

# Structural basis for control of integrative and conjugative element excision and transfer by the oligomeric winged helix–turn–helix protein RdfS

Callum J. Verdonk<sup>1,2,3</sup>, Mark Agostino<sup>2</sup>, Karina Yui Eto<sup>1,2</sup>, Drew A. Hall<sup>2,4,5</sup>, Charles S. Bond<sup>1,6,\*</sup>, Joshua P. Ramsay<sup>2,4,\*</sup>

<sup>1</sup>School of Molecular Sciences, The University of Western Australia, Perth, WA 6009, Australia

<sup>2</sup>Curtin Medical Research Institute, Curtin University, Perth, WA 6102, Australia

<sup>3</sup>Centre for Crop and Disease Management, Curtin University, Perth, WA 6102, Australia

<sup>4</sup>Curtin Medical School, Curtin University, Perth, WA 6102, Australia

<sup>5</sup>Australian National Phenome Centre, Murdoch University, Perth, WA 6150, Australia

<sup>6</sup>Marshall Centre for Infectious Disease Research and Training, The University of Western Australia, Perth, WA 6009, Australia

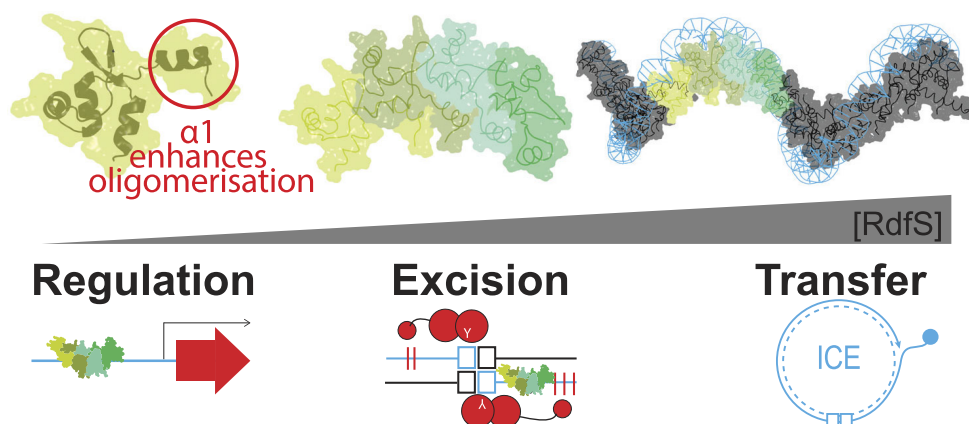
\*To whom correspondence should be addressed. Email: [charles.bond@uwa.edu.au](mailto:charles.bond@uwa.edu.au)

Correspondence may also be addressed to Joshua P. Ramsay. Email: [josh.ramsay@curtin.edu.au](mailto:josh.ramsay@curtin.edu.au)

## Abstract

Winged helix–turn–helix (wHTH) proteins are diverse DNA-binding proteins that often oligomerize on DNA and participate in DNA recombination and transcriptional regulation. wHTH recombination directionality factors (RDFs) associated with tyrosine recombinases, stimulate excision of prophage and integrative and conjugative elements (ICEs). RdfS is required for excision and conjugation of the *Mesorhizobium japonicum* R7A ICE, ICEM/Sym<sup>R7A</sup>, which carries genes for nitrogen-fixing symbiosis. We show RdfS binds to DNA regions within the IntS attachment site (*attP*) and within the *rdfS* promoter, enabling RdfS to coordinate *rdfS/intS* expression and stimulate RdfS/IntS-mediated ICEM/Sym<sup>R7A</sup> excision. Several RdfS DNA-binding sites were identified. However, no consensus motif was apparent and no individual nucleotide substitutions in *attP* prevented RdfS binding. RdfS forms extensive helical filaments in crystals, with subunits contacting via a novel  $\alpha$ 1-helix absent in other wHTH-RDFs. RdfS oligomerized in solution in the absence of DNA. Molecular dynamics simulations supported a role for the  $\alpha$ 1-helix in oligomerization and compaction of nucleoprotein complexes. Removal of RdfS- $\alpha$ 1 did not eliminate DNA-binding *in vitro* but reduced oligomerization and abolished RdfS-mediated ICEM/Sym<sup>R7A</sup> excision and conjugative transfer. We propose the novel RdfS- $\alpha$ 1 mediated oligomerization enables RdfS to specifically recognize larger DNA regions with low primary sequence conservation through an indirect readout mechanism.

## Graphical abstract



## Introduction

Integrative and conjugative elements (ICEs) are chromosomally integrating mobile genetic elements that often carry genes with selective advantages to their bacterial hosts, such as those necessary for biofilm formation, antibiotic and heavy metal

resistances, virulence and pathogenicity, iron acquisition, and symbiotic nitrogen fixation [1–10]. Genome analyses indicate that ICEs are the most abundant type of conjugative element present in prokaryote genomes [11, 12]. ICEs reside integrated within the chromosomes of their hosts but can excise from the

Received: February 10, 2025. Revised: March 12, 2025. Editorial Decision: March 14, 2025. Accepted: March 20, 2025

© The Author(s) 2025. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

chromosome to form a transient circular element that is then able to transfer a single-stranded DNA copy through conjugation. Therefore, ICE excision and circularization is a prerequisite first step in ICE transfer. For most ICEs, it is unclear how the initiation of these distinct DNA recombination and conjugation processes are coordinated at a molecular level [13, 14].

Much like bacteriophage integration and excision, recombination between the ICE and bacterial chromosome is usually site-specific and catalyzed by a tyrosine or serine recombinase, also referred to as an integrase (Int). Integrases, together with specific DNA regions called attachment (*att*) sites, form nucleoprotein complexes that orchestrate DNA strand exchanges, facilitating both integrative and excisive recombination [15, 16]. Attachment sites for integrases of the tyrosine recombinase family contain a central core sequence that acts as the DNA substrate for strand cleavage and exchange. Recombination is catalyzed by the C-terminal recombinase domain of the integrase. During ICE integration, an attachment site *attP* present on the circular excised ICE is recombined with a corresponding chromosomal *attB* site. The *attP* region also contains specific ‘arm’ (also known as ‘P’) sequences flanking the recombined ‘core’ region. These arm sites interact with the N-terminal DNA-binding domain of the integrase and support the three-dimensional organization of the nucleoprotein complex during integration and excision reactions [17, 18]. Following ICE integration, recombined *attP* and *attB* sites produce chimeric *attL* and *attR* sites, each containing a copy of the core sequence and some of the arm sites derived from *attP*. The resulting recombined core sequences within *attL* and *attR* form direct DNA sequence repeats flanking the integrated ICE [15, 16, 19].

While integrase enzymes can catalyze both integrative and excisive recombination reactions, for most tyrosine recombinases, the formation of the integration products *attL* and *attR* is favoured. Excisive recombination often requires a recombination directionality factor (RDF) protein, also known as an excisionase. Tyrosine-recombinase RDFs are typically small positively charged DNA-binding proteins that bind DNA attachment sites and, in some cases, interact with integrase proteins directly [20–24]. While RDFs are extremely diverse at the primary amino-acid sequence level, they are frequently winged helix–turn–helix (wHTH)-domain proteins [20, 25]. wHTH proteins often also have roles as transcriptional regulators [26–33] and some are regarded to be nucleoid-associated proteins [34–38] that may perform roles analogous to eukaryotic histones; binding, bending, wrapping, and/or bridging along DNA to form large nucleoid complexes [39, 40].

The symbiosis ICE of *Mesorhizobium japonicum* R7A, ICEM/Sym<sup>R7A</sup>, is a 502-kb ICE carrying genes necessary for nitrogen-fixation and symbiosis with *Lotus* spp. ICEM/Sym<sup>R7A</sup> encodes genes for root nodule formation, nitrogen fixation, metabolism, and protein secretion [6, 41, 42]. Integration and excision of ICEM/Sym<sup>R7A</sup> requires the tyrosine recombinase (integrase) IntS, which is encoded by the *intS* gene located immediately downstream of the *attL* site on the integrated ICEM/Sym<sup>R7A</sup> element. Experiments using a constructed ‘mini-ICE’ carrying only the *attP* region or the *attP-intS* region (on a plasmid unable to replicate in *Mesorhizobium*) reveal *attP* and *intS* are the only ICEM/Sym<sup>R7A</sup>-encoded factors required for efficient integration. The 283-bp *attP* sequence contains five IntS P/arm-type sequence motifs (consen-

sus TGKTGGTATC), and all five P/arm-type sites are required for efficient integration [43].

Most RDF genes are positioned nearby their cognate integrase genes. However, the ICEM/Sym<sup>R7A</sup> RDF gene *rdfS* is unique in that it is positioned as the first gene in an operon containing conjugation genes *traF-msi107* and is upstream of the origin of transfer (*oriT*) site and the conjugative relaxase gene *rlxS* [19]. This genetic organization of *rdfS* is conserved on numerous ICEs and plasmids found throughout the proteobacteria [14], suggesting the stimulation of ICEM/Sym<sup>R7A</sup> excision has become intertwined with the activation of conjugation. Interestingly, deletion of *rdfS* not only abolishes ICEM/Sym<sup>R7A</sup> excision, but also abolishes the ability of ICEM/Sym<sup>R7A</sup> to mobilize a plasmid carrying a cloned copy of the ICEM/Sym<sup>R7A</sup> *oriT* [43], suggesting RdfS is also required for activation of conjugation.

Introduction of a plasmid overexpressing *rdfS* is lethal in *M. japonicum* R7A and only cells that have lost ICEM/Sym<sup>R7A</sup>, named R7ANS, survive [19]. The R7ANS strain contains a repaired *attB* site resembling a naïve host capable of receiving ICEM/Sym<sup>R7A</sup>. Introduction of the *rdfS* gene expressed on a plasmid stimulated complete excision and loss of the *attP-intS* mini-ICE from the cell population. Introduction of a similar plasmid over-expressing *intS* only moderately increases excision frequency (measured by *attP* abundance) and loss of the mini-ICE (*attB* abundance), suggesting both RdfS and IntS together are required to strongly stimulate IntS-mediated excision [44, 45].

Several ICEs related to ICEM/Sym<sup>R7A</sup> (referred to as the ICESym family [14]) exist in a tripartite configuration. These tripartite ICEs are present as three separate DNA sections in their host chromosomes. Tripartite ICEs assemble into a single circular element prior to conjugative transfer and then separate back into three parts integrated within the chromosomes of recipients [46]. The archetypical tripartite ICE, ICEMcSym<sup>1271</sup> from *Mesorhizobium ciceri* WSM1271, encodes three site-specific tyrosine recombinases—IntS, IntG, and IntM. During ICEMcSym<sup>1271</sup> integration, the circularized tripartite ICE, which carries three distinct *attP* sites, integrates into the chromosome at one of the three different *attB* sites and then recombines twice more with the chromosome to separate the ICE into three sections [46, 47]. Tripartite ICE excision is the reverse of this process and requires an orthologue of RdfS, RdfS<sub>1271</sub>, which shares 85% amino-acid identity with RdfS. ICEMcSym<sup>1271</sup> encodes two additional RDF genes *rdfG* and *rdfM*, located adjacent to their cognate integrase genes *intG* and *intM*. Expression of RdfS<sub>1271</sub> activates transcription from both the *rdfG* and *rdfM* promoters [44], thus coordinating all three excision reactions and formation of a single circular excised ICEMcSym<sup>1271</sup> prior to transfer.

Other structurally characterized wHTH proteins have been shown to oligomerize in a head-to-tail fashion along DNA, forming variable-length nucleoprotein structures with pleiotropic roles on transcription and recombination. In this work, we interrogate the previously solved RdfS crystal structure [45] to gain insight into the structural basis for the various functions of RdfS. RdfS oligomerized in solution even in the absence of DNA, and in crystals, RdfS oligomers have been shown to form an infinite superhelical polymer [45]. Compared to other wHTH proteins, RdfS carries an additional N-terminal  $\alpha$  helix positioned at the core of the superhelical axis of RdfS polymers within crystals. DNA-binding as-

says confirmed that RdfS specifically bound DNA sites within *attP* and within the promoter regions of *rdfS* and *rdfG*. However, no obvious DNA-binding sequence motif was identified within these regions and no individual nucleotides within the *attP* site abolished binding. Removal of the unique N-terminal  $\alpha$ -helix of RdfS diminished but was not essential for DNA binding *in vitro*; however, its removal abolished the ability of RdfS activate ICEM/Sym<sup>R7A</sup> excision and conjugative transfer. We propose that the unique N-terminal  $\alpha$ -helix enhances RdfS oligomerization *in vivo* and that variable length RdfS oligomers specifically recognize diverse DNA sites on ICEs through an indirect readout DNA recognition mechanism.

## Materials and methods

### Microbiological techniques

Bacteria were cultured as previously described [19, 43, 48, 49] and further details can be found in the supplementary data. A list of strains and plasmids used in this study can be found in [Supplementary Table S1](#). Oligonucleotides and synthetic DNA are listed in [Supplementary Table S2](#). DNA extraction, plasmid construction, and  $\beta$ -galactosidase assays were carried out as previously described [49–52], and further information can be found in the supplementary data. ICEM/Sym<sup>R7A</sup> conjugation assays were carried out as described previously [46], with the exception that the R7ANS recipient strain carrying pPR3G was used in conjugation experiments and the selection medium contained gentamicin at 25  $\mu\text{g ml}^{-1}$ .

