Mechanistic basis of plasmid-specific DNA binding of the F plasmid regulatory protein, TraM

by

Yun Peng

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

Department of Biochemistry
University of Alberta

ABSTRACT

The conjugative transfer of bacterial F plasmids relies on TraM, a plasmid-encoded protein that recognizes multiple DNA sites to recruit the plasmid to the conjugative pore. In spite of the high degree of amino acid sequence conservation between TraM proteins, many of these proteins have markedly different DNA binding specificities that ensure the selective recruitment of a plasmid to its cognate pore. Here we present the structure of F TraM RHH (ribbon-helix-helix) domain bound to its *sbmA* site. The structure indicates a pair of TraM tetramers cooperatively binds an underwound *sbmA* site that contains 12 base pairs/turn. The *sbmA* is composed of 4 copies of a 5 base pair motif, each of which is recognized by an RHH domain. The structure reveals that a single conservative amino acid difference in the RHH β-ribbon between F and pED208 TraM changes its specificity for its cognate 5 base pair sequence motif. Specificity is also dictated by the positioning of 2 base pair spacer elements within *sbmA*; in F *sbmA*, the spacers are positioned between motifs 1 and 2 and motifs 3 and 4, whereas in pED208 *sbmA* there is a single spacer between motifs 2 and 3. We also demonstrate that a pair of F TraM tetramers can cooperatively bind its *sbmC* site with an affinity similar to that of *sbmA*, in spite of a lack of sequence similarity between these DNA elements. These results provide a basis for the prediction of the DNA binding properties of the family of TraM proteins.

Acknowledgement

First and foremost, I would like to thank my supervisor, Dr. Mark Glover, for his guidance and encouragement. He has been so understanding and patient with me, and it is hard to imagine a better mentor.

I really appreciate all the help from my past and present lab members (Jun, Ross, Joyce, Sheraz, Charles, Stephen, Nico, Zahra, Curtis, Danny, Steve, Lucy, Yeyun, Gina, Mike, Ruth, Barbara and Inbal), especially Jun Lu, who introduced me to my project with his extensive knowledge and invaluable experience, and Ross Edwards for his expertise in crystallography date processing.

I would like to thank my committee members, Dr. Andrew MacMillan, Dr. Tracy Raivio and Dr. Richard Fahlman, who have been very supportive through my graduate studies.

Finally, a very special thanks and love to Minchang, who has supported and encouraged me all the time.

Contents

List of Table

List of Figures

Chapter 1

Chapter 3

Abbreviations

BCA	Bicinchoninic Acid
bp	Base Pair
DTT	Dithiothreitol
EMSA	Electrophoretic Mobility Shift Assay
H-NS	Histone-like Nucleoid Structuring protein
HGT	Horizontal Gene Transfer
HTH	Helix-Turn-Helix
IHF	Integrated Host Factor
IPTG	Isopropyl-b-D-thio-galactoside
K_d	Dissociation constant
kDa	Kilodalton
L	Liter
LB	Luria-Bertani
MALLS	Multi-Angle Laser Light Scattering
oriT	Origin of Transfer
PEG	Polyethylene Glycol
PCR	Polymerase Chain Reaction
P_M	*traM* promoters
Pu	Purine
Py	Pyrimidine
RHH	Ribbon-Helix-Helix
RMSD	Root-Mean-Square-Deviation
TBE	Tris-borate EDTA
tra	Transfer
Tris	Tris(hydroxymethyl)aminomethane

Chapter 1

Introduction

1.1 Bacterial conjugation and transmission of antibiotic resistance

The epidemic spread of antibiotic resistance has become a growing health threat [1]. Although bacteria can evolve to be more resistant under the selective pressure of antibiotics through vertical transfer, horizontal gene transfer (HGT) is the primary reason for the rapid spread [2]. The mechanisms for horizontal DNA transfer are bacterial conjugation, transduction and transformation. Conjugation is the most efficient way of HGT, and therefore is an important contributor to the diversification of the bacterial genome as well as the epidemic spread of antibiotic resistance.

Bacterial conjugation, the transfer of plasmid DNA from donor to recipient cells through direct cell-to-cell contact, was discovered by Joshua Lederberg and Edward Tatum in 1946 [3]. It is often regarded as the bacterial equivalent of sexual reproduction since it is associated with exchange of genetic material. Conjugation is often beneficial to the recipient. Benefits include antibiotic resistance, xenobiotic tolerance and expansion of metabolic versatility [4]. Conjugation is responsible for many instances of antibiotic resistance, including the early outbreak of multidrug resistant *Shigella* in Japan in the 1950's. It was discovered that multiple drug resistance can be easily transferred between *Escherichia coli* and *Shigella* in the intestinal tract [5]. Conjugation can occur between similar species as well as distantly related species, disseminating the ancient antibiotic resistance genes from the species in which the resistance originally evolved, to species only recently exposed to antibiotics [6].

Donor ability is determined by conjugative plasmids called fertility plasmids. F plasmid was the first conjugative plasmid identified, as an "infectious vector" [3], and was later named fertility (F) factor [7]. It is a 100 kb circular plasmid found in *E. coli* K-12 [8], and is the prototype for fertility plasmids in Gram-negative bacteria.

During F-mediated conjugation, the donor produces a pilus that makes specific contact with the recipient cell to form a stable mating pair [9]. Upon mating signal, the plasmid is nicked and a single-stranded DNA is transferred to the recipient in a 5' to 3' direction. Both cells are capable of transfer after recircularization of the transferred strand and synthesis of complementary strand in both donor and recipient cells [8]. The 33.3 kb transfer (*tra*) region extending from 77.7 kb to 100 kb in F plasmid encodes all the sequences required for conjugative ability [8]. The F plasmid in *E. coli* can be integrated into the bacterial chromosome via insertion sequence-mediated homologous recombination. An *E. coli* strain with integrated F plasmid retains its function as a donor in conjugation, and is able to transfer chromosomal genes to recipients with high efficiency [4]. F plasmid belongs to IncF incompatibility group, and the transfer regions of the members in this group, such as the R1, R100 and pED208 plasmids exhibit a high degree of homology with each other [8].

1.2 Mechanism and regulation of F-mediated conjugation

F plasmid conjugation is a biochemically complex process involving an array of host and plasmid-encoded proteins. Direct contact between the donor and recipient is mediated via the pilus. After the mating bridge is established, a DNA-processing complex, the relaxosome creates a single-stranded nick at a unique site (the *nic* site), and unwinds the DNA. The relaxosome is also important for the recruitment of the plasmid to the conjugative pore via interactions between the coupling protein TraM and TraD, the ATPase protein that is located at the cytoplasmic side of the pore [10]. Transfer occurs through a multi-protein type IV secretion system, the transferosome, which spans the entire cellular envelope [11].

Since bacterial conjugation is a nonessential process requiring significant energy, tight regulation and quick responses to physiological and environmental stimuli enhance the survival of plasmid-bearing cells (Fig. 1.1a) [12] [10]. The 33.3 kb transfer (*tra*) operon contains all the plasmid genes necessary for F transfer, which are regulated by three plasmid-encoded proteins: TraJ, TraY and TraM [8]. TraJ is transcribed from its own promoter P_J, and is the crucial activator of the P_Y promoter, from which the majority of *tra* genes are transcribed, including *traY* [13]. The increase of TraY is thought to further activate its own transcription [14], albeit conflicting evidence indicates that TraY also auto represses P_Y [13]. In addition, TraY activates the transcription of the *traM* operon [15]. TraM binds its own promoter to repress its own expression [16]. Conjugation is also regulated by a number of host factors.

As the primary activator of *tra* operon transcription, TraJ is subject to various regulation mechanisms at transcription, post-transcription and post-translation levels [12]. Upon entry into stationary phase, host nucleoid protein H-NS represses *traJ* transcription by binding to the *traJ* promoter [13], and the RNA chaperone Hfq (host factor for Qβ replicase) represses TraJ synthesis by destabilizing *traJ* mRNA [17]. In response to heat shock at 43°C, the heat shock chaperonin GroEL protein mediates TraJ degradation [18]. The Cpx (conjugative plasmid expression) extracytoplasmic stress response of *E. coli* activates and upregulates the expression of a protease-chaperone pair, HslVU, which degrades nascent TraJ to inhibit conjugation [19]. In most F-like plasmids, *traJ* is repressed by an antisense RNA (*finP*), which is stabilized by an RNA chaperone, FinO [20]. In F, the *finO* gene is interrupted by an IS3 element, leading to derepressed *traJ* expression and F conjugation [21].

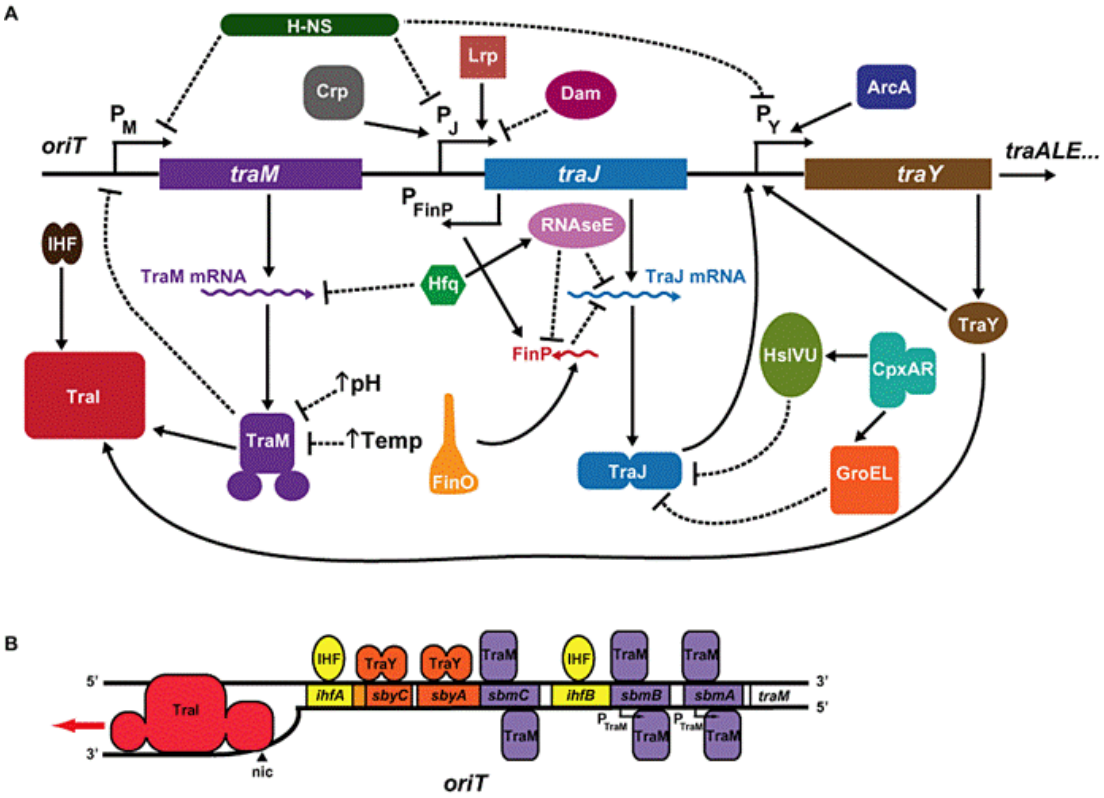

Figure 1. 1 *tra* operon regulation in F plasmid

(a) Overview of F plasmid *tra* operon regulatory factors. Positive regulatory effects are indicated by an arrow and solid lines, negative effects are indicated by a dash and dotted lines. (b) F plasmid *oriT* region with the binding sites for host and plasmid DNA-binding proteins indicated. The direction of TraI unwinding of DNA following cleavage at the *nic* site and covalent attachment to the 5′ end of DNA is indicated by a red arrow [10]. Adapted from Wong, JJW; Lu, J; Glover, JNM, "Relaxosome function and conjugation regulation in F-like plasmids - a structural biology perspective," *MOLECULAR MICROBIOLOGY*, vol. 85, pp. 602-617, 2012.

1.3 Proteins of the relaxosome

The relaxosome is a large nucleoprotein complex composed of TraI, TraY, TraM and host protein IHF (integrated host factor) assembled at the plasmid origin of transfer (*oriT*) (Fig. 1.1b) [10]. TraI is a large 192 kDa protein with two functions: relaxase and helicase. The two functional domains coordinate negatively to regulate transfer [22]. The relaxase domain of TraI cleaves one plasmid strand at the *nic* site within *oriT* and covalently links to the 5'-phosphate of the transferred strand. The helicase domain processively unwinds the DNA in a 5' to 3' direction. TraI is transported to the recipient cell along with the transferred strand and reverses the cleavage reaction to circularize the strand [23]. TraY binds to the P_Y promoter to regulate transcription of the *tra* genes [14] and facilitates the binding of TraI when bound to *oriT* as an accessory protein [24]. TraY has been predicted to be a member of the ribbon-helix-helix (RHH) family of transcriptional factors based on the sequence similarity with the Arc and Mnt repressors of bacteriophage P22 [25]. Mutations within the putative DNA-contacting β-sheet region resulted in reduced DNA binding affinity, consistent with the RHH prediction [26]. IHF is a host-encoded protein that binds two specific sites within *oriT*, and is required for relaxosome assembly and nicking [27]. TraM binds to three sites at *oriT* and enhances DNA nicking [28].

1.4 Roles of TraM protein in F conjugation

TraM is a 127 amino-acid cytoplasmic protein vital for F conjugation. Mutations that interrupt the *traM* gene reduce DNA transfer to background levels [8]. TraM is involved in relaxosome formation, coupling to the conjugative pore and regulation of the process.

1.4.1 TraM coordinates relaxosome formation

TraM, together with TraY, TraI and host protein IHF (integrated host factor), binds to the origin of transfer (*oriT*) of F plasmid to form a nucleoprotein complex, namely the relaxosome [28]. When conjugation is initiated, the topoisomerase-like domain of TraI creates a site- and strand-specific nick within *oriT* at *nic*, and the helicase domain of TraI subsequently unwinds the plasmid for transfer [29]. It has been demonstrated that TraM is a functional component of the relaxosome, since the addition of purified TraM could stimulate the TraI-catalysed transesterification reaction four- to five-fold [28]. TraM binds cooperatively to three sites within the F plasmid *oriT* region, termed *sbmA*, *sbmB* and *sbmC*, with *sbmC* located closest to the nick site, being the most important for conjugation [30]. The stimulation of relaxation may result from the significant distortion and unwinding of DNA caused by TraM binding [31]. Removal of the *sbmC* binding site abolishes the stimulation [28] and decreases transfer frequency significantly [32].

