Mechanistic insights into Xer recombination and conjugative transposition in *Helicobacter pylori*

Aleksandra Bebel
BSc(Hons), MSci

Submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy

University of Glasgow
Institute of Molecular, Cell and Systems Biology
College of Medical, Veterinary and Life Sciences

June 2015

Author’s declaration

I declare that, except where explicit reference is made to the contribution of others, that this thesis is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Signature.....

Printed Name .....

Summary

Site-specific recombinases of the Xer family are essential in most bacteria with circular chromosomes for the resolution of chromosome dimers arising after genome replication. In *Helicobacter pylori*, a gastric pathogen implicated in peptic ulcer disease and gastric cancer, the chromosome dimers are resolved by a single Xer recombinase, XerH. Interestingly, many *H. pylori* strains carry a second Xer recombinase, XerT, usually encoded on a large conjugative transposon TnPZ. Remarkably, XerT is not involved in chromosome dimer resolution, but was shown to be required for the mobilization of TnPZ *in vivo*.

In this thesis, I investigated the molecular mechanisms of XerH- and XerT-mediated recombination by combining X-ray crystallography with protein biochemistry, and microbiology. I solved the crystal structure of the XerH tetramer in a post-cleavage synaptic complex with its substrate DNA site, dif_H. To our knowledge, this is the first structure of an Xer recombinase bound to DNA. The structure and additional biochemical data provided key insights into the ordering and regulation of dif_H binding and first strand cleavage by XerH. Moreover, I investigated the regulation of XerH recombination by FtsK – a host factor usually required for Xer recombination – and found that XerH can resolve plasmids in the absence of FtsK in *E. coli*, but additional factors might be required for recombination of chromosome-borne dif_H sites. In the second part of this work, I studied the mechanism of XerT-mediated TnPZ transposition. XerT's binding and cleavage sites on transposon ends were mapped and XerT activity was reconstituted *in vitro* by trapping cleavage and strand exchange products. In addition, the complete TnPZ excision has been reconstituted *in vivo* in *E. coli*, indicating that XerT is sufficient to catalyze TnPZ mobilization. Based on the results, a testable model for TnPZ excision and integration was proposed.

In summary, this work provides valuable insights into the mechanisms of the two Xer recombinases of *H. pylori* and enhances our understanding of Xer recombination (a process essential for bacterial survival) and conjugative transposition (important in the spread of antibiotic resistance among bacteria), which in the future could help develop new therapeutic agents against deadly pathogens such as *H. pylori* or help control the spread of antibiotic resistance.

Acknowledgements

I would firstly like to thank Dr Orsolya Barabas for giving me the opportunity to study the exciting world of DNA recombination in her group. Thank you Orsolya for letting me follow my own goals while always being available for discussions, and for igniting my interest in X-ray crystallography, which in the end turns out to be pretty cool. I also wish to thank Professor Marshall Stark for all his help, advice, and guidance given over the course of my PhD, and once again for introducing me to site-specific recombination in the first place. Many thanks go the members of my Thesis Advisory Committee, Dr Teresa Carlomagno and Dr Nassos Typas, for participating in the yearly TAC meetings, during which they so generously shared their invaluable expertise.

I am grateful to the EMBL PhD programme for funding my studies and to the EMBL Graduate Office for its constant support with all the organizational issues arising over the years. I would like to thank the EMBL Core Facilities for their services and helpful advice regarding the experiments. Many thanks go to the ESRF staff for providing assistance during data collection at the synchrotron beamlines.

I would like to also acknowledge all the past and present members of the Barabas group, who always provided stimulating and enjoyable work environment. My special thanks go to Franka for our many interesting discussions and her patient help in my struggle to learn the secrets of X-ray crystallography; Ceci for taking care of everything and everyone in the lab; Anna for her help with fishing of insanely small crystals; Eike for his help during my first steps with protein work; Bianca and Pedro for preparing the first protein constructs used in this study; Banu and Reece for being great students; and Irma for being a great flatmate and making sure I always had a decent supply of Italian cookies during thesis writing. It was a great honour to work with you all!

My sincere thanks go to Dr Sebastian Glatt for all his crystallographic expertise he shared over the years, especially for collecting the first dataset from the XerH crystals, and later on for helping me to solve the structure. I would also like to thank Dr Gary Male, Dr Manuel Banzhaf, and Matylda Zietek for providing protocols for various experiments shown in this thesis. More thanks go to Dr Sean Colloms, Dr Femi Olorunniji, and Dr Stephanie Holt for discussions and advice they have given me during my work in Glasgow.

I am also grateful to all my friends at EMBL who made my time here a great, never-ending adventure. Especially the predoc class of 2011 – it was great to meet you all and together go through this journey! My utmost thanks go to the members of the College of St Flaubens, whose (not always) classy events kept life in Heidelberg the most interesting at all times.

I would like to thank my family, which always believed in my success and helped in many ways during this PhD work. And finally, I would like to thank Nik for all his love and support, and for painstakingly reading the draft of this thesis. Thank you for being super-awesome!

Table of contents

List of Tables

List of Equations

List of Figures

Abbreviations

ADP	Atomic displacement parameters
Ala/A	Alanine
Amp	Ampicillin
Arg/R	Arginine
Asn/N	Asparagine
ATP	Adenosine triphosphate
APS	Ammonium persulphate
BLAST	Basic Local Alignment Search Tool
bp	Basepair
CC	Correlation coefficient
Cm	Chloramphenicol
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytidine triphosphate
ddATP	Dideoxyadenosine triphosphate
ddCTP	Dideoxycytidine triphosphate
ddGTP	Dideoxyguanosine triphosphate
ddTTP	Dideoxythymidine triphosphate
dGTP	Deoxyguanosine triphosphate
DMSO	Dimethyl sulphoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
DTT	Dithiothreitol
dTTP	Deoxythymidine triphosphate
EDTA	Ethylenediaminetetraacetic acid
EMBL	European Molecular Biology Laboratory
Erm	Erythromycin
ESRF	European Synchrotron Radiation Facility
Fc	Calculated structure factors
Fo	Observed structure factors
Gln/Q	Glutamine
Glu/E	Glutamic acid
Gly/G	Glycine
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

His/H	Histidine
HJ	Holliday junction
I	Intensity
I3C	5-Amino-2,4,6-triiodoisophthalic acid
IDT	Integrated DNA Technologies
IMEX	Integrative Mobile Element Exploiting Xer
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Km	Kanamycin
KOPS	FtsK orienting polar sequences
LB	Lysogeny Broth
LE	Left end
LLG	Log-likelihood gain
Lys/K	Lysine
MAD	Multi-wavelength anomalous diffraction
MIR	Multiple isomorphous replacement
NCBI	The National Center for Biotechnology Information
NEB	New England Biolabs
OD_{600}	Optical density at 600 nm
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCMBS	*p*-chloromercuriphenylsulphonic acid
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
pI	Isoelectric point
PDB	Protein Data Bank
PMSF	Phenylmethanesulphonylfluoride
PZ	Plasticity zone
RE	Right end
R.m.s.	Root mean square
RNA	Ribonucleic acid
SAD	Single-wavelength anomalous diffraction
SAXS	Small-angle X-ray scattering
SDS	Sodium dodecyl sulphate
SEC	Size-exclusion chromatography
Se-Met	Seleno-methionine

Ser/S	Serine
SIR	Single isomorphous replacement
SIRAS	Single isomorphous replacement with anomalous signal
SOB	Super Optimal Broth
SSM	Secondary-structure matching
Str	Streptomycin
SUMO	Small Ubiquitin-like Modifier
TAE	Tris-acetate-EDTA
TB	Transformation buffer
TBE	Tris-borate-EDTA
TE	Tris-EDTA
TEMED	Tetramethylethylenediamine
ter	Replication terminus
Tet	Tetracycline
TFZ score	Translation function Z-score
Thr/T	Threonine
TLS	Translation/libration/screw
TnPZ	Transposable plasticity zone
Tris	Tris(hydroxymethyl)aminomethane
Trp/W	Tryptophan
TRX	Thioredoxin
Tyr/Y	Tyrosine
UV	Ultraviolet

1. General introduction

1.1 DNA recombination

All living organisms as well as many viruses carry their genetic information in the form of DNA: a string of combinations of four nucleotides (guanine, adenine, cytosine, and thymine) that can be replicated and transcribed by the cell machinery to give rise to products sustaining cell life: RNAs and proteins (Crick, 1958). The sequence of nucleotides – the genetic code – defines the variety and sequences of RNAs and proteins, determining contents of the cell and therefore its identity. One striking feature of the DNA is that this simple four-letter code can be easily modified through mutation of the bases by external factors or by mistakes arising during DNA replication, which provides the main driving force for evolution of new proteins, and as a consequence has had major impact on arising of new species, from simple anaerobic bacteria to humans (Fisher, 1930; Soskine and Tawfik, 2010).

However, *de novo* mutations are only one type of changes that the DNA can be subjected to. Of similar importance are larger DNA recombination events that include DNA inversions, duplications, and deletions, as well as insertions of foreign DNA molecules (such as viral DNA or transposons) and DNA exchange between organisms. If such DNA recombination processes could occur spontaneously, they would have disastrous effects on the stability of the DNA, leading to DNA breaking and knotting, rendering it no longer functional as a carrier of the genetic material. Therefore, DNA recombination processes occur in a controlled fashion and are only carried out by specific enzymes (recombinases) often acting under the control of other regulatory proteins (for example, DNA helicases) or the cell replication machinery.

