2009

Structural Studies of Three Factors That Affect the Prokaryotic Transcription Cycle; Microcin J25, LAMBDA Q and T4 GP33

Kelly-Anne Fisi Twist

STRUCTURAL STUDIES OF THREE FACTORS THAT AFFECT THE PROKARYOTIC TRANSCRIPTION CYCLE; MICROCIN J25, LAMBDA Q AND T4 GP33.

A Thesis Presented to the Faculty of

The Rockefeller University

in Partial Fulfillment of the Requirements for

the degree of Doctor of Philosophy

by

Kelly-Anne Fisi Twist

June 2009

STRUCTURAL STUDIES OF THREE FACTORS THAT AFFECT THE PROKARYOTIC TRANSCRIPTION CYCLE; MICROCIN J25, LAMBDA Q AND T4 GP33.

Kelly-Anne Fisi Twist Ph.D.

The Rockefeller University 2009

This thesis examines three factors that affect the prokaryotic transcription cycle; Microcin J25, a peptidyl inhibitor of RNAP; Q antiterminator protein, a bacteriophage protein that modifies RNAP to readthrough termination signals in phage late genes; and Gp33, a T4 phage protein that is a co-activator of T4 phage late gene transcription.

We determined that the Microcin J25 peptide (MccJ25) has a lassoed-tail structure using mass spectrometry and NMR. The 21 amino acid MccJ25 peptide is produced from a 58 amino acid pro-peptide, McjA, that is truncated and post-translationally modified by two maturase enzymes. We synthesized McjA and constructed a protein-A tagged biosynthesis cassette to be used in the analysis of MccJ25 maturation.

Escherichia coli (*Ec*) RNAP has not, thus far, been amenable to crystallization despite considerable effort. One potential hindrance to crystallization may be the carboxy terminal domain of the alpha subunit (α-CTD, residues 249-329) that is connected to an amino-terminal domain (α-NTD, residues 8-235) by a 14-residue linker. We successfully expressed, purified and crystallized a version of *Ec* RNAP without the α-CTD.

The Q antiterminator proteins from lambdoid phage, modify *Ec* RNAP to readthrough termination sites within the phage late gene. A structure of Q protein could reveal how Q antitermination occurs and may provide insight into the mechanism of termination. We developed an improved expression and purification protocol for λQ, and performed biophysical characterization on both λQ and the related 82Q.

The T4 phage protein Gp33 is a co-activator that interacts with both *Ec* RNAP and *T4* DNA replication machinery to upregulate transcription from T4 late promoters. Gp33 is thought to have an analogous function to E. coli $\sigma^{70}4$ in that it binds the beta flap of RNAP and helps recruit RNAP to promoters. *Ec* RNAP bound to Gp33 is recruited to the promoter by a unique mechanism through a protein-protein interaction between Gp33 and the DNA-bound sliding clamp of the T4 DNA replisome, Gp45. We obtained crystals of Gp33 bound to the *Ec* Beta flap that diffracted to 3.0 Ångstroms. The preliminary structure shows Gp33 as a mainly helical protein that wraps around the Beta flap.

Dedicated to Malcolm and Leona Twist. <3.

ACKNOWLEDGMENTS

I would likely to gratefully acknowledge my thesis advisor, Dr Seth Darst, for his mentorship and for his encouragement to tackle difficult projects. I also thank my FAC members, Tarun Kapoor, Brian Chait and Tom Muir, who were always available for consultation.

I would like to thank all the members of the Darst, Muir and Chait laboratories who have helped me throughout the past few years. In particular I want to thank Markus Kalkum and Jen Ottesen who worked with me on solving the structure of the microcin J25 peptide and were my first direct mentors at Rockefeller University.

I also want to extend deeply felt gratitude to Lars Westblade, Valerie Lamour and Elizabeth Campbell for their experimental guidance and encouragement and Lars, in particular, for his scientific enthusiasm and inspiring discussions.

I am also grateful to the Dean's office and the staff at Rockefeller who make being a student here such a wonderful experience.

I have enjoyed the privilege of earning my PhD at Rockefeller University.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1: General introduction

1.1 A structure based description of prokaryotic gene transcription

Gene transcription in all organisms is carried out by DNA-dependent RNA polymerases. In bacteria all genes are transcribed by the evolutionary conserved RNA polymerase core enzyme (RNAP), a protein complex of about 400 kDa with the subunit composition of two alpha, beta, beta prime and omega ($\alpha_2\beta\beta'\omega$) [1]. RNAP core is catalytically competent to elongate RNA chains from a DNA template. RNAP core can interact non-specifically with DNA but must bind an additional factor, sigma (σ), to form the holoenzyme capable of specific initiation at promoters [1]. The crystal structures of various σ's, the complete RNAP core, RNAP holoenzyme, and RNAP elongating complexes (containing DNA template and RNA transcript) have provided a structural basis to interpret decades of biochemical and genetic analyses into the mechanism and regulation of the transcription cycle, (initiation, elongation and termination; outlined in figure 1).

Core assembly begins with the dimerization of the alpha subunits via their amino terminal domains (α-NTD), which provides a scaffold for further assembly of the complex [2]. The β subunit associates to form the $\alpha_2\beta$ subassembly, which then binds the β′-ω subassembly to form the complete core [3]. The overall structure of the core resembles a crab claw, with β and β′ each forming one pincer [4; 5; 6]. The catalytic site is located deep within the main channel formed between the pincers. The α-NTDs and ω are bound on the outside of the claw and have no direct interaction with the catalytic residues.

Figure 1.1 Steps in the transcription cycle.

The crystal structure of the claw-shaped core RNAP (1HQM) is shown with two sub-assemblies*. In the cartoons core RNAP is grey with the β flap highlighted in blue, σ is orange, the template DNA is dark green, the non-template DNA is light green*2.

Transcription begins when core RNAP binds σ and becomes competent to bind the promoter. The DNA is then melted to form the transcription bubble with the template strand entering the main channel and the non-template strand guided around the outside of the enzyme. NTPs access the catalytic site via the secondary channel and those that correctly base-pair to the template DNA are stabilized by interactions with both the DNA and the RNAP. Phosphodiesters bonds are then catalyzed between the NTPS to form the mRNA transcript. This transcript clashes with the σ3.2 loop and is dislodged from the catalytic site in several rounds of abortive initiation. When the transcript reaches 15-17 nt it displaces the σ3.2 loop and exits via the RNA exit channel beneath the β flap dislodging σ4 and clearing the promoter. σ is eventually released as the enzyme enters productive elongation. Transcription continues until RNAP encounters a termination signal and releases the mRNA transcript and the DNA. Both RNAP core and σ are recycled to begin the pathway again.

Further descriptions are provided in text.

* Shown for illustrative purposes, these structures are based on 1HQM not independently derived structures

*2 Cartoons are from Murakami Darst The Wholo Story, 2002.

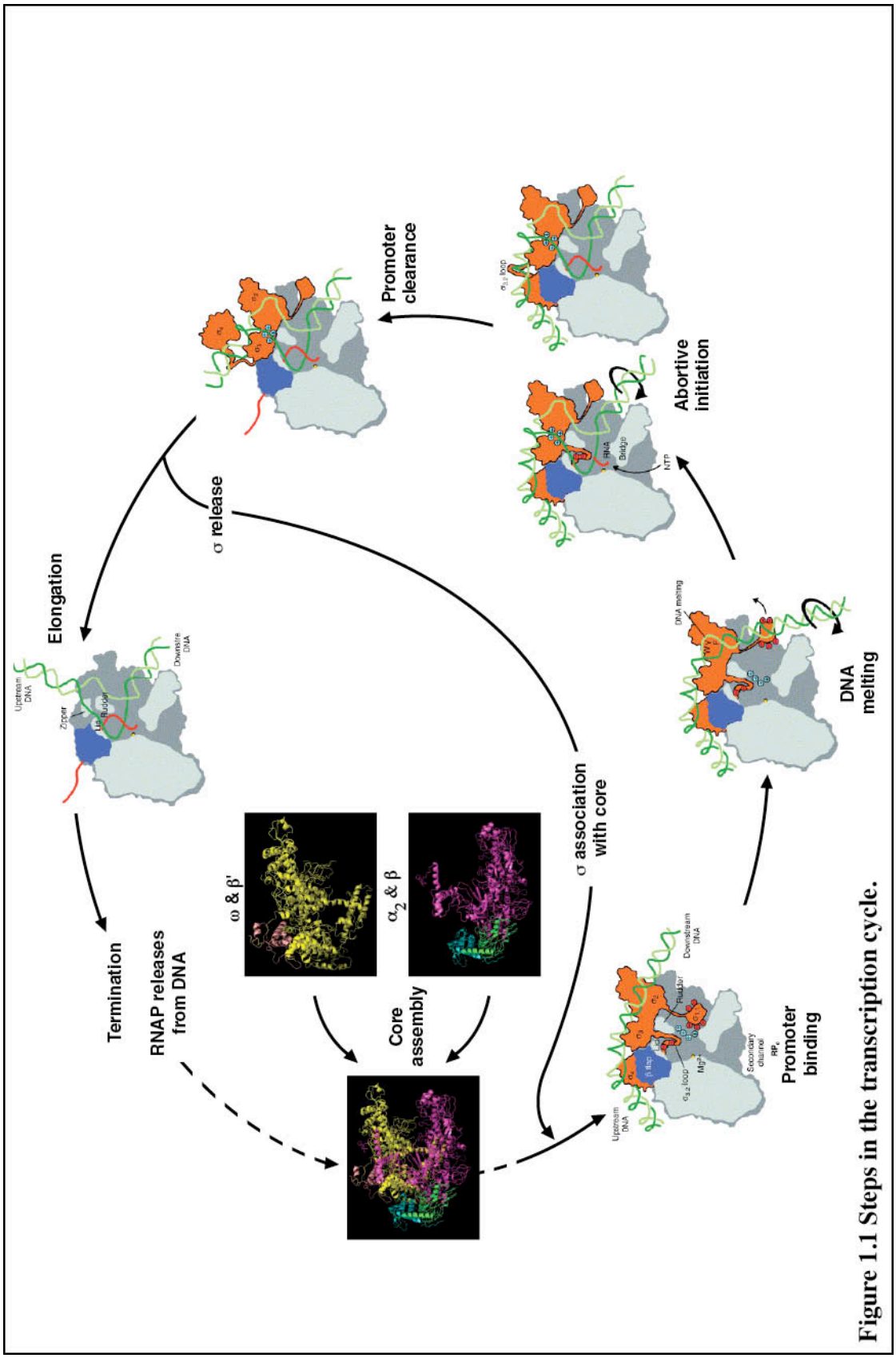


Figure 1.1 Steps in the transcription cycle.

The RNAP core is approximately 150 Å long, 115 Å tall and 115 Å wide, with an internal channel that flexes as wide as 27 Å [4]. The internal channel is split by a "wall" formed by two β′ helices (β′F & β′G) that extend between the two lobes and separate the main channel that accommodates downstream DNA and the RNA-DNA hybrid, from the secondary channel that allows nucleotides access to the active site [7; 8].

Initiation of transcription begins with the holoenzyme binding promoter DNA to form the closed complex. The specific interactions between the holoenzyme and the promoter DNA are provided by the σ subunit, which also mediates DNA melting [5; 9].

σ's are key regulators of transcription. All bacteria contain a primary σ factor that directs transcription of essential housekeeping genes. Most bacteria also contain alternative σ's that recruit RNAP to their cognate promoters to direct transcription in response to various environmental or morphological cues [10]. σ's are, in turn, regulated by anti-σ's, anti-anti-σ's and σ-appropriators [11]. Primary σ's recognize promoters containing two hexameric consensus motifs; the -10 element (or Pribnow box, consensus sequence TATAAT), and the -35 element (consensus sequence TTGACA), centered, respectively, 10 bp and 35 bp upstream of the transcription start site, +1[5]. A structure-based sequence alignment based on Thermus aquatics (Taq) σ^A, a primary σ factor, defines four conserved regions, σ1.1, σ2, σ3 & σ4, that form distinct structural domains connected by flexible linkers [9]. Biochemical and genetic experiments indicated that σ2.4 bound the -10 element and σ4.2 bound the -35 element and that σ formed extensive interactions with RNAP [12]. Free, full-length σ is unable to bind DNA because σ1.1 stabilizes a compact form of σ that cannot

bind DNA [13]. However, in the holoenzyme, σ is in an extended conformation with σ2, σ3 & σ4 splayed across the upstream face of RNAP, forming extensive protein-protein interactions. The σ/RNAP interactions position σ2.4 and σ4 appropriately to make specific interactions with the -10 and -35 promoter elements, respectively [6; 14]. σ2 is bound to the β′ subunit coiled-coil and lid [6; 14]. σ3 is bound near the downstream opening of the claw and interacts with features from both the β and β' subunits including the β flap and the β' zipper, zinc-binding domain, lid, coiled-coil and lid [6; 14]. σ3 and σ 4 are linked by a long linker that winds through the main channel, forming a hairpin loop near the active site and occupying the RNA exit channel beneath the β flap [6; 14].

The RNAP holoenzyme has an immobile core and "mobile modules" that move as rigid bodies with respect to each other, allowing the claw to open so that RNAP can flexibly interact with DNA [14] These mobile modules are the σ4 bound to the β flap, the β' clamp bound to the σ2- σ3 domains, and two portions of the β pincer; β1 and β2 [14]. These mobile modules would position the σ2 and σ4 domains to interact with the -10 and -35 element to form the closed complex [6; 14].

Once the closed complex has formed it undergoes a series of isomerization steps that lead to the open complex with a melted DNA transcription bubble that extends from nucleotides -11 to +4 with respect to the start site at +1 [15]. In the open complex DNA is bent to curve around the RNAP. The DNA melting starts at -11 and is mediated by σ2.3. As the strands separate the non-template strand is held by a groove on the outside of the enzyme formed between the β1 and β2 domains while the template strand enters the

catalytic site via a protein tunnel formed by parts of σ, β and β' [1].

Before RNAP clears the promoter it undergoes multiple rounds of abortive initiation whereby short (2-12 nucleotides) RNA transcripts sterically clash with the σ3-σ4 linker in the RNA exit channel [14]. Region 3.2 of σ stabilizes the initiating nucleotide [29]. Promoter escape occurs when a longer RNA transcript displaces this linker from the exit channel, and during early elongation the transcript must dislodge σ4 from the beta flap to emerge from the exit channel, and begin the process of σ release [7]. During elongation the transcription bubble moves along the DNA and the non-template DNA and template DNA re-anneal upon exiting the enzyme [7].

The elongation complex eventually terminates in either a rho-dependent or rho-independent manner [16]. DNA encoded termination signals lead to the formation of RNA stem-loop structures that cause the RNAP to pause, and are thought to destabilize the nucleic acid-RNAP interaction such that RNAP is released and recycled.

1.2 Thesis outline

Factors that modulate, inhibit or subvert the transcriptional machinery are interesting in their own right and can provide additional insight into the transcription cycle. This thesis examines three such factors; Microcin J25, a peptidyl inhibitor of RNAP; Q antiterminator protein, a bacteriophage protein that modifies RNAP to readthrough termination signals in phage late genes; and gp33, a T4 phage protein that is a co-activator of T4 phage late gene transcription.

Mircocin J25 (MccJ25) is a small peptide produced by *Escherichia coli* (*Ec*) that inhibits RNAP within other *Ec* bacteria lacking the MccJ25 immunity efflux pump [17; 18; 19]. We determined that this peptide has a remarkable lassoed-tail structure using mass spectrometry and NMR [20; 21; 22]. This work is described in chapter 2.

The 21 amino acid MccJ25 peptide is produced from a 58 amino acid pro-peptide, McjA, that is truncated and post-translationally modified by two maturase enzymes [23, 24]. We synthesized McjA and constructed a protein-A tagged biosynthesis cassette to be used in the analysis of MccJ25 maturation. However, despite considerable effort, we could not develop a system to overexpress these three enzymes. These efforts are described in chapter 3.

Genetic and biochemical evidence suggest that MccJ25 binds with the secondary channel [17; 18; 19; 25]. Indeed, these studies provide further support for this channel being the nucleotide access pathway to the catalytic core. A crystal structure of MccJ25 bound to

RNAP would reveal the mechanistic details of inhibition, but MccJ25 does not inhibit either of the bacterial RNAPs that have been crystallized. It would be very useful to have a crystal structure of the very well characterized Escherichia coli (*Ec*) RNAP for structural studies with the various factors such as mccJ25 that specifically target *Ec* RNAP. The successful expression, purification and crystallization of *Ec* RNAP is described in chapter 4.

The Q antiterminator proteins from lambda phage, and related phages, modify *Ec* RNAP to readthrough termination sites within the phage late genes [26]. It is thought that termination at these sites is caused by mRNA hairpin loop formation as at standard intrinsic termination sites, but the molecular mechanism of Q antitermination is not understood. A structure of Q protein could reveal how Q antitermination occurs and may provide insight into the mechanism of termination. To this end, I developed an improved expression and purification protocol for Q, and performed biophysical characterization. These efforts are detailed in chapter 5.

The T4 phage protein Gp33 is a coactivator that interacts with both *Ec* RNAP and T4 DNA replication machinery to upregulate transcription from T4 late promoters [27]. Gp33 is thought to have an analogous function to *Ec* $\sigma^{70}4$ in that it binds the beta flap of RNAP and helps recruit RNAP to promoters [28]. In contrast to $\sigma^{70}4$, however, T4 Gp33 does not interact with DNA, and T4 late promoters do not have a -35 element. *Ec* RNAP bound to Gp33 is recruited to the promoter by a unique mechanism through a protein-protein interaction between Gp33 and the DNA-bound sliding clamp of the T4 DNA replisome,

gp45 [27]. We obtained crystals of Gp33 bound to the *Ec* beta flap that diffracted to 3.0 Ångstroms and were able to solve the structure. The preliminary structure is discussed in chapter 6.

1.3 References

1. Darst, S. A. (2001). Bacterial RNA polymerase. In *Curr Opin Struct Biol*, Vol. 11, pp. 155-62.

2. Zhang, G. & Darst, S. A. (1998). Structure of the Escherichia coli RNA polymerase alpha subunit amino-terminal domain. In *Science*, Vol. 281, pp. 262-6.

3. Mathew, R. & Chatterji, D. (2006). The evolving story of the omega subunit of bacterial RNA polymerase. In *Trends Microbiol*, Vol. 14, pp. 450-5.

4. Zhang, G., Campbell, E. A., Minakhin, L., Richter, C., Severinov, K. & Darst, S. A. (1999). Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution. In *Cell*, Vol. 98, pp. 811-24.

5. Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O. & Darst, S. A. (2002). Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. In *Science*, Vol. 296, pp. 1285-90.

6. Vassylyev, D. G., Sekine, S., Laptenko, O., Lee, J., Vassylyeva, M. N., Borukhov, S. & Yokoyama, S. (2002). Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. In *Nature*, Vol. 417, pp. 712-9.

7. Korzheva, N., Mustaev, A., Kozlov, M., Malhotra, A., Nikiforov, V., Goldfarb, A. & Darst, S. A. (2000). A structural model of transcription elongation. In *Science*, Vol. 289, pp. 619-25.

8. Vassylyev, D. G., Vassylyeva, M. N., Perederina, A., Tahirov, T. & Artsimovitch, I. (2007). Structural basis for transcription elongation by bacterial RNA polymerase. In *Nature*, Vol. 448, pp. 157-162.

9. Campbell, E. A., Muzzin, O., Chlenov, M., Sun, J. L., Olson, C. A., Weinman, O., Trester-Zedlitz, M. L. & Darst, S. A. (2002). Structure of the bacterial RNA polymerase promoter specificity sigma subunit. In *Mol Cell*, Vol. 9, pp. 527-39.

10. Gruber, T. & Gross, C. (2003). Multiple sigma subunits and the partitioning of bacterial transcription space. In *Annu Rev Microbiol*, Vol. 57, pp. 441-66.

11. Campbell, E. A., Westblade, L. F. & Darst, S. A. (2008). Regulation of bacterial RNA polymerase sigma factor activity: a structural perspective. In *Curr Opin Microbiol*, Vol. 11, pp. 121-7.

12. Gross, C. A., Chan, C., Dombroski, A., Gruber, T., Sharp, M., Tupy, J. & Young, B. (1998). The functional and regulatory roles of sigma factors in transcription. In *Cold Spring Harbor Symposia on Quantitative Biology*, Vol. 63, pp. 141-55.

13. Schwartz, E. C., Shekhtman, A., Dutta, K., Pratt, M. R., Cowburn, D., Darst, S. & Muir, T. W. (2008). A full-length group 1 bacterial sigma factor adopts a compact structure incompatible with DNA binding. In *Chemistry & Biology*, Vol. 15, pp. 1091-103.

14. Murakami, K. S., Masuda, S. & Darst, S. A. (2002). Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 A resolution. In *Science*, Vol. 296, pp. 1280-4.

15. Paget, M. S. & Helmann, J. D. (2003). The sigma70 family of sigma factors. In *Genome Biol*, Vol. 4, pp. 203.

16. Henkin, T. M. (1996). Control of transcription termination in prokaryotes. *Annu Rev Genet* **30**, 35-57.

17. Mukhopadhyay, J., Sineva, E., Knight, J., Levy, R. M. & Ebright, R. H. (2004). Antibacterial peptide microcin J25 inhibits transcription by binding within and obstructing the RNA polymerase secondary channel. In *Mol Cell*, Vol. 14, pp. 739-51.

18. Yuzenkova, J., Delgado, M., Nechaev, S., Savalia, D., Epshtein, V., Artsimovitch, I., Mooney, R. A., Landick, R., Farias, R. N., Salomon, R. & Severinov, K. (2002). Mutations of bacterial RNA polymerase leading to resistance to microcin j25. *J Biol Chem* **277**, 50867-75.

19. Delgado, M. A., Rintoul, M. R., Farias, R. N. & Salomon, R. A. (2001). Escherichia coli RNA polymerase is the target of the cyclopeptide antibiotic microcin J25. *J Bacteriol* **183**, 4543-50.

20. Wilson, K. A., Kalkum, M., Ottesen, J., Yuzenkova, J., Chait, B. T., Landick, R., Muir, T., Severinov, K. & Darst, S. A. (2003). Structure of microcin J25, a peptide inhibitor of bacterial RNA polymerase, is a lassoed tail. *J Am Chem Soc* **125**, 12475-83.

21. Rosengren, K. J., Clark, R. J., Daly, N. L., Goransson, U., Jones, A. & Craik, D. J. (2003). Microcin J25 has a threaded sidechain-to-backbone ring structure and not a head-to-tail cyclized backbone. *J Am Chem Soc* **125**, 12464-74.

22. Bayro, M. J., Mukhopadhyay, J., Swapna, G. V., Huang, J. Y., Ma, L. C., Sineva, E., Dawson, P. E., Montelione, G. T. & Ebright, R. H. (2003). Structure of antibacterial peptide microcin J25: a 21-residue lariat protoknot. *J Am Chem Soc* **125**, 12382-3.

23. Solbiati, J. O., Ciaccio, M., Farías, R. N., González-Pastor, J. E., Moreno, F. & Salomón, R. A. (1999). Sequence analysis of the four plasmid genes required to produce the circular peptide antibiotic microcin J25. In *J Bacteriol*, Vol. 181, pp. 2659-62.

24. Duquesne, S., Destoumieux-Garzón, D., Zirah, S., Goulard, C., Peduzzi, J. & Rebuffat, S. (2007). Two enzymes catalyze the maturation of a lasso peptide in Escherichia coli. In *Chemistry & Biology*, Vol. 14, pp. 793-803.

25. Adelman, K., Yuzenkova, J., La Porta, A., Zenkin, N., Lee, J., Lis, J. T., Borukhov, S., Wang, M. D. & Severinov, K. (2004). Molecular mechanism of transcription inhibition by peptide antibiotic Microcin J25. In *Mol Cell*, Vol. 14, pp. 753-62.

26. Yang, X. J., Goliger, J. A. & Roberts, J. W. (1989). Specificity and mechanism of antitermination by Q proteins of bacteriophages lambda and 82. *J Mol Biol* **210**, 453-60.

27. Sanders, G. M., Kassavetis, G. A. & Geiduschek, E. P. (1997). Dual targets of a transcriptional activator that tracks on DNA. In *EMBO J*, Vol. 16, pp. 3124-32.

28. Nechaev, S., Kamali-Moghaddam, M., André, E., Léonetti, J. P. & Geiduschek, E. P. (2004). The bacteriophage T4 late-transcription coactivator gp33 binds the flap domain of Escherichia coli RNA polymerase. In *Proc Natl Acad Sci USA*, Vol. 101, pp. 17365-70.

29. Nickels BE, Garrity SJ, Mekler V, Minakhin L, Severinov K, Ebright RH. Hochschild A. (2005) The interaction between sigma70 and the beta-flap of Escherichia coli RNA polymerase inhibits extension of nascent RNA during early elongation. In *Proc Natl Acad Sci USA*. vol. 102, pp 4488-93.

Chapter 2: The Structure of Microcin J25, a Peptide inhibitor of Bacterial RNA Polymerase, is a Lassoed Tail

J Am Chem Soc. 2003 Oct 15;125(41):12475-83. Structure of microcin J25, a peptide inhibitor of bacterial RNA polymerase, is a lassoed tail. Wilson KA, Kalkum M, Ottesen J, Yuzenkova J, Chait BT, Landick R, Muir T, Severinov K, Darst SA.

2.1 Introduction

DNA-dependent RNAP, the central enzyme of bacterial gene expression, is a complex molecular machine [1,2]. The catalytically competent core RNAP has subunit composition $\alpha_2\beta\beta\omega$ and a total molecular weight near 400 kDa. In contrast to the ribosome, surprisingly few small-molecule inhibitors specific for bacterial RNAPs have been described, despite their potential as tools to investigate RNAP mechanism as well as for antibacterial drugs. Only one inhibitor, rifampicin (or its derivatives), has reached medical use [3], where it is a key component of anti-tuberculosis therapy. Thus, recent findings that an antibacterial peptide, Microcin J25 (MccJ25), functions through inhibition of the bacterial RNAP [4,5] has generated wide interest.

MccJ25 is a 21-residue, ribosomally-synthesized peptide [6,7] produced by strains of *Ec* harboring a plasmid-borne synthesis, maturation, and export system [8]. MccJ25 exhibits bacteriocidal activity against a range of Gram-negative bacterial species [6]. Genetic and

biochemical results suggest that MccJ25 inhibits transcription by binding within the RNAP secondary channel [5], a 12-15 Å diameter channel that branches off from the main active site channel, through which the nucleotide substrates can diffuse to the enzyme active site [9,10]. Binding of MccJ25 within this channel could block nucleotide substrates from entering the enzyme active site. An important step towards further understanding and manipulating the activity of MccJ25 is the characterization of its chemical and three-dimensional structure.

On the basis of biochemical and NMR studies, the primary structure of MccJ25 was reported to be a 21-residue, head-to-tail cyclic peptide, [*cyclo*(-G^{1}GAGHVPEYF^{10}VGIGTPISFY^{20}G-), or $\overline{G^{1}\text{GAGHVPEYF}^{10}\text{VGIGTPISFY}^{20}\text{G}}$, using a numbering scheme where G^{1} is the N-terminally coded amino acid of the mature peptide] and the three-dimensional solution structure was reported [7,11]. Here, we start with the key observation that synthetic peptides based on this reported structure are biologically inactive. We show, using a combination of biochemical studies, mass spectrometry, and NMR, that the reported cyclic structure is incorrect, and that the peptide contains an extraordinary structural fold. The structure contains a lactam linkage between the α-amino group of Gly1 and the γ-carboxyl of Glu8, forming an 8-residue ring (Gly1-Glu8) that we term the lariat ring. The 'tail' (Tyr9-Gly21) passes through the lariat ring, with Phe19 and Tyr20 straddling each side of the lariat ring, sterically trapping the tail within the ring in a noncovalent interaction that is resistant to strongly denaturing conditions. We call this structure a lassoed tail.

2.2 Experimental Section

2.2.1 General Materials and Methods. All amino acid derivatives and resins were purchased from Novabiochem (San Diego, CA) except Boc-L-Glu(OAllyl)-OH and Boc-Gly-PAM resin, which were purchased from Neosystems (Strasbourg, France). All other chemicals were purchased from Sigma-Aldrich Chemical Co (St. Louis, MO). Thermolysin was purchased from Calbiochem (Darmstadt,Germany). Analytical gradient HPLC was performed on a Hewlett-Packard 1100 series instrument withdetection at 214 and 280 nm. Analytical HPLC was performed on a Vydac C18 column (5 micron, 4.6 x 150 mm) at a flow rate of 1 mL/min. Preparative HPLC was routinely performed on aWaters DeltaPrep 4000 C18 column (15-20 micron, 50 x 250 mm) at a flow rate of 5 mL/min. Twosolvent systems were used on linear gradients; 1) 0.1% aqueous TFA (solvent A) vs. 90% Acetonitrile plus 0.1% TFA (solvent B); 2) 0.05% aqueous TFA (solvent C) vs. methanol plus 0.05% TFA (solvent D).

2.2.2 Bacterial culture and microcin purification. MccJ25 was purified from minimal M9 media plus1% glucose and 1 mg/ml thiamine as previously described 7 except that the *E.coli* strain, SBG231(AB259) *sjmA1*; spontaneous MccJ25 resistant mutant, β'T931I; [4]) was transformed with the pTUC200 plasmid; a pBR322 backbone with a 13 Kb insert encompassing *McJ* genes *A, B, C,* and *D.*

2.2.3 Activity assays. For the *in vitro* abortive initiation transcription assay, *Ec* RNAP holoenzyme was combined with the T7 A1 promoter-containing DNA fragment, CpA primer, and [α-32P]UTP in the presence and absence of the natural and synthetic microcin variants at indicated concentrations [5]. Reactions were incubated at 37°C for 10 min, and the products were resolved by denaturing PAGE and visualized by autoradiography. For the *in vivo* assay, an agar overlay assay test was used [5]. Double dilutions of MccJ25 variants (10 μL) were applied to a solid LB plate, allowed to dry, and overlaid with 2 mL top agar (0.7%) containing a 200 μL stationary phase innoculate of sensitive (AB259) or resistant (SBG231) *Ec*. The plates were incubated overnight at 37°C and examined for zones of inhibition.

2.2.4 Peptide Synthesis. *General*: All peptides were synthesized according to the *in-situ* neutralization/HBTU activation protocol for Boc SPPS [12]. Following chain assembly, global de-protection and cleavage from the support was achieved by treatment with HF containing 4% v/v *p*-cresol, for 1 hour at 0°C. Following removal of the HF, the crude peptide products were precipitated and washed with anhydrous cold Et_2O before being dissolved in aqueous acetonitrile (50% B) and lyophilized.

2.2.4 a) Peptide 1: Linear Microcin = NH-GGAGHVPEYFVGIGTPISFYG-OH. The peptide was synthesized (0.5 mmole scale) on a PAM resin. Following chain assembly and global deprotection/cleavage, the crude material was purified by RP-HPLC (50-60% D over 60 min.) to give the desired material, peptide **1**, which was characterized by MS; found *m/z* 2125.71 ± 0.05, expected 2125.36, and MS^2; fragmented with appropriate b

and y ions series for this sequence.

2.2.4 b) Peptide 2: Cyclic cysteine Microcin = GGAGHVPEYFVGIGTPICFYG. The synthesis of circular peptide **2** by native chemical ligation [13] required the preparation of a linear peptide α-thioester precursor corresponding to the permuted and mutated microcin sequence; CFYGGGAGHVPEYFVGIGTPI, where the underlined residue indicates a Ser-Cys mutation. This sequence was synthesized (0.25 mmole scale) on a 3-mercaptopropioamide derivatized MBHA resin 13. Following chain assembly and global deprotection/cleavage, the 3-mercaptopropioamide-(DNP-His)-peptide (found *m/z* 2393.16 ± 0.74; expected 2394) was cyclized by dissolving the crude lyophilized material (50 mg, 20.9 mmole) in ligation buffer (6 M guanidine-HCl, 100 mM Tris, pH 8.5, 1 mM EDTA, 0.1 M mercaptoethanesulfoninc acid) to a final concentration of 400 μM. The reaction was deemed complete after stirring at room temperature for 16 hours, at which point the reaction mixture was purified directly by preparative RP-HPLC (27-43% B over 60 min.). This yielded the desired cyclic, fully deprotected peptide (8 mg, 3.8 μmole) with 95% purity and 16% yield (from the 3-mercaptopropioamide-[DNP-His]-peptide), which was characterized by MS; found *m/z* 2123.18 ± 0.71, expected 2123.

2.2.4 c) Peptide 4: Lariat Microcin = GGAGHVPEYFVGIGTPISFYG. The peptide was synthesized (0.5 mmole scale) on a PAM resin. In order to allow on-resin cyclization between the α-amino and Glu8 γ-carboxyl groups, the side-chain of Glu8 was orthogonally protected as an allyl ester. Following chain assembly and N^{α} deprotection, the allyl group was removed by treatment with $Pd(PPh_3)^4$. The desired lactam linkage was then formed on the solid phase using PyBOP (5 eq.) as the activating agent and

either DMF or DMSO as the solvent. Cyclization was monitored using the quantitiative ninhydrin test, which indicated that the coupling efficiency reached a maximum of ~50% after three 15 minute coupling procedures. Following global deprotection/cleavage, the crude material was purified by RP-HPLC (50-60% D over 60 min.) to give the desired material, peptide **4**, (4.5 mg, 2.1 μmole; 2% yield, based on original loading of the resin), which was characterized by MS; found *m/z* 2107.03 ± 0.01, expected 2107.03, and MS^2; fragmented with appropriate b and y ions series for this sequence (Figure 2C). The purified material contained a small amount of a peptidic contaminant with *m/z* 2244.

2.2.5 Thermolysin cleavage of MccJ25.

2.2.5 a) ***Extensive digestion*****:** MccJ25 (10 ng, 4.7 pmole) in 5 μL water, was digested with 1000 U thermolysin (suspended in 10 mM Tris-HCl pH 8.0) for two hours at room temperature. An aliquot of this digestion mixture (1 μL) was diluted ten-fold with aqueous 0.1% TFA, absorbed to a ZipTip microcolumn (Millipore), washed with 20 μl of 0.1% aqueous TFA, and eluted with 1 μL half-saturated 2,5-dihydroxy benzoic acid solution (see above) directly onto a MALDI target for analysis.

2.2.5 b) ***Minimal digestion*****.** MccJ25 (7.4 mg, 3.5 mmole) in 20 μL 18.75 mM ammonium bicarbonate, pH 8.0, was digested with 1 U thermolysin (suspended in 10 mM Tris-HCl pH 8.0) for one hour at room temperature. The digestion was stopped by addition of 25 μL 100 mM EDTA. The digestion mixture was diluted with methanol to a final concentration of 45% methanol, centrifuged at 12,000 rpm for 2 min, and the supernatant warmed at 45°C for 5 min. The singly-cleaved peptide was purified by analytical RP-

HPLC on a 30 min, 45-60% methanol gradient at 45°C to give the desired material at ~83% purity, which was characterized by MS; found *m/z* 2125.01 ± 0.01, expected 2125.36, and MS^2; found appropriate y and b series for a single cleavage between between Phe10 and Val 11. This peptide was sequenced by Edman degradation at the Protein Chemistry Laboratory of The University of Texas Medical Branch, Galveston, TX.

2.2.6 Treatment of Digested MccJ25 with denaturing agents. The remaining extensive digestion mix (from above) was diluted 10-fold with 0.1% aqueous TFA, and 4 μL aliquots were incubated with each of the following treatments at both room temperature and 95°C for 1 hour; 1 M NaOH, 5 M guanidine-HCl, 50% TFA, or 1% SDS (all final concentrations). Each sample was then diluted five-fold with 0.1% aqueous TFA, except the 50% TFA condition, which was diluted with H_20. A portion of each sample (1 μL) was then analyzed by MALDI.

2.2.7 Mass spectrometry. Aliquots (1 μL) of peptide solution (0.5 – 0.1 pmol/μL) were applied onto the MALDI compact disk-sample plate [14] and mixed with equal volumes of half-saturated 2,5-dihydroxy benzoic acid (DHB) in 50% methanol/20% acetonitrile/0.1% TFA and allowed to dry. Accurate masses of the peptides **1**-**4**, and of proteolytic cleavage products, were measured using a prototype MALDI-QqTOF instrument [15]. The MS^2 and MS^3 experiments were performed as desribed [16] using a prototype MALDIquadrupole ion trap [14]. Ion injection durations were 200-400 msec. Resonant excitation was performed using empirically optimized instrument specific

parameters (isolation width: 2.5 Da, normalized collision energy: 35-38%, activation Q: 0.240, activation time: 300 msec). Spectral acquisition times varied from 30 sec to 1!min in order to achieve acceptable signal/noise ratios.