1.4.2 TraM mediates relaxosome-transferosome contacts

During F conjugation, the single-stranded DNA is transferred through a type IV secretion pore spanning the cell membranes [33]. The relaxosome generally locates at the mid- or quarter cell position, whereas the transferosome is formed at the cell membrane, implicating a mechanism required to bring the relaxosome to the transferosome for conjugation [11]. It was hypothesized that the plasmid is recruited to the transferosome through interactions between the TraM and the TraD coupling protein, an inner membrane component of the pore(Fig. 1.2b) [34]. TraM specifically contacts TraD, and mutations reducing the binding affinity of TraM for TraD result in decreases of conjugation efficiency [35]. This interaction has been proposed to

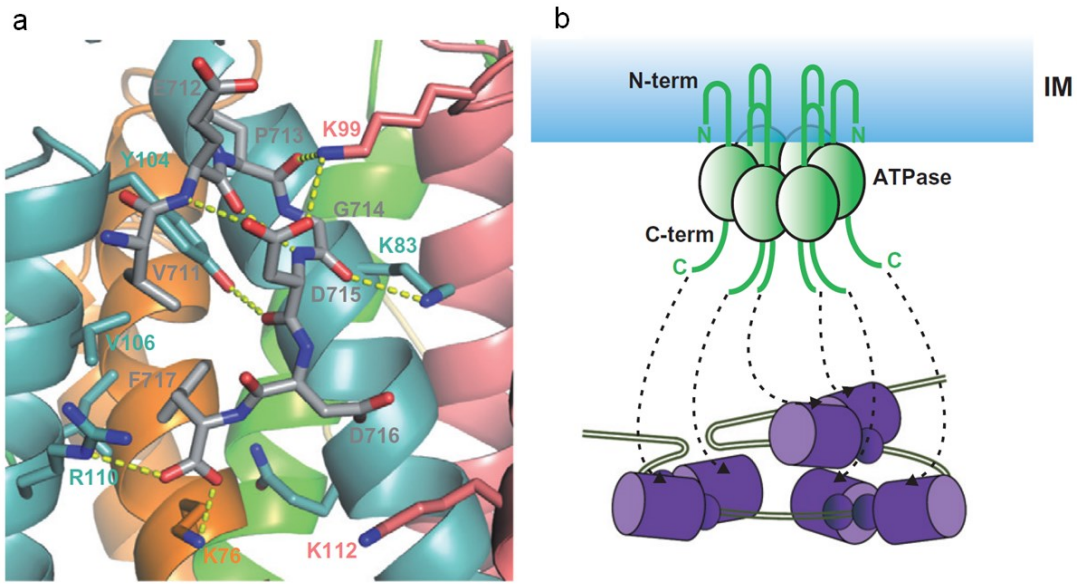

Figure 1. 2 TraM binding to TraD

(a) Detailed view of the structure of the F TraD C-terminal peptide (grey sticks) bound to the TraM C-terminal domain. (b) Model of TraM avidity effect in binding to TraD. IM, inner membrane. TraD is shown in green, and TraM in purple. TraM N-terminal domains are shown as ellipsoids, and TraM C-terminal domains are shown as cylinders. Multiple TraM tetramers are bound to three sbmA sites at *oriT* in a compact arrangement due to nucleosome-like DNA wrapping. The localized concentration of TraM tetramers facilitates interaction between multiple TraM binding sites and multiple TraD C-termini [10]. Adapted from Wong, JJW; Lu, J; Glover, JNM, "Relaxosome function and conjugation regulation in F-like plasmids - a structural biology perspective," *MOLECULAR MICROBIOLOGY*, vol. 85, pp. 602-617, 2012.

conduct the signal of cell-to-cell contact to the relaxosome to trigger DNA transfer [10].

1.4.3 Regulation of plasmid gene expression by TraM

At the level of transcription, TraM represses its own expression by binding to sites within the *traM* strong promoters (P_M) [16] to prevent potential toxicity from overexpression. Unregulated P_M in a plasmid that mimics F leads to decreased cell growth rate of host cells and reduced levels of F conjugation [36]. Cytotoxicity may result from the nonspecific binding to DNA when presented at high concentration [36, 30]. TraM also negatively regulates conjugation in a manner independent of gene regulation through its sensitivity to environmental stresses [37]. Lu *et al.* have shown that TraM is in equilibrium between deprotonated and protonated forms, and the protonated state is the active form for binding TraD [37]. Glu88, buried in the hydrophobic core of the TraM C-terminal helical bundle, is deprotonated at non-optimal pH or temperature. The deprotonation leads to destabilization of the TraM structure, thereby repressing conjugation. The sensitivity of TraM to elevated pH and temperature suggests a sensor to quickly repress conjugation under unfavourable conditions [37].

1.5. Structural basis of TraM function

F TraM contains two functional domains, an N-terminal DNA-binding domain and a C-terminal tetramerization domain [38]. The DNA-binding domain forms a dimer, and mediates specific recognition within *oriT* [39]. The C-terminal domain binds TraD, the inner membrane component of the conjugative pore, to mediate relaxosome-transferosome contacts [40]. Tetramerization is also required for efficient DNA binding as mutations in the C-terminal region affect binding to DNA fragments containing *sbmA, B* and *C* [39]. Multi-angle laser light scattering (MALLS)

and electrophoretic mobility shift assay (EMSA) indicated that a pair of TraM tetramers bind to a single *sbm* site. The binding is cooperative since no intermediate species is observed in EMSA, even at low TraM concentration [31].

The crystal structures of the C-terminal tail of TraD bound to the F TraM tetramerization domain [40] and full-length TraM from F-like plasmid pED208 bound to its cognate *sbmA* site (Fig. 1.3a) [31] revealed intriguing specificity and regulatory mechanisms of TraM.

1.5.1 The N-terminal DNA-binding domain

Members of the TraM family proteins are quite conserved throughout the N-terminal domains. Sedimentation analysis shows that both F and R1 TraM N-terminal domains form dimers [38, 41]. Mutational analysis of F TraM defined a 10 aa N-terminal region crucial for DNA binding, indicating the participation of this region in direct contact with DNA [39]. The distribution pattern of the residues involved in DNA recognition lead to a hypothesis that the F TraM DNA-binding domain forms a ribbon-helix-helix (RHH) domain, which is common in bacterial transcriptional repressors [31, 42]. Indeed, the structure of pED208 TraM bound to its *sbmA* DNA reveals that the N-terminal regions dimerize to form RHH folds, and the β-ribbon region provides critical residues for specific DNA recognition that make direct contacts with DNA base pairs [31].

The pED208 TraM structure solved by Wong *et al*. shows two tightly intertwined RHH motifs in each dimer, and the β-strands at the N terminus of each monomer pair to form an anti-parallel β-sheet [31]. Three amino acid side chains protruding from each β-strand Lys3, Gln5 and Tyr7, interact with the palindromic GANTC motif in a sequence-specific and symmetric manner through hydrogen

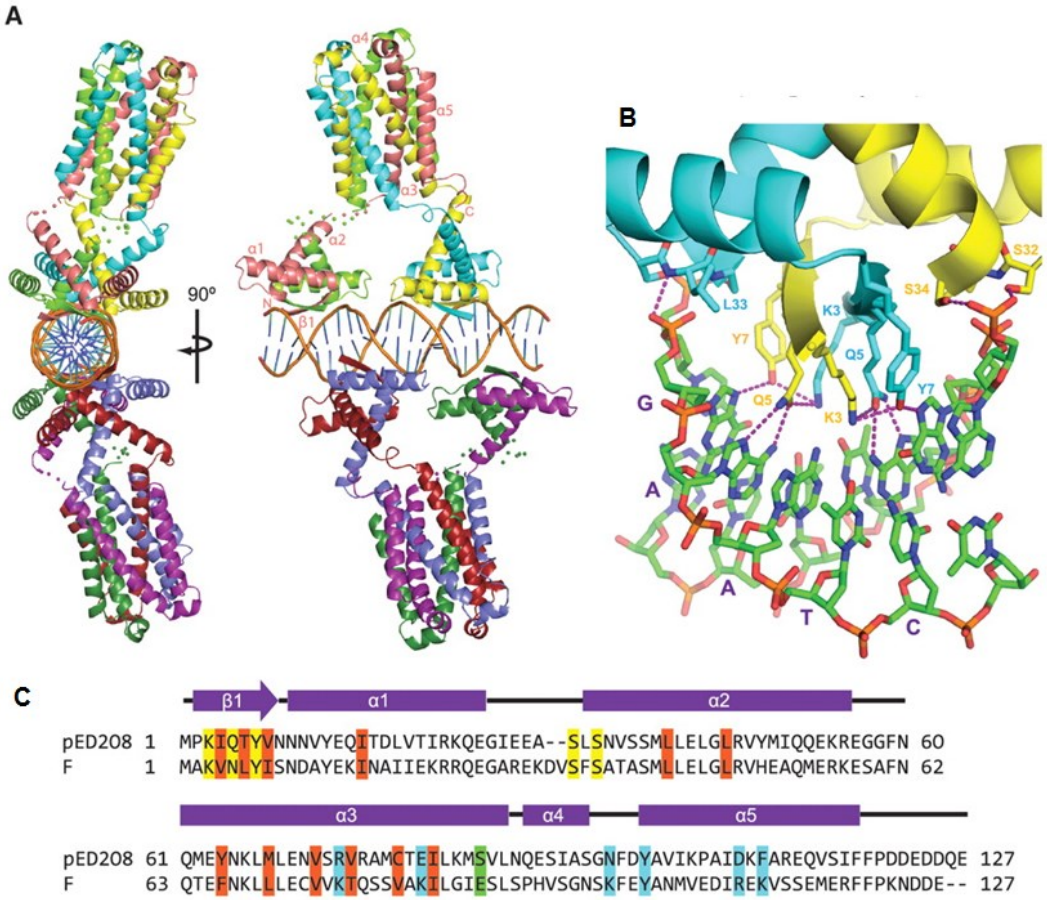

Figure 1. 3 Crystal structure of pED208 TraM bound to *sbmA*

(a) Orthogonal views of the overall structure of the TraM–*sbmA* complex. The α-helices and β-strands are indicated. Disordered linkers are indicated by spheres, one for each Cα that could not be refined. (b) Interactions between the N-terminal domain of TraM and DNA. Hydrogen bonds are indicated by purple-dashed lines. The GAATC-binding motif consisting of bases G7 to T11 of *sbmA* is indicated with purple letters. (c) Sequence alignment of pED208 TraM and F TraM. Secondary-structure elements are indicated. Conserved hydrophobic core residues are highlighted in orange, DNA-contacting residues in yellow, TraD C-terminal tail-contacting residues in cyan and position 88, responsible for protonation-mediated destabilization of the F TraM tetramer, in green [31]. Adapted from Wong, JJW; Lu, J; Edwards, RA; Frost, LS; Glover, JNM, "Structural basis of cooperative DNA recognition by the plasmid conjugation factor, TraM," *Nucleic Acids Research*, vol. 39, pp. 6775-6788, 2011.

bonds within the DNA major groove (Fig 1.3b). The N-terminus of the second helix α2 also makes non-specific contacts with the DNA phosphate backbone on either side of the major groove. This interaction orients and facilitates β-sheet binding, anchoring the RHH dimer to the DNA [31].

1.5.2 The C-terminal tetrameriztaion domain and TraD interactions

Lu *et al.* presented the crystal structure of the F TraM C-terminal domain, residues 58 to 127. Each of the four intertwined protomers donates a pair of α-helices to form a symmetric eight-stranded helical bundle. The N-terminal helix of each protomer composes a parallel, four-stranded coiled coil at the center of the structure, and the C-terminal helix of each protomer wraps around the outside of the structure, antiparallel to the N terminal helices [37]. The crystal structure of the TraM C-terminal domain bound to the TraD C-terminal tail provided structural evidence of specific contacts between the relaxosome and transferosome [40]. The TraD peptide forms a β-turn and interacts with the highly basic pocket on each of the four symmetry-related faces of the TraM tetramer. The side-chain of the C-terminal phenylalanine of TraD is critical for TraM recognition, which binds to the hydrophobic center of the TraM pocket. The C-terminal carboxylate of TraD also directly recognizes the guanidinium group of Arg110 and the amino group of Lys76 of TraM [40].

1.5.3 Overall structure of pED208 TraM-*sbmA* complex

The crystal structure of pED208 TraM bound to its cognate *sbmA* shows two TraM tetramers bind *sbmA*, which is consistent with the MALLS (multi-angle laser light scattering) and EMSA results of F TraM [31]. The two tetramers bind to opposite sides of the DNA without direct contact between the tetramers (Fig.1.3a). Each tetramer provides two RHH dimers to specifically recognize the palindromic

GANTC motif, which is repeated four times within the *sbmA* site. The *sbmA* DNA used for crystallization is 24 bp, containing a 2 bp spacer at the center between the second and third GANTC motifs, and a flanking bp at each end of the DNA [31].

1.5.4 Cooperative binding mechanism of pED208 TraM

Cooperative binding of DNA is common in transcription factors, and is generally mediated through oligmerization of the DNA-binding proteins. However, the cooperative binding of pED208 TraM is not realized through direct contact between the two tetramers, but rather acts through the DNA itself [31]. Analysis of the DNA structure in the complex demonstrates that it is significantly distorted and underwound, resulting in an average helical twist of 32° compared to 36° per base pair in standard B-DNA. The unwound DNA double helix, which is about 12 bp per turn, aligns the alternating GANTC motifs on the same side of the DNA to facilitate the binding of two RHH dimers from the same tetramer. In this way, the binding of one tetramer to one side of the *sbmA* site would unwind and align the unbound GANTC sites on the other side of the DNA for recognition of the second TraM tetramer, promoting the cooperative binding of *sbmA* [31]. This unwinding has been proved to be crucial for cooperative recognition, since a *sbmA* derivative containing only one pair of GANTC motifs separated by 12 bp binds TraM poorly and reduction of the spacing to 11 bp partially rescued the binding. Further reduction to 10 bp spacing abolished binding. The loss of interaction is explained by modeling of two RHH mounted to the GANTC motifs on the same side of DNA separated by 10 bp, which leads to severe clashes between the α1 helixes and α1- α2 loops [31].

In addition, the DNA helix axis is kinked compared to the straight axis in B-form DNA. Contacts between TraM α2 helix N-termini and the DNA widen and deepen the major groove. The repulsive interactions between acidic residues in the

α1- α2 loop and the DNA phosphate backbone leads to DNA bending to minimize the unfavorable interaction. The DNA backbone is pushed away from one RHH to help wrap the next GANTC motif around the second RHH domain [31].

1.6 Ribbon-helix-helix superfamily

The ribbon-helix-helix (RHH) superfamily of transcription factors is widespread in prokaryotes. The functional unit of the RHH fold is a dimer, which pairs through N-terminal β-strands to form an antiparallel β-sheet [42]. The two α-helices following the β-strand are similar to the helix-turn-helix (HTH) α2-loop-α3 fold. It is speculated that the RHH motif is a derivative of the HTH domain by converting the N-terminal helix to a β-strand [43].

RHH motifs recognize DNA in a sequence-specific manner. The β-sheet makes specific contacts with DNA bases in the major groove, pointing three amino acid side chains from each strand into the groove. The N terminus of α2 forms non-specific contacts with the DNA phosphate backbone (Fig. 1.4) [42].

Although the overall DNA-binding mode of each RHH dimer is the same, the details of the specific DNA base contacts are not conserved across the RHH superfamily. Each member recognizes a unique operator sequence (or sequences) [42]. It seems plausible as the three DNA-contacting residues from the β-strand vary across the family. However, NikR, CopG and CcdA utilize the same amino-acid side chains to specifically recognize distinct DNA sequences.

