

# The interaction of *Bacillus subtilis* $\sigma^A$ with RNA polymerase

Elecia B. Johnston, Peter J. Lewis, and Renate Griffith\*

Discipline of Biological Sciences, School of Environmental and Life Sciences, The University of Newcastle, Callaghan, NSW, Australia

Received 6 July 2009; Revised 20 August 2009; Accepted 21 August 2009

DOI: 10.1002/pro.239

Published online 4 September 2009 proteinscience.org

**Abstract:** RNA polymerase (RNAP) is an essential and highly conserved enzyme in all organisms. The process of transcription initiation is fundamentally different between prokaryotes and eukaryotes. In prokaryotes, initiation is regulated by  $\sigma$  factors, making the essential interaction between  $\sigma$  factors and RNAP an attractive target for antimicrobial agents. Our objective was to achieve the first step in the process of developing novel antimicrobial agents, namely to prove experimentally that the interaction between a bacterial RNAP and an essential  $\sigma$  factor can be disrupted by introducing carefully designed mutations into  $\sigma^A$  of *Bacillus subtilis*. This disruption was demonstrated qualitatively by Far-Western blotting. Design of mutant  $\sigma$ s was achieved by computer-aided visualization of the RNAP- $\sigma$  interface of the *B. subtilis* holoenzyme (RNAP +  $\sigma$ ) constructed using a homology modeling approach with published crystal structures of bacterial RNAPs. Models of the holoenzyme and the core RNAP were rigorously built, evaluated, and validated. To allow a high-quality RNAP- $\sigma$  interface model to be constructed for the design of mutations, a crucial error in the *B. subtilis*  $\sigma^A$  sequence in published databases at amino acid 165 had to be corrected first. The new model was validated through determination of RNAP- $\sigma$  interactions using targeted mutations.

**Keywords:** RNA polymerase; sigma factor; homology model; antimicrobials; protein-protein interactions

## Introduction

RNA polymerase (RNAP) is the enzyme responsible for transcription in all organisms. Although RNAP is structurally conserved among all domains of life,<sup>1</sup> there is considerable difference in how the enzyme functions in eukaryotes and prokaryotes. The active site is highly conserved but the factors that regulate transcription are very different.

Bacterial RNAP core enzyme consists of five subunits: two  $\alpha$ , and single  $\beta$ ,  $\beta'$ , and  $\omega$  subunits. A sixth

subunit,  $\sigma$ , binds to the core to form what is known as the holoenzyme (HE) enabling the enzyme to bind to specific promoter sequences on the DNA and to initiate transcription.<sup>2</sup> This is the first step of the transcription cycle which consists of three stages: initiation, elongation, and termination. Upon initiation of transcription, RNAP frequently produces short RNA transcripts, two to nine nucleotides in length, without clearing the promoter region (abortive initiation). Once the RNA chain reaches  $\sim 12$  nucleotides, the transcription complex becomes more stable allowing  $\sigma$  to stochastically dissociate and RNA chain elongation to continue.<sup>3-5</sup>  $\sigma$  Factors are a large family of proteins which are essential for initiation and specificity of transcription. In addition to a "housekeeping"  $\sigma$  factor, several additional  $\sigma$  factors may be present in a cell at any one time. The additional  $\sigma$  factors generally become activated due to specific environmental signals, leading to the transcription of specific genes, such as those involved in the stress response/

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

**Abbreviations:** core, core RNA polymerase; DDM, distance difference matrix; HE, RNA polymerase holoenzyme; HtH, helix-turn-helix; RNAP, RNA polymerase.

Grant sponsor: ARC; Grant number: DP0664370; Grant sponsor: NHMRC; Grant number: 455597.

\*Correspondence to: Renate Griffith, School of Medical Sciences/Pharmacology, University of New South Wales, Sydney, NSW 2052, Australia. E-mail: r.griffith@unsw.edu.au

pathogenicity. In *Bacillus subtilis*, the primary “house-keeping”  $\sigma$  factor is  $\sigma^A$ .<sup>6</sup>

High resolution structures of different forms of RNAP have been solved, including core,<sup>7,8</sup> HE,<sup>3,9</sup> and an elongation complex.<sup>10</sup> Although these high resolution structures are all from extremophiles, due to high levels of sequence conservation, the information obtained is transferable to other bacterial RNAPs. The structure and function of bacterial RNAP HEs have been reviewed recently.<sup>11</sup> The highly conserved bacterial  $\sigma$  factors consist of four major conserved regions, designated 1–4. There are many contacts between  $\sigma$  and the core enzyme, but the most extensive interaction occurs between the polar surface of  $\sigma$  region 2.2 and a solvent exposed coiled-coil region of the RNAP  $\beta'$  subunit.<sup>3,9</sup>

Because RNAP is an essential and highly conserved enzyme, it has become an attractive target for the development of new antimicrobial drugs. The mechanism of initiation, as regulated by  $\sigma$  factors, is unique to prokaryotes, so their interaction with RNAP represents an excellent target. This target is exploited by natural anti-sigma factors in Gram positive and negative organisms, and also in drug discovery programs.<sup>12,13</sup> *B. subtilis* has been extensively studied as a paradigm for  $\sigma$ -factor regulated gene expression (particularly during sporulation) (e.g., as shown in Ref. 14). It is a harmless soil bacterium but is closely related to many important pathogens in the low G+C group of Gram positive bacteria (firmicutes). Creation of high-quality models of the RNAP HE enables their use as a tool to design transcription initiation inhibitors. Key residues involved in the interaction between  $\sigma$  factors and the RNAP core can then be specifically targeted which may lead to improved models and aid rational design of inhibitors of this essential interaction as potential antibiotics with a novel mechanism of action.

*B. subtilis* RNAP core and HE have been modeled previously.<sup>15</sup> The interface between  $\beta'$  and  $\sigma^A$  was investigated and peptide mimics proposed to compete with  $\sigma^A$  for binding to the core enzyme. Because of software limitations and sequencing errors in  $\sigma^A$  used in the previous study, we have re-investigated this essential interaction. We show that the previous (incorrect) sequence of  $\sigma^A$  resulted in a flawed analysis of its interaction with RNAP. Our new model, and the importance of the corrected amino acid sequence, is supported by a protein–protein interaction assay. Our investigation of the interface between RNAP and  $\sigma^A$  will now permit accurate rational design of inhibitors of this interaction.

