scientific reports

OPEN

Crystal structure report of the ImmR transcriptional regulator DNA-binding domain of the *Bacillus subtilis* ICE*Bs1* transposon

Rosanna Caliandro[1,2], Iñaki de Diego[1,3] & F. Xavier Gomis-Rüth[1✉]

***Bacillus subtilis* is a commensal member of the human oral and gut microbiomes, which can become infectious to immunocompromised patients. It possesses a conjugative transposon, ICE*Bs1*, which includes > 20 genes and can be passed by horizontal gene transfer to other bacteria, including pathogenic *Bacillus anthracis* and *Listeria monocytogenes*. ICE*Bs1* is regulated by the ImmR/ ImmA tandem, which are a transcriptional repressor that constitutively blocks transcription and a metallopeptidase that acts as anti-repressor and inactivates ImmR by proteolytic cleavage. We here report the production and purification of 127-residue ImmR from ICE*Bs1* and the crystal structure of its DNA-binding domain. It features a five-helix bundle centred on a helix-turn-helix motif potentially binding the major grove of double-stranded target DNA. ImmR shows structural and mechanistic similarity with the *B. subtilis* SinR repressor, which is engaged in sporulation inhibition.**

Bacillus subtilis is a Gram-positive bacterium found in the gastrointestinal tract[1] and the oral cavity of humans[2], for which it is classified as generally regarded as safe (GRAS)[3]. However, it has been occasionally associated with food poisoning that leads to diarrhea, sickness, and fever in immunocompromised patients[4–6]. Moreover, it shares many features with members of the *Bacillus Cereus Group*[7], which includes human pathogens such as *Bacillus anthracis* and *B. cereus*.

Conjugative transposons, also referred to as "integrative and conjugative elements" (ICEs), are widespread mobile genetic elements that integrate into the genome of bacteria and provide extra functionalities[8]. They can be excised and shared with other bacteria through conjugation, thus contributing to genome plasticity and the spreading of antibiotic resistance and virulence factors across species[9]. One such ICE from *B. subtilis* is ICE*Bs1*[10], which can be transferred to *B. anthracis, Bacillus licheniformis*, and *Listeria monocytogenes*[11]. It spans 20 kb and contains over 20 genes transcribed from the P*xis* promoter, which code for the excisionase Xis, the relaxase NicK, the regulator RapI, and the regulatory peptide PhrL, among others[9,12]. Regulation of ICE*Bs1* is exerted by the Int integrase, the ImmR transcriptional repressor, and the ImmA anti-repressor metallopeptidase, which are counter-transcribed from the P*immR* promoter of the transposon[9]. By binding to six sites within the regulatory regions of both promoters, ImmR exerts a repressing function that ensures that a single stable copy of ICE*Bs1* is maintained in the cell in the quiescent state[13]. In contrast, if the global DNA damage response is launched or if potential recipient cells lacking the transposon are nearby, ImmA inactivates ImmR by proteolytic cleavage, which unleashes ICE*Bs1* expression and promotes transposon transfer[11,14].

ImmR is a 127-residue intracellular protein (UniProt[15] access code P96631), which was identified as a transcriptional regulator based on sequence similarity with bacteriophage(-like) double-stranded(ds) DNA-binding repressors[10,11,14]. It was predicted to encompass a DNA-binding domain (DBD) with a helix-turn-helix (HTH) motif within its first 61 residues[11,14]. Moreover, the protein was annotated within UniProt as a "HTH-type transcriptional regulator" based on PROSITE-ProRule annotation (PRU00257[16]) but experimental validation

[1]Proteolysis Laboratory, Department of Structural Biology, Molecular Biology Institute of Barcelona (CSIC), Barcelona Science Park, C/Baldiri Reixac, 15-21, 08028 Barcelona, Catalonia, Spain. [2]Present address: LenioBio GmbH, Erkrather Str. 401, 40231 Düsseldorf, Germany. [3]Present address: Sample Environment and Characterization Group, European XFEL GmbH, Holzkoppel 4, 22869 Schenefeld, Germany. ✉email: xgrcri@ibmb.csic.es

Dataset	
Beam line (synchrotron)	ID23-2 (ESRF)
Space group/protomers per a.u	I2/2
Cell constants (a, b, and c in Å; β in °)	54.03, 48.34, 64.46, 97.47
Wavelength (Å)	0.87260
Measurements/unique reflections after anisotropy cut-off	26,247/9322
Resolution range (Å) (outermost shell)[a]	43.97–2.10 (2.15–2.10)
Spherical/ellipsoidal completeness (%)[b]	94.0 (71.5)/94.7 (80.8)
R_{merge}/R_{pim}/CC($^1/_2$)[c]	0.255 (0.617)/0.186 (0.433)/0.904 (0.591)
$<I>/\sigma(I)$[d]/average multiplicity	3.3 (1.7)/2.8 (2.9)
Overall anisotropy B-tensor	19.4, 9.7, 33.41
Resolution range used for refinement (Å)	43.97–2.10
Reflections used (test set)	8867 (454)
Crystallographic R_{factor} (free R_{factor})[c]	0.259 (0.346)
Non-H protein atoms/waters per a.u	1062/142
Rmsd from target values	
Bonds (Å)/angles (°)	0.014/1.29
Average B-factor (Å^2)	12.8
Protein contacts and geometry analysis[e]	
Ramachandran favoured/allowed/outliers/all analysed	126 (100%)/0/0/126
Bond-length/bond-angle/chirality/planarity outliers	2/0/0/0
Side-chain outliers	8 (7.1%)
All-atom clashes/clashscore[e]	24/10.5
RSRZ outliers[e]/F_o:F_c correlation	5 (3.9%)/0.84
PDB access code	7T8I