RHH transcription factors mediate a diversity of functions through a wide range of regulatory mechanisms. Arc repressor of bacteriophage P22 regulates lytic growth cycle by monomer-dimer equilibrium and cooperative binding of two Arc RHH dimers [44] [45]. CopG from *Streptococcus* species and Omega from *inc 18*

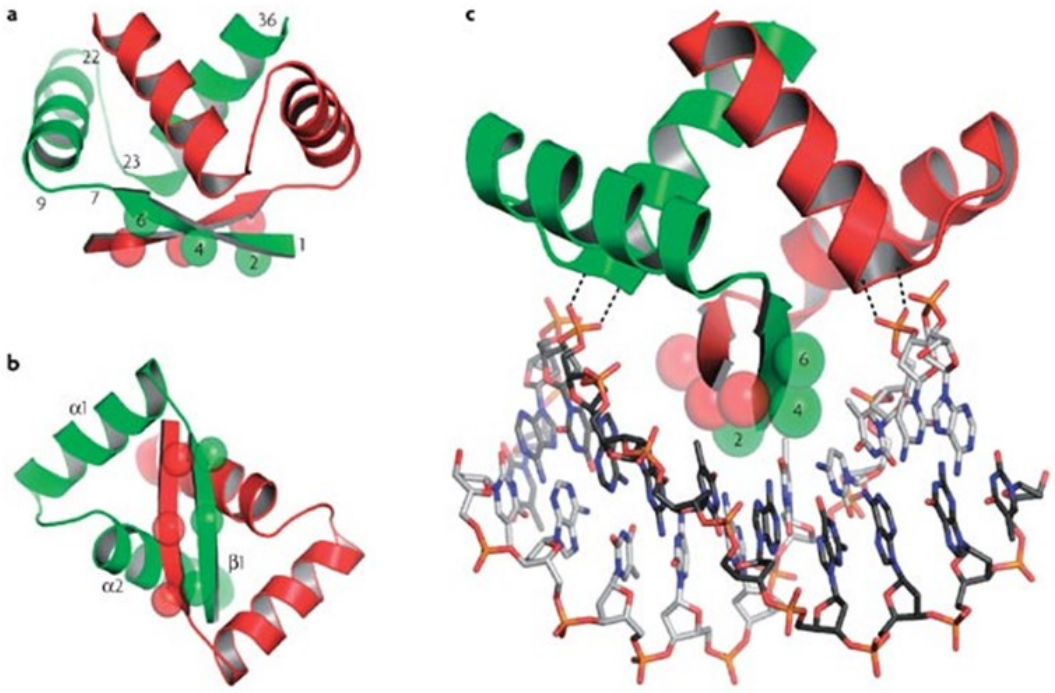

Figure 1. 4 Three views of the RHH dimer (RHH_2) domain

Three views of the RHH dimer (RHH_2) domain are shown as cartoons and coloured by subunit. The three amino-acid positions from each subunit that make sequence-specific nucleotide base contacts are shown as semi-transparent spheres. (a) Reference positions within the RHH motif are numbered on the green subunit. (b) The secondary structure elements are labelled on the green subunit. (c) Shows the interaction of a RHH2 domain with a DNA operator. Nonspecific anchoring contacts between the N terminus of the second α-helix and the DNA phosphate backbone are shown as dashed black lines. Specific base contacts are made by positions 2, 4 and 6 of each subunit from the β-sheet within the DNA major groove [42]. Adapted from Schreiter, Eric R.; Drennan, Catherine L, "Ribbon–helix–helix transcription factors: variations on a theme," *Nature Reviews Microbiology*, vol. 5, pp. 710-720, 2007.

family plasmids in gram-positive bacteria control plasmid copy number in a manner similar to Arc repressor. NikR repressor in *Escherichia coli* regulates nickel concentration via nickel-induced conformational change to repress transcription of a gene encoding a nickel-import protein [46]. The regulation of methionine biosynthesis in *Escherichia coli* is mediated by the MetJ repressor. The binding of the MetJ RHH to DNA is dependent on the binding of the regulatory ligand S-adenosylmethionine [47]. RHH folds are also common in conjugative regulatory proteins, such as TrwA of plasmid R388 [48], TraY of F plasmid [49] and VirC2 of Ti plasmid [50].

1.7 Thesis overview

This thesis investigates the structural basis for the DNA binding specificities of different TraM proteins. We determined the crystal structure of F TraM RHH bound to its *sbmA*, which is very similar to pED208 TraM bound to its cognate DNA, even though the recognized motif sequences vary. The F TraM binding affinities of an extensive set of *sbmA* mutants are tested by Electrophoretic mobility shift assay (EMSA) to study the determinants of TraM specificity. The results demonstrate that binding specificity is achieved not only by recognition of short repeating 5 base pair motifs by the individual RHH domains, but also by the spacing of these motifs within the *sbmA* element. The most important F TraM binding site *sbmC* is also characterized. This work provides a basis to begin to understand the binding properties of the large family of TraM proteins that have been identified through genomic sequencing efforts.

Chapter 2

Materials and methods

Contributions: Dr. Ross Edwards performed structure solution and refinement.

2.1 Bacterial strains, plasmids, and plasmid construction

The following *Escherichia coli* strains were used: DH5α [F$^-$Δ*lacU169* (*Φ80lacZΔM15*) *supE44 hsdR17 recA1 endA1 gyrA96*(Nal^r) *thi-1 relA1*] [51], BL21-DE3 [F$^-$ *ompT hsdS*B (rB$^-$mB$^-$) *gal dcm*] (Invitrogen). Plasmid pRFM200 was described previously [39]. The ~0.2-kb *Nde*I-*Bam*HI fragment of the PCR product amplified from pRFM200 by using primers JLU609 (5'-AAAAATTTTCATATGGCTAAGGTGAACCTGTATATCAGC-3') and JLU606 (5'-TAGGATCCTTATTACATCTGAGCCTCATGTACACG-3') was ligated to the 2.4-kb *Nde*I-*Bam*HI fragment of pT7-7, resulting in plasmid pJLM609.

2.2 Overexpression and purification of the N-terminal domain and full-length F TraM

BL21-DE3 cells containing pJLM609 (for F TraM$^{1-54}$) or pRFM200 (for full-length F TraM), were grown in 1 L of LB broth containing ampicillin at 37°C with vigorous shaking. After 3 hours, isopropyl-β-D-thiogalactopyranoside (IPTG) was added to the final concentration of 0.2 mM. The culture was grown overnight at 20°C before harvesting. The detailed purification procedures were described previously [31]. Purified protein concentration was determined by BCA assays (Pierce) following manufacturer's instructions.

2.3 Oligonucleotide DNA purification and annealing

Synthetic DNA oligonucleotides were purified as previously described [50]. Oligonucleotide DNA solutions were quantified by absorbance at 260 nm. Each double-stranded DNA was generated by annealing of purified complementary DNA oligonucleotides in buffer containing 10 mM Tris, pH 7.5 and 50 mM NaCl.

2.4 Electrophoretic mobility shift assay

EMSAs were performed on synthesized DNA as indicated using purified, full-length F TraM protein. Double-stranded DNAs were radiolabeled by using T4 polynucleotide kinase (New England Biolabs) and [γ-^{32}P]ATP. Binding reactions contained 50 mM Tris, pH 7.5, 1 mM DTT, 10% glycerol, and 1 μg of polydI·dC (Roche). To each binding reaction containing the indicated concentration of TraM or its mutant protein, 0.1 nM of ^{32}P-labeled DNA was added and incubated for 30 min at room temperature. The mixtures were run on a TBE-buffered 15% polyacrylamide (29:1) gel at 100 V for 100 min at 4 °C. Bands were visualized and quantified by phosphor screen and ImageQuant. K_D values were fit using the equation:

$$Fraction\ DNA\ bound = \frac{[TraM]}{K_D * [TraM]}$$

where fraction DNA bound was calculated from the intensity of the band corresponding to the TraM-DNA complex.

2.5 Crystallization and data collection

The purified F TraM N-terminal domain (TraM$^{1-54}$) was concentrated to 3.2 mg/mL in 0.5 M ammonium acetate, 1 mM DTT, and 1 mM spermine, and was further mixed with 0.5 mM *sbmA* DNA at 6:1 ratio (28 bp sbmA sequence: 5'-ATACCGCTAGGGGCGCTGCTAGCGGTGC-3':5'-GCACCGCTAGCAGCGCCCCT AGCGGTAT-3'). 5 μL of the protein-DNA mixture was mixed with 5 μL of reservoir solution consisting of 0.1 M Tris, pH 7.0 and 17.5% PEG1000. Crystals of the TraM$^{1-54}$-*sbmA* complex were obtained by vapor diffusion method in sitting drops at room temperature after one week, which were subsequently soaked in reservoir solution plus 20% glycerol for 20 minutes before flash-frozen in liquid nitrogen. 0.6 μL similar mixture containing TraM$^{1-54}$ and *sbmC* DNA was mixed with 0.6 μL of reservoir

solution consisting of 1 M ammonium phosphate dibasic, 100 mM sodium citrate, pH 5.5 and 200 mM NaCl. Crystals consisting of only TraM$^{1-54}$ were obtained under this condition and flash-frozen directly in liquid nitrogen without cryoprotectant. Diffraction data of TraM$^{1-54}$ and the TraM$^{1-54}$-*sbmA* complex were collected from exposure of single crystals at SIBYLS Beamline of the Advanced Light Source, Lawrence Berkeley National Laboratory to 2.0 Å and 3.11 Å, respectively. Data were processed and scaled with HKL2000 [52].

2.6 Structure solution and refinement

The structure of TraM$^{1-54}$ bound to *sbmA* was solved by molecular replacement in Phaser [53] by using the crystal structure of pED208 TraM bound to its cognate DNA as the initial search model [31]. Manual model building in Coot [54] combined with iterative cycles of refinement in Phenix [55] yielded a final model at 3.11 Å resolution.

The structure of TraM$^{1-54}$ was solved by molecular replacement in Phaser [53] using the TraM$^{1-54}$ dimer from the the TraM$^{1-54}$-*sbmA* complex structure as the initial search model. ARP/wARP [56] combined with manual model building and iterative cycles of refinement in Phenix yielded a final model at 2.0 Å. The high-resolution N-terminal TraM structure was further used to refine the α1- α2 loop region of TraM$^{1-54}$ in the TraM$^{1-54}$ -*sbmA* complex. The molecular structure figures were prepared by using PyMOL (Schrodinger, http://www.pymol.org).

Chapter 3

Results and Discussion

Although the structure of pED208 TraM bound to its cognate *sbmA* site provides a satisfactory explanation for how pED208 TraM binds its *sbmA*, the puzzle persists as the very similar F TraM binds to an apparently unrelated *sbmA*. F TraM and pED208 TraM are 38% identical at the amino acid sequence level. In particular, the three residues (Lys3-Gln5-Tyr7) pED208 TraM utilizes to directly contact DNA are almost identical to the corresponding residues of F TraM (Lys3-Asn5-Tyr7). However, the two proteins bind distinct *sbmA* elements. The specificity provides the basis for exclusive transfer of their own plasmids rather than other plasmids in the family. To understand how a single conservative amino acid substitution leads to completely different DNA binding specificity, we determined the structure of F TraM RHH bound to its *sbmA*. The functionally more important F TraM binding site *sbmC*, which bears little sequence similarity to F *sbmA*, was also characterized.

3.1 Crystallization and structure determination of F TraM RHH domain bound to *sbmA*

The RHH DNA binding domains of F and pED208 TraM are very similar (44% overall sequence identity, Fig. 3.1a). In particular, key residues at positions 3, 5 and 7, which extend from the RHH β-strand and contact the DNA bases are very similar, with only a substitution at residue 5 (Asn in F; Gln in pED208). Moreover, the SxSx motif that caps the N-terminus of α2 and contacts the DNA backbone is also conserved between the two proteins. Both of these motifs were found to be highly mutated in a screen for F TraM mutants deficient in autoregulation and conjugative transfer [57] [39]. In spite of the marked similarity of these two proteins, they bind to poorly related DNA target sequences (Fig. 3.1b). The highest affinity binding sites identified for both proteins, *sbmA*, each are 24-26 base pairs in length with obvious inverted repeat symmetry, however, the striking repeating GANTC

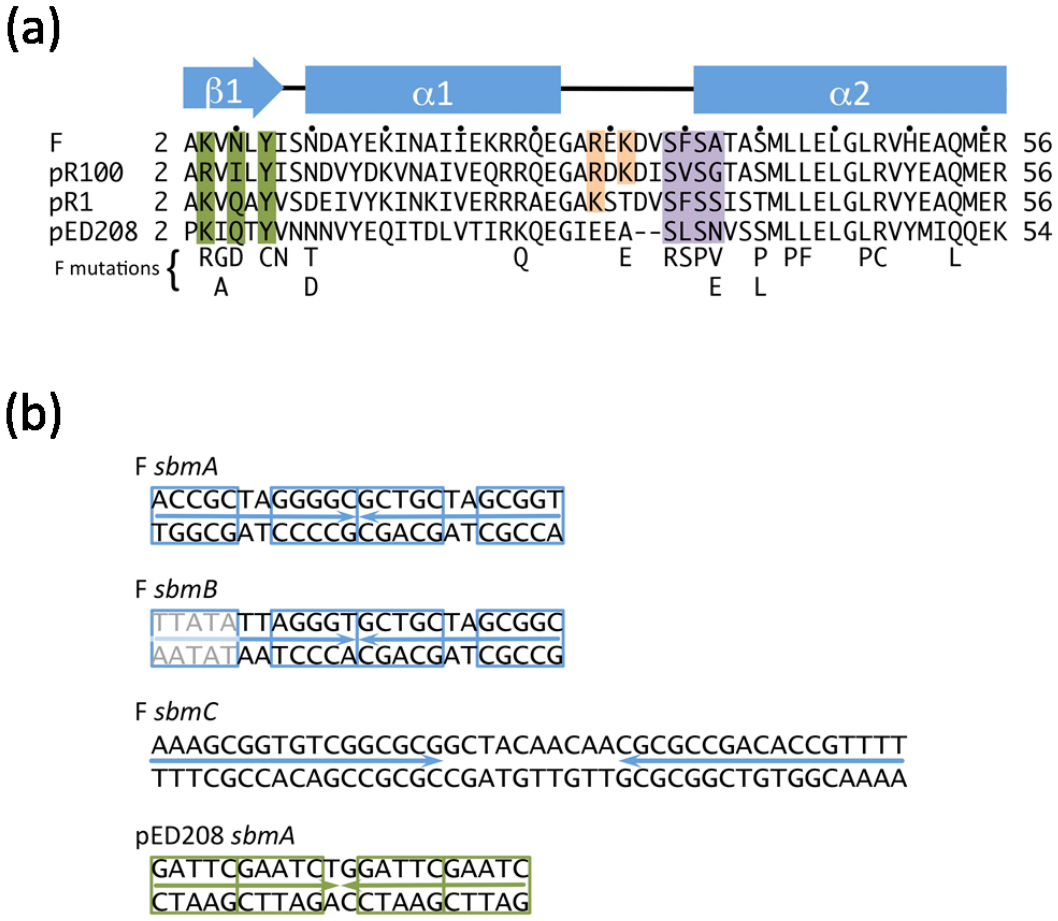

Figure 3. 1 Comparison of TraM RHH domains and DNA binding sites

(a) Sequence alignment of the RHH domains of TraM proteins from four plasmids whose DNA binding properties have been studied experimentally. Residues highlighted in green are predicted to protrude from the N-terminal β-ribbon and mediate sequence specific interactions with the DNA. Residues highlighted in purple cap the N-terminus of α2 and contact the DNA phosphate backbone. Residues highlighted in orange may make long-range electrostatic interactions with the DNA backbone. Underneath the alignment are mutations that impact F TraM conjugation activity [57] [39]. (b) DNA binding sites for F and pED208 TraM proteins. Inverted repeat symmetry in each binding site is indicated by arrows, 5 base pair motifs are boxed. The non-consensus motif in F *sbmB* is faded.

motifs that are recognized by individual RHH dimers in the pED208 *sbmA* are not present in F *sbmA*. Moreover, while F *sbmA* and *sbmB* are clearly related at the sequence level, there is little detectable similarity between these sites and F *sbmC*.

In order to gain structural insight into how F TraM binds its cognate DNA, we set out to determine the crystal structure of a complex of F TraM bound to its *sbmA* site. We were unable to crystallize full length F TraM bound to *sbmA*, likely due to the flexibility of the region linking the N-terminal RHH region to the C-terminal tetramerization domain. We instead crystallized the isolated RHH domain ($TraM^{1-54}$) bound to *sbmA*. We previously showed that this domain binds in a highly cooperative manner to *sbmA*, albeit at reduced affinity compared to intact TraM [31]. The structure of this complex was determined to 3.11 Å resolution by molecular replacement using the pED208 TraM-*sbmA* complex as a search model (Fig. 3.2). We also determined the structure of the free form of the F RHH domain to 2.0 Å resolution. This high resolution model was particularly useful to define the structure of the α1- α2 loop, whose density was difficult to interpret in the lower resolution DNA complex. The statistics are shown in Table 3.1. Comparison of the apo- and DNA-bound forms of the F TraM RHH domain reveals a very similar structure (~0.9 Å RMSD over all Cα atoms) with most of the differences in the α1- α2 loop.