Apart from its importance in introduction of genetic changes that promote differentiation and speciation, DNA recombination also plays an important role in processes such as DNA repair (homologous recombination and non-homologous end joining; Davis and Chen, 2013; Li and Heyer, 2008), maintenance of genome integrity, and propagation of viruses and mobile genetic elements (site-specific recombination and transposition; Curcio and Derbyshire, 2003; Grindley *et al*., 2006). Additionally, DNA recombination systems have

been adapted and employed in genetic research to create genome modifications (random and targeted deletions, insertions, and substitutions) of organisms ranging from bacteria such as *E. coli*, spanning model organisms such as yeast, worm, and fly, to mammals including mice and rats (reviewed in Segal and Meckler, 2013). The extensive use of various DNA recombination tools in model organism genetics is pushing forward the interest in finding more efficient and well-controllable DNA recombination systems that could eventually be confidently applied in human gene therapy.

1.2 Site-specific recombination

One type of DNA recombination is called site-specific recombination (reviewed in Grindley *et al.,* 2006). In site-specific recombination, two DNA sites with specific sequences are brought together and reciprocally exchanged by proteins termed site-specific recombinases, resulting in formation of two recombinant DNA sites (Figure 1-1a). The reaction proceeds by sequential cleavage and ligation of the two pairs of DNA strands with conservation of the phosphodiester bond energy. First, the DNA sugar-phosphate backbone is cleaved by a direct attack of the catalytic residue of the recombinase protein, and the scissile phosphate is transferred to the catalytic residue hydroxyl group, releasing a free DNA hydroxyl group. In a subsequent ligation reaction, the previously released DNA hydroxyl group attacks the phosphodiester bond between the protein and the partner DNA, which releases the recombinase and generates the recombined products (Figure 1-1b). The result of the recombination reaction, depending on the position and arrangement of the recombined sites, can be insertion, excision, or inversion of the DNA segment between two recombination sites, as shown in Figure 1-1c and d. The reactions resulting in insertion and excision are commonly employed to integrate or excise phage DNA or resolve circular DNA dimers arising after replication. Reactions resulting in inversion can be used to alter gene expression or to shuffle genes.

Site-specific recombinases can be divided into two families based on the catalytic residue used for the nucleophilic attack: the tyrosine recombinases, also known as the λ integrase family (Nunes-Duby *et al.,* 1998), and the serine recombinases, also known as the resolvase/invertase family (Smith and Thorpe, 2002). Members of both families are widely spread amongst bacteria.

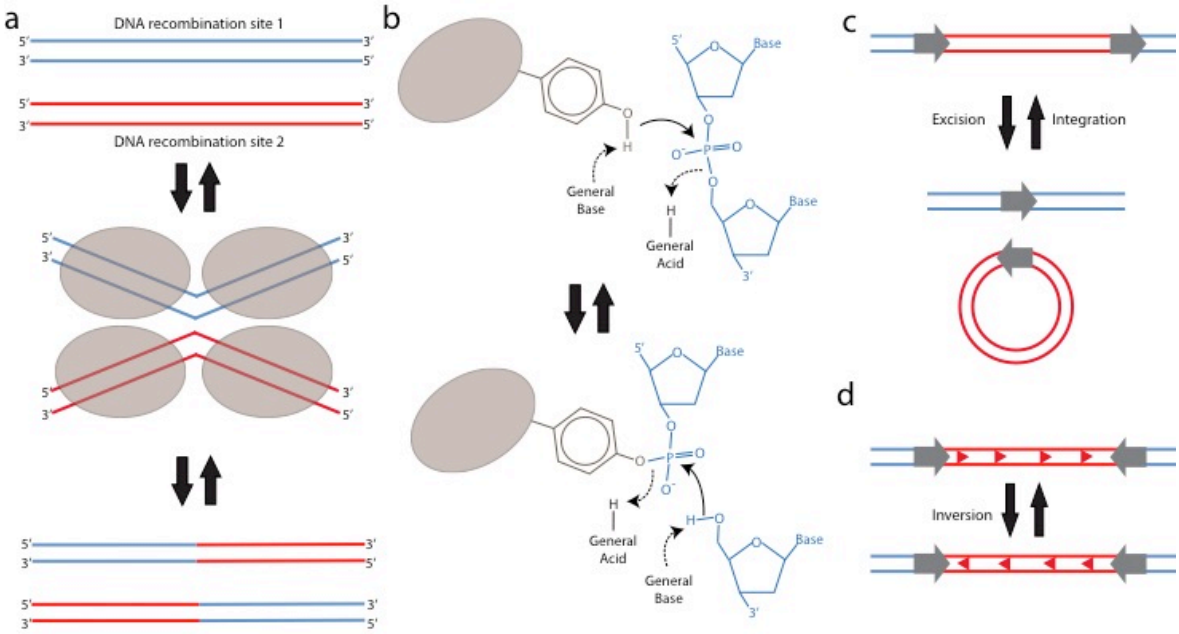


Figure 1-1: Features of site-specific recombination. a) Schematic view of site-specific recombination. Two DNA recombination partners (blue and red lines) are brought together in a synaptic complex of four recombinase monomers (grey spheres); the DNA sites are sequentially cleaved and ligated to produce two recombinant DNA sites. b) The mechanism of nucleophilic attack of the catalytic residue (here: tyrosine) of a site-specific recombinase (grey sphere) on the DNA sugar-phosphate backbone. The hydroxyl group of the tyrosine attacks the phosphate, leading to the formation of a phosphotyrosyl bond and displacement of the DNA 5'-hydroxyl group. The reaction is reversible: the DNA hydroxyl can re-attack the phosphotyrosyl bond, which will result in religation of the DNA and release of the recombinase. Based on Grindley *et al.,* 2006. c) and d) the possible outcomes of site-specific recombination reactions. If the recombination sites on one DNA molecule are arranged in a head-to-tail orientation, the DNA segment between the two sites will be excised (c). Reciprocal reaction between two sites on separate DNA molecules will result in DNA insertion (c). If the recombination sites on one DNA molecule are arranged in a head-to-head orientation, the DNA segment between the two sites will be inverted (d).

At first glance, both tyrosine and serine recombinases follow the same steps during the recombination reaction. Two monomers of a recombinase bind specifically to a defined recombination site, then two such bound sites are brought together to form a tetrameric synaptic complex. The recombinase cleaves and ligates the DNA, giving rise to recombinant DNA sites within the synaptic complex (Grindley *et al.,* 2006). However, the mechanistic details of recombination employed by the two families of recombinases differ greatly and the differences are summarized in Table 1-1. In serine recombination, the simultaneous nucleophilic attack of serine residues belonging to all four recombinase monomers breaks all four strands of the DNA and creates 5'-phosphoserine linkages. The DNA is cleaved with two-nucleotide stagger between the cleavage positions on the top and bottom strands, creating 3' overhangs (Reed and Grindley, 1981). The strands are then exchanged by a subunit rotation mechanism (bringing together two half-sites belonging to separate DNA recombination sites) and subsequently rejoined (Stark *et al.,* 1989). On the other hand, in tyrosine recombination two pairs of DNA strands are cleaved one pair at a

time resulting in formation of 3'-phosphotyrosyl bonds. The DNA is cleaved with six- to eight-nucleotide stagger, resulting in 5' overhangs. Unlike serine recombination, cleavage of the first strand in some tyrosine recombination systems might be possible in a dimeric complex, before assembly of the tetrameric synaptic complex (Voziyanov *et al.,* 1996). Despite these differences, both recombinase families produce equivalent, highly specific recombination products and therefore the members of the families are involved in similar biological functions.

Table 1-1: Differences between mechanisms of tyrosine and serine site-specific recombination.

Feature	Tyrosine recombinases	Serine recombinases
Catalytic residue	Tyrosine	Serine
Complex required for cleavage activity	Dimer bound to one site or tetrameric synaptic complex	Tetrameric synaptic comples
Number of DNA strands broken at a time	2	4
Covalent intermediate formed	3' phosphotyrosine	5' phosphoserine
Distance between cleavage sites on one DNA site	6-8 nucleotides	2 nucleotides
DNA overhang after cleavage	5'	3'
Mechanism	Sequential cleavage with Holliday Junction intermediate	Simultaneous cleavage with subunit rotation

1.3 Tyrosine recombination

1.3.1 Characteristics of tyrosine recombinases

All tyrosine recombinases consist of at least two distinct protein domains: the carboxy-terminal (C-terminal) catalytic domain and the amino-terminal (N-terminal) DNA-binding domain (Nunes-Duby *et al.,* 1998). The two domains form a C-shaped clamp that encircles the recombined DNA. In some cases, these two domains are accompanied by an additional domain binding accessory DNA sequences or having unknown function. The DNA binding domain shows little conservation between different tyrosine recombinases, consistent with a broad DNA sequence specificity within this family of proteins. The catalytic domain of tyrosine recombinases shows a well-conserved fold despite the low (15-30%) amino acid sequence identity between different family members (Nunes-Duby *et al.,* 1998). This domain is shared with another family of DNA-cleaving enzymes, the type IB topoisomerases that are involved in releasing DNA supercoiling (Cheng *et al.,* 1998). The

catalytic domain invariably contains the catalytic tyrosine residue and four other residues: two arginines, a histidine, and a lysine, together forming the RKHRY motif that constitutes the active site of the recombinase (Figure 1-2; Esposito and Scocca, 1997; Gopaul and Duyne, 1999; Nunes-Duby *et al.,* 1998). The additional residues in the catalytic site are required to coordinate the scissile phosphate, promote nucleophilic attack of the tyrosine hydroxyl, accept the proton from it (general base), stabilize the transition state, and protonate the released free hydroxyl (general acid; Krogh and Shuman, 2000, 2002; Nagarajan *et al.,* 2005; Stivers *et al*., 2000; Tian *et al.,* 2005). 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 resulted in catalytically inactive protein in all cases studied (summarized in Nunes-Duby *et al.*, 1998). In most recombinases, an additional histidine or tryptophan residue contributes to the assembly of the active site; however, this residue is not well conserved and can be mutated without the severe phenotype seen when any other residue of the RKHRY catalytic pentad is mutated (summarized in Nunes-Duby *et al.*, 1998).

