2006

Structural Analysis of an RNA Polymerase Sigma Factor and Its Anti-sigma: sigma28/FLGM

Margareta K. Sörenson

STRUCTURAL ANALYSIS OF AN RNA POLYMERASE σ FACTOR AND ITS ANTI-σ: σ^{28}/FLGM

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Margareta K. Sörenson

June 2006

STRUCTURAL ANALYSIS OF AN RNA POLYMERASE σ FACTOR AND ITS ANTI-σ: σ^{28}/FLGM

Margareta K. Sörenson, Ph.D.

The Rockefeller University 2006

Eubacterial sigma (σ) factors are required during the initiation of transcription, where they play key roles in promoter recognition, melting, and transcriptional regulation. σ is a dissociable subunit of RNA polymerase (RNAP), and sequence-specifically recognizes promoters, but only after associating with RNAP to form the holoenzyme. In the holoenzyme, the major σ domains (σ_2, σ_3, and σ_4) are spread across the surface, with the promoter binding surfaces solvent exposed and appropriately positioned for simultaneous recognition of the -10 and -35 promoter elements. A ~30 residue linker connecting σ_3 and σ_4 is unfolded and threads through the interior of the enzyme. The structure of free σ is unknown, but biochemical studies suggest that binding to core induces a large conformational change in σ. The flagellar σ factor, σ^{28} (FliA, σ^{D}), is inhibited by the anti-σ factor, FlgM, which both prevents and destabilizes its association with core RNAP.

NMR analysis of *Salmonella typhimurium* (St) FlgM suggests that it is unstructured in isolation, but gains structure in the C-terminal half upon binding to σ^{28}.

This work describes the 2.3 Å x-ray crystal structure of the σ^{28}/FlgM complex from *Aquifex aeolicus* (Aa). Both halves of FlgM are ordered and involved in inhibition of σ^{28} by sterically occluding its core binding surfaces, and by stabilizing it in an inactive conformation. Modeling suggests that FlgM can interact with σ^{28} in the holoenzyme, explaining its ability to destabilize the σ^{28} holoenzyme. The entire σ^{28} is ordered, and provides the first structure of an intact σ factor in the absence of RNAP. There are extensive interdomain contacts, and the promoter binding surfaces of σ_2 and σ_4 are occluded by the σ_3-σ_4 linker and σ_3, respectively. The σ_3-σ_4 linker forms a bent α-helix, which is incompatible with the holoenzyme conformation, suggesting that a helix-coil transition may accompany holoenzyme formation.

Double cysteine mutants of σ^{28}, predicted to form interdomain disulfides in this conformation, form even in the absence of FlgM, suggesting that free σ^{28} adopts this conformation. The disulfide bonded species predominate at equilibrium, indicating that they form in the major solution conformation.

Till min mamma, Barbro, och min pappa, Sverre
för allt dom lärt mig, särskilt för musiken.

ACKNOWLEDGEMENTS

I would like to sincerely thank my thesis advisor, Seth Darst for his mentorship and guidance, but also for the independence he entrusted me with. I would also like to thank the members of my faculty advisory committee: Erec Stebbins, especially for help with crystallographic issues, Fred Cross, whose advice both on science and on career choices, I value greatly, David Eliezer, for his NMR perspective, and Carol Gross for serving as my external examiner, and for her encouragement and advice, on sigma factors and beyond.

I want to thank all my friends and colleagues in the Darst lab, for advice and mentoring, but also for company on the way to Java Girl. I particularly want to thank Bill computational, Mark for transcription assays, Liz for mentoring and for being a great "personal trainer", and Katsu, Shoko and Oriana for mentoring me when I first joined.

I'd like to thank Soumya Ray for his invaluable help and advice throughout these projects. I have really benefited from his ability to find simple ways that work.

I would also like to thank Brian Chait for being one of the reasons that I came to Rockefeller, and for always keeping his lab open to me.

I want to thank my former mentors at Hunter College, Shirley Raps and Klaus Grohmann, for their past and present support and encouragement that has meant so much to me.

Finally, I would like give my love and thanks to all my friends and family members for their support.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iv

TABLE OF CONTENTS vi

LIST OF FIGURES viii

LIST OF TABLES ix

1. INTRODUCTION 1

1.1 The σ subunit of bacterial RNA polymerase is required for initiation at promoters. 1

1.2 σ factors consist of several, flexibly linked domains 5

1.3 Region 1.1 inhibits DNA binding in primary σ factors. 7

1.4 The σ_3-σ_4 linker is unfolded in the holoenzyme. 8

1.5 Anti-σ factors are diverse in structure and inhibitory mechanism 11

1.6 σ^{28} activates transcription of late flagellar genes upon secretion of its anti-σ factor, FlgM. 14

2. EXPERIMENTAL PROCEDURES 22

2.1 σ^{28}/FlgM Structure Determination 22

2.2 Disulfide Analysis 28

3. CRYSTAL STRUCTURE OF THE σ^{28}/ FLGM COMPLEX 32

3.1 Aquifex aeolicus σ^{28}/ FlgM were chosen based on similarity to their St homologs and their crystallizability 32

3.2 Purification and crystallization of the σ^{28}/FlgM complex. 37

3.3 Phase determination 47

3.4 Structure solution, model building and refinement 54

4. ANALYSIS OF THE σ^{28}/FLGM CRYSTAL STRUCTURE 60

4.1 Overall Structure 60

4.2 Structure of σ^{28} 61

4.3 Structure of FlgM 69

4.4 Interactions between σ^{28} and FlgM, and the mechanism of σ^{28} inhibition 72

4.5 Conformation of σ^{28} *86*

5. DISULFIDE CROSS-LINKING INDICATES THAT FREE AND FLGM BOUND σ^{28} ADOPT SIMILAR CONFORMATIONS **99**

5.1 Introduction *99*

5.2 Results *101*

5.3 Discussion *112*

APPENDIX 1: EXPRESSION, PURIFICATION AND ACTIVITY OF RECOMBINANT RNA POLYMERASE FROM *AQUIFEX AEOLICUS* **119**

A1.1 Introduction *119*

A1.2 Experimental Procedures *120*

A1.3 Results & Discussion *123*

APPENDIX 2: ASSESSMENT OF INTRAMOLECULAR DISULFIDE FORMATION BETWEEN σ_3 AND σ_4 OF E.COLI σ^{70}. **130**

A2.1 Introduction *130*

A2.2 Experimental Procedures *130*

A2.3 Results & Discussion *132*

BIBLIOGRAPHY **136**

LIST OF FIGURES

Figure 1.1 Domain organization in σ^A
Figure 3.1 σ^{28} sequence alignment
Figure 3.2 FlgM sequence alignment
Figure 3.3 Aa σ^{28}/FlgM complex after gel filtration
Figure 3.4 Initial crystals of Aa σ^{28}/FlgM
Figure 3.5 Crystal Form I
Figure 3.6 Crystal Form II
Figure 3.7 Crystal Form III
Figure 3.8 Diffraction Pattern from Crystal Form I
Figure 3.9 Sites of introduced methionines.
Figure 4.1 Structure of the Aa σ^{28}/FlgM complex.
Figure 4.2 Structural superposition
Figure 4.3 Comparison of σ^{28} and σ^A.
Figure 4.4 FlgM structure and σ^{28} interactions.
Figure 4.5 Interaction types
Figure 4.6 Conserved interacting residues.
Figure 4.7 Steric occlusion of σ^{28} core binding surfaces by FlgM.
Figure 4.8 Promoter binding surfaces of σ^{28} are buried in interdomain interfaces.
Figure 4.9 Conformational change in σ^{28} upon holoenzyme formation.
Figure 5.1 Location of introduced cysteines in σ^{28}
Figure 5.2 Mass spectra of CNBr digested disulfide mutants
Figure 5.3 SDS-PAGE analysis of disulfide mutants.
Figure A1.1 Aa RNAP purificiation scheme 1.
Figure A1.2 Aa RNAP purification scheme 2
Figure A1.3.Aa RNAP purification scheme 3
Figure A1.4 Activity of recombinant Aa RNAP
Figure A2.1 Intramolecular disulfide formation in Ec σ^{70}
Figure A2.2 Disulfide formation under equilibrium conditions

LIST OF TABLES

Table 3.1 Crystallographic statistics
Table 4.1 Summary of σ^{28}/FlgM interactions
Table 4.2 Genetic and structural comparison of σ^{28}/FlgM interactions
Table 5.1 MALDI-TOF Measured and Calculated Masses of CNBr Fragments.
Table 5.2 Distances between mutated residues in the FlgM and holoenzyme bound conformations.
Table 5.3 Relative disulfide populations at equilibrium.

1. INTRODUCTION

1.1 The σ subunit of bacterial RNA polymerase is required for initiation at promoters.

All genes in eubacteria are transcribed by a single RNA polymerase (RNAP), a ~400 kDa multi-subunit enzyme. The catalytic core of RNAP (subunit composition $\alpha_2\beta\beta'\omega$) is structurally homologous to the eukaryotic RNA polymerase II (Zhang, Campbell et al. 1999; Cramer, Bushnell et al. 2001). The sigma (σ) subunit of RNAP, in contrast, which is required for initiation of transcription at promoters, has no structurally close counterpart in eukaryotes. σ reversibly associates with core RNAP to form the initiation competent holoenzyme ($\alpha_2\beta\beta'\omega\sigma$; Burgess et al., 1969; Travers & Burgess, 1969). In the holoenzyme, sigma sequence-specifically recognizes promoters and plays an essential role in the melting of double stranded DNA (reviewed in (Gross, Chan et al. 1998)) .

In addition to its direct involvement in the initiation process, σ is also a key protein in the regulation of gene expression. σ recruits the polymerase to different promoters at different rates, both by sensing the intrinsic promoter strength, which is determined by the degree to which the sequence specific elements conform to the consensus, and by sensing the presence of

activators or repressors (reviewed in (Dove, Darst et al. 2003)). Moreover, a given bacterial species has several σ factors which differ in promoter specificity. Typically, a single, primary σ factor is responsible for the transcription of genes expressed during vegetative growth. The prototypical primary σ factor is *Escherichia coli* (Ec) σ^{70}, and most structural information comes from its thermophilic homologs *Thermus aquaticus* (Taq) σ^{A} and *Thermus thermophilus* (Tth) σ^{A}. Bacteria also have a number of alternative σ factors, which are specific for sets of genes required only under special circumstances, such as the flagellar and the sporulation regulons. The activity and cellular concentration of alternative σ factors are highly regulated, and this regulation frequently involves post-translational inhibition by a specific anti-σ factor, which binds directly to its cognate σ factor (Hughes, 1998).

All primary and most alternative σ factors share sequence homology with σ^{70}. The only known exception is the alternative σ factor, σ^{54}, which also differs in its mechanism of initiation in various ways, including its requirement for an enhancer-binding activator ATPase (Zhang, Chaney et al. 2002). Members of the σ^{70} family have up to four regions of high sequence conservation (Lonetto, Gribskov et al. 1992) (Gruber and Bryant 1997), and these approximately correspond to structural domains, $\sigma_{1.1}$, σ_2, σ_3 and σ_4

(Malhotra, Severinova et al. 1996; Campbell, Muzzin et al. 2002), which are linearly arranged along the polypeptide chain, and connected by flexible linkers (Figure 1.1). This arrangement was initially inferred from

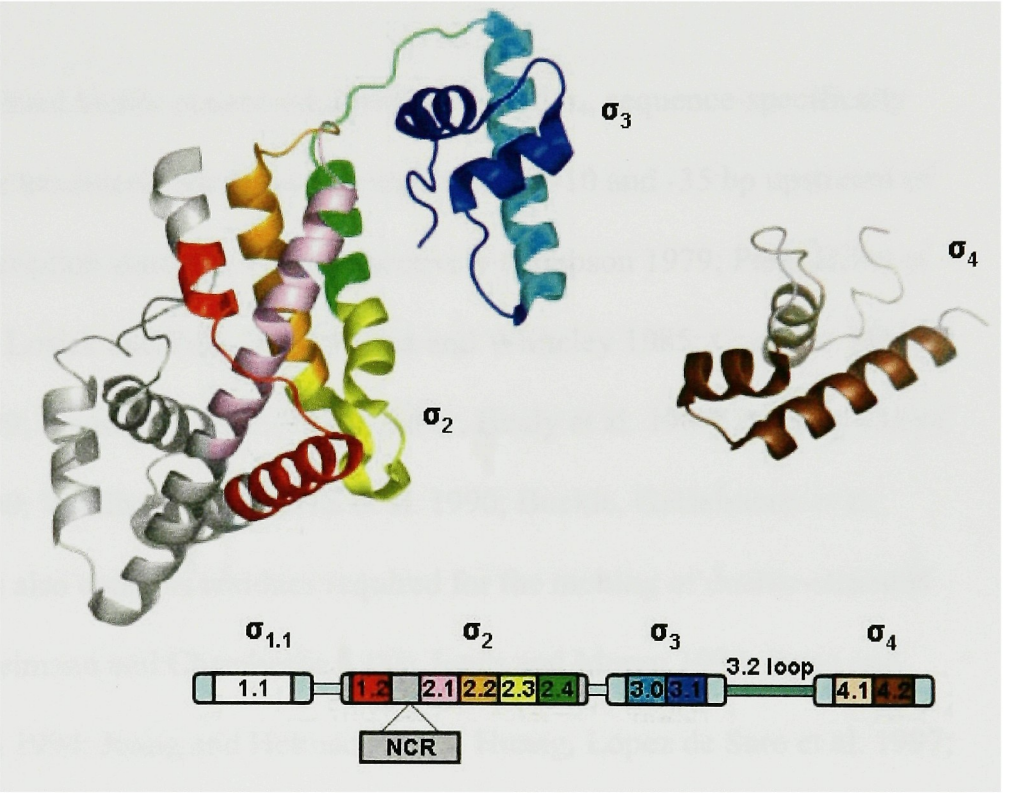


Figure 1.1 Domain organization in Taq σ^A

The crystal structures of the major Taq σ^A domains are shown as a ribbon diagram (Campbell, Muzzin et al. 2002). Conserved regions are color coded as indicated in the diagram on the bottom, which represents the linear structure from N to C terminus, with the thicker segments representing domains, and the thinner segments representing loops. The non-conserved region (NCR) of σ_2 is shown as a grey insertion.

limited proteolysis (Severinova, Severinov et al. 1996), and later confirmed by x-ray crystal structures (Malhotra, Severinova et al. 1996; Campbell, Muzzin et al. 2002; Murakami, Masuda et al. 2002; Vassylyev, Sekine et al. 2002).

The most highly conserved domains, σ_2 and σ_4, sequence-specifically recognize hexameric promoter elements located -10 and -35 bp upstream of the transcription start site (+1), respectively (Simpson 1979; Park, Hillel et al. 1980; Losick and Pero 1981; Hilton and Whiteley 1985; Gardella, Moyle et al. 1989; Siegele, Hu et al. 1989; Zuber, Healy et al. 1989; Daniels, Zuber et al. 1990; Waldburger, Gardella et al. 1990; Buckle, Geiselmann et al. 1991). σ_2 also contains residues required for the melting of double-stranded DNA (Helmann and Chamberlin 1988; Jones and Moran 1992; Rong and Helmann 1994; Juang and Helmann 1995; Huang, Lopez de Saro et al. 1997; Young, Anthony et al. 2001). $\sigma_{1.1}$ is present only in primary σ factors, and has been shown to inhibit promoter binding of free σ^{70} (Dombroski, Walter et al. 1992; Dombroski, Walter et al. 1993; Camarero, Shekhtman et al. 2002). σ_3 is also a DNA binding domain, which recognizes the "extended -10 element" located adjacent to the -10 element in some promoters (Keilty and Rosenberg 1987; Kumar, Malloch et al. 1993). In addition, all sigma

domains interact with the core polymerase, although σ_2 makes the most extensive interactions (Murakami, Masuda et al. 2002).

1.2 σ factors consist of several, flexibly linked domains.

The crystal structures of Taq and Tth σ^A holoenzyme ($E\sigma^A$) have been determined, providing the structure of the entire σ^A molecule bound to RNAP, except for the N-terminal region 1.1, which was either disordered or truncated (Murakami, Masuda et al. 2002; Vassylyev, Sekine et al. 2002). In addition, the structure of the holoenzyme bound to a "fork-junction" DNA fragment, a melted promoter mimic, has been determined to 6.5 Å resolution (Murakami, Masuda et al. 2002). In these structures, σ^A adopts an elongated conformation, with its domains spread along one face of the enzyme, and with the DNA binding surfaces exposed and appropriately spaced for recognition of the promoter -35 and -10 elements. The σ_3-σ_4 linker threads through the polymerase near the active site, and occupies the RNA exit channel, through which the nascent RNA is extruded during elongation (Murakami, Masuda et al. 2002; Vassylyev, Sekine et al. 2002).

Despite extensive attempts, crystals of an isolated, intact σ factor have not yet been obtained. Its large size, high α-helical content, and tendency to

form oligomers at high concentrations also make it a challenging target for NMR. Consequently, the structure of free σ remains unknown. Crystal structures of all the major σ domains (σ_2, σ_3 and σ_4) have been determined (Figure 1.1), both in isolation (Malhotra, Severinova et al. 1996; Campbell, Muzzin et al. 2002; Li, Stevenson et al. 2002), and in the presence of anti-σ factors (Campbell, Masuda et al. 2002; Campbell, Tupy et al. 2003). In addition, the structure of σ_4 bound to promoter DNA has been determined to high resolution (Campbell, Muzzin et al. 2002). Structural superposition of the individual sigma domains determined from different contexts reveals that the domain structures are well conserved, even between the most highly divergent sigma factors. The only known exception is the NMR structure of σ_4 bound to the anti-σ factor AsiA, in which the domain structure of σ_4 is significantly altered in the presence of AsiA (Lambert, Wei et al. 2004) .

Biophysical studies comparing free and polymerase-bound Ec σ^{70} suggest that σ^{70} undergoes a large conformational change upon binding to the polymerase. Inter-domain distance measurements obtained by luminescence resonance energy transfer (LRET) indicate that the distance between σ_2 and σ_4 increases significantly upon binding to the polymerase (Callaci, Heyduk et al. 1999). In addition, solvent exposure measurements for individual residues within the DNA binding surfaces of σ_2 and σ_4 have

demonstrated that residues in the DNA binding surfaces are buried in free σ, but become exposed in the holoenzyme (Callaci, Heyduk et al. 1998; Callaci, Heyduk et al. 1999). The observation that the domain structures of σ bound to RNAP (Murakami, Masuda et al. 2002; Vassylyev, Sekine et al. 2002) compared to those of isolated domains (Malhotra, Severinova et al. 1996; Campbell, Muzzin et al. 2002) are essentially identical, along with the presence of flexible linkers between the domains, suggest that the conformational change is primarily due to a repositioning of the domains, rather than to intra-domain changes.

1.3 Region 1.1 inhibits DNA binding in primary σ factors.

Although σ makes most of the sequence specific promoter contacts in the holoenzyme, free σ lacks sequence specific promoter binding activity (Dombroski, Walter et al. 1992). Studies of N-terminally truncated Ec σ^{70} revealed that removal of the first 50 amino acids (half of region 1.1) allows promoter binding of free σ, and removal of an additional 130 residues (regions 1.1 and 1.2) increases this effect by an order of magnitude (Dombroski, Walter et al. 1993) . In addition, a polypeptide containing the N-terminal 100 amino acids inhibits -35 binding by σ_4 when added *in trans*,

but not -10 binding by σ_2, thus pointing towards a mechanism in which region 1.1 masks the DNA binding surface of σ_4. NMR analysis of a σ^{70} homolog from *Thermotoga maritima*, however, in which conserved region 4.2 was selectively labeled, failed to find evidence for a region 1.1/4.2 interaction, suggesting that the interaction is not direct (Camarero, Shekhtman et al. 2002). Region 1.1 is not required for transcriptional activity in vitro, and is completely absent in the primary σ factor of *Bacteroides fragilis* (Vingadassalom, Kolb et al. 2005). The structure of region 1.1 has never been observed. Although this segment was present in the crystallized Tth holoenzyme, it was disordered and thus invisible in the electron density. In free σ, this segment is highly sensitive to proteolysis.

1.4 The σ_3-σ_4 linker is unfolded in the holoenzyme.

σ_3 and σ_4 are connected by a linker segment of approximately 30 residues. In primary σ factors, this segment contains conserved region 3.2 and has thus been referred to as the $\sigma_{3.2}$-loop. In alternative σ factors, a segment of similar length is typically present, but sequence alignments do not support the existence of region 3.2. For simplicity, this segment will therefore be referred to here as the σ_3-σ_4 linker in both primary and

alternative σ factors. During attempts to crystallize full-length Taq σ^A, this region was proteolytically degraded *in situ*, yielding separate crystals of σ_{2-} σ_3 and of σ_4 (Campbell, Muzzin et al. 2002). Consequently the structure of this segment has not been determined in free σ, but based on its proteolytic sensitivity, it appears to be unstructured.

In the crystal structures of the Taq and Tt σ^A holoenzymes, σ_3 and σ_4 are separated by 45Å (Murakami, Masuda et al. 2002; Vassylyev, Sekine et al. 2002). The σ_3-σ_4 linker segment connecting these is mostly unfolded, and unlike the σ domains which are on the surface of the holoenzyme, the linker segment passes through the interior of the enzyme, extending near the active site, and threading through the RNA exit channel through which the transcript is extruded during elongation.

The positioning of the linker provides the basis for abortive initiation. During initiation of transcription, the holoenzyme repeatedly synthesizes and releases premature transcripts of 2-12 nt. This abortive initiation phase ends when the polymerase escapes the promoter and enters processive elongation, and this transition is usually accompanied by σ release (Murakami, Masuda et al. 2002). Examination of the structure suggests that this phenomenon is due to a steric clash between the σ_3-σ_4 linker and the nascent transcript. Modeling reveals that a transcript longer than 6 nt would clash with the σ_3-σ_4

linker, suggesting that at least this part of σ must dissociate in order to synthesize longer transcripts. This is further supported by experimental evidence. Mutation of certain residues in the N-terminal part of the linker greatly reduces the amount of abortive product relative to the full-length transcript (Sen, Nagai et al. 1998), and a C-terminally truncated mutant lacking both the σ_3-σ_4 linker and σ_4 is transcriptionally active, but produces no detectable abortive products (Murakami, Masuda et al. 2002). The activity of this mutant in *in vitro* transcription assays is significantly lower than the wildtype, but activity comparable to wild type is restored upon the addition of excess initiating dinucleotide, suggesting that σ_3-σ_4 linker contributes to the affinity of the initiating nucleotide (Campbell, Muzzin et al. 2002). In the structure of the Tth holoenzyme a portion of the σ_3-σ_4 linker forms a hairpin, which protrudes into the active site cleft, close to the initiating nucleotide binding site, consistent with this idea (Vassylyev, Sekine et al. 2002).The σ_3-σ_4 linker may also contribute to the affinity of σ to core RNAP, since a deletion mutant for this region of Ec heat shock σ factor σ^{32} is transcriptionally active, but has a decreased affinity for core RNAP (Zhou, Walter et al. 1992).

1.5 Anti-σ factors are diverse in structure and inhibitory mechanism

The first crystal structure of a σ/anti-σ complex to be determined was that of *Bacillus stearothermophilus* σ^F bound to its anti-σ factor, SpoIIAB (Campbell, Masuda et al. 2002). σ^F is required during sporulation and initiates a transcription cascade in the fore-spore by activating the transcription of several other sporulation specific σ factors (Errington 1993). The activity of σ^F is negatively regulated by SpoIIAB (Duncan and Losick 1993). In the crystal structure, SpoIIAB sterically occludes the core binding surface of σ_3, thus sequestering σ^F from core by direct steric competition. SpoIIAB is a serine kinase, which can exist in either ATP or ADP bound form, where the ATP bound form has higher affinity for σ^F. The crystal structure is of the low affinity complex with the ADP bound form of SpoIIAB, and in this structure, σ_3 is the only part of σ^F which is interacting with SpoIIAB. Surprisingly, σ_2 and σ_4 were invisible in the structure, despite being present in the crystals. Genetic and proteolytic studies of σ^F in complex with the ATP bound form of SpoIIAB have suggested that σ_2 and σ_4 are also involved in the high affinity interaction (Decatur and Losick 1996; Campbell and Darst 2000). Thus there may be two distinct inhibitory complexes, as the difference in affinity also implies.

The extracytoplasmic function (ECF) σ factors constitute a large subgroup of the σ^{70} family, marked by their high sequence divergence and small size (reviewed in (Helmann 2002)). Many ECF σ factors have envelope related functions, such as secretion, transport and periplasmic stress response, and their activity is frequently regulated by specific transmembrane anti-sigma factors, which have an intracellular anti-sigma domain and a periplasmic sensor domain. One such pair is σ^E and RseA from Ec. σ^E responds to stress that affects the folding of periplasmic proteins. Its specific inhibitor, RseA is a transmembrane protein situated in the inner membrane. Its soluble N-terminal domain binds to σ^E on the cytoplasmic side, while its C-terminal domain senses periplasmic stress. Such stress triggers the degradation of RseA and subsequent release and activation of σ^E, which activates the transcription of genes, whose products serve to mitigate the effect of protein unfolding in the periplasm. The crystal structure of Ec σ^E bound to the N-terminal anti-σ domain of RseA (RseA-N) has been determined (Campbell, Tupy et al. 2003). σ^E lacks σ_3, and σ_2 connects to σ_4 via a linker segment, which is disordered in the structure. Despite the low degree of sequence conservation, the domain structures are well conserved with those of Taq σ^A. RseA-N has a novel all α-helical fold

and is sandwiched between σ_2 and σ_4, directly blocking the major core RNAP binding surfaces.

A completely different inhibitory strategy has been observed for the 10 kDa phage T4 encoded anti-σ factor, AsiA. This is an anomalous anti-σ factor in that it alters the promoter specificity of σ rather than preventing its association with core RNAP, and unlike most endogenous anti-σ factors, it targets a primary σ factor, Ec σ^{70}. AsiA inhibits transcription from bacterial and early phage promoters, but together with the phage protein MotA, it also stimulates transcription from middle phage promoters (Orsini, Ouhammouch et al. 1993). Although AsiA binds to free σ^{70}, it does not prevent binding to core RNAP. Rather it alters the association with core RNAP by selectively disrupting the interaction between σ_4 and the flap domain of core RNAP (Simeonov, Bieber Urbauer et al. 2003; Gregory, Nickels et al. 2004). This interaction is required to obtain the appropriate distance between σ_2 and σ_4 for -35 recognition (Kuznedelov, Minakhin et al. 2002), and recognition of promoters containing a -35 element is thus selectively disrupted, while the ability to transcribe from extended -10 promoters (which do not depend on the sigma4/-35 element interaction) is retained (Severinova, Severinov et al. 1998). NMR structures have been determined both of AsiA in isolation (Urbauer, Simeonov et al. 2002) and in complex with σ_4 (Lambert, Wei et al.

2004) AsiA is an all α-helical protein and contains a helix-turn-helix DNA binding motif. It is a homodimer in solution, but forms a heterodimer with σ^{70} through the same interface. The interaction induces a conformational change in σ_4, which includes both rearrangement of the helices and partial unfolding (Lambert, Wei et al. 2004).

In summary, anti-σ factors of known structure are completely unrelated, although they may employ similar strategies of inhibition, such as the steric blocking of core binding surfaces employed by both SpoIIAB and RseA. This reflects the diversity of signals that the different anti-sigmas must respond to.

1.6 σ^{28} activates transcription of late flagellar genes upon secretion of its anti-σ factor, FlgM.

The bacterial flagellum allows for rapid, directed cellular translocation. This large organelle is membrane anchored, and extends as a long helical filament on the extra-cellular side. The major structural components are the hook-basal body (HBB), which spans the inner and outer bacterial membranes and forms a hook-shaped structure on the extra-cellular side, and the filament, which extends outward from the hook. The filament is

a rigid assembly of a single protein, flagellin. In addition, motor and switch proteins associated with HBB generate torque and control the direction of rotation, respectively. Finally, junction and capping proteins associate with the cell proximal and distal ends of the filament respectively (Macnab 2003). All of the extracellular subunits are exported via a ~3 nm wide channel (Morgan, Macnab et al. 1993; Mimori, Yamashita et al. 1995; Morgan, Owen et al. 1995), which penetrates the HBB and extends throughout the filament. This secretion system is a type III export apparatus and has homology to the virulence factor secretion systems of many bacterial pathogens.

Flagellar synthesis in St requires approximately 50 genes, which can be hierarchically classified as early, middle and late, where the early and middle genes are σ^{70} dependent, while the late genes require the alternative σ factor, σ^{28} (FliA, also called σ^{D} in gram (+) species) , for transcription (Ohnishi, Kutsukake et al. 1990). The early genes are transcribed in response to environmental cues, and their products, FlhC and FlhD, are transcriptional activators of the middle genes. The middle genes encode all the structural components of the HBB, as well as σ^{28} along with its specific anti-σ factor, FlgM , which binds to σ^{28} with a 1:1 stoichiometry, sequestering it from core RNAP, and preventing late gene expression (Gillen and Hughes 1991; Gillen

and Hughes 1991; Ohnishi, Kutsukake et al. 1992). The late genes are transcribed only when the HBB has been assembled in intact form, and their products include flagellin, chemotactic and motility proteins, cap protein and hook-associated proteins (Kutsukake, Ohya et al. 1990). If any component of the HBB is absent or defective, the late genes are never transcribed. An explanation for this coupling between morphology and transcription was initially proposed in a review by Blair and Dutcher (Blair and Dutcher 1992), who suggested that FlgM may serve as a secretion substrate for the HBB, but only once it has been fully assembled in intact form. This elegant mechanism was subsequently confirmed experimentally to be correct (Hughes, Gillen et al. 1993; Kutsukake 1994).

σ^{28} is clearly a σ^{70}-family member, but lacks both conserved region 1 and N-terminal non-conserved sequence, thus making it much more compact than σ^{70}, while still containing the most highly conserved domains. Homologs of St σ^{28} exist in a broad range of gram(+) and gram(-) bacteria, and most are involved in flagellar synthesis, although a homolog in *Streptomyces coelicolor* functions in the development of aerial hyphae (Chater, Bruton et al. 1989). Transcription of flagellar genes also requires σ^{54} in many species. In *Vibrio cholerae*, for example, the middle genes are σ^{54} dependent, and can be divided into two temporal classes, while the late

genes are σ^{28} dependent (Prouty, Correa et al. 2001). Other species have similar systems in place, but with individual variations (Colland, Rain et al. 2001; Prouty, Correa et al. 2001; Dasgupta, Wolfgang et al. 2003; Hendrixson and DiRita 2003) .

