Published in final edited form as:

Proteins. 2015 June ; 83(6): 1137–1150. doi:10.1002/prot.24811.

Solution Structure and Properties of AlgH from *Pseudomonas aeruginosa*

Jeffrey L. Urbauer^{1,2,*}, Aaron B. Cowley², Hayley P. Broussard¹, Henry T. Niedermaier¹, and Ramona J. Bieber Urbauer^{1,2}

¹The Department of Chemistry, The University of Georgia, Athens, Georgia 30602-2556

²The Department of Biochemistry and Molecular Biology, The University of Georgia, Athens, Georgia, 30602-7229

Abstract

In *Pseudomonas aeruginosa*, the *algH* gene regulates the cellular concentrations of a number of enzymes and the production of several virulence factors, and is suggested to serve a global regulatory function. The precise mechanism by which the *algH* gene product, the AlgH protein, functions is unknown. The same is true for AlgH family members from other bacteria. In order to lay the groundwork for understanding the physical underpinnings of AlgH function, we examined the structure and physical properties of AlgH in solution. Under reducing conditions, results of NMR, electrophoretic mobility, and sedimentation equilibrium experiments indicate AlgH is predominantly monomeric and monodisperse in solution. Under non-reducing conditions intra- and intermolecular disulfide bonds form, the latter promoting AlgH oligomerization. The high-resolution solution structure of AlgH reveals alpha/beta-sandwich architecture fashioned from ten beta strands and seven alpha helices. Comparison with available structures of orthologues indicates conservation of overall structural topology. The region of the protein most strongly conserved structurally also shows the highest amino acid sequence conservation and, as revealed by hydrogen-deuterium exchange studies, is also the most stable. In this region, evolutionary trace analysis identifies two clusters of amino acid residues with the highest evolutionary importance relative to all other AlgH residues. These frame a partially solvent exposed shallow hydrophobic cleft, perhaps identifying a site for intermolecular interactions. The results establish a physical foundation for understanding the structure and function of AlgH and AlgH family proteins and should be of general importance for further investigations of these and related proteins.

Keywords

NMR; protein structure; stability; hydrogen exchange; analytical ultracentrifugation; *Pseudomonas*; bacteria; regulation; virulence; evolutionary trace

*To whom correspondence should be addressed: Jeffrey L. Urbauer, The University of Georgia, The Department of Chemistry, 140 Cedar Street, Chemistry Bldg., Athens, GA 30602-2556, urbauer@uga.edu, Voice: 706 542 7922, FAX: 706 542 9454. Aaron B. Cowley's current address is Captozyme, 5745 SW 75th St. #298, Gainesville, FL, 32068-5504, USA.

Institution at which the work was performed: The University of Georgia, Athens, Georgia

Additional Supporting Information may be found in the online version of this article.

INTRODUCTION

Pseudomonas aeruginosa is a pervasive Gram-negative bacterium, with an unusual ability to colonize many diverse habitats and environments. For instance, it can grow in distilled water and in jet airplane fuel, and with or without oxygen^{1–3}. *Pseudomonas aeruginosa* is both a human and plant pathogen. It is responsible for some readily treatable external infections such as rashes and external ear infections, however, *Pseudomonas aeruginosa* is also responsible for many chronic and life-threatening infections in immunocompromised individuals and in cases where the physical barriers to infection (mucous membranes, skin) are compromised⁴. These include infections of the heart (endocarditis), burn wound infections, respiratory tract infections, central nervous system infections (meningitis), eye infections (endophthalmitis, bacterial keratitis), bone and joint infections (osteochondritis, vertebral osteomyelitis), infections associated with cancers (necrotising enterocolitis) and AIDS, in addition to infections of surgical implants and catheters⁴. *Pseudomonas aeruginosa* infections also account for high percentages of all infections acquired in hospitals (10%), including pneumonia (17%) and infections at surgical sites (11%)^{5–7}. Chronic *Pseudomonas aeruginosa* infections are perhaps most often associated with cystic fibrosis sufferers. Because of mutations in the cystic fibrosis transmembrane regulator, which functions as a chloride ion channel, thick mucus secretions accumulate in the lungs and other internal organs, which undermine the ability of cystic fibrosis patients to clear their lungs productively, resulting in recurring bacterial infections of the lungs^{8,9}.

Pseudomonas aeruginosa is a formidable pathogen. It relies on a remarkable collection of virulence factors to initiate and establish an infection, including exoproteases, phospholipases, lipopolysaccharide, phenazines, rhamnolipids, exopolysaccharides, multiple quorum sensing systems and many others. Moreover, it possesses an intrinsic arsenal of machinery to thwart antibiotic therapies, including multiple drug efflux pump systems, outer membrane impermeability mechanisms, and β -lactamases, in addition to its uncanny ability to acquire resistance^{10–16}. Interestingly, the *Pseudomonas aeruginosa* genome is quite large at approximately 6.2 million base pairs (compare to *E. coli*, with 4.6 million base pairs), coding for nearly six thousand proteins¹⁷. The ability of *Pseudomonas aeruginosa* bacteria to adapt and survive in diverse and harsh environments reflects the evolutionary advantages conferred by the large and complex genome for survival and adaptability¹⁷. According to the Centers for Disease Control and Prevention (CDC), multidrug-resistant *Pseudomonas aeruginosa* (MDRPA), like other antibiotic resistant bacteria such as methicillin-resistant *Staphylococcus aureus* (MRSA), is classified as a serious health threat. For instance, MDRPA strains account for a large percentage (13%) of all serious healthcare-associated *Pseudomonas aeruginosa* infections¹⁸.

Among the virulence factors secreted by *Pseudomonas aeruginosa* is the exopolysaccharide alginate^{19,20}. Alginate provides a natural physical barrier for the bacteria, it is antiphagocytic, it provides resistance to opsonization, it mitigates the effectiveness of aminoglycoside antibiotics, and it inhibits neutrophil and lymphocyte functions^{21–27}. Typically, *Pseudomonas aeruginosa* strains do not produce much alginate (these strains are known as nonmucoid), but harsh conditions, such as those in the cystic fibrosis lung, select for so-called mucoid strains that overproduce alginate^{19,20}. Alginate production is also

important for formation of *Pseudomonas aeruginosa* biofilms. Although mature *Pseudomonas aeruginosa* biofilms grown in the lab may lack alginate, the importance and significance of alginate for biofilm development is unquestioned²⁸.

Regulation of alginate production in *Pseudomonas aeruginosa* has been closely investigated for decades, given the role and importance of alginate as a virulence factor and in biofilm formation, and the potential to control *Pseudomonas aeruginosa* infections by mediating or stemming alginate production²⁹. Among the many *Pseudomonas aeruginosa* proteins involved in the biosynthesis of alginate or its regulation, one of these, AlgH, remains largely uncharacterized. The gene coding for the AlgH protein, *algH*, was first identified by Chakrabarty and colleagues and was shown to be required, along with *algQ*, for alginate production^{30,31}. In addition, the *algH* gene is involved in regulating the activities of the enzymes nucleoside diphosphate kinase and coenzyme A synthetase and the production of other virulence factors including siderophores, rhamnolipid biosurfactants and proteases, suggesting the AlgH protein may serve a global regulatory function^{30,31}. Given the potential importance of AlgH in regulating virulence, understanding its physical characteristics, and ultimately its specific functions, are meaningful goals.

AlgH is a member of a family of proteins of unknown function (Pfam PF02622, domain of unknown function DUF179, InterPro IPR003774, protein of unknown function UPF0301). Although sometimes annotated as transcription or translation regulators, there is no direct evidence to support this, so the precise biological functions of these proteins have yet to be determined. Previously, in order to begin to characterize its physical characteristics, we produced and purified the protein and determined its NMR chemical shift assignments^{32,33}. Here we present a high-resolution solution structure of AlgH and compare it to structures of some homologues/orthologues from other bacteria determined recently by structural genomics initiatives. We also characterize the stability of the protein and its behavior in aqueous solution. Using evolutionary trace analysis, we have also begun to identify important amino acid residues in the proteins and how they might participate structurally and functionally.

Given the serious health threat posed by antibiotic resistance in *Pseudomonas aeruginosa* and other pathogenic bacteria, targeting virulence represents an attractive strategy, complementary to conventional antibiotic therapies for controlling pathogenesis. The advantage of the approach is that the evolutionary pressure for selecting resistance is decreased^{34,35}. Proteins such as AlgH that regulate virulence are, therefore, appealing targets potentially, and understanding their structures and physical properties are important for target development.

MATERIALS AND METHODS

AlgH Production

Recombinant *Pseudomonas aeruginosa* (strain PA01) AlgH protein (UniProt entry Q9RQ16) was expressed in *Escherichia coli* and purified as described previously^{32,33}. Production of AlgH uniformly isotopically labeled with ¹⁵N or with both ¹⁵N and ¹³C was accomplished using minimal media with uniformly ¹³C-labeled glucose and uniformly ¹⁵N-

labeled NH_4Cl (Isotec inc., Miamisburg, OH) as the sole carbon and nitrogen sources, respectively^{32,33}. Dithiothreitol (DTT, 1.0 mM) was included in all buffer solutions in all steps of the purification process to inhibit intra- and intermolecular disulfide bond formation³². Common chemicals for these procedures and those described below were purchased from established commercial sources.