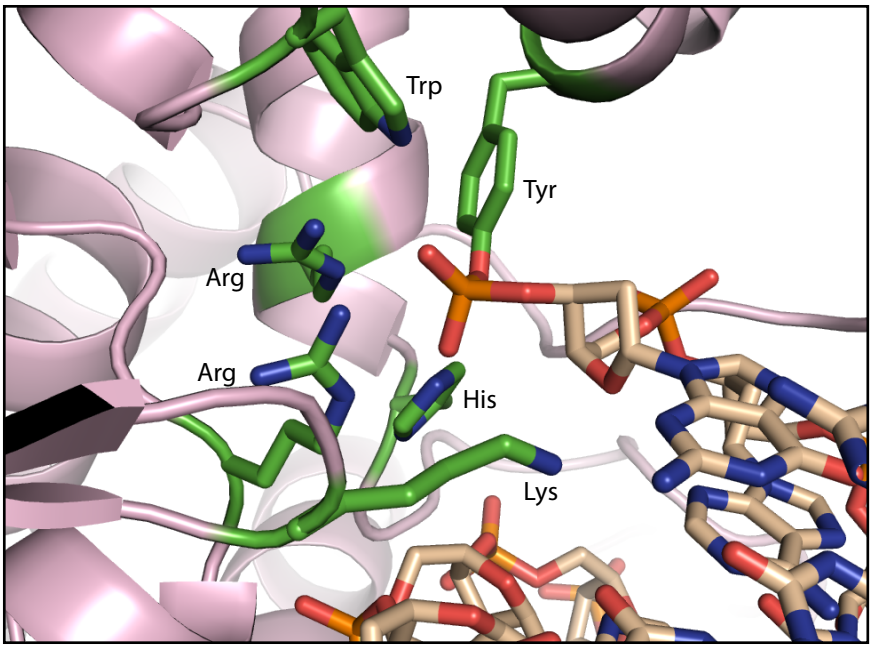


Figure 1-2: Cartoon representation of the catalytic site of tyrosine recombinases, shown for the Cre recombinase (Protein Data Bank (PDB) entry 1CRX, Guo *et al.,* 1997). Highly conserved catalytic residues are shown in green. The DNA and the catalytic residues of Cre are shown in stick representation with atomic colouring as follows: oxygen – red; nitrogen – blue; phosphorus – orange; carbon – beige (DNA) or green (protein). The catalytic tyrosine is attached to the DNA phosphate by a phosphotyrosyl bond. The two arginines, a histidine, a lysine, and a tryptophan residues support catalysis.

1.3.2 Mechanism of tyrosine recombination

Tyrosine recombinases bind to well-defined DNA sites that consist of two binding arms of around 13 basepairs (bp) separated by a central region of 6-8 bp. These sites consist of specific sequences different for each tyrosine recombinase and essential for the activity of that recombinase. A schematic overview of the mechanism of tyrosine recombination, derived from many years of structural and biochemical characterization of these enzymes, is shown in Figure 1-3 (Gopaul and Van Duyne, 1999). Two monomers of the recombinase bind to one recombination site, each monomer forming a C-shaped clamp around the DNA double helix. The two recombinase-bound sites are brought together to form a tetrameric synaptic complex that is competent for recombination. In this complex, at any given time only two non-adjoining monomers are arranged in a catalytically active conformation allowing DNA cleavage, known as the "half-of-the-sites reactivity". The two active monomers can cleave the DNA at the 5' boundary of the central region by nucleophilic attack of the tyrosine on the DNA sugar-phosphate backbone, resulting in the formation of a covalent 3'-phosphotyrosyl bond and release of a free 5'-hydroxyl. The DNA is resealed by a nucleophilic attack of the free 5' hydroxyl on the phosphotyrosyl bond of the recombination partner DNA. This "strand exchange" step leads to formation of a four-way DNA junction called a Holliday Junction (HJ; Holliday, 1964). Resolution of the HJ into recombination products follows when the second pair of monomers adopts an active conformation as a result of a so-called "isomerization switch" (Gopaul and Duyne, 1999). The second pair of monomers performs analogous cleavage and rejoining reactions as above, resulting in the formation of two recombined DNA sites.

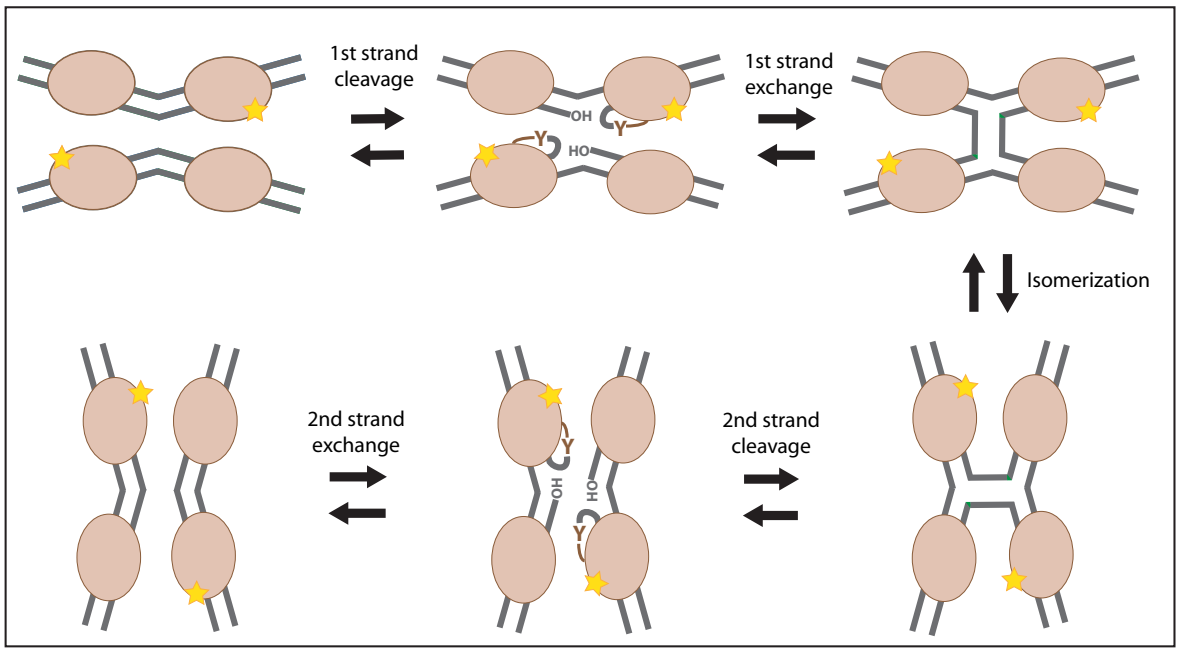


Figure 1-3: The mechanism of tyrosine recombination. Synapsis of two recombinase-bound sites is followed by cleavage and ligation of the first two DNA strands by two active recombinase monomers (marked by stars), which leads to formation of a four way HJ intermediate. Resolution of the HJ follows when the second pair of monomers becomes active as a result of an isomerisation event. Based on Gopaul and Van Duyne, 1999.

1.3.3 Regulation of tyrosine recombination

Tyrosine recombinases can cleave and ligate chromosomal DNA. If uncontrolled, this could have detrimental effects on the host cell: introducing DNA breaks, inversions, and duplications, complicating DNA topology (and therefore preventing DNA replication and segregation), changing of gene copy number, and uncontrolled propagation of mobile genetic elements, based on the known functions of tyrosine recombinases (reviewed in Grindley *et al.,* 2006). Because of this, tyrosine recombinase activity is tightly controlled by a variety of mechanisms throughout the recombination process.

The fact that tyrosine recombinases can only bind to specific sequences ensures that only recombination on legitimate substrates can occur. Additionally, HJ formation and resolution requires sequence identity between the central regions of the two recombining DNA substrates (Burgin and Nash, 1995; Nunes-Duby *et al.,* 1995), strengthening the specificity of recombination. Furthermore, in most cases recombination can only occur in the fully assembled tetrameric synaptic complex, ensuring that potentially dangerous DNA cleavage is performed only when the reaction is ready to proceed all the way to the

products, although in some cases first strand cleavage of a single site bound by a recombinase dimer might be possible (Ghosh *et al.,* 2005; Voziyanov *et al.,* 1996).

