The mechanism of a formaldehyde-sensing transcriptional regulator

Katie J. Denby1,*, Jeffrey Iwig2,†, Claudine Bisson1,*, Jodie Westwood1, Matthew D. Rolfe1, Svetlana E. Sedelnikova1, Khadine Higgins3,‡, Michael J. Maroney3, Patrick J. Baker3, Peter T. Chivers3,4 & Jeffrey Green1

Most organisms are exposed to the genotoxic chemical formaldehyde, either from endogenous or environmental sources. Therefore, biology has evolved systems to perceive and detoxify formaldehyde. The frmRA(B) operon that is present in many bacteria represents one such system. The FrmR protein is a transcriptional repressor that is specifically inactivated in the presence of formaldehyde, permitting expression of the formaldehyde detoxification machinery (FrmA and FrmB, when the latter is present). The X-ray structure of the formaldehyde-treated Escherichia coli FrmR (EcFrmR) protein reveals the formation of methylene bridges that link adjacent Pro2 and Cys35 residues in the EcFrmR tetramer. Methylen bridge formation has profound effects on the pattern of surface charge of EcFrmR and combined with biochemical/biophysical data suggests a mechanistic model for formaldehyde-sensing and derepression of frmRA(B) expression in numerous bacterial species.

All organisms must adapt to withstand the detrimental effects of toxic chemicals. Some of these toxic compounds, such as reactive oxygen and nitrogen species are generated endogenously, as a result of metabolic processes; others are natural or synthetic products that are present in the environment. Many of these chemicals have broad reactivity, which presents biological systems with a challenge to specifically sense and then respond to their presence before the cell suffers irreversible damage. Formaldehyde is a toxic electrophilic chemical that is generated endogenously by many organisms1,2. For example, formaldehyde is generated: (i) as an intermediate in methylotrophic metabolism; (ii) in the degradation of glycine, either by the glycolytic byproduct methylglyoxal or by Fenton chemistry; (iii) in the degradation of heme, during iron acquisition by some Gram-positive bacteria; (iv) by lipid peroxidation of sugars; (v) by the demethylation of histones; (vi) as a product of methylated-DNA repair by AlkB; and (vii) by the action of N-methyltryptophan oxidase (SolA)3–4. Consequently, biological systems are often exposed to both endogenous and exogenous sources of formaldehyde.

Formaldehyde mediates its toxic effects by chemically modifying vital cell components, including DNA and proteins, thereby leading to cellular dysfunction. Formaldehyde-mediated genotoxicity is caused by the formation DNA-DNA and DNA-protein cross-links, as well as covalent DNA monoadducts1,5–7. In addition, formaldehyde is able to covalently modify proteins, inhibiting their functions7,8. The life-threatening damage caused by the broad chemical reactivity of formaldehyde has driven the evolution of mechanisms to detoxify formaldehyde and counteract its detrimental effects2. In order to regulate expression of these detoxification systems, it is imperative to maintain specific response-regulators in the cell that can perceive the presence of formaldehyde and induce protective systems before significant damage to vital cell components occurs.

The starting point for the work reported here was the observation that formaldehyde is generated when the model bacterium Escherichia coli adapts to the presence of the alternative electron acceptor trimethylamine-N-oxide6. An inability to respond (by induction of the frmRAB operon) to this endogenous formaldehyde challenge resulted in growth inhibition, rather than growth promotion, when anaerobic E. coli cultures were provided with trimethylamine-N-oxide6. The frmRAB operon codes for: a regulator, FrmR (EcFrmR); a formaldehyde dehydrogenase, FrmA; and an S-formylglutathione hydrolase, FrmB9. In

1Department of Molecular Biology and Biotechnology, University of Sheffield, Sheffield, S10 2TN, UK. 2Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine in St. Louis, St. Louis, MO, 63110, USA. 3Department of Chemistry, University of Massachusetts-Amherst, Amherst, MA 01003, USA. 4Departments of Biosciences and Chemistry, Durham University, Durham, DH1 3LE, UK. 5Present address: Carmot Therapeutics, Inc. San Francisco, CA 94158, USA. 6Present address: Department of Chemistry, Salve Regina University, Newport, Rhode Island 02840, USA. 7These authors contributed equally to this work. Correspondence and requests for materials should be addressed to P.J.B. (email: p.baker@sheffield.ac.uk) or P.T.C. (email: peter.chivers@durham.ac.uk) or J.G. (email: jeff.green@sheffield.ac.uk)
RepoRts

RcnR family of transcriptional repressors

13, 14. Most of the characterized members of this family are metal-ion binding in other CsoR/RcnR family members. W-X-Y-Z fingerprint (blue font) are indicated. Position W is shown to incorporate both Pro2 and Ser/His3 residues, Pro2 (brown background), Cys35 and Cys70 (yellow background) are indicated. Residues of the residue, which is hereafter regarded as spatially part of position W, for metal-coordination 17, 21. CsoR binds Cu(I) and InrS use all four positions to coordinate their cognate metal, and in addition RcnR uses the addition to the absolute conservation at position X, there is a highly conserved His residue at position Y (Fig. 1a).

The crystal structure of a metal-sensing W and Cys35 at position X; Fig. 1)23. The crystal structure of a metal-sensing protein reveals strong conservation in the first 63 amino acids (67% identical, 89% similar) but weaker conservation thereafter (21% identical; 39% similar; Fig. 1b). There is limited conservation between the two proteins within the W-X-Y-Z fingerprint (blue font) are indicated. Position W is shown to incorporate both Pro2 and Ser/His3 (as indicated by parentheses; see text for details). Residues on blue backgrounds have been implicated in DNA-binding in other CsoR/RcnR family members.

E. coli, and many other organisms, formaldehyde in the cytosol reacts with the major reductant glutathione, yielding S-hydroxymethylglutathione, which is then oxidized by FrmA to S-formylglutathione 11, 12. In E. coli S-formylglutathione is hydrolyzed to formate and glutathione by the product of the final gene of the frmRAB operon, FrmB, or by a second, less-efficient hydrolase YetG; some bacteria, such as the closely related Salmonella enterica serovar Typhimurium, lack frmB but possess yetG. Thus, the proteins coded by the frmRAB operon and yetG permit bacteria to sense and detoxify formaldehyde.

The first gene of the E. coli frmRAB operon is the regulatory protein EcFrmR. EcFrmR is a member of the CsoR/RcnR family of transcriptional repressors 13, 14. Most of the characterized members of this family are metal-ion sensors; the properties of Cu(I) (CsoR, RicR), Ni(II) (InrS) and Ni(II)/Co(II) (RcnR, DmeR) sensors in diverse bacterial species have been reported, as well as those of the sulfite/sulfide sensor CstR 15-19. Amino acid sequence alignments of the metal-sensing members of the family revealed characteristic amino acid signatures (known as the W-X-Y-Z fingerprint) that are associated with metal-binding (Fig. 1a). The availability of structure-function data now allows the W-X-Y-Z fingerprint to be considered spatially, centered about the absolutely conserved Cys residue (position X, Fig. 1a; residue 35 in RcnR and FrmR), rather than as a primary sequence motif 17, 20. In addition to the absolute consensus at position X, there is a highly conserved His residue at position Y (Fig. 1a). RcnR and InrS use all four positions to coordinate their cognate metal, and in addition RcnR uses the N-terminal residue, which is hereafter regarded as spatially part of position W, for metal-coordination 21, 22. CsoR binds Cu(I) via the X-Y-Z positions (Fig. 1a). Amino acid residue variation at different positions within the fingerprint is implicated in signal specificity by coupling metal-coordination preferences to the allosteric network connecting the metal- and DNA-binding residues of the proteins 23. Whilst the Ni(II)/Co(II) and Cu(I)-responsive family members are relatively well-characterized, little is known about signal perception and DNA-binding mechanisms of the remaining proteins (CstR and FrmR), although the Cys residues at positions X and Z in S. aureus CstR, and Pro2 and Cys35 (positions W and X) of Salmonella enterica serovar Typhimurium FrmR (hereafter StyFrmR) have been implicated in sulfide and formaldehyde sensing, respectively (Fig. 1a) 19, 22, 23.

