Structural basis of conjugative DNA transfer mediated by MobM, a prototype of the major relaxase family of *Staphylococcus aureus*

Radoslaw Pluta

TESI DOCTORAL UPF/ 2014

DIRECTORS DE LATESI:
Prof Miquel Coll & Dr D Roeland Boer

Structural and Computational Biology Department
Institute for Research in Biomedicine (IRB Barcelona)
Instituto de Biología Molecular de Barcelona (IBMB-CSIC)

Dedykuję tę pracę tym, którzy najbardziej troszczyli się o mnie i mój rozwój i mieli najwiekszy wpływ na to kim dziś jestem.
Mamie, babci Anieli i dziadkowi Janowi.

Acknowledgements

This work has been possible thanks to the collective work of many people, who contributed to the final outcome of my thesis. I would like to acknowledge them, as well as other people not involved in the project, but having positive impact on me during my years spent at IRB:

- Prof Miquel Coll - for giving me the opportunity to pursue a PhD in his group, for guiding and supporting my scientific development and, last but not least, for giving me some space to for my own initiatives.
- Dr Roeland Boer, who taught, supervised, supported, mentored, provoked, motivated, pressed (when needed ;)) and caffeinated with me through my everyday lab life and during our synchrotron trips and overnight shifts with the hard rock music and topics.
- Other members of the MobM project: Prof Manuel Espinosa, Dr Fabian Lorenzo-Diaz and Cris Fernandez-Lopez from the MobM project homeland (CIB-CSIC) for molecular biology aspects of the project and fruitful discussions; Dr Silvia Russi and Rosa Perez-Luque from our group, who both worked on MobM project before I joined the lab.
- Prof Joan Guinovart, IRB Director - for his incredible efforts towards making IRB such a great and lively place to work in such a short time.
- Members of my Thesis Advisory Committee – Prof Baldomero Oliva, Dr Maria Sola, and temporary members - Dr Pau Bernado, Prof Ignasio Fita and Prof Modesto Orozco.
- Dr Alicia Gausch for her very useful suggestion about crystal soaking at different pH.
- Members of the Crystallographic Platform - Joan, Sonia, Robert, and members of the Purification Facility - Jenny and Isabel for their great help.

- Members of the IRB Administration, especially Clara and Patricia, for their very friendly and helpful attitude.
- La Caixa Foundation for their International PhD Fellowship Programme.
- Former and current members of the Coll Group and synchrotron trips colleagues: Albert, Anna, Ariadna, Carlo, Carme, Clara, Cristina, Delphine, Diana, Esther, Fabio, Juliana, Mailys, Marta, Montse, Nayibe, Pablo, Robert, Roeland, Rosa, Salva, Simone, Susan and Zuza.
- Friends and colleagues from the IRB/PCB community, 2nd IRB PhD Symposium Committee, Student Council, FC Sparta (Moulecule, IRBeasts) football team, basketball team for many great professional, social and personal moments during my Barcelona chapter.

Abstract

MobM relaxase from the promiscuous antibiotic resistance plasmid pMV158 is a prototype of the Mob_Pre/MOB_V family of relaxases, the major family of relaxases found in *Staphylococcus aureus.* Staphylococcal infections cause the highest number of lethal cases among antibiotic-resistant bacterial infections. Relaxases initiate the conjugative DNA transfer, a major route for the antibiotic resistance acquisition in bacteria, by nicking their substrate DNA through formation of a covalent DNA-relaxase adduct and terminate the transfer in the recipient cells by rejoining ends of the linearized plasmid. MobM forms a DNA-histidine adduct, unique to MOB_V relaxases, instead of a DNA-tyrosine adduct, thus representing a distinct category of relaxases with specialization towards the transfer of short mobile genetic elements in Gram-positive pathogenic bacteria**.** MobM overall fold resembles the fold of other structurally characterized relaxases, although some important structural differences are present. Molecular basis for the MobM processing of plasmid origin of transfer and active site mechanism are described herein.

Resumen

La relaxasa MobM del promiscuo plásmido de resistencia a antibióticos pMV158 es un prototipo de la familia Mob_Pre/MOB_V de relaxasas, la mayor familia de relaxasas se encuentran en *Staphylococcu aureus*. Las infecciones por estafilococos causan el mayor número de casos mortales entre las infecciones bacterianas resistentes a los antibióticos. Relaxases iniciar la transferencia conjugativa de ADN, una ruta el más frecuente para la adquisición de resistencia a antibióticos por bacterias, por mellar su ADN sustrato mediante la formación de un aducto covalente de ADN-relaxasa y terminan la transferencia en las células receptoras por reincorporarse extremos del plásmido linealizado. MobM forma un aducto de ADN-histidina, único para MOB_V relaxases, en lugar de un aducto de ADN-tirosina, lo que representa una

categoría distinta de relaxases con especialización hacia la transferencia de elementos genéticos móviles cortos en bacterias patógenas Gram-positivas. MobM estructura general se asemeja a la de otras veces relaxases caracterizan estructuralmente, aunque algunas diferencias estructurales importantes están presentes. Base molecular para el procesamiento de origen del plásmido por MobM y mecanismo de sitio activo se describe en esta thesis.

Publications

Radoslaw Pluta*, D. Roeland Boer*, Cris Fernández-López, Silvia Russi, Fabián Lorenzo-Díaz, Rosa Pérez-Luque, Manuel Espinosa, Miquel Coll. **Structures of DNA-MobM relaxase complexes reveal a histidine/metal catalysis for DNA cleavage and ligation**. Submitted to *Nature Structural and Molecular Biology*

∗ These authors contributed equally to this work

Radoslaw Pluta*, D. Roeland Boer*, Silvia Russi, Fabián Lorenzo-Díaz, Cris Fernández-López, Rosa Pérez-Luque, Manuel Espinosa, Miquel Coll. **Structural basis of antibiotic resistance transfer mediated by MobM, a prototype of the major family of relaxases found in *Staphylococcus aurues***. In preparation

∗ These authors contributed equally to this work

Cris Fernández-López, Radoslaw Pluta, Rosa Pérez-Luque, Lorena Rodríguez-González, Manuel Espinosa, Miquel Coll, Fabián Lorenzo-Díaz, D. Roeland Boer. **Functional Properties and Structural Requirements of the Plasmid pMV158-Encoded MobM Relaxase Domain**. *J Bacteriol.* 2013; 195(13):3000-8.

Preface

The following text is a **reproduction** of a news release of the World Health Organization Media Centre.
(http://www.who.int/mediacentre/news/releases/2014/amr-report/en/)

WHO's first global report on antibiotic resistance reveals serious, worldwide threat to public health

New WHO report provides the most comprehensive picture of antibiotic resistance to date, with data from 114 countries.

30 APRIL 2014 | GENEVA - A new report by WHO–its first to look at antimicrobial resistance, including antibiotic resistance, globally–reveals that this serious threat is no longer a prediction for the future, it is happening right now in every region of the world and has the potential to affect anyone, of any age, in any country. Antibiotic resistance–when bacteria change so antibiotics no longer work in people who need them to treat infections–is now a major threat to public health.

"Without urgent, coordinated action by many stakeholders, the world is headed for a post-antibiotic era, in which common infections and minor injuries which have been treatable for decades can once again kill," says Dr Keiji Fukuda, WHO's Assistant Director-General for Health Security. "Effective antibiotics have been one of the pillars allowing us to live longer, live healthier, and benefit from modern medicine. Unless we take significant actions to improve efforts to prevent infections and also change how we produce, prescribe and use antibiotics, the world will lose more and more of these global public health goods and the implications will be devastating."

Key findings of the report

The report, "Antimicrobial resistance: global report on surveillance", notes that resistance is occurring across many different infectious agents but the report focuses on antibiotic resistance in seven different bacteria responsible for common, serious diseases such as bloodstream infections (sepsis), diarrhoea, pneumonia, urinary tract infections and gonorrhoea. The results are cause for high concern, documenting resistance to antibiotics, especially “last resort” antibiotics, in all regions of the world.

Key findings from the report include:

- Resistance to the treatment of last resort for life-threatening infections caused by a common intestinal bacteria, *Klebsiella pneumoniae*–carbapenem antibiotics–has spread to all regions of the world. *K. pneumoniae* is a major cause of hospital-acquired infections such as pneumonia, bloodstream infections, infections in newborns and intensive-care unit patients. In some countries, because of resistance, carbapenem antibiotics would not work in more than half of people treated for *K. pneumoniae* infections.

- Resistance to one of the most widely used antibacterial medicines for the treatment of urinary tract infections caused by *E. coli*–fluoroquinolones–is very widespread. In the 1980s, when these drugs were first introduced, resistance was virtually zero. Today, there are countries in many parts of the world where this treatment is now ineffective in more than half of patients.

- Treatment failure to the last resort of treatment for gonorrhoea–third generation cephalosporins–has been confirmed in Austria, Australia, Canada, France, Japan, Norway, Slovenia, South Africa, Sweden and the United Kingdom. More than 1 million people are infected with gonorrhoea around the world every day.

- Antibiotic resistance causes people to be sick for longer and increases the risk of death. For example, people with MRSA (methicillin-resistant *Staphylococcus aureus*) are estimated to be 64% more likely to die than people with a non-resistant form of the infection. Resistance also increases the cost of health care with lengthier stays in hospital and more intensive care required.

Ways to fight antibiotic resistance

The report reveals that key tools to tackle antibiotic resistance–such as basic systems to track and monitor the problem–show gaps or do not exist in many countries. While some countries have taken important steps in addressing the problem, every country and individual needs to do more.

Other important actions include preventing infections from happening in the first place–through better hygiene, access to clean water, infection control in health-care facilities, and vaccination–to reduce the need for antibiotics. WHO is also calling attention to the need to develop new diagnostics, antibiotics and other tools to allow healthcare professionals to stay ahead of emerging resistance.

This report is kick-starting a global effort led by WHO to address drug resistance. This will involve the development of tools and standards and improved collaboration around the world to track drug resistance, measure its health and economic impacts, and design targeted solutions.

How to tackle resistance

People can help tackle resistance by:

- using antibiotics only when prescribed by a doctor;
- completing the full prescription, even if they feel better;
- never sharing antibiotics with others or using leftover prescriptions.

Health workers and pharmacists can help tackle resistance by:

- enhancing infection prevention and control;
- only prescribing and dispensing antibiotics when they are truly needed;
- prescribing and dispensing the right antibiotic(s) to treat the illness.

Policymakers can help tackle resistance by:

- strengthening resistance tracking and laboratory capacity;
- regulating and promoting appropriate use of medicines.

Policymakers and industry can help tackle resistance by:

- fostering innovation and research and development of new tools;
- promoting cooperation and information sharing among all stakeholders.

The report–which also includes information on resistance to medicines for treating other infections such as HIV, malaria, tuberculosis and influenza–provides the most comprehensive picture of drug resistance to date, incorporating data from 114 countries.

Abbreviations

bp – base pair
CA – community-acquired
CDC – Centers for Disease Control and Prevention (USA)
CDD – Conserved Domains Database
DNA – deoxyribonucleic acid
dsDNA – double stranded deoxyribonucleic acid
Dtr – DNA-transfer replication
DTT – dithiothreitol
EDTA – ethylenediaminetetraacetic acid
ESRF – European Synchrotron Radiation Facility
G- – Gram-negative
G+ – Gram-positive
GTAs – gene transfer agents
HAI – healthcare-acquired infection
HUH – two histidines separated by a hydrophobic residue
ICEs – integrative and conjugative elements
IPTG – isopropyl-β-D-thiogalactopyranoside
IR – inverted repeat
MGEs – mobile genetic elements
MOB – plasmid mobility enzyme
MPD – 2-methyl-2,4-pentanediol
MPF – mating pair formation
MRSA – methicillin-resistant *Staphylococcus aureus*
MVs – membrane vesicles
NCBI – National Center for Biotechnology Information (USA)
NES – nicking enzyme of *Staphylococcus aureus*
nic – specific site where a relaxase cleaves the DNA substrate
OD – optical density
oligo – oligonucleotide
OMVs – outer membrane vesicles
oriT – origin of transfer
PDB – Protein Data Bank
PEG – polyethylene glycol

pH – negative log (base 10) of hydronium ions molar concentration
pKa – acid dissociation constant
RCR – rolling-circle replication
RMSD – root-mean-square deviation
ssDNA – single stranded deoxyribonucleic acid
SAD – Single wavelength Anomalous Diffraction
SDS-PAGE – sodium dodecyl sulphate-polyacrylamide gel electrophoresis
T4CP – type IV secretion system coupling protein
T4SS – type IV secretion system
Tris – tris(hydroxymethyl)aminomethane
WHO – World Health Organization
wt – wild type

CONTENTS

CHAPTER 1
INTRODUCTION

1.1. Horizontal Gene Transfer

Horizontal gene transfer (**HGT**) is a process of genes transfer between organisms in ways other than traditional reproduction (sexual or asexual). By contrast, the transmission of genes from the parental generation to offspring via reproduction is called vertical gene transfer. HGT is considered to be a major factor influencing evolution of many organisms, especially of prokaryotes, as many genes have been exchanged rather than evolved by mutations (Frost *et al.*, 2005; Gogarten *et al.*, 2009; Treangen and Rocha 2011). It was therefore proposed by McInerney and colleagues to view genes (or DNA segments) as "public goods, available for all organisms to integrate into their genomes" (McInerney *et al.*, 2010). A consequence of above is the primary role of HGT in the acquisition and spread of antibiotic resistance in bacteria, an issue that, in the light of lack of development and/or approval of new antibiotics, is becoming a major medical threat to human health (Furuya and Lowy, 2006; CDC, 2013; WHO, 2014). Although HGT takes place in all three kingdoms of life, for many years most of our understanding of HGT was profoundly shifted towards the prokaryotic organisms (Thomas and Nielsen, 2005; Keeling and Palmer, 2008; Wozniak and Waldor, 2010; Popa and Dagan, 2011; Zhaxybayeva and Doolittle, 2011). Interestingly, two of the most extensive and ancient examples of HGT include acquisition of chloroplasts and mitochondria, which according to the endosymbiotic theory originated as bacterial endosymbionts of a progenitor of the eukaryotic cell (Blanchard and Lynch, 2000). Recent expansion of genome sequencing projects produced cumulative evidences for HGT between prokaryotic and eukaryotic organisms (Ros and Hurst, 2009;

Alsmark *et al.,* 2013, Gilbert and Cordaux, 2013). Among studies focused on defining the HGT events in humans, one found that DNA transfer between bacteria (*Acinetobacter* and *Pseudomonas*) and human somatic cells was elevated in cancer cells (myeloid leukemia and stomach cancer, respectively), raising questions about possible links between the human microbiome and cancer (Riley *et al.,* 2013). Furthermore, two studies reported phage and bacterial protein-DNA-covalent-adducts delivery to the mammalian and human nucleus, respectively, potentially opening new avenues for applications in gene delivery (Schröder *et al.,* 2011; Redrejo-Rodríguez *et al.,* 2012; Llosa *et al.,* 2012). Recent years shed also more light on the trends and barriers of HGT – it was shown that the ecology (common niche), rather than phylogeny or geography, is the most important driving force of HGT (Smillie *et al.,* 2011, Langille *et al.,* 2012; Gilbert and Cordaux, 2013). In addition to the three well established canonical routes of HGT (conjugation, transformation and transduction), modification of them and/or new mechanisms (gene transfer agents, nanotubes, membrane vesicles, serial transduction) are being better understood (see below for more details). Besides the implications on human health, understanding of HGT is also important because it is widely used in molecular biology, biotechnology and gene therapy (Green and Sambrook, 2012; Lacroix *et al.,* 2008; Aronovich *et al.,* 2011).

Classical routes of HGT

The three classical mechanisms of horizontal gene transfer are conjugation, transduction and transformation. They have been invaluably important for the development of molecular biology. The first two processes seem to be associated with mobile genetic elements (MGEs) and widespread among different species and kingdoms (Frost *et al.,* 2005). MGEs can be described as segments of DNA that encode for proteins (among them DNA processing enzymes) that mediate the movement of DNA within genomes (intracellular mobility; transposition) or between cells (intercellular mobility) (Fig. 1). MGEs usually include also genes that encode for proteins that provide

advantage to the acceptor cell, e.g. antibiotic resistance, metabolic enzymes.

Conjugation (unlike transduction or transformation) requires a direct physical contact between the participating cells. The DNA transfer occurs via mobile genetic elements (MGEs) that either reside externally to the bacterial chromosome (plasmids) or are integrated within it (integrons, mobilizable genomic islands (MGIs) and integrative conjugative elements (ICEs) also known as conjugative transposons) (Cambray *et al.,* 2010; Waldor 2010; Wozniak and Waldor, 2010; Daccord *et al.,* 2013). After the transfer, MGEs can be incorporated into the recipient chromosomes by homologous recombination via insertion sequences (IS) or other sequences that carry required similarity.

Transduction is a gene transfer process mediated by bacterial viruses (bacteriophages) that often results in the bacterial host cell lysis (Frost *et al.,* 2005). Cell recognition is mediated through interaction with specific receptors found on its surface. In addition to carrying their own DNA, phages can carry pieces of the DNA taken accidentally from the cells in which they matured. During transduction DNA integration into the host genome is mediated by the phage-encoded enzymes.

Transformation is a process in which gene transfer is mediated by the uptake of naked DNA (DNA without associated proteins or other biomolecules) from the environmental. It involves the DNA transfer between closely related species during the competence state of some naturally transformable bacteria (Chen and Dubnau, 2004; Frost *et al.,* 2005). The process is mediated by several dozen chromosomally encoded proteins, including the type IV pillus and type II secretion system. In some species, DNA molecules carrying the uptake signal sequences (USSs) can be integrated into the chromosomes, generally via homologous recombination (Snyder *et al.,* 2007).

New mechanisms of HGT

New routes of HGT are being discovered, potentially expanding our understanding of the significance and universality of this process in the unicellular world.

Gene transfer agents (GTAs) are bacteria-encoded phage-like elements that transfer random and specific pieces of the bacterial genome (Lang *et al.*, 2012; Guy *et al.*, 2013). Genes encoding GTAs were found in virtually every sequenced member of the alpha-proteobacteria order *Rhodobacterales* (Lang and Beatty, 2007).

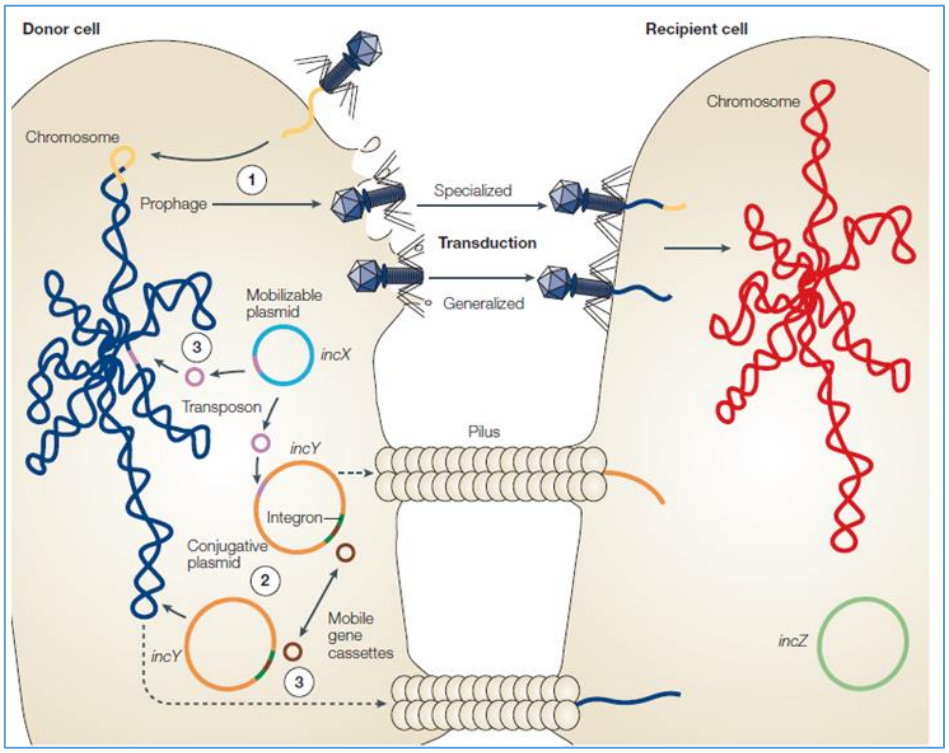


Figure 1. **Inter- and intra- cellular transfer of Mobile Genetic Elements (MGEs) in bacterial cells.** Figure from Frost *et al.*, (2005). **(1) Transduction** - the DNA genome (yellow) of a temperate phage integrates into the chromosome (dark blue) as a prophage; later prophage genes within the host genome are expressed, resulting in the production of phage particles and replication of the phage genome. Successively, packaging of the phage and host DNA (specialized transduction) occurs with occasional packaging of host DNA only (generalized transduction). Subsequently phage lyses the cell, and infects a recipient cell in which the novel DNA recombines into the recipient host cell chromosome (red). **(2) Conjugation –** self-transmissible large, low-copy number conjugative plasmids (orange) and integrative conjugative elements (ICEs; not shown) produce pili and type IV secretion system (T4SS) channels to initiate and form a connection with the recipient cells and to transfer themselves to them. Alternatively, a mobilizable plasmid or mobilizable genomic island (MGI) or a very big fragment of the bacterial chromosome can be transferred. In the recipient cells these genetic elements either insert into the chromosome or replicate independently if are not incompatible with the resident plasmids (light

green). Conjugative MGEs of Gram-positive bacteria (not shown) do not use pili. **(3) Transposition** - Transposons (pink) integrate into new sites on the chromosome or plasmids by non-homologous recombination. Integrons (dark green) use similar mechanisms to exchange single gene cassettes (brown).

Moreover, new and different GTAs are being found them in other distant groups (Lang *et al.,* 2012). GTAs were shown to be major contributors to microbial marine evolution with nearly 50% of the cultivable natural oceans microbial community confirmed as gene recipients (McDaniel *et al.,* 2010). It is presumed that GTAs require lysis for release from cells and such scenario was shown for *Rhodobacter capsulatus* RcGTA (Fogg *et al.,* 2012).

Nanotubes are discovered for the first time in 2011 new types of intercellular bacterial connectors (Dubey and Ben-Yehuda, 2011). Simultaneous inter- and intra-species nanotubes were observed in mixed populations of the three bacterial species (*B. subtilis, S. aureus* and *E. coli*) and GFP transfer between them was visualized. Moreover, the transfer of a non-conjugative plasmid was demonstrated between two *B. subtilis* strains when grown on solid medium.

Outer membrane vesicles (OMVs) in Gram-negative and **membrane vesicles (MVs)** in Gram-positive bacteria appear to be another platform for the exchange of biomolecules among bacteria (Lee, 2012; Berleman and Auer, 2013). OMVs were shown to carry chromosomal, plasmid, and phage DNA that can then merge with nearby cells.

Serial transduction is a HGT mechanism that connects both membrane vesicles and viral infection elements (Chiura *et al.,* 2011). Chiura and colleagues shown that membrane vesicle-like particles from sea water carrying large DNA segments are capable of cell transduction. Surprisingly, the membrane vesicle-like particles receiving cells were subsequently capable of production of similar membrane vesicles.

1.2. Bacterial Conjugation

The ability of bacteria to transfer their genetic material through conjugation was discovered nearly 70 years ago (Lederberg and Tatum 1946). From the three classical mechanisms of horizontal gene transfer, conjugation is thought to be quantitatively more important (Halary *et al.*, 2010). Moreover, conjugation is the route allowing the simultaneous transfer of larger amounts of DNA and between the least related cells. Plasmids, integrons, mobilizable genomic islands (MGIs) aka integrative mobilizable elements (IMEs) and integrative conjugative elements (ICEs) aka conjugative transposons use conjugation to be transferred to other cells (Cambray *et al.*, 2010; Smillie *et al.*, 2010; Wozniak and Waldor, 2010; Guglielmini *et al.*, 2011; Daccord *et al.*, 2013). Plasmids and integrative conjugative elements (ICEs) are the major players in conjugation landscape. Interestingly, it was proposed that plasmids often become ICEs, and/or vice-versa and that plasmids and ICEs might be just "the two faces shown by a very similar type of element" (Guglielmini *et al.*, 2011).

Conjugative DNA transfer is not limited only to inter-bacterial gene exchange; *Agrobacterium tumefaciens* uses conjugation as delivery tool of genetic information to take control over plant cells causing a tumor formation called the crown gall disease known (Gelvin, 2003). *Agrobacterium*-mediated T-DNA transfer to plant cells and its integration into the genome became widely used in research and biotechnology (Lacroix *et al.*, 2008). Interestingly, conjugative DNA transfer into human cells by the bacterial pathogen *Bartonella henselae* was recently reported and discussed as a potential new tool for DNA delivery into specific human cell types (Schröder *et al.*, 2011; Llosa *et al.*, 2012).

To date, a lot of efforts have been made towards the understating of DNA transfer of Gram-negative bacteria (de la Cruz *et al.*, 2010). On the other hand, much less information is available for Gram-positive bacteria (Grohmann *et al.*, 2003; Alvarez-Martinez and

Christie, 2009). The main difference in conjugal DNA transfer between these groups is believed to arise in the way these cells form cell-cell contact to initiate the transfer process. Gram-negative bacteria have a periplasmic space between an outer and an inner (cytoplasmic) membrane, which contain a thin layer of peptidoglycan. In contrast, Gram-positive bacteria have only one membrane, the cytoplasmic one, which is covered by a thick outer layer of peptidoglycan.

Modular and molecular organization of conjugation

Plasmids (and other MGEs that undergo HGT via conjugation) can be classified into three types: mobilizable, conjugative (self-transmissible) and non-transmissible (non-mobilizable, non-conjugative) (Fig. 2) (Smillie *et al.*, 2010). MGEs of the first two types contain the mobility set of genes (MOB aka Dtr - DNA-transfer replication) and cognate origin of transfer (*oriT*) DNA segment that allow for DNA nicking/processing and replication. The self-transmissible (conjugative) type elements include, in addition to above, genes for a transmembrane mating apparatus (MPF - mating pair formation), that encode for type IV secretion system (T4SS) machinery that physically connects the lumen of donor and recipient cells (Fig. 3 and Fig. 4). The MOB machinery prepares the single-stranded DNA (ssDNA) for the transfer. The process includes the formation of the relaxosome, a nucleoprotein complex. Two most crucial components of the relaxosome are a relaxase protein, which initiates and terminates conjugative DNA processing, and the DNA *oriT*, which is to be nicked and bound covalently by the relaxase. Another MGE-encoded component called coupling protein (T4CP) interacts with the relaxosome complex and the mating machinery of T4SS to recruit (or couple) the substrate DNA to the transmembrane intercellular apparatus and to facilitate transfer of the relaxase and a ssDNA to a recipient cell.

In contrast to self-transmissible conjugative plasmids, mobilizable plasmids are transferred only in the presence of T4SS encoded by other genetic elements. Plasmids that neither have the *oriT* nor the cognate relaxase are thought to be non-transmissible.

However, it was shown recently that the *Bacillus subtilis* conjugative transposon ICE*Bs1* can mobilize plasmids lacking

dedicated mobilization-*oriT* functions (Lee *et al.*, 2012). Lee and colleagues showed that examined plasmids require for the transfer ICE*Bs1*-encoded T4SS machinery and coupling protein (T4CP), but they do not need the ICE*Bs1* conjugative relaxase or co-transfer of ICE*Bs1*. Authors suggested that in those plasmids the function of conjugative relaxases and their cognate *oriTs* could be replaced by replicative relaxases. Furthermore, certain genomic islands that contain a cryptic *oriT* within their segment, were shown to be mobilized by hijacking the conjugative machinery of a co-resident Integrative Conjugative Element (Waldor, 2010; Daccord *et al.*, 2013). These results blur the current categorization of mobilizable and non-mobilizable plasmids, ICEs and other MGEs that undergo horizontal gene transfer via conjugation. Furthermore, they indicate that conjugative MGEs that escape traditional classification systems may play a role in HGT even to a bigger extent than previously classified and recognized.

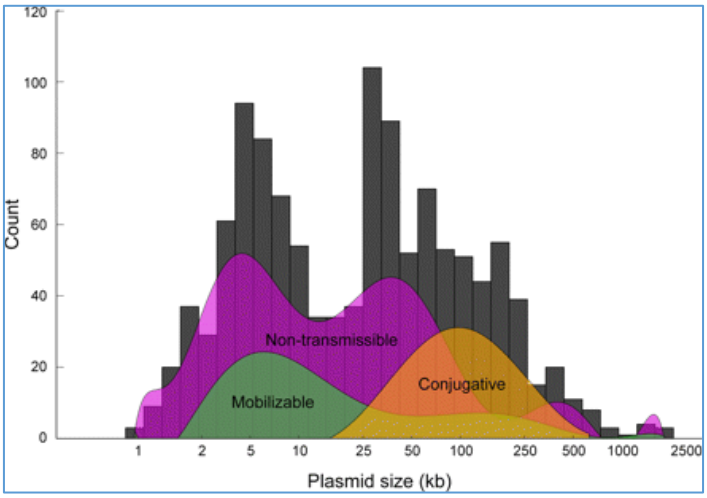


Figure 2. Distribution of conjugative, mobilizable, and non-transmissible (non-conjugative, non-mobilizable) plasmids according to plasmid size. Figure from Smillie *et al.*, (2010). (Curves were created from a polynomial interpolation of the histograms of each class).

