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., Buse Isbilir
born in Ankara, Turkey
Oral examination: 15th September 2021

Structural Characterization of Higher Order Complex Formation in Conjugative Transposition of an Antibiotic Resistance Carrying Mobile Genetic Element

Referees: Prof. Dr. Robert Russell

Dr. Martin Beck

SUMMARY

The spread of antibiotic resistance genes (ARGs) has become one of the biggest health care challenges in the last decades. Conjugative transposons (CTns) are a major class of mobile genetic elements that can transfer antibiotic resistance genes between bacterial genomes. For CTn transposition, DNA cleavage and joining reactions are catalyzed by a transposon-encoded integrase enzyme (Int). However, transposition doesn't involve only Int but requires rather complex machinery, which employs multiple CTn- and host-encoded factors assembled in distinct higher order protein-DNA complexes. These assemblies act as hubs to regulate recombination both spatially and temporally. Despite available biochemical and structural data about some CTn Int proteins, the coordination and regulation of the recombination reaction by higher order complex formation are not well understood. During my doctoral studies, I aimed to decipher the structural and functional principles of higher order nucleoprotein assemblies in the transposition of the *GISul2* element, a wide host range CTn that propagates diverse antibiotic resistance genes in pathogenic Gram-negative bacteria.

In the context of this work, several structures of transposon excision complexes formed at different steps of the reaction were determined using Cryo-EM, which revealed distinct molecular assemblies. First, the structure of the right transposon end (RE) complex showed that RE DNA is bent by integration host factor (IHF) to ~160°, which allows Int_{GISul2} to bridge its arm and core DNA sites. The formation of this complex strictly requires the presence of the host-encoded accessory protein IHF, indicating strong coordination between the host cell state and CTn mobilization. In the second part, I present the structural characterization of the left transposon end (LE) complex. The cryo-EM map revealed that seven excisionase (Xis) molecules bend the LE DNA to form a loop that is tethered by integrases. LE complex formation depends on DNA bending by the transposon-encoded Xis protein, a directionality factor of recombination, which is required for excision but inhibits integration. In the third part of the thesis, I describe structural insights into how right and left transposon ends come together to form a synaptic complex during transposon excision. This synaptic complex was captured using "suicide" DNA substrates, which can stall the recombination reaction after cleavage of the first DNA strand pair. Unexpectedly, this approach resulted in the formation of two different synaptic complexes, one including RE and LE, and the other one containing two RE molecules, which together elucidate the assembly process and catalytic activation of the native synaptic excision complex.

Together our results shed light on the principles of how the CTn integrase cooperates with accessory DNA bending proteins to coordinate active nucleoprotein complex assembly during excision and integration. These insights offer new knowledge about the mechanisms of antibiotic resistance spreading in Gram-negative pathogens. I hope that the better understanding of the CTn movement will help develop new strategies for tackling antibiotic resistance spread in the future.

ZUSAMMENFASSUNG

Die Ausbreitung von Antibiotikaresistenzgenen (ARGs) ist in den letzten Jahrzehnten zu einer der größten Herausforderungen im Gesundheitswesen geworden. Conjugative Transposons (CTns) sind eine wichtige Klasse von mobilen genetischen Elementen, die Antibiotikaresistenzgene zwischen bakteriellen Genomen übertragen können. Bei der CTn-Transposition werden die DNA-Spaltungs- und Verbindungsreaktionen durch ein vom Transposon kodiertes Integrase-Enzym (Int) katalysiert. An der Transposition ist jedoch nicht nur Int beteiligt, sondern es ist eine ziemlich komplexe Maschinerie erforderlich, die mehrere CTn- und wirtskodierte Faktoren einsetzt, die verschiedene Protein-DNA-Komplexen höherer Ordnung ausweist. Diese Komplexe fungieren als Knotenpunkte, um die Rekombination sowohl räumlich als auch zeitlich zu regulieren. Trotz verfügbarer biochemischer und struktureller Daten über einige CTn-Int-Proteine, ist die Koordination und Regulation der Rekombinationsreaktion durch die Bildung von Komplexen höherer Ordnung nicht gut verstanden. Während meines Promotionsstudiums war es mein Ziel, die strukturellen und funktionellen Prinzipien von Nukleoprotein-Assemblierungen höherer Ordnung bei der Transposition von *GIsul2* zu entschlüsseln, einem weit verbreiteten CTn, das diverse Antibiotikaresistenzgene in pathogenen gramnegativen Bakterien propagiert.

Im Rahmen dieser Arbeit wurden mehrere Strukturen von Transposon-Exzisionskomplexen, die in verschiedenen Schritten der Reaktion gebildet werden, mittels Cryo-EM bestimmt, die unterschiedliche molekulare Assemblierungen aufzeigten. Zunächst zeigte die Struktur des Komplexes des rechten Transposon-Endes (RE), dass die RE-DNA durch den Integration Host Factor (IHF) auf ~160° gebogen wird, was es Int_{GISul2} ermöglicht, seine Arm- und Core-DNA-Stellen zu überbrücken. Die Bildung dieses Komplexes erfordert zwingend die Anwesenheit des vom Wirt kodierten akzessorischen Proteins IHF, was auf eine starke Koordination zwischen dem Zustand der Wirtszelle und der Mobilisierung von CTn hinweist. Im zweiten Teil stelle ich die strukturelle Charakterisierung des Komplexes des linken Transposonendes (LE) vor. Die Cryo-EM-Karte zeigt, dass sieben Excisionase (Xis)-Moleküle die LE-DNA zu einer Schleife biegen, die von Integrasen angebunden wird. Die Bildung des LE-Komplexes hängt von der DNA-Biegung durch das transposon-kodierte Xis-Protein ab, einem Direktionalitätsfaktor der Rekombination, der für die Exzision erforderlich ist, aber die Integration hemmt. Im dritten Teil der Arbeit beschreibe ich strukturelle Erkenntnisse darüber, wie rechte und linke Transposonenden während der Transposonexzision zusammenkommen und einen synaptischen Komplex bilden. Dieser synaptische Komplex wurde mit Hilfe von "Suicide" DNA substrates eingefangen, die die Rekombinationsreaktion nach Abspaltung des ersten DNA-Strangpaares unterdrücken können. Unerwartet führte dieser Ansatz zur Bildung von zwei verschiedenen synaptischen Komplexen, von denen einer RE und LE enthält und der andere

zwei RE-Moleküle, die zusammen den Assemblierungsprozess und die katalytische Aktivierung des nativen synaptischen Exzisionskomplexes aufklären.

Zusammen werfen unsere Ergebnisse ein Licht auf die Prinzipien, wie die CTn-Integrase mit akzessorischen DNA-Biegeproteinen kooperiert, um die aktive Nukleoprotein-Komplex-Assemblierung während der Exzision und Integration zu koordinieren. Diese Erkenntnisse bieten neues Wissen über die Mechanismen der Ausbreitung von Antibiotikaresistenzen in gramnegativen Pathogenen. Ich hoffe, dass das bessere Verständnis der CTn-Bewegung dazu beitragen wird, in Zukunft neue Strategien zur Bekämpfung der Ausbreitung von Antibiotikaresistenzen zu entwickeln.

ACKNOWLEDGEMENTS

Foremost, I wish to thank Prof. Dr. Orsolya Barabas for accepting me to her lab and giving me adventurous projects. It was great to share both good and bad moments together, to share excitement and happiness or to get support from you. In addition to excellent scientific advice, you were always there for me with every aspect of my life. Thank you very much for providing me great opportunities and broadening my horizon, I always feel excited to learn new concepts from you and share your enthusiasm towards science.

Besides my advisor, I would like to thank all the members of my Thesis Advisory Committee, Dr. Martin Beck, Prof. Dr. Rob Russel, and Dr. Marco Marcia for all the support and recommendations during my PhD. Many thanks to Prof. Dr. Michael Knop for joining my Thesis Defense Committee. I also want to thank the EMBL PhD program for funding my studies and EMBL Graduate Office for their constant support during my PhD. I should also thank the EMBL Core Facilities for their service and endless support to our research for all these years. I would like to thank Brice for supporting my crystallization experiments and Martin for being there for any EM related problem. I want to show my deep gratitude to EMBL Cryo-EM Service Platform, Wim and Felix, who are great teachers and introduced me to the electron microscopy world. I also want to thank Syhamal, Yashar, Inessa and Herman for sharing their knowledge and helping me whenever I need when I was starting my cryo-EM journey.

My special thanks go to all past and present members of the Barabas Lab both EMBL and Geneva versions, who have touched my life! I have my special and sincere thanks to Gera, for reviving my enthusiasm for science and being a great discussion buddy! I have learnt a lot from you during our collaboration. I find your approach towards science is very well-balanced and this will be a great example for me during my scientific carrier. I hope we will share the excitement of many more achievements! Many thanks to my first supervisor Natalia! Thank you very much for all your support during my PhD and showing me how to be persistent and a strong in any condition. Thank you very much for always being there and for your invaluable friendship! Ceci, thank you very much for being a caring lab manager. It was always really nice to chat with you about both scientific and non-scientific topics over the years as bench mates. Vladimir, thanks for being there for any kind of passionate scientific discussion! Thanks to your great jokes, now I know that physicists have the funniest jokes ever! Mate, thanks for bringing fresh energy and fun to our lab. Especially in the stressed times, it is always fun to laugh at your jokes. Thanks to Freddy for being my first student and handling my supervision! Thanks Irma and Anna for being in the lab and showing me how to do science in a fun way, despite all the difficulties that it brings. My special thanks go to my dear friend Carlos. I am really thankful that we were together in this rollercoaster (even though you don't like rollercoasters :). Your company was always comforting, caring and fun. I am grateful that we spent this adventure together with you. I learnt a lot from you about how to be patient and nice whatever comes in life.

Last but not least, I would like to thank my family. Sizlerin desteği sayesinde bu tezi yazabiliyorum. Bana verdiğiniz duygusal destek herşeye bedel. Canım annem, her türlü zorluğun cesurca aşılabileceğini gösterdiği için, canım babam, hayat ne gösterirse göstersin yine de gülebilmek gerektiğini gösterdiği için vee tatlı, minik kardeşim bana sınırsız sevgi verdiğin ve desteğini hep hissettirdiği için. Dedelerim, babaannem ve anneannem ailemizin koca çınarları olarak hep yanımızda olduğunuz için. Canım dostum Çisotum, mesafeler ne kadar uzak olursa olsun, sımsıcak elini hep omzumda hissettirdiğin için çok minnettarım. Sizlerin desteği sayesinde bu macerayı da burada noktalayabiliyorum. Beraber daha nice maceralara…

Ve en derin teşekkürlerim sevgili eşim Özü Özdeş Koç'a, bütün doktoram boyunca ve özellikle tez yazma dönemimde bana sağladığın destek, sabır ve şefkat için çok teşekkür ederim. Başımı yasladığım omzun olmasa bu günleri görebileceğimi hiç sanmam.

CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

2D	two dimensional
3D	three dimensional
A	Adenine
aa	amino acid
AB	arm binding domain
Arg, R	Arginine
ARG	antibiotic resistance gene
ATP	Adenosine triphosphate
att	attachment site
attB	bacterial attachment site
attL	left attachment site
attP	phage attachment site
attR	right attachment site
bp	Base pair
C	Cytosine
CAT	catalytic domain
CB	core binding domain
CDC	Center for Disease Control
CI	circular intermediate
cryo-EM	Cryogenic-Electron Microscopy
CTF	contrast transfer function
CTn	conjugative transposon
CV	column volume
Da	Dalton
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
Glu, E	Glutamic acid
EDTA	Ethylenediaminetetraacetic acid
EMBL	European Molecular Biology Laboratory
EMSA	Electrophoretic mobility shift assay
FAM	Fluorescein amidite
Fis	Factor for Inversion Stimulation
FSC	Fourier Shell Correlation
G	Guanine
His, H	Histidine
HJ	Holliday Junction
HT	horizontal gene transfer
ICE	Integrative and Conjugative Element
IHF	Integration Host Factor
Int	Integrase
Int_{GISul2}	*GISul2* integrase
Int_{λ}	λ integrase
IPTG	Isopropyl β-D-1-thiogalactopyranoside

IS	Insertion sequences
Lys, K	Lysine
Kan	Kanamycin
kb	kilo basepair
kDa	kilodalton
keV	kiloelectron volt
LB	Luria-Bertani Broth
LE	left transposon end
MGE	mobile genetic element
min	minute
mRNA	messenger ribonucleic acid
nm	nanometer
oriT	origin of transfer
P	Proline
Phe, F	Phenylalanine
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDB	Protein Data Bank
Gln, Q	Glutamine
RE	right transposon end
RNAse	ribonuclease
rpm	round per minute
rRNA	ribosomal ribonucleic acid
SEC	size exclusion chromatography
Ser, S	Serine
T	Thymine
TAE	Tris-acetate-EDTA
TBE	Tris-borate-EDTA
Tm	melting temperature
Tn	transposon
TR	Tyrosine recombinase
Tyr, Y	Tyrosine
V	Valine
WHO	World Health Organization
Xis	Exicionase
YJ	Y- junction

1 INTRODUCTION

1.1 The Antibiotic Resistance Crisis

Antibiotics are chemicals, which combat bacteria either by killing them or slowing down their growth and spread. Antibiotics have existed for billion years and they have provided selective benefits to various organisms, which produce them to inhibit or eliminate bacteria in their environment that compete for the same resources. The use of antibiotics by humanity have started in early civilizations by treating infections with plant extracts and molds. It has stepped into a golden era, with the discovery of penicillin by Alexander Fleming in 1928 (Fleming, 1929) and streptomycin by Selman Waksman in 1943 (Waksman, 1952), which revolutionized medicine.

During the long existence of antibiotics, bacteria have developed defense mechanisms against antibiotics. Bacterial resistance is often encoded in antibiotic resistance genes (ARGs), which express key proteins in various pathways to overcome the harmful effect of antibiotics. The ARGs can be inherited to the offspring by vertical transfer. In addition, bacteria can also share resistance genes with each other by horizontal transfer, which constitutes the main source of ARG spread. While ARGs and their spread have existed for a long time, it has become one of the biggest health care challenges in the last few decades due to a tremendous increase in the rate of dissemination. It has been reported that globally there are 700 thousand people dying in a year, due to infections with antibiotic resistant bacteria. With the slow speed of development of new antibiotics and the growing resistance among bacteria even to last resort antibiotics, the World Health Organization (WHO) states that we are approaching a “post-antibiotic era”, where antibiotics won’t be effective and due to this reason, common infections can become deadly again (WHO, 2019). Indeed, according to the predictions of the Centers for Disease Control and Prevention (CDC), by the year 2050, antibiotic resistant bacteria will kill 10 million people annually worldwide (O’Neill, 2014), exceeding cancer-related fatalities.

The main reason behind the increasing rate of ARG spread over the last decades is the frequent misuse of antibiotics in medicine and agriculture. With the overuse of antibiotics, selective pressure has increased the abundance of resistance genes in bacterial populations. The impact of antibiotic overuse is further enhanced due to the fact that the majority of antibiotics are excreted unchanged and introduced directly into the environment. Indeed, it has been shown with soil samples taken in the Netherlands that the detected ARGs of major antibiotic classes significantly increased between 1940 to

2008, for example, tetracycline resistance genes being 15 times more detected in 2008 than in 1970s (Knapp et al., 2010).

To tackle the antibiotic resistance crisis, we must understand how bacteria acquire and transfer resistance genes. Once the mechanisms and tools used by bacteria to disseminate the resistance genes are well comprehended, new actions can be taken to limit their rapid spread. To elucidate the mechanism of antibiotic resistance spread, during my doctoral studies I have focused on understanding the molecular mechanisms of conjugative transposons, a group of mobile genetic elements that provide prime vehicles for ARG spreading. For this aim, I have worked with the *GISul2* conjugative transposon, which is a widely distributed, sulfonamide resistance gene-carrying element. I aimed to gain new insights into conjugative transposon movement by investigating the integration/excision reactions of this model element, with the ultimate aim to develop new tools to slow down the antibiotic resistance spread.

1.1.1 Mechanisms of Antibiotic Action and Bacterial Resistance

Antibiotics target vital pathways in bacteria, including cell wall synthesis, protein expression and nucleic acid biogenesis.

Antibiotics that target cell-wall synthesis are used the most extensively, and due to the absence of cell wall in human cells, they have an intrinsic selectivity against bacteria. β-lactam compounds are some of the oldest antibiotics, which target cell wall biogenesis (Tipper et al., 1985). They inhibit the formation of the peptidoglycan layer, the main component of the bacterial cell wall, eventually leading to the lysis of bacterium. Glycopeptides are another type of antibiotic, which also inhibit cell wall synthesis by binding to the precursors of peptidoglycan subunits (Binda et al., 2014).

Inhibition of protein synthesis is another mechanism used by antibiotics, whose main target is bacterial ribosomes. Aminoglycosides are positively charged molecules, which interact with the 30S subunit of the ribosome and cause misreading or premature termination of translation of mRNA (Krause et al., 2016). Tetracyclines also bind to the 30S ribosomal subunit to prevent tRNA binding to the A site (Chopra and Roberts, 2001). In turn, chloramphenicol, macrolides and oxazolidinones inhibit the activity of the 50S ribosomal subunit by binding to the 23S rRNA (Hermann, 2005).

Another pathway that antibiotics target is bacterial DNA replication. Quinolones bind to DNA gyrase and topoisomerase IV to inhibit their DNA cleavage and joining activity (Pham et al., 2019). Additionally, targeting nucleotide metabolism to interfere with nucleic acid

synthesis and other cellular functions is another approach used by antibiotics. Folic acid pathway inhibitors, such as sulfonamides and trimethoprim, provide examples of this type of antibiotics (Fanning et al., 2014).

While antibiotics inhibit essential pathways to slow bacterial growth, there are many resistance mechanisms that bacteria have developed to escape the effect of these agents. The first approach is to prevent the accumulation of antibiotics in the cell by either decreasing their uptake or increasing their efflux. Most of the drug molecules are transferred by porins through diffusion to the cytoplasm. To decrease the uptake of the negatively charged antibiotics, e.g. β-lactams, which can cross the membrane through porins, bacteria reduce the number of porin channels available for their entry to the cell (Fernandez and Hancock, 2012). Another mechanism used to decrease the antibiotic concentration in the cytoplasm involves pumping the antibiotics out of the cell. Efflux pumps found in the cytoplasmic membrane pump out antibiotics before they reach a critical concentration in the cell (Peterson and Kaur, 2018). While these pumps can be specific to a particular antibiotic, there are multi-drug pumps that can transport a variety of antibiotics (Guilfoile and Hutchinson, 1991; Li et al., 2014). To make antibiotics ineffective, another method that bacteria employ is modifying the binding sites of the antibiotic on its target protein or complex. Since the antibiotic interactions with their target sites are usually specific, even minor alterations in the site can disrupt the drug-protein interaction (Peterson and Kaur, 2018). In some cases, instead of modifying the target site, bacteria modify the chemical structure of the antibiotics. With these modifications, the activity of antibiotics is impaired by prohibiting their target site binding. Various specific enzymes in the bacteria, like β-lactamases, aminoglycoside-modifying enzymes, and chloramphenicol acetyltransferases (AACs), perform such modifications (Peterson and Kaur, 2018).

1.2 Antibiotic Resistance Spread by Horizontal Gene Transfer

Once antibiotic resistance genes emerge, they can spread rapidly among bacterial communities due to efficient gene transfer mechanisms that bacteria possess. By vertical transfer, bacteria pass ARG to their offspring and create stable resistance strains. Additionally, they can also transfer genes to other bacterial cells, which are not related by parent-offspring relationship, by horizontal gene transfer (HT). In fact, HT is a main cause for rapid ARG spreading, since it allows bacteria to exchange ARGs among diverse species. While it takes some generations of bacteria to evolve new ARGs via mutagenesis, they can immediately reach to the ARG pool of their community through HT. There are three mechanisms for HT in bacteria: transduction, transformation and conjugation, which I will discuss in more details in the next sections.

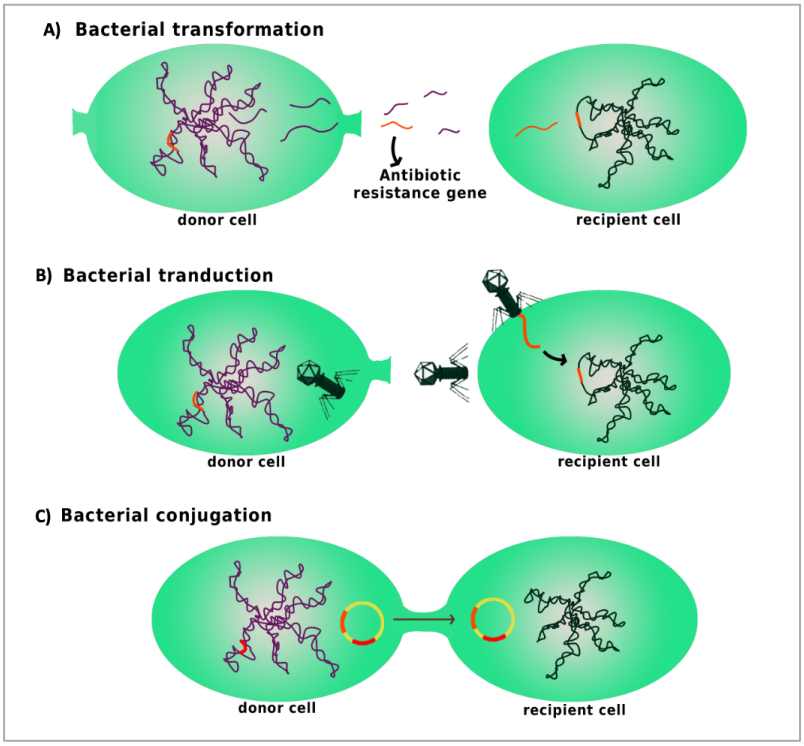


Figure 1-1: The mechanisms of horizontal gene transfer (HT). There are three canonical ways for HT: A) transformation, B) transduction, and C) conjugation. (Based on Geenen et al., 2010). The transfer of an ARG (in red) is depicted between donor and recipient bacterial cells.

1.2.1 Transformation

Transformation is the process where bacteria uptakes extracellular DNA from its environment (Figure 1-1 A). For transformation, bacteria need to be "competent", which is the state where competence related genes are expressed to enable DNA uptake (Johnston et al., 2014). The acquired extracellular DNA can be stabilized in the bacteria either by integrating into the genome or into a self-replicating plasmid, a process called genetic transformation (Palchevskiy and Finkel, 2006). Otherwise, it can also be used as a nutrient source for bacteria in an environment with limited resources.

In bacteria, the competence for transformation can be induced naturally or artificially. It is generally a transient state, which is induced by alterations in the growth state of the bacteria and established by the expression of a group of "competence specific genes". These genes encode for proteins which take part in the transport of DNA across the bacterial membranes (Claverys and Martin, 2003). There are different protein complexes involved in the uptake of both single and double stranded DNA either in sequence dependent or independent manner (Blokesch, 2016). During transformation, a DNA fragment binds to receptors found on the bacterial cell surface and passes through the

inner cell membrane in single or double stranded form via a DNA translocator (Sun, 2018). The DNA uptake machinery is formed by different set of proteins in Gram-negative and Gram-positive bacteria, due to the difference of their membrane organization (Claverys and Martin, 2003).

Exogenous DNA comes from lysed microbes or from active secretion of bacteria and includes plasmid DNA circles and fragmented linear DNA. This DNA may contain diverse ARGs, conferring resistance to various antibiotics. Since natural transformation does not discriminate about the source of the incoming DNA, it enables resistance transfer among wide range of bacteria. Although the real impact of transformation on resistance spread is not well known, several clinically relevant antibiotic-resistant pathogens are capable of DNA uptake and natural transformation, including *Acinetobacter, Haemophilus, Neisseria, Pseudomonas, Staphylococcus,* and *Streptococcus* (Johnston et al., 2014; Traglia et al., 2014).

1.2.2 Transduction

Transduction is the process where phages carry bacterial genes from one bacterium to another (Figure 1-1 B). Diverse DNA segments can be transferred this way, including genes that are advantageous to recipient and promote its survival. During infection, phages bind to the surface of a bacterium and inject their DNA into the host cell cytoplasm. Some phages can integrate their genome into host genome, others replicate themselves without integrating (Howard-Varona, 2017). After the phage components are produced in the host cell, the newly replicated genome is packed into a capsid. During this packing, in addition to the phage genome, bacterial genes can also be packed mistakenly, which is the basis of transduction.

There are two kinds of transduction, which are generalized and specialized transduction. Generalized transduction allows the transfer of any bacterial gene, while specialized transduction can transfer only specific bacterial genes. Specialized transduction can only occur in phages, which integrate their genome in the host genomes. It happens due to imprecise excision of the phage genome, which results in the inclusion of flanking bacterial DNA sequences in the new capsid (Chiang et al., 2019). The transferred bacterial DNA segments are situated close to the phage insertion site and are transferred fused to the phage genome. In turn, generalized transduction involves the packaging of any bacterial DNA in a separate fragment. Following their production, both generalized and specialized transducing particles rely on normal phage infection processes to deliver the mistakenly packaged host DNA to the next bacterium. In order for a bacterium to maintain the new

genetic information, the DNA must be incorporated into the host chromosome or be maintained as a self-replicating element like a plasmid (Chiang et al., 2019).

Recent findings suggest that phages may play a more significant role in the emergence and spread of ARGs than previously expected. An extensive study using metagenomic approaches revealed the presence of ARG-like sequences in the virome of activated sludge, which conferred resistance to several antibiotics, including tetracycline, ampicillin, and bleomycin (Parsley et al., 2010). However, most phages are species-specific, and, therefore, they do not seem to majorly contribute to the transfer of ARGs between distant species, where other mechanisms play more important roles (Koskella and Meaden, 2013).

1.2.3 Conjugation

Conjugation is the process where DNA is actively transferred from one bacterium to another via a direct physical connection (Figure 1-1 C). This connection can be a pilus or adhesin which provides a safe passage for DNA between bacteria (Llosa et al., 2002). Conjugation is typically induced by conjugative plasmids and conjugative transposons, which encode the required genes for forming the conjugation channels. In Gram-positive bacteria, the contact between donor and recipient is formed via surface adhesins, whereas in Gram-negative strains the interaction relies on conjugative pili (Llosa et al., 2002).

Conjugation is mainly mediated by Type IV secretion system. The process starts with a relaxase enzyme nicking one DNA strand at a specific sequence, called the origin of transfer (oriT), which becomes covalently attached to DNA. After unwinding the double strand, the nicked strand is replicated. Once the full strand is replicated, the ends are rejoined and form a circular single-stranded DNA. The coupling proteins recognize and direct the peeled-off relaxase-DNA complex to membrane where it gets transferred to a recipient bacterium via a pilus or surface adhesin (Cabezon et al., 2015).

Conjugation is a common mechanism of HT in bacteria and it is instrumental for the spread of ARG. In fact, it is considered to have a major role in HT of ARGs. While transduction is presumably a ‘side effect’ of the erroneous phage packaging, conjugation is a dedicated method for the genetic exchange. Furthermore, multiple types of conjugation machineries that can form physical contacts between distantly related bacteria have been reported, which enables efficient gene transfer across a broad range of bacterial hosts (Dahlberg et al, 1998).

1.3 Mobile Genetic Elements in ARG Spread

Mobile genetic elements (MGE) are DNA segments that can move from one place to another either within a genome or between genomes. Due to their mobility, MGEs contribute largely to the capture, accumulation and dissemination of ARG. Bacterial MGEs include insertion sequences, transposons, genomic islands, integrons, plasmids and integrative conjugative elements. Below I will summarize the characteristics of the major MGE groups that exist in bacterial genomes and are important for ARG spread.

Insertion Sequences (ISs) are comparatively small elements that typically encode only one protein, a transposase, that is required for mobilization of the element. ISs are flanked by terminal inverted repeats, which are DNA sequences that the transposase recognizes and binds to (Siguier et al., 2015). Although they carry only a transposon gene, they can still contribute to ARG spread, because they can form so called "composite transposons", which consist of two ISs that surround chromosomal genes. In this way, ISs can create new composite transposons entrapping and mobilizing ARGs into pathogens. IS6 family elements provide an example of these group of MGE, which have significant roles in various ARG spread in both Gram-negative (IS1216) and Gram-positive (IS257) bacteria (Partridge et al., 2018).

Unit transposons (or just transposons) are another class of MGEs, which can relocate themselves with the help of self-encoded transposases. They are larger than ISs and are also flanked by terminal inverted repeats. In addition to the transposase gene, they carry regulatory genes and "passenger" genes, such as ARGs. In particular, ARGs are often associated with Tn*3* family transposons (Nicolas et al., 2015). For example, the Tn*1546* transposon from this family carries vancomycin resistance genes (*vanA* gene cluster) and was found to cause Van-resistance in the multidrug-resistant MRSA, where vancomycin is a last-resort treatment option. Another transposon family, the Tn*7*-like transposons, includes members associated with antibiotic resistance, such as the penicillin carrying Tn*552* from *Staphylococcus aureus* (Partridge et al., 2018).

Plasmids are self-replicating extrachromosomal DNA circles, which vary in size from less than a kilobase to several megabases (Shintani et al., 2015). Their ability to self-replicate allows them to exist outside of the bacterial chromosome. They can be transferred vertically to the next generations. Some plasmids also have conjugation and mobilization function, which allows them to autonomously move horizontally. The replication and conjugation genes required for the lifecycle of the plasmids are altogether called as "backbone" (Thomas, 2000). Additionally, plasmids often carry ARGs and can mobilize other MGEs, such as ISs, transposons, and integrons, which in turn can also carry their own ARGs. In this way, plasmids can harbor a wide range of ARG cargos. There are many examples of emerging multi-resistance plasmids found in *Enterobacteriaceae* (A/C, F, HI, I2

plasmids), *P. aeruginosa* (IncP-2 plasmid), *Staphylococci* and *Enterococci* carrying diverse set of ARGs (Partridge et al., 2018).

Although plasmids are powerful vectors for carrying ARGs, they do not integrate in bacterial genomes, for that reason they can only confer temporary resistance to their hosts. On the other hand, conjugative transposons, another class of MGEs, can integrate into host genomes and autonomously move between diverse bacterial cells via conjugation.

1.3.1 Conjugative Transposons

Conjugative transposons (CTn; also called integrative conjugative elements, ICEs) are a major class of MGEs that are particularly efficient in spreading antibiotic resistance (Bi et al., 2012; van Hal et al., 2016). They often harbor various resistance genes, and they can autonomously move between diverse bacterial cells (Salyers et al., 1995). They move involving a phage-like excision/insertion process, which creates a circular DNA intermediate with plasmid-like conjugative properties.

Many CTns exhibit a relatively low level of insertion site specificity, whereas others have more stringent requirements and integrate predominantly into specific genomic loci. CTns can be found in both Gram-negative and Gram-positive bacteria, with their sizes ranging from 18 to more than 600 kilobase pair (kb). They are flanked by terminal inverted repeats, (RE and LE). Between these end sequences, they contain several genetic modules which encode for specific functions: (i) Transposition module contains a transposase gene (called integrase) and accessory factors (i.e. excisionase) required for DNA excision and/or integration; (ii) Conjugation module, contains genes required for mobilization and conjugation of the element to a recipient cell and (iii) Cargo module, which encodes for passenger functions such as ARG and virulence genes (Burrus et al., 2002).

CTns are generally inserted in the host genome and their movement is initiated under specific conditions. As a first step of their movement, the CTn DNA is excised, and it forms a covalently closed circular intermediate (CI) (Figure 1-2). The excision of the element is performed by the self-encoded transposase, called integrase (Int), with the help of the self-encoded (excisionase) and possibly some host encoded accessory factors (ie. IHF, HU, Fis, discussed in Section 1.5). Together Int and accessory factors are thought to form higher order nucleoprotein assemblies on both RE and LE, where the Integrase performs site-specific recombination reactions, breaking and rejoining the DNA ends of the CTn (Salyers et al., 1995). Most CTn integrases belong to the tyrosine recombinase family, with some examples also found among the large serine recombinase family.

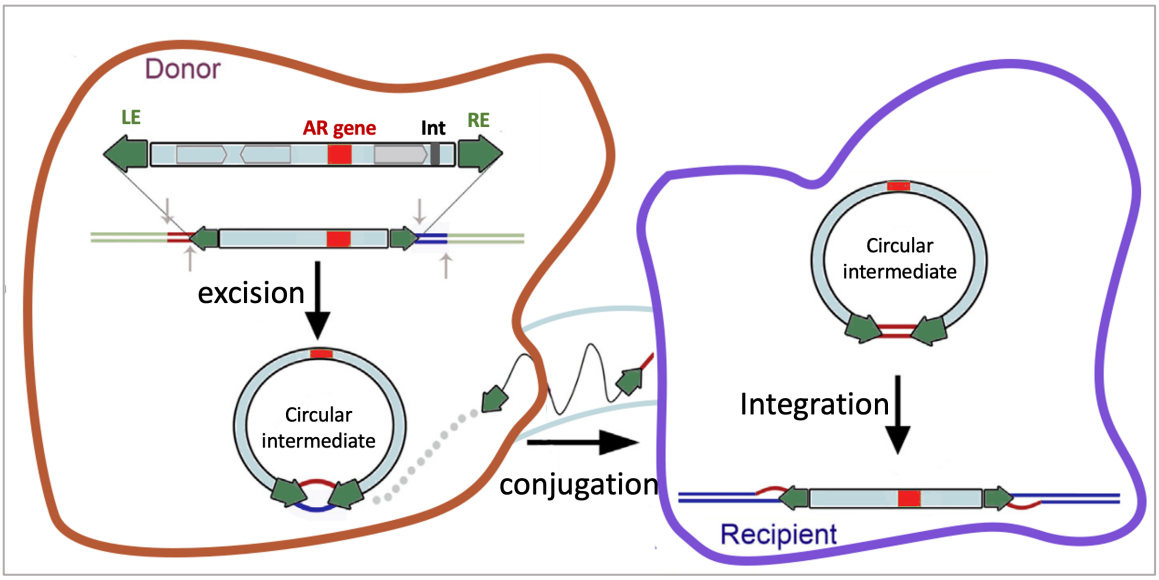


Figure 1-2: The movement of conjugative transposons (CTn). The CTn is depicted as a DNA segment containing various genes (grey arrows), including an antibiotic resistance gene (ARG, in red), flanked by attachment sites (RE and LE, green). The transposition of the CTn starts with excision of the element, which is catalyzed by the element encoded Int (in dark grey). The recombination of RE and LE produces a CI, where these two sequences are connected with a crossover region (5-7 bp, blue and dark red lines). The CI is transferred to a recipient cell via conjugation and integrates into the genome or a plasmid at an insertion site. Based on a figure from Rubio-Cosials et al., 2018.

The integrase excises the CTn in a CI form, which can either integrate elsewhere in the same cell (intracellular transposition) or transfer by conjugation to a recipient cell (Figure 1-2). For conjugation, the required genes are normally encoded in the conjugation module of the element. To start conjugative transfer, the CI is cleaved at the origin of transfer (oriT) sequence by the specific, CTn encoded relaxase enzyme, which stays attached to the cleaved DNA strand. Like classical conjugative plasmids, most CTns use rolling circle replication while they conjugate (Wright and Grossman, 2016). The relaxase bound DNA strand is then brought to the conjugation bridge (a pilus or adhesin) with the help of coupling proteins. A single strand of the CI DNA is then transferred into a recipient bacterial cell (Llosa et al., 2002). Next, the circular CTn finds a suitable integration site in the recipient cell's genome, where it integrates, again with the help of the CTn integrase protein. Due to powerful combination of their conjugation and transposition/integration features, CTns are effective tools for ARG spreading. In the following sections, I will describe four well-known CTn families, which were shown to spread ARGs.

1.3.1.1 Tn*916* Family

The term conjugative transposon was first used to describe Tn*916*, a MGE found in Gram-positive bacteria. The Tn*916* family unites CTns that have similar conjugation and transposition modules. In addition to Tn*916* itself, this family includes Tn*6000*, Tn*5397*, Tn*5801* and Tn*1549*. While the first four elements encode tetracycline resistance, Tn*1549* encodes for vancomycin resistance genes.

All members of the Tn*916* family encode an integrase belonging to the tyrosine site-specific recombinase family, which catalyzes the excision of the element by recombining the transposon ends (RE and LE) of the element to form a CI. The excised CI can then be integrated in various sites with an AT-rich sequence either in the same cell or after conjugation in a recipient cell (Poyart-Salmeron et al., 1989). Site-specific excision and low specificity integration allow these elements to efficiently spread among diverse genomes. Indeed, it has been shown that Tn*916* elements are found in a broad range of various bacterial phyla (Smyshlyaev et al., 2021).

1.3.1.2 SXT/R391 Family

CTn elements from the SXT/R391 family are found in diverse species of Gram-negative bacteria (Burrus et al., 2002). This CTn family includes more than 30 elements that have been detected in clinical and environmental isolates of several proteobacterial species around the world.

The prototype SXT element is 100 kb long and was originally discovered in *Vibrio cholerae*. SXT encodes resistance to several antibiotics, including sulfamethoxazole and trimethoprim (which the SXT abbreviation stands for) that had been previously used to treat cholera (Burrus and Waldor, 2003). Members of the SXT/R391 family encode a nearly identical integrase (Int), which is a tyrosine site-specific recombinase. It catalyzes the excision and site-specific integration of the elements at the end of *prfC*, a conserved chromosomal gene that encodes the peptide chain release factor 3 (Hochhut and Waldor, 1999). As in the Tn*916* family, these CTns also encode an excisionase protein, in addition to integrase, in their transposition module (Ryan et al., 2016).

1.4 DNA Recombination

DNA recombination is the process that all life forms employ to exchange DNA. It requires a specialized protein called recombinase, which interacts with two DNA sites and catalyzes their breakage and rejoining. DNA recombination can be classified into two groups: homologous and site-specific recombination. Homologous recombination requires homology between two DNA sites for recombination to take place. Organisms employ this type of recombination during meiosis to exchange DNA segments between homologous chromosomes. In addition, it is one of the main mechanisms used in DNA repair in all kingdoms of life. In turn, site-specific recombination requires a specific signal sequence to recombine two DNA segments. It is one of the main mechanisms used for integration and excision of the MGEs and phages. In the following sections, I will summarize the mechanistic principles of site-specific recombination.

1.4.1 Site-Specific Recombination

Site-specific recombination includes various reactions, which lead to exchange of DNA segments between two specific sites. It requires two DNA partners and a site-specific recombinase that catalyzes the breaking and rejoining reactions of DNA without the need of high-energy cofactors such as ATP (Grindley et al., 2006).

Depending on the orientation and localization of the recombination sites, site-specific recombination can have three outcomes: If the recombination sites are located in different DNA molecules, process results in integration of a segment. Excision happens if the recombination sites are located on the same molecule and oriented in the same direction; if they are inversely oriented, then the DNA segment is inverted as a result of recombination.

Site-specific recombinases are classified into two big families based on their mechanistic principles: the tyrosine (Tyr, Y) recombinase family and the serine (Ser, S) recombinase family. These families are named after the conserved catalytic residue, which starts the reaction by attacking the sugar-phosphate backbone of the DNA. Although, both recombinase families catalyze the formation of similar recombination products, they are not related in amino acid sequence and employ distinct molecular mechanisms to recombine DNA segments. In serine recombination, four recombinase monomers break four DNA strands, prior to strand exchange and ligation. The DNA is cleaved with two-nucleotide stagger by creating 3’ overhangs (summarized in Grindley et al., 2006). Once all strands are cleaved, two protein subunits rotate relative to the other two, and the strands are exchanged and religated. On the other hand, tyrosine recombinases break and rejoin

single strands in pairs to form a HJ intermediate, as described in more detail in the next section.

1.4.2 Tyrosine Recombination

Tyrosine recombinases (TR) are a large family of recombinases, which catalyze the recombination of DNA segments in various biological processes. They are widespread in prokaryotes but are also found in archaea and eukaryotes (Poulter and Butler, 2021).

TRs usually comprise three functional domains: The C-terminal catalytic domain (CAT) catalyzes DNA cleavage and joining reactions during recombination. While the overall fold of the CAT is well preserved, there is limited sequence conservation across the TR family outside of the active site region (Nunes-Düby et al., 1998). The subsequent core-binding domain (CB) binds to the recombination signal DNA sequence-specifically, allowing to position the recombinase on the so called core recombination DNA sites. Many TRs are composed of only CAT and CB domains, whereas others possess an additional arm-binding (AB) domain. AB is the most variable domain in TRs with remarkable structural and sequence diversity (Smyshlyaev et al., 2021). The AB domain binds to regulatory arm DNA sites, which are located at some distance from the core DNA sites. AB-containing TRs need to bind both core and arm sites simultaneously to recombine DNA, which brings an extra layer of regulation for the functioning of these TRs. Based on the presence of AB, TRs can be classified into two big families: simple TRs, and AB-containing TRs (Smyshlyaev et al., 2021). While the Cre and Flp recombinases are the well-known examples of simple TRs (Van Duyne, 2015; Jayaram et al., 2015), the λ phage integrase (Int_{λ}) is the most studied representative of AB-containing TRs (Van Duyne, 2005).

1.4.3 Mechanism of Tyrosine Recombination

The recombination starts with the binding of two TR molecules to the core recombination DNA site, which features two palindromic core repeats. Each TR monomer forms a C-shaped clamp around one core repeat. The protein binding sites are separated by a central region called the crossover region. This region is 5-8 bp long and it is the place where DNA cleavage and exchange take place. The two recombining DNAs are then brought together to form a tetrameric synaptic complex, which adopts a semi-fourfold planar conformation to allow proteins to efficiently exchange the DNA strands (Figure 1-3).

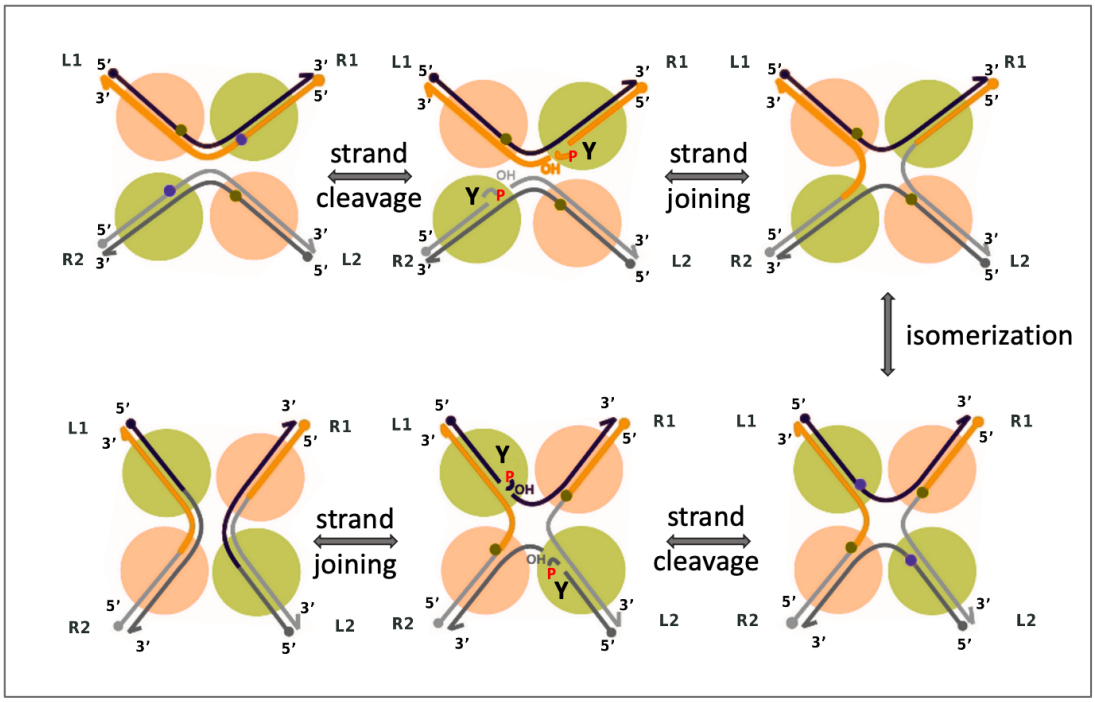


Figure 1-3: Overview of the tyrosine recombination pathway. Recombinases (orange and green circles) bind to their recognition sequence in each DNA (black, orange and grey lines) and form a tetrameric synaptic complex, in which only two protein subunits are active. Then, the first DNA cleavage is performed (see blue dots) forming a 3'-phosphotyrosine linkage (marked with red P) in two protein monomers and free 5' OHs are released (OH). In the next step, the 5' OH attacks the 3'-phosphotyrosine linkage of the other DNA partner, leading to strand exchange. This generates a Holliday Junction (HJ) intermediate. After isomerization, the other protein pair is activated and cleaves the other strands of the DNA (see new position of the blue dots), which leads to resolution of the HJ intermediate and formation of the recombined products (based on Grindley et al., 2006).

