The crystal structure of YycH involved in the regulation of the essential YycFG two-component system in Bacillus subtilis reveals a novel tertiary structure

Hendrik Szurmant, Haiyan Zhao, Michael A. Mohan, James A. Hoch and Kottayil I. Varughese

Protein Sci. 2006 15: 929-934
doi:10.1110/ps.052064406

PROTEIN STRUCTURE REPORT

The crystal structure of YycH involved in the regulation of the essential YycFG two-component system in *Bacillus subtilis* reveals a novel tertiary structure

HENDRIK SZURMANT,[1] HAIYAN ZHAO,[1] MICHAEL A. MOHAN, JAMES A. HOCH, AND KOTTAYIL I. VARUGHESE

Division of Cellular Biology, Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, California 92037, USA

(RECEIVED December 27, 2005; FINAL REVISION January 11, 2006; ACCEPTED January 12, 2006)

Abstract

The *Bacillus subtilis* YycFG two-component signal transduction system is essential for cell viability, and the YycH protein is part of the regulatory circuit that controls its activity. The crystal structure of YycH was solved by two-wavelength selenium anomalous dispersion data, and was refined using 2.3 Å data to an *R*-factor of 25.2%. The molecule is made up of three domains, and has a novel three-dimensional structure. The N-terminal domain features a calcium binding site and the central domain contains two conserved loop regions.

Keywords: YycH; signal transduction; YycF, YycG; histidine kinase; response regulator; calcium binding

Supplemental material: see www.proteinscience.org

Living organisms have to sample their environment and respond to it in order to maximize their chance for survival. Bacteria commonly employ so-called two-component systems to achieve this task (for review, see Hoch 2000). Typically, a transmembrane protein, the sensor histidine kinase with an extracytoplasmic sensing domain and a cytoplasmic kinase domain, is activated by autophosphorylation in response to a stimulus. The kinase phosphoryl group is transferred to a cytoplasmic response regulator protein to which it is mated, that upon phosphorylation modulates expression levels of target genes.

The YycFG two-component system is the only signal transduction system in *Bacillus subtilis* known to be essential for cell viability (Fabret and Hoch 1998). This system is highly conserved among the firmicutes, and essentiality has been shown for multiple organisms, including the important pathogens *Staphylococcus aureus* and *Streptococcus pneumoniae* (Martin et al. 1999; Wagner et al. 2002). With the exception of the Streptococci, two unique open reading frames, *yycH* and *yycI,* are always found immediately downstream of the response regulator protein YycF and the histidine kinase protein YycG, and they are organized within the same operon (Ng and Winkler 2004; Szurmant et al. 2005). The YycH protein was implicated in regulating the activity of YycG (Szurmant et al. 2005). YycH is a 458 amino acid protein that was shown to be located outside of the

[1]These authors contributed equally to this work.

Reprint requests to: Kottayil I. Varughese, Division of Cellular Biology, Department of Molecular and Experimental Medicine, MEM-116, The Scripps Research Institute, 10550 N. Torrey Pines Road, La Jolla, CA 92037, USA; e-mail: kiv@scripps.edu; fax: (858) 784-7966.

Article and publication are at http://www.proteinscience.org/cgi/doi/10.1110/ps.052064406.

cytoplasmic membrane and tethered to it by a putative transmembrane helix predicated to range from amino acids 12–31. To gain further insight into the mode of function for YycH, a truncated version of the protein, missing 38 N-terminal residues, was expressed and crystallized. The crystals diffracted to a resolution of 2.3 Å and the structure was solved by the multiple anomalous diffraction (MAD) technique (Hendrickson 1991), using selenomethionine (SeMet) protein crystals.

Results and Discussion

Amino acid sequence comparisons

In order to identify YycH orthologs, the *B. subtilis* protein was used as a query in a BLASTP (Altschul et al. 1990) search. Nineteen putative YycH protein sequences of various Gram-positive organisms were found and analyzed for their genomic location of their genes. All of the genes were found to be located downstream of the highly conserved histidine kinase *yycG* genes, confirming that all are true orthologs. The amino acid sequences were aligned using the program CLUSTLW (Thompson et al. 1999) (see supplemental figure S1). The YycH protein shows a very low amino acid sequence conservation even among *Bacillus* species (Fig. 1), not atypical for extra-cytoplasmically located proteins. For instance, only 25% of all amino acid residues in the *B. subtilis* and *Bacillus anthracis* orthologs are identical. Indeed, among all species the few invariant residues include two proline residues (Pro65 and Pro430) as well as one invariant tyrosine residue (Tyr413) toward the C terminus of the protein (see supplemental figure S1). This suggests that YycH is not an enzyme but rather acts through interactions with other molecules, potentially as a signal sensor and/or a modulator of an enzymatic activity. Interestingly, the transmembrane domain includes the remaining invariant residues and shows an unusual high conservation, in particular when compared to the rest of the molecule. The presence of several conserved hydrophilic residues, in particular a serine and a threonine residue within an SXXX(T/S) motif, suggests that YycH might form either a homo or hetero-oligomer through transmembrane helix interaction (Dawson et