Table 1. Crystallographic data. *a.u.* asymmetric unit, *rmsd* root-mean square deviation, *RSRZ* real-space R-value Z-score. [a]Values for data processing in parenthesis refer to the outermost resolution shell if not otherwise indicated. [b]According to *Mrfana* within *Staraniso*[21]. [c]For definitions, see[22]. [d]$<I>/\sigma(I)$ of unique reflections after merging according to *Mrfana*. [e]According to the wwPDB Validation Service (https://wwpdb-validation.wwpdb.org/validservice).

is missing. We hereby report the recombinant protein production and purification of ImmR and the crystal structure determination of its DBD.

Results and discussion

Structure analysis of the ImmR-DBD. Full-length ImmR of *B. subtilis* was produced by recombinant overexpression in *Escherichia coli* and purified through two chromatography steps. Apparently suitable crystals were routinely obtained but diffraction was consistently restricted to 7–8 Å. Eventually, crystals diffracting to around 2 Å were measured back in 2013 at the ESRF synchrotron beamline ID23-2 (Table 1). However, these crystals suffered from high mosaicity and anisotropy. Moreover, diffraction showed diffuse streaks in several regions of the reciprocal space, potentially arising from planar or linear lattice defects, so that individual diffraction spots were not properly resolved. Given the absence of heavy-atom/ion derivatives or a suitable model for molecular replacement, the project was discontinued until this year, when a predicted model for full-length 127-residue ImmR was obtained with *AlphaFold*[17]. This model divides into a compact high-confidence (∅pLDDT = 96.7%; see[17] for definition) N-terminal DBD (M^1–G^{63}) and a loose C-terminal domain (K^{64}–E^{127}) containing two large isolated α-helices (K^{64}–K^{88} and E^{103}–K^{126}), which was predicted with lower overall confidence (∅pLDDT = 74.8%). This result motivated us to reprocess the original diffraction data with up-to-date software.

Data processing with *Xds*[18] and *Dials*[19] failed in our hands to yield data that would enable crystallographic refinement. Eventually, *iMosflm*[20] processing, which reportedly deals better with data with large mosaicity and ΔΦ values, followed by anisotropy correction with *Staraniso*[21], enabled us to get a suitable reflection file for model refinement. This processing yielded comparably high values for the R_{merge} parameter[22] (see Table 1) but absence of twinning and translational non-crystallographic symmetry. Subsequently, the structure was solved by molecular replacement. While no solution satisfying the packing criterium was obtained for the full-length searching model, two clear solutions were found for the DBD model alone. These solutions showed values for the refined translation functions of 9.9 and 20.4, respectively, and a final log-likelihood gain of 356. After successive rounds of model building and refinement, the final model comprised residues M^1–K^{64} of molecules A and B, plus respective N-terminal alanines (A^0) from the purification tag[23], and 167 solvent molecules. The final values for R_{factor} and R_{free}[22] were comparably high for a dataset to 2.1 Å resolution (Table 1), which we attribute to the above crystal pathologies. This notwithstanding, the final (2mF_{obs} − DF_{calc})-type Fourier map was of excellent quality for both molecules (Fig. 1A), as were the general model validation parameters (Table 1), so that we