Recombination is initiated with the cleavage of one strand in each DNA substrate precisely at the 5' boundaries of the crossover region. For this, the nucleophilic tyrosine attacks the sugar-phosphate backbone of DNA and the scissile phosphate is transferred to the tyrosine hydroxyl group releasing a free hydroxyl group on the cleaved DNA strand (Figure 1-4). As a result of this attack, a covalent protein-DNA phosphotyrosine linkage is formed at the 3' side of the nick on both DNA sites, which preserves the bonding energy coming from the DNA backbone. As a next step, the free 5' OH of the nicked DNA strands exchange places inside the tetrameric recombination synapse and attack the phosphotyrosine linkages of the opposing DNA molecules, generating a Holliday Junction (HJ) DNA structure (Figure 1-3). Next, the recombination complex undergoes conformational isomerization, which changes the activity of its protein subunits, making the previously active monomers inactive and vice versa. Then, the same DNA cleavage and strand exchange reactions are performed by the newly activated recombinase pair on the other strands. At the end of the reaction, HJ is resolved to generate the recombined products and the recombinase proteins are released (Grindley et al., 2006).

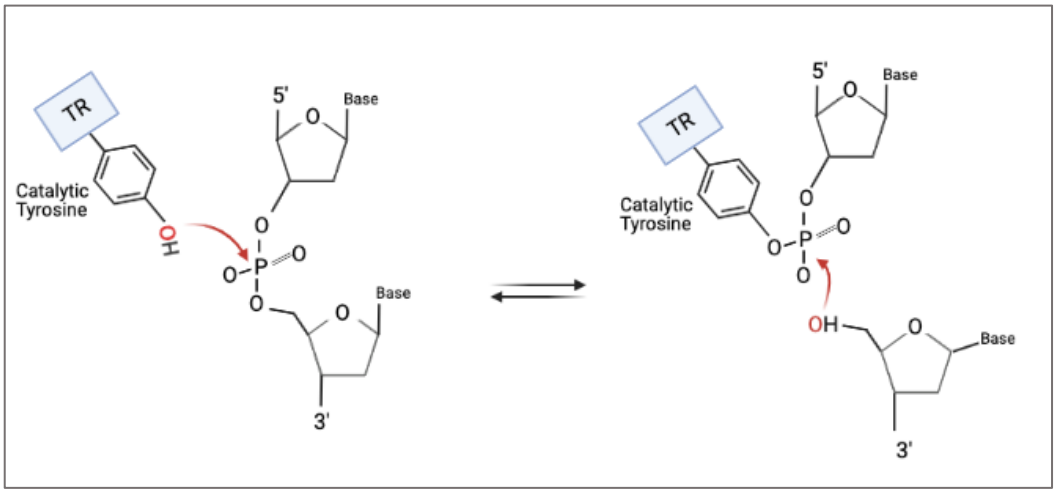


Figure 1-4: The transesterification reaction of tyrosine recombinases (TR). Schematic representation of the nucleophilic attack catalyzed by a TR (in blue). The hydroxyl group of the catalytic tyrosine attacks the scissile phosphate of DNA. As a result of this reaction, a covalent 3'-phosphotyrosine bond is formed and a free 5'-OH is released. This 5'-OH then attacks the 3'-phosphotyrosine intermediate formed on the partner DNA strand, leading to ligation of the first recombined strand pair (based on a figure from Grindley et al., 2006).

While the nucleophilic attack is performed by a tyrosine sidechain, a number of the residues in the enzyme active site help stabilize the reaction intermediates. These include two arginines, a histidine, and a lysine, which together with the nucleophile tyrosine form a conserved RKHRY motif (Grainge & Jayaram, 1999). The RKHR residues are required to coordinate the scissile phosphate, promote nucleophilic attack of the tyrosine hydroxyl, stabilize the transition state, and protonate the released free hydroxyl. Consequently, mutations of any of these residues are detrimental to the recombination reaction rates and mutating the tyrosine or any of the arginines, in particular, results in catalytically inactive protein. Typically, the active site of a TR is assembled within one protein monomer, although strand cleavage can be stimulated by allosteric contacts with an adjacent monomer. This is referred to as active site assembly *in cis*. However, in some TRs the active site is assembled using residues from both monomers, called assembly *in trans*. While the Cre recombinase and the λ integrase are good examples for *cis* assembly, Flp is one of the rare recombinases which assembles *in trans* (Grindley et al., 2006).

1.4.4 Prototypes of the TR Family

1.4.4.1 Cre Recombinase

The Cre recombinase is a well-known member of the TR family, which is extensively used as a genetic tool in eukaryotes. It is considered as one of the model TRs, due to a great extent of biochemical and structural data, which improved the understanding of tyrosine recombination (van Duyne, 2015).

The Cre recombinase was first found in bacteriophage P1, where it acts to resolve fused dimers of the circular phage genome after replication (Ikeda and Tomizawa, 1968). Cre binds to 34-bp long recombination sites, which are called loxP. The loxP site contains two 13-bp core repeats placed as inverted repeats. Between these sites, there is a 6-bp crossover region, where the cleavage takes place. Cre is a simple TR, which has only the CB and CAT domains, without an AB domain (Meinke et al., 2016) (Figure 1-5).

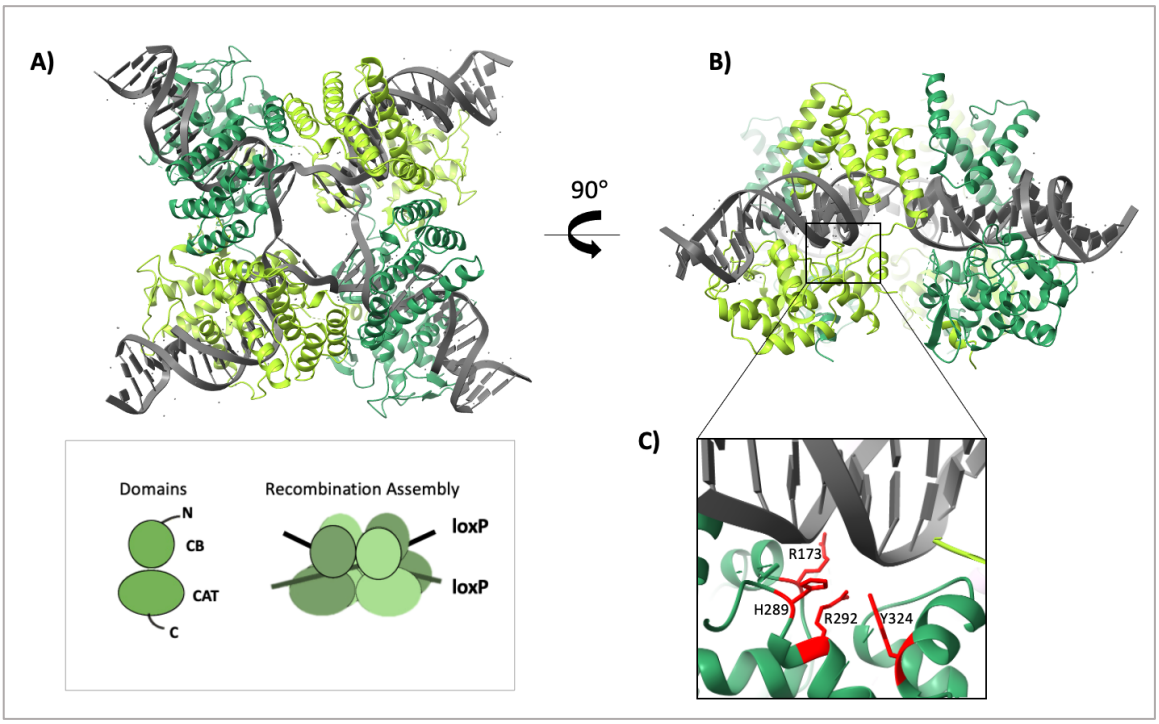


Figure 1-5: The crystal structure of the synaptic Cre recombinase (green) in complex with loxP sites (grey). A) and B) show the top and side views of the tetrameric complex, respectively. C) depicts a zoom up of the catalytic site (red), which includes the nucleophile Y324 residue (PDB 2CRX) (Gopaul et al., 1998). The schematic representation of the domain architecture and the recombination assembly of Cre recombinase is shown at the left bottom corner.

Several three-dimensional (3D) structures of the Cre recombinase are available, which visualize distinct snapshots along the recombination pathway. With the support of biochemical data, these structures have described all the steps of Cre recombination and it now stands as a model system for the TR family. The first crystal structure has revealed four Cre molecules bound to two loxP sites in a synaptic complex. Two Cre molecules bind to one loxP site slightly asymmetrically, despite the symmetric inverted repeat architecture of the site (Guo et al., 1997). One Cre molecule on each loxP site was trapped in a covalent intermediate, where the DNA scissile phosphate is covalently linked to the catalytic tyrosine residue (Y324). The small asymmetry of the complex allowed to create one active monomer on each loxP site, with its nucleophile tyrosine positioned close to the DNA cleavage site, allowing cleavage and covalent intermediate formation. In the other Cre monomer, Y324 is more distant to the cleavage site, preventing cleavage of the second DNA strand. Binding of Cre was found to bend the loxP site by inducing a kink adjacent to the scissile phosphate; this results in unstacking of bases, which lowers the energy barrier for subsequent strand exchange (Guo et al., 1997).

Various structures of Cre synaptic complexes have shown that the tetrameric complex is in a nearly square planar conformation (Figure 1-5) (Guo et al., 1997; Gopaul et al., 1998; Guo et al.,1999; Ennifar et al., 2003; Ghosh et al., 2007). Two loxP DNA sites are located in antiparallel orientation, where the two Cre dimers bound to each DNA site are related by rotational two-fold symmetry. Recombination starts with first cleavage at the bottom DNA strands (Hoess et al., 1986) and continues with the strand exchange, creating a HJ intermediate. The crystal structures of Cre bound to HJ intermediate showed that the crossover region was unstacked but fully base paired. While the bottom strands are sharply kinked, the top strands have no kinks (Gopaul et al., 1998; Martin et al., 2002; Ghosh et al., 2005). The overall arrangement of the Cre-HJ complex is remarkably similar to the synaptic complex structures, suggesting that all DNA cleavage, strand exchange and ligation reactions can occur efficiently, without major conformational changes.

1.4.4.2 Flp Recombinase

Flp is one of the few known eukaryotic TRs. It is encoded by the 2 µm plasmid of *Saccharomyces cerevisiae*, where it contributes to the regulation of plasmid copy number by inverting a DNA segment flanked by specific recombination sites (Broah, 1981). The natural DNA target for Flp-mediated recombination, frt (Flp recognition target) comprises one dispensable and two essential 13-bp Flp binding sites separated by an 8-bp spacer (Senecoff et al., 1985).

Flp is highly diverged in sequence from the prokaryotic TRs (its catalytic domain shares only 13% sequence identity with Cre's), and its active site is assembled differently. It performs DNA cleavage *in trans*, meaning that the catalytic tyrosine (Y343) is supplied to the active site from a neighboring monomer (Lee et al., 1999). Like Cre, Flp is also a simple TR, composed of only CAT and CB domains, and it does not require any accessory components for recombination.

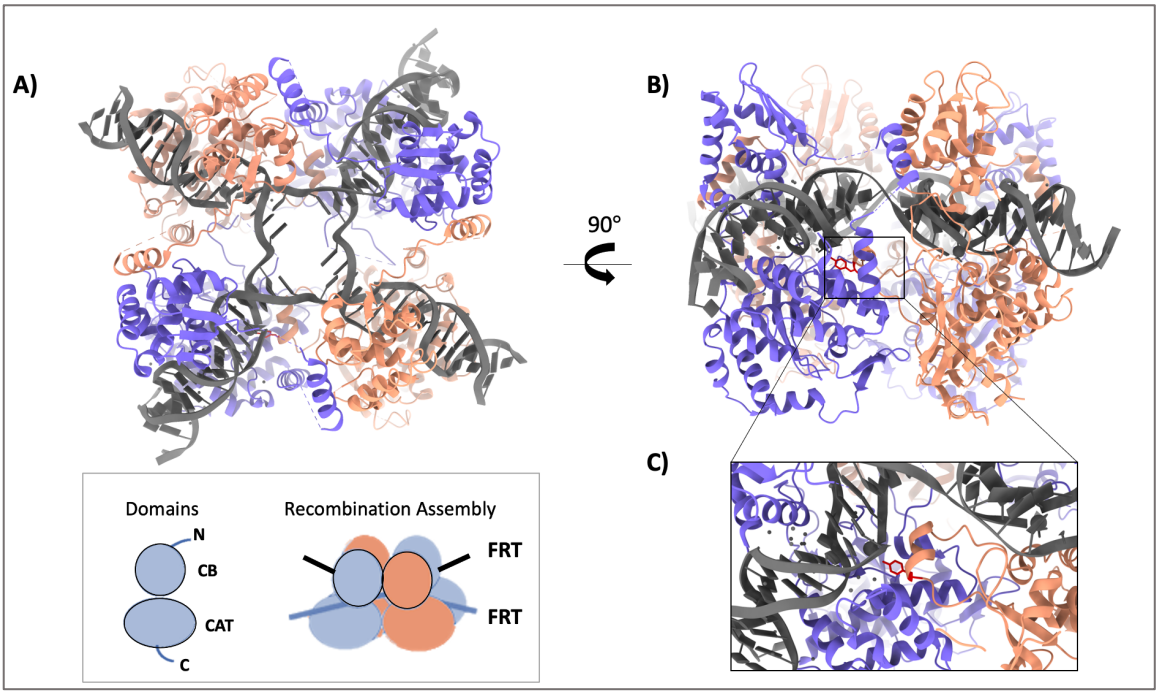


Figure 1-6: The crystal structure of the synaptic Flp recombinase (orange and blue) with frt sites (grey). A) and B) show the top and side views of the tetrameric synaptic complex with two frt sites. C) depicts a zoom up into the catalytic site, which includes the catalytic Y343 residue (PDB 1FLO) (Chen et al., 2000). The schematic representation of the domain architecture and the recombination assembly of Flp recombinase is shown at the left bottom corner.

While the recombination reaction is overall quite similar to the one performed by Cre, the crystal structures of Flp provided unique mechanistic details about the *trans* active site assembly. The active site of one Flp monomer is completed by a protruding helix (annotated as helix M) from the neighboring monomer, which carries the catalytic tyrosine (Y343) (Figure 1-6) (Chen et al., 2000). When compared to Cre, Flp has a longer linker right before M-helix, allowing it to cross between protomers in the complex (Chen et al., 2000). The position of this helix is stabilized by various inter-subunit interactions, and the incoming tyrosine itself is positioned by H309 of the receiving monomer.

In the tetrameric recombination assembly, as in Cre, two Flp monomers are in an active conformation, which are located diagonally in the synaptic tetramer, one on each frt site. While in the active monomer, the M-helix is positioned close to the scissile phosphate, in

the inactive monomer it is mostly disordered. Like in Cre, the frt sites synapse in antiparallel fashion and form a tetramer in a nearly square planar conformation (Figure 1-6). The active and inactive catalytic sites differ mainly in the location of the nucleophilic tyrosine, which are located near the DNA in the active catalytic centers while absent in the others (Chen et al., 2000).

1.4.4.3 Tn*1549* Integrase

The Tn*1549* element is a CTn from the Tn*916* family, which has a major role in the global spread of vancomycin resistance among many different strains of Gram-positive bacteria including Enterococci and Clostridium species (Launay et al., 2006; van Hal et al., 2016). It is a 34-kb long element, flanked by terminal imperfect inverted repeats. Tn*1549* encodes for an integrase (Int_{Tn1549}), which catalyzes the DNA cleavage and joining reactions during transposition. Int_{Tn1549} is an AB-containing TR, which binds to core recombination sites found at the transposon ends. This site contains two 11-bp long inverted core repeats, flanking a crossover site. In addition to the core site, Int_{Tn1549} needs to bind regulatory arm sites which need to be facilitated by accessory factors.

Int_{Tn1549} is so far the only structurally characterized TR from a CTn (Figure 1-7) (Rubio-Cosials et al., 2018). The crystal structure of Int_{Tn1549} revealed a dimeric state of two integrases bound to one DNA molecule containing left and right-end core repeats, flanking a 5-bp crossover region at the center, resembling an excised circular transposon intermediate (Figure 1-7). In this structure, the flexible AB domain of Int_{Tn1549} was truncated to facilitate crystallization. A homodimer of Int_{Tn1549} is bound to the RE and LE core repeats, forming a near-perfect two-fold symmetric assembly. The CB and CAT domains form a clamp around the DNA, which is bent at the crossover site. The crossover site is largely melted and distorted due to Int_{Tn1549}-binding. Moreover, it was shown that destabilization of one base-pair through base-flipping right after each transposon-end at the CR, is crucial for proper strand exchange (Figure 1-7).

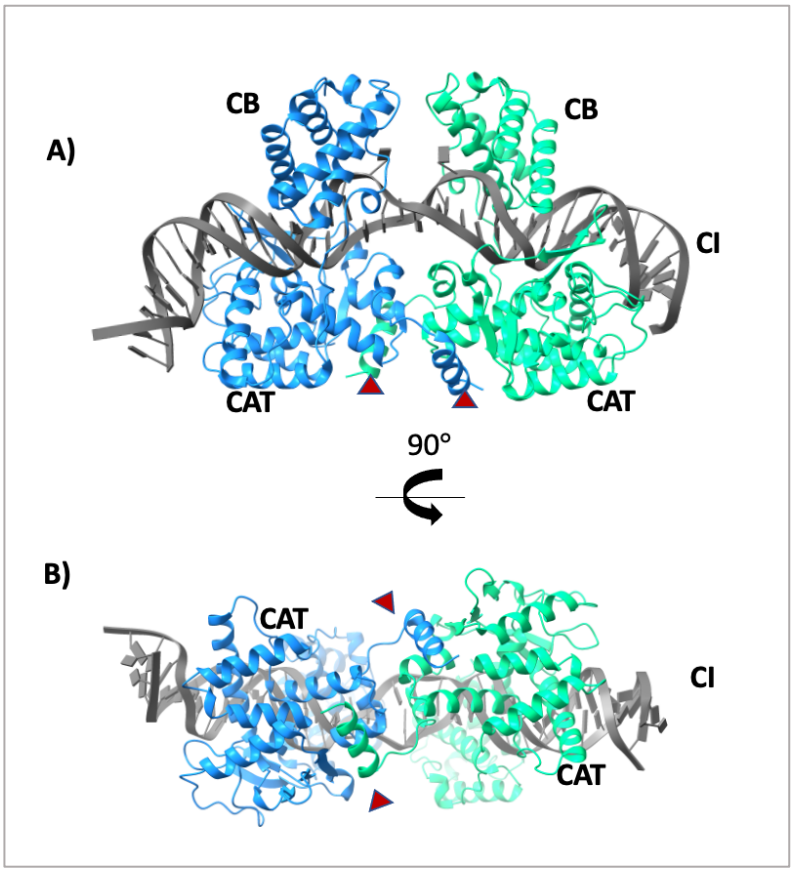


Figure 1-7: The crystal structure of the Tn1549 Integrase (blue and green) in complex with CI DNA (grey). A) shows the side view of the dimeric complex. B) shows the bottom view with 90° rotation. The exchanged C-terminal helices are highlighted with red arrows (PDB 6EMZ, Rubio-Cosials et al., 2018).

The active site of Int_{Tn1549} is assembled *in cis*, with the catalytic residues provided by the same subunit. In the dimeric complex, the C-terminal helices are exchanged reciprocally to stabilize complex formation, leading to an inactive state due to the positioning of the catalytic tyrosine in each monomer. In the tetrameric assembly, this exchange is also expected to stabilize complex formation, but it will occur in a cyclic manner, leading to the activation of two non-neighboring monomers (Rubio-Cosials et al., 2018).

1.4.4.4 λ Integrase

The λ Integrase (Int_{λ}) is another founding member of the tyrosine recombinase family. It is encoded by the Bacteriophage λ, where it catalyzes excision and integration of the phage genome into the *Escherichia coli* host genome (Lederberg, 1953). The binding sites of Int_{λ} are called attachment sites (att). For integration, it recombines attP, in the phage genome with attB, a specific site in the bacterial genome. As a result of integration, the phage genome is inserted into the bacterial genome and flanked by two att sites called attR and

attL.

The Int$_\lambda$ assembles its active site *in cis* similarly to the Cre recombinase with all active site residues supplied by one protomer. The att sites contain two inverted integrase binding sites connected with a 7-bp crossover region, where the cleavage takes place (Landy and Ross, 1977). DNA cleavage and exchange are initiated on the top strand, and then proceed to the bottom strand (Christ and Droge, 1999).

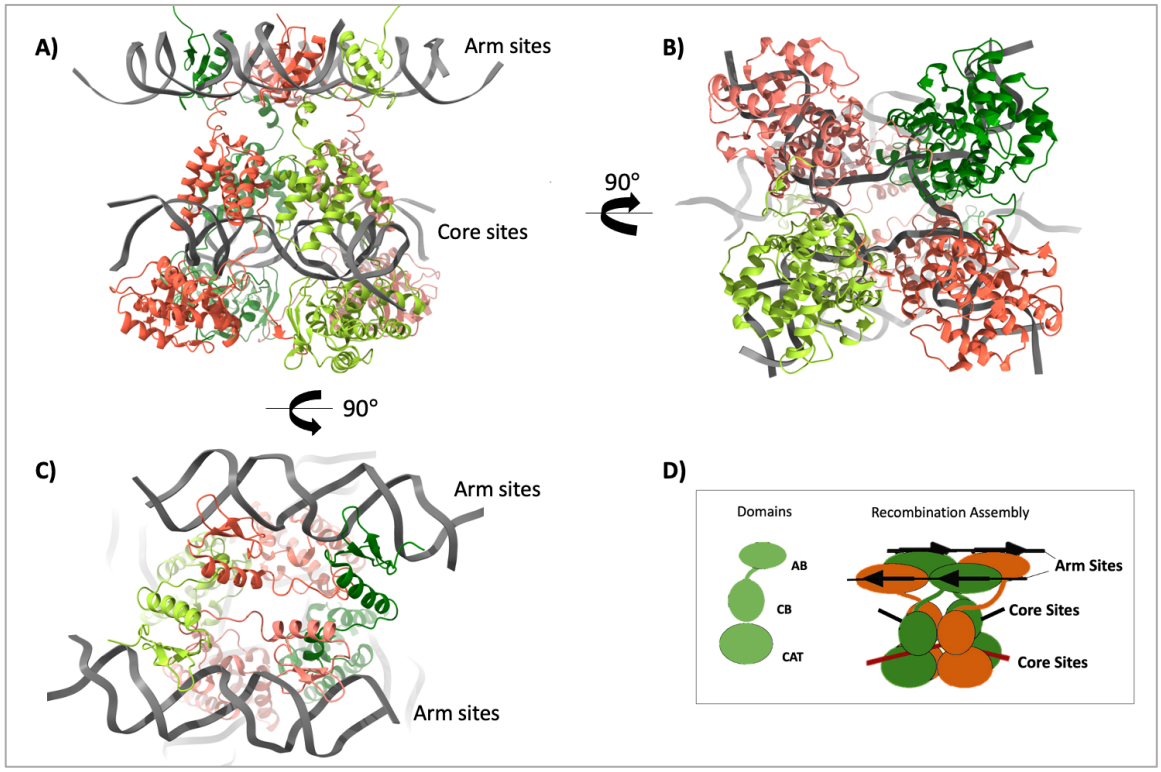


Figure 1-8: The crystal structure of the λ integrase (shades of green and red) in complex with HJ and arm DNA (grey). A) and B) show the side and bottom views of the tetrameric complex, highlighting how the CAT domain is assembled on HJ. C) shows the top view, showing how the AB domains are bound to two antiparallel arm DNA sites (PDB 1Z1B) (Biswas et al., 2005). D) shows the schematic representation of the domain architecture and the recombination assembly of λ integrase.

The Int$_\lambda$ is a complex TR, containing CAT, CB and AB domains (Figure 1-8 D). The AB domain confers extra regulation to λ recombination compared to Cre and Flp. AB binds to regulatory 'arm' DNA sites that are present in the phage genome and plays a crucial role in driving recombination. The CAT and CB domains cannot carry out effective recombination without it. In the absence of AB, the Int$_\lambda$ can resolve a HJ intermediate, but it cannot carry out the complete recombination reaction (van Duyne, 2005). Arm sites are located distantly from the recombination core repeats in the phage genome. During recombination, the Int$_\lambda$ must simultaneously bind both the arm and core repeats, which are brought to close proximity with the help of accessory proteins (Landy, 2015). Successful

recombination by Int$_\lambda$ depends on intricate interactions between the Int$_\lambda$, both arm sites and core repeats, and a number of helper proteins assembled together in a highly ordered recombination complex.

While Int$_\lambda$ binds to core repeats with low affinity, it binds to arm sites with higher affinity by the AB domain and glue the recombination complex together (Ross and Landy, 1983). The crystal structure of the Int$_\lambda$ bound to arm sites and HJ forming core DNA showed that the arm binding domains form a tightly packed dimer of dimers that is connected by a short linker to the core binding (CB) domains (Figure 1-8) (Biswas et al., 2005). While this enzyme follows a more complex pathway than Cre and Flp, the organization of the CAT and CB domains is overall fairly similar in all synaptic and HJ DNA complex structures (Biswas et al., 2005).

1.5 Higher Order Protein DNA Complexes in Tyrosine Recombination

High-precision DNA rearrangements generally require the formation of a specific nucleoprotein architecture, where protein and DNA form a distinct higher-order assembly. The complex TRs are good examples of this phenomenon. For these cases, the recombination machinery can assemble on DNA only when it is bent in a particular conformation to bring all the necessary elements to the right constellation for activity. These distinct DNA conformations are achieved with the help of accessory proteins, which bend and twist DNA. Since the particular DNA conformation is critical for the formation and correct action of these complexes, the action of accessory proteins brings a new level of control to recombination.

The accessory proteins that assist DNA recombination are generally small, positively charged proteins, which can alter DNA topology. They can bind DNA either in a sequence specific or unspecific fashion and they can bend DNA at diverse angles. Besides regulating DNA rearrangements, some of them are essential for shaping genome architecture, regulating replication, transcription and DNA repair (summarized in Flores-Ríos *et al.*, 2019).

While bacteria harbor such DNA bending proteins for their own genome organization and regulation, various MGEs and phages also encode similar proteins that promote their mobilization. These elements, mainly CTns and lysogenic phages, typically rely on a set of such accessory proteins, including self-encoded and host-derived factors to control the path and efficiency of site-specific DNA recombination during their integration and excision.

In the following sections, I will describe three well-known accessory proteins that are frequently employed in site-specific DNA recombination of MGEs in Gram-negative bacteria.

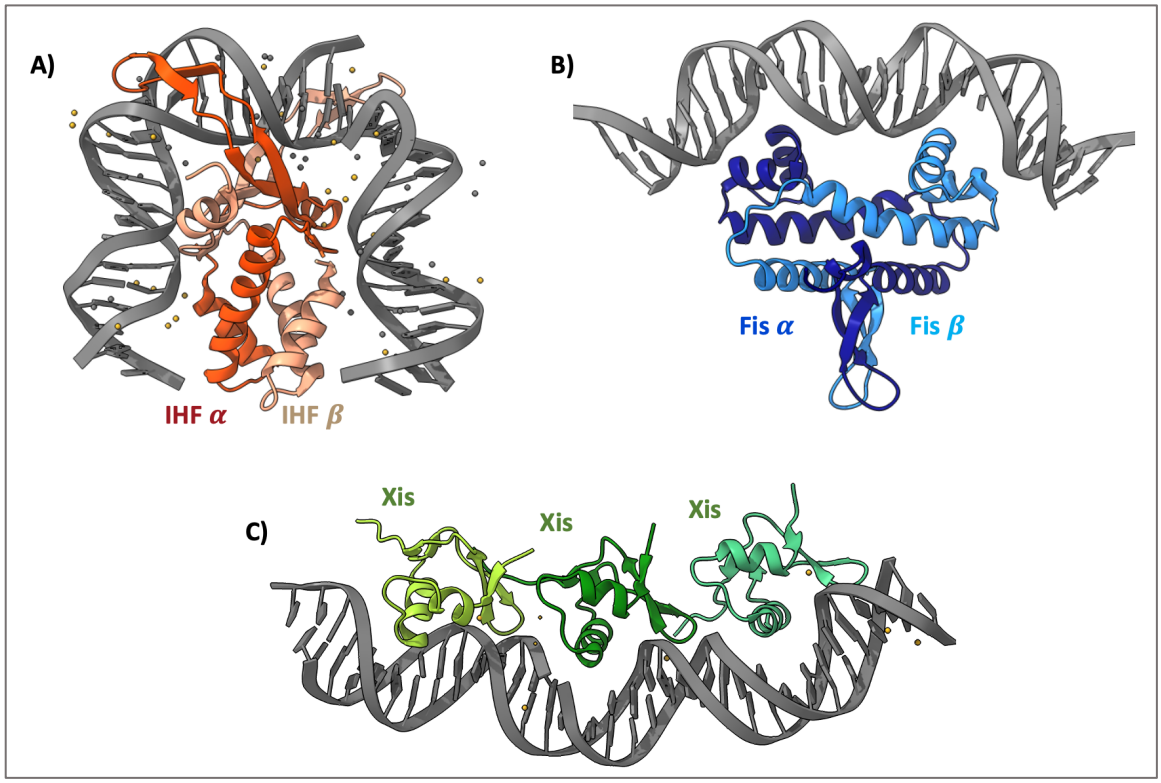


Figure 1-9: The accessory factors IHF, Fis and Xis. Crystal structures of A) IHF (PDB 1OUZ, Rice et al., 1996), B) Fis (PDB 3JR9, Stella et al., 2010) and C) Xis (PDB 2IEF, Abbani et al., 2005) in complex with DNA (grey). The IHF heterodimer can induce a U-turn in DNA, whereas the Fis heterodimer bends DNA less prominently. MGE-encoded Xis molecules bind one after another, introducing various bending degrees in a cooperative fashion.

1.5.1 Integration Host Factor (IHF)

Integration Host Factor (IHF) is a small DNA binding protein, which is found in Gram-negative bacteria. It is an abundant architectural protein, which changes the conformation of DNA by introducing a sharp bend. Although it was first discovered as a host factor required for λ phage integration, it assists many processes that involve higher-order protein DNA complexes, such as replication, transcription regulation and a variety of site-specific DNA recombination pathways (Nash and Robertson, 1981; Goosen et al., 1995).

IHF can induce a U-turn on DNA by bending it more than 160 degrees, which makes it one of the sharpest benders (Figure 1-9 A). It is a heterodimeric ~20 kDa protein composed of α and β subunits. The two subunits have similar fold (with ~30% sequence identity) and are

intertwined to form a compact core, from which two long β hairpin arms extend. The 3D structure of IHF is closely related to that of HU, which is the most abundant histone-like protein of *E. coli.* The β-stranded arms of IHF wrap around the DNA minor groove and induce two sharp kinks by interrupting base stacking via intercalation of proline residues. The bend is further stabilized by interactions of the positively charged IHF core with the sugar-phosphate backbone of both arms of the U-shaped DNA structure (Rice et al., 1996).

IHF exhibits weak sequence specificity and can bind many different sequences with different affinities. It occupies 25 bp of DNA, of which only 9 bp are conserved (Goodrich *et al*., 1990). Many IHF binding sites include an A-tract (four or more consecutive A bases). If this tract is not present, its addition increases the affinity of IHF to this substrate DNA.

1.5.2 Excisionase (Xis)

Excisionase (Xis) is a small sequence-specific DNA binding protein, which controls the directionality of recombination reactions performed by phage- and CTn-related tyrosine recombinases (Burrus et al., 2002). Available data indicate that Xis bends MGE DNA to facilitate the assembly of a productive excision complex, whereas it inhibits the formation of an integrative complex (Abremski and Gottesman, 1982).

There are different Xis proteins found in various CTns (i.e. Tn*1549*, Tn*1545*, Tn*916*, etc.) and phage-related recombination systems (λ, L5, T12, etc.), which are significantly variable in their sequences (Delavat et al., 2017).

Most Xis proteins contain a winged-helix motif, which interacts with the major and minor grooves of the DNA through a single α-helix and a loop structure, respectively (Figure 1-9 C, Sam et al., 2004). Several Xis binding sites are usually placed side-by-side in the DNA so that multiple Xis monomers form a DNA-bending filament. It has been shown that Xis proteins from different MGEs can bend DNA at different angles. While a single Xis protein bends DNA by 20-40°, a Xis filament (including 3 or 4 Xis molecules) can change the DNA conformation drastically (Abbani et al., 2007).

1.5.3 The Factor for Inversion Stimulation (Fis)

Fis is a small and versatile accessory protein, known to contribute to many different DNA metabolism related processes and to affect global transcription and bacterial chromosome organization (Grainger and Busby, 2008). It is a DNA bending protein, which functions as a

homodimer binding to a 17 bp-long AT-rich consensus sequence via recognizing the shape of its minor groove (Figure 1-9 B) (Cheng et al., 2000; Stella et al., 2010).

Fis was first identified for its role in regulating recombination reactions carried out by the DNA invertase, Hin (Haykinson et al., 1996). Other activities for Fis were identified later, including transcriptional activation of a wide number of promoters, repression of another set of promoters, cofactor for DNA replication, cell division and chromosome segregation (Lewis and Hatfull, 2001). Its expression is maximal at the beginning of the exponential growth phase, because it plays an important role in boosting the expression of genes involved in supplying components of the translation machinery. It is also known to act as a superhelicity monitor of the densely packed bacterial chromosome. Furthermore, Fis family proteins are known to regulate site-specific DNA recombination of MGEs and phages. In *E. coli,* Fis plays a role in λ prophage lysogeny maintenance and dynamics, stimulating both excision and integration of the phage genome. Fis also participates in the mobilization of other elements like the transposable phage Mu, the P1 phage, the Tn5 transposon, IS50 and Class1 integrons (Weinrecih and Reznikof, 1992).

1.5.4 Regulating Directionality by Using Higher-Order Nucleoprotein Architecture: λ Integrase as a Prototype

Effective integration and excision of the phage genome is critical for the establishment and termination of the lysogenic life cycle during bacteriophage λ propagation. These reactions are catalyzed by the λ integrase (Int_{λ}) and are tightly regulated in response to a variety of physiological and environmental signals.

The most intricately controlled feature of the process is the directionality of the reaction. Since integration and excision are isoenergetic reactions, which are simply the reverse of each other at the DNA level, the desired directionality of the reaction must be ensured by an additional mechanism. For this, λ requires the formation of two distinct nucleoprotein complexes to carry out either integration or excision.

As described above (1.4.2.2.3), each end of the phage genome (attR and attL) contains a core and an arm site, which are distantly located from each other and need to be brought closer to simultaneously bind to the integrase tetramer. To form a functional complex, the CB and CAT domains of Int_{λ} bind to the core site, and the AB domain binds to the arm sites. While the λ genome has only one core site in either attR/attL, it has multiple arm sites at each end, which are differentially used to form integration and excision complexes. To bring different arm sites, which are located in different places and orientations, near the respective core sites, the system requires different DNA topologies to be established with

the help of different accessory proteins. In this way, integration and excision complexes can be differentiated from each other to control the reaction outcomes via precise positioning of different regulatory arm sites. This control makes Int_{λ} recombination a conditional, effectively irreversible molecular switch in bacteriophage life cycle.

Int_{λ} cooperates with both host and self-encoded accessory proteins to drive its integration and excision reactions (Figure 1-10). The phage-encoded Xis protein is the most important factor for determining the directionality of the reaction. It interacts with three consecutive Xis binding sites in attR. If it is present, then the excision complex can assemble, and the integrative complex formation is inhibited. Host-encoded IHF binds both to attR and attL and is essential for both integration and excision complex formation. Fis is known to specifically stimulate integration, via binding next to the Xis binding sites in attR to execute DNA bending and complement for the function of Xis during integration.

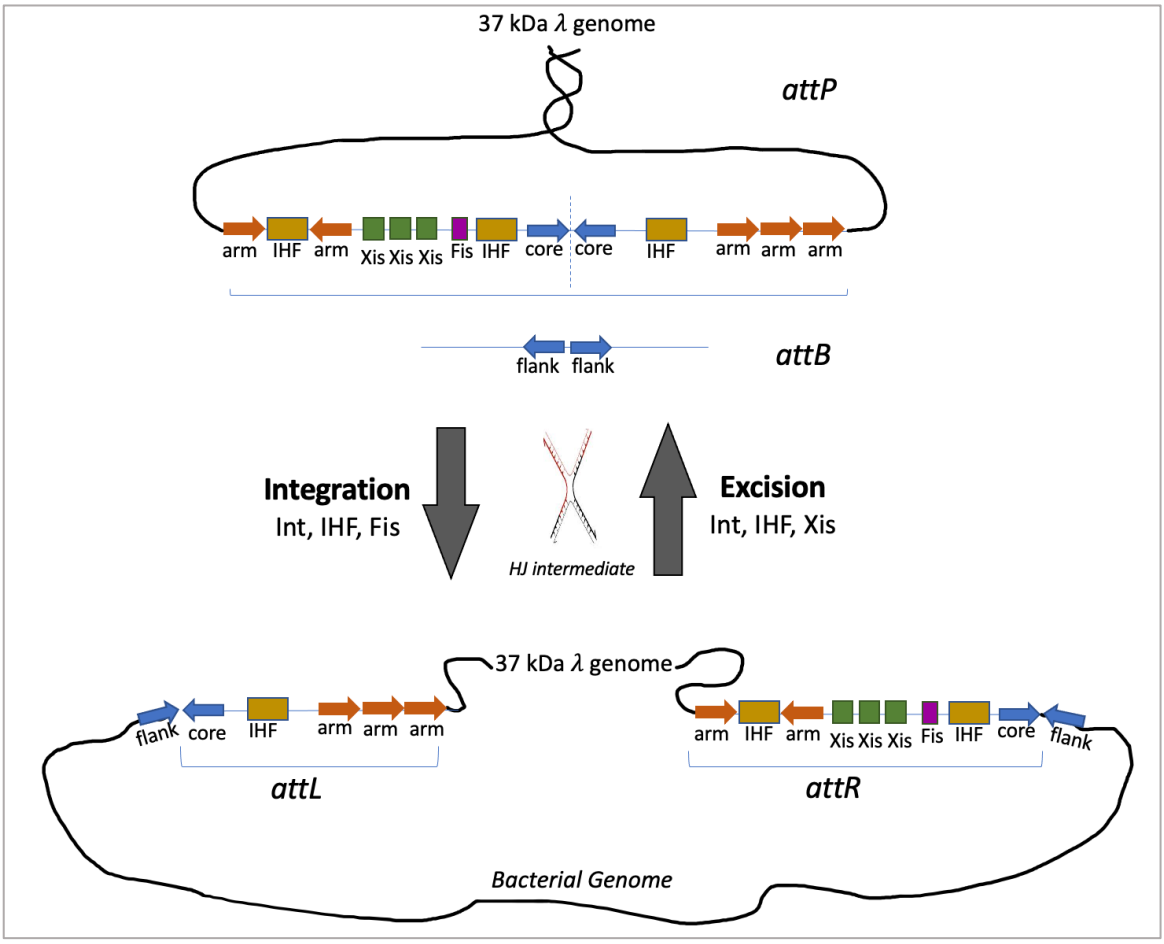


Figure 1-10: The organization of λ phage attachment sites during integration and excision. The schematic representation of att sites is shown. Each att site contains one core site (with an inverted repeat structure, blue arrows) and multiple arm sites (orange arrows). Binding sites for Fis (purple), Xis (green) and IHF (yellow) are also shown. At the top, the attP is shown, which is formed on the circular phage genome. The

recombination of attP with attB (bacterial att site) results in integration. For integration Int and IHF are sufficient, while excision also requires Xis (based on Tong et al., 2014).

The first and the only structure of a higher order nucleoprotein assembly formed during site-specific tyrosine recombination was determined for the excision complex of Int$_\lambda$ by cryo-EM at 11 Å resolution, which was trapped on the HJ DNA intermediate (Figure 1-11) (Laxmikanthan et al., 2016). At the center of the complex, four integrase molecules are bound to the HJ DNA, with the CB and CAT domains siting on the core DNA sites in the four arms of the HJ. To bring the arm binding sites to the Int$_\lambda$ AB domains, DNA is bent sharply by IHF next to the core sites in both attR and attL. In attL, the bend induced by IHF is sufficient to bring the arm sites to the integrase tetramer. However, for attR, an extra twist is required to reach the respective Int$_\lambda$ AB domain, which is mediated by three consecutively bound Xis proteins (Figure 1-11). While most of the components of the excision complex are resolved in the final EM map, the AB domain of one of the integrase subunits is not present. It has been proposed that this domain was not able to bind to its arm site, since Xis was occupying this part of DNA. Without binding to DNA, AB was disordered and could not be observed in the final EM map (Laxmikanthan et al., 2016).

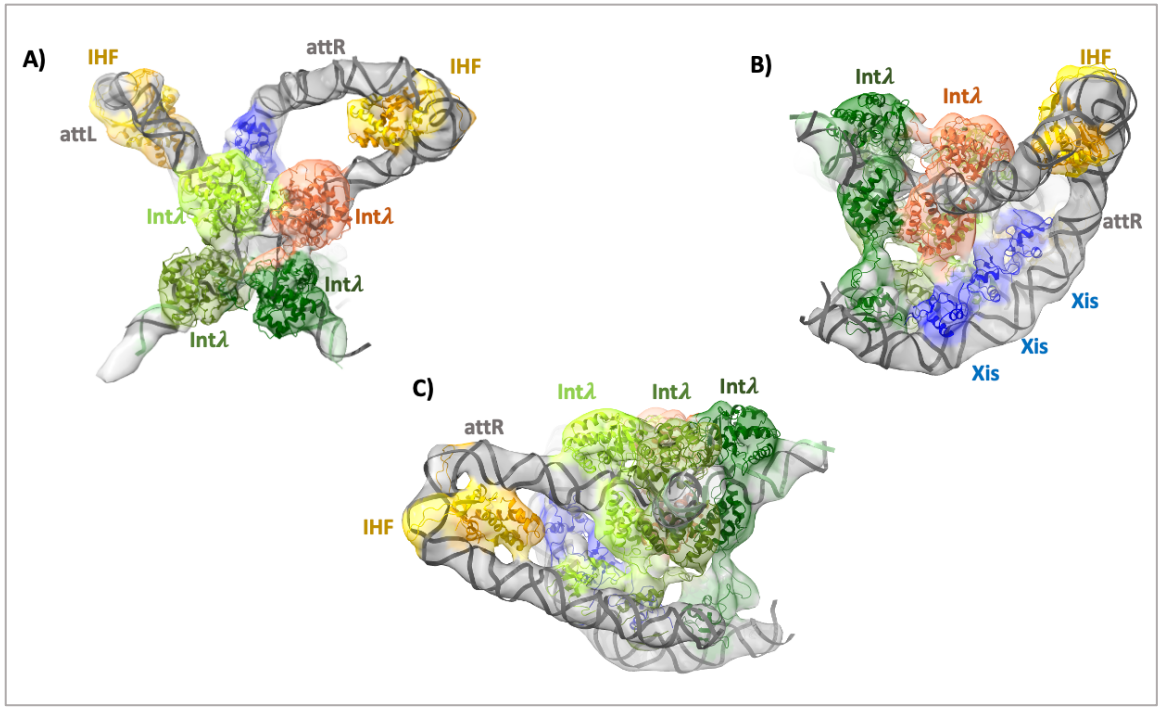


Figure 1-11: The cryo-EM map of the Int$_\lambda$ with HJ DNA intermediate. The cryo-EM map of the complex is shown as transparent surface. The right attachment site (attR, in grey) is bent by IHF (in yellow) and three Xis molecules (in blue), whereas the left attachment site (attL, in grey) is bent only by IHF. The ends of the bent DNA are connected by the integrase tetramer (in shades of green and orange). In the center of the CAT domain tetramer, the partially recombined Holliday Junction intermediate is visible (PDB 5J0N, Laxmikanthan et al., 2016).

1.6 *GISul2* element: a Genomic Island Carrying Sulphonamide Resistance Gene

1.6.1 Identification of Int$_{GISul2}$ as the Most Wide-spread Integrase in Gram-negative Bacteria

As mentioned in the previous sections, TRs are frequently found on mobile genetic elements that carry ARGs. Indeed, they can mobilize diverse ARG-carrying elements, including Integrons and conjugative transposons (Partridge et al., 2018).

To study the contribution of TRs to ARG-spread systematically, Georgy Smyshlyaev, a postdoc in the Barabas laboratory, developed a computational method that identifies horizontally transferred genes in bacterial genomes. In brief, the method quantifies how often a gene is found in unrelated species. With this pipeline, he identified TRs that are the most widely distributed in bacteria. The most ubiquitous TR was found to be the Integron integrase (IntI), which is found in many Gram-negative bacteria. IntI is associated with Integron gene cassettes, which contain diverse ARGs and play an important role in bacterial adaptation and resistance dissemination (Partridge et al., 2018). IntI catalyzes the recombination of gene cassettes within Integrons, enabling the activation of dormant ARG reservoirs on demand and the incorporation of new ARGs from environmental DNA pools. The intercellular mobility of Integrons does not depend on the activity of the IntI, rather they are propagated on transposons or by plasmids (Gillings, 2014). The integrase of the well-characterized tetracycline resistance carrying Tn*916* was identified as the second most abundant TR overall and the most widely distributed CTn-borne TR, with a very broad appearance in Gram-positive bacteria (Franke and Clewell, 1981). The third most widely distributed TR was an uncharacterized integrase from the *GISul2* element. Int$_{GISul2}$ was the most widespread CTn-like integrase found in Gram-negative bacteria (Smyshlyaev *et al.*, unpublished data).