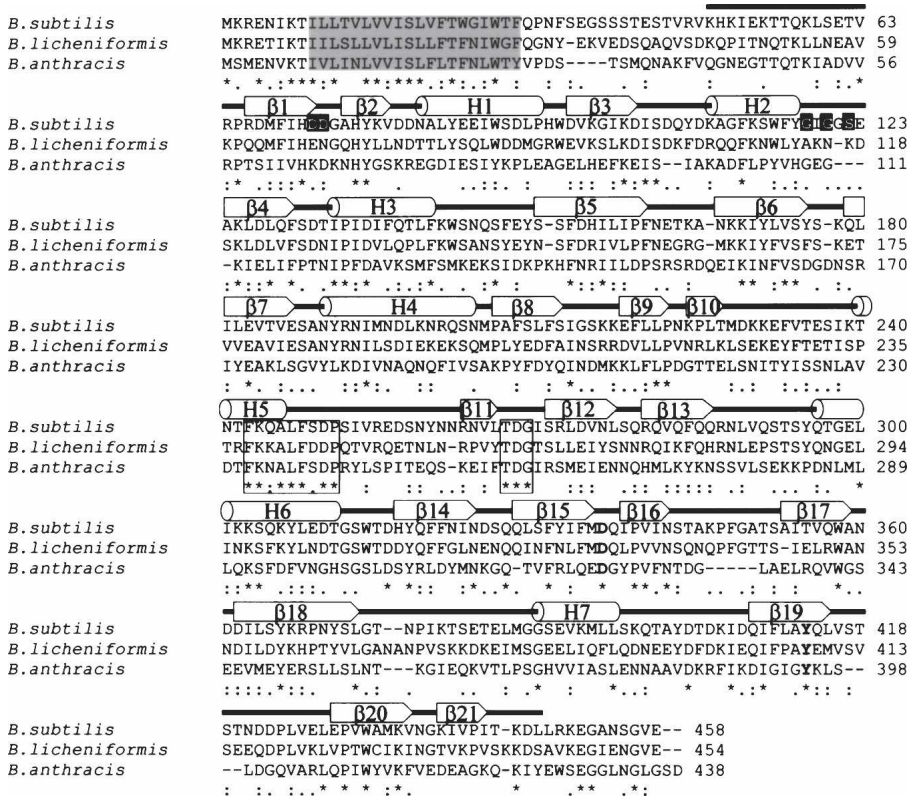

Figure 1. YycH sequence alignment and secondary structure. YycH orthologs from three Bacilli species were aligned using the program CLUSTALW. The secondary structure is shown above the protein sequences. The putative transmembrane domain is shaded in gray. Calcium binding residues are underlain in reverse video. Two conserved loop regions are framed. The invariant residue Tyr413 and the conserved residue Asp336 that are in hydrogen bond distance are in bold. Note that the first three residues of *B. subtilis* YycH were omitted from the alignment and therefore the sequence of *B. subtilis* YycH starts with residue 4.

al. 2002). The soluble protein, truncated for its transmembrane domain and used for crystallization appeared as a monomer in solution as measured by native gel electrophoresis following the method by Hedrick and Smith (1968) (data not shown).

Description of the overall structure

The molecule is made up of three domains, and it is shaped like a pyramid (Fig. 2A). The domain on the top is pyramidal in appearance, while the other two are globular in shape. Eleven N-terminal residues (39–49), 11 C-terminal residues (448–458), and a total of 33 residues from four internal loop regions (257–265, 287–290, 377–382, and 416–429) were not visible in the electron density maps. A brief description of the structure of the three individual domains is given below.

Domain 1

The architecture of the N-terminal domain (residue 39–207) is defined by a highly twisted central β-sheet, which is made up of seven antiparallel β-strands (β1–β7) (Fig. 2A). Two long antiparallel α-helices (H1 and H4) and a short α-helix (H3) lie on one side of the sheet and another short helix (H2) lies on the other side. The next segment contains two antiparallel (β8 and β9) strands followed by a very long loop (residue 223–239) running to the top of the pyramid.

