PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

STRUCTURE NOTE

High resolution crystal structure of Sco5413, a widespread actinomycete MarR family transcriptional regulator of unknown function

Tracey A. Holley,[1] Clare E. M. Stevenson,[1] Mervyn J. Bibb,[2] and David M. Lawson[1]*

[1] Department of Biological Chemistry, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, United Kingdom

[2] Department of Molecular Microbiology, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, United Kingdom

ABSTRACT

The crystal structure of Sco5413 from *Streptomyces coelicolor* A3(2) has been determined at 1.25 Å resolution, the highest resolution reported for a MarR family transcriptional regulator. Putative orthologs are encoded by the majority of sequenced actinomycete genomes, and may play roles in regulating growth and antibiotic production, but they have yet to be assigned a precise function. Sco5413 forms a homodimer and, through comparisons with other MarR family protein structures, we postulate that it adopts a conformation compatible with DNA binding, and that a channel at the dimer interface, lined by well-conserved residues, is the binding site of an unidentified effector ligand.

Proteins 2013; 81:176–182.

Key words: protein–DNA interactions; protein–ligand interactions; winged helix fold; antibiotic biosynthesis; *Streptomyces coelicolor.*

INTRODUCTION

MarR family transcriptional regulators (MFRs) are one-component response regulators that have input and output domains fused together in a single polypeptide chain. All MFRs characterized to date function as homodimers and they control gene expression by interacting with pseudopalindromic DNA sequences through conserved winged-helix DNA-binding motifs. The majority act as repressors, and their DNA binding affinities are allosterically modulated either through the binding of small molecule ligands or by sensing reactive oxygen species via cysteine residues.[1] These cytoplasmic proteins are widespread in bacteria and archaea and control a range of biological functions, including resistance to antibiotics,[2] oxidative stress response,[3–5] and pathogenesis[6], therefore they are of significant clinical interest. Moreover, the prevalence of MarR proteins in the streptomycetes, which are the source of numerous therapeutic agents, raises the possibility of exploiting them to manipulate the natural product profiles of these important bacteria.

The *Streptomyces coelicolor* A3(2) genome encodes 42 MFRs, the majority of which are uncharacterized. Of

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

Grant sponsors: Biotechnology and Biological Sciences Research Council (a studentship to TAH, and grant BB/J004561/1 to the John Innes Centre), John Innes Foundation.

*Correspondence to: David M. Lawson, Department of Biological Chemistry, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK. E-mail: david.lawson@jic.ac.uk

Received 7 September 2012; Accepted 2 October 2012
Published online 8 October 2012 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.24197

these, Sco5413 (UniProtKB entry Q9L2A7) was selected for study because it is one of the most highly represented MFRs in sequenced actinobacterial genomes: reciprocal BLAST (http://web.expasy.org/blast) searches predicted that it has apparent orthologs in roughly 60% of these (68 out of 112 genomes as of July 2011; data not shown). None of these orthologs have been characterized and the genomic context of the *sco5413* gene offers no functional clues since it is flanked on both sides by hypothetical genes (see the *Streptomyces* Annotation Server: http://strepdb.streptomyces.org.uk). However, a *sco5413* deletion mutant displayed a growth medium-dependent phenotype that differed from the wild-type, producing pigmented antibiotic prematurely and ectopically, and yielding smaller colonies (Holley *et al.*, unpublished observations). These observations indicate a key role for Sco5413 in regulating actinomycete physiology, displaying pleiotropic effects on both growth and antibiotic production.

We are taking an integrated approach to elucidate the precise biological function of Sco5143. Thus far, we have determined the crystal structure of a truncated version of the Sco5413 protein at 1.25 Å resolution, which is the highest resolution MFR structure currently in the Protein Data Bank (PDB; http://www.rcsb.org/pdb), and represents an important first step toward fully understanding Sco5143. In particular, the structure reveals a homodimer in a conformation that appears to be pre-configured for DNA binding, and a distinct channel lined by well-conserved residues that likely represents a binding site for ligands that allosterically modulate the affinity of Sco5143 for cognate DNA sequences.

