

Crystallographic Structure of SurA, a Molecular Chaperone that Facilitates Folding of Outer Membrane Porins

Eduard Bitto and David B. McKay¹
Department of Structural Biology
Stanford University School of Medicine
Stanford, California 94305

Summary

The SurA protein facilitates correct folding of outer membrane proteins in gram-negative bacteria. The sequence of *Escherichia coli* SurA presents four segments, two of which are peptidyl-prolyl isomerases (PPIases); the crystal structure reveals an asymmetric dumbbell, in which the amino-terminal, carboxy-terminal, and first PPIase segments of the sequence form a core structural module, and the second PPIase segment is a satellite domain tethered ~ 30 Å from this module. The core module, which is implicated in membrane protein folding, has a novel fold that includes an extended crevice. Crystal contacts show that peptides bind within the crevice, suggesting a model for chaperone activity whereby segments of polypeptide may be repetitively sequestered and released during the membrane protein-folding process.

Introduction

The *Escherichia coli* *surA* (for “survival”) gene was first identified as a gene whose disruption impaired cell survival in stationary phase [1]. The SurA protein was subsequently shown to be involved in the process of folding and assembly of outer membrane porins [2–4]. Experiments that tracked the maturation of LamB, a trimeric outer membrane porin, showed that, for this particular substrate, SurA specifically facilitates the conversion of apparent unfolded monomers to folded monomers, which then assemble into unstable trimers, which, in turn, convert to stable trimers, in distinct steps [4]. SurA therefore functions as a molecular chaperone that facilitates correct folding of outer membrane proteins. Although disruption of the *surA* gene alone is not lethal, it has been shown that simultaneous null mutations of *surA* and *ppiD*, a gene that also encodes a PPIase involved in maturation of outer membrane proteins, is lethal, indicating that *surA* is responsible for an essential *E. coli* activity that is encoded by redundant genes [5].

The gene sequence of *surA* reveals four distinct regions following a 20-residue leader sequence: an amino-terminal segment (denoted “N” hereafter), which has no identifiable sequence similarity to other proteins besides other SurA homologs, two ~ 100 -residue peptidyl prolyl isomerase (PPIase) domains of the parvulin PPIase family [6] (denoted “P1” and “P2,” respectively), and a carboxy-terminal segment (“C”) of ~ 35 residues (Figure 1A). The presence of two PPIase domains suggested that SurA might facilitate membrane protein folding

through a proline isomerization activity. The specific PPIase activity of full-length SurA protein on a model protein substrate (reduced, carboxymethylated S54G/P55N ribonuclease T1 [RNase T1]) is of the same order as that of parvulin [7]. It has been shown that the PPIase activity of SurA resides exclusively in the second parvulin-like domain; the P1 domain, by itself, is devoid of PPIase activity [4]; the PPIase activity of P2, alone on the model RNase T1 substrate, is two orders of magnitude lower than that of full-length SurA, and the activities of SurA and the P2 domain, alone on small peptide substrates, are approximately equal. However, both parvulin-like domains can be deleted, or, alternatively, the PPIase activity of the second domain can be abolished by mutagenesis, and the resultant modified SurA proteins still impart essentially wild-type *surA* activities in vivo with respect to outer membrane integrity and induction of the σ^E -dependent stress response; they also retain in vitro chaperone activity, monitored as the ability to suppress aggregation of thermally denatured citrate synthase [7]. These data suggest that the chaperone activity of SurA, which facilitates membrane protein folding, resides predominantly, if not exclusively, in the N and C domains and that the PPIase activity is either not essential for this activity or is complemented by other PPIases in the periplasm.

In contrast to molecular chaperones in the cytoplasm, which utilize ATP in their activities (e.g., DnaK, GroEL, ClpX, ClpA, and HslU), periplasm-resident SurA functions in the absence of ATP or other accessory sources of chemical energy. Also in contrast to molecular chaperones that facilitate correct folding of cytoplasmic proteins, there is little information on the structures or mechanisms of chaperones that facilitate membrane protein folding. Notably, the N and C domains of SurA, which appear to be sufficient for its membrane protein-folding activity, show no sequence similarity to proteins of known structure, so that inferences of SurA structure cannot be made from currently available data. In this context, we have solved the crystallographic structure of the *E. coli* SurA protein. We find that it has a bipartite domain structure and that one domain has a novel protein fold that may be imperative for its chaperone activity.

Results

The crystallographic structure of the *E. coli* SurA protein has been solved using multiwavelength anomalous dispersion (MAD) phasing with data from crystals of selenomethionine (SeMet)-labeled protein (Table 1) and refined with data to 3.0 Å resolution from a crystal of native protein (Table 2). The crystal has four independent SurA monomers per asymmetric unit. Within structural domains the molecular structure is well defined; polypep-

Key words: survival protein A (SurA); periplasmic molecular chaperone; outer membrane protein folding; gram-negative bacteria; X-ray crystallography; peptidyl-prolyl isomerase