### DNA-binding assays

Electrophoretic mobility shift assays (EMSAs) were performed as described previously [49, 53]. DNA containing the ICEM/Sym<sup>R7A</sup> *attP* and *PrdfS* were amplified by polymerase chain reaction (PCR) using IRDye800-labelled DNA oligonucleotides. Binding reactions were carried out using labelled DNA concentrations of 10 nM as described previously [49], and detailed in the supplementary data. Surface plasmon resonance (SPR)-based DNA-footprinting assays were performed using the Re-Useable DNA-Capture Technique (ReDCaT) [54] on the Biacore T200 (GE Healthcare) as described previously [49, 53] with protein concentrations at 1  $\mu\text{M}$ . The oligonucleotide array for the *attP*, *PrdfS*, and *oriT* DNA regions were constructed using the Perl script *poop.pl* [54].

### Expression and protein purification

Purification of RdfS is described elsewhere [45]. 6HRdfS<sub>1271</sub> and RdfS<sub>13–89</sub> were purified using the same methods. All protein purifications were carried out on the ÄKTA pure chromatography system (GE Healthcare), using a 5 ml HisTrap column (Cytiva) for immobilized metal affinity chromatography and a HiLoad 16/600 Superdex 200 (GE Healthcare) for size-exclusion chromatography.

### Gel-filtration chromatography

Analytical gel-filtration was performed with a Superdex 75 10/300 GL (Cytiva) column equilibrated with size exclusion chromatography (SEC) buffer (50 mM Tris-HCl, 300 mM NaCl, 5% (*v/v*) glycerol; pH 7.4) on an ÄKTA pure chromatography system (GE Healthcare) at 4°C with a flow rate of 0.5 ml min<sup>-1</sup>. Each protein was concentrated to ~7 mg ml<sup>-1</sup> and injected in aliquots of 50  $\mu\text{l}$ . Wavelengths of 230,

260, and 280 nm were used to visualize the interactions. Further details are available in the supplementary data.

### Small-angle X-ray scattering

Small-angle X-ray scattering (SAXS) data were collected on the SAXS/WAXS beamline at the Australian Synchrotron [55] with continuous data collection using a PILATUS 1M detector [56], using methods similar to those previously described [45]. Collection and processing parameters are shown in [Supplementary Table S3](#). Further methodology can be found in the supplementary data.

### Molecular modelling and simulation

A model of RdfS binding to DNA was generated by adapting homology modelling methods previously described for generating DNA-protein complexes [57]. The model was generated through aligning monomers of RdfS (PDB 8DGL) to the complex of the *Streptomyces* protein BldC with DNA (PDB 6AMA) [36]. Molecular dynamics simulations were performed using GROMACS 2020.3 [58] patched with PLUMED 2.6.1 [59]. Full details of molecular modelling and simulations are described in the supplementary data.

### Quantitative PCR

Quantitative PCR assays to determine ICEM/Sym<sup>R7A</sup> excision rates were performed as previously described [19, 49]. Details of culture growth and primer efficiencies can be found in the supplementary data.

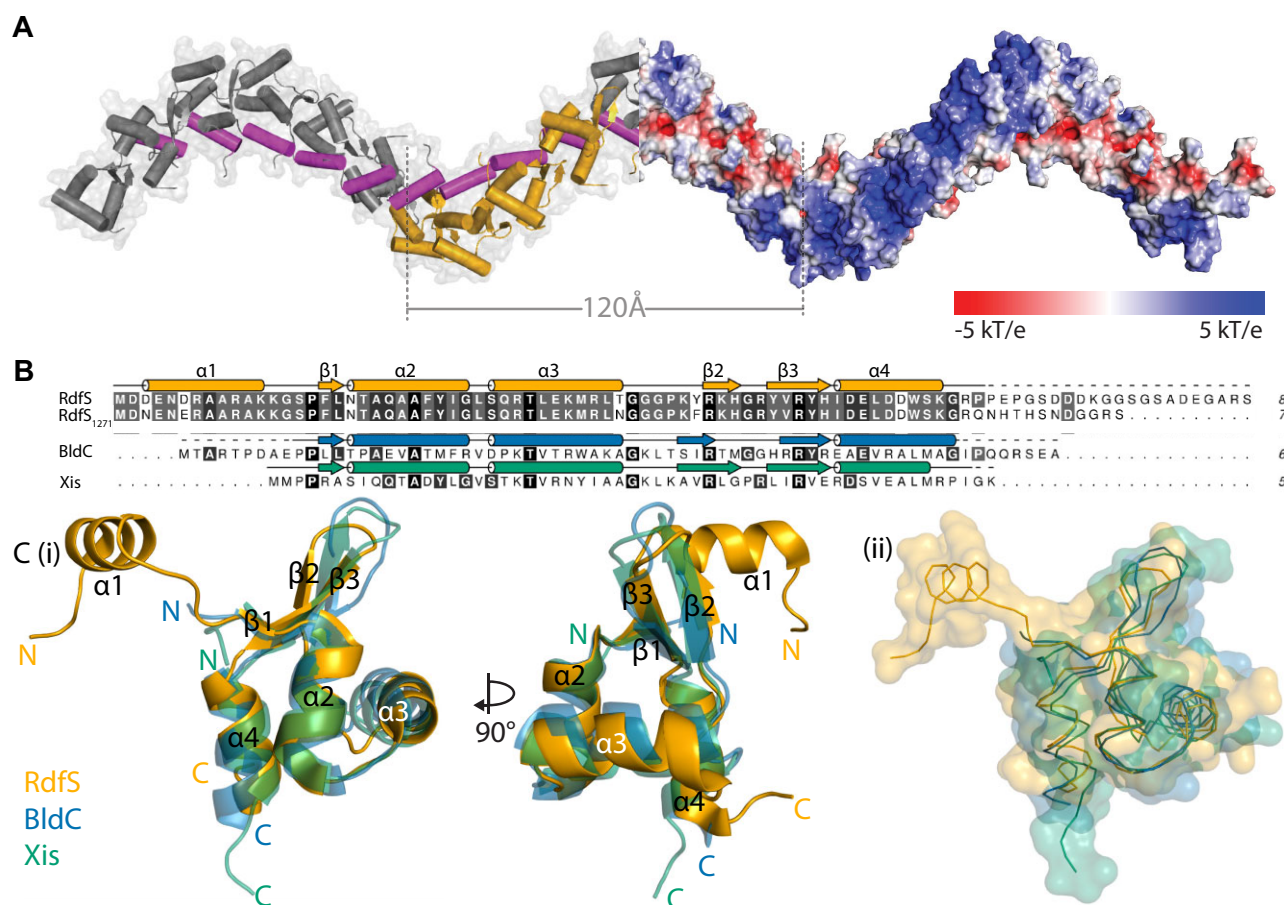
## Results

### RdfS is an oligomeric wHTH-domain with a unique N-terminal $\alpha$ -helix

We recently solved the X-ray crystal structure for RdfS, revealing it forms a head-to-tail oligomer containing four RdfS molecules within the asymmetric unit (ASU), arranged with eight-fold helical symmetry within the crystal [45] (PDB entry 8DGL; Fig. 1A). The infinite head-to-tail oligomerization of RdfS protomers results in a superhelical protein filament extending throughout the crystal. Each of the four independent molecules observed share an effectively identical structure of four  $\alpha$ -helices and a single three-stranded  $\beta$ -sheet:  $\alpha$ 1 (RdfS residues 3–12),  $\beta$ 1 (17–18),  $\alpha$ 2 (19–28),  $\alpha$ 3 (30–40),  $\beta$ 2 (47–49),  $\beta$ 3 (52–56), and  $\alpha$ 4 (57–65) (Fig. 1B). In these crystals the C-terminal 21 residues (24%) of RdfS are completely disordered. The primary amino-acid sequence of RdfS (residues 17–66) is most similar to the group 17 HTH domain, which includes several other excisionase proteins (InterPro IPR041657). Alignments of the RdfS structure with Mycobacterium virus Pukovnik excisionase (RDF) Xis or the MerR-family *Streptomyces* BldC protein (Fig. 1C) exhibit  $\text{C}\alpha$  root mean square deviations (RMSD) of <2 Å.

The superhelical RdfS filament is stabilized by two protein-protein interfaces. The first interface involves the unique RdfS N-terminal  $\alpha$ -helix ( $\alpha$ 1) from Asp3-Lys12, which sits in a hydrophobic pocket within the adjacent RdfS protomer (Fig. 2A). As far as the authors are aware, the  $\alpha$ 1 helix is absent from all other structurally characterized wHTH-domain proteins. The  $\alpha$ 1 charged sidechains from Glu4 and Arg7 form salt-bridges with the adjacent protomers' charged sidechains His56 and Asp58, respectively. Hydrogen bonds are

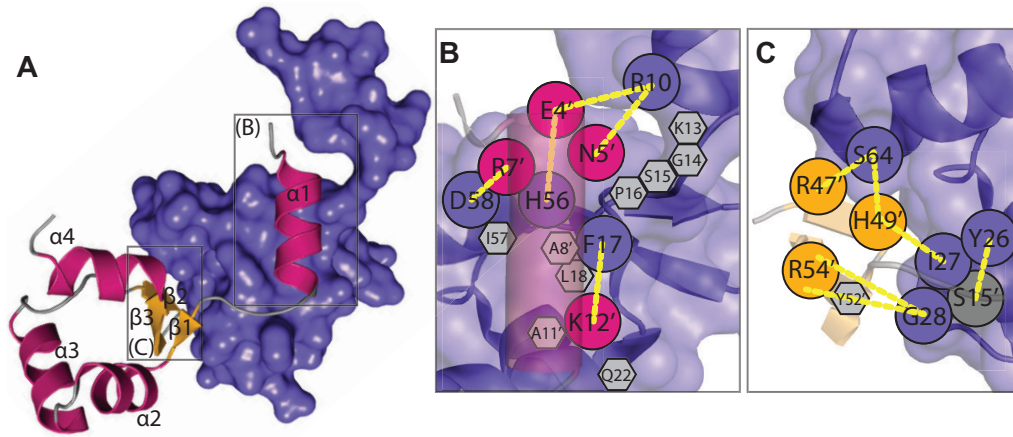




**Figure 1.** X-ray crystal structure of RdfS and comparisons to other wHTH-containing proteins. **(A)** Extended RdfS crystal structure revealing superhelical filaments. Electrostatic surface representation of 24 RdfS molecules is shown in red and blue at the right half of the helix, representing electronegative and electropositive regions, respectively (shown as a  $-5$  to  $5$  kBT/ec gradient). The RdfS asymmetric unit of four molecules shown is depicted in cartoon format in yellow. The central  $\alpha 1$  helix is shown in magenta throughout the length of one full turn of the helix (eight RdfS molecules) spanning  $\sim 120$  Å. **(B)** Amino acid alignment of ICEMISym<sup>R7A</sup> RdfS and closely related *Mesorhizobium* RdfS homologue from *M. ciceri* WSM1271 (RdfS<sub>1271</sub>). Secondary structural representation shown above based on the X-ray crystal of R7A RdfS. Other HTH 17 grouped proteins included are *Streptomyces venezuelae* BldC [36] and Mycobacterium virus Pukovnik Xis [60] with their relative secondary structure features shown above each sequence. **(C)** **(i)** Ribbon model of RdfS (PDB 8DGL), BldC (PDB 6AMK), and Xis (PDB 4J2N) aligned to show similar features with each N- and C-terminus marked, and each location of the  $\alpha$ -helices and  $\beta$ -strands for RdfS labelled to match panel (B). **(ii)** Surface representation of RdfS, BldC, and Xis, with ribbon-backbone shown to highlight overall similarity between each structure [in same orientation as panel (C) (i)]. The RMSD between the wHTH domains of RdfS and BldC, and RdfS and Xis, are  $1.66$  Å (46 atoms) and  $1.12$  Å (39 atoms), respectively.

also formed between the  $\alpha 1$  Asn5 and Lys13 sidechains interacting with the backbones of Phe17 and Arg10 from the neighbouring protomer. The  $\alpha 1$ -mediated interaction between molecules is primarily stabilized by the nine hydrophobic residues which span the surface of  $\alpha 1$  and the neighbouring RdfS protomer (Fig. 2B). The overall surface area of  $\alpha 1$  is  $1286$  Å<sup>2</sup> and with an interface area of  $463$  Å<sup>2</sup> with the adjacent RdfS molecule it is the major source of interaction area between RdfS molecules in the crystal structure. The second interface is between the wing region ( $\beta 2$ – $\beta 3$ ) of one RdfS molecule and the helices  $\alpha 2$  and  $\alpha 4$  of the neighbour; and with fewer hydrophobic residues at this interface, hydrogen bonding is the primary mediator of interface stability. The positively charged side-chains from strands  $\beta 2$  (Arg47 and His49) and  $\beta 3$  (Arg54) of RdfS interact with the backbone oxygens of  $\alpha 2$  (Gly28 and Ile27) and  $\alpha 4$  (Ser64) of the neighbouring molecule (Fig. 2C). This wing– $\alpha 2/\alpha 4$  interaction has a lower interface area of  $\sim 400$  Å<sup>2</sup> compared with the  $\alpha 1$ –RdfS interactions, suggesting that it may contribute more weakly to association.