1.6.2 The *GISul2* Element

GISul2 is a 15.5 kb long MGE, which was first identified in gamma-proteobacterial species (Nigro and Hall, 2011). It is inserted at the 3' end of the *guaA* gene in several distant Gram-negative bacterial genomes and is present in plasmids (Nigro and Hall, 2011). The mobilization of this element was through to be catalyzed by the self–encoded putative integrase Int$_{GISul2}$ (Figure 1-12). It also encodes an excisionase Xis$_{GISul2}$, as well as replication- (*repA* and *repC*) and conjugation-related genes (*trb J, K* and *L*) (Figure 1-12). The presence

of conjugation-related genes suggests that *GISul2* is a conjugative transposon, and the presence of replication related genes raises the possibility that the element can replicate independently once it is excised. As cargo, it encodes arsenic (*ars B, C, H* and *R*) and sulphonamide (*sul2*) resistance genes, which are located on a resident *ISCR2* element captured within *GISul2* (Hamidian & Hall, 2016).

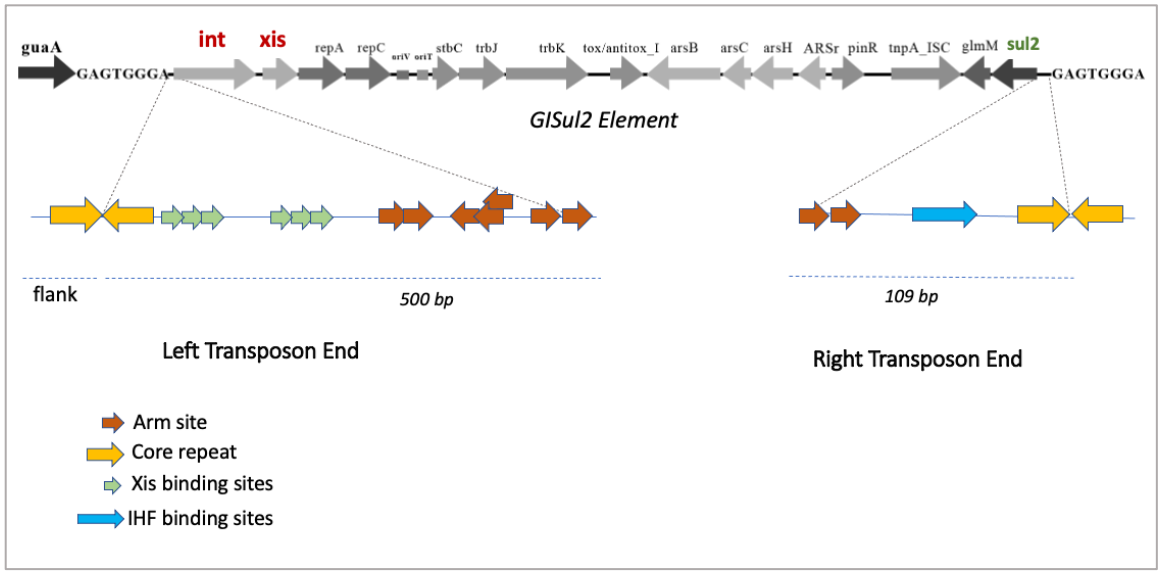


Figure 1-12: The genetic architecture of the *GISul2* element. The genes that *GISul2* element encodes are shown as blue arrows with the gene names. The element is flanked by left and right transposon ends. Each transposon end has distinct set of binding sites for the integrase, including arm (in red) and core repeats (in yellow), and for the accessory factors, IHF (in blue) and Xis (in green).

Int_{GISul2} is an uncharacterized TR from the IntSXT/R391 family. The members of the family are found on MGEs in various Gram-negative bacteria including *Vibrio cholerea*, which confer resistance to antibiotics and help their spread (Waldor *et al*., 1996). These integrases share conserved TR domains: CAT, CB and AB. While most IntSXT/R391 family members seem to promote integration of their transposon at the 5' end of the *prfC* gene, which encodes for peptide chain release factor 3 (Hochhut & Waldor, 1999), the *GISul2* element is mostly located at the 3' end of the GMP synthase gene (*guaA)* (Nigro & Hall, 2011). *guaA* is found in the *gua* operon, which is responsible for the synthesis of guanosine monophosphate. The 3' end of *guaA* (right before its stop codon) is known to be a hotspot for integration for various mobile elements (Song *et al*., 2012).

Like other CTns, the *GISul2* element is flanked by terminal imperfect inverted repeats, which are specifically recognized by the proteins required for excision and integration of the element (Figure 1-12). Based on sequence conservation among *GISul2*-related elements, the essential segment of the transposon ends was predicted (Smyshlyaev *et al.*, unpublished data). The right end (RE) is approximately 110 bp long and includes

recognition sequences for Int$_{GISul2}$ as well as for the bacterial protein IHF. There are two directly oriented arm sites located in the 5' section of RE, a putative core repeat, where CB and CAT bind, is located at the 3' end. Between the arm and core repeats, there is a predicted IHF binding site. For the left end (LE) of the transposon, the predicted architecture is a bit more complicated. In addition to one putative core repeat, several arm sites could be predicted. Between these sites, there are two Xis binding patches, each containing three direct repeats of the Xis binding site. This complex architecture of protein binding sites on both transposon ends suggests a well-regulated recombination pathway, which employs intricate higher order nucleoprotein complexes to precisely control MGE excision and integration.

1.6.3 Structural Insights into Int$_{GISul2}$ Transposon End Complexes

Although the presence of *GISul2* in multiple genomes indicates the mobility of this element, the activity of Int$_{GISul2}$ was not experimentally confirmed. Work in our lab has recently shown that Int$_{GISul2}$ is capable of excising a DNA fragment flanked by the specific transposon ends of the *GISul2* element, when it is co-expressed with Xis (G. Smyshlyaev, unpublished data).

To shed light onto the structural principles of the activity of Int$_{GISul2}$, the enzyme was crystallized bound to DNA substrates containing conserved transposon end sequences (G. Smyshlyaev, unpublished data). The first structure, solved by X-ray crystallography at 2.9 Å resolution, contains Int$_{GISul2}$ bound to a symmetrized RE core DNA sequence, which resembles the excised circular transposon intermediate (Figure 1-13 A). In this structure, the flexible AB domain of Int$_{GISul2}$ was truncated to facilitate crystallization. The structure showed similar architecture to the Tn*1549* integrase – CI complex, which was characterized earlier in our lab (Rubio-Cosials et al., 2018). A homodimer of Int$_{GISul2}$ is bound to the RE core DNA, forming a near-perfect two-fold symmetric assembly. The CB and CAT domains form a clamp around the DNA. Notably, the active sites are assembled *in trans*, with most catalytic residues provided by one subunit of the dimer and the nucleophile tyrosine donated by the other. This arrangement is unlike in Int$_{1549}$, and rather resembles the active site geometry of the Flp recombinase discussed above (Section 1.4.4.2).

The second structure was solved at 3 Å resolution and contains full-length Int$_{GISul2}$ in complex with two DNA molecules, representing the symmetrized core DNA sequence and the arm site of RE, respectively (Figure 1-13 B). As in the previous structure, two Int$_{GISul2}$ molecules bind one core DNA molecule, but the full-length protein forms a remarkably asymmetric arrangement. The CB domain of one subunit is dislocated from the core DNA to enable simultaneous binding of two directly repeated arm DNA sites in the complex.

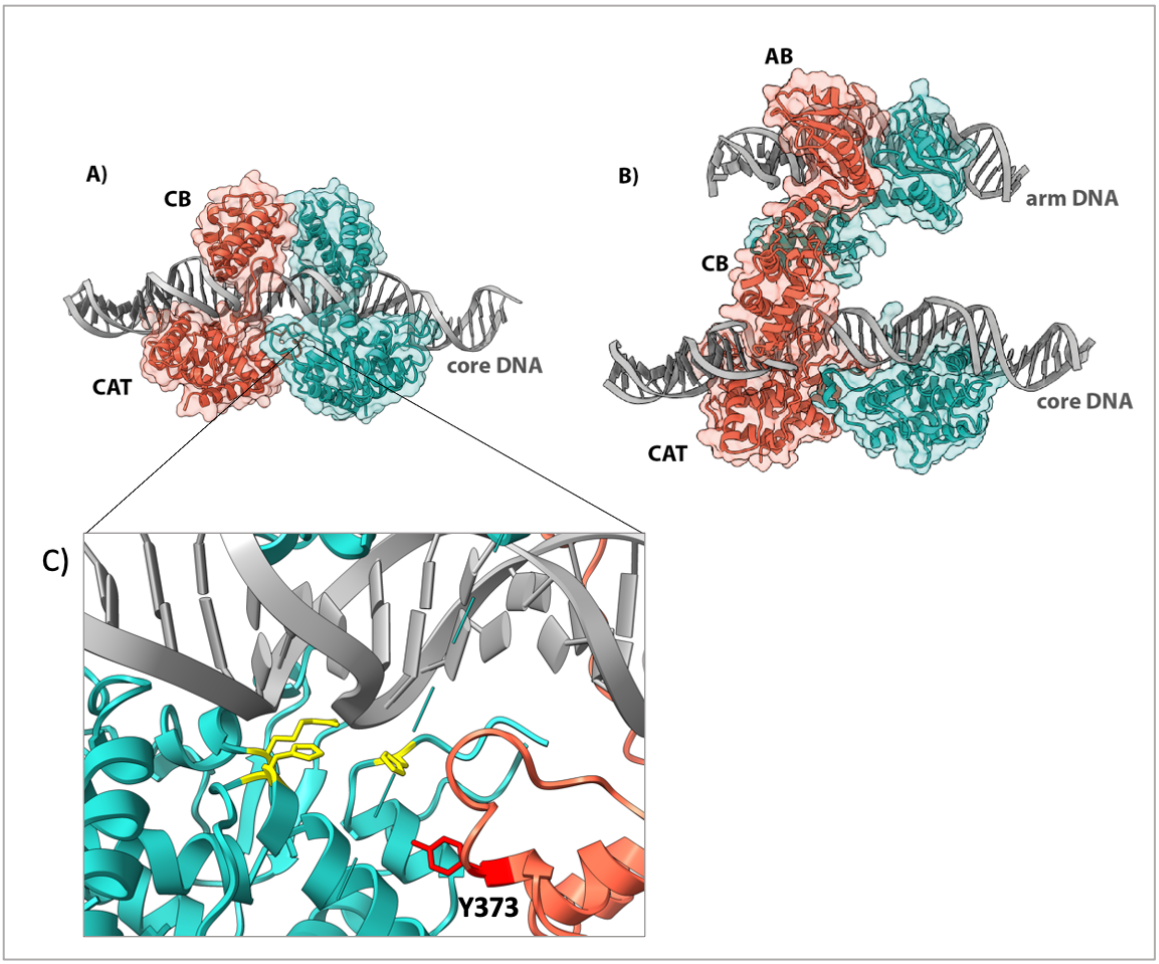


Figure 1-13: Crystal structures of Int_{GISul2} - DNA complexes. Structure of A) the core binding (CB) and catalytic (CAT) domains of Int_{GISul2} bound to symmetrized RE core DNA at 2.9 Å resolution and B) full-length Int_{GISul2} in complex with symmetrized core DNA and arm DNA sequences at 3 Å resolution. Individual protomers of Int_{GISul2} are shown in cyan and orange and the DNA molecules are depicted as grey cartoon. C) shows a zoom-up of one active site of the enzyme, with residues forming the catalytic center represented as sticks. While cis acting residues are shown in yellow, the trans acting tyrosine is shown in red.

1.7 Aim and Goals of the Thesis

The aim of this project was to understand the mechanism of *GISul2* transposon mobilization with a specific focus on identifying the mechanistic features that have contributed to the dissemination of this transposon to a wide range of hosts. Ultimately, this knowledge will help to understand the general rules of CTn movement and promote the development of new strategies to slow down the CTn-dependent ARG spreading.

To shed new light onto the mechanisms of *GISul2*, I aimed to help characterize the excision and integration of the element. While previously solved crystal structures provided detailed understanding about Int_{GISul2} organization, it was still unknown how accessory proteins regulate higher order complex formation at the transposon ends. Thus, the next aim was to structurally characterize protein-DNA complexes formed on complete native transposon ends, focusing on the following three goals:

1) Structural characterization of the transposon right end complex.

2) Structural characterization of the transposon left end complex.

3) Structural characterization of the synaptic complex, where RE and LE are brought together to form an active recombination complex.

By achieving the above-mentioned goals, we wanted to understand how a specific higher order complex formation drives the recombination towards the desired products, what the roles of the accessory factors are in this process, and, finally, how host cells control or contribute to the mobilization of their transposon residents. With this knowledge, we hoped to provide new mechanistic insights into the biology of CTn-mediated ARG spread, and of tyrosine recombinase-mediated DNA rearrangements in general.

I joined this project to help elucidating higher order complex formation in *GISul2* contributing my experience in sample preparation for cryo-EM and data analysis. The first and the third parts of the project were designed and performed together with Dr. Georgy Smyshlyaev (postdoctoral fellow Barabas laboratory, EMBL Heidelberg).

2 RESULTS

2.1 The Right End Complex of the *GISul2* Element

To decipher how accessory proteins facilitate and regulate the transposition of GISul2, we aimed to structurally characterize the higher order nucleoprotein complex, assembled at the transposon ends during excision. Given the large size and potential flexibility of the complexes, we decided to attempt structural characterization using cryo-electron microscopy (cryo-EM). In the next chapter, I will describe the results for the Int_{GISul2} right end (RE) complex assembly.

2.1.1 *In vitro* Reconstitution of the RE Complex

To gain structural insights into *GISul2* RE assembly, the complex was reconstituted *in vitro*. To define the required minimum length of the RE, *GISul2* elements found in different genomes were aligned (Figure 2-1) by Dr. Georgy Smyshlyaev. Based on the sequence conservation, the minimum essential RE was predicted to be 110 bp. This contains core and arm DNA sites, as well as an IHF binding site. The core site is located at the transposon boundary, with a well conserved core repeat and crossover region (5'-GAGTGG-3') inside the transposon and a less conserved core repeat in the flanking genomic DNA. Two arm sites are located as direct repeats one after another 60 bp apart from the core site. To bring arm site and core site close in space for Int_{GISul2} binding, RE is expected to be bent by an IHF heterodimer, which has a conserved binding site between the arm and core sites.

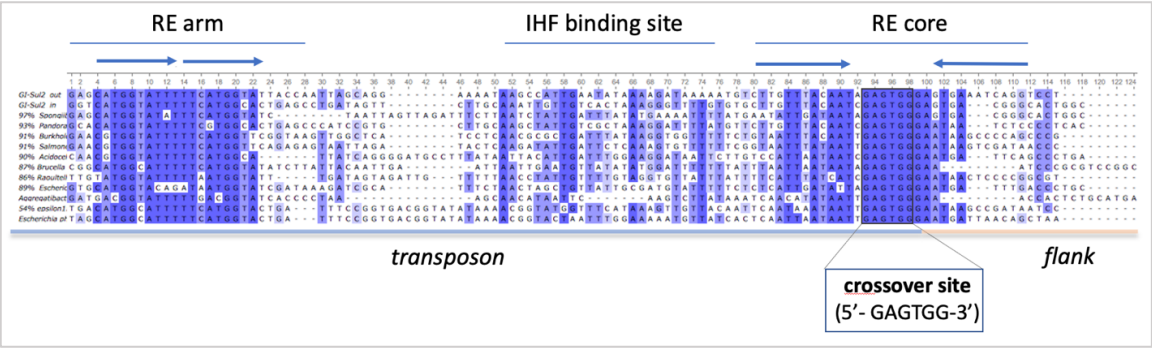


Figure 2-1: The right end (RE) of the GISul2 element. RE was identified based on sequence conservation among *GISul2* copies found in different genomes. Conserved bases are highlighted in blue, the more intense color shows the higher conservation. There are two predicted arm sites, and a putative core site that features imperfect inverted core repeats intersected with the strictly conserved crossover region (5'-GAGTGG-3') in the middle. An AT-rich IHF binding site was also predicted between the arm and core sites. (Analysis performed by Georgy Smyshlyaev).

The protein sequence of Int_{GISul2} is identical in all copies of the element and it is 45 kDa in size. Int_{GISul2} binds to RE arm and core DNA sites as a homodimer in the crystal structure. The purification of Int_{GISul2} from *E. coli* overexpression was optimized previously by Georgy Smyshlyaev in our lab. Preparing a complex of active Int_{GISul2} with its native transposon end can result in DNA cleavage, which may eventually result in structural heterogeneity of the sample. Therefore, to reconstitute a stable homogeneous complex, a catalytically inactive version of Int_{GISul2} was purified, containing a point mutation of an essential arginine residue in the enzyme active site (R241K). The two subunits (α and β) of *E. coli* IHF were produced by co-expression and purified from *E. coli*). For the transposon end DNA design, the *GISul2* element copy from the genome of *Achromobacter xylosoxidans* was used. The 110-bp RE sequence, containing predicted arm sites, core repeats and IHF binding sites, was ordered as two complementary single stranded oligonucleotides, which were later annealed to form the complete double stranded RE DNA.

To generate the complex, RE DNA was mixed first with purified IHF (1:1.5 DNA:protein ratio), and then with purified integrase (in 1:3 DNA: protein ratio). The components were mixed in a buffer containing 500 mM NaCl and dialyzed to low salt containing buffer (50 mM NaCl) to enable gradual formation of a stable complex and prevent protein aggregation. The final complex was analyzed by a size exclusion chromatography (SEC) to verify complex formation and assess the sample homogeneity. SEC showed a single peak, indicating that a homogenous nucleoprotein complex was formed (Figure 2-2). The gradual decrease of salt concentration was important for the complex formation because preparation of the complex directly in low salt containing buffer without dialysis lead to pronounced heterogeneity of the complex.

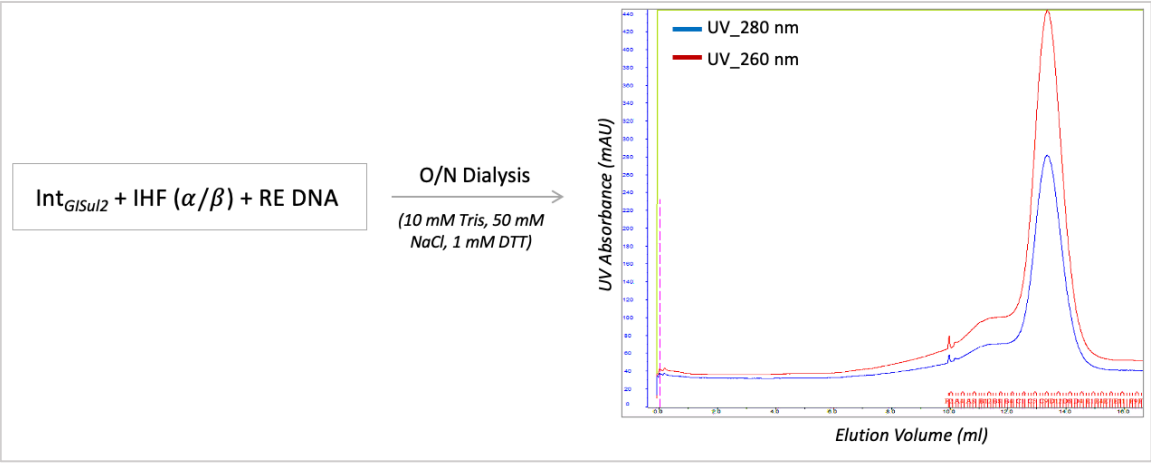


Figure 2-2: In vitro reconstitution of the RE complex. The complex was prepared by mixing catalytically inactive Int_{GISul2} (R241K), IHF and the 110bp RE DNA in 500 mM NaCl containing buffer and then dialyzed into 50 mM NaCl containing buffer to allow protein-DNA complex formation. The dialyzed complex was loaded

on Superose 6 10/300 SEC column. The resulting chromatogram is shown on the left, revealing one major peak eluting at 13.7 ml.

2.1.2 Cryo-EM Grid Preparation and Data Collection

The final complex was next used for cryo-EM grid preparation. Although the integrase dimer and IHF are too small to produce high contrast images by cryo-EM, the presence of DNA greatly increases the contrast, so that the particles can be identified in vitreous ice (Figure 2-3 A). Different concentrations of the complex, addition of various detergents, diverse plunging parameters and a number of grid types were screened to achieve the best ice and particle distribution on the cryo-EM grids. The best particle distribution and ice formation was achieved with UltrAuFoil grids with R2/2 hole size and 200 mesh size (Quantifoil GmbH) without using any detergent. An initial test dataset was collected with the optimized grid in Talos Arctica (200keV, Falcon III detector) with total electron dose of 30 e/Å^2 at pixel size of 1.21 Å/pixel.

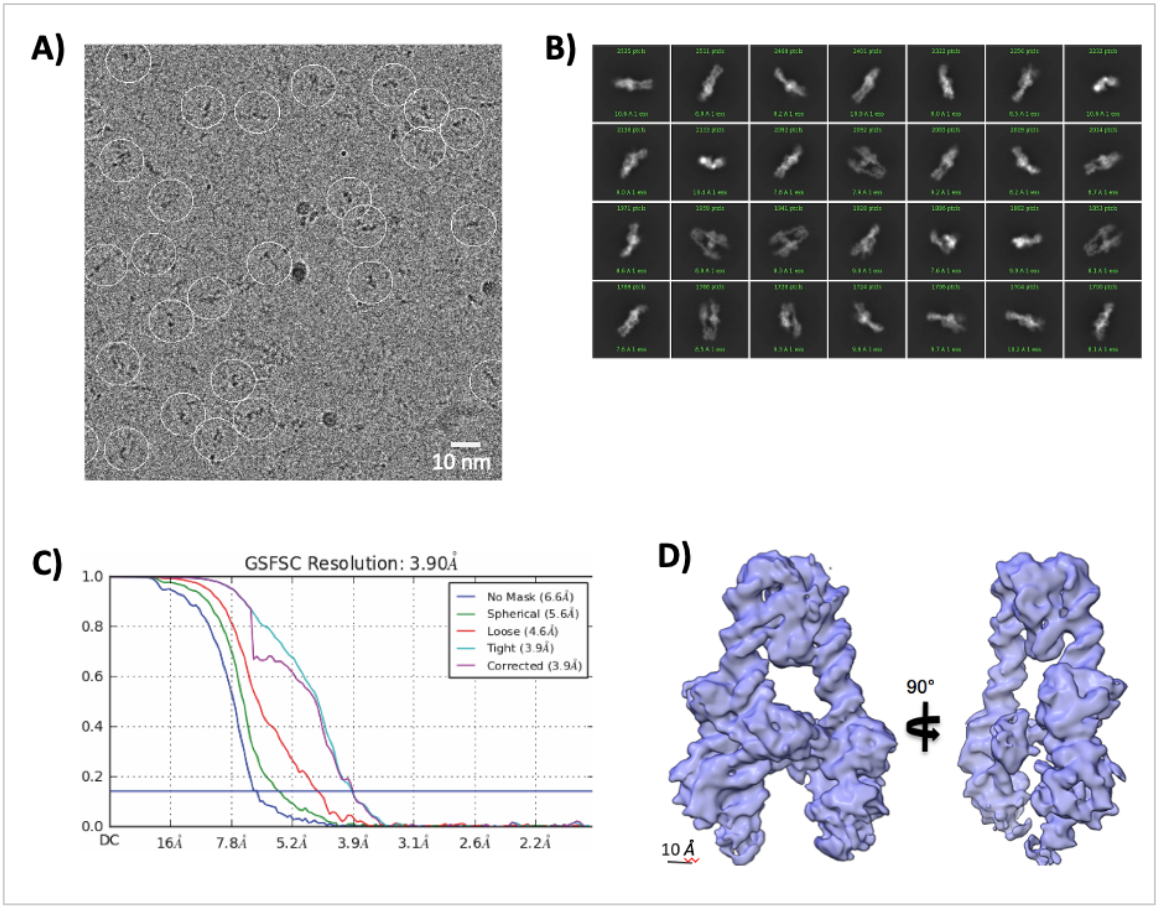


Figure 2-3: Cryo-EM analysis of the RE complex. A) A representative cryo-EM micrograph after motion correction. The micrograph was acquired with the pixel size of 0.81 Å; picked particles are shown in white circles. B) Representative 2D-classes of the RE complex. C) Fourier shell correlation (FSC) curve showing 3.9 Å resolution calculated based on the 0.143 gold standard FSC cut-off. D) The cryo-EM map of the RE complex.

This test dataset was processed with Cryosparc2 software (Punjani et al., 2017) and resulted in a map reconstruction at 5.9 Å resolution. To achieve higher resolution with the same grid preparation, a full dataset was collected in Titan Krios (300 keV, Gatan K2 detector) with total dose of 44 e/Å^2 at a pixel size of 0.81 Å/pixel. The collected 15 000 movies were processed with Cryosparc2 software (Punjani et al., 2017). As a first step movies were patch-motion corrected to align the frames without any shift. Later, CTF was estimated with patch-CTF estimation. The micrographs were manually curated based on relative ice thickness and CTF fit resolution to eliminate micrographs which have thick ice. Around 1000 particles were manually picked from 20 micrographs and classified to create initial 2D classes. In the next step, particles were picked using templates produced from some of the 2D classes generated in the previous step. The resulting particles were extracted and classified, which yielded 2D classes with high-resolution features (Figure 2-3 B). These classes corresponded to various views of the complex, without any preferred orientations. The junk particles were carefully removed with some more rounds of 2D and 3D classifications. With the final particle set, containing 268 000 particles, the homogenous refinement is performed, which gave a reconstruction at 3.9 Å overall resolution (Figure 2-3 C, D).

2.1.3 The Cryo-EM Structure of the RE Complex

The final cryo-EM map has revealed strong density for sharply bent DNA, the ends of which are connected with a protein density. To build a structural model of the complex, the crystal structure of the full-length Int$_{GISul2}$ in complex with core and arm DNA and the crystal structure of the *E. coli* IHF-DNA complex were used (PDB ID 1OUZ). The crystal structures were initially fit into our EM map using rigid body fitting. This showed that IHF fits very well into the protein density at the tip of the bent DNA shape. For Int$_{GISul2}$ and its bound DNA, it was clear that there were significant differences compared to the crystal structure. Therefore, to achieve a better fit, Int$_{GISul2}$ and its bound DNA fragments were divided into three segments, each containing one Int domain and its corresponding DNA part and fit them separately into the EM map.

The resulting model revealed that the RE DNA is bent by IHF to ~160°, bringing the arm and core DNA sites in close proximity. This arrangement allows Int$_{GISul2}$ to bind and bridge both DNA sites together (Figure 2-4). Int$_{GISul2}$ has similar structural configuration as in the full-length crystal structure: the CB of one subunit is dislocated from the core DNA site, which is required to simultaneously accommodate two directly orientated arm DNA sites and the inverted core repeats in the core DNA site by the same integrase dimer.

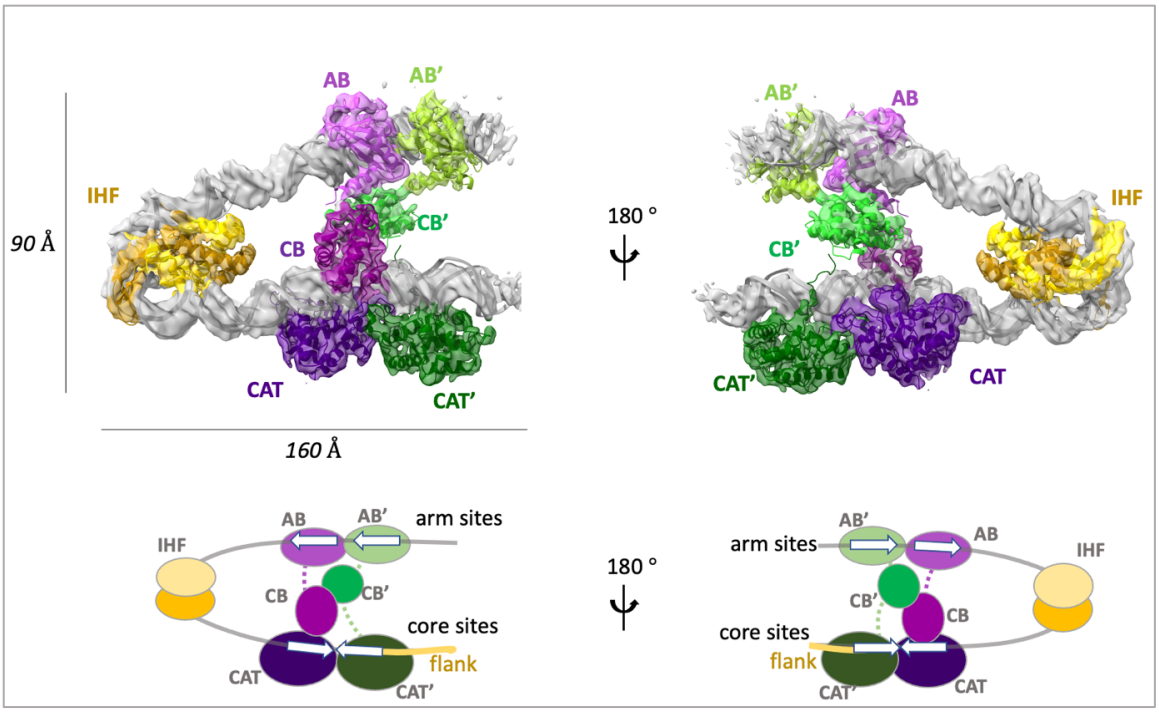


Figure 2-4: Cryo-EM structure of the *GISul2* RE complex. The cryo-EM map of the complex (transparent surface) with the crystal structures of IHF and Int_{GISul2} fitted (cartoon) are shown from two orientations in the top panel. Schematic representation of the RE complex is shown in the bottom panel. The map shows density corresponding to DNA (in grey), which is bent by IHF (in yellow) and bridged by an Int_{GISul2} dimer (purple and green tones). The Int_{GISul2} subunit (purple) located on the core repeat inside the transposon end DNA (purple) has CAT and CB stably bound to its respective binding site. The second subunit (green) binds at the core repeat in the flanking genomic DNA only via CAT and its CB (CB') domain is dislocated from the DNA.

Although the overall domain organization is similar, the AB and CB domains are positioned somewhat differently relative to each other, when compared to the crystal structure (Figure 2-5). The biggest difference is in the localization of the arm DNA, and consequently the AB domain, which are rotated by ~15° to better connect with IHF-bound DNA in the EM map (Figure 2-5 B). This is probably due to the fact that in the crystal structure arm and core sites were supplied as separate DNA molecules, whereas in the EM structure these sites are connected via IHF-bound DNA. The physical constraints present in the native complex with full RE DNA likely enforce observed positioning of the arm DNA site and the AB domain in the EM structure.

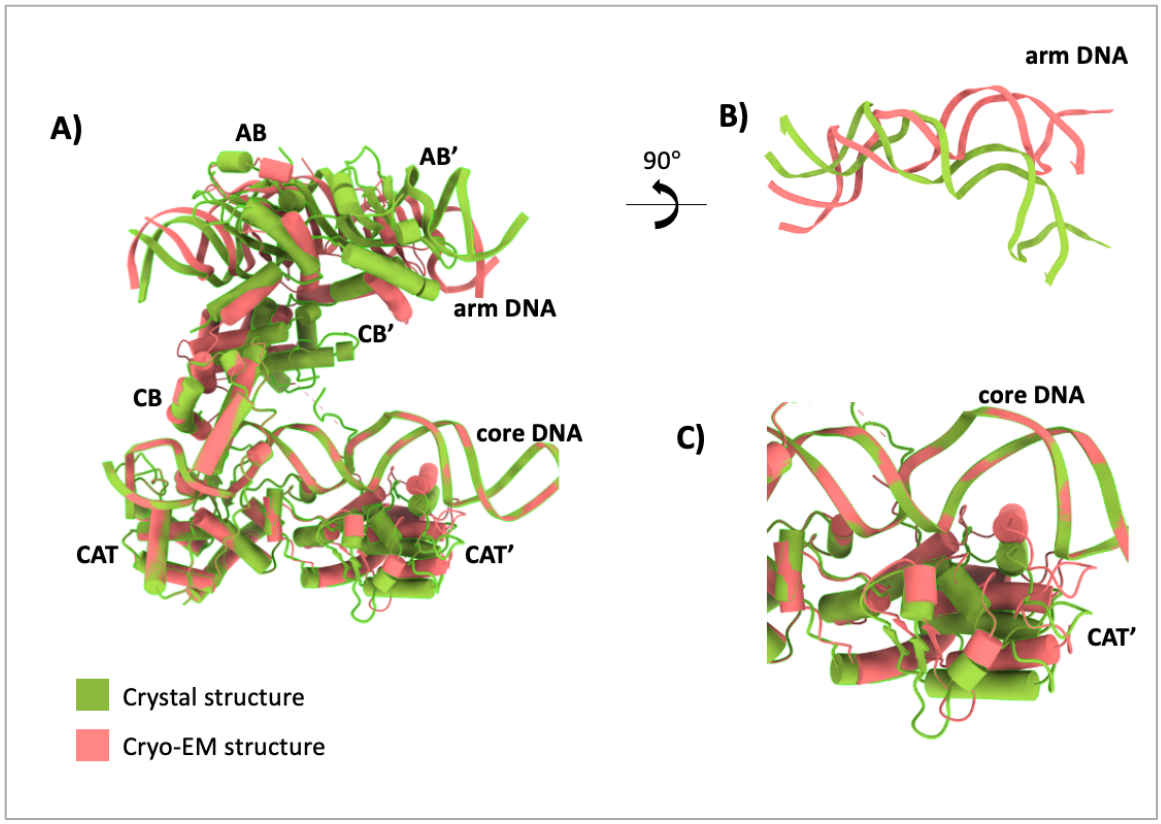


Figure 2-5: Comparison of the crystal (green) and cryo-EM (salmon) structures of the Int_{GISul2} -RE complex. The two structures were overlayed by using Chimera-Matchmaker. A) While overall domain arrangement is preserved, the arm DNA is rotated by ~15° between the two structures. B) The change in arm DNA orientation is depicted from the top view. C) The CAT domain that is bound to the core repeat at the flank interacts more loosely in the cryo-EM structure than in the crystal structure.

Another difference concerns the CAT domain (Figure 2-5 C). In the crystal structure, symmetrized core DNA, which contains two identical core repeats derived from the transposon end, was used to promote complex homogeneity and crystallization. In the EM structure, the native transposon sequence was used, where the core DNA site includes part of the conserved transposon end and the flanking genomic DNA region. While the transposon end contains a strong Int_{GISul2} binding site, the flanking DNA sequence is variable (depending on the actual target genome) and usually features a less ideal core repeat for Int_{GISul2} binding. Consistently, the CAT domain that binds to the core repeat in the transposon end looks similar in the two structures, whereas the other CAT seems to interact less tightly with the DNA in the cryo-EM structure. These differences are likely due to the sequence differences in the core repeats of the native RE substrate. Alternatively, the different localization of arm and core sites in these structures may contribute to moving the CAT away from its DNA.

2.1.4 Flexibility of the RE Complex and Focused Refinements

To determine the flexibility of the complex, 3D variability analysis was performed which is found in Cryosprac2 (Punjani et al., 2017). In this analysis, instead of classifying particles into distinct states, they are used to produce continuous series of 3D structures, which captures the distinct conformations that are present in the data. With that, two different significant movements in the RE complex were captured. The first one involves a rotation of the arm site DNA. During this movement, the arm DNA swings out of the main plane of the complex and the dislocated CB domain moves closer to the core repeat in the flank DNA (Figure 2-6, Component 1). The second movement involves the two DNA pieces bridged by the Int_{GISul2} dimer, which approach and relax with respect to each other (Figure 2-6, Component 2). The analysis has shown that there are particles with different opening angle of the two DNA "legs", suggesting that these can come closer together and move further apart in solution.

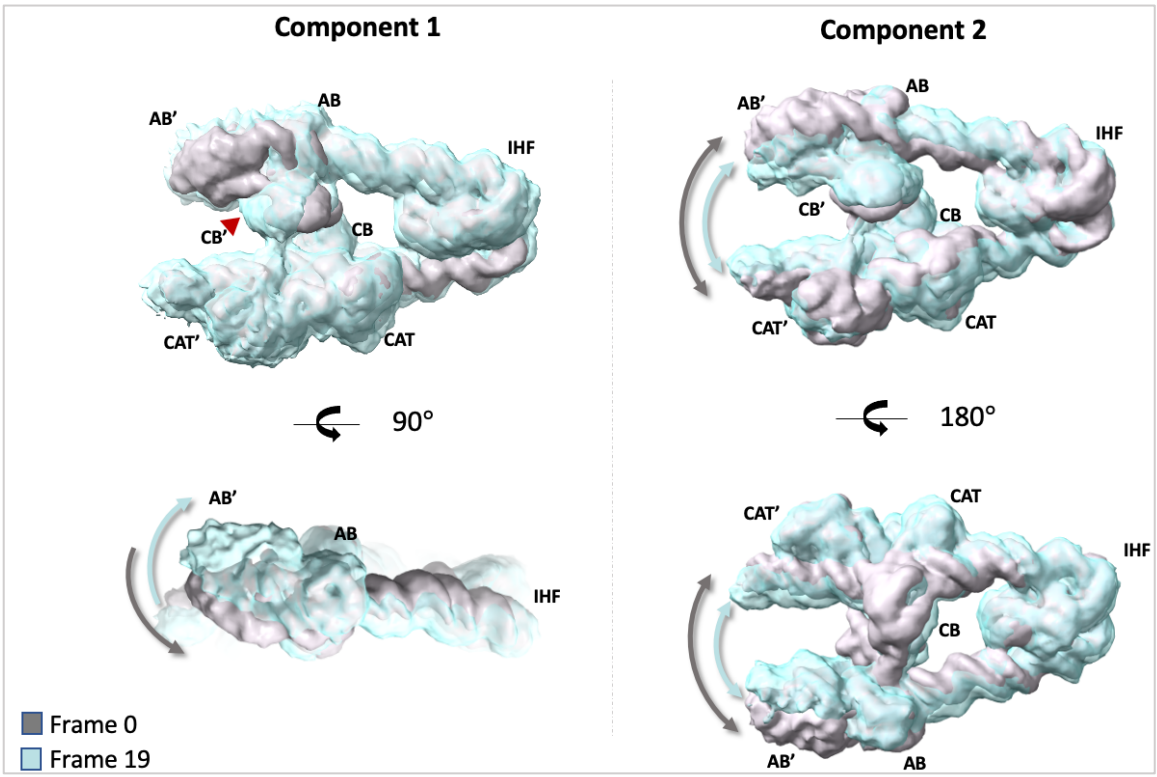


Figure 2-6: The flexibility of the RE Complex. The 3D-variability analysis (done with Cryosparc2) showed two types of movement in the RE complex, named as component 1 and 2. The two terminal conformations on the movement trajectory are shown in grey and blue, respectively. Component 1 reveals a rotation of the arm site DNA perpendicular to the plane of the complex. Component 2 shows closing of the two DNA ends, with the arm and site DNA moving closer to each other.

The observed inherent flexibility of domains in the complex limits the final resolution of the EM map, because their continuous movement hampers proper alignment of the particles and results in a blurred average. To achieve higher resolution, one needs to separately align the immobile segments of the complex to construct a map. For that reason, local refinements were performed by focusing on a specific stable region, while masking out the rest of the signal. This is performed by constructing a mask for the region of interest and subtracting the signal coming from outside of the mask. This way, the signals coming from the specified region can be aligned in local refinement.

Based on the 3D variability analysis, IHF with its bound DNA, AB with the arm DNA and CAT-CB with the core DNA were assigned as three rigid bodies that move with respect to each other and should be aligned separately. First, individual masks for CAT and AB segments were constructed to perform local refinement for these regions. Even though it is recommended to use a minimum size of 200 kDa for local refinement, we proceed with our rather small masks with the hope that the presence of DNA would facilitate the alignment and lead to an improvement in map resolution. Since DNA has a stronger signal, the particles are almost always aligned based on DNA, which may promote domain alignment as well, provided they are stably positioned on DNA. Indeed, our focused alignment with a mask including the two CAT domains the core site DNA and the stably bound CB domain resulted in a map at 3.47 Å resolution (Figure 2-7 B).

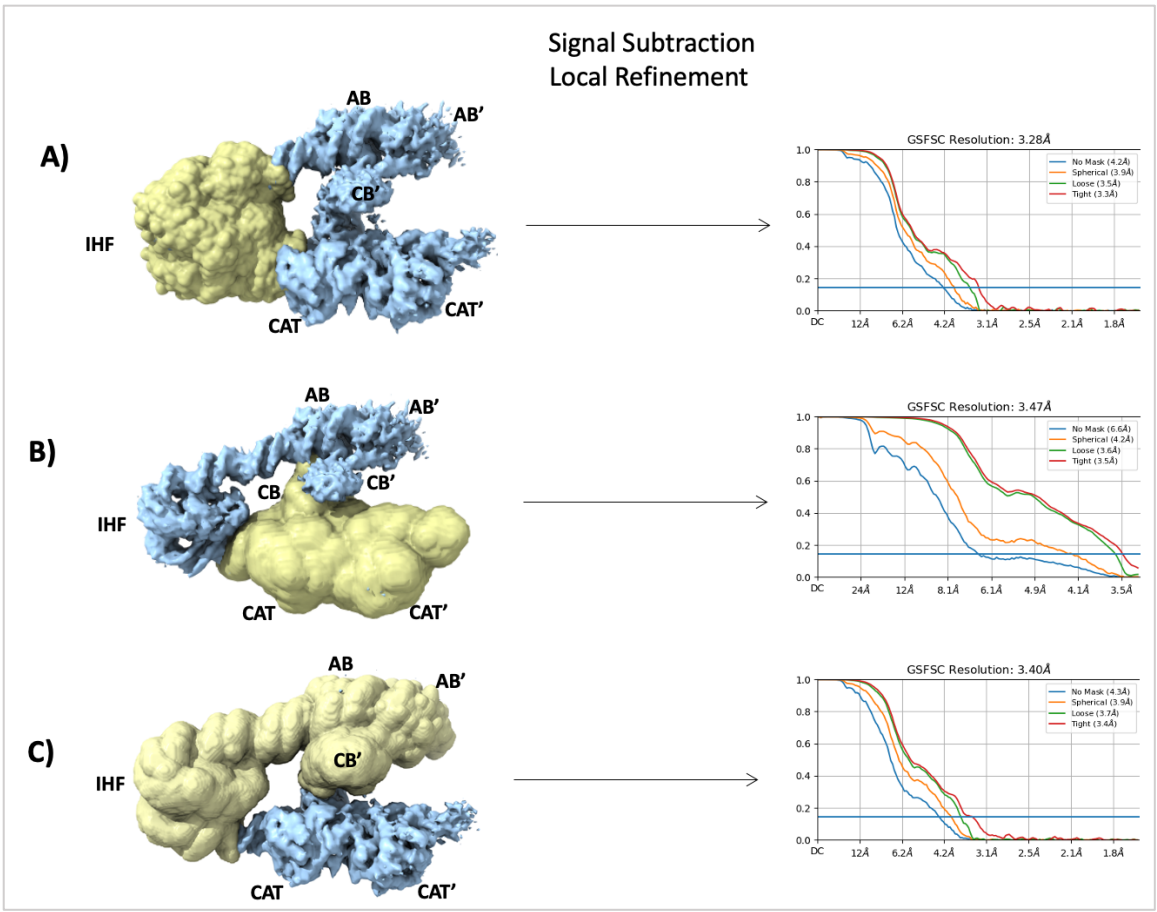


Figure 2-7: Local Refinement of the RE Complex. The masks designed for local refinement are shown in yellow on the left. These covered A) IHF with its bound DNA, B) the core site DNA with the bound CB and CAT domains, and C) arm DNA with the AB domains, the dislocated CB' and the IHF-DNA segment. FSC curves on the right show the improved resolution after signal subtraction and local refinement.

For the AB-DNA segment, the designed compact mask was too small to allow successful alignment once the rest of the volume was subtracted. Therefore, the IHF-DNA region was further included with AB and arm sites, which did not seem to move with respect to each other. Local refinement with this enlarged mask then gave a map at 3.4 Å resolution (Figure 2-7 C). Similarly, a separate mask for the IHF-DNA segment was constructed to perform local refinement, which resulted in a map at 3.28 Å resolution (Figure 2-7 A).

2.1.5 The U-Turn Induced by IHF

In the RE complex, DNA is bent by IHF by ~160°, allowing Int_{GISul2} to connect arm and core DNA sites. Its presence is essential for the correct formation of the RE complex. The local refinement provided a 3.4 Å-resolution map of IHF region (Figure 2-8), in which some of

the amino acid side chains can be differentiated. The previously solved IHF crystal structure was used for the initial fit to the EM map, which was then further optimized with real space refinement in Phenix. While the IHF backbone was well placed using rigid body movement, fitting of the DNA density required additional manual refinement, mainly due to a nick in the DNA used in the crystal structure.