StyFrmR retains 3 of the metal-binding ligands (His3, Cys35 and His60, with Glu instead of His64) that are present in the paralogous Salmonella Ni(II)-binding regulator RcnR (Fig. 1a) 22. Recent in vitro studies showed that StyFrmR binds Co(II), Cu(I) and Zn(II), but the binding affinities were weaker than those of the biological sensors of these metal ions in Salmonella, such that StyFrmR was outcompeted by CaeR, RcnR, ZntR and ZurR for their cognate metals 22. Thus in vivo, StyFrmR was able to act as a formaldehyde-responsive transcription regulator, but was unable to act as a metal ion sensor 22. Further work showed that formaldehyde-sensing by StyFrmR is specific, direct and requires two residues of the W-X-Y-Z fingerprint (the N-terminal Pro at position W and Cys35 at position X; Fig. 1) 23. The crystal structure of a metal-sensing StyFrmR variant (StyFrmR(E64H)), created by site-directed mutagenesis, revealed that it consists of a homotetrameric disc with a surface-exposed positively-charged region that is predicted to participate in DNA-binding 23. Hence, the fold of StyFrmR resembles other members of the CsoR/RcnR family 20, 22.

Amino acid sequence alignment of the E. coli K-12 FrmR (EcFrmR) and StyFrmR proteins reveals strong conservation in the first 63 amino acids (67% identical, 89% similar) but weaker conservation thereafter (21% identical; 39% similar; Fig. 1b). There is limited conservation between the two proteins within the W-X-Y-Z fingerprint. Both have Cys and His at positions X and Y, as well as Pro at residue 2 (position W). However, EcFrmR differs with Ser not His at residue 3 (a component of position W; note that the equivalent of His3 is involved in metal-binding in RcnR and InrS proteins) and Thr not Glu at position Z. Notably, EcFrmR possesses an additional non-conserved Cys residue (Cys70) (Fig. 1b). The differences between StyFrmR and EcFrmR along with the
biological imperative to mount an effective response to endogenous sources of formaldehyde, as evidenced by the observation that the **frmRAB** operon was essential for adaptation of *E. coli* to growth on trimethylamine-\(N\)-oxide, prompted an investigation of the *Ec* FrmR protein. This is worthy of investigation because, although the activities of detoxifying enzymes, such as FrmA and FrmB, have been established, the mechanism(s) used by regulatory proteins to perceive and respond to formaldehyde are poorly understood. Hence, the aim of the work reported here was to provide new insight into the mechanism used by biological systems to sense the presence of the highly reactive and toxic chemical, formaldehyde.

**Results and Discussion**

**EcFrmR is a formaldehyde sensor.** Herring and Blattner showed that EcFrmR is necessary for repression of the **frmRAB** promoter (**P**\(_{\text{frm}}\)) in *E. coli*, and that the operon is induced in the presence of formaldehyde. This is worthy of investigation because, although the activities of detoxifying enzymes, such as FrmA and FrmB, have been established, the mechanism(s) used by regulatory proteins to perceive and respond to formaldehyde are poorly understood. Hence, the aim of the work reported here was to provide new insight into the mechanism used by biological systems to sense the presence of the highly reactive and toxic chemical, formaldehyde. *In vivo* and *in vitro* data show that EcFrmR senses formaldehyde directly, with no metal-dependence, via the formation of inter-subunit methylene bridges between adjacent Pro2 and Cys35 residues. This formaldehyde-mediated cross-linking remodels the surface of the tetrameric *Ec* FrmR disc resulting in de-repression of **frmRAB** expression by promoting disassociation of the **P**\(_{\text{frm}}\)-**Ec**FrmR complex.

**Figure 2.** EcFrmR-mediated repression of **frmRAB** expression is relieved by formaldehyde. (a) Cultures of *E. coli* PC677 carrying **P**\(_{\text{frm}}\)-**frmR-lacZ** (open bars) or **P**\(_{\text{frm}}\)-**frmRstop-lacZ** (gray bars) were grown as described in Methods in the absence and presence of the indicated concentrations of formaldehyde. (b) *β*-Galactosidase activities (Miller units plotted on a log scale) were measured as a proxy for *in vivo* transcription from the **frmRAB** promoter. Activities were normalized to that measured in the absence of formaldehyde. For both panels, the error bars represent the standard deviation from the mean (n = 3).

**Results and Discussion**

**EcFrmR is a formaldehyde sensor.** Herring and Blattner showed that EcFrmR is necessary for repression of the **frmRAB** promoter (**P**\(_{\text{frm}}\)) in *E. coli*, and that the operon is induced in the presence of formaldehyde. This is worthy of investigation because, although the activities of detoxifying enzymes, such as FrmA and FrmB, have been established, the mechanism(s) used by regulatory proteins to perceive and respond to formaldehyde are poorly understood. Hence, the aim of the work reported here was to provide new insight into the mechanism used by biological systems to sense the presence of the highly reactive and toxic chemical, formaldehyde. *In vivo* and *in vitro* data show that EcFrmR senses formaldehyde directly, with no metal-dependence, via the formation of inter-subunit methylene bridges between adjacent Pro2 and Cys35 residues. This formaldehyde-mediated cross-linking remodels the surface of the tetrameric EcFrmR disc resulting in de-repression of **frmRAB** expression by promoting disassociation of the **P**\(_{\text{frm}}\)-EcFrmR complex.

**Results and Discussion**

**EcFrmR is a formaldehyde sensor.** Herring and Blattner showed that EcFrmR is necessary for repression of the **frmRAB** promoter (**P**\(_{\text{frm}}\)) in *E. coli*, and that the operon is induced in the presence of formaldehyde. This is worthy of investigation because, although the activities of detoxifying enzymes, such as FrmA and FrmB, have been established, the mechanism(s) used by regulatory proteins to perceive and respond to formaldehyde are poorly understood. Hence, the aim of the work reported here was to provide new insight into the mechanism used by biological systems to sense the presence of the highly reactive and toxic chemical, formaldehyde. *In vivo* and *in vitro* data show that EcFrmR senses formaldehyde directly, with no metal-dependence, via the formation of inter-subunit methylene bridges between adjacent Pro2 and Cys35 residues. This formaldehyde-mediated cross-linking remodels the surface of the tetrameric EcFrmR disc resulting in de-repression of **frmRAB** expression by promoting disassociation of the **P**\(_{\text{frm}}\)-EcFrmR complex.

**Results and Discussion**

**EcFrmR is a formaldehyde sensor.** Herring and Blattner showed that EcFrmR is necessary for repression of the **frmRAB** promoter (**P**\(_{\text{frm}}\)) in *E. coli*, and that the operon is induced in the presence of formaldehyde. This is worthy of investigation because, although the activities of detoxifying enzymes, such as FrmA and FrmB, have been established, the mechanism(s) used by regulatory proteins to perceive and respond to formaldehyde are poorly understood. Hence, the aim of the work reported here was to provide new insight into the mechanism used by biological systems to sense the presence of the highly reactive and toxic chemical, formaldehyde. *In vivo* and *in vitro* data show that EcFrmR senses formaldehyde directly, with no metal-dependence, via the formation of inter-subunit methylene bridges between adjacent Pro2 and Cys35 residues. This formaldehyde-mediated cross-linking remodels the surface of the tetrameric EcFrmR disc resulting in de-repression of **frmRAB** expression by promoting disassociation of the **P**\(_{\text{frm}}\)-EcFrmR complex.

**Results and Discussion**

**EcFrmR is a formaldehyde sensor.** Herring and Blattner showed that EcFrmR is necessary for repression of the **frmRAB** promoter (**P**\(_{\text{frm}}\)) in *E. coli*, and that the operon is induced in the presence of formaldehyde. This is worthy of investigation because, although the activities of detoxifying enzymes, such as FrmA and FrmB, have been established, the mechanism(s) used by regulatory proteins to perceive and respond to formaldehyde are poorly understood. Hence, the aim of the work reported here was to provide new insight into the mechanism used by biological systems to sense the presence of the highly reactive and toxic chemical, formaldehyde. *In vivo* and *in vitro* data show that EcFrmR senses formaldehyde directly, with no metal-dependence, via the formation of inter-subunit methylene bridges between adjacent Pro2 and Cys35 residues. This formaldehyde-mediated cross-linking remodels the surface of the tetrameric EcFrmR disc resulting in de-repression of **frmRAB** expression by promoting disassociation of the **P**\(_{\text{frm}}\)-EcFrmR complex.
be directly or indirectly perceived by cytosolic FrmR\textsuperscript{26-28}. Formaldehyde induced the highest β-galactosidase activity, but acetaldehyde, methylglyoxal and glyoxal also induced expression, albeit to lesser extents (Fig. 2b). This contrasts with StyFrmR which did not respond to acetaldehyde (in the same type of experiment), suggesting sequence differences between the two proteins around the sensory site could affect selectivity. Bulky aldehydes, such as furaldehyde and tribromoacetaldehyde, were unable enhance lacZ expression. Overall, these responses suggest that, although induction of frmRAB is not absolutely specific, formaldehyde is by far the most effective inducer, consistent with the specificity of the detoxification machinery (FrmRA and FrmB/YelG). However, it is possible that induction results from EcFrmR responding to the formation of S-hydroxymethylglutathione in the cytoplasm rather than formaldehyde per se.

