

## Purification and Biophysical Characterization of the CapA Membrane Protein FTT0807 from *Francisella tularensis*

Jose M. Martin-Garcia,<sup>†,§</sup> Debra T. Hansen,<sup>‡,§</sup> James Zook,<sup>†,§</sup> Andrey V. Loskutov,<sup>‡</sup> Mark D. Robida,<sup>‡,§,||</sup> Felicia M. Craciunescu,<sup>‡,§</sup> Kathryn F. Sykes,<sup>‡,§</sup> Rebekka M. Wachter,<sup>†,§</sup> Petra Fromme,<sup>†,§</sup> and James P. Allen<sup>\*,†,§</sup>

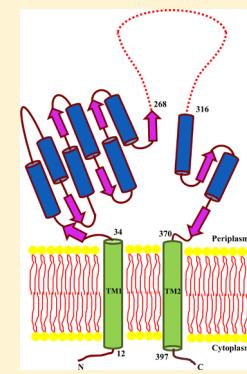
<sup>†</sup>Department of Chemistry and Biochemistry, Arizona State University, Tempe, Arizona 85287, United States

<sup>‡</sup>Center for Innovations in Medicine, Biodesign Institute, Arizona State University, Tempe, Arizona 85287, United States

<sup>§</sup>Center for Membrane Proteins in Infectious Diseases, Arizona State University, Tempe, Arizona 85287, United States

### Supporting Information

**ABSTRACT:** The *capA* gene (FTT0807) from *Francisella tularensis* subsp. *tularensis* SCHU S4 encodes a 44.4 kDa integral membrane protein composed of 403 amino acid residues that is part of an apparent operon that encodes at least two other membrane proteins, CapB, and CapC, which together play a critical role in the virulence and pathogenesis of this bacterium. The *capA* gene was overexpressed in *Escherichia coli* as a C-terminal His<sub>6</sub>-tagged fusion with a folding reporter green fluorescent protein (frGFP). Purification procedures using several detergents were developed for the fluorescing and membrane-bound product, yielding approximately 30 mg of pure protein per liter of bacterial culture. Dynamic light scattering indicated that CapA-frGFP was highly monodisperse, with a size that was dependent upon both the concentration and choice of detergent. Circular dichroism showed that CapA-frGFP was stable over the range of 3–9 for the pH, with approximately half of the protein having well-defined  $\alpha$ -helical and  $\beta$ -sheet secondary structure. The addition of either sodium chloride or calcium chloride at concentrations producing ionic strengths above 0.1 M resulted in a small increase of the  $\alpha$ -helical content and a corresponding decrease in the random-coil content. Secondary-structure predictions on the basis of the analysis of the sequence indicate that the CapA membrane protein has two transmembrane helices with a substantial hydrophilic domain. The hydrophilic domain is predicted to contain a long disordered region of 50–60 residues, suggesting that the increase of  $\alpha$ -helical content at high ionic strength could arise because of electrostatic interactions involving the disordered region. CapA is shown to be an inner-membrane protein and is predicted to play a key cellular role in the assembly of polysaccharides.



The bacterium *Francisella tularensis* is an intracellular member of the gamma Proteobacteria and the causative agent of tularemia, a disease in humans and many other species.<sup>1–3</sup> Because of the low dose sufficient for infection, the subspecies *tularensis* is one of the most virulent bacteria known and is classified as a class A bioterrorism agent in the United States. *F. tularensis* employs bacterial surface proteins to colonize the host cells by binding to specific host-cell receptors.<sup>4,5</sup> After entry, the bacterium can escape from the phagosomal membrane for replication and survival in the cytoplasm.<sup>4,5</sup> During the last 2 decades, biochemical and genetic analyses show that the bacterium's surface structure is among the factors that can contribute to pathogenesis and virulence of *F. tularensis*.<sup>6,7</sup> The surface structure is composed of an array of macromolecules including proteins, lipids, sugars, and nucleic acids. The dynamic processes of the surface include membrane turnover, alterations in gene-expression profiles, capsular formation, and retractable elements that modify the envelope structure to allow bacteria to adapt to changing environments.<sup>8–11</sup> Bacteria use surface structure components to hide from the immune system, attach to host cells, secrete effector molecules, and cope with physiological stresses. In *F.*

*tularensis*, the role of the surface in virulence is linked to the production of a capsule and an unusual lipopolysaccharide (LPS).<sup>8,11</sup>

Despite the disease implications of *F. tularensis*, the mechanism of virulence is poorly understood. Proteins involved in virulence have been identified on the surface of *F. tularensis*,<sup>12–16</sup> but their specific roles are largely undefined. The gene loci FTL\_1416, FTL\_1415, and FTL\_1414, which have been recently designated as *capB*, *capC*, and *capA*, respectively,<sup>7</sup> is highly conserved in the genus *Francisella* and is necessary for full virulence in the LVS vaccine strain of *F. tularensis* subsp. *holartica*.<sup>6</sup> These genes encode three membrane proteins: CapB (FTL\_1416) of 44.9 kDa and 405 amino acid residues, CapC (FTL\_1415) of 16.7 kDa and 154 amino acid residues, and CapA (FTL\_1414) of 44.4 kDa and 403 amino acid residues. Although the functions of the CapB, CapC, and CapA proteins have not been unambiguously identified, studies carried out in animal models have

**Received:** December 9, 2013

**Revised:** March 3, 2014

**Published:** March 4, 2014



demonstrated that deletion of these three proteins resulted in significant attenuation of bacterial virulence in LVS.<sup>6–8,17–19</sup> Unmarked deletion of CapBCA in type A strain SCHU S4 resulted in significant attenuation in virulence as well. However, the level of attenuation in SCHU S4 was much less than in LVS.<sup>6</sup> It has not yet been determined what stage of infection is impacted by the loss of these proteins, with possibilities including escape from the phagosomes to the cytosol, phagosomal maturation, replication, or reinfection. The original annotation of the *capBCA* genes as being putative capsule genes was based on their sequence homology to the poly- $\gamma$ -glutamate (PGA) capsule locus of *Bacillus anthracis*.<sup>7</sup> Whole-genome sequencing revealed that CapB and CapC proteins in *F. tularensis* LVS share a 38 and 29% amino acid identity, respectively, to the proteins CapB and CapC found in the poly- $\gamma$ -glutamate (PGA) capsule biosynthetic locus of *B. anthracis*.<sup>7,20</sup> In *B. anthracis*, CapB and CapC are believed to form a tight membrane-associated complex to catalyze the synthesis of PGA,<sup>21</sup> whereas CapA is suggested to be a transporter.<sup>22</sup>

PGA-based capsules have been identified in many Gram-positive bacteria, such as *B. anthracis* (see ref 23 for a review) and *Staphylococcus epidermidis*,<sup>24</sup> but are not typically found in Gram-negative bacteria because most of these bacteria possess a polysaccharide-based capsule. Genome-sequencing studies have revealed orthologues for PGA production proteins in several Gram-negative bacteria, including *Idiomarina loihiensis*,<sup>25</sup> *Rhodopirellula baltica*,<sup>26</sup> *Leptospira interrogans*,<sup>27</sup> *Oceanobacillus iheyensis*,<sup>28</sup> and *Desulfitobacterium hafniense*.<sup>29</sup> In contrast, the nature of the capsule in *F. tularensis* is still a subject of great controversy. Electron microscopy analyses support the existence of a capsule in *F. tularensis*,<sup>9</sup> whereas others suggest that the LPS O-antigen might be the only capsule-like structure, with CapA, CapB, and CapC playing no role.<sup>8</sup>

To understand the role of CapA in the virulence of *F. tularensis*, we characterized this protein from the highly virulent strain *F. tularensis* subsp. *tularensis* SCHU S4. In this study, the *capA* gene was cloned and overexpressed in *Escherichia coli* as a C-terminal His<sub>6</sub>-tagged folding reporter GFP fusion protein. A purification protocol for CapA was developed, and the purified protein was subjected to biophysical characterization using different techniques, including SDS-PAGE gel electrophoresis, size-exclusion chromatography (SEC), circular dichroism (CD), and dynamic light scattering (DLS). The experimental results of this study are discussed in terms of a proposed role for CapA in the virulence of *F. tularensis*.

## MATERIALS AND METHODS

**Materials.** All of the materials used are given in the Supporting Information.

**Cloning of CapA-frGFP.** The SCHU S4 genomic DNA was provided from Drs. C. Rick Lyons, Terry H. Wu, and Jason Zsemlye (Center for Infectious Disease and Immunity, University of New Mexico, Albuquerque, NM). The sequences of the CapA membrane protein (FTT0807) from *F. tularensis* subsp. *tularensis* SCHU S4 and of the folding reporter GFP<sup>30</sup> were cloned into the pRSET B vector (Life Technologies, Carlsbad, CA) using the In-Fusion cloning kit (Clontech, Mountain View, CA). The sequence in the pRSET B vector immediately between the nucleotide 5' of the ATG start codon and the sequence 5'-GATCCGGCTGCT-3' near the T7 terminator was replaced by the following insert. The insert contained, from 5' to 3', the full-length, unaltered sequence of *capA*, the linker GGATCAGCAGGTTCCGCTGCTG-

GTTCTGGCGAATTG encoding Gly-Ser-Ala-Gly-Ser-Ala-Ala-Gly-Ser-Gly-Glu-Phe, the sequence encoding the full-length, unaltered frGFP, the sequence CACCAACCACACCAC encoding the hexa-histidine tag, and the noncoding sequence TAATAATAAAAGGGCGAATTCCAGCACACTGGCGGCCGTTACTAGTG. The corresponding parent expression vector for this construct was called pRSET-natGFPHis ("nat" for expression of the natural N-terminus of the target protein). The folding reporter GFP carries a destabilizing mutation that has been shown to allow folding and fluorescence of GFP only when its N-terminal-fused partner target protein is also folded.<sup>30,31</sup> In addition, the fusion of frGFP provided a means to determine the localization of CapA in the different cellular membranes, as has been reported for membrane proteins of *E. coli*.<sup>32–35</sup>

**Cloning of frGFP.** The folding reporter GFP-His<sub>6</sub> without *capA* was also cloned into pRSET B. The sequence in pRSET B immediately between the nucleotide 5' of the ATG start and the sequence 5'-GATCCGGCTGCT-3' near the T7 terminator was replaced by the following insert. The insert contained, from 5' to 3', the full-length, unaltered sequence of frGFP,<sup>30</sup> the sequence CACCAACCACACCAC encoding the hexahistidine tag, and the noncoding sequence TAATAATAAAAGGGCGAATTCCAGCACACTGGCGGCCGTTACTAGTG.