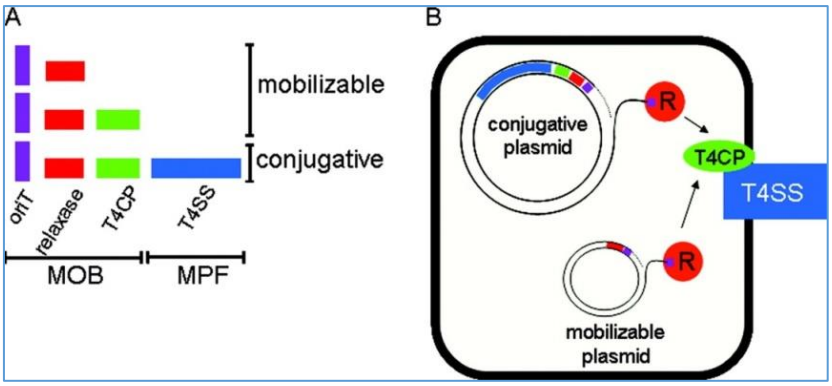


Figure 3. Modular organization of plasmids. (A) Figure from Smillie *et al*., (2010). Schematic view of the genetic constitution of transmissible plasmids. Conjugative plasmids code for the four components of a conjugative machinery: an origin of transfer (*oriT*) (violet), a relaxase (R) (red), a type IV coupling protein (T4CP) (green), and a type IV secretion system (T4SS) (blue). Mobilizable plasmids contain just a MOB module (with or without the T4CP) and need the MPF module delivered in *trans* **(B)** Scheme of some essential steps in the process of conjugation. The conjugation initiation is established when the relaxase cleaves a specific *nic* site within *oriT*. Subsequently, the DNA strand that was nicked and covalently bound (Tyr-DNA 5' end adduct) by the relaxase is displaced from the second strand by a helicase that is either encoded within the relaxase protein or delivered as a separate protein (encoded by MGE or host). Successively, the nucleoprotein complex (relaxosome) is delivered by the T4CP to the T4SS. Finally, the relaxosome is transported to the recipient cell. The energy needed to pump the relaxase-DNA complex through the T4SS channel is delivered by the ATPase activity of the T4CP (VirD4) and T4SS ATPases (VirB4 and VirB11) (Alvarez-Martinez and Christie, 2009).

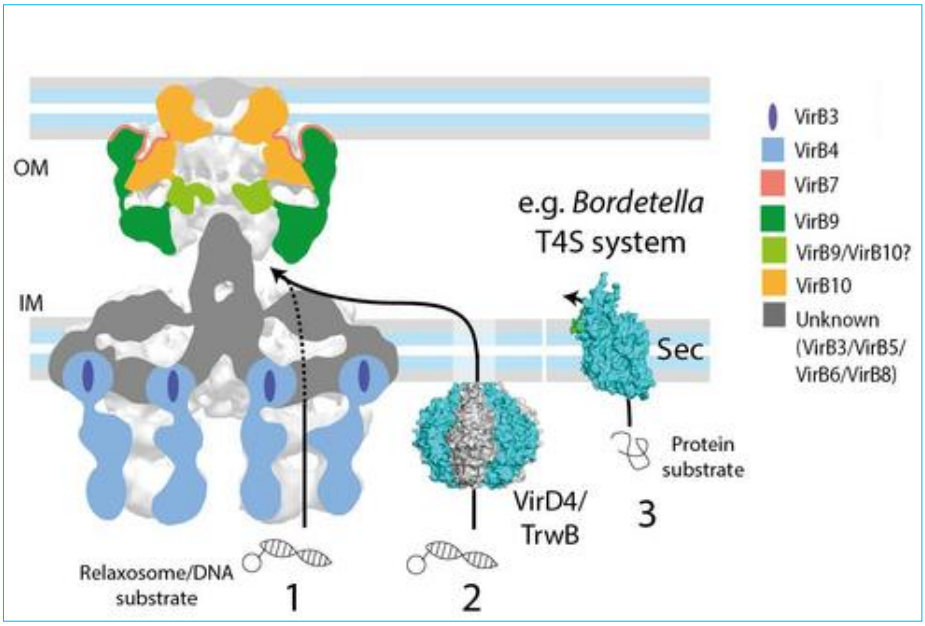


Figure 4. T4SS substrates entry routes. Figure from Low *et al.*, (2014). A scheme showing three possible entry routes across the inner membrane (IM) into the T4SS core complex inner chamber. Substrates composed of a DNA-relaxase complex or a DNA molecule only are probably pumped by the coupling protein VirD4 (route 2) into the periplasm, although passage by T4SS VirB4 protein (route 1) may currently not be discounted. T4S systems that translocate only protein substrates use the Sec pathway to transfer effectors across the inner membrane (route 3).

1.3. Relaxase Families

A classification of plasmid conjugative relaxases based on sequence alignment defined six relaxase families: MOB_P, MOB_F, MOB_V, MOB_Q, MOB_H, and MOB_C (Francia *et al.,* 2004; Garcillan-Barcia *et al.,* 2009; Smillie *et al.,* 2010) (Fig. 5). Above analysis was done using iterative BLASTP searches of the N-terminal 300-residue sequences of six prototypical relaxases versus a database of 1,730 plasmids. The following threshold e-values and archetype relaxases were used: 1e-8 for TrwC_pR388 MOB_F family, 1e-4 for TraI_pR27 MOB_H family, 1e-4 for TraI_pRP4 MOB_P family, 1e-4 for MobA_pRSF1010 MOB_Q family, 1e-5 for MobM_pMV158 MOB_V family and 1e-4 for MobC_pCloDF13 MOB_C family). A similar analysis for Integrative and Conjugative Elements (ICEs) was performed and extended the number of families to eight, with MOB_T and MOB_B being the new groups (Guglielmini *et al.,* 2011). Among other interesting correlations and findings, Guglielmini and colleagues showed that mobilizable elements outnumber self-transmissible conjugative elements in both ICEs and plasmids, which implies an extensive in *trans* use of T4SS by the mobilizable elements. Interestingly, except for a less studied MOB_B group, each of MOB families has its equivalent in protein domain databases, e.g. a search in the NCBI Conserved Domain Database (CDD) (Marchler-Bauer *et al.,* 2011) outputs: TrwC (pfam08751) family and relax_trwC (TIGR02686) family for MOB_F family, MobA_MobL (pfam03389) for MOB_Q, Relaxase (pfam03432) for MOB_P, Mob_Pre (pfam01076) for MOB_V, TraI_2 (pfam07514) and ICE_TraI_Pfluor (TIGR03760) for MOB_H, Replic_Relax (pfam13814) for MOB_C, Rep_trans (pfam02486) for MOB_T.

Structural biology of relaxases

Known MOB relaxases use at least three nuclease folds: the HUH endonuclease fold (MOB_F, MOB_Q, MOB_P, MOB_V), the PD-(D/E)XK restriction endonucleases fold (MOB_C) and the HD-hydrolase fold (MOB_H) (Garcillán-Barcia *et al.,* 2009; Chandler *et al.,* 2013; Francia *et al.,* 2013). These three structurally different types of

relaxases use a Tyr residue to covalently bind and nick the substrate DNA. X-ray crystal structures are available only for relaxases of two families, multi-tyrosine (Y2) MOB_F relaxases: TrwC_pR388 (Guasch *et al.*, 2003; Boer *et al.*, 2006) TraI_pF (Datta *et al.*, 2003; Larkin *et al.*, 2005) and TraI_pCU1 (Potts *et al.*, 2010), and single-tyrosine (Y1) MOB_Q relaxases: MobA_pR1162 (Monzingo *et al.*, 2007) and NES_pLW043 (Edwards *et al.*, 2013) (Fig. 5).

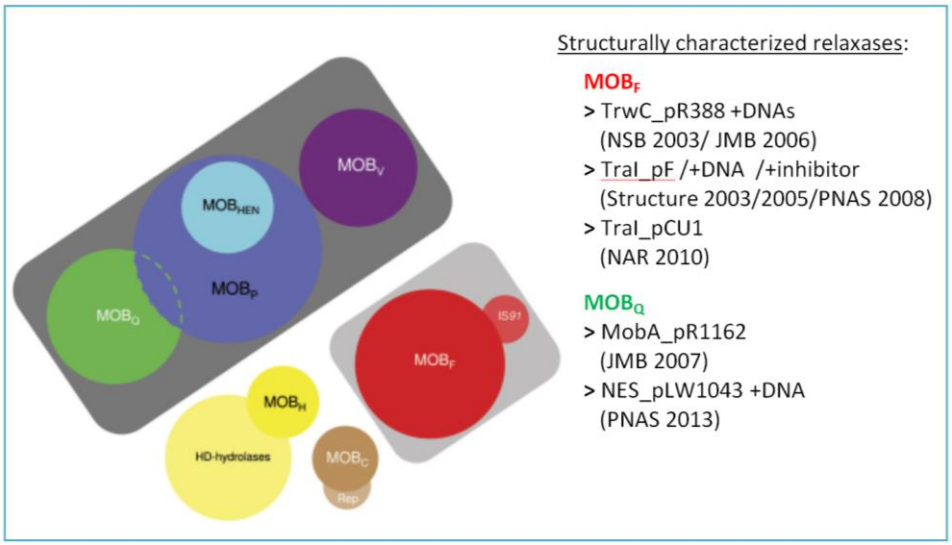


Figure 5. Schematic view of the relaxase families relationships. Figure adapted from Garcillán-Barcia *et al.*, (2009). Dark-grey cluster marks single active-site-tyrosine (Y1) families. Light-grey cluster marks multiple active-site-tyrosines (Y2) families. For the other families not enough biochemical data is known to assign them without problems. Size of a circle corresponds to the family size. Right side presents list of structurally described relaxases and details of their crystal structure composition.

MOB_V family of relaxases

MOB_V relaxases found in over 100 plasmids were shown to be overrepresented and Firmicutes and Bacteroidetes (where they are associated with antibiotic resistance genes), but they are also found in a small extent in Proteobacteria, Cyanobacteria and Spirochaetes

(Garcillán-Barcia *et al.,* 2009; Smillie *et al.,* 2010). MOB_{V1} subfamily of the prototypical MobM_pMV158 is the most populated group and gathers relaxases of plasmids belonging to several Inc groups from Firmicutes. MOB_V relaxase-*oriT* systems tend to be simple, as they employ no auxiliary proteins and contain relatively short *oriT*s of less than 100 bp (Smillie *et al.,* 2010). MOB_V relaxases are found almost exclusively among small mobilizable plasmids (Fig. 6) and are never associated with a cognate T4CP (Smillie *et al.,* 2010). Moreover, MOB_V relaxases are very promiscuous in their use of different subtypes of the T4SS/MPF apparatus (Farias and Espinosa, 2000; Francia *et al.,* 2004). Noteworthy, CDD (Marchler-Bauer *et al.,* 2011) lists (as for August 2014) 4045 specific (with an e-value threshold 4.9e-43) and 8061 general (with an e-value threshold 1.0e-2) protein hits for the CDD equivalent of the MOB_v family, the Mob_Pre family (Table 1; note that these numbers are constantly growing and that the values found in the table may be different already). Similarly to the analysis of MOB_V relaxases from plasmids, the Mob_Pre relaxases are profoundly found in Firmicutes (98% of specific hits found among bacteria).

Except for the most divergent MOB_{V5} subfamily that lacks Motif I, all other MOB_V clades are characterized by two conserved sequence motifs, Motif I - HxxR and Motif III - HxDEx(T/S/N)PHUH (Fig. 7), with clade $MOB_{V4.1}$ containing a modified version of the latter - $HKDEx_{8–24}PNxHUH$ (Garcillan-Barcia *et al.,* 2009; for clarity, note that Motif I in MOB_{V3} clade is in Garcillan-Barcia *et al.,* (2009) incorrectly assigned to a 'IxxRxxE' conserved sequence motif). When comparing the conserved motifs of MOB_V family to the ones of other families of HUH relaxases (MOB_F, MOB_Q, MOB_P) it seems obvious that they are quite different (Garcillan-Barcia *et al.,* 2009). The most striking difference is the lack of conserved tyrosine among MOB_V relaxases, although another landmark of HUH-type MOB relaxases, the metal chelating histidine triad of Motif III is preserved. Interestingly, single residue mutations of all tyrosine residues in the MOB_{V2} relaxase Mob_pBHR1 did not alter plasmid mobilization frequencies, although DNA cleavage/relaxation activity was not assayed (Szpirer *et al.,* 2001). This observation led authors to the suggestion that the Mob_pBHR1 relaxase might have more than one catalytic Tyr residue (Y2 relaxases) or alternatively that a catalytic mechanism other than

that of the known relaxases is involved. Based on mutagenesis experiments, authors hypothesized that D120 and/or E121 may be the catalytic residue(s) (Szpirer *et al.*, 2001). These residues correspond to the invariant Asp and Glu residues of MOB_V Motif III.

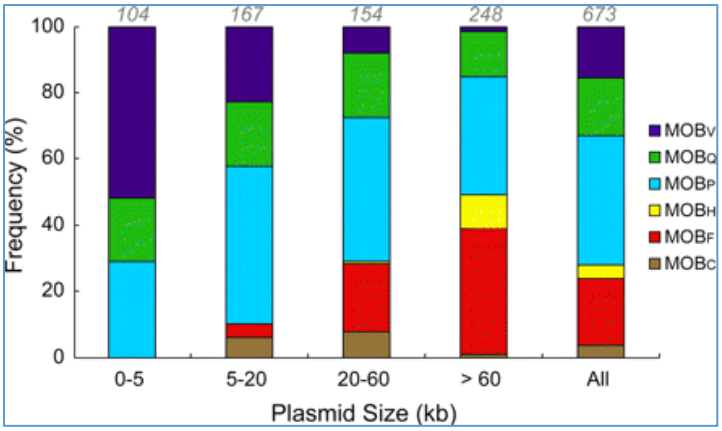


Figure 6. Distribution of relaxase families according to plasmid size. Figure from Smillie *et al.*, (2010). Each bar shows the frequency of each relaxase family for the given plasmid size range. On the top of each bar number of included plasmids is shown.

Relaxases occurrence in pathogenic Gram-positive bacteria

According to minimal estimates of the USA Center for Disease Control and Prevention (CDC), each year in USA more than 2,000,000 people are infected with antibiotic-resistant microorganism and at least 23,000 of such cases cause patient death (CDC Threat Report, 2013). Lethal antibiotic-resistant infections with Gram-positive bacteria of *Staphylococcus, Streptococcus* and *Enterococcus* genera account for 20,000 cases and majority of them (11,000) are attributed to methicillin-resistant *Staphylococcus aureus* (MRSA; a term used to describe *S. aureus* strains resistant to penicillins). MRSA was traditionally particularly problematic in hospitals, where patients with weakened immune system and open wounds have higher risk of infection than the healthy individuals in general public, however there are growing evidences that MRSA dynamically evolves and can be

now community-acquired, thus it can infect young and healthy people in households and community places and indeed in many urban regions CA-MRSA infections now appear to be endemic and cause most MRSA infections (David and Daum, 2012; David *et al.*, 2014; Uhlemann *et. al*, 2014). Further, many more people suffer or die from health conditions resulted from medical complications related to antibiotic-resistant infection. Moreover, almost 250,000 people each year need hospitalization (and 14,000 of them die) due to *Clostridium difficile* infections, which mostly happen to people that had invasive antibiotics treatment due to recent medical care

The major route of acquisition of antibiotic resistance by HGT is bacterial conjugation. The process can be enhanced when a bacterial defense against exogenous DNA is weakened. For example, certain clinical MRSA strains lack type III-like restriction endonuclease systems and are therefore highly susceptible to the gene transfer from other species (Corvaglia *et al.*, 2010). A host jump can constitute one of the most dramatic examples of genetic adaptation. In addition to be a human pathogen, *S. aureus* is also a major cause of skeletal infections of poultry, which are of high economic burden on the worldwide poultry industry. The majority of *S. aureus* isolates from broiler chickens are the descendants of a single human-to-poultry host jump that happened approximately 40 years ago (Lowder *et al*., 2009).

NCBI Conserved Domains Database (data from August 2014) outputs the Mob_Pre/MOBv relaxases as the dominating type of HUH conjugative relaxases found in *S. aureus* (Table 1). Furthermore, the top listed organism for the Mob_Pre/MOB_V family is *S. aureus* (69% of Mob_Pre family specific hits found in bacteria come from *S. aureus*). Interestingly, bacterial human pathogens from *Staphylococcus*, *Streptococcus* and *Enterococcus* genera that were listed in the 2013 CDC Threat Report (CDC, 2013) are very common hosts for relaxase genes (and therefore their cognate plasmids/ICEs) of Mob_Pre/MOB_V and Rep_trans/MOB_T families, accounting respectively for 83% and 71% of specific hits found for these families in all bacteria (Table 1). Noteworthy, the number of Mob_Pre family members found in *S. aureus* in recent years is growing exponentially (Table 2). The same scenario is observed only for one more MOB family, the Rep_trans/MOBT family (data not shown).

CDD Family (MOB Family)	Threshold, specific and general	*S. aureus*	*Streptococcus*			*Enterococcus*		Sum of CDC G+ pathogens (columns III - VIII)	Firmicutes	Bacteria	Fraction of *S. aureus* to all bacteria	Fraction of CDC G+ pathogens	
			agalactiae	*pneumoniae*	*pyogenes*	*faecalis*	*faecium*					among Firmicutes	among all bacteria
Mob_Pre (MOB_V)	4.9e-43 1.0e-2	2790 3211	202 253	1 1	12 12	177 200	90 312	3272 3989	3954 6294	4045 8061	0.69 0.40	0.83 0.63	0.81 0.49
Relaxase (MOB_P)	5.8e-51 1.0e-2	259 305	0 360	52 283	0 105	20 437	105 305	436 1795	791 5376	5975 17419	0.04 0.02	0.55 0.33	0.07 0.10
MobA_MobL (MOB_Q)	4.9e-70 1.0e-2	259 299	15 15	76 165	8 18	32 39	16 25	406 561	1060 2667	1292 7649	0.20 0.04	0.38 0.21	0.31 0.07
TrwC (MOB_F)	3.8e-52 1.0e-2	0 0	0 0	0 0	0 0	0 0	0 0	0 0	8 68	1672 6002	0.00 0.00	0.00 0.00	0.00 0.00
TraI_2 (MOB_H)	1.8e-94 1.0e-2	0 6	0 0	0 0	0 0	0 365	0 257	0 628	0 1821	1732 5856	0.00 0.00	0.00 0.34	0.00 0.11
Replic_relax (MOB_C)	8.5e-29 1.0e-2	0 22	0 0	0 46	0 0	0 277	0 9	0 354	583 2246	954 3737	0.00 0.01	0.00 0.16	0.00 0.09
Rep_trans (MOB_T)	2.5e-8 1.0e-2	3446 3465	566 568	143 172	15 15	939 954	883 906	5992 6080	8390 8697	10354 11248	0.33 0.30	0.71 0.70	0.58 0.54

Table 1. Distribution of relaxases in Gram-positive human pathogens that were listed in the CDC Threat Report "Antibiotic resistance threats in the United States, 2013"; Based on the NCBI Conserved Domain Database, Data from August 2014.

Mob_Pre (MOB_V)	threshold e-value: specific and general	31.12.2011	31.12.2012	31.12.2013	31.08.2014
Staphylococcus aureus	4.9e-43 1.0e-2	95 165	153 300	403 614	2790 3211
Bacteria	4.9e-43 1.0e-2	408 1257	584 1928	1599 5016	4045 8061
Faction of S. aureus among all bacteria	4.9e-43 1.0e-2	0.23 0.13	0.26 0.16	0.25 0.12	0.69 0.40

Table 2. Number of Mob_Pre (MOB_V) family members found in *Staphylococcus aureus* by the end of last three years. Based on the NCBI Conserved Domain Database, Data from August 2014.

1.4. Plasmid pMV158-encoded MobM Relaxase

Taking into account the medical relevance, unknown catalytical mechanism and lack of the structural information for any of the MOBv/Mob_Pre relaxases, we decided to carry out an attempt to structural and functional characterize the prototypical member of the family, the MobM relaxase from pMV158 plasmid.

MobM from the promiscuous streptococcal antibiotic resistance pMV158 plasmid is a 494-residue protein that forms dimers in solution (de Antonio *et al.,* 2004). The MobM relaxase binds strongly to pMV158 origin of transfer (*oriT*), retains in a stable complex with the target DNA upon DNA nicking, expresses the optimal nicking activity at pH 6.5, and is capable of nicking *oriTs* of other MOBv plasmids (Fernández-López *et al.,* 2013). The N-terminal 194-residues constitute one the 18 prototypical members of the Mob_Pre family of relaxases (Fig.8; NCBI Conserved Domains Database; Marchler-Bauer *et al.,* 2011). MobMN199 construct (residues 2-199; Met1 is eliminated *in vivo*) is monomeric in solution, retains binding properties on the oligonucleotides that mimic the pMV158 origin of transfer (*oriT*) and nicking activity on plasmid supercoiled DNA (Lorenzo-Diaz *et al.,* 2011). MobM relaxase was shown to require Mg/Mn ions for the DNA cleavage, however only Mn ions bind to the protein with high affinity and stabilize MobM supporting its enhanced resistance to the thermal denaturation (Lorenzo-Diaz *et al.,* 2011). The MobM C-terminal domain is involved in protein dimerization through a putative leucine zipper motif localized within one of three predicted coild coils regions (Fig.9) and is proposed to anchor the protein to the cell membrane and to participate in interactions with T4CP and T4SS (de Antonio *et al.,* 2004; Lorenzo-Diaz *et al.,* 2011). In contrast to the N-terminal relaxase/nuclease domain that is clearly defined as a member a Mob_Pre/MOBv family (specific, the C-terminal domain is predicted to be only distantly related to the chromosome segregation proteins, which would be in line with the MobM membrane association. The low specificity family score of the C-terminal domain of MobM suggests that of the C-terminal domains of Mob_Pre/MOBv relaxases have

small sequence conservation pressure and therefore presumably lack any enzymatic activity (Fig. 8).

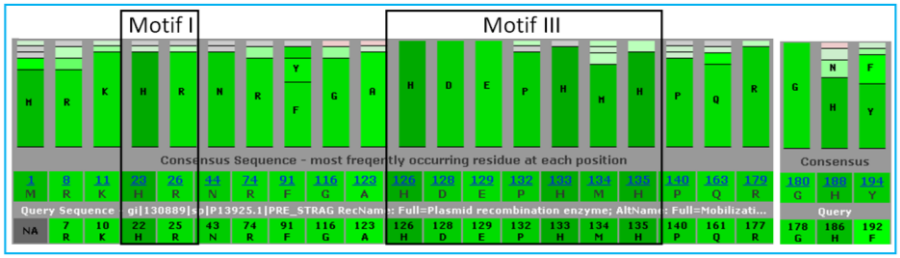


Figure 7. Sequence consensus and Motif I and III of the Mob_Pre/MOB$_V$ family. In the upper part, residue frequency bars are presented; the darkness of green colour correlates with the highest scores in the position-specific scoring matrix (PSSM). The figure was created with the NCBI PSSM Viewer.

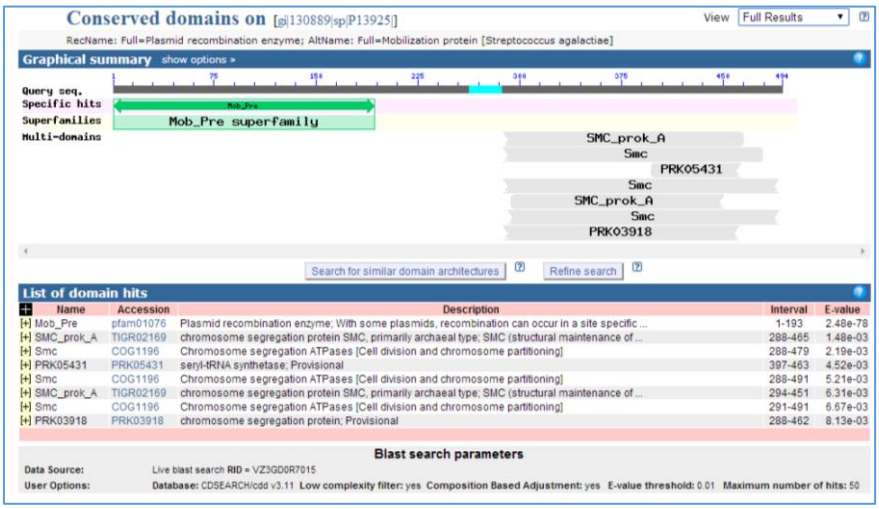


Figure 8. Conserved Domain Database analysis of the MobM relaxase. The search outputs for MobM two domains, the N-terminal highly ranked Mob_Pre domain and the C-terminal lowly ranked C-terminal domain of chromosome segregation protein.

Figure 9. Pfam graphical representation of the MobM relaxase. Mob_Pre – a short name of the "plasmid recombination enzyme" family; grey rectangles – predicted disordered regions; green rectangles - predicted coiled coil regions.

CHAPTER 2
OBJECTIVES

The main objective of this dissertation is to provide detailed structural analysis of the prototypical representative of MOB_V/Mob_Pre family of relaxases, the MobM relaxase encoded by the mainly Gram-positive promiscuous mobilizable antibiotic resistance plasmid pMV158. In other words, to propose, based on the structural, literature, mutational and conservation data, the mechanism of action of the MOBV relaxases during initial steps of the conjugative DNA transfer. Further, to provide structure-function comparison of predominantly Gram-positive MobM relaxase (MOB_V family) to the structurally described conjugative relaxases, which are either found predominantly in Gram- bacteria (MOB_F relaxases) or are found in both Gram+ and Gram- bacteria (MOB_Q relaxases).

CHAPTER 3
RESULTS

3.1. Structures of DNA-MobM relaxase complexes reveal a histidine/metal catalysis for DNA cleavage and ligation

The following manuscript was submitted to Nature Structural and Molecular Biology as a brief communication:

Structures of DNA-MobM relaxase complexes reveal a histidine/metal catalysis for DNA cleavage and ligation

Radoslaw Pluta[1,2,+], D. Roeland Boer[1,2,+], Cris Fernández-López[3], Silvia Russi[1,2,§], Fabián Lorenzo-Díaz[3,§], Rosa Pérez-Luque[1,2], Manuel Espinosa[3], Miquel Coll[1,2,*]

[1]Institute for Research in Biomedicine (IRB Barcelona), Baldiri Reixac 10-12, 08028 Barcelona, Spain

[2]Institute of Molecular Biology of Barcelona (IBMB-CSIC), Baldiri Reixac 10-12, 08028 Barcelona, Spain

[3]Centro de Investigaciones Biológicas (CIB-CSIC), Ramiro de Maeztu, 9, 28040 Madrid, Spain.

[+] These authors contributed equally

* Corresponding author

[§]Present address: SR - SLAC National Accelerator Laboratory, 2575 Sand Hill, Menolo Park, CA 94025, USA; FLD - Instituto Universitario de Enfermedades Tropicales y Salud Pública de Canarias, Centro de Investigaciones Biomédicas de Canarias (CIBICAN), Spain

A new class of endonuclease/ligase is presented, exemplified by the conjugative relaxase MobM of the MOB_V family. This enzyme uses a single metal ion and a histidine active residue to produce a covalent phosphohistidine bond that leads to a protein-DNA adduct. Five protein-DNA complex snapshots outline the reaction mechanism, which is validated by biochemical approaches.

Main

Endonuclease/ligase MobM is encoded by the promiscuous plasmid pMV158 and it recognizes and nicks the pMV158 *oriT* region (Fig. 1) on one of its strands.[1] Endonucleases/ligases of this type are called relaxases since they relax the supercoiled DNA and initiate its mobilization to another bacterial host of the same or different species. This process of horizontal gene transfer is responsible for rapid bacterial evolution and adaptation and the spread of antibiotic resistance among pathogenic bacteria. Relaxases and other enzymes of the HUH endonuclease superfamily depend on a single metal ion for the activation of the scissile phosphate and on an amino acid that attacks the phosphor atom to form a pentacovalent intermediate, thus leading to DNA cleavage and protein-5'-DNA adduct formation.[2,3] The reverse reaction –ligation– occurs in the recipient cell and restores the circular DNA. The catalytic residue, which performs the nucleophilic attack and to which the DNA remains covalently attached during the transfer process, has been described as a tyrosine residue. Both mutational and structural studies of relaxases of the MOB_Q and MOB_F families confirm this notion.[4-8]

Full-length MobM is a 57.9-kDa protein[9] that belongs to the MOB_V family of relaxases.[10] The N199 construct used in this study retains DNA-binding and relaxase activities on supercoiled plasmid DNAs.[11] However, mutation of all tyrosine residues of a related MOB_V member did not alter plasmid mobilization frequencies, although relaxation activity was not assayed;[12] this observation led to the notion that a catalytic mechanism (a Glu residue was proposed) other than that of the known relaxases was involved.

