

Structure analysis of DNA relaxases, the key enzymes of bacterial conjugation:
TraA and its N-terminal relaxase domain of the Gram-positive plasmid
pIP501 show specific *oriT* binding and behave as dimers in solution

Vorgelegt von
Mgr inz. Jolanta Kopec

von der Fakultät für Prozesswissenschaften
Der Technischen Universität Berlin
Zur Erlangung des akademischen Grades
Doktor Ingenieur
- Dr. ing. -

Promotionsausschuss:

Vorsitzender:	Prof. Dr. Helmut Görisch
1. Gutachter:	Prof. Dr. Ulrich Szewzyk
2. Gutachter:	PD Dr. Elisabeth Grohmann
3. Gutachter:	Prof. Dr. Walter Keller (Karl-Franzens-Universität, Graz)

Tag der wissenschaftlichen Aussprache: 13. Juli 2006

Berlin 2006
D 83

*Nothing in the world can take the place of persistence.
Talent will not; nothing is more common than unsuccessful people with talent.
Genius will not; unrewarded genius is almost a proverb.
Education will not; the world is full of educated derelicts.
Persistence and determination alone are omnipotent,*

John Calvin Coolidge, Jr.

Parts of the thesis were published in:

1. **Jolanta Kopec**, Alexander Bergmann, Gerhard Fritz, Elisabeth Grohmann and Walter Keller. *TraA and its N-terminal relaxase domain of the Gram-positive plasmid pIP501 show specific oriT binding and behave as dimers in solution*. Biochem. J. (2005) **387**, 401-409
2. Brigitta Kurenbach, **Jolanta Kopec**, Marion Mägdefrau, Kristin Andreas, Christine Bohn, Mohammad Yaser Abajy, Walter Keller and Elisabeth Grohmann. *The TraA relaxase autoregulates the first putative type IV secretion system from Gram-positive pathogens*. Microbiol. (2006) **152**, 637-645

Acknowledgements

I would like to thank all members of Environmental Ecology (Institute of Environmental Technology, Technical University Berlin) and Structural Biology Group (Institute of Chemistry, Karl-Franzens-Universität Graz), especially the group leaders – Prof. Dr. Ulrich Szewzyk and Prof. Dr. Christoph Kratky.

I owe particular gratitude to the supervisors of my doctoral thesis, PD Dr. Elisabeth Grohmann and Prof. Dr. Walter Keller for the interesting research subject, their support and for giving me the possibility to work independently.

My grateful thanks are due to my colleagues for all the fruitful discussions about science and life: Dr. Monika Oberer, Dr. Brigitta Kurenbach, mgr inż. Andrzej Łyskowski, Mag. Tea Pavkov, Dr. Maria Alexandrino.

I also want to thank my family for the support and stamina, they have given me.

I want to thank Dipl. Ing. Christine Bohn for letting me be a part of her family during my stay in Berlin.

This work would not be possible without the financial support from NaFöG and Berliner Programm zur Förderung der Chancengleichheit für Frauen in Forschung und Lehre.

And last, but not least, I express my gratitude to Dipl. Ing. Yvonne Werzer for letting me be a part of her life.

Table of contents

1	Introduction	10
1.1	Plasmids	11
1.1.1	pQE expression system	11
1.2	Bacterial cell wall	13
1.3	Bacterial conjugation	14
1.3.1	Conjugation in Gram-negative bacteria	15
1.3.2	Conjugation in Gram-positive bacteria	17
1.4	Type IV secretion system	17
1.5	Plasmid pIP501	21
1.6	Origins of transfer of the RSF1010 family	22
1.7	Relaxases	23
2	Aims of the work	26
3	Materials	27
3.1	Chemicals	27
3.2	Kits	28
3.3	Enzymes	28
3.4	Protein standards	29
3.5	DNA standards	30
3.6	Chromatographic resins	30
3.7	Instruments	31
3.8	Oligonucleotides	32
3.9	Plasmids and vectors	33
3.10	Microorganisms	33
3.11	PCR programs	34
3.12	Buffers	34
3.13	Media	34
4	Methods	35
4.1	DNA methods	35
4.1.1	Preparative plasmid isolation	35
4.1.2	Cloning	35
4.1.2.1	Cloning of wild type <i>traA</i>	35
4.1.2.2	Cloning of <i>traAN</i> ₂₆₈	36
4.1.2.3	Cloning of <i>traAN</i> ₂₄₆	36

4.1.3	Plasmid DNA mini-prep.....	37
4.1.4	Isolation of plasmid RSF 1010	37
4.1.5	Preparation of hairpin_ <i>nic</i> oligonucleotide for crystallization trials	38
4.2	Microbiological methods.....	38
4.2.1	Preparation of <i>E. faecalis</i> JH2-2 (pIP501) cell lysate	38
4.2.2	Preparation and transformation of <i>E. coli</i> BL21(DE3) competent cells.....	38
4.2.3	Preparation of <i>E. coli</i> XL1Blue competent cells	39
4.2.4	Transformation of <i>E. coli</i> XL1Blue competent cells.....	39
4.3	Screening of transformants.....	39
4.3.1	PCR test	39
4.3.2	Restriction analysis.....	40
4.3.3	DNA labeling.....	40
4.4	Biochemical methods	40
4.4.1	TraA, TraAN ₂₆₈ and TraAN ₂₄₆ expression	40
4.4.2	TraA, TraAN ₂₆₈ and TraAN ₂₄₆ purification.....	41
4.4.3	Orf7 constructs expression and purification.....	42
4.4.3.1	GST-Orf7ΔTM	42
4.4.3.2	7His-Orf7	42
4.4.3.3	MBP-Orf7ΔTM	42
4.4.4	Zymogram analysis	43
4.4.5	Crystallization trials.....	43
4.4.5.1	Vapor diffusion.....	43
4.4.5.2	Microbatch.....	43
4.5	Biophysical methods	44
4.5.1	Circular Dichroism Spectroscopy (CD)	44
4.5.2	Secondary structure determination and thermal stability	44
4.5.3	Cross-linking experiments of TraA and TraAN ₂₄₆	45
4.5.4	Electrophoretic Mobility Shift Assays (EMSA).....	46
4.5.5	EMSA with dimeric forms of TraA and TraAN ₂₄₆	47
4.5.6	DNase I footprint of <i>tra</i> region of pIP501	48
4.5.7	Cleavage assay.....	48
4.5.8	Small Angle X-Ray Scattering (SAXS) measurements and molecule shape calculation.....	49
4.6	Bioinformatic methods	51

4.6.1	Primary sequence analysis.....	51
4.6.2	Sequence based secondary structure prediction of proteins	51
4.6.3	Protein tertiary structure prediction.....	52
4.6.4	Topology prediction	52
4.6.5	Prediction of secondary structure of DNA	52
4.6.6	Txt2dic program	52
5	Results	53
5.1	<i>oriTs</i> of RSF1010 family.....	53
5.2	TraA, TraAN ₂₆₈ and TraAN ₂₄₆ expression and purification.....	55
5.3	The relaxase contains a mixed α/β -fold	55
5.4	TraA and TraAN ₂₄₆ unfold at the same temperature	56
5.5	The relaxase domain is stabilized through DNA-binding	57
5.6	TraA and TraAN ₂₄₆ form dimers in solution.....	57
5.7	The shape of the TraAN ₂₄₆ molecule shows a dimeric form.....	58
5.8	TraA and TraAN ₂₄₆ show nicking activity in vitro.....	59
5.9	TraA and TraAN ₂₄₆ bind specifically to <i>oriT</i> _{pIP501}	60
5.10	The TraA relaxase binds to the <i>P_{tra}</i> promoter.....	64
5.11	DNase I footprint of <i>tra</i> _{pIP501}	65
5.12	Crystallization of the TraA:hairpin_ <i>nic</i> complex.....	66
5.13	Analysis of other components of the <i>tra</i> _{pIP501} operon.....	67
5.13.1	Orf7 is a VirB1 homolog.....	68
5.13.1.1	Expression of Orf7 constructs	69
5.13.1.2	Orf7 has lytic transglycosylase activity.....	71
5.13.2	3D structure prediction of the transglycosylase domain of Orf7	72
5.13.3	Orf10 is a VirD4 homolog.....	73
5.13.4	3D structure prediction of Orf10	74
5.14	Txt2dic – hand shake between Jasco and dicropot	76
6	Conclusions	78
7	Discussion.....	79
8	Outlook	83
9	References	85
10	Appendices	92
10.1	List of abbreviations	92
10.2	Protein sequence statistics (ProtParam (www.expasy.org)).....	93

10.3	Secondary Structure Prediction (PSIPRED).....	96
10.4	Accession numbers of Orf2-Orf15	98
10.5	The Delphi code of txt2dic program.....	98

Abstract

TraA is the DNA relaxase encoded by the broad-host-range Gram-positive plasmid pIP501. It is the second relaxase characterized from plasmids originating from Gram-positive organisms. TraA (654 amino acids) and the N-terminal domain (246 amino acids), termed TraAN₂₄₆, were expressed as 6×His-tagged fusions and purified. Small-angle X-ray scattering and chemical cross-linking proved that TraAN₂₄₆ and TraA form dimers in solution. Both proteins revealed *oriT*_{pIP501} cleavage activity on supercoiled plasmid DNA *in vitro*. *oriT* binding was demonstrated by electrophoretic mobility shift assays. Radiolabeled oligonucleotides covering different parts of *oriT*_{pIP501} were subjected to binding with TraA and TraAN₂₄₆. The K_D of the protein-DNA complex encompassing the inverted repeat (IR), the nick site and additional 7 bases, was found to be 55 nM for TraA, and 26 nM for TraAN₂₄₆. The unfolding of both protein constructs was monitored by measuring the change in the circular dichroism (CD) signal at 220 nm upon temperature change. The unfolding transition of both proteins occurred at around 42°C. CD spectra measured at 20°C showed 30 % α-helix and 13 % β-sheet for TraA and 27 % α-helix and 18 % β-sheet content for the truncated protein. Upon DNA-binding an enhanced secondary structure content and increased thermal stability was observed for the TraAN₂₄₆ protein suggesting an induced-fit mechanism for the formation of the specific relaxase-*oriT* complex. Both proteins bind to promoter region of *tra* operon and therefore regulate its transcription.

Expression of soluble Orf7 protein was achieved, the protein turned out to possess lytic transglycosylase (LT) activity. 3D structure models were calculated for Orf10 and LT domain of Orf7. These models can be used as search models in crystal structure solution by molecular replacement method.

1 Introduction

Since Sir Alexander Flemming in the early 20th century discovered penicillin, the war between bacteria and humans began. The antibiotic saved the lives of millions of people during the World War II, but bacteria evolved a mechanism to inactivate the antibiotics. New antibiotic substances have been discovered or synthesized but bacteria always seem to find a way to survive. The first resistant strains were isolated from patients in hospitals, where antibiotic substances were extensively used. Introduction of antibiotics in clinical use results in quick and efficient response of resistance genes development (Table 1), resistant bacteria are usually found just a few years after the first clinical use. Only vancomycin had been effective for 15 years, but eventually the bacteria found the way to inactivate also this antibiotic (Schwarz, S. and Chaslus-Dancla, 2001).

Table 1. Time spans between discovery, first clinical use and occurrence of resistant bacteria (adapted from Schwarz, S. and Chaslus-Dancla (2001).

Antibiotic	Discovery/production	Introduction into clinical use	Occurrence of resistant bacteria
Penicillin	1940	1943	1940
Streptomycin	1944	1947	1947, 1956
Tetracycline	1948	1952	1956
Erythromycin	1952	1955	1956
Vancomycin	1956	1972	1987
Nalidixic acid	1960	1962	1966
Gentamycin	1963	1967	1970
Fluoroquinolones	1978	1982	1985

Streptococcal resistance was proved to be directly associated with antibiotic selection pressure (Albrich *et al.*, 2004). Decrease of antibiotics usage increases the susceptibility of the pathogens (Seppala *et al.*, 1997), but the reduction of use of an antibiotic must happen before the resistant strains are ubiquitous and the strains acquired resistance against multiple antibiotics (Swartz, 1997). One of the most efficient ways to disseminate antibiotic resistance between bacterial cells is conjugation.

1.1 Plasmids

Plasmids are extrachromosomal, usually circular, double stranded (ds) DNA molecules, which replicate independently from chromosomal DNA. The size of plasmids range from 1 to over 400 kbp (kilo base pairs). Usually resistance traits, in form of plasmid encoded resistance genes, are horizontally transferred between bacterial cells.

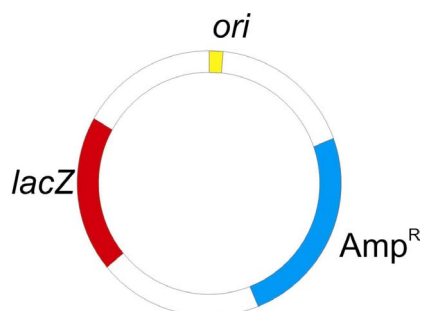


Figure 1. Scheme of a plasmid. Origin of replication (*ori*) is coloured yellow, the ampicillin resistance gene (*Amp^R*) is marked blue and the gene responsible for lactose catabolism (*lacZ*) is marked red.

Each plasmid, see Figure 1, has an origin of replication (*ori*) and most of the plasmids carry a genes coding for a catabolic enzyme, for example *lacZ*, or an antibiotic resistance gene, in this case coding for β -lactamase (*Amp^R*).

Plasmids are widely used for cloning purposes. They are quite easy to handle and there are well-established protocols for cloning and transformation.

1.1.1 pQE expression system

Heterologous protein expression systems have many advantages over protein isolation from the original source. The most important one is the abundance; the protein of interest might be expressed in minimal quantities in its original cells, whereas recombinant expression may reach more than 50% of total protein in *E. coli* (Studier and Moffatt, 1986).

The easiest way of producing recombinant proteins is to express them in *E. coli*. The metabolic pathways of this organism are well known and genetically modified strains, optimized for protein expression, are available. For example, *E. coli* strain BL21(DE3) (Invitrogen) can be used for protein expression under control of T7/T7*lac* promoter. This strain is lon and Omp protease deficient rendering the protein purification easier, since proteolytic digest of the protein of interest is reduced.

The simplicity of cloning and site-directed mutagenesis allowed the construction of fusion proteins, consisting of a tag and the protein of interest. The tag can be a short polypeptide sequence or a large protein. The tags are used during protein purification to perform affinity chromatography. His-tag is one of the most frequently used tags, consisting of 5 to 8

consecutive histidine residues. They interact with divalent cations, for example Ni^{2+} or Cu^{2+} , which are immobilized on a chelating gel matrix.

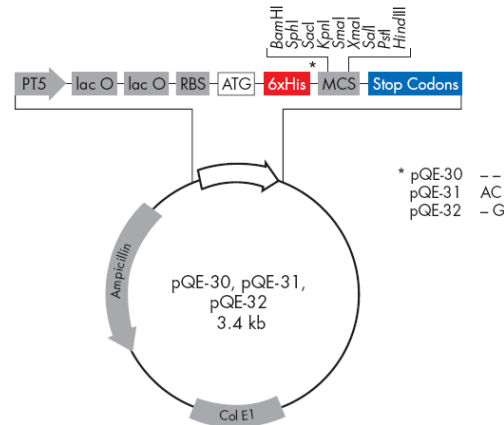


Figure 2. pQE-30 series vectors for N-terminal 6xHis tag constructs

pQE-30 series plasmids (Figure 2) allow expression of protein of interest in fusion with 6xHis. The series consists of three vectors, permitting cloning in all three reading frames. The 6xHis tag at the N-terminus of the protein can influence crystallization of the purified protein. The tag can be cleaved off using a specific protease if the specific cleavage sequence is placed between the tag and the target protein. pQTEV vector has the TEV (Tobacco Etch Virus) protease cleavage site cloned between 7xHis tag and the multiple cloning site (Figure 3).

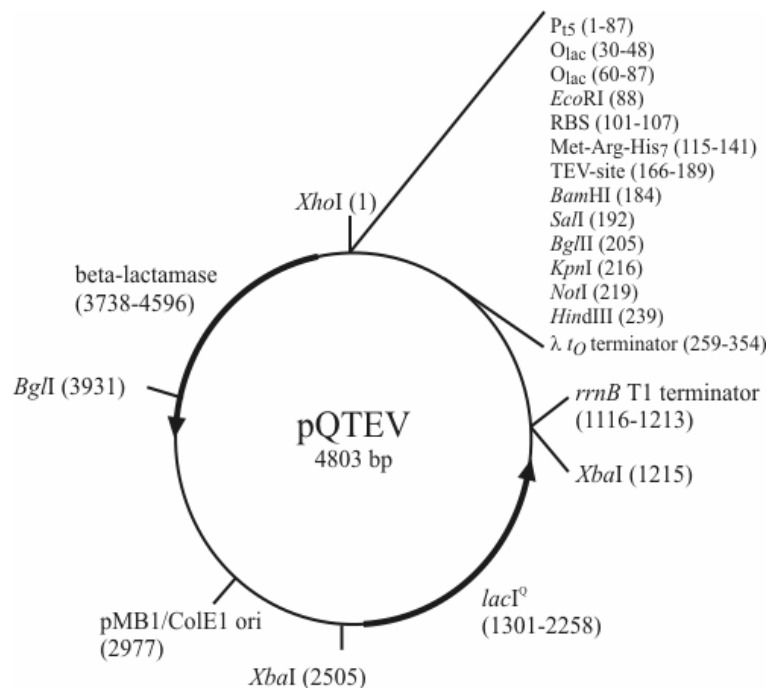


Figure 3. The vector map of pQTEV expression plasmid

1.2 Bacterial cell wall

The difference between Gram-positive (Gram +) and Gram-negative (Gram -) bacterial cells is based on the structure of their cell walls.

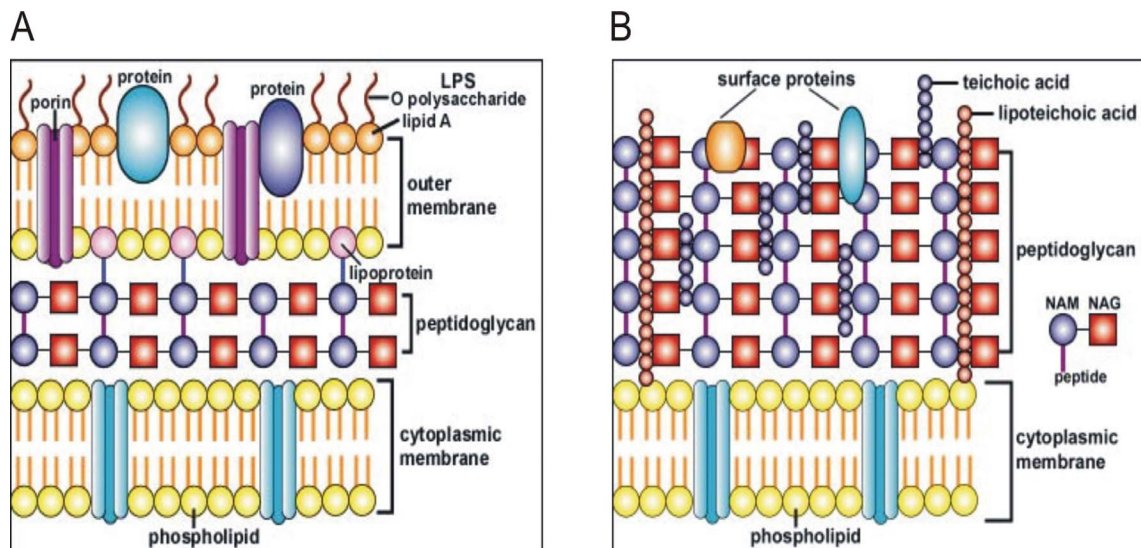


Figure 4. Structure of bacterial cell wall. A: Gram - cell wall. B: Gram + cell wall. Adapted from <http://www.cat.cc.md.us/courses/bio141/lecguide/unit1/prostruct/u1fig9b.html>

Figure 4 shows the schematic representations of Gram + and Gram – cell wall structures. The cell envelope of Gram – bacteria consists of a cytoplasmic membrane, which is a phospholipid bilayer, periplasmic space, in which a thin layer of peptidoglycan is present and an outer membrane, in which proteins are embedded or form channels through the membrane. The outer layer of the Gram - cell consists of LPS (lipopolysaccharides). The cell wall of Gram + cells consists of a cytoplasmic membrane and a thick layer of peptidoglycan. The peptidoglycan is composed of chains of alternating N-acetylglucosamine (NAG) and N-acetylmuramic acid (NAM), cross-linked between NAMs through peptide linkages, see Figure 5.

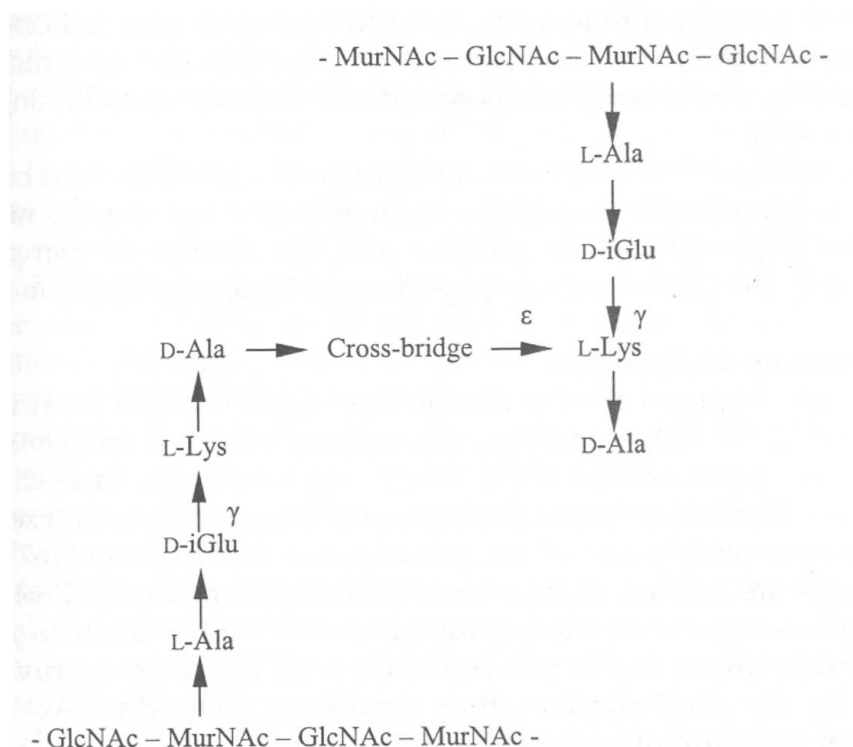


Figure 5. Primary structure of typical peptidoglycan of *E. faecalis*. Adapted from Coyette and Hancock (2002)

The scaffold made from sugar moieties is the same for all bacteria, but the cross-links differ in different species. Most of the Gram - bacteria have an L-ala-D-Glu-*meso*-diaminopimelic acid-D-Ala stem, cross-linked through a peptide bond between the ϵ -amino group of *meso*-Diamino-pimelic acid and the carboxy-group of D-Ala of the next unit (Schleifer and Kandler, 1972). The peptide stem of Gram + genera (*Enterococcus*, *Streptococcus* and *Staphylococcus*) is composed of L-Ala-D-Gln-L-Lys-D-Ala-D-Ala and the side chain, attached through the ϵ -NH₂ group of lysine is different for different species – *Enterococcus faecalis* – L-Ala₂, *Enterococcus faecium* – D-Asx and *Staphylococcus aureus* – Gly₅ (Schleifer and Kandler, 1972; Bouhss *et al.*, 2001; Bouhss *et al.*, 2002; Coyette and Hancock, 2002; Arbeloa *et al.*, 2004).

1.3 Bacterial conjugation

The most frequent way how bacteria acquire antibiotic resistance genes is conjugation. It is a specific, unidirectional process of horizontal gene transfer, during which plasmid DNA is transferred from donor to recipient cell. The DNA transfer requires physical cell-to-cell contact, unlike other horizontal gene transfer processes – transduction and transformation.

A channel spanning the cell envelope has to be formed to provide the way for DNA transfer. This is done by mating pair formation (mpf), a multiprotein complex. The proteins taking part

in *mpf* are encoded on the plasmid, which is transported through the channel. This multiprotein complex is joined with a DNA-protein complex (relaxosome) through a coupling protein. The relaxosome is a specific complex of origin of transfer (*oriT*) and the conjugative relaxase.

Relaxase is an enzyme, which catalyses the cleaving-joining reaction of the phosphodiester bond at nick site (*nic*) which lies within the *oriT*. The relaxase is covalently bound to the 5' terminus of the cleaved single strand. The nucleoprotein complex is then transported through the pore between the donor and recipient cell. The transferred single strand DNA (T-strand) is recircularized by the relaxase in the recipient cell and the complementary strands are synthesized in both, donor and recipient cells.

1.3.1 Conjugation in Gram-negative bacteria

The conjugation in Gram - bacteria is known quite well. Most information on conjugative processes was gained from experiments performed on F plasmid and plasmids RP4 and R388, all of them are *E. coli* plasmids (Frost *et al.*, 1994; Lanka and Wilkins, 1995; Pansegrau and Lanka, 1996a; Zechner *et al.*, 2000; Llosa *et al.*, 2002).

A schematic model of the conjugation process, based on the knowledge about RP4 plasmid, is presented in Figure 6.

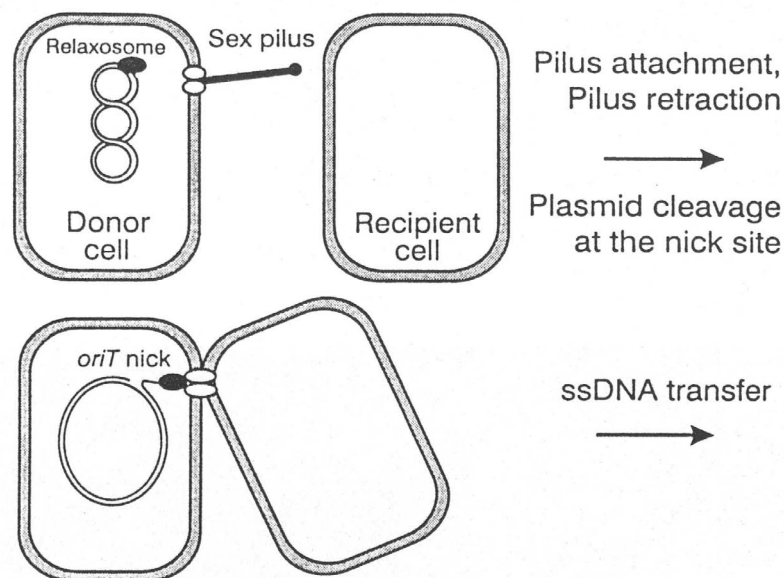


Figure 6. (Continued)

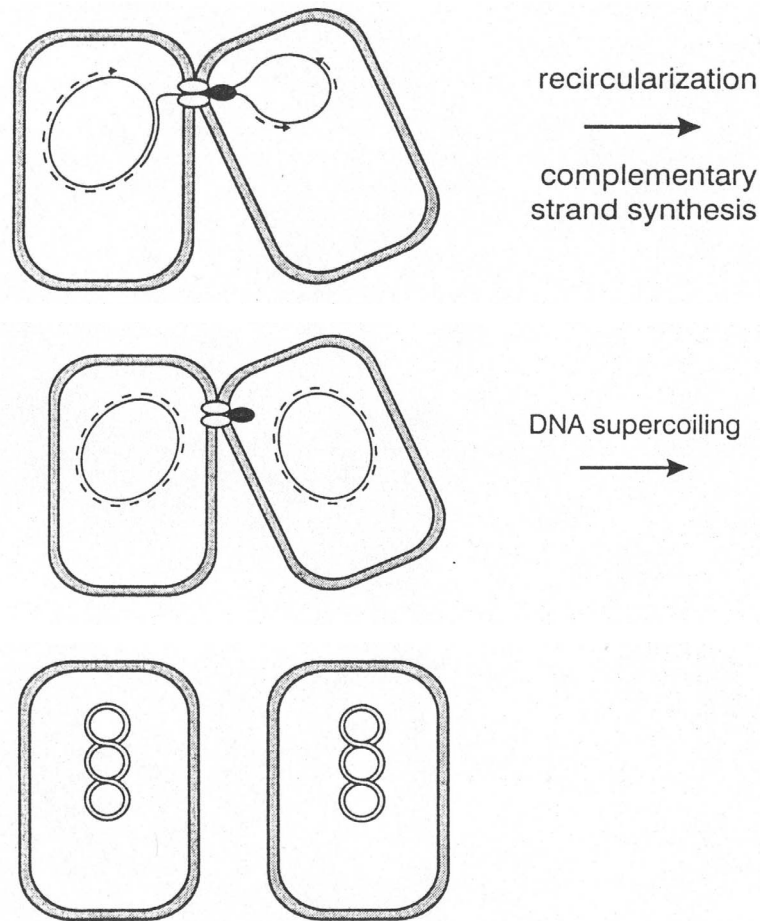


Figure 6. Model of bacterial conjugation. Adapted from Pansegrau and Lanka (1996a).

The process starts with pilus attachment to the recipient cell, followed by retraction of the pilus thus establishing an intimate contact between the cells. Then, the mating pore is formed and the DNA is nicked. Afterwards, the ssDNA (single strand DNA) is transferred through the pore, recircularized and the complementary strand is synthesized. Later, the DNA is supercoiled and both cells contain the conjugative plasmid.

Transfer functions of RP4 plasmid are divided into two regions – Tra1 and Tra2. Tra2 codes entirely for proteins involved in mating pair formation and entry exclusion. Tra1 region is dedicated to DNA processing proteins, including relaxosome proteins, a DNA topoisomerase and a DNA primase (Pansegrau and Lanka, 1996a).

The sequence of the mating pair formation region of R388 includes 11 genes named *trwN* to *trwD* as well as four additional open reading frames (ORFs) involved in entry exclusion and regulation functions (de Paz *et al.*, 2005). The DNA region involved in conjugal DNA metabolism consists of three genes (Llosa *et al.*, 1994): *trwA* (relaxation enhancer, transcriptional repressor of *trwABC* operon (Moncalian *et al.*, 1997)), *trwB* (coupling protein (Cabezón *et al.*, 1997)) and *trwC* (relaxase, helicase (Llosa *et al.*, 1996)). The crystal

structures of TrwB (Gomis-Ruth and Coll, 2001) and TrwC (Guasch *et al.*, 2003) are known. A high-throughput conjugation assays were conducted, which allowed identification of a specific inhibitor of the R388 conjugation process – DHCA (dehydrocrepenynic acid), a C₁₈, *cis*-unsaturated (9, 12, 14) fatty acid (Fernandez-Lopez *et al.*, 2005). The authors of this study propose that the fatty acid inhibits the DNA transfer replication (Dtr) machinery of R388 plasmid.

1.3.2 Conjugation in Gram-positive bacteria

The major difference in the conjugation process of Gram + and Gram – bacteria is the way to establish the cell-cell contact. Gram – bacteria produce pili, which bring the conjugating cells together by retraction, the same might be happening in Gram + bacteria, pili-like structures have been found in Gram + *Corynebacterium diphtheriae* (Ton-That and Schneewind, 2004), Group B *Streptococcus* (Lauer *et al.*, 2005), *Streptococcus pyogenes* (Mora *et al.*, 2005).

Some enterococci plasmids respond to pheromones, peptides consisting of seven or eight amino acids, produced by plasmid free cells. It is usually the case for *Enterococcus faecalis*, whereas no pheromone-responding plasmids are found in *Enterococcus faecium* or other enterococci. To provide specificity of the pheromone-plasmid interaction, the plasmid codes for a surface lipoprotein, which binds the cognate peptide (Clewell and Francia, 2004).

The plasmid transfer of non-pheromone responding plasmids in Gram + bacteria is described in Grohmann *et al.*, (2003).

1.4 Type IV secretion system

Type IV secretion systems (T4SS) comprise macromolecular transfer systems ancestrally related to conjugation systems of Gram + and Gram – bacteria (Ding *et al.*, 2003). They are able to transfer virulence factors across membranes of pathogenic bacteria and transport DNA from bacterial to bacterial or eukaryotic cells, disseminating antibiotic resistance or diseases, e.g. pertussis.

The T4SS can be classified according to amino acid sequence similarity. The systems related to VirB/D4 T4SS of *Agrobacterium tumefaciens* belong to the IVA systems (Figure 7a). Another group of putative T4SS based on protein sequence similarity, including bacterial pathogens like *Bordetella bronchiseptica*, *Brucella* spp., *Bartonella* spp. and *Actinobacillus actinomycetemcomitans* and non-pathogenic bacteria like *Caulobacter crescentus*, *Sinorhizobium meliloti*, *Rhizobium etli* and *Thiobacillus ferrooxidans* (Figure 7b) and systems

related to Dot/Icm T4SS of *Legionella pneumophila* (Segal and Shuman, 1999) are classified as IVB (Figure 7c) (Christie and Cascales, 2005).