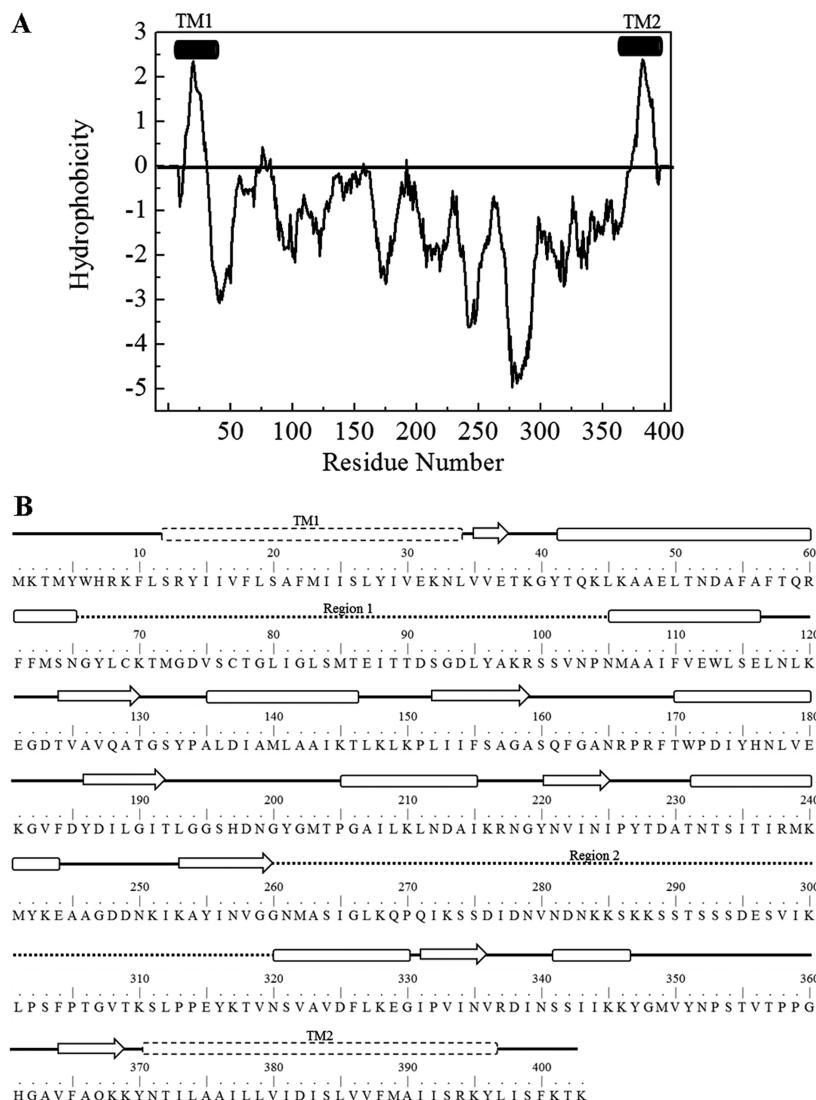
**Preparation of the Crude Membrane Fraction of CapA-frGFP.** Fractions containing membranes were prepared following standardized protocols for bacterial membrane proteins<sup>31,36,37</sup> with a few modifications. All of the details of this preparation are included in the Supporting Information. After isolation, the membrane pellet was flash-frozen in liquid nitrogen and stored at -80 °C. Then this fraction was thawed, resuspended, and homogenized in 50 mM sodium phosphate, pH 7.3 and 0.3 M sodium chloride prior to extraction.

**Detergent Solubilization of CapA-frGFP.** The CapA-frGFP protein was successfully solubilized with the following detergents and concentrations: 1%  $\beta$ -DDM, 1%  $\beta$ -DM, 2% CYMAL-6, and 2% CHAPS. Further information on the procedure is included in the Supporting Information. Final suspensions were analyzed by SDS-PAGE.

**Purification of CapA-frGFP.** The CapA-frGFP protein was first purified by nickel affinity chromatography in the presence of each of the detergents tested. Then, the protein was further purified by SEC using a Superdex-200 HR 10/300 gel-filtration column. All protein fractions from both purification steps were analyzed for purity by SDS-PAGE. More detailed information on the purification protocol can be found in the Supporting Information.

**Purification of the Folding Reporter GFP.** The folding reporter GFP was purified by nickel affinity chromatography followed by a second purification step by SEC using a Superdex-200 HR 10/300 gel-filtration column. All elution fractions from both purification steps were analyzed for purity by SDS-PAGE. More detailed information on the purification protocol can be found in the Supporting Information.

**SDS-PAGE and Western Blot Analyses.** Detection of proteins was accomplished by SDS-PAGE using silver stain and Western blotting as detection methods. SDS-PAGE was performed with Tricine-SDS gels with stacking and resolving gels of 4 and 12%, respectively. For further information on staining methods, see the Supporting Information. The mass of the purified CapA-frGFP protein was verified to be 72 860 Da by MALDI-TOF mass spectroscopy using sinapinic acid (SA)



**Figure 1.** (A) Kyte and Doolittle hydrophobicity plot<sup>83</sup> of the CapA membrane protein. Transmembrane (TM) regions are indicated as black cylinders. These TM regions were predicted to form  $\alpha$ -helices by using several membrane topology prediction programs as described in the text. (B) Amino acid sequence and domain structure of the CapA membrane protein. Predicted secondary structures using the GOR algorithm<sup>38</sup> are represented as solid lines (loops), enclosed rectangles ( $\alpha$ -helices), and block arrows ( $\beta$ -sheets). The dashed rectangles indicate the transmembrane helices TM1 and TM2. Larger loops corresponding to regions 1 and 2 are represented by dotted lines.

as matrix (Applied Biosystems DE-STR MALDI-TOF, Mass Spectrometry service, Arizona State University).

**Protein Sample Preparation, Circular Dichroism (CD), and Dynamic Light Scattering (DLS).** All information concerning sample preparation, CD, and DLS is given in the Supporting Information.

**Proteinase K Digestion.** Digestion of CapA-frGFP was carried out with proteinase K. To adapt to the extreme proteolytic sensitivity of proteins with intrinsically disordered regions, different reaction tubes of 0.5 mL each were prepared with different amounts of protein (25, 50, and 250  $\mu$ g) and incubated in proteolysis buffer (50 mM Tris, pH 8.0, 200 mM NaCl, 5 mM CaCl<sub>2</sub>) with increasing amounts of proteinase K (3.5, 31, 100, 200, 300, and 600 ng) at different incubation times (30, 20, 10, and 2 min) on ice. The proteolytic reactions were stopped by adding PMSF to a final concentration of 5 mM. The mass of the cleavage products obtained in the digestion were directly verified by MALDI-TOF mass spectroscopy using sinapinic acid (SA) as matrix (Applied

Biosystems DE-STR MALDI-TOF, Mass Spectrometry service, Arizona State University). Samples were also analyzed by SDS-PAGE gel electrophoresis and visualized by Coomassie blue or Western blotting using antibodies against GFP.

**Bioinformatics and Disorder Analyses.** All of the information concerning bioinformatics and disorder analyses is given in the Supporting Information.

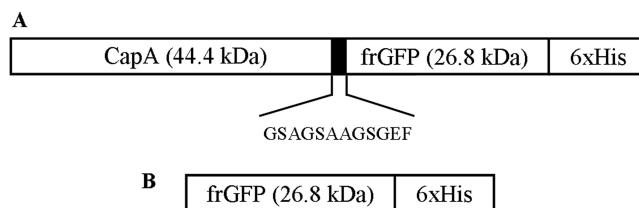
## RESULTS

**Membrane Protein Topology and Secondary-Structure Predictions of CapA.** The sequence of CapA shows regions of predominantly hydrophilic amino acids as well as regions that contain mainly hydrophobic amino acid residues. To examine which regions of CapA are embedded in the cell membrane and their in/out orientation relative to the membrane, we predicted the membrane topology using several bioinformatics algorithms. All programs identified an N-terminal transmembrane helix, termed TM1, formed by amino acids 12–34, a large hydrophilic region formed by

residues 35–369, and a C-terminal transmembrane helix, termed TM2, formed by amino acids 370–397 (Figure 1A). In addition to this, all predictors were in agreement on the overall orientation of the transmembrane helices relative to the membrane. As shown in Figure S1 (which shows more detailed information on topology derived from each program), the CapA protein is predicted to have its N- and C-termini oriented toward the cytoplasm; therefore, the large hydrophilic domain must be oriented toward the periplasm.

Secondary-structure predictions were performed using several predictors such as GOR, NetSurfP, PSIPRED, and PORTER.<sup>38–41</sup> The average of these analyses predicts the CapA protein to be mostly  $\alpha$ -helical ( $35.2 \pm 1.5\%$ ), with a high content of random coil ( $54.4 \pm 1.1\%$ ) and a small  $\beta$ -sheet content ( $10.4 \pm 0.4\%$ ), where the error represents the standard deviation from the results of the four programs. All of the algorithms predicted a very similar distribution pattern of the secondary-structure elements. As Figure 1B shows, for the secondary-structure elements predicted by the GOR algorithm,<sup>38</sup> that the CapA protein is predicted to contain a total of 11  $\alpha$ -helices, with nine being located in the hydrophilic domain, and eight short  $\beta$ -strands. A large majority of the predicted random coil is located in the regions formed by residues 65–105 and 260–320.

**Initial Expression of CapA-frGFP.** CapA was expressed in *E. coli* from the pRSET-natGFPHis vector to produce a C-terminally 6-histidine-tagged frGFP fusion protein (Figure 2).



**Figure 2.** (A) Schematic representation of the expressed CapA-frGFP fusion protein (72.8 kDa). CapA was fused to the His-tagged folding reporter GFP via a 12 amino acid linker (black rectangle). (B) Schematic representation of the expressed folding reporter GFP (26.8 kDa).

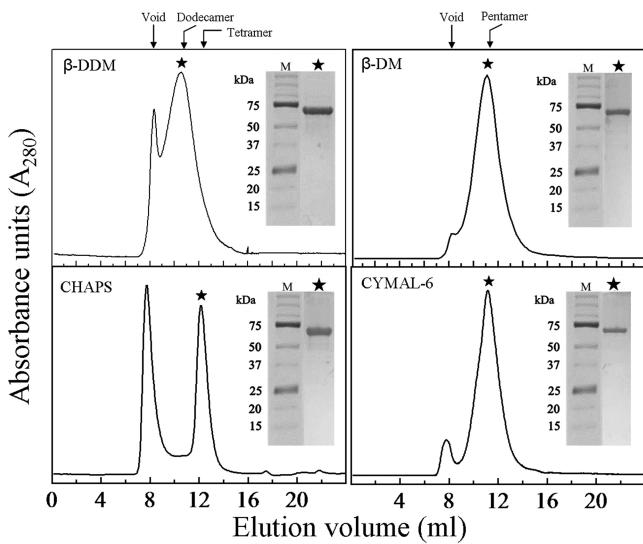
This construct results in the expression of CapA with an unmodified N-terminus for natural targeting to the membrane. The folding reporter GFP carries a destabilizing mutation that allows folding and fluorescence only when its N-terminally fused partner target protein is also folded.<sup>30,31</sup> The reporter GFP was used to identify the cellular location of CapA, as GFP becomes fluorescent only when located in the cytoplasm and not in the periplasmic space.<sup>33,34</sup> The expression of the fusion protein was monitored by western blotting using antibodies against GFP on whole cells after growth at 20 °C (Figure S2A). In addition, uninduced cells and cells without the plasmid were also grown under the same conditions for negative controls. In induced cells, besides the full-length product at about 73 kDa, several fragmented proteins were also visible, including a GFP-sized fragment at about 25 kDa, indicating proteolysis of the fusion joint. As it was expected, a lower intensity band for the fragment of the fusion protein was also detected on the gel for uninduced cells, and no bands were observed for the cells without the plasmid. Whole cells were also positively monitored by fluorescence emission on an advanced bio-spectrum imaging system using an emission filter between 500

to 523 nm (UVP, LLC, Upland, CA). The high emission fluorescence observed indicates that GFP is correctly folded and located in the cytoplasm (Figure S2B). This confirms that CapA is an inner-membrane protein with its N- and C-termini oriented toward the cytoplasm and a very large hydrophilic domain located in the periplasmic region.