Here we present five crystal structures of MobMN199 in complex with various substrates that represent the extruded DNA hairpin of the *oriT* region, which is recognized and cleaved by the relaxase. The structures outline the reaction mechanism and show that the catalytic residue for nucleophilic attack on the scissile phosphate and DNA-protein adduct formation is a histidine instead of a tyrosine. This is the first description of an enzyme with a metal/histidine catalytic mechanism for DNA cleavage and ligation.[13] Mutational studies confirm this finding, and a series of crystallographic protein-DNA snapshots outline how H22 attacks the metal-polarized phosphodiester bond.

In some enzymes, the imidazole group of histidine acts as a nucleophile in phosphate transfer reactions[14], and phosphohistidine adducts occur in two-component signal transduction proteins and other systems[15], although in none of these reactions is DNA the substrate. Some members of the phospholipase D (PLD) superfamily which have nuclease activity, hold catalytic histidines but bear no sequence or structure relationship with relaxases and are metal-ion independent enzymes.[16,17]

The substrate oligonucleotides used for our structural analysis (Suppl. Table S1) mimic various states of the relaxation reaction. The nic0 oligonucleotide ends at the cleavage site and represents the 5' reaction product. The nic0-PO and nic0-SPO oligonucleotides, respectively, include the scissile phosphate and its thiophosphate analog. Finally, nic0+1 includes a thymidine beyond the cleavage site.

The overall structure shows that the protein has the relaxase α/β plate fold, reminiscent of the DNA polymerase Klenow fragment (Fig. 1).[3] Details of the structure and DNA binding will be discussed elsewhere (RP, DRB, SR, CFL, FLD, RPL, ME & MC, in preparation). Here we will focus on the placement of the DNA in the active site. T25 is highly solvent-exposed and marks the start of a U-turn that redirects the ssDNA towards the active site, thus allowing G26 to make a non-Watson Crick (wobble) base pair with T23 (Fig. 1b). This interaction directs the 3' oxygen of G26 towards the metal ion and places the consecutive scissile phosphate within reach of the catalytic H22.

Residues H126, H133, H135 and E129 coordinate a Mn^{2+} ion with octahedral geometry (Figs. 1 & 2), the identity of which was unequivocally confirmed (Suppl. Fig S3). The fifth metal ligand is invariably the 3' oxygen of G26. The sixth ligand is the $N\varepsilon^2$ nitrogen atom of the catalytic H22, although this interaction is weak and its occurrence depends on the stage of the reaction. It is therefore not essential for metal coordination but is related to the activation of $N\varepsilon^2$ as a nucleophile since the deprotonated form of this atom and hence one of the histidine tautomers is favoured by the interaction with the metal. The interaction also helps to position correctly the $N\varepsilon^2$ for attacking the scissile phosphate.

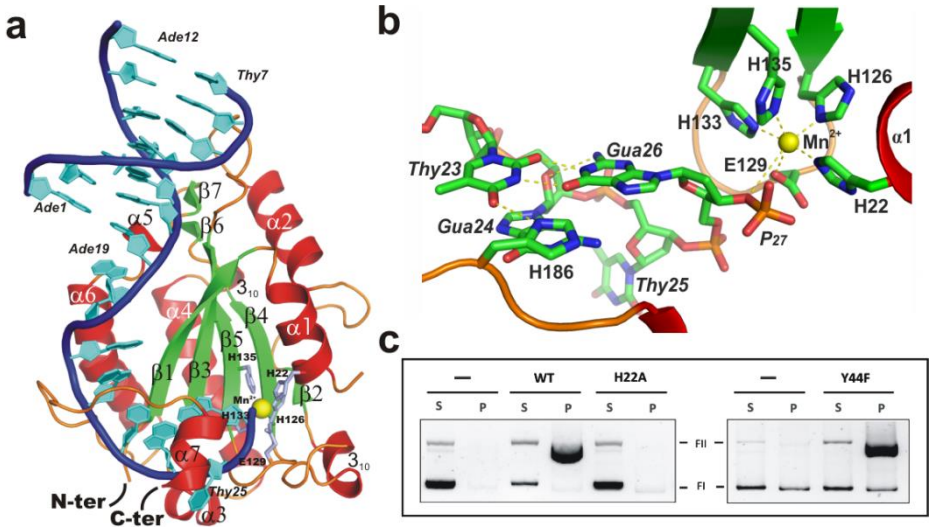


Figure 1. MobM-DNA complex. a) Overall structure of the MobM-Nic0 complex. The active site contains a Mn2+ ion, coordinated by a His-triad and E129, and the nucleophilic H22. **b**) Detail of the hydrogen bonding scheme of G26 in the active site in the Nic0+1 structure. The Mn2+-bound phosphate moiety of T27 is in close proximity to H22. **c**) Agarose gel electrophoresis showing the formation of the covalent adduct in relaxosome formation by MobM (wild-type and Y44F) proteins with the pMV158 circular plasmid through the appearance of an intermediate band, which is absent for the H22 mutant.

Consistent with this feature, an inward movement of H22 with increasing pH is observed, e.g. when comparing the Nic0 pH 4.6

structure with that of Nic0 pH 5.5 (Suppl. Fig. 1d-e), and the structure of Nic0-PO pH 4.6 with that of Nic0-SPO pH 6.8 (Fig. 2a-b). In fact,

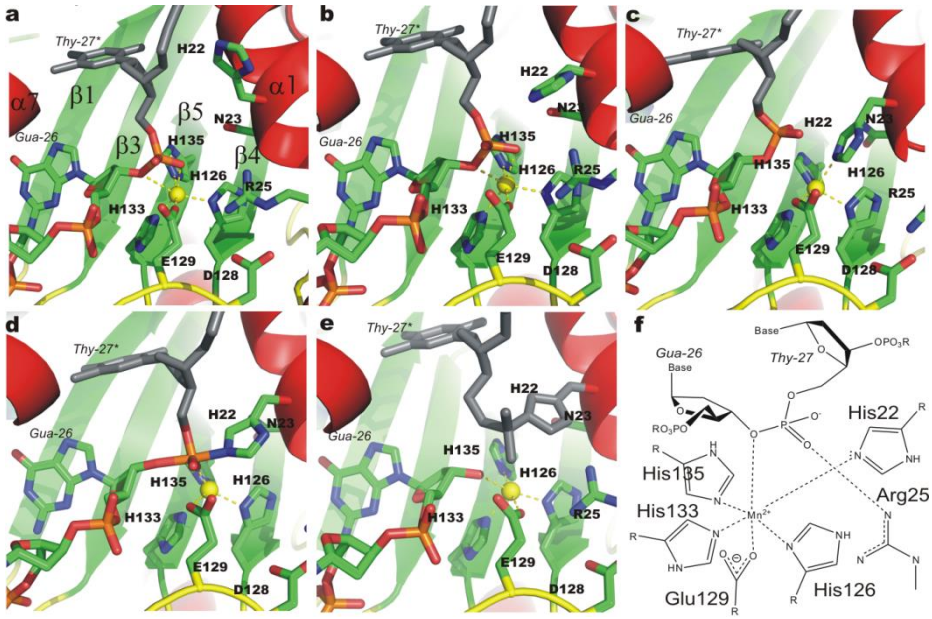


Figure 2. The active sites of the X-ray structures and pentacovalent intermediate, chronologically ordered. Modeled atoms not present in the crystal structures are shown in gray. Panels **a-c)** and **e)** are reproduced in Suppl. Fig. 1**a-d)**, which additionally shows the electron densities. **a)** Nic0-PO at pH 4.6: H22 is outward, **b)** Nic0-SPO at pH 6.8: H22 moves partially inward, **c)** Nic0+1 at pH 6.5: H22 is now in position for nucleophilic attack. **d)** The modelled pentacovalent intermediate e) Nic0 at pH 5.5: H22 switches position, taking the bound (modelled) DNA along. **f)** The proposed reaction mechanism of MobM based on the snapshots.

MobM is about 25% less efficient at low pH.[18] In the Nic0+1 (pH 6.5) structure, H22 is inward and coordinated to the Mn^{2+}. The scissile phosphate is displaced upwards compared to the Nic0-PO and Nic0-SPO structures and interacts with the plane of the imidazole ring of H22. Its phosphorus atom is found 3.7 Å from H22 $N\varepsilon^2$, whereas the G26 3' oxygen is 3.6 Å from Mn^{2+} (Fig. 2c). A slight turn of the H22 imidazole ring would allow the nucleophilic attack to occur (Fig. 2d, see Suppl. Movie). The Nic0+1 structure therefore represents a snapshot of the complex immediately before the cleavage reaction, with

all the key residues in place. The observation that the nick does not occur is in accordance with biochemical data showing that N199 does not cut ssDNA.[11] Supercoiled DNA may induce the cut by imposing required steric strain and/or induce correct placement of assisting residues, and we have evidence that the binding of MobM to its target induces extrusion of the nick-containing hairpin (FLD and ME, unpublished). After nicking, outward movement of H22 would withdraw the cleaved, covalently bound DNA, thus preventing the backward reconstitution reaction and allowing the entry of newly synthesized DNA for the final closing reaction (Fig. 2e).

The H22Y and H22A mutants confirmed the essential role of the H22 residue. *In vitro* formation of covalent wild-type protein-DNA adducts in relaxosomes was readily achieved, whereas in H22A (Fig. 1c and Suppl. Fig. S2) and H22Y mutants (Suppl. Fig. S2) protein-DNA adducts did not form. This was not the case for the Y44F mutant, which behaved like the wild-type (Fig. 1c and Suppl. Fig. S2) (Y44 was the predicted catalytic residue since it is conserved and is the most N-terminal tyrosine[9,10]). Furthermore, construction of plasmid pMV158H22A (a single amino acid change, H22A, in the entire MobM protein) led to total failure to be mobilized, whereas a mobilization frequency of 2.5×10^{-4} was found for the pMV158 wild-type plasmid. In agreement with previous data,[19] not a single transconjugant (frequency $< 1.5 \times 10^{-10}$, the experimental detection limit) was rescued in four independent experiments.

A mechanism explaining the histidine-mediated attack on the phosphodiester bond is schematized in Fig. 2f. Phosphohistidine adducts are less stable than phosphotyrosine ones, the phosphoryl group being readily transferred to a second substrate. The physiological consequences of this difference for DNA transfer remain to be unveiled.

Acknowledgements

This work was supported by the Spanish Ministry of Economy and Competitiveness (BFU2008-02372/BMC, CSD 2006-23, BFU2011-22588, CSD-2008-00013, INTERMODS and BFU2010-19597), the Generalitat de Catalunya (SGR2009-1309) and the European Commission (GA No. 260644). RP is the recipient of a "La Caixa"/IRB Barcelona PhD Fellowship. We thank Dr. W.T.Chan for constructing plasmid pMV158H22A, L. Rodríguez for help with protein purification, the staff of the Platform for Automated Crystallization (IBMB, Barcelona, Spain), and David Aparicio and Luca Martinelli for collecting the Nic0+1 diffraction data.

References

1. ***Guzmán, L. & Espinosa, M. J. Mol. Biol. 266, 688-702 (1997).***
2. ***Chandler, M. et al. Nat. Rev. Microbiol. 11, 525-38 (2013).***
3. ***Guasch, A. et al. Nat. Struct. Biol. 10, 1002-10 (2003).***
4. ***Boer, R. et al. J. Mol. Biol. 358, 857-69 (2006).***
5. ***Larkin, C. et al. Structure 13, 1533-1544 (2005).***
6. ***Monzingo, A.F., Ozburn, A., Xia, S., Meyer, R.J. & Robertus, J.D. J. Mol. Biol. 366, 165-178 (2007).***
7. ***Potts, R.G., Habibi, S., Cheng, Y., Lujan, S.A. & Redinbo, M.R. Nucl. Acids Res. 38, 1-15 (2010).***
8. ***Edwards, J.S. et al. Proc. Natl. Acad. Sci. U. S. A. (2013).***
9. ***de Antonio, C., Farias, M.E., de Lacoba, M.G. & Espinosa, M. J. Mol. Biol. 335, 733-743 (2004).***
10. ***Garcillán-Barcia, M.P., Francia, M.V. & de la Cruz, F. FEMS Microbiol. Rev. 33, 657-687 (2009).***
11. ***Lorenzo-Diaz, F. et al. Nucleic Acids Res. 39, 4315-29 (2011).***
12. ***Szpirer, C.Y., Faelen, M. & Couturier, M. J. Bacteriol.183, 2101-10 (2001).***
13. ***Yang, W. Q. Rev. Biophys. 44, 1-93 (2011).***
14. ***Fersht, A. Structure and Mechanism in Protein Science - A Guide to Enzyme Catalysis and Protein Folding, (W.H. Freeman and Co, New York, 1999).***
15. ***Puttick, J., Baker, E.N. & Delbaere, L.T. Biochim. Biophys. Acta 1784, 100-5 (2008).***

16. Stuckey, J.A. & Dixon, J.E. Nat. Struct. Biol. 6, 278-84 (1999).
17. Sasnauskas, G. et al. Nucleic Acids Res. 38, 2399-410 (2007).
18. Fernandez-Lopez, C. et al. Plasmid 70, 120-30 (2013).
19. Lorenzo-Diaz, F. & Espinosa, M. Plasmid 61, 65-70 (2009).

Online Methods

Bacterial strains, plasmids and mutagenesis

Escherichia coli BL21 (DE3) (λDE3 (*lac*I*lac*UV5-T7 gene 1 *ind*1 *sam*7 *nin*5) F$^-$*dcmomp*T*hsd*S ($r_B^-m_B^+$) *gal*) was used to purify MobM and MobMN199 proteins. *E. coli* B834 (DE3) (λDE3 (*lac*I*lac*UV5-T7 gene 1 *ind*1 *sam*7 *nin*5) F$^-$*dcmomp*T*hsd*S($r_B^-m_B^+$) *gal met*) was employed to purify MobMN199 with SeMet[20]. *E. coli* M15 / pREP4 (*NaIS, StrS, RifS, Thi–, Lac–, Ara+, Gal+, Mtl–, F–, RecA+, Uvr+, Lon+*) was used to express the His-tagged MobMN199 mutants H22A, H22Y and Y44F. Purified pMV158 supercoiled DNA and other plasmids used for expression or as substrates of MobM and MobMN199 proteins were prepared as reported[21,22]. To obtain the MobMN199 variants, point mutations were introduced using the GeneTailor Site-Directed Mutagenesis System (Invitrogen). Details of the cloning and production will be published elsewhere (SR, DRB, RP, CFL, FLD, RPL, ME & MC, in preparation).

Relaxation activity assays

Assays of scDNA relaxation by purified proteins were performed essentially as reported.[23,24] Purified DNA from pMV158 (500 ng) was treated with the indicated amounts (Fig. S2) of purified MobMN199 proteins (wt or mutants). Samples were incubated and treated as reported[21]. The generation of open circular forms (FII) by the nicking activity of the various relaxase variants was tested by electrophoresis on 1 % (w/v) agarose gels and staining with ethidium bromide 1 μg ml^{-1}.

Pull-down assays

Supercoiled pMV158 DNA (500 ng) was relaxed with 200 nM of proteinMobMN199 (wt or mutants), and stable DNA-protein complexes were precipitated by treatment with 180 mMKCl, 0ºC, 10 min. Samples were centrifuged and the pellets (p) and supernatants (s) were subsequently incubated with proteinase K as reported[25]. Separation of the plasmid forms was achieved by electrophoresis on 1% agarose in the presence of 1 μg/ml ethidium bromide, at 100 V 1 h.

Protein-DNA complex preparation

For the preparation of SeMet-labeled MobM-DNA complex, protein in buffer A (500mMNaCl, 20mMTris-HCL pH7.6, 1mM EDTA, 1% glycerol, 1mM DTT) was mixed at 1:1.2 protein:DNA stoichiometry with a previously annealed oligonucleotide encompassing the 26 nucleotides upstream of the oriT cleavage site (see Table S1, all oligonucleotides from Biomers, Ulm, Germany) and incubated on ice for 1h. For the preparation of the other complexes, a similar procedure was used but mixing MobMN199 in buffer B (100mMNaCl, 20mM Tris-HCl, pH7.5, 15mM $MnCl_2 \cdot 4H_2O$, 1% glycerol, 1mM DTT) with the respective dsDNAs (Table S1). All oligonucleotides were previously annealed, except for nic0-SPO because the thiophosphate group was not stable at high temperatures. The resulting complexes were purified by size-exclusion chromatography and concentrated to 4 mg/ml.

Crystallization

Crystals of the SeMetMobM-DNA complex were grown by sitting-drop vapor diffusion in buffer A against a crystallization buffer containing 100mMNaAc, pH4.6 and 10% PEG 6000, at 4ºC. Crystals of MobM with the Nic0, Nic0-PO and Nic0-SPO dsDNAs were grown by sitting-drop vapor diffusion of the complex in sample buffer B at 20ºC against a crystallization buffer containing 100mMNaAc, pH4.6 and 18-23% PEG 6000. Micro seeding was used for Nic0-SPO and Nic0-PO crystal growth. Crystals of the Nic0+1 complex were obtained in a similar way against a crystallization buffer containing 100mM MES, pH6.5, 200mM $MgAc_2 \cdot 4H_2O$ and 20% PEG 8000. All crystals

were harvested in cryosolutions containing crystallization cocktail supplemented with 15-25% glycerol.

For the Nic0 pH 5.5 and Nic0-SPO pH 6.8 structures, the initial crystals were soaked in cryosolutions in which the original 100mMNaAc, pH 4.6 buffer was replaced by 100mM NaAc buffer at pH 5.5 and 6.8, respectively.

X-ray data collection and refinement

The details of the data collection at beamlines ID14-4, ID23-1 and ID23-2 of the European Synchrotron Radiation Facility (ESRF) and at beamline PROXIMA1 of the French synchrotron facility Soleil are summarized in Table S2. All data were processed with iMOSFLM[26] and scaled with SCALA.[27] The experimental phases were obtained by the SAD method, using the SeMet-DNA26 data set, collected at the Se absorption edge (0.9795Å). Seven Se sites were located in the asymmetric unit using the program SHELXD[28,29], and protein phases were calculated with PHASER,[30] improving the initial electron density map by density modification with the program PIRATE[31]. 30% of the polypeptidic chain was automatically traced with the program RESOLVE[32] and completed manually with the program COOT.[33] This initial SeMet structure was trimmed, removing the side-chain of H22, the active site metal ion and bases G8-A9-A10-T11 and G25, before using it for molecular replacement with PHASER[30] against the data of the remaining complexes.

Refinement of all structures was performed with the program REFMAC 5.5.0102,[34] interspersed with manual adjustment of the model to the electron density map using the program COOT.[33] Further model improvement and validation was done using the MolProbity web server[35]. Final refinement statistics are shown in Table S2. All structures and corresponding structure factors were deposited in the PDB and the corresponding codes are given in Table S2.

Unambiguous assignment of the identity of the Mn^{2+} ion

Crystals of the Nic0 complex were diffracted at BM14 (ESRF, Grenoble, France). An x-ray absorption scan revealed an absorption peak near the Mn K-edge (E=6.550 keV). Two complete anomalous

datasets were measured at E=6.519 keV and E=6.550 keV, respectively. The anomalous map calculated for the peak dataset (Suppl. Fig. S3) showed a strong peak at the position of the active site metal ion, whereas no discernable contribution was observed for the lower energy near-edge data. Since no other element has such absorption behavior at these wavelengths, the element of the atom bound at the active site is Mn.

Reaction mechanism and movie

The individual x-ray structures were used to model the reaction mechanism. The structures were first ordered on the basis of the state they best represent. Addition of the DNA downstream of nucleotide G26 and the scissile phosphate, and a reorientation of H22 were the only modifications required to be able to fully model the sequence of events, as shown in Fig. 2. The pentacoordinated intermediate was derived from the structure containing the scissile phosphate moiety coordinated to the metal ion (Nic0-PO and Nic0-SPO), with the H22 side-chain moved to the 'in'-position and the phosphate moiety adjusted to a planar pentacoordinated configuration. Morphing of the transition between the intermediate snapshot models was done with Chimera[36] and the still pictures were produced with PYMOL.[37] The movie mpg file was then prepared using the LINUX utility program ffmpeg.

Others

In vivo mobilization of pMV158 and derivatives between pneumococcal strains was done essentially as described[38]; all these assays were performed at least four times under independent conditions.

Bibliography

20. Studier, F.W., Rosenberg, A.H., Dunn, J.J. & Dubendorff, J.W. *Meth. Enzymol.***185**, 60-89 (1990).
21. del Solar, G., Díaz, R. & Espinosa, M. *Mol. Gen. Genet.***206**, 428-435 (1987).

22. de Antonio, C., Farias, M.E., de Lacoba, M.G. & Espinosa, M. *J. Mol. Biol.***335**, 733-743 (2004).
23. Lorenzo-Díaz, F. et al. *Nucl. Acids Res.***39**, 4315-4329 (2011).
24. Guzmán, L. & Espinosa, M. *J. Mol. Biol***266**, 688-702 (1997).
25. Trask, D.K., DiDonato, J.A. & Muller, M.T. *EMBO J***3**, 671-676 (1984).
26. Leslie, A.G.W. & Powell, H.R. Processing Diffraction Data with Mosflm. in *Evolving Methods for Macromolecular Crystallography*, Vol. 245 (eds. Read, R.J. & Susman, J.L.) 41-51 (2007).
27. Evans, P. *Acta Crystallogr. D Biol. Crystallogr.***62**, 72-82 (2006).
28. Uson, I. & Sheldrick, G.M. *Curr. Opin. Struct. Biol.***9**, 643-8 (1999).
29. Schneider, T.R. & Sheldrick, G.M. *Acta Crystallogr. D Biol. Crystallogr.***58**, 1772-9 (2002).
30. McCoy, A.J. et al. *J. Appl. Crystallogr.***40**, 658-674 (2007).
31. Cowtan, K. *Acta Crystallogr. D Biol. Crystallogr.***57**, 1435-44 (2001).
32. Terwilliger, T.C. *Methods Enzymol.***374**, 22-37 (2003).
33. Emsley, P. &Cowtan, K. *Acta Crystallogr. D Biol. Crystallogr.***60**, 2126-32 (2004).
34. Murshudov, G.N., Vagin, A.A. &Dodson, E.J. *Acta Crystallogr. D Biol. Crystallogr.***53**, 240-55 (1997).
35. Davis, I.W. et al. *Nucleic acids research***35**, W375-83 (2007).
36. Pettersen, E.F. et al. *J. Comput. Chem.***25**, 1605-12 (2004).
37. DeLano, W.L. *on World Wide Web http:/www.pymol.org* (2002).
38. Lorenzo-Diaz, F. & Espinosa, M. *J. Bacteriol.***191**, 720-7 (2009).

Supplementary material for

Structures of DNA-MobM relaxase complexes reveal a histidine/metal catalysis for DNA cleavage and ligation

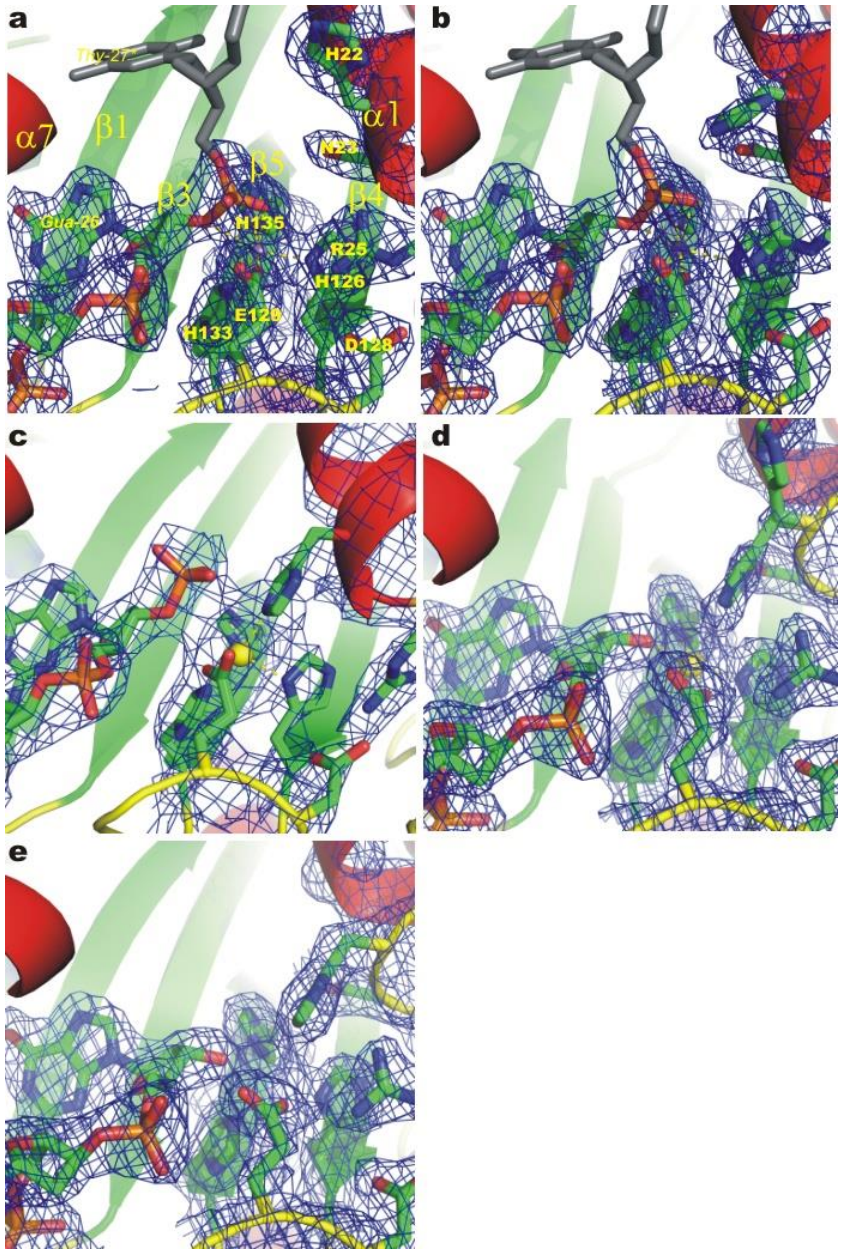
Radoslaw Pluta, D. Roeland Boer, Cris Fernández-López, Silvia Russi, Fabián Lorenzo-Díaz, Rosa Pérez-Luque, Manuel Espinosa, Miquel Coll

Name	Sequence
nic0	5'-ATAAAGTATAGTGTG-3' 3′-TATTTCA-5′
nic0-PO	5'-ATAAAGTATAGTGTGp-3' 3′-TATTTCA-5′
nic0-SPO	5'-ATAAAGTATAGTGTGp^{S}-3' 3′-TATTTCA-5′
nic0+1	5'-ATAAAGTATAGTGTGT-3' 3′-TATTTCA-5′
pMV158 (bases 3571-3597) Numbering used:	5′-ACTTTATgaatATAAAGTATAGTGTGT-3′ | | | | | | | 1 5 7 12 15 20 25

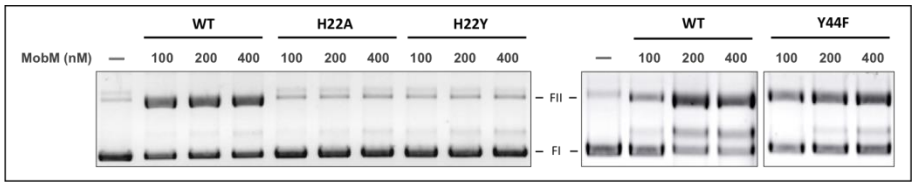
Table S1. Sequence of the oligonucleotides used for crystallizations and the oriT sequence from the pMV158 plasmid (omitted residues from the hairpin loop region are lowercase). The base numbering shown is used for all oligonucleotides.

