Published online 19 April 2013 *Nucleic Acids Research, 2013, Vol. 41, No. 11* ***5679–5691***
doi:10.1093/nar/gkt277

The actinobacterial transcription factor RbpA binds to the principal sigma subunit of RNA polymerase

Aline Tabib-Salazar[1], Bing Liu[2], Philip Doughty[1], Richard A. Lewis[1], Somadri Ghosh[1], Marie-Laure Parsy[2], Peter J. Simpson[2], Kathleen O'Dwyer[2], Steve J. Matthews[2] and Mark S. Paget[1,*]

[1]School of Life Sciences, University of Sussex, Falmer, Brighton BN1 9QG, UK and [2]Department of Life Sciences, Imperial College London, London SW7 2AZ, UK

Received February 19, 2013; Revised March 26, 2013; Accepted March 28, 2013

ABSTRACT

RbpA is a small non–DNA-binding transcription factor that associates with RNA polymerase holoenzyme and stimulates transcription in actinobacteria, including *Streptomyces coelicolor* and *Mycobacterium tuberculosis.* RbpA seems to show specificity for the vegetative form of RNA polymerase as opposed to alternative forms of the enzyme. Here, we explain the basis of this specificity by showing that RbpA binds directly to the principal σ subunit in these organisms, but not to more diverged alternative σ factors. Nuclear magnetic resonance spectroscopy revealed that, although differing in their requirement for structural zinc, the RbpA orthologues from *S. coelicolor* and *M. tuberculosis* share a common structural core domain, with extensive, apparently disordered, N- and C-terminal regions. The RbpA–σ interaction is mediated by the C-terminal region of RbpA and σ domain 2, and *S. coelicolor* RbpA mutants that are defective in binding σ are unable to stimulate transcription *in vitro* and are inactive *in vivo*. Given that RbpA is essential in *M. tuberculosis* and critical for growth in *S. coelicolor*, these data support a model in which RbpA plays a key role in the σ cycle in actinobacteria.

INTRODUCTION

The high G+C, Gram-positive actinobacteria are of immense medical and industrial importance. For example, although the *Streptomyces* genus is a key source of bioactive compounds, including antibiotics, immunosuppressants and anti-cancer drugs, *Mycobacterium tuberculosis* has maintained its global notoriety, currently causing 9 million cases and 1.5 million tuberculosis deaths per year (1). As well as being a major target for anti-mycobacterial agents, such as rifampicin, RNA polymerase (RNAP) also has a major influence on the control of antibiotic biosynthesis in *Streptomyces*, and mutations in this enzyme complex can both positively and negatively impact production (2,3). Therefore, a deeper understanding of RNAP and its regulation in the *Actinobacteria* phylum can contribute to both the production and development of new medicines.

Transcription initiation in bacteria occurs in several steps. First, an RNAP holoenzyme binds to a promoter element to form an initial closed complex in which the DNA remains double-stranded. This triggers a process of isomerization in which the DNA unwinds, and localized DNA melting reveals the template strand to the active site of the enzyme, giving rise to an open complex. Subsequently, short abortive transcripts are produced in a cyclical manner before the RNAP escapes into the elongation phase (4). The RNAP holoenzyme can be divided into core (five subunits, 2α, β, β′ and ω), which is catalytically active for elongation, and a dissociable σ factor subunit that is required for promoter recognition. The σ subunit provides critical DNA-binding determinants in both closed and open complexes, but it stochastically dissociates from elongating RNAP soon after promoter escape (5,6). This scheme gives rise to a 'σ cycle' whereby the dissociated σ factor enters a pool of σ factors that compete for binding to core RNAP (7). Bacteria are, therefore, able to tune gene expression by controlling the cellular level and composition of σ factors (8). All bacteria seem to contain a single essential σ factor that is orthologous to σ^{70} of *Escherichia coli* and directs the transcription of most 'housekeeping' genes during exponential growth, such as those involved in ribosome production and central metabolism. Structural studies on the principal σ factors of *E. coli* and *Thermus*

*To whom correspondence should be addressed. Tel: +44 1273 877764; Fax: +44 1273 678433; Email: m.paget@sussex.ac.uk

sp., either alone or in the context of holoenzyme, revealed four domains, $\sigma_{1.1}$, σ_2, σ_3 and σ_4, that are interconnected by flexible linkers and correspond to conserved regions 1.1, 1.2–2.4, 3.0–3.1 and 4.1–4.2, respectively (9–12). The σ_2, σ_3 and σ_4 domains each comprise key contact points for both core enzyme and promoter DNA, whereas $\sigma_{1.1}$ is thought to maintain free σ in a compact non–DNA-binding form through interaction with σ_2 and σ_4 (12).

In addition to a principal and essential σ factor, bacteria usually contain one or more alternative σ factors that can redirect the transcription machinery to specific regulons, allowing the cell to respond to a wide variety of environmental, physiological or developmental cues (8). Most of these are related to σ^{70} and have been classified into four main groups (Groups 1–4) based on phylogeny and biological function (13): Group 1 σ factors are the σ^{70} orthologues; Group 2 σ factors are closely related to Group 1 σ factors but have specialized non-essential functions; Group 3 σ factors are structurally and functionally diverse but include all three globular domains (σ_2–σ_4); Group 4 comprises the highly diverse extracytoplasmic function (ECF) σ factors that lack both Region 1 and the σ_3 globular domain. The control of alternative σ factor activity is particularly well established and often occurs at the post-translational level through the action of anti-σ factors that bind to σ and prevent its association with core RNAP (14). The σ factor is released after the inactivation of the anti-σ factor through competition with an alternative binding partner (partner-switching), regulated proteolysis or an allosteric change in response to the direct sensing of a signal. Although the σ subunit can act as an essential contact point for DNA-binding activators, proteins that bind to and activate free σ factors are rare. One example, however, is the small protein Crl, which stimulates the activity of the stationary phase regulator σ^S in *E. coli* and *Salmonella* (15) through direct interaction, both with free σ and as part of the holoenzyme (16,17). The role of Crl seems to be to improve the association of σ^S with core RNAP through interaction with the σ_2 globular domain (18–21).

RbpA is an RNAP-binding protein that is confined to, and widespread within, the actinobacteria, and evidence is accumulating that it plays a crucial role in transcription initiation. RbpA was discovered in the antibiotic-producing bacterium *Streptomyces coelicolor* where the 124 amino acid protein was shown to confer basal levels of resistance to rifampicin *in vivo* and stimulate transcription *in vitro* (22,23). The *S. coelicolor* RbpA (designated $RbpA^{Sc}$ for clarity) co-purifies as a major component of RNAP holoenzyme, suggesting a broad role in transcription, although stimulation of transcription was previously only demonstrated using the ribosomal RNA promoter *rrnDp3* (22). The *rrnDp3* promoter is dependent on the principal σ factor in *S. coelicolor*, σ^{HrdB}, and $RbpA^{Sc}$ failed to stimulate transcription from a promoter dependent on the alternative σ factor σ^R (22). Although *S. coelicolor* mutants that lack $RbpA^{Sc}$ grow slowly, they are viable. However, it was recently demonstrated that RbpA (Rv2050; designated $RbpA^{Mt}$ for clarity) is essential in *M. tuberculosis* (24). Furthermore, $RbpA^{Mt}$ was shown to stabilize the principal σ^A-RNAP holoenzyme and stimulate transcription from wide range of σ^A-dependent promoters *in vitro*, suggesting a general role in transcription initiation (25). Two biochemical studies have proposed that the primary contact point for $RbpA^{Mt}$ is the β subunit. However, although one study proposed binding close to the active-site cleft near the rifampicin-binding site (26), the other study proposed that $RbpA^{Mt}$ contacts the sandwich-barrel hybrid motif located on the surface of the β subunit (25). In either study, the apparent specificity of $RbpA^{Mt}$ for the principal σ factor was not explained, although it was proposed that $RbpA^{Mt}$ might allosterically modify core RNAP to improve its interaction with σ (25). Here, we show that RbpA binds specifically to the principal and certain related σ factors of *S. coelicolor* and *M. tuberculosis*, and that $RbpA^{Sc}$ is present at a σ^{HrdB}–RNAP transcription initiation complex *in vivo*. The orthologous RbpA proteins from these organisms comprise a core domain, for which we present the high-resolution solution structures, together with extensive N- and C-terminal regions that are disordered in the free proteins. Finally, we map the interaction between *S. coelicolor* σ^{HrdB} and $RbpA^{Sc}$, and we show that binding occurs primarily between the σ_2 domain of σ^{HrdB} and the C-terminal region of $RbpA^{Sc}$. Our data support a model in which RbpA plays a key role in the σ cycle of the vegetative RNAP in actinobacteria, possibly acting as a chaperone-like protein to aid the formation of active holoenzyme, which might be particularly important during stress.

MATERIALS AND METHODS

Strains, plasmids and growth conditions

Strains and plasmids are listed in Supplementary Table S1, and oligonucleotides used for the construction of plasmids are listed in Supplementary Table S2. *E. coli* K-12 strain DH5α was used for general cloning and plasmid propagation, and *E. coli* ET12567 (pUZ8002) was used to conjugate plasmids into *S. coelicolor* (27). *S. coelicolor* A3(2) strains were cultivated on mannitol–soya (MS) agar or in a modified yeast extract–malt extract (YEME, 10% sucrose and 0.01% antifoam) broth as previously described (27). To construct *S. coelicolor* S129, a Δ*rbpA*::*apr* allele was introduced into the *S. coelicolor* J1981 (*rpoC*-His_6) genome by double crossover recombination as described previously (22). To construct pSX190, the *S. coelicolor rbpA* gene and promoter region were amplified using oligonucleotide primers rbpA_H3_rev and DT1a (22), thereby introducing a HindIII site immediately upstream of the stop codon, then fused to a triple FLAG tag sequence in the integrative vector pSET152. Site-directed mutagenesis of the C-terminal region of *rbpA* was performed in pSX512, and the *rbpA* derivatives were cloned along with the native promoter into pSETΩ as a BglII fragment.