St FlgM is 97 residues long, and has been shown in NMR studies to lack stable structure in isolation, but to gain structure in the C-terminal half upon binding to σ^{28}, while the N-terminal half remains unstructured (Daughdrill, Chadsey et al. 1997). Transient α-helical structure in the C-terminal half has been detected even in the absence of σ^{28} (Daughdrill, Hanely et al. 1998), and an increase in C-terminal structure has also been observed by NMR in living cells and under crowded solution conditions (Dedmon, Patel et al. 2002). The lack of structure in free FlgM has been suggested to aid its passage through the narrow channel of the HBB during export (Daughdrill, Chadsey et al. 1997)

Based on experiments assessing the ability of a set of truncated FlgM mutants to inhibit σ^{28} *in vivo*, the σ^{28} binding domain was proposed to lie between residues 64 and 88 (Iyoda and Kutsukake 1995). Examination of the data, however, reveals that while the C-terminal half is absolutely essential for inhibition, truncations in the N-terminal half significantly weaken the inhibition. Several point mutations in FlgM which attenuate σ^{28}

inhibition have also been isolated, and these range between residues 58 and 62 (Daughdrill, Chadsey et al. 1997). The N-terminal half of FlgM has been shown to be necessary and sufficient for secretion, although no specific secretion signal has been identified (Iyoda and Kutsukake 1995). Together, these observations have painted a picture of FlgM as a modular polypeptide in which the C-terminal half inhibits σ^{28}, while the N-terminal half is required for secretion.

A number of truncated versions of σ^{28} were supplied on a plasmid, and tested for their ability to titrate endogenous FlgM sufficiently for endogenous σ^{28} to become transcriptionally active (Kutsukake, Iyoda et al. 1994). An N-terminally truncated mutant containing σ_4 almost completely titrated FlgM, whereas C-terminal truncations had no effect, suggesting that σ_4 is both necessary and sufficient for interaction with FlgM. A number of σ^{28} point mutants with transcriptional activity in the presence of FlgM have also been isolated (Kutsukake, Iyoda et al. 1994; Chadsey and Hughes 2001), and while most of these are situated in σ_4, several have also been identified in σ_2 and σ_3, leading to the proposal that the interaction is multipartite. Many of these mutants were confirmed *in vitro* to cause a binding defect (Chadsey and Hughes 2001).

Surface plasmon resonance (SPR) measurements have shown that the dissociation rate of σ^{28} from the σ^{28} holoenzyme ($E\sigma^{28}$) increases in the presence of FlgM. This, along with evidence that FlgM associates weakly with the σ^{28} holoenzyme, led to the proposal that FlgM can attack and destabilize the σ^{28} holoenzyme (Chadsey, Karlinsey et al. 1998). Mutants in which this effect was attenuated were identified in both σ^{28} (Chadsey and Hughes 2001) and FlgM (Chadsey, Karlinsey et al. 1998), and were located in region 4 and the C-terminal half of the respective polypeptides. This has led to a model in which an initial interaction between σ^{28} region 4 and the C-terminal half of FlgM occurs while σ^{28} is still bound to core RNAP, and allosterically weakens the interaction between σ^{28} and core RNAP, thus increasing the dissociation rate of the holoenzyme.

The *in vitro* promoter binding properties of free σ^{28} have been investigated in several species, and the results are contradictory. Although σ^{28} lacks the autoinhibitory region 1.1, the promoter binding activity of St σ^{28} is one order of magnitude lower than the analogous construct of σ^{70}, and removal of a short non-conserved N-terminal segment does not restore promoter binding. Thus, there appears to be a mechanism of autoinhibition that does not involve region 1.1 (Dombroski, Walter et al. 1993). In another study of St σ^{28}, promoter binding was detected in the presence, but not in the

absence of RNAP (Chadsey, Karlinsey et al. 1998). In contrast, *Bacillus subtilis* σ^D has been reported to bind promoters and produce hypersensitive sites around the -10 and -35 elements in DNA footprinting experiments. To obtain these results, however, a 50,000-fold excess of σ^D to promoter DNA was used. Given the large deviation from physiological condition, the biological relevance of these results is unclear (Chen and Helmann 1995). The *Aquifex aeolicus* σ^{28} homolog has been reported to bind DNA both in the presence and absence of Ec RNAP, although promoter specificity could not be demonstrated (Studholme and Buck 2000).

While much is known about the flagellar gene regulatory system and the σ^{28}/FlgM interaction, the molecular details of the interaction, and thus the mechanism of inhibition, has not previously been determined. The paucity of structural information and the conformational flexibility of isolated σ factors have also limited our understanding of autoinhibition. Here, the co-crystal structure of the σ^{28}/FlgM complex provides a detailed understanding of the inhibitory mechanism employed by FlgM. In addition, it provides a novel σ conformation, in which the domains are close together, the promoter binding surfaces are buried, and the linker domain is α-helical. Disulfide crosslinking of free and FlgM-bound σ^{28} indicates that this conformation is also well populated in free σ^{28}, thus providing a novel

explanation for autoinhibition and illustrating the magnitude of the conformational change that is likely to accompany binding of σ^{28} to core RNAP.

2. EXPERIMENTAL PROCEDURES

2.1 σ^{28}/FlgM Structure Determination

Cloning and Mutagenesis

The *fliA* (σ^{28}) and *flgM* genes were PCR amplified from Aa genomic DNA (gift of K.O. Stetter, University of Regensburg, Germany) and subcloned into a pET-21a-based (Novagen) expression vector such that the two proteins were coexpressed from a single promoter (Campbell and Darst 2000). In the resulting vector (pKMS5), σ^{28} was expressed carrying an N-terminal, thrombin-cleavable hexahistidine tag. For phasing purposes, a triple methionine σ^{28} mutant (I77M, L122M, L191M) was engineered using the Quikchange Multi Kit (Stratagene, La Jolla, CA). DNA sequencing (DNA Sequencing Facility, The Rockefeller University) was used to confirm the correctness of each construct.

Expression and Purification

The σ^{28}/FlgM complex was expressed and purified according to the protocol described for σ^{E}/RseA-N (Campbell, Tupy et al. 2003) with the following exceptions: BL21-CodonPlus® (DE3)-RIL (Stratagene, La Jolla, CA) cells were used, expression duration was 4 hr, and 0.5% glycerol was

used in the nickel chromatography buffers. The complex containing the σ^{28}-3M mutant was prepared as a SeMet derivative by suppression of methionine biosynthesis (Doublie 1997). The purification protocol was identical to that of the native, wild-type complex. The expected masses of σ^{28} and FlgM in both the native and mutant SeMet derivative complexes were confirmed by MALDI-TOF mass spectrometry.

Crystallization and Structure Determination

Crystallization of the native σ^{28}/FlgM complex was achieved by hanging drop vapor diffusion. Crystals appeared under several different conditions. Two different crystal forms were utilized for structure determination.

Crystal form I: protein solution (1 µl at 65 mg/mL in 20 mM Tris-HCl [pH 8.0], 20 mM NaCl, 2 mM DTT) was mixed with 1 µl of precipitant solution [18% PEG 8000 (w/v), 0.1 M sodium cacodylate (pH 6.5), 0.2 M calcium acetate], followed by 0.3 µl additive solution (30% v/v of isopropanol) and allowed to equilibrate against well solution (0.9 ml precipitant solution and 0.1 ml additive solution) at 4°C. Plate-like, birefringent crystals appeared within a few days and reached full size within 1 to 2 weeks. Trimethyl-lead acetate (1 µl at 2 mM in well solution) was

added to the drop 9 days prior to freezing. Although no ordered lead was detected in the electron density, the diffraction quality improved with this soak. Prior to data collection, the crystals were incubated for a few minutes in cryoprotectant solution [20% glycerol (v/v), 16.2% PEG 8000 (w/v), 80 mM calcium acetate, 25 mM MES (pH 6.5), 3% (v/v) isopropanol], flash frozen in liquid ethane, and stored in liquid nitrogen.

Crystal form II was obtained at the same conditions except that 0.3 μl of 30% (w/v) 6-aminocaproic acid was also added to the drop. Very large hexagonal rods appeared overnight with this additive, and these diffracted to 3.25 Å. Crystals of form II were frozen by gradually replacing the mother liquor with 18% 1,2,6-trihydroxyhexane and 82% well solution, flash freezing in liquid ethane, and storing in liquid nitrogen.

Discounting the N termini, the native σ^{28}/FlgM complex contains only one methionine. Site-directed mutagenesis was therefore used to introduce three additional methionines into σ^{28} by substituting Ile77, Leu122, and Leu191, positions where methionine occurs in σ^{28} orthologs (Figures 3.1 and 3.9). The SeMet-complex containing the triple methionine σ^{28} mutant (σ^{28}-3M) produced isomorphous crystals (both form I and form II) by microseeding with native crystals. The crystallization conditions were the

same, except that 13% PEG 8000 (w/v) was used, the protein concentration was 54 mg/mL, and the trimethyl-lead acetate soak was omitted. The cryoprotectant solutions were the same as for the native crystals, except that 13% PEG 8000 (w/v) (forms I and II) and 30% glycerol (v/v) (form II) were used.

All data were indexed and scaled using HKL2000 and Scalepack (Otwinowski and Minor 1997). The structures were phased with data from the SeMet complex containing the σ^{28}-3M mutant using SOLVE/RESOLVE (Terwilliger 1999; Terwilliger 2000; Terwilliger 2002) . The map quality was excellent for both datasets. For the form I crystals, SOLVE located 17 of a possible 24 selenium sites, and 4 of a possible 6 sites were located for crystal form II, corresponding to all methionines except the N-terminal ones.

A partial model was built into the 2.3 Å resolution form I map, combining parts of models generated using RESOLVE (Terwilliger 2002) and ARP/wARP (Perrakis, Morris et al. 1999). Additional model building was done using O (Jones, Zou et al. 1991). Initial iterative cycles of refinement and model building against the SeMet(peak) amplitudes and SIGMAA-weighted phase combination were performed using CNS (Adams, Pannu et al. 1997). Molecular replacement was then used to position the

model against the native amplitudes, and the refinement was completed. Noncrystallographic restraints were not incorporated during any part of the refinement. PROCHECK (Laskowski, MacArthur et al. 1993) revealed 94.4% of residues in most favored (ϕ, ψ) regions, 5% in additional allowed regions, 0.4% in generously allowed regions, and 0.2% (2 residues) in disallowed regions, and an overall G factor of 0.5.

For the form II crystals, an initial model was manually built using homology models of the major domains as a starting point, created from the Taq σ^A domain structures (PDB ID 1KU2 and 1KU3; (Campbell, Muzzin et al. 2002) using SWISS-MODEL (Schwede, Kopp et al. 2003). One molecule from the refined 2.3 Å structure was then placed in the density by superposition onto the manually built model, and this was improved further by iterative model building and CNS refinement using native amplitudes and SIGMAA-weighted phase combination. PROCHECK (Laskowski, MacArthur et al. 1993) revealed 81.5% of residues in most favored (ϕ, ψ) regions, 17.5% in additional allowed regions, 0.4% in generously allowed regions, and 0.4% (one residue) in disallowed regions, and an overall G factor of 0.2.

Calculation of Relative Solvent Accessibilities

Side chains of the positions in σ^{28} denoted in Figure 4.9, corresponding to the study of (Callaci, Heyduk et al. 1998) Callaci et al., were changed to cysteine *in silico* using O (Jones, Zou et al. 1991). The accessible surface area (A) of the side chain atoms for these residues from the conformation of σ^{28} in the σ^{28}/FlgM complex and in the holoenzyme conformation was calculated using GRASP (Nicholls, Sharp et al. 1991). In the case of σ in the holoenzyme conformation, the presence of core RNAP did not affect any of the surface areas for the selected positions, so differences in solvent-accessible surface areas arose due to the conformational change going from the σ^{28}/FlgM conformation to the holoenzyme conformation. The relative solvent accessibility was calculated as A_H/A_F where A is the solvent accessible surface area of the cysteine side-chain, and the subscript denotes the context such that A_H refers to σ^{28} in the modeled holoenzyme, and A_F to σ^{28} in a hypothetical free state, where its conformation is the same as in the FlgM complex, but where FlgM has been omitted in the calculation. A value close to 1 indicates no change in solvent accessibility in going from the σ^{28}/FlgM conformation to the holoenzyme conformation, while a value greater than one indicates an increase in solvent accessibility in the holoenzyme compared with the σ^{28}/FlgM conformation.

2.2 Disulfide Analysis

Mutagenesis

The program MODIP (Sowdhamini, Srinivasan et al. 1989) provided a list of pairs of residues which if mutated to cysteine are predicted to form disulfides, based on the crystal structure of Aa σ^{28} (pdb ID 1rp3). Four double cysteine mutants of Aa σ^{28} were constructed in a triple methionine mutant I77M/L122M/L191M (σ^{28}-3M) background. σ^{28}-3M was PCR amplified from the coexpression construct described previously (Sorenson, Ray et al. 2004) and ligated using the NdeI and BamHI sites of vector pSKB-2, which is a derivative of pET28 with N-terminal hexahistidine tag, but where thrombin cleavage site has been replaced with a Prescission protease X (PPX) site. The cysteine mutations were introduced by site-directed mutagenesis, giving the quintuple mutants: σ^{28}-3M/R31C/P202C (31/202), σ^{28}-3M/R31C/E211C (31/211), σ^{28}-3M L133C/S218C (133/218) and σ^{28}-3M/ Y140C/L225C (140/225). The correct DNA sequences were confirmed by sequencing.

Expression and Purification

σ^{28}-3M and each of the four disulfide mutants were expressed in BL21-CodonPlus® (DE3)-RIL (Stratagene, La Jolla, CA) except for 31/211

which was expressed in 4L Rosetta™(DE3)pLysS (Novagen). The cells were grown in LB with shaking at 37°C. At OD_{600} of 0.4-0.6 expression was induced by adding 1 mM IPTG and the temperature was lowered to 19°C. The cells were harvested by centrifugation after 14-16 hrs. Aa FlgM in pET-21a (without any tag) was expressed separately in the same way. The pellets were resuspended in 20 mM TrisHCl pH 8.0, 0.5 M NaCl and a protease inhibitor cocktail containing 174 μg/mL PMSF, 312 μg/mL Benzamidine, 5 μg/mL Chymostatin, 5 μg/mL Leupeptin, 1 μg/mL Pepstatin and 10 μg/mL Aprotinin. The cells were lysed with a French Press, and the soluble fraction removed by centrifugation. Half the soluble lysate of each mutant was mixed with soluble FlgM lysate (corresponding to 2L culture) while the other half was kept separate. The lysates were then incubated at 65°C for 45 min to precipitate heat sensitive Ec proteins. Precipitated contaminants were removed by centrifugation, and the soluble protein was purified using nickel chromatography. For the FlgM complexes, this step removes excess FlgM, yielding a 1:1 σ^{28}/FlgM complex. The nickel eluates were further purified on a Superdex 75 gel filtration column in 20 mM TrisHCl pH 8.0, 0.5 M NaCl. The pure protein was concentrated to between 8 and 70 mg/mL in the same buffer. No reducing agent was added at any time during the purification, and disulfides formed during purification without the aid of a catalyst.

CNBr cleavage and MALDI-TOF mass spectrometry

CNBr crystals were dissolved in 20 μL of 1 mg/mL protein in 70% trifluoroacetic acid (TFA), and then incubated for 19 hrs in the dark at room temperature. The samples were subsequently dried using a SpeedVac and redissolved in 20μL water. For matrix assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF-MS), the samples were first diluted to 1 mg/mL in 67% Acetonitrile and 0.1% TFA, and then diluted to 0.2 mg/mL in the same solvent saturated with alpha-cyano-4-hydroxycinnamic acid (4-HCCA) and immediately spotted on a MALDI sample plate as described (Cadene and Chait 2000). For the reduced samples, 1 μL 1*M* β-mercaptoethanol (BME) was added per 10 μL 1mg/mL protein, and ammonium hydroxide was added until the pH ≥8. The samples were incubated for ~ 1 hr at 25°C and prepared for MALDI-TOF-MS as described above.

SDS-PAGE analysis

Concentrated protein stock was diluted in 20 m*M* TrisHCl pH 8.0, 0.5 *M* NaCl to 0.25-0.5 mg/mL. The protein was then mixed 4:1 with 5X SDS sample buffer (with or without BME, as noted), heated to ~100°C for 1 min prior to loading. NuPage 10% Bis-tris precast gels (Invitrogen) were used

with MOPS running buffer (50 m*M* MOPS, 50 m*M* Tris Base, 0.1% SDS, 1m*M* EDTA, pH 7.7). The gels were run at 200V for 1-2 hrs and stained with Coomassie brilliant blue.

Glutathione catalyzed disulfide equilibrium measurements

The procedure was essentially as described (Dorigo, Schalch et al. 2004). Concentrated protein was diluted to 1mg/mL in 20 m*M* Tris pH 8, 0.5 *M* NaCl, 100 mM DTT, and incubated at 37°C for 90 min. 50 uL of reduced protein was then dialyzed against >1L 10 m*M* Tris pH 7.5, 0.5 *M* NaCl, 0.1 m*M* EDTA for 4 hrs with a buffer change after 2 hrs at 4°C. 10 µL dialyzed protein was removed and alkylated by adding 2.5 µL 50 m*M* iodoacetamide (IAM) and incubated for 1 hr in the dark at room temperature. 20 µL of the remaining protein was mixed with 5 µL of 250 m*M* TrisHCl pH 9.0, 0.5 *M* NaCl, 2.5 m*M* reduced glutathione and 2.5 m*M* oxidized glutathione and incubated for 16 hrs at room temperature, and then alkylated with IAM as described above.

3. CRYSTAL STRUCTURE OF THE σ^{28}/ FLGM COMPLEX

3.1 *Aquifex aeolicus* σ^{28}/ FlgM were chosen based on similarity to their St homologs and their crystallizability.

Although the most extensively studied σ^{28}/FlgM complex is that of *Salmonella typhimurium* (St), σ^{28}/FlgM from the hyperthermophilic eubacterium, *Aquifex aeolicus* (Aa), was the focus of this structural study for three reasons. First, we were aware that crystallization of the St complex had already been attempted unsuccessfully by a number of groups. Second, proteins from thermophilic species frequently show greater propensity for crystallization than their mesophilic counterparts, at least anecdotally. Notable examples include bacterial RNAPs (multiple structures of thermophilic Taq and Tt RNAPs are available (Zhang, Campbell et al. 1999; Murakami, Masuda et al. 2002; Murakami, Masuda et al. 2002; Vassylyev, Sekine et al. 2002), while the prototype *Escherichia coli* (Ec) RNAP has resisted extensive crystallization attempts) and ribosomes (the structures of the small, 30S subunit, and the intact ribosome, come from Tt (Wimberly, Brodersen et al. 2000; Yusupov, Yusupova et al. 2001). Finally, recombinant Aa σ^{28} had already been characterized, and its ability to restore motility in an

Figure 3.1 Amino acid sequence alignment for σ^{28}.

Shown is a sequence alignment of flagellar σ's. The sequences are presented in one-letter amino acid code and are identified by the species in the rightmost column. Numbers at the beginning of each line indicate amino acid positions relative to the start of each mature protein sequence. Numbers at the top indicate the amino acid positions in Aa σ^{28}. Amino acid identity in >50% of the sequences is indicated by a red background, amino acid similarity by a black background. Groups of residues considered similar are: ST, RK, DE, NQ, FYW, and ILVM. Gaps are indicated by dashes. Helices in the Aa σ^{28} structure (see Chapter 4) are indicated above the sequences as rectangles, loops as a solid line. The histogram at the top represents the level of sequence identity at each position. Sequence identity of 100% is represented by a tall red bar, less than 20% is represented by a small blue bar, intermediate levels are represented by orange, light green, and light blue bars. The σ conserved regions 2.1-4.2 (Lonetto, Gribskov et al. 1992) are denoted by the colored bars above the sequences, and labeled at the top. Residues mutated to methionine in σ^{28}-3M are marked by asterisks.

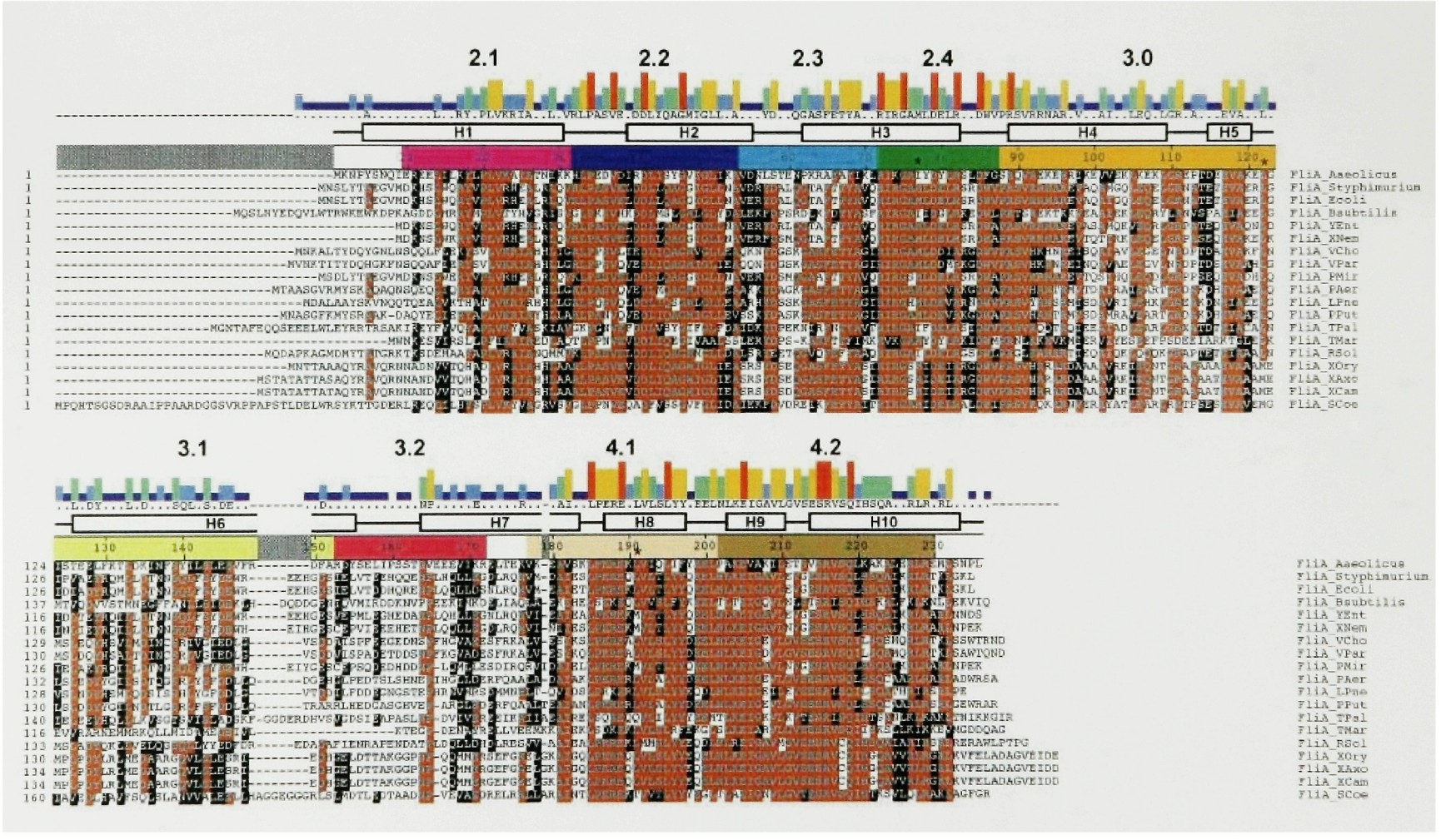
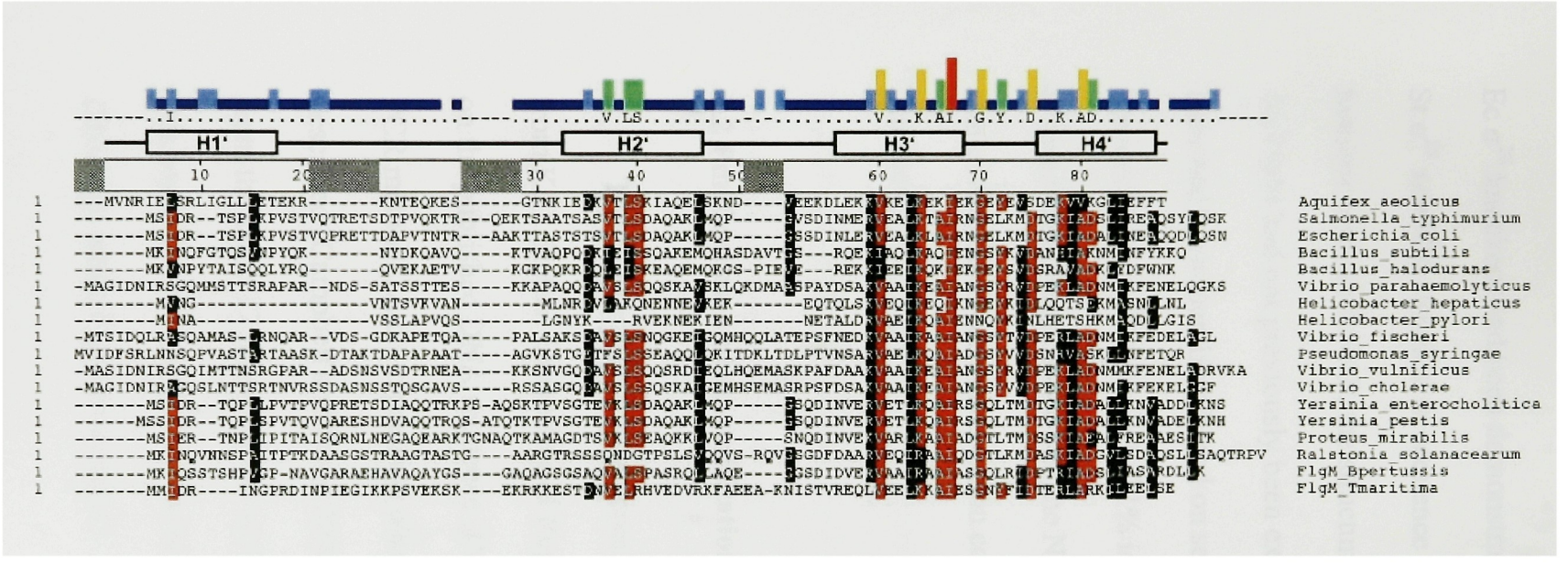

2.1
2.2
2.3
2.4
3.0
H1
H2
H3
H4
H5
3.1
3.2
4.1
4.2
H6
H7
H8
H9
H10
FliA_Aaeolicus
FliA_Styphimurium
FliA_Ecoli
FliA_Bsubtilis
FliA_YEnt
FliA_XNem
FliA_VCho
FliA_VPar
FliA_PMir
FliA_PAer
FliA_LPne
FliA_PPut
FliA_TPal
FliA_TMar
FliA_RSol
FliA_XOry
FliA_XAxo
FliA_XCam
FliA_SCoe

Figure 3.2 Amino acid sequence alignment for FlgM.

Shown is a sequence alignment of FlgM homologs from diverse bacteria. The sequences are presented in one-letter amino acid code and are identified by the species in the rightmost column. Numbers at the top indicate the amino acid positions in Aa FlgM. Amino acid identity in >50% of the sequences is indicated by a red background, amino acid similarity by a black background. Groups of residues considered similar are: ST, RK, DE, NQ, FYW, and ILVM. Gaps are indicated by dashes. Helices in the Aa FlgM structure are indicated above the sequences as rectangles, loops as a solid line. The histogram at the top represents the level of sequence identity at each position. Sequence identity of 100% is represented by a tall red bar, less than 20% is represented by a small blue bar, intermediate levels are represented by orange, light green, and light blue bars.


H1'
H2'
H3'
H4'
Aquifex_aeolicus
Salmonella_typhimurium
Escherichia_coli
Bacillus_subtilis
Bacillus_halodurans
Vibrio_parahaemolyticus
Helicobacter_hepaticus
Helicobacter_pylori
Vibrio_fischeri
Pseudomonas_syringae
Vibrio_vulnificus
Vibrio_cholerae
Yersinia_enterocolitica
Yersinia_pestis
Proteus_mirabilis
Ralstonia_solanacearum
FlgM_Bpertussis
FlgM_Tmaritima

Ec σ^{28} knockout had been demonstrated (Studholme and Buck 2000). Aa and St σ^{28} are 34% identical in sequence (Figure 3.1), and since Aa σ^{28} can function with Ec RNAP, the structure of σ^{28} is also likely to be conserved. Aa FlgM had not previously been expressed or characterized, although the gene had been identified based on sequence homology. Aa FlgM is less well conserved than σ^{28}, with only 21% identity to St FlgM. Sequence alignment (Figure 3.2) reveals that while the N-terminal region of FlgM is poorly conserved, the C-terminal portion contains a stretch of well conserved residues.

3.2 Purification and crystallization of the σ^{28}/FlgM complex.

The *fliA* (σ^{28}) and *flgM* genes were cloned from genomic DNA (a gift from K.O. Stetter, University of Regensberg, Germany) using primers based on the published DNA sequence (Deckert, Warren et al. 1998) and standard PCR methods. A co-expression plasmid was generated as previously described (Campbell and Darst 2000), in which both genes are expressed from the same T7 promoter but each has its own ribosomal binding site. In this construct, σ^{28} has an N-terminal, thrombin-cleavable hexa-histidine tag (His-tag), while FlgM is untagged. The complex could be overexpressed in

Ec at high levels, but as noted previously for Aa σ^{28} alone (Studholme and Buck 2000), satisfactory expression required extra copies of tRNA genes which are rare in Ec, but common in AT-rich genomes such as that of Aa. These were provided in the BL21-CodonPlus® (DE3)-RIL strain (Stratagene, La Jolla, CA). The complex was almost exclusively in the soluble fraction, and could be purified in three major steps: heat treatment, nickel-chelating chromatography, and gel filtration. The clarified lysate was heated to 65 °C, leading to the selective precipitation of most Ec proteins. The complex was then co-purified using nickel-chelating chromatography. FlgM was retained in approximately stoichiometric amounts, and since it does not alone bind to the nickel column, this confirmed the formation of a stable complex. The His-tag was subsequently cleaved from the N-terminus of σ^{28} using thrombin, and subtracted along with uncleaved protein on a second nickel-chelating column. The complex was ammonium sulfate precipitated, and redissolved in a small volume in order to concentrate it. Aggregates and excess σ^{28} were removed by gel filtration (Figure 3.3), and the elution volume indicated a 1:1 stoichiometry. Finally, the purified protein was concentrated by centrifugal filtration to 30 mg/ml for crystallization screening.

