Identification and Structure of the Anti-sigma Factor-binding Domain of the Disulphide-stress Regulated Sigma Factor $\sigma^R$ from *Streptomyces coelicolor*

Wei Li, Clare E. M. Stevenson, Nicolas Burton, Piotr Jakimowicz, Mark S. B. Paget, Mark J. Buttner, David M. Lawson and Colin Kleanthous*

1School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK
2Department of Biological Chemistry, John Innes Centre, Norwich NR4 7IH, UK
3Department of Molecular Microbiology, John Innes Centre, Norwich NR4 7UH, UK
4School of Biological Sciences, University of Sussex, Falmer, Brighton BN1 9QG, UK

The extracytoplasmic function (ECF) sigma factor $\sigma^R$ is a global regulator of redox homeostasis in the antibiotic-producing bacterium *Streptomyces coelicolor*, with a similar role in other actinomycetes such as *Mycobacterium tuberculosis*. Normally maintained in an inactive state by its bound anti-sigma factor RsrA, $\sigma^R$ dissociates in response to intracellular disulphide-stress to direct core RNA polymerase to transcribe genes, such as *trxBA* and *trxC* that encode the enzymes of the thioredoxin disulphide reductase pathway, that re-establish redox homeostasis. Little is known about where RsrA binds on $\sigma^R$ or how it suppresses $\sigma^R$-dependent transcriptional activity. Using a combination of proteolysis, surface-enhanced laser desorption ionisation mass spectrometry and pull-down assays we identify an N-terminal, 10 kDa domain (sRN) that encompasses region 2 of $\sigma^R$ that represents the major RsrA binding site. We show that sRN inhibits transcription by an unrelated sigma factor and that this inhibition is relieved by RsrA binding, reaffirming that region 2 is involved in binding to core RNA polymerase but also demonstrating that the likely mechanism by which RsrA inhibits $\sigma^R$ activity is by blocking this association. We also report the 2.4 Å resolution crystal structure of sRN that reveals extensive structural conservation with the equivalent region of $\sigma^{70}$ from *Escherichia coli* as well as with the cyclin-box, a domain-fold found in the eukaryotic proteins TFIIB and cyclin A. $\sigma^{70}$ has a propensity to aggregate, due to steric complementarity of oppositely charged surfaces on the domain, but this is inhibited by RsrA, an observation that suggests a possible mode of action for RsrA which we compare to other well-studied sigma factor-anti-sigma factor systems.

© 2002 Elsevier Science Ltd. All rights reserved

*Corresponding author*

Keywords: structure; RsrA; *S. coelicolor*; sigma factor; X-ray crystallography

Introduction

Promoter recognition by bacterial RNA polymerase requires the core enzyme ($\alpha_2\beta\beta'\omega = E$) to associate with a specificity subunit called $\sigma$. The first $\sigma$ factor to be characterised was $\sigma^{70}$, the principal, essential $\sigma$ factor of *Escherichia coli*. Other members of the $\sigma^{70}$ family, known as alternative $\sigma$ factors, each confer a different promoter specificity on the holoenzyme (Eo), and bacteria use alternative $\sigma$ subunits to control expression of specific sets of genes (regulons). Alignment of members of the $\sigma^{70}$ family has identified four major conserved regions (1–4), each further divided into subregions. The binding motifs that recognise
the −10 and −35 promoter sequences are located in regions 2.4 and 4.2, respectively, whereas the interface of σ with core is more extensive, with regions 2–4 all involved.

It is common for bacteria to express alternative σ factors before they are needed, but maintain them in an inactive state until their release is triggered by appropriate conditions. This inhibition is achieved through the reversible action of inhibitor proteins known as anti-σ factors. Whereas all members of the σ70 family are homologous, anti-σ factors belong to a rapidly increasing number of phylogenetically unrelated families. Although σ factor:anti-σ factor interaction has been analysed in detail in only a few cases, most particularly σE:SpoIIAB in Bacillus subtilis, σ28:FlgM in Salmonella typhimurium and σ70:AsiA in E. coli, it is already clear that inhibition can be achieved through a variety of distinct mechanisms.

In Streptomyces coelicolor, σR activates expression of a regulon of >30 genes that help the bacterium survive “disulphide stress”, the unwanted formation of disulphide bonds in the cytoplasm. σR target genes include trxBA and trxC, which encode the enzymes of the thioredoxin disulphide reductase pathway. The activity of σR is controlled by the anti-σ factor RsrA, a redox-sensitive, zinc metalloprotein that modulates σR activity in response to changes in the thiol-disulphide redox balance of the cytoplasm. Under reducing conditions, RsrA binds to σR to form a 1:1 complex, preventing it from activating transcription. Oxidative stress induces intramolecular disulphide bond formation in RsrA, causing it to dissociate and thereby release σR to activate its target genes. Increased trxBA and trxC expression in turn leads to the thioredoxin-dependent reduction of oxidised RsrA back to its σR-binding conformation, shutting off σR-dependent transcription.

The σR–RsrA system also exists in other actinomycetes, including pathogens like Mycobacterium tuberculosis, where it is named σH-RshA. RsrA is the first-described member of a new and growing family of anti-σ factors, the zinc-containing anti-σ factor (ZAS) family, that control the activity of certain members of the extracytoplasmic function (ECF) subfamily of sigma factors.

