Dissertation

submitted to the

Combined Faculties of Natural Sciences and Mathematics

of the Ruperto Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

presented by

M.Sc., Carlos Alberto Rojas Cordova

born in Lima, Peru

Oral examination: 23rd February 2021

STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF CONJUGATIVE TRANSPOSITION FOR THE DESIGN OF NOVEL ANTIBIOTIC RESISTANCE TRANSFER INHIBITORS

Referees: Dr. Christoph W. Müller

Prof. Dr. Matthias P. Mayer

SUMMARY

Antibiotic resistance (ABR) is currently one of the most significant global health challenges. In addition to the rapid development of resistance against new antibiotics, the transfer of existing ABR genes between bacteria leads to the growth of difficult-to-treat multidrug-resistant opportunistic pathogens, causing millions of infections and thousands of deaths every year worldwide. Mobile genetic elements (MGEs) provide a powerful mechanism to transfer ABR genes, because they can move across bacterial cells and species and carry ABR cargos within their sequence. However, their mechanisms of transfer are incompletely understood. Therefore, I investigated the molecular mechanism of a prominent but poorly characterized MGE, the vancomycin resistance carrying conjugative transposon (CTn) Tn*1549* from *Enterococcus spp*. This element is responsible for propagating resistance to this last-resort antibiotic in a wide range of intestinal bacteria. I focused my work on the integrase (Int) and excisionase (Xis) proteins, which are responsible for performing all DNA cleavage and joining reactions during Tn*1549* transposition.

In the first part, I reconstituted the complex of Int with a four-way Holliday Junction (HJ) DNA molecule *in vitro* and solved its crystal structure at 3.3 Å resolution. This is the first CTn integrase structure trapped with this reaction intermediate, showing that these enzymes assemble a stable tetramer to recombine two DNA substrates, in a similar way as the site-specific tyrosine recombinases. Comparison of both enzyme families shows that a cyclic exchange of the C-terminal protein segments promotes tetramerization in all cases. I further characterized the structure of the accessory Xis protein at 1.5 Å resolution. In the second part of my work, I validated the structure by performing HJ resolution experiments *in vitro.* I found that Int can resolve HJ intermediates both to products and to substrates, likely due to missing regulatory factors. This reaction leads to DNA products containing up to 3 nt long unpaired regions, reflecting Int's ability to insert its cargo DNA at diverse genomic sequences. In the third part of my work, I show that novobiocin, an aminocoumarin antibiotic, can inhibit Int tetramerization and HJ resolution *in vitro*, highlighting the importance of the tetrameric state for the transposition reaction.

This work sheds light on an essential step of Tn*1549* transposition and its regulation. Moreover, it highlights crucial similarities with site-specific recombinases, increasing our understanding of conjugative transposition. As a proof-of-principle, inhibition of Int activity by novobiocin may open doors to develop potent CTn inhibitors as a new strategy to limit ABR spreading among bacteria.

ZUSAMMENFASSUNG

Antibiotikaresistenz (ABR) ist heutzutage eine der größten globalen Herausforderungen im Gesundheitswesen. Neben der raschen Entwicklung von Resistenzen gegen neue Antibiotika, der Transfer bestehender ABR-Gene zwischen Bakterien führt zum Wachstum multiresistenter opportunistischen Krankheitserreger, die schwer zu behandeln sind. Solche Erreger verursachen Millionen von Infektionen und führen zum Tod tausender Menschen jedes Jahr weltweit. Mobile genetische Elemente (MGEs) bieten einen leistungsfähigen Mechanismus zur Übertragung von ABR-Genen, da sie sich zwischen Bakterienzellen und unterschiedlichen Bakterienarten hinwegbewegen können und ABR-Ladungen in ihrer Sequenz tragen können. Diese Übertragungsmechanismen sind jedoch bis heute unvollständig verstanden.

Aus diesem Grund untersuchte ich den molekularen Mechanismus eines prominenten, aber wenig charakterisierten MGE. Ein aus Enterokokken stammenden Konjugierenden Transposon (CTn) Tn*1549*, welches Vancomycin Resistenzgene in sich trägt. Dieses Element ist für die Ausbreitung der Resistenz gegen dieses Reserveantibiotikum in einer Vielzahl von Darmbakterien verantwortlich. Ich konzentrierte meine Arbeit auf die Proteine, Integrase (Int) und Excisionase (Xis), welche für die Durchführung aller DNA-Spaltungs- und Verbindungsreaktionen während der Tn*1549* Transposition verantwortlich sind.

Im ersten Teil meiner Arbeit rekonstituierte ich einen Komplex, bestehend aus Int mit DNA, in ihrer „Holliday-Junction“ Struktur (HJ), *in vitro* und bestimmte seine Kristallstruktur bei 3.3 Å Auflösung. Dies ist die erste CTn-Integrase Struktur, welche in diesem Reaktionszwichenprodukt eingefangen wurde. Es zeigt, dass diese Enzyme einen stabilen Tetramer-Komplex zusammensetzen können, um zwei DNA-Substrate zu rekombinieren und dass diese Reaktion auf ähnlicher Weise wie bei den ortsspezifischen Tyrosin-Rekombinasen abläuft. Ein Vergleich beider Enzymfamilien zeigt, dass ein zyklischer Austausch der C-terminalen Proteinsegmente in allen Fällen die Tetramerisierung fördert. Ferner charakterisierte ich die Struktur des akzessorischen Xis-Proteins bei einer Auflösung von 1,5 Å.

Im zweiten Teil meiner Arbeit validierte ich die Struktur durch HJ-Auflösungsexperimente *in vitro*. Ich fand heraus, dass Int HJ-Zwischenprodukte sowohl zu Produkten als auch zu Substraten rekombinieren kann, wahrscheinlich aufgrund fehlender regulatorischer Faktoren. Diese Reaktion führt zu DNA-Produkten, die bis zu 3 nt lange ungepaarte

Regionen enthalten, was die Fähigkeit von Int widerspiegelt, seine Fracht-DNA an verschiedenen genomischen Sequenzen zu integrieren.

Im dritten Teil meiner Arbeit zeige ich, dass Novobiocin, ein Aminocumarin-Antibiotikum, die Int-Tetramerisierung und die HJ-Auflösung *in vitro* hemmen kann, was die Bedeutung des Tetramer-Zustands für die Transpositionsreaktion hervorhebt.

Diese Arbeit beleuchtet einen wesentlichen Schritt der Tn*1549*-Transposition und ihrer Regulation. Darüber hinaus werden entscheidende Ähnlichkeiten mit ortsspezifischen Rekombinasen hervorgehoben, wodurch unser Verständnis der konjugierenden Transposition verbessert wird. Als Grundsatzbeweis kann die Hemmung der Int-Aktivität durch Novobiocin Türen öffnen, um wirksame CTn-Inhibitoren als neue Strategie zur Begrenzung der ABR-Ausbreitung unter Bakterien zu entwickeln.

ACKNOWLEDGEMENTS

First of all, I would like to especially thank Dr. Orsolya Barabas for giving me the opportunity to start and do my PhD project in her lab. I would like to thank you for introducing me into the world of DNA transposons and X-ray crystallography. With your endlessly knowledge in the structural and transposon field, you were always able to support me during difficult times of my research. Moreover, you also had ears for problems outside the lab environment and I am very thankful for your support in difficult times, especially during this year.

I would also like to thank my Thesis Advisory Committee, Prof. Dr. Matthias Mayer, Dr. Christoph Müller, and Dr. Nassos Typas, for their time and invaluable support during our annual meetings.

I want to thank the EMBL PhD program for the fellowship that supported my studies, and the EMBL Graduate Office for its constant support and for offering that many courses, especially this year. I should also thank the EMBL Core Facilities for their service and support during the years. Especially I wish to thank you Brice, for your support in this four-year long crystallization journey. I appreciate a lot that you always had an open ear and advice in the never-ending possibilities of crystallization trials. Merci beaucoup. In relation to the crystallization part of my work, I also wish to thank the ESRF and DESY staff for their assistance during data collection at the synchrotron beamlines.

I would like to express my immense gratitude to former and present members of the Barabas Lab. Each of you provided a unique flavor to the work environment, making it a very enjoyable place to do research. I would like to especially thank Buse for simply being there in the lab. It was nice not to be the only new PhD student in the lab. Thank you also for sharing the ups and downs that we encountered during our PhD student's life. I also want to really thank Gera for sharing his knowledge and support in questions related to bioinformatics and for all the discussions related to bacterial transposons and of course also for introducing me to the EMBL Beer Sessions. I would like to thank Vladimir for his perfectionism and for sharing his crystallographic expertise with me. Special thanks go to Ceci. Thanks to you, the lab was always kept in order; but also, for sharing your knowledge on many techniques that helped me save valuable time. To my students Marius and Robert, thank you for your interest on my project and for the hard work. I also want to thank Anna, for teaching me many techniques I needed later during my PhD and for helping me at the start of my PhD project; Lotte for giving me specific advice in questions related also to my

PhD Project; Natalia for always being there for a coffee and your kindness; Irma for your directness and for always sharing these delicious cookies with us.
My sincere thanks go to Drew for collaborating with me and helping to expand the last section of this thesis. Without it, this story would feel incomplete.
What would be EMBL without the possibilities it offers to meet awesome and diverse people. I met and did new friends here, hopefully for life, that I would also like to thank, simply for spending time with me and being there during good and bad times. I would like to thank Dani and Lalo for their hospitality and the funny moments we spent together; Vilma and Diana for always organizing dinners and brunches so that we don't lose sight from each other. To my Predoc course friends Samantha and Esther, thank you for sharing your culture and for being there from the very beginning; Paul, thank you for spending time in invaluable discussions about life and the future, and for your care in general. Muchas gracias a ti Hugo ☺. I want to thank you for your company and support during the last years and wish you all the best in China.

Finalmente, me gustaria agradecerle a mi familia por su apoyo en general. Especialmente a ti mama, gracias por tu esfuerzo que nos saca adelante y por dejarme tomar mis propias deciciones pero siempre teniendo buenos consejos.
Finally, I would like to thank my family for their support. My mother, thank you for your hard work and letting me decide and take my own path in life.

Insbesondere widme ich diese Arbeit Franz und Frau Fiedler. Auch wenn ihr nicht mehr bei uns seid, ohne euch, hätte ich es nicht bis hierhergeschafft. Ich danke euch von ganzem Herzen und werde euch immer in Erinnerung halten.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

aa	amino acid
AB	Arm-binding domain
ABR	antibiotic resistance
Amp	Ampicillin
AMR	Antimicrobial resistance
APS	Ammonium persulfate
Arg/R	Arginine
Asn/N	Asparagine
Asp/D	Aspartic acid
ATP	Adenosine triphosphate
bp	Base pair
C-terminus	Carboxy terminus
CAT	Catalytic domain
CB	Core binding domain
CI	Circular intermediate
Cm	Chloramphenicol
CTn	Conjugative transposon
Da	Dalton
DBD	DNA binding domain
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DNAse	Deoxyribonuclease
DR	Direct repeat
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EMBL	European Molecular Biology Laboratory
ESRF	European Synchrotron Radiation Facility
FDA	Food and drug administration
for	forward
FT	Fourier transform
Glu/E	Glutamic acid
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HGT	Horizontal gene transfer
His/H	Histidine
HIV	Human immunodeficiency virus
HJ	Holliday Junction
HTH	Helix-Turn-Helix
ICE	Integrative and conjugative element
IDSA	Infectious Diseases Society of America
IDT	Integrated DNA Technologies
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Int	Integrase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IR	Inverted repeat
IR_L	Inverted repeat left
IR_R	Inverted repeat right
ITC	Isothermal Titration Calorimetry
kbp	Kilobase pair
K_D	Dissociation constant
keV	Kiloelectron volt
Km	Kanamycin
LA	Left end arm
LB	Lysogeny Broth
LE	Left end
LLG	Log-likelihood gain
LSPPs	Lincosamides, streptogramins, phenicols, and pleuromutilins
Lys/K	Lysine
Met/M	Methionine

MDR	Multidrug-resistant
MGE	Mobile genetic element
Mw	Molecular weight
MR	Molecular replacement
MRSA	Methicillin-resistant *Staphylococcus aureus*
M.Sc.	Master of Science
NanoDSF	Nano Differential Scanning Fluorimetry
N-terminus	Amino-terminus
NCBI	The National Center for Biotechnology Information
NEB	New England Biolabs
OD_{600}	Optical density at 600 nm
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PBPs	Penicillin-binding-proteins
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PEG	Polyethylene glycol
pI	Isoelectric point
PMSF	Phenylmethanesulphonylfluoride
PTC	Peptidyl transferase center
RA	Right end arm
RE	Right end
rev	reverse
RF	Restriction free (cloning)
RNA	Ribonucleic acid
RNase	Ribonuclease
rpm	round per minute
SDS	Sodium dodecyl sulfate
SEC	Size exclusion chromatography
SeMet	Selenomethionine
Ser/S	Serine
SOB	Super Optimal Broth
spp.	several species
T	Target site
TAE	Tris-acetate-EDTA
TBE	Tris-borate-EDTA
TCEP	Tris (2-carboxyethyl) phosphine
TE	Transposable element
TE buffer	Tris-EDTA buffer
TEMED	Tetramethylethylenediamine
Tet	Tetracycline
T_m	Melting temperature
Tris	Tris(hydroxymethyl)aminomethane
Trp/W	Tryptophan
TSD	Target site duplication
Tyr/Y	Tyrosine
US	United States
USD	United States Dollar
UV	Ultraviolet
VR	Vancomycin resistance
VRE	Vancomycin-resistant enterococci
WHO	World Health Organization
XDR	Extremely-drug resistant
Xis	Excisionase

1 INTRODUCTION

1.1 The antibiotic resistance challenge

Antibiotics are chemical compounds used to treat bacterial infections in humans and animals in healthcare and agriculture. Their mode of action or antimicrobial effect is based on inhibiting bacteria's growth or killing these microorganisms. They can be of synthetic or natural origin (Davies, 2010; Peterson and Kaur, 2018). Antimicrobials of synthetic origin were discovered first: Salvarsan (marketed by Hoechst) was developed against syphilis by P. Ehrlich and co-workers, and later sulphonamides proved to be effective against a variety of bacterial and parasitic diseases (Aminov, 2010). After the discovery of penicillin in 1928 by A. Fleming (Fleming, 1929) and streptomycin in 1943 by S. Waksman, which led to a breakthrough in the treatment of tuberculosis (Waksman, 1952), these natural compounds launched the golden age of antibiotic discovery from living organisms (microbes like bacteria or fungi) (Walsh, 2000). However, bacteria can naturally develop resistance against antibiotics, for example through point mutations, selection pressure, and the acquisition of mobile genetic elements (MGEs). The misuse of antibiotics in healthcare and agriculture can help accelerate this process (WHO, 2020).

For this reason, antibiotics can become ineffective and there is the need for the steady development of new ones. However, the development of new antibiotics has enormously decreased and the number of multidrug-resistant (MDR) ‘superbugs’ has increased over time, creating the perfect storm (Cooper and Shlaes, 2011). As a consequence, antibiotic resistance (ABR) is currently one of the most significant global health-care challenges. It spreads rapidly, resulting in the continuous emergence of highly virulent pathogens and MDR ‘superbugs’ (WHO, 2019). Nowadays, we are already losing up to 700 thousand human lives yearly due to drug-resistant diseases. Furthermore, it is predicted that death rates may approach 10 million people annually by 2050, surpassing cancer death numbers. Apart from that, many secondary health effects add up to the life costs: cancer therapies (chemotherapy), medical interventions like cesarean sections and organ transplantations become more dangerous due to an increased risk of untreatable bacterial infections. All this is predicted to cost the world economy up to 100 trillion USD from now until 2050 (Jim O’Neill, 2014; Sugden, Kelly, and Davies, 2016).

To respond to this increasing threat, the World Health Organization (WHO) has developed an action plan to tackle antimicrobial resistance (AMR) in bacteria as well as in parasites

(malaria), viruses (human immunodeficiency virus, HIV), and fungi, which can also develop drug-resistance. In short, the plan consists of investing into research and development of new drugs, but also into dissemination prevention and global education to tackle the problem world-wide (WHO, 2015). In conclusion, it is essential to find new ways to delay or completely stop ABR spreading in the next years to come.

While AMR covers the full range of species that can develop resistance against a specific drug, the term ABR is explicitly used for bacteria. I will continue to use the term ABR in this thesis, because the presented work focuses specifically on antibiotic resistance spreading among bacteria.

One of the biggest problems today is health-care-associated infections caused by multidrug- or even pan-resistant bacteria. These bugs are not only resistant to one, but often to many antibiotic classes, and in some cases, even to all antibiotic classes (called pan-resistance). The most important and dangerous pathogens were classified by the Infectious Diseases Society of America (IDSA) in the USA as the ESKAPE group: *Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Acinetobacter baumannii*, *Pseudomonas aeruginosa,* and *Enterobacter* species (Rice, 2008). Recently, the WHO went a step further and classified the most dangerous pathogens worldwide to have a global priority list and help guide future research and development of antibiotics. They identified and classified the most important drug-resistant bacteria according to the species and type of resistance and ordered them into three priority tiers: Priority 1 (Critical) – *Acinetobacter baumannii* (carbapenem-resistant), *Pseudomonas aeruginosa* (carbapenem-resistant) and *Enterobacteriaceae* [including *Klebsiella pneumonia*, *Escherichia coli*, *Enterobacter spp.*, *Serratia* spp., *Proteus spp.*, *Providencia spp.*, *Morganella spp.*] (carbapenem-resistant, third-generation cephalosporin-resistant); Priority 2 (High) - *Enterococcus faecium* (vancomycin-resistant), *Staphylococcus aureus* (methicillin-resistant, vancomycin-intermediate and resistant), *Helicobacter pylori* (clarithromycin-resistant), *Campylobacter* (fluoroquinolone-resistant), *Salmonella spp.* (fluoroquinolone-resistant) and *Neisseria gonorrhoeae* (third-generation cephalosporin-resistant, fluoroquinolone-resistant); Priority 3 (Medium) - *Streptococcus pneumoniae* (penicillin-non-susceptible), *Haemophilus influenzae* (ampicillin-resistant) and *Shigella spp.* (fluoroquinolone-resistant) (Tacconelli et al., 2017). This updated list should help researchers worldwide to focus their efforts on developing new drugs and on understanding fundamental questions related to the most critical threats in antibiotic resistance spreading.

In the following paragraphs, I will give a short introduction to the most important antibiotics, the resistance mechanisms that bacterial pathogens use against them, the pathways involved in resistance spreading, and the role of MGEs in disseminating resistance genes among bacteria. Thereby, I will focus on a specific MGE class, the conjugative transposons (CTns, also called integrative and conjugative elements – ICEs) and the model element Tn*1549*, which plays a major role in vancomycin resistance spreading among Gram-positive bacteria. Tn*1549* has been studied in the Barabas lab as a model CTn over the last years (Lambertsen et al., 2018; Rubio-Cosials et al., 2018), but many questions related to this MGE class are still open. Thus, I decided to pursue my doctoral studies in the lab and continue studying this CTn to tackle the ABR spreading problem.

1.2 Antibiotics and resistance

The word "antibiotic" was coined by S. Waksman, the discoverer of streptomycin, to describe the specific effect and activity of a natural or chemical compound. Such molecules can either inhibit growth or kill bacterial cells (Waksman and Flynn, 1973). With the pass of the time, certain antibiotics have been repurposed for antiviral, antitumor, or anticancer treatments. Thus, the term antibiotic should not be used in such cases (Demain and Sanchez, 2009). Antibiotic discovery and resistance emergence are going hand in hand in the course of history. Sulfonamides were introduced in 1935, and resistance against this antibiotic was reported a few years later (WM and LA, 1942). After discovering penicillin in 1928, a penicillinase enzyme was spotted a few years later, even before introducing the antibiotic into the clinic (Frère, Sauvage, and Kerff, 2015). Another example is streptomycin, which was introduced in 1944 for tuberculosis treatment, and shortly after *M. tuberculosis* strains resistant against this drug appeared (Musser, 1995).

Bacteria can acquire resistance through point mutations or by sequestering a gene that confers resistance (i.e. resistance genes) through horizontal gene transfer (HGT). Moreover, they can be intrinsically resistant to certain antibiotic classes. For example, Gram-negative bacteria are resistant to glycopeptides because of their specific cell-wall architecture (Zeng et al., 2016). Antibiotic-producing bacteria, such as *Streptomyces*, show self-resistance against their own antibiotic compounds (Peterson and Kaur, 2018). Apart from that, bacteria can show tolerance or persistence to high antibiotic concentrations by slowing down their own growth or by protecting themselves via biofilm formation (Brauner et al., 2016).

In the next section, I will summarize the most critical antibiotic types and their mode of action, explain the resistance mechanisms against these compounds and describe one of the most urgent MDR pathogens in Gram-positive bacteria, which carries the CTn Tn*1549* that I chose as a model system for my doctoral studies.

1.2.1 Antibiotic's mode of action

Most antibacterial drugs have one of three common targets: (1) the bacterial protein synthesis pathway, (2) DNA replication and repair pathways and (3) the bacterial cell-wall biosynthesis pathway (Walsh, 2000). A list of the most common antibiotic classes, their molecular targets and typical examples are provided in Table 1-1 (Coates, Halls, and Hu, 2011).

1.2.1.1 Targeting the protein synthesis pathway

The ribosome is the main target for many antibiotic classes (Arenz and Wilson, 2016). For example, macrolides, lincosamides, streptogramins, phenicols, and pleuromutilins, and oxazolidinones interact at similar locations with the 23S rRNA and related proteins in the peptidyl transferase center in the 50S ribosome subunit, sterically blocking peptide transfer and inhibiting the elongation process (Walsh, 2000; Fyfe et al., 2016; Schwarz et al., 2016). Aminoglycosides inhibit translocation during translation elongation through interactions with the 16S rRNA in the 30S subunit (Moazed and Noller, 1987; Houghton et al., 2010). Tetracyclines inhibit protein biosynthesis by hindering the aminoacyl-tRNA positioning at the ribosome's entry site through binding to the 30S subunit (Chopra and Roberts, 2001).

1.2.1.2 Targeting the DNA replication and repair pathway

Quinolones and fluoroquinolones can have a bactericidal effect on Gram-negative and Gram-positive bacteria by blocking DNA replication and repair pathways indirectly (Table 1.1). They interact with two different bacterial topoisomerase II family enzymes, DNA gyrase and topoisomerase IV. Quinolone binding traps these enzymes at an intermediate step of their reaction, covalently bound to DNA after double-strand cleavage has occurred. This trapped complex inhibits the repair of the DNA double-strand breaks and hinders the DNA replication or transcriptional machinery to move forward, leading to permanent DNA double-strand breaks and induction of an SOS-response. If the DNA repair machinery cannot cope with the DNA damage, this will lead to bacterial cell death (Hooper and Jacoby, 2016).

1.2.1.3 Targeting the cell-wall biosynthesis pathway

Several antibiotic classes target the bacterial cell-wall biosynthesis pathway. Most heavily used are the β-lactam antibiotics, which can be employed against both Gram-negative and Gram-positive bacteria. Their β-lactam ring interacts with transpeptidase enzymes (also called penicillin-binding proteins, PBPs), which are necessary for crosslinking the peptide chains in the cell wall's peptidoglycan layer. These antibiotics serve as pseudosubstrates (Tipper and Strominger, 1965) that get covalently bound through acylation of a serine residue in the transpeptidase active site (Georgopapadakou, Hammarstrom and Strominger, 1977). This very slowly reversible reaction prevents the protein to function correctly. Without a suitable fraction of covalently linked peptide chains, the cell wall loses strength and becomes more prone to osmotic lysis. Thus, β-lactam antibiotics have a bactericidal effect (Walsh, 2000; Bush and Bradford, 2016).

Another antibiotic that targets the bacterial cell-wall is the lipopeptide, daptomycin. Its mode of action is still not fully elucidated. Nevertheless, it is proposed that it binds and disrupts the cell wall of Gram-positive bacteria, leading to cell lysis and death. Daptomycin is currently used as a last resort antibiotic against MRSA and vancomycin-resistant *Enterococci* (VRE) infections (Miller, Bayer, and Arias, 2016).

The polymyxins – polymyxin B and colistin (polymyxin E) – were used before the 1980s. Later on, their usage was stopped due to their high neural and renal toxicity. A renaissance of this drug class started in the last decade due to the expansion of MDR bacteria, and now it is being used as last resort treatment despite the side effects. The exact mechanism of polymyxins is still under discussion, but the widely accepted mode of action implies targeting of the cytoplasmic cell membrane of Gram-negative bacteria leading to its disruption or permeabilization and bacterial cell death as the outcome (Trimble et al., 2016).

The glycopeptide antibiotics target the cell-wall biosynthesis in Gram-positive bacteria (Figure 1-2B). However, instead of mimicking the peptide substrate, they interact with the peptide itself through hydrogen-bond interactions, by binding at the D-alanyl-D-alanine termini of the lipid II monomer and inhibit its interaction to the transpeptidase (substrate sequestration) and the transglycosylase (steric hindrance). This inhibitory mechanism also leads to the same outcome as for the β-lactam antibiotics: The peptidoglycan layer loses mechanical strength due to fewer crosslinks, resulting in osmotic lysis and usually bacterial cell death (Barna and Williams, 1984; Zeng *et al.*, 2016).

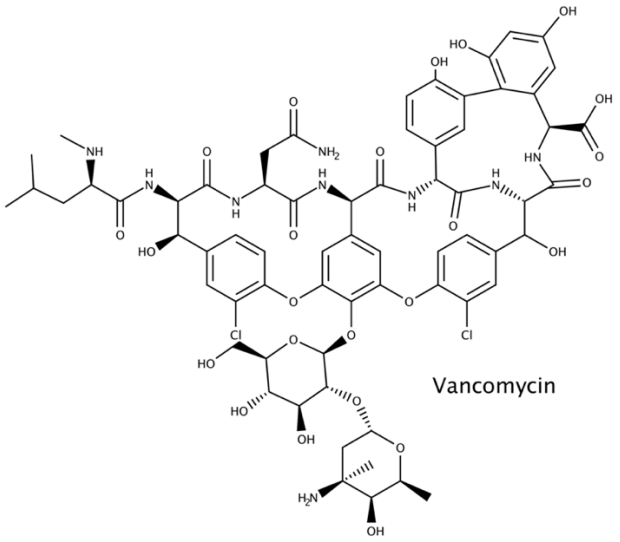

Figure 1-1: Structure of the glycopeptide antibiotic vancomycin (Barna and Williams, 1984).

Vancomycin (Figure 1-1) is one of the most prominent examples of glycopeptide antibiotics. It was discovered in 1953 from *Amycolatopsis orientalis* by E.C Kornfeld (Griffith, 1981). Its clinical use got approved by the FDA already in 1955 to treat penicillin-resistant Staphylococci. Its structural core is composed of a heptapeptide with five aromatic residues that are heavily cross-linked. Moreover, the peptide is glycosylated at residue four with a disaccharide D-glucosyl-2,1-D-vancosamine (Barna and Williams, 1984). It is used as a last-resort antibiotic against severe MDR infections, including drug-resistant *Enterococcus*, *Clostridium* and methicillin-resistant *S. aureus* (MRSA) (Rubinstein and Keynan, 2014). Vancomycin is mostly given intravenously due to is hydrophilic character and large size that makes it difficult to penetrate the intestinal barrier. Early clinical data showed drug-induced nephrotoxicity (toxic to the kidney) and ototoxicity (toxic to the ear) that turned out to be related to impurities and has been overcome through an improved production and purification procedure (Kahne *et al.*, 2005; Rubinstein and Keynan, 2014; Zeng *et al.*, 2016).

1.2.2 Mechanisms of antibiotic resistance in bacteria

Bacteria seem to be very good at finding ways of how to protect themselves against antibiotics or to make them ineffective. It builds on i) their ability to accumulate genomic mutations over time that at some time point confer resistance to a specific antibiotic, allowing the bacteria to grow in the particular environment; and ii) the ability to exchange genomic information (including ABR genes) through HGT (explained later in detail in

section 1.3.1). Below, I will discuss the main antibiotic resistance mechanisms: drug permeability and efflux, drug degradation or modification, and target modification or protection (Walsh, 2000; Peterson and Kaur, 2018).

1.2.2.1 Drug permeability and efflux

Bacteria can be intrinsically resistant to specific classes of antibiotics. Specially Gram-negative bacteria with their second cell membrane, are more difficult to penetrate by some drugs, such as glycopeptides and lipoglycopeptides (Zeng et al., 2016). Another type of intrinsic resistance can be gained by expressing efflux pumps that can transport bacterial drugs as a secondary effect (so-called gene "repurposing"). Usually, these pumps perform the transport of heavy metals out of the cell and are also needed for nodulation and cell division. In pathogenic isolates, however, overexpression of such genes leads to antibiotic resistance due to their ability to transport various antibiotics out of the cell, for example, in *E. coli*, *P. aeruginosa*, *S. typhimurium,* among others (Dantas and Sommer, 2012).

Another example is quinolone resistance. Active efflux leads to low-level resistance in Gram-positive bacteria. While in Gram-negatives, reduced cell permeability and active efflux can provide low to high levels of quinolone resistance (Hooper and Jacoby, 2016).

1.2.2.2 Drug degradation or modification

Antibiotic modification is a common strategy used by bacteria to render drugs ineffective. For example, aminoglycoside resistance is achieved mainly through antibiotic modification, both in Gram-positive and in Gram-negative bacteria. For that, different types of aminoglycoside modifying enzymes are used, with N-acetyltransferases, O-phosphotransferases and O-adenyltransferases representing three big subfamilies (Houghton *et al.*, 2010).

β-lactamases are another large protein family that can degrade antibiotics. In Gram-negative bacteria, the most common resistance mechanism is the hydrolysis of β-lactam antibiotics by β-lactamases. Nowadays, more than 2000 β-lactamases have been identified. These can be classified either in classes A-D based on protein sequence homology or in groups 1-4 based on the hydrolysis reaction they perform and on their respective substrates (Jacoby, 2006). The reaction mechanism is similar to the one between β-lactams with their target proteins (see section 1.2.1.3). However, after acylation, β-lactamases can hydrolyze the drug, rendering it inactive and, at the same time, regenerating themselves to perform the next hydrolysis reaction (Drawz and Bonomo, 2010).

Class	Example	Molecular Target	Active in
Aminoglycosides	streptomycin, neomycin, kanamycin, paromomycin, gentamicin, tobramycin, amikacin, netilmicin, spectinomycin, sisomicin, dibekacin, isepamicin	protein biosynthesis	Gram-(-) and (+)
β-Lactams	penicillin G, penicillin V, methicillin, oxacillin, cloxacillin, dicloxacillin, nafcillin, ampicillin, amoxicillin, carbenicillin, ticarcillin, mezlocillin, piperacillin, azlocillin, temocillin	cell-wall biosynthesis	Gram-(-) and (+)
β-Lactamase inhibitors	clavulanate, sulbactam, tazobactam, avibactam, vaborbactam	cell-wall biosynthesis	Gram-(-)
Carbapenems	imipenem, meropenem, ertapenem, doripenem	cell-wall biosynthesis	Gram-(-) and (+)
Cephalosporins	first-generation: cephalothin, cephapirin, cephradine, cephaloridine, cefazolin second generation: cefamandole, cefuroxime, cephalexin, cefprozil, cefaclor, loracarbef, cefoxitin, cefmetazole third generation: cefotaxime, ceftizoxime, ceftriaxone, cefoperazone, ceftazidime, cefixime, cefpodoxime, ceftibuten, cefdinir fourth generation: cefpirome, cefepime fifth-generation: ceftaroline fosamil, ceftobiprole, cefiderocol[#]	cell-wall biosynthesis	Gram-(-) and (+)
(Lipo-)Glycopeptides	vancomycin, teicoplanin, telavancin, oritavancin, dalbavancin	cell-wall biosynthesis	Gram-(+)
Ketolides	telithromycin	protein biosynthesis	Gram-(-) and (+)
Lincosamides	lincomycin, clindamycin	protein biosynthesis	Gram-(+)
Lipopeptides	daptomycin[#]	cell-wall biosynthesis	Gram-(+)
Macrolides	erythromycin, azithromycin, clarithromycin	protein biosynthesis	Gram-(-) and (+)
Monobactams	aztreonam	cell-wall biosynthesis	Gram-(-) and (+)
Oxazolidinones	linezolid	protein biosynthesis	Gram-(+)
Phenicols	chloramphenicol[$]	protein biosynthesis	Gram-(-) and (+)
Pleuromutilins	retapamulin[&]	protein biosynthesis	Gram-(+)
(Fluoro-)Quinolones	nalidixic acid, oxolinic acid, norfloxacin, pefloxacin, enoxacin, ofloxacin/levofloxacin, ciprofloxacin, temafloxacin, lomefloxacin, fleroxacin, grepafloxacin, sparfloxacin, trovafloxacin, clinafloxacin, gatifloxacin, moxifloxacin, sitafloxacin	DNA repair and replication pathway	Gram-(-) and (+)
Rifamycins	rifampicin, rifapentine, rifabutin, bezoxazinorifamycin, rifaximin	DNA transcription pathway	Gram-(-)[%] and (+)
Streptogramins	quinupristin[#], dalfopristin[#], pristinamycin	protein biosynthesis	Gram-(+)
Sulfonamides	sulphanilamide, *para*-aminobenzoic acid, sulfadiazine, sulfisoxazole, sulfamethoxazole, sulfathalidine	folate synthesis pathway	Gram-(-) and (+)
Tetracyclines	tetracycline, chlortetracycline, demeclocycline, minocycline, oxytetracycline, methacycline, doxycycline, tigecycline#, eravacycline[#]	protein biosynthesis	Gram-(-) and (+)
Cationic peptides/polymyxins	polymyxin B[#], colistin[#] (polymyxin E)	cytoplasmic cell membrane	Gram-(-)
Others	metronidazole, trimethoprim, fosfomycin (phosphonic acid),	DNA replication pathway folate synthesis pathway cell-wall biosynthesis	Gram-(-) and (+) Gram-(-) and (+) Gram-(-) and (+)

Table 1-1: List of antibiotics commonly used in the clinic, divided by classes (Coates, Halls, and Hu, 2011).#: last-resort antibiotic. $: Only in exceptional, extreme life-threatening cases, too many adverse effects. &: the first antibiotic of its type to be approved for humans. Limited to topical treatment of impetigo by MSSA. % only a few Gram-negative pathogens, Neisseria meningitides, N. gonorrhoeae, and Hemophilus influenzae.

While class A, C, and D enzymes utilize a serine nucleophile in their active sites and are structurally similar, class B β-lactamases need a metal-ion as a cofactor (like Zn^{2+}) for the hydrolysis reaction (also called Metallo-β-lactamases). Examples of clinically relevant enzymes include TEM, SHV, and CTX-M class A β-lactamases, which can be chromosomally-encoded or be found in plasmids (Bonomo, 2017; Peterson and Kaur, 2018).

1.2.2.3 Target modification or protection

On the other hand, Gram-positive bacteria frequently employ a target protection mechanism by utilizing different transpeptidase proteins with a lower binding affinity to β-lactam antibiotics. MRSA is the most prominent example. This bacterial pathogen is resistant to many penicillins, including methicillin and cephalosporins, because it has acquired the *mecA* gene, which encodes a different type of transpeptidase, the PBP2a enzyme. This protein has a much lower binding affinity to β-lactams than the usual PBPs and can perform its function also in presence of the antibiotic (Bonomo, 2017).

Quinolone resistance frequently arises through target modification. Point mutations in the GyrA or ParC subunits of gyrase or topoisomerase IV, respectively, can lead to a decrease in quinolone binding affinity to the protein-DNA complex. Quinolone resistance can happen in both Gram-negative and Gram-positive bacteria. The more point mutations bacteria accumulate in these specific enzymes over time, the stronger the resistance phenotype. Because inhibition of either protein can induce cell-death, resistant bacteria accumulate mutations in both proteins (Hooper and Jacoby, 2016). Moreover, plasmid-encoded resistance of quinolone was recently discovered. It is based on the *qnr* genes that confers resistance in two ways: by protecting GyrA from quinolone binding and by modification of specific quinolone drugs. These plasmids can additionally encode quinolone efflux pumps (Strahilevitz *et al.*, 2009).

1.2.2.3.1 Glycopeptide resistance

A different type of target modification, which leads to drug protection, is glycopeptide resistance (also called vancomycin resistance). In this case, binding of the glycopeptide, such as vancomycin, is reduced by the production of altered peptidoglycan precursors (Figure 1-2C). All genes necessary for vancomycin resistance are organized in operons, which can be encoded on the bacterial chromosome or carried on plasmids. The most prominent are the *vanA* and *vanB* gene clusters. Their functionality and organization are similar. Both resistance types encode a dehydrogenase (VanH), a ligase (VanA or VanB)

and a dipeptidase (VanX) that are conserved among them. On the other side, regulatory proteins like VanR and VanS show only low percentage identity (~30%). Another difference is that in the *vanB* type these regulatory proteins are induced by vancomycin but not by teicoplanin. The *vanA* type resistance shows induction by both antibiotics (Courvalin, 2006).

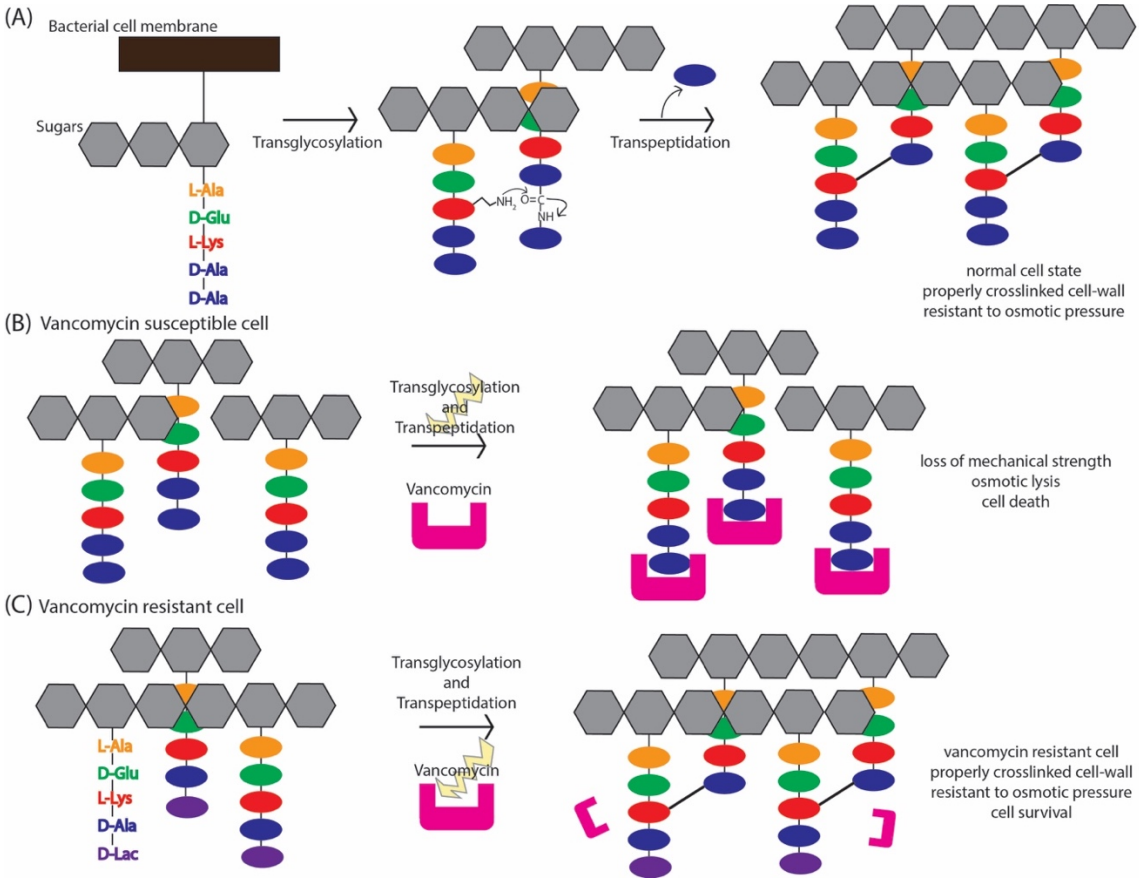

Figure 1-2: Bacterial cell-wall biosynthesis inhibition by vancomycin and its resistance mechanism. (A) Transglycosylation and transpeptidation will crosslink peptidoglycan precursors and render the cell-wall resistant to osmotic pressure and ensure cell survival. (B) In the presence of vancomycin (magenta), the transglycosylation and transpeptidation step will be inhibited by coverage of the D-ala-D-ala ends (blue) through vancomycin interaction, leading to a cell-wall that loses mechanical strength and is prone to osmotic lysis, what leads to cell death. (C) Vancomycin-resistant bacterial cells can produce peptidoglycan precursors that end in D-ala-D-lac (purple) (for example in vanB-type) instead of D-ala-D-ala. Vancomycin and similar antibiotics will have a lower binding affinity to this modified peptidoglycan ends, rendering them ineffective; ensuring proper peptidoglycan crosslinking and cell survival (Walsh, 2000; Courvalin, 2006).

The resistance mechanism is based on production of peptidoglycan precursors with new ends at the carboxy-terminus (D-alanyl-D-lactate or D-alanyl-D-serine). In the case of *vanB*-type resistance, VanH reduces pyruvate to D-lac, which is ligated afterwards to D-ala by VanB to replace the D-ala-D-ala dipeptide (target of glycopeptide antibiotics) with D-ala-D-lac. VanX eliminates D-ala-D-ala dipeptides through hydrolysis and VanY (a carboxypeptidase) fulfills the duty of hydrolyzing D-ala-D-ala pentapeptide ends. This process leads to a modified peptidoglycan precursor and to a massive decrease in the

binding affinity of vancomycin and similar antibiotics, rendering them ineffective (Courvalin, 2006; Zeng *et al.*, 2016).

The two main resistance types (*vanA* and *vanB*) pose considerable problems in hospitals currently, because they frequently occur in drug-resistant *Enterococcus* (VRE) and MRSA (VRSA) strains, for which vancomycin is commonly used as a last resort antibiotic. It has been shown that the HGT of vancomycin resistance can occur via mobile genetic elements (MGE). In these specific cases, *vanA* is mobilized by Tn*1546* and related elements via conjugative plasmids, and *vanB* is transferred by the CTns Tn*1547* and Tn*1549* (Zeng et al., 2016).

1.2.3 MDR Gram-negative and Gram-positive bacteria

Many pathogens mentioned in the ESKAPE- and WHO-lists are Gram-negative bacteria. They are more challenging to treat because of their second cell membrane, which makes them intrinsically resistant to several drug classes that cannot penetrate the outer membrane and reach their particular targets. Among them are the pathogenic carbapenem-resistant *P. aeruginosa, A. baumannii and Enterobacteriaceae* like *K. pneumoniae*. They belong to the group of hospital-associated infections (also called nosocomial infections) and are considered extremely-drug resistant (XDR) pathogens. That means that these bacteria are resistant to almost all of the first-line antibiotics commonly used in the clinic and also to carbapenems that are used as the second-line drugs (Tacconelli *et al.*, 2017). These XDR-bacteria need to be treated nowadays with a combination of aminoglycosides and last-resort antibiotics like polymyxins (polymyxin B or colistin) and the tetracycline tigecycline, which can show severe side-effects. Moreover, recent clinically relevant strains sometimes show a pan-resistant phenotype that includes resistance to the last-resort antibiotics (carbapenems, aminoglycosides, polymyxins and tigecycline) (Theuretzbacher, 2017; Karakonstantis, Kritsotakis and Gikas, 2020).

MDR Gram-positive bacteria are also of great concern. Among them are vancomycin-resistant *Enterococcus* (VRE), methicillin-resistant *S. aureus* (MRSA), and penicillin-non-susceptible *S. pneumoniae* (Tacconelli et al., 2017). Other drug-resistant pathogens include *Staphylococcus epidermidis* (Namvar *et al.*, 2014), *Streptococcus agalactiae* (Bolukaoto *et al.*, 2015), *Clostridium difficile* (Spigaglia, 2016), and *Listeria monocytogenes* (Morvan *et al.*, 2010). Their MDR strains are treated with newly developed antibiotics like 5th generation cephalosporins (ceftaroline and ceftobiprole), oxazolidinones (tedizolid phosphate), quinolones (besifloxacin, delafloxacin, ozenoxacin), tetracyclines

(omadacycline), and glycopeptides (dalbavancin, telavancin, oritavancin) (Jubeh, Breijyeh and Karaman, 2020).

Among the mentioned MDR pathogens, VRE infections are worrisome because they are intrinsically resistant to many antibiotics and available treatment options are limited. There have been cases where VRE strains have shown resistance to last-resort antibiotics like linezolid, daptomycin, tigecycline, and quinopristin-dalfopristin (Zhou *et al.*, 2020). Thus, there is an urgent need to reduce the spreading of these MDR strains, for example, through better screening and detection methods as well as the discovery of alternative drug targets, which may lead to more accurate early detection in infected patients and improved treatment options.

1.2.3.1 *Enterococcus faecalis* and *Enterococcus faecium*

Enterococci were discovered in human feces 121 years ago. They were first designated to the genus Streptococci (Murray, 1990) and, more than 80 years later, transferred to the genus Enterococci based on new biochemical data (Schleifer and Kilpper-Balz, 1984). Among the enterococcal family, there are two prominent species, *E. faecalis* and *E. faecium,* which are responsible for most nosocomial infections in humans. There are two subpopulations of *E. faecalis* and *E. faecium*, the benign commensals that live in the human intestine and the hospital-adapted pathogenic bacteria that have acquired antibiotic resistance genes via MGEs (plasmids and transposons). The second group is of great concern worldwide, especially VRE, which has an MDR-phenotype (Arias and Murray, 2012). They can cause urinary infections, sepsis, and endocarditis, jeopardizing patients' safety, especially immunocompromised and transplant convalescents (Agudelo Higuita and Huycke, 2014).

Vancomycin resistance (section 1.2.1.3) was discovered already in the 1980s (Leclercq *et al.*, 1988) and the appearance of ampicillin resistant *E. faecium* (AREfm) that adapted and survived under hospital-settings paved the way for the emergence of vancomycin-resistant *E. faecium* (VREfm) later. This successful *E. faecium* subpopulation (AREfm) has ampicillin- and other antibiotic resistance genes. Moreover, it has inserted pathogenicity islands that include genes for enhanced biofilm formation and colonization (Gao, Howden, and Stinear, 2018). Thus, the gain of vancomycin resistance led to an MDR 'superbug' that is of great concern. The numbers of VREfm infections have been increasing in the last 20 years (Figure 1-3). This is a worldwide trend. Since then, the ratio of *E. faecium* to *E.*

faecalis in patients has changed dramatically in favor of the former, and it is increasing steadily (Zhou *et al.*, 2020).

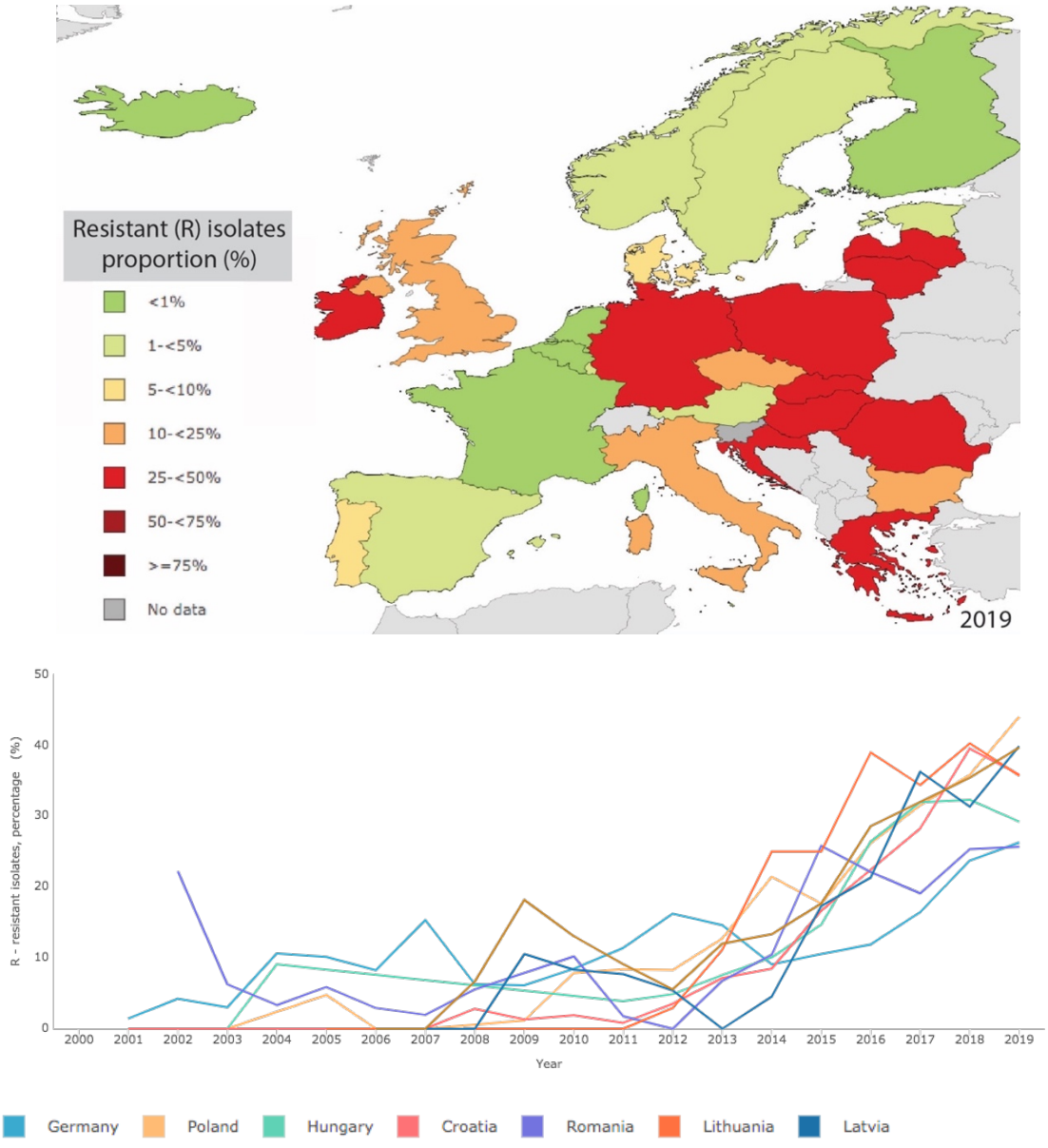

Figure 1-3: Monitoring data for vancomycin resistant E. faecium isolates in Europe. The data shows the percentage rates of VREfm isolates in Europe for 2019 (as a map in color-code - top) and plotted for selected European countries showing the highest increases in resistant isolates in the last 20 years (bottom). Dataset provided by ECDC based on data provided by WHO and Ministries of Health from affected countries. Data and images obtained from http://atlas.ecdc.europa.eu/public/index.aspx.

1.3 Antibiotic resistance spreading among bacteria

1.3.1 The pathways of horizontal gene transfer

Bacteria transfer their genetic information to their daughter cells in a so-called vertical manner, like all living organisms, through DNA replication and cell division. They are also capable of exchanging DNA with other cells via a process termed horizontal gene transfer (HGT). HGT provides important ways for bacteria to evolve and adapt rapidly to its environment. Different HGT pathways promote DNA exchange between cells and shape the bacterial genome (Hall et al., 2020). There are three canonical ways of gene transfer: transduction (Touchon, Moura de Sousa and Rocha, 2017), transformation (Ambur et al., 2016), and conjugation (Cabezón et al., 2015). Other non-canonical HGT pathways include the transfer of bacterial DNA through phage capsids (gene transfer agents, GTAs) (Lang, Zhaxybayeva, and Beatty, 2012), the usage of membrane vesicles (MV) that can transport not only DNA but also proteins, RNA, metabolites, and signaling molecules (Berleman and Auer, 2013), and through nanotubes (Dubey and Ben-Yehuda, 2011). The importance of HGT in bacterial evolution is reflected in estimates that up to 25% of some genomes arose from HGT (Ambur et al., 2016).

1.3.1.1 Transduction

Transduction happens when a virus transfers foreign DNA into a new host cell. Prominent examples of viruses that are capable of transduction in bacteria are the bacteriophages lambda in *E. coli* (Zeng et al., 2010) and CTXΦ in *Vibrio cholerae* (Waldor and Mekalanos, 1996). Usually, after infection, the phage enters a lytic cycle and hijacks host cell machineries to replicate its own genome and make new phage particles. This is accompanied by host DNA degradation and bacterial cell lysis, releasing the new virions. Phages can also enter into a lysogenic cycle that involves integration of the phage chromosome into the host genome. Here, the so-called prophage can stay dormant for longer time and can be transmitted vertically. Upon prophage activation, the phage genome is excised from the bacterial chromosome and starts the lytic cycle again, leading to phage replication and host cell lysis. During these processes two phage-mediated HGT events can occur: i) Specialized transduction occurs when a part of the flanking chromosomal DNA is excised together with the prophage, packaged in the phage particles and transferred to new host cells. ii) Generalized transduction happens when the phage capsid encapsulates small

pieces of bacterial DNA and transfers it into a new recipient cell (Touchon, Moura de Sousa, and Rocha, 2017).

1.3.1.2 Transformation

Transformation is the action of uptake of foreign DNA from the environment. It was the first discovered HGT mechanism (Griffith, 1928). DNA can be released to the environment from dying and self-lysed cells, phage particles, or through ejection from living cells. In order for natural transformation to occur, the bacterial cell needs to attain a so-called state of "competence" (Mell and Redfield, 2014). This means that the cell is ready for extracellular DNA uptake and involves expression of a number of specific genes. The process is well regulated and time-limited, depending on the bacterial growth state and specific environmental conditions. First, DNA is recognized and bound by a DNA receptor at particular cell surface sites. Single DNA strand then passes through the cytoplasmic membrane via a DNA translocase, while the other strand is degraded by nucleases. Natural transformation of plasmid DNA seems to be less efficient than for linear segments, due to the need for single-stranded DNA in the translocation process (Thomas and Nielsen, 2005). Natural transformation has been detected in diverse bacteria and archaea. Many pathogenic bacterial strains (*Campylobacter*, *Haemophilus*, *Helicobacter*, *Neisseria*, *Pseudomonas,* and *Staphylococcus,* among others) can use this HGT mechanism, which may help them to uptake of antibiotic resistance genes and develop MDR superbugs (Thomas and Nielsen, 2005).

1.3.1.3 Conjugation

Bacterial conjugation (Lederberg and Tatum, 1946) represents perhaps the most used HGT mechanism for ABR spreading among bacteria (Mazel and Davies, 1999; Grohmann et al., 2018). The conjugation process allows the transfer of ssDNA from a donor bacterium to a recipient cell. For that, a specific multi-protein transporter complex, which creates cell to cell contact between donor and recipient bacterium, needs to be established (Thomas and Nielsen, 2005). This complex machinery is often encoded in MGEs, such as autonomous replicating and conjugative plasmids (Smillie et al., 2010) and in conjugative transposons (CTns/ICEs) (Wozniak and Waldor, 2010).

Such conjugation systems are considered a subfamily of the type IV secretion system (T4SS) of bacterial transporters. Their feature is that they can also transport DNA apart from protein effectors (Zechner, Lang and Schildbach, 2012). T4SSs have been shown to fulfill many functions in both Gram-negative and in Gram-positive bacteria. Apart from

DNA transport between bacteria and to the extracellular space, they are also crucial for the delivery of effector proteins into eukaryotic cells (Sámano-Sánchez and Gibson, 2020), contribute to biofilm formation, and are important in defense mechanisms against other bacteria by delivering toxins (Grohmann et al., 2018).

All necessary factors for DNA processing prior transfer and for its transport to a new bacterial cell are encoded in two regions : i) the Dtr-operon (DNA transfer replication), that includes the relaxase and other accessory proteins for DNA preparation, and ii) the Mpf-operon (mating pair formation) that encodes proteins for T4SS- and conjugative pilus-formation (Fernández-López *et al.*, 2006).

The conjugation process starts with formation of the relaxosome complex. A relaxase protein binds at a specific DNA site (origin of transfer, oriT) in the MGE. It nicks the DNA at the *nic* site in a particular strand, with the help of auxiliary proteins that promote DNA bending and unwinding, forming a so-called relaxosome complex. The relaxase uses the hydroxyl group from its catalytic tyrosine for nucleophilic attack at the DNA phosphodiester bond and becomes covalently attached to the ssDNA by forming a phosphotyrosyl linkage (De La Cruz et al., 2010).

An actual model of how this process may continue is based on extensive biochemical and structural work, recently reviewed in (Cabezón *et al.*, 2015; Ilangovan, Connery and Waksman, 2015). In short, translocation by the conjugative T4SS is proposed to continue in the donor cytoplasm by interaction of the relaxosome with the coupling protein (T4CP) of T4SS (Cabezón, Ignacio Sastre and De La Cruz, 1997), guiding the ssDNA through the translocase core. Whereby, T4CP may use the energy of ATP-hydrolysis for pumping ssDNA into the channel of the T4SS complex (Cabezon and de la Cruz, 2006). The covalently bound relaxase is thought to be co-translocated to the recipient cell in an unfolded state. The presence of internal translocation signals needed for T4SS recognition and transfer in relaxases supports this idea (Redzej *et al.*, 2013), but the mechanism is not yet clear.

Following ssDNA translocation, it is believed that the relaxase will recognize the *nic* site again and circularize the ssDNA in the recipient cell (Draper *et al.*, 2005). The remaining MGE strand is simultaneously replicated in the donor cell, so that the MGE is both sustained and transmitted in the process (Cabezón et al., 2015).

1.3.2 Mobile genetic elements (MGEs)

Mobile genetic elements (MGEs) are one of the main means how bacteria become multi-resistant or even pan-resistant (Partridge *et al.*, 2018). These discrete pieces of DNA possess inherent mobility or can be moved around by other elements in the genome. Their discovery earned Barbara McClintock the Nobel prize in Medicine in 1983 (Mcclintock, 1983). In bacteria, intracellular mobility happens when such an element moves from one location to another in the genome, from the genome to a plasmid or from a plasmid to the genome or another plasmid. However, some elements can also move between cells, from one bacterium's genome to another through horizontal gene transfer (HGT). Many bacterial MGEs carry various types of genes that can positively impact the host's growth and survival. These genes include ABR genes and therefore MGEs can act as vehicles for resistance transfer. In bacteria, there are several main MGE classes, which I will describe in more details below.

1.3.2.1 Insertion Sequences

Insertion sequences (ISs) are the most simple and shortest (0.7 – 2.5 kb) autonomous MGEs. Classical ISs only contain the genes necessary for their own transposition, typically only one transposase (*tnp*) gene, which protein product catalyzes DNA cleavage and joining for genomic excision and integration. The IS ends are marked by specific DNA sequences, which interact with the Tpase and feature inverted terminal repeats (ITR). IS transposases (Tpases) can be classified into two major families (Mahillon and Chandler, 1998).

The most abundant, the DDE- (Asp, Asp, and Glu) Tpases use two metal ions as catalysts for DNA hydrolysis and strand transfer reactions (Hickman and Dyda, 2014). Prominent IS families that carry DDE transposases include IS*6*, IS*630*, IS*4*, IS*3*, IS*21,* and IS*30,* among others (Siguier et al., 2015). The second class of Tpases belongs to the HUH nuclease protein superfamily, members of which contain a histidine (H) -hydrophobic (U) -histidine (H) triad in their active sites and use a tyrosine (Y) residue as the nucleophile. The most prominent families with such Tpase are IS*200*/IS*605* (Barabas et al., 2008) and IS*91* (Chandler et al., 2013).

Traditional ISs do not carry accessory genes, such as toxins or ABR genes. However, they can move such cargos, when two copies of related ISs surround an external gene, forming a so-called "composite transposon" that can move as a single unit (Razavi et al., 2020). Some examples of ABR-carrying ISs have been detected in pathogenic strains (Partridge *et al.*, 2018). In particular, IS*26* from the IS*6* family has been shown to frequently carry ABR

genes in Gram-negative bacteria. Its success is likely due to two main features. First, it contains a -35 promoter sequence at both ITR ends, which can enhance gene expression if inserted close to a -10 promoter sequence upstream of a gene, such as an ABR gene (Lee, Hopkins and Syvanen, 1990). Second, IS*26* can move as a so-called translocatable unit, which includes its own sequence and the adjacent region (up to the next IS*26* junction in a composite transposon). This pathway enables efficient accumulation of ABR genes in a single transposable unit (Harmer, Moran and Hall, 2014). In Gram-positive pathogens, other members of the IS6 family, such as IS*257* (IS*431*) behaves similar to IS*26* and can carry many different ABR genes (Needham, Noble and Dyke, 1995).

1.3.2.2 Transposons

Transposons are discrete DNA segments that can relocate in genomes, inserting themselves at diverse sites using self-encoded transposase proteins. They generally comprise ITRs at their ends and an array of genes encoding for their own transposase, regulatory genes, and passenger (or cargo) genes, including antibiotic resistance genes (Chandler, 2016). Here, I will focus on the two transposon families, Tn*3* and Tn*7*.

Tn*3* family transposons contain a large transposase gene (*tnpA*) and ITRs at the transposon ends that are around 38 bp in length (Nicolas *et al.*, 2015). Tn*1546* is the most prominent example of a Tn*3*-like transposon in Gram-positive bacteria. It is responsible for vancomycin resistance (*vanA*- type) dissemination in *Enterococci* and vancomycin resistance transfer to MRSA, which led to some of the VRSA outbreaks (Zeng et al., 2016). Moreover, Tn*3*-like transposons can become more complex and acquire additional ABR genes through IS insertions (Partridge *et al.*, 2018).

Tn*7* transposons carry the *tnsABCDE* gene cluster (*tni* region) necessary for transposition and uses a 'cut-and-paste' mechanism for moving to a new genomic site. It encodes ITRs of approximately 28 bp length and additional internal TnsB binding sites. TnsA and TnsB form a heteromeric complex that executes Tn*7* excision. TnsD and TnsE are required for target site selection for insertion. TnsC is a regulator that connects TnsA/B with TnsD or TnsE. When using TnsD, Tn*7* inserts at a specific site, called *attTn7*, located downstream of the *glmS* gene in the chromosome of Gram-negative bacteria. In turn, TnsE directs Tn*7* to insert into random sites in conjugative plasmids, promoting its dissemination to other bacterial cells and species (Peters and Craig, 2001). Prominent examples of Tn*7*-related transposons include Tn*552* from *S. aureus* that encodes (*blaZ*) resistance to penicillin and *A. baumannii* resistance islands AbaR and AbGRI1 (Partridge *et al.*, 2018; Bi *et al.*, 2019).

1.3.2.3 Plasmids

Plasmids are non-chromosomal DNA elements (with a size from a few kb to several mb) that can replicate and appear in Gram-negative and Gram-positive bacteria. They can carry ABR genes and mobilize other MGE (such as ISs, transposons, and integrons), which contain ABR genes themselves, making some of these plasmids vast MDR harbors. Another characteristic of plasmids is that they can be transferred horizontally via conjugation (see section 1.3.1.3) if they encode proteins for mobilization (mobilizable plasmids) or for autonomous conjugation (conjugative plasmids) (Thomas, 2000). The necessary genes are found at specific transfer (*tra*) regions composed of Mpf- and Dtr- operons (see section 1.3.1.3). Here, I highlight a few examples of resistance carrying plasmids in Enterococci and their role in ABR spreading, with a focus on vancomycin resistance.

In Enterococcus spp*.,* resistance plasmids are divided into three groups, Rep_3, Inc18, and RepA_N families, depending on their replication factors (Jensen *et al.*, 2010). A prominent example of the Inc18-family is the conjugative pRE25 plasmid that encodes macrolide-lincosamide-streptogramin (MLS)- (*ermB*), chloramphenicol- (*cat*), kanamycin-neomycin- (*aphA3*), aminoglycoside- (*aadE*), and streptothricin- (*sat4*) resistance (Schwarz, Perreten and Teuber, 2001). pRE25-like plasmids are also involved in disseminating vancomycin-resistance type A (*vanA*) to MRSA strains with the help of Tn*1546*-like transposons (Zhu *et al.*, 2010). Pheromone responsive conjugative plasmids of the RepA_N family can also carry several ABR genes, like streptomycin- (*aadE*), kanamycin-neomycin- (*aphA3*), and MLS- (*ermB*) resistance, but also glycopeptide- (*vanA*) resistance. pRUM-like plasmids and megaplasmids from the RepA_N family spread glycopeptide- (*vanA*) resistance with the help of Tn*1546*-like transposons in *E. faecium* (Hegstad *et al.*, 2010). These examples highlight that several plasmids promote vancomycin resistance transfer among Enterococci. However, plasmids do not stably integrate into bacterial genomes and therefore confer only temporarily resistance. In contrast, conjugative transposons, which share several features with conjugating plasmids, can create a stable resistant phenotype (see next section).

1.3.2.4 Conjugative transposons

Conjugative transposons – CTns (also referred to as integrative and conjugative elements; ICEs), constitute a class of autonomous MGEs. They play an important role in bacterial adaptation and are very efficient in ABR spreading, because they can harbor many types of antibiotic resistance genes, autonomously shuttle these between diverse bacterial cells, and

stably integrate them into the recipient genomes. These MGEs are found in Gram-positive and Gram-negative bacteria, and their size can range from 18 to more than 600 kbp. Their genetic composition is commonly divided into modules, which perform specific functions in the CTn life cycle: i) excision/integration functions (transposition module), ii) plasmid-like conjugation and maintenance functions (conjugation module), and iii) an accessory genes module that can encode ABR or virulence genes, among others (Burrus *et al.*, 2002). CTns are typically found inserted in the host genome. CTn movement (Figure 1-4) starts under specific cellular conditions, when excision of the element leads to the formation of a closed circular intermediate (CI) DNA molecule. Excision occurs through recombination of specific DNA sites at the transposon ends (inverted repeats, IRs). These DNA sequences are sometimes called attachment sites (*attL* and *attR*). Recombination reactions are performed by an enzyme called integrase (Int), a tyrosine or serine site-specific recombinase family protein that is encoded on the element itself. This creates the CI intermediate, a sealed dsDNA molecule with the CTn ends joined together, connected by a 5-7 bp long non-homologous overlap or crossover region (CR). This specific site can sometimes be called *attP*. Afterwards, the conjugative transfer of the CI occurs.

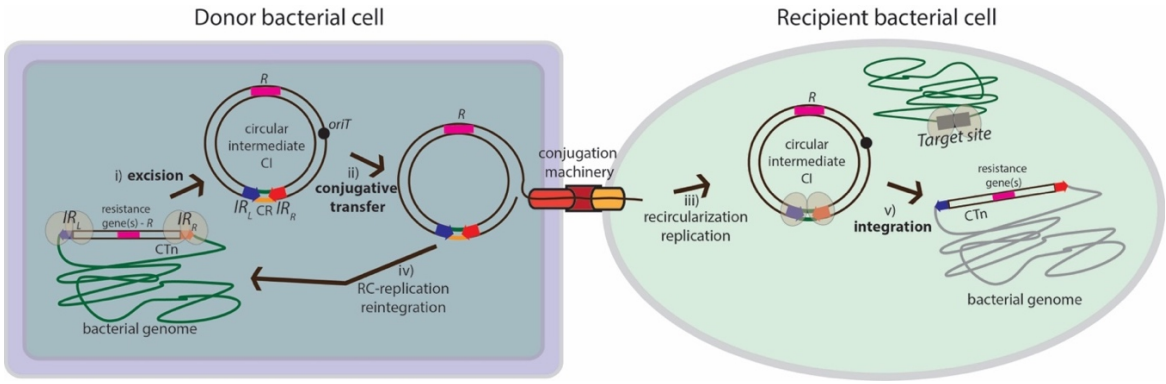

Figure 1-4: The transposition and conjugation pathway of conjugative transposons (CTns). (i) The transposition reaction starts by CTn excision from the donor's cell genome. For that, the CTn integrase (green ovals) binds and performs the recombination reactions at the transposon ends (IR_L in blue and IR_R in red). This leads to the formation of a circular intermediate (CI) CTn that is connected by a 5-7 bp long non-homologous crossover region (CR, top strand in green, bottom strand orange). (ii) The CI will be subsequently nicked at oriT by a relaxase and initiate the conjugation process (see section 1.3.1.3) and be transferred as ssDNA to the recipient cell. (iii) The CTn in its ssDNA form is proposed to be circularized by the relaxase protein and the dsDNA CI generated later via rolling circle replication in both, (iv) donor and (iii) recipient cell. (v) Afterwards, the CI will find a target site (black rectangle) and integrate with the help of the CTn integrase and probably transposon- and host-encoded auxiliary factors. Some CTns integrate at specific target sites, others are more promiscuous and integrate at various target sites. (iv) The reformed dsDNA CI in the donor cell will also be reintegrated in the genome in a similar manner. If the CTn carries antibiotic resistance determinants (R, rectangle in magenta), it will be co-transferred and -integrated in the recipient cell and maintained in the donor cell, leading to the ABR spreading.

It has been shown that many CTns contain an origin of transfer (*oriT*) sequence that is necessary for initiating the conjugative transfer. Like conjugative plasmids, many CTns encode their own relaxase that cleaves a specific strand of the *oriT*, leading to relaxosome formation. This protein-DNA complex is then brought to the mating pore or conjugation machinery (usually related to the T4SS family) with the help of a coupling protein. A single strand of the CI DNA is then transferred into a recipient bacterial cell, but the mechanism is still not well understood (see section 1.3.1.3). Nevertheless, it is assumed that rolling circle replication generates the dsDNA CI both in the donor and recipient cells (Burrus, 2017). Next, the circular CTn finds a suitable target site (integration site, *attB*) in the recipient's genome, where it is integrated again with the help of the CTn integrase protein and probably other accessory factors (transposon- or host-encoded). Many CTns integrate at specific genomic locations, usually close to tRNA genes. Others can be more promiscuous and show integration into many different target site sequences (Johnson and Grossman, 2015). Tn*916* is a prototype of CTn elements. It was originally found in *Enterococcus*, but related elements are present in a diverse range of bacteria (Roberts and Mullany, 2009). Usually, Tn*916* confers resistance to tetracycline (*tet*(M)), but similar elements, like Tn*1545* can additionally confer resistance to MLS (*erm*(B)) and kanamycin/neomycin (*aphA-3*) antibiotics (Cochetti *et al.*, 2008). Another related element is Tn*1549* that confers resistance to vancomycin (see section 1.2.2.3). Further details of the modularity and mobilization mechanisms of Tn*916* and Tn*1549* are described in sections 1.5.1 and 1.6.

1.4 Mechanisms of site-specific DNA recombination

In general, DNA recombination is a way to create DNA rearrangements that exists in all three kingdoms of life. There are two basic types of DNA recombination: homologous and site-specific recombination. Homologous recombination depends on inter- or intramolecular sequence homology and is applied, for example, during meiosis and DNA repair. In turn, site-specific recombination requires specific recombinase proteins and specific DNA sequences in both recombination substrates. Site-specific recombination is frequently used by viruses and transposons, leading to deletions, insertions, duplications, and inversions of specific DNA segments. This work concentrated on the study of conjugative transposons that undergo site-specific recombination during their excision and integration in the bacterial genome. Thus, I will next explain more details about the mechanisms of site-specific recombination.

1.4.1 Site-specific recombination

Site-specific recombination is a process that takes place during integration, excision, and inversion of specific DNA segments. This process requires a specialized site-specific recombinase protein that belongs to the tyrosine- or serine recombinase family. These enzyme families got their names based on the catalytic residue responsible for cutting and joining the DNA during the recombination. The recombinase protein is responsible for recognizing the recombined DNA sites and for catalyzing all cleavage and ligation reactions. The mechanisms of both families have some similarities and differences, exemplified in Table 1-2: Comparison of both (tyrosine- and serine-) site-specific recombination mechanisms. In short, tyrosine recombinases perform the strand cleavage and exchange reactions in a sequential manner. First, one strand pair is cleaved and recombined, creating a four-way Holliday junction (HJ) intermediate (Liu and West, 2004). Then, recombination of the second pair of DNA strands completes the process (Grindley, Whiteson and Rice, 2006).

In turn, serine recombinases perform the necessary reactions by cleaving all four DNA strands at once, followed by subunit rotation that enables simultaneous strand exchange and religation of all strands. This simplified view of the site-specific DNA recombination mechanism is, in reality, much more complex. For several recombination systems it has been shown that the reactions are regulated in a sophisticated manner, through various accessory proteins and DNA sites (Grindley, Whiteson and Rice, 2006).

Table 1-2: Comparison of both (tyrosine- and serine-) site-specific recombination mechanisms.

Feature	Tyrosine recombinases	Serine recombinases
Catalytic residue	Tyrosine	Serine
DNA strands cleaved at a time	2	4
Type of covalent intermediate	3'-phosphotyrosine	5'-phosphotyrosine
Crossover region length	6-8 nucleotides	Two nucleotides
The polarity of free DNA hydroxy group after cleavage	5'-OH	3'-OH
Synaptic complex	Tetrameric	Tetrameric
Mechanism	Sequential cleavage at opposite sites with HJ intermediate	Simultaneous cleavage with subunit rotation

1.4.2 Tyrosine site-specific recombination

Tyrosine recombinases are prominent in prokaryotes but can also be found in archaea and eukaryotes. From a structural point of view, they share a conserved domain arrangement, but their sequences can be very distinct from each other. They possess a C-terminal catalytic

domain (CAT) with a well-conserved active site (Nunes-Düby *et al.*, 1998). Generally, the fold of the CAT-domain is the most conserved part of this recombinase family. In addition, all tyrosine recombinases contain a second domain that binds to the core DNA recombination sites (called core-binding domain, CB), but is less well conserved than CAT (Grindley, Whiteson and Rice, 2006). A well-known example of a tyrosine recombinase with these two domains is the Cre recombinase (Duyne, 2015). Finally, some members of the family contain a third domain preceding the usual CB, called the arm binding domain (AB). This domain does not participate directly in the recombination reactions at the core sites, rather it plays a regulatory role by binding additional arm DNA sites (Grindley, Whiteson, and Rice, 2006) nearby the core sites. A well-studied example for this group of recombinases is the λ Integrase (Landy, 2015).