The mechanism of "half-of-the-sites reactivity" (as shown in Figure 1-3) ensures that problematic double-strand DNA breaks are avoided, as only one strand of each DNA molecule in the synaptic complex is cleaved at a time. This mechanism is enforced by two different conformations of the protein in the synaptic complex: one active conformation, with the catalytic tyrosine placed in line for direct attack of the DNA sugar-phosphate backbone, and one inactive conformation where the tyrosine is placed too far from the DNA backbone for nucleophilic attack (Guo *et al.,* 1997). These two conformations are brought about by interactions between protruding C-terminal segments of one monomer and the catalytic domain of the neighbouring monomer, repeated in a cyclic fashion around the tetrameric complex (Figure 1-4; Biswas *et al.,* 2005; Chen *et al*., 2000; Guo *et al.,* 1997). These regulatory protein segments (α-helix or β-strand) contain or are directly attached to the catalytic tyrosine and therefore their position is crucial for determining the position of the tyrosine. The cyclic assembly functions such that when two non-adjoining monomers are positioned in the active conformation, their C-terminal segments affect the neighbouring monomers so that they assume the inactive conformation, and *vice versa* (Ferreira *et al*., 2003). This mechanism ensures that only two monomers in the synaptic complex are active at any given time – which ones is governed by the DNA bending upon protein binding, or DNA binding order (Grindley *et al.,* 2006).

Finally, the requirement for HJ isomerization provides another level of recombination regulation since HJ isomerization governs whether the HJ intermediates are converted into products or revert back to substrates. The mechanism of the HJ isomerization step is not fully understood and it appears that there is no single common mechanism involved, but rather that various recombinases employ different mechanisms to facilitate HJ isomerization (discussed in Grindley *et al*., 2006). For example, HJ resolution in λ integrase recombination requires the arm-binding domain and its bound arm site (Biswas *et al*., 2005), while isomerization in Cre is favoured by the nearly square planar conformation of the synaptic complex throughout the reaction (Gopaul and Van Duyne, 1999). In addition, regulation of HJ isomerization by specific host factors ensures that recombination only happens in the desired spatial and temporal context.

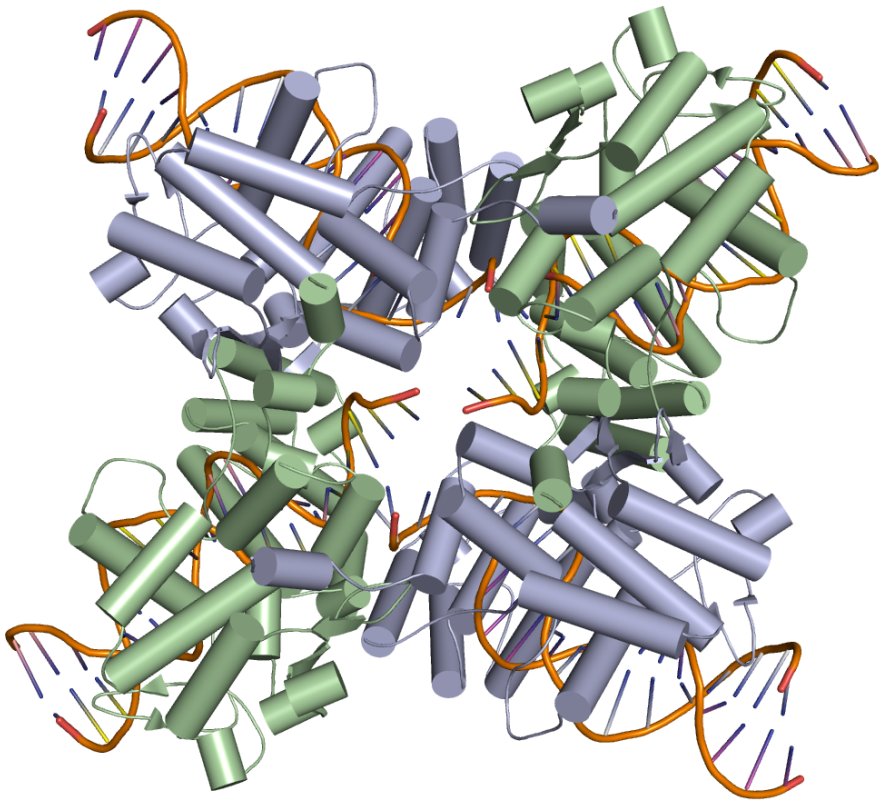


Figure 1-4: Cartoon representation of the circular arrangement of tyrosine recombinases within the synaptic complex, shown for the Cre recombinase (PDB 1CRX, Guo *et al.,* 1997). The two monomers in an active conformation are shown in blue, the two monomers in an inactive conformation are shown in green. The α-helices are shown as cylinders. The regulatory C-terminal α-helices can be seen protruding from each monomer into the body of a neighbouring monomer in a cyclic fashion, contributing to the allosteric regulation of subunit activity ("half-of-the-sites reactivity" phenomenon).

1.3.4 Members of the tyrosine recombinase family

Members of the tyrosine recombinase family are widely spread amongst prokaryotes and Archaea, but are also present in eukaryotes, for example in yeast. More than 100 members of the family have been identified (Nunes-Duby *et al.,* 1998), with many more expected to be found. Some examples of better-studied family members, together with their biological functions, are shown in Table 1-2. The most studied members of the family include the λ integrase, Cre, and Flp, as well as XerC/D (discussed later).

Table 1-2: Examples of tyrosine recombinases and their function.

Recombinase	Source	Biological function
λ integrase	Bacteriophage λ	Excision and integration of phage λ genome
Integrase of Tn916/Tn1545/Tn1549	Bacterial conjugative transposons	Excision and integration of conjugative transposons
HP1 integrase	Bacteriophage HP1	Excision and integration of phage HP1 genome
Cre	Bacteriophage P1 plasmids	Phage P1 plasmid dimer resolution
XerC/D	Bacterial chromosomes	Chromosome dimer resolution; Excision and insertion of mobile elements such as phage CTX
TnpI of Tn4430	Transposon Tn4430	Resolution of cointegrates formed during Tn4430 transposition
FimB, FimE	*E. coli* chromosome	Alteration of gene expression through inversion of *fimS* site
Flp	Yeast 2-μm plasmid	Regulation of copy number of yeast 2-μm plasmid

1.3.4.1 λ integrase

λ integrase, to which the tyrosine recombinase family owes its other name, binds and recombines sites *attP* (present on bacteriophage λ) and *attB* (present on the bacterial chromosome). As a result of recombination, the phage is integrated into the bacterial genome, flanked by two newly formed attachment sites *attL* and *attR*, enabling lysogenic phage maintenance (Landy, 1989). The transition to lytic growth results in excision of the phage by λ integrase-mediated recombination of *attL* and *attR* and reconstruction of the *attP* and *attB* sites. In addition to the previously described DNA-binding and catalytic domains, λ integrase and related phage integrases contain an additional N-terminal DNA-binding domain called the arm-binding domain, which recognizes additional regulatory DNA sequences within the phage genome, called arm sites (Radman-Livaja *et al*., 2003). Successful recombination by λ integrase depends on intricate topological interactions between the integrase protein, both arm and core binding sites, and host factors assembled together in a highly regulated recombination complex (Biswas *et al*., 2005; Franz and Landy, 1995).

1.3.4.2 Cre

The Cre recombinase is a well-studied recombinase that is extensively used as a genetic tool, particularly for conditional gene inactivation or other DNA rearrangements in mice and other model organisms (Nagy, 2000). Cre is considered a model tyrosine recombinase due to an abundance of available biochemical data and numerous crystal structures capturing various steps of the recombination pathway (Ennifar *et al.,* 2003; Gopaul *et al.,* 1998; Guo *et al*., 1997, 1999; Martin *et al*., 2002). Cre is encoded on bacteriophage P1 and

it is required for maintaining the monomeric genome of the phage (Abremski *et al*., 1983). It acts by recombining two *loxP* sites that are present on the dimeric form of the phage genome (Hoess *et al.,* 1982). Cre consists of a catalytic and a DNA binding domains only, and does not require any host factors or additional regulatory DNA sequences for successful recombination (Abremski and Hoess, 1984).

1.3.4.3 Flp

Flp is one of a few known eukaryotic tyrosine recombinases. It is encoded by the 2 μm plasmid of the yeast *Saccharomyces cerevisiae*, where it contributes to the regulation of plasmid copy number by inverting a DNA segment flanked by two FRT recombination sites, resulting in changed orientation of the replication fork and consequently in multiple replication events (Volkert and Broach 1986). Similarly to Cre, Flp has been used as a genetic tool, for example in mouse genetics. However, Flp shares very low sequence similarity with other tyrosine recombinases (13% sequence identity between Cre and Flp catalytic domains) and, strikingly, the catalytic tyrosine is supplied to the active site *in trans* from a neighbouring monomer (Dixon *et al.,* 1995).

1.4 Chromosomal dimer resolution by Xer recombinases

Chromosome dimers occasionally arise in organisms with circular genomes such as bacteria and Archaea. This phenomenon affects about 10-15% of dividing *Escherichia coli* cells and is only observed in strains containing a functional *recA* gene, suggesting that formation of chromosomal dimers is directly related to homologous recombination (Perals *et al.,* 2000; Steiner and Kuempel, 1998). Homologous recombination occurs as a DNA repair mechanism during genome replication when a stalled replication fork needs to be restarted, as shown in Figure 1-5. Homologous recombination machinery promotes the strand invasion and formation of a HJ intermediate, which is resolved by the RuvABC complex into functional replication fork that can be restarted (Seigneur *et al.,* 1998). Depending on the orientation of HJ resolution by the RuvABC complex, homologous recombination can result in a "crossing over" event and formation of the dimeric chromosome.