2.2.8 NMR Spectroscopy. MccJ25 (4 mg, 1.9 μmole) and synthetic lariat (4 mg, 1.9 μmole) were dissolved in 500 μL *d*6-DMSO, and spectra were acquired on a Bruker DMX500 spectrometer. Spectra were processed using NMRPipe[17] and analyzed using NMRView [18]. The synthetic linear peptide (2.2 mg, 1.0 μmole) was dissolved in 500 μL *d*3-CD_3OD, and spectra were acquired on a Bruker DPX400 spectrometer, and processed using XwinNMR. All spectra were acquired at 298K.

Sequence-specific backbone and sidechain assignments for MccJ25 were derived from TOCSY and NOESY (250 ms, 175 ms, 100 ms mixing time) data. A total of 274 NOESY peaks were assigned. Crosspeak intensities were measured at a NOESY mixing time of 175 ms, and NOE-derived constraints were generated and sorted into strong (1.8-2.8Å), medium (1.8-3.4Å), or weak (1.8-5.0Å) bins. Protons were not assigned stereospecifically; spectrally differentiable protons were collapsed to pseudoatoms, yielding 192 constraints and distance corrections were handled internally within the software suite CNS [19].

2.2.9 Structure Determination. A total of 192 distance constraints [of which 79 were intra-residue, 59 sequential, 9 medium range (|i - j| < 4 Å) and 45 long range (|i - j| >/= 4

Å)] were utilized in distance geometry simulated annealing calculations [20] using the standard CNS protocol, with a modified protein parameter file to allow for the isopeptide amide bond between Gly1 and Glu8. A total of 75 embedded substructures were generated from the lariat peptide in an extended starting conformation. All embedded starting structures, irrespective of energies, were used to generate 75 final structures by NOE-restrained molecular dynamics simulated annealing, followed by an additional minimization step with reduced dependence on NOE constraints. The 20 lowest energy structures were selected for analysis. No NMR constraint violations >0.5 Å were observed in these 20 structures. A Ramachandran plot revealed that 46.2% of the dihedral angles fell in the most favored region, 53.9% fell in additionally allowed regions, and the remaining 7.7% fell in the generously allowed region.

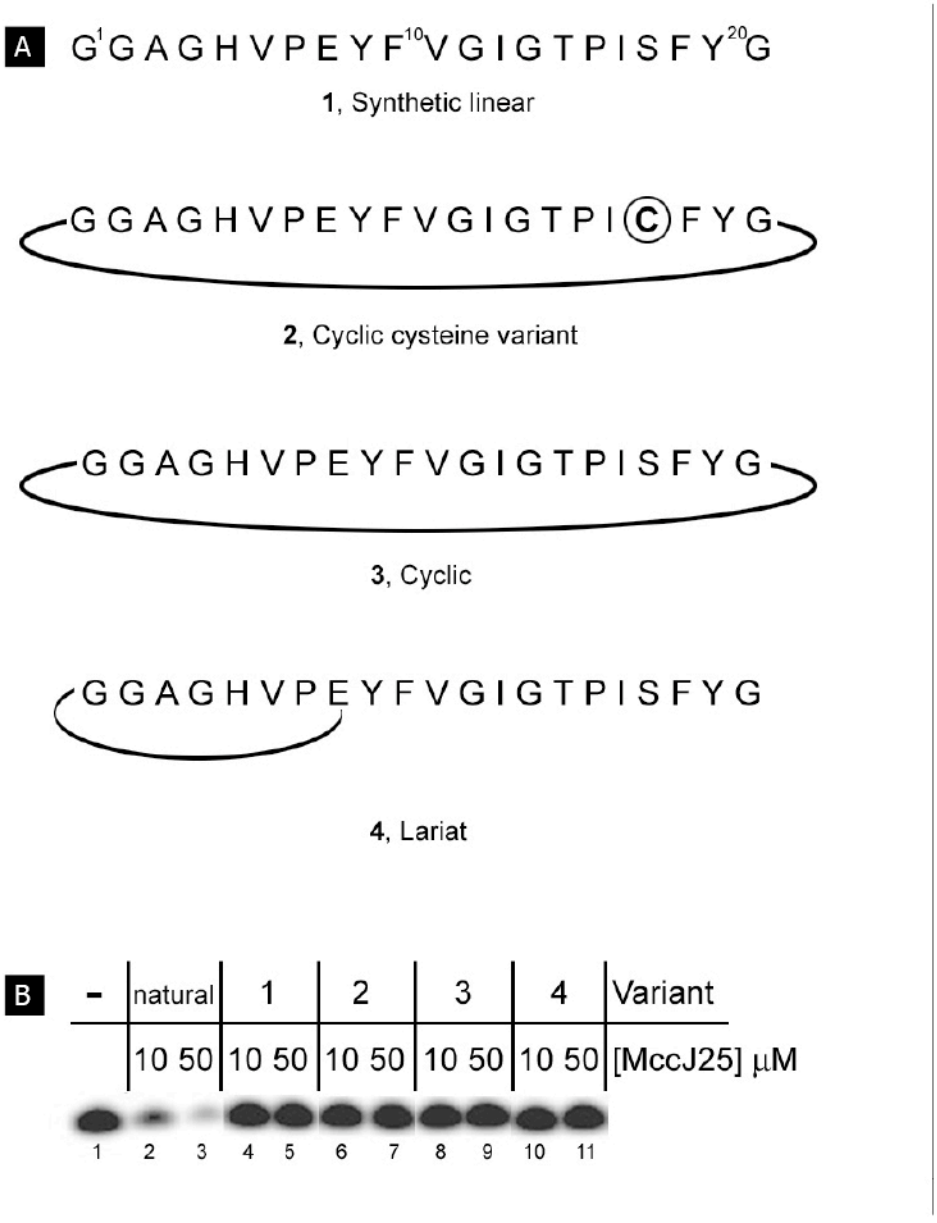


Figure 2.1 Structure and activity of four microcin variants.

(a) Primary structures of four microcin variants, the synthetic linear peptide (1) is numbered according to the encoded sequence of the mature MccJ25. (b) MccJ25 inhibits E. coli RNAP in an in vitro abortive initiation transcription assay, rendering the polymerase unable to extend a CpA primer with [α32P]UTP (compare lanes 2 and 3 with lane 1). The four MccJ25 variants have no effect on RNAP activity (lanes 4–11).

2.3 Results and Discussion

2.3.1 The structure of MccJ25 requires reassessment. The biological activity of totally synthetic versions of MccJ25 was assessed in order to evaluate methods for the preparation of large amounts of highly pure, biologically active MccJ25 required for structural studies of the RNAP/MccJ25 complex (Figure 2.1a). An intramolecular version of native chemical ligation was used to synthesize peptide **2** (Figure 2.1a), where Ser18 of the coded MccJ25 sequence (**1**, Figure 2.1a) was mutated to provide the required N-terminal Cys residue [21]. In addition, cyclic peptide **3** (Figure 2.1a), with the coded MccJ25 sequence obtained by a native chemical ligation scheme followed by desulfurization [22], was a generous gift of P.E.Dawson.

Two tests of biological activity were used, *in vivo* (killing of sensitive bacterial cells) and *in vitro* (inhibition of transcription activity by purified *Ec* RNAP) [5]. Each of the synthetic peptides **1**, **2**, and **3**, were totally inactive *in vivo* (not shown) and *in vitro* (Figure 2.1b). For instance, 10 μM MccJ25 (purified from MccJ25 producing *Ec*) [7] reduced transcription to 43% of the level in the absence of MccJ25 (Figure 2.1b, lanes 1 and 2), while 50 μM MccJ25 nearly completely inhibited transcription (19%, lane 3). In contrast, concentrations of **2** and **3** up to 400 μM (the highest concentration tested) showed no inhibition.

The behavior of MccJ25 and **3** was also compared by analytical reversed phase HPLC. The retention time of MccJ25 was 13.8 min on a 45 min, 50-65% methanol gradient using a C18 column, whereas the retention time of **3** was 21.0 min.

Finally, MccJ25 and **3** were compared by mass spectrometric analysis (Figure 2.2). The *m/z* of MccJ25 and **3** were both measured to be 2107.0, which is 18 Da lower than that of the linear form (peptide **1**). The MALDI-ion trap mass spectrometric fragmentation spectrum (MS^2 spectrum) of MccJ25 was characterized by an intense peak arising through the loss of water from the parent ion and a series of low intensity b-ions (b_8 to b_{20}; Figure 2.2a). This spectra was compared to the MS^2 spectrum of **3** (Figure 2.2b), which displayed a low intensity series of fragment ions characteristic of a cyclic peptide [16], with fragment masses distinct from those of MccJ25.

In summary, peptide **3** did not behave like MccJ25 in three significant tests. First, **3** was inactive in *in vivo* and *in vitro* assays of biological activity at concentrations up to 400 μM, while MccJ25 showed significant transcription inhibition at 10 μM (Figure 2.1b). Second, **3** and MccJ25 had significantly different retention times by reversed phase HPLC. And third, the MS^2 spectra of **3** and MccJ25 were very different (Figures 2.2a, b). Together, these observations conclusively establish that the structure of MccJ25 is not **3**.

2.3.2 Mass spectrometric analysis of MccJ25. Further fragmentation of the smallest observable ion of the MccJ25 MS^2 series, the b_8-ion (Figure 2.2a, magnified in Figure 2.3a), led to an MS^3 spectrum characteristic of the unbranched, 8-residue cyclic moiety shown in Figure 2.3b [16]. These data (Figures 2.3a and 2.3b) are consistent with the presence of an isopeptide bond between Gly1 and Glu8, as confirmed in subsequent experiments (see later). We therefore synthesized a peptide that incorporated this putative Gly1 to Glu8amide linkage ('lariat', peptide **4** of Figure 2.1a). However, the lariat

peptide **4** showed no *in vivo* (not shown) or *in vitro* activity (Figure 2.1b). MS^2 fragmentation of **4** was compared to that of MccJ25 taken under identical instrumental settings (Figure 2.2a versus 2.2c). The same b-ion series was observed for both peptides, but, surprisingly, the intensity of the lariat-derived series was approximately two orders of magnitude higher (when normalized to the intense water loss peak) than that obtained from MccJ25. Comparison of the spectra in Figure 2.2, all of which were derived from peptides with identical molecular masses, demonstrated that the synthetic peptides, **3** and **4**, were structurally distinct from MccJ25. More detailed comparison between the MS^2 spectra of natural MccJ25 and **4** revealed several peaks unique to MccJ25 (colored red in Figure 2.3a). For example, the unique peak at *m/z* 1948.9 appeared to arise from the loss of $G^{14}T^{15}$ from the parent ion at *m/z* 2107.0. This loss can only be explained if residues 16–21 were connected in some way to the N-terminal part of the peptide. We indicate this unknown N- to C-terminal "connection" by the dashed line in Figures 2.2, 2.3, and 2.4. Further fragmentation of the MS^2 fragment at *m/z* 1948.9 (MS^3 experiment) confirmed this connectivity through observation of the sequential losses of $P^{16}I^{17}$ and $S^{18}FYG^{21}$ (Figure 2.3c). These data demonstrate that the C-terminal part of MccJ25 (beyond residue 17) is attached to the N-terminal part (before residue 12) firmly enough to survive both the MALDI process and the subsequent collision induced dissociation. The observation of other unique MS^2 fragments at *m/z* 1778.6 (loss of $G^{12}IGT^{15}$), *m/z* 1738.6 (loss of $G^{14}TPI^{17}$), and *m/z* 1568.7 (loss of $G^{12}IGTPI^{17}$) further supports the existence of the putative connection between the C- and N-terminal parts. This connection is mysterious, as it results in no apparent change in mass.

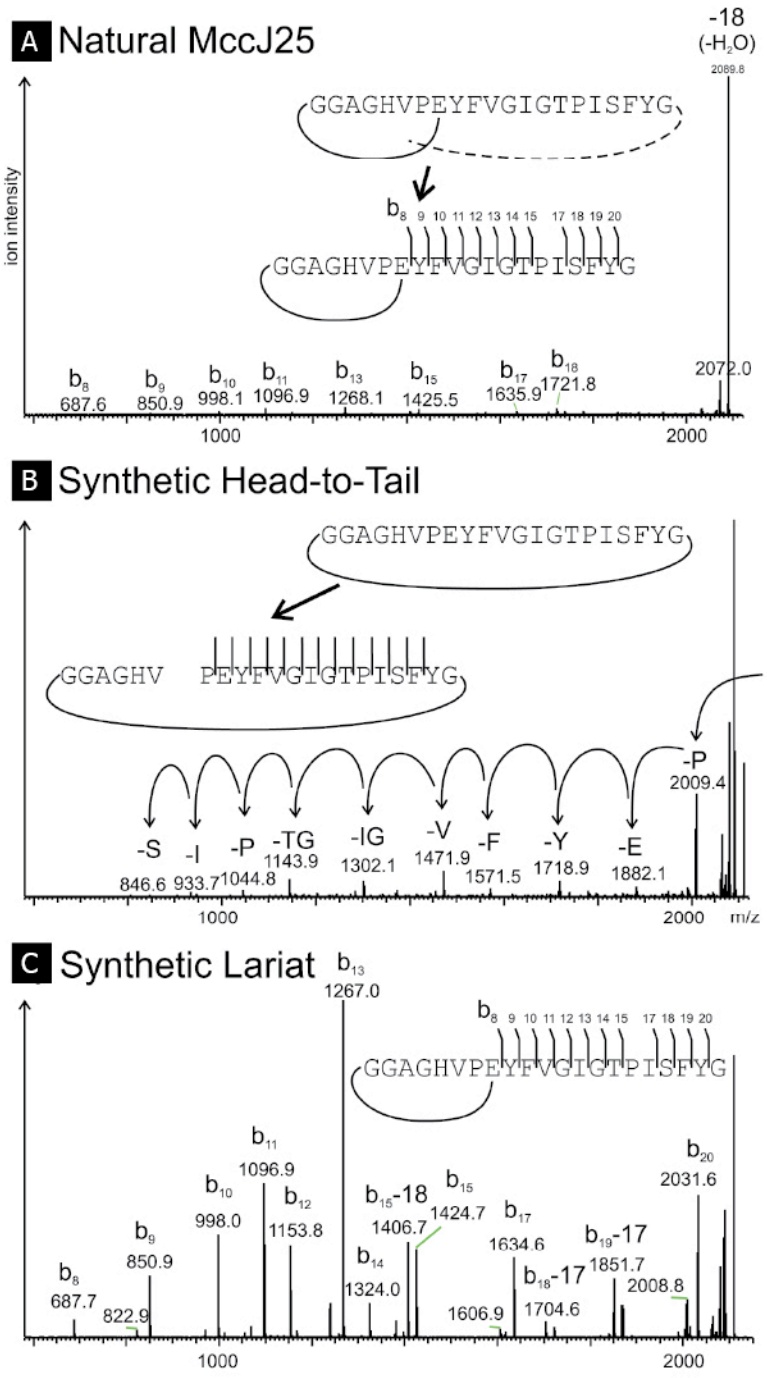


Figure 2.2 Comparison of MALDI-ion trap MS2 spectra obtained from (a) natural MccJ25 and the synthetic microcin-isomers (b) 3 and (c) 4. In each case, the parent ion at m/z 2107.0 was selected and fragmented using identical instrument settings. The arrows in spectrum (b) indicate sequential residue losses.

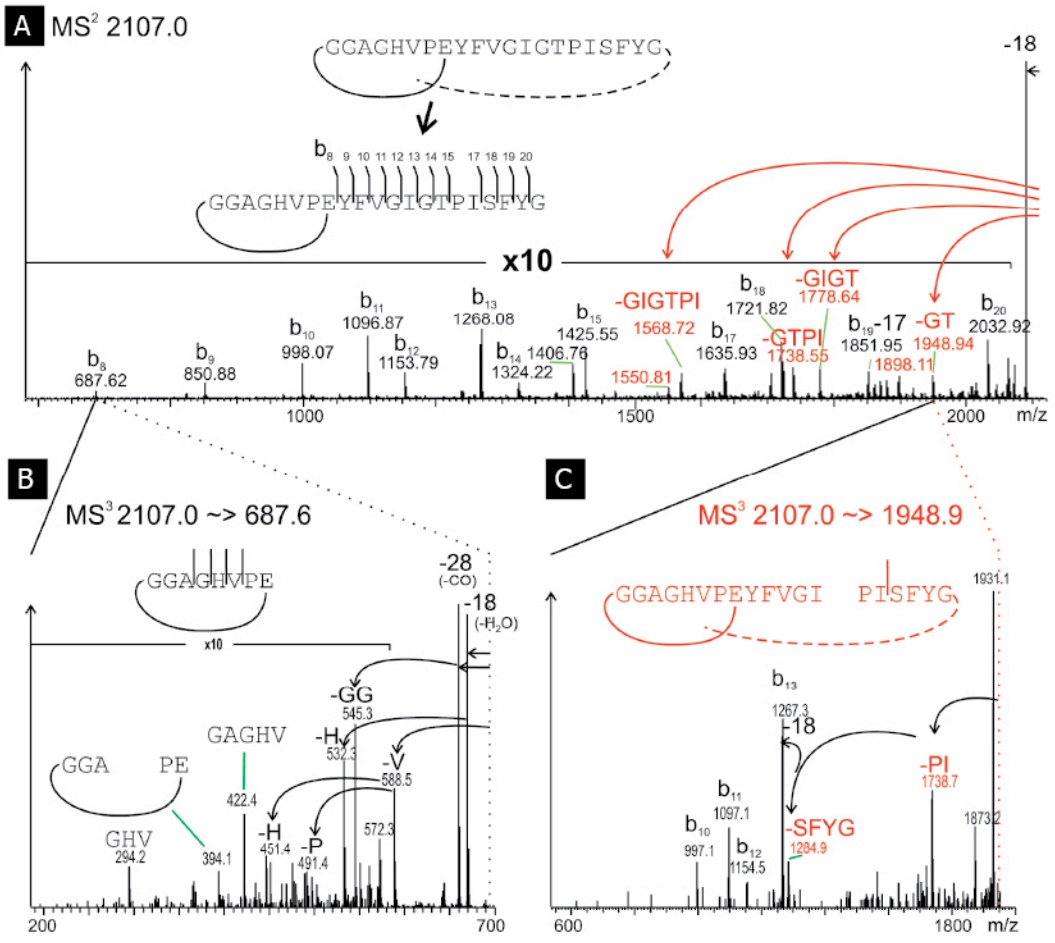


Figure 2.3

(a) MS2 spectrum of MccJ25 from Figure 2a with peak intensities amplified by a factor of 10. Signals that do not belong to the b-ion series are labeled in red. (b) MS3 spectrum of the b8-fragment, characterized by nonsequential residue losses that provide evidence for the existence of the lariat-ring. (c) MS3 spectrum of a nonb-ion fragment that originates from the loss of G14T15 from the MccJ25-parent ion, indicating the existence of the noncovalent connection between the N-to-C-terminal parts of the molecule.

Further insight into structural differences between MccJ25 and **4** was obtained by mass spectrometric analysis of proteolytic fragments derived from thermolysin digestion. Treatment of MccJ25 with low quantities of thermolysin resulted in cleavage of the peptide bond between Phe10 and Val11, yielding a single chemical species with *m/z* 2125.0, as reported previously (in the context of the erroneous structure **3**) [7,23]. This result is inconsistent with the primary structure of **4**, since cleavage between Phe10/Val11 would yield two peptide fragments. Experiments with larger amounts of thermolysin produced three major degradation products at *m/z* 1968.9, 1487.7 and 1015.5 (Figure 2.4a). The ion at *m/z* 1015.5 corresponds to the truncated lariat moiety containing the first eight amino acids and the short $Y^{9}F^{10}$-tail, $\overline{G^{1}GAGHVPEYF^{10}}$. The ion at *m/z* 1487.7 corresponds to the protonated truncated lariat moiety, $\overline{G^{1}GAGHVPE}YF^{10}$, plus the peptide $S^{18}FYG^{21}$. The latter 4-residue peptide contains a free amino- and carboxyl-terminus, as verified by subjecting the peptide at *m/z* 1487.7 to mass spectrometric fragmentation (Figure 2.4e). Likewise, the ion at *m/z* 1968.9 arose from protonated $\overline{G^{1}GAGHVPEYF^{10}}$ + $I^{13}GTPISFYG^{21}$. Thus, we find that the unique connection between the C- and N-terminal portions of MccJ25 is at least partially stable under the conditions of enzymatic digestion. The nature of this connection was further investigated by treatment of the thermolysin digested peptide with denaturing agents. Treatment with 1% SDS, 6 M guanidine-HCl, or 0.1% TFA had no effect on the connection, even after one hour incubations at 95°C, implying the connection was stable to these denaturing treatments.

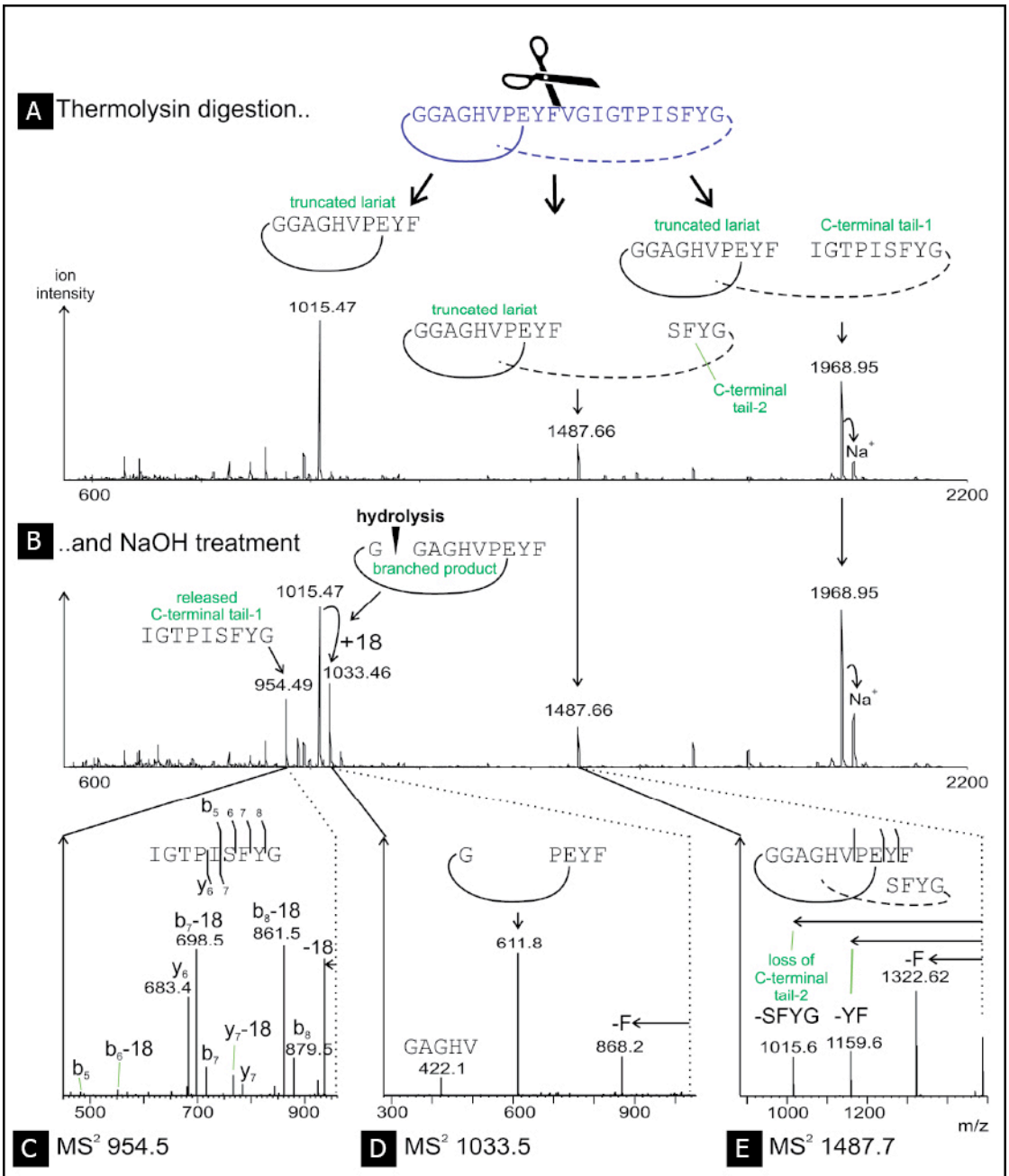


Figure 2.4 MALDI-QqTOF single-stage mass spectra of thermolysin digested MccJ25 – (a)before and (b) after NaOH treatment.

The ion trap MS2 spectra below provide evidence for the release of the C-terminal tail (shown in (c)) due to hydrolysis of the circular portion (as revealed by spectrum (d)). The MS2 analysis of the smallest proteolytic product that bears a portion of the trapped C-terminus is shown in (e).

Additional information on the chemical nature of the mysterious N- to C-terminal connection was obtained by subjecting the thermolysin generated MccJ25-digest to partial hydrolysis with NaOH (Figure 2.4b). The peptide bond between Gly1 and Gly2, within the lariat ring, was labile to base treatment, leaving a branched product with a mass 18 Da higher than that of the truncated lariat (Figure 2.4b, peak doublet at *m/z* 1015.47 and 1033.46). The identity of this branched product was verified by MS2 fragmentation of the *m/z* 1033.46 ion (Figure 2.4d). In addition a peak at *m/z* 954.49 appeared, which we identified as the C-terminal tail, $I^{13}GTPISFYG^{21}$ (Figure 2.4c). These observations are consistent with hydrolysis of the ring, allowing release of the C-terminal tail. This hypothesis was further supported by the lack of peak doublets for the products at *m/z* 1968.95 and 1487.66 (Figure 2.4b), indicating that these structures retained the intact ring and associated C-terminal tail. Finally, under MS^2 conditions, the labile bond between Val6 and Pro7 dissociated (verified by MS^3 experiments), leading to release of $S^{18}FYG^{21}$ (Figure 2.4e).

2.3.3 NMR analysis of MccJ25. The structure of MccJ25 was further analysed using ^{1}H-NMR experiments. No sequential NOE crosspeaks were observed between Gly21 and Gly1, either in our experiments or in previously reported experiments, [7] an observation inconsistent with structure **3**. However, strong NOEs between Gly1 NH and the sidechain protons of Glu8 were observed in NOESY spectra of MccJ25 (Figure 2.5a), consistent with the formation of the lactam linkage.

The ^{1}H-NMR spectra of MccJ25 displayed significant chemical shift dispersion, typical of a highly structured peptide (Figure 2.5b). While the chemical shifts of the NH protons ranged from 6.8 to 9.4 ppm, the NH chemical shifts of residues 10-17 were all clustered between 7.9 and 8.5; the large chemical shift dispersion was due primarily to residues 1-9 and 18-21. This disparity is especially interesting in light of the MS^2 analysis of the thermolysin cleavage experiments, which indicated an association between the lariat residues and the C-terminus of the peptide (Figure 2.4). While peptide **4** also contains the lactam linkage between Gly1 and Glu8, the NH chemical shifts ranged only from 7.3 to 8.6 ppm, typical of a random coil structure (Figure 2.5b). Similarly, the spectra of linear peptide **1** showed NH chemical shifts ranging from 7.7 to 8.5 ppm (data not shown).

The wide dispersion of chemical shifts allowed for the unambiguous assignment of 274 peaks in the NOESY spectrum of MccJ25. Several interesting long-range interactions were observed between the Cterminal tail of the peptide and the N-terminal lariat ring. In particular, Phe19 and Tyr20 define two independent clusters of long-range NOEs. For example, NOE connectivities were observed between the ring protons of Tyr20 and the sidechains of Val6 and Glu8 but not Pro7, and between the Tyr20 NH and backbone protons of 'lariat' residues 5-8. On the other hand, the ring protons of Phe19 interacted solely with the sidechains of 'tail' residues 16-18 and 'lariat' residue Pro7 (but not Val6 or Glu8), and the protons of Phe19 interacted with backbone protons of 'lariat' residues 3-7. NOE crosspeaks between the side-chains of Phe19 and Tyr20 were conspicuously absent.

A model structure consistent with the NMR distance information was calculated. Specifically, 192 NOEderived distance constraints were utilized in hybrid distance geometry simulated annealing, using the covalent structure of **4** as the starting point. An ensemble of 20 structures was generated that converged to a common fold with RMSD for backbone atoms of 1.73 +/- 0.54 Å, and for heavy atoms of 2.38 +/- 0.55Å (Figure 2.6a). Residues 18-20 and 1-8 represented an area of well-defined structure, with backbone (including Glu8 sidechain) RMSD of 0.91 +/- 0.24 Å and heavy atom RMSD of 1.36 +/- 0.27 Å.

The calculated structure closest to the mean of the ensemble was selected as a representative structure (Figures 2.6b, 2.6c, 2.6d). The clusters of interactions involving Phe19 and Tyr20 define an unusual structure, in which the tail residues 18-21 are threaded through the lariat ring formed by the backbone of residues 1-7 plus the sidechain of Glu8. The bulky sidechains of Phe19 and Tyr20 straddle the lariat, protruding on either face of the loop.

Two distinct topologies are possible for the tightened noose structure, depending on which direction the C-terminal tail passes through the lariat ring. In the low-energy structures, only one topology was observed (Figure 2.6). This topology is consistent with the NOE clusters described above, where interactions between tail residues Phe19 and Tyr20 with lariat residues Val6 and Pro7, respectively, define opposite faces of the lariat ring (Figure 2.6b).

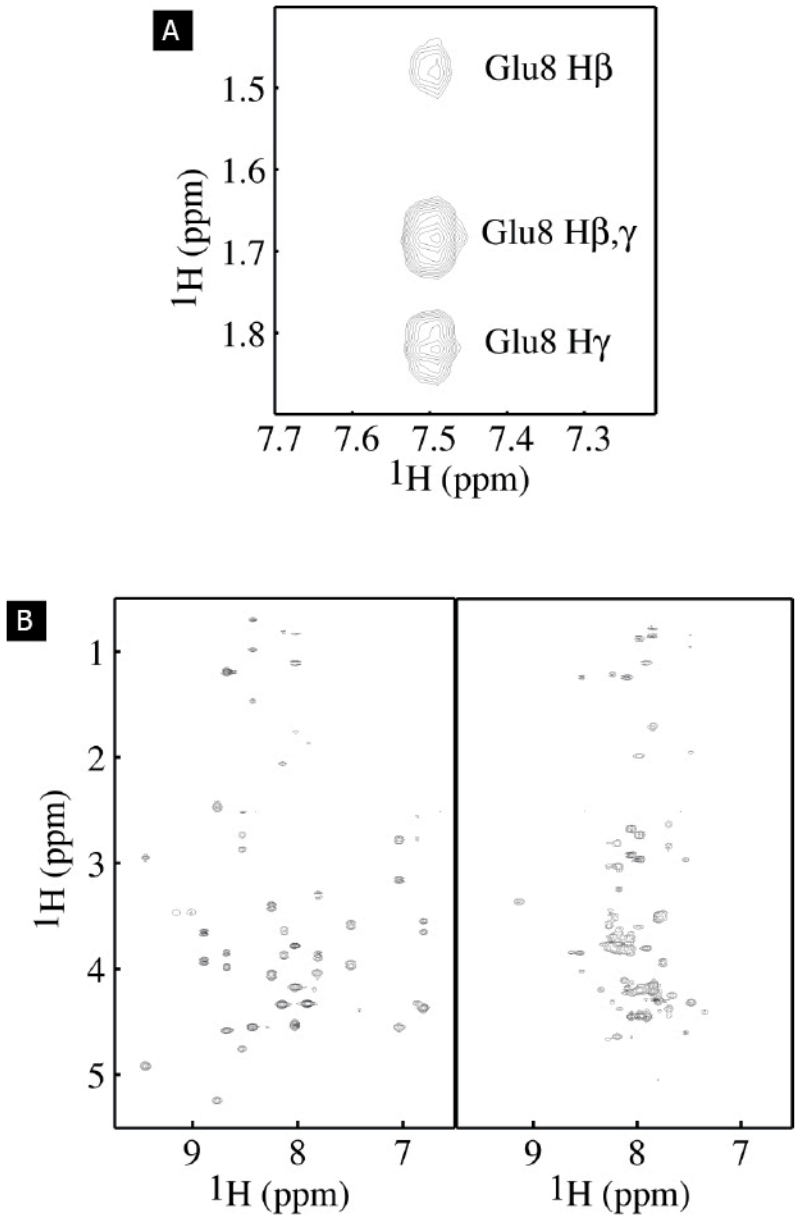


Figure 2.5

(a) NOE cross-peaks between Gly1 HN and Glu8 side chain indicating the isopeptide bond forming the lariat ring. (b) Amide regions of the TOCSY spectra of MccJ25 (left) and synthetic lariat peptide 4 (right). For MccJ25, peaks corresponding to residues 1–9 and 18–21 are in red; residues 10–17 are in black, and the residual solvent ridge is in blue.

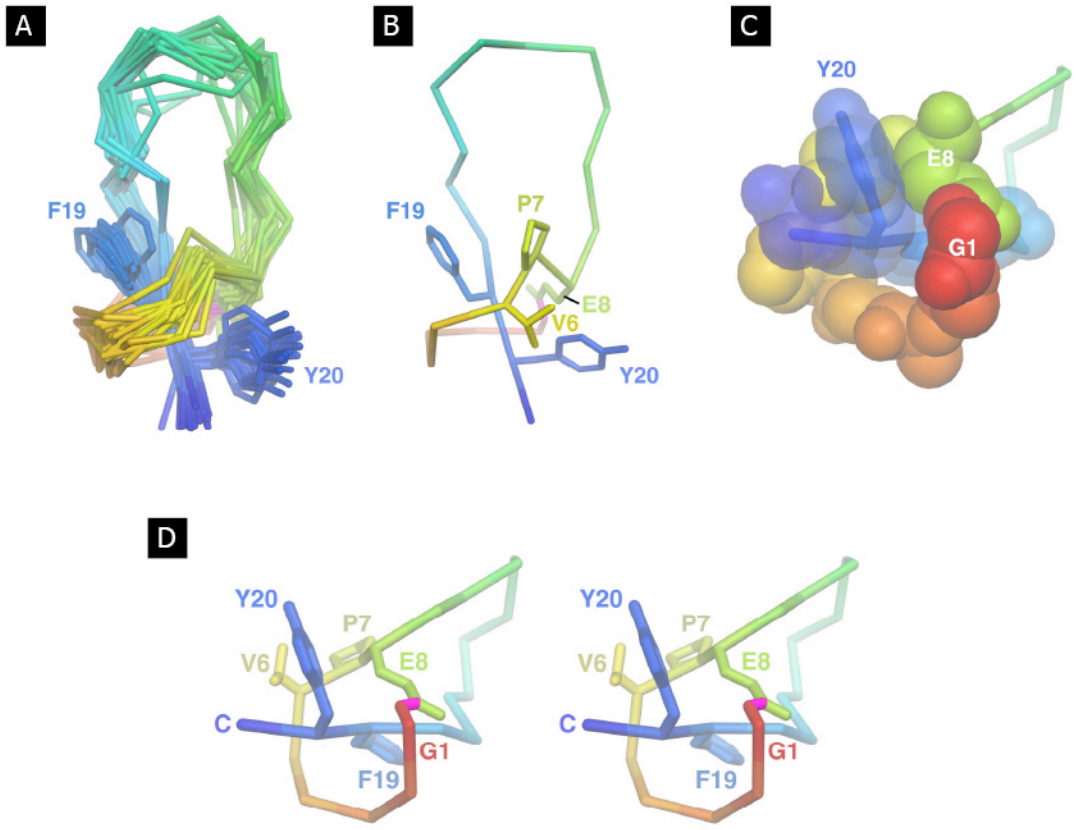


Figure 2.6 NMR-based structural model of MccJ25.

In each view, the peptide chain is colored in a ramp (red-orange-yellow-green-blue) from the N-terminal Gly1 (red) to the C-terminal Gly21 (blue), with the isopeptide bond linking the α-amino group of Gly1 with the γ-carboxyl of Glu8 colored magenta. (a) Ensemble of 20 models resulting from distance geometry simulated annealing calculations. The side chains of E8, F19, and Y20 are shown and labeled. (b) Representative structure, shown in the same view as part a. The side-chains of V6, P7, E8, F19, and Y20 are shown. (c) The lariat ring (G1-E8) and the C-terminal tail (S18-G21) are shown as CPK atoms, with the C-terminal tail rendered transparent so the backbone is visible. Selected side-chains are labeled. (d) Stereoview with selected side-chains shown and labeled.