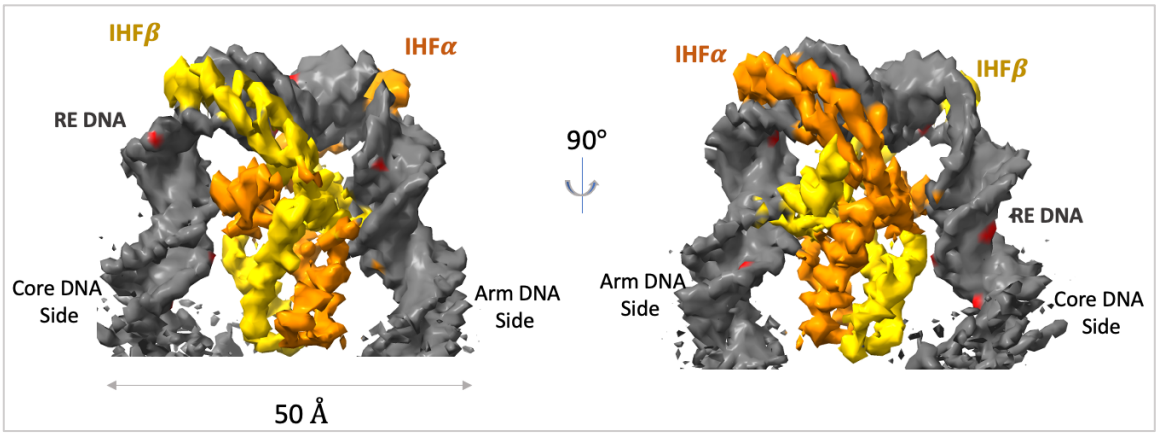


Figure 2-8: The cryo-EM map of IHF and its bound DNA. The sharply bent DNA is shown in grey. IHFα (orange) and IHFβ (yellow) are intertwined, forming the IHF body between the almost parallel DNA stems. The two extended arms of the protein wrap DNA along the minor grooves.

In the EM map, the long β-ribbon arms of IHF wrap DNA from the minor grooves and the body of it interacts with the bent DNA as it was described in the crystal structure (Figure 2-9) (Rice et al., 1996). It achieves this dramatic DNA bend with two sharp kinks, which are located 9-bp apart. These sharp turns are executed by proline residues (P65 in IHFα, P64 in IHFβ) at the tips of the β-ribbon arms of IHF through intercalation between DNA bases (Figure 2-9). Proline residues are typically used for DNA intercalation in IHF/HU family proteins, due to their favorable wide shape which allows hydrophobic contact with DNA bases.

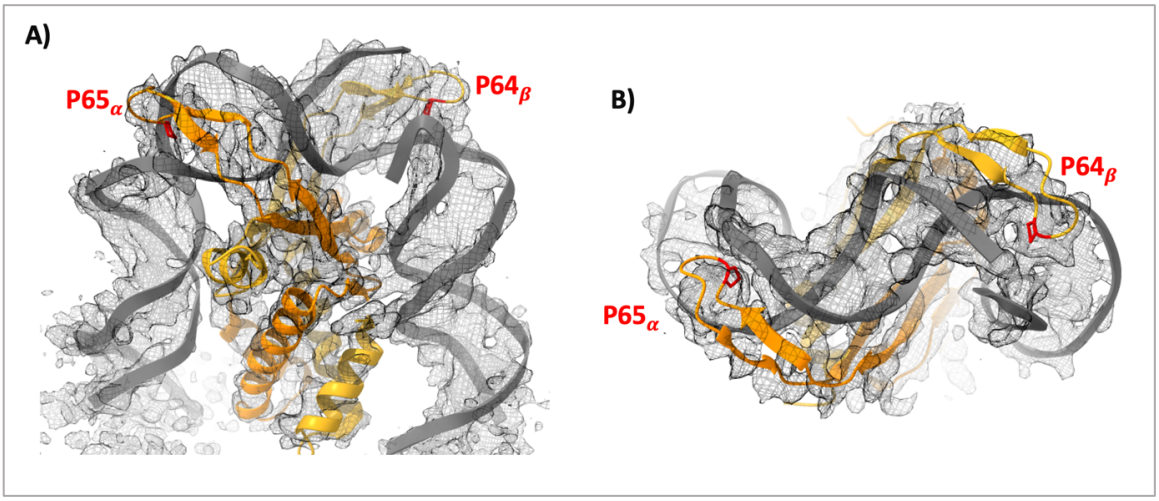
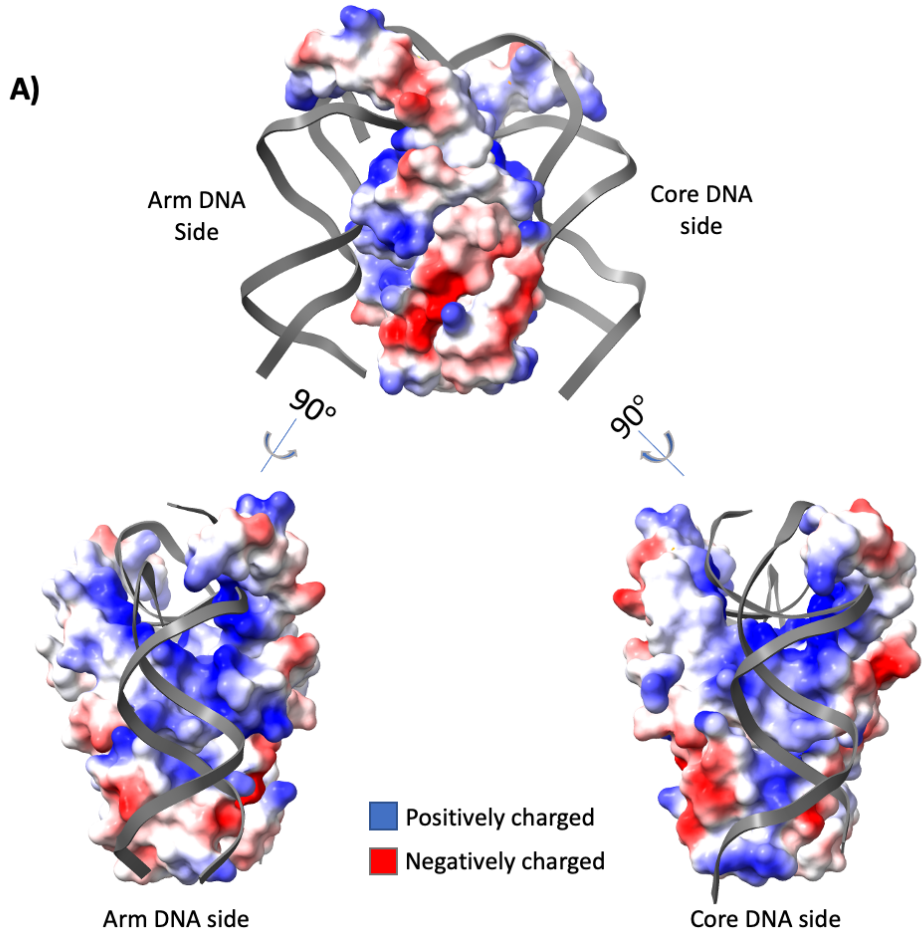


Figure 2-9: Conserved proline residues in IHF intercalate between DNA bases. Prolines found at the tip of βribbon arms of both IHF (yellow and orange) subunits are shown in red (P65α and P64β). While A) shows the side view, B) shows the top view. The electron density map is shown as a grey mesh.

While the kinks are executed by the arms, the body of the IHF heterodimer stabilizes the two stems of the U-shaped DNA. This is confirmed by the IHF electrostatic charge distribution plot, which shows that its body is mostly positively charged at the surfaces where it interacts with the DNA backbone (Figure 2-10 A). In addition to overall positive charge distribution, the IHF body forms clamps to hold DNA more tightly (Rice et al., 1996). In the EM structure, on the core DNA side α helix 1 from IHFβ (α1 on Figure 2-10 C) and helix 3 from IHFα (α3 on Figure 2-10 C) form a clamp around the DNA stem that goes towards the core site, with S47α positioned between the clamp jaws in the DNA minor groove (Figure 2-10 C). This interaction is formed with the conserved A-tract, which was shown to have distinct structural features, like a narrow minor groove (DiGabriele and Steitz, 1993), which enables the DNA to fit in the protein clamp (Rice et al., 1996). On the DNA stem towards the arm site, the clamp is formed by α-helix 1 from IHFα and helix 3 from IHFβ (Figure 2-10 B), with R46β placed in the minor groove in the middle (Figure 2-10 B). While on the core DNA side S47α doesn't contact the bases, on the arm DNA side R46β makes specific base interactions.



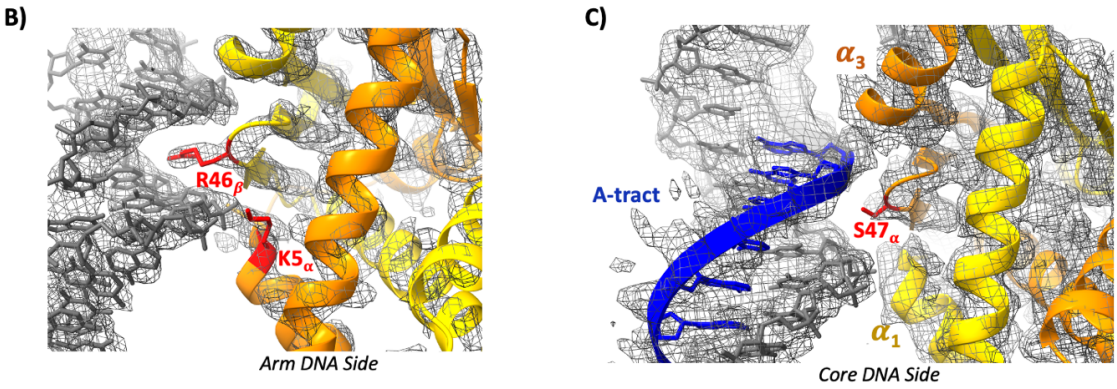


Figure 2-10: The body of IHF stabilizes the bent DNA. A) The electrostatic surface distribution of IHF is shown, with positively charged residues in blue and negatively charged residues in red. DNA is shown in grey. The right and left sides of the IHF body are positively charged and interact with the bent DNA backbone. B) On the arm DNA side of the IHF binding sequence, α3 (IHFβ) and α1 (IHFα) form a clamp around the DNA, with R46 inserts in the minor groove between the two claws of the clamp. C) On the core DNA side of IHF binding sequence, the clamp is less pronounced and is formed by IHF by α3 (IHFα) and α1 (IHFβ), with S47 inserting in the minor groove in the middle. The electron density map is shown as a grey mesh in panels B and C.

IHF's DNA interaction heavily relies on indirect sequence readout to recognize its binding site. It mostly interacts with the phosphodiester backbone of the DNA and makes very few base specific interactions. However, there are some conserved nucleotides in the regions where extensive protein-DNA contacts are formed. The IHF binding sequences found in the *GISul2* element and in λ genome were aligned (Figure 2-11). The A-tract is mostly conserved between the two sequences, only in *GISul2* it is interrupted by a single C base (Figure 2-11 A). The conservation of this sequence carries high importance, since it is involved in clamp recognition and its absence was shown to impair IHF's DNA binding (Hales et al., 1994). In addition to the A-tract, there are additional conserved sequence elements. On the arm DNA side, the sequences before and after the DNA kink are identical in the two IHF binding sites, and part of the DNA sequence between the two DNA kinks is also conserved. Specific nucleotides between the kinks make hydrogen bonds with R60 and R63 from the IHFα arm, likely promoting recognition of these sequences (Figure 2-11 C). When IHF binding at the two DNA stems is compared, the arm side DNA, which contains more conserved sequences, seems to interact more tightly than the core side DNA which contains the A-tract.

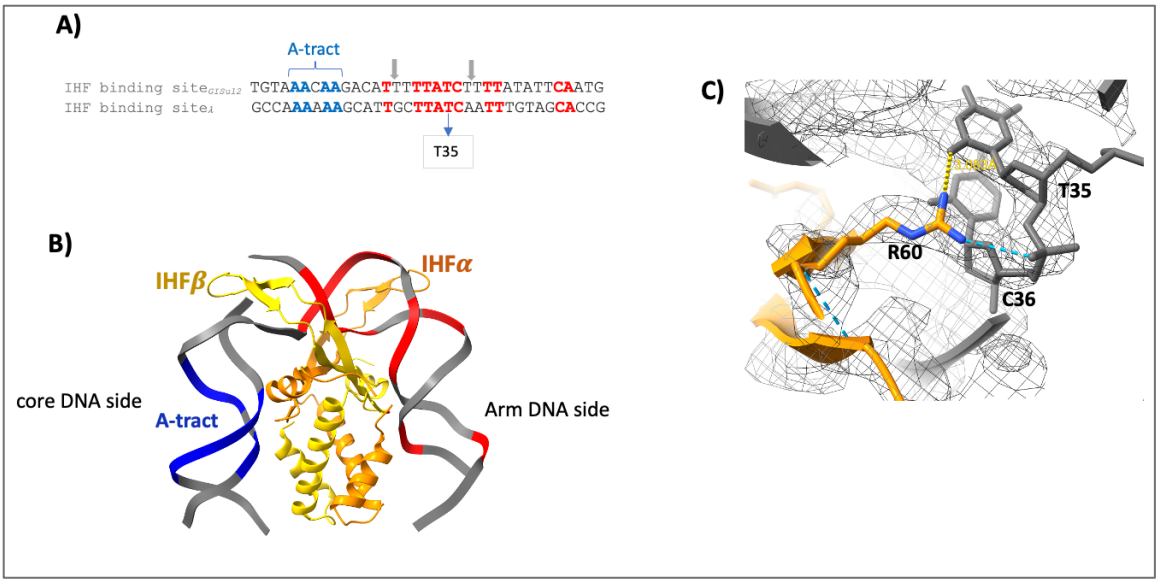


Figure 2-11: Comparison of the IHF binding sites of *GISul2* and the λ phage. A) Sequence alignment of the binding sites. The A-tract is shown in blue and other conserved sequences are shown in red. Arrows indicate where the proline residues intercalate the DNA. B) The conserved nucleotides are marked on the IHF-DNA structure. The color scheme is the same as in A). C) A zoomed view of the interaction between R60 (IHFα) and T35 from the IHF binding site.

2.1.6 CAT and CB of Int$_{GISul2}$

In the RE complex, a Int$_{GISul2}$ dimer bridges the RE arm and core sites. As mentioned above, due to the directionality of arm and core repeats, one of the CB domains is detached from the DNA. To achieve a higher local resolution, a mask which excludes the dislocated CB and includes two CAT domains and one CB together with the core DNA site was designed (Figure 2-7 B). With the local refinement, a 3.4 Å resolution map was obtained for this region of the structure. The previously solved crystal structure was fitted and refined with real space refinement in Phenix.

In the EM structure, the CAT domain bound to the core repeat in the flank DNA (Figure 2-12, light green, CAT') interacts with DNA less extensively, when compared to the CAT and CB domains of the other Int$_{GISul2}$ subunit located on the transposon end core repeat (dark green, CAT), which form a closed C- shaped clamp around the DNA. The tight DNA binding of this subunit is further facilitated by the loop between CB and CAT, which interacts with the DNA through residues R203 and F200 (Figure 2-12 A,D). This connecting loop is not visible in the loosely bound Int$_{GISul2}$ subunit. The active site of the CATs is assembled *in trans* and it is composed of H338, R241, R341, H364, Q359 from one subunit and nucleophile Y373, which is donated by the neighboring CAT (Figure 2-12 B). The resolution of the map allowed us to locate the side chains of all the catalytic residues except R241, which was

mutated to alanine to create an inactive Int_{GISul2} variant for this complex formation (Figure2-12 B).

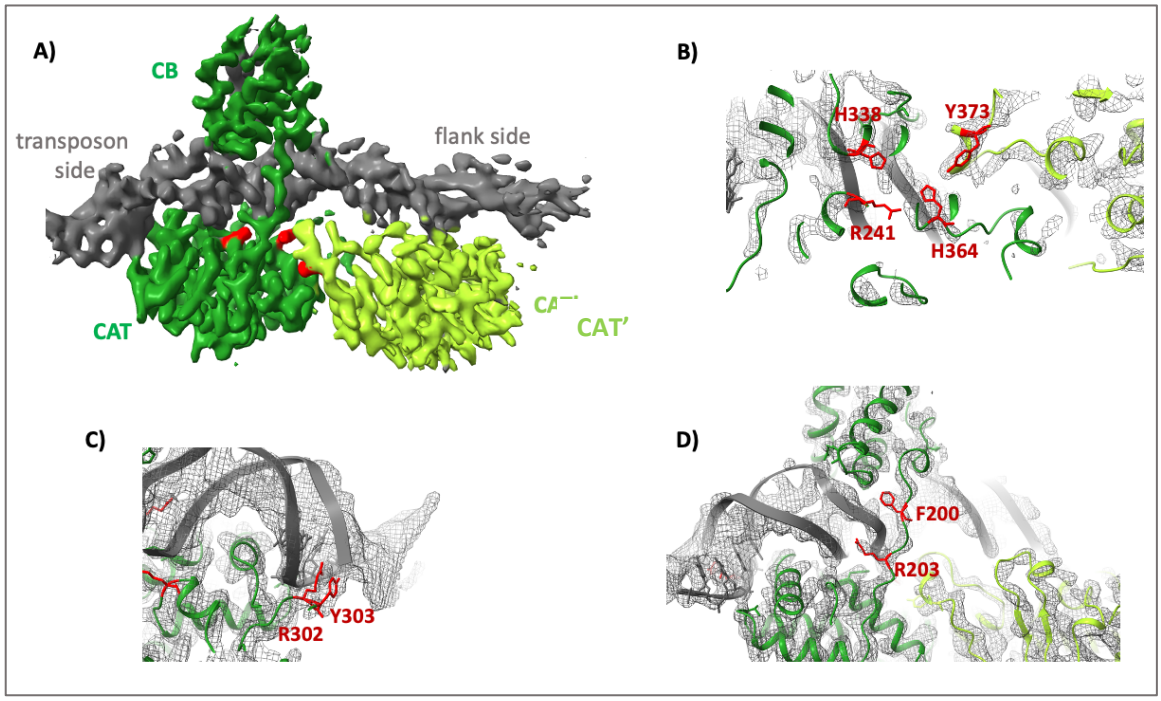


Figure 2-12: Refined structure of the CAT-CB-core DNA segments. A) The EM map obtained after local refinement is shown (DNA is in grey, Int_{GISul2} subunit stably bound to transposon DNA is in dark green, the CAT positioned on the flank DNA is in light green). B) Zoom up of the Int_{GISul2} active site. The residues forming the active site are shown in red sticks and the EM map is depicted as grey mesh. The active site is composed of H338, R241, R341, H364, Q359 and the nucleophile Y373, donated by the neighboring CAT. C) The interaction of R302 and Y303 residues with DNA bases stabilizes the DNA binding. D) The residues F200 and R203 found in the loop between CAT and CB provide extensive contact with DNA.

The CAT domain located at the core DNA repeat in the transposon end interacts more extensively with DNA when compared to the CAT' interaction with the flank DNA. Superimposition of the two CATs and their bound core DNA repeats, revealed essentially no change in the 3D organization of the domains, including virtually identical positioning of the Y373 nucleophile on the M-helix (Figure 2-13). Notably, the core repeat in the transposon end is 40° bent, while the repeat at the flank region DNA is straight (Figure 2-13 A). Since the two CAT domains have the same structure, this differential bending may be explained by CB binding. Additionally, the AT-rich sequence of the transposon end core repeat can facilitate DNA bending. H166 from the CB domain makes hydrogen bonds with the first conserved A-T base pair, in the core DNA site next to the crossover region. Additionally, some hydrophobic amino acids from CB, including V126 and F200, which point towards bases in the major groove, seem to contribute to the observed DNA bending. To further stabilize the bent DNA, R302 and Y303 residues from CAT interact with the conserved 10^{th} and 11^{th} (TT) bp after the crossover region (Figure 2-12 C). When inspecting

the CAT active site, it appears that the 40° bend of the DNA brings the cleavage site of the DNA closer to the catalytic site of CAT when compared to CAT' (Figure 2-13 A).

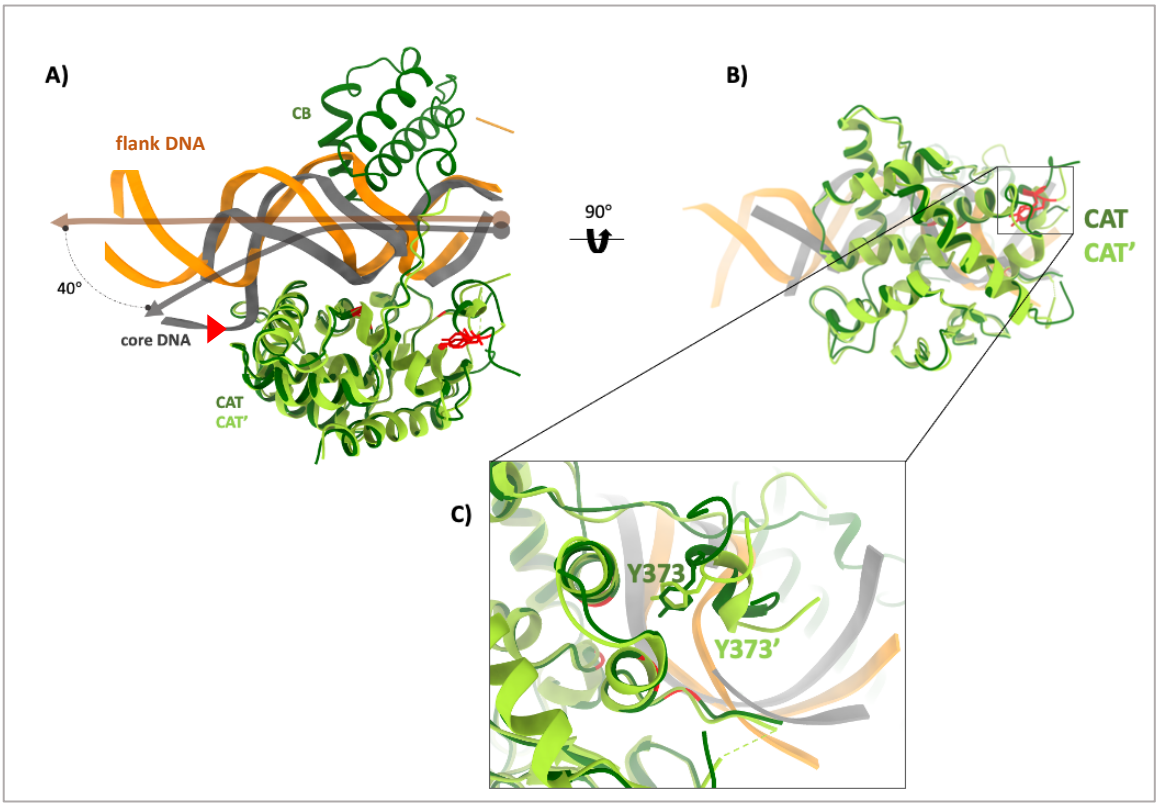


Figure 2-13: The CB binding bends DNA. A) Two CATs and their bound DNA are superimposed (transposon end DNA is in grey, Int subunit bound to transposon core repeat is in dark green, the CAT' positioned on the flank DNA is in light green, flank DNA is in orange). B) The superimposition from A) is shown with a 90° rotation. C) Zoom up of the alignment of catalytic Y373 residues in CAT and CAT'.

2.1.7 AB Bends Arm Sites

In the initial EM volume, the AB domains were the most flexible and mobile part of the structure. Because of its mobility, in the overall refined structure, this was the part with the lowest resolution. Following focused refinement, the resolution was slightly improved. Among the two AB domain, the one that is connected to the stable DNA bound CB domain is more stable and better resolved, suggesting that the fixed CB position also stabilizes AB localization (Figure 2-14 dark green).

There are two arm sites located in the same direction in RE end, which are bound by two AB domains back-to-back in the same direction. AB binding induces a 30° bend in the arm site DNA (Figure 2-15 B). The AB bends the DNA mainly by docking its beta strands into the DNA major groove and widens it (Figure 2-14 B). AB-DNA binding is stabilized by many

connections of several arginine residues with the DNA backbone (mainly R41 and R57). Since we do not see side chains for the AB region in the EM map, this information is largely derived from the fitted crystal structure but agrees with the EM data within its limits.

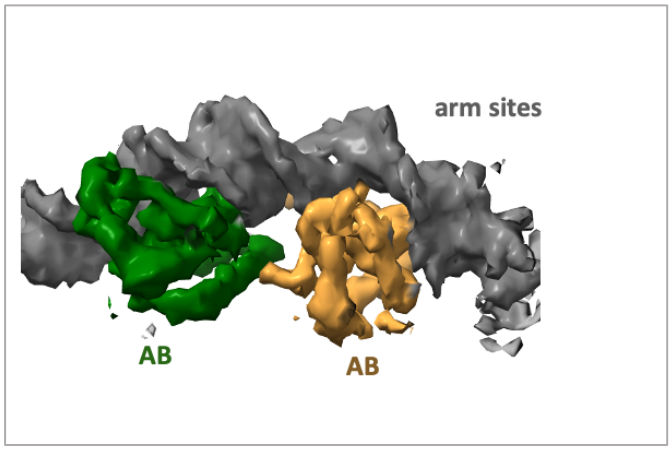


Figure 2-14: DNA binding of the Arm Binding Domain (AB) of Int_{GISul2}. A) The EM map is shown, with arm site DNA in grey. The AB on the left side (dark green) is connected to the CB domain that is stably positioned on the core site, while the other AB (yellow) is connected to the dislocated CB.

ABs are the most diverse domains of TRs (Smyshlyaev et al., 2021). AB of Int_{GISul2} is composed of a four stranded β-sheet between two α-helices. When compared the AB of Int_{λ}, it has an extra α-helix ($\alpha 1$) and a β-strand ($\beta 1$) (Figure 2-15). The additional helix and strand are found in all the SXT family integrases (Smyshlyaev et al., 2021). In contrast to AB of Int_{GISul2}, Int_{λ} AB does not induce a bend in the arm site DNA, which can be due to absence of extra β-strand resulting in a narrower β-sheet (Figure 2-15).

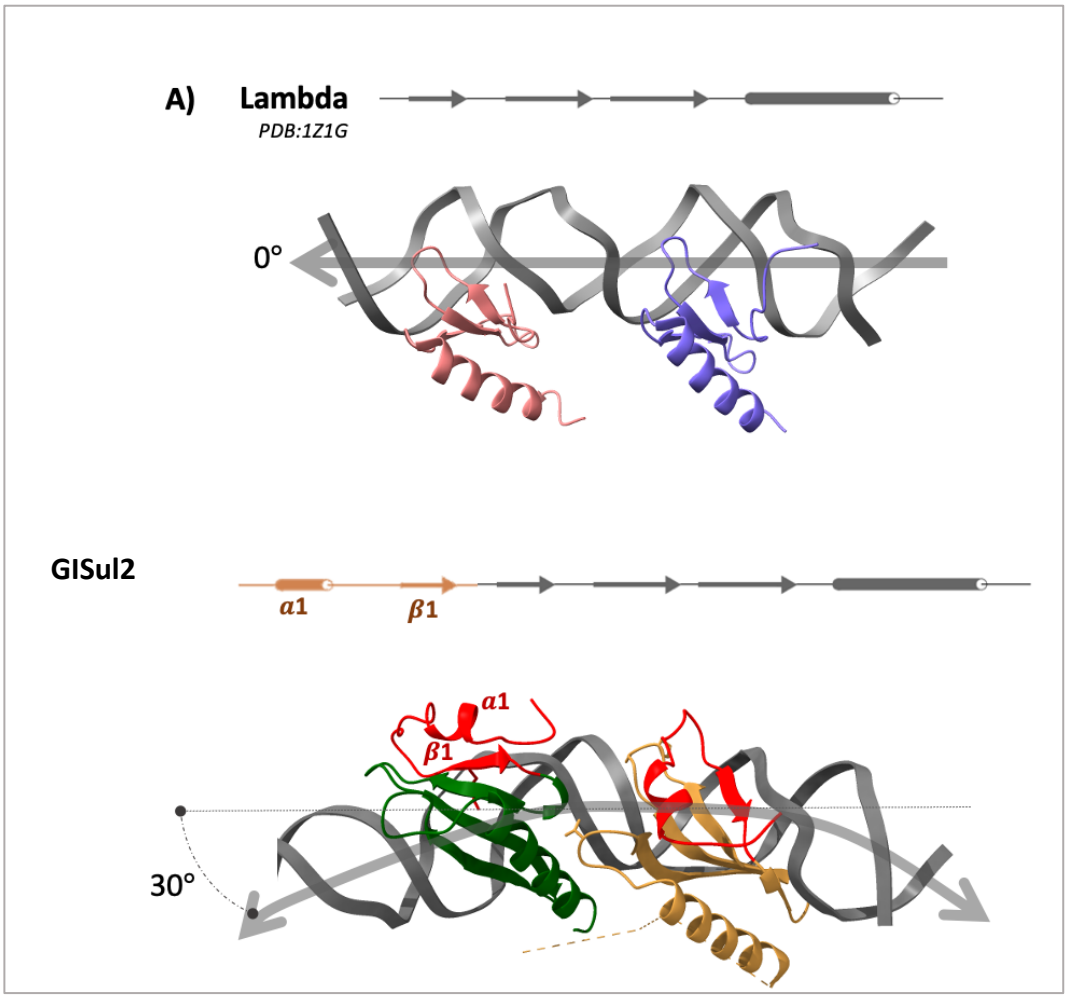


Figure 2-15: Comparison of DNA binding by the Int_{GISul2} and Int_{λ} AB domains. The diagrams show the secondary and the tertiary structures of Int_{λ} (A) and Int_{GISul2} (B) AB. The extra helix (α1) and a β-strand (β1) of Int_{GISul2} are shown in red.

2.1.8 Comparison of the Int_{GISul2} with known TR structures

Out of four model TRs, in terms of domain organization and primary sequence, λ integrase (Int_{λ}) is the closest one to Int_{GISul2} (Smyshlyaev et al., 2021). The comparison of the structures revealed that they share very similar fold in the CB and CAT domains (Figure 2-16). The only difference is in the positioning of α-helix, which carries the nucleophile tyrosine (noted as M-helix), and the following helix (noted as N-helix). Since Int_{GISul2} assembles its active site in *trans*, the M-helix has a long and flexible linker which is important for its mobility when supplying the tyrosine residue to the neighboring monomer; while Int_{λ} assembles its active site *in cis* and doesn't have a flexible linker before the catalytic tyrosine carrying helix (Figure 2-16).

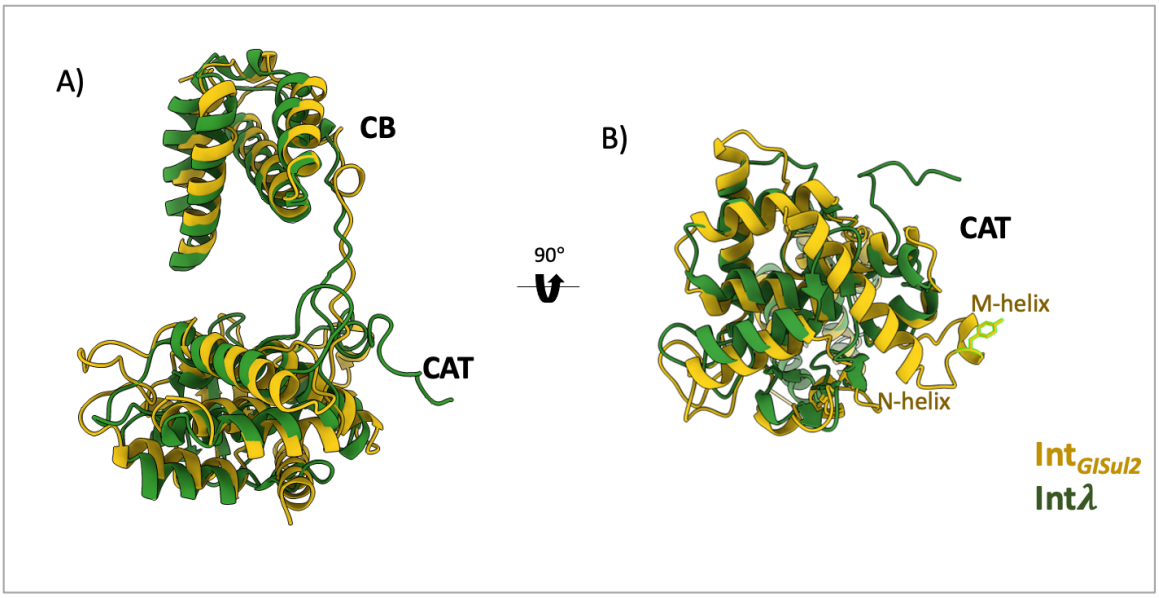


Figure 2-16: Comparison of the CB and CAT domain structures from Int$_{GISul2}$ and Int $_{\lambda}$. A) CB and CAT of Int$_{GISul2}$ (yellow) and Int $_{\lambda}$ (green, PDB ID 1Z1G (Biswas et al., 2005) are superimposed. B) The superimposition from the bottom view (with 90° rotation). The nucleophile tyrosine is shown as sticks.

The next comparison was done with the Cre recombinase. Cre is quite different from other bacterial TRs and phylogenetic analysis has shown that it does not cluster with any known bacterial TR (Smyshlyaev et al., 2021). Thus, the structural similarity between Int$_{GISul2}$ and Cre is less pronounced. Although the CATs share a similar fold, the localization of the C-terminal helices (M and N) is different as stated in Int$_{\lambda}$ comparison. Differences in the structure of the two CB domains are even more substantial, with Int$_{GISul2}$ having a more compact CB when compared to Cre and CB of Cre has an extra α-helix.

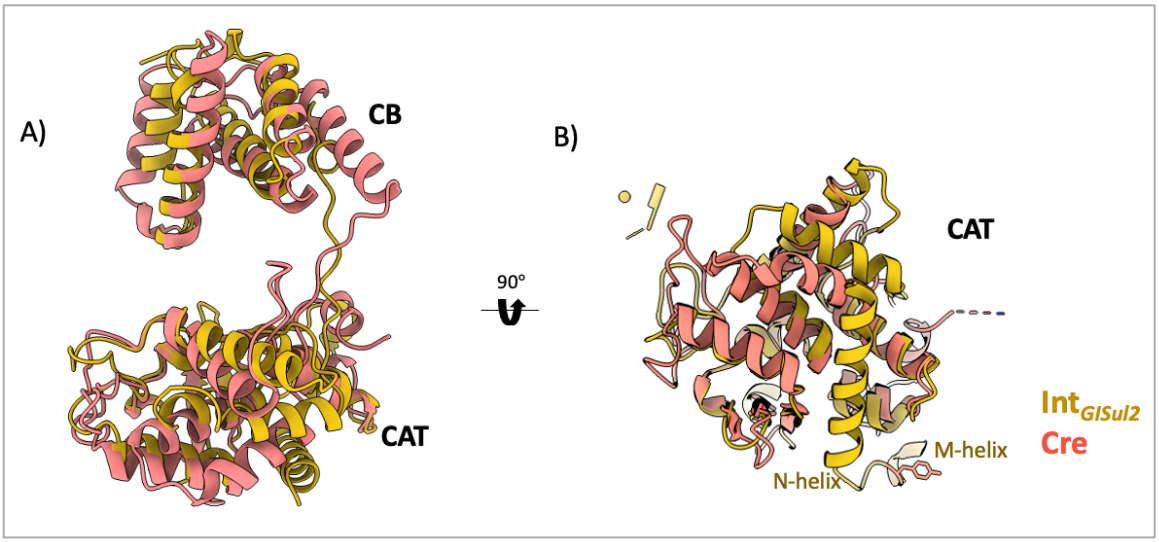


Figure 2-17: Comparison of Int$_{GISul2}$ and Cre structures. A) CB and CAT of Int$_{GISul2}$ (yellow) and Cre (salmon, PDB ID 2CRX (Gopaul et al., 1998) are superimposed. B) The superimposition from the bottom view (with 90° rotation).

Another comparison is performed between Int_{GISul2} and the integrase of Tn*1549* element (Int_{Tn1549}). As in the case of Int_{λ}, these two CTn integrases share highly similar fold in the CB and CAT domains (Figure 2-18). The main difference is in the positioning of the two C-terminal helices of the proteins. First, the helices that carry the nucleophile tyrosine (M in Int_{GISul2}) are placed differently, due to their distinct active site assemblies (*trans* in Int_{GISul2}, *cis* in Int_{Tn1549}). Additionally, the very C-terminal helices are altered. While N-helix swaps back to its own protein subunit and docks on its surface in Int_{GISul2}, the equivalent segment (C-terminal helix) is donated to the neighboring subunit in Int_{Tn1549} (Figure 2-18 B). This helix stabilizes the dimer formation of Int_{Tn1549} and mediates the formation of an auto-inhibitory state, which prevents premature DNA cleavage (Rubio-Cosials et al., 2018).

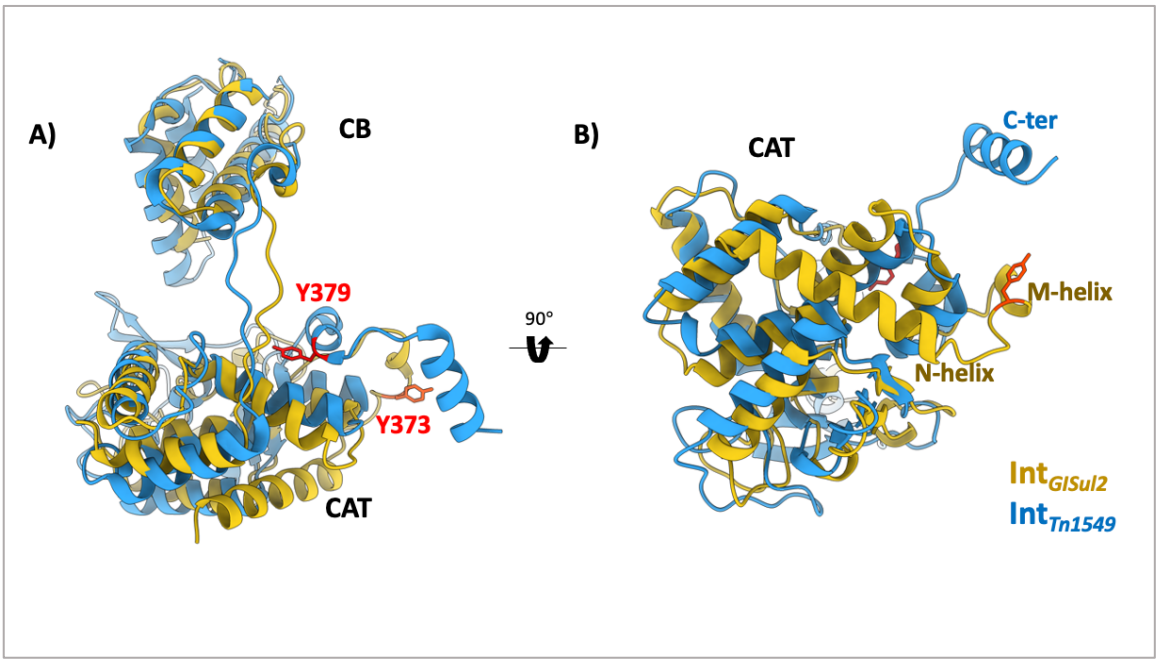


Figure 2-18: Comparison of Int_{GISul2} with Int_{Tn1549}. A) CB and CAT of Int_{GISul2} (yellow) and Int_{Tn1549} (blue, PDB ID 6EN0 (Rubio-Cosials et al., 2018) are superimposed. The catalytic residues are shown in red - Y379 in Int_{Tn1549} and Y373 in Int_{GISul2}. B) The superimposition from the bottom view (with 90° rotation).

The last comparison was done with the Flp recombinase. Since Flp is a eukaryotic TR, the similarity with Int_{GISul2}, was not expected to be high. Nevertheless, the *trans* assembly of the active site is a common feature of both enzymes. Although, the domain composition and localization are matching, their secondary structures are rather dissimilar (Figure 2-19 A). Strikingly, the M-helixes carrying the active tyrosine residues are located quite similarly in both proteins. This helix is connected to the CAT domain with flexible loops in both cases, which favors its movement to the neighboring CAT. In Flp, the M-helix, is stabilized by an interaction with a histidine residue (H309) in the receiving subunit, which is also the case in Int_{GISul2} (Figure 2-19 C).

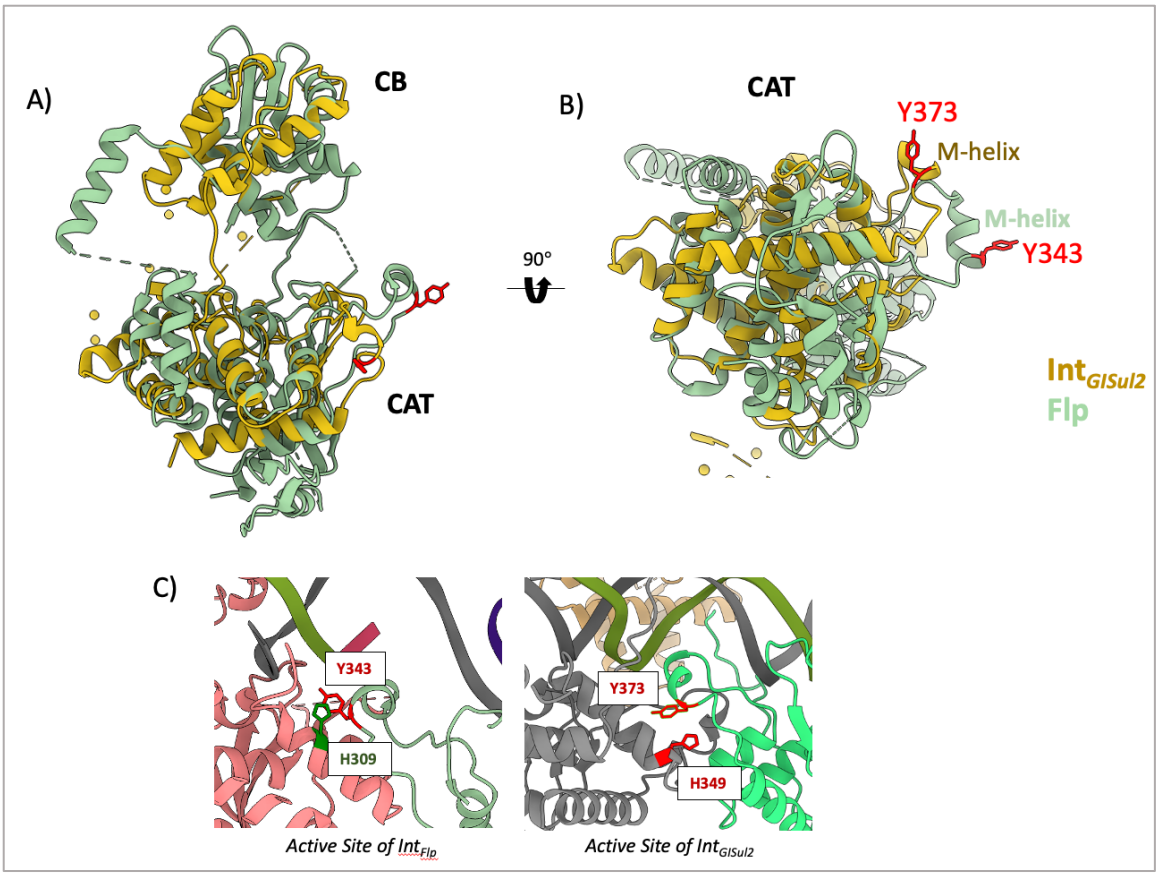


Figure 2-19: The comparison of Int_{GISul2} with Flp recombinase. A) The CB and CAT of Int_{GISul2} (yellow) with Flp (green) PDB ID 1FLO (Chen et al., 2000) are superimposed. B) The superimposition with 90° rotation. C) The catalytic tyrosines are stabilized by histidine residues by stacking.

2.1.9 Comparison of the RE Complex Assembly of *GISul2* with the λ System

In the previous sections, Int_{λ} was introduced as a model system for the higher order complex formation to regulate recombination. Although, CTns are known to use phage like recombination machinery, the mechanistic similarities and the differences between two systems were not known. Therefore, once the structure of the RE complex was obtained, it was compared with the Int_{λ} system.

The left end of the λ phage genome (attL) was shown to require only IHF for folding and Int_{λ} binding, while its right end (attR) requires Xis in addition to IHF. Therefore, I compared the *GISul2* RE complex structure to the λ attR assembly. Since there is no structure for the separate attL- Int_{λ} complex in the λ system, I used the excision assembly structure, which contains both attL and attR and was solved at 11 Å resolution. Overlay of the EM volumes shows that the IHF segment is essentially the same, including the density of the protein and its bound DNA. This is consistent with the fact that the authors also used *E. coli* IHF

protein in the λ complex. The λ excision assembly structure was solved at rather low resolution, for that reason a detailed comparison cannot be performed. However, one can already see that the overall localization of the integrase subunits has some similarities, but the localization of AB and CAT are quite distinct. In particular, arm sites are located much further away from the IHF binding site and the CAT domains seem to interact more tightly and occupy less DNA in the *GISul2* structure.