Protein overexpression and purification

Core RNAP was prepared from late-exponential YEME cultures of *S. coelicolor* S129. Harvested

mycelium was re-suspended in binding buffer (20 mM Tris–HCl, pH 7.5, 500 mM NaCl, 20 mM imidazole, 3 mM 2-mercaptethanol, 5% glycerol and 25 μg/ml phenylmethanesulphonyl fluoride) plus complete ethylenediaminetetraacetic acid (EDTA)-free protease inhibitor (Roche). Mycelium was lysed by passing twice through a French press at 10 000 kPa before centrifugation to remove the cell debris. The lysate was immediately loaded on a Ni-IDA sepharose column and washed with 10 column volumes of binding buffer followed by 5 column volumes buffer A (20 mM Tris–HCl, pH 7.5, 3 mM 2-mercaptoethanol, 5% glycerol and 150 mM NaCl) containing 20 mM imidazole, then eluted with buffer A containing 250 mM imidazole. The eluate was immediately applied to a HiTrap Heparin HP column (GE Healthcare) equilibrated with buffer A, and protein was eluted from the column using a salt gradient (20 mM Tris–HCl, pH 7.5, 3 mM 2-mercaptoethanol, 5% glycerol and 1 M NaCl). RNAP fractions were dialysed against buffer C (20 mM Tris–HCl, pH 7.5, 3 mM 2-mercaptoethanol, 5% glycerol and 50 mM NaCl) and further purified by mono-Q (GE Healthcare) ion-exchange chromatography. Core-rich RNAP fractions were eluted in tail fractions from the mono-Q column. Native $RbpA^{Sc}$, hexahistidine (His_6)-tagged $RbpA^{Sc}$ and σ^{HrdB} were prepared as described previously (22). For *in vitro* transcription reactions, the His_6-tag on $RbpA^{Sc}$ was removed by treatment with thrombin, and the protein was purified by gel filtration. For expression of truncated σ^{HrdB} domain fragments (σ_2, σ_2–σ_4, σ_3–σ_4 and σ_4) with N-terminal His_6-tags, polymerase chain reaction (PCR) was used to amplify the corresponding *hrdB* regions with primers incorporating N-terminal NdeI and C-terminal XbaI sites and cloned into pET15b. Expression was achieved in *E. coli* BL21λDE3 (pLysS), and fragments were purified using Ni-NTA spin columns (Qiagen) as recommended by the manufacturer. For NMR analysis, the full-length *M. tuberculosis rbpA* gene (Rv2050) was amplified by PCR incorporating an N-terminal NdeI site and C-terminal BamHI site and cloned into pET20b to give pSX500. To solve the structure of the core domain of $RbpA^{Sc}$, a truncated gene encoding the first 75 amino acids was chemically synthesized with codon optimization (Eurofins) for expression in *E. coli* and subsequently cloned into pET15b to give pSX505. *E. coli* BL21λDE3 (pLysS) containing pSX500 or pSX505 was grown in L-broth to OD_{600} of 3–5, then transferred to optimized high-cell density isopropyl-β-D-thiogalactopyranoside (IPTG)-induction minimal media (28,29) containing 0.1% $^{15}NH_4Cl$ and 1% ^{13}C-glucose, grown for 2 h at 37°C, before induction with 1 mM IPTG at 30°C for 3 h. Oligonucleotides used to amplify each gene or domain are listed in Supplementary Table S2.

In vitro transcription reactions

In vitro transcription assays were carried out using *S. coelicolor* S129 core RNAP and purified native σ^{HrdB} and $RbpA^{Sc}$ proteins, essentially as described previously (30). Templates containing promoters were amplified by PCR using primers listed in Supplementary Table S2. Assays (10 μl) were performed at 30°C, and they contained 5 nM DNA template in reaction buffer (50 mM Tris–HCl, pH 8, 10 mM $MgCl_2$, 200 mM KCl, 0.1 mg/ml of bovine serum albumin and 10 mM β-mecaptoethanol). σ^{HrdB}, $RbpA^{Sc}$ and core RNAP were mixed in RNAP dilution buffer (10 mM Tris–HCl, pH 8, 10 mM KCl, 0.4 mg/ml of bovine serum albumin, 0.1 mM Na_2EDTA, 0.1% Triton X-100 and 10 mM β-mecaptoethanol) (30) for 15 min on ice, and reactions were initiated by the addition of nucleotide mix (1 mM adenosine triphosphate, GTP, CTP and 50 μM UTP), including 5 μCi [α-^{32}P] UTP (>800 Ci/mmol, GE Healthcare), and incubated for 15 min. Reactions were halted by addition of an equal volume of *in vitro* loading dye [80% (w/v) formamide, 0.01% (w/v) xylene cyanol and 0.01% (w/v) bromophenol blue], separated on denaturing 6% polyacrylamide gels and quantified by phosphorimaging.

Bacterial two-hybrid analysis

The complete *rbpA* genes from *S. coelicolor* and *M. tuberculosis*, as well as truncated derivatives as indicated in figure legends, were PCR amplified then cloned into pUT18, generating *rbpA*-T18 translational fusions to the T18 catalytic domain of *Bordetella pertussis* adenylate cyclase. The σ factor genes from *S. coelicolor* (*hrdA*, *hrdB*, *hrdC*, *hrdD*, *sigB*, *sigE*, *sigR* and *whiG*) and *M. tuberculosis* (*sigA* and *sigB*), as well as truncated derivatives of *hrdB*, were cloned into pKT25, generating T25-σ translational fusions to the T25 catalytic domain. The amino acids included in each truncated hybrid construct are indicated in the figure legends, and the oligonucleotides used for PCR amplification are indicated in Supplementary Table S2. Functional complementation between chimeric proteins was evaluated in *E. coli* BTH101 by performing β-galactosidase measurements on permeabilized overnight cultures grown in L-broth containing 0.5 mM IPTG. One unit of β-galactosidase activity corresponds to 1 nmol of *o*-nitrophenyl-β-D-galactoside (ONPG) hydrolysed per min at 28°C. Results were presented either as Miller units (31) or as % Miller units relative to a control.

In vitro pull-down assays

Experiments were performed using Dynabeads® His-Tag Isolation and Pulldown beads (Invitrogen) essentially as recommended by the manufacturer. Test proteins were mixed on ice in binding buffer (50 mM sodium phosphate, pH 8, 300 mM NaCl and 0.01% Tween®-20), then added to 20 μl magnetic Dynabeads and left to incubate for 30 min at ambient temperature with shaking. Samples were washed four times, and then eluted in 100 μl elution buffer (binding buffer plus 300 mM imidazole) for 15 min.

Chromatin immunoprecipitation–quantitativePCR

Strains were grown in modified YEME to mid-late exponential phase (OD_{450}: 2.0), treated with rifampicin (140 μg/ml) for 20 min, then formaldehyde [1.1% (v/v); Sigma] for a further 20 min, with continued shaking at 200 rpm. After quenching with 0.5 M glycine for 5 min at 30°C, mycelium was collected by centrifugation, washed

twice in phosphate-buffered saline and stored at −20°C. Chromatin extraction and immunoprecipitation protocols were essentially as described previously (32). Mycelial pellets were subjected to 3 × 90 s of cryogenic grinding (Retsch MM301, 30 Hz) then resuspended in 9 ml of IP buffer (32), and 3-ml aliquots were sonicated (35% amplitude) for 4, 8 and 12 cycles of 15 s ON and 45 s OFF. After clearing by centrifugation, 70 µl of samples were incubated with proteinase K (Roche), de-cross-linked and then analysed by gel electrophoresis (32). Samples comprising chromatin fragments of ~0.5 kb were precleared by incubation (rotation for 4 h at 4°C) with either protein A or protein G magnetic beads (NEB) before the addition of antibodies. Immunoprecipitations were conducted with M2 mouse monoclonal anti-FLAG antibody (Sigma), monoclonal 8RB13 anti-*E. coli* RNAP β subunit antibody (Abcam) or a custom rabbit polyclonal antiserum raised against σ^{HrdB} (Eurogentec). Corresponding mock immunoprecipitations (IPs) with no added antibody were performed as negative control. After incubation with antibodies (rotation for 12 h at 4°C), immunoprecipitations were performed using protein G magnetic beads (RbpA^{Sc} and RNAP; NEB) or protein A magnetic beads (σ^{HrdB}; NEB). Subsequent steps to dissociate cross-links and purify DNA were as described previously (32), and DNA was resuspended in 25 µl of ultrapure water. Quantitative real-time PCR was performed using the standard curve absolute quantification approach using serial dilutions of sonicated *S. coelicolor* M145 genomic DNA to generate standard curves for each of the four primer pairs used (Supplementary Table S2). qPCR reactions were set-up in 0.1 ml of MicroAMP FAST optical 96-well reaction plates (Applied Biosystems) using GoTaq qPCR reaction mix (Promega), 10 ng of template DNA and each primer at 0.15 pmol/µl. The data were analysed using a modified method to that previously described (33). The mock immunoprecipitation (IP) signal was subtracted from the signal obtained using the corresponding immunoprecipitated sample DNA as input to correct for background signal. To allow visual comparison of the relative enrichments obtained using the three different antibodies, the background corrected signals obtained using a particular antibody were expressed relative to the highest corrected signal obtained using either of the four primer pairs.

Metal analysis

Native (un-tagged) RbpA proteins were produced in *E. coli* and purified by ion-exchange chromatography and gel filtration, before concentration as described previously (22). Approximately 0.5 mg of protein was vacuum-dried then re-suspended in 86 µl of 70% nitric acid (metal analysis grade) and heated for 10 min at 105°C. Water (metal analysis grade) was added to give a final concentration of 2% nitric acid, and inductively coupled plasma mass spectrometry was performed using an Agilent 7500 series instrument. A solution containing 1 ppm of each metal to be analysed (Mn, Fe, Ni, Cu and Zn) was used as standard, and values were normalized against protein storage buffer alone. Experiments were performed in duplicate.

NMR and solution structure determination

Samples were buffer-exchanged into 100 mM MES buffer, pH 6.0, with 150 mM KCl. NMR experiments were performed on >200 µM samples of uniformly ^{15}N- and ^{13}C-labelled protein.

The backbone resonance assignments of RbpA^{Sc} and RbpA^{Mt} were completed by using the standard triple-resonance assignment approach (34), and data were analysed using in-house algorithms with the program NMRView (35). Hα and Hβ assignments were obtained using HBHA(CBCACO)NH. The side-chain assignments of the two proteins were extended using HCCH TOCSY and (H)CC(CO)NH TOCSY. The distance restraints were obtained from ^{1}H-^{15}N/^{13}C nuclear Overhauser effect spectroscopy (NOESY)–heteronuclear single- or multiple-quantum coherence (HSQC/HMQC) experiments (mixing time, 100 ms at 800 MHz). Data were collected on Bruker AvanceIII (600 MHz) and AvanceII (800 MHz) spectrometers equipped with TCI and TXI cryoprobes, respectively. The ARIA protocol was used for completion of the NOE assignment and structure calculation (36). Dihedral angle restraints derived from TALOS were also incorporated in the calculation (37). The frequency window tolerances for assigning NOEs were ±0.05 ppm for direct proton dimensions and ±0.05 ppm for indirect proton dimensions, and ±0.5 ppm for nitrogen dimensions and ±1.3 ppm for carbon dimensions. One hundred structures were calculated in the final iteration, and these had no NOE violations >0.5 Å and no dihedral angle violations >5°. The 10 lowest energy structures were selected for further analysis and figure generation.