2.4 Conclusions

Here, we establish that the previously proposed structure for MccJ25 [7,11] is incorrect. Synthetic peptides based on this structure (**2**, with a single Ser-Cys mutation at position 18, and **3**, identical to the proposed structure) differ from MccJ25 in three important ways. First, they are biologically inactive (Figure 2.1b). Second, they behave differently by reversed phase HPLC. Third, they behave differently upon mass spectrometric fragmentation analysis (Figure 2.2). We show that MccJ25 contains a lactam linkage between the α-amino group of Gly1 and the γ-carboxyl of Glu8, forming an 8-residue ring (Gly1-Glu8, the lariat ring). Our data suggests that the C-terminal 'tail' of MccJ25 (Tyr9-Gly21) is trapped within the lariat ring. Release of the C-terminus is prevented by steric hindrance and requires the cleavage of at least one peptide bond, which must occur either within the ring or, based on the NMR data, between residues Phe19 and Tyr20. The tail-free lariat observed at *m/z* 1015.47 (Figures 4a and 4b) may have arisen from thermolysin digestion of the bond between Phe19 and Tyr20, but additional experiments are required to establish this hypothesis.

Our MccJ25 model (Figure 2.6) is consistent with all of our biochemical and mass spectrometric fragmentation results, and is strongly supported by the NMR analysis. In retrospect, the previously published experimental support for **3** as the structure of MccJ25 was not conclusive [7]. We re-evaulate the previous interpretations in light of our new results in the following four points:

1) The N- and C-termini of MccJ25 were blocked to Edman and carboxypeptidase Y sequencing, respectively [6,7]. In fact, the blocked N-terminus is due to the lactam linkage with the g-carboxyl of Glu8. The inability to sequence from the C-terminus by carboxypeptidase Y is not evidence that the C-terminus is engaged in a chemical linkage, but can also be explained by steric considerations since the large protease must be able to bind the C-terminus of its substrate. In our model of the MccJ25 structure, the two C-terminal residues of the peptide are not freely accessible, being engaged in non-covalent interactions with the lariat ring (Figure 2.6).

2) Partial acid hydrolysis of MccJ25 yielded peptide fragments, two of which were analyzed by N-terminal sequencing (fragment H-9^{II}, corresponding to G^{1}GAGHVPEYF^{10}VG, and H-10^{II}, corresponding to Y^{9}FVGIG) and these results were cited in support of **3** [7]. Since these peptide fragments did not contain a peptide bond linking Gly21 with Gly1 (Figure 2.1a), they neither support nor contradict **3**. However, the observed fragments are consistent with our model with the assumption that the harsh acid hydrolysis conditions cleaved the lariat -ring lactam linkage.

3) Thermolysin digestion, which cleaved between Phe10 and Val11, yielded a single peptide [7], as expected from **3**. However, a more thorough analysis of the results of thermolysin digestion (Figure 2.4) supports our model.

4) The most compelling evidence that the structure of MccJ25 corresponded to **3** was the previously published N-terminal sequence analysis, by Edman degradation, of

thermolysin cleaved MccJ25 [7]. The proteolytic cleavage of the peptide bond between Phe10 and Val11 was interpreted to generate a linear peptide with a free N-terminus, allowing N-terminal sequencing to begin at Val11. Blond et al. (1999) [7] report the following sequencing results: $V_{11}GIGTPISFY_{20}G_{21}G_1GAG_4$. Sequencing through the C-terminal Gly21 to Gly1 and on to Gly4 supports **3** and contradicts our model. For this reason, we analysed our own sample of thermolysin-cleaved MccJ25 by Edman degradation. We obtained the sequence from Val11 through Gly21, followed by several cycles indicating Gly, which is expected behavior for the C-terminal residue of a peptide (J. S. Smith, personal communication). Importantly, in cycles subsequent to Gly21, none of the amino acids expected in the N-terminal region of MccJ25 (other than Gly) were observed. For example, Ala or His (expected at positions 3 and 5 of the coded MccJ25 sequence, Figure 1a) were not observed. Thus, we believe our Edman degradation results are consistent with our model, since the precise identification of the C-terminal residue by this technique is imprecise.

Although the previously published NMR structure for MccJ25 [11,23] does not have the correct covalent linkages, the previous structure and our structure (Figure 2.6) are quite similar when considering the well defined portion of our model. The α-carbon backbone of residues 4-8 and 19-20 of our model can be superimposed onto the corresponding atoms of the previous structure with an RMSD of 1.13 Å.

In principal, the final test of the MccJ25 structure should come from its total chemical synthesis and comparison of its biological activity with MccJ25. Due to the nature of the

structure, with its sterically trapped tail, total chemical synthesis has so far proven unsuccessful (data not shown). In this case, the pathway to the final structure must be critical, since the MccJ25 precursor must fold into a pre-defined conformation prior to the lariat ring cyclization to allow steric trapping of the C-terminal tail (Figure 2.6). Pre-folding of the MccJ25 structure may be a function of the McJA pro-protein, or may be assisted by one of the processing enzymes.

Synthesis of active MccJ25 *in vivo* requires the action of two processing enzymes, McJB and McJC [8,24,25,26] , which act on the McJA pro-protein to produce the mature MccJ25 peptide. McJB and McJC are both necessary and sufficient for the *in vitro* maturation of MccJ25 [25,26]. Interestingly, McJC shows sequence and predicted structural homology [27] with a group of related ATP/Mg2+-dependent enzymes including class B asparagine synthetases (that convert aspartic acid to asparagine; [28,29]) and β-lactam synthetase (catalyzes the formation of the b-lactam ring in clavulanic acid) [30,31], suggesting McJC is an ATP/Mg2+-dependent enyzme that catalyzes the formation of the lactam ring found in MccJ25. A sequence analysis of McJB revealed two putative functional domains; an N-terminal domain with weak homology to mammalian adenosine kinases and a C-terminal domain with homology to cysteine preoteases [26] suggesting that McJB cleaves the McJA pro-protein.

The lassoed-tail structure has been observed previously in small peptide inhibitors [32]. In some cases, this architecture appears to be supported by a network of covalent linkages, including disulfide bonds [32,33] , while in one case the tail appears to be sterically trapped

within the ring, as is the case with MccJ25 [34].

An NMR analysis of the thermolysin-cleaved MccJ25 found that the two-chain peptide retained the structure of the native MccJ25 in the structural "core" consisting of the 8-residue lariat and the lassoed four residue tail ($S_{18}FYG_{21}$) whereas the cleaved beta-hairpin made up by residues 11-16 was disordered [35]. This structured core was apparent in our analyses and remained intact in the gas phase (Fig. 2.4).

Two shorter variants of the two-chain peptide produced by extended thermolysin digestion MccJ25(Δ11-12) [36] and MccJ25(Δ13-17) [37] were found to retain *in vitro* transcription inhibition activity but were inactive *in vivo* suggesting that the core is sufficient to bind and inhibit RNAP whereas the beta-hairpin is necessary for import. In support of this a recent systematic structure-activity analysis found that the beta-hairpin residues could be substituted for other amino acids without any effect on in vitro transcription [38]. This structure-activity analysis further defined two groups of residues essential for 1) production; Gly1, Gly2 and Glu8 and 2) *in vitro* inhibition of RNAP; Gly4, Pro7, Tyr9, Phe10, and Phe19 [38]. These two sets of residues were inferred to form surface recognition determinants for 1) the MccJ25 maturation enzymes and 2) RNAP [38] but an alternative interpretation is that the second set of residues are essential for the lassoed-tail structure; both Gly and Pro affect local conformational flexibility and may affect the flexibility of the lariat, and Phe19 could only be substituted by either Trp or Tyr reflecting the importance of a bulky residue to sterically hinder the release of the tail. Thus, the inference that these residues form a hydrophobic recognition patch [38] will have

to be evaluated by further studies.

Modeling studies indicated MccJ25 (Figure 2.6b) is an appropriate size and shape to bind in and block the RNAP secondary channel, as suggested from biochemical and genetic experiments [5, 39]. However, these models were based on the intact MccJ25 and need to be re-evaluated with regard to the rigid core of MccJ25 being necessary and sufficient for RNAP inhibition. The structure of either MccJ25 or the MccJ25 "core" bound to RNAP would provide further insights to the mechanism of inhibition.

Our investigations into the processing pathway of MccJ25 will be discussed in chapter 3.

Acknowledgment. We are indebted to D. Craik for major contributions to our thinking on this problem. We thank R. Salomon for providing plasmids for MccJ25 production, and P. E. Dawson for the gift of cyclized peptide. N-terminal sequencing by Edman degradation was done by J. S. Smith at the University of Texas Medical Branch Biomolecular Resource Facility Protein Chemistry Lab, Galveston, TX. Figure 6 was generated using DINO (http://www.dino3d.org). J. O. was supported by The Rockefeller University's Merck Postdoctoral Fellowship Program. This work was supported by NIH grants RR00862 and CA89810 (B. T. C.), GM38660 (R. L.), GM59908 and GM55843 (T. M.), GM64530 (K. S.), and GM61898 (S. A. D.).

2.5 References

1. Murakami, K.; Darst, S. A. *Current Opinion in Structural Biology* **2003**, *13*, 31-39.

2. Darst, S. A. *Current Opinion in Structural Biology* **2001**, *11*, 155-162. 16

3. Sensi, P. *Reviews of Infectious Diseases* **1983**, *5, Supp. 3*, 402-406.

4. Delgado, M. A.; Rintoul, M. R.; Farias, R. N.; Salomon, R. A. *J Bacteriol* **2001**, *183*, 4543-4550.

5. Yuzenkova, J.; Delgado, M.; Nechaev, S.; Savalia, D.; Epshtein, V.; Artsimovitch, I.; Mooney, R. A.; Landick, R.; Farias, R. N.; Salomon, R.; Severinov, K. *J Biol Chem* **2002**, *277*, 50867-50875.

6. Salomon, R. A.; Farias, R. N. *J Bacteriol* **1992**, *174*, 7428-7435.

7. Blond, A.; Peduzzi, J.; Goulard, C.; Chiuchiolo, M. J.; Barthelemy, M.; Prigent, Y.; Salomon, R. A.; Farias, R. N.; Moreno, F.; Rebuffat, S. *Eur J Biochem* **1999**, *259*, 747-755.

8. Solbiati, J. O.; Ciaccio, M.; Farias, R. N.; Gonzalez-Pastor, J. E.; Moreno, F.; Salomon, R. A. *J Bacteriol* **1999**, *181*, 2659-2662.

9. Zhang, G.; Campbell, E. A.; Minakhin, L.; Richter, C.; Severinov, K.; Darst, S. A. *Cell* **1999**, *98*, 811-824.

10. Korzheva, N.; Mustaev, A.; Kozlov, M.; Malhotra, A.; Nikiforov, V.; Goldfarb, A.; Darst, S. A. *Science* **2000**, *289*, 619-625.

11. Blond, A.; Cheminant, M.; Segalas-Milazzo, I.; Peduzzi, J.; Barthelemy, M.; Goulard, C.; Salomon, R.; Moreno, F.; Farias, R.; Rebuffat, S. *Eur J Biochem* **2001**, *268*, 2124-2133.

12. Schnolzer, M.; Alewood, P.; Jones, A.; Alewood, D.; Kent, S. B. *Int. J. Pept. Protein Res.* **1992**, *40*, 180-193.

13. Camarero, J. A.; Cotton, G. J.; Adeva, A.; Muir, T. W. *Journal of Peptide Research* **1998**, *51*, 303- 316.

14. Krutchinsky, A. N.; Kalkum, M.; Chait, B. T. *Analytical Chemistry* **2001**, *73*, 5066-5077.

15. Krutchinsky, A. N.; Zhang, W.; Chait, B. T. *Journal of the American Society of Mass Spectrometry* **2000**, *11*, 493-504.

16. Kalkum, M.; Lyon, G. J.; Chait, B. T. *Proceedings of the National Academy of Sciences U.S.A.* **2003**, *100*, 2795-2800.

17. Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277-293.

18. Johnson, B. A.; Blevins, R. A. *J. Biomol. NMR* **1994**, *4*, 603-614.

19. Adams, P. D.; Pannu, N. S.; Read, R. J.; Brunger, A. T. *Proceedings of the National Academy of Sciences USA* **1997**, *94*, 5018-5023.

20. Nilges, M.; Clore, G. M.; Gronenborn, A. M. *FEBS Lett.* **1988**, *229*, 317-324.

21. Camarero, J. A.; Muir, T. W. *Chem. Commun.* **1997**, *15*, 1369-1370.

22. Yan, L. Z.; Dawson, P. E. *J. Am. Chem. Soc.* **2001**, *123*, 526-533.

23. Blond, A.; Cheminant, M.; Destoumieux-Garzon, D.; Segalas-Milazzo, I.; Peduzzi, J.; Goulard, C.; Rebuffat, S. *Eur J Biochem* **2002**, *269*, 6212-6222.

24. Solbiati, J. O.; Ciaccio, M.; Farias, R. N.; Salomon, R. A. *J Bacteriol* **1996**, *178*, 3661-3663.

25. Clarke, DJ; Campopiano, DJ. *Org Biomol Chem.* **2007** 5(16), 2564-6.

26. Duquesne S, Destoumieux-Garzón D, Zirah S, Goulard C, Peduzzi J, Rebuffat S. *Chem Biol.* **2007** 14(7), 793-803.

27. Rost, B.; Sander, C. *Proteins* **1994**, *19*, 55-72.

28. Richards, N. G.; Schuster, S. M. *Adv. Enzymol. Relat. Areas Mol. Biol.* **1998**, *72*, 145-198.

29. Scofield, M. A.; Lewis, W. S.; Schuster, S. S. *J Biol Chem* **1990**, *265*, 12895-12902.

30. Miller, M. T.; Bachmann, B. O.; Townsend, C. A.; Rosenzweig, A. C. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 14752-14757.

31. Miller, M. T.; Bachmann, B. O.; Townsend, C. A.; Rosenzweig, A. C. *Nature Struct. Biol.* **2001**, *8*, 684-689.

32. Katahira, R.; Yamasaki, M.; Matsuda, Y.; Yoshida, M. *Bioorganic & Medicinal Chemistry* **1996**, *4*, 121-129.

33. Frechet, D.; Guitton, J. D.; Herman, F.; Faucher, D.; Helynck, G.; Monegier du Sorbier, B.; Ridoux, J. P.; James-Surcouf, E.; Vuilhorgne, M. *Biochemistry* **1994**, *33*, 42-50.

34. Katahira, R.; Shibata, K.; Yamasaki, M.; Matsuda, Y.; Yoshida, M. *Bioorganic & Medicinal Chemistry* **1995**, *3*, 1273-1280.

35. Rosengren, KJ; Blond, A; Afonso, C; Tabet, JC; Rebuffat, S; Craik, DJ. *Biochemistry.* **2004** 43 (16), 4696-702.

36. Bellomio, A; Vincent, PA; de Arcuri, BF; Salomón, RA; Morero, RD; Farías, RN. *Biochem Biophys Res Commun.* **2004** 24;325(4), 1454-8.

37. Semenova, E; Yuzenkova,Y; Peduzzi, J; Rebuffat, S; Severinov K. *J Bacteriol.* **2005** 187(11), 3859-63.

38. Pavlova O, Mukhopadhyay J, Sineva E, Ebright RH, Severinov K. *J Biol Chem.* **2008** 283(37), 25589-95.

39. Mukhopadhyay, J; Sineva, E; Knight, J; Levy, RM; Ebright, RH; *Mol Cell.* **2004**, 14(6), 739-51.

Chapter 3: The biosynthetic pathway of Microcin J25

3.1 Introduction

MccJ25 contains an internal lactam linkage between the α-amino group of Gly1 and the γ-carboxyl of Glu8. The tail (Tyr9-Gly21) passes through the ring (Gly1-Glu8), with Phe19 and Tyr20 straddling each side of the ring, sterically trapping the tail in a noncovalent interaction we call a lassoed tail [1, 2, 3]. The biosynthesis of MccJ25 is dependent on a plasmid-borne synthesis, maturation, and export system that involves just four genes, *mcjABCD,* found within a 4.8 Kb contiguous sequence on the naturally occurring 50 Kb plasmid, see table 3.1 [4] . The *mcjA* gene encodes a 58-residue precursor that includes the 21 amino acids found in the mature form of MccJ25 at the carboxyl-terminus, preceded by 37 amino acids that are cleaved during maturation [5]. The *mcjB* and *mcjC* genes encode maturase enzymes essential for microcin biosynthesis [5, 6]. The *mcjD* gene has high homology to ABC exporters and is believed to export MccJ25, thus conferring host immunity [5]. The processing pathway of MccJ25 is particularly attractive for study as the genetic system has been defined and includes only two maturase enzymes, McjB and McjC (See fig. 3.1).

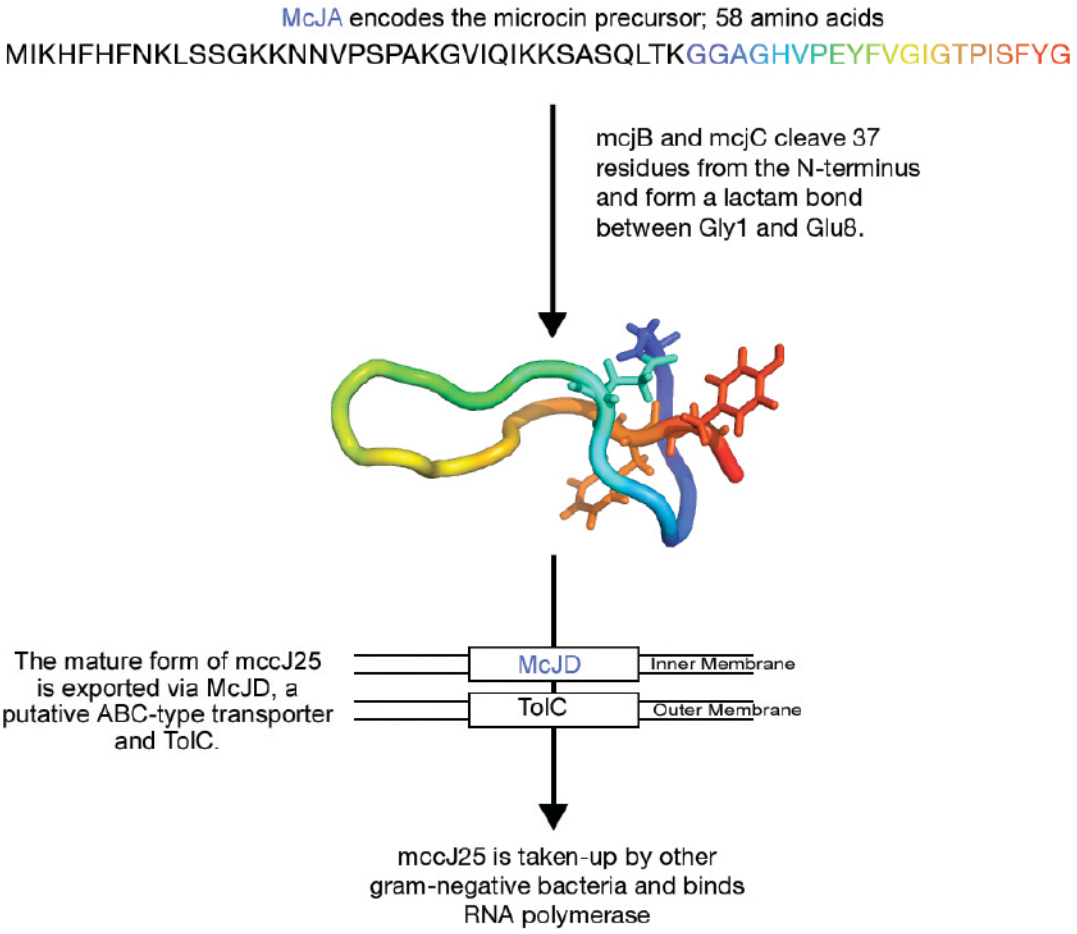


Figure 3.1 Biosynthesis, export and uptake of mccJ25.

Synthesis of active MccJ25 in vivo requires the action of two processing enzymes, mcJB and mcJC which act on the mcJA pro-protein to produce the mature mccJ25 peptide.

MccJ25 bactericidal activity acts against a range of Gram-negative bacterial species [7] and this activity could potentially be exploited as a drug. However, such use would be fairly restricted as uptake requires the FhuA receptor protein and the TonB pathway [8], and activity is confined to Gram negative RNAP. Nevertheless, an understanding of the biosynthetic pathway for MccJ25 may allow the development of novel drugs incorporating the lassoed tail structure. This secondary structure is not novel and has been observed in other peptidyl inhibitors [9] -it allows short peptides to adopt defined tertiary structures akin to mini-proteins and confers an impressive stability, rendering them impervious to denaturing agents, heat and some proteases [3].

Table 3. 1 Characteristics of the ORFs of the MccJ25 cluster.

Protein	**Gene size (bp)**	**Predicted protein size**		**Homology**	**Proposed function**
		Amino acids	**kDa**		
McjA	174	58	6.2	Other lariat peptides	MccJ25 precursor
McjB	624	208	24.6	Adenosine kinase	MccJ25 maturation
McjC	1542	514	58.7	b-lactam synthetase	MccJ25 maturation
McjD	1,740	580	65.4	ABC exporters	MccJ25 export

We set out to biochemically characterize this biosynthetic pathway with the goal of crystallizing the enzymes to gain a detailed knowledge of MccJ25 maturation. The first part of our experimental plan was to set-up an *in vitro* maturase assay using a synthetic peptide corresponding to the full-length McjA precursor protein with McjB and McjC purified as recombinant proteins. We hoped to follow intermediate steps in the maturation using mass spectrometry. We then intended to crystallize McjB and McjC to obtain detailed information about the catalytic steps in biosynthesis.

The *mcjB*, *mcjC and mcjD* genes are encoded on a polycistronic mRNA, the Shine-Delgarno sequence of *mcjC* is before the stop codon of *mcjB*, and only two base pairs separate each of the three open reading frames [6; 10] These characteristics suggest that expression of these genes will be coupled at both the transcription and translation levels, and that the products may functionally and physically interact [11; 12]. Therefore, we tried to co-express these proteins as a binary complex and as a ternary complex with *mcjD*.

Despite extensive efforts, we were unable to overexpress *mcjB* or *mcjC* to the levels required for structural studies. Meanwhile, other groups were successful at setting up *in vitro* McjA maturation systems. A simple *in vitro* assay using a membrane fraction extract from cells carrying a plasmid-borne MccJ25 operon was able to mature a recombinant 58-amino acid mcjA precursor protein to the active 21-residue, cyclized form [13]. A more sophisticated *in vitro* assay using purified hexahistidine-tagged mcjB and McjC demonstrated that McjB and McjC were necessary and sufficient to catalyze the

maturation [6]. Our results are presented below and are discussed in context of the other groups' successes.

3.2 Methods

3.2.1 Peptide Synthesis. *General*: The synthesis of the full length 58-amino acid precursor peptide required the preparation of a 32-amino acid a-thioester peptide and a 26-amino acid N-terminal-cysteine peptide which were joined by native chemical ligation [14]. Both peptides were synthesized according to the *in-situ* neutralization/HBTU activation protocol for Boc SPPS [15]. Following chain assembly, global de-protection and cleavage from the support was achieved by treatment with HF containing 4% v/v *p*-cresol, for 1 hour at 0˚C. Following removal of the HF, the crude peptide products were precipitated and washed with anhydrous cold Et_2O before being dissolved in aqueous acetonitrile (45% v/v) and lyophilized.

Peptide 1: α-Thioester peptide = MIKHFHFNKLSSGKKNNVPSPAKGVIQIKKSA**.**
The peptide was synthesized (0.3 mmole scale) on a 3-mercaptopropioamide derivatized MBHA resin [14]. Following chain assembly and global deprotection/cleavage, the crude material was purified by RP-HPLC (17.1-30.6% acteonitrile over 60 min.) to give the desired material, which was characterized by ESI-MS; found mass 3622, expected 3621.

Peptide 2: Cysteinyl peptide = CQLTKGGAGHVPEYFVGIGTPISFYG. The native chemical ligation reaction requires an N-terminal cysteine residue, therefore the N-terminal fragment was synthesized with a mutated sequence where the underlined residue

indicates a Ser-Cys mutation. The peptide was synthesized (0.3 mmole scale) on PAM resin. Following chain assembly and global deprotection/cleavage, the crude material was purified by RP-HPLC (18-36% acteonitrile over 45 min.) to give the desired material, which was characterized by ESI-MS; found mass 2698, expected 2698.

Native chemical ligation. Only a small scale ligation has been completed. The purified lyophilized peptides were dissolved in ligation buffer (6 M guanidine-HCl, 100 mM Tris, pH 8.5, 1 mM EDTA, 200 mM mercaptoethanesulfoninc acid), peptide 1 was dissolved at a final concentration of 1.6 mM, peptide 2 was dissolved at a final concentration of 5 mM. The reaction was deemed complete after stirring at room temperature for 66 hours, at which point the reaction mixture was analyzed by analytical RP-HPLC (20-40% B over 60 min.). This yielded the desired full-length peptide, which was characterized by MS; found mass 6216 expected 6215.

3.2.2 Cloning and protein expression in standard pET vectors

A plasmid carrying the MccJ25 expression cassette, pTUC100, was a gift from Raul Salomon. The microcin expression cassette was cut from pTUC100 and cloned into a pBR322 backbone via the HindIII/SalI sites to form pKW2.

Each of the MccJ25 maturase genes were PCR amplified from pTUC100 as a single or multi-gene ORF and cloned into a pET28a-based vector between NdeI/BamHI sites as a hexahistidine-fusion protein to generate pSKB2-*mcjB*, pSKB2-*mcjC*, pSKB2-mcj*BC*, pSKB2-mcj*BCD*. An additional co-expression system was made by first cloning the

PCR amplified *mcjC* gene into pet21a to generate pET21a-*mcjC*, and then PCR amplifying this plasmid from upstream of the rbs and subcloning this fragment into pSKB2-*mcjB* to generate pSKB2-*mcjB*-rbs-*mcjC*.

For a typical expression trial *Ec Rosetta2(DE3*) (Novagen) cells were transformed with each of these plasmids and grown in LB plus 50 μg ml^{-1} kanamycin at 37°C to mid-log phase and then expression was induced with 1 m*M* isopropyl β-d-thiogalactopyranoside for 1-4 hours. Cells were harvested by centrifugation, resuspended in SDS-PAGE loading buffer and heated at 95°C and examined for protein content by SDS-PAGE followed by Coomassie blue staining. In other expression trials various factors differed; *Ec BL21(DE3)* was used, MagicMedia (Invitrogen) or M9 media were used, and cells were grown at various temperatures ranging from 18°C - 37°C. In some cases the cells were lysed by sonication in lysis buffer and separated into soluble and insoluble fractions by centrifugation.

3.2.3 Cloning and protein expression of MBP-mcjA

The *mcjA* gene was PCR amplified from pTUC100 and cloned into the pMAL-c2X plasmid (NEB) and expressed in *Ec ER2508* cells as described above for the pET vectors in BL21(DE3).

3.2.4 Cloning and protein expression of protein A-tagged mcjB within pTUC200

A proteinA-tagged McjB (prA-McjB) fusion was made using overlapped PCR fragments. Fragment A consisted of DNA upstream of the *mcjB* gene with 5' HindIII site and 3'

DNA that had homology to fragment B. Fragment B consisted of the *prA* gene with homology to fragment A at the 5' end and homology to fragment C at the 3' end. Fragment C consisted of the *mcjB* gene downstream DNA to the unique BstXI site with a 5' NheI site homology to fragment B. The three fragments were fused together and extended with a short PCR cycle followed by amplification of the entire insert using primers matching to the 5' end of fragment A and the 3' end of fragment C. The blunt ended PCR product was ligated into the pCR-Blunt plasmid then cloned into pKW2 via the HindIII/BstXI sites replacing the untagged mcjB within the mccJ25 expression cassette to generate pKW4.

Coli#55_prA::rpoC are an environmental Ec strain modified to encode protein-A tagged *rpoB* described in Savalia et al. J Mol Biol. 2008 377(3): 774–789. *Ec MG1655 or Coli#55_prA::rpoC* cells were transformed with pKW4 and grown in mimimal M9 media at 37°C for 12 hours beyond stationary phase to induce microcin expression. Cells were harvested by centrifugation, washed once with ice-cold 10% (v/v) glycerol, and frozen as pellets in liquid nitrogen. Cells were cryogenically disrupted with a Retsch MM301 mixer mill (Retsch) that was maintained at liquid nitrogen temperature, and stored at minus 80°C.

3.2.5 Western blot analysis of protein A-tagged mcjB

Frozen cell lysate (0.5 g) were suspended in 2.5 ml of extraction buffer (20 mM Hepes (pH 7.4); 2 mM $MgCl_2$; 150 mM NaCl) supplemented with 200 μg/mL PMSF and 4

μg/mL pepstatin with or without detergent; SDS v/v 0.1%, 1%, octylglucoside v/v 0.5%, 1%, 2%, Tween v/v 0.1%, 0.5%, 1%, Triton v/v 0.1%, 0.5%, 1%.

Resuspended cells were mixed with 0.002 % (w/v) DNase I (Sigma-Aldrich) and slowly spun at room temperature for 15 minutes. The lysate was clarified by centrifugation for 10 min at 13,000 rpm; the supernatant was decanted and centrifuged for another 10 min at 13,000 rpm and decanted again to generate the soluble fraction and insoluble pellet. The two fractions were dissolved in SDS-PAGE loading buffer and heated at 95 °C for five minutes and then resolved by SDS-PAGE with 4-12 % (w/v) Bis-Tris polyacrylamide gels (Invitrogen). Proteins were visualized by Coomassie blue staining.

These protein samples were also examined by Western Blotting with anti-prA antibodies using the Western Lightning chemiluminescence kit (Perkin Elmer).

3.3 Results

3.3.1 McjA was insoluble when expressed as a hexahistidine fusion protein, and was degraded when expressed as a MBP fusion protein.

We first tried to obtain the full-length McjA protein by expressing it as a hexahistidine fusion protein (his_6-McjA) with the intention of purifying it with a Ni+-affinity column. Unfortunately, the vast majority of his_6-McjA protein was insoluble under standard inductions conditions and when expressed at low temperature (see fig. 3.2).

To obtain soluble McjA protein we cloned and expressed *mcjA* as a maltose-binding-protein fusion (MBP-McjA). Initial expression trials indicated that MBP-McjA was expressed to a reasonable level using a standard induction protocol. However, it was apparent from the SDS-PAGE analysis that the mcjA portion of the MBP-McjA fusion protein was degraded during expression such that one could see a doublet of bands on a SDS PAGE gel corresponding to MBP-McjA (48.7 kDa) and MBP alone (42.5 kDa, see fig. 3.3, lane 3). In the subsequent attempts to purify MBP-McjA we found that we were unable to obtain the desired full-length MBP-McjA.

Figure 3.2 McjA is insoluble even when expressed at low temperature.

A) Time trial of expression at 37°C. BL21(DE3) cells induced to express His6-mcjA were harvested at various timepoints and lysed by sonication. The soluble and insoluble fractions were analyzed by SDS-PAGE (Phast gels 8-25% acrylamide, stained with Coomassie blue). Lanes: 1) Uninduced cells, soluble fraction, 2) 1 hour induction soluble fraction, 3) 1 hour induction insoluble fraction, 4) 3 hour induction, soluble fraction, 5) 3 hour induction, insoluble fraction, 6) 4 hour induction, soluble fraction, 7) 4 hour induction, insoluble fraction,

B) Temperature trial of expression. BL21(DE3) cells were grown at various temperatures, induced to express His6-mcjA harvested and lysed by sonication. The soluble and insoluble fractions were analyzed by SDS-PAGE (Phast gels 8-25% acrylamide). Lanes: 1) 37°C, 4 hour induction soluble fraction, 2) 37°C, 4 hour induction insoluble fraction, 3) 30°C, 5 hour induction soluble fraction, 2) 30°C, 5 hour induction insoluble fraction, 1) 25°C, 11 hour induction soluble fraction, 2) 25°C, 11 hour induction insoluble fraction.

The samples corresponding to soluble and insoluble fractions are indicated by S and I respectively. Red arrowheads indicate the bands assuned to correspond to His_6mcjA that has a molecular weight of 8.6 kDa.

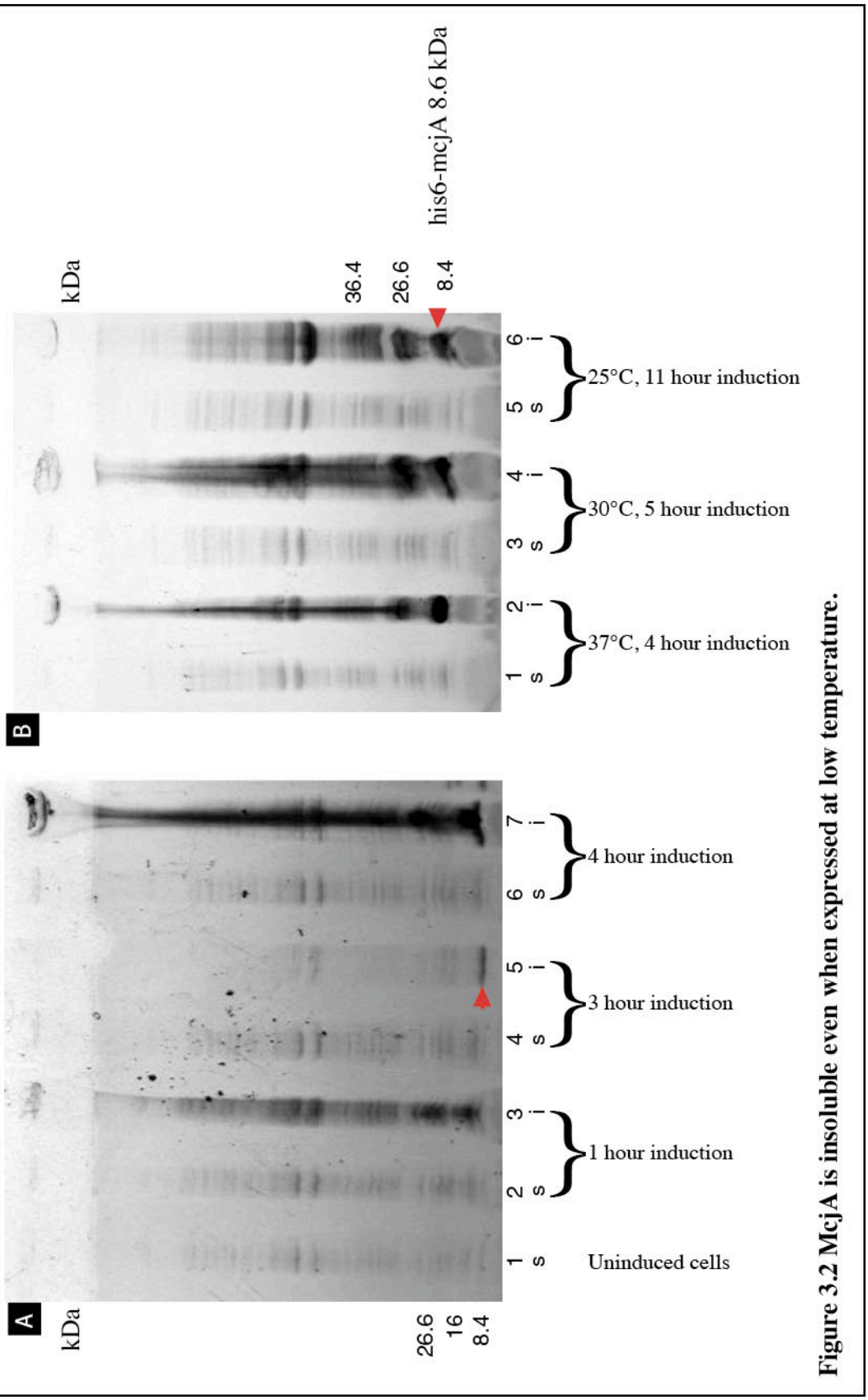


Figure 3.2 McjA is insoluble even when expressed at low temperature.