The N-terminal domain possesses a calcium binding site and five oxygen atoms from this domain coordinate with the Ca^{++} atom in octahedral geometry (Fig. 2B). Carbonyls of the three residues Gly118, Gly120, Ser122, and carboxylate oxygen atoms of the side chains of Asp72 and Asp73 are involved in calcium binding. Surprisingly, one of the coordination sites of calcium is unoccupied. As this site is on the surface of the molecule, it is likely to be occupied by a water molecule. On the other hand, the calcium atom could strengthen the interactions of YycH with another protein or ligand by coordinating to one of its oxygen atoms. The presence of Ca^{++} provides rigidity to the loop region carrying residues 118–123 as it coordinates to three carbonyl oxygen atoms from this region. It provides additional stability to this domain by acting as connecting link between two loop regions (Fig. 2B).

Domain 2

The pyramid domain (residue 240–376) is also dominated by a β-sheet comprising seven antiparallel β-strands (β11–β15, β17, and β18), which forms one side of the domain. The other side has two helices (H5 and

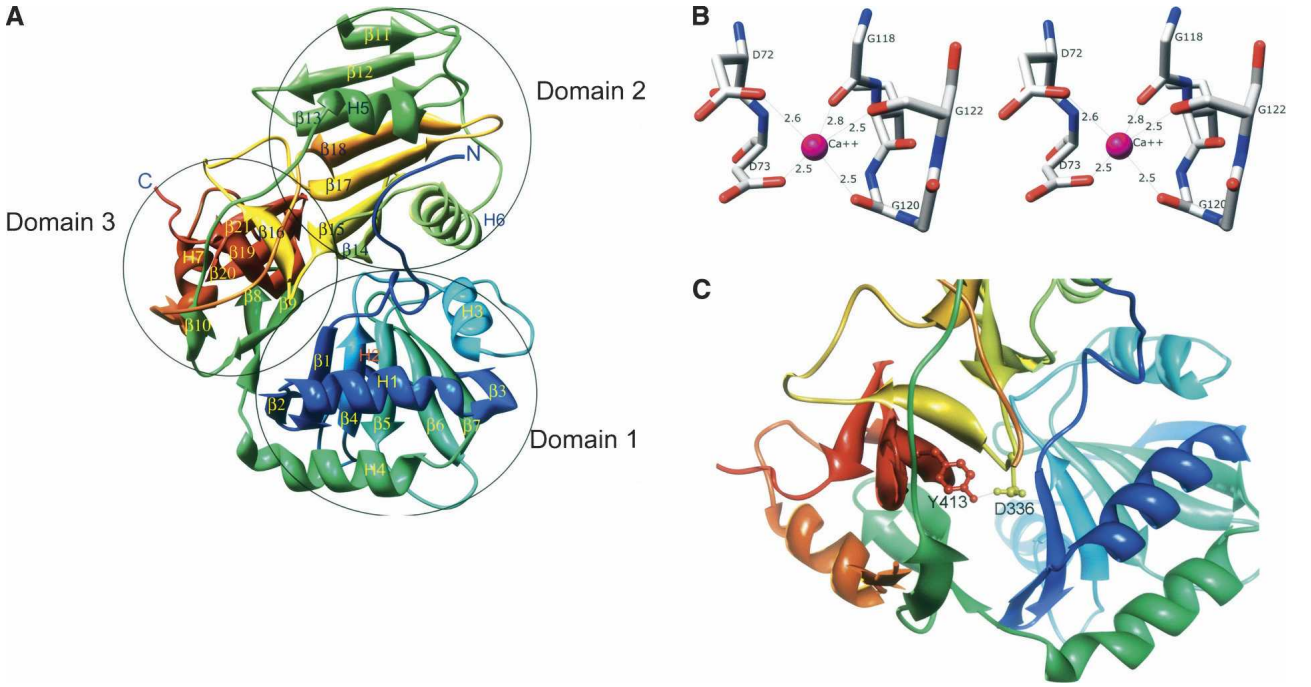

Figure 2. Crystal structure of YycH. (*A*) The molecular architecture of the YycH molecule is depicted in the rainbow color code, with the N-terminal in blue and the C-terminal in red. The three domains are encircled. (*B*) Shown is a stereoscopic view of the calcium binding site. The Ca^{++} is coordinated to five oxygen atoms from domain 1. The apex of the coordination octahedron is unoccupied. The Ca^{++} coordination site is formed by two adjacent loop regions located between strands β5 and β6 and between β7 and α-helix H4. (*C*) The invariant residue Tyr413 forms a hydrogen bond with conserved residue Asp336, and thereby stabilizes the interaction between domains 2 and 3.