Figure 1. Identification of σRN, the RsrA binding domain on σR. (a) Limited trypsin digestion profile of σR under native conditions. Three major fragments were generated, two from the N terminus (σRN) and one from the C terminus (σRC) that were mapped by N-terminal sequencing of blotted fragments (indicated by the positions of the arrows) and SELDI-MS; the C termini of the fragments are predicted on the basis of the cleavage specificity of trypsin and correlation of observed and expected molecular masses. The numbered sections of σR represent the characteristic regions of sigma factors (after Lonetto et al.). (b) SELDI-MS of σR and its proteolytic fragments. (i) The major σR tryptic fragments that are detected using a normal phase (NP20) chip, where both σRN and σRC are evident. (ii)–(iv) Data collected from IMAC3 (immobilized metal affinity capture) chips. In (ii), the chip was loaded with a mixture of His-RsrA and the σR limited tryptic digest (∼20 μM). In (iii), the σR tryptic digest was added without His-RsrA while in (iv) only His-RsrA was loaded onto the chip. In all cases the chips were washed extensively with buffer prior to laser desorption (see Materials and Methods). Only σRN fragments were retained on the His-RsrA-IMAC3 chip, indicating an association between the domain and the anti-sigma factor.
Characterisation of the RsrA-binding Domain of \( \sigma^R \)

Figure 2. Far-UV CD spectrum of overexpressed and purified His-\( \sigma^{RN} \). His-\( \sigma^{RN} \) (3.8 \( \mu M \)) was dialysed into 50 mM potassium phosphate buffer (pH 5.4) and the spectrum collected in a 1 mm pathlength cell from 190–260 nm. The resulting spectrum was baseline-subtracted and the secondary structure composition estimated as \( \sim 68\% \alpha \)-helix with little or no \( \beta \)-structure.

There is currently no information on how ZAS proteins inhibit their target sigma factors, and since the mechanisms of action of anti-sigma factors are diverse, it is impossible to extrapolate from other well-studied systems. Hence, we embarked on a series of biochemical experiments to delineate the binding site on \( \sigma^R \) for its ZAS protein, RsrA, and to demonstrate how this association inhibits \( \sigma^R \)-dependent transcriptional activity. Through this work we identify an RsrA binding domain on \( \sigma^R \), determine its three-dimensional structure by X-ray crystallography, and show that its structural conservation goes beyond the sigma factor family to include eukaryotic transcription factors and signalling proteins. Our data illustrate yet another variation on the rich mechanistic themes by which sigma factors are inhibited by their anti-sigma factor proteins.

**Results**

**Identification of the RsrA-binding domain on \( \sigma^R \)**

\( \sigma^R \) (25 kDa) and RsrA (12 kDa) were overproduced in *E. coli* and purified as described and in Materials and Methods. Both heterologously produced proteins are functional *in vitro*, demonstrated for \( \sigma^R \) through transcription run-off assays directed by \( \sigma^R \)-specific promoters, and for RsrA by suppression of this activity when in the reduced state.

Proteolysis has been used extensively to delineate anti-sigma factor binding sites on sigma factors, either through the identification of sites of protease protection or fragments that bind the anti-sigma factor protein. Adopting the later approach, we used proteolysis to define regions of \( \sigma^R \) that interact with RsrA, identifying initially the major trypsin-generated fragments of \( \sigma^R \) in the absence of its anti-sigma factor. Although there are 28 theoretical trypsin cleavage sites in the \( \sigma^R \) sequence we found that limited digestion with trypsin under native conditions (0.4\% (w/w) at 25 °C, pH 7.5 for 30 minutes) generated only three major species, as deduced by SDS-PAGE (data not shown). These were readily mapped to the sequence of \( \sigma^R \) through a combination of N-terminal sequencing of the blotted fragments and analysis of their molecular weights by surface-enhanced laser desorption ionisation mass spectrometry (SELDI-MS) (Figure 1(a) and (b), (i)). Two of the fragments originate from the N terminus of the protein and correspond to a 9.5–10 kDa domain that encompasses region 2 and part of region 1, which we term \( \sigma^{RN} \). The third fragment corresponds to a ~6 kDa domain that encompasses regions 3.2, 4.1 with some of 4.2 and comes from the C-terminal half of \( \sigma^R \), which we term \( \sigma^{RC} \). Together, the two proteolytically defined domains account for approximately two-thirds of the \( \sigma^R \) sequence (Figure 1(a)).

To determine if either of these domains interacts with RsrA we performed a pull-down assay using a SELDI-IMAC chip (see Materials and Methods). This approach capitalises on the ability to attach a bait protein to the metal-chelate surface of the IMAC chip through an engineered histidine tag and to then identify proteins that bind specifically to such an activated surface by SELDI-MS. In this instance, His-tagged RsrA and a limited trypsic digest of \( \sigma^R \) were loaded onto an IMAC chip and the chip washed repeatedly with binding buffer prior to the identification of bound fragments by SELDI-MS. Both N-terminal \( \sigma^{RN} \) fragments were retained specifically on the His-RsrA-activated chip, with no detectable retention of \( \sigma^{RC} \) (Figure 1(b), (ii)). Since both \( \sigma^{RN} \) fragments were identified, it is reasonable to assume that the deletion of the C-terminal six residues does not affect the ability of this domain to bind RsrA. Control experiments in which either His-RsrA or \( \sigma^R \)-specific trypsin fragments alone were bound to IMAC chips, washed repeatedly and then analysed by SELDI-MS, indicated that retention of the \( \sigma^{RN} \) fragments required bound RsrA (Figure 1(b), (iii) and (iv)). We conclude that \( \sigma^{RN} \) is a stable N-terminal 10 kDa domain, encompassing regions 2.1–2.4 of \( \sigma^R \), which also represents a major interaction site for RsrA.

