Identification of Residues in DmsD for Twin-Arginine Leader Peptide Binding, Defined through Random and Bioinformatics-Directed Mutagenesis[†]

Catherine S. Chan,[‡,§] Tara M. L. Winstone,[‡,§] Limei Chang,[§] Charles M. Stevens,[§,||] Matthew L. Workentine,[§] Haiming Li,[§] Ying Wei,[⊥] Mary J. Ondrechen,[⊥] Mark Paetzel,[||] and Raymond J. Turner*[,§]

Department of Biological Sciences, 2500 University Drive Northwest, University of Calgary, Calgary, Alberta T2N 1N4, Canada, Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada, and Department of Chemistry and Chemical Biology, Northeastern University, Boston, Massachusetts 02115

Received October 23, 2007; Revised Manuscript Received January 7, 2008

ABSTRACT: The twin-arginine translocase (Tat) system is used for the targeting and translocation of folded proteins across the cell membrane of most bacteria. Substrates of this system contain a conserved "twin-arginine" (RR) motif within their signal/leader peptide sequence. Many Tat substrates have their own system-specific chaperone called redox enzyme maturation proteins (REMPs). Here, we study the binding of DmsD, the REMP for dimethyl sulfoxide reductase in *Escherichia coli*, toward the RR-containing leader peptide of the catalytic subunit DmsA. We have used a multipronged approach targeted at the amino acid sequence of DmsD to define residues and regions important for recognition of the DmsA leader sequence. Residues identified through bioinformatics and THEMATICS analysis were mutated using site-directed mutagenesis. These DmsD residue variants were purified and screened with an *in vitro* dot-blot far-Western assay to analyze the binding to the DmsA leader sequence. Degenerative polymerase chain reaction was also used to produce a bank of random DmsD amino acid mutants, which were then screened by an *in vivo* bacterial two-hybrid assay. Using this hybrid method, each DmsD variant was classified into one of three groups based on their degree of interaction with the DmsA leader (none, weak, and moderate). The data from both the *in vitro* and *in vivo* analyses were then applied to a model structure of DmsD based on the crystal structure of the *Salmonella typhimurium* homologue. Our results illustrate the positions of important DmsD residues involved in binding the DmsA leader peptide and identify a "hot pocket" of residues important for leader binding on the structure of DmsD.

Mapping the site of protein interactions can aid in the understanding of the function of a protein. While structural techniques, such as X-ray crystallography, can provide a visual representation of the residues involved in protein interactions, this technique is not always amenable to every system. Assuming that an appropriate screening system is available, mutagenesis of the protein(s) of interest is a useful method to identify residues or regions involved in the interactions. Mutagenesis can be site-specific or random, and the choice between the two depends upon the amount of background information that is available. Site-specific mutagenesis requires enough information to predict residues that are suspected to be involved in binding. The random approach is often useful when little background information is available. Characterizing the DmsD/DmsA leader interaction has been a focus of our research (*1–3*). In this study, we describe our findings in identifying residues in DmsD important for DmsA leader binding using site-specific or random mutagenesis followed by *in vitro* dot-blot far-Western or *in vivo* bacterial two-hybrid screens. Furthermore, the success of using primary sequence analysis (bioinformatics) or tertiary structure analysis [electrostatics-based theoretical microscopic titration curves (THEMATICS)[1] (*4*)] of a modeled structure of DmsD to predict residues important for binding is compared.

Dimethyl sulfoxide (DMSO) reductase is an enzyme present in many bacteria and is involved in anaerobic respiration using DMSO as its final electron acceptor (reviewed in ref *5*). DMSO is reduced to DMS in the following reaction: $Me_2SO + 2e^- + 2H^+ \leftrightarrow Me_2S + H_2O$. DMSO reductase is a heterotrimeric enzyme consisting of the DmsABC subunits. DmsA is the catalytic subunit containing a molybdopterin (MoPt) cofactor, and DmsB is the electron-transfer subunit containing four [4Fe−4S] centers. The DmsAB dimer is targeted toward the cytoplas-

[†] This work was supported by Canadian Institute of Health Research grants to R.J.T. and M.P. M.P. is also supported by the National Science and Engineering Research Council of Canada (NSERC), the Canadian Foundation of Innovation, and the Michael Smith Foundation for Health Research. T.M.L.W. and C.S.C. are funded by CGS-D and PGS-D scholarships, respectively, from NSERC.

* To whom correspondence should be addressed: Department of Biological Sciences, BI 156 Biological Sciences Building, University of Calgary, 2500 University Dr. NW, Calgary, Alberta T2N 1N4, Canada. Telephone: 1-403-220-3581. Fax: 1-403-220-9311. E-mail: turnerr@ucalgary.ca.

[‡] Both authors to be considered as the first author.

[§] University of Calgary.

[||] Simon Fraser University.

[⊥] Northeastern University.

[1] Abbreviations: $DmsA_L$, DmsA leader peptide; DMSO, dimethyl sulfoxide; REMP, redox enzyme maturation protein; Tat, twin-arginine translocase; THEMATICS, theoretical microscopic titration curves; WT, wild type.

10.1021/bi702138a CCC: $40.75
Published on Web 02/05/2008

mic membrane via the twin-arginine translocation (Tat) system, where it attaches to the membrane-anchor subunit, DmsC (*6*). *Escherichia coli* DmsA contains a SRRgLVK sequence in its N-terminal leader peptide that has been identified as the twin-arginine motif that serves as the targeting signal toward the Tat translocon [the identified conserved motif for Tat-dependent substrates is S/TRRxFLK (*7*)]. It was found that the leader peptide of DmsA is cleaved during assembly, and its cleavage requires the presence of the Tat apparatus (*6, 8*).

E. coli DmsD was identified as a system-specific chaperone involved in the biogenesis of DMSO reductase that binds the RR-containing leader peptide of DmsA (*1, 9*). Further studies showed that DmsD interacts with the cytoplasmic membrane of *E. coli* under anaerobic conditions and its localization is likely through association with the TatBC subunits of the Tat apparatus (*10*). The above results also implicate a targeting role for DmsD. Through bioinformatic sequence analyses, DmsD was found to belong to a group of proteins considered system-specific chaperones involved in assembly, maturation, folding, and targeting of bacterial redox enzymes and has been given a collective term of redox enzyme maturation proteins (REMPs) (*11*). The DmsD *E. coli* REMPs is a member of the TorD superfamily of proteins, which include TorD, YcdY, and NarJ. TorD is the REMP for trimethylamine-*N*-oxide reductase, and NarJ is the REMP for the cytoplasmic nitrate reductase, while the function of YcdY remains unknown. Both TorD and NarJ have been shown to bind the N-terminal leader/region of the subunits TorA and NarG, respectively (*12–14*).

Using the site-specific mutagenesis and *in vitro* dot-blot far-Western approach, we identify 11 residues within DmsD that are important for DmsA twin-arginine leader peptide binding. Upon mapping these residues onto an *E. coli* DmsD modeled structure, we found that most of the residues identified in this study are clustered in a "hot pocket" on the surface of the molecule. A comparison of the methods used to predict the target mutations revealed that all, except two residues that were important for binding, were predicted by bioinformatic analysis. THEMATICS, an electrostatics-based analysis, predicted two residues important for binding, plus an additional unique residue for binding not predicted by sequence alignment. The random mutagenesis and *in vivo* bacterial two-hybrid approach identified several residues that appear to be important for DmsA leader binding, many of which were common with those observed identified by computational approaches. Through a combination of these methods, we now have the identification of many residues involved in a putative leader peptide-binding site on DmsD.

EXPERIMENTAL PROCEDURES

Molecular Visualizations. To date, structures of several REMP homologues to DmsD that are available in the Protein Data Bank include *Shewanella massilia* (SmTorD) 1n1c at 2.4 Å, *Salmonella typhimurium* LT2 DmsD (StDmsD) 1s9u at 1.38 Å resolution, and *Archaeoglobus fulgidus* DSM4304 (AfDmsD) 2o9X at 3.40 Å and AF0160 2idg at 2.69 Å. The best homology with the *E. coli* protein studied here was StDmsD (1s9u), which shows 77.8% sequence identity and greater than 90% sequence similarity. The only challenge was a six-residue region (117–122) having no electron density in the StDmsD structure. Homology models were produced by SWISS-MODEL (*15*) and CPHmodels 2.0 (*16*), with the six-residue portion generated by ProModII in SWISS-MODEL. The models were judged by the evaluation programs Verify 3D (*17, 18*), Prosa2 (*19*), and Anolea (*20*). Structural alignment of EcDmsD and StDmsD was performed in VMD (*21*). The surface/binding-pocket analysis was carried out using CASTp (*22*) with a 1.4 Å probe radius.

