

Diverse sequences are functional at the C-terminus of the *E. coli* periplasmic chaperone SurA

Qian Chai, Brent Ferrell, Meng Zhong, Xinyi Zhang,
Cui Ye and Yanan Wei¹

Department of Chemistry, University of Kentucky, 305 Chemistry-Physics
Building, Lexington, KY 40506-0055, USA

¹To whom correspondence should be addressed.
E-mail: yinan.wei@uky.edu

Received August 23, 2013; revised January 31, 2014;
accepted February 7, 2014

Edited by Gideon Schreiber

SurA is a major periplasmic molecular chaperone in *Escherichia coli* and has been shown to assist the biogenesis of several outer membrane proteins. The C-terminal fragment of SurA folds into a short β -strand, which forms a small three-stranded anti-parallel β -sheet module with the N-terminal β -hairpin. We found that the length of the C-terminal fragment, rather than its exact amino acid composition, had a big impact on SurA function. To investigate the determinant factor of the C-terminal sequence, we created a library of SurA constructs randomized in the last 10 residues. We screened the library and randomly analyzed 19 constructs that displayed SurA activity. The C-termini of these constructs shared little sequence similarity, except that β -strand-forming residues were preferentially enriched. Three SurA constructs were expressed and purified for structural characterization. Circular dichroism and fluorescence spectroscopy analyses revealed that their structures were similar to the structure of the wild-type SurA. Our results suggest that for scaffolding purpose proteins may tolerate various sequences provided certain general requirements such as hydrophobicity and secondary structure propensity are satisfied. Furthermore, the sequence tolerance of SurA at the C-terminus indicates that this area is not likely to be involved in substrate binding.

Keywords: mutation/periplasmic molecular chaperone/
random sequence library/sequence diversity

Introduction

The periplasm is the aqueous space in gram-negative bacterial cells which separates the inner membrane from the outer membrane. It is a gel-like crowded space filled with macromolecules including proteins and polysaccharides (Wulfing and Pluckthun, 1994). In prokaryotic cells, outer membrane proteins (OMPs) are synthesized by ribosomes in the cytoplasm and transported in the unfolded form across the inner membrane through translocons (Bos and Tommassen, 2004; Ruiz *et al.*, 2006; Bos *et al.*, 2007; Gold *et al.*, 2007). Before folding and inserting into the outer membrane, OMPs need to travel across the periplasmic space, which relies on the assistance of periplasmic molecular chaperones. Periplasmic molecular

chaperones are proteins that bind to and stabilize the nascent OMP to prevent misfolding and precipitation, and potentially also assist the insertion of the proteins into their final destination in the outer membrane (Missiakas *et al.*, 1996; Bernstein, 2000; Duguay and Silhavy, 2004).

Several periplasmic molecular chaperones have been identified, including DegP, Skp, FkpA and SurA (Dominique *et al.*, 1996; Rizzitello *et al.*, 2001; Mogensen and Otzen, 2005; Sklar *et al.*, 2007). SurA is the primary periplasmic molecular chaperone that facilitates the folding and assembling of many OMPs in gram-negative bacteria (Tormo *et al.*, 1990; Lazar and Kolter, 1996; Rouviere and Gross, 1996; Bitto and McKay, 2003, 2004; Hennecke *et al.*, 2005; Behrens-Kneip, 2010; Palomino *et al.*, 2011; Vertommen *et al.*, 2009). It is highly conserved in most gram-negative bacteria (Rouviere and Gross, 1996). Deletion of the *surA* gene in *Escherichia coli* leads to a decrease in outer membrane density and an increase in bacterial drug susceptibility (Dominique *et al.*, 1996; Tamae *et al.*, 2008). SurA is a small protein of 45 kDa. The structure of SurA is composed of four domains, a N-terminal domain (NT) followed by two peptidyl-prolyl isomerases domains (P1 and P2) and a long α -helix terminated with a short β -strand at the C-terminus (CT) (Fig. 1A) (Bitto and McKay, 2002; Xu *et al.*, 2007). P1 and P2 domains have been shown to be dispensable and their deletion has little effect on SurA chaperone activities (Behrens *et al.*, 2001; Justice *et al.*, 2005). The chaperone activity relies on the small N- and C-terminal domains that contain 170 residues. In an attempt to identify key amino acids that are involved in SurA–OMP interaction, we have previously conducted site-directed mutagenesis and replaced 23 conserved residues in the N- and C-terminal domains (Zhong *et al.*, 2013). None of the residues investigated appeared to be critical for substrate binding. In the process of creating SurA point mutations, we stumbled across a SurA mutant that was truncated at the CT, likely due to a frame shift randomly occurred during the whole-plasmid PCR experiment. Truncations at the CT, even of just a few residues, led to mutants with little or no activity. However, the actual amino acid composition of the C-terminal fragment was very tolerant to mutations. Here, we identified C-terminal sequences capable of complementing the function of SurA from a randomized library and characterized selected SurA constructs.

