA* **106**, 5557–5562 (2009).
32. Brynildsen, M.P. & Liao, J.C. An integrated network approach identifies the isobutanol response network of *Escherichia coli*. *Mol. Syst. Biol.* **5**, 277 (2009); doi:10.1038/msb.2009.34.
33. Rutherford, B.J. *et al.* Functional genomic study of exogenous n-butanol stress in *Escherichia coli*. *Appl. Environ. Microbiol.* **76**, 1935–1945 (2010).
34. Miyawaki, O. & Tatsuno, M. Thermodynamic analysis of alcohol effect on thermal stability of proteins. *J. Biosci. Bioeng.* **111**, 198–203 (2011).
35. Neidhardt, F.C., VanBogelen, R.A. & Vaughn, V. The genetics and regulation of heat-shock proteins. *Annu. Rev. Genet.* **18**, 295–329 (1984).
36. Neidhardt, F.C. *et al.* Identity of the B56.5 protein, the A-protein, and the *groE* gene product of *Escherichia coli*. *J. Bacteriol.* **145**, 513–520 (1981).
37. McNulty, B.C., Young, G.B. & Pielak, G.J. Macromolecular crowding in the *Escherichia coli* periplasm maintains α-synuclein disorder. *J. Mol. Biol.* **355**, 893–897 (2006).
38. Tapley, T.L., Franzmann, T.M., Chakraborty, S., Jakob, U. & Bardwell, J.C.A. Protein refolding by pH-triggered chaperone binding and release. *Proc. Natl. Acad. Sci. USA* **107**, 1071–1076 (2010).

ONLINE METHODS

Mutagenesis and selection to identify new periplasmic folding factors. Bacteria containing the plasmids pBR322bla::GS linker Im7 L53A I54A and pBAD33 dsbA::GS linker Im7 L53A I54A, which encode sandwich fusions of Im7 L53A I54A with β-lactamase and DsbA, respectively, were subject to EMS (Sigma) mutagenesis[39], and mutants resistant to 1,500 μg ml^{-1} penicillin and 0.5 mM $CdCl_2$ (Sigma) were selected. Western blotting of total cell extracts was used to detect the levels of β-lactamase–Im7 fusion protein present in these mutant strains. Following transformation with plasmids that encode wild-type Im7 and various destabilized Im7 mutants expressed under the pTrc promoter, periplasmic extracts were prepared as previously described[40], and the pattern of soluble periplasmic protein expression was examined on SDS gels (Invitrogen). The level of Im7 and Spy present in these extracts was quantified using a Bioanalyzer 2100 (Agilent) as described below.

Quantification of proteins. The amount of the sandwich fusion protein β-lactamase–Im7 L53A I54A in SQ1306 and EMS1–EMS10 was quantified by western blot with anti–β-lactamase antibody (Millipore) using whole cell lysate as described previously[3] with minor modifications. To quantify plasmid-encoded Im7 (wild-type and variants) expressed in the absence of the fusion, cells were grown to mid-log phase in LB medium at 37 °C. Im7 protein expression was induced with 2 mM IPTG for 2 h. Addition of IPTG also induces the expression of Spy when the pTrc-spy plasmid is present. Periplasmic extractions were prepared as described previously[40], and the proteins were separated using a Bioanalyzer 2100 (Agilent) with the High Sensitivity Protein 250 kit (also from Agilent) and the conditions specified by the manufacturer. To visualize the tiny signal corresponding to Im7 present in the absence of Spy expression, a pico labeling protocol was applied (Agilent technical note, publication number 5990-3703EN). Prior to labeling, periplasmic extractions were diluted to different extents to ensure the linear relationship between Im7 amount and signal. Protein ratios were determined by integrating the trough-to-trough peak area. The high sensitivity of the 250 kit and the pico labeling protocol allowed the visualization and quantification of the very small amount of Im7 present in the absence of Spy, an amount that was not detectable either on Coomassie blue–stained SDS-PAGE gels or on the Agilent Bioanalyzer using either the 80 or 230 protein kits. Although excellent for determining the ratio of a single defined protein present in different sample preparations, we reasoned that differences in labeling efficiencies between different proteins would make this protocol suboptimal for percentages of the total protein. Instead, the percentage of Spy and Im7 was quantified in overproducing strains using the protein 80 kit (Agilent), without labeling. Areas corresponding to Spy or Im7 peaks were compared to the total area of all proteins present in the periplasmic extract. In the absence of Spy, the Im7 peak was not detectable; therefore, the percentage of Im7 was calculated by dividing the Im7 percentage in the presence of Spy overexpression with the fold increase of Im7 level. For example, Im7 is 10% of total periplasm in SQ1405. Im7 expression increased 34-fold in SQ1405 compared to SQ1413. The percentage of Im7 in SQ1413 was therefore calculated to be 0.3%. The percentage of Spy present in periplasmic extracts in different EMS strains was quantified similarly with the protein 80 kit.