On the basis of the identification of important residues in the procedures below, the type of functional effect was colored accordingly and structures were rendered using Pymol (*23*). Prior sequence entries of DmsD in databanks, such as GenBank of National Center for Biotechnology Information (NCBI) (P69853), included a three-residue addition of MLG at the N terminus, whereas the correct sequence was verified by N-terminal sequencing to begin at MTH (*1*). As of May 2007, the common databanks, such as Genbank and EcoGene (EG13847), appear to have the corrected DmsD sequence. All structural demonstrations and references to residues used in this paper are numbered according to the correct DmsD sequence starting with residues MTH.

Bioinformatic and THEMATICS Analysis of DmsD. Sequence alignment of the REMPs DmsD, TorD, NarJ, and YcdY in our previous study identified two key sequence motifs (*11*). Sequence alignment of only DmsD sequences was performed using the online algorithm Mafft version 5.23 (*24*) using default settings. Alignment containing secondary-structure information as in Figure 2 was performed using ESPript version 2.2 (*25*) using the alignment file generated from Mafft and the PDB file of modeled EcDmsD. Similarity calculations based on each column of aligned residues was performed as %Equivalent (global score of 0.8 and difference score of 0.5) in ESPript.

THEMATICS (*4, 26, 27*) computations were performed on the template PDB 1s9u and on the model structure obtained for *E. coli* DmsD. The hydrogen atoms were built into the structure using TINKER (*28*) and the OPLS-UA force field (*29, 30*). The dielectric constants were assumed to be 20 for the protein interior and 80 for the solvent. The finite difference Poisson–Boltzmann procedure embedded in the University of Houston Brownian Dynamics (UHBD) program (*31*) was used to obtain the electrical potential function. Theoretical titration curves, expressed as the average charge C as a function of pH, were computed for each ionizable residue by the program HYBRID (*32*). Positive residues were selected using Ko's method (*4*); Z scores were computed with a cutoff value of 0.99 (*27*). Note that THEMATICS examines ionizable residues (Arg, Asp, Cys, Glu, His, Lys, Tyr, and the N and C termini) only. Identified residues were then assigned to clusters, where any positive residue is a cluster member if it is within 9 Å of another cluster member. For purposes of cluster definition, distances between residues are measured between the side-chain atoms that represent the center of charge on each residue. Clusters containing two or more residues constitute predictions of local catalytic and/or binding sites.

Plasmid Constructs and Site-Directed Mutagenesis of DmsD. A DmsA leader peptide (residues 1−43) glutathione *S*-transferase fusion protein ($DmsA_L$/GST) was produced from pTDMS24 (*1*). Single-residue mutations in His_6-T_7/DmsD were generated using Quickchange II site-directed

mutagenesis kit (Stratagene,) with 25 ng of pTDMS67 (*2*) as the template following instructions of the manufacturer and primers described in Supplementary Table S2 in the Supporting Information. The residue choice for single mutations were selected on the basis of a guide for "safe" substitutions for site-directed mutagenesis (*33*). The resulting recombinant plasmids (Supplementary Table S2 in the Supporting Information) were isolated using a Plasmid Midi kit (Qiagen) and verified by sequencing (University Core DNA Services, University of Calgary, Calgary, Canada) prior to transformation into *E. coli* C41(DE3) cells (*34*).

Dot-Blot Far-Western Detection of DmsD Mutants Binding to DmsA Leader. Cells expressing $DmsA_L$/GST or His_6-T_7/DmsD [wild type (WT) or mutant] were grown, and each protein was purified as described in ref *2*. The terminology $DmsA_L$ or DmsA leader infers that $DmsA_L$/GST was used, while DmsD infers that His_6-T_7/DmsD was used. Frozen DmsD protein was thawed in a room-temperature water bath for 5 min, vortexed, and centrifuged at 10000*g* for 30 min at 4 °C to remove any precipitated material. DmsD protein concentrations were determined with a Bradford assay (BioRad) and diluted to 0.25 mg/mL with TBS buffer (50 mM Tris-HCl at pH 8.0 and 200 mM NaCl). DmsD samples were serially diluted 1.5-fold across 8 rows of a 96-well microtiter plate and then applied to a prewet nitrocellulose membrane (0.45 μm, BioRad) using a BioDot apparatus (BioRad) according to the protocol of the manufacturer. Each blot contained one row of WT DmsD, one row of buffer blank (TBS), and six rows of different DmsD variants. Blots were blocked overnight (16–20 h) at room temperature with TBS containing 5% (w/v) skim milk. Blots were washed twice with TBST [50 mM Tris-HCl at pH 8.0, 200 mM NaCl, and 0.05% (v/v) Tween-20] and then twice with TBS. They were incubated with 25 μg/mL $DmsA_L$/GST in TBS for 1 h at room temperature, followed by two washes with TBST prior to incubation with mouse monoclonal anti-GST (1:5000) (Novagen) in TBST containing 1% milk for 1 h. Blots were then washed twice with TBST and incubated with goat-antimouse-HRP conjugate (1:3000; BioRad) in TBST containing 1% milk for 1 h, followed by two washes each with TBST and TBS. Development was performed with a colorimetric HRP substrate kit (BioRad) for image capture and analysis.

Immunoblot Image Capture and Analysis of Pixel Intensities. The developed blots were photographed with a Kodak Gel Logic 100 imaging system. Kodak 1D Image Analysis software was used to determine the intensity of each spot as follows. A template was made in which the region of interest (ROI) was set just outside the area of each spot for all 96 spots of the blot. All 96 ROIs were analyzed to determine the experimental mean pixel intensity (mPI) within the defined area. A background mPI was also determined from the mean of the pixel intensities found within the perimeter of the ROI. The background mPI was subtracted from the experimental mPI to give a corrected mPI. The corrected mPI of the WT with the maximum applied amount (25 μg) was used to normalize all other corrected mPI values.

$$\text{corrected mPI} = \text{experimental mPI} - \text{background mPI}$$

$$DmsA_L \text{ relative binding} = (\text{corrected mPI})_{\text{mutant},n}/(\text{corrected mPI})_{\text{WT,max}}$$

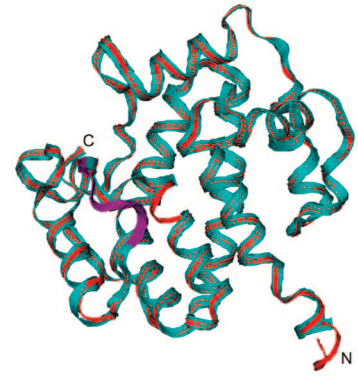

FIGURE 1: Structural alignment of *E. coli* modeled DmsD and the template *S. typhimurium* DmsD to which EcDmsD was modeled. EcDmsD is colored cyan, and StDmsD (1s9u) is colored red. The six-residue portion (117–122) that was not available from the StDmsD model template but was modeled for EcDmsD is shown in purple.

The normalized values or "$DmsA_L$ relative binding" were plotted against the amount of DmsD protein applied to the spot producing a binding curve for each variant protein relative to the WT. Each variant was assessed on 5−10 blots, and the mean binding curve and standard deviation was determined. Sample binding curves are shown in Figure 4A. The mean (and standard deviation) of the relative binding at the highest amount of protein applied (25 μg) was then used to compare all 20 single DmsD residue mutants relative to the WT in a bar graph format (Figure 4B). Each spot on the blots corresponds to ~0.07 cm^2. The maximum amount of protein able to be bound to this area of nitrocellulose is 5.6−7.1 μg (BioRad Blot Guide), and this corresponds well with the saturation point on the binding curves (Figure 4A).

Construction of DmsD Random Mutant Library. A DmsD fusion with the *Bordetella pertussis* adenylate cyclase T25 fragment was generated first by amplifying *dmsD* using pTDMS67 (*2*) as the template and primers TDMS112 and TDMS113 (Supplementary Table S2 in the Supporting Information). The PCR fragment was ligated into pKNT25 (*35, 36*) and verified via sequencing, and the plasmid was named pDmsDT25. The addition of green fluorescent protein (GFP) at the end of the above construct was performed by amplification of *gfp* by polymerase chain reaction (PCR) using the primers GFP1 and GFP2 (Supplementary Table S2 in the Supporting Information) with pHTK1 (*37*) as the template. The PCR product was digested with *Cla*I and ligated into *Cla*I-digested pDmsDT25, creating plasmid pDmsDT25GFP. The GFP was added to prescreen the DmsD random mutant library for frame shifts or stop codons. Thus, the presence of GFP fluorescence would verify that the DmsD sequence was completely translated.

