AMERICAN SOCIETY FOR MICROBIOLOGY | mBio®

RESEARCH ARTICLE

The ChiS-Family DNA-Binding Domain Contains a Cryptic Helix-Turn-Helix Variant

Catherine A. Klancher,[a] George Minasov,[b,c] Ram Podicheti,[d] Douglas B. Rusch,[d] Triana N. Dalia,[a] Karla J. F. Satchell,[b,c] Matthew B. Neiditch,[e] Ankur B. Dalia[a]

[a]Department of Biology, Indiana University, Bloomington, Indiana, USA
[b]Center for Structural Genomics of Infectious Diseases, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA
[c]Department of Biochemistry and Molecular Genetics, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA
[d]Center for Genomics and Bioinformatics, Indiana University, Bloomington, Indiana, USA
[e]Department of Microbiology, Biochemistry, and Molecular Genetics, New Jersey Medical School, Rutgers Biomedical Health Sciences, Newark, New Jersey, USA

ABSTRACT Sequence-specific DNA-binding domains (DBDs) are conserved in all domains of life. These proteins carry out a variety of cellular functions, and there are a number of distinct structural domains already described that allow for sequence-specific DNA binding, including the ubiquitous helix-turn-helix (HTH) domain. In the facultative pathogen *Vibrio cholerae*, the chitin sensor ChiS is a transcriptional regulator that is critical for the survival of this organism in its marine reservoir. We recently showed that ChiS contains a cryptic DBD in its C terminus. This domain is not homologous to any known DBD, but it is a conserved domain present in other bacterial proteins. Here, we present the crystal structure of the ChiS DBD at a resolution of 1.28 Å. We find that the ChiS DBD contains an HTH domain that is structurally similar to those found in other DNA-binding proteins, like the LacI repressor. However, one striking difference observed in the ChiS DBD is that the canonical tight turn of the HTH is replaced with an insertion containing a β-sheet, a variant which we term the helix-sheet-helix. Through systematic mutagenesis of all positively charged residues within the ChiS DBD, we show that residues within and proximal to the ChiS helix-sheet-helix are critical for DNA binding. Finally, through phylogenetic analyses we show that the ChiS DBD is found in diverse proteobacterial proteins that exhibit distinct domain architectures. Together, these results suggest that the structure described here represents the prototypical member of the ChiS-family of DBDs.

IMPORTANCE Regulating gene expression is essential in all domains of life. This process is commonly facilitated by the activity of DNA-binding transcription factors. There are diverse structural domains that allow proteins to bind to specific DNA sequences. The structural basis underlying how some proteins bind to DNA, however, remains unclear. Previously, we showed that in the major human pathogen *Vibrio cholerae*, the transcription factor ChiS directly regulates gene expression through a cryptic DNA-binding domain. This domain lacked homology to any known DNA-binding protein. In the current study, we determined the structure of the ChiS DNA-binding domain (DBD) and found that the ChiS-family DBD is a cryptic variant of the ubiquitous helix-turn-helix (HTH) domain. We further demonstrate that this domain is conserved in diverse proteins that may represent a novel group of transcriptional regulators.

KEYWORDS DNA-binding proteins, molecular genetics, structural biology

The intestinal pathogen *Vibrio cholerae* natively resides in the aquatic environment and can cause disease if ingested in contaminated food or drinking water. In the aquatic environment, *V. cholerae* commonly associates with the chitinous surfaces of crustacean zooplankton (1). Chitin is an abundant source of carbon and nitrogen for

Citation Klancher CA, Minasov G, Podicheti R, Rusch DB, Dalia TN, Satchell KJF, Neiditch MB, Dalia AB. 2021. The ChiS-family DNA-binding domain contains a cryptic helix-turn-helix variant. mBio 12:e03287-20. https://doi.org/10.1128/mBio.03287-20.

Invited Editor Mark Goulian, University of Pennsylvania

Editor Edward G. Ruby, University of Hawaii at Manoa

Address correspondence to Ankur B. Dalia, ankdalia@indiana.edu.

Received 18 November 2020
Accepted 10 February 2021
Published 16 March 2021

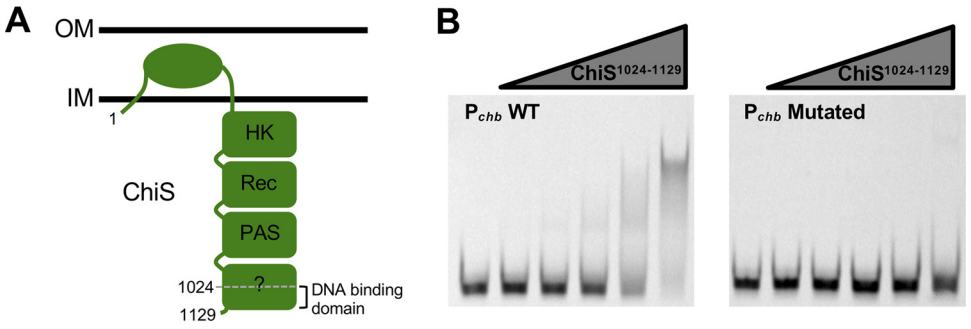

FIG 1 The C terminus of ChiS (ChiS$^{1024–1129}$) is sufficient to bind P_{chb}. (A) Diagram of the domain architecture for the hybrid histidine kinase ChiS. ChiS contains a histidine kinase (HK) domain, a receiver domain (Rec), a Per-Arnt-Sim (PAS) domain, and a domain that does not have homology to known domains. Residues 1024 to 1129 were previously shown to be sufficient to bind P_{chb} *in vivo* (5). (B) A fragment of the ChiS C terminus (ChiS$^{1024–1129}$) was purified and assessed for DNA binding activity by EMSA. Purified protein was incubated with the indicated Cy5-labeled 60-bp probes containing sequence from P_{chb} encompassing the two ChiS binding sites (CBSs). The probe sequence was WT (P_{chb} WT) or the CBSs were both mutated (P_{chb} Mutated). See Fig. S1A for a promoter map and the probe sequences used. The concentrations of ChiS used (from left to right) were 0 nM, 25 nM, 50 nM, 100 nM, 200 nM, and 400 nM. Data are representative of two independent experiments.

marine bacteria, including *V. cholerae* (2, 3). In addition, chitin serves as a cue to induce horizontal gene transfer by natural transformation in this species (4). Thus, *Vibrio*-chitin interactions are critical for this facultative pathogen to thrive and evolve in its environmental reservoir.

