FOR THE RECORD

Crystallization, X-ray studies, and site-directed cysteine mutagenesis of the DNA-binding domain of OmpR

ERIK MARTÍNEZ-HACKERT,^{1,2} SUSAN HARLOCKER,³ MASAYORI INOUE,³
HELEN M. BERMAN,² AND ANN M. STOCK^{1,3,4}

¹ Center for Advanced Biotechnology and Medicine, Piscataway, New Jersey 08854

² Department of Chemistry, Rutgers University, Piscataway, New Jersey 08854

³ Department of Biochemistry, University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey 08854

⁴ Howard Hughes Medical Institute

(RECEIVED March 20, 1996; ACCEPTED May 3, 1996)

Abstract: A C-terminal fragment of the transcription factor OmpR has been crystallized using the sitting drop vapor-diffusion method. Crystals belong to the trigonal spacegroup $P3_12$ with cell dimensions $a = b = 54.4$ Å, $c = 135.5$ Å, and $\gamma = 120.00^\circ$. A second crystal form has been obtained by soaking this crystal form in a cryo-buffer and flash-cooling to 108 K in a cold nitrogen stream. Crystals belong to the trigonal spacegroup $P3_12$ with cell dimensions $a = b = 108.07$ Å, $c = 131.81$ Å, and $\gamma = 120.00^\circ$. Both crystal forms diffract to at least 2.3 Å at a synchrotron light source. Single-site cysteine mutations have been introduced to provide mercury-binding sites for multiple isomorphous replacement.

Keywords: bacterial transcription factor; osmoregulation; response regulator; two-component system

Escherichia coli respond to changes in osmotic pressure by altering the expression of two porin proteins, OmpF and OmpC, located in the outer membrane (Csonka & Hanson, 1991). The differential expression of the *ompF* and *ompC* genes is regulated by two proteins that belong to the ubiquitous two component family of regulatory proteins, the histidine kinase EnvZ and the response regulator OmpR (Forst & Inouye, 1988; Pratt & Silhavy, 1995). Response regulator proteins are characterized by a conserved regulatory domain of approximately 130 amino acids and often contain a C-terminal effector domain. The N-terminal regulatory domain catalyzes its autophosphorylation using the high-energy phosphoryl group from the histidine kinase. The activity of the C-terminal or downstream effector domain is controlled by the phosphorylation state of the regulatory domain (Stock et al., 1990; Bourret et al., 1991; Parkinson & Kofoed, 1992). In many cases, response regulators are transcription fac-

tors with the C-terminal end comprising a DNA recognition domain.

OmpR belongs to a subfamily of at least 40 response regulators with homologous DNA-binding domains of approximately 110 amino acids. The mechanism of DNA interaction and transcriptional activation of the OmpR family has not been elucidated (Maeda & Mizuno, 1990; Huang et al., 1994; Rampersaud et al., 1994). This family of transcription factors does not have sequence homology with DNA-binding proteins of known three-dimensional structure, thus a novel DNA-binding motif might be expected. The structure determination of the C-terminal domain of OmpR, OmpRc, should define a protein fold and provide insight regarding DNA recognition and perhaps the mechanism of transcriptional regulation for a large family of transcription factors. Crystals of the DNA binding domain of OmpR have been previously reported (Kondo et al., 1994). These crystals belong to spacegroup $P3_121$ with cell dimensions $a = b = 60.4$ Å, $c = 58.8$ Å, and $\gamma = 120.00^\circ$ and diffract to 3.0 Å resolution. We have obtained two new crystal forms of an OmpR C-terminal fragment that includes amino acids 130–239. Here we report conditions for high-resolution data collection and a strategy to introduce single-site cysteine mutations for use in heavy-atom derivatization.