¹Correspondence: dave.mckay@stanford.edu

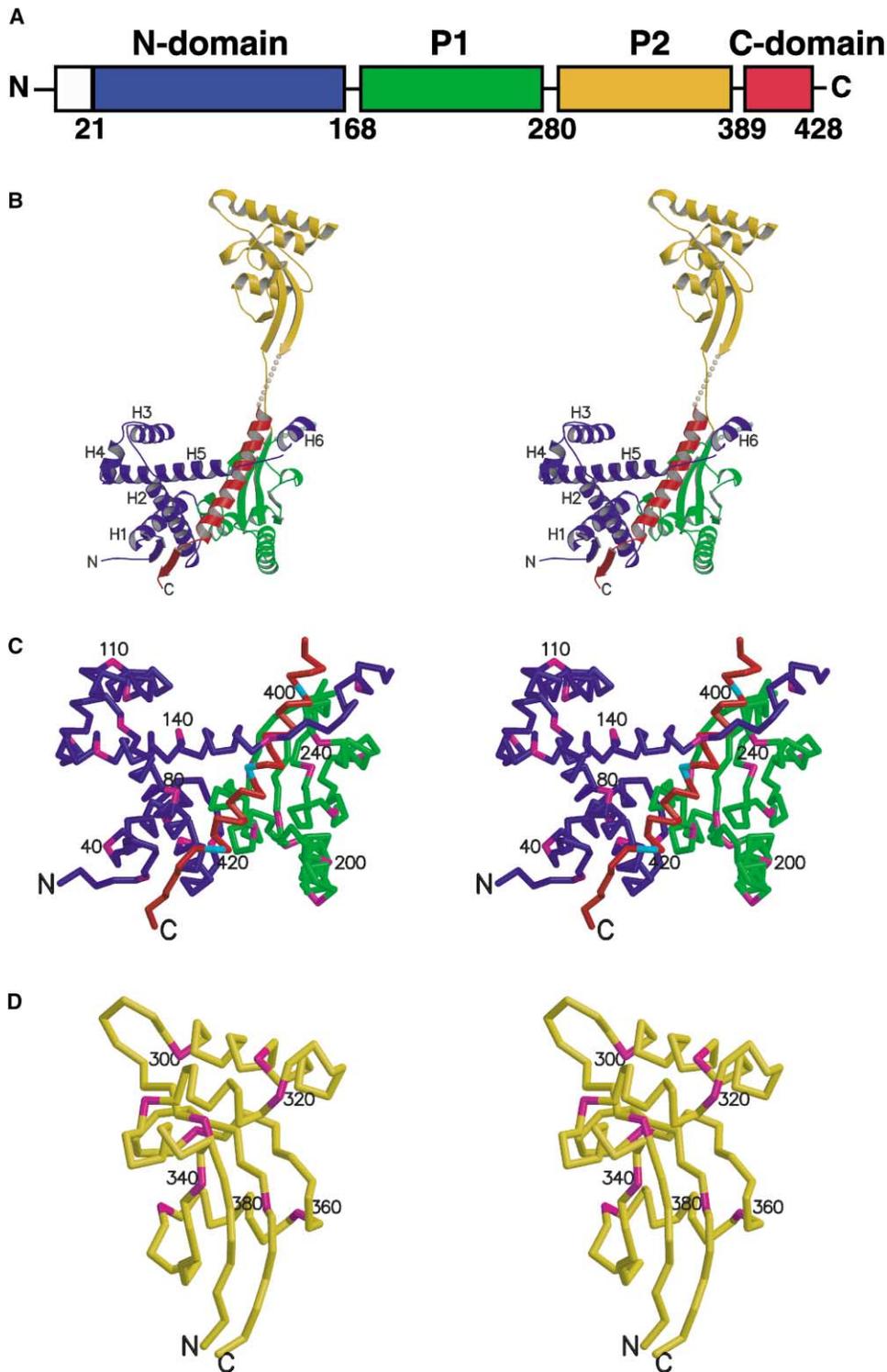


Figure 1. Structure of SurA

(A) Schematic diagram of the modular sequence domains of SurA. Leader sequence, white; N domain, blue; P1 PPIase domain, green; P2 domain, gold; C domain, green. Amino acid numbers shown below the diagram correspond to the beginning of the N domain, the middle of linkers between successive domains, and the end of the C domain, respectively.

(B) Stereo ribbon drawing of SurA protomer A, with the same color-coding convention as in (A). Helices of the N domain are numbered. Polypeptide connections that could not be traced are indicated with dotted lines.

(C) Alpha carbon trace of core module, in approximately same orientation as in (B). Every 10th α carbon is highlighted (magenta in N and P1 domains; cyan in C domain); selected residues are numbered.

(D) Alpha carbon trace of P2 domain, in approximately same orientation as in (B). Every α carbon is highlighted in magenta; selected residues are numbered.

Table 1. Data Collection and Phasing Statistics for SeMet-SurA

Data Statistics									
Wavelength (Å)	Resolution (highest shell) (Å)		Completeness (%)	R_{sym}^a	f'^b	f''^b			
$\lambda 1 = 0.97922$ (edge)	30–3.55 (3.63–3.55)		0.920 (0.936)	0.097 (0.292)	–8.87	–11.45			
$\lambda 2 = 0.97913$ (peak)	30–3.50 (3.58–3.50)		0.920 (0.934)	0.103 (0.330)	–7.08	–11.08			
$\lambda 3 = 0.91162$ (remote)	30–3.50 (3.58–3.50)		0.917 (0.927)	0.082 (0.313)	0.00	–6.31			
Diffraction Ratios and Phasing Statistics									
Wavelength (Å)	Anomalous Diffraction Ratios ^c			Phasing Power ^d					
	$\lambda 1$	$\lambda 2$	$\lambda 3$	(+) Friedel Mate	(–) Friedel Mate				
$\lambda 1$	0.086	0.035	0.067	1.78	2.49				
$\lambda 2$		0.090	0.060	1.52	2.30				
$\lambda 3$			0.068	reference	1.30				
Figure of Merit <m>									
Resolution (Å)	30.0–7.54	7.54–6.00	6.00–5.25	5.25–4.77	4.77–4.43	4.43–4.17	4.17–3.96	3.96–3.79	overall
<m>	0.765	0.781	0.718	0.666	0.625	0.5856	0.513	0.466	0.643

^a $R_{\text{sym}} = \sum |I_{\text{hkl}} - \langle I_{\text{hkl}} \rangle| / \sum \langle I_{\text{hkl}} \rangle$, where I_{hkl} is the single value of measured intensity of hkl reflection and $\langle I_{\text{hkl}} \rangle$ is the mean of all measured value intensity of hkl reflection. Bijvoet measurements were treated as independent reflections for the MAD phasing data sets.

^b Values of f' and f'' were initially estimated from an EXAFS scan and refined in CNS.

^c Anomalous diffraction ratio values equal $\langle \Delta|F| \rangle^2 / \langle |F|^2 \rangle^{1/2}$, where $\Delta|F|$ is the dispersive (off-diagonal element) or Bijvoet (diagonal element) difference, computed between 30.0 and 3.50 Å resolution.

^d Phasing Power equals $\langle |F_{\text{a}}| \rangle / E$, where $\langle |F_{\text{a}}| \rangle$ is the rms structure factor amplitude for anomalous scatterers and E is the estimated lack of closure error. Phasing power is listed for each lack of closure expression between the reference data set ([+]Friedel mate at $\lambda 3$) and the (+) or (–) Friedel set at each wavelength. Phasing powers were calculated using all data between 30.0 and 3.5 Å.

tide linkers between domains are poorly ordered and, in many cases, could not be traced unambiguously (as described in detail in Experimental Procedures). The backbone conformation of a single residue, Asn33, lies

in the disallowed region of the Ramachandran diagram; the conformation of this residue is similar in all four SurA protomers and is a reproducible result of refinement, regardless of the starting conformation. The backbone conformations of all other residues lie in allowed regions of the Ramachandran diagram. The average coordinate uncertainty, estimated by the crossvalidated Luzzati method, is 0.50 Å [8]; the average B factor for the structure is 68 Å².

Table 2. Data Collection and Refinement Statistics for Native SurA

Data Collection	
Wavelength (Å)	1.033
Resolution range (last shell) (Å)	30.0–3.00 (3.05–3.00)
Observations (total/unique)	217,260/46,785
Completeness (%)	92.9 (85.2)
Average I/σ	16.2 (3.9)
R_{sym}^a	0.077 (0.348)
Refinement	
Resolution range (last shell) (Å)	30.0–3.00 (3.11–3.00)
R_{cryst}^b	0.228 (0.347)
R_{free}^b	0.283 (0.417)
Number of reflections (working set)	38,116
Number of reflections (test set)	3,753
Number of protein atoms	11,920
Average B value, main chain atoms (Å ²)	63.6
Average B value, all protein atoms (Å ²)	67.5
Rmsd bond length (Å)	0.008
Rmsd angles (°)	1.38

Data collection statistics were computed with SCALEPACK [13], and refinement statistics were computed with CNS [14], as described in Experimental Procedures.