Materials and methods

Cloning, mutation, expression and purification

Escherichia coli surA-null strain BW25113 Δ *surA* was obtained from the Coli Genetic Stock Center at Yale University, where *surA* was deleted from the genomic DNA and replaced with a kanamycin-resistance cassette. The plasmid containing *surA* gene, pMal-SurA, was constructed in a previous study (Zhong *et al.*, 2013). pMal-SurA contains the full-length *surA* gene including the signal peptide. SurA expression and purification were performed as described (Zhong *et al.*, 2013).

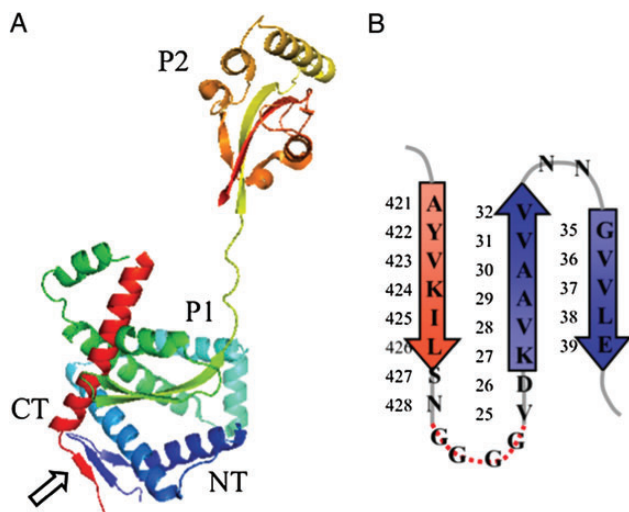


Fig. 1. (A) Structure of SurA created from 1M5Y.pdb (Bitto and McKay, 2002). Color coding is blue to red from N to C termini. (B) Peptide sequence used in the secondary structure prediction as described in the text. The C-terminal fragment (red) was linked to the N-terminal β -hairpin (blue) via a tetra-glycine linker.

To enable purification, a histag containing six histidines was inserted in-frame before the stop codon of *surA* gene using the Quikchange site-directed mutagenesis kit (Agilent Technologies) following the manufacturer's instruction. To create truncated SurA constructs, a stop codon was introduced into the respective position in the *surA* gene using the same method.

CD and fluorescence spectroscopy

Circular dichroism (CD) was performed on a JASCO J-810 spectrometer (JASCO, Inc.). Fluorescence wavelength scans were performed on a LS-55 fluorescence spectrometer (PerkinElmer, Inc.). Scans were performed at room temperature. Protein samples were dialyzed overnight into a low salt buffer (25 mM phosphate buffer, 100 mM NaCl, pH 7.5) before the experiments. Blank scans were performed with the exterior dialysis buffer and subtracted from the measured data.

SurA activity assays

Two methods were used to determine SurA activity, a drug susceptibility assay and a cell growth assay.

For the drug susceptibility assay, BW25113 Δ *surA* cells were transformed with pMal-SurA plasmids containing different SurA constructs. Following transformation, individual colonies were picked and grown to an optical density at 600 nm (OD_{600}) of 0.1–0.15. Cells were diluted 1000-fold and a 1 μ l spot was applied to Luria Bertani (LB)-agar plates containing 50 μ g/ml ampicillin and the indicated concentration of novobiocin, followed by overnight incubation at 37°C. Cell growth was examined the next morning. The lowest concentration of novobiocin that completely inhibited the growth of bacteria was recorded as the minimum inhibitory concentration (MIC) for the corresponding SurA construct.

The cell growth assay under basic conditions was conducted as described in the literature (Lazar et al., 1998). Briefly, single colonies of BW25113 Δ *surA* transformed with plasmid encoding SurA constructs were picked and cultured overnight with shaking (250 rpm) at 37°C. In the next morning, cell cultures were diluted 50-fold and cultured to log phase

($OD_{600} \sim 0.8$) before being diluted 10-fold into a basic LB medium (containing 100 mM glycine/NaOH buffer, pH 9.0). Cell growth under basic conditions was monitored by recording their OD_{600} at the indicated time into incubation.