Results and discussion: OmpRc, the C-terminal domain of OmpR, was obtained from *E. coli* BL21(DE3) (Studier & Moffatt, 1986) cells containing a T7 expression vector, pJES307 (Tabor & Richardson, 1985), with an insert encoding residues Ala 130–Ala 239 and an initiator methionine. Approximately 6.0 g of cells from 3 L of culture were lysed by sonic disruption, resuspended in 20 mM Tris·HCl, 2 mM 2-mercaptoethanol (β ME), pH 8.6, and membranes were removed by ultracentrifugation. The OmpRc protein was purified using a 2.5×20 cm Q Sepharose Fast Flow (Pharmacia) column equilibrated in 20 mM Tris·HCl, 2 mM β ME, pH 8.6. The protein was eluted with a 400-mL linear gradient of 0–150 mM NaCl in 20 mM

Reprint requests to: Ann M. Stock, CABM/UMDNJ, 679 Hoes Lane, Piscataway, New Jersey 08854; e-mail: stock@mbcl.rutgers.edu.

Tris·HCl, 2 mM β ME, pH 8.6. The fractions containing OmpRc protein were analyzed by SDS-PAGE, pooled, and concentrated by precipitation with 60% saturated ammonium sulfate. The precipitate was collected by centrifugation at $20,000 \times g$ and resuspended in 4 mL 20 mM Tris·HCl, 200 mM KCl, 2 mM β ME, pH 7.5. The filtered sample was applied to a HiLoad 26/60 Superdex G75 pg FPLC column (Pharmacia) equilibrated in 20 mM Tris·HCl, 200 mM KCl, 2 mM β ME, pH 7.5. OmpRc eluted in one peak in a volume consistent with a monomeric species of approximately 12,000 Da. Purity was assessed by Coomassie blue-stained SDS-polyacrylamide gels.

Initial crystallization conditions were found using the sparse matrix approach in a hanging drop vapor-diffusion experiment (Jancarik & Kim, 1991; Cudney et al., 1994). OmpRc is crystallized at 4 °C from a reservoir buffer containing 18% PEG 8000, 400 mM CaCl_2 , 100 mM cacodylate, 2 mM β ME, pH 6.5, using freshly purified protein concentrated with Centricon 10 units (Amicon) to 45–60 mg/mL in 20 mM BisTris·HCl, 100 mM KCl, 2 mM β ME, pH 6.5. Crystals are now grown in a sitting drop vapor-diffusion setup, where 3 μ L of protein solution are mixed with 3–5 μ L of reservoir solution on a watch glass with 1 μ L of reservoir buffer in a 35 $\times$ 10 mm petri dish sealed with parafilm. Crystals appear after 1–2 weeks (Fig. 1A). The divalent metal ion Ca^{2+} can be substituted by Mg^{2+} or Ba^{2+} without significant effect on crystal nucleation, growth, or crystal shape. Crystals grown under these conditions belong to the trigonal spacegroup $P3_112$, with cell dimensions $a = b = 54.4$ Å,

$c = 135.5$ Å, and $\gamma = 120.00^\circ$. The spacegroup was determined from a precession photograph of the $hk1$ zone.

For crystallization by the macroseeding method, a small wild-type crystal is washed in reservoir buffer and transferred, using a 0.3-mm glass capillary, to a sitting drop setup containing 3 μ L of protein solution mixed with 2 μ L of reservoir solution, yielding a final drop volume of approximately 6 μ L. Crystals appear after 1–3 days. Macroseeding yielded crystals reproducibly, but of inferior size and shape, often producing highly aggregated crystals or few well-shaped but small crystals.

While characterizing native crystals and screening for heavy-atom derivatives, we found that cell axis c could vary between 134.5 Å and 138.5 Å for data collected at room temperature, making it difficult to scale data sets from different crystals. Collecting data at cold nitrogen temperature was attempted in order to overcome the crystal to crystal cell variability problem, and also to be able to collect multiple-wavelength (MAD) and single-crystal high-resolution data at a synchrotron light source. Prior to freezing, crystals are first transferred into a stabilizing solution that contains 25% PEG 8000, 400 mM CaCl_2 , 100 mM cacodylate, pH 6.5, then slowly overlaid and soaked with solutions containing stabilizing buffer plus addition of either I, 40% glycerol; or II, 30% ethylene glycol and 10% glycerol. Cryoprotected crystals are equilibrated for at least 30 min in the cryo buffer, mounted on a nylon loop, and frozen directly in a cold nitrogen stream at a temperature of 108 K. Cell axes a and b double upon freezing in the presence of buffer I, generating a