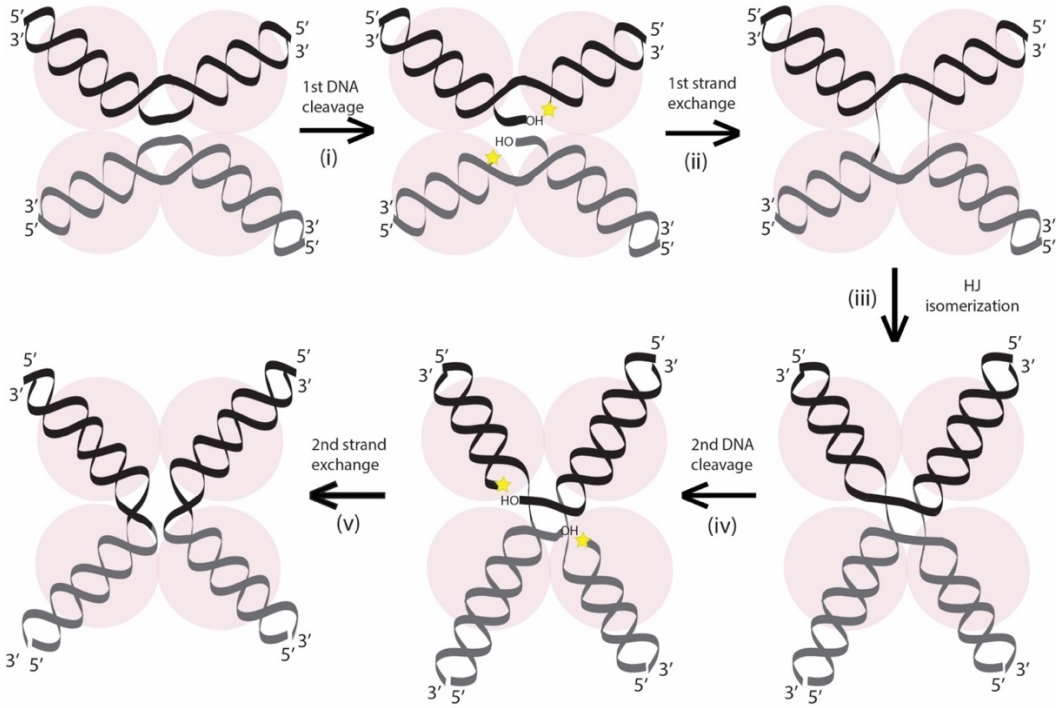

Figure 1-5: The tyrosine recombination pathway. (i) Two tyrosine recombinase dimers (light gray circles) will bind a specific target site each (in black and dark gray) and come together, leading to a synaptic complex formation. Afterwards, the first DNA cleavage reaction will occur at one strand in each DNA substrate leading to the formation of a 3'-phosphotyrosine protein-DNA linkage (yellow stars) and the release of a free 5'-OH group at each cleavage site. (ii) Then, the reaction proceeds through an exchange of the cleaved DNA strands. Each free 5'-OH group will attack the 3'-phosphotyrosine bond in the recombination partner DNA molecule, leading to formation of a four-way Holliday Junction (HJ) intermediate. (iii) HJ isomerization will activate the second protein pair, leading to the second round of (iv) cleavage and (v) strand exchange, generating the final recombined products (Grindley, Whiteson and Rice, 2006).

The general reaction mechanism of tyrosine recombinases is quite well understood nowadays, based on many biochemical and structural studies in the last three decades (Figure 1-5). First, two recombinase molecules recognize and bind to each specific

recombination DNA substrates. For that, each recombinase monomer interacts with DNA by forming a C-shaped clamp that encloses the DNA at the center. Each DNA site is constituted of an inverted repeat, with a 5-8 bp long central region (called the crossover region) and engages two recombinase molecules at the repeats. Two DNA substrates then come together to form a synaptic complex with four recombinase monomers. In this complex, DNA cleavage can occur at one strand in each DNA substrate precisely at the 5' boundaries of the crossover region. The trans-esterification reaction (Figure 1-6) is performed by a catalytic tyrosine in the recombinase active site, leading to the formation of a covalent protein-DNA 3'-phosphotyrosine linkage and the release of a free 5'-hydroxy group (Grindley, Whiteson and Rice, 2006). The reaction is supported by the well-conserved catalytic pentad RKHRH in the active site pocket, which are necessary to coordinate and prepare the scissile phosphate (Chen and Rice, 2003). Another characteristic is that the last helix (or beta-strand for λ Int) at the C-terminal CAT-domain of each monomer swaps and interacts with its neighbor partner in a circular manner, stabilizing synaptic complex formation.

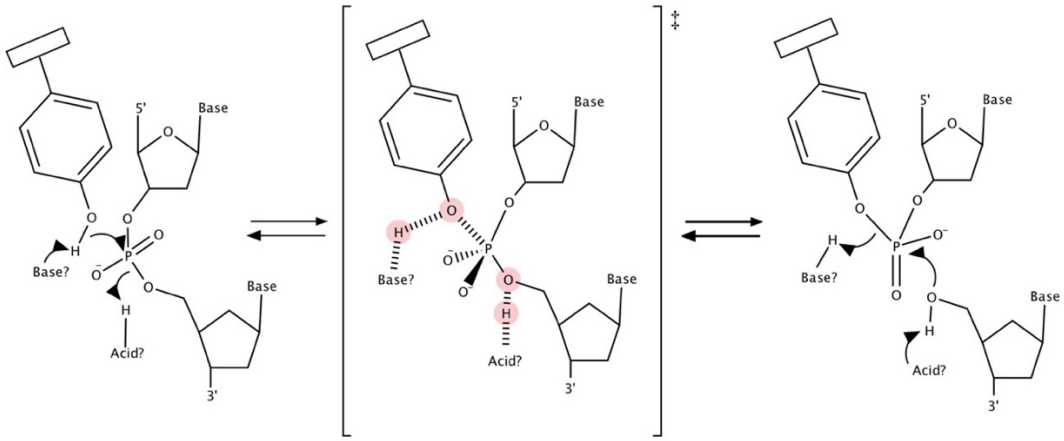

Figure 1-6: The trans-esterification reaction performed by tyrosine recombinases. Schematic of the nucleophilic attack performed by a site-specific tyrosine recombinase (rectangle). The scissile phosphate of the DNA substrate will be attacked by the tyrosine's hydroxyl group. The reaction goes through a transition state, where the backbone phosphate may be negatively charged. The reaction is supported by a proton acceptor (base) and a proton donor (acid). Finally, a 3'-phosphotyrosine bond will be formed, and a free 5'-OH will be released. The reaction is reversible. Based on (Grindley, Whiteson and Rice, 2006).

Afterwards, the reaction proceeds through an exchange of the cleaved DNA strands. Each free 5'-hydroxy group then attacks the 3'-phosphotyrosine bond in the recombination partner DNA molecule, leading to formation of an HJ intermediate. Then, the HJ will isomerize, leading to activation of the second pair of recombinase subunits in the tetramer, which did not participate in the first round of recombination reactions. These then perform

the same reaction steps of strand cleavage and exchange, leading to HJ resolution and product formation (Grindley, Whiteson, and Rice, 2006).

1.4.3 Prototypes of the tyrosine recombinase family

1.4.3.1 Cre recombinase

The Cre recombinase is a widely known genetic engineering tool, used in a wide range of organisms (Wirth *et al.*, 2007). Moreover, it is perhaps the best-studied tyrosine recombinase. with a large pool of biochemical and structural data available. Therefore, Cre has played an immense role in elucidating the mechanisms of site-specific recombination (Grindley, Whiteson, and Rice, 2006).

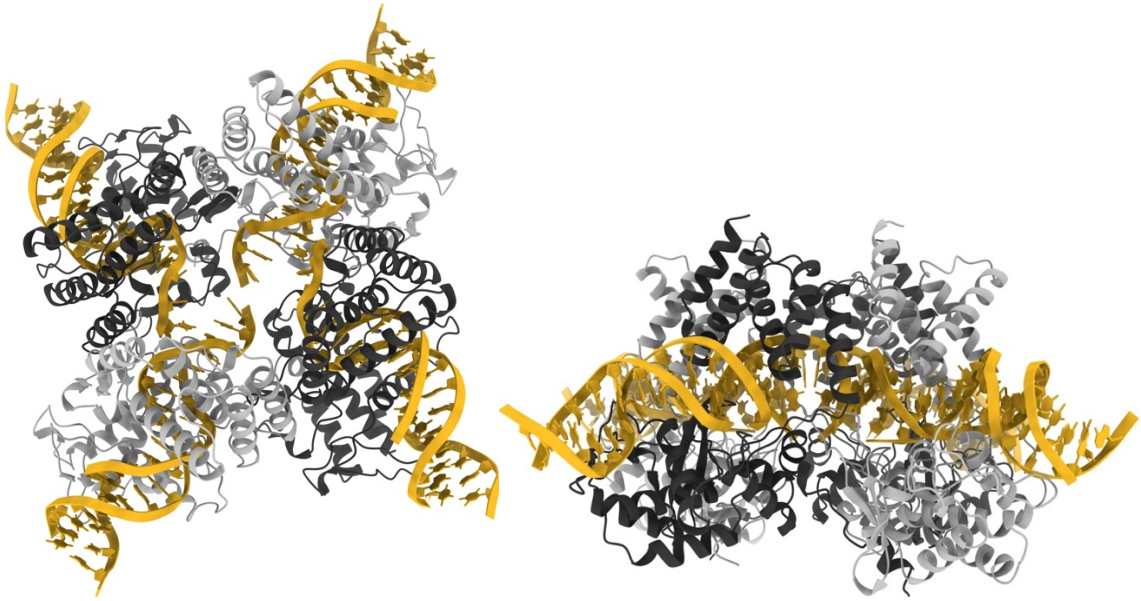

Figure 1-7: The crystal structure of the post-cleavage Cre-loxP complex. Cartoon representation of the synaptic Cre complex (top view left, side view right). Two Cre monomers bound to the same loxA site (golden) are shown in grey and light grey, pdb: 1crx (Guo, Gopaul and Van Duyne, 1997).

In nature, Cre is encoded in the bacteriophage P1, and its role is to resolve erroneously formed dimers of the P1 genome to monomers before cell division. For that, Cre recombines two identical 34-bp long sites called *loxP*, composed of two 14-bp inverted repeats (Van Duyne, 2001). Cre is composed of two domains, an N-terminal DNA binding domain and a C-terminal catalytic domain. The *loxP* sites are divided by an asymmetric 6-bp crossover region (CR). The first Cre-DNA crystal structure showed that one Cre molecule binds to a single *loxP* site, by encircling the DNA with both protein domains and forming interactions through both the DNA minor and major grooves. Binding of a second Cre monomer at *loxP* occurs cooperatively, leading to increased binding affinity. In order for recombination to occur, two Cre-loxP dimers associate in an antiparallel fashion

generating a two-fold symmetric tetrameric synaptic complex. This assembly is needed before DNA cleavage and strand exchange reactions can happen (Duyne, 2015).

Further biochemical and structural studies showed that the first strand exchange reaction occurs specifically at the bottom strand (Lee and Sadowski, 2003), due to preferred DNA bending at the left half-sites of *loxP*. This biased bending, results in the activation of Cre monomers at the right half-sites, leading to cleavage and strand exchange at the bottom strands of *loxP* (Ghosh, Guo, and Van Duyne, 2007).

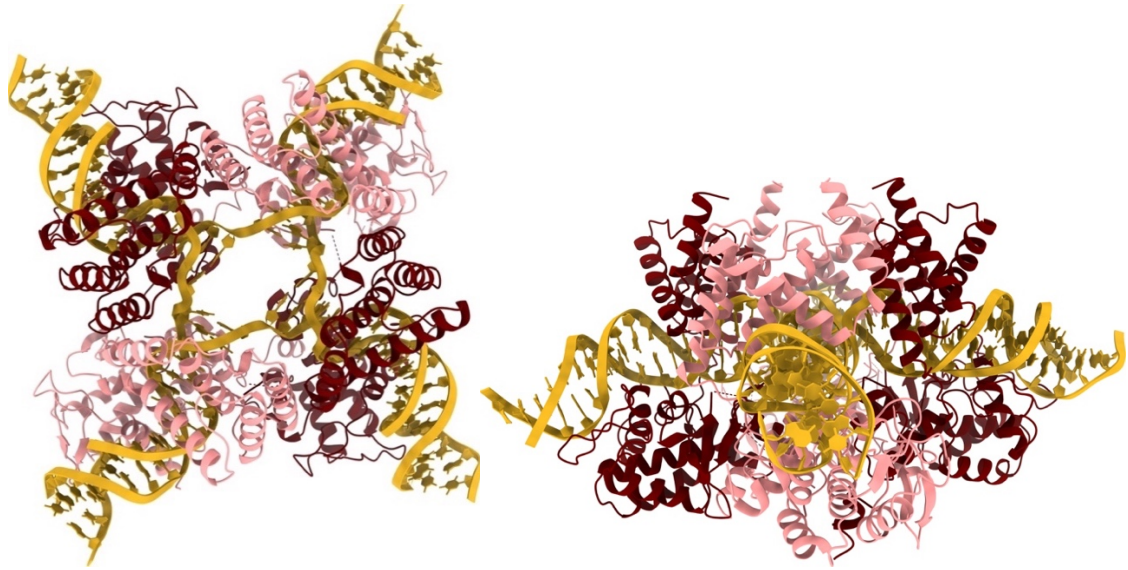

Figure 1-8: The crystal structure of the Cre-HJ DNA complex. Cartoon representation of the tetrameric Cre complex (top view left, side view right). Two Cre monomers bound to the same loxA site (golden) are shown in firebrick and pink, pdb: 2crx (Gopaul, Guo and Van Duyne, 1998).

A critical requirement in the Cre-loxP recombination pathway is that the crossover regions of the recombining DNA sites need to be homologous/identical to each other (Hoess, Wierzbicki, and Abremski, 1986). This homology is needed for efficient strand exchange and ligation during the recombination reactions (Nunes-Düby, Azaro, and Landy, 1995). Cre's active catalytic site contains various residues that are well conserved among the tyrosine recombinase family: Y324 is the active nucleophile that cleaves the DNA at the phosphate backbone, whereas other residues activate the scissile phosphate and stabilize the reaction intermediates during the reaction (Duyne, 2015). Following initial cleavage, Cre subunits form a 3'-phosphotyrosine bond and release a 5'-hydroxyl group in the DNA, as shown by various crystal structures that trapped the covalent intermediate complex, see Figure 1-7 (Guo, Gopaul and Van Duyne, 1997; Ennifar *et al.*, 2003). Next, strand exchange between the free 5'-hydroxy groups and the phosphotyrosine bonds on the partner *loxP* site result in forming an HJ intermediate, which has also been observed structurally, see Figure 1-8 (Gopaul, Guo and Van Duyne, 1998). This HJ can then isomerize, which

leads to activation of the Cre monomers bound at the left half-sites. The second pair of strand cleavage and exchange reactions at the top strands leads to HJ resolution and the formation of the recombined *DNA* products (Pinkney *et al.*, 2012).

1.4.3.2 Xer recombinase

Xer recombinases are necessary for bacteria and archaea to resolve erroneous chromosome dimers after replication (Blakely *et al.*, 1993). The XerC-XerD system in *E. coli* carries out recombination on *dif* recombination sites (Blakely *et al.*, 1993) in the bacterial genome to untangle physically-linked sister chromosomes before cell division. Recent results from our lab revealed the structures of XerH from *Helicobacter pylori*, giving new insights into how this protein family works (Bebel *et al.*, 2016). The first structure trapped a XerH-*dif*$_H$ synaptic complex in a pre-cleavage state (Figure 1-9). It revealed a surprisingly different tetrameric assembly compared to Cre-*loxP* and other available tyrosine recombinase-DNA complexes. The *dif*$_H$-DNA was found to be nearly straight, while other complexes contained strongly bent DNA. Nevertheless, the active site conformation resembled the typical tyrosine recombinase catalytic pocket, including the conserved residues R213, K239, H309, R312, H335, and the catalytic tyrosine Y344 (Figure 1-10). In this structure, Y344 was far away from the scissile phosphate in both monomers, indicating that some conformational changes needed to happen before DNA cleavage. A stimulatory factor that may be required for the activation is FtsK, which has been shown to promote XerH recombination in *H. pylori in vivo*.

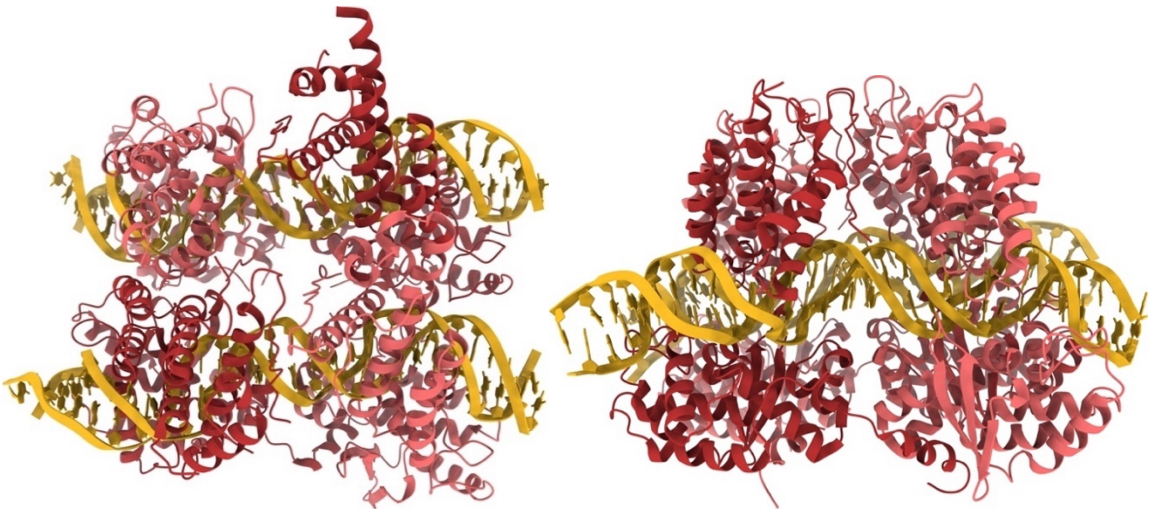

Figure 1-9: The crystal structure of the pre-cleavage XerH-dif$_H$ synaptic complex. Cartoon representation of the tetrameric XerH complex (top view left, side view right). Two XerH monomers bound to the same dif$_H$ site (golden) are shown in brown and light brown, pdb: 5jk0 (Bebel et al., 2016).

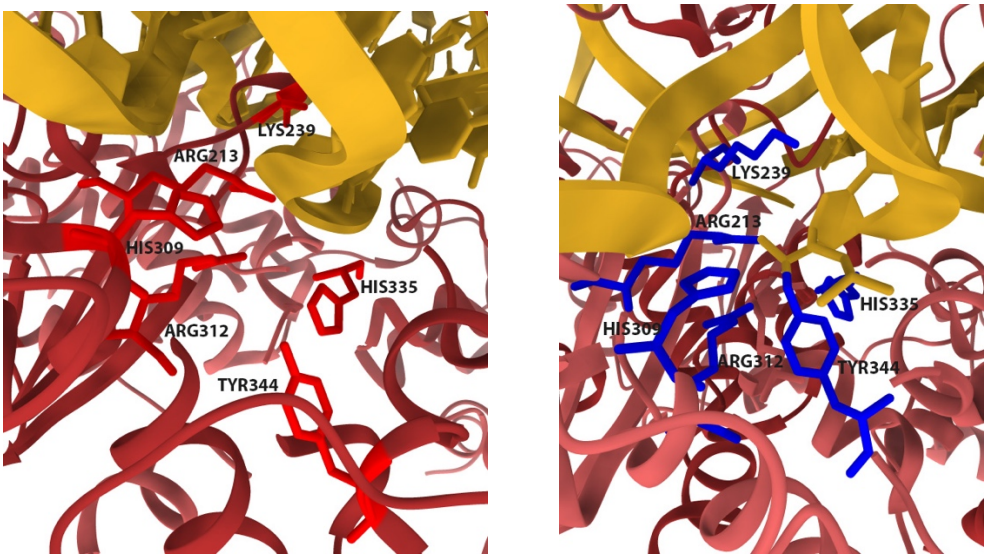

Figure 1-10: Catalytic pocket of the XerH monomer. Cartoon representation of the catalytic pocket of an inactive XerH monomer (left, catalytic residues in red) and the active site with a covalently 3'-phosphotyrosine bond (right, catalytic residues in blue) (Bebel et al., 2016).

The second structure trapped a XerH-dif_H synaptic complex in a post-cleavage state (Figure 1-11). Here, the dif_H DNA molecules are strongly and asymmetrically bent through the introduction of a kink, near the cleavage sites at the crossover region boundaries. The structure is more compact than in the pre-cleavage state due to significant rotations in all XerH subunits that led to more pronounced monomer-monomer interactions. These rearrangements further resulted in repositioning of the protein's C-terminal tails, moving the catalytic tyrosine residue Y344 close to the actives sites in two subunits of the tetrameric complex and allowing the formation of covalent phosphotyrosyl bonds (Figure 1-10, right). The other two monomers remained in an inactive conformation, like in the pre-cleavage synaptic complex. These results were in good agreement with previous data for other tyrosine recombinases and highlighted the special role of DNA bending in Xer activation during cell division. Another characteristic feature of the Xer structures is the intrinsic property of the system for asymmetric DNA bending, which leads to cleavage preference at the left dif_H sites in the first round and dictates the order of cleavage and strand exchange reactions.

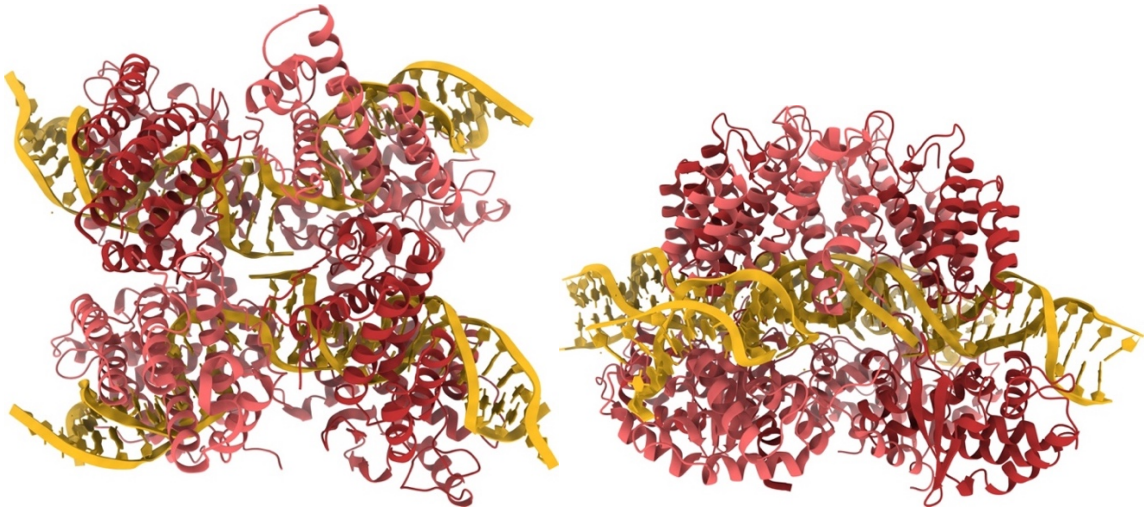

Figure 1-11: The crystal structure of the post-cleavage XerH-dif$_H$ synaptic complex. Cartoon representation of the tetrameric XerH complex (top view left, side view right). Two XerH monomers bound to the same dif$_H$ site (golden) are shown in brown and light brown, pdb: 5jk0 (Bebel et al., 2016).

1.4.3.3 Lambda integrase

The λ Integrase (Int) is encoded by bacteriophage λ and is responsible for its excision from and integration into the chromosome of *Escherichia coli* (Landy, 1989). This recombinase is very well studied, and various biochemical and structural studies have led to a good understanding of its mechanism. Its reaction differs from the Cre recombinase in that it is very well regulated and directional, where excision and integration reactions are irreversible and dependent on physiological signals of the host and on environmental cues (Landy, 2015).

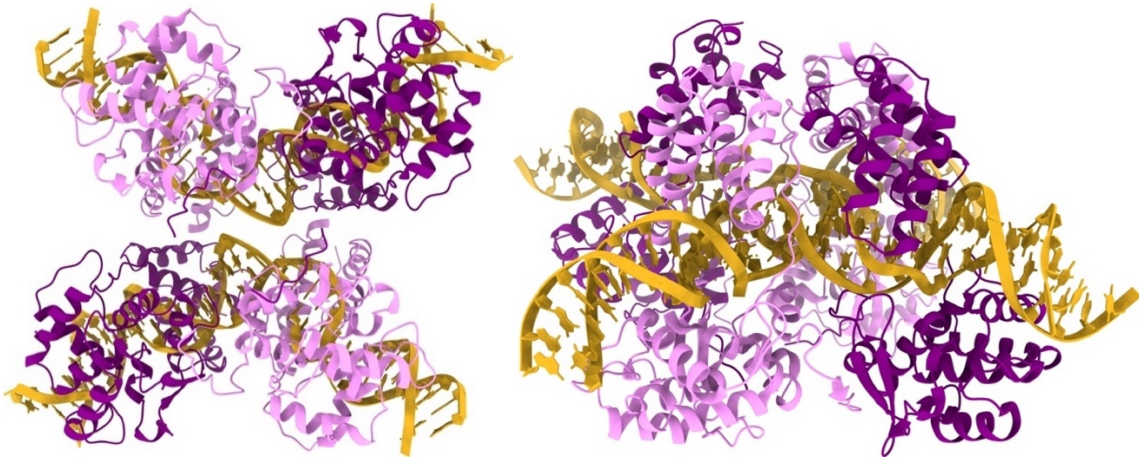

Figure 1-12: The crystal structure of the λ post-cleavage synaptic complex. Cartoon representation of the synaptic λ-Int complex (top view left, side view right). Two λ-Int monomers bound to the same core DNA site (COC', golden) are shown in purple and light purple, pdb: 1z19 (Biswas et al., 2005).

λ Int is composed of three protein domains. An arm binding domain (AB) that interacts with arm binding sequences inside the bacteriophage. A core-binding (CB) domain recognizes inverted repeat DNA sequences in the phage or the bacterial chromosome to

perform their recombination. The repeats of each site are separated by an identical 7-bp crossover region. Finally, λ Int contains a catalytic domain (CAT) that is responsible for performing all DNA cleavage and strand exchange reactions.

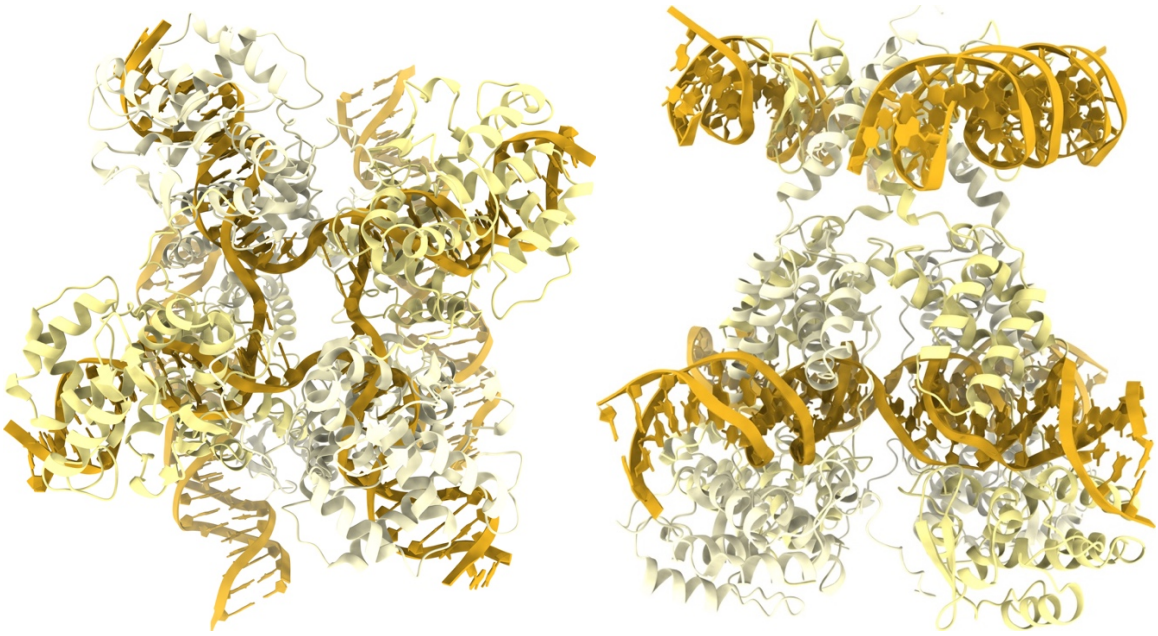

Figure 1-13: The crystal structure of the λ post-strand exchange HJ complex. Cartoon representation of the tetrameric λ-Int complex (bottom view left, side view right). Two λ-Int monomers bound to the same core DNA site (COC', golden) are shown in lemon and light lemon, arm DNA sites (P'1-P'2) are also in gold, pdb: 1z1b (Biswas et al., 2005).

In order for recombination to happen, λ Int needs to form a tetrameric complex, including two DNA substrates. For integration, λ Int binds to so-called *attP* (on the phage chromosome) and *attB* (on the bacterial chromosome) sites, whereas *attL* and *attR* sites at the pro-phage flanks are used for excision. Each *att* site binds two integrase monomers, which come together to form a recombination competent synaptic complex. Afterwards, λ Int performs the two rounds of cleavage and strand exchange reactions sequentially, in a similar way as it was shown for Cre. First cleavage and strand exchange reactions happen at the top strands of both sites during excision and integration, indicating that the reactions are not the reverse of one another. During these steps, i) a 3'-phosphotyrosyl bond is created, which is further ii) resolved by the attack of the partner 5'-hydroxyl groups leading to iii) the formation of the HJ intermediate. The structure of the post-cleavage synaptic complex, representing step i) has been reported (Biswas *et al.*, 2005). In this structure, N-terminally truncated λ Int (lacking the AB domain) forms a tetramer by binding to two core sites. This structure represents the reaction step after the first cleavage reaction but before strand exchange, because the strands ending with a free 5'-OH are still base paired. Two λ Int monomers are in an active state and have cleaved the DNA creating 3'-phosphotyrosyl

bonds, while the other two monomers are in an inactive state. Moreover, the DNA molecules are strongly bent in an asymmetric fashion, bringing the cleaved 5' ends close to the neighbor partner's 3'-phosphotyrosine, what may facilitate strand exchange and ligation. Bending might originate from a kink at the bottom, non-cleaved strand, 2 nt after the cleavage site (Figure 1-12).

Step ii) has been characterized by the post-strand exchange complex structure, which is composed of the full-length (FL) λ Int and DNA molecules representing the core and arm sites from *attP* (Biswas *et al.*, 2005). Compared to the synaptic post-cleavage complex, the free 5' ends have exchanged their positions and base-paired with the non-cleaved partner strands. HJ formation was prevented by the addition of a 5' phosphate group. The DNA kink has shifted to the junction center, leading to a stronger bending that brings the bottom strand cleavage sites closer to each other. Thus, this structure showed that arm DNA binding promotes strand exchange, but further remodeling is still needed to bring the bottom cleavage sites even closer before the second-strand cleavage and exchange reaction can occur (Figure 1-13).

Figure 1-14: The crystal structure of the λ Int-HJ-arm DNA complex. Cartoon representation of the tetrameric λ-Int-HJ-arm DNA complex (bottom view left, side view right). λ-Int monomers are colored in green and light green, HJ and arm DNA sites in gold, pdb: 1z1g (Biswas et al., 2005).

Step iii) has been characterized by the structure of the HJ complex composed of the FL λ Int tetramer, HJ DNA and *attP* arm DNA (Biswas *et al.*, 2005) (Figure 1-14). In this state, the arm DNA sites are binding in a unique orientation, positioning the active Int subunits ready for HJ resolution towards products (Biswas *et al.*, 2005; Radman-Livaja *et al.*, 2005). The overall reaction of λ Int recombination is more complex than for the simpler recombinases like Cre or Xer. For efficient and correct functioning, λ Int needs additional factors, namely IHF, FIS (host-encoded), and λ Xis (bacteriophage-encoded), which help

dictate the directionality of the process towards excision or integration. While IHF is necessary for both reactions, FIS stimulates integration, and λ Xis inhibits it. On the other hand, λ Xis and FIS, together with IHF, promote excision. These small proteins drive DNA bending and help bring the arm DNA sites close to the core binding sites (Landy, 2015).

1.5 Conjugative Transposons encoding members of the tyrosine recombinase family

In recent decades many different CTns (or ICEs) have been discovered and studied. An excellent recent review summarizes their characteristics (Johnson and Grossman, 2015). These MGEs are particularly gaining attention because they can move ABR genes between bacteria (in both Gram-positive and Gram-negative bacteria), which may give them an evolutionary advantage (Partridge *et al.*, 2018). As explained in section 1.3.2.4, these elements are usually found inserted in the bacterial genome. However, they need to excise from it to be horizontally transferred to a new host, where they need to integrate again to fix themselves in the recipient's genome. Both reactions are often performed by self-encoded recombinases, called integrases. This work is focused on studying integrases of the tyrosine recombinase family, for which it has been proposed that they perform the excision and integration reactions following the canonical pathway of site-specific tyrosine recombinases (Hickman and Dyda, 2014).

1.5.1 Tn*916* and Tn*916-like* family

The Tn*916* family of CTns is prevalent among environmental, commensal, and pathogenic Gram-positive bacteria. Among them, Tn*916* was the first discovered CTn (in *E. faecalis*) and is the archetype of the CTn family (Franke and Clewell, 1981). This element has a size of 18 kb, and it is composed of four functional genetic modules: i) a conjugation module, ii) a regulation module, iii) a cargo module, and iv) a transposition module (Clewell *et al.*, 1995).

The regulation module encodes the genes necessary for transcriptional regulation, including induction and repression of genes essential for CTn transfer. The conjugation module encodes the proteins necessary for CTn transfer to a new host, including a specific relaxase (*orf20*) that nicks Tn*916* in the CI form at the origin of transfer (*oriT*) to initiate ssDNA transfer (Rocco and Churchward, 2006). In Tn*916*, the cargo module encodes for tetracycline resistance – *tet*(M), but related elements can harbor different ABR gene types and encode other functions (Roberts and Mullany, 2011). This reflects the ability of Tn*916*-

like CTns to acquire new accessory genes and makes them perfect vectors for the dissemination of diverse genetic traits.

The transposition module encodes the necessary proteins for performing the excision and integration reactions (Caparon and Scott, 1989). Typically, these are Int that catalyzes the recombination reactions and a helper protein called excisionase (Xis) that seems to specifically facilitate excision (Poyart-Salmeron *et al.*, 1989).

Recent sequencing results have shown that Tn*916* elements can be found in a broad range of various bacterial phyla. This could be a result of the unspecific integration site selection of these elements. Tn*916* prefers to integrate into AT-rich sites without a strong sequence specificity, which allows it to efficiently spread to diverse genomes (Roberts and Mullany, 2009).

1.6 Conjugative transposition of Tn*1549*

Tn*1549* was identified as the primary source of vancomycin resistance type B (*vanB*) (van Hal *et al.*, 2016), and it has been detected in many different strains in Gram-positive bacteria (including *Enterococcus* and *Clostridium* among others) (Launay *et al.*, 2006). Tn*1549* has a size of 34 kb and belongs to the Tn*916*-like family of CTns. It is composed of three functional modules (Figure 1-15): i) conjugation and mobilization, ii) resistance, and iii) transposition modules, Table 1-3 (Garnier *et al.*, 2000). The CTn contains at its ends so-called imperfect inverted repeats (IR) that are AT-rich (Lambertsen *et al.*, 2018). The resistance module encodes a VanB2 resistance operon, which confers resistance to vancomycin, a glycopeptide that inhibits cell wall synthesis in Gram-positive bacteria (see section 1.2.1.3) and is used as a last-resort antibiotic for multi-drug resistant infections (Rubinstein and Keynan, 2014). Thus, dissemination of resistance against it is a major concern (Arias and Murray, 2012; Zhou *et al.*, 2020).

The transposition module shows high similarity to the prototypical module of Tn*916* (Roberts and Mullany, 2011). It encodes two proteins: the integrase (Int) and excisionase (Xis), which together enable all steps of transposition. Tn*1549* Int is also a member of the large tyrosine site-specific recombinase family and is responsible for catalyzing DNA cleavage and rejoining reactions at the transposon ends (Rubio-Cosials *et al.*, 2018). In contrast, Xis is a small accessory DNA bending protein that promotes efficient and accurate transposition (Lambertsen *et al.*, 2018).

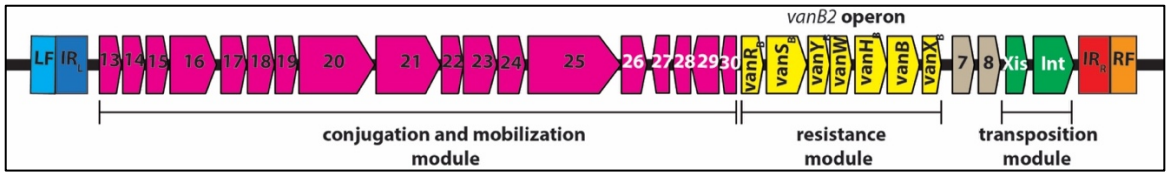

Figure 1-15: The conjugative transposon Tn1549. The CTn it composed of three functional modules: a conjugation and mobilization module (cherry arrows, orf13-30), a resistance module containing the VanB2 operon (yellow arrows) and a transposition module encoding Xis and Int proteins (green arrows). The transposon ends are found at inverted repeat left (IR_L) and inverted repeat right (IR_R), and represented as rectangles (blue and red, respectively). Left (LF) and right (RF) flank sequences are represented as rectangles (light blue and light red, respectively). A detailed description of the genes from Tn1549 can be found in Table 1-3 (Garnier et al., 2000).

The integrase (Int) is 397 amino acids long and has a molecular weight of 37 kDa. Int is composed of three domains: i) the N-terminal arm binding domain (AB) binds conserved, sub-terminal DNA sequences (arm sites) that are found inside the transposon, ii) the core binding domain (CB) binds to the specific transposon ends (also called inverted repeats, IR or core DNA sites) and to the target DNA sites during integration, and iii) the catalytic domain (CAT) at the C-terminus is responsible for all chemical reactions. CB and CAT domains are well conserved among tyrosine recombinases. The AB domain is only present in a subfamily of tyrosine recombinases, and it is not necessary for DNA recombination reactions *in vitro* (Rubio-Cosials *et al.*, 2018).

Recent studies in our lab suggested that Tn*1549* transposition occurs via the following overall pathway (see Figure 1-4): i) it starts with Int-mediated recombination at the transposon ends excising a sealed circular intermediate (CI) with a heteroduplex crossover region of 5-7 bp length linking the two transposon ends. ii) Conjugative transfer then carries the CI to a new recipient cell as single-stranded DNA. There, it gets replicated to create a double-stranded CI with a homoduplex crossover region (CR). iii) Finally, Int catalyzes the integration of the CI at random locations in AT-rich target DNA (consensus sequence TTTT-N_6-AAAA) (Lambertsen *et al.*, 2018; Rubio-Cosials *et al.*, 2018).

*Table 1-3: Overview and description of ORFs from Tn1549. The table was updated from (Garnier et al., 2000). Protein sequences were submitted for a BlastP search against non-redundant protein sequences (nr) to look for updates or new information related to characterized protein domains that may show similarities to uncharacterized ORFs in the conjugative transposon Tn1549 (highlighted with *, GenBank accession number AF192329.1).*

ORF	Size (aa)	Description
ORF 13	400	Predicted helix-turn-helix (HTH) domain, which is similar in *Agrobacterium tumefaciens* plasmid pTi*
ORF 14	71	It has a domain of unknown function (DUF5348)*
ORF 15	159	It has a domain of unknown function (DUF3801)*
ORF 16	565	TrsK-like protein (similar to VirD4, a coupling protein in Gram-negative bacteria using the T4SS for conjugation)
ORF 17	288	Unknown function
ORF 18	181	Low similarity to MunI methyltransferase of *Mycoplasma sp.*
ORF 19	141	Similarity to PrgI protein family, which are transmembrane proteins and form part of bacterial T4SS*
ORF 20	800	Low similarity to TrsE protein, necessary for conjugative transfer of *S. aureus* plasmid pGO1
ORF 21	664	Similarity to Tn*916*, ORF14 shows conservation to the C40 peptidase family, which are bacterial cell wall modifying enzymes. Maybe it is necessary for opening the recipient bacterial cell membrane*
ORF 22	84	It has a domain of unknown function (DUF4315)*
ORF 23	410	It has a domain of unknown function (DUF4366), also a SH3 domain necessary for protein-protein interaction*
ORF 24	694	Low identity to DNA topoisomerase III from *B. subtilis*
ORF 25	1307	Low similarity with a domain of a LtrC-like protein that plays a role in the conjugation of plasmid pSK41 from *S. aureus*
ORF 26	315	Unknown function
ORF 27	149	Low similarity to Tn*916* ORF8, that regulates transcription of genes downstream of it
ORF 28	443	Similarity to the relaxase protein of *S. aureus*, necessary for relaxosome complex formation and conjugation
ORF 29	110	MobC-like protein, well conserved that plays a role in relaxosome complex formation as an auxiliary protein. There is also conservation with a gene in *A. tumefaciens* plasmid Ti*
ORF 30	124	HTH domain with unknown function*
vanRB	220	DNA-binding response regulator protein (OmpR family)
vanSB	447	Sensor protein
vanYB	268	carboxypeptidase
vanW	275	Vancomycin resistance accessory protein
vanHB	323	dehydrogenase
vanB	342	D-ala-D-lac ligase
vanXB	202	D-ala-D-ala dipeptidase
ORF 7-Tn*1549*	143	Domain similarity to sigma-70 family of RNA polymerase sigma factors (initiation factors)*
ORF 8-Tn*1549*	76	Predicted helix-turn-helix (HTH) domain, common for transcriptional regulatory proteins*
Xis-Tn*1549*	66	Excisionase (Xis), accessory protein, plays a role in the directionality of recombination reaction, promotes excision
Int-Tn*1549*	397	Integrase (Int), tyrosine recombinase, responsible for excision and integration reactions of Tn*1549*

1.6.1 The structure of a Tn*1549* Int-CI DNA complex

Recent work in the lab has revealed the structure of Int bound to a CI DNA (Rubio-Cosials *et al.*, 2018). The structure consists of the catalytically inactive and truncated protein variant Int^{82N}R225K (aa 82-397) attached to a 44 bp long DNA substrate that mimics the CI (Figure 1-16). The protein construct lacks the AB-domain and it has an arginine to lysine mutation at residue 225 that renders the protein inactive. This means, the protein cannot perform cleavage and strand exchange reactions because it is missing an important amino acid from the conserved catalytic pocket (see section 1.4.2). The DNA sequence contains 11 bp long segments of the left and right IRs, separated by a 5 bp homoduplex crossover region (CR). This specific sequence was called CI5. The 2.8 Å structure revealed an Int^{82N}R225K homodimer bound to one CI5 DNA molecule. It shows how each monomer recognizes and binds to one IR in a symmetric fashion despite some base pair differences in the inverted repeat sequences. The CR at the center is not bound by protein. Overall, binding is achieved via formation of a C-shaped clamp by CB and CAT domains around the CI DNA, which is typical for the tyrosine recombinase family.

The Int-fold resembles the structural fold of other tyrosine recombinases. The C-terminal helix αM is exchanged between the two monomers. Although a similar intermolecular swapping of C-terminal tails was also observed in other tyrosine recombinase structures, like in Cre, XerH, and λ Int, these occurred in tetrameric complexes cyclically. While here, the C-terminal tail interaction is reciprocal and plays an essential role in holding the dimer together.

Another unexpected part of the Int structure is a β-hairpin structure, found inserted in a loop region of the conserved CAT domain fold of tyrosine recombinases. Sequence analysis showed that this feature is shared within the Tn*916*-like family of transposases. The structure indicates that this hairpin interacts directly with CI DNA and helps induce a small bend. Biochemical assays showed that this hairpin is important for Int activity because deletion mutants showed decreased recombination activity *in vitro*.

Overall, the CI DNA has a nearly straight conformation in the structure. This is in contrast to previously characterized tyrosine recombinase synaptic complexes, which showed that strong DNA bending is necessary before DNA cleavage can occur. At the CR the DNA is melted, distorted and shows remarkable unwinding. This DNA distortion enables Int to recognize substrates with variable length and sequence at the CR. This DNA flexibility was supported by biochemical data and two additional structures, Int-CI6a (pdb: 6en1) and Int-

CI6b (pdb: 6en2), where the CI contained a CR of 6 bp length. This opening is achieved by inducing base flipping at each transposon end, which is pivotal for proper strand exchange (supported by biochemical assays). This happens by insertion of the R153 side-chain (CB-domain) at both transposon ends, leading to flipping of the first base on both sides of the CR. The flipped-out bases are stabilized via pi-stacking interactions with Y160 (CB-domain).

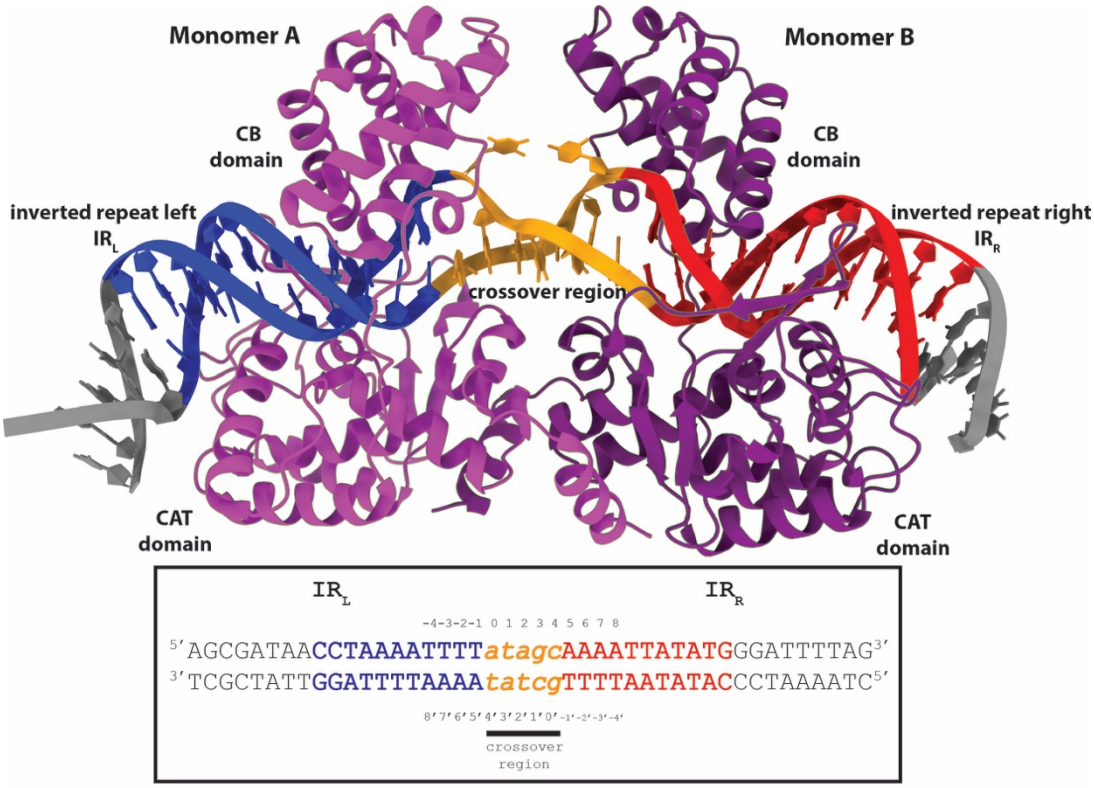

Figure 1-16: The crystal structure of the Int-CI5 DNA complex. Cartoon representation of the Int-CI5 structure (top, Int monomer A in light purple and monomer B in purple). Each monomer is composed of a core-binding (CB) and catalytic (CAT) domain that are binding a circular intermediate (CI) DNA, composed of inverted repeat left (blue) and right (red) core sequences connected by a crossover region (CR, in orange). DNA ends are colored in gray, pdb: 6emz. The sequence of the CI DNA is shown (bottom, same color code) (Rubio-Cosials et al., 2018).

With respect to the catalytic active site, Int shows an overall conserved arrangement with many conserved residues of the tyrosine recombinase family: R225 (mutated to K in the structure), H344, R347, and H370. However, one peculiar feature of Tn*916*-like family Int enzymes is the presence of two tyrosines near the active site, Y379 and Y380. The structure shows that Y379 is closer to the scissile phosphate than Y380, but it is still too distant for performing catalytic activity. In turn, Y380 is hidden in a hydrophobic pocket.

Notably, biochemical data showed that single Y379F or Y380F mutants can still cleave DNA, but Y380F cannot proceed to strand exchange *in vitro*. Consistently, *in vivo* experiments showed that the Y379F mutant could perform excision, while Y380F could

not excise the CTn Tn*1549* (Lambertsen *et al.*, 2018). Thus, Y380 seems to be essential for performing the strand exchange reactions.

Additionally, the structure showed that the catalytic pocket is positioned around the phosphate backbone one nucleotide inside the transposon end. Subsequent biochemical data confirmed that the DNA cleavage occurs at this position. This particular cleavage position creates a single TA base-pair complementarity with the target DNA and helps find the ligation site for strand re-joining. These features appear to be conserved in Tn*916*-like conjugative transposons, likely helping to overcome non-homology of their substrates, thus allowing these elements to integrate at various genomic sites.

However, these structures showed a pre-synaptic complex, with one protein dimer bound to a single DNA substrate in an autoinhibited conformation, leaving many critical questions about the subsequent reaction steps.

1.6.2 Importance of the C-terminal helix in Int activity

The Int-CI5 structure revealed that the αM-helix at the Int C-terminus plays a vital role in complex stabilization through interactions with the surface of the neighboring Int monomer (Figure 1-17). This C-terminal helix-partner surface interaction locks the protein dimer in an inactive conformation due to positioning the proposed catalytic Y380 far from the active site and further hindering the formation of a synaptic complex. Moreover, residues involved in these interactions are mostly conserved in the Tn*916*-like family of transposases.

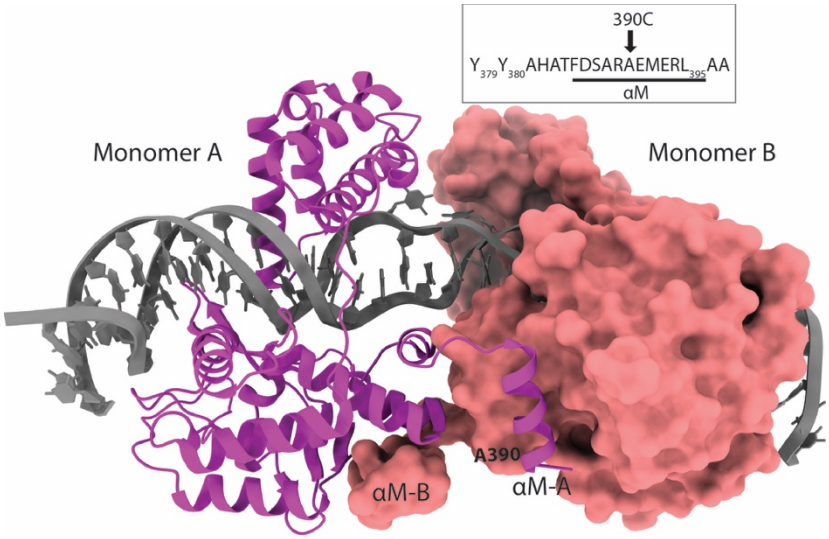

Figure 1-17: The interaction of the C-terminal αM helix in the pre-synaptic Tn1549 structure. Monomer A (light purple, cartoon representation), monomer B (pink, surface representation), and DNA (grey). The location of the C-terminal truncation used in biochemical experiments is highlighted in the model structure and in the sequence. This truncated protein variant was called Int^{82N}390C, pdb: 6emz (Rubio-Cosials et al., 2018).

In accordance to these observations, further experiments using a C-terminally truncated Int variant (Int^{82N}390C) showed increased cleavage and strand exchange activity *in vitro* compared to the wild type protein, probably due to Y380 being more flexible to move and perform the necessary recombination reactions. These results indicate a regulatory feature of the CTn Int family, highlighting the C-tail's importance in stabilizing a pre-catalytic dimer complex. Probably, this autoinhibitory state may help in hindering undesired cleavage of the CI site until a proper target site sequence has been found in order to proceed to synaptic complex formation and DNA recombination (Rubio-Cosials *et al.*, 2018).

1.6.3 Model for Tn*1549* integration and its comparison to site-specific recombination

The above studies proposed that Tn*1549* integration occurs via the following pathway: It starts with Int-mediated recombination between CI and target site sequences that have non-homologous crossover region (CR) sequences. For that, CI and AT-rich target sites are recognized by Int dimers, through indirect sequence readout and specific interactions between the β-hairpin structure and the inverted repeats in the CI or similar sequences in the target site. These binding interactions lead to DNA distortion and opening at the crossover region through base flipping. Afterwards, we expect that conformational changes will occur upon synaptic complex formation, probably involving the C-terminal protein segments. This will lead to rearrangements at the catalytic sites, which activate the protein. Next, Int cleaves one DNA strand at the CI and one at target sequences, creating two 3'-phosphotyrosyl bonds and freeing two 5'-hydroxyl groups. For that, it cleaves at the scissile phosphate between two thymines, one nucleotide inside the inverted repeats in the transposon ends or the target site. The free 5'-OH groups will then attack the 3'-phosphotyrosyl bonds on the partner DNA molecule and resolve the Int-DNA linkages. The CI 5'-OH will go to the target site's 3'-phosphotyrosyl bond and the target 5'OH to the CI 3'-phosphotyrosine. Base pairing of the terminal thymine in the incoming 5' strand with the unpaired adenine at the uncleaved strand helps correct strand exchange and promotes the ligation reaction. This reaction generates a HJ intermediate that will be resolved through a second-round of strand cleavage and exchange reactions resembling the first ones. This time the reactions are performed by the two monomers that were inactive in the first round, leading to transposon integration. The integrated transposon will contain heteroduplex regions of 5-7 bp length between the transposon ends and flanking bacterial DNA sequences, which may be repaired later by the DNA repair machinery or through DNA

replication. We assume that the excision reaction will follow a similar pathway (Lambertsen *et al.*, 2018; Rubio-Cosials *et al.*, 2018).

Comparing the available biochemical and structural data on the Tn*1549* integrase with site-specific tyrosine recombinases some clear parallels can be identified: i) A similarity from a sequence and structural point of view is the conserved CAT domain fold and catalytic pocket, including the same essential residues. ii) A tyrosine nucleophile catalyzes the DNA cleavage and rejoining reactions without the need for other energy sources (like cofactors) in all cases. iii) Both systems cleave DNA at similar positions with in a 5-7 bp staggered manner.

Nevertheless, many key features of the conjugative transposition process remain unclear. For instance, previous mechanistic insights into canonical recombinases cannot explain how conjugative transposons can integrate to many sites in diverse genomes, although this is the key for their broad spreading across an incredibly diverse range of bacteria. In both systems 5-8 bp at the crossover region are exchanged during the recombination reaction. However, in the canonical system the exchanged sequences maintain homology, whereas conjugative transposons create a heteroduplex region after the excision and integration reactions. The Int-CI structures revealed that active DNA melting promotes such process, but it will be very interesting to see the differences in HJ DNA structure in the two systems. The structure of Tn*1549* Int in an autoinhibitory state was surprising and different from all other site-specific recombinase structures obtained so far. λ Int, Cre, and XerH pre-cleavage complexes were always trapped in a synaptic tetrameric conformation. It will be exciting to see how synaptic complex formation occurs in Tn*1549* and what role Int's C-tail plays in this process.

Therefore, understanding Tn*1549* transposition still requires structural visualization of the proposed steps, including a synaptic complex with covalently bound DNA and the Int-HJ state, which would help to clarify many open questions of the conjugative transposition process.

1.7 Aims and objectives of this study

The thesis focuses on studying the integration's mechanism of the CTn Tn*1549* from *E. faecalis*. These insights will be critical to help understanding conjugative transposon-mediated ABR transfer and may open up novel avenues to limit ABR's further spread among bacteria in the future. I aim to elucidate the unknown steps of conjugative transposition, using an integrated approach consisting of (i) biochemical characterization of elementary transposition steps *in vitro* and (ii) high-resolution structure determination of Int-DNA complexes trapped at critical intermediary steps of the pathway.

My specific aims are:

- Structural characterization of specific Int-DNA complexes to elucidate unknown steps in the transposon integration reaction
- Molecular and structural characterization of the accessory protein Xis
- Biochemical characterization of HJ resolution (*in vitro*)
- Molecular characterization of an Integrase Inhibitor

2 RESULTS

2.1 Structural and biochemical characterization of Tn*1549* transposon integration

The crystal structure of the Tn*1549* Int in complex with a DNA substrate mimicking the excised transposon CI (hereafter referred to as the Int-CI structure) revealed a protein homodimer bound to a single DNA substrate in a pre-integration state (see section 1.6.1) (Rubio-Cosials *et al.*, 2018). While this first structure provided important insights into the early steps preceding integration, it did not contain target DNA, leaving the mechanistic principles of target recognition and integration unknown. Furthermore, it was clear that the complex needed significant conformational rearrangement for catalytic activation and its unexpected dimeric state raised the question to what extent conjugative transposition resembles the recombination pathway of tyrosine recombinases. Therefore, I investigated the biochemical pathway of Tn*1549* integration and aimed to identify conditions for trapping protein-DNA complexes during integration for subsequent structure determination.

2.1.1 Characterization of a catalytically active Int-DNA complex

To capture and crystallize Tn*1549* Int in a catalytically active state before integration, I first aimed to reconstitute a post-cleavage protein-DNA complex. For that, I used so-called suicide variants of CI DNA substrates (sCI; for more details, see section 2.1.1.2) and the previously characterized protein constructs Int^{82N} and Int^{82N}390C (Rubio-Cosials *et al.*, 2018).

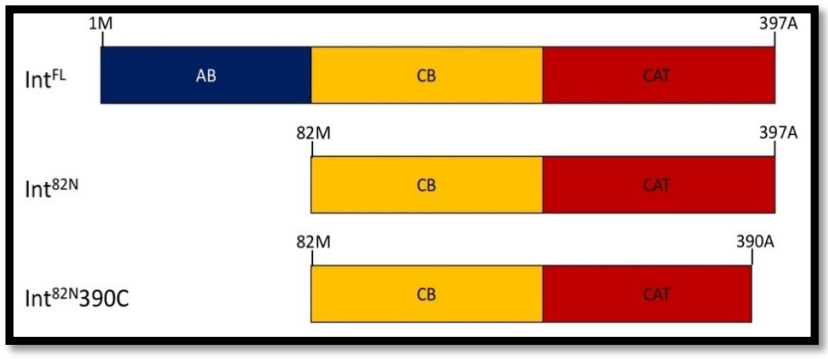

Figure 2-1: Schematics of the Tn1549 Int constructs. Int^{FL} starts at amino acid (aa) 1M and ends at aa 397A. Int^{82N} starts at aa 82M and ends at aa 397A. Int^{82N}390C starts at aa 82M and ends at aa 390A. The N in Int^{82N} stands for N-terminally truncated. The C in Int^{82N}390C stands for C-terminally truncated. AB = arm-binding domain, CB = core binding domain, CAT =catalytic domain.

Int^{82N} is a truncated version of Int that is missing the AB domain responsible for arm DNA binding (Figure 2-1). This variant was chosen because it contains all essential domains for recombination *in vitro* and is more stable in complex with DNA compared to the full length Int (Int^{FL}). The Int^{82N}390C variant (Figure 2-1) additionally lacks part of the C-terminal helix, which plays an essential role in protein dimerization and autoinhibition, as seen in the previous Int-CI crystal structure (Rubio-Cosials *et al.*, 2018). Due to weakened autoinhibitory interactions, Int^{82N}390C has a hyperactive phenotype in DNA cleavage and recombination *in vitro* (Rubio-Cosials *et al.*, 2018).

2.1.1.1 Expression and purification of Int^{82N} and Int^{82N}390C

Int^{82N} and Int^{82N}390C constructs were overexpressed in *E. coli* as described in section 4.2.9.1 and purified as described in section 4.2.9.2. I obtained around 5 mg protein for each construct in high purity after size-exclusion chromatography (Figure 2-2). I used the peak fractions for biochemical assays and complex formation for further crystallization experiments.

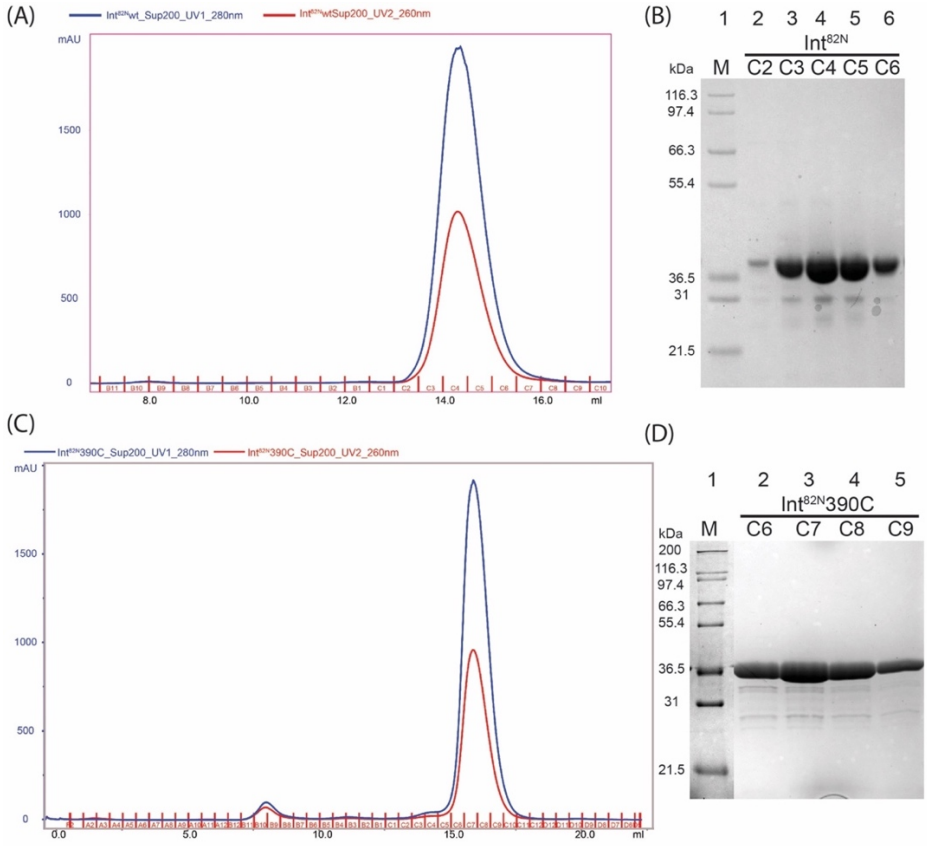

Figure 2-2: Typical purification of Int^{82N} and Int^{82N}390C constructs. The last step of (A) Int^{82N} and (C) Int^{82N}390C purification: Size-exclusion-chromatography (SEC) on a Superdex^{TM} 200 gel filtration column. The chromatogram shows the prominent elution peak that corresponds in size to a dimer of Int^{82N} (72 kDa) or a monomer of Int^{82N}390C (35 kDa). UV absorbance at 280 nm is shown in blue, absorbance at 260 nm is shown in red. (B) and (D) SDS-PAGE analysis of the fractions collected from SEC step, stained with Coomassie Blue for protein visualization. Marker (Mark12^{TM}) positions are labelled on the left in kDa.

2.1.1.2 Design of suicide DNA substrates

In general, suicide DNA substrates have this name because they stall recombination at specific reaction steps. Here, I used suicide substrates that contain a nick in one of the DNA strands some nucleotides after the cleavage site. After DNA cleavage, the nucleotide(s) coming after the cleavage site will diffuse away, leading to a trapped covalent protein-DNA intermediate. The free 5'-OH of the cleaved DNA-strand will be too short to either reverse the reaction by attacking back at the 3'-phosphotyrosine-bond on the same strand or to reach the other protein-DNA linkage at the recombination partner molecule for subsequent strand exchange.

First, I designed suicide DNA substrates based on the CI sequence (called sCI, Figure 2-3A), which mimics the junction of the excised transposon intermediate. The CI sequence comprises the transposon ends (inverted repeat left, IR_L and inverted repeat right, IR_R) joined by a crossover region (CR) in the middle that can be 5 or 6 nucleotides long. For these substrates, the nick was placed 2 nt downstream of the cleavage site (Figure 2-3B). I also tested the addition of "flap" sequences of 1 or 2 nt length. These extended sequences at the 5' end of the second fragment of the nicked strand should occupy the free space left behind the diffused nucleotides after cleavage occurred and may help to stabilize the resulting complex. The 5' end of the flaps were phosphorylated to prevent any further reaction to occur.

Second, I designed and tested different palindromic suicide CI substrates based on the IR_L-sequence (IR_L-CR-IR_L) or the IR_R-sequence (IR_R-CR-IR_R), with a crossover region of 6 nt length, called palCI6 (Figure 2-3C). These palCI sequences had two nicks placed either 1 or 2 nt after the cleavage sites. The idea of having 2 nicks in the substrate helped in obtaining the crystal structure of the post-cleavage XerH-*dif*$_H$LP synaptic complex (Bebel *et al.*, 2016). Therefore, I added a nick near the cleavage sites in both strands to increase the percentage of covalent intermediate product formation. Furthermore, I also tested substrates with different 5'-phosphorylated flaps. Finally, for crystallization experiments, I added a T-T mismatch or a TA-overhang to selected oligos to promote formation of DNA mediated crystal contacts (Figure 2-3A and D). The T-T mismatch was helpful for obtaining crystals of the Int-CI5 complex (Rubio-Cosials *et al.*, 2018).

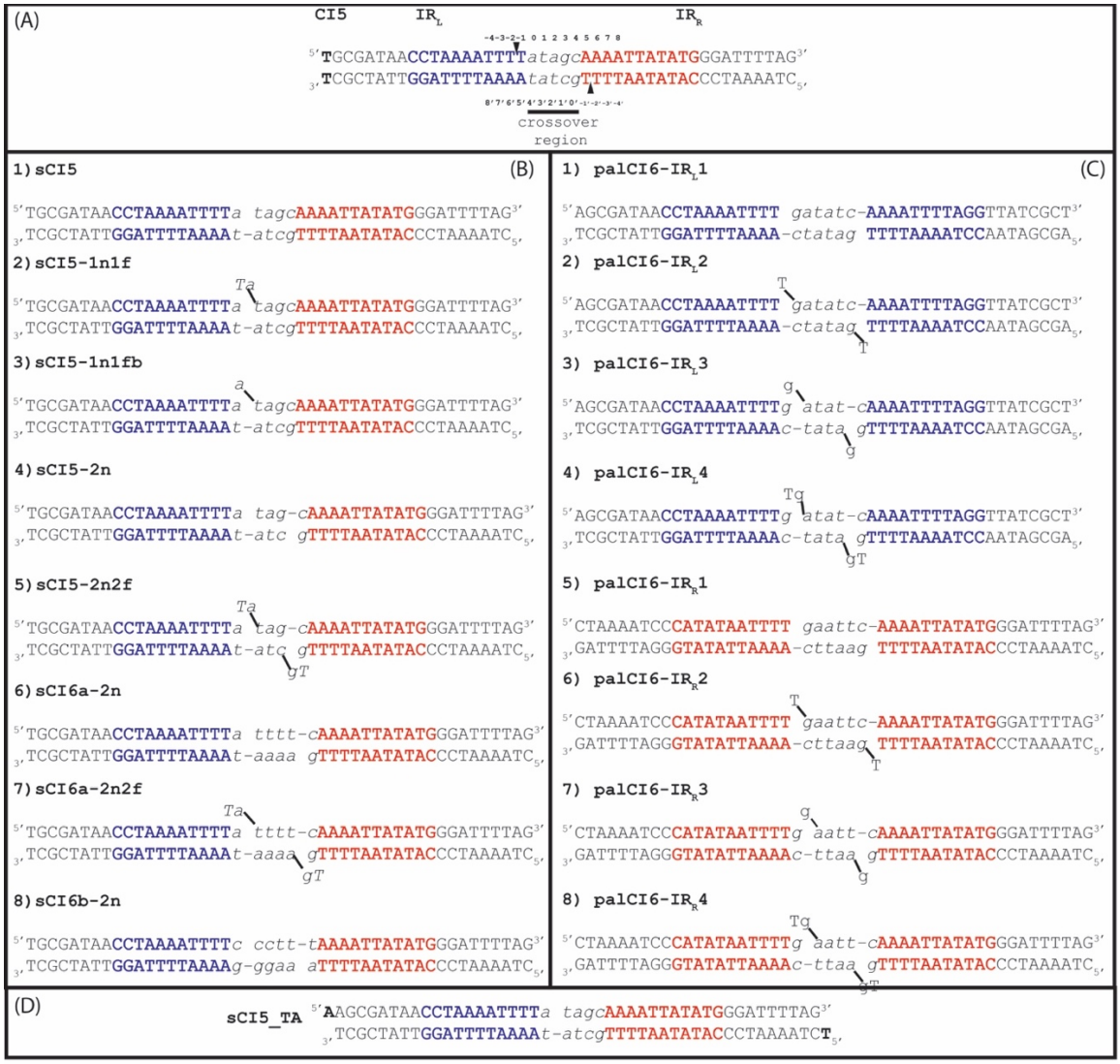

Figure 2-3: Suicide DNA substrates used for activity assays and crystallization experiments with Int^{82N} and Int^{82N}390C. (A) Circular intermediate sequence (CI) composed of IR$_L$ (blue), crossover region (5 or 6 nt long, black, lowercase and italic) and IR$_R$ (red) with a T-T mismatch (bold) at IR$_L$, as it was used before for crystallization experiments in (Rubio-Cosials et al., 2018). The cleavage position is highlighted with black arrowheads, between -2 and -1 position at the top strand of IR$_L$ or between -1' and -2' position at the IR$_R$'s bottom strand. (B) List of suicide sCI5 and sCI6 substrates used in biochemical assays. They varied in the crossover region's length with either 5 or 6 nt length; the nick was always positioned 2 nt downstream of the cleavage, indicated with a gap in the sequences. In some cases, an additional flap sequence of 1 or 2 nt length was used on the nicked strand to fill the gap after cleavage. (C) Additional palindromic (pal) suicide substrates used to test cleavage activity. The substrates were based on either the IR$_L$ (blue) or IR$_R$ (red) sequences, with a 6 nt long crossover region, the nick position was either 1 or 2 nt after the cleavage site, and the flap was 0, 1 or 2 nt long. (D) The sequence of the sCI5_TA substrate used for cleavage and crystallization experiments. It contains a T-overhang at the 5' end and an A-overhang at the 5' end to promote crystallization.

2.1.1.3 DNA cleavage activity of Int^{82N} and Int^{82N}390C on suicide substrates

To find the most suitable conditions to trap a post-cleavage Int-DNA complex for crystallization experiments, I tested the ability of various sCI and palCI substrates (Figure

2-3) to produce stable covalent protein-DNA intermediates with Int^{82N} and Int^{82N}390C (see assay design in section 4.3.5). First, I tested Int^{82N} and Int^{82N}390C on sCI5 and sCI6 substrates. These experiments showed that both proteins cleaved all substrates and produced some amount of covalent complex. The activity of Int^{82N}390C was higher than that of Int^{82N}, as expected due to lack of the autoinhibitory C-terminal helix (Figure 2-4A). Using DNA binding assays, I also showed that In^{82N}390C forms higher oligomeric complexes with these suicide substrates, while Int^{82N} binds most probably only as a dimer (Figure 2-4B).

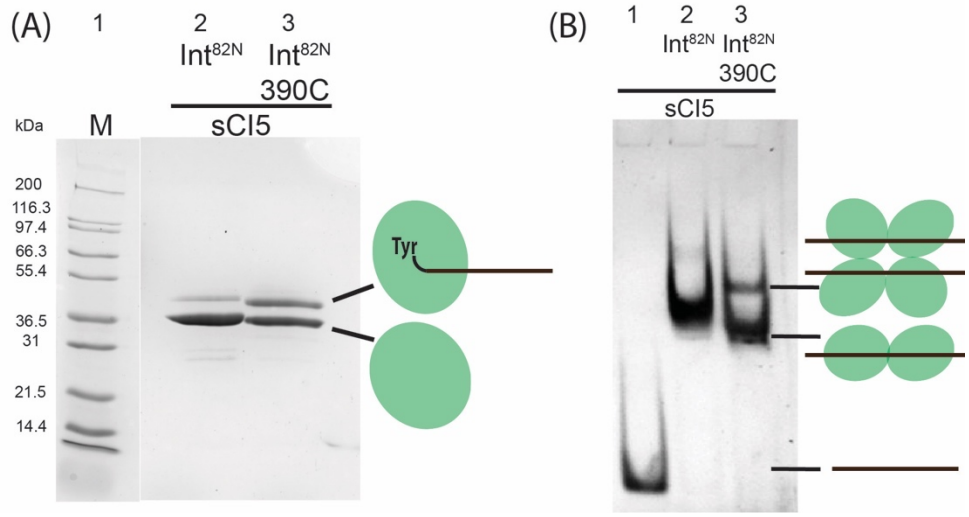

Figure 2-4: Probing Int^{82N} and Int^{82N}390C covalent intermediate formation with suicide substrates. (A) Cleavage assays with Int^{82N} and Int^{82N}390C on suicide DNA. Upon cleavage, Int becomes covalently attached to the suicide CI DNA substrate (green oval with black DNA piece attached to the catalytic tyrosine residue), which can be separated from the unreacted protein (green oval alone) on denaturing SDS-PAGE. The DNA substrate tested here is sCI5 (Figure 2-3B1). (B) Native PAGE of Int^{82N}-sCI5 (lane 2) and In^{82N}390C-sCI5 (lane 3) complexes. sCI5 alone runs faster as a single band (lane 1), while Int^{82N}-sCI5 runs slower also as a single band. Int^{82N}390C-sCI shows two bands that could stand for a dimer- and a tetramer-complex (as indicated with the schematics).

However, covalent intermediate formation remained partial for all protein and sCI DNA variants. Next, I tested Int^{82N} and Int^{82N}390C cleavage activity on various palCI sequences (Figure 2-3C) to find a substrate for Int^{82N} that may lead to higher product formation. Experiments with Int^{82N} and palCI6 showed no or low-level covalent product formation (Figure 2-5A-C). Thus, none of the palCI substrates were used with Int^{82N} for further crystallization experiments. With Int^{82N}390C, palCI6 cleavage experiments showed no covalent product formation (Figure 2-5A, lanes 8 and 9), low rate of covalent intermediate formation (Figure 2-5D, lanes 3 and 4), or unspecific cleavage and product formation (Figure 2-5C, lane 3 and D, lanes 3-6). Only the palindromic substrate based with the IR$_L$-sequence, two nicks and two flaps (palCI6-IR$_L$-4) gave the expected covalent intermediate product, with a single band on an SDS-PAGE gel (Figure 2-5D, lane 7).