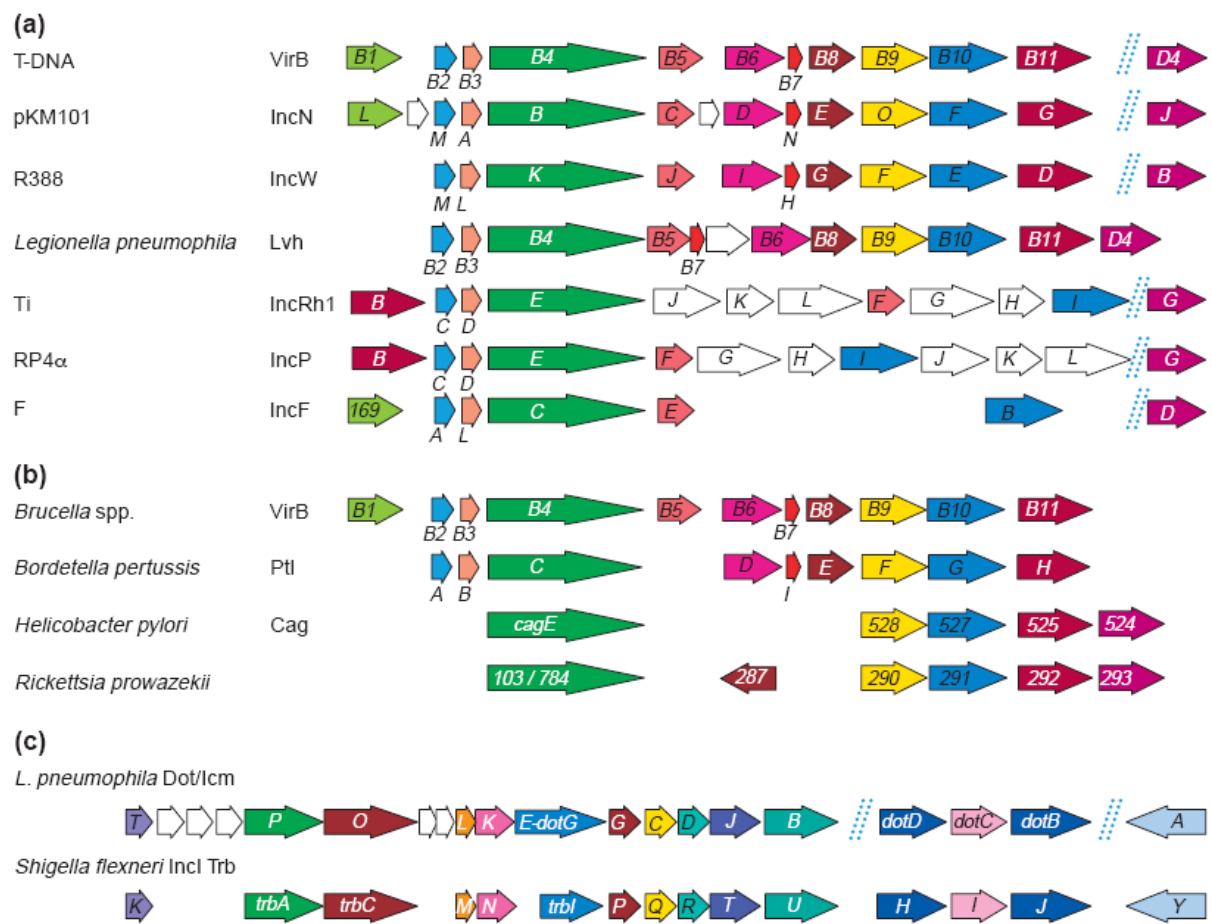


Figure 7. Sequence homology within type IV secretion systems. Adapted from Schroder and Lanka (2005)

The T4SSs can also be grouped according to their functions:

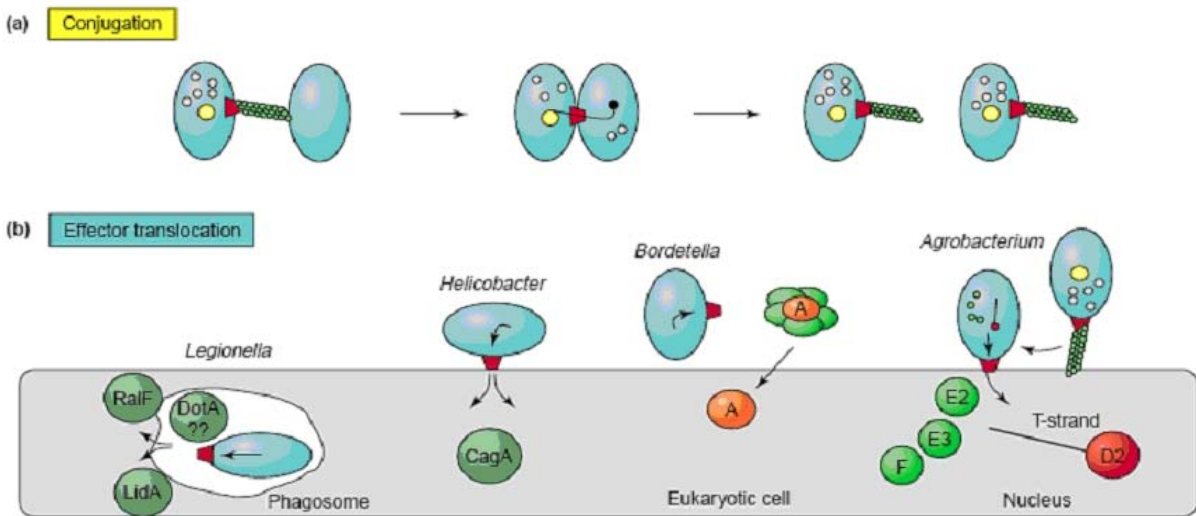


Figure 8. (Continued)

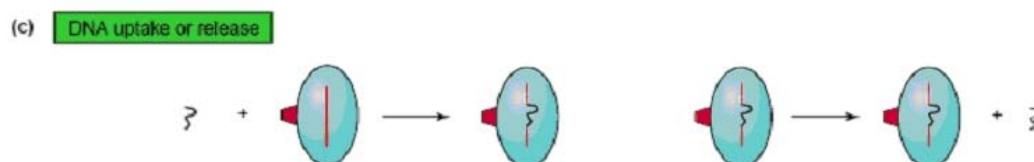


Figure 8. The functions of T4SS. Adapted from Ding *et al.*, (2003)

One group mediates conjugative transfer of mobile elements to bacterial cells (Figure 8a). “Effector translocation” systems transmit protein or DNA to eukaryotic cells (Figure 8b). These two groups rely on direct contact between donor and recipient cells, whereas the third group – “DNA uptake or release” systems exchange genetic information with the extracellular environment (Figure 8c) (Christie and Cascales, 2005).

The functions of the proteins of the mpf complex of *Agrobacterium tumefaciens* Ti plasmid are already known or postulated, as reviewed in Schroder and Lanka (2005), Table 2.

Table 2. The functions of VirB/D4 complex components (Schroder and Lanka, 2005)

Component	Function
VirB1	Perforation of the cell wall
VirB2	Structural subunit of pili
VirB3	Outer-membrane component
VirB4	Motor of secretion ^A
VirB5	Component of pilus, cell adhesion
VirB6	Modulator of secretion channel ^A
VirB7	Lipoprotein connecting pilus and core complex
VirB8	Bridge through periplasm
VirB9	Outer-membrane anchor
VirB10	Energy sensor gating the mpf channel
VirB11	NTPase fuelling the secretion machinery
VirD4	Cytoplasmic gate to the secretion channel

^A – Postulated function

There are two working models presenting the mechanism of macromolecule transport through T4SS, Figure 9, (Atmakuri *et al.*, 2004).

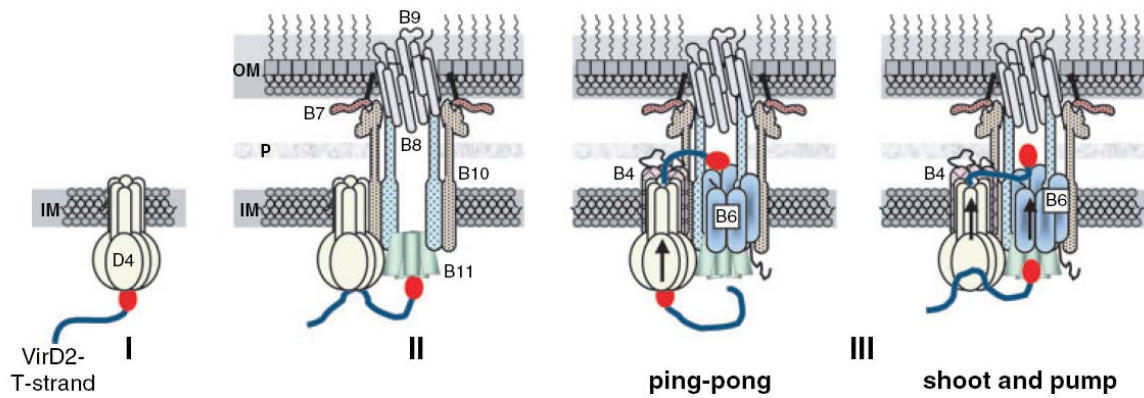


Figure 9. Steps of DNA transfer through VirB/D4 T4SS. Adapted from Atmakuri *et al.*, (2004)

The first two steps of the transfer are the same for both models – VirD2 covalently bound to T-DNA is recruited by VirD4, then the VirD2-T-DNA complex is transferred to VirB11. In the ‘ping-pong’ model, the VirD2 is partially unfolded and transferred back to VirD4, whereas in the ‘shoot and pump’ model, VirB11 unfolds VirD2, interacts with other Vir proteins to deliver the VirD2 across the inner membrane, at the same time VirD4 uses ATP energy to pump the DNA strand across the membrane. Then the VirD2-T-DNA complex enters the secretion pilus built of VirB2 and VirB9 subunits (Atmakuri *et al.*, 2004).

Transfer DNA immunoprecipitation assay (TrIP) of the *A. tumefaciens* VirB/D4 system provided new insight into the pathway of DNA translocation (Cascales and Christie, 2004). Figure 10 depicts the model of the translocation. The dark gray components interact with the VirD2-T-DNA and the others (light gray) take part in machine biogenesis and/or regulate substrate passage through the channel (Cascales and Christie, 2004).

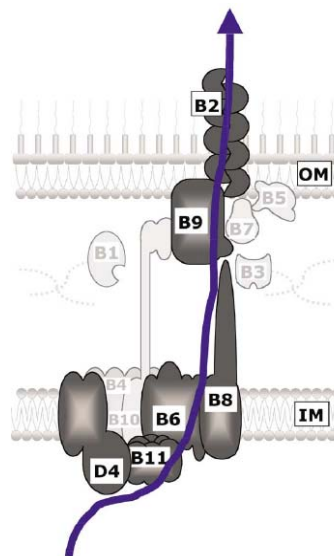


Figure 10. The T-strand translocation pathway, as identified by immunoprecipitation. IM - inner membrane, OM - outer membrane. Adapted from Christie and Cascales (2005)

The VirD2-T-DNA complex transport begins with interaction of T-DNA with VirD4. The T-DNA is then transferred to VirB11, VirB7 is essential for this step, probably by stabilizing the multiprotein complex, enabling effective contact between VirD4 and VirB11. VirB6 and VirB8 mediate the next step, the transfer through the inner membrane. This step is mediated by ATP binding or hydrolysis by VirD4, VirB4 and VirB11. VirB2 and VirB9 take part in the last step of VirD2-T-DNA transfer. VirB3, VirB5 and VirB10 do not interact with T-DNA, but are essential for substrate trafficking through the periplasm (Christie and Cascales, 2005).

1.5 Plasmid pIP501

pIP501 is a conjugative plasmid originally isolated from *Streptococcus agalactiae* (Horodniceanu *et al.*, 1976), a Gram + pathogenic bacterium. This plasmid belongs to the Inc18 incompatibility group and carries a chloramphenicol (*cat*) resistance gene as well as a resistance gene against the macrolide-lincosamide-streptogramin B group of antibiotics (*ermB*). pIP501 can be transferred by conjugation not only to other Gram + bacteria including multicellular *Streptomyces lividans*, but also to Gram – *E. coli* (Kurenbach *et al.*, 2003).

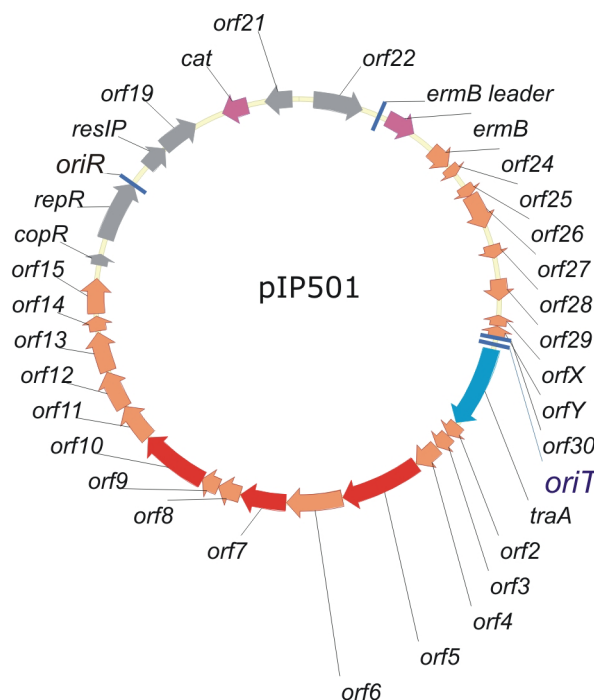


Figure 11. Schematic view of plasmid pIP501.

Sequencing of pIP501 (Kurenbach *et al.*, 2003; Thompson and Collins, 2003) revealed that the transfer region extends over half of the plasmid, 15 kbp out of 30.6 kbp. The operon consists of 15 open reading frames (*orfs*), starting at *traA* through *orf15* (Kurenbach *et al.*, 2006) (see Figure 11). Three *orfs* (*orf5*, *orf7* and *orf10*) are marked red on Figure 11, they code for protein products showing sequence homology to proteins involved in VirB/D4 type

IV secretion systems. Orf5_{pIP501} shows homology to VirB4. Orf10 is homologous to VirD4 and Orf7 to VirB1 (Grohmann *et al.*, 2003; Grohmann, 2005).

1.6 Origins of transfer of the RSF1010 family

The *oriT* of pIP501 has been mapped and the function of Orf1 was determined to be a relaxase (Wang and Macrina, 1995a). The full length Orf1 (TraA) and the putative relaxase domain, first 293 amino acids (aa) were expressed as GST fusions and their activity was confirmed *in vitro* (Kurenbach *et al.*, 2002).

Each relaxase recognizes a specific DNA sequence and cleaves strand- and site-specifically a phosphodiester bond at nick site within the *oriT*, which is the only sequence required for conjugal DNA transport (Pansegrau and Lanka, 1996a).

The *oriT* of pIP501 (Wang and Macrina, 1995b) belongs to the RSF1010-*oriT* family (Zechner *et al.*, 2000), along with *oriT*s of RSF1010 (Scholz *et al.*, 1989) and R1162 (Brasch and Meyer, 1986) (*E. coli*), pTF1 (Drolet *et al.*, 1990) (*Thiobacillus ferrooxidans*), pTiC58 (Cook and Farrand, 1992) (*Agrobacterium tumefaciens*), pSC101 (Bernardi and Bernardi, 1984) (*Salmonella*), pRE25 (Schwarz, F. V. *et al.*, 2001) (*Enterococcus faecalis*) and pGO1 (Climo *et al.*, 1996) (*Staphylococcus aureus*). The RSF1010-family *oriT*s aligned using T-Coffee algorithm (Notredame *et al.*, 2000), available at <http://www.ch.embnet.org/software/TCoffee.html>, is shown on Figure 12.

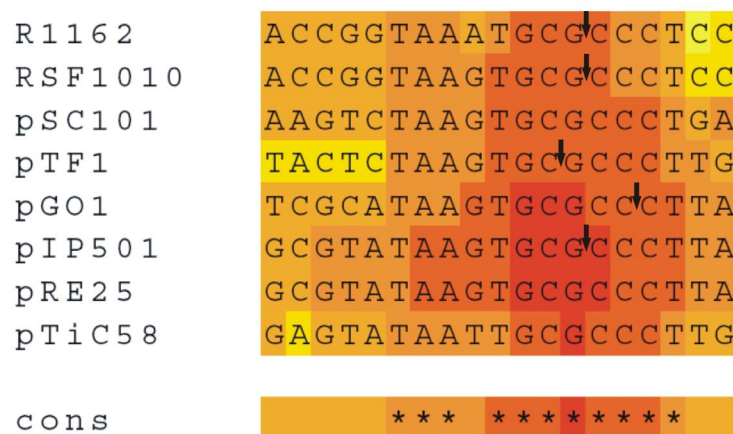


Figure 12. Alignment of *oriT* regions of RSF1010 family. The known nick sites are marked with a vertical arrow. Modified from Grohmann *et al.* (2003)

The highly conserved *oriT* region is presented on red background, the 5' – TGC GCCCT – 3' sequence is present in all *oriT*s in this family. The *nic* sites in five *oriT*s of the RSF1010 family are known, they are in the center ± 2 of the consensus sequence.

1.7 Relaxases

Relaxases are essential for conjugative plasmid transfer; they start the process by nicking plasmid DNA by trans-esterification (Byrd and Matson, 1997).

The relaxases possess three motifs (Zechner *et al.*, 2000), two of which are conserved in RSF1010 relaxase family (IncQ-type). The tyrosine in motif I (Figure 13) is responsible for the catalytic activity of the protein and its linkage to the 5' terminus of the cleaved strand (Pansegrau *et al.*, 1993). Motif II takes part in non-covalent interactions with the 3' terminus of the cleaved strand (Pansegrau and Lanka, 1996a). Motif III forms another part of the catalytic center with two conserved histidine residues throughout all conjugative relaxases and also other rolling circle initiator proteins. There are two models for the role of the motif III: i) the two histidine residues are involved in metal ion coordination (Ilyina and Koonin, 1992), ii) one His activates the active-site tyrosine residue by proton abstraction, and the other gives a proton to form the free 3' hydroxyl terminus of the DNA, in the rejoining reaction the roles could be reversed (Pansegrau and Lanka, 1996a).

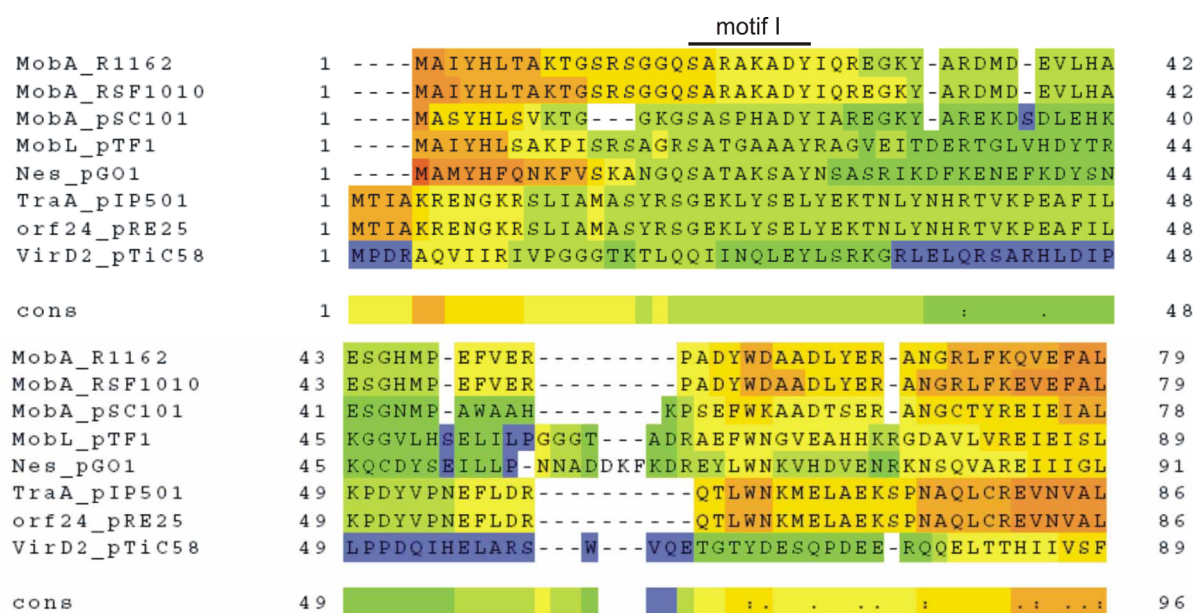


Figure 13. (Continued)

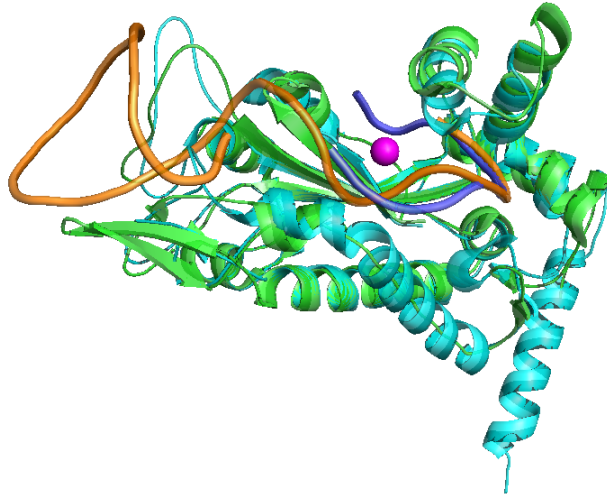


Figure 14. Superposition of TrwC (green) and TraI (cyan)

Figure 14 presents a superposition of the two structures, TrwC (pdb 1QX0) (Guasch *et al.*, 2003) and TraI (pdb 2A0I) (Larkin *et al.*, 2005). Both relaxase domains have been crystallized with the respective *oriT* fragment. Additionally, the TraI structure without the target DNA is also known (pdb 1P4D) (Datta *et al.*, 2003). In both structures, metal ions are present. TrwC contains Zn^{2+} and TraI contains Mg^{2+} . The metal ions are positioned in the same place in the superimposed structures, represented as a pink ball in Figure 14. The part of the ssDNA 5' of the *nic* sites, shown as orange and blue ribbons, interacts with the DNA binding clefts.

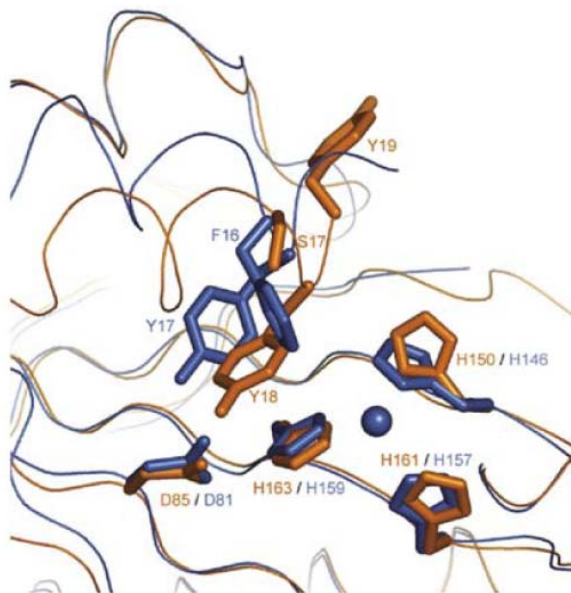


Figure 15. Active sites of TrwC (orange) and TraI (blue). Adapted from Larkin *et al.* (2005)

The active sites of TrwC and TraI, see Figure 15, contain the three histidine residues, the metal ions, represented as a blue ball, and the catalytic tyrosine.

2 Aims of the work

The goal of the study is to supply new biophysical and structural data concerning TraA and its relaxase domain. In particular, the DNA binding properties of the proteins were to be examined.

Preliminary work on protein expression and purification of the proteins involved in the putative T4SS, as well as the protein sequence analysis of the *tra* region of the pIP501 were set as further goals.

3 Materials

3.1 Chemicals

Acrylamide	Fluka
Agarose	Eurogentec
Ammoniumpersulfate	Sigma
Ampicillin	Sigma
Bisacrylamide	Fluka
Brij 58	Fluka
Bromophenol Blue	Fluka
BugBuster® HT	Novagen
B-PER (protein extraction reagent)	Pierce
CaCl ₂	Sigma
DTT (Dithiothreitol)	Sigma
EDTA (Ethylenediaminetetraacetic acid Disodium salt Dihydrate)	Sigma
Ethidium bromide	Sigma
Formamide	Roth
Glycerol	Merck
Hionic fluor	Perkin Elmer
Imidazole	Fluka
IPTG (Isopropyl β-D-1-thiogalactopyranoside)	Roth
LiCl	Sigma
MgCl ₂	Fluka
MOPS (3-Morpholinopropanesulfonic acid)	Fluka
NaOH	Merck
Paraffin oil	Merck
Phenol/chloroform/isoamylalcohol (25:24:1)	Roth
QA-Agarose TM (low melting)	Qbiogene
RbCl	Fluka
SDS (Sodium dodecyl sulfate)	Biomol
Silicone oil	Merck

Soluene-350	Perkin Elmer
Streptomycine	Sigma
Sucrose	Fluka
Trizma base	Riedel-de Haën
Xylene cyanol	Fluka
α - ³² P-ATP	Hartmann Analytic GmbH
α - ³² P-ATP	NEN (PerkinElmer)
β -mercaptoethanol	Roth
γ - ³² P-ATP	Hartmann Analytic GmbH
γ - ³² P-ATP	NEN (PerkinElmer)

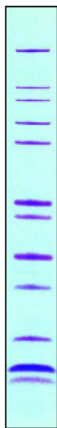
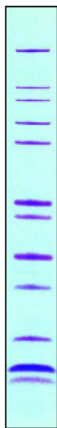
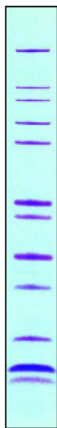
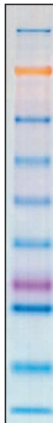
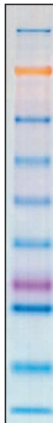
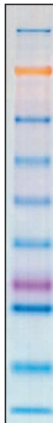
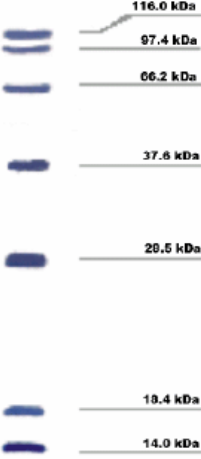
3.2 Kits

BCA Protein Assay	Pierce, Perbio Science
Big Dye Sequencing kit	Applied Biosystems
GenElute™ Plasmid Miniprep Kit	Sigma
Plasmid DNA Maxi kit	Qiagen
RotiTransform kit	Roth
Wizard® SV Gel and PCR Clean-Up kit	Promega
Cycle Sequencing Kit	Fermentas

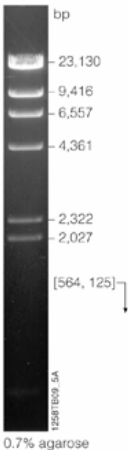
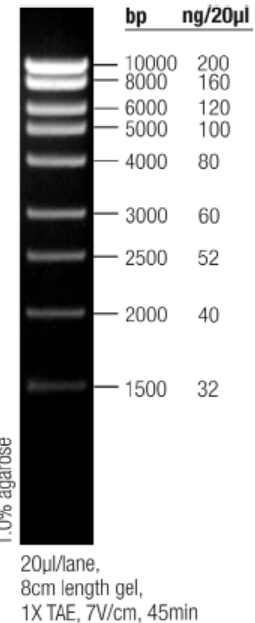
3.3 Enzymes

GenTherm™ DNA polymerase	Rapidozym
<i>Bam</i> HI	Promega
DNase I	Fermentas
<i>Hind</i> III	Promega
<i>Sal</i> I	Promega
T4 DNA ligase	Roche

3.4 Protein standards

Mark12™	Invitrogen	<table><tr><th rowspan="2">Protein</th><th colspan="2">Approx. Molecular Weights (kDa)</th></tr><tr><th>NuPAGE® MES</th><th>Tris-Glycine</th></tr><tr><td> Myosin</td><td>200</td><td>200</td></tr><tr><td>β-galactosidase</td><td>116.3</td><td>116.3</td></tr><tr><td>Phosphorylase B</td><td>97.4</td><td>97.4</td></tr><tr><td>BSA</td><td>66.3</td><td>66.3</td></tr><tr><td>Glutamic dehydrogenase</td><td>55.4</td><td>55.4</td></tr><tr><td>Lactate dehydrogenase</td><td>36.5</td><td>36.5</td></tr><tr><td>Carbonic anhydrase</td><td>31.0</td><td>31.0</td></tr><tr><td>Trypsin inhibitor</td><td>21.5</td><td>21.5</td></tr><tr><td>Lysozyme</td><td>14.4</td><td>14.4</td></tr><tr><td>Aprotinin</td><td>6.0</td><td>6.0</td></tr><tr><td>Insulin B chain</td><td>3.5</td><td rowspan="2">Unresolved Insulin</td></tr><tr><td>Insulin A chain</td><td>2.5</td></tr></table> <p>NuPAGE® 4-12% Bis-Tris Gel with MES stained with Coomassie® stain. 4-20% Tris-Glycine Gel stained with Coomassie® stain.</p>	Protein	Approx. Molecular Weights (kDa)		NuPAGE® MES	Tris-Glycine	 Myosin	200	200	β-galactosidase	116.3	116.3	Phosphorylase B	97.4	97.4	BSA	66.3	66.3	Glutamic dehydrogenase	55.4	55.4	Lactate dehydrogenase	36.5	36.5	Carbonic anhydrase	31.0	31.0	Trypsin inhibitor	21.5	21.5	Lysozyme	14.4	14.4	Aprotinin	6.0	6.0	Insulin B chain	3.5	Unresolved Insulin	Insulin A chain	2.5																															
Protein	Approx. Molecular Weights (kDa)																																																																								
	NuPAGE® MES	Tris-Glycine																																																																							
 Myosin	200	200																																																																							
β-galactosidase	116.3	116.3																																																																							
Phosphorylase B	97.4	97.4																																																																							
BSA	66.3	66.3																																																																							
Glutamic dehydrogenase	55.4	55.4																																																																							
Lactate dehydrogenase	36.5	36.5																																																																							
Carbonic anhydrase	31.0	31.0																																																																							
Trypsin inhibitor	21.5	21.5																																																																							
Lysozyme	14.4	14.4																																																																							
Aprotinin	6.0	6.0																																																																							
Insulin B chain	3.5	Unresolved Insulin																																																																							
Insulin A chain	2.5																																																																								
SeeBlue® Plus2	Invitrogen	<table><tr><th rowspan="2">Protein</th><th colspan="5">Approximate Molecular Weights (kDa)</th></tr><tr><th>Tris-Glycine</th><th>Tricine</th><th>NuPAGE® MES</th><th>NuPAGE® MOPS</th><th>NuPAGE® Tris-Acetate</th></tr><tr><td> Myosin</td><td>250</td><td>210</td><td>188</td><td>191</td><td>210</td></tr><tr><td>Phosphorylase</td><td>148</td><td>105</td><td>98</td><td>97</td><td>111</td></tr><tr><td>BSA</td><td>98</td><td>78</td><td>62</td><td>64</td><td>71</td></tr><tr><td>Glutamic Dehydrogenase</td><td>64</td><td>55</td><td>49</td><td>51</td><td>55</td></tr><tr><td>Alcohol Dehydrogenase</td><td>50</td><td>45</td><td>38</td><td>39</td><td>41</td></tr><tr><td>Carbonic Anhydrase</td><td>36</td><td>34</td><td>28</td><td>28</td><td>n/a</td></tr><tr><td>Myoglobin Red</td><td>22</td><td>17</td><td>17</td><td>19</td><td>n/a</td></tr><tr><td>Lysozyme</td><td>16</td><td>16</td><td>14</td><td>14</td><td>n/a</td></tr><tr><td>Aprotinin</td><td>6</td><td>7</td><td>6</td><td>n/a</td><td>n/a</td></tr><tr><td>Insulin, B Chain</td><td>4</td><td>4</td><td>3</td><td>n/a</td><td>n/a</td></tr></table> <p>NuPAGE® Novex Bis-Tris 4-12% Gel ©1999-2002 Invitrogen Corporation. All rights reserved.</p>	Protein	Approximate Molecular Weights (kDa)					Tris-Glycine	Tricine	NuPAGE® MES	NuPAGE® MOPS	NuPAGE® Tris-Acetate	 Myosin	250	210	188	191	210	Phosphorylase	148	105	98	97	111	BSA	98	78	62	64	71	Glutamic Dehydrogenase	64	55	49	51	55	Alcohol Dehydrogenase	50	45	38	39	41	Carbonic Anhydrase	36	34	28	28	n/a	Myoglobin Red	22	17	17	19	n/a	Lysozyme	16	16	14	14	n/a	Aprotinin	6	7	6	n/a	n/a	Insulin, B Chain	4	4	3	n/a	n/a
Protein	Approximate Molecular Weights (kDa)																																																																								
	Tris-Glycine	Tricine	NuPAGE® MES	NuPAGE® MOPS	NuPAGE® Tris-Acetate																																																																				
 Myosin	250	210	188	191	210																																																																				
Phosphorylase	148	105	98	97	111																																																																				
BSA	98	78	62	64	71																																																																				
Glutamic Dehydrogenase	64	55	49	51	55																																																																				
Alcohol Dehydrogenase	50	45	38	39	41																																																																				
Carbonic Anhydrase	36	34	28	28	n/a																																																																				
Myoglobin Red	22	17	17	19	n/a																																																																				
Lysozyme	16	16	14	14	n/a																																																																				
Aprotinin	6	7	6	n/a	n/a																																																																				
Insulin, B Chain	4	4	3	n/a	n/a																																																																				
Protein Molecular Weight Marker	Jena Bioscience	 <p>116.0 kDa 97.4 kDa 66.2 kDa 37.5 kDa 20.5 kDa 13.4 kDa 14.0 kDa</p>																																																																							

3.5 DNA standards

λ DNA/ <i>Hind</i> III	Promega	 <p>bp</p> <p>23,130 9,416 6,557 4,361 2,322 2,027 [564, 125]</p> <p>0.7% agarose</p>
Mass Ruler™ high range	Fermentas	 <p>bp ng/20µl</p> <p>10000 200 8000 160 6000 120 5000 100 4000 80 3000 60 2500 52 2000 40 1500 32</p> <p>1.0% agarose</p> <p>20µl/lane, 8cm length gel, 1X TAE, 7V/cm, 45min</p>

3.6 Chromatographic resins

HiTrap Heparin 5 mL column	Amersham Biosciences
HiTrap Chelating 5 mL column	Amersham Biosciences
Resource Q 6 mL column	Amersham Biosciences
GSTrap FF 1 mL column	Amersham Biosciences
Amylose resin	New England Biolabs

3.7 Instruments

Typhoon 9400 imaging system	Amersham Biosciences
Storage Phosphor GP Screen	Amersham Biosciences
Chromatography systems	ÄKTApurifier (Amersham Biosciences) ÄKTAprime (Amersham Biosciences) BioLogic (Bio-Rad)
Centrifuges	Centrifuge 5417C (Eppendorf) Microfuge R (Beckman) Sorvall RC-5B (DuPont) Sorvall Evolution RC
Rotors	GSA for Sorvall RC-5B SS-34 for Sorvall RC-5B SLC 4000 for Sorvall Evolution RC SA-300 for Sorvall Evolution RC SH-3000 Sorvall Evolution RC
PCR cyclers	T Gradient Thermoblock (Biometra) Primus (MWG Biotech) Primus 96 (MWG Biotech)
Concentrators	Amicon Ultra concentrators (Millipore) Ultrafree-MC concentrators (Millipore)
Filters	0.22 µm sterile filters (Osmonics) 0.45 µm sterile filters (Osmonics) 0.45 µm membrane filters (Millipore)
Centrivap concentrator	Labconco
Protein electrophoresis unit	Biorad
DNA electrophoresis unit	Biorad
Incubators	Unitron incubator (INFORS) Unimax 1010 shaker (Heidolph) Inkubator 1000 (Heidolph)
Crystallisation robot	Oryx6 (Douglas Instruments)
Crystallisation plates	Douglas Vapor Batch Plate (Douglas Instruments) CrystalClear Strips (Hampton Research)

DNA sequencer	ABI Prism 310 (Perkin-Elmer) Model S2 (Gibco)
Spectrophotometer	Cuvette test LKT photometer (Dr. Lange) DU-64 (Beckman) Spectronic® 20 Genesys™ (Spectronic Instruments)
CD spectrometer	Jasco J-715

3.8 Oligonucleotides

Oligonucleotides were purchased from MWG Biotech AG, Ebersberg, Germany, VBC-Genomics, Vienna, Austria and TIB Molbiol, Berlin, Germany.