Initial crystal screening of 96 conditions (Hampton Research, Aliso Viejo, CA) was performed at 22°C and 4°C, 15 mg/ml and 30 mg/ml protein, and with and without the His-tag on σ^{28}. Hanging drop vapor diffusion was used, in which the protein solution was mixed 1:1 (usually 1 µl of each) with a precipitant solution on a siliconized cover slip. The cover slip was then inverted and used to seal a grease lined reservoir containing 1 ml of the precipitant solution.

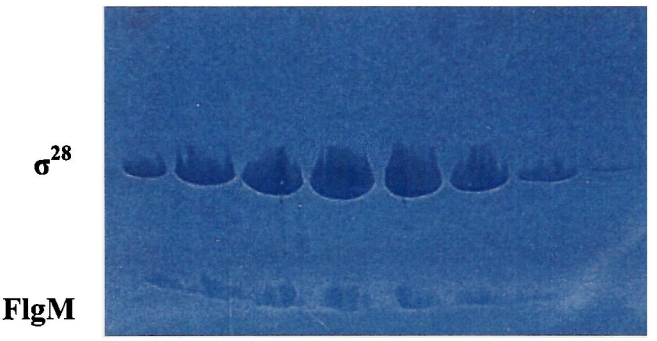


Figure 3.3 The purified σ^{28}/FlgM complex

Superdex 75 peak fractions of the σ^{28}/FlgM complex visualized on an SDS-PAGE phast gel (Amersham).

The hanging drop is not in contact with the reservoir solution, but in the sealed compartment, water is slowly transferred from the drop to the reservoir via vapor diffusion (osmosis), until the system reaches osmotic

equilibrium, at which the precipitant concentration in the drop is close to that in the well. This method allows for the screening of a range of precipitant concentrations in each drop, over the course of a few weeks.

In the initial screen, virtually all drops were still clear after two weeks, indicating that a higher protein concentration was required to reach saturation. The screen was repeated with 65 mg/ml protein. At this concentration, the protein precipitated or formed a phase separation under many conditions, particularly at 4°C. Initial crystals of the complex without a His-tag appeared after 13 days at 4°C in the presence of 18% PEG 8000, 0.1 M sodium cacodylate pH 6.5 and 0.2M calcium acetate. These crystals were highly reproducible, but diffracted only very weakly when tested at room temperature using a home x-ray-source. They grew out of a protein skin with an uneven, rod- shaped appearance, and only very weak birefringence (Figure 3.4).

The initial crystallization conditions were optimized by screening additives (Hampton Research, Aliso Viejo, CA). These screens yielded several new crystal forms, most notably birefringent, plate-like crystals which appeared with 3% isopropanol (IPA) (Crystal Form I) (Figure 3.5) and very large hexagonal crystals, which grew in the presence of 6-

aminocaproic acid (Crystal Form II) (Figure 3.6). Phase separation was present prior to crystallization for all crystal forms.

Figure 3.4 Initial crystals of σ^{28}/FlgM.

Crystal Form I

Crystal form I belongs to space group $P2_1$ with four complexes per asymmetric unit (ASU). The ASU is the smallest unit from which the entire crystal can be generated by symmetry operations (Blow 2002). There are no symmetry constraints within the asymmetric unit, and each complex within the ASU can thus be regarded as an independent structure. The crystals of form I were highly reproducible, but their formation was strongly dependent on both the IPA concentration and the temperature. Increasing the IPA to 4%

led to too much nucleation, and extremely thin crystals, and beyond this concentration denaturation of the protein and formation of a brown precipitate occurred.

A slight decrease in the IPA concentration led to the appearance of the original crystal form. The nucleation of these crystals was also very

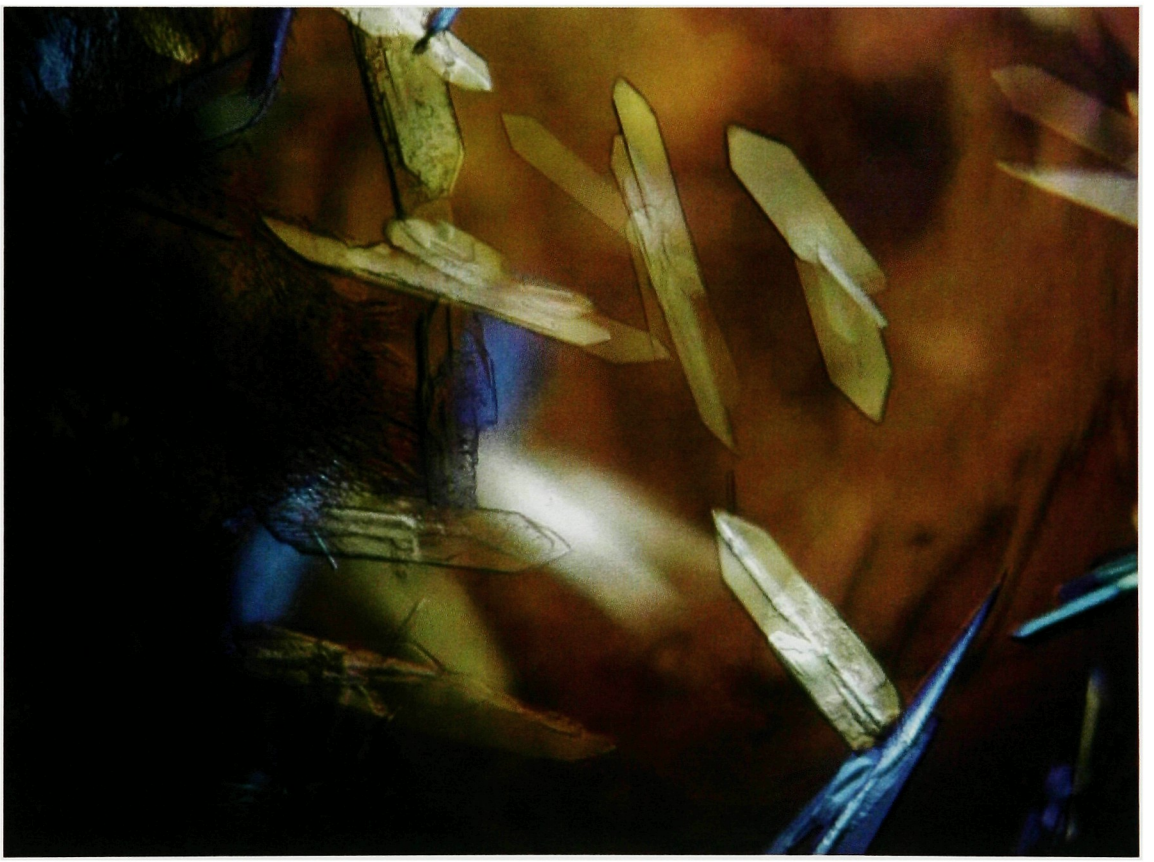


Figure 3.5. Crystal Form I

These crystals were obtained by adding isopropanol to the initial condition. This drastically improved both the appearance and the diffraction.

temperature sensitive: when 24 drops were set up on ice, (rather than inside the 4°C cold-room), the first 6 yielded fewer and larger crystals than normally obtained, while the remaining drops had no crystals at all. This could be explained by a temperature dependent decrease in the nucleation rate.

The form I crystals were so robust that they were difficult to shatter for seeding experiments, and they were stable in a precipitant solution containing 20% glycerol, thus avoiding extensive screening for a cryoprotectant. An initial diffraction pattern was recorded of a frozen crystal using synchrotron radiation. The diffraction was strong but anisotropic with reflections extending to 1.8 Å in one direction, and to about 2.3 Å in the other. The diffraction pattern of most crystals tested was also double (split spots), indicating an internal crystal defect. Although many crystals were obviously double by visual inspection, many which appeared single still gave rise to double diffraction. Because of this problem, about 5-10 crystals had to be tested for each crystal suitable for data collection.

Crystal Form II

The type II crystals (Figure 3.6) belong to space group $P3_221$ with a single complex in the ASU, and a solvent content of about 75%. These

crystals were not reproducible using the original protocol, but this problem could be overcome by adding 3% IPA prior to the addition of 6-aminocaprioc acid. Interestingly, the order of addition was crucial: adding the IPA before the 6-aminocaproic acid gave crystals in almost every

Figure 3.6. Crystal Form II

This large crystal grew upon the addition of 6-aminocaproic acid to the initial condition.

attempt, whereas the reverse order never yielded crystals. In all cases, the additives were layered onto the drop and mixed by diffusion rather than pipetting, as pipet mixing completely prevented crystallization. IPA could be replaced with the structurally similar propylene glycol, although unfortunately not at concentrations where it would be useful as a cryoprotectant.

The greatest challenge presented by crystal form II was cryoprotection. These crystals were highly sensitive even to minor changes in the mother liquor, and were damaged by precipitant solutions containing every standard cryoprotectant tested. The damage occurred regardless whether the cryoprotectant was introduced by slow serial transfers to solutions containing increasing amounts of cryoprotectant or if they were quickly "dunked" into a solution containing the final concentration of cryoprotectant. Further, more extensive screening for cryoprotectants eventually showed that the crystals could tolerate a precipitant solution containing 18% trihydroxyhexane. Even with this cryoprotectant, some damage occurred with rapid introduction. A successful procedure for introducing this cryoprotectant was eventually developed by Soumya Ray (Lansbury lab, Harvard University), who found that slow exchange of the

mother liquor of the original drop for a reservoir solution containing 18% trihydroxyhexane prior to flash-freezing in liquid ethane gave only minimal damage. Over the course of ~1 hr. a fraction (~0.2 μL) of the drop was removed by pipetting and an equal volume of cryoprotectant solution was added at 10 min. intervals. Flash-frozen crystals prepared in this way were of sufficient quality for data collection and structure solution.

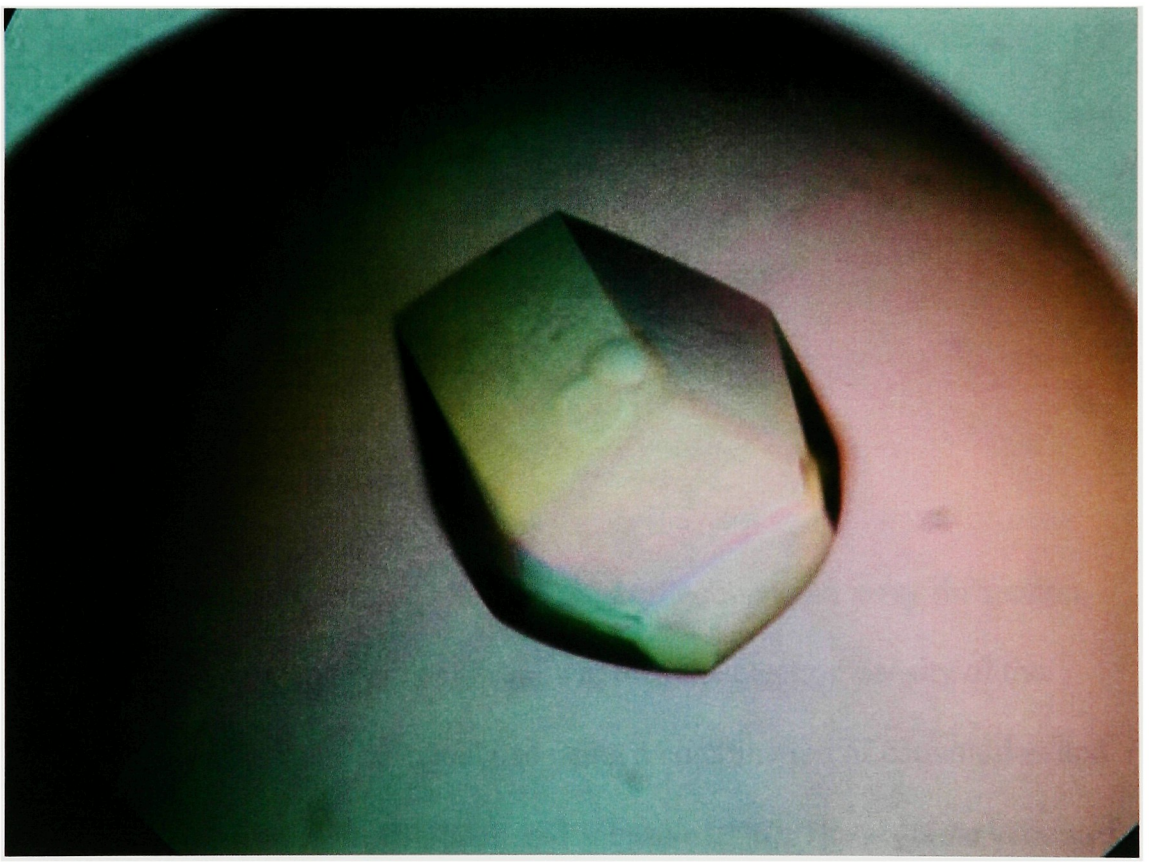


Figure 3.7 Crystal Form III

This crystal only diffracted to 4Å, despite its promising appearance.

Crystal Form III

A third crystal form was obtained from the initial screen (30 mg/mL protein) after about 6 months at 4°C (Figure 3.7). These large, single, hexagonal crystals appeared at two different protein concentrations in the presence of 2.0 *M* Ammonium sulfate, 0.1*M* Na Citrate pH 5.6 and 0.2 *M* K/Na tartrate. Despite their promising appearance, these diffracted only to 4Å. Interestingly, indexing and scaling revealed that these were near isomorphous to crystal form II, even though the external appearance, growth conditions, sensitivity and diffraction limit were all different. These crystals were also significantly less sensitive than those of form II and could be cryoprotected with either glycerol or sodium citrate.

3.3 Phase determination

If the amplitude and phase of each reflection in an x-ray diffraction pattern of a molecular crystal is known, then the electron density of the diffracting object can be calculated using a combination of convolution and Fourier transformation (Blundell and Johnson 1976). The absolute values of the amplitudes are obtained from the intensities of the reflections in the native crystal diffraction pattern. All information about the relative phases of

these reflections, however, is lost in this experiment, and has to be obtained by indirect means.

Currently, the most common method for obtaining experimental phase information is by introducing an electron-rich element (heavy atom) into the protein crystal and utilizing anomalous scattering due to x-ray absorption by this element in order to derive phases for the protein scattering. Under non-absorbing conditions, scattered reflections occur in symmetry related pairs with equal amplitudes but phases of opposite signs, so called Friedel pairs (Blow 2002). If an element in the crystal absorbs x-rays at the particular wavelength used in the experiment however, the absorption causes the amplitudes of the Friedel pairs to differ. Because diffraction is additive, the anomalous scattering from the heavy atoms can be isolated and used to determine the location of these heavy atoms (substructure) in the crystal. Once these sites have been determined, their calculated diffraction can be used as a reference in order to derive phases for the rest of the diffracting object (Hendrickson and Ogata 1997). Since anomalous diffraction is strongly dependent on the wavelength, data collection at several wavelengths and utilization of the dispersive differences between different wavelengths further aids in accurate phase determination. This method, in which data is usually collected at least three wavelengths, is called

multiwavelength anomalous dispersion (MAD) (Hendrickson and Ogata 1997). A significant benefit of this method is that all data required for the structure solution can be obtained from a single crystal, thus avoiding problems with non-isomorphism.

In practice, a heavy atom derivative for MAD phasing is usually obtained by either making a selenomethionine (SeMet) derivative of the native protein where methionine is replaced by SeMet during expression, or by soaking the heavy atom into the crystal. The advantage of making a SeMet derivative is that selenium is certain to be introduced into the protein at specific sites. Phasing using heavy atoms introduced by soaking, by contrast, requires that the introduced element preferentially binds to specific sites in the protein with high occupancy.

Heavy atom soaking of the Aa σ^{28}/FlgM crystals did not yield a useful derivative.

The σ^{28}/FlgM complex contains only one internal methionine in a total of 324 residues. Since this is insufficient for SeMet MAD phasing, heavy atom soaking was initially attempted to obtain a derivative of the crystal form I. Despite testing a large number of heavy atoms, no useful derivative was identified. There were several complicating factors. First, σ^{28}

and FlgM both lack cysteines, which are some of the most common ligands for heavy atoms, such as mercury and gold. Second, the split crystal defect made many of the soaked crystals unsuitable for data collection. To avoid this problem, small fragments were cut from larger crystals and used for soaking. Although these were less likely to give rise to double diffraction, they also diffracted with low intensity, thus preventing detection of weak anomalous signal. Third, the tendency of these crystals to decay upon prolonged x-ray exposure prevented enough images to be collected to obtain sufficient redundancy to detect anomalous differences, particularly since the low symmetry space group ($P2_1$) requires a large number of images to be collected. A full data set was collected on a crystal soaked in trimethyl lead acetate, and although no ordered lead was present, this crystal gave single diffraction of high quality to 2.3 Å (Figure 3.8).

Methionines introduced by mutagenesis allowed for phase determination by SeMet MAD.

Since the heavy atom soaks failed to yield a useful derivative, an alternative approach was attempted, in which methionines were introduced into σ^{28} by mutagenesis in order to allow for SeMet derivatization (Figure 3.9). Sites for Met substitution were chosen at positions where methionine occurs in an

alignment of σ^{28} from different species (residues marked by asterisks in Figure 3.1). Sequence alignment with Taq σ^{A} and analysis of the crystal structures of its domains (Campbell, Muzzin et al. 2002) also provided clues to which of the sites identified by sequence alignment would likely have the least thermal disorder and thus give the strongest anomalous diffraction (based on the atomic B factors). A triple methionine mutant, σ^{28}/ I77M/

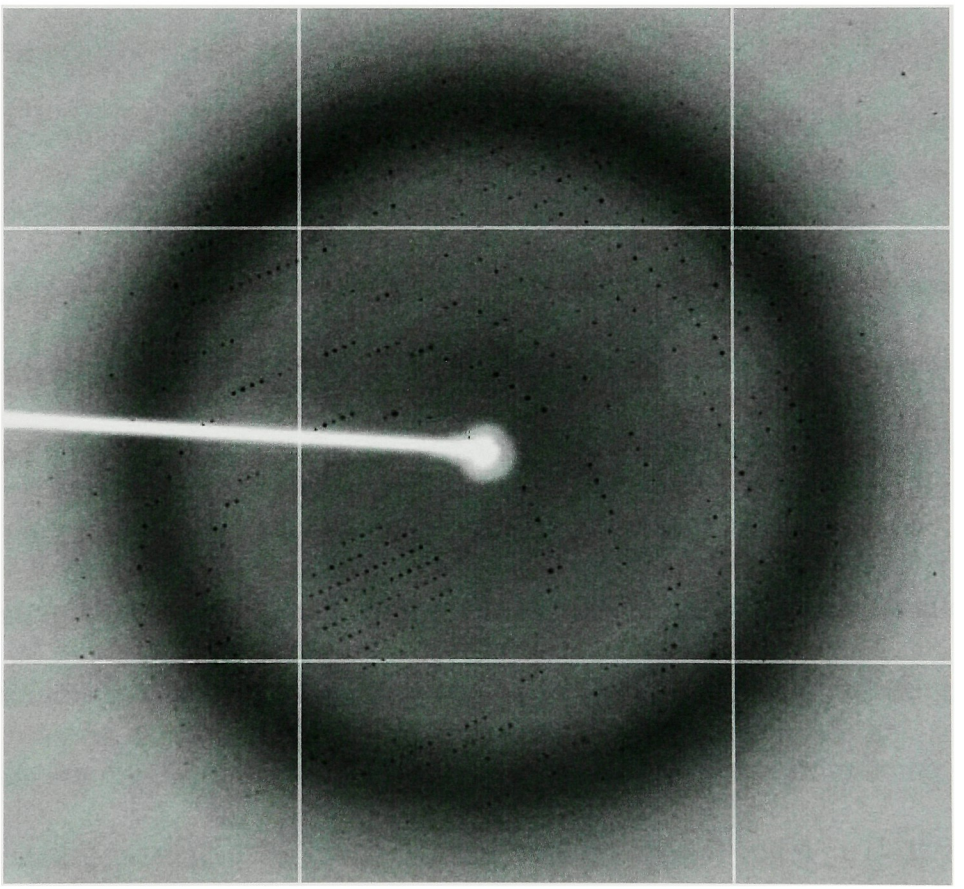


Figure 3.8 Diffraction Pattern from Crystal Form I

L122M/ L191M (σ^{28}-3M) was thus engineered, and a SeMet derivative of this mutant was expressed and purified. Crystals of this derivative did not nucleate spontaneously, but could be obtained by microseeding with native crystals. In this method, a crystal is first crushed in its mother liquor and large fragments are removed by centrifugation. The supernatant contains microscopic nuclei, which can be added to fresh crystallization drops. In this case, the SeMet crystals obtained by microseeding were isomorphous to the native crystals for all three crystal forms.

SeMet MAD data at three wavelengths as well as native data were collected using synchrotron radiation for all three crystal forms at synchrotron beam lines 19ID (Advanced photon source, Argonne) and X25 (National synchrotron light source, Brookhaven) (Table 3.1). While the type II and type III crystals had sufficient anomalous signal, the type I crystals did not. The problem was not likely to be due to a lack of order of the selenium atoms in the protein, since they gave rise to sufficient anomalous signal in the other crystal form. Data analysis did not indicate twinning, which is a crystal defect that can prevent anomalous signal from being detected. Comparison with the data set from the type II and type III crystals did reveal one significant difference: the signal/noise of the diffraction data was significantly greater for these crystal forms. Although the larger crystals

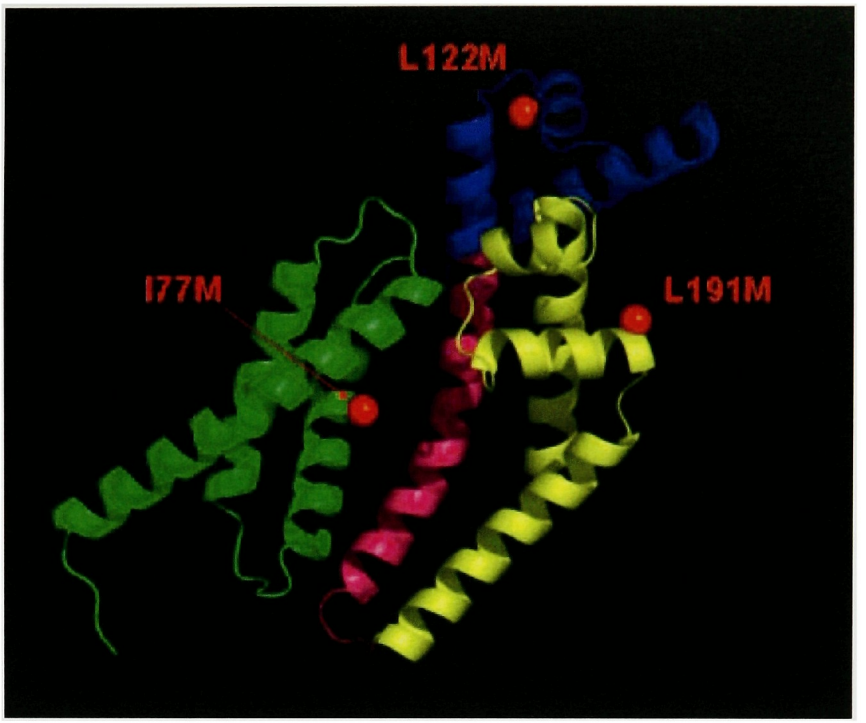


Figure 3.9 Location of introduced methionines.

The locations of the methionines introduced into σ^{28} for phasing are shown in the crystal structure (Chapter 4, (Sorenson, Ray et al. 2004)). The structural elements of σ^{28} are color coded as follows: σ_2, green; σ_3, blue; σ_3- σ_4 linker, pink; σ_4, yellow. The sites of introduced methionines are marked by red spheres.

diffracted to lower resolution due to greater internal thermal motion, the intensity of the diffraction, which is proportional to the number of unit cells contained in the crystal, was much greater due to their large size. The type I crystals used for the SeMet data collection, by contrast, were small

fragments of crystals, and the intensity of the diffraction, and thus the signal to noise was significantly smaller. Since the anomalous difference has to be greater than the noise in order to be detectable, this was presumably because the anomalous signal was weaker than the noise for the small crystals. In order to overcome this, optimization of different precipitant and seed concentrations eventually produced large, single SeMet derivative crystals of Type I. These took approximately one month to reach full size. A single crystal was used for data collection at three wavelengths, and the resulting data sets had a resolution limit of 2.3 Å and had sufficient anomalous signal to allow for structure solution.

3.4 Structure solution, model building and refinement.

Sructure of σ^{28}/FlgM to 3.25 Å

Since crystal form I did initially not yield sufficient anomalous signal for structure solution, the structure was first solved from crystal form II to a resolution limit of 3.25 Å. Four selenium sites were located and initial phases were calculated with SOLVE. The resulting map was correct by visual inspection, with clear α-helices and separate protein and solvent regions. The phases were improved further using RESOLVE, which uses

algorithms that reward those phases that result in clearly defined solvent and protein regions, and protein-like features (Terwilliger 2001). The density-modified map was excellent, with continuous density and low background. Homology models of the major domains, σ_2, σ_3 and σ_4, produced based on the domain structures of Taq σ^A (Campbell, Muzzin et al. 2002) using SWISSMODEL (Schwede, Kopp et al. 2003), could be docked into the density, and used as a starting point for building. The selenium sites were also useful in this process. After placing the domains into density, the remaining density could be assigned to FlgM and the σ_3-σ_4 linker. FlgM could be seen wrapping around σ^{28}, and the σ_3-σ_4 linker segment, surprisingly, was α-helical, forming an extension of the C-terminal helix of σ_3 and continuing as a short loop connecting with σ_4.

FlgM could be distinguished as four α-helices connected by loops. The density for the C-terminal helix was excellent, and the C-terminus could easily be identified by the two adjacent phenylalanine residues adjacent to the C-terminal threonine. Building from this C-terminal helix towards the N-terminus, a helix with poorly defined density formed the other part of the C-terminal half. Surprisingly, given that the N-terminal half of St FlgM has been shown to be disordered in NMR experiments, the remaining helices constituted the N-terminal half of FlgM. Individual residues in these helices

could be assigned based on the primary sequence, side-chain density and secondary structure prediction data. Building of the N-terminal helices revealed that FlgM was fully ordered and interacting with σ_2 with the exception of a loop of ~10 residues connecting these helices.

Once the structure was solved at higher resolution from crystal form I (see next section), a complex from the refined model of this structure was superimposed onto the existing 3.25 Å model. The superimposed model was then adjusted to better fit the density in O, and refined using CNS. Refinement serves to maximize the agreement between the model and the experimental data, as well as to remove geometrical errors, such as steric clashes between atoms and improper bond angle and lengths. The final model had an R_{cryst}/R_{free} of 24.1/27.1%.

Structure of σ28/FlgM to 2.3 Å

The structure of σ^{28}/FlgM from crystal form I was solved to 2.3 Å using SOLVE (Terwilliger; Terwilliger and Berendzen 1999). RESOLVE (Terwilliger 2001; Terwilliger 2002; Terwilliger 2003) was subsequently used for both density modification and automatic model building, and the resulting map was of excellent quality. The asymmetric unit contained four complexes, which all looked essentially like that of the 3.25 resolution

Table 3.1 Crystallographic Analysis

CRYSTAL	Form I (Native)	Form I (SeMet)		
$P2_1/a$ = 76.4 Å, b = 119.7 Å, c = 100.1 Å, β = 107.0°				
X25, NSLS, Brookhaven National Laboratory				
λ (Å)	0.94900 (native)	0.97911 (peak)	0.94900 (native)	0.97911 (peak)
Resolution (Å)	30-2.3 (2.38-2.30)	30-2.3 (2.38-2.30)	30-2.3 (2.38-2.30)	30-2.3 (2.38-2.30)
No. of Refls. (Total/Unique)	264,914/ 76,085	282,008/ 145,667	264,914/ 76,085	282,008/ 145,667
Completeness (%)	99.9 (99.9)	96.1 (91.3)	99.9 (99.9)	96.1 (91.3)
I/σ(I)	16.0 (2.7)	14.6 (1.7)	16.0 (2.7)	14.6 (1.7)
R_{sym} (%)[a]	6.7(46.4)	5.5 (45.1)	6.7(46.4)	5.5 (45.1)
Figure of Merit		0.40		
Refinement Resolution (Å)	30-2.3	30.2.3	30-2.3	30.2.3
No. of Residues	1,211		1,211	
No. of Waters	73		73	
R_{crystt}/R_{free}[b]	24.2/26.2		24.2/26.2	
Rmsd bond lengths	0.007		0.007	
Rmsd bond angles	1.1		1.1	

CRYSTAL	Form II (Native)	Form II (SeMet)		
$P3_221/a$ = 118.1 Å, c = 87.4 Å				
19ID, APS, Argonne				
λ (Å)	0.96486 (native)	0.97969 (peak)	0.97989 (inflection)	0.96486 (remote)
Resolution (Å)	30-3.25 (3.37-3.25)	30-3.5 (3.62-3.50)	30-3.5 (3.62-3.50)	30-3.5 (3.62-3.50)
No. of Refls. (Total/Unique)	55,313/ 11,003	48,089/ 16,874	48,062/ 16,862	48,572/ 17,037
Completeness (%)	98.1 (92.0)	99.6 (100.0)	99.6 (100.0)	99.7 (99.9)
I/σ(I)	15.3 (3.6)	18.2 (3.1)	18.1 (3.1)	16.6 (2.6)
R_{sym} (%)[a]	8 (32.4)	6.6(38.7)	6.6(40.0)	7.3(48.8)
Figure of Merit		0.42		
Refinement Resolution (Å)	30-3.25			
No. of Residues	300			
No. of Waters	0			
R_{cryst}/R_{free}[b]	24.09/27.01			
Rmsd bond lengths	0.008			
Rmsd bond angles	1.4			

[a] $R_{sym} = \Sigma|I - \langle I \rangle| / \Sigma I$, where I is observed intensity and $\langle I \rangle$ is average intensity obtained from multiple observations of symmetry related reflections; [b] $R_{cryst} = \Sigma ||F_{observed}| - |F_{calculated}|| / \Sigma |F_{observed}|$, $R_{free} = R_{cryst}$ calculated using 5% random data omitted from the refinement.

structure. A large portion of the model was built automatically, but manual model building was required for many of the side-chains, and for the backbone in regions of poor density. ARP/wARP (Perrakis, Morris et al. 1999) was also used as an alternative model building tool, and the initial model was a combination of segments built by RESOLVE, ARP/wARP, and manually. This model was improved by refinement and iterative building. The final model of the triple mutant was subsequently used to solve the structure of the wild-type protein, using a native data set to 2.3 Å.