Two DmsD random mutant libraries were generated using a GeneMorph II EZClone Domain Mutagenesis kit (Stratagene). For the two libraries generated, high and low rate, upon sequencing, it was noticed that the high-rate library contained up to 12 mutated residues, while the low-rate library contained 2−3 mutations.

Isolation and Analysis of DmsD Random Mutants That Do Not Interact with the DmsA Leader Based on Bacterial Two-Hybrid Screening. The leader peptide of DmsA fused to the N terminus of the T18 fragment of adenylate cyclase was generated as above using primers TDMS110 and

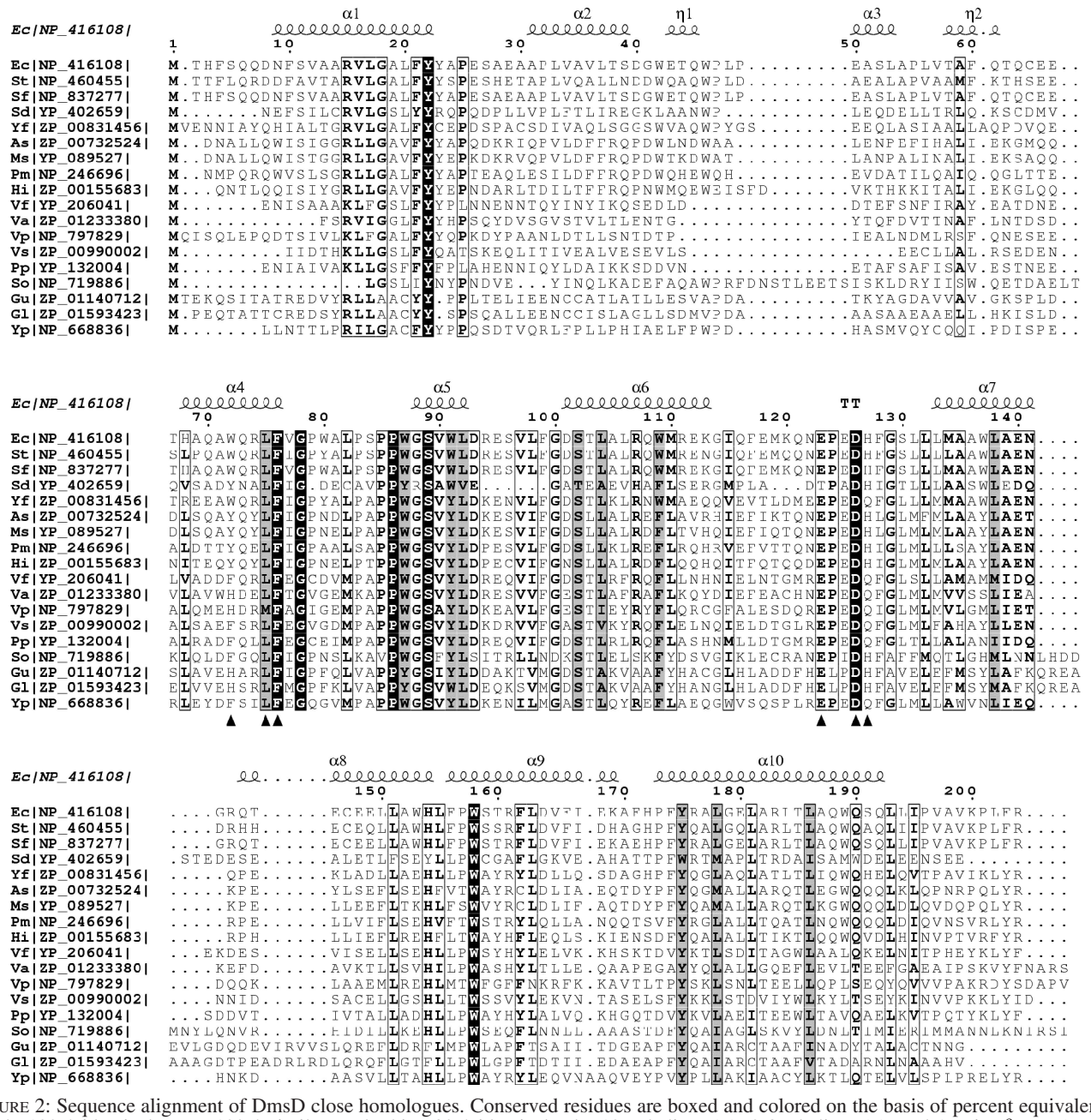

FIGURE 2: Sequence alignment of DmsD close homologues. Conserved residues are boxed and colored on the basis of percent equivalence: black, 100% identical; gray, 100% similar; and white, 80% identical. The 10 α helices (α), 2 3_{10} helices (η), and 1 strict β turn (TT) of EcDmsD based on its modeled structure are indicated above. The REMP conserved motifs WxxLF and E(P/x)(x/P)DH are indicated by black triangles below.

TDMS111 and pTDMS24 (*1*) as the template (Supplementary Table S2 in the Supporting Information). The resulting PCR fragment was ligated into pUT18 (*35, 36*) to generate the plasmid pDmsALT18. The plasmid pDmsALT18 was transformed into the *E. coli* strain BTH101 (a *cya*-deficient strain), followed by secondary transformation of the mutant pDmsDT25GFP plasmid library (prescreened for GFP fluorescence). Cells were plated on MacConkey agar plates containing 1% (w/v) maltose, 100 μg/mL ampicillin, and 50 μg/mL kanamycin and incubated at 30 °C for 3 days. White and pale pink colonies were restreaked onto fresh plates and incubated at 30 °C for 3 days. Plasmids from colonies displaying weakened or the loss of interaction by their inability to develop into red colonies were sequenced.

The isolated plasmids, shown to not interact with the $DmsA_L$ as evidenced by a white or pale pink colony, were then individually retransformed into *E. coli* strain BTH101 already containing pDmsALT18 and plated onto MacConkey agar plates containing maltose, ampicillin, and kanamycin as above. The percentage of colonies remaining white after 3 days were determined and used to categorize the DmsD variants generated. If 100% of the colonies remained white, the variant was categorized as a noninteractor. If >50% of the colonies remained white, then the variant was considered to maintain a "weak" interaction with the $DmsA_L$, while <50% white meant that the mutant DmsD molecule produced a "moderate" interaction with $DmsA_L$. A mutant that produces 0% white colonies (i.e., 100% red) is assumed to have near WT interaction levels.

In Vivo Analysis of DmsA Leader Interactions with Site-Directed DmsD Mutants. Site-directed mutants screened by the dot-blot far-Western were selected for analysis by the