RESULTS

RbpA^{Sc} activates a range of σ^{HrdB}-dependent promoters

Previously, using total RNAP holoenzyme preparations, we showed that RbpA^{Sc} stimulated transcription *in vitro* from the σ^{HrdB}-dependent promoter *rrnDp3*, but not from a σ^{R}-dependent promoter, *rbpAp* (22). Initially, we extended this work by showing that RbpA increased the maximal activity of σ^{HrdB}–RNAP at the *rrnDp3* promoter by titrating increasing concentrations of σ^{HrdB} in the presence and absence of excess RbpA^{Sc} (Figure 1A). We then tested whether the RbpA^{Sc} activity was specific to *rrnDp3* or whether it played a more general role in activating σ^{HrdB}-dependent promoters. Transcription run-off assays were performed with a selection of promoters that we or others have predicted to be dependent on σ^{HrdB}, based on the putative −10 promoter elements: *sacAp* (38), *tuf3p* (39), *relAp2* (40), *atpIp* (unpublished observations) and *rplJp* (41). In each case, RbpA^{Sc} stimulated transcription in assays consisting of core RNAP, σ^{HrdB} and linear DNA templates (Figure 1B). RbpA^{Sc} also activated transcription initiation in single-round transcription assays, which suggests that the stimulatory effect is not a consequence of increased recycling or elongation (data not shown). Together with the previous demonstration that RbpA^{Mt} stimulates a range of σ^{A}-dependent promoters with no apparent specificity (25),

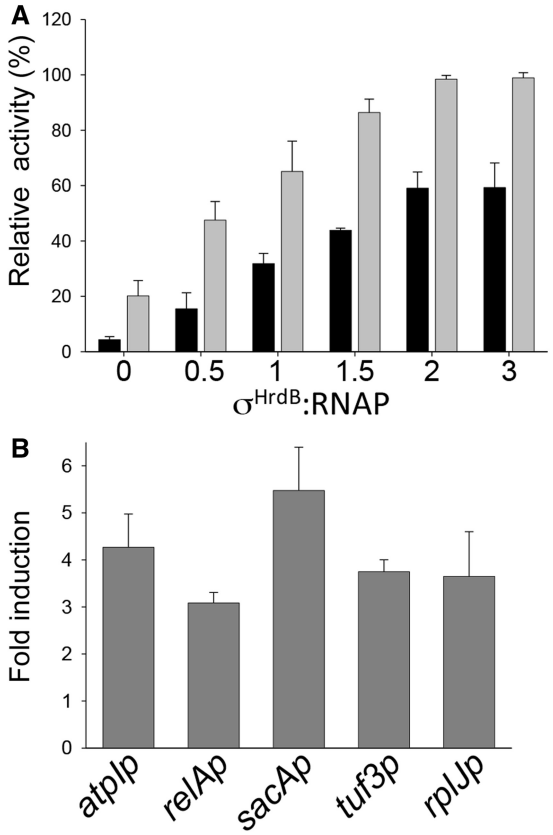

Figure 1. RbpA^{Sc} activates transcription from a range of σ^{HrdB}-dependent promoters *in vitro*. (**A**) Multi-round *in vitro* transcription reactions using the *rrnDp3* promoter as template (22) and including core RNAP (75 nM) σ^{HrdB} at σ^{HrdB}: RNAP ratios of 0, 0.5, 1, 1.5 or 2, or 3, in the presence (grey bars) or absence (black bars) of excess RbpA^{Sc} (40 μM). *rrnDp3* transcription products are normalized relative to the highest signal. Note that these preparations of core RNAP have residual levels of σ^{HrdB}. (**B**) Multi-round *in vitro* transcription reactions contained core RNAP (75 nM) σ^{HrdB} (375 nM), RbpA^{Sc} (750 nM) and DNA templates generated by PCR using primers listed in Supplementary Table S2. Data are presented as fold-difference relative to reactions lacking RbpA. Transcript levels were quantified by phosphorimaging from triplicate data, and standard deviation is indicated.

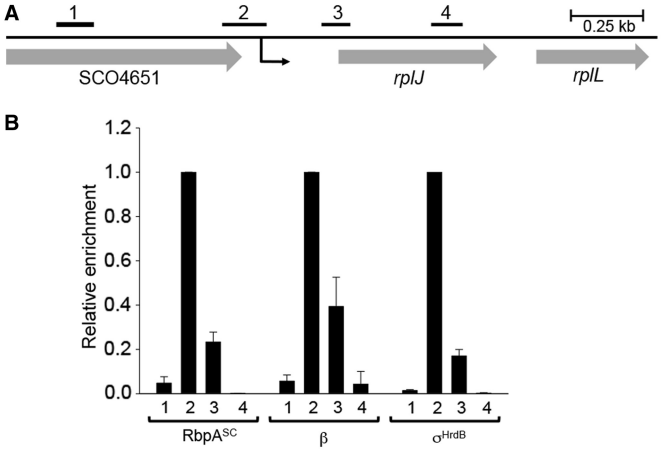

Figure 2. Localization of RbpA^{Sc} at a σ^{HrdB}-dependent promoter *in vivo*. (**A**) Schematic of the *rplJ* region. The transcription start point (+1) of *rplJp* is indicated by a black arrow located 240 bp upstream from the *rplJ* start codon. The bars above the genes show the relative positions of PCR products used for ChIP–qPCR that are centred with respect to +1 as follows: 1, −653 bp; 2, −79 bp; 3, +235 bp; 4, +235 bp. (**B**) Occupancy of RbpA^{Sc}, RNAP β subunit, and σ^{HrdB} at the indicated regions in *S. coelicolor* S129 (pSX190), after treatment with rifampicin. Immunoprecipitations were performed using monoclonal anti-β, polyclonal anti-σ^{HrdB} and anti-FLAG antibody to detect RbpA–FLAG. To allow comparison of RbpA^{Sc}, β and σ^{HrdB} localization, after absolute quantitation of co-immunoprecipitated DNA and background correction for each antibody, enrichment is presented relative to the highest corrected signal obtained using either of the four primer pairs. Standard deviations (calculated for two biological replicates) are indicated.

these data support the idea that RbpA is a general activator of the principal RNAP holoenzyme in actinobacteria.

Although it is known that RbpA^{Sc} is present in purified RNAP preparations, it has not yet been shown to be present in RNAP–promoter complexes *in vivo*. We, therefore, decided to test this using chromatin immunoprecipitation combined with quantitative PCR (ChIP–qPCR). The presence of RbpA^{Sc}, σ^{HrdB} and the β subunit of RNAP was analysed at the *rplJp* promoter. Immunoprecipitation of β and σ^{HrdB} was performed using a commercially available monoclonal antibody (42) and a polyclonal antibody, respectively. To immunoprecipitate RbpA^{Sc}, the *rbpA* gene was C-terminally tagged with tandem FLAG epitopes and integrated into the genome of the *S. coelicolor rbpA* mutant S129 at the ϕC31 attachment site using the vector pSET152. The resulting construct, pSX190, fully restored normal growth rate to S129, indicating that that the epitope tag did not impede normal RbpA^{Sc} function (data not shown). *S. coelicolor* S129 (pSX190) was grown to mid-exponential phase before treatment with rifampicin to inhibit global RNA synthesis by trapping RNAP at promoters (43). As expected, the *rplJ* promoter region (qPCR product 2; Figure 2), centred −79 bp upstream from the transcription start point, was highly enriched for both σ^{HrdB} and the β subunit, with much lower enrichment seen for control regions centred −653 bp upstream, or +235 bp and +608 bp downstream, of the transcription initiation site, suggesting that σ^{HrdB}–RNAP was indeed trapped at the *rplJ* promoter (Figure 2). The enrichment patterns seen using the anti-FLAG antibody were similar to those obtained with the σ^{HrdB} and the β antibodies, indicating that RbpA^{Sc} is present in these transcription initiation complexes, although it should be noted that differences in antibody–epitope affinities prevent the relative proportion of initiation complexes that contain RbpA^{Sc} from being assessed. Control anti-FLAG immunoprecipitation experiments using an equivalent strain in which RbpA^{Sc} was untagged showed no enrichment above that seen with the 'no antibody' control, confirming that this signal is specific to RbpA^{Sc} (data not shown).

RbpA binds specifically to principal and related σ factors

We considered the possibility that the apparent specificity of RbpA activity towards promoters recognized by the principal σ factors might be determined by a direct interaction with the σ factor itself. To test this idea, we applied

a bacterial two-hybrid (BACTH) assay based on the *B. pertussis* adenylate cyclase (44). The *S. coelicolor rbpA* gene was fused to the gene that encodes the T18 domain of adenylate cyclase such that $RbpA^{Sc}$ was at the N-terminus of the fusion protein. A range of σ factor genes were fused to the gene that encodes the T25 domain, such that the σ factors were located at the C-terminus of the fusion proteins. Initially we tested for interaction between $RbpA^{Sc}$ and the following *S. coelicolor* σ factors: σ^{HrdB} (Group 1); σ^{HrdA}, σ^{HrdC} and σ^{HrdD} (Group 2); σ^{WhiG} and σ^{B} (Group 3); and σ^{E} and σ^{R} (Group 4) (Figure 3A). The Group 1 and Group 2 σ factors were N-terminally truncated such that they included only conserved regions 1.2–4.2, which comprise structural domains σ_2, σ_3 and σ_4. This was necessary because full-length T25–*hrdB* fusions seemed to be toxic to *E. coli*. A positive result was detected for σ^{HrdB} and σ^{HrdA}, whereas none of the remaining σ factors bound to $RbpA^{Sc}$. This suggests that $RbpA^{Sc}$ interacts only with principal and closely related σ factors. To broaden this idea, we tested for interaction between $RbpA^{Mt}$ and *M. tuberculosis* σ factors σ^{A} (Group 1) and σ^{B} (Group 2). Binding was detected in both cases (Figure 3A), which suggests that the interaction with principal and related σ factors is a general feature of RbpA proteins in actinobacteria. In principal, the interactions observed during *in vivo* two-hybrid analysis could involve *E. coli* RNAP in hybrid higher order complexes. To confirm that the interaction between $RbpA^{Sc}$ and σ^{HrdB} does not involve any other factors, the two proteins were analysed, separately or mixed, by gel filtration chromatography (Figure 3B). σ^{HrdB} (55.9 kDa) and $RbpA^{Sc}$ (14.1 kDa) eluted at different volumes with apparent molecular masses of 183.2 and 29.9 kDa, respectively, suggesting trimeric and dimeric complexes. When mixed together, a complex of the two proteins eluted close to the σ^{HrdB} peak with an apparent mass of 211.4 kDa. Although it is clear that the two proteins interact, the difference in size between σ^{HrdB} and $RbpA^{Sc}$, and our suspicion that σ^{HrdB} might run anomalously during gel filtration, makes it difficult to predict precise stoichiometry at this point.