^a $R_{\text{sym}} = \sum |I_{\text{hkl}} - \langle I_{\text{hkl}} \rangle| / \sum \langle I_{\text{hkl}} \rangle$, where I_{hkl} is the single value of measured intensity of hkl reflection, and $\langle I_{\text{hkl}} \rangle$ is the mean of all measured value intensity of hkl reflection.

^b $R_{\text{cryst}} = \sum |F_{\text{obs}} - F_{\text{calc}}| / \sum F_{\text{obs}}$, where F_{obs} is the observed structure factor amplitude and F_{calc} is the structure factor calculated from model. R_{free} is computed in the same manner as is R_{cryst} , with the test set of reflections.

Structure of the SurA Protomer

The four segments that are described in the *E. coli* SurA sequence as separate domains (N, P1, P2, and C; Figure 1A) reveal themselves in the SurA tertiary structure in two distinct modules, the larger of which includes the N, P1, and C segments (hereafter referred to as the “core module,” or “core domain”) and the smaller of which is the P2 domain, which is connected to the core module by two extended segments of polypeptide ~25–30 Å in length (Figure 1). Consequently, SurA is an “asymmetric dumbbell.” Within the core module, the N domain begins with a pair of short antiparallel β strands followed by six α helices. Following that, a short linker leads to the P1 domain, which shares the topology of the parvulin-related family of PPIases that was first described in the structures of human Pin1 [9] and Par14 [10] proteins. The C domain consists of a long α helix (residues 396–422), sandwiched between the N and P1 domains, followed by a short β strand (residues 423–426) that runs antiparallel to the initial β strand of the N domain. The surface of the helix of the C domain is almost entirely buried, with approximately 60% (~2400 Å²) of surface area interfacing with the N domain and 40% in contact with the P1 domain; the helix is an integral element of the

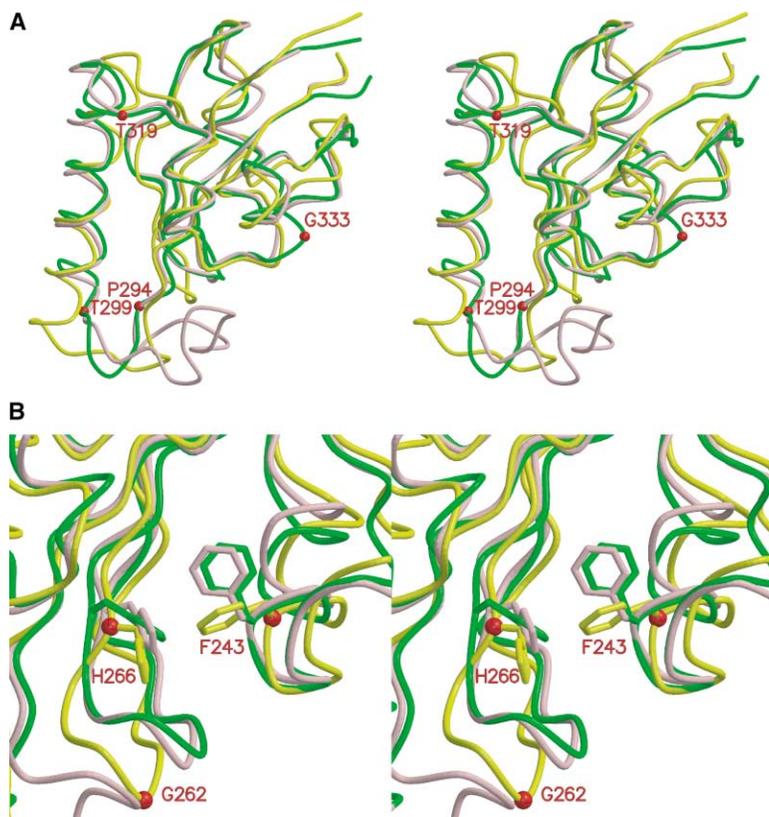


Figure 2. Superposition of P1, P2, and Human Pin1 PPLase Domains

P1, yellow; P2, green; Pin1, gray.

(A) Tube drawing of α carbon backbone of complete domain.

(B) Close-up view of active site region, showing backbone and selected side chains.

core module. There is an extended crevice walled by the helices of the N domain, as well as by the P1 and C domains. A search of the Protein Data Bank for structures that are similar to the N domain, with the Dali server (<http://www2.ebi.ac.uk/dali>), yielded no convincingly positive hits; the N domain (or, alternatively, the N and C domains) apparently forms a unique fold.

The P2 domain also has a parvulin fold; the P1 and P2 domains share 32% sequence identity with each other; their C_{α} backbones superimpose with a root-mean-square deviation (rmsd) in position of 1.5 Å. The P2 domain is a closer homolog of Pin1 than is the P1 domain, having both a higher sequence identity than P1 with the Pin1 protein (36% versus 23%) and a closer structural superposition (1.0 Å versus 1.5 Å rmsd of C_{α} positions; Protein Data Bank 1PIN coordinates were used for Pin1) (Figure 2). There are three positions at which P1 and P2 differ by insertions/deletions: (1) an extended loop following the first β strand of the domain; this loop is larger in P1 (residues 184–191) than in P2 (residues 295–299); Pin1 has an extended peptide segment at this position (residues 65–81) that contributes ligands for phosphate binding near the active site; (2) a deletion of a single residue in P1 that aligns with Thr319 in P2 and Glu101 in Pin1; this site is remote from the active site; (3) an insertion of a single residue, Gly333, in P2 relative to P1 and Pin1; this residue borders the active site cleft.

Within the core module, the catalytic cleft of P1 is partially occluded by the helix of the C domain. Further, interactions with the C helix displace the β strands at the top of the active site cleft, as compared with their

conformations in P2 and Pin1. This propagates conformational alterations within the substrate binding region of P1, including displacement of the side chains of Phe243 and the catalytic His266 residue from their conformations in P2 and Pin1. The combination of occlusion and distortion of the active site cleft of P1 by the C domain helix may rationalize the apparent lack of PPLase activity of this domain in full-length SurA protein. It does not, however, provide an obvious explanation for the lack of PPLase activity in the isolated P1 domain [4, 7].