	Nic0 pH4.6	Nic0 pH5.5	Nic0-PO pH4.6	Nic0-SPO pH6.8	Nic0+1 pH6.5
PDB code	4LVI	4LVJ	4LVK	4LVL	4LVM
			Data collection		
Space group	$P6_122$	$P6_122$	$P6_122$	$P6_122$	$P2_12_12$
Cell dimensions					
a, b, c (Å)	111.1, 111.1, 90.7	112, 112, 90.6	112.7, 112.7, 91.7	112.4, 112.4, 90.6	174.9, 118.3, 204.0
α, β, γ (°)	90, 90, 120	90, 90, 120	90, 90, 120	90, 90, 120	90.0, 90.0, 90.0
Resolution (Å)	48.1-1.90 (2.00-1.90)	41-2.17 (2.29-2.17)	35.6-2.38 (2.51-2.38)	41.1-2.2 (2.32-2.2)	28.4-3.10 (3.18-3.10)
R_{merge}	0.11 (0.65)	0.15 (0.70)	0.11 (0.67)	0.16 (0.63)	0.09 (0.42)
R_{meas}	0.12 (0.78)	0.17 (0.80)	0.11 (0.72)	0.18 (0.72)	0.11 (0.48)
$\langle I / \sigma(I) \rangle$	8.2 (1.4)	7.3 (1.8)	12.2 (2.8)	7.0 (2.4)	9.4 (2.9)
Completeness (%)	99.8 (99.6)	98.2 (100.0)	99.3 (97.8)	100 (100)	98.7 (98.1)
Redundancy	5.0 (3.3)	4.2 (4.2)	8.4 (8.1)	4.1 (4.2)	4.1 (4.2)
			Refinement		
Resolution (Å)	48.1-1.90	35.2-2.17	35.6-2.37	36.8-2.20	28.4-3.1
No. reflections	25154	16928	13543	16824	13500
R_{work} / R_{free}	20.6/24.8	18.6/23.2	20.8/25.6	17.9/22.5	25.5/29.4
No. atoms	2224	2219	2109	2279	4031
Protein	1634	1661	1572	1672	3144
DNA	449	449	453	453	864
Ligands/water	141	109	84	154	23
			***B*-factors**		
Wilson	28.3	28.3	52.9	26.9	76.9
Overall	28.5	26.9	23.3	22.9	58.7
Protein	27.8	27.1	24.4	22.4	59.9
DNA	24.9	26.3	20.2	23.6	57.5
Ligands/water	47.9	27.1	23.2	26.1	37.2
			R.M.S. deviations		
Bond lengths (Å)	0.014	0.021	0.013	0.017	0.008
Bond angles (°)	1.740	2.133	1.633	1.781	1.409
			MolProbity scores		
Overall score	1.6 (91th%)	1.8 (93th%)	1.2 (100^{th}%)	1.5 (99^{th}%)	9.6 (96^{th}%)
All-atom clashscore	3.8 (98th%)	9.7 (90th%)	2.1 (100^{th}%)	4.7 (98^{th}%)	9.6 (96^{th}%)
Bad rotamers (%)	1.8	1.2	0.6	0.57	2.6
Ramachandran outlier (%)	0	0	0	0	0
Ramachandran favoured (%)	96.2	96.3	96.3	96.4	93.4

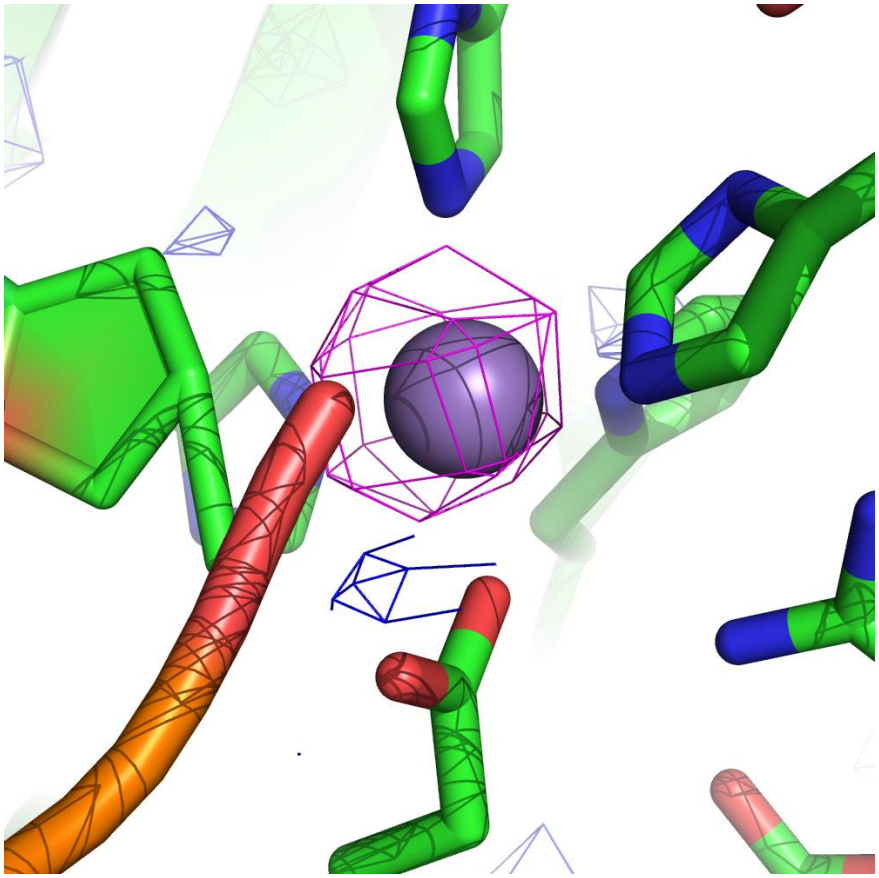
Table S2. Data collection and refinement statistics



Suppl. Fig. S1. The active sites of the five X-ray structures as shown in Fig.2 (main text), here including the $2F_o$-F_c electron density around the relevant residues. Modeled atoms not present in the crystal structures are shown in gray. **a**) Nic0-PO at pH 4.6: H22 is outward. **b**) Nic0-SPO at pH 6.8: H22 moves partially inward. **c**) Nic0+1: H22 is now in the inward position and ready for nucleophilic attack. **d**) Nic0 at pH 4.6, H22 is partially in the inward and partially in the outward position. **e**) The structure of Nic0 at pH 5.5, in which H22 is fully occupied in the inward position.



Suppl. Fig. S2. Separation of supercoiled and relaxed DNA plasmid forms by electrophoresis upon generation of relaxosomes by the wild-type FL protein and H22 (left) and Y44 (right) mutants. FII and FI depict relaxed and supercoiled DNA, respectively.



Suppl. Fig. S3. Anomalous maps from data measured at 6.550 keV, the Mn peak (magenta, 5σ, well above the noise level) and at 6.519 keV, near-edge low energy (blue, contoured at 2σ, near the noise level).

3.2. Structural basis of antibiotic resistance transfer mediated by MobM, a prototype of the major family of relaxases found in *Staphylococcus aureus*

The following manuscript is in preparation for submission:

Structural basis of antibiotic resistance transfer mediated by MobM, a prototype of the major family of relaxases found in *Staphylococcus aureus*

Radoslaw Pluta[1,2,+ §], D. Roeland Boer[1,2+, §], Silvia Russi[1,2, §], Fabián Lorenzo-Díaz[3,§], Cris Fernández-López[3], Rosa Pérez-Luque[1,2], Manuel Espinosa[3], Miquel Coll[1,2,*]

[1]Institute for Research in Biomedicine (IRB Barcelona), Baldiri Reixac 10-12, 08028 Barcelona, Spain
[2]Institute of Molecular Biology of Barcelona (IBMB-CSIC), Baldiri Reixac 10-12, 08028 Barcelona, Spain
[3]Centro de Investigaciones Biológicas (CIB-CSIC), Ramiro de Maeztu, 9, 28040 Madrid, Spain.

[+] These authors contributed equally
* Corresponding author

[§]Present address: RP - International Institute of Molecular and Cell Biology in Warsaw; ul. ks. Trojdena 4, 02-109 Warsaw, Poland; DRB – ALBA Synchrotron, Carretera BP 1413, 08290 Cerdanyola del Vallès, Spain; SR - SLAC National Accelerator Laboratory, 2575 Sand Hill, Menolo Park, CA 94025, USA; FLD - Instituto Universitario de Enfermedades Tropicales y Salud Pública de Canarias, Centro de Investigaciones Biomédicas de Canarias (CIBICAN), Spain

Keywords: bacterial conjugation, Gram-positive bacteria, antibiotic resistance Staphylococcus aureus, MOB_V family of relaxases, Mob_Pre, plasmid pMV158, relaxase-DNA binding, histidine/metal ion nuclease

Abstract

Methicillin-resistant *Staphylococcus aureus* (MRSA) infection ranks among the leading causes of death and *S. aureus* strains evolve towards infecting healthy individuals in non-hospital environments. Conjugative transfer of plasmid DNA is the main route for the antibiotic resistance acquisition. Relaxases initiate the DNA transfer by cleaving DNA through a formation of covalent relaxase-DNA adduct and terminate the transfer in the recipient cells by re-joining ends of the linearized plasmid. MobM protein from the promiscuous antibiotic resistance plasmid pMV158 is a prototype of the Mob_Pre/MOB_V family of relaxases, the major family of relaxases found in *S. aureus.* Previously, we showed that MobM forms a DNA-histidine adduct, unique to MOB_V relaxases, instead of a DNA-tyrosine adduct, thus representing a distinct category of relaxases. Here, we present crystal structures of the MobM-DNA complexes and describe molecular basis for the MobM processing of plasmid DNA. Interestingly, we observe for the first time that the flexibility of two relaxase regions that lock the DNA in the active site can lead to specific DNA-DNA interactions between two relaxase-DNA complexes, which result in the active site inhibition. Together, our data reveal the structural basis for antibiotic resistance transmission by the major *S. aureus* family of relaxases and offer a starting point for structure-based inhibitors design.

Introduction

Mobile genetic elements (MGEs) compose up to 20% of the bacterial genome and constitute the bacterial horizontal gene, which is considered a major driving force in bacterial evolution (de la Cruz and Davies, 2000; Thomas, 2000; Boucher *et al.,* 2003; Wozniak and Waldor, 2010). Acquisition of MGEs by bacteria is mainly achieved via the conjugative DNA transfer of plasmids and integrative and conjugative elements (ICEs), which are involved in transmission of genes that confer threat to human health, such as antibiotic resistance and virulence genes, leading to strains that are both more resistant and more virulent, which is posing a real threat to the public health systems (Baquero, 2004; Burrus and Waldor, 2004; CDC, 2013; Lindsey, 2010;

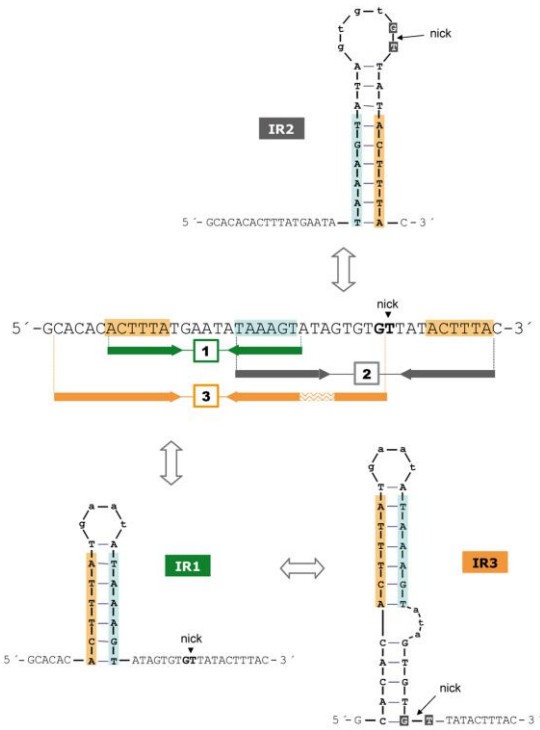
WHO, 2014). Moreover, trans-kingdom conjugative DNA transfer is also well-known (Christie, 2004) and widely used for research and biotechnology as in case of *Agrobacterium*-mediated T-DNA transfer to plant cells (Gelvin, 2003; Lacroix *et al.*, 2008). Recently, conjugative DNA transfer into human cells by the bacterial pathogen *Bartonella henselae* has been demonstrated and described as a potential tool for DNA delivery into human cells (Schröder *et al.*, 2011; Llosa *et al.*, 2012). In general, most efforts up to now have been made towards the understating of DNA transfer of Gram-negative (G-) bacteria (de la Cruz *et al.*, 2010). Much less information is available for Gram-positive (G+) bacteria (Alvarez-Martinez and Christie, 2009; Goessweiner-Mohr *et al.*, 2013), even though G+ drug-resistant infections constitute a major health threat. According to estimates of the Center for Dissease Control and Prevention (CDC), each year in USA more than 2,000,000 people are infected with antibiotic-resistant microorganisms and infections with G+ bacteria of *Staphylococcus*, *Streptococcus* and *Enterococcus* genera account for 20,000 out of 23,000 lethal antibiotic-resistant infections (CDC, 2013). More than a half of them are attributed to methicillin-resistant *Staphylococcus aureus* (MRSA). Furthermore, many more people suffer or die from health conditions resulted from medical complications related to antibiotic-resistant infection.

The key player in conjugative DNA transfer is an endonuclease/ligase termed relaxase, which cleaves supercoiled plasmid DNA in a strand and sequence specific manner, assists with the transfer of the nicked single DNA strand to the recipient cell and finally ensures re-joining of the nicked DNA. (Pansegrau and Lanka, 1996a, 1996b). Conjugation begins with the relaxase recognizing a DNA hairpin within the origin of transfer (*oriT*), upstream the nick site (*nic*). Nucleophilic attack at the *nic* site generates a covalent linkage between the relaxase and the 5'-phosphate DNA end, leaving the free 3'-hydroxyl end to serve as a primer for DNA replication in a process called conjugative rolling-circle replication (RCR) (Gomis-Rüth and Coll, 2006; Chandler *et al.*, 2013). The cut DNA strand (called T strand) is then unwound through RCR and transferred to the recipient cell. In the classical view, transfer of the ssDNA-relaxase complex from the donor to the recipient cell occurs through a type IV secretion

system (T4SS) with the aid of a coupling protein (T4CP) that delivers the protein-DNA adduct from the cytoplasm to the T4SS membrane channel. However, recent findings suggest that the T4SS secretion chamber could be also accessed by relaxosome from the periplasm (Low *et al.,* 2014). Conjugative plasmids encode for all three elements, while smaller mobilizable plasmids lack genes encoding the T4SS/T4CP machinery and hitchhike T4SS/T4CP of the conjugative plasmids. Based on their 300-residue N-terminal sequences, conjugative relaxases were divided into six plasmid/ICE-based families: MOB_F, MOB_Q, MOB_P, MOB_V, MOB_H, MOB_C and further expanded by two more ICE-based families: MOB_T, MOB_B (Francia, 2004; Garcillán-Barcia *et al.,* 2009; Guglielmini *et al.,* 2011). The best studied relaxases belong to the superfamily of HUH endonucleases, which are characterized by the presence of the 'HUH motif' (two His residues used for metal coordination, separated by a hydrophobic residue) and the active site nucleophilic residue that nicks the ssDNA. The latter was assigned as a 'Y motif', containing either one (Y1) or two (Y2) active-site Tyr residues, which delivers the nucleophilic hydroxyl group(s) (Chandler *et al.,* 2013). However, we have shown recently that the MOB_V relaxases use histidine imidazole nitrogen as the nucleophile (Pluta *et al.,* submitted). The classical 'HUH motif' and active site nucleophiles were experimentally determined for MOB_F, MOB_Q, MOB_P, MOB_V families (Garcillán-Barcia *et al.,* 2009, Chandler *et al.,* 2013; Pluta *et al.,* submitted) and the X-ray crystal structures of members of three families are available: MOB_F family TrwC_pR388 (Guasch *et al.,* 2003; Boer *et al.,* 2006), TraI_pF (Datta *et al.,* 2003; Larkin *et al.,* 2005) and TraI_pCU1 (Potts *et al.,* 2010); MOB_Q family MobA_pR1162 (Monzingo *et al.,* 2007) and NES_pLW043 (Edwards *et al.,* 2013) and MOB_V family MobM_pMV158 (Pluta *et al.,* submitted; this work). MobM relaxase of the mobilizable promiscuous resistance plasmid pMV158 (5,5kb) serves as the prototype of the MOB_V family of relaxases (Francia *et al.,* 2004; Garcillán-Barcia *et al.,* 2009). The usage of histidine imidazole nitrogen nucleophile rather than tyrosine hydroxyl one is most striking difference between the MOB_V relaxases and relaxases from other families (Pluta *et al.,* submitted). MOB_V relaxases dominate among the smallest plasmids (<5kb), are frequently found in medium-size plasmids (5-20kb) and are

almost absent in the big plasmids (>60kDa) (Smillie *et al.*, 2010). The full length MobM relaxase is a 494 residue protein which dimerizes in solution through the C-terminal domain (de Antonio *et al.*, 2004). The MobM199 relaxase domain (residues 2-199) used in this study is monomeric in solution, binds plasmid *oriT* DNA and retains nuclease activity on supercoiled plasmid DNA, but not on ssDNA oligonucleotides mimicking the *oriT*, for which residues 200 to 243 are needed for DNA nicking (Lorenzo-Díaz *et al.*, 2011; Fernández-López *et al.*, 2013). The pMV158 *oriT* is relatively complex: it contains three inverted repeats, IR1, IR2 and IR3 that can form DNA hairpins (Fig. 1) (Lorenzo-Díaz *et al.*, 2011). The IR1 hairpin can be viewed as a substructure of the IR3 hairpin, in which a 7 base pairs (bp) stem of the IR1 hairpin is followed by a 3-base bulge (IR3 ATA bulge) located within the IR3 right arm (IR3-R) and terminated by an additional 5 bp stem. Due to sequence overlap, formation of the upstream to the *nic* site hairpin (IR1 or IR3) excludes formation of the downstream IR2 hairpin (Fig. 1). The pMV158 *oriT* is highly conserved among the pMV158 family of plasmids (Grohmann *et al.*, 2003).

To gain a deeper insight into the conjugative gene transfer mediated by MOB_V relaxases we undertook structural characterization of the N-terminal relaxase domain of MobM protein. In our previous paper we focused on the description of mechanism of histidine-metal ion mediated DNA nicking, the first description of a nuclease with such mode of action (Pluta *et al.*, submitted), here, we present four structures



oriT sequence of pMV158; coordinates 3566-3606	3566 3571 3576 3581 3586 3591 3596 3601 3606
dna26 oligo; crystal structures numbering	5′-CACAC**ACTTTATGAATATAAAGTATAGTGTG***TTATACTTTA-3′ **1 6 11 16 21 26**

Figure 1. Inverted repeats (IR) and possible DNA hairpins within the plasmid pMV158 origin of transfer (*oriT*). The three inverted repeats (IR1, IR2 and IR3) are indicated by arrows within a 43 bp region (the nick site is indicated by a vertical arrowhead). Shadowed bases are complementary to each other. Three alternative hairpin structures could be formed, the IR1 hairpin being a sub-structure of the IR3 hairpin, and the IR2 hairpin hindering the possibility of formation of the IR1/3 hairpins (and vice versa). Numbering of the plasmid pMV158 nucleotides within the *oriT* and crystal structures is presented at the bottom of the figure.

of an inhibited state of MobM relaxase domain bound to a 26-base oligonucleotide, which together with five MobM-DNA structures from our previous study, allow us to describe the molecular basis of *oriT* DNA processing by MobM. Moreover, our structural data let us suggest roles of active site conserved residues that assist the nucleophilic His22 in DNA cleavage and hypothesise a relaxase -

plasmid autoinhibitory regulatory mechanism. Finally, we discuss structural similarities and differences of DNA interacting regions of relaxases from different MOB families.

Results

MobM represents the major *S. aureus* family of conjugative relaxases

MobM N-terminal 194-residues are classified by the NCBI Conserved Domains Database (CDD) (Marchler-Bauer *et al.*, 2013) as a domain that belongs to the Mob_Pre family. As for August 2014, the family includes more than 4,000 or 8,000 specific (4.9e-43) or general (1.0e-2) members, that are found almost exclusively in Firmicutes, mainly in *Staphylococcus aureus*, but also in *Streptococcus agalactiae*, *Enterococcus faecalis* and *Enterococcus faecium* (Table 1). Noteworthy, infections with these bacteria are responsible for the majority of lethal Gram-positive drug-resistant infections in USA (CDC, 2013). Mob_Pre relaxases cover 84% of all HUH relaxases found in *S. aureus.* Moreover, selection of the *S. aureus* as the host for MGEs that carry Mob_Pre relaxases seems much stronger than for relaxases from other families, resulting in 69% of specific Mob_Pre family members being reported as found in *S. aureus*, the second highest score being 3,5-fold lower (20% of MobA_MobL relaxases are found in *S.aureus*). In general, specific members of the Mob_Pre family have the highest tendency (81%) among HUH relaxases to be found in the CDC-listed Gram-positive human pathogens than to be found in other bacteria. What is more, due to recent studies focused on the genomic characterisation of drug resistant *S. aureus* isolates, the number of Mob_Pre relaxases found in *S. aureus* is growing massively in recent years (Table 2). These data indicate that MGEs carrying Mob_Pre/MOBv relaxases (and accompanying them virulence/adaptation genes) may constitute a strong beneficial genetic cargo for G+ human bacterial pathogens and particularly for drug-resistant *S. aureus* strains.

Crystal structures MobMN199-DNA complexes

A 26-base oligonucleotide named DNA26 (bases 6-31 of the *oriT*; herein numbered 1-26; Fig.1) comprising the 26 bases upstream the *nic* site (IR1 hairpin + 8 bases downstream) and the MobM N-terminal relaxase domain (residues 2-199) were used for the complex co-crystallization. Four distinct crystal forms were obtained with the MobM199-DNA26 complex. The structures were determined to 2.0-2.5 Å resolution range and the experimental phases were obtained by the SAD method, using the selenomethonine protein derivate data set. Data collection and refinement statistics are shown in the table Final refinement statistics of both structures are shown in table 3. The description is complemented by structural information from our recent work in which MobM199 was crystallized with annealed oligonucleotides, similar to DNA26, that miss however the hairpin tip 4-base loop fragment (Nic0 oligonucleotide) and are extended at the 3' end (Nic0-PO, Nic0-SPO and Nic0+1 oligonucleotides) (Pluta *et al.*, submitted). The overall MobM relaxase domain structure shows that the protein has the relaxase α/β plate fold, reminiscent of the DNA polymerase Klenow fragment and is similar to other relaxases of the rep/mob family (Fig. 2 and 3). It consists of five anti-parallel β-strands (β1, β3, β5, β4, β2) which are flanked on each side by a pair of packed α-helices, α1/α2 and α4/α6 respectively. The structure can be presented as a left hand, where the 'palm' (central β-strands, loop β4-β5 and loop α5-α6 and helices α2, α3, α4, α6 and $3_{10}2$) serves as a platform for ssDNA docking and metal ion coordination by a histidine triad (H126, H133 and H135), a landmark of the conjugative HUH relaxases and a unique to MOB_V relaxases glutamate (E129), the 'thumb' (C-terminal elongated loop α6-α7 and helix α7) wraps on ssDNA and directs it into the active site, the 'little finger' (β-turn loop α1-$3_{10}1$) closes the active site from the top and bears the nucleophilic His22, the 'middle finger' (loop α2-β6) grabs the DNA hairpin minor groove and the

CDD Family (MOB Family)	Threshold, specific and general	*Staphylococcus aureus*	*Streptococcus*			*Enterococcus*		CDC G+ pathogens; columns 3-8	Firmicutes	Bacteria	*S. aureus* /all bacteria	CDC G+ pathogens /all bacteria
			agalactiae	*pneumoniae*	*pyogenes*	*faecalis*	*faecium*					
Mob_Pre (MOB_V)	4.9e-43	2790	202	1	12	177	90	3272	3954	4045	0.69	0.81
	1.0e-2	3211	253	1	12	200	312	3989	6294	8061	0.40	0.49
Relaxase (MOB_P)	5.8e-51	259	0	52	0	20	105	436	791	5975	0.04	0.07
	1.0e-2	305	360	283	105	437	305	1795	5376	17419	0.02	0.10
MobA_MobL (MOB_Q)	4.9e-70	259	15	76	8	32	16	406	1060	1292	0.20	0.31
	1.0e-2	299	15	165	18	39	25	561	2667	7649	0.04	0.07
TrwC (MOB_F)	3.8e-52	0	0	0	0	0	0	0	8	1672	0.00	0.00
	1.0e-2	0	0	0	0	0	0	0	68	6002	0.00	0.00

Table 1. Distribution of HUH relaxases in G+ bacteria that are listed in the CDC Threat Report "Antibiotic resistance threats in the United States, 2013". Based on data from the NCBI Conserved Domain Database, data from August 2014.

Mob_Pre (MOB_V)	Threshold, specific and general	By 31.12.2011	By 31.12.2012	By 31.12.2013	By 31.08.2014
Staphylococcus aureus	4.9e-43	95	153	403	2790
	1.0e-2	165	300	614	3211
Bacteria	4.9e-43	408	584	1599	4045
	1.0e-2	1257	1928	5016	8061
Faction of S. aureus among all bacteria	4.9e-43	0.23	0.26	0.25	0.69
	1.0e-2	0.13	0.16	0.12	0.40

Table 2. Number of Mob_Pre family members found in *S. aureus* and all bacteria in time scales. Based on data from the NCBI Conserved Domain Database.

‘index finger’ (hairpin β6-β7 and helix α5) fixes the DNA hairpin backbone and major groove from the opposite side (Fig. 4).

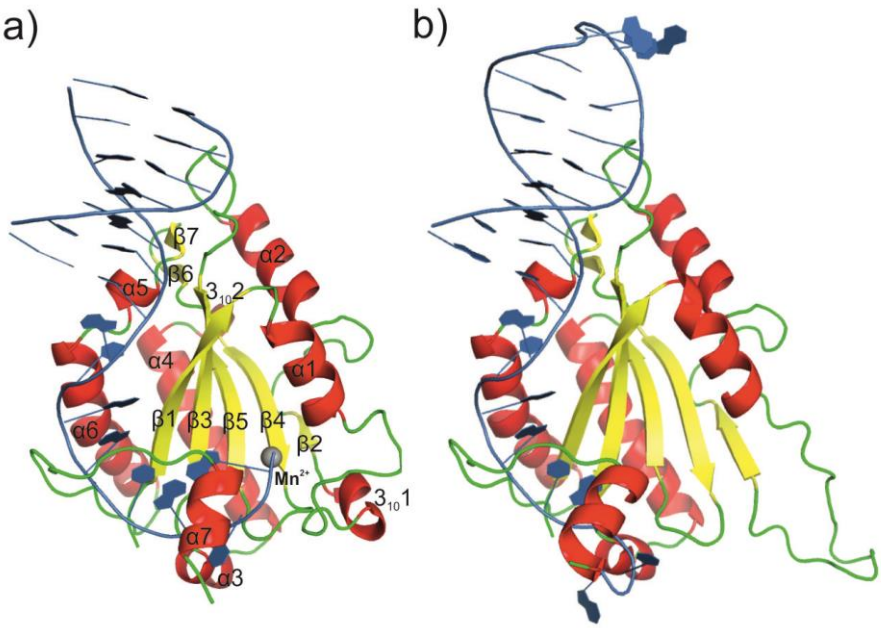


Figure 2. **Cartoon representations of two types of MobM199-DNA complexes. A)** and **b)** represents the Nic0-SPO and DNA26mono3 structures, respectively. The DNA molecule is shown as sticks in light blue, whereas the protein is represented by a cartoon drawing colored according to the secondary structure. The secondary structure elements labelling of the DNA26mono3 structure was omitted for clarity, as it is the same as in the Nic0-SPO structure, except for two 3_{10} helices, which are not present in the DNA26mono3 structure.

The DNA hairpin binds to a long track of positively charged MobM residues stacking out from the protein surface (Fig. 5). This is followed by residues interacting with the ATA budge of the IR3 hairpin which initiate separation of downstream DNA strands and a long C-terminal loop (the thumb) wrapping on separated DNA strand, which is finally locked in the active site by a loop coming from the opposite site (the little finger) (Fig. 4 and Fig. 5). Although extensive interactions between MobM and the hairpin backbone

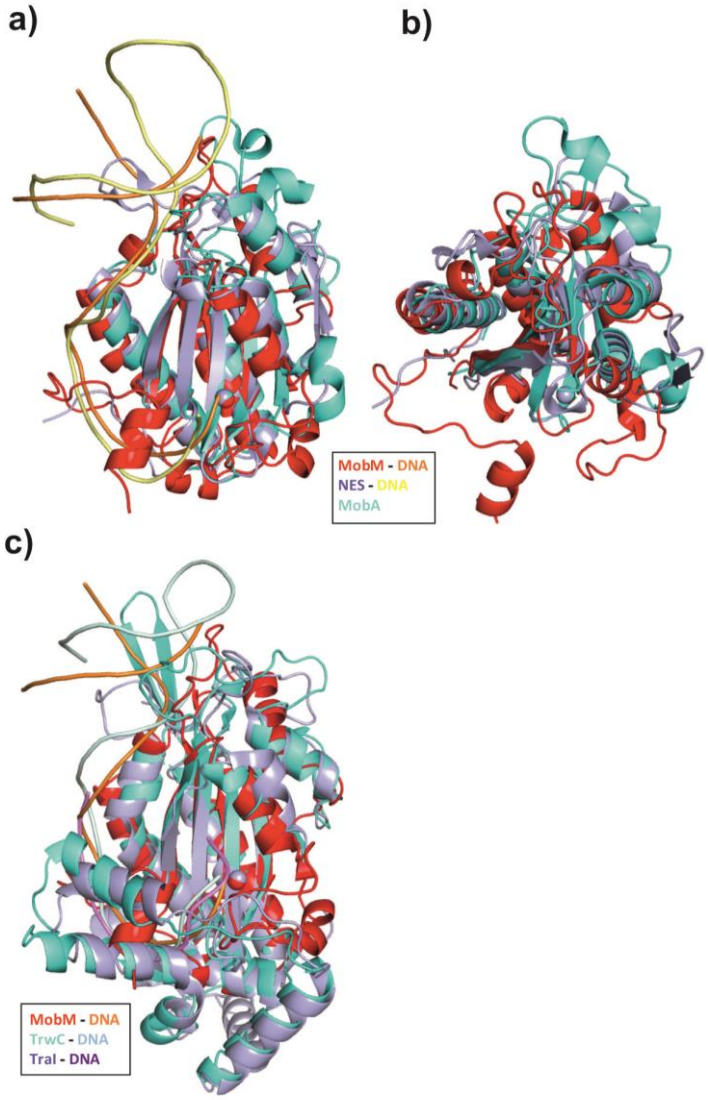


Figure 3. Superposition of MobM and other relaxases. a) Front and **b)** top view (DNA strand omitted for clarity) on superimposition with MOB_Q relaxases MobA (PDB: 2NS6) and NES (PDB: 4HT4). MobM Nic0-SPO structure was used for the analysis (PDB: 4LVL). **c)** front view on superimposition with MOB_F relaxases TraI (PDB: 2A0I) and TrwC (PDB: 2CDM).