Typically, wHTH proteins contact both the DNA major groove through their recognition helix and the DNA minor groove with their  $\beta$ -sheet ‘wing’. The positions of the expected major and minor DNA groove binding regions of the wHTH motif and electrostatic potential along the RdfS polymer are consistent with RdfS being capable of binding long stretches of double-stranded DNA (dsDNA) as an oligomer. The tentative recognition helix  $\alpha 3$  contains exposed positively charged basic residues Arg32, Lys36, and Arg38 and exposed acidic residues Glu35 and Gln31, which may stabilize the protein–DNA interface. The arrangement of basic amino acids on the external surface results in a strongly positive electrostatic potential (Fig. 1A) and a continuous DNA-binding interface running the length of the superhelical protein filament. Superimposing RdfS monomers with a selection of crystal structures of wHTH:DNA complexes (Supplementary Fig. S1) revealed the orientation of the RdfS  $\alpha 3$  and  $\beta 2$ – $\beta 3$  wing in the crystal structure resembles the orientation of DNA-binding components of other members of the wHTH family.



**Figure 2.** The interface between adjacent RdfS molecules in the crystal. **(A)** Ribbon diagram of RdfS molecule interacting with adjacent molecule as surface representation (purple). Ribbon colour scheme shown is strand; yellow, helix; magenta, and loop; grey. Schematic showing: **(B)** Close-up of RdfS N-terminus  $\alpha$ -helix ( $\alpha 1$ ) (transparent magenta cylinder) interacting residues. **(C)** RdfS  $\beta 1$ – $\beta 2$ – $\beta 3$  'wing' region interacting with an adjacent RdfS protomer. Diagram residues coloured as in panel (A), hydrogen bonds which stabilize the interface shown as yellow. Hydrophobic contact residues shown as grey hexagons.

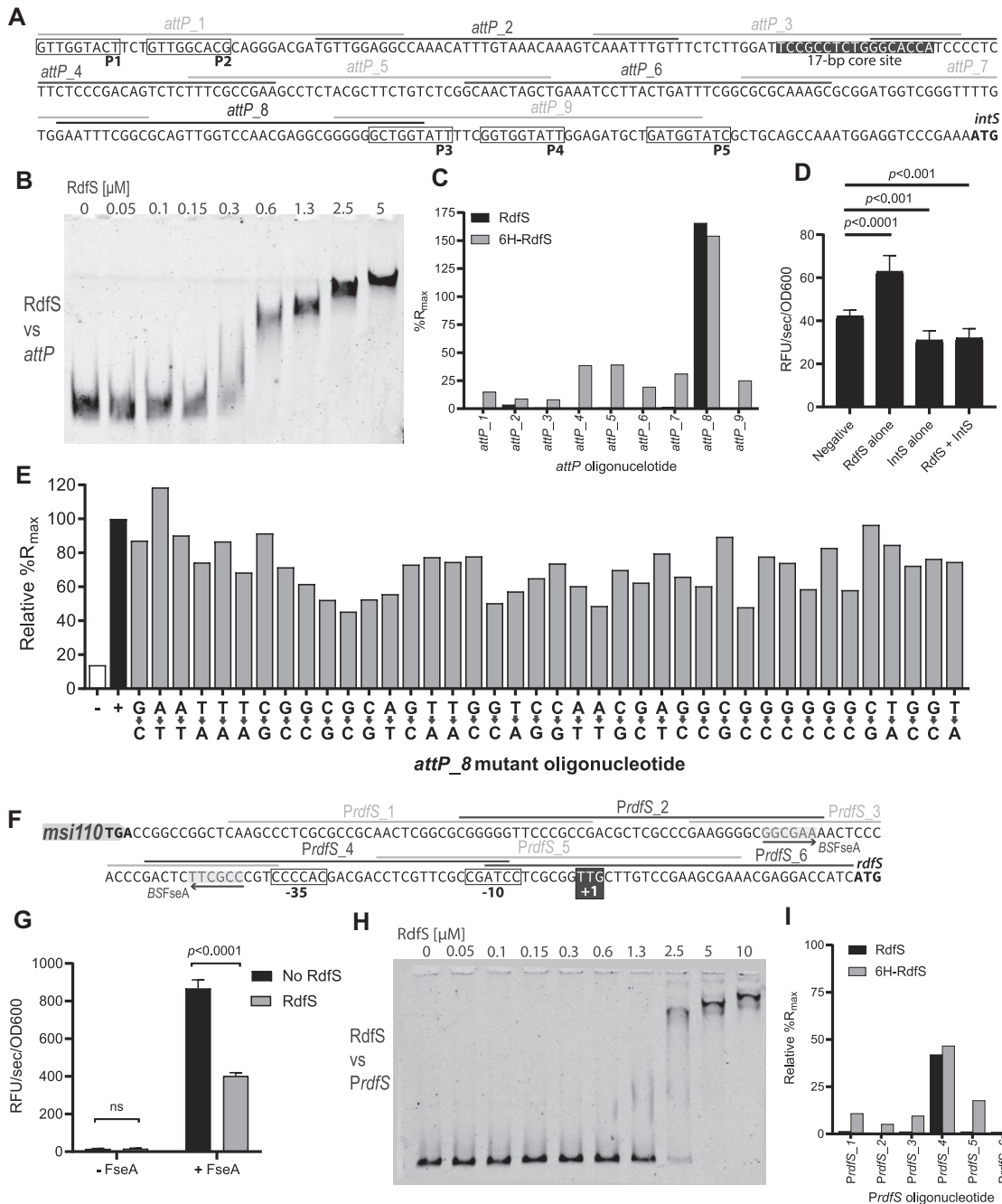
### RdfS binds the *attP* site upstream of *intS* and upstream of promoters *PrdfS* and *PrdfG*

Characterized RDFs have been shown to bind specific DNA regions within their cognate *att* sites and promote excise recombination catalyzed by their cognate integrase. Consistent with RdfS playing a similar role in ICEM/Sym<sup>R7A</sup> excision, overexpression of RdfS stimulates excision and loss of ICEM/Sym<sup>R7A</sup> from cells. Likewise, overexpression of RdfS in the presence of a mini-ICE plasmid carrying only the *attP*–*intS* region also results in loss of the mini-ICE, indicating that RdfS likely binds the *att* regions of ICEM/Sym<sup>R7A</sup> to stimulate IntS-mediated excision. To test whether RdfS could bind the *attP* region present on the circularized ICEM/Sym<sup>R7A</sup>, RdfS was purified as an N-terminal hexahistidine fusion protein (6H-RdfS) and the 6H tag was removed following purification. EMSAs were carried out with the 283-bp *attP* region (Fig. 3A). EMSAs revealed RdfS retarded the migration of 10 nM *attP* when supplied at concentrations of 0.3–0.6  $\mu$ M. The *attP* fragment was increasingly retarded with increased RdfS concentrations, suggesting that additional RdfS molecules were potentially oligomerizing on the RdfS–DNA complex (Fig. 3B). While the ratio of RdfS:DNA required to retard *attP* was relatively high, RdfS was unable to shift a 100-bp control DNA region (100-bp of the downstream *intS* gene) even with 10  $\mu$ M of RdfS (Supplementary Fig. S2), confirming that RdfS binding to the *attP* region was specific. To refine the location of RdfS binding to *attP*, SPR-based DNA-footprinting assays were carried out. A tiled array of 40-bp double-stranded oligonucleotides spanning the *attP* region was designed, with each dsDNA oligonucleotide overlapping the sequence of the adjacent oligonucleotide by 10 bp (Fig. 3A). SPR assays with these oligonucleotides revealed that purified RdfS (1  $\mu$ M) induced a strong binding response with a single dsDNA oligonucleotide, *attP*\_8, the sequence of which overlaps the putative IntS arm site 'P3' towards the 3' end of the *attP* region. The SPR binding response (% $R_{\max}$ ) values were 150%–250% of the theoretical  $R_{\max}$  (across multiple experiments), suggesting more than one RdfS molecule was bound to *attP*\_8. Due to the apparent complexity of RdfS–DNA interactions we were unable to determine association/dissociation kinetics from SPR data. Interestingly, when 6H-RdfS was used in SPR experi-

ments, binding responses with *attP*\_8 were unchanged; however, binding to each of the other oligonucleotides increased, suggesting that the 6H tag slightly increased non-specific DNA binding (Fig. 3C).

We attempted to further delineate the RdfS binding region within the *attP*\_8 sequence by carrying out SPR experiments with a 20-bp oligonucleotide array covering the *attP*\_8 sequence, with each oligonucleotide overlapping the adjacent oligonucleotides by 15 bp. However, none of these 20-bp dsDNA regions achieved >45%  $R_{\max}$  observed for *attP*\_8 (Supplementary Fig. S2), suggesting that either the optimal RdfS binding region in *attP*\_8 was larger than 20-bp and/or that the positioning of the oligonucleotide did not capture the entire binding motif. To determine whether there were any critical nucleotides targeted by RdfS within the 40-bp *attP*\_8 oligonucleotide, we created 40 new variants of *attP*\_8 dsDNA oligonucleotide, each with a single-base-pair substitution to its complement nucleotide. Strikingly, no single nucleotide substitution abolished RdfS binding specificity, although some nucleotide changes reduced the relative % $R_{\max}$  to ~50%–60% that of the native sequence (Fig. 3E). Together, these data suggest that RdfS oligomers may bind the *attP* sequence over a relatively large DNA-binding region, and that while the binding to this region is specific, no individual nucleotides in the region are absolutely critical for RdfS binding.

Due to the position of *attP*\_8 sequence directly upstream of *intS*, we wondered whether RdfS and/or IntS might regulate transcription of the *intS* gene. In previous work, a plasmid pFUS2P1-P5 was constructed carrying the *attP* region with the *intS* gene replaced by *lacZ*, thus forming an *attL*-integrated transcriptional fusion carrying the *PintS* promoter [43]. Therefore, we used this strain to report on the expression of *intS* in the presence and absence of *rd*fS. Measured  $\beta$ -galactosidase expression from the *PintS*–*lacZ* fusion was increased ~53% above background levels when *rd*fS was expressed from plasmid pSRKrdfS, indicating RdfS activates transcription of *intS* (Fig. 3D). In contrast,  $\beta$ -galactosidase expression was decreased ~25% below background levels in the presence of the *intS*-expressing plasmid pJJ611, revealing that IntS negatively regulates its own expression. When both pSRKrdfS and pJJ611 were introduced,  $\beta$ -galactosidase



**Figure 3.** RdfS DNA-binding to the ICEM/Sym<sup>R7A</sup> attP and PrdfS. **(A)** Sequence of the minimal ICEM/Sym<sup>R7A</sup> attP DNA region required for integration, highlighting the arrangement of putative IntS-binding P sites P1–P5 (labelled boxes) and central recombination region (including the putative IntS ‘core’ site, white text, black background). Oligonucleotides (40 bp) used in SPR-based DNA-footprinting experiments are indicated by alternating grey and black lines above the attP sequence, labelled attP\_1–attP\_9. **(B)** EMSA using fluorescently labelled attP DNA (10 nM) and two-fold increasing concentrations of RdfS. **(C)** SPR DNA-footprinting for RdfS and 6H-RdfS using individual 40-bp oligonucleotides from the attP region. Oligonucleotide attP\_8 consistently exhibited the largest %R<sub>max</sub> response and was therefore used as a positive control in all subsequent SPR footprinting experiments. **(D)**  $\beta$ -Galactosidase activity in strains carrying a chromosomally integrated *PintS::lacZ* fusion (R7ANS::pFUS2P1-P5). Cultures were tested in the presence of RdfS (pSRKrdfs) and/or IntS (pJJ611), or with empty vectors (pSRKKm and pFAJ1708). **(E)** Analysis of SPR responses for dsDNA attP\_8 oligonucleotides containing single base-pair substitutions. For each nucleotide position a substitution is indicated, where each base was switched for its complementary base on both strands of the dsDNA oligonucleotide. Each substitution was sequentially tested in SPR DNA-footprinting in a single run and the positive control attP\_8 (+) and negative control attP\_2 (–) oligonucleotides (shown as black and white bars) were run every 10th cycle and used to normalize drifting data over the duration of the experiment. **(F)** Sequence of the ICEM/Sym<sup>R7A</sup> PrdfS region. Each oligonucleotide tested in SPR DNA-footprinting are shown as alternating grey and black bars above the sequence (labelled PrdfS\_1–PrdfS\_6). The PrdfS transcriptional activator FseA binding site (BSFseA) [61] is shaded with a grey background and the inverted repeat are shown by arrows. The –35, –10 regions and +1 transcriptional start site for PrdfS estimated from previous 5’ RACE experiments are indicated. **(G)**  $\beta$ -Galactosidase activity of stationary-phase R7ANS cultures containing PrdfS-lacZ (pSDrdfs-lacZ) plasmid constructs in the absence of FseA (– FseA) or containing PrdfS-lacZ with the presence of FseA (pFseArdfs-lacZ: + FseA). Cells additionally contained vectors pSRKKm (No RdfS: black) or pSRKrdfs (RdfS: grey). **(H)** EMSA with RdfS and the PrdfS region shown in panel (F). **(I)** SPR DNA-footprinting of RdfS and 6H-RdfS against PrdfS 40-bp oligonucleotides. Values are shown as relative %R<sub>max</sub> to the attP\_8 control oligonucleotide (i.e. where attP\_8 has a %R<sub>max</sub> of 100%).



expression was similar to the strain carrying the *intS*-plasmid pJJ611 alone, indicating that the negative autoregulation by IntS was dominant over the observed transcriptional activation by RdfS. In summary, the results of these experiments are consistent with a model where RdfS activates transcription from *PintS* and once IntS is produced, IntS negatively autoregulates its own expression.