METHODS

Protein expression and purification

sco5413 of *S. coelicolor* encodes a 169-residue protein with a molecular mass of 18,413.2 Da. Initial attempts at crystallizing full-length Sco5413 did not yield promising results. Bioinformatic analysis using the DISOPRED2 server (http://bioinf.cs.ucl.ac.uk/disopred) predicted disorder at the N-terminus. We therefore truncated the protein at its N-terminus in an attempt to enhance crystallization. The new translational start site for *sco5413* was chosen to correspond to that annotated for its ortholog in *Streptomyces avermitilis* (Sav2834; UniProtKB entry Q82JC1) which is 90% identical, but starts at the equivalent of residue 26 of the Sco5413 sequence. Hereafter we refer to N-terminally truncated Sco5413 as Sco5413tr. Truncation was achieved by amplifying the nucleotide sequence corresponding to residues 25–169 of Sco5413 from *S. coelicolor* M145 genomic DNA and cloning it into the pOPINF expression vector (http://www.oppf.ox.ac.uk/OPPF/protocols/cloning.jsp) using the In-Fusion cloning system (Clontech). The pOPINF*sco5413tr* construct encoded residues 25–169 of the wild-type Sco5413 sequence with an N-terminal, 3C protease-cleavable, His_6-tag of sequence MAHHHHHHSSGLEVLFQGP. This construct was introduced into *Escherichia coli* BL21(DE3)-pLysS cells by transformation. The protein was expressed by inoculating 1 L of Luria-Bertani medium containing 100 mg carbenicillin and 30 mg chloramphenicol with 10 mL of an overnight culture of these cells, and then growing at 310 K with shaking until the optical density of the sample, measured at a wavelength of 600 nm, reached a value of 0.5. Protein expression was induced by the addition of isopropyl β-D-thiogalactopyranoside to a final concentration of 0.5 m*M*, and shaking was continued at 303 K for a further 3 h. The cells were harvested by centrifugation at 277 K, then resuspended in lysis buffer [25 m*M* Tris-HCl pH 8.0, 0.5*M* NaCl, 40 m*M* imidazole, 1% (vol/vol) Tween-20] and disrupted by sonication. The insoluble fraction was removed by centrifugation at 277 K, and the supernatant was applied to a 1 mL His-Trap Chelating HP Column (GE Healthcare). The column was washed with buffer A (25 m*M* Tris-HCl pH 8.0, 0.5*M* NaCl, 40 m*M* imidazole) to remove any loosely bound proteins, before a linear gradient of imidazole (40–500 m*M*) in buffer A was applied to elute Sco5413tr. The fractions containing Sco5413tr (confirmed by sodium dodecyl sulfate polyacrylamide gel electrophoresis) were pooled then buffer-exchanged and concentrated using an Amicon Ultra-15 10 kDa MW cutoff concentrator (Millipore) into tag-cleavage buffer [25 m*M* Tris-HCl pH 8.0, 0.5*M* NaCl, 2% (vol/vol) β-mercaptoethanol] to a final concentration of 2 mg mL^{-1}. His-tagged 3C-Express Protease (Expedeon) was then added at a ratio of 12 μg protease mg^{-1} of Sco5413tr, and the reaction carried out for 16 h at 277 K. This removed the first 17 amino acids from the protein leaving just two amino acids of the tag (Gly-Pro) before residue 26 of the wild-type sequence. The resultant protein was thus 147 residues in length with a calculated molecular mass of 15,800.1 Da. Cleaved protein was isolated from the reaction solution by a second pass through the His-Trap Column, and the flow-through was collected and applied to a Superdex 75 HiLoad HP gel-filtration column (GE Healthcare) which had been pre-equilibrated with 50 m*M* 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid pH 7.5, 200 m*M* NaCl. Sco5413tr eluted at a volume that corresponded to a molecular mass of ~40 kDa, being somewhat larger than the expected value for a homodimer of 31.6 kDa. The protein was concentrated to ~10 mg mL^{-1} using an Amicon Ultra-4 10 kDa MW cutoff concentrator, with a final yield of approximately 3 mg, and judged to be greater than 95% pure by SDS-PAGE.