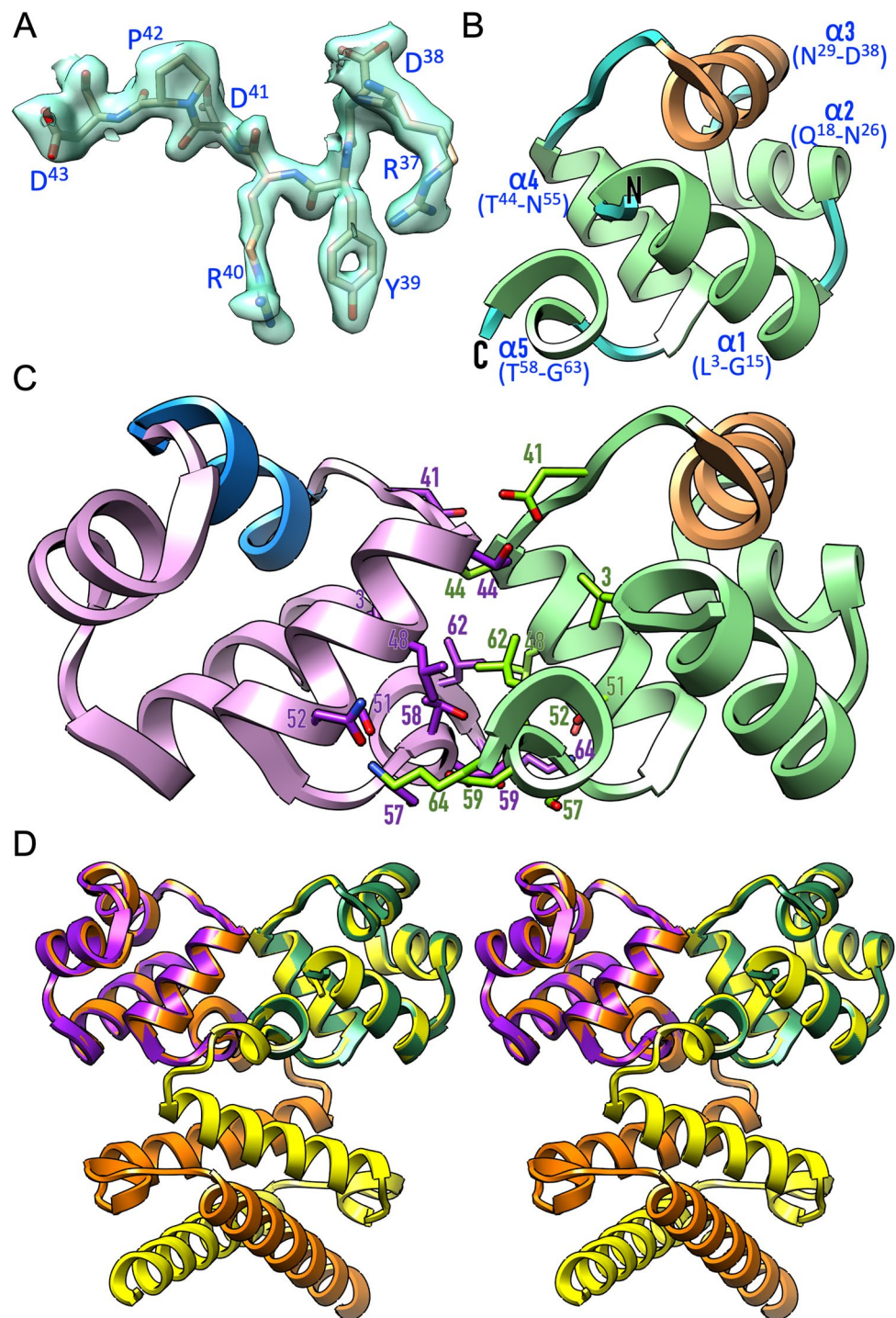**Figure 1.** Structure of *B. subtilis* ImmR. **(A)** Representative fragment of the final $(2mF_{obs} - DF_{calc})$-type Fourier map displayed at 1 σ above threshold as a semi-transparent turquoise surface, superimposed with segment R^{37}–D^{43} of the final refined experimental structure. **(B)** Ribbon-type plot of the ImmR DBD, which consists of five helices (α1–α5). The recognition helix is shown in sandy brown. **(C)** Dimeric arrangement of the ImmR DBD in the crystals. Each protomer is shown in one colour (light green and plum), except the recognition helices (sandy brown and dodger blue). Residues from each protomer contributing to the interface are shown as sticks, with carbons coloured as the respective ribbon, and labelled. **(D)** Superposition in cross-eye stereo of the experimental DBD dimer (chains in purple and green) and the predicted *AlphaFold* dimer of the full-length structure (chains in orange and yellow).

are confident that our experimental structure provides a valuable model for the protein. Remarkably, the final Fourier map did not show relevant density beyond K[64] of either protomer, which would be compatible with the full-length molecule set up for crystallisation being proteolytically processed after this lysine by a trypsin-like contaminant. This hypothesis is supported by calculation of the Matthews-coefficient[24], which would be 1.4 Å^3/Da (14% solvent contents) for the full-length protein, which is unlikely. In contrast, the values for the current experimental model (2.8 Å^3/Da; 57% solvent contents) are in accordance with the literature[25].

Description of the ImmR-DBD.

The protein is a compact almost spherical pentahelical bundle (α1–α5) cohered by a central hydrophobic core, in which the N- and the C-terminal helices are nearly antiparallel, so that the chain termini are adjacent (Fig. 1B). Helices α2–α4 form a flap that folds back onto the two terminal helices. Overall, the five helices are connected by short linkers of 2-to-5 residues and each helix is approximately perpendicular to the preceding one. Following the nomenclature of HTH$_{GBB}$-DBDs[26], helices α2 and α3 would correspond to the "positioning helix" and the "recognition helix" of the HTH-motif engaged in double-stranded DNA recognition.

The two protomers in the asymmetric unit (a.u.) are related by a dyad, which gives rise to an interface of 573 Å^2 (Δ^iG = − 2.1 kcal/mol; Δ^iG P-value = 0.424[27]). The interface involves 56 and 49 atoms of 18 residues of molecules A and B, respectively, which overall perform nine hydrogen bonds, as well as symmetric hydrophobic interactions between 11 residues of either molecule. The main participating residues are L^3, D^{41}, T^{44}, L^{47}, L^{48}, S^{51}, N^{52}, T^{58}, D^{59}, L^{62}, and K^{64} (Fig. 1C), which are provided by helices α4 and α5 plus the linker preceding α4. Finally, the experimental structure is in very good agreement with the predicted dimer (Fig. 1D). Indeed, the 130 residues of the former coincided with the predicted model with a core *rmsd* of 0.43 Å. Moreover, this superposition further revealed that the C-terminal α-helix of the full-length protein would clash with a symmetric DBD mate, which further supports that the crystal only contained the DBD (see "Structure analysis of the ImmR-DBD" section).