H6) and an N-terminal loop (Fig. 2A). The top of the pyramid contains two surface-exposed loops of increased conservation, particularly in *Bacillus* species (Fig. 1). These loops are immediately following α-helix 5 and β-strand 11, respectively. Overall conservation of the protein is low, and most conserved residues are buried, and are therefore likely important for structural integrity. This suggests that the conserved loops may be important for the function of YycH.

Domain 3

The third domain consists of seven β-strands (β8–β10, β16, and β19–21) and an α-helix (H7) (Fig. 2). Unlike the other two domains, the β-strands are not arranged as a single sheet. Interestingly the domain includes internal sequences in addition to the C-terminal residues 380–458. This domain has strong interactions with both the pyramid domain and the base domain. Therefore, domain 3 should be considered as the clamp between domains 1 and 2 (Fig. 2A). Located within this domain is the aforementioned invariant residue Tyr413. This tyrosine is within hydrogen bond distance of the relatively conserved residue Asp336 further stabilizing the interaction between the central and the C-terminal domain (Fig. 2C).

The YycH structure has a novel protein fold. Neither the entire structure of YycH nor the individual domains are similar to any of the structures deposited in the Protein Data Bank. The closest structural homologs identified by the DALI server (Holm and Sander 1993) is a hypothetical protein (PDB code 1TU1) with a *Z* score of 3.8 and an RMSD of 3.5 Å. The *Z* score for individual domains are 2.6, 4.1, and 2.2 for domains 1, 2, and 3, respectively. The corresponding RMSDs are 3.8 Å, 3.5 Å, and 3.0 Å, respectively. Visual analysis supports virtually no structural homology between these proteins and YycH. Therefore, all three YycH domains are clearly novel structures.

Surface analysis of YycH

The three domains of the molecule are packed to ensure some flexibility for the molecule. The buried surface areas among the three domains are relatively low. Between domains 1 and 2, 6.8% of the surface area of domain 1 (equal to 7.0% of the surface area of domain 2) is buried. Between domains 1 and 3 only 4.4% of domain 1 area (equal to 7.1% of the area of domain 3) is buried. Between domains 2 and 3, 7.3% of domain 2 (equivalent of 11.5% of domain 3) is buried.

As domain 3, through its interaction maintains the relative orientations of the other domains, the disruption of these interactions would take the molecule from the observed "closed" state to a possible "open" state. The interaction surfaces between the domain 1 and domain 2 are, in support of this hypothesis, unusually hydrophilic. At this point it is not clear if such flexibility has any bearing on its function.

Analysis of the molecular surface did not yield any striking features. No hydrophobic patches often implicated in protein–protein interactions were observed. There are however, two regions where an aromatic side chain positions itself next to another aromatic residue (residues Phe150 with Tyr152 and residues Phe233 with Phe348, respectively). Since these residues are not conserved even among Bacilli species, they are unlikely to have an important function.

Summary

In summary, features of the protein that give a link to its mode of action are the Ca^{++} binding site in the N-terminal domain, the conserved loop region in the central domain, and possibly the interactions of the C-terminal domain with the previous two, that could allow for an opening of the molecule if disrupted. Given the low number of invariant residues YycH is likely to act through protein interaction either with the histidine kinase YycG or with the YycI protein with which it is always paired. It is unlikely for this protein to have any enzymatic activity.

Materials and methods

Cloning, expression, and purification of YycH

The *B. subtilis* gene encoding for aminoacids 39–458 of *B. subtilis* YycH was PCR amplified introducing 5′ NdeI and 3′ BamHI sites by using the following oligonucleotides (restrictions sites in bold): yycHf: 5′-CCAAACTTTTCT**CATATG**TCATCCTCAACAGAG-3′; yycHr: 5′-GGAAGGCAA**GGATCC**AGATTGATTTTGTC-3′.

The PCR fragment was digested with the above restriction enzymes and cloned into the same sites of the vector pET28a (Novagen) generating pJS06. The construct was confirmed by DNA sequencing. pJS06 was transformed into *Escherichia coli* strain BL21(DE3) (Stratagen) for overexpression of unlabeled and into B834(DE3) (Stratagen) for overexpression of selenomethionine-labeled *yycH*. YycH produced from these strains is missing the 35 N-terminal amino acids, including an N-terminal transmembrane domain and includes a thrombin cleavable N-terminal his-tag.