Artificial overexpression of RdfS from a plasmid is lethal in R7A cells and leads to curing of ICEM/Sym<sup>R7A</sup> from cells [19]. However, we have never observed natural curing of ICEM/Sym<sup>R7A</sup> from cells during culture, suggesting that RdfS might negatively regulate its own transcription from *PrdfS* and prevent *rdfS* overexpression. To explore this, a previously constructed *PrdfS-lacZ* fusion was used to report on *PrdfS* transcription using  $\beta$ -galactosidase assays [50]. Since there is negligible expression from *PrdfS* in the absence of the transcriptional activator FseA, a plasmid containing both an IPTG-inducible copy of *fseA* and a *PrdfS-lacZ* transcriptional fusion was used to estimate transcription from *PrdfS* (Fig. 3G). A second plasmid carrying *rdfS* also under control of the *lac* promoter (pSRKrdfS) was used to express RdfS. As previously demonstrated, strong *PrdfS-lacZ* expression was observed from the pFseArdfS-*lacZ* plasmid because FseA activates transcription from *PrdfS*. However, when RdfS was additionally expressed from plasmid pSRKrdfS in the same cells,  $\beta$ -galactosidase activity was reduced  $\sim 54\%$ , indicating RdfS also binds *PrdfS* and attenuates the transcriptional activation by FseA.

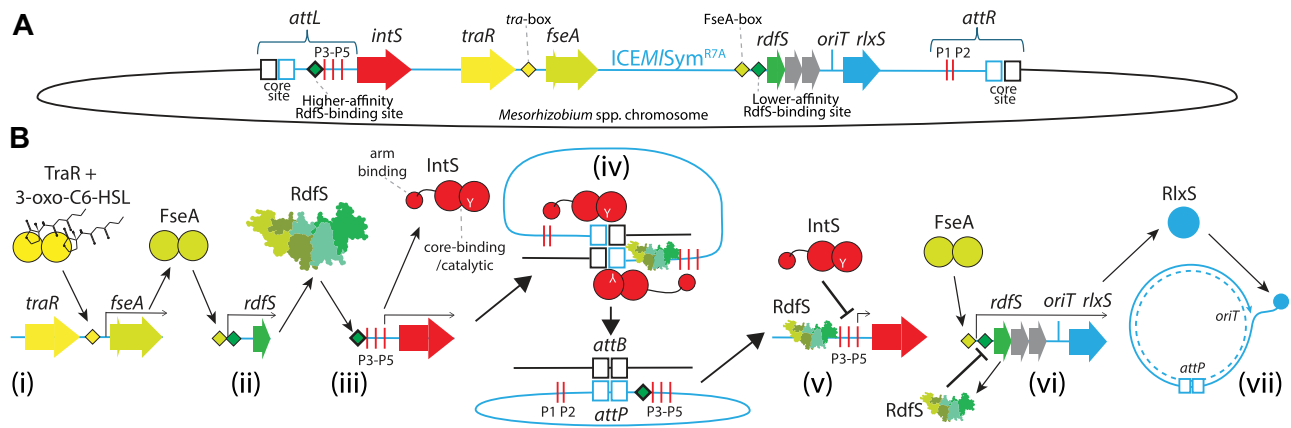
EMSAs were performed with purified RdfS and a 162-bp region containing *PrdfS*. RdfS began to reduce migration of this DNA region when supplied at 1.3  $\mu$ M. This concentration is approximately double the concentration of RdfS required to shift the *attP* region, indicating RdfS has an apparent weaker affinity for its own promoter than *attP* (Fig. 3H). We next again used SPR-footprinting with 40-bp tiled arrays of the *PrdfS* region to identify RdfS binding region(s). A single 40-bp dsDNA oligonucleotide *PrdfS*\_4 was bound by RdfS. In comparison to the 40-bp *attP* region identified previously, the 40-bp *rdfS* binding response was 52% of the theoretical  $R_{\max}$ , which is  $\sim 43\%$  that of the binding response observed for the *attP*\_8 oligonucleotide (Fig. 3I). The sequence of *PrdfS*\_4 overlapped the  $-35$  and  $-10$  region of *PrdfS*, immediately downstream of the FseA binding site, consistent with RdfS repressing transcription by potentially occluding RNA polymerase from *PrdfS* (Fig. 3A). Shorter 20-bp dsDNA oligonucleotides derived from the *PrdfS*\_4 sequence did not respond to RdfS in SPR experiments, again suggesting the optimal RdfS DNA binding site may be larger than 20 bp (Supplementary Fig. S2). Overall, the results of these experiments support a model of ICEM/Sym<sup>R7A</sup> excision in which RdfS first binds and stimulates *intS* transcription and stimulates IntS-mediated excision, after which IntS represses its own transcription from *PintS*. Once RdfS concentration increases above a certain threshold, RdfS binds *PrdfS* and represses its own expression, tentatively explaining why RdfS-dependent growth inhibition and loss of ICEM/Sym<sup>R7A</sup> has not been observed without artificial overexpression of *rdfS* (Fig. 4).

RdfS and RdfS<sub>1271</sub> of the tripartite-ICE ICEM/Sym<sup>1271</sup> share 85% amino-acid identity over their aligned length and otherwise only differ in the length and sequence of their C termini (Fig. 1B). Previously we have demonstrated RdfS<sub>1271</sub> activates transcription from the *PrdfG* and *PrdfM* promot-

ers of ICEM/Sym<sup>1271</sup> [44], stimulated tripartite ICE excision through the sequential activities of integrases IntS<sub>1271</sub>, IntG, and IntM. To test if RdfS<sub>1271</sub> exhibited similar DNA-binding characteristics to RdfS, hexahistidine-tagged RdfS<sub>1271</sub> (6HRdfS<sub>1271</sub>) was expressed and purified from *Escherichia coli*. 6HRdfS<sub>1271</sub> was tested in SPR experiments using the tiled oligonucleotide arrays designed for ICEM/Sym<sup>R7A</sup> *attP* and *PrdfS* regions as shown in Fig. 3A and F. Similar to R7A RdfS, 6HRdfS<sub>1271</sub> bound the ICEM/Sym<sup>R7A</sup> *attP*\_8 and *PrdfS*\_4 oligonucleotides with the highest relative % $R_{\max}$  (Supplementary Fig. S3A), indicating an identical DNA-binding specificity on ICEM/Sym<sup>R7A</sup>, despite RdfS<sub>1271</sub> originating from ICEM/Sym<sup>1271</sup>. Next, tiled oligonucleotide arrays were designed spanning the sequences of ICEM/Sym<sup>1271</sup> promoters *PrdfS*, *PrdfG*, and *PrdfM*. For the ICEM/Sym<sup>1271</sup> *PrdfS*, 6HRdfS<sub>1271</sub> bound a dsDNA oligonucleotide in similar relative position within the *PrdfS*<sub>1271</sub> promoter region as RdfS did with its cognate promoter (offset 5' by 8 bp; Supplementary Fig. S3B) confirming both RdfS and RdfS<sub>1271</sub> bind in similar positions overlapping their own gene promoters. 6HRdfS<sub>1271</sub> bound two distinct regions within the *PrdfG* promoter region (Supplementary Fig. S3C). One of these was centred 170 bp upstream of the *rdfG* start codon, while the other was positioned directly upstream of the *rdfG* start codon, overlapping the putative ribosome-binding site. Only very weak binding was observed for any of the oligonucleotides in the *PrdfM* array (Supplementary Fig. S3D); however, this is consistent with very weak transcriptional activation of the *PrdfM* promoter by RdfS<sub>1271</sub> compared to much stronger activation of the *PrdfG* promoter [44].

### RdfS oligomerizes alone in solution and together with DNA

Most wHTH protein that form head-to-tail oligomers do so only in the presence of their cognate DNA sites [34]. The crystal structure of RdfS indicates that it can, at least in the crystal environment, form extensive oligomers in the absence of any DNA. We carried out analytical gel-filtration experiments to determine the oligomerization states of RdfS in solution. In the absence of DNA, RdfS eluted as an asymmetric peak with a leading edge corresponding to a maximum of  $\sim 40$  kDa (the expected mass of a tetramer) and a long trailing edge suggesting the presence of conformational or compositional heterogeneity (a flexible oligomer, and the presence of smaller oligomers such as trimers or dimers) (Fig. 5A and B). To validate the potential oligomeric properties of RdfS when binding DNA, we performed size-exclusion chromatography-coupled synchrotron SAXS (SEC-SY-SAXS) of 6H-RdfS bound to *attP*\_8. Although there were multiple long trailing peaks for the 6H-RdfS-*attP*\_8 sample in the SEC elution absorbance profiles, there was only one clear peak for the average X-ray scattering intensity profile (Fig. 5C), suggesting a single stable complex was being formed for 6H-RdfS-*attP*\_8. Analysis of the 6H-RdfS-*attP*\_8 complex indicated it shows properties of a monodisperse sample. The normalized Kratky plot indicates a flexible extended particle, with a peak above  $\sim 1.1$  and  $qR_g$  beyond  $\sqrt{3}$  returning towards zero at high  $qR_g$ , indicating elongation [62–64], commensurate with the likely shape of a nucleoprotein complex (Supplementary Fig. S4). SAXS-derived molecular weight estimates using size and shape [65] indicated a 6H-RdfS-*attP*\_8



**Figure 4.** Stepwise induction of ICEM/Sym<sup>R7A</sup> excision and transfer coordinated by IntS and RdfS. **(A)** A simplified depiction of integrated ICEM/Sym<sup>R7A</sup> (light blue) including the *attL* and *attR* regions, the putative arm-type IntS-binding sites P1–P5 and other genes and binding sites involved in coordinating ICEM/Sym<sup>R7A</sup> excision (as labelled). **(B)** A stepwise model of ICEM/Sym<sup>R7A</sup> excision and transfer. **(i)** In transfer-active cells, the quorum sensing transcriptional activator TraR together with 3-oxo-C6-HSL binds a *tra*-box upstream of *fseA* and stimulates transcription (*fseA* is present as a two-part open-reading frame *msi172-msi171* and the FseA protein is translated following a programmed ribosomal frameshift). **(ii)** An FseA dimer binds the FseA box upstream of *PrdfS* and activates transcription of *rdfs* and downstream genes involved in transfer (*traF*, *msi107*, and *rlxS*). **(iii)** RdfS (shown as tetramer in shades of green), either as monomers or potentially as preformed oligomers bind the *attL* region of ICEM/Sym<sup>R7A</sup> upstream of the P3 arm-type IntS binding site. RdfS stimulates transcription of *intS* and production of IntS proteins. **(iv)** IntS proteins bind the core regions and arm-type sites present within *attL* and *attR* and together with RdfS, IntS catalyzes the excision of ICEM/Sym<sup>R7A</sup> through recombination of *attL* and *attR*, producing *attP* and *attB*. **(v)** Once excised, *attP*-bound IntS protein prevents further transcription of *intS*, likely keeping the ratio of IntS:RdfS low, preventing reintegration of ICEM/Sym<sup>R7A</sup>. **(vi)** Once RdfS concentration increases above a certain threshold, RdfS binds the apparent lower-affinity binding site within *PrdfS*, repressing transcription from this promoter and thus preventing RdfS-mediated growth inhibition. **(vii)** Finally, following some unknown signal indicating a mating-pair has formed with a recipient cell, conjugation commences with RlxS nicking the *oriT* and piloting an ssDNA copy of ICEM/Sym<sup>R7A</sup> through the type-IV secretion system to the recipient cell.

mass of 85 655 Da, while concentration-independent estimation with Consensus Bayesian Assessment [66] suggested a complex mass of 83 125 Da (21.55% probable; credibility interval of 75 300–86 950 Da at 91.27% probability). These molecular weight estimates indicate a tetramer of 6H-RdfS bound to the dsDNA 40-mer *attP*<sub>8</sub>; however, they predict a mass slightly larger than the expected molecular weight of 76.5 kDa, corresponding to four copies of 6H-RdfS and *attP*<sub>8</sub>. The slight over-estimation for the molecular mass of the 6H-RdfS-*attP*<sub>8</sub> nucleoprotein complex is likely due to the increased scattering intensities of more electron-dense nucleic acids [67]. Generally adjusting for the difference in electron intensities for nucleic acids [68], the molecular weight estimates from the scattering approximates 71.5 kDa—within 7% of the expected molecular weights (Supplementary Table S3). *Ab initio* bead models from the scattering approximates an envelope shape which superficially matches a generated RdfS–DNA molecular dynamics model (Fig. 5D). While fits into these bead models are poor, this is unsurprising given the absence of the 6H tag and disordered C-termini in the molecular dynamics-generated RdfS–DNA model—which is reflected in the residuals at low and mid-*q* (Supplementary Fig. S4). We concluded from these data that RdfS is binding as a tetramer to the 40-bp *attP*<sub>8</sub> DNA target, suggesting 10-bp DNA per molecule of RdfS, comparable with previously characterized WTH proteins which form multimers on DNA [36, 69–71].