Crystallization and X-ray data collection

All crystallizations were performed at a constant temperature of 291 K. Screening was carried out with an OrxyNano robot (Douglas Instruments Ltd) in 96-well sitting-drop MRC plates (Molecular Dimensions) using a variety of commercial screens (Molecular Dimensions

Table I
X-Ray Data Collection and Refinement Statistics for Sco5413tr

	Native[a]	Thimerosal derivative
Data collection		
Beamline	I02, Diamond Light Source, UK	I24, Diamond Light Source, UK
Wavelength (Å)	0.9100	1.0072
Detector	Quantum 315 CCD	Pilatus 6M
Resolution range (Å)	69.72–1.25 (1.28–1.25)	31.55–2.26 (2.32–2.26)
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell parameters (Å)	a = 63.40, b = 65.65, c = 69.72	a = 62.66, b = 63.10, c = 68.71
Total no. of measured intensities	669,686 (35277)	415,840 (17449)
Unique reflections	81,107 (5923)	12,870 (724)
Multiplicity	8.3 (6.0)	32.3 (24.1)
Mean $I/\sigma(I)$	11.5 (2.6)	26.3 (5.4)
Completeness (%)	100.0 (100.0)	97.4 (77.5)
R_{merge}[b]	0.088 (0.709)	0.095 (0.943)
R_{meas}[c]	0.096 (0.868)	0.098 (0.983)
$CC_{1/2}$[d]	0.998 (0.714)	0.999 (0.931)
Wilson B value (Å^2)	14.6	49.6
Refinement		
Resolution range (Å)	47.80–1.25 (1.28–1.25)	—
Reflections: working/free[e]	76966/4062	—
R_{work}/R_{free}[f]	0.147/0.175 (0.205/0.257)	—
Ramachandran plot: favored/allowed/disallowed[g] (%)	100/0/0	—
R.m.s. bond distance deviation (Å)	0.010	—
R.m.s. bond angle deviation (°)	1.409	—
No. of protein residues: chain A/chain B	140 (25–109, 114–168)/139 (25–109, 115–168)	—
No. of water molecules/chloride ions	372/2	—
Mean B factors: protein/water/overall (Å^2)	17.7/31.5/19.8	—
PDB accession code	4B8X	—

Values in parentheses are for the outer resolution shell.
[a]Collected in two passes at medium and high resolution from the same crystal.
[b]$R_{merge} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl} \sum_i I_i(hkl)$.
[c]$R_{meas} = \sum_{hkl} [N/(N-1)]^{1/2} \times \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the ith observation of reflection hkl, $\langle I(hkl)\rangle$ is the weighted average intensity for all observations i of reflection hkl and N is the number of observations of reflection hkl.
[d]$CC_{1/2}$ is the correlation coefficient between intensities from random halves of the dataset.
[e]The data set was split into "working" and "free" sets consisting of 95 and 5% of the data respectively. The free set was not used for refinement.
[f]The R-factors R_{work} and R_{free} are calculated as follows: $R = \sum(| F_{obs} - F_{calc} |) / \sum|F_{obs}| \times 100$, where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively.
[g]As calculated using MolProbity.[15]

and Qiagen). Drops consisted of 0.3 μL protein solution mixed with 0.3 μL precipitant solution with a reservoir volume of 50 μL. Crystals grew from several conditions, with the most promising appearing in 0.2*M* potassium nitrate, 20% (wt/vol) PEG 3350 (JCSG screen, Molecular Dimensions). These conditions were then optimized manually in a 24-well hanging-drop vapor diffusion format using VDX plates (Hampton Research) with a reservoir volume of 1 mL, and drops consisting of 1 μL protein and 1 μL precipitant. The final conditions of 0.2*M* potassium nitrate, 25% (wt/vol) PEG 3350, 6% (vol/vol) MPD, 15% (vol/vol) glycerol, 1 m*M* spermidine, yielded single crystals with dimensions of up to 200 × 200 × 100 μm^3 within 24 h. The composition of the precipitant was such that crystals could be flash-cooled without further cryoprotection.

Crystals were manipulated and flash-cooled in liquid nitrogen using LithoLoops (Molecular Dimensions). The mounted crystals were stored in Unipuck cassettes (MiTeGen) prior to transport to the synchrotron. For experimental phasing, a single crystal was soaked overnight at 291 K in the optimized precipitant solution containing saturating amounts of thimerosal, then back-soaked for 20 min in precipitant solution lacking the compound, before cooling. Crystals were subsequently transferred robotically to the goniostat on the beamline at the Diamond Light Source (Oxfordshire, UK) and maintained at 100 K with a Cryojet cryocooler (Oxford Instruments). Native diffraction data were recorded on beamline I02 using a Quantum 315 CCD detector (ADSC) in two passes, at medium (2.32 Å) and high (1.25 Å) resolutions, respectively. A highly redundant derivative data set was recorded on beamline I24 using a Pilatus 6M detector (Dectris) to 2.26 Å resolution at the L_{III} X-ray absorption edge of mercury (wavelength = 1.0072 Å). All X-ray data were processed using the *XIA2* expert system[7] and the resultant data collection statistics are summarized in Table I. The crystals belonged to space group $P2_12_12_1$ with approximate cell parameters of a = 63, b = 66, c = 70 Å.