## Results

### sigA sequencing

When sequencing multiple clones containing *B. subtilis* *sigA* or regions of *sigA* (data not shown), a “mutation” was always observed in region 2.2. To investigate this further, *sigA* from *B. subtilis* strains 168, 168CA,



**Figure 1.** Alignment of region 2.2 of 15  $\sigma$  factors. Three-letter organism codes are as follows: Bsu, *B. subtilis*; Bha, *Bacillus halodurans*; Sau, *Staphylococcus aureus*; Mtu, *Mycobacterium tuberculosis*; Sco, *Streptomyces coelicolor*; Taq, *Thermus aquaticus*; Tth, *Thermus thermophilus*; Cmu, *Chlamydia muridarum*; Eco, *Escherichia coli*; Pae, *Pseudomonas aeruginosa*; Ret, *Rhizobium etli*; Tma, *Thermatoga maritima*; Mge, *Mycoplasma genitalium*. The asterisk indicates the incorrect amino acid for *B. subtilis* in sequence databases.

and SB19 (refer Table SI in Supporting Information) was sequenced. All three strains showed the same sequence, all with a single amino acid different to that found in the *B. subtilis* 168CA genome sequence SubtiList web server (<http://genolist.pasteur.fr/SubtiList/>). An example of a sequencing run is shown in the Supporting Information (Fig. S1). A substitution of G for T gives rise to a glutamine (CAG) in position 165, which was previously thought to be a histidine (CAT).<sup>16</sup>

To further support our data, a sequence alignment of sigma factors from a diverse range of Gram positive and negative bacteria was performed ( $\sigma$  region 2.2 shown in Fig. 1). The histidine (H) present in the *B. subtilis* *sigA* database sequence is highlighted with an asterisk. In all other sequences, including the less closely related alternative factors  $\sigma^F$  (*sigF*, *B. subtilis*) and  $\sigma^{32}$  (*rpoH*, *Escherichia coli*), the amino acid at the equivalent position is always glutamine (Q). Given our sequencing results and the alignments, we propose that *B. subtilis*  $\sigma^A$  also contains a glutamine at amino acid position 165. Subsequent projects to re-sequence the *B. subtilis* 168 genome have confirmed our results.<sup>17</sup> Comparison of the other RNAP subunits showed there were no other errors in sequencing.

### *B. subtilis* RNAP model construction and structure validation

Homology models of RNAP core and HE subunits were created individually and then combined to form the complete enzyme. The crystal structures of

*Thermus aquaticus* core<sup>8</sup> and *Thermus thermophilus* HE<sup>3</sup> were used as templates. The major subunits ( $\alpha$ ,  $\beta$ ,  $\beta'$ ) of *B. subtilis* RNAP share about 45% identity and 65% similarity with the corresponding *T. aquaticus* and *T. thermophilus* subunits (refer Table SII in the Supporting Information for a full identity and similarity list).

The enzyme models were subjected to structure clean-up and optimization. After each minimization set, the quality of the model was examined by Ramachandran analysis<sup>18–20</sup> (refer Fig. S2 in the Supporting Information). A check of the templates showed that only 54% of *T. aquaticus* core residues were in the favored regions of the Ramachandran plot (81% in the allowed regions), while for *T. thermophilus* HE 86% of residues were in the favorable region (97% in the allowed regions). This very low score for the *T. aquaticus* core is not ideal for a template, but this is the only core RNAP structure available that is not in complex with an antibiotic. The *T. thermophilus* HE score of 86% is very high considering the large size of the enzyme. The scores for the homology models are comparable to the crystal structure scores, suggesting the models are of good quality. These scores were obtained for models refined by minimization of the sidechains with the backbone constrained. Further minimization without constraints led to deterioration of the backbone secondary structures for both our models and the crystal structures. Only models minimized with the backbone constrained were used in further analysis. An alternative approach to enzyme construction was also undertaken to investigate if there was a difference between minimizing the subunits first and then putting together the enzyme, or putting the enzyme together from unminimized structures and then minimizing. This second procedure was performed to account for possible changes in the structure of each subunit under the influence of the surrounding subunits. However, the same result was obtained (data not shown).

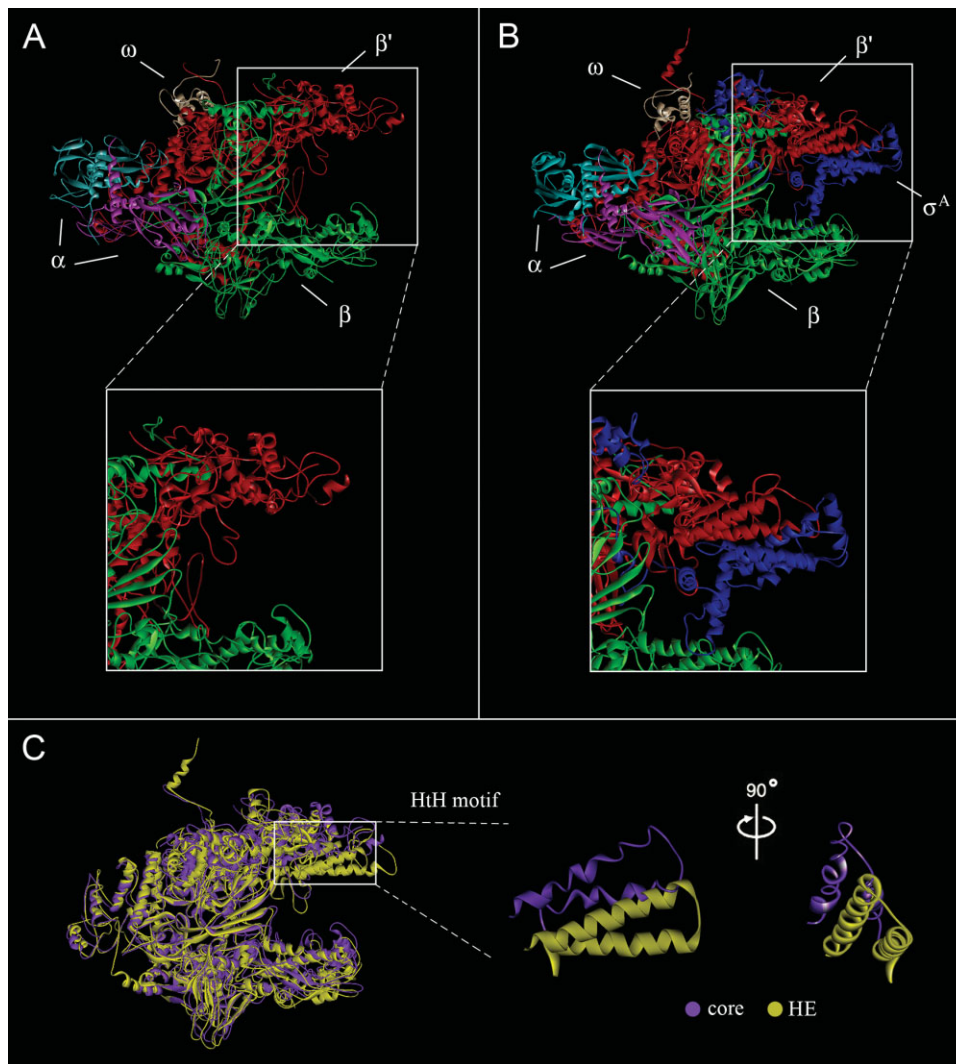
The final *B. subtilis* RNAP models, with sidechains minimized, are shown in Figure 2. Panel A shows the core enzyme while Panel B shows the HE. Parts of the  $\beta$  and  $\beta'$  subunits where there are gaps in the template structure (>10 amino acids) are not displayed, because the models are not reliable under these circumstances. Sequence alignments available as Supporting Information (A1–A4) show these regions. Both models show the classic crab claw shape made with the two  $\alpha$  subunits (cyan and pink) at the back, and the  $\beta$  (green) and  $\beta'$  (red) subunits forming the two halves of the DNA-binding channel.  $\omega$  (cream) is positioned on top of  $\beta'$  and  $\sigma^A$  (blue) is on the top front of the HE only and protrudes across the channel. The region where the major interaction between  $\sigma^A$  and the core occurs is shown in the insets of Panels A and B. The interacting helix-turn-helix (HtH) motif of  $\beta'$ (red) can be seen clearly in the Panel B inset.