Data set [a]	SeMet-MobMN198-26DNA	MobMN198-26DNAortho	MobMN198-26DNAmono2
Data collection			
Beamline	ID4-4	ID23-1	ID23-2
Unit Cell (Å)	a=54.50, b=65.19, c=76.99	a=54.40, b=65.25, c=76.16	*a=43.57, b=52.85, c=56.04, β=95.88°*
Space Group	$P2_12_12_1$	$P2_12_12_1$	$P2_1$
Wavelength (Å)	0.97950	1.0065	0.8726
Resolution range [b]	30.00-2.60 (2.74-2.60)	76.25-2.10 (2.21-2.10)	55.73-2.00 (2.11-2.00)
n° of observed reflections [b]	51817 (7669)	94475 (13927)	70243 (7909)
n° of unique reflections [b]	8878 (1270)	16388 (2338)	17056 (2268)
Completeness [b]	99.8 (100.0)	99.8 (100.0)	98.6 (90.7)
R_{merge} [b,c]	7.7 (27.7)	8.0 (34.8)	9.3 (43.6)
$I/\sigma(I)$ [b]	15.8 (5.9)	15.6 (5.2)	10.8 (2.4)
Multiplicity [b]	5.8 (6.0)	5.8 (6.0)	4.1 (3.5)
Anomalous completeness [b]	99.8 (100.0)	-	-
Anomalous multiplicity [b]	3.2 (3.2)	-	-
Refinement			
Protein atoms [d]		1389	1410
DNA atoms [d]	Not refined	535	535
Water O atoms [d]		98	125
n° of reflections used in refinement [e]		15555	16174
R factor (%) [e]		21.2	20.7
R_{free} (%) [e]	Not refined	26.7	27.8
Rmsd from target values			
Bonds (Å)		0.013	0.005
Angles (°)		1.76	0.92
Mean *B* factor (Å^3)		28.5	26.9

[a] See experimental procedures for explanation. [b] Outermost resolution shell values in parenthesis.
[c] $R_{merge} = [\Sigma_{hkl}\Sigma_i |I_i(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl}\Sigma_i I_i(hkl)]\times 100$. [d] Per asymmetric unit. [e] $R_{factor}=[\Sigma_{hkl} ||F_{obs}| - k|F^{calc}|| / \Sigma_{hkl}|F_{obs}|]\times 100$; R_{free}, same for a test set of reflections not used during refinement.

Table 3. Data collection and refinement statistics.

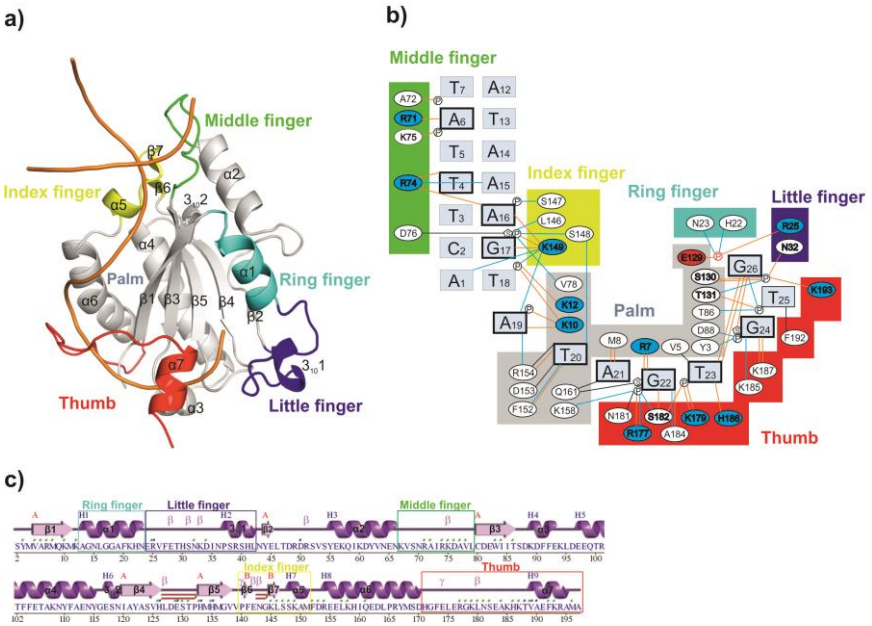


Figure 4. MobM-DNA interactions and secondary structure elements. The different regions (palm and fingers) interacting with the DNA are indicated by a colour scheme. **a)** Cartoon drawing showing the different regions (palm and fingers) interacting with the DNA, which represented by a gold-colored trace of the backbone. **b)** Scheme showing protein-DNA interactions is detail; bonds colouring: brown – hydrogen bonding, blue – water mediated hydrogen bonding, black – nonbonded contact to DNA (< 3.35A); residues that use their side chain to form hydrogen bonds to DNA are indicated in bold and positively and negatively charged residues that do that are additionally indicated by blue and red background, respectively; nucleotides that are involved in hydrogen bonding (to protein or DNA) through their bases are indicated in bold; the scissile phosphate is shown in red. **c)** Secondary structure representation and residues contacts to DNA and to metal. Secondary structures: helices are labelled H1-H9, strands by the sheets they form (A and B). Motifs: β - beta turn, γ - gamma turn, - beta hairpin. Residue contacts: - to DNA; - to metal (panel c was generated by the PDBsum server (Laskowski, 2009) and then modified). The Nic0-SPO structure was used for the analysis, except for the R71-Ade6 and A72-P_7 hydrogen bonds, which were taken from the DNA26mono3 structure, because in the Nic0-SPO structure due to modified DNA loop they were limited to non-bonded steric interactions between R71 and Thy phosphate (P_7); look at Fig.11 for the details of R71 side chain location within the both structures.

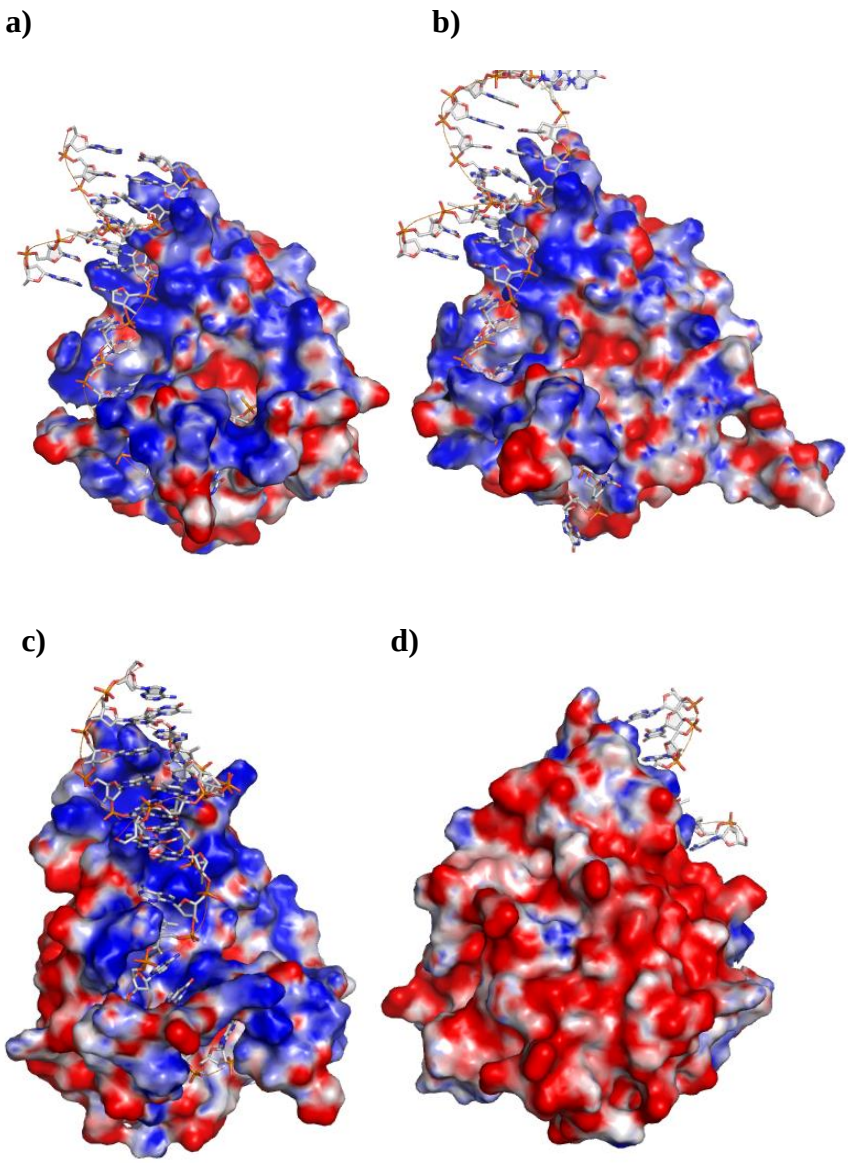


Figure 5. MobM electrostatic surface potential. **a)** front, **c)** side and **d)** back view on the Nic0-SPO structure, respectively; **b)** front view on the DNA26mono3 structure. Positive and negative charge shown in blue and red, respectively, the DNA shown as sticks with phosphates contacted in a cartoon tube presentation.

occur here, only few specific interactions with the bases of the nucleotides occur in that region (Fig. 4 and Fig. 6). Exceptions are

residues R71, R74 of the middle finger and K149 of the index finger. In MobM crystal structures the guanidinium carbon atom of R71 is being shifted up to 3 Å and the guanidinium group forms either a hydrogen bond to Ade6 (Fig. 4) or a water-mediated hydrogen bond to Ade14 or to Thy13 displaying the R71 flexibility of movement and interactions establishment (not shown). R74 forms hydrogen bonds to bases of both sides of dsDNA, i.e. to Thy4 and Ade16, plus a water-mediated one to Ade15. At the other side of the DNA hairpin, the index finger K149 side chain establishes several interactions with DNA bases: a hydrogen bond to Ade16 and Gua17 and a water mediated one to Ade1 (Fig. 4 and Fig. 6).

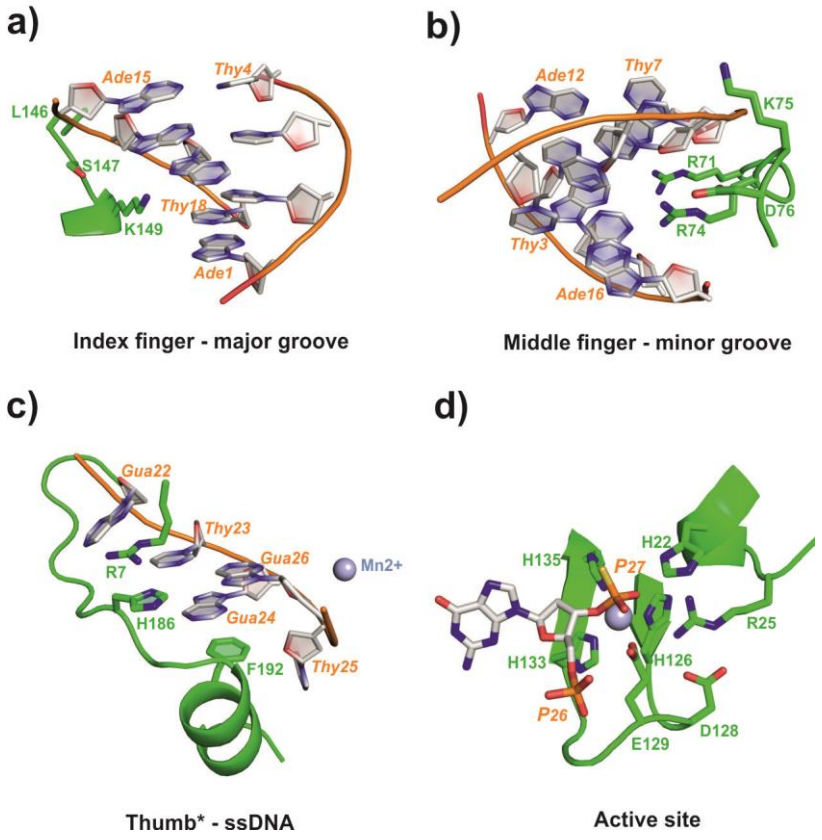


Fig. 6. Cartoon and sticks representation of the MobM-DNA interactions. Panels show **a)** index finger, **b)** middle finger, **c)** thumb (plus R7 from the palm) and **d)** active site (sphere represent she Mn^{2+} ion). The Nic0-SPO structure was used for the analysis.

Moreover, K149 peptide bond forms water-mediated hydrogen bonds to Ade19 and the DNA phosphate backbone. In addition to hydrogen interactions with phosphates of Gua17 and Ade18, the peptide bond of K149 additionally forms a water mediated hydrogen bond to Ade19.

Whereas only two residues (R74 and K149) were found to make profound specific interactions with the IR1 hairpin, a radically different situation arises for the ssDNA bases downstream to the IR1, i.e. Ade19-Thy20-Ade21 of the ATA bulge and Gua22-Thy23-Gua24-Thy25-Gua26 of the right arm of the IR3 stem (IR3-R). With the exception of Thy25, which is solvent-exposed, all bases turn towards the protein surface and are in fact embedded into the protein surface, whereas the phosphate backbone turns outward towards the solvent (Fig. 4 and Fig. 7). K10 and K12 of strand β1 (left edge of the palm) are placed next to phosphate group of Thy18, the last base in double stranded configuration. Additionally, K12 interacts with the peptide carboxyl group of D76, and K10 establishes a hydrogen bond to Ade19 and marks the point at which residues of the 'palm' start to interact with ssDNA bases turning them towards its surface. Other specific interactions of the palm region are double hydrogen bonds of M8 peptide bond to Ade21 and of R7 side chain to Gua22, which is further stabilized by hydrogen contacts with peptide bonds of S182 and A184 the thumb. At bases Gua22 and Thy23, the thumb folds over the DNA backbone, interacting with every phosphate or nucleobase till the Gua26 phosphate (Fig. 4 and Fig. 7). The enclosure of Gua22 phosphate is stabilized by diverse interactions with R177, N181 and S182, whereas the phosphate of Thy23 is steadied by a single hydrogen bond to S182 and a double one to K179. Further downstream, the role of conserved H186 needs to be highlighted, since it is the last specific interaction between a protein side chain and a DNA base: the H186 nitrogen Nδ1 atom forms a hydrogen bond with oxygen O4 of Thy23 while its imidazole ring stacks with the guanidinium group of R7, which itself forms hydrogen bonds with the base of Gua22 (Fig. 4, Fig. 6 and Fig 7). The last protein -

nucleobase specific interactions are found for Gua24 and K187 peptide bond, however a number of hydrogen bond contacts between protein residues and Gua24 and Gua26 phosphate groups are also found in this region (Y3, K185 and N32, S130, T131, K193, respectively). The only protein-DNA link established between solvent exposed Thy25 and MobM is the stacking interaction Thy25-F192, which however seems important due to the fully conserved aromatic character of Mob_Pre family residues at that position (Fig. 8). Noteworthy, strong and specific base-base interactions mediated by Gua26 and Thy23 that form the U-turn of the DNA and place the substrate in the right position for reaction to occur are found just before the scissile phosphate (Fig. 4, Fig. 6 and Fig 7). Furthermore, the scissile phosphate interacts with the active site metal ion and forms hydrogen bonds with R25 and E129 (Fig. 4 and Fig. 6).

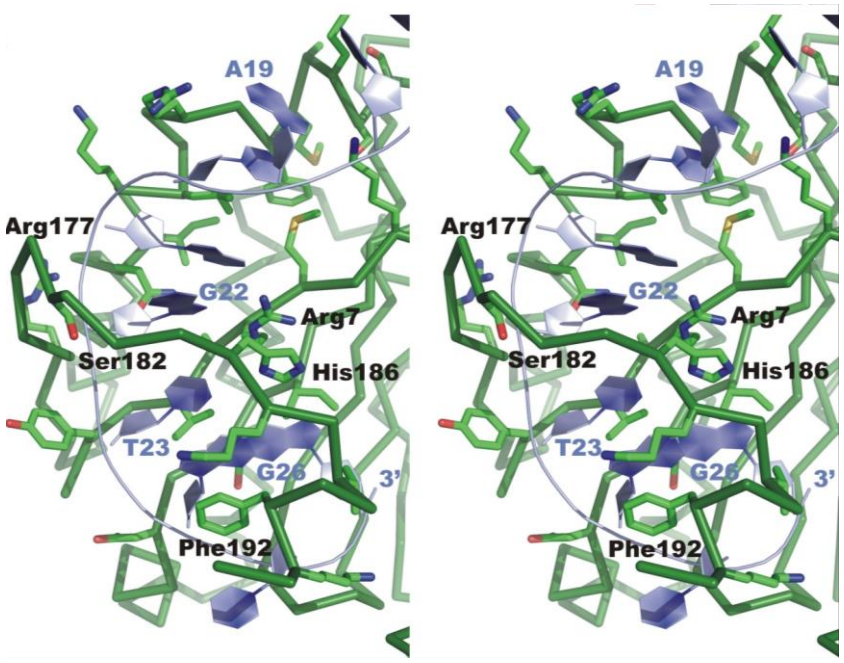


Figure 7. Stereo cartoon-ribbon-sticks view on MobM-ssDNA interactions. The Nic0 structure was used for the analysis.

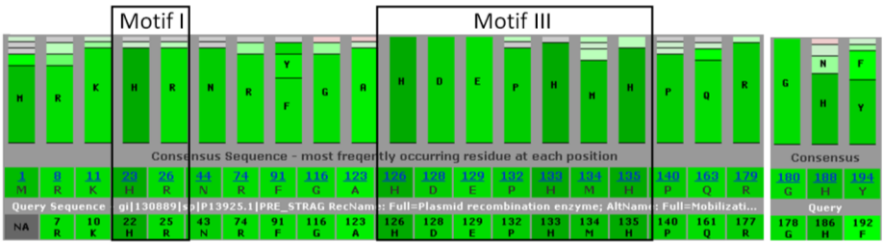


Figure 8. Sequence consensus with indicated Motif I and Motif III of the Mob_Pre/MOB$_V$ family of relaxases. In the upper part, residue frequency bars are presented; the darkness of green colour correlates with the highest scores in the position-specific scoring matrix (PSSM). The figure was created with the NCBI PSSM Viewer.

Active site residues and active site inhibition

In the MobM-Nic structures, intrastrand interactions between Thy23 and Gua26 that from a wobble (non-Watson Crick) Thy:Gua base pair mark a sharp U-turn that positions the scissile phosphate within reach of the catalytic H22 (Fig. 4, Fig. 6 and Fig. 7) (Pluta *et al*, submitted). The U-turn formation was previously observed in MOB$_F$ relaxases (Guasch *et al.*, 2003; Larkin *et al.*, 2005), but not in MOB$_Q$ relaxases (Edwards *et al.*, 2013). The active site center metal ion, Mn2+ interacts with the histidine triad (H126, H133 and H135) and E129 from the protein side and with the 3' oxygen of Gua26 and one of the free oxygen atoms of the scissile phosphate from the DNA side (Fig. 6). Moreover, the metal-interacting scissile phosphate oxygen forms a hydrogen bond to arginine R25, which in turn makes a hydrogen bond with E129 and a salt bridge with D128, by which fixing the β4-β5 loop. Interestingly, a highly conserved N43 is also involved in the β4-β5 loop fixation by extensive hydrogen bond interactions with residues D128 and L127 (not shown). Additionally, R25 establishes a hydrogen bond to the catalytic H22 peptide bond. Very different situation is observed in the MobM-DNA26 structures, in which the two last DNA bases, Thy25 and Gua26, are located outside the active site (Fig. 2 and Fig. 5).

This is due to the placement of the DNA hairpin loop (Gua8'-Ade9'-Ade10'-Thy11') of a neighbouring complex (molecule B) into the active site of MobM (molecule A) (Fig. 9). An extensive array of intermolecular interactions between DNA molecules is observed, i.e. ssDNA bases Thy23-Gua24-Thy25 of one complex interact with hairpin loop bases Gua8'-Ade9'-Ade10' of a neighbouring complex (Fig. 9). Interestingly, Gua8' of the MobM-DNA26 structures is found in the same position as Gua26 of the MobM-Nic0 structures (in both cases the Gua base form a wobble base pair with Thy23), while the phosphate group of Ade9' is found in the position of the metal ion site observed in the Nic0 structures (Fig. 9). Thus several DNA-sequence encoded interactions between the IR1/IR3 hairpin loop and the IR3-R region occur, allowing the tertiary configuration found in the DNA26 structures. The insertion of the DNA hairpin loop region into the active site

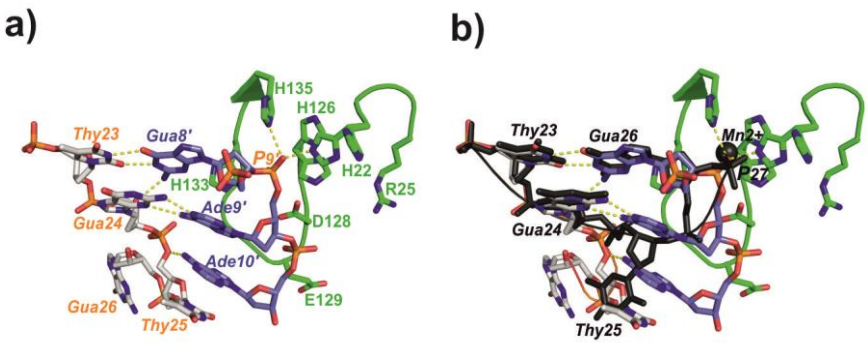


Figure 9. Inhibited active site configuration in the DNA26mono3 structure. **a)** Protein residues and DNA bases from one complex molecule are shown in green and grey. DNA bases from the adjacent complex molecule are shown in violet. Alternative conformations for residues with partial occupancies (H22 and H126) are presented. Hydrogen bonds between the receiving and the incoming molecule are marked by yellow dashed lines. Phosphate that occupies the metal ion position is labelled P9'. **b)** Superposition of DNA bases and an metal ion from the Nic0-SPO structure on the DNA26mono3 structure. The Nic0-SPO structure is shown in black and the DNA26mono3 structure is presented as in panel a).

Figure 10. Superposition of MobM-DNA26 structures. Cartoon presentation of MobM199 from the DNA26mono1 structure (blue) in complex with DNA hairpin from all DNA26 structures (pink) and symmetry aligned adjacent MobM-DNA26 complexes (green and orange) from all DNA26 structures.

causes its disruption: i) metal-coordinating histidines H126 and H133 adopt a side chain conformations ii) the hydrogen bond network established between residues H22, R25, D128 and E129 is interrupted by the displacement of the palm bottom region (β4-β5 loop) and the 'little finger' (Fig. 2, Fig. 9 and Fig. 10). The observed situation happens in all four MobM-DNA26 crystal structures, which were obtained from crystals grown in different conditions and with different space groups (Fig. 10; Materials and Methods; Table 3).

Discussion

We focused on the promiscuous antibiotic resistance plasmid pMV158 encoded MobM relaxase enzyme, a prototype of the Mob_Pre/MOB_V family of relaxases, which is the major family of relaxases found in *S. aureus*. Methicillin-resistant S. aureus (MRSA) kills the highest number of people among antibiotic resistant bacteria and recently became a major concern in worldwide healthcare system (CDC, 2013; WHO, 2014). Recently, we showed that MobM is a one-metal-ion dependent histidine nuclease and outlined the DNA cleavage reaction mechanism by histidine 22 (Pluta *et al*, submitted). Here, we present structural data focused on the specificity of the interaction between MobM and its cognate *oriT* and describe how MobM-induced *oriT* unwinding may take place. The roles of conserved active site residues that assist histidine 22 are also discussed. Furthermore, we observe for the first time that DNA intermolecular interactions within two *oriTs* can lead to formation of multiple relaxase-DNA complexes in which a specific and upstream to the *nic* site DNA fragment from one complex prevents the DNA *nic* site from another complex from placement into the active site. The observed scenario could constitute an auto-regulatory mechanism and provide information for structure-based design of molecules that would block the conjugative DNA transfer. Finally, we show that DNA binding motifs of MobM contain substantial differences in comparison to other solved relaxase structures (Datta *et al.*, 2003; Guasch *et al.*, 2003; Monzingo *et al.*, 2007; Potts *et al.*, 2010; Edwards *et al.*, 2013).

The pMV158-like origins of transfer with the three overlapping inverted repeats (IR1, IR2 and IR3) are unique among known *oriT* sites, although the function of IR2 remains unknown. The IR1 hairpin is a substructure of the long bulged IR3 hairpin, which is computationally predicted to be the most thermodynamically stable among hairpins that can be formed within the $oriT_{pMV158}$ (Lorenzo-Díaz *et al.*, 2011; Zuker 2003; Gruber *et*

al., 2008). The pBBR1 plasmid and the Tn5520 transpozon, two members of the MOB_V family studied in some detail also include IR1/IR3-like sequences with potential to form long imperfect DNA hairpin (Szpirer *et al.*, 2001; Vedantam *et al.*, 2006). It was shown that MobM199 forms high affinity complexes with the *oriT*, IR3 and IR1+8 oligonucleotides, but not with the IR1 hairpin, IR2 hairpin, IR3 hairpin right arm (IR3-R) or IR3 hairpin left arm (IR3-L) oligonucleotides (Lorenzo-Díaz *et al.*, 2011). The minimal binding sequence and the highest binding affinity was assigned to the IR3 element missing the first five bases, that can be also presented as the IR1 hairpin extended by 8 bases (IR1+8) belonging to the IR3-R. Similarly, TrwC, TraI_pF and NES also require for their optimal DNA binding the presence of IR and downstream single-stranded bases (Guasch *et al.*, 2003; Williams and Schildbach 2006; Edwards *et al.*, 2013). Interestingly, the formation of the DNA hairpin is not required for the initial nicking of *oriT*, as shown for plasmids with deletion of the IR left arm, but the re-circularization of the transferred strand by relaxase requires both arms of the IR to fold into a hairpin (Becker and Meyer, 2000; Furuya and Komano, 2000; Gonzalez-Perez *et al.*, 2007). Above suggest that for the DNA nicking reaction to occur a transient relaxase binding is sufficient, but for the reverse, the re-joining of the covalently bound to relaxase DNA 5' end and the free DNA 3' end, the hairpin formation that firmly binds to a relaxase is a must. In the case of plasmids R100 and R388 that encode MOB_F relaxases footprinting experiments showed that the *nic* region melted only upon binding of the relaxase, although the left arm of the IR was not needed for melting (Fukuda and Ohtsubo, 1997; Guasch *et al.*, 2003). Interestingly, the $oriT_{pMV158}$ GC content (24%) is much lower than the GC content of *oriT* sites from other MOB families, and the low GC content can promote the opening of the DNA double helix to allow intra-strand base pairing leading to formation of hairpins (Bikard *et al.*, 2010). Therefore, the $oriT_{pMV158}$ could undergo initial melting and hairpin formation prior to the relaxase binding. We propose that MobM binds with its index and middle

finger the top part of the long bulged IR3 hairpin and subsequently, starting at the ATA bulge, melts the IR3 hairpin base till the formation of the IR1 hairpin and placement of the downstream IR3-R strand into the active site path. It would be valuable to perform footprinting analysis on *oriT*$_{pMV158}$ and MobM-*oriT*$_{pMV158}$ interactions to reveal weather that is the case and how MobM binding shapes the equilibrium of IR1-IR2-IR3 hairpins formation. A careful inspection of *oriT* sites of plasmids encoding relaxases TrwC, TraI_F, NES and MobA shows the potential for formation of long bulged hairpins within *oriT* regions of these plasmids, however (with exception of pLW1043/NES) with less favourable thermodynamics (Fig.11). Again, the pUC1 plasmid for which it was shown that the pUC1 encoded TraI relaxase binds cognate and random oligonucleotides weakly and in a sequence-independent manner (Nash *et al.*, 2009), is an exception and the *oriT*$_{pUC1}$ cannot form long imperfect hairpins (not shown). Intriguingly, a four base (GCAA to CGTT) substitution within the left arm of the *oriT*$_{pF}$ IR, which was intended to disrupt the hairpin formation, resulted in a 3-fold increase in plasmid mobilization efficiency (Williams and Schildbach, 2006). Analysis of the mutated *oriT*$_{pF}$ sequence reveals that the introduced substitution extended the IR1-like hairpin by 1bp and introduced a bulge into the IR1-like hairpin (Fig.11), which suggest that long bulged hairpins could constitute more efficient substrates for *oriT* processing by relaxases.