Name	Sequence	Position
<i>traA_fr</i>	GAGGTGATAGGATCCGCAATCTT	1382–1404 ^A
<i>traA_rev</i>	TTTCATTTTAGGTACCTCTTGTTTTT	3383–3408 ^A
<i>traAN₂₆₈_rev</i>	GTTTTGGTCGACTTTTAATAC	2213–2233 ^A
<i>traAN₂₄₆_rev</i>	CATTGTAACTTTGAAGCTTTAAATT	2131–2162 ^A
<i>seq_pQE_fr</i>	CCCGAAAAGTGCCACCTG	3370–3887 ^C
<i>seq_pQE_rv</i>	GTTCTGAGGTCATTACTGG	225–243 ^C
42-mer_fr	CTAAGGGCGCACTTATACGCAGTAACTTCGTTACTTCGTATT	1256–1297 ^A
42-mer_rv	AATACGAAGTAACGAAGTTACTGCGTATAAGTGCGCCCTTAG	1256–1297 ^A
hairpin	ATACGAAGTAACGAAGTTACTGCGTAT	1270–1296 ^A
hairpin G>T	ATACGAAGTAACGAAGTTACTTCGTAT	1270–1296 ^A
hairpin A>C	ATACGCAGTAACGAAGTTACTGCGTAT	1270–1296 ^A
hairpin_nic	ATACGAAGTAACGAAGTTACTGCGTATAAGTGCG	1263–1296 ^A
prom_35_fr	TATTAAGGATTCCCTTAGATTATTTACTAA	1230–1259 ^A
prom_35_rv	TTAGTAAATAATCTAAGGGAATCCTTAATA	1230–1259 ^A
prom_10_fr	GCACTTATACGCAGTAACTT	1264–1283 ^A
prom_10_rv	AAGTTACTGCGTATAAGTGC	1264–1283 ^A
prom_fr	ATTCCCTTAGATTATTTACTAAGGGCGCACTTATACG	1238–1274 ^A
prom_rv	CGTATAAGTGCGCCCTTAGTAAATAATCTAAGGGAAT	1238–1274 ^A
42_random	AGAATGTTGAACGACTAACTAAAAATAGACTAGCGGTTGAAT	2179–2220 ^B

FP_forward	TTCCATGAAATACGCCTCCT	1097–1116 ^A
FP_reverse	CAAACACGCTTGACGATAAAA	1326–1346 ^A

^A – pIP501 (Acc. No. L39769)

^B – pIP501 (Acc. No. AJ301605)

^C – pQE30 (QIAGEN)

The restriction enzyme recognition sequences are underlined.

The exchanged bases are bold.

3.9 Plasmids and vectors

Name	Size [kb]	Description/Resistance	Reference
pIP501	30.6	<i>tra</i> ⁺ , Cm ^R , MLS ^R	(Evans and Macrina, 1983)
pQE30	3.4	<i>Ptac</i> ::6xHis, Amp ^R	QIAGEN
pQTEV	4.8	<i>Ptac</i> ::7xHis, <i>lacI</i> ^q , Amp ^R	gift from K. Büssow Max-Planck-Institut für Molekulare Genetik Proteinstrukturfabrik Heubnerweg 6 14059 Berlin Germany http://www.molgen.mpg.de/~buessow
pVA2241	6.9	pDL277 encompassing 309 bp <i>oriT</i> _{pIP501} , <i>tra</i> ⁺ , Sp ^R	(Wang and Macrina, 1995a)
pDL277	6.6	<i>tra</i> ⁺ , Sp ^R	(LeBlanc <i>et al.</i> , 1992)
RSF1010	8.7	Sm ^R , mob ⁺	(Scherzinger <i>et al.</i> , 1992)

3.10 Microorganisms

<i>Escherichia coli</i>		
XL1 blue	<i>recA1</i> , <i>endA1</i> , <i>gyrA96</i> , <i>thi-1</i> , <i>hsdR17</i> , <i>supE44</i> , <i>relA1</i> , <i>lac</i> [F ⁻ <i>proAB lacI</i> ^q Z, M15 Tn10 (Tc ^R)]	Stratagene
BL21(DE3)	F ⁻ , <i>ompT hsdS</i> _B (r _B ⁻ m _B ⁻), <i>gal</i> , <i>dcm</i> , (DE3)	Novagen
BL21(DE3) CodonPlusRIL	F ⁻ <i>ompT hsdS</i> (r _B ⁻ m _B ⁻) <i>dcm</i> ⁺ Tet ^r <i>gal</i> λ (DE3) <i>endA</i> Hte [<i>argU ileY leuW</i> Cam ^r]	Stratagene
<i>Enterococcus faecalis</i>		
JH2-2	Rif ^R , Fus ^R	(Jacob and Hobbs, 1974)

3.11 PCR programs

traA and *traAN*₂₆₈

94°C 2 min

94°C 30 sec
50°C 2 min
72°C 2 min

 30x
 72°C 7 min

Footprint

94°C 2 min

94°C 30 sec
53°C 45 sec
72°C 30 sec

 30x
 72°C 7 min

*traAN*₂₄₆

94°C 2 min

94°C 30 sec
49°C 1 min 45 sec
72°C 1 min 30 sec

 30x
 72°C 7 min

Insert check

94°C 2 min

94°C 30 sec
52°C 2 min
72°C 2 min

 30x
 72°C 7 min

3.12 Buffers

TAE	40 mM Tris/acetate 1 mM EDTA pH 8.0	TBE	50 mM Tris/borate 1 mM EDTA pH 8.0
TE (10:1)	10 mM Tris/HCl 1 mM EDTA pH 8.0		
Running buffer (SDS-PAGE)	200 mM glycine 25 mM Tris base 0.2% SDS pH=8.3	2x Protein loading buffer	12.5 mM Tris/HCl 2% (w/v) SDS 5% (v/v) β-mercaptoethanol 20 % (v/v) glycerol 0.001% (w/v) bromophenyl blue pH 6.8

3.13 Media

LB	10 g/L trypton 5 g/L yeast extract 10 g/L NaCl	2xYT	18 g/L trypton 8 g/L yeast extract 5 g/L NaCl
----	--	------	---

For solid media, 1.5 % w/v agar was added.

4 Methods

4.1 DNA methods

4.1.1 Preparative plasmid isolation

Plasmids were isolated using the Qiagen Plasmid DNA Maxi kit, according to the protocol provided by the manufacturer. An 8-hour culture was diluted 1:500 to 100 mL selective LB medium and incubated overnight at 37°C with vigorous shaking. The cells were harvested at 6000 *g* for 15 minutes at 4°C. The pellet was suspended in 10 mL buffer P1 (50 mM Tris/HCl, 10 mM EDTA, 100 µg/mL RNase, pH 8.0). 10 mL buffer P2 (200 mM NaOH, 1% w/v SDS) were added and the suspension was mixed by inverting 6 times and incubated at room temperature for 15 minutes. 10 mL chilled buffer P3 (3 M potassium acetate pH 5.5) were added, and mixed by 6 times inverting. The lysate was poured onto a QIAfilter cartridge and incubated for 15 minutes at room temperature. The QIAGEN-tip 500 was equilibrated with 10 mL QBT (750 mM NaCl, 50 mM MOPS, 15 % isopropanol v/v, 0.15 % Triton x-100 v/v, pH. 7.0) buffer, the column was emptied by gravity flow. The lysate was filtered through the QIAfilter cartridge and loaded onto the QIAGEN-tip. The column was washed twice with 30 mL buffer QC (1 M NaCl, 50 mM MOPS, 15 % isopropanol v/v, pH 7.0). The DNA was eluted with 15 mL buffer QF (1.25 M NaCl, 50 mM Tris/HCl, 15 % isopropanol v/v, pH 8.5) and precipitated with 10.5 mL room temperature isopropanol. The suspension was centrifuged at 15 000 *g* for 30 minutes at 4°C. The supernatant was decanted and the pellet was washed with 5 mL 70 % ethanol and centrifuged at 15 000 *g* for 30 minutes at 4°C. The pellet was dried under vacuum for 20 minutes, and then dissolved in 200 µL sterile water. The quality of the plasmid DNA was analyzed by agarose gel electrophoresis.

4.1.2 Cloning

4.1.2.1 Cloning of wild type *traA*

The *traA* fragment of plasmid pIP501 (GenBank® accession no. L39769) was amplified by PCR using GenTherm™ DNA polymerase and specific primers *traA*_fr and *traA*_rev with the *traA* PCR program. The PCR product was purified using the *Wizard® SV Gel and PCR Clean-Up* kit and cut with the restriction enzymes *Bam*HI and *Kpn*I. Expression vector pQE30, which is designed to express proteins with an N-terminal 6×His tag, was also cut with *Bam*HI and *Kpn*I. After restriction, plasmid DNA was loaded onto a 0.7 % low melting agarose gel

and run for 2 hours at 5 V/cm in TAE buffer. The plasmid band was cut out from the gel, and plasmid DNA was eluted using the *Wizard® SV Gel and PCR Clean-Up* kit, according to the protocol provided by supplier. The ligation mixture, consisting of a 1:1 molar ratio of insert to plasmid, was incubated at 4°C overnight with T4 DNA ligase and used for transformation of *Escherichia coli* XL1Blue cells. The transformants were first screened by PCR. PCR positive colonies were plated on selective LB plates, and plasmid DNA was isolated using TELT method, see 4.1.3. Plasmid DNA mini-prep. Plasmid DNA was cut with 0.02 U *Bam*HI and *Hind*III for 1 hour, the lengths of the resulting fragments were analyzed on a 1% agarose gel. If the correct length was obtained (2.0 kbp), then the plasmid DNA was isolated using the GenElute™ Plasmid Miniprep kit. The sequence of both strands of the recombinant plasmid pQE30-*traA* was verified using a BigDye™ Terminator Cycle Sequencing Kit and an ABI Prism 310 automated sequencer.

4.1.2.2 Cloning of *traAN*₂₆₈

A 851-bp *traAN*₂₆₈ fragment encoding the first 268 amino acids of TraA was PCR-amplified using *traA_fr* forward primer and *traAN*_{268_rev} reverse primer with the *traAN*₂₆₈ PCR program. The PCR product was cloned into the unique *Bam*HI and *Sal*I sites of pQE30. All following cloning steps were performed as described for *traA* cloning. The correct length of the *traAN*₂₆₈ fragment is 1247 bp.

4.1.2.3 Cloning of *traAN*₂₄₆

A 780 bp *traAN*₂₄₆ fragment encoding the first 246 amino acids of TraA was PCR-amplified using *traA_fr* forward primer and *traAN*_{246_rev} reverse primer with the *traAN*₂₄₆ program. The PCR product was cloned into the unique *Bam*HI and *Hind*III sites of pQE30. All following cloning steps were performed as described for *traA* cloning. The correct length of the *traAN*₂₄₆ fragment is 1176 bp.

All three constructs were also inserted into the pQTEV expression vector using the same PCR conditions and restriction enzymes. The *E. coli* BL21-CodonPlus®(DE3) RIL cells were transformed with the expression vectors.

4.1.3 Plasmid DNA mini-prep

Mini plasmid DNA preparations were performed using the TELT method (Medina-Acosta and Cross, 1993). Plasmid harboring *E. coli* cells were scraped off from a selective LB plate and suspended in 1 mL sterile 0.9 % NaCl. The cells were centrifuged for 5 minutes at 5 000 *g*. The cell pellet was resuspended in 400 µL TELT buffer (50 mM Tris/HCl pH 7.5, 62.5 mM EDTA, 2.5 M LiCl, 0.4 % Triton X-100). Lysozyme was added to a final concentration of 2.5 mg/mL. Lysis continued for 10 minutes at room temperature. The suspension was heated to 95°C for 90 seconds and centrifuged at 12 000 *g* for 15 minutes. The pellet was removed with a sterile toothpick, and the plasmid DNA was precipitated with 250 µL isopropanol. Plasmid DNA was collected by centrifugation at 14 000 *g* for 15 minutes at 4°C. The DNA pellet was washed with 80 % ethanol, centrifuged again, dried under vacuum and dissolved in 50 µL sterile water.

4.1.4 Isolation of plasmid RSF 1010

1 L LB medium with 50 µg/mL streptomycin was inoculated with 10 mL of an overnight culture of *E. coli* DH5α RSF1010 and incubated until OD₆₀₀ reached 1. The cells were harvested at 6 000 *g* for 15 minutes at 4°C and suspended in 60 mL solution I (5 mM sucrose, 10 mM EDTA, 25 mM Tris, pH 8.0). 120 mL solution II (0.2 M NaOH, 1% (w/v) SDS) were added and after addition of 100 mL solution III (3 M sodium acetate, pH 4.8), an incubating step for 30 minutes on ice followed (Birnboim, 1983). The solution was centrifuged at 16 000 *g* for 30 minutes at 4°C and the supernatant was dialysed against 0.5 M potassium acetate pH 5.5. The dialysate was centrifuged at 16 000 *g* for 30 minutes at 4°C to remove any particulate material and loaded onto a Resource Q 6 mL anion exchange column equilibrated with 0.5 M potassium acetate pH 5.5, mounted on the ÄKTAprime purification system. RNA was eluted in 3 column volumes (CV) of 25 mM Tris-HCl, 0.75 M NaCl, 1 mM EDTA pH 8.0 and 5 CV of 25 mM Tris-HCl, 0.5 M NaCl, 1 mM EDTA pH 8.0. Plasmid DNA was eluted by applying 1 CV of elution buffer (50 mM phosphate, 1 M NaCl pH 10), the process was paused for 5 minutes and the elution was continued for further 4 CV. The column was cleaned with 5 CV of cleaning buffer (0.5 NaOH, 2 M NaCl). The elution was performed at 0.2 mL/min, whereas all other chromatographic steps were done at 4 mL/min. Absorbance at 280 nm and conductivity were monitored during the purification. The elution peak was pooled, dialysed against sterile bidistilled (milliQ) water and concentrated in Centrивap concentrator (speedvac) (Bhikhabhai *et al.*, 2000). The pellet was dissolved in 200 µL sterile

water. 10 μ L were loaded onto a 0.7 % agarose gel and run at 5 V/cm for 2 hours in TAE buffer at 4°C. The whole isolation procedure of RSF 1010 plasmid DNA was carried out at 4°C.

4.1.5 Preparation of hairpin_*nic* oligonucleotide for crystallization trials

436.2 nmol oligonucleotide were dissolved in 100 μ L milliQ water. 800 μ L of 10 mM NaOH, 100 mM NaCl were added to the solution. The sample was loaded onto a 6 mL ResourceQ column equilibrated with 90 % buffer A (10 mM NaOH) +10 % buffer B (10 mM NaOH, 1 M NaCl). 1 mL fractions were collected in tubes containing 50 μ L 1 M HEPES pH 7.5. The peak fractions were pooled and dialyzed twice against 3 L milliQ water. The oligonucleotide was vacuum-dried and the pellet was dissolved in 150 μ L milliQ water. The hairpin was annealed by heating to 95°C for 5 minutes and fast cooling on ice.

4.2 Microbiological methods

4.2.1 Preparation of *E. faecalis* JH2-2 (pIP501) cell lysate

A single *E. faecalis* JH2-2 (pIP501) colony was lysed in 20 μ L 50 mM NaOH and 0.25% SDS. The suspension was heated to 95°C for 15 minutes and diluted with 180 μ L sterile water. For each PCR reaction, 1 μ L of the lysate was used.

4.2.2 Preparation and transformation of *E. coli* BL21(DE3) competent cells

Competent cells were prepared using the RotiTransform kit, following the protocol provided by the supplier. 50 mL LB medium were inoculated with 2 mL overnight culture and grown at 37°C with shaking until OD₆₀₀ reached 0.4. 2x 10 mL culture were centrifuged at 3000 g for 5 minutes at 4°C, each pellet was suspended in 500 μ L cold LB medium and transferred to a microcentrifuge tube. 450 μ L ice-cold Roti-transform 1 were added to each tube and the suspension was gently vortexed. 50 μ L Roti-transform 2 were added to each tube and the suspension was vortexed again. 180 μ L competent cells each were aliquoted into ice cold microcentrifuge tubes. The tubes were immersed in liquid nitrogen; the competent bacteria were kept frozen at -80°C. To transform the cells, 1 μ L plasmid DNA was added to an aliquot of competent cells, gently mixed and incubated on ice for 30 minutes. The suspension was plated onto selective LB agar plate and incubated overnight at 37°C.

4.2.3 Preparation of *E. coli* XL1Blue competent cells

Competent cells were prepared with the RbCl method (Tang *et al.*, 1994).

3 mL overnight culture of *E.coli* XL1blue cells were added to 250 mL LB medium and grown at 37°C with shaking until OD₆₀₀ reached 0.4-0.6. During subsequent steps, all instruments used were chilled on ice. The cells were harvested at 5 000 *g* for 5 minutes. The pellet was suspended in 100 mL cold TFB I buffer (30 mM potassium acetate, 100 mM RbCl, 10 mM CaCl₂, 50 mM MgCl₂, 15% glycerol, pH 5.8) and kept on ice for 5 minutes. After centrifugation at 5 000 *g* for 5 minutes, the pellet was resuspended in 2 mL TFB II buffer (10 mM MOPS, 75 mM CaCl₂, 10 mM RbCl, 15 % glycerol, pH 6.5) and left on ice for 30 minutes. The suspension was divided into 100 µL aliquots, frozen in liquid nitrogen and stored at -80°C.

4.2.4 Transformation of *E. coli* XL1Blue competent cells

100 µL competent cells were thawed on ice. 1-2 µL of plasmid DNA (~10 ng) or the ligation reaction were added to the thawed cells and left on ice for another 20 minutes. The cells were heat shocked for exactly 90 seconds and immediately put on ice for 5 minutes. 900 µL room temperature LB medium were added to the cells. The cells were vigorously shaken at 37°C for one hour. 5 x 200 µL were plated on selective LB agar plates and incubated overnight at 37°C.

4.3 Screening of transformants

4.3.1 PCR test

The transformants were picked with a sterile toothpick from transformation plate and transferred in a PCR tube, containing 5 µL water. 5 µL of 2x PCR reaction mixture containing seq_pQE_fr and seq_pQE_rev primers were added to the PCR tube. The “insert check” program was used for amplification of the DNA fragment. The whole PCR reaction was loaded onto a 1 % agarose gel and run for 1 hour at 7 V/cm in TAE buffer. The gel was stained with ethidium bromide (0.5 µg/mL) and analysed under UV-light. If the length of the resulting DNA fragment was correct, the clone was further checked by plasmid DNA digestion with the appropriate restriction enzymes.

4.3.2 Restriction analysis

20 µL plasmid DNA from TELT isolation (see chapter 4.1.3. Plasmid DNA mini-prep) were treated with 0.2 U *Bam*HI and *Hind*III for 2 hours at 37°C. 10 µl were loaded onto a 1 % agarose gel and run for 90 minutes at 8 V/cm in TAE buffer. If the resulting length of the DNA insert was correct, the clone was stored in the strain collection. This strain was used for protein expression.

4.3.3 DNA labeling

Synthetic oligonucleotides (HPLC-grade) were purchased from VBC-GENOMICS (Vienna, Austria). 10 pmol of each oligonucleotide were labeled with 3 µL of [γ^{32} P]ATP (1.11×10^{14} Bq/mmol) with 10 U T4 polynucleotide kinase (Roche Diagnostics, Mannheim, Germany) for 30 minutes at 37°C. The kinase was inactivated by heating to 95°C for 5 minutes. Unbound [γ^{32} P] ATP was removed by a Sephadex G-50 (Amersham Biosciences, Uppsala, Sweden) column: 10 g Sephadex G-50 were suspended in 300 mL TE (10:1) and autoclaved. The supernatant was removed and the beads were washed three times with sterile water. Finally, the beads were suspended in 300 mL TE (10:1). 4 mL of the suspension was poured into a glass Pasteur pipette filled with sterile glass wool. The tip was plugged with parafilm until the beads had settled, then the column was equilibrated with 10 mL TE (10:1). The resin was not allowed to dry. 10 µL of dextran blue (1mg/mL) were added to the labeling reaction and loaded onto the Sephadex column. The oligonucleotides were eluted with TE (10:1) buffer in 100 µL fractions. Each fraction was measured with a Geiger counter. The oligonucleotides and dextran blue elute at exclusion volume of the column. The blue fractions with the highest radioactivity were taken for the experiments.

4.4 Biochemical methods

4.4.1 TraA, TraAN₂₆₈ and TraAN₂₄₆ expression

E. coli XL1 Blue (pQE30-*traA*), *E. coli* XL1 Blue (pQE30-*traAN*₂₆₈) and *E. coli* XL1 Blue (pQE30 - *traAN*₂₄₆) cells were inoculated into LB broth containing 100 mg/L ampicillin and 40 mg/L tetracycline and incubated with shaking overnight at 37°C. The overnight cultures were diluted 100-fold with LB broth containing the respective antibiotics, and the cultures were shaken at 37°C until a OD₆₀₀ of 0.5–0.6 was reached. Transcription of the 6×His-tag fusion proteins from the T5 promoter was induced by addition of

isopropyl β -D-thiogalactopyranoside (IPTG) to a final concentration of 1 mM. Incubation continued for 4 h. The cells were harvested by centrifugation at 6 000 *g* at 4°C for 15 min.

E. coli BL21-CodonPlus[®] (DE3) RIL (pQTEV-*traA*), *E. coli* BL21-CodonPlus[®] (DE3) RIL (pQTEV- *traAN*₂₆₈) and *E. coli* BL21-CodonPlus[®] (DE3) RIL (pQTEV- *traAN*₂₄₆) cells were inoculated into LB broth containing 100 mg/L ampicillin and 50 mg/L chloramphenicol and incubated with shaking overnight at 37°C. The overnight cultures were diluted 100-fold with selective LB medium and the cultures were shaken at 37°C until a OD₆₀₀ of 0.5–0.6 was reached. Protein expression was induced by addition of 1 mM IPTG. After 4 hours of induction, the cells were harvested by centrifugation at 6 000 *g* at 4°C for 15 min.

4.4.2 TraA, TraAN₂₆₈ and TraAN₂₄₆ purification

The cell pellets were suspended in BugBuster[®] HT buffer following the instructions provided by the supplier. The clarified supernatant was adjusted to 5 mM imidazole, 50 mM Tris HCl pH 8.0 and 0.5 M NaCl and loaded onto a Ni²⁺ charged HiTrap Chelating column to bind the His-tagged protein. The column was washed with several volumes of binding buffer (50 mM Tris HCl pH 8.0, 0.5 M NaCl, 5 mM imidazole) and finally TraA, TraAN₂₆₈ or TraAN₂₄₆ were eluted with an imidazole gradient (50 mM Tris HCl pH 6.0, 0.5 M NaCl, 250 mM imidazole). The fractions containing the desired protein were dialyzed against a buffer consisting of 20 mM Tris HCl pH 8.0, 300 mM NaCl, 1 mM EDTA and 0.5 mM DTT. This buffer was used as loading buffer for purification on a HiTrap Heparin column. The target protein was eluted with a NaCl gradient (20 mM Tris HCl pH 8.0, 1 M NaCl, 1 mM EDTA and 0.5 mM DTT). Samples were taken at the different purification steps and loaded onto SDS polyacrylamide gels to follow the purification process. The 7xHis-TEV- tagged constructs, expressed from pQTEV vector, were purified the same way.

TraA was concentrated to 12 mg/ml, TraAN₂₄₆ to 6 mg/ml and TraAN₂₆₈ to 5 mg/ml. The concentration of the proteins was determined by UV₂₈₀ and by the BCA Protein Assay procedure using bovine serum albumin as a standard. The protein solutions were diluted for further experiments.

4.4.3 Orf7 constructs expression and purification

4.4.3.1 GST-Orf7 Δ TM

Expression of GST tagged Orf7 lacking the transmembrane helix (GST-Orf7 Δ TM) in 1 L LB medium was induced by 1 mM IPTG when OD₆₀₀ reached 0.6. After 4 hours of expression, cells were harvested at 6 000 g. The cell pellet was suspended in 40 mL 20 mM phosphate buffer containing 200 mM NaCl, 1 mM EDTA and 0.5 mM DTT, pH 6.8. The cells were lysed by lysozyme added to the suspension to 0.5 mg/mL. After 3 hours of incubation on ice, the suspension was homogenized using a 20 mL syringe and 0.7 mm needle to shear the DNA. The suspension was centrifuged at 16 000 g at 4°C for 20 minutes. The supernatant was loaded onto 1 mL GStrap FF column; the protein was eluted with 10 mM glutathione.

4.4.3.2 7His-Orf7

7His-Orf7 expression was performed in 2xYT containing 0.5% glucose and 100 µg/mL ampicillin. The culture was shaken at 37°C until OD₆₀₀ reached 1. At this point IPTG was added to a final concentration of 1 mM. Two hours later the cells were harvested, yielding 5 g wet cells per 1 L of culture.

The cell pellet was suspended in 20 mL B-PER bacterial cells extraction solution, incubated at 20°C for 20 minutes and finally centrifuged at 20 000 g for 20 min at 4°C. The supernatant was loaded onto a Ni²⁺ charged HiTrap Chelating column equilibrated with buffer A (20 mM phosphate, 10 mM imidazole, 300 mM NaCl, 0.1% Triton-X100, pH 8.0). The protein was eluted with an imidazole gradient (20 mM phosphate, 250 mM imidazole, 300 mM NaCl, 0.1% Triton-X100, pH 8.0)

4.4.3.3 MBP-Orf7 Δ TM

E. coli XL10gold pMAL-*orf7* Δ TM culture was grown in LB at 37°C until OD₆₀₀ reached 0.6. IPTG was added to 0.5 mM and the expression was carried out for 2 hours. The cells were harvested at 6 000 g for 10 minutes at 4°C.

Cell pellet was resuspended in lysis buffer (20 mM Tris, 200 mM NaCl, 1 mM EDTA, 1 % Triton X-100, pH 7.0). Lysozyme was added to a final concentration of 1 mg/mL. The suspension was sonicated in a sonic water bath for 30 min and centrifuged at 30 000 g for 30 minutes at 4°C. The supernatant was applied to 1.5 mL amylose resin equilibrated with lysis buffer and shaken at 4°C for 2 hours. The beads were centrifuged for 2 min at 3 000 g, and the supernatant was discarded. The beads were washed twice with wash buffer

(20 mM Tris, 200 mM NaCl, 1 mM EDTA, 0.1 % Triton X-100, pH 7.0) and the protein was eluted with 10 mM maltose.

4.4.4 Zymogram analysis

Analysis of peptidoglycan degradation was conducted as previously described (Zahrl *et al.*, 2005) with minor modifications. 0.05 % peptidoglycan purified from *E. faecalis* JH2-2 was incorporated into a 15 % polyacrylamide gel containing 0.02 % SDS. The samples were diluted with sample buffer (50 mM Tris/HCl pH 8.0, 0.04 % SDS, 10 % glycerol, 0.01 % bromophenol blue). The gel was run in running buffer (25 mM Tris, 20 mM glycine, 0.02 % SDS, pH 8.0) at 10 W until the bromophenolblue band reached the end of the gel. The gel was washed three times with water and transferred to refolding buffer (20 mM Tris-malonate pH 6.8, 1mM MgCl₂, 0.1 % Triton X-100) and incubated overnight at 4°C.

The gel was stained with 0.1% methylene blue in 0.01% KOH for 1 hour at 4°C and destained with water for 1 hour and photographed against sun light.

4.4.5 Crystallization trials

4.4.5.1 Vapor diffusion

The proteins were purified to ~90 %, concentrated to 12 mg/ml (TraA) 6 mg/mL (TraAN₂₆₈) and 7 mg/mL (TraAN₂₄₆) and dialyzed against 20 mM Tris/HCl, 300 mM NaCl, 1 mM EDTA. Hampton Research Crystal Screen and Crystal Screen 2 were set up as sitting drops in CrystalClear Strips. TraA was also set up with hairpin_*nic* oligonucleotide in 1:1.2 molar ratio using Hampton Research NATRIX screen. The TraA protein (7 mg/mL) was mixed with the oligonucleotide and the complex was formed by heating the mixture to 42°C and slowly cooling down to room temperature.

The 2 µL drops consisted of 50% protein or protein:DNA solution and 50% screening solution. The plates were incubated at 20°C and 4°C.

4.4.5.2 Microbatch

For microbatch crystallization setups, 500 nL drops were made using the Oryx crystallization robot and the screening solutions used in vapor diffusion experiments. Additionally the Index screen (Hampton Research) was used. The drops were formed in a Douglas Vapor Batch Plate and covered with 3:1 v/v paraffin oil : silicone oil and incubated at 20°C.

4.5 Biophysical methods

4.5.1 Circular Dichroism Spectroscopy (CD)

CD spectroscopy was used to estimate secondary structure contents of the protein. It measures the difference in absorbance of left- (A_l) and right- (A_r) circularly polarized light,

$$\Delta A = A_l - A_r = \varepsilon_l Cl - \varepsilon_r Cl = \Delta \varepsilon Cl$$

where $\varepsilon_{l,r}$ is the decadic molar extinction coefficient of the solute for l or r circularly polarized light, C is the molar concentration of the chiral substance and l is the path length of the light passing through the medium. The spectropolarimeter measures $\Delta \varepsilon$, however another measure of CD is used in biochemical literature, namely ellipticity Θ , which is calculated as:

$$[\Theta] = 3298 \Delta \varepsilon$$

Usually the mean residue concentration for the far-UV CD of polypeptides and proteins is used. For the near-UV region, the protein molarity is often reported (Woody, 1996). Mean residue concentration was used throughout this work.

CD measurements were performed on a Jasco J-715 spectropolarimeter using 0.02 cm and 0.1cm path length cylindrical cells. Far-UV CD spectra were recorded from 260 to 195 nm. Each spectrum was recorded as an average of three scans taken with the following parameters: Step resolution 0.5 nm, speed 50 nm/min, response 1 sec and bandwidth 1 nm. The resulting spectra were baseline corrected by subtracting the buffer signal. The protein samples were prepared in 50 mM Tris HCl buffer pH 7.5 containing 250 mM NaCl. Protein concentrations used were 4.61×10^{-5} M, 2.02×10^{-5} M, and 5.7×10^{-5} M for TraA, TraAN₂₄₆ and the TraAN₂₄₆-DNA complex (1:1), respectively. As DNA target, the annealed 42-mer oligonucleotide containing the inverted repeat structure, the *nic* and additional 7 bases was used. The CD signal was converted to mean residue ellipticity $[\Theta]$.

4.5.2 Secondary structure determination and thermal stability

CD is one of the most sensitive techniques for determination of secondary structures of proteins as well as monitoring structural changes upon denaturation or binding a ligand. By looking at a CD spectrum one can determine the α helical and β strand contents of the protein. The spectra exhibit specific CD bands; all α -helix protein spectrum shows minima at 222 and 208 nm and a maximum at 190 nm, whereas an all- β strand protein spectrum has a minimum at 215 nm and a maximum at 195 nm. The calculation of secondary structures gets more complicated, if the protein consists of α -helices and β -strands in different proportions. Much

work has been done to design algorithms, which calculate the secondary structure fractions. The algorithms range from simple, where the CD spectrum is considered as a linear combination of basis spectra, recorded for short, synthetic polypeptides, known to form an α -helix or β -strand, to more sophisticated, which use CD spectra of proteins with known 3D structures as references (Venyaninov and Yang, 1996).

The secondary structure content of the three protein constructs was calculated using K2d algorithm available at <http://www.embl-heidelberg.de/~andrade/k2d.html> (Andrade *et al.*, 1993; Merelo *et al.*, 1994) and in Dichroprot 2000 program (Deleage and Geourjon, 1993). CD data obtained at 20°C were used as input for the program. Thermal denaturation curves were collected using a 0.02 cm water-jacket cylindrical cell thermostated by an external computer-controlled water bath. The data were recorded in the range of 20 to 60°C at 208 nm by a step scan procedure: heating rate 20°C/hour, response 1 sec, recorded interval 0.2°C. The thermal denaturation curves were fitted with a sigmoidal function and T_M was determined from the point of inflection using Origin™ 5.0 (Microcal™, Northampton, MA, USA). The temperature scans were normalized according to their initial CD₂₀₈ signal, as determined from the least squares fit and the curves were smoothed using a moving window average algorithm with a window size of 10 points.

4.5.3 Cross-linking experiments of TraA and TraAN₂₄₆

The oligomerization state of proteins can be determined using cross-linkers. They make a covalent bond between two molecules. The length of a cross-linker is an important factor. Proteins, which interact more closely, can be cross-linked by a shorter cross-linker, whereas longer cross-linker has to be used to confirm long distance relationship between the two molecules. If the protein molecules are close together, then the cross-linker can trap the complex by a covalent bond. The oligomerization state can be followed by running the complexes on an SDS-PA gel.

The easiest and cheapest way to cross-link proteins is to use glutaraldehyde, it is also used for immobilization of proteins, low concentrations of that compound preserve activity of the immobilized proteins (Migneault *et al.*, 2004).

Although the molar ratio between glutaraldehyde and protein during protein-protein cross-linking is high (1000:1), the method is mild enough not to destroy the 3D structure of the protein. Glutaraldehyde is sometimes used to stabilize protein crystals used for X-ray structure determination (Lusty, 1999; Reményi *et al.*, 2001).