**EcFrmR binds directly to the frmRAB promoter (P\textsubscript{frm}) to repress transcription.** The effect of signal molecules on DNA-binding affinity and transcriptional regulation provides a sensitive measure of allosteric effectiveness. Bio-Layer Interferometry (BLItz) measurements were used to investigate interactions between the EcFrmR protein and frmRAB promoter DNA (P\textsubscript{frm}). Under these conditions a K\textsubscript{d} for EcFrmR binding to immobilized P\textsubscript{frm} DNA was ~220 nM, determined from an overall on-rate constant (k\textsubscript{on}) of ~13000 M\textsuperscript{-1} s\textsuperscript{-1} and an overall off-rate constant (k\textsubscript{off}) of ~0.003 s\textsuperscript{-1} at 20 °C (Fig. 3a; Table 1). This K\textsubscript{d} value is similar to that reported for StyFrmR (~100 nM)\textsuperscript{23}, but should be considered as an upper limit because the potential effects of low level metal contamination and protein oxidation (see below) on DNA-binding are unknown. Nevertheless, this interaction was specific because the K\textsubscript{d} for EcFrmR binding at an unrelated promoter DNA fragment (E. coli ydhY) was only ~3600 nM (Table 1). Pre-treatment of EcFrmR protein with excess formaldehyde for 3 min essentially abolished specific binding to P\textsubscript{frm} (Fig. 3a). Exposure of the pre-formed P\textsubscript{frm}-EcFrmR binary complex to increasing concentrations of formaldehyde resulted in disassociation of the complex (Fig. 3b). The P\textsubscript{frm}-EcFrmR disassociation curves were fitted to a single exponential function. Disassociation of the P\textsubscript{frm}-EcFrmR-complex exhibited a linear dependence on formaldehyde concentration, with a rate constant of ~4 M\textsuperscript{-1} s\textsuperscript{-1} at 20 °C (Fig. 3c; Table 1). Accordingly, *in vitro* transcription reactions showed that synthesis of the frmRAB transcript was inhibited in the presence of EcFrmR and that this inhibition was relieved when EcFrmR was treated with formaldehyde (Fig. 3d). EcFrmR did not affect transcription from the E. coli ndh promoter, indicating that the effects of EcFrmR and formaldehyde on frmRAB transcription were specific (Fig. 3d). Therefore, the *in vivo* and *in vitro* data showed that EcFrmR is a repressor of frmRAB expression that responds directly to formaldehyde by disassociation of the P\textsubscript{frm}-EcFrmR complex, allowing the expression of the formaldehyde detoxification system.

**Properties of isolated EcFrmR.** Liquid chromatography-mass spectrometry (LC-MS) showed that EcFrmR lacked an N-terminal methionine (confirmed by N-terminal amino acid sequencing) and had the expected monomeric molecular mass of 10186.50 Da. EcFrmR eluted from a calibrated size exclusion chromatography column at a volume indicative of a tetramer and this assignment was confirmed by analytical ultracentrifugation, which yielded a mass of 44.9 kDa (Fig. S1a). Thus, like other members of the CsoR/RcnR family, EcFrmR is a homotetramer. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry column at a volume indicative of a tetramer and this assignment was confirmed by analytical ultracentrifugation. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively coupled plasma mass spectrometry. The isolated protein was metal-free as judged by inductively couple
and would be antagonistic to the deployment of the detoxification machinery. Therefore, it was concluded that formaldehyde-sensing and DNA-binding are not metal-(Zn(II)-) dependent and, based on the affinity of EcFrmR (and StyFrmR) for Zn(II), it is likely that Zn(II) could only inhibit FrmR activity in vivo under conditions when Zn(II) homeostasis is severely perturbed (Fig. S2; Table 1)22.

Table 1. Rate constants for EcFrmR DNA interactions. *The BLItz data used to calculate the kinetic parameters shown are provided in Table S4. The data were fitted to a 1:1 binding model to derive $k_f$, $k_r$ and $K_d$ values using all the sample data simultaneously (Global fitting). **Value and standard error.

---

**Figure 3. Formaldehyde enhances disassociation of the Pfrm−EcFrmR complex.** (a) Bio-Layer Interferometry (BLItz) assays. Reactions to evaluate the interaction of biotin-labeled Pfrm DNA, immobilized on a streptavidin probe, with EcFrmR were carried out with 10 different concentrations of EcFrmR (Table S4A). Representative traces for EcFrmR (6.16 μM tetramer; black line, 0.88 μM tetramer; red line), as well as EcFrmR pre-treated with 200-fold molar excess of formaldehyde (0.88 μM tetramer; blue line), and EcFrmR binding at a non-target DNA (PpydhY, 0.88 μM EcFrmR tetramer; green line) are shown. (b) Pre-formed Pfrm−EcFrmR complexes were exposed to 10 different concentrations (Table S4C) of formaldehyde and disassociation curves were recorded. Traces for 0 (black); 0.05 mM (orange); 0.25 mM (gray); 0.62 mM (yellow); 1.25 mM (blue); 3.69 mM (green); 4.92 mM (dark blue); 7.38 mM (brown) are shown. (c) Single exponential fits to formaldehyde disassociation curves were used to obtain the observed rate constants ($k_{obs}$) which were plotted against formaldehyde concentration to obtain the apparent second order rate constant. (d) Inhibition of frmRAB transcription by EcFrmR in vitro is relieved by formaldehyde. Reaction conditions are described in the Methods section. Left panel, Pfrm; right panel, Pndh. Lanes 1, RNA size markers, top to bottom: 600, 500, 400, 300, 200, 100 bases; Lanes 2, no EcFrmR; lanes 3, 1 nM EcFrmR tetramer; lane 4, 1 nM EcFrmR tetramer pre-treated with 200-molar excess formaldehyde. The locations of the frmR and ndh are indicated.
Identification of EcFrmR residues necessary for formaldehyde sensing. Site-directed mutagenesis of the amino acids of the W-X-Y-Z fingerprint has revealed the importance of these residues for the function of CsoR/RcnR family proteins (Fig. 1a). Recently, Pro2 (position W) and Cys35 (position X) of StyFrmR have been shown to be required for formaldehyde-responsiveness in vivo and in vitro, whereas Glu64 (position Z) was not.

Therefore, to identify EcFrmR residues necessary for formaldehyde sensing, site-directed mutation of the Pfrm-frmR-lacZ construct (see above) was undertaken. β-Galactosidase activity measurements showed that, like StyFrmR, EcFrmR(P2A) and EcFrmR(C35A) failed to respond to formaldehyde, confirming that Pro2 and Cys35 (W and X positions in the CsoR/RcnR family fingerprint; Fig. 1a) are essential for perception of formaldehyde (Fig. 4a). Replacement of His60 (position Y) resulted in high basal activity that was further enhanced in the presence of formaldehyde, whereas replacement of Thr64 (position Z) had no effect on the function of EcFrmR (Fig. 4a). Additionally, by analogy to RcnR metal-sensing residues 17, insertion of an Ala codon before Pro2 (EcFrmR(A2*)) also resulted in formaldehyde insensitivity, albeit with higher basal Pfrm activity, implicating the Pro2 imino group in formaldehyde sensing (Fig. 4a). Replacement of the only other cysteine residue (Cys70) in EcFrmR did not impair the response to formaldehyde (Fig. 4a). The conclusion that Pro2 and Cys35 are required for EcFrmR to respond to formaldehyde was supported by the formaldehyde-sensitivity of E. coli strains expressing EcFrmR(P2A) and EcFrmR(C35A) in place of EcFrmR. Cultures expressing these variants exhibited enhanced sensitivity to formaldehyde, consistent with the low frmR expression observed in the reporter fusion experiments, presumably arising from constitutive repression frmR expression (Fig. 4a and b; Table 2).