Similar structure.

A search with *Dali* identified several members of the "434 Cro family" of HTH-DBDs from bacteria or bacteriophages[26] as structurally related. Closest similarity was found with 111-residue SinR from *B. subtilis*, followed by the C2 repressor of *Salmonella* bacteriophage P22 (PDB 1ADR[28]), CylR2 of *Enterococcus faecalis*, and DdrO of *Deinococcus geothermalis* (Fig. 2A).

In all structures, the first four helices have a very similar arrangement (Fig. 2A), and significant differences are only found in the respective fifth helices. These have variable length and are shifted along the polypeptide chain in the different structures, which supports that the minimal functional unit for these domains is a four-helix bundle[26]. Moreover, SinR, CylR2, and DdrO evince dimeric crystal structures that are equivalent to that of ImmR (Fig. 2B). In the case of SinR, this dimeric arrangement was functionally validated through the crystal structure of a dsDNA complex[29] and further suggests that ImmR may oligomerize for the production of DNA-loop structures similar to SinR[30]. We constructed a homology model for the DNA-complex of the ImmR-DBD dimer based on the SinR complex (Fig. 2C). Accordingly, the DNA major groove would be contacted through the recognition helices, and flanking helices α3 and α4 would play a supportive role. Putative residues engaged in binding would encompass T^{17}–E^{20}, N^{29}–N^{31}, S^{33}–Y^{35}, R^{37}, and Y^{39}–D^{43} of either protomer.

Remarkably, archetypal 434 Cro repressor just spans the pentahelical HTH-DBD[31] but other family members are C-terminally extended and comprise additional domains. This is the case for SinR, which has two helices engaged in dimerization and binding to other proteins (PDB 1B0N[32]) that are very similar to the *AlphaFold* prediction for ImmR (see "Structure analysis of the ImmR-DBD" section). Given that SinR is currently the closest structural relative of ImmR, both C-terminal regions may have similar functions. Indeed, ImmA inactivates ImmR through cleavage at F^{95}–M^{96}, which is in the linker between the two predicted helices[14]. This would be consistent with the protein:dsDNA complex falling apart upon cleavage, thus releasing transcriptional repression.

Materials and methods

Protein production and purification.