Chitin is sensed in *V. cholerae* by the hybrid histidine kinase ChiS (5–7). In response to chitin, ChiS activates the expression of the chitin utilization program. This regulon includes the *chb* operon, which is required for the uptake and degradation of the chitin disaccharide chitobiose. In a recent study, we showed that unlike most histidine kinases, ChiS is capable of directly binding to DNA to regulate the expression of the *chb* operon (5). This finding was particularly surprising because ChiS is not predicted to contain a DNA-binding domain via primary sequence homology (BLAST [8]) or structural predictions (Phyre2 [9]). In the current study, we sought to understand the structural basis for ChiS DNA binding. To that end, we determined the structure of the ChiS DNA-binding domain (DBD) and found that it contains a distinct variant of the canonical helix-turn-helix domain, which we term a helix-sheet-helix.

RESULTS AND DISCUSSION

The C terminus of ChiS (ChiS$^{1024–1129}$) is sufficient to bind P_{chb}. Previous work from our group demonstrated that ChiS is a noncanonical hybrid histidine kinase that contains a DBD at its C terminus (Fig. 1A) (5). In that study, we found that the C-terminal 106 amino acids of ChiS (ChiS$^{1024–1129}$) were necessary and sufficient to bind to the *chb* promoter *in vivo*. We further showed that ChiS binds directly to two binding sites within the *chb* operon promoter (P_{chb}) to activate the expression of this locus. To confirm that ChiS$^{1024–1129}$ was sufficient to bind DNA, we purified this domain and tested its DNA-binding activity *in vitro* by electrophoretic mobility shift assays (EMSAs). We found that ChiS$^{1024–1129}$ bound to a wild-type P_{chb} promoter probe, but not to a probe in which the two ChiS binding sites were mutated, suggesting that this domain is sufficient to bind to DNA in a sequence-specific manner (Fig. 1B and Fig. S1). Thus, based on our *in vivo* and *in vitro* analysis, we refer to ChiS$^{1024–1129}$ as the ChiS DBD.

Identification of positively charged residues in the ChiS DBD that are critical for DNA binding and transcriptional activation of P_{chb}. As mentioned above, ChiS is not predicted to contain a DNA-binding domain. This is based on *in silico* searches using the primary sequence of the empirically determined ChiS DBD. With BLAST, no conserved domains were detected in the ChiS DBD (8). Further, Phyre2-predicted structural models were of very low confidence, and none of the hits identified contained a known DNA-binding domain (9).

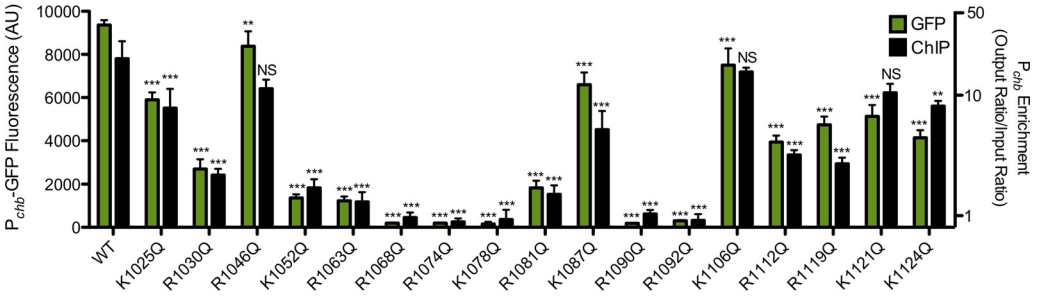

FIG 2 Identification of positively charged residues in the ChiS DBD that are critical for DNA binding and transcriptional activation of P_{chb}. All lysines and arginines in the ChiS DNA-binding domain were individually mutated to a glutamine and ChiS was assessed for (i) transcriptional activation of a P_{chb}-GFP reporter (green bars; left *y* axis) and (ii) ChiS binding to P_{chb} *in vivo* by chromatin immunoprecipitation (ChIP) (black bars; right *y* axis). ChiS can be activated with its native inducer, chitin, or by deletion of its periplasmic regulator, chitin binding protein (CBP); here, ChiS was activated artificially by deleting CBP. Data are the results of at least three independent biological replicates and are means and standard deviations (SD). Statistical markers above the bars indicate comparisons to the WT made by one-way analysis of variance (ANOVA) with Tukey's posttest. ***, $P < 0.001$; **, $P < 0.01$; NS, not significant.

To characterize interactions between the ChiS DBD and DNA, we first tried to identify residues important for DNA binding. The positively charged amino acids arginine (R) and lysine (K) commonly interact with the negatively charged DNA backbone and can also make critical contacts with nucleotide bases (10). Thus, we mutated every R and K residue in the ChiS DBD to a glutamine (Q), to ablate their charge but maintain, to a reasonable extent, the steric properties of the side group. To determine how these mutations affected ChiS activity, we introduced them into full-length FLAG-tagged ChiS (5) and assessed the ability of each mutant to bind to DNA *in vivo* (by chromatin immunoprecipitation, or ChIP) and to activate P_{chb} expression (using a P_{chb}-green fluorescent protein [GFP] reporter). Full-length ChiS was used for these experiments because this construct is functional for both DNA binding and transcriptional activation of P_{chb}, whereas the ChiS DBD is functional only for DNA binding (5). We found that all mutations to ChiS reduced P_{chb}-GFP activation to various degrees (Fig. 2). Most mutants were able to facilitate partial activation of P_{chb} and, correspondingly, partially enriched for P_{chb} by ChIP, indicating that they bound to the promoter *in vivo*. Some mutants (R1068Q, R1074Q, K1078Q, R1090Q, and R1092Q) did not bind to P_{chb} DNA *in vivo* and resulted in complete loss of P_{chb} expression. All mutants still produced ChiS protein, as assessed by Western blotting analysis (Fig. S2); however, we cannot exclude the possibility that these single amino acid substitutions result in protein misfolding. Collectively, these data identify a subset of positively charged residues in the ChiS DBD that are likely critical for DNA binding and subsequent transcriptional activation of the *chb* operon.

Structure of the ChiS DNA-binding domain reveals a variant of the helix-turn-helix. We next sought to determine the structure of the ChiS DBD to further explore how ChiS interacts with DNA. Since no structures for close sequence homologs were available in the Protein Data Bank (PDB) to serve as search models for molecular replacement, we used the single-wavelength anomalous dispersion (SAD) technique to determine initial phases. Selenomethionine (Se-Met) was used as the replacement for methionine. Anomalous data were collected from a single crystal (see Table S1 in the supplemental material). The crystal diffracted to 1.28 Å resolution and belonged to the orthogonal $C222_1$ space group with unit cell parameters of $a = 51.91$ Å, $b = 78.61$ Å, $c = 72.37$ Å, and $\alpha = \beta = \gamma = 90.00°$. There was one polypeptide chain in the asymmetric unit. The structure includes 105 of 106 residues of the protein (1024 to 1128), two uncleavable residues of the purification tag, four sulfate ions (SO_4^{2-}), one 2-(2-hydroxyethyloxy)ethanol molecule (PEG), two formic acid molecules (FMT), and 200 water molecules (HOH). Only the C-terminal E1129 was disordered in the structure and was not included in the final model.