3.2 Overall structure of F TraM RHH bound to *sbmA*

The crystal structure of F TraM RHH bound to *sbmA* is in many ways similar to the overall structure of the pED208 ortholog. In both cases four RHH dimers bind to 5 base pair sequence motifs arranged along the *sbmA* DNA such that motifs 1 and 3 bind RHH dimers on one side of the DNA, while motifs 2 and 4 each bind RHH dimers on the opposite side of the DNA (Fig. 3.3 a). As in pED208, F motifs 1 and 3 and motifs 2 and 4 are separated by 12 base pairs and the *sbmA* DNA adopts

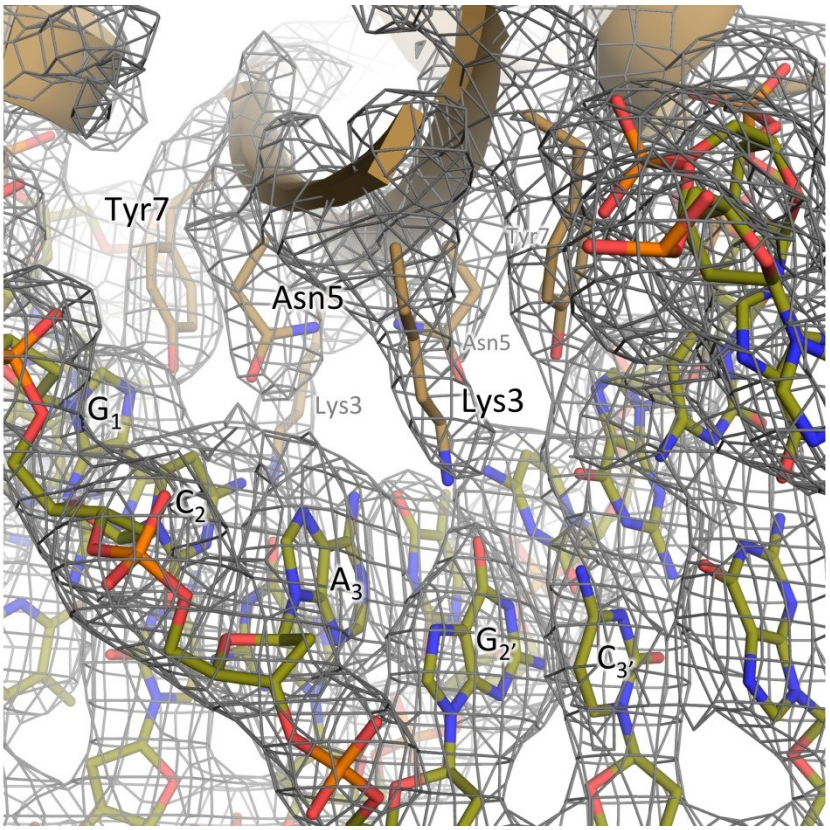

Figure 3. 2 Sample electron density of the F TraM RHH – *sbmA*

Electron density is shown for the major groove of the DNA showing the nucleotide bases of one $G_1C_2A_3G_{2'}C_{1'}$ motif as well as the RHH β-ribbon and side chains that contact the DNA motif base pairs. 2Fo-Fc density at 3.11 Å resolution is contoured at 1.7 σ cutoff.

Table 1: Crystallographic Statistics and Refinement

	TraM$^{1-54}$ *sbmA*	**TraM$^{1-54}$**
Wavelength (Å)	1.11587	1.07808
Resolution range (Å)	46.96 - 3.11 (3.22 - 3.11)	36.95 - 2.00 (2.07 - 2.00)
Space group	P 43 21 2	R 3 :H
Unit cell	87.1 87.1 217.8 90 90 90	115.7 115.7 54.7 90 90 120
Total reflections	241579	100949
Unique reflections	15828 (1523)	18421 (1843)
Multiplicity	15.2	5.5
Completeness (%)	99.9 (98.8)	99.8 (99.8)
Mean I/ (I)	33.0 (3.2)	17.7 (2.7)
Wilson B-factor	88.1	35.6
Linear R-factor[1]	0.087 (0.788)	0.073 (0.740)
Square R-factor[2]	0.045 (0.562)	0.053 (0.655)
R-work	0.2237 (0.2617)	0.1922 (0.2608)
R-free	0.2316 (0.2922)	0.2304 (0.2775)
Number of non-hydrogen atoms	4280	1643
macromolecules	4280	1598
ligands		20
water	0	25
Protein residues	465	209
RMS(bonds)	0.003	0.011
RMS(angles)	0.59	1.08
Ramachandran favored (%)	90	97
Ramachandran outliers (%)	0	0
Clashscore	5.00	6.88
Average B-factor	120.6	51.1
macromolecules	120.6	51.0
ligands		63.0
solvent		49.2

[1]Linear R-factor = $\Sigma|I - \langle I \rangle| / \Sigma(I)$
[2]Square R-factor = $\Sigma|I - \langle I \rangle|^2 / \Sigma(I^2)$

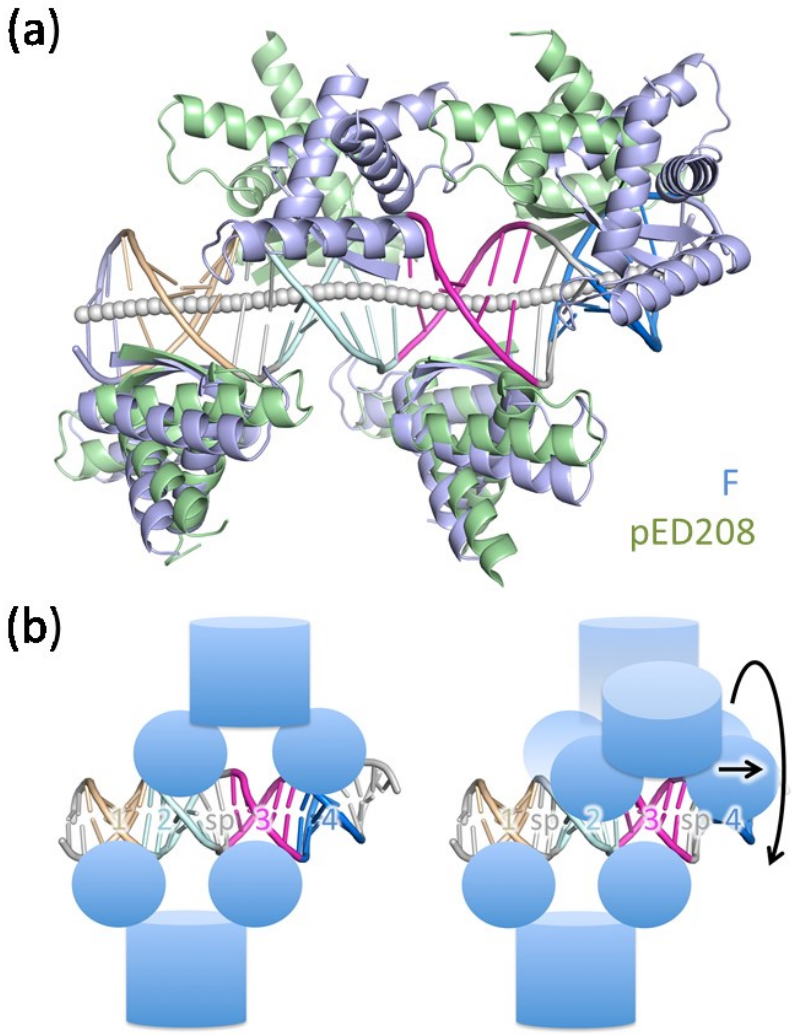

Figure 3. 3 Overall structure of F TraM RHH domains bound to *sbmA*, and comparison with pED208 TraM

(a) Alignment of the structures of the F TraM RHH domain (mauve) bound to its *sbmA*, with pED208 TraM RHH domain (green) bound to its *sbmA*. The DNA shown is the F *sbmA* DNA, the pED208 DNA is not shown. Two RHH domains from the same tetramer were aligned (bottom of the figure), to highlight the shift of the RHH domains from the second tetramer on the opposite side of the DNA. The DNA helix axis as calculated using Curves+ (http://bisi.ibcp.fr/tools/curves_plus/) is illustrated as a series of spheres. (b) Cartoon illustrating the two base pair shift in the positioning of pED208 TraM on its *sbmA* (left) compared with F TraM (right). Note that in these figures, the 4 5 base pair motifs are color coded: motif1 – beige; motif2 – cyan; motif3 – magenta; motif4 – blue. Spacer elements are colored grey.

an underwound, kinked conformation (31.7^{o} average twist/bp) to position the pairs of motifs on the same side of the DNA. The fact that this unwinding is accomplished by binding of the F RHH domains alone indicates that DNA unwinding does not require that pairs of RHH dimers be linked in a tetramer. As in the pED208 case, cooperative recognition of the F *sbmA* element by a pair of F TraM tetramers is likely facilitated by binding-induced unwinding and distortion of the DNA.

In addition to the 5 base pair motifs contacted by the RHH dimers, both *sbmA* elements contain 2 base pair spacer elements that maintain the 12 base pair spacing between 5 base pair motifs bound by the same tetramer. Strikingly, the positioning of these spacers is different in the F and pED208 systems. In pED208 *sbmA*, there is a single 2 base pair spacer at the centre of the site. In F there are two spacers, one between motifs 1 and 2 and a second between motifs 3 and 4 (Fig. 3.1b). This difference results in a shift of the relative positions of the two TraM tetramers on the DNA between the F and pED208 systems. This shift can be visualized in an alignment where the RHH domains of one tetramer from the pED208 complex are superimposed on one pair of RHH domains in the F structure (Fig. 3.3a). In this alignment the second pair of F RHH domains are translated and rotated 2 base pairs along the DNA compared to the pED208 RHH domains (Fig. 3.3b).

3.3 F TraM RHH domains recognize a PuCNGPy motif

The pED208 TraM RHH binds a $G_1A_2N_3T_{2'}C_{1'}$ motif that is precisely conserved in its *sbmA*, while each F RHH binds a less well conserved $Pu_1C_2N_3G_{2'}Py_{1'}$ motif (where Pu indicates either G or A, and Py indicates either C or T; Fig. 3.1b). In both F and pED208 cases, the motif is symmetrical about the central, nonconserved base pair, and we have numbered the bases within the motif

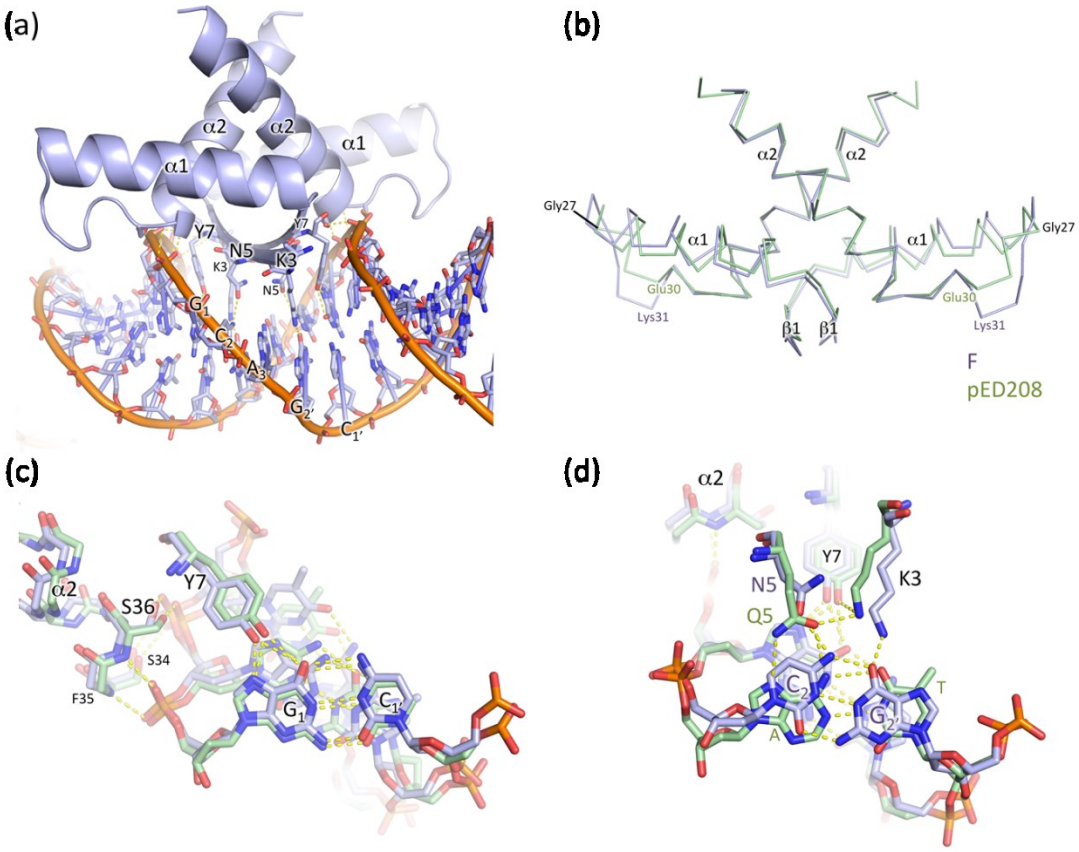

Figure 3. 4 Structure of the TraM RHH domain and interactions with DNA

(a) Overview of the structure of a single F TraM RHH domain bound to DNA. (b) Alignment of the F and pED208 RHH domains illustrating the overall strong similarity in backbone structure, and the difference in the α1-α2 loop.