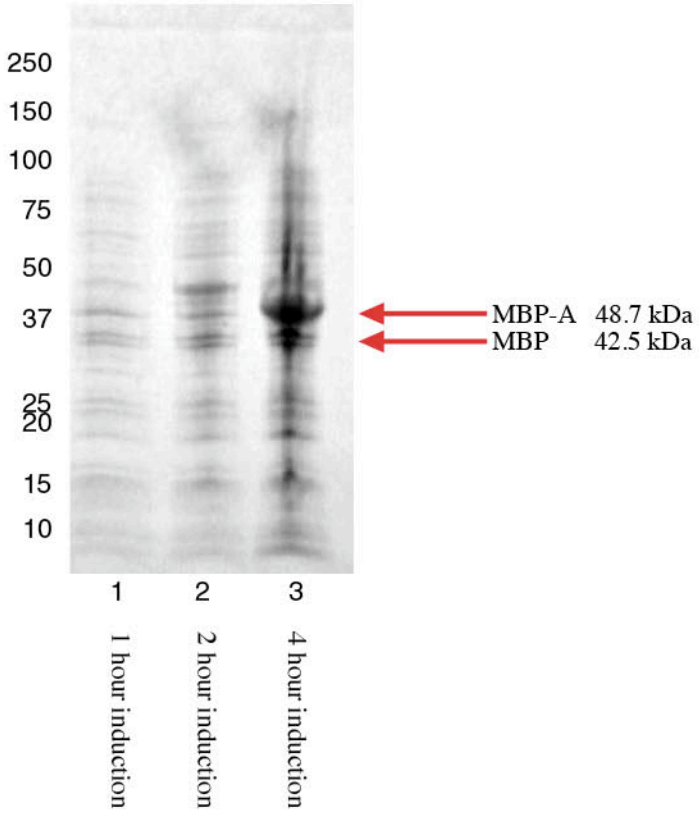
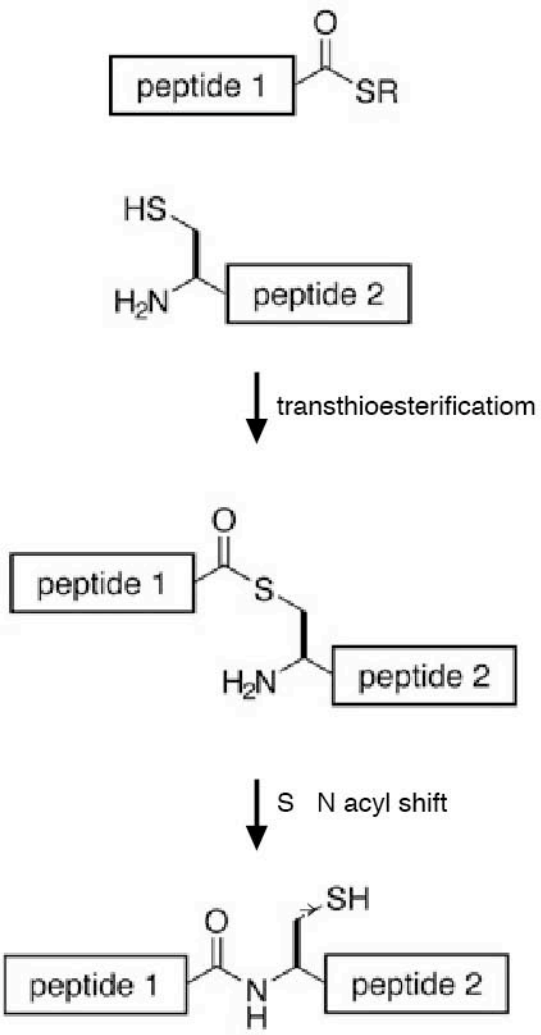


Figure 3.3 MBP-McjA fusion protein is rapidly degraded.
ER2508 *Ec* cells were induced to express a MBP-mcjA fusion protein, then harvested by centrifugation and analyzed by SDS-PAGE (Phast gels 8-25% acrylamide stained with Coomassie blue). Lanes: 1) MBP-mcjA uninduced 2) MBP-mcjA 2 hours after induction 3) MBP-mcjA 4 hours after induction.

Peptide 1: Thioester peptide = **MIKHFHFNKLSSGKKNNVPSPAKGVIQIKKSA**

Peptide 2: Cysteinyl peptide = **C**QLTKGGAGHVP**E**YFVGIGTPISFYG



Final peptide: McjA =
MIKHFHFNKLSSGKKNNVPSPAKGVIQIKKSAS**QLTKGGAGHVP**E**YFVGIGTPISFYG**

Figure 3.4 Peptide synthesis scheme.

Two peptides (Peptide 1 and 2) were synthesized by SPPS with the requisite terminal groups for native chemical ligation as shown in the reaction scheme. After ligation these two peptides formed the final peptide corresponding to McjA the 58-residue MccJ25 precursor.

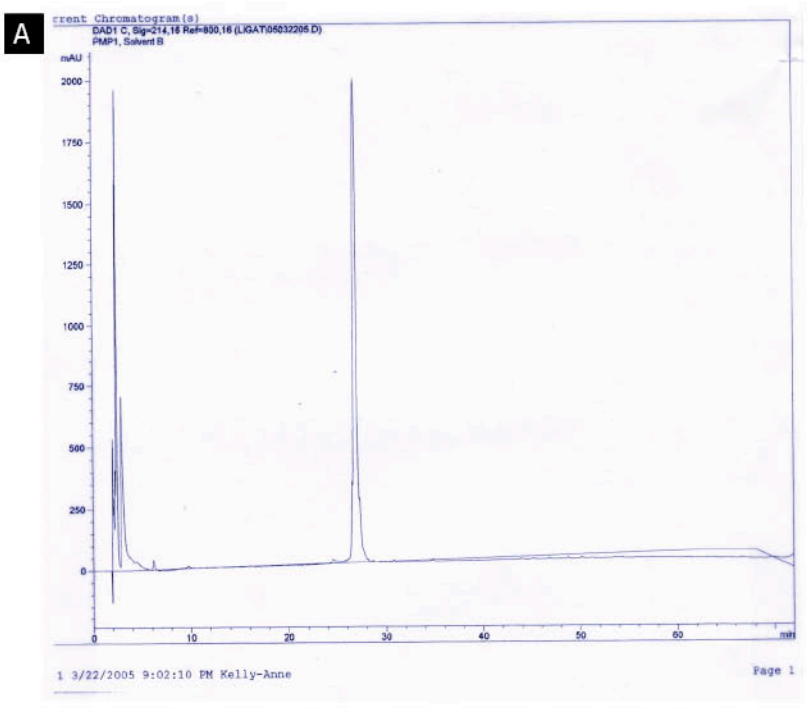
3.

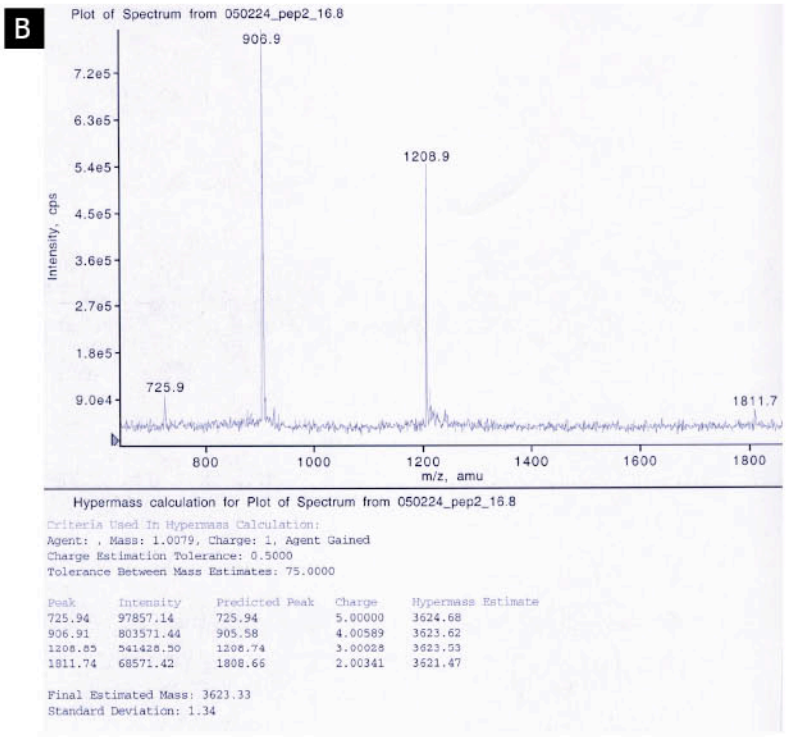
3.2 McjA was synthesized in two pieces and then ligated using native chemical ligation to produce the full-length peptide.

As we were unable to obtain McjA as an expressed recombinant fusion protein, we decided to use solid phase peptide synthesis. McjA is a 58-amino acid peptide that would be technically difficult to synthesize as a single peptide. Therefore, we synthesized the peptide in two pieces; peptide 1, a 32-residue thioester corresponding to the amino-terminus and peptide 2, a cysteine-substituted 26-residue peptide corresponding to the carboxyl-terminus (see fig. 3.4). Both peptides were obtained in high purity as assessed by analytical HPLC (see fig. 3.5 and 3.6). The purified peptides were joined by native chemical ligation and the final product was purified by preparative HPLC to obtain 2.3 mg of full-length mcjA at about 70% purity (see fig. 3.7, 3.8 and 3.9). (This final yield was from a small scale preparation using only a portion of the purified peptides.) This peptide was sent to our collaborator (Konstantin Severinov) for further characterization.

3.3.3 The *mcJC* gene was incorrectly defined in the literature and actually extends beyond the published stop codon.

The microcin J25 operon was sequenced [10], however, after a number of expression constructs showed an apparent frameshift error in the *mcJC* gene we sent the pTUC200 plasmid for additional sequencing. We found that the mcjC gene extended to 1542 bp and encoded 514 amino acids rather than the previously published 1326 bp. This extension results in just 2 bp separating the *mcjC* gene from the *mcjD,* just as only 2 bp separates the *mcjB* and *mcjC* genes.





Hypermass calculation for Plot of Spectrum from 050224_pep2_16.8

Criteria Used In Hypermass Calculation:
Agent: , Mass: 1.0079, Charge: 1, Agent Gained
Charge Estimation Tolerance: 0.5000
Tolerance Between Mass Estimates: 75.0000

Peak	Intensity	Predicted Peak	Charge	Hypermass Estimate
725.94	97857.14	725.94	5.00000	3624.68
906.91	803571.44	905.58	4.00589	3623.62
1208.65	541428.50	1208.74	3.00028	3623.53
1811.74	68571.42	1808.66	2.00341	3621.47

Final Estimated Mass: 3623.33
Standard Deviation: 1.34

Figure 3.5 Peptide 1 was purified by HPLC.

A) Analytical HPLC trace and B) ESI-MS spectrum for Peptide 1: α-Thioester peptide, mass 3622, expected 3621.

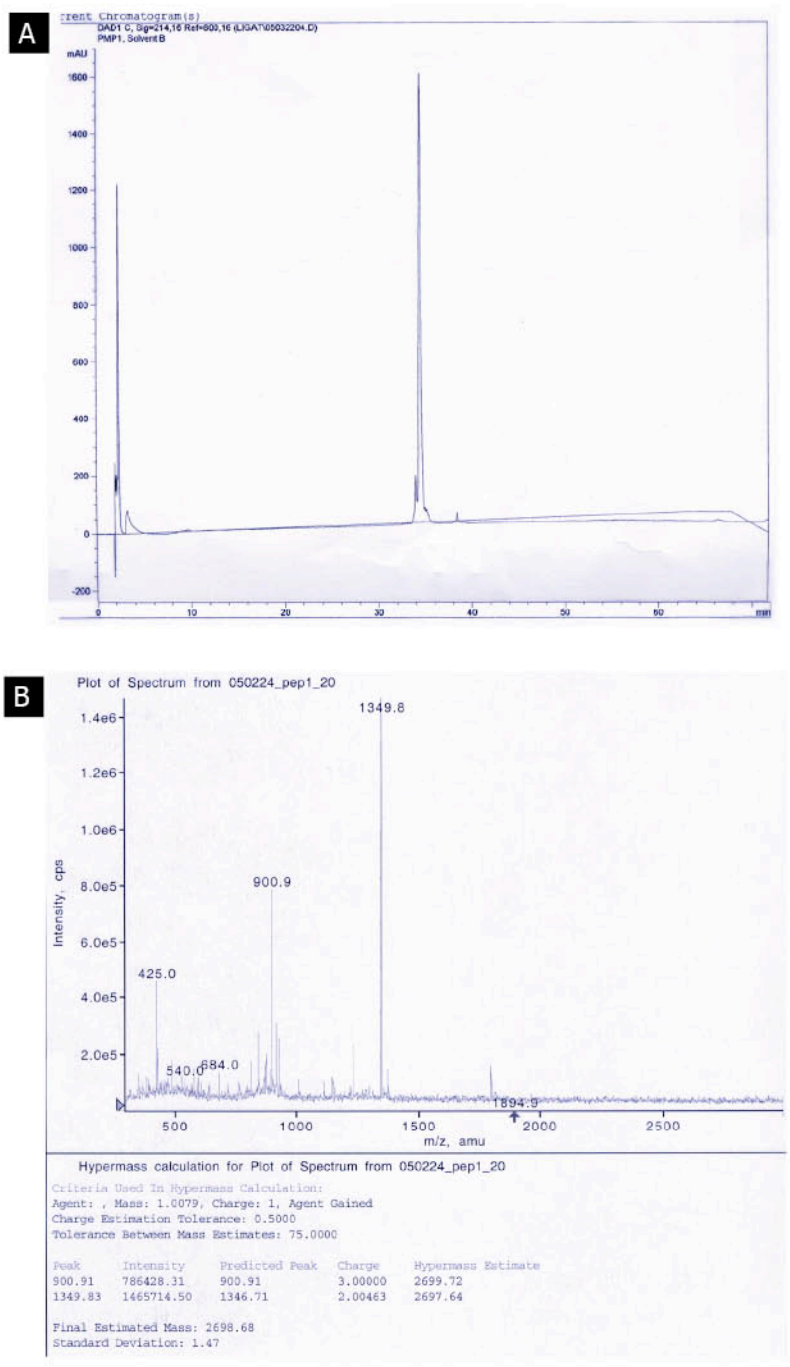


Figure 3.6 Peptide 2 was purified by HPLC.

A) Analytical HPLC trace and B) ESI-MS spectrum for Peptide 2: Cysteinyl peptide mass 2698, expected 2698.

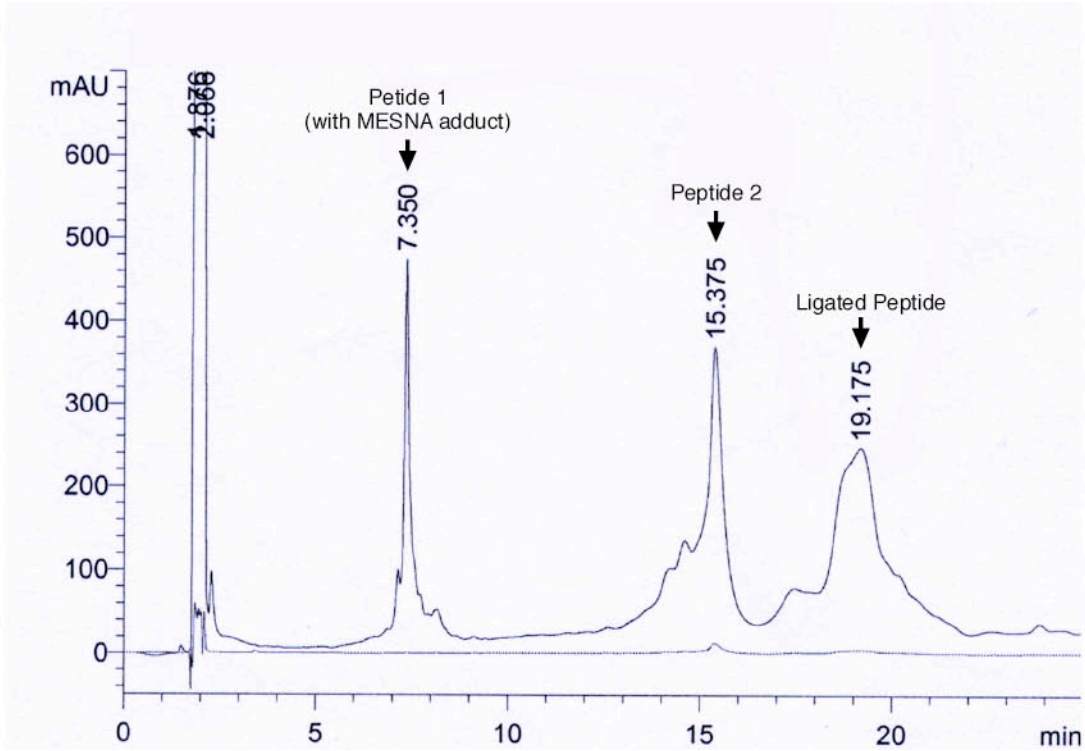


Figure 3.7 Analytical HPLC trace of the native chemical ligation reaction of Peptide 1 and Peptide 2.
The reaction mix analyzed by HPLC shows that the reaction was incomplete. UV absorbance peaks (wavelength = 216 nm) corresponding to each of the peptides and the final ligated peptide are indicated.

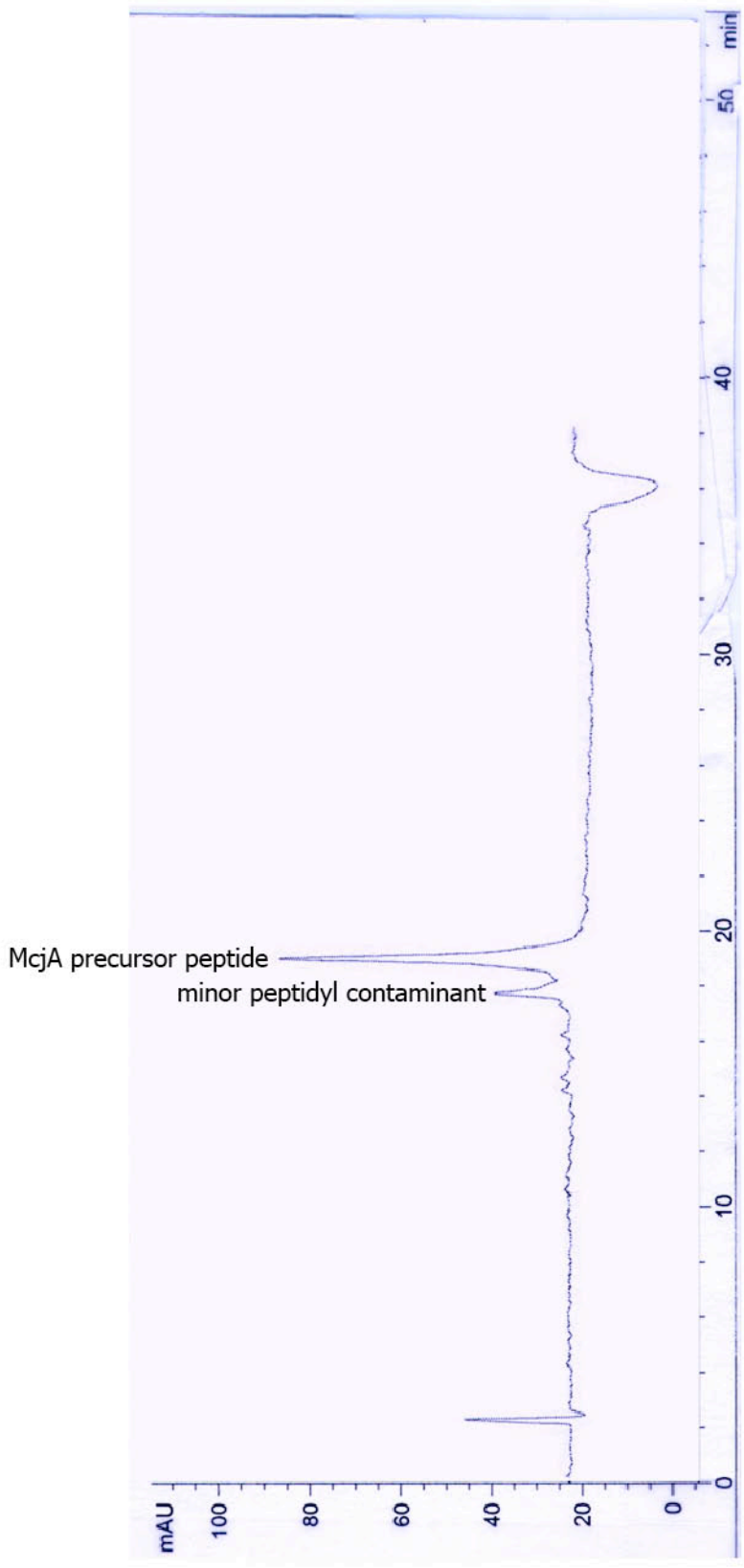


Figure 3.8 Analytical HPLC trace of the full length McjA precursor peptide.

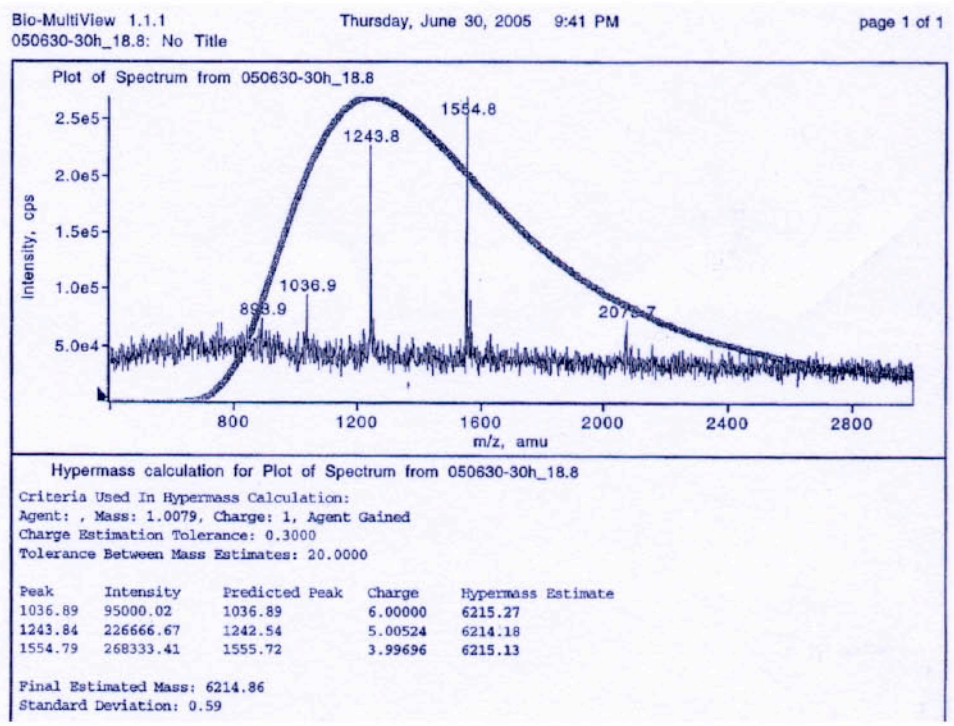


Figure 3.9 ESI-MS of the full length McjA precursor peptide.
Measured mass 6216 Da, expected mass 6215 Da.

3.3.4 The expression of recombinant MccJ25 maturase enzymes was undetectable in cell lysates analyzed by SDS-PAGE and stained with Coomassie Blue.

Each of the MccJ25 maturase genes (*mcjB* & *mcjC*) were cloned into expression plasmids as both single and multi-gene operons as described in table 3.2. These constructs were tested for expression under standard conditions or in non-standard conditions with various temperatures and various amounts of inducer (IPTG), and even substituting LB with the autoinducing Magicmedia (Invitrogen), but there was no expression detectable by Coomassie staining of cell lysates separated by of SDS-PAGE. An example expression trial is shown in figure 3.10.

3.3.5 Rare codon usage in *mcjB* and *mcjC* is high.

The poor expression of *mcjB* and *mcJC* was surprising considering that these are *Ec* proteins being expressed in *Ec* cells. Such poor expression is sometimes associated with heterologous expression when the origin genome has a different codon bias to the expression host. Therefore, we analyzed the codon usage of these genes by comparing codon usage to that in genes highly expressed during exponential phase (*E.coli* Codon Usage Analysis 2.0 by Morris Maduro).

The *mcjB* gene contains 208 codons, 25% of these are below a 10% threshold for codon usage. This includes 5 pairs of adjacent rare codons and 1 cluster of three in a row. The *mcJC* gene contains 531 codons, 23% these are below a 10% threshold for codon usage

with 17 adjacent pairs of rare codons, 3 clusters of three in a row and 1 cluster of four in a row. For comparison, the subunits of the *Ec* RNAP contained far fewer rare codons with each subunit scoring between 1% and 7% (see Table 3.3).

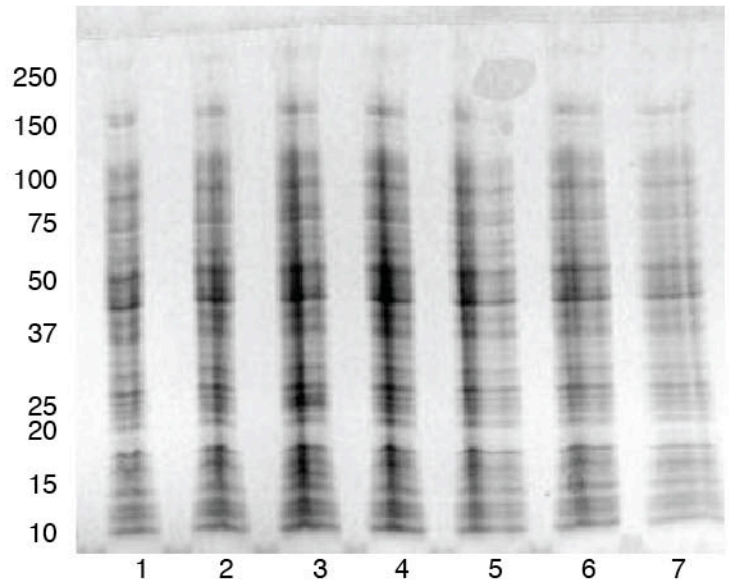


Figure 3.10 There is no apparent expression of recombinant hexahistidine-tagged microcin proteins

In this example Rosetta2(DE3) cells transformed with each of the recombinant expression vectors were grown in autoinducing MagicMedia at 37°C and after 24 hours the cells were harvested by centrifugation, resuspended in SDS-loading buffer and analysed by SDS-PAGE. Lanes: 1) Rosetta2(DE3) cells (control) 2) His6-McjA 3) His6-mcjB 4) His6-mcjC 5) His6-mcjBC 6) His6-mcjBCD 7) His6-mcjBrbsC

Table 3.2 Plasmids used in the study of the MccJ25 biosynthesis pathway.

Name	Description	Expression products
pTUC100	pTUC200 HindIII digest fragment inserted into pBR322.	MccJ25 operon
pKW2	pTUC100 HindIII/SalI digest fragment inserted into pBR322.	MccJ25 operon
pKW4	pKW2 with a substituted protein A-tagged McjB	MccJ25 operon with proteinA-tagged McjB
pMAL_A	*mcjA* inserted into the NEB pMAL-C2X plasmid	MBP-McjA
pSKB2-mcjA	*mcjA* inserted into the Novagen pet28a-based plasmid, pSKB2, via NdeI/BamHIII sites.	McjA
pSKB2-mcjB	*mcjB* inserted into the Novagen pet28a-based plasmid, pSKB2, via NdeI/BamHIII sites.	McjB
pSKB2-mcjC	*mcjC* inserted into the Novagen pet28a-based plasmid, pSKB2, via NdeI/BamHIII sites.	McjC
pSKB2-mcjBC	*mcjB and mcjC* inserted as a polycistronic sequence into the Novagen pet28a-based plasmid, pSKB2, via NdeI/BamHIII sites.	McjB & McjC
pSKB2-mcjB-rbs- mcjC	*mcjB* inserted into the Novagen pet28a-based plasmid, pSKB2, via NdeI/BamHIII sites.	McjB & McjC
pSKB2-mcjBCD	*mcjB, mcjC and mcjD* inserted as a polycistronic sequence into the Novagen pet28a-based plasmid, pSKB2, via NdeI/BamHIII sites.	McjB & McjC & McjD

Table 3. 3 Rare codon bias in the microcin maturase and export enzymes.

Gene	Size of protein; number of codons, molecular weight (kDa),	Fraction of codons below 10% usage in highly expressed proteins.	Number of adjacent pairs of rare codons	Number of adjacent triplets of rare codons	Number of additional clusters
mcjB	209, 24.6	25%	5	1	0
mcjC	514, 58.7	23%	17	3	1 cluster of 4
mcjD	581, 65.4	30%	25	12	1 cluster of 5, 1 cluster of 6
Ec rpoA	329, 36.5	7%	2	1	0
Ec rpoB	1343, 150.6	2%	1	0	0
Ec rpoC	1407, 155.1	1%	1	0	0
Ec rpoZ	92, 10.2	3%	0	0	0
Ec rpoD	614, 70.2	3%	1	0	0

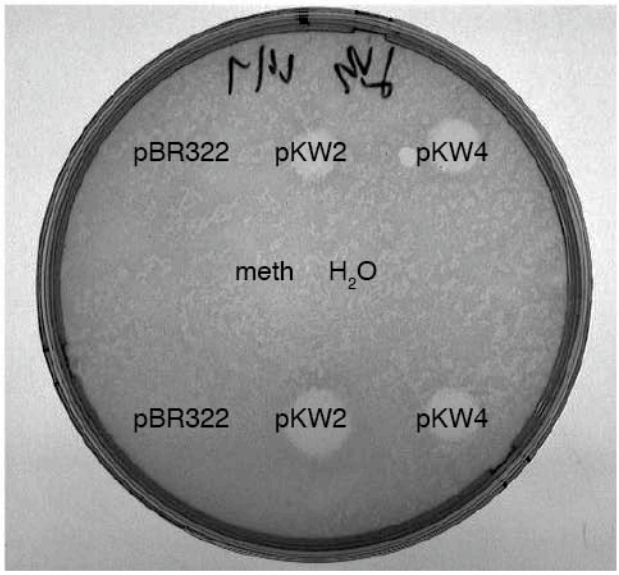
3.3.6 Western blotting could follow protein A-tagged McjB expression.

One method to visualize low abundance protein expression is to use Western Blot analysis. Rather than raise antibodies to the MccJ25 maturase enzymes, the McjB protein was expressed as a protein-A fusion protein (prA-McjB) that could be probed using anti-prA antibodies.

The *prA-mcjB* fusion gene was subcloned into the native MccJ25 expression cassette to generate plasmid pKW4. To test for competent MccJ25 expression this construct was transformed into MG1655 cells, grown to stationary phase and the growth media was applied to a C8 cartridge and a MccJ25 containing fraction was eluted with 100% methanol. This fraction was confirmed to contain MccJ25 by mass spectrometric analysis (data not shown) and by bioassay (see fig 3.11).

A time course of PrA-McjB expression, examined by Western Blot analysis, revealed that McjB expression was switched on during early-exponential phase and continued to accumulate throughout later growth phases (see fig. 3.12). The expression levels of *mcjB* were compared to expression levels of *rpoC* during this same experiment by expressing pKW4 in an *Ec* strain modified with a chromosomal *proteinA::rpoC* gene (*Coli#55_prA::rpoC*). McjB protein levels were comparable to those of the *Ec* RNAP beta' subunit – a surprisingly high expression level considering that the *Ec* RNAP is an abundant protein (see fig. 3.12B the higher band corresponds to prA-rpoC and the lower band corresponds prA-mcjB).

Further experiments indicated that PrA-McjB was insoluble in the standard lysis buffer; therefore we tested whether adding detergents might improve its solubility. While a mild detergent, octylglucoside, actually made the protein more insoluble, SDS and Tween did improve the solubility (see fig. 3.13).



3.11 Bioassay of microcin activity in media from cells transfomed with either pBR322, pKW2 or pKW4.

MG1655 cells transformed with pBR322, pKW2 or pKW4 plasmids were grown in M9 minimal media. The media was applied to a C8 cartridge and bound peptides were eluted with 80% methanol. These fractions were lyophilized by speedvac and then resuspended in either 1 mL of H2O or 1 mL 8% methanol, and 2 µL of each was spotted onto an LB-agar plate and overlaid with agar containing microcin sensitive MG1655 cells. The pale circles are areas where growth of bacteria in this top agar was inhibited.

Figure 3.12 Growth curve and western blot analysis of prA_mcjB expression compared to prA_Ec _rpoC.
Coli#55 cells with a chromosomal rpoC gene tagged with prA were transformed with the pKW4 plasmid and grown in minimal media to trigger expression of the MccJ25 biosynthesis cassette. Aliquots of the cells were taken at various timepoints and were analyzed by SDS-PAGE followed by Western blot using the anti-prA antibody.
A) Growth curve B) SDS-PAGE analysis of whole cell extracts collected at the timepoints indiacted C) Western blot analysis of the same samples in A.
Lanes in A & B: (1) Coli#55_prA::rpoC at 51.5 hours post innoculation (2-8) Coli#55_prA::rpoC transfected with pKW4 at various timepoints post-innoculation; (2) 3 hours (3) 5 hours (4) 6.5 hours (5) 8 hours (6) 13 hours (7) 25 hours (8) 51.5 hours

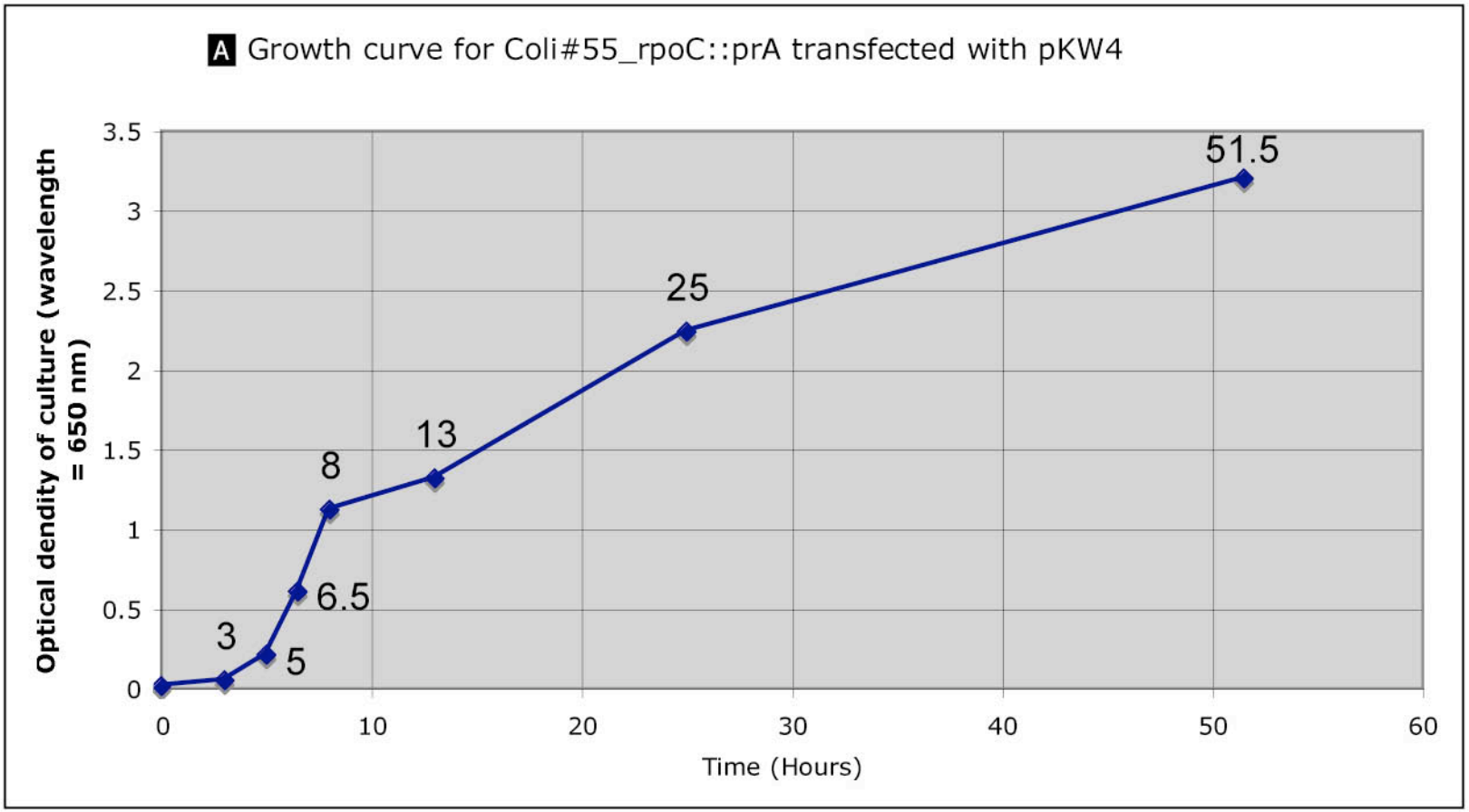


Figure 3.12 Growth curve and western blot analysis of prA_mcjB expression compared to prA_Ec _rpoC.