Crystal Packing and the Differences in Conformation between Protomers

There are four independent SurA protomers in the crystallographic asymmetric unit (denoted A–D). There are significant differences in conformation between the protomers, which can most easily be described in the context of the crystal packing that gives rise to them. We observe two predominant sets of crystal-packing interactions: those between core modules and those between P2 domains. First, the core modules pack with an approximate noncrystallographic 2-fold screw axis relating successive modules; this is accomplished by helix 6 of the N domain of one molecule binding in an extended cleft in the core domain of the adjacent molecule (Figure 3). Second, P2 domains of two pairs of molecules related by a crystallographic 2-fold axis cluster to form a “crystal-packing tetramer”; in one case, the P2 domains of two “B” protomers pack with P2 domains of two “A” protomers, but in such a manner that the B molecule domains cluster closer to each other than the A molecule domains; the minimum C_{α} - C_{α} dis-

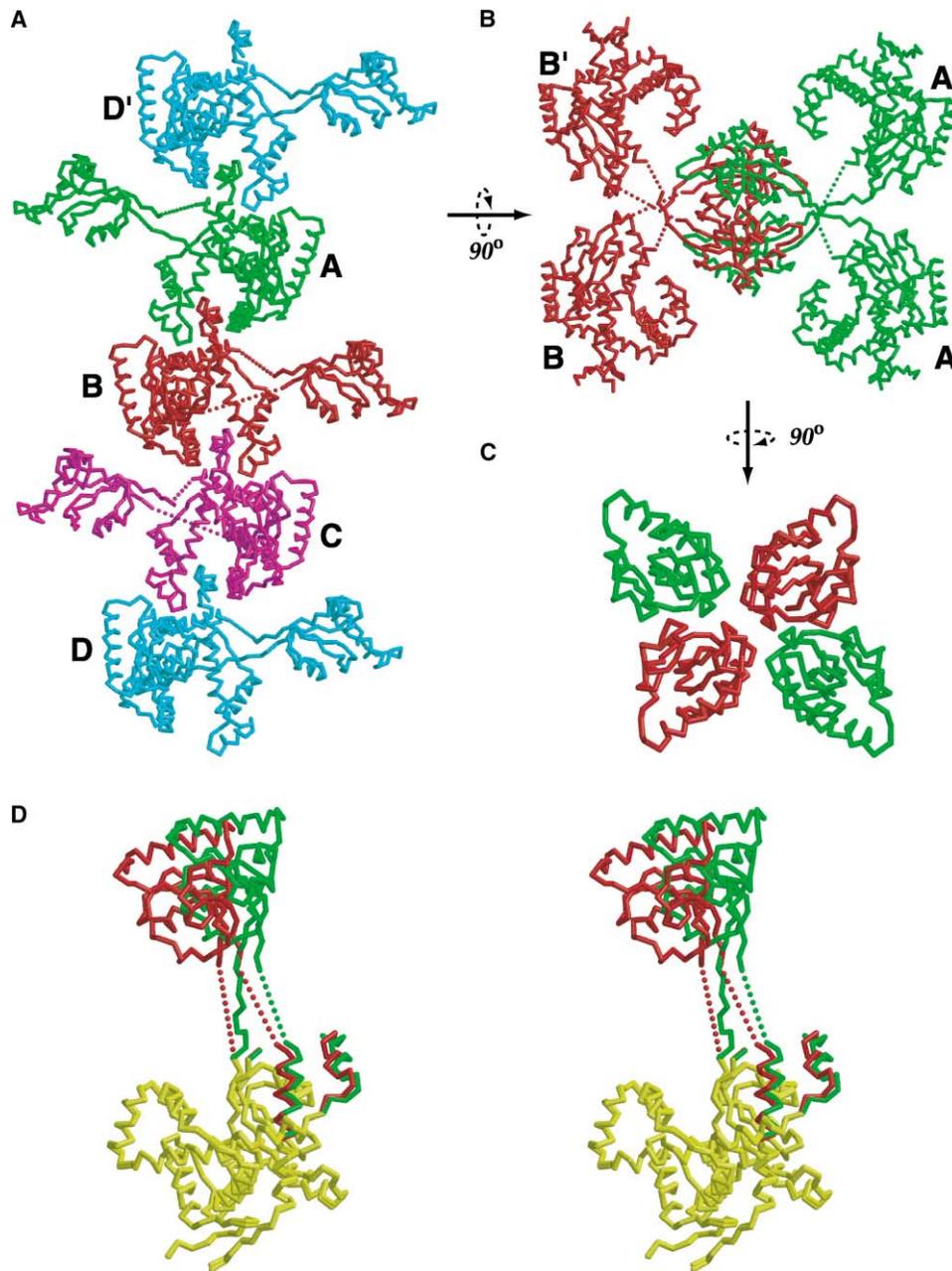


Figure 3. Crystal Packing

Alpha carbon backbone traces of SurA protomers are shown; dotted lines indicate the connections between the core domains and P2 domains that could not be traced.

(A) Interaction of core modules. The view is approximately perpendicular to the crystallographic a-c plane. Four crystallographically independent protomers, plus a fifth related by translational symmetry (top and bottom protomers D' and D), are shown.

(B) Interaction of P2 domains. The view is approximately perpendicular to the crystallographic b-c plane. Two A protomers (green) and two B protomers (red) are shown; a crystallographic 2-fold axis runs horizontally in the plane of the figure.

(C) Packing of four P2 domains from (B). The view is perpendicular to the crystallographic a-b plane.

(D) Superposition of A and B protomers, colored green (A) and red (B) where the structures differ and yellow where they are essentially the same.

tance between the P2 domains of the two B molecules is $\sim 6\text{--}7$ Å, while the minimum $C_{\alpha}\text{--}C_{\alpha}$ distance between P2 domains of A protomers is $\sim 15\text{--}16$ Å. Also, the P2 domains of pairs of "C" and "D" protomers cluster to form a similar tetramer, with the domains of the C pro-

tomers close to each other, in the same manner as the P2 domains of the B protomers, and the domains of the D protomers more-distantly separated. As a consequence of the crystal packing, the A and D protomers are similar to each other in conformation, as are the B