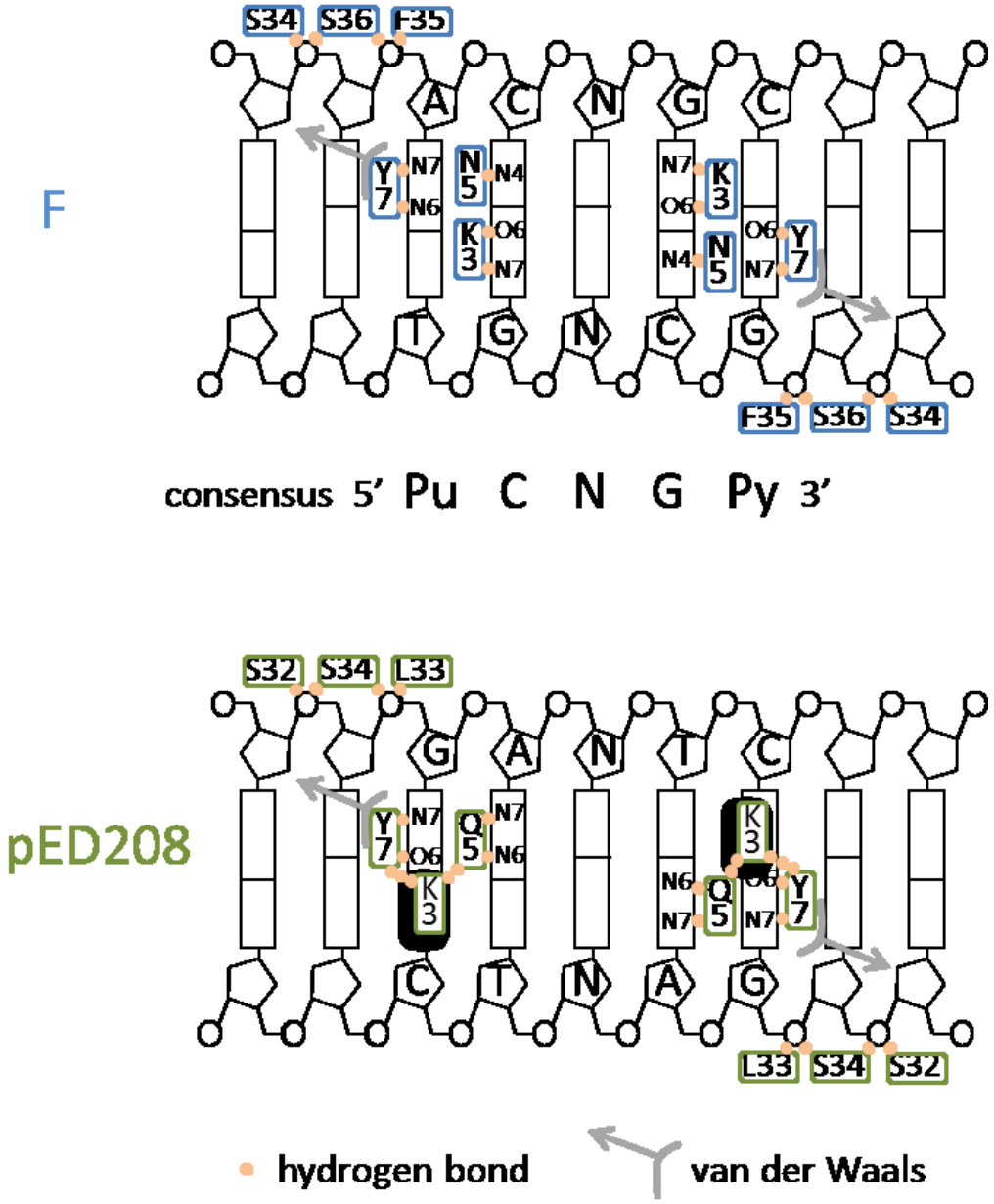

Figure 3. 5 Recognition of the 5 base pair motifs by the TraM RHH domains

The schematic diagram illustrates the major hydrogen bonding and van der Waals interactions observed between amino acids and regions of the 5 base pair motifs for both the F (top) and pED208 (bottom) systems. The consensus sequence of the 5 base pair motif in the F system is shown.

to reflect this symmetry. Comparison of the F and pED208 structures suggests a structural basis for these differences (Fig. 3.4, 3.5). In pED208, the primary β-strand DNA contacting residues form a hydrogen-bonded network in which the conserved Lys3 does not directly interact with DNA but instead hydrogen bonds with Gln5 and Tyr7 to orient these residues for recognition of the GANTC motif. Gln5 recognizes the A at position 2/2' via a pair of hydrogen bonds, while Tyr7 recognizes the G at position 1/1'. In the F system, the loss of the methylene group at residue 5 (Gln – Asn) means that the shorter Asn can no longer hydrogen bond with Lys3. The base at position 2 is pulled closer to the β-strand allowing hydrogen bonding with Asn and as a result, Lys3 now also contacts the same base pair. The end result is that both Lys3 and Asn5 work together to recognize the conserved G-C pair at position 2/2' in the motif (Fig. 3.4d). Tyr7 recognizes the Pu position of the F motif through hydrogen bonding interactions with the major groove face of the purine base (Fig. 3.4c). The fact that either a G-C or A-T pair seems to be recognized at this position may be a result of the fact that in F, Tyr7 can act as either a hydrogen bond donor (to recognize O6 of G) or as a hydrogen bond acceptor (to recognize N6 of A). In contrast, the Tyr7 interaction with Lys3 in pED208 TraM ensures that Tyr7 can only act as a hydrogen bond donor to the DNA, thereby explaining the conserved G-C pair at the ends of the **G**ANT**C** motif.

3.4 Design of a set of *sbmA* mutants to test TraM binding specificity

We utilized a series of *sbmA* mutants to evaluate the relative contributions of the different nucleotides in each of the 5 base pair motifs, as well as the spacer elements, to TraM binding affinity, assessed by an electrophoretic mobility shift assay (EMSA, see Fig. 3.6 and 3.7). In these experiments we used a symmetrized

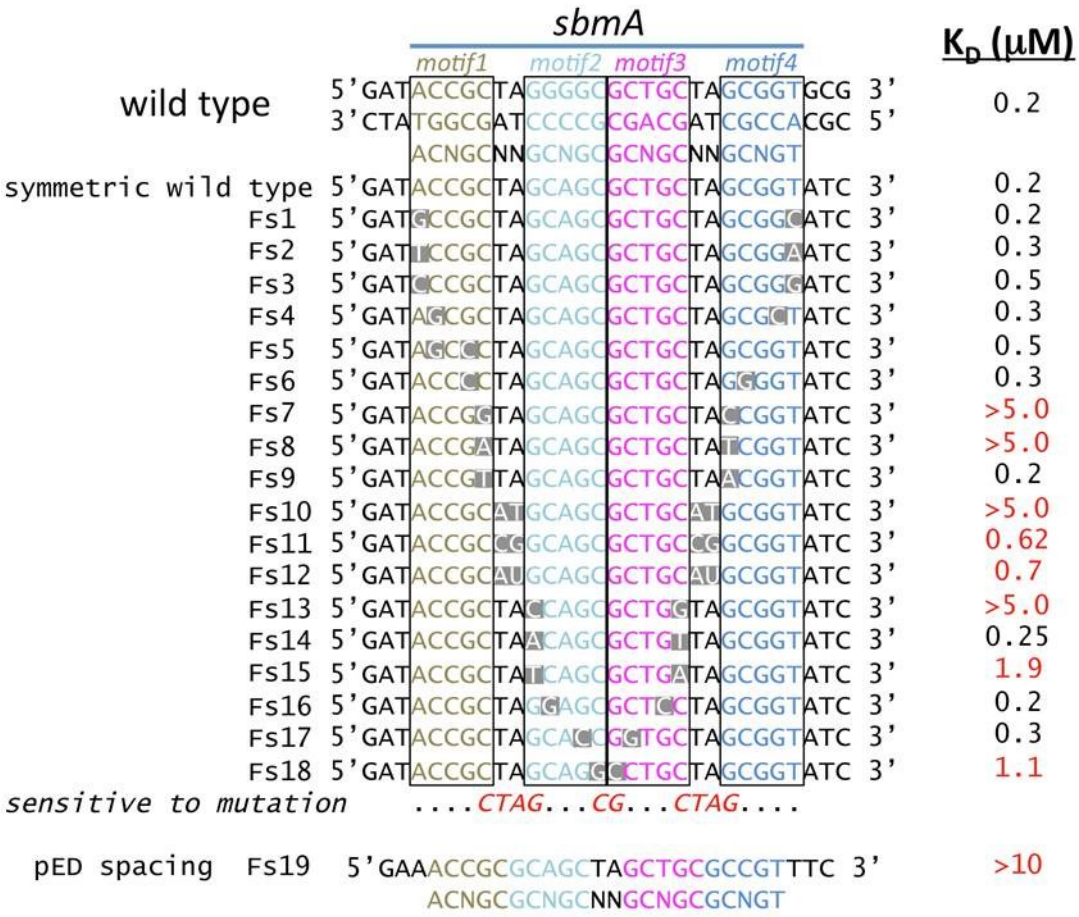

Figure 3. 6 Probing the determinants of F TraM DNA binding specificity

(a) F *sbmA* mutants used in the determination of F TraM DNA binding specificity. Top is shown the wild type F *sbmA* sequence with individual 5 base pair motifs color-coded. Below is shown the sequences (top strands only) of the mutant series, based on a symmetric version of the F TraM. On the right of the panel are the K_D values calculated from EMSA with the values differing by >3-fold from wild type colored red.

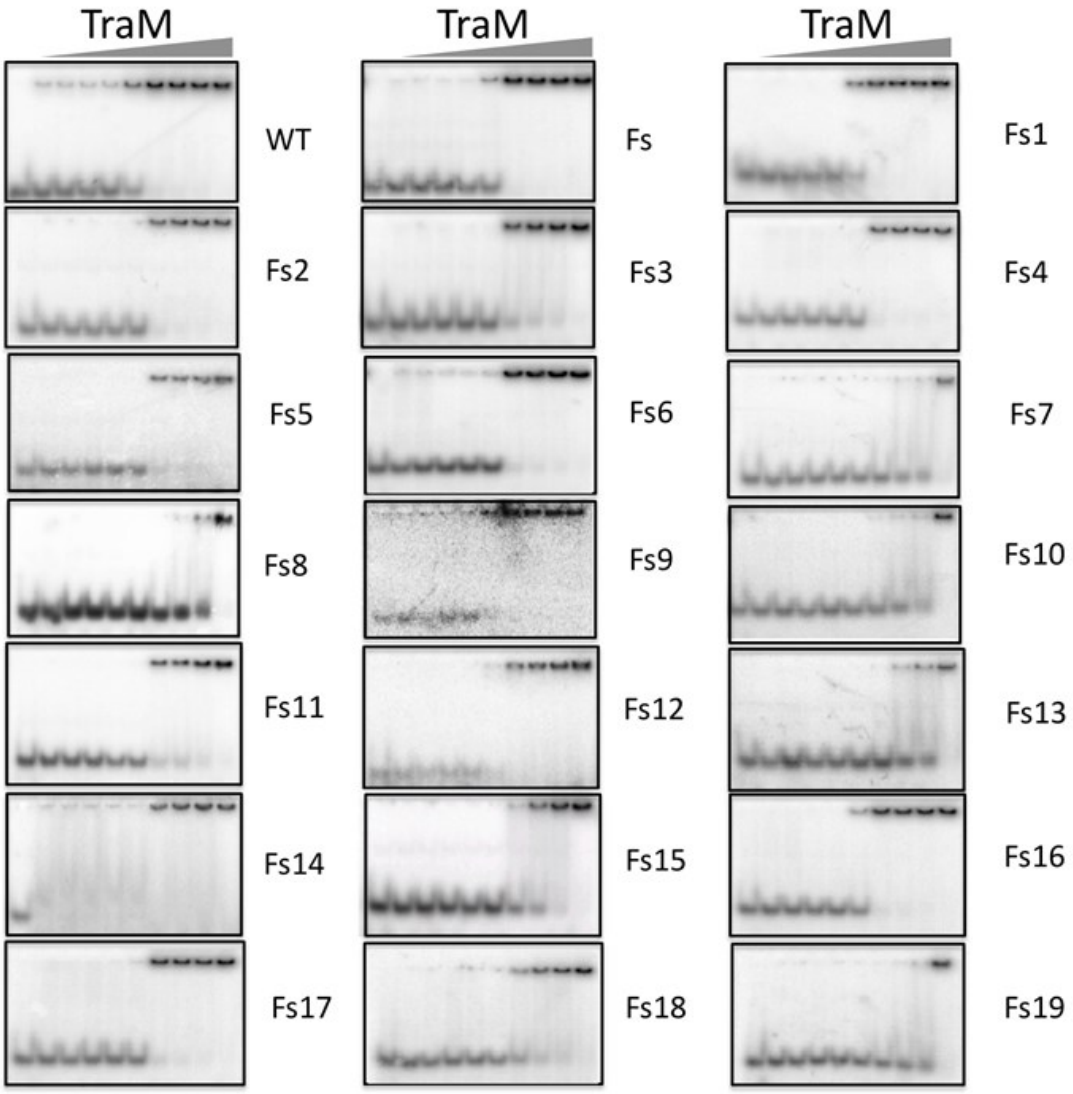

Figure 3. 7 Assessment of the binding of F TraM to a series of *sbmA* variant DNAs by electrophoretic mobility shift assay (EMSA)

F TraM was titrated against the indicated DNA (see Fig. 3.6 for DNA sequences). 0.1 nM DNA was used in each reaction. The concentrations of TraM protein in each lane are 0, 1, 3.5, 10, 35, 100, 350, 1000, 3500, 10000 nM.

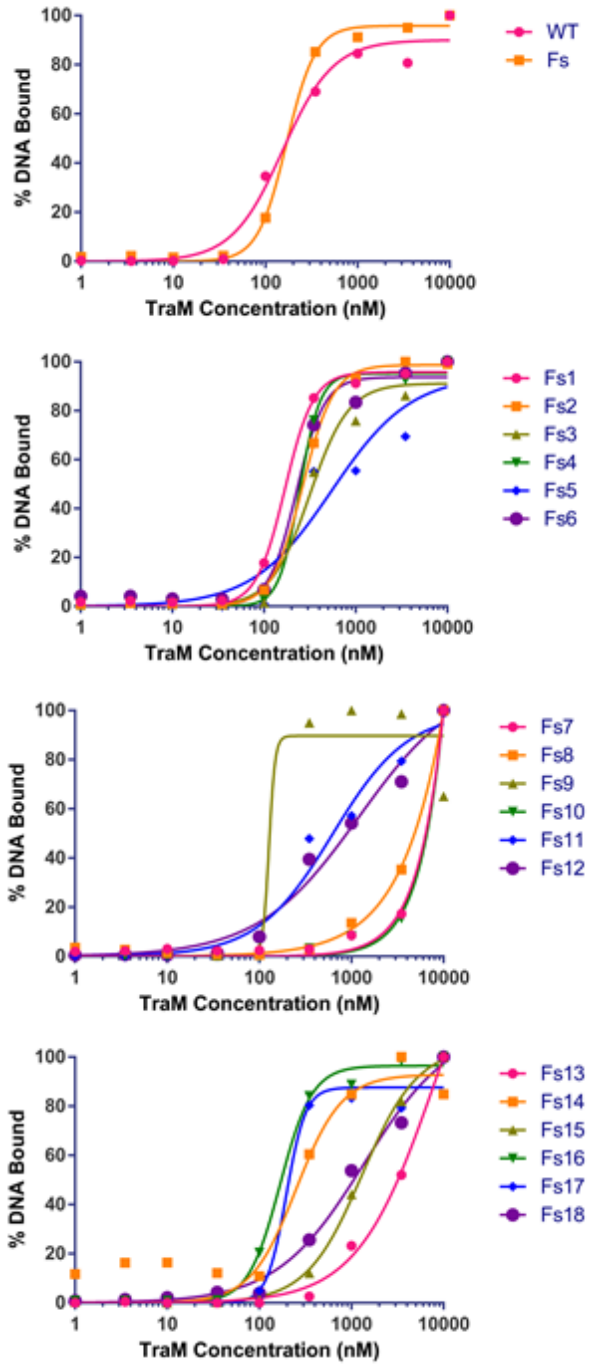

Figure 3. 8. Quantitation of EMSA data of F TraM binding to a series of *sbmA* variant DNAs

Using the EMSA data shown in Fig. 3.7, the % DNA bound at each point was assessed and the apparent K_D for each variant was determined as described in Materials and Methods.

sbmA in which motif 2 was mutated from 5'-GGGGC-3' to 5'-GCAGC-3' to conform to the consensus (Fig. 3.6). Symmetric mutations were made throughout the *sbmA* element at all positions except for the nonconserved base pair at the central 3 position of each motif which is not directly contacted by TraM. Both the wild type *sbmA* and the symmetric *sbmA* were bound by F TraM with similar affinities of 0.2 μM. The results for the mutation of each of the conserved base pairs in the motif, as well as for the mutation of the spacer elements are described in detail below.