SDS-PAGE and Mass Spectrometry

Samples of AlgH were subjected to standard 15% SDS-PAGE. For normal visualization, proteins in the gels were stained with Coomassie Brilliant Blue. For samples for mass spectrometry analysis, the proteins were stained with Pierce Zinc Reversible Stain (Thermo Scientific, Rockford, IL). The protein bands were excised from the gel, diced into small pieces and added to a Nanosep Centrifugal Filter Unit with a 300 kDa molecular weight cutoff Omega membrane (Pall Corporation, Ann Arbor, Michigan). To extract the protein from the gel slices, 300 μl of elution buffer (0.25 M Tris-HCl, pH 6.8, and 0.1% (w/v) SDS) was added to the gel pieces in the filter, vortexed, and centrifuged for 20 min at 14,000 $\times$ g at ambient temperature. The filtrate was recovered and transferred to a fresh tube. This process was repeated one additional time. The filtrates for each sample were combined and concentrated to approximately 25 μl using an EMD Millipore Microcon Centrifugal Filter Unit with a 10 kDa molecular weight cutoff membrane centrifuged at 10,000 $\times$ g at ambient temperature (Fisher Scientific, Pittsburgh, PA). The samples were analyzed by mass spectrometry (MALDI-TOF). The MALDI target was prepared by spotting 0.8 μl of matrix (20 mg/ml of 2,5-dihydroxybenzoic acid in 1:1 acetonitrile and 0.1% trifluoroacetic acid) and allowed to dry. Then 0.8 μl of additional matrix was respotted on the plate with 0.8 μl of protein sample that was extracted from the gel slices and dried. Mass spectra were acquired in linear mode using a Bruker Autoflex (TOF) mass spectrometer (Bruker Daltonics Inc., Billerica, MA) equipped with a pulsed nitrogen laser (337 nm). The raw data were analyzed with Compass for Flex series 1.4, version 3.4 build 70 (Bruker Daltonics Inc., Billerica, MA).

NMR Spectroscopy

NMR spectra were acquired using Varian INOVA NMR spectrometers (600 and 900 MHz) with cryogenically cooled triple resonance probes. Samples of AlgH for NMR experiments consisted of 1.0–1.3 mM AlgH protein (^{15}N - or ^{13}C , ^{15}N -labeled) in a buffer consisting of 10 mM potassium phosphate, 10 mM potassium chloride, 5 mM DTT, 0.02% sodium azide and 8% D_2O , pH 7.0 (uncorrected for the isotope effect). Sample temperature was maintained at 25 $^{\circ}\text{C}$. Sample volumes were approximately 300 μL in susceptibility-matched NMR tubes (Shigemi Inc.). External Na^+DSS^+ in D_2O ($^2\text{H}_2\text{O}$) was used as the ^1H chemical shift reference (0.00 ppm). The ^{15}N and ^{13}C chemical shifts were referenced indirectly in the standard manner³⁶. Felix 2000 (Accelrys, San Diego, CA) was used for processing, visualization and analysis of all NMR data. Chemical shifts for the main chain and aliphatic side chains of AlgH were assigned previously³³. Chemical shifts of aromatic side chains were assigned using $(\text{H}\beta)\text{C}\beta(\text{C}\gamma\text{C}\delta)\text{H}\delta$ and $(\text{H}\beta)\text{C}\beta(\text{C}\gamma\text{C}\delta\text{C}\epsilon)\text{H}\epsilon$ experiments³⁷, a 3D ^1H -TOCSY-relayed ct- ^{13}C , ^1H]HMQC experiment³⁸, and a NOESY spectrum (see below). Stereospecific assignments for Val and Leu methyl groups were determined experimentally using fractional ^{13}C -labeling as described previously^{39,40}. Of these methyl groups, 71%

were assigned stereospecifically (65% for Leu, 88% for Val). All of the ^1H chemical shifts for side chain $-\text{NH}_2$ groups of Asn and Gln residues were assigned stereospecifically. This was accomplished as described previously⁴¹. Overall, the completeness of assignments is 93% (not including assignments for nitrogen in proline residues or assignments for nuclei whose signals generally are not observed in the spectra, which include those for the N-terminal methionine and side chain $-\text{NH}_2$ groups of lysine and arginine residues). All chemical shifts have been deposited (BMRB entry 6644). For structure calculations, distance restraints were based on a 3D NOESY-(^1H , ^{15}N)-HSQC spectrum, a 3D NOESY-(^1H , ^{13}C)-HSQC spectrum, and a 3D NOESY-(^1H , ^{13}C)-HSQC spectrum of the aromatic region, acquired with mixing times of 80 ms, 140 ms and 120 ms, respectively.

Residual Dipolar Couplings

Main-chain, one-bond ^1H - ^{15}N residual dipolar couplings ($^1D_{\text{NH}}$) were measured from 2D ^1H , ^{15}N -HSQC NMR spectra, recorded without ^1H decoupling during ^{15}N evolution, for samples of AlgH partially aligned in solution (Figure S1). The temperature of the samples was maintained at 25 °C. Partial alignment was effected using 20 mg/mL Pf1 phage (ASLA Ltd.). The remaining sample/buffer conditions are as described above. The axial and rhombic components of the alignment tensor were estimated initially using the program PALES⁴² and refined by a grid search⁴³ in the structure calculation process using CNS.

Hydrogen-Deuterium Exchange Experiments

Amide hydrogen-deuterium exchange (HX) experiments were conducted by diluting a concentrated solution of AlgH in H_2O (buffer conditions as described above) into an identical buffer in D_2O . The final solution contained 0.7 mM AlgH and 90% D_2O . After dilution, the sample was transferred into an NMR tube, the temperature of the sample was allowed to equilibrate in the NMR magnet (approximately 10 minutes), and a series of two-dimensional ^1H , ^{15}N -HSQC NMR spectra were recorded sequentially over a period of approximately two days (approximately 20 minutes per spectrum) in order to monitor the decreases in signal intensities accompanying exchange of amide hydrogen for solvent deuterium. Observed hydrogen exchange rate constants (k_{ex}) were determined by fitting the time-dependent decay of signal intensities to the following three-component single exponential equation:

$$I(t) = I_{\infty} + A \exp(-k_{\text{ex}} t)$$

where $I(t)$ is the signal intensity at time t , I_{∞} is the signal intensity at infinite time, A is the maximum intensity change for a given signal, given by $I(t)$ at $t = 0$ minus I_{∞} , and k_{ex} is the site-specific observed rate constant for exchange. The intrinsic (random coil, amino acid sequence-dependent) rate constants for exchange (k_{int}) were determined from reference rates defined by Englander and coworkers^{44,45}. These were corrected for the ionizations and concentrations of H_2O and D_2O , temperature, and the glass pH electrode response to D_2O ^{46,47}. Given EX2 (bimolecular) exchange conditions, the equilibrium constant (K_{HX}) and free energy (ΔG_{HX}) for the events that allow hydrogen to exchange were calculated as shown previously^{48,49}:

$$K_{\text{HX}} = k_{\text{ex}}/k_{\text{int}}$$
$$\Delta G_{\text{HX}} = -RT \ln K_{\text{HX}}$$

The overall conformational stability was calculated as the average of the largest, identical (within the statistical errors) values of G_{HX} ^{50,51}, weighted by the uncertainties in the values of G_{HX} using standard methods⁵². Corrections to the conformational stability, to account for proline isomerization, were made as described previously^{50,51}. Proline cis/trans peptide bond conformation designations were made based on patterns of NOE crosspeaks⁵³ and predictions from chemical shifts⁵⁴ using the PROMEGA web server.

Structure Calculations and Analyses

Signals/peaks in 3D NOESY spectra were identified manually and peak lists with intensities were generated from them. Along with assigned chemical shifts, these were used to iteratively define distance restraints and generate initial structures in an automated manner using the CANDID module of the torsion angle dynamics program CYANA 2.0^{55–61}. Dihedral angle restraints were derived from values of assigned chemical shifts using TALOS+^{62,63}. Hydrogen bond restraints were derived based on the results of hydrogen exchange experiments described above. The distance restraints, dihedral angle restraints, and hydrogen bond restraints were used to calculate 500 initial structures starting from an extended structure using CNS^{64,65}. These structures were refined with inclusion of the residual dipolar couplings. The 20 structures with the lowest energies were selected for analysis. Final analyses of the structures comprising this ensemble were performed with CNS, MOLMOL⁶⁶ and PROCHECK-NMR^{67,68}. Secondary structure assignments were made based on STRIDE⁶⁹, PROMOTIF⁷⁰, and patterns of NOEs. Topology maps were created based on output from the HERA program⁷¹. Molecular models were generated with MOLMOL and PyMOL⁷². Restraints for the structure calculations, atomic coordinates for the ensemble of 20 structures (ordered by increasing total energy, lowest first), and related information have been deposited in the BMRB (entry 6644) and PDB (PDB ID 2MUI).