Cross-linking of the relaxase was performed as previously described (Verdino *et al.*, 1999) with minor modifications. The reaction volume of 50 µl consisted of 0.5 mg/ml protein, 100 mM Bicine pH 7.5, 300 mM NaCl, 1 mM DTT and respective concentration of glutaraldehyde (0.001 %, 0.01 %, 0.05 % and 0.1 % v/v). The reaction was stopped after 15 minutes by adding 1 M glycine pH 8.0 to a final concentration of 140 mM. The samples were incubated for another 5 minutes. The proteins were precipitated with 400 µl cold acetone for 2 hours at -20°C and centrifuged at 15000 g for 15 minutes at room temperature. Prior to loading onto an 18 % SDS polyacrylamide gel, the pellets were dissolved in loading buffer (100 mM Tris-HCl pH 6.8, 4 % w/v SDS, 0.2 % w/v bromophenol blue, 20 % v/v glycerol, 200 mM β-mercaptoethanol) and heated to 100°C for 5 minutes.

4.5.4 Electrophoretic Mobility Shift Assays (EMSA)

Electrophoretic mobility shift assay allows studying DNA-protein interactions. The protein-DNA complex is loaded onto a native polyacrylamide gel along with DNA alone (as a control) and if the reaction conditions permitted the binding, then the complex is easily distinguished from the unbound DNA, because the mobility of the complex is reduced in comparison with the DNA alone.

In this study [³²P]-labeled oligonucleotides were used as target DNA. The radioactively labeled oligonucleotides were annealed in the following way: oligonucleotides were diluted in TE buffer (10 mM Tris pH 8.0, 1 mM EDTA), completely denatured (5 min at 95°C) and annealed by quick cooling on ice. The annealed oligonucleotides were applied immediately to the EMSA or stored at 4°C.

To obtain double stranded substrate for the experiment, the forward strand was radioactively labeled and mixed in a 1:1 molar ratio with complementary unlabeled oligonucleotide in TE 10/1 buffer (10 mM Tris, pH 8.0, 1 mM EDTA). The mixture was heated to 95°C for 5 minutes and annealed slowly by cooling down in the heat block to room temperature.

20 µl binding mixtures contained 10 fmol (0.5 nM) radiolabeled oligonucleotides and increasing TraA or TraAN₂₄₆ concentrations in 20 mM Tris HCl pH 7.5, 0.1 mM EDTA and 200 mM NaCl. EMSA experiments with hairpin and hairpin_{nic} oligonucleotides were performed in the presence of 1 mg/ml bovine serum albumin. All oligonucleotides used in the EMSAs are shown in Figure 24. After 30 minutes of incubation at 42°C the samples were loaded on a 10 % native polyacrylamide gel and run 90 minutes at constant voltage of 80 V in ice cold 22.5 mM Tris-borate, 0.5 mM EDTA buffer (pH 8.0). The amounts of bound and unbound oligonucleotides were determined with the liquid scintillation counter (Liquid

Scintillation Analyser Tri-Carb 1900 CA, Packard Instrument, Groningen, The Netherlands) and/or AlphaEaseFC™ software (Biozym Scientific, Hess Oldendorf, Germany). For scintillation counting, each lane was cut out of the gel, and cut into 16 pieces. Each piece was solubilized with 0.5 ml Soluene-350 and incubated at 50°C for 3 hours. 10 ml Hionic-Fluor were added to each vial. The samples were vortexed and allowed to adapt to room temperature before counting. The dissociation constant (K_D) of the DNA-protein complex was estimated to be equal to the protein concentration at which half of the oligonucleotide was bound. As a negative control a randomly chosen 42-mer 42_random oligonucleotide with no sequence identity with the pIP501 *oriT* encompassing nucleotides 2179 through 2220 on pIP501 (Acc. No. AJ301605) - was used. To test binding to IncP group (RP4) *oriT*, the oligonucleotide *oriT* 2: 5'-CAGCCGGGCAGGATAGGTGAAGT-3 (Gotz *et al.*, 1996), for binding of pMV158-type *oriT* the oligonucleotide 5'-ATAAAGTATAGTGTGTTATACTTTAT-3' (nucleotide positions 3582 to 3607 GenBank Acc. No. X15669, the last T was exchanged for a C) was used. Competition assays were done in the same way as described for other EMSA experiments: unlabeled 42-mer as specific competitor or the negative control 42-mer oligonucleotide as unspecific competitor were used. The polyacrylamide gels were exposed to Hyperfilm™-βmax at 4°C overnight.

4.5.5 EMSA with dimeric forms of TraA and TraAN₂₄₆

The 42-mer oligonucleotide comprising nick site and inverted repeat structure was also used in EMSA with dimeric forms of TraA and TraAN₂₄₆. The optimal glutaraldehyde concentrations to form protein dimers were obtained from cross-linking experiments. When 0.5 mg/mL TraA was incubated with 0.05 % glutaraldehyde for 15 minutes, ca. 70 % of the protein preparation formed dimers, whereas 0.5 mg/mL TraAN₂₄₆ had to be incubated with 0.1 % glutaraldehyde, so that the majority of the preparation was dimeric. 20 µl binding mixtures contained 10 fmol (0.5 nM) radiolabeled 42-mer oligonucleotide and dimeric TraA or TraAN₂₄₆ (625 nM) in 20 mM Tris HCl pH 7.5, 0.1 mM EDTA and 200 mM NaCl. After 30 minutes of incubation at 42°C the samples were loaded on a 10 % native polyacrylamide gel and run 90 minutes at constant voltage of 80 V in ice cold 22.5 mM Tris-borate, 0.5 mM EDTA buffer (pH 8.0). The gel was exposed to a storage phosphor screen GP overnight and scanned by Typhoon 9400.

4.5.6 DNase I footprint of *tra* region of pIP501

DNA-protein interactions can be detected on nucleotide level. The DNA, with one strand fluorescently or radioactively labeled, is mixed with the protein; the interactions can be detected by adding a chemical, which unspecifically cuts the DNA, e.g. DNase I or hydroxyl radical. The DNA is then run on a denaturing polyacrylamide gel and the bases, which are protected from the damaging substance by the bound protein, are not visible on the gel, compared to control DNA without bound protein.

5' radioactively labeled FP_forward and unlabeled FP_reverse primer or 5' labeled FP_reverse and unlabeled FP_forward primer were used to generate a 250-bp DNA fragment encompassing *P_{tra}* and *oriT* (nucleotides 1097 through 1346, Acc. No. L39769). PCR program "footprint" (see chapter 3.11 PCR programs) was used, resulting in labeled coding and non-coding strand (nicked strand), respectively.

Prior to the DNase I footprint reaction the PCR fragments were heated to 95°C for 10 minutes and allowed to cool down to 37°C. 10 ng of the 250-bp fragment (6.5 nM) were incubated at 37°C for 10 minutes with increasing TraA concentrations ranging from 200 nM to 2 µM in binding buffer (20 mM Tris / HCl pH 8.0, 200 mM NaCl, 0.1 mM EDTA) in a total volume of 10 µl. 0.01 Kunitz units of DNase I were added to each sample. The reaction was stopped after 1 minute for naked DNA and after 3 minutes in the presence of TraA with 10 µl of stop buffer (100 µg ml⁻¹ yeast tRNA, 30 mM EDTA, 1 % SDS, 200 mM NaCl) (Leblanc and Moss, 1994). DNA was extracted with phenol/chloroform/isoamylalcohol (25:24:1 v/v/v) and precipitated with ethanol. The pellets were air dried, dissolved in loading buffer (33 mM NaOH, 32 % formamide, 0.1 % bromophenol blue, 0.1 % xylene cyanol) and loaded onto a 5 % denaturing polyacrylamide gel containing 8 M urea. Reference sequencing reactions were prepared with the Cycle Sequencing Kit with primers FP_forward and FP_reverse. Only the nucleotide numbers are shown in Figure 25. The gel was exposed overnight to a Storage Phosphor GP Screen and read by the Typhoon 9400 imaging system.

4.5.7 Cleavage assay

Plasmids pVA2241 and pDL277 were isolated from *E. coli* DH5α cultures using Sigma GenElute™ Plasmid Miniprep Kit. The plasmid preparations were loaded onto a 0.7% agarose gel and run at 5 V/cm for 3 hours in TAE buffer. Supercoiled plasmid was cut out from the gel and extracted using Wizard® SV Gel and PCR Clean-Up kit.

Plasmid pVA2241 (6.9 kb) (Wang and Macrina, 1995a) was used as a substrate for the cleavage reaction. This plasmid contains the complete *oriT*_{PIP501} region. Control reactions were performed with the mobilisable IncQ plasmid RSF1010 (8.7 kb) (Scherzinger *et al.*, 1992) and the *oriT*_{PIP501}⁻ plasmid pDL277 (Krah and Macrina, 1989; Wang and Macrina, 1995a). The reaction mixture (20 µl) contained 20 mM Tris HCl pH 7.8, 100 mM NaCl, 15 mM MgCl₂, 0.1 mM EDTA, 0.1 % Brij 58, 200 ng bovine serum albumin, 150 ng of each plasmid and 20 or 80 fold molar excess of TraA or TraAN₂₄₆. The samples were incubated at 42°C for 60 minutes and the reactions were stopped by adjusting the mixture to 10 mM EDTA, 0.5 % SDS and 0.5 mg/ml proteinase K. After further incubation at 37°C for 30 minutes (Scherzinger *et al.*, 1992), the mixtures were subjected to electrophoresis in 0.7 % agarose gels in TAE buffer at 7 V/cm for 4 hours. Gels were stained with ethidium bromide and documented. The amount of cleaved pVA2241 DNA was determined with AlphaEaseFC™ software (Biozym Scientific, Hess Oldendorf, Germany).

4.5.8 Small Angle X-Ray Scattering (SAXS) measurements and molecule shape calculation

Small Angle X-Ray Scattering patterns are used to examine the shape of a protein in solution. When a solution of macromolecules is exposed to X-Ray radiation, the intensity of the scattering $I(s)$ can be recorded.

$$s = \frac{4\pi \sin \Theta}{\lambda},$$

where 2Θ is the angle between incident and scattered beam and λ is the wavelength of radiation. The recorded intensity is proportional to squared momentum transfer (s^2) (Svergun *et al.*, 2001).

Random orientation of the macromolecule results in isotropic distribution of intensity, which is proportional to the scattering from one molecule averaged over all orientations.

Important structural information can be extracted from the scattering curve.

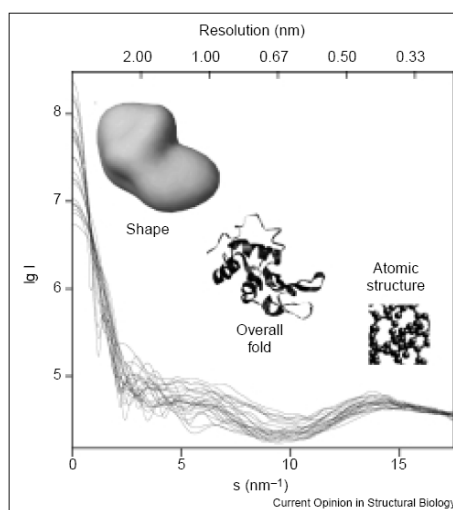


Figure 16. Small-Angle X-Ray Scattering curve. Adapted from Svergun and Koch (2002)

In each part of the scattering curve, different features of the macromolecule are represented, Figure 16. The data between s 0 and 5 nm^{-1} contain information about shape, data between 5 and 10 nm^{-1} characterize the overall fold and the data recorded for higher s include atomic structure information (Svergun and Koch, 2002).

The scattering data contain only the amplitude of the scattered waves and not the phase. To reconstruct the 3D density map, the phase problem has to be solved. As in the direct-methods approach employed in crystallography, the 3D reconstruction for SAXS data uses constraints to distinguish physical and non-physical solutions. One of the constraints is the positivity of the electron density map; the other is compactness and connectivity.

There are several computer programs, which calculate the shape of the macromolecule *ab initio* from SAXS data, namely ATSAS suite (Svergun *et al.*, 2004), Sax3D (Walther *et al.*, 2000), DALAI_GA (Chacon *et al.*, 1998; Chacon *et al.*, 2000).

Three concentrations of TraAN₂₄₆ were prepared (4.8 mg/mL, 2.4 mg/mL and 1.2 mg/mL) in buffer containing 20 mM Tris/HCl pH 7.5, 300 mM NaCl, 1 mM EDTA and 1 mM DTT. The measurements were performed at the X33 beam line at EMBL Hamburg Outstation.

The resulting data were processed with the ATSAS 2.1 (Svergun *et al.*, 2004) program suite. The primary data were merged using PRIMUS (Konarev *et al.*, 2003) and the P(R) function was calculated using GNOM (Svergun, 1992). The bead model was calculated with GASBOR (Svergun *et al.*, 2001), and averaged using DAMAVER (Volkov and Svergun, 2003).

4.6 Bioinformatic methods

4.6.1 Primary sequence analysis

The primary sequence analysis was done using program ProtParam (available at <http://ca.expasy.org/tools/protparam.html>).

4.6.2 Sequence based secondary structure prediction of proteins

One of the first secondary structure prediction algorithms was designed by Chou and Fasman in 1974 (Chou and Fasman, 1974). The authors assigned helix and β -sheet potential to each amino acid based on statistical analysis of 15 proteins with known structures. This easy calculation is even now available at www.expasy.org/tools/protscale.html as one of the possible sequence analysis methods.

With increasing computational power, new methods were designed. PHD algorithm uses sequence profiles and neural networks to predict secondary structure (Rost and Sander, 1993). The training set consists of 130 protein chains of known structure; the pairwise alignment of the used sequences shows a maximum of 25 % identity. The overall accuracy of prediction is around 69 %. For a new protein sequence one can expect 61 % - 79 % correctly predicted secondary structure, in particular when structures of several homologous proteins are known. The secondary structure prediction server using PHD method is available under <http://www.PredictProtein.org>.

PSIPRED is another secondary structure prediction algorithm. It is based on neural network evaluation of profiles generated by local pairwise alignments (Jones, 1999). It is available at <http://bioinf.cs.ucl.ac.uk/psiform.html>.

Somewhat smaller accuracy is achieved by JPred, an algorithm, which calculates a consensus prediction based on results obtained from 6 different servers (Cuff *et al.*, 1998).

An interactive secondary structure prediction server Network Protein Sequence @nalysis (NPS@), available at http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/-npsa_seccons.html (Combet *et al.*, 2000), where up to 12 different methods can be chosen. A consensus prediction is then displayed along with the results of the chosen algorithms.

The secondary structure of TraA, TraAN₂₆₈ and TraAN₂₄₆ were calculated using PSIPRED and NPS@.

4.6.3 Protein tertiary structure prediction

Protein 3D structures provide the answers to functional and mechanistic questions. Without them, it is impossible to say, how the given protein works. Nuclear magnetic resonance (NMR) and X-ray crystallography are the methods for elucidation of 3D structures on atomic level. Proteins with similar amino acid sequences have also similar three-dimensional structures, that is why it is possible to predict a 3D structure from an amino acid sequence, if there is a structure of a similar protein available in the Protein Data Bank (pdb) (Berman *et al.*, 2000).

The structures of Orf10 and the lytic transglycosylase domain of Orf7 were predicted using 3Djigsaw (Bates *et al.*, 2001) <http://www.bmm.icnet.uk/servers/3djigsaw/> and SWISS-model (Schwede *et al.*, 2003) <http://swissmodel.expasy.org/> servers.

4.6.4 Topology prediction

Subcellular localization of proteins can provide insight into putative function of the protein. The localization can be predicted using PSORTb (<http://www.psort.org>) (Gardy *et al.*, 2005). The program takes into account sequence features known to influence the localization within the cell.

Predictions of transmembrane helices were done using several algorithms: PHDhtm (Rost *et al.*, 1996), HMMTOP (Tusnady and Simon, 2001), TMPred (Hofmann and Stoffel, 1993), TMAP (Milpetz *et al.*, 1995; Persson and Argos, 1996; Persson and Argos, 1997), TopPred (von Heijne, 1992; Claros and von Heijne, 1994), SOSUI (Scholz *et al.*, 1989) and DAS (Cserzo *et al.*, 1997).

4.6.5 Prediction of secondary structure of DNA

All secondary structure predictions of DNA were calculated with mFOLD (SantaLucia, 1998; Zuker, 2003) available at <http://www.bioinfo.rpi.edu/applications/mfold/>.

4.6.6 Txt2dic program

The file format obtained from the CD measurement is not readable by Dicroprot 2000 (Deleage and Geourjon, 1993), a program for secondary structure content calculation from CD spectra. To convert the file to the desirable format, a program was written in Borland Delphi 6, namely txt2dic. The Delphi code can be found in appendix.

5 Results

5.1 *oriTs* of RSF1010 family

The *nic* sites in five *oriTs* of the RSF1010 family are known, they lay in the center of the consensus sequence (see Figure 12). Nonetheless, the specificity of relaxase binding must confer to something else than this sequence, since e.g. the relaxase of pIP501 does not nick the RSF1010 plasmid, see chapter 5.8. In the near neighborhood of the consensus sequence, there are inverted repeat (IR) sequences, which potentially can form secondary structures, hairpins (Figure 17).

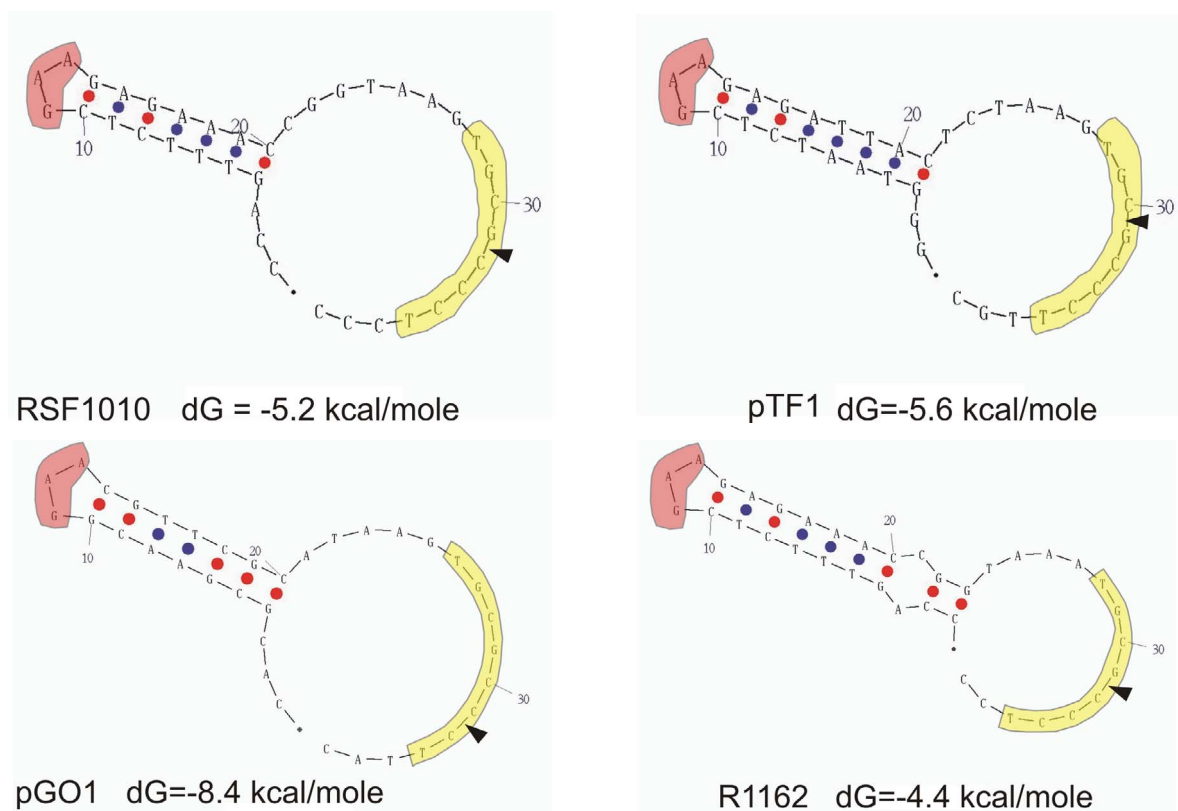


Figure 17. (Continued)

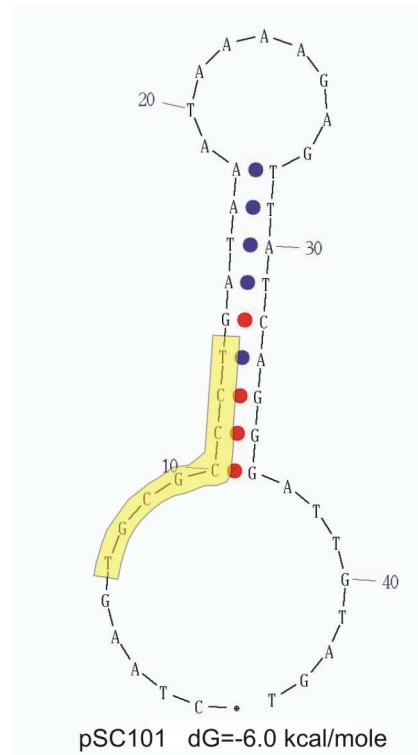
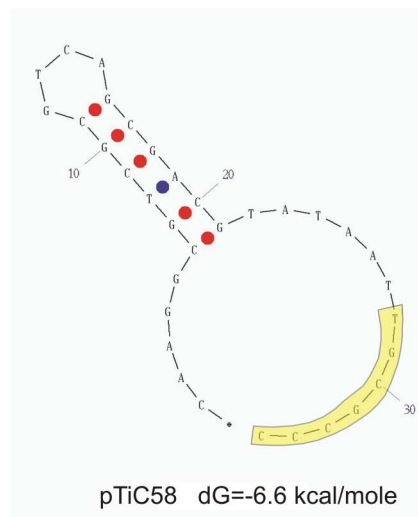
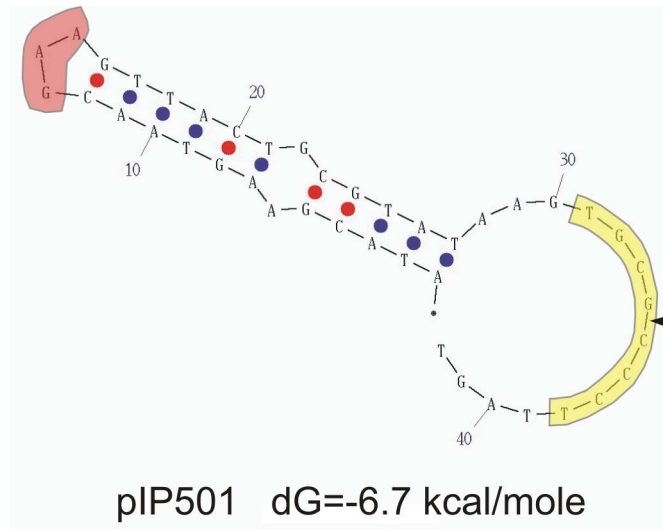


Figure 17. RSF1010-family *oriTs* forming hairpin structures.

Figure 17 presents the folds, which were predicted at 37°C in 150 mM NaCl (Peyret, 2000). The consensus sequence is marked with yellow background; the known *nic* sites are shown with a black arrow and the conserved 5'-GAA-3' sequence is marked with red background. Free energy values vary from -4.4 kcal/mole for the R1162 hairpin to -8.4 kcal/mole for the pGO1 hairpin. The lower the free energy value, the more stable the hairpin is.

5.2 *TraA*, *TraAN*₂₆₈ and *TraAN*₂₄₆ expression and purification

TraA, *TraAN*₂₆₈ and *TraAN*₂₄₆ were expressed successfully as 6xhis tag fusions from *E. coli* XL1blue cells. The soluble protein were purified using two-step affinity chromatography and the purity of the sample was analyzed by SDS-PAGE (Figure 18).

While the *TraAN*₂₄₆ band migrates at 32 kDa in the denaturing PAGE according to its expected molecular mass of 31.6 kDa, the *TraA* band migrates at approximately 66 kDa, significantly faster than expected for its calculated molecular mass of 78.6 kDa. This behaviour may be due to the basic nature of the *TraA* protein (theoretical pI=8.89) and has been previously described for the HMG (high mobility group) proteins (Johns, 1982).

The two-step purification procedure yielded highly pure protein. Expression of *TraA*, *TraAN*₂₆₈ and *TraAN*₂₄₆ resulted in 1.4 g cells per L of culture. The yield of purified protein was approximately 3 mg of *TraA* and 1.5 mg of *TraAN*₂₆₈ (data not shown) and *TraAN*₂₄₆ per L of culture.

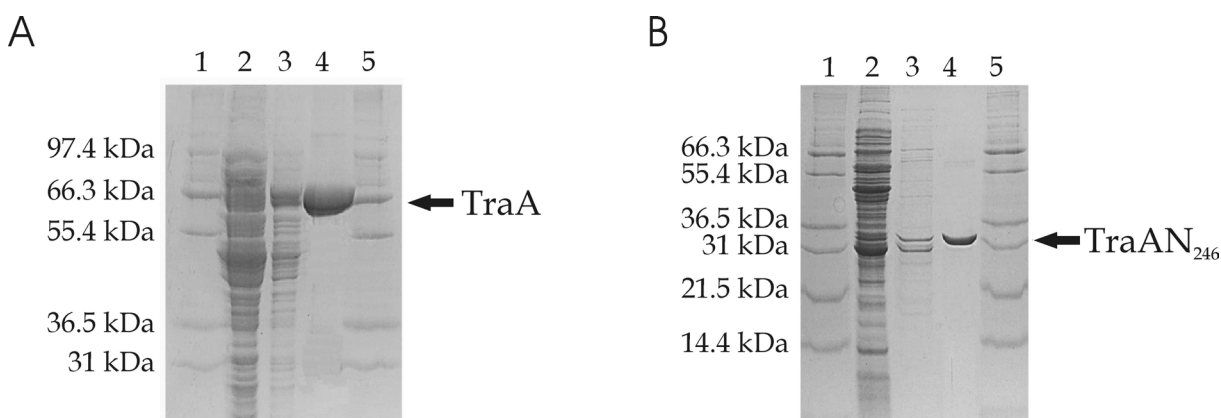


Figure 18. Protein purification. Samples were applied to 12% (w/v) SDS-PA gel (*TraA*, A) or 18% (w/v) (*TraAN*₂₄₆, B) polyacrylamide gel stained with Coomassie Brilliant Blue. Lanes 1 and 5 - Mark12 Protein standard, lane 2 - clarified lysate, lane 3 - eluate after the HiTrap Chelating Ni²⁺ charged column, lane 4 - eluate after the HiTrap heparin column.

5.3 *The relaxase contains a mixed α/β -fold*

In order to check the structural integrity and to compare the fold content of the full-length and the C-terminal truncated *TraA* proteins the far UV circular dichroism (CD) spectra of purified *TraA*, *TraAN*₂₄₆ and *TraAN*_{246-oriT} DNA complex were recorded at room temperature.

The spectra were normalized for protein concentration and residue number and superimposed (Figure 19 A), showing mainly α -helical features for both proteins and a higher fold content for the *TraAN*₂₄₆ protein.

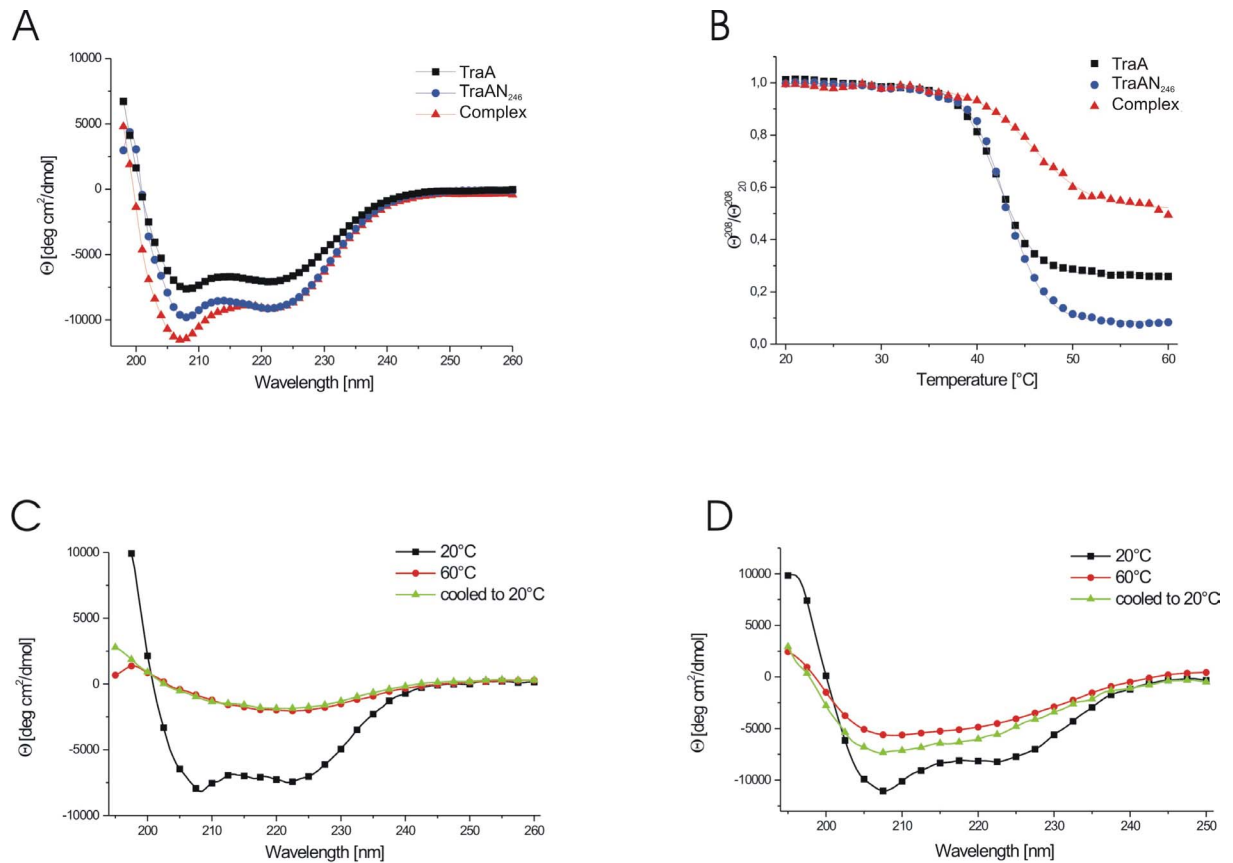


Figure 19. Circular dichroism. A. Spectra of TraA, TraAN₂₄₆ and TraAN₂₄₆-oriT complex at 20°C, B. Thermal unfolding of TraA, TraAN₂₄₆ and TraAN₂₄₆-oriT complex. Heat denaturation of TraA (C) and TraAN₂₄₆ (D), spectra were taken at 20°C, followed by heating to 60°C and cooling down to 20°C

To quantify the secondary structure fractions in the two proteins the K2d program (Andrade *et al.*, 1993) was used. The results obtained for TraA revealed 30 % α -helix, 13 % β -sheet and 57 % random coil, whereas TraAN₂₄₆ showed a higher β -sheet content (27 % α -helix and 18 % β -sheet). This distribution indicates that the β -sheet elements are mainly concentrated in the N-terminal relaxase domain and that the C-terminal domain of TraA constitutes a predominantly α -helical fold. These results are in very good agreement with secondary structure predictions (see appendix) using the program PSIPRED (McGuffin *et al.*, 2000), which yield a mixed α/β -fold for the relaxase domain (41 % α , 15 % β) and an almost entirely α -helical prediction for the C-terminal part (residues 247-654: 65 % α -helix, 4.8 % β -strand).

5.4 TraA and TraAN₂₄₆ unfold at the same temperature

The thermal stability of TraA and TraAN₂₄₆ was investigated. Temperature scans were performed on the CD-spectropolarimeter following a step-scan procedure, where the CD

spectra of the proteins were recorded at 10 degree intervals upon heating and cooling. In addition, the CD signal at 208 nm was recorded every 0.2 degrees over the whole temperature range (Figure 19 B). The transition temperature (T_M) for both proteins was 42°C and the unfolding curves are very similar in shape, indicating a single domain transition or at least a highly cooperative unfolding procedure. TraA precipitated in the measuring cell when the temperature scan approached 60°C and the corresponding CD spectrum shows a major reduction of the CD signal. No refolding was observed upon cooling to room temperature (Figure 19 C). In contrast the TraAN₂₄₆ protein retained some secondary structure when heated to 60°C, but the minimum of the spectrum was shifted towards shorter wavelengths, as expected for a higher random coil fraction (Figure 19 D). Only a slight refolding effect could be observed for TraAN₂₄₆ upon cooling, which means that the heat denaturation is essentially irreversible for both TraA and TraAN₂₄₆.

5.5 *The relaxase domain is stabilized through DNA-binding*

DNA-binding proteins often undergo conformational rearrangements upon DNA-binding (Patel *et al.*, 1990; Weiss *et al.*, 1990). The effect of *oriT*-TraAN₂₄₆ complex formation was investigated by CD-spectroscopy. At room temperature, the CD spectrum of the 1:1 stoichiometric TraAN₂₄₆-DNA complex showed a 20 % increase of the CD-signal at 208 nm (Figure 19 A). In order to analyze the effect of DNA-binding on the thermodynamic behavior of TraAN₂₄₆ a temperature scan of the complex from 20 to 65°C was performed (Figure 19 B). The T_M of the complex, as determined by non-linear least-squares fit using a Boltzmann function, was 48.8°C indicating a significant stabilization of the relaxase domain. Approaching the high temperature end of the scan the protein precipitated and the unfolding procedure was irreversible.

5.6 *TraA and TraAN₂₄₆ form dimers in solution*

The quaternary structures of the TraA protein and the N-terminal domain TraAN₂₄₆ were investigated by chemical cross-linking experiments.