**Characterisation of the \( \sigma^{RN} \) domain and its complex with RsrA**

In order to obtain quantities of \( \sigma^{RN} \) sufficient for biophysical analysis, the corresponding fragment of sigR encoding its 88 amino acid residues (residues 23–110) was subcloned into pET15b, thereby placing an N-terminal hexa-histidine tag onto the domain (His-\( \sigma^{RN} \)) (see Materials and Methods). Circular dichroism (CD) spectroscopy of purified His-\( \sigma^{RN} \) at pH 5.4 indicated that the domain is folded and composed of ~68% \( \alpha \)-helix, with little or no \( \beta \)-sheet structure (Figure 2(a)). We
found that preparations of His-σRN (with or without the histidine tag) were more stable at low pH (6) but prone to aggregation at neutral pH unless imidazole or RsrA was included in the buffer (data not shown). We return to this issue below.

Overexpressed His-σRN was capable of binding non-His-tagged RsrA in vitro. Figure 3 shows the results of a pull-down assay (at pH 7.5 in the presence of 100 mM imidazole) in which His-σRN and RsrA were added to Ni²⁺-chelate beads, the beads washed extensively with binding buffer, and then finally eluted with EDTA. RsrA and His-σRN only co-eluted during the final EDTA strip (Figure 3(a), lane 5). Omission of His-σRN from the Ni²⁺-charged resin (Figure 3(a), lane 4) demonstrated that retention of RsrA on the beads required bound His-σRN. Complex formation between overexpressed His-σRN and reduced RsrA was also stable enough to be detected by gel-filtration chromatography (data not shown).

Interaction of σRN with core RNA polymerase is inhibited by RsrA

σRN is an autonomous, stably folded domain that contains sequences (region 2.4) required for binding to the −10 box of a promoter but lacks key sequences in the C-terminal half of σH that are needed for −35 recognition. Hence, it was to be expected that σRN could not initiate transcription from sigR-dependent promoters, and preliminary experiments indicated that this was indeed the case (data not shown). Binding to core RNA polymerase by bacterial sigma factors involves conserved regions of the protein that have been delineated for many different sigma factors. The presence of at least one of these (region 2.1) suggested that σRN might be able to associate with the enzyme even though it could not initiate transcription.

To test this hypothesis we investigated whether His-σRN could compete with the unrelated Streptomyces sigma factor, σH, for the σ binding site on core RNA polymerase. σH plays an important role in the general stress response in Streptomyces.28

Figure 4. RsrA competes with core RNA polymerase for σRN binding. In vitro transcription of the ctc promoter by σH. Lane 1, no σH control; lanes 2–4 contain 1.9 pmol σH plus 0, 9.5 and 19-fold of His-σRN to σH; lanes 5–7, contain the same amount of σH and His-σRN as in lane 4 but include two-, four- and eightfold excess (relative to His-σRN) of reduced RsrA. The amount of transcript in lanes 2–7 was quantified by densitometry and plotted as the percentage of σH-dependent transcription observed in lane 2. The decrease in σH-dependent transcription when His-σRN was added (lanes 3 and 4) indicates that this small domain competes effectively for core RNA polymerase binding (although itself cannot activate transcription) but that this inhibition is relieved by RsrA.
Transcription from a $\sigma^H$-dependent promoter (Figure 4, lane 2) was largely inhibited when a molar excess of His-$\sigma^RN$ was included in the assay (Figure 4, lane 4), consistent with $\sigma^RN$ competing with $\sigma^H$ for the sigma factor binding site on core RNA polymerase. Importantly, the majority of $\sigma^H$-dependent transcription could be recovered by titrating-in an excess of reduced RsrA relative to His-$\sigma^RN$ (Figure 4, lanes 5–7). Taken together, these results suggest that $\sigma^RN$ retains the capacity to bind core RNA polymerase but that this association is likely inhibited by RsrA.

Crystallization and structure of $\sigma^{RN}$

Crystallographic information is currently available for $\sigma^{70}$ from *E. coli* and $\sigma^A$ from *Thermus aquaticus* both bound to core RNA polymerase and in the unbound form, and for $\sigma^F$ from *B. stearothermophilus* in complex with its anti-sigma factor SpoIIAB. No structural information is yet available for any ECF sigma factor. We attempted to grow crystals of intact His-$\sigma^R$ that had been purified through Ni-affinity chromatography and the N-terminal histidine tag removed by thrombin digestion (see Materials and Methods). Hexagonal-shaped crystals were obtained after only a few days using polyethylene glycol (PEG) 8000 as precipitant at pH 8 and at 4°C. However, subsequent analysis of the crystals by SDS-PAGE indicated that they did not consist of intact $\sigma^R$ but rather of a proteolytic digestion product identical with the 10 kDa N-terminal fragment generated by trypsin, $\sigma^{RN}$. It seems that thrombin digestion of the intact $\sigma^R$ continued during crystallisation experiments and that crystals were only formed after the formation of $\sigma^{RN}$. Similar proteolytic susceptibility during crystallisation has been noted for $\sigma^A$ from *T. aquaticus*.6

We solved the structure of $\sigma^{RN}$ by multiple anomalous displacement (MAD) phasing to a resolution of 2.7 A˚ using intact selenomethionine-labelled $\sigma^R$ as the starting material, which also became proteolysed during crystallisation. The structure was subsequently refined against native data collected to 2.4 A˚ resolution. Crystals were also obtained for overexpressed $\sigma^{RN}$ but these did not improve the data quality (data not shown). Of the 88 amino acid residues that comprise $\sigma^{RN}$ only 78 could be unambiguously placed in the electron density; the residues that are missing from the final model originate primarily from the C terminus, suggestive either of static or dynamic disorder in the crystal. $\sigma^{RN}$ is an antiparallel three-helix bundle with three residues of 3 10 helix in the loop joining helices 2 and 3 (Figure 5). $\alpha$-Helix accounts for 68% of the observed $\sigma^{RN}$ domain structure, identical with the predicted helical content from CD. Helix 1 is the longest and displays a pronounced kink approximately halfway along its length.