His-tagged *yycH* was overexpressed by growing 2 L of cells in Luria Bertani broth (unlabeled) or in seleno-methionine supplemented minimal medium to an A_{600} of 0.6 and subsequent induction with 1 mM IPTG for 12 h at 25°C under aeration (250 rpm). Cells were pelleted by centrifugation at 6000*g* for 10 min and washed once with buffer A (50 mM Tris-HCl, 300 mM NaCl at pH 8.0). This process typically yielded 6 g of cells which were stored at −80°C until further processing.

To purify YycH, cells were resuspended in five times their volume buffer A and lysed by repeated passage through a

French press. All subsequent steps were performed at 4°C. The cell lysate was cleared by ultracentrifugation at 50,000*g* for 1 h. Five milliliters of Ni-NTA-Agarose (Qiagen), equilibrated in buffer A was added to the cleared lysate and incubated for 12 h on a rocking platform. The suspension was poured into a column and washed with 100 mL buffer A. YycH eluted in buffer A supplemented with 30 mM imidazole. To remove the N-terminal his-tag, YycH was dialyzed twice in buffer B (50 mM Tris-HCl, 50 mM NaCl at pH 8.0) for 12 h each in the presence of 1 U Thrombin per 1 mg of YycH. Remaining cleaved his-tag and undigested protein was removed by addition of Ni-NTA resin, and thrombin was removed by addition of *p*-amino-benzamidine agarose (Sigma) and subsequent centrifugation following a 3-h incubation time. The supernatant was passed over a desalting column equilibrated in buffer C (5 mM Tris-HCl, 50 mM NaCl at pH 8.0) and concentrated to 7 mg/mL for subsequent crystallization trials. This purification procedure yielded in excess of 100 mg YycH essentially pure as judged by SDS-PAGE with subsequent Coomassie staining.

Protein crystallization

Crystals were grown using hanging drop techniques. The initial screening was performed with kits I and II from Hampton Research Inc., as well as Wizard kits I and II. One condition that yielded small crystals was subsequently optimized to 0.1 M Tris-HCl (pH 7), 18% PEG 3350, 100 mM calcium acetate to obtain good diffraction quality crystals. The presence of at least 100 mM calcium acetate was necessary to achieve crystallization. Seleno-methionine-labeled YycH was also crystallized under the same conditions, but it required slightly less PEG 3350 (17%–17.5%) than the native protein. Under these conditions crystals appeared within 2–3 d, and typically achieved maximum size (0.3× 0.3× 0.4 mm) after 7–10 d.

Data collection and structure determination

As the crystals were found to be intergrown, a single crystal was carefully separated from a doublet comprising two single crystals. The separated crystal was then picked from the drop with a nylon loop and transferred to a cryoprotectant solution composed of equilibration buffer supplemented with 10% PEG200. Crystals were then flash frozen directly in a liquid nitrogen vial and transferred to a 100 K cryostream. Diffraction data were collected on beamline 8.3.1 at the Advanced Light Source using an ADSC Quantum-210 CCD detector. Diffraction data of the Seleno-Met crystal were collected at two wavelengths to a resolution of 2.3 Å using the inverse beam geometry in wedges of 30° (Table 1). One data set was collected at wavelength = 0.97958 Å between the selenium edge and the inflection points. The remote data set was collected at a higher wavelength (1.01978 Å). The data sets were processed with MOSFLM (Leslie 1992). The native data were collected to a resolution of 2.3 Å, and were processed and scaled with DENZO1.97 and SCALEPACK (Otwinowski and Minor 1997). Out of the total of eight selenium atoms in the protein, seven were located using the program SOLVE (Terwilliger and Berendzen 1999). Protein phases to 2.3 Å were calculated using SOLVE, and solvent flattening calculations were performed using RESOLVE (Terwilliger 2000), which located a total of 190 residues in the automated mode. An additional 184 residues were fitted manually using the electron density map produced by RESOLVE. Refinement of the structure was carried out using the program package CNS1.1 (Brunger et al. 1998). The final refinement was performed using 2.3 Å native data. The final model, comprising 374 residues, one calcium ion, and 94 water molecules, satisfies the