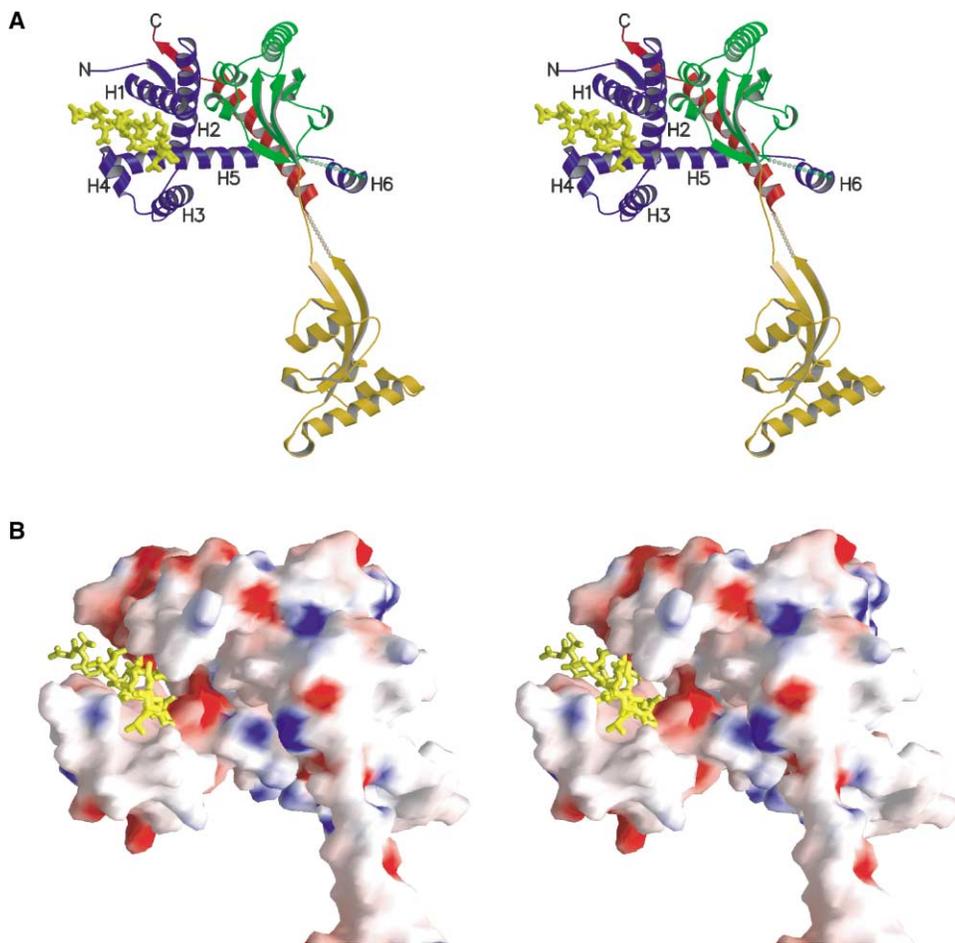


Figure 4. Intermolecular Peptide Binding Interaction

(A) Ribbon drawing showing the peptide from a neighbor molecule binding to crevice in the core domain. The color-coding and labeling of helices of the N domain are the same as in Figure 1.

(B) Surface and electrostatic potential of the core domain, in approximately the same orientation as in (A), computed with the program GRASP [25].

and C protomers, but there are major differences between the A/D conformations and the B/C conformations. The most apparent difference is in the orientation of the P2 domain with respect to the core module; in the B and C protomers, it is “swung” approximately 8° – 10° from its placement in the A and D protomers. This does not imply, however, that SurA consists of two static domains connected by a flexible linker. We also see conformational differences within the core module, which can be described as a rotation of both the C-terminal helix and helix 6 of the N domain relative to the rest of the module; these two segments of the core shift in concert with, and in the same angular direction as, the P2 domain.

Candidate Polypeptide Binding Site Suggested by Crystal Packing

The crystallographic packing interactions between adjacent core modules reveals a candidate peptide binding channel of potential relevance to the chaperone activity

of SurA. The segment of peptide (residues 153–164) that tethers a core module to its neighbor binds as an α helix approximately 15 Å in length. It binds in a crevice formed by the N domain, the helices of which envelop the peptide on three sides (Figure 4). On one side of the peptide, helices H1 and H2 of the N domain pack with the P1 and C domains, as if to form a stable “wall”; on the other side, helices H3 and H4 form a “flap” that is unconstrained by intramolecular packing; helix H5 extends under the peptide and can be thought of as a “floor” of the crevice. Approximately 1700 Å² of peptide surface area is buried in the binding pocket.

The channel in which the peptide from the adjacent molecule binds extends ~ 50 Å, running along the floor helix of the N domain to the P1 domain and the helix of the C domain. Hence, although the segment of peptide that is bound in the crystal only spans ~ 15 Å, significantly longer segments of peptide could potentially be accommodated. The flap of the N domain has a nonpolar surface patch facing the interior of the crevice, including

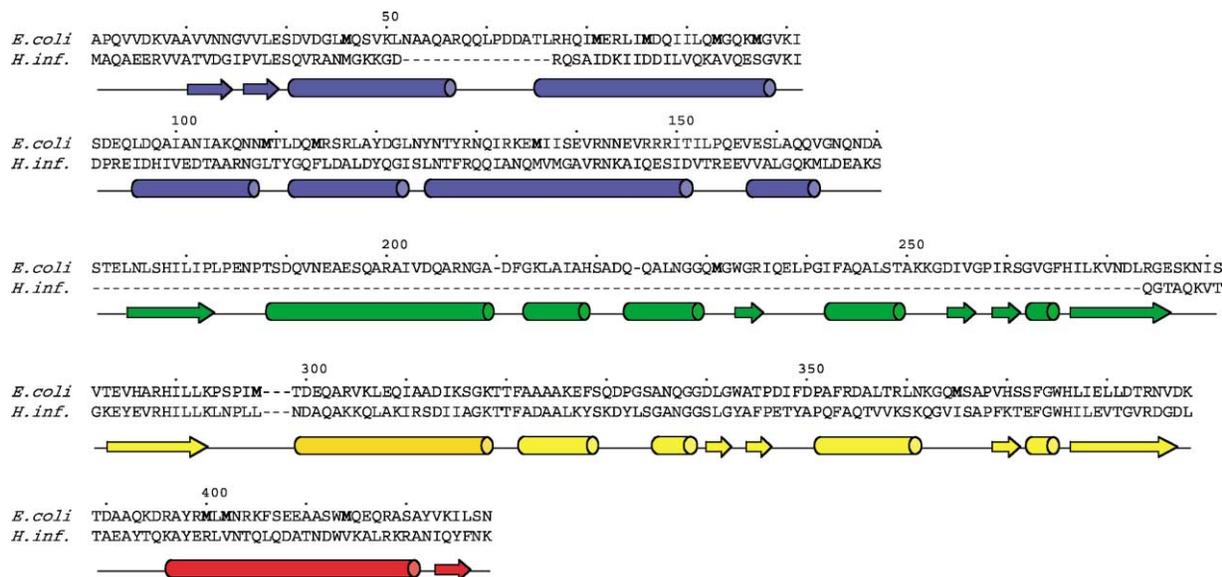


Figure 5. Sequence and Secondary Structure of Mature SurA

Color-coding is the same as in Figure 1. Methionines, whose SeMet analogs were used in placing the sequence in the model, are highlighted. Vertical alignment of the two parvulin domain sequences corresponds to their structural alignment. The alignment of the *H. influenzae* SurA sequence with that of *E. coli*, based on the alignment of available SurA homologs (data not shown), is included.

a hydrophobic pocket that binds the side chain of Leu153 of the peptide; otherwise, the channel displays no unusual surface characteristics.