Structural homologues of $\sigma^{RN}$ in the Protein Data Bank (PDB) were identified using the DALI server. This yielded several significant matches with fragments of other structures (Figure 6), the best match being that with the structure of $\sigma^{70}$ (PDB accession code 1SIG). A least-squares superposition of the $\sigma^{70}$ structure onto $\sigma^{RN}$, on the basis of the main-chain atoms from 75 equivalent residues, gave an overall root-mean-square deviation (rmsd) of 2.9 A˚. It is notable that the 3 10 helix and the kinked helix, which encompasses region 2.1, are also preserved in the $\sigma^{70}$ structure. Other
significant matches were found with cyclin A (PDB accession code 1VIN), rmsd of 2.4 Å over 68 residues, and transcription factor TFII B (PDB accession code 1AIS), rmsd of 1.7 Å over 59 residues, although in both of these the kink in the first helix is not present (Figure 6).

Protein–protein interactions in the crystal and implications for the mode of action of RsrA

In the σRN crystal, the molecules pack as a network of non-intersecting fibres that run along the 2-fold screw axes of the unit cell (Figure 7). The asymmetric unit contains two copies of σRN that are related to one another by a 90° rotation about the 2-fold screw axis and a quarter cell edge translation along the axis. Further 90° rotations and quarter cell edge translations in both directions relate adjacent molecules in neighbouring asymmetric units. Consequently, there is a 4-fold non-crystallographic screw axis coincident with the crystallographic 2-fold screw axis. This is apparent in a self-rotation function (calculated using data in the range 10.0–3.5 Å resolution) as a peak at 15% of the origin height that coincides with the crystallographic 2-fold screw axis. Thus the crystallographic and non-crystallographic interfaces along the fibre are virtually indistinguishable.

The total buried surface area at the interfaces between adjacent molecules amounts to 810 Å². The surfaces are remarkably complementary in terms of shape (a shape complementarity index of 0.64 makes them comparable to antibody/antigen complexes) and charge (Figure 7). These characteristics likely explain the propensity of σRN to precipitate at neutral pH and the higher order oligomers that are readily detected in chemical cross-linking experiments (data not shown).

The protein–protein interface observed in the σRN fibre is unlikely to be biologically significant given that the domain represents less than 40% of intact σ and that σ itself is not prone to precipitation at physiological pH. However, the self-association provides clues as to how RsrA might bind and sequester σ in an inactive, heterodimeric complex. This stems from the fact that we found RsrA to be very effective at preventing the aggregation of σRN. In our handling of purified σRN, we found that the domain (with or without

**Figure 7.** Protein–protein interactions in the σRN crystal. (a) Packing of molecules in the fibre. Crystallographically equivalent molecules are shown in the same colour and are related by the 2-fold screw axis. Adjacent molecules are related by a non-crystallographic 4-fold screw axis. (a)–(g) The central pair of molecules from the fibre in (a) rotated by 90° about the vertical axis in opposite directions so that their interacting surfaces face the viewer. (b), (d) and (f) show the left-hand molecule, (c), (e) and (g) show the right-hand molecule. (b) and (c) are in ribbon representation and use the same colours as Figure 6. (d) and (e) show the electrostatic surface potentials as red, white and blue for values less than −10 kT, neutral and greater than 10 kT, respectively. (f) and (g) show the interaction surfaces. Those regions of the surface that are closer than 2 Å away from the neighbouring surface are coloured yellow, and the remainder is coloured grey. (a)–(c) were generated as for Figure 6. (d)–(g) were produced using GRASP.

**Figure 8.** RsrA prevents the self-association of σRN. A mixture of His-σRN and σRN in 20 mM Tris–HCl (pH 7.5), 0.1 M NaCl, 400 mM imidazole was dialysed overnight at 4 °C against the same buffer, but lacking imidazole, in the presence (lanes 3 and 4) or absence (lanes 1 and 2) of a twofold molar excess of RsrA. The addition of BSA to dialysed His-σRN/σRN served as a negative control (lanes 5 and 6). Soluble (S) and pellet (P) fractions from each of the three samples were then analysed by SDS-20% PAGE. The experiment demonstrates that His-σRN/σRN readily precipitates at neutral pH, even in the presence of BSA, but that this is prevented by the inclusion of RsrA. The σRN preparation used in the experiment was a mixture of histidine and non-histidine-tagged protein (resulting from prolonged storage of His-σRN), demonstrating that both are prone to aggregation. The asterix indicates an impurity in the preparation.
the histidine tag) was only soluble at neutral pH in the presence of high concentrations (>100 mM) of imidazole. Removal of the imidazole by dialysis against Tris–HCl at pH 7.5 results in the precipitation of both $\sigma^R$ and His-$\sigma^R$ (Figure 8, lane 2). Aggregation was largely abolished by the presence of RsrA during dialysis but not by the inclusion of BSA (Figure 8, lanes 3–6). Since RsrA binds specifically to $\sigma^R$ (see Figures 1 and 3), this effect is most readily interpreted in terms of RsrA inhibiting the self-association of $\sigma^R$. Since RsrA is an acidic protein (pI, 4.8) we speculate that it prevents $\sigma^R$ self-association by binding to its electropositive surface (Figure 7(e)).