A) Growth curve of Coli#55 transfected with pKW4 plasmid.

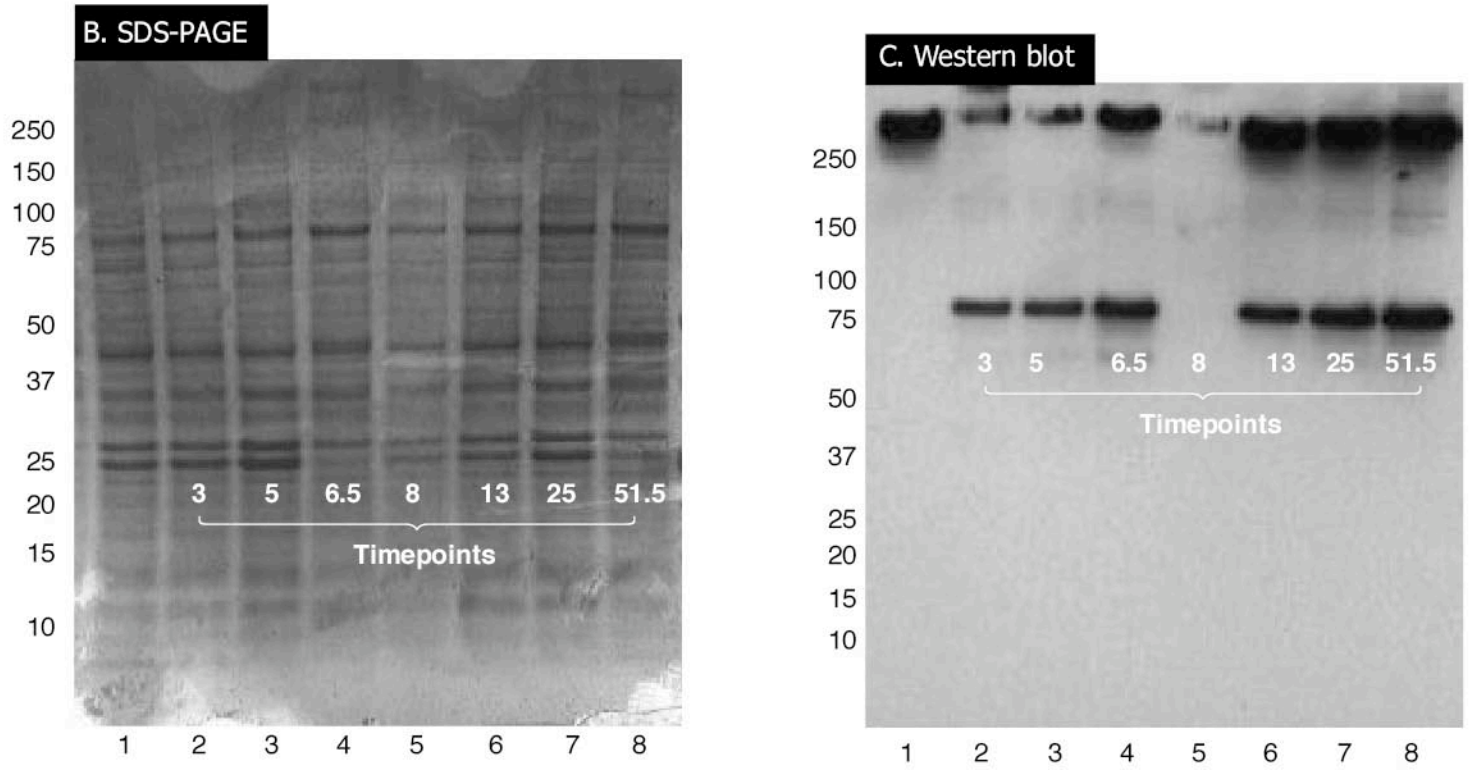


Figure 3.12 Growth curve and western blot analysis of prA_mcjB expression compared to prA_Ec _rpoC. B) SDS-PAGE analysis C) Western Blot analysis.

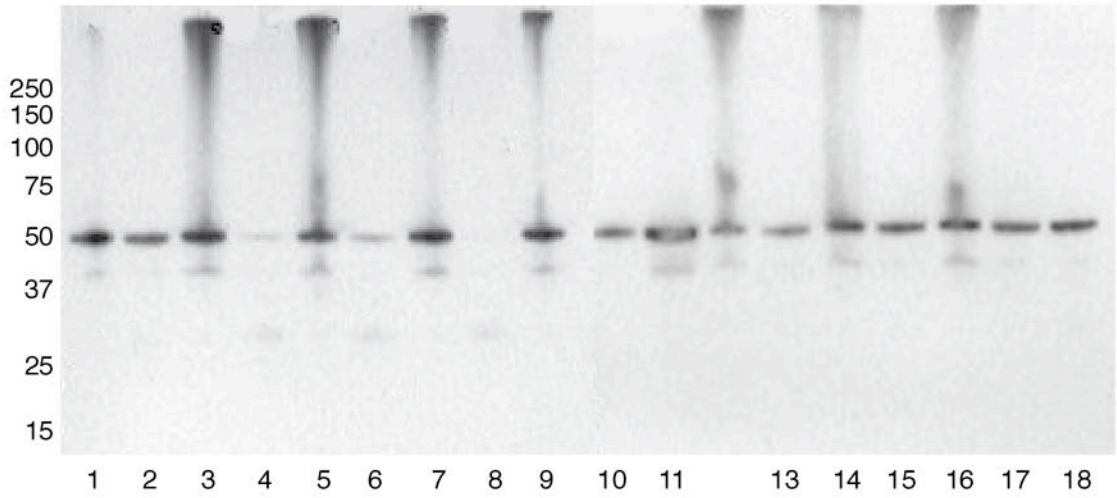


Figure 3.13 Western blot of prA-mcjB extracted in buffers containing detergents.

Cryogenically crushed cell pellets were resuspended in buffer with and without the addition of detergent and then clarified by centrifugation. The resultant soluble extracts and the insoluble pellets were boiled in SDS loading buffer and analysed by SDS-PAGE (NuPage 4-12% Bis-Tris gels) followed by Western blotting with anti-prA antibody.

Lanes: 1) no detergent insoluble fraction, 2) no detergent soluble fraction, 3) 0.5% octylglucoside insoluble fraction, 4) 0.5% octylglucoside soluble fraction, 5) 1% octylglucoside insoluble fraction, 6) 1% octylglucoside soluble fraction, 7) 2% octylglucoside insoluble fraction, 8) 2% octylglucoside soluble fraction, 9) 0.1% SDS insoluble fraction, 10) 1% SDS insoluble fraction, 11) 1% SDS soluble fraction, 12) 0.1% Tween insoluble fraction, 13) 0.1% Tween soluble fraction, 14) 0.5% Tween insoluble fraction 15) 0.5% Tween soluble fraction, 16) 1% Tween insoluble fraction, 17) 1% Tween soluble fraction, 18) 0.1% SDS soluble fraction.

3.4 Discussion

The lack of detectable protein expression from our hexahistidine recombinant expression vectors impeded any further progress in this project. Why were we unable to detect protein expression when other groups were able?

Duquesne et al. [6] used very similar pET vectors and used the standard expression strain BL21(DE3) to successfully express hexahistidine-tagged McjA, McjB and McjC. The most likely explanation for their success is that they used more discriminating tools for detecting expression and were satisfied with very low yields. Proteins were detected by tryptic digest of gel bands followed by mass spectrometry. Their reported yields were 40 ug/L for His_6-McJB and 200 ug/L for His_6-McjC after a single chromatographic purification step. These yields and this one-step purification would not suffice for crystallography purposes but were sufficient for a biochemical analysis.

In a similar strategy to our PrA-McJB strategy, Clarke and Campopiano [16] appended a hexahistidine-tag to the *mcjC* gene within the native operon and followed expression and purification by Western blot analysis. They did not obtain the maturase enzymes in a biochemically pure state, but rather obtained a detergent extract of the membrane fraction in which they detected His_6-McjC by Western lot. This extract was successfully used *in vitro* to mature recombinant McJA to the active form.

In contrast, Duquesne et al. [6] obtained soluble proteins by inducing at expression at low temperature, –15°C, and purified both His_6-McjB and His_6-McjC by standard

chromatographic techniques, which they then combined in an *in vitro* MccJ25 maturase assay using His_6-McjA as the substrate.

Although Duquesne et al. [6] purified His_6-McjA from the soluble fraction they found that the protein was highly susceptible to cellular proteases and was rapidly degraded, similar to our finding that the soluble MBP-mcjA protein was rapidly degraded. Clarke and Campopiano [16] obtained recombinant mcjA by expressing it as a keto-steroid isomerse fusion (KSI-mcjA) that was targeted to inclusion bodies and was extracted with 6 M Guanidine HCl. In retrospect we should have worked with the insoluble His_6-McjA protein rather than undergo the arduous task of chemical synthesis.

From our chemical synthesis efforts we obtained approximately 2.3 mg of 70% pure peptide. Again with the benefit of hindsight, we should have further purified this peptide and retained it for biochemical analysis. It would have been interesting to examine whether this peptide had any intrinsic structure by NMR analysis. Duquesne et al. [6] reported that the McjA protein has some sequence similarity to a serine peptidase and may have some autolytic activity. They also found that a linear peptide corresponding to the 21 residues found in MccJ25 was not a substrate for the McjB-McjC maturase enzymes. It would be interesting to probe which residues of the 37-amino acid leader are important for maturation. A mutational analysis of the 21 C-terminal residues of McjA corresponding to MccJ25 found that Gly38, Gly39 and Glu45 (Gly1, Gly2 and Glu8 in MccJ25) were absolutely restricted to the wildtype residues for microcin production *in vivo* [17]. The catalysis of the lactam bond between Gly1 and Glu8 must therefore require

the small, flexible di-Gly and the longer sidechain of Glu (even a Glu>Asp substitution was not tolerated).

Even though both Duquesne et al. [6] and Clarke and Campopiano [16] were able to produce MccJ25 *in vitro,* there is still no detailed understanding of the MccJ25 biosynthesis. The actual steps in the maturation are not known, the roles of each protein have not been defined, and no intermediary products have been characterized. Our prA-McJB construct could allow the MccJ25 maturase complex to be immunoisolated on Dynabeads (Invitrogen) conjugated with IgG and might provide an alternative *in vitro* assay that would not require the individual recombinant proteins to be expressed and purified. Crystal structures of McJB and McJC could still provide mechanistic details that are lacking.

Both the *mcJB* and *mcJC* genes have a very high rare codon bias, which is thought to strain the translation machinery and result in poor and faulty expression [18]. This rare codon usage may be an evolved strategy for ensuring that MccJ25 production is low enough to ensure that it is efficiently removed from the host cell before it can target an essential cell function; transcription by RNAP. Thus, generating sufficient McjB and McjC for crystallography studies may require synthetic gene design to remove the rare codon bias, *in vitro* gene expression to prevent problems with toxicity, and a system that couples transcription to translation to promote protein solubility (eg. [19; 20,]). These solutions are becoming cheaper and more efficient and may be a feasible option for future studies.

3.5 References

1. Rosengren, K. J., Clark, R. J., Daly, N. L., Goransson, U., Jones, A. & Craik, D. J. (2003). Microcin J25 has a threaded sidechain-to-backbone ring structure and not a head-to-tail cyclized backbone. *J Am Chem Soc* **125**, 12464-74.

2. Bayro, M. J., Mukhopadhyay, J., Swapna, G. V., Huang, J. Y., Ma, L. C., Sineva, E., Dawson, P. E., Montelione, G. T. & Ebright, R. H. (2003). Structure of antibacterial peptide microcin J25: a 21-residue lariat protoknot. *J Am Chem Soc* **125**, 12382-3.

3. Wilson, K., Kalkum, M., Ottesen, J., Yuzenkova, J., Chait, B. T., Landick, R., Muir, T., Severinov, K. & Darst, S. A. (2003). Structure of microcin J25, a peptide inhibitor of bacterial RNA polymerase, is a lassoed tail. In *J. Am. Chem. Soc.*, Vol. 125, pp. 12475-83.

4. Solbiati, J. O., Ciaccio, M., Farías, R. N. & Salomón, R. A. (1996). Genetic analysis of plasmid determinants for microcin J25 production and immunity. In *J Bacteriol*, Vol. 178, pp. 3661-3.

5. Solbiati, J. O., Ciaccio, M., Farías, R. N., González-Pastor, J. E., Moreno, F. & Salomón, R. A. (1999). Sequence analysis of the four plasmid genes required to produce the circular peptide antibiotic microcin J25. In *J Bacteriol*, Vol. 181, pp. 2659-62.

6. Duquesne, S., Destoumieux-Garzón, D., Zirah, S., Goulard, C., Peduzzi, J. & Rebuffat, S. (2007). Two enzymes catalyze the maturation of a lasso peptide in Escherichia coli. In *Chemistry & Biology*, Vol. 14, pp. 793-803.

7. Salomón, R. A. & Farías, R. N. (1992). Microcin 25, a novel antimicrobial peptide produced by Escherichia coli. In *J Bacteriol*, Vol. 174, pp. 7428-35.

8. Salomon, R. A. & Farias, R. N. (1995). The peptide antibiotic microcin 25 is imported through the TonB pathway and the SbmA protein. *J Bacteriol* **177**, 3323-5.

9. Katahira, R., Yamasaki, M., Matsuda, Y. & Yoshida, M. (1996). MS-271, a novel inhibitor of calmodulin-activated myosin light chain kinase from Streptomyces sp.--II. Solution structure of MS-271: characteristic features of the "lasso' structure. *Bioorg Med Chem* **4**, 121-9.

10. Solbiati, J. O., Ciaccio, M., Farias, R. N., Gonzalez-Pastor, J. E., Moreno, F. & Salomon, R. A. (1999). Sequence analysis of the four plasmid genes required to produce the circular peptide antibiotic microcin J25. *J Bacteriol* **181**, 2659-62.

11. Torgov, M. Y., Janzen, D. M. & Reddy, M. K. (1998). Efficiency and frequency of translational coupling between the bacteriophage T4 clamp loader genes. *J Bacteriol* **180**, 4339-43.

12. Oppenheim, D. S., Bennett, G. N. & Yanofsky, C. (1980). Escherichia coli RNA polymerase and trp repressor interaction with the promoter-operator region of the tryptophan operon of Salmonella typhimurium. *J Mol Biol* **144**, 133-42.

13. Clarke, D. J. & Campopiano, D. J. (2007). Maturation of McjA precursor peptide into active microcin MccJ25. In *Org Biomol Chem*, Vol. 5, pp. 2564-6.

14. Camarero, J. A., Cotton, G. J., Adeva, A. & Muir, T. W. (1998). Chemical ligation of unprotected peptides directly from a solid support. *J Pept Res* **51**, 303-16.

15. Schnolzer, M., Alewood, P., Jones, A., Alewood, D. & Kent, S. B. (1992). In situ neutralization in Boc-chemistry solid phase peptide synthesis. Rapid, high yield assembly of difficult sequences. *Int J Pept Protein Res* **40**, 180-93.

16. Clarke, D. J. & Campopiano, D. J. (2007). Maturation of McjA precursor peptide into active microcin MccJ25. *Org Biomol Chem* **5**, 2564-6.

17. Pavlova, O., Mukhopadhyay, J., Sineva, E., Ebright, R. H. & Severinov, K. (2008). Systematic structure-activity analysis of microcin J25. *J Biol Chem* **283**, 25589-95.

18. Kane, J. F. (1995). Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli. *Curr Opin Biotechnol* **6**, 494-500.

19. Makarova, O. V., Makarov, E. M., Sousa, R. & Dreyfus, M. (1995). Transcribing of Escherichia coli genes with mutant T7 RNA polymerases: stability of lacZ mRNA inversely correlates with polymerase speed. *Proc Natl Acad Sci U S A* **92**, 12250-4.

20. Iskakova, M., Szaflarski, W., Dreyfus, M., Remme, J. & Nierhaus, K. (2006). Troubleshooting coupled in vitro transcription-translation system derived from Escherichia coli cells: synthesis of high-yield fully active proteins. In *Nucleic Acids Research*, Vol. 34, pp. e135-e135.

Chapter 4. Expression, purification and crystallization of *Escherichia coli* ΔαCTD-RNAP

4.1 Introduction

Gene transcription in all organisms is carried out by DNA-dependent RNA polymerases (RNAPs). The crystal structures for two thermophilic eubacterial, multi-subunit RNAPs have been solved; *Thermus aquaticus (Taq)*[1] and *Thermus thermophilus (Tth)*[2]. The genes encoding RNAP have been highly conserved throughout evolution and these structures have provided immense insight into the transcription mechanism. However, the model organism for prokaryotic cells is *Escherichia coli* (*Ec*), a mesophilic, Gram-negative bacteria. More research has probably been done on this organism than any other, and the *Ec* RNAP and regulation of the transcription system is by far the best studied and characterized. The *Taq* RNAP structure has been fit into a 15-Å resolution cryo-electron microscopy reconstruction of the *Ec* RNAP revealing that these highly conserved enzymes are globally similar apart from the addition of three large insertions in the *Ec* β and β' subunits [3]. A higher resolution crystal structure of *Ec* RNAP would allow finer details of these differences to be examined and would provide insight into functional differences between these enzymes. Crystals of *Ec* RNAP could also be soaked with various co-factors that have been found to specifically affect transcription in *E. coli.* For instance, there would be the potential for crystals to be soaked with inhibitors such as microcinJ25, which does not inhibit either *Taq* or *Tth*.

Ec RNAP has not, thus far, been amenable to crystallization despite considerable effort.

One potential hindrance to crystallization may be the carboxyl-terminal domain of the alpha subunit (α-CTD, residues 249-329) that is connected to an amino-terminal domain (α-NTD, residues 8-235) by a 14-residue linker [4]. This long, flexible linker tethers each α-CTD monomer to the main body of the core enzyme, allowing interaction with upstream DNA elements or DNA-bound activators to upregulate transcription [5]. These interactions are essential *in vivo* but this flexibility is a potential source of disorder in crystals. In all of the crystal structures of *Taq* or *Tth* RNAP, the □-CTDs are disordered. In contrast, the α-NTD is structurally integrated into the main body of the core enzyme. Indeed, the α-NTD homodimer serves as a scaffold for the assembly of the core enzyme [6]. *Ec* RNAP lacking the α-CTD (RNAPΔα-CTD) has been previously produced by *in vitro* assembly from individual recombinant subunits (eg. Tang et al. [12]), however, such preparations are unlikely to result in well-folded, homogenous preparations suitable for crystallization.

With this in mind, we set out to develop an expression and purification system for *Ec* RNAP lacking the α-CTD (RNAPΔα-CTD) to facilitate crystallization. This system would be based on plasmid expression in *Ec* and therefore needed to consider that the host RNAP subunits could intermingle with plasmid-encoded subunits. Thus, the host *Ec* strain would have to be engineered to contain a matching α-CTD on the chromosome and the α-CTD would have to be removed post-expression because it is essential *in vivo*. Therefore, we used an inducible expression plasmid with the *Ec* RNAP genes encoding β, β' and ω with a α subunit modified with a protease restriction site encoded within the α-NTD-CTD linker for removal of the α–CTD, and a decahistidine-tag appended to the

amino-terminus for purification; α_ppx234-241_CTD$(His)_{10}$. Expression from this plasmid was induced in a specialized host strain chromosomally modified to express the same α. The *Ec* RNAP produced from this system was screened for candidate crystallization conditions and a resulting "hit" was obtained, reproduced and improved, but was not ultimately suitable for data collection.

4.2 Methods

4.2.1 Expression and purification of *Ec* RNAPΔα-CTD

A recombinant *rpoA* gene was designed with the CTD-NTD linker (234-241) substituted with a PreScission Protease™ (Amersham) cleavage site, LEVLFQGP, and with an uncleavable decahistidine tag appended 3' to the *rpoA* gene followed by a downstream thermostable kanamycin resistance cassette. The resultant DNA and protein sequences are appended to this chapter (Appendix 4.6). This gene was first cloned into a pet21a vector for trial expression and cleavage and was then sub-cloned into an *E. coli* expression vector, pVS6 (a gift from Robert Landick) to generate pVS10. Both these plasmids and the pACYCDuet1_rpoZ plasmid encoding *E. coli rpoZ* were provided by Lars Westblade. *The BL21(DE3) rpoA_ppx_CTD(his)$_{10}$* strain was provided by Imran Husnain and Mark Thomas.

BL21(DE3)_rpoA_ppx_CTD(his)10 transformed with both pVS10 and pACYCDuet1_rpoZ were grown at 37°C in LB plus 200 ug/mL ampicillin to mid-log phase, then the temperature was reduced to 30°C and expression induced by addition of isopropyl β-d-thiogalactopyranoside to 1 mM. After 4 hours induction, cells were

harvested by centrifugation, resuspended in lysis buffer (50 mM Tris-HcL, pH 8, 5% glyverol, 1 mM EDTA, pH 8.0, 5 mM DTT, 1 mM PMSF, EDTA-free protease-inhibitor cocktail (Sigma) and lysed by French Press. DNA binding proteins were precipitated from the clarified cell lysate with polymin P added to a final concentration of 0.6%. The polymin P precipitate was pelleted by centrifugation and then washed with TGED buffer (10 mM Tris-HCl, pH 8.0, 5% glycerol, 0.1 mM EDTA, 1 mM DTT) containing 0.5 M NaCl. This wash step was repeated. Proteins were then eluted from the polymin P pellet with TGED buffer containing 1 M NaCl. The elution step was repeated twice and pooled. Protein in the eluant was precipitated by the addition of ammonium sulfate to a final concentration of 350 g/L. The ammonium sulfate precipitate was pelleted by centrifugation, resuspended in IMAC-A buffer (20 m*M* Tris pH 8.0, 1 M NaCl, 5% glycerol, 0.5 m*M* β-mercaptoethanol) and applied to an an IMAC HiTrap column (GE Healthcare) charged with Ni^{2+} ions and pre-equilibrated with IMAC-A buffer. The column was washed with 20 column volumes of IMAC-A plus 5 mM imidazole and then 10 column volumes each of IMAC-A plus 20 mM imidazole, IMAC-A plus 25 mM imidazole, IMAC-A plus 60 mM imidazole, IMAC-A plus 80 mM imidazole, IMAC-A plus 100 mM imidazole. Proteins bound to the column were eluted with IMAC-A plus 250 mM imidazole. The eluted proteins were digested overnight with PreScission protease (GE Healthcare) and simultaneously dialyzed against IMAC-A buffer at 4°C before being applied to a second, subtractive, IMAC HiTrap column to remove non-specific and uncleaved protein. The flowthrough was dialyzed against TGED buffer containing 0.1 M NaCl and applied to a BioRex70 (Bio-Rad) column thad had been preequilibrated with TGED buffer containing 0.1 M NaCl. The column was washed with

two column volumes of TGED buffer containing 0.1 M NaCl and bound proteins were eluted with a linear gradient of 0.2 M – 1 M NaCl in TGED buffer. Fractions containing core RNAP were precipitated with the addition of ammonium sulfate to the final concentration of 350 g/L. The ammonium sulfate precipitate was pelleted by centrifugation, resuspended in TGED buffer containing 0.5 M NaCl, and further purified by size-exclusion chromatography on a SD200 column (GE Healthcare).

Fractions containing RNAP were dialysed into storage buffer; 150 mM NaCl, 10 mM TrisCl pH 8.0, 1% glycerol, 5 mM DTT, 10 mM MgCl2, 5 mM B-ME and concentrated using centrifugal concentrators (Vivaspin, 100, 000 MWCO).

4.2.2 Crystallization of *Ec* RNAPΔα-CTD

An extensive number of vapor diffusion crystallization trials were conducted using various commercial screens; JCSG+, Protein Complex, MPD, PEG suite II (all from Qiagen Research), Natrix, MembFac, Classics Lite, PEG-ION (Hampton Research). A single candidate condition was identified; MembFac#23, 0.1 M ADA buffer pH 6.5, 0.1 M lithium sulfate, 12% polyethylene glycol 4,000, 2% (v/v) isopropanol. Further vapor diffusion crystallization trials were setup based on MembFac#23 varying the temperature, pH and polyethylene glycol 4,000 concentration and screening commercial additive screens; Additive Screen, Detergent Screens I-III and Silver Bullets (Hampton Research) and OptiSalts (Qiagen Research).

Seedbeads (Hampton Research) were used to make seed stock from crystals grown over

MembFac#23 and dilutions of this seed stock were used as an epitaxial seed to re-screen the MembFac, Natrix, and PEG/IONs (Hampton Research) suites as well as grid screens around MembFac#23.

4.2.3 Data collection and processing

Diffraction images were collected at the at X3A, National Synchrotron Light Source (NSLS), Brookhaven National Laboratory (BNL) and at the 24-ID-E, Advanced Photon Source, (APS), Argonne National Laboratory (ANL). The data was processed with HKL2000.

4.3 Results

4.3.1 *Ec* RNAPΔα-CTD with stochiometric amounts of ω was purified in quantities suitable for crystallography

The linker region between the α-CTD and α-NTD has been shown to tolerate substitutions, therefore a recombinant *rpoA* gene was designed to incorporate a PreScission protease (ppx) cleavage site within this linker as well as a decahistidine tag at the amino terminus to facilitate purification of this modified RNAP and removal of the cleaved α-CTD by metal-affinity chromatography. This modified gene *rpoA_ppx234-241_CTD(his)10* was cloned into a standard expression vector, pet21a (Novagen), and the expression and purification of this subunit alone was tested. The α_ppx234-241_CTD$(His)_{10}$ protein was well expressed under standard expression conditions and

cleavage with ppx went close to completion (data not shown). Therefore, this modified gene was sublconed into a recombinant RNAP expression vector replacing the standard *rpoA* gene to generate pVS10. The expression and cleavage of α_ppx234-241_CTD$(His)_{10}$ within the context of the core RNAP was also very good, however, the purified protein contained full-length α from the BL21(DE3) chromosomal expression. Therefore, a strain of BL21(DE3) incorporating the modified *rpoA* gene was generated by our collaborators; Imran Husnain and Mark Thomas.

Ec RNAPΔα-CTD expressed from pVS10 in *BL21(DE3)_rpoA_ppx234-241(his)10* was purified by an arduous protocol (see fig. 4.1 and methods for details) that resulted in highly pure protein. In the initial preparations we suspected that the ω subunit was present in sub-stochiometric amounts based on SDS-PAGE and Coomassie blue staining of purified protein. This has been previously observed for preparations of RNAP from non-recombinant expression systems. This was confirmed by Electrospray Ionisation–Mass Spectrometric analysis (ESI-MS) of the intact RNAPΔα-CTD. We compared RNAP prepared from either endogenous expression or recombinant expression with and without a supplementary ω-expression plasmid (rRNAPΔα-CTD and rRNAPΔα-CTD +ω). In all three cases we found two series of ions corresponding to the intact RNAP and RNAP minus the 10 kDa ω subunit. For the endogenous RNAP and for rRNAPΔα-CTD +ω the peaks corresponding to RNAP minus ω were much smaller than those corresponding to the intact RNAP. However, in the case of the recombinant RNAP prepared without a supplementary ω-expression plasmid the two peaks were equal in size. Although ESI-MS is not strictly quantitative it does give some indication to the

relative quantities of two species. In light of these results, we included a tandem ω expression plasmid, pACYC_rpoZ in our preparations, to increase the ω content. Typical yields for the highly pure RNAPΔα-CTD were 3 mg/L of culture.

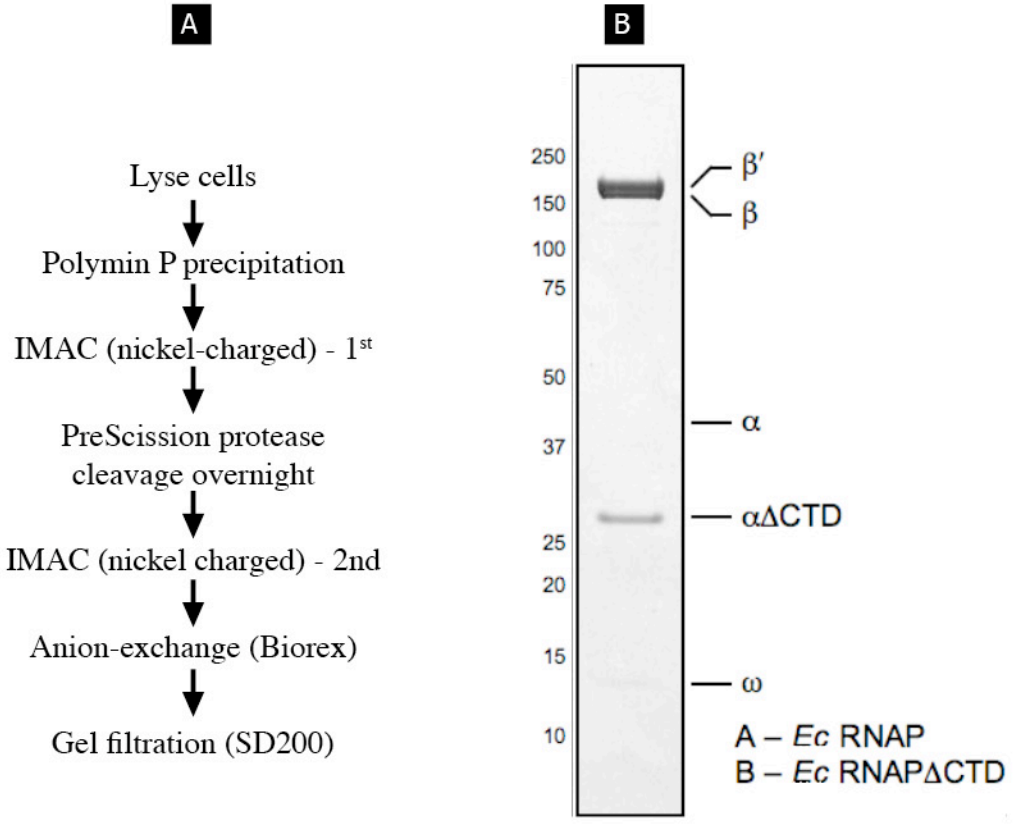
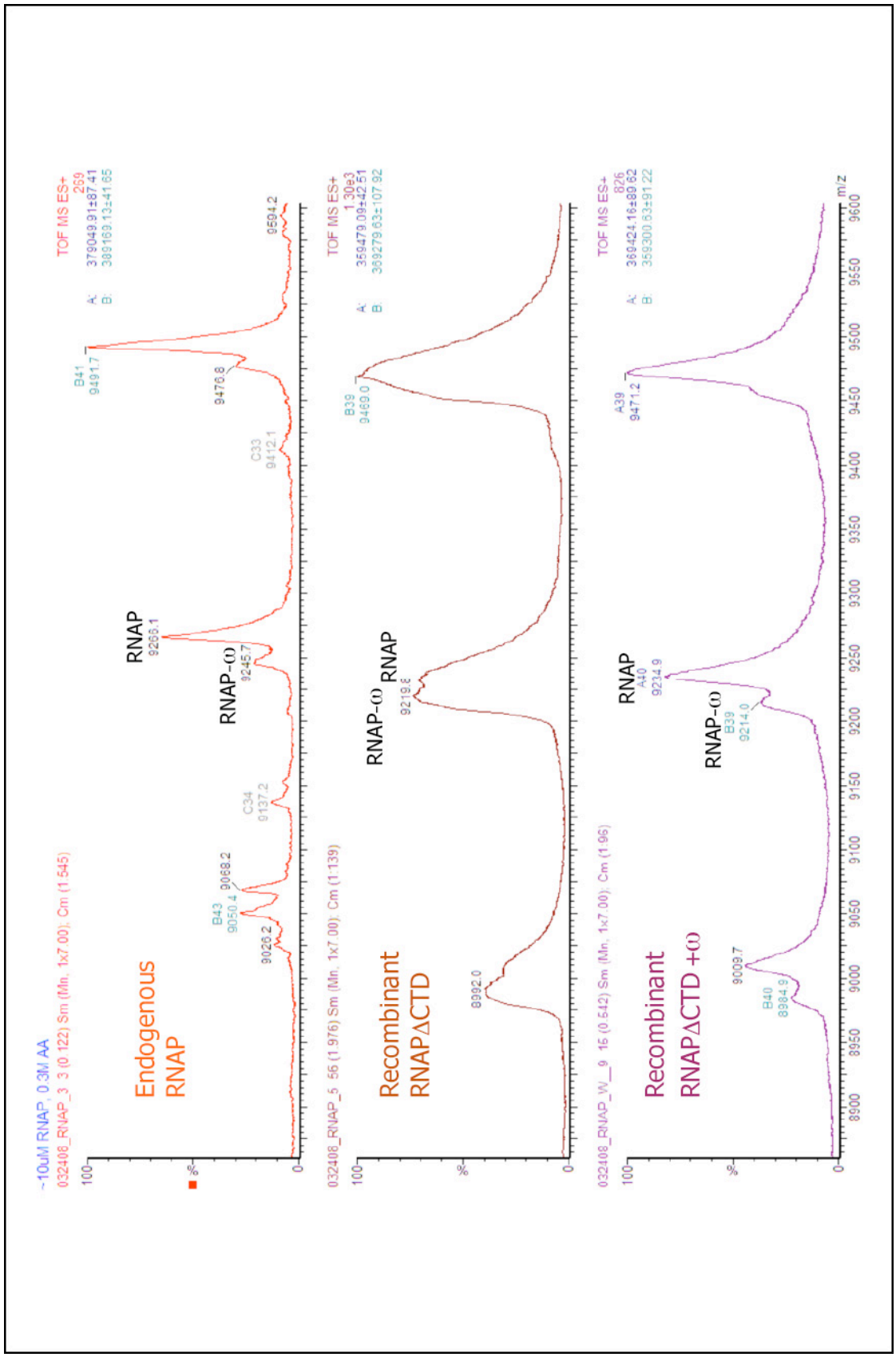


Figure 4.1 A) Purification protocol and B) SDS-PAGE analysis of the final purified protein.

Figure 4.2 The omega content of our initial preparations was sub-stochiometric.

ESI-MS spectra from three different Ec RNAP preparations; A) endogenous RNAP, B) recombinant RNAP from the pVS10 plasmid, and C) recombinant RNAP from the pVS10 plasmid supplemented with the pACYC_rpoZ plasmid. The most abundant species of ions in each each spectrum corresponds to the core RNAP while the smaller peak to the left of this main series corresponds to the RNAP-ω. In A) core RNAP (ion series B), calculated mass 388,981 Da, measured mass, 389,169 Da, in B) RNAP (ion series B) calculated mass 368,811 Da, measured mass, 369, 279 Da, in C) RNAP (ion series A) calculated mass 368,811 Da, measured mass, 369, 424 Da. RNAP-ω in all spectra differs from RNAP by approximakely 10 kDa; an ion from each series is labeled in each spectrum. (Spectra obtained by Zachary Quinkert).


~10uM RNAP, 0.3M AA
032408_RNAP_3 3 (0.122) Sm (Mn, 1x7.00); Cm (1:545)
TOF MS ES+
269
A: 379049.91±87.41
B: 389169.13±41.65
Endogenous
RNAP
RNAP
9266.1
RNAP-ω
9245.7
B41
9491.7
9476.8
C33
9412.1
9594.2
B43
9050.4
9068.2
C34
9137.2
9026.2
032408_RNAP_5 56 (1.976) Sm (Mn, 1x7.00); Cm (1:139)
TOF MS ES+
1.30e3
A: 359479.09±42.51
B: 389279.63±107.92
Recombinant
RNAPΔCTD
RNAP-ω
RNAP
9219.8
B39
9469.0
8992.0
032408_RNAP_W_ 9 16 (0.542) Sm (Mn, 1x7.00); Cm (1:96)
TOF MS ES+
826
A: 369424.16±89.62
B: 359300.83±91.22
Recombinant
RNAPΔCTD +ω
RNAP
A40
9214.9
A39
9471.2
RNAP-ω
B39
9214.0
9009.7
B40
8984.9
%
100
0
8900
8950
9000
9050
9100
9150
9200
9250
9300
9350
9400
9450
9500
9550
9600
m/z

4.3.2 Crystals of *Ec* RNAPΔα-CTD diffracted to 10 Ångstroms, and after various improvements, to 8 Ångstroms.

An initial screen of 96 crystallization reagents (Hampton Research Natrix and MembFac) produced a single candidate hit – small urchins. These urchins diffracted weakly to 10 Å at the X3A beamline at BNL. Further screening of additional 600+ conditions failed to produce an improved candidate condition. Initially these crystals were only reproducible from some preparations and not others. After the addition of the extra ω expression cassette these crystals could be reliably reproduced from every preparation.