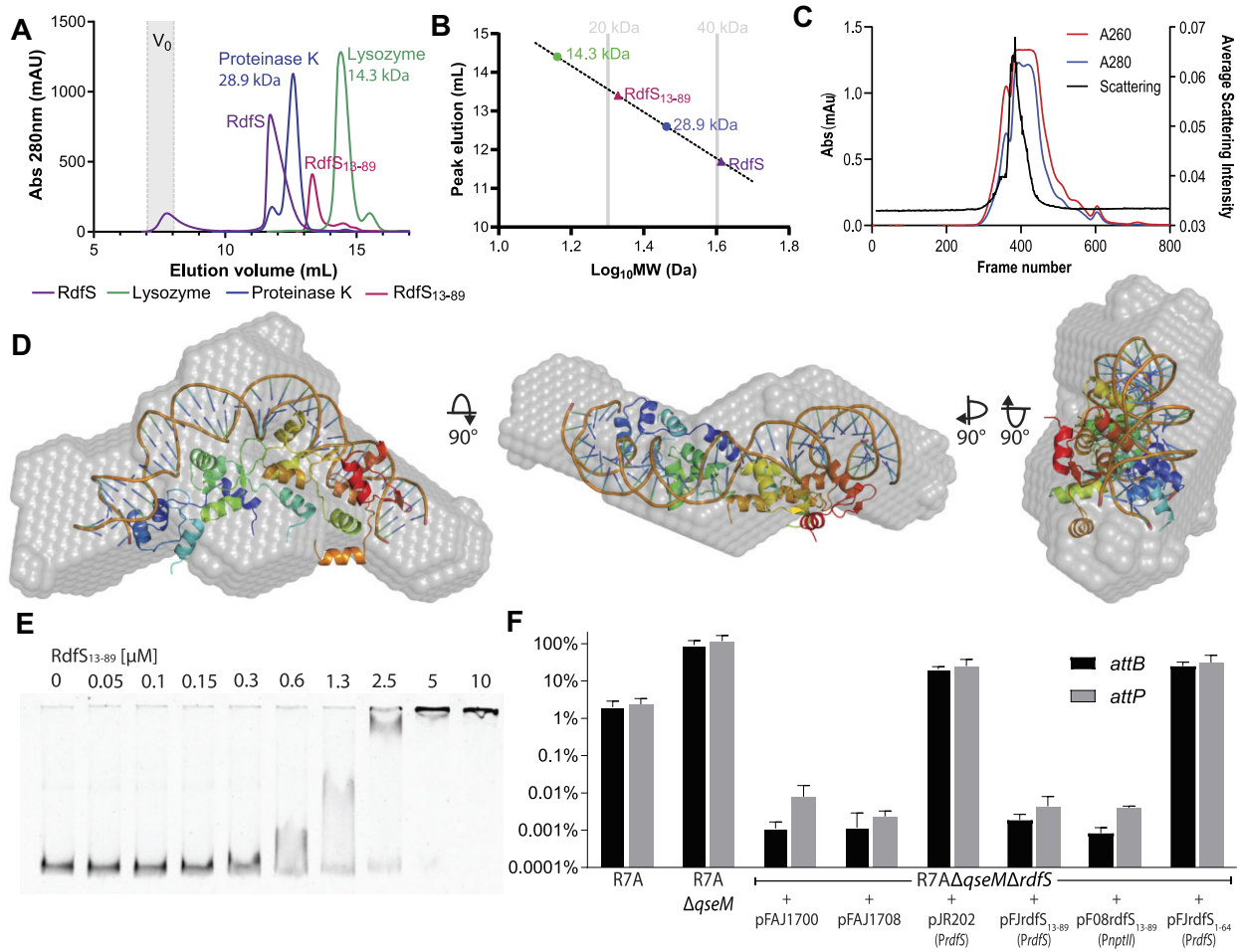
### Molecular dynamics simulations suggest that the N-terminal helix stabilizes RdfS oligomers and compacts the RdfS quaternary structure

Despite exhaustive attempts, we were unable to produce crystals of RdfS complexed with DNA, so we used *in silico* methods to assess how RdfS tetramers might bind

with DNA targets and to probe whether the additional N-terminal  $\alpha$ 1 helix of RdfS might be important for these interactions. Extracting the DNA regions from WTH–DNA nucleoprotein crystal structures (Supplementary Fig. S1) and superimposing them on the finalized RdfS ASU tetramer (Fig. 6A), we observed a strongly curved DNA bend path. To ensure this DNA bending was physically realistic, we first compared it to other structures with curved DNA and noted the bend to have similar parameters to nucleosomal DNA (Fig. 6B). We then performed molecular dynamics simulations on protein-only tetramers of both full length and N-terminally truncated RdfS<sub>13–89</sub> (Supplementary Fig. S5A and B). The RMSDs (Supplementary Fig. S5C and D) over the simulation time courses for each tetramer suggest that RdfS<sub>13–89</sub> yields a more flexible tetramer than the full length RdfS; this is further supported by cluster analysis, which indicates a vastly greater number of clusters for RdfS<sub>13–89</sub> versus RdfS (Supplementary Table S4). The distance between the centres of mass of the capping molecules over the simulation time courses for each model, as well as within the middle structure of the largest cluster (Supplementary Fig. S5E and F), indicate that RdfS<sub>13–89</sub> yields a slightly more compact tetramer compared to the full-length RdfS tetramer.

RdfS–DNA complexes for RdfS and RdfS<sub>13–89</sub> generated by homology modelling (Fig. 6C) were subjected to adiabatic-biased molecular dynamics (ABMD), during which RdfS monomer-monomer interactions were targeted to replicate those observed in RdfS ASU tetramer (see supplementary data for full details). RdfS–DNA complexes showed structural similarity (RMSD  $\sim$ 0.3 nm) to the protein-only RdfS crystal structure, indicating successful targeting of this protein assembly using ABMD (Fig. 6D). Subsequent unbiased simulations following the ABMD simulations suggested the general stability of full-length RdfS in this complex, main-



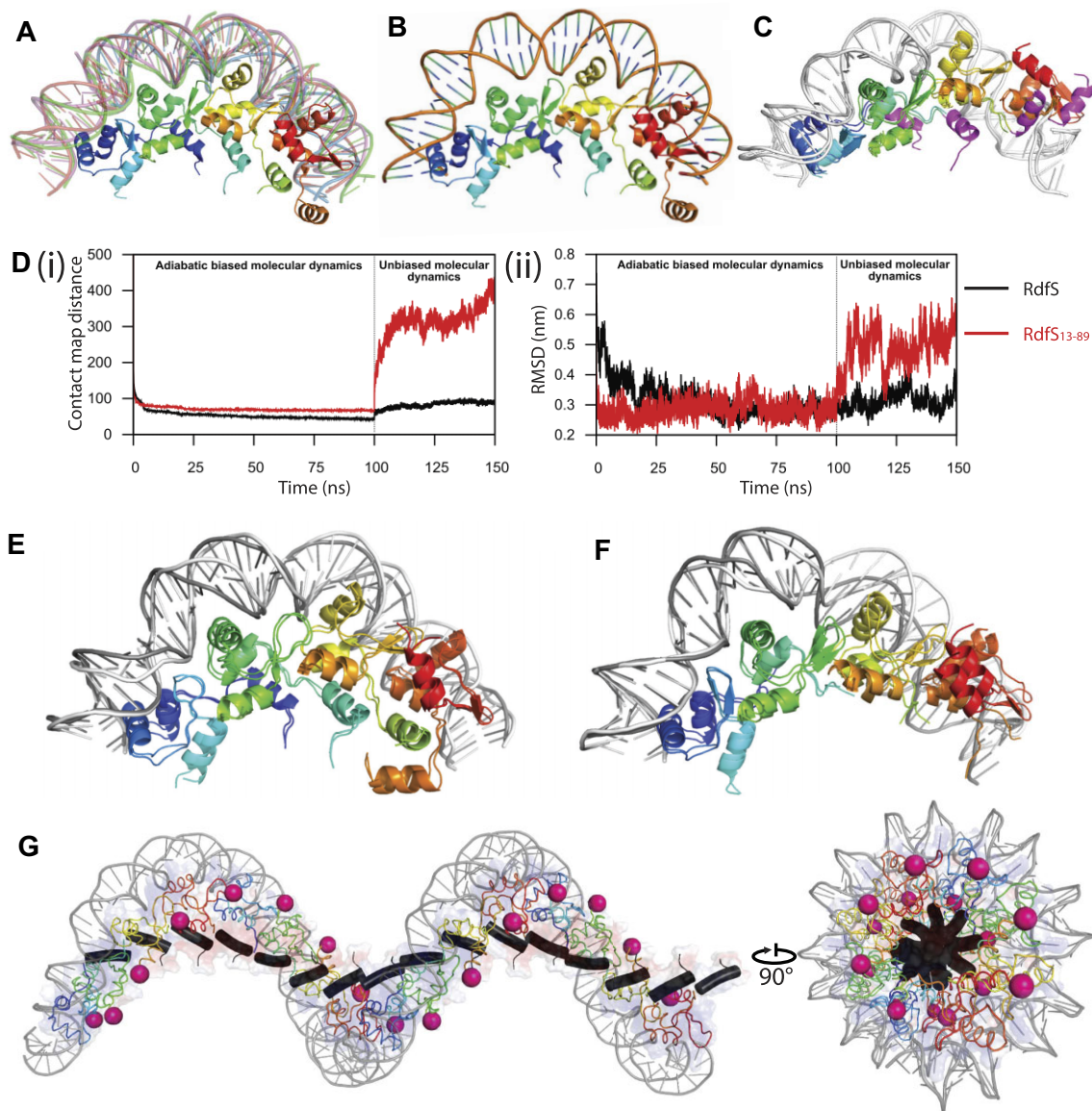


**Figure 5.** RdfS oligomerization, DNA binding and *in vivo* functions are impacted by deletion of  $\alpha 1$ . **(A)** Comparison of analytical size exclusion elution profiles of RdfS and RdfS<sub>13-89</sub>. RdfS elutes larger than 28.9 kDa, indicating likely oligomeric states of tetramers >40 kDa, with some soluble aggregation/large multimer complexes appearing at the void volume (V<sub>0</sub>; shown as grey).  $\alpha 1$ -truncated RdfS (RdfS<sub>13-89</sub>) elutes between 14.3 and 28.9 kDa in size, indicating dimers or trimers in solution (chromatography peak slightly larger than 20 kDa). **(B)** Log<sub>10</sub> plot of SEC data from (A) for estimation of RdfS/RdfS<sub>13-89</sub> complex sizes. **(C)** Ultraviolet elution profile showing absorbance at 260 (red) and 280 (blue) nm, and average intensity of scattering of the 6H-RdfS-*attP*<sub>8</sub> (black). Each frame represents 1 s of exposure. **(D)** Most representative (as determined by DAMAVER) DAMMIF *ab initio* envelope bead model of the 6H-RdfS-*attP*<sub>8</sub> complex, with a representative atomic structure of the RdfS-DNA complex from molecular dynamics simulations unbiased adiabatic model shown within. This atomized model excludes the extended N- and C-termini present in the scattering data and is only shown as a representative of a superficial visualization of a potential nucleoprotein arrangement fit to the *ab initio* envelope. **(E)** EMSA showing RdfS<sub>13-89</sub> retarding the migration of the fluorescently labelled *attP* region. The N-terminal truncated RdfS<sub>13-89</sub> remained capable of binding *attP* DNA with ~50% the affinity of RdfS. Much like the EMSAs with RdfS, migration of *attP* was increasingly retarded with increasing concentrations of RdfS<sub>13-89</sub>, suggesting RdfS<sub>13-89</sub> was still likely capable of oligomerizing on DNA with increasing concentrations. **(F)** The relative abundance (relative to the R7A chromosome copy number) of *attP* and *attB* (present only when ICEM/Sym<sup>R7A</sup> is excised) were measured by quantitative PCR to determine the frequency of ICEM/Sym<sup>R7A</sup> excision. Plasmids expressing RdfS or truncated RdfS variants RdfS<sub>13-89</sub> and RdfS<sub>1-64</sub> were introduced to determine if each variant was functional in activating ICEM/Sym<sup>R7A</sup> excision. The R7AΔ*qseM*Δ*rdS* background was used as it is derepressed at a regulatory level for ICEM/Sym<sup>R7A</sup> excision and conjugation but lacks *rdS* gene required to excise and transfer [52]. The promoter used to drive expression of each of the RdfS variants is shown in brackets under the name of each plasmid.

taining a shape with similar curvature to nucleosomal DNA (Fig. 6E). In contrast, the complex with RdfS<sub>13-89</sub> was conformationally unstable in unbiased simulations, with the distal two RdfS<sub>13-89</sub> molecules dislodging from the DNA (Fig. 6F), suggesting that the  $\alpha 1$  helix is required for stable protein-protein and/or protein-DNA interactions. Extension of this 40-bp DNA model to a continuous RdfS superhelix highlights the possibility that RdfS and DNA can form an extensive polymeric complex (Fig. 6G), where the N-termini assemble along the superhelical axis of the elongated nucleoprotein and the C-termini face outward toward the external surface, with an overall left-handed twist consistent with negative supercoiling.