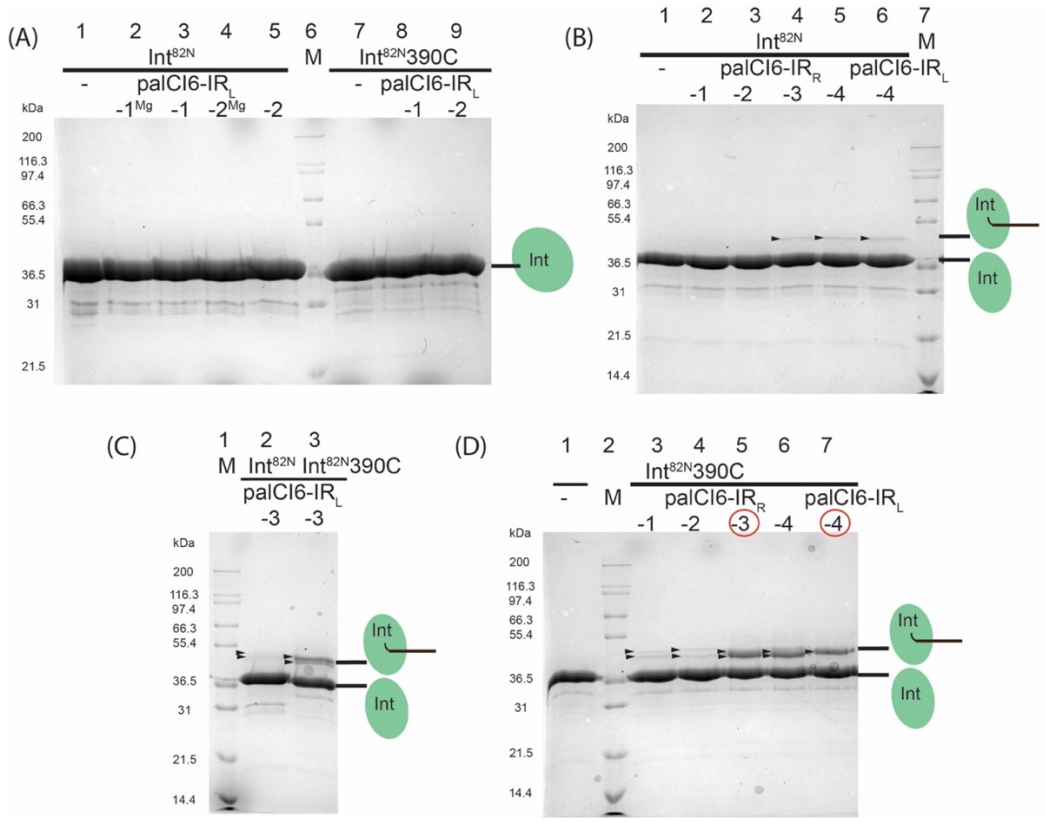

Figure 2-5: Probing Int^{82N} and Int^{82N}390C covalent intermediate formation with palCI6 substrates. SDS-PAGE gels showing Int^{82N} and In^{82N}390C cleavage on various palCI6 substrates. Upon cleavage, Int becomes covalently attached to the suicide CI DNA substrate (green oval with DNA piece in brown), which can be separated from the unreacted protein (green oval alone) on denaturing SDS-PAGE. The DNA substrates tested here are different palindromic suicide substrates; see Figure 2-3. M = protein marker (Mark12^{TM}), - = negative control with protein alone. (A) palCI6-IR$_L$-1 and -2 substrates were tested and don't show any covalent intermediate product for either protein variant (lanes 2-5 and 8-9). For Int^{82N}, an activity buffer with 10 mM $MgCl_2$ (final concentration) was tested without positively affecting the reaction (lanes 2 and 4). (B) Int^{82N} cleavage activity was tested on palCI6-IR$_R$-1, 2, 3, and 4, and palCI6-IR$_L$-4 substrates. Low level of product formation could be observed for palCI6-IR$_R$-3 (lane 4), palCI6-IR$_R$-4 (lane 5), and palCI6-IR$_L$-4 (lane 6). (C) Int^{82N} and Int^{82N}390C cleavage activity was tested on palCI6-IR$_L$-3 substrates. Low level of product formation could be observed by using Int^{82N} (lane 2); for Int^{82N}390C, there was more covalent intermediate product (lane 3). However, for both unspecific cleavages occurred, because at least two product bands were detected. (D) Int^{82N}390C cleavage activity was tested on palCI6-IR$_R$-1, 2, 3, and 4, and palCI6-IR$_L$-4 substrates. Only little product formation could be observed for palCI6-IR$_R$-1 and -2 (lanes 3 and 4), while product formation for palCI6-IR$_R$-3 and 4 (lanes 5 and 6) and palCI6-IR$_L$-4 (lane 7) was higher. Here also unspecific cleavage and product formation could be observed for all substrates, except for palCI6-IR$_L$-4.

2.1.1.4 Crystallization of various Int^{82N}-sCI and Int^{82N}390C-sCI complexes

After probing covalent intermediate formation with various Int^{82N}- and Int^{82N}390C-sCI/palCI combinations, I chose the complexes that showed higher covalent product formation for further crystallization experiments. First, I tested the Int^{82N}-sCI5 complex (Figure 2-4A, lane 2) for crystallization. I prepared complexes and performed high-throughput crystallization screens as described in sections 4.4.2.1 and 4.4.2.2. The most promising hits were manually optimized (see section 4.4.2.2). Best crystals appeared after

a few days and were shaped as small needles or plates form. After cryoprotection, their diffraction was tested as described in section 4.4.2.3. The best diffracting crystal grew in condition 0.2 M NaCl, 0.1 M Bis-tris pH 5.7, 27.5% (w/v) PEG 3350 at 20 °C and diffracted to a maximum of 4 Å. However, the diffraction pattern showed overlapping diffraction lattices, likely due to the multimeric nature of the crystals (Figure 2-6A). To test if the crystals contained both protein and DNA, I ran some of them on an SDS-PAGE gel and visualized with silver-staining. The gel showed only one single band for the protein alone and not the expected band of the covalent intermediate (Figure 2-6B). Thus, I assumed that the crystals did not capture the intended post-cleavage complex. Unfortunately, further manual crystal optimization also did not lead to single crystals. Therefore, I stopped further optimization of the crystals for this specific complex.

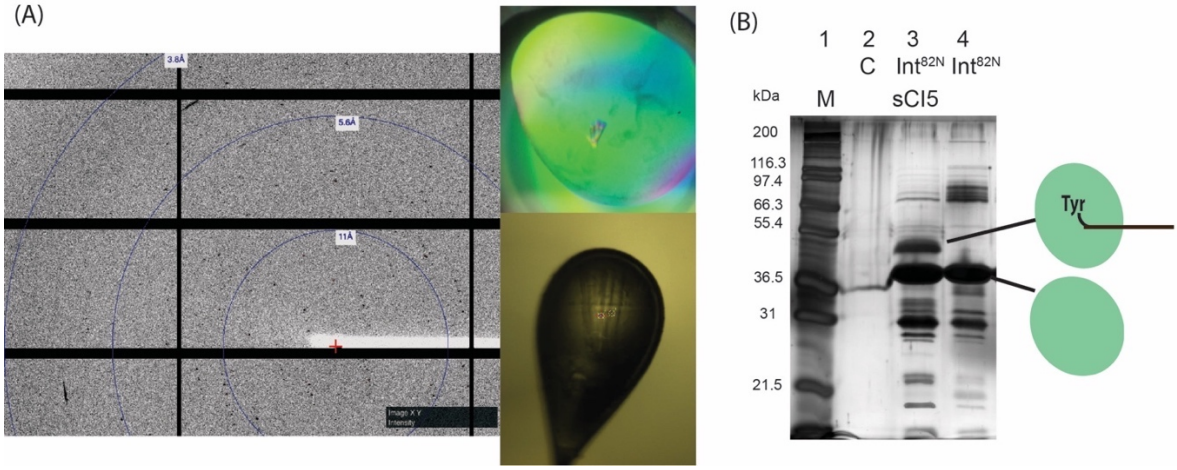

Figure 2-6:Crystallization and X-ray diffraction of the Int^{82N}-sCI5 complex. (A) Diffraction image of Int^{82N}-sCI5 crystals. The crystals were grown at RT as described in section 4.4.2.2 (A-top right corner). The multimeric crystal diffracted up to 4 Å resolution. Image of the same crystal in the frozen loop (A-bottom right corner). (B) SDS-PAGE of Int^{82N}-sCI5 crystals. Crystals from the Int^{82N}-sCI5 complex were washed and dissolved in water and run on an SDS-PAGE gel (lane 2) with the Int^{82N}-sCI5 complex (lane 3) and Int^{82N} alone (lane 4) as controls. Protein and DNA were visualized with silver staining.

Second, I tested the Int^{82N}390C-sCI5 complex for crystallization (Figure 2-4A, lane 3). For that, I prepared the complex, as described in section 4.4.2. I tested different complex concentrations in high-throughput crystallization trials, using various commercial screens listed in Table 6-3 (appendix), both at 7 °C and 20 °C. Nevertheless, this complex didn't yield any initial crystal hits.

Third, I tested the Int^{82N}390C-palCI6-IR$_L$-4 and the Int^{82N}390C-palCI6-IR$_R$-3 complexes for crystallization (Figure 2-5D, lanes 5 and 7). These complexes produced the highest amounts of covalent complexes in my biochemical tests, as shown above. palCI6-IR$_L$-4 showed high amounts of product formation but with some unspecific bands at the expected

size (Figure 2-7A, lane 4). palCI6-IR$_R$-3 showed two product bands, but the unspecific product band was very weak (Figure 2-7A, lane 8). I prepared the complexes, as described in section 4.4.2 and set up high-throughput crystallization trays using the commercial screens listed in Table 6-3, but none of the complexes yielded any initial crystal hits.

Finally, I went back to the sCI5 design, and added a TA-overhang to its ends (Figure 2-3D), which may help in crystallization by creating packing interactions and/or stabilizing crystal contacts. With this complex, I observed ~50% covalent protein-DNA product, as shown in Figure 2-7B, lanes 7 and 8. The Int^{82N}390C-sCI5_TA complex was prepared and used for high-throughput crystallization trials using various commercial screens listed in Table 6-3, as described in section 4.4.2. I obtained a single initial hit, which I optimized as described in 4.4.2.2. Resulting crystal hits (Figure 2-7B) were tested for diffraction (as described in section 4.4.2.3), but no diffraction was observed. As a control experiment, I tested crystal growth with sCI5_TA-DNA alone in the same conditions and observed DNA crystals after a few days. This indicates that my initial hits with the Int^{82N}390C-sCI5_TA complex contained DNA alone without the Int^{82N}390C protein. Thus, I stopped further optimization of these crystals.

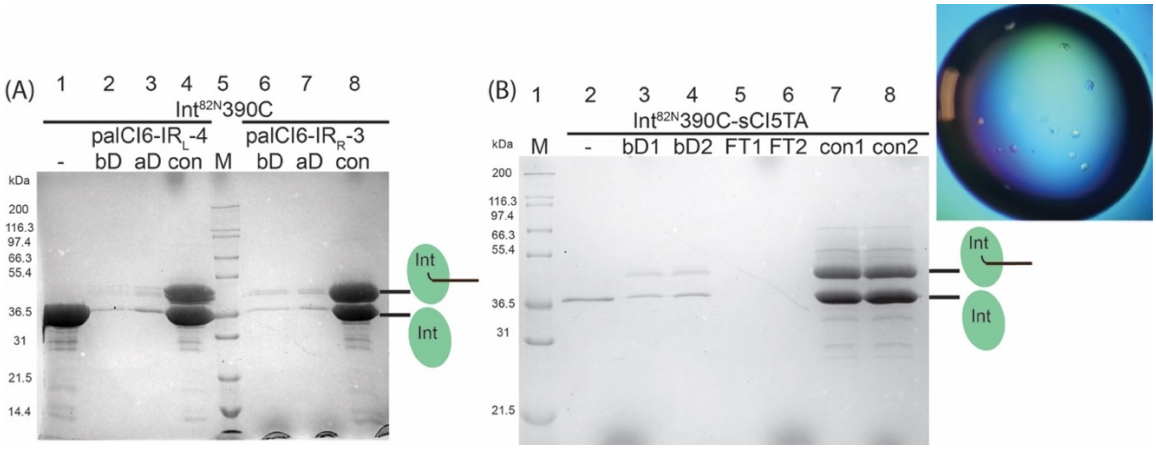

Figure 2-7: SDS-PAGE analysis of various In^{82N}390C-DNA complexes that were used further for crystallization experiments. M = protein marker (Mark12^{TM}), - = negative control, protein alone, bD = before dialysis, aD = after dialysis, con = concentrated complex, FT = concentrator flow-through. (A) Int^{82N}390C cleavage activity on palCI6-IR$_L$-4 and palCI6-IR$_R$-3 substrates reached 50%. (B) Int^{82N}390C cleavage activity on sCI5_TA substrate reached 50% with few unspecific cleavage products. Image of Int^{82N}390C-sCI5_TA complex crystals (right corner) in manual drops with "flower" shape.

2.1.1.5 Discussion

2.1.1.5.1 Approaches to crystallize Int in an active post-cleavage state

In my first aim, I wanted to structurally characterize specific Int-DNA complexes by X-ray crystallography to elucidate unknown steps in the transposon integration reaction. In this specific part of my thesis, I tested trapping an Int-DNA complex in an active post-cleavage state. As discussed in section 1.6.3, this complex is expected to show a synaptic tetrameric assembly of Int with two DNA substrates, in which two monomers should be covalently bound to DNA, as predicted from the mechanism of the site-specific tyrosine recombinase family.

I reconstituted post-cleavage protein-DNA complexes using two different Int constructs and various DNA substrates. By systematically probing the activity of the Int^{82N} and Int^{82N}390C constructs in diverse conditions *in vitro*, I found that Int^{82N} had low cleavage activity (Figure 2-4). Nevertheless, initial crystallization trials showed that Int^{82N} could crystallize with nicked DNA substrates (Figure 2-6A). However, the crystals did not capture the post-cleavage state (Figure 2-6B). Thus, to optimize trapping, I further tested DNA variants with nicks at either the IR_L or IR_R and nicked palindromic repeats that consisted of two IR_L or IR_R sequences (Figure 2-3). This strategy was successfully used before for Cre and XerH complexes (Guo, Gopaul, and Van Duyne, 1997; Bebel *et al.*, 2016). Nevertheless, Int^{82N} activity was also low for these substrates (Figure 2-5).

Although the Int^{82N}-sCI5 crystals didn't trap the post-cleavage complex, it can still be that the protein was crystallized in a DNA bound state that differs from the Int-CI5 structure (Rubio-Cosials *et al.*, 2018). These crystals were the only ones that showed promising diffraction up to 4 Å from all crystallization attempts, including the ones using Int^{82N}390C. Thus, it may be interesting to further improve crystallization of this complex or test crystallization of sCI5 with an inactive Int mutant and see if they help to capture a pre-catalytic synaptic complex, as it was shown for Cre, λ Int and XerH recombinases (Guo, Gopaul and Van Duyne, 1997; Biswas *et al.*, 2005; Bebel *et al.*, 2016). This could be possible because nicks can help in DNA bending by reducing the necessary energy for this process and may help induce synaptic complex formation.

In parallel, I concentrated on testing the crystallization of the hyperactive truncated mutant Int^{82N}390C with different suicide substrates. Int^{82N}390C showed good activity with various suicide substrates in comparison to Int^{82N} *in vitro* (Figure 2-4 and Figure 2-7). Indeed, these substrates showed maximal covalent intermediate formation leading to 50% yield of

protein-DNA covalent intermediate (Figure 2-7). Nevertheless, Int^{82N}390C showed unspecific cleavage of the DNA substrates, reflected by the formation of different size covalent intermediate products (Figure 2-7). This indicates that the truncation at the C-terminal αM-helix leads to increased activity, but also results in reduced precision of the cleavage position. The reason for such unspecific cleavage may be that shortening of the αM-helix reduced the contacts with the surface of the neighboring protein monomer, leading to an increased flexibility of the C-terminal region.

The difficulties in crystallizing the post-cleavage complex using the Int^{82N}390C variant may lay in the flexibility of the truncated C-tail. In most tyrosine recombinases, the C-terminal helices play an essential role in holding the active synaptic tetramers together. If the interactions on this site are too weak, the synaptic complex may only form transiently to perform cleavage but may fall apart quickly after. This will lead to a mix of dimers and tetramers in the solution, which will reduce the probability of crystal nucleation events.

2.1.2 Structural characterization of the Int-HJ DNA complex

Another option to visualize an integration competent Int-DNA complex, apart from trapping the post-cleavage state with suicide DNA substrates, was to reconstitute an Int-HJ DNA complex. This complex constitutes an intermediate step during transposon integration, as proposed in (Rubio-Cosials *et al.*, 2018). For that, I designed an HJ DNA substrate based on the CI and an integration target site sequence (Figure 2-9) obtained from *in vivo* integration assays (Lambertsen *et al.*, 2018). The HJ-DNA contains four Int binding sites, which can help to drive the formation of a tetrameric state. To reconstitute a stable and homogenous Int-HJ complex, I used a catalytic mutant (R225K) that cannot perform DNA cleavage reactions. The protein also lacked the flexible AB domain, which was shown to be dispensable for Int activity *in vitro* (Rubio-Cosials *et al.*, 2018).

2.1.2.1 Expression and purification of Int^{82N}R225K

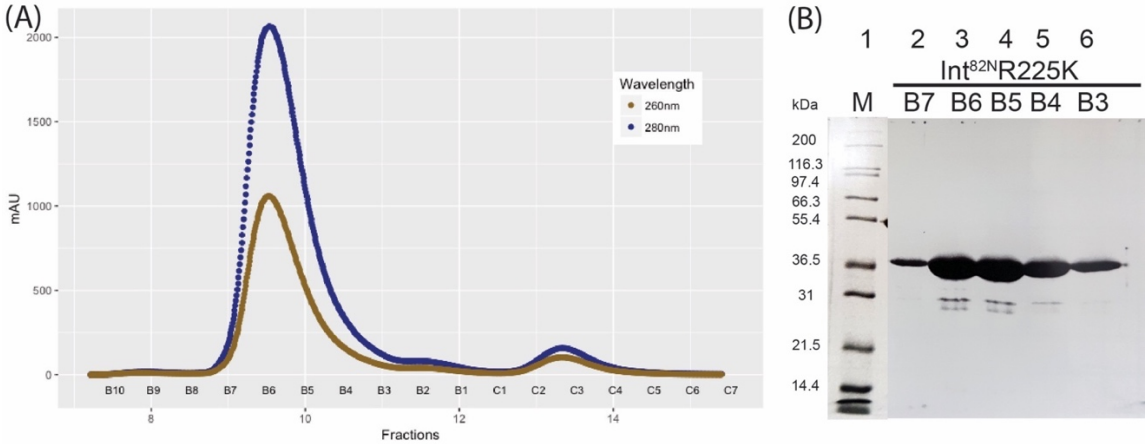

Figure 2-8: Purification of Int^{82N}R225K. (A) SEC on a Superdex^{TM} 75 column. The chromatogram shows a prominent elution peak corresponding to a dimer of Int^{82N}R225K (72 kDa). UV absorbance at 280 nm is shown in blue, absorbance at 260 nm is shown in red. (B) SDS-PAGE analysis of the fractions collected from the SEC step, stained with Coomassie Blue for protein.

The Int^{82N}R225K construct was overexpressed in *E. coli,* as described in section 4.2.9.1 and purified as described in section 4.2.9.2. I obtained around 5 mg protein in 99% purity after size-exclusion chromatography (Figure 2-8). The peak fractions were used further for DNA binding assays and crystallization experiments.

2.1.2.2 HJ DNA substrate design

The HJ-DNA comprises four stems, each containing one Int binding site. These binding sites include the IR$_L$) and IR$_R$ from the transposon CI and AT-rich sequences from the integration target, left (T$_L$) and right (T$_R$) from the insertion site. The central region of the

HJ was mutated to obtain a fully base-paired and stable HJ intermediate with the CR equally distributed among all the four stems. Similar designs were previously used for obtaining Cre- and λ Int-HJ structures (Gopaul, Guo and Van Duyne, 1998; Biswas *et al.*, 2005). This HJ was called HJ1 (Figure 2-10), which I used for initial binding experiments and crystallization trials.

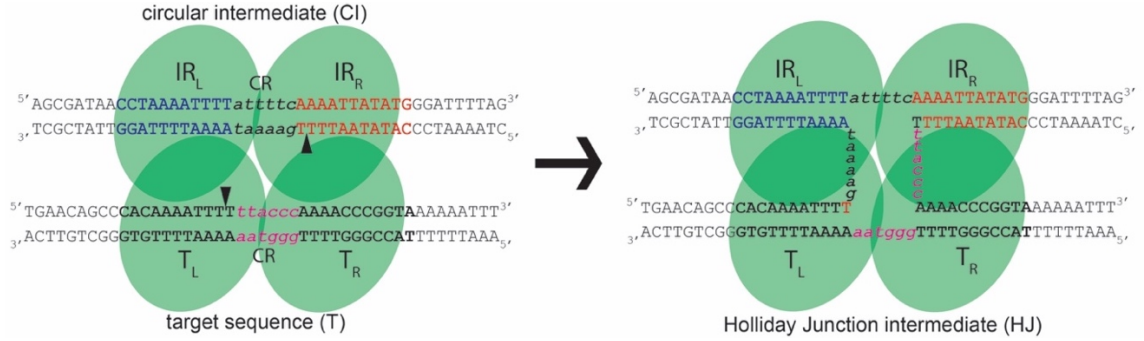

Figure 2-9: Schematics of HJ intermediate formation. The HJ intermediate is formed by recombination of a transposon CI and a selected target site (T) from (Lambertsen et al., 2018). The two stems of the CI, containing IR_L (blue) and IR_R (red), are connected via a 6nt CR (lowercase, italic) to the stems of the target site, T_L and T_R (bold). The scheme assumes that first cleavage and strand exchange occur at IR_R and T_L.

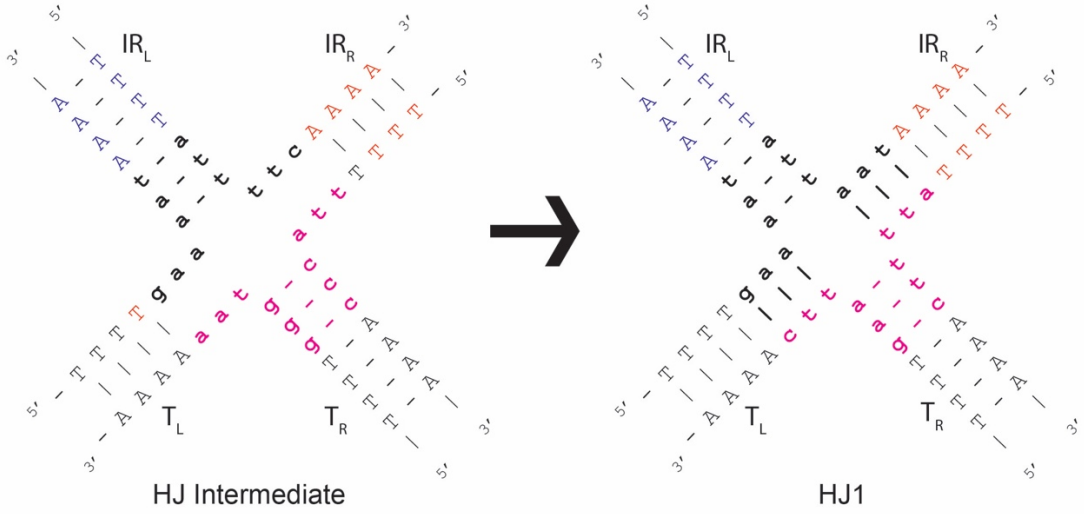

Figure 2-10: Design of the HJ1 substrate used for binding and crystallization experiments. Only the central region sequence is shown; the sequence continues in all four directions as shown in Figure 2-9. Bases at the center (lowercase) were mutated to obtain a stable and fully base-paired HJ.

Apart from this original HJ1 design, I designed different HJ1 variants to increase the crystallization probability of the Int-HJ complex. For example, I tested T or TA overhangs at the 5'-end in various configurations or added GA-mismatches to the ends of all four HJ-DNA stems (Figure 2-11A). Moreover, HJ substrates with a CR of 5 bp length were created

(Figure 2-11B). This substrate type should have significance *in vivo* because sequencing experiments have shown that Tn*1549* can integrate into target sites, where the CR is only 5 bp long (Lambertsen et al., 2018).

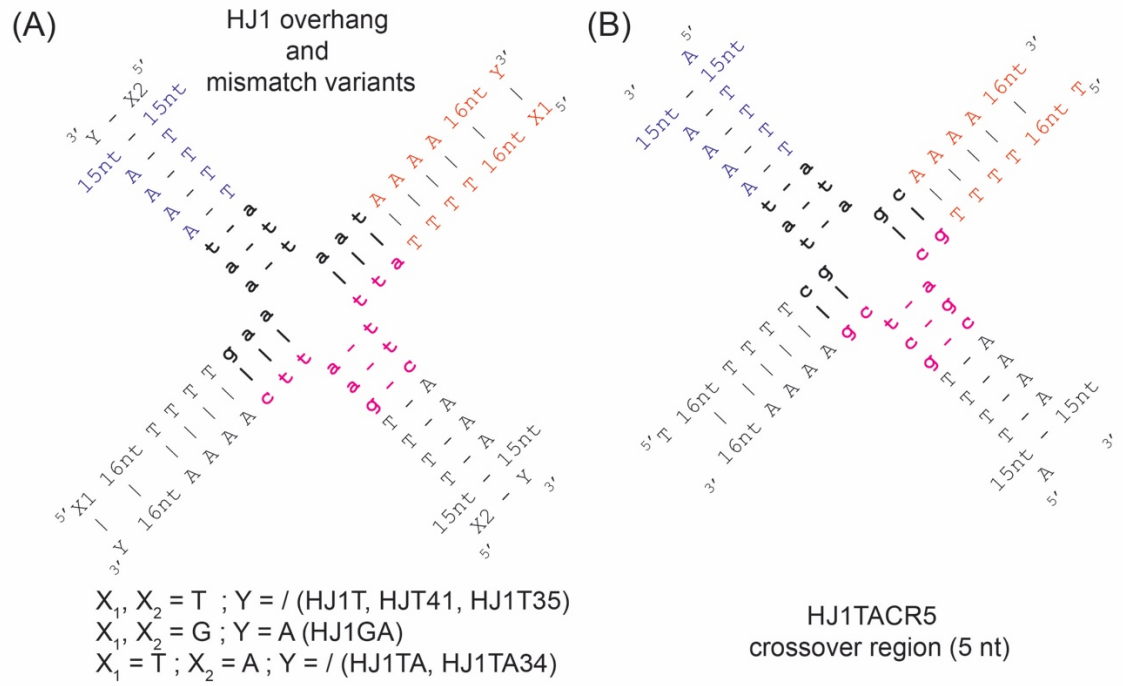

Figure 2-11: HJ1 variants used in this study. (A) HJs with different overhangs and mismatches at the HJ ends: Different overhangs were added to increase the crystallization probability by promoting base-stacking/base-pairing for better crystal packing and stability. Overhangs tested were 5'T-overhangs in all 4 stems (HJ1T) and alternating 5'T/5'A-overhangs (HJ1TA). Also, 5'G/3'A-mismatches in all 4 stems (HJ1GA) were tested. (B) HJ1 variant used in this study with a CR of 5 bp, called HJ1TACR5. The CR region is highlighted in bold, lowercase letters (black and magenta).

The HJ1 substrate design assumed that the first cleavage and strand exchange reactions occurred at IR_R and T_L. This assumption stemmed from earlier biochemical experiments with separated transposon end sequences, so called half sites (Rubio-Cosials et al., 2018). Nevertheless, the correct order of strand cleavage and exchange reactions has not yet been fully confirmed. Thus, I also designed a second set of HJ variants called HJ_L, which assumed a different order of strand cleavage and exchange reactions, with first reactions happening at IR_L and T_R (Figure 2-12 and Figure 2-13).

Figure 2-12: Schematics of HJ_L intermediate formation. The HJ is formed by recombination of a transposon CI and a selected target site (T) from (Lambertsen et al., 2018). The two stems of the CI, containing IR_L (blue) and IR_R (red), are connected via a 6nt CR (lowercase, italic) to the stems of the target site, T_L and T_R (bold). The scheme assumes that first cleavage and strand exchange occur at IR_L and T_R.

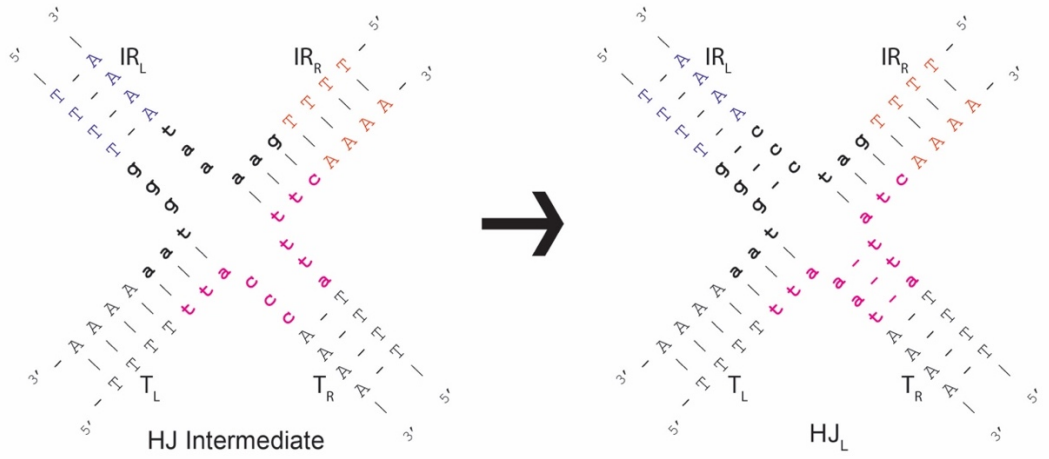

Figure 2-13: Design of the HJ_L substrate for binding and crystallization experiments. Only the central region sequence is shown; the sequence continues in all four directions as shown in Figure 2-12. Bases at the center (lowercase) were mutated to obtain a stable and fully base-paired HJ.

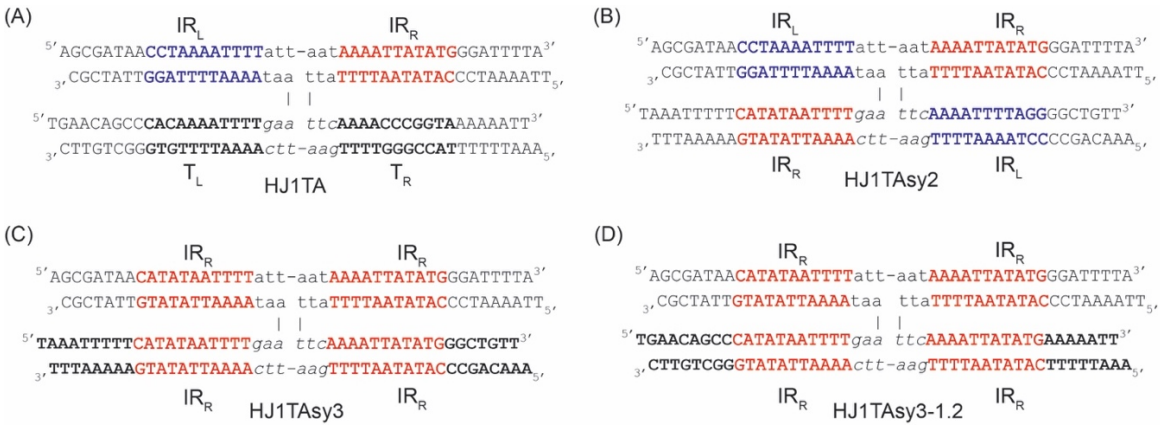

Figure 2-14: Various symmetric HJ designs used in this study. The target sites were substituted by IR sites to help improve Int binding, complex stability and symmetry. (a) The original HJ1TA sequence (B) HJ1TAsy2 contains two IR_L and two IR_R sequences that alternate with each other. (C) HJ1TAsy3 contains four IR_R sites. (D) HJ1TAsy3-1.2 also contains four IR_R core sequences, but with different external sequences in two stems (bold) to better resemble HJ1TA.

Introducing palindromicity into the HJ design was another way to try to increase complex stability and symmetry for crystallization. For this, I designed various HJs with multiple IR_L and/or IR_R sites (Figure 2-14). The IR sequences were chosen as repeating units, because it has been shown that Int can bind strongly to these sites, while target site binding may be weaker.

2.1.2.3 Analysis of Int^{82N}R225K binding to HJ DNA

Next, I tested annealing of HJ1-DNA *in vitro* to assess the quality and quantity of HJ1-DNA formation in various conditions. Overall, buffers with lower NaCl concentrations or no $MgCl_2$ led to heterogenous product formation and little amount of the expected HJ1 (Figure 2-15, lanes 4 and 8). In turn, HJ1 was obtained in high amounts (> 95% product) after annealing in the buffer: 1x Tris-EDTA (TE) buffer, 100 mM NaCl, 10 mM $MgCl_2$ (Figure 2-15, lane 12). This is demonstrated by the formation of a single, slow-migrating main DNA product band on a native PAGE gel when mixing all four DNA strands, whereas controls that contained only two or three strands produced several faster migrating bands. Thus, I used HJ1 directly for binding assays, complex formation, and crystallization experiments.

Using the HJ1 substrate, I next characterized Int^{82N}R225K-HJ1 complex formation by electrophoretic mobility shift assay (EMSA). This indicated the formation of a tetrameric complex when mixing DNA and protein in 1:4 molar ratio (Figure 2-15B, lane 3). However, the complex was not very stable, as it did not elute together on analytical size-exclusion chromatography (SEC, not shown). Nevertheless, the complex looked suitable for performing initial crystallization experiments.

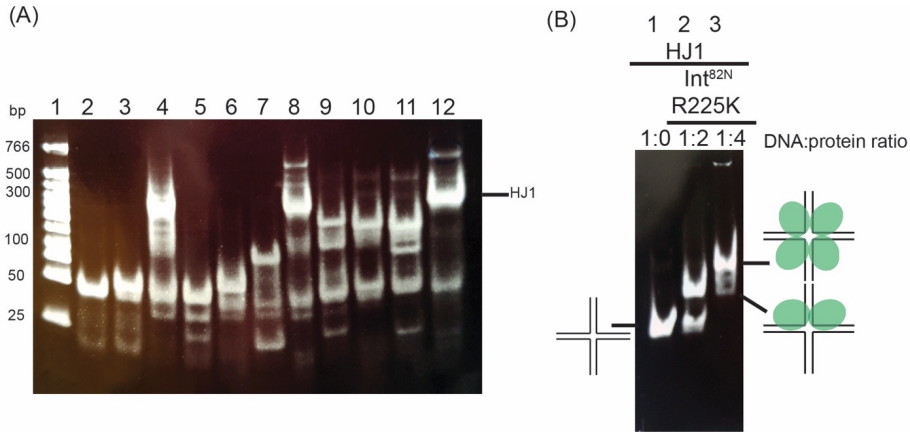

Figure 2-15: Characterization of HJ1 annealing and Int binding. (A) Native gel after HJ1 annealing in different buffers. 1 = marker (Low MW DNA Ladder – NEB). Controls: 2 and 3 =dsDNA substrates with similar length to the HJ DNA strands; 4 = all strands of HJ1 mixed in TE buffer. Different oligonucleotide combinations using the different component strands (a-d) of HJ1: 5 = HJ1a + HJ1b, 6 = HJ1a + HJ1c, 7 = HJ1b + HJ1d, 9 = HJ1a + HJ1b +HJ1d, 10 = HJ1a + HJ1c + HJ1d, 11 = HJ1b + HJ1c + HJ1d. 8 = HJ1

in TE buffer with additional 100 mM NaCl. 12 = HJ1 in TE buffer with additional 100 mM NaCl and 10 mM $MgCl_2$. The sequences of the oligos used can be found in Table 6-2, appendix. (B) EMSA of Int^{82N}R225K-HJ1 complex. Native PAGE of HJ1 alone (lane 1) and mixed with Int in 1:2 (lane 2) and 1:4 (lane 3) molar ratios (buffer :250 mM NaCl, 10% glycerol, 50 mM HEPES pH7.5, 10 mM $MgCl_2$, 5 mM DTT).

2.1.2.4 Crystallization of the Int^{82N}R225K-HJ DNA complex

After assessing HJ DNA annealing and characterizing the binding affinity of Int^{82N}R225K to HJ DNA, I continued with high-throughput crystallization experiments to obtain initial crystal hits. Each different HJ DNA substrate, was assessed for proper annealing on a native gel before complex formation. The various Int-HJ DNA complexes were initially characterized on a native gel before crystallization trials took place. Complexes showing high levels of tetramer formation on native gels were prepared as described in section 4.4.3.1. High-throughput crystallization trays were set up using various commercial screens suited for protein-DNA complexes, as described in section 4.4.3.2. Initial crystallization attempts using the HJ1TA design and its variants are listed in Table 2-1. I tested different HJ substrates that differed in their stem-lengths (for example, HJ1T, HJ1T41, HJT35), their endings (5'-end T, TA, or GA-mismatches), or their crossover regions (CR of 6 and 5 bp). I also tested a HJ DNA substrate that resembles the reaction intermediate when cleavage occurs first at the IR_L of the CI and T_R of the target sequence (HJ_L, Figure 2-13). Moreover, I tested different protein constructs in complex with HJ1 DNA. In addition to the Int^{82N}R225K variant, I used two catalytic mutant constructs that, besides the N-terminal deletion, also have a C-terminal truncation (aa391-aa397). Those were Int^{82N}390C-R225K and Int^{82N}390-2YF. The Int^{82N}390C truncation was previously shown to promote tetramer formation upon DNA binding, suggesting that it could be a better construct for tetrameric complex formation.

From all these initial crystallization experiments, the Int^{82N}R225K-HJ1TA complex showed the most promising initial hits that appeared in condition A1 from the PEGs screen (QIAGEN) at 7 °C. These thin, roundish crystals appeared after 3-4 days and showed diffraction up to 12 Å. The identified condition contained 0.1 M Sodium acetate pH 4.6 and 40% (v/v) PEG 200 as a precipitant. The initial condition was chosen for further manual optimization in 24-well plates, as described in section 4.4.3.2. The best crystals grew using drops of 1 µl + 2 µl size (precipitant:complex) in manual plates, with 0.1M Sodium acetate pH 4.6 and a range of 41 – 48% PEG 200 at 7 °C, by using the hanging-drop method. The crystals were bigger but still had a roundish- and flat-shape like the initial hits. They were prepared for data collection, as described in section 4.4.3.2.

Table 2-1: List of various Int-HJ1TA complexes used for initial crystallization trials and the crystallization conditions tested. The full list can be seen in the appendix, Table 6-4.

Complex	Screens	T	Complex Buffer
Int^{82N}R225K-HJ1TA (no glycerol) (hit)	JCSG+, PEGs, The LMB screen, MIDAS, MORPHEUS, MORPHEUS II, MORPHEUS III, XP screen, The BCS Screen, Additive Screen (PEGs-A1), Silver-Bullets, User Screen (PEGs-A1)	7 °C	200 mM NaCl, 20 mM HEPES pH 7.5, 5 mM $MgCl_2$, 0.5 mM TCEP
Int^{82N}R225K-HJ1TA (no glycerol)	JCSG+, PEGs, The LMB screen, MIDAS, MORPHEUS, MORPHEUS II, MORPHEUS III,	20 °C	200 mM NaCl, 20 mM HEPES pH 7.5, 5 mM $MgCl_2$, 0.5 mM TCEP
Int^{82N}R225K-HJ1TA (no glycerol)	JCSG+, Index, PEGs, PEGs II, The LMB screen, MIDAS, XP screen, The BCS Screen	7 °C	200 mM NaCl, 25 mM Sodium acetate pH 5.5, 5 mM $MgCl_2$, 0.5 mM TCEP
Int^{82N}R225K-HJ1TA (no glycerol)	PEGs, PEGs II, The LMB screen, MIDAS, XP screen, The BCS Screen	20 °C	200 mM NaCl, 25 mM Sodium acetate pH 5.5, 5 mM $MgCl_2$, 0.5 mM TCEP
Int^{82N}R225K-HJ1TA-CR5-43	PEGs	7 °C	200 mM NaCl, 20 mM HEPES pH 7.5, 5 mM $MgCl_2$, 0.5 mM TCEP
Int^{82N}R225K-HJ1TA-CR5-44	PEGs	7 °C	200 mM NaCl, 20 mM HEPES pH 7.5, 5 mM $MgCl_2$, 0.5 mM TCEP
Int^{82N}R225K-HJ1TA-Sy3	JCSG+, PEGs, PEGs II, MORPHEUS, The LMB Screen, Index, The BCS Screen, User Screen (PEGs A1 based), Additive Screen (PEGs A1 based)	7 °C	200 mM NaCl, 20 mM HEPES pH 7.5, 5 mM $MgCl_2$, 0.5 mM TCEP
Int^{82N}R225K-HJ1TA-Sy3	JCSG+, PEGs, The BCS Screen	7 °C	200 mM NaCl, 25 mM Sodium acetate pH 5.5, 5 mM $MgCl_2$, 0.5 mM TCEP
Int^{82N}R225K-HJ1TA-Sy3-1.2	User Screen (PEGs A1 based)	7 °C	200 mM NaCl, 20 mM HEPES pH 7.5, 5 mM $MgCl_2$, 0.5 mM TCEP

2.1.2.5 Data collection and structure solution

I collected X-ray diffraction datasets at beamlines of the European Synchrotron Radiation Facility (ESRF) in Grenoble, France and German Electron Synchrotron (DESY) in Hamburg, Germany. The best datasets were collected at beamline P13 at DESY (Figure 2-16), as described in section 4.4.3.3. Most of the tested crystals showed poor diffraction, with only rare examples diffracting beyond 5 Å. The best diffracting crystals grew in condition 0.1M Sodium acetate pH 4.6 and 47% (v/v) PEG 200. The datasets were processed with the XDS software (Kabsch *et al.*, 2010) (see section 4.4.3.4). The best dataset had significant diffraction intensity up to 3.3 Å. Nevertheless, the data statistics (such as signal to noise ratio, R-factors and CC1/2) worsened rapidly at high resolution (Table 2-2). A closer look at the diffraction images showed tailed diffraction spots pointing at crystal mosaicity and crystal disorder (Rupp, 2010).

Table 2-2: Statistics table of the processed Int-HJ1TA dataset in XDS from the CORRECT.LP file. The data was cut at 3.3 Å. I/σ(I) values are a measure of signal-to-noise ratio. CC(1/2) is a statistical indicator of data consistency.

SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= -3.0 AS FUNCTION OF RESOLUTION

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR COMPARED expected	I/SIGMA	R-meas	CC(1/2)	Anomal Corr	SigAno	Nano
9.85	4492	1366	1447	94.4%	4.2%	4.0%	4432	28.51	5.0%	99.7*	-17	0.740	833
6.99	8438	2531	2637	96.0%	3.9%	4.2%	8342	24.34	4.7%	99.7*	-21	0.699	1619
5.71	11891	3296	3341	98.7%	7.5%	7.4%	11872	13.48	8.8%	99.5*	-12	0.735	2526
4.95	13366	3977	4049	98.2%	12.4%	12.4%	13244	7.72	14.8%	99.0*	-6	0.736	2627
4.43	16134	4480	4548	98.5%	19.8%	20.0%	16086	5.15	23.3%	98.1*	-7	0.704	3441
4.04	16508	4926	5033	97.9%	44.8%	44.7%	16399	2.29	53.4%	92.9*	-3	0.665	3170
3.74	19033	5389	5491	98.1%	85.2%	84.6%	18994	1.29	100.5%	81.6*	-1	0.632	3910
3.50	20936	5771	5876	98.2%	125.5%	127.6%	20884	0.91	147.4%	63.6*	-2	0.613	4502
3.30	19765	5986	6166	97.1%	189.9%	194.4%	19617	0.58	227.4%	30.2*	0	0.621	3683
total	130563	37722	38588	97.8%	10.1%	10.2%	129870	5.98	12.0%	99.8*	-5	0.668	26311

Another problem I observed was the anisotropic diffraction pattern. A dataset shows anisotropy when it diffracts to a higher resolution in a specific direction than in other directions during the diffraction experiment. Moderate cases do not cause significant difficulties but, when the cases are severe, refinement can be challenging and the electron density map may miss high-resolution features in some parts of the structure (Zwart, Grosse-Kunstleve, 2005). We analyzed the Int-HJ1TA datasets using the STARANISO server from Global Phasing Ltd. (http://staraniso.globalphasing.org/). The results showed that the dataset was strongly anisotropic (Figure 2-17). Thus, the data was scaled anisotropically and cut elliptically at a resolution of 3.3 Å. This helped to improve the data statistics overall, especially in the high-resolution shells (see Table 2-3). Other challenges in data processing that may have compromised data quality included the low symmetry (P1) of the unit cell and the modest redundancy of the collected data.

The anisotropically scaled and merged dataset was then used to solve the phase problem by molecular replacement in Phaser (McCoy *et al.*, 2007). As model I used one Int subunit and the bound DNA fragment from the previously solved Int-CI5 structure (Rubio-Cosials *et al.*, 2018). The best solutions were then further used for model building in COOT (Emsley *et al.*, 2010) and refinement in PHENIX (Liebschner *et al.*, 2019) (see section 4.4.3.6). The molecular replacement solution constructs the crystal lattice without significant clashes and reconstructs the observed electron density well in most parts. The current best refined model shows R-factors that convincingly confirm its overall validity. However, both R-factors remained relatively high (R_{work} = 0.3250 and R_{free} = 0.3710) for diffraction data at 3.3 Å (see Table 2-4). This may be due to the limits of data quality and/or inherent flexibilities in the structure, as discussed below.

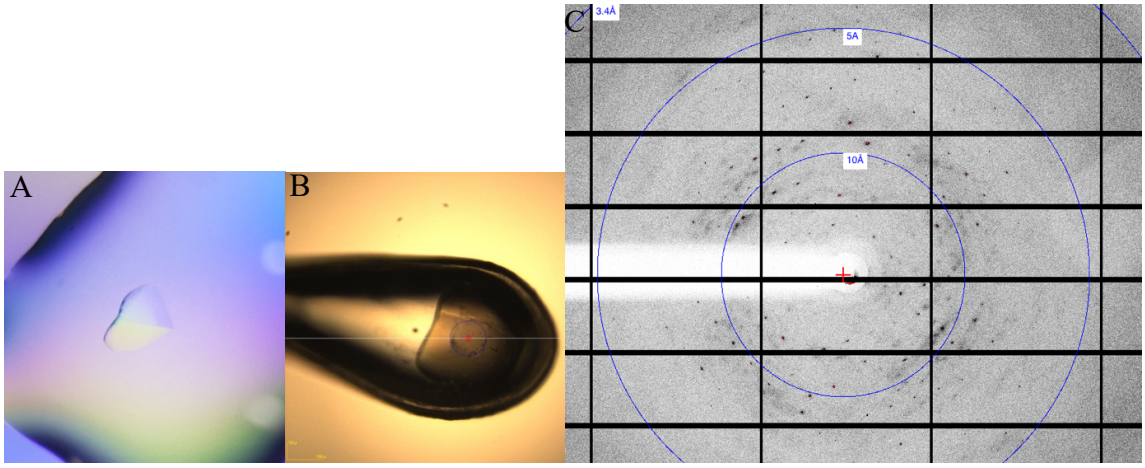

Figure 2-16: Crystallization and data collection of Int in complex with HJ1TA substrate. (A) Image of one Int-HJ1TA crystal in the manual sitting-drop and (B) in the loop prepared for data collection. (C) Example diffraction pattern obtained from the crystal shown in (A).

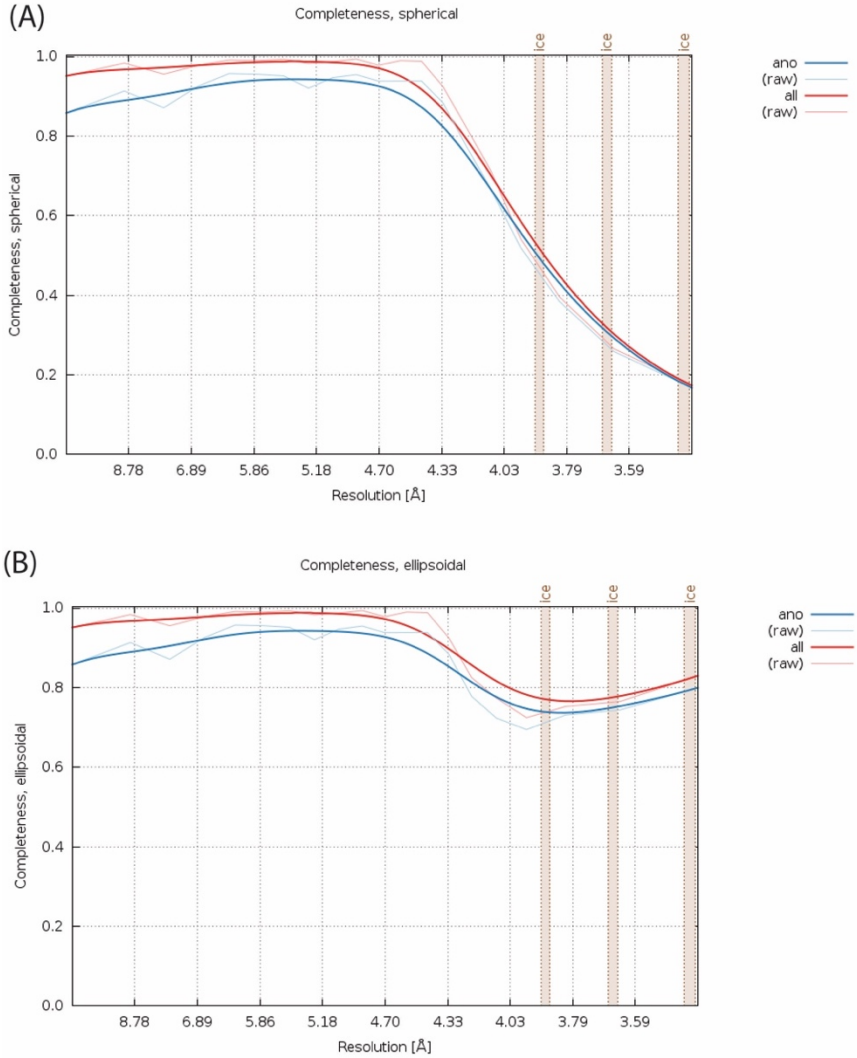

Figure 2-17: Anisotropy correction of the Int-HJ1TA dataset. Analysis using the STARANISO server from Global Phasing Ltd. (http://staraniso.globalphasing.org/). Comparison of dataset completeness (A - spherical and B - ellipsoidal) plotted against resolution in Å after anisotropic scaling. After this process, the server will provide a .mtz file containing the corrected structure factors.

Table 2-3: Statistics table of the merged Int-HJ1TA dataset after anisotropic scaling in STARANISO.

	Resolution		#uniq	#Rfac	Rmerge	Rmeas	Rpim	#Isig	I/sigI	Compl. Spher. all	Compl. Spher. ano	Multiplicity all	Multiplicity ano	CC(1/2)	#CCAno	CC(ano)	SigAno	Compl. Ellip. all	Compl. Ellip. ano
	94.178 -	10.172	1244	4003	0.042	0.050	0.027	1244	27.754	0.9511	0.8578	3.26	1.72	0.9960	739	0.1031	0.783	0.9511	0.8578
	10.172 -	8.133	1246	4236	0.038	0.045	0.024	1246	26.155	0.9834	0.9132	3.41	1.77	0.9972	827	-0.2036	0.697	0.9834	0.9132
	8.133 -	7.098	1245	3954	0.040	0.048	0.026	1245	21.115	0.9555	0.8711	3.24	1.72	0.9965	760	-0.1740	0.664	0.9555	0.8711
	7.098 -	6.452	1247	4400	0.059	0.069	0.036	1247	16.332	0.9780	0.9255	3.55	1.83	0.9935	927	-0.1738	0.710	0.9780	0.9255
	6.452 -	5.994	1245	4556	0.078	0.091	0.047	1245	12.513	0.9912	0.9570	3.66	1.86	0.9927	980	-0.2461	0.715	0.9912	0.9570
	5.994 -	5.649	1244	4380	0.107	0.127	0.067	1244	9.284	0.9897	0.9554	3.52	1.79	0.9862	900	-0.0787	0.736	0.9897	0.9554
	5.649 -	5.374	1247	3947	0.115	0.139	0.077	1247	7.855	0.9928	0.9514	3.19	1.64	0.9870	710	-0.1727	0.755	0.9928	0.9514
	5.374 -	5.145	1245	4176	0.123	0.145	0.077	1245	7.435	0.9826	0.9203	3.40	1.77	0.9889	863	-0.1445	0.729	0.9826	0.9203
	5.145 -	4.942	1246	4366	0.148	0.175	0.092	1246	6.344	0.9873	0.9469	3.53	1.81	0.9871	922	-0.0506	0.764	0.9873	0.9469
	4.942 -	4.777	1244	4439	0.163	0.193	0.101	1244	5.857	0.9936	0.9545	3.57	1.82	0.9852	931	-0.0902	0.739	0.9936	0.9545
	4.777 -	4.626	1245	4504	0.192	0.226	0.117	1245	5.440	0.9780	0.9379	3.63	1.86	0.9742	982	-0.0904	0.761	0.9780	0.9379
	4.626 -	4.495	1245	4460	0.227	0.266	0.138	1245	4.270	0.9897	0.9380	3.60	1.85	0.9771	961	-0.0749	0.741	0.9897	0.9380
	4.495 -	4.378	1246	4499	0.319	0.375	0.195	1246	3.258	0.9881	0.9381	3.61	1.85	0.9481	955	-0.0245	0.738	0.9881	0.9381
	4.378 -	4.264	1247	4226	0.327	0.390	0.209	1247	3.061	0.9244	0.8836	3.40	1.74	0.9510	824	-0.0127	0.782	0.9244	0.8836
	4.264 -	4.151	1245	4073	0.502	0.600	0.323	1245	2.045	0.8250	0.7780	3.31	1.71	0.8931	775	0.0194	0.778	0.8250	0.7780
	4.151 -	4.034	1245	4072	0.501	0.599	0.324	1245	2.031	0.7147	0.6682	3.30	1.71	0.8787	786	-0.1073	0.746	0.7733	0.7229
	4.034 -	3.899	1246	4352	0.490	0.580	0.306	1246	2.244	0.5429	0.5207	3.50	1.79	0.8623	894	-0.0257	0.769	0.7243	0.6946
	3.899 -	3.738	1245	4477	0.621	0.730	0.380	1245	1.891	0.3957	0.3840	3.60	1.82	0.8060	956	-0.0818	0.753	0.7525	0.7310
	3.738 -	3.539	1246	4614	0.644	0.753	0.388	1246	1.856	0.2657	0.2585	3.71	1.88	0.7736	1037	-0.0209	0.765	0.7649	0.7438
	3.539 -	3.302	1245	4109	0.992	1.188	0.644	1245	1.130	0.1733	0.1675	3.32	1.69	0.5266	801	0.0031	0.771	0.8297	0.7995
Total:	94.178 -	3.302	24908	85843	0.075	0.090	0.048	24908	8.393	0.6484	0.6158	3.47	1.78	0.9972	17530	0.0484	0.745	0.9075	0.8616

Table 2-4: Crystallographic data collection (taken from Table 2-2) and refinement statistics for Int-HJ1TA dataset. R.m.s = root mean square.

	Int-HJ1TA
Data collection	
Space group	P 1
Cell dimensions	
a, b, c (Å)	80.185, 96.26, 102.489
a, b, g (°)	77.836, 66.991, 65.464
Wavelength (Å)	0.9763
Resolution (Å)	94.18 – 3.302 (3.539 – 3.302)
R-merge	0.075 (0.992)
R-meas	0.090 (1.188)
I / sigma (I)	8.393 (1.13)
CC1/2	0.9972 (0.5266)
Completeness (ellipsoidal, %)	90.75 (82.97)
Redundancy	3.47 (3.32)
Refinement	
Resolution	94.18 – 3.302
No. reflections	24817 (562)
R_{work} / R_{free}	0.3250 / 0.3710
No. atoms	12252
Macromolecule	12252
Water	-
Average B-factor	125.95
Protein	125.95
Water	-
R.m.s deviations	
Bond lengths (Å)	0.013
Bond angles (°)	1.35
Ramachandran favored (%)	88.51
Ramachandran allowed (%)	9.68
Ramachandran outliers (%)	1.60
Rotamer outliers (%)	6.50

In accordance with the predicted unit cell content, I could place four Int monomers with four CI5 DNA half sites in the asymmetric unit, forming a tetrameric complex. The four DNA molecules come together at the complex's center to form the HJ DNA junction. The CB-domains and the HJ-stems seemed to fit quite well into the electron density (Figure 2-18). However, the density was weaker at the CAT domains (Figure 2-19).

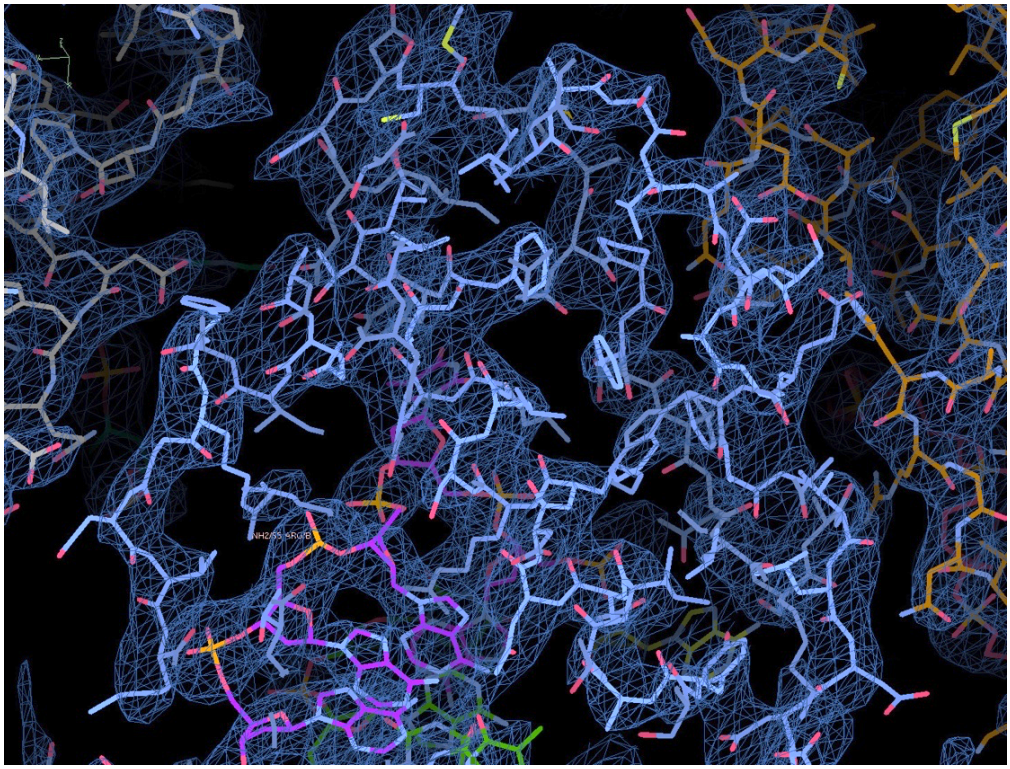

Figure 2-18: 2Fo-Fc electron density map at contour level of 1 sigma showing one CB domain region of the Int-HJ1TA structure.

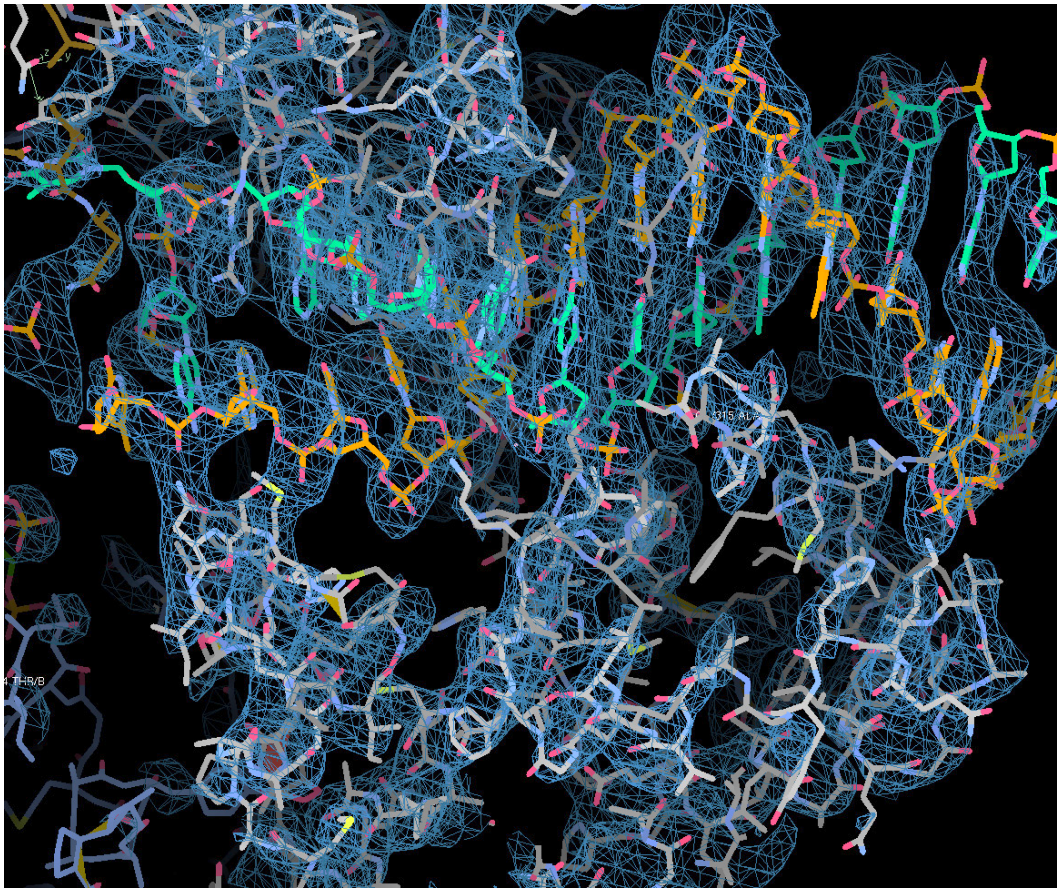

Figure 2-19: 2Fo-Fc electron density map at contour level of 1 sigma showing one CAT domain region and the bound HJ DNA stem in the Int-HJ1TA structure.

2.1.2.6 Overview of the architecture of the Int-HJ complex

Although the structure could not be refined to R-factors expected at 3.3 Å resolution so far, probably due to limitations in the diffraction data, the current model still reveals important new insights. The structure shows four Int^{82N} monomers bound to one HJ1TA DNA molecule forming a tetrameric assembly (Figure 2-20). The protein monomers are related with imperfect four-fold non-crystallographic symmetry. The protein used for crystallization consists of 316 amino acid residues, which includes an N-terminal serine resulting from SenP2 cleavage of the 6xHis-SUMO tag. In the current model, various side chains could not be adequately placed due to inconclusive density. Most of these are found in the CAT domain and belong mostly to flexible loops or linker regions.

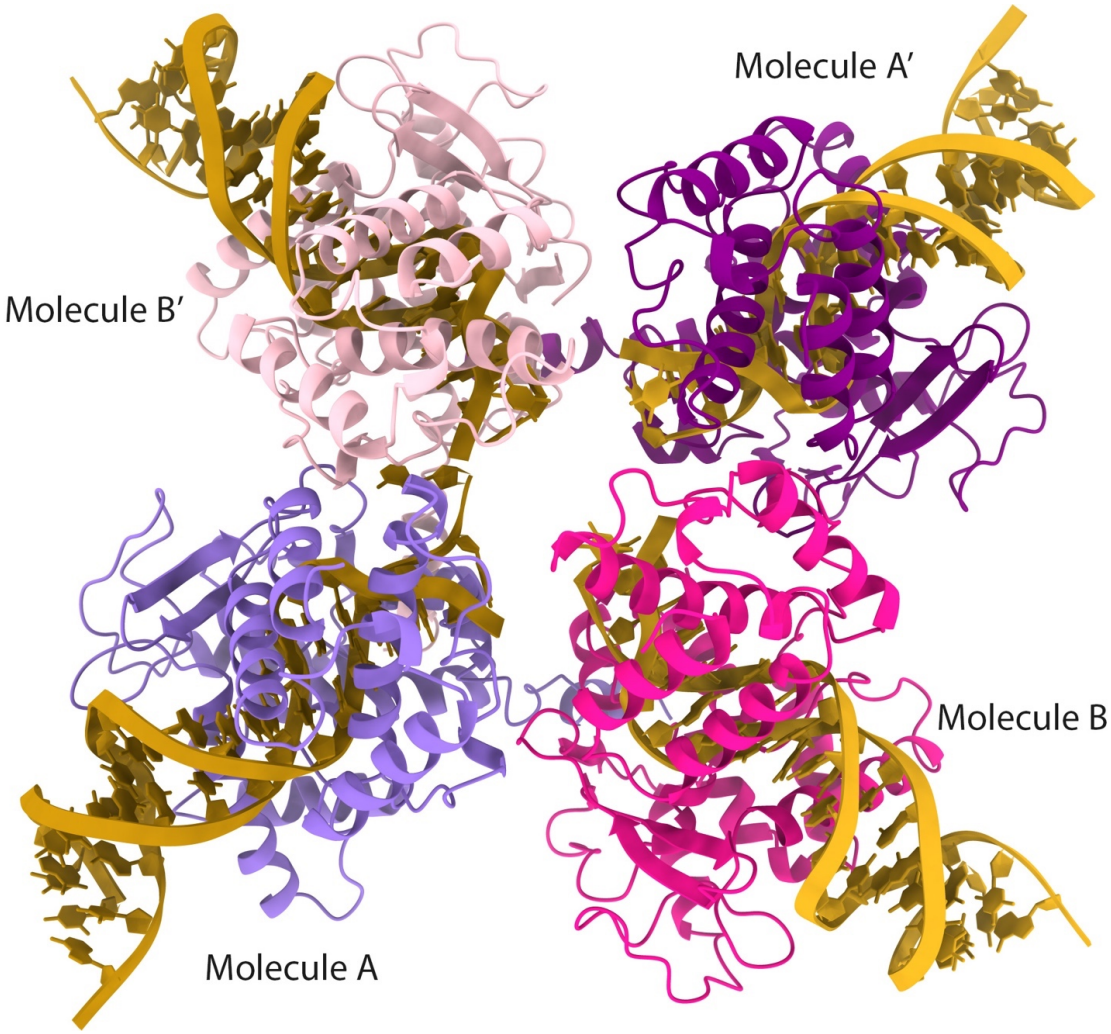

Figure 2-20: Overview of the Int-HJ1TA complex structure, shown in cartoon representation. Four Int molecules (shades of pink and purple) are bound to one HJ1TA DNA (golden). Non-crystallographic symmetry related monomers are highlighted with similar colors: molecule A (light purple) and molecule A' (purple), molecule B (pink), and molecule B' (light pink).

The HJ1TA DNA contains four 21 bp / 22 bp long stems , which meet in the middle of the complex. The stems comprise the inverted repeats (IR_L and IR_R) and the target sites (T_L and T_R) with alternating T- and A-overhangs at their 5'-ends (Figure 2-11). The core of the IRs and T-sites are well resolved and well visible in the density map. At the center of the complex, I observe weak density connecting the stems and the exact path of the DNA could not be unambiguously built (Figure 2-21). Similarly, the ends of the stem sequences (i.e. the last 3 bp on each stem) could not be resolved in the density maps. Moreover, although density for the core DNA parts was well visible, the resolution was not sufficient enough to assign the specific bases. Thus, DNA building largely relied on the previously solved Int-CI5 structure, leaving the assignment of the four HJ stems ambiguous.

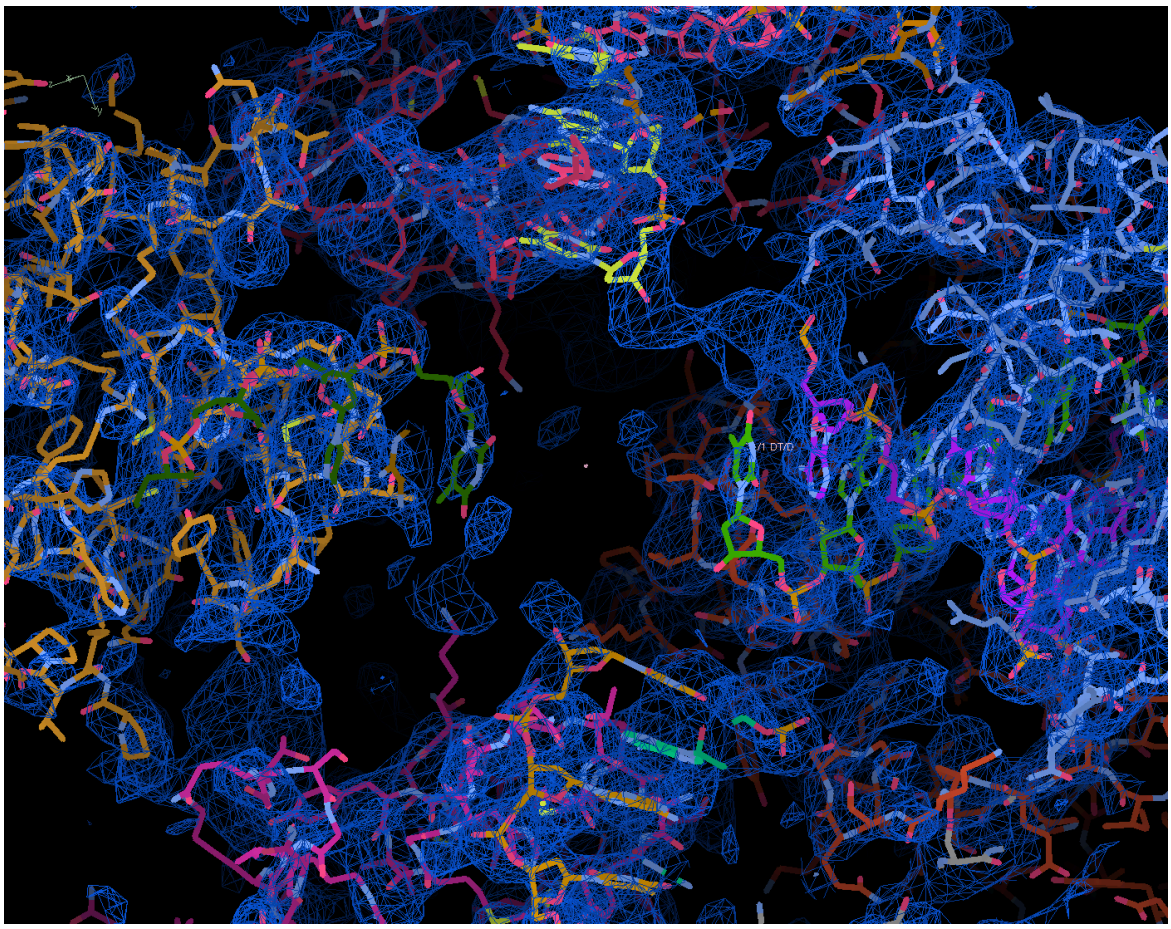

Figure 2-21: Close-up of the Int-HJ1TA structure showing the HJ center. 2Fo-Fc electron density map at contour level of 1 sigma showing very weak signal at the center of the HJ molecule where the stems should meet in the Int-HJ1TA structure.

Moreover, it can be observed that within the tetrameric complex, the Int^{82N} monomers are arranged into a cyclic assembly with their C-terminal helices swapped to the neighboring monomers in an clockwise manner (Figure 2-22, bottom view, from the CAT domains). This is similar to the arrangement seen before in synaptic structures of λ Int, Cre, and XerH

(Gopaul, Guo and Van Duyne, 1998; Biswas *et al.*, 2005; Bebel *et al.*, 2016). This is in contrast to the Int-CI5 structure, which showed a two-fold symmetric dimer on linear dsDNA, with the C-terminal helices exchanged reciprocally within the dimer. Not all residues of the C-terminal helices could be built well in the current model, but their circular swapping is well supported with electron density (Figure 2-23).

The relative positioning of the Int molecules also changed in the Int-HJ1TA tetramer compared to the Int-CI5 structure. In fact, the protein subunits underwent a large 45 degree rotation relative to each other, thereby creating a roughly four-fold symmetric overall arrangement. In the new conformation, each Int molecule interacts with one stem of the HJ. Protein-protein contacts between neighboring monomers through the CB and CAT domains further stabilize the assembly (Figure 2-20).

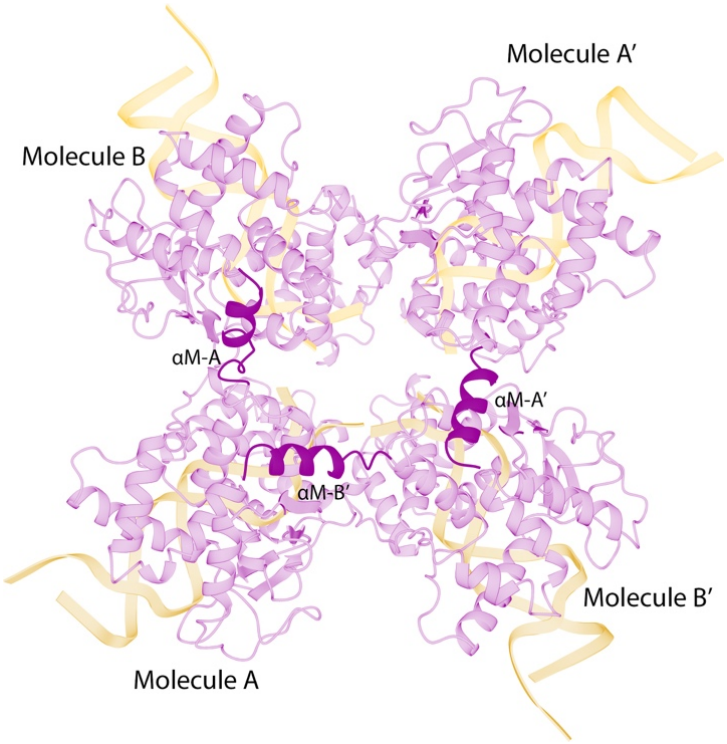

Figure 2-22: C-terminal αM helices interaction in the Int-HJ1TA crystal structure. Cartoon representation of the Int-HJ1TA complex, Int monomers in purple, HJ DNA in golden. αM-C tails are highlighted in dark purple (bottom view). The αM-helix of monomer B is missing.

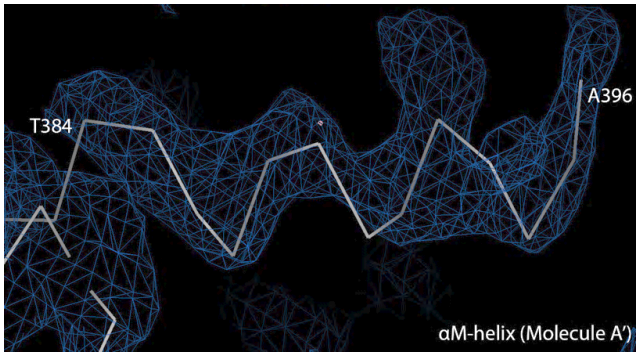

Figure 2-23: Close-up of the Int-HJ1TA structure showing the C-tail of monomer A'. 2Fo-Fc electron density map at contour level of 1.2 sigma showing enough signal to build part of the αM-helix in the Int-HJ1TA structure. One example is given from three αM-helices that could be placed and had enough electron density from a total of four.

2.1.2.7 Structure of the Int molecules and their interaction with DNA

Within the limits of data quality and map interpretability, overall fold of the Int monomers in the Int HJ1TA structure resembles the Int-CI5 structure (Figure 2-24A). Each monomer contains two domains that are connected by a flexible linker (Figure 2-25A). The CB domain (residues 82-166) at the N-terminus is equivalent to the CB domains of Cre, λ Int, and XerH recombinases (Gopaul, Guo, and Van Duyne, 1998; Biswas *et al.*, 2005; Bebel *et al.*, 2016). In the Tn*1549* Int, the CB domain consists of four α-helices like in the λ Int (Figure 2-26A). In XerH, CB has two additional helices between Int's αA and αB helices (Figure 2-26C) and Cre has an extra helix at the end of the CB domain, before the inter-domain linker (Figure 2-26B). This is in agreement with earlier observations that the CB domain is less well conserved between tyrosine recombinases than the CAT domain, although the overall fold is similar.