The initial hit was improved by screening around the crystallization condition, varying the pH and precipitant concentration. This led to larger clusters of very thin plates. We then seeded this improved condition with seed stock made from the original urchins and obtained single plates (see fig. 4.3). The largest of these plates measured 10 x 30 x 5 μm and diffracted to 8 Å at the 24-ID-E micro-beamline at ANL (see fig. 4.4). The images collected were indexed and the crystals had a P2 primitive monoclinic lattice with unit cell dimensions a= 130.9 Å, b= 173.5 Å, c = 241.1 Å, $\alpha = \beta = 90$ $\gamma = 95.79$, volume =5,447,723 $Å^3$. These dimensions would allow for 3 copies of RNAP in the asymmetric unit with 48.5% solvent content and a Matthews coefficient of 2.39 $Å^3$ /Dalton which are within the empirically observed range for protein crystals [11] and thus the unit cell is consistent with the crystals containing *Ec* RNAPΔα-CTD.

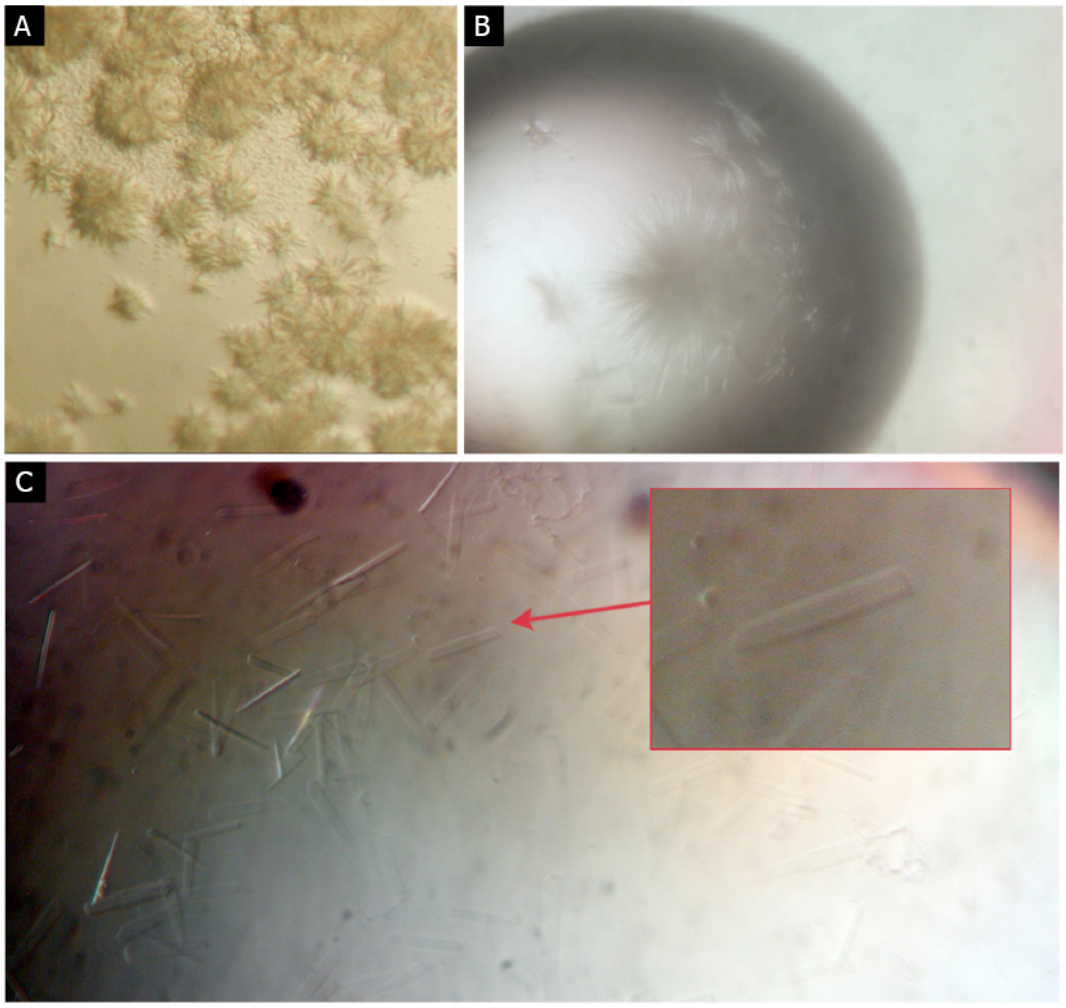


Figure 4.3 Crystals of Ec RNAPΔα-CTD were obtained initially as small urchins (A) and then improved to larger clusters (B) and to plates (C).

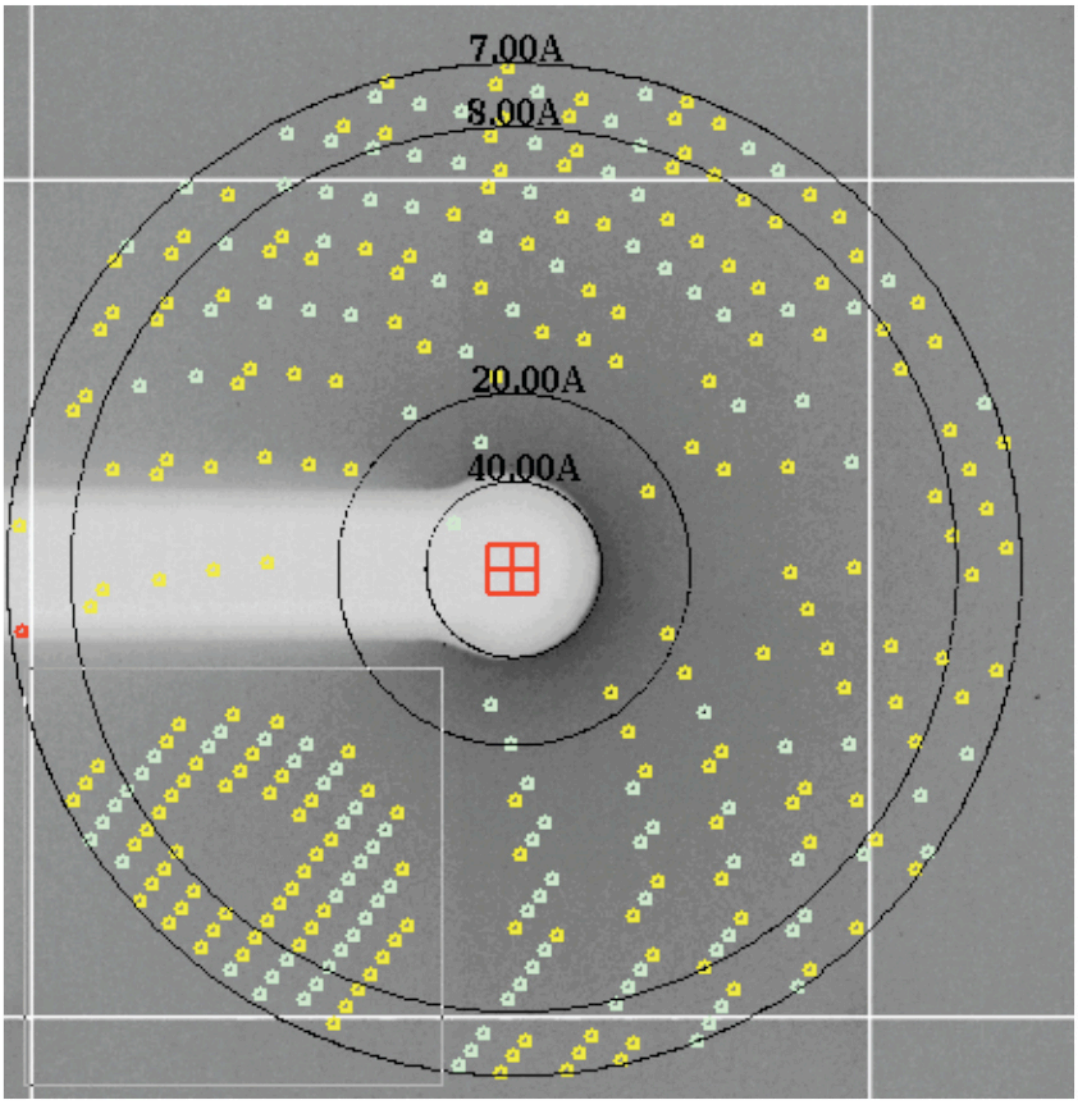


Figure 4.4 The Ec RNAPΔα-CTD crystal plates diffracted to 8 Ångstroms. Shown here is a single diffraction image.

4.4 Discussion

We obtained crystals of *Ec* RNAPΔα-CTD that diffracted weakly. To our knowledge, this is the first time that diffracting crystals of *Ec* RNAP have been obtained. However, these crystals were not suitable for data collection as they required long exposure times and suffered from radiation damage.

That addition of the extra ω-expression plasmid improved the reproducibility of obtaining crystals was not surprising as this subunit is believed to have a key chaperone role and be important for core assembly [7].

We have tried nearly every trick in the book to further optimize these crystals to no avail. Further modifications of the protein construct might be necessary for any further improvements in diffraction. One feature of RNAP that might be a concern for our preparations is the β' zinc-binding domain. This domain was disordered in the Taq core structure [8], was "difficult to interpret" in the Taq holoenzyme [9] and had no secondary structure in the Tth holoenzyme [2]. The zinc bound by this domain acts as a chaperone to keep this region structured [10]. This may be an issue in our purification protocol as we use 1-5 mM DTT in the buffers throughout the preparation except for the IMAC stage, and the final protein buffer contained both 5 mM B-ME and 5 mM DTT -reducing agents that may strip zinc from the protein. Another modification that might improve diffraction is the removal of dispensable region 2 from within the β subunit, we recently solved this

structure (see Chapter 6) and found that this non-conserved insert forms two long helices that extend away from the globular core of RNAP and might prevent tight crystal packing. There has been no function assigned to this region and it can be removed without any affect on *in vitro* function [13].

The thermophilic RNAP structures were solved despite regional areas of disorder due to the α-CTD and the β' zinc-binding domain. Thermophilic proteins are more amenable to crystallography as they must have more rigid structures to cope with life at high temperatures. It may be useful to search for a thermophilic relative of *Ec* that is susceptible to drugs, inhibitors and protein factors that target gram-negative bacteria, for future structural studies. The structure of the *Ec* RNAP, of course, remains an ambitious but worthy goal.

4.5 References

1. Zhang, G., Campbell, E. A., Minakhin, L., Richter, C., Severinov, K. & Darst, S. A. (1999). Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution. *Cell* **98**, 811-24.

2. Vassylyev, D. G., Sekine, S., Laptenko, O., Lee, J., Vassylyeva, M. N., Borukhov, S. & Yokoyama, S. (2002). Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. In *Nature*, Vol. 417, pp. 712-9.

3. Darst, S. A., Opalka, N., Chacon, P., Polyakov, A., Richter, C., Zhang, G. & Wriggers, W. (2002). Conformational flexibility of bacterial RNA polymerase. In *Proc Natl Acad Sci USA*, Vol. 99, pp. 4296-301.

4. Husnain, S. I., Meng, W., Busby, S. J. & Thomas, M. S. (2004). Escherichia coli can tolerate insertions of up to 16 amino acids in the RNA polymerase alpha subunit inter-domain linker. *Biochim Biophys Acta* **1678**, 47-56.

5. Ebright, R. H. & Busby, S. (1995). The Escherichia coli RNA polymerase alpha subunit: structure and function. *Curr Opin Genet Dev* **5**, 197-203.

6. Zhang, G. & Darst, S. A. (1998). Structure of the Escherichia coli RNA polymerase alpha subunit amino-terminal domain. In *Science*, Vol. 281, pp. 262-6.

7. Mathew, R. & Chatterji, D. (2006). The evolving story of the omega subunit of bacterial RNA polymerase. *Trends Microbiol* **14**, 450-5.

8. Zhang, G., Campbell, E. A., Minakhin, L., Richter, C., Severinov, K. & Darst, S. A. (1999). Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution. In *Cell*, Vol. 98, pp. 811-24.

9. Murakami, K. S., Masuda, S. & Darst, S. A. (2002). Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 A resolution. In *Science*, Vol. 296, pp. 1280-4.

10. Markov, D., Naryshkina, T., Mustaev, A. & Severinov, K. (1999). A zinc-binding site in the largest subunit of DNA-dependent RNA polymerase is involved in enzyme assembly. *Genes Dev* **13**, 2439-48.

11. Kantardjieff, KA., Rupp, B., (2003) *Matthews coefficient probabilities: Improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals.* Protein Sci. 12(9):1865-71.

12. Tang, H., Severinov, K., Goldfarb, A., Ebright, RH. (1995) *Rapid RNA polymerase genetics: one-day, no-column preparation of reconstituted recombinant Escherichia coli RNA polymerase.* Proc Natl Acad Sci USA. **23**;92(11):4902-6.

13. Severinov K, Mustaev A, Kashlev M, Borukhov S, Nikiforov V, Goldfarb A. *(*1992) *Dissection of the beta subunit in the Escherichia coli RNA polymerase intodomains by proteolytic cleavage.* J Biol Chem. **25**;267(18):12813-9.

Chapter 5: Studies of Q proteins from lambda and 82 phages

Introduction

Phage invade bacterial cells and subvert host cell machinery, including RNAP, for their own needs. The *Lambda* and *82* phage Q antitermination proteins (λQ and 82Q) bind and remodel host *Ec* RNAP so that the Q-RNAP complex transcribes through termination sequences in the late gene operon. The *Lambda* and *82* Q proteins share no sequence homology but are functional homologs -the details described below are for λQ but are very similar for the two proteins (reviewed in [1]).

λQ acts at the phage late promoter (λPR') via the Q utilization sequence (*qut*) that contains two crucial elements; a Q binding element (*QBE*) between the -35 and -10 promoter elements and a downstream pause-inducing element. The σ^{70}-RNAP holoenzyme ($E\sigma^{70}$) initiates transcription from this promoter but encounters the pause-inducing element after transcribing 16-17 nt. The pause-inducing element is a -10-like sequence that interacts with $\sigma^{70}2$. In the absence of Q, $E\sigma^{70}$ pauses for about 30 seconds and then continues transcription until it reaches the termination sequence, TR', 200 nt downstream. Q has both anti-pausing and anti-termination activity, thus in the presence of Q the pause is shortened and a Q-RNAP complex exits the pause and transcribes through TR'.

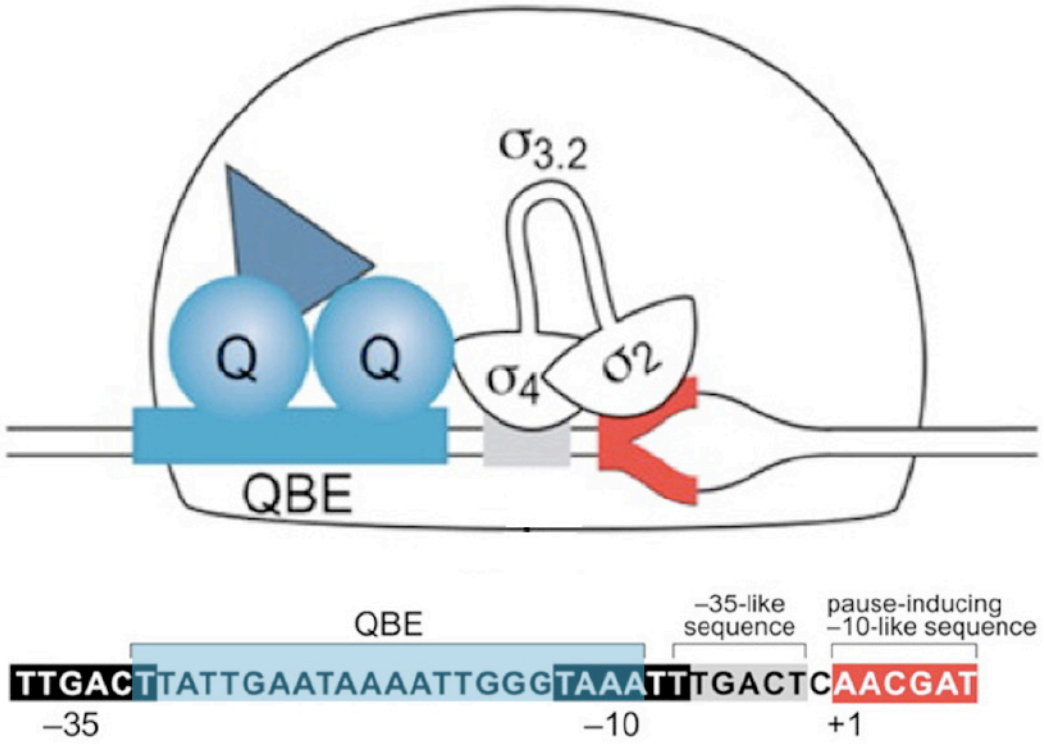


Figure 5.1 The Q utilization sequence and the Q-engaged complex

Q binds the QBE (blue) and interacts with both $\sigma^{70}4$ and the β flap (blue triangle). $\sigma^{70}4$ is displaced from the β flap and binds a -35-like sequence (grey) stabilized by its interaction with Q. $\sigma^{70}2$ binds the -10-like sequence (red). The DNA sequences of the Q utilization are shown below. (Figure adapted from Nickels et al. (2006).

The mechanisms for Q anti-pausing and anti-termination are not fully understood but multiple lines of evidence have led to the following model (see fig. 5.1, reviewed in [1], presented in [2]). In $E\sigma^{70}$, the $\sigma^{70}4$ is bound to the β flap and $\sigma^{70}2$ is bound to the β coiled-coil, positioned to interact with the -35 and -10 elements [3; 4]. At *qut*, the $E\sigma^{70}$ must clear the promoter in order for Q to bind the QBE. Thus, Q encounters $E\sigma^{70}$ where the $\sigma^{70}3.2$ is presumed to be displaced from the RNA exit channel and the $\sigma^{70}4$-β flap interaction is destabilized due to the presence of a short RNA transcript [3]. Q then interacts with $\sigma^{70}4$,

stabilizing $\sigma^{70}4$ binding to a -35-like element that is just 1 bp upstream from the pause-inducing -10 element [5]. This allows Q to bind the β-flap and may place structural tension on σ^{70} that promotes its release from RNAP. Q forms a stable component of the ensuing elongating complex and can travel with RNAP up to 22 kilobases [6].

We were able to purify λQ in a soluble form that was highly active. Unfortunately, we were unable to generate crystals of this protein suitable for data collection.

5.2 Methods

5.2.1 *λQ* Protein expression and purification

The *λQ* gene was PCR amplified from pACQ (gift from Bryce Nickels) and subcloned into a pET21a (Novagen) vector between NdeI/BamHI sites to generate pET21a_λQ.

BL21(DE3) cells transformed with pet21a_λQ were grown in LB plus 200 μg ml^{-1} ampicilin at 37°C to mid-log phase, then the temperature was lowered to 16°C and expression was induced with the addition of isopropyl β-d-thiogalactopyranoside to 1mM. The cells were harvested 16 hours after induction by centrifugation, resuspended in buffer A (10 m*M* Tris pH 8.0, 200 m*M* NaCl, 5% glycerol, 0.1 m*M* EDTA, 10 mM DTT) supplemented with 1 m*M* PMSF and 0.5X EDTA-free protease-inhibitor cocktail (Sigma) and lysed using a French press. The clarified lysate was applied to a heparin column equilibrated with lysis buffer, the column was washed with 30 column volumes (CV) of buffer A and then bound proteins were eluted with a linear gradient of buffer A plus 0.2 – 0.6 M NaCl over 20 CV. The fractions containing λQ were pooled and diluted so that the final NaCl content of the buffer was approximately 150 mM. This diluted sample was applied to a HiTrap SP column equilibrated in buffer A, washed with 10 CV buffer A and bound proteins were eluted using bufferA plus a linear gradient of NaCl 100 mM – 1M over 20 CV. Fractions containing λQ were pooled and concentrated using centrifugal concentrators (Vivaspin, 5,000 MWCO) and then finally purified by size-exclusion chromatography on a SD75 column (GE Healthcare) equilibrated in buffer A plus 500 mM NaCl.

Fractions containing λQ were pooled and dialyzed into storage buffer; 10 mM Tris-Cl, pH8.0, 200 mM NaCl, 10 mM DTT, snap frozen in liquid N_2 and stored at -80°C.

Plasmids encoding mutant *λQ;* Q(T101I,A160V,T165A), QE134K & QE141K, were gifts from Bryce Nickels. Truncated and mutant versions of *λQ* were similarly cloned and expressed.

5.2.1 *82Q* Protein expression and purification

The *82Q* gene was PCR amplified from p82Qhis_5 (gift from Jeffrey Roberts) and subcloned into a pET29a (Novagen) vector between NdeI/BamHI sites to generate pET29a_82Q, and into a pET28a-based vector between NdeI/BamHI sites as a hexahistidine-fusion protein to generate pSKB2-82Q. The $his_6$82Q protein was PCR amplified from pSKB2-82Q and cloned into pACYCDuet1_rpoZ via NotI/EcoRI sites. 82Q was expressed from pET29a_82Q and purified using the λQ protocol described above.

5.2.2 *λQ* + β flap and 82*Q* + β flap protein expression

An *Ec rpoB* gene fragment encoding amino acids 831-1057 (β flap) was PCR amplified from an *Ec* RNAP expression vector, pVS10 and cloned into a pet29a (Novagen) vector between NdeI/BamHI sites to generate pet29a_βflap. The *λQ* gene was PCR amplified from plasmid pACQ, and cloned into a pET28a-based vector between NdeI/BamHI sites as a hexahistidine-fusion protein to generate pSKB2-*λQ*. The hexahistidine-*λQ* fusion

was in turn PCR amplified along with upstream ribosomal binding site (rbs) and cloned into pET29a-β flap between EcoRI/NotI sites to generate pet29a-β flap-*λQ*. The β flap-82Q co-expression vector, pet29a-βflap was made in an identical manner.

BL21(DE3) cells transformed with either construct were grown in LB plus 50 μg ml^{-1} kanamycin at 37°C to mid-log phase, then the temperature was lowered to 16°C and expression was induced with 0.1 m*M* isopropyl β-d-thiogalactopyranoside for 16 hours.

5.2.3 *λQ* + σ^{70}4 protein expression

The pSKB2Ecσ^{70}4 expressing his_6Ec σ^{70} residues 541–613 was a gift from other members of the laboratory [7].

The his_6Ecσ^{70}4 gene and upstream rbs was PCR amplified from pSKB2Ecσ^{70}4 and cloned into the pet21a_*λQ* plasmid to generate pet21a_*λQ* _ $his_6\sigma^{70}$4. BL21(DE3) cells transformed with pet21a_*λQ* _ $his_6\sigma^{70}$4were grown in LB plus 200 μg ml^{-1} ampicillin at 37°C to mid-log phase, and expression was induced with 1 m*M* isopropyl β-d-thiogalactopyranoside for 3 hours. Cells were lysed by French Press and the lysate was clarified by centrifugation at 10,000 rpm for 20 min.

5.2.4 ICP-MS analysis of zinc content

λQ protein samples were analysed for metal content by ICP-MS at the University of Wisconsin Soil & PlantAnalysis Lab, located in Madison, WI

5.2.5 Limited Proteolysis, N-Terminal Sequencing, and Mass Spectrometry

Proteolytic digests were carried out in 20 mM Tris-HCl (pH 8), 50 mM NaCl, 5% (v/v) glycerol, 0.1mM EDTA. To form the λQ-DNA complex λQ was incubated with a 2.5-fold excess of Qut-containing DNA on ice for 30 minutes. 1 nanomole of λQ or λQ-DNA complex was digested in a 20 uL volume with differing concentrations of trypsin (1-10 picomoles), and chymotrypsin (1-15 picomoles) for 60 min at 25°C. Reactions were terminated by the addition of Laemmli loading buffer and boiling. Reaction products were analyzed on denaturing 4-12% gradient NuPage Bis-Tris gels (Invitrogen) followed by Coomassie blue staining.

Preparative samples for N-terminal sequencing were separated on 4-12% gradient NuPage Bis-Tris gels (Invitrogen) and blotted onto PVDF membrane (Bio-Rad). Fragments were excised and submitted for N-terminal sequence analysis at the protein sequencing facility at the University of Texas Medical Branch, Galveston, TX.

Samples for MALDI-MS were prepared using the ultra-thin layer method [8]. The MALDI plate was covered with a saturated solution of 4HCCA (1:3:2 (v/v) formic acid:water:isopropanol) allowed to dry for a few seconds and then the excess was wiped off. The samples were then spotted this plate, allowed to crystallize on the matrix and excess liquid was removed by vacuum. The spots were then overlaid with a drop of cold 0.1% TFA solution that was allowed to stand for 15 seconds before being removed by vacuum. MALDI-MS data was obtained on a Voyager linear time-of-flight instrument.

5.2.6 Electrospray Ionisation - Mass Spectrometry (ESI-MS)

Zachary Quinkert in the laboratory of Brian Chait conducted the ESI-MS analyses. Samples were were buffer exchanged into 0.5 M ammonium acetate using microcentrifuge columns (Bio-Spin, BioRad Laboratories, Hercules, CA) at room temperature prior to injection.

5.3 Results

5.3.1 Expression and purification of soluble λQ

The λQ protein was reported to be largely insoluble and has previously been purified from inclusion bodies and re-folded (eg. [9]). Re-folded denatured proteins can aggregate, which may explain why previous attempts to crystallize λQ failed [10]. Therefore, we modified the standard procedure by inducing expression for an extended time at a lower temperature with a low concentration of IPTG; 16 hours, 16°C, 0.1 mM IPTG. These modifications are thought to slow the rate of target gene transcription to synchronize with host translation and chaperone functions. We were thus able to obtain Q protein in a soluble form that could be purified to high homogeneity by standard chromatographic techniques without a re-folding step (see fig. 5.2). Typically we obtain a yield of 2 mg of λQ at >95% purity per liter of culture. This protein was found to be highly active in antitermination assays (personal communication; C. Roberts & P. Deighan).

5.3.2 Four strategies for crystallization of λQ were pursued

With soluble protein in hand we pursued four strategies for crystallization. The first was to screen λQ alone**.** This strategy was quickly dismissed as the λQ precipitated in many of the crystal screen conditions. The remaining strategies involved forming a complex between λQ and either 1) the QBE DNA, 2) $\sigma^{70}4$, or 3) the β flap. Protein complexes can be more amenable for crystallization as they form more globular forms with less disorder. Furthermore, it has been observed that in crystals of DNA-protein complexes the DNA drives crystallization by packing end-to-end to form a pseudo-continuous helix [11].

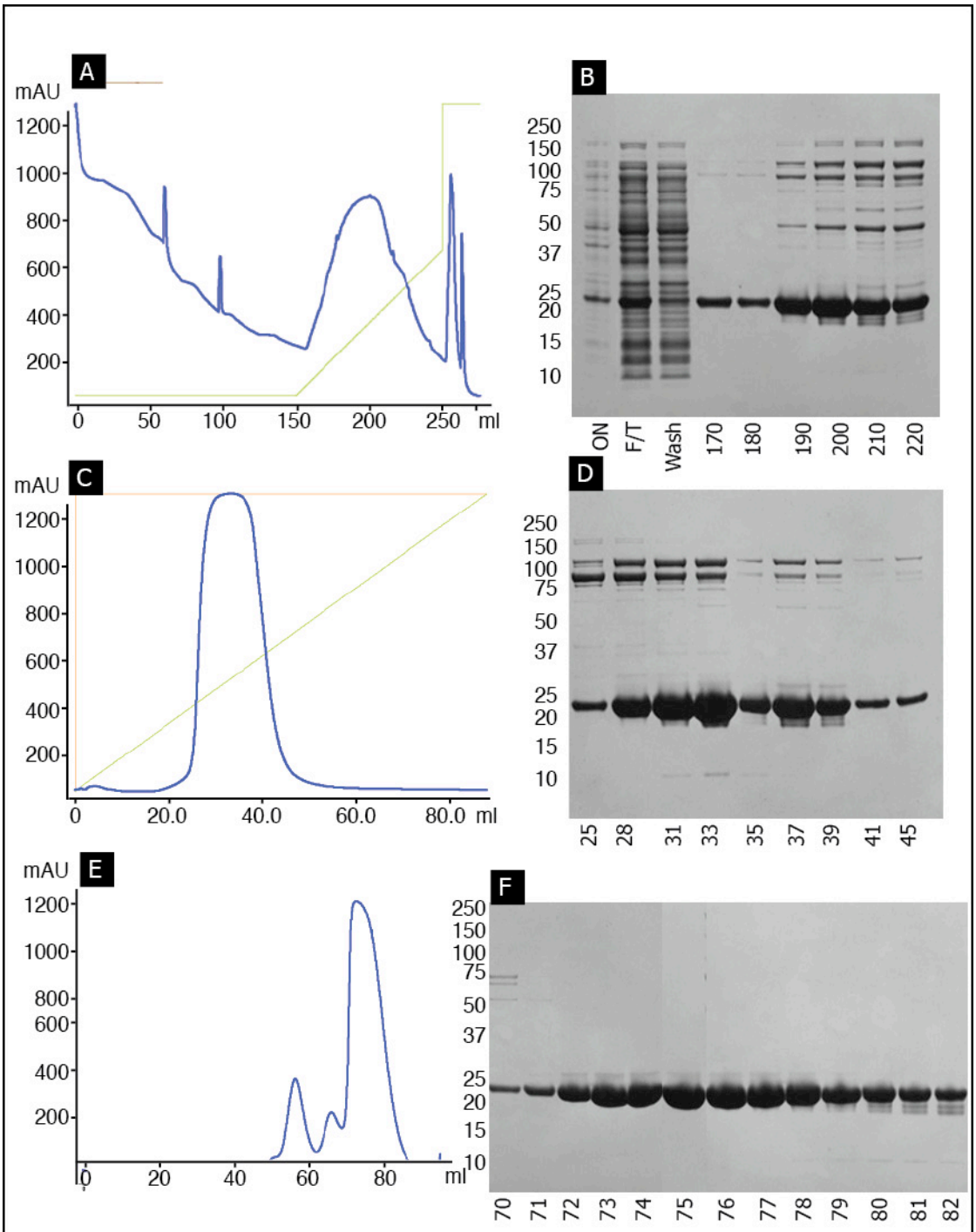


Figure 5.2 Purification of soluble λQ; chromatogram traces and SDS-PAGE analyses for the three column process.

A & B) Heparin column, C&D) Anionic exchange column, E&F) Gel filtration Chromotogram traces show UV traces for each purification step. Fractions were collected during each purification step and analysed by SDS-PAGE (NuPage 4-12% Bis-Tris gels) and stained with Coomassie blue dye.

5.3.3 Limited proteolysis indicates that the N-terminal region of λQ is unstructured.

Limited proteolysis has been used extensively in crystallography to probe for protease-resistant fragments likely to be in a compact globular form that facilitates ordered crystal packing. Secondary structure predictions suggest that the N-terminal region of λQ is unstructured, and this is supported by an observation that λQ was degraded during crystallization trials to a truncated form missing up to 56 N-terminal residues (Santangelo, PhD thesis). Therefore, we used various concentrations of trypsin, and chymotrypsin to probe the protease susceptibility of λQ in free-form and bound to DNA prior to setting up crystallization trials.

We found that λQ–DNA was more protease-resistant than free λQ but eventually digested in a similar pattern (see fig. 5.3). We mapped the resulting protein fragments using mass spectrometry and Edman sequencing. As expected, these fragments were missing residues from the N-termini. We repeated these experiments several times and based on these analyses, three truncated versions of Q, starting at residue 10, 32 and 56 were cloned into the Novagen™ pET 21a vector and all three were found to have a similar expression and solubility to the full length Q. The three truncated Q proteins were purified using the same protocol as for the full-length protein.

5.3.4 λQΔHis56 bound to DNA crystallizes but the crystals are not suitable for data collection.

The λQΔHis56 protein was incubated with four different DNA species corresponding to the Qut with additional nucleotides forming blunt or overhanging ends to facilitate crystal packing (see fig. 5.4 A). The resulting complex was used in crystallization trials using a sparse matrix designed for protein-DNA complexes, the Nucleix suite (Qiagen). Several candidate crystal forms were obtained including rectangular prisms that seemed very promising (see fig. 5.4 B). However, these crystals proved to contain only DNA and the various other "hits" were likewise found to contain just DNA, were too small for analysis or only diffracted weakly (eg. fig. 5.4 C).

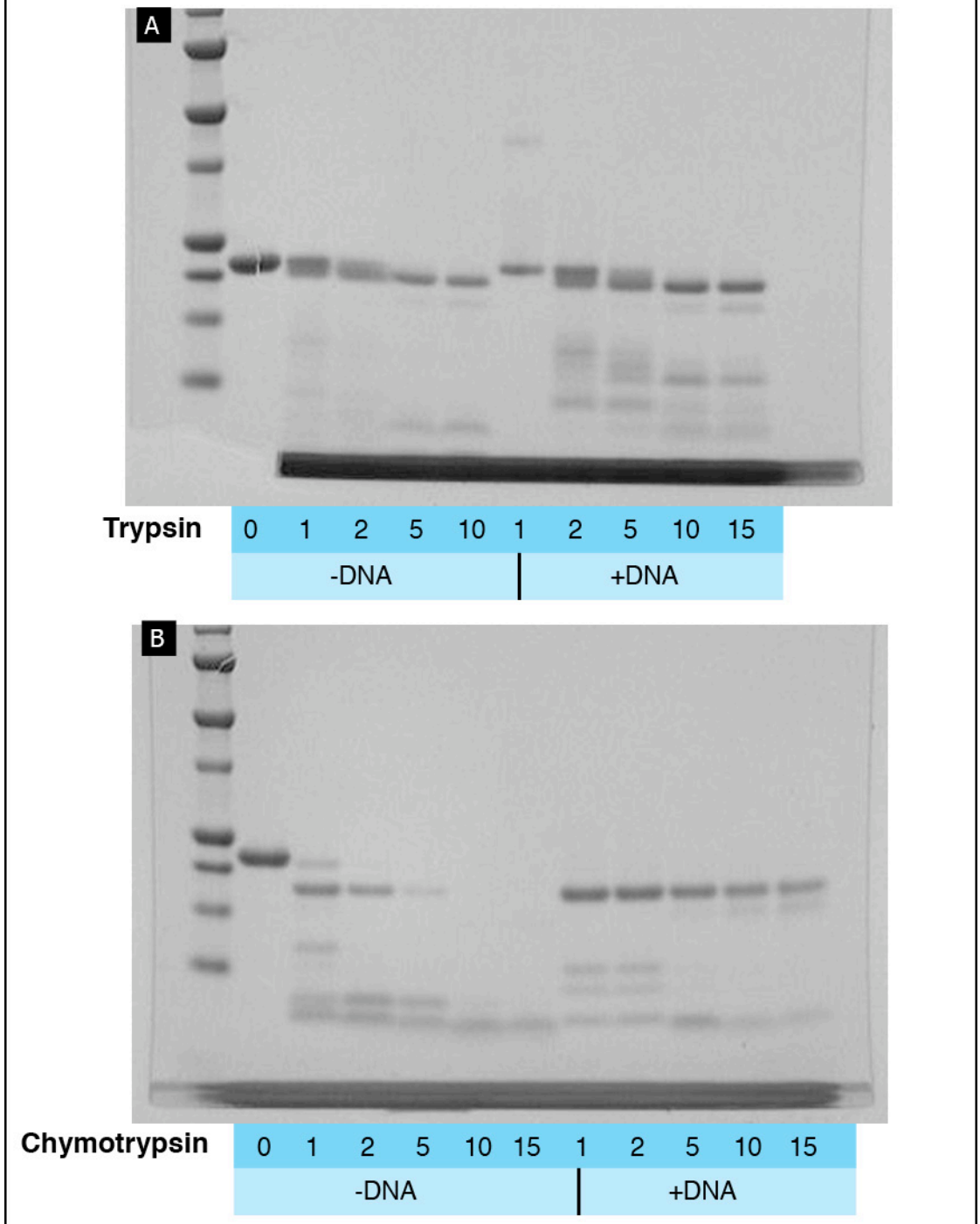


Figure 5.3 Proteolysis of λQ and λQ bound to DNA.