Although the native and SeMet crystals were essentially isomorphous, phases were obtained by molecular replacement in order to avoid any potential problems due to the slight differences in unit cell parameters. The structure of the wild-type protein was subsequently refined and side-chain conformations were manually adjusted. The final model had an R_{cryst}/R_{free} of 24.2/26.2% (pdb ID 1rp3). The crystallographic statistics are provided in Table 3.1.

4. ANALYSIS OF THE σ^{28}/FLGM CRYSTAL STRUCTURE

4.1 Overall Structure

The structure reveals the entire σ^{28} molecule, comprising σ_2 (green, Figure 4.1), σ_3 (blue), and σ_4 (yellow). Connecting σ_3 and σ_4 is the σ_3-σ_4 linker (grey). In stark contrast to the extended arrangement of these domains in the σ^A-holoenzyme (Murakami, Masuda et al. 2002; Vassylyev, Sekine et al. 2002), the three σ^{28} domains pack together in a tight, compact unit in the σ^{28}/FlgM complex, with the extended FlgM molecule wrapped around the outside (red, Figure 4.1A). Both proteins are entirely α-helical (σ^{28}, H1-H10; FlgM, H1'-H4').

In total, five crystallographically independent σ^{28}/FlgM complexes were solved, with no major conformational differences between any of the complexes. Comparing α-carbon positions over 279 residues between each of the five complexes (σ^{28} residues 10-224 and FlgM residues 4-17, 35-50, and 55-88), the maximum root-mean-square-deviation (rmsd) between any two complexes is 1.8 Å. Excluding σ^{28} residues 148-173 (a relatively variable segment, see below), the maximum rmsd is 1.0 Å, indicating that the structures represent the σ^{28}/FlgM complex in solution and are not perturbed by crystal packing.

Both the superposition and the B-factors reveal that the most flexible region of σ^{28} (Figure 4.2 A) is the σ_3-σ_4 linker. FlgM (Figure 4.2 B) has greater flexibility overall, compared to σ^{28}, with the exception of the H'4. H'1 through H'3 are all rather flexible, and the loop connecting H'1 and H'2 contains a completely disordered segment of ~10 residues, while the loop connecting H'2 and H'3 lacks interpretable density in two chains, and is highly flexible in the remaining two.

4.2 Structure of σ^{28}

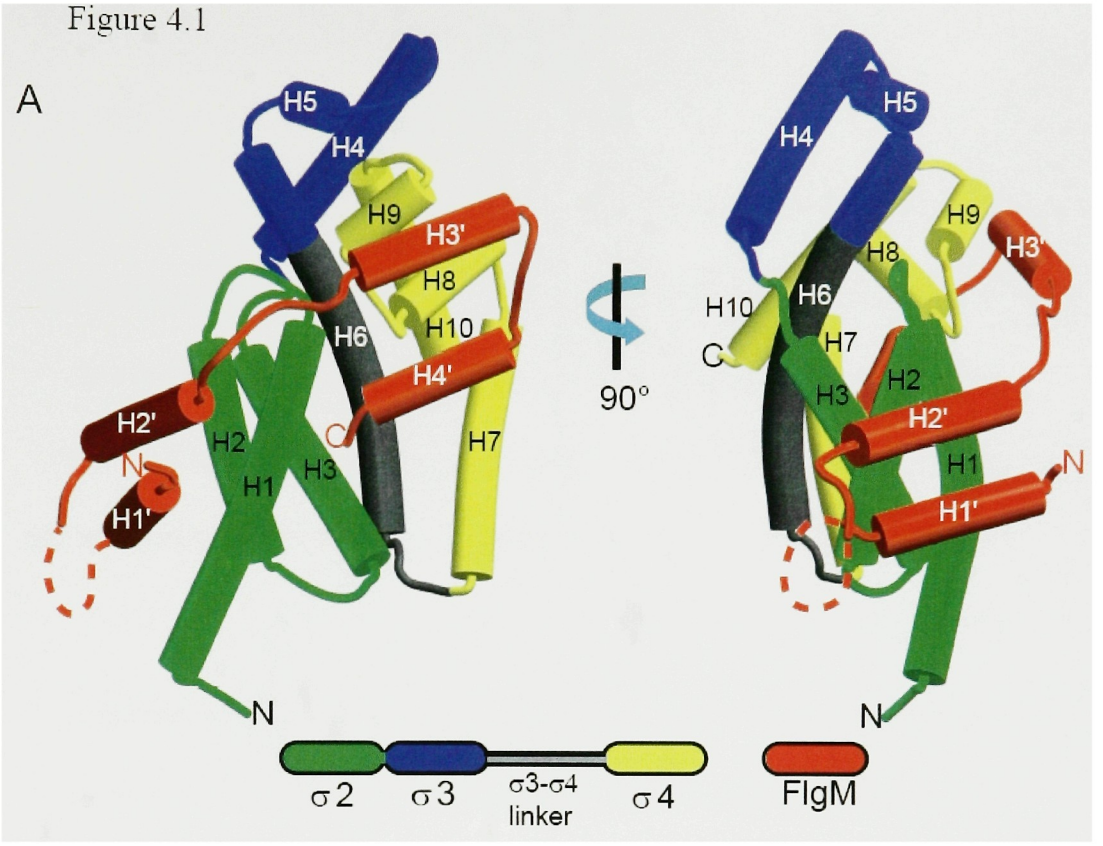
The structures of the individual σ^{28} domains, $\sigma^{28}{}_2$, $\sigma^{28}{}_3$, and $\sigma^{28}{}_4$, are essentially identical to those observed in structures of other σ^{70}-family members (Malhotra, Severinova et al. 1996; Campbell, Masuda et al. 2002; Campbell, Muzzin et al. 2002; Li, Stevenson et al. 2002; Murakami, Masuda et al. 2002; Vassylyev, Sekine et al. 2002; Campbell, Tupy et al. 2003). In the RNAP σ^{A}-holoenzyme, these three domains are extended across the surface of the core RNAP (Figure 4.9; (Murakami, Masuda et al. 2002; Vassylyev, Sekine et al. 2002)). Connecting σ_3 and σ_4, which are separated by 45 Å in the holoenzyme, is the σ_3-σ_4 linker, that penetrates into the RNAP active site channel, contributing to stabilization of the initiation

Figure 4.1. Structure of the Aa σ^{28}/FlgM complex.

A) Two views (related by a 90° rotation as shown) of the Aa σ^{28}/FlgM complex. Helices (shown as cylinders) are labeled H1-H10 in σ^{28} and H1'-H4' in FlgM. The domain architecture and color-coding are illustrated schematically at the bottom. Structural elements of σ^{28} are color coded as follows: σ_2, green; σ_3, blue; σ_3-σ_4 linker, grey; σ_4, yellow. FlgM is colored red. The loop between H1' and H2' of FlgM was not modeled, and is shown as a dotted line.

B) Section of the 2.3 Å-resolution $2F_o$-F_c map, contoured at 1σ, showing the segment joining the σ_3-σ_4 linker and σ_4. The model is shown as sticks and color-coded according to σ conserved regions (region 3.1, orange; 3.2, red; 4.1, green; (Lonetto, Gribskov et al. 1992)) as in Figure 4.3.

Figure 4.1


A
H5
H4
H9
H3'
H8
H6
H10
H4'
H2'
H2
N
H1
H3
H1'
C
H7
N
90°
H5
H4
H9
H3'
H8
H10
H6
C
H7
H2
H3
H2'
H1
N
H1'
N
σ2
σ3
σ3-σ4
linker
σ4
FlgM

B

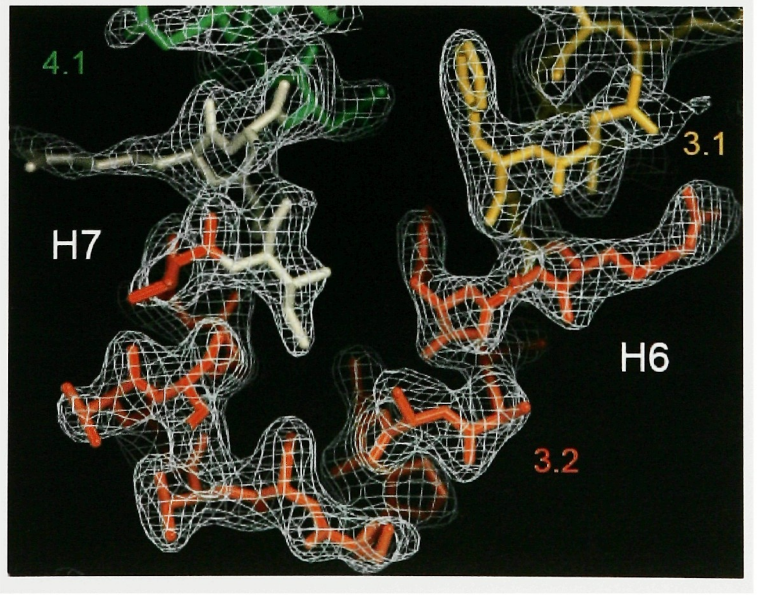
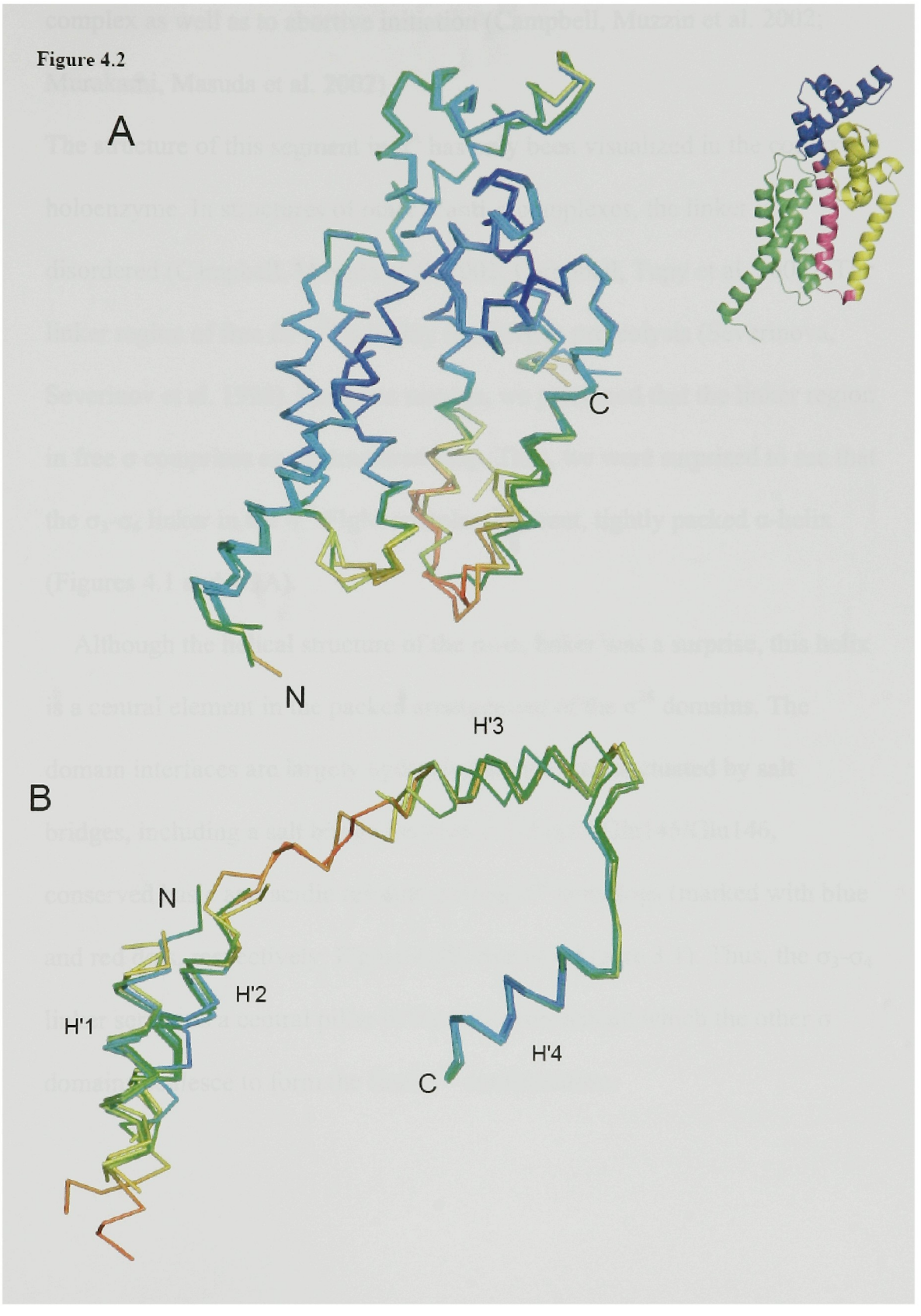

4.1
3.1
H7
H6
3.2

Figure 4.2. Structural superposition

A) Superposition of the four σ^{28} molecules from the 2.3 Å structure. The chains are rainbow colored by B-factor from indigo (lowest) to red (highest). The ribbon diagram to the right shows one chain of σ^{28} in the same orientation, and color coded as in Figure 4.1A, except for the σ_3-σ_4 linker, which is pink.

B) Superposition of the four FlgM molecules from the 2.3 Å structure. The chains are colored by B-factor as in A. The termini and helices are labeled

Figure 4.2



complex as well as to abortive initiation (Campbell, Muzzin et al. 2002; Murakami, Masuda et al. 2002)

The structure of this segment in σ^A has only been visualized in the context of holoenzyme. In structures of other σ:anti-σ complexes, the linker is disordered (Campbell, Masuda et al. 2002; Campbell, Tupy et al. 2003). The linker region of free *Ec* σ^{70} is highly sensitive to proteolysis (Severinova, Severinov et al. 1996). For these reasons, we presumed that the linker region in free σ comprises an unstructured loop. Thus, we were surprised to see that the σ_3-σ_4 linker in the σ^{28}/FlgM complex is a bent, tightly packed α-helix (Figures 4.1 and 4.3A).

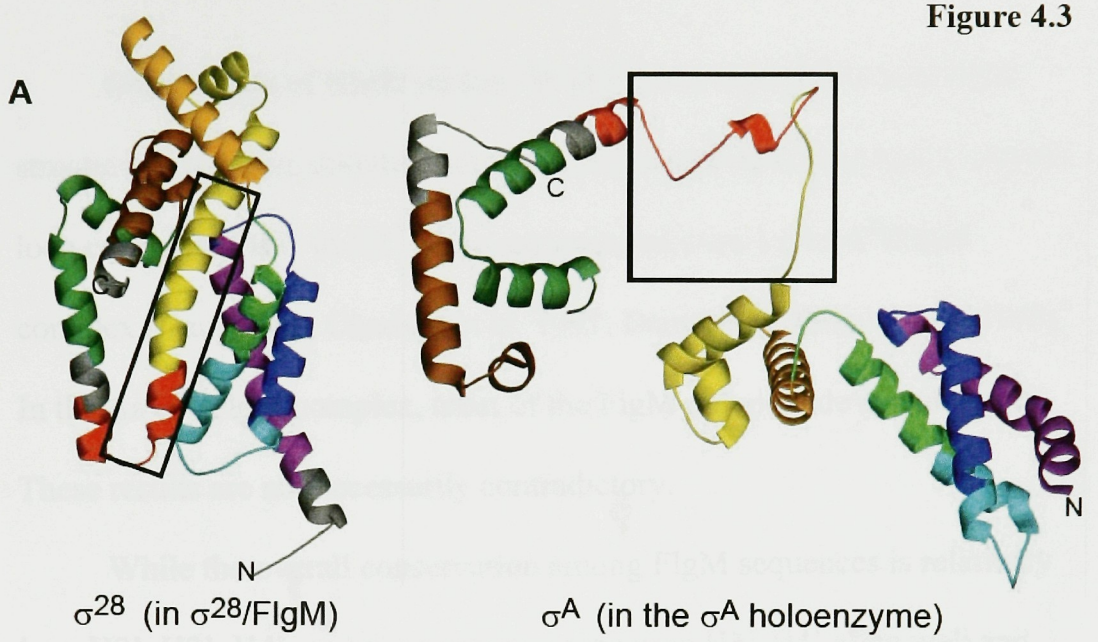
Although the helical structure of the σ_3-σ_4 linker was a surprise, this helix is a central element in the packed arrangement of the σ^{28} domains. The domain interfaces are largely hydrophobic, but are punctuated by salt bridges, including a salt bridge between σ^{28}-Arg82:Glu145/Glu146, conserved basic and acidic residues among σ^{28} homologs (marked with blue and red dots, respectively, Figure 4.3B. See also Figure 3.1). Thus, the σ_3-σ_4 linker serves as a central pillar of the structure, around which the other σ domains coalesce to form the final σ^{28} conformation.

Figure 4.3. Comparison of σ^{28} and σ^{A}.

A) Ribbon diagrams showing the conformations of Aa σ^{28} in the σ^{28}/FlgM complex (left) and Tth σ^{A} from the σ^{A}-holoenzyme (right; PDB ID 1IW7; (Vassylyev, Sekine et al. 2002)). On the left, FlgM has been removed; on the right, the N-terminal segment of σ^{A} which lacks a counterpart in σ^{28}, as well as the core RNAP, have all been removed to facilitate comparison. Conserved regions of σ (Lonetto, Gribskov et al. 1992) are color coded as shown in Figure 4.3 B. The segment which differs in secondary structure between the two structures (the σ_3-σ_4 linker of σ^{28} and the $\sigma_{3.2}$-loop of σ^{A}) is boxed.

B) Sequence alignment of σ^{28} homologs, and structure-based alignment with Tth σ^{A}. Conserved regions 2.1-4.2 (Lonetto, Gribskov et al. 1992) are denoted as colored bars above the sequences. A total of 23 σ^{28} sequences were aligned (Figure 3.1) using Clustal (Gish and States 1993), but only Aa, St, and *B. subtilis* (Bs) are displayed. The σ^{A} was aligned with Aa σ^{28} by structural superposition of conserved domains. Conservation in the complete σ^{28} alignment is denoted by green (high) or grey (moderate) shading. Those residues which are also conserved in σ^{A} (based on an alignment of 47 primary σ's) are also shaded. The segment which differs in secondary structure is boxed (corresponding to the boxed segments in Figure 2A). Residues that interact with FlgM in all four complexes of the form I crystals (< 4 Å) are highlighted in magenta. Blue and red dots above the sequences denote conserved basic and acidic residues (respectively) that participate in a salt bridge stabilizing the σ^{28} conformation.

Figure 4.3



B

```
                                    2.1                                 2.2
  1 -------------MKNPYSNQIEREELILKYLPLVKAIATNIKKHLPEDVDIRDLISYGVIGLIKAVDNL Sig28 Aa
  1 -----------MNSLYTAEGVMDKHSLWQRYVPLVRHEALRLQVRLPASVELDDLLQAGGIGLLNAVDRY Sig28 St
  1 MQSLNYEDQVLWTRWKEWKDPKAGDDLMRRYMPLVTYHVGRISVGLPKSVHKDDLMSLGMLGLYDALEKF Sig28 Bs
174 -------------LHIAREGEAARQHLIEANLRLVVSIAKKYTGRG---LSFLDLIQEGNQGLIRAVEKF SigA  Tt

                  2.3                 2.4                           3.0
 58 STENPKRAEAYIKLRIKGAIYDYLRSLDF---GSRQVREKERRIKEVVEKLKEKLGREPTDEEVAKELG- Sig28 Aa
 60 DALQGTAFTTYAVQRIRGAMLDELRSRDW---VPRSVRRNAREVAQAMGQLEQELGRNATETEVAERLG- Sig28 St
 71 DPSRDLKFDTYASFRIRGAIIDGLRKEDW---LPRTSREKTKKVEAAIEKLEQRYLRNVSPAEIAEELG- Sig28 Bs
228 EYKRRFKFSTYATWWIRQAINRAIADQARTIRIPVHMVETINKLSRTARQLQQELGREPTYEEIAEAMGP SigA  Tt

                       3.1                           3.2
124 -ISTEELFKTLDKINFSYILSLEEVFRDFARDYS-------ELIPSST--NVEEEVIKRELTEKVKEAVS Sig28 Aa
126 -IPVAEYRQMLLDTNNSQLFSYDEWREEHGDSI-----ELVTEEHQQE--NPLHQLLEGDLRQRVMDAIE Sig28 St
137 -MTVQDVVSTMNEGFFANLLSIDEKLHDQDDGE---NIQVMIRDDKNV--PPEEKIMKDELIAQLAEKIH Sig28 Bs
298 GWDAKRVEETLKIAQEPVSLETPIGDEKDS-FYG-------DFIPDEHLPSPVDAATQSLLSEELEKALS SigA  Tt

                  4.1                           4.2
184 KLPEREKLVIQLIFYE----ELPAKEVAKILETSVSRVSQLKAKALERLREMLSNPL-------       Sig28 Aa
188 SLPEREQLVLTLYYQE----ELNLKEIGAVLEVGESRVSQLHSQAIKRLRTKLGKL--------       Sig28 St
201 ELSEKEQLVVSLFYKE----ELTLTEIGQVLNLSTSRISQIHSKALFKLKNLLEKVIQ------       Sig28 Bs
360 KLSEREAMVLKLRKGLIDGREHTLEEVGAFFGVTRERIRQIENKALRKLKYHESRTRKLRDFLD       SigA  Tt
```

4.3 Structure of FlgM

On the basis of NMR studies, St FlgM was reported to lack stable structure in solution, and the first 40 residues, corresponding to H1' and the loop connecting H1' to H2', were unstructured even in the σ^{28}/FlgM complex (Daughdrill, Chadsey et al. 1997; Daughdrill, Hanely et al. 1998). In the Aa σ^{28}/FlgM complex, most of the FlgM polypeptide is structured. These results are not necessarily contradictory.

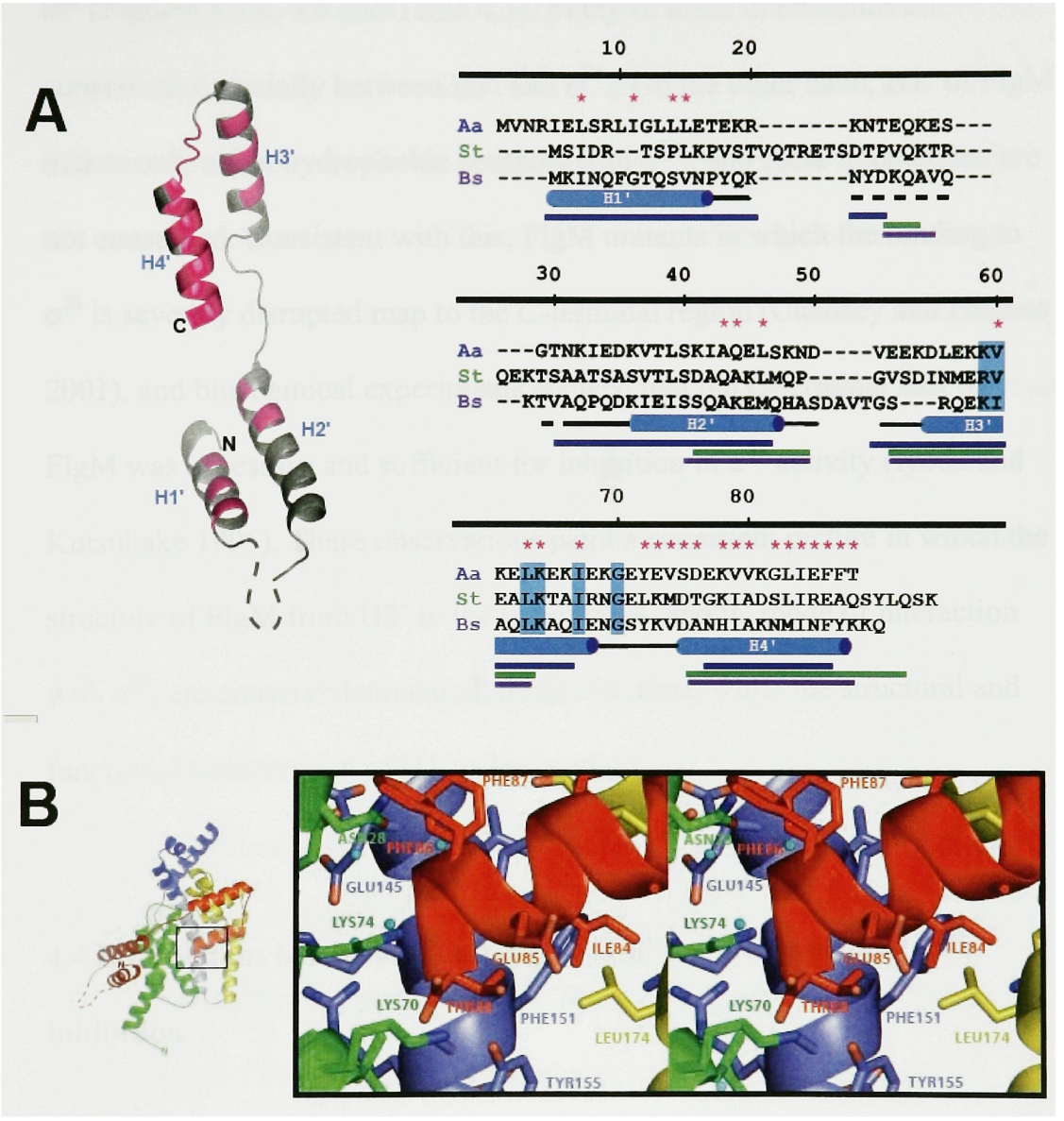
While the overall conservation among FlgM sequences is relatively low, H2', H3', H4', and the segment connecting H3'-H4' align well and contain clusters of conserved residues (Figures 4.4A and 4.4C). On the other hand, sequences N-terminal of H2' align very poorly, requiring large gaps, and FlgM from some bacteria appears to lack sequences corresponding to H1' all together. A secondary structure prediction algorithm predicted all four helices in Aa FlgM, and helices corresponding to H2', H3', and H4' in other FlgM sequences, but not H1' (Figure 4.4A; (Rost and Sander 1993)).

Figure 4.4. FlgM structure and σ^{28} interactions.

A) **Structure and sequence alignment of FlgM.** (Left) Ribbon diagram of FlgM from the σ^{28}/FlgM complex. The backbone is grey, except for residues interacting with σ^{28} in all four non-crystallographically related complexes (< 4 Å), which are colored pink. The segment not modeled is shown as a dotted line. (Right) Alignment of FlgM sequences. A total of 18 FlgM sequences were aligned (Figure 3.2) using WUBLAST (Gish 1996-2004), but only Aa, St, and Bs are shown. Numbering is according to the Aa sequence. A pink asterisk above a residue denotes interaction with σ^{28}. Cyan shading of residues denotes high conservation in the complete alignment (>80% identity). The average secondary structure of the four complexes in the 2.3 Å structure is shown below, with blue cylinders denoting α-helices, thick black lines denoting loops, and the dotted line denoting the disordered segment. Predicted helices (Rost and Sander 1993) are shown as thick lines below the alignment, top-bottom: Aa (blue), St (green), Bs (purple).

B) **Interactions between the C-terminal helix of FlgM (H4') and σ^{28}.** (Left) Structural overview of the σ^{28}/FlgM complex (same view and color scheme as Figure 4.1A, left). The boxed region is shown magnified in the stereo view. The backbone and side-chain carbon atoms are colored as follows: FlgM, red; σ_2 green, σ_3-σ_4 linker, slate blue; σ_4, yellow. Side chain oxygen and nitrogen atoms are colored red and blue respectively. Water molecules are displayed as cyan spheres.

Figure 4.4


A
H3'
H4'
C
N
H2'
H1'
10
20
Aa MVNRIELSRLIGLLLETEKR-------KNTEQKES---
St ----MSIDR--TSPLKPVSTVQTRETSDTPVQKTR---
Bs ----MKINQFGTQSVNPYQK-------NYDKQAVQ---
H1'
30
40
50
60
Aa ---GTNKIEDKVTLSKIAQELSKND----VEEKDLEKKV
St QEKTSAATSASVTLSDAQAKLMQP-----GVSDINMERV
Bs --KTVAQPQDKIEISSQAKEMQHASDAVTGS---RQEKI
H2'
H3'
70
80
Aa KELKEKIEKGEYEVSDEKVVKGLIEFFT
St EALKTAIRNGELKMDTGKIADSLIREAQSYLQSK
Bs AQLKAQIENGSYKVDANHIAKNMINFYKKQ
H4'
B
PHE87
GLU145
LYS74
GLU85
ILE84
LYS70
PHE151
LEU174
TYR155

FlgM helices H2', H3', H4', and the segment connecting H3'-H4', make hydrophobic, hydrogen bonding, and favorable electrostatic interactions with σ^{28} (Figures 4.3A, 4.4 and Table 4.1). Many of these interactions are conserved, especially between H3' and $\sigma^{28}{}_4$. On the other hand, H1' of FlgM makes only a few hydrophobic contacts with $\sigma^{28}{}_2$, and these interactions are not conserved. Consistent with this, FlgM mutants in which the binding to σ^{28} is severely disrupted map to the C-terminal region (Chadsey and Hughes 2001), and biochemical experiments showed that the C-terminal half of FlgM was necessary and sufficient for inhibition of σ^{28} activity (Iyoda and Kutsukake 1995). These observations paint a consistent picture in which the structure of FlgM from H2' to the C-terminus, and its mode of interaction with σ^{28}, are conserved among all motile bacteria, while the structural and functional conservation of H1' is less certain.

4.4 Interactions between σ^{28} and FlgM, and the mechanism of σ^{28} inhibition

The FlgM structure can be neatly divided in half, with the N-terminal half comprising H1'-H2', and the C-terminal half comprising H3'-H4' (Figure 4.4A) FlgM interacts with σ^{28} in a multipartite manner, with FlgM

helices H1'-H2' interacting with σ_2 (buried surface area of 934 $Å^2$), and FlgM H3'-H4' interacting with σ_4 (buried surface area of 1,263 $Å^2$). FlgM does not interact with σ_3. At the very C-terminus of H4', FlgM residues also interact with σ_2 and the σ_3-σ_4 linker. Thus, the C-terminal helix (H4') of FlgM interacts with three σ^{28} structural elements (Figure 4.4B). This helix sits in a cavity lined by two σ_4 helices (H7 and H8). At the bottom of the cavity, the helical dipole and the negative charge of the carboxyl-terminus (Thr88) are stabilized by two Lys residues from σ_2 (Lys70 and the conserved Lys74), Phe87 makes hydrophobic interactions with residues in the σ_3-σ_4 linker, Phe86 interacts with σ_2, and Glu85 makes a favorable electrostatic interaction with Lys74.