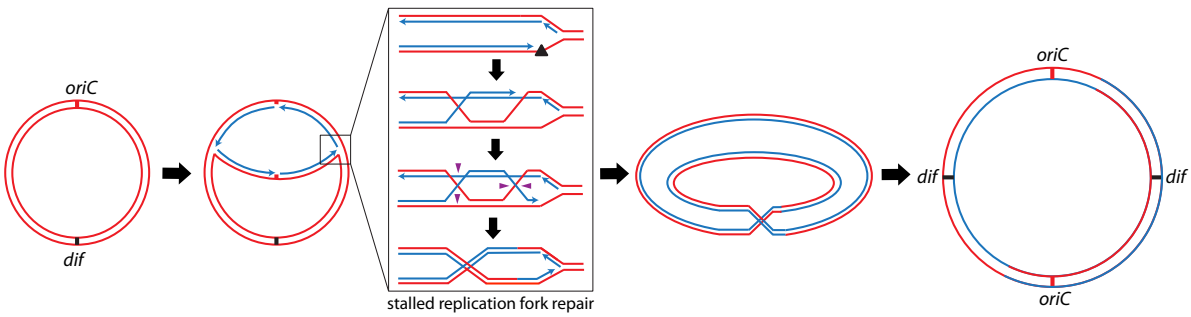


Figure 1-5: Formation of chromosome dimers as a result of a homologous recombination event at the replication fork. Replication of bacterial circular chromosomes starts at the origin of replication *oriC* and proceeds bidirectionally towards the terminus region containing a *dif* site. If the replication fork is stalled, the homologous recombination machinery will restart it by crossing-over the parent (red) and daughter (blue) strands, as shown in the black box. These events might result in formation of two daughter chromosomes that are linked at the cross-over region, forming chromosome dimers that need to be resolved before cell division.

The dimeric form of the chromosome prevents normal DNA segregation and successful cell division (discussed in Lesterlin *et al.,* 2004). In cells with chromosomal dimers, DNA replication and segregation cannot proceed until completion, leaving the two daughter chromosomes linked by DNA passing through the division septum (Hendricks *et al.,* 2000). As cell division proceeds, the DNA is cut generating double-strand breaks in the chromosome. This double strand break is then processed by RecBCD, resulting in DNA degradation and induction of the SOS response, which inhibits cell division (Hendricks *et al.,* 2000). This results in formation of multi-cellular filaments and eventually cell death.

As a result, bacteria and Archaea with circular genomes employ a mechanism dedicated to the resolution of chromosome and plasmid dimers into monomers prior to genome segregation and cell division (Blakely *et al*., 1991; Summers and Sherratt, 1984). The key enzymes involved in this process are called Xer recombinases and belong to the tyrosine family of site-specific recombinases. Recombinases of the Xer family can be found in most bacteria with circular chromosomes (Carnoy and Roten, 2009), and their mutations exhibit phenotypes consistent with an inability to resolve chromosomal dimers (Blakely *et al.,* 1991; Blakely *et al.*, 1993). In addition, in some species the mutations in Xer genes are associated with unexplained phenotypes (for example, lack of viability of XerD mutants in *Staphylococcus aureus*) suggesting that Xer recombinases might play additional roles in other cellular processes (Chalker *et al.*, 2000).

Interestingly, Xer recombinases are the only site-specific recombinases identified to date that are essential for bacterial growth and have a clear function associated with chromosome maintenance.

1.5 *E. coli* XerC/D as a model Xer recombination system

The process of Xer recombination has been extensively studied in *E. coli* where it is carried out by two closely related (37% sequence identity) proteins XerC and XerD (Blakely *et al.,* 1993). Absence of either Xer protein results in abolished recombination *in vitro* and lack of dimer resolution *in vivo* (Blakely *et al*., 1991; Kuempel *et al.,* 1991). This system is unique amongst the site-specific recombinases in that it requires two proteins acting together as heterodimers for productive recombination. However, the exact reason for employing two distinct proteins in this process is not fully understood.

XerC was first identified as a recombinase that resolves plasmid dimers, which arise during plasmid replication in a similar manner as chromosome dimers (Colloms *et al*., 1990). Two prominent examples of plasmid dimer resolution by XerC/D recombination come from plasmids ColE1, with its recombination site *cer*, and pSC101, with its site *psi*. Each site is present in only one copy in the respective plasmid (Cornet *et al*., 1994; Summers and Sherratt, 1984). In both cases, accessory host factors and DNA sequences are required for productive recombination by XerC/D, ensuring that only two sites present on the same plasmid molecule can be recombined, resulting in plasmid dimer resolution while preventing plasmid dimer formation.

It was later shown that XerC/D recombination is also responsible for chromosome dimer resolution in *E. coli*. The chromosomal XerC/D recombination site is called *dif. E. coli* genome contains a single *dif* site, located in the replication terminus region of the chromosome (Kuempel *et al.,* 1991). The site shows high conservation among bacteria despite highly variable genomic context surrounding the site (Carnoy and Roten, 2009). No accessory DNA sequences are required for XerC/D recombination at *dif* sites (Leslie and Sherratt, 1995). To initiate recombination, XerC and XerD bind as a heterodimer to one 28-bp *dif* site (Kuempel *et al*., 1991). The *dif* site consists of two 11-bp half-sites that constitute two recombinase binding sites (the left half-site for XerC monomer and the right half-site for XerD), to which the recombinases bind cooperatively (Figure 1-6a; Blakely *et*

al., 1993). The two half-sites are separated by a central region of six bp that marks the XerC/D cleavage positions (Blakely *et al.,* 1997). The 28-bp *dif* site is sufficient for XerC/D recombination *in vivo* and *in vitro* (Leslie and Sherratt, 1995).

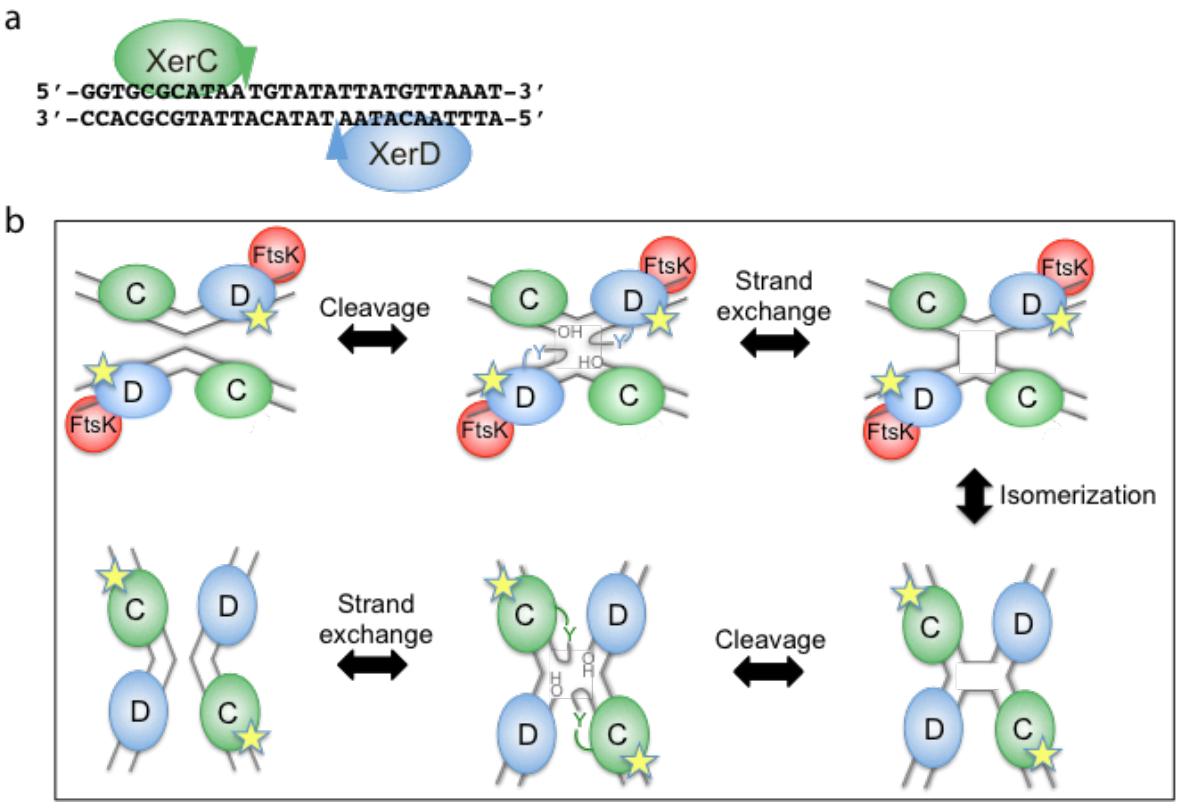


Figure 1-6: Features of XerC/D recombination. a) The XerC/D recombination site, *dif*. The site consists of two binding arms, each binding one recombinase only. The site is cleaved with a six-nucleotide stagger as indicated by triangles. b) Mechanism of XerC/D recombination. Synapsis of two recombinase-bound sites is followed by activation of XerD by direct interaction with FtsK. XerD as the active monomer (indicated by stars) performs DNA cleavage and ligation of the first two DNA strands, leading to formation of a four-way HJ intermediate. Resolution of the HJ follows when XerC monomers become active following a HJ isomerization event.