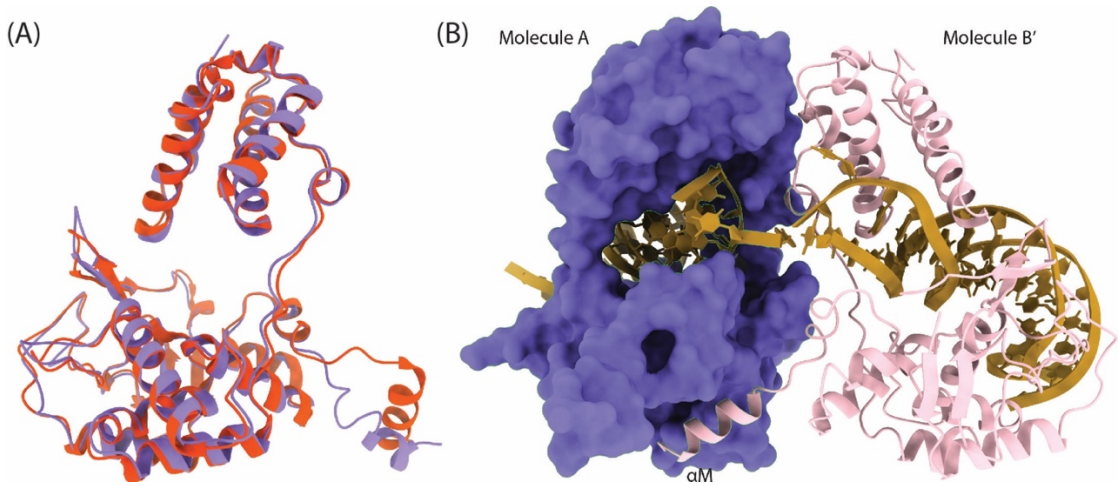

Figure 2-24: Superimposition of one Int-HJ monomer with one Int-CI5 monomer and the C-terminal αM helix interaction in the Int-HJ1TA complex. (A) Superimposition of one Int-HJ1TA monomer (light purple) with one Int-CI5 monomer (red). (B) The C-terminal αM helix interaction in the Int-HJ1TA complex. The αM-helix of monomer B' (light pink) interacts with monomer A (surface representation, light purple) by binding in a small cleft on its surface. Only half of the Int-HJ complex is shown. DNA (golden).

The CAT domain (residues 191-397) is composed of nine helices and four β-strands (Figure 2-25A). These includes the unique β-hairpin, which is only conserved in the family of Tn*916*-like transposases (Rubio-Cosials *et al.*, 2018). As in the Int-CI5 structure, the C-terminal helix (αM) of each monomer interacts with a neighboring monomer, binding into the same cleft in the CAT domain (Figure 2-24B). However, the orientation of αM has changed relative to the Int core, following the major change observed in relative orientation of the monomers (Figure 2-24A).

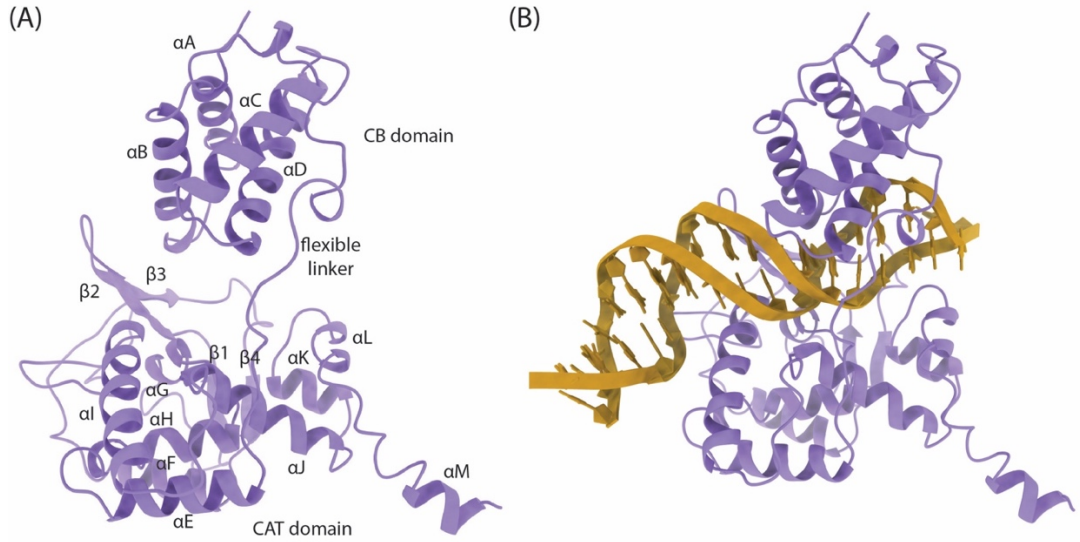

Figure 2-25: Overview of one Int monomer from the Int-HJ1TA structure, alone and with one HJ stem. (A) Cartoon representation of the structure of one Int monomer (light purple) alone and (B) with one stem of HJ DNA (golden, right) from the tetrameric Int-HJ structure. The α-helices order is indicated by letters, β-sheet strands are numbered.

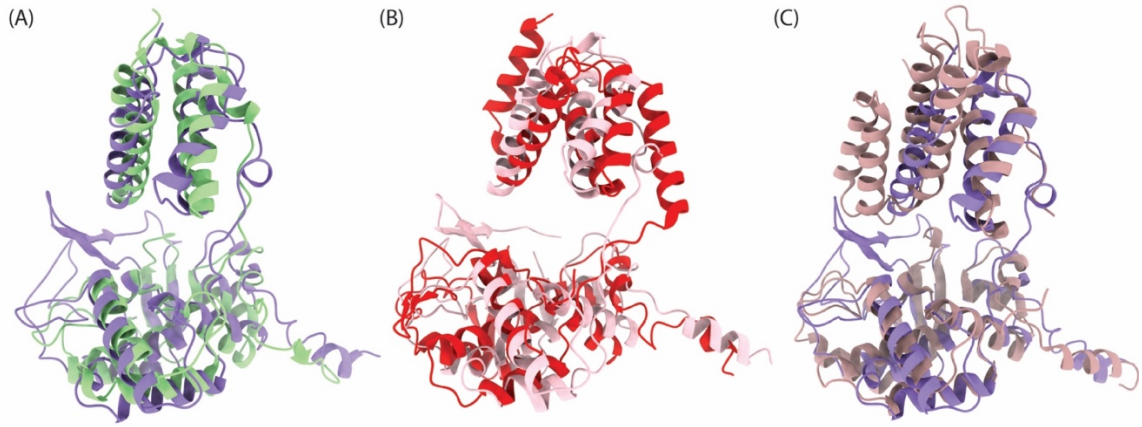

Figure 2-26: Superimpositions of one Int-HJ monomer with structurally related tyrosine recombinase structures. Cartoon representation of various superimpositions using an Int^{82N} monomer (light purple or light pink) from the Int-HJ1TA structure with one monomer of (A) the λ Int-HJ complex (light green), pdb 1z1g (Biswas et al., 2005); (B) the Cre-HJ complex (red), pdb 2crx (Gopaul, Guo and Van Duyne, 1998); (C) the post-cleavage XerH-dif$_H$ synaptic complex (pink), pdb 5jjv (Bebel et al., 2016).

The overall C-shaped clamp structure of Int has also been preserved and each monomer interacts with DNA in a similar way as in the Int-CI5 structure (Figure 2-25B). The most significant interactions are formed with the core DNA sequences in each HJ stem. The CB domain forms numerous contacts in the DNA major groove and CAT forms mostly non-specific contacts at the minor groove. The β-hairpin and the interdomain linker also contribute to DNA binding, as seen in the Int-CI5 structure.

2.1.2.8 Description of the HJ DNA arrangement

Due to the low-resolution map, it was not possible to assign the correct DNA sequences in the density and the central region density is not clearly defined (Figure 2-21). Nevertheless, the core DNA sites could be modeled in all four stems of the HJ (Figure 2-19), which allows to compare the HJ conformation compared to other tyrosine recombinase structures.

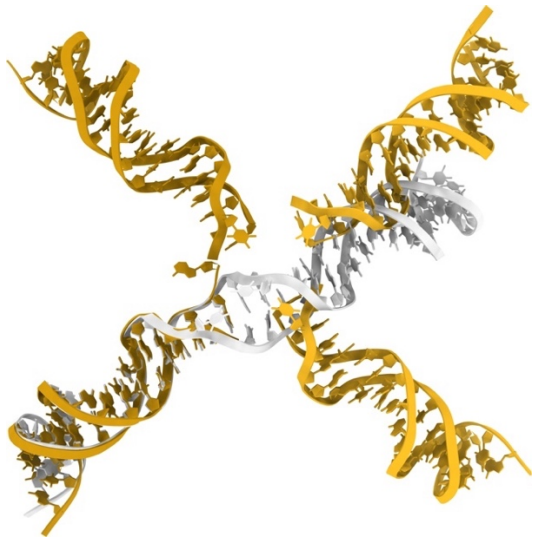

Figure 2-27: Superimposition and comparison of HJ1TA-DNA with CI5 DNA. Cartoon representation of HJ1TA DNA (golden) from the Int-HJ1TA structure, superimposed with the CI5 DNA (silver) from the Int-CI5 structure (Rubio-Cosials et al., 2018). The HJ1TA-DNA has an overall four-fold rotational symmetry and a nearly square planar conformation, whereas the CI5 DNA is linear and almost entirely straight.

The HJ1TA DNA has an overall four-fold rotational symmetry with a nearly square planar conformation (Figure 2-27). Comparison of HJ1TA with CI5 from the Int-CI5 complex (Rubio-Cosials *et al.*, 2018), makes it evident that a strong DNA bending will need to occur upon synaptic complex formation to change the nearly straight conformation of the CI5 DNA towards an HJ-like arrangement of the neighboring stems (Figure 2-27).

The HJ conformation observed in the Int- HJ1TA structure is more similar to the ones from the λ Int-HJ and the λ Int-COC' post-strand exchange complexes (Biswas et al., 2005) than

to the state in the Cre-HJ structure (Figure 2-28) (Gopaul, Guo and Van Duyne, 1998). However, Int seems to bend two HJ stem pairs stronger than in both λ Int structures.

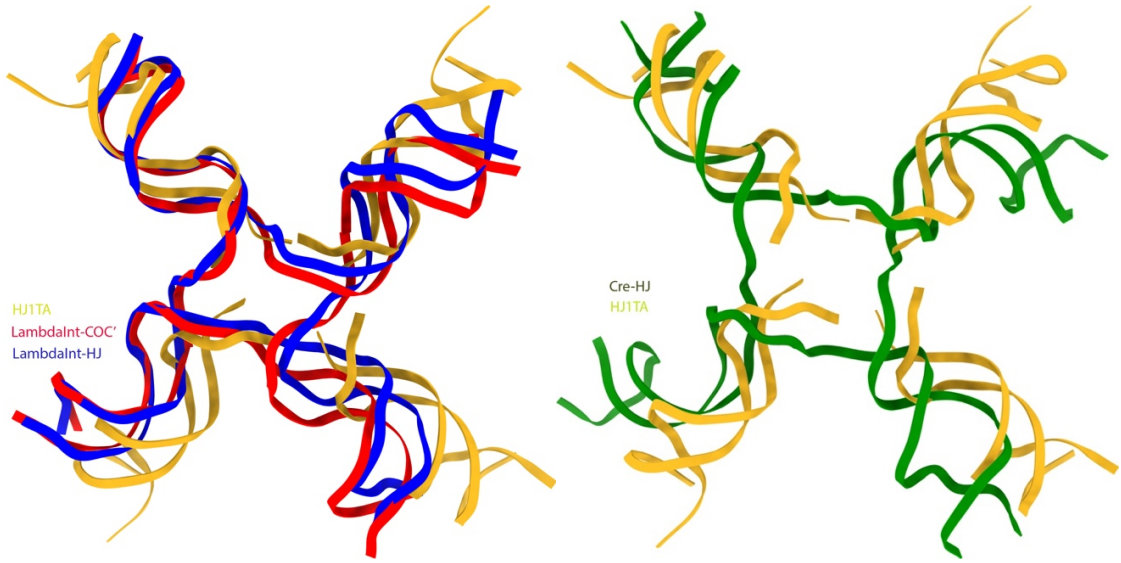

Figure 2-28: Cartoon representation of HJ1TA DNA (golden) from the Int-HJ1TA structure superimposed to HJ-DNA from the λ Int-HJ (blue) and the λ Int-COC' post-strand exchange (red) complexes (Biswas et al., 2005) (left) and with HJ-DNA (green) from the Cre-HJ structure (Gopaul, Guo and Van Duyne, 1998) (right).

2.1.2.9 Discussion

2.1.2.9.1 Problems in data processing and refinement

All diffraction images of the Int-HJ1TA crystals showed some problematic characteristics. First, diffraction was observed only to moderate resolution, roughly 3.3 - 4 Å. Second, the reflection spots had large size and elongated shape. These features indicate high crystal mosaicity and poor order within the crystals. Consistently, XDS and Xtriage outputs highlighted strong drop in completeness and CC1/2 values with increasing R factors beyond 5 Å. A third problem that could already be detected in the images was diffraction anisotropy. Therefore, we corrected the Int-HJ1TA dataset for anisotropy and cut at a resolution of 3.3 Å.

As mentioned above, the datasets show high-resolution reflections, but many are weak. Thus, another approach to improve data quality was to merge several datasets collected on different Int-HJ1TA crystals. With the help of a postdoc in the lab, Vladimir Arinkin, we merged various datasets using BLEND from the CCP4 suite (Foadi *et al.*, 2013). The process is described in section 4.4.3.7. The best merged datasets were then used for molecular replacement and initial refinement, as described above. The resulting model

fitted well in the density and showed similar features as described in section 2.1.2.6. The electron density map better-resolved side chains in the CB domains and the HJ DNA seemed somewhat better resolved. Nevertheless, problematic regions in the CAT domain did not improve. Refinement of these models is in progress but has similar challenges as seen before with diffraction data from a single crystal.

The above limitations in the diffraction data likely contribute to the relatively high R-factors observed during refinement (R_{work} = 0.3250 and R_{free} = 0.3710). Another problem might lie in the intrinsic characteristics of the structure, such as disorder or flexibility of protein segments and connecting linkers, leading to an ensemble of conformations across the crystal (Rupp, 2010).

Due these problems, in some parts of the structure, especially in the CAT domain, the electron density is weak and noisy. This made it difficult to place protein side chains and HJ1TA bases correctly. Some of these problems could be overcome by using a merged dataset for structure solution and refinement. Nevertheless, the core of the model, meaning the main protein chain and the core DNA sites fit well in the density map. Thus, the structure already provides insights into the overall complex architecture, including the Int monomer fold, the HJ1TA DNA conformation, tetramer assembly and the most apparent protein-protein and protein-DNA interactions. Moreover, we can compare the structure already with other structures of the tyrosine recombinase family and detect similarities and differences in the conjugative transposase's integration pathway.

2.1.2.9.2 Attempts to improve the Int-HJ crystals

The crystallization of the Int-HJ complex was not an easy task. The behaviors of the various Int-HJ complexes depended heavily on the exact HJ DNA variant used. Some of the complexes precipitated during complex formation, others during concentration, which sometimes led to reproducibility problems. Moreover, some complexes tended to precipitate after concentration if kept on ice or at RT. Luckily, the Int-HJ1TA complex behaved nicely and could be used for initial crystallization trials. Final crystals grew up to 100-300 µm size and showed up to 3.3 Å resolution. Nevertheless, I encountered several issues during the data processing (discussed in section 2.1.2.9.1). Thus, I decided to continue with crystallization experiments, in parallel with data analysis, to obtain better-diffracting crystals.

I decided to keep the TA-overhang design but changed other parts of the HJ1TA sequence. The first change was to use HJs with a crossover region of 5 bp (HJ1TACR5, Figure 2-11).

Although this would lead to more asymmetric complex, the protein-DNA assembly should have a more compact character. This could increase the intrinsic stability of the complex and may lead to a better crystal packing and higher resolution diffraction. A second approach was to introduce symmetry in the DNA and use HJs with the core IR sequences in all four arms (Figure 2-14). Many of the complexes were rescreened to find new crystallization conditions, but this did not lead to any further hits.

However, I obtained crystals with Int-HJ1TACR5, HJ1TAsy3, and HJ1TAsy3-1.2 complexes in the same condition as for HJ1TA. Diffraction experiments showed an estimated diffraction up to 4 - 6 Å resolution. Datasets were collected and processed, as described in section 2.1.2.5, when the crystals diffracted to around 4-5 Å. Unfortunately, all collected datasets showed similar issues as observed before for the Int-HJ1TA datasets. The R-factors stalled at even higher values this time, and there were fewer features resolved in both CB and CAT domains, probably also due to the lower resolution. Thus, crystallization with alternative HJ DNA sequences did not help overcome the problem of weak and anisotropic diffraction in my experiments so far.

Another reason for the diffraction problems could be weak binding to the poorly conserved T-sites in the HJ stems. Nevertheless, crystals of Int-HJ1TAsy3 that contained four IR_R sites did not show any improvement. Moreover, all four CAT domains in the Int-HJ1TA density map show similar issues, and not only the monomers binding the T-sites. Thus, possible weaker binding of the T-sites cannot account alone for the weak density at high resolution. Together with the previous observations for the Int-HJ1TA complex these results indicate that the problems result from intrinsic features of the protein-DNA complex.

2.1.2.9.3 Comparison of the Int-HJ and the Int-CI structures

When comparing the Int-HJ1TA and the Int-CI5 structures, the most obvious observation is that the first structure comprises four Int monomers binding one HJ1TA molecule. In comparison, the Int-CI5 structure shows an Int dimer bound to CI5 DNA in a pre-catalytic inactive state (Figure 2-29). As expected, the monomer fold did not change, and the C-clamp form of the protein has been maintained when binding to the HJ1TA DNA (Figure 2-25). To accommodate four Int monomers on the HJ1TA DNA, monomer B in the Int-CI5 structure rotated by about 45° to be positioned on the second HJ stem. Thus, while the monomers were interacting in a ‘face to face’ fashion in the dimer structure, four monomers interact more in a “side to side” way in the Int-HJ1TA structure. In the CAT domain, changes in the inter-molecular arrangement drive the C-terminal αM-helices to interact with

the neighboring monomer in a cyclic manner (Figure 2-22). These contacts of the C-tail are supporting tetrameric complex formation.

At this step in the structure refinement, it is challenging to judge if the protein active sites are in a catalytically active or inactive state. Several catalytic residues, including the catalytic tyrosine, are not sufficiently resolved in the current density maps. The position of the R153 side chains also cannot be deduced with certainty now, leaving it unclear if the unique base flipping observed at the cleavage site in the Int-CI5 structure is preserved in the HJ complex.

Nevertheless, the overall shape of the Int-HJ structure is well supported by the density maps, albeit the structure would benefit from further improvement, especially for high-resolution conclusions.

Figure 2-29: Comparison of the Int-HJ1TA and Int-CI5 crystal structures. Superimposition of Int-HJ1TA (protein in purple and DNA golden) and Int-CI5 (red, pdb 6emz) crystal structures.

2.1.2.9.4 Comparison of the Int-HJ structure to known tyrosine recombinase structures

The crystal structure of Int in a tetrameric complex with an HJ intermediate containing CI and integration sites resembles one of the expected steps during conjugative transposon integration. The structure presented in this work reflects similarities compared to other members of the tyrosine recombinase family that have been structurally characterized before. To date, two site-specific tyrosine recombinases have been trapped in a similar HJ intermediate conformation: Cre with six different HJ-bound structures (Gopaul, Guo and Van Duyne, 1998; Martin *et al.*, 2002; Martin, Chu, and Baldwin, 2003; Ghosh *et al.*, 2005) and λ Int bound to a native immobile HJ (Biswas *et al.*, 2005). Apart from that, the post-cleavage synaptic complex of XerH-*dif*$_H$ also showed an overall similar tetrameric assembly (Bebel *et al.*, 2016).

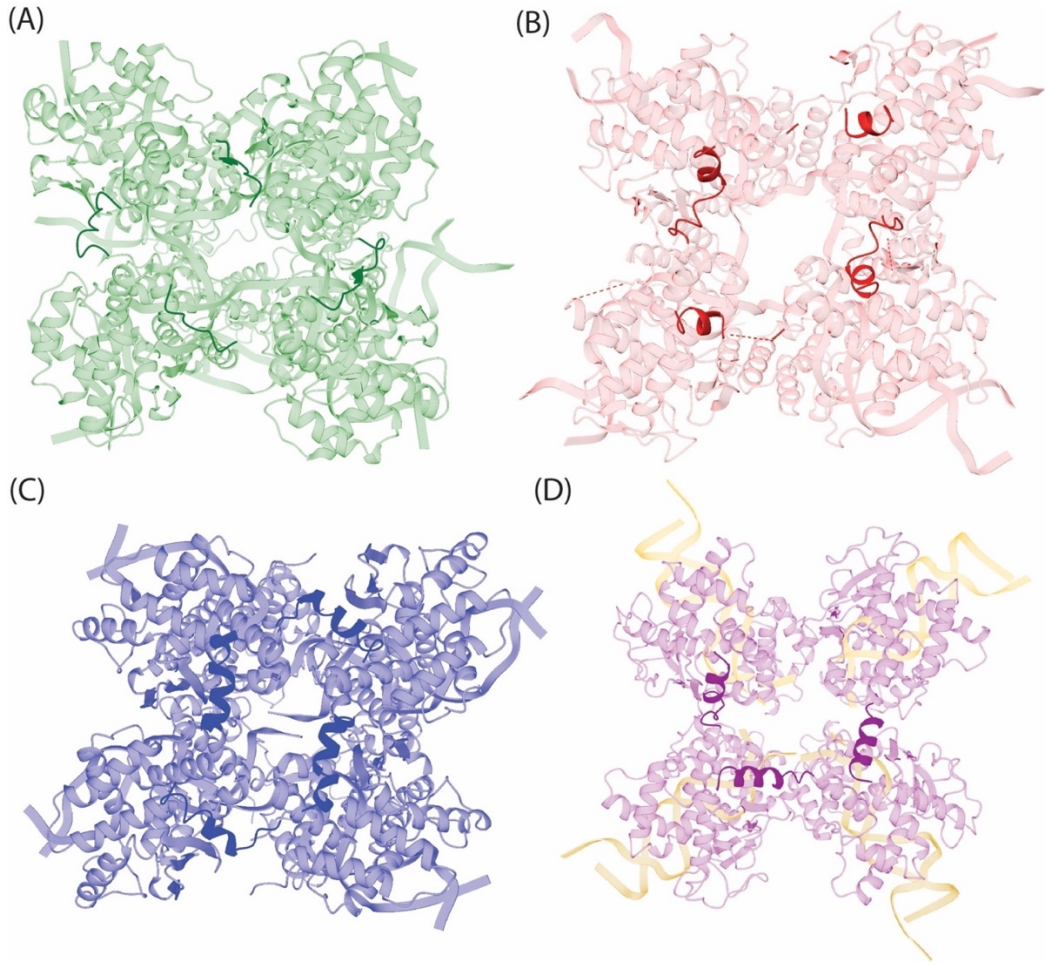

Figure 2-30: Comparison of the cyclic C-terminal domain exchange in the Int-HJ1TA structure with other tyrosine recombinase structures. Cartoon representation of (A) the λ Int-HJ complex (green), pdb 1z1g (Biswas et al., 2005); (B) the Cre-HJ complex (pink), pdb 2crx (Gopaul, Guo and Van Duyne, 1998); (C) the post-cleavage XerH-dif$_H$ synaptic complex (blue), pdb 5jjv (Bebel et al., 2016) and (D) the Tn1549 Int-HJ1TA complex (purple, this study). The circularly swapped C-termini are highlighted by showing darker color tones.

In all the tyrosine recombinase-HJ complex structures, the protein forms a stable tetramer in catalytically active conformation. DNA cleavage was avoided either by using a catalytic mutant protein or by introducing changes at the DNA sequence. In this study, I used a catalytic Int mutant (Int^{82N}R225K) and an immobile HJ to produce diffracting crystals and achieve structure determination. The overall conformation of my Int-HJ structure resembles equivalent structures of other tyrosine recombinases. Four recombinase monomers are bound to one HJ DNA substrate. The C-terminal helices interact with the neighboring subunits in a cyclic manner (Figure 2-30).

The overall Int monomer conformation and its binding to the specific HJ1TA in a C-clamp shape also resemble the other tyrosine recombinase structures. Int is structurally closest to the core fold of CB and CAT domains of the λ Int-HJ structure (Figure 2-26A). Both CB domains contain four α helices and their superimposition showed a nearly perfect fit. In turn, superimpositions with Cre and XerH highlighted differences in the CB domain (Figure 2-26B and C), where the number of α helices varies. This observation is in agreement with previous protein sequence alignments that showed that the CB domain is moderately well conserved in the tyrosine recombinase family (Rubio-Cosials *et al.*, 2018). On the other hand, all CAT domains also show good superimposition.

In conclusion, the Int-HJ1TA complex structure of the CTn Tn*1549* integrase reveals that several similarities to the tyrosine recombinase super-family. Moreover, it confirms multiple aspects of the proposed integration pathway, including that integration reactions indeed proceed through a tetrameric complex with a HJ substrate. The differences in target selection and cleavage specificity, including the recombination of non-homologous substrates, are probably overcome by different means: for example, by the introduction of the extra β-hairpin structure in the CAT domain or active opening of the CR region. Specific amino acid changes, that are not well-conserved, may play a role in unique protein-DNA interactions that still need to be discovered. Moreover, like λ Int, Tn*1549* Int contains an extra N-terminal domain that is missing in this structure. but may play a role in target site selection and in driving the chemical reactions forward.

2.1.3 Structural studies of Xis - an accessory protein in Tn*1549* transposition

For some members of the site-specific tyrosine recombinase family, previous studies have shown that accessory factors may be needed to make excision and integration reactions efficient, like in the case of λ-phage Int. One of them is the self-encoded excisionase (λ-Xis), which bends DNA and helps bringing the arm binding sites to λ Int's AB-domain (Landy, 2015). Xis is a recombination directionality factor (RDF). Similar proteins are found in numerous bacteriophages, transposons, and conjugative plasmids and are thought to help controlling the directionality of recombination reactions (Lewis and Hatfull, 2001). Apart from Xis, other accessory, host-encoded factors, such as IHF and Fis, are sometimes required for efficient excision or integration of various CTns and phages in Gram-negative bacteria. For instance, for the λ-phage, λ-Xis is necessary for excision and inhibits integration, while IHF and Fis stimulate both excision and integration (Landy, 2015).

Tn*1549* encodes an excisionase protein (Xis) in its transposition module (see section intro) (Garnier *et al.*, 2000). Tn*916*, the prototype of the Tn*916*-like family of conjugative transposons, also encodes a Xis^{916} protein (Roberts and Mullany, 2009). Xis^{916} binds to specific sequences in the transposon ends, which are located between the terminal core sites and the arm sites (Connolly, Iwahara and Clubb, 2002). Previous studies have shown that Xis^{916} is necessary for Tn*916* excision *in vivo* (Su and Clewell, 1993). It enhances excision when binding at the left transposon end, but inhibits the reaction when interacting with the right end (Connolly, Iwahara and Clubb, 2002). Similarly, for Tn*1549*, experiments in *E.coli* have shown that Xis is necessary for accurate excision and productive transposition (Lambertsen *et al.*, 2018). However, the exact molecular function of Xis remained ambiguous. Therefore, I decided to characterize Xis biochemically and structurally.

2.1.3.1 Expression and purification of Xis

The Xis construct was overexpressed in *E. coli,* as described in section 4.2.9.1 and purified as described in section 4.2.9.2. The Xis construct used in this study includes an N-terminal serine residue left from SenP2 cleavage of the 6xHis-SUMO tag. Moreover, it has two additional amino acids (L and E) at the C-terminus, and thus, it has a molecular weight of 8.2 kDa and a length of 69 aa compared to it is original sequence (7.87 kDa, 66 aa). Protein without this extra C-terminus rendered the construct unpurifiable. I obtained around 5 mg of over 99% pure protein after size-exclusion chromatography from 1 L culture (Figure

2-31). The peak fractions were used further for binding assays and crystallization experiments.

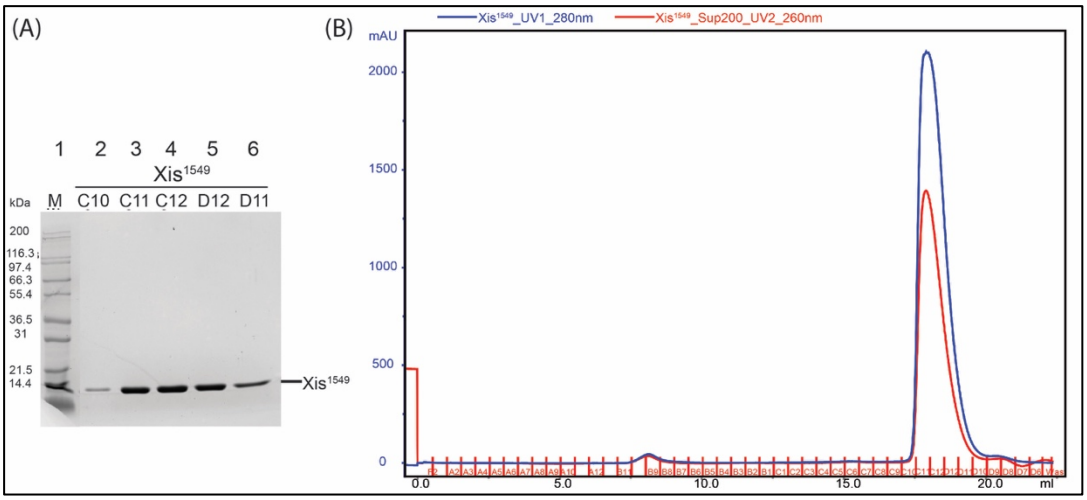

Figure 2-31: Purification of Tn1549 Xis. (A) SDS-PAGE analysis of the fractions collected after SEC, stained with Coomassie Blue for protein. Marker (Mark12™) band sizes are indicated on the left in kDa. (B) SEC on a Superdex 200 10/300 column. The chromatogram shows a prominent elution peak corresponding to a monomer of Xis (8.2 kDa). UV absorbance at 280 nm is shown in blue, absorbance at 260 nm is shown in red.

2.1.3.2 Design and validation of Xis-DNA binding sites

Previous studies of Tn*916* have shown that Xis^{916} has binding sites close to the transposon termini, as seen for λ-Xis in bacteriophage λ (Figure 2-32). For Xis^{916}, three consecutive binding sites have been documented (Connolly, Iwahara and Clubb, 2002). Tn*916* and Tn*1549* belong to the big family of Tn*916*-like CTns, and their excisionase proteins show high level of sequence similarity (Figure 2-33). Nevertheless, a quick sequence comparison at the Tn*916* and Tn*1549* transposon ends did not reveal similar Xis^{916} binding sites in the Tn*1549* ends.

Therefore, a bioinformatic approach to predict conserved Xis binding sites was performed, with the help of postdoc Georgy Smyshlyaev from the Barabas Lab. The process consisted of performing several sequence alignments to compare defined regions at the transposon ends. For that, available Tn*916*-like transposon sequences from the ICEberg database https://db-mml.sjtu.edu.cn/ICEberg/index.php (Liu *et al.*, 2019), were used to look for conserved motifs. The results showed no conserved binding sites for Xis in Tn*1549* (data not shown).

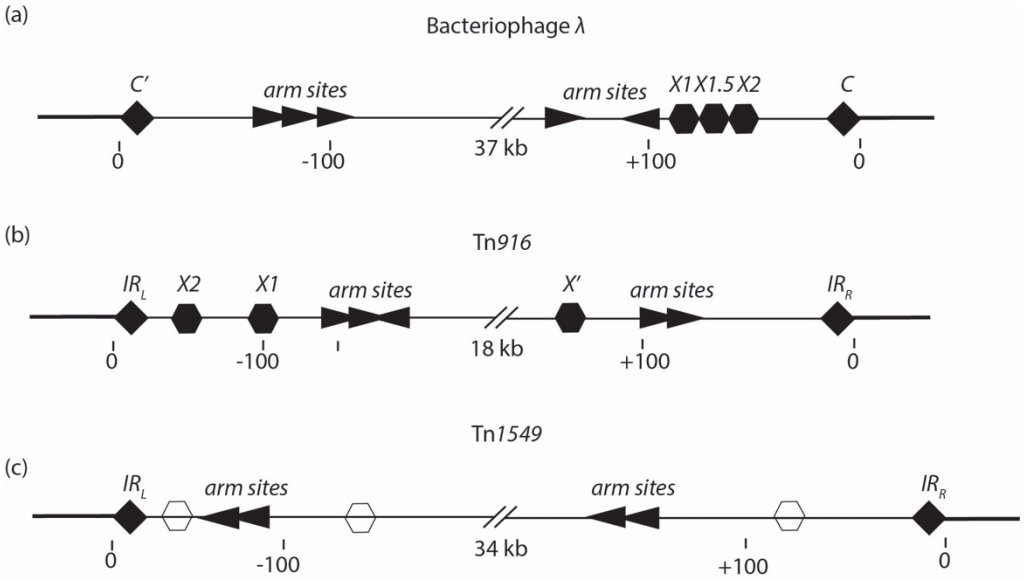

Figure 2-32: The arrangement of DNA sequences involved in protein binding in (a) bacteriophage λ, (b) Tn916, and (c) Tn1549. C/C' = bacteriophage λ ends, IR_L/IR_R = Tn916 and Tn1549 ends (black rhombus). X = Xis binding site (black hexagon), arm binding sites (black triangle). Xis binding sites are not known and only highlighted as white hexagons, assuming a similar location as in λ-phage or Tn916.

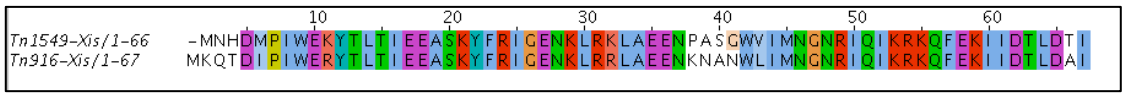

Figure 2-33: Sequence alignment of the Tn916 Xis and the Tn1549 Xis proteins. The alignments were prepared by ClustalOmega (Sievers et al., 2011) and visualized using the Jalview software (Waterhouse et al., 2009). The residues are colored according to the ClustalX scheme in Jalview. Non-conserved amino acids are not colored.

Thus, I tested Xis binding to various DNA sequences (Figure 2-34), comprising regions from both Tn*1549* transposon ends, experimentally by EMSA. First, I PCR amplified the long sequences at the right and left transposon ends, starting at the inverted repeat and spanning a few hundred bp over the arm sites. Another approach was to test binding to smaller segments of these regions. Figure 2-35 shows an example of an EMSA performed to test Xis-DNA binding. In general, none of the EMSAs did show any specific binding of Xis to the DNA substrates (long or short), even when high amounts of protein were used to overcome weak binding (Figure 2-35).

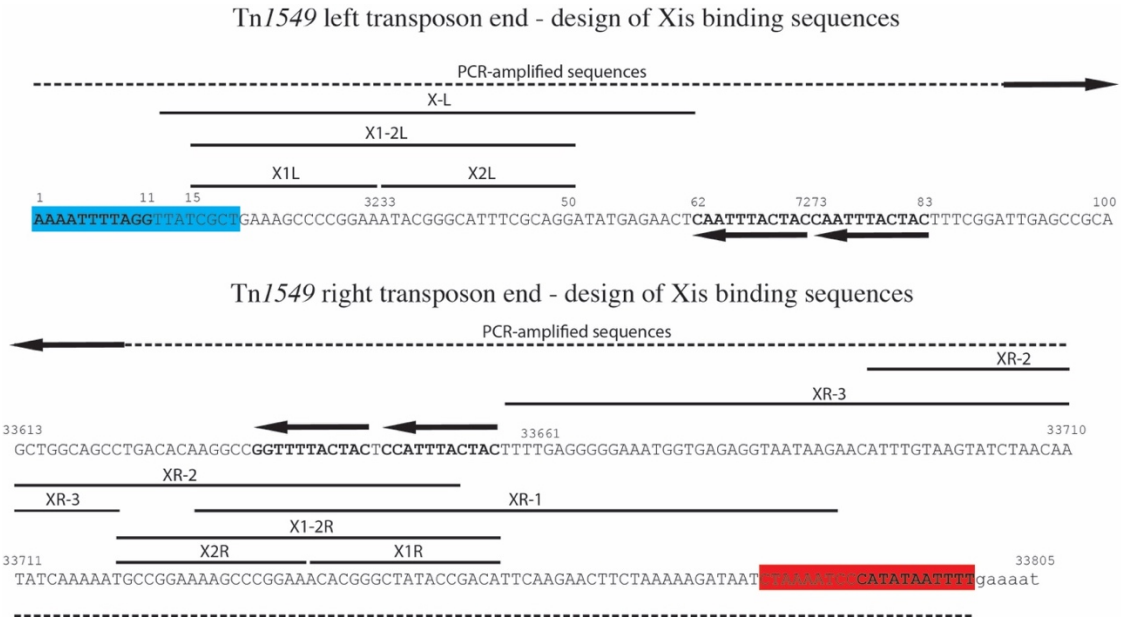

Figure 2-34: Design of the Xis substrates used for biochemical experiments. Left end sequence (top): the core IR_L site is in bold (1-11 bp) and highlighted in blue. Internal arm sites (bold) are highlighted with arrows. Right end sequence (bottom): the core IR_R site is in bold and highlighted in red. Internal arm sites (bold) are highlighted with arrows. Left (L) and right (R) end sequences used to test Xis binding are marked with black lines on top of each sequence.

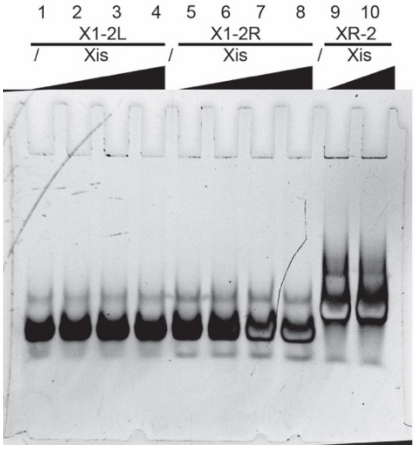

Figure 2-35: Characterization of Xis DNA binding. The native PAGE gel shows 8 µM DNA substrates (as labelled above the gel) mixed with increasing concentrations of Xis (0, 20, 40, and 60 µM). XR-2 was mixed with (0, 20 µM) protein. DNA was stained with SybrGold.

2.1.3.3 Crystallization of Xis

After protein purification, Xis was prepared for crystallization, as described in section 4.4.1.1. Then, I continued with high-throughput crystallization experiments and obtained many hits using the JCSG+ screen at 20 °C. These crystals had already a good fishable-size (20x60x15 µm) and single-crystal, rod shape. Thus, several of them were prepared for initial diffraction experiments and data collection, as described in section 4.4.1.2.

2.1.3.4 Data collection, structure solution and refinement of Xis

The best diffracting crystals grew as single rods in 96-well plates using the sitting-drop method at 20 °C using a crystallization buffer with 0.2 M LiSO4, 0.1 M BisTris pH 5.5, 25% (w/v) PEG 3350 (E12) from the JCSG+ Screen (Molecular Dimensions). Diffraction data was collected to 1.5 Å resolution at the beamline ID-29 at the ESRF in Grenoble (Figure 2-36, section 4.4.1.3).

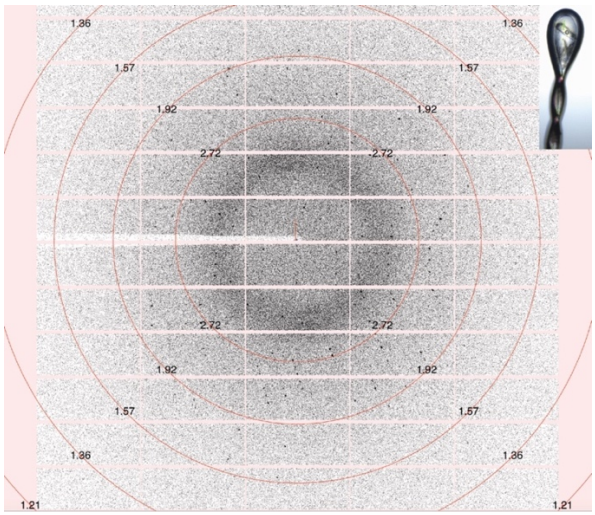

Figure 2-36: Crystallization and X-ray data collection of Xis. Image of one Xis crystal in the loop, prepared for data collection (right top corner). Example diffraction pattern obtained from the crystal shown. Resolution rings are given in Å.

The dataset was processed using XDS (Kabsch *et al.*, 2010) (section 4.4.1.4). Initial structure solution via molecular replacement using Phaser (McCoy *et al.*, 2007) and previously determined Xis structures (like Xis^{916}, pdb:1y6u; λ-Xis, pdb:1rh6, among others) as input did not work. Therefore, the structure solution was performed by using ARCIMBOLDO Lite (Rodríguez *et al.*, 2009) in the CCP4 suite (section 4.4.1.5). It is an ab initio phasing method, based on a combination of the location of model fragments like small polyalanine α-helices with Phaser and density modification with SHELXE (Thorn and Sheldrick, 2013). We searched initially for 2 ideal polyalanine α-helices of 12 residues each with Phaser and iterative density modification and autotracing with SHELXE. The initial α-helical model was then manually extended and the Xis sequence was docked with COOT (Emsley *et al.*, 2010). The structure was solved in space group P65. The final model was obtained through iterative cycles of model building in COOT and refinement in PHENIX (Liebschner *et al.*, 2019) (section 4.4.1.6). Refinement has finally R-factors of R_{work} = 0.1342 and R_{free} = 0.1628 (Table 2-5), indicating a very good correlation between the model and the observed diffraction data (Figure 2-37). The stereochemistry of the model solution was evaluated by the information obtained from the Ramachandran plot, which

showed that 100% of the protein residues are found in the favored Ramachandran region. Moreover, the low root mean square (R.m.s) deviations (0.006 Å for bond lengths and 0.75° for bond angles) together with none Ramachandran outliers, indicate good protein geometry.

Table 2-5: Crystallographic data collection and refinement statistics for the Xis structure. R.m.s = root mean square.

	Native
Data collection	
Space group	P 65
Cell dimensions	
a, b, c (Å)	58.32, 58.32, 26.35
a, b, g (°)	90, 90, 120
Wavelength (Å)	1.07227
Resolution (Å)	19.09 - 1.5 (1.554 – 1.5)
R-merge	0.08822 (1.162)
R-meas	0.09322 (1.227)
I / sigma (I)	13.94 (1.92)
CC1/2	0.999 (0.722)
Completeness (%)	99.92 (99.88)
Redundancy	9.7 (9.8)
Refinement	
Resolution	19.09 - 1.5
No. reflections	8376 (840)
R_{work} / R_{free}	0.1342 / 0.1628
No. atoms	653
Macromolecule	598
Water	55
Average B-factor	13.76
Protein	12.43
Water	28.24
R.m.s deviations	
Bond lengths (Å)	0.006
Bond angles (°)	0.75
Ramachandran favored (%)	100.00
Ramachandran outliers (%)	0.00

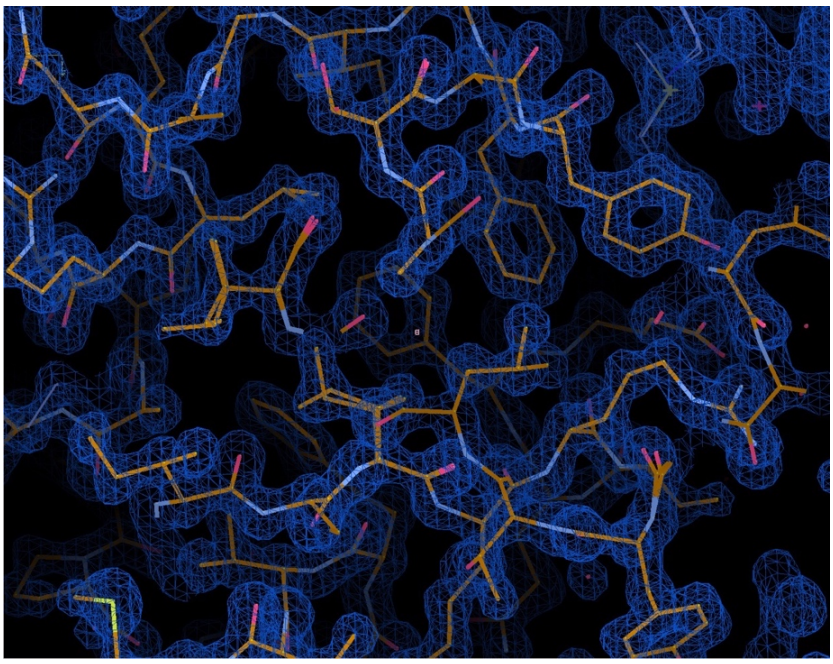

Figure 2-37: 2Fo-Fc electron density map at contour level of 1.5 sigma for a region of the Xis structure after refinement at 1.5 Å resolution.

2.1.3.5 Overview of the Xis structure

The structure shows a single molecule of Xis (Figure 2-38) in the crystal asymmetric unit. The protein used for crystallization consisted of 69 amino acid residues, which includes an N-terminal serine residue left from SenP2 cleavage of the 6xHis-SUMO tag. The first three amino acids at the N-terminus could not be built due to poor density in this region, probably resulting from flexibility. All other residues could be resolved in the final refined model. For some residues, even alternative rotamer conformations could be observed in the high-resolution data.

Figure 2-38: Overview of the Xis structure (golden), shown in cartoon representation. HTH = Helix-turn-helix is formed by αA and αB. The wing (W) motif is created by β-strands β2 and β3 connected by a loop.

The Xis molecule is composed of three α-helices and three β-strands (Figure 2-39). It has the conserved winged-helix fold, typical for many excisionase proteins. The fold is formed by a β-sheet, composed of three antiparallel β-strands, a helix-turn-helix (HTH) motif inserted between β-strands β1 and β2 and a C-terminal α-helix (αC).

The HTH motif is formed by αA (I15-F22) and αB (E26-E35), where the turn is composed of 3 residues (R24-G26), which points in the opposite direction from the β-sheet of the wing. The antiparallel β-sheet is formed by β1 (T12-T14), β2 (V42-N45), and β3 (R48-R53). Moreover, there is a 3_{10}-helix (P6-K10) located before β1, where P6 forms H-bond with E9 and I7 with K10 in an i+3 manner, as expected for this type of secondary structures.

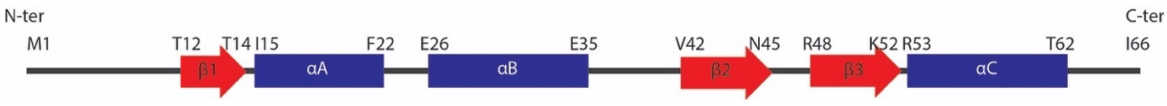

Figure 2-39: Secondary structure of Xis. It is composed of three α-helices and three β-strands, which form a conserved winged-helix fold typical for the excisionase family (α-helices in blue boxes, β-strands in red arrows).

2.1.3.6 Discussion

2.1.3.6.1 Predicting Tn*1549* Xis binding sites

Studies of λ-Xis, have shown that this type of protein binds cooperatively to DNA. Bacteriophage λ has three consecutive binding sites together building a 30 bp total length (Figure 2-32), where the first and second sites have very similar sequences (Abbani *et al.*, 2007). Xis^{916} also binds 30-40 bp long sequences, as shown by footprinting experiments (Connolly, Iwahara and Clubb, 2002). However, the Xis^{916} sites do not share strong sequence conservation among themselves. The bioinformatic prediction of Xis binding sites based on similarities to other excisionase sites or on internal sequence conservation was not successful, suggesting a relaxed specificity of DNA binding by this protein as compared to λ-Xis. Biochemical assays also did not show detectable Xis binding to various DNA sequences encompassing the regions between the core- and internal arm sites of Tn*1549*.

Similar to my results for Xis, Xis^{916} showed weak binding to its specific sites in EMSA experiments (Hinerfeld and Churchward, 2001). Thus, DNase I protection experiments had to be performed to find and confirm Xis^{916} binding sites (Connolly, Iwahara and Clubb, 2002). Xis^{916} and Xis have a similar overall fold and feature high sequence identity (around 95%, Figure 2-33). Thus, it is surprising that Xis^{916} and Xis sites differ that much and do not show conserved features. Future DNase I footprinting assays may help to define if Xis has specific binding sites in the transposon end DNA and compare them to the ones found in Tn*916*.

2.1.3.6.2 Tn*1549* Xis structure in the context of other Xis proteins

Although Xis is very similar to Xis^{916} at the sequence level (Figure 2-33), their DNA binding sites are not conserved. The structure of Xis^{916} was solved previously by NMR (Abbani, Iwahara and Clubb, 2005). A superimposition shows a good fit between the two structures, as expected from their high sequence similarity (Figure 2-40). Thus, small differences at the sequence level may be the reason for the divergence of their DNA binding sites. Alternatively, both Xis proteins may rely on indirect sequence readout for DNA binding, which is consistent with the dissimilarity of multiple Xis-binding sites even within the same element.

Although a protein sequence alignment of diverse members of the excisionase family shows that these proteins can vary significantly at the sequence level, their molecular

structures are generally very similar (Figure 2-41). They share the conserved winged-helix fold that is characteristic for this protein family in bacteria and bacteriophages (Abbani, Iwahara and Clubb, 2005).

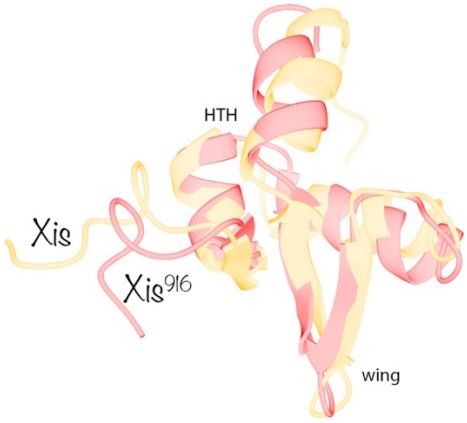

Figure 2-40: Superimposition of Xis and Tn916 Xis structures. Cartoon representation of Xis (golden, this study) and Tn916 Xis (red, pdb:1y6u) structures.

Excisionase proteins bind to DNA by inserting their wing region into the DNA minor groove and one helix of the HTH domain into the major groove. The structure of λ-Xis in complex with DNA revealed four arginine residues that render the protein surface positively charged and enable interactions with the DNA sugar-phosphate backbone (Sam *et al.*, 2004).

Superposition of the Xis structure with the Xis^{916} and the λ-Xis-DNA structures (Figure 2-42) (Abbani, Iwahara and Clubb, 2005) shows that the λ-Xis residue R22, which is important for HTH-major groove interactions, is well conserved in both Xis^{916} and Xis (residues R31 and R30, respectively). Two more of the four essential DNA-interacting arginine residues in λ-Xis are also located in the HTH αB. The λ-Xis R23 was replaced in Xis with K31, whereas R26 has changed in both CTn excisionases to glutamate (E35 and E34 in Xis^{916} and Xis, respectively), switching even the charge from positive to negative.

The fourth important arginine residue in λ-Xis, R39 is located at the second protein-DNA binding interface, mediating the interaction between the DNA minor groove and the wing domain. This residue is conserved in both Xis^{916} and Xis (residues R49 and R48, respectively). This highlights the wing domain's importance in DNA recognition and supports the idea that Xis may bind DNA in a very similar manner.

In conclusion, Tn*1549* Xis forms weak interaction with its DNA sites, similarly to what had been reported for the Xis^{916} protein. It shares the conserved winged-helix fold with other

Xis proteins in prokaryotes. Moreover, the λ-Xis, Xis^{916}, and Xis structures share residues in the predicted DNA interaction regions. For the Xis^{916} and Xis proteins, only a single amino acid change (R to K) in the HTH αB region could be identified, which is unlikely to be responsible for their different binding sites.

Another interesting aspect of Xis-DNA interactions is binding cooperativity. In the case of λ-Xis, the X1 and X2 sites are very similar and show strong binding in EMSAs, while X1.5 can be bound by λ-Xis only in cooperative manner, with X1 and X2 sites present and preoccupied by λ-Xis. Thus, cooperative binding might compensate for the rather low DNA affinity of CTn Xis proteins, perhaps playing an even more important role for these proteins. A reason for weak Xis-DNA interaction may be that excisionase proteins from CTns play a somewhat different role in transposition, as it was already proposed for Xis^{916} (Abbani, Iwahara and Clubb, 2005). They may not regulate excision as strongly as they do in the λ bacteriophage system, thus, allowing CTns to move more freely between bacteria.

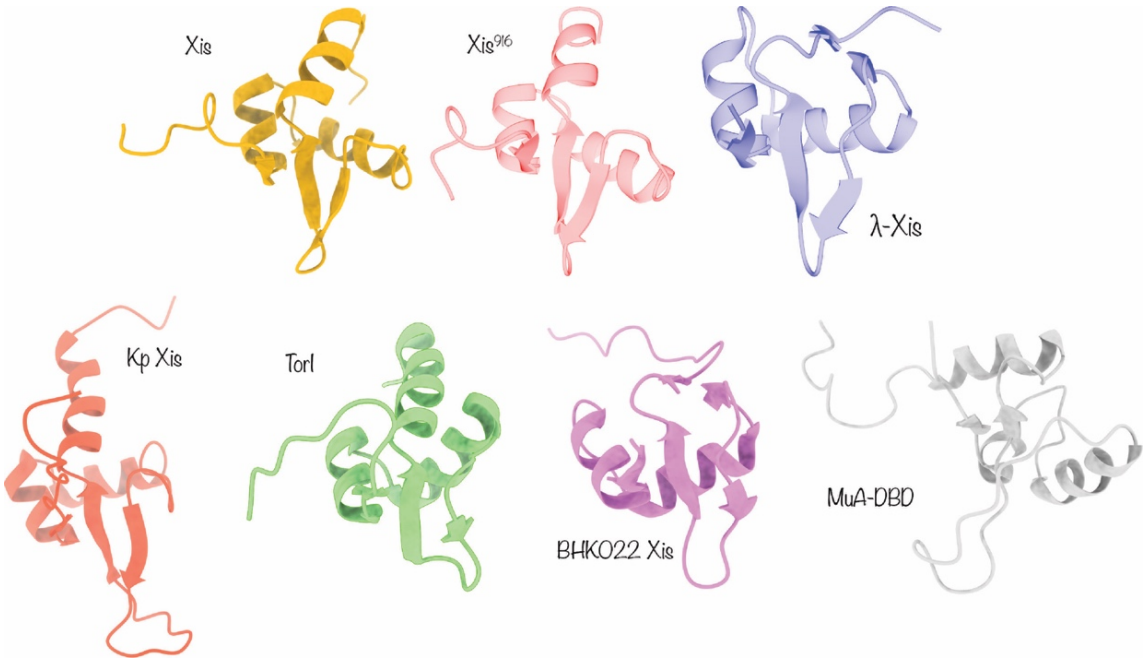

Figure 2-41: Cartoon representation of various excisionase proteins and DNA-binding domains that share the conserved winged-helix domain. The structures are Tn1549 Xis (golden, this study), Xis^{916} (pink, pdb:1y6u), λ-Xis (blue, pdb:1rh6), putative Xis from K. pneumoniae (red, pdb:2kvv), atypical Xis from the response regulator TorI (green, pdb:1z4h), Bacteriophage HK022-Xis (violet, pdb:1pm6), and MuA-DBD (grey, pdb:1tns), starting from the top left. DBD = DNA binding domain.

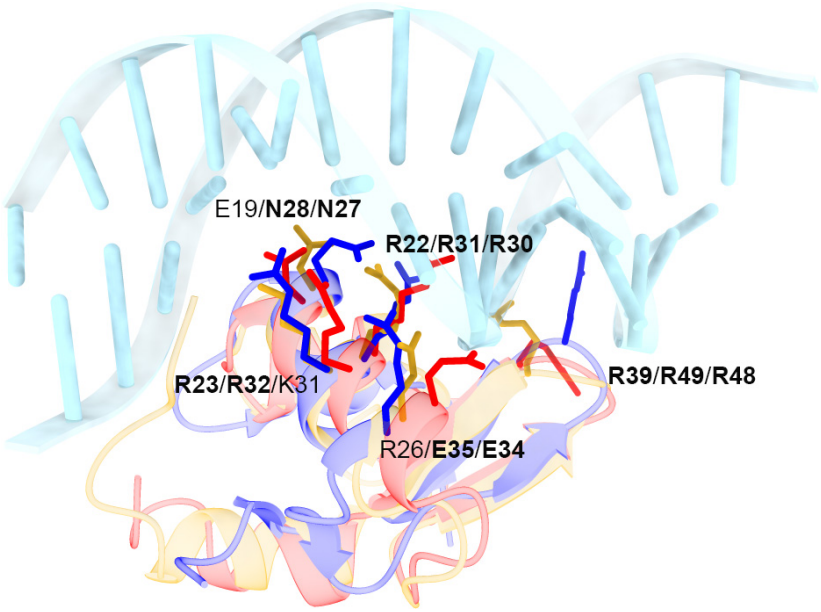

Figure 2-42: Superimposition of λ-Xis-DNA, Xis[916] and Xis structures. Cartoon representation of λ-Xis-DNA (blue, pdb:1rh6), Xis[916] (red, pdb:1y6u), and Xis (golden, this study). Amino acids that are involved in DNA recognition and binding in λ-Xis and their structural equivalents in Xis[916] and Xis are shown in stick representation. Highlighted residues are named in the following order: λ-Xis, Xis[916] and Xis. Residues conserved in at least two proteins are in bold.

2.2 Characterization of HJ resolution in Tn*1549* transposition

Int from Tn*1549* is proposed to perform the recombination reactions for transposon excision and integration in a similar way to other site-specific tyrosine recombinases. Nevertheless, Int shows some differences, characteristic to CTn integrases (see section 1.6.3). One of the main differences being that these proteins can recombine non-homologous sites leading to DNA products that have mismatches at their crossover regions (Figure 2-43).

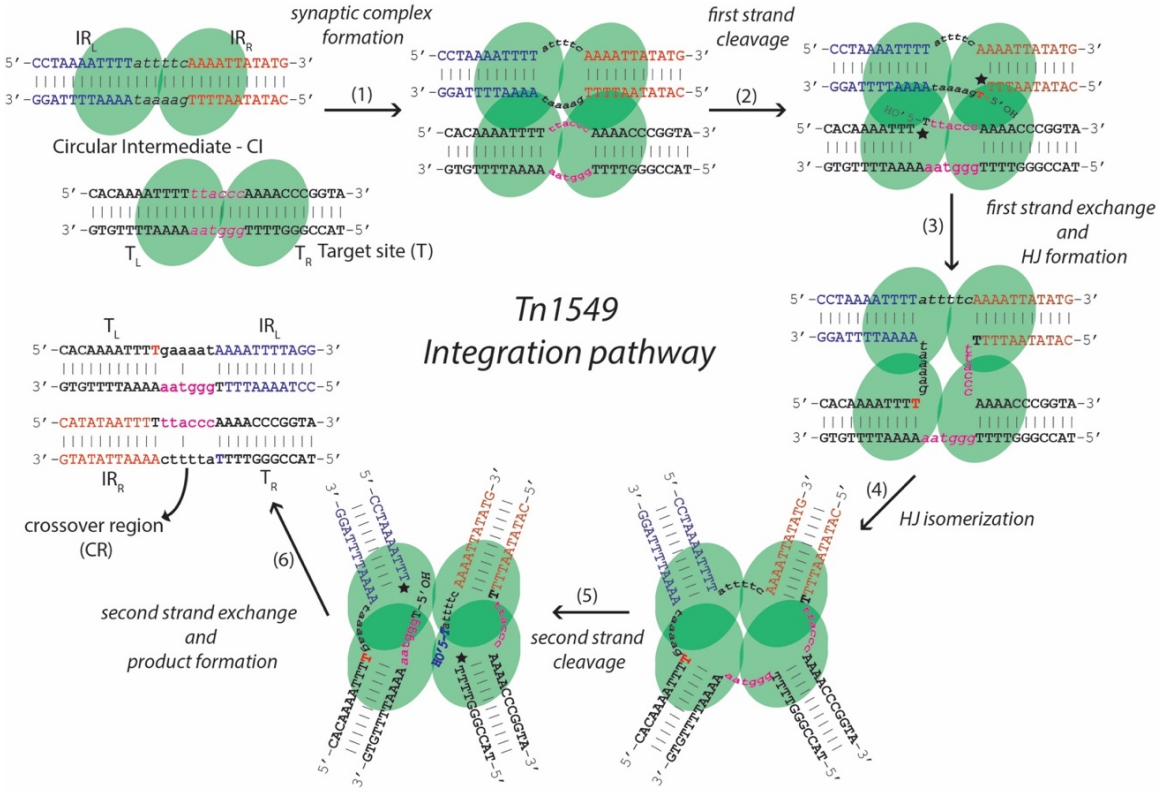

Figure 2-43: Schematic representation of the proposed Tn1549 integration pathway. One Int dimer binds at the CI and another at the target site, leading to DNA opening at the crossover region (1). After synaptic complex formation, the first cleavage (2) and strand exchange (3) reactions occur, leading to HJ intermediate formation. After isomerization of the HJ intermediate (4), the second round of strand cleavage (5) and exchange reactions (6) will occur, leading to recombined products and transposon integration. The recombined sites will have a mismatched CR of up to 6 bp length. In this example, CRs with a 5 bp long mismatch are formed. Base-pairing is highlighted by dashes between the opposing DNA strands.

Previous studies in the lab revealed the structure of an Int-CI complex (Rubio-Cosials *et al.*, 2018), in which the integrase is bound as a dimer to an almost straight DNA molecule, forming an inactive pre-catalytic state. This observation is in stark contrast to previous structures of other site-specific tyrosine recombinases, where synaptic complexes, composed of four protein monomers and two DNA molecules, were trapped in the pre-cleavage state with strongly bent DNA (see section 1.4.3). Thus, these results raised the

question whether elementary recombination steps may differ between CTn integrases and site-specific recombinases. In the first part of this work, I could show that Int can form stable complexes with HJ DNA (see section 2.1.2.3). Therefore, to complement structural work on the Int-HJ complex, I next aimed to test if Int can resolve the integrative HJ intermediate into recombined products. As HJ resolution is an essential step in canonical tyrosine recombination, validating the ability of Int to perform this reaction would confirm the involvement of HJ formation in CTn transposition.

2.2.1 Resolution of a symmetrized HJ substrate

To probe Int activity on HJ DNA, I set up a HJ resolution assay, which allowed me to monitor the resolution of a preformed HJ construct by Int (see section 4.3.7). The HJ substrates were designed to mimic the intermediate of transposon CI integration and contained four double-stranded DNA stems representing the two transposon ends (IR_L and IR_R) and the integration target sites (T_L and T_R). Upon addition of Int, the reaction could go in two directions: either to products, where T_L is linked to IR_L and IR_R to T_R, representing CTn integration; or back to substrates, where a CI with IR_L-IR_R link and the target site T_L-T_R are formed (Figure 2-44).

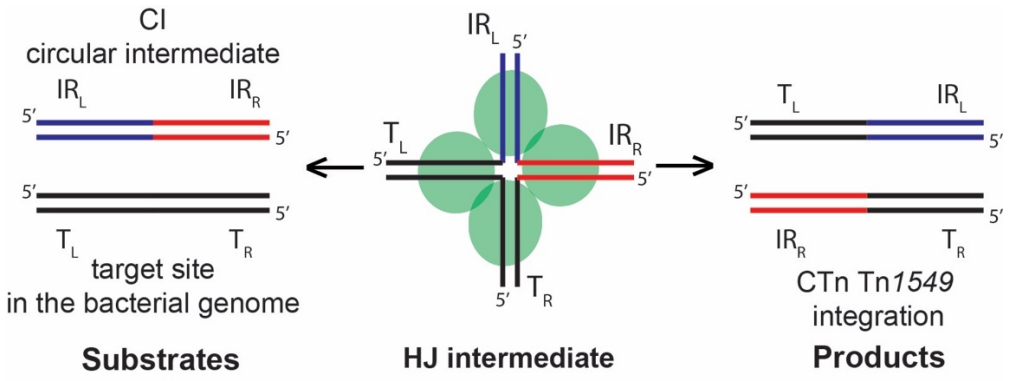

Figure 2-44: Schematics of HJ resolution. The HJ intermediate is composed of four stems, IR_L (blue), IR_R (red), and target sites T_L and T_R (black). Recombination in direction to products will lead to T_L-IR_L and IR_R-T_R sites. HJ resolution back to substrates will lead to the formation of the circular intermediate with joined transposon ends (IR_L-IR_R) and the bacterial target site (T_L-T_R).

As a first variant, I used the HJ1 DNA design from the Int-HJ crystallization experiments (see section 2.1.2.2). The construct had to be slightly modified for the HJ resolution assay to confidently separate various potential products of the reaction. For that, I differentially extended the lengths of all four HJ stems. Then, the four-way HJ DNA intermediate was radiolabeled with ^{32}P at either one or two 5' ends. In this way, DNA size changes could be

monitored on a sequencing-grade TBE-Urea gel. The first HJ construct was called HJ1-A (Figure 2-45A).

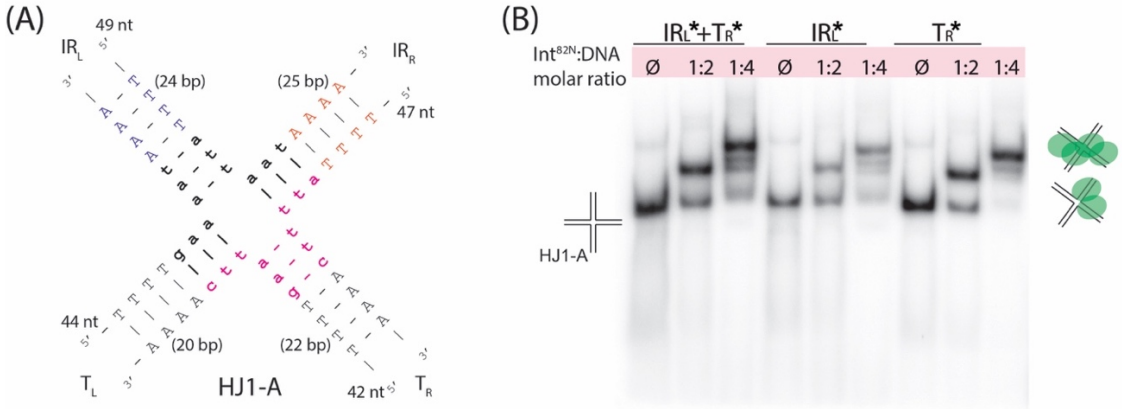

Figure 2-45: (A) Design of the HJ1-A construct for HJ resolution experiments. The HJ DNA mimics the intermediate of CTn integration after the first round of cleavage and strand exchange occurred at IR_R and T_L. The CR bases were mutated to obtain a stable and fully base-paired HJ. HJ stems had different lengths allowing to follow changes in the sequence length on a denaturing TBE-Urea gel. Stem lengths are given in bp. Oligonucleotide length for each DNA strand is given next to the 5'ends (in nt). Only the central sequence of the HJ DNA is shown. (B) The native PAGE gel shows Int^{82N} binding to HJ1-A at molar ratios of 1:2 and 1:4 (DNA to protein). HJ1-A constructs were radiolabeled at the IR_L or T_R arm or both.

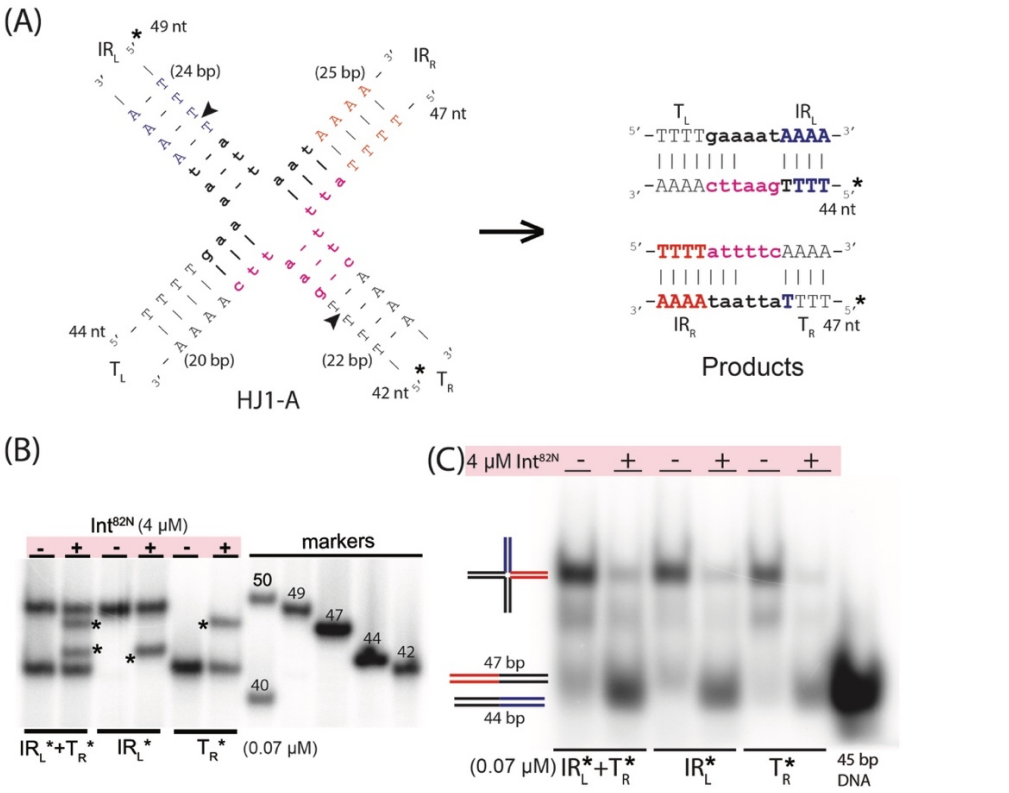

Figure 2-46: Characterization of Int's HJ resolution activity on HJ1-A in direction to products.

(A) Scheme of HJ1-A and its products (IR_L in blue, IR_R in red, and T-sites in black). The exact sequence is only shown for the central regions. The CI crossover region is in black and the target site CR is in magenta. Base-pairing is highlighted with dashes. Int-mediated cleavage and strand exchange at the site indicated by black arrows leads to the final integration products (T_L-IR_L, 44 nt, and IR_R-T_R, 47 nt).

(B) Denaturing sequencing PAGE gel showing resolution of HJ1-A intermediates in direction to products. Products are highlighted with a star on the gel.

(C) Native PAGE gel showing the radiolabeled DNA products that were also used in the sequencing gel. HJ1-A remained intact in reactions without (-) Int^{82N}. When protein was added (+) formation of dsDNA products can be observed, confirming HJ resolution.

PAGE analysis of Int^{82N}-HJ1-A complexes on native gels with radioactively labeled HJ DNA showed that Int^{82N} binds to the four-way junction and forms a shift, corresponding to a tetramer at 1:4 DNA:protein ratio (Figure 2-45B). Next, I tested HJ1-A resolution in both directions, to products (Figure 2-46B and C) and back to substrates (Figure 2-47B and C) by differentially labelling the various HJ1-A stems. The results show that Int^{82N} can assemble a functional complex on HJ1-A and catalyze formation of integration products. Int^{82N} resolved HJs leading to products with a mismatch at the crossover region, as expected for the native reaction (Rubio-Cosials *et al.*, 2018). Notably, Int^{82N} supported HJ1-A resolution back to the original recombination substrates with similar efficiency, suggesting that the protein does not preferentially drive the reaction forward to the native recombination products.

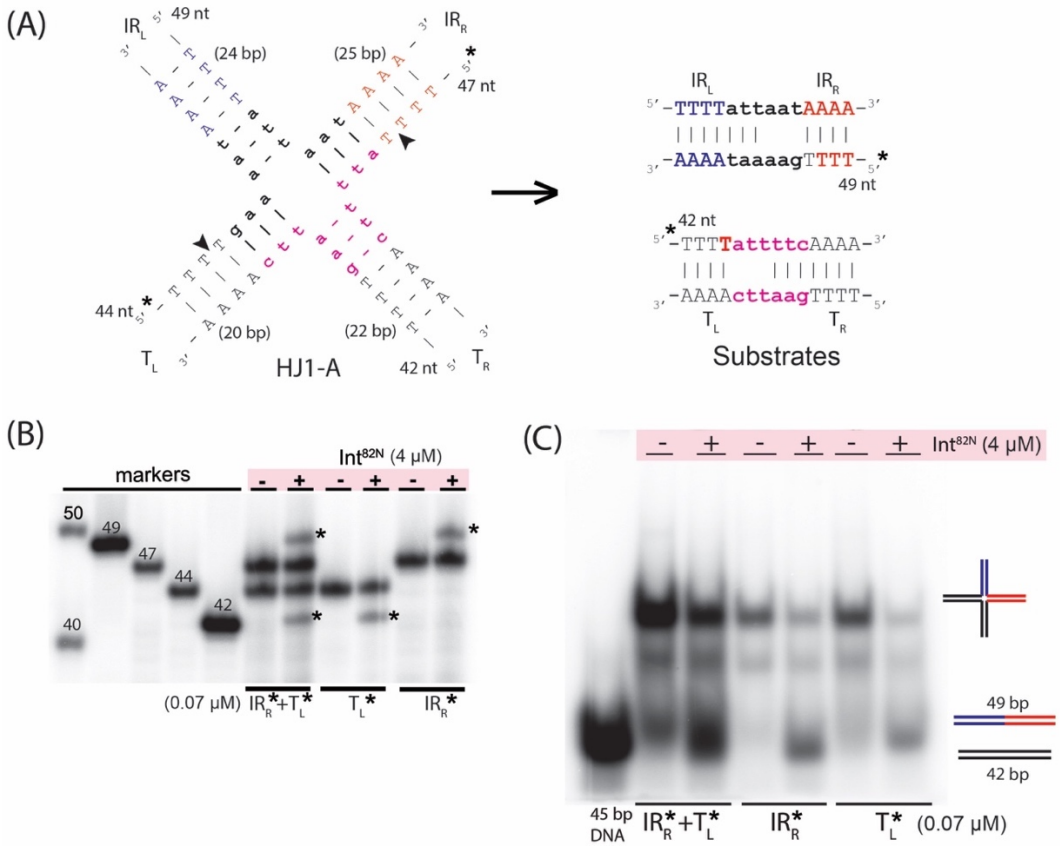

Figure 2-47: Characterization of Int's HJ resolution activity on HJ1-A in direction to substrates.

(A) Scheme of the HJ1-A and its products (IR$_L$ in blue, IR$_R$ in red, and T-sites in black). The exact sequence is only shown for the central regions. The CI crossover region is in black and the target site CR is in magenta.

Base-pairing is highlighted with dashes. Int-mediated cleavage and strand exchange at the site indicated by black arrows goes backward to substrates (IR_L-IR_R, 49 nt, and T_R-T_L, 42 nt).

(B) Denaturing sequencing PAGE gel showing resolution of HJ1-A intermediates in direction to substrates. Products are highlighted with a star on the gel.

(C) Native PAGE gel showing the radiolabeled DNA products that were also used in the sequencing gel. HJ1-A remained intact in reactions without (-) Int^{82N}. When protein was added (+) formation of dsDNA products can be observed, confirming HJ resolution.

2.2.2 Role of the crossover region (CR) in HJ resolution

While the first assays showed that Int can bind and resolve HJ DNA, several questions remained open. In particular, I introduced artificial base-pairing in the first HJ design to stabilize the junction, so it remained unclear how lack of DNA homology affects HJ resolution. I decided to address this question by changing the CR of HJ1-A to create HJs with different unpaired crossover regions (Figure 2-48). HJ1-B contained the native sequences that are expected to form after the first strand cleavage and recombination reactions. For HJ1-C, I introduced base mismatches in two stems to stabilize the unpaired intermediate.