The structure of the ChiS DBD revealed that it contains a fold that is reminiscent of

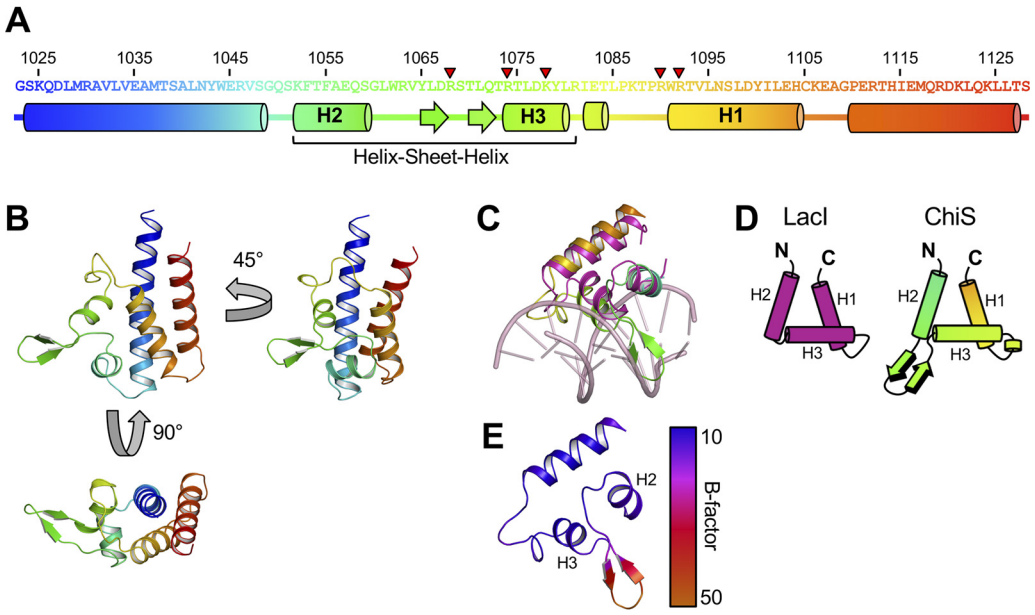

FIG 3 Structure of the ChiS DNA-binding domain reveals a variant of the helix-turn-helix. (A) Domain architecture of the ChiS DNA-binding domain. The primary sequence of the ChiS DBD is shown. Helices are depicted as cylinders, while sheets are depicted as arrows. The five R and K residues found to be critical for DNA binding are denoted by red arrowheads. (B) Crystal structure of the ChiS DNA-binding domain. The structural elements are color-coded as depicted in the primary sequence in panel A. (C) Alignment of the ChiS trihelical bundle (rainbow) with the LacI trihelical bundle bound to the LacI operator site (PDB code 1EFA; pink). Alignment of alpha carbons gave an RMSD of 3.514. (D) Cartoon representations of the trihelical bundle from LacI and ChiS. Helices are labeled with nomenclature presented in reference 11. (E) Structure of the ChiS trihelical bundle colored to represent the B factor. Helices found in the helix-sheet-helix motif (H2 and H3) are indicated.

the canonical helix-turn-helix (HTH) used by diverse DNA-binding proteins (Fig. 3A and B). The basic HTH domain consists of a trihelical bundle where the second and third helix encompass the namesake helix-turn-helix (11). The two helices that compose the HTH are connected via a relatively short linker that forms a sharp turn, which is a characteristic feature of this domain. Helix 3 from the HTH is generally inserted into the major groove of DNA, forming the principal DNA-protein interface. Alignment of the trihelical bundle from ChiS with the DNA-bound structure of the LacI repressor (PDB code 1EFA [12]; root mean square deviation [RMSD] of modeled C_α carbons = 3.514) revealed a similar spatial arrangement for each helix (Fig. 3C). In addition, LacI and ChiS have similar electrostatic properties, suggesting that a positively charged protein interface that interacts with DNA is a conserved feature of both proteins (Fig. S3). Notably, however, the ChiS HTH has an insertion containing two anti-parallel β-strands connected by a turn between helix 2 and helix 3 that form a β-sheet (Fig. 3B to D). Structural insertion between these helices is not typical; thus, the sheet feature found here is a distinct variant of the HTH which we refer to as a helix-sheet-helix. Comparison of the ChiS trihelical bundle to other structures in the PDB using the DALI server (13) did not reveal any structures that resemble the helix-sheet-helix described here, suggesting that this structure represents a new variant of the HTH.

Alignment of the ChiS DBD to LacI also revealed that the sheet within the ChiS helix-sheet-helix domain runs along the major groove (Fig. 3C and 4A), though it sterically conflicts with the DNA bases. This may suggest that the ChiS DBD takes on a slightly different conformation when bound to DNA. Consistent with this idea, the β-sheet insertion has the highest B factor (a measure of structural motion) in the ChiS DBD structure, indicating that it is relatively flexible (Fig. 3E). Despite the elevated B factor in this region, an omit map indicates that the antiparallel beta strands of the sheet are strongly supported by the data collected (Fig. S4). We speculate that this