**Optimization of CapA-frGFP Expression and Purification.** Terrific broth medium was chosen for optimal cell growth based on a significantly increased cell growth compared to Luria–Bertani medium. Optimal expression of the CapA-frGFP fusion protein was achieved at 20 °C with 12–15 h duration of the expression induction phase. The inclusion of a His<sub>6</sub> tag allowed purification of the protein by affinity chromatography. Five detergents,  $\beta$ -DDM,  $\beta$ -DM, CHAPS, CYMAL-6, and LDAO, were screened for their efficiency to extract and purify CapA-frGFP. Membrane preparations were solubilized in 50 mM sodium phosphate, pH 7.3, and 0.3 M sodium chloride containing one of the following detergents: (i) 1%  $\beta$ -DDM, (ii) 1%  $\beta$ -DM, (iii) 1% CYMAL-6, (iv) 2% CYMAL-6, (v) 1% CHAPS, (vi) 2% CHAPS, and (vii) 1% LDAO. The CapA-frGFP protein could be extracted effectively using all detergents (Figure S3A). To quantify the efficiency of each of the detergents to extract the fusion protein from the membranes, GFP absorbance measurements at 490 nm were carried out (Figure S3B). The most efficient detergent was  $\beta$ -DDM, with 46.5 mg of extracted protein from 1 L of TB medium, followed by  $\beta$ -DM with 38.5 mg, and there were no significant differences between CHAPS, CYMAL-6, and LDAO.

The supernatants containing detergent-solubilized protein were purified by Ni-affinity chromatography in the presence of a given detergent followed by elution with 500 mM imidazole. The purity of each of the eluted fractions was analyzed by SDS-PAGE and western blot (Figure S4). The CapA-frGFP protein could be efficiently purified using most of the detergents (data not shown). The exception was LDAO, where the protein precipitated shortly after elution from the Ni-IMAC column. Although the detergents CHAPS and CYMAL-6 were less effective in extracting the protein, their use was continued to explore their effects on the homogeneity, the conformational stability, and the oligomeric state of the CapA-frGFP protein. For each detergent, elution fractions containing CapA-frGFP were combined and dialyzed extensively against 20 mM HEPES, pH 7.0, and 0.3 M sodium chloride to remove the imidazole prior to biophysical characterization.

**Gel-Filtration Analysis.** SEC was used to purify the solubilized CapA-frGFP and to assess its oligomeric state. Gel-filtration analysis was carried out in 20 mM HEPES, pH 7.0, and 0.3 M sodium chloride plus the appropriate detergent at the following concentrations: 0.05%  $\beta$ -DDM, 0.3%  $\beta$ -DM, 0.1% CYMAL-6, and 1% CHAPS. The SEC chromatograms showed that CapA-frGFP was eluted in two major fractions: one as an aggregated form from peak 1 (column void volume) and the other as nonaggregated from peak 2 (Figure 3). All fractions of the second peak of the SEC for all detergents showed a very high purity upon SDS-PAGE analysis (Figure 3). Using the monomeric molecular mass of 72.9 kDa (as described above) and considering the following micelle sizes of 72, 45, and 46 kDa for  $\beta$ -DDM,  $\beta$ -DM, and CYMAL-6, respectively,<sup>42</sup> and 6 kDa for CHAPS,<sup>43</sup> the CapA-frGFP-detergent complex is estimated to be a heptamer in the presence of  $\beta$ -DDM, a pentamer with  $\beta$ -DM and CYMAL-6, and a tetramer in the presence of CHAPS.



**Figure 3.** Analytical gel-filtration analysis of CapA-frGFP with the indicated detergents in 20 mM HEPES, pH 7.0, and 0.3 M sodium chloride. The arrows indicate the estimated elution position of the void volume and of the oligomer species: tetramer (CHAPS), pentamer ( $\beta$ -DM, and CYMAL-6), and dodecamer ( $\beta$ -DDM). The panels show silver-stained SDS-PAGE of CapA-frGFP from the nonaggregated peak (indicated by a star) following analytical gel filtration. M: protein molecular weight marker.

**Dynamic Light Scattering (DLS).** To characterize the size of the solubilized protein-detergent micelles, the hydrodynamic radius of CapA-frGFP was analyzed by DLS in the presence of each detergent. The oligomeric state of CapA-frGFP was found to be dependent on the detergent used for purification (Figure 4). The hydrodynamic radii of the major species were 19.3, 10.4, and 9.5 nm for  $\beta$ -DDM,  $\beta$ -DM, and CYMAL-6, respectively, corresponding to oligomers of 24, 8–9, and 8–9 identical protein subunits. The smallest radius of 6.9 nm was found in the presence of CHAPS, corresponding to an oligomer of 3–4 identical protein subunits. The estimation of the oligomerization state of CapA-frGFP was done as described in the previous section.

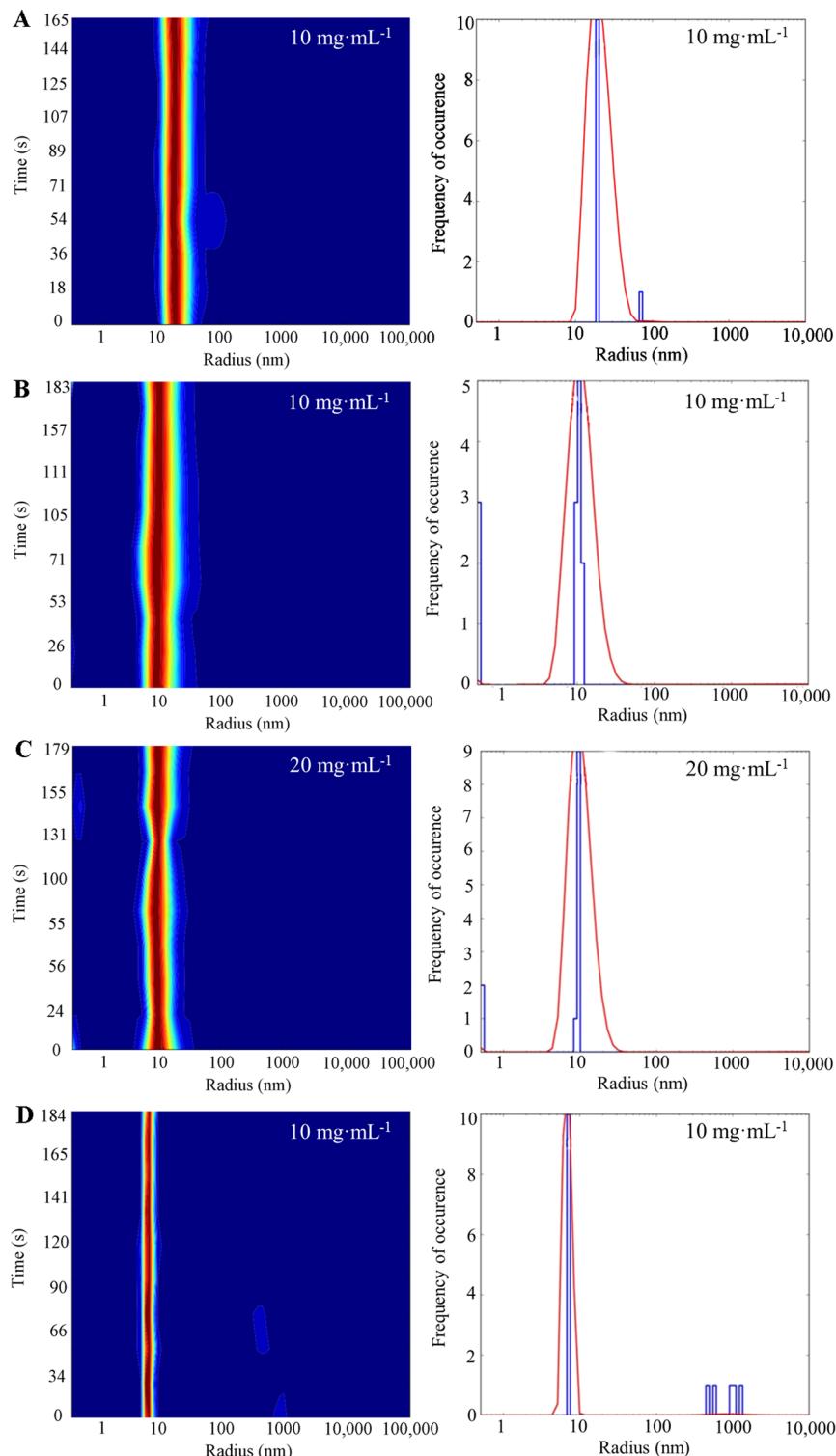
The samples were homogeneous in size at low and moderate protein concentrations (5–20 mg mL<sup>-1</sup>), but larger forms were observed at high protein concentrations (>20 mg mL<sup>-1</sup>). For example, DLS measurements revealed particle sizes of the protein–CHAPS complex with hydrodynamic radii of 6.9–8.9 nm (with estimated molecular weights of 263–473 kDa) within a protein concentration range of 5–20 mg mL<sup>-1</sup> (Figure S3). Considering 72.9 kDa as the monomeric CapA-frGFP molecular mass and a micelle size of 6 kDa for CHAPS detergent,<sup>43</sup> the oligomeric state of the CapA-frGFP protein is estimated to be between a trimer and a hexamer at low protein concentrations. At higher protein concentrations (25 and 30 mg mL<sup>-1</sup>), increasing amounts of larger oligomeric species were observed (Figure S5). Taken together, the detergents investigated in this study are suitable for stabilizing CapA-frGFP in a stable and monodisperse solution (polydispersity <15%) up to 10 mg mL<sup>-1</sup> for  $\beta$ -DDM,  $\beta$ -DM, and CYMAL-6 and 20 mg mL<sup>-1</sup> for CHAPS.

**Effects of Detergents and pH on Secondary Structure.** The far-UV circular dichroism spectrum of the CapA-frGFP protein was initially measured between 180 and 260 nm in the presence of  $\beta$ -DM in 20 mM HEPES, pH 7.0, and 0.3 M

sodium chloride at 20 °C using a reduced-path-length cell (0.01 cm) (Figure S6). The spectrum showed a strong CD signal with a positive band at 193 nm and a broad negative feature between 208 and 240 nm with two minima near 208 and 222 nm, which is characteristic of a protein with a large fraction of  $\alpha$ -helices. Analysis of secondary-structure content provided the following values:  $32.5 \pm 2.1\%$   $\alpha$ -helix,  $20.7 \pm 1.5\%$   $\beta$ -sheet, and  $46.8 \pm 3.2\%$  random coil. For most measurements, circular dichroism spectra measurements were performed using a regular-path-length cell (0.1 cm), whose use was restricted to between 205 and 260 nm and yielded essentially equivalent spectra and secondary-structure estimations. The spectra for the maltoside detergents  $\beta$ -DDM,  $\beta$ -DM, and CYMAL-6 were largely similar and yielded average secondary-structure values of  $27 \pm 1.8\%$   $\alpha$ -helix,  $23 \pm 1.2\%$   $\beta$ -sheet, and  $50 \pm 2.8\%$  random coil (Figure 5). The use of CHAPS resulted in a single unresolved peak near 220 nm, and the spectral analysis yielded a higher value of 5% for the  $\alpha$ -helical component with a corresponding 5% lower value for random coil (Figure 5). These values are comparable to the predicted secondary-structure content of 24, 23, and 53% for  $\alpha$ -helix,  $\beta$ -sheet, and random-coil content, respectively, based upon the protein sequence. No significant changes in the CD spectra were observed for CapA-frGFP in the presence of  $\beta$ -DDM and 300 mM NaCl at pH values ranging from 3.0 to 9.0, although the data indicate a small decrease of approximately 5% in the  $\alpha$ -helical content and a corresponding increase in  $\beta$ -sheet content as the pH increased from 3 to 9 (Figure S7). Thus, these results indicate that the fusion protein CapA-frGFP is stable in a wide range of pH from 3.0 to 9.0. This unusual stability of the CapA-frGFP protein might be attributed to the presence of the folding reporter GFP protein.