1.6 Mechanism of XerC/D recombination

The mechanistic model of XerC/D recombination (Figure 1-6b), derived from the model of tyrosine recombination and supported by biochemical studies of XerC/D, describes a step-wise process, in which initially XerC and XerD bind their respective *dif* half-sites to form a synaptic complex consisting of an XerC/D heterotetramer and two *dif* sites (Blakely *et al.,* 1993). As for other tyrosine recombinases, the complex is expected to contain two protein monomers arranged in an active conformation while the other two protein monomers remain inactive, allowing for highly orchestrated cleavage of four *dif* strands, two strands

at a time. In case of XerC/D recombination, one of the proteins (for example XerC) shows the active conformation while the other (XerD) remains inactive. In productive recombination of chromosomal *dif* sites, the cleavage of the first strands is performed by XerD monomers upon activation by accessory protein FtsK (Aussel *et al.,* 2002; Grainge *et al.,* 2011). As in other tyrosine recombinases, the hydroxyl group of the catalytic tyrosine of the XerD pair of monomers attacks the DNA phosphate backbone, leading to the formation of a phosphotyrosyl bond and leaving a free 5' hydroxyl group. This hydroxyl group attacks the phosphotyrosyl bond on the other *dif* site leading to strand exchange and formation of the HJ intermediate. As a result of conformational changes within the HJ intermediate (isomerization), the XerC pair of monomers acquires the active conformation. A second set of cleavage and strand exchange reactions performed by the XerC monomers on the second pair of DNA strands results in resolution of the HJ and formation of two recombined products (Grainge *et al*., 2011).

Altogether, the proposed mechanism of XerC/D recombination at *dif* follows the canonical tyrosine recombination pathway, with a distinct feature of using two homologous recombinases that are regulated by an accessory host factor FtsK. It is worth mentioning that XerC/D recombination at *cer* only requires catalytic activity of XerC (although both XerC and XerD have to be present to form the synaptic complex) and is independent of FtsK (Colloms *et al.,* 1997; Recchia *et al.,* 1999). In reactions at *cer*, XerC performs the first cleavage and strand exchange, and the resulting HJ is resolved by the host machinery via an unknown mechanism.

1.7 Regulation of XerC/D by FtsK

In order to assure that during XerC/D recombination at the chromosomal *dif* site only the desired recombination products are produced, specific pairs of monomers within the recombination complex have to be sequentially active. This is achieved through specific activation of XerD by an accessory protein, FtsK. FtsK is a well-conserved protein essential for cell division, although its function in Xer recombination differs between bacterial species (Sciochetti *et al.,* 2001). In *E. coli,* FtsK (Figure 1-7) contains an N-terminal transmembrane domain (≈200 amino acids) containing five transmembrane helices that anchor the protein at the division septum, a long (≈600 amino acids) linker of unclear function, and a C-terminal ATP-dependent helicase domain (≈500 amino acids).

Deletions of FtsK N-terminal domain in *E. coli* are lethal (Wang and Lutkenhaus, 1998), whereas deletions of the C-terminal domain result in a phenotype consistent with inability to resolve chromosome dimers (Recchia *et al.,* 1999; Steiner *et al.,* 1999).

Consistent with this phenotype, the C-terminal helicase domain ($FtsK_C$, Figure 1-7) was shown to be sufficient *in vivo* and *in vitro* for enabling XerC/D recombination at *dif* (Aussel *et al.,* 2002; Barre *et al.,* 2000). This domain consists of α, β, and γ subdomains, of which the α and β subdomains form the main helicase ring, while the γ domain is important for DNA binding and interactions with the Xer recombinases. Further studies indicated that only the very C-terminal γ subdomain of the helicase domain is required to activate XerD cleavage *in vitro* and *in vivo*, while the α and β subdomains are required to maintain the correct DNA topology during XerC/D recombination (Grainge *et al*., 2011). The specific interaction between XerD and FtsKγ in *E. coli* has been mapped to residues 282-292 (RQxxQQ motif) on the XerD, and residues 1277-1282 (TEKRKA) of the FtsKγ (Yates *et al.,* 2006). Mutations of these residues result in abolished XerC/D recombination *in vitro* and *in vivo*. Additionally, the interaction between FtsK and XerD in *E. coli* is considered species-specific as the FtsK homologue from *Haemophilus influenza* does not support *E. coli* XerC/D recombination and *vice versa* (Yates *et al.,* 2003).

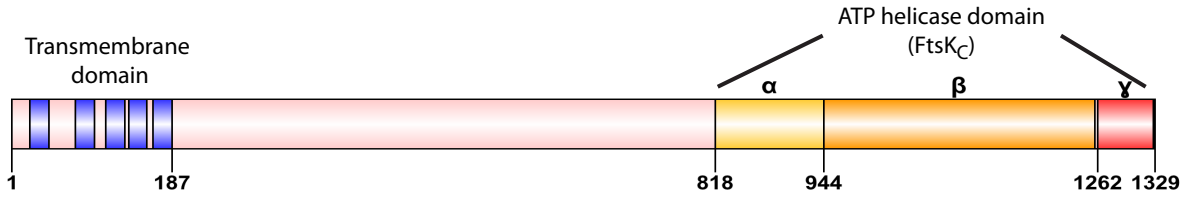


Figure 1-7: Domain representation of the FtsK protein from *E. coli*. The transmembrane helices within the N-terminal transmembrane domain are shown in blue. The C-terminal ATP helicase domain ($FtsK_C$) consists of three subdomains α (yellow), β (orange), and γ (red).

The exact mechanism by which FtsK regulates XerC/D recombination at *dif* is still not fully understood. It was proposed that FtsK is guided to *dif* by FtsK-orienting polar sequences (KOPS). The KOPS are octameric DNA repeats that show an orientation biased in one direction, with the bias switched in the replication terminus region as if pointing towards the *dif* site (Levy *et al.,* 2005). FtsK binds to KOPS and due to their polarity the helicase domain translocates the DNA until it encounters a *dif* site (Graham *et al.,* 2010; Sivanathan *et al.,* 2009). Since FtsK is anchored at the septum through its N-terminal domain, this brings two *dif* sites, one from each daughter chromosome, to the division

septum. At this point, the *dif* sites are considered to be bound by one XerC/D heterodimer each and synapsed in a conformation that allows only unproductive cleavage, HJ formation, and HJ resolution back to pre-recombination substrates by XerC monomers (Barre *et al.,* 2000). Through an unknown mechanism, interaction with FtsKγ remodels the synaptic complex to enable cleavage by XerD and productive recombination can proceed.

Altogether, regulation of XerC/D recombination by FtsK ensures unidirectionality of the process, tight spatial and temporal coupling of cell division with chromosomal dimer resolution, and correct localization of the chromosome for post-recombination genome segregation.

1.8 Integrative mobile elements exploiting Xer (IMEX)

Although the indispensable role of Xer recombinases is to resolve chromosome and plasmid dimers, these proteins (in particular XerC and XerD) are also involved in integration of various mobile genetic elements such as CTXϕ, VGJϕ, or TLC collectively termed integrative mobile elements exploiting Xer (IMEX) into the bacterial genome (Das *et al*., 2013). Each of these elements hijacks bacterial Xer recombinases to integrate its genome into the bacterial chromosome using self-contained Xer binding sites. In the single-stranded DNA bacteriophage CTX from *Vibrio cholerae* the XerC/D binding site is formed upon formation of a DNA hairpin (Val *et al.,* 2005). The site contains a small hairpin within the central region between the two binding arms while the binding arms themselves closely resemble the binding arms for XerC and XerD in the chromosomal *dif* site (Val *et al.,* 2005). In turn, bacteriophage VGJ is a double-stranded DNA virus that contains a more canonical XerC/D recombination site (Das *et al.,* 2013). In both cases, the central region is not homologous to that of *V. cholerae dif* site. Interestingly, toxin-linked cryptic elements (TLC) from *V. cholerae* typically contain XerC binding site and the central region that are fully conserved with comparison to *dif*, while there is no binding site for XerD (Das *et al*., 2013).

Such recombinase hijacking was not observed for other tyrosine recombinases, which are normally associated with only one recombination function and are encoded on the element concerned; for example, Cre exclusively recombines *loxP* sites of the P1 plasmid, whilst λ integrase inserts and excises phage λ DNA by interactions with specific *att* sites only.

XerC and XerD can interact with genomic *dif* sites to resolve chromosome dimers, *cer* and *psi* sites to resolve plasmid dimers, and various sites on mobile elements to insert and excise them to and from the bacterial genome. This renders the Xer recombinase family exceptionally diverse in terms of function and specificity amongst site-specific recombinases.

1.9 Single protein Xer recombination systems

The XerC/D system from *E. coli* provides a paradigm for Xer recombination, but not all bacteria employ an equivalent system with two Xer proteins resolving chromosome dimers. In fact, a large number of species employs phylogenetically distant single-recombinase systems such as XerS in *Lactococcus* species or XerH in *Helicobacter* and *Campylobacter* species (Figure 1-8; Carnoy and Roten, 2009; Le Bourgeois *et al.,* 2007; Leroux *et al.,* 2013). The studies of these systems suggest that they might employ a different mechanism to ensure correct ordering of DNA binding and cleavage events (Leroux *et al.,* 2013; Nolivos *et al.,* 2010). Similarly, the role of FtsK in these systems is yet to be determined.