The ImmR gene was amplified from *Bacillus subtilis* strain 168 using 5′-CAATCATATGAGCCTAGGCAAACGATTAAAAGAAG-3′ and 5′-CAATCTCGAGTCAC TCTTTC TTCTTTAATTCGTCAATG-3′ as forward and backward primers, respectively. The PCR product was cloned into the pCri8b vector using *Nde*I and *Xho*I restriction sites, which attaches an N-terminal hexahistidine (His$_6$)-tag followed by a tobacco-etch virus (TEV) recognition sequence to the target protein[23]. The plasmid was transformed into *Escherichia coli* BL21 (DE3) cells, which were grown at 20 °C in Luria Bertani medium containing ampicillin (30 μg/mL) and chloramphenicol (34 μg/mL) under agitation (220 rpm) until an OD$_{600}$ of 0.6–1.0 was reached. Expression was then induced by adding 400 μM isopropyl-β-D-thiogalactopyranoside, and the culture was incubated for further 12 h. Cells were harvested by centrifugation at 4000×*g* for 15 min at 4 °C and resuspended in lysis buffer (20 mM Tris–HCl pH 7.5, 5 mM magnesium chloride, 20 mM imidazole, 10 μg/mL DNAse). Cells were lysed in a cell disruptor (Constant Systems, Ltd.), and the lysate was clarified by centrifugation for 1 h at 30,000×*g* at 4 °C. Sodium chloride (1.5 M) was then added to the supernatant and incubated at room temperature for 45 min prior to nickel nitrilotriacetic affinity chromatography purification (NiNTA resin from Invitrogen). The resin had been preequilibrated with buffer A (20 mM Tris–HCl pH 7.5, 1.5 M sodium chloride, 20 mM imidazole), and the protein was eluted with buffer B (20 mM Tris–HCl pH 7.5, 1.5 M sodium chloride, 300 mM imidazole). The protein was then dialysed against buffer A to remove excess of imidazole and incubated with His$_6$-tagged TEV protease at a 1:10 molar ratio over night at 4 °C to cleave the N-terminal His$_6$-tag. The protein solution was then reapplied to the NiNTA resin pre-equilibrated as before to remove the TEV protease, the cleaved His$_6$-tags and non-cleaved N-terminally His$_6$-tagged ImmR. The flow through was collected and concentrated to ~2 mL using a Vivaspin 20 ultrafiltration device of 5-kDa cut-off (Sartorius). The sample was then run through a Superdex 200 16/60 column (GE Healthcare), which had been attached to an
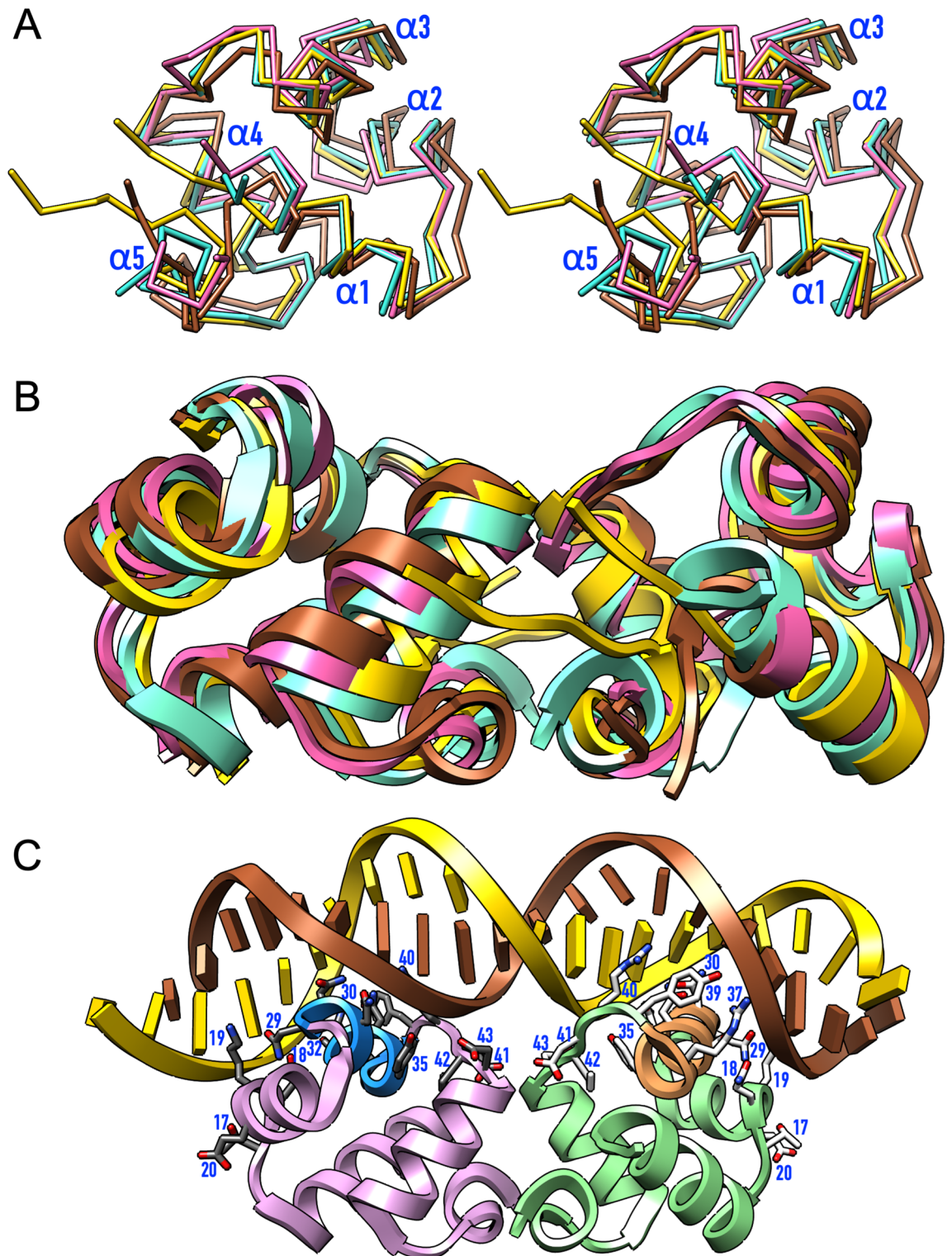
Figure 2. Structural similarities and presumable DNA binding. **(A)** Superposition in cross-eye stereo of the Cα-traces of the monomer of ImmR (aquamarine), which is shown in the orientation of Fig. 1B, onto SinR (hot pink; 62 aligned residues with ImmR show a core *rmsd* of 1.11 Å, 35% sequence identity, and a *Dali* Z-score of 11.4; PDB 1B0N[32]), CylR2 (gold; 56 residues, 0.91 Å, 23%, 11.0; PDB 1UTX[45]), and DdrO (sienna; 63 residues, 1.64 Å, 33%, 10.9; PDB 6JQ1[46]). The five helices are labelled. **(B)** Superposition of the DBD dimers of ImmR (aquamarine), SinR (hot pink; *rmsd* = 1.92 Å for 118 aligned residues; PDB 3ZKC[29]), CylR2 (gold; *rmsd* = 1.45 Å for 119 aligned residues; PDB 1UTX) and DdrO (sienna; *rmsd* = 2.12 Å for 119 aligned residues; PDB 6JQ1). **(C)** Homology model of the protein:dsDNA complex of ImmR based on the structure of the equivalent SinR complex (PDB 3ZKC[29]). The orientation of the protein is the same as in Fig. 1C. Protein residues hypothetically participating in the protein:dsDNA interface are shown as sticks with white and grey carbons for either protomer, respectively, and labelled.

ÄKTA liquid chromatography system (GE Healthcare) and equilibrated with buffer C (20 mM Tris–HCl pH 7.5, 1 M sodium chloride). Fractions corresponding to the protein of interest were collected, and the protein purity and molecular mass (theoretic value 14.8 kDa) were assessed through SDS-PAGE. Protein concentration was determined with a Nanodrop spectrophotometer (Thermo Fisher Scientific) using the theoretical absorption coefficient ($\varepsilon = 7450$ M^{-1} cm^{-1}) calculated by *ProtParam* within *Expasy*[33]. Protein identity was confirmed by peptide mass fingerprinting analysis at the Protein Chemistry Service and the Proteomics Facilities of the Centro de Investigaciones Biológicas (Madrid, Spain). Briefly, samples were subjected to 10% SDS-PAGE, and gels were stained for 5 min with freshly prepared Coomasie Blue Stain (0.1% solution in 40% methanol/10% acetic acid) and destained for 15 min in 50% methanol. Gel bands were excised with a clean razor blade and placed in a 1.5-mL Eppendorf tube with 50 μL H_2O for wet shipment.