Accessible Surface Area Analyses

The method of Lee and Richards was used to calculate the solvent accessible surface areas using the program NACCESS^{73,74}. The probe size was 1.4 Å (default). These were calculated for each of the 20 structures in the AlgH ensemble and averaged. The average relative side chain accessibilities are reported as percentages. Because these are relative to the side chain accessibility in an extended chain tripeptide model (Ala-Xaa-Ala), values greater than 100% are possible due to local geometry differences in the folded protein compared to the model.

Analytical Ultracentrifugation

Sedimentation equilibration experiments were conducted essentially as we have described previously^{75,76} using a Beckman Optima XL-A centrifuge and a four-position AN-60-Ti rotor. AlgH samples (5.6, 14, and 22 μM) were prepared in reference buffer (10 mM potassium phosphate, 10 mM potassium chloride, 1 mM DTT, pH 7.0). Typically, the use of

DTT is not advised due to the absorbance at 280 nm when oxidized. Here the DTT concentration was low and the AlgH samples were completely reduced before use, so the concentration of oxidized DTT initially was negligible and the oxidation during the experiment nominal and insignificant. Sample and reference solutions were loaded into the channels on opposing sides of six-sector Epon/charcoal centerpieces in each of three separate cells. The experiments were performed at 4 °C and three different rotor speeds (16000, 20000 and 30000 rpm). After equilibration at each speed, data (absorbance versus radial position) were recorded (280 nm, 0.001 cm increment, four-point averaging). Maximum optical density cutoff was set at 0.9. The UltraScan 7.0 software was used for data visualization and analysis^{77–79}. The partial specific volume for AlgH (0.735906 mL/g), the extinction coefficient (18,020 M⁻¹ cm⁻¹), and buffer density (0.9997 g/mL) at 20 °C, and temperature corrected values as necessary, were estimated using the UltraScan utilities^{80–82}. Using global fitting methods as implemented in UltraScan⁸³ the data were fit to likely noninteracting models (single-, two- and three-component, ideal) and equilibrium models (monomer-dimer, monomer-dimer-trimer). These models have been described in detail previously^{77,79}. Monte Carlo analyses (10,000 iterations) employing the UltraScan utility were used to determine confidence limits for fitted parameters⁸⁴. Fits were evaluated based on analysis of variance, residuals, and values of fitted parameters.

Sequence and Structure Comparisons and Analyses

Seven high resolution structures of five AlgH homologues, putative orthologues from bacteria, were identified in the PDB (Table S1). Amino acid sequences for these were obtained from the UniProt databases. Multiple and pairwise amino acid sequence alignments were performed with the CLUSTAL W algorithm⁸⁵ using the PBIL CLUSTALW server and default parameters. Structure alignments were performed using the combinatorial extension method, which returns an optimal global alignment, as implemented in the CEAlign algorithm in PyMOL^{86,87}.

Evolutionary Trace Analysis

Real-valued evolutionary trace analysis was used to evaluate the relative importance of amino acid residues in AlgH and homologues^{88–90}. Amino acid sequences for AlgH homologues were identified using BLAST and a custom NCBI sequence database⁹¹. Sequences with BLAST E-values > 0.05 were rejected, as were those more than 98% identical to the AlgH sequence. Sequences shorter than 80% of the AlgH sequence were removed, as were those where the fractional length of the BLAST high-scoring segment pair (HSP) compared to the AlgH sequence length was 0.5 or less. The remaining sequences were aligned with CLUSTAL W, and sequences were removed if the aligned length was less than 0.9 times the AlgH sequence length. The remaining sequences were then realigned. A final group of 763 acceptable, unique sequences was used for the trace, which was performed as described previously⁸⁹ using the Universal Evolutionary Trace web tool (<http://mammoth.bcm.tmc.edu/uet/>).

RESULTS AND DISCUSSION

Redox-dependent properties of AlgH

Recombinant, wild type *Pseudomonas aeruginosa* AlgH protein was expressed in *E. coli* and purified classically (no affinity tags) as we described earlier³². For all purification steps, excess (1 mM) DTT was included in all buffer solutions. Long-term storage (−80 °C) buffers for AlgH also include 1 mM DTT.

A reducing agent is necessary to inhibit formation of both intra- and intermolecular disulfide bonds in AlgH, the latter giving rise to disulfide-linked AlgH dimers and high mass oligomers (Figure 1, Figure S2). Presumably, under non-reducing conditions, these disulfide bonds are formed when the two cysteine residues in AlgH, C70 and C152, located at opposite ends of the folded protein, are transiently solvent exposed. The process is slow at neutral pH and without added oxidant. When subjected to denaturing (SDS) PAGE (Figure 1), AlgH protein purified and stored under reducing conditions at neutral pH appears homogeneous with an electrophoretic mobility consistent with a monomeric AlgH mass (~20.2 kDa/mol). The mass is confirmed by mass spectrometry. The observed chemical shift dispersion and linewidths of signals observed in the two-dimensional ¹H, ¹⁵N-HSQC NMR spectrum (Figure 1) indicate that, under reducing conditions, AlgH is monodisperse and monomeric, as we have reported previously^{32,33}. When aged at room temperature without reductant, SDS-PAGE analysis indicates that many higher mass species are formed and increased linewidths in NMR spectra, combined with loss of chemical shift dispersion, indicate both higher mass and loss of tertiary structure. An abundant species of apparent lower mass is also observed on the SDS-PAGE gel for the aged sample in the absence of reductant. Mass spectrometry indicates that this species is also monomeric AlgH. We reason that this species is AlgH with an intramolecular disulfide bond between C70 and C152, and that the added structural constraint conferred by the disulfide bond leads to an anomalous electrophoretic mobility. Addition of a reducing agent (DTT) to the sample aged under non-reducing conditions reverses the effects of aging (Figure 1), giving rise to monomeric, monodisperse AlgH, which supports the above contentions.

Results of sedimentation equilibrium experiments with AlgH, under reducing conditions, indicate that the protein is predominantly monomeric and monodisperse but with small amounts of higher mass species present. Sedimentation equilibrium data for AlgH were fit to noninteracting and equilibrium models (Table S2). The fit of the data to a single component, ideal model resulted in the highest variance and very poor agreement between the fitted and actual molecular masses. Somewhat lower variances were observed for fits to monomer-dimer and monomer-dimer-trimer equilibrium models. The fitted masses were also significantly higher than the AlgH monomer mass, although they were closer to actual values than was observed for the fit to the single component model. The dissociation constants would predict that for a 1 mM sample of AlgH, about two-thirds of these molecules would be in dimers, which is not consistent with the NMR results and with the SDS-PAGE results. Nominally lower variances were observed for the fits to the noninteracting models compared to the equilibrium models. Comparison of the raw data to the fitted values for the two-component, noninteracting model indicates the residuals are

minimized and random (Figure S3). The noninteracting models also produced the best agreement between actual and fitted masses. Analysis of the ratios of the integrals of the scans for the lower and higher mass components does not indicate increasing proportion of high to low mass components at the higher speeds and higher AlgH concentrations used in the experiments, which supports the noninteracting models compared to the equilibrium models⁷⁹. The results indicate that, under reducing conditions, AlgH is predominantly monomeric but that small quantities of other species, most likely disulfide-linked dimers, are present⁹².

Reversible redox-dependent AlgH oligomerization suggests the potential for redox mediated AlgH function and that AlgH might function as a redox sensor. It should be noted that the two cysteine residues in AlgH, C70 and C152, are not highly conserved, and evolutionary trace analysis (see below) indicates C70 to be one of the least important residues in AlgH, with C152 also of marginal importance. Thus, suppositions regarding redox control of AlgH function do not necessarily extend to all AlgH family members.

AlgH solution structure

The structure of monomeric AlgH protein in solution was determined using experimental distance and angle restraints derived from NMR spectroscopy (Figure 2). The high-resolution structural models indicate AlgH is comprised of ten beta strands ($\beta 1 - \beta 10$), seven alpha helices ($\alpha 1 - \alpha 7$) and one short region of partial helical character ($\alpha 0$), arranged in an alpha/beta-sandwich architecture (Figure 2A–D). The topology designation for the 10 strand, mixed type beta sheet is $-1 -1 -3x 1 0x -2x 1 2x 4x$ ⁹³. An ensemble of the 20 lowest energy refined structures was selected for analysis (Table 1). The analysis indicates very good agreement between the atomic coordinates for the models in the ensemble, as the RMSD from the mean structure for residues 10–175 is just 0.26 ± 0.05 Å for main chain atoms and 0.59 ± 0.06 Å for heavy atoms. This agreement is illustrated in the superposition of these 20 models (Figure 2B). The helices and sheets are well formed, well restrained and well ordered. The N-terminal residues are not well restrained, so there is little agreement among structures for this region. There is also significant variation in the positioning of the C-terminal region, and some variability among the structures in the loop between helices $\alpha 1$ and $\alpha 2$ (residues 61–67, Figure 2B–D). Plots of mean global displacement for each residue and NOE-derived distance restraint density per residue are shown in Figure S4. Other positive indicators of the quality of the structural models, including small deviations from idealized covalent geometry, a favorable distribution of main chain phi/psi dihedral angle pairs (Ramachandran analysis), and agreement with experimental restraints, also serve to validate the models.