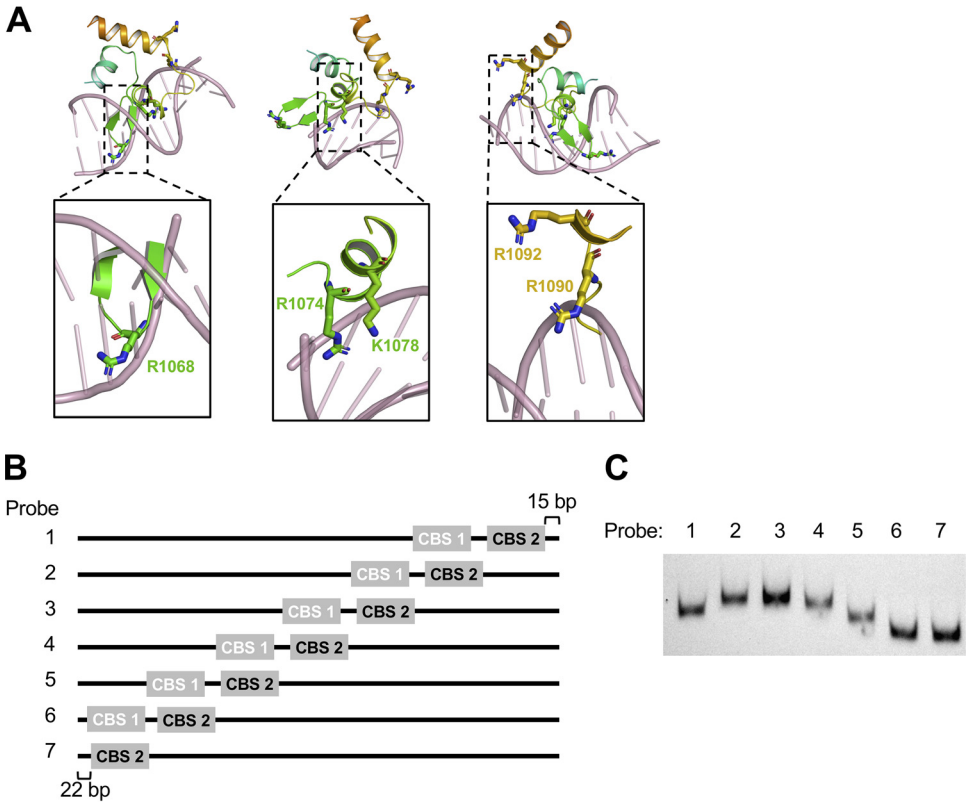

FIG 4 ChiS may bind to intrinsically bent DNA. (A) Model of the ChiS trihelical bundle bound to double-stranded DNA from the alignment shown in Fig. 3C. Side chains for the residues critical for DNA binding (R1068, R1074, K1078, R1090, and R1092) are indicated. (B) Diagram of the 7 distinct 230-bp probes used in panel C. ChiS binding site 1 (CBS 1) was mutated (white text), and ChiS binding site 2 (CBS 2) was left intact (black text). CBS 2 was shifted by 30 bp between probes. (C) The DNA probes diagrammed in panel B were labeled with Cy5 and separated by native PAGE in the absence of ChiS protein.

β-sheet is stabilized in the major groove when the ChiS DBD is bound to DNA. The unique helix-sheet-helix feature of the ChiS C-terminal domain may also explain why it was not previously identified as a DBD by structure prediction algorithms like Phyre2.

ChiS may bind to intrinsically bent DNA. Above, we identified five residues (R1068, R1074, K1078, R1090, and R1092) that were critical for the ChiS DBD to bind to DNA. Mapping these residues onto the ChiS DBD structure revealed that all five residues were found within the trihelical bundle that forms the helix-sheet-helix (Fig. 4A), which is consistent with this domain playing a critical role in DNA binding. Specifically, these residues were located in the β-sheet of the helix-sheet-helix (R1068), helix 3 (R1074, K1078), and helix 1 (R1090, R1092).

Most residues critical for DNA binding activity (R1068, R1074, K1078, and R1090) were in close proximity to DNA on our modeled alignment. Based on the model, we can speculate on the DNA contacts made by these residues. R1068 is found in the sheet of the helix-sheet-helix, which, as stated above, sterically conflicts with DNA bases on our modeled alignment. Thus, it is unclear whether R1068 would make contact with the DNA backbone or with the nucleotide bases. R1074 and K1078 model closest to the nucleotide bases, suggesting that these residues may be critical for base pair recognition. R1090, on the other hand, potentially makes contacts with the DNA backbone.

While the above-mentioned residues modeled closely to DNA, one residue (R1092) was distant from the DNA (Fig. 4A). Many transcription factors bend DNA upon binding to their target site (14, 15). Thus, one possible explanation for the critical role of R1092 is that the P_{chb} promoter is bent when bound by ChiS, which would allow R1092 to

come into close contact with DNA. To test this idea, we carried out a classic *in vitro* gel mobility shift assay to test DNA bending (16). This assay operates on the basis that the location of a bend within a DNA molecule alters its mobility during native PAGE analysis (17, 18). DNA probes that contain a bend in the middle of the probe exhibit the lowest mobility, while probes with the bend closer to one end show the highest mobility. Thus, we designed 7 DNA probes of equal length that gradually shifted the position of the ChiS binding sites within the *chb* promoter (Fig. 4B and Fig. S1). First, we ran these probes in the absence of ChiS protein and found that they ran at different mobilities where the probes with the ChiS binding sites in the middle exhibited the lowest mobility (Fig. 4C). This suggested that the *chb* promoter likely has an intrinsic bend that is centered around the ChiS binding sites. The mobility pattern observed for these DNA probes did not change when they were incubated with the purified ChiS DBD (Fig. S5), suggesting that binding of the DNA probe by ChiS does not further bend the promoter. We propose that the *chb* promoter has an intrinsic bend, which may allow residues in the ChiS DBD, like R1092, to directly interact with DNA. The intrinsic bend found in the *chb* promoter may increase the affinity of ChiS for this region of DNA; indeed, DNA bending has been shown to increase the affinity of certain transcription factors for their DNA binding site (19).

The ChiS-family DNA-binding domain is associated with variable domain arrangements in diverse proteins. Above, we show that the ChiS DBD represents a cryptic variant of an HTH domain. As noted previously, the ChiS DBD is found in proteins other than homologs of ChiS (5). To more fully catalog proteins that contain this domain, we generated a profile hidden Markov model (HMM) to the ChiS DBD and screened for its presence among eubacterial genomes. A profile HMM is a position-specific scoring system that can effectively encode the variation in a training set of representative peptide sequences and then find similar sequences from a much larger and more distantly related data set compared to tools that do not require training, such as BLAST (20, 21).

This analysis revealed that the ChiS DBD is present in diverse proteobacterial genomes (Data Set S1). The vast majority of hits from our search were direct homologs of ChiS (3,242/3,829 [84.7%]); however, many proteins exhibited distinct domain architectures (587/3,829 [15.3%]) (Fig. 5A). Strikingly, the ChiS DBD was found exclusively at the C terminus in all of these proteins and was commonly associated with sensory domains (Fig. 5A). The residues found to be critical for DNA binding in Fig. 2 had various degrees of conservation with ChiS-family DBD-containing proteins (Fig. 5B). R1068 is poorly conserved, suggesting that this residue may be involved in sequence-specific interactions with DNA. Consistent with this idea, R1068 may interact closely with the nucleotide bases (Fig. 3A). R1074 and R1092 are somewhat conserved and K1078 and R0190 are very well conserved across several proteins. This suggests that these residues of the ChiS-family DBD may be required for general DNA interactions and do not contribute to sequence specificity. In general, the helix-sheet-helix is highly conserved across these diverse proteins (Fig. 5B; Data Set S1), and even the most dissimilar ChiS DBD homolog (MAC43155.1; bit score of 43.5; 22.6% identical and 43.4% similar to the ChiS DBD) still threaded (9) remarkably well onto the trihelical bundle of the ChiS DBD structure (RMSD of modeled C_α carbons = 0.002) (Fig. S6). Thus, we suggest that ChiS is the founding member for a new group of DNA-binding transcription factors whose activity is regulated by diverse sensory inputs.