1 nanomole each of λQ and λQ bound to DNA were proteolysed with increasing amounts of A) trypsin (1-10 picomoles) and B) chymotrypsin (1-15 picomoles) and analyzed by SDS-PAGE using 4-12% NuPage Bis-Tris gels.

5.3.5 λQ binds zinc.

We were surprised to obtain multiple crystal forms that contained DNA alone as this has not occurred in other protein-DNA complex projects in the laboratory. Furthermore, the λQ had been shown to be active in anti-termination assays and was shown to bind DNA. λQ is presumed to bind zinc via a zinc-finger motif (cys 118, 121, 144, and 147; $CX_2CX_{22}CX_2C$) [12], and is also known to be sensitive to oxidation. However, zinc binding has not been established experimentally. Zinc often has a structural role and zinc content has been shown to be sensitive to both oxidation and to chelation by reducing agents used in purification buffers [13]. Therefore, we modified the buffers in the λQ purification and then submitted samples of the resulting protein for ICP-MS analysis of zinc content. λQ was found to contain 0.8 – 2.0 molecules of zinc per λQ monomer depending on the buffer content (see table 5.1). These results were somewhat inconsistent but could be interpreted as λQ tightly binding one zinc, presumably in the zinc-finger, and loosely binding a second zinc elsewhere. There is no obvious second site for zinc-binding although a sequence analysis tool ZincFinder [14] gave weak predictive scores for His 56, His122, Cys191 and Glu195 (see fig. 5.5).

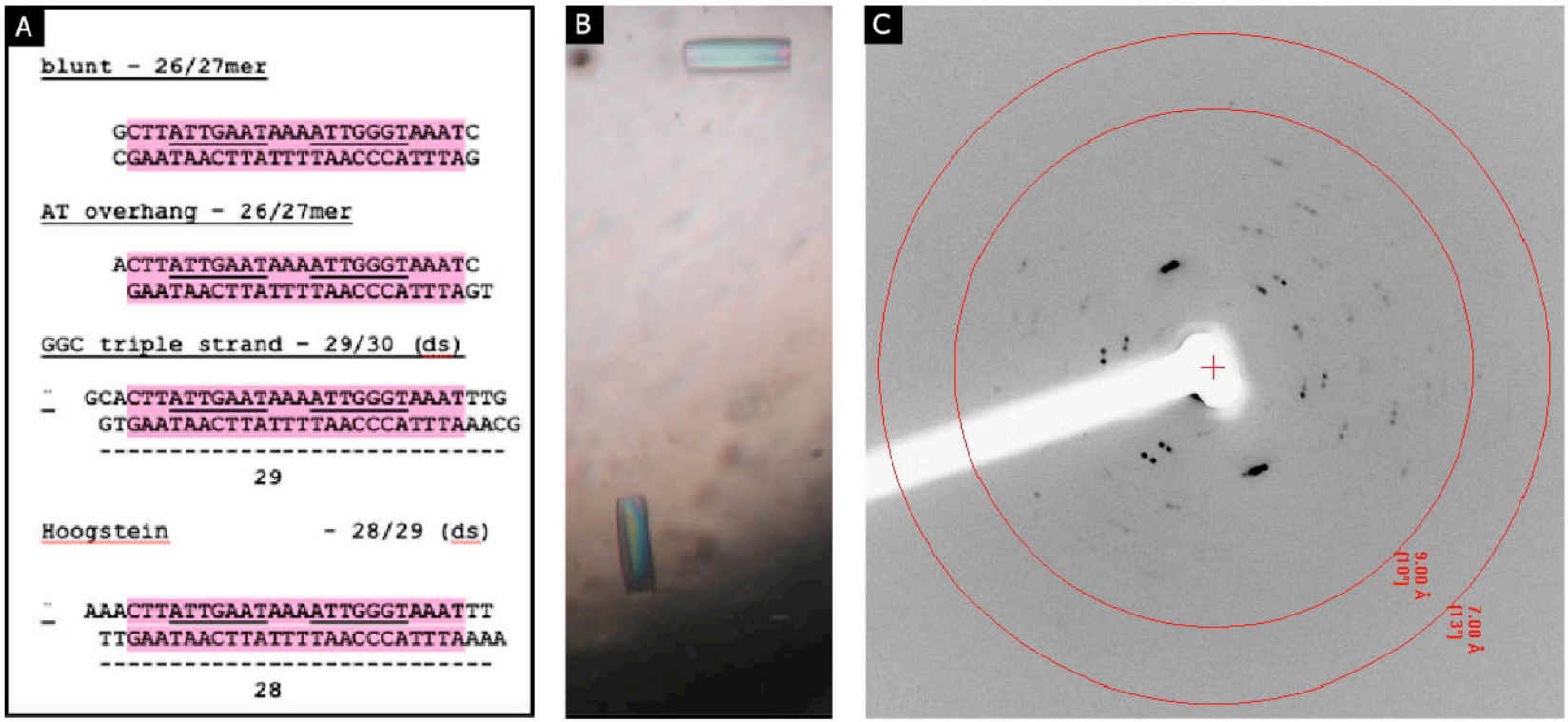


Figure 5.4 Crystallizing λQΔHis56 bound to DNA.

A) A series of double-stranded oligonucleotides were designed to incorporate the Q binding element (defined by Guo and Roberts 2004) and meet established guidelines to facilitate crystallization (Tan et al. 2000). B) λQΔHis56 bound to the blunt-ended DNA produced an initial hit of small rectangular prisms in Nucleix condition #96. These crystals were reproduced but were found to only contain DNA. C) Another hit in JCSG+ condition #38 weakly diffracted to about 8 Angstroms.

Table 5.1 ICP-MS analysis of zinc binding by λQ

Buffer conditions	Reducing agent in purification buffers	Reducing agent in storage buffer	[Lambda Q] mg/mL, μM	[Zinc bound] ppm, μM	**Zinc/λQ**
Standard	10 mM DTT	10 mM DTT	3.08, 137	10.59, 166	1.2
10 μM Zinc added	2 mM B-ME	2 mM B-ME	3.32, 148	12.02, 184	0.8
10 μM Zinc added	2 mM B-ME	10 mM DTT	2.79, 124	16.59, 254	2.0
Degassed,10 μM Zinc added	10-15 mM B-ME	10 mM B-ME	2.31, 102.8	12.31, 188	1.8
Degassed,10 μM Zinc added		10 mM DTT	2.25, 100.1	8.22, 125	0.8
Degassed,10 μM Zinc added*	10 mM B-ME	7.5 mM TCEP	1.32, 58.7	5.54, 84.8	1.4
Degassed,10 μM Zinc added*	10 mM B-ME	7.5 mM B-ME	1.32, 58.7	5.42, 82.9	1.4
Degassed,10 μM Zinc added*	10 mM B-ME	7.5 mM DTT	1.48. 65.9	5.15, 78.7	1.2

*Protein from this preparation was dialyzed before storage in two steps; for step 1 the buffer contained 10 mM B-ME and no zinc and thus removed any unbound zinc, for step 2 cthe buffers ontained the various reducing agents listed.

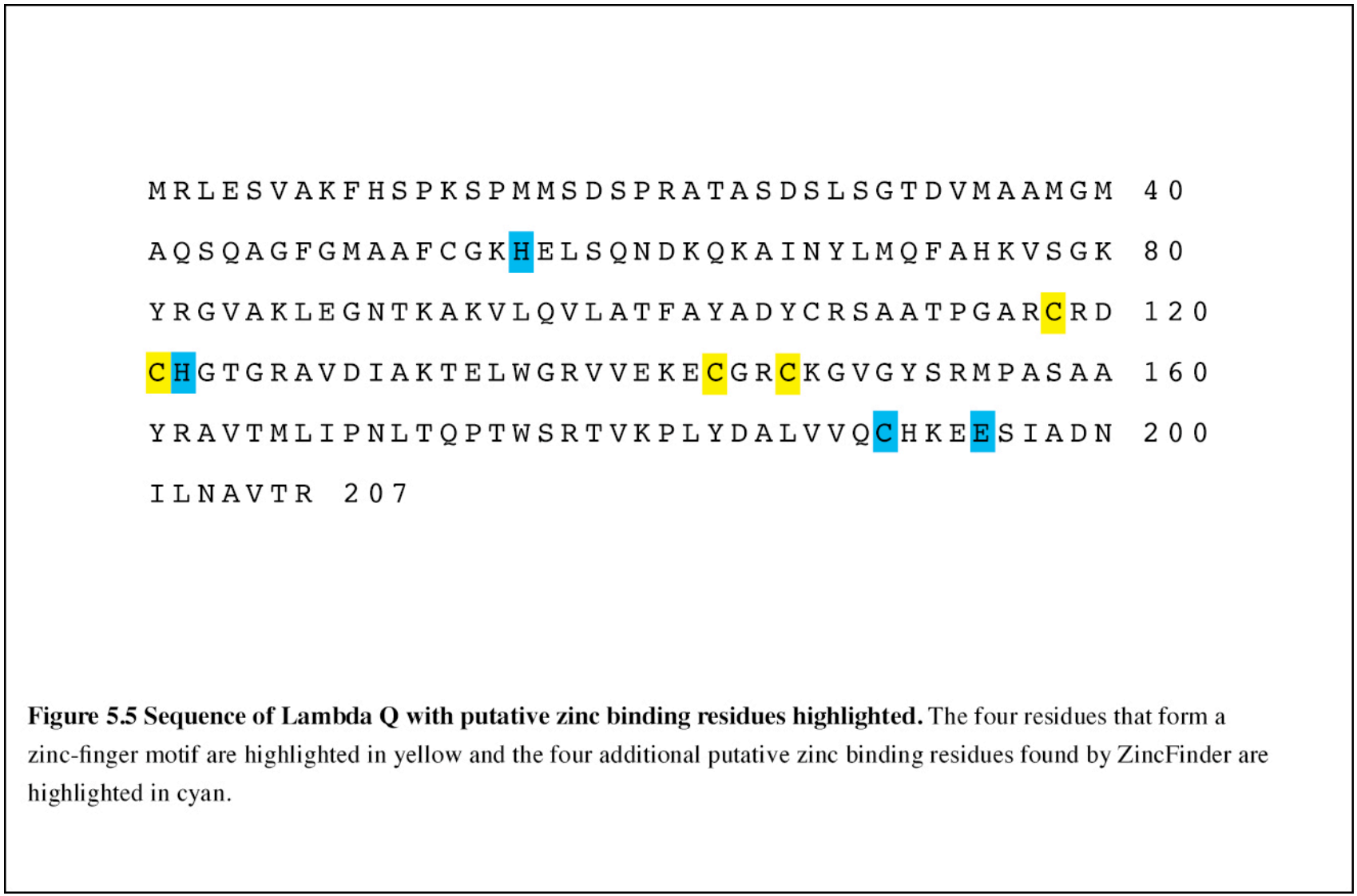


Figure 5.5 Sequence of Lambda Q with putative zinc binding residues highlighted. The four residues that form a zinc-finger motif are highlighted in yellow and the four additional putative zinc binding residues found by ZincFinder are highlighted in cyan.

5.3.6 λQ does not form a stable complex with $\sigma^{70}4$ in solution.

It has been shown that λQ interacts with $\sigma^{70}4$. The $\boldsymbol{\sigma^{70}}4$ alone is insoluble but was successfully co-crystallized with Rsd by another member of the laboratory [7]. We hoped that co-expressed λQ and his_6-$\sigma^{70}4$ would form a stable, soluble complex. However, the co-expressed proteins did not stably interact and his_6-$\sigma^{70}4$ was found mostly in the insoluble fraction and λQ in the soluble fraction (see fig. 5.6). Clarified cell lysate containing the co-expressed proteins was applied to a metal affinity column but no detectable his_6-$\sigma^{70}4$ or his_6-$\sigma^{70}4$/λQ complex bound (data not shown).

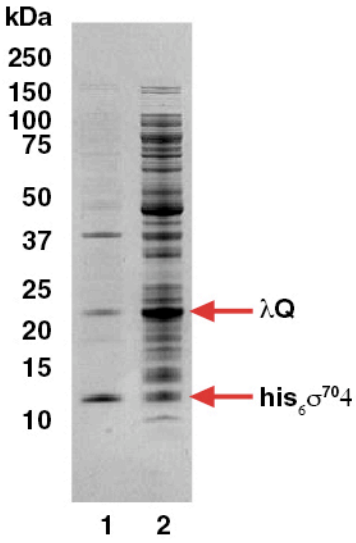


Figure 5.6 The co-expressed λQ and $his_6\sigma^{70}4$ do not form a stable complex. Cell lysate from BL21(DE3) cells transformed with a λQ-$his_6\sigma^{70}4$ co-expression plasmid was separated into insoluble (lane 1) and soluble fractions (lane 2) by centrifugation and analyzed by SDS-PAGE using 4-12% NuPage Bis-Tris gels. The bands assumed to correspond to λQ (mw 22.47 kDa) and $his_6\sigma^{70}4$ (mw 11.5 kDa) are indicated.

5.3.7 82Q co-purifies with RNAP.

Due to the variable results for λQ zinc binding and the unsuccessful crystallization trials, we decided to investigate whether another Q protein might be more suited for our studies. The lambdoid Phage 82 has a Q protein with no sequence similarity to λQ but with analogous anti-pausing and anti-termination functions. This protein has no cysteine residues and does not bind zinc. 82Q was cloned, expressed and purified in a similar manner to λQ. However, this protocol did not generate pure 82Q but rather enriched for an 82Q-RNAP complex (see fig. 5.6). An analysis by ESI-MS indicated that 82Q binds RNAP as a dimer to give a combined mass of 441,781 Da, (measured 444,232, see fig 5.7).

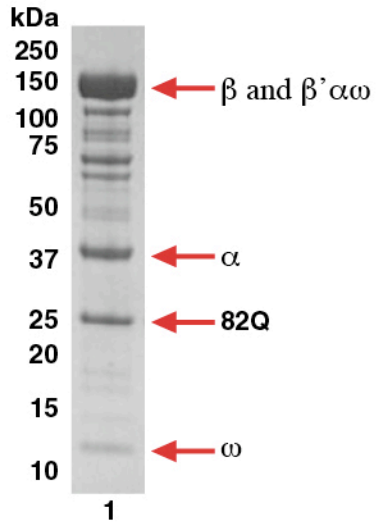


Figure 5.7 82Q co-purifies with RNAP

82Q prepared with the same protocol used for λQ (heaprin column, anioic exchange column and gel filtration) co-purified with RNAP. Shown here is the final "pure" fraction analyzed bySDS-PAGE using 4-12% NuPage Bis-Tris gels. The bands presumed to correspond to 82Q (mw 26.4 kDa) and the RNAP subunits αββ'ω are indicated.

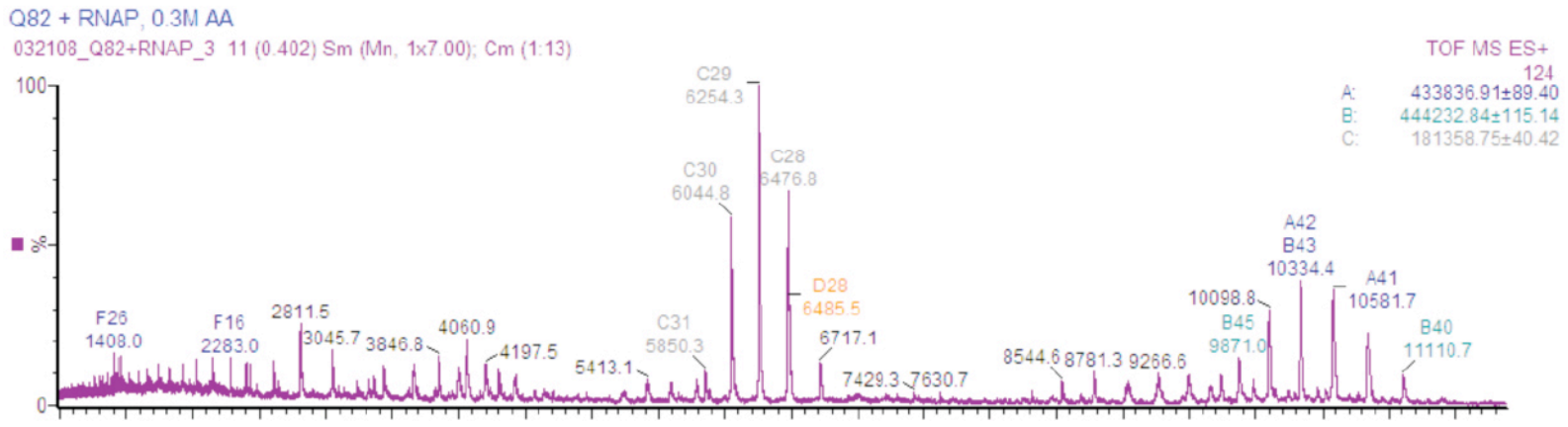


Figure 5.8 ESI-MS spectra of the 82Q-RNAP complex.

ESI-MS spectrum of the 82Q-RNAP complex; ion-series A corresponds to core-ω+a dimer of Q (theroretical mass 431,616 Da, measured mass mass 433,886 Da); series B corresponds to core+a dimer of Q (theroretical mass 441,781 Da, masured mass 444,232 Da), ion series C corresponds to an αβ complex (theroretical mass 187,144 Da, masured mass 181,358 Da)

5.3.8 λQ and 82Q *may* form stable complexes with the β flap.

The λQ interaction with $\sigma^{70}4$ is expected to be transient; σ^{70} is displaced from the elongating RNAP whereas λQ remains bound. Indeed, λQ binds the β flap and competes with $\sigma^{70}4$ for the same binding surface; the flap tip helix [2]. We co-expressed both λQ and 82Q with the β flap. When these proteins are co-expressed at 30°C there is no apparent complex formation. However, when these proteins are co-expressed at 16°C it is possible that they do form complexes (see fig. 5.8 and fig. 5.9). The His_6-λQ and the β flap are too close in size to distinguish them by SDS-PAGE, but the His_6-82Q and β flap are distinguishable and protein bands matching these proteins masses are eluted from ni+-beads analysed by SDS-PAGE. This crude purification is not definitive but is suggestive that 82Q, at least, does stably interact with the β flap.

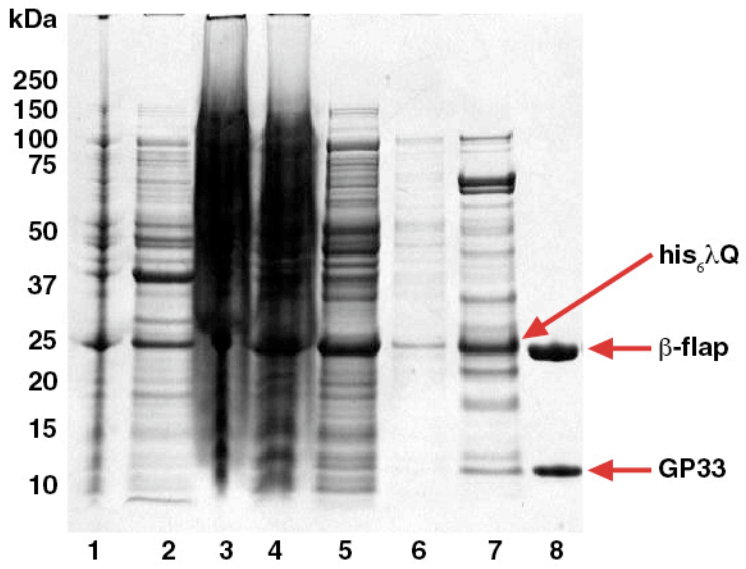


Figure 5.9 The co-expressed his_6λQ may form a stable complex with the β flap. Clarified cell lysate from BL21(DE3) cells transformed with a his_6λQ-β flap co-expression plasmid were incubated with Ni+-agarose-beads and then bound proteins were washed with low imidazole buffer and eluted with high imidazole buffer. Protein samples were analyzed by SDS-PAGE and tained with Coomassie blue dye(4-12% NuPage Bis-Tris gels) lanes; 1) whole cell extract, 2) insoluble fraction, 3) soluble fraction, 4) unbound proteins, 5) 5 mM imidazole wash 1, 6) 5 mM imidazole wash 2, 7) 250 mM imidazole elution, 8) purified gp33-β flap complex (included here as a marker).

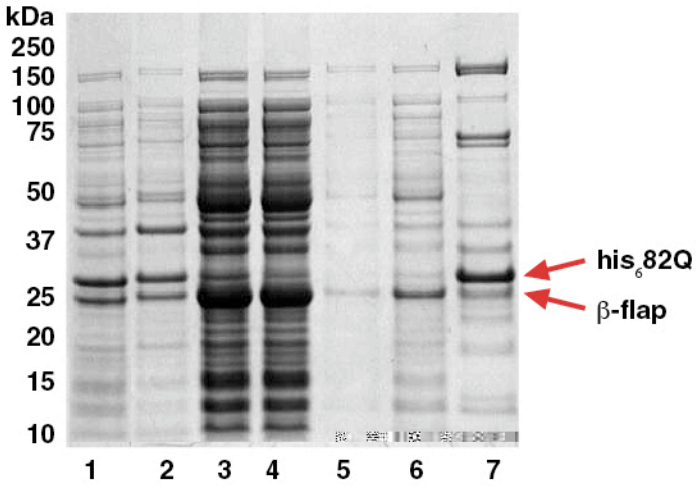


Figure 5.10 The co-expressed $his_6$82Q may form a stable complex with the β flap. Clarified cell lysate from BL21(DE3) cells transformed with a his_6-λQ-β flap co-expression plasmid were incubated with Ni+-agarose-beads and then bound proteins were washed with low imidazole buffer and eluted with high imidazole buffer. Protein samples were analyzed by SDS-PAGE (4-12% NuPage Bis-Tris gels) lanes;

1) whole cell extract, 2) insoluble fraction, 3) soluble fraction, 4) unbound proteins, 5) 5 mM imidazole wash 1, 6) 5 mM imidazole wash 2, 7) 250 mM imidazole elution. The bands assumed to correspond to $his_6$82Q (mw 27.5 kDa) and the β flap (mw 25 kDa) are indicated.

5.4 Discussion

The Q proteins were amongst the first proteins found to direct transcription by affecting the elongation step. A structure of Q may reveal the mechanism by which these proteins affect the RNAP to allow anti-transcription. We were unable to generate a crystal structure of Q but hope that the information generated by our studies will be useful for further work.

Although we were able to purify soluble λQ, this protein may not have a stable tertiary structure, and in particular may be prone to losing a loosely bound zinc molecule. We obtained weakly diffracting crystals of the λQΔHis56 protein bound to DNA but due to time constraints these crystals were not improved nor analyzed to confirm that they contained the intact λQΔHis56 protein.

We were unable to obtain a λQ-σ^{70}4 complex, possibly because this interaction is transient or possibly because the interaction is mediated by DNA [5]. The co-expression of either λQ or 82Q with the β flap may be a promising route to follow. The λQ interaction with the β flap can be strengthened by the mutation of three residues; T101I, A160V, T165A [2] and this mutant should be used in future work. The λQ(T101I, A160V, T165A) was shown to bind RNAP from solution *in vitro* without any requirement for the *Qut* DNA sequence.

The 82Q also had a robust interaction with the β flap interaction in bacterial two-hybrid assays [2]. While the λQ(T101I, A160V, T165A) was shown to bind RNAP from solution

in vitro, we found that 82Q co-purified with RNAP suggesting that 82Q is able to bind RNAP from solution *in vivo* without any dependence on phage DNA elements. A co-expression system for 82Q + RNAP has been made by cloning the $His_6$82Q gene into the pACYC-rpoZ plasmid described in chapter 4 which is compatible with the *Ec* RNAP expression vector, pVS10. This system could be used to produce the 82Q-RNAP complex for further analysis. It would be important to test whether this 82Q-RNAP complex was active in anti-termination assays as at this stage we do not know whether it is the biologically relevant complex.

Considering the difficulties we have had in obtaining a crystal structure of *Ec* RNAP alone, it seems unlikely that one could obtain a crystal structure of the 82Q-RNAP complex. However, the 82Q-RNAP complex may be suitable for cryo-Electron Microscopy (Cryo–EM). Previous Cryo-EM efforts in our laboratory focused on the (refolded) lambda Q protein and were unsuccessful – lambda Q was incubated with RNAP helical crystals but there was no density corresponding to lambdaQ in the resulting cryo-EM maps (personal communication N. Opalka). Either the 82Q or the λQ(T101I, A160V, T165A) with their higher affinity for RNAP may be more successful for this approach. Cryo-EM could provide a low resolution structure that would at least indicate where Q binds RNAP and may provide some low resolution details about Q function.

5.5 References

1. Roberts, J. W., Yarnell, W., Bartlett, E., Guo, J., Marr, M., Ko, D. C., Sun, H. & Roberts, C. W. (1998). Antitermination by bacteriophage lambda Q protein. *Cold Spring Harb Symp Quant Biol* **63**, 319-25.

2. Deighan, P., Diez, C. M., Leibman, M., Hochschild, A. & Nickels, B. E. (2008). The bacteriophage lambda Q antiterminator protein contacts the beta-flap domain of RNA polymerase. *Proc Natl Acad Sci U S A* **105**, 15305-10.

3. Murakami, K. S., Masuda, S. & Darst, S. A. (2002). Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 A resolution. In *Science*, Vol. 296, pp. 1280-4.

4. Korzheva, N., Mustaev, A., Kozlov, M., Malhotra, A., Nikiforov, V., Goldfarb, A. & Darst, S. A. (2000). A structural model of transcription elongation. In *Science*, Vol. 289, pp. 619-25.

5. Nickels, B. E., Roberts, C. W., Sun, H., Roberts, J. W. & Hochschild, A. (2002). The sigma(70) subunit of RNA polymerase is contacted by the (lambda)Q antiterminator during early elongation. *Mol Cell* **10**, 611-22.

6. Deighan, P. & Hochschild, A. (2007). The bacteriophage lambdaQ anti-terminator protein regulates late gene expression as a stable component of the transcription elongation complex. In *Mol Microbiol*, Vol. 63, pp. 911-20.

7. Patikoglou, G. A., Westblade, L. F., Campbell, E. A., Lamour, V., Lane, W. J. & Darst, S. A. (2007). Crystal structure of the Escherichia coli regulator of sigma70, Rsd, in complex with sigma70 domain 4. *J Mol Biol* **372**, 649-59.

8. Fenyo, D., Wang, Q., DeGrasse, J. A., Padovan, J. C., Cadene, M. & Chait, B. T. (2007). MALDI sample preparation: the ultra thin layer method. *J Vis Exp*, 192.

9. Marr, M. T., Datwyler, S. A., Meares, C. F. & Roberts, J. W. (2001). Restructuring of an RNA polymerase holoenzyme elongation complex by lambdoid phage Q proteins. *Proc Natl Acad Sci U S A* **98**, 8972-8.

10. Santangelo, T., Unpublished PhD Thesis. Cornell University.

11. Tan, S., Hunziker, Y., Pellegrini, L. & Richmond, T. J. (2000). Crystallization of the yeast MATalpha2/MCM1/DNA ternary complex: general methods and principles for protein/DNA cocrystallization. In *J Mol Biol*, Vol. 297, pp. 947-59.

12. Guo, J. & Roberts, J. W. (2004). DNA binding regions of Q proteins of phages lambda and phi80. *J Bacteriol* **186**, 3599-608.

13. Campbell, E. A., Greenwell, R., Anthony, J. R., Wang, S., Lim, L., Das, K., Sofia, H. J., Donohue, T. J. & Darst, S. A. (2007). A conserved structural module regulates transcriptional responses to diverse stress signals in bacteria. In *Mol Cell*, Vol. 27, pp. 793-805.

14. A. Passerini, C. A., S. Menchetti, A. Rosato, P. Frasconi,. (2007). Predicting zinc binding at the proteome level. In *BMC Bioinformatics*, Vol. 8:39.

Appendix 4.6 Sequences of the modified *rpoA* gene used to produce Ec. RNAPΔαCTD.

Plasmid backbone is pET21a (ampicillin resitance). Fragment cloned between *Nde* I (CATATG) and *Blp* I site (GCTNAGC).

1. DNA encoding *E. coli* alpha subunit in blue coloured text; PPX site (between residues 234-241) in red colored text. Uncleavable decahistidine tag appended to the 3 prime end of *rpoA* in green coloured text; cloned between *Nhe*I and *Nco*I / *Bam*HI sites.*PPX PreScission™ Protease Amersham cuts between Gln and Gly of **LEVLFQGP**

2. DNA encoding thermostable kanamycin resistance cassette in orange colored text

3. DNA immediately downstream of *T. thermophilus* HB8 *rpoC* (purple colored text) – this is present as the starter plasmid fro this construct was pRecT.thHB8*rpoC*::*PrA*.

CATATGcagggttctgtgacagagtttctaaaaccgcgcctggttgatatcgagcaagtgagttcgacgcacgccaaggtg
acccttgagcctttagagcgtggctttggccatactctgggtaacgcactgcgccgtattctgctctcatcgatgccgggttgcgc
ggtgaccgaggttgagattgatggtgtactacatgagtacagcaccaaagaaggcgttcaggaagatatcctggaaatcctgctc
aacctgaaagggctggcggtgagagttcagggcaaagatgaagttattcttaccttgaataaatctggcattggccctgtgactgc
agccgatatcacccacgacggtgatgtcgaaatcgtcaagccgcagcacgtgatctgccacctgaccgatgagaacgcgtctat
tagcatgcgtatcaaagttcagcgcggtcgtggttatgtgccggcttctacccgaattcattcggaagaagatgagcgcccaatc
ggccgtctgctggtcgacgcatgctacagccctgtggagcgtattgcctacaatgttgaagcagcgcgtgtagaacagcgtacc
gacctggacaagctggtcatcgaaatggaaaccaacggcacaatcgatcctgaagaggcgattcgtcgtgcggcaaccattct
ggctgaacaactggaagctttcgttgacttagaagttttatttcagggtccagtgaaagaagagaaaccagagttcgatccgatcc
tgctgcgccctgttgacgatctggaattgactgtccgctctgctaactgccttaaagcagaagctatccactatatcggtgatctggt
acagcgtaccgaggttgagctccttaaaacgcctaaccttggtaaaaaatctcttactgagattaaagacgtgctggcttcccgtg
gactgtctctgggcatgcgcctggaaaactggccaccggcaagcatcgctgacgagGCTAGCcatcaccatcaccatcat
caccatcaccattaaCCATGGGGATCCccgggagtataacagaaaccttaaggcccgaccgc**ttgaca**agggcgc
gtgaggttt**tacgat**agcgccgg***a***tgcggggaaaaagggctcctttgggggttttccccgcaccgggcggacctgggcgg
agaggaaacgcggcaactcgcccgtctcgggttcccgcccacgacccttaaggaggtgtgaggcatat**atg**aatggaccaata
ataatgactagagaagaaagaatgaagattgttcatgaaattaaggaacgaatattggataaatatggggatgatgttaaggctatt
ggtgtttatggctctcttggtcgtcagactgatgggccctattcggatattgagatgatgtgtgtcatgtcaacagaggaagcagag
ttcagccatgaatggacaaccggtgagtggaaggtggaagtgaatttttatagcgaagagattctactagattatgcatctcaggt
ggaatcagattggccgcttacacatggtcaatttttctctattttgccgatttatgattcaggtggatacttagagaaagtgtatcaaac
tgctaaatcggtagaagcccaaaagttccacgatgcgatttgtgcccttatcgtagaagagctgtttgaatatgcaggcaaatggc
gtaatattcgtgtgcaaggaccgacaacatttctaccatccttgactgtacaggtagcaatggcaggtgccatgttgattggtctgc
atcatcgcatctgttatacgacgagcgcttcg**gtc**ttaactgaagcagttaagcaatcagatcttccttcaggttatgaccatctgtg
ccagttcgtaatgtctggtcaactttccgactctgagaaacttctggaatcgctagagaatttctggaatgggattcaggagtggac
agaacgacacggatatatagtggatgtgtcaaaacgcataccatttt**taa**GGTACCgcccgccccgcaaggtggaggccc
ccggtttccccgggggccttttttccggcataatgggcccgtgcgcgcccggttgtacgccgcgttccgtcaggtgggcgagga
cctcttcgcccagggcctcatctccgccacggccgggaacttctccgtgcgcaccaaggggggttcctcatcaccaaaagcg
gggtgcagaaggcgaggctcaccccggaggacctcctggaggtgcccttggaagggcccatccccgaaggggcgagcgtg

gagagcgtggtccaccgggaggtctaccgcaggaccggggcccgggccctggtccacgcccacccccgggtggccgtggc
cctctccttccacctctcccgcctaaggcctttggacctcgagggccagcactacctgaaggaggtcccggtcctcgcccccaa
gacggtctccgccacggaggaggcggccttgagcgtggccgaggccttgcgggagcaccgggcctgcctcctgaggggggc
acggggccttcgccgtgggcctgaaggaggccccggaggaggcccttctcgaggcctacggcctcatgaccaccctggagg
agagcgcccagatcctcctctaccaccgcctctggcaggggcggggcccgccttgggggtggggaatgagggtcctcttc
gtggaggggaaggaccgggaggccctggtggccctggcggaggccctgccccatccctactggctcctcgaggGAATT
CGCTGAGC

Modified RpoA has a mass of 37, 930 Da. PPX cleavage results in two subunits CTD 11, 521 Da
NTD 26, 298 Da

Alpha_ppx_CTD_(His)10: 37, 930 Da
MQGSVTEFLKPRLVDIEQVSSTHAKVTLEPLERGFGHTLGNALRRILLSSMPGCA
VTEVEIDGVLHEYSTKEGVQEDILEILLNLKGLAVRVQGKDEVILTLNKSGIGPVT
AADITHDGDVEIVKPQHVICHLTDENASISMRIKVQRGRGYVPASTRIHSEEDERP
IGRLLVDACYSPVERIAYNVEAARVEQRTDLDKLVIEMETNGTIDPEEAIRRAATI
LAEQLEAFVDLEVLFQGPVKEEKPEFDPILLRPVDDLELTVRSANCLKAEAIHYIG
DLVQRTEVELLKTPNLGKKSLTEIKDVLASRGLSLGMRLENWPPASIADEASHHH
HHHHHHH.

Wildtype alpha 36, 510 Da
MQGSVTEFLKPRLVDIEQVSSTHAKVTLEPLERGFGHTLGNALRRILLSSMPGCA
VTEVEIDGVLHEYSTKEGVQEDILEILLNLKGLAVRVQGKDEVILTLNKSGIGPVT
AADITHDGDVEIVKPQHVICHLTDENASISMRIKVQRGRGYVPASTRIHSEEDERP
IGRLLVDACYSPVERIAYNVEAARVEQRTDLDKLVIEMETNGTIDPEEAIRRAATL
AEQLEAFVDLRDVRQPEVKEEKPEFDPILLRPVDDLELTVRSANCLKAEAIHYIG
DLVQRTEVELLKTPNLGKKSLTEIKDVLASRGLSLGMRLENWPPASIADE.

Chapter 6: Gp33 – a co-activator with a novel *modus operandi.*

6.1 Introduction

T4 bacteriophage infect *Ec* and hijack the host RNAP, re-directing it to transcribe the T4 genome. The general mechanisms for switching transcription to the T4 genome begins with the T4 DNA being distinguished from the host DNA by its glucosylated, hydroxymethylated cytosines [1]. This allows the T4 protein Alc to specifically terminate transcription on *Ec* DNA and allows phage nucleases to specifically digest host DNA. A finer mechanism is provided by the T4 proteins Alt and ModA, which ADP-ribosylate Arg265 of the RNAP α-CTD subunits, thus preventing activation by DNA UP-elements or DNA bound catabolite activator protein [1]. Alt also ADP-ribosylates other arginine residues on RNAP and other host proteins [1].

The temporal regulation of T4 genes involves manipulation and mimicry of σ^{70} function; promoter recognition and DNA melting. The domain organization of the σ^{70} family has been characterized based by sequence alignment [2], proteolysis [3], and structure [4; 5; 6; 7; 8]. There are four well-conserved domains; σ1 functions as an autoinhibitor [8], σ2 recognizes the -10 element (TATAAT) and mediates melting of double stranded DNA [4; 6; 7], σ3 binds extended -10 elements (TGnTATAAT) [9; 10] and σ 4 binds the -35 element (TTGACA).