Creation of the random peptide library

A stop codon was introduced into the *surA* gene after the codon for Arg418 to create a SurA construct with 10 amino acid deletion at the CT (denoted construct EC, and the corresponding plasmid pMal-EC). To construct the library, first a unique NheI digestion site was introduced into plasmid pMal-EC. A silent mutation was created in which the codon of Ala411 in *surA* gene in the plasmid was switched from GCA into GCT. It formed a NheI recognition site together with the S412 codon, AGC. The resultant plasmid was digested using NheI and EcoRI and purified by agarose gel electrophoresis. A high-performance liquid chromatography-purified DNA oligonucleotide library was obtained from IDT (Coralville, IO). Sequences in the 96-mer library were randomized in the middle: 5'-CGTTTGGATTAAACCCGCTAGCTGGATGCAGGAACAA CGT-30nt-TAACTGAATTCATCCAGTGTAGTCGT-3'. The NheI and EcoRI restriction sites were underlined and the stop codon bolded. The sequence leading to the random 30-mer, starting from the NheI site, translates into ASWMQEQR, corresponding to residues 411–418 of SurA. The library was amplified by polymerization chain reaction with a 20-mer primer complementary to the CT of the sequence. The PCR product was digested using NheI and EcoRI, column-purified and ligated into a similarly digested and purified EC vector. The ligation product was transformed into BW25113 Δ *surA*. The transformed cells were plated on an LB-agar plate containing 150 μ g/ml novobiocin and allowed to grow for 48 h at 37°C. The resulting colonies were picked and grown in LB media and the *surA*-containing plasmid was extracted. Following sequencing, the plasmid DNA was re-transformed into BW25113 Δ *surA* and at least three colonies for each mutant were tested for MIC using novobiocin.

To determine the portion of random sequences that could at least partially restore SurA activity, after ligation and transformation, 200 μ l of cell culture was withdrawn and subjected to 10-fold serial dilutions. Each dilution sample (100 μ l) was plated on an LB-agar plate supplemented with 100 μ g/ml ampicillin (A) or an LB-agar plate supplemented with 100 μ g/ml ampicillin and 150 μ g/ml novobiocin (A&N). The number of colonies on each plate was counted after incubation, and the ratio between the number on Plate A&N and the number on Plate A was estimated to be the portion of random sequences that could partially or fully restore SurA activity. This estimation was repeated three times using samples from different batches of digestion and ligation experiments to avoid bias.

Analysis of SurA expression by western blot

One colony was picked from each mutant and grown overnight in 15 ml of LB media at 37°C with shaking. The OD_{600} was measured and 13.5 ml of each culture was centrifuged at 5500 g for 20 min. The cell pellet was re-suspended in 8 ml of Tris-sucrose solution (20% sucrose, 30 mM Tris, 1 mM EDTA, pH 8.0) and incubated on ice with shaking for 10–20 min. The cell-containing solution was centrifuged at 8000 g for 20 min. The cell pellet was then re-suspended in a normalized volume of 5 mM $MgSO_4$ based on the OD_{600} of each culture. The resulting supernatant was analyzed using sodium dodecylsulfate-polyacrylamide gel

electrophoresis on a 10% gel. Anti-SurA western blot was conducted as described (Zhong *et al.*, 2013).

Secondary structure prediction

We used the online server of NetSurfP to predict the secondary structure of the C-terminal-randomized sequences (Petersen *et al.*, 2009). The CT fragment is part of a three-stranded anti-parallel β -sheet involving a β -sheet hairpin at the N-terminus (Fig. 1A). Therefore, in folded SurA the structural neighbor of the C-terminal fragment is the N-terminal β -sheet hairpin. To make it possible to predict the secondary structure of the C-terminal sequence in the context of the three-stranded β -sheet, we created peptides in which the N- and C-terminal sequences were linked together for the sake of computation (Fig. 1B). For example, for the wild-type SurA sequence, we connected the last eight amino acid residues of the protein (A421-N428, underlined) with the N-terminal sequence (V25–E39, italic) via a tetra-glycine linker to yield AYVKILSNGGGGVDKVAAVVNGVVLE (Fig. 1B). This peptide was used in the secondary structure prediction using the online server of NetSurfP (Petersen *et al.*, 2009). To confirm that the tetra-glycine linker did not introduce a strain that might affect the prediction, we have also tried longer linkers containing six or eight glycines and did not observe a significant effect on the calculation result. The sums of α -helix, β -strand and coil forming probabilities of residues 422–426 are listed in Table II in columns α , β and coil, respectively.

Results

The CT is critical for SurA function

The structure of SurA contains four domains, but only the N- and C-terminal domains are critical for its chaperone activity (Behrens *et al.*, 2001; Justice *et al.*, 2005). The C-terminal domain composes of the last 33 amino acid residues of the protein. In an earlier study, we have mutated 11 conserved and semi-conserved residues in the C-terminal domain but none of them has a significant effect on SurA activity (Zhong *et al.*, 2013). However, we found that the length of the C-terminal tail was critical for activity. Here, we created a series of truncation mutations in which the indicated number of residues were removed from the sequence. To examine the effect of C-terminal truncation on the function of SurA, we monitored the drug susceptibility levels of BW25113 Δ surA strains containing plasmids encoding each SurA mutant (Table I). The same strains containing plasmids encoding the wild-type SurA or the empty vector were used as the positive and negative controls, respectively. We found that the removal of as little as five residues from the CT reduced the MIC of novobiocin to a level close to that of the negative control.