Fig. 1. Crystals from OmpRc wild-type and mutant proteins grown by sitting drop vapor-diffusion as described in the text from 18% PEG 8000, 400 mM CaCl_2 , 100 mM cacodylate, 2 mM β ME, pH 6.5. **A:** Wild-type OmpRc, crystals grow to 0.8-mm length in the largest dimension. **B:** Mutant A139C crystals closely resemble wild-type crystals. **C:** Mutant R150C crystals are smaller and diffract more weakly. **D:** Mutant K144C crystals only grow as crystal aggregates after macroseeding. All images represent equal magnification.

new crystal form that belongs to spacegroup $P3_112$, with cell dimensions $a = b = 108.07$ Å, $c = 131.81$ Å, and $\gamma = 120.00^\circ$. Crystals frozen in the presence of buffer II belong to spacegroup $P3_112$, with cell dimensions $a = b = 53.54$ Å, $c = 131.37$ Å, and $\gamma = 120.00^\circ$ (Table 1).

Image plate (IP) data have been taken from crystals that were mounted on a nylon loop and frozen at 108 K in a cold nitrogen stream. IP data have been collected on an R-Axis IIC detector using mirror-focused and Ni-filtered $\text{CuK}\alpha$ radiation from a Rigaku RU-200 rotating anode source operated at 5 kW. IP data have also been collected at the Brookhaven National Synchrotron Light Source (NSLS) beamline X4A at a wavelength of 0.9793 Å. IP data were integrated using the program DENZO, and SCALEPACK was used to estimate the degree of mosaic spread (Otwinowski, 1993). Data reduction for all data was performed with the programs ROTAVATA/AGROVATA from the CCP4 suite of programs (CCP4, 1994).

Despite of an exhaustive search, no useful heavy-atom derivatives could be found. SeMet substituted protein was produced and MAD data were collected at the Se edge. Although a measurable, wavelength-dependent anomalous signal was observed, it was not possible to locate any one of five seleniums per polypeptide chain. We therefore decided to introduce cysteines into the protein to covalently attach mercury-containing compounds at specific sites for solution of the phase problem.

A number of structure determinations have relied on the introduction of cysteines to obtain useful heavy-atom derivatives, and several criteria have been used to select mutation sites (reviewed in Forest & Schutt, 1992). The success rate for obtaining a useful heavy-atom derivative has been approximately 20%. Rather than introducing cysteines at entirely random positions, but also to perhaps improve the probabilities for obtaining a useful derivative, we chose a different rational approach. The sequences of OmpR and homologous proteins were aligned and homologues were searched for cysteine residues. Amino acid residues at positions in OmpR corresponding to cysteines in other homologues were altered with the expectation that cysteine substitutions at these sites might be tolerated (Fig. 2). Some of the

mutations suggested by this approach were quite extraordinary changes, such as lysine to cysteine or arginine to cysteine substitutions. A total of five mutations were made according to the described rationale. The sequence database used in the alignment has been enlarged since these experiments were designed, and additional OmpR homologues have been found to contain cysteines, increasing the number of suggested mutation sites to 11.

Oligonucleotide-directed mutagenesis was performed using a two-step PCR-based technique (Higuchi et al., 1988). Plasmid pORC DNA was used as template with the 5' primer ATATACAT ATGGCACCGTCACAGGAAG containing an *Nde* I restriction site, and the 3' primer GTTAGCAGCCGGATCCTCATGCTT TAG containing a *Bam* H I restriction site. Internal oligonucleotides contained the following mutagenic triplets flanked on both sides by 10–12 nucleotides corresponding to the anticoding or coding strand sequences of *ompR*, respectively: A139C, TGT, and ACA; K144C, TGC, and GCA; R150C, TGC, and GCA; M152C, TGT, and ACA; and V173C, TGC, and GCA. PCR was performed in a 30-cycle reaction using native *pfu* polymerase (Stratagene) in its supplied buffer. The final products were digested with *Nde* I and *Bam* H I, the 338-bp fragment was isolated by preparative gel electrophoresis, then ligated into *Nde* I and *Bam* H I-digested pJES307 vector DNA using T4 DNA ligase. The resulting plasmids were used to first transform DH5 α cells, and later BL21(DE3) cells that served as the host for high-level expression. Sequences of the entire mutated genes were confirmed by DNA sequencing using the dideoxy chain termination method and Sequenase polymerase (U.S. Biochemicals).