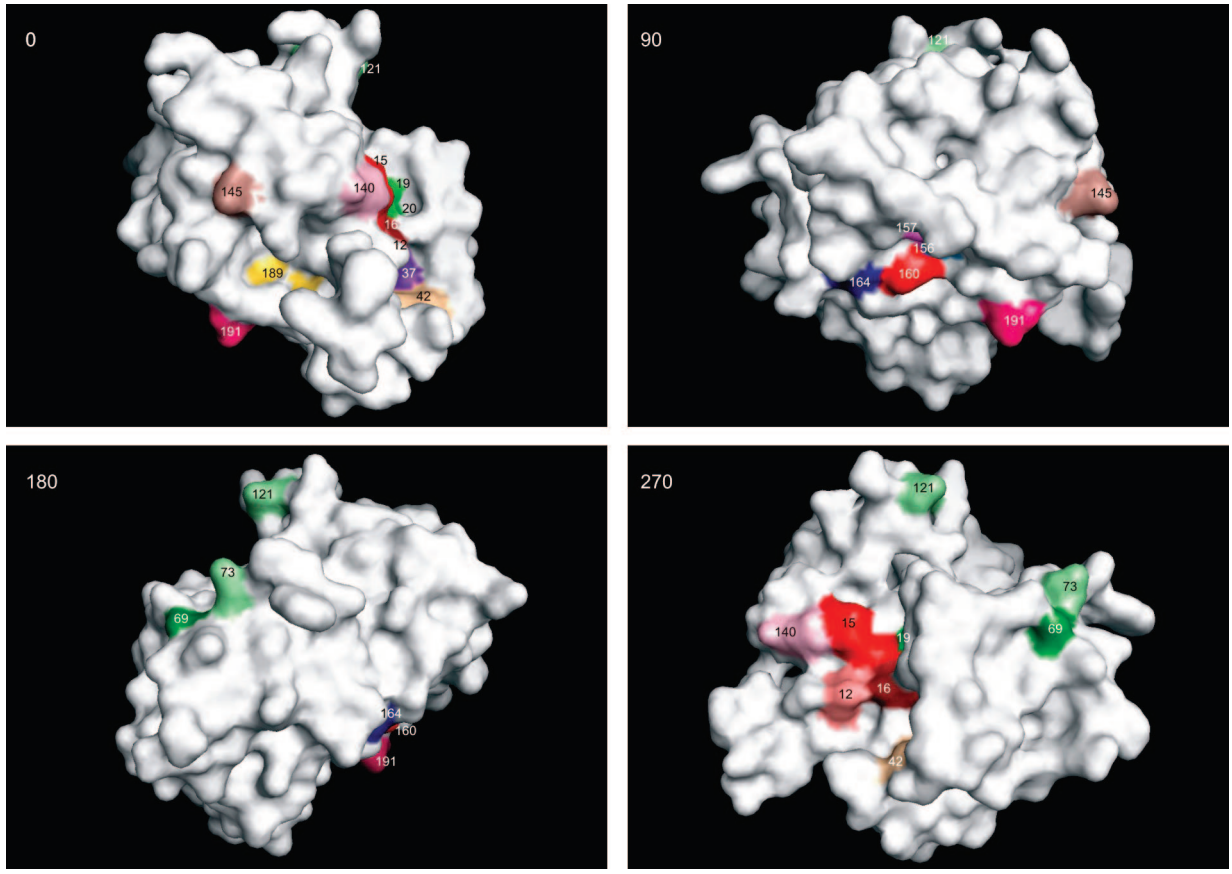

FIGURE 3: Model of DmsD highlighting the reduced pool of 22 mutated amino acids isolated from random mutagenesis, showing a decreased interaction with $DmsA_L$. Only mutations identified from DmsD mutants containing between 2 and 6 residue mutations were highlighted on this model. The four-panel image follows a 90°, right-hand rotation of the starting image (0). The color scheme seen in the images is V12D, salmon; R15C, red; V16E, ruby; A19T, green; L20Q, forest; L37P, purple blue; W42C, wheat; A69T, forest; Q73L, pale green; Q121L, pale green; E140K, light pink; T145P, dark salmon; F156S and F156L (mutated twice), sky blue; P157L, dirty violet; T160I, red; D164G, deep blue; W189L, yellow orange; and S191A, hot pink. Residues L17V and A170P are located on the surface of the molecule but not seen in the figure, whereas L75S and F128Y are not located on the surface of the molecule.

bacterial two-hybrid screen. Select mutants were chosen on the basis of their observed binding effect from the dot-blot far-Western assay, representative of 0.25- or 0.5-fold of, equal to, or greater than WT binding of $DmsA_L$. These mutants were subcloned to generate a T25 fusion similar to that described above and then screened by doubly transforming the resulting plasmids and pDmsALT18 into *E. coli* BTH101 cells. As a negative control, a plasmid containing the leucine zipper fused to T18, pUT18-zip (*35, 36*) transformed with WT DmsD, and pDmsDT25 was also included. Screening based on counting the percentage of white colonies over 3 days was performed along with assaying the activity of β-galactosidase as described by Miller (*38*).

RESULTS

Modeling of E. coli DmsD. A homology model of *E. coli* DmsD (EcDmsD) was based on the crystal structure of *S. typhimurium* DmsD (StDmsD) 1s9u, which was solved to 1.38 Å resolution. A ClustalW sequence alignment between StDmsD and EcDmsD, along with the structural information provided by the template structure StDmsD (1s9u) was used as the input data for modeling. A structural alignment of the two is shown in Figure 1, according to VMD (*21*), the alignment has an root-mean-square (rms) value of 0.4 Å, indicating that the two have a good fit. When the template is compared to the model, there is one six-residue portion from EcDmsD (residues 117–122) that is not resolved in the StDmsD structure (Figure 1, purple). This portion was generated by the ProModII component of the SWISS-MODEL (*15*) system, where similar motifs are identified in existing structures and incorporated into the model as a α-C trace. The side chains are then built around the backbone from a library of allowed rotamers, such that steric clashes are minimized. This model was judged by Verify3D (*18*) to give an average score of 0.39, and Anolea (*20*) analysis gave −904.425 E/kT units as the total energy for the model. Thus, we were able to exchange the *E. coli* sequence into the StDmsD structure with a high degree of confidence. On the basis of the model, DmsD has an α-helical bundle protein fold with 10 α helices and 2 3_{10} helices (Figure 1 and Supplementary Figure S1 in the Supporting Information).

Bioinformatic Analysis of DmsD. Sequence alignments of members of the TorD superfamily of redox enzyme maturation proteins as in ref *11* identified two conserved motifs: $_{72}$**WQRLF** and $_{123}$**EPEDH** (numbered according to DmsD sequence), where conserved residues of the motifs are bolded. Upon further bioinformatic analysis of select DmsD homologues, another set of residues appeared to be conserved (Figure 2) and, of these residues, L75, F76, E123, P124, and E126 from the two REMP motifs were commonly identified to be >80% identical in the DmsD homologues.

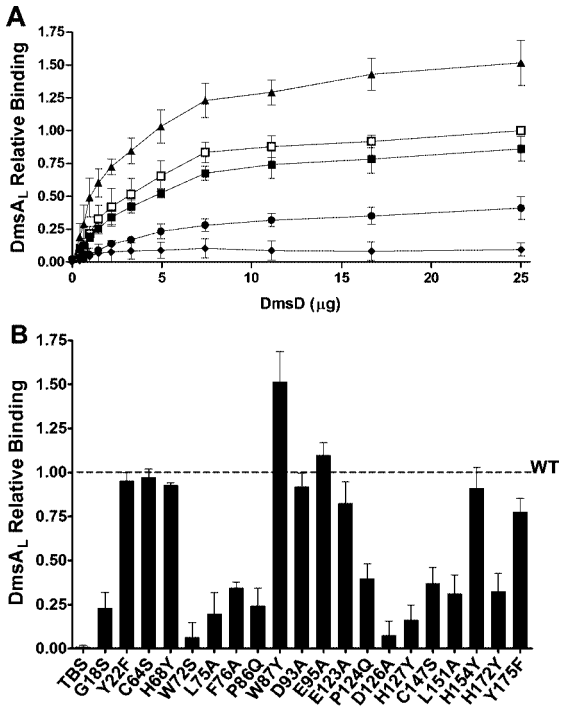

FIGURE 4: DmsA leader peptide binding to DmsD and site-specific DmsD mutants. (A) Sample binding curves showing representative behavior of various DmsD mutants compared to DmsD WT binding. (B) Binding of 20 single DmsD mutants at saturation relative to WT DmsD.

Analysis of DmsD Structure by THEMATICS for Possible Ligand-Binding Sites. THEMATICS is an electrostatics-based computational method for the identification of interaction sites in protein structures originally developed as a method to locate catalytic sites. The method also locates recognition sites and local sites for ligand binding (*27, 39*). The method locates interaction sites correctly for 93% of protein structures in an annotated database of enzymes. Even though THEMATICS predicts ionizable residues only, the sensitivity to catalytic residues is equal to or better than other 3D-structure-based methods. The main advantage of THEMATICS is that precision is higher in the selection of active residues and false-positive rates are lower compared to the other methods (*27*). Six residues of DmsD were identified by THEMATICS for potential activity. These were H68, D93, E95, D126, H127, and H172. Four of the six residues were also identified by the sequence alignment approaches described above, but H68 and H172 were unique to this approach (Table 1).