Table I. MIC of novobiocin for BW25113 Δ surA strains containing SurA with the indicated number of amino acids truncated from the CT

Mutation	MIC (μ g/ml)
Wild type	320
Δ surA	40
20 residues	40
11 residues	40
10 residues (EC)	40
5 residues	40
4 residues	160
2 residues	320

Identification of C-terminal sequences that complement SurA function

The data obtained from site-directed mutagenesis indicated that SurA is a very robust protein and can accommodate many different mutations in the C-terminal fragment without compromising its function. However, the length of the C-terminal region seemed to be important. To examine whether there is any sequence preference at the C-terminal end, we created a randomized sequence library and searched for sequence compositions that support SurA function.

Toward this end, a *surA* vector (pMal-EC) was constructed in which a stop codon was inserted right after the codon of residue R418. This SurA construct lacks the last 10 residues and behaved like the negative control (Table I). Using pMal-EC as the template, a library of SurA constructs randomized in the last 10 residues were created as described in the section Materials and methods. We used a novobiocin concentration (150 μ g/ml) that inhibits the growth of BW25113 Δ surA strains containing plasmid pMal-EC, but not the strain containing plasmid pMal-SurA, to identify sequences at the CT that at least partially restore SurA activity. To test the relative frequency of mutants that could at least partially recover SurA function, the number of colonies that grew with novobiocin selection pressure was divided by the number of colonies that grew without the selection pressure. The average ratio was \sim 1.5%.

Colonies that grew on plates containing novobiocin were cultured for plasmid extraction. The plasmid was transformed back into BW25113 Δ surA competent cells, and then three colonies for each plasmid were picked from the agar plate for MIC measurement. This was done to ensure that the resistance to novobiocin was due to the plasmid and not a genomic mutation. Most colonies examined did derive their resistance from the SurA gene in the plasmid, confirming the feasibility of using colony numbers to estimate the rate of successful sequences as described above. In total, 19 colonies were randomly picked and analyzed, which displayed a range of activity in the presence of novobiocin (Table II). Their sequences were determined and are shown in Table II. Among the 19 constructs characterized, five exhibited activity close to that of the wild type SurA in the novobiocin MIC assay (marked with +++), seven was 50% active (++), and another seven was 25% active (+).

Relative surA expression levels among mutants

In order to rule out the protein expression level as a source for the observed differences in activity among the mutants, the expression levels of these mutants in BW25113 Δ surA were examined under the basal condition. The periplasmic fraction was isolated from overnight cultures of BW25113 Δ surA containing the indicated plasmid and analyzed using an anti-SurA Western blot (Fig. 2). The results indicated that the expression levels for the mutants and the wild-type SurA were not drastically different. In addition, the variation of expression levels was not correlated with the observed novobiocin resistance, further confirming that the observed difference in activity was not a result of difference in expression levels.

Effect of mutants in complementing growth defect at high pH

A phenotype of BW25113 Δ surA strain is the compromised growth at basic pH (Lazar *et al.*, 1998). As shown in Fig. 3A, while *surA* knockout strain grew well under the neutral pH, it

Table II. Sequences and properties of SurA constructs (residues 419–428). MIC of novobiocin (Nov.), growth under basic pH (Gr.) and secondary structure properties are also shown

ID	Nov.	Gr.	Sequences										ΣP_{α}^*	ΣP_{β}^*	ΔP^*	α^{**}	β^{**}	coil ^{**}
			419	420	421	422	423	424	425	426	427	428						
WT	+++	++	A	S	A	Y	V	K	I	L	S	N	5.16	6.30	1.14	0.089	3.817	1.094
M7	+++	++	A	C	A	T	V	V	W	G	G	S	4.77	6.50	1.73	0.015	3.284	1.700
M8	+++	++	L	L	A	V	V	I	M	S	V	Q	5.27	7.29	2.02	0.007	4.436	0.557
M9	+++	++	F	F	C	R	C	F	L	T	S	D	4.84	5.90	1.06	0.013	3.782	1.206
M15	+++	++	L	G	T	T	V	I	M	P	P	/	4.75	6.74	1.99	0.016	2.865	2.120
M4	+++	+	A	C	Q	L	V	I	L	S	E	R	5.61	6.41	0.80	0.036	4.096	0.869
M6	++	++	D	L	L	R	R	Y	V	S	S	S	4.12	5.46	1.34	0.209	2.466	2.326
M3	++	+	L	S	T	H	I	S	V	V	S	V	5.31	6.33	1.02	0.013	3.451	1.535
M5	++	+	R	R	A	W	F	F	L	C	H	G	5.49	6.27	0.78	0.010	3.946	1.043
M10	++	+	F	N	T	V	V	V	Y	L	S	S	5.37	7.46	2.09	0.006	4.593	0.401
M11	++	+	A	L	A	W	V	I	Y	F	C	F	5.01	7.01	2.00	0.008	4.291	0.700
M12	++	+	A	R	V	V	G	V	V	P	M	E	4.54	6.38	1.84	0.014	3.582	1.403
M16	++	+	V	S	S	K	V	S	L	C	I	R	5.11	5.63	0.52	0.093	3.351	1.555
M2	+	+	I	R	G	P	V	V	R	V	V	Q	4.80	6.47	1.67	0.024	3.625	1.351
M13	+	+	D	K	T	L	I	C	Y	E	H	C	4.25	5.67	1.42	0.105	3.565	1.330
M17	+	+	M	T	V	Y	I	R	L	T	H	S	4.56	6.21	1.65	0.009	4.150	0.841
M18	+	+	T	I	C	H	V	A	I	P	R	G	5.42	5.55	0.13	0.058	2.673	2.270
M14	+	+	Q	K	A	R	M	M	G	G	G	G	4.25	5.86	1.61	0.103	1.603	3.294
M19	+	-	A	R	I	F	V	R	L	F	G	L	5.51	6.33	0.82	0.052	4.144	0.804
M1	+	-	A	V	G	R	V	A	T	P	S	M	4.79	5.34	0.55	0.103	1.269	3.628
C1	-	-	A	S	A	S	A	E	A	A	S	N	6.67	3.89	-2.78	0.935	0.751	3.312
C2	-	-	A	S	A	E	A	E	A	A	S	N	7.41	3.43	-3.98	0.633	1.175	3.192
C3	-	-	A	S	A	E	E	E	A	A	S	N	7.49	2.72	-4.77	1.438	0.559	2.999
C4	-	-	A	S	A	E	A	E	A	E	S	N	7.49	2.72	-4.77	0.327	1.516	3.158
CS1 ^{***}			3	3	7	4	9	3	4	3	0							
CS2 ^{***}			1	2	5	2	2	2	2	1	0	0						