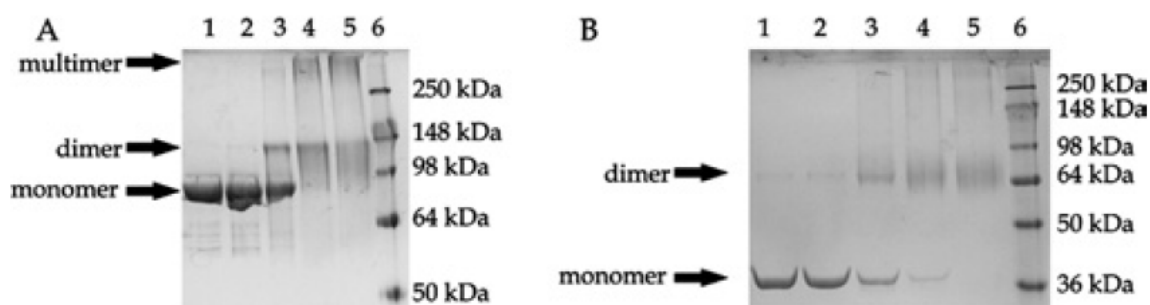


Figure 20. Chemical cross-linking of TraA (A) and TraAN₂₄₆ (B). Lane 1 - no glutaraldehyde added, lane 2 - 0.001 %, lane 3 - 0.01 %, lane 4 - 0.05 %, lane 5 - 0.1 % glutaraldehyde, lane 6 - protein marker SeeBlue[®] plus2.

Using increasing amounts of glutaraldehyde as cross-linking reagent it can be seen that both the full-length TraA and the TraAN₂₄₆ form dimers. In lane 1 of Figure 20A and B the proteins without cross-linking reagent were loaded on an 18 % SDS-polyacrylamide gel, showing the expected positions for the monomeric proteins. With the lowest glutaraldehyde concentration (0.001 %, lane 2 in Figure 20 A and B) a faint band appears at the expected position corresponding to the double molecular mass of the monomers (expected molecular mass: 157.1 kDa for TraA and 63.3kDa for TraAN₂₄₆). With increasing glutaraldehyde concentration (0.01 %, 0.05 % and 0.1 % in lanes 3 to 5, Figure 20) the monomer bands gradually vanish and the dimer bands become the predominant species. In the case of TraA the intensity of the dimer-bands decreases at high cross linker concentration (Figure 20 A lanes 4 and 5) and a high molecular mass species appears indicative of unspecific cross-linking.

5.7 The shape of the TraAN₂₄₆ molecule shows a dimeric form

The shape of the TraAN₂₄₆ molecule was calculated by Small Angle X-ray Scattering (SAXS).

The data obtained from a sample containing 4.8 mg/ml TraAN₂₄₆ showed aggregation of the protein and were not taken for shape determination, only data measured at 2.4 mg/mL and 1.2 mg/mL were used.

The merged scattering curve and P(R) function are presented on Figure 21. The radius of gyration, calculated from Guinier fit is 6.5 nm and the intensity extrapolated to 0 angle at concentration 0 (I_0) is 22202. Comparing I_0 of TraAN₂₄₆ and standard protein (BSA), the molecular weight of the species present in solution can be calculated:

$$MW_{TraAN_{246}} = \frac{I_0(TraAN_{246})}{I_0(BSA)} * MW_{BSA} = \frac{22202}{19910} * 66 \text{ kDa} = 73 \text{ kDa}$$

The theoretical molecular weight of the TraAN₂₄₆ dimer is 67 kDa.

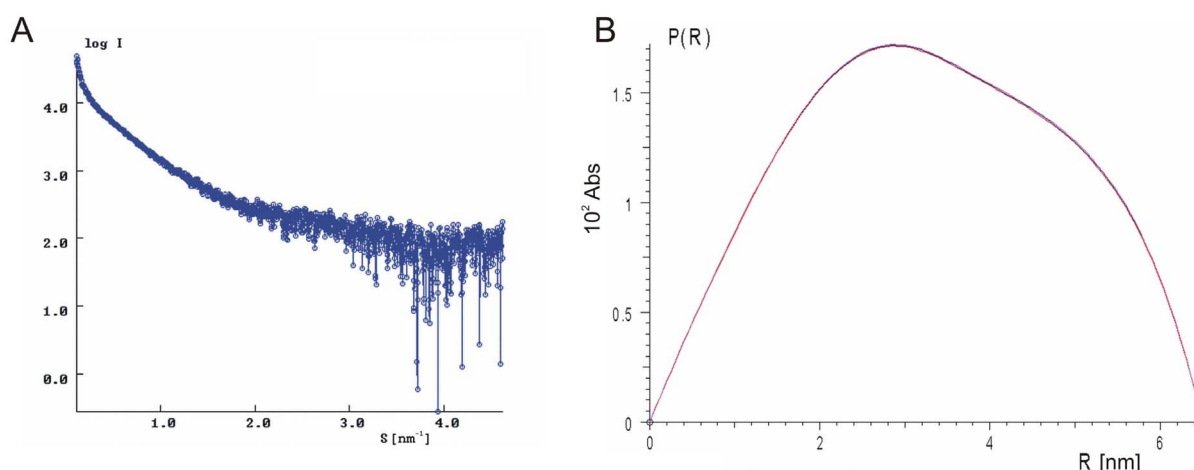


Figure 21. Small Angle X-ray Scattering. A. Scattering curve. B. Pair distance distribution function.

The final envelope was determined as an average of 15 calculations. Figure 22 shows the calculated surface of TraAN₂₄₆ with an evident symmetry.

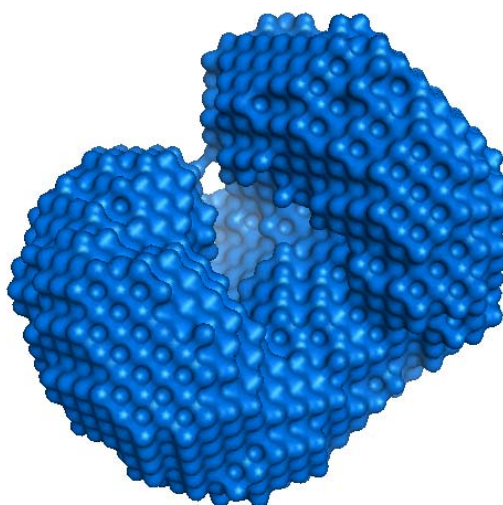


Figure 22. The shape of the TraAN₂₄₆ molecule in solution.

5.8 TraA and TraAN₂₄₆ show nicking activity in vitro

Relaxation activities of TraA and TraAN₂₄₆ proteins on supercoiled plasmid DNA were tested. The *oriT*_{plP501} containing plasmid pVA2241 (6.9 kb) was used as substrate for the reaction. The plasmid DNA subjected to the reaction contained ~73 % supercoiled form (ccc). TraA and TraAN₂₄₆ caused an increase of the amount of relaxed form (oc). The optimal cleavage conditions for GST-TraA and GST-TraA* (amino terminal 293 amino acids of

TraA) had been determined before (Kurenbach *et al.*, 2002). Under these conditions, the maximum conversion of the supercoiled plasmid DNA into the relaxed form obtained by incubation of the plasmid with TraA reached 78 % oc form (Figure 23, lane 3) and with TraAN₂₄₆ around 49 % (Figure 23, lane 5).

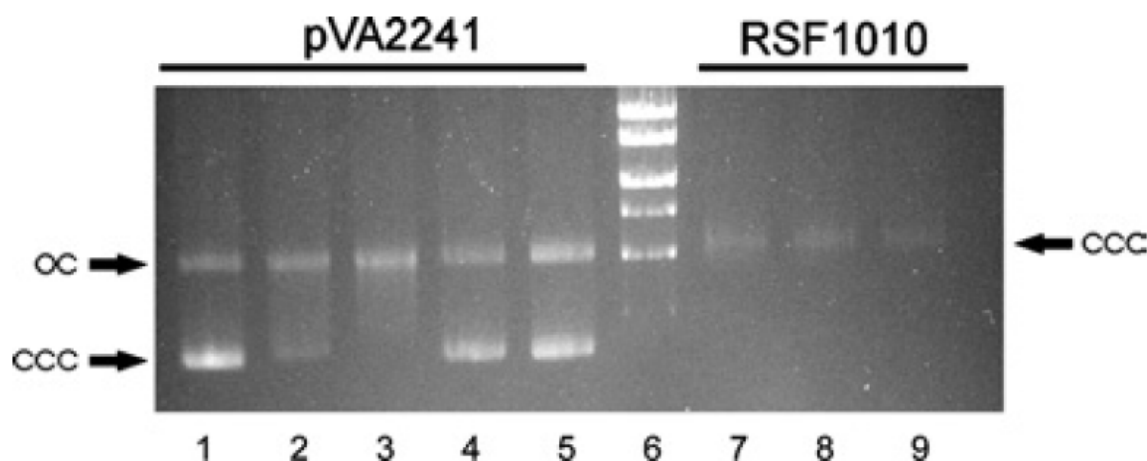


Figure 23. Cleavage assay. Lane 1 - pVA2241 without protein, lane 2 - pVA2241 with 20-fold excess of TraA, lane 3 - pVA2241 with 80-fold excess of TraA, lane 4 - pVA2241 with 20-fold excess of TraAN₂₄₆, lane 5 - pVA2241 with 80-fold excess of TraAN₂₄₆, lane 6 - Mass Ruler™ high range, lane 7 - RSF1010 with 80-fold excess of TraA, lane 8 - RSF1010 with 80-fold excess of TraAN₂₄₆, lane 9 - RSF1010 without protein. ccc- covalently closed circular plasmid DNA, oc – open circular plasmid DNA.

No relaxation was observed for the IncQ plasmid RSF1010 (8.7 kb) belonging to the same *oriT* nick region family as pIP501 (Zechner *et al.*, 2000) and the *oriT*_{pIP501}- plasmid pDL277 (6.6 kb) (Krah and Macrina, 1989; Wang and Macrina, 1995a) (data not shown). In comparison with previous data, showing that the first 293 amino acids of TraA are sufficient for *oriT*_{pIP501} cleavage (Kurenbach *et al.*, 2002), TraAN₂₄₆ (amino-terminal 246 amino acids of TraA) now represents the smallest known TraA portion exhibiting specific transesterase activity.

5.9 TraA and TraAN₂₄₆ bind specifically to *oriT*_{pIP501}

The oligonucleotides used for the band shift experiments are shown in Figure 24.

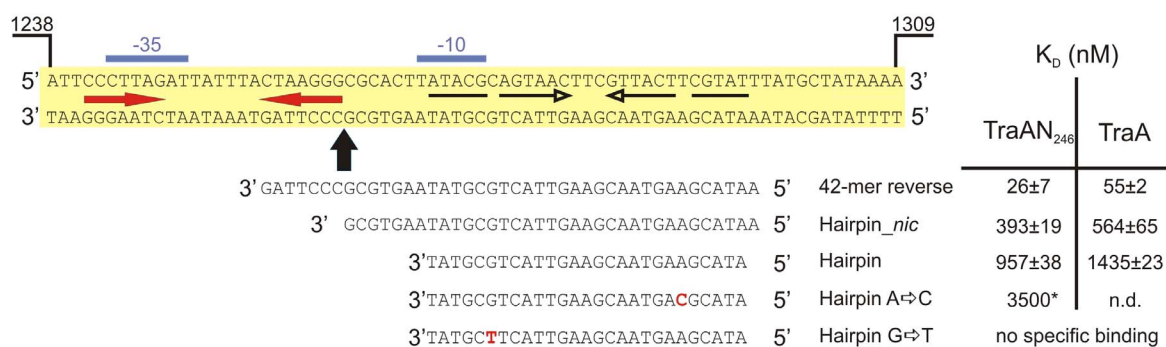


Figure 24. Oligonucleotides used in *oriT* binding assays. The *oriT*_{pIP501} region consists of a perfect inverted repeat (IR-1) (red horizontal arrows), an imperfect inverted repeat (IR-2) (black horizontal arrows) and the *nic* site (vertical arrow), *P*_{tra}, -35 and -10 regions are indicated. The dissociation constants estimated for oligonucleotides covering IR-2 are presented next to the respective oligonucleotides. The exchanged bases forming perfect hairpins are shown in red

The 42-mer oligonucleotide, Figure 25, comprises the sequence of the DNA strand, which is nicked *in vivo* by the TraA protein (Wang and Macrina, 1995a). It contains an inverted repeat, which might exhibit a stem-loop structure with 11 base pairs and a bulge in the middle. Its annealing temperature is $T_M = 66.2^\circ\text{C}$ and $\Delta G = -5.42$ kcal/mol at 42°C , as calculated using mFold (SantaLucia, 1998; Zuker, 2003). There is a single-strand region comprised of the nick site and additional 7 bases at the 3' terminus (nucleotides 1256 through 1297 on pIP501, Acc. No. L39769).

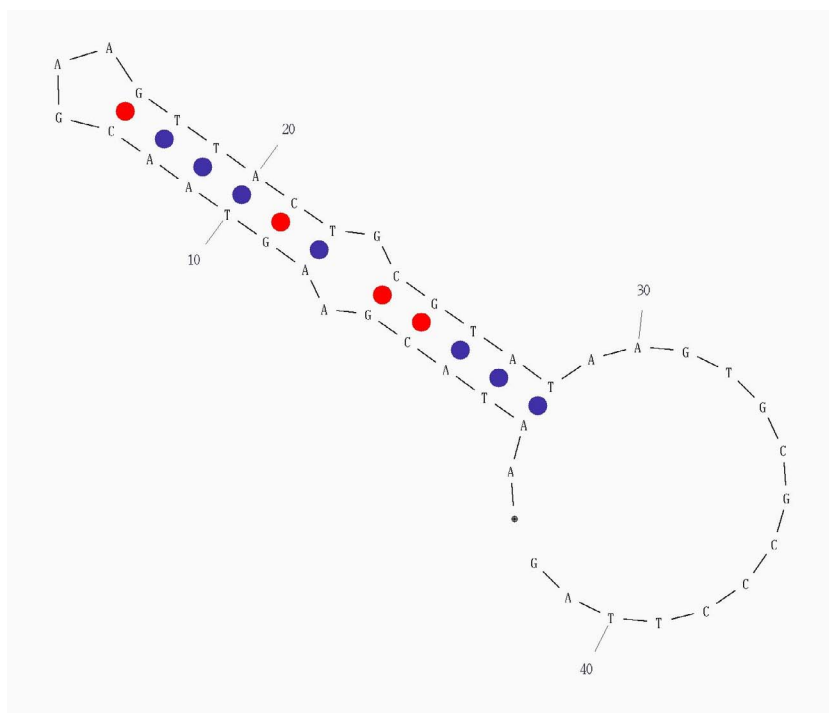


Figure 25. The calculated secondary structure of the 42-mer.

The 42-mer oligonucleotide-protein binding reactions produce more than one shifted species. Low concentrations of TraA favour formation of a slowly migrating complex (complex II in Figure 26A), higher concentrations result in an increasing fraction of the faster migrating complex (complex I) and at high ($>5 \mu\text{M}$) TraA concentrations, mostly aggregates are formed, incapable to enter the gel.

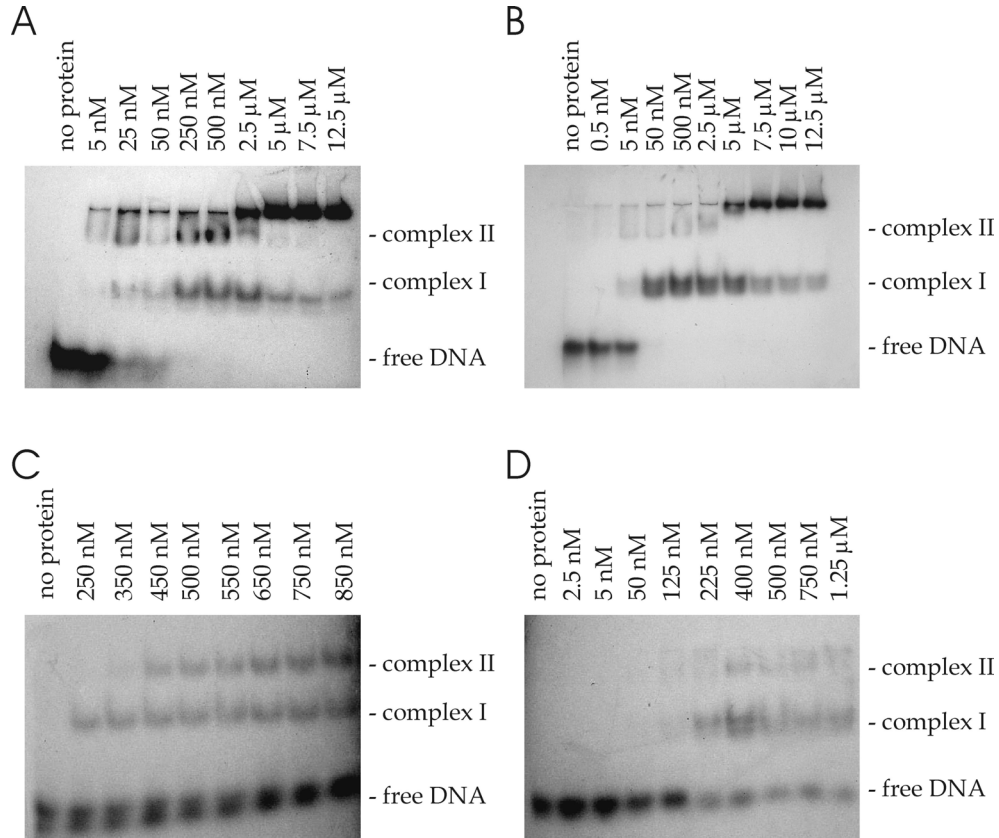


Figure 26. Band shift analysis. Increasing amounts of TraA (A and C) and TraAN₂₄₆ (B and D) were subjected to binding to 42-mer (A and B) and hairpin_*nic* (C and D) oligonucleotide.

At low protein concentrations predominantly complex I is formed for TraAN₂₄₆ (Figure 26 B). As for TraA, also high concentrations of TraAN₂₄₆ built up large aggregates that were unable to enter the gel. In order to identify the nature of the shifted complexes, band shift experiments were repeated with chemically cross-linked protein preparations, where the main fraction of the protein formed dimers. Using the cross-linked TraA protein, the complex II in Figure 26 A is identified as the TraA-dimer DNA complex (Figure 27). When using the cross-linked TraAN₂₄₆ sample, however, the shifted band coincides with the complex I band in Figure 26 B. Complex II in Figure 27 therefore represents a higher oligomer of TraAN₂₄₆, which was also observed in solution in the SAXS experiment of this protein, or an unspecific binding of more than one TraAN₂₄₆-dimer to one oligonucleotide, respectively.

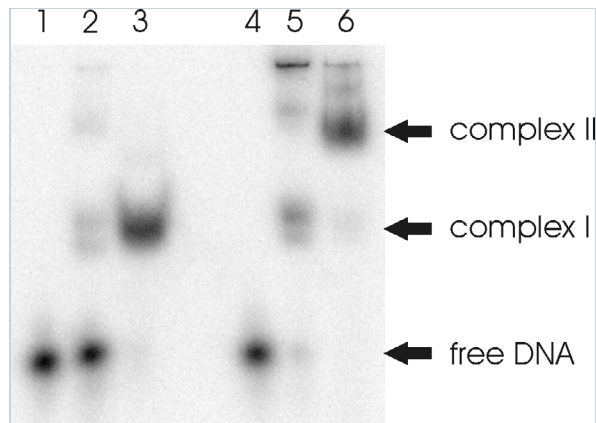


Figure 27. Band shift of 42-mer oligonucleotide with native and cross-linked proteins. 0.5 nM 42-mer was incubated with 1.25 μ M native protein or 625 nM dimeric protein at 42°C for 30 min. Lane 1 - free DNA; lane 2 - 42-mer + native TraAN₂₄₆; lane 3 - 42-mer + dimeric TraAN₂₄₆; lane 4 - free DNA; lane 5 - 42-mer + native TraA; lane 6 - 42-mer + dimeric TraA.

The 42-mer oligonucleotide is bound by TraA and TraAN₂₄₆ with a dissociation constant of 55 and 26 nM, respectively. TraA and TraAN₂₄₆ bind the hairpin_{nic} oligonucleotide generating both retarded complexes (Figure 26 C and D), but with lower affinity than the 42-mer resulting in a K_D of 564 nM for TraA and 393 nM for TraAN₂₄₆. The dissociation constant increases approximately another 2.5 fold, when only the hairpin is available. Since the wild type hairpin contains a mismatch (see Figure 24), perfect hairpins were generated and used in the EMSA experiments, too. The A \rightarrow C perfect hairpin is bound to TraAN₂₄₆ approximately 4 times weaker (K_D =3.5 μ M) than the wild type hairpin (K_D =957 nM). Applying high protein concentrations to oligonucleotides containing the hairpin without flanking sequences (original bulged hairpin and hairpin with A \rightarrow C exchange) generates complex I and a complex not entering the gel. Incubation of TraA and TraAN₂₄₆ with the perfect hairpin (G \rightarrow T exchange in the left half repeat of IR) resulted exclusively in a high molecular mass DNA-protein complex not entering the gel (data not shown).

Putative cross-reactivity of TraA and TraAN₂₄₆ on *oriT*s of different *oriT* nick region families (Zechner *et al.*, 2000) was investigated by applying oligonucleotides containing *oriT*_{pMV158} and *oriT*_{RP4}, respectively, to the EMSAs. Even at a protein/DNA ratio of 35,000 the pIP501 DNA relaxase did not bind to the heterologous origins (data not shown). Specificity of TraA and TraAN₂₄₆ binding was further tested by EMSAs with a randomly selected 42-mer oligonucleotide. The oligonucleotide was incubated with increasing TraA or TraAN₂₄₆ concentrations up to 5 μ M, no binding was observed. In competition assays the labeled 42-mer oligonucleotide (*oriT* 42-mer or random 42-mer) was first incubated with TraA or TraAN₂₄₆ and then with different concentrations of the respective unlabeled 42-mer

oligonucleotide. A 250-fold excess of the specific competitor completely inhibited binding, whereas a 1000-fold excess of unspecific competitor DNA (random 42-mer) did not affect binding of the TraA and TraAN₂₄₆ proteins (data not shown). This clearly demonstrates that the pIP501 TraA relaxase binds specifically to its cognate *oriT* region.

5.10 The TraA relaxase binds to the *P_{tra}* promoter

The compact organisation of the pIP501 *oriT* region (Figure 28), in the sense that the *P_{tra}* -10 and -35 box overlap with the left half repeat of inverted repeat structures (IR-1 and IR-2), presumably representing the TraA recognition and binding site, makes autoregulation of the *tra* operon by the TraA relaxase likely.

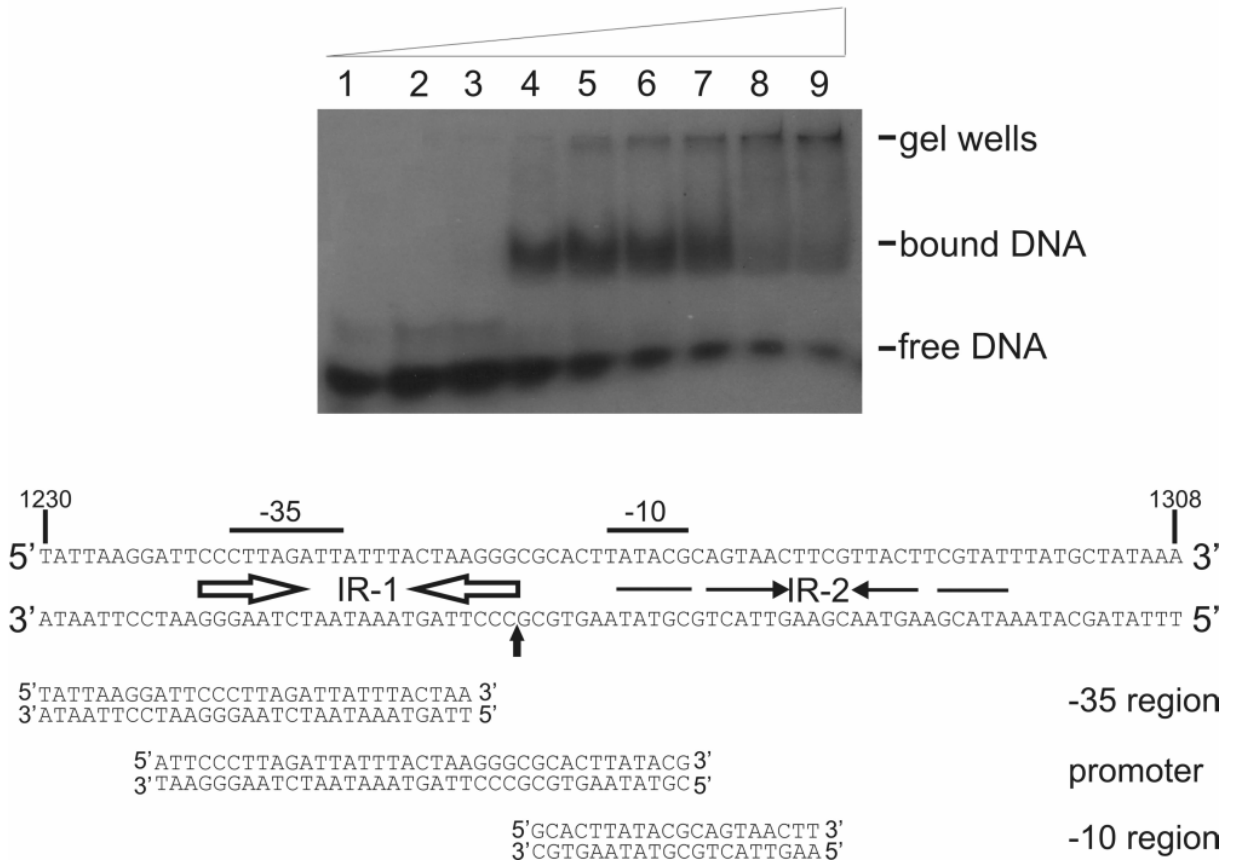


Figure 28. EMSA of the -10 promoter region fragment with TraAN₂₄₆. Increasing TraAN₂₄₆ concentrations (0, 5, 50, 200, 350, 500, 650, 800 nM, 1 μ M) were incubated with 10 fmol DNA of the -10 promoter region fragment at 37°C for 30 min, and loaded onto a 10% native polyacrylamide gel. The *oriT*_{pIP501} region, containing the *P_{tra}* promoter, a perfect inverted repeat (IR-1) (open horizontal arrows), an imperfect inverted repeat (IR-2) (solid horizontal arrows) and the *nic* site (vertical arrow), are indicated. The promoter fragment and the -35 DNA fragment are also shown.

To study relaxase binding to the P_{tra} promoter, three DNA fragments were selected, the first comprising the whole promoter (-35 and -10 region), the second only the -35 and the third only the -10 region. The shortest aminoterminal portion of TraA exhibiting relaxase activity, the 246 amino acid-TraAN₂₄₆ protein, was used in the bandshift assays. The oligonucleotide representing the *tra* coding strand was 5' labeled and annealed to the complementary unlabeled strand to generate double strand substrates for the EMSAs. Exemplarily, the data with the -10 region fragment are shown. Applying increasing TraAN₂₄₆ concentrations to this fragment, one retarded DNA-protein complex was detected (Figure 28).

Incubation of the -35 region and the promoter fragment with TraAN₂₄₆ concentrations > 1.0 μ M resulted in a complete shift represented by exclusively large protein-DNA complexes not entering the gel. Binding affinity for the -35 region and for the whole promoter region was lower than for the -10 region fragment. This could be due to presence of the complete left half repeat of the inverted repeat structure (IR-2) in the -10 region fragment.

The complete left half repeat of IR-2 was present in all single strand oligonucleotides shown to bind TraAN₂₄₆ and TraA. An oligonucleotide comparable with the -10 region fragment, but additionally containing the right half repeat, resulted in similar binding affinity. TraA showed similar binding affinities for the different promoter fragments as the aminoterminal domain TraAN₂₄₆.

Incubation of TraAN₂₄₆ concentrations up to 4 μ M with a 42-mer control fragment (8,000-fold excess of protein) resulted in no visible shift. It can be concluded, that the relaxase binds to the P_{tra} promoter region. This is in good agreement with relaxase-binding to oligonucleotides comprising i) the complete IR-2 structure, ii) IR-2 and the region until the *nic*-site, and iii) IR-2, *nic*-site and seven further 5' bases.

5.11 DNase I footprint of *tra*_{pIP501}

DNase I footprinting analyses with a 250-bp DNA fragment comprising P_{tra} and the complete IR-1 and IR-2 structures showed protection of both the P_{tra} -35 and -10 region with hypersensitive sites on the non-cleaved strand close to the *nic* site (nucleotide numbers 1253 and 1256), at the *nic* site (nucleotide number 1262) and two nucleotides downstream the -10 region. On the cleaved strand DNase I protection extends eight nucleotides to the *nic* site with the *nic*-site itself as hypersensitive site (Figure 29). The DNase I hypersensitive sites could be generated by conformational change of the *oriT* region induced by TraA binding resulting in better exposure to DNase I attack.

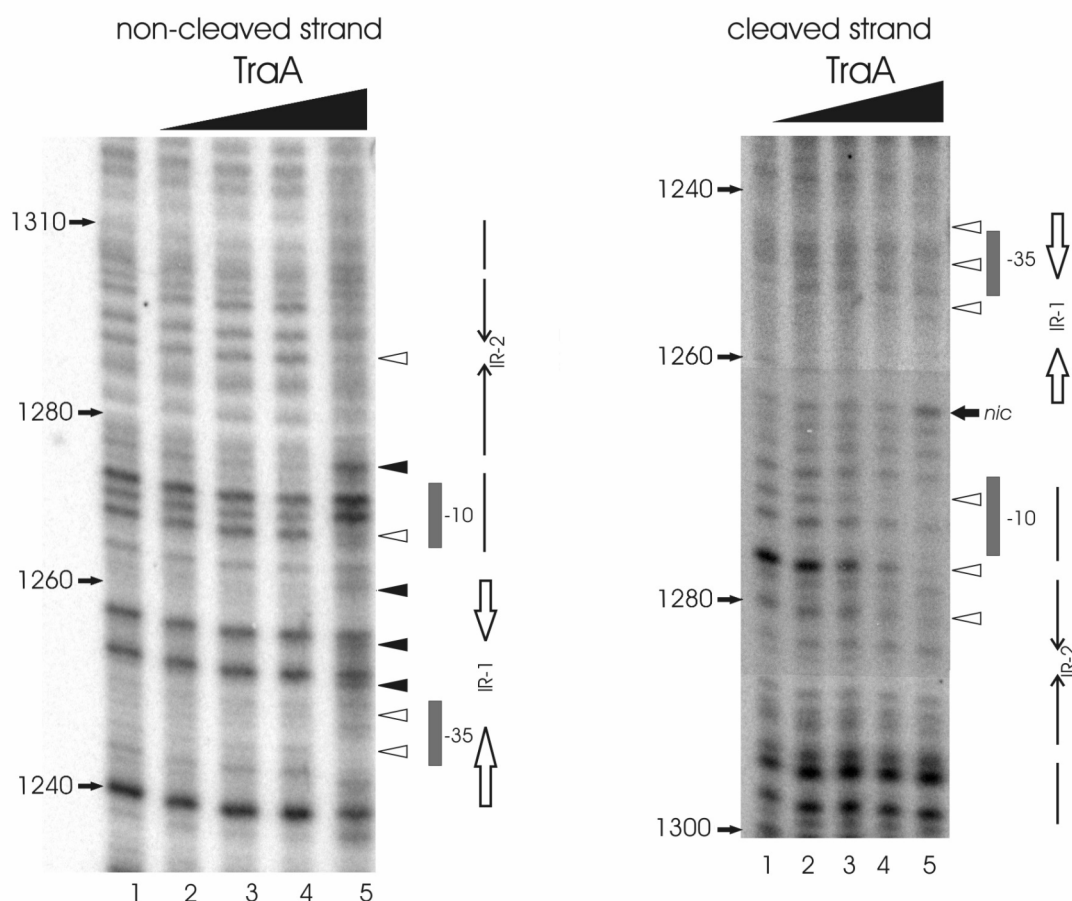


Figure 29. DNase I footprinting analysis of the *oriT* region of pIP501. 100 ng of the 250 bp *oriT* fragment were incubated with increasing TraA concentrations: lane 1, no TraA added; lane 2, 200 nM TraA; lane 3, 500 nM TraA; lane 4, 1 mM TraA; lane 5, 2 mM TraA. The nucleotide positions on the pIP501 sequence (Acc. No. L39769) are shown on the left; the -10 and -35 *P_{tra}* regions are marked by filled rectangles. The open and solid arrows indicate the inverted repeats IR-1 and IR-2, respectively. The open and solid arrowheads to the right of each panel show the bands decreased and increased in intensity, respectively, with increasing concentrations of TraA. The *nic* site is marked with a horizontal arrow. In the right-hand panel, the gel image is composed of three parts with different brightness and contrast for better clarity.

The data indicate that the left half repeat of IR-1 and IR-2 are the preferential binding sites for the TraA relaxase. Binding of TraA to its target DNA would be a prerequisite for the recognition and cleavage of DNA at the 5'-CpG-3' dinucleotide in the *nic*-site, which would remain accessible to the enzymatic activity of TraA.

5.12 Crystallization of the TraA:hairpin_{nic} complex

TraA protein was concentrated to 7 mg/mL in buffer containing 50 mM Tris/HCl, 300 mM NaCl and 1 mM EDTA, pH 7.8, mixed with purified and annealed hairpin_{nic}

oligonucleotide (see chapter 4.1.5) in 1:1.2 molar ratio, with oligonucleotide excess. To form the complex, the mixture was heated to 42°C for 5 minutes and slowly cooled down to room temperature. Vapor diffusion crystallization trials were set up with NATRIX screen. A crystal (Figure 30 A) and spherulites (Figure 30 B) were found in the drops after 2 weeks of incubation at 20°C.

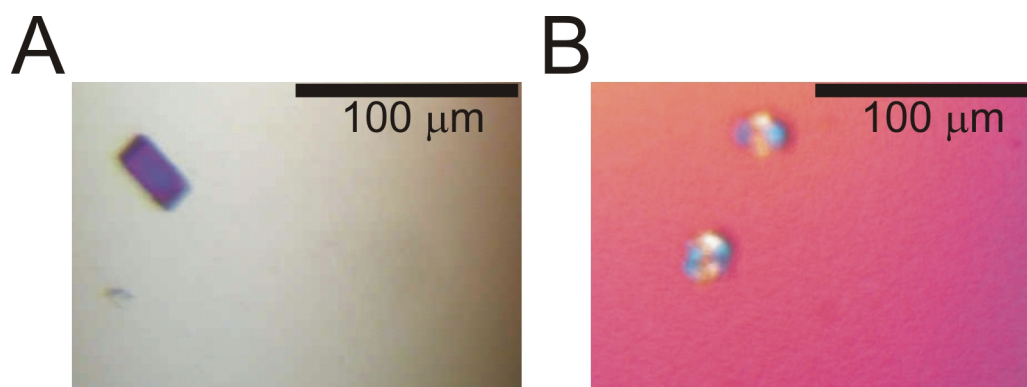


Figure 30. TraA-hairpin_nic crystallization trials. (A) Crystal, (B) Spherulites.