In Vivo Two-Hybrid Screening of the DmsD Random Mutant Library. A bacterial two-hybrid method was employed to screen the library of random DmsD mutants for the loss of interaction with the DmsA leader peptide (DmsA$_L$). In this system, *E. coli* BTH101 (a strain with a *mal* reporter gene and deficient of adenylate cyclase) was transformed with a plasmid expressing DmsA$_L$ fused with the T18 fragment of *Bordtella pertussis* adenylate cyclase. These cells were then transformed again with DmsD fused to the T25 fragment of adenylate cyclase and GFP. Inclusion of GFP was to rule out false-negative interactions because of nonsense or frame-shift mutations by verifying fluorescence of GFP. An interaction between DmsD and DmsA$_L$ would bring the two adenylate cyclase fragments

Table 1: Residues in DmsD Involved in DmsA Leader Binding[a]

	bioinformatics			
residue	REMP[b]	DmsD[c]	THEMATICS	random[d] library screen
G18		×		
Y22		×		●
C64				
H68			×	
W72	×			
L75	×	×		●
F76	×	×		●
P86		×		
W87		×		●
D93		×	×	
E95		×	×	●
E123	×	×		●
P124	×	×		
D126	×	×	×	
H127	×		×	
C147				
L151		×		
H154		×		
H172			×	
Y175		×		

[a] A bank of 20 residues was selected on the basis of bioinformatic alignment and THEMATICS electrostatics analyses and then targeted for site-directed mutagenesis (×). Their effect on DmsA leader binding was assayed by an *in vitro* dot-blot far-Western screen, where residues showing a negative effect upon binding by this screen are bolded. Using a random mutagenesis library of DmsD, followed by an *in vivo* bacterial two-hybrid screening, residues identified by this screen that are in common with the *in vitro* screen are indicated (●). [b] REMP = conserved residues from the two motifs identified in the REMP family of proteins. [c] DmsD > 80% conserved residues based on sequence alignment of all prokaryotic DmsD sequences in Figure 1. [d] These residues were identified in DmsD mutants containing 2−12 total mutations. For details, see Supplementary Table S2 in the Supporting Information.

(T18 and T25) together, leading to a proportional production of cAMP for the duration of the interaction, thus turning on the *mal* operon, which results in red colonies in the presence of maltose. This procedure was used to screen the DmsD random mutant library, whereby a loss of or weaker interaction with DmsA$_L$ results in white and pale pink colonies. Both a high and low mutation rate library was screened. Upon sequencing, it was noticed that the high-rate library contained up to 12 mutated residues, while the low-rate library contained a maximum of 3 mutations. This *in vivo* approach would find mutations that affect not only leader binding directly but also DmsD residues that may allow for the interaction with other molecules that have an effect on the leader binding to DmsD.

To classify the interactions of each of the DmsD random mutants with the DmsA$_L$, a second two-hybrid rescreen was performed and the percentage of white colonies was scored. Because each transformation contains only one mutant DmsD plasmid, the total number of white colonies can be counted on each plate and scored as the percentage of these out of the total number of colonies that were used to classify the mutants as noninteractors, weak interactors, or moderate interactors (Supplementary Table S1 in the Supporting Information). Of the 31 colonies expressing mutant variant proteins examined, 15 were found to abolish the interaction with DmsA$_L$, while the remaining 16 were found to allow for weak or moderate interactions to persist with DmsA$_L$ under the conditions of the second screen (Supplementary Table S1 in the Supporting Information). These were classified on the basis of all white colonies (no interaction),

more than 50% white colonies (weak interaction), or less than 50% white colonies (moderate interaction). In the 15 interaction-deficient mutants, each of the isolated colonies contained 2−12 mutated residues in DmsD, covering a total of 73 residues. Furthermore, some mutations were found to be common to more than one clone. The reduced pool of 15 noninteracting mutants was aligned with wild-type DmsD with ClustalW (*40*) and shown in Supplementary Figure S2 in the Supporting Information. On the basis of these results, putative hot spots affecting peptide binding were observed in residues F10−Y22, A28−E42, E66−W87, Q108−Q116, A142−E143, H154−T160, K169−L178, and L186−S191.

The isolated mutants were highlighted in the EcDmsD model based on the three classifications above and displayed in 90° right-handed rotations (Supplementary Figure S3 in the Supporting Information). Most of the mutated residues identified from random mutagenesis appeared on the surface of DmsD, whereas residues 60, 74, 78, 91, 131, 132, 135, 138, 141, 158, and 178 were buried within the core (Supplementary Table S1 in the Supporting Information). On the basis of the EcDmsD model, residues 192, 13, 142, 143, and 18 were found clustered together (panels N0 and N270 in Supplementary Figure S3 in the Supporting Information). Similar results were found for residues 179 and 59 (panels N90 and N180 in Supplementary Figure S3 in the Supporting Information); 159, 163, 167, and 193 (panel N90 in Supplementary Figure S3 in the Supporting Information); 69 and 174 (panel N180 in Supplementary Figure S3 in the Supporting Information); 25, 79, 90, and 126 (panel N270 in Supplementary Figure S3 in the Supporting Information); and 41, 42, and 64 (panel N270 in Supplementary Figure S3 in the Supporting Information).

To further refine this data, the mutant pool was reduced to those containing 6 or fewer point mutations per DmsD molecule in an attempt to reduce the multiple residue noise. This reduced the mutant pool from 15 to 7 with 22 amino acids mutated but did not show any linear sequence. However, when the mutations were mapped onto the EcDmsD structure, clustering of identified residues appeared in the N-terminal lobe of DmsD (Figure 3).

In Vitro Analysis of Specific DmsD Single-Mutant Interaction with the DmsA Leader Peptide. DmsD single mutants were analyzed with an *in vitro* dot-blot far-Western assay. The choice of residues for mutation was based on bioinformatics and THEMATICS results. Seven residues from the REMP conserved motifs were targeted for mutation: W72, L75, F76, E123, P124, D126, and H127. Nine more were chosen based on the DmsD homologue alignments: G18, Y22, P86, W87, D93, E95, L151, H154, and Y175, while five others overlapped with the REMP conserved motifs. Two residues were mutated on the basis of the THEMATICS analysis: H68 and H172. The two cysteine residues in DmsD (C64 and C147) were also targeted for mutation. The 20 chosen residues (Table 1) were mutated using site-directed mutagenesis, and the individual amino acid variants DmsD proteins were expressed and purified. Cell yields and total soluble protein yields for DmsD WT and all 20 residue mutants were approximately equal. Protein yields of residue mutants were similar to the WT, with notable exceptions being G18S, W72S, L75A, P86Q, D126A, H127Y, L151A, and H172Y. These eight mutant DmsD proteins were produced at levels of less than 25% of the WT DmsD protein and required a concentration step prior to performing the dot-blot far-Western analysis. Coincidentally, these low accumulating mutants also bound the DmsA leader peptide ($DmsA_L$) at weaker levels than the WT (see below). However, these amino acid substitution derivatives of DmsD showed similar circular dichroism spectra to the WT, suggesting that they are folding in a similar fashion (data not shown).

To assess the effect of each DmsD mutation upon binding $DmsA_L$, a dot-blot far-Western previously described (*3*) was performed. Purified WT DmsD and DmsD single-mutant proteins were spotted onto nitrocellulose membrane at equivalent concentrations. To quantitate the binding of each DmsD mutant relative to DmsD WT, the relative pixel intensities were normalized to that of the WT over a range of applied DmsD protein. This generated binding curves for each residue mutant variant (examples of such curves are shown in Figure 4A). All binding curves (mutant and WT) appeared to reach saturation with between 5 and 10 μg of applied protein (this also corresponds with the maximum amount of protein able to bind to the spot surface area of nitrocellulose). To compare each of the mutants for their ability to bind $DmsA_L$, the relative binding at 25 μg (more than 2 times the saturation point) was plotted on a bar graph (Figure 4B). Of the 20 single mutants assayed for binding, 11 were shown to bind to the $DmsA_L$ peptide below 0.5-fold of WT DmsD, 8 were shown to bind at approximately WT levels (0.75−1.25-fold), and 1 "hyper-binder" was identified (bound at ~1.5-fold as compared to the WT). DmsD residues with a strong reduction in binding (less than 0.5-fold relative to the WT) included G18S, W72S, L75A, F76A, P86Q, P124Q, D126A, H127Y, C147S, L151A, and H172Y. Residues that caused a 0.25-fold reduction or enhancement relative to the WT (i.e., little or no effect upon binding) included Y22F, C64S, H68Y, D93A, E95A, E123A, H154Y, and Y175F. The hyper-binder was W87Y DmsD (Figure 4B).

Using the $DmsA_L$ binding information for each of the DmsD mutants, the residues were highlighted on the structural model of EcDmsD according to their binding effect. Apparent surface-located residues with a strong negative effect upon binding (i.e., very weak binders) included W72S, F76A, P86Q, P124Q, H127Y, C147S, L151A, and H172Y (parts A and B of Figure 5). A cluster of residues showing a putative interaction "hot pocket" includes W72S, F76A, P86Q, P124Q, and H127Y (Figure 5A). Residue H172Y is nearby this pocket, while C147S and L151A are found on the opposite side (160° rotation) of the structure (Figure 5B). Furthermore, three nonsurface residues were also found to have a strong negative effect upon binding: G18S, L75A, and D126A. Within the DmsD-modeled structure, L75A and D126A are located just below the hot pocket residues F76 and H127, respectively, while G18S is close to the core of the protein. The portion of EcDmsD (residues 117–122) that was not resolved in the StDmsD structure maps to one edge of the putative signal peptide-binding pocket before helix 8. DmsD-mutated residues that had little or no effect on $DmsA_L$ binding were found primarily on the surface of the protein (Y22F, C64S, H68Y, D93A, E95A, E123A, and H154Y). Y175F was the only nonsurface mutation in DmsD that showed only a small reduction in $DmsA_L$ binding.