Structural superimposition of MobM to known structures of relaxases, MOB$_F$ family relaxases (TrwC_pR388, TraI_pF and TraI_pCU1) and MOB$_Q$ family relaxases (MobA_pR1162, NES_pLW043) demonstrate the structural equivalence among MobMN199 and those proteins (2.9 - 4.1Å Cα RMSD), in spite of a very low sequence homology (9 - 13% sequence identity for structurally superimposable residues) (Fig. 3). There are however evident structural differences that distinguish members of these conjugative HUH relaxase families. Notably, the C-terminal thumb directing ssDNA to the active site is in MOB$_F$ relaxases formed by a much longer (and mainly α-helical) protein fragment (TrwC

residues 210-265 vs. MobM residues 175-196) (Fig. 3). In case of MOB_Q relaxases this element lacks structural information, since the MobA structure is based on a shorter protein construct (MobA_1-184) and in the NES structure (NES_2-195) the C-terminus of the protein (NES_1-220) used for crystallization is untraceable. The second evident difference is that MobM (and NES) lack an extensive β-hairpin structure that interacts with the tip of the DNA hairpin in TrwC, and which in MobM would be located between the helix α4 and the strand β4 (TrwC_Q120-I145 to MobM N114-I120). As a result, the tip region of the DNA hairpin is not covered by MobM and protrudes from the complex more prominently than for TrwC.

There are two regions of highly specific interactions found in MobM199-DNA structures. The first one, related to the DNA hairpin binding is based on interactions of the middle finger R71 and R74 residues to bases in the minor grove and of the index finger K149 to bases in the major grove. The second region is formed between the palm/thumb interface and the ssDNA bases downstream to the IR1 hairpin (Fig. 4). Interestingly, R74 is a part of a common β-turn RxD/N motif that relaxases use to penetrate the DNA hairpin minor groove in the relaxase-DNA hairpin crystal structures (MobM_74-76: RKD; NES_78-80: RKN; TrwC_75-77: RQD) (Fig. 12). Superposition to TraI_F and MobA structures that were solved without bound DNA hairpin also reveals presence of such motif (TraI_F_68-70: RMD; MobA_66-68: RAN). The TraI_pCU1 relaxase, which is an exception in that matter and lacks the RxD/N motif, was shown to bind DNA weakly in a sequence-independent manner and presumably need aid of an additional proteins to initiate sequence-specific DNA binding (Nash *et al.*, 2009). The first and the third residue of the RxD/N motif enter the minor groove, while the n+1 residue is placed between phosphates of the DNA backbone (Fig. 12). In addition to R74, K75 and D76 residues of the RxD/N motif, MobM engages R71, and TrwC R68 into their middle finger set of interactions with the minor groove. Interestingly, in the MOB_{Q1} MobA relaxase structure an arginine

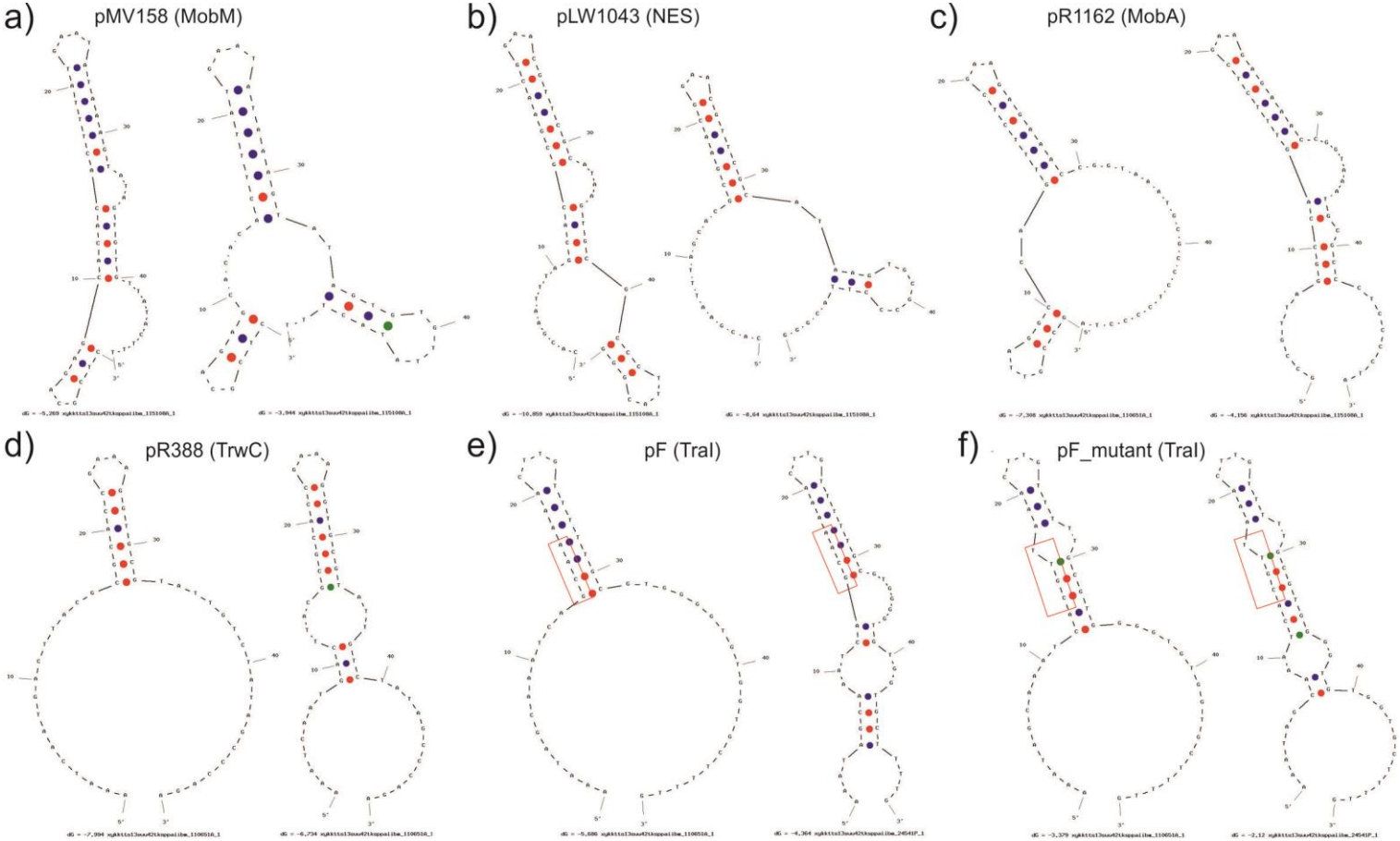

a)
pMV158 (MobM)
b)
pLW1043 (NES)
c)
pR1162 (MobA)
d)
pR388 (TrwC)
e)
pF (TraI)
f)
pF_mutant (TraI)

Figure. 11. Comparison of IR1 and IR3 like structures within *oriT* regions of different plasmids. A) pMV158, **b)** pLW1043, **c)** pR1162, **d)** pR288, **e)** and **f)** pF wt and mutant, respectively, with highlighted mutation (GCAA to CGTT) resulting in alternative bulged structure. For each plasmid a 50-base sequence that spans 40 bases upstream and 10 bases downstream from the nic site (position 40) was taken for the analysis. Note that in each panel the first structure from the right is the structure with preferable formation in terms of thermodynamic (below each structure the free energy is shown). Structures were predicted using the Integrated DNA Technologies' Oligo Analyzer, which is based on the UNAFold software (Markham and Zuker, 2008). The used parameters were as following: 37^{o}C, 200mM Na, 1mM Mg;
https://eu.idtdna.com/analyzer/Applications/OligoAnalyzer

(R143) from a distant sequence region superimposes with MobM R71 (Fig 12). At the opposite face of the hairpin, where MobM locates its index finger, interactions unique to each of the relaxases are found. MobM (MOB_{V1} subfamily) enters the major groove using just one residue (K149) from a short 4-residue helix α5, which defines the minimal major groove interacting element among all discussed relaxases (Fig. 12). NES relaxase (MOB_{Qu} subfamily) places here K150, R151, N154 and Y156 from an 8-residue β8-β9 hairpin, while MobA relaxase (MOB_{Q1} subfamily) expands interactions in this area even further: in addition to a NES-like element (MobA_149-156: PEKGGAQK vs. NES_149-156: PKRKGNDY) and K161 that structurally superimposes on the MobM K149, an 11-residue insertion emerges here for MobA (MobA_131-149: **D**GIERPAAQWFKRYNGKT**P** vs. NES_142-149: **D**KNNEFE**P**), of which MobA R143 structurally superimposes on R71 of the MobM middle finger, as mentioned above. Therefore, it could be that MobA pools together the dsDNA interacting elements that are found separately in MobM and NES relaxases. For $MOB_{F1.1}$ relaxases TrwC and TraI_pCU1 and $MOB_{F1.2}$ relaxase TraI_F, an equivalent to MobM K149 is also found (K181, K180 and K179, respectively). Additionally, in TraI_F relaxase a 9-residue insertion is observed at this area (TraI_F_175-187: **L**SSDKVGKTGFI**E** vs. TrwC_180-183:

LKN**D**), however the DNA-TraI_F complex crystal structure does not include dsDNA hairpin element, therefore the interactions between this 9-residue insertion and dsDNA can be only assumed. Moreover, as exemplified by the TrwC-DNA hairpin structure, MOB_F relaxases have additional β-hairpin structures for interactions with the DNA hairpin major groove, in which the main roles are played (in TrwC) by R128 and K130. Further down, interaction between the left edge of the palm and the IR3-R ATA bulge (Ade19-Thy20-Ade21) are probably important for unwinding of the IR3 hairpin stem base, which would initially involve binding of the bulge, followed by establishment of specific interactions between MobM and downstream bases, that would lead to the strand separation. In MobM-DNA structures, a substantial kink is observed in the DNA backbone between Thy18-Ade19 and Thy20-Ade21 nucleotides (Fig. 7), which disrupts the stacking of respective bases. Further down, the enclosure of Gua22 and downstream bases by the thumb locks the separated DNA strand into a path towards the active site (Fig. 6 and Fig. 7). In contrast, in other relaxase structures, the locking system is based on nucleobases embracement within protein clefts: bases in positions -3 and -5 to the *nic* site for MOB_F relaxases (TrwC: **T**G**T**CT/A, TraI_pF: **G**G**T**GT/G) and in position -3 for MOB_Q relaxase (NES: GT**G**CG/C) are deeply embedded in protein clefts and may provide a mechanism for discrimination of similar but not the cognate *oriT* (Guasch *et al*., 2003; Edwards *et al*., 2013; Carballeira *et al.*, 2014). Curiously, it was shown for NES and TraI_pF that weakening the formation of a stable relaxase-DNA complex (by use of relaxase or DNA mutants) results in the nicking-joining equilibrium being shifted towards the cleavage reaction (Edwards *et al.,* 2013) and ultimately in the plasmid mobilization failure (Williams and Schildbach 2006; Edwards *et al.,* 2013). For MobM the conformational changes induced by placement of the thumb α6-α7 loop on the separated DNA strand could involve relocation of the C-terminal domain with respect to the DNA hairpin-binding part of the protein. Indeed, it was observed that upon formation of the functional relaxosome (i.e.

metal ion, relaxase and substrate DNA), residues 200-243 of the MobM linker region undergo substantial conformational changes (Fernández-López *et al.*, 2013). Such movement could serve as a sensory mechanism of IR3 hairpin recognition, allowing appropriate response of the C-terminal domain in further steps of MobM-mediated DNA transfer. The loop-lock also ensures the formation of the appropriate geometry for subsequent nicking activity. Further downstream, the importance of Gua26 (Gua8' in DNA26 structures) is highlighted, since this base is bound in a very similar way in both types of MobM-DNA structures (Fig. 9), indicating that this region within the MobM protein constitutes a high affinity binding site for guanidine. Taking into account that Gua26 is a part of the non-canonical Thy-Gua base pair found at the U-turn just before the *nic* site in other structures of relaxases (TrwC and TraI_pF), this specificity would for the larger part, together with the binding of the bases proceeding Gua26, explain the specific nicking/joining activity at the *nic* site within the plasmid *oriT*. Indeed, elimination of the Gua-Thy base pair formation (Gua to Cyt substitution) in plasmid F dramatically reduced plasmid mobilization (when introduced into plasmid F *oriT*) and significantly reduced binding and prevented DNA nicking when introduced into an oligonucleotide (Stern and Schildbach, 2001; Williams and Schildbach, 2006). What is more, a systematic mutagenesis using degenerate oligonucleotide libraries revealed that the U-turn formation is a crucial element of *oriT* processing by TrwC relaxase and found that the *oriT* sequence can accept mutation within the U-turn bases (Gua:Thy to Cyt:Gua) as long as they can form the turn with the complex with TrwC (Carballeira *et al.*, 2014).

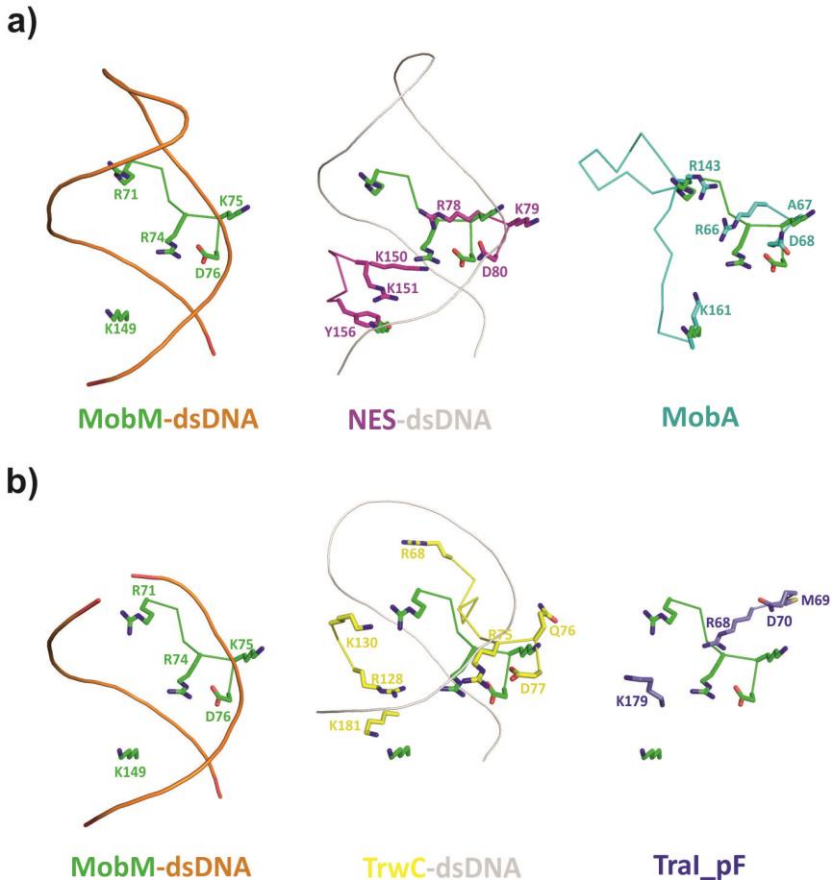


Figure 12. Comparison of DNA hairpin interacting elements of MobM and other relaxases. a) MobM-DNA26mono3 structure and structures of MOB_Q relaxases. **b)** MobM-Nic0-SPO structure and structures of MOB_F relaxases.

Our work reveals roles of conserved residues belonging to the conserved motifs of MOB_V relaxases, Motif I and Motif III (Fig. 8) (Garcillán-Barcia *et al.*, 2009). In addition to the catalytic H22, Motif I includes R25 that forms a slat bridge with aspartate D128 and a hydrogen bond with E129 of Motif III. Moreover, R25 makes hydrogen bond to the scissile phosphate and its guanidinium group superimposes to the TraI_pF K265 side chain nitrogen atom, which in the TraI-DNA structure interacts with the scissile phosphate (Fig.

13). Interestingly, a K265A mutation was shown to cause 20-fold decrease in ssDNA oligonucleotide binding affinity, although the mutant enzymatic activity was not tested (Larkin et al 2005). In type II topoisomerases, which do not belong to the HUH endonucleases, but their catalytic mechanism is related to that of HUH nucleases, the conserved arginine R781 contacts and stabilizes the covalent phosphor-tyrosine moiety as shown in the crystal structure of type II topoisomerase in complex with DNA (Schmidt *et al.,* 2010). R781 mutations to other than lysine residues dramatically reduces DNA cleavage and relaxation activity of this enzyme (Liu and Wang, 1998; Okada *et al.,* 2000). We propose that in MobM R25 fixes the scissile phosphodiester bond for the incoming nucleophilic attack of histidine H22 and support the penta-covalent transition state stabilization and perhaps aids in the phosphohistidine adduct stabilization. Furthermore, our study confirms and expands previous analysis that suggested the superposition of an equivalent metal ion situated on the leaving group site ('B site') in one-metal-ion and two-metal-ion catalysis and the occupation of the second metal-ion site ('A site') by a positively charged/polar protein moiety in one-metal-ion nucleases (Yang, 2008; 2011). In MobM, the 'A site' metal ion being functionally replaced by MobM R25 and TraI K265 (Fig. 13). Intriguingly, members of the one-metal-ion ββα-Me superfamily do not show a positively charged side chain in the metal ion 'A site' (Yang 2008), however a structurally conserved position of an amide bond is present here instead (Fig. 13). Moreover, negatively charged D128 and E129 of the Motif III, which are not observed in other know relaxase structures, contribute to the active site chemistry by assisting in the metal ion ligation and fixation/stabilization of the R25 and of the scissile phosphate. A mutational analysis of the plasmid pBHR1 MOB_V family Mob relaxase shown that the glutamate and aspartate residues of the Motif III are essential for plasmid mobilization and conjugal transfer (Szpirer *et al.* 2001). Regarding the inhibited active site arrangement in MobM199-DNA26 structures, the sequence-encoded intermolecular interactions suggest that the binding mode

observed in the DNA26 structures may reflect bona fide DNA-DNA and DNA-protein binding that could play a role in MobM activity regulation (although one cannot rule out that the observed situation is a crystallographic artefact). Such molecular active site occupation could set MobM catalytic activity in a resting state and represent a plasmid/relaxase copy number-dependent inhibition of MobM mediated activity. Further experiments will be required to determine the relevance of the loop insertion in the active site of the DNA26 structures.

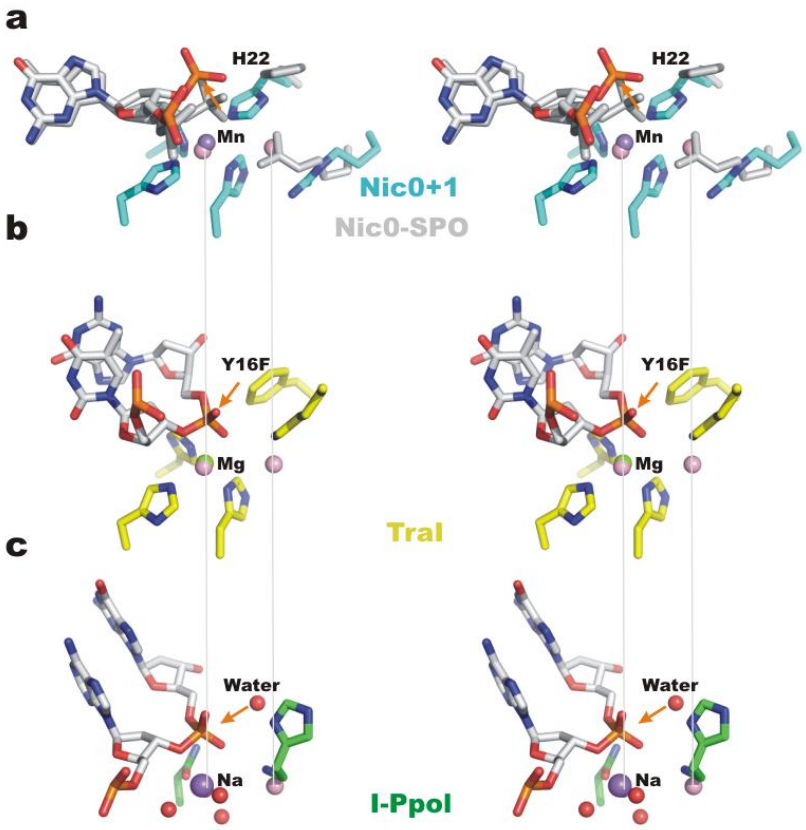


Figure 13. Comparison of the active sites of one metal ion nucleases. Histidine relaxase MobM and tyrosine relaxase TraI from the HUH endonuclease superfamily and homing endonuclease I-PpoI from the ββα-Me superfamily are

shown. The A site and B site (see text) metal ions of two-metal dependent *Bacillus halodurans* RNAse H (PDB 1zbl) are shown as pink spheres. **a)** Active sites of MobM Nic0+1 and Nic0-SPO (shown in gray). **b)** Active site of plasmid F relaxase TraI (PDB 2a0i). **c)** Active site of the homing endonuclease I-Ppol (PDB 1cz0). The histidine side chain that activates the nucleophilic water molecule and the histidine amide bond that superimposes with the metal ion A site are shown.

An interesting difference between tyrosine and histidine relaxases is that the tyrosine relaxases are commonly found in both Gram-positive and Gram-negative bacteria, while the histidine relaxases are almost exclusively found in Gram-positive Firmicutes phylum (Supplementary material; digital version). Furthermore, a study focused on plasmids analysis shown that MOB_V relaxases are evolutionary selected by small mobilizable plasmids of 5-10 kb length (Smillie *et al.* 2010). Such selection could be related to the high energy of P-N bond between the scissile phosphate and the catalytic histidine, which makes it readily transferable to other groups (Kee and Muir, 2012; Robinson and Stock, 1999). In such scenario, the use of histidine could facilitate the DNA closing reaction, but on the other hand would also limit the DNA cargo to short strands, since the P-N bond would be prone to breakage if the DNA transfer takes too much time. MOBv/Mob_Pre relaxases are also found in long plasmids as in the case of closely related plasmids pSK41 (46.4 kb) and pGO1 plasmid (54 kb), the latter being the prototype for class III staphylococcal plasmids (Caryl and O'Neill, 2009). However, in these multi-drug resistance/multi-relaxase plasmids, the Mob_Pre relaxase gene is located (together with aminoglycoside and bleomycin resistance genes, respectively) within a transposable 5 kb cassettes-like module flanked by the IS257-mediated recombination elements. Another characteristic chemical feature of the P-N bond is the acidic liability. The existence of MOB_V relaxases almost exclusively in G+ bacteria could be related to the fact that G- bacteria have periplasm, and this bacterial compartment has no pH homeostasis maintenance, i.e. pH of the periplasm is the same as pH of the environment that

surrounds bacteria (Slonczewski *et al.*, 2009). Therefore, for G-bacteria, conjugative DNA transfer mediated by a histidine relaxase would fail in an acidic environment, due to P-N bond exposure to low pH and breakage during the transfer of a relaxase-DNA covalent complex through the periplasm.

In summary, work presented in this paper provides the first structural analysis of protein-DNA interactions of a member of the MOB_V/Mob_Pre family of relaxases, which is the major family of relaxases found in *Staphylococcus aureus* isolates.

Materials and Methods

Protein expression and purification

The N-terminal relaxase domain of MobM (MobMN198, comprising residues 1 to 198) was expressed and purified according to published procedures (Guzman & Espinosa, 1997). Overexpression in *E. coli* strain BL21(DE3) strain was induced with 1 mM IPTG and after 30 minutes of incubation, 200 μg/ml of rifampicilin was added, continuing the incubation in the same conditions for 90 minutes. Cells were harvested, lysed with a French press, and proteins pelleted with 70 % ammonium sulphate. After dialysis with 0.3 M NaCl, the protein was loaded onto an Affi-Gel heparin-agarose column (Biorad) equilibrated in the same buffer. MobM was eluted with a 0.3-0.8 M NaCl linear gradient and protein containing fractions, as judged from SDS-PAGE, were pooled and further purified using a HiLoad Superdex 200 gel filtration chromatography and a HPLC (ÅKTA, Amersham) system. The protein fractions were collected and further analyzed with SDS-PAGE for purity. The protein was concentrated to 5 mg/ml using 3 kDa molecular weight cut-off centrifugal units Macrosep (Pall). The SeMet MobMN198 was produced according to Budisa *et al.* (1995) with minor changes. A pre-culture of *E. coli* strain B834 (DE3), which is a methionine auxotroph, was prepared in rich minimal medium. Then, the cells were harvested, re-suspended in the same

medium without methionine, and grown during 30' to exhaust intracellular methionine. This culture was diluted in rich minimal medium supplemented with SeMet at a final concentration of 50 ug/ml. Overexpression and purification of the SeMet N-terminal MobM were carried out as before.

Crystallization and structure determination

Details of obtaining structures of MobM-Nic0 complexes are given elsewhere (Pluta *et al*, submitted). In order to prepare the MobM-DNA26 complex, solutions of MobMN199 in 500 mM NaCl, 20 mM Tris pH 7.6, 1% glycerol, 1 mM EDTA and 1mM DTT, and the annealed DNA oligonucleotide, encompassing the *nic* sequence 26 nucleotides upstream from the cleavage site (5'-ACTTTATGAATATAAAGTATAGTGTG-3'), dissolved in the same buffer and salt, were mixed in 1:1.2 ratio (protein/DNA). All oligonucleotides were obtained from Biomers (Ulm, Germany). The resulting complexes were purified by size-exclusion chromatography and concentrated to 5 mg/ml. Crystals of the SeMet derivative of MobM in complex with DNA26 were obtained by vapour diffusion, at 4ºC, from a sitting drop crystallization setup using a crystallization buffer containing 60% MPD, 0.2 M $MgCl_2$ and 0.1 M sodium acetate pH 4.6. Four different crystalline forms of the native MobM-DNA26 complex were obtained by vapour diffusion, at 4ºC, from a sitting drop crystallization setup using crystallization buffers that contained 0.1 M sodium acetate pH 4.6 and additionally: (i) 10% (w/v) PEG1500, 3000 or 6000 (DNA26mono1); (ii) 10% isopropanol (DNA26mono2); (iii) 8% PEG1000, PEG 4000 or PEG8000 (DNA26mono3); (iv) 60% MPD, 0.2 M $MgCl_2$ (DNA26orto). Crystals appeared, in all cases, after 30-35 days and continued growing to a maximum size after 40 days.

The details of the data collections, carried on at beamlines ID14-4, ID23-1 and ID23-2 of the European Synchrotron Radiation Facility (ESRF), are summarized in Table 1. The experimental phases were obtained by the SAD method, using the orthorhombic **SeMet-DNA26-ortho** data set, collected at the Se absorption edge

(0.9795Å). Seven Se sites were located in the asymmetric unit with the program SHELXD (Uson and Sheldrick, 1999; Schneider and Sheldrick, 2002) and protein phases were calculated with PHASER (Read, 2001), improving the initial electron density map by density modification with the program PIRATE (Cowtan, 2002). 30% of the polypeptide chain was automatically traced with the program RESOLVE (Terwilliger, 2004) and completed manually with the program COOT (Emsley and Cowtan, 2004) to a final model that included residues S2-L16 and N43-A195. The refinement of the structure was performed on the 2.1 Å resolution native DNA26-ortho dataset (Table 3) with the program REFMAC5.5.0102 (Murshudov et al., 1997), interspersed with the manual adjustment of the model to the electron density map using the program COOT. In later stages of refinement, model validation using the MolProbity web server was used to improve the model where appropriate. The four other crystal forms were solved by molecular replacement with the program PHASER (Read, 2001) using the orthorhombic crystal structure as model. The refinement of these structures was performed using the procedures described before. TLS rigid body groups were obtained from the TLMSD web server (Painter and Merritt, 2006). A complete model (residues S2-D198) could be built for the DNA26-mono3. Final refinement statistics of structures are shown in Table 3.

Miscellaneous

All structural representations in figures were generated with Pymol. Superpositions were based on secondary structure element algorithm matching implemented in Coot.