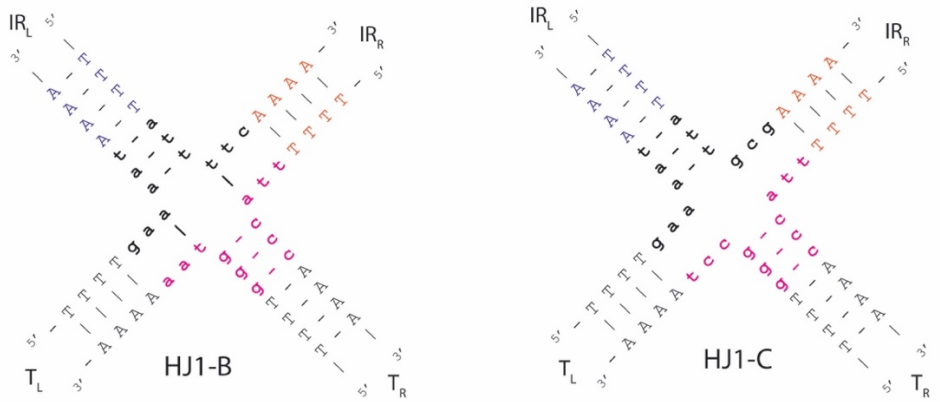

Figure 2-48: Design of HJ1-B and HJ1-C constructs for HJ resolution experiments. Both HJs are variants of HJ1-A. HJ1-B contains the native CR sequences. HJ1-C has base mismatches in two stems to stabilize the intermediate. Only the HJ's central region is shown for comparison. IR_L in blue, IR_R in red, and T-sites in black. CR from the CI sequence is in black and lowercase, and the CR from the target sequence is in magenta and lowercase.

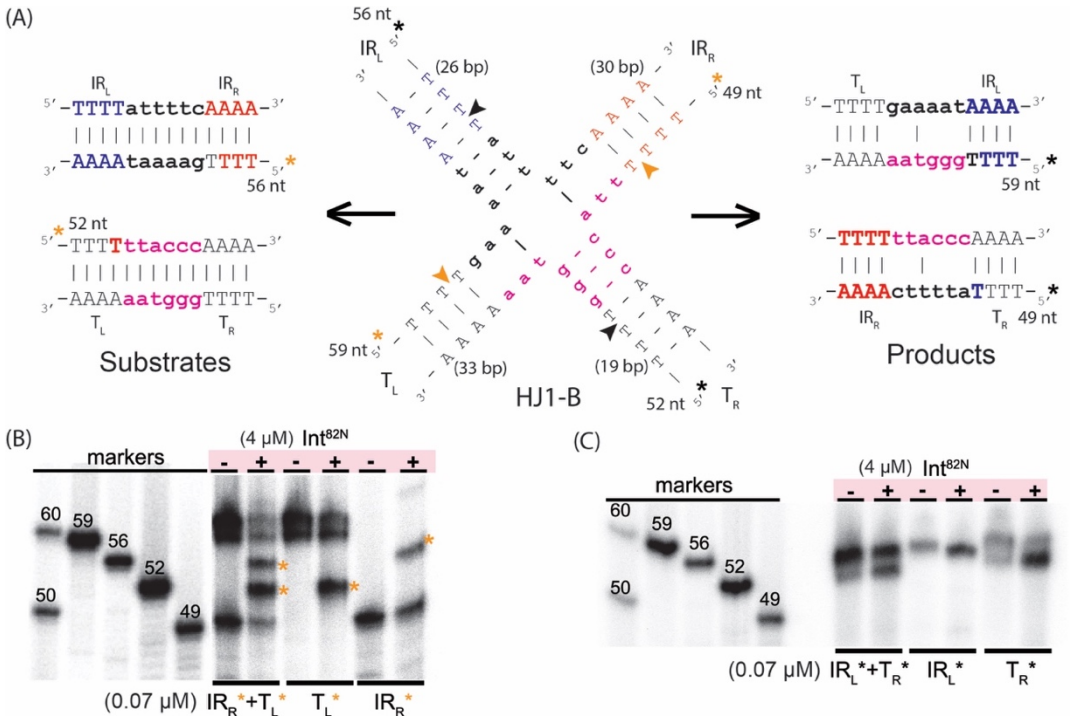

Figure 2-49: Characterization of Int's HJ resolution activity on HJ1-B.

(A) Scheme of HJ1-B resolution (IR_L in blue, IR_R in red, and T-sites in black). The exact sequence is only shown for the central regions. The CI crossover region is in black and the target site CR is in magenta. Base-pairing is highlighted with dashes. Int-mediated cleavage and strand exchange at the site indicated by black (products) or orange (substrates) arrows leads to the final integration products (T_L-IR_L, 59 nt, and IR_R-T_R, 49 nt, right) or goes backward to substrates (IR_L- IR_R, 56 nt, and T_L-T_R, 52 nt, left). Black and orange stars highlight the ^{32}P radiolabeled DNA strands.

(B) Denaturing sequencing PAGE gel showing resolution of HJ1-B intermediates in direction to substrates. Products are highlighted with a star (orange) on the gel.

(C) Denaturing sequencing PAGE gel showing resolution of HJ-B intermediates in direction to products. In this case, only the initially radiolabeled oligos can be observed and no product bands are seen at the expected size. - = no protein was added to the reaction (negative control), + = protein was added.

Both HJs (HJ1-B and -C) showed resolution back to substrates when incubated with Int^{82N} (Figure 2-49B and Figure 2-50B). On the other hand, resolution to products could not be observed (Figure 2-49C and Figure 2-50C). These experiments indicate that Int^{82N} do not intrinsically prefer to resolve the HJ1 DNA to the native products. Instead the direction of resolution may be biased by the number of possible base pairs in the CRs of the resulting dsDNA molecules. In the case of HJ1-B, resolution in direction to substrates leads to a fully base-paired CR, while resolution in the other direction would generate a 5 bp long mismatch (Figure 2-49A). In the case of HJ1-C, resolution back to substrates leads to 3 base-pairs in the CR, whereas resolution in direction to products would create a CR with a 6 bp long mismatch (Figure 2-50A).

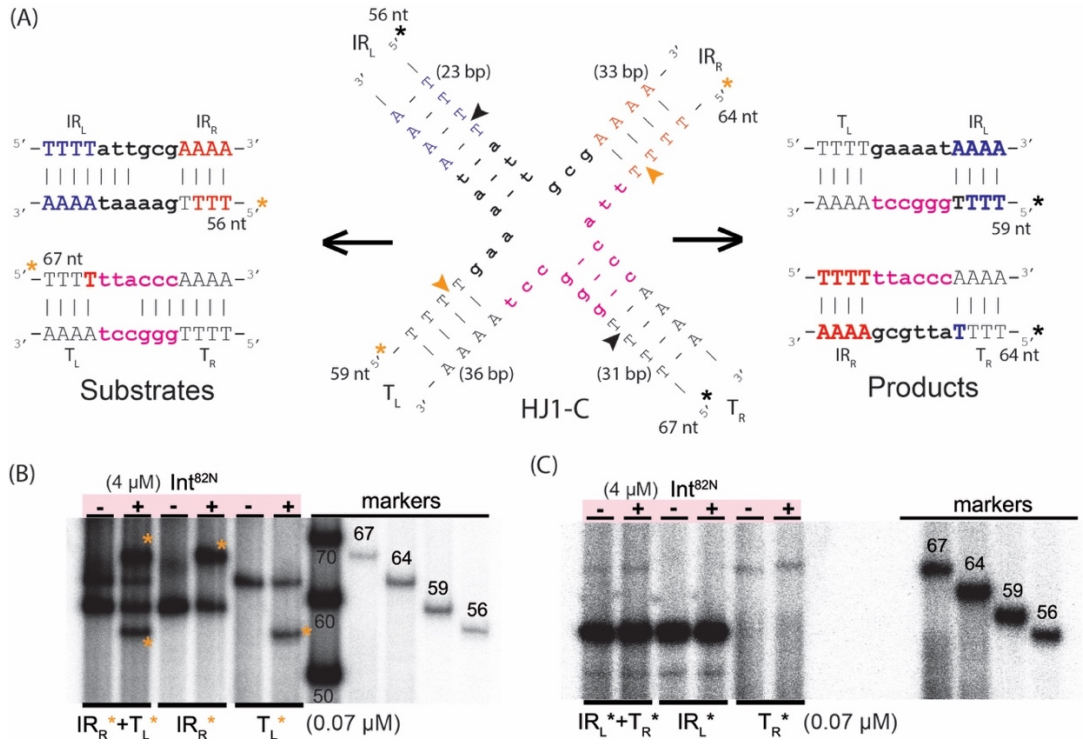

Figure 2-50: Characterization of Int's HJ resolution activity on HJ1-C.

(A) Scheme of HJ1-C resolution (IR_L in blue, IR_R in red, and T-sites in black). The exact sequence is only shown for the central regions. The CI crossover region is in black and the target site CR is in magenta. Base-pairing is highlighted with dashes. Int-mediated cleavage and strand exchange at the site indicated by black (products) or orange (substrates) arrows leads to the final integration products (T_L-IR_L, 59 nt, and IR_R-T_R, 64 nt, right) or goes backward to substrates (IR_L- IR_R, 56 nt, and T_L-T_R, 67 nt, left). Black and orange stars highlight the ^{32}P radiolabeled DNA strands.

(B) Denaturing sequencing PAGE gel showing resolution of HJ1-C intermediates in direction to substrates. Products are highlighted with a star (orange) on the gel.

(C) Denaturing sequencing PAGE gel showing resolution of HJ-C intermediates in direction to products. In this case, only the initially radiolabeled oligos can be observed and no product bands are seen at the expected size. - = no protein was added to the reaction (negative control), + = protein was added.

2.2.3 Order of strand exchange and the directionality of HJ resolution

Until now, I had analyzed the resolution of one possible HJ intermediate of the Tn*1549* integration reaction, which was designed based on the assumption that the first cleavage and strand exchange reactions occurred at the IR_R and T_L sites. However, as discussed above, a different cleavage order remains possible (section 2.1.2.2), which would lead to a different HJ intermediate. Such difference in HJ sequence could also influence the directionality of HJ resolution. Thus, I designed a second construct, called HJ_L, which resembles the second possible HJ intermediate of the integration reaction. This design is based on the assumption that the first cleavage and strand exchange reactions occur at the

IR_L of the CI and T_R of the target site. HJ_L was also used for crystallization experiments. For more details on its design, see section 2.1.2.2.

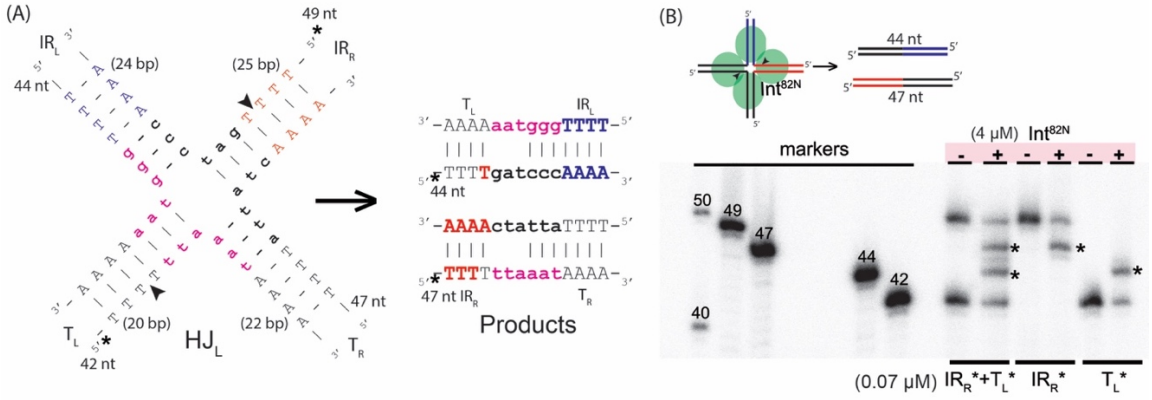

Figure 2-51: Characterization of Int's HJ resolution activity on HJ_L in direction to products.

(A) Scheme of HJ_L and its products (IR_L in blue, IR_R in red, and T-sites in black). The exact sequence is only shown for the central regions. The CI crossover region is in black and the target site CR is in magenta. Base-pairing is highlighted with dashes. Int-mediated cleavage and strand exchange at the site indicated by black arrows leads to the final integration products (T_L-IR_L, 44 nt, and IR_R-T_R, 47 nt). Both products contain a 3 bp mismatch at the CR. Black star highlights the ^{32}P radiolabeled DNA strands.

(B) Denaturing sequencing PAGE gel showing resolution of HJ_L intermediates in direction to products. Products are highlighted with a star on the gel. - = no protein was added to the reaction (negative control), + = protein was added.

Incubation of HJ_L with Int^{82N} showed efficient resolution towards integration products (Figure 2-51B) and minimal resolution in direction to substrates (Figure 2-52B). Thus, HJ_L was preferentially resolved towards products, despite that in this case both reactions lead to dsDNA with the same 3 bp mismatch in the CR. These results are in contrast with the HJ1-A resolution assay, which does not show any preference in the reaction outcomes.

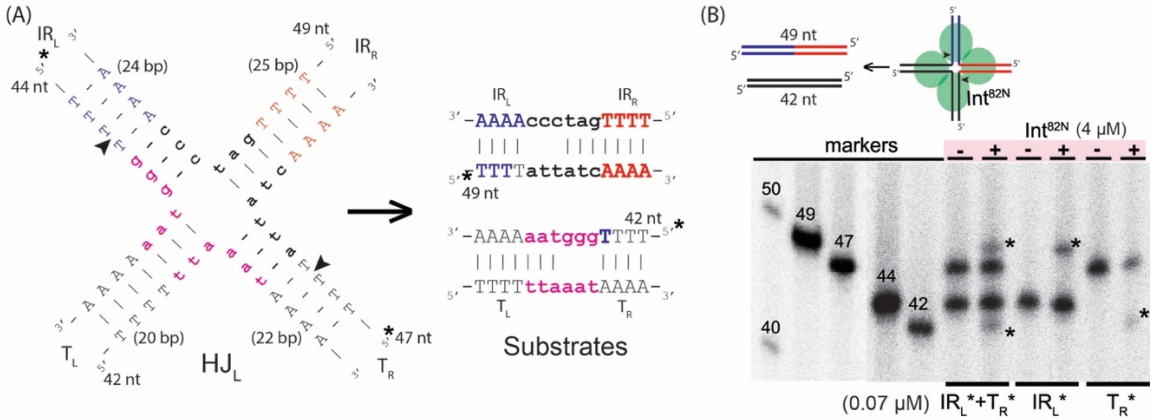

Figure 2-52: Characterization of Int's HJ resolution activity on HJ_L in direction to substrates.

(A) Scheme of HJL and its products (IRL in blue, IRR in red, and T-sites in black). The exact sequence is only shown for the central regions. The CI crossover region is in black and the target site CR is in magenta. Base-pairing is highlighted with dashes. Int-mediated cleavage and strand exchange at the site indicated by black arrows goes backward to substrates (IRL-IRR, 49 nt, and TR-TL, 42 nt). Both products contain a 3 bp mismatch at the CR. Black star highlights the ^{32}P radiolabeled DNA strands.

(B) Denaturing sequencing PAGE gel showing resolution of HJL intermediates in direction to substrates. Products are highlighted with a star on the gel. - = no protein was added to the reaction (negative control), + = protein was added.

Next, to support this finding I created a HJ_L-C variant, by incorporating the mismatched CR design of HJ1-C into HJ_L (see section 2.2.2). In this case, in contrast to the HJ_L resolution assay, sequencing gels showed resolution in direction to substrates (Figure 2-54B), but not towards products (Figure 2-53B). This result is similar to the observations with HJ1-B and HJ1-C resolution, indicating that the *in vitro* HJ resolution reactions with Int^{82N} might be biased by base-pairing in the resulting products. As for HJ1-C, resolution of HJ_L-C to products leads to 6 mismatches, whereas resolution to substrates creates 3 bps and only 3 mismatches. Alternatively, the lack of base-pairing near the cleavage sites in specific HJ_L-C stems (Figure 2-54A), like for the HJ-B and HJ-C constructs, might help the integrase to open DNA and perform strand exchange more easily at these sites, thereby artificially biasing the reactions towards resolution to substrates.

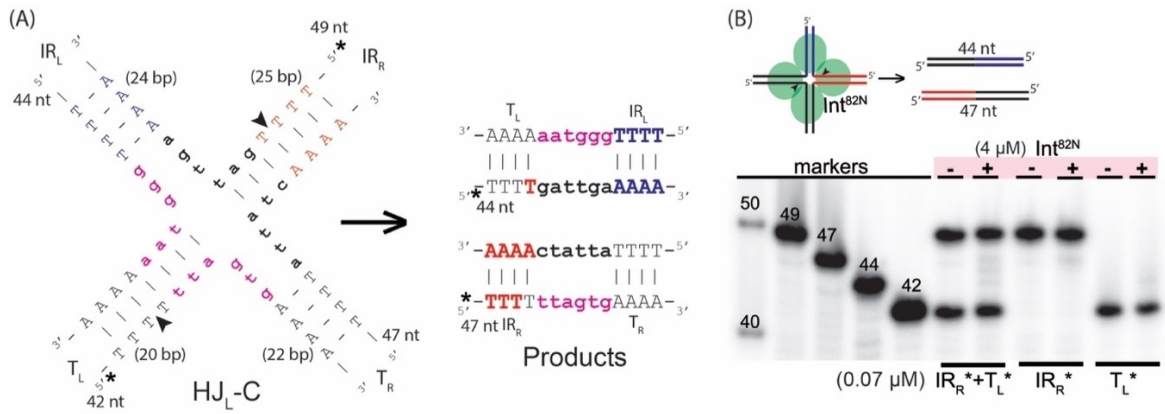

Figure 2-53: Characterization of Int's HJ resolution activity on HJL-C in direction to products.

(A) Scheme of HJL-C and its products (IRL in blue, IRR in red, and T-sites in black). The exact sequence is only shown for the central regions. The CI crossover region is in black and the target site CR is in magenta. Base-pairing is highlighted with dashes. Int-mediated cleavage and strand exchange at the site indicated by black arrows leads to the final integration products (TL-IRL, 44 nt, and IRR-TR, 47 nt). Both outcomes would contain a 6 bp mismatch at the CR. Black star highlights the ^{32}P radiolabeled DNA strands.

(B) Denaturing sequencing PAGE gel showing resolution of HJL-C intermediates in direction to products. In this case, only the initially radiolabeled oligos can be observed and no product bands are seen at the expected size. - = no protein was added to the reaction (negative control), + = protein was added.

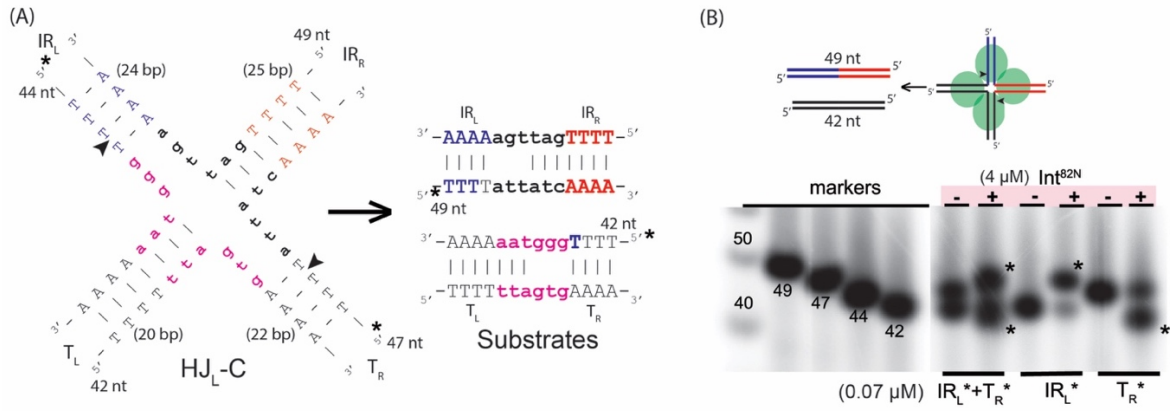

Figure 2-54: Characterization of Int's HJ resolution activity on HJ$_L$-C in direction to substrates.

(A) Scheme of HJ$_L$-C and its products (IR$_L$ in blue, IR$_R$ in red, and T-sites in black). The exact sequence is only shown for the central regions. The CI crossover region is in black and the target site CR is in magenta. Base-pairing is highlighted with dashes. Int-mediated cleavage and strand exchange at the site indicated by black arrows goes backward to substrates (IR$_L$-IR$_R$, 49 nt, and T$_R$-T$_L$, 42 nt). Both products contain a 3 bp mismatch at the CR. Black star highlights the ^{32}P radiolabeled DNA strands.

(B) Denaturing sequencing PAGE gel showing resolution of HJ$_L$-C intermediates in direction to substrates. Products are highlighted with a star on the gel. - = no protein was added to the reaction (negative control), + = protein was added.

2.2.4 The role of Int's AB domain and arm DNA binding sites in HJ$_{short}$ resolution directionality

Testing Int's activity on various HJ constructs indicated that the direction of HJ resolution might be biased due to the CR design. Therefore, I decided to use the full-length protein (Int^{FL}). Int^{FL} contains the AB domain that is missing in the Int^{82N} variant (Figure 2-1) and is expected to interact with arm DNA sequences inside the transposon ends (Figure 2-55). In the λ-bacteriophage system, interaction of λ-Int's AB domain with its arm sites has been shown to play a role in regulating the directionality of λ-phage integration (Landy, 2015). Thus, probing of Tn*1549* Int's AB domain with its specific arm sites in HJ resolution may help to guide resolution towards integration products.

First, DNA binding studies were performed to confirm the interaction of Int^{FL} with HJ and arm DNA molecules. For that, the Int^{FL} constructs were overexpressed and purified as described in section 4.2.9.2. As HJ DNA, I used the HJ1 construct (see section 2.1.2.2). The arm DNA sequences were predicted using a bioinformatic approach (by postdoc Georgy Smyshlyaev from the Barabas Lab, Figure 2-55). The left arm (LA) site is located inside the left transposon end, and it is composed of two direct repeats of 11 bp length each. The right arm (RA) site is found inside the right transposon end and is composed of 2 imperfect direct repeats of 11 bp length, with a single nucleotide insertion between the

repeats (Figure 2-56). The LA and RA repeats are similar but differ in the last three base pairs.

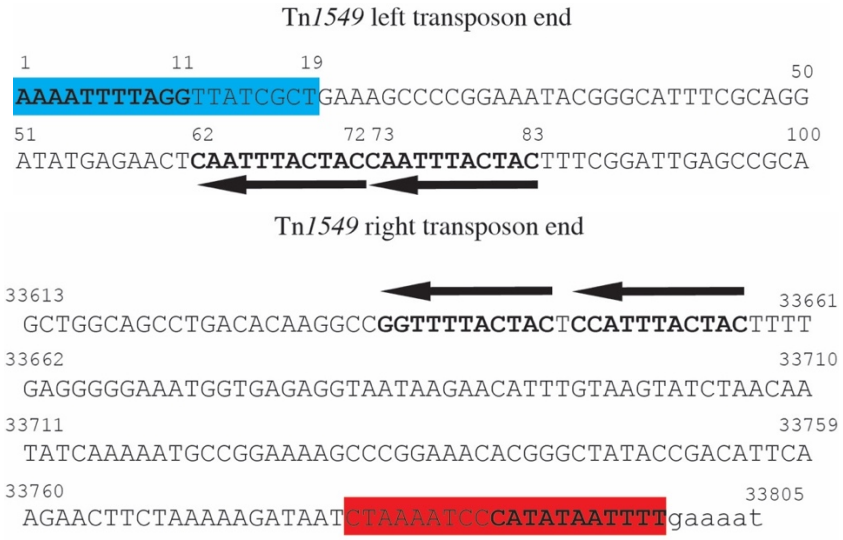

Figure 2-55: The Tn1549 end sequences. Left end sequence (top): the core IR_L binding site is in bold (1-11 bp), and the sequence segment used for experimental work is highlighted in cyan. Internal arm binding sites (bold) are composed of two direct repeats (11 bp each) and marked with arrows. Right end sequence (bottom): the core IR_R binding site is in bold (33789-33799 bp), and the sequence used to design substrates for experimental work is highlighted in red. Internal arm binding sites (bold) are composed of two imperfect direct repeats (11 bp each) separated by one nucleotide and highlighted with arrows. The last six bases (in lowercase) represent a crossover region derived from the genomic flank at a specific donor site.

The initial characterization of various Int^{FL}R225K-DNA complexes by EMSA showed the following: I observed stable complex formation at 1:2 molar ratios of HJ1 DNA with Int^{FL}R225K (Figure 2-57A, lane 2). Higher protein molar ratios led to precipitation during incubation, and neither DNA alone nor in the complex could be observed in the gel (Figure 2-57A, lane 4). Adding an excess of arm LA or RA DNA substrate to the Int^{FL}R225K-HJ1 complex resulted in a characteristic supershift (Figure 2-57A, lanes 7 and 8) that differed from the shift observed, when RA- or LA-DNA was mixed alone with Int^{FL}R225K (Figure 2-57A, lanes 5 and 6). Also, Int^{FL}R225K-CI5-arm DNA complex formation was tested to confirm AB-arm DNA interaction. Indeed, a characteristic supershift could be observed on native PAGE gels when Int^{FL}R225K-CI5 complex was mixed with LA or RA DNA together (Figure 2-57B, lanes 8 - 10), while Int^{FL}R225K with CI5 DNA alone precipitated, probably during incubation, and could not be observed in the gel (Figure 2-57B, lane 7).

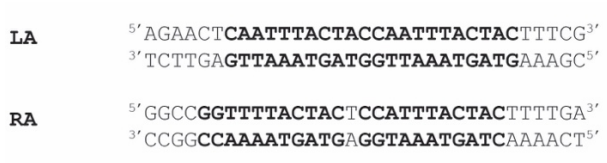

Figure 2-56: Arm DNA substrates used for binding and HJ-resolution experiments with Int^{FL}R225K and Int^{FL}, respectively. LA - left arm (33 bp); RA - right arm (33 bp).

Another way to analyze Int^{FL}R225K-HJ-arm DNA complexes and confirm the previously observed complex formation by EMSA, was by performing analytical size exclusion chromatography (see section 4.3.8). Mixing Int^{FL}R225K with HJ1 DNA resulted in substantial precipitation. Nevertheless, analysis of the remaining sample showed a small shift in the elution volume to larger oligomeric assemblies compared to the HJ1 DNA control, indicating protein binding to the HJ DNA (Figure 2-58). The Int^{FL}R225K-HJ1-arm DNA complexes precipitated to a lesser extent and showed a more pronounced shift to larger oligomeric assemblies compared to the Int^{FL}R225K-HJ1 DNA sample (Figure 2-58). Thus, I decided Int^{FL}-HJ-arm DNA complexes could be formed and used for HJ-resolution analysis.

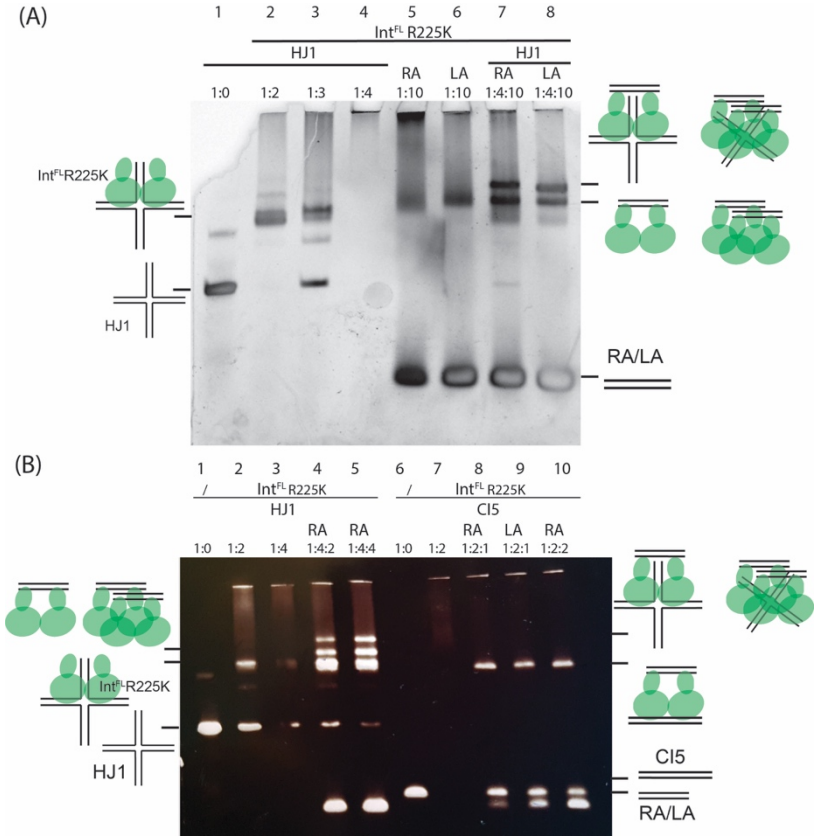

Figure 2-57: Characterization of various Int^{FL}R225K-DNA complexes by EMSA. (A) Native PAGE gel showing Int^{FL}R225K-HJ1 complexes mixed at different (1:0, 1:2, 1:3, 1:4) molar ratios of DNA (2 µM, constant) to protein; Int^{FL}R225K-arm DNA and Int^{FL}R225K-HJ1-arm DNA at 1:10 and 1:4:10 molar ratios

with arm DNA in excess. (B) Native PAGE gel showing Int^{FL}R225K-HJ1 complexes mixed at different (1:0, 1: 1:4) molar ratios of DNA to protein and with RA DNA in 1:4:2 or 1:4:4 molar ratio. The second half of the native gel shows complexes of Int^{FL}R225K-CI5 or Int^{FL}R225K-CI5-arm DNA at different molar ratios.

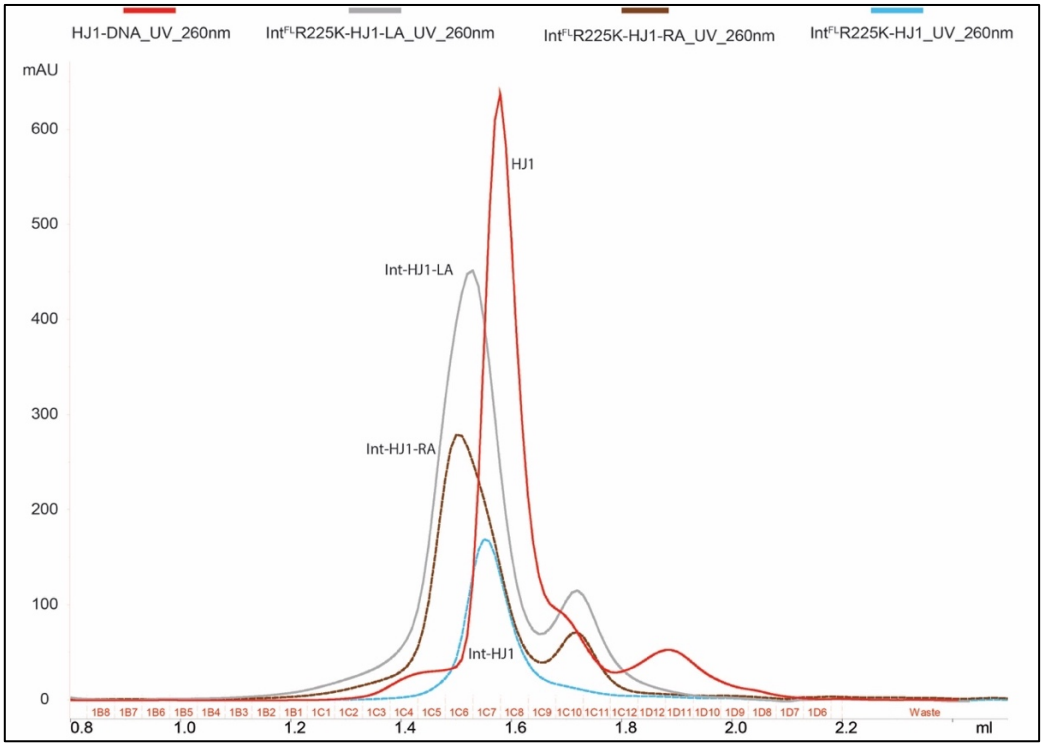

Figure 2-58: Overlay of chromatograms from analytical SEC runs with various Int^{FL}R225K-HJ complexes and controls. Comparison of HJ1 DNA alone (red, 1.58 ml) with complexes Int^{FL}R225K-HJ1 (cyan, 1.55 ml), Int^{FL}R225K-HJ1-LA (grey, 1.53 ml), and Int^{FL}R225K-HJ1-RA (brown, 1.50 ml). Control with arm DNA alone is not shown but revealed an elution peak at 1.72 ml for either LA or RA DNA. For all runs, a Superose 6 Increase column was used. For clarity, only chromatograms with UV absorbance at 260 nm are shown.

After confirming Int^{FL}-HJ-arm DNA complex formation and interaction of Int^{FL} with its predicted arm sites, I tested the role of the AB domain-arm sites interaction in HJ resolution assays by using the HJ1-B and HJ1-C constructs (Figure 2-48), as it was done previously for Int^{82N} (see section 2.2.2). Both HJ constructs were preferred over HJ1-A because they resemble closer the real HJ integration intermediate.

HJ resolution assays with either HJ1-B or HJ1-C constructs showed resolution back to substrates by using wild type Int^{FL} alone or in combination with arm sites (Figure 2-59B) and Figure 2-60B). On the other hand, resolution to products could not be observed with none of the HJ constructs (Figure 2-61B and Figure 2-62B). This data indicates that Int^{FL} with arm DNA has similar activity in the assay as Int^{82N} (see section 2.2.2) and arm DNA binding to the Int AB domain alone is not sufficient to promote preferential resolution of HJ1 to the native integration products.

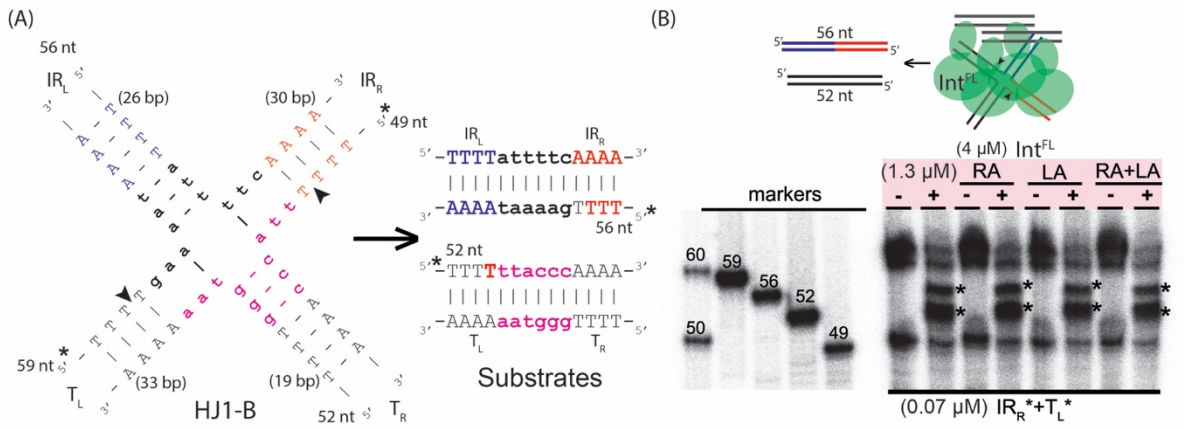

Figure 2-59: Characterization of Int^{FL}'s HJ resolution activity on HJ1-B in direction to substrates.

(A) Scheme of HJ1-B resolution by Int^{FL} (IR$_L$ in blue, IR$_R$ in red, and T-sites in black). The exact sequence is only shown for the central regions. The CI crossover region is in black and the target site CR is in magenta. Base-pairing is highlighted with dashes. Int-mediated cleavage and strand exchange at the site indicated by black arrows goes backward to substrates (IR$_L$- IR$_R$, 56 nt, and T$_L$-T$_R$, 52 nt, left). Both products contain a fully base-paired CR. Black stars - the ^{32}P radiolabeled DNA strands.

(B) Denaturing sequencing PAGE gel showing resolution of HJ1-B intermediates in direction to substrates by Int^{FL} with unlabeled arm DNA. Products are highlighted with a star on the gel. - = no Int^{FL} protein was added to the reaction (negative control), + = Int^{FL} protein was added.

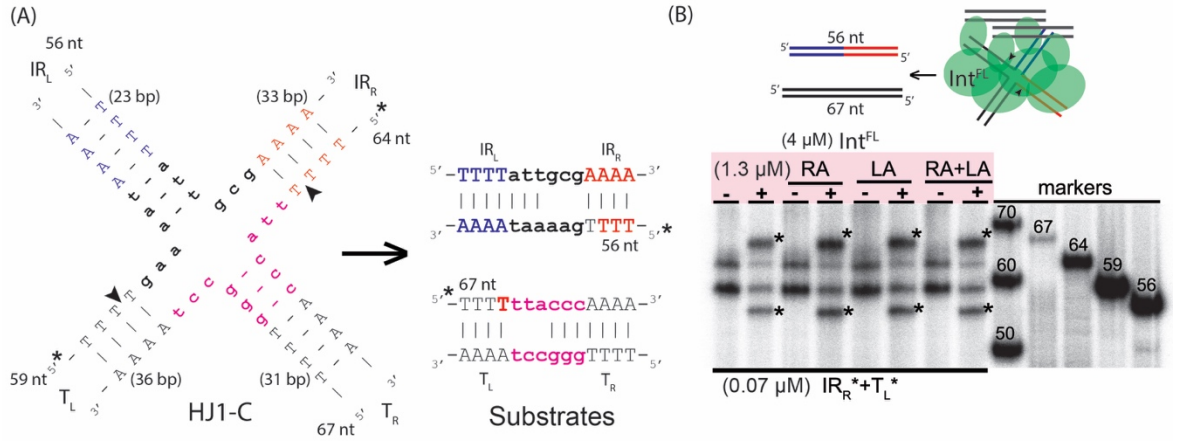

Figure 2-60: Characterization of Int^{FL}'s HJ resolution activity on HJ1-C in direction to substrates.

(A) Scheme of HJ1-C resolution by Int^{FL} (IR$_L$ in blue, IR$_R$ in red, and T-sites in black). The exact sequence is only shown for the central regions. The CI crossover region is in black and the target site CR is in magenta. Base-pairing is highlighted with dashes. Int-mediated cleavage and strand exchange at the site indicated by black arrows goes backward to substrates (IR$_L$- IR$_R$, 56 nt, and T$_L$-T$_R$, 67 nt, left). Both products would contain a 3 bp mismatch at the CR. Black stars - the ^{32}P radiolabeled DNA strands.

(B) Denaturing sequencing PAGE gel showing resolution of HJ1-C intermediates in direction to substrates by Int^{FL} with unlabeled arm DNA. Products are highlighted with a star on the gel. - = no Int^{FL} protein was added to the reaction (negative control), + = Int^{FL} protein was added.

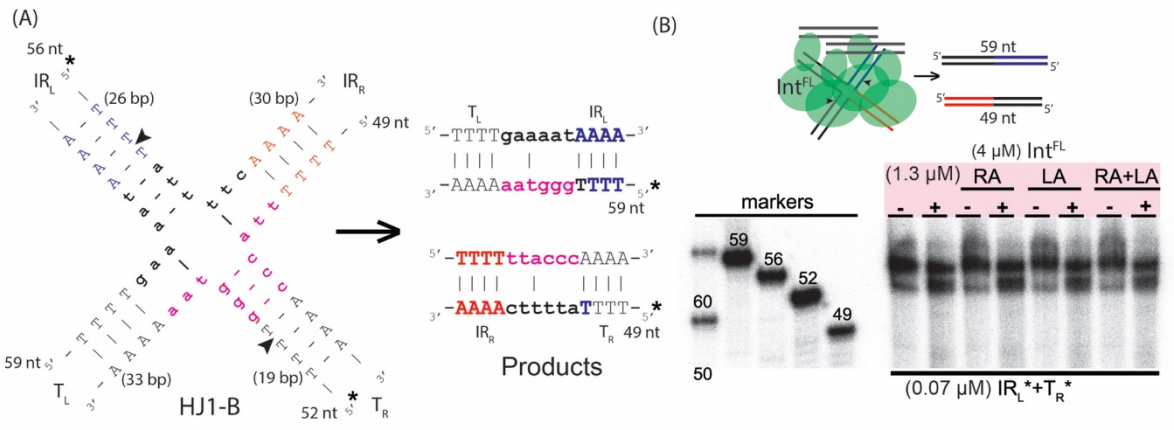

Figure 2-61: Characterization of Int^{FL}'s HJ resolution activity on HJ1-B in direction to products.

(A) Scheme of HJ1-B resolution by Int^{FL} (IR$_L$ in blue, IR$_R$ in red, and T-sites in black). The exact sequence is only shown for the central regions. The CI crossover region is in black and the target site CR is in magenta. Base-pairing is highlighted with dashes. Int-mediated cleavage and strand exchange at the site indicated by black arrows should lead to the final integration products (T$_L$-IR$_L$, 59 nt, and IR$_R$-T$_R$, 49 nt, right). Both products would contain a 5 bp mismatch at the CR. Black stars - ^{32}P radiolabeled DNA strands.

(B) Denaturing sequencing PAGE gel showing resolution of HJ-B intermediates in direction to products by Int^{FL} with unlabeled arm DNA. In this case, only the initially radiolabeled oligos can be observed and no product bands are seen at the expected size. - = no Int^{FL} protein was added to the reaction (negative control), + = Int^{FL} protein was added.

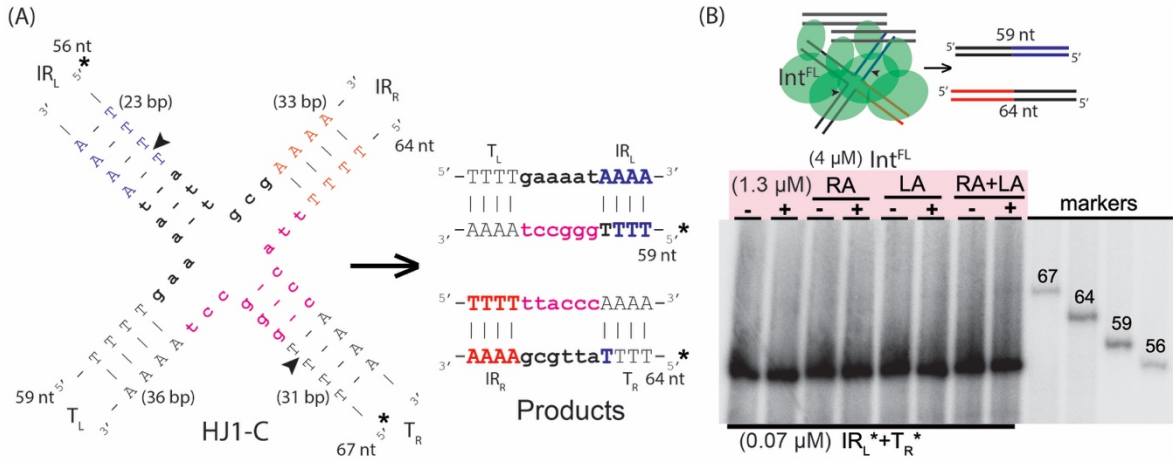

Figure 2-62: Characterization of Int^{FL}'s HJ resolution activity on HJ1-C in direction to products.

(A) Scheme of HJ1-C resolution by Int^{FL} (IR$_L$ in blue, IR$_R$ in red, and T-sites in black). The exact sequence is only shown for the central regions. The CI crossover region is in black and the target site CR is in magenta. Base-pairing is highlighted with dashes. Int-mediated cleavage and strand exchange at the site indicated by black arrows should lead to the final integration products (T$_L$-IR$_L$, 59 nt, and IR$_R$-T$_R$, 64 nt, right). Both products would contain a 6 bp mismatch at the CR. Black stars - the ^{32}P radiolabeled DNA strands.

(B) Denaturing sequencing PAGE gel showing resolution of HJ-C intermediates in direction to products by Int^{FL} with unlabeled arm DNA. In this case, only the initially radiolabeled oligos can be observed and no product bands are seen at the expected size. - = no Int^{FL} protein was added to the reaction (negative control), + = Int^{FL} protein was added.

2.2.5 Resolution of a long HJ intermediate

The initial HJ resolution assay was helpful to answer the question if Int could resolve a HJ intermediate, similar to other members of the tyrosine recombinase family. Nevertheless, this assay did not clarify the requirements and regulation of specific reaction steps. Previous results indicated that the direction of HJ resolution might be biased by the specific CR designs used. Another shortcoming of the assay was that it lacked the native regulatory context of transposon integration. Although the activity with full length Int and regulatory arm DNA was tested, the arm site was supplied in a separate DNA molecule as the core recombination sites. Moreover, Xis and its DNA sites were not present in the reactions. However, all such elements were implicated in regulating the directionality of λ-phage integration in earlier studies (Landy, 2015).

Thus, I next generated HJs with longer stems, called HJ_{long} (Figure 2-63). The HJ_{long} central region is equivalent to the HJ1-A design and contains the IR_L, IR_R, and target site sequences. Apart from that, this substrate contains the transposon arm sites and putative Xis binding sites.

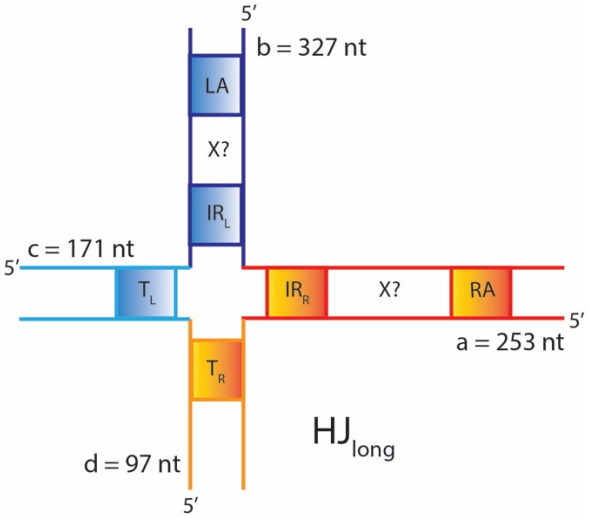

Figure 2-63: Schematics of the HJ_{long} DNA used for resolution experiments. The HJ DNA imitates the intermediate of the CTn integration reaction after the first round of cleavage and strand exchange occurs at IR_R and T_L. The central region bases were mutated to obtain a stable and fully base-paired HJ. HJ stems had different lengths allowing to follow changes in the sequence length on a denaturing Urea-TBE gel. Oligonucleotide lengths (in nt) are next to the 5'ends. The construct contains the core sequences from the CI (IR_L and IR_R) and target site (T_L and T_R). Moreover, it includes the arm sites LA and RA. Putative Xis binding sites are highlighted with an X? in the CI stems.

HJ_{long} formation was performed, as explained in sections 4.3.1.2 and 4.3.2. Afterwards, I performed HJ resolution assays with Int^{82N} and various Int^{FL} constructs. The assay is described in section 4.3.7. The Int^{FL} variants included the wild type protein (Int^{FL}) and two

mutants. Int^{FL}Y380F has a mutation in the catalytic tyrosine nucleophile to phenylalanine and Int^{FL}Y379F has the Y to F mutation in the neighboring tyrosine. It has been shown that both single tyrosine mutants are able to perform DNA cleavage, but Int^{FL}Y380F is inactive in strand exchange reactions *in vitro* (Rubio-Cosials *et al.*, 2018). Therefore, testing HJ resolution ability of both mutants may help to characterize the tyrosine's specific roles in the recombination reaction further. Moreover, I tested HJ resolution with these Int^{FL} variants with or without Xis to investigate its possible role in integration. As controls, I performed the reaction with the catalytic inactive mutant Int^{FL}R225K and Xis alone.

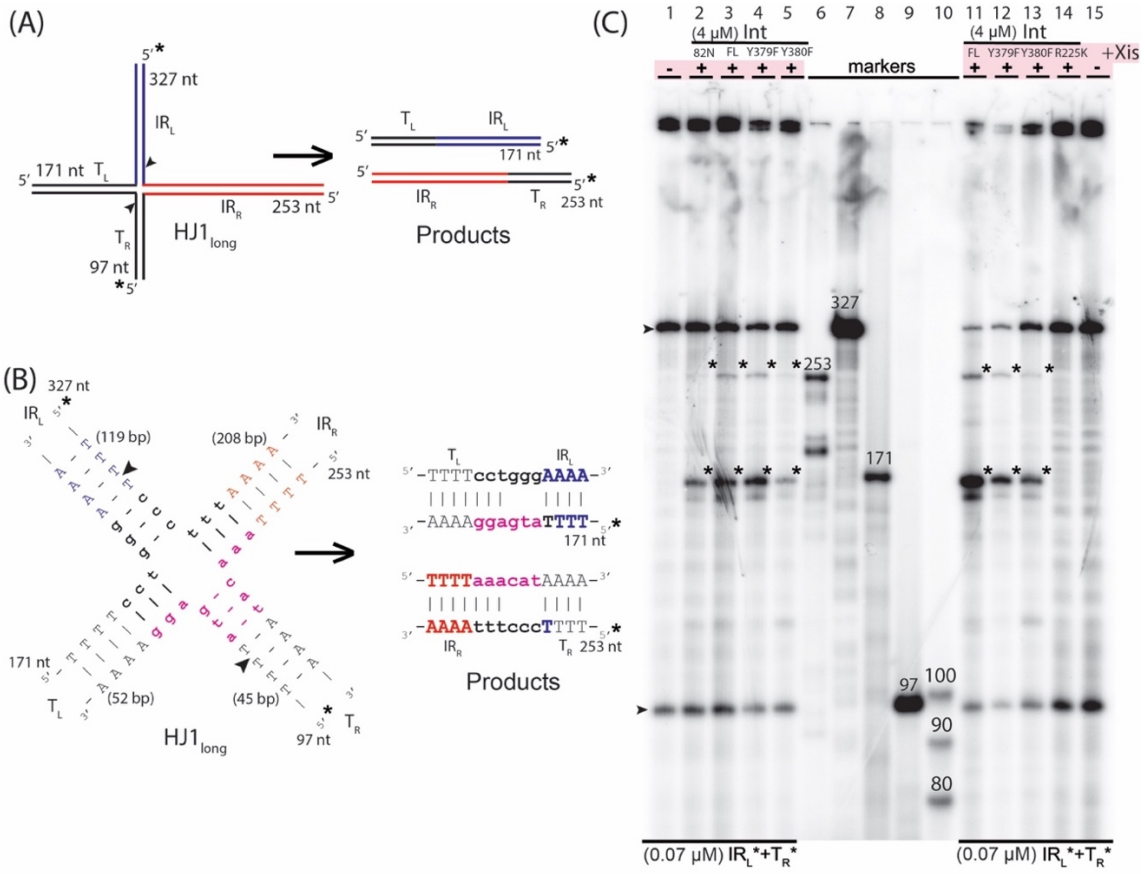

Figure 2-64: HJ$_{long}$ resolution towards integration products. (A) and (B) show the schematics of the reaction.

(A) Scheme of HJ$_{long}$ and its products (IR$_L$ in blue, IR$_R$ in red, and T-sites in black). After HJ resolution, IR$_L$ is connected to T$_L$ and IR$_R$ to T$_R$. Black star highlights the ^{32}P radiolabeled DNA strands.

(B) The exact sequence is only shown for the central regions. The CI crossover region is in black and the target site CR is in magenta. Base-pairing is highlighted with dashes. Oligo length in nt is marked at the 5'-end of each sequence. Stem lengths are given in bp in parenthesis. Int-mediated cleavage and strand exchange at the site indicated by black arrows leads to the final integration products (IR$_L$-T$_L$, 171 nt, and IR$_R$-T$_R$, 253 nt). Both products contain a 3 bp mismatch at the CR. Black star - ^{32}P radiolabeled strands.

(C) Denaturing sequencing PAGE gel showing HJ$_{long}$ intermediates resolution with ^{32}P radiolabeled IR$_L$ and T$_R$ stems by various Int constructs with or without Xis. The position of the labeled HJ strands is marked with arrows next to the gel. Product bands run at the expected size of 171 nt for IR$_L$-T$_L$ and 253 nt for IR$_R$-T$_R$, highlighted with black stars on the gel. - = no Int protein was added to the reaction (negative control), + = Int protein was added.

The results showed that both Int^{82N} and Int^{FL} constructs could resolve the long HJ towards integration products (Figure 2-64C, lanes 2 and 3), while Int^{FL} was substantially more efficient. The mutant constructs, Int^{FL}Y379F, and Int^{FL}Y380F also showed some activity (Figure 2-64C, lanes 4 and 5), whereas the Int^{FL}R225K catalytic mutant was entirely inactive. The addition of Xis appeared to further stimulate HJ resolution towards integration products for all three Int^{FL} constructs (Figure 2-64C, lanes 11, 12, and 13). Remarkably, none of the proteins resolved the HJ_{long} towards substrates (Figure 2-65). These results indicate that the presence of the arm DNA sequence, in the correct architectural context, helps to drive recombination towards transposon integration, but further experiments are needed to confirm this conclusion.

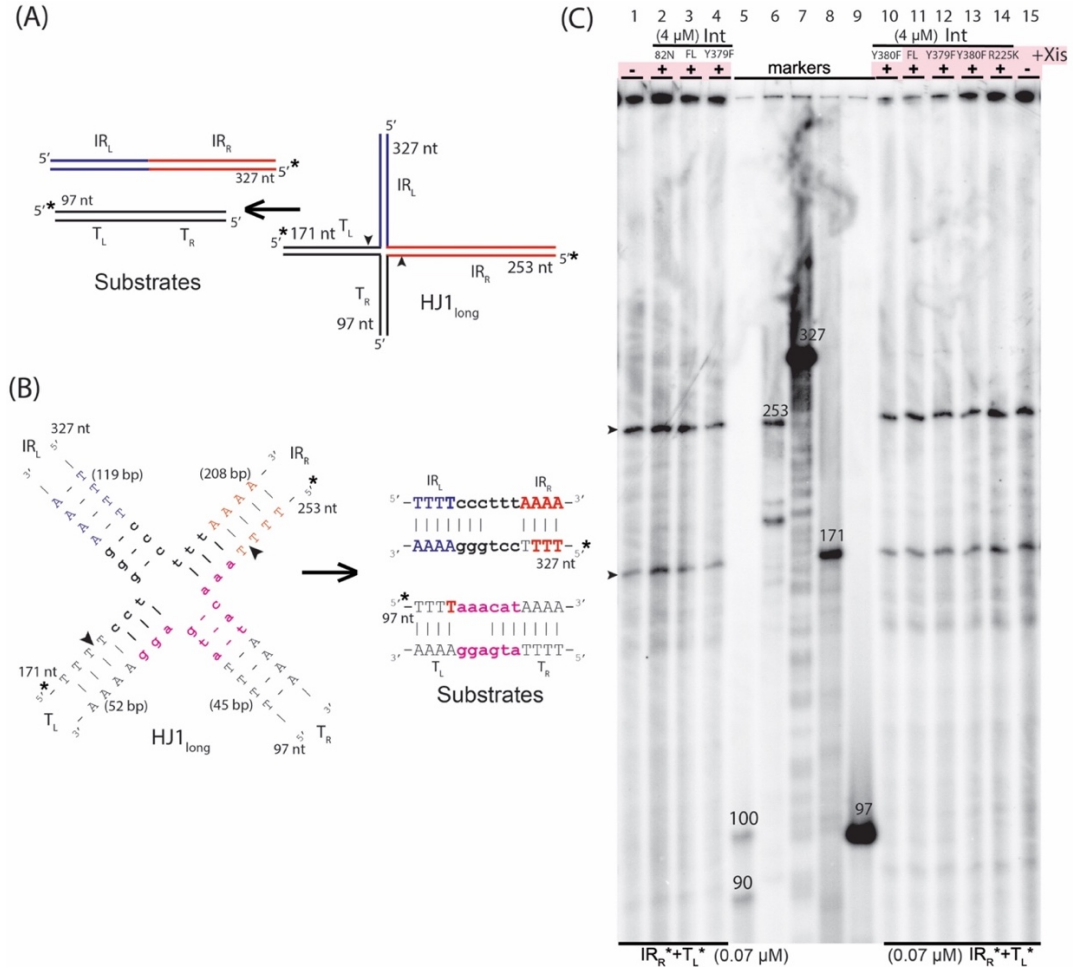

Figure 2-65: HJ_{long} resolution towards integration substrates. (A) and (B) show the schematics of the reaction.

(A) Scheme of HJ_{long} and its products (IR_L in blue, IR_R in red, and T-sites in black). After HJ resolution, IR_L is connected with IR_R and T_L with T_R to reform the CI and target sequences. Black star highlights the ^{32}P radiolabeled DNA strands.

(B) The exact sequence is only shown for the central regions. The CI crossover region is in black and the target site CR is in magenta. Base-pairing is highlighted with dashes. Oligo length in nt is marked at the 5'-end of each sequence. Stem lengths are given in bp in parenthesis. Int-mediated cleavage and strand exchange at the site indicated by black arrows should go backward to substrates: CI (IR_L-IR_R, 327 nt) and the target site (T_L-T_R, 97 nt). Both products contain a 3 bp mismatch at the CR. Black star - ^{32}P radiolabeled strands.

(C) Denaturing sequencing PAGE gel showing HJ_{long} intermediates resolution with ^{32}P radiolabeled IR_R and T_L stems by various Int constructs with or without Xis. The position of the labeled HJ strands is marked with arrows next to the gel. In this case, no product bands can be observed at the expected size of 327 nt for IR_L-IR_R and 97 nt for T_L-T_R, only the initially radiolabeled oligos. - = no Int protein was added to the reaction (negative control), + = Int protein was added.

2.2.6 Discussion

2.2.6.1 The role of the crossover region in the directionality of HJ resolution

To study specific steps of the Tn*1549*'s integration pathway, I have set up a HJ resolution assay that allowed me to reconstitute the second cleavage and strand exchange reaction steps. This assay mimics the recombination of the Tn*1549* circular intermediate (IR_L-IR_R) with an integration site (T_L-T_R) in the bacterial chromosome, which will lead to the creation of new DNA sites (T_L-IR_L and IR_R-T_R) and transposon insertion (see Figure 2-43).

Initial experiments helped me to characterize Int's ability to resolve short HJ substrates. Indeed, Int can resolve fully base-paired, short HJ DNA. However, Int resolves these DNA in both directions and not only towards products as expected. This may be because regulatory elements were lacking and/or because the crossover region was mutated, which may influence the assay results.

In the native reaction, the CR is already opened in the pre-cleavage state, as shown in the Int-CI structure (Rubio-Cosials *et al.*, 2018). This facilitates the attack of the free 5'-OH group on the 3'-phosphotyrosyl bond of the recombination partner DNA for strand exchange. Moreover, this unwinding allows Int to accommodate substrates with different CRs, which in turn leads to mismatched HJ intermediates. Therefore, I decided to investigate if non-homology at the CR may affect the directionality of HJ resolution.

For that, I tested HJs that had mismatched CRs at the IR_R- and T_L-stems (Figure 2-48). Experiments with these DNA variants surprisingly showed that Int prefers to resolve these HJs back to integration substrates (CI and target site). A potential explanation for this observation could be that Int prefers to cleave and recombine at the sites, where the mismatches are located, rather than resolving the HJ at the fully base-paired sites. For the latter, additional base-pair opening would be necessary, which would require additional energy investment and lead to dsDNA products with 6 bp long mismatches. Interestingly, resolution assays with Int^{FL} and the arm sites (in separate DNA molecules) showed the same results as with Int^{82N}, indicating that the sole presence of the arm DNA sequence does not change the reaction outcome.

Another possibility for the observed unexpected directionality in the above HJ resolution reactions could be a bias by our initial assumption for the order of the strand exchange reactions in Tn*1549* integration. Thus, I created a new set of HJ constructs to probe this possibility. First resolution assays with fully base paired HJ_L showed preferred recombination towards products over substrates. Nevertheless, by having a closer look at the junction's core region, I noticed that the IR_L stem contained a CR of three GC base-pairs, while the other arms were AT-rich. Thus, HJ_L resolution may be biased to products because Int would need to melt more GC base-pairs for the resolution towards substrates. However, the CI sequence information for these experiments was obtained from *in vivo* assays (Lambertsen *et al.*, 2018), indicating that Int can effectively recombine GC-rich substrates, at least in bacterial cells. Therefore, HJ resolution with these specific settings *in vitro* may be missing essential components that help overcome the energy barrier for melting such sequences during the reaction, such as DNA supercoiling or additional transposon- or host-derived factors.

In summary, Int can resolve short HJs with four base-paired stems in both directions, creating dsDNA products with a 3 bp mismatch. Asymmetric HJs with two based-paired and two mismatched segments are preferentially cleaved at the non-base paired sites. This leads to dsDNA products that are fully base-paired in the CR or contain a 3 bp long heterology, but never go in the direction where a 6 bp mismatch would be made. These results show that *in vitro*, Int alone prefers to recombine in the direction where less energy is required. The integration reaction therefore must be highly regulated and supported by accessory elements to drive its outcome effectively towards products *in vivo*. Most likely, Int can only overcome the energy barrier of correct strand exchange together with all its regulatory components and factors so as to drive the reaction towards transposon insertion.

2.2.6.2 Comparison of short and long HJ resolution reactions

As summarized in the previous chapter, HJ resolution experiments with a short HJ substrate led to the conclusion that Int alone cannot intrinsically perform its recombination reaction in a specific direction. Thus, I decided to create a longer HJ that contained the regulatory arm DNA sites and putative Xis binding sequences. Furthermore, I used the full-length Int protein including the AB domain.

Initial results showed that HJ_{long} could be resolved towards products and not back to substrates, reflecting the expected directionality of Tn*1549*'s integration reaction. Another interesting result was that the HJ_{long} construct could be resolved in the direction where the

HJ stems had a GC-rich crossover region. This observation is in strong contrast to the results with the short HJs, which could not be resolved well, when the CR was GC-rich. Therefore, this result confirms that Int can resolve full HJs with GC-rich CRs and that additional regulatory elements may facilitate this process.

The wild type protein Int^{FL} showed higher efficiency than the truncated Int^{82N} variant and the Int^{FL} mutants (Int^{FL}Y379F and Int^{FL}Y380F). Whereby, Int^{FL}Y379F showed more activity than Int^{FL}Y380F, in concordance with previous observations from previous *in vitro* and *in vivo* experiments. These showed that Y380 is essential for strand exchange reactions and promotes recombination natively in bacteria. However, when Y380 is mutated, Y379 can take over its role in DNA cleavage *in vitro* (Rubio-Cosials *et al.*, 2018).

On the other hand, the addition of the transposon-encoded accessory factor Xis to all three Int^{FL} constructs led to increased product formation. This outcome was unexpected because previous results from bacteriophage λ have shown that λ-Xis inhibits integration *in vivo* (Landy, 2015). In turn, Xis from Tn*916* showed unambiguous results in *in vivo* excision assays. It seems to induce (Connolly, Iwahara and Clubb, 2002) or inhibit excision (Hinerfeld and Churchward, 2001), depending whether it binds at the left- or right-arm binding sites, respectively. Furthermore, Tn*916* Xis does not seem to impede integration *in vivo* (Marra *et al.*, 1999). *In vivo* studies for Tn*1549* Xis showed that this accessory protein is necessary for proper transposition (Lambertsen *et al.*, 2018). My results on HJ_{long} resolution seem to support the role of Xis in influencing the reaction and driving it towards integration. Nevertheless, more experiments are needed to validate and expand these initial results.

In summary, although HJ_{long} resolution needs further investigation, initial results indicate that the arm DNA sites, and the accessory DNA bending protein Xis, help drive recombination towards transposon integration.

2.2.6.3 HJ resolution by Int and other tyrosine recombinases

The recombination of two DNA sites by tyrosine recombinases has been extensively studied in diverse systems (Grindley, Whiteson and Rice, 2006), such as λ-Integrase (Landy, 2015), Cre recombinase (Duyne, 2015) and Xer recombinases. These studies have shown that all site-specific tyrosine recombinases share a common feature: the strict requirement for homology at the crossover region (also called overlap) between the recombination sites. While λ-Int and Cre need the whole overlap region to be homologous in both DNA sequences, conjugative transposon integrases are different. They can resolve

HJs leading to heterologous dsDNA products at the crossover region, as long as there is a microhomology region that is needed for the first strand exchange reaction. The requirement for CR homology in conjugative transposition has been studied for the Bacteroides conjugative transposon CTnDOT (Wood and Gardner, 2015). Here, recombination strictly requires a conserved GC dinucleotide inside the CR. Previous results for Tn*1549* Int showed that it needs only one single nucleotide to be homologous in the exchanged DNA parts, usually a thymine base that is situated one nucleotide inside the transposon end (Rubio-Cosials *et al.*, 2018).

Interestingly, in CTnDOT, the order of strand exchange seems to be dictated by base pairing of the GC dinucleotide located on the top strands, where the first cleavage happens. Previous studies suggest that IntDOT performs the first round of strand exchange in a homology-dependent manner at the top strands where the conserved GC dinucleotide is located (Wood and Gardner, 2015). HJ resolution experiments with IntDOT have shown that short HJ constructs can only be resolved back to substrates, when product formation would lead to mismatched CRs, probably due to the GC dinucleotide base-pairing that influences the reaction *in vitro* and because arm sites and accessory factors are missing (Kim and Gardner, 2011). Thus, IntDOT is not able to perform the second homology-independent strand exchange round. Remarkably, addition of separate DNA molecules containing the arm site sequences showed increased formation of products with mismatches at the CR, helping to overcome the GC-dinucleotide bias of the reaction (Ringwald, Yoneji and Gardner, 2017). Similar experiments using Tn*1549* Int^{FL} and arm sites did not help to drive the reaction in direction to integration products and rather lead mostly to substrate formation.

IntDOT experiments with full HJ_{long} constructs, containing arm-binding and BHFa-sites (Bacteroides host factor, an IHF-homologue) showed preferential resolution back to substrates, when BHFa was missing. However, resolution in direction to products happens when BHFa is present (Ringwald, Yoneji and Gardner, 2017). On the contrary, initial results with Int showed preferred resolution of HJ_{long} in direction to products without the need for an accessory factor, although Xis seems to stimulate the reaction.

The results of the HJ_{long} experiments with Int are fascinating because they are pointing to the idea that directionality of the reaction may be dictated by the arm sites, in a way similar to λ-Int, where asymmetry of regulatory sites arrangement dictates the order of the HJ resolution reaction.

2.3 Molecular characterization of an Integrase inhibitor

CTns are essential vehicles for the transfer of ABR determinants (see section 1.3.2). Therefore, finding conjugative transposition inhibitors may pave a new way to limit antibiotic resistance spreading among bacteria. Recent work in the lab has developed nucleic acid and peptide molecules that can inhibit Int activity *in vitro* (Rubio-Cosials et al., 2018). This work used a rational design approach, based on the previous Int-CI structure and aimed to engineer molecules that block the Int-DNA complex in an inactive state.

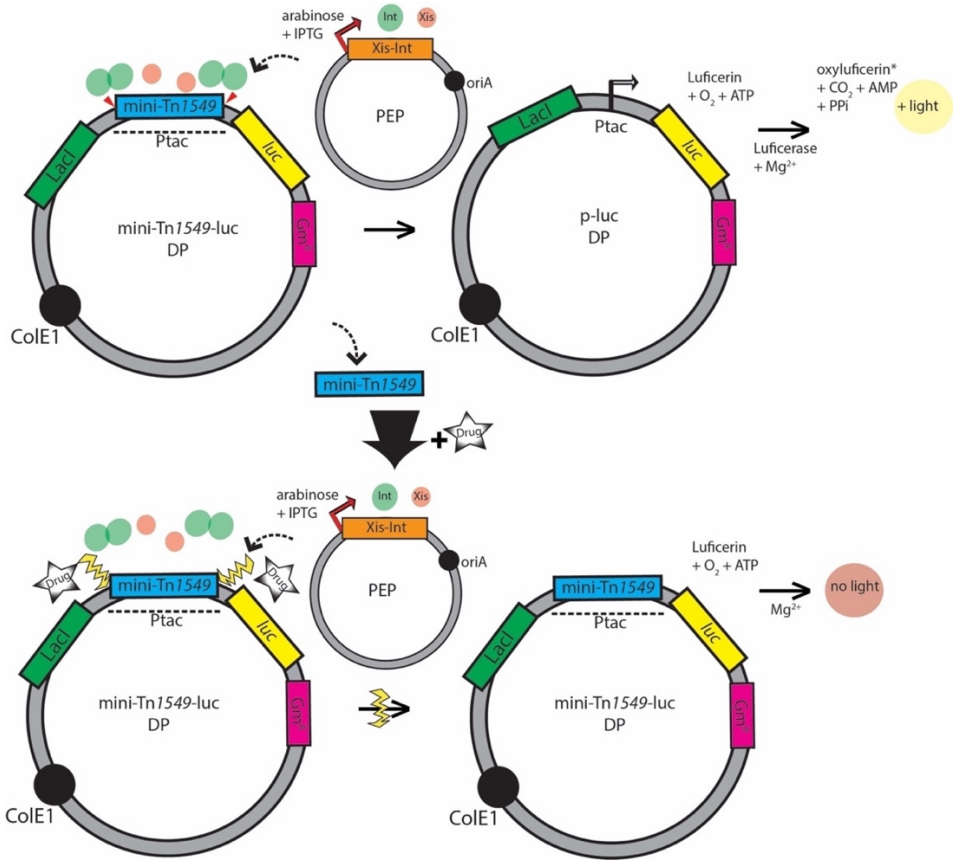

Figure 2-66: Schematics of the luciferase reporter assay. The donor plasmid (DP) contains the mini-Tn1549 transposon with all DNA sequences required for proper excision. This plasmid also has a luciferase gene (luc) after the mini-Tn1549 sequence. The second plasmid (PEP) encodes the Xis and Int proteins under the control of an arabinose inducible promoter. After Xis and Int protein expression, the mini-Tn1549 transposon is excised leading to luciferase expression and light emission. Inhibition of transposon excision leads to reduction of light emission.

Additionally, Dr. Lotte Lambertsen in the lab established a luciferase-based screening assay (Figure 2-66) to search for compounds that can inhibit Tn*1549* transposition in bacteria. This assay monitors excision of a mini-Tn*1549* element placed inside the promoter region of a luciferase gene (*luc*) on a donor plasmid in *E. coli*. Expression of the Xis and Int proteins by arabinose induction leads to mini-Tn*1549* excision and promoter regeneration

inducing luciferase expression. In the presence of luciferin, ATP, and oxygen, the expressed luciferase will emit well quantifiable bioluminescent light. In the presence of a drug that inhibits the mini-Tn*1549* excision, light-emission will be strongly diminished or absent. Using this assay, our collaborators from Gerry Wright's Lab at McMaster University in Hamilton, Canada, have identified several inhibitor candidates that prevent transposon excision *in vivo* (unpublished data). Therefore, I next wanted to validate the inhibitory effect and probe the mode-of-action of some of these compounds by using *in vitro* activity assays that I developed.

2.3.1 The effect of inhibitor candidates on Int function *in vitro*

With the help of Drew Hansen from the Wright lab, I selected four drug compounds that inhibited transposition in *E. coli* (from the luminescence-based inhibitor screens). These included the fluoroquinolones, ciprofloxacin, levofloxacin and norfloxacin as well as novobiocin, an aminocoumarin antibiotic. To probe the effect of these candidates on Int function, we used some of the activity assays *in vitro* that I previously described. So far, these assays can only test the Int activity and not the role of Xis. In particular, we performed strand-exchange and HJ-resolution assays to characterize the ability of the compounds to inhibit these specific reaction steps. The strand-exchange and HJ resolution assays are described in sections 4.3.6 and 4.3.7 respectively.

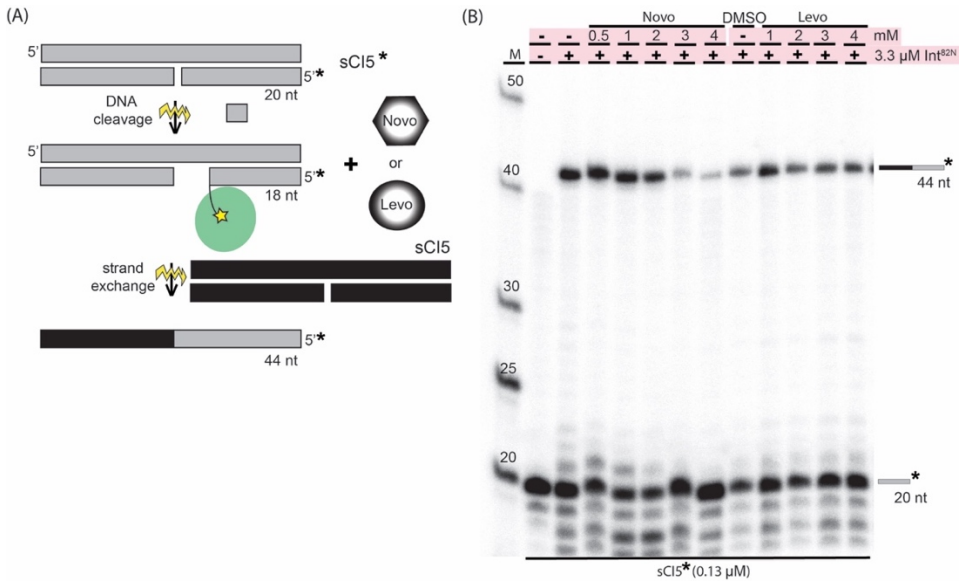

Figure 2-67: Strand-exchange assay with drug candidates selected from chemical screens: novobiocin and levofloxacin.

(A) Schematics of the strand-exchange reaction. The 5'-^{32}P-labeled, suicide substrate (sCI5, grey bar, 20 nt) is mixed with Int (green oval) that will cleave the nicked substrate strand (18 nt). In the presence of additional unlabeled DNA (sCI5 in this case, black bar), the protein can perform the ligation reaction leading to a

recombined product (44 nt). In the presence of a drug (novobiocin, hexagon or levofloxacin, oval) that antagonizes the reaction, cleavage and/or recombination reactions will be inhibited, leading to a reduction in product formation.

(B) Denaturing TBE-Urea PAGE gel shows the effect of two drug candidates on Int^{82N} strand-exchange activity with ^{32}P-radiolabeled sCI5 substrates. While levofloxacin does not inhibit the reaction, novobiocin inhibits Int activity at high concentrations. Strand-exchange reactions were performed with constant concentrations of Int^{82N} (3.3 μM), ^{32}P-labeled sCI5 (0.13 μM), and unlabeled sCI5 (6.67 μM); and increasing concentrations of novobiocin (0.5 - 4 mM) or levofloxacin (1 - 4 mM). - = no Int protein or drug was added, + = Int protein was added. All concentrations are final concentrations in the reaction volume. Novobiocin was dissolved in water and levofloxacin in 100% DMSO. Positive control: DMSO (6.7 %). Data produced jointly with Drew Hansen from the Wright lab.

First, we performed strand exchange assays with all four candidates by testing the effect of each compound in the reaction. Initial results showed that most of the candidates did not inhibit Int's activity. Ciprofloxacin and norfloxacin did not show inhibition of this reaction step (not shown). Similarly, strand exchange experiments with increasing concentrations of levofloxacin did not show significant inhibition of the reaction (Figure 2-67B). In contrast to the fluoroquinolones, novobiocin showed dose dependent inhibition, starting at 3 mM final concentration (Figure 2-67B).