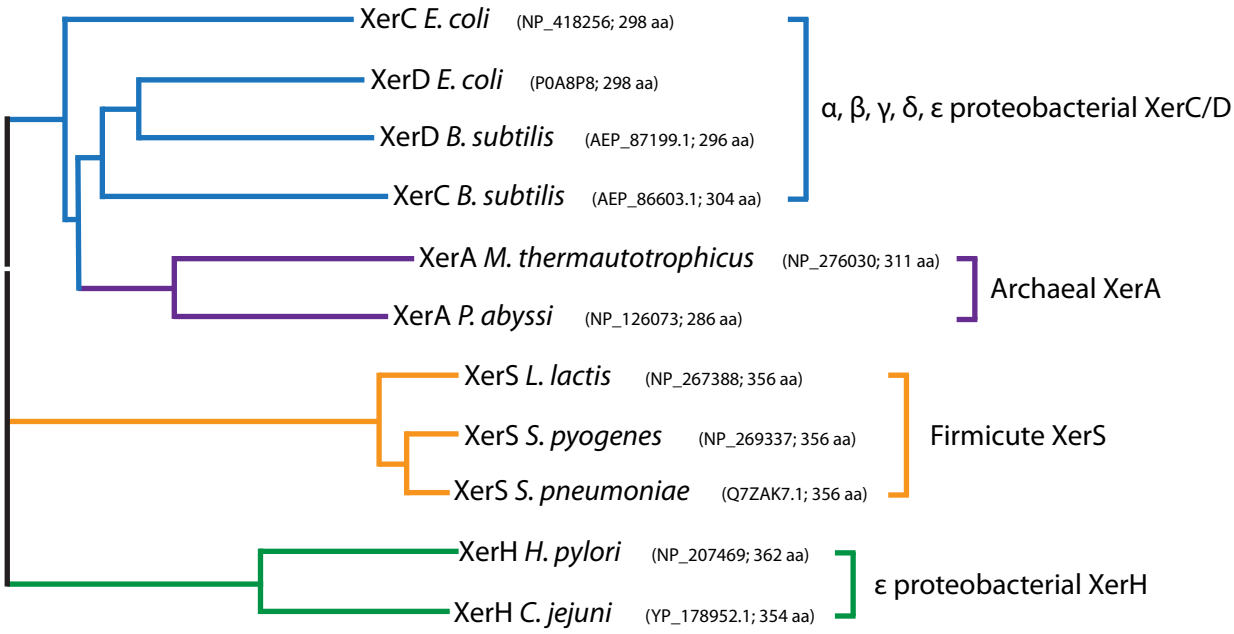


Figure 1-8: Phylogenetic tree of Xer recombinases. The National Center for Biotechnology Information (NCBI) accession number and the size of the protein in amino acids (aa) are shown in brackets. The phylogenetic tree was prepared in MEGA6 (Tamura *et al.,* 2013) using the Neighbour-Joining method (Saitou and Nei, 1987). Phylogenetic branches are marked with different colours. The length of each branch corresponds to the phylogenetic distance.

1.9.1 XerS

An example of a single-recombinase Xer system can be found in *Lactococci* and *Streptococci*, where the recombinase XerS acts on a conserved site dif_{SL}, identified by comparative genomics based on its localization and a single appearance in streptococcal genomes (Le Bourgeois *et al*., 2007). In a range of *in vivo* experiments, XerS was shown to be the only streptococcal recombinase required for chromosome dimer resolution by dif_{SL} recombination (Le Bourgeois *et al*., 2007). XerS recombination was also shown to depend on the presence of *E. coli* $FtsK_C$ in recombination assays in this bacterium, or the streptococcal $FtsK_C$ in *S. pneumoniae* recombination assays (Le Bourgeois *et al.,* 2007). As $FtsK_C$ from *E. coli* was sufficient to support XerS recombination, it was suggested that the XerS-$FtsK_C$ interaction was not species-specific. As XerS does not contain the sequences corresponding to the XerD region that directly interacts with FtsKγ (the C-terminal subdomain of $FtsK_C$), it seems likely that XerS does not require direct activation by interactions with FtsKγ. Instead, $FtsK_C$ might be important in spatial and temporal regulation of XerS recombination, as well as in ensuring resolution of chromosome dimers rather than their formation (Le Bourgeois *et al.,* 2007). Finally, XerS from *Lactococcus lactis* was shown to bind and cleave dif_{SL} *in vitro* in an asymmetric fashion with a bias towards one of the dif_{SL} binding arms; however, the full recombination reaction could not be reconstituted *in vitro* even in the presence of $FtsK_C$ (Nolivos *et al.,* 2010).

1.9.2 XerH

XerH found in *Campylobacter jejuni* and *Helicobacter pylori* species is another example of a single Xer recombinase (Debowski *et al.*, 2012; Leroux *et al.,* 2013). *Campylobacter* XerH was shown to bind and cleave a computationally predicted dif_H site *in vitro* (Leroux *et al.,* 2013). *Campylobacter* XerH supported recombination of two dif_H sites located on a plasmid in *E. coli*, even in the absence of *E. coli* $FtsK_C$, suggesting that XerH can perform cleavage and strand exchange of at least one pair of DNA strands without $FtsK_C$ (Leroux *et al.,* 2013). However, in these assays recombination products were present in only small amounts after a large number of *E. coli* generations. Recombination by XerH could not be detected in *in vitro* conditions (Leroux *et al.,* 2013).

1.9.3 XerA

In the kingdom of Archaea, all identified Xer recombination systems employ a single Xer recombinase called XerA (Cortez *et al*., 2010). Amongst these, XerA from *Pyrococcus abyssi* has been studied in most detail. This XerA was shown to recombine archaeal *dif* sites located on plasmids *in vivo* and bind and cleave them *in vitro* (Cortez *et al.,* 2010). Interestingly, an FtsK homologue could not be found in Archaea, suggesting that these organisms might apply a different mechanism to regulate chromosome dimer resolution.

1.10 Xer recombinases of *Helicobacter pylori*

Helicobacter pylori is a gram-negative, helically shaped bacterium belonging to the class of ε proteobacteria, which includes genera such as *Helicobacter* and *Campylobacter*. *H. pylori* is a gastric pathogen implicated in peptic ulcer disease and gastric cancer, the latter making it the only known organism classified as a primary cause of cancer (Eslick *et al.,* 1999; Marshall and Warren, 1984). It is estimated that up to two thirds of the human population are infected with *H. pylori*, with most infections being asymptomatic (Brown, 2000). Currently, symptomatic infections of *H. pylori* are efficiently cured with broad spectrum antibiotic cocktails; however, the lack of effective vaccines and the spread of antibiotic resistance genes both pose a threat, rendering *H. pylori* a highly relevant medical research target (Trust *et al*., 2001).

The genomes of many *H. pylori* strains have been sequenced, revealing a small, compact genome (1.67 Mbp, compared to *E. coli* 4.64 Mbp), rich in strain-specific gene islands that contribute to *H. pylori* pathogenicity and fitness (Alm *et al.,* 1999; Baltrus *et al.* 2009; Oh *et al.,* 2006; Tomb *et al*., 1997). The studies shed light on features important for colonization of the stomach, which due to its extremely low pH presents an unfavourable environment for bacterial growth.

Computational analysis of *H. pylori* genomes revealed that this bacterium carries two genes potentially encoding Xer recombinases. *H. pylori* XerH, a close homologue of *C. jejuni* XerH, is present on the chromosome of all *H. pylori* strains. The putative XerH recombination site, dif_H, was identified by comparative genome analysis (Carnoy and Roten, 2009). Deletion of XerH results in slower cell growth, increased DNA content, and

increased UV sensitivity, a phenotype similar to that observed for XerC/D deletions in *E. coli* and consistent with the role of XerH in chromosome dimer resolution (Debowski *et al*., 2012). Deletions of *H. pylori* $FtsK_C$ showed similar phenotype, suggesting that analogously to XerC/D and XerS recombination, FtsK might be involved in the process of chromosome dimer resolution by XerH. Additionally, as the *H. pylori* genome lacks homologues of topoisomerase IV (*parC* and *parE*), a potential role of XerH in chromosome decatenation was also discussed (Debowski *et al*., 2012).

Interestingly, many *H. pylori* strains carry a second Xer recombinase, XerT, usually within a large genetic island called transposable Plasticity Zone (TnPZ; Fischer *et al.,* 2010; Kersulyte *et al.,* 2009). The presence of TnPZs, 30-65 kbp genetic islands containing up to 45% strain-specific genes, slightly increased the fitness of *H. pylori* in mice (Kersulyte *et al*., 2009). XerT is not essential for *H. pylori* growth and its deletions do not cause any growth disadvantage (Debowski *et al*., 2012). In fact, in many *H. pylori* strains, only inactive remnants of XerT can be found (Kersulyte *et al*., 2009). Thus, XerT is not required for chromosome dimer resolution at dif_H and no other conserved *dif*-like site for XerT binding could be identified. Instead, XerT was proposed to act as a conjugative transposase: an enzyme required for mobilization and transfer of TnPZ between *H. pylori* strains (Kersulyte *et al*., 2009). If confirmed, this would classify TnPZ as a special case of an IMEX encoding its own Xer recombinase (Das *et al.,* 2013).