### The N-terminal helix of RdfS is not essential for DNA-binding *in vitro* but is essential for *in vivo* activation of ICEM/Sym<sup>R7A</sup> excision and conjugative transfer

Noting the importance of contacts involving the N-terminal  $\alpha 1$  helix from RdfS, and its influence on stable DNA binding revealed by both the crystal structure and the molecular dynamics simulations, we decided to test experimentally if  $\alpha 1$  was essential for oligomerization or DNA binding. A construct expressing a truncation of RdfS with residues 2–12 removed (RdfS<sub>13-89</sub>) was generated and the resulting protein purified. Analytical gel filtration experiments revealed RdfS<sub>13-89</sub> was present mostly as dimers with an approximate



**Figure 6.** RdfS modelled with DNA. In all panels, the RdfS ASU tetramer is coloured in a blue-to-red rainbow from N- to C-terminal. **(A)** Visual representation of RdfS tetramer with an ensemble of fragments from crystallographically determined DNA for wHTH proteins: MtaN (red; PDB 1R8D), and filament forming wHTH proteins: BldC (blue; PDB 6AMK) and Xis (purple and green; 2IEF and 1RH6, respectively). Accompanying alignment to each structure individually can be found in [Supplementary Fig. S1](#). **(B)** RdfS ASU tetramer modelled with nucleosome DNA (extracted from PDB 5GSE [72]) closely resembles the fragment ensemble model generated in panel (A). **(C)** Overlaid models of DNA-bound RdfS and RdfS<sub>13-89</sub> initially prepared via molecular modelling based on BldC-DNA (PDB 6AMA). The additional  $\alpha 1$  (not present in RdfS<sub>13-89</sub>) shown in magenta. **(D)** **(i)** Distance in the contact map from the target state with respect to the RdfS crystal structure and **(ii)** RMSD over the adiabatic biased molecular dynamics (first 100 ns) and the subsequent unbiased simulation (final 50 ns) for DNA-bound full length RdfS (black) and RdfS<sub>13-89</sub> (red). **(E)** Overlay of the representative structure of DNA-bound RdfS from adiabatic biased molecular dynamics (with DNA shown in white) with the representative structure following unbiased molecular dynamics simulation (with DNA shown in grey). See in-text and supplementary data for further molecular dynamics analysis. **(F)** Overlay of the representative structure of DNA-bound RdfS<sub>13-89</sub> from adiabatic-biased molecular dynamics (ABMD) (with DNA shown in white) with the representative structure following unbiased molecular dynamics simulation (with DNA shown in grey). **(G)** Minimized model of extended RdfS filament (16 total molecules) coloured with electrostatic surface, and DNA derived from unbiased molecular dynamics simulation shown in panel (E). RdfS  $\alpha 1$  helix coloured in black and a magenta sphere marks the C $\alpha$  of Pro66, showing the relative arrangement of the N- and C-termini of the RdfS superhelical structure.

molecular mass at  $\sim 20$  kDa (Fig. 5A). EMSAs were then carried out with RdfS<sub>13-89</sub> against the *attP* region. A concentration of  $1.3 \mu\text{M}$  RdfS<sub>13-89</sub> was required to shift *attP* DNA, which is approximately twice the concentration required for the wild-type RdfS to shift the same DNA. Interestingly, we observed no significant reduction of the multi-levelled migration effect (Fig. 5E), suggesting that additional RdfS<sub>13-89</sub> molecules can oligomerize independently of  $\alpha 1$  when bind-

ing DNA. Therefore, it seems that removal of the  $\alpha 1$  reduces protein oligomerization, but DNA binding propensity is reduced, and further oligomerization likely still occurs in conjunction with DNA. This was not surprising given other wHTH domains do not have the unique RdfS  $\alpha 1$  helix and yet are capable of oligomerizing on DNA, and RdfS maintains the same, albeit smaller, oligomerization interface as these other proteins.

As RdfS<sub>13–89</sub> was still capable of binding DNA *in vitro*, we wondered if an allele encoding this truncated version would be capable of activating ICEM/Sym<sup>R7A</sup> excision and conjugative transfer. Versions of the *rdfS* gene encoding N- and C-terminal truncations in RdfS, each positioned downstream of the native *rdfS* ribosome-binding site and *PrdfS* promoter, were constructed and cloned into plasmid pFAJ1700. To test the ability of these truncated alleles to stimulate ICEM/Sym<sup>R7A</sup> excision, we introduced these plasmids into *M. japonicum* R7A and used quantitative PCR assays to measure ICEM/Sym<sup>R7A</sup> excision. The R7AΔ*qseM*Δ*rdfS* background was used as it is derepressed for activation of ICEM/Sym<sup>R7A</sup> excision and transfer through deletion of the antiactivator gene *qseM*, but incapable of ICEM/Sym<sup>R7A</sup> excision or transfer due to the deletion of *rdfS*. R7AΔ*qseM*Δ*rdfS* cells containing empty-vector controls exhibited population excision frequencies <0.01% (Fig. 5F). When R7AΔ*qseM*Δ*rdfS* was used as a donor in conjugation experiments, transfer of ICEM/Sym<sup>R7A</sup> was not detected (detection limit of  $1.3 \times 10^{-9}$ ) (Supplementary Table S5). In the presence of a plasmid carrying the wild-type RdfS allele expressed from *PrdfS* (pJR202), the proportion of chromosomes with an excised ICEM/Sym<sup>R7A</sup> increased to ~20% and conjugative transfer of ICEM/Sym<sup>R7A</sup> was restored to an average of  $1.2 \times 10^{-4}$  exconjugants per donor. In contrast, no restoration of excision or conjugative transfer was observed in the presence of α1-truncated RdfS<sub>13–89</sub> expressed from pFJrdfS<sub>13–89</sub>. We wondered if overexpression of this *rdfS*<sub>13–89</sub> allele might be able to compensate for the reduced DNA-binding exhibited by this protein; however, even when *rdfS*<sub>13–89</sub> was expressed from the strong constitutive *nptII* promoter on pF08rdfS<sub>13–89</sub>, no restoration of excision or conjugative transfer of ICEM/Sym<sup>R7A</sup> was observed suggesting RdfS-mediated excision of ICEM/Sym<sup>R7A</sup> requires the RdfS α1-helix. We also constructed a separate plasmid coding for a version of RdfS deleted for the C-terminal 25 amino acids, RdfS<sub>1–64</sub> (residues 64–89 are random coil or unresolved in the RdfS crystal structure). In contrast to the N-terminal truncation, C-terminally truncated version of RdfS restored excision and conjugation frequencies to a similar level as full-length RdfS, suggesting that the disordered C-terminus of RdfS is not critical for excision and conjugation functions.

## Discussion

The diverse molecular roles of the RdfS homologues encoded by *Mesorhizobium* spp. ICEs have been previously reported [19, 43, 44]; however, their structural basis has remained uncharacterized. In this work, we analysed the X-ray crystal structure of *apo* RdfS and discover that the left-handed superhelical filaments formed in crystals are also represented in solution where, at the concentrations tested, RdfS forms tetramers both in the presence and absence of DNA. The head-to-tail oligomerization of RdfS molecules is likely enhanced by the unique N-terminal α-helix not present in other characterized wHTH proteins. DNA-binding assays confirmed RdfS bound the ICEM/Sym<sup>R7A</sup>*attP* site within a 40-bp region directly upstream of the putative IntS P3 arm-site, supporting its role as an RDF/excisionase for IntS-mediated ICEM/Sym<sup>R7A</sup> excision. RdfS binding to *attP* also activated transcription from the *PintS* promoter, which is likely silenced immediately following ICEM/Sym<sup>R7A</sup> excision through negative autoregulation by the IntS protein. DNA binding assays revealed RdfS also binds its own promoter *PrdfS* and thus negatively au-

to-regulates its own expression. The apparent weaker affinity of RdfS for *PrdfS* likely facilitates the establishment of an equilibrium concentration of RdfS in transfer-active cells, which ensures ICEM/Sym<sup>R7A</sup> excision and conjugation functions are activated without causing the growth inhibition and ICEM/Sym<sup>R7A</sup>-curing observed following artificial *rdfS* overexpression. Molecular dynamics simulations predicted the unique N-terminal α1-helix of RdfS was likely required for optimal oligomerization and compaction of higher-order RdfS–DNA complexes. Interestingly, removal of the α1-helix reduced DNA binding and oligomerization *in vitro*, but this same modification abolished all detectable activity of RdfS *in vivo*. In summary, the RdfS protein is a unique member of the wHTH family with a novel oligomer-stabilizing α1-helix. We propose that RdfS oligomerizes in a concentration-dependent manner and such oligomerization facilitates the stepwise binding to various target sites required for ICEM/Sym<sup>R7A</sup> excision and conjugation functions.

The identified RdfS target region *attP*<sub>8</sub>, overlaps the presumed promoter region of *intS*. Our *PintS*–*lacZ* fusion experiments revealed RdfS activates *intS* transcription. In isolation it could be interpreted that the sole role of RdfS therefore is to transcriptionally activate IntS expression rather than participate as RDF. However, overexpression of IntS alone does not stimulate the observed ~100% ICEM/Sym<sup>R7A</sup> excision when both RdfS and IntS are expressed together [19], confirming RdfS is required to stimulate excision directly as observed for other RDFs [15, 20, 73, 74]. Our experiments also showed IntS negatively regulates its own expression. This presumably prevents RdfS from overinducing *PintS*, which could shift the ratio of RdfS:IntS in favour of ICEM/Sym<sup>R7A</sup> integration. Therefore, our data are consistent with a model in which RdfS expression transiently activates IntS expression and then IntS and RdfS together excise ICEM/Sym<sup>R7A</sup>, after which IntS expression is inhibited by IntS, which a ratio of RdfS:IntS that maintains ICEM/Sym<sup>R7A</sup> in an excised state in preparation for conjugative transfer.

Other RDFs such P2 Cox [75], phage 186 Apl [76], P4 Vis [77], WΦ Cox [70], and P22 Cox [78] form higher-order complexes on *att* DNA. The increasing DNA retardation with RdfS concentration observed in EMSAs here supports RdfS forming variable-length multimers along the *attP* DNA region. Our SPR experiments revealed RdfS exhibits specificity for a 40-bp region *attP*<sub>8</sub> and smaller portions of this DNA region were not bound as strongly, suggesting the natural footprint for RdfS binding may be close to this size (at least at this site). SEC-SY-SAXS analysis of 6H-RdfS indicated a stable complex with *attP*<sub>8</sub> as a tetramer (4 RdfS: 1 *attP*<sub>8</sub> dsDNA), with one RdfS molecule per 10-bp of DNA. Preparations of purified RdfS tended to precipitate at higher concentrations, suggesting larger oligomers may form in solution but are unstable in the conditions tested. It seems possible that in more suitable conditions RdfS may form larger multimers together with longer DNA regions.

Much like RdfS, the P2 bacteriophage RDF Cox<sub>P2</sub> is a wHTH domain protein that oligomerizes *in vitro* and its oligomerization is essential for its roles in P2 recombination and regulation [75]. However, Cox<sub>P2</sub> encodes an additional C-terminal helix beyond the corresponding α4 helix in RdfS (Supplementary Fig. S6). The two C-terminal α-helices of Cox<sub>P2</sub> enable it to form a continuous left-handed superhelical structure in crystals. This arrangement of Cox<sub>P2</sub> enables the formation of oligomeric filaments which compact on DNA,



mimicking histone architecture [79]. It seems likely that the C-terminal helices of Coxp<sub>2</sub> and the N-terminal  $\alpha$ 1 helix of RdfS function in an analogous manner, enabling protein–protein contacts along DNA to compact and transform overall morphology of the nucleoprotein filament. Much like the removal of the  $\alpha$ 1 helix of RdfS, mutations within the C-terminal helices of Coxp<sub>2</sub> abolish biological activity [24]. It seems possible that like the Coxp<sub>2</sub> C-terminal helices, the RdfS  $\alpha$ 1 helix may facilitate nucleoprotein contraction and that this contraction may be critical to its functions as an RDF and a transcription factor. Such DNA compaction is supported by our molecular dynamic simulations (Fig. 6), although remains experimentally untested.

Despite identifying several DNA-binding targets for RdfS, we were unable to identify consensus binding motifs. Our SPR experiments also indicated no individual nucleotide changes abolished RdfS binding to *attP*<sub>8</sub>, although most changes reduced the binding response. The *Streptomyces* BldC protein is a WHTH-domain regulator with structural similarity to ICEM/Sym<sup>R7A</sup> RdfS (Fig. 1). BldC binds multiple DNA targets and oligomerizes in a head-to-tail manner. Whilst BldC can bind a 22-bp DNA-target with 9-bp direct repeats (*whiI*) as a dimer, the mechanism of binding to the larger pseudo-continuous *smeA-sffA* promoter differs. BldC cooperatively forms extended filaments along the *smeA-sffA* region with one DNA-turn between each bound BldC molecule [36]. This resembles our proposed model of DNA binding by RdfS (Supplementary Fig. S6). Interestingly, binding of BldC to the *smeA-sffA* region appears not to be strictly dependent on DNA sequence-consensus, suggesting the target DNA structure could be more crucial than sequence—a phenomenon referred to as indirect readout [80, 81]. A similar indirect readout type of DNA-binding mechanism is described for another BldC orthologue from *Actinoplanes missouriensis*, which binds AT-rich regions [82], and for phage  $\lambda$  Xis which forms cooperative filaments along DNA [69]. Unlike RdfS however, *Streptomyces* BldC remains a monomer in solution and the multimeric superhelical behaviour is dependent on the presence of DNA—indicating BldC must be anchored to its cognate DNA to assemble into higher-order oligomers [36]. BldC primarily forms its head-to-tail oligomers due to its hydrophobic core and cross-contacts between Glu and Arg between protomers at the  $\beta$  wing-HTH region. Although the relative position of these residues differs in RdfS, equivalent contacts are made at the  $\beta$  wing-HTH, although many more RdfS intermolecular contacts are made with the  $\alpha$ 1 (Fig. 2B and C). Our analytical SEC and molecular dynamics suggest that RdfS oligomerization is mediated independent of DNA and RdfS–DNA binding stability is enhanced by the N-terminal  $\alpha$ 1 helix. Our data here indicate that RdfS can oligomerize in solution, so it seems possible that RdfS may oligomerize *in situ* prior to DNA binding and as such, have the capability to recognize larger and more dispersed DNA recognition motifs.