To identify the contribution of the frGFP protein to the spectra of the CapA-frGFP fusion protein, CD measurements were performed on purified folding reporter GFP (Figure S8). The far-UV CD spectrum of the frGFP, in the absence of detergent, showed a single minimum at 216 nm, which is characteristic of a  $\beta$ -sheet structure. The averaging of the secondary-structure values obtained from spectral analysis using CDPro software<sup>44</sup> indicates the presence of  $12 \pm 0.9\%$   $\alpha$ -helix,  $55 \pm 1.4\%$   $\beta$ -sheet, and  $33 \pm 2.0\%$  random coil, in agreement with earlier studies.<sup>45</sup> No significant difference was observed with the addition of 0.05%  $\beta$ -DDM, 0.3%  $\beta$ -DM, 0.1% CYMAL-6, or 1% CHAPS at 20 °C and pH 7.0. Likewise, no significant differences were evident when the NaCl concentration was varied from 0 to 300 mM or when the pH was changed from 3 to 9. Thus, the effect of detergents and ionic strength on the conformational stability of the CapA-frGFP can be attributed to arising from changes in the secondary structure of CapA and not frGFP.

**Effect of Ionic Strength on the Secondary Structure.** To investigate the effect of ionic strength on the secondary structure, CD measurements were performed with CapA-frGFP in the presence of either sodium chloride or calcium chloride at concentrations ranging from 0 to 300 mM (Figure 6). For  $\beta$ -DDM with NaCl or CaCl<sub>2</sub>, the spectral band from 210 to 220 nm, which is associated with the presence of  $\alpha$ -helices, is present at ionic strengths of 0.1 M and above, reaching a maximum at 0.2 M (0.2 M NaCl and 0.075 M CaCl<sub>2</sub>), but greatly diminishes in amplitude below 0.1 M. For CHAPS, the protein has a slightly higher  $\alpha$ -helical content, as described earlier, but the spectrum has the same dependence upon ionic strength for both NaCl and CaCl<sub>2</sub>, with the spectral features

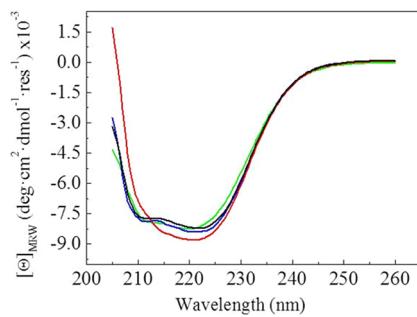


**Figure 4.** Size distribution comparison of CapA-frGFP in the presence of detergents: (A)  $\beta$ -DDM, (B)  $\beta$ -DM, (C) CHAPS, and (D) CYMAL-6. All experiments were performed at 20 °C in 20 mM HEPES, pH 7.0, and 0.3 M sodium chloride. The protein concentration was 10 mg mL<sup>-1</sup> for the detergents  $\beta$ -DDM,  $\beta$ -DM, and CYMAL-6 and 20 mg mL<sup>-1</sup> for CHAPS.

being lost when the ionic strength is decreased below 0.1 M. Thus, the presence of high concentration of salt leads to a substantial increase in the secondary-structure content of the CapA-frGFP protein at neutral pH, irrespective of the detergent used.

**Thermal Stability.** To assess the thermal stability of the CapA-frGFP protein in the presence of each detergent, CD

measurements were conducted at sample temperatures ranging from 4.5 to 90 °C. For all of the detergents, the ellipticity at 220 nm decreased with increasing temperatures, indicating a decrease in the  $\alpha$ -helical content (Figure 7). For  $\beta$ -DDM, the spectral features started to change above 25 °C. In contrast, the protein was more stable in the detergents  $\beta$ -DM, CYMAL-6, and CHAPS, as the spectra were independent of the



**Figure 5.** Effect of detergents on far-UV CD spectra of CapA-frGFP. All spectra were recorded at 20 °C in 20 mM HEPES, pH 7.0, and 0.3 M sodium chloride at the following detergent concentrations: 0.05%  $\beta$ -DDM (green line), 0.3%  $\beta$ -DM (black line), 1% CHAPS (red line), and 0.08% CYMAL-6 (blue line).

temperature up to 37 °C. Using the ellipticity at 220 nm as a marker for helicity, thermal-denaturation profiles were determined for each detergent, yielding apparent melting temperatures of 36, 41, 42, and 48 °C for  $\beta$ -DDM,  $\beta$ -DM, CYMAL-6, and CHAPS, respectively. Thus, the samples purified using CHAPS have the highest melting temperature and  $\alpha$ -helical content.

The stability of CapA-frGFP was also examined by SDS-PAGE with the sample prepared over the temperature range of 4–95 °C (Figure 8). Different samples of CapA-frGFP were prepared and subjected to different temperatures before SDS-PAGE. Unheated samples (4 and 20 °C) and those that were heated at moderate temperatures (36 and 56 °C) showed the presence of two bands on the gel. The upper band indicates a  $M_r$  value close to the true molecular weight (Figure 8, lanes 1–4), and the lower band underestimates the true  $M_r$  value by approximately 8000 Da (Figure 8, lanes 5–7). However, the

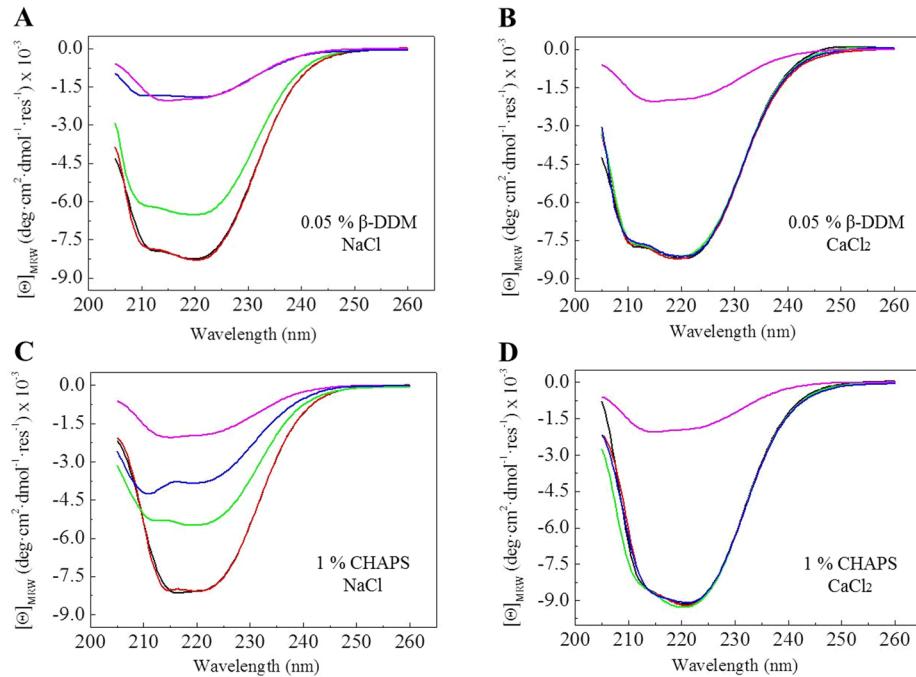
lower band disappeared after heating the samples at temperatures  $\geq 76$  °C. Such behavior indicates a compact, SDS-resistant, secondary structure that results in the protein migrating through the gel faster than protein unfolded by a combination of SDS and heat. This behavior has also been described for SDS-PAGE analyses of TonB-dependent receptors,<sup>46,47</sup> a variety of porins,<sup>48</sup> and the outer-membrane protein BtuB.<sup>49</sup>

### Prediction of a Long Disordered Region in the CapA Protein

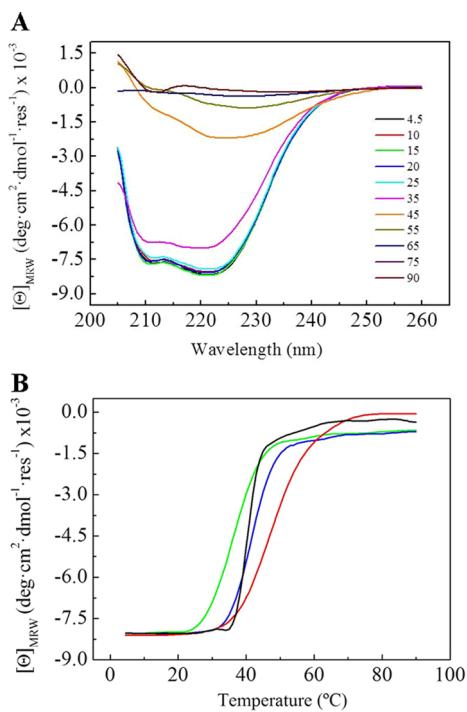
The sequence of CapA was analyzed using several programs that predicted the presence of disordered regions in proteins such as the ammonia channel from *E. coli* and the outer-membrane protein G from *E. coli*.<sup>50,51</sup> On the basis of these programs, the region of CapA formed by residues 260–315 is disordered or has increased propensity for disorder (Figures 9A,B). Using a composite profiler tool (Figure 9C), the disordered region is depleted of order-promoting residues (W, C, F, Y, V, L, and H) and enriched in disorder-promoting residues (S, P, D, and K). The abundance of charged residues (K and D) and the lack of hydrophobic residues (W, F, and Y) indicate a high net charge and a low mean hydrophobicity, respectively. Such a combination has been correlated with intrinsic disorder in proteins and thus can be used to discriminate intrinsically disordered from folded proteins.<sup>52</sup> On the basis of a charge-hydrophobicity phase space distribution, a clear tendency was revealed for intrinsic disorder, with CapA being located in the region of folded proteins (Figure 9D). The high content of polar and charged residues of this unstructured region indicates that it may be highly solvent exposed, suggesting that the secondary structure may be influenced by the ionic strength, as has been reported for other proteins with disordered regions.<sup>53–55</sup>

### Experimental Evidence of an Intrinsically Disordered Region in the CapA Protein

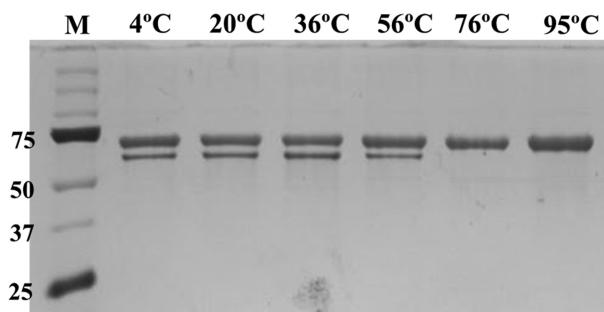
A distinctive feature of proteins



**Figure 6.** Effect of ionic strength on conformational stability of the CapA-frGFP protein: (A) 0.05%  $\beta$ -DDM and NaCl, (B) 0.05%  $\beta$ -DDM and CaCl<sub>2</sub>, (C) 1% CHAPS and NaCl, and (D) 1% CHAPS and CaCl<sub>2</sub>. Far-UV CD experiments were recorded at salt concentrations ranging from 0 to 300 mM (0 mM, pink line; 50 mM, blue line; 100 mM, green line; 200 mM, red line; and 300 mM, black line). These experiments were performed in 20 mM HEPES, pH 7.0, at 20 °C. The protein concentration was 0.5 mg mL<sup>-1</sup>.