Stability of AlgH

Site-resolved stabilities (Figure 3), and the overall conformational stability (free energy of unfolding) of AlgH were determined by hydrogen-deuterium exchange⁵⁰. As implemented, the fastest hydrogen exchange rates that can be measured are approximately 0.01 min^{-1} , which dictates a rough lower limit for measurable free energies of hydrogen exchange (ΔG_{HX}) of $\sim 15 \text{ kJ/mol}$. Sites with no reported ΔG_{HX} (Figure 3A) indicate either hydrogens that exchange too quickly to measure, proline residues, or hydrogens for which peak overlap

in the NMR spectra prohibits unambiguous rate measurement. Under the conditions these experiments were performed, the exchange process is EX2, as the intrinsic rate constants for exchange can vary by more than an order of magnitude for neighboring sites in AlgH with statistically identical measured values of G_{HX} ⁵⁰.

The most stable sites in the AlgH protein include C152 (in β 10), N46 (turn between β 3 and β 4) and G42 (β 3) (Figure 3A, 3B). The overall conformational stability of AlgH, calculated as the weighted average of the measured values of G_{HX} for these sites, is 37.6 ± 1.2 kJ/mol (8.99 kcal/mol) in 90% D₂O and uncorrected for isomerization of peptide bonds to proline (Xaa-Pro peptide bonds). For proteins with proline residues, the transient unfolding sampled by hydrogen exchange does not allow enough time for the equilibrium populations of *cis* and *trans* isomers to be established in the unfolded state, because refolding is too fast. Thus, conformational stabilities measured by hydrogen exchange in proteins with proline residues are biased and must be corrected for isomerization of the Xaa-Pro bond in the unfolded state^{50,51,94}. AlgH has 14 proline residues. Based on NOE crosspeaks and chemical shifts, these are all predicted to be *trans* isomers in the folded protein (all are *trans* in the reported 20 structure ensemble). Assuming all 14 of the Xaa-Pro bonds are *trans*, the corrected conformational stability is 34.0 ± 1.2 kJ/mol (8.13 kcal/mol). So, this might be considered a lower limit. It is also important to note that a given protein may be, unpredictably, more or less stable in D₂O compared to H₂O⁵¹.

Stable regions of the AlgH protein correspond generally to areas of well-defined secondary structure (Figure 3A, 3B, Figure S5). The most stable secondary structure elements are β 2 and β 3 (33.0 ± 1.0 kJ/mol and 33.1 ± 0.1 kJ/mol, respectively). The turn between β 3 and β 4 (I45–R47) is also quite stable (33.5 ± 1.8 kJ/mol). Similarly, the remaining strands and helices are quite stable (25–30 kJ/mol) with the exception of β 4 and β 7, which are somewhat less stable (16–18 kJ/mol), and α 0, α 2, and α 7, for which stabilities cannot be determined (but are expected to be low). As shown in Figure 3B, the antiparallel beta strands β 10, β 1, β 2, β 3 and β 9, at one end of the large beta sheet, along with the turn between β 3 and β 4 and the alpha helix (α 4) connecting β 9 and β 10, form the stable core of the protein.

The most stable regions of the AlgH protein are also those with the least side chain solvent accessibility (Figure 3A) and the least solvent exposed hydrophobic side chain area (Figure 3C). There is little solvent exposed side chain area for the residues in the stable beta sheet region (β 10, β 1, β 2, β 3 and β 9). Of all the amino acids in this region, there is only a single hydrophobic side chain with a relative solvent accessible area of 50% or more (M41, on the solvent-facing side of β 3). This lack of hydrophobic side chain solvent accessibility most likely contributes significantly to the high local stability of this region of AlgH. In contrast, the C-terminal region of AlgH is highly solvent exposed, and from residue 175 (C-terminal end of α 6) to residue 189 (C-terminus), includes seven highly solvent exposed hydrophobic side chains, which surely contribute significantly to the relative instability of the C-terminal region. Many of the less stable loop regions of AlgH also include one or more residues with highly solvent exposed hydrophobic side chains (Figure 3C).

AlgH sequence and structure comparisons with homologues

Currently there are seven high resolution structures of five bacterial AlgH homologues (putative orthologues) in the PDB (Table S1). An alignment of the amino acid sequences for these five proteins is shown in Figure S6. The statistics for this alignment of these sequences is shown in Table S3, as are the statistics for all pair-wise sequence alignments. The overall sequence similarity is a relatively low 23% (8% identical, 15% strongly similar). The pairwise sequence similarities are naturally somewhat higher. For the proteins from Gram-negative bacteria, they range from 73% (47% identical, 26% strongly similar) to 58% (32% identical, 26% strongly similar). One of the proteins is a Gram-positive protein (Q6NEA9, from *Corynebacterium diphtheria*), and its sequence shows less similarity to the others, ranging from 42% (20% identical, 22% similar) to 37% (20% identical, 17% similar).

The results of the multiple sequence alignment mapped onto the AlgH structure (Figure 4) help establish regions of conserved structure and correlations with stability. It is instructive to compare the sequence conservation to the measured local stabilities, even though the six aligned sequences represent only a relatively limited cross-section of bacterial sequence space (Gram-negative bacteria from four orders of the phylum Proteobacteria, and one Gram-positive bacterium, *Corynebacterium diphtheria*, from the phylum Actinobacteria). Most of the highest sequence conservation occurs in the region of AlgH that is also the most stable: the section of the beta sheet formed by $\beta 10$, $\beta 1$, $\beta 2$, $\beta 3$, $\beta 9$ and helix $\alpha 4$ connecting $\beta 9$ and $\beta 10$. In addition, there is a very highly conserved sequence (G78, G79, P80 in AlgH) constituting the turn connecting $\beta 5$ and $\beta 6$. This turn is also quite stable; the local stability of G79 is very high (> 30 kJ/mol), and although the local stability of G78 could not be unambiguously determined due to signal overlap in the NMR spectra, the apparent rate of hydrogen-deuterium exchange is slow, and the overlapping signals are in loops that are clearly unstable, so the apparent slow exchange (high stability) is most likely due to G78. There are, however, some highly conserved residues in some regions of the protein that are decidedly unstable. For instance, the region of partial helical character, $\alpha 0$, is not very stable, but F25, in $\alpha 0$, is highly conserved. Likewise, the C-terminal region is not stable, but G187, the antepenultimate residue, is highly conserved.

There is considerable secondary and tertiary structure similarity among this group of proteins, as expected. The overall topologies are similar, with strong conservation of secondary and tertiary structure (Figure S7). The matrix of pairwise structural alignments (Table S3) shows RMSD values ranging from ~ 2 to 5 Å. The poorest RMSD values are for alignments with the structures of the proteins from *Shewanella oneidensis* (2GZO) and *Haemophilus ducreyi* (2DO8). These are NMR structures from structural genomics initiatives. Although generally the folds agree with those of the other structures, many of the secondary structure elements are not well defined and, overall, the structures have not been highly refined. The agreement among the remaining structures is much better. As expected, the regions of the proteins that are most similar structurally are those showing the highest amino acid sequence similarity. This is exemplified by the alignment of the AlgH structure with the structure of the homologue from *Vibrio cholerae* (2HAF), which is reasonably representative of this group of pairwise structure alignments (RMSD is 3.3 Å). The structures show strong similarity for the region of the beta sheet formed by $\beta 10$, $\beta 1$, $\beta 2$, $\beta 3$,

$\beta 9$ and helix $\alpha 4$ connecting $\beta 9$ and $\beta 10$. The poorest agreement is for the $\alpha 1$ -loop- $\alpha 2$ region, which is the case for all of the pairwise alignments of these proteins. There also are significant differences among these proteins in the positioning of the region of partial helical character ($\alpha 0$). Conversely, there is good agreement in the structure and positioning of helices $\alpha 3$, $\alpha 5$ and $\alpha 6$, even though the amino acid sequence conservation is not high for these helices.

Evolutionary trace analysis

Evolutionary trace analysis is a powerful method, based on a combination of entropic and evolutionary contributions, for determining the relative evolutionary importance of residues in proteins, thereby identifying important residues^{88,89}. It has been used successfully in numerous studies to predict and identify sites important in proteins for structure and function, to direct protein mutation, and it can assist in studies of protein design, redesign and annotation^{95–98}.