### **RNAP is a dynamic structure: Changes from core to HE**

Apart from the obvious addition of  $\sigma^A$ , the core enzyme undergoes a number of changes when forming the HE. The major change is highlighted in Panel C of Figure 2. The top half of the “claw” of the HE (yellow) has closed down to hold  $\sigma^A$  compared to the more open channel in the core enzyme (purple). Showing the HtH motif only from two different angles emphasizes the fact that this region of the HE has moved down in comparison to the core, rather than rotated. The presence of more disordered regions with less secondary structure in the core enzyme may be explained by the lower resolution and quality of the core crystal structure template. To analyze the flexibility of RNAP in more detail, all subunits were subjected to distance difference matrix (DDM) analysis.<sup>21</sup> This allows evaluation of relative changes in the position of each amino acid C $\alpha$  atom with respect to all other C $\alpha$  atoms in each subunit. As has been suggested before (reviewed in Ref. 11), our models confirm that the two  $\alpha$  subunits, the  $\omega$  subunit and the C-terminal halves of the  $\beta$  and  $\beta'$  subunits move as rigid bodies upon binding of  $\sigma$ . This leads to DDMs with no significant changes within any of the subunits. The DDMs of the N-terminal halves of  $\beta$  and  $\beta'$  (refer Fig. S3 in the Supporting Information) did show changes between the core and HE. However, the largest of these changes are due to the insertions in the *B. subtilis* sequence when compared to the template sequences, and are thus not meaningful.

### **The interface between $\beta'$ and $\sigma^A$**

We minimized the *B. subtilis* homology model and the *T. thermophilus* crystal structure in exactly the same way, so any differences observed are meaningful. The major interaction between  $\beta'$  and  $\sigma^A$  is between a coiled-coil region of  $\beta'$ (comprising a HtH motif) and region 2.2 of  $\sigma^A$  (a single helix), highlighted red and blue, respectively, in the center of Figure 3. Panels A–D of Figure 3 show the amino acids forming hydrogen bonds (A and C) and hydrophobic interactions (B and D) in both the *B. subtilis* and *T. thermophilus* structures. Our *B. subtilis* model has an extensive network of proposed hydrogen bonds as shown in Panel A. Interestingly, one of these is formed by  $\sigma^A$  Q165, which was previously thought to be a histidine (refer previous section), and when the enzyme is modeled with H in position 165, this hydrogen bond is not formed (data not shown). The hydrogen bond network seen in the *T. thermophilus* crystal structure [Fig. 3(C)] is even more extensive, with more amino acids involved. It has been suggested, that no single molecular cause can explain the greater thermal stability of thermophilic proteins, and that they rely instead on combinations of stabilizing effects.<sup>22</sup> On the other hand, older studies proposed that a greater number of specific interactions, particularly electrostatic interactions and hydrogen bonds, are largely responsible for the higher stability of thermophilic proteins.<sup>23,24</sup> The



**Figure 2.** Homology models of *B. subtilis* RNAP core (A) and RNAP HE (B). The inset panels show regions of RNAP that change upon binding of  $\sigma$ , from core to HE. The enzyme subunits are color coded as indicated. Panel C: Superimposition of the two RNAP models, using the  $\alpha$  subunits. Only the  $\beta$  and  $\beta'$  subunits are shown.

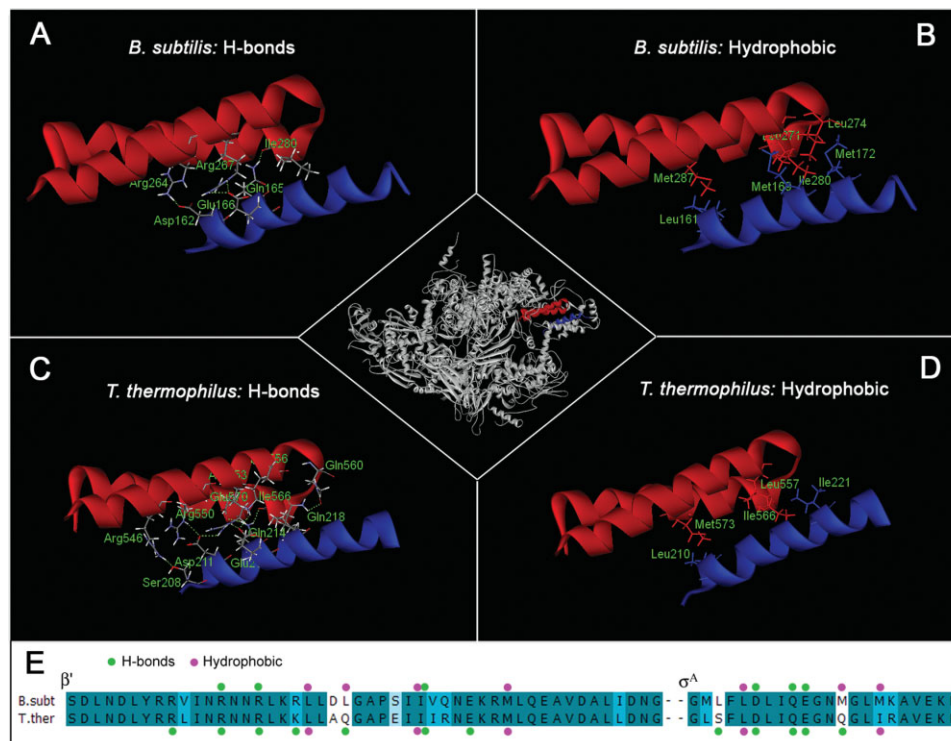
alignment in Panel E summarizes the amino acids involved in hydrogen bonds (green dots) and shows that corresponding amino acids in *B. subtilis* and *T. thermophilus* are implicated, with *T. thermophilus* having an extra four amino acids from  $\beta'$  and an extra two from  $\sigma^A$  involved. A detailed list of predicted hydrogen bonds is given in Table I, the predicted hydrogen bond lengths are listed as an indication of the likely strength of the bonds. While more hydrogen bonds are predicted for *T. thermophilus*, most of these are expected to be weak ( $>2 \text{ \AA}$ ).