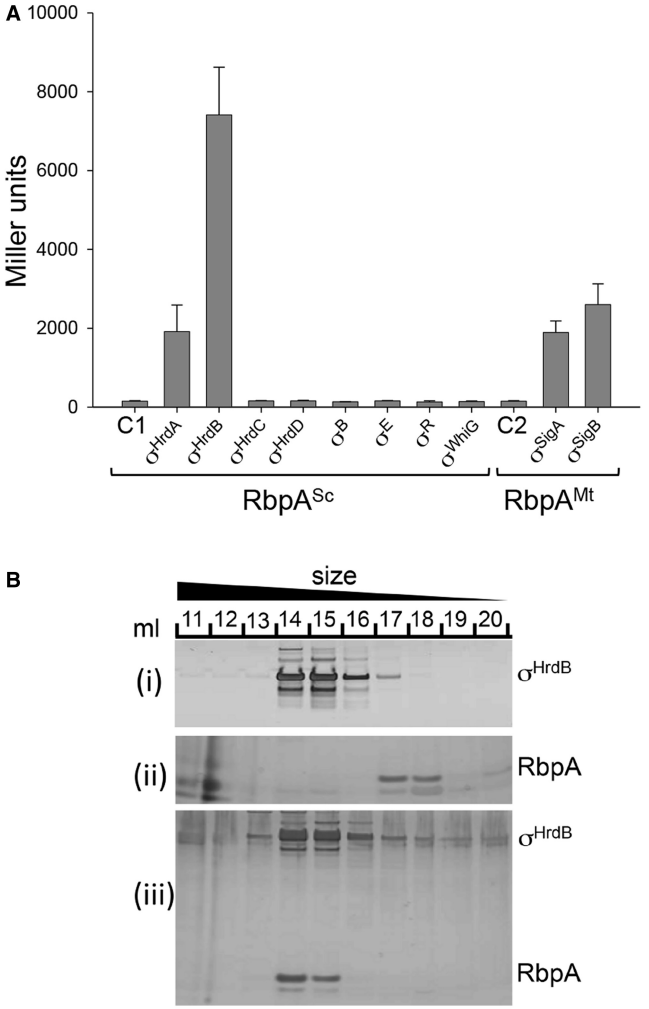

Figure 3. RbpA binds to the principal and closely related σ factors. (**A**) BACTH analysis of RbpA from *S. coelicolor* ($RbpA^{Sc}$) and *M. tuberculosis* ($RbpA^{Mt}$) interactions with σ factors from *S. coelicolor* (Group 1, σ^{HrdB}; Group 2, σ^{HrdA}, σ^{HrdC} and σ^{HrdD}; Group 3, σ^{HrdB} and σ^{WhiG}; Group 4, σ^{E} and σ^{R}) and *M. tuberculosis* (Group 1, σ^{A}; Group 2, σ^{B}). The *rbpA* genes were fused to the T18 subunit of *B. pertussis* adenylate cyclase, whereas the σ factor genes were fused to the T25 subunit. Groups 1 and 2 σ factor hybrid fusions included only σ domains σ_2, σ_3 and σ_4, whereas the remaining σ factor fusions were full length. β-galactosidase assays were performed in triplicate, and standard deviations are indicated. Control strains contained C1, pUT18-$rbpA^{Sc}$, pKT25; C2, pUT18-$rbpA^{Mt}$, pKT25. (**B**) Co-elution of $RbpA^{Sc}$ and σ^{HrdB} during size-exclusion chromatography. (i) σ^{HrdB} (5 nmol), (ii) $RbpA^{Sc}$ (5 nmol) or a mixture of the two (iii) were passed through a Superose 6 10/300 GL size-exclusion column, and 1 ml of fractions (indicated by numbers) were separated by sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS–PAGE) and silver stained.

RbpA orthologues differ in their zinc content

Protein sequence alignments suggested that the RbpA orthologues from *S. coelicolor* and *M. tuberculosis* might differ in their metal content (Figure 4). *S. coelicolor* $RbpA^{Sc}$ contains three cysteine residues, Cys35, Cys56 and Cys59, and a histidine His38, each of which are widely conserved (e.g. in *Bifidobacterium* sp. and *Frankia* sp.) and arranged as a putative C(H/C)CC zinc-ribbon motif. Surprisingly, $RbpA^{Mt}$, along with RbpA orthologues from *Corynebacterium* sp. and *Nocardia* sp., lacks all of these residues apart from the Cys56-equivalent cysteine. To investigate metal content, we isolated $RbpA^{Sc}$ and $RbpA^{Mt}$ from overproducing *E. coli* strains in their native (untagged) forms. Inductively coupled plasma mass spectrometry (ICP-MS), set to detect Mn, Fe, Ni, Cu and Zn, revealed zinc as the only metal present at significant quantities in $RbpA^{Sc}$, at a stoichiometry of 0. 83 (±0.04) mol zinc mol^{-1} protein. Conversely, no metal was detected at significant levels in $RbpA^{Mt}$. To test the importance of the presumed cysteine zinc ligands in $RbpA^{Sc}$, each was mutated to alanine, and the resultant genes were introduced into *S. coelicolor* S101 (Δ*rbpA::hyg*) using the integrative expression vector pIJ6902 to test for complementation. In each case, the mutant genes failed to restore normal growth rate (Supplementary Figure S1), suggesting that each residue is crucial for the function or stability of the protein. Each

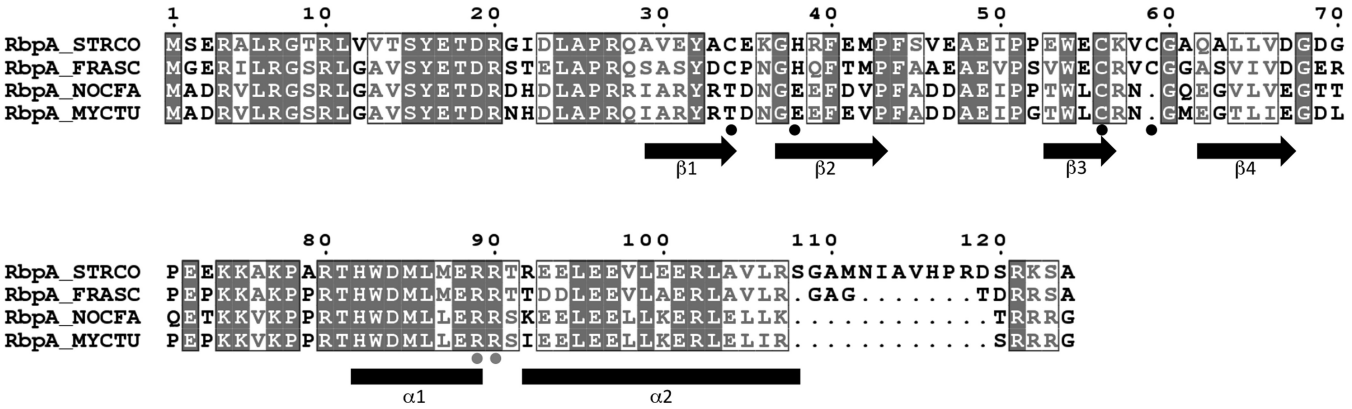

Figure 4. Alignment of the amino acid sequences of selected RbpA orthologues. STRCO, *S. coelicolor*, NP_625703; FRASC, *Frankia* sp. CcI3, YP_481367; NOCFA, *Nocardia farcinica*, YP_119325; *M. tuberculosis* H37Rv, NP_216566. Residues that coordinate a zinc atom in $RbpA^{Sc}$, or form a stabilizing hydrogen bonding network in $RbpA^{Mt}$, are indicated with black dots. Arginine residues that when mutated prevent σ^{HrdB} interaction are indicated by grey dots (Figure 6). The β-strand secondary structure elements in the RCD domain are indicated, as are predicted α-helices in the C-terminal region.

of the three histidine residues in $RbpA^{Sc}$ was also changed to alanine, although in each case the mutant protein complemented the S101 strain (Supplementary Figure S1), suggesting that the protein retains function even when the functional requirement of zinc is conferred solely by the cysteine ligands.

The solution structure of the core domain of RbpA from *M. tuberculosis* and *S. coelicolor*

To investigate why $RbpA^{Sc}$ and $RbpA^{Mt}$ differ in their zinc content and to dissect the structure–function relationship of RbpA, we used NMR spectroscopy to determine the solution structure of each protein. The dispersion of resonances in proton NMR spectra recorded on both RbpA proteins clearly showed that the proteins were at least partially structured (data not shown), although the *M. tuberculosis* protein gave better spectra and was, therefore, our initial focus. To determine the structured portions of the protein, attempts were made to assign backbone amides. Despite efforts to prevent proteolysis, the samples degraded rapidly, and complete assignments were obtained only for the core domain. Resonance line widths and chemical shift dispersion indicate that regions outside these domains are highly disordered. The solution structure of the core domain was then determined using standard multidimensional NMR spectroscopy methods. After this, we determined the structure of the $RbpA^{Sc}$ core domain using a truncated clone that included $RbpA^{Sc}$ residues 1–75. The structures of the RbpA orthologues (Figure 5; see Supplementary Table S3 for NMR and refinement statistics) reveal a compact core domain (residues 28–72; designated the RbpA core domain, RCD) comprising four β strands that form a partially open β barrel. Residues 29–34, 37–43, 53–56 and 62–67 make up strands β1, β2, β3 and β4, respectively. In $RbpA^{Sc}$, the three Cys residues cluster together with His38 to form the likely metal-binding site, which is, together with the presence of Zn, consistent with a CHCC Zn-ribbon fold, a major subclass of the zinc-finger family, particularly common in proteins involved in translation and transcription (45). In $RbpA^{Mt}$, in place of the key residues (Cys34, His38, Cys56 and Cys59) are the following conserved hydrophilic residues: Thr34, Glu38, Cys56 and Asn58, which form a stabilizing hydrogen bond network (Figure 5E and F). Interestingly, although the RCD domain of $RbpA^{Sc}$ is largely negatively charged, the flexible N-terminal tail (1–26) is positively charged (Figure 5G), which may implicate this region in an interaction with DNA in transcription initiation complexes. Although the classical zinc-finger structure in the $RbpA^{Sc}$ RCD may suggest that it is specialized for DNA binding, the electrostatic potential and the absence of the Zn binding in other orthologues suggest that it is primarily a structural scaffold. The β-barrel structure is seen in other proteins (Supplementary Figure S2), including domains from mini-chromosome maintenance complexes (46), ribosomal protein L27 (47) and polypeptide chain release factors (48). Although the precise role of this domain in these proteins is unclear, they, along with adjacent sequences, seem to contribute to stabilizing either protein–protein or protein–RNA interactions. Interestingly, for example, the ribosomal protein L27 has a disordered N-terminal region (residues 1–19) in the oligonucleotide-free form, but in complex it forms an extended structure and interacts with the acceptor stem of the P-site tRNA and contributes to an increase in catalysis (49).