**Discussion**

We have defined the anti-sigma factor binding domain on the ECF sigma factor $\sigma^R$ from *S. coelicolor*. $\sigma^R$ regulates the transcription of many genes involved in the maintenance of thiol-disulphide redox homeostasis, while its anti-sigma factor RsrA is the redox sensor. Trypsin treatment of $\sigma^R$ under native conditions generated two major fragments, from the N and C terminus, respectively, with only $\sigma^R$ able to bind RsrA (Figure 1). While this demonstrates that $\sigma^R$ is most likely the major binding site for RsrA we cannot discount the possibility that additional contacts are made to other regions but are lost during proteolysis. Over-expressed $\sigma^R$ inhibits transcription by an unrelated sigma factor (Figure 4), suggesting that $\sigma^R$ is an autonomous domain able to bind at the sigma factor binding site on core RNA polymerase. Reduced RsrA abolished this effect, implying that RsrA may inhibit $\sigma^R$-dependent transcriptional activity by inhibiting its binding to polymerase. The self-association of $\sigma^R$, evident in solution as a propensity to aggregate and observed crystallographically as $\sigma^R$ fibres (Figure 7), is due to electrostatically complementary surfaces in the domain coming together, a phenomenon that is largely prevented by the binding of RsrA (Figure 8).

The crystal structure of $\sigma^R$ revealed a similar amount of $\alpha$-helix to that suggested by circular dichroism, and that this domain is structurally similar to the equivalent region in $\sigma^F$ from *E. coli* (Figure 6), consistent with the presence of well-defined and conserved regions involved in binding to core RNA polymerase (region 2.1), DNA melting (region 2.3) and −10 recognition (region 2.4). $\sigma^R$ also shows strong structural similarity to a fold that is shared by the eukaryotic signalling protein cyclin A and transcription factor TFIIIB, termed the cyclin box fold. The cyclin box fold is characterised by five helices, three of which are found in $\sigma^R$. The cyclin box fold seems to be a general adapter motif employed in transcriptional regulation, as well as cell cycle control, able to bind a diverse set of proteins and DNA. Although a prokaryotic protein, the presence of the cyclin box fold within $\sigma^R$ is consistent with these properties, since it binds other proteins (core RNA polymerase and RsrA) and DNA.

Genome sequencing has revealed that the ECF subfamily of sigma factors is far larger than first imagined; *S. coelicolor* alone has 49 members. Overlapping promoter specificity, where more than one ECF sigma factor recognizes a single promoter, is an emerging theme. For example, almost 50% of promoters recognized by EorK are also recognized by at least one other form of holoenzyme, and $\sigma^R$ and $\sigma^W$ of *B. subtilis* have overlapping specificity at several promoters. In both cases, the −10 region appears to play an important role in promoter selectivity. The $\sigma^R$ structure, and the identification of residues along the likely DNA-binding face of the holoenzyme, will aid the design of experiments directed towards understanding the mechanism of promoter selection by $\sigma^R$ and other ECF sigma factors.

**Comparison with other $\sigma$ factor:anti-$\sigma$ factor complexes**

One of the best understood sigma factor:anti-sigma factor interactions is that of $\sigma^F$ and its complex with SpoIIAB. $\sigma^F$ initiates forespore-specific gene expression during sporulation in *B. subtilis* and forms a 1:2 complex with its anti-sigma factor SpoIIAB. In the crystal structure of the low-affinity (SpoIIAB(ADP)) form of the complex, region 3 of $\sigma^F$ makes specific contacts with each of the SpoIIAB monomers, while the remaining 80% of $\sigma^F$ is disordered. Genetic and biochemical data indicate that SpoIIAB also contacts regions 2 and 4 of $\sigma^F$ in the high-affinity (SpoIIAB(ATP)) form of the complex. Direct occlusion of a core RNA polymerase binding surface appears to be the basis of SpoIIAB’s anti-$\sigma$ activity.

In *S. typhimurium*, $\sigma^{25}$, which transcribes flagellar and chemotaxis genes, is regulated by a cognate anti-sigma factor, FlgM. In contrast to the mode of inhibition of $\sigma^F$ by SpoIIAB, FlgM actively dissociates the $\sigma^{28}$ holoenzyme to yield a 1:1 $\sigma^{28}$:FlgM complex. Genetic and biochemical data suggest that FlgM contacts multiple regions of free $\sigma^{28}$, including 2.1, 3.1 and 4, whereas anti-holoenzyme activity appears to be mediated solely through its interaction with region 4.

During infection of *E. coli* by T4 bacteriophage, the phage-encoded protein AsiA inhibits host transcription by binding to $\sigma^F$ to form a 1:1 complex, interacting specifically with region 4. AsiA is a symmetric dimer with the residues that form the hydrophobic dimer interface also involved in binding $\sigma^F$, suggesting that an AsiA protomer is displaced from the dimer through its interaction with $\sigma^F$. Unlike the mode of action of SpoIIAB and FlgM, AsiA can form a stable complex with the $\sigma^F$ holoenzyme. Available evidence suggests that AsiA modifies $\sigma^F$ function, masking the −35 helix-turn-helix DNA binding motif in region 4 while bound to the holoenzyme.
The present work, along with that reported previously, suggests that the anti-ECF action of the ZAS protein RsrA is distinct from that of other anti-sigma factors. Unlike the 1:2 complex of the α5-SpoIIB complex, α5 forms a 1:1 complex with RsrA and so, since RsrA is a monomer (data not shown), dimer dissociation as seen with AsiA is unlikely. Region 2 plays a prominent role in the interaction of both α5 and α28 with their anti-sigma factors, which is also the case for α5. Region 4, which plays a central role in the regulation of α5, α28 and α70 by anti-sigma factors, may not be involved in binding to RsrA, since the C-terminal domain, αRC, which includes region 4, did not interact with RsrA (Figure 1).