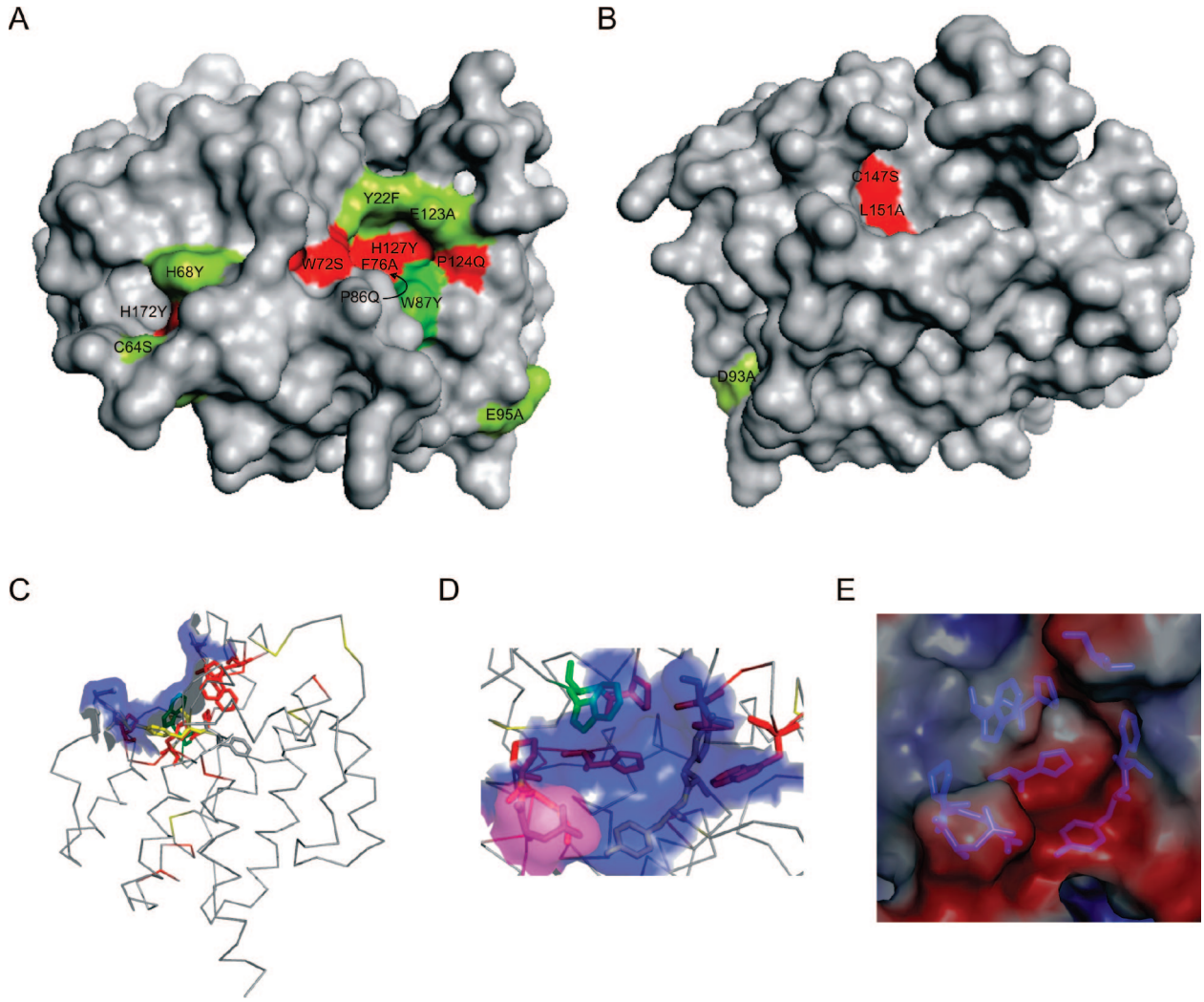

FIGURE 5: EcDmsD model of the DmsA leader binding site from *in vitro* assayed site-directed mutants. (A and B) Surface model of DmsD mutants showing less than 0.5-fold relative binding, colored red. Mutants showing WT binding (0.75−1.25-fold) were colored lime, and mutants showing binding above 1.25-fold relative to the WT were colored green. B is a 160° rotation about the *y* axis of A. (C) Cα trace of DmsD with the molecule surface for a pocket (side view) that maps to many of the mutations that effect signal peptide binding, rendered in blue. The pocket residues are shown as sticks. (D) Close up (top) view of the pocket, with each pocket residue labeled. The portion that is part of a six-residue stretch (117–122) not available from the 1s9u template but was modeled for *E. coli* DmsD is shown in magenta. (E) Electrostatic surface of the pocket, with red representing a negative charge, blue representing a positive charge, and gray representing a neutral charge.

CASTp (*22*) analysis of the molecular surface of the *E. coli* DmsD model with a 1.4 Å probe radius identified several potential binding clefts and pockets; however, a single pocket with an area of 179 Å^2 and volume of 305.2 Å^3 was partially comprised of $DmsA_L$ binding residues as identified by the above experimental data (parts C and D of Figure 5). The residues that line this CASTp predicted pocket are F21, Y22, W72, V77, F76, P86, W87, N122, E123, P124, and H127. The dimensions of the pocket are approximately 12.5 × 9.5 Å, with a maximum depth of approximately 4 Å. The base of the pocket is made up of H127 and Y22, and the walls are formed by the remainder of the residues. These are mainly hydrophobic and aromatic residues, with the exception of the polar residues N122, E123, H127. Electrostatic analysis of the molecular surface shows a patch of negative charges (Figure 5E) that may function to complement the positive charge of the arginines in the N region of the DmsA signal peptide.

In Vivo Analysis of a Specific DmsD Single-Mutant Interaction with the DmsA Leader Peptide. Select mutants from the above *in vitro* studies representing 0.25- or 0.5-fold, near WT, and greater than WT binding of $DmsA_L$ were chosen for screening by the two-hybrid *in vivo* method. An interaction between the DmsD mutants and $DmsA_L$ was quantitated by counting the number of white colonies after 1 day (results not shown) and measuring the levels of β-galactosidase activity in Miller units. The percentage of white colonies was much higher than WT DmsD for the W72S, L75A, F76A, D126A, and H127Y DmsD mutants, indicating that these mutations had reduced $DmsA_L$ binding. Mutants W87Y, E123A, and P124Q had similar amounts of white colonies to that of the WT after 1 day. By the second day, no white colonies remained for any of the DmsD mutants (all had turned red), indicating that none of the single mutations were able to completely abolish the interaction with $DmsA_L$. The negative control containing the leucine zipper of GCN4 (*35, 36*) instead of $DmsA_L$ produced white colonies throughout the screening time. When β-galactosidase activity was measured, WT DmsD interacted with $DmsA_L$ to produce ~500 Miller units of activity (Figure 6),

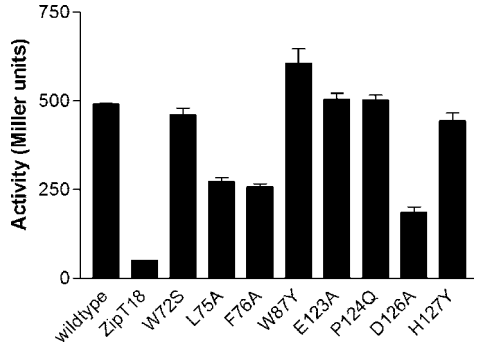

FIGURE 6: Activity assays for β-galactosidase from bacterial two-hybrid screens of DmsD mutants binding to the DmsA leader peptide. The strength of DmsD−DmsA leader binding was related to the level of β-galactosidase activity because an interaction of the two restored the ability of the cell to produce cyclic-AMP and, subsequently, β-galactosidase. As a negative control, the leucine zipper (ZipT18) instead of DmsA leader was also tested with WT DmsD.

whereas the negative control (ZipT18) yielded only ~50 Miller units. All DmsD mutants produced activity higher than that of the negative control, indicating that the binding of the $DmsA_L$ was not completely blocked by any single mutation by *in vivo* evaluation. Mutants L75A, F76A, W87Y, E123A, and D126A showed similar effects of $DmsA_L$ binding as observed by the *in vitro* dot-blot studies. Contrarily, mutants W72S, P124Q, and H127Y showed similar activity levels near that of WT DmsD (Figure 6), whereas they bound $DmsA_L$ at levels less than 0.5-fold compared to the WT from the *in vitro* studies (Figure 4B).