Discussion

It has been shown that the combination of the N and C domains of SurA is both necessary and, to a large extent, sufficient for in vivo complementation of SurA activity, as well as for in vitro chaperone activity, assayed as suppression of aggregation of heat-denatured citrate synthase [7]. The structure clarifies the basis for this observation; the N and C domains are entwined in a core module, which, presumably, is responsible for the chaperone activity of SurA. The C domain is an integral part of the module and appears to be indispensable for proper folding. The P1 domain is also part of the core module of *E. coli* SurA but, to a large extent, appears to be dispensable. The N and C domains alone, expressed as a fusion in vivo, complement wild-type *surA* activity [7]; further, expressed as a recombinant construct in *E. coli*, the fusion is stable and can be purified to homogeneity (E.B. and D.B.M., unpublished data); the N and C domains appear to form a stable structure in absence of P1. In this context, it is notable that (1) some SurA homologs (e.g., those of *Haemophilus influenzae* [11] and *Pasteurella multocida* [12]) have only a single PPIase domain, (2) this single PPIase domain shares greater sequence similarity with the P2 domain of *E. coli* SurA than with P1, suggesting that it may also be a peripheral domain, and (3) the homologs with a single PPIase domain have a second deletion within their sequences, corresponding to the region of the N domain that contacts the P1 domain in *E. coli* SurA (Figure 5). These data inspire the conjecture that there may be

natural variants of SurA having core modules constituted of N and C domains alone, while retaining a similar peripheral PPIase domain. If so, it would reinforce the suggestion that the N and C domains alone can form a standalone functional module.

The extended crevice within the core module is suggestive of a peptide binding channel, wherein segments of polypeptide of a target protein could be bound and released during a folding transition. The intermolecular crystal-packing interaction, wherein a segment of peptide of a neighbor molecule is bound in the channel, endorses this suggestion. Ironically, the bound peptide segment is in an α -helical conformation, while the predominant mature secondary structure of the membrane proteins with which SurA is known to interact is β strand; the structure presented here, by itself, does not immediately resolve this paradox.

The function of the P2 domain and its satellite position in the SurA molecule also raise a perplexing dilemma. When it is deleted or when its PPIase activity is abolished by mutagenesis, the resulting derivative still complements wild-type SurA activity in several assays [7]. However, the suggestion that it is completely dispensable is counterintuitive, when considered in the context of the relatively high conservation of the SurA modular structure at the sequence level. Several possibilities, in addition to dispensability, may be suggested for its function. For example, (1) after the core module binds a peptide, the P2 domain may condense onto it, thereby “trapping” the peptide; (2) the P2 domain may provide a second peptide binding site “at a distance” that works in concert with peptide binding by the core module, so that two disparate segments of unfolded polypeptide of a target could be bound and stabilized simultaneously; (3) it may assert a PPIase activity synergistically with the core module—the core module might recruit and

tether a substrate by binding one segment of a polypeptide, allowing the P2 domain to scan for proline residues elsewhere in the polypeptide. It is notable that the PPIase activity of the P2 domain on short peptide substrates is approximately equal to that of full-length SurA, while the activity on a model RNase T1 substrate only approaches that of SurA in constructs where the P2 domain is linked to at least a fragment of the core module, such as the P1 domain or the N and C domains [7]. These data suggest a bifunctional mechanism whereby the core module (or subfragment thereof) would select and bind a target polypeptide in order to enhance the efficiency with which the P2 domain could assert a PPIase activity on it.

Biological Implications

The SurA protein facilitates correct folding of outer membrane proteins in the periplasm of gram-negative bacteria [4]. It accomplishes this in the absence of ATP or other sources of chemical energy; the mechanism by which this is done is not known. Although the SurA sequence includes two PPIase domains, the PPIase activity is apparently dispensable for SurA function in vivo (or, at a minimum, is encoded by redundant PPIase genes) [7]; SurA has an alternative chaperone mechanism that is largely encoded in the non-PPIase domains of its sequence.

SurA has a bipartite structure, with a core module that includes the N, P1, and C domains of the sequence and a satellite P2 domain. The structure reveals a novel protein fold within the core module and clarifies why the C domain, which is an integral part of the core module structure, is essential for activity. The core module has an extended crevice that is suggestive of a peptide binding channel; crystal-packing interactions, in which a segment of peptide from a neighbor molecule is bound in this crevice, reinforces this suggestion. The structure suggests a chaperone mechanism whereby extended segments of polypeptide could be sequentially or repetitively sequestered by SurA and then released, possibly to suppress aggregation or unproductive digressions from a folding pathway. The binding specificity of the core module, with respect to both polypeptide sequence and secondary structure, is still an unanswered question, as is the participation (or lack thereof) of the peripheral PPIase domain in chaperone activity. The SurA structure now provides a framework for addressing these questions.

Experimental Procedures

Protein Expression, Purification, and Crystallization

The *surA* gene was PCR-amplified from genomic DNA of *E. coli* K-12 strain MG1655, and the fragment encoding mature SurA protein (amino acid residues 21–428; SwissProt P21202) was cloned into the TYB1 plasmid of the IMPACT expression system (New England Biolabs) by standard molecular biology techniques. This plasmid was used to express protein in *E. coli* BL21(DE3); cells were grown at 37°C in Luria-Bertrani (LB) media supplemented with 100mg/l ampicillin to a cell density corresponding to $A_{600} = 0.6$. SurA expression was induced at this point by the addition of 0.4 mM isopropyl-β-D-thio-galactoside (IPTG), and cells were grown for another 3 hr at 25°C. Cells were then harvested by centrifugation, resuspended in 20mM Tris (pH 8.0), 500 mM NaCl, and 2 mM EDTA (column

buffer), and either frozen or processed immediately, as follows. Cells were supplemented with PMSF, lysed by sonication, and centrifuged at $20,000 \times g$ for 40 min. The supernatant was applied to a 10 ml chitin beads column (New England Biolabs) preequilibrated with column buffer. The column was washed extensively (50–100 column volumes) with column buffer; then, 15 ml of column buffer supplemented with 50 mM DDT was applied, and the column was incubated at room temperature for 16 hr. SurA was eluted with 10 ml of 10 mM MOPS (pH 7.0) and concentrated to ~20 mg/ml. Typical protein yield from this protocol was 2–5 mg SurA per liter of cell culture.

To prepare selenomethionine (SeMet)-labeled SurA, 2 liters of BL21(DE3) cells with the expression plasmid were grown in LB supplemented with 100 mg/l ampicillin at 37°C to a cell density of $A_{600} = 1.0$. Cells were then harvested and resuspended in 2 liters of M9 salts. After 15 min of incubation, the culture was supplemented with 8 g glucose, 2 mg thiamine, 100 mg of D-lysine, D-threonine, D-valine, D-phenylalanine, D-isoleucine, and D-leucine, 200 mg of D,L-selenomethionine, plus MgSO₄ and CaCl₂ to final concentrations of 2 mM and 0.1 mM, respectively. Protein expression was induced 15 min later by the addition of 0.4 mM IPTG, the culture was grown for 7 hr at 25°C, and SeMet-SurA was purified by the protocol described above.