* P_{α} and P_{β} are cumulative α -helix and β -strand secondary structure propensity of residues 422–426 (boxed) (Chou and Fasman, 1974). ΔP is the difference between cumulative β -strand and α -helix propensity.

**Cumulative probability for α -helix (α), β -strand (β), and coil (coil) of residues 422–426 (boxed). (Petersen et al., 2009).

***Conservation score calculated using Jalview (Waterhouse et al., 2009). CS1 is the score among SurA homologs from different organisms (Behrens-Kneip, 2010). CS2 is the score among *E. coli* SurA and the 19 mutants M1–19.

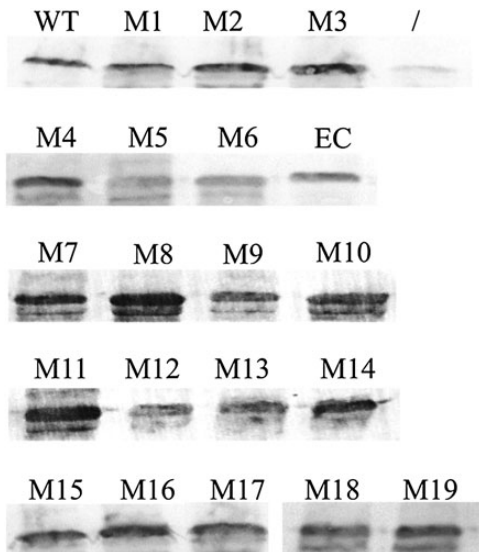


Fig. 2. Anti-SurA western blotting experiment showing that all SurA constructs had similar expression levels.

failed to grow at pH 9. We transformed plasmid-encoded SurA variants into BW25113 Δ surA and monitored their growth for 6 h (Fig. 3B). Cells transformed with plasmid encoding the wild-type SurA or empty pMal vector were used as positive and negative controls, respectively. Different SurA variants complemented the growth to different levels. While several

constructs supported the growth to a level similar to the wild-type SurA, there were also ones that had little effect in improving growth of the knockout strain. For comparison purpose, we categorized the activities into three groups: constructs that support the strain to grow to an OD₆₀₀ of at least 50% of the positive control were marked with two plus signs ‘++’. Constructs that support the growth to an OD₆₀₀ of less than three times the density of the negative control were marked by a minus sign ‘-’. Constructs behaved in between were marked with a single plus sign ‘+’. These marks are shown in Table II. Overall, the activities of the constructs determined by these two orthogonal methods correlated very well. Strains that were more sensitive to novobiocin also grew worse under basic pH conditions. The only cases of mismatch were observed for M4 and M6. While the novobiocin MIC was higher in M4, it grew worse than M6.

Characterization of selected SurA constructs

Three SurA mutants, M1, M4 and M6, were chosen for expression and structural characterization. These mutants have high, intermediate and low activities when compared with the wild-type SurA (Table II). All mutants were expressed at similar levels and could be purified similarly as the wild-type SurA.