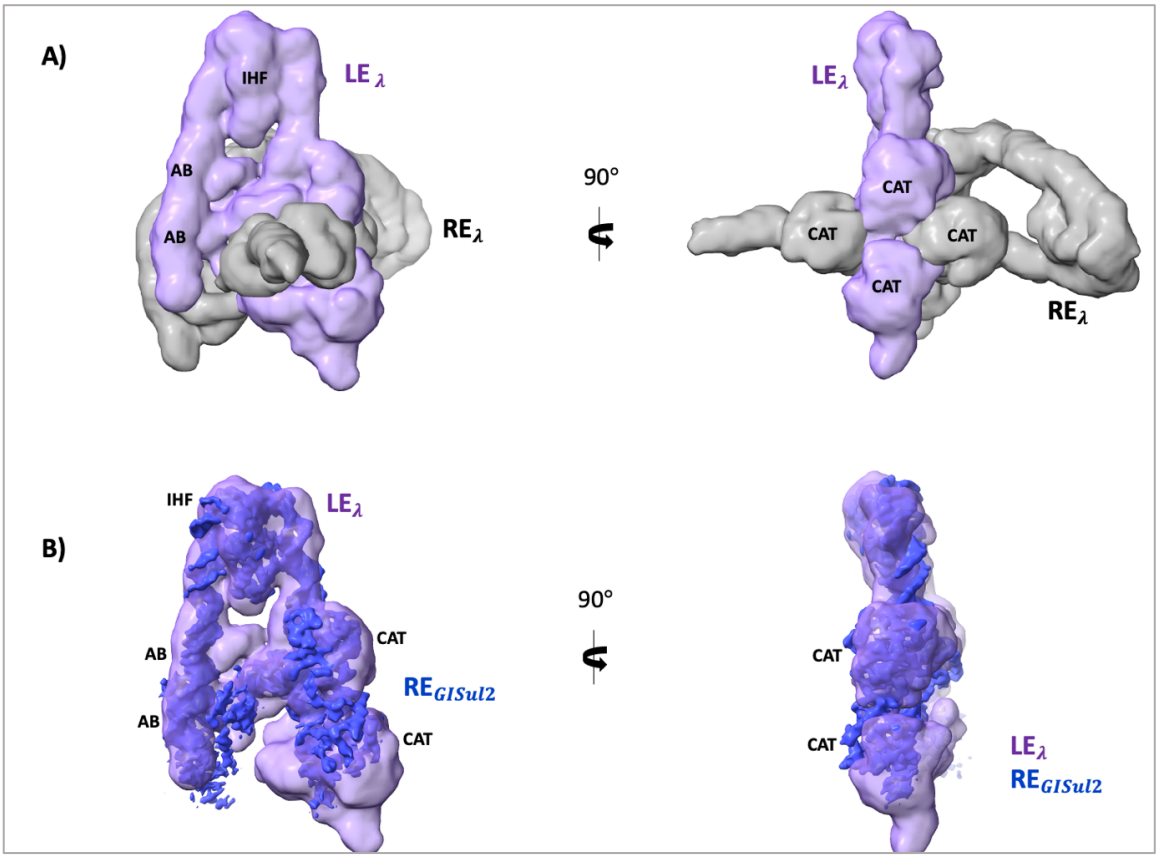


Figure 2-20: Comparison of *GISul2* RE Complex with the λ Excision Complex. A) The excision complex of λ ; attL DNA and bound proteins are colored in purple, whereas attR DNA and the proteins on it are shown in grey (Laxmikanthan et al., 2016). B) The overlay of λ attL DNA (in purple) with *GISul2* RE DNA (in blue).

2.1.10 Conclusions

The Right End complex of the *GISul2* element have revealed how Int_{GISul2} assembles on distantly located arm and core sites with the help of IHF. In the structure, while two Int_{GISul2} subunits are bound to two core repeats and two arm sites, due to the inverted orientation of one of the core sites, the overall assembly is asymmetric and the CB at the inverted core site is detached from the DNA. This dislocation is likely driven by the AB domains, which interact strongly with the arm sites and dictate tandem Int dimerization. Then, to bind to

the inverted core repeats, one Int molecule must “turn around” to place its CAT domain properly to execute DNA cleavage and recombination. Due to these acrobatics CB is dislocated, but CAT binds to the inverted core DNA. The interaction of CAT with its neighboring subunit, mainly through the exchanged M-helixes, together with some non-specific DNA interactions, seems to be sufficient to keep it connected to the recombination site.

While the core repeat at transposon end is highly conserved between various *GISul2* copies, its equivalent in the flanking DNA, which is inversely oriented, is not well conserved. Given that CB seems to be responsible for the vast majority of sequence-specific DNA interactions in all characterized TRs, its dislocation mechanism, may enable the *GISul2* element to tolerate different genomic sites and form a potent complex on the RE regardless of the flank sequences.

The dislocation of CB may also serve as a great link to the following steps of the recombination. For the formation of an active recombination complex, the RE complex needs to synapse with the LE complex. The dislocated CB can scan DNA for another core site to form the synaptic complex. Once CB recognizes a core sequence on the LE, it can attract its CAT to this new site and pull together two transposon ends. To perform such acrobatics, the inherent flexibility of the complex is functionally important, so as to allow dynamic binding and stepwise engagement with another core site for synaptic complex formation.

In the RE complex, IHF plays an indispensable role of bringing arm and core sites to close proximity for Int_{GISul2} binding. The DNA bending by IHF needs to be done at a very well-defined position in order to align the arm and core sites with a proper angle, degree and position to allow their simultaneous interaction with both Int_{GISul2} subunits. The exact position of the IHF binding site between arm and core sites carries high importance. Indeed in the λ system, moving the IHF binding site even by one bp reduces the recombination efficiency (Nunes-Duby et al., 1995).

Due to IHF’s importance in the complex formation, IHF regulation allows to directly control the recombination of the element. Therefore, by changing the levels of IHF, the host cell may be able to control the mobilization of the elements. From another perspective, the element can use IHF levels as a signal to determine the favorable time for its mobilization and continuation of its life cycle. Ints’s dependence on host-encoded accessory proteins may help linking the regulation and direction of CTn transposition to the physiology of the host cell. For example, IHF levels increase in the stationary phase by 5 to 7-fold and decrease at the end of the stationary phase, and IHF was shown to upregulate stationary phase genes. On the other hand, for the λ system, it is known that high levels of IHF concentration can have an inhibitory effect on excision *in vitro*. Thus, it seems reasonable

to propose that during the stationary phase, the host does not favor mobilization of mobile elements and it will also not be useful for the element, since its host will not be able support its mobilization with its reduced metabolism.

2.2 Characterization of the LE Complex of the *GISul2* Element

Once the structure of right transposon end complex of the *GISul2* element was obtained, I aimed to get insights into the left transposon end (LE) complex assembly to have a full picture about the presynaptic complexes on both transposon ends. In this chapter, I will describe mechanistic insights into the *GISul2* LE complex assembly.

2.2.1 *In vitro* Reconstitution of the LE Complex

The LE of the *GISul2* element was predicted based on the identification and characterization of conserved repetitive sequences (by Georgy Smyshlyaev) near the left transposon terminus (Figure 2-21 A). This showed that the core recombination site is located at the transposon boundary, with one core repeat placed in the transposon and one in the flanking genomic DNA. The crossover site between the two core repeats is situated one codon before the stop codon of the upstream *guaA* gene. Hence, the LE flank sequence includes the 3' end of the conserved *guaA* gene and is thus conserved among different *GISul2* copies. In contrast to the rather simple architecture of the arm sites in RE, there are a total of seven arm sites predicted in LE, which are positioned in various locations and orientations. Six Xis binding sites were also predicted between the arm and core sites, forming two separate patches (Figure 2-21 A). In contrast to RE, there are no IHF binding sites predicted in LE.

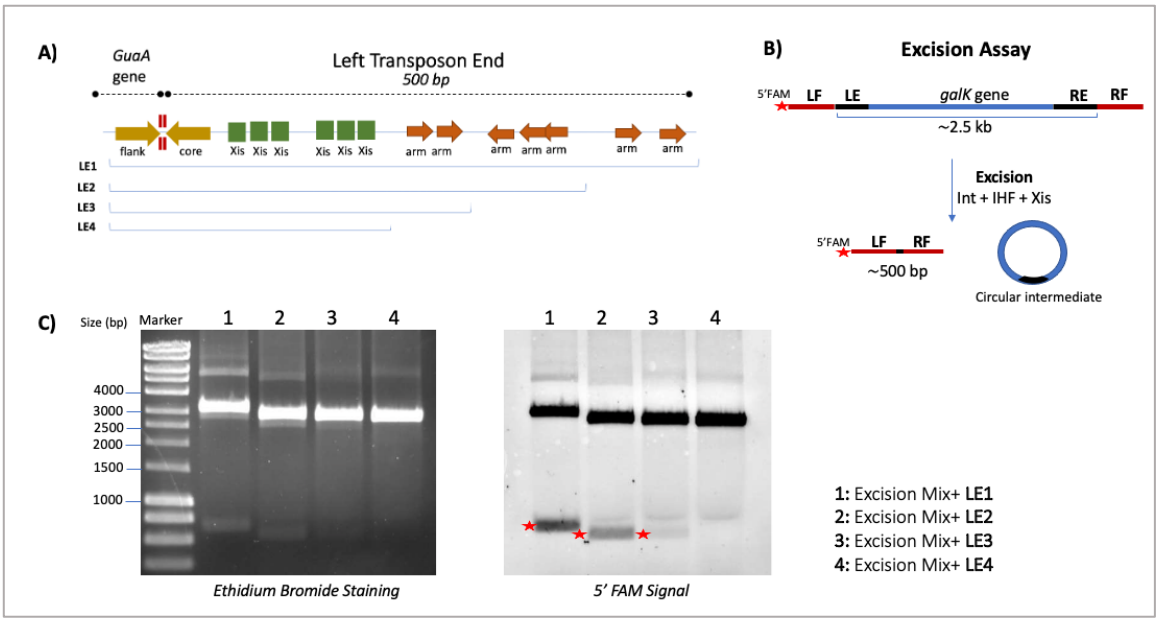


Figure 2-21: Architecture and activity of the left transposon end. A) Scheme of the *GISul2* LE. Various LE DNA constructs (LE1, LE2, LE3, LE4), containing different number of arm sites are marked. B) Design of the *in vitro* excision assay for *GISul2*. The linear DNA containing the mini-transposon sequence (blue and black) and flanking sequences (red) was amplified with 5' FAM labeled primer. The excision of CI is detected with the

appearance of an additional fluorescent band, containing the two flank sites. C) The excision assay was performed with four different LE constructs and analyzed by electrophoresis in 1% agarose gel. The ethidium bromide-stained gel is shown on the left, while the fluorescent (5' FAM, carboxyfluorescein) signal is shown on the right (excitation at 473 nm). The excision products are marked with red stars. The excision assay was performed with Lorenzo Rossi.

Prior to structural studies, the next step was to determine the minimum functional LE sequence required for excision. For this purpose, an *in vitro* excision assay was developed to detect the excision of the element from fluorescently labelled linear DNA. In this assay, we used specific mini-transposon constructs, which included the RE and LE of the *GISul2* element, flanking a *galK* gene (Figure 2-21 B). A DNA segment, comprising the mini- *GISul2* transposon and its flanking regions (total ~3 kb), was amplified by PCR with carboxyfluorescein (5' FAM) containing primers. The amplified sequence represented the transposon as integrated in its native site in the bacterial genome. Following incubation of the DNA fragment with Xis, IHF and Int_{GISul2}, we observed an additional fluorescent band with lower molecular weight (Figure 2-21 C), which corresponds well to the expected size of the two flank sequences fused together after excision of the fragment between RE and LE (Figure 2-21 C, marked with red star). With this assay, we tested the requirement of predicted regulatory LE sequences in excision by gradually removing pairs of arm sites. These experiments were performed with a MSc student in the lab, Lorenzo Rossi. When only the most proximal arms sites were retained, excision efficiency was still detected but reduced. In turn, removal of these sites completely inhibited excision (Figure 2-21 C-Lane 4), indicating that the first two LE arms are most essential for the reaction.

To gain structural insights into the LE assembly of Int_{GIsul2}, the complex was reconstituted *in vitro* using the shortest functional LE design, which was around 200 bp long and included the core site, two Xis binding patches and the first two arm sites (Figure 2-21 A; LE3). Next, I aimed to identify which components are required for LE complex formation and analyze complex heterogeneity using size exclusion chromatography (SEC). However, this gave ambiguous results with multiple peaks. As an alternative method, then electrophoretic mobility shift assay (EMSA) was performed to screen and optimize complex formation.

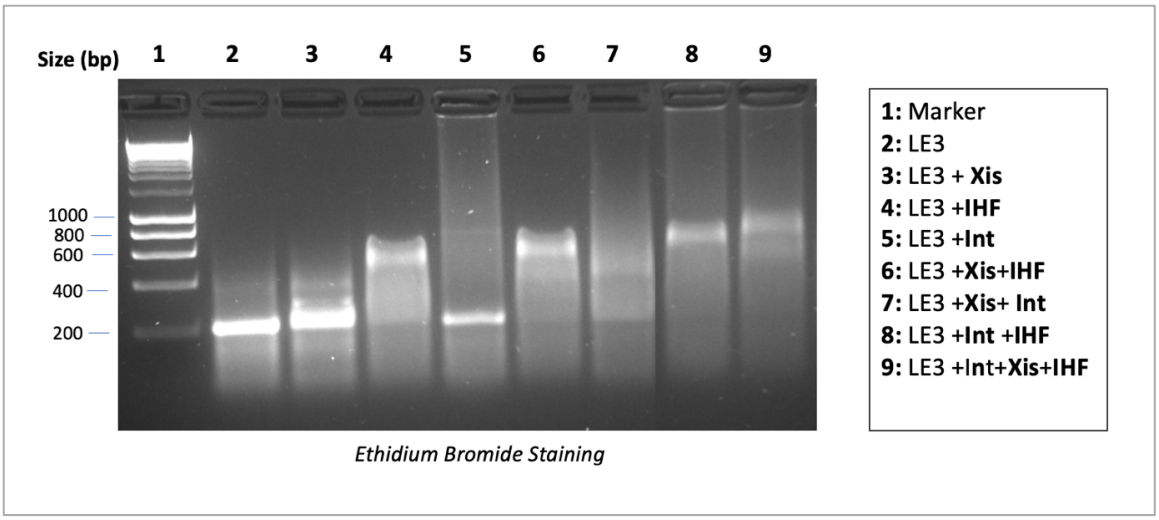


Figure 2-22: LE complex formation with different *GISul2* proteins and host factors. Electrophoretic Mobility Shift Assay (EMSA) with the LE3 DNA design (see Figure 2-21). The assay was performed by mixing LE3 DNA with the respective proteins, listed on the right for each sample. The complexes were analyzed by electrophoresis on native 1% agarose gel, bands were visualized by EtBr staining. The retarded mobility of the DNA indicates complex formation.

LE3 DNA was mixed with the respective proteins depicted in Figure 2-22 and diluted to a final buffer containing 100 mM NaCl to allow protein-DNA complex formation. After two-hour incubation at room temperature, the mixtures were analyzed by electrophoresis on a 1% agarose gel to detect and compare the mobility of LE DNA. In this assay, protein binding causes retardation of the electrophoretic mobility of the DNA, leading to an observed shift in its position on the gel. Although there was no IHF binding site identified in LE, IHF was still added because it was reported to bind to a broad range of binding sequences (Goodman et al., 1999). In the EMSA, we observed that LE DNA was shifted with addition of Xis_{GIsul2} or IHF separately (Figure 2-22; lane 3, 4), indicating their binding to the DNA. However, Int_{GIsul2} alone induced no shift and did not seem to bind strongly to the DNA. When Xis_{GIsul2} was added together with Int_{GIsul2}, the DNA shifted, but did not create a homogenous band, instead we observed a smear. Once IHF was added together with Int_{GIsul2}, the DNA shifted as a single band, indicating homogenous complex formation (Figure 2-22; lane 8). With the addition of Xis_{GIsul2} to this mix, the band shifted even more, suggesting formation of a quaternary complex, with LE3, IHF, Xis_{GIsul2} and Int_{GIsul2} (Figure 3-3, lane 9). In our subsequent reconstructions of the LE complex, we used the protein/DNA ratios optimized in this assay (Int: IHF: Xis: DNA, 2:2:7:1.5). For cryo-EM grid preparation, the components were mixed at the optimized ratios in 500 mM NaCl containing buffer and dialyzed overnight to 50mM NaCl containing buffer to promote the formation of an ordered protein-DNA complex.

2.2.2 Cryo-EM Grid Preparation and Data Collection

The dialyzed complex was applied to cryo-EM grids and then plunged them into liquid ethane. The same type of grids, conditions and parameters were used, which were previously optimized for the RE complex (Section 2.1.2). The grids were first screened with a Talos Arctica microscope to optimize ice thickness and particle distribution. Due to the big size of LE DNA, we were able to easily identify the particles with the looped DNA from the micrographs already prior to data processing (Figure 2-23 A).

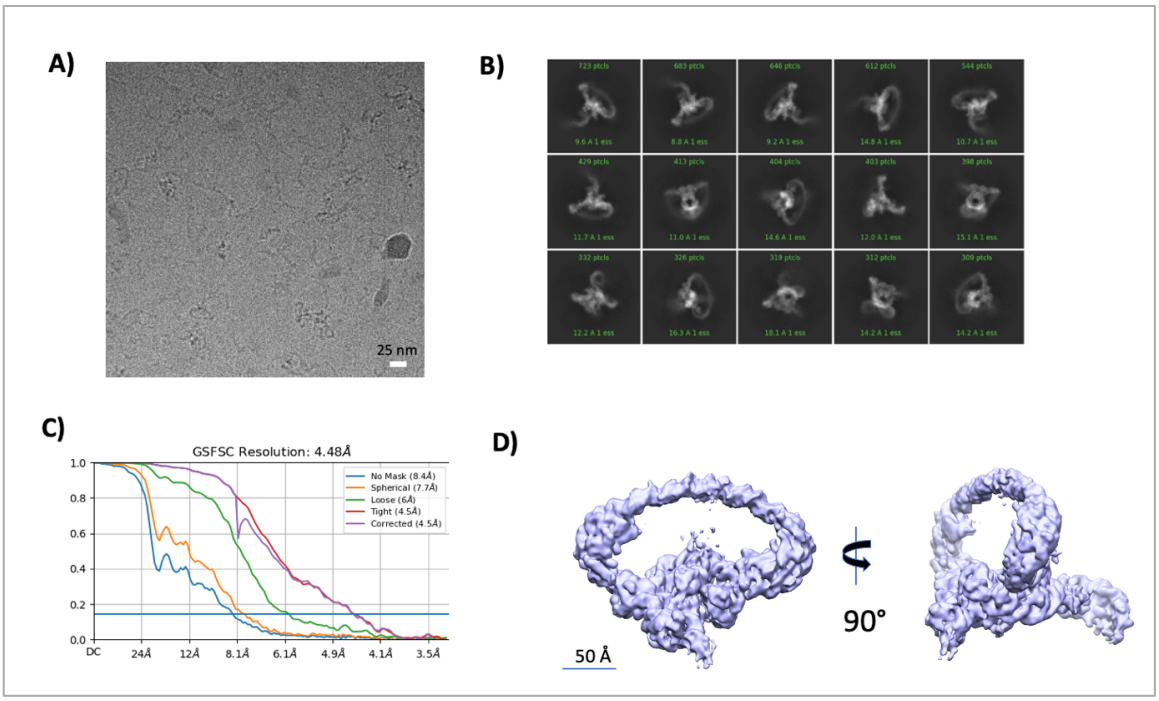


Figure 2-23: The LE complex cryo-EM data processing. A) A representative cryo-EM micrograph after motion correction. The micrograph was acquired with the pixel size of 0.81 Å. B) Representative 2D-classes of the LE complex. C) Fourier shell correlation (FSC) curve showing the 4.5 Å resolution calculated based on 0.143 gold standard FSC cut-off. D) The cryo-EM map of the LE complex.

To obtain a high-resolution structure, we then collected a dataset in a Titan Krios (300 keV, Gatan K2 detector) microscope with a total electron dose of 41 e/Å2 at pixel size of 0.81 Å/pixel. The collected 8600 movies were processed with Cryosparc2 software (from Strucutra). Like in data processing of the RE complex, movies were patch-motion corrected and patch-CTF estimation was performed. After manual curation of micrographs based on relative ice thickness and CTF-fit resolution, around 1000 particles were manually picked from 20 micrographs and classified to create initial 2D classes. In the next step, particles were picked, using the templates generated from some of the 2D classes produced in the previous step. The resulting particles were extracted and classified, which yielded 2D classes with high-resolution features (Figure 2-23 B). These classes corresponded to various views of the complex, without any preferred orientations. The junk particles were carefully

removed with more rounds of 2D and 3D classifications. With the final particle set, containing 15 330 particles, the homogenous refinement was performed, which gave a reconstruction at 4.5 Å overall resolution (Figure 2-23 C,D).

2.2.3 The Cryo-EM Structure of the LE Complex

The final cryo-EM map revealed a widely looped DNA, the ends of which are connected with protein density. To build a structural model of the complex, the crystal structure of the full-length Int_{GISul2} bound to core and arm DNA and the crystal structure of Xis_{GISul2} were used (G. Smyshlyaev, unpublished data). For Int_{GISul2}, the domains were separated and fitted individually, due to their significantly different localization as compared to the crystal structure.

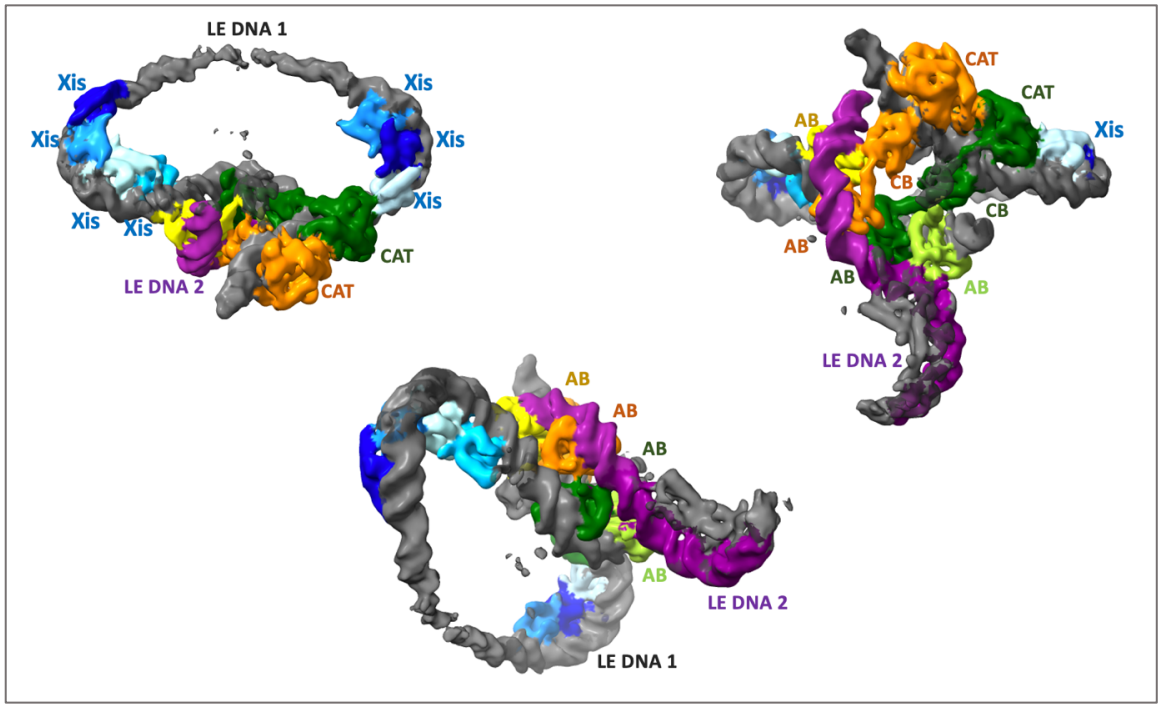


Figure 2-24: The segmented cryo-EM map of the LE complex. Three different views are shown. LE DNA 1 (in grey) is widely looped by a total of seven Xis $_{GISul2}$ molecules (in shades of blue) and connected by four Int_{GISul2} subunits (in yellow, orange, light and dark green). The second DNA fragment, LE DNA2 (shown in purple) is connected via AB-DNA interactions.

The resulting EM model has shown that seven Xis molecules bend the LE DNA to form a loop that is tethered by integrases. In contrast to the sharp DNA bend observed in the RE complex, in the LE complex, Xis bending results in a widely looped DNA. This shape is formed by a total of seven Xis monomers, which are bound to the DNA in two patches: the first one is located adjacent to the core DNA site and contains three Xis monomers, and

the second patch is adjacent to the arm DNA site and consists of four Xis proteins. Int_{GISul2} bridges the two ends of the looped DNA fragment. Two CAT domains bind to the core DNA repeats together with their CB domains, and the AB domains of the same proteins bind to the arm sites on the other end of the loop. Surprisingly, the AB domains are bound not only to one but to two separate DNA molecules (Figure 2-24, shown in gray and in purple). While one LE DNA is looped (in grey, marked as LE DNA 1) and connected on two ends by Int_{GISul2}, another LE DNA molecule is brought to the complex solely via interactions with the AB domains and only the arm DNA part of this second DNA is visible in the density (in purple, marked LE DNA 2). In total four AB domains hold the two DNA molecules together by forming a filamentous assembly (Figure 2-24). Two of these ABs come from the Int_{GISul2} molecules bound to the LE core site, whereas for the other two, no connecting CB and CAT density can be detected. This is supposedly due to the flexibility of those domains in the observed structure. Although the interaction of two LE DNA molecules in one nucleoprotein complex is not expected during transposon excision, the observed assembly likely mimic the first step of synaptic complex formation, where arm sites from one LE and one RE DNA are connected through similar AB interactions.

2.2.4 The organization of Int_{GISul2} Domains in the LE Complex

The overall organization of the Int_{GISul2} domains in the LE complex is distinct from the RE complex. In this structure, two CAT and two CB domains bind to the core repeats located in the transposon end and flank region, and a total of four AB domains bind to four arm DNA sites (Figure 2-25). The AB of the Int_{GISul2} subunit bound to the flank region interacts with the arm site located at the other end of the same LE DNA molecule (Figure 2-25, dark green Int). However, the AB that is connected to the CB-CAT module sitting on the transposon end (in orange) cannot reach the second arm site of this LE DNA. Instead, it binds to the arm site of another DNA molecule in the complex. While these two Int molecules each occupy one arm site in two distinct DNA fragments, the remaining two arm sites are bound by the AB domains of two additional Int_{GISul2} molecules (light green and yellow). However, these molecules do not manage to capture any core sites, and they are not visible in our cryo-EM map, probably due to lack of stable conformation.

The reason why the LE complex did not form a typical presynaptic assembly with just one DNA molecule is provided by the localization of arm, core and Xis binding sites in LE. With the two wide turns introduced by Xis proteins, the arm sites are not positioned above core repeats, where both sites could be simultaneously reached by Int_{GISul2} dimer, instead they are shifted to the right (Figure 2-25 C). This orientation of the arm and core sites, only one Int monomer can bind to both the arm and core sites of the same DNA molecule. In turn, the Int_{GISul2} subunit bound to the second core repeat, recruits another DNA molecule,

where it finds other arm sites. In this way, the Int_{GISul2} AB domains drive synapsis of two DNA molecules.

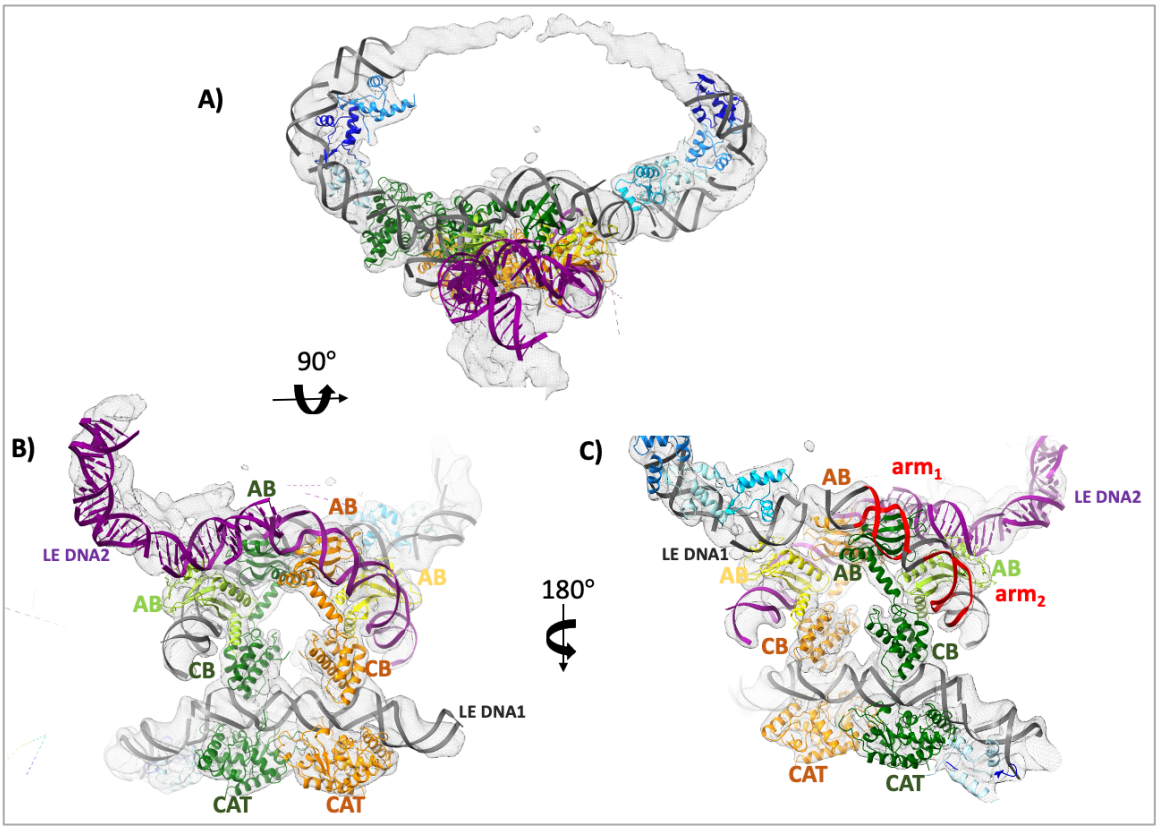


Figure 2-25: The Cryo-EM structure of the LE complex. A) Overview of the structure. Crystal structures of Xis and Int_{GISul2} were fitted in the cryo-EM map (transparent surface) and are represented by cartoons. The LE DNA1 (grey) is looped by seven Xis monomers (blue), which are organized in two patches. A second DNA fragment, LE DNA2, is shown in purple. B) Bottom view of the LE complex with 90° rotation around the horizontal axis. Two Int_{GISul2} molecules are fully traced, with their CB-CAT domains bound to the core repeats on LE DNA1, one AB domain bound to an arm site in the LE DNA1, and the other AB bound to an arm site in LE DNA2. Additional two AB domains bind to LE DNA1 and 2 but their CB and CAT domains are not visible in the map. C) Top view of the complex with 180° rotation around the vertical axis from B. Two arm DNA sites in LE DNA1 are highlighted in red.

When compared to the RE complex, the CAT domains are organized similarly but the CB domains are located differently in LE. In the RE structure, one of the CB was dislocated from the core DNA to allow AB binding to directly oriented arm sites. In the LE case, AB binding to an additional DNA molecule allows to conveniently accommodate both CBs on one core DNA site. These AB-mediated rearrangements suggest that arm site positioning and orientation has a central role in structuring the inner architecture of *GISul2* complexes during transposition, which is reminiscent of previous reports for the λ phage.

2.2.5 The AB Filament

In the EM structure, ABs are organized in a way that they form a filament-like structure that glues two DNA fragments together. The AB filament is assembled on four arm sites from two DNA molecules, which are positioned in an antiparallel fashion. Each DNA contains two arm sites oriented as direct repeats. Four AB domains bind primarily to the four arm sites, each time introducing a slight bend in the DNA. Additionally, the back side of each AB interacts with the second DNA molecule in the filament, in a segment adjacent to the arm sites (Figure 2-26 B).

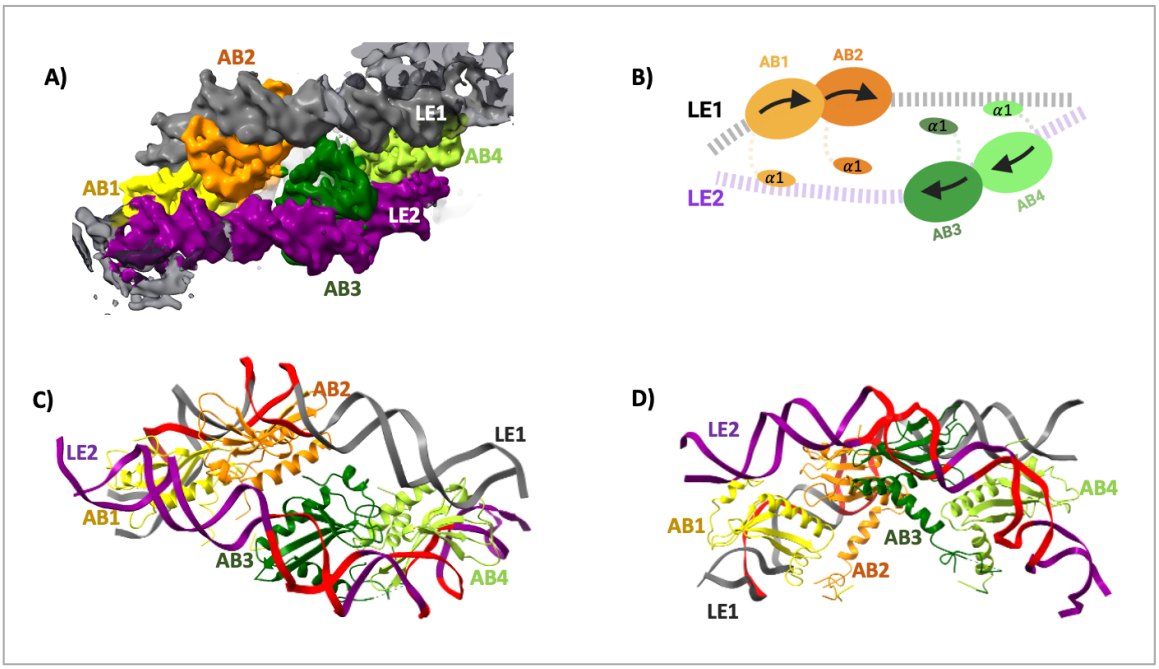


Figure 2-26: The AB Filament Structure. A) Segmented cryo-EM map of the AB filament. LE DNA1 (in grey) and LE DNA2 (in purple) are brought together by the interaction of four AB domains. B) Schematic diagram of the AB filament. ABs are shown as large ovals, with their N-terminal alpha helices (α1) depicted as smaller ovals connected to the ABs. The directionalities of the arm DNA sites are marked by arrows. C) Structural model of the AB filament. The same color scheme is used as in A), and the arm sites are colored in red. D) Side view of the AB filament, after 90° rotation around the horizontal axis from C.

AB filament formation is mediated by two distinct DNA binding regions of the AB domain. First, it binds to arm sites in a sequence specific manner, using its four stranded beta-sheet, which docks into the major groove of the arm site DNA. Additionally, it interacts non-specifically with the second DNA molecule with its N-terminal alpha helix ($\alpha 1$). This way, AB can simultaneously bind to two DNA molecules and bring them together. The $\alpha 1$ of AB2 and AB3 (numbered as in the Figure 2-26 B,) is located near the major groove of the partner DNA, while AB1 and AB4 place their $\alpha 1$ in the DNA minor groove. AB1 and AB4 form stable interactions through the T6 and T8 residues contacting the DNA bases and R11 and K9 stabilizing the backbone. The $\alpha 1$ from AB2 and AB3 does not seem to interact as tightly with the neighbor DNA molecule, probably because the curvature induced by AB binding

brings it further away in this region of the filament. Hence, to form a stable AB filament with two DNA fragments, at least four AB domains are need, since the stable $\alpha 1$ AB-DNA interactions occur for the peripheral (AB1 and AB4), but not for the central (AB2 and AB3) ABs.

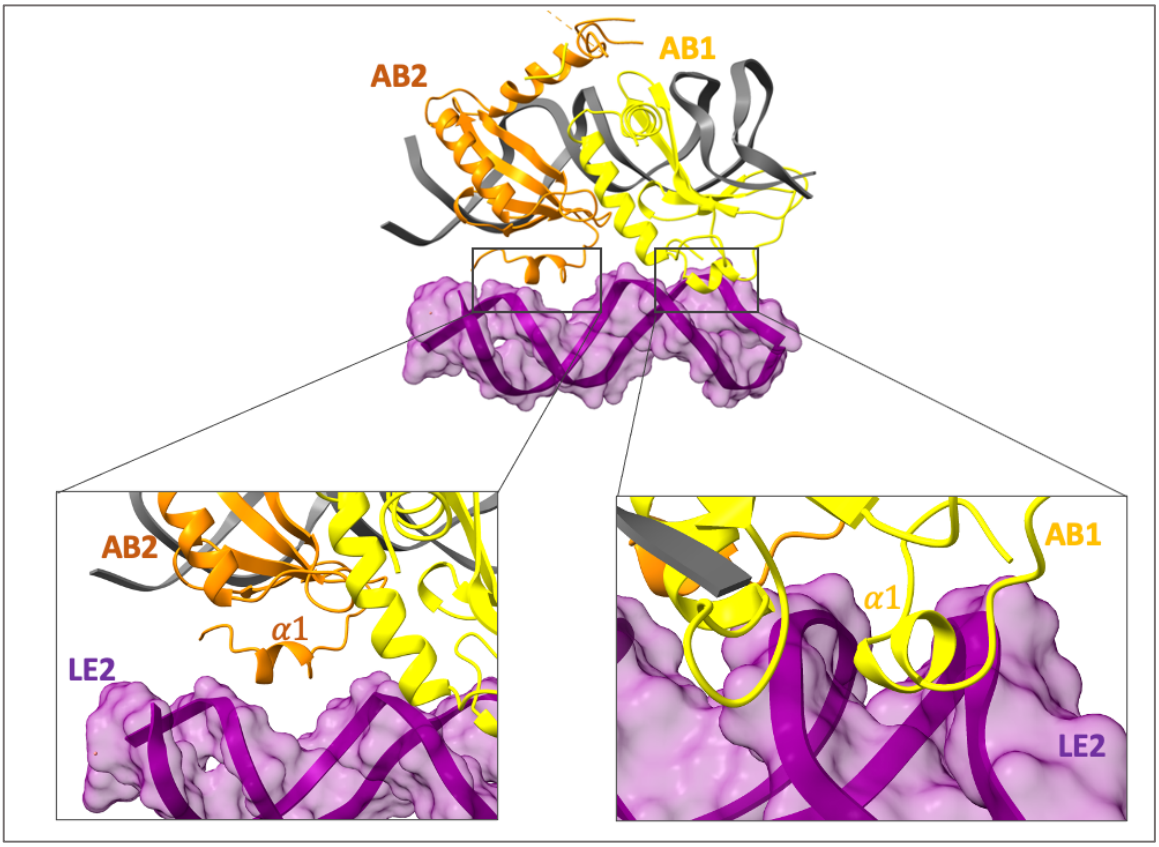


Figure 2-27: The Interaction of AB domains with DNA through α1 helix. The coloring is the same as in the Figure 2-26. The interaction of α1 with the partner DNA drives the formation of the AB filament with two DNA molecules. Since AB binding bends DNA, the α1-mediated interactions are not the same for all ABs. While the peripheral AB1 and AB4 tightly interact with DNA (bottom right), AB2 and AB3 are more loosely connected (bottom left).

Thus, in Int_{GISul2} $\alpha 1$ of the AB seems to play a central role in the formation of the arm filament. As previously noted in Section 2.1.7, this helix is only found in ABs of the SXT family TRs. Indeed, for example in the λ system, AB does not form any bilateral filament-like organization (Biswas et al., 2005).

2.2.6 The Characterization of Xis Bending

In the LE structure, Xis is crucial for DNA looping and for bridging the ends of the DNA together to enable Int_{GISul2} binding. The Xis binding sites are organized in two patches, which are separated by 44 bp and contain three and four Xis binding sites, respectively. At each of these patches the DNA bends cooperatively upon Xis binding and essentially inverts its direction. By having two Xis-mediated turns in the LE DNA, its direction changes two times, eventually coming back to the same place with the same polarity.

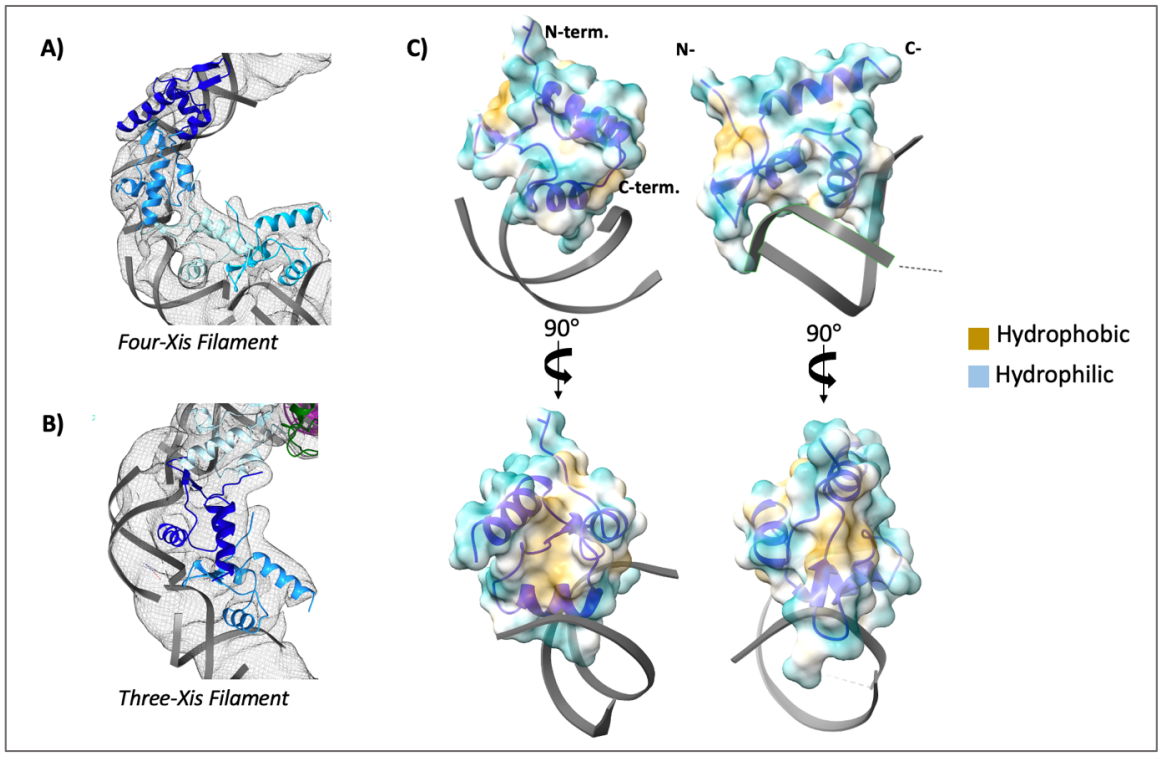


Figure 2-28: Cooperative DNA binding by Xis. A) and B) show two different Xis patches (see figure 2-25), where Xis molecules bind to DNA and form filaments. A) The Xis filament formed adjacent to the LE arm sites. B) The Xis filament located next to the core DNA site. C) The surface hydrophobicity is shown for two adjacent Xis molecules, which are positioned head-to-tail fashion on the DNA. The interaction interface is shown in the 90° rotated view at the bottom. Yellow patches depict the hydrophobic surfaces; blue represents hydrophilic parts of the proteins.

Xis is composed of three α-helices and a two-stranded β-sheet. It binds to a set of consecutive binding sites on the LE DNA and makes tight interactions between individual subunits, forming a filament-like structure (Figure 2-28 A, B). In the filament, protein monomers bind in a head-to-tail fashion, with the β-sheet of the leading molecule docked to the next molecule (Figure 2-28 C). The interaction of two molecules is strengthened by hydrophobic patches found in the interaction surfaces (Figure 2-28 C).

In the λ phage, Xisλ plays a significant role in formation of the excision complex. Together with IHF, they bend the attR DNA. Additionally, Xis was proposed to interact with the AB domain of Intλ and help to drive synaptic complex formation. In the case of *GISul2,* AB filament formation has a primary role in driving the synapsis, and Xis does not seem to be required for this purpose.