In this study, we characterized the first member of the ChiS-family of DBDs. Though many DBDs have been extensively studied, this work demonstrates that subtle structural variants of canonical DBDs can be difficult to identify by structural prediction algorithms, like Phyre2. Further, the findings here suggest that many proteins with DBDs may currently elude detection. As many residues required for DNA binding in the ChiS DBD are well conserved, our data suggest that there is a common mechanism of binding DNA among the ChiS-family DBDs. Our work also indicates that the canonical tight turn of the HTH is not a critical feature for sequence-specific DNA binding and further highlights the diversity in structural solutions that can allow this type of

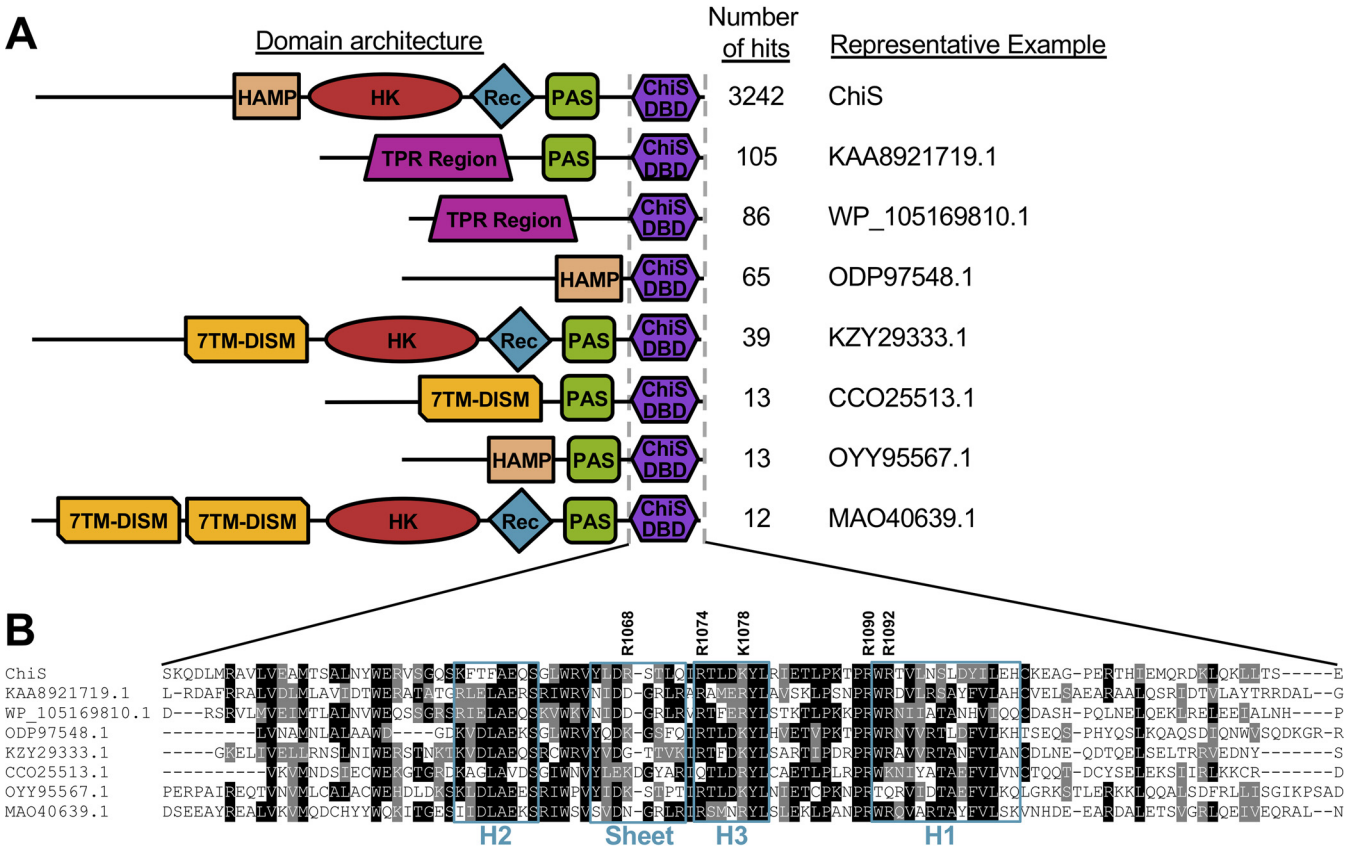

FIG 5 The ChiS-family DBD is found in diverse proteins with distinct domain architectures among proteobacterial genomes. (A) Diagrams of the most abundant protein architectures containing the ChiS-family DBD. Protein domains shown are HAMP, histidine kinase (HK), receiver (Rec), Per-Arnt-Sim (PAS), tetratricopeptide repeat (TPR), 7-transmembrane receptors with diverse intracellular signaling modules (7TMR-DISM), and the ChiS-family DNA-binding domain (ChiS DBD). For a complete list of hits containing the indicated architectures see Data Set S1. (B) Alignment of the primary sequences of the ChiS-family DBD in the indicated proteins. Residues in black are identical, while those in gray are similar. The sequences for helix 1 (H1), helix 2 (H2), helix 3 (H3), and the sheet of the helix-sheet-helix are boxed in teal. The five R and K residues found to be critical for DNA binding in ChiS are also indicated.

activity. While we have generated a putative model of the ChiS DBD bound to DNA in this study, it remains unclear how the sheet within the helix-sheet-helix contributes to sequence-specific DNA binding. Solving the structure of a ChiS-family DBD bound to DNA will be the focus of future work.

MATERIALS AND METHODS

Bacterial strains and culture conditions. All *V. cholerae* strains used in this study are derived from the El Tor strain E7946 (22). *V. cholerae* strains were grown in LB medium and on LB agar supplemented when necessary with carbenicillin (20 μg/ml), kanamycin (50 μg/ml), spectinomycin (200 μg/ml), and/or trimethoprim (10 μg/ml). See Table S2 for a detailed list of mutant strains used in this study.