**Figure 7.** Effect of temperature on the conformational stability of the CapA-frGFP fusion protein. (A) Temperature dependency of the far-UV CD spectra of CapA-frGFP in 20 mM HEPES, pH 7.0, and 0.3 M sodium chloride, 0.05%  $\beta$ -DDM, and 0.5 mg mL $^{-1}$  protein. Eleven spectra from 4.5 to 90 °C were recorded at increasing intervals of 5–10 °C. (B) Thermal-denaturation curves of CapA-frGFP in the presence of  $\beta$ -DDM (green line),  $\beta$ -DM (black line), CHAPS (red line), and CYMAL-6 (blue line) measured at 220 nm and at a rate of temperature change of 1 °C min $^{-1}$ . Samples were heated from 4.5 to 90 °C. The apparent midpoint temperature of the protein melting-point transition was  $T_m = 36.1$  ( $\beta$ -DDM), 42.1 ( $\beta$ -DM), 48.3 (CHAPS), and 40.8 °C (CYMAL-6).



**Figure 8.** Analysis of secondary structure as identified by SDS-PAGE. Silver-stained SDS 12% PAGE gel of CapA-frGFP after incubation for 1 h at 4, 20, 36, and 56 °C; for 30 min at 76 °C; and for 10 min at 95 °C.

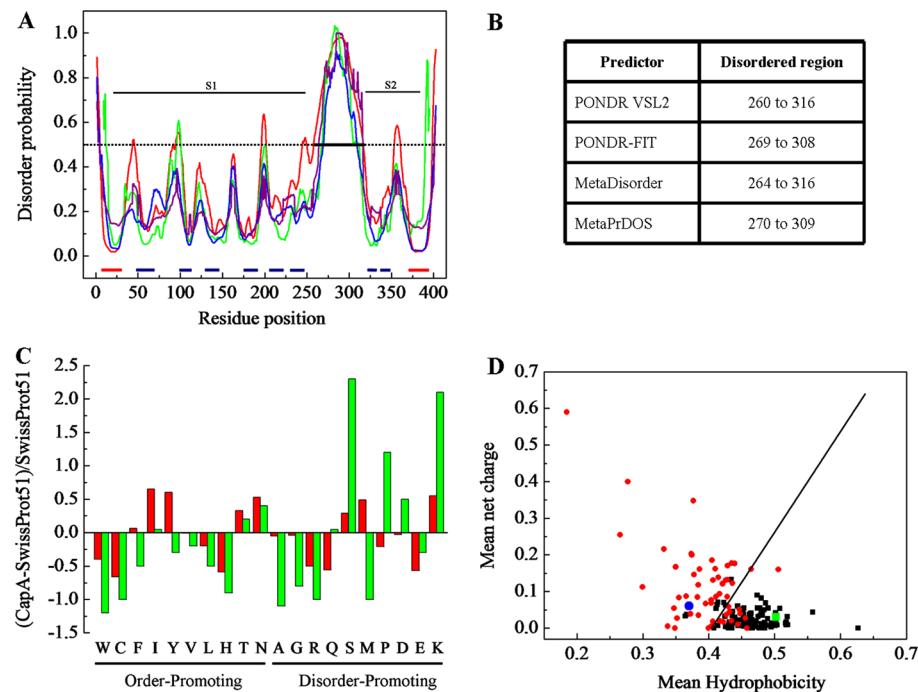
that possess intrinsically disordered regions is their sensitivity to proteolysis,<sup>56</sup> which results in their rapid and complete degradation by proteases even under the usual conditions of limited proteolysis. Limited proteolysis occurs preferentially at those loops that display inherent conformational flexibility, whereas the protein core remains quite rigid and thus resistant to proteolysis.<sup>57–61</sup> Therefore, to identify the presence of an intrinsically disordered region in CapA experimentally, proteolysis experiments were performed using proteinase K,

which has been described to hydrolyze flexible and therefore disordered regions.<sup>62–65</sup>

Incubation of purified CapA-frGFP with increasing concentrations of proteinase K ranging from 0.06 to 1.2 ng  $\mu$ L $^{-1}$  at incubation times of 10, 20, and 30 min led to the formation of many degradation products (data not shown). Under these conditions, proteinase K degrades the full-length CapA-frGFP protein into small fragments of molecular weights  $\leq$ 15 kDa. Two more fragments at about 27 kDa are apparent, most likely containing the fragments of frGFP alone and linker-frGFP. However, a different protein-degradation pattern was observed when the amount of proteinase K was decreased to 0.007 ng  $\mu$ L $^{-1}$  and the incubation time was reduced to 2 min (Figure 10A, lane 1). Under these conditions, the presence of the prominent band at about 73 kDa indicates that the full-length CapA-frGFP is mostly resistant to this protease treatment. In addition, two additional groups of bands are also visible on the gel: two bands migrating between 37 and 50 kDa and two more bands migrating between 25 and 37 kDa. The fragments corresponding to these four bands were also detected, along with the fragments of the singly and doubly charged ions of the full-length protein (73 and 36.5 kDa, respectively), on MALDI-TOF experiments (Figure 10A). Western blot analysis with antibodies specific for GFP revealed that, besides the fragment at 73 kDa for the full-length protein, only the upper bands contain frGFP (Figure 10A, lane 2). Despite the high number of predicted cleavage sites on CapA-frGFP for proteinase K (a total of 326), our results with limited proteolysis demonstrates the presence of two particularly susceptible locations: cleavage site 1 between residue Ser275 and Asp276 and cleavage site 2 between residue Thr291 and Ser292 (Figure 10B). The fragments lacking frGFP that migrate between 25 and 37 kDa (lower bands in Figure 10A, lane 1) can be identified as fragments 1 (Figure 10B) with observed molecular weights of 30.4 and 32.2 kDa. The fragments that contain frGFP and migrate between 37 and 50 kDa (upper bands in Figure 10A, lane 1) can be identified as fragments 2 (Figure 10B) with observed molecular weights of 40.8 and 42.6 kDa. Thus, these results are consistent with the region of the protein between residues 268 and 316 as being highly disordered and therefore prone to proteolytic cleavage.

## DISCUSSION

Despite their key role in virulence, the characterization of the inner- and outer-membrane-associated proteins of *F. tularensis* is still limited because of the challenge of isolating these proteins in a pure, monodisperse, stable, and functional state. Toward this end, we have developed a purification protocol coupled with biophysical characterization and sequence analyses of the integral membrane protein CapA from the highly virulent SCHU S4 strain of *F. tularensis*. To generate large amounts of pure membrane protein, we have developed a simple and very efficient method of purification of the CapA protein tagged with the folding reporter GFP.<sup>30,31</sup> Optimal solubilization conditions were found using the detergents  $\beta$ -DDM,  $\beta$ -DM, CYMAL-6, and CHAPS (Figure S3). Preparation yields were 15 mg of purified CapA-frGFP per 1 L of culture with CYMAL-6 and CHAPS detergents and 30 mg of purified CapA-frGFP per 1 L of culture using  $\beta$ -DDM and  $\beta$ -DM. The oligomeric state of the CapA-frGFP protein was shown to be detergent-dependent (Figure 4). The estimated values range from an oligomer consisting of 3–4 identical protein subunits in the presence of CHAPS to a larger oligomeric state observed



**Figure 9.** Prediction of intrinsically disordered residues in the CapA membrane protein. (A) Plot comparing the prediction by PONDR VSL2 (red line), PONDR-FIT (green line), MetaPrDOS (blue line), and MetaDisorder (purple line). Structured domains are shown by black horizontal lines (S1 and S2). The transmembrane regions are shown by red horizontal lines. Structured, nontransmembrane regions are shown by blue horizontal lines. The dashed line at 0.5 on the y axis is the threshold line for disordered/ordered residues. Residues with a score above this line are predicted to be disordered, and residues with a score below this line are predicted to be ordered. (B) Predicted disordered regions on CapA. (C) Composition profile of full-length CapA (red bars) and of the predicted disordered region (green bars). Analysis was performed using the composite profiler tool and the SwissProt51 database. Horizontal bars indicate order-promoting and disorder-promoting residues. (D) Charge-hydrophobicity phase space for full-length CapA (green square) and for the predicted disordered region (blue dot). Red dots depict intrinsically disordered proteins described in the literature (data taken partially from ref 52). Black squares depict natively folded proteins obtained from the PDB database. The solid black line represents the border between intrinsically disordered proteins and folded proteins.

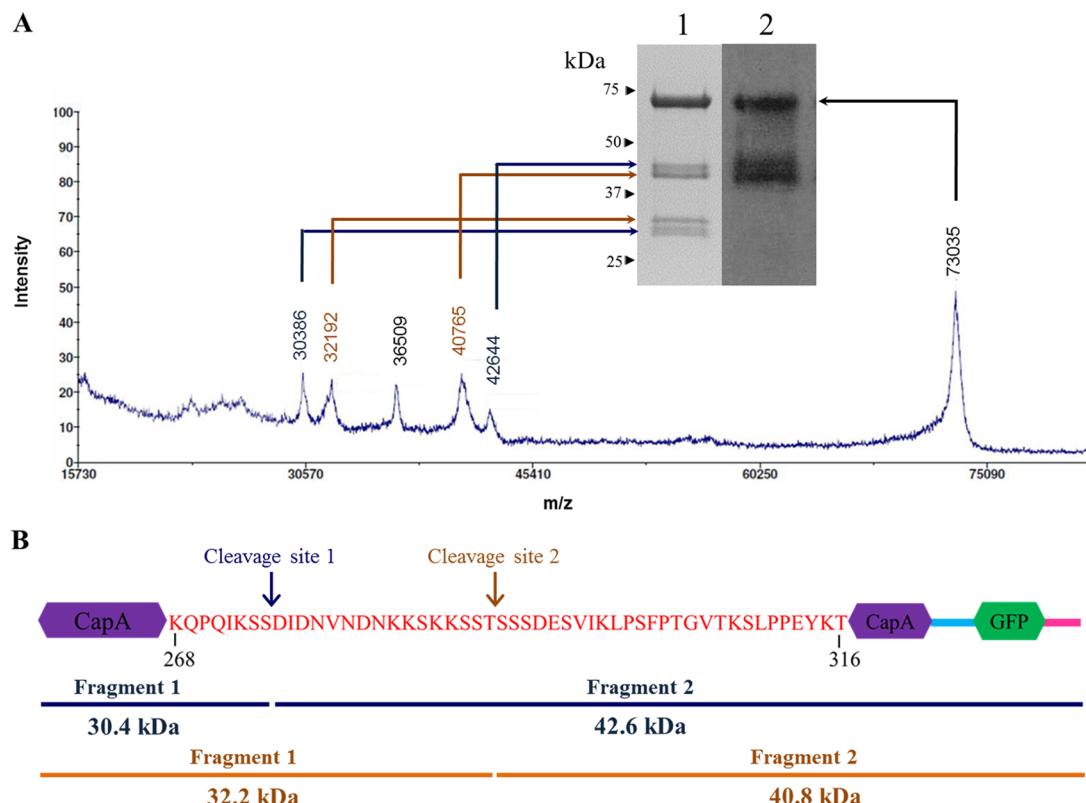
in the presence of  $\beta$ -DDM with approximately 24 identical copies of the CapA-frGFP protein. Despite the large changes in the oligomeric states formed in the presence of different detergents, all CapA-frGFP protein complexes were shown to be highly monodisperse (polydispersity <15%) with all of the detergents investigated.