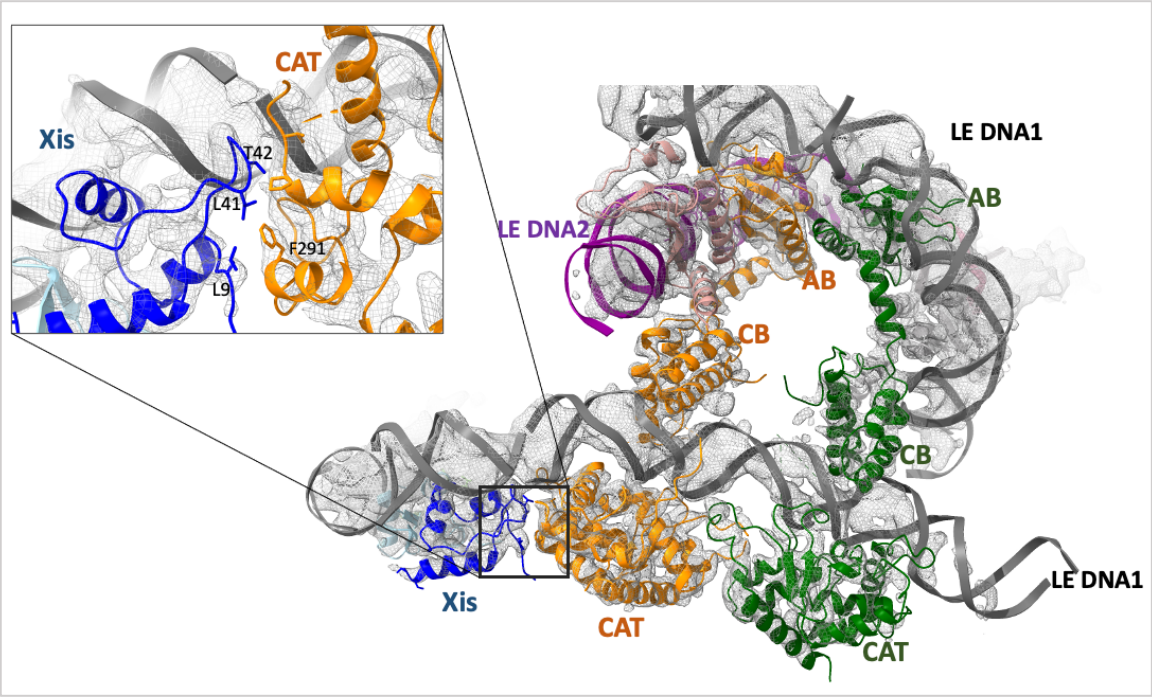


Figure 2-29: The Int$_{GISul2}$ CAT - Xis Interaction. The CAT domain bound to the transposon end core site (in orange) interacts with the adjacent Xis molecule (in dark blue). Insert: Zoom-up of the hydrophobic residues located at the interaction surface.

Surprisingly, in the LE structure Xis interacts with one of the Int$_{GISul2}$ CAT domains. Given the proximity of their binding sites, the two proteins are located closely on DNA and interact thorough hydrophobic residues on their proximal surfaces (Figure 2-29). This interaction may have a significance for the stabilization of the overall DNA conformation. It anchors the looped DNA at a fixed position relative to the core DNA site and may increase the affinity of CAT to its DNA binding site, thereby promoting both the correct formation and activity of the excision complex.

2.2.7 Flexibility of the LE Complex

The flexibility of the LE complex was analyzed by 3D-variablity analysis (CryoSparc2). This analysis has shown that the central part of the DNA loop is the most flexible part of the complex. In fact, the LE DNA is bent by Xis at the two ends, and the middle part, which is not bound by any proteins, can probably have various conformations, and thus is not well resolved in the final EM map.

When the final EM map was checked at different thresholds, it was observed that the middle of the loop was connected to the arm DNA site by weak electron density. As the gap between the loop and the arm DNA narrows down, the extra density appears to connect the loop with the integrase arm filament. This additional density could be an IHF molecule, which was present in the reconstituted LE complex but not detected in the map so far. In this position, IHF could bridge and further stabilizing the loop. However, this density was not well resolved, and it is not possible to reliably fit any structure.

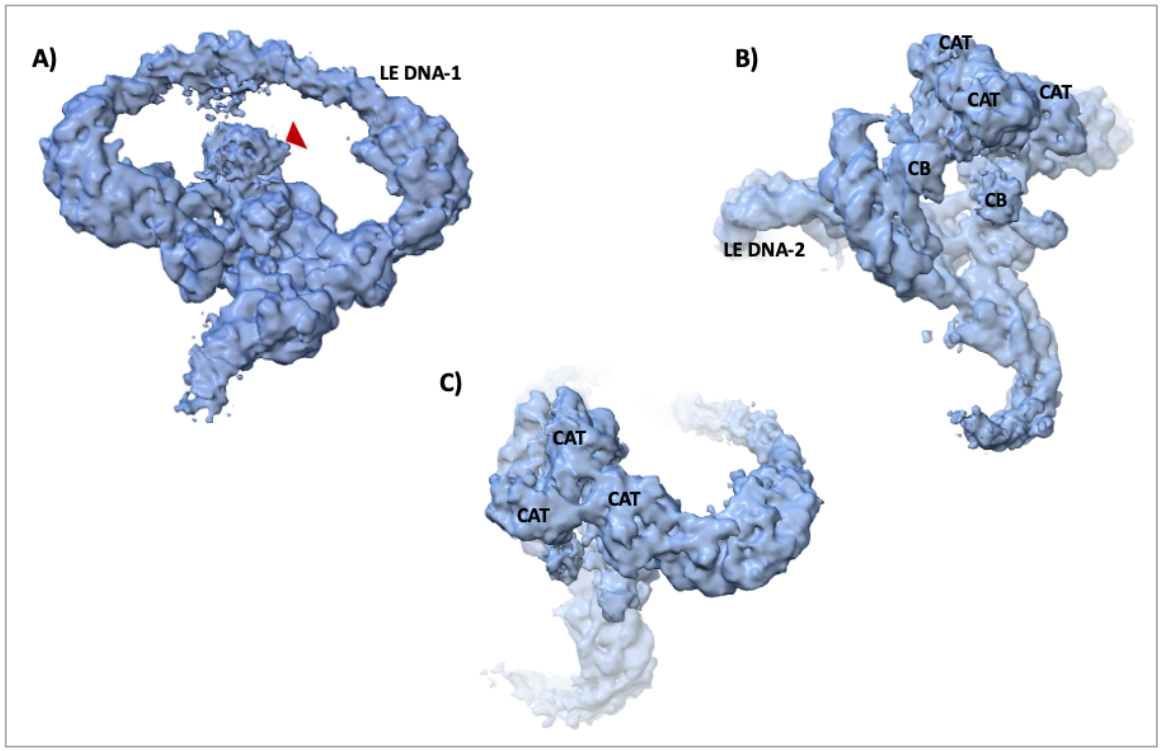


Figure 2-30: 3D variability analysis of the LE Complex. A) Extra density appearing in the middle of the LE DNA1 loop (shown with a red arrow). B) and C) Presence of additional density corresponding to the third CAT in a subset of particles.

In addition to of the DNA loop, there was mobility observed in the Integrase-bound region of the complex. Upon careful assessment, the ambiguity could be attributed to the presence of a third CAT domain in a subset of the EM particles (Figure 6-B, C). This CAT

belongs to the one of the additional AB domains, whose CAT was not observed in the original density reconstruction due to the low particle number representing this density. Together the three CAT domains create a trimeric configuration will be discussed in more details in the next chapter.

2.2.8 Conclusion and Discussion

The left end (LE) complex of the *GISul2* element has revealed how Int_{GISul2} assembles on distantly located arm and core sites with the help of Xis-mediated DNA bending. Surprisingly, in this structure two separate DNA molecules were captured in the LE complex, possibly revealing the first step of synaptic complex formation. Only one of the two Int_{GISul2} monomers bound to the LE core DNA site can reach and simultaneously bind to the arm site from the same DNA molecule. In turn, the second monomer recruits another LE DNA fragment and binds to its arm DNA site. This drives the formation of a filamentous AB-DNA structure, which contain in total four AB domains, and glues two DNA molecules.

This structure shows that the LE complex tends to seek binding and dimerization with a second piece of arm site containing DNA via AB-DNA interactions. Given the high level of DNA sequence identity between LE and RE arm DNA, the second DNA fragment in the LE structure may actually represent how the RE DNA can be brought in close proximity to the LE. This may provide an elegant mechanism for the synapsis of two transposon ends via the AB filament formation. The observed AB filament provides an exciting and novel mechanism to form the synaptic complex. Previously in the λ structure, an interaction between Xis and the AB domain of Int was suggested to drive synaptic complex formation. With the current structure, it was clear that the *GISul2* element follows a strikingly different way for the synapsis of two DNA fragments, in which each AB domain can bind two separate DNA sites at the same time.

Flexibility analysis of the LE complex revealed an additional density, which connects the DNA loop to the arm sites. This density can be IHF, since its presence is essential for LE complex formation in the EMSA assay, and therefore included it in the complex preparation. The lack of IHF in the final structure could also be explained by its role in the earlier stages of the LE complex assembly and subsequent dissociation. The density observed in the middle of the DNA loop does not fit to a canonical “fully-wrapped” mode of IHF binding, as observed, in the RE complex. However, IHF was predicted to have various binding modes, including a so-called “bridged” mode, where IHF could connect two segments of DNA, offering a better fit to our observed electron density (Yoshua et al, 2020).

2.3 Characterization of the Synaptic Complex involved in *GISul2* Excision

Following the characterization of RE and LE complexes, we aimed to get insights into how these two ends come together to form a synaptic complex during transposon excision. While RE and LE structures gave hints to predict the final assembly of the synaptic excision complex, the exact arrangement remained elusive. In this chapter, I will describe the experimental setup that we used to obtain structures of the synaptic complex and the mechanistic insights that these structures revealed.

2.3.1 *In vitro* Reconstitution of the RE-LE Synaptic Complex

During the excision of *GISul2*, the complexes formed on RE and LE are expected to come together to form an excision-competent nucleoprotein assembly. To capture a stable synaptic complex *in vitro*, the components of the RE and LE complexes were mixed with modified DNA substrates, which were designed to trap recombination before resolution to products via several approaches.

In the first approach, the catalytically inactive Int_{GISul2} was used with RE DNA having a mismatched crossover region (Figure 2-31, Design1). This DNA design produces a DNA fragment which is more flexible at the core site and can be easily bent to promote the formation of a synaptic complex. Use of the Int_{GISul2} mutant will then trap the complex prior to cleavage of the first DNA strand. To form the complex, RE and LE complexes were prepared separately, mixed and dialyzed to a buffer with lower salt concentration overnight. The complex was probed directly on EM grids, by collecting a test dataset in the Talos Arctica microscope. In this data collection, the structures of RE and LE complexes were obtained separately, while there was no synaptic complex observed.

Based on the previous LE structure, it was predicted that the synapsis of RE and LE complexes will be driven mainly by AB filament formation. This assembly requires bipartite interaction of the Int_{GISul2} AB domain with the arm site of one DNA molecule and the adjacent region of the other DNA in the synapse. However, in the previous RE DNA design, there was no adjacent DNA that could be used for the formation of the required assembly since this construct was ending right after the arm sites. Therefore, the previous RE DNA substrate was extended with additional 20 bp from upstream of the arm site (Figure 2-31, Design 2). To further promote correct synapsis and disfavor LE-LE self-assembly, the procedure of complex formation was also changed. The RE complex was pre-formed using catalytically inactive Int_{GISul2} as before, followed by addition of the LE complex components separately to the RE complex. The mixture was then dialyzed into 50 mM NaCl containing

buffer overnight. With the dialyzed complex, another test dataset was collected, which resulted in separate RE and LE complexes as in the previous case.

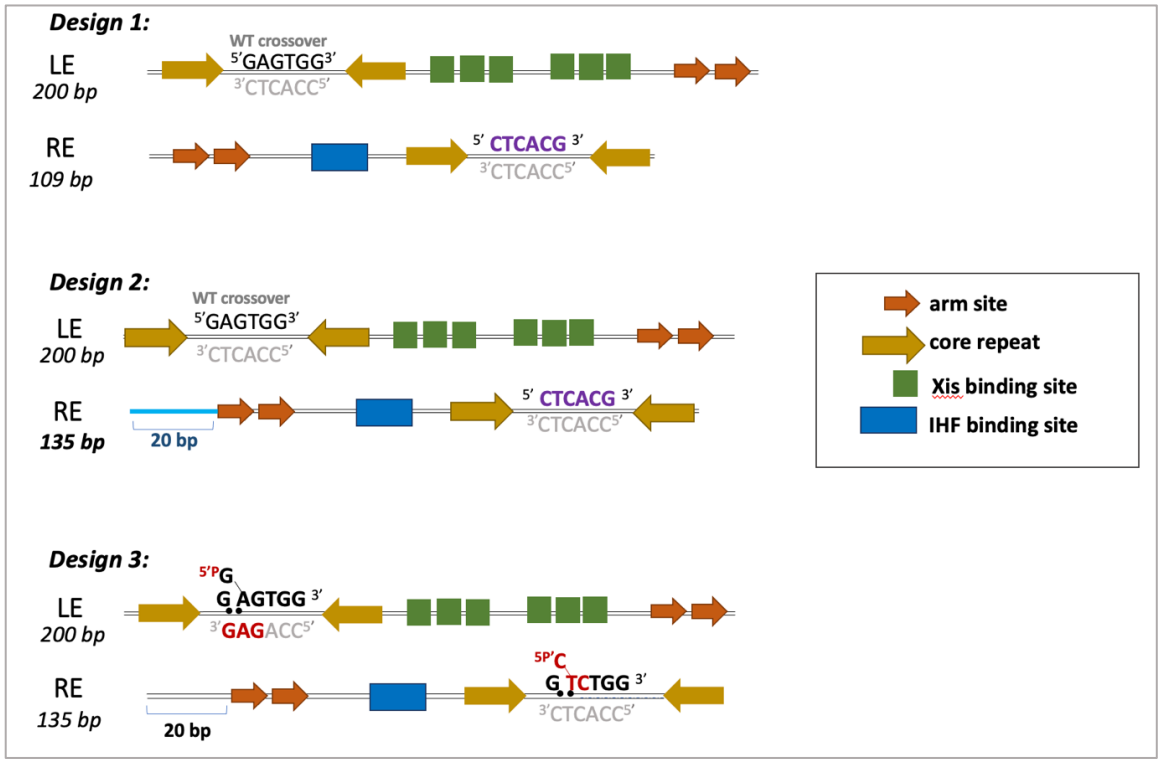


Figure 2-31: DNA designs to trap RE-LE synaptic complex. The Int_{GISul2} binding sites are shown as arrows (core repeats in yellow and arm sites in orange) and accessory factor binding sites are marked as boxes (IHF binding site in blue and Xis binding site in green) on the RE and LE. In Design 1, the LE DNA is 200 bp and RE is 109 bp long. While for LE the wildtype crossover site (capital letters) is used, the RE crossover site is mutated in the top strand, shown in purple. In Design 2, the same LE construct is used, whereas RE is extended by 20 bp upstream from arm sites (in blue). In Design 3, the same length of RE and LE constructs were used as in Design 2. The DNA constructs were prepared with a "suicide" design, each containing a nick and an extra nucleotide in the top strand. Following the nick, the DNA is 5' phosphorylated to trap the reaction after first strand cleavage. The mutated nucleotides are shown in red.

Since previous constructs designed for capturing the synaptic complex prior to first cleavage did not result in a stable complex, the next approach was to trap the synaptic complex after first strand cleavage. For this purpose, "suicide" substrates were used, which were designed to stall the recombination with wild type Int_{GISul2}. In this design, one DNA strand contained a nick in the backbone and an extra nucleotide, after the natural cleavage site in both LE and RE (Figure 2-32). This construction allows Int_{GISul2} to attack and cleave its cleavage site, by cutting off a single nucleotide in each nicked DNA strand. These extra nucleotides then diffuse away and can be replaced with an equivalent from the oligo placed after the nick. To prevent the reaction to proceed or revert, the oligo after the nick was 5' phosphorylated to preclude its attack on the Int-DNA phosphotyrosine linkage and trap the

complex in the covalent intermediate state after first cleavage. Although the nicked strands cannot go forward with chemistry, they can still exchange between the recombining molecules and form a four-way DNA structure, thanks to the presence of complementary DNA bases in their crossover regions. However, the intermediate won't be a full Holliday Junction, since the strands cannot be ligated.

In addition to "suicide" design, the crossover regions of both LE and RE were mutated to have three non-complementary base pairs (Figure 2-32, in red) in such a way that the new sequences can form complementary base pairs with the recombining DNA partner after strand exchange. In this setup, RE and LE have mismatched and unstable crossover regions, whereas the produced four-way DNA will be fully complimentary and stable, which may promote its trapping.

To form the suicide LE and RE DNA substrates, three oligos were annealed for each. At first, the RE complex was prepared at room temperature with two-hour incubation to allow the formation of the trapped complex. Later, the components of the LE complex were added separately to the preformed RE complex and dialyzed at room temperature to gradually lower the salt concentration in the buffer.

2.3.2 Cryo-EM Grid Preparation and Data Collection

The dialyzed complex was then used for cryo-EM grid preparation. For this, previously optimized parameters were used (Section 2.1.2). The best particle distribution and ice formation was achieved with UltrAuFoil grids with R2/2 hole size and 200 mesh size (from Quantifoil), without using any detergent.

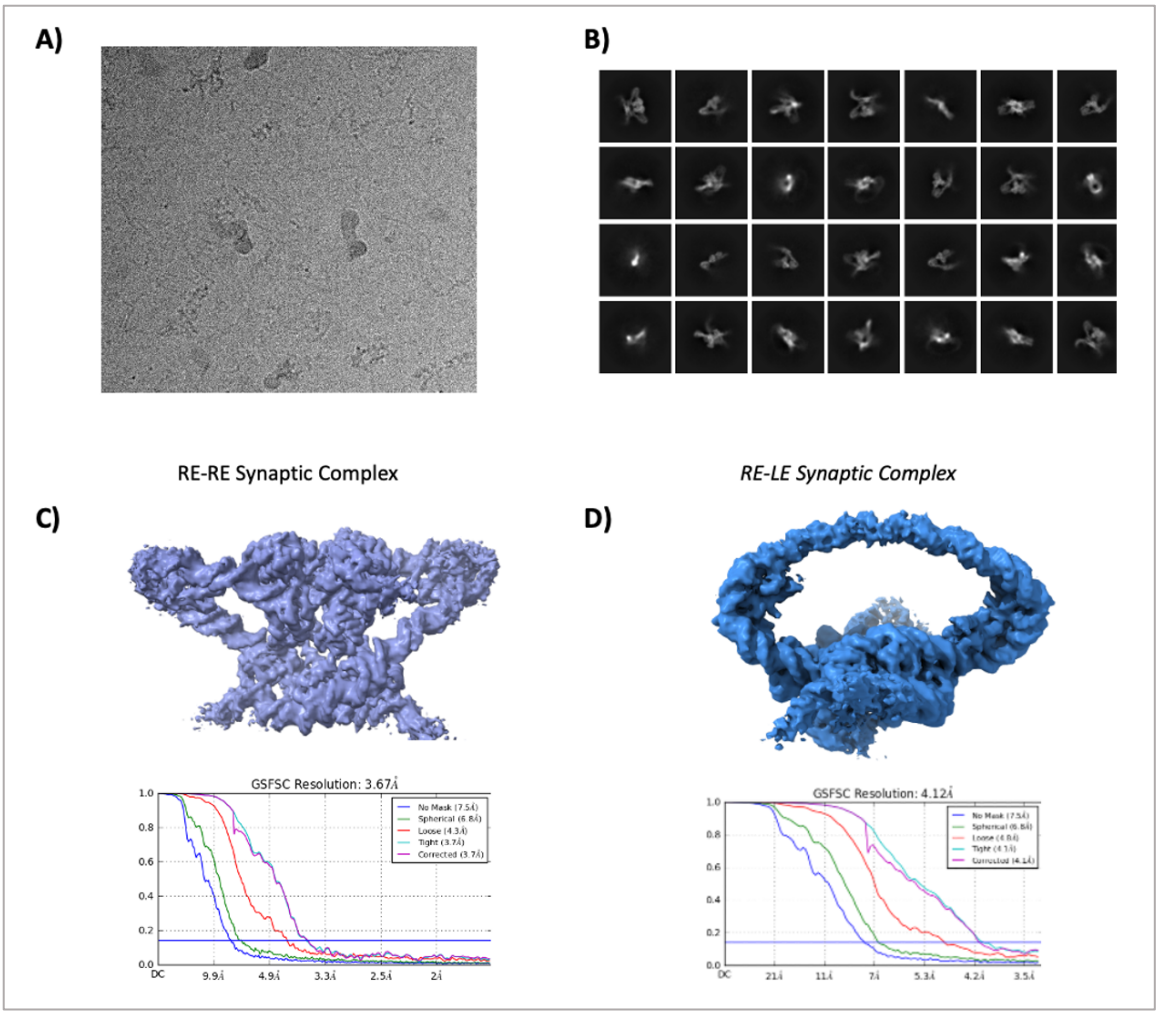


Figure 2-32: Cryo-EM analysis of the synaptic complexes. A) A representative cryo-EM micrograph after motion correction. The micrograph was acquired with the pixel size of 0.81 Å. B) Representative 2D-classes of this dataset are shown. C) Cryo-EM reconstruction of the RE-RE synaptic complex is shown on the top, and Fourier shell correlation (FSC) curve showing 3.67 Å resolution calculated based on 0.143 gold-standard FSC cut-off is depicted below. D) Cryo-EM reconstruction of the RE-LE synaptic complex is shown on the top and (FSC) curve showing 4.12 Å resolution is depicted below.

A full dataset was collected in Titan Krios (300 keV, K2 detector) microscope, with a total electron dose of 48 e/Å^2 at a pixel size of 0.81 Å/pixel. The collected 11 600 micrographs were processed with Cryosparc2 software (from Structura). As in previous cryo-EM dataset processing, movies were patch-motion corrected and CTF was estimated with patch-CTF estimation. The micrographs were manually curated based on relative ice thickness and CTF-fit resolution, to eliminate the micrographs with thick ice. The initial set of particles were picked with template-based particle picking, using 2D projections of the LE complex structure as templates. Particles were extracted and 2D classification yielded classes, which have high-resolution features (Figure 2-32 B). Once ab-initio volume reconstruction was performed, there were two different complexes detected in this dataset: the first one contained two RE DNA molecules assembled in one complex, and the second one had

synapsed one LE and one RE DNA. With the grouped particles corresponding to the two different complexes, 2D classes were produced containing more views of the two complexes. By using these 2D classes as templates, more particles were picked, and they were separately cleaned with 2D and 3D classifications. After final refinements, for the complex which had two RE complexes synapsed, a cryo-EM reconstruction at 3.9 Å resolution was obtained with 81 900 particles by using homogenous refinement (Figure 2-32 C). For the other complex, containing synapsed RE and LE complexes, a cryo-EM map was obtained at 4.12 Å resolution with 80 819 particles (Figure 2-32 D).

2.3.3 The Cryo-EM Structure of the RE-RE Synaptic Complex

The cryo-EM data processing revealed two different complexes in the synaptic complex sample. While one complex contains an RE DNA synapsed with LE (RE-LE), as expected; the other complex features two interacting RE DNA molecules (RE-RE).

In the RE-RE synaptic complex, both DNA molecules are sharply bent by IHF, and all four DNA ends are connected in the center by an Int_{GISul2} tetramer (Figure 2-33). The DNA bending of IHF is the same as in the pre-synaptic RE complex (Section 2.1.3). The synapsis of two RE DNAs is mediated mainly by AB and CAT domains of the Int_{GISul2} tetramer. AB forms a filament-like structure and glues the arm DNA sites. At the same time, CAT forms an almost square planar tetramer, which brings the two core DNA sites together (Figure 2-33 B). The AB filament architecture in the RE-RE complex is essentially identical to the one observed in the LE complex (Section 2.2.5), and it is arranged in a way that it connects two different RE DNA molecules in an antiparallel fashion. The filament is formed by four AB domains binding to arm sites on two DNA fragments, each containing two arm sites oriented as direct repeats. Each AB binds to an arm site in a sequence specific manner and interacts non-specifically with the second DNA molecule, connecting them directly in a tightly packed arrangement (Figure 2-33 C).

To build a structural model of the complex, the structure of previously solved pre-synaptic RE complex was used. For Int_{GISul2}, the CB and CAT domains were separated and fitted individually, due to their significantly different localization as compared to the previous structure.

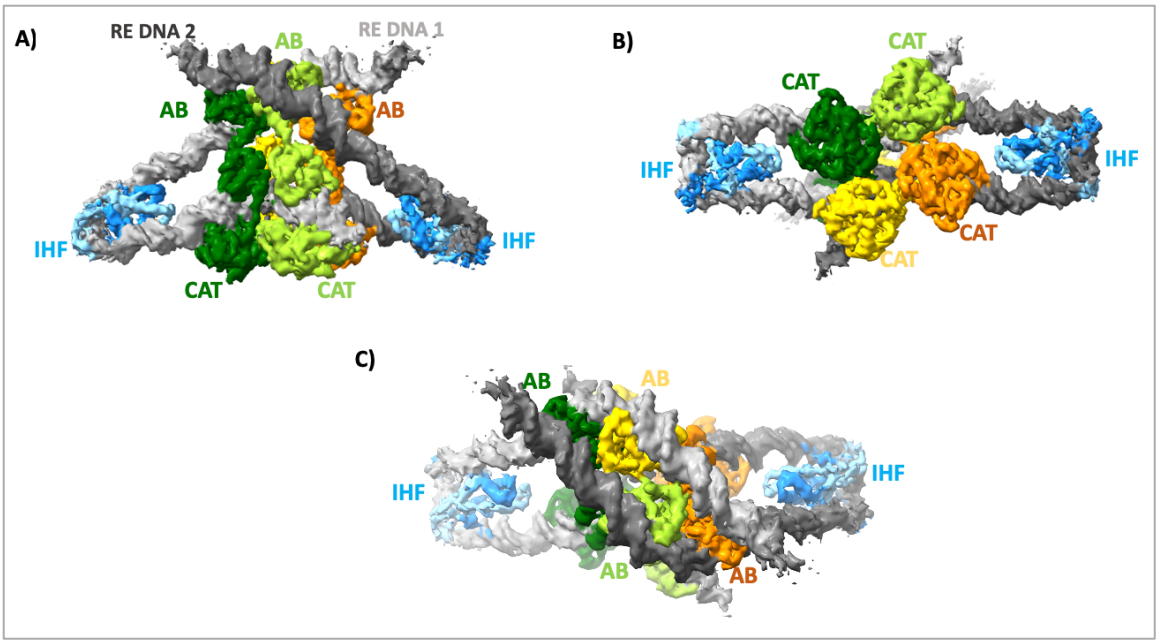


Figure 2-33: The segmented cryo-EM map of the RE-RE Synaptic Complex. The two RE DNA molecules are shown in dark and light grey, respectively. The IHF heterodimer is depicted in two shades of blue. The four Int_{GISul2} monomers are shown in yellow, orange, dark and light green. A) Side view of the complex showing two RE complexes synapsed via interactions of the CAT and AB domains. B) Bottom view highlights the tetrameric CAT assembly. C) The AB filament is shown from the top view.

2.3.3.1 Comparison with the Pre-Synaptic RE Complex

In the cryo-EM map, the Int_{GISul2} core of the complex has better resolution when compared to the IHF and its bound DNA. To obtain a better resolved reconstruction for the Int_{GISul2} core, local refinement was performed with a mask including only the Int_{GISul2} tetramer and its directly bound DNA segments, which resulted in a map at 3.6 Å resolution.

This structure revealed a complete tetrameric organization of Int_{GISul2} for the first time. When compared to the pre-synaptic RE-complex, all CBs bind to core DNA sites forming closed clamps around the core repeats together with their respective CAT domains (Figure 2-33). This can be achieved by AB of the previously dislocated subunit to bind to arm site on the other RE DNA molecule. To exemplify based on Figure 2-33, on RE DNA 1, the CB of the light green colored Int_{GISul2} subunit is expected to be dislocated in the pre-synaptic version of the complex. In the tetrameric form, the AB of this subunit binds to the arm site on RE DNA 2 (which is in antiparallel orientation), allowing its CB to be located on the core site (Figure 2-33 A). As a result, in the tetrameric arrangement, ABs of an Int_{GISul2} pair bind to different DNA molecules, whereas CB-CAT bind to the same core site.

In the tetrameric arrangement of CATs, both core DNAs are bent at the crossover region, allowing the formation of almost square planar configuration of CATs. In the EM density,

while the uncleaved DNA strand at the crossover site can be traced clearly, the nicked strand is not visible at the crossover region.

When the tetrameric Int_{GISul2} structure in the RE-RE synapse was aligned with the Int_{GISul2} dimer in the pre-synaptic RE complex, the positioning of CAT and CB domains are identical on the core repeat inside the RE transposon end (Figure 2-34 A). However, the localization of the flank DNA segment and the bound CB and CAT domains are markedly different. In the synaptic complex CB is bound to core site, while in the pre-synaptic complex CB is dislocated. The different positioning of flank DNA bound CAT is due to 60° bending of the core site DNA, pivoting at the crossover region (Figure 2-34 B). Due to the DNA bending, two neighboring CAT domains interact at a different angle to each other, and different residues are exposed to their interaction interface, when compared to the pre-synaptic RE complex.

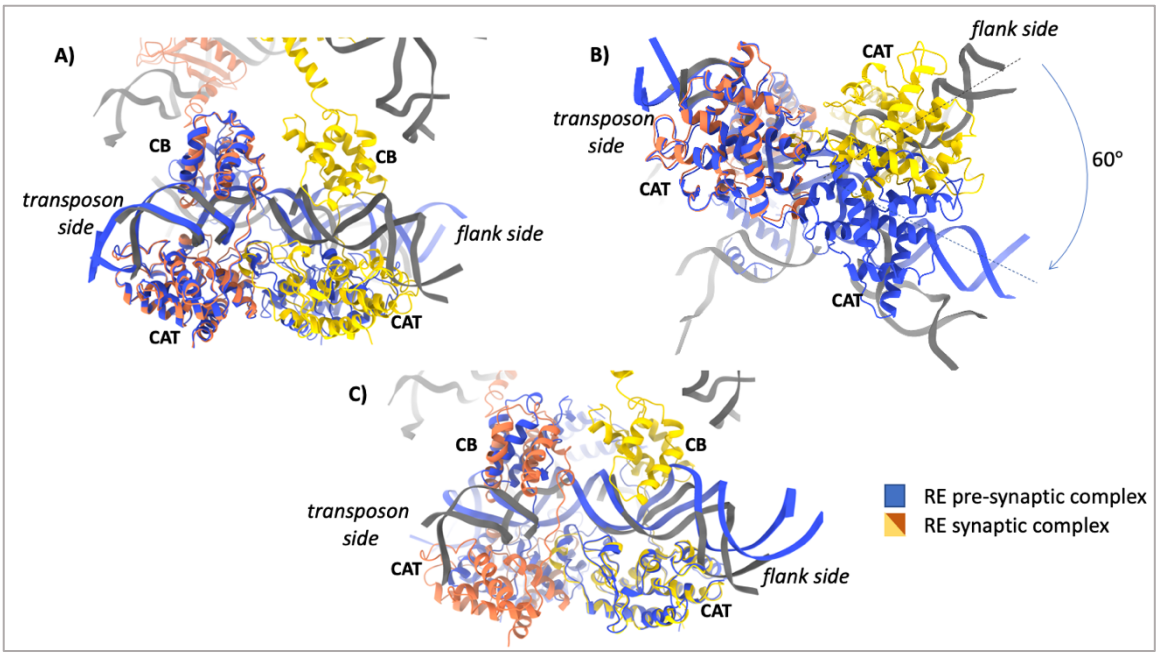


Figure 2-34: The comparison of CB-CAT positioning in the pre-synaptic and synaptic RE complexes. The pre-synaptic RE complex is shown in blue, while the synaptic complex is shown in orange and yellow, with DNA in grey. A) Superposition of the Int_{GISul2} subunit bound to the core DNA repeat in the transposon end shows identical positioning of the CAT and CB domains. B) Bottom view of the superposition from A) shows that the core site DNA is bent by 60° when a tetrameric synaptic complex is formed. C) Superposition of the Int_{GISul2} subunit bound to the core repeat in the RE flank shows similar positioning of the CAT domains. While CB is dislocated from the DNA at this site in the pre-synaptic complex, it binds tightly after the synapse. The comparison also shows DNA bending in the flank DNA, upon CB binding in the synaptic complex.

As a consequence of the change in CAT orientation in the RE-RE synaptic structure, the enzyme active sites are formed quite differently than in the pre-synaptic RE complex.

Although the *trans* active site assembly is preserved, the pre-synaptic complex is in an inactive state, whereas the synaptic complex can be accepted as an active state due to the positioning of nucleophile tyrosine (Y373) (Figure 2-35 A). In the synaptic complex Y373 is positioned close to scissile phosphate (1.95 Å distance) in two subunits of the Int_{GISul2} tetramer, when compared to the pre-synaptic RE complex (8 Å distance) (Figure 2-35). This change is achieved with the help of several factors. First, due to the presence of the nick, DNA is more flexible in the crossover region. This allows the DNA to bend upon the formation of a CAT tetramer by approximately 60°. Since the CAT domains are consequently rotated to a different angle with respect to each other, the M-helix carrying tyrosine also changes its position and inserts into the active site formed in the receiving CAT domain. Finally, another factor which affects the positioning of the catalytic Y373 with respect to the cleavage site in the DNA, is the binding of the CB domain to both core repeats in the recombination site. This CB binding tilts DNA towards CAT and brings the M-helix towards the scissile phosphate (Figure 2-34 C).

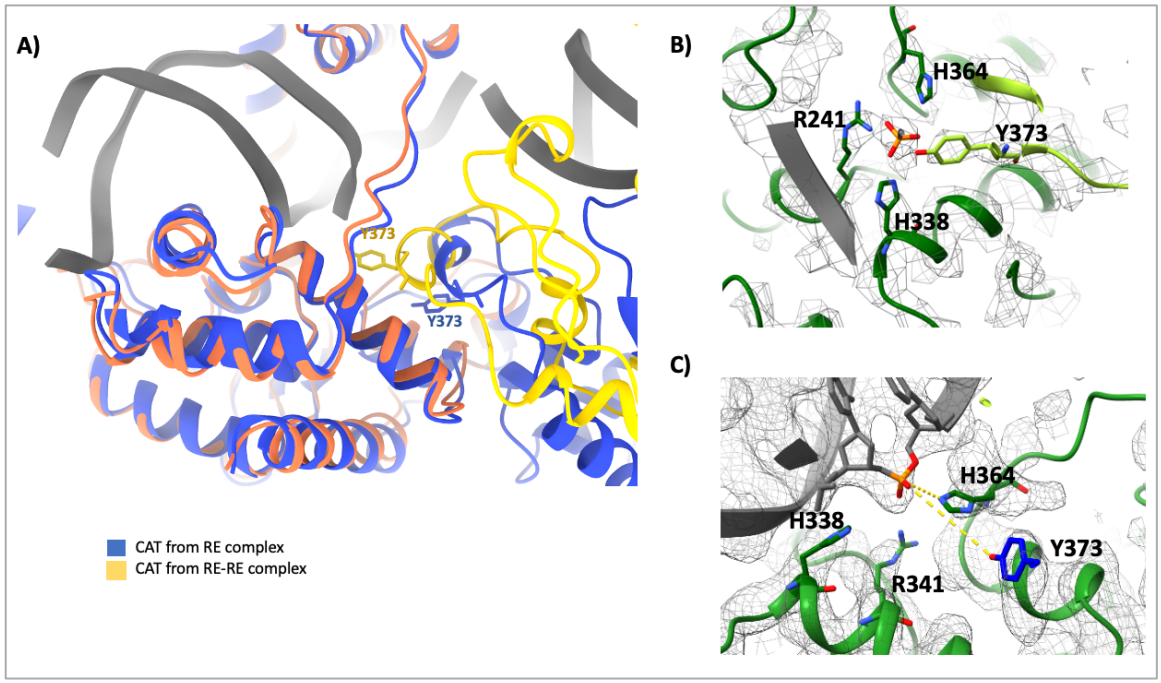


Figure 2-35: Int_{GISul2} active site formation in trans. A) Comparison of the active sites in the pre-synaptic (in blue) and synaptic complex (in yellow). The positioning of the catalytic tyrosine (Y373) is shown in sticks representation in both structures. B) Zoom-up of the active site formed in the RE-RE synaptic complex with the residues H364, R241, H338 and Y373, and the scissile phosphate is shown in the middle. The cryo-EM map is depicted as grey mesh. C) The active site in the RE pre-synaptic complex.

Due to changed position of the M-helix, the interactions made to stabilize it are also different in the two RE complexes. In the pre-synaptic complex, Y373 interacts with H349 of the receiving CAT by aromatic stacking in the inactive state. In the synaptic complex, as

Y373 comes closer to the scissile phosphate, H349 interacts with Y378 and contributes to the active positioning of the M-helix.

2.3.3.2 The Organization of Int_{GISul2} Tetramer

The tetrameric arrangement of the CAT domains is held together by a cyclic exchange of the M-helices, which carry the catalytic tyrosine residue (Y373) (Figure 2-34 D). The M-helices that are swapped to a CAT domain located on the same DNA strand, approach the active site of their partner subunit and donate their Y373 to complete its catalytic centre. In turn, the M-helices that are swapped to a CAT located on the other DNA across the synapse, are further away and cannot reach the respective active sites. The EM density of these distant M-helices cannot be fitted well with a helical model, perhaps indicating a partially unfolded structure. The above observations are consistent with stepwise activation of two Int subunits at once and suggest that the first set of catalytic sites (with the visible and closely located M-helix) are in an active conformation, while the other two are inactive. This cyclic exchange of the M-helices appears to be the main interaction holding the CAT tetramer together.

In the case of the Flp recombinase, it was also observed that the M-helix was exchanged between CAT domains in a cyclic fashion (Figure 2-34 C). As for *GISul2*, two M-helices approach their neighboring CAT domains closely and contribute to the active site formation, whereas the other two are unfolded. In the other known TR tetramer structures (Cre, λ, Xer), the catalytic tyrosine carrying helices are not exchanged between different recombinase subunits, consistent with their *cis* active site arrangement. Instead, these proteins exchange their subsequent, very C-terminal segments in a similar cyclic fashion, which again allows to hold the CAT tetramer together and regulate the activation of different protein monomers in the tetramer by differentially localizing the connected the M-helices in the complex (Figure 2-34).

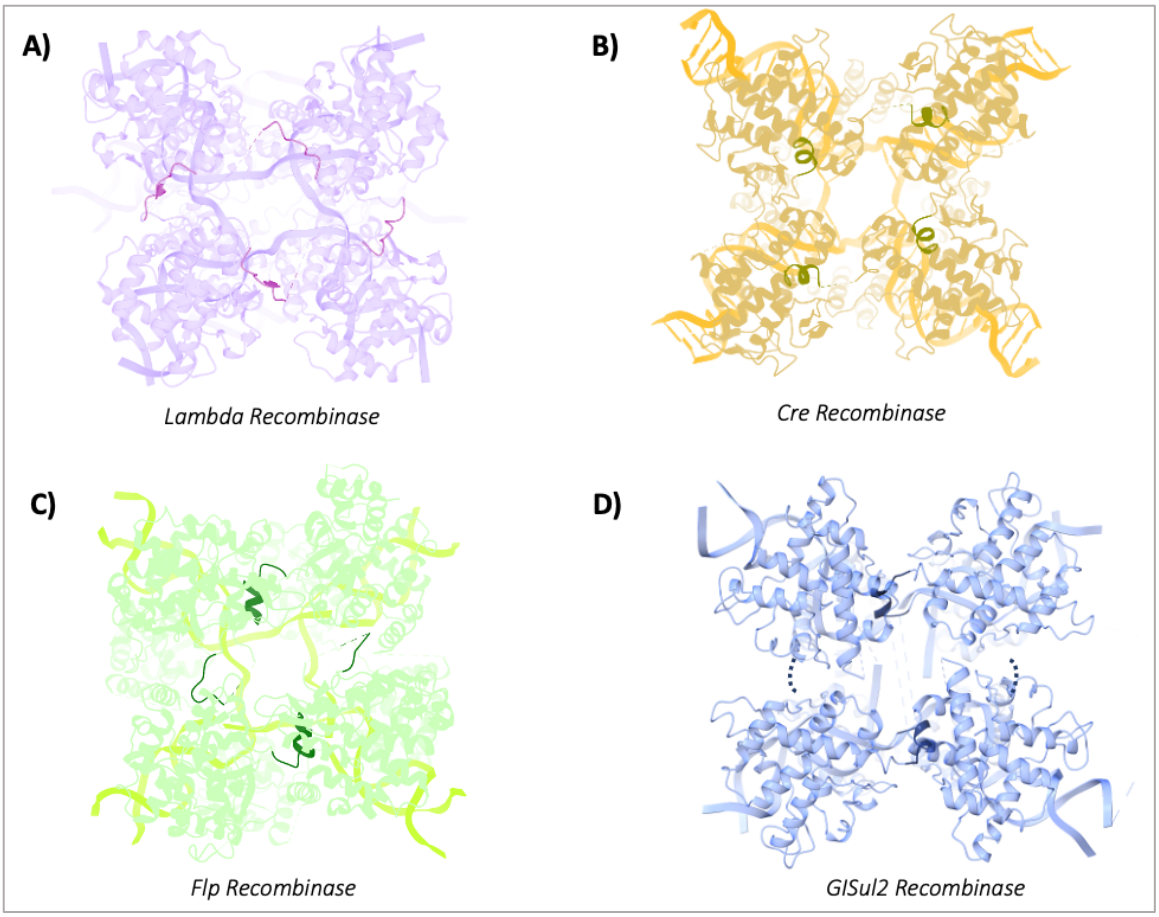


Figure 2-34: Tetramer formation of previously characterized TRs. The swapped protein segments are highlighted with stronger colors. Cyclic exchange of the C-terminal segments is shown for A) the λ integrase (the swapped segment is a short β-strand) and B) the Cre recombinase (final α-helix is swapped). Cyclic exchange of the M-helices is shown for C) the Flp recombinase and D) the *GISul2* integrase. While this helix is folded in two subunits and contributes to the formation of active catalytic centers, two inactive subunits receive unfolded M-helices.

In addition to the CAT and AB interactions, the Int_{GISul2} tetramer is held together by additional cross-domain interactions. Strong interactions are formed between the AB and CB domains of two Int monomers sitting on different DNA molecules (Figure 2-35). These include interactions between R49(AB)-E142(CB) and R91(AB)-D147(CB), which stabilize tetramer formation and may guide the positioning of the CB domains in well-defined locations, as described above within all our structures.

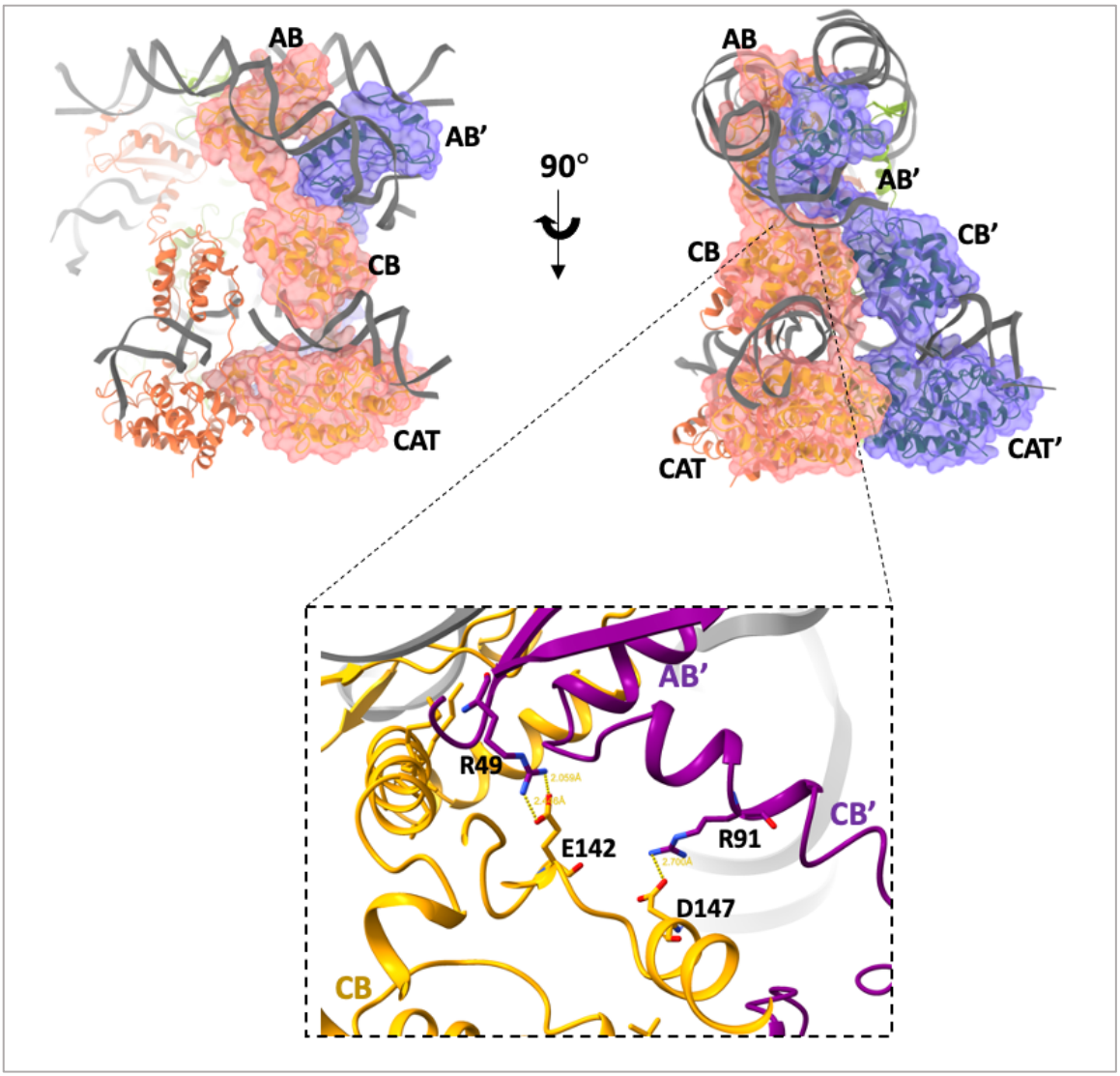


Figure 2-35: Securing the Int_{GISul2} Tetramer with AB-CB Interactions. The Int_{GISul2} subunits bound to two different RE DNA molecules are shown in orange and purple. While CAT and CB of these subunits bind separate DNA molecules, their AB domains interact with arm sites located on the same DNA. Residues involved in the interaction are shown in the zoomed view (sticks).