A summary of the interactions observed in the 2.3Å structure are shown in Table 4.1. The types of interactions are displayed graphically in Figure 4.5. Although most of the interface is hydrophobic, both the N and C terminal halves of FlgM also interact through hydgrogen bonding and salt bridges. Interacting residues which are conserved amongst σ^{28} orthologs are highlighted in Table 4.1, and displayed graphically in red, in Figure 4.6. Most of these residues cluster in σ_4, except for two residues in σ_2, Lys74, which interacts with the C-terminus of FlgM, and S44 (not visible in Figure4.6), which interacts with H'3.

Table 4.1. Summary of σ^{28}/FlgM interactions.

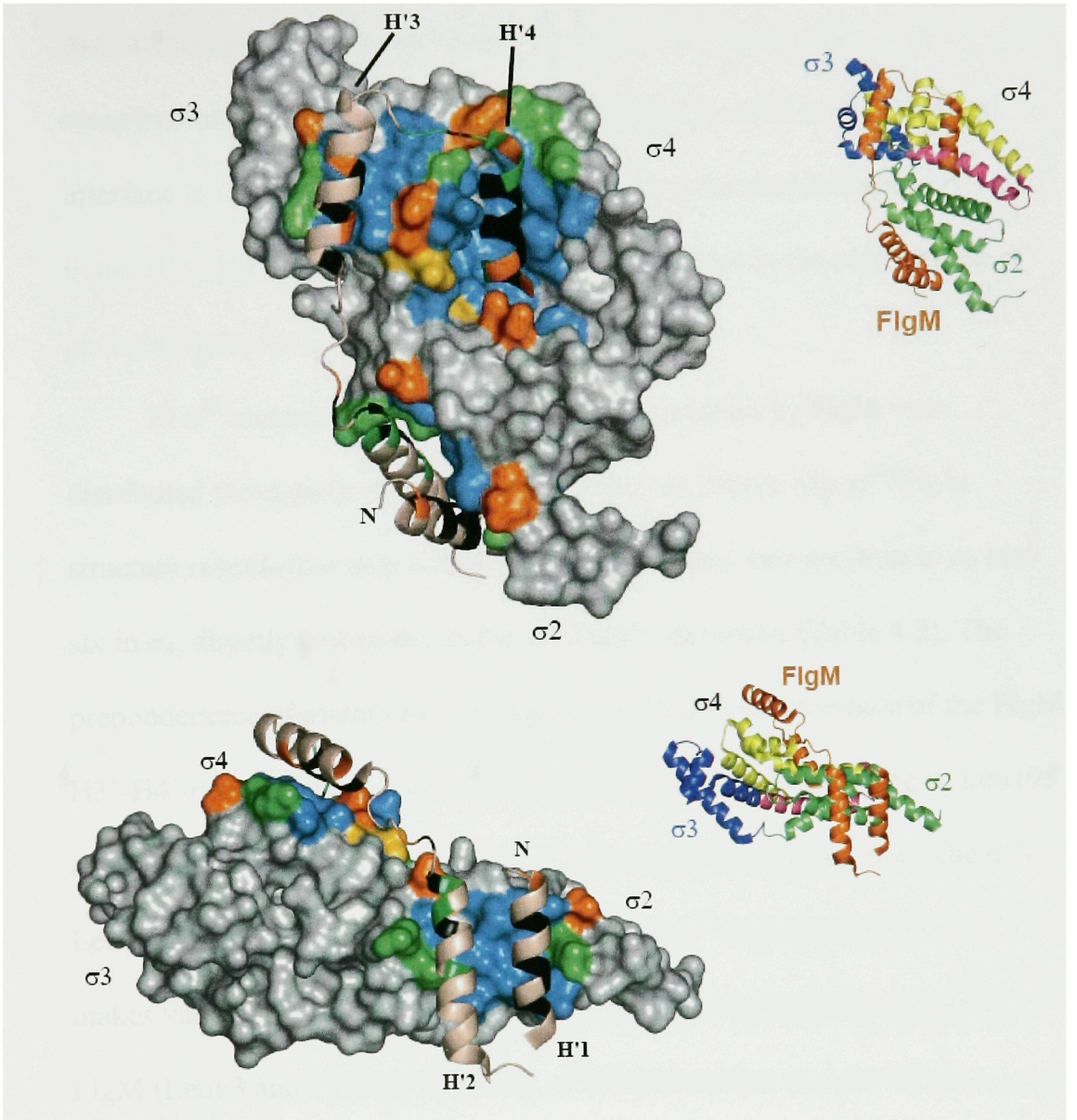
σ^{28}	FlgM	Type[1]
R11	L15	HB
I15	I11	Hy
I15	L15	Hy
L16	L7	Hy
L19	L46	Hy
L19	I11	Hy
K23	L46	Hy, HB
K23	N49	Hy
N28	F86	W
N28	F86	Hy
I29	F86	W
H32	F86	Ar
R40	A43	Hy
R40	Q44	HB
I43	A43	Hy
S44	L39	HB
S44	A43	Hy
V47	L39	Hy
I48	L39	Hy
I51	L14	Hy
K70	T88	SB
K74	T88	SB
K74	F86	W
K74	F86	HB
K74	E85	SB
Y78	F86	W
Y78	F86	Ar
I141	F86	Hy
I141	F87	Hy
L144	F87	Hy
E145	F87	Hy
F148	F87	Ar
F148	T88	Hy
T175	I84	Hy
V178	V80	Hy
K179	V80	Hy
V182	V79	Hy
V182	V80	Hy

V182	L83	Hy
S183	D76	HB
K190	V74	Hy
K190	V79	Hy
K190	D76	SB
L191	V74	Hy
I193	V79	Hy
Q194	V79	Hy
Q194	Y72	HB
Q194	V74	Hy
Q194	E73	HB
Q194	K78	Hy
Q194	S75	HB
L195	Y72	Hy
F197	G82	Hy
F197	F86	Hy
Y198	K78	Hy
Y198	G82	Hy
E199	K59	SB
E199	Y72	HB
E199	L63	Hy
E200	F86	W
L201	L63	Hy
E205	V60	Hy
K208	K64	HB
I209	V60	Hy
I209	I67	Hy
I209	K64	Hy
E211	K64	SB
L228	L83	Hy

Interactions were calculated using Contact (CCP4) (1994), as pairs of residues separated by 4Å or less. Only interactions observed in at least 3 of the 4 complexes in the 2.3Å structure are listed. Conservation is marked by shading, and color coded as follows: yellow = moderate, orange = high, red = absolute (based on Clustal W sequence alignments of σ^{28} and FlgM orthologs, with 23 sequences in alignment). Interaction types are abbreviated as follows: Hy = hydrophobic, HB = Hydrogen bond, SB = Salt bridge, Ar= Aromatic stacking, W = Water mediated.

Figure 4.5. Interaction types

σ^{28} is represented as a grey molecular surface and FlgM as a pink ribbon. Interacting residues are colored according to interaction type as follows: hydrophobic or aromatic, blue (σ^{28}) or black (FlgM); hydrogen bonds, green; salt bridges, red; water-mediated hydrogen bonds, orange. Interactions with the C-terminal (top) and N-terminal (bottom) halves of FlgM are displayed separately. A smaller ribbon diagram in the same orientation is shown for each, and color coded by domain as in Figure 4.1, except the σ_3-σ_4 linker is shown in pink.



σ28
FlgM
Interaction Types:
Hydrophobic/Aromatic (σ28)
Hydrophobic/Aromatic (FlgM)
Hydrogen bond
Salt bridge
Water mediated hydrogen bond

Figure 4.5

Conserved interacting residues in FlgM (Figure 4.6, purple) are all located in its C-terminal half, with one conserved face of H'3 interacting with a complementary conserved patch on σ_4, forming the most well conserved interface in the interaction. In addition, some conserved residues in H'4 and in the H'3-H'4 loop interact with conserved residues in the surrounding σ^{28} pocket (Figure 4.6).

St σ^{28} mutants defective for negative regulation by FlgM were distributed throughout σ^{28} (Chadsey and Hughes 2001). The σ^{28}/FlgM structure reveals that only a subset of these mutants, one position in σ_2 and six in σ_4, directly participate in the σ^{28}/FlgM interaction (Table 4.2). The preponderance of mutations in σ_4 again confirms the importance of the FlgM H3'-H4'/σ_4 interaction. Two of the positions in σ_4, corresponding to Leu195 and Ile209 of Aa σ^{28}, disrupt conserved hydrophobic interactions. The σ^{28}-Leu195, which is absolutely conserved among σ^{28} sequences (Figure 3.1), makes van der Waal's interactions with conserved hydrophobic residues of FlgM (Leu63 and Tyr72). Similarly, σ^{28}-Ile209 makes van der Waal's interactions with FlgM-Val60 (conserved as Val or Ile) and the absolutely conserved FlgM-Ile67 (Figure 3.2). Consistent with this, the mutants at these two σ^{28} positions are among the most resistant to FlgM inhibition *in vivo*,

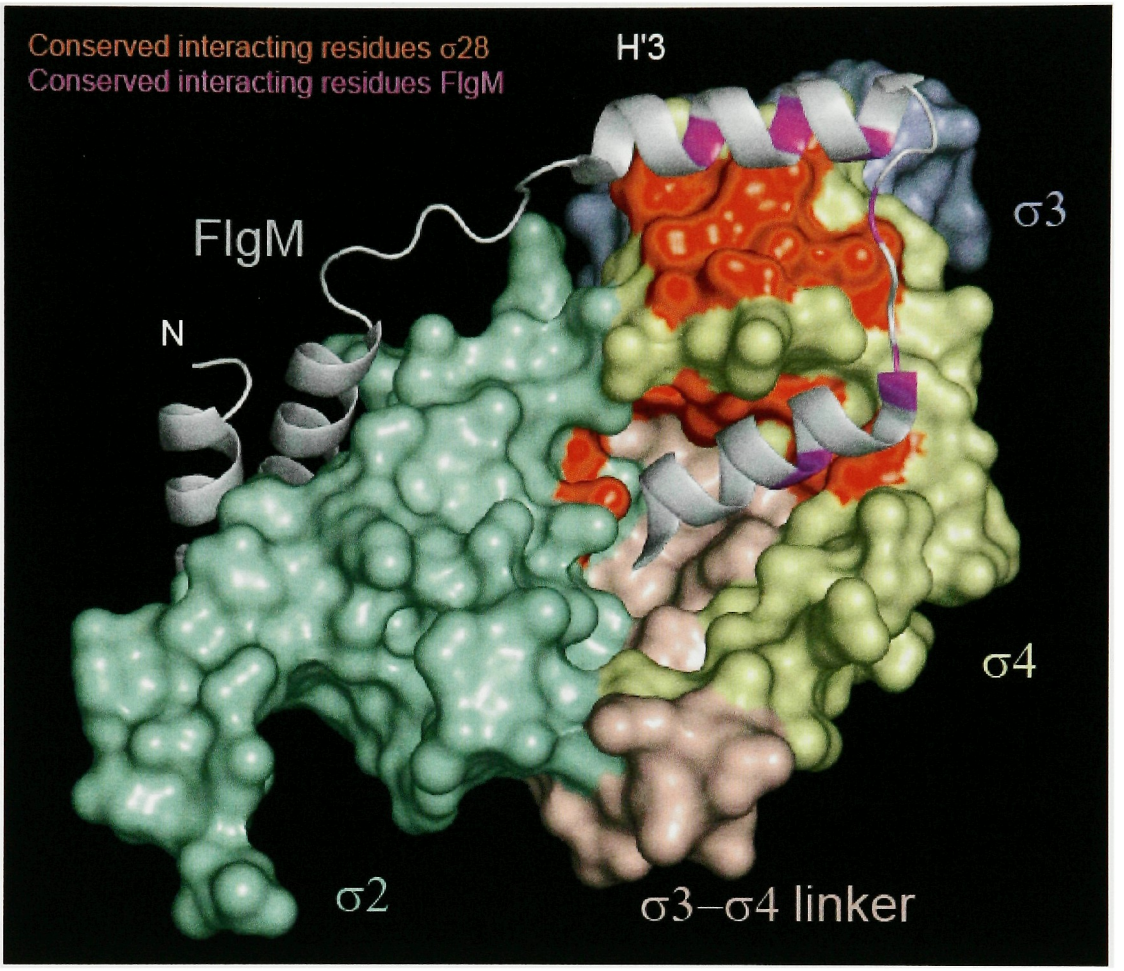


Figure 4.6 Conserved interacting residues.

The location of conserved interacting residues in the crystal structure is shown. σ^{28} is represented as a surface, and FlgM as a ribbon. The domains of σ^{28} have been color coded as follows: σ_2 is pale green, σ_3 is pale blue, the σ_3-σ_4 linker is pink, and σ_4 is pale yellow. Conserved residues corresponding to those highlighted in Table 4.1 have been colored red in σ^{28} and purple in FlgM.

Table 4.2 σ^{28}/FlgM interactions, genetic and structural results

S. typhimurium substitution(s) causing defect in FlgM inhibition[1]	corresponding *A. aeolicus* σ^{28} residue[2]	σ^{28} conserved region	role in σ^{28}/FlgM structure
H14D/N/Y	E12	2	interacts with FlgM
V33E	K31	$\sigma_{\square}$	interacts with σ_4
M104T	V102	σ_3	near interface with σ_4
N114K/L	E112	σ_3	interacts with σ_4
L124F	L122	σ_3	σ_3 hydrophobic core
T138I	I136	σ_3-σ_4 linker	interacts with σ_3
N139I/K	Y137	σ_3-σ_4 linker	interacts with σ_4
P190Q/S	P186	σ_4	turn between σ^{28} H7-H8
V196G	V192	σ_4	σ_4 hydrophobic core
T198K	Q194	σ_4	interacts with FlgM
L199Q/R	L195	σ_4	interacts with FlgM
Q202L/R	Y198	σ_4	interacts with FlgM
E203D	E199	σ_4	interacts with FlgM
N206K	P202	σ_4	interacts with σ_2
E209D	E205	σ_4	interacts with FlgM
V213E/G	I209	σ_4	interacts with FlgM
S226N/R	A222	σ_4	interacts with σ_3-σ_4 linker
R231Q	R227	σ_4	-

[1](Chadsey and Hughes 2001).
[2]See Figure 3.1.

and have the most severe FlgM binding defects *in vitro* (Chadsey and Hughes 2001). Interestingly, most of the remaining mutants that do not directly participate in the FlgM interaction are involved in interdomain interfaces within σ^{28}, and may destabilize the σ^{28} conformation required for the FlgM complex.

Although all of the σ structural domains and linker regions interact with core RNAP (Murakami, Masuda et al. 2002; Vassylyev, Sekine et al. 2002), the most prominent interface occurs between σ_2 (mainly an α-helix comprising conserved region 2.2, H2 in σ^{28}) and the RNAP β'-coiled-coil (Arthur and Burgess 1998; Arthur, Anthony et al. 2000; Murakami, Masuda et al. 2002). In the packed conformation of σ^{28}, the β'-coiled-coil binding determinant in $\sigma^{28}{}_2$ is facing outwards but is directly occluded by H1'-H2' of FlgM (Figure 4.7A).

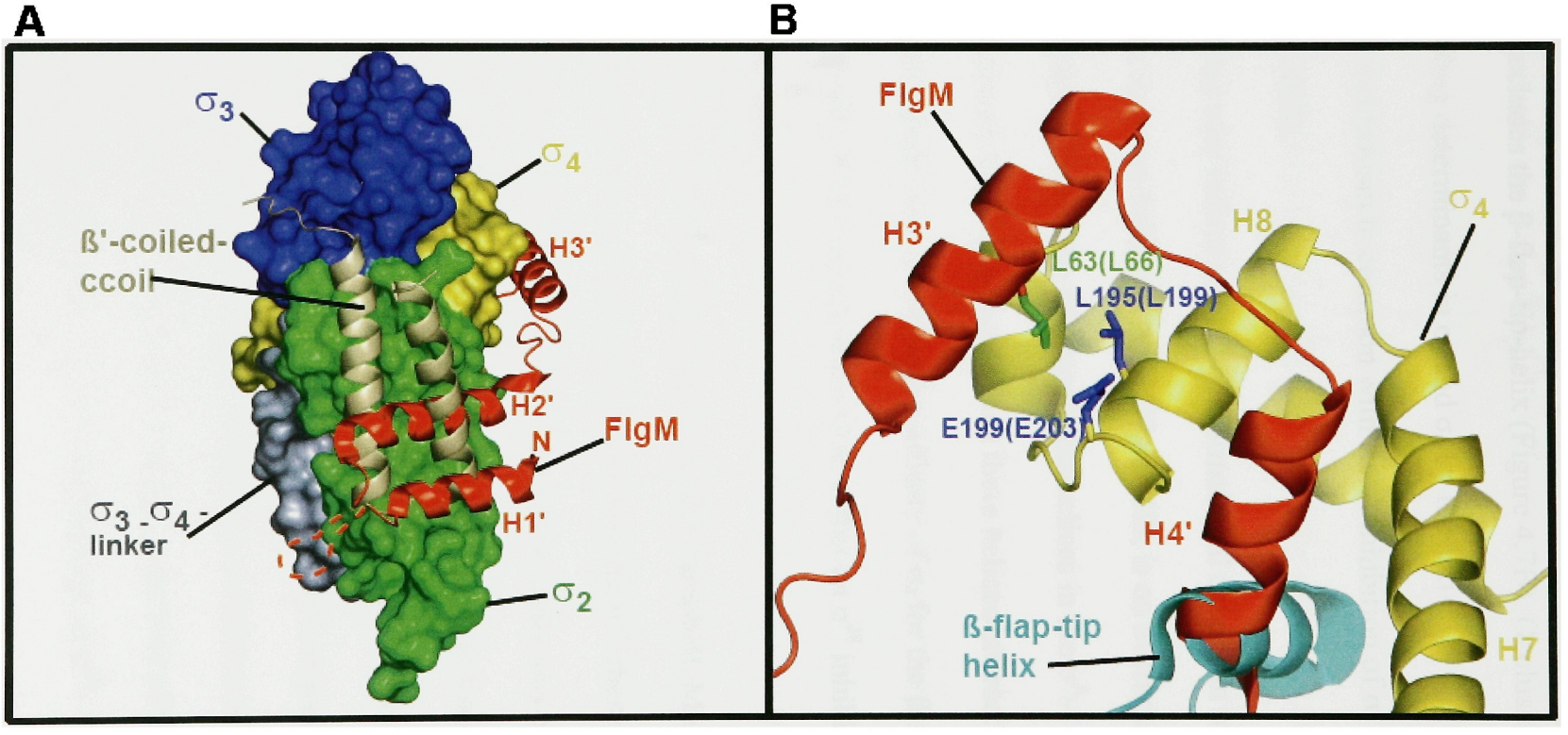
In the RNAP holoenzyme, σ_4 interacts with the core RNAP by clamping onto the β-flap-tip-helix (Kuznedelov, Minakhin et al. 2002; Murakami, Masuda et al. 2002; Vassylyev, Sekine et al. 2002). The binding of FlgM

Figure 4.7. Steric occlusion of σ^{28} core binding surfaces by FlgM.

A) Steric clash between FlgM H1'-H2' and the β'-coiled-coil of core RNAP. View of the σ^{28}/FlgM complex (same view as shown in the right of Figure 4.1A), color-coded as in Figure 4.1A, with σ^{28} shown as a molecular surface and FlgM shown as a backbone ribbon. The β'-coiled-coil (tan ribbon) is shown in the context of the holoenzyme (positioned as described in the text), illustrating the steric clash with FlgM.

B) Steric clash between FlgM H4' and the β-flap-tip-helix of core RNAP. View of the $\sigma^{28}{}_4$/FlgM (H3'-H4') interaction, with the β-flap-tip-helix shown in the context of the holoenzyme (positioned as described in the text). The proteins are shown as backbone ribbons, color-coded as follows: $\sigma^{28}{}_4$, yellow; FlgM, red; β-flap-tip-helix, cyan. Three residues important for FlgM attack and dissociation of σ^{28}-holoenzyme (Chadsey, Karlinsey et al. 1998; Chadsey and Hughes 2001) interact directly in the structure, and are shown as sticks and colored green (FlgM) or blue (σ^{28}). These are labeled to indicate the 1-letter amino acid code and the residue number in Aa and St (in parentheses).

Figure 4.7


A
σ3
σ4
ß'-coiled-
ccoil
H3'
H2'
N
FlgM
σ3 -σ4-
linker
H1'
σ2
B
FlgM
H3'
H8
σ4
L63(L66)
L195(L199)
E199(E203)
H4'
ß-flap-tip
helix
H7

H3'-H4' to $\sigma^{28}{}_4$ occludes the β-flap-tip-helix (Figure 4.7B). Occlusion of the primary core binding determinants in σ_2 and σ_4 by the anti-σ is an inhibition mechanism in common with RseA inhibition of σ^E (Campbell et al., 2003).

FlgM binding may induce a small conformational change in $\sigma^{28}{}_4$ (the angle between the first two helices of $\sigma^{28}{}_4$, H7 and H8, is on average 11° greater than the angle between the corresponding helices in the σ^A-holoenzyme structures). Since residues in both these helices interact with the β-flap-tip-helix, this angle may influence the affinity of σ_4 for the β-flap-tip-helix. However, the $\sigma^{28}{}_4$/FlgM interaction and its effect on σ^{28} inhibition may be even more complex.

Like other anti-σ factors investigated in detail (Campbell, Masuda et al. 2002; Campbell, Tupy et al. 2003), FlgM can bind free σ^{28}, thereby preventing its interaction with core RNAP. Unlike these other anti-σ's, however, FlgM can also form a ternary complex with the σ^{28}-holoenzyme, thereby destabilizing the σ^{28}/RNAP interaction, as shown by an increased dissociation rate of the σ^{28}/FlgM complex from core RNAP compared with σ^{28} alone (Chadsey, Karlinsey et al. 1998). Among the collection of σ^{28} and FlgM mutants that destabilize the σ^{28}/FlgM complex, only four of these (two in σ^{28} and two in FlgM) attenuate the FlgM destabilization of σ^{28}-

holoenzyme as well (σ^{28}-Leu195, σ^{28}-Glu199, FlgM-Leu63, and FlgM-Val79; (Chadsey and Hughes 2001). All of these mutants are at the interface between $\sigma^{28}{}_4$ and FlgM H3'-H4', and three of these residues interact directly in some of complexes in the assymmetric unit (see 3.2) (Figure 4.7B).

The interaction of $\sigma^{28}{}_2$ with the RNAP β'-coiled-coil or FlgM H1'-H2' appears to be mutually exclusive (Figure 4.7A). In contrast, the structure suggests that $\sigma^{28}{}_4$ might be able to interact simultaneously with FlgM and the RNAP, since only the C-terminal end of FlgM H4' sterically clashes with the β-flap-tip-helix and since FlgM H3' and H4' are flexibly linked and do not interact with each other (Figure 4.7B).

These observations suggest a docking mechanism that explains the ability of FlgM to attack and destabilize the σ^{28}-holoenzyme, consistent with but extending previous models (Chadsey, Karlinsey et al. 1998; Chadsey and Hughes 2001). In the RNAP holoenzyme, the simultaneous, independent binding of the discrete structural elements of σ to different parts of core RNAP provides high-affinity binding without any one interaction between the core and an individual σ element being particularly stable, since complete dissociation of σ would require simultaneous release of every element (Murakami and Darst 2003). Nevertheless, enzyme 'breathing', or dissociation and re-association of the individual, tethered σ domains could

occur. In this scenario, the initial σ^{28}-holoenzyme/FlgM ternary complex could form through the interaction of FlgM H3' with $\sigma^{28}{}_4$. Upon release of $\sigma^{28}{}_4$ from the β-flap-tip-helix due to enzyme breathing, FlgM H4' could complete the $\sigma^{28}{}_4$/FlgM H3'-H4' interaction, forming the σ^{28}-holoenzyme/FlgM ternary complex observed by (Chadsey, Karlinsey et al. 1998), preventing reassociation of $\sigma^{28}{}_4$ with the β-flap-tip-helix (Figure 4.7B). The loss of the $\sigma^{28}{}_4$/β-flap-tip-helix interaction would increase the dissociation rate for the rest of the σ^{28}/FlgM complex from the core RNAP (Chadsey, Karlinsey et al. 1998). Upon complete dissociation of σ^{28}/FlgM from the core RNAP, the σ^{28} structural domains could fold into their packed conformation and FlgM H1'-H2' could complete the interaction, preventing $\sigma^{28}{}_2$ from reassociating with the β'-coiled-coil.

4.5 Conformation of σ^{28}

In addition to the direct observation of σ^{28}/FlgM interactions and insight into the FlgM inhibition mechanism described above, the novel conformation of the σ factor observed in the σ^{28}/FlgM structure has implications for σ factor regulation in general.

The promoter-binding surfaces of σ^{28} are masked in interdomain interfaces

All of the sequence-specific protein/DNA contacts with the conserved -10 and -35 promoter elements are mediated by σ_2 and σ_4, respectively (Murakami, Masuda et al. 2002). Nevertheless, free σ factors generally do not specifically bind promoter DNA, or do so extremely weakly (Gross, Chan et al. 1998). In the case of free σ^{28}, very weak promoter specific DNA binding has been reported (Dombroski, Walter et al. 1993; Chen and Helmann 1994), but this conflicts with other studies (Chadsey, Karlinsey et al. 1998). Even if σ^{28} is capable of promoter-specific binding on its own, this is not comparable to the specific interaction with promoter DNA in the σ^{28}-holoenzyme.

While an autoinhibitory role for the N-terminal σ region 1.1 has been delineated (Dombroski et al., 1992; Camarero et al., 2002), $\sigma_{1.1}$ is found only in primary σ's. Moreover, even in primary σ's, inhibitory mechanisms that are independent of $\sigma_{1.1}$ appear to operate. For instance, σ_2 exhibits sequence-specific -10 element interactions in complex with core RNAP but not on its own (Severinova, Severinov et al. 1996; Young, Anthony et al. 2001). This effect is somewhat mysterious since the structure of σ_2 does not change significantly upon binding core RNAP (Malhotra, Severinova et al. 1996;

Campbell, Muzzin et al. 2002; Murakami, Masuda et al. 2002; Vassylyev, Sekine et al. 2002) .

In the σ^{28}/FlgM complex, σ^{28} is arranged in a tightly packed, compact unit with significant interdomain interfaces. In this configuration, the important core binding determinants of σ_2 and σ_4 are on the surface of the σ, but are occluded by FlgM (Figure 4.7). Although there is no direct data specifying promoter binding determinants of σ^{28}, the close structural relationship with primary σ's, and the analogous arrangement of the conserved promoter elements recognized by σ^{28} in similarly spaced -10 and -35 elements (Gilman, Wiggs et al. 1981; Helmann and Chamberlin 1987) allow us to infer that promoter recognition occurs through the same protein regions. In the complex with FlgM, these promoter binding regions of $\sigma^{28}{}_2$ and $\sigma^{28}{}_4$ are not directly occluded by FlgM, but are held in a conformation where the spacing between them is incompatible with

the 78 Å spacing between the -10 and -35 elements (assuming B-form DNA). Moreover, both promoter binding regions of σ^{28} are buried in interdomain interfaces (Figure 4.8).

The orientation of the -10 element DNA with respect to $\sigma^{28}{}_2$ was modeled by superimposition of $\sigma^{28}{}_2$ onto $\sigma^{A}{}_2$ of the Taq RNAP holoenzyme/fork-junction DNA structure (PDB ID 1L9Z; (Murakami, Masuda et al. 2002).

The resulting position of the DNA, in the context of the σ^{28} conformation in the σ^{28}/FlgM complex, reveals a steric clash with the σ_3-σ_4 linker (Figure 4.8A). Most of the σ^{28} region 2.3 and 2.4 residues that correspond to σ^{A} residues implicated in -10 element recognition and melting are buried in the interface of $\sigma^{28}{}_2$ with the σ_3-σ_4 linker.

The orientation of the -35 element DNA with respect to $\sigma^{28}{}_4$ was modeled by superimposition of $\sigma^{28}{}_4$ onto $\sigma^{A}{}_4$ of the Taq $\sigma^{A}{}_4$/-35 element DNA structure (PDB ID 1KU7; (Campbell, Muzzin et al. 2002). The resulting position of the DNA, in the context of the σ^{28} conformation in the σ^{28}/FlgM complex, reveals a steric clash with $\sigma^{28}{}_3$ (Figure 4.8B). Most of the $\sigma^{28}{}_4$ residues that correspond to $\sigma^{A}{}_4$ residues involved in -35 element recognition (Campbell, Muzzin et al. 2002) are buried in the interface of $\sigma^{28}{}_4$ with $\sigma^{28}{}_3$. The recognition helix of the $\sigma^{28}{}_4$ helix-turn-helix motif (H10) is accommodated in an acidic groove on the surface of $\sigma^{28}{}_3$ that is reminiscent of the DNA major groove (Figures 4.8B and 4.8C).

Five acidic residues exposed at the surface of the $\sigma^{28}{}_3$ groove contribute to the negative charge (Aa σ^{28}-Glu112, Asp115, Glu116, Glu117, and Glu134), two of which are conserved among σ^{28}'s (Asp115 and Glu117; Figure 3.1). Intriguingly, the acidic surface charge is a conserved feature of group 1 σ's as well. In the structure-based alignment of Taq σ^{A} with σ^{28}, Taq σ^{A} residues

Figure 4.8 Promoter binding surfaces of σ^{28} are buried in interdomain interfaces.

A) Occlusion of the -10 element binding surface of $\sigma^{28}{}_2$ by the σ_3-σ_4-linker. Elements of σ^{28} (σ_2, green ribbon; σ_3, transparent blue molecular surface; σ_3-σ_4 linker, transparent red molecular surface) are shown in the context of the σ^{28}/FlgM complex. The -10 element DNA (shown as grey phosphate backbone worms, positioned as described in the text) is shown in the context of the Taq σ^A-holoenzyme/fork-junction complex (Murakami, Masuda et al. 2002). The -10 element binding surface of σ_2 on the surface of H3 is buried in the interface with the σ_3-σ_4 linker in the σ^{28} conformation, resulting in a steric clash with the modeled DNA.