Structure solution

The native and the thimerosal derivative data were combined to solve the structure by the Single

Isomorphous Replacement with Anomalous Scattering (SIRAS) method in PHENIX.[8] Two clear mercury sites were located. This solution gave a figure-of-merit (FOM) of 0.318 at 2.29 Å resolution after phasing, which increased to 0.671 after density modification, assuming two copies of the Sco5413tr monomer in the asymmetric unit (giving a corresponding solvent content of 46%). The resultant phases were fed into AutoBuild,[9] which produced a partial model comprised of 238 residues with R_{work} and R_{free} values of 0.283 and 0.345, respectively, and a FOM of 0.616 at 2.29 Å resolution. The model contained two clear subunits, and restrained refinement using REFMAC5[10] against the 1.25 Å resolution native data, gave R_{work} and R_{free} values of 0.306 and 0.323 respectively, and a FOM of 0.553. Phases calculated from this model were subsequently density-modified in PARROT,[11] with twofold averaging to give a FOM of 0.755 at 1.25 Å resolution. Electron density maps calculated from these phases were readily interpretable, so rather than rebuild the existing model, we elected to use automated model building from scratch with BUCCANEER[12]; this yielded a model consisting of 286 residues fitted to sequence (although 15 of these were later removed) with R_{work} and R_{free} values of 0.279 and 0.299 respectively, and a FOM of 0.802. At this stage, automated water fitting with ARP/wARP[13] introduced 354 water molecules, giving a model with R_{work} and R_{free} values of 0.206 and 0.236, respectively, and a FOM of 0.863. Thereafter, several iterations of model building in COOT[14] and restrained refinement (with anisotropic temperature factors) in REFMAC5, yielded the final model consisting of 279 residues (representing 96% of the expected wild-type sequence), 372 water molecules, and two chloride ions, with corresponding R_{work} and R_{free} values of 0.147 and 0.175 respectively, and a FOM of 0.920 at 1.25 Å resolution (refinement statistics are summarized in Table I). MolProbity[15] was used to validate the model before deposition in the PDB with accession code 4B8X. All structural figures were generated using PyMOL.[16]