Next, I tested the effect of the drugs on HJ resolution by adding the compounds in increasing concentrations to the reaction (17 – 700 μM). The results for levofloxacin and ciprofloxacin did not show inhibition of the reaction (Figure 2-68). However, novobiocin, could inhibit HJ resolution both in the direction to substrates and products. It showed complete inhibition of the reaction starting at 2 mM drug concentration (Figure 2-68A and B).

Afterwards, we tested if novobiocin could inhibit Int activity in covalent intermediate product formation using cleavage assays *in vitro* (Figure 2-69, see section 4.3.5). For that, we used the Int^{82N}390C construct because this protein variant is hyperactive in DNA cleavage and can help to detect smaller changes in reaction efficiency (see section 2.1.1.3). Again, we could observe inhibition of the reaction, starting at 2 mM drug concentration.

Taking all results together, I chose novobiocin for further detailed characterization by performing different biochemical and biophysical assays *in vitro*.

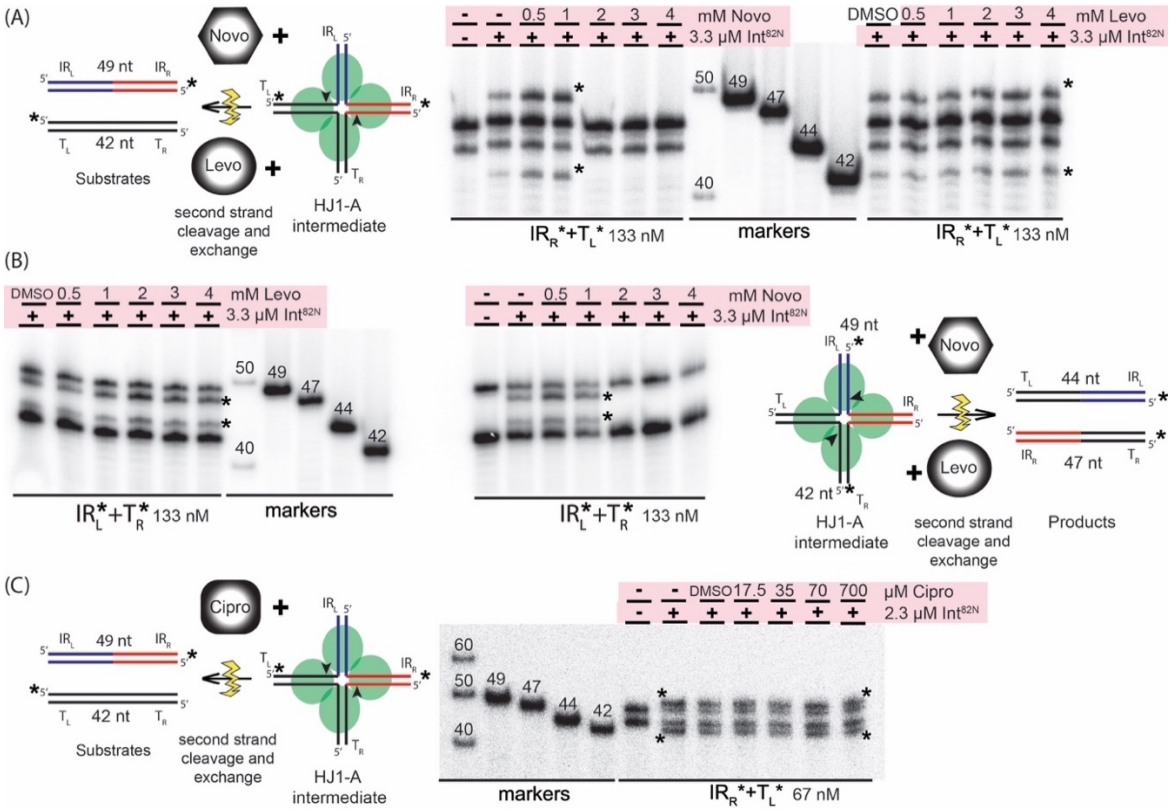

Figure 2-68: HJ-resolution assays with drug candidates selected from chemical screens: novobiocin, levofloxacin, and ciprofloxacin.

For all reactions, their schematics are on the left (A and C, direction to substrates) or right (B, direction to products) side of the gel. The 5'-^{32}P-labeled HJ substrate (HJ1-A) at two stems is mixed with Int (green oval) to resolve the junction towards dsDNA products or back to substrates. In the presence of a drug (novobiocin, hexagon; levofloxacin, oval or ciprofloxacin, roundish rectangle) that antagonizes the reaction, HJ resolution will be inhibited, leading to a reduction in product formation.

(A) Denaturing TBE-Urea PAGE gel shows the effect of novobiocin on Int^{82N} HJ resolution activity towards substrates (left). Complete inhibition starts at 2 mM novobiocin concentration. Levofloxacin does not inhibit the reaction (right).

(B) Denaturing TBE-Urea PAGE gel shows the effect of novobiocin on Int^{82N} HJ resolution activity towards products (right). Complete inhibition starts at 2 mM novobiocin concentration. Levofloxacin does not impede the reaction (left).

(C) Denaturing TBE-Urea PAGE gel shows the effect of ciprofloxacin on Int^{82N} HJ resolution activity towards substrates. Ciprofloxacin does not inhibit the reaction.

All HJ resolution reactions were performed with constant concentrations of Int^{82N} and ^{32}P-labeled HJ-DNA; and increasing drug concentrations. - = no Int protein or drug was added, + = Int protein was added. All concentrations are final concentrations in the reaction volume. Product bands on the gel are highlighted with a black star.

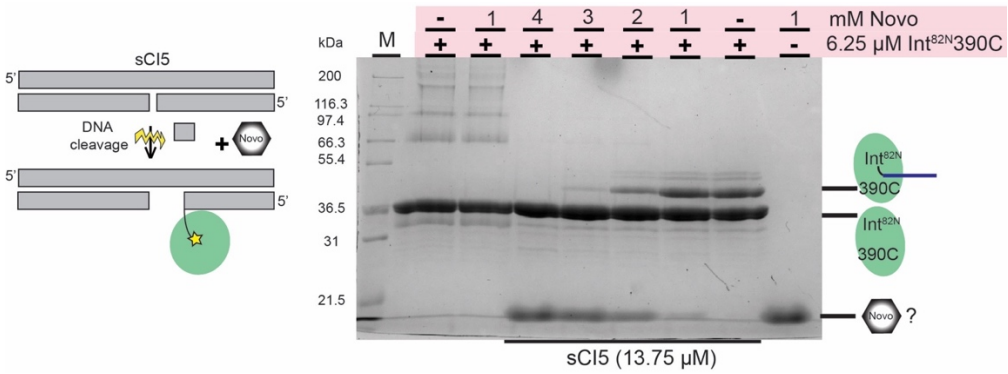

Figure 2-69: Covalent-intermediate assay with novobiocin. Schematics of the reaction (left): A nicked, suicide DNA substrate (sCI5, grey bar) is mixed with Int (green oval), leading to DNA cleavage. Two nucleotides will diffuse away, leading to a covalently trapped protein-DNA product that can be separated from unreacted protein on an SDS-PAGE gel. In the presence of Novobiocin (hexagon), cleavage reaction will be inhibited, leading to a reduction in product formation.

SDS-PAGE gel shows the ability of novobiocin to inhibit the cleavage activity of Int^{82N}390C on sCI5 substrates. Cleavage reactions with constant concentrations of Int^{82N}390C (6.25 μM) and sCI5 (13.75 μM); and increasing concentrations of novobiocin (1 - 4 mM). - = no Int protein or drug was added, + = Int protein was added. All concentrations are final concentrations in the reaction volume. For this assay, the hyperactive variant Int^{82N}390C was used because it shows more product formation. Data produced jointly with Drew Hansen from the Wright lab.

2.3.2 Characterization of the effect of novobiocin on Int recombination

First, I decided to test the novobiocin inhibition of strand exchange by using a broader concentration range, starting at very low drug concentrations (0.0125 – 4 mM). This experiment showed that novobiocin inhibits the reaction in a concentration-dependent manner, and complete inhibition is reached only at high drug concentrations (~2 mM) (Figure 2-70).

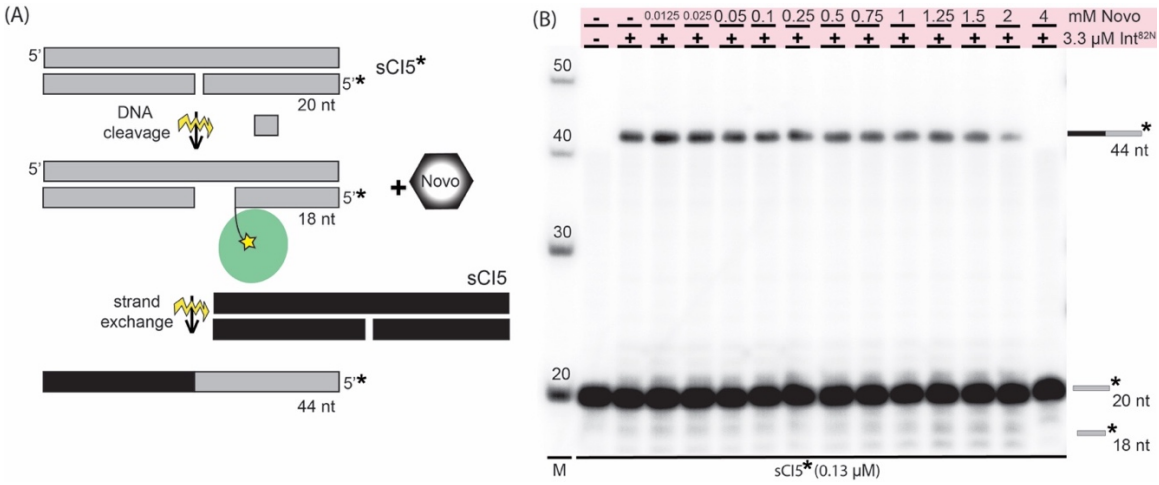

Figure 2-70: Strand-exchange assay with novobiocin, using a broader concentration range of the drug candidate.

(A) Schematics of the strand-exchange reaction (for details see Figure 2-67).

(B) Denaturing TBE-Urea PAGE gel shows the effect of novobiocin on Int^{82N} strand-exchange activity with ^{32}P-radiolabeled sCI5 substrates. Novobiocin inhibits Int activity in a concentration-dependent manner. Strand-exchange reactions were performed with constant concentrations of Int^{82N} (3.3 µM), ^{32}P-labeled sCI5 (0.13 µM), unlabeled sCI5 (6.67 µM); and increasing concentrations of novobiocin (0.0125 - 4 mM). - = no Int protein or drug was added, + = Int protein was added. All concentrations are final concentrations in the reaction volume.

Next, I tested HJ-resolution inhibition by using a broader drug concentration range (0.0125 – 2 mM). In this experiment, I noticed an increase in product formation at 0.5 – 1 mM novobiocin, followed by Int^{82N} activity reduction at 1.25 mM, before complete inhibition was reached at 1.5 mM novobiocin concentration (Figure 2-71B). Therefore, surprisingly novobiocin exhibits a dualistic behavior, and may inhibit or increase Int^{82N} activity depending on the concentration. Moreover, this result was highly reproducible in several HJ resolution experiments.

Then, I decided to test if novobiocin may be inhibiting the reaction by interfering with protein-DNA binding. EMSA assays with CI5 and HJ DNA (Figure 2-72) showed that protein-DNA complex formation is strongly reduced at novobiocin concentrations similar to ones used in the strand-exchange experiments. DNA binding decreases in a concentration-dependent manner and shows almost complete loss of binding at 4 mM drug concentration (Figure 2-72A). Similar step-wise inhibition could be also observed for Int binding to HJ-DNA (Figure 2-72B).

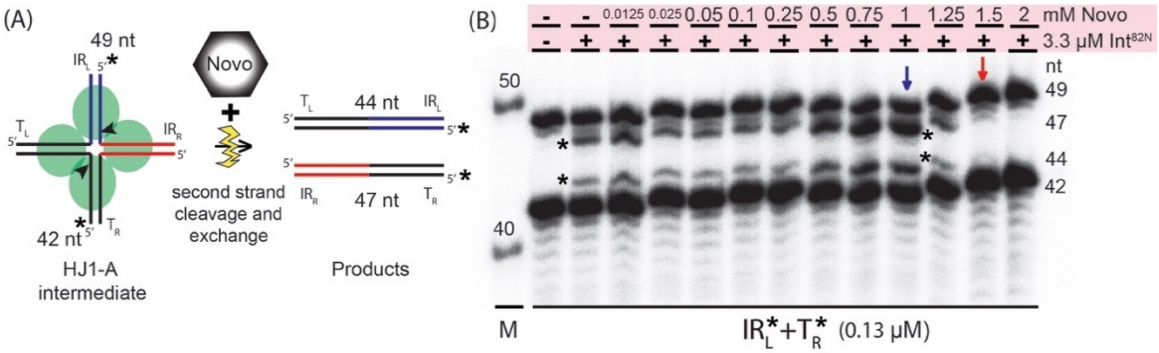

Figure 2-71: HJ-resolution assay with novobiocin, using a broader concentration range of the drug candidate.

(A) Schematics of the HJ-resolution assay. The 5'-^{32}P-labeled HJ substrate (HJ1-A) at two stems is mixed with Int (green oval) to resolve the junction towards dsDNA products. In the presence of novobiocin (hexagon) that antagonizes the reaction, HJ resolution will be inhibited, leading to a reduction in product formation.

(B) Denaturing TBE-Urea PAGE gel shows the effect of novobiocin on Int^{82N} HJ resolution activity towards products. Inhibition starts at 1.25 mM drug concentration and reaches completeness between 1.5 - 2 mM novobiocin (red arrow). There seems to be increased Int^{82N} activity at 0.5 - 1 mM drug concentration (blue arrow). All HJ resolution reactions were performed with constant concentrations of Int^{82N} (3.3 µM) and ^{32}P-labeled HJ-DNA (0.13 µM); and increasing concentrations of novobiocin (0.0125 - 2 mM). - = no Int protein or drug was added, + = Int protein was added. All concentrations are final concentrations in the reaction volume. Product bands on the gel are highlighted with a black star. M = marker.

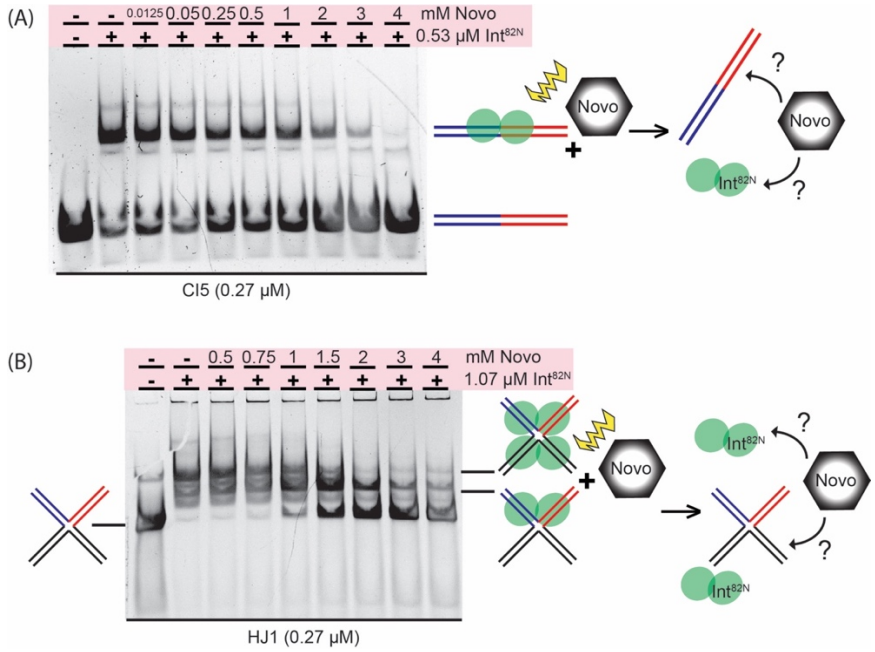

Figure 2-72: Characterization of Int^{82N}-CI-novobiocin and Int^{82N}-HJ-novobiocin complexes by EMSA.

(A) The native gel shows novobiocin's ability to hinder Int^{82N}-CI5 binding and complex formation in a concentration-dependent manner. Binding assays were performed with constant concentrations of CI5 (0.27 µM) and Int^{82N} (0.53 µM) in 1:2 molar ratios; and increasing concentrations of novobiocin (0.0125 - 4 mM).

(B) The native gel shows novobiocin's ability to disturb Int^{82N}-HJ binding and complex formation in a concentration-dependent manner. Binding assays were performed with constant concentrations of HJ1 (0.27 µM) and Int^{82N} (1.07 µM) in 1:4 molar ratios; and increasing concentrations of novobiocin (0.5 - 4 mM).

- = no Int protein or drug was added, + = Int protein was added. All concentrations are final concentrations in the reaction volume.

To learn more about how novobiocin interacts with Int, I aimed to characterize the Int-novobiocin complex formation by using different biophysical methods with the help of Dr. Kathryn Perez from the Protein Expression and Purification Core Facility (PEP-CORE) at EMBL. We performed initial Isothermal titration calorimetry (ITC) experiments to study novobiocin-binding to Int^{82N} (see section 4.6.1). ITC measures the energy required to keep the temperature constant during a titration experiment - i.e. to compensate for the heat change that results from the direct interaction of molecules in the system, allowing the calculation of binding constants and other thermodynamic parameters. For that, we titrated a constant concentration of Int^{82N} (20 µM) with increasing novobiocin concentrations. We performed two titration experiments that differed in the novobiocin concentrations used (0.2 and 1 mM). The raw data (Figure 2-73) did not show any heat changes during the measurement (none or very tiny peaks visible). Thus, no binding isotherm could be calculated, as shown in the Wiseman plots (Figure 2-73). In conclusion, at these specific conditions and molecule concentrations, no protein-drug binding could be detected.

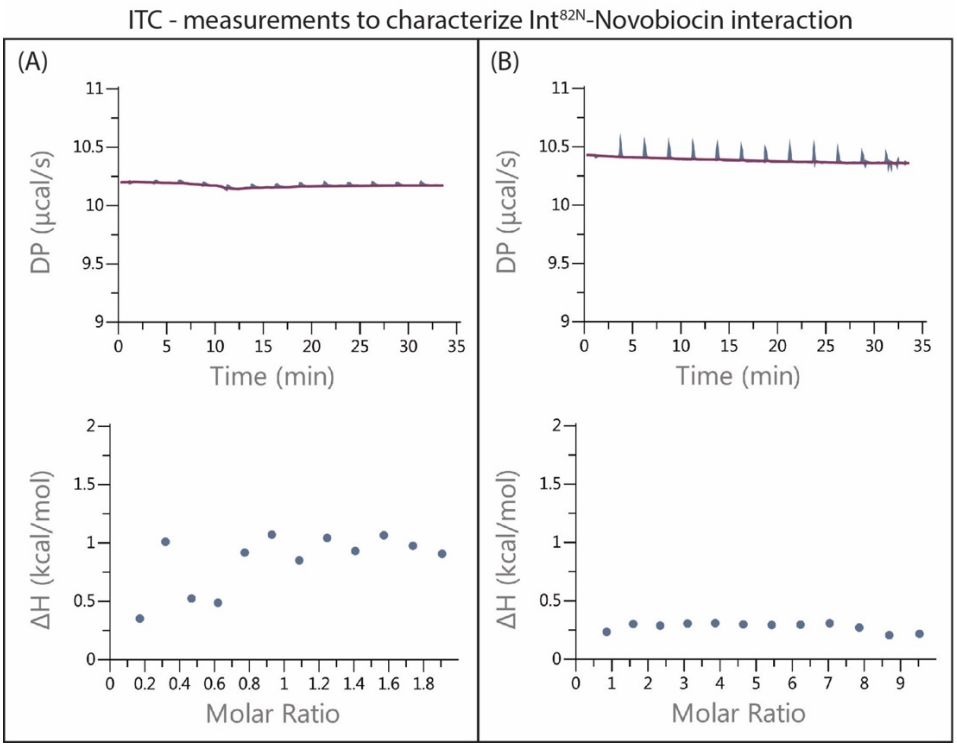

Figure 2-73: Isothermal titration calorimetry (ITC) measurements for characterization of Int^{82N}-novobiocin interaction. The top panels show the raw data, differential power (DP) curve. The bottom panels plot the integrated heat peaks (raw data) against the molar ratio of the drug (Wiseman plot); calculation of the binding isotherms was not possible. Int^{82N} in constant concentration (20 µM), titrated with increasing amounts of novobiocin using 0.2 mM (A) and 1 mM (B) stock concentrations.

Next, we tested protein-drug interaction by performing Nano differential scanning fluorimetry (NanoDSF) experiments (see section 4.6.2). This method helps to calculate the protein's melting temperature (T_m) by measuring changes in the intrinsic fluorescence of proteins (originating mainly from tryptophan and tyrosine residues) during heat induced unfolding. For that, we measured tryptophan fluorescence emission at 330/350 nm wavelength (excitation at 280 nm). We observed a melting temperature for Int^{82N} alone at 46 °C (data not shown). Then, we measured fluorescence of Int^{82N} in melting curves with various novobiocin concentrations (0.5, 1, and 4 mM). We used drug concentrations of up to 4 mM to reconstitute the inhibitory concentration range observed in the activity assays. We observed a small decrease of the T_m to 43 °C, at 0.5 mM novobiocin. However, unfortunately, the presence of novobiocin largely reduced tryptophan fluorescence, and we saw a very weak signal already at 0.5 mM drug concentration. For higher compound concentrations, practically no tryptophan signal was visible. This may be explained by quenching of the signal due to overlapping tryptophan emission (330/350 nm) and

novobiocin excitation range, which absorbs UV-light in the 300-390 nm range. These results showed that NanoDSF cannot be used for Int-novobiocin characterization.

Table 2-6: Summary of melting temperatures (T_m in °C) from four different Thermofluor runs for analysis of Int-novo interactions by increasing the drug concentration. Run 1: 350 mM NaCl, 50 mM HEPES pH 7.5, 0.25 mM TCEP. Run 2: 175 mM NaCl, 50 mM HEPES pH 7.5, 0.25 mM TCEP. Run 3 and 4: 350 mM NaCl, 50 mM HEPES pH 7.5, 5 % Glycerol, 0.25 mM TCEP. Controls: c1 – Sypro alone; c2 – novo (1 mM) with Sypro. Both controls did not show a significant fluorescence signal.

Novobiocin concentration	**Run 1 (T_m)**	**Run 2 (T_m)**	**Run 3 (T_m)**	**Run 4 (T_m)**
0 µM	44	42	45	46
10 µM	44	42	45	44
20 µM	44	42	45	45
100 µM	43	41	45	45
500 µM	40	40	43	43
1 mM	-	-	39	39
1.5 mM	-	-	29/40	28
2 mM	-	-	(23)	27
3 mM	-	-	(29)	(37)
4 mM	-	-	(33)	(33)

Finally, I investigated protein-drug interaction using thermofluor assays (see section 4.6.3). This assay can measure the protein's thermostability and helps to calculate the protein's melting temperature by using a fluorescent dye, which interacts with the hydrophobic parts of the unfolding protein that get exposed during temperature increasing. These experiments showed that novobiocin decreases the T_m of Int, starting at 0.1 mM novobiocin with a constant concentration of Int (20 µM) (Table 2-6, Figure 2-74). At high drug concentrations (1 – 4 mM), part of the protein precipitated after mixing. Thus, melting points could not be calculated for this range due to signal saturation (Figure 2-74A).

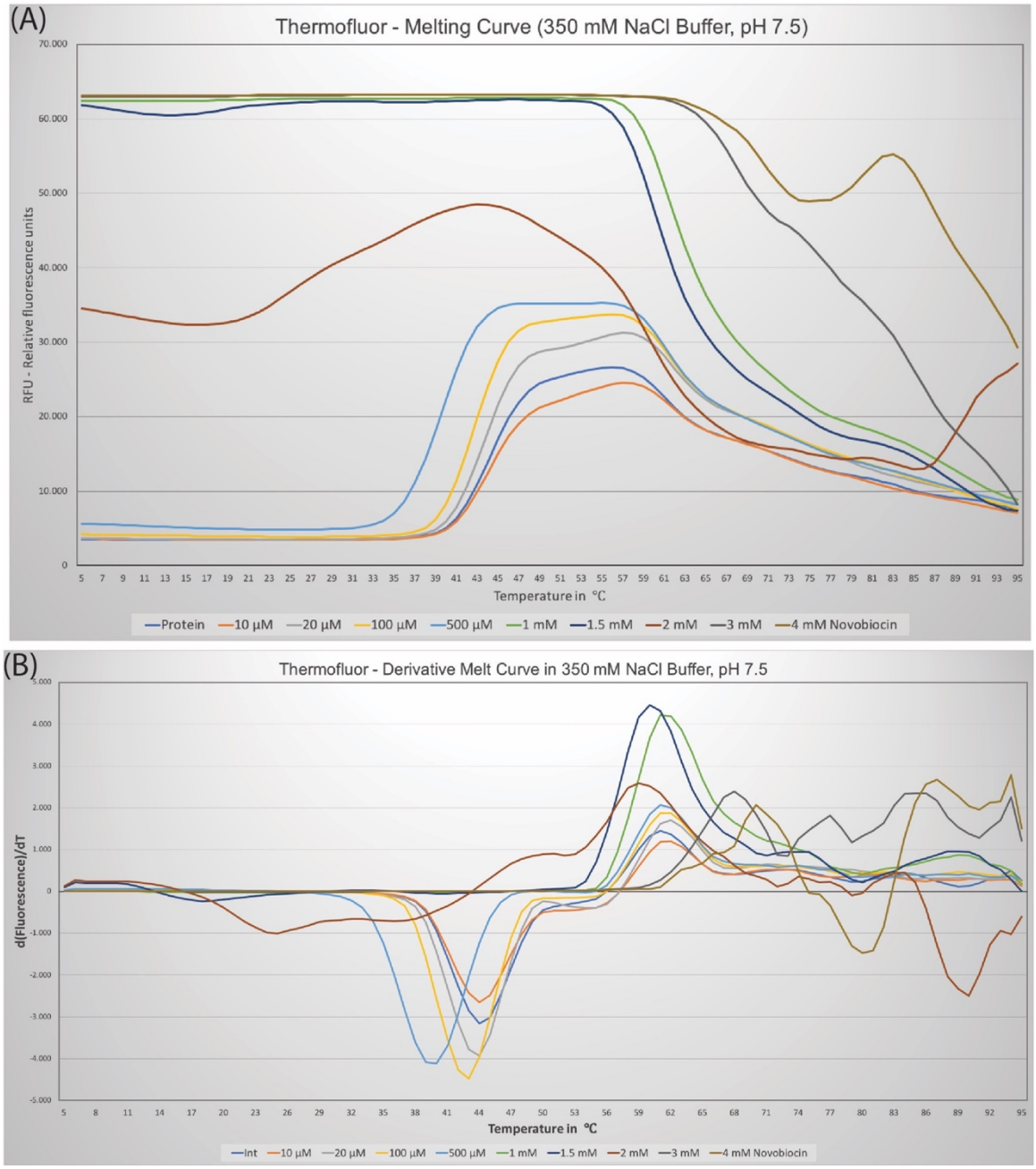

Figure 2-74: Thermofluor assay results showing the effect of novobiocin on Int's thermostability by increasing drug concentrations. (A) Relative fluorescence units (RFU) - and (B) first derivative - melting curves after a thermofluor run to analyze Int-novo interactions (run 1).

2.3.3 Discussion

2.3.3.1 Fluoroquinolones do not inhibit Tn1549 Int activity *in vitro*

The fluoroquinolones norfloxacin, ciprofloxacin and levofloxacin were promising candidates found in previous chemical screens. They are also well known for inhibiting the bacterial DNA gyrase and topo IV proteins (type IIA topoisomerases). However, they did not reveal inhibition in *in vitro* activity assays with Int^{82N} using either CI5 or HJ DNA

substrates, even at very high drug concentration. For these compounds, it is proposed that they inhibit topoisomerase proteins by binding to a protein-DNA complex in a particular manner, interacting with the cleaved DNA and trapping the reaction at the covalent protein-DNA intermediate (Laponogov *et al.*, 2009; Bax *et al.*, 2010). This complex could resemble one of the Int-DNA recombination steps that also shows a covalently linked protein-DNA intermediate during the reaction (see section 2.1.1), but my experiments indicate that Tn*1549* inhibition does not involve fluoroquinolone-mediated trapping at this step of the reaction.

Given the fact that *in vivo* assays were performed in *E. coli* cells, where fluoroquinolone antibiotics can inhibit gyrase, we hypothesize that Tn*1549* inhibition may be indirect in these cases and gyrase inhibition may indirectly influence mini-transposon excision. To test this possibility, follow-up experiments were recently performed in fluoroquinolone-resistant bacterial cells that encode a gyrase mutant (data not shown). These experiments indicated that excision inhibition may have involved gyrase inhibition, suggesting that gyrase might somehow support transposon excision. However, we cannot exclude so far that fluoroquinolones also have a direct effect on some of the so far unexplored steps of Int action.

2.3.3.2 Novobiocin inhibits several steps of the transposition reaction at high concentrations

The only candidate that showed an inhibitory effect in my *in vitro* assays was novobiocin (Figure 2-75). This chemical, also known as Albamycin, is an antibiotic of the aminocoumarin class (Heide, 2009). It was used previously against bacterial infections of mainly Gram-positive bacteria. But it has been recently withdrawn from the market by the Food and Drug Administration due to poor pharmacological properties and safety concerns (FDA, 2011).

Its mode of action consists of inhibiting the bacterial DNA gyrase, similar to the fluoroquinolone antibiotic family. But instead of binding to the GyrA monomer, novobiocin interacts with the GyrB ATPase subunit. It competes for the ATP binding site leading to energy depletion of the enzyme, which is needed for GyrA to perform the DNA cleavage reactions. Moreover, novobiocin binding inhibits GyrB dimerization. The crystal structure of the GyrB domain solved together with this drug shows that novobiocin interacts with a critical arginine residue (R136) that is necessary for protein dimerization. Naturally occurring coumarin-resistant mutants have mutations at R136 (Holdgate *et al.*, 1997).

Later, it was found that this compound can also modulate the function of Hsp90 molecular chaperons in eukaryotic cells by interacting with their ATP-binding C-terminal domains. This made novobiocin a potent cancer drug candidate (Li *et al.*, 2009).

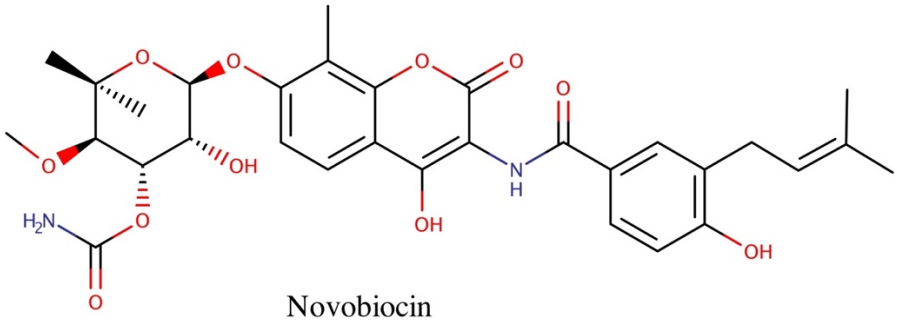

Figure 2-75: Structure of novobiocin (Boles and Taylor, 1975), an aminocoumarin antibiotic and inhibitor of the bacterial topoisomerase, DNA gyrase.

The binding affinity of aminocoumarin antibiotics to bacterial gyrases is robust, showing equilibrium dissociation constants (K_D) in the low nM range (around 10 nM) (Maxwell and Lawson, 2003). In contrast, for the novobiocin-Int interactions, my initial biophysical experiments and *in vitro* activity assays are pointing to a very weak affinity. ITC experiments did not detect binding at up to 1 mM drug concentrations, but biochemical activity and DNA binding was compromised at 0.5 – 2 mM concentration.

The thermofluor assays showed that novobiocin reduces the T_m of Int. This observation was more pronounced, starting at 0.5 mM drug concentrations. This suggests that binding of the compound might reduce conformational stability of the protein, leading to easier unfolding at lower temperatures. Consistently, I noticed rapid protein precipitation a few minutes after protein-drug incubation at higher concentrations.

It is possible that novobiocin may interact stronger with an Int-DNA complex. Future biophysical interaction studies of Int-DNA (CI5 or HJ1) complexes with novobiocin will be needed to test this possibility and shed more light on the inhibition mechanism.

Novobiocin was recently withdrawn from the market and it is no longer used as an antibiotic in humans. Nevertheless, its activity against Int could help to discover specific transposase inhibitory mechanisms and promote the design of new drug candidates against AB spreading. Although the biophysical experiments did not show strong binding of novobiocin, it can provide a starting point (lead compound) for the design of more potent and safer antagonists.

Furthermore, the novobiocin experiments provide interesting mechanistic clues about Int action. For example, EMSAs showed that novobiocin does not block completely the binding of Int to HJ DNA even at 4 mM compound concentration (Figure 2-72B). An Int dimer can still bind to HJ DNA even at 2 – 4 mM, while the tetramer complex band disappears at around 1.5 – 2 mM concentration. On the other hand, Int binding to CI5 DNA is already fully inhibited at 3 – 4 mM drug concentration (Figure 2-72A). Interesting is that inhibition of HJ resolution by Int also starts at around 1.5 – 2 mM drug concentration (Figure 2-71). Thus, the binding assays indicate that novobiocin is interfering with protein tetramerization at approximately 1.5 mM concentration. Moreover, it suggests that Int dimer-HJ complexes cannot resolve HJ-DNA and that a tetrameric complex of Int is necessary to perform this step of the recombination reaction. This observation strongly supports our Int-HJ tetramer structure, indicating that we trapped a real intermediate of the CTn recombination pathway.

The results of the strand-exchange and the binding assays show a clear concentration-dependent interaction of the drug with the protein-DNA complex at increasing inhibitor concentrations (Figure 2-70). However, visible changes are only observed, starting from 0.5 mM concentration onwards. In contrast, HJ resolution shows a more complex response to novobiocin concentration (Figure 2-71). In the μM range, Int's HJ-resolution activity seems to remain unchanged, then novobiocin increases resolution efficiency at a range of 0.5 – 1 mM drug concentration. Finally, a further increase in the drug concentration leads to inhibition of the reaction. The unexpected finding of increased HJ resolution efficiency at specific drug concentrations still remains enigmatic and requires further studies.

In summary, novobiocin helped us to provide new insights into the importance of the Int tetramer in HJ resolution, confirming structural and biochemical experiments discussed in the previous chapters (see sections 2.1.2 and 2.2). Nevertheless, many questions remain open about the exact mechanism of Int inhibition by novobiocin. Thus, further studies are needed to pinpoint the exact way this drug interacts with the Int-DNA complex. Nevertheless, novobiocin may serve as a starting point and be a lead compound for future development of new drug candidates against ABR spreading.

3 GENERAL DISCUSSION AND CONCLUSIONS

3.1 The Tn*1549* integration pathway

The work presented in this thesis investigated different aspects of the integration reaction of the conjugative transposon Tn*1549* from *E. faecalis*. For that, the proteins Int and Xis were studied in detail because they are responsible for performing all reactions necessary to excise the element from the host genome and integrate into the recipient's chromosome (Lambertsen *et al.*, 2018; Rubio-Cosials *et al.*, 2018). Conjugative transposons are proposed to be major players in spread of ABR among bacteria, promoting the emergence of MDR pathogens (Partridge *et al.*, 2018). Tn*1549* carries resistance to vancomycin (Garnier *et al.*, 2000), an antibiotic that is used as a last-resort treatment, for example, against MRSA infections (Zeng *et al.*, 2016). Tn*1549* can integrate at various AT-rich target sites that are not conserved (Lambertsen *et al.*, 2018). To achieve this, its recombinase (Int) can perform the necessary recombination reactions of two DNA sites without the necessity of homology, which is in strong contrast to the related site-specific tyrosine recombinases. Until now, Int was only trapped in an inactive pre-synaptic complex, as seen in its crystal structure that contains two Int monomers bound to a circular intermediate DNA (Rubio-Cosials *et al.*, 2018). A striking feature was that the structure trapped a dimer of Int molecules instead of a tetrameric synaptic complex, seen before for other members of the site-specific recombinase family. This opened many questions regarding further steps of transposition. Therefore, in this work, I wanted to answer how similar or different CTn integrases and site-specific tyrosine recombinases are concerning the following steps of recombination (discussed in section 1.6.3). Shedding light on more steps of the mechanistic pathway of CTn integration was my main aim in order to better understand these promiscuous MGEs and maybe enable the design of novel antibacterial compounds to limit the spread of ABR among bacteria.

This work concentrated on investigating specific steps of the integration pathway by studying Int and Xis proteins. For that, I addressed questions concerning how Int specifically recognizes, cleaves, and recombines DNA-intermediates of the integration reaction. Moreover, I tried to find out how the directionality of the reaction is controlled. In particular, I wanted to find out if an active synaptic complex of Int and CI DNA may feature a tetrameric state. Moreover, I tested if Int can form stable complexes with HJ DNA, which is a key intermediate of the reaction, and learn how such assemblies may look at the

molecular level. I also tested the accessory protein Xis for DNA binding and its role in the regulation of HJ resolution. Moreover, I aimed to characterize this Xis at the structural level. Answers to these questions could apply more generally to other CTn systems, where the lack of structural information on the protein-DNA recombination complex assembly has, so far, prevented full characterization of the recombination process.

Biochemical and structural studies in this work showed that, indeed, the Tn*1549* integration mechanism seems to resemble several features of the canonical tyrosine recombination pathway. On one side, obtaining the crystal structure of Int covalently bound to CI-DNA was so far not successful due to low activity of the Int^{82N} construct, flexibility of the Int^{82N}390C variant, and/or heterogeneity of the samples (section 2.1.1). On the other hand, I could show stable Int-HJ complex formation and solved the crystal structure of this assembly (section 2.1.2). The solution shows four Int monomers bound to one HJ molecule, forming a semi-symmetric tetrameric complex (Figure 2-20) that is characteristic for members of the tyrosine recombinase family, like λ Int and Cre recombinase (section 1.4.3). Another resemblance is the cyclic exchange of the C-terminal tails, which is important for complex stability and probably for allosteric regulation of protein activity (Figure 2-30). Identification of specific differences, for example, in protein-DNA interactions or in the catalytic pocket, will require further work to improve the resolution of the structure, either by overcoming problems during structure refinement or by producing better diffracting crystals (section 2.1.2.9.1).

Nevertheless, I could reveal mechanistic features unique to Int and this specific family of conjugative transposases in biochemical studies of HJ resolution. Int exhibited its ability to resolve HJ intermediates with a non-homologous overlap region. However, this reaction is deregulated on short HJ DNA molecules and can proceed in both directions (section 2.2.1). Preferred resolution towards products, could not be achieved using the full-length integrase and arm DNA sites added on a separate DNA piece (section 2.2.4). However, protein-DNA complexes including full-length Int, arm DNA and CI or HJ DNA could be reconstituted *in vitro* (section 2.2.4), showing that Int behaves like λ Int and can simultaneously bind to arm DNA sites and recombination substrates. These results are in contrast with previous HJ-resolution experiments with IntDOT, which showed that protein binding to arm sites led to a slight increase in product formation (section 2.2.6.3). This highlights that despite the overall structural conservation, tyrosine recombinase proteins differ in how they interact with their particular DNA sites and in how the reaction steps are controlled.

Nevertheless, initial experiments using long HJ intermediates that contain the arm DNA sites in their sequence, showed preferred resolution in direction to products. This preference further increased by adding the accessory factor Xis (section 2.2.5). This striking result indicates a new role of Xis in guiding the directionality of the recombination in CTn transposition. This is unlike its typical role observed in the bacteriophage-λ, where Xis promotes excision and inhibits integration (Landy, 2015). Studies of Tn*916* Xis already hinted at the possibility that excisionase proteins from this CTn family may differ from the λ Xis prototype and may fulfill slightly different roles in the recombination reactions (section 2.2.6.2).

Binding studies showed that Tn*1549* Xis interacts weakly with its prospective binding sites at the transposon ends, similar to Tn*916* Xis. This is again unlike λ Xis, which binds strongly to its DNA sites *in vitro* (section 2.1.3.6.1). Obtaining the crystal structure of Xis showed the conserved winged helix domain fold that is needed for DNA-recognition (Figure 2-38). Structurally and sequentially, Tn*1549* and Tn*916* Xis proteins are well conserved. They have only a single amino acid difference at the predicted protein-DNA binding interface (section 2.1.3.6.2). Nevertheless, their binding sites are not conserved. Interestingly, despite the weak DNA binding affinity, the Tn*1549* excisionase was able to resolve CTn HJs in HJ_{long} resolution experiments, indicating that transient interaction is enough to help in these reactions *in vitro*.

Studies of Tn*1549* integration inhibition *in vitro* with novobiocin displayed the ability of this drug candidate to hinder Int activity (section 2.3.2). Novobiocin led to inhibition of DNA binding, cleavage, and strand-exchange with CI substrates and hampered HJ DNA resolution. However, inhibition was observed only at relatively high novobiocin concentrations, starting at 0.5 mM onwards and reaching full inhibitory capacity at 2 - 4 mM compound, depending on the assay and the recombination step. Int-drug interaction could not be detected via ITC or NanoDSF at the optimal conditions of these methods (section 2.3.2). However, full biophysical characterization of Int-novobiocin and Int-DNA-novobiocin complexes still need to be performed. Nevertheless, these studies helped to confirm the direct effect of novobiocin on Int-mediated transposon mobilization. Furthermore, the experiments showed that novobiocin primarily hampers Int tetramerization and validated the importance of the tetramer assembly for HJ resolution (section 2.3.3.2).

3.2 An updated integration model for Tn*1549*

By combining all the results obtained in this study, an updated model for Tn*1549* integration can be presented (Figure 3-1). This model integrates the results of my thesis work with the previous model from (Rubio-Cosials *et al.*, 2018), which was introduced in section 1.6.3.

1. A protein dimer binds to the AT-rich inverted repeats (IR_L and IR_R) in the excised circular transposon intermediate.

2. A second dimer probably binds to a similar AT-rich sequence, the integration target site, in the bacterial chromosome (T_L and T_R). In both complexes, Int will melt the DNA at the crossover region and persist in a pre-catalytic inactive state until the two complexes come together to form a synaptic complex with four Int monomers and two DNA molecules. Synaptic complex assembly is likely facilitated by the interaction of Int's AB domain with specific arm DNA sites inside the transposon (composed of direct repeats, shown in section 2.2.4), which probably connects the two dimeric complexes in the tetrameric state. As demonstrated by my HJ resolution assays, presence of the arm sites in the recombined DNA in accurate special arrangement with the core recombination sites is critical to drive the reaction to the final products. Tetramer assembly may further be supported by Xis, as indicated by my HJ_{long} resolution results (section 2.2.5). Other DNA bending proteins, like HU, may additionally promote or regulate DNA recombination, as suggested for the Tn*916* system *in vivo* (Connolly, Iwahara and Clubb, 2002). These factors can help bring arm- and core-binding sites together by interacting with respective binding sites, usually located between the Int binding sites in the transposon DNA, as shown in the λ-bacteriophage system (Landy, 2015) and discussed in 2.1.3.6.

3. Tetrameric complex formation and conformational rearrangements, like the cyclic exchange of the C-terminal domains, lead to catalytic activation of two Int monomers, one at each recombination site. The first cleavage and strand exchange reaction occur at IR_R of the transposon CI and T_L of the integration target site. This leads to the formation of the HJ DNA intermediate. The crystal structure of the Int-HJ complex solved in this thesis, shows four Int molecules binding to a four-way DNA junction composed of CI (IR_L-IR_R) and target (T_R-T_L) sequences. Each Int monomer interacts with one stem of the HJ DNA, forming a four-fold semi-symmetric assembly, with the C-terminal tails swapped between the subunits in a circular fashion.

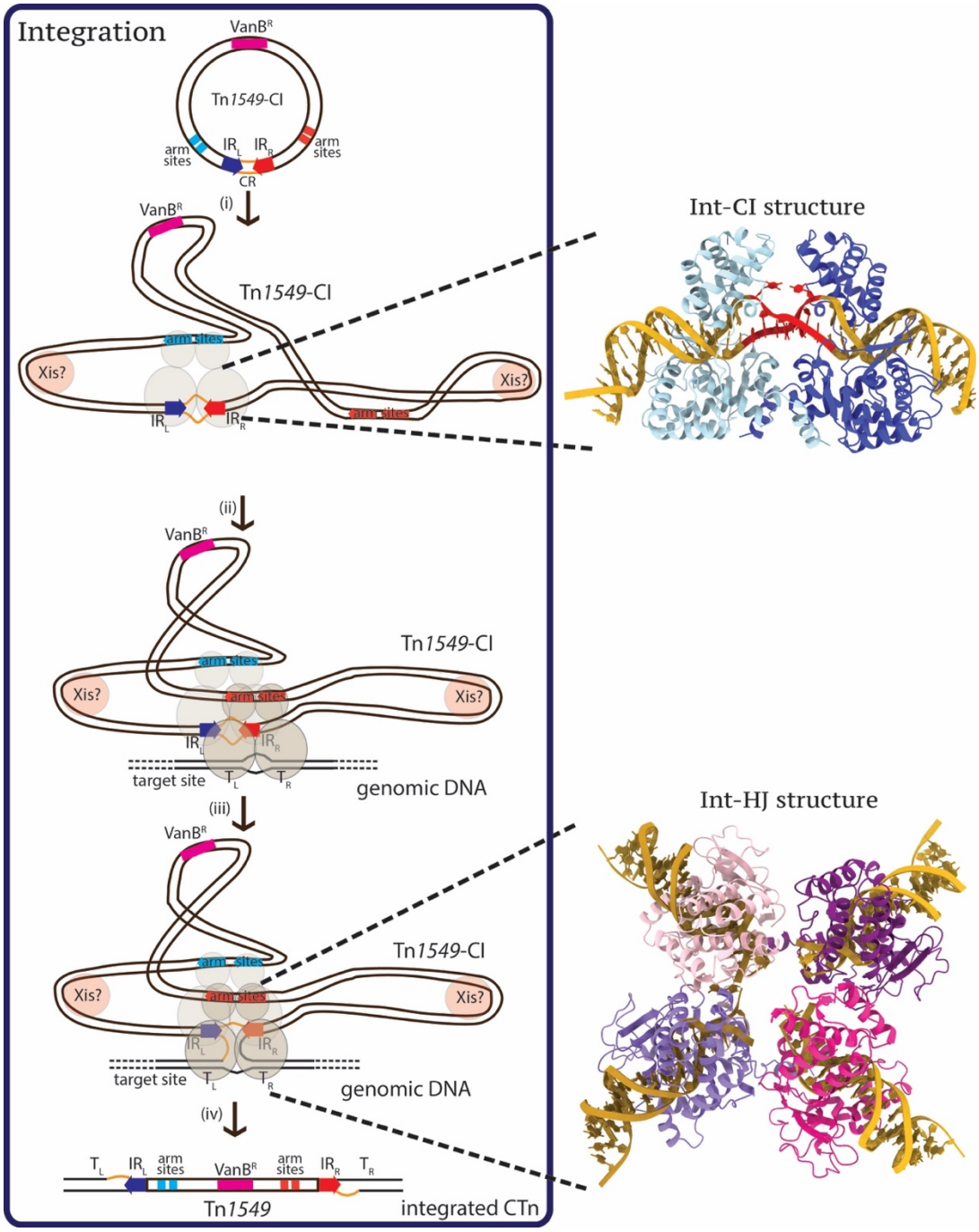

Figure 3-1: Updated integration model of Tn1549. Tn1549 is in its circular intermediate (CI) state. The CI central region is composed of IR_L (blue arrow) and IR_R (red arrow), connected by a crossover region (CR, orange lines). It also encodes arm sites composed of direct repeats (DR, in light-blue or light-red for the left and right sequences). VanB resistance cassette is marked as a rectangle (magenta). (i) A dimer of Int (semi-transparent, grey ovals: big oval represents CB-CAT domain, small oval shows AB domain) binds to the CI in an inactive state. Crystal structure of an Int dimer containing two CB-CAT domains bound to the core CI sequence was solved previously (Rubio-Cosials et al., 2018). The process of synaptic complex formation and tetramerization will be supported by bending of the left- and right-sites in the CI region probably with the help of Xis or other host-encoded accessory factor (circle, light red) by binding prospective sequences between the Int-binding sites and bringing these sites together to Int. Int's AB domains bind to the arm DNA sites, whereby left- or right arm sites could bind to the Int-CI dimer. Here, Int-binding to the left arm sites is shown as an example. (ii) The second pair of Int proteins bound to an AT-rich target site is brought together with the Int-CI complex via interactions of the AB domains. During this process, exchange of the C-terminal domains occurs on a circular fashion and stabilizes the tetrameric complex. (iii) These protein-DNA rearrangements lead to the activation of two monomers at IR_R and T_L, inducing cleavage and strand exchange

reactions and leading to the formation of a HJ intermediate. The Int-HJ complex was crystallized, and its structure was solved in this work. It contains four CB-CAT monomers and a HJ DNA composed of the core CI and target sequences. (iv) After further protein-DNA rearrangements, the other pair of monomers will be activated and lead to the second round of cleavage and strand exchange reactions, resulting in transposon integration into the recipient's genome. This reaction produces sort non-homologous regions at the flanks of the transposon ends (highlighted by the difference in colors, orange for the CR of the CI and black for the target site).

4. Afterwards, further rearrangements at the protein- and DNA-level are expected to happen, leading to activation of the second pair of monomers that did not participate in the first round of cleavage reactions. These adjustments trigger the second round of cleavage and strand exchange reactions at IR_L of the CI and at T_R of the target, leading to resolution of the HJ intermediate (shown in section 2.2.5). This last step of the reaction leads to the generation of recombined products and transposon integration into the recipient genome.

With respect to HJ resolution, canonical site-specific recombinases are assumed to follow a so-called "strand-swapping" model, melting only 2-3 bp of the overlap region at the cleavage site and annealing with the complementary sequence during the strand exchange reaction, as proposed in (Van Duyne, 2001). This model cannot be followed in the CTn system because the central parts of the recombination sites are non-homologous. As shown in the Int-CI structure, the crossover region is actively melted by Int already prior to first DNA cleavage, which likely enables strand exchange in absence of complementary base pairing on the partner strands. Consistently, the central part of the junction DNA in the Int-HJ structure is not well-resolved, probably due to flexibility, which supports the idea that the CR bases are melted throughout the recombination process. Thus, although Int requires one base-pair microhomology for finding and attacking the correct 3'-phosphotyrosine bond on the recombination partner DNA, the complex intermediates have to maintain perfect conditions to keep the DNA melted. This will help to drive the reaction forward to generate non-homologous products and avoid reformation of the base-pairs at the original recombination sites reverting the reaction to substrates (CI and target).

As shown by my HJ_{long} resolution experiments, the AB-arm site interactions probably dictate the reaction's directionality and drive the outcome of the reaction and favor integration. Similar results were shown for the λ-bacteriophage system (Landy, 2015), but specific characteristics of the regulation might differ. It will be of great interest to discover if specific differences in this regard also help to overcome the challenge of recombining non-homologous sites. The last step of the reaction will lead to the generation of recombined products and transposon integration into the recipient genome.

3.3 Future directions

The conjugative transposon Tn*1549* has been discovered more than 20 years ago, and the Tn*916*-like family is known for more than three decades. Nevertheless, many questions regarding this MGE family remain open. The presented work offers a basis for further studies of the Int-Tn*1549* recombination system. Furthermore, the results presented in this study introduce new questions that can be addressed in the future by combining structural and molecular biology approaches.

The current work provided insights into one specific step of the integration mechanism of conjugative transposition, namely by solving the Int-HJ1TA crystal structure (Figure 2-20). This confirmed that CTn integrases go through an Int-HJ tetramer state to recombine non-homologous DNA sites and insert the transposon into the bacterial genome, resembling the mechanism of the site-specific tyrosine recombinases. From a structural point of view, several further intermediates of the integration and the excision reactions remain to be elucidated to fully map the transposition process and answer if the other reaction intermediates resemble the conserved mechanism of tyrosine recombination. In particular, it would be very valuable to crystallize Int in a catalytically active state and covalently bound to DNA. This would allow us to learn what drives tetramerization and how the opened DNA at the junction changes upon cleavage. I started working in this direction and discussed possible ways to continue these efforts in section 2.1.1.5.

In the second part of my work (section 2.2.4), I was able to demonstrate that full-length Int binds to arm DNA sites and forms stable complexes with circular intermediate and HJ DNA. These results open doors for further crystallization and cryoEM-based structure analysis to provide more insights into transposition regulation in the future. Information about the AB domain – arm DNA interface will help to elucidate its specific role in guiding the reaction's directionality, like in the λ-bacteriophage system. New advances in single-particle cryoEM are making it possible to solve structures of smaller sizes. A dimer of Int^{82N} with its suicide DNA substrate has a size of around 90 kDa, which may still be too small, but complexes that include the full-length protein with CI or HJ DNA and the arm sites in a tetrameric state would have a suitable size of more than 200 kDa. Another way to trap bigger complexes for single-particle cryoEM experiments would be to use longer DNAs that contain core and arm sites in one molecule. By using such DNA substrates, it would be possible to trap the full-length integrase with excisionase and maybe other accessory factors like HU and form complete transpososome complexes. Such a large transpososome

complex was structurally characterized by cryo-EM previously for the λ-bacteriophage system (Laxmikanthan *et al.*, 2016).

Furthermore, it would be great to improve the resolution of the Int-HJ structure to enable characterization of specific protein-DNA interactions and determine protein residues that may play a significant role in HJ formation and resolution. For this, further refinement trials should be performed with the merged dataset to improve the Int-HJ structure's resolution and electron density map. Another approach could also be using single-particle cryoEM, which should be feasible at the 190 kDa size of the tetrameric assembly.

It would be also of great interest to find conditions or Int mutants that lead to trapping HJ-DNA intermediates. Such an assay is missing, and HJ formation could not be demonstrated *in vitro* for any CTn systems to date. Further work on the HJ_{long} resolution experiments will also help elucidate the role of essential protein residues that may play a role in the last step of the recombination reaction. This assay will also help to clarify the role of Xis in integration, to characterize the directionality of integration in more detail and test the function of host encoded proteins, such as HU.

Concerning novobiocin binding and action, biophysical characterization should be continued to find the reasons for its inhibitory role in different steps of the transposition reaction. In particular, analyzing Int-DNA complexes together with novobiocin will be required. It will also be interesting to test other compounds of the aminocoumarin family using *in vivo* and *in vitro* assays and examine their possible influence on excision and integration. An integration assay *in vivo* has been previously set up in the lab (Lambertsen *et al.*, 2018) and could be used to monitor and validate future drug candidates.

3.4 Conclusions

This thesis describes the results obtained during my Ph.D. work on the proteins Int and Xis. These proteins are responsible for performing the excision and integration reactions of the vancomycin resistance carrying conjugative transposon Tn*1549*, initially found in *E. faecalis*. Conjugative transposons can transfer ABR genes because they can move across bacterial cells and species and carry ABR cargos within their sequence. Tn*1549* is responsible for propagating resistance to vancomycin, a last-resort antibiotic, in a wide range of intestinal bacteria.

My work focused on investigating the Int enzyme, a member of the broad tyrosine recombinase family, and its role in transposon integration. Int is responsible for performing the excision and integration reactions of this MGE, and its mechanism seems to be

conserved among diverse CTn elements that use homologous tyrosine recombinases in the bacterial kingdom. Therefore, Int represents an exciting research target that may enable the development of CTn inhibitors as a new strategy to limit ABR spreading among bacteria.

In this study, the crystal structure of the Int recombinase from Tn*1549* in complex with an integrative Holliday junction DNA intermediate was determined. As far as we know, this represents the first structure of any CTn integrase in complex with HJ DNA. The structure offers unique insights into the mechanism of conjugative transposition of Tn*1549* and related MGEs in general. Moreover, this structure confirmed that CTn integrases form stable tetramer complexes in a fairly similar manner to the well-known site-specific recombinases λ Int, Cre, and XerH.

In the second part of my work, I validated the Int-HJ structure by showing that Int can resolve HJ-DNA *in vitro*. This reaction leads to DNA products containing up to 3 nt long heterologous overlaps, highlighting Int's ability to insert its cargo DNA in a promiscuous manner. My data indicates that a preferred directionality towards products (meaning integration) may be established when all necessary factors are present in the correct configuration, including the accessory factor Xis. I solved the crystal structure of Xis, which showed the conserved winged-helix fold typical for this protein family. The DNA binding sites of Tn*1549* Xis are not conserved with Xis from Tn*916*, although their DNA-binding interface differs only in one critical residue, indicating that these proteins employ an indirect sequence readout.

In the third part of my work, I showed that novobiocin, a Tn*1549* transposition-inhibitor candidate, can hinder Int tetramerization on HJ DNA and HJ resolution *in vitro*, highlighting the importance of the tetrameric state during the transposition reaction.

In conclusion, my work has brought new insights into various aspects of the mechanism of conjugative transposition and I hope that these insights may help develop drug candidates for combating antibiotic resistance spread among bacteria in the future.

4 MATERIALS AND METHODS

4.1 Materials

4.1.1 Chemicals and reagents

Unless otherwise provided, all chemicals used in this thesis were acquired from Sigma-Aldrich, Merck Millipore or Molecular Dimensions. All restriction enzymes were supplied by New England Biolabs (NEB) apart from DpnI, which was purchased from Thermo Scientific. NEB and Thermo Scientific provided enzymes and reagents used for molecular biology as stated in the text. Antibiotics were purchased from Sigma-Aldrich and Carl Roth. All buffers and stocks were made using deionized, distilled water and sterilized by autoclaving or filtering and degassed when required.

4.1.2 Bacterial growth media

All media for bacterial growth used in this study were prepared and distributed by EMBL Media Kitchen Facility. Liquid *E. coli* cultures were grown in Luria-Bertani (LB) medium containing 10 g tryptone, 5 g yeast extract, and 5 g sodium chloride per liter (pH 7.2) or in Super Optimal Broth (SOB or SOC) medium containing 20 g tryptone, 5 g yeast extract, 0.585 g sodium chloride and 0.186 g potassium chloride per liter (pH 7.2). Depending on the bacterial strain and plasmid(s), the following antibiotics were supplied to the growth media: kanamycin (kanamycin sulphate, 50 μg/ml), ampicillin (ampicillin sodium salt, 100 μg/ml) and chloramphenicol (33 μg/ml). Bacterial cell growth was monitored through optical density measurement at the wavelength of 600 nm (OD_{600}).

4.1.3 Bacterial strains

The *E. coli* strains used in this study were the following:

E. coli strain XL10-Gold - Genotype: Tet^R Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F' proAB lacIqZDM15 Tn10 (Tet^R) Amy Cm^R] (Stratagene/Novagen).

E. coli strain BL21 (DE3) – Genotype: F- *ompT gal dcm lon hsdSB*(*rB- mB-*) λ(DE3) (Stratagene/Novagen).

E. coli strain BL21 (DE3) pLysS – Genotype: F- *ompT gal dcm lon hsdSB*(*rB- mB-*) λ(DE3) pLysS (Cm^R) (Stratagene/Novagen).

E. coli strain Top10 – Genotype: F- *mcrA* Δ(*mrr-hsdRMS-mcrBC*) φ80lacZΔM15 Δ*lacX74 recA1 araD139* Δ(ara-leu)7697 *galU galK rpsL* (StrR) *endA1 nupG*. The strain was supplied by EMBL Protein Expression and Purification Core (PEP-core) Facility.

All *E. coli* strains were maintained as glycerol stocks in 35% glycerol (final concentration) and stored at -80 °C.

4.1.4 Plasmids

Table 4-1 contains a list of plasmids used in this study together with the description, carried antibiotic resistance, origin of replication, and source of each plasmid. All plasmids were purified with the GeneElute Plasmid DNA Miniprep Kit (Sigma-Aldrich) using the manufacturer's protocol. Plasmid DNA was stored in deionized, distilled water at -20 °C.

Table 4-1: Plasmids used in this study.

Plasmid	Description	Resistance*	Origin	Source
pETM22	T7 expression vector with N-terminal 6xHis-TRX tag	Km	pBR322	PEP-core EMBL Heidelberg
pETM28	T7 expression vector with N-terminal 6xHis-SUMO tag	Km	pBR322	PEP-core EMBL Heidelberg
pETM28-Int^{FL}	T7 expression vector encoding 6xHis-SUMO-Int^{FL}	Km	pBR322	(Rubio-Cosials *et al.*, 2018)
pETM28-Int^{FL} (R225K)	T7 expression vector encoding 6xHis-SUMO-Int^{FL} R225K	Km	pBR322	(Rubio-Cosials *et al.*, 2018)
pETM28-Int^{FL} (Y379F)	T7 expression vector encoding 6xHis-SUMO-Int^{FL} Y379F	Km	pBR322	(Rubio-Cosials *et al.*, 2018)
pETM28-Int^{FL} (Y380F)	T7 expression vector encoding 6xHis-SUMO-Int^{FL} Y380F	Km	pBR322	(Rubio-Cosials *et al.*, 2018)
pETM28-Int^{FL} (Y379F-Y380F)	T7 expression vector encoding 6xHis-SUMO-Int^{FL} Y2F	Km	pBR322	(Rubio-Cosials *et al.*, 2018)
pETM28-Int^{FL} (390C)	T7 expression vector encoding 6xHis-SUMO-Int^{FL} 390C	Km	pBR322	(Rubio-Cosials *et al.*, 2018)
pETM28-Int^{82N}	T7 expression vector encoding 6xHis-SUMO-Int^{82N}	Km	pBR322	(Rubio-Cosials *et al.*, 2018)
pETM28-Int^{82N} (R225K)	T7 expression vector encoding 6xHis-SUMO-Int^{82N} R225K	Km	pBR322	(Rubio-Cosials *et al.*, 2018)
pETM28-Int^{82N} (390C)	T7 expression vector encoding 6xHis-SUMO-Int^{82N} 390C	Km	pBR322	(Rubio-Cosials *et al.*, 2018)
pETM28-Int^{82N} (390C-R225K)	T7 expression vector encoding 6xHis-SUMO-Int^{82N}390C-R225K	Km	pBR322	This study
pETM28-Int^{82N} (390C-Y379F-Y380F)	T7 expression vector encoding 6xHis-SUMO-Int^{82N} 390C-Y2F	Km	pBR322	This study
pETM28-Xis	T7 expression vector encoding 6xHis-SUMO-Xis	Km	pBR322	Barabas Lab (unpublished)
pBAD-XIS-INT	Protein expression plasmid (PEP) containing *xis* and *int* cloned from Tn*1549* in *E. faecalis* 268-10. Plate always on LB with 0.2% glucose.	Amp	p15A	(Lambertsen *et al.*, 2018)
Donor plasmid with wt mini-Tn*1549* (DP) #105	Mini-Tn*1549* consist of the left and right end of Tn*1549* from *E. faecalis* 268-10 inserted on each side of a Cm resistance gene (cat). Flanking DNA from *Enterococcus* is inserted with the transposons ends. The mini-Tn*1549* and the flanking DNA is placed on a donor plasmid.	Cm	ColE1	(Lambertsen *et al.*, 2018)
Donor plasmid with wt mini-Tn*1549* (DP) #437	Inserted Gm resistance on backbone of mini-Tn*1549* donor plasmid	Cm, Gm	ColE1	(Lambertsen *et al.*, 2018)
pETM28_CI	Plasmid containing the circular intermediate (CI) sequence of Tn*1549*, including transposon ends and internal arm binding sites. LE – 216 bp, RE – 219 bp. Crossover region from the LF: cccttt	Km	pBR322	This study
pETM28_LF-LE_AGG	Plasmid containing the left flank (LF) and transposon LE sequences of Tn*1549*. LE – 216 bp, LF – 500 bp. LF sequence comes from left part of a specific integration site from the pOX-plasmid (TTACCC). Crossover region from the LF: cccAGG, is mutated for HJ_{long} formation.	Km	pBR322	This study
pETM28_RE-RF_ATG	Plasmid containing the right flank (RF) and transposon RE sequences of Tn*1549*. RE – 223 bp, RF – 500 bp. RF sequence comes from right part of a specific integration site from the pOX plasmid (TTACCC). Crossover region from the RF: ATGttt, is mutated for HJ_{long} formation.	Km	pBR322	This study
pOX38-Km	F-based conjugative plasmid used for conjugation-based integration assay. In this study, it was used only for cloning of LF and RF.	Km	F	(Lambertsen *et al.*, 2018)

* Km-kanamycin; Cm-chloramphenicol; Amp-ampicillin, Gm-gentamicin

4.1.5 Oligonucleotides and primers

Oligonucleotides used as DNA substrates in biochemical assays, crystallization and binding studies were purchased from Integrated DNA Technologies (IDT, Leuven, Belgium) and were resuspended in Tris-EDTA (TE) buffer [10 mM tris(hydroxymethyl)aminomethane (Tris), pH 8, and 1 mM ethylenediaminetetraacetic acid (EDTA)] to a final concentration of 100 µM or 1 mM. They are shown in the results sections for individual experiments.

Oligonucleotides used as primers in polymerase chain reaction (PCR) are listed in Table 4-2. Melting temperature of the primers was estimated using the IDT OligoAnalyzer tool (http://eu.idtdna.com/calc/analyzer) or the restriction free cloning tool (https://www.rf-cloning.org/index.php), depending on the cloning method. Primers were resuspended in distilled water. Oligonucleotides and primers were purchased with standard desalting and without modifications unless otherwise stated.

*Table 4-2: Primers for PCR reactions used in this study. * Primers are from (Lambertsen et al., 2018).*

No	Primer	Sequence	Purpose
1	IntR225K_390Cstop_for	/Phos/CACAGTTCGCTAATCTTCAGACCGGTGCCCA	Site-directed mutagenesis of Int^{82N}390C-R225K
2	Int2YF_390Cstop_for	/Phos/GCAATATTACCATGACCCTGAACTTTTTTGCCCATGCAACCTTTGATAGCGCACG	Site-directed mutagenesis of Int^{82N}390C-2YF
3	P2_for*	GCGGGATCCTGTTCTCCCAT	Test for CI of mini-Tn1549
4	P3_rev*	ACGCAAGCTTCGATTCCGCAAG	Test for CI of mini-Tn1549
5	P46_for*	GTCACGCTCGTCGTTTGGTATGGC	Test for presence of mini-Tn1549 in donor plasmid
6	P47_rev*	GCGCCGACATCATAACGGTTCTGG	Test for presence of mini-Tn1549 in donor plasmid
7	CI1_for	GCAGCAGCCATCATCATCATCATCACAGGCGGGATCCTGTTCTCCC	Cloning of CI sequence into pETM28_CI
8	CI1_rev	ACCTTCAGGTTGATGTGGTCATTCATATGGACGCAAGCTTCGATTCCG	Cloning of CI sequence into pETM28_CI
9	CI1_TL_for	GGCTTTCAGCGATAACCTAAAATTTTCCCTTTAAAATTTTGTGGGCTGTTCATTTAC	Cloning of LF sequence into pETM28_ LF-LE
10	CI1_TL_rev	CCACCTTCAGGTTGATGTGGTCATTCATATGGCCCGTTACCCTTCCTG	Cloning of LF sequence into pETM28_LF-LE
11	CI1_TR_for	GCAGCAGCCATCATCATCATCATCACAGGTGACGTAACGGGTGATACC	Cloning of RF sequence into pETM28_RE-RF
12	CI1_TR_rev	TCAAGAACTTCTAAAAAGATAATCTAAAATCCCATATAATTTTAAAGGGAAAACCCGGTAAAAAATTTTGCG	Cloning of RF sequence into pETM28_RE-RF
13	HJ1c_long_AGG	/Phos/GCGATAACCTAAAATTTTCCCAGGAAAATTTTGTGGGCTGTTC	Site-directed mutagenesis of pETM28_LF-LE-AGG
14	HJ1a_long_ATG	/Phos/CAAAATTTTTTACCGGGTTTTATGTTTAAAATTATATGGGATTTTAG	Site-directed mutagenesis of pETM28_RE-RF-ATG
15	HJ1a_long_forP	/Phos/GACACACCTGTCCTGG	Amplification of RE-RF from pETM28_RE-RF-ATG
16	HJ1a_long_rev	AGAAATGGAACGGCTGG	Amplification of RE-RF from pETM28_RE-RF-ATG
17	HJ1b_long_for	TATGGTTCAGCATATATGCG	Amplification of CI region from pETM28_CI
18	HJ1b_long_revP	/Phos/AGAAATGGAACGGCTGG	Amplification of CI region from pETM28_CI
19	HJ1c_long_forP	/Phos/TATGGTTCAGCATATATGCG	Amplification of LF-LE region from pETM28_LF-LE-AGG
20	HJ1c_long_rev	AACCTGCCGCACTGA	Amplification of LF-LE region from pETM28_LF-LE-AGG
21	HJ1a2_long_forP	/Phos/TGACACACCTGTCCT	Amplification of RE-RF from pETM28_RE-RF-ATG
22	HJ1a3_long_forP	/Phos/ACACCTGTCCTGGGCAGAA	Amplification of RE-RF from pETM28_RE-RF-ATG
23	HJ1a2_long_rev	AAGGGCAGAAATGGAACG	Amplification of RE-RF from pETM28_RE-RF-ATG
24	HJ1b2_long_revP	/Phos/AAGGGCAGAAATGGAACG	Amplification of CI region from pETM28_CI

4.2 Molecular Biology Methods

4.2.1 Constructs for protein overexpression

The open reading frames for *xis*- and *int*-Tn*1549* genes (NCBI accession numbers AAF72367.1 and AAF72368.1 respectively) have been previously identified through sequencing (Garnier *et al.*, 2000).

The full-length *int*- and *xis*-Tn*1549* genes from *E. faecalis* were codon-optimized for expression in *E. coli* and synthetized by GeneArt (Thermo Fisher). The int^{FL}-, int^{82N}- and *xis*-Tn*1549* genes were then cloned into expression vector pETM28 by restriction cloning, using BamHI/XhoI restriction sites, for overexpression in *E. coli* prior to this thesis work, to give vectors pETM28-Int^{FL}, pETM28-Int^{82N} (Rubio-Cosials *et al.*, 2018) and pETM28-Xis (unpublished).

4.2.2 Polymerase chain reaction (PCR)

PCR amplifications needed for different purposes were performed using primers in Table 4-2. Primers were resuspended in distilled water to a final concentration of 100 µM and stored at -20 °C. For all reactions, primer stocks of 10 µM were used. The PCR mix consisted of 1x Phusion High-Fidelity Buffer (NEB), 2 units of Phusion® High-Fidelity DNA Polymerase (NEB), 200 µM dNTPs (Bioline), 200 µM of each forward and reverse primer, and 20-150 ng of template DNA (either plasmid or genomic) in a final reaction volume of 50 µl. The reaction conditions are shown in Table 4-3. The reaction products correctness was checked through agarose gel electrophoresis. Otherwise the products were purified using the GenElute PCR Clean-Up Kit (Sigma-Aldrich) following the manufacturer's instructions, or through agarose gel electrophoresis and gel extraction, when further cloning steps were needed.

Table 4-3: Thermocycling conditions for PCR.

Step	Temperature	Time
1. Initial Denaturation	98 °C	30 seconds
35 cycles (step 2-4):		
2. Denaturation	98 °C	5 seconds
3. Annealing	50-72 °C	5 seconds
4. Extension	72 °C	30 seconds/kbp
5. Final extension	72 °C	10 minutes
6. Hold	4 °C	hold

4.2.3 Restriction-free (RF) cloning

The primers for restriction-free (RF) cloning were designed in a way, that they contain complementary sequences to the desired insert and target plasmid (vector). They were

generated using the primer design tool at www.rf-cloning.org. All DNA concentrations were determined by UV spectroscopy at 260 nm using a NanoDrop instrument (Thermo Scientific).

The first round of PCR was set up as stated in Table 4-4. The thermocycling conditions are shown in Table 4-5. The product, so-called MegaPrimer, was purified using the GenElute PCR Clean-Up Kit (Sigma-Aldrich) following the manufacturer's instructions. The purified PCR product was used as primer and the target plasmid as template in a second PCR reaction, described in Table 4-6. The thermocycling conditions for the second reaction are explained in Table 4-7. The product was purified using the GenElute PCR Clean-Up Kit (Sigma-Aldrich) and eluted in 17 µl distilled water. Template plasmid in the eluate was digested with 2 µl of 10x FastDigest Buffer and 2 µl of FastDigest DpnI restriction enzyme (both from Thermo Scientific) at 37 °C for at least 3 hours. 5 µl of the sample were used for transformation of *E. coli* XL10-Gold chemically competent cells (Stratagene).

Table 4-4: PCR set up 1. Volumes and final concentration of each component used for a PCR 1.

Component	50 µl Reaction	Final Concentration
Distilled water	to 50 µl	-
2x Phusion Flash High-Fidelity PCR Master Mix (Thermo Scientific)	25 µl	1x
10 µM Forward primer	2.5 µl	500 nM
10 µM Reverse primer	2.5 µl	500 nM
Template DNA (20 ng)	0.5 µl (variable)	-

Table 4-5: Thermocycling conditions in restriction-free cloning PCR 1.

Step	Temperature	Time
1. Initial Denaturation	98 °C	20 seconds
35 cycles (step 2-4):		
2. Denaturation	98 °C	1 second
3. Annealing	50-60 °C	5 seconds
4. Extension	72 °C	30 seconds/kbp
5. Final extension	72 °C	5 minutes
6. Hold	4 °C	hold

Table 4-6: PCR set up 2. Volumes and final concentration of each component used for a PCR 2.

Component	50 µl Reaction	Final Concentration
Distilled water	to 50 µl	-
2x Phusion Flash High-Fidelity PCR Master Mix (Thermo Scientific)	25 µl	1x
MegaPrimer (350 ng)	2.5 µl (variable)	-
Destination vector (50 ng)	1 µl (variable)	-

Table 4-7: Thermocycling conditions in restriction-free cloning PCR 2.

Step	Temperature	Time
1. Initial Denaturation	98 °C	20 seconds
18 cycles (step 2-4):		
2. Denaturation	98 °C	1 second
3. Annealing	60-65 °C	5 seconds
4. Extension	72 °C	1 min/kbp
5. Final extension	72 °C	10 minutes
6. Hold	4 °C	hold

Usually 2 to 3 colonies were picked, inoculated into 5 ml LB medium containing appropriate antibiotics and grown overnight at 37 °C. The day after, the bacterial cells were harvested, and the plasmid DNA was extracted using GenElute Plasmid DNA Miniprep Kit (Sigma-Aldrich) and sent for DNA sequencing (performed by GATC Biotech, later Eurofins) for validation of positive clones.