Generating mutant strains. *V. cholerae* mutant constructs were generated using splicing-by-overlap extension exactly as previously described (23). See Table S3 for all of the primers used to generate mutant constructs in this study. Mutant *V. cholerae* strains were generated by chitin-dependent natural transformation and cotransformation exactly as previously described (24). Mutant strains were confirmed by PCR and/or sequencing.

Cloning and protein production and purification. The *chiS*$^{1024–1129}$ construct was cloned into an Amp^r pET15b-based vector using the FastCloning method (25). This vector appended a tobacco etch virus (TEV) cleavable 6× His tag onto the N terminus of ChiS$^{1024–1129}$. Vector and inserts were amplified using the primers listed in Table S3. The plasmid was transformed into *Escherichia coli* BL21(DE3) (Magic) cells (26), and the protein was expressed in M9 medium (high-yield M9 Se-Met medium; Medicilon, Inc.). The starting overnight culture was grown in LB medium supplemented with 130 μg/ml ampicillin and 50 μg/ml kanamycin at 37°C and 220 rpm. The next day, M9 medium supplemented with 200 μg/ml ampicillin and 50 μg/ml kanamycin was inoculated with the overnight culture (1:100 dilution) and incubated at 37°C and 220 rpm. Protein expression was induced at an optical density at 600 nm (OD_{600}) of 1.8 to 2.0 by the addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside, and the culture was further incubated at 25°C and 200 rpm for 14 h (27). The cells were harvested by centrifugation at 6,000 × *g* for

10 min, resuspended to 0.2 g/ml in lysis buffer (50 mM Tris [pH 8.3], 0.5 M NaCl, 10% glycerol, 0.1% IGEPAL CA-630), and frozen at −30°C until purification.

Frozen pellets were thawed and sonicated at 50% amplitude, in a 5-s-on, 10-s-off cycle for 20 min at 4°C. The lysate was clarified by centrifugation at 18,000 × *g* for 40 min at 4°C, and the supernatant was collected. The protein was purified in one step by immobilized-metal affinity chromatography (IMAC) followed by size exclusion chromatography using an ÅKTAxpress system (GE Healthcare) as previously described with some modifications (28). The cell extract was loaded into a His-Trap FF (nickel-nitrilotriacetic acid [Ni-NTA]) column with loading buffer [10 mM Tris-HCl (pH 8.3), 500 mM NaCl, 1 mM tris(2-carboxyethyl) phosphine (TCEP), 5% glycerol], and the column was washed with 10 column volumes of loading buffer and 10 column volumes of washing buffer (10 mM Tris-HCl [pH 8.3], 1 M NaCl, 25 mM imidazole, 5% glycerol). Protein was eluted with elution buffer (10 mM Tris [pH 8.3], 500 mM NaCl, 1 M imidazole), loaded onto a Superdex 200 26/600 column, separated in loading buffer, collected, and analyzed by PAGE. The 6× His tag was cleaved with recombinant TEV protease in a ratio of 1:20 (protein to protease) overnight at room temperature. The cleaved protein was separated from uncleaved protein, recombinant TEV protease, and 6× His tag peptide by Ni-NTA affinity chromatography using loading buffer followed by loading buffer with 25 mM imidazole. The cleaved protein was collected in the flow-through fraction in both the loading buffer and the loading buffer with 25 mM imidazole. Both fractions were analyzed by PAGE for 6× His tag cleavage, concentrated to 6 to 8 mg/ml, and set up for crystallization.

Crystallization, data collection, structure solution, and refinement. The protein from both fractions (collected in flowthrough and in 25 mM imidazole) was set up at 6 to 8 mg/ml in loading buffer containing 0 or 500 mM NaCl as 2-μl crystallization drops (1 μl protein in 1 μl reservoir solution) in 96-well plates (Corning) using commercial Classics II, PACT, and JCSG+ (Qiagen) crystallization screens. A diffraction-quality crystal of the protein collected with 25 mM imidazole grown from the condition with 0.2 M lithium sulfate, 0.1 M bis-Tris (pH 5.5), and 25% (wt/vol) PEG 3350 (Classics II; no. 74) was flash frozen in liquid nitrogen for data collection.

The crystals were screened, and data were collected at the Life Sciences-Collaborative Access Team (LS-CAT) beamline F at the Advanced Photon Source (APS) of the Argonne National Laboratory. A total of 300 diffraction images were indexed, integrated and scaled using HKL-3000 (29). The structure was determined with the HKL-3000 structure solution package using anomalous signal from selenomethionine (Se-Met). The initial model went through several rounds of refinement in REFMAC v. 5.8.0258 (30) and manual corrections in Coot (31). The water molecules were generated using ARP/wARP (32), and ligands were added to the model manually during visual inspection in Coot. Translation-libration-screw (TLS) groups were created by the TLSMD server (33), and TLS corrections were applied during the final stages of refinement. MolProbity (34) was used for monitoring the quality of the model during refinement and for the final validation of the structure. Structural diagrams were drawn from PDB files using the PyMOL Molecular Graphics System v2.4 (Schrödinger, Inc.).

EMSAs. Binding reaction mixtures contained 10 mM Tris HCl (pH 7.5), 1 mM EDTA, 10 mM KCl, 1 mM dithiothreitol (DTT), 50 μg/ml bovine serum albumin (BSA), 0.1 mg/ml salmon sperm DNA, 5% glycerol, a 1 nM concentration of a Cy5-labeled DNA probe, and purified ChiS DBD at the indicated concentrations (diluted in 10 mM Tris [pH 7.5], 10 mM KCl, 1 mM DTT, and 5% glycerol). Reaction mixtures were incubated at room temperature for 20 min in the dark and then electrophoretically separated on polyacrylamide gels in 0.5× Tris-borate-EDTA (TBE) buffer at 4°C. Gels were imaged for Cy5 fluorescence on a Typhoon-9210 instrument. Cy5-labeled P_{chb} probes were made by Phusion PCR, where Cy5-dCTP was included in the reaction mixture at a level that would result in incorporation of 1 or 2 Cy5-labeled nucleotides in the final probe as previously described (23).

Measuring GFP reporter fluorescence. GFP fluorescence was determined essentially as previously described (35). Briefly, single colonies were picked and grown in LB broth at 30°C for 18 h. Cells were then washed and resuspended to an OD_{600} of 1.0 in Instant Ocean medium (7 g/liter; Aquarium Systems). Then, fluorescence was determined using a BioTek H1M plate reader with excitation set to 500 nm and emission set to 540 nm.