The putative complex crystallized in a drop containing NATRIX screening solution 47 (50 mM Tris/HCl pH 8.5, 100 mM KCl, 10 mM MgCl₂, 30% PEG 400). Unfortunately this crystal did not diffract and was not reproducible. The spherulites were formed in a drop with NATRIX solution 29 (50 mM HEPES, pH 7.0, 10 mM MgCl₂, 4M LiCl) and were reproducible, but no crystals were obtained upon optimization.

5.13 Analysis of other components of the *tra_{PIP501}* operon

The *tra_{PIP501}* operon consists of 15 open reading frames (*orfs*) (Kurenbach *et al.*, 2006). For the accession numbers see Appendix 8.3.

Amino acid sequences of Orf2-Orf15 were analysed to predict their localization within the cell and the presence of transmembrane helices.

Table 3. Topology prediction of Orf2-Orf15

	PHDhtm	HMMTOP	Psort	TMPred	TMAP	TopPred	SOSUI	DAS
Orf2	48-65 89-106	12-31 48-66 85-103	Cytoplasmic membrane	10-30 47-66 88-104	10-29 40-68 80-107	12-32 46-66 84-104	11-33 48-70 85-107	16-24 47-62 87-102
Orf3	49-73 78-95	47-66 71-90	Unknown	70-88	59-87	67-87	56-78	48-86 (two TMH with a short linker)

Orf4	Soluble	Soluble	Unknown	Soluble	Soluble	Soluble	Soluble	Soluble
Orf5	Soluble	Soluble	Cytoplasmic	Soluble	Soluble	Soluble	Soluble	Soluble
Orf6	216-233	Soluble	Unknown	218-235	210-236	217-237	216-238	Non-TM-protein
Orf7	18-35	19-35	Unknown	19-35	10-37	16-36	14-36	19-35
Orf8	11-28	8-27	Unknown	10-29	8-36	9-29	7-29	9-27, potential signal peptide
Orf9	16-35 65-83	12-36 67-86	Cytoplasmic membrane	12-29 68-86	10-38 65-92	11-31 67-87	18-39 67-89	10-36 67-88
Orf10	Soluble	Soluble	Cytoplasmic	Soluble	Soluble	Soluble	Soluble	Soluble
Orf11	45-62	47-64	Unknown	46-64	37-65	45-65	42-63	46-60
Orf12	42-59 72-95 120-141 177-201 206-223 239-258 266-283	71-93 118-136 167-191 198-222 241-259 286-308	Cytoplasmic membrane	77-93 114-136 177-194 196-222 243-259 262-284	68-96 113-141 169-189 196-216 235-261 266-289	73-93 117-137 174-194 203-223 240-260	75-97 115-136 173-195 201-223 237-259 263-285	76-93 176-220 241-258 268-277
Orf13	168-185	168-186	Unknown	168-186	164-188	166-186	167-189	168-185
Orf14	Soluble	Soluble	Unknown	Soluble	Soluble	Soluble	Soluble	Soluble
Orf15	261-278	4-21 261-278	Cell wall	4-24 260-278	1-23	4-24 259-279	Soluble	8-23 potential signal peptide 262-274

The numbers in Table 3 refer to amino acid numbers of postulated transmembrane helices in respective proteins. For the topology prediction programs see chapter 4.6.4.

The localization of Orf15 is not easily predictable; Psort predicted its localization in the cell wall, and SOSUI predicted, that the protein is soluble. This protein has an extraordinary sequence, it has 7 consecutive PVD tripeptides followed by 37 repeats of the PTE tripeptide.

5.13.1 Orf7 is a VirB1 homolog

The sequence of Orf7 was analyzed using Protein Families database of alignments and hidden Markov models (Bateman *et al.*, 2004). The protein consists of two domains – SLT (soluble lytic transglycosylase) and CHAP (cysteine, histidine-dependent amidohydrolase/peptidase) (Figure 31).

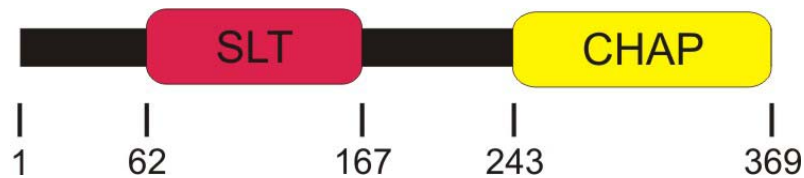


Figure 31. Schematic representation of domain distribution within Orf7. Numbers refer to amino acid residues of Orf7.

5.13.1.1 Expression of Orf7 constructs

Cloning of pGEX-6P-2-*orf7* and pMAL-c2x-*orf7* was performed in the diploma thesis of Karsten Arends in Dr. Grohmann's lab at TU Berlin.

Topology prediction showed a probable transmembrane helix at the N-terminus of the protein, see Table 3. *orf7* lacking the transmembrane helix (*orf7*ΔTM) was cloned into pMAL-c2x (New England Biolabs) and pGEX-6P-2 (Amersham Biosciences) vectors. Trial protein expressions were conducted with both plasmids. Visible expression with pGEX-6P-2-*orf7*ΔTM plasmid was achieved, when expression was induced by 1 mM IPTG at OD₆₀₀ = 0.6, followed by further incubation for 4 hours at 37°C. The GST-Orf7ΔTM protein (63 kDa) was expressed from 1 L *E. coli* BL21 CodonPlus RIL culture and purified using 1 mL GSTrap column (Figure 32).

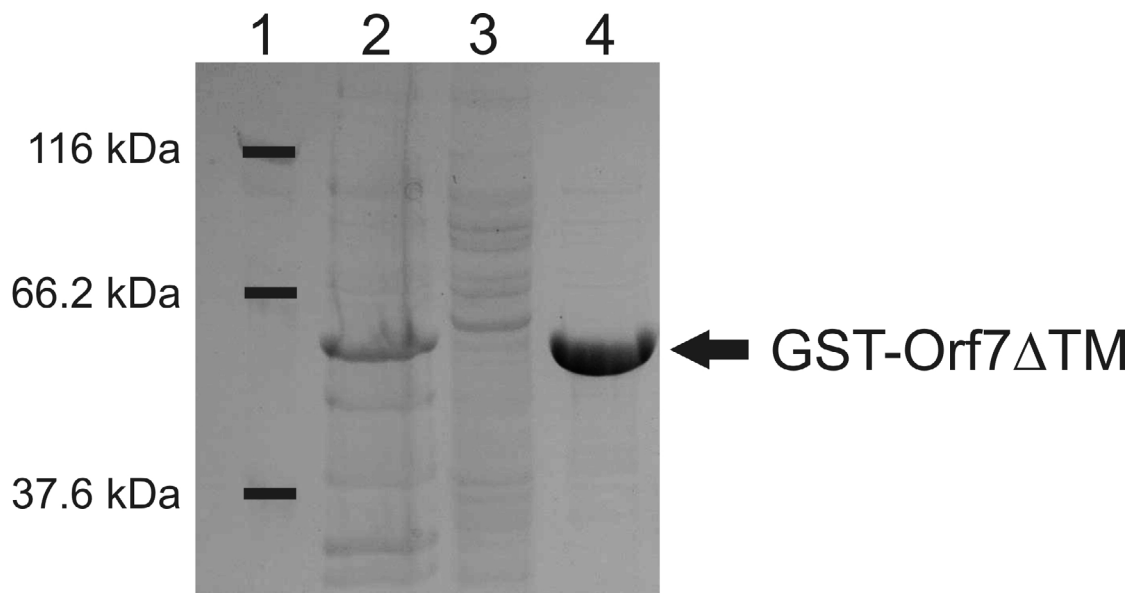


Figure 32. GST-Orf7ΔTM purification. 12 % SDS-PAGE, lane 1 - Protein Molecular Weight Marker (Jena Biosciences), lane 2 – *E. coli* BL21 CodonPlus RIL pGEX-6P-2-*orf7*ΔTM crude cell lysate, lane 3 – flow through after GSTrap column, lane 4 - protein eluted from GSTrap column.

The best Orf7 expression from pMAL-c2x-*orf7* Δ TM was achieved after 2 hours of expression induced by 0.5 mM IPTG at OD₆₀₀ = 0.6. The protein was mostly insoluble, when extracted with 50 mM phosphate buffer, 300 mM NaCl, 1 mM EDTA pH7.5. Other buffers were tried (Lindwall *et al.*, 2000), soluble protein was recovered from buffer containing 100 mM sodium phosphate, 50 mM (NH₄)₂SO₄, 1% Triton X-100, pH 7.5. The MBP-Orf7 Δ TM protein (78 kDa) was purified by affinity chromatography on amylose resin (Figure 33).

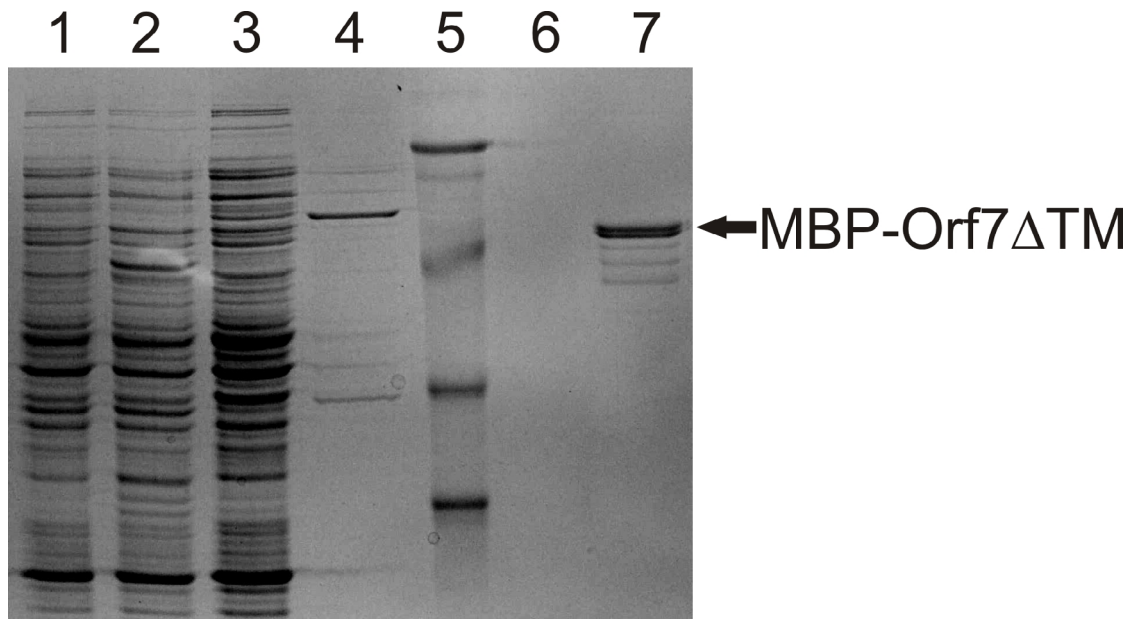


Figure 33. MBP-Orf7 Δ TM purification. Lane 1 – non-induced cells, lane 2 – cell lysate, lane 3 – non-bound fraction, lane 4 –wash 1, lane 5 – Jena Molecular Weight Marker, lane 6 – wash 2, lane 7 – MBP-Orf7 Δ TM eluate

Orf7 was also expressed from pQTEV vector as 7xHis-tag fusion. The first expression experiments resulted in cell lysis upon induction with IPTG. Therefore, the culture was grown to OD₆₀₀ of 1 (late exponential growth phase) before induction and the cells were harvested 2 hours later. The protein (44 kDa) was enriched to ~50 % total protein on Ni²⁺ charged HiTrap Chelating column (Figure 34). Because of the cell lysis, due to Orf7 activity, the protein yield is quite small, from 1 L of culture ~0.5 mg can be extracted.

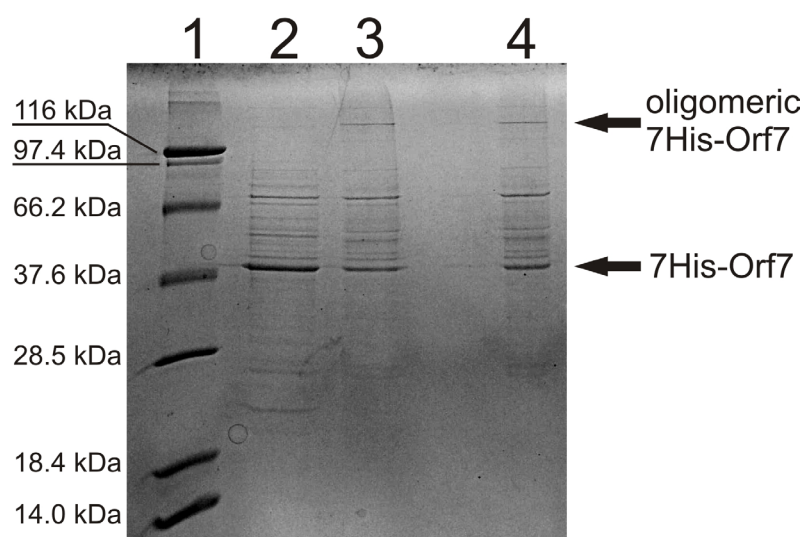


Figure 34. 7His-Orf7 after Ni^{2+} charged HiTrap Chelating column. Lane 1 - Jena Molecular Weight Marker, lane 2 - reduced, denatured sample, lane 3 - non-reduced, denatured sample, lane 4 - non-reduced, 0.02 % SDS sample.

The samples for the SDS-PAGE were prepared in different loading buffers (Figure 34). Thick band representing 7His-Orf7 protein was detected, when the protein sample was reduced with 5 % β -mercaptoethanol and denatured by heat and 1 % SDS (Figure 34, lane 2). This band gets thinner, when the sample was heated with 1 % SDS but no reducing agent, a high molecular weight band, marked “oligomeric 7His-Orf7”, appears instead (Figure 34, lane 3). Similar result was obtained with a protein sample treated only with 0.02 % SDS, but not heated and not reduced (Figure 34, lane 4). This shows that the 7His-Orf7 protein probably forms oligomeric structures in solution.

5.13.1.2 Orf7 has lytic transglycosylase activity

The lytic transglycosylase activity of Orf7 was assayed using zymograms. The gels with copolymerized *E. faecalis* JH2-2 peptidoglycan were run in 25 mM Tris, 20 mM glycine, pH 8.0, containing different concentrations of SDS (0.1 to 0.01 %). The proteins present in the gel were refolded with different buffers (20 mM Tris-malonate pH 8.0, 20 mM sodium phosphate buffer pH 5.5 or 20 mM citrate buffer pH 6.0) containing various concentrations of NaCl (0-200 mM) and/or MgCl_2 (0-20 mM), but the GST-Orf7 ΔTM band always became dark blue, when stained with methylene blue, whereas the positive control (lysozyme) band turned out to be transparent. These experiments show that the GST-Orf7 ΔTM protein does not possess the transglycosylase activity, since only undigested peptidoglycan can be stained with methylene blue. However, MBP-Orf7 ΔTM was able to leave a transparent band on a

zymogram gel, which means that the peptidoglycan was digested (personal communication, Mohammad Yaser Abajy, Dr. Grohmann, TU Berlin).

5.13.2 3D structure prediction of the transglycosylase domain of Orf7

No structure of a transglycosylase originating from Gram + bacteria is known. There are only few structures of lytic transglycosylases originating from Gram - bacteria (*E. coli*, *N. gonorrhoeae*) published (van Asselt *et al.*, 1999a; van Asselt *et al.*, 1999b; van Asselt *et al.*, 2000; Leung *et al.*, 2001). The so far crystallized transglycosylases contain goose-type lysozyme-like domain, derived from goose-egg-white-lysozyme (GEWL) (Koraimann, 2003). Alignment of SLT domain of Orf7 with GEWL and Slt70, *E.coli* transglycosylase, shows three conserved motifs: ES (26-27)-EXXIXXG (72-78)-AYNXGXGXV (108-116) (Figure 35).

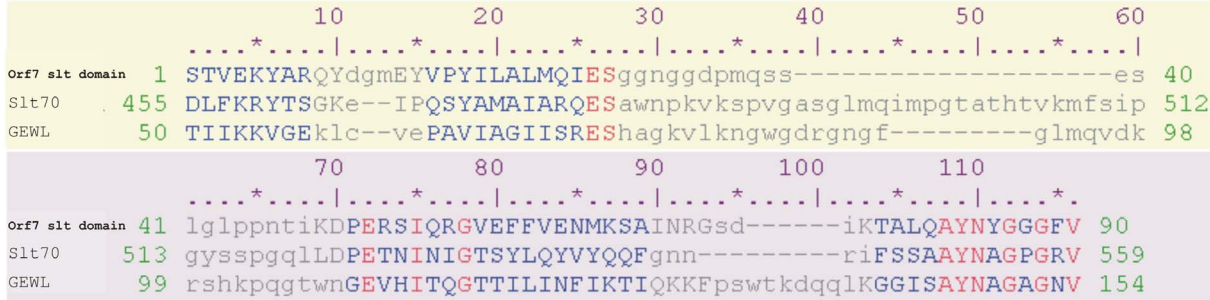


Figure 35. Sequence alignment of the SLT domain of Orf7, Slt70 and GEWL

The structure of the lytic transglycosylase domain of Orf7 was predicted using 3Djigsaw (Bates *et al.*, 2001) using SLT70 (pdb:1qsa) and GEWL (pdb:154l) as model structures.

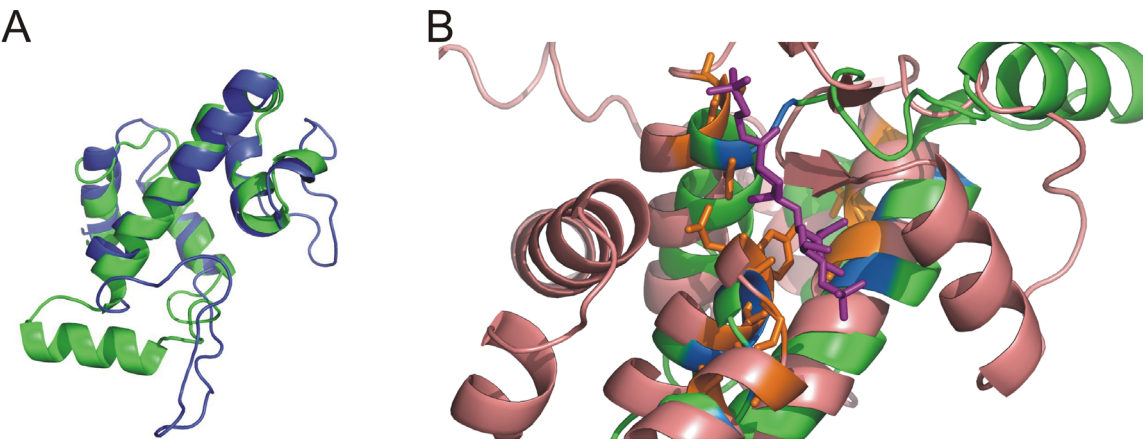


Figure 36. Model of the SLT domain of Orf7. (A) Superimposed models. (B) Active center

The resulting models are presented in Figure 36 A. The model calculated using GEWL is colored blue and the model calculated using Slt70 is green. The biggest difference between

the models lies in one loop present in the GEWL structure which forms an α -helix in the Slt70 structure. The three conserved motifs (Figure 35) are present in the active center of the enzyme. The green structure presented in Figure 36 B, is the modelled SLT domain and the light magenta one is Slt70 complexed with bulgecin A, an inhibitor, colored purple. The conserved motifs are colored blue and orange, respectively.

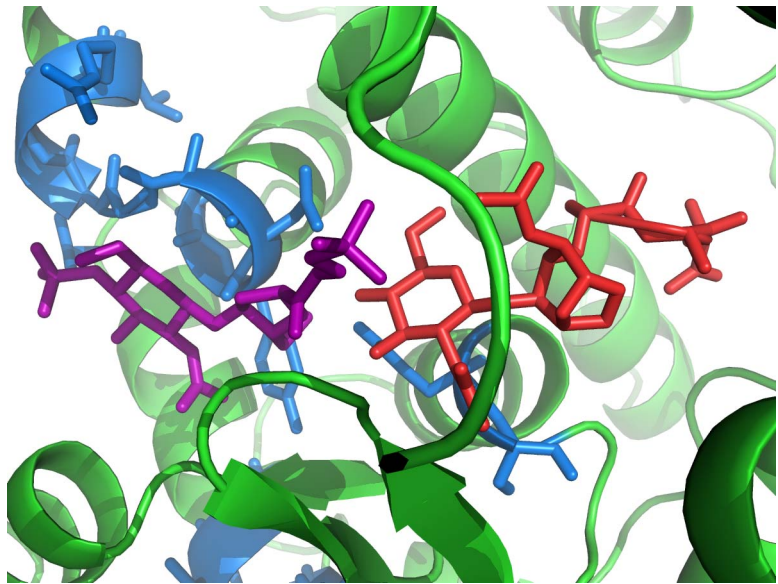


Figure 37. Slt70 binding sites for muropeptide and bulgecin A.

Bulgecin A binds to Slt70 not in the same site as the product of peptidoglycan cleavage (Figure 37). Slt70 structure is presented in Figure 37 in green, the conserved motifs are blue. Bulgecin A (purple) interacts with the motif 3, whereas 1,6-anhydromurotripeptide (red) interacts with motif 1. The binding sites are close to each other, but far enough to consider Bulgecin A as a noncompetitive inhibitor of Slt70 (Templin *et al.*, 1992).

5.13.3 Orf10 is a VirD4 homolog

The amino acid sequence of Orf10 shows homology to the VirD4 protein, encoded by Ti plasmid (Figure 38). VirD4 is the coupling protein, joining the relaxosome with the cell-wall-spanning transfer channel. VirD4 is also one of the putative energetic components required for Ti plasmid DNA transfer; it binds ATP by the Walker A motif (Atmakuri *et al.*, 2004).

VirD4	MNSSKTSPQRM TLSIVCSLAAGFCAASCVTFRRGFNGEAIMTFDVFAFWYETPLYLGYA	60
Orf10	-----	
VirD4	STVFWRGLSVVIFTSLAVLSSQLIISLRNQKHHGTARWAEIGEMQHAGYLQRYSRIGKPI	120
Orf10	-----MTSLLAESDGLILGKFPS-----	18
	:*** . *. **:.	


```

VirD4      FGKTCGPLWFGSYLTNGEQPHSLVVAAPTRAGKGVGVGIVIPTLLTFKG-SVIALDVKGELFE 179
Orf10      -----GKVVIQPEDSKISNRNIFVGGPGSFKTQSYVLPNVVNNRSTSIVVTDPKGEIYE 73
          * : : . : : : : ** . : * . * : * : : . : * : : * * : : *

VirD4      LTSRARKASGDAVFKFSPLDSEKTHCYNPVLDIATLPPERQFTETRRRLAANLITAKGKG 239
Orf10      LTNEIKKAQGFKTVVIN-FKDFLLSSRYNPLLYIRKSNDTNKIANVIVSAKNDPKRKDFW 132
          **.. : **.* .. : : : : : : : : : : : : : : : : : : : : * * . *.

VirD4      AEGFIDGARDLFFVAGILTCIERGTPTIGAVYDLFAQPGEKYK-----LFAQLAEESLNK 293
Orf10      FNAQLN-LLNTLIKVVYFEYEPSARTIEGILDFLEEFDPYNEEGVSELDEQFERLPDGH 191
          : : : : : : : : * : : ** : : * : : : : : : : * * : : . : :

VirD4      EAQRIFDNMAGNDTKILTSTSVLGDGGLNLWADPLIKAATSRSDFSVYDLRRKKTCTIYL 353
Orf10      EAKRSYYLGFRQAQSEARNIVISLLTTLQDFVDKEVSEFTSTNDFFEELGTSKICLYV 251
          **.* : : : . . : : : * : : * : : ** .* . : * . * * : :

VirD4      CVSPNDLEVLAPLMRLLFQQQLVLSILQRS--LPGKDECHEVLFLLDDEFKHLGKLEAETAI 411
Orf10      LISPLDRTWDG-LVNLFQQMFTELYFLGDKHNAKLPVPLVMLLDEFVNLGYFPTYENFL 310
          : ** * . * : : * : : : : * . . : : : * : : * : : * : :

VirD4      TTIAGYKGRFMFIQSLSALTGTYYDAGKQNFLSNTGVQVFMATADDETPTYISKAIGEY 471
Orf10      ATCRGYRISVSTILQSLPQGFELYGDKKFKAIIGNHAIKICLGGVEETTAEYFSRQVNDT 370
          : * ** : . * : ** . * * : : : * : : : : . : : * . * : : :

VirD4      TFKARSTSYSQASMFHDNIQISDQGAALLRPEQVRLDDQSEIVLIK---GRPPLKLRKV 528
Orf10      TIKVYTGGTSESKTSAKTTNRSGSKSESYGQKRRLITEGEVINLQQEENGRKSIVLIDG 430
          * : * . : . * : : . : : * : : : : : * : : : : * * : : * * : :

VirD4      QYYSDRTLKRLFERQMGSLEPAPMLMSDYSNDQVQSHLAEIANFNEDAAPRNRTVAEDH 588
Orf10      KPYMLR--KTPQFELFGNLLKKHEISQQDYISSQTEFAKEYIEELEKQHKQRKVQLASAP 488
          : * * * . : ** : : : . ** ..* : : : : : * : : : : * : :

VirD4      GSVKVGADIPERVMGINGDEDQADAREIPPESVVPPELTALAAQQQLLDQIIALQQRSR 648
Orf10      IFHEKKQEDDLNKKEINLPEQPIESEVTEEE----EEKELSMSDILSSFDEGIEENTANN 544
          : : . ** * : : . * * * : : . : : * : : .

VirD4      SAPAQPAK 656
Orf10      DNSELPF- 551
          . . *

```

Figure 38. Amino acid sequence alignment of Orf10 and VirD4.

The amino acid sequences of the two proteins show a few patterns involved in ATP binding and cleavage, as analyzed with PROSCAN (Combet *et al.*, 2000). Apart from Walker motif A, shown on red background in Figure 38, there are ATP synthase α and β subunits signature (yellow background), with a conservation of 73 % and 71 % in VirD4 and Orf10, respectively, and ATP-binding cassette signature (gray background), conserved in 70 % and 72 %, respectively. Preserved functional motifs suggest similar function of VirD4 and Orf10.

5.13.4 3D structure prediction of Orf10

The structure of the Orf10 homolog, TrwB encoded on plasmid R388 is known (Gomis-Ruth *et al.*, 2001). The structure was used as a model for the 3D structure prediction of Orf10 (Figure 39).

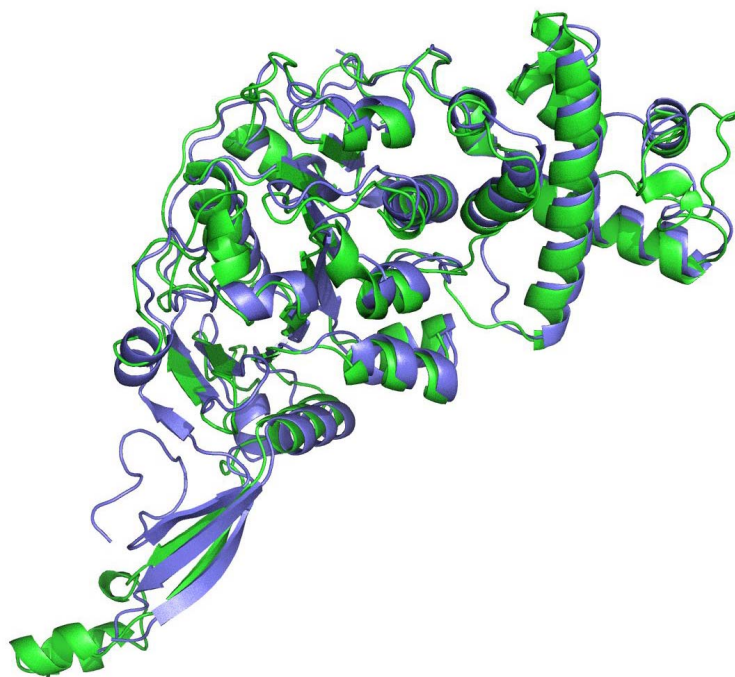


Figure 39. Modelled structure of Orf10. TrwB is the parent structure (blue) of the Orf10 model (green).

The predicted Walker motif A and B in the modelled Orf10 structure are situated in close proximity of the ADP site of TrwB (Figure 40).

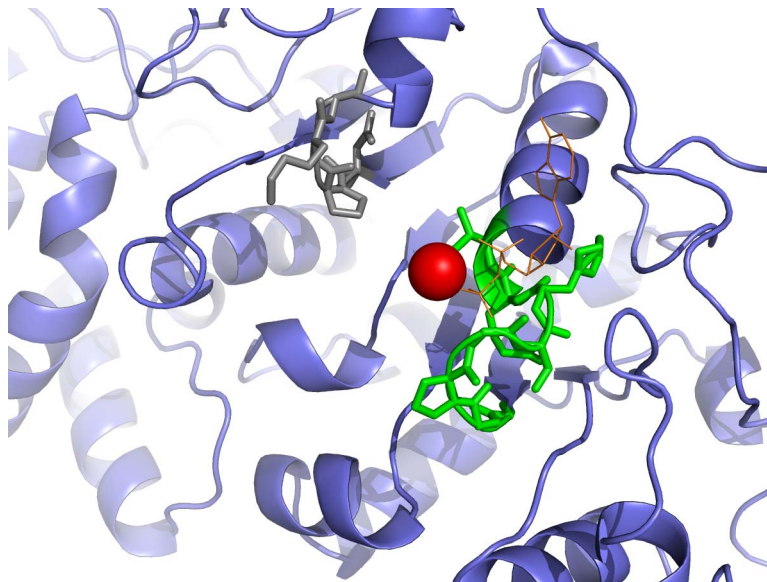


Figure 40. The nucleotide binding site of the modelled Orf10 structure. The Walker motif A is colored green, Walker motif B is colored gray, ADP is orange and Mg^{2+} is a red sphere.

The magnesium ion and ADP molecule were taken from the aligned structure of TrwC complexed with ADP and Mg^{2+} and included into modelled Orf10.

The structural similarity of TrwB with ring helicases suggests that the transferred ss DNA might pass through the central channel of the TrwB hexamer, thereby entering the

translocation apparatus connecting donor and recipient cell. ATP hydrolysis would provide the energy to pump the ss DNA through the TrwB channel (Gomis-Ruth *et al.*, 2001). It was recently shown, that that TrwB is a DNA-dependent ATPase (Tato *et al.*, 2005).

No structural information about interactions of coupling proteins with the relaxosome or mating pair formation proteins is known to date.

5.14 Txt2dic – hand shake between Jasco and dicroprot

The raw data format of the Jasco-910 spectrometer is a binary or ASCII file, left panel of Figure 39.

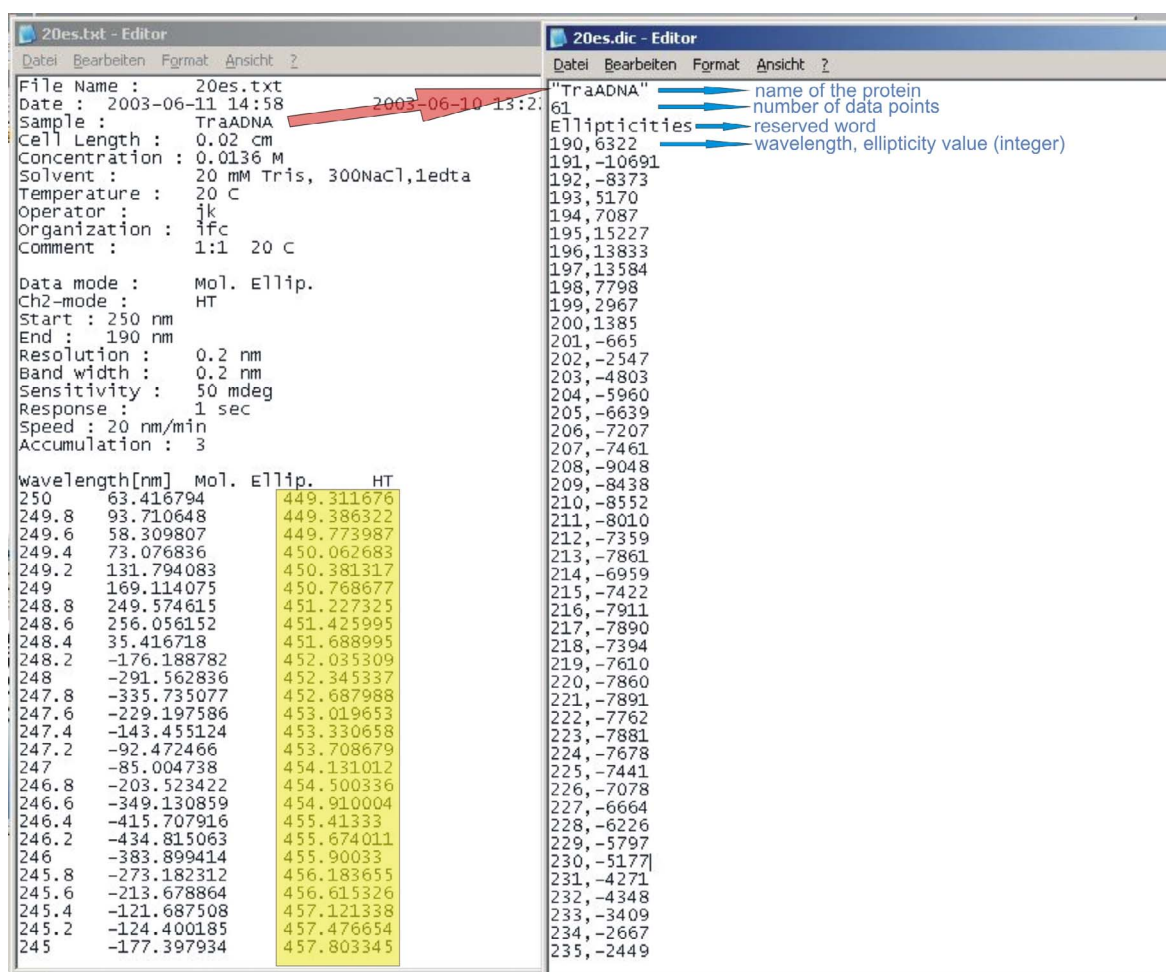


Figure 41. File formats. Example ASCII file resulting from CD measurement (left panel), dic file used as input for dicroprot (right panel). The name of the sample is automatically copied from ASCII file to .dic file (red arrow).