The conformational stability of the CapA-frGFP protein was examined using CD (Figures 5–7). Use of the detergents  $\beta$ -DDM,  $\beta$ -DM, and CYMAL-6 showed similar spectral features, with two large peaks of negative ellipticity centered at 210 and 220 nm, which is characteristic of a protein with a large fraction of  $\alpha$ -helices. The use of CHAPS resulted in a notable change in the observed spectrum and a corresponding small increase in the  $\alpha$ -helical content. The CapA-frGFP protein was found to be stable throughout a large pH range (Figure S7). The thermal stability of the secondary structure was examined by native SDS-PAGE at different temperatures and revealed that the higher  $\alpha$ -helical content corresponds with an increased stability of the CapA-frGFP, as measured by the melting temperature (Figure 8).

**Disordered Region of CapA.** To maintain its stability, the CapA-frGFP fusion protein requires an ionic strength of 0.2 M or higher, indicating that electrostatic interactions are playing a role in stability of the protein (Figure 6). This effect suggested that CapA has a disordered region, as many proteins with disordered regions can be transformed into more ordered conformations when electrostatic repulsion is reduced by the binding of oppositely charged ions.<sup>66–69</sup> An analysis of the sequence and proteolysis experiments (Figures 9 and 10)

revealed the disordered region to involve residues 268–316. This region contains an abundance of charged amino acids (nine K residues, four D residues, and two E residues), making it very flexible and hydrophilic, with a relatively high positive net charge (+3), which is characteristic of solvent-exposed regions. Disordered regions have been implicated as being binding sites of proteins to their target molecules,<sup>70–74</sup> suggesting that this region of CapA may interact with other proteins. These interacting proteins may include CapB and CapC, as a study of the protein–protein interactions using 2D BN/SDS-PAGE gels analysis in the LVS vaccine strain of *F. tularensis* identified a complex with a size of 400 kDa formed by these three proteins.<sup>75</sup> We propose that CapA interacts with CapB and CapC through the disordered region of CapA; work is underway in our laboratories to identify the specific interactions.

**Role of the CapA Protein in Polysaccharide Formation.** Gram-negative bacteria have a more complex cell wall envelope than their Gram-positive counterparts. The outer and inner membranes of Gram-negative bacteria are separated from the plasma membrane by a prominent periplasmic space. Integral membrane proteins found in the outer membrane are mainly  $\beta$ -barrel outer-membrane proteins ( $\beta$ -barrel OMPs). These  $\beta$ -barrel OMPs mainly function as pores to allow the transport of nutrients, solutes, and waste products in and out of the cell. In contrast, integral membrane proteins found in the inner membrane are mostly  $\alpha$ -helical inner-membrane proteins ( $\alpha$ -helical IMPs). Among other functions,  $\alpha$ -helical IMPs have been described to be involved in the biosynthesis and export of



**Figure 10.** Evidence of an intrinsically disordered region in CapA. (A) Limited proteolysis of CapA-frGFP protein. CapA-frGFP protein at 0.5 mg mL<sup>-1</sup> was digested with 0.007 ng μL<sup>-1</sup> of proteinase K on ice. Digestion was stopped after 2 min and visualized by SDS-PAGE using Coomassie blue stain (lane 1) and western blot with anti-GFP antibody (lane 2). Fragments generated after proteolysis are marked by long arrows. The upper band at about 73 kDa corresponds to the uncleaved protein. The molecular masses of the marker are indicated by triangles. (B) Schematic representation of the CapA-frGFP protein. The linker and the 6x His-tag are in blue and pink, respectively. The sequence is not to scale. The hypothetical cleavage sites, the two fragments, and the molecular masses of each fragment are indicated.

the polysaccharide-based capsule in Gram-negative bacteria.<sup>76</sup> In this work, we have combined the use of membrane topology predictors with frGFP as a C-terminal reporter to demonstrate that the CapA protein is an IMP.

The CapA protein has two transmembrane helices located at the N- and C-termini of the protein and a very large hydrophilic domain of approximately 340 amino acid residues between TM1 and TM2 (Figure 1A). This topology is similar to that of a family of proteins called polysaccharide co-polymerases (PCPs).<sup>77</sup> PCPs, also known as BY-kinases (for bacterial tyrosine kinases),<sup>78</sup> belong to a superfamily of membrane proteins usually encoded by genes in the large operons involved in the export of polymeric compounds such as LPS, O-antigen polysaccharides, and capsule polysaccharides (CPS) in both Gram-negative and Gram-positive species.<sup>76,77,79,80</sup> The PCPs are located in the cytoplasmic membrane and are distinguished by their common membrane topology of one transmembrane helix located at the amino terminus region and another at the carboxyl terminus region. Between the transmembrane helices is a large hydrophilic region ranging from 130 to 400 amino acid residues, which, for several Gram-negative bacteria, is located on the periplasmic side of the inner membrane and contains regions with significant α-helical content.<sup>81,82</sup>

Three crystallographic structures of PCP proteins of the subfamily PCP1 are currently available.<sup>80</sup> The periplasmic domains of FepE (*E. coli*), WzzB (*Salmonella typhimurium*), and WzzE (*E. coli*), despite a limited sequence identity (15–20%), all exhibit a very similar 3D fold that can be divided structurally

into two domains: an α/β base domain and a protruding α-helical domain. In light of this structural information and in order to explore wider structural similarities within the PCP family, secondary-structure predictions have been used by Morona and co-workers to analyze representatives of each PCP subfamily and to compare them with the known atomic resolution structures of PCP1 proteins.<sup>77</sup> The proteins Wzc from *E. coli*, ExoP from *Sinorhizobium meliloti*, and CspC from *Streptococcus pneumoniae*, which are representative of the PCP2 subfamily, were shown to have a very similar membrane topology as the PCP1 group, namely, two TM helices flanking a hydrophilic domain of about 300 residues for Wzc and ExoP and 130 residues for CspC.

This conservation of topology suggests that all PCPs adopt a similar fold and supports the identification of CapA from *F. tularensis* as a PCP. The secondary-structure predictions indicate the presence of β-strands at similar topological locations as in PCP proteins: a β strand at the N-terminal end of the hydrophilic region distal to TM1 and another β strand at the C-terminus of this region proximal to TM2. Another similarity with the PCP1 group is the long flexible connections between the secondary-structure elements. CapA is predicted to contain several additional β-strands and fewer α-helices than that found in PCP1 proteins, as has been also observed for the capsule proteins CapA from *S. aureus* and CspC from *S. pneumoniae*. Although the sequence of the hydrophilic domain of CapA from *F. tularensis* does not match any structurally characterized protein, the pairwise sequence

identity between the sequences of CapA and the PCP proteins WzzE and WzzB (from PCP1 group) and Wzc and ExoP (from PCP2 group) calculated with BLAST reaches from 11 to 15% for their hydrophilic segments alone, which would be in agreement with the limited sequence identity described earlier for this protein family. In addition, sequence analysis of the periplasmic regions of PCP proteins has revealed the presence of random-coil regions,<sup>77,79</sup> as found for CapA from *F. tularensis*. Examination of the periplasmic region using disorder predictors has demonstrated that this region is potentially disordered for several members of the PCP family (Figure S9), which would be in agreement with the weak electron density observed for these regions in their crystallographic structures.<sup>80</sup>

## CONCLUSIONS

The CapA membrane protein from *F. tularensis* can be purified as a stable, monodisperse solution. These preparations have allowed a detailed biophysical characterization, and crystallization trials are underway to determine the 3D structure. The sequence analyses predict transmembrane helices at the N- and C-termini regions and a mostly  $\alpha$ -helical secondary structure with several  $\beta$ -strands. The sequence analyses and proteolysis experiments demonstrate that CapA has a disordered region in the hydrophilic domain at residues 268–316 that becomes more ordered at high ionic strength. CapA probably interacts with CapB and CapC through this disordered region; these interactions are currently being investigated in our laboratory. CapA is shown to be an inner membrane protein with a topology characteristic of polysaccharide co-polymerases. On the basis of all of these properties, we propose CapA to be a co-polymerase involved in polysaccharide assembly in the cell, with loss of this function contributing to the observed loss of virulence when the *capBCA* locus is deleted.

## ASSOCIATED CONTENT

### Supporting Information

Material and methods, western blot analysis of the detergent screening of the CapA-GFP protein, SDS-12% PAGE of the CapA-GFP protein after purification in the presence of  $\beta$ -DM with Ni-IMAC, size distribution of the CapA-GFP protein in the presence of CHAPS detergent at different protein concentrations, far-UV CD experiment of CapA-GFP at different pH values, far-UV CD experiments of the folding reporter GFP, and predictions of intrinsically disordered regions of PCP proteins. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Author

\*Telephone: (480) 965-8241. Fax: (480) 965-2747. E-mail: JAllen@asu.edu.

### Present Address

<sup>†</sup>M.D.R.: Ventana Medical Systems, Inc., 1910 East Innovation Drive, Tucson, Arizona 85755, United States.

### Author Contributions

Kathryn F. Sykes supervised the GFP fusion constructs. Mark D. Robida did cloning and expression of the CapA membrane protein in pRSET-natGFPHis. Andrey V. Loskutov designed and constructed the Waldo-linker-GFP insert and was involved in the construction of the pRSET-natGFP vector. Felicia M. Craciunescu cloned the GFP-His into pRSET B vector. Debra T. Hansen and Andrey V. Loskutov co-wrote the methods

sections describing the clones. James Zook purified the GFP-His protein. Rebekka M. Wachter, Debra T. Hansen, and James Zook participated in revising the article critically for important intellectual content. Jose M. Martin-Garcia, Petra Fromme, and James P. Allen contributed to the conception and design of the experiments. Jose M. Martin-Garcia acquired and analyzed the data. Jose M. Martin-Garcia, James P. Allen, and Petra Fromme contributed equally to the interpretation of the data. Jose M. Martin-Garcia and James P. Allen contributed equally to the writing of the manuscript, with comments from all authors.

### Funding

This work was funded by National Institutes of Health award U54 GM094599, the center for Membrane Proteins in Infectious Diseases (MPID), which is funded as part of the NIH PSI:Biology Initiative.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We thank all of the staff at the MPID Center (Arizona State University, Tempe, USA), particularly, Tina Esquerra, the manager of the Center, and Katerina Dörner for her many helpful comments and suggestions.