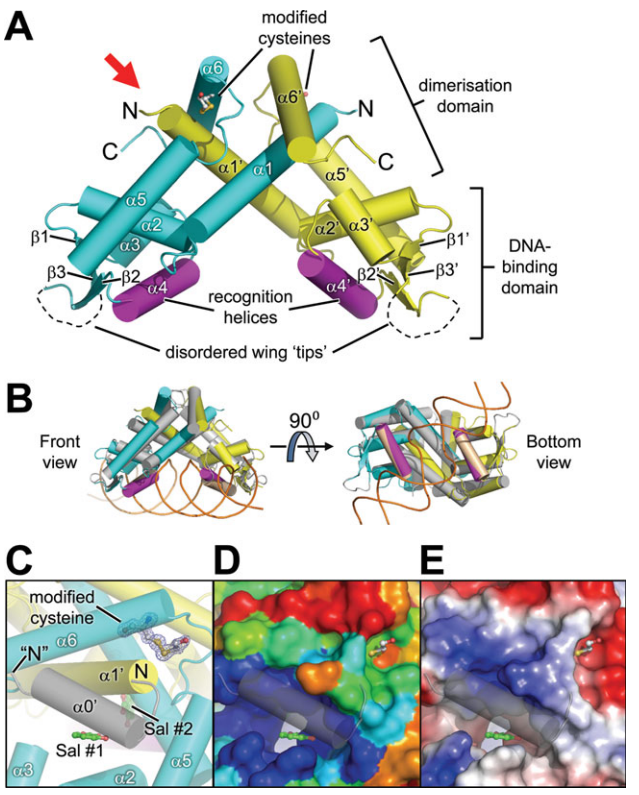

Figure 1

X-ray crystal structure of N-terminally truncated Sco5413. **A**: Cartoon representation of the Sco5413tr homodimer with key features labeled. The two subunits are distinguished by cyan and yellow coloration, and the two recognition helices are shown in magenta. The red arrow denotes the direction of view for panels C–E. **B**: Superposition of Sco5413tr on *B. subtilis* OhrR-*ohrA* operator complex (PDB code: 1Z9C),[3] where Sco5413 is colored as in panel A, OhrR is shown in gray with the recognition helices in pale brown, and the DNA is represented as an orange ribbon. **C**: Close-up of N-terminal region showing the dispositions of the two salicylate molecules (labeled Sal #1 and Sal #2) from the MTH313-salicylate complex (PDB code: 3BPX)[21] after superposition on Sco5413 (shown in semi-transparent cartoon). The predicted location of the putative α0 N-terminal helix is shown in grey (N indicates the N-terminus of the Sco5143tr structure, whereas "N" indicates the predicted N-terminus of the full-length protein). Also shown is 1.25 Å resolution omit electron density for the β-mercaptoethanol adduct of C149 contoured at 3σ. **D**: A molecular surface colored by conservation based on a multiple sequence alignment of 61 sequences vs. Sco5413 (see main text for details), where dark blue indicates strict conservation through to red, which indicates high sequence variability. Note that C149 lies in a region of low sequence conservation, whilst Sal #1 sits in a region of very high sequence conservation. **E**: A molecular surface colored by vacuum electrostatic potential (where red = electronegative, white = neutral, blue = electropositive), showing that Sal #1 sits in a region with both non-polar and electronegative character.

RESULTS AND DISCUSSION

Sco5413tr structure

The Sco5413tr homodimer adopts the roughly triangular overall shape characteristic of MFRs, with approximate dimensions of 67 × 45 × 35 Å^3 [width × height × depth; as viewed in Fig. 1(A)]. Each subunit is comprised of six α-helices (60% of model), three 3_{10}-helices (5% of model), and a three-stranded, antiparallel β-sheet (9% of model), which fold into two discrete domains: a dimerization domain and a DNA-binding domain. The dimerization domain comprises helices α1, α5, and α6, where α1 interdigitates between α5′ and α6′ of the opposing subunit to give a dimer interface covering some 2015 Å^2 and representing 22% of the total solvent accessible surface per subunit (as calculated using the PISA server; http://www.ebi.ac.uk/msd-srv/prot_int). The DNA binding domain comprises helices α2, α3, and α4 and the β-sheet, and displays the topology of the canonical winged helix motif,[17] where the loop between β2 and β3 corresponds to wing W1, although, in common with