Crystallisation and data collection.　Pure protein in 20 mM Tris–HCl pH 7.5, 100 mM sodium chloride was concentrated to 6.5 mg/mL and employed to screen crystallisation conditions applying the sitting-drop vapor diffusion method at the Automated Crystallography Platform (https://www.ibmb.csic.es/en/facilities/automated-crystallographic-platform). Crystallization solutions were prepared with a Freedom EVO robot (Tecan) and pipetted into the reservoir wells of 96×2-well MRC crystallization plates (Innovadyne Tech.). Nanodrops consisting of 100 nL of each reservoir solution and protein solution were dispensed by a Cartesian Microsys 4000 XL robot (Genomic Solutions) into the shallow wells of the crystallization plates, which were stored at 4 °C or 20 °C in thermostatic crystal farms (Bruker). Upscaling and optimization were performed by sitting-drop vapor diffusion, using 2 μL protein solution and 1 μL precipitant solution in 24-well Cryschem crystallization plates (Hampton Research).

Suitable crystals of ImmR-DBD were obtained with 18% (w/v) PEG 3350, 10 mM magnesium chloride, 50 mM Tris–HCl pH 8.5 as reservoir solution. Crystals were harvested with cryo-loops (Molecular Dimensions), cryoprotected, flash-vitrified in liquid nitrogen, and stored for data collection. X-ray diffraction data were recorded at 100 K on a 225-mm MARMOSAIC CCD detector (MAR Research) at the ID23-2 beamline[34] of the ESRF synchrotron (Grenoble, France). Crystals were indexed as space group I2, with two protomers per a.u.. Diffraction data were processed with programs *iMoslfm*[20] and *Staraniso*[21], which included the *Mrfana* analysis routine, to obtain structure-factor amplitudes in MTZ-format for the *Phenix*[35] and *Ccp4*[36] suites of programs. Data were further assessed with *Xtriage*[37] within *Phenix* and *Pointless*[38] within *Ccp4*. Statistics on data collection and processing are provided in Table 1.

Structure solution and refinement.　The structure was solved by molecular replacement with the *Phaser* program[39] using a homology model for the ImmR-DBD monomer obtained with *AlphaFold*[17]. These calculations yielded two unique solutions at Eulerian angles (in °) $\alpha = 116.1$, $\beta = 73.5$, $\gamma = 25.4$ (fractional cell coordinates 0.214, 0.998, 0.333) and $\alpha = 296.7$, $\beta = 73.6$, $\gamma = 25.2$ (fractional cell coordinates 0.815, 0.893, 0.331), respectively, which are related by a dyad parallel to cell axis c. The associated values for the translation functions after refinement were 9.9 and 20.4, respectively, and the final log-likelihood gain was 356. The adequately rotated and translated molecules were subjected to the *Autobuild*[40] protocol within *Phenix*, which yielded a Fourier map of high quality for manual model building with the *Coot* program[41]. The latter alternated with crystallographic refinement using the *Refine* protocol of *Phenix*[35], which included hydrogens in riding positions and translation/libration/screw-motion plus non-crystallographic symmetry restraints, until completion of the model. Table 1 provides essential statistics on the final refined model, which was validated trough the wwPDB validation service (https://validate-rcsb-1.wwpdb.org/validservice). The coordinates can be retrieved from the Protein Data Bank (www.pdb.org) under access code 7T8I.

Miscellaneous.　Structural relatives were identified through the *Dali*[42] server (ekhidna2.biocenter.helsinki.fi/dali). Structure superpositions were calculated with *Ssm*[43] in *Coot*. Figures were prepared using *Chimera*[44]. Protein interfaces and intermolecular interactions were analyzed using *PDBePISA* (www.ebi.ac.uk/pdbe/pisa)[27] and verified by visual inspection. The interacting surface of a complex was taken as half the sum of the buried surface areas of either molecule. A homology model of the complex between the ImmR-DBD dimer and target dsDNA was obtained by superposing the ImmR dimer onto the SinR dimer within its experimental DNA complex (PDB 3ZKC[29]). This model is provided as Supplementary File 1. The ImmR chain was then slightly readjusted manually with *Coot* and geometry-minimised with the same program to iron out clashes and unfavourable side-chain conformations. The dsDNA part was kept intact.

Data availability

The coordinates and structure factors generated during the current study are available from the Protein Data Bank (www.pdb.org) under access code 7T8I.

References

1. Hong, H. A. *et al. Bacillus subtilis* isolated from the human gastrointestinal tract. *Res. Microbiol.* **160**, 134–143. https://doi.org/10.1016/j.resmic.2008.11.002 (2009).
2. Jain, K., Parida, S., Mangwani, N., Dash, H. R. & Das, S. Isolation and characterization of biofilm-forming bacteria and associated extracellular polymeric substances from oral cavity. *Ann. Microbiol.* **63**, 1553–1562. https://doi.org/10.1007/s13213-013-0618-9 (2013).