Reaction of EcFrmR with formaldehyde. The modification of EcFrmR by formaldehyde, and the competition with Zn(II), was analyzed by LC-MS. Without formaldehyde, the major species corresponded to the EcFrmR monomer lacking the N-terminal methionine (10186.60 Da; predicted mass 10186.50 Da) with lower amounts of a disulfide-linked dimer (20371.02 Da; predicted unmodified dimer mass 20373.20 Da) (Table 2). After exposure to formaldehyde (8-fold molar excess) for 3 min the EcFrmR monomer was still detected.
The formaldehyde-exposed FrmR tetramer is asymmetric. One face of the tetramer (A/B face) is comprised from a platform of five parallel helices with a hole at the centre (Fig. 5b). This arrangement differs from that seen in CsoR and Sty

α

in CsoR and

Sty

from a platform of five parallel helices with a hole at the centre (Fig. 5b). This arrangement differs from that seen

α

FrmR and

Sty

Ec

(10187.02 Da) along with new species of molecular mass 20396.87 Da, corresponding to an FrmR dimer plus an additional mass of 23.67 Da, and 40768.73 Da, corresponding to a tetramer with an additional mass of 22.30 Da (Table 2; Fig. S3). The former modified species represents FrmR dimers linked by two intermolecular methylene bridges (−CH₂−) (net mass gain of 2 × 12 Da per FrmR dimer); the latter modified species can be accounted for by an FrmR tetramer in which each subunit participates in only one methylene bridge (total of two in the tetramer) and these cross-linked dimers are held together by one disulfide bond (Table 2; Fig. S3). Incubation with stoichiometric Zn(II) (4 Zn(II) per FrmR tetramer) before or after formaldehyde treatment, yielded dimeric species of molecular masses 20396.99 Da and 20396.95 Da, respectively, indicating that Zn(II) ions did not prevent formaldehyde-dependent methylene bridge formation under these conditions.

Formaldehyde modification of FrmR(P2A) and FrmR(C35A) was examined because they failed to respond to formaldehyde in vivo (Fig. 4). Neither mutation affected the oligomeric state of the protein, as both variants eluted as tetramers upon size exclusion chromatography. LC-MS showed the presence of disulfide-linked FrmR(P2A) dimers (20319.69 Da; predicted unmodified mass 20320.8 Da) in the absence or presence of formaldehyde (Table 2). However, exposure to formaldehyde without quenching permitted the detection of a monomeric FrmR(P2A) species (10190.71 Da) with additional mass 29.91 Da, equivalent to an hydroxymethylated FrmR(P2A) monomer (Table 2; Fig. S3). The former modified species represents FrmR dimers linked by two intermolecular methylene bridges (−CH₂−) (net mass gain of 2 × 12 Da per FrmR dimer); the latter modified species can be accounted for by an FrmR tetramer in which each subunit participates in only one methylene bridge (total of two in the tetramer) and these cross-linked dimers are held together by one disulfide bond (Table 2; Fig. S3). Incubation with stoichiometric Zn(II) (4 Zn(II) per FrmR tetramer), before or after formaldehyde treatment, yielded dimeric species of molecular masses 20396.99 Da and 20396.95 Da, respectively, indicating that Zn(II) ions did not prevent formaldehyde-dependent methylene bridge formation under these conditions.

The formaldehyde-modified FrmR was determined to a resolution of 2.7 Å (protein databank identifier PDB: 5LCY) where the equivalent α helix is domain swapped onto the opposite 1-

α

residues 10–63 (Fig. 1b).

Table 2. Liquid chromatography mass spectrometry analyses of FrmR proteins treated with formaldehyde. aDifference between predicted mass and measured mass. bInterpretation of mass difference; disulfide bond, −2.2 Da; methylene bridge, net mass gain 12 Da; hydroxymethylation, net mass gain 30 Da. cExposure to formaldehyde without quenching.

| Protein sample                      | Measured mass (Da) | Relative abundance (a. u.) | Mass difference (Da) | Comment                                      |
|-------------------------------------|--------------------|----------------------------|----------------------|----------------------------------------------|
| EcFrmR                              | 10186.60           | 2.0 × 10⁷                   | 0.1                  | FrmR monomer                                 |
|                                     | 20371.02           | 4.4 × 10⁶                   | −2.18                | FrmR disulfide-linked dimer                   |
|                                     | 10187.02           | 1.8 × 10⁶                   | 0.52                 | FrmR monomer                                 |
| HCHO-treated EcFrmR                 | 20396.87           | 2.0 × 10⁵                   | 23.67                | FrmR dimer with 2 methylene bridges           |
|                                     | 40768.73           | 2.0 × 10⁵                   | 22.30                | Two FrmR dimers, each with 1 methylene bridge, linked to form a tetramer by 1 disulfide bond |
| HCHO-treated EcFrmR plus Zn(II)     | 10187.01           | 1.8 × 10⁶                   | 0.51                 | FrmR monomer                                 |
|                                     | 20396.95           | 3.0 × 10⁵                   | 23.75                | FrmR dimer with 2 methylene bridges           |
| Zn(II)-treated EcFrmR plus HCHO     | 10187.04           | 1.8 × 10⁶                   | 0.54                 | FrmR monomer                                 |
|                                     | 20396.99           | 4.0 × 10⁵                   | 23.79                | FrmR dimer with 2 methylene bridges           |
| HCHO-treated EcFrmR(P2A)            | 20319.69           | 4.0 × 10⁵                   | −1.11                | FrmR(P2A) disulfide-linked dimer              |
| HCHO-treated EcFrmR(P2A)            | 10190.71           | 6.8 × 10⁵                   | 29.91                | Hydroxymethylated FrmR(P2A) monomer           |
| HCHO-treated EcFrmR(C35A)           | 20307.40           | 2.8 × 10⁵                   | −1.40                | FrmR(C35A) disulfide-linked dimer             |
| HCHO-treated EcFrmR(C35A)           | 20307.21           | 1.7 × 10⁵                   | −1.59                | FrmR(C35A) disulfide-linked dimer             |

The structural response of FrmR to formaldehyde modification and a mechanism for formaldehyde sensing. The crystal structure of formaldehyde-exposed FrmR was determined to a resolution of 2.7 Å (protein databank identifier PDB: 5LB). Like other members of the CsoR/RcnR family, FrmR oligomerizes to form a disc-like tetramer, constructed from two homodimers, each of which forms one face of the disc. Each subunit consists of three helices (α1, residues 2–30; α2, residues 35–68; and α3, residues 73–91), linked by two short loops (L1, residues 31–34; and L2, residues 69–72), that are arranged as a flattened S-shape (Fig. 5a). The α3 helices slot together at the homodimer interface, such that each face of the tetramer is formed from a platform of five parallel helices with a hole at the centre (Fig. 5b). This arrangement differs from that seen in CsoR and StyR (PDB: 5CY) where the equivalent α3 helix is domain swapped onto the opposite face of the tetramer (Fig. S4). This difference between FrmR and StyR (PDB: 5E4H) is striking especially as the conserved regions of the α1 and α2 helices of these proteins superpose well, superposition of the Cα atoms of residues 10–36 (α1-α2) of the uncross-linked or cross-linked chains of FrmR with the uncross-linked StyR yielded root mean square deviation (RMSD) values of ~1.0 Å for both. However, superposition of the Cα atoms from the full length chains (either cross-linked or uncross-linked) gave RMSD values of ~5 Å, reflecting the different organisation of α3 within the tetramer, which could be a consequence of the differences in amino acid sequence in the region spanning the terminus of α2, L2 and the beginning of α3 (Fig. 1b).

The formaldehyde-exposed FrmR tetramer is asymmetric. One face of the tetramer (A/B face) is comprised of an unmodified homodimer with electron density visible for residues Lys9–Lys91 of both polypeptides with a disordered N-terminal region (residues Pro2–Lys8). Electron density corresponding to Zn(II) or other metal ions was not detected; however the W–X–Y–Z fingerprint residues in the unmodified homo-dimer are located in...
similar positions in 3D space to those of the CsoR proteins from *Geobacillus thermodenitrificans* and *Thermus thermophilus* and some of these residues are likely to constitute the EcFrmR Zn(II) binding site (Fig. S5)\(^{24,25}\). The XAS data supports the assignment of Cys35 and His60 as Zn(II) ligands (Fig. S1d). The residue at position Zn(Thr54) does not interact with other residues of the W-X-Y-Z motif in either form of the EcFrmR dimer, consistent with the lack of effect of the T64A mutation on formaldehyde-dependent de-repression of *P*\(_{\text{frm}}\) (Fig. 4a).