Evolving a capacity to recognize larger DNA recognition sites would presumably afford RdfS with a larger DNA footprint within which binding specificity could evolve. Such a capability may have been critical for the evolutionary success of ICEM/Sym<sup>R7A</sup> in competition with the multitude of related ICEs present in the genus *Mesorhizobium*. Proteins closely related to RdfS are encoded by conjugative elements throughout the genus *Mesorhizobium*. Indeed, all identified ICEs (average 1.2 ICEs per genome [14]) and most identified conjugative plasmids in *Mesorhizobium* carry a homo-

logue of RdfS encoded in a similar genetic context upstream of the transfer genes *traF* and *rlxS*. Phylogenetic comparisons of the core conjugation/transfer genes of all *Mesorhizobium* ICEs places all symbiosis-gene carrying ICEs (both monopartite and tripartite) into a monophyletic group [14], which we have named the ICESym family. *Mesorhizobium* ICEs outside this clade exhibit a more diverse evolutionary history but nonetheless carry the same set of core conjugation/transfer genes, including *rdfS*. Strikingly, we observed here that only RdfS homologues encoded by the ICESym family contain the  $\alpha$ 1 helix (Supplementary Fig. S7). All other RdfS homologues encoded by *Mesorhizobium* ICEs and plasmids, while exhibiting ~61% similarity over the WHTH domain region with ICEM/Sym<sup>R7A</sup> RdfS, do not encode the  $\alpha$ 1 helix. We speculate therefore that evolution of this helix coincided with the expansion and dissemination of the ICESym family throughout the genus *Mesorhizobium* and that the additional RdfS helix may have reduced regulatory interference from other RdfS homologues present in *Mesorhizobium* genomes. This model is also consistent with the broader evolutionary distribution of ICESym family ICEs compared to other ICE families in the genus [14, 83].

Several lines of evidence point to RdfS having a role in the regulation of conjugation. The position of the *rdfS* gene upstream of *traF* and *rlxS* conjugation genes is conserved across related ICEs and plasmids throughout the proteobacteria [14, 19]. This contrasts with nearly all other identified RDF genes, which are typically located adjacent to their cognate integrase gene [20]. Deletion of *rdfS* also abolishes ICEM/Sym<sup>R7A</sup> conjugation; however, this is not surprising given the requirement for ICEM/Sym<sup>R7A</sup> excision. More surprising is that the *rdfS* deletion also abolishes the ability of the ICEM/Sym<sup>R7A</sup> conjugation system to mobilize a plasmid carrying the ICEM/Sym<sup>R7A</sup> *oriT*, a process which does not physically require ICEM/Sym<sup>R7A</sup> to excise. This suggests RdfS more directly activates conjugation functions. Furthermore, in these experiments, low-level expression of *rdfS* from the *lac* promoter restored ICEM/Sym<sup>R7A</sup> excision but did not restore conjugation [43], suggesting conjugation might require a greater level of *rdfS* expression. In conjugation experiments here, we introduced *rdfS* expressed from its native promoter and in contrast to our previous work, this restored both ICEM/Sym<sup>R7A</sup> excision and conjugation. These observations suggest that a greater concentration of RdfS is required to activate conjugation than stimulate excision. We speculate that the RdfS regulon extends beyond the *attP* and *PrdF*/*PrdG* regions identified in this work and that at higher concentrations RdfS activates or derepresses additional ICE genes more directly involved in conjugative transfer. The concentration-dependent binding and oligomerization of RdfS with various DNA targets described in this work fits well with such a model. In future experiments, genome-wide protein–DNA interaction assays such as ChIP-Seq in conditions of transient RdfS overexpression may unveil additional targets of the multifunctional RdfS.

## Acknowledgements

This research made use of the SAXS/WAXS beamlines at the Australian Synchrotron, part of ANSTO. The authors thank the SAXS/WAXS team Dr Nigel Kirby and Dr Tim Ryan for their expertise and training resources. C.J.V. acknowledges Dr Katarzyna Ignasiak and Dr Andrew Marshall for their assis-

tance with the SAXS experiments. C.J.V. thanks Prof. Phoebe Rice and Prof. Nicholas Dixon for their feedback on an earlier version of this manuscript. The authors thank Dr Rob Steuart from Curtin Medical Research Institute for technical support with the Biacore T200 system. Bacterial expression vector pETM11 was a gift from Dr Gunter Stier (EMBL, Heidelberg, Germany; <https://www.embl.de>).

**Author contributions:** Callum J. Verdonk (Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing—original draft, Writing—review & editing), Mark Agostino (Formal analysis, Investigation, Methodology), Karina Yui Eto (Methodology, Supervision), Drew Hall (Methodology, Supervision), Charles S Bond (Conceptualization, Formal analysis, Funding acquisition, Methodology, Resources, Supervision, Validation, Writing—original draft, Writing—review & editing), Joshua P Ramsay (Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Resources, Supervision, Validation, Writing—original draft, Writing—review & editing)

## Supplementary data

Supplementary data is available at NAR online.

## Conflict of interest

None declared.

## Funding

J.P.R. is the recipient of an Australian Research Council Future Fellowship (Project ID FT170100235) funded by the Australian Government. C.J.V. is the recipient of an Australian Government Research Training Program (RTP) Scholarship administered by the University of Western Australia. C.J.V. and J.P.R. received funding from Curtin Medical School, Curtin University.

## Data availability

Small-angle x-ray scattering data and fits were deposited at SASBDB under the accession code SASDW97 (<https://www.sasbdb.org/data/SASDW97/>).

## References

1. Beaber JW, Hochhut B, Waldor MK. Genomic and functional analyses of SXT, an integrating antibiotic resistance gene transfer element derived from *Vibrio cholerae*. *J Bacteriol* 2002;184:4259–69. <https://doi.org/10.1128/JB.184.15.4259-4269.2002>
2. Drenkard E, Ausubel FM. *Pseudomonas* biofilm formation and antibiotic resistance are linked to phenotypic variation. *Nature* 2002;416:740–3. <https://doi.org/10.1038/416740a>
3. Peters SE, Hobman JL, Strike P *et al.* Novel mercury resistance determinants carried by IncJ plasmids pMERPH and R391. *Molec Gen Genet* 1991;228:294–9. <https://doi.org/10.1007/BF00282479>
4. Ravatn R, Studer S, Springael D *et al.* Chromosomal integration, tandem amplification, and deamplification in *Pseudomonas putida* F1 of a 105-kilobase genetic element containing the chlorocatechol degradative genes from *Pseudomonas* sp. Strain B13. *J Bacteriol* 1998;180:4360–9. <https://doi.org/10.1128/JB.180.17.4360-4369.1998>
5. Finan TM. Evolving insights: symbiosis islands and horizontal gene transfer. *J Bacteriol* 2002;184:2855–6. <https://doi.org/10.1128/JB.184.11.2855-2856.2002>
6. Sullivan J, Ronson C. Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a *phe*-tRNA gene. *Proc Natl Acad Sci USA* 1998;95:5145–9. <https://doi.org/10.1073/pnas.95.9.5145>
7. Burrus V, Marrero J, Waldor MK. The current ICE age: biology and evolution of SXT-related integrating conjugative elements. *Plasmid* 2006;55:173–83. <https://doi.org/10.1016/j.plasmid.2006.01.001>
8. Dimopoulou ID, Kartali SI, Harding RM *et al.* Diversity of antibiotic resistance integrative and conjugative elements among haemophili. *J Med Microbiol* 2007;56:838–46. <https://doi.org/10.1099/jmm.0.47125-0>
9. Heather Z, Holden MT, Steward KF *et al.* A novel streptococcal integrative conjugative element involved in iron acquisition. *Mol Microbiol* 2008;70:1274–92. <https://doi.org/10.1111/j.1365-2958.2008.06481.x>
10. Colombi E, Straub C, Künzel S *et al.* Evolution of copper resistance in the kiwifruit pathogen *Pseudomonas syringae* pv. actinidiae through acquisition of integrative conjugative elements and plasmids. *Environ Microbiol* 2017;19:819–32. <https://doi.org/10.1111/1462-2920.13662>
11. Osborn MA, Böltner D. When phage, plasmids, and transposons collide: genomic islands, and conjugative- and mobilizable-transposons as a mosaic continuum. *Plasmid* 2002;48:202–12. [https://doi.org/10.1016/S0147-619X\(02\)00117-8](https://doi.org/10.1016/S0147-619X(02)00117-8)
12. Guglielmini J, Quintais L, Garcillán-Barcia MP *et al.* The repertoire of ICE in prokaryotes underscores the unity, diversity, and ubiquity of conjugation. *PLoS Genet* 2011;7:e1002222. <https://doi.org/10.1371/journal.pgen.1002222>
13. Delavat F, Miyazaki R, Carraro N *et al.* The hidden life of integrative and conjugative elements. *FEMS Microbiol Rev* 2017;41:512–37. <https://doi.org/10.1093/femsre/fux008>
14. Colombi E, Perry BJ, Sullivan JT *et al.* Comparative analysis of integrative and conjugative mobile genetic elements in the genus *Mesorhizobium*. *Microbial Genomics* 2021;7:e000657. <https://doi.org/10.1099/mgen.0.000657>
15. Groth AC, Calos MP. Phage integrases: biology and applications. *J Mol Biol* 2004;335:667–78. <https://doi.org/10.1016/j.jmb.2003.09.082>
16. Johnson C, Grossman A. Integrative and conjugative elements (ICEs): what they do and how they work. *Annu Rev Genet* 2015;49:577–601. <https://doi.org/10.1146/annurev-genet-112414-055018>
17. Radman-Livaja M, Biswas T, Ellenberger T *et al.* DNA arms do the legwork to ensure the directionality of lambda site-specific recombination. *Curr Opin Struct Biol* 2006;16:42–50. <https://doi.org/10.1016/j.sbi.2005.12.003>
18. Radman-Livaja M, Biswas T, Mierke D *et al.* Architecture of recombination intermediates visualized by in-gel FRET of  $\lambda$  integrase-Holliday junction-arm DNA complexes. *Proc Natl Acad Sci USA* 2005;102:3913–20. <https://doi.org/10.1073/pnas.0500844102>
19. Ramsay JP, Sullivan JT, Stuart GS *et al.* Excision and transfer of the *Mesorhizobium loti* R7A symbiosis island requires an integrase IntS, a novel recombination directionality factor RdfS, and a putative relaxase RlxS. *Mol Microbiol* 2006;62:723–34. <https://doi.org/10.1111/j.1365-2958.2006.05396.x>
20. Lewis JA, Hatfull GF. Control of directionality in integrase-mediated recombination: examination of recombination directionality factors (RDFs) including Xis and Cox proteins. *Nucleic Acids Res* 2001;29:2205–16. <https://doi.org/10.1093/nar/29.11.2205>
21. Lewis JA, Hatfull GF. Control of directionality in L5 integrase-mediated site-specific recombination. *J Mol Biol*