Figure 3 (Panels B and D) shows the residues involved in hydrophobic interactions between  $\beta'$  and  $\sigma^A$ . They are illustrated as sidechains from each subunit extending towards each other. This occurs in two areas along the interface in both the *B. subtilis* model (B) and the *T. thermophilus* crystal structure (D), with more *B. subtilis* amino acids predicted to be involved in hydrophobic interactions. In *T. thermophilus*, three residues from  $\beta'$  and two from  $\sigma^A$  interact and in *B.*

*subtilis*, two additional amino acids from each chain are involved (Panel E, pink dots). The extra amino acids involved in *B. subtilis* are  $\sigma^A$  M169 and  $\beta'$  L274. Interestingly, the corresponding amino acid in *T. thermophilus* at both these positions is a polar glutamine, and these two amino acids are instead involved in hydrogen bonds.

#### Comparison of new HE model to previous modeling

When comparing the new *B. subtilis* holoenzyme model from this study to a previous model,<sup>15</sup> several important interactions not observed previously were found between the residues at the  $\beta'$ - $\sigma^A$  interface, as summarized in Table SIII in the Supporting Information. One difference in particular is the presence of  $\sigma^A$  Q165 in the predicted hydrogen bond network of the model created in this study. Q165 is the residue previously thought to be a histidine. With the correct amino acid now in this position, a more extensive hydrogen



**Figure 3.** The major interaction interface between  $\beta'$  and  $\sigma^A$ . The center shows the RNAP HE with the  $\beta'$  coiled-coil in red and  $\sigma^A$  region 2.2 in blue. The amino acids forming the hydrogen bond network in the *B. subtilis* HE model and the *T. thermophilus* crystal structure are shown in Panels A and C, hydrophobic interactions in *B. subtilis* and in *T. thermophilus* are in Panels B and D. A summary of the interactions is shown in the alignment in E, with hydrogen bonds and hydrophobic interactions marked with green and pink dots, respectively.

bond network is predicted. The interface in our new HE model is also more similar to the interface present in the *T. thermophilus* crystal structure (as shown in Fig. 3). This probable similarity between distantly related organisms is significant and we suggest it further validates the use of a non-pathogenic model organism closely related to pathogens for experimental design of inhibitors.

#### Design of peptides to inhibit HE formation

A necessary first step to test the theory that disrupting binding of  $\sigma^A$  to  $\beta'$  will inhibit transcription initiation, is to create mutants of  $\sigma^A$  in the 2.2 region by mutating the residues that are proposed to be important in

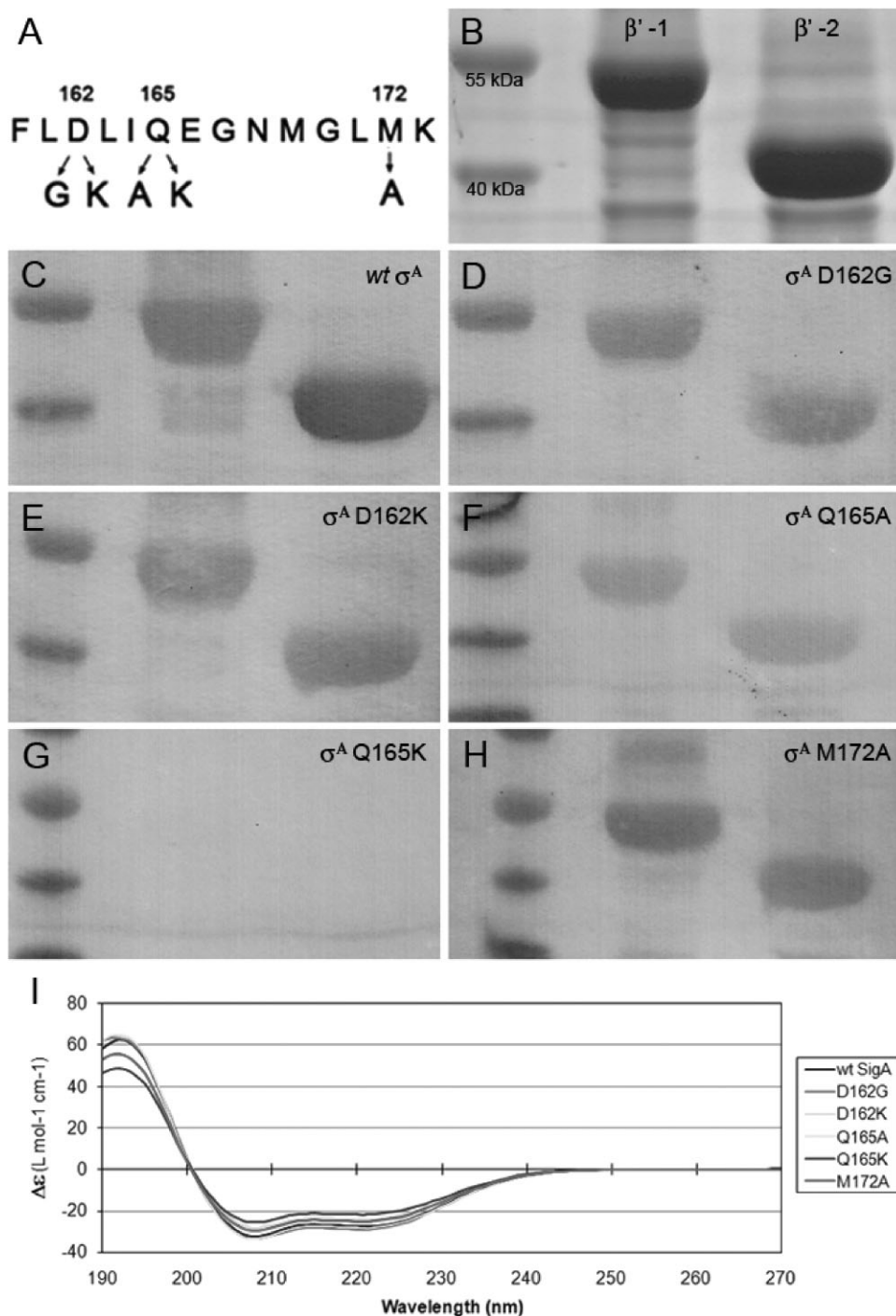
the interaction (as shown in Fig. 3) and observing their effect on binding. Inhibitors were designed based on  $\sigma^A$  and not  $\beta'$ , as  $\sigma^A$  is limiting in the cell. An initial set of mutations were chosen to have either minimal (D162G, Q165A, and M172A) or dramatic (D162K, Q165K) effects on binding, as illustrated in Figure 4(A).

#### $\sigma^A$ Mutants have altered binding capacity for $\beta'$

Mutations were introduced into *B. subtilis sigA* using site-directed mutagenic PCR. The mutated and wild-type genes were inserted into a T7 expression vector and overproduction of proteins carried out in *E. coli*.