Two crystal forms of SurA grew in hanging drops from 0.75 M (NH₄)₂SO₄ and 100 mM sodium citrate (pH 5.6). The predominant crystal form, hexagonal bipyramids, with dimensions of ~0.25 mm × 0.25 mm × 0.50 mm, grew readily at 18°C and 30°C but diffracted poorly. A second form grew under identical conditions as thin plates, with maximum thickness of ~10–15 μm; these crystals, which grew optimally at 30°C, diffracted to ~3.0 Å resolution and were utilized for the structure determination. SeMet-SurA crystals were obtained by microseeding crystallizations with crystals of the second form.

Data Collection

For data collection, native and SeMet-labeled crystals were transferred to a stabilization solution of 1.2 M (NH₄)₂SO₄ and 100 mM sodium citrate (pH 5.6), adapted stepwise to a cryoprotectant consisting of the stabilization solution plus 30% ethylene glycol in six equal increments of ethylene glycol concentration, and flash-frozen in a stream of nitrogen gas at ~100 K. Candidate heavy-atom-derivative crystals were soaked in solutions consisting of the stabilization solution plus heavy atom; a heavy-atom compound that proved useful for the structure determination was terpyridyl platinum chloride (Terpy-Pt) at 5 mM concentration. Native crystals are orthorhombic, space group F222, $a = 158.82 \text{ \AA}$, $b = 223.41 \text{ \AA}$, and $c = 279.72 \text{ \AA}$, with four molecules per asymmetric unit. SeMet-SurA crystals have slightly different cell parameters: $a = 158.40 \text{ \AA}$, $b = 222.92 \text{ \AA}$, and $c = 280.31 \text{ \AA}$. The crystal on which the final native dataset was collected was first transferred to 1.2 M Li₂SO₄ before cryoprotection.

Diffraction data were collected on beamlines (BL) 9-1, 9-2, and 11-1 of the Stanford Synchrotron Radiation Laboratory (SSRL) and BL 5.0.2 of the Lawrence Berkeley Advanced Light Source Laboratory (ALS). Datasets that ultimately contributed to the structure determination were as follows: (1) native data to 3.0 Å resolution, collected on SSRL BL 11-1 at $\lambda = 0.965 \text{ \AA}$; (2) data on a Terpy-Pt heavy-atom derivative, collected on SSRL BL 9-1 at $\lambda = 0.971 \text{ \AA}$; (3) three-wavelength data on SeMet-labeled crystals, collected on SSRL BL 9-2 at $\lambda = 0.979126 \text{ \AA}$, 0.979222 \AA , and 0.911618 \AA for selenium absorption peak, dispersive edge, and remote energies, respectively; (4) two single-wavelength datasets on SeMet-labeled crystals, collected at the selenium adsorption peak to 3.5 Å and 3.65 Å resolution on SSRL BL 9-2 and ALS BL 5.0.2, respectively. All data were processed with DENZO and scaled with SCALEPACK [13].

Structure Determination and Analysis

Crystallographic calculations were carried out with the program CNS [14], versions 1.0 and 1.1. MAD and SAD datasets were locally scaled with the program SOLVE [15]. Model building was effected with the program O [16, 17]. Superposition of structures was done with programs in CCP4 [18] and the Uppsala Software Factory [19]. Molecular drawings were made with MOLSCRIPT [20]; map figures

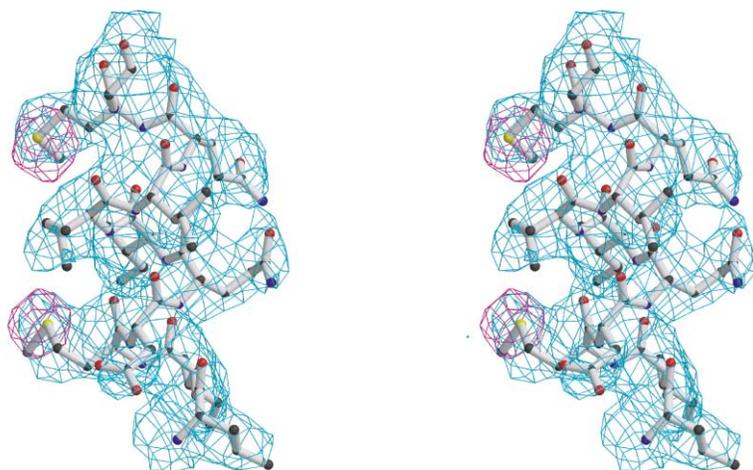


Figure 6. Electron Density Maps

Residues 74–84 of the A protomer. Cyan; $2F_o - F_c$ simulated-annealing omit map, contoured at 1.3σ ; magenta; anomalous difference Fourier map with data from SeMet-labeled SurA, contoured at 7.3σ .

were made with CONSCRIPT [21]; figures were rendered with RASTER3D [22].

Manual and automated Patterson search methods proved fruitless in attempts to find sites of either heavy atoms in isomorphous difference Pattersons or selenium sites in single-wavelength or multi-wavelength anomalous difference Pattersons. However, the use of the direct methods-based program Shake-and-Bake (SnB, version 2.1 [23, 24]) yielded solutions. Eight platinum sites were found for the Terpy-Pt heavy-atom derivative, and 22 consensus selenium sites of a possible 56 (14 per protomer times 4 protomers per asymmetric unit) were identified after multiple runs on the two independent SAD datasets with data to 3.75 or 4.0 Å. Independent solution of the Terpy-Pt derivative by difference Fourier with phases derived from the selenium sites, and vice versa, verified the correctness of the solutions. Additional selenium sites were identified with anomalous difference and log-likelihood gradient Fourier maps; 49 selenium sites were ultimately identified.

MAD phases computed with CNS [14] with the 49 selenium sites gave an overall figure of merit of 0.64 to 3.8 Å resolution; phasing statistics are summarized in Table 1. The initial experimental map was improved by solvent flipping and partial averaging: the selenium sites defined noncrystallographic symmetry operators between the four protomers in the asymmetric unit for a large, globular domain (subsequently found to be the large domain or “core module” of SurA, accounting for 63% of the residues); these were used for averaging this fragment of the molecule. After averaging and solvent modification, the polypeptide backbone of much of the large domain was traceable because of its high fractional content of α helix. Primary structure assignment was facilitated by the constraint of placing methionine residues at the selenium sites, 11 or 12 of which were identified for each large domain (Figure 6). Two parvulin-like domains per protomer were also identified; tracing the connectivity of these domains was facilitated by the similarity with two parvulin homologs of known structure, Pin1 (Protein Data Bank 1PIN) and hPar14 (Protein Data Bank 1EQ3 and 1FJD). The two parvulin domains of SurA, which show about 30% sequence identity, were differentiated by the differences in position of their (seleno)methionine residues.