4.2.4 Site-directed mutagenesis

The site-directed mutagenesis method or loop-in protocol was performed in order to generate different point mutants of the different proteins used in this study. For that a single 5' phosphorylated primer was used to amplify the vector of interest with the desired point mutation in the middle. The primer's melting temperature (T_M) was calculated using the IDT OligoAnalyzer tool (http://eu.idtdna.com/calc/analyzer). The primer's T_M should be around 58 °C at each end and lay around of 68 °C for the full sequence. The mismatch should be surrounded by 10 – 25 nucleotides on either side and the 3'-end should be a G or a C. The PCR mutagenesis reaction was performed as described in Table 4-8. The thermocycling conditions are shown in Table 4-9. The amplification product is a single-stranded DNA plasmid that contains the desired mutation and was sealed via ligation by Taq DNA ligase in the same reaction. The PCR product was purified using the GenElute PCR Clean-Up Kit (Sigma-Aldrich) and eluted in 17 µl distilled water. Template plasmid in the eluate was digested with 2 µl of 10x FastDigest Buffer and 2 µl of FastDigest DpnI restriction enzyme (both from Thermo Scientific) at 37 °C for at least 3 hours. 5 µl of the sample were used for transformation of *E. coli* XL10-Gold chemically competent cells (Stratagene). Sequencing by GATC Biotech was performed to confirm correct clones.

Table 4-8: PCR set up for site-directed mutagenesis.

Component	50 µl Reaction	Final Concentration
Distilled water	to 50 µl	-
5x HF DNA Polymerase Buffer with $MgCl_2$	10 µl	1x
10x Taq DNA Ligase Buffer	5 µl	1x
dNTPs (10 mM)	1 µl	200 µM
Template DNA (100 – 300 ng)	1 µl (variable)	-
Phusion® HF DNA Polymerase (2 units)	1 µl	-
Taq DNA Ligase (1 unit)	1 µl	-
5'-P-primer (10 µM)	2.5 µl	500 nM

Table 4-9: Thermocycling conditions in site-directed mutagenesis.

Step	Temperature	Time
1. Initial Denaturation	95 °C	1 min
30 cycles (step 2-4):		
2. Denaturation	95 °C	30 seconds
3. Annealing	55 °C	30 seconds
4. Extension	65 °C	1min/kbp
5. Hold	4 °C	hold

4.2.5 Agarose gel electrophoresis

PCRs and digested products were analyzed on 0.5% - 1.2% (w/v) agarose gels in 1x Tris-acetate-EDTA (TAE) buffer (40 mM tris, 20 mM acetic acid, and 1 mM EDTA). The samples were loaded in 1x GelPilot Loading Dye (QIAGEN). 5 µl of DNA marker (HyperLadder™ 1kb, Bioline) were loaded on the gel. Gels were run with 100 V for 30 min – 1 hour. The DNA was visualized by UV-light (UV transilluminator, Alpha Innotech) with ethidium bromide staining. The gels were documented by a gel documentary system (AlphaImager® HP software, Fischer Scientific).

When necessary, the bands of interest were extracted from the gel and purified using a GenElute DNA Gel Extraction Kit (Sigma-Aldrich) following the manufacturer's instructions. Purified DNA was either directly used for further cloning steps or stored at -20 °C.

4.2.6 Transformation of competent cells

4.2.6.1 Electro-competent cells

50 µl of frozen electro-competent *E. coli* cells were thawed on ice and mixed with up to 5 µl of Gibson-cloning mix. The cells were then transferred into a 0.1 cm electroporation cuvette (Bio-Rad) and eletroporated using MicroPulser™ Electroporator (Bio-Rad). 300 µl of SOC medium were added to the eletroporated cells and the mix was transferred into an eppendorf tube. Cells were incubated at 37 °C for 1 hour with shaking. Afterwards all the mix was plated on LB-agar plates containing appropriate antibiotics and incubated at 37 °C overnight.

4.2.6.2 Chemically competent cells

50 µl frozen chemically competent *E. coli* cells were thawed on ice and mixed with 1 µl of plasmid DNA or 5 µl PCR-mix. The cells were kept on ice for 15 min, followed by heat-shock incubation at 42 °C for 45 sec. After heat shock samples were put on ice for 7 min, mixed with 300 µl of SOC medium and incubated at 37 °C for 1 hour with shaking. Afterwards transformed cells were plated on LB-agar plates containing appropriate antibiotics and incubated at 37 °C overnight.

4.2.7 Plasmid DNA extraction

E. coli cells were grown in 5 ml LB medium (for MINI-prep) or 50 mL LB medium (for MIDI-prep) with appropriate antibiotics at 37 °C overnight. Cells were harvested by

centrifugation for 10 min at 3000 g at 4 °C. Plasmid DNA was purified using the GeneElute Plasmid DNA Miniprep Kit (Sigma-Aldrich) or the GeneElute Plasmid DNA Midiprep Kit (Sigma-Aldrich) using the manufacturer's protocol. Plasmid DNA was eluted and stored in deionized, distilled water at -20 °C.

4.2.8 DNA sequencing

DNA sequencing performed by GATC Biotech, later Eurofins, validated plasmids used in this study. For that, 20 µl DNA (min 1000 ng) of each sample was sent. The sequences were inspected using ApE – A plasmid Editor, Blast and the ExPASy (https://web.expasy.org/translate/) translator tool.

4.2.9 Protein overexpression and purification

4.2.9.1 Protein overexpression in E. coli

All protein constructs used in this study were overexpressed in *E. coli* strains BL21 (DE3) or BL21 (DE3) pLysS from pET vectors under a T7 promoter. The expression vectors were transformed into chemically competent cells and plated on LB-agar plates containing appropriate antibiotics and incubated at 37 °C overnight. Fresh transformants or glycerol stocks were used to inoculate pre-cultures (usually 80 ml) of LB medium with appropriate antibiotics and incubated at 37 °C overnight with shaking (200 rpm). For large-scale expression, 500 ml of LB medium with appropriate antibiotics was inoculated with 10 ml pre-culture and grown at 37 °C and 180 rpm until an OD_{600} of 0.6-0.8 was reached. Afterwards the temperature was reduced to 18 °C and after 20 min (or a final cell-duplication) expression was induced by addition of 1 mM (final concentration) isopropyl β-D-1- thiogalactopyranoside (IPTG). Cells were grown at 18 °C and 180 rpm for 18 hours and then harvested by centrifugation (4000 g, 30 min, 4 °C) the next day. The obtained pellet was then resuspended and washed once with 25 ml 1x phosphate buffered saline (PBS) and pelleted at 4000g for 30-60 min at 4 °C. Afterwards the pellet was flash-frozen in liquid nitrogen and stored at -80 °C until further use or immediately used for purification. To assess proper expression, samples before addition of IPTG and before pelleting were collected and kept for analysis by SDS-PAGE (see 4.2.10).

4.2.9.2 Protein purification of Int and Xis Tn1549 protein constructs

All Int and Xis Tn*1549* constructs described in this study were overexpressed as fusion proteins with an N-terminal 6xHis-SUMO affinity and solubility tag. They were purified through a three-step purification scheme, including 1) first purification by nickel affinity

chromatography, 2) tag cleavage and removal by second nickel affinity chromatography, and 3) size exclusion chromatography (SEC) on a gel filtration column. All purification steps were performed using the ÄKTApurifier Protein Purification System (GE Healthcare) at 4 °C. The chromatogram was collected at 280 nm (mainly protein absorption) and 260 nm (mainly DNA absorption). The system was operated, and the chromatograms collected and analyzed using UNICORN Control Software (GE Healthcare).

4.2.9.2.1 Sample preparation

The cell pellets were thawed one ice for 30 min – 1 hour and resuspended in 45 ml Lysis buffer (Table 4-10). The resuspended cells were then lysed by sonication using a Branson Sonifier 250 set to 50% output control, in cycles of 5 seconds sonication and 15 seconds rest on ice for a total active sonication time of 3 min 30 seconds. The crude lysate was clarified by centrifugation (30000 g, 30 min, 4°C). The supernatant was filtered through 0.22 µm disposable filter and loaded onto a 5 ml HisTrap HP column (GE Healthcare), which was previously equilibrated in Loading buffer (Table 4-10) for the first affinity purification.

Table 4-10: Composition of purification buffers for Tn1549 Int and Xis constructs.

Buffer	Composition
Purification of all Tn*1549* Int constructs	
Lysis buffer	50 mM Hepes, 750 mM NaCl, 50 mM imidazole, 5% Glycerol, 1 mM DTT, 1 tablet of cOmplete Protease Inhibitor Cocktail (Roche), 50 µg/ml RNaseA (Roche), 50 µg/ml DNaseI (Roche), pH 7.5
Loading buffer A	50 mM HEPES, 750 mM NaCl, 50 mM imidazole, 5% Glycerol, 1mM DTT, pH 7.5
Elution buffer B	50 mM HEPES, 750 mM NaCl, 500 mM imidazole, 5% Glycerol, 1mM DTT, pH 7.5
Dialysis buffer	50 mM HEPES, 750 mM NaCl, 5% Glycerol, 5 mM DTT, pH 7.5
Gel filtration buffer 1	50 mM HEPES, 750 mM NaCl, 5% Glycerol, 5 mM DTT, pH 7.5
Gel filtration buffer 2	50 mM HEPES, 250 mM NaCl, 10% Glycerol, 5 mM DTT, pH 7.5
Purification of Tn*1549* Xis construct	
Lysis buffer	50 mM BisTris, 750 mM NaCl, 50 mM imidazole, 1 mM DTT, 1 tablet of cOmplete Protease Inhibitor Cocktail (Roche), 50 µg/ml RNaseA (Roche), 50 µg/ml DNaseI (Roche), pH 6.0
Loading buffer	50 mM BisTris, 750 mM NaCl, 50 mM imidazole, 1 mM DTT, pH 6.0
Elution buffer	50 mM BisTris, 750 mM NaCl, 500 mM imidazole, 1 mM DTT, pH 6.0
Dialysis buffer	50 mM BisTris, 750 mM NaCl, 5 mM DTT, pH 6.0
Gel filtration buffer	50 mM BisTris, 750 mM NaCl, 5 mM DTT, pH 6.0

4.2.9.2.2 First HisTrap purification

For the first HisTrap purification, a 5 mL HisTrap HP column (GE Healthcare) was used. The column was prewashed with 5 column volumes (CV) of distilled water, 5 CV of buffer A (Loading buffer, Table 4-10), 5 CV of buffer B (Elution buffer, Table 4-10) and finally equilibrated back with 5 CV of buffer A prior injection of the filtered supernatant. The protein was injected into the column with a flow rate of 1.5 ml/min. Afterwards the column was washed with 8-12 CV of buffer A or until the absorbance reached the baseline. All

subsequent steps were carried out at a flow rate of 2 ml/min. Bound protein was eluted with buffer B by increasing the imidazole concentration in a gradient manner (10% to 100% Elution buffer, Table 4-10, reach 100% in 20 CV time). Int Tn1549 constructs eluted between 162.5-275 mM imidazole, while Xis Tn1549 protein eluted later, around 400 mM imidazole. Fractions of 2 ml volume each were collected during elution. Afterwards, the column was washed with 5 CV buffer B to remove all remaining impurities. Then, the column was washed with 5 CV buffer A, before a final wash with 5 CV distilled water for storage at 4 °C. Flow-through (FT) fraction after protein injection, wash-fraction (W) and elution fractions of interest were prepared accordingly and analyzed on SDS-PAGE (see 4.2.10) to identify the fractions containing the protein of interest.

For 6xHis-SUMO-tag removal, fractions containing protein of interest were pulled together and the protein concentration measured. Afterwards, SenP2 protease (PEPcore, EMBL-Heidelberg) was added in a 1:100 molar ratio and transferred into pre-washed Spectra/Por® Dialysis Membrane tubing (molecular weight cutoff of 3.5 kDa, Spectrumlabs). The sample was then dialysed against Dialysis buffer (Table 4-10) overnight at 4 °C in order to remove imidazole and allow for protease cleavage.

4.2.9.2.3 Second HisTrap purification

The dialysed protein was injected into the newly equilibrated 5 mL HisTrap column in buffer A (see Table 4-10) with a flow rate of 1.5 ml/min. Afterwards the column was washed with 8-12 CV of buffer A or until the absorbance reached the baseline. All subsequent steps were carried out at a flow rate of 2 ml/min. Elution was performed with buffer B by increasing the imidazole concentration in a gradient manner (10% to 100% Elution buffer, Table 4-10, reach 100% in 20 CV time). Cleaved protein (all Int and Xis Tn1549 constructs without 6xHis-SUMO-tag) eluted during the wash-step (W, 50 mM imidazole) due to not being able to bind strongly to the column anymore. Elution was carried out to check for cleaved 6xHis-SUMO-tag, SenP2 protease and amount of non-cleaved cleaved protein. Afterwards, the column was washed with 5 CV buffer B to remove all remaining impurities. Then, the column was washed with 5 CV buffer A, before a final wash with 5 CV distilled water for storage at 4 °C. For long-term storage the column was put in 20% ethanol after washing with distilled water. FT-, W- and elution fractions of interest were prepared accordingly and analyzed on SDS-PAGE (see 4.2.10).

4.2.9.2.4 Size-exclusion chromatography

The caught W-fraction, that contained the protein of interest, was concentrated for gel filtration using a Vivaspin® Turbo 15 concentrator (molecular weight cutoff of 3.5 kDa for Xis and 10 kDa for Int constructs, Sartorius stedim biotech) by centrifugation at 3500 g at 4 °C until the volume reached approximately 500 µl or a concentration of 10 mg/ml. The gel filtration column (HighLoad™ 10/300 Superdex™ 200, GE Healthcare) was equilibrated with 1.5 CV distilled water followed by 1 CV Gel filtration buffer (-1 for crystallization experiments, -2 for biochemistry and long-term storage, see Table 4-10). The sample was loaded into a 500 µl-sample loop and injected into the column with a flowrate of 0.3 ml/min. The protein was eluted in an isocratic manner with 1 CV Gel filtration buffer 1 (flowrate 0.3 ml/min). With this method proteins are being separated by size and not by binding affinity. Elution was fractionated and samples were collected in 0.5 ml volumes. The fractions were analyzed by SDS-PAGE (see 4.2.10) to detect the protein fractions of interest. The peak fractions were used immediately for crystallization experiments (see 4.4).

When the protein was eluted in Gel filtration buffer 2, the peak fractions were pooled together and concentrated as desired using a Vivaspin® Turbo 15 concentrator (molecular weight cutoff of 3.5 kDa for Xis and 10 kDa for Int constructs, Sartorius stedim biotech), flash-frozen in liquid nitrogen and stored at -80 °C for later use.

4.2.10 Sodium dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was performed to control protein expression, yield and purity. Protein samples were mixed with 2x SDS loading buffer [100 mM Tris-Cl, pH 6.8, 12% glycerol, 4% SDS, 0.01% bromophenol blue, and 0.5 M dithiothreitol (DTT)], heated to 98 °C for 3-5 min and loaded on SDS-PA gels [12% resolving gel: 0.375 M Tris-Cl, pH 8.8, 0.1% SDS, 12% Acrylamide: Bisacrylamide (37.5:1), 0.1% (w/v) ammonium persulfate (APS), 0.16% tetramethylethylenediamine (TEMED); 4% stacking gel: 0.125 M Tris-Cl, pH 8.8, 0.1% SDS, 4% Acrylamide: Bisacrylamide (37.5:1), 0.1% APS, 0.16% TEMED]. General volumes used for SDS-PAGE are described in Table 4-11. Gels were run for 50 min at 180 V in 1x Laemmli buffer (prepared by the EMBL Media Kitchen Facility). When necessary, precast NuPAGE 4-12% Bis-Tris Gels (Invitrogen) were used for electrophoresis according to the manufacturer's instructions. 7 µl protein marker (Mark12™, Thermo Fisher) was

loaded for size determination. The protein gels were washed and heated 2 times with distilled water and then stained with Coomassie staining solution (0.075% (w/v) Coomassie Brilliant Blue G250 (Sigma Aldrich) and 0.1% (v/v) HCl) by heating for 30 sec in a microwave and shaking for 5-15 min. Afterwards the gels were distained by washing and heating in distilled water for 30 sec in a microwave and incubation for 30 min with shaking. The gels were analyzed under white light on a transilluminator (Alpha Innotech) and documented using AlphaImager® HP software (Fischer Scientific).

Table 4-11: Sample volumes used for SDS-PAGE.

Sample name	Sample Volume	2x SDS Loading Buffer	Distilled water	Volume loaded
Before Induction (BI)	cell pellet	35 µl	35 µl	2-3 µl
After Induction (AI)	cell pellet	35 µl	35 µl	2-3 µl
Flow-through (FT)	5 µl	5 µl	-	2-3 µl
Wash-fraction (W)	5 µl	5 µl	-	10 µl
Elution fractions (A1, A2...)	5 µl	5 µl	-	10 µl
Before dialysis (BD)	10 µl	10 µl	-	10 µl
After dialysis (AD)	10 µl	10 µl	-	10 µl
Concentrated protein	2 µl	5 µl	3 µl	10 µl

The molecular weight and theoretical extinction coefficient of the purified proteins was estimated using the ProtParam bioinformatics tool provided by ExPASy (https://www.expasy.org/). Protein samples were quantified based on their UV absorption at 280 nm wavelength measured using a Nanodrop spectrophotometer (Thermo Fisher Scientific).

4.3 Biochemical methods

4.3.1 Annealing of DNA substrates

4.3.1.1 Double-stranded DNA

All double-stranded DNA (dsDNA) substrates used in this study were annealed by mixing oligonucleotides (stocks at 100 µM, 500 µM, or 1 mM concentration, resuspended in TE buffer, see 4.1.5) in equimolar ratios in 1x Annealing buffer (see Table 4-12), followed by incubation at 98 °C for 3 min and step-wise temperature reduction (1 °C/min, from 98-12 °C) in a PCR-machine.

4.3.1.2 Holliday Junction (HJ) DNA

The HJ substrates used in this study were prepared by annealing four oligonucleotides (stocks at 100 µM or 1 mM concentration, resuspended in TE buffer, see 4.1.5) in equimolar ratios in 1x HJ-annealing buffer (see Table 4-12), followed by incubation at 98 °C for 3 min and step-wise temperature reduction (1 °C/min, from 98-12 °C) in a PCR-machine.

HJ_{long} substrates were prepared by mixing equimolar ratios of 1 µM ssDNA strands (see 4.3.2) with synthetic DNA in ultramer quality (IDT, Leuven, Belgium) in excess (2 µM), following the same procedure as with short HJ substrates.

For radiolabeled oligonucleotides, the necessary oligos were first labeled as described in 4.3.3, followed by equimolar mixing and annealing as described above.

Table 4-12: Composition of annealing buffers.

Buffer	Composition
5x Annealing buffer	250 mM NaCl, 1xTE [10 mM tris(hydroxymethyl)aminomethane (Tris), pH 8, and 1 mM ethylenediaminetetraacetic acid (EDTA)]
5 x HJ-annealing buffer	500 mM NaCl, 50 mM $MgCl_2$, 1xTE [10 mM tris(hydroxymethyl)aminomethane (Tris), pH 8, and 1 mM ethylenediaminetetraacetic acid (EDTA)]

4.3.2 Generation of long single-stranded (ss) DNA

Generation of ssDNA was necessary for formation of long HJ (HJ_{long}) substrates that resemble the intermediate of the Tn*1549* integration reaction and contain all Int and prospective Xis binding sites in cis. First, a CI-plasmid was generated (Figure 4-1, 1). For that, the *in vivo* mini-Tn*1549* excision assay (see section 4.5.1) was performed to obtain the mini-transposon in the circular intermediate (CI) form. CI-DNA found in the supernatant fraction from opened bacterial cells (see section 4.5.1), was used as template for cloning part of it into a pETM28 vector by restriction free cloning (4.2.3), using specific primers (Table 4-2). The result was a new plasmid, called pET28M-CI (Table 4-1), containing the CI's core sequences.: the transposon left end (LE or IR_L), a crossover region (CR) of 6 bp length coming from the left flank (LF) site, the transposon right end (RE or IR_R) and prospective Xis-binding sites (X). Afterward, the pET28M-CI plasmid was used as a vector for the creation of two additional plasmids: pET28M-LF-LE and pETM28_RE-RF. For that, specific sequences, flanking an integration site (TTACCC) from the pOX38-Km plasmid, were amplified and inserted into the pET28M-CI plasmid via restriction free cloning (4.2.3). This specific site had shown many integration events *in vivo* (Lambertsen *et al.*, 2018). Finally, by site-specific mutagenesis both plasmids, pET28M-LF-LE and pETM28_RE-RF, were mutated at the crossover region to obtain pET28M-LF-LE_AGG and pETM28_RE-RF_ATG (Figure 4-1, 2 and 3). This was necessary in order to obtain later a fully based-paired and symmetric HJ_{long} substrate at the overlap region.

Table 4-13: Thermocycling conditions for PCR amplification of HJ_{long} strands.

Step	Temperature	Time
1. Initial Denaturation	98 °C	30 seconds
35 cycles (step 2-4):		
2. Denaturation	98 °C	5 seconds
3. Annealing	59 °C	5 seconds
4. Extension	72 °C	10 seconds
5. Final extension	72 °C	3 minutes
6. Hold	4 °C	hold

In order to produce ssDNA, first, the plasmids (Figure 4-1) were used for dsDNA amplification, whereby one strand was amplified using a 5'-phosphorylated primer and the other using a non-modified primer. For the HJ1a_long-strand, primers HJ1a2_long_forP and HJ1a2_long_rev were used; for the HJ1b_long-strand, primers HJ1b_long_for and HJ1b2_long_revP, and for the HJ1c_long-strand, primers HJ1c_long_forP and HJ1c_long_rev. After PCR amplification (4.2.2, Table 4-13), the DNA products were purified using the GenElute PCR Clean-Up Kit (Sigma-Aldrich) following the manufacturer's instructions and their concentrations measured using a NanoDrop instrument (Thermo Scientific). Then, Lambda Exonuclease (NEB) treatment was performed. Lambda Exonuclease degrades preferentially 5'-phosphorylated strands from dsDNA substrates (5' to 3' direction), leading to non-phosphorylated ssDNA products. Whereby, 5'-OH ends are digested 20X slower than the 5'-phosphorylated ends.

Table 4-14: Lambda Exonuclease treatment reaction set up

Component	20 µl Reaction	Final concentration
10x Lambda Exonuclease Reaction Buffer	2 µl	1x
Purified dsDNA substrate (50 – 150 ng/µl)	16 µl	~ 80 ng/µl
Lambda Exonuclease (5 Units/µl)	1 µl	~ 0.25 Units/µl
Distilled water	1 µl	-

The reaction was prepared in a 20 µl reaction mix, as stated in Table 4-14. The samples were incubated at 37 °C for 10 – 15 min depending on the DNA length. Heat-inactivation of the exonuclease was performed by incubation at 75 °C for 10 min. The ssDNA products were checked on an 1% agarose gel (section 4.2.5). ssDNA product, that needed to be 5'-^{32}P-phosphorylated, was treated afterwards as described in section 4.3.3. Finally, HJ_{long} substrate annealing was performed (section 4.3.1.2) and used for HJ_{long} resolution assays (section 4.3.7).

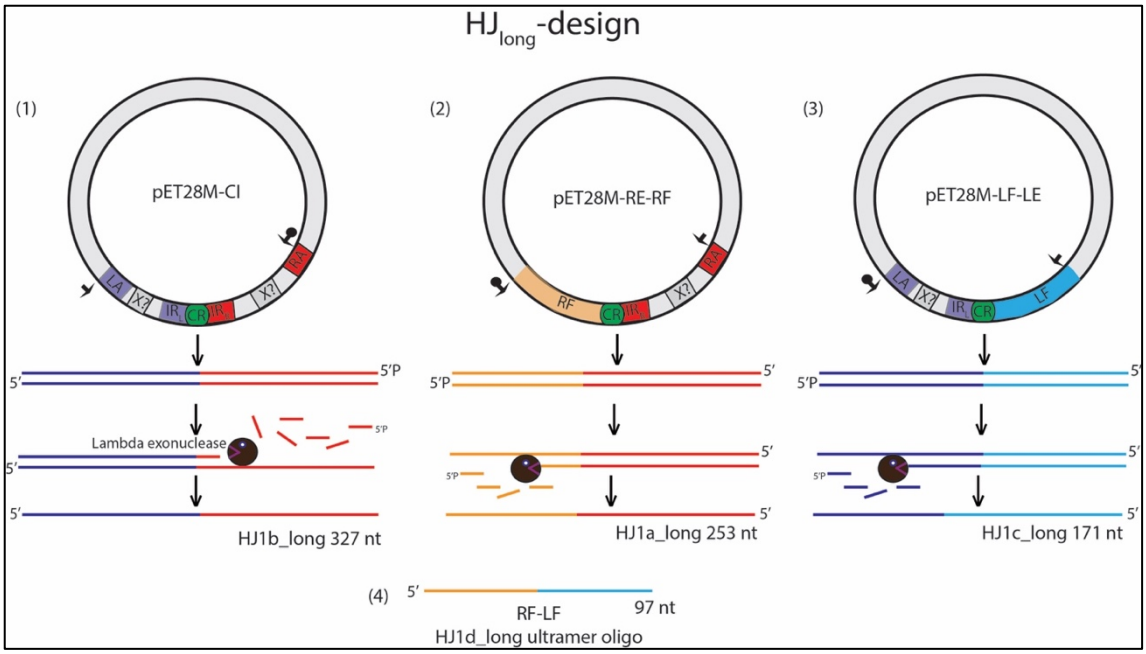

Figure 4-1: Preparation scheme of ssDNA substrates for HJ_{long} formation. Three plasmids were cloned that contained the necessary sequences for HJ_{long} formation of the Tn1549 integration intermediate: circular intermediate sequence (1), right flank and right transposon end sequences (2) and left flank and left transposon end sequences (3). The fourth strand (RF-LF) was ordered as ultramer oligo from IDT, later also the LF-LE strand. For PCR amplification a pair of primers was used, where one primer was 5'-phosphorylated (black arrow with circle) and the other was not (black arrow). Afterward, Lambda exonuclease (black ball) treatment will preferably degrade the 5'-phosphorylated strands, leading to a ssDNA unphosphorylated strand that can be used further for 5'-^{32}P-labeling.

Table 4-15: Ultramer oligos (IDT) used for HJ_{long} formation.

Oligo name	Sequence (5'-3')	Length in nt
HJ1c_long_ CCTGGG	AACCTGCCGCACTGACTGTCCAGGGTAAATGAACAGCCCACAAAATTTTCCTGGGAA AATTTTAGGTTATCGCTGAAAGCCCCGGAAATACGGGCATTTCGCAGGATATGAGAACT CAATTTACTACCAATTTACTACTTTCGGATTGAGCCGCATATATGCTGAACCATA	171
HJ1d_long	ACACCTGTCCTGGGCAGAACGCAAAATTTTTTACCGGGTTTTATGAGGAAAATTTTGTG GGCTGTTCATTTACCCTGGACAGTCAGTGCGGCAGGTT	97

4.3.3 Radioactive labeling of DNA substrates

All oligonucleotides used in labeling reactions were unphosphorylated at the 5'-end. The labeling reaction was prepared in 10 µl reaction mix as stated in Table 4-16. For radiolabeling of ssDNA, 1 µM product was used in 20 µl reaction mix. The reactions were incubated at 37 °C for 1 hour, afterwards 15 µl (for HJ annealing with 2 radiolabeled oligos), 30 µl (for dsDNA or HJ annealing with 1 radiolabeled oligo) or 40 µl (for single-stranded DNA, 10/60 or 20/100 marker) of TE buffer was added. Heat-inactivation of the kinase was performed by incubation at 80 °C for 30 minutes. In order to remove free [γ-^{32}P]-ATP (PerkinElmer), the samples were applied to Micro Bio-Spin® Chromatography Columns (Bio-Rad) and purified following the manufacturer's instructions. Afterwards, the radiolabeled oligos were used for dsDNA or HJ DNA annealing as described in 4.3.1.1 or 4.3.1.2 respectively.

Table 4-16: Radioactive labeling reaction set up.

Component	10 µl Reaction	Final concentration
10x T4 Polynucleotide Kinase Reaction Buffer (NEB)	1 µl	1x
T4 Polynucleotide Kinase (NEB)	1 µl	10 Units
Oligonucleotide (Marker)	2 µl (2.5 µl)	20 µM
[γ-^{32}P]-ATP (PerkinElmer)	2.5 µl	92.5 MBq (2.5 mCi)/ml
Distilled water	3.5 µl (3 µl)	-

The Oligo Length Standards 10/60 and 20/100 Ladder markers (IDT, Leuven, Belgium) were labeled as described above for single oligonucleotides. Afterwards 10/60, 20/100 and single-stranded DNA markers were prepared as described in Table 4-17.

Table 4-17: Radiolabeled marker set up.

Component	Volume	Final concentration
Labeled single-stranded DNA marker/ 10/60 or 20/100 Ladder marker	10 µl/ 40 µl	-
2x Formamide loading buffer [1x TBE (100 mM Tris base, 100 mM boric acid, 2mM EDTA), 90% Formamide, 0.005% xylene cyanol and 0.005% bromophenol blue]	250 µl/ 500 µl	-

4.3.4 Electrophoretic mobility shift assay (EMSA)

The electrophoretic mobility shift assay (EMSA) or DNA binding assay was performed in 1x Binding buffer (see Table 4-18). The dsDNA and HJ-DNA concentrations used for each assay, but also protein concentrations varied, and they are described in the appropriate results section. The reactions were incubated for 30 min at 4 °C and loaded onto a 6 % native gel without addition of a loading dye. The composition of a 6 % native Tris-borate-EDTA (TBE) gel is shown in Table 4-19. The gels were pre-run for 60 min at 100 V at room temperature (RT). After sample loading the gels were run for 40-45 min at 100 V at RT. The gels were stained with 1x SYBR® Gold Nucleic Acid Gel Stain (Life technologies) for 15 min, visualized in the UV transilluminator (Alpha Innotech) and documented using AlphaImager® HP software (Fischer Scientific).

For EMSAs performed using radiolabeled DNA, reactions were treated as described above. Only the imaging step differed, for that gels were imaged with a Typhoon FLA 7000 phosphoimager (GE Healthcare).

Table 4-18: Composition of various buffers for binding and/or activity assays using Xis and Int Tn1549 constructs.

Buffer	Composition
5x Binding Buffer	125 mM HEPES, 625 mM NaCl, 25% glycerol, 5 mM DTT, pH 7.5
5x Binding Buffer - 2	125 mM HEPES, 625 mM NaCl, 25% glycerol, 5 mM DTT, 50 mM $MgCl_2$, pH 7.5
5x Activity Buffer	125 mM HEPES, 625 mM NaCl, 25% glycerol, 5 mM DTT, 5 mM EDTA, 50 mM $MgCl_2$, pH 7.5
5x Activity Buffer - 2	125 mM HEPES, 625 mM NaCl, 25% glycerol, 5 mM DTT, 50 mM $MgCl_2$, pH 7.5

Table 4-19: Composition of the 6% native TBE polyacrylamide gel.

Component	Amount per 16 ml
10x TBE buffer	1.6 ml
30% Acrylamide: Bisacrylamide (37.5:1)	3.2 ml
Distilled water	11.2 ml
10% (w/v) APS	160 µl
TEMED	16 µl

4.3.5 Covalent intermediate assay

The covalent intermediate or *in vitro* cleavage assay used in this study is based on (Rubio-Cosials *et al.*, 2018). Int^{82N} and Int^{82N}(390C) Tn*1549* activity *in vitro* was tested with different "suicide" DNA substrates to decide which complex was more suitable for further crystallization experiments, later the assay was used for testing novobiocin's inhibitory effect on Int. The assay design is described in Figure 4-2. The DNA substrates used in the assay contained a nick in top, bottom or both DNA strands. The sequence contained the IR_L and IR_R sites of the circular intermediate (CI), for more detailed information see section 2.1.1.2 in results. The cleavage reaction was mixed as described in Table 4-20. It contained 1x activity buffer (see Table 4-18) in a 20 µl final reaction volume. The reaction mix was incubated for 3-4 hours at 37 °C. Afterwards 2x SDS loading buffer was added to the samples to stop the reaction, heated to 98 °C for 3-5 min and loaded on 12% SDS-PA gels for analysis*. 7 µl protein marker (Mark12™, Thermo Fisher) was loaded for size determination. DNA-free protein** was used as negative control. Protein gels were stained with Coomassie Brilliant Blue G250 (Sigma Aldrich) for detection of DNA-free protein and covalent protein-DNA complexes, and documented as described in 4.2.10. A shift to a higher molecular weight position indicated the presence of protein-DNA covalent intermediate compared to the band of DNA-free protein.

*10-20 µl of each sample was loaded depending on the well size.

**Protein used for negative control was always the same construct as for the protein-DNA reactions.

Table 4-20: Covalent intermediate assay reaction set up for analysis of cleavage activity.

Component	20 µl Reaction	Final concentration
5x Activity Buffer	5 µl	1x
"Suicide" DNA substrate	4 µl	10 µM
Protein	x µl	20 µM
Distilled water	to 20 µl	-
Drug candidate	y µl	(0 – 4 mM)

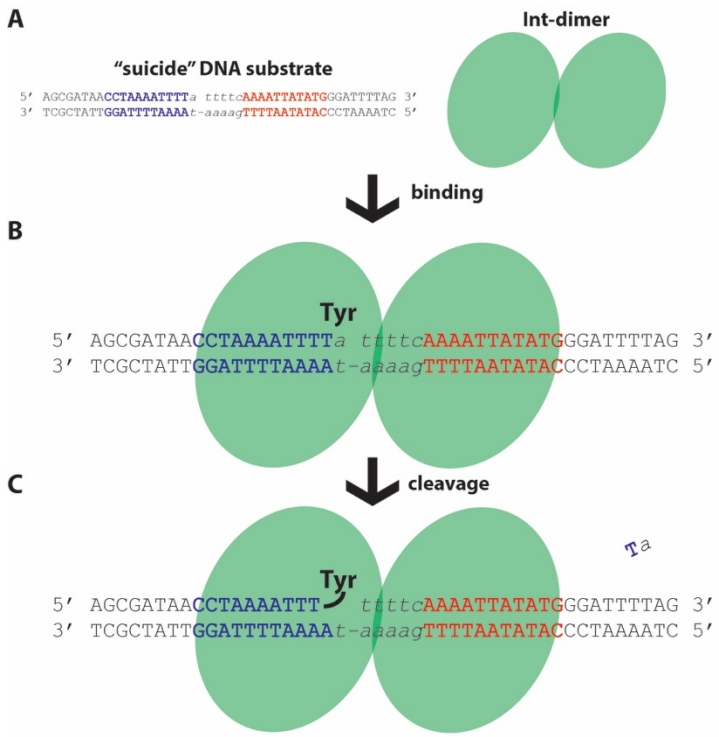

Figure 4-2: Covalent intermediate assay: (A) Int Tn1549 dimer (green ovals) will bind to circular intermediate (CI) DNA composed of IR_L and IR_R sites, divided by a crossover region (CR) in the center of the sequence. This so-called "suicide" DNA substrate has a nick on the IR_L, two nucleotides after the expected cleavage site (T'T). (B) Upon protein binding, the catalytic active residue (Tyr) will be able to cleave at the IR_L (C) leading to a 3'-phosphotyrosine bound DNA and diffusion of the 2 free nucleotides (Ta). By running the protein-DNA bound form on SDS-PA gel, a difference between protein-free and protein-DNA bound form can be observed due to a change on the molecular weight.

4.3.6 DNA cleavage and strand-exchange assay

The *in vitro* DNA cleavage and strand-exchange assay used in this study is based on (Rubio-Cosials *et al.*, 2018). The assay was used for testing inhibitor candidates on Int activity *in vitro*. For that, Int^{82N} and two suicide DNA substrates, based on the CI sequence, were used, where one suicide substrate was 5'-^{32}P-labeled and other was not. After Int^{82N} cleavage of the radiolabeled DNA substrate, a free 5'-hydroxy group will be released that will act as a nucleophile and attack the unlabeled partner, leading to strand exchange. The generated recombined product can be detected on Urea-TBE sequencing gels, because the oligo's length differs after the reaction. Suicide DNA substrates were used to facilitate the reaction. The DNA substrates used in the assay contained a nick at the top strand. The sequence contained the IR_L and IR_R sites of the CI with a crossover region of 5 bp (CI5).

50 µM Int^{82N} were mixed with 2 µM 5'-^{32}P-labeled CI5, unlabeled CI5 substrate in excess (50-fold) and different concentrations of the drug candidates in a final volume of 15 µl containing 1x activity buffer – 2 (Table 4-18). The reaction mix was incubated for 2-4 hours

at 37 °C. The reactions were stopped with Proteinase K digestion at 45 °C for 30 min – 1h (Table 4-22). DNA products were precipitated with NaAc/EtOH in the presence of 24 µg/ml glycogen (Thermo Fischer) (Table 4-23) and dried DNA was resuspended in 5 µl distilled water. Afterwards the samples were mixed with 5 µl 2x formamide loading buffer. The samples were heated to 98 °C for 5 min and analyzed by electrophoresis. For that, the samples were run on denaturing 12% PAGE TBE-Urea gels at 24 W for 2 h and imaged using a Typhoon FLA-7000 phosphoimager (GE Healthcare). The 10/60 or 20/100 DNA ladders (IDT) were used as markers.

4.3.7 HJ resolution assay

A four-way HJ DNA intermediate was 5'-end-radiolabeled with ^{32}P (described in 4.3.3) at either one or two ends to monitor the size of the single-stranded (ss) DNA after performing the recombination reaction. For that, the HJ was designed to mimic the intermediate product after the first strand exchange and recombination reaction between CI6a and a specific target sequence (described in more detail in results 2.2.1), assuming initial cleavage at IR_R and T_L. All four HJ arms had different lengths, so that resolution into substrates or products could be observed by a change in the DNA length of the labeled oligo(s) on sequencing-grade Urea-TBE gels. Non-labeled arms had a 5'-phosphorylation modification to avoid the formation of unspecific products. The HJ was formed by annealing as described in section 4.3.1.1. Proper HJ formation was confirmed by EMSA (4.3.4).

50-80 µM Int (Int^{82N}, Int^{FL} or other mutants) were mixed with 1 µM 5'-^{32}P-labeled HJ (short or long variant) in a final volume of 15 µl containing 1x activity buffer – 2 (Table 4-18). For the HJ-resolution inhibition experiments, specific amounts of the drug candidate were added without changing the final volume (drug concentrations are given in the respective results section). The reaction mix was incubated for 2-4 hours at 37 °C. The reactions were stopped with Proteinase K digestion at 45 °C for 30 min – 1h (Table 4-22). DNA products were precipitated with NaAc/EtOH in the presence of 24 µg/ml glycogen (Thermo Fischer) (Table 4-23) and dried DNA was resuspended in 5 µl distilled water. Afterwards the samples were mixed either with 5 µl 2x formamide loading buffer (for denaturing gels) or 3 µl 3x glycerol loading buffer (for native gels). The samples for denaturing gels were heated to 98 °C for 5 min and analyzed by electrophoresis. For that, the samples were run on denaturing 12% PAGE TBE-Urea gels at 24 W for 2 h (6-8% PAGE TBE-Urea gels, 40-45 W for HJ_{long}) and imaged using a Typhoon FLA-7000 phosphoimager (GE Healthcare). The 10/60 or 20/100 DNA ladders (IDT) were used as markers. Also 5'-end-radiolabeled

oligos that resembled the recombined products or 5'-end-radiolabeled oligos that were also used for HJ formation, were used as markers for size estimation.

Table 4-21: HJ resolution assay reaction set up, with and without drug.

Component	15 µl Reaction	Initial concentration
5x Activity Buffer - 2	3 µl	5x
5'-^{32}P-labeled HJ DNA substrate	1 µl	1 µM
Int Protein	x µl	50 - 80 µM
Distilled water	to 15 µl	-
Drug candidate	y µl	-

Table 4-22: Proteinase K treatment reaction set up.

Component	200 µl Reaction	Initial concentration
Sample reaction	15 µl	-
Proteinase K (Carl Roth)	10 µl	10 mg/ml
2x Proteinase K buffer (20 mM Tris-HCl pH 7.5, 5 mM EDTA, 2% SDS)	100 µl	2x
Distilled water	75 µl	-

Table 4-23: NaAc/EtOH precipitation reaction set up.

Component	821 µl Reaction	Initial concentration
Reaction after Proteinase K treatment	200 µl	-
Glycogen	1 µl	20 µg
Sodium acetate	20 µl	3 M
EtOH	600 µl	100%

Table 4-24: Composition of loading buffers used for experiments using radiolabeled oligos.

Buffer	Composition
2x Formamide loading buffer	90% formamide, 1x TBE, 0.01% bromophenol blue, 0.01% xylene cyanol
3x Glycerol loading buffer	85% glycerol, 1x TBE, 0.01% bromophenol blue, 0.01% xylene cyanol

Table 4-25: Composition of a denaturing 12% PAGE TBE-Urea gel.

Component	Amount per 30 ml
Urea	15 g
10x TBE buffer	3 ml
40% Acrylamide: Bisacrylamide (19:1)	9 ml
Distilled water	6 ml
10% (w/v) APS	500 µl
TEMED	10 µl

4.3.8 Analytical size exclusion chromatography (SEC)

Int^{82N}R225K-HJ DNA complexes were analyzed by analytical size exclusion chromatography (SEC) on Superdex 200 3.2/300 gel filtration column (GE Healthcare) and ternary complexes of Int^{FL}R225K with CI or HJ-DNA and arm DNA substrates were analyzed by SEC on a Superose 6 Increase 3.2/300 gel filtration column (GE Healthcare) using ÄKTAettan or ÄKTAmicro liquid chromatography systems (GE Healthcare) following the manufacturer's instructions.

The protein-DNA complexes were prepared as follows: In 10 ml complex buffer (see Table 4-27), purified Int^{FL}R225K protein (final concentration 0.26 µM) was mixed first with CI (0.13 µM) or HJ DNA (0.065 µM) in 2:1 or 4:1 protein:DNA molar ratio. The complex was incubated at 4 °C for 30 min. Afterward, the arm DNA (0.13 µM) was added to have a final protein:DNA ratio of 2:1:1 or 4:1:2, respectively. After further incubation of 30 min at 4 °C, the samples were dialysed step-wise against crystallization buffers A, B, and C2 (see table X). The first dialysis step was carried out for 30 min, the second for one and a half hours, and the third step overnight at 4 °C. The complexes were then concentrated using Vivaspin® Turbo 15 concentrator (molecular weight cutoff of 10 kDa, Sartorius stedim biotech) and afterwards Amicon® Ultra – 0.5 ml Centrifugal Filters (molecular weight cutoff of 10 kDa, Merck Millipore) if necessary, following the manufacturer's instructions until a volume of around 100 µl was reached. The samples were filtered afterwards using Centrifugal Filter Units (Millipore).

For the Int^{82N}R225K-HJ DNA complexes most steps were the same, only the dialysis buffer differed. The sample was dialysed to crystallization buffer C (see Table 4-27). DNA only controls were prepared directly in 100 µl volumes at the desired concentrations and then treated like the protein-DNA complexes. If necessary, they were also concentrated using Amicon® Ultra – 0.5 ml Centrifugal Filters (molecular weight cutoff of 3.5 kDa, Merck Millipore) to reach the desired volume.

After complex concentration, 30 µl of the complex was injected onto the analytical gel filtration column, previously equilibrated with crystallization buffer C or C2. The elution was performed at 0.05 ml/min for 1.5 CV with the same buffer. Protein and DNA absorbance chromatograms at 280 nm and 260 nm were collected, respectively and analyzed using the UNICORN Control Software (GE Healthcare). This was important for comparison of the elution peaks between complexes and controls. When necessary, the samples were further analyzed by EMSA using a 6% native TBE gel, as described in section 4.3.4.

4.4 X-ray crystallography methods

4.4.1 Crystallization and structure solution of Tn*1549* Xis

4.4.1.1 Preparation of Xis protein for crystallization

For crystallization experiments, the Xis protein was purified until the SEC step, as stated in section 4.2.9.2. The protein purity was determined by SDS-PAGE analysis. Then, the protein was concentrated to 3.6 mg/ml using Amicon® Ultra – 0.5 ml Centrifugal Filters (molecular weight cutoff of 3.5 kDa, Merck Millipore). The buffer of the concentrated Xis protein was equal to the gel filtration buffer: 50 mM BisTris, 750 mM NaCl, 5 mM DTT, pH 6.0.

4.4.1.2 Crystallization of Xis

In order to obtain crystals from Xis, the protein solution had to reach a supersaturated state. The most common method used in X-ray crystallography is the vapor diffusion method, that was also applied in this study. By doing so, the concentration of the macromolecules and the precipitant increase in parallel leading to a supersaturated state of the protein in the solution that can lead to crystal nucleation and growth.

The initial crystallization screens were performed with help of the EMBL Crystallization Facility. These included the commercially available JCSG+ (Molecular Dimensions) and XP (Jene Bioscience) screens. The screens were prepared in 96-well MRC sitting drop Crystallization Plates™ (Molecular Dimensions) using a Mosquito Crystallization Robot (TTP LabTech). For each condition, two different crystallization drop sizes were set up, consisting of 0.1 µl of the reservoir precipitant solution and 0.1 µl or 0.2 µl of the protein solution. For each screen, two plates were made and used for crystallization screening at either 7 °C or 20 °C. The sealed plates were handled, stored and automatically imaged using an RI182 or R1000 (Formulatrix, Inc., U.S.) imaging system at either 7 °C or 20 °C, respectively.

Initial crystal hits were obtained in condition: 0.2 M LiSO4, 0.1 M BisTris, pH 5.5, 25% (w/v) PEG 3350, using the JCSG+ screen at 20 °C. These crystals had already a single-crystal shape. Thus, many of them were harvested and used for X-ray diffraction data collection. Xis crystals obtained from 96-well plates were fished using 0.05 – 0.4 mm CryoLoops (Hampton Research), transferred into the same well-solution containing 20% glycerol as cryoprotectant and then immediately transferred into liquid nitrogen-containing

MicroTubes (Hampton Research). The crystals were transferred to the X-ray source in Cryogenic Dewar Flasks (TED PELLA) under cryogenic conditions.

Table 4-26: List of conditions and screens used for initial crystallization experiments for Xis.

Complex	Screens	T	Crystallization Buffer
Xis (3.6 mg/ml)	JCSG+, XP Screen	7 °C, 20 °C	750 mM NaCl, 50 mM BisTris pH 6.0, 0.5 mM TCEP

4.4.1.3 Data collection

X-ray diffraction data from native Xis crystals were collected by rotation method on tuneable beamline ID-29 at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France. Usually an initial characterization experiment of the crystal is performed. For that, 4 diffraction images were collected at four different rotation positions (*phi* angles) to estimate the diffractions limits and determine the point group. The dataset was collected following the strategy designed by EDNA (diffraction characterization and data collection strategy software). The collection strategy for the best dataset was the following: P3, 1800 images, 0.1° oscillation, 0.037 sec exposure.

4.4.1.4 Data processing

The Xis dataset was integrated and merged in XDS, part of the XDS Program Package (Kabsch *et al.*, 2010). Moreover, resolution cut-off limits (up to 1.3 Å) were estimated from merging and CC(1/2) statistics produced by XDS. Afterwards the dataset was internally scaled in XSCALE from the XDS Program Package, producing a final table of data statistics. The reflection file was then converted to mtz format using XDSCONV from the XDS Program Package, and the resulting reflection file was used for further structure solution through molecular replacement.

4.4.1.5 Molecular replacement

Initial structure solution trials via molecular replacement using Phaser (McCoy *et al.*, 2007) and input models from the PDB-database, using previously determined Xis structures (like Xis[916], pdb:1y6u; λ-Xis, pdb:1rh6, among others), was not successful. Therefore, initial phases of Xis diffraction data were derived using ab initio phasing and chain tracing in ARCIMBOLDO (Lite) (Rodríguez *et al.*, 2009) from the CCP4 suite. This method is based on a combination of the location of model fragments like small α-helices with Phaser and density modification with SHELXE (Thorn and Sheldrick, 2013). We searched initially for 2 ideal polyalanine α-helices of 12 residues each with Phaser and iterative density modification and autotracing with SHELXE. Afterwards, a first *B*-value refinement of the

traced atoms was performed. The initial solution gave a fast fully traced polyalanine chain covering nearly the full-length of the Xis protein with a good model-density coverage. The model solution was used further for Xis sequence docking and side chain autofit in COOT (Emsley *et al.*, 2010).

4.4.1.6 Refinement and validation

The obtained model from section 4.4.1.5, the reflection file from the data processing step and the sequence of Xis were used for further refinement in Phenix (Liebschner *et al.*, 2019). In order to improve the fit of the model to the density, real-space and reciprocal-space refinement rounds were carried out iteratively until a final satisfactory model was obtained. The real-space refinement was performed with help of the COOT software (Emsley *et al.*, 2010), and included manual model rebuilding and automated real-space refinement. For these operations, 2Fo-Fc maps were contoured at 1.5 sigma and Fo-Fc maps at 3.5 sigma. Each round of optimization in COOT comprised visual inspection of all residues for their proper fit into the electron density, building new residues where free electron density was found, correction of protein geometry, and building of alternate conformations of some side chains. The reciprocal-space refinement was performed using phenix.refine from the Phenix suite. As refinement strategy, in the first round initial rigid-body fitting to the data was performed. Then, during each refinement round, XYZ coordinates (reciprocal-space), individual B-factors, and the translation/libration/screw (TLS) parameters were refined in 2 – 3 consecutive cycles. For the targets and weighting part X-ray/stereochemistry weights, X-ray/atomic displacement parameters (ADP) restraints and experimental phase restraints were chosen. Other options that were selected for refinement were: the automatically correction of N/Q/H errors and in a later stage of the process, water molecules were built and validated manually in COOT. Also, the option of automatically adding hydrogens to the model was selected in a late step of the refinement process. By following this process, a final model could be obtained after extensive refinement with several building cycles, leading to a model with R-factors of $R_{work} = 0.1342$ and $R_{free} = 0.1628$.

4.4.2 Crystallization experiments of Int^{82N} and Int^{82N}390C-DNA complexes

4.4.2.1 Preparation of complexes for crystallization

For crystallization experiments, Int^{82N} and Int^{82N}390C proteins were purified until the SEC step, as stated in section 4.2.9.2. The proteins purity was determined by SDS-PAGE analysis. DNA substrates were prepared by annealing specific oligos in high concentrations (200 µM), as described in 4.3.1. Later, each protein construct was mixed with suicide DNA substrate at a 2:1.2 molar ratio in complex buffer (final volume of 10 ml) and incubated for 30 min at RT. The complex was then directly dialysed to crystallization buffer 1 or 2 (see Table 4-27) at RT, overnight. Then, the protein-DNA complex was concentrated to ~5-10 mg/ml using a Vivaspin® Turbo 15 concentrator (molecular weight cutoff of 10 kDa, Sartorius stedim biotech) and afterward, Amicon® Ultra – 0.5 ml Centrifugal Filters (molecular weight cutoff of 10 kDa, Merck Millipore) if necessary, until the calculated volume was reached to have the desired protein amounts. The concentration of the protein was estimated assuming that the used protein amount remained the same after complex formation and concentration. Many of the tested complexes with their crystallization buffers and screens are listed in Table X (appendix).

Table 4-27: Complex formation, dialysis and crystallization buffers for diverse Int-DNA complexes.

Buffers for Int-DNA complexes	Composition
Complex buffer	1xPBS, 600 mM NaCl, 10% Glycerol, pH 7.4
Int-suicide DNA complexes	
Activity and crystallization buffer 1	25 mM Hepes, 125 mM NaCl, 5% Glycerol, 0.25 mM TCEP, pH 7.5
Activity and crystallization buffer 2	25 mM Hepes, 125 mM NaCl, 0.25 TCEP, pH 7.5
Int^{82N}R225K-HJ complexes	
Dialysis buffer A1	20 mM Hepes, 500 mM NaCl, 5% Glycerol, 5 mM $MgCl_2$, 0.25 mM TCEP, pH 7.5
Dialysis buffer B1	20 mM Hepes, 350 mM NaCl, 5% Glycerol, 5 mM $MgCl_2$, 0.25 mM TCEP, pH 7.5
Dialysis and crystallization buffer C1	20 mM Hepes, 200 mM NaCl, 10% Glycerol, 5 mM $MgCl_2$, 0.5 mM TCEP, pH 7.5
Dialysis and crystallization buffer C2	20 mM Hepes, 200 mM NaCl, 5% Glycerol, 5 mM $MgCl_2$, 0.5 mM TCEP, pH 7.5
Dialysis and crystallization buffer C3	20 mM Hepes, 200 mM NaCl, 5 mM $MgCl_2$, 0.5 mM TCEP, pH 7.5
Acetate dialysis buffer A	25 mM Sodium Acetate, 500 mM NaCl, 5% Glycerol, 5 mM $MgCl_2$, 0.25 mM TCEP, pH 5.5
Acetate dialysis buffer B	25 mM Sodium Acetate, 350 mM NaCl, 5% Glycerol, 5 mM $MgCl_2$, 0.25 mM TCEP, pH 5.5
Acetate dialysis and crystallization buffer C1	25 mM Sodium Acetate, 200 mM NaCl, 5% Glycerol, 5 mM $MgCl_2$, 0.5 mM TCEP, pH 5.5
Acetate dialysis and crystallization buffer C2	25 mM Sodium Acetate, 200 mM NaCl, 5 mM $MgCl_2$, 0.5 mM TCEP, pH 5.5
Int^{FL}R225K-DNA ternary complexes	
Dialysis buffer A2	25 mM Hepes, 500 mM NaCl, 5% Glycerol, 0.25 mM TCEP, pH 7.5
Dialysis buffer B2	25 mM Hepes, 350 mM NaCl, 5% Glycerol, 0.25 mM TCEP, pH 7.5
Dialysis buffer A3	25 mM Hepes, 500 mM NaCl, 5% Glycerol, 5 mM $MgCl_2$, 0.25 mM TCEP, pH 7.5
Dialysis buffer B3	25 mM Hepes, 350 mM NaCl, 5% Glycerol, 5 mM $MgCl_2$, 0.25 mM TCEP, pH 7.5
Dialysis and crystallization buffer D1	25 mM Hepes, 125 mM NaCl, 5% Glycerol, 0.5 mM TCEP, pH 7.5
Dialysis and crystallization buffer D2	25 mM Hepes, 125 mM NaCl, 5% Glycerol, 5 mM $MgCl_2$, 0.5 mM TCEP, pH 7.5
Dialysis and crystallization buffer D3	25 mM Hepes, 125 mM NaCl, 5 mM $MgCl_2$, 0.5 mM TCEP, pH 7.5

4.4.2.2 Crystallization of Int^{82N} and Int^{82N}390C-DNA complexes

In order to obtain crystals from Int in an active state, the protein-DNA solution had to reach a supersaturated state. The most common method used in X-ray crystallography is the vapor diffusion method, that was also applied in this study. By doing so, the concentration of the macromolecules and the precipitant increase in parallel leading to a supersaturated state of the protein in the solution that can lead to crystal nucleation and growth.

The initial crystallization screens were performed with help of the EMBL Crystallization Facility. The tested screens are listed in Table 6-3, appendix. They were prepared in 96-well MRC sitting drop Crystallization Plates^{TM} (Molecular Dimensions) using a Mosquito Crystallization Robot (TTP LabTech). For each condition, two different crystallization drop sizes were set up, consisting of 0.1 µl of the reservoir precipitant solution (70 µl) and 0.1 µl or 0.2 µl of the protein solution. For each screen, if possible, two plates were made and used for testing crystal growth at 7 °C and 20 °C. The sealed plates were handled, stored and automatically imaged using an RI182 or R1000 (Formulatrix, Inc., U.S.) imaging system at either 7 °C or 20 °C, respectively.

For the Int^{82N}-sCI5 complex, crystal hits appeared in the JCSG+-D12 condition (0.2 M NaCl, 0.1 M Bis-tris pH 5.5, 25% (w/v) PEG 3350) at 20 °C. They were chosen for manual optimization by screening around the original condition in 24 well-plates at 20 °C and using the hanging-drop method. For each condition, two crystallization drops were set up, consisting of 1 µl of the reservoir precipitant solution and 1 µl or 2 µl of the protein-DNA complex solution. Crystals grew as multimeric needles/plates in many conditions in the 24 well-plates after a few days.

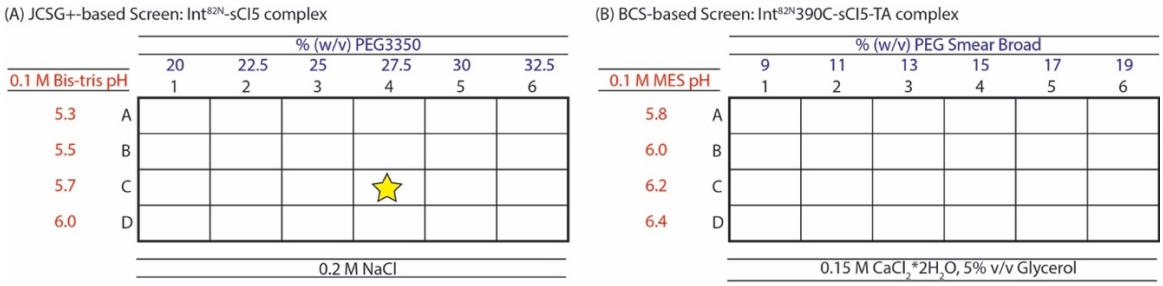

Figure 4-3: Schemes of optimization screens for Int^{82N}-sCI5 and Int^{82N}390C-sCI-TA complexes. They show the crystallization parameters in 24-well crystallization plates. Star shows the condition that led to crystals that diffracted up to 4 Å for the Int^{82N}-sCI5 complex.

In parallel to diffraction experiments, Int^{82N}-sCI5 crystals obtained after optimization were checked for the presence of the protein-DNA complex. For that, a few crystals were transferred to drops of reservoir solution to wash soluble and precipitated protein-DNA complex away. Afterwards, the crystals were dissolved in distilled water, and analyzed by

SDS-PAGE (see section 4.2.10) following silver-staining using the Pierce™ Silver Stain Kit (Thermo-Fischer) according to the manufacturer's instructions.

For the Int^{82N}390-sCI5-TA complex, crystal hits appeared in the BCS-D8 condition (0.15 M Calcium chloride dihydrate, 0.1 M MES pH 6.2, 15% v/v PEG Smear Broad, 5% v/v Glycerol) at 7 °C. The crystals grew after 4 days and had a thin plate and/or thin rod form. Apart from that, they were very small and stopped growing 2-3 days after. They were chosen for manual optimization by screening around the original condition in 24 well-plates at 7 °C and using the hanging-drop method. For each condition, two crystallization drops were set up, consisting of 1 µl of the reservoir precipitant solution and 1 µl or 2 µl of the protein-DNA complex solution. Different protein-DNA complex concentrations were screened. Few hits appeared as small, thin and, shiny "flower-shaped" crystals in many conditions in the 24 well-plates after two weeks or longer.

Table 4-28: Composition of PEG Smear Broad used in crystallization optimization. The PEG Smear is from the BCS screen (Molecular Dimensions).

PEG Smear	Composition
50% v/v PEG Smear Broad	4.55% v/v PEG 400
	4.55% v/v PEG 500 MME
	4.55% v/v PEG 600
	4.55% w/v PEG 1000
	4.55% w/v PEG 2000
	4.55% w/v PEG 3350
	4.55% w/v PEG 4000
	4.55% w/v PEG 5000 MME
	4.55% w/v PEG 6000
	4.55% w/v PEG 8000
	4.55% w/v PEG 10000

In order to have better crystal nucleation and growth that may lead to bigger crystal size, the streak seeding method was used (also called microseeding). Some of the crystals, obtained from the 24-well screen (Figure 4-3B), were used for microseeding. They were crushed in reservoir solution from the B4 condition into tiny, micro crystals with a probe and transferred to a Seed Bead tube (Hampton Research) in a total volume of 50 µl, following the manufacturer's instructions. In an attempt to obtain more single crystals, the seed stock was used directly on fresh reservoir-complex drops, repeating the 24-well screen from Figure 4-3B. For that, streak seeding was performed by running the tip of the cat whisker in a straight manner across the middle of the drop. This method did not lead to increased nucleation or crystal growth at all. Thus, the procedure was not repeated with sequential dilutions of the original seed stock.

In parallel, as a control, sCI5_TA-DNA drops were used for manual screening as mentioned above for the protein-DNA complex. DNA crystal growth was tested around the original

condition in the 24-well plate format at 7°C. DNA crystals grew after a few days, confirming that very probably the initial hits were DNA crystals without the Int^{82N}390C protein.

4.4.2.3 Diffraction experiments of Int^{82N} and Int^{82N}390C-DNA crystals

Because the multimeric needles/plates were too thin and too close together to be broken and tested as single crystals, the whole Int^{82N}-sCI5 crystals were fished and tested for diffraction. They were fished using 0.05 – 0.2 mm CryoLoops (Hampton Research), transferred into the same well-solution containing 10% glycerol as cryoprotectant and then immediately transferred into liquid nitrogen-containing MicroTubes (Hampton Research). The crystals were transferred to the X-ray source in Cryogenic Dewar Flasks (TED PELLA) under cryogenic conditions. The best diffracting crystal grew in condition 0.2 M NaCl, 0.1 M Bis-tris pH 5.7, 27.5% (w/v) PEG 3350 at 20 °C. X-ray diffraction experiments from native Int^{82N}-sCI5 crystals were performed by rotation method on fixed beamline ID23-2 at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France. The initial characterization experiment of the crystal was performed by collecting 4 diffraction images at four different rotation positions (*phi* angles) to determine its diffraction limits and to calculate the collection strategy, if strong diffraction was detected. The best crystal showed diffraction at around 4 Å, but with overlapping patterns due to the use of multi-crystals. Thus, EDNA was not able to calculate a collection strategy and no dataset was collected.

Int^{82N}390-sCI5-TA complex crystals from both plate types (96- and 24-well) were tested for diffraction. They were prepared and characterized for diffraction as described above for Int^{82N}-sCI5 crystals. The crystals did not diffract and thus, no dataset could be collected.

4.4.3 Crystallization experiments of Int^{82N}R225K-HJ DNA complexes

4.4.3.1 Preparation of Int^{82N}R225K-HJ DNA complexes for crystallization

Here, the specifics for the Int^{82N}R225K-HJ1TA complex will be explained, that led to good diffracting crystals and structure solution. This general procedure was applied to all tested Int-HJ complexes and differed only in the buffers and concentrations used for crystallization.

For crystallization experiments, Int^{82N}R225K protein was purified until the SEC step, as stated in section 4.2.9.2. The protein purity was determined by SDS-PAGE analysis.

HJ1TA-DNA substrate was prepared by annealing specific oligos in high concentrations (200 μM), as described in 4.3.1. It should be mentioned, PAGE purified or normal oligos led to similar diffraction. Afterwards, the protein construct was mixed with HJ1TA DNA substrate at 3.5:1 molar ratio in complex buffer (final volume of 10 ml) and incubated for 30 min at 4 °C. Fresh TCEP was added to the complex to achieve a 0.25 mM final concentration. Then, the complex was step-wise dialysed: first to dialysis buffer A1 for 30 min, at 7 °C; then the complex was transferred to dialysis buffer B1 for one and a half hours, at 7 °C. Finally, the complex was dialysed to crystallization buffer C3 at 7 °C, ON (see Table 4-27).

The Int^{82N}R225K-HJ1TA complex was concentrated to ~5-10 mg/ml using a Vivaspin® Turbo 15 concentrator (molecular weight cutoff of 10 kDa, Sartorius stedim biotech) and afterwards, an Amicon® Ultra – 0.5 ml Centrifugal Filter (molecular weight cutoff of 10 kDa, Merck Millipore) if necessary, until the calculated volume was reached to have the desired protein amounts. The concentration of the protein was estimated assuming that the used protein amount remained the same after complex formation and concentration. Many of the tested complexes with their crystallization buffers and screens are listed in Table 6-4 (appendix).

4.4.3.2 Crystallization of Int^{82N}R225K-HJ DNA complexes

The initial crystallization screens were performed as described in section 4.4.2.2. For the Int^{82N}R225K-HJ1TA complex initial hits were obtained in 96-well plates at 1:1 drop ratio, using the PEGs screen (QIAGEN), at 7 °C. Thin, roundish, single crystals grew after 3 – 4 days in a condition containing 0.1 M Sodium acetate pH 4.6 and 40% (v/v) PEG 200. This condition was chosen for manual optimization by screening around the original condition in 24 well-plates at 7 °C and by using two different vapor diffusion methods: the sitting-drop method and the hanging-drop method. For the sitting-drop method, two crystallization plates were set up. In the first plate, drops were pipetted, consisting of 1 μl of the reservoir precipitant solution (500 μl) and 1 μl of the protein-DNA complex solution. In the second plate, 2 μl of the protein-DNA complex was pipetted to 1 μl of the reservoir precipitant solution (500 μl). For the hanging-drop method, in each condition, up to three crystallization drops were set up, consisting of 1 μl of the reservoir precipitant solution (500 μl) and 1 μl, 1.5 μl or 2 μl of the protein-DNA complex solution. Also, other drop ratios of 1.5:1.5 μl, 2:3 μl or 0.5:1 μl of reservoir solution:complex and more were tested in further optimizations. Reproducible crystals grew in a wide range of conditions surrounding the

original crystallization reservoir, using both vapor-diffusion methods. The crystals appeared after 4 – 7 days. They were big and had a roundish and flat form, moreover they showed birefringence. Crystals that started growing faster, already after 1 – 2 days, led to smaller and multimeric crystals. Bigger crystals continued growing up to one month after nucleation.

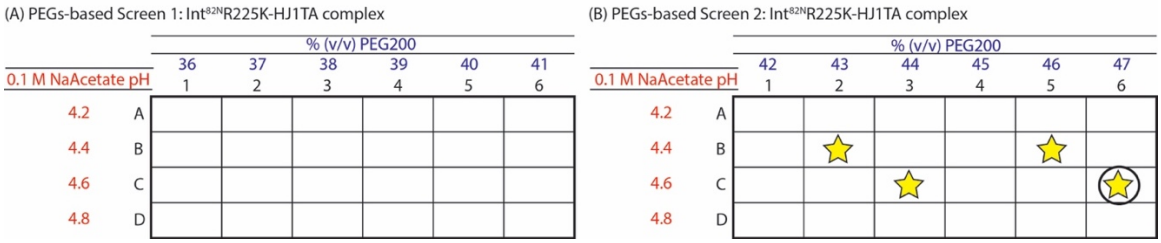

Figure 4-4: Schematics of optimization screens for the Int^{82N}R225K-HJ1TA complex. They show the crystallization parameters in 24-well crystallization plates. Yellow star indicates conditions that led to crystal hits that were tested in diffraction experiments and led to many data sets. Yellow star encircled highlights the condition (0.1M Sodium acetate pH 4.6, 47% (v/v) PEG 200) that led to the best diffracting crystals, used for data collection and structure solution. PEG 200 for optimization screenings was purchased from NeXtal (QIAGEN).

DNA-only controls were performed by setting up specific drops of reservoir (around the hit condition) and HJ1TA DNA (in crystallization buffer C3) in 24-well manual plates, as described above for the protein-DNA complex. This control did not lead to crystallization of HJ1TA-DNA alone.

Int^{82N}R225K-HJ1TA crystals obtained from 96-well and manual plates were fished using 0.05 – 0.4 mm CryoLoops (Hampton Research) and immediately transferred into liquid nitrogen-containing MicroTubes (Hampton Research). The crystals were transferred to the X-ray source in Cryogenic Dewar Flasks (TED PELLA) under cryogenic conditions.

4.4.3.3 Data collection

X-ray diffraction datasets were collected at beamlines of the European Synchrotron Radiation Facility (ESRF) in Grenoble, France and German Electron Synchrotron (DESY) in Hamburg, Germany. The best X-ray diffraction data from native Int^{82N}R225K-HJ1TA crystals were collected by rotation method on tuneable beamline P13 at DESY. Crystals underwent an initial characterization experiment before data collection. For that, 4 diffraction images were collected at four different rotation positions (*phi* angles) to estimate the diffraction limits and determine the point group. Several datasets were collected following the strategy designed by EDNA (diffraction characterization and data collection strategy software). The collection strategy for the best dataset was the following: P1, 3600 images, 0.1° oscillation, 0.04 sec exposure time and beam transmission of 11.93%.

4.4.3.4 Data processing

The Int^{82N}R225K-HJ datasets were analyzed and integrated in XDS, part of the XDS Program Package (Kabsch *et al.*, 2010). Initial XDS processing of the best Int^{82N}R225K-HJ1TA dataset showed that the crystal had most probably P1 space group with the cell dimensions of 80.185, 96.27, and 102.50 Å. The data was cut following the CC(1/2) criterion (Karplus and Diederichs, 2012). The study shows that resolution shells that are below the $I/\sigma(I) \geq 2$ criterion but still have strong CC(1/2) values might contain useful data. Therefore, data up to 3.3 Å resolution with acceptable CC(1/2) value (30.2) in the highest resolution shell were included for further processing (Table 2-2, results 2.1.2.5). The overall completeness of the dataset was 97.8%, with low redundancy of 3.5-fold. The R-factors (R-factor, also called R-merge, and R-meas) indicate the data's internal consistency. They should be as low as possible in high-resolution shells and lower than 10% for the entire dataset and the lowest resolution shell. Here, the criteria fit the low-resolution shell, but it raised significantly with higher resolution giving values over 189.9% (R-merge) and 227.4% (R-meas) in the highest resolution shell (3.5 - 3.3 Å). Also, the signal to noise ratio worsened rapidly at high resolution, from an $I/\sigma(I)$ of 2.29 at 4.04 Å to already 0.58 at 3.3 Å resolution.

These unexpected issues including the high R-values at high resolution may show that the dataset had some anisotropy issues. This means that diffraction during data collection was stronger in some directions than in other directions, which may lay in the crystal's nature. Moreover, the dataset's overall R-merge was calculated at 10.1% and R-meas at 12%, indicating some problems. Therefore, the dataset was analyzed and anisotropically scaled and merged, producing a final table of data statistics (Table 2-3, results 2.1.2.5) and an MTZ file for further structure solution, using the STARANISO server from Global Phasing Ltd. (http://staraniso.globalphasing.org/). The dataset turned out to be strongly anisotropic, showed by the spherical completeness (64.84%) of the data after scaling and merging (Table 2-3). Thus, the data was scaled anisotropically and cut elliptically at a resolution of 3.3 Å. This helped to enhance the data statistics, especially in the high-resolution shells, with an improved ellipsoidal completeness (90.75%, Table 2-3).

The resulting reflection file was used for further structure solution through molecular replacement.

4.4.3.5 Molecular replacement

The anisotropically scaled and merged dataset from STARANISO was used further as input for molecular replacement in Phaser (McCoy *et al.*, 2007) to solve the Int^{82N}R225K-HJ1TA complex structure. Matthews coefficients were calculated with Xtriage in Phenix (Zwart, Grosse-Kunstleve, 2005), indicating that the unit cell is most probably occupied by 4 monomers bound to 1 HJ-DNA molecule per asymmetric unit. Thus, I searched for four Int^{82N}R225K molecules and four CI-half sites in molecular replacement. The modified Int^{82N}R225K-CI5 structure [PDB: 6emz; (Rubio-Cosials *et al.*, 2018)] was used as a search model that consisted of one protein-DNA subunit of the structure and led to a reasonable solution with high (log-likelihood gain) LLG- and (translation function Z) TFZ scores that lay over the confidence limit of 120 and 8, respectively. After initial rigid-body refinement with PHENIX (Liebschner *et al.*, 2019), the result showed 4 Int monomers bound to 4 CI5 DNA half-sites forming a tetrameric complex. Moreover, the four DNA molecules seemed to come together at the complex's center to form the DNA junction. Thus, I concluded that the initial model fitted reasonably well into the electron density. However, the density was weaker at the CAT domains.

4.4.3.6 Refinement and validation

The current model was obtained by alternating intensive model building in COOT (Emsley *et al.*, 2010) and refinement runs in PHENIX (Liebschner *et al.*, 2019). In order to improve the fit of the model to the density, real-space refinement and reciprocal-space refinement rounds were carried out iteratively. The real-space refinement was performed with help of COOT and included manual model rebuilding and automated real-space refinement. For these operations, 2Fo-Fc maps were contoured at 1.0 sigma and Fo-Fc maps at 3.0 sigma. Each round of optimization in COOT comprised visual inspection of all residues for their proper fit into the electron density, building new residues where free electron density was found, correction of protein geometry, and sometimes deletion of side chains where density was missing. The reciprocal-space refinement was performed using phenix.refine from the Phenix suite. As refinement strategy, in the first round, initial rigid-body fitting to the data was performed. Then, during each refinement round, XYZ coordinates (reciprocal-space), group B-factors, and the translation/libration/screw (TLS) parameters were refined in 2 – 3 consecutive cycles. For the targets and weighting part, X-ray/stereochemistry weights, X-ray/atomic displacement parameters (ADP) restraints and experimental phase restraints were chosen. Also, the use of secondary structure restraints was tested. Other options that

were selected for refinement was the automatically correction of N/Q/H errors. By following this process, an actual model could be obtained after refinement with several building cycles, using diffraction data up to 3.3 Å resolution with still high R-values (R_{work} = 0.3250 and R_{free} = 0.3710). Further refinement is work in progress.

4.4.3.7 Merging of Datasets with BLEND

Datasets collected from single Int-HJ1TA crystals showed weak high-resolution reflections and low redundancy that hampered the structure solution process. Therefore, various datasets collected from Int-HJ1TA crystals, grown in the same crystallization drop or similar conditions in the same manual plate, were analyzed and merged with the help of Vladimir Arinkin (a postdoc in the lab) using the software BLEND from the CCP4 suite (Foadi *et al.*, 2013). This approach may improve the quality of the data at high resolution and help in structure solution.

As input, unscaled integrated datasets (INTEGRATE.HKL) from XDS were used. A total of 11 different datasets were analyzed and clustered in BLEND, based on statistical descriptors that are calculated from each dataset (analysis mode). The results from the analysis mode run were then used in the synthesis mode run. There, BLEND scaled together the previously clustered datasets and gave as output merged reflection files in MTZ format that were used further in molecular replacement and refinement trials for Int-HJ1TA structure solution.

4.4.3.8 Miscellaneous

Structural figures of models and superimpositions were generated with UCSF ChimeraX (http://www.rbvi.ucsf.edu/chimerax) (Pettersen *et al.*, 2020). Superimpositions of the obtained structural models with other models, shown in this work, were generated in COOT using secondary-structure matching (SSM) superimposition function (Krissinel and Henrick, 2004).

4.5 Microbiology methods

4.5.1 *In vivo* mini-Tn*1549* excision assay in *E. coli*

The mini-Tn*1549* excision assay *in vivo* (Lambertsen *et al.*, 2018) was performed in order to obtain the CI-form of the mini-transposon for further cloning experiments and insert it into a pET28M-vector. The *E. coli* Top10 cells were already transformed with the PEP plasmid (protein-expressing plasmid) for expression of Xis and Int proteins under the control of a P_{BAD} promoter and a DP plasmid (donor plasmid) encoding the mini-Tn*1549*. Upon protein induction and expression, Xis and Int can excise the mini-transposon from the DP plasmid and create the circular intermediate *in vivo*.

In day one, 5 ml LB (with 0.2% glucose) were inoculated with *E. coli* Top 10 cells (glycerol stock) and proper antibiotics (Amp and Cm) and grown at 37 °C, overnight. The next day, the ON-culture was used to inoculate fresh 10 ml LB (with 0.2% glucose and proper antibiotics) at OD_{600} of 0.05. The culture was grown for 2 h at 37 °C until the mid-exponential growth phase was reached. Afterward, protein expression was induced by changing the medium to 10 ml LB (with 0.2% arabinose, and Amp and Cm antibiotics). The culture was diluted to an OD_{600} of 0.1 and grown for 3 h at 37 °C.