The homodimer forming the other face of the tetramer (A′/B′ face) is sandwiched against the first via a network of hydrophobic packing interactions between the secondary structural elements. In contrast to the A/B face, clear electron density is present for residues Pro2-Lys8 of both subunits on the A′/B′ face, resulting in an ordered extension to the N-terminal region of α1. The Pro2′-N atom is located within ~2.5 Å of the Cys35-5G atom of the corresponding subunit on the opposite face of the tetramer (Fig. 5b). There is extra density between these two atomic positions, indicating the presence of the formaldehyde cross-link, and a methylene bridge (−CH\(_2\)-), as indicated by the LC-MS data, has been modeled at this location (Fig. 5c).

The crystal structure of EcFrmR reveals both the free and signal-triggered states of the protein. One face of the EcFrmR tetramer represents the conformation of the protein in the absence of formaldehyde (A/B face with a disordered Pro2), whilst the other face represents both the conformation that has responded to formaldehyde and has formed the methylene bridge (A′/B′ face with an ordered Pro2) (Fig. 5b). ‘On’- and ‘off’-states of an RcnR/CsoR family member from a single source have not been observed previously and the structure presented here facilitates a detailed analysis of the conformational changes that occur upon reaction of EcFrmR with formaldehyde to inhibit DNA-binding. Unlike EcFrmR, the N-terminal region of StyFrmR(E64H) is visible in the absence of the methylene bridge\(^{21}\). This showed that the StyFrmR(E64H) N-terminal Pro2′ is located close to Cys35 (~3 Å between Pro2′-N and Cys35-5G atoms), poised to facilitate formaldehyde-specific cross-linking. With the exception of the domain (A′) face of α2, the N-term superposes well on the StyFrmR structure (RMSD ~1.0 Å). This suggests that a remarkably subtle change in the locations of Pro2′ and Cys35 and the constraints imposed by the covalent methylene bridge cross-link contribute to creating a distinct difference in the relative conformation and orientation of the subunits in the cross-linked face compared to those of the uncross-linked face. While the position of α3 is similar on both faces of the tetramer, there is a translational movement of α1 and α2 on the A′/B′ face, which slide across the equivalent helices on the opposite face by ~1.5 α-helical turns in response to formation of the methylene bridge. In addition, the C-terminal half of α2 twists and buckles towards α3, which alters the packing of these two helices, generating a 10° off-angle between α3 and α2. This movement changes the overall size and shape of the A′/B′ face, which expands and elongates by ~10 Å on the diagonal in response to formaldehyde (Fig. 5b; Animation S1). Notably, expansion of the envelope of CsoR in response to Cu(I) was apparent in small angle X-ray scattering studies\(^{22}\). Furthermore, in addition to an H-bond network, involving His60-Tyr34-Glu80 that is thought to be involved in signal (Cu(I)) perception but not signal binding by *M. tuberculosis* and *G. thermodenitrificans* CsoR proteins, the N-terminal region of *G. thermodenitrificans* CsoR becomes ordered over the Cu(I)-binding site\(^{21}\). The stable ordering of the N-terminal regions of EcFrmR and CsoR proteins upon signal perception establishes new interactions between the α1 and α2 helices of these proteins (Fig. S6). Thus, signal perception by EcFrmR results in a H-bond interaction between Glu7 (α1) and Ser39 (α2), which is not present in the uncross-linked form of EcFrmR, StyFrmR(E64H) structure or in the Cu(I)-free form of *S. lividans* CsoR. The same two residue positions of Cu(I)-loaded *G. thermodenitrificans* CsoR (Glu22-Arg74) participate in an electrostatic interaction. Moreover, an equivalent interaction occurs in *M. tuberculosis* CsoR, but involving a slightly different residue positions in α2 (Lys8-Glu63). Thus, it is suggested that N-terminal ordering and the establishment of interactions between α1 and α2 could be a common feature of signal perception and transduction in the CsoR/RcnR family that has only been revealed because both ‘on’ and ‘off’ states are captured in the EcFrmR structure.

The different conformational states of the two faces of formaldehyde-treated EcFrmR have a dramatic effect on the pattern of surface charge and the position of protrusions on each face of the tetramer (Fig. 5b). There are two patches of positive charge either side of the central hole on both faces of the tetramer, which contain several residues implicated in DNA-binding (Arg14, Arg16, Gln41, Arg46 and Lys91)\(^{24,30,31}\). The differences in size and shape of the two faces of EcFrmR in the crystal structure mean that these patches are separated by ~35 Å on the A/B face but by ~45 Å on the A′/B′ face. Whilst an asymmetric tetramer has been captured in the crystal structure, modeling suggests that both faces of EcFrmR could adopt the compact DNA-binding conformation in the absence of formaldehyde-induced cross-linking. However, models of a fully cross-linked tetramer, in which subunits A and A′ and B and B′ were both linked by two methylene bridges, as implied by the LC-MS data, contained many inter-subunit clashes, indicating that both faces of EcFrmR might not be able to adopt the extended conformation whilst retaining the tetrameric state. The presence of only two methylene bridges in the structure of the EcFrmR tetramer raises the possibility that, at least under some conditions, the formation of the third and fourth cross-links might be subject to negative cooperativity, as observed for binding the third and fourth Zn(II) ions to EcFrmR (see above) and that the fully cross-linked protein might dissociate into dimers incapable of DNA-binding. The negative cooperativity is also consistent with the effects of the H60A mutation on *P*\(_{\text{frm}}\)-lacZ activity (Fig. 4a). This mutant shows high constitutive promoter activity, similar to the stop codon mutant (Fig. 2a). However, the maximum induction by formaldehyde is substantially lower. This contrasts with the ability of Ni(II) and Co(II) repress *P*\(_{\text{frm}}\) to the same extent as a stop codon mutant of RcnR\(^{16}\). The His60 residue of EcFrmR is within H-bond distance of Cys35 in the unmodified dimer (3.2 Å), and makes no obvious interactions with side chain or main chain residues in the methylene-brided structure. Hence, the H60A mutation may mimic the modified form of the protein at all four sites, instead of two, resulting in greater de-repression.

It has been suggested that CsoR/RcnR family members recognize their DNA target by a combination of: (i) shape selectivity, resulting from the propensity of the central GC tract to adopt the A-DNA form; (ii) specific interactions with the flanking inverted repeats; and (iii) non-specific binding to distant DNA that might result in DNA-wrapping\(^{20,31}\). Operator sequences for CsoR/RcnR family members have been classified into two groups:
Figure 5. Structure of EcFrmR. (a) Cartoon representations of uncross-linked (left) and cross-linked (right) EcFrmR monomers colored blue (N-terminal) to red (C-terminal). Secondary structure elements (α-helices, α1 to α3; loops, L1 and L2) are labeled and the amino acid residues (single letter code, P2, C35 and C70) involved in cross-linking and disulfide bond formation are shown as sticks. The disordered N-terminal region in the uncross-linked subunit is represented by the blue dashed line. (b) A comparison of the overall size and shape of the uncross-linked (left) and cross-linked (right) faces of the EcFrmR tetramer. The upper images show the arrangement of the helices on each face of the tetramer, the positions of the methylene bridges (P2'-C35) and the Cys70-Cys70' disulfide bonds (S-S). The homodimer (A/B) on the uncross-linked face is drawn in shades of green and the cross-linked face (A'/B') in shades of orange. The middle images show the expansion of the surface envelope upon cross-linking (black double headed arrow drawn between Arg14 Cα atoms, highlighted in pink). The lower images show the surface-charge on either side of the tetramer (red represents negative charge, blue positive charge and white neutral). (c) Section of the 2Fo-Fc map between chains A and A' obtained when the coordinates for Pro2 and the methylene bridge were omitted from the refinement (black mesh, contoured at 1σ). Residues are indicated by their single letter codes.
type I sites consist of GC tracts (3–8 bp) flanked by AT-rich inverted repeats; type II sites have shorter interrupted GC tracts\(^{30,31}\). The *E. coli* *P*\(_{frm}\) contains a large inverted repeat centered at \(-29\) relative to the predicted transcriptional start site (Fig. 6a). This region contains tandem type I FrmR-binding sites consisting of 9 bp GC-rich tracts flanked by ATAC/GTAT inverted repeats (Fig. 6a). The *E. coli* *rcnR-A* intergenic region also possesses tandem type I RcnR-binding sites containing a TACTGGGGGGNAGTA motif, which imparts some A-form DNA character on this region of DNA, and one RcnR tetramer binds at each site on the same face of the DNA helix\(^{31}\). The recently reported structure of the site-directed mutant *Sty* FrmR(E64H), which responds to Co(II), Zn(II) and formaldehyde in vivo showed the presence of positively-charged surface patches for protein that had not been exposed to formaldehyde\(^{22,23}\). As noted above for the uncross-linked surface of *Ec* FrmR (Fig. 5b), the *Sty* FrmR(E64H) positively-charged patches were also separated by \(-35\) Å (distance between the Arg14 C\(\alpha\) atoms)\(^{23}\). These data were used to inform models of the *P*\(_{frm}\)-*EcFrmR* complexes with A- and B-form DNA (Fig. 6). Models with both A- and B-form DNA suggested that residues (Lys10, Arg14, Arg16 and Arg17) forming the two positively-charged protrusions on the A/B face could interact with the major grooves of the DNA, with Lys91 from the central hole packing into the minor groove, but the better fit is with B-form DNA. It is noted that in the bacterial cell *P*\(_{frm}\) DNA is unlikely to be exclusively in A- or B-form, but more likely a hybrid that will be further distorted upon binding of the FrmR tetramer. Nevertheless, the \(-45\) Å that separates the positively-charged patches on the cross-linked A'/B' face precludes convincing interactions with either A- or B-form DNA. These differences suggest a mechanism for de-repression of the *frmRAB* promoter in response to formaldehyde. In the DNA bound state, Lys91 is held in a surface-exposed position by an ion pair interaction between Arg14 and the C-terminal carboxyl. Upon cross-linking, \(\alpha1\) and \(\alpha2\) are pulled away from the centre of the face of the tetramer, with loop 2 acting as a pivot point. This motion pulls Arg14 away from Lys91, breaking the interaction with the C-terminal carboxyl, which causes it to rotate by \(-90\)° into the central hole, forming a new interaction with the guanidyl group of Arg46. The surface-exposed side-chain of Lys91 moves into the hole, burying the NZ atom so that it is no longer available to interact with DNA (Fig. S7). In addition, the motion of \(\alpha1\) and \(\alpha2\) increases the distance between the cluster