The right panel of Figure 41 shows the input file required for dicroprot 2000 (Deleage and Geourjon, 1993), a secondary structure estimation program. The three first lines have a defined format, the name of the protein must be in the first line, the second line is the number of data points in the file and the third line is the reserved word “ellipticities”. Manual

preparation of such a file from the raw data of the measurement is somewhat laborious. In the input file, the data points have to be integer values of wavelengths and ellipticities. Although the ellipticity values can be easily rounded to integer values in any spreadsheet program, deleting points collected for not integer values of wavelengths [nm] and calculating number of data points, may be problematic, especially when it has to be done for multiple files. The program txt2dic takes care of this task. The user only has to choose the ASCII file to convert, and click the “convert” button (Figure 42).

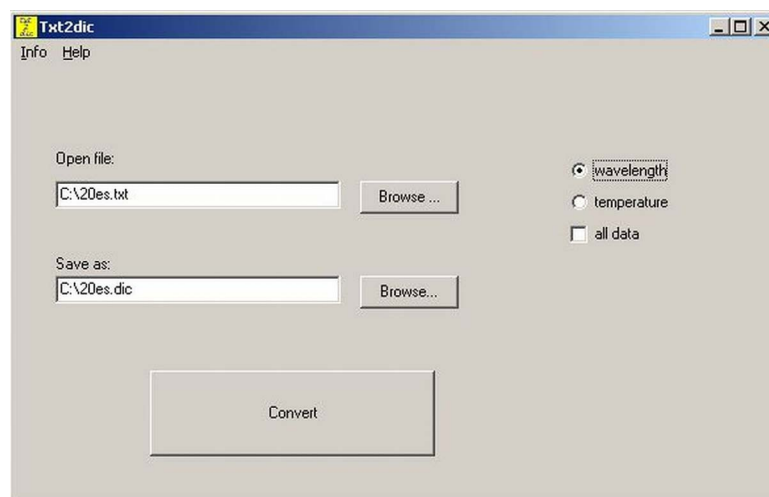


Figure 42. Txt2dic application window.

By clicking the upper “Browse ...” button, the user can choose the txt file to be converted. The program’s input is the text file, see Figure 41, left panel, reads the sample name, and copies it to the first line of resulting file and puts it in “”. Only the data points collected for integer values of wavelengths [nm] are rounded to integer and taken to .dic file. The data points are counted and ordered in ascending order of wavelength. The HT values present in the txt file, marked yellow in the left panel of Figure 41, are omitted.

By default, the resulting file is saved in the same directory and with the same name, but with extension “.dic”.

A .dic file prepared this way can be used as input file for dichroprot 2000 program (Deleage and Geourjon, 1993). With this program, one can analyse secondary structure content of the samples from their CD spectra. It implements 6 ways of deconvoluting CD spectra, one of them is k2D algorithm, which was used to deconvolute the CD spectra of the TraA and TraAN₂₄₆.

6 Conclusions

The relaxase domain of the TraA protein could be limited to 246 N-terminal amino acids. The TraA and TraAN₂₄₆ show similar fold and TraAN₂₄₆ demonstrates higher secondary structure content. The function of the C-terminal part of the TraA protein cannot be determined from amino acid sequence comparison with proteins of known functions (BLAST search), but it was shown that its presence increases the overall activity. The transition temperatures of the two proteins are equal. *oriT*_{pIP501} DNA stabilizes the relaxase domain and induces more fold in the protein. Among oligonucleotides, tested for relaxase binding, the 42-mer oligonucleotide covering an inverted repeat, *nic* site and additional 7 bases showed the best affinity. The full-length protein has lower affinity for the *oriT* than the relaxase domain alone, although the binding of both protein constructs is quite strong. TraA and TraAN₂₄₆ form dimers in solution.

Orf7 proved to possess lytic transglycosylase activity as predicted by BLAST search. This indicates that this protein might be vital for conjugative plasmid pIP501 transfer.

7 Discussion

TraA is the second relaxase characterized from plasmids originating from Gram + organisms. The Gram + conjugative transfer machinery might require fewer components than the type IV secretion systems in Gram - bacteria; presumably because it is easier to cross the barrier of the Gram + cell wall (see Figure 4). Nevertheless, the relaxase, the key enzyme initiating the DNA transfer, seems to follow a quite conserved mechanism in Gram + as well as Gram - bacteria. The 654 amino acid TraA protein from plasmid pIP501 belongs to the IncQ type family of conjugative DNA relaxases. TraA exhibits a two-domain structure like other relaxases, where the relaxase function has been proposed to reside in the N-terminal domain. The function of the C-terminal domain is unknown. It has been shown previously that the N-terminal 293 amino acids of the TraA protein are sufficient for cleavage activity *in vitro* (Kurenbach *et al.*, 2002). The results of *in vitro* cleavage experiments performed in this work show that an even shorter construct, the C-terminal truncation mutant TraAN₂₄₆ contains the functional relaxase domain. In addition, TraAN₂₄₆ exhibits a similar fold and slightly higher secondary structure content than the full length TraA. The analysis of secondary structure content yields a mixed α -helix/ β -sheet fold for both TraA proteins, but with higher β -sheet content for the relaxase domain TraAN₂₄₆. This is in good agreement with the results of sequence based secondary structure predictions, which suggest an α/β fold for the N-terminal domain, but a predominantly α -helical C-terminal domain.

Thermal denaturation curves are very similar for TraA and TraAN₂₄₆ and the transition temperature was nearly identical for both proteins ($T_M = 42^\circ\text{C}$ for TraA and 43°C for TraAN₂₄₆). Interestingly, the temperature range of $42\text{--}45^\circ\text{C}$ was found to be optimal for the transesterase activity of the glutathione-S-transferase-TraA fusion (Kurenbach *et al.*, 2002). Temperatures below 37°C and above 50°C resulted in a significant decrease of cleavage activity (Kurenbach, 2004). The optimization of the gel retardation assays also yielded 42°C as the optimal temperature. Incubation at 30° resulted in no visible binding and at 37°C less DNA was bound by the TraA proteins than at 42°C . Presumably, the elevated temperatures, which trigger a partial unfolding of the relaxase, are necessary to enable the DNA-substrate to enter the active site.

Upon binding of the 42-mer oligonucleotide comprising the inverted repeat and the nick site the relaxase domain exhibited an increased CD signal at 208 nm and the thermal unfolding transition was shifted by 6°C , indicating a stabilization of the relaxase domain.

The first observation suggests a conformational rearrangement of TraAN₂₄₆ upon binding the *oriT* site DNA reminiscent of an induced fit mechanism, which has been described for several DNA-binding proteins, e.g. bZIP transcription factors, where the basic domain is unfolded in the unbound state and only acquires its α -helical conformation upon binding to the DNA (Patel *et al.*, 1990; Weiss *et al.*, 1990) or the conjugation protein TraM, which features an N-terminal recognition helix, that is unfolded in the free state (Verdino *et al.*, 1999). The second observation, the increased thermal stability of the complex, is also a commonly observed phenomenon of DNA binding proteins (Greenfield *et al.*, 2001). The higher transition temperature of the complex as well as the broadening of its unfolding transition (Figure 19B) indicates an intimate contact of the 42-mer oligonucleotide with the putative binding pocket of the relaxase domain.

Comparison of TraA and TraAN₂₄₆ binding to the 42-mer and the hairpin_{nic} oligonucleotides shows that both proteins form two distinct complexes (complex I and II, Figure 26). The highest binding affinity was observed for the 42-mer oligonucleotide. Deletion of the seven bases proximal to the nick site results in a 15-fold reduction of TraAN₂₄₆ binding affinity, suggesting that this region comprises a part of the *oriT* which is crucial for high affinity relaxase binding. Deletion of another seven bases (5' to the nick site) generates the hairpin oligonucleotide which shows a further ~2.5 fold reduction of the affinity to the relaxase protein. Similar results were published for the relaxase domain of TrwC (Guasch *et al.*, 2003), whose binding affinities to the hairpin and a corresponding hairpin_{nic} oligonucleotide differ ~7 fold.

Gel retardation experiments with idealized hairpins showed that the bulge in the original imperfect hairpin is required for efficient specific binding of the protein: The A \rightleftharpoons C exchange in the right half repeat resulted in a 4 fold reduction of TraAN₂₄₆ binding affinity. High molecular mass DNA-protein complexes are formed. At high protein concentrations complex I is visible. An even more drastic effect on protein binding is exerted by the G \rightleftharpoons T exchange in the left half repeat. Only applying high protein concentrations (≥ 2.5 μ M TraA and ≥ 3.75 μ M TraAN₂₄₆) to this idealized hairpin results in visible DNA-binding. However, exclusively high molecular mass protein-DNA complexes not entering the gel are formed. Since neither complex I nor II occur, the guanine in the left half repeat of the IR might play an important role in specific binding.

Unexpectedly, the relaxase domain TraAN₂₄₆ had a higher binding affinity than the full-length TraA protein with all specific oligonucleotides tested in this study (Figures 26 B and D). The

reason for this difference may be either the reduced accessibility of the binding site in the relaxase domain due to steric interference by the C-terminal domain of TraA or a conformational change in the binding site induced by the larger C-terminal domain.

The investigation of the aggregation state of the relaxase domain and the whole TraA protein by chemical cross-linking and small angle X-ray scattering proved that TraAN₂₄₆ as well as TraA form dimers in solution at concentrations used for the experiments (16 - 39 μ M in the case of TraAN₂₄₆ and 6.4 - 59 μ M in the case of TraA). Provided that the dimerization domain is opposite to the DNA-binding site of the relaxase, these results could explain the occurrence of two complex bands in the EMSA experiments, which were performed at much lower protein concentrations than the dimerization experiments.

Furthermore, the existence of TraA dimers has important implications for the mechanism of this enzyme, especially for its proposed activity of religation of the nicked plasmid: The product of the cleavage reaction is a covalent adduct of the relaxase bound to the 5'-phosphoryl at the nick site (Pansegrau and Lanka, 1996b). Genetic evidence for a rolling-circle-replication-like (RCR) mechanism was presented for the transfer of the cleaved strand in IncF and IncQ systems (Gao *et al.*, 1994; Rao and Meyer, 1994). This mechanism requires a second cleavage reaction in order to generate a free 3'-hydroxyl group for religation (Wilkins and Lanka, 1993). Relaxases of the IncF and IncW type feature a tandem repeat of tyrosines which are both active in the DNA cleavage reaction (Zechner *et al.*, 2000). For the relaxase TrwC of plasmid R388 de la Cruz and coworkers have shown that two out of the four tyrosines are actively involved in cleavage and strand transfer of linear oligonucleotides, but only the first (Y18) is able to cleave supercoiled DNA containing the R388 *oriT* (Grandoso *et al.*, 2000). From these data the authors have developed a model, where the first active tyrosine (Y18) catalyses the initial cleavage reaction, whereas the other (Y26) is responsible for the second cleavage and thus the termination of conjugation. According to this model relaxases from the IncF and IncW type are active in initiation as well as termination of conjugation as monomers.

The relaxases TraI (RP4) and MobA (RSF1010) of the IncP and IncQ families, on the other hand, have been shown to carry only one active site tyrosine for the DNA strand transfer reaction (Pansegrau *et al.*, 1993; Scherzinger *et al.*, 1993). Therefore, the monomeric form of these relaxases is insufficient for the second cleavage and it was postulated for TraI of RP4 that it acts as a dimer in termination of transfer DNA replication (Pansegrau and Lanka, 1996b). The tendency of the full length TraA and the TraA relaxase domain to form stable dimers is consistent with the requirements of the second cleavage to be performed in close

proximity to the site of the first cleavage/transesterification reaction in order to enable efficient DNA-religation at termination of strand transfer. For a precise description of the spatial arrangement of the active sites within the dimer and the elucidation of the reaction mechanism 3D-structure information is needed. Using the data obtained from this work crystallization trials of the relaxase domain alone and its specific DNA complexes have been initiated. The TrwC protein was crystallized with an oligonucleotide covering the inverted repeat and the *nic* site (Guasch *et al.*, 2003). An equivalent oligonucleotide, hairpin_*nic* (Figure 24), was used for crystallization trials of the TraA-oligonucleotide complex, resulting in a non-reproducible and non-diffracting single crystal.

The cleavage assay employing RSF1010 plasmid and TraA turned out to be negative (Figure 23). This could be explained by the geometry of the respective hairpin (Figure 17). The RSF1010 hairpin is perfect and shorter, in comparison to the pIP501 hairpin, which has a bulge and is longer. Because of the unpaired nucleotides, the pIP501 hairpin might be bent, and this might be an important feature for TraA binding. The –GAA– triplet in the loop of the non-cleaved strand is protected from DNase I digestion by TraA (Figure 29). Therefore, the symmetry of the protein dimer is broken upon DNA binding.

The G present in the bulge of the imperfect hairpin of *oriT*_{pIP501} must be important for TraA binding because in the DNase I footprint assay it is protected by TraA (Figure 29), and when exchanged to T (Figure 24), no specific binding by TraA and TraAN₂₄₆ is observable.

The two functions of TraA, nicking of the transferred strand and regulation of the *tra* operon transcription are exerted at a pretty short DNA fragment (Figure 24). The two functions are moderated by the temperature. The cleavage reaction is very inefficient at 37°C, and the maximum of cleavage is achieved at 42-45°C (Kurenbach, 2004). The EMSAs with the oligonucleotides covering the hairpin structure were performed at 42°C, because the binding was undetectable at 37°C, whereas the binding to the promotor region is quite strong at 37°C. This implies existence of two binding sites for the relaxase on *oriT*_{pIP501}. This may be due to partial unfolding of the protein, the T_M of TraA is 42°C and of TraAN₂₄₆-42mer complex is 48°C. The secondary structure of the *oriT* (Figure 24), may also be altered due to temperature change.

The characterization of other proteins encoded on *tra* operon of pIP501 has been also carried out. The Orf7 protein, a lytic transglycosylase, was expressed and purified as GST, 7His and MBP tag fusion proteins. The GST-Orf7 Δ TM construct turned out to be inactive in zymogram assays. This might be due to false disulfide bridge formation. Orf7 has a putative disulfide

bridge between Cys₂₆₃ and Cys₃₄₃, as predicted by DISULFIND (Vullo and Frasconi, 2004), but a false Cys-Cys is formed in GST-Orf7 Δ TM protein, where the disulfide bridge is formed between Cys present in GST and Cys₂₆₃ of Orf7, causing a wrong fold and inactivity of the protein. The MBP-Orf7 Δ TM turned out to be active in a zymogram test (Mohammad Yaser Abajy, personal communication), as well as in radioactive muramidase assays. During the expression of 7His-Orf7 the *E. coli* XL10 gold cells lyse, this can be an indirect evidence of Orf7 lytic transglycosylase activity.

The calculated models of Orf10 and the Slt domain of Orf7 can be used as search models for molecular replacement solution of Orf10 and Orf7 crystal structures.

8 Outlook

Crystallization trials of the full-length TraA as well as TraAN₂₄₆ will be carried on. To facilitate the crystallization, the binding partner will be co-crystallized, this may be a ds DNA fragment covering -10 promotor region or a protein partner binding to the relaxase with high affinity. From protein-protein interaction assays performed in Dr. Grohmann's lab the putative protein partner will be selected. Initial experiments show strong interaction between TraA and Orf10. Before setting up crystallization trials, the correct stoichiometric molar ratio has to be carefully determined (Stura *et al.*, 2001).

The other components of the *tra* operon will be also considered for structural studies. Computer-assisted prediction of transmembrane helices helped to design soluble protein constructs. The protein constructs will also be cloned into variety of expression vectors, maltose binding protein is known to increase solubility of proteins (Goh *et al.*, 2003). Solubility of the proteins can be also enhanced by point mutations, exchanging cysteins for serins prevents the protein from improper folding due to disulfide bridges formation, lysine residues present on protein surface may disturb crystal contacts and can be exchanged by alanine residues (Derewenda, 2004). Large-scale expression of soluble constructs will be performed. The purification of the proteins is simplified due to the tags (7His, GST or MBP). There are specific protease cleavage sites between the tag and the target protein in the proteins expressed from pQTEV, pGEX-6P-2 and pMAL-c2x expression plasmids. pQTEV inserts tobacco etch virus (TEV) protease cleavage site, pGEX-6P-2 inserts PreScission[®] protease (GE Healthcare) cleavage site and pMAL produces Factor Xa cleavage site. The tags will be removed, the optimal cleavage conditions will be found by changing the cleavage buffer and

protein:protease ratio (Gruswitz *et al.*, 2005). The proteins alone and/or with binding partners will be set up for crystallization.

The proteins will also be subjected to biochemical characterization. The activity assay of Orf5 and Orf10, putative ATPases, will be performed using different methods. For example in radiometric analysis, [γ^{32} -P] labeled ATP is added to the protein, the hydrolysis reaction is stopped with EDTA and the products are spotted onto a polyethylenimine-cellulose plate and thin layer chromatography (TLC) is developed in LiCl/formic acid. The efficiency of the reaction can be estimated by quantitation of radioactivity distribution on the TLC plate (Weinstock *et al.*, 1981).

The proteins, activity of which cannot be guessed from amino acid sequence, will also be expressed, purified and queued for structural studies. When their structures are known, proteins with similar fold may be found using DALI server (<http://www.ebi.ac.uk/dali/>) (Holm and Sander, 1995; Holm and Sander, 1999), and the function shall be derived from the 3D structure.

This will provide insight into pIP501 plasmid transfer on atomic level.

The biological functionality of each *orf* in plasmid transfer can be determined in conjugation assays performed with deletion mutants. Plasmids lacking 1 of the 15 *orfs* can be inspected with respect to their conjugative transfer. Proteins, which are necessary for transfer, can be determined from these experiments.

When the structures of the interacting proteins are known, new point mutants can be designed, in which the interaction will be disturbed. The biological significance of the mutation can be then examined *in vivo*.

9 References

- Albrich, W. C., Monnet, D. L. and Harbarth, S. (2004). Antibiotic selection pressure and resistance in *Streptococcus pneumoniae* and *Streptococcus pyogenes*. *Emerg Infect Dis* **10** (3): 514-7.
- Andrade, M. A., Chacon, P., Merelo, J. J. and Moran, F. (1993). Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. *Protein Eng* **6** (4): 383-90.
- Arbeloa, A., Hugonnet, J. E., Sentilhes, A. C., Josseaume, N., Dubost, L., Monsempes, C., Blanot, D., Brouard, J. P. and Arthur, M. (2004). Synthesis of mosaic peptidoglycan cross-bridges by hybrid peptidoglycan assembly pathways in gram-positive bacteria. *J Biol Chem* **279** (40): 41546-56.
- Atmakuri, K., Cascales, E. and Christie, P. J. (2004). Energetic components VirD4, VirB11 and VirB4 mediate early DNA transfer reactions required for bacterial type IV secretion. *Mol Microbiol* **54** (5): 1199-211.
- Bateman, A., Coin, L., Durbin, R., Finn, R. D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L., Studholme, D. J., Yeats, C. and Eddy, S. R. (2004). The Pfam protein families database. *Nucleic Acids Res* **32** (Database issue): D138-41.
- Bates, P. A., Kelley, L. A., Maccallum, R. M. and Sternberg, M. J. (2001). Enhancement of protein modeling by human intervention in applying the automatic programs 3D-JIGSAW and 3D-PSSM. *Proteins Suppl* **5**: 39-46.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Res* **28** (1): 235-42.
- Bernardi, A. and Bernardi, F. (1984). Complete sequence of pSC101. *Nucleic Acids Res* **12** (24): 9415-26.
- Bhikhabhai, R., Ollivier, M. and Blanche, F. (2000). A novel, rapid process for purification of plasmids for gene therapy. Life Science News Amersham Biosciences.
- Birnboim, H. C. (1983). A rapid alkaline extraction method for the isolation of plasmid DNA. *Methods Enzymol* **100**: 243-55.
- Bouhss, A., Josseaume, N., Allanic, D., Crouvoisier, M., Gutmann, L., Mainardi, J. L., Mengin-Lecreulx, D., Van Heijenoort, J. and Arthur, M. (2001). Identification of the UDP-MurNAc-pentapeptide:L-alanine ligase for synthesis of branched peptidoglycan precursors in *Enterococcus faecalis*. *J Bacteriol* **183** (17): 5122-7.
- Bouhss, A., Josseaume, N., Severin, A., Tabei, K., Hugonnet, J. E., Shlaes, D., Mengin-Lecreulx, D., Van Heijenoort, J. and Arthur, M. (2002). Synthesis of the L-alanyl-L-alanine cross-bridge of *Enterococcus faecalis* peptidoglycan. *J Biol Chem* **277** (48): 45935-41.
- Brasch, M. A. and Meyer, R. J. (1986). Genetic organization of plasmid R1162 DNA involved in conjugative mobilization. *J Bacteriol* **167** (2): 703-10.
- Byrd, D. R. and Matson, S. W. (1997). Nicking by transesterification: the reaction catalysed by a relaxase. *Mol Microbiol* **25** (6): 1011-22.
- Cabezón, E., Sastre, J. I. and De La Cruz, F. (1997). Genetic evidence of a coupling role for the TraG protein family in bacterial conjugation. *Mol Gen Genet* **254** (4): 400-6.
- Cascales, E. and Christie, P. J. (2004). Definition of a bacterial type IV secretion pathway for a DNA substrate. *Science* **304** (5674): 1170-3.

- Chacon, P., Diaz, J. F., Moran, F. and Andreu, J. M. (2000). Reconstruction of protein form with X-ray solution scattering and a genetic algorithm. *J Mol Biol* **299** (5): 1289-302.
- Chacon, P., Moran, F., Diaz, J. F., Pantos, E. and Andreu, J. M. (1998). Low-resolution structures of proteins in solution retrieved from X-ray scattering with a genetic algorithm. *Biophys J* **74** (6): 2760-75.
- Chou, P. Y. and Fasman, G. D. (1974). Prediction of protein conformation. *Biochemistry* **13** (2): 222-45.
- Christie, P. J. and Cascales, E. (2005). Structural and dynamic properties of bacterial type IV secretion systems (review). *Mol Membr Biol* **22** (1-2): 51-61.
- Claros, M. G. and Von Heijne, G. (1994). TopPred II: an improved software for membrane protein structure predictions. *Comput Appl Biosci* **10** (6): 685-6.
- Clewell, D. B. and Francia, M. V. (2004). Conjugation in Gram-Positive Bacteria. Plasmid Biology. G. J. Philips. Washington, DC, ASM Press 227-56.
- Climo, M. W., Sharma, V. K. and Archer, G. L. (1996). Identification and characterization of the origin of conjugative transfer (oriT) and a gene (nes) encoding a single-stranded endonuclease on the staphylococcal plasmid pGO1. *J Bacteriol* **178** (16): 4975-83.
- Combet, C., Blanchet, C., Geourjon, C. and Deleage, G. (2000). NPS@: network protein sequence analysis. *Trends Biochem Sci* **25** (3): 147-50.
- Cook, D. M. and Farrand, S. K. (1992). The oriT region of the *Agrobacterium tumefaciens* Ti plasmid pTiC58 shares DNA sequence identity with the transfer origins of RSF1010 and RK2/RP4 and with T-region borders. *J Bacteriol* **174** (19): 6238-46.
- Coyette, J. and Hancock, L. E. (2002). Enterococcal cell wall *In* M.S. Gilmore (ed.), The Enterococci Pathogenesis, Molecular Biology, and Antibiotic Resistance, ASM Press, Washington, D.C., U.S.A.: 177-218.
- Cserzo, M., Wallin, E., Simon, I., Von Heijne, G. and Elofsson, A. (1997). Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. *Protein Eng* **10** (6): 673-6.
- Cuff, J. A., Clamp, M. E., Siddiqui, A. S., Finlay, M. and Barton, G. J. (1998). JPred: a consensus secondary structure prediction server. *Bioinformatics* **14** (10): 892-3.
- Datta, S., Larkin, C. and Schildbach, J. F. (2003). Structural insights into single-stranded DNA binding and cleavage by F factor TraI. *Structure (Camb)* **11** (11): 1369-79.
- De Paz, H. D., Sangari, F. J., Bolland, S., Garcia-Lobo, J. M., Dehio, C., De La Cruz, F. and Llosa, M. (2005). Functional interactions between type IV secretion systems involved in DNA transfer and virulence. *Microbiology* **151** (Pt 11): 3505-16.
- Deleage, G. and Geourjon, C. (1993). An interactive graphic program for calculating the secondary structure content of proteins from circular dichroism spectrum. *Comput Appl Biosci* **9** (2): 197-9.
- Derewenda, Z. S. (2004). Rational protein crystallization by mutational surface engineering. *Structure* **12** (4): 529-35.
- Ding, Z., Atmakuri, K. and Christie, P. J. (2003). The outs and ins of bacterial type IV secretion substrates. *Trends Microbiol* **11** (11): 527-35.
- Drolet, M., Zanga, P. and Lau, P. C. (1990). The mobilization and origin of transfer regions of a *Thiobacillus ferrooxidans* plasmid: relatedness to plasmids RSF1010 and pSC101. *Mol Microbiol* **4** (8): 1381-91.
- Evans, R. P., Jr. and Macrina, F. L. (1983). Streptococcal R plasmid pIP501: endonuclease site map, resistance determinant location, and construction of novel derivatives. *J Bacteriol* **154** (3): 1347-55.
- Fernandez-Lopez, R., Machon, C., Longshaw, C. M., Martin, S., Molin, S., Zechner, E. L., Espinosa, M., Lanka, E. and De La Cruz, F. (2005). Unsaturated fatty acids are inhibitors of bacterial conjugation. *Microbiology* **151** (Pt 11): 3517-26.

- Frost, L. S., Ippen-Ihler, K. and Skurray, R. A. (1994). Analysis of the sequence and gene products of the transfer region of the F sex factor. *Microbiol. Rev.* **58** (2): 162-210.
- Gao, Q., Luo, Y. and Deonier, R. C. (1994). Initiation and termination of DNA transfer at F plasmid oriT. *Mol Microbiol* **11** (3): 449-58.
- Gardy, J. L., Laird, M. R., Chen, F., Rey, S., Walsh, C. J., Ester, M. and Brinkman, F. S. (2005). PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. *Bioinformatics* **21** (5): 617-23.
- Goh, L. L., Loke, P., Singh, M. and Sim, T. S. (2003). Soluble expression of a functionally active Plasmodium falciparum falcipain-2 fused to maltose-binding protein in Escherichia coli. *Protein Expr Purif* **32** (2): 194-201.
- Gomis-Ruth, F. X. and Coll, M. (2001). Structure of TrwB, a gatekeeper in bacterial conjugation. *Int J Biochem Cell Biol* **33** (9): 839-43.
- Gomis-Ruth, F. X., Moncalian, G., Perez-Luque, R., Gonzalez, A., Cabezon, E., De La Cruz, F. and Coll, M. (2001). The bacterial conjugation protein TrwB resembles ring helicases and F1-ATPase. *Nature* **409** (6820): 637-41.
- Gotz, A., Pukall, R., Smit, E., Tietze, E., Prager, R., Tschape, H., Van Elsas, J. D. and Smalla, K. (1996). Detection and characterization of broad-host-range plasmids in environmental bacteria by PCR. *Appl Environ Microbiol* **62** (7): 2621-8.
- Grandoso, G., Avila, P., Cayon, A., Hernando, M. A., Llosa, M. and De La Cruz, F. (2000). Two active-site tyrosyl residues of protein TrwC act sequentially at the origin of transfer during plasmid R388 conjugation. *J Mol Biol* **295** (5): 1163-72.
- Greenfield, N., Vijayanathan, V., Thomas, T. J., Gallo, M. A. and Thomas, T. (2001). Increase in the stability and helical content of estrogen receptor alpha in the presence of the estrogen response element: analysis by circular dichroism spectroscopy. *Biochemistry* **40** (22): 6646-52.
- Grohmann, E. (2005). Mating cell-cell channels in conjugating bacteria. Cell-Cell Channels. F. Baluska, D. Volkmann and P. W. Barlow. Georgetown, Texas, Landes Biosciences.
- Grohmann, E., Muth, G. and Espinosa, M. (2003). Conjugative plasmid transfer in gram-positive bacteria. *Microbiol Mol Biol Rev* **67** (2): 277-301.
- Gruswitz, F., Frishman, M., Goldstein, B. M. and Wedekind, J. E. (2005). Coupling of MBP fusion protein cleavage with sparse matrix crystallization screens to overcome problematic protein solubility. *Biotechniques* **39** (4): 476, 8, 80.
- Guasch, A., Lucas, M., Moncalian, G., Cabezas, M., Perez-Luque, R., Gomis-Ruth, F. X., De La Cruz, F. and Coll, M. (2003). Recognition and processing of the origin of transfer DNA by conjugative relaxase TrwC. *Nat Struct Biol* **10** (12): 1002-10.
- Hofmann, K. and Stoffel, W. (1993). TMBASE - A database of membrane spanning protein segments. *Biol Chem Hoppe-Seyler* **374**: 166.
- Holm, L. and Sander, C. (1995). Dali: a network tool for protein structure comparison. *Trends Biochem Sci* **20** (11): 478-80.
- Holm, L. and Sander, C. (1999). Protein folds and families: sequence and structure alignments. *Nucleic Acids Res* **27** (1): 244-7.
- Horodniceanu, T., Bouanchaud, D. H., Bieth, G. and Chabbert, Y. A. (1976). R plasmids in Streptococcus agalactiae (group B). *Antimicrob Agents Chemother* **10** (5): 795-801.
- Ilyina, T. V. and Koonin, E. V. (1992). Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. *Nucleic Acids Res* **20** (13): 3279-85.
- Jacob, A. E. and Hobbs, S. J. (1974). Conjugal transfer of plasmid-borne multiple antibiotic resistance in Streptococcus faecalis var. zymogenes. *J Bacteriol* **117** (2): 360-72.

- Johns, E. W. (1982). The HMG Chromosomal Proteins. History, definitions and problems. E. W. Johns. London and New York, Academic Press, 1–9.
- Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* **292** (2): 195-202.
- Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. and Svergun, D. I. (2003). PRIMUS: A Windows PC-based system for small-angle scattering data analysis. *J Appl Cryst* **36** (5): 1277-82.
- Koraimann, G. (2003). Lytic transglycosylases in macromolecular transport systems of Gram-negative bacteria. *Cell Mol Life Sci* **60** (11): 2371-88.
- Krah, E. R., 3rd and Macrina, F. L. (1989). Genetic analysis of the conjugal transfer determinants encoded by the streptococcal broad-host-range plasmid pIP501. *J Bacteriol* **171** (11): 6005-12.
- Kurenbach, B. (2004) Konjugativer DNA Transfer zwischen Gram-positiven und Gram-negativen Bakterien: Transferkomponenten des Multiresistenzplasmids pIP501 aus *Streptococcus agalactiae*, Ph.D. Thesis, Technische Universität Berlin
- Kurenbach, B., Bohn, C., Prabhu, J., Abudukerim, M., Szewzyk, U. and Grohmann, E. (2003). Intergeneric transfer of the Enterococcus faecalis plasmid pIP501 to Escherichia coli and Streptomyces lividans and sequence analysis of its tra region. *Plasmid* **50** (1): 86-93.
- Kurenbach, B., Grothe, D., Farias, M. E., Szewzyk, U. and Grohmann, E. (2002). The tra region of the conjugative plasmid pIP501 is organized in an operon with the first gene encoding the relaxase. *J Bacteriol* **184** (6): 1801-5.
- Kurenbach, B., Kopec, J., Magdefrau, M., Andreas, K., Keller, W., Bohn, C., Abajy, M. Y. and Grohmann, E. (2006). The TraA relaxase autoregulates the putative type IV secretion-like system encoded by the broad-host-range Streptococcus agalactiae plasmid pIP501. *Microbiology* **152** (3): 637-45.
- Lanka, E. and Wilkins, B. M. (1995). DNA processing reactions in bacterial conjugation. *Annu Rev Biochem* **64**: 141-69.
- Larkin, C., Datta, S., Harley, M. J., Anderson, B. J., Ebie, A., Hargreaves, V. and Schildbach, J. F. (2005). Inter- and intramolecular determinants of the specificity of single-stranded DNA binding and cleavage by the F factor relaxase. *Structure (Camb)* **13** (10): 1533-44.
- Lauer, P., Rinaudo, C. D., Soriani, M., Margarit, I., Maione, D., Rosini, R., Taddei, A. R., Mora, M., Rappuoli, R., Grandi, G. and Telford, J. L. (2005). Microbiology: Genome analysis reveals pili in group B streptococcus. *Science* **309** (5731): 105.
- Leblanc, B. and Moss, T. (1994). DNase I footprinting. *Methods in molecular biology (Clifton, N.J.)* **30**: 1-10.
- Leblanc, D. J., Lee, L. N. and Abu-Al-Jaibat, A. (1992). Molecular, genetic, and functional analysis of the basic replicon of pVA380-1, a plasmid of oral streptococcal origin. *Plasmid* **28** (2): 130-45.
- Leung, A. K., Duewel, H. S., Honek, J. F. and Berghuis, A. M. (2001). Crystal structure of the lytic transglycosylase from bacteriophage lambda in complex with hexa-N-acetylchitohexaose. *Biochemistry* **40** (19): 5665-73.
- Lindwall, G., Chau, M.-F., Gardner, S. R. and Kohlstaedt, L. A. (2000). A sparse matrix approach to the solubilization of overexpressed proteins. *Protein Eng* **13** (1): 67-71.
- Llosa, M., Bolland, S. and De La Cruz, F. (1994). Genetic organization of the conjugal DNA processing region of the IncW plasmid R388. *J Mol Biol* **235** (2): 448-64.
- Llosa, M., Gomis-Ruth, F. X., Coll, M. and De La Cruz Fd, F. (2002). Bacterial conjugation: a two-step mechanism for DNA transport. *Mol Microbiol* **45** (1): 1-8.