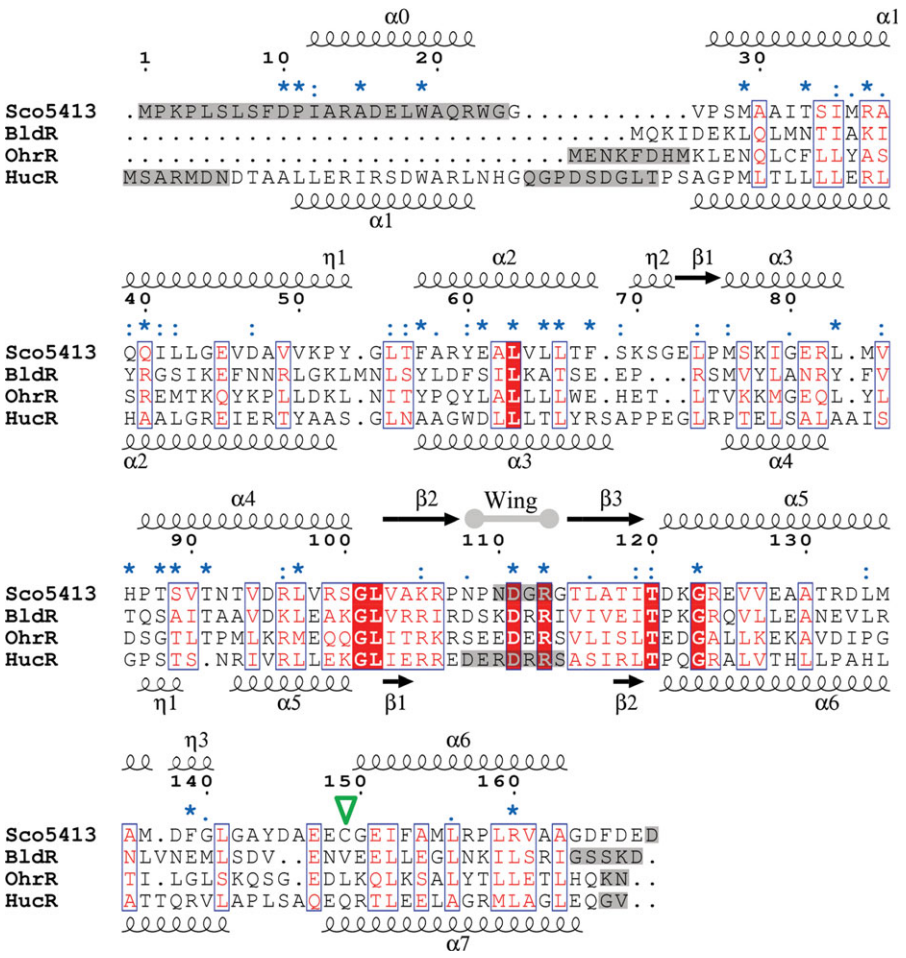

Figure 2

Structure-based multiple sequence alignment of Sco5413 with *Sulfolobus solfataricus* BldR (PDB code: 3F3X),[19] *Bacillus subtilis* OhrR bound to DNA (PDB code: 1Z9C),[3] and *Deinococcus radiodurans* HucR (PDB code: 2FBK).[20] The initial alignment was generated using the SSM server (http://www.ebi.ac.uk/msd-srv/ssm). This was subsequently adjusted manually with reference to the superposed structures, and then displayed using ESPript 2.2 (http://espript.ibcp.fr/ESPript/ESPript). Strictly conserved residues are highlighted with red shaded boxes, and moderately conserved residues are boxed. Secondary structure elements for Sco5413 are shown above the alignment, whilst those for HucR are shown below, where α = α helix, β = β strand, η = 3_{10} helix. Gray shading indicates regions of sequence that were not modeled in the X-ray structures and thus α0 in Sco5413 is only a prediction. The inverted green triangle marks the location of C149 in Sco5413. Characters shown in blue along the top of the alignment denote conservation scores based on a multiple sequence alignment of 61 sequences vs. Sco5413 (see main text for details; where "*" = strictly conserved, ":" = well conserved, and "." = moderately conserved). Note that many of the residues conserved in this latter alignment are not conserved in the three structural homologs.

other MFR protein structures, there is no equivalent of the second wing of the motif, W2. Moreover, as is often the case for MFR structures determined in the absence of bound DNA, the extremities of the W1 wings are disordered in Sco5413tr. In the structures of *Bacillus subtilis* OhrR (PDB code: 1Z9C)[3] and SlyA bound to DNA (PDB code: 3Q5F),[6] the W1 wings form important minor groove interactions.

During model building, it was apparent that C149 in both subunits was modified to *S*,*S*-(2-hydroxyethyl)thiocysteine due to the β-mercaptoethanol present in the His-tag cleavage buffer, which gave an excellent fit to electron density maps [Fig. 1(C) and Supporting Information Fig. S3]. Whilst this may have been indicative of an unusually reactive cysteine, we dismissed it as being a functionally significant residue because it is not conserved amongst closely related protein sequences [Figs. 1(D) and 2], and its location at the N-terminal end of helix α6 [Fig. 1(A,C)] does not correspond to any of the locations of the important cysteine residues seen so far in characterized MFRs.[3–5]

Structural homologs of Sco5413tr

The PFAM database (http://pfam.sanger.ac.uk; release Pfam 26.0, November 2011) reports 58 structures for

MarR family proteins (PF01047), whilst interrogation of the Protein Data Bank using the DALI server (http://ekhidna.biocenter.helsinki.fi/dali_server) retrieves 61 non-redundant entries with high structural similarity to Sco5413tr (Z scores $>$ 10.0). The subsequent comparisons were restricted to selected published structures that gave Z scores greater than 13.0, which are summarized in Supporting Information Table SI. A structure-based sequence alignment of a subset of these structures versus Sco5413 is shown in Figure 2. The superpositions made by DALI are based on a single subunit, and thus were not representative of the structures of the dimers, which may not superpose well due to variations in their quaternary structures. We thus compared monomer and dimer superpositions using Secondary Structure Matching[18] for the selected structural homologs (Supporting Information Table S1). This analysis showed that whilst Sco5413tr was most similar to *Sulfolobus solfataricus* BldR (PDB code: 3F3X)[19] at the level of the subunit based on Z score, the conformation of the dimer was somewhat different from that of BldR [root mean square deviation (rmsd) = 2.68 Å; Supporting Information Fig. S1(A)], and distinctly different from that of *E. coli* MarR (PDB code: 1JGS; rmsd = 3.80 Å),[2] but similar to that of *B. subtilis* OhrR when it is complexed to the *ohrA* operator DNA [rmsd = 1.88 Å; Fig. 1(B)]. Thus we propose that in our crystal structure, Sco5413tr is pre-configured to bind to DNA. In particular, the recognition helices are appropriately disposed relative to one another to engage with consecutive major grooves in DNA.