Following expression and purification, all three SurA mutant proteins were analyzed using CD and fluorescence spectroscopy for potential structural changes. The CD plots of all three proteins superimposed well with the spectrum of the wild-type SurA, indicating a lack of change in the secondary structure

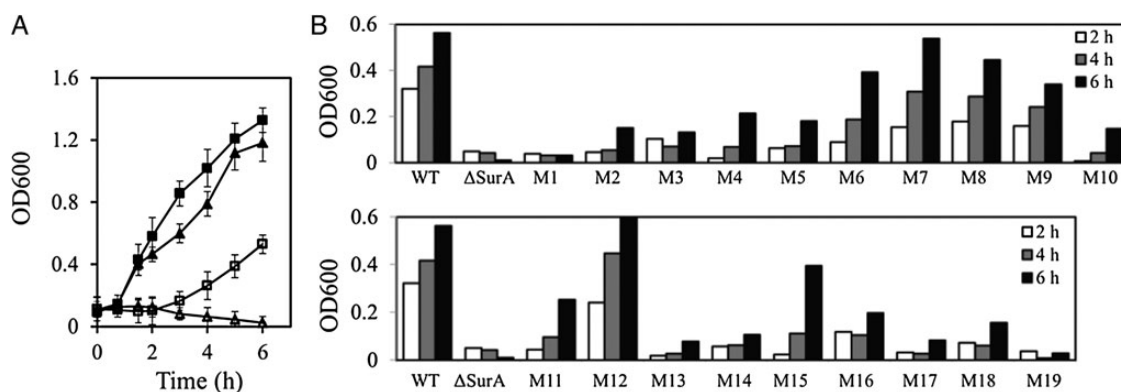


Fig. 3. SurA activity assay. (A) Growth of BW25113 (squares) and BW25113ΔsurA (triangles) under neutral (black) or basic (white) pH. (B) Growth of BW25113ΔsurA strains containing plasmid-encoded SurA constructs monitored at 2 (white), 4 (grey) or 6 (black) hours. ΔSurA is BW25113ΔsurA containing the empty vector.

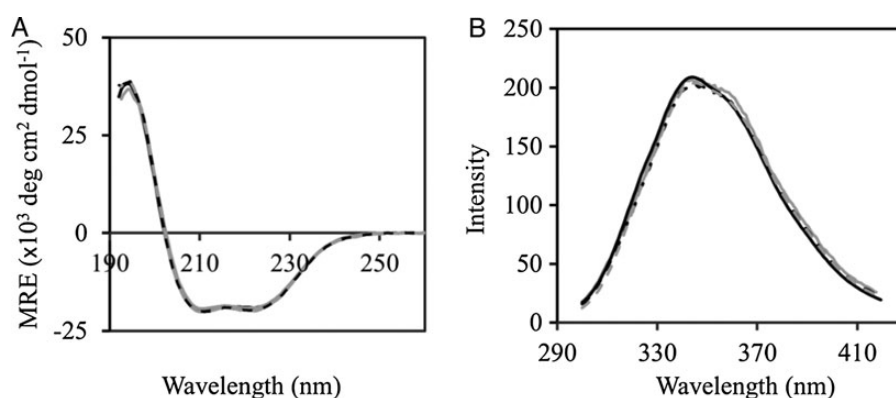


Fig. 4. Structure characterization of wild-type SurA (black continuous) and mutants M1 (black dashed), M4 (grey dashed) and M6 (grey continuous). (A) CD spectra shown in the unit of mean residue ellipticity (MRE). (B) Fluorescence emission spectra. All traces were normalized to the concentration of each protein. Excitation wavelength was 280 nm.

composition (Fig. 4A). To probe the tertiary structure, we collected intrinsic fluorescence spectra for all three mutants and compared them with the spectrum of the wild-type SurA (Fig. 4B). When excited at 280 nm, the emission spectra of all proteins tested were very similar. The small difference between the plots could be due to the difference in amino acid composition of the four proteins. For example, the wild-type and M6 SurA contain one aromatic amino acid (Tyr) in the last 10 residues, while M1 and M4 do not contain any aromatic amino acid in this stretch (Table II).

Discussion

The last 10 residues of the wild-type SurA form a structure containing a short β -strand which associates with an N-terminal β -hairpin to form an anti-parallel β -sheet (Fig. 1B). In Table II, the fragment that forms the C-terminal β -strand in the wild-type SurA (residues 422–426) is highlighted by a box. A rough estimation indicated that $\sim 1.5\%$ of all possible amino acid compositions at this deca-peptide could support at least partial SurA activity. It is apparent that a vast number of sequences could be tolerated by the protein at this region. Nineteen colonies containing SurA mutants that survived in the presence of novobiocin were randomly picked and analyzed.