ChIP-qPCR assays. Chromatin immunoprecipitation (ChIP) assays were carried out exactly as previously described (5). Briefly, overnight cultures were diluted to an OD_{600} of 0.08 and then grown for 6 h at 30°C. Cultures were cross-linked using 1% paraformaldehyde, then quenched with a 1.2 molar excess of Tris. Cells were washed with PBS and stored at −80°C overnight. The next day, cells were resuspended in lysis buffer {1× FastBreak cell lysis reagent [Promega], 50 μg/ml lysozyme, 1% Triton X-100, 1 mM PMSF, and 1× protease inhibitor cocktail; 100× inhibitor cocktail contained 0.07 mg/ml phosphoramidon [Santa Cruz], 0.006 mg/ml bestatin [MP Biomedicals/Fisher Scientific], 1.67 mg/ml AEBSF [4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride; DOT Scientific], 0.07 mg/ml pepstatin A [Gold Bio], 0.07 mg/ml E64 [Gold Bio]} and then lysed by sonication, resulting in a DNA shear size of ~500 bp. Lysates were incubated with anti-FLAG M2 magnetic beads (Sigma) and washed to remove unbound proteins, and then bound protein-DNA complexes were eluted off with SDS. Samples were digested with proteinase K, and then cross-links were reversed. DNA samples were cleaned up and used as the template for quantitative PCR (qPCR) using iTaq universal SYBR green supermix (Bio-Rad) and primers specific for the genes indicated (see Table S3 for primers) on a Step-One qPCR system. Standard curves of genomic DNA were included in each experiment and were used to determine the abundance of each amplicon in the input (derived from the lysate prior to ChIP) and output (derived from the samples after ChIP). Primers to amplify *rpoB* served as a baseline control in this assay because ChiS does not bind this locus. Data are reported as fold enrichment, which is defined as the ratio of P_{chb}/*rpoB* found in the output divided by the same ratio found in the input.

Western blot analysis. Strains were grown as described for ChIP assays, pelleted, resuspended, and boiled in 1× SDS-PAGE sample buffer (110 mM Tris [pH 6.8], 12.5% glycerol, 0.6% SDS, 0.01% bromophenol blue, and 2.5% β-mercaptoethanol). Proteins were separated by SDS-polyacrylamide gel electrophoresis, then transferred to a polyvinylidene difluoride (PVDF) membrane, and probed with rabbit polyclonal anti-FLAG (Sigma) or mouse monoclonal anti-RpoA (BioLegend) primary antibodies. Blots were then incubated with horseradish peroxidase (HRP)-conjugated anti-rabbit or anti-mouse immunoglobulin secondary antibodies, developed using Pierce ECL 529 Western blotting substrate (Thermo Fisher), and imaged on a ProteinSimple Fluorchem E instrument.

Bioinformatic identification of eubacterial proteins with putative ChiS DBD domains. The DBD sequence segments from the protein sequences of seven ChiS DNA-binding domain homologs (THB81618.1, OGG93021.1, OUR95018.1, WP_084205767.1, ODU31202.1, WP_070993003.1, and WP_078715702.1) (5) were aligned using MUSCLE version 3.8.31 (36). The resulting multiple-sequence alignment was turned into a profile HMM which was searched against the eubacterial subset (taxonomy ID: 2) of the NCBI nonredundant protein sequence database using HMMER version 3.2.1 (http://hmmer.org/), requiring the alignment length to be at least 90. Among the hits, protein sequences tagged as "partial" in their FASTA headers were excluded. Domain architectures for the remaining hits were obtained from the NLM conserved-domain database (37). Any protein hits with regions aligned to the DNA-binding domain HMM overlapping with known annotated functional domains were excluded. The resulting ChiS DBD homolog protein sequences were clustered using cd-hit v. v4.8.1-2019-0228 (38) (parameters: -M, 0; -g, 1; -s, 0.8; -c, 0.4; -n, 2; -d, 500). Clusters identified by cd-hit were further grouped together by manually analyzing the domain architecture of hits as shown in Fig. 5A. Only clusters containing 10 or more representatives were grouped, while the remaining proteins were left unassigned. For a list of all proteins containing a putative ChiS DBD, see Data Set S1.

Data availability. The structure was deposited in the Protein Data Bank (https://www.rcsb.org/) with the assigned PDB code 7KPO.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

DATA SET S1, XLSX file, 0.4 MB.
FIG S1, PDF file, 0.05 MB.
FIG S2, PDF file, 0.1 MB.
FIG S3, PDF file, 2 MB.
FIG S4, PDF file, 2.6 MB.
FIG S5, PDF file, 1.2 MB.
FIG S6, PDF file, 0.3 MB.
TABLE S1, DOCX file, 0.01 MB.
TABLE S2, DOCX file, 0.02 MB.
TABLE S3, DOCX file, 0.02 MB.

ACKNOWLEDGMENTS

We thank Dipankar Sen, Julia van Kessel, and Ryan Chaparian for helpful discussions.

This work was supported by grant R35GM128674 from the National Institutes of Health (to A.B.D.) and, in part, with Federal funds from the Department of Health and Human Services, National Institutes of Health, National Institute of Allergy and Infectious Diseases, under contract no. HHSN272201700060C. This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science user facility operated for the DOE Office of Science by Argonne National Laboratory under contract no. DE-AC02-06CH11357. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (Grant 085P1000817). This research was supported in part by Lilly Endowment, Inc., through its support for the Indiana University Pervasive Technology Institute.

REFERENCES

1. Pruzzo C, Vezzulli L, Colwell RR. 2008. Global impact of Vibrio cholerae interactions with chitin. Environ Microbiol 10:1400–1410. https://doi.org/10.1111/j.1462-2920.2007.01559.x.
2. Hunt DE, Gevers D, Vahora NM, Polz MF. 2008. Conservation of the chitin utilization pathway in the Vibrionaceae. Appl Environ Microbiol 74:44–51. https://doi.org/10.1128/AEM.01412-07.
3. Nahar S, Sultana M, Naser MN, Nair GB, Watanabe H, Ohnishi M, Yamamoto S, Endtz H, Cravioto A, Sack RB, Hasan NA, Sadique A, Huq A, Colwell RR, Alam M. 2011. Role of shrimp chitin in the ecology of toxigenic Vibrio cholerae and cholera transmission. Front Microbiol 2:260. https://doi.org/10.3389/fmicb.2011.00260.
4. Meibom KL, Blokesch M, Dolganov NA, Wu CY, Schoolnik GK. 2005. Chitin induces natural competence in Vibrio cholerae. Science 310:1824–1827. https://doi.org/10.1126/science.1120096.
5. Klancher CA, Yamamoto S, Dalia TN, Dalia AB. 2020. ChiS is a noncanonical DNA-binding hybrid sensor kinase that directly regulates the chitin utilization program in Vibrio cholerae. Proc Natl Acad Sci U S A 117:20180–20189. https://doi.org/10.1073/pnas.2001768117.