B) Occlusion of the -35 element binding surface of $\sigma^{28}{}_4$ by the σ_3. Elements of σ^{28} (σ_3, blue backbone ribbon; σ_3-σ_4 linker, red backbone ribbon; σ_4, yellow backbone ribbon) are shown in the context of the σ^{28}/FlgM complex. The -35 element DNA (shown as a molecular surface, with the nontemplate strand colored beige and the template strand colored grey, positioned as described in the text) is shown in the context of the Taq $\sigma^A{}_4$/-35 element DNA structure (Campbell, Muzzin et al. 2002). The -35 element recognition helix (H10) of the σ_4 helix-turn-helix motif fits in a groove on the surface of σ_3 in the σ^{28} conformation, resulting in a steric clash with the modeled DNA.

C) Acidic groove of σ_3. The σ_3 domains of Aa σ^{28} (left) and Taq σ^A (right; (Campbell, Muzzin et al. 2002)) are shown as molecular surfaces, color-coded according to the electrostatic surface distribution (-8 kT, red; 0, white; 8 kT, blue) calculated using GRASP (Nicholls, Sharp et al. 1991). The position of $\sigma^{28}{}_4$ in the context of the σ^{28}/FlgM complex is shown as a yellow ribbon. Conserved acidic residues discussed in the text are labeled.

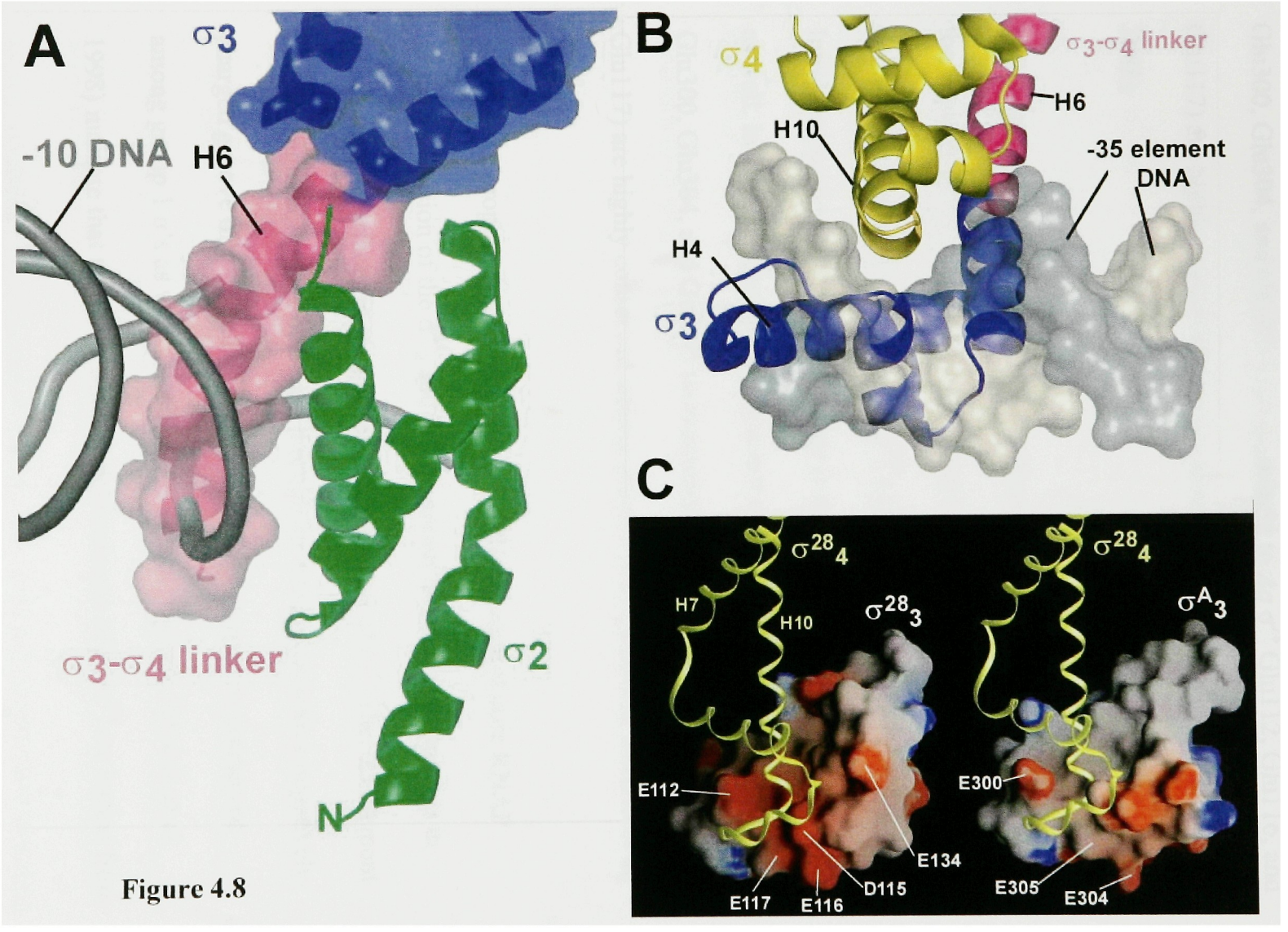

A
σ3
-10 DNA
H6
σ3-σ4 linker
σ2
N
B
σ4
σ3-σ4 linker
H6
H10
-35 element
DNA
H4
σ3
C
σ28 4
H7
H10
σ28 3
σA 3
E112
E300
E134
D115
E117
E116
E305
E304

Figure 4.8

Glu300, Glu304, and Glu305 (corresponding to *Aa* σ^{28} Glu112, Glu116, and Glu117) are highly conserved among group 1 σ's (Campbell, Muzzin et al. 2002)

Five acidic residues exposed at the surface of the $\sigma^{28}{}_3$ groove contribute to the negative charge (Aa σ^{28}-Glu112, Asp115, Glu116, Glu117, and Glu134), two of which are conserved among σ^{28}'s (Asp115 and Glu117; Figure 3.1). Intriguingly, the acidic surface charge is a conserved feature of group 1 σ's as well. In the structure-based alignment of Taq σ^{A} with σ^{28}, Taq σ^{A} residues Glu300, Glu304, and Glu305 (corresponding to *Aa* σ^{28} Glu112, Glu116, and Glu117) are highly conserved among group 1 σ's (Campbell, Muzzin et al. 2002).

Thus, we propose that σ^{28} is inhibited in the absence of core RNAP through occlusion of the $\sigma^{28}{}_2$ and $\sigma^{28}{}_4$ promoter binding determinants in interdomain interfaces in the packed conformation of σ^{28}. The -35 element binding region of $\sigma^{28}{}_4$ is occluded through an interaction with a negatively-charged groove of $\sigma^{28}{}_3$, and this characteristic of σ_3 may be conserved among group 1 σ's as well. Biochemical data of (Callaci, Heyduk et al. 1998) indicate that the DNA binding determinants of Ec σ^{70} are buried from

solvent in free σ^{70}, but exposed in the holoenzyme, lending further support to the idea that this mechanism for inhibition extends to group 1 σ's.

The structure of a free group 1 σ remains to be determined, and its precise relationship to the conformation of σ^{28} is unclear. Nevertheless, the group 1 σ DNA binding determinants could be occluded through some other arrangement of the σ domains, or even through oligomerization of the free σ's. In this scenario, one must consider that the autoinhibitory function of region 1.1 may not be direct (Camarero, Shekhtman et al. 2002), but rather region 1.1 may stabilize a conformational state of the group 1 σ's that is recalcitrant to DNA binding.

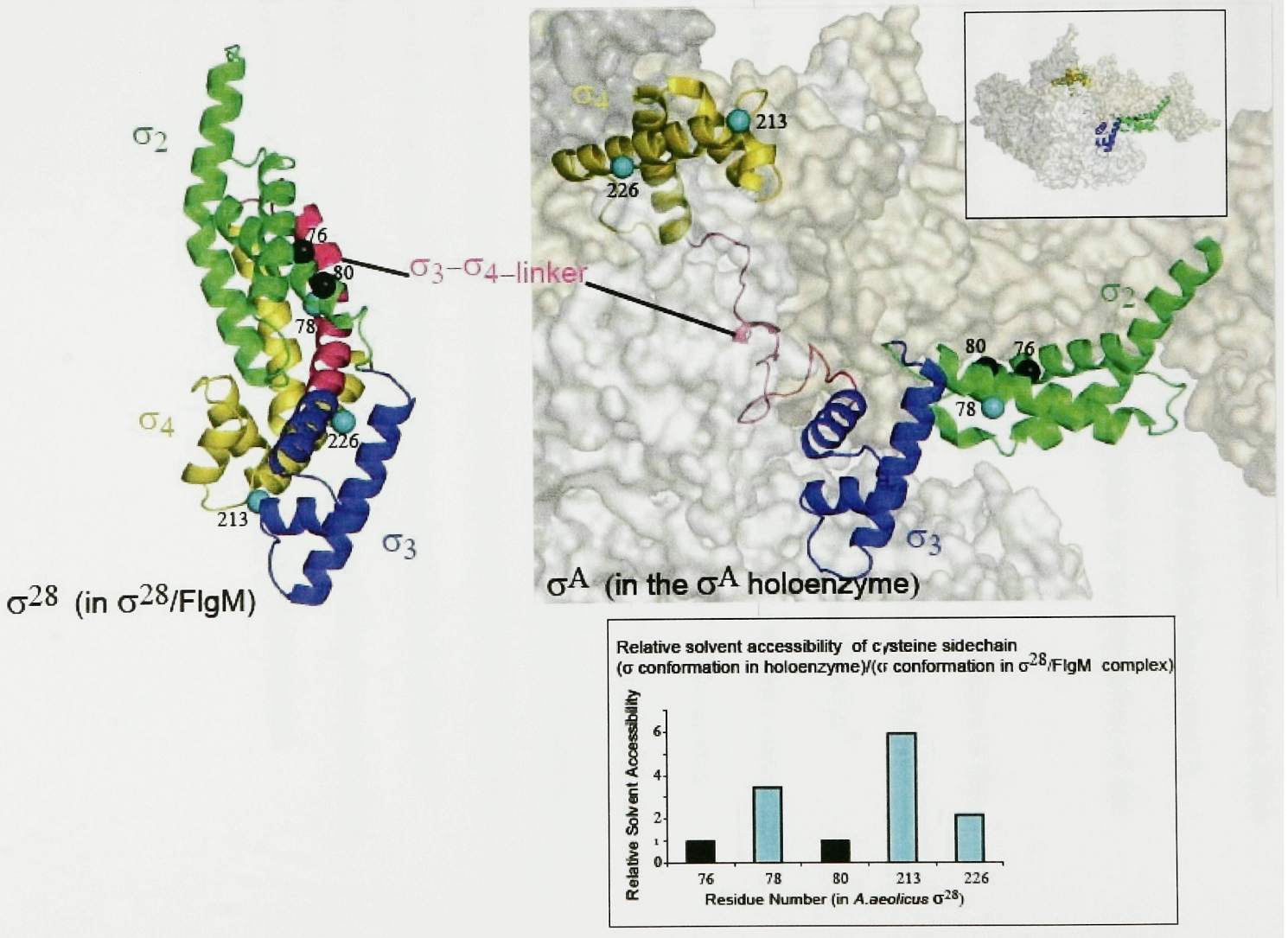
The conformation of free σ^{28} and formation of the RNAP holoenzyme

Implicit in the idea that the DNA binding determinants of σ are occluded in free σ is that these become exposed in the RNAP holoenzyme through a large conformational change. The close structural similarity between the σ_2, σ_3, and σ_4 domains of σ^{28} and σ^{A}, combined with structures of σ^{A}-holoenzyme (Murakami, Masuda et al. 2002; Vassylyev, Sekine et al. 2002), allowed us to construct a model of the σ^{28}-holoenzyme, which can be compared with σ^{28} in the FlgM complex. To construct the σ^{28}-holoenzyme model, domains $\sigma^{28}{}_2$, $\sigma^{28}{}_3$, and $\sigma^{28}{}_4$ were superimposed on the corresponding

domains of σ^A in the *Tth* (*Tth*) holoenzyme structure (Vassylyev, Sekine et al. 2002). When the σ_3 domains of σ^{28} in the FlgM complex and in the holoenzyme model are superimposed, the movements required to adopt the holoenzyme conformation can be calculated: σ_2 undergoes a rotation of 135° and a translation of 44 Å, while σ_4 undergoes a rotation of 179° and a translation of 49 Å. The separation between the centers of gravity of σ_2 and σ_4 changes from 25 Å in the FlgM complex to 66 Å in the σ^{28}-holoenzyme, an increase of 41 Å. This conformational change is qualitatively consistent with the luminescence resonance energy transfer study of (Callaci, Heyduk et al. 1999), where a significant increase in the measured distance between fluorescent probes placed in σ_2 and σ_4 was observed upon core RNAP binding, although the measured distances themselves do not correspond to the RNAP holoenzyme structures (Murakami et al. 2002a, Vassylyev et al., 2002). The σ_3-σ_4 linker could not be modeled by superposition, so to fully illustrate the conformational change in σ upon core binding, Figure 4.9 shows a comparison between the conformation of σ^{28} (in the σ^{28}/FlgM complex) and σ^A (in the *Tth* σ^A-holoenzyme; Vassylyev et al., 2002), with the σ_3 domain shown in the identical orientation to highlight the relative motions of σ_2 and σ_4 between the two conformations. The N-terminal segment of σ^A not present in σ^{28} has been omitted to facilitate the

Figure 4.9. Conformational change in σ^{28} upon holoenzyme formation. The conformations of σ in the σ^{28}/FlgM complex (top left) and in the RNAP holoenzyme (top right) are compared. The Aa σ^{28} from the σ^{28}/FlgM complex (left), and Tth σ^{A} (right) in the holoenzyme (Vassylyev, Sekine et al. 2002) are shown as backbone ribbons and color-coded as in Figure 4.1A, except for the σ_3-σ_4 linker, which is red. On the left, FlgM has been omitted; on the right, residues 74-173 of Tth σ^{A} (which have no counterpart in σ^{28}) are omitted. On the right, the core RNAP is shown as a molecular surface (β', beige; β, light grey; αI, αII, and ω, dark grey). The small inset shows the same view of the entire holoenzyme for reference. Spheres denote the α-carbon positions of residues where the reactivity of cysteine substitutions was compared in free and core-bound Ec σ^{70} (Callaci, Heyduk et al. 1998). The residue numbers are (Aa σ^{28}/Tth σ^{A}/Ec σ^{70}): (76/246/438), (78/248/440), (80/250/442), (213/393/583), (226/406/596). The positions are denoted with Aa σ^{28} numbering. The experimental results of Callaci et al. (Callaci, Heyduk et al. 1998) are denoted qualitatively by the color-coding; increased solvent accessibility in holoenzyme compared to free σ is shown in cyan, no change in solvent accessibility is shown in black. Relative solvent accessibility was calculated for cysteine substitutions modeled *in silico* at these same positions in σ^{28} in the σ^{28}/FlgM conformation and in the holoenzyme conformation (see 2.1). These results, displayed in the histogram, show good correlation with the experimental results of Callaci et al. (1998) (Callaci, Heyduk et al. 1998).

Figure 4.9


σ2
76
80
78
σ4
226
213
σ3
σ28 (in σ28/FlgM)
σ3–σ4–linker
σ4
213
226
σ2
80
76
78
σ3
σA (in the σA holoenzyme)
Relative solvent accessibility of cysteine sidechain
(σ conformation in holoenzyme)/(σ conformation in σ28/FlgM complex)
Relative Solvent Accessibility
0
1
2
4
6
76
78
80
213
226
Residue Number (in A.aeolicus σ28)

comparison. The helical σ_3-σ_4 linker of σ^{28} in the FlgM complex (red in Figure 4.9) is incompatible with the σ^{28}/core RNAP complex. The helix would severely clash with the RNAP, and its presence would also not allow the separation or orientation of σ_3 and σ_4 required in the holoenzyme. A helix-coil transition in this region is thus proposed for the formation of the σ^{28}-holoenzyme.

Using a library of single-cysteine mutants of Ec σ^{70}, changes in the degree of solvent exposure of σ^{70} regions were tested by comparing the incorporation of a cysteine-reactive probe in free and core-bound σ^{70} (Callaci, Heyduk et al. 1998). The results of this study are broadly consistent with the changes between σ^{28} in the FlgM complex and σ^{28}-holoenzyme (Figure 4.9). It is difficult to extrapolate the structure of σ^{28} to that of free σ^{70}, since the σ_3-σ_4 linker, whose structure is pivotal to this conformation, lacks any detectable sequence conservation (Figure 4.3B). It is noteworthy however, that the acidity of this segment is conserved in both σ^{28} and σ^{70}. The average calculated pI of this segment is 3.67 in primary σ's and 4.22 in σ^{28}'s, suggesting that the function of this segment may be conserved.

Conclusion

In the σ^{28}/FlgM complex, a novel conformation of the σ factor is observed in which the σ^{28} structural elements pack together in a compact unit, coalesced around the helical σ_3-σ_4 linker. The extended FlgM molecule wraps around the outside and occludes the core RNAP binding determinants on $\sigma^{28}{}_2$ and $\sigma^{28}{}_4$. Characteristics of the structure suggest how FlgM forms a ternary complex with, and destabilizes, the σ^{28}-holoenzyme. The extensive interdomain interactions within σ^{28} suggest that this conformation of σ^{28} is stable in solution even in the absence of FlgM. In this inactive conformation of the σ, the promoter binding determinants are buried in interdomain interfaces, preventing DNA binding, while the core binding surfaces are exposed, allowing σ to readily interact with core RNAP, consistent with the behavior of free σ. Formation of the holoenzyme requires a helix-coil transition in the σ_3-σ_4 linker, as well as large rearrangements of the σ domains with respect to each other, exposing the promoter binding determinants. We suggest that these principles of σ factor regulation revealed for σ^{28} may apply to group 1 σ's as well.

5. DISULFIDE CROSS-LINKING INDICATES THAT FREE AND FLGM BOUND σ^{28} ADOPT SIMILAR CONFORMATIONS

5.1 Introduction

As described in detail in Chapter 4, σ^{28} in complex with FlgM complex is fully ordered and adopts a compact conformation, in which the promoter binding surfaces of σ_2 and σ_4 are masked by σ_3 and the σ_3-σ_4 linker respectively, and where the σ_3-σ_4 linker forms an α-helical extension of the C-terminal helix of σ_3. The compactness of this conformation makes it a plausible candidate for the conformation of free σ^{28}. Since it is also consistent with biochemical studies of Ec σ^{70} (Callaci, Heyduk et al. 1998; Callaci, Heyduk et al. 1999; Sorenson, Ray et al. 2004), it may also represent a conformation adopted by primary sigma factors.

In order to test whether σ^{28} adopts this conformation in the absence of FlgM, we engineered a set of double cysteine mutants of σ^{28}, predicted to form interdomain disulfides provided σ^{28} alone adopts the FlgM bound conformation. We demonstrate that these disulfides form both in the presence and absence of FlgM, indicating a similar conformation.

Specifically we show that σ_2, σ_3 and the σ_3-σ_4 linker each can be specifically crosslinked to σ_4. Since the maximum $C_{\beta i}$- $C_{\beta j}$ for disulfide formation is 4.7 Å (Dani, Ramakrishnan et al. 2003), this demonstrates that the close interdomain interactions observed in the FlgM complex also occur in free σ^{28}. In addition, when the crosslinked residue in σ_2 is paired with a residue on the opposite face of σ_4 instead, no significant disulfide formation occurs, showing that this process is not simply the result of random, nonspecific interactions, made possible by interdomain flexibility. The disulfide connecting σ_3 and σ_4 involves a residue in the -35 binding surface of σ_4, thus implicating σ_3 in the autoinhibition of -35 binding. The cysteine connecting the σ_3-σ_4 linker with σ_4 is situated in the α-helical segment of the linker, suggesting that at least part of it is α-helical even in the absence of FlgM.

Quantification of disulfide formation of two of the mutants under equilibrium conditions indicates that the disulfide bonded species represents a major conformation. Together, these data are consistent with the proposal that the conformation of free σ^{28} is similar to that observed in the FlgM complex. This suggests that autoinhibition of DNA binding in free σ^{28} is accomplished by steric occlusion of the promoter binding surfaces as well as by a suboptimal distance between σ_2 and σ_4, and that the σ_3-σ_4 linker is at least partially α-helical in free σ^{28}.

5.2 Results

Disulfide mutant design and preparation

Four double cysteine mutants of Aa σ^{28} were constructed in a triple methionine mutant background, σ^{28}-3M (see 2.2), in order to allow detection of disulfide bonds by CNBr cleavage (CNBr specifically cleaves peptide bonds C-terminal of methionine residues). Comparison of the wild-type and σ^{28}-3M crystal structures reveals that the structure is unperturbed by the mutations (Sorenson, Ray et al. 2004). Three of the double cysteine mutants were predicted by the program MODIP (Dani, Ramakrishnan et al. 2003) to form interdomain disulfides, based on the structure of σ^{28} bound to FlgM. The σ_2-31/σ_4-202 disulfide connects σ_2 and σ_4, the σ_3-133/σ_4-218 disulfide connects the σ_3 and the -35 binding surface of σ_4, and the σ_{linker}-140/ σ_4-225 disulfide connects the σ_3-σ_4 linker segment to σ_4 (Figure 5.1A and B). σ_2-31/σ_4-211 is not predicted to form a disulfide in the FlgM bound conformation, and serves as a negative control. σ^{28}-3M and the four double cysteine mutants were each purified, both in the absence and the presence of FlgM. The FlgM binding of the mutants was unperturbed. The only mutant that differed in its behavior from σ^{28}-3M was the control, σ_2-31/σ_4-211,

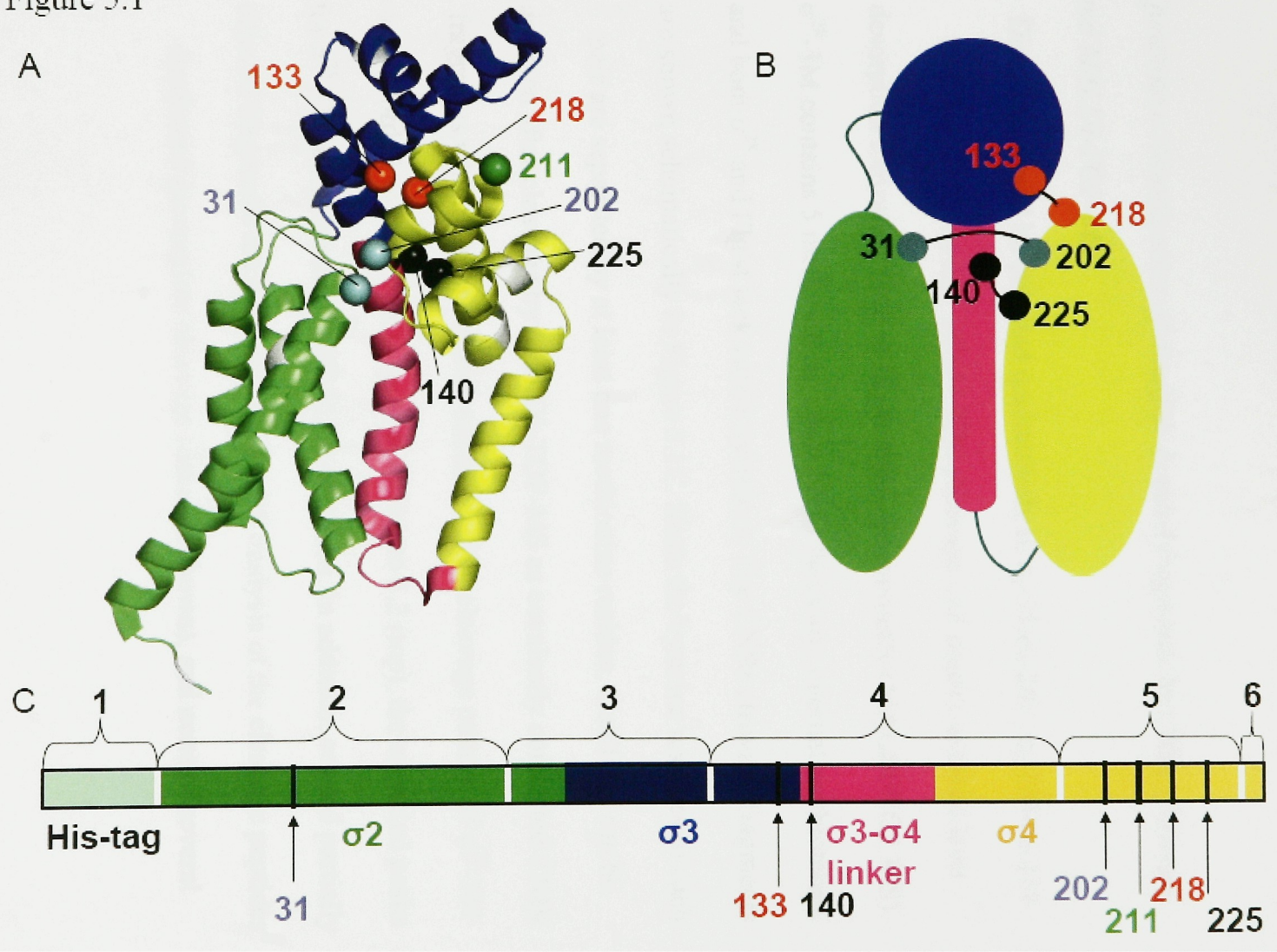
Figure 5.1. Location of introduced cysteines in σ^{28}.

A) Aa σ^{28} from the σ^{28}/FlgM co-crystal structure (1rp3, (Sorenson, Ray et al. 2004)) is shown as a ribbon. FlgM has been omitted. Structural elements are color coded as follows: σ2, green; σ3, blue; σ3- σ4 linker, pink; σ4, yellow. The C_β atoms of sites mutated to cysteine are shown as spheres, and color coded as follows: C31 and C202, grey; C133 and C218, red; C140 and C225, black; C211 (control) in green. Methionines (CNBr cleavage sites) are white.

B) Schematic diagram showing the interdomain connections made by the disulfides. The color coding is the same as in A.

C) Linear representation of σ^{28}, color coded as in A. The CNBr cleavage fragments are enclosed in numbered brackets. Cysteine sites are indicated by vertical black lines and numbered arrows. The numbers are color coded by disulfide pair as in A.

Figure 5.1


A
133
218
211
31
202
225
140
B
133
218
31
202
140
225
C
1
2
3
4
5
6
His-tag
σ2
σ3
σ3-σ4
linker
σ4
31
133
140
202
211
218
225

which was toxic to the expression host, and had a tendency to form intermolecular disulfides leading to aggregation during purification.

Qualitative identification of disulfide bonded fragments by CNBr cleavage and MALDI-TOF-MS

Disulfides could be detected qualitatively in σ_2-31/σ_4-202, σ_3-133/σ_4-218 and σ_{linker}-140/ σ_4-225 using CNBr cleavage and matrix-assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF MS). σ^{28}-3M contains 5 methionines, while FlgM lacks any internal methionines, and both σ^{28} and FlgM lack native cysteines. The CNBr cleavage fragments are shown schematically in Figure 5.1C. Since the cysteine residues of each pair are separated by at least one methionine residue, disulfides can be detected after cleavage by MALDI-TOF-MS as covalently connected CNBr fragments which disappear with reduction. The cleavage pattern of σ^{28}-3M was recorded as a reference. As seen in Figure 5.2 (top), the expected peaks for σ^{28}-3M are present in the MALDI spectrum. In addition, several partially cleaved fragments are present. SDS-PAGE analysis of the cleavage peptides confirmed incomplete cleavage (data not shown), and none of several

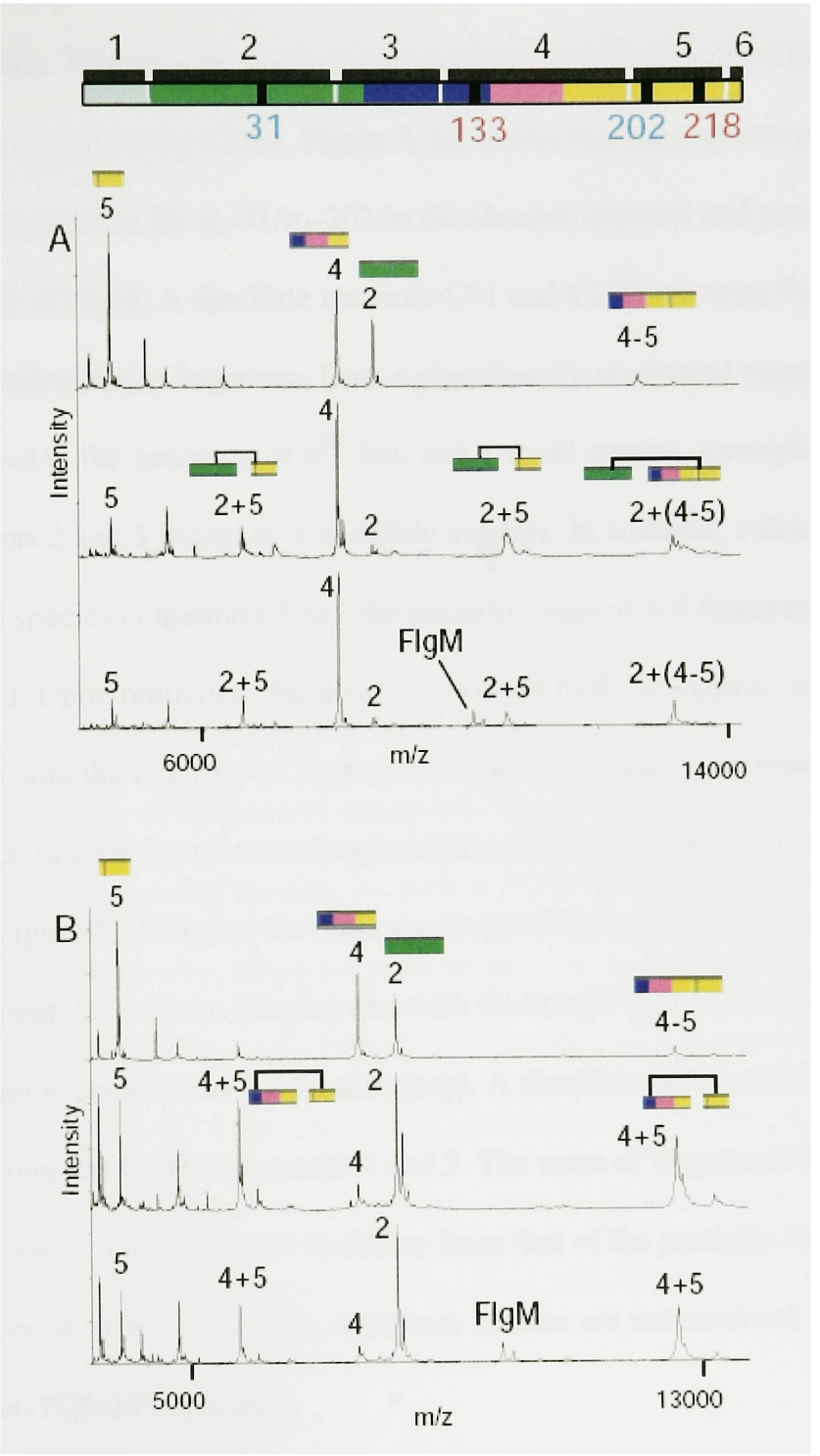
Figure 5.2. Mass spectra of CNBr digested disulfide mutants.