**Table I.** Summary of Amino Acids and Atoms Involved in the Predicted Hydrogen Bond Network Between  $\beta'$  and  $\sigma^A$  in *B. subtilis* and *T. thermophilus*

$\sigma^A$ amino acid	<i>B. subtilis</i>				Distance (Å)	<i>T. thermophilus</i>				Distance (Å)
	Atom	$\beta'$ amino acid	Atom			$\sigma^A$ amino acid	Atom	$\beta'$ amino acid	Atom	
D162	OD2	R264	HH21	1.8	S208	OG	R546	HH12	1.7	
E166	OE2	R267	HE	1.9	D211	OD1	R550	HE	2.2	
E166	OE2	R267	HH21	2.4	E215	OE2	R553	HE	2.3	
Q165	HE22	I280	O	1.9	D211	OD1	R553	HH12	2.4	
					E215	OE2	K556	HZ2	2.3	
					Q214	HE21	E570	OE2	2.5	
					Q214	HE22	I566	O	1.9	
					Q218	HE21	Q560	OE1	2.4	



**Figure 4.** Evaluation of the RNA polymerase binding capacity of  $\sigma^A$  mutants. (A) Schematic showing the different  $\sigma^A$  mutations generated. (B) Coomassie blue stain of the  $\beta'$  fragments used in the Far-Western blots. Far-Western blots show binding of wild-type  $\sigma^A$  (C),  $\sigma^A$  D162G (D),  $\sigma^A$  D162K (E),  $\sigma^A$  Q165A (F),  $\sigma^A$  Q165K (G), and  $\sigma^A$  M172A (H) to the  $\beta'$  fragments. Each blot shows a molecular weight ladder in the first lane followed by two lanes of  $\beta'$ -bound  $\sigma^A$ . (I) Circular dichroism analysis of  $\sigma^A$  and mutants.

In contrast to well established protocols used to renature  $\sigma$  expressed as an insoluble form in inclusion bodies,<sup>25</sup> we optimized conditions so that wild-type and mutant  $\sigma^A$  proteins were overproduced in soluble form. Wild-type  $\sigma^A$  produced by our optimized methods has been shown to be functional in transcription assays.<sup>26</sup>

Binding of  $\sigma^A$  proteins to  $\beta'$  was tested by Far-Western blotting. Two fragments of  $\beta'^{27}$  were used to test the binding of  $\sigma^A$  and its mutant forms. These fragments overlap, so both contain the coiled-coil region of interest, as shown in Figure S4 of the Supporting Information. Far-Western blotting is a well established method for identifying protein-protein

interactions and has been used in previous studies to identify where  $\sigma^{70}$  binds to RNAP in *E. coli*<sup>28</sup> and where NusA binds to RNAP in *B. subtilis*.<sup>27</sup>

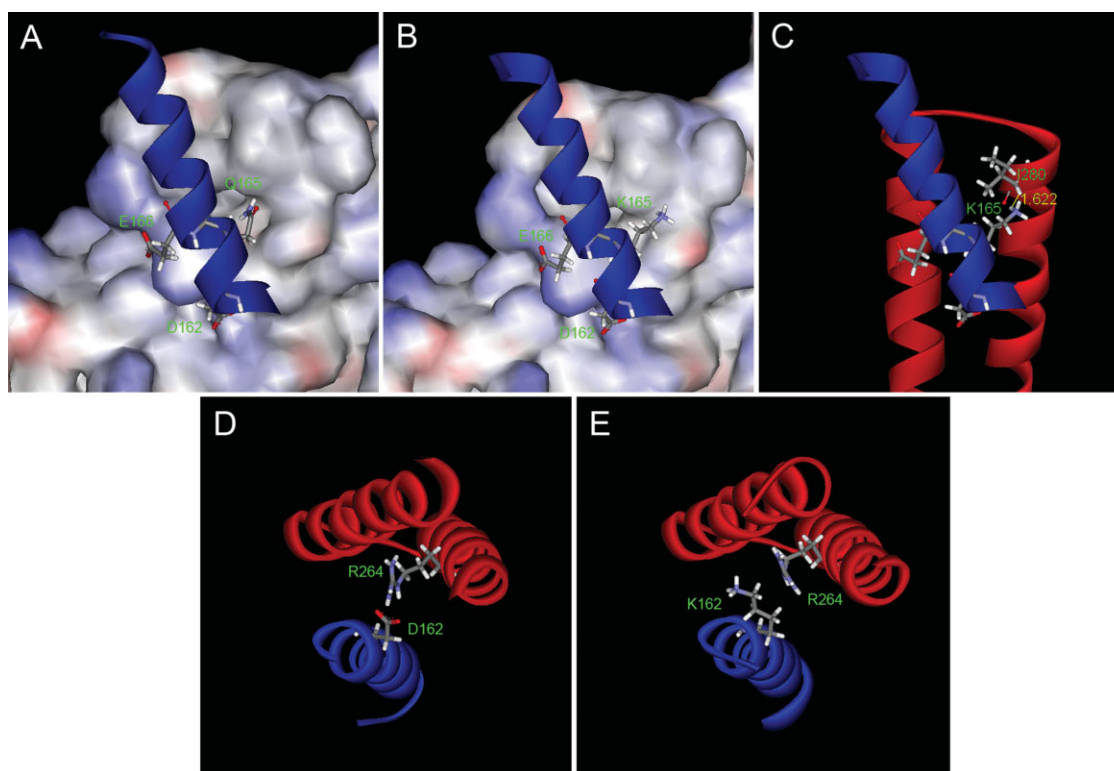
The two  $\beta'$  fragments are shown in Figure 4(B), with molecular weights of  $\sim 40$  and 55 kDa when compared with the molecular weight marker (first lane of each panel). Upon addition of wild-type  $\sigma^A$  to the membrane, binding to both  $\beta'$  fragments was observed as shown in Figure 4(C) (second and third lanes). By comparing the wild-type data to that of the mutant proteins, qualitative comparisons can be made as to their ability to bind to  $\beta'$ .  $\sigma^A$  mutations D162G, D162K, and M172A (Panels D, E, and H) do not appear to have a large effect on the ability of  $\sigma^A$  to bind to  $\beta'$ . On the other hand, mutating amino acid 165 drastically reduces the capacity for  $\sigma^A$  binding. Q165K (Panel G) abolishes binding altogether, while Q165A (Panel F) clearly shows a reduced level of binding.

To confirm that the altered binding was due to the change in the  $\sigma^A$ - $\beta'$  interface and not due to protein misfolding, circular dichroism (CD) was performed to compare the structures of the  $\sigma^A$  proteins. All mutants and the wild-type  $\sigma^A$  displayed the same CD spectra (Panel I, Fig. 4) indicating that all proteins were folded correctly.