Protein data bank accession codes

Atomic coordinates have been deposited with the RCSB PDB under accession codes 1UUU, 1VVV, 1XXX, 1YYY and 1ZZZ for the 26DNA-mono1, 26DNA-mono2, 26DNA-mono3 and 26DNA-ortho structures, respectively.

References

Alvarez-Martinez CE, Christie PJ (2009) Biological diversity of prokaryotic type IV secretion systems. Microbiol Mol Biol Rev 73: 775-808.

Baquero F (2004) From pieces to patterns: evolutionary engineering in bacterial pathogens. Nature Rev Microbiol 2: 510-518.

Bikard D, Loot C, Baharoglu Z, Mazel D (2010) Folded DNA in action: hairpin formation and biological functions in prokaryotes. Microbiol Mol Biol Rev. 2010 74:570-88.

Becker EC, Meyer RJ (2000) Recognition of *oriT* for DNA processing at termination of a round of conjugal transfer. J Mol Biol 300: 1067-1077.

Boer R, Russi S, Guasch A, Lucas M, Blanco AG *et al.,* (2006) Unveiling the molecular mechanism of a conjugative relaxase: The structure of TrwC complexed with a 27-mer DNA comprising the recognition hairpin and the cleavage site. J Mol Biol 358(3): 857-869.

Budisa N, Steipe B, Demange P, Eckerskorn C, Kellermann J *et al.,* (1995) High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in *Escherichia coli*. Eur J Biochem 230: 788-796.

Boucher Y, Douady CJ, Papke RT, Walsh DA, Boudreau ME *et al.,* (2003) Lateral gene transfer and the origins of prokaryotic groups. Annual review of genetics 37: 283-328.

Burrus V, Waldor MK (2004) Shaping bacterial genomes with integrative and conjugative elements. Res Microbiol 155(5): 376-386.

Carballeira JD, González-Pérez B, Moncalián G and de la Cruz F (2014) A high security double lock and key mechanism in HUH relaxases controls oriT-processing for plasmid conjugation. Nucleic Acids Res. pii: gku741v1-gku741

Caryl JA, O'Neill AJ (2009) Complete nucleotide sequence of pGO1, the prototype conjugative plasmid from the Staphylococci. Plasmid 2009: 35-38.

CDC (2013), Antibiotic resistance threats in the United States, 2013, http://www.cdc.gov/drugresistance/threat-report-2013/

Chandler M, de la Cruz F, Dyda F, Hickman AB, Moncalian G, Ton-Hoang B (2013) Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. Nat Rev Microbiol 11: 525-538.

Christie PJ (2004) Type IV secretion: the Agrobacterium VirB/D4 and related conjugation systems. Biochim Biophys Acta 1694(1-3): 219-234.

Cowtan K (2002) Generic representation and evaluation of properties as a function of position in reciprocal space. J Appl Cryst 35: 655-663.

Datta S, Larkin C, Schildbach JF (2003) Structural insights into single-stranded DNA binding and cleavage by F factor TraI. Structure 11: 1369-1379.

de Antonio C, Farias ME, de Lacoba MG, Espinosa M (2004) Features of the plasmid pMV158-encoded MobM, a protein involved in its mobilization. J Mol Biol 335(3): 733-743.

de la Cruz F, Davies J (2000) Horizontal gene transfer and the origin of species: lessons from bacteria. Trends Microbiol 128: 128-133.

de la Cruz F, Frost LS, Meyer RJ, Zechner EL (2010) Conjugative DNA metabolism in Gram-negative bacteria. FEMS Microbiology Reviews 34(1): 18-40.

del Solar G, Díaz R, Espinosa M (1987) Replication of the streptococcal plasmid pMV158 and derivatives in cell-free extracts of *Escherichia coli*. Mol Gen Genet 206: 428-435.

Dostál L, Shao S, Schildbach JF (2011) Tracking F plasmid TraI relaxase processing reactions provides insight into F plasmid transfer. Nucleic Acids Res 39: 2658–2670

Edwards JS, Betts L, Frazier ML, Pollet RM, Kwong SM et. al (2013). Molecular basis of antibiotic multiresistance transfer in Staphylococcus aureus. Proc Natl Acad Sci U S A. 110: 2804-9.

Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60(Pt 12 Pt 1): 2126-2132.

Farías ME, Espinosa M (2000) Conjugal transfer of plasmid pMV158: uncoupling of the pMV158 origin of transfer from the mobilization gene *mobM*, and modulation of pMV158 transfer in *Escherichia coli* mediated by IncP plasmids. Microbiology 146: 2259-2265.

Fernández-López C, Pluta R, Pérez-Luque R, Rodríguez-González L, Espinosa M *et al.*, (2013) Functional properties and structural requirements of the plasmid pMV158-encoded MobM relaxase domain. J Bacteriol 195: 3000-3008.

Francia MV, Varsaki, A., Garcillán-Barcia, M. P., Latorre, A., Drainas, C., de la Cruz, F. (2004) A classification scheme for mobilization regions of bacterial plasmids. FEMS Microbiology Reviews 28: 79-100.

Francia MV, Clewell DB, de la Cruz F, Moncalián G (2013) Catalytic domain of plasmid pAD1 relaxase TraX defines a group of relaxases related to restriction endonucleases. Proc Natl Acad Sci U S A. 110:13606-11.

Fukuda Ha, Ohtsubo E (1997) Roles of Tra I protein with activities of cleaving and rejoining the single-stranded DNA in both initiation and termination of conjugal DNA transfer. Genes to Cell 2: 735-751.

Furuya N, Komano T (2000) Initiation and termination of DNA transfer during conjugation of IncI1 plasmid R64: roles of two sets of inverted repeat sequences within *oriT* in termination of R64 transfer. J Bacteriol 182: 3191- 3196.

Garcillán-Barcia MP, Francia MV, de la Cruz F (2009) The diversity of conjugative relaxases and its application in plasmid classification. FEMS Microbiol Rev 33(3): 657-687.

Gelvin SB (2003) Agrobacterium-mediated plant transformation: the biology behind the "gene-jockeying" tool. Microbiol Mol Biol Rev 67: 16-37.

Gomis-Rüth FX, Coll M (2006) Cut and move: protein machinery for DNA processing in bacterial conjugation. Curr Opin Struct Biol 16: 744-752.

González-Pérez B, Lucas M, Cooke LA, Vyle JS, de la Cruz F *et al.,* (2007) Analysis of DNA processing reactions in bacterial conjugation by using suicide oligonucleotides. EMBO J 26: 3847-3857.

Grohmann E, Guzmán LM, Espinosa M (1999) Mobilisation of the streptococcal plasmid pMV158: interactions of MobM protein with its cognate *oriT* DNA region. Mol Gen Genet 261: 707-715.

Grohmann E, Muth G, Espinosa M (2003) Conjugative plasmid transfer in Gram-positive bacteria. Microbiol Mol Biol Rev 67: 277-301.

Goessweiner-Mohr N, Arends K, Keller W, Grohmann E (2013). Conjugative plasmid transfer in Gram-positive bacteria. Plasmid. 70: 289-302

Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL (2008) The Vienna RNA Websuite. Nucleic Acids Res 36: W70–W74.

Guasch A, Lucas M, Moncalian G, Cabezas M, Perez-Luque R *et al.,* (2003) Recognition and processing of the origin of transfer DNA by conjugative relaxase TrwC. Nat Struct Biol 10(12): 1002-1010.

Guglielmini J, Quintais L, Garcillán-Barcia MP, de la Cruz F, Rocha EP (2011) The repertoire of ICE in prokaryotes underscores the unity, diversity, and ubiquity of conjugation. PLoS Genet 7: e1002222.

Guzmán L, Espinosa M (1997) The mobilization protein, MobM, of the streptococcal plasmid pMV158 specifically cleaves supercoiled DNA at the plasmid *oriT*. J Mol Biol 266: 688-702.

Jones BV, Sun F, Marchesi JR (2010) Comparative metagenomic analysis of plasmid encoded functions in the human gut microbiome. BMC Genomics 11: 46

Kee JM, Muir TW (2012) Chasing phosphohistidine, an elusive sibling in the phosphoamino acid family. ACS Chem Biol 7: 44-51.

Kopec J, Bergmann A, Fritz G, Grohmann E, Keller W (2005) 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 387: 401-409.

Larkin C, Datta S, Harley MJ, Anderson BJ, Ebie A *et al.,* (2005) Inter- and Intramolecular Determinants of the Specificity of Single-Stranded DNA Binding and Cleavage by the F Factor Relaxase. Structure 13(10): 1533-1544.

Laskowski RA (2009) PDBsum new things. Nucleic Acids Res. 37: D355–359

Lacroix B, Kozlovsky SV, Citovsky V (2008) Recent patents on agrobacterium-mediated gene and protein transfer, for research and biotechnology. Recent Pat DNA Gene Seq 2: 69-81.

Ley RE, Turnbaugh PJ, Klein S, Gordon JI. (2006) Microbial ecology: human gut microbes associated with obesity. Nature 444: 1022-1023.

Lindsay JA (2010) Genomic variation and evolution of Staphylococcus aureus. Int J Med Microbiol 300:98-103

Llosa M, Gomis-Ruth FX, Coll M, de la Cruz F (2001) Bacterial conjugation: a two-step mechanism for DNA transport. Mol Microbiol 45: 1-8.

Llosa M, Schröder G, Dehio C (2012) New perspectives into bacterial DNA transfer to human cells. Trends Microbiol 8: 355-359.

Lorenzo-Díaz F, Espinosa M (2009) Lagging strand DNA replication origins are required for conjugal transfer of the promiscuous plasmid pMV158. J Bacteriol 191: 720–727.

Lorenzo-Díaz F, Dostál L, Coll M, Schildbach JF, Menendez M *et al.,* (2011) The MobM-relaxase domain of plasmid pMV158: thermal stability and activity upon Mn2+ and specific DNA binding. Nucleic Acids Res 39:4315-29

Low HH, Gubellini F, Rivera-Calzada A, Braun N, Connery S et al., (2104) Structure of a type IV secretion system. Nature 508: 550-3.

Lucas M, González-Pérez B, Cabezas M, Moncalián G, Rivas G *et al.*, (2010) Relaxase DNA binding and cleavage are two distinguishable steps in conjugative DNA processing that involve different sequence elements of the *nic* site. J Biol Chem 285: 8918-8926.

Markham NR and Zuker M (2008) UNAFold: software for nucleic acid folding and hybridization. Methods Mol Biol. 453:3-31.

Monzingo AF, Ozburn A, Xia S, Meyer RJ, Robertus JD (2007) The structure of the minimal relaxase domain of MobA at 2.1 Å resolution. J Mol Biol 366: 165-178.

Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr 53(Pt 3): 240-255.

Painter J, Merritt EA (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr D Biol Crystallogr 62(Pt 4): 439-450.

Pansegrau W, Lanka E (1996a) Mechanisms of initiation and termination reactions in conjugative DNA processing: independence of tight substrate binding and catalytic activity of relaxase (TraI) of IncPa plasmid RP4. J Biol Chem 271: 13068-13076.

Pansegrau W, Lanka E (1996b) Enzymology of DNA strand transfer by conjugative mechanisms. Prog Nucl Acid Res Mol Biol 54: 197-251.

Potts RG, Habibi S, Cheng Y, Lujan SA, Redinbo MR (2010) The mechanism and control of DNA transfer by the conjugative relaxase of resistance plasmid pCU1. Nucl Acids Res doi:10.1093/nar/gkq303: 1-15.

Priebe SD, Lacks SA (1989) Region of the streptococcal plasmid pMV158 required for conjugative mobilization. J Bacteriol 171: 4778-4784.

Puttick J, Baker EN, Delbaere LT (2008) Histidine phosphorylation in biological systems. Biochimica et biophysica acta 1784(1): 100-105.

Read RJ (2001) Pushing the boundaries of molecular replacement with maximum likelihood. Acta Crystallogr D Biol Crystallogr 57(Pt 10): 1373-1382.

Riley DR, Sieber KB, Robinson KM, White JR, Ganesan A, Nourbakhsh S, Dunning Hotopp JC (2013) Bacteria-human somatic cell lateral gene transfer is enriched in cancer samples. PLoS Comput Biol 9: e1003107.

Robinson VL, Stock AM. (1999) High energy exchange: proteins that make or break phosphoramidate bonds. Structure 7: R47-53.

Russi S, Boer R, Coll M (2008) Molecular Machinery for DNA Translocation in Bacterial Conjugation. In: Lipps G, editor. Plasmids: Current Research and Future Trends. Bayreuth (Germany): Caister Academic Press. pp. 183-213.

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Scherzinger E, Lurz R, Otto S, Dobrinski B (1992) In vitro cleavage of double- and single-stranded DNA by plasmid RSF1010-encoded mobilization proteins. Nucl Acids Res 20: 41-48.

Schneider TR, Sheldrick GM (2002) Substructure solution with SHELXD. Acta Crystallogr D Biol Crystallogr 58(Pt 10 Pt 2): 1772-1779.

Schroder G, 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.

Schröder G, Schuelein R, Quebatte M, Dehio C (2011) Conjugative DNA transfer into human cells by the VirB/VirD4 type IV secretion system of the bacterial pathogen Bartonella henselae. Proc Natl Acad Sci USA 108:14643-14648.

Slonczewski JL, Fujisawa M, Dopson M, Krulwich TA (2009) Cytoplasmic pH measurement and homeostasis in bacteria and archaea. Adv Microb Physiol 55: 1-79.

Smillie C, Garcillán-Barcia MP, Francia MV, Rocha EP, de la Cruz F (2010) Mobility of Plasmids. Microbiol Mol Biol Rev 74: 434–452.

Stern JC, Schildbach JF (2001) DNA recognition by F Factor TraI36: Highly sequence-specific binding of single-stranded DNA.Biochemistry, 40, 11586–11595.

Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW (1990) Use of T7 RNA polymerase to direct expression of cloned genes. Meth Enzymol 185: 60-89.

Szpirer CY, Faelen M, Couturier M (2001) Mobilization function of the pBHR1 plasmid, a derivative of the broad-host-range plasmid pBBR1. J Bacteriol 183: 2101-2110.

Terwilliger T (2004) SOLVE and RESOLVE: automated structure solution, density modification and model building. J Synchrotron Radiat 11(Pt 1): 49-52.

Thomas CM (2000) The horizontal gene pool. Amsterdam: Harwood Academic Publishers. 419 p.

Uhlemann AC, Dordel J, Knox JR, Raven KE, Parkhill J et al. (2014) Molecular tracing of the emergence, diversification, and transmission of S. aureus sequence type 8 in a New York community. Proc Natl Acad Sci U S A 111:6738-43

Uson I, Sheldrick GM (1999) Advances in direct methods for protein crystallography. Curr Opin Struct Biol 9(5): 643-648.

Vedantam G, Knopf S, Hecht DW (2006) *Bacteroides fragilis* mobilizable transposon Tn*5520* requires a 71 base pair origin of transfer sequence and a single mobilization protein for relaxosome formation during conjugation. Mol Microbiol 59(1): 288-300.

Wallace AC, Laskowski RA, Thornton JM (1995) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng 8(2): 127-134.

Williams SL, Schildbach JF (2006) Examination of an inverted repeat within the F factor origin of transfer: context dependence of F TraI relaxase DNA specificity. Nucl Acids Res 34(2): 426-435.

WHO (2014), Antimicrobial resistance: global report on surveillance 2014, http://www.who.int/drugresistance/documents/surveillancereport/en/

Wozniak RAF, Waldor MK (2010) Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. Nat Rev Microbiol. 8:552-63.

Yang W (2008) An equivalent metal ion in one- and two-metal-ion catalysis. Nat Struct Mol Biol.15:1228-31

Yang W (2011) Nucleases: diversity of structure, function and mechanism. Q Rev Biophys. 44:1-93.

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31: 3406-15

Fernández-López C, Pluta R, Pérez-Luque R, Rodríguez-González L, Espinosa M, Coll M, Lorenzo-Díaz F, Boer DR. Functional properties and structural requirements of the plasmid pMV158-encoded MobM relaxase domain. J Bacteriol. 2013; 195(13):3000-8.

doi: 10.1128/JB.02264-12

3.3. Functional Properties and Structural Requirements of the Plasmid pMV158-Encoded MobM Relaxase Domain

The chapter 3.3 has been published in Journal of Bacteriology in with the following details:

Functional Properties and Structural Requirements of the Plasmid pMV158-Encoded MobM Relaxase Domain

Cris Fernández-López[a], Radoslaw Pluta[b,c], Rosa Pérez-Luque[b,c], Lorena Rodríguez-González[a], Manuel Espinosa[a], Miquel Coll[b,c], Fabián Lorenzo-Díaz*[d,] D. Roeland Boer*[b,c]

[a] Centro de Investigaciones Biológicas, CSIC, Madrid, Spain.
[b] Institute for Research in Biomedicine (IRB Barcelona), Barcelona, Spain
[c] Institute of Molecular Biology of Barcelona (IBMB-CSIC) Barcelona, Spain
[d] Instituto Universitario de Enfermedades Tropicales y Salud Pública de Canarias,Tenerife, Spain

Received 9 January 2013, Accepted 20 April 2013, Published ahead of print 26 April 2013
Address correspondence to D. Roeland Boer, rbocri@ibmb.csic.es, or Fabián Lorenzo-Díaz, florenzo@ull.edu.es.

doi:10.1128/JB.02264-12

CHAPTER 4
DISCUSSION

The work presented in this thesis provides the first structural analysis of a representative of Mob_Pre/MOB_V family of relaxases, the MobM protein. The combinatorial approach of X-ray crystallography, computational analysis and protein-DNA biochemistry has led to new insights in metal-dependent histidine-mediated DNA cleavage, relaxase-DNA adduct stabilization, active site auto-inhibition by specific intermolecular interactions, relaxase-induced melting of the long bulged DNA hairpin and in the dissection of the relaxase N-terminal sequence into two distinct regions, the nuclease and the hinge regions. Moreover, literature and database-based analyses of MOB families let to the description of features unique to MOB_V relaxases and have shown for the first time that the MOB_V/Mob_Pre relaxases are the major relaxases found *Staphylococcus aureus*, which is a cause of the highest number of lethal antibiotic-resistant infections in USA (CDC, 2013). Importantly, the results presented here can contribute to the understanding of *S. aureus* preference towards use of histidine relaxases and may provide an avenue for the process of drug discovery against *S. aureus*. Nowadays, bacterial resistance is a major health threat (CDC, 2013; WHO, 2014) and MRSA is the major killer among antibiotic-resistant infections in USA.

The chapter 3.1 focuses on the description of MobM structure-function catalytic mechanism of DNA nicking. Our structural analysis together with in vitro and in vivo experiments show that MobM is the first example of a metal-dependent nuclease with a nucleophilic histidine that proceeds through a unique to MOB_V relaxases covalent DNA-histidine adduct. The use of a histidine nitrogen atom instead of an oxygen atom of a tyrosine/serine residue makes MobM the first example of a DNA cleavage-and-ligation enzyme that operates through a formation of the phosphor-nitrogen bond. Chemical properties of the P-N bond, i.e. high energy and liability at low pH, which are different to the properties of the P-O bond, may lead to the design of enzymes with

such operating characteristics. Interestingly, importance of histidine phosphorylation in biological systems, another process where such bond is formed, is currently under growing interest (Kee and Muir 2012; Piggott and Attwood 2013). Histidine is a special amino acid due to its high reactivity that leads to a variety of roles it can play in proteins, especially as a key residue in many enzyme active sites where it can form a metal chelation cluster (e.g. HUH nucleases) or serves as any of the components of acid-base catalysis (e.g. serine proteases, ββα-Me nucleases or PLD nucleases) (Fersht, 1999), however it's most common role is to be a general base (hydrogen bond acceptor) and strengthen the nucleophilic property of another residue or water. The pKa values of nucleophiles found in nucleases are 6 (His), 10 (Tyr), 12-14 (2'-OH of ribose), 13 (Ser) and 16 (H_2O) (Yang, 2011). Importantly, histidine is the only nucleophile, which pKa is close to the neutral pH and therefore the only one that does not require deprotonation for activation, however when it adopts its unprotonated form then it becomes a particularly strong nucleophile.

In addition to the histidine triad, a landmark of the HUH-type conjugative relaxases, the active sites of MOB_V relaxases have a fourth residue for the metal ligation, a glutamate residue (H126, E129, H133 and H135 in MobM). Intriguingly, another pMV158-encoded HUH endonuclease, the replicative relaxase RepB, also employs an additional negatively charged residue for the manganese ion coordination (H39, D42, H55 and H57). Taking into account that RepB is a tyrosine relaxase one can assume that the use of an extra carboxylate group in the active site of MobM and RepB is related to the manganese ion coordination and activated active site fixation (e.g. β4-β5 loop) rather than the use of histidine as a nucleophile. Additionally, in MobM E129 interacts with the scissile phosphate and R25, which contacts the scissile phosphate and forms a salt bridge with D128 of β4-β5 loop. Mutational study performed on another MOB_V relaxase, the Mob_pBHR1 protein, shown that the Mob_pBHR1 residues D120 and E121 (equivalents of MobM D128 and E129) are essential for plasmid pBHR1 conjugal transfer (Szpirer *et al.*, 2001). It would be interesting to make a comparative mutagenesis analysis between the histidine relaxase MobM and the tyrosine relaxase RepB. How does the fourth metal ligand influence the cleavage/ligation reaction mechanism? Is it needed for the selective binding of manganese over magnesium? Interestingly, both theses metals support cleavage (data not shown), but if

magnesium can also support histidine-DNA adduct stabilization and DNA ligation is unknown. The MobM H22Y mutant turned out to be inactive for DNA nicking and plasmid transfer, but would perhaps RepB-inspired engineering of the MobM active site result in MobM becoming a tyrosine relaxase and if so, then how it would affect the pMV158 plasmid transfer?

Chapter 3.2 includes a detailed structural description of the MobM relaxase and MobM-DNA interactions, including interactions that result in MobM active site inhibition, which could possibly serve as molecular basis of a relaxase/plasmid auto-regulatory system. Moreover, the analysis of the DNA binding motifs of MobM and relaxases from other MOB families is presented here. A model where MobM recognizes elongated bulged IR3 DNA hairpin within the *oriT*, subsequently melts its base and guides the ssDNA region into the active site is proposed. The comparison of the MobM structure to structures of relaxases from other MOB families demonstrate high structural similarity among these proteins, in spite of a very low sequence homology between MobM and other relaxases. Two major differences in protein fold of relaxases from different MOB families lay in the way DNA binding is supported. Namely, MobM thumb that directs the ssDNA to the active site is made of an extended loop, while in MOB_F relaxases (TrwC and TraI) it is built by a much longer, mainly α-helical protein fragment. In case of MOB_Q relaxases this part lacks structural description, since the MobA structure is based on a construct (MobA_1-184) that terminates too early to capture the thumb and in the NES structure the 25 terminal residues of the NES_1-220 construct is untraceable in the electron density map, i.e. it is either degraded or in a conformation that is too much flexible to form repetitive pattern in the crystal lattice. The second evident difference is the presence of extensive secondary structures that interact with the tip of the DNA hairpin in MOB_F TrwC relaxase, and are absent in structures/sequences of MOB_V MobM and MOB_Q NES relaxases. Surprisingly, the biggest and the most biotechnologically used family of relaxases, the MOB_P family, to which belong relaxases from the Agrobacterium transformation system, still lacks a structural description.

MobM-DNA specific interactions within the hairpin stem region are limited to a few nucleotides contacting two highly conserved residues (R74 and K149). This is however a region for

which all but one structurally described relaxases use a common β-turn RxD/N motif (MobM_74-76: RKD) for the penetration of DNA hairpin minor groove. Interestingly, the TraI_pCU1 relaxase that is the exception and lacks the RxD/N motif was shown to binds its cognate DNA with low affinity and suggested to need aid of additional partners to initiate the high affinity sequence-specific DNA binding (Nash *et al*., 2009). On the other hand, motifs for interactions with the hairpin major grove vary even inside individual MOB families, e.g. different motifs are used by $MOB_{F1.1}$ subfamily (TrwC and TraI_pCU1) and by $MOB_{F1.2}$ subfamily (TraI_pF). Moreover, MOB_F relaxases use a unique to their family extensive β-hairpin motif to anchor themselves in the major grove. MobM achieves binding to the major grove by just one conserved lysine (K149) from a 4-residue α-helix, which defines the minimal major groove interacting element among structurally characterized relaxases. Not surprisingly, NES relaxase deletion mutants that lacked regions responsible for stem binding, analogous to MobM R74 and K149 areas, were shown to have low DNA affinity, disrupted DNA cleavage-religation equilibrium and reduced conjugative DNA transfer (Edwards *et al*., 2013). Downstream to the hairpin stem, the ssDNA region is where there the specific protein-DNA interactions are abundant and the specificity of relaxase-plasmid systems is achieved. In MobM-DNA structures, substantial kink is observed in the DNA backbone between Thy18-Ade19 and Thy20-Ade21 nucleotides of the ATA bulge, and the same situation is observed for analogous bases in TrwC and NES relaxases. Further down, the enclosure of Gua22 and of downstream bases by the thumb, made of an extended loop, wraps on the ssDNA and locks the separated strand into a path towards the active site ensuring the formation of the appropriate geometry for subsequent nicking activity. For MobM the conformational changes triggered by the thumb placement on the separated DNA strand could induce the C-terminal domain relocation with respect to the relaxase domain, which is the subject of the chapter 3.3. In contrast, the locking system in other relaxases is based mainly on nucleotides anchoring in protein cavities as observed for bases in positions -3 and -5 to the *nic* site in the TrwC- and TraI-DNA structures and in position -3 in the NES-DNA structure. Further downstream, the importance of Gua26 or Gua8’ in the active site inhibited structures is highlighted, since this base is bound in a very similar way in both types of MobM-DNA structures presented in this work, suggesting

that this region of the MobM relaxase provides a high affinity binding site for guanidine. Taking into account that Gua26 forms the non-canonical Thy-Gua base pair (bases positions -2 and 0 to the *nic* site) found at the U-turn that directs the scissile phosphate into the reach of the nucleophilic histidine, this guanidine-specificity would together with the binding of the bases proceeding Gua26 provide the specificity for the MobM nicking/ligating activity. Indeed, elimination of the non-canonical Gua-Thy base pair (Gua to Cyt substitution) in plasmid F system dramatically reduced plasmid transfer, killed the oligonucleotide nicking and diminished protein-DNA binding (Stern and Schildbach 2001; Williams and Schildbach 2006). Interestingly, we observe for the first time that the flexibility of the C-terminal thumb and the specific protein-DNA and DNA-DNA interactions can lead to the formation of a complex where one of the active site is inhibited by DNA from the neighbouring molecule. We speculate that the observed metal-binding site occupation by a phosphodiester moiety and active site disruption may serve as an auto-regulatory mechanism. Moreover, it also provides a starting point for computer-aided pharmacophore and/or structure-based design of drugs targeting conjugative DNA transfer. Conventional antibiotics kill bacteria by targeting essential cell molecules and therefore exert strong selective pressure for the development and maintenance of resistance. By contrast, an antimicrobial agent that inhibits virulence without a profound lethal cytotoxic effect would generate minimal selective pressure for the development of resistance and would be selective towards a pathogen (Clatworthy *et al.,* 2007; Rasko and Sperandio, 2010). Targeting relaxases could not only provide a way to reduce/impede the spread of antibiotic resistance, but can also lead to relaxase-dependent bacterial cell lethality, as was unexpectedly shown by Lujan *et al.* in a work, where presumably interaction between bisphosphonates and TraI_pF relaxase in cells carrying plasmid F resulted in a competitive disadvantage relative to cells without conjugative plasmids (Lujan *et al*., 2007).

Another important observation stated in the chapter 3.2 is the prevalence of MOB_V relaxases in Gram+ antibiotic-resistant human pathogens, particularly in *S. aureus* in which almost exclusively MOB_V histidine conjugative HUH-fold relaxases and the MOB_T tyrosine replicative/conjugative DDE-fold relaxases are being found. Interestingly members of these two families can be

encoded within one MGE, as in case of 3.9 kb spectinomycin resistance plasmid pDJ91S from methicillin-resistant *S. aureus* (MRSA) human and animal isolates (Jamrozy *et al*., 2014). It would be interesting to analyse all *S. aureus* MGEs to see if these relaxases are often co-found in MRSA strains. Structural characterization of a MOB_T relaxase is another interesting case to study. Structure prediction and functional analysis of a MOB_C family member, another MOB family with the DDE-fold was performed recently (Francia *et al*., 2013), but the mechanism of catalysis of a DDE-relaxase remains unclear. It would be interesting to spot the differences between the MOB_T and MOB_C relaxases as the latter family seems to be 'invisible' for Staphylococcus and Streptococcus genera, in spite of presence of MOB_C relaxases is other Firmicutes genera.