Figure 6. **Modeling the P\(_{frm}\)-EcFrmR complex.** (a) The DNA sequence of the *frmRAB* promoter region (P\(_{frm}\)) contains tandem *EcFrmR* binding sites consisting of ATAC/GTAT inverted repeats (bold) separated by G/C-rich tracts (italic) that form a larger inverted repeat (convergent red arrows). The size of *EcFrmR* (subunits colored in shades of green and orange as in Fig. 5b) suggests that two tetramers could bind to the *frmRAB* promoter region. One *EcFrmR* tetramer (side view) is shown on the top face of the DNA sequence and the other (top view) behind the DNA sequence, offset by approximately a quarter turn relative to the first tetramer. (b) Models of binary complexes formed from *EcFrmR* and A- or B-form DNA. One of the tandem *EcFrmR* binding sites of P\(_{frm}\) (dark gray) is modeled as A- (left) and B-form (right) DNA. *EcFrmR* is shown as surface representation with subunits colored in shades of green (uncross-linked A/B face) and orange (cross-linked A'/B' face), with the amino acid side-chains on the A/B face that are implicated in DNA-binding highlighted in blue.
of residues within the positively-charged protrusions by ~10 Å, such as Lys10, Arg14, Arg16 and Arg17 can no longer interact with the major groove, thus breaking the complex between EcFrmR and DNA. The centers of the tandem binding sites at P_frm are separated by 31 bp compared to 19 bp for the rcnR-Å intergenic region, raising the possibility that, unlike RcnR, both faces of a single EcFrmR tetramer could participate in binding to the tandem sites at P_frm.

Concluding remarks. EcFrmR is the first example of a CsoR/RcnR family protein where the asymmetry of the tetramer in the crystal structure reveals the conformational changes induced by signal perception that lead to de-repression of target promoters. In vivo and in vitro experimental evidence show that the EcFrmR specifically reacts with the toxic chemical formaldehyde, resulting in the formation of inter-molecular methylene bridges between adjacent Pro2 and Cys35 residues. In the absence of formaldehyde the StyFrmR(E64H) structure indicates that the N-terminal regions are ordered, such that the N-terminal Pro residues are in close proximity to Cys35 of an adjacent subunit23. The conservation of His60 and its proximity to Cys35 in the adjacent subunit suggests that it could act to abstract a proton from the thiol group to facilitate the initial reaction with formaldehyde forming an S-hydroxymethyl adduct (LC-MS data Table 2). The initial hydroxymethylated Cys35 residues appear to be associated with one surface of the EcFrmR disc permitting nucleophilic attack by Pro2 residues associated with the opposite face of the EcFrmR disc resulting in methylene bridge formation (detected as the oxidized EcFrmR tetramer with additional mass 22.3 Da in LC-MS; Table 2, Fig. S3); an example of an N-terminal Pro residue acting as a nucleophile has been reported for the DNA-repair enzyme, formamidopyrimidine-DNA glycosylase2. The covalent capture of the Pro2 residues of one face of EcFrmR could contribute to the disordering of the uncross-linked N-terminal regions, which is apparent in the crystal structure reported here (Fig. S3). Alternatively, fully uncross-linked EcFrmR could differ in the degree of flexibility in their N-terminal regions despite the high level of sequence conservation (Fig. 1b). The N-terminal disordering captured in the formaldehyde-treated EcFrmR crystal structure could account for the negative cooperativity observed for Zn(II) binding and perhaps in formaldehyde reactivity. Formaldehyde-induced methylene bridge formation (two per tetramer) expands the A'/B' surface of EcFrmR resulting in a similar change in size to that observed for G. thermodenitrificans CsoR upon Cu(I) binding24. Interestingly, CsoR does not use direct coordination of the signal molecule via the N-terminus, unlike FrmR and RcnR. An allosteric network has been proposed for CsoR that involves an interaction between His60 and two other residues (Tyr34 and Glu80, numbered based on the N-terminus of FrmR and RcnR)33. However, these interactions are not present in FrmR and RcnR, suggesting that the allosteric network may initiate at the N-terminus upon signal perception. This notion is consistent with the observation that the G. thermodenitrificans CsoR structure shows the N-terminus capping the Cu(I)-binding site, but not directly coordinating the Cu(I) atom24. Thus, despite specific differences between RcnR, FrmR, and CsoR in the connectivity between signal perception residue(s) and the common DNA-binding interface of the proteins, an ordering of the N-terminal region and α1 may be a common feature of signal perception in this protein family. The inability of Zn(II) to drive EcFrmR off DNA is likely to be because Zn(II) binding cannot order the N-terminal region of α1. Ultimately, all four EcFrmR N-terminal Pro2 residues can be methylene bridged to their partner Cys35 residues resulting in derepression of frmRA expression (detected as EcFrmR dimers + 24 Da in LC-MS; Table 2; Fig. S3).

Finally, it is notable that the second EcFrmR Cys residue (Cys70), which is not conserved in StyFrmR, can form a disulfide bond linking loops 2 of the A/B’ and the B/A’ chains. The effect of this disulfide on the ability of all four EcFrmR subunits to undergo formaldehyde modification to fine-tune the EcFrmR response has not been assessed here, but the potential for EcFrmR to act as a redox sensor, through the formation of a disulfide bond, is of interest because formaldehyde exposure is associated with glutathione-depletion and oxidative stress in higher organisms34–38. Work is on-going to determine the possible role of glutathione (S-hydroxymethylglutathione is formed in E. coli exposed to formaldehyde) and oxidative stress (disulfide bond formation) in regulating EcFrmR activity in vivo as well as to determine the structure of the P_frm-EcFrmR complex.