Top) Linear structure, color coded as in Figure 1, and with the disulfide fragments indicated by horizontal heavy black lines.

A) Cleavage fragments of σ^{28}-3M (Top), σ_2-31/σ_4-202 (Center), and σ_2-31/σ_4-202 in complex with FlgM (Bottom). Peaks are labeled with the fragment number and with colored bars corresponding to the linear diagram on top.

B) Cleavage fragments of σ^{28}-3M (Top), σ_3-133/σ_4-218 (Center), and σ_3-133/σ_4-218 in complex with FlgM (Bottom).

Figure 5.2


1
2
3
4
5
6
31
133
202
218
A
5
4
2
4-5
Intensity
4
5
2+5
2
2+5
2+(4-5)
4
FlgM
5
2+5
2
2+5
2+(4-5)
6000
m/z
14000
B
5
4
2
4-5
Intensity
5
4+5
2
4+5
4
2
5
4+5
FlgM
4+5
4
5000
m/z
13000

attempted modifications to the cleavage protocol led to complete cleavage (these included the addition of denaturants and addition of CNBr to saturation). This may be due to the presence of oxidized methionines, which are not recognized by CNBr. Figure 5.2A shows the mass spectrum of the cleavage peptides for σ_2-31/σ_4-202 in the absence (center) and presence (bottom) of FlgM. A disulfide between C31 and C202 connects fragments 2 and 5. Indeed these fragments have a significantly weakened signal as compared to the spectrum of σ^{28}-3M, and a novel species corresponding to fragments 2 and 5 linked by a disulfide appears. In addition, a disulfide bonded species containing 2 and the partially cleaved 4-5 fragment is detected. Upon reduction, the disulfide bonded peaks disappear, (data not shown), and the intensity of fragments 2 and 5 increases. The measured and expected masses for relevant fragments are displayed in Table 5.1.

Figure 5.2B shows the mass spectrum of the oxidized σ_3-133/σ_4-218 CNBr peptides without (center) and with (bottom) FlgM. The σ^{28}-3M CNBr spectrum is provided as a reference (top). A disulfide between C133 and C218 connects CNBr fragments 4 and 5. The mass of fragments 4 and 5 connected by a disulfide (4+5) differs from that of the partially cleaved fragment (4-5) by only 16 Da, and these species are not resolved in the MALDI-TOF-MS spectrum

Table 5.1 Measured and Calculated Mass of CNBr Fragments.

Mutant	Fragment # (residues)	Calculated Mass (Da)	Observed Mass (Da)	ΔMass (Da)
σ_2-31/σ_4-202	2 (2-77)	8,574.1	8,573.3	-0.8
	5 (192-231)	4,588.5	4,588.3	-0.2
	2+5	13,160.6	13,162	1.4
σ_3-133/σ_4-218	4 (123-191)	8,041.1	8,040.3	-0.8
	5 (192-231)	4598.5	4598.1	-0.4
	4+5	12,637.6	12,637.8	0.2
σ_{linker}-140/ σ_4-225	4 (123-191)	7991.1	7990.5	-0.6
	5 (192-231)	4572.4	4572.3	-0.1
	4+5	12,561.5	12,562.0	0.5

Since their size and amino acid compositions are identical, however, the increase in the peak height for this species in the σ_3-133/σ_4-218 spectrum compared to the σ^{28}-3M can be interpreted quantitatively as the appearance of the disulfide bonded species. As seen in 5.2.B (center), this peak height increases significantly, while the peak heights for the individual fragments, 4 and 5 are reduced. The σ_{linker}-140/ σ_4-225 disulfide also connects fragments 4 and 5, and the MALDI spectra for this mutant are very similar to those of σ_3-133/σ_4-218 (data not shown). Due to the presence of intermolecular disulfides in the control mutant, σ_2-31/σ_4-211, MALDI-TOF analysis was not expected to give an informative result.

Quantitative disulfide analysis by SDS-PAGE

The mobility of each disulfide mutant on a denaturing, non-reducing gel increased significantly upon disulfide formation, allowing resolution and quantitative analysis of the disulfide bonded versus the reduced species for each disulfide mutant. The size of the shift is proportional to the distance between the two cysteines, with the longest range mutant, σ_2-31/σ_4-202, giving rise to the largest shift. As seen in Figure 5.3, all disulfide mutants migrate with the same mobility as σ^{28}-3M when reduced, but shift downward when oxidized, with the longest range mutant (σ_2-31/σ_4-202) giving rise to the largest shift. In the absence of FlgM, disulfide formation is essentially complete for all three disulfide mutants. σ_2-31/σ_4-202 forms disulfide as readily in the presence of FlgM. The disulfide bonded species predominate for both σ_3-133/σ_4-218 and σ_{linker}-140/ σ_4-225 even in the presence of FlgM, but significant amounts of reduced protein is still detectable. The control mutant, σ_2-31/σ_4-211, does not form significant amounts of intramolecular disulfide regardless of FlgM, although a very faint band corresponding to this species is visible in Figure 5.3B, in the absence of FlgM.

Figure 5.3

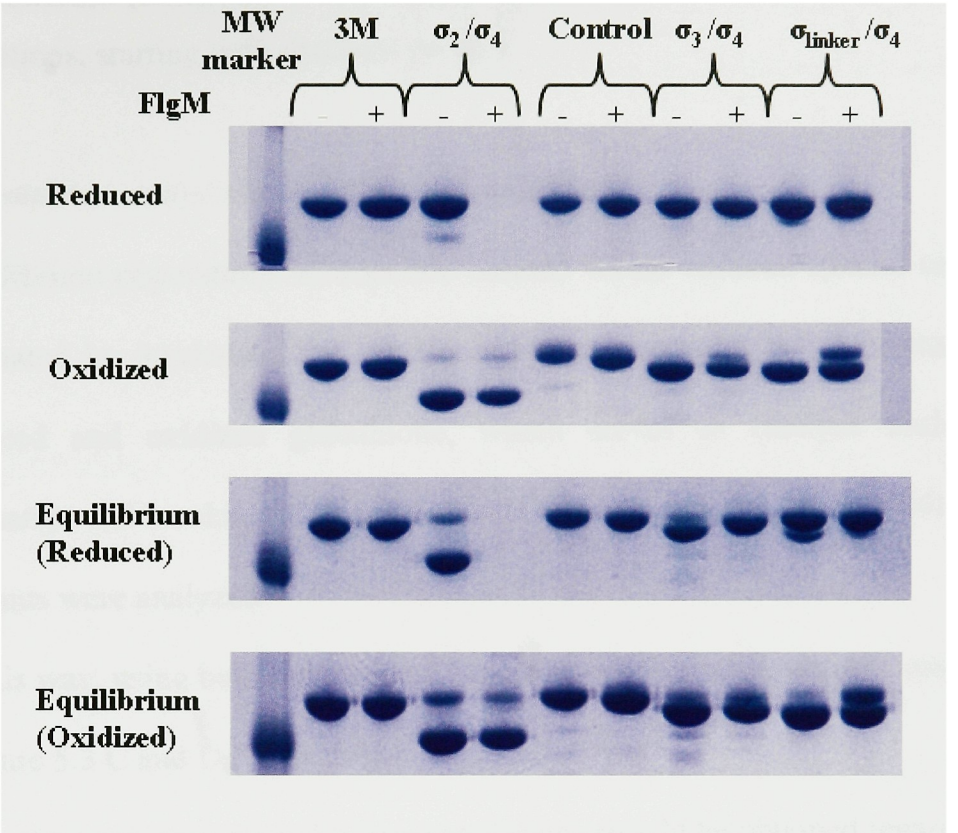


Figure 5.3. SDS-PAGE analysis of disulfide mutants

Intramolecular disulfide bonded versus reduced σ^{28} were separated on a 10% NuPAGE® Novex Bis-Tris denaturing gel. See also Table 5.3 for relative integrated quantities. **Reduced**: After reduction with DTT. σ^{28}-3M and the 4 double cysteine mutants are each shown with and without FlgM as indicated. The names have been abbreviated as follows: 3M=σ^{28}-3M, σ_2/σ_4 = σ_2-31/σ_4-202, σ_3/σ_4= σ_3-133/σ_4-218, σ_{linker}/σ_4 = σ_{linker}-140/ σ_4-225 and control= σ_2-31/σ_4-211. **Oxidized:** Samples (as above) that were allowed to oxidize during purification. **Equilibrium (reduced)**: After incubation under glutathione equilibrium conditions, starting with reduced protein.

Equilibrium (oxidized): After incubation under glutathione equilibrium conditions, starting with oxidized protein.

Quantitative analysis of equilibrium populations

Equilibrium populations of disulfide bonded versus reduced species can be estimated by incubating the sample in buffer containing a 1:1 mixture of reduced and oxidized glutathione, which serves to catalyze both the formation and breaking of disulfide bonds (Dorigo, Schalch et al. 2004). The mutants were analyzed

in this way, using both reduced and oxidized samples as the starting material (Figure 5.3 C and D, Table 5.3).

Since the same equilibrium populations should be obtained regardless of the oxidation state of the starting material, this served as a control for buried sites. For σ_2-31/σ_4-202, the protein was mostly oxidized both in the presence and absence of FlgM, and this result did not depend on the oxidation state of the starting material. Similarly, the disulfide bonded species predominated for σ_3-133/σ_4-218 in the absence of FlgM, both when starting with reduced and oxidized protein. In the presence of FlgM however, the amount of oxidation of this mutant was similar to that of the starting material, suggesting that the cysteines were inaccessible to the catalyst. The same was true for the σ_{linker}-140/ σ_4-225 mutant with or without

FlgM, and consequently these were not amenable to this type of analysis. No disulfide formation was detected upon glutathione incubation in the control mutant,σ_2-31/σ_4-211.

5.3 Discussion

We have demonstrated that engineered, long-range, interdomain disulfide bridges predicted to form in σ^{28} bound to FlgM, also form in free σ^{28}, indicating that the solution conformation of σ^{28} is similar to that observed in the crystal structure of the σ^{28}/FlgM complex. This conformation differs significantly from the σ^{A} holoenzyme structures, in which σ^{A} adopts an elongated conformation, where the domains are spread across one face of the enzyme and each are engaged in interaction with the core RNAP, while making no interdomain contacts. The linker segment is mostly unfolded, and passes from σ_3 through the interior of the enzyme near the active site, and threads through the RNA exit channel to connect with σ_4. It is reasonable to assume that σ^{28} adopts a similar holoenzyme conformation, since the structures of the individual domains of Aa σ^{28} are very similar to those of Taq σ^{A}, and since both recognize -10 and -35 promoter elements, thus requiring a similar interdomain spacing. The linker of σ^{A}, however, bears no structural resemblance to the α-helical linker

conformation in the σ^{28}/FlgM complex. Since detectable sequence conservation is also absent, and since a similar function has not been established, it is not possible to predict whether the structural differences are strictly context dependent, or if they are innately different. Since it is reasonable to assume that σ_3 and σ_4 bind to the holoenzyme in the same manner in both σ factors however, the structure of the linker segment connecting these is constrained by their positions, and modeling suggests that a helical conformation is incompatible with such positioning.

We have directly shown that σ_2 and σ_4 can be disulfide bonded in the free σ^{28}, and equilibrium measurements suggest that this represents the favored conformation. The maximum distance between the Cα atoms of two cysteines to form disulfides is 7Å (Dani, Ramakrishnan et al. 2003), and in the σ^{28}/FlgM complex the Cα atoms of R31 and P202 are 5.36Å apart. In the Taq holoenzyme (Murakami, Masuda et al. 2002), the corresponding residues, G219 and T397, are 63.26 Å apart, appropriately spaced for -10 and -35 recognition (Table 5.2).

The close proximity of these domains in free σ^{28} is thus incompatible with simultaneous recognition of these promoter elements, providing one explanation for autoinhibition, and suggesting that these residues should move apart by ~58Å upon binding to core RNAP.

Table 5.2 Distances between mutated residues in the FlgM and holoenzyme bound conformations.

Residue Pair		C_α- C_α Distance (Å)	
Aa σ^{28}	Taq σ^A	σ^{28} (FlgM complex)	σ^A (holoenzyme)
K31/P202	G219/T397	5.36	63.26
L133/S218	L323/R413	6.56	57.08
Y140/L225	V330/L420	5.93	51.21

A study using luminescent resonance energy transfer between introduced probes has shown that the distance between σ_2 and σ_4 increases by 15Å in Ec σ^{70}(Callaci, Heyduk et al. 1999). Although the direction of the movement is consistent with this, our data suggest a greater magnitude. Alternatively, the average distance between these domains in free σ may be greater than that trapped by the disulfide, giving rise to a smaller net movement upon binding to core, but the equilibrium measurements argue against this. Another possibility is that the conformation of Ec σ^{70} is altogether different from that of σ^{28}. Alternatively, energy transfer methods may be useful in comparing relative distance, but not in determining absolute distances. Comparison of the C_α- C_α distances between the Taq σ^A residues corresponding to the other

two disulfide pairs with those in the σ^{28}/FlgM complex also reveals large differences in the two conformations (Table 5.2).

We have also shown that σ_3 and σ_4 can be disulfide crosslinked in the σ_3-133/σ_4-218 mutant, and that this disulfide is favored at equilibrium in free σ^{28}. The larger proportion of disulfide bonded species in free σ^{28}, indicates that disulfide formation in this mutant is more rapid in the absence of FlgM (Figure 5.3 B). This suggests that the proximity of the domains is similar, but that rigidity imposed by FlgM decreases the rate of disulfide formation. Under equilibrium conditions, mostly oxidized protein was obtained for free σ^{28}, regardless of whether the protein was reduced or oxidized at the beginning of the experiment (Figure 5.3 C and D, Table 5.3).

For the FlgM bound species, by contrast, the population after incubation under equilibrium conditions resembled the starting material, suggesting that the site was inaccessible to the catalyst (Figure 5.3 C and D). FlgM does not directly occlude this site in the structure, but it may constrain interdomain flexibility which otherwise would allow access to the catalyst. S218 of Aa σ^{28} corresponds to R413 of Taq σ^{A} which has been observed in a co-crystal structure to interact with the -35 element (Campbell, Muzzin et al. 2002).

Table 5.3 Relative disulfide populations at equilibrium.

Mutant	FlgM	Starting Population	% Disulfide	Std. dev.(%)
σ_2/σ_4	-	Red	87.9	3.6
		Ox	89.5	1.4
	+	Red	76.1	1.1
		Ox	99.1	1.5
σ_3/σ_4	-	Red	74.3	4.1
		Ox	82.4	2.6
	+	Red	27.8	0.5
		Ox	74.6	9.4
σ_{linker}/σ_4	-	Red	17.2	1.9
		Ox	88.7	2.7
	+	Red	4.9	2.8
		Ox	71.2	0.5

The table shows the relative populations of reduced and disulfide bonded species after incubation under equilibrium conditions (see Experimental Procedures). The % Disulfide values are averages of gel densitometric measurements performed in triplicate.

Although the -35 binding surface of σ^{28} has not been established, it is reasonable that σ^{28} binds to the -35 element using the same surface of σ_4. Interaction with L133 of σ3 thus buries the -35 binding surface, directly occluding it, and providing a role for σ_3 in auto-inhibition. L133 is located on an acidic face of σ_3, and the acidity of this surface is conserved in primary σ factors, as noted previously (Sorenson, Ray et al. 2004). Thus this mode of autoinhibition may be more general. Since the interaction between

1.1 and σ_3 has been shown to be indirect, one possibility is that 1.1 acts indirectly, by stabilizing an interaction between σ_3 and σ_4, similar to that observed in σ^{28}.

Finally, we have demonstrated using mutant σ_{linker}-140/ σ_4-225 that a cysteine introduced into the σ_3-σ_4 linker segment crosslinks to a cysteine in the C-terminal helix of σ_4, both in the FlgM complex and in free σ^{28}. In the holoenzyme, the corresponding part of the linker is unfolded. Since the crosslink requires this residue to be on the appropriate side of the linker helix, its formation supports the idea that it is α-helical even in solution. The lower portion of the helix may still be unfolded in free σ^{28}, but since formation of this and the σ_3-133/σ_4-218 disulfides also require that the orientation of σ_4 is similar to that in the FlgM complex, both to σ_3 and to the linker, the conformation of this segment is constrained. In the FlgM complex, the linker segment sterically occludes the -10 binding surface of σ_2. Although crosslinking between these two elements is not directly shown in this study, the fact that both σ_2 and the linker crosslink to σ_4 in the same way as in the FlgM complex, suggests that they are also in a similar orientation with respect to each other, thus indirectly implicating the linker segment in the autoinhibition of -10 binding. The disulfide formation was

more rapid in free σ^{28} than in the FlgM complex for this mutant as well, again suggesting greater flexibility but a similar conformation of free σ^{28}.

Together these data suggest that the conformation of free σ^{28} resembles that observed in the FlgM complex, but that it is less rigid. This conformation provides an explanation for the autoinhibition of DNA binding and provides insight into the conformational change that accompanies binding to RNAP. Auto-inhibition is accomplished both by suboptimal interdomain distances and by steric occlusion of the -35 binding surface, and possibly also of the -10 binding surface. These results may extend to primary sigma factors as well.

APPENDIX 1: EXPRESSION, PURIFICATION AND ACTIVITY OF RECOMBINANT RNA POLYMERASE FROM *AQUIFEX AEOLICUS*

A1.1 Introduction

The RNA polymerase (RNAP) from Aa was cloned and purified for use in crystallographic studies. Although its overall structure is expected to be similar to the previously observed Taq and Tth structures, the gram (-) Aa is more closely related to Ec than the *Thermus* species, and has σ^{54} and σ^{28} factors, which are absent in *Thermus*. Thus, it is useful for crystallization of the holoenzymes of these alternative sigma factors. In addition, it contains insertions homologous to those in the Ec polymerase, which are also absent in Thermus, and given that it is more similar to the Ec polymerase, it may also bind other ligands which bind to the Ec but not the Taq polymerase. The structure of a chimeric holoenzyme consisting of Aa σ^{28} and Taq core RNAP was attempted previously (Murakami 2002). The complex yielded crystals which diffracted to ~6Å and a map of good quality could be obtained by molecular replacement using the Taq RNAP as a search model. This map contained clear density for σ^{28}, and σ_2 and σ_3 could be identified by docking

previously determined domain structures. These domains were bound in a completely different fashion than observed in Taq σ^A, with σ_2 approximately bound in the σ_3 binding site and the σ_2 binding site occupied by σ_3. Since the transcriptional activity of this chimeric enzyme is also very poor, we suspected that the structure was not biologically relevant, and decided to attempt the structure of the Aa holoenzyme instead. Recombinant expression of Aa RNAP was necessary, since the strain is not available. Here we present the expression, purification and activity of the recombinant enzyme.

A1.2 Experimental Procedures

Expression

All subunits of the Aa RNAP were recombinantly expressed on a single pET28 derivative plasmid, pET28-rpoABZCAa, with a C-terminal hexahistidine tag on the β' subunit (Severinov 2004). The subunits were expressed in BL21-CodonPlus® (DE3)-RIL (Stratagene, La Jolla, CA). The cells were shaken in LB at 37°C for 20 hrs without induction, as leaky expression gave sufficient quantities. Induction with IPTG actually decreased the overall yield, since this arrested cell growth.

Purification

Scheme 1

Cells were pelleted after expression, resuspended in Buffer A (20 mM Tris pH 8, 5% Glycerol, 0.5 M NaCl, 1mM β-ME) and protease inhibitor cocktail (Sigma), and lysed with a French Press. The clarified lysate was incubated at 75°C for 35 min. to precipitate heat-sensitive Ec proteins. The precipitated fraction was removed by centrifugation, and the supernatant was loaded on a chelating HiTrap column, precharged with Ni^{2+}, washed with 20% Buffer B (20 mM Tris pH 8, 5% Glycerol, 0.5 M KCl, 100 mM Imidazole, 1mM β-ME), and eluted with 100% Buffer B. The eluted fraction was further purified by gel filtration (Superdex 200) in TGED, 0.5 M NaCl (1L).

Scheme 2

Cells were pelleted after expression, resuspended in Buffer A' (TGED, 1 mM EDTA, 50 mM NaCl) and lysed and heat purified as in Scheme 1. Protease inhibitor cocktail (Sigma) to 1X and a home-made protease inhibitor cocktail (174 μg/mL PMSF, 312 μg/mL Benzamidine, 5 μg/mL Chymostatin, 5 μg/mL Leupeptin, 1 μg/mL Pepstatin and 10 μg/mL Aprotinin) were both added to the resuspension buffer. The supernatant was

loaded on a 60 ml Q column, washed with 15% Buffer B', and eluted with a gradient from 15-47% Buffer B' (TGED 1 mM EDTA/1M NaCl). The protein eluted in wash and that eluted during the gradient were purified further separately on heparin (eluted with a gradient from 0-100% B') and gel filtration (Superdex 200) (Figure A1.2)

Scheme 3

Cells were pelleted and heat purified as in scheme 2. The heat purified fraction was loaded on a heparin Hi-trap column, washed with 13% B' until baseline, and eluted with linear gradient from 13% B' - 100% B'. The eluted fraction was dialyzed against TGED, 20 mM NaCl, which caused it to precipitate. The protein rapidly redissolved when the NaCl concentration was increased to 0.1 M NaCl. It was loaded on a Q Hitrap column, eluted with a100 ml TGED gradient from 20 mM to 1M NaCl. The polymerase peak fractions were pooled and diluted 12/50 with TGED without salt to NaCl final of < 60 mM and loaded on the Q column again. The protein was eluted with a shallow TGED gradient from 20mM to 0.5M NaCl. The eluted peak was concentrated and purified by gel filtration (Superdex 200).

Abortive transcription assay

The assay was performed as described previously (Campbell, Muzzin et al. 2002), except that the temperature was 65°C. In the reactions indicated, the *fliC* promoter was used in place of T7 A1. In the reactions containing *fliC*, the dinucleotide primer was ApC and the radioactive substrate was [α-P^{32}]ATP. The *fliC* promoter was synthesized based on the published sequence (Deckert, Warren et al. 1998; Studholme and Buck 2000).

Crystallization

Needle-like crystals were reproducibly obtained of protein from Scheme 2, wash fraction. 1 µl protein solution (in 20 mM Tris-HCl [pH 8.0], 20 mM NaCl, 2 mM DTT) was mixed with 1 µl of precipitant solution [0.1M KCl, 5 mM magnesium sulfate, 50 mM HEPES pH 7.0, 5% MPD] and allowed to equilibrate against 1 ml precipitant solution solution .

A1.3 Results & Discussion

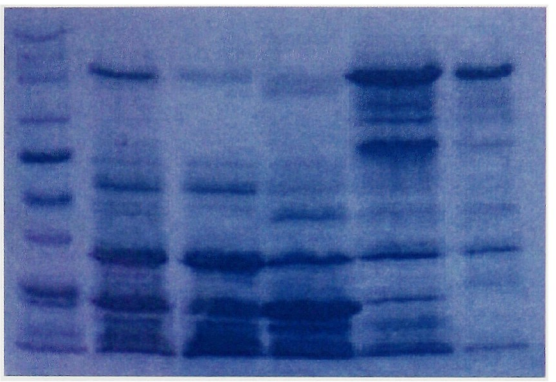
Here we describe the expression and purification of recombinant Aa RNAP. The purified protein was active in transcription assays together with

Aa σ^{28}. In addition, needle-like crystals of the core RNAP were obtained from one of the purified batches of protein.

The subunits were expressed from a single plasmid by leaky expression. Although induction with IPTG also resulted in expression, it caused the cells to stop growing, and gave a lower total yield. The α-subunit was expressed in large excess compared to the larger subunits, and was removed during purification. If greater stoichiometric balance could be achieved during expression, this may improve the overall yield of the function RNAP.

Three basic purification schemes were attempted. Scheme 1 utilizes binding of the hexahistidine tag on the β'-subunit to a nickel chelating resin. Although this removes a large amount of Ec proteins as well as excess α and ω subunits, the binding to the resin was rather poor, and a significant portion of the RNAP was found in the flow-through and wash fractions (Figure A1.1a). In addition, use of nickel precludes the use of EDTA in order to inhibit proteases. The final product of Scheme 1 was also significantly contaminated with other proteins (Ec proteins or degradation products), despite purification by gel filtration and ion exchange following the nickel step (Figure A1.1b)

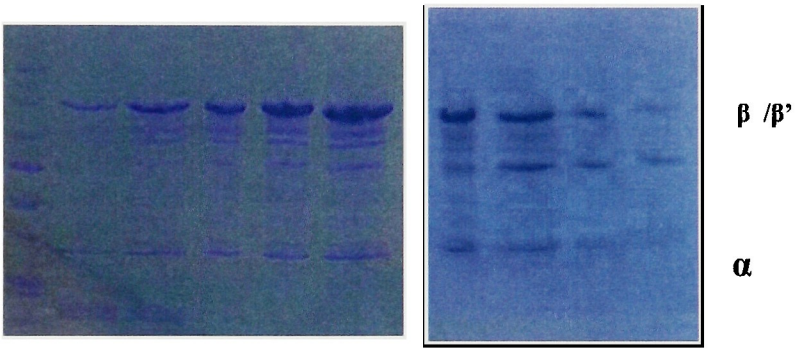
a)



β (167 kDa) /β' 178 kDa)

α

b)



β /β'

α

Figure A1.1. Aa RNAP purificiation scheme 1

a) Heat and nickel purification of Aa RNAP. From left to right the lanes contain 1: MW marker, 2: heat purified RNAP, 3: Nickel flow-through, 4: Nickel wash, 5: Nickel eluate fraction 1, 6: Nickel eluate fraction 2.b) RNAP peak eluted from the gel filtration column. The gel contains peak fractions in order of collection from left to right.

Scheme 2 utilized Q and Heparin columns together, and 1 mM EDTA was included in order to inhibit proteases. The binding to the Q column was

rather poor, resulting in a large portion of the RNAP in the wash fraction, and this was further purified separately from the fraction of RNAP which bound to the Q column. The wash fraction (Figure A1.2) reproducibly formed needle-like crystals (see A1.2 Experimental Procedures).

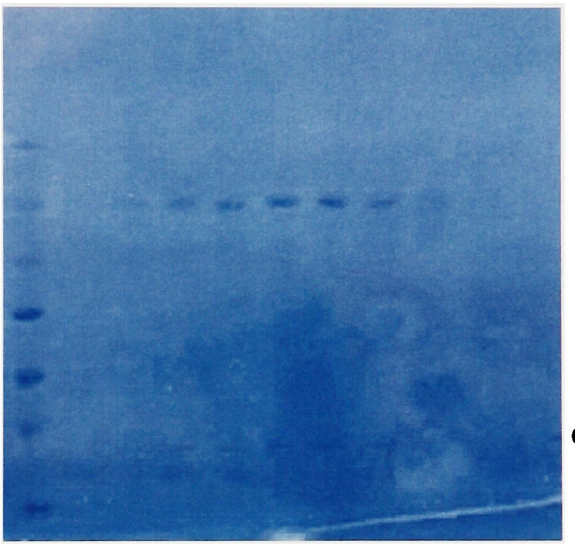


Figure A1.2 Aa RNAP purificiation scheme 2.
Gel filtration fractions of the part of the protein that eluted from the Q column in the wash.

The Q column used in this purification was regenerated prior to the next purification, and very little protein was obtained in the wash fraction when the purification was repeated. Thus, the crystallizable protein batch could not be reproducibly prepared, and no other batch of protein yielded crystals (including the product of Scheme 3).

Scheme 3 utilizes the heparin column as the first column, since the polymerase binds strongly to this column. This was followed by ion exchange using a Q HiTrap column. A shallow salt gradient removed a major contaminant, and this purification scheme gave the most pure product of the schemes attempted (Figure A1.3).

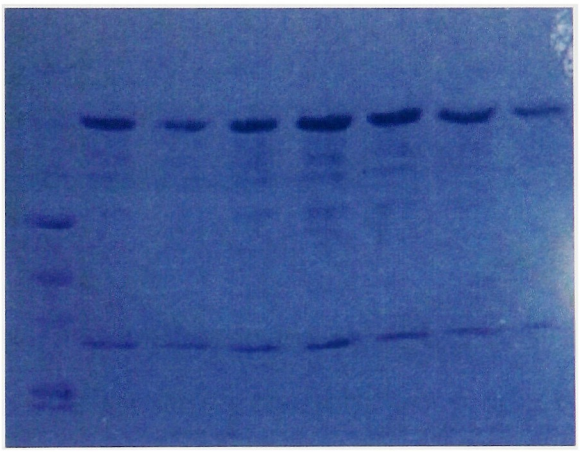


Figure A1.3 Aa RNAP purificiation scheme 3
Gel filtration fractions

Finally, activity of the purified RNAP was demonstrated in abortive transcription assays (Figure A1.4). The polymerase was tested with both Aa σ^{28} and Taq σ^{A} and on both the σ^{28} dependent fliC promoter and the σ^{A} dependent T7 A1 promoter. Taq core polymerase was used as a control. As seen in Figure A1.4, the only active combination was Aa RNAP with Aa σ^{28}.