### Modeling of $\sigma^A$ mutants

Mutations were introduced into the *B. subtilis* HE model to observe their effects on the  $\sigma^A$ - $\beta'$  interface. In wild-type  $\sigma^A$  Q165 is proposed to dip into a hydrophobic pocket [Fig. 5(A)] and form a hydrogen bond with the mainchain oxygen of  $\beta'$  I280 (refer also Fig. 3 and Table I). With lysine (K) in its place, the sidechain does not appear to be able to fit into the pocket [Fig. 5(B)], which would lead to a loss of the hydrogen bond, as well as the hydrophobic interactions between the sidechain of Q165 and the pocket. Our model also suggests the possibility of a clash with  $\beta'$  I280 [Fig. 5(C)]. When alanine (A) is at this position, decreased binding was seen rather than complete abolishment of the interaction [Fig. 4(F)]. Alanine has a short, hydrophobic sidechain, so some hydrophobic interactions with the pocket would be possible, but the hydrogen bond would be lost.

Far-Western blots also showed that mutating the very highly conserved D162 did not greatly interfere in the ability of  $\sigma^A$  to bind [Fig. 4(D,E)]. In the wild-type model, D162 forms a hydrogen bond with  $\beta'$  R264 [Fig. 5(D)]. Mutations in this position appear to be tolerated because the interaction is towards the end of the interface. Hence, the lysine in the D162K mutant may be able to bend out of the way, with binding still



**Figure 5.** Modeling of the interaction of mutant  $\sigma^A$  with RNAP.  $\sigma^A$  region 2.2 is shown in blue and the surface of  $\beta'$  is colored according to the electrostatic potential in Panels A and B. The interacting region of the  $\beta'$  subunit is represented by a red ribbon in Panels C–E. Wild-type Q165 fits into a hydrophobic pocket on the surface of  $\beta'$  (A). With K165 modeled in its place, the sidechain cannot fit into the pocket (B) and there is a clash with I280 of  $\beta'$  (C, clash shown in yellow). Wild-type D162 forms a hydrogen bond with  $\beta'$  R264 (D). With K modeled into position 162 (E) binding can still occur because the sidechain bends away from the interface.

possible across the rest of the interface (Fig. 5, compare E to D). However, there would still be a loss of a hydrogen bond, which may translate to slightly weaker binding using more sensitive assays. Similarly, the M172A mutation does not show dramatically decreased interaction, possibly because it is also positioned at the end of the interface [Fig. 3(B)].

## Discussion

We detected a mutation in the published sequence of *B. subtilis*  $\sigma^A$ , and this was recently confirmed by other authors.<sup>17</sup> The correct amino acid at position 165, in region 2.2 of  $\sigma^A$  is glutamine (Q), not histidine (H), as previously thought.<sup>16</sup> This is significant for the current work, because the major interaction between  $\beta'$  and  $\sigma^A$  in the RNAP holoenzyme involves  $\sigma^A$  region 2.2.

Previously, a homology model of *B. subtilis* HE was published<sup>15</sup> which included the incorrect amino acid in the  $\sigma^A$  sequence. This model has now been corrected and extensively refined and validated. A model of the core enzyme has also been built and compared to the HE. The differences between core and HE models showed that the  $\beta$  and  $\beta'$  subunits move as rigid bodies (as has previously been suggested<sup>11</sup>), rather than the subunits being flexible within themselves. The comparison between the  $\beta'$  and  $\sigma^A$  interface in the models and the crystal structures used as templates has yielded interesting results. The crystal structures are both of RNAP from thermophiles, where a greater number of hydrogen bonds seem to be involved in the  $\sigma^A$ - $\beta'$  interaction than in the mesophilic *B. subtilis*, where more hydrophobic interactions appear to be involved.

Our models propose that the corrected amino acid Q165 is crucial for the  $\sigma^A$ - $\beta'$  interaction. We designed several mutants to test our model and show that targeted disruption of residues at the  $\sigma^A$ - $\beta'$  interface could lead to decreased interaction.

$\sigma^A$  mutants were prepared and used to demonstrate that mutation of Q165 dramatically reduces the ability of  $\sigma^A$  to bind to  $\beta'$ . Binding was significantly decreased for the Q165A mutant, and was undetectable in the Q165K mutant. Previous studies have also shown that mutation of the equivalent residue in *E. coli*  $\sigma^{32}$  (Q80; an alternate  $\sigma$  factor) to R or N, reduces core association markedly and greatly lowers the level of transcription.<sup>29</sup> Similar results were seen with an *E. coli*  $\sigma^{70}$  Q406A mutant (equivalent to *B. subtilis* Q165A).<sup>30</sup> Interestingly, this study also had a D403N mutation (equivalent to *B. subtilis*  $\sigma^A$  D162) which showed that 9 times the amount of mutant protein was required to achieve the same level of binding as wild-type  $\sigma^{70}$ . Our results did not show a significant effect on binding when mutating this residue, but this could be due to the lack of sensitivity of our assay. Importantly, mutation of this residue in *E. coli* did not abolish the interaction completely, which fits well with

our modeling of mutant D162K. Additionally, mutation of hydrophobic residues along the interface in *E. coli*  $\sigma^{70}$  had little or no effect on core binding. This was also observed in this study with the M172A mutation.<sup>30</sup>

Combining computer-modeling and *in vitro* binding assays, we have thus been able to take the first steps towards validating the  $\sigma^A$ - $\beta'$  interaction in bacterial RNAP holoenzymes as a potential new target for the development of antibacterial agents with a novel mode of action.

## Materials and Methods

### Sequence and template acquisition

All sequences of *B. subtilis* proteins were obtained from the SubtiList web server ([genolist.pasteur.fr/SubtiList](http://genolist.pasteur.fr/SubtiList)). Gene names were as follows:  $\alpha$ -*rpoA*,  $\beta$ -*rpoB*,  $\beta'$ -*rpoC*,  $\omega$ -*yloH*,  $\sigma^A$ -*sigA*. Protein sequences from all other organisms were obtained from NCBI ([www.ncbi.nih.gov](http://www.ncbi.nih.gov)). These represent RNAP subunits from a range of Gram positive and Gram negative bacteria.

Crystal structure templates were obtained from the Protein Data Bank (PDB)<sup>31</sup> ([www.rcsb.org/pdb](http://www.rcsb.org/pdb)). The template for core RNAP was of *T. aquaticus* core,<sup>8</sup> PDB code 1HQM, chains A–E; and the template for the HE was of *T. thermophilus* HE,<sup>3</sup> PDB code 1IW7, chains A–F.

### DNA sequencing

*sigA* from three different strains of *B. subtilis* was amplified by PCR and then sequenced (Australian Genome Research Facility, Brisbane, Australia). Two of the strains used were "wild-type" 168 laboratory strains from different research groups; this is important as the provenance of strains is often hard to determine. Strain 168, a highly transformable tryptophan auxotroph, was originally obtained by X-ray mutagenesis in the 1940s<sup>32</sup> and has since been passaged through many laboratories, making unambiguous establishment of strain provenance problematic. This accounts for slight differences between strains/stocks between labs. The strains used in this study included: 168 (laboratory stock, P.J. Lewis), 168CA,<sup>33</sup> the type of strain used in the *B. subtilis* genome sequencing project and SB19 (R.G. Wake, The University of Sydney). Refer Supporting Information, Table SI for strain details.