Real-valued evolutionary trace was used to predict the relative importance of the residues in AlgH (Figure 5, Table S4). The top-ranked (most important) residues are G78 and G79 (100% in the histogram in Figure 5A). Although the evolutionary trace is not simply a measure of sequence conservation, it is worthwhile noting that these two glycine residues are conserved in every one of the 763 sequences used for the trace. The seventh and eighth most important residues are P80 (97.2%) and G131 (96.7%), respectively. These four residues (G78, G79, P80, and G131) cluster at one edge of the region of the beta sheet formed by $\beta 10$, $\beta 1$, $\beta 2$, $\beta 3$, $\beta 9$ and helix $\alpha 4$ connecting $\beta 9$ and $\beta 10$. The three sequential residues (G78, G79, P80) combine to effect the turn between $\beta 5$ and $\beta 6$ (Figure 5A–5C). The two glycine residues in this sequence are nearly completely solvent inaccessible, whereas the proline side chain shows significant solvent accessibility (Figure 3A, Figure 5C). This is typical for proline residues, which are found commonly in sections of turns that are accessible to solvent⁹⁹. The remaining glycine residue in this cluster, G131, is likewise nearly completely solvent inaccessible and is centered at the bend in strand $\beta 9$, apparently facilitating maintenance of this bend in the strand. It also is most likely critical that glycine occupy position 131 because of its small size, as it adjoins $\beta 6$, facilitating a bend in that strand and maintaining the sheet. A residue with a larger side chain would presumably strongly destabilize the sheet in this region. Finally, G131 is packed tightly against G78 and G79, such that a residue with a larger side chain at this position would most likely destabilize the turn between $\beta 5$ and $\beta 6$. Thus, the region of the large beta sheet in the AlgH family proteins comprised by the juncture between $\beta 5$, $\beta 9$, and $\beta 6$, appears to be a structurally fundamental element, maintained and stabilized by these four critical residues.

Of the twelve residues in AlgH identified as the most important by evolutionary trace analysis, eight of these (W149, W135, G42, Q139, E143, F25, L140, and G39) comprise a single cluster in the stable region of the beta sheet formed by $\beta 10$, $\beta 1$, $\beta 2$, $\beta 3$, $\beta 9$ and helix $\alpha 4$ connecting $\beta 9$ and $\beta 10$, with one residue of the cluster (F25) contributed by the short region of partial helical character designated as $\alpha 0$ (Figure 5C). Residues W149 ($\beta 10$) and W135 ($\beta 9$) are the third and fourth most important residues, respectively, by evolutionary trace (99.5%, 98.9%). The side chains of these residues form a predominantly T-shaped “ π -

stacking” geometry that presumably contributes to the stability of this region of the protein¹⁰⁰. Moreover, the side chain of F25 (10th most important residue, 95.6%, located in $\alpha 0$) is likewise involved in a π -stacking arrangement with W149 (Figure 5C, 5D). The agreement in the orientation and arrangement of these three side chains in the X-ray crystal structures of the AlgH homologues from *Acinetobacter baylyi*, *Vibrio cholerae* and *Corynebacterium diphtheria* and the solution structure of AlgH, overall, is extremely good, with the exception of the positioning of F25 in the AlgH structure (Figure 5E). The NOE restraint density from residue 18 through residue 25 is relatively low, and the average displacements of the residues in this region in the AlgH ensemble are correspondingly relatively large (Figure S4). The positioning of $\alpha 0$ and the residues surrounding it are, accordingly, not exceptionally well defined. This is most likely the cause for the apparent differences between the AlgH structure and the X-ray structures of the homologues in the positioning of F25. In both the AlgH structure and these X-ray structures (in Figure 5E), the π -stacking geometry for W149 and W135 is T-shaped, and from the X-ray crystal structures it is apparent that the geometry for the stacking of F25 and W149 is likewise T-shaped. In addition to this trio of π -stacked aromatic residues, this cluster includes G42, Q139, E143, L140 and G39, which are the 5th (98.3%), 6th (97.8%), 9th (96.1%), 11th (95.0%) and 12th (94.5%) most important residues, respectively. G39 and G42 ostensibly play structural roles, stabilizing $\beta 3$ and the large β sheet, with G39 also probably assisting in facilitating the turn between $\beta 2$ and $\beta 3$. Likewise, L140 appears to be important structurally; the side chain of L140 is not solvent accessible, and there are many contacts between this side chain and hydrophobic side chains on the opposing beta sheet that function to position $\alpha 4$ and stabilize the interaction of $\alpha 4$ with the beta sheet. Of the two remaining residues in the cluster, E143 also appears to perform an important structural role. The E143 side chain is nearly completely protected from solvent; it is partially inserted between the side chains of both W135 and W149 and is involved in contacts with each. It appears to assist in maintaining and reinforcing the T-stacked geometry of the W135 and W149 side chains, therefore contributing to structural stabilization. The side chain of the remaining residue in this cluster, Q139, is highly exposed (~72%) to solvent. It does not appear to be important structurally.

Side chains in these two important clusters are arranged such that a shallow cleft is formed on the AlgH surface (Figure 5C). The most solvent exposed side chains are those from F25 (~38%), Q139 (~72%) and P80 (~68%). The cleft exists in the X-ray crystal structures of the AlgH homologues also, but the solvent exposure is somewhat less overall. The cleft is very hydrophobic, with a polar component at one end of the cleft contributed by Q139. So, although the high importance of most of the residues in these clusters can be rationalized as due primarily to structural concerns, the cleft might also play a functional role, perhaps serving as a site for intermolecular interaction.

The majority of the important residues identified by the evolutionary trace analysis in AlgH are included in the region of the beta sheet formed by $\beta 10$, $\beta 1$, $\beta 2$, $\beta 3$, $\beta 9$ and helix $\alpha 4$ connecting $\beta 9$ and $\beta 10$ (Figure 5), which is also the most stable part of the protein (Figure 3). Interesting exceptions are the three residues at the C-terminus (G187, H188, and A189). Residue A189 is the 14th most important residue in AlgH. Residues H188 and G187 are the

39th and 43rd most important, respectively. Given the high solvent exposure and limited stability of the C-terminal region, it is difficult to envision these residues being important for structural reasons, and tempting to speculate that they comprise an intermolecular recognition element.

CONCLUSIONS

Our results demonstrate that the AlgH protein is a stable, well-structured protein in solution under reducing conditions. It shares an overall structural topology with putative orthologues from other bacteria, which includes one highly conserved region or domain that is also the most stable part of the protein. Although monomeric and monodisperse in solution under reducing conditions, reversible redox-sensitive AlgH oligomer formation is indicated, as is intramolecular disulfide bond formation, which suggest possible redox sensing functions for the protein and the potential to regulate AlgH function according to the redox state of the bacterium. However, sequence conservation and evolutionary trace results suggest this potential for redox-regulated function is not a general characteristic of the AlgH family. In the stable, conserved region, evolutionary trace identifies twelve amino acid residues in two clusters as the relatively most important residues and predicts a potential intermolecular recognition site. The results should impact future studies to reveal the mechanism by which AlgH and AlgH family proteins function and the underlying structural and physical origins, which should be beneficial for subsequent efforts to target virulence.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by the NIH (R21 AI070933 to JLU, F32 AI065070 to ABC), by the University of Georgia Office of the Vice President for Research and by the University of Georgia Office of the Provost. The research utilized resources of the Southeast Collaboratory for Biomolecular NMR funded in part by the NIH (P41 GM066340 to James H. Prestegard), and the Proteomic and Mass Spectrometry (PAMS) Core Facility funded in part by the NIH (S10 RR028859 to I. Jonathan Amster). This study utilized the high-performance computational capabilities of the Helix Systems at the National Institutes of Health (NIH), Bethesda, MD (<http://helix.nih.gov>), the PROMEGA web server from the Bax group at the NIH (<http://spin.niddk.nih.gov/bax/nmrserver/promega/>), the CLUSTALW web server at the Pôle BioInformatique Lyonnais (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_clustalw.html), the UniProt amino acid sequence databases (<http://www.uniprot.org>), the RCSB Protein Data Bank (<http://www.rcsb.org/pdb>), the Lichtarge Computational Biology Lab Evolutionary Trace web utilities (<http://mammoth.bcm.tmc.edu>). The authors thank Dr. Dennis Phillips, director of the PAMS facility for acquiring the mass spectrometry data and for his expertise. The authors disclose that Dr. Aaron Cowley currently is CEO of Captozyme, a Biotechnology company, but Captozyme has no financial interest in the research described in this manuscript.