DISCUSSION

Two conserved motifs were previously identified within the EcDmsD sequence $_{72}$WxxLF and $_{123}$E(P/x)(x/P)DH (*11*). In this study, all three of the conserved residues in the first motif were mutated (W72S, L75A, and F76A) and found to have a very detrimental effect on $DmsA_L$ binding by the *in vitro* dot-blot far-Western method. These three residues are located at the end of helix 4. W72 and F76 are located on the same helix face (*4*) one rotation apart and surface-exposed, while the L75 side chain is not on the surface of the DmsD model. Four residues within the second motif were mutated (E123A, P124Q, D126A, and H127Y), and while E123A DmsD binding to $DmsA_L$ was not greatly affected, the other three DmsD mutants bound $DmsA_L$ at less than 0.5-fold relative to WT DmsD. These four residues are located within the longest loop of the DmsD molecule between helices 6 and 7. Both conserved motifs are in close proximity in the three-dimensional structure (~10–19 Å from the C_α atoms between the conserved residues from both motifs). Six residues from these motifs W72, L75, F76, P124, D126, and H127 along with P86 and W87 were found to be important for $DmsA_L$ binding and are within close proximity in the model, thus making up a putative RR-leader binding "hot pocket" (Figures 4B and 5A).

Despite observing detrimental effects by mutants W72S, P124Q, and H127Y on $DmsA_L$ binding using the *in vitro* dot-blot far-Western approach, the *in vivo* bacterial two-hybrid assay showed that these three mutants had little or no effect upon binding when detecting the interaction as determined by β-galactosidase activities. However, when assaying the effect of binding based on the percentage of colonies undergoing color changes within 1 day, only P124Q gave similar levels of binding as WT DmsD. Thus, the overall rate of cAMP production correlates better with the *in vitro* dot-blot far-Western assay results. The inconsistency between the two methods of assaying the two-hybrid screen and the difference between the dot-blot far-Western screens suggests that these mutants may suffer from stability effects. Although it is more likely that *in vivo*, other as yet unknown secondary-interacting proteins or biochemical events modulate the interaction.

Recently, a crystal structure was solved for the *Archaeoglobus fulgidus* protein AF0173 (PDB 2o9X) (*41*). The AF0173 protein has sequence and structural similarity to the TorD family of REMPs. The hot pocket identified in this study appears to have some similarity to a surface funnel that the authors identified in the AF0173 structure, which was hypothesized to potentially make up the leader binding pocket for this protein. Although no experimental evidence was presented to support their idea, it is lined with the same conserved residues identified in our study. D126 and H127 residues homologous within EcTorD were shown to be important for binding the leader peptide TorA (*42*) because mutation of either residue caused the K_D to be increased greater than 2-fold above that of the WT TorD protein for a short synthetic peptide composed of residues 10–36 of the TorA sequence. This suggests that the hot pocket identified here for EcDmsD may be a common theme for leader peptide binding in all proteins belonging to the REMP TorD superfamily (i.e., DmsD, TorD, NarJ, and YcdY of *E. coli*). It is also interesting to note that the domain-swapped dimer structure of SmTorD (PDB 1n1c) appears to maintain the putative leader binding pocket defined here (however, it is made up of two separate polypeptide chains).

Two tryptophan residues in DmsD were found to be important for $DmsA_L$ binding. The mutant W72S showed a $DmsA_L$ relative binding of 0.08-fold, while DmsD W87Y bound at greater than 1.5-fold of the WT level (Figure 4B). These two tryptophan residues are both found in very close proximity within the identified hot pocket of the EcDmsD modeled structure (Figure 5A) yet have completely opposite effects on $DmsA_L$ binding. This could be due to the choice of mutation because one tryptophan (W87) was mutated to a tyrosine and greatly enhanced binding, while the other (W72) was mutated to a serine and almost abolished binding. An independent study based on antibody/antigen interactions showed that tyrosine residues are very important for protein–protein interactions because of their ability to form multiple types of contacts including hydrogen bonds and polar and hydrophobic interactions (*43*). Interestingly, the residue aligning with W87 of EcDmsD is found to be a tyrosine in EcTorD, SmTorD, and AF0173 protein (which is a NarJ homologue), in addition to three of the DmsD homologues listed in Figure 2 (SdDmsD, GuDmsD, and GlDmsD). However, residues aligning with W72 of EcDmsD are more variable among DmsD homologues, with W, Y, F, or H found in this position (Figure 2).

Random mutagenesis showed that numerous residues appear to be involved in DmsA leader binding (Figure 3 and Supplementary Table S1 in the Supporting Information). Because ~90% of mutated residues identified from this method were not buried within the core of DmsD, it is

unlikely that their mutation affected the global folding or stability of the protein (capitalized residues in Supplementary Table S1 in the Supporting Information). This method also identified several residues seen in the bioinformatics approaches that are important for $DmsA_L$ binding (Table 1). These residues were isolated from mutagenesis reactions that resulted in 2−12 total mutations in DmsD. Because the combinatorial effects of multiple mutations could affect the binding of $DmsA_L$ more severely, it is difficult to draw conclusions on individual residues. However, several residues were identified by random mutagenesis but not by bioinformatic-directed mutagenesis as important for binding (Supplementary Table S1 in the Supporting Information). When the frequencies of specific mutants in each isolated random mutagenized clone were analyzed, there was some clustering of mutation spots, which is also apparent that the clusters span the entire length of DmsD (Supplementary Figure S2 in the Supporting Information). When the mutation pool was reduced to those containing between 2 and 6 mutations and the residues were mapped on the structure, we were able to see a clustering around the N-terminal lobe (Figure 3). Although it would be difficult for one to predict binding residues based on the random library results alone, this approach still provides useful clues, particularly, in situations where limited homology and structural information is available.

The site-directed approach relied on bioinformatic sequence alignments of DmsD close homologues and related REMPs to select residues to target for mutation (Table 1). A total of 16 residues were selected on the basis of bioinformatic analyses along with the two cysteine residues in DmsD. According to the results obtained from the *in vitro* dot-blot far-Western approach, all except one (E123) of the seven residues identified from the bioinformatic alignment of REMP family proteins (*11*) were shown to be important for DmsA leader binding. When DmsD close homologues were aligned, residues G18, P86, and L151 were uniquely predicted and found to be important for binding. The THEMATICS analysis of the protonation properties predicted six ionizable residues to be involved in DmsA leader binding, four of which were in common with those predicted by bioinformatic analyses (D93, E95, D126, and H127). Of the two remaining residues predicted by THEMATICS, H172 was also shown to be important for DmsA leader binding. Each of the three computational methods identified common residues involved in binding, yet each approach also identified unique residues separate from the other methods (Table 1).

In this study, we used three approaches with the goal to try to find residues and regions important in DmsA twin-arginine leader peptide binding to its cognate REMP DmsD. A mixture of random and site-directed mutagenesis was employed. Both approaches were successful. The random approach gave some ambiguity when analyzing the results because of multiple mutations within each clone; however, the site-directed approach required the target residues to be predicted in some way. We used two methods for prediction, a bioinformatic analysis requiring only primary sequence information and the THEMATICS approach, which required three-dimensional structural data. These approaches led to the identification of mutation-sensitive hot spots, providing putative residues for the leader peptide. This study demonstrates that each method was approximately 50% successful and that multiple approaches lead to a more complete understanding of the protein−peptide interface. Other DmsD residues selected for mutation, which were not found to be important for binding the leader peptide, may be involved in other protein–protein interactions or other ligand-binding events.

ACKNOWLEDGMENT

We thank Claire Wood for assistance during her undergraduate work.