Methods

Bacterial strains, plasmids and oligonucleotides. These are listed in Supplementary Table S1. The P_frm-frmA-lacZ plasmid was constructed from pPC163, which contains the promoters and lacZ ligated into pACYC18427. A fragment consisting of 499 bp of DNA located upstream of frmA, which included frmR, was ligated into the Eagl and SalI sites of pPC163, replacing the SalI cassette. For overproduction of EcFrmR, the frmR gene was amplified from E. coli K12 genomic DNA with primers J1174 and J1175, and ligated into pET22b at the Ndel and NcoI sites. Mutations in frmR were introduced by overlap PCR or by the Quikchange protocol (Stratagene) using the appropriate oligonucleotide primers (Table S1). Formaldehyde-sensitivities of FrmR and selenomethionine (Se-Met) incorporated EcFrmR protein. For isolation of EcFrmR and the P2A and C35A variants, cultures of the E. coli expression strains (JRG6782, 6783 and 6784) were grown at 37 °C in LB medium containing ampicillin (100 μg ml−1), to an OD of ~0.6, at which point 1 mM IPTG was added and the cultures were incubated for a further 3 h. To obtain selenomethionine (Se-Met) incorporated EcFrmR, cultures were grown to OD ~0.6 in LB, then pelleted and resuspended into Se-Met over-expression medium (60 mM K2HPO4, 8 mM (NH4)2SO4, 33 mM KH2PO4, 2 mM tri-sodium citrate, 54 mM glycerol, 4 mM adenosine, 2 mM guanosine, 4 mM thymine, 4 mM uracil, 4 mM MgSO4, 12 mM thiamine, L-lysine (100 mg l−1),
L-phenylalanine (100 mg·l⁻¹), L-threonine (100 mg·l⁻¹) L-isoleucine (50 mg·l⁻¹). L-leucine (50 mg·l⁻¹), L-valine (50 mg·l⁻¹) and seleno L-methionine (40 mg·l⁻¹) before further incubation for 1 h at 37°C prior to induction of EcFrmR protein expression with IPTG. Cells were lysed after re-suspension in Buffer A (50 mM Tris, 0.1 M NaCl; pH 8.0) by sonication (Soniprep 150 ultrasonic disintegrator) at ~16 microns for 2 cycles of 20 s. The lysate was cleared by centrifugation (10 min, 7000 g) and the resulting cell-free extract was applied to a Heparin-HP column (GE Healthcare) and EcFrmR was eluted using a NaCl gradient (0 to 0.5 M in 50 ml) in Buffer A. Fractions containing EcFrmR protein were combined and dialyzed overnight against 50 mM Tris-HCl buffer, pH 8.0 at 4°C. Micro-crystals of EcFrmR were collected by centrifugation and dissolved in 1 ml of 1 M NaCl in buffer A (see above). The sample was then applied to a HiLoad Superdex 200 column (GE Healthcare) equilibrated with 50 mM Tris–HCl, 0.5 M NaCl (pH 8.0). EcFrmR-containing fractions were pooled and the purity of the samples was assessed by SDS-polyacrylamide gel electrophoresis. For crystallization, protein was concentrated to 10 mg ml⁻¹ using a VivaSpin device with MWCO 30000 (GE Healthcare). The authenticity of EcFrmR was confirmed by total amino acid analysis (Alta Biosciences; University of Birmingham, UK) allowing calibration of EcFrmR protein concentrations measured by either the Bio–Rad protein reagent or using the calculated molar extinction coefficient for EcFrmR²⁹,⁴⁰.

Zn(II) binding Assays. Zn(II) titrations, XAS and EXAFS analyses are described in Supplementary material.

Analytical ultracentrifugation. Details are provided in the supplementary material.

Estimation of the number of reactive thiols in EcFrmR. Reactive protein thiols were estimated by the method of Thelander⁴¹. In a Quartz cuvette, a total reaction volume of 1 ml was prepared consisting of purified protein (100 μl) diluted with protein elution buffer (50 mM Tris, 0.5 M NaCl (pH 8.0) and 200 μl DTNB solution (0.2 ml 0.4% DTNB in ethanol and 1.8 ml 1 M Tris–HCl, pH 8.0). Absorbance at 412 nm was measured and the 2-nitrothiobenzoate concentration was used to calculate the number of reactive sulfydryl groups.

Mass spectrometry. Protein samples (~12 μM) for LC-MS were incubated with either an 8-fold or 40-fold molar excess of formaldehyde to EcFrmR tetramer concentration at room temperature. After 3 min, reactions were quenched with 10 mM glycine. Samples were loaded onto an Agilent 1260 Infinity liquid chromatograph fitted with an Agilent Extended C18 column (2.1 mm × 50 mm) and eluted with a gradient of 5-95% acetonitrile in 0.1% formic acid at 400 μl min⁻¹ over 8 min. The eluate was directly coupled to an Agilent 6530 Q-TOF mass spectrometer fitted with an electrospray ionisation (ESI) source for determination of the masses of species detected in the chromatograph.

For ICP-MS, EcFrmR (200 μM) in 50 mM Tris (pH 8.0) buffered containing 0.5 M NaCl or buffer was incubated with concentrated nitric acid (1:1 ratio) at 60°C for 1 h. Samples were cooled, diluted with DH₂O (up to 10 ml) and filtered before analysis on a Perkin Elmer NexIon ICP-MS system. Ions were quantified using a dilution series of certified multi-element reference standard (Sigma-Aldrich). Counts per second values for elements in the buffer and protein samples were then compared with the calibration curve to determine actual concentrations.

Bio-layer interferometry (BLItz). Biotinylated-promoter DNA for frm (Pfrm) and ydhY (PydhY) was amplified from E. coli MG1655 genomic DNA by PCR using appropriate oligonucleotide primers (Table S1). Where indicated, EcFrmR tetramers were incubated with formaldehyde (up to an 800-fold molar excess) for 3 min (before quenching with 10 mM glycine) or Zn(II) (16-fold molar excess) at room temperature, prior to measurements. Analysis of the interaction between purified EcFrmR and biotinylated-promoter DNA was carried out using the BLItz system (FortéBio), at 20°C. Streptavidin (SA) biosensors (FortéBio) were hydrated by soaking the tips in 250 μl protein elution buffer (50 mM Tris–HCl, pH 8.0 containing 0.5 M NaCl) for at least 10 min. The measurements were made using the Advanced Kinetics function of the BLItz Pro 1.1.0.31 software (FortéBio). The sequence for each run was as follows: (1) baseline step with protein elution buffer for 60 s, (2) loading step with 100 nM biotinylated DNA in buffer for 240 s, (3) baseline step with buffer for 60 s, (4) association step with various concentrations of purified EcFrmR in buffer for 240 s, (5) dissociation step with buffer (or formaldehyde-containing buffer) for 120 s. Baseline and dissociation steps were carried out by placing the biosensor in a black 500 μl Eppendorf tube filled with 250 μl buffer. Loading and association steps were carried out by placing the biosensor in the drop holder containing either 5 μl of P frm DNA or EcFrmR protein depending on the step. For each run a new biosensor was used. The BLItz Pro 1.1.0.31 software using global fitting and correcting both association and dissociation curves was used for analysis.

β-Galactosidase reporter experiments. Starter cultures of E. coli PC677 (ΔfrmR) containing the appropriate reporter plasmid (wild type or mutant) were grown aerobically in LB medium with chloramphenicol (0.1 mM) to an OD₆₀₀ ~0.5. These cultures were then used to inoculate 1.87 ml of LB medium with 0.1 mM chloramphenicol in capped microcentrifuge tubes in triplelicate with aldehyde compounds as indicated. Cultures were grown overnight (12–15 h) before measurement of β-galactosidase activity⁴². The maximal aldehyde concentrations that inhibited growth by <15% were: 700 μM formaldehyde, 7 mM acetaldehyde, 100 μM chloroacetaldehyde, 11 μM tribromoacetaldehyde, 4 mM propionaldehyde, 10 mM fururaldehyde, 3.5 mM glyoxal, 350 μM methylglyoxal and 500 μM glutaraldehyde.

In vitro transcription reactions. The promoter and part of the coding regions of frmR and ndh (~200 bp upstream of the start codon to ~190 bp (frmR) or ~100 bp (ndh) into the gene) were amplified from E. coli MG1655 genomic DNA using appropriate oligonucleotides (Table S1). These DNA fragments (~0.1 pmol) were incubated for 30 min at 37°C in a 10.5 μl reaction volume containing 40 mM Tris-HCl, pH 8.0, 10 mM MgCl₂,
1 mM dithiothreitol, 75 mM KCl, 0.1 mM EDTA, 5% glycerol, 250 μg/ml bovine serum albumin, 20 units of RiboLoc RNase inhibitor (Fermentas), 1 pmol E. coli RNA polymerase holoenzyme (New England Biolabs, Inc.) and 0 or 1 nM EcFrmR tetramer. EcFrmR was reduced with 10 mM dithiothreitol and when required treated with 200-fold molar excess of formaldehyde for 5 min at room temperature, before quenching with 10 mM glycine. Transcription was initiated by the addition of 2 μl solution containing UTP at 50 μM; ATP, CTP and GTP at 1 mM; and 2.5 μCi of [α-32P]UTP (800 Ci mmol−1 PerkinElmer Life Sciences), followed by incubation for 15 min at 37 °C. Reactions were terminated by the addition of 12.5 μl Stop/Loading dye solution (95% formamide, 20 mM EDTA, pH 8, 0.05% bromophenol blue, 0.05% xylene cyanol). Samples (10 μl) of each reaction were loaded onto a 6% acrylamide, 1x TBE, 8 M urea gel and analysed using a phosphorimagery (Typhoon 7000; GE Healthcare). Markers (0.1–1 kb) were prepared using Perfect RNA Marker template mix (Novagen). A 20 μl reaction containing 0.75 μg of RNA template mix, 80 mM HEPES, pH 7.5, 12 mM MgCl2, 10 mM NaCl, 10 mM dithiothreitol, 2 mM ATP, 2 mM GTP, 2 mM CTP, 0.1 mM UTP, 5 μCi of [α-32P]UTP (800 Ci mmol−1 PerkinElmer Life Sciences), 20 units of RiboLoc RNase inhibitor (Fermentas) and 50 units of T7 RNA polymerase (Novagen), was incubated for 1 h at 37 °C, before storing at −20 °C. Markers from ~20 ng template were used for gel calibration.