Crystallization. Initial crystallization conditions of the mature form of Spy (residues 1–138) with N-terminal His_6 tag were identified using the $AmSO_4$ suite (Qiagen). Spy contains 161 residues of which 23 constitute the signal peptide that directs Spy to the periplasm. Our residue numbers refer to the mature, periplasmic form of Spy, which contains 138 residues; the N-terminal His_6 tag is not numbered. The SeMet-substituted protein was crystallized under the same conditions. The best diffracting crystals were obtained by mixing 1 μl protein in the final buffer with 1 μl of reservoir solution containing 0.3 M $CdCl_2$ and 2.4 M $AmSO_4$ under the vapor diffusion hanging-drop method. Prior to data collection, the crystals were cryoprotected in paratone and flash frozen in liquid nitrogen.

Data collection and processing. The diffraction data were collected from a single crystal at the Canadian Macromolecular Crystallography Facility-1 beamline at the Canadian Light Source (CLS) at the selenium absorption edge (wavelength of 0.9792 Å) to 2.6 Å resolution at 100 K. Data were processed and scaled using DENZO and SCALEPACK in the HKL2000 suite[41].

Structure determination and refinement. The structure was solved by the single anomalous diffraction method. Of the 22 expected selenium atoms in the asymmetric unit, eight were located and refined using autoSHARP[42]. These sites were used to obtain preliminary phases. The model was built manually in Coot[43] using the selenium sites as reference points. Several cycles of refinement using REFMAC5[44] followed by model rebuilding were carried out. The final refinement was done with PHENIX[45] and included the translation-libration-screw model for thermal motions. The His_6 tag and the first 28 residues as well as the last 14 C-terminal residues are disordered. The final model includes the central 96 residues for each monomer. Of these, 11 long side chains in molecule A and 15 in molecule B were only partially modeled. The pertinent data are shown in **Table 2**. The model has good geometry as analyzed by Molprobity[46]; 96.9% (309/319) of all residues are in the favored region and 100% of all residues are in the allowed region.

Additional methods. A detailed description of all other methods is given in **Supplementary Methods**.

39. Miller, J.H. in *A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for Escherichia coli and Related Bacteria* (Cold Spring Harbor Laboratory Press, Woodbury, NY, 1992).
40. Hiniker, A. & Bardwell, J.C.A. *In vivo* substrate specificity of periplasmic disulfide oxidoreductases. *J. Biol. Chem.* **279**, 12967–12973 (2004).
41. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Meth. Enzymol.* **276**, 307–326 (1997).
42. Vonrhein, C., Blanc, E., Roversi, P. & Bricogne, G. Automated structure solution with autoSHARP. *Methods Mol. Biol.* **364**, 215–230 (2007).
43. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
44. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* **53**, 240–255 (1997).
45. Adams, P.D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
46. Chen, V.B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).

doi:10.1038/nsmb.2016