3. Errington, J. & Aart, L. T. V. Microbe profile: *Bacillus subtilis*: model organism for cellular development, and industrial workhorse. *Microbiology (Reading)* **166**, 425–427. https://doi.org/10.1099/mic.0.000922 (2020).

4. Logan, N. A. *Bacillus* species of medical and veterinary importance. *J. Med. Microbiol.* **25**, 157–165. https://doi.org/10.1099/00222615-25-3-157 (1988).

5. Pavić, S. *et al.* An outbreak of food poisoning in a kindergarten caused by milk powder containing toxigenic *Bacillus subtilis* and *Bacillus licheniformis*. *Archiv f. Lebensmittelhyg.* **56**, 20–22 (2005).

6. Jeon, Y. L. *et al.* Combined *Bacillus licheniformis* and *Bacillus subtilis* infection in a patient with oesophageal perforation. *J. Med. Microbiol.* **61**, 1766–1769. https://doi.org/10.1099/jmm.0.042275-0 (2012).

7. Ehling-Schulz, M., Lereclus, D. & Koehler, T. M. The *Bacillus cereus* group: Bacillus species with pathogenic potential. *Microbiol. Spectr.* https://doi.org/10.1128/microbiolspec.GPP3-0032-2018 (2019).

8. Johnson, C. M. & Grossman, A. D. Integrative and conjugative elements (ICEs): What they do and how they work. *Annu. Rev. Genet.* **49**, 577–601. https://doi.org/10.1146/annurev-genet-112414-055018 (2015).

9. Bose, B. & Grossman, A. D. Regulation of horizontal gene transfer in *Bacillus subtilis* by activation of a conserved site-specific protease. *J. Bacteriol.* **193**, 22–29. https://doi.org/10.1128/JB.01143-10 (2011).

10. Burrus, V., Pavlovic, G., Decaris, B. & Guedon, G. The ICE*St1* element of *Streptococcus thermophilus* belongs to a large family of integrative and conjugative elements that exchange modules and change their specificity of integration. *Plasmid* **48**, 77–97. https://doi.org/10.1016/s0147-619x(02)00102-6 (2002).

11. Auchtung, J. M., Lee, C. A., Monson, R. E., Lehman, A. P. & Grossman, A. D. Regulation of a *Bacillus subtilis* mobile genetic element by intercellular signaling and the global DNA damage response. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 12554–12559. https://doi.org/10.1073/pnas.0505835102 (2005).

12. Lee, C. A., Auchtung, J. M., Monson, R. E. & Grossman, A. D. Identification and characterization of int (integrase), xis (excisionase) and chromosomal attachment sites of the integrative and conjugative element ICEBs1 of *Bacillus subtilis*. *Mol. Microbiol.* **66**, 1356–1369. https://doi.org/10.1111/j.1365-2958.2007.06000.x (2007).

13. Auchtung, J. M., Lee, C. A., Garrison, K. L. & Grossman, A. D. Identification and characterization of the immunity repressor (ImmR) that controls the mobile genetic element ICEBs1 of *Bacillus subtilis*. *Mol. Microbiol.* **64**, 1515–1528. https://doi.org/10.1111/j.1365-2958.2007.05748.x (2007).

14. Bose, B., Auchtung, J. M., Lee, C. A. & Grossman, A. D. A conserved anti-repressor controls horizontal gene transfer by proteolysis. *Mol. Microbiol.* **70**, 570–582. https://doi.org/10.1111/j.1365-2958.2008.06414.x (2008).

15. UniProt, C. UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489. https://doi.org/10.1093/nar/gkaa1100 (2021).

16. Sigrist, C. J. A. *et al.* New and continuing developments at PROSITE. *Nucl. Acids Res.* **41**, D344–D347. https://doi.org/10.1093/nar/gks1067 (2013).

17. Jumper, J. *et al.* Highly accurate protein structure prediction with *AlphaFold*. *Nature* **596**, 583–589. https://doi.org/10.1038/s41586-021-03819-2 (2021).

18. Diederichs, K. Crystallographic data and model quality. *Methods Mol. Biol.* **1320**, 147–173. https://doi.org/10.1007/978-1-4939-2763-0_10 (2016).

19. Winter, G. *et al.* DIALS: Implementation and evaluation of a new integration package. *Acta Crystallogr. Sect. D* **74**, 85–97. https://doi.org/10.1107/S2059798317017235 (2018).

20. Powell, H. R., Battye, T. G. G., Kontogiannis, L., Johnson, O. & Leslie, A. G. W. Integrating macromolecular X-ray diffraction data with the graphical user interface *iMosflm*. *Nat. Protoc.* **12**, 1310–1325. https://doi.org/10.1038/nprot.2017.037 (2017).

21. Tickle, I. J. *et al.* STARANISO (Global Phasing Ltd., 2018).

22. Arnold, E. *et al.* (eds) *International Tables for Crystallography. Volume F*: *Crystallography of Biological Macromolecules* 64–74 (Wiley, 2012).

23. Goulas, T. *et al.* The pCri System: A vector collection for recombinant protein expression and purification. *PLoS ONE* **9**, e112643. https://doi.org/10.1371/journal.pone.0112643 (2014).