## ABBREVIATIONS

CMC, critical micelle concentration; SDS, sodium dodecyl sulfate; GFP, green fluorescent protein; frGFP, folding reporter green fluorescent protein; SEC, size-exclusion chromatography; MALDI-TOF, matrix assisted laser desorption ionization time of flight; SA, sinapinic acid; PVDF, poly(vinylidene difluoride); DLS, dynamic light scattering; CD, circular dichroism; TM, transmembrane; LPS, lipopolysaccharide; PGA, poly- $\gamma$ -glutamate; LVS, live vaccine strain; CPS, capsule polysaccharides; OMP, outer membrane protein; IMP, inner membrane protein; PCP, polysaccharide co-polymerases

## REFERENCES

- (1) Ellis, J., Oyston, P. C., Green, M., and Titball, R. W. (2002) Tularemia. *Clin. Microbiol. Rev.* 15, 631–646.
- (2) Oyston, P. C., Sjostedt, A., and Titball, R. W. (2004) Tularemia: Bioterrorism defence renews interest in *Francisella tularensis*. *Nat. Rev. Microbiol.* 2, 967–978.
- (3) Sjostedt, A. (2007) Tularemia: History, epidemiology, pathogen physiology, and clinical manifestations. *Ann. N.Y. Acad. Sci.* 1105, 1–29.
- (4) Jones, C. L., Napier, B. A., Sampson, T. R., Llewellyn, A. C., Schroeder, M. R., and Weiss, D. S. (2012) Subversion of host recognition and defense systems by *Francisella* spp. *Microbiol. Mol. Biol. Rev.* 76, 383–404.
- (5) Moreau, G. B., and Mann, B. J. (2013) Adherence and uptake of *Francisella* into host cells. *Virulence* 4, 826–832.
- (6) Su, J., Asare, R., Yang, J., Nair, M. K., Mazurkiewicz, J. E., Abu-Kwaik, Y., and Zhang, J. R. (2011) The *capBCA* locus is required for intracellular growth of *Francisella tularensis* LVS. *Front. Microbiol.* 2, 83–1–83–13.
- (7) Su, J., Yang, J., Zhao, D., Kawula, T. H., Banas, J. A., and Zhang, J. R. (2007) Genome-wide identification of *Francisella tularensis* virulence determinants. *Infect. Immun.* 75, 3089–3101.
- (8) Apicella, M. A., Post, D. M., Fowler, A. C., Jones, B. D., Rasmussen, J. A., Hunt, J. R., Imagawa, S., Choudhury, B., Inzana, T. J., Maier, T. M., Frank, D. W., Zahrt, T. C., Chaloner, K., Jennings, M. P., McLendon, M. K., and Gibson, B. W. (2010) Identification, characterization and immunogenicity of an O-antigen capsular polysaccharide of *Francisella tularensis*. *PLoS One* 5, e11060-1–e11060-18.

- (9) Bandara, A. B., Champion, A. E., Wang, X., Berg, G., Apicella, M. A., McLendon, M., Azadi, P., Snyder, D. S., and Inzana, T. J. (2011) Isolation and mutagenesis of a capsule-like complex (CLC) from *Francisella tularensis*, and contribution of the CLC to *F. tularensis* virulence in mice. *PLoS One* 6, e19003-1–e19003-14.
- (10) Bos, M. P., Robert, V., and Tommassen, J. (2007) Biogenesis of the Gram-negative bacterial outer membrane. *Annu. Rev. Microbiol.* 61, 191–214.
- (11) Gunn, J. S., and Ernst, R. K. (2007) The structure and function of *Francisella* lipopolysaccharide. *Ann. N.Y. Acad. Sci.* 1105, 202–218.
- (12) Huntley, J. F., Conley, P. G., Hagman, K. E., and Norgard, M. V. (2007) Characterization of *Francisella tularensis* outer membrane proteins. *J. Bacteriol.* 189, 561–574.
- (13) Ludu, J. S., de Bruin, O. M., Duplantis, B. N., Schmerk, C. L., Chou, A. Y., Elkins, K. L., and Nano, F. E. (2008) The *Francisella* pathogenicity island protein PdpD is required for full virulence and associates with homologues of the type VI secretion system. *J. Bacteriol.* 190, 4584–4595.
- (14) Melillo, A., Sledjeski, D. D., Lipski, S., Wooten, R. M., Basrur, V., and Lafontaine, E. R. (2006) Identification of a *Francisella tularensis* LVS outer membrane protein that confers adherence to A549 human lung cells. *FEMS Microbiol. Lett.* 263, 102–108.
- (15) Nano, F. E. (1988) Identification of a heat-modifiable protein of *Francisella tularensis* and molecular cloning of the encoding gene. *Microb. Pathog.* 5, 109–119.
- (16) Sandstrom, G., Lofgren, S., and Tarnvik, A. (1988) A capsule-deficient mutant of *Francisella tularensis* LVS exhibits enhanced sensitivity to killing by serum but diminished sensitivity to killing by polymorphonuclear leukocytes. *Infect. Immun.* 56, 1194–1202.
- (17) Jia, Q., Lee, B. Y., Bowen, R., Dillon, B. J., Som, S. M., and Horwitz, M. A. (2010) A *Francisella tularensis* live vaccine strain (LVS) mutant with a deletion in capB, encoding a putative capsular biosynthesis protein, is significantly more attenuated than LVS yet induces potent protective immunity in mice against *F. tularensis* challenge. *Infect. Immun.* 78, 4341–4355.
- (18) Michell, S. L., Dean, R. E., Eyles, J. E., Hartley, M. G., Waters, E., Prior, J. L., Titball, R. W., and Oyston, P. C. (2010) Deletion of the *Bacillus anthracis* capB homologue in *Francisella tularensis* subspecies tularensis generates an attenuated strain that protects mice against virulent tularemia. *J. Med. Microbiol.* 59, 1275–1284.
- (19) Weiss, D. S., Brotcke, A., Henry, T., Margolis, J. J., Chan, K., and Monack, D. M. (2007) *In vivo* negative selection screen identifies genes required for *Francisella* virulence. *Proc. Natl. Acad. Sci. U.S.A.* 104, 6037–6042.
- (20) Larsson, P., Oyston, P. C., Chain, P., Chu, M. C., Duffield, M., Fuxelius, H. H., Garcia, E., Halltorp, G., Johansson, D., Isherwood, K. E., Karp, P. D., Larsson, E., Liu, Y., Michell, S., Prior, J., Prior, R., Malfatti, S., Sjostedt, A., Svensson, K., Thompson, N., Vergez, L., Wagg, J. K., Wren, B. W., Lindler, L. E., Andersson, S. G., Forsman, M., and Titball, R. W. (2005) The complete genome sequence of *Francisella tularensis*, the causative agent of tularemia. *Nat. Genet.* 37, 153–159.
- (21) Ashiuchi, M., Soda, K., and Misono, H. (1999) A poly-gamma-glutamate synthetic system of *Bacillus subtilis* IFO 3336: Gene cloning and biochemical analysis of poly-gamma-glutamate produced by *Escherichia coli* clone cells. *Biochem. Biophys. Res. Commun.* 263, 6–12.
- (22) Ashiuchi, M., Nawa, C., Kamei, T., Song, J. J., Hong, S. P., Sung, M. H., Soda, K., and Misono, H. (2001) Physiological and biochemical characteristics of poly gamma-glutamate synthetase complex of *Bacillus subtilis*. *Eur. J. Biochem.* 268, 5321–5328.
- (23) Candela, T., and Fouet, A. (2006) Poly-gamma-glutamate in bacteria. *Mol. Microbiol.* 60, 1091–1098.
- (24) Kocianova, S., Vuong, C., Yao, Y., Voyich, J. M., Fischer, E. R., DeLeo, F. R., and Otto, M. (2005) Key role of poly-gamma-DL-glutamic acid in immune evasion and virulence of *Staphylococcus epidermidis*. *J. Clin. Invest.* 115, 688–694.
- (25) Hou, S., Saw, J. H., Lee, K. S., Freitas, T. A., Belisle, C., Kawarabayasi, Y., Donachie, S. P., Pikina, A., Galperin, M. Y., Koonin, E. V., Makarova, K. S., Omelchenko, M. V., Sorokin, A., Wolf, Y. I., Li, Q. X., Keum, Y. S., Campbell, S., Denery, J., Aizawa, S., Shibata, S., Malahoff, A., and Alam, M. (2004) Genome sequence of the deep-sea gamma-proteobacterium *Idiomarina loihiensis* reveals amino acid fermentation as a source of carbon and energy. *Proc. Natl. Acad. Sci. U.S.A.* 101, 18036–18041.
- (26) Glockner, F. O., Kube, M., Bauer, M., Teeling, H., Lombardot, T., Ludwig, W., Gade, D., Beck, A., Borzym, K., Heitmann, K., Rabus, R., Schlesner, H., Amann, R., and Reinhardt, R. (2003) Complete genome sequence of the marine planctomycete *Pirellula* sp. strain 1. *Proc. Natl. Acad. Sci. U.S.A.* 100, 8298–8303.
- (27) Ren, S. X., Fu, G., Jiang, X. G., Zeng, R., Miao, Y. G., Xu, H., Zhang, Y. X., Xiong, H., Lu, G., Lu, L. F., Jiang, H. Q., Jia, J., Tu, Y. F., Jiang, J. X., Gu, W. Y., Zhang, Y. Q., Cai, Z., Sheng, H. H., Yin, H. F., Zhang, Y., Zhu, G. F., Wan, M., Huang, H. L., Qian, Z., Wang, S. Y., Ma, W., Yao, Z. J., Shen, Y., Qiang, B. Q., Xia, Q. C., Guo, X. K., Danchin, A., Saint Girons, I., Somerville, R. L., Wen, Y. M., Shi, M. H., Chen, Z., Xu, J. G., and Zhao, G. P. (2003) Unique physiological and pathogenic features of *Leptospira interrogans* revealed by whole-genome sequencing. *Nature* 422, 888–893.
- (28) Takami, H., Takaki, Y., and Uchiyama, I. (2002) Genome sequence of *Oceanobacillus iheyensis* isolated from the Iheya Ridge and its unexpected adaptive capabilities to extreme environments. *Nucleic Acids Res.* 30, 3927–3935.
- (29) Nonaka, H., Keresztes, G., Shinoda, Y., Ikenaga, Y., Abe, M., Naito, K., Inatomi, K., Furukawa, K., Inui, M., and Yukawa, H. (2006) Complete genome sequence of the dehalorespiring bacterium *Desulfobacterium hafniense* Y51 and comparison with *Dehalococcoides ethenogenes* 19S. *J. Bacteriol.* 188, 2262–2274.
- (30) Waldo, G. S., Standish, B. M., Berendzen, J., and Terwilliger, T. C. (1999) Rapid protein-folding assay using green fluorescent protein. *Nat. Biotechnol.* 17, 691–695.
- (31) Drew, D., Lerch, M., Kunji, E., Slotboom, D. J., and de Gier, J. W. (2006) Optimization of membrane protein overexpression and purification using GFP fusions. *Nat. Methods* 3, 303–313.
- (32) Daley, D. O., Rapp, M., Granseth, E., Melen, K., Drew, D., and von Heijne, G. (2005) Global topology analysis of the *Escherichia coli* inner membrane proteome. *Science* 308, 1321–1323.
- (33) Drew, D., Sjostrand, D., Nilsson, J., Urbig, T., Chin, C. N., de Gier, J. W., and von Heijne, G. (2002) Rapid topology mapping of *Escherichia coli* inner-membrane proteins by prediction and PhoA/GFP fusion analysis. *Proc. Natl. Acad. Sci. U.S.A.* 99, 2690–2695.
- (34) Feilmeier, B. J., Isemlinger, G., Schroeder, D., Webber, H., and Phillips, G. J. (2000) Green fluorescent protein functions as a reporter for protein localization in *Escherichia coli*. *J. Bacteriol.* 182, 4068–4076.
- (35) Rapp, M., Drew, D., Daley, D. O., Nilsson, J., Carvalho, T., Melen, K., De Gier, J. W., and Von Heijne, G. (2004) Experimentally based topology models for *E. coli* inner membrane proteins. *Protein Sci.* 13, 937–945.
- (36) Arnold, T., and Linke, D. (2008) The use of detergents to purify membrane proteins. *Curr. Protoc. Protein Sci.* 4, 1–30.
- (37) Drew, D., Newstead, S., Sonoda, Y., Kim, H., von Heijne, G., and Iwata, S. (2008) GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in *Saccharomyces cerevisiae*. *Nat. Protoc.* 3, 784–798.
- (38) Garnier, J., Gibrat, J. F., and Robson, B. (1996) GOR method for predicting protein secondary structure from amino acid sequence. *Methods Enzymol.* 266, 540–553.
- (39) Jones, D. T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 292, 195–202.
- (40) Petersen, B., Petersen, T. N., Andersen, P., Nielsen, M., and Lundgaard, C. (2009) A generic method for assignment of reliability scores applied to solvent accessibility predictions. *BMC Struct. Biol.* 9, S1-1–S1-10.
- (41) Pollastri, G., and McLysaght, A. (2005) Porter: A new, accurate server for protein secondary structure prediction. *Bioinformatics* 21, 1719–1720.
- (42) Oliver, R. C., Lipfert, J., Fox, D. A., Lo, R. H., Doniach, S., and Columbus, L. (2013) Dependence of micelle size and shape on