Four of the five mutant proteins could be produced at high levels in the T7 expression system and could be purified under the described conditions. M152C was not expressed at significant levels and therefore was not purified. Of the four mutant proteins that were purified, A139C and R150C yielded, under the described crystallization conditions, large and well-shaped crystals isomorphous to wild-type crystals (Figs. 1B and 1C). One mutant, K144C, produced, with application of macroseeding, multiple crystal aggregates (Fig. 1D). The fourth mutant, V173C, precipitated during crystallization setups and did not

Table 1. Data collection statistics for OmpRc data collected at 108 K on image plate detectors

Source	R-axis IIC ^a	NSLS I ^{a,b,d}	NSLS II ^{c,d}	R-axis IIC ^a
$a = b$ (Å)	53.54	53.54	108.07	53.40
c (Å)	131.37	130.61	131.81	130.89
No. Reflections	4,753	10,678	34,786	3,087
Resolution (Å)	2.75 (2.8–2.75)	2.1 (2.16–2.1)	2.3 (2.36–2.3)	3.2 (3.28–3.20)
R_{merge}^e	0.038 (0.13)	0.054 (0.21)	0.050 (0.27)	0.065 (0.181)
I/Sigma	15.0 (5.2)	7.7 (3.4)	11.4 (2.8)	10.3 (4.2)
Completeness	96.6 (92.2)	93.2 (91.0)	96.7 (90.8)	85.3 (88.3)
Multiplicity	3.4 (2.5)	3.0 (2.3)	3.9 (2.6)	2.7 (2.6)
Protein	Wild type	Wild type	Wild type	A139C
Derivative				HgOAc ₂
R_{deriv}^f				0.170

^a Cryobuffer: 25% PEG 8000, 30% ethylene glycol, 10% glycerol, 0.4 M CaCl_2 , 0.1 M cacodylate, pH 6.5.

^b Partial reflections included in this data set.

^c Cryobuffer: 25% PEG 8000, 40% glycerol, 0.4 M CaCl_2 , 0.1 M cacodylate, pH 6.5.

^d NSLS data collected at a wavelength of 0.9793 Å.

^e $R_{\text{merge}} = \sum (|I_j - \langle I_j \rangle|) / \sum \langle I_j \rangle$, where I_j is the intensity of observed reflection j and $\langle I_j \rangle$ is the mean intensity for reflection j .

^f $R_{\text{deriv}} = \sum \text{abs}(FPH - FP) / \sum (FP)$.

			*	*	*	*	*
OmpR	(130-239)	APSQEEAVIA	FGKFKLNLGT	REMFRE.DEP	MPLTSGEFAV	LKALVSHPRE	
VirG	(151-267)	.RSKDRRSFC	FTDWTNLNRQ	RRLMSEAGGE	VKLTAGFNL	LLAFLEKPRD	
TorR	(127-230)	QPHTQDNLYR	FAGYCLNVSR	HTLERD.GEP	IKLTRAHEYEM	LVAFVTNPGE	
SphR	(157-257)	...ETPAVLR	YEGLKLFPEE	CRVLLD.DRE	LTLSPKEFRL	LELFMRHPRR	
VanR	(126-231)	VKEQNENVIV	HSGLVINVNT	HECYLN.EKQ	LSLTPTEFSI	LRILCENKGN	
OmpR		PLSRDKLMNL	ARGREYSAME	.RSIDVQISR	LRRMVEEDPA	HPRYIQTVWG	LGYVFVPDGS KA
VirG		VLSREQLLIA	SRVRDEEVYD	.RSIDVLILR	LRRKLEADPS	SPQLIKTARG	AGYFFDADVQ VSHGGTMAA
TorR		.LSRERLLRM	LSARRVENPD	LRTVDVLIRR	LRHKLSAD..	...LLVTQHG	EGYFLAADVC
SphR		VWSRDQLLEK	IWGIDFMGDS	.KTIDVHIRW	LREKIEANPS	NPSYLLTVRG	FGYRLG
VanR		VVSSELLFHE	IWGDEYFSKS	NNTITVHIRH	LREKMNDTID	NPKYIKTVWG	VGYKIEK