αRC can inhibit sigma-dependent transcription and RsrA can inhibit this effect, suggesting that RsrA prevents RNA polymerase binding by α5 (Figure 4), the binding site on the sigma factor likely comprising a positively charged surface (Figure 7e). This site is composed of α3 and the C-terminal end of α1 of αRC, the former incorporating residues involved in DNA melting and −10 recognition (regions 2.3 and 2.4), the latter, part of the RNA polymerase binding site (region 2.1). Hence, RsrA likely obscuresα5 sequences involved in binding both promoter DNA and RNA polymerase. This form of sigma factor inhibition would represent a novel mechanistic variation in sigma factor:anti-sigma factor interactions and may underpin the regulation of other ZAS protein—ECF sigma factor complexes.

Materials and Methods

Protein production

RsrA was purified as described by Kang et al. T7-based overexpression of α5 was accomplished essentially as described by Paget et al., but where α5 carried an N-terminal hexa-histidine tag and an intervening thrombin site, transformed into E. coli BL21DE3/plysS for unlabelled cells or BL21DE3 met−/plysS for selenomethionyl-labelled cells. For the expression of unlabelled protein, cells were grown at 37°C in LB medium containing 200 μg/ml ampicillin and 30 μg/ml chloramphenicol to mid-log and then induced for three hours with IPTG (1 mM). For selenomethionyl-labelled protein, an overnight LB culture was harvested, washed with 10 ml M9 medium and then resuspended in 5 ml M9 medium, 3 ml of which was used to inoculate a 11 culture of M9 medium containing 200 μg/ml ampicillin and 30 μg/ml chloramphenicol and supplemented with 2 mM MgSO4, 10 mg thiamine, 0.1 ml M CaCl2, 10 g glucose, 40 mg of 19 amino acid residues (except methionine) and 60 mg selenomethionine (SeMet). This culture was grown at 37°C and induced at mid-log with 1 mM IPTG for 20 hours. Harvested cells were disrupted using either a Soniprobe sonicator (Dawes) at 4°C or by French pressure, and clarified supernatant passed down a nickel-affinity column (POROS 20 MC, PerSeptive Biosystems) in binding buffer (20 mM Tris–HCl (pH 7.5), 0.5 M NaCl, 5 mM imidazole) at room temperature. Bound α5 was eluted with 50–600 mM imidazole gradient on a BioCad SPRINT (PerSeptive Biosystems) using a flow rate of 10 ml/minute. Following thrombin digestion to remove the hexa-histidine tag (0.5 units thrombin/mg α5 for two hours at 27°C), the protein was further purified by gel-filtered chromatography using a Superdex 200 26/60 column (Pharmacia Biotech) equilibrated in 50 mM Tris–HCl (pH 7.5), 1 mM EDTA. Pure α5 fractions were pooled, concentrated and stored in 50% (v/v) glycerol at −20°C.

The N-terminal domain of α5, corresponding to residues 23–110 (α28), was generated with a hexa-histidine tag (His-α28) by cloning the corresponding DNA into pET15b (pET15b-sigRN), and purified as described above for α5 often without the inclusion of the thrombin cleavage or gel-filtration steps, the protein being essentially pure after Ni2+-affinity chromatography. His-α5 was prone to aggregation at pH 7.5 unless imidazole was included in the buffer and so was generally stored at −20°C as pooled fractions from the Ni2+-column (final imidazole ~300 mM) and included 50% glycerol. The protein was soluble at pH 7.5 but only in the presence of RsrA (see the text for details), alternatively the protein remained soluble in low pH buffers such as sodium acetate or sodium phosphate (pH 5.4).

All protein molecular masses were verified either by electrospray ionisation mass spectrometry (VG Platform; Micromass, UK) or by SELDI-MS (Ciphergen Biosystems, USA) and were within 1–2 Da of the expected mass. Protein concentrations were determined from molar absorption coefficients derived from amino acid analysis (α5, ε280 = 27,682 M−1 cm−1; α28, ε280 = 16,298 M−1 cm−1).

Tryptsin digestion and mass spectrometry

α5 was digested with trypsin (0.4%, Sigma) for 30 minutes at 25°C in 50 mM Tris–HCl (pH 7.5) and the digestion mixture subjected to SDS-16% (w/v) PAGE. Major proteolytic fragments were blotted onto PVDF membrane and subjected to N-terminal sequencing (Alta Bioscience, Birmingham, UK). SEDL-MS was generally used to determine fragment masses using either normal phase silicon oxide surface (NP1 and NP20) chips, for simple verification of peptide masses, or immobilised metal affinity capture (IMAC3) chips for the identification of peptide fragments binding to RsrA. In the latter experiments, 10 mM DTT reduced His-RsrA and a α5 trypsin digestion mixture (each ~20 μM) were incubated at 30°C for one hour before spotting onto an IMAC3 chip pre-loaded with NiSO4 and washed repeatedly with binding buffer (× 5) and distilled water (× 3) before detection of captured peptides by SELDI-MS. In all experiments, 3,5-dimethoxy-4-hydroxycinnamic acid or α-cyano-4-hydroxycinnamic acid were used as the matrix and the endonuclease domain of colicin E9 (15,088 Da) and the immunity protein Im9 (9882.6 Da) used as calibrants of molecular mass.