Despite the predicted disorder for the N-terminal region of Sco5413, secondary structure prediction using PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred) suggests that residues 11–22 are α-helical. An additional N-terminal helix is observed in the structure of *Deinococcus radiodurans* HucR (PDB code: 2FBK),[20] another close structural homolog of Sco5413 [Supporting Information Table SI and Fig. S1(B)]. We thus used a homology model of full-length Sco5413, generated by the PHYRE server (http://www.sbg.bio.ic.ac.uk/~phyre) and based on the HucR structure, to predict the location of this putative N-terminal helix, which we term α0. After superposition on the Sco5413tr subunit, the additional helix lies roughly antiparallel to α1, and spans a wide depression in the surface of the truncated protein [Fig. 1(C–E)].

Sequence conservation and a potential ligand binding site

BLAST searching reveals some 89 amino acid sequences that share at least 50% identity with Sco5413 and, of these, 19 sequences are 90% or more identical. Thus, in order to use amino acid conservation as a means of predicting functional importance, we needed to filter out sequences that had not significantly diverged from that of Sco5413. We therefore ran a redundancy filter at 90% sequence similarity on the top 100 BLAST hits. This left 61 sequences with identities to Sco5413 in the range 46–89%. These sequences were then aligned to that of Sco5413 using CLUSTALW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2). A total of 25 residues were strictly conserved in this alignment (Fig. 2). These were seen to cluster either on the predicted DNA interaction surface, in particular around the N-terminal half of the α4 recognition helices [Supporting Information Fig. S2(B)], or in two symmetry-equivalent intersubunit channels created at the dimer interface involving residues mainly from helices α1′ and α2, and partially occluded by the putative helix α0′ [Fig. 1(C–E)]. Several MFRs have been co-crystallized with salicylate, although the biological relevance of some of the binding sites is questionable in those cases where the complexes were obtained at very high ligand concentrations, and where the ligands do not occupy well-defined pockets, e.g. *E. coli* MarR (250 m*M* sodium salicylate; PDB code: 1JGS).[2] By contrast, the crystal structure of MTH313 from *Methanobacterium thermoautotrophicum* with bound salicylate was determined at a significantly lower ligand concentration (40 m*M* sodium salicylate) and the ligand binds asymmetrically, occupying two discrete non-equivalent pockets in the homodimer (PDB code: 3BPX).[21] After superposition of the MTH313 ligand-bound structure on Sco5413tr, it can be seen that one of these sites [labelled Sal #1 in Fig. 1(C–E) and Supporting Information Fig. S1(C)] corresponds roughly to the constriction within the Sco5413tr channel that is lined by conserved residues. Taken together, these observations suggest that this site may represent an allosteric effector binding site. An analysis of the electrostatic surface in this region revealed both non-polar and electronegative character [Fig. 1(E)], perhaps indicative of an ability to bind amphipathic ligands. In the full-length protein, it is conceivable that access to this site may be restricted from the direction of view shown in Figure 1(C–E) by the putative α0 N-terminal helix, which may have a gating function, whilst at the same time providing additional residues for ligand binding. It is notable that two of the strictly conserved residues from the above alignment, A15 and W19 (Fig. 2), have been modeled by PHYRE on the inward facing side of α0. Indeed, W19 is structurally conserved in HucR (W20; HucR numbering) and has been implicated in binding the effector molecule urate.[22] Alternatively, if α0 does not move, the effector site of Sco5413 could be accessed through the intersubunit channel from the opposite direction via the twofold axis of the homodimer.

CONCLUSIONS

We have determined the crystal structure of N-terminally truncated Sco5413 from *S. coelicolor*, which represents the highest resolution MFR structure determined to

date, and provides a three-dimensional platform for the interrogation of both sequence and structural databases. In particular, comparisons with other known MFR structures indicate that the protein adopts a conformation that would be compatible with DNA binding, and that a channel located at the dimer interface lined by well-conserved residues may represent a binding site for an amphipathic effector molecule. We will use virtual ligand screening to identify candidate ligands for this site and validate these using biophysical methods and further crystallographic studies. Moreover, experiments to define the Sco5413 regulon, using bioinformatics, chromatin immunoprecipitation, and electrophoretic mobility shift assays, are in hand. In aggregate, these data will enable a comprehensive understanding of this widespread actinomycete transcription factor that is implicated in the regulation of growth and antibiotic production.