- Llosa, M., Grandoso, G., Hernando, M. A. and De La Cruz, F. (1996). Functional domains in protein TrwC of plasmid R388: dissected DNA strand transferase and DNA helicase activities reconstitute protein function. *J Mol Biol* **264** (1): 56-67.
- Lusty, C. J. (1999). A gentle vapor-diffusion technique for cross-linking of protein crystals for cryocrystallography. *Journal of Applied Crystallography* **32** (1): 106-12.
- Mcguffin, L. J., Bryson, K. and Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics* **16** (4): 404-5.
- Medina-Acosta, E. and Cross, G. A. (1993). Rapid isolation of DNA from trypanosomatid protozoa using a simple 'mini-prep' procedure. *Mol Biochem Parasitol* **59** (2): 327-9.
- Merelo, J. J., Andrade, M. A., Prieto, A. and Moran, F. (1994). Proteinotopic feature maps. *Neurocomputing* **6** (4): 443-54.
- Migneault, I., Dartiguenave, C., Vinh, J., Bertrand, M. J. and Waldron, K. C. (2004). Comparison of two glutaraldehyde immobilization techniques for solid-phase tryptic peptide mapping of human hemoglobin by capillary zone electrophoresis and mass spectrometry. *Electrophoresis* **25** (9): 1367-78.
- Milpetz, F., Argos, P. and Persson, B. (1995). TMAP: a new email and WWW service for membrane-protein structural predictions. *Trends Biochem Sci* **20** (5): 204-5.
- Moncalian, G., Grandoso, G., Llosa, M. and De La Cruz, F. (1997). oriT-processing and regulatory roles of TrwA protein in plasmid R388 conjugation. *J Mol Biol* **270** (2): 188-200.
- Mora, M., Bensi, G., Capo, S., Falugi, F., Zingaretti, C., Manetti, A. G., Maggi, T., Taddei, A. R., Grandi, G. and Telford, J. L. (2005). Group A Streptococcus produce pilus-like structures containing protective antigens and Lancefield T antigens. *Proc Natl Acad Sci U S A* **102** (43): 15641-6.
- Notredame, C., Higgins, D. G. and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302** (1): 205-17.
- Pansegrau, W. and Lanka, E. (1996a). Enzymology of DNA transfer by conjugative mechanisms. *Prog Nucleic Acid Res Mol Biol* **54**: 197-251.
- Pansegrau, W. and Lanka, E. (1996b). Mechanisms of initiation and termination reactions in conjugative DNA processing. Independence of tight substrate binding and catalytic activity of relaxase (TraI) of IncPalph plasmid RP4. *J Biol Chem* **271** (22): 13068-76.
- Pansegrau, W., Schroder, W. and Lanka, E. (1993). Relaxase (TraI) of IncP alpha plasmid RP4 catalyzes a site-specific cleaving-joining reaction of single-stranded DNA. *Proc Natl Acad Sci U S A* **90** (7): 2925-9.
- Patel, L., Abate, C. and Curran, T. (1990). Altered protein conformation on DNA binding by Fos and Jun. *Nature* **347** (6293): 572-5.
- Persson, B. and Argos, P. (1996). Topology prediction of membrane proteins. *Protein Sci* **5** (2): 363-71.
- Persson, B. and Argos, P. (1997). Prediction of membrane protein topology utilizing multiple sequence alignments. *J Protein Chem* **16** (5): 453-7.
- Peyret, N. (2000). Prediction of Nucleic Acid Hybridization: Parameters and Algorithms. *PhD dissertation, Wayne State University, Department of Chemistry, Detroit, MI.*
- Rao, X. M. and Meyer, R. J. (1994). Conjugal mobilization of plasmid DNA: termination frequency at the origin of transfer of plasmid R1162. *J Bacteriol* **176** (19): 5958-61.
- Reményi, A., Pohl, E., Schöler, H. R. and Wilmanns, M. (2001). Crystallization of redox-insensitive Oct1 POU domain with different DNA-response elements. *Acta Crystallographica Section D: Biological Crystallography* **57** (11): 1634-8.
- Rost, B., Fariselli, P. and Casadio, R. (1996). Topology prediction for helical transmembrane proteins at 86% accuracy. *Protein Sci* **5** (8): 1704-18.

- Rost, B. and Sander, C. (1993). Improved Prediction of Protein Secondary Structure by Use of Sequence Profiles and Neural Networks. *Proc Natl Acad Sci U S A* **90** (16): 7558-62.
- Santalucia, J., Jr. (1998). A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc Natl Acad Sci U S A* **95** (4): 1460-5.
- Scherzinger, E., Kruft, V. and Otto, S. (1993). Purification of the large mobilization protein of plasmid RSF1010 and characterization of its site-specific DNA-cleaving/DNA-joining activity. *Eur J Biochem* **217** (3): 929-38.
- Scherzinger, E., Lurz, R., Otto, S. and Dobrinski, B. (1992). In vitro cleavage of double- and single-stranded DNA by plasmid RSF1010-encoded mobilization proteins. *Nucleic Acids Res* **20** (1): 41-8.
- Schleifer, K. H. and Kandler, O. (1972). Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriol Rev* **36** (4): 407-77.
- Scholz, P., Haring, V., Wittmann-Liebold, B., Ashman, K., Bagdasarian, M. and Scherzinger, E. (1989). Complete nucleotide sequence and gene organization of the broad-host-range plasmid RSF1010. *Gene* **75** (2): 271-88.
- Schroder, G. and Lanka, E. (2005). The mating pair formation system of conjugative plasmids-A versatile secretion machinery for transfer of proteins and DNA. *Plasmid* **54** (1): 1-25.
- Schwarz, F. V., Perreten, V. and Teuber, M. (2001). Sequence of the 50-kb conjugative multiresistance plasmid pRE25 from *Enterococcus faecalis* RE25. *Plasmid* **46** (3): 170-87.
- Schwarz, S. and Chaslus-Dancla, E. (2001). Use of antimicrobials in veterinary medicine and mechanisms of resistance. *Vet Res* **32** (3-4): 201-25.
- Schwede, T., Kopp, J., Guex, N. and Peitsch, M. C. (2003). SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res* **31** (13): 3381-5.
- Segal, G. and Shuman, H. A. (1999). Possible origin of the *Legionella pneumophila* virulence genes and their relation to *Coxiella burnetii*. *Mol Microbiol* **33** (3): 669-70.
- Seppala, H., Klaukka, T., Vuopio-Varkila, J., Muotiala, A., Helenius, H., Lager, K. and Huovinen, P. (1997). The effect of changes in the consumption of macrolide antibiotics on erythromycin resistance in group A streptococci in Finland. Finnish Study Group for Antimicrobial Resistance. *N Engl J Med* **337** (7): 441-6.
- Studier, F. W. and Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* **189** (1): 113-30.
- Stura, E. A., Graille, M., Taussig, M. J., Sutton, B., Gore, M. G., Silverman, G. J. and Charbonnier, J.-B. (2001). Crystallization of macromolecular complexes: Stoichiometric variation screening. *Journal of Crystal Growth* **232** (1-4): 580-90.
- Svergun, D. I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Cryst* **25** (pt 4): 495-503.
- Svergun, D. I. and Koch, M. H. (2002). Advances in structure analysis using small-angle scattering in solution. *Curr Opin Struct Biol* **12** (5): 654-60.
- Svergun, D. I., Koch, M. H. J., Volkov, V. V., Kozin, M. B., Malfois, M., Konarev, P. V., Petoukhov, M. V. and Sokolova, A. V. (2004). All That SAS (ATSAS) a data analysis program suite.
- Svergun, D. I., Petoukhov, M. V. and Koch, M. H. (2001). Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* **80** (6): 2946-53.
- Swartz, M. N. (1997). Use of antimicrobial agents and drug resistance. *N Engl J Med* **337** (7): 491-2.

- Tang, X., Nakata, Y., Li, H. O., Zhang, M., Gao, H., Fujita, A., Sakatsume, O., Ohta, T. and Yokoyama, K. (1994). The optimization of preparations of competent cells for transformation of *E. coli*. *Nucleic Acids Res* **22** (14): 2857-8.
- Tato, I., Zunzunegui, S., De La Cruz, F. and Cabezon, E. (2005). TrwB, the coupling protein involved in DNA transport during bacterial conjugation, is a DNA-dependent ATPase. *Proc Natl Acad Sci U S A* **102** (23): 8156-61.
- Templin, M. F., Edwards, D. H. and Holtje, J. V. (1992). A murein hydrolase is the specific target of bulgecin in *Escherichia coli*. *J Biol Chem* **267** (28): 20039-43.
- Thompson, J. K. and Collins, M. A. (2003). Completed sequence of plasmid pIP501 and origin of spontaneous deletion derivatives. *Plasmid* **50** (1): 28-35.
- Ton-That, H. and Schneewind, O. (2004). Assembly of pili in Gram-positive bacteria. *Trends Microbiol* **12** (5): 228-34.
- Tusnady, G. E. and Simon, I. (2001). The HMMTOP transmembrane topology prediction server. *Bioinformatics* **17** (9): 849-50.
- Van Asselt, E. J., Dijkstra, A. J., Kalk, K. H., Takacs, B., Keck, W. and Dijkstra, B. W. (1999a). Crystal structure of *Escherichia coli* lytic transglycosylase Slt35 reveals a lysozyme-like catalytic domain with an EF-hand. *Structure Fold Des* **7** (10): 1167-80.
- Van Asselt, E. J., Kalk, K. H. and Dijkstra, B. W. (2000). Crystallographic studies of the interactions of *Escherichia coli* lytic transglycosylase Slt35 with peptidoglycan. *Biochemistry* **39** (8): 1924-34.
- Van Asselt, E. J., Thunnissen, A. M. and Dijkstra, B. W. (1999b). High resolution crystal structures of the *Escherichia coli* lytic transglycosylase Slt70 and its complex with a peptidoglycan fragment. *J Mol Biol* **291** (4): 877-98.
- Venyaninov, S. Y. and Yang, J. T. (1996). Determination of Protein Secondary Structure. Circular Dichroism and the Conformational Analysis of Biomolecules. G. D. Fasman. New York and London, Plenum Press 69-107.
- Verdino, P., Keller, W., Strohmaier, H., Bischof, K., Lindner, H. and Koraimann, G. (1999). The essential transfer protein TraM binds to DNA as a tetramer. *J Biol Chem* **274** (52): 37421-8.
- Volkov, V. V. and Svergun, D. I. (2003). Uniqueness of ab initio shape determination in small-angle scattering. *J Appl Cryst* **36** (3 I): 860-4.
- Von Heijne, G. (1992). Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. *J Mol Biol* **225** (2): 487-94.
- Vullo, A. and Frasconi, P. (2004). Disulfide connectivity prediction using recursive neural networks and evolutionary information. *Bioinformatics* **20** (5): 653-9.
- Walther, D., Cohen, F. E. and Doniach, S. (2000). Reconstruction of low-resolution three-dimensional density maps from one-dimensional small-angle X-ray solution scattering data for biomolecules. *Journal of Applied Crystallography* **33** (2): 350-63.
- Wang, A. and Macrina, F. L. (1995a). Characterization of six linked open reading frames necessary for pIP501-mediated conjugation. *Plasmid* **34** (3): 206-10.
- Wang, A. and Macrina, F. L. (1995b). Streptococcal plasmid pIP501 has a functional oriT site. *J Bacteriol* **177** (15): 4199-206.
- Weinstock, G. M., Mcentee, K. and Lehman, I. R. (1981). Hydrolysis of nucleoside triphosphates catalyzed by the recA protein of *Escherichia coli*. Characterization of ATP hydrolysis. *J Biol Chem* **256** (16): 8829-34.
- Weiss, M. A., Ellenberger, T., Wobbe, C. R., Lee, J. P., Harrison, S. C. and Struhl, K. (1990). Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA. *Nature* **347** (6293): 575-8.

- Wilkins, B. M. and Lanka, E. (1993). DNA processing and replication during plasmid transfer between Gram-negative bacteria. *Bacterial Conjugation*. D. B. Clewell. New York, Plenum Publishing Corp 105-36.
- Woody, R. W. (1996). Theory of Circular Dichroism of Proteins. *Circular Dichroism and the Conformational Analysis of Biomolecules*. G. D. Fasman. New York and London, Plenum Press 25-67.
- Zahrl, D., Wagner, M., Bischof, K., Bayer, M., Zavec, B., Beranek, A., Ruckenstein, C., Zarfel, G. E. and Koraimann, G. (2005). Peptidoglycan degradation by specialized lytic transglycosylases associated with type III and type IV secretion systems. *Microbiology* **151** (11): 3455-67.
- Zechner, E. L., De La Cruz, F., Eisenbrandt, R., Grahn, A. M., Koraimann, G., Lanka, E., Muth, G., Pansegrau, W., Thomas, C. M., Wilkins, B. M. and Zatyka, M. (2000). Conjugative-DNA transfer processes. *The Horizontal Gene Pool. Bacterial Plasmids and Gene Spread*. C. M. Thomas. Amsterdam, Harwood Academic Publishers 87–174.
- Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31** (13): 3406-15.

10 Appendices

10.1 List of abbreviations

3D	Three-dimensional
aa	Amino acid
Acc. No.	Accession Number
Amp	Ampicillin
bp	Base pair
BSA	Bovine serum albumine
<i>cat</i>	Chloramphenicol resistance gene
ccc	Covalently closed circular (supercoiled) plasmid DNA
Cm	Chloramphenicol
dsDNA	Double-strand DNA
Dtr	DNA transfer replication
EDTA	Ethylenediaminetetraacetic acid
Fus	Fusidic acid
fw	Forward
GST	Glutathione S-transferase
His-tag	6 or 7 histidine tag
Inc	Incompatibility group
IPTG	Isopropyl β -D-1-thiogalactopyranoside

kbp	kilo (1000) base pairs
LPS	Lipopolisaccharide
MBP	Maltose binding protein
MLS	Macrolide-Lincosamid-Streptogramin B antibiotics group (e.g. erythromycin)
MOPS	3-(N-Morpholino)propanesulfonic acid
mpf	Mating pair formation
NAG	N-acetylglucosamine
NAM	N-acetyl muramic acid
oc	Open circular form of a plasmid
OD ₆₀₀	Optical density at 600 nm
<i>orf</i>	Open reading frame
<i>oriT</i>	Origin of transfer
PA	Polyacrylamide
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
rev	Reverse
Rif	Rifampicin
SDS	Sodium dodecyl sulfate
SLT	Specialized lytic transglycosylase
Sm	Streptomycin
Sp	Spectinomycin
ssDNA	Single-strand DNA
T-strand	Transfer-DNA
T4SS	Type IV secretion systems
Tc	Tetracyclin
TEV	Tobacco etch virus
TLC	Thin layer chromatography
<i>tra</i>	Transfer
Tris	Tris-(hydroxymethyl)-aminomethane
X ^R	Resistance against substance X

10.2 Protein sequence statistics (*ProtParam* (www.expasy.org))

TraA

Number of amino acids: 672
Molecular weight: 78560.6
Theoretical pI: 8.89
Amino acid composition:

Amino acid	No.	%	Amino acid	No.	%	Amino acid	No.	%	Amino acid	No.	%
Ala (A)	25	3.7	Gln (Q)	34	5.1	Leu (L)	65	9.7	Ser (S)	48	7.1
Arg (R)	33	4.9	Glu (E)	68	10.1	Lys (K)	84	12.5	Thr (T)	22	3.3
Asn (N)	54	8.0	Gly (G)	21	3.1	Met (M)	16	2.4	Trp (W)	4	0.6
Asp (D)	40	6.0	His (H)	16	2.4	Phe (F)	26	3.9	Tyr (Y)	15	2.2
Cys (C)	2	0.3	Ile (I)	49	7.3	Pro (P)	17	2.5	Val (V)	33	4.9

Total number of negatively charged residues (Asp + Glu): 108
Total number of positively charged residues (Arg + Lys): 117

Formula: C3466H5605N979O1062S18
Total number of atoms: 11130

Extinction coefficients:
Conditions: 6.0 M guanidinium hydrochloride
0.02 M phosphate buffer
pH 6.5

Extinction coefficients are in units of M-1 cm-1 .
The first table lists values computed assuming ALL Cys residues appear as half cystines, whereas the second table assumes that NONE do.

	276	278	279	280	282
	nm	nm	nm	nm	nm
Ext. coefficient	43495	43527	42935	42080	40520
Abs 0.1% (=1 g/l)	0.554	0.554	0.547	0.536	0.516
	276	278	279	280	282
	nm	nm	nm	nm	nm
Ext. coefficient	43350	43400	42815	41960	40400
Abs 0.1% (=1 g/l)	0.552	0.552	0.545	0.534	0.514

Instability index:
The instability index (II) is computed to be 42.13
This classifies the protein as unstable.
Aliphatic index: 84.12
Grand average of hydropathicity (GRAVY): -0.843

TraAN₂₆₈

Number of amino acids: 296
Molecular weight: 34372.9
Theoretical pI: 6.55
Amino acid composition:

Amino acid	No.	%	Amino acid	No.	%	Amino acid	No.	%	Amino acid	No.	%
Ala (A)	14	4.7	Gln (Q)	11	3.7	Leu (L)	29	9.8	Ser (S)	20	6.8
Arg (R)	16	5.4	Glu (E)	28	9.5	Lys (K)	28	9.5	Thr (T)	9	3.0
Asn (N)	24	8.1	Gly (G)	12	4.1	Met (M)	8	2.7	Trp (W)	3	1.0
Asp (D)	19	6.4	His (H)	14	4.7	Phe (F)	7	2.4	Tyr (Y)	8	2.7
Cys (C)	1	0.3	Ile (I)	21	7.1	Pro (P)	9	3.0	Val (V)	15	5.1

Total number of negatively charged residues (Asp + Glu): 47
Total number of positively charged residues (Arg + Lys): 44

Formula: C1508H2410N438O463S9
Total number of atoms: 4828

Extinction coefficients: Conditions: 6.0 M guanidinium hydrochloride
0.02 M phosphate buffer
pH 6.5 Extinction coefficients are in units of M-1 cm-1 .
The first table lists values computed assuming ALL Cys residues appear as half cystines, whereas the second table assumes that NONE do.

	276	278	279	280	282
	nm	nm	nm	nm	nm
Ext. coefficient	27800	28000	27740	27310	26400
Abs 0.1% (=1 g/l)	0.809	0.815	0.807	0.795	0.768

	276	278	279	280	282
	nm	nm	nm	nm	nm
Ext. coefficient	27800	28000	27740	27310	26400
Abs 0.1% (=1 g/l)	0.809	0.815	0.807	0.795	0.768

Instability index: The instability index (II) is computed to be 31.83
This classifies the protein as stable.
Aliphatic index: 85.30
Grand average of hydropathicity (GRAVY): -0.802

TraAN₂₄₆

Number of amino acids: 266
Molecular weight: 30938.9
Theoretical pI: 7.27
Amino acid composition:

Amino acid	No.	%	Amino acid	No.	%	Amino acid	No.	%	Amino acid	No.	%
Ala (A)	14	5.3	Gln (Q)	8	3.0	Leu (L)	22	8.3	Ser (S)	17	6.4
Arg (R)	16	6.0	Glu (E)	25	9.4	Lys (K)	25	9.4	Thr (T)	9	3.4
Asn (N)	22	8.3	Gly (G)	12	4.5	Met (M)	8	3.0	Trp (W)	3	1.1
Asp (D)	16	6.0	His (H)	14	5.3	Phe (F)	8	3.0	Tyr (Y)	7	2.6
Cys (C)	1	0.4	Ile (I)	18	6.8	Pro (P)	8	3.0	Val (V)	13	4.9

Total number of negatively charged residues (Asp + Glu): 41
Total number of positively charged residues (Arg + Lys): 41

Formula C1356H2152N400O412S9

Total number of atoms: 4329

Extinction coefficients:
Conditions: 6.0 M guanidinium hydrochloride
0.02 M phosphate buffer
pH 6.5

Extinction coefficients are in units of M⁻¹ cm⁻¹.

The first table lists values computed assuming ALL Cys residues appear as half cystines, whereas the second table assumes that NONE do.

	276	278	279	280	282
	nm	nm	nm	nm	nm
Ext. coefficient	26350	26600	26395	26030	25200
Abs 0.1% (=1 g/l)	0.852	0.860	0.853	0.841	0.815

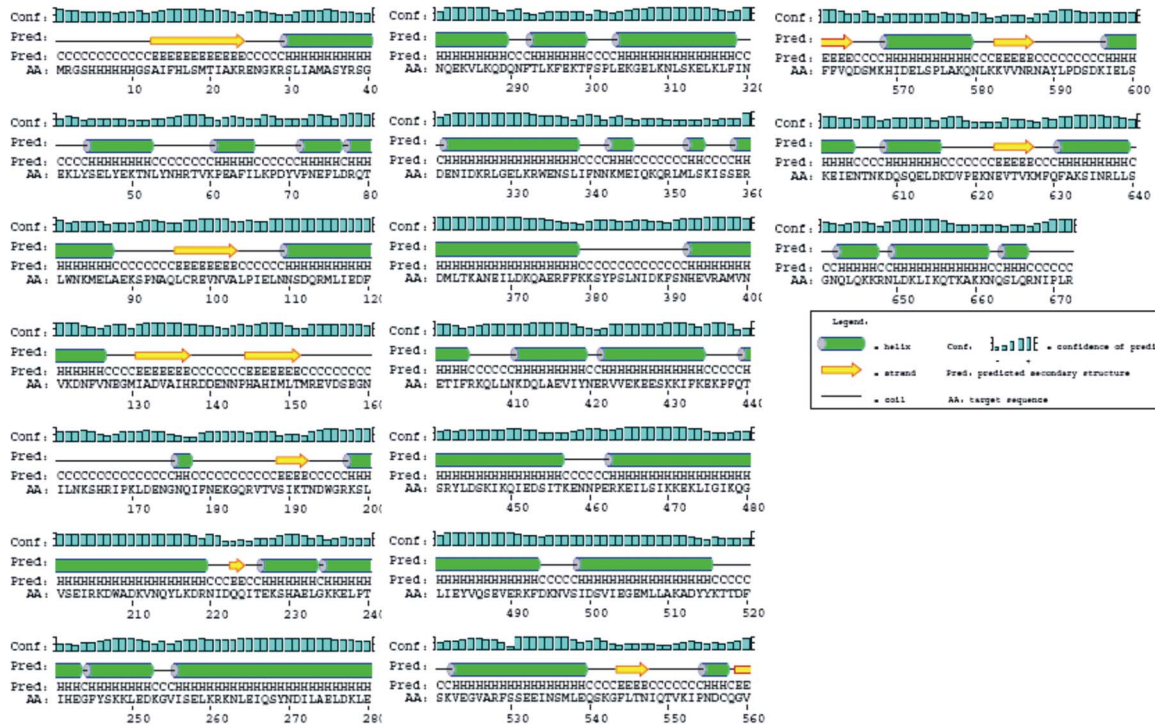
	276	278	279	280	282
	nm	nm	nm	nm	nm
Ext. coefficient	26350	26600	26395	26030	25200
Abs 0.1% (=1 g/l)	0.852	0.860	0.853	0.841	0.815

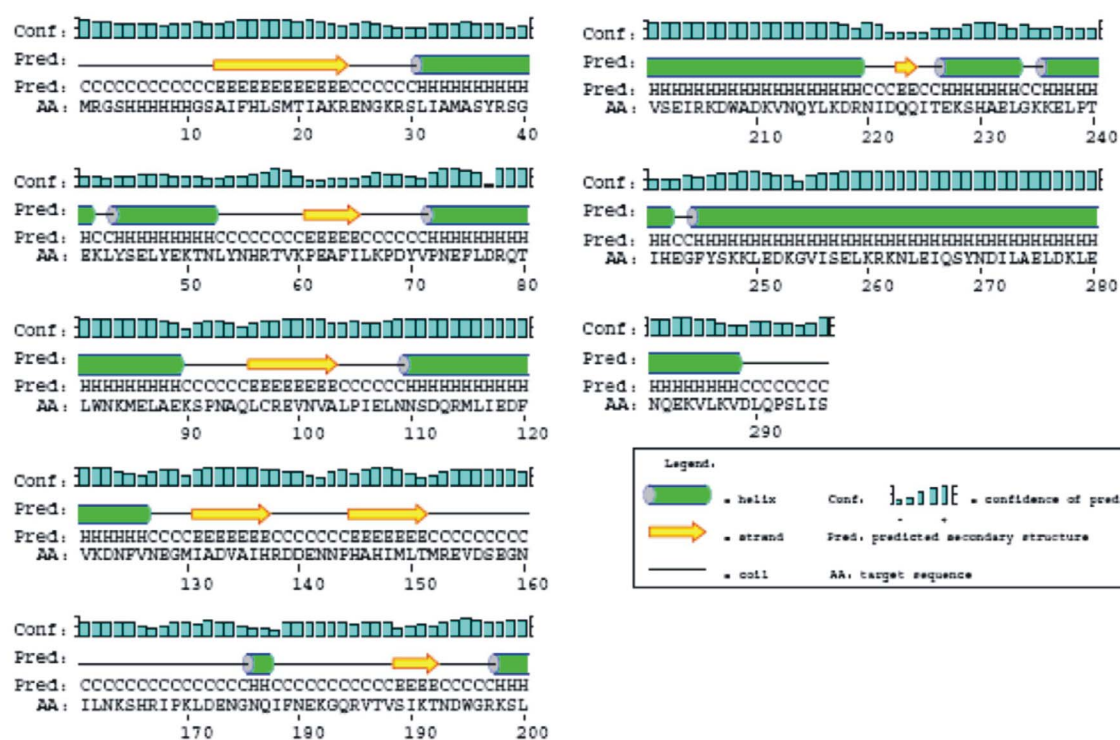
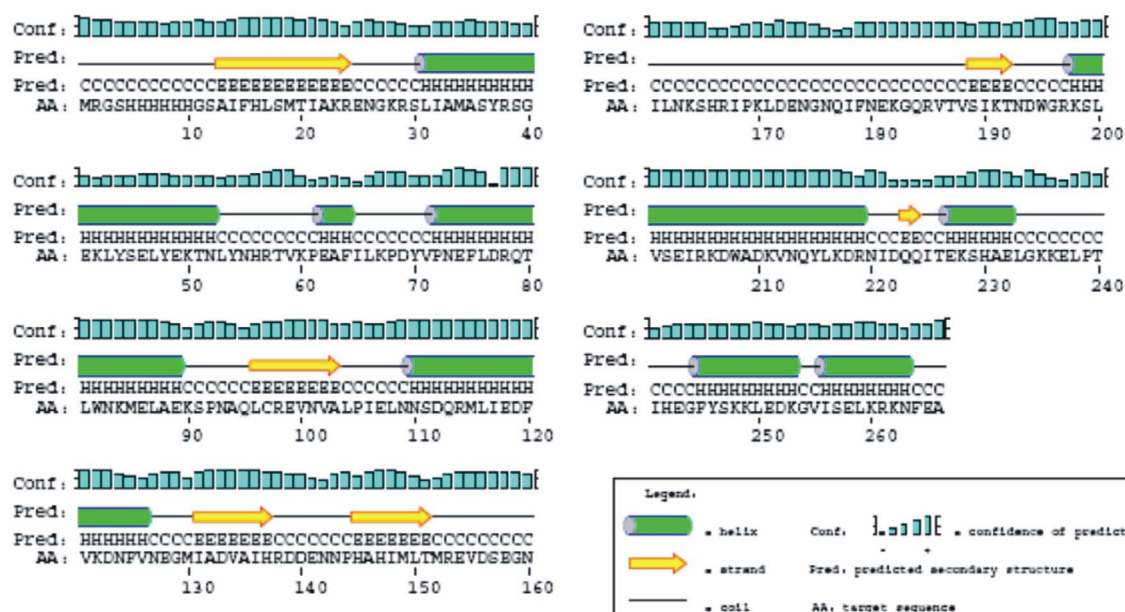
Instability index:
The instability index (II) is computed to be 29.67
This classifies the protein as stable.

Aliphatic index: 78.08
Grand average of hydropathicity (GRAVY): -0.856

10.3 Secondary Structure Prediction (PSIPRED)

TraA



TraAN₂₆₈TraAN₂₄₆

10.4 Accession numbers of Orf2-Orf15

Orf2	AAA99467	Orf7	CAD44387	Orf12	CAD44392
Orf3	AAA99468	Orf8	CAD44388	Orf13	CAD44393
Orf4	AAA99469	Orf9	CAD44389	Orf14	CAD44394
Orf5	AAA99470	Orf10	CAD44390	Orf15	CAD44395
Orf6	AAA99471	Orf11	CAD44391		

10.5 The Delphi code of txt2dic program

```
program txt2dic;
uses
  Forms,
  main in 'main.pas' {Form1};
{$R *.res}

begin
  Application.Initialize;
  Application.Title := 'txt2dic';
  Application.CreateForm(TForm1, Form1);
  Application.Run;
end.

unit main;

interface

uses
  Windows, Messages, SysUtils, Variants, Classes, Graphics, Controls,
  Forms,
  Dialogs, StdCtrls, ActnList, Menus, ComCtrls;

type
  TForm1 = class(TForm)
    OpenDialog1: TOpenDialog;
    Button1: TButton;
    Button2: TButton;
    Edit1: TEdit;
    Label1: TLabel;
    Label2: TLabel;
    Edit2: TEdit;
    SaveDialog1: TSaveDialog;
    Button3: TButton;
    MainMenu1: TMainMenu;
    Info1: TMenuItem;
    Help1: TMenuItem;
    Memo1: TMemo;
    RadioButton1: TRadioButton;
    CheckBox2: TCheckBox;
    RadioButton2: TRadioButton;
    HeaderControl1: THeaderControl;
    procedure Button1Click(Sender: TObject);
    procedure Button2Click(Sender: TObject);
    procedure Button3Click(Sender: TObject);
    procedure Info1Click(Sender: TObject);
  private
    { Private declarations }
  end;
end;
```

```

public
{ Public declarations }
end;

var
  Form1: TForm1;
  Filename, filename_save:string;
  plik :TextFile ;
  licznik,maxi,mini,k,dlugosc,ilosc,i : integer;
implementation
{$R *.dfm}

procedure TForm1.Button1Click(Sender: TObject);
var
  kropka,ss,znak,sample,sam,res,rs:string;
  //nr wiersza, nr znaku
  line:array[1..5000,1..3] of string;
  elp,elp1:array[1..5000,1..2]of real;
  resol, smooth:real;
  l,maxl,dl, smooth_int,a,b:integer;
begin
  if length(Filename)=0 then
    MessageBox(0,'You need to specify txt file first','Warning',MB_OK)
  else
    begin
      AssignFile(plik,Filename);
      Reset(plik);
      licznik:=0;
      i:=0;
      Memol.Clear;
      maxl:=1;

      while not Eof(plik) do
        begin
          ReadLn(plik,ss);
          inc(i);
          dlugosc:=length(ss) ;
          k:=1;
          kropka:=DecimalSeparator;
          for licznik:=1 to dlugosc do
            begin
              znak:=ss[licznik];
              if i>20 then if znak='.' then znak:=kropka;
              if znak=#9 then inc(k)
              else
                line[i,k]:=line[i,k]+znak;
            end;
            maxi:=i
          end;
          CloseFile(plik);
        end;

        for i:=1 to maxi do
          begin
            if line[i,1]='Sample :' then
              begin
                sam:=line[i,2];
                for l:=1 to 10 do
                  Sample:=sample+sam[l];
                end;
            if line[i,1]='Resolution :' then
              begin
                res:=line[i,2];
                dl:=length(res);
                for l:=1 to dl-2 do
                  begin

```

```

        if res[l]='.' then res[l]:=DecimalSeparator;
        rs:=rs+res[l];
    end;
    resol:=StrToFloat(rs);
end;
end;

i:=1;

if RadioButton1.Checked=true
then
begin
    repeat
        i:=i+1
    until line[i,1]='Wavelength[nm]';
    mini:=i+1
    end
else
begin
    repeat
        i:=i+1
    until line[i,1]='Temperature';
    mini:=i+1;
    end;
end;

memo1.Lines.add(concat(' ',sample,' '));
memo1.Lines.append('Ellipticities');

for i:=mini to maxi do
begin
    for k:=1 to 2 do
        elp[i-mini+1,k]:=StrToFloat(line[i,k]);
    end; {for}

    l:=1;

if Checkbox2.Checked=false then
begin
    for i:=mini to maxi do
    begin
        if round(elp[i-mini+1,1]*10)mod round(1/resol) = 0 then
        begin
            elp1[l,1]:=elp[i-mini+1,1];
            elp1[l,2]:=round(elp[i-mini+1,2]);
            l:=l+1;
        end;
        maxl:=l-1
    end; {for}
end {if}

else
begin
    for i:=mini to maxi do
    begin
        elp1[l,1]:=elp[i-mini+1,1];
        elp1[l,2]:=elp[i-mini+1,2];
        l:=l+1;
        maxl:=l-1
    end; {for}
end {else};

if checkbox2.Checked then
begin
    for l:=maxl downto 1 do
memo1.Lines.Append(FloatToStr(elp1[l,1])+';'+FloatToStr(elp1[l,2]));
    end
else
begin

```



```

        for l:=maxl downto 1 do
memol.Lines.Append(FloatToStr(elp1[l,1])+',' +FloatToStr(elp1[l,2]));
        end;
        filename_save:=Edit2.Text;
        memol.Lines.Insert(1,FloatToStr(maxl));
        memol.Lines.SaveToFile(filename_save);
end;

procedure TForm1.Button2Click(Sender: TObject);
begin
if OpenFileDialog1.Execute then
    filename := OpenFileDialog1.FileName;
    filename_save:='';
    Edit1.Text:=filename;
    dlugosc:=length(filename);
    for i:=1 to dlugosc-3 do
        filename_save:=filename_save + filename[i];
        filename_save:=filename_save+'dic';
    Edit2.Text:='';
    Edit2.Text:=filename_save;
end;

procedure TForm1.Button3Click(Sender: TObject);
begin
    if SaveDialog1.Execute then
        begin
            Edit2.text:='';
            filename_save:=SaveDialog1.FileName;
            Edit2.Text:=filename_save;
            if Edit2.Text='' then filename_save:='no_name.dic'
            else filename_save:=Edit2.Text;
            SaveDialog1.FileName:=filename_save;
        end
end;

procedure TForm1.Info1Click(Sender: TObject);
begin
    MessageBox(0,'This program converts ascii file,'+#10+'obtained from
CD spectrum,'+#10+'into dic file, which can be read by
dicroprot.','Txt2Dic v0.8',MB_OK);
end;
end.

```