The C-terminal region of $RbpA^{Sc}$ is required for interaction with σ^{HrdB}

Although the NMR structures of the RbpA core domain did not include the C-terminal region, and indeed suggested that this region was disordered, sequence-based structural predictions using the Phyre 2 package (50) suggested the presence of two α-helices (designated α1 and α2), the second of which would be particularly amphipathic with a negatively charged face (Figure 4). To identify the region of $RbpA^{Sc}$ that interacts with σ^{HrdB}, we used BACTH to test three truncated *S. coelicolor rbpA*-T18 fusions for interaction with

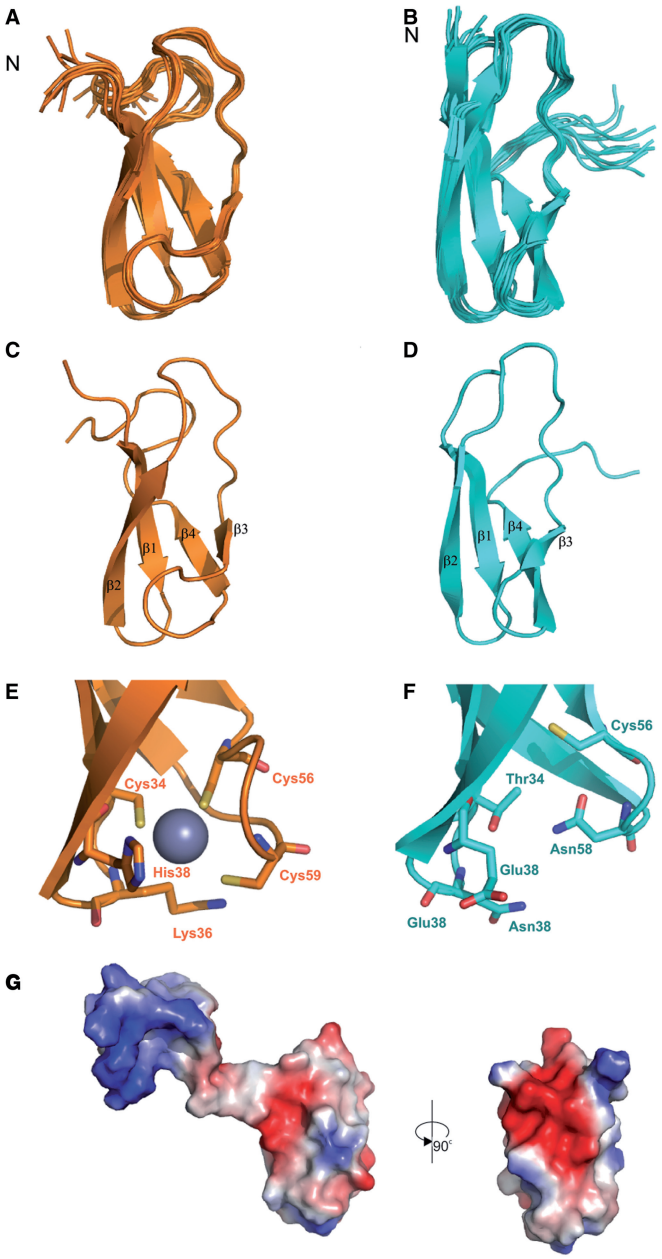

Figure 5. Solution structures of RbpA from *S. coelicolor* ($RbpA^{Sc}$) and *M. tuberculosis* ($RbpA^{Mt}$). NMR ensemble of the structure calculation for $RbpA^{Sc}$ (orange, **A**) and $RbpA^{Mt}$ (cyan, **B**). Ribbon representation of the structure of $RbpA^{Sc}$ (**C**) and $RbpA^{Mt}$ (**D**). Comparison of the Zn-binding site of $RbpA^{Sc}$ (**E**) with the equivalent residues in $RbpA^{Mt}$ (**F**). Electrostatic surface potential for $RbpA^{Sc}$ (**G**) shown in two orientations. The flexible N-terminus in $RbpA^{Sc}$ is included in the orientation shown on the left and removed for clarity on the right.

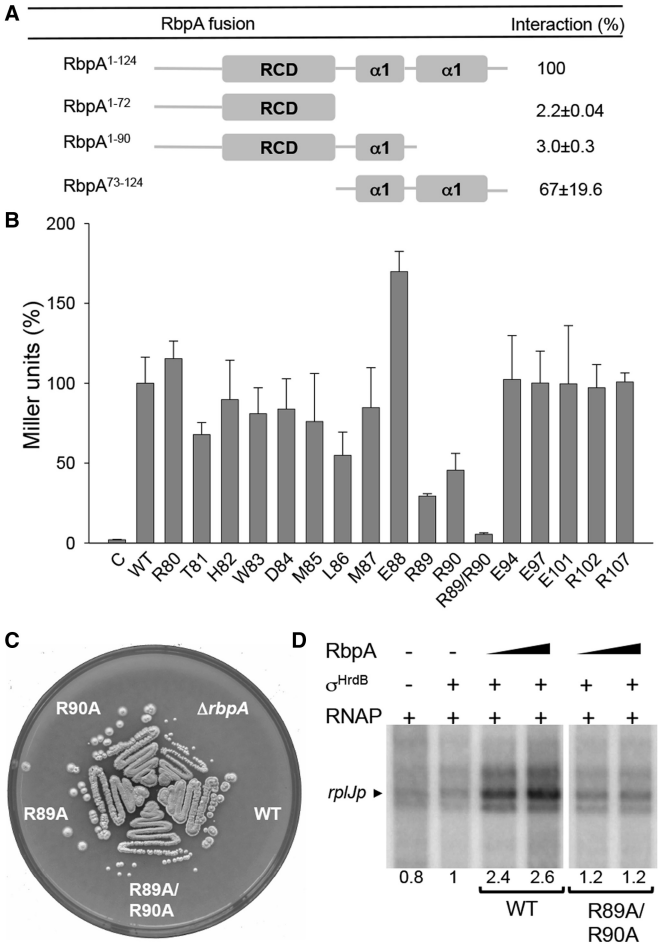

Figure 6. The C-terminal region of $RbpA^{Sc}$ is necessary and sufficient for σ^{HrdB} interaction. (**A**) A schematic diagram indicating the regions of $RbpA^{Sc}$ tested for interaction with σ^{HrdB} using BACTH analysis. The $RbpA^{Sc}$ fragments and σ^{HrdB} σ_2–σ_4 were fused to the T18 and T25 domains of *B. pertussis* adenylate cyclase, respectively. (**B**) BACTH analysis of alanine point mutations in conserved residues in the C-terminal region of $RbpA^{Sc}$. C, control strain containing pKT25–*hrdB* and pUT18. For (A) and (B), experiments were performed in triplicate (standard deviations indicated), and data are presented as % Miller units relative to results obtained with the full-length wild-type $RbpA^{Sc}$. Control strains with pKT25–*hrdB* and pUT18 exhibited <1% the activity of the full-length $RbpA^{Sc}$ interaction. (**C**) Growth of S129 (Δ*rbpA*), S129 containing pSETΩ::*rbpA* (WT), or equivalent constructs with the *rbpA* mutations as indicated. Strains were streaked to MS agar and photographed after 4 days incubation at 30°C. (**D**) *In vitro* transcription from the *rplJ* promoter in the presence of $RbpA^{Sc}$ or $RbpA^{Sc}$ R89A/R90A mutant proteins. Multi-round *in vitro* transcription reactions contained core RNAP (50 nM) σ^{HrdB} (250 nM), $RbpA^{Sc}$ (250 nM or 1.25 μM) and a DNA template generated by PCR using primers listed in Supplementary Table S2. Transcript levels were quantified by phosphorimaging, and data are presented relative to reactions performed in the presence of σ^{HrdB} but the absence of RbpA.

T25-σ^{HrdB} (Figure 6): $RbpA^{Sc}$ lacking both predicted C-terminal helices (residues 1–72, $RbpA^{1-72}$); $RbpA^{Sc}$ lacking only α2 ($RbpA^{1-90}$); and $RbpA^{Sc}$ lacking the N-terminal region and RCD domain but including α1 and α2 ($RbpA^{73-124}$). In contrast to $RbpA^{1-72}$ and $RbpA^{1-90}$ for which an interaction with σ^{HrdB} was not detected, $RbpA^{73-124}$ did interact, indicating that the C-terminal helical region of $RbpA^{Sc}$ is necessary and sufficient for binding. To identify critical residues involved in this interaction, we performed alanine-scanning mutagenesis, focusing on highly conserved residues in the C-terminal region. Apparent binding effects were observed for the highly conserved ERR motif (amino acid residues 88–90). Although binding seemed to be improved for the E88A mutation, binding was significantly reduced for R89A and R90A mutants, and a double R89A/R90A

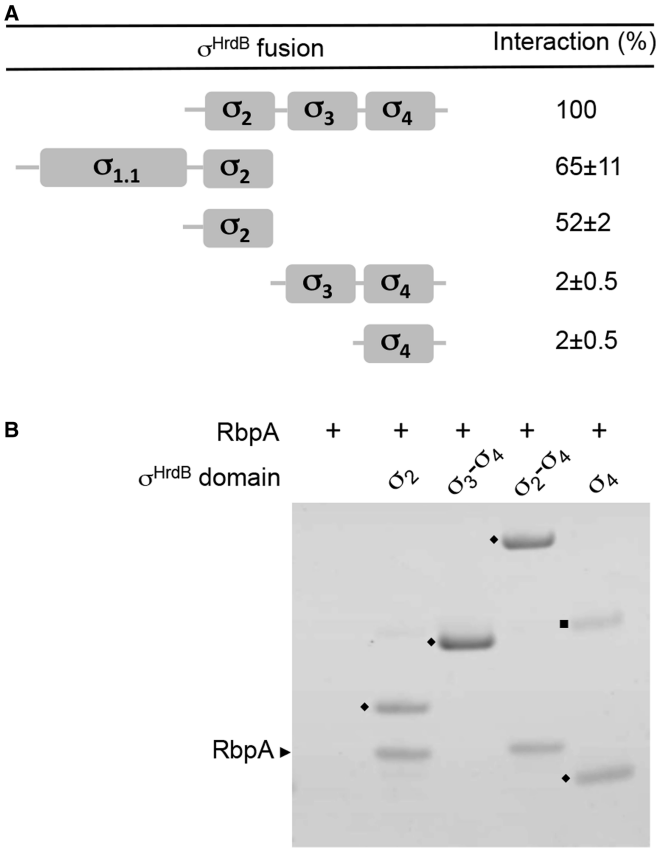

Figure 7. RbpA^{Sc} binds to the σ_2 domain of σ^{HrdB}. (**A**) A schematic diagram indicating the four conserved regions/globular domains of σ^{HrdB} together with BACTH interaction data between truncated T25–σ^{HrdB} fusions and RbpA^{Sc}–T18. The T25–σ^{HrdB} fusions included the following amino acids: $\sigma_{1.1}$–σ_2, 1–347; σ_2, 211–347; σ_3–σ_4, 348–511; σ_2–σ_4, 211–511; σ_4, 435–511. Control strains with pKT25 and pUT18–*rbpA* exhibited <1% the activity of the σ_2–σ_4 interaction. Experiments were performed in triplicate, and standard deviations are indicated. (**B**) Interaction between His$_6$-tagged σ^{HrdB} fragments and RbpA^{Sc} as judged using *in vitro* pull-down experiments. Purified His$_6$-tagged σ^{HrdB} fragments (closed diamond) were mixed with RbpA^{Sc} (~0.5–1 μM each) before purification using Ni-affinity magnetic beads. Eluted proteins were separated by 4–12% Bis–Tris SDS–PAGE and stained using Coomassie brilliant blue. RbpA^{Sc} is indicated with a black arrowhead. Closed square, unknown contaminating protein.