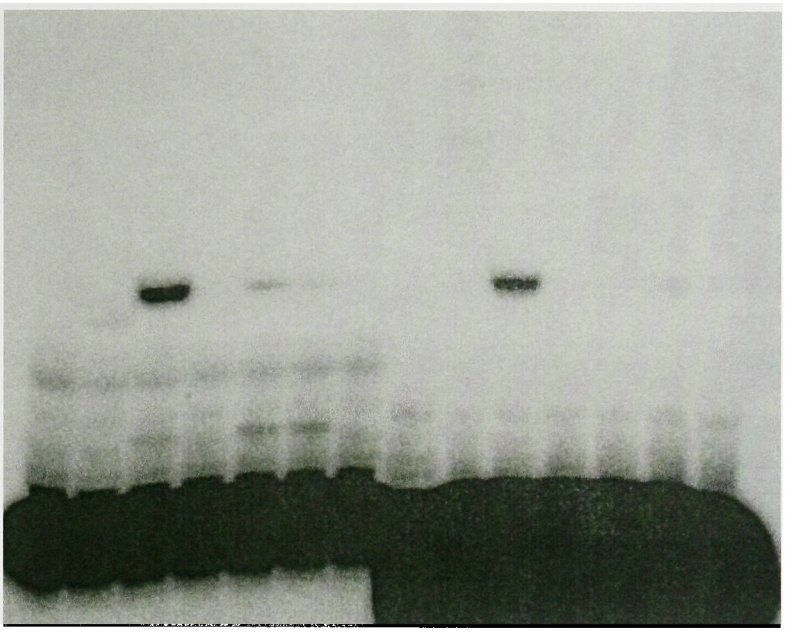


Figure A1.4 Activity of recombinant Aa RNAP

An autoradiogram of the abortive transcription products is shown. Lanes 1-7 all contain the σ^{28} dependent fliC promoter and [α-P^{32}]ATP . The lanes from left to right contain 1: no protein, 2: Aa core RNAP, 3: Aa core RNAP + Aa σ^{28}, 4: Aa core + Taq σ^{A}, 5: Taq core RNAP + Aa σ^{28}, 6: Taq core RNAP + Taq σ^{A}, 7: Aa σ^{28}. Lanes 8-14 contain the σ^{A} dependent T7 A1 promoter and [α-P^{32}]CTP. The protein composition in 8-14 is the same as in 1-7.

This activity was seen both on the fliC and the T7 A1 promoters, suggesting that it is non-specific (possibly due to single stranded DNA contamination). It does nevertheless demonstrate that the catalytic activity of the recombinant polymerase is intact. Since Taq RNAP with σ^{A} was not

active on the T7 A1 promoter, as would be expected, the protein used may have been of insufficient quality.

In summary, we have demonstrated that pure, recombinant, catalytically active Aa RNAP can be obtained in sufficient quantities for crystallization trials. Further studies are required in order to obtain x-ray quality crystals both of the core RNAP and of complexes with σ^{28} and σ^{54}.

APPENDIX 2: ASSESSMENT OF INTRAMOLECULAR DISULFIDE FORMATION BETWEEN σ_3 AND σ_4 OF E.COLI σ^{70}.

A2.1 Introduction

The conformation of free σ^{28} described in chapter 5 is consistent with biochemical observations for free Ec σ^{70}. In order to test whether Ec σ^{70} adopts a conformation similar to that of σ^{28}, we constructed a double cysteine mutant of Ec σ^{70} analogous to σ_3-133/σ_4-218 of Aa σ^{28}. This mutant, $\sigma^{70}{}_3$-498/ $\sigma^{70}{}_4$-588 was purified and assayed for disulfide formation.

A2.2 Experimental Procedures

Mutagenesis

The mutants were constructed in a background where the three native cysteines had been mutated to serine (Callaci, Heyduk et al. 1998). This mutant was cloned into a pET28 derivative vector with an N-terminal hexahistidine tag, and two novel cysteines, L498C and R588C were

introduced by site-directed mutagenesis. These sites correspond to Aa Sigma-28 residues 133 and 218, based on sequence alignment. The expected sequence was confirmed by sequencing.

Protein expression and purification

The protein was transformed into BL21 codon(+) RIL and protein expression was induced at mid log phase with 1 mM IPTG. After induction the temperature was lowered from 37 to 19 degrees and the cells were harvested by spinning after 17 hrs. The cells were purified as described for

Disulfide detection and equilibrium analysis.

DTNB analysis (Riddles, Blakeley et al. 1983) did not reveal the presence of free cysteines. Separation on a 7% NuPAGE® Novex Tris Acetate gel (Invitrogen) revealed a mobility shift upon the addition of reducing agent, indicating the presence of intramolecular disulfide bridges (Figure A2.1). Equilibrium analysis was performed as in section 2.2 and this resulted in the formation of intramolecular disulfides (Figure A2.2).

A2.3 Results & Discussion

A double mutant analogous to σ_3-133/σ_4-218 of Aa σ^{28} (section 2.2) was constructed in Ec σ^{70} in a cysteine-free background (Callaci, Heyduk et al. 1998). The double mutant was over-expressed and purified to homogeneity. Size exclusion chromatography revealed that 68% of the protein existed as large molecular weight aggregates, and SDS-PAGE of the aggregated protein in the absence and presence of reducing agent revealed that a large portion of the aggregates contained intermolecular disulfides. This indicates either that one or both cysteines are solvent exposed, or that unfolding occurred. This is in contrast to σ_3-133/σ_4-218 where only a minor portion of the protein aggregated when purified in the same manner (see section 2.2). 27% of the protein eluted from the gel filtration column as a dimer, and only a small portion of this fraction contained intramolecular disulfides, while most of it contained the expected intramolecular disulfide between C498 and C588 (Figure A2.1).

Incubation of this fraction under equilibrium conditions where disulfide bond formation and breaking are catalyzed gave mostly intramolecular disulfide (Figure A2.2). This is in contrast to σ_3-133/σ_4-218 where the same experiment gave mostly intramolecular disulfide. Although these results indicate that formation of a disulfide between C498 and C588 is

possible, it does not show that this is a favored species as with σ_3-133/σ_4-218. It is possible that the intramolecular disulfides form in misfolded protein, and that this protein has a less stably folded structure than σ^{28}.

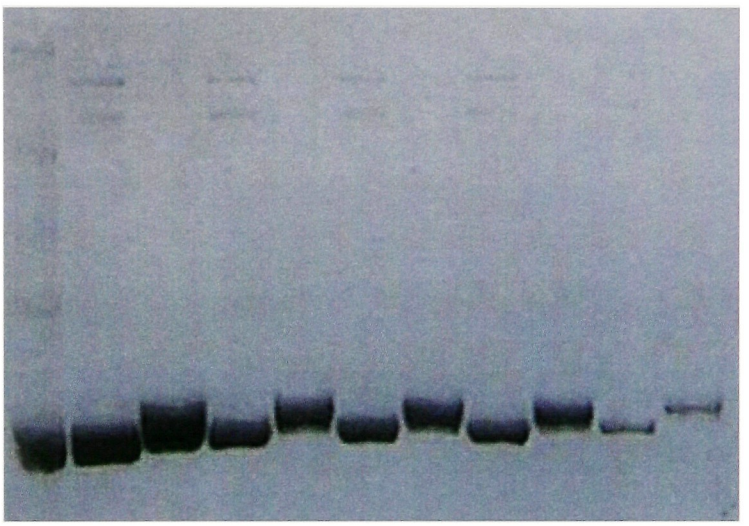


Figure A2.1 Intramolecular disulfide formation in Ec σ^{70}

The non-aggregated protein fraction contained the expected intramolecular disulfide, as detected by a gel shift in the absence of reducing agent on a nonreducing, denaturing gel. The gel (left to right) shows MW marker, followed by alternating lanes of oxidized and reduced protein (starting with oxidized), with decreasing protein concentration from left to right.

Incubation of this fraction under equilibrium conditions where disulfide bond formation and breaking are catalyzed gave mostly intramolecular disulfide (Figure A2.2). This is in contrast to σ_3-133/σ_4-218 where the same experiment gave mostly intramolecular disulfide. Although

these results indicate that formation of a disulfide between C498 and C588 is possible, it does not show that this is a favored species as with σ_3-133/σ_4-218. It is possible that the intramolecular disulfides form in misfolded protein, and that this protein has a less stably folded structure than σ^{28}.

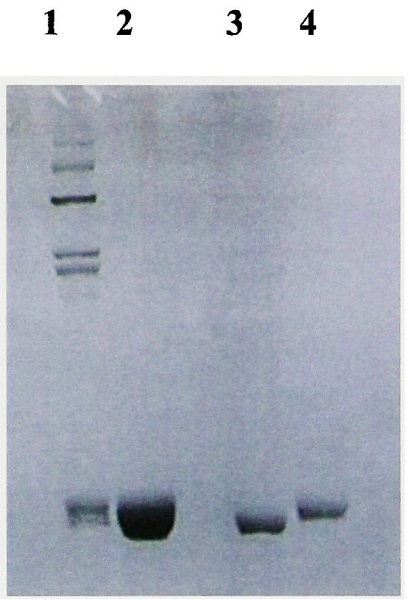


Figure A2.2 Disulfide formation under equilibrium conditions
1: protein after incubation under disulfide equilibrium conditions, 2: reduced protein (overloaded), 3: oxidized protein , 4: reduced protein

Although the same conformation may be adopted in σ^{70} as observed in σ^{28}, it appears to be much more transient and less stable. Another possible explanation for the difference in behavior compared to σ_3-133/σ_4-218 is that the sequence alignments used to determine the sites of cysteine introduction

is unreliable. Residue 588 is situated in a highly conserved part of σ_4, and the alignment in this region is likely to be reliable. Aa σ^{28}residue 133 however falls in a poorly conserved segment. 498 is the corresponding residue in Ec σ^{70} based on a structural alignment of Aa $\sigma^{28}$and Taq σ^{A}, and a sequence alignment of primary sigma factors. A large Clustal W alignment of $\sigma^{28}$and σ^{70} from different species aligns residue 133 of Aa $\sigma^{28}$with 496 of Ec σ^{70}. The certainty of this assignment is thus low, and if incorrect would provide an alternative explanation for the difference in behavior between the two homologs.

BIBLIOGRAPHY

(1994). "The CCP4 suite: programs for protein crystallography." Acta Crystallogr D Biol Crystallogr **50**(Pt 5): 760-3.

Adams, P. D., N. S. Pannu, et al. (1997). "Cross-validated maximum likelihood enhances crystallographic simulated annealing refinement." Proc Natl Acad Sci U S A **94**: 5018-5023.

Arthur, T. M., L. C. Anthony, et al. (2000). "Mutational analysis of beta '260-309, a sigma 70 binding site located on Escherichia coli core RNA polymerase." Journal of Biological Chemistry **275**(30): 23113-9.

Arthur, T. M. and R. R. Burgess (1998). "Localization of a sigma70 binding site on the N terminus of the Escherichia coli RNA polymerase beta' subunit." Journal of Biological Chemistry **273**(47): 31381-7.

Blair, D. F. and S. K. Dutcher (1992). "Flagella in prokaryotes and lower eukaryotes." Curr Opin Genet Dev **2**(5): 756-67.

Blow, D. (2002). Outline of crystallography for biologists. Oxford, Oxford Univeristy Press.

Blundell, T. L. and L. N. Johnson (1976). Protein Crystallography. New York, Academic Press.

Buckle, M., J. Geiselmann, et al. (1991). "Protein-DNA cross-linking at the lac promoter." Nucleic Acids Res **19**: 833-840.

Cadene, M. and B. T. Chait (2000). "A robust, detergent-friendly method for mass spectrometric analysis of integral membrane proteins." Anal Chem **72**(22): 5655-8.

Callaci, S., E. Heyduk, et al. (1998). "Conformational changes of Escherichia coli RNA polymerase sigma70 factor induced by binding to the core enzyme." J Biol Chem **273**(49): 32995-3001.

Callaci, S., E. Heyduk, et al. (1999). "Core RNA polymerase from E. coli induces a major change in the domain arrangement of the sigma 70 subunit." Mol Cell **3**(2): 229-38.

Camarero, J. A., A. Shekhtman, et al. (2002). "Autoregulation of a bacterial sigma factor explored by using segmental isotopic labeling and NMR." Proc Natl Acad Sci U S A **99**(13): 8536-41.

Campbell, E. A. and S. A. Darst (2000). "The anti-sigma factor SpoIIAB forms a 2:1 complex with sigma(F), contacting multiple conserved regions of the sigma factor." J Mol Biol **300**(1): 17-28.

Campbell, E. A., S. Masuda, et al. (2002). "Crystal structure of the Bacillus stearothermophilus anti-sigma factor SpoIIAB with the sporulation sigma factor sigmaF." Cell **108**(6): 795-807.

Campbell, E. A., O. Muzzin, et al. (2002). "Structure of the bacterial RNA polymerase promoter specificity sigma subunit." Mol Cell **9**(3): 527-39.

Campbell, E. A., J. L. Tupy, et al. (2003). "Crystal structure of Escherichia coli sigmaE with the cytoplasmic domain of its anti-sigma RseA." Mol Cell **11**(4): 1067-78.

Chadsey, M. S. and K. T. Hughes (2001). "A multipartite interaction between Salmonella transcription factor sigma28 and its anti-sigma factor FlgM: implications for sigma28 holoenzyme destabilization through stepwise binding." J Mol Biol **306**(5): 915-29.

Chadsey, M. S., J. E. Karlinsey, et al. (1998). "The flagellar anti-sigma factor FlgM actively dissociates Salmonella typhimurium sigma28 RNA polymerase holoenzyme." Genes Dev **12**(19): 3123-36.

Chater, K. F., C. J. Bruton, et al. (1989). "The developmental fate of S. coelicolor hyphae depends upon a gene product homologous with the motility sigma factor of B. subtilis." Cell **59**(1): 133-43.

Chen, L. and J. D. Helmann (1994). "The Bacillus subtilis sigma D-dependent operon encoding the flagellar proteins FliD, FliS, and FliT." J Bacteriol **176**(11): 3093-101.

Chen, Y. F. and J. D. Helmann (1995). "The Bacillus subtilis flagellar regulatory protein sigma D: overproduction, domain analysis and DNA-binding properties." J Mol Biol **249**(4): 743-53.

Colland, F., J. C. Rain, et al. (2001). "Identification of the Helicobacter pylori anti-sigma28 factor." Mol Microbiol **41**(2): 477-87.

Cramer, P., D. A. Bushnell, et al. (2001). "Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution." Science **292**(5523): 1863-76.

Dani, V. S., C. Ramakrishnan, et al. (2003). "MODIP revisited: re-evaluation and refinement of an automated procedure for modeling of disulfide bonds in proteins." Protein Eng **16**(3): 187-93.

Daniels, D., P. Zuber, et al. (1990). "Two amino acids in an RNA polymerase sigma factor involved in the recognition of adjacent base pairs in the -10 region of a cognate promoter." Proc Natl Acad Sci U S A **87**(20): 8075-9.

Dasgupta, N., M. C. Wolfgang, et al. (2003). "A four-tiered transcriptional regulatory circuit controls flagellar biogenesis in Pseudomonas aeruginosa." Mol Microbiol **50**(3): 809-24.

Daughdrill, G. W., M. S. Chadsey, et al. (1997). "The C-terminal half of the anti-sigma factor, FlgM, becomes structured when bound to its target, sigma 28." Nat Struct Biol **4**(4): 285-91.

Daughdrill, G. W., L. J. Hanely, et al. (1998). "The C-terminal half of the anti-sigma factor FlgM contains a dynamic equilibrium solution structure favoring helical conformations." Biochemistry **37**(4): 1076-82.

Decatur, A. L. and R. Losick (1996). "Three sites of contact between the Bacillus subtilis transcription factor sigmaF and its antisigma factor SpoIIAB." Genes Dev **10**(18): 2348-58.

Deckert, G., P. V. Warren, et al. (1998). "The complete genome of the hyperthermophilic bacterium Aquifex aeolicus." Nature **392**(6674): 353-8.

Dedmon, M. M., C. N. Patel, et al. (2002). "FlgM gains structure in living cells." PNAS **99**(20): 12681-12684.

Dombroski, A. J., W. A. Walter, et al. (1993). "Amino-terminal amino acids modulate sigma-factor DNA-binding activity." Genes Dev **7**(12A): 2446-55.

Dombroski, A. J., W. A. Walter, et al. (1992). "Polypeptides containing highly conserved regions of transcription initiation factor sigma 70 exhibit specificity of binding to promoter DNA." Cell **70**(3): 501-12.

Dorigo, B., T. Schalch, et al. (2004). "Nucleosome arrays reveal the two-start organization of the chromatin fiber." Science **306**(5701): 1571-3.

Doublie, S. (1997). "Preparation of selenomethionyl proteins for phase determination." Methods in Enzymology **276**: 523-530.

Dove, S. L., S. A. Darst, et al. (2003). "Region 4 of sigma as a target for transcription regulation." Mol Microbiol **48**(4): 863-74.

Duncan, L. and R. Losick (1993). "SpoIIAB is an anti-sigma factor that binds to and inhibits transcription by regulatory protein sigma F from Bacillus subtilis." Proc Natl Acad Sci U S A **90**(6): 2325-9.

Errington, J. (1993). "Bacillus subtilis sporulation: regulation of gene expression and control of morphogenesis." Microbiol Rev **57**(1): 1-33.

Gardella, T., H. Moyle, et al. (1989). "A mutant Escherichia coli sigma 70 subunit of RNA polymerase with altered promoter specificity." J Mol Biol **206**(4): 579-90.

Gillen, K. L. and K. T. Hughes (1991). "Molecular characterization of flgM, a gene encoding a negative regulator of flagellin synthesis in Salmonella typhimurium." J Bacteriol **173**(20): 6453-9.

Gillen, K. L. and K. T. Hughes (1991). "Negative regulatory loci coupling flagellin synthesis to flagellar assembly in Salmonella typhimurium." J Bacteriol **173**(7): 2301-10.

Gilman, M. Z., J. L. Wiggs, et al. (1981). "Nucleotide sequences of two Bacillus subtilis promoters used by Bacillus subtilis sigma-28 RNA polymerase." Nucleic Acids Res **9**(22): 5991-6000.

Gish, W. (1996-2004). "http://blast.wustl.edu."

Gish, W. and D. J. States (1993). "Identification of protein coding regions by database similarity search." Nat Genet **3**(3): 266-72.

Gregory, B. D., B. E. Nickels, et al. (2004). "A regulator that inhibits transcription by targeting an intersubunit interaction of the RNA polymerase holoenzyme." Proc Natl Acad Sci U S A **101**(13): 4554-9.

Gross, C. A., C. Chan, et al. (1998). "The functional and regulatory roles of sigma factors in transcription." Cold Spring Harb Symp Quant Biol **63**: 141-55.

Gruber, T. M. and D. A. Bryant (1997). "Molecular systematic studies of eubacteria, using sigma70-type sigma factors of group 1 and group 2." J Bacteriol **179**(5): 1734-47.

Helmann, J. D. (2002). "The extracytoplasmic function (ECF) sigma factors." Adv Microb Physiol **46**: 47-110.

Helmann, J. D. and M. J. Chamberlin (1987). "DNA sequence analysis suggests that expression of flagellar and chemotaxis genes in Escherichia coli and Salmonella typhimurium is controlled by an alternative sigma factor." Proc Natl Acad Sci U S A **84**(18): 6422-4.

Helmann, J. D. and M. J. Chamberlin (1988). "Structure and function of bacterial sigma factors." Annu Rev Biochem **57**: 839-72.

Hendrickson, W. A. and C. M. Ogata (1997). "Phase determination from multiwavelength anomalous diffraction measurements." Methods Enzymol **276**(A): 494-523.

Hendrixson, D. R. and V. J. DiRita (2003). "Transcription of sigma54-dependent but not sigma28-dependent flagellar genes in Campylobacter jejuni is associated with formation of the flagellar secretory apparatus." Mol Microbiol **50**(2): 687-702.

Hilton, M. D. and H. R. Whiteley (1985). "UV cross-linking of the Bacillus subtilis RNA polymerase to DNA in promoter and non-promoter complexes." J Biol Chem **260**(13): 8121-7.

Huang, X., F. J. Lopez de Saro, et al. (1997). "sigma factor mutations affecting the sequence-selective interaction of RNA polymerase with -10 region single-stranded DNA." Nucleic Acids Res **25**(13): 2603-9.

Hughes, K. T., K. L. Gillen, et al. (1993). "Sensing structural intermediates in bacterial flagellar assembly by export of a negative regulator." Science **262**(5137): 1277-80.

Iyoda, S. and K. Kutsukake (1995). "Molecular dissection of the flagellum-specific anti-sigma factor, FlgM, of Salmonella typhimurium." Mol Gen Genet **249**(4): 417-24.

Jones, C. H. and C. P. Moran, Jr. (1992). "Mutant sigma factor blocks transition between promoter binding and initiation of transcription." Proc Natl Acad Sci U S A **89**(5): 1958-62.

Jones, T. A., J.-Y. Zou, et al. (1991). "Improved methods for building protein models in electron density maps and the location of errors in these models." Acta Cryst. **A47**: 110-119.

Juang, Y. L. and J. D. Helmann (1995). "Pathway of promoter melting by Bacillus subtilis RNA polymerase at a stable RNA promoter: effects of temperature, delta protein, and sigma factor mutations." Biochemistry **34**(26): 8465-73.

Keilty, S. and M. Rosenberg (1987). "Constitutive function of a positively regulated promoter reveals new sequences essential for activity." J Biol Chem **262**(13): 6389-95.

Kumar, A., R. A. Malloch, et al. (1993). "The minus 35-recognition region of Escherichia coli sigma 70 is inessential for initiation of transcription at an "extended minus 10" promoter." J Mol Biol **232**(2): 406-18.

Kutsukake, K. (1994). "Excretion of the anti-sigma factor through a flagellar substructure couples flagellar gene expression with flagellar assembly in Salmonella typhimurium." Mol Gen Genet. **243**(6): 605-612.

Kutsukake, K., S. Iyoda, et al. (1994). "Genetic and molecular analyses of the interaction between the flagellum-specific sigma and anti-sigma factors in Salmonella typhimurium." Embo J **13**(19): 4568-76.

Kutsukake, K., Y. Ohya, et al. (1990). "Transcriptional analysis of the flagellar regulon of Salmonella typhimurium." J Bacteriol **172**(2): 741-7.

Kuznedelov, K., L. Minakhin, et al. (2002). "A role for interaction of the RNA polymerase flap domain with the sigma subunit in promoter recognition." Science **295**(5556): 855-7.

Lambert, L. J., Y. Wei, et al. (2004). "T4 AsiA blocks DNA recognition by remodeling sigma70 region 4." Embo J **23**(15): 2952-62.

Laskowski, R. A., M. W. MacArthur, et al. (1993). "PROCHECK—a program to check the stereochemical quality of protein structures." Journal of Applied Crystallography **26**: 283-291.

Li, W., C. E. M. Stevenson, et al. (2002). "Identification and Structure of the Anti-sigma Factor-binding Domain of the Disulphide-stress Regulated Sigma Factor [sigma]R from Streptomyces coelicolor." Journal of Molecular Biology **323**(2): 225-236.

Lonetto, M., M. Gribskov, et al. (1992). "The sigma 70 family: sequence conservation and evolutionary relationships." J Bacteriol **174**(12): 3843-9.

Losick, R. and J. Pero (1981). "Cascades of Sigma factors." Cell **25**(3): 582-4.

Macnab, R. M. (2003). "How bacteria assemble flagella." Annu Rev Microbiol **57**: 77-100.

Malhotra, A., E. Severinova, et al. (1996). "Crystal structure of a sigma 70 subunit fragment from E. coli RNA polymerase." Cell **87**(1): 127-36.

Mimori, Y., I. Yamashita, et al. (1995). "The structure of the R-type straight flagellar filament of Salmonella at 9 A resolution by electron cryomicroscopy." J Mol Biol **249**(1): 69-87.

Morgan, D. G., R. M. Macnab, et al. (1993). "Domain organization of the subunit of the Salmonella typhimurium flagellar hook." J Mol Biol **229**(1): 79-84.

Morgan, D. G., C. Owen, et al. (1995). "Structure of bacterial flagellar filaments at 11 A resolution: packing of the alpha-helices." J Mol Biol **249**(1): 88-110.

Murakami, K. S. and S. A. Darst (2003). "Bacterial RNA polymerases: the wholo story." Curr Opin Struct Biol **13**(1): 31-9.

Murakami, K. S., S. Masuda, et al. (2002). "Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex." Science **296**(5571): 1285-90.

Murakami, K. S., S. Masuda, et al. (2002). "Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 A resolution." Science **296**(5571): 1280-4.

Murakami, K. S., Sorenson, M. K. & S.A. Darst (2002). Unpublished data.

Nicholls, A., K. Sharp, et al. (1991). "Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons." Proteins **11**: 281-296.

Ohnishi, K., K. Kutsukake, et al. (1990). "Gene fliA encodes an alternative sigma factor specific for flagellar operons in Salmonella typhimurium." Mol Gen Genet **221**(2): 139-47.

Ohnishi, K., K. Kutsukake, et al. (1992). "A novel transcriptional regulation mechanism in the flagellar regulon of Salmonella typhimurium: an

antisigma factor inhibits the activity of the flagellum-specific sigma factor, sigma F." Mol Microbiol **6**(21): 3149-57.

Orsini, G., M. Ouhammouch, et al. (1993). "The asiA gene of bacteriophage T4 codes for the anti-sigma 70 protein." J Bacteriol **175**(1): 85-93.

Otwinowski, Z. and W. Minor (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods in Enzymology:Macromolecular Crystallography, part A. R. M. Sweet. **276:** 307-326.

Park, C. S., Z. Hillel, et al. (1980). "DNA strand specificity in promoter recognition by RNA polymerase." Nucleic Acids Res **8**(23): 5895-912.

Perrakis, A., R. J. Morris, et al. (1999). "Automated protein model building combined with iterative structure refinement." Nature Struct. Biol. **6**: 458-463.

Prouty, M. G., N. E. Correa, et al. (2001). "The novel sigma54- and sigma28-dependent flagellar gene transcription hierarchy of Vibrio cholerae." Mol Microbiol **39**(6): 1595-609.

Riddles, P. W., R. L. Blakeley, et al. (1983). "Reassessment of Ellman's reagent." Methods Enzymol **91**: 49-60.

Rong, J. C. and J. D. Helmann (1994). "Genetic and physiological studies of Bacillus subtilis sigma A mutants defective in promoter melting." J Bacteriol **176**(17): 5218-24.

Rost, B. and C. Sander (1993). "Prediction of protein secondary structure at better than 70% accuracy." J Mol Biol **232**(2): 584-99.

Schwede, T., J. Kopp, et al. (2003). "SWISS-MODEL: An automated protein homology-modeling server." Nucleic Acids Res **31**(13): 3381-5.

Sen, R., H. Nagai, et al. (1998). "Reduction in abortive transcription from the lambdaPR promoter by mutations in region 3 of the sigma70 subunit of Escherichia coli RNA polymerase." J Biol Chem **273**(16): 9872-7.

Severinov, K. K. K. (2004). Unpublished data.

Severinova, E., K. Severinov, et al. (1998). "Inhibition of Escherichia coli RNA polymerase by bacteriophage T4 AsiA." J Mol Biol **279**(1): 9-18.

Severinova, E., K. Severinov, et al. (1996). "Domain organization of the Escherichia coli RNA polymerase sigma 70 subunit." J Mol Biol **263**(5): 637-47.

Siegele, D. A., J. C. Hu, et al. (1989). "Altered promoter recognition by mutant forms of the sigma 70 subunit of Escherichia coli RNA polymerase." J Mol Biol **206**(4): 591-603.

Simeonov, M. F., R. J. Bieber Urbauer, et al. (2003). "Characterization of the interactions between the bacteriophage T4 AsiA protein and RNA polymerase." Biochemistry **42**(25): 7717-26.

Simpson, R. B. (1979). "The molecular topology of RNA polymerase-promoter interaction." Cell **18**: 277-285.

Sorenson, M. K., S. S. Ray, et al. (2004). "Crystal structure of the flagellar sigma/anti-sigma complex sigma(28)/FlgM reveals an intact sigma factor in an inactive conformation." Mol Cell **14**(1): 127-38.

Sowdhamini, R., N. Srinivasan, et al. (1989). "Stereochemical modeling of disulfide bridges. Criteria for introduction into proteins by site-directed mutagenesis." Protein Eng **3**(2): 95-103.

Studholme, D. J. and M. Buck (2000). "The alternative sigma factor sigma(28) of the extreme thermophile Aquifex aeolicus restores motility to an Escherichia coli fliA mutant." FEMS Microbiol Lett **191**(1): 103-7.

Terwilliger, T. C. "www.solve.lanl.gov."

Terwilliger, T. C. (1999). "Reciprocal-space solvent flattening." Acta Cryst. **D55**: 1863-1871.

Terwilliger, T. C. (2000). "Maximum likelihood density modification." Acta Cryst. **D56**(965-972).

Terwilliger, T. C. (2001). "Maximum-likelihood density modification with pattern recognition of structural motifs." Acta Cryst. **D57**: 1755-1762.

Terwilliger, T. C. (2002). "Automated main-chain model-building by template-matching and iterative fragment extension." Acta Crystallographica **D59**: 34-44.

Terwilliger, T. C. (2003). "Automated side-chain model building and sequence assignment by template matching." Acta Cryst. **D59**: 45-49.

Terwilliger, T. C. and J. Berendzen (1999). "Automated MAD and MIR structure solution." Acta Crystallographica **D55**: 849-861.

Urbauer, J. L., M. F. Simeonov, et al. (2002). "Solution structure and stability of the anti-sigma factor AsiA: implications for novel functions." Proc Natl Acad Sci U S A **99**(4): 1831-5.

Vassylyev, D. G., S. Sekine, et al. (2002). "Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution." Nature **417**(6890): 712-9.

Vingadassalom, D., A. Kolb, et al. (2005). "An unusual primary sigma factor in the Bacteroidetes phylum." Mol Microbiol **56**(4): 888-902.

Waldburger, C., T. Gardella, et al. (1990). "Changes in conserved region 2 of Escherichia coli sigma 70 affecting promoter recognition." J Mol Biol **215**(2): 267-76.

Wimberly, B. T., D. E. Brodersen, et al. (2000). "Structure of the 30S ribosomal subunit." Nature **407**(6802): 327-39.

Young, B. A., L. C. Anthony, et al. (2001). "A coiled-coil from the RNA polymerase beta' subunit allosterically induces selective nontemplate strand binding by sigma(70)." Cell **105**(7): 935-44.

Yusupov, M. M., G. Z. Yusupova, et al. (2001). "Crystal structure of the ribosome at 5.5 A resolution." Science **292**(5518): 883-96.

Zhang, G., E. A. Campbell, et al. (1999). "Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution." Cell **98**(6): 811-24.

Zhang, X., M. Chaney, et al. (2002). "Mechanochemical ATPases and transcriptional activation." Mol Microbiol **45**(4): 895-903.

Zhou, Y. N., W. A. Walter, et al. (1992). "A mutant sigma 32 with a small deletion in conserved region 3 of sigma has reduced affinity for core RNA polymerase." J Bacteriol **174**(15): 5005-12.

Zuber, P., J. Healy, et al. (1989). "Mutation changing the specificity of an RNA polymerase sigma factor." J Mol Biol **206**: 605–614.