6. Li X, Roseman S. 2004. The chitinolytic cascade in Vibrios is regulated by chitin oligosaccharides and a two-component chitin catabolic sensor/kinase. Proc Natl Acad Sci U S A 101:627–631. https://doi.org/10.1073/pnas.0307645100.
7. Meibom KL, Li XB, Nielsen AT, Wu CY, Roseman S, Schoolnik GK. 2004. The Vibrio cholerae chitin utilization program. Proc Natl Acad Sci U S A 101:2524–2529. https://doi.org/10.1073/pnas.0308707101.
8. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.
9. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc 10:845–858. https://doi.org/10.1038/nprot.2015.053.
10. Luscombe NM, Laskowski RA, Thornton JM. 2001. Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. Nucleic Acids Res 29:2860–2874. https://doi.org/10.1093/nar/29.13.2860.
11. Aravind L, Anantharaman V, Balaji S, Babu MM, Iyer LM. 2005. The many faces of the helix-turn-helix domain: transcription regulation and beyond. FEMS Microbiol Rev 29:231–262. https://doi.org/10.1016/j.femsre.2004.12.008.
12. Bell CE, Lewis M. 2000. A closer view of the conformation of the Lac repressor bound to operator. Nat Struct Biol 7:209–214. https://doi.org/10.1038/73317.
13. Holm L. 2020. DALI and the persistence of protein shape. Protein Sci 29:128–140. https://doi.org/10.1002/pro.3749.
14. Vamosi G, Rueda D. 2018. DNA bends the knee to transcription factors. Biophys J 114:2253–2254. https://doi.org/10.1016/j.bpj.2017.10.047.
15. van der Vliet PC, Verrijzer CP. 1993. Bending of DNA by transcription factors. Bioessays 15:25–32. https://doi.org/10.1002/bies.950150105.
16. Kim J, Zwieb C, Wu C, Adhya S. 1989. Bending of DNA by gene-regulatory proteins: construction and use of a DNA bending vector. Gene 85:15–23. https://doi.org/10.1016/0378-1119(89)90459-9.
17. Koo HS, Wu HM, Crothers DM. 1986. DNA bending at adenine·thymine tracts. Nature 320:501–506. https://doi.org/10.1038/320501a0.
18. Wu HM, Crothers DM. 1984. The locus of sequence-directed and protein-induced DNA bending. Nature 308:509–513. https://doi.org/10.1038/308509a0.
19. Afek A, Shi H, Rangadurai A, Sahay H, Senitzki A, Xhani S, Fang M, Salinas R, Mielko Z, Pufall MA, Poon GMK, Haran TE, Schumacher MA, Al-Hashimi HM, Gordân R. 2020. DNA mismatches reveal conformational penalties in protein-DNA recognition. Nature 587:291–296. https://doi.org/10.1038/s41586-020-2843-2.
20. Yoon BJ. 2009. Hidden Markov models and their applications in biological sequence analysis. Curr Genomics 10:402–415. https://doi.org/10.2174/138920209789177575.
21. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402. https://doi.org/10.1093/nar/25.17.3389.
22. Miller VL, DiRita VJ, Mekalanos JJ. 1989. Identification of toxS, a regulatory gene whose product enhances toxR-mediated activation of the cholera toxin promoter. J Bacteriol 171:1288–1293. https://doi.org/10.1128/jb.171.3.1288-1293.1989.
23. Dalia AB, Lazinski DW, Camilli A. 2013. Characterization of undermethylated sites in Vibrio cholerae. J Bacteriol 195:2389–2399. https://doi.org/10.1128/JB.02112-12.
24. Dalia AB. 2018. Natural cotransformation and multiplex genome editing by natural transformation (MuGENT) of Vibrio cholerae. Methods Mol Biol 1839:53–64. https://doi.org/10.1007/978-1-4939-8685-9_6.
25. Li C, Wen A, Shen B, Lu J, Huang Y, Chang Y. 2011. FastCloning: a highly simplified, purification-free, sequence- and ligation-independent PCR cloning method. BMC Biotechnol 11:92. https://doi.org/10.1186/1472-6750-11-92.
26. Kwon K, Peterson SN. 2014. High-throughput cloning for biophysical applications. Methods Mol Biol 1140:61–74. https://doi.org/10.1007/978-1-4939-0354-2_5.
27. Millard CS, Stols L, Quartey P, Kim Y, Dementieva I, Donnelly MI. 2003. A less laborious approach to the high-throughput production of recombinant proteins in Escherichia coli using 2-liter plastic bottles. Protein Expr Purif 29:311–320. https://doi.org/10.1016/S1046-5928(03)00063-9.
28. Shuvalova L. 2014. Parallel protein purification. Methods Mol Biol 1140:137–143. https://doi.org/10.1007/978-1-4939-0354-2_10.
29. Minor W, Cymborowski M, Otwinowski Z, Chruszcz M. 2006. HKL-3000: the integration of data reduction and structure solution–from diffraction images to an initial model in minutes. Acta Crystallogr D Biol Crystallogr 62:859–866. https://doi.org/10.1107/S0907444906019949.
30. Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA. 2011. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr 67:355–367. https://doi.org/10.1107/S0907444911001314.
31. Emsley P, Cowtan K. 2004. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60:2126–2132. https://doi.org/10.1107/S0907444904019158.
32. Morris RJ, Perrakis A, Lamzin VS. 2003. ARP/wARP and automatic interpretation of protein electron density maps. Methods Enzymol 374:229–244. https://doi.org/10.1016/S0076-6879(03)74011-7.
33. Painter J, Merritt EA. 2006. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr D Biol Crystallogr 62:439–450. https://doi.org/10.1107/S0907444906005270.
34. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. 2010. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 66:12–21. https://doi.org/10.1107/S0907444909042073.
35. Dalia AB. 2016. RpoS is required for natural transformation of Vibrio cholerae through regulation of chitinases. Environ Microbiol 18:3758–3767. https://doi.org/10.1111/1462-2920.13302.
36. Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:113. https://doi.org/10.1186/1471-2105-5-113.
37. Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler GH, Song JS, Thanki N, Yamashita RA, Yang M, Zhang D, Zheng C, Lanczycki CJ, Marchler-Bauer A. 2020. CDD/SPARCLE: the conserved domain database in 2020. Nucleic Acids Res 48:D265–D268. https://doi.org/10.1093/nar/gkz991.
38. Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658–1659. https://doi.org/10.1093/bioinformatics/btl158.