Abbreviations

BMRB	biological magnetic resonance data bank
DTT	dithiothreitol
HSQC	heteronuclear single quantum coherence
MALDI	matrix assisted laser desorption ionization

MSE	selenomethionine
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
PAGE	polyacrylamide gel electrophoresis
PBIL	Pôle BioInformatique Lyonnais
PDB	protein data bank
RMSD	root mean square deviation
SDS	sodium dodecyl sulfate
TROSY	transverse relaxation optimized spectroscopy

References

1. Das N, Chandran P. Microbial degradation of petroleum hydrocarbon contaminants: an overview. *Biotechnol Res Int*. 2011; 941810:1–12.
2. Favero MS, Carson LA, Bond WW, Petersen NJ. *Pseudomonas aeruginosa*: growth in distilled water from hospitals. *Science*. 1971; 173:836–838. [PubMed: 4999114]
3. Zannoni D. The respiratory chains of pathogenic pseudomonads. *Biochim Biophys Acta*. 1989; 975:299–316. [PubMed: 2667644]
4. Pollack, M. *Pseudomonas aeruginosa*. In: Mandell, GL.; Bennet, JE.; Dolin, R., editors. *Principles and Practice of Infectious Diseases*. 4. Vol. 2. New York: Churchill Livingstone; 1995. p. 1980-2003.
5. Morrison AJ Jr, Wenzel RP. Epidemiology of infections due to *Pseudomonas aeruginosa*. *Rev of Infect Dis*. 1984; 6 (Suppl 3):S627–S642. [PubMed: 6443765]
6. National Nosocomial Infections Surveillance S. National Nosocomial Infections Surveillance (NNIS) System Report, data summary from January 1992 through June 2004, issued October 2004. *Am J Infect Control*. 2004; 32:470–485. [PubMed: 15573054]
7. Richards MJ, Edwards JR, Culver DH, Gaynes RP. Nosocomial infections in medical intensive care units in the United States. *National Nosocomial Infections Surveillance System Critical care medicine*. 1999; 27:887–892.
8. Collins FS. Cystic fibrosis: molecular biology and therapeutic implications. *Science*. 1992; 256:774–779. [PubMed: 1375392]
9. Govan JR, Deretic V. Microbial pathogenesis in cystic fibrosis: mucoid *Pseudomonas aeruginosa* and *Burkholderia cepacia*. *Microb Rev*. 1996; 60:539–574.
10. Breidenstein EB, de la Fuente-Nunez C, Hancock RE. *Pseudomonas aeruginosa*: all roads lead to resistance. *Trends Microbiol*. 2011; 19:419–426. [PubMed: 21664819]
11. Obritsch MD, Fish DN, MacLaren R, Jung R. National surveillance of antimicrobial resistance in *Pseudomonas aeruginosa* isolates obtained from intensive care unit patients from 1993 to 2002. *Antimicrob Agents Chemother*. 2004; 48:4606–4610. [PubMed: 15561832]
12. Obritsch MD, Fish DN, MacLaren R, Jung R. Nosocomial infections due to multidrug-resistant *Pseudomonas aeruginosa*: epidemiology and treatment options. *Pharmacotherapy*. 2005; 25:1353–1364. [PubMed: 16185180]
13. Poole K. Efflux-mediated multiresistance in Gram-negative bacteria. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiol and Infect Dis*. 2004; 10:12–26.

14. Poole K. Aminoglycoside resistance in *Pseudomonas aeruginosa*. Antimicrob Agents Chemother. 2005; 49:479–487. [PubMed: 15673721]
15. Schweizer HP. Efflux as a mechanism of resistance to antimicrobials in *Pseudomonas aeruginosa* and related bacteria: unanswered questions. Genet Mol Res. 2003; 2:48–62. [PubMed: 12917802]
16. Strateva T, Yordanov D. *Pseudomonas aeruginosa* - a phenomenon of bacterial resistance. J Med Microbiol. 2009; 58:1133–1148. [PubMed: 19528173]
17. Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrenner P, Hickey MJ, Brinkman FS, Hufnagle WO, Kowalik DJ, Lagrou M, Garber RL, Goltry L, Tolentino E, Westbrook-Wadman S, Yuan Y, Brody LL, Coulter SN, Folger KR, Kas A, Larbig K, Lim R, Smith K, Spencer D, Wong GK, Wu Z, Paulsen IT, Reizer J, Saier MH, Hancock RE, Lory S, Olson MV. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. Nature. 2000; 406:959–964. [PubMed: 10984043]
18. Prevention CfDCA. Antibiotic Resistance Threats in the United States. Washington, D.C: 2013. p. 1-114.
19. Evans LR, Linker A. Production and characterization of the slime polysaccharide of *Pseudomonas aeruginosa*. J Bacteriol. 1973; 116:915–924. [PubMed: 4200860]
20. Linker A, Jones RS. A new polysaccharide resembling alginic acid isolated from pseudomonads. J Biol Chem. 1966; 241:3845–3851. [PubMed: 5916397]
21. Baltimore RS, Cross AS, Dobek AS. The inhibitory effect of sodium alginate on antibiotic activity against mucoid and non-mucoid strains of *Pseudomonas aeruginosa*. J Antimicrob Chemother. 1987; 20:815–823. [PubMed: 2964429]
22. Baltimore RS, Mitchell M. Immunologic investigations of mucoid strains of *Pseudomonas aeruginosa*: comparison of susceptibility to opsonic antibody in mucoid and nonmucoid strains. J Infect Dis. 1980; 141:238–247. [PubMed: 6444976]
23. Bayer AS, Speert DP, Park S, Tu J, Witt M, Nast CC, Norman DC. Functional role of mucoid exopolysaccharide (alginate) in antibiotic-induced and polymorphonuclear leukocyte-mediated killing of *Pseudomonas aeruginosa*. Infect Immun. 1991; 59:302–308. [PubMed: 1898898]
24. Jensen ET, Kharazmi A, Lam K, Costerton JW, Hoiby N. Human polymorphonuclear leukocyte response to *Pseudomonas aeruginosa* grown in biofilms. Infect Immun. 1990; 58:2383–2385. [PubMed: 2114367]
25. Mai GT, Seow WK, McCormack JG, Thong YH. In vitro immunosuppressive and anti-phagocytic properties of the exopolysaccharide of mucoid strains of *Pseudomonas aeruginosa*. Int Arch of Allergy Appl Immunol. 1990; 92:105–112. [PubMed: 2122958]
26. Mai GT, Seow WK, Pier GB, McCormack JG, Thong YH. Suppression of lymphocyte and neutrophil functions by *Pseudomonas aeruginosa* mucoid exopolysaccharide (alginate): reversal by physicochemical, alginase, and specific monoclonal antibody treatments. Infect Immun. 1993; 61:559–564. [PubMed: 8423085]
27. Schwarzmans S, Boring JR. Antiphagocytic Effect of Slime from a Mucoid Strain of *Pseudomonas aeruginosa*. Infect Immun. 1971; 3:762–767. [PubMed: 16558051]
28. Wozniak DJ, Wyckoff TJ, Starkey M, Keyser R, Azadi P, O'Toole GA, Parsek MR. Alginate is not a significant component of the extracellular polysaccharide matrix of PA14 and PAO1 *Pseudomonas aeruginosa* biofilms. Proc Natl Acad Sci USA. 2003; 100:7907–7912. [PubMed: 12810959]
29. Ramsey DM, Wozniak DJ. Understanding the control of *Pseudomonas aeruginosa* alginate synthesis and the prospects for management of chronic infections in cystic fibrosis. Mol Microbiol. 2005; 56:309–322. [PubMed: 15813726]
30. Schlichtman D, Kubo M, Shankar S, Chakrabarty AM. Regulation of nucleoside diphosphate kinase and secreted virulence factors in *Pseudomonas aeruginosa*: roles of algR2 and algH. J Bacteriol. 1995; 177:2469–2474. [PubMed: 7730279]
31. Sundin GW, Shankar S, Chugani SA, Chopade BA, Kavanaugh-Black A, Chakrabarty AM. Nucleoside diphosphate kinase from *Pseudomonas aeruginosa*: characterization of the gene and its role in cellular growth and exopolysaccharide alginate synthesis. Mol Microbiol. 1996; 20:965–979. [PubMed: 8809750]