mutant was completely defective in σ^{HrdB} binding. To test the importance of this interaction in the function of RbpA^{Sc}, we cloned the mutant genes into the integrative plasmid pSETΩ and tested for complementation of *S. coelicolor* S129 (Δ*rbpA::apr*). Although the introduction of wild-type *rbpA* and the E88A mutant (data not shown) completely restored normal growth to S129, the *rbpA* R89A/R90A mutant did not. The individual R89A and R90A mutants also failed to completely restore normal growth, each exhibiting a slight, but reproducible, reduction in colony size and a delay in sporulation timing. As the inability of the R89A/R90A double mutant to complement *S. coelicolor* S129 might be due to protein instability *in vivo*, we overexpressed the protein in *E. coli* and tested whether the pure protein would stimulate transcription *in vitro*. Using the *rplJp* promoter as template, the RbpA^{Sc} R89A/R90A mutant failed to stimulate transcription, which suggests that a direct interaction with σ^{HrdB} is required for the action of RbpA^{Sc}.

RbpA binds to the σ_2 domain of the principal sigma factor

The different domains of σ have distinct functions; therefore, to better understand the role of RbpA in transcription initiation, we sought to localize the interaction to individual domain(s) of σ^{HrdB}. The σ_2, σ_3 and σ_4 structural domains of σ^{HrdB} were predicted based on the structures of σ^A from *Thermus* sp. (11,51), and five overlapping T25–*hrdB* fusions were constructed and tested for interaction with the *rbpA*–T18 fusion in BACTH analysis (Figure 7). No interaction was detected for T25–*hrdB* (σ_3–σ_4) or T25–*hrdB* (σ_4). However, all fragments that included σ_2 interacted, including T25–*hrdB* (σ_2) (amino acid residues 211–347), which comprises conserved regions 1.2–2.4. To confirm these data, each of the *hrdB* constructs used in the BACTH analysis was overexpressed with an N-terminal His$_6$-tag, and the corresponding proteins purified, then tested for interaction with RbpA^{Sc} using an *in vitro* pull-down assay (apart from $\sigma_{1.1}$–σ_2, which was insoluble). The pure σ^{HrdB} fragments (domains σ_2, σ_2–σ_4, σ_3–σ_4 and σ_4) were mixed with native RbpA^{Sc} for 15 min and then isolated using magnetic Ni-affinity beads. RbpA^{Sc} was co-isolated with the σ_2 and σ_2–σ_4 fragments of σ^{HrdB} but not with σ_3–σ_4 and σ_4, confirming a direct and specific interaction with the σ_2 domain. Using both BACTH and *in vitro* pull downs, we also detected equivalent interactions between the σ_2 domain of σ^A and the RbpA^{Mt} (data not shown). Furthermore, when equivalent σ_2 and RbpA proteins from *S. coelicolor* and *M. tuberculosis* are co-expressed in *E. coli*, with one partner His$_6$-tagged, they co-elute as a complex during Ni-affinity chromatography (data not shown). Taken together, these data suggest that the σ_2–RbpA interaction involving the principal σ factor is conserved across the actinobacteria.

DISCUSSION

Here, we provide an explanation for why the RNA polymerase-binding protein RbpA stimulates transcription from promoters that are dependent on the principal σ factor in *S. coelicolor* and *M. tuberculosis* (σ^{HrdB} and σ^A), but not from promoters dependent on alternative σ factors. Although RbpA^{Sc} was originally discovered as a component of RNAP holoenzyme, and it is demonstrated here to be present at transcription initiation complexes in *S. coelicolor*, we found that RbpA^{Sc} can also form a complex with the free principal σ factor. We failed to detect RbpA^{Sc} interactions with selected Group 3 (σ^B and σ^{WhiG}) and Group 4 (σ^E and σ^R) σ factors in *S. coelicolor*, although we found that RbpA^{Sc} and RbpA^{Mt} can bind to the Group 2 σ factors, σ^{HrdA} in *S. coelicolor* and σ^B in *M. tuberculosis*, respectively. In *S. coelicolor*, the three Group 2 σ factors (σ^{HrdA}, σ^{HrdC} and σ^{HrdD}) are collectively non-essential (52), as is σ^B in *M. tuberculosis* (53), which implies that the interaction with, and activation of, the essential primary σ factor underlies the

critical importance of RbpA for growth. Consistent with this, an $RbpA^{Sc}$ mutant that failed to bind σ^{HrdB} ($RbpA^{R89A/R90A}$) did not stimulate transcription *in vitro* and could not complement an *S. coelicolor rbpA* null mutant. Given that σ^{HrdB}–$RbpA^{Sc}$ complexes are stably maintained during gel filtration and Ni-affinity chromatography, we speculate that the two proteins are able to form an initial binary complex before the formation of holoenzyme. One role of RbpA might, therefore, be to act in a chaperone-like fashion to remodel σ for optimal binding to core RNAP, and thereby be an integral component of the 'σ cycle' in actinobacteria. This is consistent with the observation that $RbpA^{Mt}$ stabilizes *M. tuberculosis* σ^{A}–RNAP holoenzyme, but it argues for a more direct role rather than a proposed allosteric effect through changing the conformation of the β lobes (25). Such a role might account for the induction of *rbpA* during stress, as also proposed by others (25). Although *rbpA* is continually expressed in unstressed *S. coelicolor* cultures, consistent with its general importance for growth (23), transcription is also induced in response to disulphide stress, directed by the alternative ECF σ factor σ^{R} (23). In *M. tuberculosis,* the *rbpA* gene is also induced by stress through the σ^{R} orthologue, σ^{H} (54). The induction of *rbpA* might, therefore, ensure that the increased cellular levels of free alternative σ factors during stress do not adversely affect σ^{HrdB}/σ^{A}-dependent 'house keeping' gene expression, and it may also help the cell to reset the normal state once the stress has been resolved. Interestingly, it was recently discovered that the gene that encodes σ^{HrdB} is also controlled by two ECF σ factors: σ^{ShbA} during normal growth and σ^{R} during disulphide stress, which would also help to maintain σ^{HrdB} activity when stress-induced alternative σ factors accumulate in the cell (55,56). Furthermore, the Group 2 σ factors σ^{HrdD} in *S. coelicolor* and σ^{B} in *M. tuberculosis* are induced by stress (23,57), which might also contribute to maintaining the 'house keeping' duties of RNAP during stress.

Although RbpA might play an important role in promoting the formation of fully active σ^{HrdB}/σ^{A} holoenzyme, we show that RbpA is also poised to play additional roles in the subsequent stages of transcription initiation, such as promoter binding, and the formation and stability of open complexes. Our ChIP–qPCR data suggest that $RbpA^{Sc}$ is present at RNAP–promoter complexes *in vivo* that have been trapped using rifampicin. As rifampicin does not seem to influence promoter binding or open complex formation, but rather blocks transcription when the nascent transcript is ~3 nt in length (58), RbpA might play a role in any of the preceding stages of initiation. We mapped the primary binding site of $RbpA^{Sc}$ to the σ_2 domain of σ^{HrdB} (conserved regions 1.2 and 2.1–2.4), although we cannot rule out additional weaker or transient contacts elsewhere on the protein. The σ_2 domain has several key roles in transcription initiation. As well as forming the largest interface between σ and RNAP (10), it confers promoter specificity for the −10 element and provides aromatic side chains that nucleate and maintain the strand-separated state during the formation of the open complex (6,59). Furthermore, variation in the sequence of the discriminator element that lies between the −10 element and the +1 transcription start point can influence stabilizing interactions with σ region 1.2, which is exploited by factors such as ppGpp that further destabilize open complexes (60). An understanding of the mechanism of action of RbpA is an important future goal.

In a similar manner to RbpA, the Crl protein of enteric bacteria interacts with the σ_2 domain of the stationary phase σ factor σ^{S} and stimulates holoenzyme formation, thereby contributing to the ability of σ^{S} to compete with σ^{70} during stationary phase or under stress, despite its lower abundance (19,21). Crl seems to be specific for σ^{S}, and although *crl* mutants were originally discovered by their defect in curli fimbriae formation, they are in fact defective in the expression of a large subset of σ^{S}-dependent genes. Nonetheless, it was found that Crl also improved the activity of σ^{70}–RNAP when mixed with σ^{70} before formation of holoenzyme (18), suggesting that Crl might also play a non-essential role in formation of the vegetative holoenzyme. Although Crl and RbpA are not obviously related in protein sequence, the two proteins might play an analogous role in transcription initiation, for example, through modulating the interaction of σ factors with RNAP and promoting the formation and activation of certain holoenzyme complexes.

How do we reconcile our observations that RbpA interacts with the σ subunit with previous proposals that $RbpA^{Mt}$ principally interacts with the β-subunit of RNAP (25,26)? First, it should be noted that previous studies involved attaching chemical probes to the single cysteine residue in $RbpA^{Mt}$ located in the RCD, which does not interact with σ. Moreover, the cysteine is likely to be important in proper folding of the RCD (Figure 5), which might have caused anomalous results, although evidence was provided that the labelled proteins were functional. Alternatively, and perhaps more likely, RbpA might interact with both σ and β, through distinct regions of RbpA. For example, although the σ_2 and the β sandwich-barrel hybrid motif are distantly located in the holoenzyme structure, the disordered regions of RbpA would have significant conformational freedom to bridge between these regions. In the structurally related ribosomal protein L27, the N-terminal 1–19 residues are unstructured in its free form, but extend to a distance of >35 Å from the core domain when assembled with the ribonucleoprotein complex.