The T4 genome has three classes of promoters; early, middle and late. T4 early genes have strong promoters consisting of a -35 element similar to *Ec* (consensus GTTTAC)

and a very strong -10 element that can extend even beyond the *Ec* extended -10 elements (eg. GTGGTATTAT), and are transcribed by σ^{70}-holoenzyme with and without ADP-riboslyations [1]. The general modifications combined with strong promoters are sufficient to support transcription of the T4 early genes [1].

Middle gene promoters also have strong -10 elements but the -35 element is functionally replaced by the Mot box, a 9 bp sequence centered at -30 that is recognized by T4-encoded motA [11]. MotA and AsiA are two T4 early gene products that modify the σ^{70}-holoenzyme so that it preferentially transcribes T4 middle genes. AsiA acts as an inhibitor by binding to $\sigma^{70}4$ and preventing it from interacting with -35 elements, and works in concert with MotA as an appropriator to recruit the σ^{70}-holoenzyme to middle gene promoters [11; 12]. AsiA and MotA also increase the rate that σ^{70}-holoenzyme clears the promoter and begins elongation [13].

The T4 late gene promoters consist of only an extended -10-like element (TATAAATA)[14]. Transcription from these promoters does not rely on the σ^{70} but rather on two T4 proteins, Gp55 and Gp33, that together form a functional σ analogue (see figure 6.1) [15; 16]. The -10 element is recognized by Gp55, a very divergent member of the σ^{70} family less than one third the size of σ^{70}[17]. Gp55 has weak sequence homology to σ2 and, like σ2, binds the RNAP β' coiled-coil [18]. Gp55 bound to *Ec* RNAP is sufficient for specific initiation at late promoters *in vitro* [16]. Gp33 has no homology to σ, yet binds to the RNAP β flap like σ4 and activates transcription[15]. This activation requires active

DNA replication and a subunit of the T4 DNA replisome, gp45 [19].

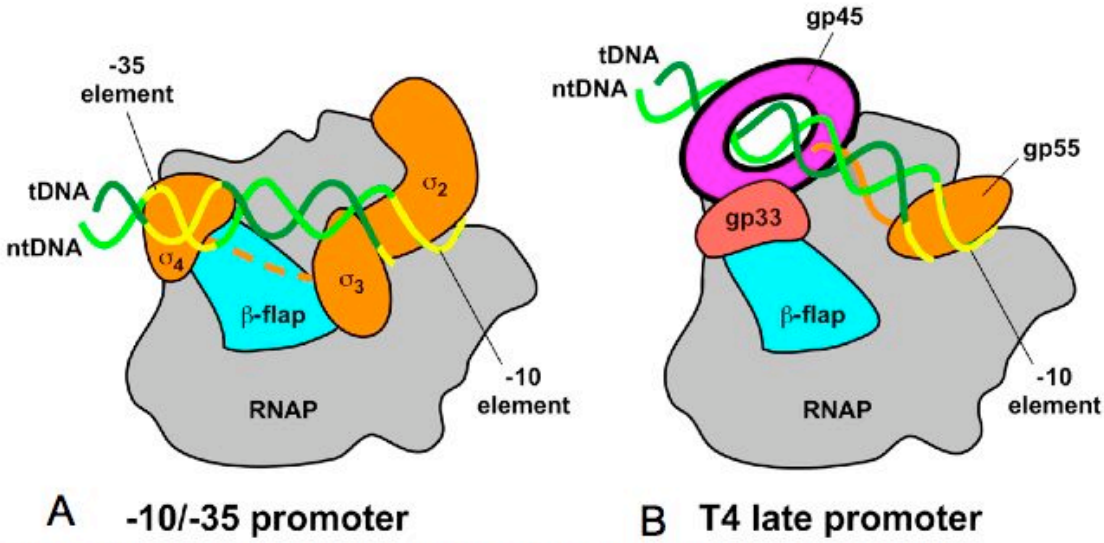


Figure 6.1 Comparison between a canonical σ70 promoter complex and a T4 promoter complex.
A) A model of a canonical σ70 -10/-35 promoter complex; σ2 is bound to the RNAP β' coiled-coil and interacts with the –10 element, while σ4 is bound to the RNAP β-flap, and recognizes the –35 element
B) A model of a T4 late promoter; gp55 has a homologous role to σ2 and similarly binds the β' coiled-coil and interacts with the –10 element, while gp55 is also bound to the RNAP β-flap there is no -35 element and instead gp55 iprovides an addional indirect contact with the DNA by interacting with gp45, the T4 DNA sliding clamp. tDNA = template DNA, ntDNA = non-template DNA

Gp45 is a sliding clamp protein, a subunit of the phage DNA replisome that is loaded onto DNA at the primer-template junctions in the replication fork by the clamp loader complex Gp44/Gp62 [20]. Trimers of Gp45 form a ring that encircles and slides along DNA, tethering the catalytic subunit of DNA polymerase, Gp43, to the DNA [21]. Gp45 actually binds both Gp55 and Gp33 to synergistically activate transcription, but the interaction with Gp33 has more substantial effect than the interaction with Gp55 [19]. The binding determinant for Gp45 in both these proteins (as well as Gp43) is a carboxy-terminal hydrophobic "tail" with the conserved motif (S/T)LDFL(F/Y/L) [19; 22; 23]. Gp45

can slide along the DNA to activate transcription from late gene promoters more than 1 kb from where it is initially loaded and can carry both Gp33 and Gp55 so that they too track along the DNA [24; 25].

The crystal structures of a number of σ^{70} family members have been solved and it has been observed that even distantly related members have highly conserved σ2 and σ4 structures [5]. It has been proposed that σ2 and σ4 are the minimal requirements for σ to recognize a bipartite promoter [5], but in the case of the T4 late gene promoters there is only a single DNA element for σ to recognize and thus a single σ domain is sufficient. Gp33 may be a functional analogue to σ4 however it has a distinct modus operandi; it does not bind DNA and instead provides an additional DNA-interaction by binding to Gp45. Therefore, it would not be expected to be structurally similar to σ4.

Gp33 is thought to be unstructured or in an extended alpha-helical conformation in solution [26], so we set-out to crystallize Gp33 in complex with the β flap. This complex was well behaved in solution and could be readily purified using standard chromatographic techniques in quantities suitable for crystallography. Here, I describe the preliminary results for the 3.3 Å-resolution X-ray crystal structure of the Gp33/*Ec* beta flap(831-1057) complex.

6.2 Methods

6.2.1 Protein expression and purification

An *Ec rpoB* gene fragment encoding amino acids 831-1057 (beta flap) was PCR amplified from an *Ec* RNAP expression vector, pVS10 and subcloned into a pet29a (Novagen) vector between NdeI/BamHI sites to generate pet29a_betaflap. The *T4 gp33* gene was PCR amplified from plasmid pGP33NHis (gift from Vikas Jain, Peter Geiduschek) and subcloned into a pET28a-based vector between NdeI/BamHI sites as a hexahistidine-fusion protein to generate pSKB2-GP33. The hexahistidine-Gp33 fusion was in turn PCR amplified and subcloned into pET29a-Betaflap between EcoRI/NotI sites to generate pet29a-betaflap-GP33.

BL21(DE3) cells transformed with pet29a-Betaflap-GP33 were grown in LB plus 50 μg ml^{-1} kanamycin at 37°C to mid-log phase, then the temperature was lowered to 16°C and expression was induced with 0.5 m*M* isopropyl β-d-thiogalactopyranoside for 16 hours. The cells were then harvested by centrifugation, resuspended in IMAC-A buffer (20 m*M* Tris pH 8.0, 500 m*M* NaCl, 5% glycerol, 0.5 m*M* β-mercaptoethanol) supplemented with 1 m*M* PMSF and EDTA-free protease-inhibitor cocktail (Sigma) and lysed using a French press. The clarified lysate was loaded onto an IMAC HiTrap column (GE Healthcare) charged with Ni^{2+} ions and pre-equilibrated with IMAC-A buffer. The column was washed with 10 column volumes of IMAC-A plus 5 mM imidazole, followed by 10 column volumes of IMAC-A plus 25 mM imidazole and then bound proteins were eluted with IMAC-A plus 250 mM imidazole. The eluted proteins were

digested overnight with PreScission protease (GE Healthcare) and simultaneously dialyzed against IMAC-A buffer at 4°C before being applied to a second, subtractive, IMAC HiTrap column to remove non-specific and uncleaved protein. The flowthrough was precipitated by adding ammonium sulfate to 60% saturation, pelleted by centrifugation and then resuspended in gel filtration buffer (10 mM Tris-HCl pH 8, 5% glycerol, 0.1 mM EDTA, 1 mM DTT). Finally, the GP33 - beta flap complex was purified by size-exclusion chromatography on a SD75 column (GE Healthcare).

Selenomethionyl-substituted protein was produced using the same protocol as for the native protein except that cells were grown in M9 minimal media supplemented with selenomethionine 10 mg/mL (CalBiochem) and 2.5 mg/mL isoleucine, lysine, and threonine to block endogenous methionine biosynthesis [27]. The selenomethionyl-substituted protein was purified using the same protocol as for the native protein except that the flowthrough from the second IMAC column was concentrated using centrifugal concentrators (Vivaspin, 5, 000 MWCO).

Both native and selenomethionyl-substituted protein were dialysed into a final storage buffer consisting of 10 mM Tris pH 8.0, 150 mM NaCl , 1% glycerol, 1 mM β-mercaptoethanol and 1 mM DTT.

6.2.2 Crystallization, data collection and processing.

Crystallization screening trials were set up using vapor difusion with the JCSG+ screen (Qiagen Research). 12 candidate conditions were identified; JCSG+ conditions #10, 12, 14, 15, 24, 59, 70, 74, 76, 78, 79, 80. Crystals suitable for structure determination were grown using hanging-drop vapor diffusion by mixing 1 μL of protein solution with 1 μL of reservoir solution above 0.5 ml of reservoir solution (0.1 M Tris pH 8.5, 0.1 M triethylamine N-oxide, 20% polyethylene glycol 2,000 monomethylether) with protein concentrations between 7.5 – 12 mg/mL at 22°C. These crystals belonged to the orthorhombic space group $I2_12_12_1$, had 67% solvent content and contained one molecule of each protein per asymmetric unit. Crystals were cryoprotected by slow exchange of crystallization solution with 0.1 M triethylamine N-oxide, 23% polyethylene glycol 2,000 monomethylether, 20% ethylene glycol and subsequent flash-freezing in liquid nitrogen.

A full dataset was collected on a single crystal to 3.1 Å at X25, National Synchrotron Light Source (NSLS), Brookhaven National Laboratory (BNL). The data was processed with HKL2000 and SCALEPACK. A homology model of the beta flap was built using MODELLER [28] with the taq core RNAP crystal structure [29] as the template and Ec residues 831-891, 911-930, 1037-1057 aligned with the evolutionary conserved taq residues 703-763, 783-802, 807-830. The preliminary phases were obtained from molecular replacement using PHASER.

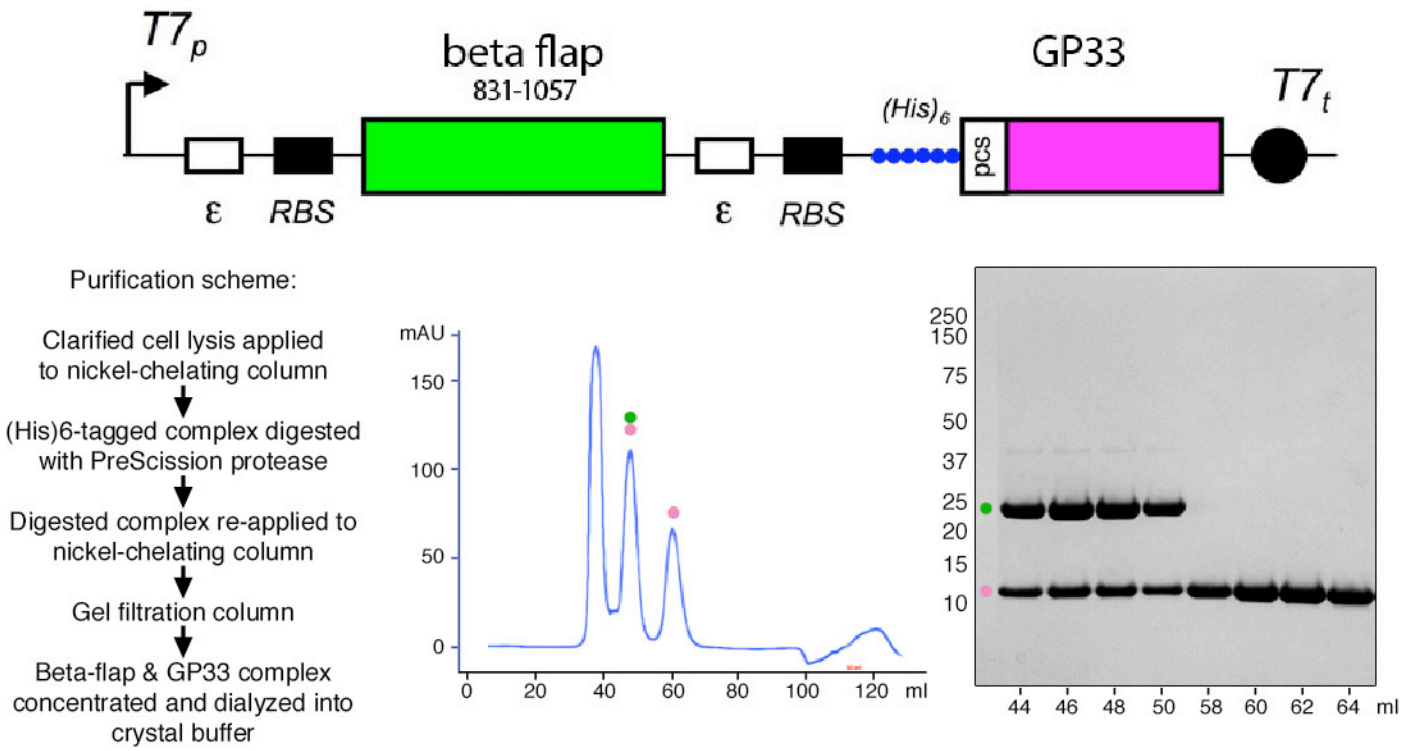


Figure 6.2 Purification scheme, gel filtration chromatography trace and SDS-PAGE analysis of the beta flap-GP33 complex. A) Schematic of the co-expression operon used to produce the beta flap- GP33 complex. B) Steps in the purification scheme. C) Chromatography trace of the gel filtration step of the purification with the peak containing the beta-flap-GP33 complex indicated by the pink and green circles and the peak containing the GP33 protein alone indicated by the pink circle. D) SDS-PAGE analysis of the fractions eluted from the gel filtration column showing fractions from each peak contain (pink circle indicates GP33, green circle indicates the beta flap.)

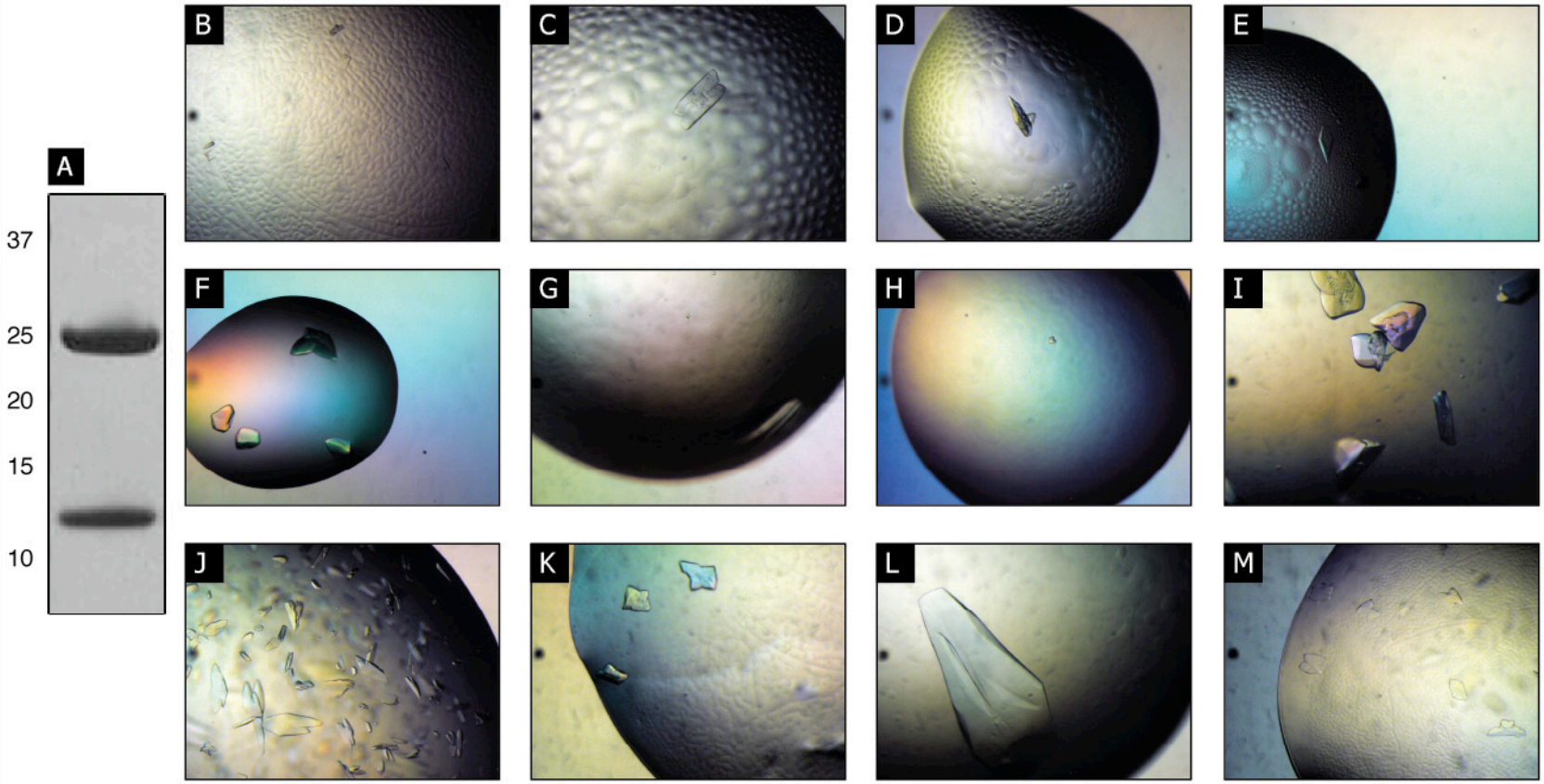


Figure 6.3 The beta flap – GP33 complex crystallized in 12 conditions in a commercial sparse matrix screen.

A) 10 µg of the purified complex was analyzed by SDS-PAGE and was seen to be free of contaminants. B-M) Crystals obtained in the JCSG+ sparse matrix screen; conditions 10, 12, 14, 15, 24, 59, 70, 74, 76, 78, 79, 80.

6.3 Preliminary results and discussion

6.3.1 Crystallization, structure determination and preliminary structure of the β flap (831-1057)/gp33 complex.

The β flap (residues 831-1057) was co-expressed with hexahistidine-tagged Gp33 and the soluble complex was purified to homogeneity using standard chromatographic techniques that included removal of the hexahistidine tag (see fig. 6.2 and methods). Typical yield was about 7 mg/L culture.

Crystals were obtained in 12 conditions from the JCSG+ screen (Hampton Research). These crystals varied from fine needle clusters to hexagonal plates as shown in figure 6.3.

A complete data set to 3.3 Å-resolution was collected from the hexagonal plates (approximate dimensions 200 x 40 x 10 μm) produced in condition #72 (a single diffraction image is shown in fig. 6.4). The data collection statistics are summarized in table 6.1. The crystals contained one 38.4 kDa Beta flap/Gp33 complex per asymmetric unit with a solvent content of 67%. The structure was solved using molecular replacement using an *Ec* homology model based on a portion of the Taq β flap (residues 831-890/912-935/1043-1057, 1L9Z) as the search model.

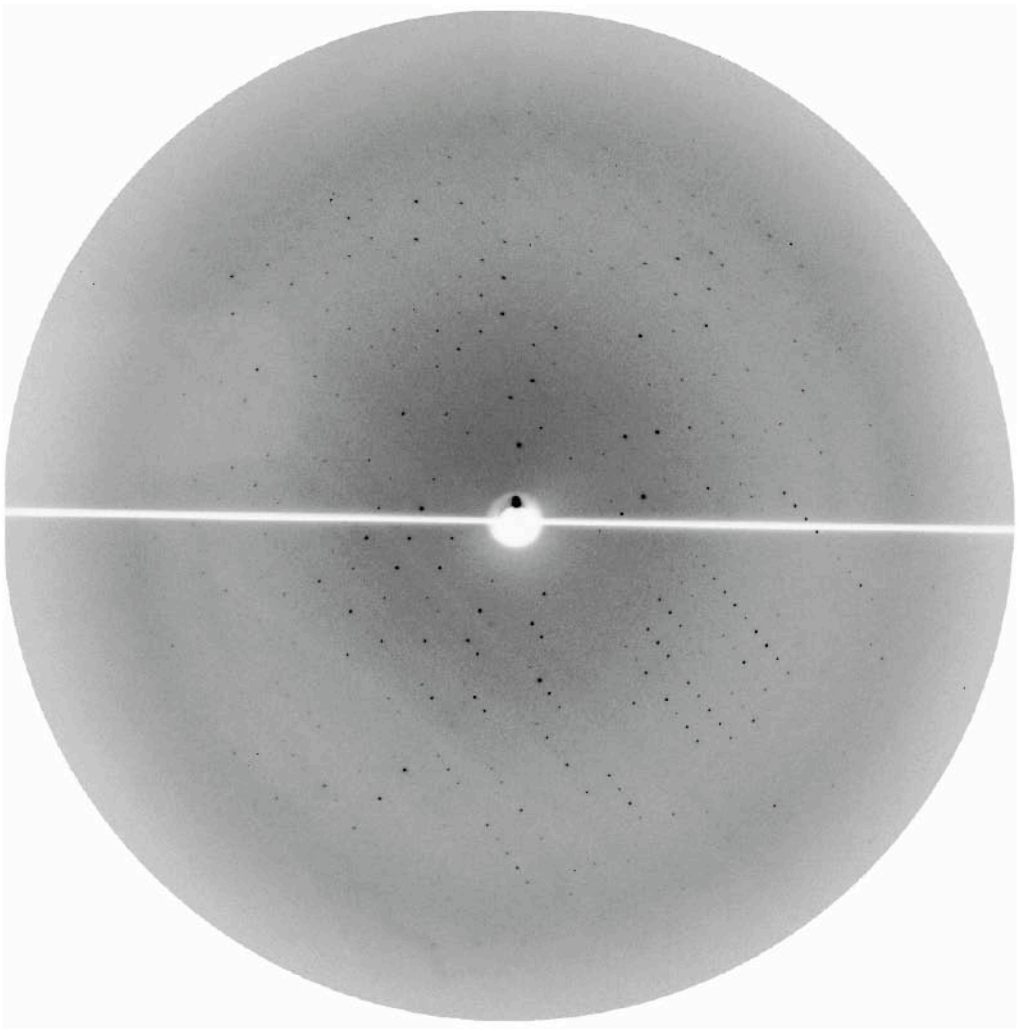


Figure 6.4 A single diffraction image collected from a native Betaflap/Gp33 crystal at the X25 beamline at NSLS, BNL.

We were able to do some very preliminary building into the electron density map (see fig. 6.5). There is electron density around the beta flap-tip helix that we assume to correspond to Gp33. This area of the map was not sufficiently clear to allow us to build into it, however, when we tried to superimpose the flap-tip helix/σ4 from the *taq* holoenzyme structure [6] it is obvious that the density does not correspond to a σ4 structural mimic.

We were able to build two long helices into the map in the region corresponding to *Ec* dispensable region 2 (DR2). This region is not conserved in the taq enzyme and is seen here for the first time. *Ec* DR2 forms two long, parallel helices connected at the end by a short helix**.**

Table 6.1 Diffraction and data statistics for E. coli β-flap/gp33 crystals.

	Native crystals	**Selenomethionine-substituted crystals**
Crystallization condition	0.1 M Tris pH 8.5, 0.1 M triethylamine N-oxide, 20% polyethylene glycol 2,000 monomethylether	0.2 M potassium citrate 20% PEG 3350
Space group	I222	I222
Unit cell	a = 55.04 Å	a = 55.218 Å
	b = 109.93 Å	b = 112.250 Å
	c = 165.73 Å	c = 164.847 Å
β-flap/gp33 complexes (38.4 kDa)/AU	1	1
Solvent content	65%	65%
Resolution (Å)	3.3	3.0
I/σI	21.0 (2.6)	21.9 (2.6)
Rsym	0.086 (0.473)	0.053 (0.412)
Completeness (%)	98.5 (90.2)	98.5 (90.2)
Redundancy	7.5 (6.6)	3.0 (2.7)

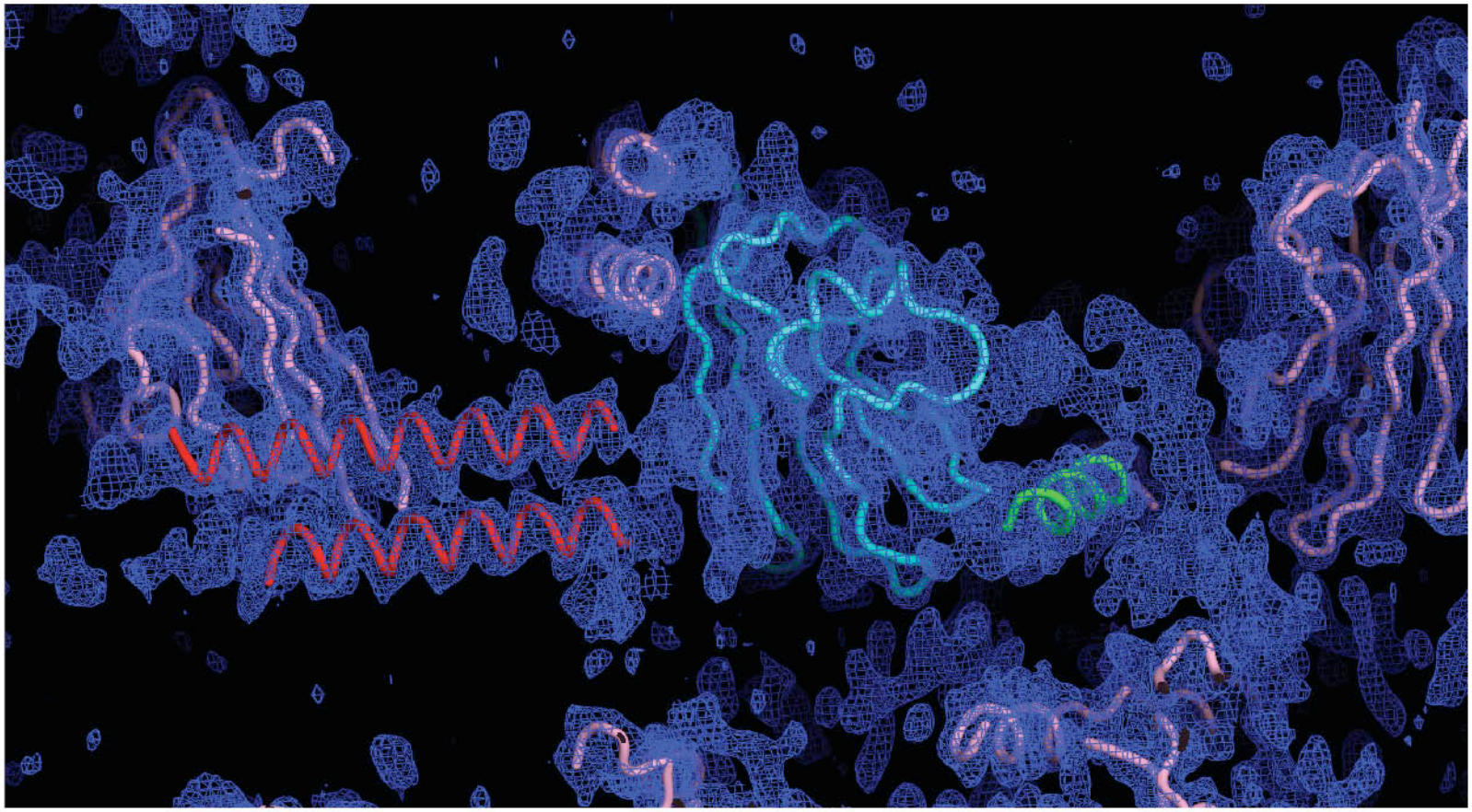


Figure 6.5 A view of the electron density map and preliminary model

Cyan: homology model of E. coli RNAP flap (based on the Taq RNAP beta flap, Taq RNAP beta residues 703-830). This contains E. coli RNAP beta subunit residues 831-890/912-935/1043-1057. This is what was used as the MR search model. Red: two long alpha-helices that belong to DR2.

6.5 Further structural analysis

We have obtained multiple data sets from crystals containing selenomethionine-substituted protein and will be using these to improve the phases of the current structure. We hope that these experimental phases will improve the electron density map and we will be able to build the Gp33 molecule as well as the missing beta-flap residues. The data collection statistics for the selenomethionine-substituted protein crystals is summarized in table 6.1.

6.6 References

1. Miller, E. S., Kutter, E., Mosig, G., Arisaka, F., Kunisawa, T. & Rüger, W. (2003). Bacteriophage T4 genome. In *Microbiol Mol Biol Rev*, Vol. 67, pp. 86-156, table of contents.

2. Lonetto, M., Gribskov, M. & Gross, C. A. (1992). The sigma 70 family: sequence conservation and evolutionary relationships. In *J Bacteriol*, Vol. 174, pp. 3843-9.

3. Severinova, E., Severinov, K., Fenyö, D., Marr, M., Brody, E. N., Roberts, J. W., Chait, B. T. & Darst, S. A. (1996). Domain organization of the Escherichia coli RNA polymerase sigma 70 subunit. In *J Mol Biol*, Vol. 263, pp. 637-47.

4. Malhotra, A., Severinova, E. & Darst, S. A. (1996). Crystal structure of a sigma 70 subunit fragment from E. coli RNA polymerase. In *Cell*, Vol. 87, pp. 127-36.

5. Campbell, E. A., Tupy, J. L., Gruber, T. M., Wang, S., Sharp, M. M., Gross, C. A. & Darst, S. A. (2003). Crystal structure of Escherichia coli sigmaE with the cytoplasmic domain of its anti-sigma RseA. In *Mol Cell*, Vol. 11, pp. 1067-78.

6. Murakami, K. S., Masuda, S. & Darst, S. A. (2002). Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 A resolution. In *Science*, Vol. 296, pp. 1280-4.

7. Vassylyev, D. G., Sekine, S., Laptenko, O., Lee, J., Vassylyeva, M. N., Borukhov, S. & Yokoyama, S. (2002). Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 A resolution. In *Nature*, Vol. 417, pp. 712-9.

8. Schwartz, E. C., Shekhtman, A., Dutta, K., Pratt, M. R., Cowburn, D., Darst, S. & Muir, T. W. (2008). A full-length group 1 bacterial sigma factor adopts a compact structure incompatible with DNA binding. In *Chemistry & Biology*, Vol. 15, pp. 1091-103.
9. Barne, K. A., Bown, J. A., Busby, S. J. & Minchin, S. D. (1997). Region 2.5 of the Escherichia coli RNA polymerase sigma70 subunit is responsible for the recognition of the 'extended-10' motif at promoters. In *EMBO J*, Vol. 16, pp. 4034-40.

10. Campbell, E. A., Muzzin, O., Chlenov, M., Sun, J. L., Olson, C. A., Weinman, O., Trester-Zedlitz, M. L. & Darst, S. A. (2002). Structure of the bacterial RNA polymerase promoter specificity sigma subunit. In *Mol Cell*, Vol. 9, pp. 527-39.

11. Hinton, D., Pande, S., Wais, N., Johnson, X. B., Vuthoori, M., Makela, A. & Hook-Barnard, I. (2005). Transcriptional takeover by sigma appropriation: remodelling of the sigma70 subunit of Escherichia coli RNA polymerase by the bacteriophage T4 activator MotA and co-activator AsiA. In *Microbiology (Reading, Engl)*, Vol. 151, pp. 1729-40.

12. Campbell, E. A., Westblade, L. F. & Darst, S. A. (2008). Regulation of bacterial RNA polymerase sigma factor activity: a structural perspective. In *Curr Opin Microbiol*, Vol. 11, pp. 121-7.

13. Adelman, K., Brody, E. N. & Buckle, M. (1998). Stimulation of bacteriophage T4 middle transcription by the T4 proteins MotA and AsiA occurs at two distinct steps in the transcription cycle. In *Proc Natl Acad Sci USA*, Vol. 95, pp. 15247-52.

14. Nechaev, S. & Geiduschek, E. P. (2008). Dissection of the bacteriophage T4 late promoter complex. In *J Mol Biol*, Vol. 379, pp. 402-13.

15. Nechaev, S., Kamali-Moghaddam, M., André, E., Léonetti, J. P. & Geiduschek, E. P. (2004). The bacteriophage T4 late-transcription coactivator gp33 binds the flap domain of Escherichia coli RNA polymerase. In *Proc Natl Acad Sci USA*, Vol. 101, pp. 17365-70.

16. Kassavetis, G. A. & Geiduschek, E. P. (1984). Defining a bacteriophage T4 late promoter: bacteriophage T4 gene 55 protein suffices for directing late promoter recognition. In *Proc Natl Acad Sci USA*, Vol. 81, pp. 5101-5.

17. Léonetti, J. P., Wong, K. & Geiduschek, E. P. (1998). Core-sigma interaction: probing the interaction of the bacteriophage T4 gene 55 promoter recognition protein with E.coli RNA polymerase core. In *EMBO J*, Vol. 17, pp. 1467-75.

18. Wong, K. (2003). Mutational and Functional Analysis of a Segment of the Sigma Family Bacteriophage T4 Late Promoter Recognition Protein gp55. In *Journal of Biological Chemistry*, Vol. 278, pp. 7073-7080.

19. Sanders, G. M., Kassavetis, G. A. & Geiduschek, E. P. (1997). Dual targets of a transcriptional activator that tracks on DNA. In *EMBO J*, Vol. 16, pp. 3124-32.

20. Spiering, M. M., Nelson, S. W. & Benkovic, S. J. (2008). Repetitive lagging strand DNA synthesis by the bacteriophage T4 replisome. In *Mol Biosyst*, Vol. 4, pp. 1070-4.

21. Moarefi, I., Jeruzalmi, D., Turner, J., O'Donnell, M. & Kuriyan, J. (2000). Crystal structure of the DNA polymerase processivity factor of T4 bacteriophage. In *J Mol Biol*, Vol. 296, pp. 1215-23.

22. Wong, K. & Geiduschek, E. P. (1998). Activator-sigma interaction: A hydrophobic segment mediates the interaction of a sigma family promoter recognition protein with a sliding clamp transcription activator. In *J Mol Biol*, Vol. 284, pp. 195-203.

23. Alley, S. C., Jones, A. D., Soumillion, P. & Benkovic, S. J. (1999). The carboxyl terminus of the bacteriophage T4 DNA polymerase contacts its sliding clamp at the subunit interface. In *J Biol Chem*, Vol. 274, pp. 24485-9.

24. Herendeen, D. R., Kassavetis, G. A., Barry, J., Alberts, B. M. & Geiduschek, E. P. (1989). Enhancement of bacteriophage T4 late transcription by components of the T4 DNA replication apparatus. In *Science*, Vol. 245, pp. 952-8.

25. Herendeen, D. R., Kassavetis, G. A. & Geiduschek, E. P. (1992). A transcriptional enhancer whose function imposes a requirement that proteins track along DNA. In *Science*, Vol. 256, pp. 1298-303.

26. Shao, W., Kearns, D. R. & Sanders, G. M. (1997). Secondary structure of T4 gene 33 protein. Fourier transform infrared and circular dichroic spectroscopic studies. In *Int J Biol Macromol*, Vol. 20, pp. 115-21.

27. Doublie, S. (1997). Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol* **276**, 523-30.

28. Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M. Y., Pieper, U. & Sali, A. (2007). Comparative protein structure modeling using MODELLER. *Curr Protoc Protein Sci* **Chapter 2**, Unit 2 9.

29. Zhang, G., Campbell, E. A., Minakhin, L., Richter, C., Severinov, K. & Darst, S. A. (1999). Crystal structure of Thermus aquaticus core RNA polymerase at 3.3 A resolution. In *Cell*, Vol. 98, pp. 811-24.