### Homology modeling

Multiple sequence alignments were completed for each subunit using the ClustalW2 program<sup>34</sup> ([www.ebi.ac.uk/Tools/clustalw2](http://www.ebi.ac.uk/Tools/clustalw2)) and then submitted to the SWISS-MODEL server<sup>35</sup> (<http://swissmodel.expasy.org>) using the "alignment interface." In the case of  $\beta'$  (*rpoC*) and  $\omega$  (*yloH*), the position of gaps would not allow a model to be created, so the gaps in the



alignment had to be manipulated manually using Swiss-Pdb Viewer<sup>35</sup> and then submitted to SWISS-MODEL using the project mode. Alignments for these subunits are provided in the Supporting Information. Core  $\beta'$  had to be split into two parts to obtain the homology model.

### Structure assembly and clean-up

To create the entire enzyme, each subunit of *B. subtilis* core and HE was imported on top of the crystal structure template to check for correct positioning. The template was then deleted, leaving behind *B. subtilis* RNAP models only.

Swiss-Pdb Viewer<sup>35</sup> was used to clean-up models before further refinement. Amino acids with steric clashes were selected and then "fixed" (repaired) using the "quick and dirty" method. This was repeated until there was no improvement in the number of amino acids with clashes.

### Structure minimization

Energy minimization of the structures was carried out using Discovery Studio 2.0 (Accelrys). Conditions were set by the CHARMM forcefield<sup>36</sup> as implemented in the Accelrys software. First, hydrogens were added to the model or crystal structure and the remainder of the structure was constrained (fixed position in space) to refine the positioning of the hydrogens. Next, the backbone was constrained and minimization carried out to refine the positioning of the sidechains; and finally, the entire structure was minimized with no constraints. For mutation analysis, single amino acids were substituted and minimized alone (i.e., with the rest of the enzyme constrained) followed by minimization of sidechains within 10 Å of the  $\sigma^A$ - $\beta'$  interface (backbone constrained).

Each set of refinements involved first minimizing with the Steepest Descent algorithm until the maximum derivative was less than 0.1 kJ/(mol Å); and then minimizing with the Adopted-Basis Newton Raphson (ABNR) algorithm until the maximum derivative was less than 0.01 kJ/(mol Å). All other parameters were kept at default values.

### Structure evaluation

Quality of models was checked using online programs, including MolProbity<sup>18</sup> (<http://kinemage.biochem.duke.edu>) and PROCHECK<sup>19</sup> (Protein Data Bank). These packages make use of Ramachandran plots<sup>20</sup> and clash-scores, and determine outliers in areas such as bond lengths, angles, and spatial arrangement. To assess subunit flexibility, the ProFlex2.0 program<sup>21</sup> was used to obtain distance difference matrix (DDM) plots. The  $\beta$  and  $\beta'$  subunits were split into two parts because the program cannot evaluate chains longer than 999 amino acids.

### Strain construction

A full strain and plasmid list can be found in the Supporting Information, Table SI.

*E. coli* DH5 $\alpha$  was used for all cloning procedures. *B. subtilis sigA* was amplified from 168 genomic DNA and inserted into pETMCSIII via the *Xba*I and *Acc*65I sites to create pNG590 (wild-type *sigA* with an N-terminal 6xHis tag to enable protein purification). PCR site-directed mutagenesis<sup>37</sup> was used to create mutants of *sigA* that were similarly inserted into pETMCSIII. The plasmids pNG570, 571, 572, 573, and 600 code for  $\sigma^A$  mutations M172A, D162G, D162K, Q165A, and Q165K, respectively.

### Protein overproduction, purification, and evaluation

*E. coli* strain C41(DE3) (Table SI) was transformed with each *sigA*-containing plasmid and resulting colonies placed into 1L of Luria-Bertani broth (LB) containing 100  $\mu$ g/mL ampicillin and grown shaking at 37°C. At  $A_{600}$  0.6–1.0 isopropyl- $\beta$ -D-thiogalactoside (IPTG) was added to 0.5 mM and shaking continued at room temperature for further 5 h. Cells were harvested by centrifugation and stored at –80°C.

Cells were resuspended in 20–40 mL buffer A (20 mM NaH<sub>2</sub>PO<sub>4</sub>, 500 mM NaCl, 20 mM imidazole, pH 7.5) and lysed using an EmulsiFlex-C5 homogenizer (Avestin). Proteins were purified by affinity chromatography using a 1 mL HisTrap HP column (GE Healthcare) and an AKTA FPLC system (GE Healthcare). The cell lysate (soluble fraction) was applied to the column then washed with 6.6% buffer B (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 500 mM NaCl, 400 mM imidazole, pH 7.5) and eluted with a step to 100% buffer B. The protein peak was collected and dialyzed into storage buffer by two step dialysis (step 1: 50 mM NaH<sub>2</sub>PO<sub>4</sub>, 150 mM NaCl, 1 mM EDTA, 0.1 mM dithiothreitol, pH 7.5; step 2: 50 mM NaH<sub>2</sub>PO<sub>4</sub>, 150 mM NaCl, 0.1 mM dithiothreitol, 30% glycerol, pH 7.5) and stored at –80°C.

To check for correct protein folding, CD spectra were recorded from 300 nm to 190 nm on a JASCO J-810 spectropolarimeter, with proteins diluted to ~0.25 mg/mL in 10 mM NaH<sub>2</sub>PO<sub>4</sub>.

### Far-Western blotting

pNG482 and pNG483<sup>27</sup> containing overlapping  $\beta'$  fragments (both with the coiled-coil region of interest) were transformed into BL21(DE3)pLysS. Colonies were added to 10 mL LB containing 100  $\mu$ g/mL ampicillin and grown at 37°C until  $A_{600}$  reached 0.5–0.8, where IPTG was then added (to 0.5 mM) and the culture grown for a further 3 h at 37°C; 1 mL aliquots were centrifuged and cells resuspended to an  $A_{600}$  of 10 in TES [200 mM Tris-HCl (pH 7.5), 5 mM EDTA, 100 mM NaCl]; 5–10  $\mu$ L of each cell lysate was subjected to SDS-PAGE and then transferred to a nitrocellulose

membrane. Following transfer, the membrane was washed three times (5 min each) in PBS and three times in FWB (20 mM Tris, 100 mM NaCl, 0.5 mM EDTA, 10% glycerol, 0.1% Tween-20). The membranes were blocked and proteins refolded by immersing in FWB + 5% skim milk (w/v) for 1 h while shaking. The membrane was then probed with a  $\sigma^A$  protein ( $\sim 10 \mu\text{g}$  protein added to FWB + 5% skim milk) for 1 h and then washed three times in FWB. Rabbit polyclonal anti- $\sigma^A$  antibody (M. Yudkin, University of Oxford) diluted 1:3000 in FWB + 5% milk was added to the membrane and left to incubate for 1 h, followed by another three washes and incubation with 1:3000 goat-anti-rabbit horseradish peroxidase (BioRad). After the three final washes, the antibody was detected using an Opti-4CN™ substrate kit (BioRad) following the manufacturer's instructions.