Several observations could be made by comparing the sequences. First, the lack of conserved residues or a conserved pattern among the peptides was clear. We aligned the

sequences of wild-type SurA with the 19 mutants using the online server of Clustal W2 (Larkin *et al.*, 2007) and calculated the conservation scores for each position using Jalview (Waterhouse *et al.*, 2009). The scale of the conservation score ranges between 0 and 10, with 0 being the least conserved and 10 being invariable. The conservation scores for each residue are shown in Table II (CS2). In a previous review paper, Behrens-Kneip conducted sequence alignment of *E. coli* SurA with its homologs in 11 different organisms and calculated the conservation score similarly (Behrens-Kneip, 2010). The conservation scores of the C-terminal residues are also shown in Table II (CS1). For SurA homolog alignment, V423 and A421 have the highest conservation scores of 9 and 7, respectively. Other residues have scores of 3 or 4. Overall, residues 419–428 at the CT are modestly conserved. In contrast, the conservation scores were much lower among the 20 sequences including the mutants and wild-type SurA, indicating a more diverse collection of sequence composition. Second, valine seemed to be over-represented in the table, especially in the boxed area corresponding to the β -strand. This observation prompted us to examine the secondary structure propensity of residues in these functional sequences as Val is known to be a strong β -strand former (Chou and Fasman, 1974). P_α and P_β are α -helix and β -strand parameters computed from the frequency of occurrence of each amino acid residue in the α or β conformations (Chou and Fasman, 1974). The cumulative α -helix ($\sum P_\alpha$) and β -strand ($\sum P_\beta$) parameters of residues 422–426 in each

sequence were calculated (Table II). Interestingly, all sequences in the table had a higher overall β -strand propensity as revealed by the positive ΔP values ($\Delta P = \Sigma P_{\beta} - \Sigma P_{\alpha}$), indicating an intrinsic propensity to form the β -strand. However, the exact number of ΔP did not correlate with how active the protein was.

To further examine the secondary structure preference of residues 422–426 in different mutants, we conducted secondary structure prediction using a recently developed program NetSurfP (Petersen et al., 2009). The cumulative probabilities of these five residues to form α -helix, β -strand and coil are summarized in Table II. Seventeen out of the 19 randomly picked sequences were predicted to most likely form β -strands, in good agreement with the prediction using the Chou-Fasman's secondary structure propensity values. Two sequences, M1 and M14, were predicted to have a higher tendency to form coils. This difference could be due to the possibility that certain local environmental effect might not be captured by the prediction.

To further explore the sequence requirement at the CT, we constructed four negative control sequences (C1–C4, Table II). These sequences are identical to the sequence of the wild-type, SurA except that residues 422–426 were occupied by a mixture of Ser, Ala and Glu, which according to Chou-Fasman's scale have low β -strand forming propensities (Chou and Fasman, 1974). The activities of these control proteins were measured similarly as the 19 mutants and none of them were active.

Our result echoes with the observation of the promiscuous peptide binding by the PDZ domains, which are a class of the most widespread protein–protein interacting modules (Teyra et al., 2012; Luck et al., 2012). PDZ domains bind with peptide ligands, which are usually part of transmembrane receptors or ion channels (Ivarsson, 2012). The interactions tend to be promiscuous, as one PDZ domain can commonly recognize various peptide ligands and the same peptide ligand can be recognized by different PDZ domains (Harris and Lim, 2001). Remarkably, PDZ domains generally bind the CT of target proteins in a shallow pocket in which the C-terminal residues of the ligand form an extended anti-parallel β -sheet with two β -strands of the PDZ domain (Doyle et al., 1996; Jemth and Gianni, 2007), highly resembling the interaction between the C-terminal fragment and the N-terminal β -hairpin in the structure of SurA. The backbone hydrogen bonds of the extended β -sheet have been shown to be conserved features of the canonical PDZ-ligand binding (Chi et al., 2012), which is consistent with our observation that the formation of β -strand, and thus the preservation of the backbone hydrogen bonds in the anti-parallel β -sheet, is critical for the function of SurA.

In summary, we discovered that SurA tolerated a vast diversity of sequence compositions at its CT. This observation suggests that the functional role of the CT is more likely for structural support rather than substrate recognition and interaction, which is expected to require sequence-specific information. From the protein engineering point of view, while under certain conditions a single protein sequence always folds into a specific structure, the same structure could be the folding destination of many different sequences. For example, Hecht et al. have shown, through a series of studies, that well-folded four-helix bundle proteins could be designed by a combinatorial approach (Hecht et al., 2004). Here, our result confirmed that in natural proteins the amino acid composition

of fragments that are not directly involved in activity could be varied drastically provided basic principles such as secondary structure propensities are satisfied.

Funding

This work was supported by the National Science Foundation (MCB 1158036 to Y.W.) and NASA EPSCoR program (WKURF 516212 and NNX10AV39A to Y.W.).