The model was refined with CNS version 1.1 [14] against native data to 3.0 Å resolution in cycles of simulated annealing and group B factor refinement alternated with manual rebuilding of the model. Noncrystallographic symmetry restraints were implemented when it became apparent that the four protomers in the asymmetric unit segregated into two pairs with similar structure, but with substantial differences between pairs (in the nomenclature of the coordinates, the A and D protomers are similar and the B and C protomers are similar). Each protomer has a large domain and a small domain connected by an extended linker; between similar protomers of each pair, the large domains were restrained toward identical structure, and, separately, the small domains were restrained toward identical structure; the restraint parameter was set equal to $30 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$.

In the later stages of refinement, one strand of the polypeptide linker for both the A and D protomers became traceable; this region was also then restrained toward identical conformation for these two protomers. The mature SurA protein has 408 native residues plus a carboxy-terminal Gly residue added by the intein expression system; the final crystallographic model includes residues 388 of protomer A, 377 of protomer B, 376 of protomer C, and 389 of protomer D. The first five amino-terminal residues, the last carboxy-terminal residue of the native protein, and several segments in apparently flexible linkers that connect domains could not be traced unambiguously. Structure quality assessment with the program PROCHECK indicates that >87% of the residues are in the most favorable regions and that 0.3% of the residues are in the disallowed regions of the Ramachandran plot. Refinement statistics are summarized in Table 2.

Acknowledgments

This work was supported by grant NIH-GM39928 to D.B.M. from the National Institutes of Health (NIH). Portions of this research were carried out at the Stanford Synchrotron Radiation Laboratory (SSRL), a national user facility operated by Stanford University on behalf of the U.S. Department of Energy (DOE) Office of Basic Energy Sciences, and the Advanced Light Source (ALS). The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research and by the NIH National Center for Research Resources, Biomedical Technology Program, and National Institute of General Medical Sciences; A.L.S. is supported by the DOE.

Received: May 29, 2002

Revised: July 17, 2002

Accepted: July 18, 2002

References

1. Tormo, A., Almiron, M., and Kolter, R. (1990). *surA*, an *Escherichia coli* gene essential for survival in stationary phase. *J. Bacteriol.* **172**, 4339–4347.
2. Lazar, S.W., and Kolter, R. (1996). SurA assists the folding of *Escherichia coli* outer membrane proteins. *J. Bacteriol.* **178**, 1770–1773.
3. Missiakas, D., Betton, J.M., and Raina, S. (1996). New components of protein folding in extracytoplasmic compartments of *Escherichia coli* SurA, FkpA and Skp/OmpH. *Mol. Microbiol.* **21**, 871–884.
4. Rouvière, P.E., and Gross, C.A. (1996). SurA, a periplasmic protein with peptidyl-prolyl isomerase activity, participates in the assembly of outer membrane porins. *Genes Dev.* **10**, 3170–3182.
5. Dartigalongue, C., and Raina, S. (1998). A new heat-shock gene, *ppiD*, encodes a peptidyl-prolyl isomerase required for folding

- of outer membrane proteins in *Escherichia coli*. *EMBO J.* *17*, 3968–3980.
6. Rahfeld, J.U., Rucknagel, K.P., Schelbert, B., Ludwig, B., Hacker, J., Mann, K., and Fischer, J. (1994). Confirmation of the existence of a third family among peptidyl-prolyl cis/trans isomerases. Amino acid sequence and recombinant production of parvulin. *FEBS Lett.* *352*, 180–184.
 7. Behrens, S., Maier, R., de Cock, H., Schmid, F.X., and Gross, C.A. (2001). The SurA periplasmic PPIase lacking its parvulin domains functions in vivo and has chaperone activity. *EMBO J.* *20*, 285–294.
 8. Luzzati, V. (1952). Traitement statistique des erreurs dans la détermination des structures cristallines. *Acta Crystallogr. A* *5*, 802–810.
 9. Ranganathan, R., Lu, K.P., Hunter, T., and Noel, J.P. (1997). Structural and functional analysis of the mitotic rotamase Pin1 suggests substrate recognition is phosphorylation dependent. *Cell* *89*, 875–886.
 10. Sekerina, E., Rahfeld, J.U., Muller, J., Fanghanel, J., Rascher, C., Fischer, G., and Bayer, P. (2000). NMR solution structure of hPar14 reveals similarity to the peptidyl prolyl cis/trans isomerase domain of the mitotic regulator hPin1 but indicates a different functionality of the protein. *J. Mol. Biol.* *301*, 1003–1017.
 11. Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M., et al. (1995). Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* *269*, 496–512.
 12. May, B.J., Zhang, Q., Li, L.L., Paustian, M.L., Whittam, T.S., and Kapur, V. (2001). Complete genomic sequence of *Pasteurella multocida*, Pm70. *Proc. Natl. Acad. Sci. USA* *98*, 3460–3465.
 13. Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* *276*, 307–326.
 14. Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* *54*, 905–921.
 15. Terwilliger, T.C., and Berendzen, J. (1999). Automated structure solution for MIR and MAD. *Acta Crystallogr. D Biol. Crystallogr.* *55*, 849–861.
 16. Jones, A. (1978). A graphics model building and refinement system for macromolecules. *J. Appl. Crystallogr.* *11*, 268–272.
 17. Jones, T.A., Zhou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for the building of protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* *47*, 110–119.
 18. Bailey, S. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* *50*, 760–763.
 19. Kleywegt, G.J., and Jones, T.A. (1997). Detecting folding motifs and similarities in protein structures. *Methods Enzymol.* *277*, 525–545.
 20. Kraulis, P. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* *24*, 946–950.
 21. Lawrence, M.C., and Bourke, P. (2000). CONSCRIPT: a program for generating electron density isosurfaces from Fourier syntheses in protein crystallography. *J. Appl. Crystallogr.* *33*, 990–991.
 22. Merritt, E.A., and Bacon, D.J. (1997). Raster3D. Photorealistic molecular graphics. *Methods Enzymol.* *277*, 505–524.
 23. Miller, R., Gallo, S.M., Khalak, H.G., and Weeks, C.M. (1994). SnB: crystal-structure determination via Shake-and-Bake. *J. Appl. Crystallogr.* *27*, 613–621.
 24. Weeks, C.M., and Miller, R. (1999). The design and implementation of SnB version 2.0. *J. Appl. Crystallogr.* *32*, 120–124.
 25. Nicholls, A., and Honig, B.J. (1991). A rapid finite-difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *J. Comput. Chem.* *12*, 435–445.

Accession Numbers

The SurA coordinates have been deposited in the Protein Data Bank under accession code 1M5Y.