### Acknowledgments

The authors thank M. Yudkin, University of Oxford, for kindly supplying polyclonal anti- $\sigma^A$  anti serum. E.J. was supported by a Ph.D. scholarship from the ARC and the University of Newcastle.

### References

- Ebright RH (2000) RNA polymerase: structural similarities between bacterial RNA polymerase and eukaryotic RNA polymerase II. *J Mol Biol* 304:687–698.
- Burgess RR, Travers AA, Dunn JJ, Bautz EKF (1969) Factor stimulating transcription by RNA polymerase. *Nature* 221:43–46.
- Vassilyev DG, Sekine S, Laptenko O, Lee J, Vassilyeva MN, Borukhov S, Yokoyama S (2002) Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* 417:712–719.
- Murakami KS, Masuda S, Campbell EA, Muzzin O, Darst SA (2002) Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* 296:1285–1290.
- Mooney RA, Davis SE, Peters JM, Rowland JL, Ansari AZ, Landick R (2009) Regulator trafficking on bacterial transcription units in vivo. *Mol Cell* 33:97–108.
- Gross CA, Chan C, Dombroski A, Gruber T, Sharp M, Tupy J, Young B (1998) The functional and regulatory roles of sigma factors in transcription. *Cold Spring Harb Symp Quant Biol* 63:141–155.
- Zhang G, Campbell EA, Minakhin L, Richter C, Severinov K, Darst SA (1999) Crystal structure of thermus aquaticus core RNA polymerase at 3.3 Å resolution. *Cell* 98:811–824.
- Minakhin L, Bhagat S, Brunning A, Campbell EA, Darst SA, Ebright RH, Severinov K (2001) Bacterial RNA polymerase subunit omega and eukaryotic RNA polymerase subunit RPB6 are sequence, structural, and functional homologs and promote RNA polymerase assembly. *Proc Natl Acad Sci USA* 98:892–897.
- Murakami KS, Masuda S, Darst SA (2002) Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science* 296:1280–1284.
- Vassilyev DG, Vassilyeva MN, Perederina A, Tahirov TH, Artsimovitch I (2007) Structural basis for transcription elongation by bacterial RNA polymerase. *Nature* 448:157–162.
- Murakami KS, Darst SA (2003) Bacterial RNA polymerases: the whole story. *Curr Opin Struct Biol* 13:31–39.
- Hughes KT, Mathee K (1998) The anti-sigma factors. *Annu Rev Microbiol* 52:231–286.
- Andre E, Bastide L, Michaux-Charachon S, Gouby A, Villain-Guillot P, Latouche J, Bouchet A, Gualtieri M, Leonetti JP (2006) Novel synthetic molecules targeting the bacterial RNA polymerase assembly. *J Antimicrob Chemother* 57:245–251.
- Errington J (1993) *Bacillus subtilis* sporulation: regulation of gene expression and control of morphogenesis. *Microbiol Rev* 57:1–33.
- MacDougall IJ, Lewis PJ, Griffith R (2005) Homology modeling of RNA polymerase and associated transcription factors from *Bacillus subtilis*. *J Mol Graph Model* 23:297–303.
- Kunst F, Ogasawara N, Moszer I, Albertini AM, Alloni G, Azevedo V, Bertero MG, Bessieres P, Bolotin A, Borchert S, et al. (1997) The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* 390:249–256.
- Barbe V, Cruveiller S, Kunst F, Lenoble P, Meurice G, Sekowska A, Vallenet D, Wang T, Moszer I, Medigue C, et al. (2009) From a consortium sequence to a unified sequence: the *Bacillus subtilis* 168 reference genome a decade later. *Microbiology* 155:1758–1775.
- Lovell SC, Davis IW, Arendall WB, III, de Bakker PI, Word JM, Prisant MG, Richardson JS, Richardson DC (2003) Structure validation by calpha geometry: phi, psi and Cbeta deviation. *Proteins* 50:437–450.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check stereochemical quality of protein structures. *J Appl Crystallogr* 26:283–291.
- Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963) Stereochemistry of polypeptide chain configurations. *J Mol Biol* 7:95–99.
- Keller PA, Leach SP, Luu TT, Titmuss SJ, Griffith R (2000) Development of computational and graphical tools for analysis of movement and flexibility in large molecules. *J Mol Graph Model* 18:235–241, 299.
- Razvi A, Scholtz JM (2006) Lessons in stability from thermophilic proteins. *Protein Sci* 15:1569–1578.
- Kumar S, Tsai C-J, Nussinov R (2001) Thermodynamic differences among homologous thermophilic and mesophilic proteins. *Biochemistry* 40:14152–14165.
- Vogt G, Woell S, Argos P (1997) Protein thermal stability, hydrogen bonds, and ion pairs. *J Mol Biol* 269:631–643.
- Helmann JD, Purification of *Bacillus subtilis* RNA polymerase and associated factors. *Methods in enzymology*. In: SAA S Garges, Ed. (2003) RNA polymerases and associated factors, Part C, Volume 370. Academic Press, San Diego, CA pp 10–24.
- Yang X, Lewis PJ (2008) Overproduction and purification of recombinant *Bacillus subtilis* RNA polymerase. *Protein Expr Purif* 59:86–93.
- Yang X, Molimau S, Doherty GP, Johnston EB, Marles-Wright J, Rothnagel R, Hankamer B, Lewis RJ, Lewis PJ (in press) The structure of bacterial RNA polymerase in complex with the essential transcription elongation factor NusA. *EMBO Rep* 10:997–1002.
- Arthur TM, Burgess RR (1998) Localization of a sigma70 binding site on the N terminus of the *Escherichia coli* RNA polymerase beta' subunit. *J Biol Chem* 273:31381–31387.
- Joo DM, Ng N, Calendar R (1997) A  $\sigma_{32}$  mutant with a single amino acid change in the highly conserved region 2.2 exhibits reduced core RNA polymerase affinity. *Proc Natl Acad Sci USA* 94:4907–4912.

30. Sharp MM, Chan CL, Lu CZ, Marr MT, Nechaev S, Merritt EW, Severinov K, Roberts JW, Gross CA (1999) The interface of  $\sigma$ , with core RNA polymerase is extensive, conserved, and functionally specialized. *Genes Dev* 13:3015–3026.
31. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. *Nucleic Acids Res* 28:235–242.
32. Burkholder PR, Giles NH, Jr. (1947) Induced biochemical mutations in *Bacillus subtilis*. *Am J Bot* 34:345–348.
33. Anagnostopoulos C, Spizizen J (1961) Requirements for transformation in *Bacillus subtilis*. *J Bacteriol* 81:741–746.
34. Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD (2003) Multiple sequence alignment with the clustal series of programs. *Nucleic Acids Res* 31:3497–3500.
35. Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 18:2714–2723.
36. Brooks BR, Brucoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J Comput Chem* 4:187–217.
37. Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR (1989) Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* 77: 51–59.