Samples were taken at specific time points for PCR analysis. For that, 1 ml samples were taken, OD_{600} measured and spun down 4 °C at 5000g for 5 min. The supernatant was removed, and the pellets were either directly used or stored at -20 °C. Samples were taken at timepoints, t = 0 (before the change to the arabinose medium), and at t = 3 h after induction of protein expression. The pellets were resuspended in distilled water (1xPBS for analyzing protein expression) and corrected for OD_{600} of the sample. For example, 50 µl distilled water were used for a pellet with an OD_{600} = 0.5. Then, the resuspended cells were boiled for 10 min, cooled down on ice and centrifuged. For the PCR reactions (detecting excision products or cloning of the CI intermediate), 0.5 µl supernatant were used as template.

The excised mini-Tn*1549* was detected via PCR (see section 4.2.2). The first PCR reaction detected the core sequence of the joined transposon ends (IR_L-CR-IR_R) after mini-transposon excision. For that, a primer pair (P2 and P3, see Table 4-2) were used that matched to the transposon ends. The second PCR reaction detected the presence or absence of the mini-Tn*1549* by using primers (P46 and P47, see Table 4-2) that annealed to the left and right flanks. Thus, two products of different lengths could be obtained. The PCR products were analyzed by 1% agarose gel electrophoresis (see section 4.2.5).

Table 4-29: PCR set up for detecting products of the in vivo mini-Tn1549 assay.

Component	25 µl Reaction	Final Concentration
Distilled water	to 25 µl	-
2x Phusion Flash High-Fidelity PCR Master Mix (Thermo Scientific)	12.5 µl	1x
10 µM Forward primer	1.25 µl	500 nM
10 µM Reverse primer	1.25 µl	500 nM
Template DNA (supernatant)	0.5 µl (variable)	-

Table 4-30: Thermocycling conditions for PCR to detect the excised mini-Tn1549.

Step	Temperature	Time
1. Initial Denaturation	98 °C	2 min
25 cycles (step 2-4):		
2. Denaturation	98 °C	10 seconds
3. Annealing (P2/P3) or	58 °C	10 seconds
3. Annealing (P46/P47)	60 °C	10 seconds
4. Extension	72 °C	30 - 60 seconds
5. Final extension	72 °C	2 minutes
6. Hold	4 °C	hold

4.6 Biophysical methods

4.6.1 Isothermal titration calorimetry (ITC)

ITC measures the energy that is necessary to maintain a constant temperature during a titration experiment, so that heat changes that result from the direct interaction of molecules in the system, are compensated. Therefore, ITC can indirectly measure heat absorbance or its release. This permits the calculation of binding constants and other thermodynamic parameters. Initial ITC experiments were performed to study Int^{82N}-novobiocin interactions. The measurements were performed using a MicroCal PEAQ-ITC instrument (Malvern Panalytical, Spectris) in an ITC buffer (350 mM NaCl, 50 mM HEPES pH 7.5, 0.25 mM TCEP). The stock solution of novobiocin was diluted, while the integrase solution was first dialysed and then diluted accordingly in the ITC buffer to a drug stock concentration of 200 µM (run 1) or 1 M (run 2) and a protein concentration of 20 µM, respectively, before the titration experiments. The protein sample was titrated using the compound. The experiment started with an initial injection of 2 ul novobiocin solution followed by 11 injections of 3 ul with an interval of 150 s between the injections. The data were analyzed by MicroCal PEAQ-ITC analysis software.

4.6.2 Nanoscale differential scanning fluorimetry (NanoDSF)

We tested protein-drug interaction by performing NanoDSF experiments. This method uses the intrinsic fluorescence of proteins that originate from aromatic side chains, mainly

tryptophan and tyrosine residues, to measure protein unfolding during heat induced unfolding. Upon protein denaturation, differences in the environment will lead to a variation in the fluorescence intensity. From these observations, an apparent melting temperature can be obtained. Protein-ligand interactions can lead to changes in protein stability and, thus, in its melting point. We measured Int^{82N} tryptophan fluorescence emission at 330 nm and 350 nm wavelength (excitation wavelength at 280 nm) using a Prometheus NT.48 instrument (NanoTemper Technologies). Capillaries were filled with 10 µl of i) Int^{82N}-protein (1 mg/ml), and ii) protein-novobiocin mix, having 1 mg/ml protein and varied concentration of the compound (0.5, 1 and 4 mM) in buffer (350 mM NaCl, 50 mM HEPES pH 7.5, 0.25 mM TCEP). The capillaries were located into the sample holder and the temperature was increased from 20 to 95 °C at a ramp rate of 1°C/min. Whereby, the fluorescence emission was measured once every 0.05 °C. The ratio of the recorded emission intensities (E350nm/E330nm) was plotted as a function of the temperature. The fluorescence intensity ratio and its first derivative were calculated with the PR.ThermControl version 2.1.2 software (from the manufacturer).

4.6.3 Thermofluor shift assay (TSA)

Int^{82N}-novobiocin interactions were investigated with the help of Thermofluor assays. This assay uses a fluorescent dye, in this case, SYPRO® Orange (Molecular Probes, Sigma-Aldrich), that can bind to hydrophobic parts of the protein. These regions are usually hidden inside the protein. Upon temperature increase in a step-wise manner, the protein will start unfolding, leading to exposure of its hydrophobic residues. The binding of Sypro to the protein will increase fluorescence emissions that peak when all the protein's hydrophobic regions are exposed. Afterwards, protein precipitation will occur, occluding some of the hydrophobic parts again, leading to less fluorophore-binding and reducing the fluorescence. This temperature-dependent behavior will give a so-called melting curve. The first peak of the derivative curve shows the melting temperature indicating that around 50 % of the solution's protein is unfolded. Different protein-partner interactions can increase or reduce the protein stability that may lead to a higher or lower melting point compared to the wild type protein (in a specific condition).

Int^{82N} (20 µM, final concentration in solution) was mixed with increasing drug concentrations and constant concentration of Sypro dye (5X) in a final volume of 25 µl, see Table 4-31. As controls, Sypro orange alone and with novobiocin were also tested. The samples were analyzed in a real-time PCR machine (CFX Connect Real-Time PCR

Detection System – Bio-Rad). After equilibration at 5 °C for 1 min, the melting curve was measured using a gradient with 1 °C/min from 5 °C to 95 °C and one scan/1°C. Different runs in respect to the TSA-buffer were performed (see Table 4-32). The melting and first derivative curves were obtained using the Real-time PCR machine's program CFX Maestro Software (Bio-Rad) and used for data analysis with Excel.

Table 4-31: TSA set up and final concentration of each reaction component.

Component	Final concentration
Buffer	1, 2 or 3
Protein	Int^{82N} (20 µM)
Compound	Novobiocin (0, 0.01, 0.02, 0.1, 0.5, 1, 1.5, 2, 3, 4 mM)
Sypro orange	5x
Final volume	25 µl

Table 4-32: Buffers used for different TSA runs. Runs 1 and 2 tested 2 salt concentrations and similar conditions compared to the ITC/NanoDSF experiments. Run 3 tested the buffer conditions usually used for performing activity assays in vitro.

TSA-Buffer	Composition (final concentrations/conditions)	TSA-Run
Buffer 1	350 mM NaCl, 50 mM HEPES pH 7.5, 0.25 mM TCEP	1
Buffer 2	175 mM NaCl, 50 mM HEPES pH 7.5, 0.25 mM TCEP	2
Buffer 3	125 mM NaCl, 25 mM HEPES pH 7.5, 5% glycerol, 0.25 mM TCEP	3

5 REFERENCES

Abbani, M. A. *et al.* (2007) 'Structure of the cooperative Xis-DNA complex reveals a micronucleoprotein filament that regulates phage lambda intasome assembly', *Proceedings of the National Academy of Sciences of the United States of America*, 104(7), pp. 2109–2114. doi: 10.1073/pnas.0607820104.

Abbani, M., Iwahara, M. and Clubb, R. T. (2005) 'The structure of the excisionase (Xis) protein from conjugative transposon Tn916 provides insights into the regulation of heterobivalent tyrosine recombinases', *Journal of Molecular Biology*, 347(1), pp. 11–25. doi: 10.1016/j.jmb.2005.01.019.

Agudelo Higuita, N. I. and Huycke, M. M. (2014) *Enterococcal Disease, Epidemiology, and Implications for Treatment*, *Enterococci: From Commensals to Leading Causes of Drug Resistant Infection*. Massachusetts Eye and Ear Infirmary. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24649504 (Accessed: 10 November 2020).

Ambur, O. H. *et al.* (2016) 'Steady at the wheel: Conservative sex and the benefits of bacterial transformation', *Philosophical Transactions of the Royal Society B: Biological Sciences*. Royal Society of London. doi: 10.1098/rstb.2015.0528.

Aminov, R. I. (2010) 'A Brief History of the Antibiotic Era: Lessons Learned and Challenges for the Future', *Frontiers in Microbiology*. Frontiers Research Foundation, 1(DEC), p. 134. doi: 10.3389/fmicb.2010.00134.

Arenz, S. and Wilson, D. N. (2016) 'Bacterial protein synthesis as a target for antibiotic inhibition', *Cold Spring Harbor Perspectives in Medicine*. Cold Spring Harbor Laboratory Press, 6(9), p. a025361. doi: 10.1101/cshperspect.a025361.

Arias, C. A. and Murray, B. E. (2012) 'The rise of the Enterococcus: Beyond vancomycin resistance', *Nature Reviews Microbiology*. Nature Publishing Group, 10(4), pp. 266–278. doi: 10.1038/nrmicro2761.

Barabas, O. *et al.* (2008) 'Mechanism of IS200/IS605 Family DNA Transposases: Activation and Transposon-Directed Target Site Selection', *Cell*. Cell, 132(2), pp. 208–220. doi: 10.1016/j.cell.2007.12.029.

Barna, J. C. J. and Williams, D. H. (1984) 'The Structure and Mode of Action of Glycopeptide Antibiotics of the Vancomycin Group', *Annual Review of Microbiology*. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA , 38(1), pp. 339–357. doi: 10.1146/annurev.mi.38.100184.002011.

Bax, B. D. *et al.* (2010) 'Type IIA topoisomerase inhibition by a new class of antibacterial agents', *Nature*. Nature Publishing Group, 466(7309), pp. 935–940. doi: 10.1038/nature09197.

Bebel, A. *et al.* (2016) 'Structural snapshots of Xer recombination reveal activation by synaptic complex remodeling and DNA bending', *eLife*. eLife Sciences Publications Ltd, 5(DECEMBER2016). doi: 10.7554/eLife.19706.

Berleman, J. and Auer, M. (2013) 'The role of bacterial outer membrane vesicles for intra- and interspecies delivery', *Environmental Microbiology*. John Wiley & Sons, Ltd, 15(2), pp. 347–354. doi: 10.1111/1462-2920.12048.

Bi, D. *et al.* (2019) 'Large-Scale Identification of AbaR-Type Genomic Islands in Acinetobacter baumannii Reveals Diverse Insertion Sites and Clonal Lineage-Specific Antimicrobial Resistance Gene Profiles', *Antimicrobial agents and*

chemotherapy. NLM (Medline), 63(4). doi: 10.1128/AAC.02526-18.

Biswas, T. *et al.* (2005) 'A structural basis for allosteric control of DNA recombination by λ integrase', *Nature*. Nature Publishing Group, 435(7045), pp. 1059–1066. doi: 10.1038/nature03657.

Blakely, G. *et al.* (1993) 'Two related recombinases are required for site-specific recombination at dif and cer in E. coli K12', *Cell*. Elsevier, 75(2), pp. 351–361. doi: 10.1016/0092-8674(93)80076-Q.

Boles, M. O. and Taylor, D. J. (1975) 'The crystal structure of novobiocin', *Acta Crystallographica Section B Structural Crystallography and Crystal Chemistry*. International Union of Crystallography (IUCr), 31(5), pp. 1400–1406. doi: 10.1107/s0567740875005286.

Bolukaoto, J. Y. *et al.* (2015) 'Antibiotic resistance of Streptococcus agalactiae isolated from pregnant women in Garankuwa, South Africa', *BMC Research Notes*. BioMed Central Ltd., 8(1). doi: 10.1186/s13104-015-1328-0.

Bonomo, R. A. (2017) 'β-Lactamases: A focus on current challenges', *Cold Spring Harbor Perspectives in Medicine*. Cold Spring Harbor Laboratory Press. doi: 10.1101/cshperspect.a025239.

Brauner, A. *et al.* (2016) 'Distinguishing between resistance, tolerance and persistence to antibiotic treatment', *Nature Reviews Microbiology*. Nature Publishing Group, pp. 320–330. doi: 10.1038/nrmicro.2016.34.

Burrus, V. *et al.* (2002) 'Conjugative transposons: The tip of the iceberg', *Molecular Microbiology*. John Wiley & Sons, Ltd, pp. 601–610. doi: 10.1046/j.1365-2958.2002.03191.x.

Burrus, V. (2017) 'Mechanisms of stabilization of integrative and conjugative elements', *Current Opinion in Microbiology*. Elsevier Ltd, 38, pp. 44–50. doi: 10.1016/j.mib.2017.03.014.

Bush, K. and Bradford, P. A. (2016) 'β-lactams and β-lactamase inhibitors: An overview', *Cold Spring Harbor Perspectives in Medicine*. Cold Spring Harbor Laboratory Press, 6(8), p. a025247. doi: 10.1101/cshperspect.a025247.

Cabezón, E. *et al.* (2015) 'Towards an integrated model of bacterial conjugation', *FEMS Microbiology Reviews*. Oxford University Press, pp. 81–95. doi: 10.1111/1574-6976.12085.

Cabezón, E., Ignacio Sastre, J. and De La Cruz, F. (1997) 'Genetic evidence of a coupling role for the TraG protein family in bacterial conjugation', *Molecular and General Genetics*. Mol Gen Genet, 254(4), pp. 400–406. doi: 10.1007/s004380050432.

Cabezon, E. and de la Cruz, F. (2006) 'TrwB: An F1-ATPase-like molecular motor involved in DNA transport during bacterial conjugation', *Research in Microbiology*. Res Microbiol, pp. 299–305. doi: 10.1016/j.resmic.2005.12.002.

Caparon, M. G. and Scott, J. R. (1989) 'Excision and insertion of the conjugative transposon Tn916 involves a novel recombination mechanism', *Cell*. Cell, 59(6), pp. 1027–1034. doi: 10.1016/0092-8674(89)90759-9.

Chandler, M. *et al.* (2013) 'Breaking and joining single-stranded DNA: the HUH endonuclease superfamily', *Nature Reviews Microbiology*. Nature Publishing Group, pp. 525–538. doi: 10.1038/nrmicro3067.

Chandler, M. (2016) ‘Transposons: Prokaryotic’, in *eLS*. Chichester, UK: John Wiley & Sons, Ltd, pp. 1–9. doi: 10.1002/9780470015902.a0000591.pub2.

Chen, Y. and Rice, P. A. (2003) ‘New insight into site-specific recombination from Flp recombinase-DNA structures’, *Annual Review of Biophysics and Biomolecular Structure*. Annu Rev Biophys Biomol Struct, pp. 135–159. doi: 10.1146/annurev.biophys.32.110601.141732.

Chopra, I. and Roberts, M. (2001) ‘Tetracycline Antibiotics: Mode of Action, Applications, Molecular Biology, and Epidemiology of Bacterial Resistance’, *Microbiology and Molecular Biology Reviews*. American Society for Microbiology, 65(2), pp. 232–260. doi: 10.1128/mmbr.65.2.232-260.2001.

Clewell, D. B. *et al.* (1995) ‘Unconstrained bacterial promiscuity: the Tn916-Tn1545 family of conjugative transposons’, *Trends in Microbiology*. Elsevier Current Trends, pp. 229–236. doi: 10.1016/S0966-842X(00)88930-1.

Coates, A. R., Halls, G. and Hu, Y. (2011) ‘Novel classes of antibiotics or more of the same?’, *British Journal of Pharmacology*. John Wiley & Sons, Ltd, pp. 184–194. doi: 10.1111/j.1476-5381.2011.01250.x.

Cochetti, I. *et al.* (2008) ‘erm(B)-carrying elements in tetracycline-resistant pneumococci and correspondence between Tn1545 and Tn6003’, *Antimicrobial Agents and Chemotherapy*. American Society for Microbiology Journals, 52(4), pp. 1285–1290. doi: 10.1128/AAC.01457-07.

Connolly, K. M., Iwahara, M. and Clubb, R. T. (2002) ‘Xis protein binding to the left arm stimulates excision of conjugative transposon Tn916’, *Journal of Bacteriology*, 184(8), pp. 2088–2099. doi: 10.1128/JB.184.8.2088-2099.2002.
Cooper, M. A. and Shlaes, D. (2011) ‘Fix the antibiotics pipeline’, *Nature*. Nature Publishing Group, p. 32. doi: 10.1038/472032a.

Courvalin, P. (2006) ‘Vancomycin resistance in gram-positive cocci’, *Clinical Infectious Diseases*. Clin Infect Dis. doi: 10.1086/491711.

Dantas, G. and Sommer, M. O. A. (2012) ‘Context matters - the complex interplay between resistome genotypes and resistance phenotypes’, *Current Opinion in Microbiology*, pp. 577–582. doi: 10.1016/j.mib.2012.07.004.

Davies, J. (2010) ‘Origins and evolution of antibiotic resistance.’, *Microbiol. Mol. Biol. Rev.*, 74, pp. 417–433. doi: 10.1128/mmbr.00016-10.

Demain, A. L. and Sanchez, S. (2009) ‘Microbial drug discovery: 80 Years of progress’, *Journal of Antibiotics*. Nature Publishing Group, pp. 5–16. doi: 10.1038/ja.2008.16.

Draper, O. *et al.* (2005) ‘Site-specific recombinase and integrase activities of a conjugative relaxase in recipient cells’, *Proceedings of the National Academy of Sciences of the United States of America*. Proc Natl Acad Sci U S A, 102(45), pp. 16385–16390. doi: 10.1073/pnas.0506081102.

Drawz, S. M. and Bonomo, R. A. (2010) ‘Three decades of β-lactamase inhibitors’, *Clinical Microbiology Reviews*. American Society for Microbiology (ASM), pp. 160–201. doi: 10.1128/CMR.00037-09.

Dubey, G. P. and Ben-Yehuda, S. (2011) 'Intercellular nanotubes mediate bacterial communication', *Cell*. Cell Press, 144(4), pp. 590–600. doi: 10.1016/j.cell.2011.01.015.

Van Duyne, G. D. (2001) 'A structural view of Cre-loxP site-specific recombination', *Annual Review of Biophysics and Biomolecular Structure*. Annu Rev Biophys Biomol Struct, pp. 87–104. doi: 10.1146/annurev.biophys.30.1.87.

Duyne, G. D. Van (2015) 'Cre Recombinase', in *Mobile DNA III*. American Society of Microbiology, pp. 119–138. doi: 10.1128/microbiolspec.mdna3-0014-2014.

Emsley, P. *et al.* (2010) 'Features and development of Coot', *Acta Crystallographica Section D: Biological Crystallography*. International Union of Crystallography, 66(4), pp. 486–501. doi: 10.1107/S0907444910007493.

Ennifar, E. *et al.* (2003) 'Crystal structure of a wild-type Cre recombinase-IoxP synapse reveals a novel spacer conformation suggesting an alternative mechanism for DNA cleavage activation', *Nucleic Acids Research*. Oxford Academic, 31(18), pp. 5449–5460. doi: 10.1093/nar/gkg732.

FDA (2011) *Federal Register :: Determination That ALBAMYCIN (Novobiocin Sodium) Capsule, 250 Milligrams, Was Withdrawn From Sale for Reasons of Safety or Effectiveness*. Available at: https://www.federalregister.gov/documents/2011/01/19/2011-1000/determination-that-albamycin-novobiocin-sodium-capsule-250-milligrams-was-withdrawn-from-sale-for (Accessed: 28 October 2020).

Fernández-López, R. *et al.* (2006) 'Dynamics of the IncW genetic backbone imply general trends in conjugative plasmid evolution', *FEMS Microbiology Reviews*. FEMS Microbiol Rev, pp. 942–966. doi: 10.1111/j.1574-6976.2006.00042.x.

Fleming, A. (1929) 'A. Fleming,1929- antibacterial action of cultures of a penicillium.pdf', *The british journal of experimental pathology*, pp. 226–236.

Foadi, J. *et al.* (2013) 'Clustering procedures for the optimal selection of data sets from multiple crystals in macromolecular crystallography', *Acta Crystallographica Section D: Biological Crystallography*. International Union of Crystallography, 69(8), pp. 1617–1632. doi: 10.1107/S0907444913012274.

Franke, A. E. and Clewell, D. B. (1981) 'Evidence for a chromosome-borne resistance transposon (Tn916) in Streptococcus faecalis that is capable of "conjugal" transfer in the absence of a conjugative plasmid', *Journal of Bacteriology*. J Bacteriol, 145(1), pp. 494–502. doi: 10.1128/jb.145.1.494-502.1981.

Frère, J.-M., Sauvage, E. and Kerff, F. (2015) 'From "An Enzyme Able to Destroy Penicillin" to Carbapenemases: 70 Years of Beta-lactamase Misbehaviour', *Current Drug Targets*. Bentham Science Publishers Ltd., 17(9), pp. 974–982. doi: 10.2174/1389450116666151001112859.

Fyfe, C. *et al.* (2016) 'Resistance to macrolide antibiotics in public health pathogens', *Cold Spring Harbor Perspectives in Medicine*. Cold Spring Harbor Laboratory Press, 6(10), p. a025395. doi: 10.1101/cshperspect.a025395.

Gao, W., Howden, B. P. and Stinear, T. P. (2018) 'Evolution of virulence in Enterococcus faecium, a hospital-adapted opportunistic pathogen', *Current Opinion in Microbiology*. Elsevier Ltd, pp. 76–82. doi: 10.1016/j.mib.2017.11.030.

Garnier, F. *et al.* (2000) 'Characterization of transposon Tn1549, conferring VanB-type resistance in Enterococcus spp.', *Microbiology*, 146(6), pp. 1481–1489. doi: 10.1099/00221287-146-6-1481.

Georgopapadakou, N., Hammarstrom, S. and Strominger, J. L. (1977) 'Isolation of the penicillin binding peptide from D alanine carboxypeptidase of Bacillus subtilis', *Proceedings of the National Academy of Sciences of the United States of America*. Proc Natl Acad Sci U S A, 74(3), pp. 1009–1012. doi: 10.1073/pnas.74.3.1009.

Ghosh, K. *et al.* (2005) 'Peptide trapping of the holliday junction intermediate in Cre-loxP site-specific recombination', *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology, 280(9), pp. 8290–8299. doi: 10.1074/jbc.M411668200.

Ghosh, K., Guo, F. and Van Duyne, G. D. (2007) 'Synapsis of loxP sites by cre recombinase', *Journal of Biological Chemistry*. J Biol Chem, 282(33), pp. 24004–24016. doi: 10.1074/jbc.M703283200.

Gopaul, D. N., Guo, F. and Van Duyne, G. D. (1998) 'Structure of the Holliday junction intermediate in Cre-loxP site-specific recombination', *EMBO Journal*. Wiley-VCH Verlag, 17(14), pp. 4175–4187. doi: 10.1093/emboj/17.14.4175.

Griffith, B. Y. F. (1928) 'The Significance of pneumococcal types. Occurrence of a Variety of Serological Types in the Sputum from an individual case of pneumonia', *Journal of Hygiene*. European Molecular Biology Laboratory, 27(2), pp. 113–159.

Griffith, R. S. (1981) 'Introduction to vancomycin', *Reviews of Infectious Diseases*. Oxford Academic, 3(Supplement_2), pp. S200–S204. doi: 10.1093/clinids/3.Supplement_2.S200.

Grindley, N. D. F., Whiteson, K. L. and Rice, P. A. (2006) 'Mechanisms of site-specific recombination', *Annual Review of Biochemistry*. Annual Reviews , pp. 567–605. doi: 10.1146/annurev.biochem.73.011303.073908.

Grohmann, E. *et al.* (2018) 'Type IV secretion in Gram-negative and Gram-positive bacteria', *Molecular Microbiology*. Blackwell Publishing Ltd, pp. 455–471. doi: 10.1111/mmi.13896.

Guo, F., Gopaul, D. N. and Van Duyne, G. D. (1997) 'Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse', *Nature*. Nature Publishing Group, 389(6646), pp. 40–46. doi: 10.1038/37925.

van Hal, S. J. *et al.* (2016) 'Evolutionary dynamics of Enterococcus faecium reveals complex genomic relationships between isolates with independent emergence of vancomycin resistance', *Microbial genomics*. Microbiology Society, 2(1), p. e000048. doi: 10.1099/mgen.0.000048.

Hall, R. J. *et al.* (2020) 'Horizontal Gene Transfer as a Source of Conflict and Cooperation in Prokaryotes', *Frontiers in Microbiology*. Frontiers Media S.A., 11, p. 1569. doi: 10.3389/fmicb.2020.01569.

Harmer, C. J., Moran, R. A. and Hall, R. M. (2014) 'Movement of IS26-Associated antibiotic resistance genes occurs via a translocatable unit that includes a single IS26 and preferentially inserts adjacent to another IS26', *mBio*. American Society for Microbiology, 5(5). doi: 10.1128/mBio.01801-14.

Hegstad, K. *et al.* (2010) 'Mobile genetic elements and their contribution to the emergence of antimicrobial resistant

Enterococcus faecalis and Enterococcus faecium', *Clinical Microbiology and Infection*. Blackwell Publishing Ltd, pp. 541–554. doi: 10.1111/j.1469-0691.2010.03226.x.

Heide, L. (2009) 'The aminocoumarins: Biosynthesis and biology', *Natural Product Reports*. The Royal Society of Chemistry, pp. 1241–1250. doi: 10.1039/b808333a.

Hickman, A. B. and Dyda, F. (2014) 'Mechanisms of DNA Transposition'. doi: 10.1128/microbiolspec.MDNA3-0034-2014.

Hinerfeld, D. and Churchward, G. (2001) 'Xis protein of the conjugative transposon Tn916 plays dual opposing roles in transposon excision', *Molecular Microbiology*. Mol Microbiol, 41(6), pp. 1459–1467. doi: 10.1046/j.1365-2958.2001.02626.x.

Hoess, R. H., Wierzbicki, A. and Abremski, K. (1986) 'The role of the loxP spacer region in PI site-specific recombination', *Nucleic Acids Research*. Nucleic Acids Res, 14(5), pp. 2287–2300. doi: 10.1093/nar/14.5.2287.

Holdgate, G. A. *et al.* (1997) *The Entropic Penalty of Ordered Water Accounts for Weaker Binding of the Antibiotic Novobiocin to a Resistant Mutant of DNA Gyrase: A Thermodynamic and Crystallographic Study ‡*. Available at: https://pubs.acs.org/sharingguidelines (Accessed: 31 October 2020).

Hooper, D. C. and Jacoby, G. A. (2016) 'Topoisomerase inhibitors: Fluoroquinolone mechanisms of action and resistance', *Cold Spring Harbor Perspectives in Medicine*. Cold Spring Harbor Laboratory Press, 6(9). doi: 10.1101/cshperspect.a025320.

Houghton, J. L. *et al.* (2010) 'The future of aminoglycosides: The end or renaissance?', *ChemBioChem*. John Wiley & Sons, Ltd, pp. 880–902. doi: 10.1002/cbic.200900779.

Ilangovan, A., Connery, S. and Waksman, G. (2015) 'Structural biology of the Gram-negative bacterial conjugation systems', *Trends in Microbiology*. Elsevier Ltd, pp. 301–310. doi: 10.1016/j.tim.2015.02.012.

Jacoby, G. A. (2006) 'β-lactamase nomenclature', *Antimicrobial Agents and Chemotherapy*. American Society for Microbiology Journals, pp. 1123–1129. doi: 10.1128/AAC.50.4.1123-1129.2006.

Jensen, L. B. *et al.* (2010) 'A classification system for plasmids from enterococci and other Gram-positive bacteria', *Journal of Microbiological Methods*. Elsevier, 80(1), pp. 25–43. doi: 10.1016/j.mimet.2009.10.012.

Jim O'Neill (2014) *Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations*.

Johnson, C. M. and Grossman, A. D. (2015) 'Integrative and Conjugative Elements (ICEs): What They Do and How They Work', *Annual Review of Genetics*, 49(1), pp. 577–601. doi: 10.1146/annurev-genet-112414-055018.

Jubeh, B., Breijyeh, Z. and Karaman, R. (2020) 'Resistance of gram-positive bacteria to current antibacterial agents and overcoming approaches', *Molecules*. MDPI AG. doi: 10.3390/molecules25122888.

Kabsch, W. *et al.* (2010) '*XDS*', *Acta Crystallographica Section D Biological Crystallography*, 66(2), pp. 125–132. doi:

10.1107/S0907444909047337.

Kahne, D. *et al.* (2005) 'Glycopeptide and lipoglycopeptide antibiotics', *Chemical Reviews*. American Chemical Society , pp. 425–448. doi: 10.1021/cr030103a.

Karakonstantis, S., Kritsotakis, E. I. and Gikas, A. (2020) 'Treatment options for K. pneumoniae, P. aeruginosa and A. baumannii co-resistant to carbapenems, aminoglycosides, polymyxins and tigecycline: an approach based on the mechanisms of resistance to carbapenems', *Infection*. Springer Berlin Heidelberg, (0123456789). doi: 10.1007/s15010-020-01520-6.

Karplus, P. A. and Diederichs, K. (2012) 'Linking crystallographic model and data quality', *Science*. American Association for the Advancement of Science, 336(6084), pp. 1030–1033. doi: 10.1126/science.1218231.

Kim, S. and Gardner, J. F. (2011) 'Resolution of holliday junction recombination intermediates by wild-type and mutant IntDOT proteins', *Journal of Bacteriology*, 193(6), pp. 1351–1358. doi: 10.1128/JB.01465-10.

Krissinel, E. & Henrick, K. 2004. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallographica D Biological Crystallography*, 60, 2256-68. doi: 10.1107/S0907444904026460

De La Cruz, F. *et al.* (2010) 'Conjugative DNA metabolism in Gram-negative bacteria', *FEMS Microbiology Reviews*. FEMS Microbiol Rev, pp. 18–40. doi: 10.1111/j.1574-6976.2009.00195.x.

Lambertsen, L. *et al.* (2018) 'Conjugative transposition of the vancomycin resistance carrying Tn1549: enzymatic requirements and target site preferences', *Molecular Microbiology*, 107(5), pp. 639–658. doi: 10.1111/mmi.13905.

Landy, A. (1989) 'Dynamic, Structural, and Regulatory Aspects of lambda Site-Specific Recombination', *Annual Review of Biochemistry*. Annual Reviews, 58(1), pp. 913–941. doi: 10.1146/annurev.bi.58.070189.004405.

Landy, A. (2015) 'The λ Integrase Site-specific Recombination Pathway', in *Mobile DNA III*. American Society of Microbiology, pp. 91–118. doi: 10.1128/microbiolspec.mdna3-0051-2014.

Lang, A. S., Zhaxybayeva, O. and Beatty, J. T. (2012) 'Gene transfer agents: Phage-like elements of genetic exchange', *Nature Reviews Microbiology*. Nature Publishing Group, pp. 472–482. doi: 10.1038/nrmicro2802.

Laponogov, I. *et al.* (2009) 'Structural insight into the quinolone-DNA cleavage complex of type IIA topoisomerases', *Nature Structural and Molecular Biology*. Nature Publishing Group, 16(6), pp. 667–669. doi: 10.1038/nsmb.1604.

Launay, A. *et al.* (2006) 'Transfer of vancomycin resistance transposon Tn1549 from Clostridium symbiosum to Enterococcus spp. in the gut of gnotobiotic mice', *Antimicrobial Agents and Chemotherapy*. American Society for Microbiology Journals, 50(3), pp. 1054–1062. doi: 10.1128/AAC.50.3.1054-1062.2006.

Laxmikanthan, G. *et al.* (2016) 'Structure of a holliday junction complex reveals mechanisms governing a highly regulated DNA transaction', *eLife*, 5(MAY2016), pp. 1–23. doi: 10.7554/eLife.14313.

Leclercq, R. *et al.* (1988) 'Plasmid-mediated resistance to vancomycin and teicoplanin in Enterococcus faecium', *New*

England Journal of Medicine. Massachusetts Medical Society , 319(3), pp. 157–161. doi: 10.1056/NEJM198807213190307.

Lederberg, J. and Tatum, E. L. (1946) 'Gene recombination in Escherichia coli [23]', *Nature*. Eslick, A., Proc. Linn. Soc, p. 558. doi: 10.1038/158558a0.

Lee, K. Y., Hopkins, J. D. and Syvanen, M. (1990) 'Direct involvement of IS26 in an antibiotic resistance operon', *Journal of Bacteriology*. J Bacteriol, 172(6), pp. 3229–3236. doi: 10.1128/jb.172.6.3229-3236.1990.

Lee, L. and Sadowski, P. D. (2003) 'Sequence of the loxP site determines the order of strand exchange by the Cre recombinase', *Journal of Molecular Biology*. Academic Press, 326(2), pp. 397–412. doi: 10.1016/S0022-2836(02)01429-8.

Lewis, J. A. and Hatfull, G. F. (2001) 'Control of directionality in integrase-mediated recombination: Examination of recombination directionality factors (RDFs) including Xis and Cox proteins', *Nucleic Acids Research*. Oxford University Press, 29(11), pp. 2205–2216. doi: 10.1093/nar/29.11.2205.

Li, Y. *et al.* (2009) 'New developments in Hsp90 inhibitors as anti-cancer therapeutics: Mechanisms, clinical perspective and more potential', *Drug Resistance Updates*. NIH Public Access, 12(1–2), pp. 17–27. doi: 10.1016/j.drup.2008.12.002.

Liebschner, D. *et al.* (2019) 'Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix', *Acta Crystallographica Section D: Structural Biology*, 75, pp. 861–877. doi: 10.1107/S2059798319011471.

Liu, M. *et al.* (2019) 'ICEberg 2.0: An updated database of bacterial integrative and conjugative elements', *Nucleic Acids Research*. Oxford University Press, 47(D1), pp. D660–D665. doi: 10.1093/nar/gky1123.

Liu, Y. and West, S. C. (2004) 'Happy Hollidays: 40th Anniversary of the Holliday junction', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, pp. 937–944. doi: 10.1038/nrm1502.

Mahillon, J. and Chandler, M. (1998) 'Insertion Sequences', *Microbiology and Molecular Biology Reviews*. American Society for Microbiology, 62(3), pp. 725–774. doi: 10.1128/mmbr.62.3.725-774.1998.

Marra, D. *et al.* (1999) 'The frequency of conjugative transposition of Tn916 is not determined by the frequency of excision', *Journal of Bacteriology*. American Society for Microbiology, 181(17), pp. 5414–5418. doi: 10.1128/jb.181.17.5414-5418.1999.

Martin, S. S. *et al.* (2002) 'The order of strand exchanges in Cre-LoxP recombination and its basis suggested by the crystal structure of a Cre-LoxP holliday junction complex', *Journal of Molecular Biology*. Academic Press, 319(1), pp. 107–127. doi: 10.1016/S0022-2836(02)00246-2.

Martin, S. S., Chu, V. C. and Baldwin, E. (2003) 'Modulation of the active complex assembly and turnover rate by protein-DNA interactions in Cre-LoxP recombination', *Biochemistry*. American Chemical Society , 42(22), pp. 6814–6826. doi: 10.1021/bi0272306.

Maxwell, A. and Lawson, D. (2003) 'The ATP-Binding Site of Type II Topoisomerases as a Target for Antibacterial Drugs', *Current Topics in Medicinal Chemistry*. Bentham Science Publishers Ltd., 3(3), pp. 283–303. doi: 10.2174/1568026033452500.

Mazel, D. and Davies, J. (1999) 'Antibiotic resistance in microbes', in *Cellular and Molecular Life Sciences*. Birkhauser Verlag Basel, pp. 742–754. doi: 10.1007/s000180050021.

Mcclintock, B. (1983) *THE SIGNIFICANCE OF RESPONSES OF THE GENOME TO CHALLENGE*. Available at: https://www.nobelprize.org/uploads/2018/06/mcclintock-lecture.pdf

McCoy, A. J. *et al.* (2007) 'Phaser crystallographic software', *Journal of Applied Crystallography*. International Union of Crystallography, 40(4), pp. 658–674. doi: 10.1107/S0021889807021206.

Mell, J. C. and Redfield, R. J. (2014) 'Natural competence and the evolution of DNA uptake specificity', *Journal of Bacteriology*. American Society for Microbiology, pp. 1471–1483. doi: 10.1128/JB.01293-13.

Miller, W. R., Bayer, A. S. and Arias, C. A. (2016) 'Mechanism of action and resistance to daptomycin in Staphylococcus aureus and enterococci', *Cold Spring Harbor Perspectives in Medicine*. Cold Spring Harbor Laboratory Press, 6(11). doi: 10.1101/cshperspect.a026997.

Moazed, D. and Noller, H. F. (1987) 'Interaction of antibiotics with functional sites in 16S ribosomal RNA', *Nature*. Nature, 327(6121), pp. 389–394. doi: 10.1038/327389a0.

Morvan, A. *et al.* (2010) 'Antimicrobial resistance of Listeria monocytogenes strains isolated from humans in France', *Antimicrobial Agents and Chemotherapy*. Antimicrob Agents Chemother, 54(6), pp. 2728–2731. doi: 10.1128/AAC.01557-09.

Murray, B. E. (1990) 'The life and times of the enterococcus', *Clinical Microbiology Reviews*, 3(1), pp. 46–65. doi: 10.1128/CMR.3.1.46.

Musser, J. M. (1995) 'Antimicrobial agent resistance in mycobacteria: molecular genetic insights.', *Clinical microbiology reviews*. American Society for Microbiology, 8(4), pp. 496–514. doi: 10.1128/cmr.8.4.496-514.1995.

Namvar, A. E. *et al.* (2014) 'Clinical characteristics of Staphylococcus epidermidis: a systematic review.', *GMS hygiene and infection control*. GMS Hyg Infect Control, 9(3), p. Doc23. doi: 10.3205/dgkh000243.

Needham, C., Noble, W. C. and Dyke, K. G. H. (1995) 'The staphylococcal insertion sequence is257is active', *Plasmid*. Academic Press, 34(3), pp. 198–205. doi: 10.1006/plas.1995.0005.

Nicolas, E. *et al.* (2015) 'The Tn3-family of Replicative Transposons', in *Mobile DNA III*. American Society of Microbiology, pp. 693–726. doi: 10.1128/microbiolspec.mdna3-0060-2014.

Nunes-Düby, S. E. *et al.* (1998) 'Similarities and differences among 105 members of the Int family of site-specific recombinases', *Nucleic Acids Research*. Nucleic Acids Res, 26(2), pp. 391–406. doi: 10.1093/nar/26.2.391.

Nunes-Düby, S. E., Azaro, M. A. and Landy, A. (1995) 'Swapping DNA strands and sensing homology without branch

migration in λ site-specific recombination', *Current Biology*. Curr Biol, 5(2), pp. 139–148. doi: 10.1016/S0960-9822(95)00035-2.

Partridge, S. R. *et al.* (2018) 'Mobile genetic elements associated with antimicrobial resistance', *Clinical Microbiology Reviews*. American Society for Microbiology. doi: 10.1128/CMR.00088-17.

Peters, J. E. and Craig, N. L. (2001) 'Tn7: Smarter than we thought', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 2(11), pp. 806–814. doi: 10.1038/35099006.

Peterson, E. and Kaur, P. (2018) 'Antibiotic resistance mechanisms in bacteria: Relationships between resistance determinants of antibiotic producers, environmental bacteria, and clinical pathogens', *Frontiers in Microbiology*. Frontiers Media S.A., 9(NOV), p. 2928. doi: 10.3389/fmicb.2018.02928.

Pettersen EF, Goddard TD, Huang CC, Meng EC, Couch GS, Croll TI, Morris JH, Ferrin TE. (2020) 'UCSF ChimeraX: Structure visualization for researchers, educators, and developers', Protein Sci. doi: 10.1002/pro.3943.

Pinkney, J. N. M. *et al.* (2012) 'Capturing reaction paths and intermediates in Cre-loxP recombination using single-molecule fluorescence', *Proceedings of the National Academy of Sciences of the United States of America*. Proc Natl Acad Sci U S A, 109(51), pp. 20871–20876. doi: 10.1073/pnas.1211922109.

Poyart-Salmeron, C. *et al.* (1989) 'Molecular characterization of two proteins involved in the excision of the conjugative transposon Tn1545: Homologies with other site-specific recombinases', *EMBO Journal*. EMBO J, 8(8), pp. 2425–2433. doi: 10.1002/j.1460-2075.1989.tb08373.x.

Radman-Livaja, M. *et al.* (2005) 'Architecture of recombination intermediates visualized by in-gel FRET of λ integrase-Holliday junction-arm DNA complexes', *Proceedings of the National Academy of Sciences of the United States of America*. Proc Natl Acad Sci U S A, 102(11), pp. 3913–3920. doi: 10.1073/pnas.0500844102.

Razavi, M. *et al.* (2020) 'The Association between Insertion Sequences and Antibiotic Resistance Genes', *mSphere*. Edited by T. M. LaPara. mSphere, 5(5). doi: 10.1128/mSphere.00418-20.

Redzej, A. *et al.* (2013) 'Structure of a translocation signal domain mediating conjugative transfer by type IV secretion systems', *Molecular Microbiology*. John Wiley & Sons, Ltd, 89(2), pp. 324–333. doi: 10.1111/mmi.12275.

Rice, L. B. (2008) 'Federal funding for the study of antimicrobial resistance in nosocomial pathogens: No ESKAPE', *Journal of Infectious Diseases*. J Infect Dis, pp. 1079–1081. doi: 10.1086/533452.

Ringwald, K., Yoneji, S. and Gardner, J. (2017) 'Resolution of mismatched overlap Holliday junction intermediates by the tyrosine recombinase IntDOT', *Journal of Bacteriology*. American Society for Microbiology, 199(10). doi: 10.1128/JB.00873-16.

Roberts, A. P. and Mullany, P. (2009) 'A modular master on the move: the Tn916 family of mobile genetic elements', *Trends in Microbiology*. Elsevier Current Trends, pp. 251–258. doi: 10.1016/j.tim.2009.03.002.

Roberts, A. P. and Mullany, P. (2011) 'Tn916-like genetic elements: A diverse group of modular mobile elements

conferring antibiotic resistance', *FEMS Microbiology Reviews*, 35(5), pp. 856–871. doi: 10.1111/j.1574-6976.2011.00283.x.

Rocco, J. M. and Churchward, G. (2006) 'The integrase of the conjugative transposon Tn916 directs strand- and sequence-specific cleavage of the origin of conjugal transfer, oriT, by the endonuclease Orf20', *Journal of Bacteriology*. J Bacteriol, 188(6), pp. 2207–2213. doi: 10.1128/JB.188.6.2207-2213.2006.

Rodríguez, D. D. *et al.* (2009) 'Crystallographic ab initio protein structure solution below atomic resolution', *Nature Methods*. Nature Publishing Group, 6(9), pp. 651–653. doi: 10.1038/nmeth.1365.

Rubinstein, E. and Keynan, Y. (2014) 'Vancomycin revisited - 60 years later', *Frontiers in Public Health*. Frontiers Media S. A, p. 217. doi: 10.3389/fpubh.2014.00217.

Rubio-Cosials, A. *et al.* (2018) 'Transposase-DNA Complex Structures Reveal Mechanisms for Conjugative Transposition of Antibiotic Resistance', *Cell*. Elsevier, 173(1), pp. 208-220.e20. doi: 10.1016/j.cell.2018.02.032.

Rupp, B. (2010) *BIOMOLECULAR CRYSTALLOGRAPHY: Principles, Practice, and Application to Structural Biology (English Edition) eBook: Rupp, Bernhard: Amazon.de: Kindle Store*. Available at: https://www.amazon.de/-/en/gp/product/B00AQNVTOW/ref=ppx_yo_dt_b_d_asin_image_o02?ie=UTF8&psc=1 (Accessed: 5 November 2020).

Sam, M. D. *et al.* (2004) 'Crystal structure of the excisionase-DNA complex from bacteriophage lambda', *Journal of Molecular Biology*, 338(2), pp. 229–240. doi: 10.1016/j.jmb.2004.02.053.

Sámano-Sánchez, H. and Gibson, T. J. (2020) 'Mimicry of Short Linear Motifs by Bacterial Pathogens: A Drugging Opportunity', *Trends in Biochemical Sciences*. Elsevier Ltd, pp. 526–544. doi: 10.1016/j.tibs.2020.03.003.

Schleifer, K. H. and Kilpper-Balz, R. (1984) 'Transfer of Streptococcus faecalis and Streptococcus faecium to the genus Enterococcus nom. rev. as Enterococcus faecalis comb. nov. and Enterococcus faecium comb. nov.', *International Journal of Systematic Bacteriology*, 34(1), pp. 31–34. doi: 10.1099/00207713-34-1-31.

Schwarz, F. V., Perreten, V. and Teuber, M. (2001) 'Sequence of the 50-kb conjugative multiresistance plasmid pRE25 from Enterococcus faecalis RE25', *Plasmid*. Academic Press Inc., 46(3), pp. 170–187. doi: 10.1006/plas.2001.1544.

Schwarz, S. *et al.* (2016) 'Lincosamides, streptogramins, phenicols, and pleuromutilins: Mode of action and mechanisms of resistance', *Cold Spring Harbor Perspectives in Medicine*. Cold Spring Harbor Laboratory Press, 6(11), p. a027037. doi: 10.1101/cshperspect.a027037.

Sievers, F. *et al.* (2011) 'Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega', *Molecular Systems Biology*. John Wiley & Sons, Ltd, 7(1), p. 539. doi: 10.1038/msb.2011.75.

Siguier, P. *et al.* (2015) 'Everyman's Guide to Bacterial Insertion Sequences'. doi: 10.1128/microbiolspec.MDNA3-0030-2014.

Smillie, C. *et al.* (2010) 'Mobility of Plasmids', *Microbiology and Molecular Biology Reviews*. American Society for Microbiology, 74(3), pp. 434–452. doi: 10.1128/mmbr.00020-10.

Spigaglia, P. (2016) ‘ Recent advances in the understanding of antibiotic resistance in Clostridium difficile infection ’, *Therapeutic Advances in Infectious Disease*. SAGE Publications, 3(1), pp. 23–42. doi: 10.1177/2049936115622891.

Strahilevitz, J. *et al.* (2009) ‘Plasmid-mediated quinolone resistance: A multifaceted threat’, *Clinical Microbiology Reviews*. Clin Microbiol Rev, pp. 664–689. doi: 10.1128/CMR.00016-09.

Su, Y. A. and Clewell, D. B. (1993) ‘Characterization of the Left 4 kb of Conjugative Transposon Tn916: Determinants Involved in Excision’, *Plasmid*. Plasmid, 30(3), pp. 234–250. doi: 10.1006/plas.1993.1055.

Sugden, R., Kelly, R. and Davies, S. (2016) ‘Combatting antimicrobial resistance globally’, *Nature Microbiology*. Nature Publishing Group, p. 16187. doi: 10.1038/nmicrobiol.2016.187.

Tacconelli, E. *et al.* (2017) *GLOBAL PRIORITY LIST OF ANTIBIOTIC-RESISTANT BACTERIA TO GUIDE RESEARCH, DISCOVERY, AND DEVELOPMENT OF NEW ANTIBIOTICS*. Geneva. Available at: http://www.cdc.gov/drugresistance/threat-report-2013/ (Accessed: 20 August 2020).

Theuretzbacher, U. (2017) ‘Global antimicrobial resistance in Gram-negative pathogens and clinical need’, *Current Opinion in Microbiology*. Elsevier Ltd, pp. 106–112. doi: 10.1016/j.mib.2017.10.028.

Thomas, C. M. (2000) ‘Paradigms of plasmid organization’, *Molecular Microbiology*. Mol Microbiol, pp. 485–491. doi: 10.1046/j.1365-2958.2000.02006.x.

Thomas, C. M. and Nielsen, K. M. (2005) ‘Mechanisms of, and barriers to, horizontal gene transfer between bacteria’, *Nature Reviews Microbiology*. Nature Publishing Group, pp. 711–721. doi: 10.1038/nrmicro1234.

Thorn, A. and Sheldrick, G. M. (2013) ‘Extending molecular-replacement solutions with SHELXE’, *Acta Crystallographica Section D: Biological Crystallography*. International Union of Crystallography, 69(11), pp. 2251–2256. doi: 10.1107/S0907444913027534.

Tipper, D. J. and Strominger, J. L. (1965) ‘Mechanism of action of penicillins: a proposal based on their structural similarity to acyl-D-alanyl-D-alanine.’, *Proceedings of the National Academy of Sciences of the United States of America*. Proc Natl Acad Sci U S A, 54(4), pp. 1133–1141. doi: 10.1073/pnas.54.4.1133.

Touchon, M., Moura de Sousa, J. A. and Rocha, E. P. (2017) ‘Embracing the enemy: The diversification of microbial gene repertoires by phage-mediated horizontal gene transfer’, *Current Opinion in Microbiology*. Elsevier Ltd, pp. 66–73. doi: 10.1016/j.mib.2017.04.010.

Trimble, M. J. *et al.* (2016) ‘Polymyxin: Alternative mechanisms of action and resistance’, *Cold Spring Harbor Perspectives in Medicine*. Cold Spring Harbor Laboratory Press, 6(10), p. a025288. doi: 10.1101/cshperspect.a025288.

Waksman, S. A. (1952) *Streptomycin: background, isolation, properties, and utilization*. Available at: https://www.nobelprize.org/uploads/2018/06/waksman-lecture.pdf (Accessed: 18 August 2020).

Waksman, S. A. and Flynn, J. E. (1973) ‘History of the word “antibiotic”.’, *Journal of the history of medicine and allied*

sciences. Oxford Academic, 28(3), pp. 284–286. doi: 10.1093/jhmas/xxviii.3.284.

Waldor, M. K. and Mekalanos, J. J. (1996) 'Lysogenic conversion by a filamentous phage encoding cholera toxin', *Science*. American Association for the Advancement of Science, 272(5270), pp. 1910–1913. doi: 10.1126/science.272.5270.1910.

Walsh, C. (2000) 'Molecular mechanisms that confer antibacterial drug resistance', *Nature*. Nature Publishing Group, pp. 775–781. doi: 10.1038/35021219.

Waterhouse, A. M. *et al.* (2009) 'Jalview Version 2-A multiple sequence alignment editor and analysis workbench', *Bioinformatics*. Oxford Academic, 25(9), pp. 1189–1191. doi: 10.1093/bioinformatics/btp033.

WHO (2015) *Global action plan on antimicrobial resistance*. Available at: https://www.who.int/publications/i/item/global-action-plan-on-antimicrobial-resistance (Accessed: 18 August 2020).

WHO (2019) *No time to wait: securing the future from drug-resistant infections. Report to the Secretary-General of the United Nations*. Available at: https://www.who.int/antimicrobial-resistance/interagency-coordination-group/IACG_final_report_EN.pdf?ua=1 (Accessed: 19 August 2020).

WHO (2020) *Antibiotic Resistance*. Available at: https://www.who.int/news-room/fact-sheets/detail/antibiotic-resistance (Accessed: 19 August 2020).

Wirth, D. *et al.* (2007) 'Road to precision: recombinase-based targeting technologies for genome engineering', *Current Opinion in Biotechnology*. Curr Opin Biotechnol, pp. 411–419. doi: 10.1016/j.copbio.2007.07.013.

WM, K. and LA, R. (1942) 'Sulfonamide Resistance', *California and western medicine*. Cal West Med, 57(3). Available at: https://pubmed.ncbi.nlm.nih.gov/18746315/ (Accessed: 16 September 2020).

Wood, M. M. and Gardner, J. F. (2015) 'The Integration and Excision of CTnDOT', in *Mobile DNA III*. American Society of Microbiology, pp. 183–198. doi: 10.1128/microbiolspec.mdna3-0020-2014.

Wozniak, R. A. F. and Waldor, M. K. (2010) 'Integrative and conjugative elements: Mosaic mobile genetic elements enabling dynamic lateral gene flow', *Nature Reviews Microbiology*. Nat Rev Microbiol, pp. 552–563. doi: 10.1038/nrmicro2382.

Zechner, E. L., Lang, S. and Schildbach, J. F. (2012) 'Assembly and mechanisms of bacterial type IV secretion machines', *Philosophical Transactions of the Royal Society B: Biological Sciences*. Royal Society, pp. 1073–1087. doi: 10.1098/rstb.2011.0207.

Zeng, D. *et al.* (2016) 'Approved glycopeptide antibacterial drugs: Mechanism of action and resistance', *Cold Spring Harbor Perspectives in Medicine*. Cold Spring Harbor Laboratory Press, 6(12), p. a026989. doi: 10.1101/cshperspect.a026989.

Zeng, L. *et al.* (2010) 'Decision Making at a Subcellular Level Determines the Outcome of Bacteriophage Infection', *Cell*. Cell Press, 141(4), pp. 682–691. doi: 10.1016/j.cell.2010.03.034.

Zhou, X. *et al.* (2020) 'Enterococcus faecium : from microbiological insights to practical recommendations for infection control and diagnostics'. Antimicrobial Resistance & Infection Control, 1, pp. 1–13.

Zhu, W. *et al.* (2010) 'Dissemination of an Enterococcus Inc18-like vanA plasmid associated with vancomycin-resistant Staphylococcus aureus', *Antimicrobial Agents and Chemotherapy*. American Society for Microbiology Journals, 54(10), pp. 4314–4320. doi: 10.1128/AAC.00185-10.

Zwart, Grosse-Kunstleve, A. (2005) 'Xtriage and Fest: automatic assessment of X-ray data and substructure structure factor estimation', *CCP4 newsletter*. Available at: http://www.ccp4.ac.uk/newsletters/newsletter43/articles/PHZ_RWGK_PDA.pdf.

6 APPENDIX

Table 6-1: List of oligonucleotides used for activity and crystallization experiments with Int^{82N} and Int^{82N}390C. The crossover region is in lowercase. /5Phos/ = 5'-phosphorylated oligonucleotides.

Oligo name	Sequence (5'-3')	Length in nt
CI5full_bot	/5Phos/CTAAAATCCCATATAATTTTgctatAAAATTTTAGGTTATCGCT	44
CI5top_1-1	/5Phos/TGCGATAACCTAAAATTTTa	20
CI5top_2-1	/5Phos/tagcAAAATTATATGGGATTTTAG	24
CI5top_2-2	/5Phos/TatagcAAAATTATATGGGATTTTAG	26
CI5top_2-3	/5Phos/atagcAAAATTATATGGGATTTTAG	25
CI5bot_1-1	/5Phos/CTAAAATCCCATATAATTTTg	21
CI5bot_2-2	/5Phos/TgctatAAAATTTTAGGTTATCGCT	25
CI6atop_1-1	TGCGATAACCTAAAATTTTa	20
CI6atop_2-1	ttttcAAAATTATATGGGATTTTAG	25
CI6atop_2-2	TattttcAAAATTATATGGGATTTTAG	27
CI6abot_1-1	CTAAAATCCCATATAATTTTg	21
CI6abot_2-1	aaaatAAAATTTTAGGTTATCGCT	24
CI6abot_2-2	TgaaaatAAAATTTTAGGTTATCGCT	26
CI6btop_1-1	TGCGATAACCTAAAATTTTc	20
CI6btop_2-1	cctttAAAATTATATGGGATTTTAG	25
CI6bbot_1-1	CTAAAATCCCATATAATTTTa	21
CI6bbot_2-1	aagggAAAATTTTAGGTTATCGCT	24
CI6afull_bot	CTAAAATCCCATATAATTTTgaaaatAAAATTTTAGGTTATCGCT	45
palCItop_1-1	/5Phos/AGCGATAACCTAAAATTTT	19
palCItop_1-2	/5Phos/AGCGATAACCTAAAATTTTg	20
palCIbot_1-1	/5Phos/gatatcAAAATTTTAGGTTATCGCT	25
palCIbot_1-2	/5Phos/TgatatcAAAATTTTAGGTTATCGCT	26
palCI2top_1-1	/5Phos/gaattcAAAATTATATGGGATTTTAG	26
palCI2top_1-2	/5Phos/TgaattcAAAATTATATGGGATTTTAG	27
palCI2bot_1-1	/5Phos/CTAAAATCCCATATAATTTT	20
palCI2bot_1-2	/5Phos/CTAAAATCCCATATAATTTTg	21
CI5full_botT	/5Phos/TCTAAAATCCCATATAATTTTgctatAAAATTTTAGGTTATCGCT	45
CI5top_1-2	/5Phos/AAGCGATAACCTAAAATTTTa	21
CI5top_2-1	/5Phos/tagcAAAATTATATGGGATTTTAG	24

Table 6-2: List of oligonucleotides used for binding and crystallization experiments with Int^{82N}R225K. Overhangs are highlighted in bold.

Oligo name	Sequence (5'-3')	Length in nt
HJ1a	TGAACAGCCCACAAAATTTTGAAAATAAAATTTTAGGTTATCGCT	45
HJ1b	AGCGATAACCTAAAATTTTATTAATAAAATTATATGGGATTTTAG	45
HJ1c	AAATTTTTTACCGGGTTTTGAATTCAAAATTTTGTGGGCTGTTCA	45
HJ1d	CTAAAATCCCATATAATTTTATTTTCAAAACCCGGTAAAAAATTT	45
HJ1aT	TTGAACAGCCCACAAAATTTTGAAAATAAAATTTTAGGTTATCGCT	46
HJ1bT	TAGCGATAACCTAAAATTTTATTAATAAAATTATATGGGATTTTAG	46
HJ1cT	TAAATTTTTTACCGGGTTTTGAATTCAAAATTTTGTGGGCTGTTCA	46
HJ1dT	TCTAAAATCCCATATAATTTTATTTTCAAAACCCGGTAAAAAATTT	46
HJ1aT35	TGCCCACAAAATTTTGAAAATAAAATTTTAGGTTA	35

HJ1bT35	TTAACCTAAAATTTTATTAATAAAATTATATGGGA	35
HJ1cT35	TTTTTACCGGGTTTTGAATTCAAAATTTTGTGGGC	35
HJ1dT35	TTCCCATATAATTTTATTTTCAAAACCCGGTAAAA	35
HJ1aGA	GGAACAGCCCACAAAATTTTGAAAATAAAATTTTAGGTTATCGCA	45
HJ1bGA	GGCGATAACCTAAAATTTTATTAATAAAATTATATGGGATTTTAA	45
HJ1cGA	GAATTTTTTACCGGGTTTTGAATTCAAAATTTTGTGGGCTGTTCA	45
HJ1dGA	GTAAAATCCCATATAATTTTATTTTCAAAACCCGGTAAAAAATTA	45
HJ1aT41	TACAGCCCACAAAATTTTGAAAATAAAATTTTAGGTTATCG	41
HJ1bT41	TCGATAACCTAAAATTTTATTAATAAAATTATATGGGATTT	41
HJ1cT41	TATTTTTTACCGGGTTTTGAATTCAAAATTTTGTGGGCTGT	41
HJ1dT41	TAAATCCCATATAATTTTATTTTCAAAACCCGGTAAAAAAT	41
HJ6a45	AGCGATAACCTAAAATTTTGGGTAAAAAATTTTGTGGGCTGTTCA	45
HJ6b45	CTAAAATCCCATATAATTTTGATCCCAAAATTTTAGGTTATCGCT	45
HJ6c45	TGAACAGCCCACAAAATTTTTTAAATAAAACCCGGTAAAAAATTT	45
HJ6d45	AAATTTTTTACCGGGTTTTATTATCAAAATTATATGGGATTTTAG	45
HJ6a44	AGCGATAACCTAAAATTTTGGGTAAAAAATTTTGTGGGCTGTTC	44
HJ6b44	TAAAATCCCATATAATTTTGATCCCAAAATTTTAGGTTATCGCT	44
HJ6c44	GAACAGCCCACAAAATTTTTTAAATAAAACCCGGTAAAAAATTT	44
HJ6d44	AAATTTTTTACCGGGTTTTATTATCAAAATTATATGGGATTTTA	44
HJ1-8a	TGAACAGGGCATATAATTTTGAAAATAAAATTTTAGGTTATCGCT	45
HJ1-8b	AGCGATAACCTAAAATTTTATTAATAAAATTATATGGGATTTTAG	45
HJ1-8c	CTAAAATCCCATATAATTTTATTTTCAAAATTTTAGGGCTGTTCA	45
HJ1-8d	TGAACAGCCCTAAAATTTTGAATTCAAAATTATATGCCCTGTTCA	45
HJ1aTA	TGAACAGCCCACAAAATTTTGAAAATAAAATTTTAGGTTATCGC	44
HJ1bTA	AGCGATAACCTAAAATTTTATTAATAAAATTATATGGGATTTTA	44
HJ1cTA	AAATTTTTTACCGGGTTTTGAATTCAAAATTTTGTGGGCTGTTC	44
HJ1dTA	TTAAAATCCCATATAATTTTATTTTCAAAACCCGGTAAAAAATT	44
HJ1aTACR5	TGAACAGCCCACAAAATTTTcgtatAAAATTTTAGGTTATCGC	43
HJ1bTACR5	AGCGATAACCTAAAATTTTatagcAAAATTATATGGGATTTTA	43
HJ1cTACR5	AAATTTTTTACCGGGTTTTgctcgAAAATTTTGTGGGCTGTTC	43
HJ1dTACR5	TTAAAATCCCATATAATTTTgcagcAAAACCCGGTAAAAAATT	43
HJ1aTACR5B	TTGAACAGCCCACAAAATTTTcgtatAAAATTTTAGGTTATCGC	44
HJ1bTACR5B	AGCGATAACCTAAAATTTTatagcAAAATTATATGGGATTTTAG	44
HJ1cTACR5B	AAATTTTTTACCGGGTTTTgctcgAAAATTTTGTGGGCTGTTCA	44
HJ1dTACR5B	TCTAAAATCCCATATAATTTTgcagcAAAACCCGGTAAAAAATT	44
HJ1aTA34	TGCCCACAAAATTTTGAAAATAAAATTTTAGGTT	34
HJ1bTA34	AAACCTAAAATTTTATTAATAAAATTATATGGGA	34
HJ1cTA34	ATTTACCGGGTTTTGAATTCAAAATTTTGTGGGC	34
HJ1dTA34	TTCCCATATAATTTTATTTTCAAAACCCGGTAAA	34
HJ1aTAsy2	TAAATTTTTCATATAATTTTGAAAATAAAATTTTAGGTTATCGC	44
HJ1bTAsy2	AGCGATAACCTAAAATTTTATTAATAAAATTATATGGGATTTTA	44
HJ1cTAsy2	AAACAGCCCCTAAAATTTTGAATTCAAAATTATATGAAAAATTT	44
HJ1dTAsy2	TTAAAATCCCATATAATTTTATTTTCAAAATTTTAGGGGCTGTT	44
HJ1aTAsy3	TAAATTTTTCATATAATTTTGAAAATAAAATTATATGTTATCGC	44
HJ1bTAsy3	AGCGATAACATATAATTTTATTAATAAAATTATATGGGATTTTA	44
HJ1cTAsy3	AAACAGCCCATATAATTTTGAATTCAAAATTATATGAAAAATTT	44
HJ1dTAsy3	TTAAAATCCCATATAATTTTATTTTCAAAATTATATGGGCTGTT	44
HJ1aTAsy3-1.2	TGAACAGCCCATATAATTTTGAAAATAAAATTATATGTTATCGC	44

HJ1bTAsy3-1.2	AGCGATAACATATAATTTTATTAATAAAATTATATGGGATTTTA	44
HJ1cTAsy3-1.2	AAATTTTTCATATAATTTTGAATTCAAAATTATATGGGCTGTTC	44
HJ1dTAsy3-1.2	TTAAAATCCCATATAATTTTATTTTCAAAATTATATGAAAAATT	44
HJ1aTAsy3-2	TAAATTTTTCATATAATTTTGAAAATAAAATTATATGTTATCGCT	45
HJ1bTAsy3-2	AAGCGATAACATATAATTTTATTAATAAAATTATATGGGATTTTA	45
HJ1cTAsy3-2	AGAACAGCCCATATAATTTTGAATTCAAAATTATATGAAAAATTT	45
HJ1dTAsy3-2	TTAAAATCCCATATAATTTTATTTTCAAAATTATATGGGCTGTTC	45

Table 6-3: List of various Int^{82N} and Int^{82N}390C-suicide DNA-complexes used for high-throughput crystallization trials. The list includes the names of the different complexes, the commercial screens, temperature, and the specific crystallization buffer used for each complex.

Complex	Screens	T	Crystallization Buffer
Int^{82N}-sCI5	JCSG+, PEGs	7 °C, 20 °C	250 mM NaCl, 10% Glycerol, 50 mM HEPES pH 7.5, 1 mM TCEP
Int^{82N}390C-sCI5	JCSG+, PEGs, Index, Wizard_I_II, The BCS Screen	7 °C, 20 °C	250 mM NaCl, 10% Glycerol, 50 mM HEPES pH 7.5, 1 mM TCEP
Int^{82N}390C-sCI5	XP Screen	7 °C, 20 °C	50 mM NaCl, 5% Glycerol, 25 mM HEPES pH 7.5, 5 mM $MgCl_2$, 1 mM TCEP
Int^{82N}390C-palCI6-IR_L-4	JCSG+, PEGs, The LMB Screen, MORPHEUS	20 °C	125 mM NaCl, 5% Glycerol, 25 mM HEPES pH 7.5, 1 mM DTT
Int^{82N}390C-palCI6-IR_R-3	JCSG+, PEGs, The LMB Screen, MORPHEUS	20 °C	125 mM NaCl, 5% Glycerol, 25 mM HEPES pH 7.5, 1 mM DTT
Int^{82N}390C-sCI5_TA	JCSG+, PEGs	20 °C	125 mM NaCl, 5% Glycerol, 25 mM HEPES pH 7.5, 0.5 mM TCEP
Int^{82N}390C-sCI5_TA	JCSG+, PEGs, XP Screen, MORPHEUS, MORPHEUS II, The BCS Screen	20 °C	125 mM NaCl, 25 mM HEPES pH 7.5, 0.5 mM TCEP
Int^{82N}390C-sCI5_TA (hit?)	JCSG+, PEGs, XP Screen, MORPHEUS, MORPHEUS II, The BCS Screen	7 °C	125 mM NaCl, 25 mM HEPES pH 7.5, 0.5 mM TCEP

Table 6-4: List of various Int-HJ complexes used for initial crystallization trials in order to obtain crystals of the Int-HJ tetrameric complex.

Complex	Screens	T	Crystallization Buffer
Int^{82N}R225K-HJ1	JCSG+, PEGs, Wizard_I_II, Nucleix, Natrix, MORPHEUS, MPD, MIDAS	20 °C	250 mM NaCl, 10% Glycerol, 50 mM HEPES pH 7.5, 10 mM $MgCl_2$, 5 mM DTT
Int^{82N}R225K-HJ1	JCSG+, PEGs, Wizard_I_II, Nucleix, The BCS Screen, Index, MORPHEUS, XP Screen, The LMB screen, Additive screen*, The Angstrom Additive Screen*, user screens*	7 °C	250 mM NaCl, 10% Glycerol, 50 mM HEPES pH 7.5, 10 mM $MgCl_2$, 5 mM DTT
Int^{82N}R225K-HJ1	JCSG+, PEGs, Nucleix	7 °C	250 mM NaCl, 5% Glycerol, 25 mM Sodium acetate pH 5.5, 0.5 mM TCEP
Int^{82N}R225K-HJ1	JCSG+, PEGs	7 °C	200 mM NaCl, 10% Glycerol, 100 mM BisTris pH 6.0, 0.5 mM TCEP

Int^{82N}R225K-HJ1T	JCSG+, PEGs	7 °C	250 mM NaCl, 10% Glycerol, 50 mM HEPES pH 7.5, 10 mM $MgCl_2$, 1 mM TCEP
Int^{82N}R225K-HJ1Ts	JCSG+, PEGs	7 °C	250 mM NaCl, 10% Glycerol, 50 mM HEPES pH 7.5, 10 mM $MgCl_2$, 1 mM TCEP
Int^{82N}R225K-HJ1T	JCSG+, PEGs	20 °C	250 mM NaCl, 5% Glycerol, 50 mM HEPES pH 7.5, 10 mM $MgCl_2$, 1 mM TCEP
Int^{82N}R225K-HJ1Ts	JCSG+, PEGs	20 °C	250 mM NaCl, 5% Glycerol, 50 mM HEPES pH 7.5, 10 mM $MgCl_2$, 1 mM TCEP
Int^{82N}390C-R225K-HJ1	JCSG+, PEGs, The BCS Screen	7 °C, 20 °C	250 mM NaCl, 10% Glycerol, 50 mM HEPES pH 7.5, 10 mM $MgCl_2$, 5 mM DTT
Int^{82N}390C-2YF-HJ1	JCSG+, PEGs, The BCS Screen	7 °C, 20 °C	250 mM NaCl, 10% Glycerol, 50 mM HEPES pH 7.5, 10 mM $MgCl_2$, 5 mM DTT
Int^{82N}R225K-HJ1GA	JCSG+, Index, The BCS Screen	7 °C	250 mM NaCl, 10% Glycerol, 50 mM HEPES pH 7.5, 10 mM $MgCl_2$, 5 mM DTT
Int^{82N}R225K-HJ1T41 (hit?)	JCSG+, Index, The BCS Screen, XP Screen	7 °C	250 mM NaCl, 10% Glycerol, 50 mM HEPES pH 7.5, 10 mM $MgCl_2$, 5 mM DTT
Int^{82N}R225K-HJ1	JCSG+, MORPHEUS, The BCS Screen	7 °C	200 mM $MgCl_2$, 10% Glycerol, 50 mM HEPES pH 7.5, 5 mM DTT
Int^{82N}R225K-HJ6-44	JCSG+, PEGs, The BCS Screen, XP Screen, MORPHEUS	7 °C	250 mM NaCl, 10% Glycerol, 50 mM HEPES pH 7.5, 10 mM $MgCl_2$, 5 mM DTT
Int^{82N}R225K-HJ6-45	JCSG+, PEGs, The BCS Screen, XP Screen, MORPHEUS	7 °C	250 mM NaCl, 10% Glycerol, 50 mM HEPES pH 7.5, 10 mM $MgCl_2$, 5 mM DTT
Int^{82N}R225K-HJ1 (hit?)	JCSG+, User screen + Additive Screen (F5-JCSG+), PEGs, The LMB screen, The BCS Screen, XP Screen, MORPHEUS, Natrix, Nucleix, MIDAS	7 °C	200 mM NaCl, 5% Glycerol, 20 mM HEPES pH 7.5, 5 mM $MgCl_2$, 0.5 mM TCEP
Int^{82N}R225K-HJ1	JCSG+, The LMB screen, The BCS Screen, XP Screen, Natrix, Nucleix, MIDAS	20 °C	200 mM NaCl, 5% Glycerol, 20 mM HEPES pH 7.5, 5 mM $MgCl_2$, 0.5 mM TCEP
Int^{82N}R225K-HJ8	JCSG+	7 °C, 20 °C	200 mM NaCl, 5% Glycerol, 20 mM HEPES pH 7.5, 5 mM $MgCl_2$, 0.5 mM TCEP
Int^{82N}R225K-HJ1 (no glycerol)	JCSG+, PEGs, MORPHEUS, The LMB screen, The BCS Screen, XP Screen, Natrix, Nucleix, MIDAS	7 °C	200 mM NaCl, 20 mM HEPES pH 7.5, 5 mM $MgCl_2$, 0.5 mM TCEP
Int^{82N}R225K-HJ1 (no glycerol)	The LMB screen, The BCS Screen, Natrix, Nucleix, MIDAS	20 °C	200 mM NaCl, 20 mM HEPES pH 7.5, 5 mM $MgCl_2$, 0.5 mM TCEP

Int^{82N}R225K-HJ1T	The LMB screen, MIDAS, MORPHEUS, The BCS Screen	7 °C	200 mM NaCl, 20 mM HEPES pH 7.5, 5 mM $MgCl_2$, 0.5 mM TCEP
Int^{82N}R225K-HJ1T	The LMB screen, MIDAS, The BCS Screen	20 °C	200 mM NaCl, 20 mM HEPES pH 7.5, 5 mM $MgCl_2$, 0.5 mM TCEP
Int^{82N}R225K-HJ1	The LMB screen, The BCS Screen, MIDAS	7 °C	200 mM NaCl, 25 mM Sodium acetate pH 5.5, 5 mM $MgCl_2$, 0.5 mM TCEP
Int^{82N}R225K-HJ1	The LMB screen, MORPHEUS, MIDAS	7 °C, 20 °C	200 mM NaCl, 5% Glycerol, 25 mM Sodium acetate pH 5.5, 5 mM $MgCl_2$, 0.5 mM TCEP
Int^{82N}R225K-HJ1TA (no glycerol) (hit)	JCSG+, PEGs, The LMB screen, MIDAS, MORPHEUS, MORPHEUS II, MORPHEUS III, XP screen, The BCS Screen, Additive Screen (PEGs-A1), Silver-Bullets, User Screen (PEGs-A1)	7 °C	200 mM NaCl, 20 mM HEPES pH 7.5, 5 mM $MgCl_2$, 0.5 mM TCEP
Int^{82N}R225K-HJ1TA (no glycerol)	JCSG+, PEGs, The LMB screen, MIDAS, MORPHEUS, MORPHEUS II, MORPHEUS III,	20 °C	200 mM NaCl, 20 mM HEPES pH 7.5, 5 mM $MgCl_2$, 0.5 mM TCEP
Int^{82N}R225K-HJ1TA (no glycerol)	JCSG+, Index, PEGs, PEGs II, The LMB screen, MIDAS, XP screen, The BCS Screen	7 °C	200 mM NaCl, 25 mM Sodium acetate pH 5.5, 5 mM $MgCl_2$, 0.5 mM TCEP
Int^{82N}R225K-HJ1TA (no glycerol)	PEGs, PEGs II, The LMB screen, MIDAS, XP screen, The BCS Screen	20 °C	200 mM NaCl, 25 mM Sodium acetate pH 5.5, 5 mM $MgCl_2$, 0.5 mM TCEP
Int^{82N}R225K-HJ1TA-CR5-43	PEGs	7 °C	200 mM NaCl, 20 mM HEPES pH 7.5, 5 mM $MgCl_2$, 0.5 mM TCEP
Int^{82N}R225K-HJ1TA-CR5-44	PEGs	7 °C	200 mM NaCl, 20 mM HEPES pH 7.5, 5 mM $MgCl_2$, 0.5 mM TCEP
^{SeMet}Int^{82N}R225K-HJ1TA	User Screen (PEGs A1 based)	7 °C	200 mM NaCl, 20 mM HEPES pH 7.5, 5 mM $MgCl_2$, 0.5 mM TCEP
Int^{82N}R225K-HJ1TA-Sy3	JCSG+, PEGs, PEGs II, MORPHEUS, The LMB Screen, Index, The BCS Screen, User Screen (PEGs A1 based), Additive Screen (PEGs A1 based)	7 °C	200 mM NaCl, 20 mM HEPES pH 7.5, 5 mM $MgCl_2$, 0.5 mM TCEP
Int^{82N}R225K-HJ1TA-Sy3	JCSG+, PEGs, The BCS Screen	7 °C	200 mM NaCl, 25 mM Sodium acetate pH 5.5, 5 mM $MgCl_2$, 0.5 mM TCEP
Int^{82N}R225K-HJ1TA-Sy3-1.2	User Screen (PEGs A1 based)	7 °C	200 mM NaCl, 20 mM HEPES pH 7.5, 5 mM $MgCl_2$, 0.5 mM TCEP