3.4.1 Specific recognition of base pair 1

Base pair 1 in the F *sbmA* motif can either be G-C or A-T. The A-T pair is only found at the terminal bases of *sbmA*; all the other base pairs at position 1 are G-C. Mutations that replace the terminal A-T base pair with G-C do not impact binding affinity (mutant Fs1), nor does replacement of any of the internal G-C pairs at position 1 with A-T (mutants Fs9 and Fs14), indicating that Tyr7 is equally able to recognize either purine base at position 1 (Fig. 3.4c, 3.6). However, mutations that flip the position 1 base pair such that a pyrimidine base faces Tyr7 have a detrimental impact on binding. These effects are much more pronounced at the internal sites (mutants Fs7, Fs8, Fs13, Fs15, and Fs18) while similar mutations at the terminal base pairs have a reduced impact on binding (mutants Fs2 and Fs3), suggesting that recognition of the terminal base pair is more relaxed than the internal sites.

3.4.2 Specific recognition of base pair 2

Base pair 2 in *sbmA* is a C-G pair except in the wild type motif 2 where it is either a C-G or a G-C pair. The fact that the symmetric *sbmA*, in which motif 2 conforms to the consensus, exhibits the same binding affinity as the asymmetric wild type sequence, suggests that the F TraM can adjust to recognize changes at

this base pair (Fig. 3.6). We systematically substituted each C-G pair for the flipped G-C pair throughout the entire *sbmA* sequence (mutants Fs4, Fs6, Fs16, and Fs17). None of these mutations significantly impact binding affinity (<2-fold reduction in binding affinity compared to wild type). Base pair 2 is recognized by Asn5 and Lys3, and in the case of the recognition of the consensus C-G pair, Asn5 hydrogen bonds with the cytosine base, while Lys3 hydrogen bonds with the guanine (Fig. 3.4d). In motif 2, where a G-C pair is found at position 2, Asn5 hydrogen bonds with guanine N7/O6, while Lys3 approaches cytosine. The ability of Asn5 to hydrogen bond with either a cytosine or a guanine likely is due to the ability of the side chain to rotate to either present its carbonyl or its amine group to make appropriate, complimentary interactions with its base partner. Lys3, on the other hand, would only be expected to make favorable interactions with guanine. To further test the importance of base pair 2, we simultaneously flipped both base pair 2 and 2' within motif 1. This mutant (Fs5) only displayed a small (~2.5-fold) reduction in binding affinity compared to the wild type, further underlining the flexibility of recognition at this position.

3.4.3 Specific recognition of spacer elements

F *sbmA* contains two conserved TA spacer elements, which is also conserved in the left half site of F *sbmB*. In light of the flexible recognition of the 5 base pair motif elements, we asked whether the conserved spacers also contribute to TraM binding specificity. Mutation of both TA spacers to AT dramatically reduces binding (mutant Fs10), indicating that the spacers contribute directly to DNA binding specificity. The spacer base pairs do not directly hydrogen bond with residues from the RHH β-strand, however the base immediately 5' to base pair 1 is in van der Waals contact with Tyr7, and Tyr7 also contacts the deoxyribose of the nucleotide of the next 5' residue. In addition, the phosphate groups of these two residues are

each recognized by pairs of hydrogen bonds from the N-terminus of α2 (Fig. 3.4c, 3.4d and 3.5). The van der Waals contact with the adenine base immediately 5' to the purine of base pair 1 suggested that a purine might be favored in this position over a larger pyrimidine, which would be predicted to clash with Tyr7. In agreement with this, substitution of the TA spacer with a CG spacer (mutant Fs11) resulted in a smaller loss in binding affinity compared to the AT spacer substitution. Furthermore, replacement of the TA spacer with an AU spacer also resulted in a reduced loss in binding affinity compared with the AT substitution, consistent with the idea that the predicted van der Waals clash between the thymine 5-methyl in Fs11 and Tyr7 is responsible for much of the loss in binding affinity of this mutant.

The finding that the spacer elements play an important part in sequence-specific DNA binding led us to ask if the position of the spacer is also important. We constructed a version of F *sbmA* that, like pED208 *sbmA*, only contains a single 2 base pair spacer at the centre of the site (Fs19). The spacer was chosen to be TA to match the F TraM preference. Surprisingly, in spite of the match of the spacer to the consensus, binding of this DNA was barely detectable under our experimental conditions, suggesting that not only the sequence but the positioning of two spacers, one between motifs 1 and 2, and the second between motifs 3 and 4, is critical for F TraM binding (Fig. 3.6, 3.7).

To further probe the role of the β-strand residues in DNA binding specificity, we constructed a mutant form of pED208 TraM containing a glutamine-asparagine mutation at residue 5. We predicted that this mutant might bind the same sequence motif as F TraM, however, due to the overall structure of the pED208 RHH, it might preferentially bind a DNA containing these motifs spaced as in the pED208 sbmA (Fs19). EMSA experiments with this protein however showed that it does not bind

Fs19, nor is it able to bind either the F or pED208 sbmA DNAs. This could imply that the whole RHH domain acts as a unit to recognize DNA and that a simple single amino acid change cannot swap binding specificity.

3.5 A pair of TraM tetramers specifically binds *sbmC*

Previous EMSA and DNA footprinting studies indicated that TraM could bind *sbmC*, albeit at reduced affinity compared to *sbmA* [30]. These studies were however complicated by the fact that at the protein concentrations employed, there was significant non-specific binding of TraM to neighboring regions of the DNA. To clarify the nature and specificity of binding of F TraM to its *sbmC*, we utilized EMSA to characterize F TraM binding to a 50 base pair *sbmC*-containing DNA (Fig. 3.9). This DNA contains an obvious inverted repeat but does not bear significant sequence similarity to *sbmA*. The sequence only contains one pair of 5 base pair motifs that match the PuCNGPy consensus (Fig. 3.9a). EMSA revealed that this 50 base pair DNA binds F TraM with a similar affinity to *sbmA* (Fig. 3.9b). To further define the binding site, we created a series of truncated *sbmC* DNAs, with symmetric deletions from both ends, and tested the binding affinities of these DNAs for TraM by EMSA (Fig. 3.9a). Deletions down to 26 base pairs did not impact binding affinity, however removal of a single additional base pair from each end essentially abrogated binding (Fig. 3.9b). A large region of the inverted repeat is deleted in the 26 base pair minimal *sbmC*, including the PuCNGPy motifs. We next used a quantitative EMSA at high protein concentrations above the K_D to determine the binding stoichiometry. This result suggests that two F TraM tetramers (eight TraM protomers) bind *sbmC*. The fact that only a single shifted species is observed suggests that, like *sbmA*, a pair of TraM tetramers cooperatively bind *sbmC* (Fig. 3.9c).

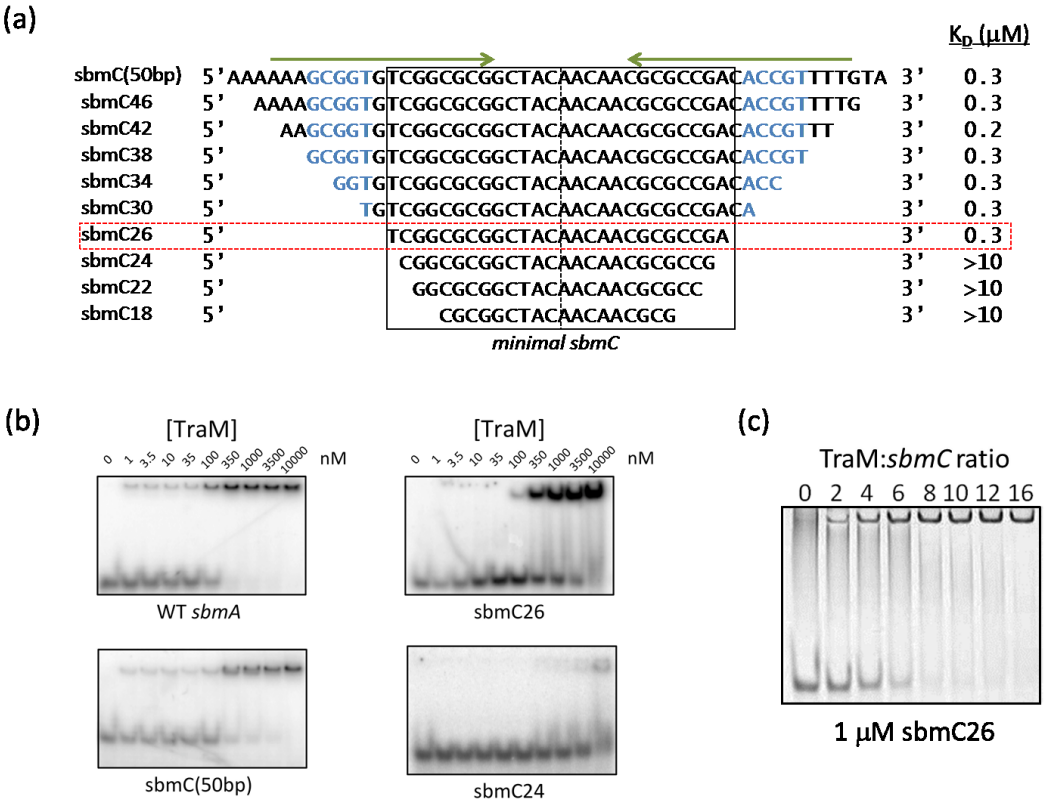

Figure 3. 9 Definition of a minimal *sbmC* that binds TraM

(a) Deletion series of DNAs starting at a 50 base pair *sbmC* (sbmC(50bp)) which was previously shown to bind F TraM and contains inverted repeat symmetry (arrows). Each of the DNAs were tested for TraM binding by EMSA and the K_Ds for each are shown in the right column. The minimal *sbmC* DNA (sbmC26) is boxed. Sequences corresponding to the PuCNGPy motif that could represent possible 5 base pair motifs for TraM binding are in blue. (b) EMSA analysis for the minimal *sbmC*, sbmC26, and the slightly shorter sbmC24. The concentrations of TraM protein in each lane are 0, 1, 3.5, 10, 35, 100, 350, 1000, 3500, 10000 nM. (c) EMSA to define the stoichiometry of F TraM:*sbmC* interactions. The titration was performed with sbmC26 at 1 μM and the ratio of F TraM protomer:*sbmC* is shown.

3.6 Analysis of the TraM protein family and DNA binding properties

To date, several hundred TraM proteins have been identified. Starting with a sequence alignment containing 221 distinct TraM orthologs, we identified 31 different classes of these proteins that significantly differ in amino acid sequence within their RHH domains (Fig. 3.10). These proteins can be further grouped into 11 classes based on the identity of the critical residues that are predicted to protrude from the β-ribbon at the 3, 5 and 7 positions to define DNA binding specificity (Fig. 3.11a). This analysis reveals a striking conservation of the DNA contacting residues. The most conserved is position 7 that is either a tyrosine or a phenylalanine. In both F and pED208 TraM this residue is a tyrosine, which recognizes the terminal base pairs of the motif and also packs against the two base pair DNA spacer element. The fact that certain TraM orthologs contain a phenylalanine at this position would indicate that the hydrogen bonding capacity of the position 7 residue is dispensable, however, the van der Waals packing against the neighboring spacer DNA may be more critical. The position 3 residue is also highly conserved, and can be either a lysine or an arginine residue. In both the F and pED208 orthologs it is a lysine, however in these proteins it plays different roles that help to define their distinct DNA binding specificities. In F, Lys3 directly contacts base pair 2 of the motif, however in pED208, it does not contact DNA but instead helps organize the other DNA contacting residues of the β-ribbon. Arginine at position 3 potentially could play a similar role to lysine; alternatively, arginine often recognizes guanine bases or pyrimidine-guanine dinucleotide steps [58] [59] [60]. Position 5 is the most variable. In F an asparagine occupies this position and directly hydrogen bonds with base pair 2, helping to define the preference for a GC at this position. In pED208 TraM a

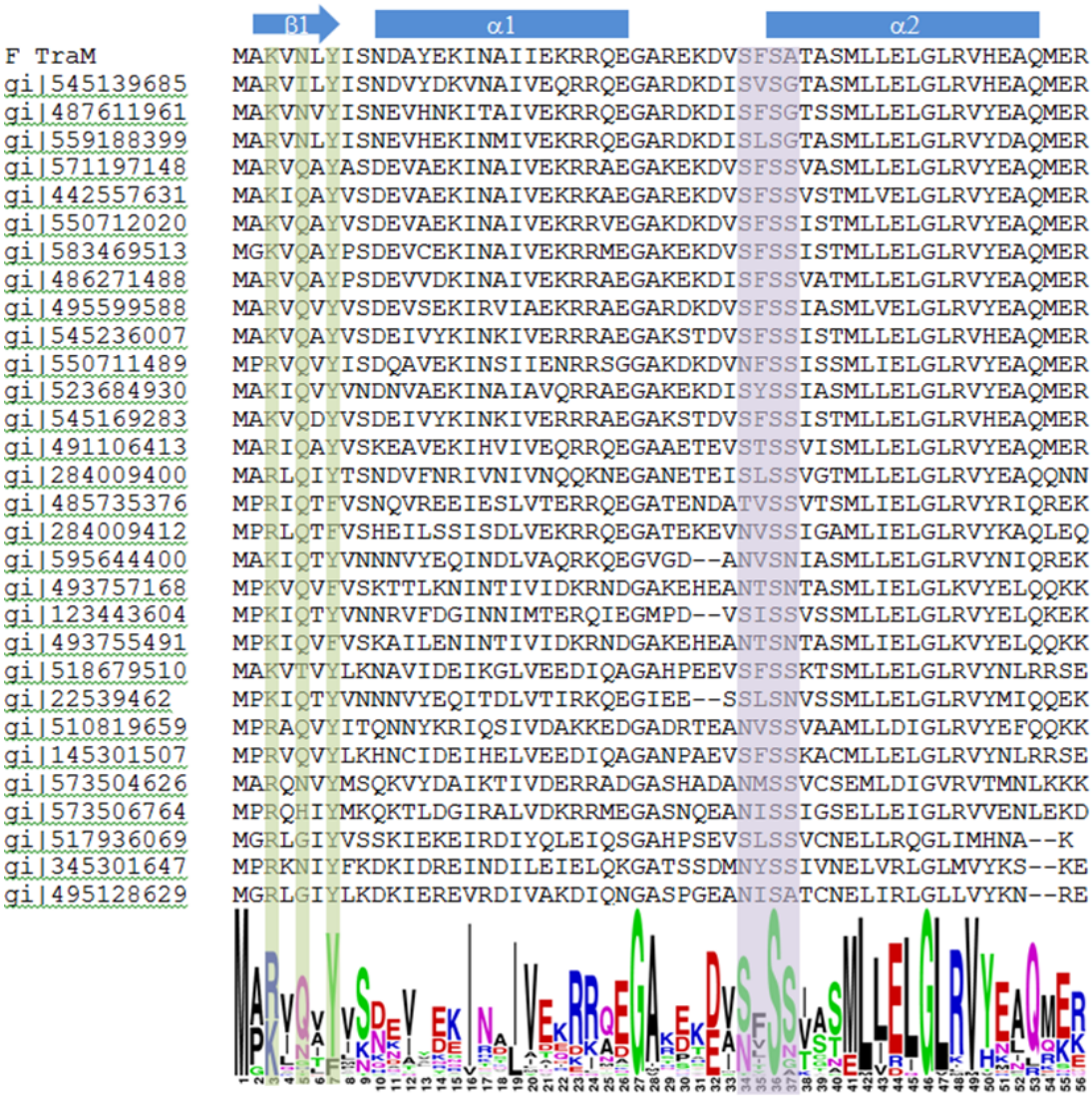

Figure 3. 10 Sequence alignment of RHH domains of 31 TraM orthologs

F TraM was used to seed a BLAST search that uncovered 221 TraM orthologs from the non-redundant protein database with >26% sequence identity compared to F TraM. This list was culled to remove highly similar proteins, resulting in a list of 31 proteins. This alignment was used to generate a sequence logo that illustrates sequence conservation for the TraM RHH domain using the Weblogo server (http://weblogo.berkeley.edu/logo.cgi) which is displayed below the alignment. Highlighted in green are β -ribbon residues that directly contact DNA via the major groove. Highlighted in purple is a cluster of residues that caps α2 and contacts the DNA phosphate backbone. Note the two residue gaps in the α1- α2 loops of the pED208-like TraM proteins (gi|595644400, gi|123443604, and gi|22539462).