Table 1. *Crystallographic data*

	λ1 (Remote)	λ2 (Peak inflection)	Native λ3
Data collection			
Wavelength	1.01978 Å	0.97958 Å	0.97625 Å
Resolution (Å)	2.28 (2.28–2.47)	2.28 (2.28–2.47)	2.15 (2.15–2.25)
Total reflections	281,281	290,190	188,482
Unique reflections	18,545	18,580	18,254
Completeness (%)	98.8 (95.7)	99.6 (99.5)	99.4
R_{merge} (%)	6.1 (31.2)	8.9 (36.9)	5.7 (24.5)
$I/\sigma(I)$	15.5 (2.7)	13.7 (2.0)	24.5 (5.6)
Phasing statistics			
Figure of merit	0.49		
Refinement statistics			
Resolution limits (Å)			2.3
Reflection in working/test sets			16,454/1800
R/R_{free} (%)			25.2/30.4
Bond (Å) angle (°) RMS			0.007/1.4
Ramachandran plot			
Most favored regions			85.4%
Additional allowed regions			11.3%
No. of residues			376 residues
No. of cations			1 Calcium
No. of solvent molecules			94 (H_2O)

quality criteria limits of the program PROCHECK (Laskowski et al. 1993)

Accession number

The atomic coordinates and structure factors for YycH (PDB ID 2FGT) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org).

Electronic supplemental material

The supplemental figure S1 shows an alignment of 19 YycH orthologs from various Gram-positive bacteria, demonstrating the low overall conservation and the location of the few invariant residues.

Acknowledgments

This research was supported, in part, by grants GM54246 and GM19416 from the National Institute of General Medical Sciences, grants AI055860 and AI052289 from National Institutes of Allergy and Infectious Diseases, the National Institutes of Health, USPHS. This is publication 18007-MEM from The Scripps Research Institute. We thank the ALS Synchrotron Source at Berkeley for providing X-ray diffraction facilities. We thank Dr. Michael Pique for his help in preparing the cover illustration.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., and Pannu, N.S. 1998. Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54:** 905–921.

Dawson, J.P., Weinger, J.S., and Engelman, D.M. 2002. Motifs of serine and threonine can drive association of transmembrane helices. *J. Mol. Biol.* **316:** 799–805.

Fabret, C. and Hoch, J.A. 1998. A two-component signal transduction system essential for growth of *Bacillus subtilis*: Implications for anti-infective therapy. *J. Bacteriol.* **180:** 6375–6383.

Hedrick, J.L. and Smith, A.J. 1968. Size and charge isomer separation and estimation of molecular weights of proteins by disc gel electrophoresis. *Arch. Biochem. Biophys.* **126:** 155–164.

Hendrickson, W.A. 1991. Determination of macromolecular structures from anomalous difraction of synchrotron radiation. *Science* **254:** 51–58.

Hoch, J.A. 2000. Two-component and phosphorelay signal transduction. *Curr. Opin. Microbiol.* **3:** 165–170.

Holm, L. and Sander, C. 1993. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233:** 123–138.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. 1993. PROCHECK—A program to check the stereochemical quality of protein structures. *J Appl Crystallogr* **26:** 283–291.

Leslie, A. 1992. Recent changes to the MOSFLM package for processing film and image plate data. *CCP4 and ESF-EACMB Newsletter on Protein Crystallography 26*.

Martin, P.K., Li, T., Sun, D., Bick, D.P., and Schmid, M.B. 1999. Role in cell permeability of an essential two-component system in *Staphylococcus aureus*. *J. Bacteriol.* **181:** 3666–3673.

Ng, W. and Winkler, M. 2004. Singular structures and operon organizations of essential two-component systems in species of *Streptococcus*. *Microbiology* **150:** 3096–3098.

Otwinowski, Z. and Minor, W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276:** 307–326.

Szurmant, H., Nelson, K., Kim, E.J., Perego, M., and Hoch, J.A. 2005. YycH regulates the activity of the essential YycFG two-component system in *Bacillus subtilis*. *J. Bacteriol.* **187:** 5419–5426.

Terwilliger, T.C. 2000. Maximum-likelihood density modification. *Acta Crystallogr. D Biol. Crystallogr.* **56:** 965–972.

Terwilliger, T.C. and Berendzen, J. 1999. Automated MAD and MIR structure solution. *Acta Crystallogr.* **D55:** 849–861.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1999. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22:** 4673–4680.

Wagner, C., Saizieu Ad, A., Schonfeld, H., Kamber, M., Lange, R., and Thompson, C.J. 2002. Genetic analysis and functional characterization of the *Streptococcus pneumoniae* vic operon. *Infect. Immun.* **70:** 6121–6128.