Finally, we try to link the use of a histidine nucleophile and the fact that 90% of conjugative HUH-type relaxases found in *S. aureus* are MOB_V relaxases. Is it the high energy of the P-N bond and the lack of the need for histidine deprotonation that make MobM-like relaxases the most attractive ones for *S. aureus*? Does the use of histidine and not tyrosine relaxases make MRSA strains more successful or there is no such correlation? Does the low number of MOB_V relaxases found in Gram- bacteria arises from to the presence of periplasmatic space in Gram- bacteria? The periplasmatic space lacks pH homeostasis regulation and can be easily acidified, which would lead to the breakage of P-N bond. Is the high energy of a P-N bond and therefore ease of bond transfer to other atoms a cause to histidine relaxases existence almost only in short MGEs (but on the other hand, their domination over tyrosine relaxases in small MGEs)? With how many different antibiotic resistance genes are MOB_V relaxases coupled? Are plasmids/ICEs with multiple relaxases from different MOB families more 'successful' than MGEs that carry single relaxases? What are the rules that drive compatibility of multiple relaxases existence in a single MGE?

In the third chapter 3.3 we show that MobM N-terminal 199- and 243-residue constructs behave as monomers that bear the DNA binding and cleavage activity and we propose that the region between residues 200 and 243, which follows the N-terminal relaxase domain, is important for ssDNA processivity and functions

as signal transducer of substrate binding to the C-terminal domain. The MobMN199 construct which covers the Pfam-defined relaxase domain (residues 1-194 for MobM) is able to nick and therefore relax supercoiled plasmid DNA, but does not cut examined ssDNA oligonucleotides, whereas the MobMN243 construct is able cut both types of DNA substrates. Thus, the nuclease activity *per se* is encoded within the first 199-residues, and residues 200-243 are needed for the ssDNA positioning for the nuclease reaction. Noteworthy, one of the steps of conjugative plasmid processing is the nicking of newly synthesized ssDNA *oriT*. The MobM200-243 region is predicted to be highly disordered and our protein stability/MS analysis showed its increased susceptibility to degradation upon the relaxosome assembly (metal ion and DNA bound), especially within a 20-reisdue region (190-210), where most of the degradation occurred. The observed results suggest that this region undergoes conformational changes during plasmid processing and may serve as a hinge allowing for a reorientation of the N- and C-terminal domains, which could transduce the 'substrate bound' signal from the relaxase N-terminal domain to the C-terminal domain, ultimately allowing appropriate response of the C-terminal domain in further steps of MobM-mediated DNA transfer. Further above, three 35-residue regions that form coiled coils (CC) are predicted and the MobM dimerization was suggested to be achieved by a putative leucine-zipper localized in one of them (Fernandez-Lopes *et al*., 2013). Moreover, the C-terminal domain of MobM is responsible for membrane association and was suggested to interact with the coupling protein (Lorenzo-Diaz *et al*., 2011), which mediates the interactions between the relaxosome and the membrane transfer apparatus (Alvarez-Martinez and Christie, 2009). A quadruple mutant (R421P, K423P, K425P, and K427P) designed to break the α-helix structure within the beginning of the third CC region abolished MobM membrane association and plasmid transfer between pneumococcal strains (de Antonio *et al*., 2004; Fernandez-Lopez *et al*. 2013). It remains unknown if the localized degradation occurs *in vivo* during the conjugative plasmid

transfer. It would be interesting to test whether the MobM-DNA complex that is being transferred into the recipient cell contains the whole MobM protein (as demonstrated for TrwC, Garcillán-Barcia *et al*., 2007) or the version in which the non-nuclease part of MobM is cut before entering the membrane channel. Among relaxases of similar to MobM domains organization, i.e. a nuclease domain followed by a domain without a catalytic activity, there are structural data available only for the C-terminal domain of the NES relaxase (Edwards *et al*., 2013), which was found to be all alpha helical and being a new fold. Similarly, the C-terminal domain of MobM is predicted to be all alpha helical. However, there are several differences between them, namely the NES C-terminal domain, which is about 150 residues longer, is predicted to has only one 21-residue CC region and there are in the NES crystal structure pairs of helices interacting with each, but none of them forming CC region. Moreover, SAXS experiments showed that the full length NES alone dynamically form range of complexes, while the full length NES-DNA complex was identified as monomeric (Edwards *et al*., 2013). On the other hand, the full length MobM was shown by analytical ultracentrifugation to be a stable dimer (de Antonio *et al*., 2004), possibly due to strong adhesion forces generated by the putative leucine fingers. Unfortunately, there is no data available about the oligomeric state of the full length MobM-DNA complex. Interestingly, computational searches describe the MobM C-terminal domain as being distantly homologous to chromosome segregation proteins, which would be in agreement with the MobM membrane association.

CHAPTER 5
CONCLUSIONS

1. Members of the MOB_V/Mob_Pre family of relaxases, of which MobM is a prototypic representative, constitute 90% of MOB HUH conjugative relaxases and 50% of all MOB relaxases found in *Staphylococcus aureus*, which is the most lethal Gram-positive drug resistant human pathogen. Moreover, among relaxase families found in Gram-positive bacteria only MOB_V relaxases are being found much more often in pathogenic bacteria than in non-pathogenic ones.

2. In spite of very low sequence homology, the overall structure of MobM relaxase domain is similar to its counterparts from MOB_F and MOB_Q relaxase families. The two major differences in protein fold between MobM and other relaxases lay in i) the thumb construction, in MobM the thumb being made of an extended loop wrapping on ssDNA and ii) in the presence (in MOB_F relaxases) of an additional β-hairpin motif for interactions with DNA hairpin major grove. The uniqueness of MobM lays in the active site composition/chemistry, i.e. MobM structure shows that MOB_V relaxases use histidine instead of tyrosine for DNA cleavage and covalent adduct formation. Moreover, they use additional glutamate for the metal coordination.

3. MobM uses canonical one-metal-ion mechanism, reminiscent of two-metal-ion catalysis, for which one of the metal ions is replaced by positively charged residue (R25 in MobM); MobM relaxase uses histidine 22 to cleave the substrate DNA and is the first example of a nuclease that utilizes a metal-dependent histidine-mediated DNA cleavage mechanism (note that metal-independent histidine nucleases

are known already) and the first example of a nuclease/ligase that utilizes a histidine-mediated DNA cleavage-and-ligation mechanism. In MobM, R25 that contacts the scissile phosphate and 'substitutes the metal ion A' is fixed in its position by a hydrogen bond with metal-ligating E129 and a salt bridge with D128. Mutations of residues equivalent to R25, D128 and E129 in other one-metal-ion nucleases result in non-functional proteins.

4. The phosphor-nitrogen bond present in MOB_V relaxase-DNA adducts, in opposite to the phosphor-oxygen bond, is characterized by the ease of energy transfer to other atoms and by the acid liability, two features which could possibly shape the small size of mobile genetic elements (MGEs) that employ histidine relaxases and the low abundance of histidine relaxases in Gram-negative bacteria, in which the periplasmatic space (it lacks pH homeostasis regulation) can expose the P-N bond to the acidic environment.

5. Origins of transfer of other plasmids have potential to form IR3-like long bulged hairpins, which in MobM is predicted to be the most thermodynamically favoured. The IR3 ATA bulge is probably used by MobM for initial melting of the IR3 stem base (formation of the IR1 hairpin) and simultaneous placement of the unwound strand in the path towards the active site.

6. MobM specific interactions with the IR1 hairpin stem are limited to a few nucleotides, while many MobM-DNA specific interactions are observed for the downstream ssDNA region, which is pushed into the active site by a thumb made of an extended loop that wraps over the ssDNA.

7. MobM and other structurally described relaxases that have high affinity for their cognate DNA use common β-turn RxD/N motif (MobM_74-76: RKD) to penetrate the DNA hairpin minor groove; on the other hand, various motifs for interactions with hairpin major grove are used by different MOB subfamilies, in MobM such motif being just one conserved lysine (K149) from a 4-residue α-helix, which defines the minimal major groove interacting element among structurally characterized relaxases.

8. As observed in MobM-DNA26 structures, the flexibility of the thumb and the little finger and the specific intermolecular DNA-DNA interactions can lead to the active site disruption and metal-binding site occupation by a phosphodiester moiety, a scenario that may underlie molecular basis of conjugative auto-regulatory mechanism (copy number-dependent inhibition or resting state just before the cleavage reaction

9. MobM region that lays at the boundary between N- and C-terminal domains (residues 200-243) shows increased susceptibility to degradation upon the relaxosome assembly, suggesting that this region may serve as a hinge allowing for a conformational change in relative orientation of the N- and C-terminal domains during substrate processing, which could function as a signal transducer of substrate binding by the relaxase domain to the C-terminal domain, which was suggested previously to interact with the plasma membrane and other proteins involved in conjugation.

10. In contrast to the relaxase-domain-only construct (MobMN199), which is able to cut plasmid supercoiled DNA, but not single-stranded DNA oligonucleotides, the construct that covers the relaxase domain and the following interdomain region (MobMN243) is capable of nicking both

types of DNA substrates, what indicates that the interdomain region is involved in the ssDNA positioning for the nuclease reaction rather than in the nicking mechanism itself (one of the steps of plasmid processing is the nicking of newly synthesized ssDNA *oriT* region).

Bibliography

Alvarez-Martinez CE, Christie PJ (2009) Biological diversity of prokaryotic type IV secretion systems. Microbiol Mol Biol Rev 73: 775-808.

Alsmark C, Foster PG, Sicheritz-Ponten T, Nakjang S, Martin Embley T, Hirt RP (2013) Patterns of prokaryotic lateral gene transfers affecting parasitic microbial eukaryotes. Genome Biol. 2013 14:R19

Aronovich EL, Scott McIvor R, Hackett PB (2011) The Sleeping Beauty transposon system: a non-viral vector for gene therapy. Hum Mol Genet. 20: R14–R20.

Blanchard JL, Lynch M (2000). Organellar genes: why do they end up in the nucleus? Trends Genet. 16: 315–20

Berleman J, Auer M (2013) The role of bacterial outer membrane vesicles for intra- and interspecies delivery. Environ Microbiol. 15:347-54.

Cambray G, Guerout AM, Mazel D (2012) Integrons. J Bacteriol Virol. 42:181-188.

CDC (2013), Antibiotic resistance threats in the United States, 2013, http://www.cdc.gov/drugresistance/threat-report-2013/

Chandler M, de la Cruz F, Dyda F, Hickman AB, Moncalian G, Ton-Hoang B (2013) Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. Nat Rev Microbiol 11: 525-538.

Chen I, Dubnau D (2004) DNA uptake during bacterial transformation. Nat Rev Microbiol. 2:241-9.

Chiura HX, Kogure K, Hagemann S, Ellinger A, Velimirov B (2011) Evidence for particle-induced horizontal gene transfer and serial transduction between bacteria. FEMS Microbiol Ecol. 76:576–591.

Christie PJ (2004) Type IV secretion: the Agrobacterium VirB/D4 and related conjugation systems. Biochim Biophys Acta 1694(1-3): 219-234.

Clatworthy AE, Pierson E, Hung DT (2007) Targeting virulence: a new paradigm for antimicrobial therapy. Nat Chem Biol. 2007 3:541-8.

Corvaglia AR, François P, Hernandez D, Perron K, Linder P, Schrenzel J (2010) A type III-like restriction endonuclease functions as a major barrier to horizontal gene transfer in clinical Staphylococcus aureus strains. Proc Natl Acad Sci U S A. 107:11954-8.

Daccord A, Ceccarelli D, Rodrigue S, Burrus V (2013) Comparative analysis of mobilizable genomic islands. J Bacteriol. 195:606-14.

David MZ, Daum RS (2010) Community-associated methicillin-resistant Staphylococcus aureus: epidemiology and clinical consequences of an emerging epidemic. Clin Microbiol Rev. 23:616-87.

David MZ, Cadilla A, Boyle-Vavra S, Daum RS (2014) Replacement of HA-MRSA by CA-MRSA infections at an academic medical center in the midwestern United States, 2004-5 to 2008. PLoS One. 9:e92760.

Datta S, Larkin C, Schildbach JF (2003) Structural insights into single-stranded DNA binding and cleavage by F factor TraI. Structure 11: 1369-1379.

Dubey GP, Ben-Yehuda S (2011) Intercellular nanotubes mediate bacterial communication. Cell. 144:590-600.

de Antonio C, Farias ME, de Lacoba MG, Espinosa M (2004) Features of the plasmid pMV158-encoded MobM, a protein involved in its mobilization. J Mol Biol 335(3): 733-743.

de la Cruz F, Davies J (2000) Horizontal gene transfer and the origin of species: lessons from bacteria. Trends Microbiol 128: 128-133.

de la Cruz F, Frost LS, Meyer RJ, Zechner EL (2010) Conjugative DNA metabolism in Gram-negative bacteria. FEMS Microbiology Reviews 34(1): 18-40.

del Solar G, Díaz R, Espinosa M (1987) Replication of the streptococcal plasmid pMV158 and derivatives in cell-free extracts of *Escherichia coli*. Mol Gen Genet 206: 428-435.

Dostál L, Shao S, Schildbach JF (2011) Tracking F plasmid TraI relaxase processing reactions provides insight into F plasmid transfer. Nucleic Acids Res 39: 2658–2670.

Edwards JS, Betts L, Frazier ML, Pollet RM, Kwong SM et. al (2013). Molecular basis of antibiotic multiresistance transfer in Staphylococcus aureus. Proc Natl Acad Sci U S A. 110: 2804-9.

Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60(Pt 12 Pt 1): 2126-2132.

Farías ME, Espinosa M (2000) Conjugal transfer of plasmid pMV158: uncoupling of the pMV158 origin of transfer from the mobilization gene *mobM*, and modulation of pMV158 transfer in *Escherichia coli* mediated by IncP plasmids. Microbiology 146: 2259-2265.

Fernández-López C, Lorenzo-Díaz F, Pérez-Luque R, Rodríguez-González L, Boer R et. al (2013) Nicking activity of the pMV158 MobM relaxase on cognate and heterologous origins of transfer. Plasmid. 70:120-30.

Fersht A (1999) Structure and Mechanism in Protein Science - A Guide to Enzyme Catalysis and Protein Folding, (W.H. Freeman and Co).

Fogg PC, Westbye AB, Beatty JT (2012) One for all or all for one: heterogeneous expression and host cell lysis are key to gene transfer agent activity in Rhodobacter capsulatus. PLoS One. 7:e43772.

Francia MV, Varsaki, A., Garcillán-Barcia, M. P., Latorre, A., Drainas, C., de la Cruz, F. (2004) A classification scheme for mobilization regions of bacterial plasmids. FEMS Microbiology Reviews 28: 79-100.

Francia MV, Clewell DB, de la Cruz F, Moncalián G (2013) Catalytic domain of plasmid pAD1 relaxase TraX defines a group of relaxases related to restriction endonucleases. Proc Natl Acad Sci U S A. 110:13606-11.

Frost LS, Leplae R, Summers AO, Toussaint A (2005) Mobile genetic elements: the agents of open source evolution. Nat Rev Microbiol. 3:722-32.

Fukuda H, Ohtsubo E (1997) Roles of Tra I protein with activities of cleaving and rejoining the single-stranded DNA in both initiation and termination of conjugal DNA transfer. Genes to Cell 2: 735-751.

Furuya EY, Lowy FD (2006) Antimicrobial-resistant bacteria in the community setting. Nat Rev Microbiol. 4:36-45.

Furuya N, Komano T (2000) Initiation and termination of DNA transfer during conjugation of IncI1 plasmid R64: roles of two sets of inverted repeat sequences within *oriT* in termination of R64 transfer. J Bacteriol 182: 3191- 3196.

Garcillán-Barcia MP, Francia MV, de la Cruz F (2009) The diversity of conjugative relaxases and its application in plasmid classification. FEMS Microbiol Rev 33(3): 657-687.

Gelvin SB (2003) Agrobacterium-mediated plant transformation: the biology behind the "gene-jockeying" tool. Microbiol Mol Biol Rev 67: 16-37.

Gilbert C, Cordaux R (2013) Horizontal transfer and evolution of prokaryote transposable elements in eukaryotes. Genome Biol Evol. 5:822-32.

González-Pérez B, Lucas M, Cooke LA, Vyle JS, de la Cruz F *et al.*, (2007) Analysis of DNA processing reactions in bacterial conjugation by using suicide oligonucleotides. EMBO J 26: 3847-3857.

Goessweiner-Mohr N, Arends K, Keller W, Grohmann E (2013). Conjugative plasmid transfer in Gram-positive bacteria. Plasmid. 70: 289-302.

Gogarten MB, Gogarten JP, Olendzenski L (2009) Horizontal Gene Transfer: Genomes in Flux. (Humana Press).

Green MR, Sambrook J (2012) Molecular Cloning: A Laboratory Manual. (CSHL Press)

Guasch A, Lucas M, Moncalian G, Cabezas M, Perez-Luque R *et al.*, (2003) Recognition and processing of the origin of transfer DNA by conjugative relaxase TrwC. Nat Struct Biol 10(12): 1002-1010.

Guglielmini J, Quintais L, Garcillán-Barcia MP, de la Cruz F, Rocha EP (2011) The repertoire of ICE in prokaryotes underscores the unity, diversity, and ubiquity of conjugation. PLoS Genet 7: e1002222.

Guy L, Nystedt B, Toft C, Zaremba-Niedzwiedzka K, Berglund EC, et al. (2013) A Gene Transfer Agent and a Dynamic Repertoire of Secretion Systems Hold the Keys to the Explosive Radiation of the Emerging Pathogen Bartonella. PLoS Genet 9: e1003393.

Guzmán L, Espinosa M (1997) The mobilization protein, MobM, of the streptococcal plasmid pMV158 specifically cleaves supercoiled DNA at the plasmid oriT. J Mol Biol 266: 688-702.

Halary S, Leigh JW, Cheaib B, Lopez P, Bapteste E (2010) Network analyses structure genetic diversity in independent genetic worlds. Proc Natl Acad Sci U S A. 107:127-32.

Jamrozy DM, Coldham NG, Butaye P, Fielder MD (2014) Identification of a novel plasmid-associated spectinomycin adenyltransferase gene spd in methicillin-resistant Staphylococcus aureus ST398 isolated from animal and human sources. J Antimicrob Chemother. 69:1193-6

Jones BV, Sun F, Marchesi JR (2010) Comparative metagenomic analysis of plasmid encoded functions in the human gut microbiome. BMC Genomics 11: 46

Kee JM, Muir TW (2012) Chasing phosphohistidine, an elusive sibling in the phosphoamino acid family. ACS Chem Biol 7: 44-51.

Keeling PJ, Palmer JD. (2008) Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet. 9: 605-18.

Lang AS, Beatty JT (2007) Importance of widespread gene transfer agent genes in alpha-proteobacteria. Trends Microbiol. 15:54-62.

Lang AS, Zhaxybayeva O, Beatty JT (2012) Gene transfer agents: phage-like elements of genetic exchange. Nat Rev Microbiol. 10:472-82.

Langille MG, Meehan CJ, Beiko RG. (2012) Human microbiome: a genetic bazaar for microbes? Curr Biol. 22:R20-2.

Larkin C, Datta S, Harley MJ, Anderson BJ, Ebie A *et al.,* (2005) Inter- and Intramolecular Determinants of the Specificity of Single-Stranded DNA Binding and Cleavage by the F Factor Relaxase. Structure 13(10): 1533-1544.

Lacroix B, Kozlovsky SV, Citovsky V (2008) Recent patents on agrobacterium-mediated gene and protein transfer, for research and biotechnology. Recent Pat DNA Gene Seq 2: 69-81.

Lederberg J, Tatum El (1946) Gene recombination in Escherichia coli. Nature. 158:558

Lee CA, Thomas J, Grossman AD (2012) The Bacillus subtilis conjugative transposon ICEBs1 mobilizes plasmids lacking dedicated mobilization functions. J Bacteriol. 194:3165-72.

Ley RE, Turnbaugh PJ, Klein S, Gordon JI. (2006) Microbial ecology: human gut microbes associated with obesity. Nature 444: 1022-1023.

Lindsay JA (2010) Genomic variation and evolution of Staphylococcus aureus. Int J Med Microbiol 300:98-103

Llosa M, Gomis-Ruth FX, Coll M, de la Cruz F (2001) Bacterial conjugation: a two-step mechanism for DNA transport. Mol Microbiol 45: 1-8.

Llosa M, Schröder G, Dehio C (2012) New perspectives into bacterial DNA transfer to human cells. Trends Microbiol 8: 355-359.

Lorenzo-Díaz F, Espinosa M (2009) Lagging strand DNA replication origins are required for conjugal transfer of the promiscuous plasmid pMV158. J Bacteriol 191: 720–727.

Lorenzo-Díaz F, Dostál L, Coll M, Schildbach JF, Menendez M *et al.,* (2011) The MobM-relaxase domain of plasmid pMV158: thermal stability and activity upon Mn2+ and specific DNA binding. Nucleic Acids Res 39:4315-29

Low HH, Gubellini F, Rivera-Calzada A, Braun N, Connery S et al., (2104) Structure of a type IV secretion system. Nature 508: 550-3.

Lowder BV, Guinane CM, Ben Zakour NL, Weinert LA, Conway-Morris A et al. (2009) Recent human-to-poultry host jump, adaptation, and pandemic spread of Staphylococcus aureus. Proc Natl Acad Sci U S A. 106:19545-50

Lucas M, González-Pérez B, Cabezas M, Moncalián G, Rivas G *et al.,* (2010) Relaxase DNA binding and cleavage are two distinguishable steps in conjugative DNA processing that involve different sequence elements of the *nic* site. J Biol Chem 285: 8918-8926.

Lujan SA, Guogas LM, Ragonese H, Matson SW, Redinbo MR (2007) Disrupting antibiotic resistance propagation by inhibiting the conjugative DNA relaxase. Proc Natl Acad Sci U S A. 104:12282-7

Marchler-Bauer A, Zheng C, Chitsaz F, Derbyshire MK, Geer LY et al. (2013) CDD: conserved domains and protein three-dimensional structure. Nucleic Acids Res.41:D348-52.

McInerney JO, Pisani D, Bapteste E, O'Connell MJ (2011) The public goods hypothesis for the evolution of life on earth. Biol Direct. 6:41

McDaniel LD, Young E, Delaney J, Ruhnau F, Ritchie KB, Paul JH (2010) High frequency of horizontal gene transfer in the oceans. Science. 330: 50.

Monzingo AF, Ozburn A, Xia S, Meyer RJ, Robertus JD (2007) The structure of the minimal relaxase domain of MobA at 2.1 Å resolution. J Mol Biol 366: 165-178.

Piggott MJ, Attwood PV (2013) Post-translational modifications: Panning for phosphohistidine. Nat Chem Biol. 9:411-2.

Popa O, Dagan T (2011) Trends and barriers to lateral gene transfer in prokaryotes. Curr Opin Microbiol. 14:615-23.

Potts RG, Habibi S, Cheng Y, Lujan SA, Redinbo MR (2010) The mechanism and control of DNA transfer by the conjugative relaxase of resistance plasmid pCU1. Nucl Acids Res doi:10.1093/nar/gkq303: 1-15.

Priebe SD, Lacks SA (1989) Region of the streptococcal plasmid pMV158 required for conjugative mobilization. J Bacteriol 171: 4778-4784.

Puttick J, Baker EN, Delbaere LT (2008) Histidine phosphorylation in biological systems. Biochimica et biophysica acta 1784(1): 100-105.

Rasko DA, Sperandio V (2010) Anti-virulence strategies to combat bacteria-mediated disease. Nat Rev Drug Discov. 9:117-28.

Read RJ (2001) Pushing the boundaries of molecular replacement with maximum likelihood. Acta Crystallogr D Biol Crystallogr 57(Pt 10): 1373-1382.

Redrejo-Rodríguez M, Muñoz-Espín D, Holguera I, Mencía M, Salas M (2012) Functional eukaryotic nuclear localization signals are widespread in terminal proteins of bacteriophages.Proc Natl Acad Sci U S A. 109:18482-7.

Riley DR, Sieber KB, Robinson KM, White JR, Ganesan A, Nourbakhsh S, Dunning Hotopp JC (2013) Bacteria-human somatic cell lateral gene transfer is enriched in cancer samples. PLoS Comput Biol 9: e1003107.

Robinson VL, Stock AM. (1999) High energy exchange: proteins that make or break phosphoramidate bonds. Structure 7: R47-53.

Ros VI, Hurst GD (2009) Lateral gene transfer between prokaryotes and multicellular eukaryotes: ongoing and significant? BMC Biol. 7:20

Russi S, Boer R, Coll M (2008) Molecular Machinery for DNA Translocation in Bacterial Conjugation. In: Lipps G, editor. Plasmids: Current Research and Future Trends. Bayreuth (Germany): Caister Academic Press. pp. 183-213.

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Scherzinger E, Lurz R, Otto S, Dobrinski B (1992) In vitro cleavage of double- and single-stranded DNA by plasmid RSF1010-encoded mobilization proteins. Nucl Acids Res 20: 41-48.

Schneider TR, Sheldrick GM (2002) Substructure solution with SHELXD. Acta Crystallogr D Biol Crystallogr 58(Pt 10 Pt 2): 1772-1779.

Schroder G, 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.

Schröder G, Schuelein R, Quebatte M, Dehio C (2011) Conjugative DNA transfer into human cells by the VirB/VirD4 type IV secretion system of the bacterial pathogen Bartonella henselae. Proc Natl Acad Sci USA 108:14643-14648.

Slonczewski JL, Fujisawa M, Dopson M, Krulwich TA (2009) Cytoplasmic pH measurement and homeostasis in bacteria and archaea. Adv Microb Physiol 55: 1-79.

Smillie C, Garcillán-Barcia MP, Francia MV, Rocha EP, de la Cruz F (2010) Mobility of Plasmids. Microbiol Mol Biol Rev 74: 434–452.

Smillie CS, Smith MB, Friedman J, Cordero OX, David LA, Alm EJ (2011) Ecology drives a global network of gene exchange connecting the human microbiome. Nature 480:241-4.

Snyder LA, McGowan S, Rogers M, Duro E, O'Farrell E, Saunders NJ (2007) The repertoire of minimal mobile elements in the Neisseria species and evidence that these are involved in horizontal gene transfer in other bacteria. Mol Biol Evol. 24:2802-15.

Stern JC, Schildbach JF (2001) DNA recognition by F Factor TraI36: Highly sequence-specific binding of single-stranded DNA.Biochemistry, 40, 11586–11595.

Studier FW, Rosenberg AH, Dunn JJ, Dubendorff JW (1990) Use of T7 RNA polymerase to direct expression of cloned genes. Meth Enzymol 185: 60-89.

Szpirer CY, Faelen M, Couturier M (2001) Mobilization function of the pBHR1 plasmid, a derivative of the broad-host-range plasmid pBBR1. J Bacteriol 183: 2101-2110.

Terwilliger T (2004) SOLVE and RESOLVE: automated structure solution, density modification and model building. J Synchrotron Radiat 11(Pt 1): 49-52.

Thomas CM, Nielsen KM (2005) Mechanisms of, and barriers to, horizontal gene transfer between bacteria. Nat Rev Microbiol. 3:711-21.

Treangen TJ, Rocha EP (2011) Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. PLoS Genet. 7:e1001284.

Uhlemann AC, Dordel J, Knox JR, Raven KE, Parkhill J et al. (2014) Molecular tracing of the emergence, diversification, and transmission of S. aureus sequence type 8 in a New York community. Proc Natl Acad Sci U S A 111:6738-43

Uson I, Sheldrick GM (1999) Advances in direct methods for protein crystallography. Curr Opin Struct Biol 9(5): 643-648.

Vedantam G, Knopf S, Hecht DW (2006) *Bacteroides fragilis* mobilizable transposon Tn*5520* requires a 71 base pair origin of transfer sequence and a single mobilization protein for relaxosome formation during conjugation. Mol Microbiol 59(1): 288-300.

Waldor MK (2010) Mobilizable genomic islands: going mobile with oriT mimicry. Mol Microbiol. 78:537-40.

Wallace AC, Laskowski RA, Thornton JM (1995) LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. Protein Eng 8(2): 127-134.

Williams SL, Schildbach JF (2006) Examination of an inverted repeat within the F factor origin of transfer: context dependence of F TraI relaxase DNA specificity. Nucl Acids Res 34(2): 426-435.

WHO (2014) Antimicrobial resistance: global report on surveillance 2014, http://www.who.int/drugresistance/documents/surveillancereport/en/

Wozniak RAF, Waldor MK (2010) Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. Nat Rev Microbiol. 8:552-63.

Uhlemann AC, Dordel J, Knox JR, Raven KE, Parkhill J et al. (2014) Molecular tracing of the emergence, diversification, and transmission of S. aureus sequence type 8 in a New York community. Proc Natl Acad Sci U S A 111:6738-43

Yang W (2008) An equivalent metal ion in one- and two-metal-ion catalysis. Nat Struct Mol Biol.15:1228-31

Yang W (2011) Nucleases: diversity of structure, function and mechanism. Q Rev Biophys. 44:1-93.

Zhaxybayeva O, Doolittle WF (2011) Lateral gene transfer. Curr Biol. 21:R242-6.