Fig. 2. Sequence alignment of the C-terminal domain of OmpR and the four homologous response regulators containing cysteines. The alignment was performed with the GCG programs PILEUP and PRETTY (Devereux et al., 1984). Asterisks indicate cysteine mutation sites; amino acid residue numbers are written in parentheses. Sequences and SWISS-PROT accession numbers: *E. coli* OmpR, p03025; *Agrobacterium tumefaciens* VirG, p06664; *E. coli* TorR, p38684; *Synechococcus sphaeroides* SphR, p39663; *Enterococcus faecium* VanR, q06239.

produce crystals under the employed conditions. One heavy-atom derivative has been prepared by soaking the mutant protein A139C with 1 mM HgOAc₂. A difference Patterson map calculated with the structure factors of wild-type and derivatized A139C is consistent with one mercury atom per asymmetric unit (Fig. 3). The heavy-atom binding site is close to a special position, and therefore only two non-origin vectors can be seen in the entire Patterson map.

We have grown three different trigonal crystal forms of OmpRc. One of these (spacegroup $P3_121$) is similar to the crystal form reported by Kondo et al. (1994) that diffracted to 3.0 Å. However, because the specific OmpR fragment used by Kondo et al. is not described, it is not clear that the forms are identical. Crystals from both other trigonal forms (spacegroup $P3_112$) reported here diffract to at least 2.3 Å resolution at a synchrotron light source. Crystal form I has cell dimensions $a = b = 53.54$ Å, $c = 131.37$ Å, and $\gamma = 120.00^\circ$. Assuming two molecules in the asymmetric unit, the Matthews coefficient is $2.3 \text{ Å}^3 \text{ Da}^{-1}$, within a range expected for globular proteins. However, derivative data indicate only one mercury per asymmetric unit. There is either one monomer with high solvent content per asymmetric unit, or a dimer with asymmetry that disrupts mercury derivatization of the cysteine in one monomer. Crystal form II, gen-

erated upon freezing, with cell dimensions $a = b = 108.07$ Å, $c = 131.81$ Å, and $\gamma = 120.00^\circ$, represents an approximate quadrupling of the volume of the asymmetric unit with respect to crystal form I. A self rotation search for crystal form II did not identify noncrystallographic symmetry. A possible explanation is that, when cell axes double, only molecules related by translation are incorporated into the new asymmetric unit.

In situations where conventional isomorphous replacement methods are problematic, both selenomethionine substitution and cysteine mutagenesis for mercury derivatization provide feasible strategies for phase determination by MAD methods. In the case of OmpRc, mercury derivatization of a substituted cysteine has provided useful phase information, whereas selenomethionine incorporation has not. Different characteristics of cysteine and selenomethionine substitutions may provide particular advantages of one approach over another. Methionines are commonly present in greater numbers than cysteines, which may allow for a strong anomalous signal, but can cause difficulties in locating seleniums in proteins with high methionine content or crystals with large asymmetric units. Another consideration is the preparation of the protein itself. Substitution with selenomethionine often significantly lowers levels of protein expression, decreases the solubility of the protein, and increases

Fig. 3. Determination of the Hg position in crystals of OmpRc A139C. The $u, v, 1/3$ Harker section of the native-derivative difference Patterson map calculated between 12 and 5 Å resolution contoured in increments of 0.5 SD, starting at 2 SD, is shown. The Hg position is $x = 0.31, y = 0.64, z = 0.009$. Data were collected from frozen crystals on an R-axisIIc detector.

oxidative damage. On the other hand, cysteine substitution generally requires multiple constructions substituted at different sites, with some having disastrous effects on protein folding or stability and others resulting in imperceptible perturbations. Unfortunately, each protein inevitably poses a unique set of challenges and the considerations presented above may only be apparent in retrospect.