ACKNOWLEDGMENTS

Authors acknowledge the support of the Diamond Light Source and the assistance of beamline scientists on stations I02 and I24. Govind Chandra, Andrew Hesketh, and Maureen Bibb are also thanked for helpful advice and further assistance.

REFERENCES

1. Perera IC, Grove A. Molecular mechanisms of ligand-mediated attenuation of DNA binding by MarR family transcriptional regulators. J Mol Cell Biol 2010;2:243–254.
2. Alekshun MN, Levy SB, Mealy TR, Seaton BA, Head JF. The crystal structure of MarR, a regulator of multiple antibiotic resistance, at 2.3 Å resolution. Nat Struct Biol 2001;8:710–714.
3. Hong M, Fuangthong M, Helmann JD, Brennan RG. Structure of an OhrR-*ohrA* operator complex reveals the DNA binding mechanism of the MarR family. Mol Cell 2005;20:131–141.
4. Newberry KJ, Fuangthong M, Panmanee W, Mongkolsuk S, Brennan RG. Structural mechanism of organic hydroperoxide induction of the transcription regulator OhrR. Mol Cell 2007;28:652–664.
5. Palm GJ, Khanh Chi B, Waack P, Gronau K, Becher D, Albrecht D, Hinrichs W, Read RJ, Antelmann H. Structural insights into the redox-switch mechanism of the MarR/DUF24-type regulator HypR. Nucleic Acids Res 2012;40:4178–4192.
6. Dolan KT, Duguid EM, He C. Crystal structures of SlyA protein, a master virulence regulator of *Salmonella*, in free and DNA-bound states. J Biol Chem 2011;286:22178–22185.
7. Winter G. Xia2: an expert system for macromolecular crystallography data reduction. J Appl Crystallogr 2010;43:186–190.
8. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr Sect D 2010;66:213–221.
9. Terwilliger TC, Grosse-Kunstleve RW, Afonine PV, Moriarty NW, Zwart PH, Hung LW, Read RJ, Adams PD. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. Acta Crystallogr Sect D 2008;64:61–69.
10. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr Sect D 1997;53:240–255.
11. Cowtan K. Recent developments in classical density modification. Acta Crystallogr Sect D 2010;66:470–478.
12. Cowtan K. The Buccaneer software for automated model building. I. Tracing protein chains. Acta Crystallogr Sect D 2006;62:1002–1011.
13. Perrakis A, Morris R, Lamzin VS. Automated protein model building combined with iterative structure refinement. Nat Struct Biol 1999;6:458–463.
14. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallogr Sect D 2004;60:2126–2132.
15. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB, 3rd, Snoeyink J, Richardson JS, Richardson DC. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res 2007;35:W375–W383.
16. DeLano WL. The PyMOL user's manual. San Carlos, CA, USA: DeLano Scientific; 2002.
17. Gajiwala KS, Burley SK. Winged helix proteins. Curr Opin Struct Biol 2000;10:110–116.
18. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr Sect D 2004;60:2256–2268.
19. Di Fiore A, Fiorentino G, Vitale RM, Ronca R, Amodeo P, Pedone C, Bartolucci S, De Simone G. Structural analysis of BldR from *Sulfolobus solfataricus* provides insights into the molecular basis of transcriptional activation in Archaea by MarR family proteins. J Mol Biol 2009;388:559–569.
20. Bordelon T, Wilkinson SP, Grove A, Newcomer ME. The crystal structure of the transcriptional regulator HucR from *Deinococcus radiodurans* reveals a repressor preconfigured for DNA binding. J Mol Biol 2006;360:168–177.
21. Saridakis V, Shahinas D, Xu X, Christendat D. Structural insight on the mechanism of regulation of the MarR family of proteins: high-resolution crystal structure of a transcriptional repressor from *Methanobacterium thermoautotrophicum*. J Mol Biol 2008;377:655–667.
22. Perera IC, Lee YH, Wilkinson SP, Grove A. Mechanism for attenuation of DNA binding by MarR family transcriptional regulators by small molecule ligands. J Mol Biol 2009;390:1019–1029.