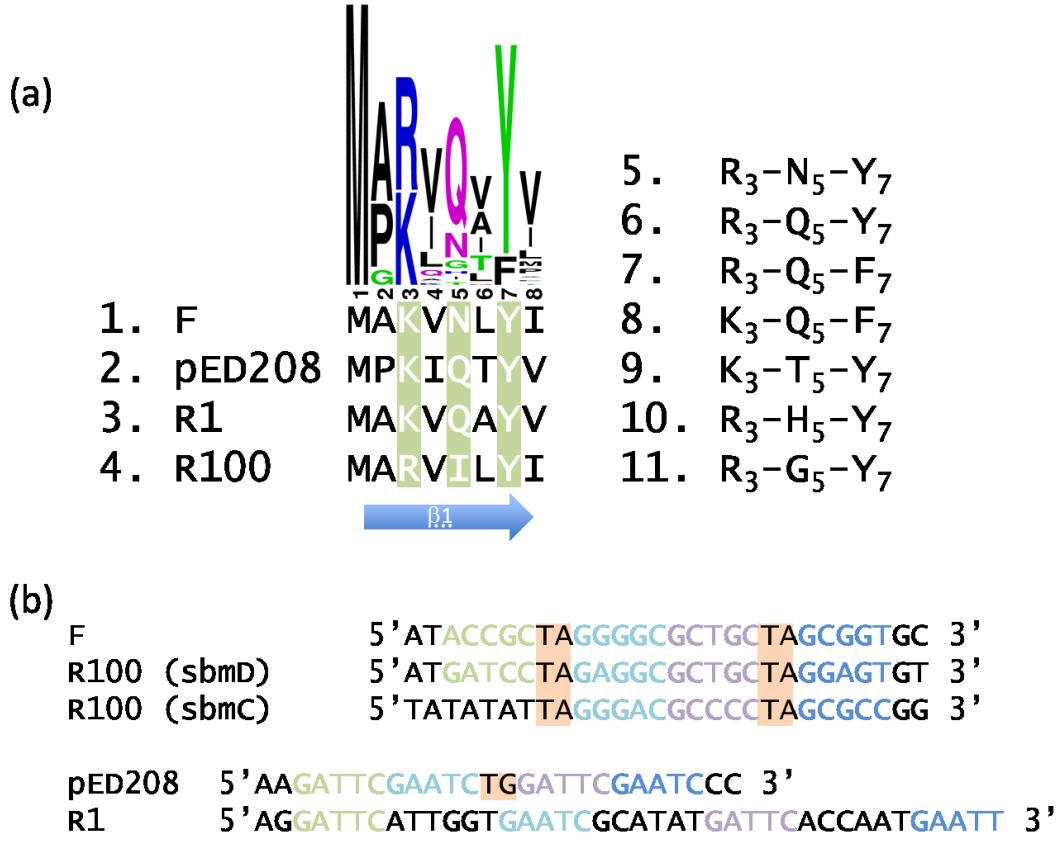

Figure 3. 11 Analysis of the diversity of TraM proteins and their interactions with DNA

(a) The 11 families of TraM proteins clustered according to their β-ribbon sequences. The families were derived from an alignment of 221 TraM orthologs, based on the identity residues 3, 5 and 7. Experimental DNA binding data exists for families 1-4 (represented by F, pED208, R1, and R100) and full β-ribbon sequences are shown for these proteins. For the other 7 families, only the residues 3, 5 and 7 are shown. A logo representing the diversity of sequences derived from these families is displayed. (b) Predicted RHH recognition motifs in experimentally determined R100 (top) and R1 (bottom) *sbm* sites.

glutamine occupies this position, which interacts with Lys3 and makes a pair of hydrogen bonds with the adenine of base pair 2 of the GANTC motif. In other TraM proteins, this position is occupied by histidine or threonine residues, which could also participate in defining DNA binding specificity through hydrogen bonding interactions with the major groove faces of the base pairs [59]. In addition, isoleucine and glycine can also be found at position 5. Isoleucine could help define DNA specificity through van der Waals interactions (eg: with 5-methyl group of thymine). In contrast, a glycine at this position could lead to a cavity at the protein-DNA interface and could allow more flexibility in the DNA contacting residues at positions 3 and 5.

Of the 11 classes of TraM that we have identified by sequence analysis, four have been experimentally investigated in terms of their DNA binding specificity. As we have shown here, F and pED208 TraM differ not only in their recognition of their 5 base pair motif elements, but also in the spacing between these elements on the DNA. We show that the β-strand residues dictate specificity for the 5 base pair motif, while other aspects of the RHH structure must govern spacer specificity so that F TraM binds elements with one TA spacer between motifs 1 and 2 and a second between motifs 3 and 4, while pED208 TraM binds elements with a single central 2 base pair spacer. The most notable structural difference between these two proteins is in the α1-α2 loop, which is two residues shorter in pED208 TraM and may be involved in defining spacer specificity (Fig. 3.4b, 3.10). The TraM protein from plasmid R1 has a similar β-ribbon to pED208 with Lys3, Gln5 and Tyr7 and, as expected, it has been shown to selectively bind to DNAs with a GANTC motif [61]. However, the remainder of the R1 RHH domain is more similar to F with 78% overall sequence identity to F and a α1- α2 loop that is more similar to F than pED208.

Intriguingly, the high affinity R1 *sbmA* site consists of four GANTC motifs, separated by 6 base pair spacers (Fig. 3.11b). This would place the motifs 11 base pairs apart on the DNA, which may alleviate the need for staggered binding of TraM tetramers to drive cooperative unwinding of the DNA. Instead, in the R1 case we imagine that two TraM tetramers may bind sequentially to *sbmA* on one side of the DNA double helix. The reduced 11 base pair spacing could potentially allow for interactions between RHH domains on the same side of the DNA that are not possible with the 12 base pair spacings in the F and pED208 systems. The TraM protein from R100 is overall very similar to F TraM (89% sequence identity) however it is predicted to recognize a different DNA motif due to the presence of Arg3 and Ile5 in the β-ribbon. Two R100 DNA elements, termed *sbmC* and *sbmD*, have been shown to bind TraM tightly [62]. Alignment of these DNA elements with F *sbmA* indicates that the TA spacer elements are conserved, as is the guanine base at the 1 position of the motif (Fig. 3.11b). This would suggest that R100 TraM might bind its DNA with the same spacing preference as F, however it would recognize a different 5 base pair motif. This would be consistent with the previous finding that a R100 TraM R3K:I5N mutant can bind F *sbmA* and can drive F plasmid conjugation in an F *traM* mutant strain [31].

Chapter 4

Conclusions

4.1 Summary of results

The crystal structure of F TraM RHH bound to its *sbmA* is very similar to the overall structure of the pED208 ortholog. Four RHH domains cooperatively bind to an underwound *sbmA* site with two RHH dimers on one side of the DNA and two on the other. Instead of a single, 2-bp spacer at the center of the pED208 *sbmA*, F *sbmA* contains two TA spacers, one between motifs 1 and 2 and a second between motifs 3 and 4. The motif bound by the F RHH (PuCNGPy) is less well conserved than the pED208 GANTC motif. Comparison of the two structures suggests a structural basis for the relaxed specificity of F.

The EMSA results of binding affinities of a series of F *sbmA* mutants demonstrated a reduced specificity for the individual pentamer motifs compared to pED208. Both the sequence and positioning of the spacers are critical for TraM binding.

The most important F TraM binding site for conjugation, *sbmC*, also contains an obvious inverted repeat but does not bear significant sequence similarity to *sbmA*. Truncation of base pairs from both ends of the site defines the minimal 26 bp *sbmC* without loss of binding affinity. EMSA results indicated that, like *sbmA*, a pair of TraM tetramers cooperatively binds *sbmC*.

Several hundred TraM proteins have been identified to date. Analysis of the predicted DNA contacting residues from the β-strand of the 221 TraM orthologs reveals a striking conservation. The most conserved is position 7, either a tyrosine or a phenylalanine. Position 3 is also highly conserved while position 5 is the most variable. In terms of DNA binding specificity, four TraM proteins have been experimentally investigated: F, pED208, R1 and R100.

4.2 Conclusions

TraM proteins in the family of F-like plasmids are responsible for the selective recruitment of the plasmid to the TraD ATPase complex that forms the cytoplasmic entrance to the conjugative pore. Taken together, our results show that TraM proteins have evolved a way to allow their RHH DNA binding modules to interact with one another to create complex DNA binding specificities that mediate plasmid specificity within the family of F-like plasmids.

The ability of F TraM to not only bind *sbmA* but also the apparently unrelated *sbmC* sequence is an intriguing feature of this protein. TraM binding at *sbmC* appears to be more critical for driving conjugation than binding at either the *sbmA* or *sbmB* sites alone [32], however, it has not been determined whether the importance of *sbmC* is due to its position close to the *nic* site, its unique sequence, or a combination of these two. It has been suggested that TraM may drive conformational changes in the DNA that could facilitate DNA unwinding near the *nic* site. Indeed, measurement of plasmid supercoiling levels in the presence of TraM indicates that TraM causes a significant unwinding of the F plasmid *in vivo* [63]. Footprinting experiments using DNAs containing *sbmC* reveal not only protection at *sbmC*, but also a periodic protection pattern every 11-12 base pairs radiating out from the *sbmC* site [30]. These results are not readily explained by our crystal structures of either F or pED208 TraM bound to their respective *sbmA* sites, and may suggest alternative modes of binding to DNA. This difference between *in vitro* and *in vivo* DNA binding patterns suggests that other factors, specified by the plasmid (TraI, TraY, DNA conformation) or its host (IHF), which are thought to initiate relaxosome formation, might influence TraM binding in preparation for plasmid transfer [28]. Whereas *sbmA* and *sbmB* are involved in regulating *traM*

expression and are probably continuously bound by TraM *in vivo*, the occupancy of *sbmC* by TraM during vegetative growth is unknown. The role of all three TraM binding sites in altering DNA conformation as well as initiating and terminating DNA transport (these sites are almost the last sequences transferred during normal F plasmid transfer) awaits detailed *in vivo* analyses.

4.3 Future directions

4.3.1 Specific roles of *sbmA*, *B* and *C* for plasmid transfer and autoregulation

Previous research showed the different functions of TraM binding sites: *sbmA* and *sbmB* overlap the *traM* promoters for autoregulating *traM* transcription, while *sbmC* plays an important role in recruitment of the relaxosome [64]. Alignment of the three sites reveals great similarities between *sbmA* and *sbmB*, but *sbmC* is apparently unrelated. It will be interesting to study if the different modes of binding of TraM to its specific target sequences have important consequences for function. A mutant F plasmid with *sbmA* sequence at the *sbmC* position can be created and tested for the ability to drive plasmid transfer.

4.3.2 Probe the effect of TraM and relaxosome formation on *oriT*

In vitro helicase assays indicated that the TraI relaxase/helicase requires extensive (30-60 bps) DNA unwinding in the AT-rich region between the *nic* site and *sbmC* to initiate large scale plasmid unwinding for conjugative transfer [65]. It has been hypothesized that an important function of TraM binding at the *sbmC* site is to drive DNA distortion that facilitates TraI binding, DNA nicking and unwinding [28] [63], presumably in response to the mating signal upon the donor-recipient contact. Plasmid supercoiling experiments have shown that TraM significantly unwinds its target plasmids *in vivo*, corresponding to ~4 turns of dsDNA [63]. While our

structures do suggest a small degree of DNA unwinding upon TraM binding to *sbmA*, this unwinding is not sufficient to explain the plasmid supercoiling results. It is likely that TraM binds *sbmC* in a totally different way or that higher order interactions between TraM and other proteins in the relaxosome could be responsible for initiation of plasmid nicking and unwinding.

4.3.3 Structural basis of DNA recognition in the TraM family

Within the IncF group, the transfer regions of different plasmids exhibit high degree of homology with each other [8], including the *traM* gene. All the known TraM sequences are highly homologous considering conservative substitutions with the possible exception of pED208 [8]. Nonetheless, complementation of the defective *traM* gene of plasmid F by other allelic forms is very limited [66], providing evidence for the specificity of TraM binding from different plasmids. Structures of other members in the TraM family will provide insights into the recognition mechanisms of different TraM proteins.

Bibliography

[1] H. C. Neu, "The Crisis in Antibiotic Resistance," *Science,* vol. 257, pp. 1064-1073, 1992.

[2] Koonin EV, Makarova KS, Aravind L, "Horizontal gene transfer in prokaryotes: quantification and classification," *Annual Review of Microbiology,* vol. 55, pp. 709-742, 2001.

[3] Lederberg J, Tatum EL, "Gene recombination in E. coli," *Nature,* vol. 158, p. 558, 1946.

[4] Holmes RK, Jobling MG, "Genetics: Exchange of Genetic Information," in *Medical Microbiology*, Galveston, Univ of Texas Medical Branch, 1996.

[5] W. T, "Infective heredity of multiple drug resistance in bacteria," *Bacteriol Rev,* vol. 27, pp. 87-115, 1963.

[6] Mazodier P, Davies J., "Gene transfer between distantly related bacteria," *Annu Rev Genet.,* vol. 25, pp. 147-71, 1991.

[7] W. Hayes, "Observations on a Transmissible Agent Determining Sexual Differentiation in Bacterium coli," *Society for General Microbiology,* vol. 8, pp. 72-88, 1953.

[8] Frost LS; Ippen-Ihler K; Skurray RA, "Analysis of the sequence and gene products of the transfer region of the F sex factor," *Microbiol Rev,* vol. 58, pp. 162-210, 1994.

[9] Willetts N ; Wilkins B, "Processing of plasmid DNA during bacterial conjugation," *Microbiol Rev,* vol. 48, pp. 24-41, 1984.

[10] Wong, JJW; Lu, J; Glover, JNM, "Relaxosome function and conjugation regulation in F-like plasmids - a structural biology perspective," *MOLECULAR MICROBIOLOGY,* vol. 85, pp. 602-617, 2012.

[11] Lawley, TD; Gordon, GS; Wright, A; Taylor, DE, "Bacterial conjugative transfer: visualization of successful mating pairs and plasmid establishment in live Escherichia coli," *MOLECULAR MICROBIOLOGY,* vol. 44, pp. 947-956, 2002.

[12] Lu, J; Peng, Y; Arutyunov, D; Frost, LS; Glover, JNM, "Error-Prone PCR Mutagenesis Reveals Functional Domains of a Bacterial Transcriptional Activator, TraJ," *Journal Of Bacteriology,* vol. 194, pp. 3670-p3677, 2012.

[13] Will WR, Lu J, Frost LS., "The role of H-NS in silencing F transfer gene expression during entry into stationary phase," *Mol Microbiol.,* vol. 54, pp. 769-82, 2004.

[14] Silverman PM, Sholl A., "Effect of traY amber mutations on F-plasmid traY promoter activity in vivo," *J Bacteriol.,* vol. 178, pp. 5787-9, 1996.

[15] Penfold SS, Simon J, Frost LS., "Regulation of the expression of the traM gene of the F sex factor of Escherichia coli," *Mol Microbiol.,* vol. 20, pp. 549-58, 1996.

[16] Di Laurenzio, Laura; Frost, Laura S.; Paranchych, William, "The TraM protein of the conjugative plasmid F binds to the origin of transfer of the F and ColE1 plasmids," *Molecular Microbiology,* vol. 6, pp. 2951-2959, 1992.