The RCD domains of *S. coelicolor* and *M. tuberculosis* have virtually identical structures, but differ in their ability to bind zinc. This correlates with the replacement of three of the likely zinc ligands in $RbpA^{Sc}$ with hydrophilic residues in $RbpA^{Mt}$ that stabilize the domain through a hydrogen bond network. As the zinc-coordinating ligands are only absent from the *Corynebacterium–Mycobacterium–Nocardia* (CMN) group of actinobacteria, but are present in certain deeply rooted actinobacteria, such as *Bifidobacterium* sp. (61), it is likely that zinc-free RbpA evolved soon after the emergence of the CMN group. The RbpA proteins, therefore, present an example of how the evolution of a structural domain can contribute to a reduction in zinc requirements, which has also been described for alternative

zinc-free ribosomal proteins (62,63). However, it is unlikely that this adaptation is linked to pathogenesis because the majority of CMN group comprises free-living non-pathogenic strains. Although mutations of the three zinc-coordinating cysteine residues seemed to inactivate or destabilize $RbpA^{Sc}$, the RCD domain does not seem to be required for σ interaction. BACTH analysis and alanine scanning mutagenesis of $RbpA^{Sc}$ revealed that the C-terminal predicted helical region was required for this interaction, and that the highly conserved amino acid residues R89 and R90 play a critical role, although it is not yet clear whether they contact σ directly, or whether their mutation disrupts the overall conformation of this region. Although the C-terminal region is predicted to be helical, NMR analysis of the full-length RbpA proteins from both *S. coelicolor* and *M. tuberculosis* suggested that this region in each protein is unstructured, and despite extensive attempts to prevent it, it is readily proteolysed to the RCD during storage. The binding to σ, therefore, likely promotes folding of the RbpA C-terminal region, thereby stabilizing the complex. Clearly, future structural characterization of RbpA either in a binary complex with σ or in higher order complexes with RNAP will be highly valuable in understanding further mechanistic aspects of the function of this protein. This is likely to be of particular importance in the development of new approaches to target the process of transcription in *M. tuberculosis* and other pathogenic actinobacteria.

ACCESSION NUMBERS

Atomic coordinates for $RbpA^{Sc}$ and $RbpA^{Mt}$ have been deposited with the Protein Data Bank under accession codes 2m6o and 2m6p, respectively. NMR chemical shifts for $RbpA^{Sc}$ and $RbpA^{Mt}$ have been deposited with the Biological Magnetic Resonance DataBank under accession codes 19148 and 19149, respectively.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–3, Supplementary Figures 1 and 2 and Supplementary References [64–73].

ACKNOWLEDGEMENTS

The authors are grateful to Chris Dadswell, Sophie Buitenhuis and Yichao Zhou for contributions to experiments.

FUNDING

A.T.-S. B.L. and R.J.L. were funded by BBSRC grant [BB/I003045 to M.S.P. and S.J.M.]; P.D. was funded by a BBSRC grant [BB/D018293/1 to M.S.P.]; A.T.-S. was funded by the University of Sussex and the Overseas Research Students Awards Scheme. Funding for open access charge: BBSRC grant [BB/1003045].

Conflict of interest statement. None declared.

REFERENCES

1. WHO. (2012) *Global Tuberculosis Report*. WHO Press, Geneva.
2. Hu,H., Zhang,Q. and Ochi,K. (2002) Activation of antibiotic biosynthesis by specified mutations in the *rpoB* gene (encoding the RNA polymerase beta subunit) of *Streptomyces lividans*. *J. Bacteriol.*, **184**, 3984–3991.
3. Zhuo,Y., Zhang,W., Chen,D., Gao,H., Tao,J., Liu,M., Gou,Z., Zhou,X., Ye,B.C., Zhang,Q. *et al.* (2010) Reverse biological engineering of *hrdB* to enhance the production of avermectins in an industrial strain of *Streptomyces avermitilis*. *Proc. Natl Acad. Sci. USA*, **107**, 11250–11254.
4. Saecker,R.M., Record,M.T. and Dehaseth,P.L. (2011) Mechanism of bacterial transcription initiation: RNA polymerase—promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *J. Mol. Biol.*, **412**, 754–771.
5. Raffaelle,M., Kanin,E.I., Vogt,J., Burgess,R.R. and Ansari,A.Z. (2005) Holoenzyme switching and stochastic release of sigma factors from RNA polymerase *in vivo*. *Mol. Cell*, **20**, 357–366.
6. Feklistov,A. and Darst,S.A. (2011) Structural basis for promoter-10 element recognition by the bacterial RNA polymerase σ subunit. *Cell*, **147**, 1257–1269.
7. Burgess,R.R., Travers,A.A., Dunn,J.J. and Bautz,E.K. (1969) Factor stimulating transcription by RNA polymerase. *Nature*, **221**, 43–46.
8. Gruber,T.M. and Gross,C.A. (2003) Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu. Rev. Microbiol.*, **57**, 441–466.
9. Young,B.A., Gruber,T.M. and Gross,C.A. (2002) Views of transcription initiation. *Cell*, **109**, 417–420.
10. Murakami,K.S., Masuda,S. and Darst,S.A. (2002) Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science*, **296**, 1280–1284.
11. Vassylyev,D.G., Sekine,S., Laptenko,O., Lee,J., Vassylyeva,M.N., Borukhov,S. and Yokoyama,S. (2002) Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature*, **417**, 712–719.
12. Schwartz,E.C., Shekhtman,A., Dutta,K., Pratt,M.R., Cowburn,D., Darst,S. and Muir,T.W. (2008) A full-length group 1 bacterial sigma factor adopts a compact structure incompatible with DNA binding. *Chem. Biol.*, **15**, 1091–1103.
13. Paget,M.S. and Helmann,J.D. (2003) The σ^{70} family of sigma factors. *Genome Biol.*, **4**, 203.
14. Österberg,S., del Peso-Santos,T. and Shingler,V. (2011) Regulation of alternative sigma factor use. *Annu. Rev. Microbiol.*, **65**, 37–55.
15. Pratt,L.A. and Silhavy,T.J. (1998) Crl stimulates RpoS activity during stationary phase. *Mol. Microbiol.*, **29**, 1225–1236.
16. Bougdour,A., Lelong,C. and Geiselmann,J. (2004) Crl, a low temperature-induced protein in *Escherichia coli* that binds directly to the stationary phase sigma subunit of RNA polymerase. *J. Biol. Chem.*, **279**, 19540–19550.
17. England,P., Westblade,L.F., Karimova,G., Robbe-Saule,V., Norel,F. and Kolb,A. (2008) Binding of the unorthodox transcription activator, Crl, to the components of the transcription machinery. *J. Biol. Chem.*, **283**, 33455–33464.
18. Gaal,T., Mandel,M.J., Silhavy,T.J. and Gourse,R.L. (2006) Crl facilitates RNA polymerase holoenzyme formation. *J. Bacteriol.*, **188**, 7966–7970.
19. Monteil,V., Kolb,A., Mayer,C., Hoos,S., England,P. and Norel,F. (2010) Crl binds to domain 2 of σ^S and confers a competitive advantage on a natural *rpoS* mutant of *Salmonella enterica* serovar Typhi. *J. Bacteriol.*, **192**, 6401–6410.
20. Monteil,V., Kolb,A., D'Alayer,J., Beguin,P. and Norel,F. (2010) Identification of conserved amino acid residues of the *Salmonella* σ^S chaperone Crl involved in Crl-σ^S interactions. *J. Bacteriol.*, **192**, 1075–1087.
21. Typas,A., Barembruch,C., Possling,A. and Hengge,R. (2007) Stationary phase reorganisation of the *Escherichia coli* transcription machinery by Crl protein, a fine-tuner of sigmas activity and levels. *EMBO J.*, **26**, 1569–1578.
22. Newell,K.V., Thomas,D.P., Brekasis,D. and Paget,M.S. (2006) The RNA polymerase-binding protein RbpA confers basal levels