2.3.3.3 Conclusion

The RE-RE synaptic complex revealed a tetrameric Int_{GISul2} assembled on two RE DNA fragments, which are bent by IHF. The RE-RE synapsis was not expected; however, the longer RE DNA design used in these experiments facilitated AB filament formation and enabled synapsis of two identical RE transposon ends. In physiological conditions, this complex probably does not form because only one RE and one LE of the *GISul2* element are present in one cell at a time. Still this complex provides many ideas about how the physiological tetrameric assembly of Int_{GISul2} may look like. Although it was previously observed how four AB domains can bring together two arm DNA segments in a filament in

the LE complex, formation of an active CAT tetramer was observed for the first time in this structure, providing unique insights into the organization of a putative functional synaptic complex.

To reconstitute a stable synaptic complex, different DNA designs were tested, and found that a "suicide" design with nicked DNA led to successful synaptic complex formation. This suggests that the nicked DNA provided a flexible substrate, which can be more easily bent and allow the formation of a stable tetrameric Int_{GISul2} assembly. The nicked DNA was designed to trap the complex after first strand cleavage. At the center of the tetramer, the uncleaved DNA strand has a continuous density, while the nicked strand is not visible at the crossover site.

Cyclic exchange of protein segments is a frequently employed mechanism to regulate active site formation and catalytic activation in TR assemblies. In the case of Cre, swapped C-terminal α-helices connect monomers in a tetrameric assembly and coordinate the activation of specific protein subunits during recombination (Martin et al., 2002). Depending on the distance between neighboring CAT domains, the swapped helices induce somewhat different conformations in the different catalytic centers around the tetramer. In the active Cre subunits, the conformation of the exchanged C-terminal helix allows the catalytic tyrosine to cleave DNA, while in the inactive subunits it prevents DNA cleavage. This mechanism ensures that only two catalytic centers are active at a time in the tetrameric assembly (called "half-of-the-sites" activity) and allows to regulate the order of strand cleavage and ligation reactions during the recombination process (Grindley et al., 2006; Martin et al., 2002). For CTn integrase Int_{Tn1549}, it was shown that C- terminal helix exchange creates an auto-inhibited state in dimeric pre-synaptic Int_{Tn1549} – DNA complexes, and it is predicted to relieve inhibition once an active tetramer is assembled (Rubio-Cosials et al., 2018). In the case of Int_{GISul2}, exchanged M-helices assist tetramer formation and ensure half-of-the-sites activity (c.f. discussion above regarding the role of the C-terminal helix in Cre). The M-helices that are donated between CAT domains bound to the same core DNA molecule approach the catalytic center of the receiving CAT, producing a complete functional active site. However, the M-helices that are swapped between two CATs located on two separate core DNA sites cannot properly reach their receiving CATs, leaving their active sites inactive. After the first DNA cleavage and strand exchange, the tetramer is expected to undergo conformational rearrangement and isomerization, where the two distant CATs can come closer, allowing the M-helix to reach and assemble the active sites in the CATs that were previously inactive. This is an elegant mechanism to ensure ordered progression of the recombination reactions during transposon excision and integration.

2.3.4 The Cryo-EM Map of RE-LE Synaptic Complex

The cryo-EM data processing revealed that there are two different complexes formed in the experiment designed to trap a synaptic Int_{GISul2} -DNA complex. In the previous section, I have discussed the RE-RE synaptic complex, and in this chapter, I will talk about the details of the RE-LE synaptic complex.

In this structure, the synapsis of one RE and one LE DNA was captured. To build a structural model of the complex, the structures of previously solved pre-synaptic RE and LE complex were used and the CB and CAT domains of Int_{GISul2} were separately fitted. The overall DNA conformations of RE and LE are similar to their presynaptic versions: LE DNA is widely looped by seven Xis molecules, and RE DNA is sharply bent by IHF (Figure 2-38 B). All four ends of the two DNA fragments are connected and bridged by an Int_{GISul2} tetramer. The synapsis of RE and LE were driven by all Int_{GISul2} domains. The four AB domains form a stable filament, which connects the arm sites of RE and LE together. Simultaneously, CB and CAT domains bridge the core sites of the two transposon ends.

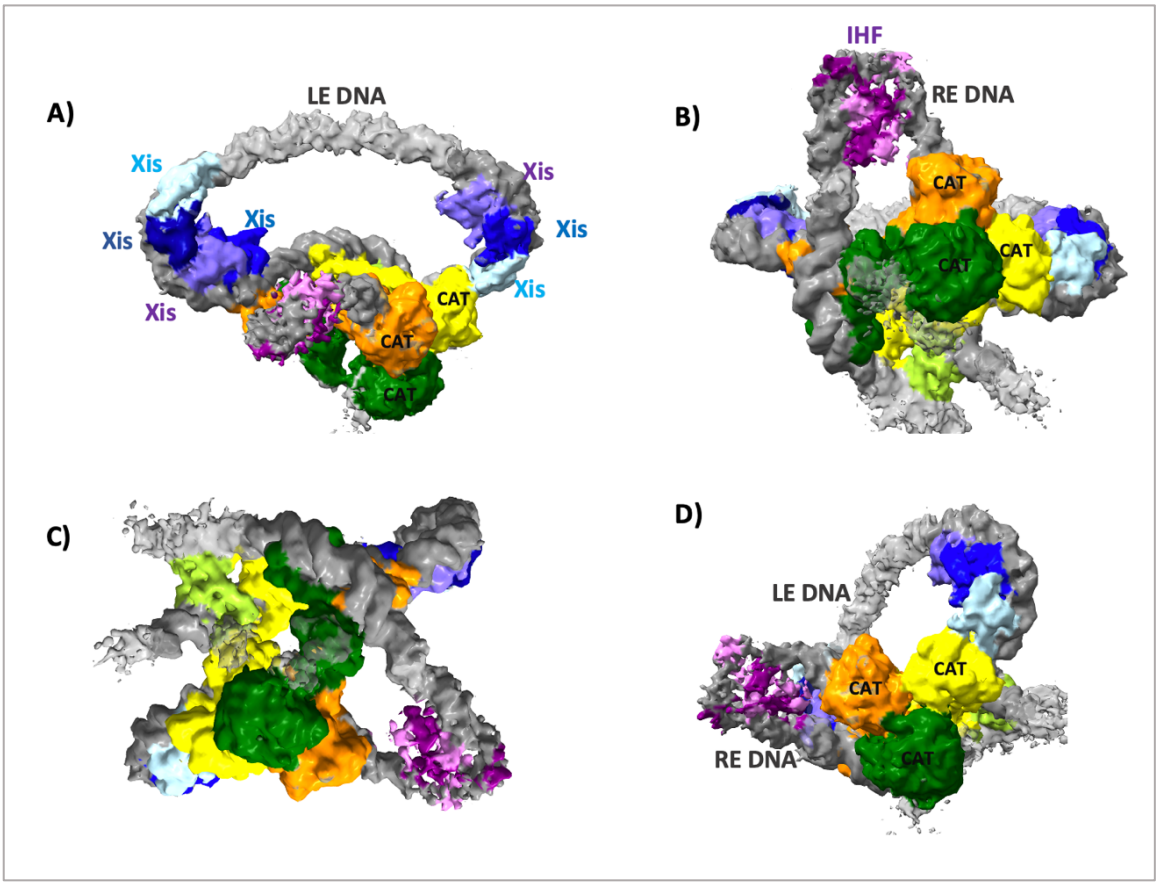


Figure 2-38: Four Views of the Segmented Cryo-EM Map of the RE-LE Synaptic Complex. A) First view shows that the LE DNA (in grey) is looped by Xis molecules (in blue shades) in two patches with three and four Xis proteins, respectively. B) View highlighting the RE part of the synaptic complex, where RE DNA is colored in

grey and IHF is in purple. C) Side view of the Int_{GISul2} tetramer (orange, yellow, dark and light green subunits) shows LE and RE connected by CATs trimerization at the bottom and by the AB filament on the top. D) Bottom view of the Int_{GISul2} assembly. Three CAT domains form a triagonal arrangement; he fourth CAT domain is not visible in the EM map.

Although, there are four AB domains forming the AB filament, surprisingly only three CB and CAT domains can be detected in the EM density. Two CB-CAT units bind to the core repeats inside the transposon ends in RE and LE (Figure 2-38 D, orange and yellow), respectively, and one CB-CAT segment binds to the core repeat in flank DNA. Due to non-continuous EM density at this region, it is not clear which flank DNA (i.e. the one from the RE or LE) observed in the structure. The reason why the fourth CB-CAT segment is not visible in the EM map can be due to a lack of stable binding to the respective flank DNA, resulting in flexibility of the corresponding protein and DNA parts. Since the flank DNA density is not well resolved, a DNA model was not built for that DNA fragment (Figure 2-39).

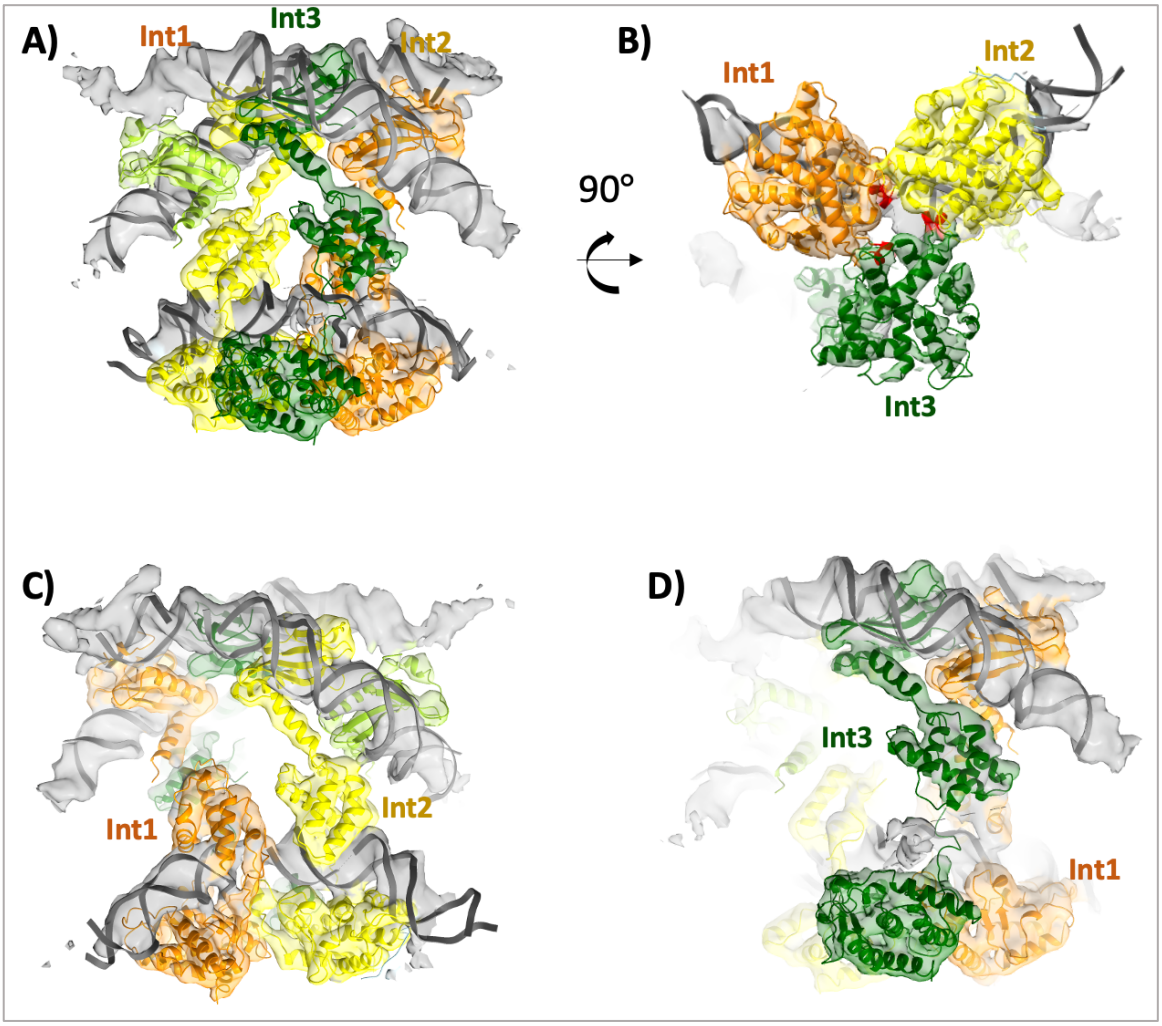


Figure 2-39: Structure of the Int_{GISul2} assembly in the RE-LE synapse. A) Crystal structures of Int_{GISul2} were fitted in the cryo-EM map (transparent) and the resulting model is shown by cartoon representation. While there are four AB domains forming the AB filament, only three copies of the CB and CAT domains are visible

in the EM density. The Int_{GISul2} subunits are numbered as Int1 (orange), Int2 (yellow) and Int3 (dark green). B) Zoom-up of the CAT domains. The exchanged M-helices are highlighted with red. C) shows the subunit Int 1 and Int 2, D) shows Int3 in focus.

In the RE-LE synaptic complex, the positioning of Int_{GISul2} on the flank DNA region (Int3) is different when compared to other Int subunits (Int1 and Int2) (Figure 2-39). Int1 and Int2 located at RE and LE core repeats, respectively, have their CAT and CB positioned on DNA by forming a C-clamp shape as in the RE-RE complex (Figure 2-39 C). While the CB of Int3 which is positioned on flank DNA is not located close to DNA density as in the case of Int1 and Int2 (Figure 2-39 D).

Like in the RE-RE synaptic complex, in RE-LE synaptic complex the Int_{GISul2} tetramer is held together by AB filament formation, cross-domain interactions and finally M-helix exchange across CAT domains. The trigonal CAT structure assembles in a way that a cyclic exchange of the M-helices maintained (Figure 2-39 B). In the RE-RE synaptic complex, this mechanism created two active catalytic centers, while the other two were inactive due to more distant positioning of the affected CAT domains with respect to each other. In the case of trimeric CAT assembly, all three domains are located in equal distance (Figure 2-39 B), which can allow the M-helices to be donated in active fashion in all three catalytic sites.

2.3.4.1 Comparison of the RE-LE and RE-RE Synaptic Complexes

Although the RE-RE and the RE-LE complexes contain different transposon ends, the way how two DNA molecules are brought together by Int_{GISul2} is similar in both cases. When the two EM maps are overlayed, the AB filament formation at the arm DNA sites looks identical in the two complexes (Figure 2-40 C). However, there is a striking difference in the organization of the CAT and CB domains (Figure 2-40 A). While in the RE-RE complex four CATs assemble on two core DNA sites with pseudo four-fold symmetry, in the RE-LE complex three CAT domains form a three-fold symmetric triangle shape. While two CAT domains (CAT1 and CAT2) are similarly located in both structures, the third one (CAT3) is located in the middle between the other two CATs in the RE-LE complex (Figure 2-40 A).

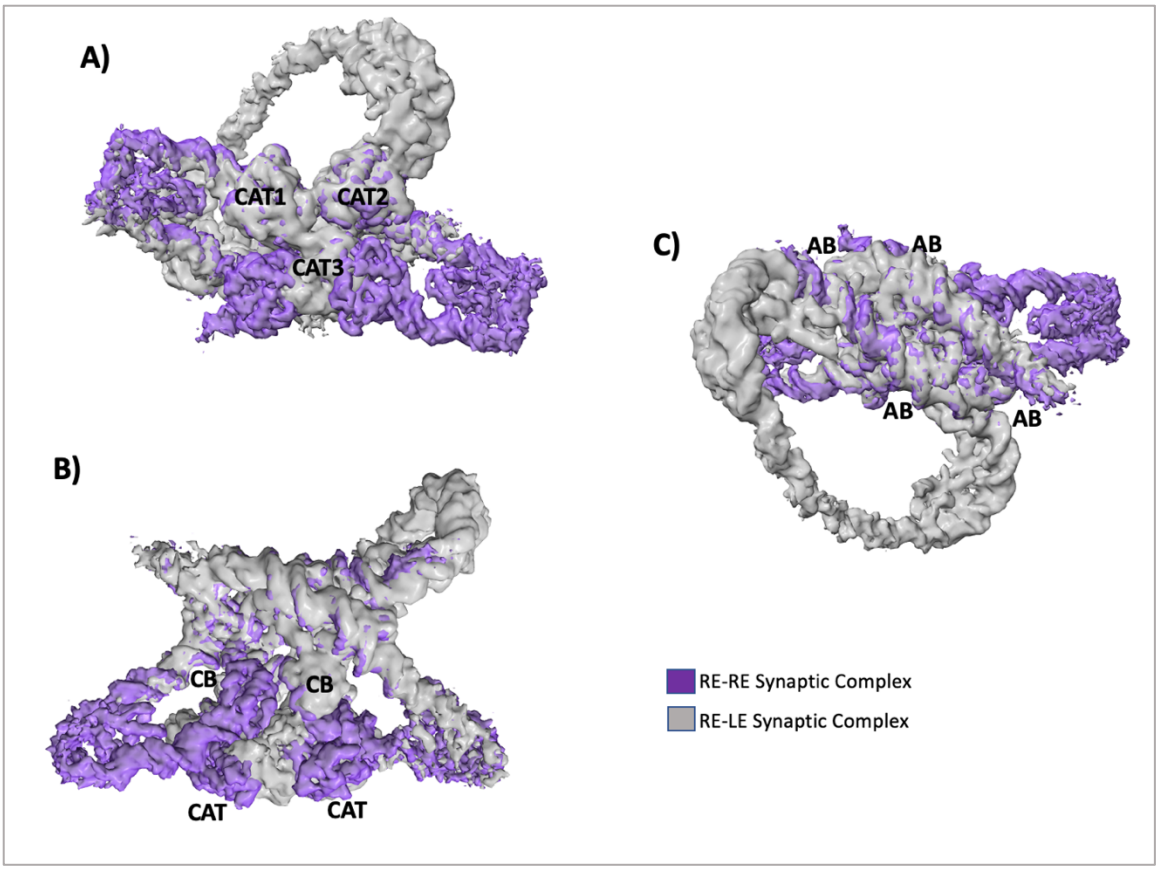


Figure 2-40: Superposition of the RE-RE and RE-LE synaptic complexes. RE-RE synaptic complex is shown in purple, while RE-LE complex is shown in grey. A), B) and C) are showing the overlay from different views. A) Bottom view of the overlayed complexes showing the change in CAT organization. While CAT1 and CAT2 are positioned identical in both structures, CAT3 is positioned in between the third and fourth CATs present in the RE-RE complex. B) Side view of the complexes. C) Top view showing the arm filament formation, which is identical in both structures.

While CAT1 and CAT2 are positioned in the same way in both complexes, the DNA sites that they bind to are different. In the RE-RE complex, CAT1 and CAT2 interact with core repeats in the transposon end and flanking DNA of the same RE DNA fragment, respectively. In turn, in the RE-LE complex, CAT1 and CAT2 are placed on the core sites inside the transposon ends of RE and LE DNA, respectively. The flank sequence that was observed can either belong to RE or LE; the connecting DNA density is not clear in the EM map.

In the suicide oligo design (Section 2.3.1), the nick in the crossover region was placed on the top DNA strand, based on the hypothesis that first cleavage occurs at the top strand (Table 2-1). This nick was designed to trap the reaction after first strand cleavage. In addition to trapping the reaction, the presence of a nick provides additional flexibility to the DNA substrate at the crossover region, which likely promotes binding and bending by

CAT and CB domains. Depending on the position of the nick, the angle and direction of DNA bending can be different, which can result in a distinct complex formation.

For the Int_{GISul2} system, which DNA strand is cleaved first at the beginning of recombination has not been clearly demonstrated yet. Thus, to rule out the possibility that false crossover design resulted in artificial formation of the trimeric CAT assembly, we have designed DNA oligos with different nick positions and mutations at the crossover region (Table 2-1). The design that we have used for the formation of the synaptic complexes assumes the first strand cleavage occurs at the top strand (Table 2-1, Design 3). To test the possibility of first bottom strand cleavage, suicide oligos with a nick design at the bottom strand was designed (Table 2-1, Design 4). As a last design, the DNA sequence at the crossover site of LE has exchanged with that of RE, to imitate the crossover site architecture of the tetrameric RE-RE synaptic complex to rule out the effect of DNA sequence in the organization of CATs.

	DNA Designs	LE crossover	RE crossover
Design 3	Nick at the top strand	5′G GAGTGG3′ 3′G---AGACC5′	5′G CTCTGG3′ 3′C---TCACC5′
Design 4	Nick at the bottom strand	5′GAGTG---G3′ 3′GAGACG C5′	5′GAGTG---G3′ 3′CTCACG C5′
Design 5	RE crossover design	5′G CTCTGG3′ 3′C---TCACC5′	5′G CTCTGG3′ 3′C---TCACC5′

Table 2-1: DNA Designs with different crossover sequences. The red color highlights mutated nucleotides, while underlined nucleotides are showing the extra ones to replace the nucleotide which will be cleaved.

With these DNA designs, we have followed the same protocol for *in vitro* complex reconstitution and grid preparation for cryo-EM data collection. However, all of them resulted in a similar RE-LE synaptic complex architecture with a trigonal CAT arrangement. Thus, this assembly appears to be independent of crossover DNA design and the position of the nick.

2.3.4.2 Trimeric Assemblies in Other TRs

Previous studies of TR mechanisms have revealed that functional recombination complexes mostly assemble as tetramers. However, there are several examples showing that trimeric assemblies can also perform junction resolution. Such assemblies were discovered on three-way DNA junctions called as Y junctions (YJ), which are shown to be natural intermediates of DNA break repair or retroviral integration (Seeman and Kallenbach, 1994). However, whether YJs can be functional transitional intermediates in TR recombination is still not known.

For example, the Cre recombinase was shown to bind and resolve a YJ structure by forming a trimeric arrangement (Woods et al., 2001). The crystal structure of the Cre - YJ complex was also solved, showing a Cre trimer arranged on the three YJ arms in a three-fold symmetric fashion (Woods et al., 2001). When this trimeric Cre structure is aligned with the CAT trimer of our *GISul2* RE-LE synaptic complex, the overall assembly looks similar (Figure 2-41). The subtle differences can be observed in the relative positing of protein subunits; while CAT1 and CAT2 positioning are similar to corresponding Cre domains, CAT3 is shifted. This can result from the different DNA architectures used in the two structures. While in the case of Cre, YJ is formed *in vitro*, in the case of *GISul2,* in addition to three DNA segments that are visible in the EM map, there is a fourth flank DNA still connected.

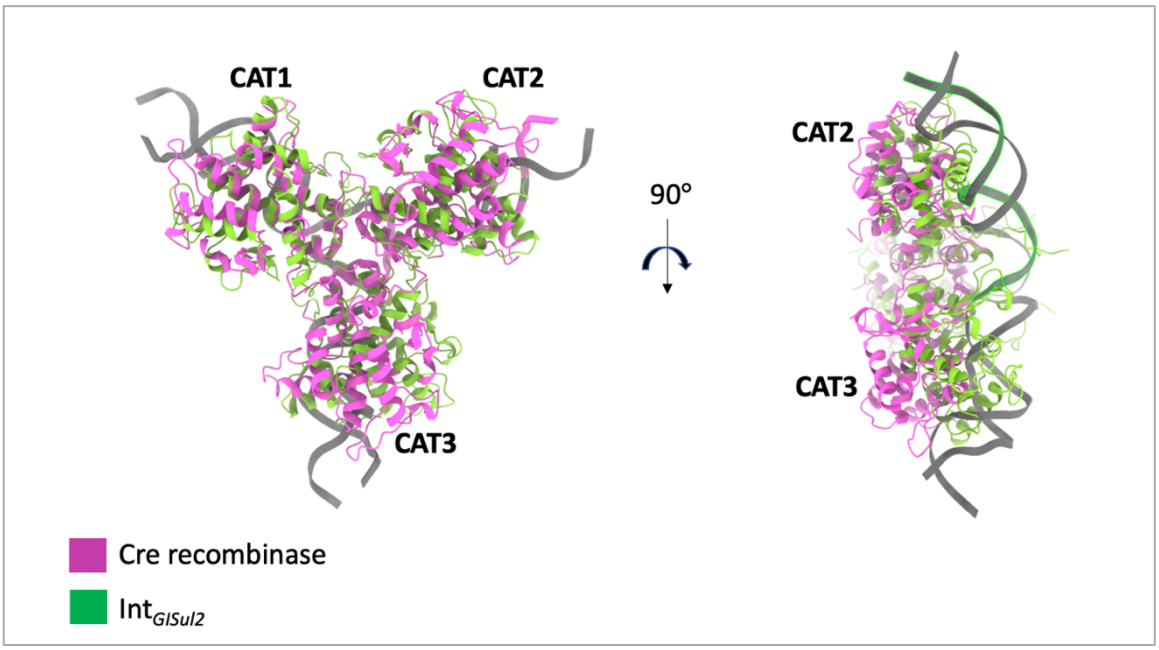


Figure 2-41: Comparison of the structure of a trimeric Cre recombinase assembly (pink) with the CAT domains of Int$_{GISul2}$ in the RE-LE synapse (green). DNA shown in grey belongs to Cre trimer (PDB ID 14FF, Woods et al., 2001). While CAT1 and CAT2 positioning is similar to its Cre counterparts, CAT3 is shifted.

In addition to Cre, Int$_\lambda$ was also shown to resolve YJs into a hairpin and a dsDNA product, albeit with a less efficiency when compared to HJ resolution (Nunes-Duby et al., 1997). For the Flp recombinase, the existence of a trimeric assembly was also proven biochemically (Lee et al., 1996). Flp can readily disassemble a synthetic three-armed DNA junction into a DNA duplex and a hairpin product, which requires the simultaneous activity of adjacent Flp protomers. It was shown that Flp resolves YJ more efficiently than Cre and Int$_\lambda$ (Conway et al., 2003). When compared to *cis*-acting TRs, *trans*-acting TRs are thought to be more prone to trimer formation due to the flexible linkers between M- and N- helix, which can allow the formation of different assemblies with different geometries.

All in all, these findings indicate that the trimeric CAT assembly can be a functional intermediate of recombination complexes, which may specialize to execute distinct steps of the reaction.

2.3.4.3 Conclusion

The RE-LE synaptic complex revealed a tetrameric Int$_{GISul2}$ bringing together one RE and one LE DNA, which are bent by IHF and Xis, respectively. While there are four ABs gluing the two DNA molecules, surprisingly, only three CAT domains are visible in the EM density.

It was previously shown that a mini transposon that contains LE DNA corresponding to the minimal DNA design, with only two arm sites, can still be excised from a plasmid *in vitro*. This suggests that the assembly that was observed in the RE-LE synaptic complex may be a physiological intermediate during excision. Catalytic activity of threefold CAT assemblies has been also shown for related TRs (Cre, Int$_\lambda$ and Flp), which they can bind and resolve three-way YJ DNA molecules. It was also proposed that *trans*-acting Flp is more efficient in resolving YJ than its *cis*-acting counterparts, due to flexible linkers connecting the exchanged M-helix. The analogous *trans* active site assembly and flexible linkers may similarly support trimer activity in the case of Int$_{GISul2}$.

Comparison of RE-RE and RE-LE synaptic complexes highlighted an identical AB filament structure, regardless of the DNA molecules used. This means that the AB filament is a stable structure which acts as the anchoring point for the assembly. However, the two structures contain markedly different arrangements in their catalytic cores. Between the DNA molecules used in the two structures the main difference is in the distance between the arm and core sites and the way that they are bent. This means that by changing the relative positioning of arm and core sites, one can change the positioning of CAT domains. For the current RE-LE complex structure, we used a LE design that includes the two nearest arm sites to the core sites at the transposon boundary. With this, we could show that this LE

design is functional in an in vitro excision assay. This means that the included arm sites in LE play an essential role in recombination and the observed trimeric CAT assembly in our structure may function in some stage of the reaction. However with the limits of the resolution, we don't know in which step of the process this complex is used. In any case, changing the architecture of the Int organization appears to be a good way of securing that the chemical reactions are directed towards desired products and do not go backwards. If we assume that using the first two LE arms Int_{GISul2} creates a trimeric CAT assembly, which performs the first strand cleavage, subsequent second strand cleavage, may involve the usage of other arm sites found in the LE, which could trigger a different CAT arrangement. The next pair of arm sites found in the LE DNA are oriented in reverse direction compared to the ones that were included in our current design (Figure 2-21 A). By using these more distant arm sites, a different synaptic complex may be assembled, in which the LE DNA arms could be positioned differently, perhaps promoting the assembly of four CAT domains on the respective RE core and LE core sites. Differential arm binding during different steps of the reaction may offer an exciting mechanism for coordinating the order of strand exchange reactions.

3 GENERAL DISCUSSION AND CONCLUSIONS

3.1 Summary of the Experimental Findings

The work that I have presented in this thesis investigated the excision of the *GISul2* transposon. For this purpose, we have structurally characterized various complexes involved in the different stages of transposon excision to understand how higher order complex formation at the transposon end helps to regulate its movement across bacterial genomes.

Conjugative transposons (CTns) play a major role in the spread of antibiotic resistance genes. During the excision and integration of an element, site-specific recombination reactions are performed by a transposon-encoded Integrase (Int) enzyme. There is significant biochemical and some structural information available about CTns and their Int proteins, which provide a basic understanding of the recombination reaction. However, CTn transposition does not involve only Int but requires rather complex machinery, which employs multiple transposon- and host-encoded factors assembled in various higher order protein-DNA complexes. Different stages of the excision and integration reactions of CTns require the formation of distinct higher order nucleoprotein complexes, which provides a new level of regulation to the recombination process both spatially and temporally. The only structurally characterized example of a higher order recombination complex belongs to the bacteriophage λ system, in which site-specific recombination of the phage genome ends is catalyzed by a dedicated integrase (Int_{λ}) protein during lysogeny and invasion. For this system, a low resolution cryo-EM map with Int_{λ} bound to a complex Holliday Junction DNA intermediate, representing a central step of the excision reaction, has been reported (Laxmikanthan et al., 2016). Apart from that, biochemical data indicate that the integration and excision of the phage genome are tightly regulated in response to a variety of physiological and environmental signals, by employing distinct higher order complexes in different stages of the reaction. Although, CTns and phages encode similar proteins to drive their mobilization, it was not known how closely the higher order complex formation of these two systems are related.

To decipher the principles of higher order complex assembly in CTn movement, we chose the *GISul2* element as a model system. *GISul2* is a sulphonamide resistance (*sul2*) carrying transposon, which was first identified in gamma-proteobacterial species (Nigro & Hall, 2011). In a recent study, the integrase encoded on *GISul2*, Int_{GISul2} was shown to be the most widespread TR-family integrase in Gram-negative bacteria (G. Smyshlyaev, unpublished data). The broad distribution of the element indicates a highly effective

mechanism for transposition and its regulation, which allows mobilization in many different bacterial hosts. Previously solved crystal structures of Int_{GISul2} in complex with short transposon DNAs segments shed light into the structural principles of its activity (G. Smyshlyaev, unpublished data). These structures improved the understanding about Int_{GISul2} organization and DNA binding; however, it remained unknown how accessory proteins coordinate and regulate higher order complex formation at the transposon ends. To see the whole picture, we aimed to structurally characterize DNA-protein complexes formed on native transposon ends during excision of the element.

We started with the characterization of the complexes formed on individual transposon ends prior to synapsis. In the first part of my thesis, we successfully determined the structure of the right transposon end (RE) complex. This complex is formed on RE DNA, which is bent by the integration host factor (IHF) protein to ~160°, allowing Int_{GISul2} to simultaneously bind and bridge arm and core DNA sites. The formation of the RE complex strictly requires the presence of the host-encoded accessory protein IHF, indicating a strong coordination between host cell state and CTn mobilization (see further discussion below). In the structure, while two Int_{GISul2} subunits are bound to pairs of core repeats and arm DNA sites, due to the inverted orientation of one core repeat, one CB domain is dislocated from its core site. This intriguing dislocation is enforced by stably bound arm sites in the complex and we propose that it plays a role in the next step of recombination. Here the RE needs to synapse with the LE to form an active recombination complex, and the dislocated CB can scan DNA for another core site to form a suitable synaptic complex. Importantly, CB dislocation precludes sequence specific recognition of the flank DNA sequence and allows the RE complex to form regardless of the site of transposon insertion in diverse genomes.

In the second part of the thesis, we presented the structural characterization of the left transposon end (LE) complex. Based on *in vitro* excision assays the shortest functional LE segment was identified, and the LE complex was reconstituted with this DNA sequence. The resulting cryo-EM map revealed that seven excisionase (Xis) molecules bend the LE DNA to create a loop that is 'connected' by integrase molecules at the ends. In contrast to the sharp bend in the RE DNA, Xis bending results in a widely looped DNA shape in the LE complex. LE complex formation strictly depends on DNA bending by the transposon-encoded Xis molecules. Xis is generally recognized as a directionality factor for recombination, which promotes excision, while it inhibits integration. Therefore, LE is expected to form distinct higher order complexes during excision and integration, with the presence or absence of Xis, respectively. Surprisingly, in the LE complex structure two DNA fragments were captured, one fully and one partially, possibly revealing the first step of synaptic complex formation. One of two Int_{GISul2} monomers bound to the LE core repeat can simultaneously reach and bind the arm site from the same DNA molecule, whereas the second monomer recruits another LE DNA fragment and binds to its arm DNA site,

triggering the formation of a filamentous AB-DNA structure. This AB filament glues two DNA strands, providing an exciting and novel mechanism to form the synaptic complex. Previously in the low-resolution structure of Int_{λ}-Holliday Junction intermediate, the interaction between Xis and AB domains was suggested to drive synaptic complex formation. In contrast, our structure shows that the Int_{GISul2} follows a different approach for synapsis, using its AB domains as a glue to pull two DNA fragments together. The AB filament formation could be a unique mechanism of Int enzymes from the Int_{SXT} family (Smyshlyaev et al., 2021), as the members of this family have additional structural elements in their AB domains, which provide a second DNA binding site for bipartite recruitment of two DNA molecules.

In the third part of the thesis, we obtained structural insights into how two transposon ends come together to form a synaptic complex during transposon excision. To trap a synaptic complex, we designed "suicide" DNA substrates, which can stall the recombination reaction after cleavage of the first DNA strand. Unexpectedly, this approach resulted in the formation of two different synaptic complexes, one including RE and LE DNA, and the other one containing two RE and RE molecules. In the RE-RE synaptic complex, both RE DNA molecules are bent sharply by IHF and all four ends are connected in the center by an Int_{GISul2} tetramer. DNA synapsis is driven by the formation of an AB filament, as seen in the LE complex, and by cyclic interactions of the CB and CAT domains at the core DNA sites. In physiological conditions, this complex probably does not form because only one RE and one LE of the element is present in one cell at a time. Still, this complex shows how an active tetrameric recombination assembly looks like. In the RE-LE synaptic complex, the overall DNA conformation of RE and LE are similar to their presynaptic complexes: LE DNA is looped by Xis, while RE DNA is sharply bent by IHF. The ends of both DNA fragments are connected by an Int_{GISul2} tetramer. The synapsis of RE and LE complexes involve all Int domains in a similar way as in the RE-RE synapse. While AB forms a stable filament which connects the arm sites of RE and LE DNA, CBs and CATs bridge the core sites of two transposon ends. Although, four AB domains form the AB filament, surprisingly only three CB and CAT domains are visible in the EM density.

In conclusion, our work provided new mechanistic insights into the biology of CTns, and of tyrosine recombinase-mediated DNA rearrangements in general. In particular, we now understand better how specific higher order complex formation can drive recombination towards different products, what the roles of the accessory factors are in this process, how they may control the timing of the reaction, and how host cells can control or contribute to the mobilization of their transposon residents. We hope that a better understanding of the CTn movement will ultimately help to develop new strategies for tackling antibiotic resistance spreading.

3.2 A model for *GISul2* Excision and its Regulation

3.2.1 An updated Model for *GISul2* Transposition

Based on our findings, we can propose the following model for explaining how higher order complex formation regulates *GISul2* transposition at different stages of the reaction.

1) The first step of excision is the formation of pre-synaptic complexes on individual transposon ends. The right end (RE) complex is formed on the right transposon end DNA once IHF binds and bends the DNA. This allows an Int_{GISul2} dimer to simultaneously bind and bridge the arm and core sites. The binding of Int_{GISul2} and IHF can happen concomitantly. For example, while Int_{GISul2} binds either the arm or core sites, IHF can bend the DNA and help Int_{GISul2} to reach and connect to the other site. In the presynaptic complex the Int_{GISul2}'s CAT domains are positioned so that the catalytic sites, which are completed by a reciprocal exchange of M-helixes between two CATs, are both in inactive state, where the catalytic tyrosine is far away from the DNA scissile phosphate. This presynaptic assembly, positions one CB domain of one Int_{GISul2} dimer detached from the core DNA repeat in the transposon flank, allowing it to recognize transposon ends in diverse genomic locations.

2) The detached CB of the RE complex scans DNA for another core site. Due to the supercoiled and condensed nature of bacterial genome, despite the long distance between RE and LE of the *GISul2* element (15.5 kb), RE can come in close proximity to the LE sequence to form a synaptic complex. Our structural data indicates that LE is not able to produce a pre-synaptic complex, because stable Int_{GISul2} binding seems to require the presence of a partner DNA. However, LE may already have Xis molecules bound to the DNA bending sites and Int_{GISul2} molecules bound to the arm sites. Once the RE complex approaches, a synaptic complex can be readily formed. The driving force for this process is the formation of the AB filament between the RE complex and LE DNA, which allows rigid positioning of the two transposon ends together. The synaptic complex contains a trimeric CAT architecture at the recombination core sites. With this arrangement, first strand cleavage is performed at the crossover regions of both RE and LE DNA. After this, strand exchange and ligation occur and a HJ DNA intermediate is formed. With the formation of this intermediate, structural isomerization occurs, which should involve major rearrangements of the core DNA sites, their bound Int_{GISul2} subunits and their connection to arm sites. In this way, the tetrameric synaptic complex can be formed on the HJ for second strand cleavage. Here, we propose that some of the Int_{GISul2}

subunits are bound to different arm sites in this complex, which are oriented in the reverse orientation in the LE, as discussed in Section 2.3.4. The isomerization step was shown to be the slowest step of recombination for the λ system, which can also be the case for *GISul2* system due to the complexity of the required rearrangements. After second strand cleavage and subsequent strand exchange, the transposon ends are fully recombined and an excised circular intermediate (CI) is produced, which is then released from the protein complex.

3) Later for the integration, there must be a higher order complex assembled on CI by using IHF as an accessory factor. Xis was shown to inhibit integration. We expect that the RE part of the CI complex looks similar to the one observed in the RE complex. On the LE part, one or multiple IHF binding sites may shape the DNA to form a distinct assembly that is activated for integration. Once such presynaptic complex is formed, an integration target DNA site can be captured and bound by the Int_{GISul2} subunits. The mechanistic details of the integration process still need to be elucidated.

3.2.2 The Regulation of *GISul2* Transposition by Accessory Factors

3.2.2.1 IHF Regulation

IHF is an essential component of the molecular complexes formed at the *GISul2* transposon ends during both integration and excision. Therefore, the formation of these complexes may depend on the cellular concentration of IHF, which is regulated by the host cell. This provides an opportunity for the bacteria to limit and control transposon movement. From another perspective, the transposon can use cellular IHF levels as a signal to determine the favorable time for its mobilization and continuation of its life cycle. Thus, dependence of Int_{GISul2} activity on host-encoded accessory proteins may help linking the regulation and direction of recombination to the physiology of the host cell. IHF levels increase in the stationary phase of bacterial growth 5 to 7-fold and decrease at the end of the stationary phase. For the λ phage, it is known that high levels of IHF concentration can also inhibit excision *in vitro* (Bushman et al., 1985). This suggests that during the stationary phase, CTn excision is not favored. Excision at this point will also not be useful for the element, because the host will not be able to support its mobilization, due to its reduced level of metabolism.

In turn integration during stationary phase may be favorable as it may help to limit the energetic cost related to CI replication.

Differential formation of distinct higher order complexes depending on IHF concentration levels can be achieved through the presence of multiple IHF binding sites with different affinities. High affinity binding sites can be occupied even at low levels of IHF and lead to the formation of molecular complexes favoring excision. In turn, the low affinity binding sites require higher IHF concentrations and may play their specific roles during integration, under higher IHF concentration.

Altogether, we suggest that IHF concentration, act as a direct molecular sensor of the cell state, regulating CTn activity according to the host life cycle.

3.2.2.2 Xis Regulation

Xis is a transposon encoded protein, which defines the directionality of the transposition reaction. As in the λ system, Xis is strictly required for *GISul2* excision, while it inhibits integration. The *xis* gene is located right after *int* in the transposon sequence. The exact promoter of *xis* is not known, but *int* and *xis* are found in frame, suggesting that they are co-regulated. The co-expression of Int and Xis proteins may help to form functional complexes during excision, while for integration, Xis needs to be downregulated. In λ, Xis levels are controlled by the Lon and FtsH proteases (Leffers and Gottesman, 1998; Landy, 2015). For the *GISul2* system, there can be a similar strategy employed to decrease Xis levels, either post- or pre-translation during integration of the element.

We have showed that the Xis binding is crucial for shaping the LE to form an active excision complex. The LE of *GISul2* includes multiple binding sites for Int_{GISul2} and for various accessory factors, allowing the formation of diverse architectures on this transposon end. Depending on the absence or presence of Xis, different DNA curvatures could be formed using different arm sites and other DNA bending proteins, which may consequently change the geometric constraints on CAT positioning and allow only a specific step of recombination to be carried out.

3.3 Future Directions

It has been known for a long time that conjugative transposons are particularly efficient in spreading antibiotic resistance. To come up with novel strategies to slow down their contribution to ARG spread, their mechanism of transposition needs to be well understood. Although there was a significant biochemical and some structural information available about their integrases, until today there was no information regarding to full protein-DNA complexes assembled during their transposition. With the work that I have presented in this thesis, we have characterized the higher-order complexes formed during excision of the *GISul2* transposon. Our work provides new mechanistic insights into the biology of CTns, and of tyrosine recombinase-mediated DNA rearrangements in general. Moreover, with the results presented in this study, new questions arise to better understand the details of the regulation of transposition by higher order complexes and the roles of the accessory factors in this process.

In order to validate our excision model, the order of cleavage reactions at the crossover sites during excision needs to be verified. Additionally, to complete the set of structural snapshots, a tetrameric complex, containing both RE and LE DNA and representing the complex during second strand cleavage needs to be obtained. For that aim, the additional arm sites found in LE can be used and the resulting complex can be characterized with cryo-EM. In addition, to check the effect of IHF concentration on excision, the excision assay can be performed with increasing IHF concentration. This can be tested with the same fluorescence-based assay that was optimized for excision, to determine the precise effect of IHF on guiding and controlling the reaction.

Once the excision reaction is fully characterized, the next step will be to continue with the characterization of integration, which is another crucial step of CTn transposition. Determining the structure of the dedicated integration complexes will help to understand how integration complex is assembled on the CI and how it captures the integration target site. For this purpose, an integration assay was already optimized to identify the minimal CI sequence that is required for integration. With the optimized length of CI, the complex will be trapped either by using a catalytically inactive Int_{GISul2} mutant or a suicide DNA oligo design, to gather structural information by cryo-EM.

Finally, the ultimate aim is to develop potential inhibitory molecules using the new insights on CTn transposition that we have obtained here. As a potential inhibitor of *GISul2* transposition, the peptide similar to M-helix, which binds and blocks the enzyme active site can be envisaged. Another idea can be to develop Xis specific inhibitors/proteases to restrain the excision of the element and limit the spreading of antibiotic resistance gene cargos.

4 MATERIALS AND METHODS

4.1 MATERIALS

4.1.1 Chemicals and Reagents

All the chemicals used in this thesis were purchased from Sigma-Aldrich and Merck Millipore. The restriction enzymes were acquired from New England Labs and Thermo Scientific. The antibiotics were purchased from Carl Roth. All the chemical stocks and buffers were prepared by using deionized, distilled, autoclaved water and filtered when it was required with 0.22 µM disposable filter.