- detergent alkyl chain length and head group. *PLoS One* 8, e62488-1–e62488-10.
- (43) McGuire, A. M., Matsuo, H., and Wagner, G. (1998) Internal and overall motions of the translation factor eIF4E: Cap binding and insertion in a CHAPS detergent micelle. *J. Biomol. NMR* 12, 73–88.
- (44) Sreerama, N., and Woody, R. W. (2000) Estimation of protein secondary structure from circular dichroism spectra: Comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem.* 287, 252–260.
- (45) Visser, N. V., Hink, M. A., Borst, J. W., van der Krogt, G. N., and Visser, A. J. (2002) Circular dichroism spectroscopy of fluorescent proteins. *FEBS Lett.* 521, 31–35.
- (46) Locher, K. P., and Rosenbusch, J. P. (1997) Oligomeric states and siderophore binding of the ligand-gated FhuA protein that forms channels across *Escherichia coli* outer membranes. *Eur. J. Biochem.* 247, 770–775.
- (47) Murphy, C. K., Kalve, V. I., and Klebba, P. E. (1990) Surface topology of the *Escherichia coli* K-12 ferric enterobactin receptor. *J. Bacteriol.* 172, 2736–2746.
- (48) Jap, B. K., Downing, K. H., and Walian, P. J. (1990) Structure of PhoE porin in projection at 3.5 Å resolution. *J. Struct. Biol.* 103, 57–63.
- (49) Taylor, R., Burgner, J. W., Clifton, J., and Cramer, W. A. (1998) Purification and characterization of monomeric *Escherichia coli* vitamin B12 receptor with high affinity for colicin E3. *J. Biol. Chem.* 273, 31113–31118.
- (50) Xue, B., Dunbrack, R. L., Williams, R. W., Dunker, A. K., and Uversky, V. N. (2010) PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. *Biochim. Biophys. Acta* 1804, 996–1010.
- (51) Xue, B., Li, L., Meroueh, S. O., Uversky, V. N., and Dunker, A. K. (2009) Analysis of structured and intrinsically disordered regions of transmembrane proteins. *Mol. BioSyst.* 5, 1688–1702.
- (52) Uversky, V. N., Gillespie, J. R., and Fink, A. L. (2000) Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins* 41, 415–427.
- (53) Nishimura, C., Uversky, V. N., and Fink, A. L. (2001) Effect of salts on the stability and folding of staphylococcal nuclease. *Biochemistry* 40, 2113–2128.
- (54) Uversky, V. N., Karnoupp, A. S., Khurana, R., Segel, D. J., Doniach, S., and Fink, A. L. (1999) Association of partially-folded intermediates of staphylococcal nuclease induces structure and stability. *Protein Sci.* 8, 161–173.
- (55) Uversky, V. N., Li, J., and Fink, A. L. (2001) Metal-triggered structural transformations, aggregation, and fibrillation of human alpha-synuclein. A possible molecular NK between Parkinson’s disease and heavy metal exposure. *J. Biol. Chem.* 276, 44284–44296.
- (56) Tompa, P. (2002) Intrinsically unstructured proteins. *Trends Biochem. Sci.* 27, 527–533.
- (57) Fontana, A., Fassina, G., Vita, C., Dalzoppo, D., Zamai, M., and Zambonin, M. (1986) Correlation between sites of limited proteolysis and segmental mobility in thermolysin. *Biochemistry* 25, 1847–1851.
- (58) Fontana, A., Polverino de Laureto, P., De Filippis, V., Scaramella, E., and Zambonin, M. (1997) Probing the partly folded states of proteins by limited proteolysis. *Folding Des.* 2, R17–R26.
- (59) Fontana, A., Zambonin, M., Polverino de Laureto, P., De Filippis, V., Clementi, A., and Scaramella, E. (1997) Probing the conformational state of apomyoglobin by limited proteolysis. *J. Mol. Biol.* 266, 223–230.
- (60) Polverino de Laureto, P., De Filippis, V., Di Bello, M., Zambonin, M., and Fontana, A. (1995) Probing the molten globule state of alpha-lactalbumin by limited proteolysis. *Biochemistry* 34, 12596–12604.
- (61) Polverino de Laureto, P., Toma, S., Tonon, G., and Fontana, A. (1995) Probing the structure of human growth hormone by limited proteolysis. *Int. J. Pept. Protein Res.* 45, 200–208.
- (62) Csizmok, V., Bokor, M., Banki, P., Klement, E., Medzihradzky, K. F., Friedrich, P., Tompa, K., and Tompa, P. (2005) Primary contact sites in intrinsically unstructured proteins: The case of calpastatin and microtubule-associated protein 2. *Biochemistry* 44, 3955–3964.
- (63) Denning, D. P., Patel, S. S., Uversky, V., Fink, A. L., and Rexach, M. (2003) Disorder in the nuclear pore complex: The FG repeat regions of nucleoporins are natively unfolded. *Proc. Natl. Acad. Sci. U.S.A.* 100, 2450–2455.
- (64) Marin, M., Thallmair, V., and Ott, T. (2012) The intrinsically disordered N-terminal region of AtREM1.3 remorin protein mediates protein-protein interactions. *J. Biol. Chem.* 287, 39982–39991.
- (65) Nyarko, A., Hare, M., Hays, T. S., and Barbar, E. (2004) The intermediate chain of cytoplasmic dynein is partially disordered and gains structure upon binding to light-chain LC8. *Biochemistry* 43, 15595–15603.
- (66) Fink, A. L., Calciano, L. J., Goto, Y., Kurotsu, T., and Palleros, D. R. (1994) Classification of acid denaturation of proteins: Intermediates and unfolded states. *Biochemistry* 33, 12504–12511.
- (67) Goto, Y., Calciano, L. J., and Fink, A. L. (1990) Acid-induced folding of proteins. *Proc. Natl. Acad. Sci. U.S.A.* 87, 573–577.
- (68) Goto, Y., and Fink, A. L. (1990) Phase diagram for acidic conformational states of apomyoglobin. *J. Mol. Biol.* 214, 803–805.
- (69) Goto, Y., Takahashi, N., and Fink, A. L. (1990) Mechanism of acid-induced folding of proteins. *Biochemistry* 29, 3480–3488.
- (70) Dunker, A. K., Cortese, M. S., Romero, P., Iakoucheva, L. M., and Uversky, V. N. (2005) Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J.* 272, 5129–5148.
- (71) Uversky, V. N. (2002) Cracking the folding code. Why do some proteins adopt partially folded conformations, whereas other don’t? *FEBS Lett.* 514, 181–183.
- (72) Uversky, V. N. (2002) Natively unfolded proteins: A point where biology waits for physics. *Protein Sci.* 11, 739–756.
- (73) Uversky, V. N. (2002) What does it mean to be natively unfolded? *Eur. J. Biochem.* 269, 2–12.
- (74) Uversky, V. N. (2011) Multitude of binding modes attainable by intrinsically disordered proteins: A portrait gallery of disorder-based complexes. *Chem. Soc. Rev.* 40, 1623–1634.
- (75) Dresler, J., Klimentova, J., and Stulik, J. (2011) *Francisella tularensis* membrane complexome by blue native/SDS-PAGE. *J. Proteomics* 75, 257–269.
- (76) Whitfield, C. (2006) Biosynthesis and assembly of capsular polysaccharides in *Escherichia coli*. *Annu. Rev. Biochem.* 75, 39–68.
- (77) Morona, R., Purins, L., Tocilj, A., Matte, A., and Cygler, M. (2009) Sequence-structure relationships in polysaccharide co-polymerase (PCP) proteins. *Trends Biochem. Sci.* 34, 78–84.
- (78) Grangeasse, C., Cozzzone, A. J., Deutscher, J., and Mijakovic, I. (2007) Tyrosine phosphorylation: an emerging regulatory device of bacterial physiology. *Trends Biochem. Sci.* 32, 86–94.
- (79) Morona, R., Van Den Bosch, L., and Daniels, C. (2000) Evaluation of Wzz/MPA1/MPA2 proteins based on the presence of coiled-coil regions. *Microbiology* 146, 1–4.
- (80) Tocilj, A., Munger, C., Proteau, A., Morona, R., Purins, L., Ajamian, E., Wagner, J., Papadopoulos, M., Van Den Bosch, L., Rubinstein, J. L., Fethiere, J., Matte, A., and Cygler, M. (2008) Bacterial polysaccharide co-polymerases share a common framework for control of polymer length. *Nat. Struct. Mol. Biol.* 15, 130–138.
- (81) Bastin, D. A., Stevenson, G., Brown, P. K., Haase, A., and Reeves, P. R. (1993) Repeat unit polysaccharides of bacteria: A model for polymerization resembling that of ribosomes and fatty acid synthetase, with a novel mechanism for determining chain length. *Mol. Microbiol.* 7, 725–734.
- (82) Paulsen, I. T., Park, J. H., Choi, P. S., and Saier, M. H., Jr. (1997) A family of gram-negative bacterial outer membrane factors that function in the export of proteins, carbohydrates, drugs and heavy metals from gram-negative bacteria. *FEMS Microbiol. Lett.* 156, 1–8.
- (83) Kyte, J., and Doolittle, R. F. (1982) A simple method for displaying the hydrophobic character of a protein. *J. Mol. Biol.* 157, 105–132.