of rifampicin resistance on *Streptomyces coelicolor*. *Mol. Microbiol.*, **60**, 687–696.
23. Paget,M.S., Molle,V., Cohen,G., Aharonowitz,Y. and Buttner,M.J. (2001) Defining the disulphide stress response in *Streptomyces coelicolor* A3(2): identification of the σ^R regulon. *Mol. Microbiol.*, **42**, 1007–1020.
24. Forti,F., Mauri,V., Dehò,G. and Ghisotti,D. (2011) Isolation of conditional expression mutants in *Mycobacterium tuberculosis* by transposon mutagenesis. *Tuberculosis (Edinb.)*, **91**, 569–578.
25. Hu,Y., Morichaud,Z., Chen,S., Leonetti,J.P. and Brodolin,K. (2012) *Mycobacterium tuberculosis* RbpA protein is a new type of transcriptional activator that stabilizes the σ^A-containing RNA polymerase holoenzyme. *Nucleic Acids Res.*, **40**, 6547–6557.
26. Dey,A., Verma,A.K. and Chatterji,D. (2010) Role of an RNA polymerase interacting protein, MsRbpA, from *Mycobacterium smegmatis* in phenotypic tolerance to rifampicin. *Microbiology*, **156**, 873–883.
27. Kieser,T., Bibb,M.J., Buttner,M.J., Chater,K.F. and Hopwood,D.A. (2000) *Practical Streptomyces Genetics*. The John Innes Foundation, Norwich.
28. Studier,F.W. (2005) Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.*, **41**, 207–234.
29. Sivashanmugam,A., Murray,V., Cui,C., Zhang,Y., Wang,J. and Li,Q. (2009) Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*. *Protein Sci.*, **18**, 936–948.
30. Hsu,L.M. (2009) Monitoring abortive initiation. *Methods*, **47**, 25–36.
31. Zhang,X. and Bremer,H. (1995) Control of the *Escherichia coli rrnB* P1 promoter strength by ppGpp. *J. Biol. Chem.*, **270**, 11181–11189.
32. Bucca,G., Laing,E., Mersinias,V., Allenby,N., Hurd,D., Holdstock,J., Brenner,V., Harrison,M. and Smith,C.P. (2009) Development and application of versatile high density microarrays for genome-wide analysis of *Streptomyces coelicolor*: characterization of the HspR regulon. *Genome Biol.*, **10**, R5.
33. McClellan,M.J., Khasnis,S., Wood,C.D., Palermo,R.D., Schlick,S.N., Kanhere,A.S., Jenner,R.G. and West,M.J. (2012) Downregulation of integrin receptor-signaling genes by Epstein-Barr virus EBNA 3C via promoter-proximal and -distal binding elements. *J. Virol.*, **86**, 5165–5178.
34. Sattler,M., Schleucher,J. and Griesinger,C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.*, **34**, 93–158.
35. Marchant,J., Sawmynaden,K., Saouros,S., Simpson,P. and Matthews,S. (2008) Complete resonance assignment of the first and second apple domains of MIC4 from *Toxoplasma gondii*, using a new NMRView-based assignment aid. *Biomol. NMR Assign.*, **2**, 119–121.
36. Rieping,W., Habeck,M., Bardiaux,B., Bernard,A., Malliavin,T. and Nilges,M. (2007) ARIA2: automated NOE assignment and data integration in NMR structure calculation. *Bioinformatics*, **23**, 381–382.
37. Cornilescu,G., Delaglio,F. and Bax,A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR*, **13**, 289–302.
38. Viollier,P.H., Nguyen,K.T., Minas,W., Folcher,M., Dale,G.E. and Thompson,C.J. (2001) Roles of aconitase in growth, metabolism, and morphological differentiation of *Streptomyces coelicolor*. *J. Bacteriol.*, **183**, 3193–3203.
39. van Wezel,G.P., Takano,E., Vijgenboom,E., Bosch,L. and Bibb,M.J. (1995) The *tuf3* gene of *Streptomyces coelicolor* A3(2) encodes an inessential elongation factor Tu that is apparently subject to positive stringent control. *Microbiology*, **141**, 2519–2528.
40. Chakraburtty,R., White,J., Takano,E. and Bibb,M. (1996) Cloning, characterization and disruption of a (p)ppGpp synthetase gene (relA) of *Streptomyces coelicolor* A3(2). *Mol. Microbiol.*, **19**, 357–368.
41. Vockenhuber,M.P., Sharma,C.M., Statt,M.G., Schmidt,D., Xu,Z., Dietrich,S., Liesegang,H., Mathews,D.H. and Suess,B. (2011) Deep sequencing-based identification of small non-coding RNAs in *Streptomyces coelicolor*. *RNA Biol.*, **8**, 468–477.
42. Bergendahl,V., Thompson,N.E., Foley,K.M., Olson,B.M. and Burgess,R.R. (2003) A cross-reactive polyol-responsive monoclonal antibody useful for isolation of core RNA polymerase from many bacterial species. *Protein Expr. Purif.*, **31**, 155–160.
43. Herring,C.D., Raffaelle,M., Allen,T.E., Kanin,E.I., Landick,R., Ansari,A.Z. and Palsson,B. (2005) Immobilization of *Escherichia coli* RNA polymerase and location of binding sites by use of chromatin immunoprecipitation and microarrays. *J. Bacteriol.*, **187**, 6166–6174.
44. Karimova,G., Pidoux,J., Ullmann,A. and Ladant,D. (1998) A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proc. Natl Acad. Sci. USA*, **95**, 5752–5756.
45. Krishna,S.S., Majumdar,I. and Grishin,N.V. (2003) Structural classification of zinc fingers: survey and summary. *Nucleic Acids Res.*, **31**, 532–550.
46. Fletcher,R.J., Bishop,B.E., Leon,R.P., Sclafani,R.A., Ogata,C.M. and Chen,X.S. (2003) The structure and function of MCM from archaeal *M. Thermoautotrophicum*. *Nat. Struct. Biol.*, **10**, 160–167.
47. Wang,H., Takemoto,C.H., Murayama,K., Sakai,H., Tatsuguchi,A., Terada,T., Shirouzu,M., Kuramitsu,S. and Yokoyama,S. (2004) Crystal structure of ribosomal protein L27 from *Thermus thermophilus* HB8. *Protein Sci.*, **13**, 2806–2810.
48. Mantsyzov,A.B., Ivanova,E.V., Birdsall,B., Alkalaeva,E.Z., Kryuchkova,P.N., Kelly,G., Frolova,L.Y. and Polshakov,V.I. (2010) NMR solution structure and function of the C-terminal domain of eukaryotic class 1 polypeptide chain release factor. *FEBS J.*, **277**, 2611–2627.
49. Selmer,M., Dunham,C.M., Murphy,F.V., Weixlbaumer,A., Petry,S., Kelley,A.C., Weir,J.R. and Ramakrishnan,V. (2006) Structure of the 70S ribosome complexed with mRNA and tRNA. *Science*, **313**, 1935–1942.
50. Kelley,L.A. and Sternberg,M.J. (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.*, **4**, 363–371.
51. Campbell,E.A., Muzzin,O., Chlenov,M., Sun,J.L., Olson,C.A., Weinman,O., Trester-Zedlitz,M.L. and Darst,S.A. (2002) Structure of the bacterial RNA polymerase promoter specificity sigma subunit. *Mol. Cell*, **9**, 527–539.
52. Buttner,M.J., Chater,K.F. and Bibb,M.J. (1990) Cloning, disruption, and transcriptional analysis of three RNA polymerase sigma factor genes of *Streptomyces coelicolor* A3(2). *J. Bacteriol.*, **172**, 3367–3378.
53. Fontán,P.A., Voskuil,M.I., Gomez,M., Tan,D., Pardini,M., Manganelli,R., Fattorini,L., Schoolnik,G.K. and Smith,I. (2009) The *Mycobacterium tuberculosis* sigma factor σ^B is required for full response to cell envelope stress and hypoxia in vitro, but it is dispensable for *in vivo* growth. *J. Bacteriol.*, **191**, 5628–5633.
54. Manganelli,R., Voskuil,M.I., Schoolnik,G.K., Dubnau,E., Gomez,M. and Smith,I. (2002) Role of the extracytoplasmic–function sigma factor σ^H in *Mycobacterium tuberculosis* global gene expression. *Mol. Microbiol.*, **45**, 365–374.
55. Otani,H., Higo,A., Nanamiya,H., Horinouchi,S. and Ohnishi,Y. (2013) An alternative sigma factor governs the principal sigma factor in *Streptomyces griseus*. *Mol. Microbiol.*, **87**, 1223–1236.
56. Kim,M.S., Dufour,Y.S., Yoo,J.S., Cho,Y.B., Park,J.H., Nam,G.B., Kim,H.M., Lee,K.L., Donohue,T.J. and Roe,J.H. (2012) Conservation of thiol-oxidative stress responses regulated by SigR orthologues in actinomycetes. *Mol. Microbiol.*, **85**, 326–344.
57. Manganelli,R., Voskuil,M.I., Schoolnik,G.K. and Smith,I. (2001) The *Mycobacterium tuberculosis* ECF sigma factor σ^E: role in global gene expression and survival in macrophages. *Mol. Microbiol.*, **41**, 423–437.
58. Campbell,E.A., Korzheva,N., Mustaev,A., Murakami,K., Nair,S., Goldfarb,A. and Darst,S.A. (2001) Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. *Cell*, **104**, 901–912.
59. Zhang,Y., Feng,Y., Chatterjee,S., Tuske,S., Ho,M.X., Arnold,E. and Ebright,R.H. (2012) Structural basis of transcription initiation. *Science*, **338**, 1076–1080.
60. Haugen,S.P., Berkmen,M.B., Ross,W., Gaal,T., Ward,C. and Gourse,R.L. (2006) rRNA promoter regulation by nonoptimal

binding of sigma region 1.2: an additional recognition element for RNA polymerase. *Cell*, **125**, 1069–1082.
61. Ventura,M., Canchaya,C., Tauch,A., Chandra,G., Fitzgerald,G.F., Chater,K.F. and van Sinderen,D. (2007) Genomics of actinobacteria: tracing the evolutionary history of an ancient phylum. *Microbiol. Mol. Biol. Rev.*, **71**, 495–548.
62. Gabriel,S.E. and Helmann,J.D. (2009) Contributions of Zur-controlled ribosomal proteins to growth under zinc starvation conditions. *J. Bacteriol.*, **191**, 6116–6122.
63. Panina,E.M., Mironov,A.A. and Gelfand,M.S. (2003) Comparative genomics of bacterial zinc regulons: enhanced ion transport, pathogenesis, and rearrangement of ribosomal proteins. *Proc. Natl Acad. Sci. USA*, **100**, 9912–9917.
64. Kelemen,G.H., Plaskitt,K.A., Lewis,C.G., Findlay,K.C. and Buttner,M.J. (1995) Deletion of DNA lying close to the glkA locus induces ectopic sporulation in *Streptomyces coelicolor* A3(2). *Mol. Microbiol.*, **17**, 221–230.
65. Babcock,M.J., Buttner,M.J., Keler,C.H., Clarke,B.R., Morris,R.A., Lewis,C.G. and Brawner,M.E. (1997) Characterization of the rpoC gene of *Streptomyces coelicolor* A3(2) and its use to develop a simple and rapid method for the purification of RNA polymerase. *Gene*, **196**, 31–42.
66. Paget,M.S., Chamberlin,L., Atrih,A., Foster,S.J. and Buttner,M.J. (1999) Evidence that the extracytoplasmic function sigma factor sigmaE is required for normal cell wall structure in *Streptomyces coelicolor* A3(2). *J. Bacteriol.*, **181**, 204–211.
67. Studier,F.W. and Moffatt,B.A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.*, **189**, 113–130.
68. Paget,M.S., Hintermann,G. and Smith,C.P. (1994) Construction and application of streptomycete promoter probe vectors which employ the *Streptomyces glaucescens* tyrosinase-encoding gene as reporter. *Gene*, **146**, 105–110.
69. Alting-Mees,M.A. and Short,J.M. (1989) pBluescript II: gene mapping vectors. *Nucleic Acids Res.*, **17**, 9494.
70. Karimova,G., Ullmann,A. and Ladant,D. (2001) Protein-protein interaction between *Bacillus stearothermophilus* tyrosyl-tRNA synthetase subdomains revealed by a bacterial two-hybrid system. *J. Mol. Microbiol. Biotechnol.*, **3**, 73–82.
71. Bierman,M., Logan,R., O'Brien,K., Seno,E.T., Rao,R.N. and Schoner,B.E. (1992) Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to Streptomyces spp. *Gene*, **116**, 43–49.
72. O'Connor,T.J., Kanellis,P. and Nodwell,J.R. (2002) The ramC gene is required for morphogenesis in *Streptomyces coelicolor* and expressed in a cell type-specific manner under the direct control of RamR. *Mol. Microbiol.*, **45**, 45–57.
73. Huang,J., Shi,J., Molle,V., Sohlberg,B., Weaver,D., Bibb,M.J., Karoonuthaisiri,N., Lih,C.J., Kao,C.M., Buttner,M.J. *et al.* (2005) Crossregulation among disparate antibiotic biosynthetic pathways of *Streptomyces coelicolor*. *Mol. Microbiol.*, **58**, 1276–1287.