Circular dichroism spectroscopy

CD data were collected on an Applied Photophysics π-180 CD spectrophotometer (Leatherhead, UK) in a 1 mm pathlength cell at 20°C using 3.8 μM His-α5 in 50 mM potassium phosphate buffer (pH 5.4). A total of 50,000 data points were collected between 190 nm and 260 nm and 20 scans averaged. Secondary structure content was estimated using the deconvolution program CDNN.
Unique reflections, BSA was added to His-dithiothreitol-reduced RsrA. In a third dialysis experiment, the presence or absence of a twofold molar excess of 4
containing 400 mM imidazole, was dialysed overnight at 4°C. The precipitant consisted of 10–15% (w/v) PEG 8000, in 100 mM Tris–HCl at pH 8.5. These conditions were supplemented with 1 mM DTT for the SeMet protein. Drops consisted of 1 μl of protein solution mixed with 1 μl of precipitant. The crystals appeared as hexagonal plates after two days, being approximately 100 μm across and 20 μm thick. In order to ascertain the precise protein sequence of the crystallized material, several crystals were dissolved and N-terminally sequenced using a Proxcel Model 491 protein sequencer (Applied Biosystems, Warrington, Cheshire), while SELDI-MS (see above) was used to ascertain the molecular weight and hence define the C terminus. The crystallized protein corresponded to the 10.3 kDa fragment of σRN, consisting of residues 23–110.

### Data collection
All crystal manipulations were performed using Hampton Research tools. Initially, crystals were cryoprotected by soaking for up to five minutes in mother liquor containing ethylene glycol in place of 25% of the buffer volume, after which they were mounted in cryoloops and flash-cooled by plunging into liquid nitrogen. They were then transferred to a dewar using the Crystal Cap system for storage and subsequent transportation to the synchrotron. For data collection, crystals were mounted on the goniometer using cryotongs and maintained at 100 K with an Oxford Cryosystems Cryostream cooler. The observed diffraction was consistent with trigonal symmetry giving approximate cell parameters of a = b = 71.6 Å, c = 102.8 Å. Systematic absences along c indicated that the space group was either P3_1_21 or P3_2_1. An estimation of solvent content gave values of 66% for two copies of σRN per asymmetric unit, and 50% for three copies.44 X-ray data for the native protein were collected on the microfocus beamline ID13 (λ = 0.782 Å) at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France, using a 165 mm MAR-Research CCD detector on beamline PX9.5 at the Synchrotron Radiation Source (SRS) at Daresbury, UK. Prior to data collection, a fluorescence scan was used to determine the best wavelengths at which to collect data. Subsequently, X-ray data were collected at wavelengths of 0.9794 Å (peak), 0.9801 Å (inflection point) and 0.8500 Å (high-energy remote). All data sets were collected in a single 180° sweep of 1 deg. images to 2.7 Å resolution.

### Data processing and structure solution
The native X-ray data were processed with DENZO and scaled and merged using SCALEPACK40 and all

<table>
<thead>
<tr>
<th>Beamline</th>
<th>Data set</th>
<th>Resolution (Å)</th>
<th>Completeness (%)</th>
<th>Redundancy</th>
<th>Rmerge</th>
</tr>
</thead>
<tbody>
<tr>
<td>PX9.5 (SRS)</td>
<td>SeMet-peak</td>
<td>0.9794</td>
<td>100.0 (100.0)</td>
<td>5.9</td>
<td>0.045 (0.282)</td>
</tr>
<tr>
<td>PX9.5 (SRS)</td>
<td>SeMet-inflection</td>
<td>0.9801</td>
<td>100.0 (100.0)</td>
<td>5.9</td>
<td>0.053 (0.324)</td>
</tr>
<tr>
<td>PX9.5 (SRS)</td>
<td>SeMet-remote</td>
<td>0.8500</td>
<td>100.0 (100.0)</td>
<td>5.9</td>
<td>0.055 (0.343)</td>
</tr>
<tr>
<td>ID13 (ESRF)</td>
<td>Native</td>
<td>0.782</td>
<td>97.9 (83.4)</td>
<td>3.8</td>
<td>0.047 (0.142)</td>
</tr>
</tbody>
</table>

$R_{merge} = \frac{\sum_{j}(I_j - \langle I \rangle)}{\sum_{j}(I_j)}$, where I is the intensity of an observation of reflection j and $\langle I \rangle$ is the average intensity for reflection j.

In vitro binding assays
Pull-down assays were performed using Ni-NTA resin (Novagen) and in 100 mM Tris–HCl (pH 7.5), 150 mM NaCl containing 100 mM imidazole. His-σRN, which remains soluble under these conditions, and reduced RsrA (non-His-tagged) were each incubated with Ni-NTA resin either individually or in complex. After washing the beads with binding buffer (see above) containing 1 mM of the reducing agent Tris (2-carboxyethyl) phosphine hydrochloride, bound proteins were eluted with strip buffer (100 mM Tris–HCl (pH 7.5), 150 mM NaCl, 100 mM EDTA) and fractions analysed by SDS–16% PAGE.

In vitro transcription run-off assays
In vitro transcription assays were performed as described31 using transcription buffer containing 1.9 pmol σRN and 0.5 pmol of the σRN-specific promoter etc.31 The effects of His-σRN and RsrA on this activity were monitored by pre-incubating these proteins, in the presence of 10 mM DTT, at 30°C for 30 minutes, and then with 1 μl of E. coli core RNA polymerase (1.5 pmol, Cambio, Cambridge, UK) for a further five minutes before adding transcription buffer and initiating transcription by the addition of an NTP mix (0.4 mM) containing [γ-32P]CTP (400 Ci/mmol). Transcripts were analysed on a 6% polyacrylamide gel containing 7 M urea.