4.1.2 Bacterial Growth Media

The bacterial growth media used for this study was prepared by EMBL Media and Kitchen Facility. *E. coli* cultures were grown in Luria-Bertani (LB) medium which contains 5 g yeast extract, 10 g tryptone and 5 g sodium chloride per liter (pH 7.2). The SOC medium is used in special cases contained 20 g tryptone, 5 g yeast extract, 0.585 g sodium chloride, and 0.186 g potassium chloride per liter (pH 7.2). LB agar plates were used as solid media to grow bacteria and they contained 10 g tryptone, 5 g yeast extract, 10 g sodium chloride, and 15 g agar per liter (pH 7.2). Depending on bacterial strain and the plasmid that it possesses, different antibiotics were added to LB-medium in the following concentrations: kanamycin at 50 µg/ml, ampicillin at 100 µg/ml and chloramphenicol at 33µg/ml.

4.1.3 Bacterial Strains

The strains of *E. coli* used for cloning and recombinant protein production were as following:

- XL10-Gold strain with the genotype: TetR Δ(mcrA)183 Δ(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F' proAB lacIqZDM15 Tn10 (TetR) Amy CmR], was supplied from Novagen.

- BL21 (DE3) strain with the genotype: F- ompT gal dcm lon hsdSB(rB- mB-) λ(DE3), was supplied by Novagen.

- E. coli strain Top10 – Genotype: F- mcrA Δ(mrr-hsdRMS-mcrBC) ϕ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara-leu)7697 galU galK rpsL (StrR) endA1 nupG was supplied by EMBL Protein Expression and Purification Core Facility (Pep-Core).

4.1.4 Plasmids

The plasmids used in this thesis are stated in Table 4-1. In the table, the name of the plasmid, its purpose of use and carried antibiotic resistance are stated.

Table 4-1: Plasmids used in the thesis

Plasmid	Description	Resistance	Origin	Source
pETM21-Int ^{WT}	Expression vector (N-terminal 6xHis-TRX tag)	Ampicillin	pBR322	EMBL PepCore
pETM21-Int^{R241K}	Expression vector (N-terminal 6xHis-TRX tag)	Ampicillin	pBR322	EMBL PepCore
pETM28- Xis	Expression vector (N-terminal 6xHis-Sumo tag)	Kanamycin	pBR322	EMBL PepCore
pETM10-IHFa	Expression vector (N-terminal 6xHis tag)	Kanamycin	pBR322	EMBL PepCore
pETM13-IHFb	Expression vector (no tag)	Kanamycin	pBR322	EMBL PepCore

4.1.5 Cryo-EM Grids

Grids used in this study for screening and data collection purposes were ordered from Quantifoil. The grid types used in this thesis are listed in Table 4-2. The material, hole and mesh sizes of the grids are also stated in the table.

Table 4-2: EM Grids used in the thesis

Grid Name	Material	Hole Size	Mesh Size
UltrAufoil	Au	R2/2	200
UltrAufoil	Au	R1.2/1.3	200
Quantifoil	Carbon	R2/1	300
Quantifoil	Carbon	R2/2	300
Quantifoil	Carbon	R2/1	300

4.2 MOLECULAR BIOLOGY METHODS

4.2.1 Polymerase Chain Reaction

PCR amplifications were needed for multiple purposes throughout the thesis. The primers were resuspended with water to a final concentration of 100 µM and dilutions are prepared at 10 µM concentration for the PCR mix. The PCR mix is prepared with 25 µl of 2X Phusion PCR Mix (supplied by EMBL PEP-Core Facility), 2.5 µl of each forward and reverse primers at 10 µM concentration, template DNA (20 ng) in total of 50 µl reaction volume. PCR of the prepared reaction was performed by following the conditions stated in Table 4-3. After PCR, to check the completeness and the specificity of the reaction, the samples were checked with agarose gel electrophoresis. For later use, they were purified by using GenElute PCR Clean-Up Kit (Sigma-Aldrich) according to manufacturer's protocol. Depending on purpose, in some cases the gel extraction was also performed.

Table 4-3: Thermocycling Steps for PCR

Step	Temperature	Duration
1.Initial Denaturation	98 °C	40 seconds
35 cycles (between steps 2-4)		
2.Denaturation	98 °C	10 seconds
3.Annealing	Variable	10 seconds
4.Extension	72 °C	1 minute per kb
5.Final Extension	72 °C	2-10 minutes
6.Hold	4 °C	-

4.2.2 Restriction Free Cloning

The restriction-free cloning was performed in a way that the designed primers have complementary sequences for the insert and the target site. The first PCR reaction is performed with the designed primers to amplify the desired insert. This reaction was set as stated in the previous section and PCR was performed by following the thermocycling steps stated in Table 4-3. After PCR, the reaction was checked with agarose gel electrophoresis to confirm the size of the amplified fragment, called as megaprimer. The PCR reaction is purified, and 2nd PCR was prepared with 25 µl of 2X Phusion PCR Mix (supplied by EMBL PEP-Core Facility), 350 ng of megaprimer, destination vector (50 ng) in total of 50 µl reaction volume. With this reaction, PCR was performed by following the

thermocycling steps stated in Table 4-4. The PCR product was purified and eluted with 17 µl distilled water. The purified DNA was mixed with 2 µl of FastDigest DpnI (Thermo Scientific) and 2 µl of 10X FastDigest buffer and incubated for 3 hours at 37 °C. 5-7 µl was used to transform *E. coli* XL10-Gold chemically competent cells, while 3 µl was used to transform electrocompetent cells.

Table 4-4: Thermocycling Steps for PCR

Step	Temperature	Duration
1.Initial Denaturation	98 °C	40 seconds
18 cycles (between steps 2-4)		
2.Denaturation	98 °C	10 seconds
3.Annealing	Variable	10 seconds
4.Extension	72 °C	7 minutes
5.Final Extension	72 °C	10 minutes
6.Hold	4 °C	-

4.2.3 Agarose Gel Electrophoresis

The agarose gels were prepared with between 0.7 - 2% (w/v) agarose in 1X Tris-Acetate EDTA (TAE) buffer (40mM Tris, 20 mM Acetic acid and 1 mM EDTA). Ethidium Bromide was added to gels and mixed while they were left for solidification. The samples were mixed with 5x loading dye (Qiagen) GelPilot Loading Dye in 1:5 ratio. The mixed samples were loaded next to 5 µl of DNA marker (Hyperladder 1kb Bioline or Easyladder). The gels were run at 120 V for 40 minutes, later visualized by UV transilluminator, Alpha Innotech. If the visualized band was required for the later steps of experiments, the bands were extracted by using GenElute Gel Extraction Kit (sigma-Aldrich) by following the manufacturer's protocol.

4.2.4 Transformation of Competent Cells

4.2.4.1 Electro-Competent Cells

The 50 µl aliquot of frozen electro-competent *E. coli* cells were thawed on ice for 10 minutes and up to 3 µl DNA was added to cells and mixed by flipping. The mixture was placed to cooled 0.1 cm electroporation cuvette (Bio-Rad). The electroporation was performed with Micro Pulser (Bio-Rad). The cells were immediately resuspended with 500 µl of SOC medium, which was pre-warmed at 37°C and transferred to an Eppendorf tube

and incubated 1 hour at 37°C with 700 rpm shaking. After incubation, cultures were plated to LB-plates containing appropriate antibiotic and incubated overnight at 37°C.

4.2.4.1 Chemo-Competent Cells

The 50 µl aliquot of frozen chemo-competent *E. coli* cells were thawed on ice for 10 minutes and up to 7 µl DNA was added to cells and mixed by flipping. The mixture was placed in water bath at 42°C and kept for 25 seconds. After heat shock cells were kept on ice for 10 minutes for recovery. Later, the cells were resuspended with 500 µl of SOC, which was pre-warmed at 37°C and transferred to an Eppendorf tube and incubated 1 hour at 37°C with 700 rpm shaking. After incubation, culture was plated to LB-plates with appropriate antibiotic and incubated overnight at 37°C.

4.2.5 Plasmid DNA Extraction

E. coli cultures with the plasmids of interest were grown in 5 ml LB medium containing the appropriate antibiotic overnight at 37°C. Next day, cells were centrifuged at 3000 g for 10 minutes at 4°C. Plasmid DNA was purified by using Gene Elute Mini Prep Kit (Sigma) following the manufacturer's protocol. In the final step, plasmid DNA was eluted with distilled water and stored at -20°C.

4.2.6 DNA Sequencing

DNA sequencing was performed by GATC Biotech or Europhins to validate the cloning. For this purpose, 15 µl of sample with 20 ng/µl concentration was sent. The sequences were checked with the APE software (A plasmid Editor).

4.2.7 Protein Overexpression and Purification

4.2.7.1 Protein Overexpression in *E. coli*

Previously described protein constructs were overexpressed in *E. coli* strains BL21 (DE3) strains. The gene of interest was cloned under T7 promoter in pET vectors. The expression vectors were transformed into chemo-competent cells, later the transformants were incubated on LB plates containing appropriate antibiotic. The plates were incubated overnight at 37°C. Next day, from the fresh colonies, 10 ml LB culture was initiated and incubated overnight at 37°C. For the large-scale expression, for each 500 ml of LB

containing the desired antibiotic, 1 ml of the overnight culture was added. The flasks were incubated at 37°C with 180 rpm shaking until the OD reaches 0.6. Once the desired OD was reached, the cultures were cooled down for 20 minutes in the incubator which was set to 17°C. Later, the cultures were induced by adding 0.5 mM IPTG and incubated for 17-19 hours with 180 rpm shaking at 17°C. To harvest, the cultures were centrifugated at 4°C at 4000 g for 30 minutes. Later, bacteria pellet was resuspended and washed with cold PBS. After the final harvest, the pellet was either used for purification and lysed immediately, or frozen in liquid nitrogen and kept at -80°C.

4.2.7.1 Protein Purification

All proteins used in this project were tagged with 6xHis tag and purified in three-step purification: 1- first nickel affinity chromatography, 2- cleavage of the tag and second nickel affinity chromatography, 3- size exclusion chromatography on SEC column. All steps of the purification were performed by using ÄKTA protein purification system (GE Healthcare) at 4°C. The chromatograms were depicted at 280 nm (for protein) and at 260 nm (for DNA, or DNA-containing complexes).

4.2.7.1.1 Sample Preparation

The buffers used during the protein purification are listed in Table 4-6. For the purification of Int and IHF, buffers were prepared with 500 mM NaCl as stated in the table, while for Xis purification 500 mM NaCl was replaced by 1M NaCl in buffers. The frozen pellets were thawed on ice for 30 minutes and resuspended with the lysis buffer (Table 4-6) until there was no clump left. The resuspended cells were lysed by sonication on ice with using Branson Sonifier 250. The sonication protocol was as follows: 3 minutes sonication time, 10 seconds on, 30 seconds off at 40% efficiency. The crude lysate was centrifuged at 17 000 rpm at 4°C for 30 minutes. The supernatant was filtered through 0.22 µM disposable filter and loaded onto 5-ml HisTrap column (GE Healthcare).

Table 4-5: Composition of Buffers used during Protein Purification

Buffer	Composition
Lysis Buffer	50 mM Tris-Cl (pH 7.5), 500 mM NaCl, 5% Glycerol, 1mM DTT, 1 table of Complete Protease Inhibitor Cocktail (Roche), 50 µg/ml RNaseA and 50 µg/ml DNaseI
Buffer A	50 mM Tris-Cl (pH 7.5), 500 mM NaCl, 5% Glycerol, 1mM DTT, 1mM DTT
Buffer B	50 mM Tris-Cl (pH 7.5), 500 mM NaCl, 5% Glycerol, 1mM DTT, 1 mM DTT, 500 mM Imidazole
SEC Buffer	50 mM Tris-Cl (pH 7.5), 500 mM NaCl, 5% Glycerol, 1mM DTT, 1mM DTT

4.2.7.1.2 First HisTrap Purification

The 5 ml-HisTrap column was used for the first HisTrap and it was washed with 60 ml of filtered distilled water, then equilibrated with 60 ml of Buffer A (Table 4-5). All the washing steps were carried out with 2.5 ml/min flow rate. Later, the lysate was loaded to the equilibrated column at 1.25 ml/min flow rate. Once the injection was completed, the flow rate was increased to 2 ml/min for the rest of the purification. After the completion of injection, the column was washed with Buffer A until the absorbance reached the baseline (10-15 CV). To elute the unspecific binders, column was washed with 5% Buffer B. The bound proteins were eluted from the column with increasing percentage of Buffer B (between 10-100%). During the elution, 3 ml fractions were collected in collection tubes. Once the elution of proteins started, the Buffer B was kept stable at that concentration for the full elution. Once the elution was over, the column was washed with 40 ml of 100% Buffer B to remove all the impurities, later 40 ml of Buffer A and 60 ml of filtered distilled water. The fractions were analyzed on SDS-PAGE and the interested fractions were pooled for tag removal.

To remove the Sumo tag, the pooled fractions were mixed with SenP2 protease (PEPcore, EMBL Heidelberg). The protease was added 1:100 molar ratio and the mix was dialyzed to SEC buffer to remove the imidazole from the mix in a dialysis membrane tubing with the molecular weight cut off 3.5 kDa (Spectrum Labs). The dialysis was performed overnight at 4 ℃ with constant mixing.

4.2.7.1.3 Second HisTrap Purification

After the dialysis, the cleaved mixture was injected on prewashed and equilibrated 5ml-HisTrap column with 1 ml /min flow rate. During this step, the flow through was also collected. Once the injection was over, the flow rate was increased to 2 ml/min for the rest of the purification. The elution was performed with a gradient of Buffer B from 10 % to 100%. All the fractions were collected as in the previous purification as 3 ml. The cleaved proteins were either eluted in the flowthrough or with 5% Buffer B, due to the cleaved 6xHis tag. The fractions eluted with higher percentage of Buffer B included the 6xHis -Sumo tag and His-tagged SenP2 protease and uncleaved protein. Once the elution was over, the column was washed with 40 ml of 100% Buffer B to remove all the impurities, later 40 ml of Buffer A and 60 ml of filtered distilled water. The fractions were analyzed on SDS-PAGE and interested fractions were pooled for tag removal. For long term storage column was kept at 20% ethanol.

4.2.7.1.3 Size-Exclusion Chromatography (SEC)

The collected fractions were pooled and concentrated to be load to SEC column. The concentration process was done with Vivaspin Turbo concentrator (for Int 10 kDa-cut off, for Xis and IHF 3.5 kDa cutoff) by centrifugation at 3500 g at 4°C until the volume of concentrated protein reached 5 ml to be to the column. The SEC column (HiLoad 16/600 Superdex 200) was equilibrated with 0.75 ml/min flow rate with SEC Buffer (Table 4-6). After equilibration, 5 ml of concentrated protein was loaded to the column with 0.9 ml/min flow rate. The proteins were separated based on their size; while the bigger ones eluted in the earlier fractions, the smaller ones were eluted in later fractions. The fractions were checked by loading to SDS-PAGE to detect the protein of interest.

4.3 BIOCHEMICAL METHODS

4.3.1 Annealing Oligos

All double stranded DNA substrates were annealed by mixing the oligonucleotides in equimolar ratios, which were previously suspended in water. The mixture was incubated for 5 min at 95°C, later the heat block was turned off to let the tube cool down gradually until it reached the room temperature. The annealed oligos were stored at -20°C.

4.3.2 Electrophoretic Mobility Shift Assay (EMSA)

The EMSA assays were performed to detect the DNA binding properties of proteins. This assay was performed with 1% agarose gel prepared with 1X TBE (Tris-borate-EDTA) buffer. The double stranded DNA fragments were mixed with proteins of interest in complex formation buffer containing 10mM Tris-Cl (pH 7.5), 50 mM NaCl, 5% Glycerol, 1mM DTT and incubated for 30 minutes at room temperature. The complexes were loaded directly to the agarose gel and run at 100 V for 40 minutes. Once the run was finished, the gels were stained in Ethidium Bromide containing TBE buffer for 30 minutes and imaged as previously described.

4.3.3 Excision Assay with Fluorescence

Excision assay with fluorescently labelled oligonucleotides were developed to observe the excision of the element in a sensitive way. The fragment required for the assay was amplified from mini-transposon containing plasmid based on the protocol shown in Section

4.2.1. The forward primer used for this assay was labelled with 5' fluorescein amidate (5'-FAM). The amplified fragment was mixed with Int_{GISul2}, Xis and IHF in a buffer containing 10mM Tris-Cl (pH 7.5), 50 mM NaCl, 5% Glycerol, 1mM DTT, 10mM $MgCl_2$ and incubated overnight at room temperature at dark. Next day, the complex was incubated with 6 µl Proteinase K and with 6 µl of 6X Proteinase K buffer at 65 °C for half an hour. Later, the sample was mixed with 5X DNA loading dye and loaded to 1% Agarose gel made with TAE Buffer. Gel was run for 40 minutes at 100 V at room temperature. Later for imaging the gel Typhoon FLA 7000 phosphoimager (GE Healthcare) was used by using laser with wavelength at 477 nm. After imaging the fluorescence signal, the gel was stained with EtBr for 30 minutes in TAE buffer containing EtBr and imaged by UV transilluminator, Alpha Innotech.

4.4 CRYO-EM METHODS

4.4.1 Sample Preparation for Cryo-EM

The complexes used in cryo-EM grid preparation were prepared differently depending on the DNA substrate used in the experiment.

To prepare pre-synaptic RE and LE complexes, the components stated in the Table 4-6 and Table 4-7 were mixed, respectively. The mixed components were placed in a dialysis bag with 3.5 kDa cut-off and dialyzed to buffer containing 10 mM Tris-Cl (pH 7.5), 50 mM NaCl, 1mM DTT overnight at 4°C.

Table 4-6: RE Complex Preparation

Component	Volume
RE DNA @25uM	16 µl
IHF (150 uM)	4 µl
Int wt (80 uM)	12 µl
dH2O	76 µl
Total Volume	108 µl

Table 4-7: LE Complex Preparation

Step	Temperature
LE DNA (50uM)	10 µl
Int R mutant (50 uM)	40 µl
Xis (70 uM)	50 µl

IHF (150 uM)	7 µl
Total Volume	107 µl

To prepare the synaptic complexes, the "suicide" DNA substrates were designed. With this design, Int needed to perform the first cleavage to get trapped as a covalent intermediate. To favor the cleavage, the following protocol was followed for the preparation of synaptic complexes.

As a first step, the components of RE complex were added as in the written order in Table 4-8 and it was incubated for 1 hour at room temperature. Later, the components of LE complex (Table 4-9) were added to previously prepared RE complex, as in the written order in Table 4-9. The mixture was placed in a dialysis bag with 3.5 kDa cut-off and dialyzed to buffer containing 10 mM Tris-Cl (pH 7.5), 50 mM NaCl, 1mM DTT. The dialysis was performed for 2 hours at room temperature, then it was dialyzed overnight at 4°C.

Table 4-8: Synaptic RE Complex Preparation

Step	Temperature
"suicide"RE DNA (30uM)	6 µl
IHF (150 uM)	3 µl
Int wt (80 uM)	6 µl
H_20	39 µl
Total	54 µl

Table 4-9: Synaptic LE Complex Preparation

Step	Temperature
"suicide"LE DNA (50uM)	4
Int wt (80 uM)	5
Xis (70 uM)	15
IHF (150 uM)	2
Total	26

Prior to plunging, the grids used in that session were glow discharged with PELCO easiGlow system for 45 seconds. The dialyzed complexes were plunge frozen by using Vitrobot Mark IV. During plunging, the manufacturer's directions were performed. The chamber of the Vitrobot was set to 10°C with 95% humidity. The liquid nitrogen was filled to Nitrogen cup and once it cooled down the inner Ethane cup and Ethane was filled to inner cup. Once the signs of ethane freezing were observed, the connecting metal piece was removed. The

container was placed in Vitrobot. The grids were plunged by following these parameters: blot force 3, blot time 0 second and wait time 0 seconds. 3.5 ul of the dialyzed complex was pipetted on the glow discharged grids and plunged. The grids were stored in grid boxes in liquid nitrogen until the imaging day.

4.4.2 Cryo-EM Data Collection

Data collection was performed mainly on two microscopes: Talos Arctica and Titan Krios. The previously plunged grids were clipped in between cartridges and clipped rings. The clipped grids were inserted into the cassette, which had capacity for 12 grids. The cassette was placed into a specific container, called as capsule, filled with liquid nitrogen. The capsule was placed in microscope and the cassette carrying the grids was docked in microscope. Both microscopes had SerialEM software to control the microscope tuning and data collection.

The grid maps were collected at low magnification (150x), to have an idea about the ice distribution on the grid. Based on these maps, the thin ice having squares were picked and square maps were collected for them at medium magnification(2600x). Form the square maps, the holes were picked, and the microscope was set for high resolution data collection. For Talos Arctica datasets, the electron dose of 30 e/Å^2 and pixel size of 1.21 Å/pixel were used 20-framed movies. In turn, for Titan Krios total dose of 44 e/Å^2 with a pixel size of 0.81 Å/pixel were used to collect 40-framed movies. The high-resolution data collection was done for the picked holes.

4.4.3 Cryo-EM Data Processing

The collected movies were processed with Cryosparc2 software (Punjani et al., 2017). As a first step, movies were patch-motion corrected to align the frames without any shift. As a second step, CTF was estimated with patch-CTF estimation. The micrographs were manually curated based on relative ice thickness and CTF fit resolution to eliminate micrographs which have thicker ice. Depending on the dataset, the templates for particle picking were produced either from a small set of manually picked particles or previously produced 2D classes. Particles were picked with reference-based particle picking and the parameters of particle picking were adjusted based on the contrast of the particles in the aligned micrographs. The picked particles were initially extracted with a box size approximately 1.5 times bigger than the longest dimension of the particles and 2D classification was performed. The junk particles were carefully removed with some more rounds of 2D and 3D classifications. The final set of particles were re-extracted with a bigger box size, 2-2.5 times bigger than the longest dimension of the particles. To produce

high resolution 3D volume, either homogenous refinement or non-uniform refinement was used depending on the case. As a final step, CTF refinement was performed.

The produced 3D cryo-EM maps were inspected by using ChimeraX (Pettersen et al., 2020) for segmentation, model fitting and superimposing. For the fitted models, the real space refinement was performed in Phenix (Liebschner et al., 2019).

5 REFERENCES

Abbani, M., Iwahara, M., & 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), 11–25.

Abremski, K., & Gottesman, S. (1982). Purification of the bacteriophage lambda xis gene product required for lambda excisive recombination. The Journal of biological chemistry, 257(16), 9658–9662.

Bi D., Xu Z., Harrison E. M., Tai C., Wei Y., He X., et al. . (2012). ICEberg: a web-based resource for integrative and conjugative elements found in Bacteria. Nucleic Acids Res. 40, D621–D626. 10.1093/nar/gkr846

Binda, E., Marinelli, F., & Marcone, G. L. (2014). Old and new glycopeptide antibiotics: Action and resistance. Antibiotics, 3(4), 572–594.

Biswas, T., Aihara, H., Radman-Livaja, M., Filman, D., Landy, A., & Ellenberger, T. (2005). A structural basis for allosteric control of DNA recombination by lambda integrase. Nature, 435(7045), 1059–1066. https://doi.org/10.1038/nature03657

Blokesch M. (2016). Natural competence for transformation. Current biology: CB, 26(21), R1126–R1130.

Burrus, V., & Waldor, M. K. (2003). Control of SXT integration and excision. Journal of Bacteriology, 185(17), 5045–5054. https://doi.org/10.1128/JB.185.17.5045-5054.2003

Burrus, V., Pavlovic, G., Decaris, B., & Guédon, G. (2002). Conjugative transposons: the tip of the iceberg. Molecular microbiology, 46(3), 601–610. https://doi.org/10.1046/j.1365-2958.2002.03191.x

Bushman, W., Thompson, J. F., Vargas, L., & Landy, A. (1985). Control of directionality in lambda site specific recombination. Science (New York, N.Y.), 230(4728), 906–911. https://doi.org/10.1126/science.2932798

Cabezón, E., Ripoll-Rozada, J., Peña, A., de la Cruz, F., & Arechaga, I. (2015). Towards an integrated model of bacterial conjugation. FEMS microbiology reviews, 39(1), 81–95. https://doi.org/10.1111/1574-6976.12085

Chen, Y., Narendra, U., Iype, L. E., Cox, M. M., & Rice, P. A. (2000). Crystal structure of a Flp recombinase-Holliday junction complex: assembly of an active oligomer by helix swapping. Molecular cell, 6(4), 885–897.

Cheng, Y. S., Yang, W. Z., Johnson, R. C., & Yuan, H. S. (2000). Structural analysis of the transcriptional activation region on Fis: crystal structures of six Fis mutants with different activation properties. Journal of molecular biology, 302(5), 1139–1151. https://doi.org/10.1006/jmbi.2000.4123

Chiang, Y. N., Penadés, J. R., & Chen, J. (2019). Genetic transduction by phages and chromosomal islands: The new and noncanonical. PLoS Pathogens, 15(8), 1–7. https://doi.org/10.1371/journal.ppat.1007878

Chopra, I., & Roberts, M. (2001). Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. Microbiology and molecular biology reviews : MMBR, 65(2), 232–260.

Christ, N., & Dröge, P. (1999). Alterations in the directionality of lambda site-specific recombination catalyzed by mutant integrases in vivo. Journal of molecular biology, 288(5), 825–836. https://doi.org/10.1006/jmbi.1999.2730

Claverys, J. P., & Martin, B. (2003). Bacterial "competence" genes: signatures of active transformation, or only remnants?. Trends in microbiology, 11(4), 161–165. https://doi.org/10.1016/s0966-842x(03)00064-7

Dahlberg, C., Bergström, M., Andreasen, M., Christensen, B. B., Molin, S., & Hermansson, M. (1998). Interspecies bacterial conjugation by plasmids from marine environments visualized by gfp expression. Molecular Biology and Evolution, 15(4), 385–390. https://doi.org/10.1093/oxfordjournals.molbev.a025935

Delavat, F., Miyazaki, R., Carraro, N., Pradervand, N., & van der Meer, J. R. (2017). The hidden life of integrative and conjugative elements. FEMS microbiology reviews, 41(4), 512–537. https://doi.org/10.1093/femsre/fux008

DiGabriele, A. D., & Steitz, T. A. (1993). A DNA dodecamer containing an adenine tract crystallizes in a unique lattice and exhibits a new bend. Journal of molecular biology, 231(4), 1024–1039. https://doi.org/10.1006/jmbi.1993.1349

Ennifar, E. et al. (2003) 'Crystal structure of a wild-type Cre recombinase-loxP 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.

Fernández, L., & Hancock, R. E. W. (2012). Adaptive and mutational resistance: Role of porins and efflux pumps in drug resistance. Clinical Microbiology Reviews, 25(4), 661–681. https://doi.org/10.1128/CMR.00043-12

Fleming, A. (1929) 'A. Fleming,1929- antibacterial action of cultures of a penicillium.pdf', The British Journal of Experimental Pathology, pp. 226–236.

Flores-Ríos, R., Quatrini, R., & Loyola, A. (2019). Endogenous and Foreign Nucleoid-Associated Proteins of Bacteria: Occurrence, Interactions and Effects on Mobile Genetic Elements and Host's Biology. Computational and Structural Biotechnology Journal, 17, 746–756.

Franke, A. E., & Clewell, D. B. (1981). Evidence for conjugal transfer of a Streptococcus faecalis transposon (Tn916) from a chromosomal site in the absence of plasmid DNA. Cold Spring Harbor symposia on quantitative biology, 45 Pt 1, 77–80. https://doi.org/10.1101/sqb.1981.045.01.014

Furuya, E. Y., & Lowy, F. D. (2006). Antimicrobial-resistant bacteria in the community setting. Nature Reviews Microbiology, 4(1), 36–45. https://doi.org/10.1038/nrmicro1325

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.

Gillings M. R. (2014). Integrons: past, present, and future. Microbiology and molecular biology reviews: MMBR, 78(2), 257–277. https://doi.org/10.1128/MMBR.00056-13

Goodman, S. D., & Kay, O. (1999). Replacement of integration host factor protein-induced DNA bending by flexible regions of DNA. The Journal of biological chemistry, 274(52), 37004–37011. https://doi.org/10.1074/jbc.274.52.37004

Goodrich, J. A., Schwartz, M. L., & McClure, W. R. (1990). Searching for and predicting the activity of sites for DNA binding proteins: compilation and analysis of the binding sites for Escherichia coli integration host factor (IHF). Nucleic acids research, 18(17), 4993–5000. https://doi.org/10.1093/nar/18.17.4993

Goosen, N., & van de Putte, P. (1995). The regulation of transcription initiation by integration host factor. Molecular microbiology, 16(1), 1–7. https://doi.org/10.1111/j.1365-2958.1995.tb02386.x

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.

Grainge, I., & Jayaram, M. (1999). The integrase family of recombinases: Organization and function of the active site. Molecular Microbiology, 33(3), 449–456. https://doi.org/10.1046/j.1365-2958.1999.01493.x

Grainger, D. C., & Busby, S. J. (2008). Global regulators of transcription in Escherichia coli: mechanisms of action and methods for study. Advances in applied microbiology, 65, 93–113. https://doi.org/10.1016/S0065-2164(08)00604-7

Grindley, N. D., Whiteson, K. L., & Rice, P. A. (2006). Mechanisms of site-specific recombination. Annual review of biochemistry, 75, 567–605. https://doi.org/10.1146/annurev.biochem.73.011303.073908

Guilfoile, P. G., & Hutchinson, C. R. (1991). A bacterial analog of the mdr gene of mammalian tumor cells is present in Streptomyces peucetius, the producer of daunorubicin and doxorubicin. Proceedings of the National Academy of Sciences of the United States of America, 88(19), 8553–8557. https://doi.org/10.1073/pnas.88.19.8553

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.

Guo, F., Gopaul, D. N., & Van Duyne, G. D. (1999). Asymmetric DNA bending in the Cre-loxP site-specific recombination synapse. Proceedings of the National Academy of Sciences of the United States of America, 96(13), 7143–7148.

Hales, L. M., Gumport, R. I., & Gardner, J. F. (1994). Determining the DNA sequence elements required for binding integration host factor to two different target sites. Journal of bacteriology, 176(10), 2999–3006. https://doi.org/10.1128/jb.176.10.2999-3006.1994

Hamidian, M., & Hall, R. M. (2016). Acinetobacter baumannii ATCC 19606 Carries *GIsul2* in a Genomic Island Located in the Chromosome. Antimicrobial agents and chemotherapy, 61(1), e01991-16. https://doi.org/10.1128/AAC.01991-16

Haykinson, M. J., Johnson, L. M., Soong, J., & Johnson, R. C. (1996). The Hin dimer interface is critical for Fis-mediated activation of the catalytic steps of site-specific DNA inversion. Current Biology, 6(2), 163–177. https://doi.org/https://doi.org/10.1016/S0960-9822(02)00449-9

Hochhut, B., and M. K. Waldor (1999). Site-specific integration of the conjugal Vibrio cholerae SXT element into prfC.Mol. Microbiol.32:99-110

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.

Howard-Varona, C., Hargreaves, K. R., Abedon, S. T., & Sullivan, M. B. (2017). Lysogeny in nature: Mechanisms, impact and ecology of temperate phages. ISME Journal, 11(7), 1511–1520. https://doi.org/10.1038/ismej.2017.16

Ikeda, H., & Tomizawa, J. (1968). Prophage P1, and extrachromosomal replication unit. Cold Spring Harbor symposia on quantitative biology, 33, 791–798. https://doi.org/10.1101/sqb.1968.033.01.091

Jayaram, M., Ma, C. H., Kachroo, A. H., Rowley, P. A., Guga, P., Fan, H. F., & Voziyanov, Y. (2015). An Overview of Tyrosine Site-specific Recombination: From an Flp Perspective. Microbiology spectrum, 3(4), 10.1128/microbiolspec.MDNA3-0021-2014. https://doi.org/10.1128/microbiolspec.MDNA3-0021-2014

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

Johnston C., Martin B., Fichant G., Polard P., Claverys J. P. (2014). Bacterial transformation: distribution, shared mechanisms and divergent control. Nat. Rev. Microbiol. 12 181–196. 10.1038/nrmicro3199

Kapoor, G., Saigal, S., & Elongavan, A. (2017). Action and resistance mechanisms of antibiotics: A guide for clinicians. Journal of anaesthesiology, clinical pharmacology, 33(3), 300–305.

Knapp, C. W., Dolfing, J., Ehlert, P. A. I., & Graham, D. W. (2010). Evidence of increasing antibiotic resistance gene abundances in archived soils since 1940. Environmental Science and Technology, 44(2), 580–587. https://doi.org/10.1021/es901221x

Koskella, B., & Meaden, S. (2013). Understanding bacteriophage specificity in natural microbial communities. Viruses, 5(3), 806–823. https://doi.org/10.3390/v5030806

Krause, K. M., Serio, A. W., Kane, T. R., & Connolly, L. E. (2016). Aminoglycosides: An Overview. Cold Spring Harbor perspectives in medicine, 6(6), a027029. https://doi.org/10.1101/cshperspect.a027029

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.

Landy, A., & Ross, W. (1977). Viral integration and excision: structure of the lambda att sites. Science (New York, N.Y.), 197(4309), 1147–1160. https://doi.org/10.1126/science.331474

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., Xu, C., Brilot, A. F., Warren, D., Steele, L., Seah, N., Tong, W., Grigorieff, N., Landy, A., & Van Duyne, G. D. (2016). Structure of a Holliday junction complex reveals mechanisms governing a highly regulated DNA transaction. eLife, 5, e14313. https://doi.org/10.7554/eLife.14313

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

Lee, J., Jayaram, M., & Grainge, I. (1999). Wild-type Flp recombinase cleaves DNA in trans. The EMBO journal, 18(3), 784–791. https://doi.org/10.1093/emboj/18.3.784

Leffers, G. G., Jr, & Gottesman, S. (1998). Lambda Xis degradation in vivo by Lon and FtsH. Journal of bacteriology, 180(6), 1573–1577. https://doi.org/10.1128/JB.180.6.1573-1577.1998

Lewis, J. A., & 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, 29(11), 2205–2216. https://doi.org/10.1093/nar/29.11.2205

Li, W., Sharma, M., and Kaur, P. (2014). The DrrAB efflux system of Streptomyces peucetius is a multidrug transporter of broad substrate specificity. J. Biol. Chem. 289, 12633–12646. doi: 10.1074/jbc.M113.536136

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.

Llosa, M., Gomis-Rüth, F.X., Coll, M. and Cruz, F.d.l. (2002), Bacterial conjugation: a two-step mechanism for DNA transport. Molecular Microbiology, 45: 1-8.

M., P. R. T., I., B. M., Alan, L., & Nancy, C. (2015). Tyrosine Recombinase Retrotransposons and Transposons. Microbiology Spectrum, 3(2), 3.2.07. https://doi.org/10.1128/microbiolspec.MDNA3-0036-2014

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.

Meinke, G., Bohm, A., Hauber, J., Pisabarro, M. T., & Buchholz, F. (2016). Cre Recombinase and Other Tyrosine Recombinases. Chemical Reviews, 116(20), 12785–12820. https://doi.org/10.1021/acs.chemrev.6b00077

Nash, H. A., & Robertson, C. A. (1981). Purification and properties of the Escherichia coli protein factor required for lambda integrative recombination. The Journal of biological chemistry, 256(17), 9246–9253.

Nicolas, E., Lambin, M., Dandoy, D., Galloy, C., Nguyen, N., Oger, C. A., & Hallet, B. (2015). The Tn3-family of Replicative Transposons. Microbiology spectrum, 3(4), 10.1128/microbiolspec.MDNA3-0060-2014. https://doi.org/10.1128/microbiolspec.MDNA3-0060-2014

Nigro, S. J., & Hall, R. M. (2011). GIsul2 , a genomic island carrying the sul2 sulphonamide resistance gene and the small mobile element CR2 found in the Enterobacter cloacae subspecies cloacae type strain ATCC 13047 from 1890 , Shigella flexneri ATCC 700930 from 1954 and Acinetobacter ba. June, 2175–2176. https://doi.org/10.1093/jac/dkr230

Nunes-Düby, S. E., Kwon, H. J., Tirumalai, R. S., Ellenberger, T., & Landy, A. (1998). Similarities and differences among 105 members of the Int family of site-specific recombinases. Nucleic acids research, 26(2), 391–406. https://doi.org/10.1093/nar/26.2.391

Palchevskiy, V., & Finkel, S. E. (2006). Escherichia coli competence gene homologs are essential for competitive fitness and the use of DNA as a nutrient. Journal of bacteriology, 188(11), 3902–3910. https://doi.org/10.1128/JB.01974-05

Parsley LC, Consuegra EJ, Kakirde KS, Land AM, Harper WF Jr, Liles MR. (2010). Identification of diverse antimicrobial resistance determinants carried on bacterial, plasmid, or viral metagenomes from an activated sludge microbial assemblage. Appl Environ Microbiol 76: 3753–3757.

Partridge, S. R., Kwong, S. M., Firth, N., & Jensen, S. O. (2018). Mobile genetic elements associated with antimicrobial resistance. Clinical Microbiology Reviews, 31(4), 1–61. https://doi.org/10.1128/CMR.00088-17

Patrick Trieu-Cuot, Cécile Carlier, Claire Poyart-Salmeron, Patrice Courvalin (1991). An integrative vector exploiting the transposition properties of Tn1545 for insertional mutagenesis and cloning of genes from Gram-positive bacteria, Gene, Volume 106, Issue 1.

Peterson, E., & Kaur, P. (2018). Antibiotic resistance mechanisms in bacteria: Relationships between resistance determinants of antibiotic producers, environmental bacteria, and clinical pathogens. Frontiers in Microbiology, 9(NOV), 1–21. https://doi.org/10.3389/fmicb.2018.02928

Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H., & Ferrin, T. E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein science: a publication of the Protein Society, 30(1), 70–82. https://doi.org/10.1002/pro.3943

Pham, J. V., Yilma, M. A., Feliz, A., Majid, M. T., Maffetone, N., Walker, J. R., Kim, E., Cho, H. J., Reynolds, J. M., Song, M. C., Park, S. R., & Yoon, Y. J. (2019). A review of the microbial production of bioactive natural products and biologics. Frontiers in Microbiology, 10(JUN), 1–27.

Punjani, A., Rubinstein, J. L., Fleet, D. J., & Brubaker, M. A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nature methods, 14(3), 290–296. https://doi.org/10.1038/nmeth.4169

Rice, P. A., Yang, S., Mizuuchi, K., & Nash, H. A. (1996). Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. Cell, 87(7), 1295–1306. https://doi.org/10.1016/s0092-8674(00)81824-3

Ross, W., & Landy, A. (1983). Patterns of lambda Int recognition in the regions of strand exchange. Cell, 33(1), 261–272. https://doi.org/10.1016/0092-8674(83)90355-0

Rubio-Cosials, A., Schulz, E. C., Lambertsen, L., Smyshlyaev, G., Rojas-Cordova, C., Forslund, K., Karaca, E., Bebel, A., Bork, P., & Barabas, O. (2018). Transposase-DNA Complex Structures Reveal Mechanisms for Conjugative Transposition of Antibiotic Resistance. Cell, 173(1), 208–220.e20.

Ryan, M. P., Armshaw, P., & Pembroke, J. T. (2016). SXT/R391 integrative and conjugative elements (ICEs) encode a novel “trap-door” strategy for mobile element escape. Frontiers in Microbiology, 7(MAY), 1–6. https://doi.org/10.3389/fmicb.2016.00829

S.J. vanHal, C.L.C. Ip, M.A. Ansari, D.J. Wilson, B.A. Espedido, S.O. Jensen, R. Bowden (2016). Evolutionary dynamics of Enterococcus faecium reveals complex genomic relationships between isolates with independent emergence of vancomycin resistance.

Salyers, A. A., Shoemaker, N. B., Stevens, A. M., & Li, L. Y. (1995). Conjugative transposons: an unusual and diverse set of integrated gene transfer elements. Microbiological reviews, 59(4), 579–590. https://doi.org/10.1128/mr.59.4.579-590.1995

Sam, M. D., Abbani, M. A., Cascio, D., Johnson, R. C., & Clubb, R. T. (2006). Crystallization, dehydration and preliminary X-ray analysis of excisionase (Xis) proteins cooperatively bound to DNA. Acta crystallographica. Section F, Structural biology and crystallization communications, 62(Pt 8), 825–828. https://doi.org/10.1107/S1744309106027643

Sam, M. D., Cascio, D., Johnson, R. C., & Clubb, R. T. (2004). Crystal structure of the excisionase-DNA complex from bacteriophage lambda. Journal of molecular biology, 338(2), 229–240.

Senecoff, J. F., Bruckner, R. C., & Cox, M. M. (1985). The FLP recombinase of the yeast 2-micron plasmid: characterization of its recombination site. Proceedings of the National Academy of Sciences of the United States of America, 82(21), 7270–7274. https://doi.org/10.1073/pnas.82.21.7270

Shintani, M., Sanchez, Z. K., & Kimbara, K. (2015). Genomics of microbial plasmids: Classification and identification based on replication and transfer systems and host taxonomy. Frontiers in Microbiology, 6(MAR), 1–16. https://doi.org/10.3389/fmicb.2015.00242

Siguier, P., Gourbeyre, E., Varani, A., Ton-Hoang, B., & Chandler, M. (2015). Everyman's Guide to Bacterial Insertion Sequences. Microbiology spectrum, 3(2), MDNA3–2014.

Smyshlyaev, G., Bateman, A., & Barabas, O. (2021). Sequence analysis of tyrosine recombinases allows annotation of mobile genetic elements in prokaryotic genomes. Molecular systems biology, 17(5), e9880. https://doi.org/10.15252/msb.20209880

Song, L., Pan, Y., Chen, S., & Zhang, X. (2012). Structural characteristics of genomic islands associated with GMP synthases as integration hotspot among sequenced microbial genomes. Computational biology and chemistry, 36, 62–70. https://doi.org/10.1016/j.compbiolchem.2012.01.001

Stella, S., Cascio, D., & Johnson, R. C. (2010). The shape of the DNA minor groove directs binding by the DNA-bending protein Fis. Genes & development, 24(8), 814–826. https://doi.org/10.1101/gad.1900610

Sun D. (2018). Pull in and Push Out: Mechanisms of Horizontal Gene Transfer in Bacteria. Frontiers in microbiology, 9, 2154. https://doi.org/10.3389/fmicb.2018.02154

Thomas, C. M. (2000) 'Paradigms of plasmid organization', Molecular Microbiology. Mol Microbiol, pp. 485–491.

Tipper D. J. (1985). Mode of action of beta-lactam antibiotics. Pharmacology & therapeutics, 27(1), 1–35. https://doi.org/10.1016/0163-7258(85)90062-2

Tong, W., Warren, D., Seah, N. E., Laxmikanthan, G., Van Duyne, G. D., & Landy, A. (2014). Mapping the λ Integrase bridges in the nucleoprotein Holliday junction intermediates of viral integrative and excisive

recombination. Proceedings of the National Academy of Sciences of the United States of America, 111(34), 12366–12371.

Traglia, G.M., Chua, K., Centrón, D., Tolmasky, M.E., and Ramírez, M.S. (2014). Whole-genome sequence analysis of the naturally competent Acinetobacter baumannii clinical isolate A118. Genome Biol. Evol. 6: 2235–2239.

Van Duyne G. D. (2005). Lambda integrase: armed for recombination. Current biology: CB, 15(17), R658–R660. https://doi.org/10.1016/j.cub.2005.08.031

Van Duyne G. D. (2015). Cre Recombinase. Microbiology spectrum, 3(1), MDNA3–2014. https://doi.org/10.1128/microbiolspec.MDNA3-0014-2014

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).

Waldor, M. K., Tscha, H., & Genetics, M. (1996). A New Type of Conjugative Transposon Encodes Resistance to Sulfamethoxazole, Trimethoprim, and Streptomycin in Vibrio cholerae O139. 178(14), 4157–4165.

Weinreich, M. D., & Reznikoff, W. S. (1992). Fis plays a role in Tn5 and IS50 transposition. Journal of bacteriology, 174(14), 4530–4537. https://doi.org/10.1128/jb.174.14.4530-4537.1992

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).

Woods, K. C., Martin, S. S., Chu, V. C., & Baldwin, E. P. (2001). Quasi-equivalence in site-specific recombinase structure and function: crystal structure and activity of trimeric Cre recombinase bound to a three-way Lox DNA junction. Journal of molecular biology, 313(1), 49–69. https://doi.org/10.1006/jmbi.2001.5012

Wright, L. D., & Grossman, A. D. (2016). Autonomous Replication of the Conjugative Transposon Tn916. Journal of bacteriology, 198(24), 3355–3366. https://doi.org/10.1128/JB.00639-16

Yoshua, S. B., Watson, G. D., Howard, J. A. L., Velasco-Berrelleza, V., Leake, M. C., & Noy, A. (2020). A nucleoid-associated protein bends and bridges DNA in a multiplicity of topological states with varying specificity. BioRxiv, 2020.04.17.047076. https://doi.org/10.1101/2020.04.17.047076