Acknowledgments: We thank Craig Ogata, Monique Floer, Rachel Kramer, and Jaroslav Vojtechovsky for assistance with data collection at NSLS; Jane Lee, Ann West, and Alejo Santa Cruz for assistance with cloning, mutagenesis, and sequencing; and Richard Leidich and Greg Listner for X-ray technical assistance. This work was supported by grants from the NIH (GM47958), the NSF, and the Sinsheimer Foundation to AMS, and by a grant from the NIH (GM21589) to H.M.B. A.M.S. is an assistant investigator of the Howard Hughes Medical Institute.

References

- Bourret RB, Borkovich KA, Simon MI. 1991. Signal transduction pathways involving protein phosphorylation in prokaryotes. *Annu Rev Biochem* 60:401–441.
- CCP4. 1994. Collaborative Computational Project N4. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D* 50:760–763.
- Csonka LN, Hanson AD. 1991. Prokaryotic osmoregulation: Genetics and physiology. *Annu Rev Microbiol* 45:569–606.
- Cudney R, Patel S, Weisgraber K, Newhouse Y, McPherson A. 1994. Screening and optimization strategy for macromolecular crystal growth. *Acta Crystallogr D* 50:414–423.
- Devereux J, Haeblerli P, Smithies O. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res* 12:387–395.
- Forest K, Schutt C. 1992. Protein engineering for structure determination. *Curr Opin Struct Biol* 2:576–581.
- Forst S, Inouye M. 1988. Environmentally regulated gene expression for membrane proteins in *Escherichia coli*. *Annu Rev Cell Biol* 4:21–42.
- Higuchi R, Krummel B, Saiki RK. 1988. A general method of in vitro preparation and specific mutagenesis of DNA fragments: Study of protein and DNA interactions. *Nucleic Acids Res* 16:7351–7367.
- Huang KJ, Schieberl JL, Igo MM. 1994. A distant upstream site involved in the negative regulation of the *Escherichia coli ompF* gene. *J Bacteriol* 176:1309–1315.
- Jancarik J, Kim SH. 1991. Sparse matrix sampling: A screening method for crystallization of proteins. *J Appl Crystallogr* 24:409–411.
- Kondo H, Miyaji TA, Suzuki M, Tate SI, Mizuno T, Nishimura Y, Tanaka I. 1994. Crystallization and X-ray studies of the DNA-binding domain of OmpR protein, a positive regulator involved in activation of osmoregulatory genes in *Escherichia coli*. *J Mol Biol* 235:780–782.
- Maeda S, Mizuno T. 1990. Evidence for multiple OmpR-binding sites in the upstream activation sequence of the *ompC* promoter in *E. coli*: A single OmpR-binding site is capable of activating the promoter. *J Bacteriol* 172:501–503.
- Otwinowski Z. 1993. Oscillation data reduction program. In: Sawyer L, Isaacs N, Bayley S, eds. *Proceedings of the CCP4 study weekend: Data collection and processing*. Warrington, UK: SERC Daresbury Laboratory. pp 56–62.
- Parkinson JS, Kofoed EC. 1992. Communication modules in bacterial signaling proteins. *Annu Rev Genet* 26:71–112.
- Pratt LA, Silhavy TJ. 1995. Porin regulon of *Escherichia coli*. In: Hoch JA, Silhavy TJ, eds. *Two-component signal transduction*. Washington, DC: American Society for Microbiology. pp 105–127.
- Rampersaud A, Harlocker SL, Inouye M. 1994. The OmpR protein of *Escherichia coli* binds to sites in the *ompF* promoter region in a hierarchical manner determined by its degree of phosphorylation. *J Biol Chem* 269:12559–12566.
- Stock JB, Stock AM, Mottonen JM. 1990. Signal transduction in bacteria. *Nature* 344:395–400.
- Studier FW, Moffatt BA. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* 189:113–130.
- Tabor S, Richardson CC. 1985. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proc Natl Acad Sci USA* 84:1074–1078.