24. Matthews, B. W. Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497 (1968).

25. Kantardjieff, K. A. & Rupp, B. Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Prot. Sci.* **12**, 1865–1871. https://doi.org/10.1110/ps.0350503 (2003).

26. Wintjens, R. & Rooman, M. Structural classification of HTH DNA-binding domains and protein-DNA interaction modes. *J. Mol. Biol.* **262**, 294–313. https://doi.org/10.1006/jmbi.1996.0514 (1996).

27. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797. https://doi.org/10.1016/j.jmb.2007.05.022 (2007).

28. Sevilla-Sierra, P., Otting, G. & Wüthrich, K. Determination of the nuclear magnetic resonance structure of the DNA-binding domain of the P22 c2 repressor (1 to 76) in solution and comparison with the DNA-binding domain of the 434 repressor. *J. Mol. Biol.* **235**, 1003–1020. https://doi.org/10.1006/jmbi.1994.1053 (1994).

29. Newman, J. A., Rodrigues, C. & Lewis, R. J. Molecular basis of the activity of SinR protein, the master regulator of biofilm formation in *Bacillus subtilis*. *J. Biol. Chem.* **288**, 10766–10778. https://doi.org/10.1074/jbc.M113.455592 (2013).

30. Milton, M. E. *et al.* The solution structures and interaction of SinR and SinI: Elucidating the mechanism of action of the master regulator switch for biofilm formation in *Bacillus subtilis*. *J. Mol. Biol.* **432**, 343–357. https://doi.org/10.1016/j.jmb.2019.08.019 (2020).

31. Wolberger, C., Dong, Y. C., Ptashne, M. & Harrison, S. C. Structure of a phage 434 Cro/DNA complex. *Nature* **335**, 789–795. https://doi.org/10.1038/335789a0 (1988).

32. Lewis, R. J., Brannigan, J. A., Offen, W. A., Smith, I. & Wilkinson, A. J. An evolutionary link between sporulation and prophage induction in the structure of a repressor:anti-repressor complex. *J. Mol. Biol.* **283**, 907–912. https://doi.org/10.1006/jmbi.1998.2163 (1998).

33. Artimo, P. *et al.* ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* **40**, W597–W603. https://doi.org/10.1093/nar/gks400 (2012).

34. Flot, D. *et al.* The ID23-2 structural biology microfocus beamline at the ESRF. *J. Synchrotron. Radiat.* **17**, 107–118. https://doi.org/10.1107/S0909049509041168 (2010).

35. van Zundert, G. C. P., Moriarty, N. W., Sobolev, O. V., Adams, P. D. & Borrelli, K. W. Macromolecular refinement of X-ray and cryoelectron microscopy structures with Phenix/OPLS3e for improved structure and ligand quality. *Structure* **29**, 913–921. https://doi.org/10.1016/j.str.2021.03.011 (2021).

36. Hough, M. A. & Wilson, K. S. From crystal to structure with CCP4. *Acta Crystallogr. sect. D* **74**, 67–67. https://doi.org/10.1107/S2059798317017557 (2018).

37. Remacle, F. (ed.) *CCP4 Newsletter on Protein Crystallography* Vol. 43, 27–35 (Daresbury Laboratory, 2005).

38. Evans, P. R. An introduction to data reduction: Space-group determination, scaling and intensity statistics. *Acta Crystallogr. Sect. D* **67**, 282–292. https://doi.org/10.1107/S090744491003982X (2011).

39. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674. https://doi.org/10.1107/S0021889807021206 (2007).

40. Terwilliger, T. C. *et al.* Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. Sect. D* **64**, 61–69. https://doi.org/10.1107/S090744490705024X (2008).

41. Casañal, A., Lohkamp, B. & Emsley, P. Current developments in *Coot* for macromolecular model building of electron cryo-microscopy and crystallographic data. *Protein Sci.* **29**, 1069–1078. https://doi.org/10.1002/pro.3791 (2020).
42. Holm, L. DALI and the persistence of protein shape. *Protein Sci.* **29**, 128–140. https://doi.org/10.1002/pro.3749 (2020).
43. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. Sect. D* **60**, 2256–2268. https://doi.org/10.1107/S0907444904026460 (2004).
44. Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25. https://doi.org/10.1002/pro.3235 (2018).
45. Rumpel, S. *et al.* Structure and DNA-binding properties of the cytolysin regulator CylR2 from *Enterococcus faecalis*. *EMBO J.* **23**, 3632–3642. https://doi.org/10.1038/sj.emboj.7600367 (2004).
46. Lu, H. *et al.* Structure and DNA damage-dependent derepression mechanism for the XRE family member DG-DdrO. *Nucleic Acids Res.* **47**, 9925–9933. https://doi.org/10.1093/nar/gkz720 (2019).

Acknowledgements

Author contributions

F.X.G.R. conceived and supervised the project; R.C. produced, purified, and crystallised the protein under supervision of I.d.D., who also collected diffraction data; F.X.G.R. solved and refined the crystal structure; and F.X.G.R. wrote the manuscript with contributions from all authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-09237-2.

Correspondence and requests for materials should be addressed to F.X.G.-R.

Reprints and permissions information is available at www.nature.com/reprints.

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