32. Bieber Urbauer RJ, Gilmore JM, Rosasco SE, Hattle JM, Cowley AB, Urbauer JL. Cloning, high yield overexpression, purification, and characterization of AlgH, a regulator of alginate biosynthesis in *Pseudomonas aeruginosa*. *Protein Expr Purif*. 2005; 43:57–64. [PubMed: 16084397]
33. Cowley AB, Urbauer RJ, Urbauer JL. ¹H, ¹³C and ¹⁵N NMR assignments for AlgH, a putative transcriptional regulator from *Pseudomonas aeruginosa*. *J Biomol NMR*. 2005; 33:74. [PubMed: 16222564]
34. Allen RC, Popat R, Diggle SP, Brown SP. Targeting virulence: can we make evolution-proof drugs? *Nat Rev Microbiol*. 2014; 12:300–308. [PubMed: 24625893]
35. National Research Council of the National Academies. Treating Infectious Diseases in a Microbial World. Washington, D.C: 2006. Committee on New Directions in the Study of Antimicrobial Therapeutics: New Classes of Antimicrobials and Immunomodulation.
36. Wishart DS, Bigam CG, Holm A, Hodges RS, Sykes BD. ¹H, ¹³C and ¹⁵N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J Biomol NMR*. 1995; 5:332. [PubMed: 22911508]
37. Yamazaki T, Forman-Kay JD, Kay LE. 2-Dimensional NMR Experiments for Correlating ¹³C-Beta and ¹H-Delta/Epsilon Chemical-Shifts of Aromatic Residues in ¹³C-Labeled Proteins Via Scalar Couplings. *J Am Chem Soc*. 1993; 115:11054–11055.
38. Zerbe O, Szyperski T, Ottiger M, Wuthrich K. Three-dimensional ¹H-TOCSY-relayed ct-[¹³C, ¹H]-HMQC for aromatic spin system identification in uniformly ¹³C-labeled proteins. *J Biomol NMR*. 1996; 7:99–106. [PubMed: 22911006]
39. Neri D, Otting G, Wuthrich K. ¹H and ¹³C NMR Chemical-Shifts of the Diastereotopic Methyl-Groups of Valyl and Leucyl Residues in Peptides and Proteins. *Tetrahedron*. 1990; 46:3287–3296.
40. Neri D, Szyperski T, Otting G, Senn H, Wuthrich K. Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional ¹³C labeling. *Biochemistry*. 1989; 28:7510–7516. [PubMed: 2692701]
41. McIntosh LP, Brun E, Kay LE. Stereospecific assignment of the NH₂ resonances from the primary amides of asparagine and glutamine side chains in isotopically labeled proteins. *J Biomol NMR*. 1997; 9:306–312. [PubMed: 20680662]
42. Zweckstetter M, Bax A. Prediction of sterically induced alignment in a dilute liquid crystalline phase: Aid to protein structure determination by NMR. *J Am Chem Soc*. 2000; 122:3791–3792.
43. Clore GM, Gronenborn AM, Tjandra N. Direct structure refinement against residual dipolar couplings in the presence of rhombicity of unknown magnitude. *J Magn Reson*. 1998; 131:159–162. [PubMed: 9533920]
44. Bai Y, Milne JS, Mayne L, Englander SW. Primary structure effects on peptide group hydrogen exchange. *Proteins*. 1993; 17:75–86. [PubMed: 8234246]
45. Connelly GP, Bai Y, Jeng MF, Englander SW. Isotope effects in peptide group hydrogen exchange. *Proteins*. 1993; 17:87–92. [PubMed: 8234247]
46. Covington AK, Robinson RA, Bates RG. Ionization Constant of Deuterium Oxide from 5 to 50 Degrees. *J Phys Chem-U.S.* 1966; 70:3820–3824.
47. Schowen, KBJ. Solvent Hydrogen Isotope Effects. In: Gandour, RDaSRL., editor. *Transition States of Biochemical Processes*. New York: Plenum; 1978. p. 225-283.
48. Englander SW, Kallenbach NR. Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Q Rev Biophys*. 1983; 16:521–655. [PubMed: 6204354]
49. Hvidt A, Nielsen SO. Hydrogen exchange in proteins. *Adv Protein Chem*. 1966; 21:287–386. [PubMed: 5333290]
50. Huyghues-Despointes BM, Pace CN, Englander SW, Scholtz JM. Measuring the conformational stability of a protein by hydrogen exchange. *Methods Mol Biol*. 2001; 168:69–92. [PubMed: 11357629]
51. Huyghues-Despointes BM, Scholtz JM, Pace CN. Protein conformational stabilities can be determined from hydrogen exchange rates. *Nat Struct Biol*. 1999; 6:910–912. [PubMed: 10504722]
52. Taylor, JR. *An Introduction to Error Analysis*. Sausalito, CA: University Science Books; 1997.

53. Wuthrich, K. NMR of Proteins and Nucleic Acids. New York: John Wiley & Sons; 1986.
54. Shen Y, Bax A. Prediction of Xaa-Pro peptide bond conformation from sequence and chemical shifts. *J Biomol NMR*. 2010; 46:199–204. [PubMed: 20041279]
55. Guntert P. Automated NMR protein structure calculation. *Prog Nucl Mag Res Sp*. 2003; 43:105–125.
56. Guntert P. Automated NMR structure calculation with CYANA. *Methods Mol Biol*. 2004; 278:353–378. [PubMed: 15318003]
57. Guntert P, Mumenthaler C, Wuthrich K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol*. 1997; 273:283–298. [PubMed: 9367762]
58. Herrmann T, Guntert P, Wuthrich K. Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J Biomol NMR*. 2002; 24:171–189. [PubMed: 12522306]
59. Herrmann T, Guntert P, Wuthrich K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J Mol Biol*. 2002; 319:209–227. [PubMed: 12051947]
60. Jee J, Guntert P. Influence of the completeness of chemical shift assignments on NMR structures obtained with automated NOE assignment. *J Struct Funct Genomics*. 2003; 4:179–189. [PubMed: 14649302]
61. Mumenthaler C, Guntert P, Braun W, Wuthrich K. Automated combined assignment of NOESY spectra and three-dimensional protein structure determination. *J Biomol NMR*. 1997; 10:351–362. [PubMed: 9460241]
62. Cornilescu G, Delaglio F, Bax A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR*. 1999; 13:289–302. [PubMed: 10212987]
63. Shen Y, Delaglio F, Cornilescu G, Bax A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR*. 2009; 44:213–223. [PubMed: 19548092]
64. Brunger AT. Version 1. 2 of the Crystallography and NMR system. *Nat Protoc*. 2007; 2:2728–2733. [PubMed: 18007608]
65. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D*. 1998; 54:905–921. [PubMed: 9757107]
66. Koradi R, Billeter M, Wuthrich K. MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph*. 1996; 14:51–55. [PubMed: 8744573]
67. Laskowski RA, Macarthur MW, Moss DS, Thornton JM. Procheck - a Program to Check the Stereochemical Quality of Protein Structures. *J Appl Crystallogr*. 1993; 26:283–291.
68. Laskowski RA, Rullmann JA, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR*. 1996; 8:477–486. [PubMed: 9008363]
69. Frishman D, Argos P. Knowledge-based protein secondary structure assignment. *Proteins*. 1995; 23:566–579. [PubMed: 8749853]
70. Hutchinson EG, Thornton JM. PROMOTIF - a program to identify and analyze structural motifs in proteins. *Protein Sci*. 1996; 5:212–220. [PubMed: 8745398]
71. Hutchinson EG, Thornton JM. HERA - a program to draw schematic diagrams of protein secondary structures. *Proteins*. 1990; 8:203–212. [PubMed: 2281084]
72. DeLano, WL. The PyMOL Molecular Graphics System. Palo Alto, CA: 2002.
73. Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. *J Mol Biol*. 1971; 55:379–400. [PubMed: 5551392]
74. Hubbard, SJ.; Thornton, JM. “NACCESS”, computer program. Department of Biochemistry and Molecular Biology, University College, London; London, U.K: 1993.
75. Gilmore JM, Bieber Urbauer RJ, Minakhin L, Akoyev V, Zolkiewski M, Severinov K, Urbauer JL. Determinants of affinity and activity of the anti-sigma factor AsiA. *Biochemistry*. 2010; 49:6143–6154. [PubMed: 20545305]

76. Urbauer JL, Adelman K, Urbauer RJ, Simeonov MF, Gilmore JM, Zolkiewski M, Brody EN. Conserved regions 4.1 and 4.2 of sigma(70) constitute the recognition sites for the anti-sigma factor AsiA, and AsiA is a dimer free in solution. *J Biol Chem.* 2001; 276:41128–41132. [PubMed: 11518715]
77. Demeler, B. UltraScan - A Comprehensive Data Analysis Software Package for Analytical Ultracentrifugation Experiments. In: Scott, DJ.; Harding, SE.; Rowe, AJ., editors. *Modern Analytical Ultracentrifugation: Techniques and Methods.* United Kingdom: Royal Society of Chemistry; 2005. p. 210-229.
78. Demeler, B. Hydrodynamic Methods. In: Rashidi, H.; Buehler, L., editors. *Bioinformatics Basics: Applications in Biological Science and Medicine.* 2. CCRC Press LLC; 2005. p. 226-255.
79. Demeler, B. UltraScan - A Comprehensive Data Analysis Software Package for Analytical Ultracentrifugation Experiments. The University of Texas Health Science Center at San Antonio, Department of Biochemistry; <http://www.ultrascan.uthsca.edu>
80. Cohn, EJ.; Edsall, JT. *Proteins, Amino Acids and Peptides as Ions and Dipolar Ions.* New York, NY: Reinhold; 1943.
81. Durchschlag, H. Specific volumes of biological macromolecules and some other molecules of biological interest. In: Hinz, H-J., editor. *Thermodynamic Data for Biochemistry and Biotechnology.* New York: Springer-Verlag; 1986. p. 45-128.
82. Laue, TM.; Shah, BD.; Ridgeway, TM.; Pelletier, SL. Computer-aided interpretation of analytical sedimentation data for proteins. In: Harding, SE.; Rowe, AJ.; Horton, JC., editors. *Analytical Ultracentrifugation in Biochemistry and Polymer Science.* Cambridge, Royal Society of Chemistry; Horton: 1992. p. 90-125.
83. Johnson ML, Correia JJ, Yphantis DA, Halvorson HR. Analysis of data from the analytical ultracentrifuge by nonlinear least-squares techniques. *Biophys J.* 1981; 36:575–588. [PubMed: 7326325]
84. Demeler B, Brookes E. Monte Carlo analysis of sedimentation experiments. *Colloid Polym Sci.* 2008; 286:129–137.
85. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994; 22:4673–4680. [PubMed: 7984417]
86. Jia Y, Dewey TG, Shindyalov IN, Bourne PE. A new scoring function and associated statistical significance for structure alignment by CE. *J Comput Biol.* 2004; 11:787–799. [PubMed: 15700402]
87. Shindyalov IN, Bourne PE. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.* 1998; 11:739–747. [PubMed: 9796821]
88. Lichtarge O, Bourne HR, Cohen FE. An evolutionary trace method defines binding surfaces common to protein families. *Journal Mol Biol.* 1996; 257:342–358. [PubMed: 8609628]
89. Mihalek I, Res I, Lichtarge O. A family of evolution-entropy hybrid methods for ranking protein residues by importance. *Journal Mol Biol.* 2004; 336:1265–1282. [PubMed: 15037084]
90. Mihalek I, Res I, Lichtarge O. Evolutionary trace report_maker: a new type of service for comparative analysis of proteins. *Bioinformatics.* 2006; 22:1656–1657. [PubMed: 16644792]
91. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 1997; 25:3389–3402. [PubMed: 9254694]
92. Philo JS. An improved function for fitting sedimentation velocity data for low-molecular-weight solutes. *Biophys J.* 1997; 72:435–444. [PubMed: 8994630]
93. Richardson JS. The anatomy and taxonomy of protein structure. *Adv Protein Chem.* 1981; 34:167–339. [PubMed: 7020376]
94. Bai Y, Milne JS, Mayne L, Englander SW. Protein stability parameters measured by hydrogen exchange. *Proteins.* 1994; 20:4–14. [PubMed: 7824522]
95. Erdin S, Lisewski AM, Lichtarge O. Protein function prediction: towards integration of similarity metrics. *Curr Opin Struc Biol.* 2011; 21:180–188.
96. Lichtarge O, Sowa ME. Evolutionary predictions of binding surfaces and interactions. *Curr Opin Struc Biol.* 2002; 12:21–27.