[17] Will WR, Frost LS. 2006., "Hfq is a regulator of F-plasmid TraJ and TraM synthesis in Escherichia coli.," *J. Bacteriol,* vol. 188, p. 124–131, 2006.

[18] Zahrl D, Wagner A, Tscherner M, Koraimann G., "GroEL plays a central role in stress-induced negative regulation of bacterial conjugation by promoting proteolytic degradation of the activator protein TraJ," *J. Bacteriol.,* vol. 189, p. 5885–5894, 2007.

[19] Lau-Wong IC, Locke T, Ellison MJ, Raivio TL, Frost LS., "Activation of the Cpx regulon destabilizes the F plasmid transfer activator, TraJ, via the HslVU protease in Escherichia coli," *Mol. Microbiol.,* vol. 67, p. 516–527, 2008.

[20] Arthur DC, et al., "FinO is an RNA chaperone that facilitates sense-antisense RNA interactions.," *EMBO J,* vol. 22, p. 6346–6355, 2003.

[21] Cheah, K.C., and Skurray, R. , "The F plasmid carries an IS3 insertion within finO," *J Gen Microbiol ,* vol. 132, p. 3269–3275, 1986.

[22] Dostál L, Schildbach JF, "Single-stranded DNA binding by F TraI relaxase and helicase domains is coordinately regulated," *J Bacteriol. ,* vol. 192, pp. 3620-8, 2010.

[23] Lubomír Dostál, Sichen Shao and Joel F. Schildbach*, "Tracking F plasmid TraI relaxase processing reactions provides insight into F plasmid transfer," *Nucl. Acids Res. ,* vol. 39, pp. 2658-2670, 2011.

[24] Howard MT, Nelson WC, Matson SW., "Stepwise Assembly of a Relaxosome at the F Plasmid Origin of Transfer," *J Biol Chem.,* vol. 270, pp. 28381-6, 1995.

[25] James U. Bowie, Robert T. Sauer, "TraY proteins of F and related episomes are members of the Arc and Mnt repressor family," *Journal of Molecular Biology,* vol. 211, no. 1, p. 5–6, 1990.

[26] Lum PL, Schildbach JF., "Specific DNA recognition by F Factor TraY involves beta-sheet residues," *J Biol Chem,* vol. 274, pp. 19644-8, 1999.

[27] Nelson WC, Howard MT, Sherman JA, Matson SW., "The traY gene product and integration host factor stimulate Escherichia coli DNA helicase I-catalyzed nicking at the F plasmid oriT," *J Biol Chem.,* vol. 270, pp. 28374-80, 1995.

[28] Ragonese, H; Haisch, D; Villareal, E; Choi, JH; Matson, SW, "The F plasmid-encoded TraM protein stimulates relaxosome-mediated cleavage at oriT through an interaction with TraI," *MOLECULAR MICROBIOLOGY,* vol. 63, pp. 1173-1184, 2007.

[29] Matson SW; Morton BS, "Escherichia coli DNA helicase I catalyzes a site- and strand-specific nicking reaction at the F plasmid oriT," *The Journal Of Biological Chemistry,* vol. 266 (24), pp. 16232-16237, 1991.

[30] R. Fekete and L. Frost, "Characterizing the DNA contacts and cooperative binding of F plasmid TraM to its cognate sites at oriT," *Journal Of Biological Chemistry,* vol. 277, pp. 16705-16711, 2002.

[31] Wong, JJW; Lu, J; Edwards, RA; Frost, LS; Glover, JNM, "Structural basis of cooperative DNA recognition by the plasmid conjugation factor, TraM," *Nucleic Acids Research ,* vol. 39, pp. 6775-6788, 2011.

[32] Fu YH; Tsai MM; Luo YN; Deonier RC, "Deletion analysis of the F plasmid oriT locus," *Journal Of Bacteriology,* vol. 173 (3), pp. 1012-1020, 1991.

[33] Lawley, TD; Klimke, WA; Gubbins, MJ; Frost, LS, "F factor conjugation is a true type IV secretion system," *FEMS Microbiology Letters ,* vol. 224 (1), pp. 1-15, 2003.

[34] Beranek A; Zettl M; Lorenzoni K; Schauer A; Manhart M; Koraimann G, "Thirty-eight C-terminal amino acids of the coupling protein TraD of the F-like conjugative resistance plasmid R1 are required and sufficient to confer binding to the substrate selector protein TraM," *Journal Of Bacteriology,* vol. 186 (20), pp. 6999-7006, 2004.

[35] Lu, J; Frost, LS, "Mutations in the C-Terminal Region of TraM Provide Evidence for In Vivo TraM-TraD Interactions during F-Plasmid Conjugation," *Journal of Bacteriology,* vol. 187, pp. 4767-4773, 2005.

[36] J. Lu, Role of traM in F conjugation, Edmonton: University of Alberta. Dept. of Biological Sciences., 2004.

[37] Lu, J; Edwards, RA; Wong, JJW; Manchak, J; Scott, PG; Frost, LS; Glover, JNM, "Protonation-mediated structural flexibility in the F conjugation regulatory protein, TraM," *EMBO JOURNAL,* vol. 25, pp. 2930-2939, 2006.

[38] Dana L. Miller and Joel F. Schildbach, "Evidence for a Monomeric Intermediate in the Reversible Unfolding of F Factor TraM," *JOURNAL OF BIOLOGICAL CHEMISTRY,* vol. 278, pp. 10400-10407, 2003.

[39] Lu, J; Zhao, W; Frost, LS, "Mutational analysis of TraM correlates oligomerization and DNA binding with autoregulation and conjugative DNA transfer," *JOURNAL OF BIOLOGICAL CHEMISTRY,* vol. 279 (53), pp. 55324-55333, 2004.

[40] Lu, J; Wong, JJW; Edwards, RA; Manchak, J; Frost, LS; Glover, JNM, "Structural basis of specific TraD-TraM recognition during F plasmid-mediated bacterial conjugation," *Molecular Microbiology,* vol. 70, pp. 89-99, 2008.

[41] Verdino, P; Keller, W; Strohmaier, H; Bischof, K; Lindner, H; Koraimann, G, "The essential transfer protein TraM binds to DNA as a tetramer," *Journal Of Biological Chemistry,* vol. 274, pp. 37421-37428, 1999.

[42] Schreiter, Eric R.; Drennan, Catherine L, "Ribbon–helix–helix transcription factors: variations on a theme," *Nature Reviews Microbiology,* vol. 5, pp. 710-720, 2007.

[43] L. Aravind, Vivek Anantharaman, Santhanam Balaji, M. Mohan Babu andLakshminarayan M. Iyer, "The many faces of the helix-turn-helix domain: Transcription regulation and beyond," *FEMS Microbiology Reviews,* vol. 29, no. 2, p. 231–262, 2005.

[44] Brown BM, Sauer RT, "Assembly of the Arc repressor-operator complex: cooperative interactions between DNA-bound dimers," *Biochemistry,* vol. 32, pp. 1354-63, 1993 .

[45] Bowie JU, Sauer RT., "Equilibrium dissociation and unfolding of the Arc repressor dimer," *Biochemistry.,* vol. 28, pp. 7139-43, 1989.

[46] Schreiter, E. R., Wang, S. C., Zamble, D. B. & Drennan, C. L. , "NikR–operator complex structure and the mechanism of repressor activation by metal ions," *Proc. Natl Acad. Sci. USA,* vol. 103, p. 13676–13681, 2006.

[47] Phillips, K. & Phillips, S. E. , "Electrostatic activation of Escherichia coli methionine repressor," *Structure,* vol. 2, p. 309–316, 1994.

[48] Moncalian, G. & de la Cruz, F. , "DNA binding properties of protein TrwA, a possible structural variant of the Arc repressor superfamily," *Biochim. Biophys. Acta,* vol. 1701, p. 15–23, 2004.

[49] Schildbach, J. F., Robinson, C. R. & Sauer, R. T., "Biophysical characterization of the

TraY protein of Escherichia coli F factor.," *J. Biol. Chem,* vol. 273, p. 1329–1333, 1998.

[50] Lu J, den Dulk-Ras A, Hooykaas PJ, Glover JN, "Agrobacterium tumefaciens VirC2 enhances T-DNA transfer and virulence through its C-terminal ribbon-helix-helix DNA-binding fold," *Proc Natl Acad Sci U S A. ,* vol. 106, pp. 9643-8, 2009.

[51] H. D., "Studies on transformation of Escherichia coli with plasmids," *J Mol Biol,* vol. 166, pp. 557-580, 1983.

[52] Otwinowski L., Minor W., "Processing of X-ray diffraction data collected in oscillation mode," *Methods Enzymol,* vol. 276, pp. 307-326, 1997.

[53] McCoy A. J., Grosse-Kunstleve R. W., Adams P. D., Winn M. D., Storoni L. C., Read R. J. , "Phaser crystallographic software," *Journal of applied crystallography,* vol. 40, pp. 658-674, 2007.

[54] E. P., "Coot: model-building tools for molecular graphics," *Acta crystallographica. Section D, Biological crystallography,* vol. 60, pp. 2126-2132, 2004.

[55] Adams P. D., Afonine P. V., Bunkoczi G., Chen V. B., Davis I. W., Echols N., et al. , "PHENIX: a comprehensive Python-based system for macromolecular structure solution.," *Acta crystallographica. Section D, Biological crystallography,* vol. 66, pp. 213-221, 2010.

[56] Morris R. J., Perrakis A., Lamzin V. S., "ARP/wARP and automatic interpretation of protein electron density maps," *Methods Enzymol,* vol. 374, pp. 229-244, 2003.

[57] Lu, J.; Fekete, R. A.; Frost, L. S., "A rapid screen for functional mutants of TraM, an autoregulatory protein required for F conjugation," *Molecular Genetics & Genomics,* vol. 269, pp. 227-233, 2003.

[58] Seeman N. C., Rosenberg J. M., Rich A., "Sequence-specific recognition of double helical nucleic acids by proteins," *Proc Natl Acad Sci U S A,* vol. 73, pp. 804-808, 1976.

[59] Luscombe N. M., Laskowski R. A., Thornton J. M. , "Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level," *Nucleic Acids Res,* vol. 29, pp. 2860-2874, 2001.

[60] Lamoureux J. S., Maynes J. T., Glover J. N. , "Recognition of 5'-YpG-3' sequences by coupled stacking/hydrogen bonding interactions with amino acid residues," *J Mol Biol,* vol. 335, pp. 399-408, 2004.

[61] Geist C, Brantl S, "TraM protein of plasmid R1: in vitro selection of the target region

reveals two consensus 7 bp binding motifs spaced by a 4 bp linker of defined sequence," *Plasmid,* vol. 59(1), pp. 20-35, 2008.

[62] Abo T; Inamoto S; Ohtsubo E, "Specific DNA binding of the TraM protein to the oriT region of plasmid R100," *Journal Of Bacteriology,* vol. 173 (20), pp. 6347-6354, 1992.

[63] Mihajlovic S., Lang S., Sut M. V., Strohmaier H., Gruber C. J., Koraimann G., et al., "Plasmid r1 conjugative DNA processing is regulated at the coupling protein interface," *J Bacteriol,* vol. 191, pp. 6877-6887, 2009.

[64] Di Laurenzio, L.; Frost, L. S.; Finlay, B. B.; Paranchych, W, "Characterization of the oriT region of the IncFV plasmid pED208," *Molecular Microbiology,* vol. 5, pp. 1779-1790, 1991.

[65] Csitkovits VC, Dermić D, Zechner EL., "Concomitant reconstitution of TraI-catalyzed DNA transesterase and DNA helicase activity in vitro," *J Biol Chem. ,* vol. 279, pp. 45477-84, 2004.

[66] Willetts N, Maule J, "Specificities of IncF plasmid conjugation genes," *Genet Res.,* vol. 47(1), pp. 1-11, 1986.

[67] Dooner, Hugo K.; Robbins, Timothy P.; Jorgensen, Richard A, "Gene transfer between distantly related bacteria," *Annual Review of Genetics,* vol. 25, pp. 147-171, 1991.

[68] Fowler T ; Taylor L ; Thompson R, "The control region of the F plasmid transfer operon: DNA sequence of the traJ and traY genes and characterisation of the traY leads to Z promoter," *Gene ,* vol. 26, pp. 79-89, 1983.

[69] Lee SH; Frost LS; Paranchych W, "FinOP repression of the F plasmid involves extension of the half-life of FinP antisense RNA by FinO," *Molecular & General Genetics,* vol. 235(1), pp. 131-139, 1992.

[70] Freifelder, David. San Francisco : W.H. Freeman, Physical biochemistry : applications to biochemistry and molecular biology, San Francisco: W. H. Freeman, 1982.

[71] H. Ye, "Simultaneous determination of protein aggregation, degradation, and absolute molecular weight by size exclusion chromatography-multiangle laser light scattering," *Analytical Biochemistry,* vol. 356, pp. 76-85, 2006.

[72] P. J. Wyatt, "Light scattering and the absolute characterization of macromolecules," *Analytica Chimica Acta,* vol. 272, p. 1–40, 1993.

[73] M. Schwab, H. Gruber and G. Hogenauer, "The TraM protein of plasmid R1 is a DNA-

binding protein," *MolecularMicrobiology,* vol. 5, pp. 439-446, 1991.

[74] HO, SN; HUNT, HD; HORTON, RM; PULLEN, JK; PEASE, LR, "Site-directed mutagenesis by overlap extension using the polymerase chain reaction," *GENE,* vol. 77, pp. 51-59, 1989.

[75] Sonya S. Penfold, John Simon and Laura S. Frost, "Regulation of the expression of the traM gene of the F sex factor of Escherichia coii," *Molecular Microbiology,* vol. 20, pp. 549-558, 1996.

[76] Guyer MS; Reed RR; Steitz JA; Low KB , "Identification of a sex-factor-affinity site in E. coli as gamma delta," *Cold Spring Harbor Symposia On Quantitative Biology ,* vol. 45, pp. 135-140, 1981.

[77] Chang AC; Cohen SN, "Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid," *Journal Of Bacteriology ,* vol. 134, pp. 1141-1156, 1978.

[78] Santos, PM; Di Bartolo, I; Blatny, JM; Zennaro, E; Valla, S, "New broad-host-range promoter probe vectors based on the plasmid RK2 replicon," *FEMS MICROBIOLOGY LETTERS,* vol. 195, pp. 91-96, 2001.

[79] Chandler, M., and D. Galas , "Cointegrate formation mediated by Tn9. II. Activity of IS1 is modulated by external DNA," *J. Mol. Biol.,* vol. 170, pp. 61-69, 1983.

[80] Hirschel, Bernard J.; Galas, David J.; Chandler, Michael, "Cointegrate Formation by Tn5, but not Transposition, is Dependent on recA," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 79, pp. 4530-4534, 1982.

[81] M. JH., Experiments in molecular genetics, New York: Cold Spring Harbor, 1972.

[82] Schwab M; Gruber H; Högenauer G, "The TraM protein of plasmid R1 is a DNA-binding protein," *Molecular Microbiology,* vol. 5 (2), pp. 439-446, 1991.

[83] J. Strutt, "On the scattering of light by small particles," *Philosophical Magazine,* vol. 41, pp. 447-454, 1871.

[84] D. Some, "Characterizing Protein–Protein Interactions Using Composition Gradient Multi-angle Light Scattering," Wyatt Technology Corporation,, Santa Barbara, Californa, USA.

[85] Yang, "Crystallization of protein-DNA complexes," 2006.

[86] Mulec J1, Starcic M, Zgur-Bertok D., "F-like plasmid sequences in enteric bacteria of diverse origin, with implication of horizontal transfer and plasmid host range.," *Curr Microbiol,* vol. 44, pp. 231-5, 2002.