SUPPORTING INFORMATION AVAILABLE

Supplementary Tables S1 and S2 and Figures S1−S3 described in the text. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Oresnik, I. J., Ladner, C. L., and Turner, R. J. (2001) Identification of a twin-arginine leader-binding protein. *Mol. Microbiol. 40*, 323–331.
2. Winstone, T. L., Workentine, M. L., Sarfo, K. J., Binding, A. J., Haslam, B. D., and Turner, R. J. (2006) Physical nature of signal peptide binding to DmsD. *Arch. Biochem. Biophys. 455*, 89–97.
3. Sarfo, K. J., Winstone, T. L., Papish, A. L., Howell, J. M., Kadir, H., Vogel, H. J., and Turner, R. J. (2004) Folding forms of *Escherichia coli* DmsD, a twin-arginine leader binding protein. *Biochem. Biophys. Res. Commun. 315*, 397–403.
4. Ko, J., Murga, L. F., Andre, P., Yang, H., Ondrechen, M. J., Williams, R. J., Agunwamba, A., and Budil, D. E. (2005) Statistical criteria for the identification of protein active sites using theoretical microscopic titration curves. *Proteins 59*, 183–195.
5. McCrindle, S. L., Kappler, U., and McEwan, A. G. (2005) Microbial dimethylsulfoxide and trimethylamine-*N*-oxide respiration. *Adv. Microb. Physiol. 50*, 147–198.
6. Sambasivarao, D., Dawson, H. A., Zhang, G., Shaw, G., Hu, J., and Weiner, J. H. (2001) Investigation of *Escherichia coli* dimethyl sulfoxide reductase assembly and processing in strains defective for the Sec-independent protein translocation system membrane targeting and translocation. *J. Biol. Chem. 276*, 20167–20174.
7. Berks, B. C. (1996) A common export pathway for proteins binding complex redox cofactors? *Mol. Microbiol. 22*, 393–404.
8. Sambasivarao, D., Turner, R. J., Simala-Grant, J. L., Shaw, G., Hu, J., and Weiner, J. H. (2000) Multiple Roles for the twin arginine leader sequence of dimethyl sulfoxide reductase of *Escherichia coli*. *J. Biol. Chem. 275*, 22526–22531.
9. Ray, N., Oates, J., Turner, R. J., and Robinson, C. (2003) DmsD is required for the biogenesis of DMSO reductase in *Escherichia coli* but not for the interaction of the DmsA signal peptide with the Tat apparatus. *FEBS Lett. 534*, 156–160.
10. Papish, A. L., Ladner, C. L., and Turner, R. J. (2003) The twin-arginine leader-binding protein, DmsD, interacts with the TatB and TatC subunits of the *Escherichia coli* twin-arginine translocase. *J. Biol. Chem. 278*, 32501–32506.
11. Turner, R. J., Papish, A. L., and Sargent, F. (2004) Sequence analysis of bacterial redox enzyme maturation proteins (REMPs). *Can. J. Microbiol. 50*, 225–238.
12. Pommier, J., Mejean, V., Giordano, G., and Iobbi-Nivol, C. (1998) TorD, a cytoplasmic chaperone that interacts with the unfolded trimethylamine *N*-oxide reductase enzyme (TorA) in *Escherichia coli*. *J. Biol. Chem. 273*, 16615–16620.
13. Chan, C. S., Howell, J. M., Workentine, M. L., and Turner, R. J. (2006) Twin-arginine translocase may have a role in the chaperone function of NarJ from *Escherichia coli*. *Biochem. Biophys. Res. Commun. 343*, 244–251.
14. Vergnes, A., Pommier, J., Toci, R., Blasco, F., Giordano, G., and Magalon, A. (2006) NarJ chaperone binds on two distinct sites of the aponitrate reductase of *Escherichia coli* to coordinate molybdenum cofactor insertion and assembly. *J. Biol. Chem. 281*, 2170–2176.

15. Guex, N., and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis 18*, 2714–2723.
16. Lund, O., Nielsen, M., Lundegaard, C., and Worning, P. (2002) CPHmodels 2.0: X3M a computer program to extract 3D models, in CASP5 Conference A102.
17. Bowie, J. U., Luthy, R., and Eisenberg, D. (1991) A method to identify protein sequences that fold into a known three-dimensional structure. *Science 253*, 164–170.
18. Luthy, R., Bowie, J. U., and Eisenberg, D. (1992) Assessment of protein models with three-dimensional profiles. *Nature 356*, 83–85.
19. Sippl, M. J. (1993) Recognition of errors in three-dimensional structures of proteins. *Proteins 17*, 355–362.
20. Melo, F., and Feytmans, E. (1998) Assessing protein structures with a non-local atomic interaction energy. *J. Mol. Biol. 277*, 1141–1152.
21. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics. *J. Mol. Graph. 14*, 27–38.
22. Binkowski, T. A., Naghibzadeh, S., and Liang, J. (2003) CASTp: Computed Atlas of Surface Topography of proteins. *Nucleic Acids Res. 31*, 3352–3355.
23. DeLano, W. L. (2002) *The PyMOL Molecular Graphics System*, DeLano Scientific, San Carlos, CA.
24. Katoh, K., Kuma, K.-i., Toh, H., and Miyata, T. (2005) MAFFT version 5: Improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res. 33*, 511–518.
25. Gouet, P., Courcelle, E., Stuart, D. I., and Metoz, F. (1999) ESPript: Analysis of multiple sequence alignments in PostScript. *Bioinformatics 15*, 305–308.
26. Ondrechen, M. J., Clifton, J. G., and Ringe, D. (2001) THEMATICS: A simple computational predictor of enzyme function from structure. *Proc. Natl. Acad. Sci. U.S.A. 98*, 12473–12478.
27. Wei, Y., Ko, J., Murga, L. F., and Ondrechen, M. J. (2007) Selective prediction of interaction sites in protein structures with THEMATICS. *BMC Bioinformatics 8*, 119.
28. Ren, P., and Ponder, J. W. (2003) Polarizable atomic multipole water model for molecular mechanics simulation. *J. Phys. Chem. B 107*, 5933–5947.
29. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys. 79*, 926–935.
30. Jorgensen, W. L., and Tirado-Rives, J. (1988) The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc. 110*, 1657–1666.
31. Madura, J. D., Briggs, J. M., Wade, R. C., Davis, M. E., Luty, B. A., Ilin, A., Antosiewicz, J., Gilson, M. K., Bagheri, B., Scott, L. R., and McCammon, J. A. (1995) Electrostatics and diffusion of molecules in solution—Simulations with the University of Houston Brownian Dynamics program. *Comput. Phys. Commun. 91*, 57–95.
32. Gilson, M. K. (1993) Multiple-site titration and molecular modeling: Two rapid methods for computing energies and forces for ionizable groups in proteins. *Proteins 15*, 266–282.
33. Bordo, D., and Argos, P. (1991) Suggestions for "safe" residue substitutions in site-directed mutagenesis. *J. Mol. Biol. 217*, 721–729.
34. Miroux, B., and Walker, J. E. (1996) Over-production of proteins in *Escherichia coli*: Mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol. 260*, 289–298.
35. Karimova, G., Dautin, N., and Ladant, D. (2005) Interaction network among *Escherichia coli* membrane proteins involved in cell division as revealed by bacterial two-hybrid analysis. *J. Bacteriol. 187*, 2233–2243.
36. Karimova, G., Ullmann, A., and Ladant, D. (2001) Protein−protein interaction between *Bacillus stearothermophilus* tyrosyl-tRNA synthetase subdomains revealed by a bacterial two-hybrid system. *J. Mol. Microbiol. Biotechnol. 3*, 73–82.
37. Tomlin, K. L., Clark, S. R. D., and Ceri, H. (2004) Green and red fluorescent protein vectors for use in Biofilm studies of the intrinsically resistant *Burkholderia cepacia* complex. *J. Microbiol. Methods 57*, 95–106.
38. Miller, J. H. (1972) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, New York.
39. Ringe, D., Wei, Y., Boino, K. R., and Ondrechen, M. J. (2004) Protein structure to function: Insights from computation. *Cell. Mol. Life Sci. 61*, 387–392.
40. Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G., and Thompson, J. D. (2003) Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res. 31*, 3497–3500.
41. Kirillova, O., Chruszcz, M., Shumilin, I. A., Skarina, T., Gorodichtchenskaia, E., Cymborowski, M., Savchenko, A., Edwards, A., and Minor, W. (2007) An extremely SAD case: Structure of a putative redox-enzyme maturation protein from *Archaeoglobus fulgidus* at 3.4 Å resolution. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 63*, 348–354.
42. Hatzixanthis, K., Clarke, T. A., Oubrie, A., Richardson, D. J., Turner, R. J., and Sargent, F. (2005) Signal peptide−chaperone interactions on the twin-arginine protein transport pathway. *Proc. Natl. Acad. Sci. U.S.A. 102*, 8460–8465.
43. Fellouse, F. A., Wiesmann, C., and Sidhu, S. S. (2004) Synthetic antibodies from a four-amino-acid code: A dominant role for tyrosine in antigen recognition. *Proc. Natl. Acad. Sci. U.S.A. 101*, 12467–12472.

BI702138A