References

- Behrens,S., Maier,R., de Cock,H., Schmid,F.X. and Gross,C.A. (2001) *EMBO J.*, **20**, 285–294.
- Behrens-Kneip,S. (2010) *Int. J. Med. Microbiol.*, **300**, 421–428.
- Bernstein,H.D. (2000) *Curr. Opin. Microbiol.*, **3**, 203–209.
- Bitto,E. and McKay,D.B. (2002) *Structure*, **10**, 1489–1498.
- Bitto,E. and McKay,D.B. (2003) *J. Biol. Chem.*, **278**, 49316–49322.
- Bitto,E. and McKay,D.B. (2004) *FEBS Lett.*, **568**, 94–98.
- Bos,M.P. and Tommassen,J. (2004) *Curr. Opin. Microbiol.*, **7**, 610–616.
- Bos,M.P., Robert,V. and Tommassen,J. (2007) *Ann. Rev. Microbiol.*, **61**, 191–214.
- Chi,C.N., Bach,A., Strømgaard,K., Gianni,S. and Jemth,P. (2012) *Biofactors*, **38**, 338–348.
- Chou,P.Y. and Fasman,G.D. (1974) *Biochemistry*, **13**, 211–222.
- Dominique,M., Jean-Michel,B. and Satish,R. (1996) *Mol. Microbiol.*, **21**, 871–884.
- Doyle,D.A., Lee,A., Lewis,J., Kim,E., Sheng,M. and MacKinnon,R. (1996) *Cell*, **85**, 1067–1076.
- Duguay,A.R. and Silhavy,T.J. (2004) *Biochim. Biophys. Acta*, **1694**, 121–134.
- Gold,V.A.M., Duong,F. and Collinson,I. (2007) *Mol. Membr. Biol.*, **24**, 387–394.
- Harris,B.Z. and Lim,W.A. (2001) *J. Cell Sci.*, **114**, 3219–3231.
- Hecht,M.H., Das,A., Go,A., Bradley,L.H. and Wei,Y. (2004) *Protein Sci.*, **13**, 1711–1723.
- Hennecke,G., Nolte,J., Volkmer-Engert,R., Schneider-Mergener,J. and Behrens,S. (2005) *J. Biol. Chem.*, **280**, 23540–23548.
- Ivarsson,Y. (2012) *FEBS Lett.*, **586**, 2638–2647.
- Jemth,P. and Gianni,S. (2007) *Biochemistry*, **46**, 8701–8708.
- Justice,S.S., Hunstad,D.A., Harper,J.R., Duguay,A.R., Pinkner,J.S., Bann,J., Frieden,C., Silhavy,T.J. and Hultgren,S.J. (2005) *J. Bacteriol.*, **187**, 7680–7686.
- Larkin,M.A., Blackshields,G., Brown,N.P., et al. (2007) *Bioinformatics*, **23**, 2947–2948.
- Lazar,S.W. and Kolter,R. (1996) *J. Bacteriol.*, **178**, 1770–1773.
- Lazar,S.W., Almiron,M., Tormo,A. and Kolter,R. (1998) *J. Bacteriol.*, **180**, 5704–5711.
- Luck,K., Charbonnier,S. and Travé,G. (2012) *FEBS Lett.*, **586**, 2648–2661.
- Missiakas,D., Betton,J.M. and Raina,S. (1996) *Mol. Microbiol.*, **21**, 871–884.
- Mogensen,J.E. and Otzen,D.E. (2005) *Mol. Microbiol.*, **57**, 326–346.
- Palomino,C., Marin,E. and Angel Fernandez,L. (2011) *J. Bacteriol.*, **193**, 5222–5230.
- Petersen,B., Petersen,T.N., Andersen,P., Nielsen,M. and Lundegaard,C. (2009) *BMC Struct. Biol.*, **9**, 51.
- Rizzitello,A.E., Harper,J.R. and Silhavy,T.J. (2001) *J. Bacteriol.*, **183**, 6794–6800.
- Rouviere,P.E. and Gross,C.A. (1996) *Genes Dev.*, **10**, 3170–3182.
- Ruiz,N., Kahne,D. and Silhavy,T.J. (2006) *Nat. Rev. Microbiol.*, **4**, 57–66.
- Sklar,J.G., Wu,T., Kahne,D. and Silhavy,T.J. (2007) *Genes Dev.*, **21**, 2473–2484.
- Tamae,C., Liu,A., Kim,K., et al. (2008) *J. Bacteriol.*, **190**, 5981–5988.
- Teyra,J., Sidhu,S.S. and Kim,P.M. (2012) *FEBS Lett.*, **586**, 2631–2637.
- Tormo,A., Almiron,M. and Kolter,R. (1990) *J. Bacteriol.*, **172**, 4339–4347.
- Vertommen,D., Ruiz,N., Leverrier,P., Silhavy,T.J. and Collet,J.F. (2009) *Proteomics*, **9**, 2432–2443.
- Waterhouse,A.M., Procter,J.B., Martin,D.M.A., Clamp,M. and Barton,G.J. (2009) *Bioinformatics*, **25**, 1189–1191.
- Wulfing,C. and Pluckthun,A. (1994) *Mol. Microbiol.*, **12**, 685–692.
- Xu,X.H., Wang,S.Y., Hu,Y.X. and McKay,D.B. (2007) *J. Mol. Biol.*, **373**, 367–381.
- Zhong,M., Ferrell,B., Lu,W., Chai,Q. and Wei,Y. (2013) *J. Bacteriol.*, **195**, 1061–1067.