97. Lichtarge O, Wilkins A. Evolution: a guide to perturb protein function and networks. *Curr Opin Struc Biol.* 2010; 20:351–359.
98. Wilkins AD, Bachman BJ, Erdin S, Lichtarge O. The use of evolutionary patterns in protein annotation. *Curr Opin Struc Biol.* 2012; 22:316–325.
99. Lins L, Thomas A, Brasseur R. Analysis of accessible surface of residues in proteins. *Protein Sci.* 2003; 12:1406–1417. [PubMed: 12824487]
100. Martinez CR, Iverson BL. Rethinking the term “pi-stacking”. *Chem Sci.* 2012; 3:2191–2201.

Figure 1.

Redox-dependent properties of AlgH. **(A)** Expansion of a ^1H - ^{15}N TROSY NMR spectrum (full spectrum shown in Figure S2, panel A) of uniformly ^{15}N -, ^{13}C -labeled AlgH. This protein was prepared under reducing conditions and then dialyzed against the normal NMR buffer, but lacking DTT, in order to remove the DTT. **(B)** Expansion of a ^1H - ^{15}N TROSY NMR spectrum (full spectrum shown in Figure S2, panel B) of the AlgH sample from panel A aged at ambient temperature in the absence of added reductant. **(C)** Expansion of a ^1H - ^{15}N TROSY NMR spectrum (full spectrum shown in Figure S2, panel C) of the aged sample from panel B after addition of DTT to a final concentration of 50 mM. **(D)** SDS-PAGE gel of AlgH samples. Lanes: M = molecular mass standards, 1 = AlgH purified under reducing conditions (then dialyzed to remove the DTT), with 100 mM DTT in the gel loading buffer, 2 = purified AlgH aged without added reductant (DTT) and without DTT in the gel loading buffer, 3 = sample from lane 2 with 100 mM DTT in the sample loading buffer.

Figure 2.

AlgH structure and topology. **(A)** Ribbon diagram for the lowest energy AlgH structure with alpha helix (α1 – α7) and beta strand (β1 – β10) designations indicated. An additional short sequence with partial helical character (α0) is also indicated. **(B)** Stereoview (cross-eyed) of the 20 lowest energy AlgH structures superimposed on the average structure. **(C)** AlgH topology map. **(D)** Secondary structure elements of AlgH aligned with the AlgH amino acid sequence.

Figure 3.

Site resolved stability of AlgH and side chain solvent accessibility. **(A)** The measured, residue-specific free energies of hydrogen exchange (ΔG_{HX} , kJ/mol) for AlgH are shown as black bars. The uncertainties, calculated by propagating the standard errors in the fitted values of the hydrogen exchange rate constants, are shown as gray caps on the bars. The secondary structure elements are indicated (helices – green cylinders, beta strands – red arrows). The dashed line is ΔG_{HX} for the most stable site (C152, 38 ± 1 kJ/mol). The open bars indicate either proline residues (P6, P18, P23, P48, P62, P66, P67, P80, P91, P123, P153, P156, P163, P164), or residues for which the data suggest the hydrogen-deuterium exchange rates are measurable, but for which peak overlap precludes an absolute rate determination (D22, V29, Q38, L43, N52, V56, Q71, G78, H90, L127, G131, W135, S169). For the 20 structures in the AlgH structure ensemble, the average relative side chain solvent accessible surface areas (RASA), as percentages, are also shown as black bars (uncertainties shown as gray caps on the bars). **(B)** The site-resolved stabilities are shown mapped onto the ribbon diagram of the lowest energy AlgH conformer of the ensemble, colored according to the magnitude of ΔG_{HX} , as indicated. The three most stable residues are indicated. **(C)** The average relative side chain solvent accessible surface areas (RASA), as percentages, are mapped onto the ribbon diagram of the lowest energy AlgH conformer of the ensemble, colored according to their magnitudes, as indicated. The side chains shown as sticks are hydrophobic side chains (except glycine and proline) with greater than 30% solvent accessible surface area.

Figure 4. Sequence and structure conservation in the AlgH family. **(A)** For AlgH and homologues for which high resolution structures are available, the amino acid sequence identities and similarities from the multiple sequence alignment using ClustalW are mapped onto the AlgH ribbon diagram of the lowest energy conformer of the ensemble. Color-coding is red=identical, green=strongly similar, and blue=weakly similar. **(B)** The structural alignment between the lowest energy AlgH conformer of the ensemble (gray) and the X-ray crystal structure of the AlgH homologue from *Vibrio cholerae* (green, PDB identifier 2HAF) exemplifies the regions of structural similarity and dissimilarity among these homologues.

Figure 5.

Real-valued evolutionary trace analysis of the AlgH family. **(A)** The histogram shows the relative importance (RI) of AlgH residues as percentages determined by real-valued evolutionary trace analysis, with 100% being the most important. The secondary structure elements of AlgH are indicated (helices – green cylinders, beta strands – red arrows). **(B)** The evolutionary trace relative importance of AlgH residues color-coded as indicated and mapped onto the ribbon diagram AlgH. **(C)** Two clusters of important residues, color-coded as in panel B, represented as spheres. One cluster is comprised of G78, G79, P80 and G131, and the other of F25, G39, G42, W135, Q139, L140, E143 and W149. These twelve residues are the most important in AlgH according to the analysis. **(D)** The π -stacking of the side chains of F25, W149, and W135, three of the most important residues in AlgH (colored as in panels B and C). **(E)** The π -stacking arrangement of F25, W149, and W135 in AlgH (gray residues) are compared to the orientations found in X-ray crystal structures of the AlgH homologues from *Vibrio cholerae* (green, PDB identifier 2HAF), *Acinetobacter baylyi* (red, 2EW0), and *Corynebacterium diphtheria* (yellow, 2GS5). For panels B–E, the lowest energy AlgH structure of the ensemble is shown.

Table I

Statistics summary for the AlgH structure ensemble

NOE-based distance restraints	
Intra-residue ($i = j$)	587
Sequential ($ i-j = 1$)	823
Medium Range ($2 \leq i-j \leq 4$)	463
Long Range ($ i-j \geq 5$)	737
Total	2610
Hydrogen bond distance restraints	152
Dihedral angle restraints (φ and ψ)	154
RDC restraints	84
RMSD from the mean structure*	
Main chain (residues 10–175)	0.26 ± 0.05 (0.36 ± 0.07)
Heavy atoms (residues 10–175)	0.59 ± 0.06 (0.87 ± 0.08)
RMSD from experimental restraints**	
NOE-based distance restraints (Å)	0.0276 ± 0.0005
Dihedral angle restraints (°)	0.47 ± 0.03
RMSD from idealized covalent geometry	
Bonds	0.0070 ± 0.00004
Angles	0.707 ± 0.005
Improper angles	0.614 ± 0.007
Ramachandran analysis, residues 10–175 (%)	
Residues in most favored regions	85.8
Residues in additional allowed regions	13.4
Residues in generously allowed regions	0.9
Residues in disallowed regions	0.0

* values in parentheses are mean global RMSD values (MOLMOL)

** for the 20 structure ensemble, there were no NOE distance restraint violations > 0.25 Å, no hydrogen bond restraint violations > 0.2 Å, no dihedral angle restraint violations $> 5^\circ$, and no RDC restraint violations > 1 Hz