σRN aggregation assay
A mixture of His-σRN and σRN (~1 mg/ml total), dissolved in 20 mM Tris–HCl (pH 7.5), 0.1 M NaCl containing 400 mM imidazole, was dialysed overnight at 4°C against 20 mM Tris–HCl (pH 7.5), 0.1 M NaCl in the presence or absence of a twofold molar excess of dithiothreitol-reduced RsrA. In a third dialysis experiment, BSA was added to His-σRN/σRN. Soluble and pellet fractions from each of the three samples were then analysed by gel-electrophoresis.

Crystallization
Thrombin-cleaved σR was concentrated to approximately 10 mg/ml, filtered through a 0.1 μl Ultrafree filter and crystals grown by vapour diffusion in hanging drops using VDX plates (Hampton Research) at 4°C. The precipitant consisted of 10–15% (w/v) PEG 8000, in 100 mM Tris–HCl at pH 8.5. These conditions were supplemented with 1 mM DTT for the SeMet protein.

### Table 1. X-ray data collection statistics

<table>
<thead>
<tr>
<th>Beamline</th>
<th>Data set</th>
<th>Resolution (Å)</th>
<th>Completeness (%)</th>
<th>Redundancy</th>
<th>Rmerge</th>
</tr>
</thead>
<tbody>
<tr>
<td>PX9.5 (SRS)</td>
<td>SeMet-peak</td>
<td>0.9794</td>
<td>100.0 (100.0)</td>
<td>5.9</td>
<td>0.045 (0.282)</td>
</tr>
<tr>
<td>PX9.5 (SRS)</td>
<td>SeMet-inflection</td>
<td>0.9801</td>
<td>100.0 (100.0)</td>
<td>5.9</td>
<td>0.053 (0.324)</td>
</tr>
<tr>
<td>PX9.5 (SRS)</td>
<td>SeMet-remote</td>
<td>0.8500</td>
<td>100.0 (100.0)</td>
<td>5.9</td>
<td>0.055 (0.343)</td>
</tr>
<tr>
<td>ID13 (ESRF)</td>
<td>Native</td>
<td>0.782</td>
<td>97.9 (83.4)</td>
<td>3.8</td>
<td>0.047 (0.142)</td>
</tr>
</tbody>
</table>

$R_{merge} = \frac{\sum_{j}(I_j - \langle I \rangle)}{\sum_{j}(I_j)}$, where I is the intensity of an observation of reflection j and $\langle I \rangle$ is the average intensity for reflection j.

Anomalous pairs treated as separate reflections for MAD data.

Figures in parentheses refer to data in the highest resolution bin.
Table 2. Model parameters and quality

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution range (Å)</td>
<td>40–2.4</td>
</tr>
<tr>
<td>Rcryst (on the basis of 95% of data; %)</td>
<td>25.8</td>
</tr>
<tr>
<td>Rmerge (on the basis of 5% of data; %)</td>
<td>27.8</td>
</tr>
<tr>
<td>Residues with most favoured 4Φ (%)</td>
<td>87.8</td>
</tr>
</tbody>
</table>

**Bond distances (Å)**
- 0.029

**Root-mean-square deviations from ideality**
- Bond angles (°) 2.44
- Chiral volume (Å³) 0.204
- Planar torsion angles (°) 4.2
- Staggered torsion angles (°) 20.8

**Average B factors (Å²)**
- Main-chain atoms 34
- Side-chain atoms 36
- Water molecules 52
- Overall 34

*The R-factors Rcryst and Rmerge are calculated as follows: R = \[
\sum (|F_{obs} - F_{calc}|) / \sum |F_{obs}| \times 100\%
\]
where F_{obs} and F_{calc} are the observed and calculated structure factor amplitudes, respectively.

*As calculated using PROCHECK.*

Model building and refinement

Model building was performed using the program O and alternated with refinement using REFMAC. The four Se positions were arranged in two pairs and, upon closer inspection, each pair was seen to be associated with an α-helix, the two atoms lying approximately two turns of helix apart. The latter observation was consistent with the likely separation of Met47 and Met54 in the structure. These positions were subsequently used to define the 2-fold non-crystallographic symmetry axis relating one protein molecule to the other. This enabled the initial electron density maps to be improved by 2-fold averaging. After model building and refinement, the final model consisted of a total of 154 residues in two chains. Molecule A comprised residues 28–102 of the full length α70, whilst molecule B contained residues 25–102. In other words, some eight residues at the C terminus of α70N were not modelled for both chains, whereas five and two residues were absent at the N termini of the A and B chains, respectively. After refinement, the final Rcryst and Rmerge values were 25.8% and 27.8%, respectively. The model quality was evaluated using PROCHECK, approximately 88% of the residues lay in the most favoured regions of the Ramachandran plot, otherwise the overall criteria required for a structure at a resolution of 2.4 Å were either satisfied or exceeded. The parameters of the final structure are summarised in Table 1.

Accession numbers

The coordinates and structure factor data for α70N have been deposited in the Protein Data Bank with accession codes 1H3L and 1H3LSF, respectively.

Acknowledgements

This work was funded by The Wellcome Trust and the BBSRC. N.B. was supported by the BBSRC-funded Cambridge and East Anglia Centre for Structural Biology. We are grateful for support and access to the SRS and particularly thank beamline scientists J. Nicholson (SRS) and A. Perrakis (ESRF). L. Mitchenall, S. Mayer and T. Clarke are acknowledged for assistance during X-ray data collection. We thank the EMBL Grenoble Outstation for supporting data collection at the ESRF under the European Union TMR/LSF Programme. Finally we acknowledge M. Naldrett for N-terminal sequencing and A. Leech for assistance with SELDI-MS data acquisition.

References


Edited by J. Karn

(Received 19 June 2002; received in revised form 30 August 2002; accepted 5 September 2002)