Crystallization and structural determination of EcFrmR. Crystals of SeMet-labeled EcFrmR (in 50 mM Hepes pH 7.5 and 0.5 mM NaCl) were grown in 0.2 M MgCl2, 0.1 M Na cacodylate pH 6.5 and 31% PEG 2000. Crystals were harvested and cryoprotected in their mother liquor with an additional 25% ethylene glycol. Crystallization and data collection were performed at beamline i03 at the Diamond Light Source (Table S3). The data were auto-processed using Xia2 (XDS/Aimless). Data reduction and cell refinement were performed using Space21 (CCP4i). Crystals were indexed in space group P3112 and AutoSharp was used to locate eight selenium sites (four full occupancy and four half occupancy). Positional and anisotropic displacement parameters were refined to an Rmerge of 0.37. Reactions were carried out to produce a poly-alanine model. The six subunits in the asymmetric unit comprised three half tetramers that were related by translational NCS, with poor electron density for one dimer. Subsequent higher resolution data (2.7 Å) were collected on beamline i03 (100 K) from a second SeMet-labeled crystal (in 200-fold molar excess of formaldehyde for 3 h before cryo-cooling. These data were in spacegroup P3 1, with a tetramer of EcFrmR tetramer. Related by translational NCS, with poor electron density for one dimer. Subsequent higher resolution data (2.7 Å) were collected from a second SeMet-labeled crystal (in 200-fold molar excess of formaldehyde for 3 h before cryo-cooling. These data were in spacegroup P3 1 with a tetramer of EcFrmR in the asymmetric unit (Table S3). The poly-alanine model was used in molecular replacement48 as a starting point to build and refine the full structure of EcFrmR using COOT45, ArpWarp49 and Refmac550. The final structure was refined to Rnative/Rfree values of 0.21 and 0.27, respectively, and was validated using Molprobity50. Structure superpositions were made using the SuperPose software51.

References
1. Yu, R. et al. Formation, accumulation, and hydrolysis of endogenous and exogenous formaldehyde-induced DNA damage. Toxicol. Sci. 116, 170–182 (2015).
2. Chen, N. H., Djoko, K. Y., Veyrier, F. J. & McEwan, A. G. Formaldehyde stress responses in bacterial pathogens. Front. Microbiol. 7, 257 (2016).
3. Trewick, S. C., Henshaw, T. F., Hausinger, R. P., Lindahl, T. & Sedgwick, B. Oxidative demethylation by Escherichia coli AlkB directly reverts DNA base damage. Nature 419, 174–178 (2002).
4. Koyama, Y. & Ohmori, H. Nucleotide sequence of the Escherichia coli alkB gene encoding a sarcosine oxidase-like protein and characterization of its product. Gene 181, 179–183 (1996).
5. Lai, Y. et al. Measurement of endogenous versus exogenous formaldehyde-induced DNA-protein crosslinks in animal tissues by stable isotope labeling and ultrasonic mass spectrometry. Cancer Res. 76, 2652–2661 (2016).
6. Kuykendall, J. R. & Bogdanoff, M. S. Efficiency of DNA-histone crosslinking induced by saturated and unsaturated aldehydes in vitro. Mutat. Res. 283, 131–136 (1992).
7. Xie, M. Z. et al. Aldehydes with high and low toxicities inactivate cells by damaging distinct cellular targets. Mutat. Res. 786, 41–51 (2016).
8. Metz, B. et al. Identification of formaldehyde-induced modifications in proteins: reactions with model peptides. J. Biol. Chem. 279, 6235–6243 (2004).
9. Denby, K. J. et al. Adaptation of anaerobic cultures of Escherichia coli K-12 to response to environmental trimethylamine-N-oxide. Environ. Microbiol. 17, 2477–2491 (2015).
10. Herrling, C. D. & Blattner, F. R. Global transcriptional effects of a suppressor tRNA and the inactivation of the regulator frmR. J. Bacteriol. 186, 6714–6720 (2004).
11. Gonzalez, C. F. et al. Molecular basis of formaldehyde detoxification. Characterization of two S-formylglutathione hydrolases from Escherichia coli, FrmB and YeiG. J. Biol. Chem. 281, 14514–14522 (2006).
12. Harms, N., Ras, J., Reijnders, W. N., van Spanning, R. J. & Stouthamer, A. H. S-formylglutathione hydrolase of Paracoccus denitrificans is homologous to D: a universal pathway for formaldehyde detoxification? J. Bacteriol. 178, 6296–6299 (1996).
13. Giedroc, D. P. & Arunkumar, A. I. Metal sensor proteins: nature’s metalloregulated allosteric switches. Dalton Trans. 7, 3107–3120 (2007).
14. Higgins, K. A. & Giedroc, D. Insights into protein allostery in the CsoR/RcnR family of transcriptional repressors. Chem. Lett. 43, 20–25 (2014).
15. Liu, T. et al. CooR is a novel Mycobacterium tuberculosis copper-sensing transcriptional regulator. Nat. Chem. Biol. 3, 60–68 (2007).
16. Iwig, J. S., Rowe, J. L. & Chivers, P. T. Nickel homeostasis in Escherichia coli: the rcnR-rcnA efflux pathway and its linkage to NikR function. Mol. Microbiol. 62, 252–262 (2006).
17. Iwig, J. S., Leitch, S., Herbst, R. W., Maroney, M. J. & Chivers, P. T. Nif(A) and Co(A) sensing by Escherichia coli RcnR. J. Am. Chem. Soc. 130, 7592–7606 (2008).
18. Foster, A. W., Pernul, R., Patterson, C. J. & Robinson, N. J. Metal specificity of cyanobacterial nickel-responsive repressor InrS: cells maintain zinc and copper below the detection threshold for InrS. Mol. Microbiol. 92, 797–812 (2014).
19. Grossoehme, N. et al. Control of copper resistance and inorganic sulfur metabolism by paralogous regulators in Staphylococcus aureus. J. Biol. Chem. 286, 13522–13531 (2011).
20. Ma, Z., Jacobsen, F. E. & Giedroc, D. P. Coordination chemistry of bacterial metal transport and sensing. Chem. Rev. 109, 4644–4681 (2009).
Structural data have been deposited in with the PDB; accession code 5LBM.

Acknowledgements
This work was supported by the Biotechnology and Biological Sciences Research Council UK through the White Rose-BBSRC Mechanistic Biology Doctoral Training Programme (BB/J014443/1) (K.J.D) and project grant BB/L008114/1 (J.G) and the National Science Foundation (USA) grant MCB0520877 (P.T.C). We thank Mr. Simon Thornpe (The University of Sheffield Faculty of Science Mass Spectrometry Centre) for LC-MS analysis; Dr. Jamie L008114/1 (J.G) and the National Science Foundation (USA) grant MCB0520877 (PTC). We thank Mr. Simon Thornpe and Dr. Deenah Osman for their insightful equipment; and Professor Nigel Robinson and Dr. Deenah Osman (University of Durham) for their insightful input and critical comments during the course of this research project.

Author Contributions
P.J.B., P.T.C and J.G. conceived the study and supervised the experiments. K.J.D., J.I., M.D.R., S.E.S. and J.W. carried out the molecular genetics, purified and determined the biochemical properties of EcFrmR proteins. M.J.M. and K.H. were responsible for the XAS experimental design, data collection, and analysis. C.B. and P.J.B. solved the structure of EcFrmR; all the authors contributed to analyzing data and writing the manuscript.

Additional Information
Accession code: Structural data have been deposited in with the PDB; accession code 5LBM.

Supplementary information accompanies this paper at http://www.nature.com/srep
Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Denby, K. J. et al. The mechanism of a formaldehyde-sensing transcriptional regulator. Sci. Rep. 6, 38879; doi: 10.1038/srep38879 (2016).

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

© The Author(s) 2016