- 2003;326:805–21.  
[https://doi.org/10.1016/S0022-2836\(02\)01475-4](https://doi.org/10.1016/S0022-2836(02)01475-4)
22. Esposito D, Scocca JJ. Purification and characterization of HP1 Cox and definition of its role in controlling the direction of site-specific recombination. *J Biol Chem* 1997;272:8660–70.  
<https://doi.org/10.1074/jbc.272.13.8660>
  23. Esposito D, Scocca JJ. The integrase family of tyrosine recombinases: evolution of a conserved active site domain. *Nucleic Acids Res* 1997;25:3605–14.  
<https://doi.org/10.1093/nar/25.18.3605>
  24. Berntsson RP, Odegrip R, Söhlén W *et al.* Structural insight into DNA binding and oligomerization of the multifunctional Cox protein of bacteriophage P2. *Nucleic Acids Res* 2014;42:2725–35.  
<https://doi.org/10.1093/nar/gkt1119>
  25. Lewis JA, Hatfull GF. Identification and characterization of mycobacteriophage L5 excisionase. *Mol Microbiol* 2000;35:350–60.  
<https://doi.org/10.1046/j.1365-2958.2000.01695.x>
  26. Aravind L, Anantharaman V, Balaji S *et al.* The many faces of the helix–turn–helix domain: transcription regulation and beyond\*. *FEMS Microbiol Rev* 2005;29:231–62.  
<https://doi.org/10.1016/j.femsre.2004.12.008>
  27. Martínez-Hackert E, Stock AM. The DNA-binding domain of OmpR: crystal structures of a winged helix transcription factor. *Structure* 1997;5:109–24.  
[https://doi.org/10.1016/S0969-2126\(97\)00170-6](https://doi.org/10.1016/S0969-2126(97)00170-6)
  28. Martínez-Hackert E, Stock AM. Structural relationships in the OmpR family of winged-helix transcription factors. *J Mol Biol* 1997;269:301–12. <https://doi.org/10.1006/jmbi.1997.1065>
  29. Brennan RG. The winged-helix DNA-binding motif: another helix–turn–helix takeoff. *Cell* 1993;74:773–6.  
[https://doi.org/10.1016/0092-8674\(93\)90456-Z](https://doi.org/10.1016/0092-8674(93)90456-Z)
  30. Lai E, Clark KL, Burley SK *et al.* Hepatocyte nuclear factor 3/fork head or “winged helix” proteins: a family of transcription factors of diverse biologic function. *Proc Natl Acad Sci USA* 1993;90:10421–3. <https://doi.org/10.1073/pnas.90.22.10421>
  31. Panis G, Franche N, Mejean V *et al.* Insights into the functions of a prophage recombination directionality factor. *Viruses* 2012;4:2417–31. <https://doi.org/10.3390/v4112417>
  32. Carpenter MR, Rozovsky S, Boyd EF. Pathogenicity island cross talk mediated by recombination directionality factors facilitates excision from the chromosome. *J Bacteriol* 2016;198:766–76.  
<https://doi.org/10.1128/JB.00704-15>
  33. Abbani M, Iwahara M, Clubb RT. The structure of the excisionase (Xis) protein from conjugative transposon Tn 916 provides insights into the regulation of heterobivalent tyrosine recombinases. *J Mol Biol* 2005;347:11–25.  
<https://doi.org/10.1016/j.jmb.2005.01.019>
  34. Dorman CJ, Schumacher MA, Bush MJ *et al.* When is a transcription factor a NAP? *Curr Opin Microbiol* 2020;55:26–33.  
<https://doi.org/10.1016/j.mib.2020.01.019>
  35. Ben-Yehuda S, Rudner DZ, Losick R. RacA, a bacterial protein that anchors chromosomes to the cell poles. *Science* 2003;299:532–6. <https://doi.org/10.1126/science.1079914>
  36. Schumacher MA, den Hengst CD, Bush MJ *et al.* The MerR-like protein BldC binds DNA direct repeats as cooperative multimers to regulate *Streptomyces* development. *Nat Commun* 2018;9:1139. <https://doi.org/10.1038/s41467-018-03576-3>
  37. Visweswariah SS, Busby SJW. Evolution of bacterial transcription factors: how proteins take on new tasks, but do not always stop doing the old ones. *Trends Microbiol* 2015;23:463–7.  
<https://doi.org/10.1016/j.tim.2015.04.009>
  38. Mejía-Almonte C, Busby SJW, Wade JT *et al.* Redefining fundamental concepts of transcription initiation in bacteria. *Nat Rev Genet* 2020;21:699–714.  
<https://doi.org/10.1038/s41576-020-0254-8>
  39. Dillon SC, Dorman CJ. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nat Rev Micro* 2010;8:185–95. <https://doi.org/10.1038/nrmicro2261>
  40. Hołówna J, Zakrzewska-Czerwińska J. Nucleoid associated proteins: the small organizers that help to cope with stress. *Front Microbiol* 2020;11:590.  
<https://doi.org/10.3389/fmicb.2020.00590>
  41. Sullivan JT, Patrick HN, Lowther WL *et al.* Nodulating strains of *Rhizobium loti* arise through chromosomal symbiotic gene transfer in the environment. *Proc Natl Acad Sci USA* 1995;92:8985–9. <https://doi.org/10.1073/pnas.92.19.8985>
  42. Kelly S, Sullivan J, Ronson C *et al.* Genome sequence of the *Lotus spp.* microsymbiont *Mesorhizobium loti* strain R7A. *Stand Genomic Sci* 2014;9:6. <https://doi.org/10.1186/1944-3277-9-6>
  43. Verdonk CJ, Sullivan JT, Williman KM *et al.* Delineation of the integrase-attachment and origin-of-transfer regions of the symbiosis island ICEM/Sym<sup>R7A</sup>. *Plasmid* 2019;104:102416.  
<https://doi.org/10.1016/j.plasmid.2019.102416>
  44. Haskett TL, Terpolilli JJ, Ramachandran VK *et al.* Sequential induction of three recombination directionality factors directs assembly of tripartite integrative and conjugative elements. *PLoS Genet* 2018;14:e1007292.  
<https://doi.org/10.1371/journal.pgen.1007292>
  45. Verdonk CJ, Marshall AC, Ramsay JP *et al.* Crystallographic and X-ray scattering study of RdfS, a recombination directionality factor from an integrative and conjugative element. *Acta Crystallogr D Struct Biol* 2022;78:1210–20.  
<https://doi.org/10.1107/s2059798322008579>
  46. Haskett TL, Terpolilli JJ, Bekuma A *et al.* Assembly and transfer of tripartite integrative and conjugative genetic elements. *Proc Natl Acad Sci USA* 2016;113:12268–73.  
<https://doi.org/10.1073/pnas.1613358113>
  47. Haskett TL, Ramsay JP, Bekuma AA *et al.* Evolutionary persistence of tripartite integrative and conjugative elements. *Plasmid* 2017;92:30–6.  
<https://doi.org/10.1016/j.plasmid.2017.06.001>
  48. Ramsay JP, Sullivan JT, Jambari N *et al.* A LuxRI-family regulatory system controls excision and transfer of the *Mesorhizobium loti* strain R7A symbiosis island by activating expression of two conserved hypothetical genes. *Mol Microbiol* 2009;73:1141–55.  
<https://doi.org/10.1111/j.1365-2958.2009.06843.x>
  49. Ramsay JP, Bastholm TR, Verdonk CJ *et al.* An epigenetic switch activates bacterial quorum sensing and horizontal transfer of an integrative and conjugative element. *Nucleic Acids Res* 2022;50:975–88. <https://doi.org/10.1093/nar/gkab1217>
  50. Ramsay JP, Tester LGL, Major AS *et al.* Ribosomal frameshifting and dual-target antiactivation restrict quorum-sensing-activated transfer of a mobile genetic element. *Proc Natl Acad Sci USA* 2015;112:4104–9. <https://doi.org/10.1073/pnas.1501574112>
  51. Ramsay JP. High-throughput  $\beta$ -galactosidase and  $\beta$ -glucuronidase assays using fluorogenic substrates. *Bio Protoc* 2013;3:e827.  
<https://doi.org/10.21769/BioProtoc.827>
  52. Ramsay JP, Major AS, Komarovskiy VM *et al.* A widely conserved molecular switch controls quorum sensing and symbiosis island transfer in *Mesorhizobium loti* through expression of a novel antiactivator. *Mol Microbiol* 2013;87:1–13.  
<https://doi.org/10.1111/mmi.12079>
  53. Yui Eto K, Kwong SM, LaBreck PT *et al.* Evolving origin-of-transfer sequences on staphylococcal conjugative and mobilizable plasmids—who’s mimicking whom? *Nucleic Acids Res* 2021;49:5177–88. <https://doi.org/10.1093/nar/gkab303>
  54. Stevenson CEM, Assaad A, Chandra G *et al.* Investigation of DNA sequence recognition by a streptomycete MarR family transcriptional regulator through surface plasmon resonance and X-ray crystallography. *Nucleic Acids Res* 2013;41:7009–22.  
<https://doi.org/10.1093/nar/gkt523>
  55. Kirby NM, Mudie ST, Hawley AM *et al.* A low-background-intensity focusing small-angle X-ray scattering undulator beamline. *J Appl Crystallogr* 2013;46:1670–80.  
<https://doi.org/10.1107/S002188981302774X>



56. Broennimann C, Eikenberry EF, Henrich B *et al.* The PILATUS 1M detector. *J Synchrotron Rad* 2006;13:120–30. <https://doi.org/10.1107/S0909049505038665>
57. Hemming IA, Clément O, Gladwyn-Ng IE *et al.* Disease-associated missense variants in ZBTB18 disrupt DNA binding and impair the development of neurons within the embryonic cerebral cortex. *Hum Mutat* 2019;40:1841–55. <https://doi.org/10.1002/humu.23803>
58. Kohnke B, Kutzner C, Grubmüller H. A GPU-accelerated fast multipole method for GROMACS: performance and accuracy. *J Chem Theory Comput* 2020;16:6938–49. <https://doi.org/10.1021/acs.jctc.0c00744>
59. Tribello GA, Bonomi M, Branduardi D *et al.* PLUMED 2: new feathers for an old bird. *Comput Phys Commun* 2014;185:604–13. <https://doi.org/10.1016/j.cpc.2013.09.018>
60. Singh S, Plaks JG, Homa NJ *et al.* The structure of Xis reveals the basis for filament formation and insight into DNA bending within a mycobacteriophage intasome. *J Mol Biol* 2014;426:412–22. <https://doi.org/10.1016/j.jmb.2013.10.002>
61. Jowsey WJ, Morris CRP, Hall DA *et al.* DUF2285 is a novel helix–turn–helix domain variant that orchestrates both activation and antiactivation of conjugative element transfer in proteobacteria. *Nucleic Acids Res* 2023;51:6841–56. <https://doi.org/10.1093/nar/gkad457>
62. Pérez J, Vachette P, Russo D *et al.* Heat-induced unfolding of neocarzinostatin, a small all- $\beta$  protein investigated by small-angle X-ray scattering. Edited by M. F. Moody. *J Mol Biol* 2001;308:721–43. <https://doi.org/10.1006/jmbi.2001.4611>
63. Bernadó P, Blackledge M. A self-consistent description of the conformational behavior of chemically denatured proteins from NMR and small angle scattering. *Biophys J* 2009;97:2839–45. <https://doi.org/10.1016/j.bpj.2009.08.044>
64. Kikhney AG, Svergun DI. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett* 2015;589:2570–7. <https://doi.org/10.1016/j.febslet.2015.08.027>
65. Franke D, Jeffries CM, Svergun DI. Machine learning methods for X-ray scattering data analysis from biomacromolecular solutions. *Biophys J* 2018;114:2485–92. <https://doi.org/10.1016/j.bpj.2018.04.018>
66. Hajizadeh NR, Franke D, Jeffries CM *et al.* Consensus Bayesian assessment of protein molecular mass from solution X-ray scattering data. *Sci Rep* 2018;8:7204. <https://doi.org/10.1038/s41598-018-25355-2>
67. Rambo RP, Tainer JA. Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* 2013;496:477–81. <https://doi.org/10.1038/nature12070>
68. Gabel F, Engilberge S, Pérez J *et al.* Medical contrast media as possible tools for SAXS contrast variation. *IUCrJ* 2019;6:521–5. <https://doi.org/10.1107/S2052252519005943>
69. Abbani MA, Papagiannis CV, Sam MD *et al.* Structure of the cooperative Xis-DNA complex reveals a micronucleoprotein filament that regulates phage lambda intasome assembly. *Proc Natl Acad Sci USA* 2007;104:2109–14. <https://doi.org/10.1073/pnas.0607820104>
70. Ahlgren-Berg A, Cardoso-Palacios C, Eriksson JM *et al.* A comparative analysis of the bifunctional Cox proteins of two heteroimmune P2-like phages with different host integration sites. *Virology* 2009;385:303–12. <https://doi.org/10.1016/j.virol.2008.12.002>
71. Connolly KM, Iwahara M, Clubb RT. Xis protein binding to the left arm stimulates excision of conjugative transposon Tn916. *J Bacteriol* 2002;184:2088–99. <https://doi.org/10.1128/JB.184.8.2088-2099.2002>
72. Kato D, Osakabe A, Arimura Y *et al.* Crystal structure of the overlapping dinucleosome composed of hexasome and octasome. *Science* 2017;356:205–8. <https://doi.org/10.1126/science.aak9867>
73. Swalla BM, Cho EH, Gumpert RI *et al.* The molecular basis of co-operative DNA binding between lambda integrase and excisionase. *Mol Microbiol* 2003;50:89–99. <https://doi.org/10.1046/j.1365-2958.2003.03687.x>
74. Laxmikanthan G, Xu C, Brilot AF *et al.* Structure of a Holliday junction complex reveals mechanisms governing a highly regulated DNA transaction. *eLife* 2016;5:e14313. <https://doi.org/10.7554/eLife.14313>
75. Eriksson JM, Haggård-Ljungquist E. The multifunctional bacteriophage P2 cox protein requires oligomerization for biological activity. *J Bacteriol* 2000;182:6714–23. <https://doi.org/10.1128/JB.182.23.6714-6723.2000>
76. Cutts EE, Barry Egan J, Dodd IB *et al.* A quantitative binding model for the Apl protein, the dual purpose recombination-directionality factor and lysis-lysogeny regulator of bacteriophage 186. *Nucleic Acids Res* 2020;48:8914–26. <https://doi.org/10.1093/nar/gkaa655>
77. Cali S, Spoldi E, Piazzolla D *et al.* Bacteriophage P4 Vis protein is needed for prophage excision. *Virology* 2004;322:82–92. <https://doi.org/10.1016/j.virol.2004.01.016>
78. Mattis AN, Gumpert RI, Gardner JF. Purification and characterization of bacteriophage P22 Xis protein. *J Bacteriol* 2008;190:5781–96. <https://doi.org/10.1128/JB.00170-08>
79. Frykholm K, Berntsson RP-A, Claesson M *et al.* DNA compaction by the bacteriophage protein Cox studied on the single DNA molecule level using nanofluidic channels. *Nucleic Acids Res* 2016;44:7219–27. <https://doi.org/10.1093/nar/gkw352>
80. Otwinowski Z, Schevitz RW, Zhang RG *et al.* Crystal structure of trp repressor/operator complex at atomic resolution. *Nature* 1988;335:321–9. <https://doi.org/10.1038/335321a0>
81. Rohs R, Jin X, West SM *et al.* Origins of specificity in protein–DNA recognition. *Annu Rev Biochem* 2010;79:233–69. <https://doi.org/10.1146/annurev-biochem-060408-091030>
82. Tezuka T, Nitta S, Ohnishi Y. Involvement of BldC in the formation of physiologically mature sporangium in *Actinoplanes missouriensis*. *J Bacteriol* 2022;204:e00189–22. <https://doi.org/10.1128/jb.00189-22>
83. Colombi E, Hill Y, Lines R *et al.* Population genomics of Australian indigenous Mesorhizobium reveals diverse nonsymbiotic genospecies capable of nitrogen-fixing symbioses following horizontal gene transfer. *Microb Genom* 2023;9:mgen000918. <https://doi.org/10.1099/mgen.0.000918>