It seems that the two Xer recombinases of *H. pylori* have evolved separately to perform different functions in *H. pylori* life cycle: XerH is the canonical chromosome dimer recombinase required for maintaining stability of the genome, while XerT mobilizes conjugative elements that spread amongst *H. pylori* strains, conferring fitness advantage together with potential unknown functions. So far, there is no indication of any cross-talk between the two proteins in *H. pylori*, apart from a slight increase in XerH recombination in XerT deletion strains, the significance of which is not yet clear (Debowski *et al.,* 2012).

1.11 Aims and objectives of this study

This PhD project focuses on the study of the two Xer recombinases from *H. pylori*, XerH and XerT. On the one hand, this study investigates XerH, an Xer recombinase performing the canonical function of chromosome dimer resolution, in order to provide better

understanding of Xer recombination in general and reveal the mechanism and regulation of single-protein Xer systems. On the other hand, it studies a putative conjugative transposon integrase, XerT, in order to gain deeper understanding of conjugative transposition, a mechanism of DNA transfer that so far is poorly understood. In addition, studying these two systems could shed light on the mechanistic and regulatory aspects of Xer recombination in *H. pylori*, which in the future could help the development of genetic tools in *Helicobacter* or the design of antibacterial agents against this deadly pathogen.

The specific aims include:

- Gaining understanding of the mechanism of XerH recombination, in particular with respect to DNA recognition, binding, and cleavage by XerH at the recombination site dif_H, through extensive biochemical characterization.

- Structural characterization of XerH in complex with DNA by X-ray crystallography, taking advantage of the fact that this system employs a single protein and therefore features higher level of symmetry that may facilitate crystallization.

- Gaining insights into regulation of XerH recombination and chromosome dimer resolution, focusing on the proposed role of FtsK as a regulator of XerH function.

- Investigating the role of XerT in mobilization of the novel conjugative transposon TnPZ: confirming the role of XerT as a transposase in *in vitro* and *in vivo* assays and characterizing the binding and cleavage events performed by XerT at the transposon ends.

2. Materials and methods

2.1 Materials

2.1.1 Chemicals and reagents

Unless otherwise indicated, all chemicals used in this study were purchased from Sigma-Aldrich or Merck Millipore. All restriction enzymes were supplied by New England Biolabs (NEB) apart from DpnI, which was purchased from Thermo Scientific. Enzymes and reagents used for molecular biology were supplied by NEB, Bioline, and Thermo Scientific as stated in the text. Antibiotics were purchased from Sigma-Aldrich and Carl Roth. All buffers and stocks were made using deionized, distilled water.

2.1.2 Bacterial growth media

E. coli cultures were handled and propagated as described in Sambrook *et al*., 2001. All media for bacterial growth were prepared by EMBL Media Kitchen Facility. Liquid *E. coli* cultures were grown in Lysogeny Broth (LB) medium containing 10 g tryptone, 5 g yeast extract, and 5 g sodium chloride per litre (pH 7.2) or in Super Optimal Broth (SOB) medium containing 20 g tryptone, 5 g yeast extract, 0.585 g sodium chloride, and 0.186 g potassium chloride per litre (pH 7.2). Solid medium for *E. coli* growth consisted of 10 g tryptone, 5 g yeast extract, 10 g sodium chloride, and 15 g agar per litre (pH 7.2). When applicable, the following antibiotics were added to the growth media: kanamycin (kanamycin sulphate, 50 μg/ml), ampicillin (ampicillin sodium salt, 100 μg/ml), chloramphenicol (33 μg/ml), streptomycin (streptomycin sulphate salt, 50 μg/ml), erythromycin (200 μg/ml), and tetracycline (tetracycline hydrochloride, 10 μg/ml). For growth of *pir* strains, the media were supplemented with 0.3 mM of thymidine. Growth of bacteria was monitored by measuring the optical density at the wavelength of 600 nm (OD_{600}).

2.1.3 Bacterial strains

E. coli strains used in this study, their genotypes, and their antibiotic resistances are listed in Table 2-1. All *E. coli* strains were maintained as glycerol stocks (1:1 ratio of bacterial culture in LB and 80% glycerol) at -80 °C. Genomic DNA from *H. pylori* strains (Table 2-1) was kindly provided by Professor Douglas Berg (Washington University, St. Louis). Strains DS941 and π1 were kindly provided by Dr Sean Colloms (University of Glasgow). Strain DS9041 was kindly provided by Professor David Sherratt (University of Oxford). The remaining strains were supplied by EMBL Protein Expression and Purification Core (PEP-core) Facility.

Table 2-1: Bacterial strains used in this study.

Strain	Genotype	Resistance[a]	Source
E. coli			
XL10 Gold	Tet^R Δ*(mcrA)183* Δ*(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac Hte* [F′ *proAB lacIqZDM15 Tn10* (Tet^R) Amy Cm^R].	Cm/Tet	Stratagene
π1	F- *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK- mK+), λ–, thyA::(erm-pir116)*	Erm	Demarre *et al*., 2004
BL21 (DE3)	F- *ompT gal dcm lon hsdSB(rB- mB-)* λ(DE3)	-	Stratagene
BL21 (DE3) pLysS	F- *ompT gal dcm lon hsdSB(rB- mB-)* λ(DE3) pLysS (Cm^R)	Cm	Stratagene
Rosetta	F- *ompT gal dcm hsdSB(rB- mB-)* λ(DE3)	-	Stratagene
DS941	AB1157 [*thr-1, araC14, leuB6*(Am), *Δ(gpt-proA)62, lacY1, tsx-33, qsr'-0, glnV44*(AS), *galK2*(Oc), *LAM-, Rac-0, hisG4*(Oc), *rfbC1, mgl-51, rpoS396*(Am), *rpsL31*(Str^R), *kdgK51, xylA5, mtl-1, argE3*(Oc), *thi-1*] *recFJ43*, *lacZΔM15*, *lacIq*	Str	Summers and Sherratt, 1988
DS9041	DS941 *fisK*::*cat1*	Cm/Str	Recchia *et al*., 1999
DS941Δ*dif*	DS941 *dif*::Km^R	Km/Str	this study
DS941dif_H	DS941 *dif* ::(dif_H-Km^R)	Km/Str	this study
DS941dif_H-Km-dif_H	DS941 *dif* ::(dif_H-Km^R-dif_H)	Km/Str	this study
DS9041dif_H-Km-dif_H	DS9041 *dif* ::(dif_H-Km^R-dif_H)	Cm/Km/Str	this study
H. pylori			
26695	Wild-type strain	-	Akopyants *et al*., 1995
P12	Wild-type strain	-	Kersulyte *et al*., 2003

[a] – Cm – chloramphenicol; Tet – tetracycline; Erm – erythromycin; Str – streptomycin; Km – kanamycin.

2.1.4 Plasmids

Table 2-2 contains a list of *E. coli* plasmids used in this study together with the description, carried antibiotic resistance, origin of replication, and source of each plasmid. The plasmid pMS183Δ was kindly provided by Professor Marshall Stark. The plasmids pKD13 and pKD46 were kindly provided by Dr Athanasios Typas. All purified plasmids were stored

in Tris-EDTA (TE) buffer (10 mM tris(hydroxymethyl)aminomethane (Tris), pH 8, and 1 mM ethylenediaminetetraacetic acid (EDTA)) at -20 °C.

Table 2-2: Plasmids used in this study.

Plasmid	Description	Resistance[a]	Origin	Source
XerH_26695_pMA-T	Carrier vector for codon-optimized XerH from strain 26695	Amp	ColE1	MrGene (Life Technologies)
pETM-28	T7 expression vector with N-terminal 6xHis-SUMO tag	Km	pBR322	PEP-core EMBL Heidelberg
pETM-22	T7 expression vector with N-terminal TRX-6xHis tag	Km	pBR322	PEP-core EMBL Heidelberg
XerH_26695_pETM-28	T7 expression vector encoding 6xHis-SUMO-XerH	Km	pBR322	This study
XerT_P12_pETM-28	T7 expression vector encoding 6xHis-SUMO-XerT	Km	pBR322	This study
XerT_P12_pETM-22	T7 expression vector encoding TRX-6xHis-XerT	Km	pBR322	This study
pBAD/MCS	araBAD expression vector	Amp	pUC	PEP-core EMBL Heidelberg
XerH_pBAD/MCS	araBAD expression vector encoding XerH	Amp	pUC	This study
XerT_pBAD/MCS	araBAD expression vector encoding XerT	Amp	pUC	This study
$FtsK_C$-A_P12_pETM-22	T7 expression vector encoding TRX-6xHis-$FtsK_C$-A	Km	pBR322	This study
$FtsK_C$-B_P12_pETM-22	T7 expression vector encoding TRX-6xHis-$FtsK_C$-B	Km	pBR322	This study
XerH-FtsKγ_pETM-28	T7 expression vector encoding 6xHis-SUMO-XerH-linker-FtsKγ	Km	pBR322	This study
XerH_Y344F_pETM-28	T7 expression vector encoding 6xHis-SUMO-XerH Y344F	Km	pBR323	This study
XerT_Y332F_pETM-22	T7 expression vector encoding TRX-6xHis-XerT Y332F	Km	pBR324	This study
XerT_R300Q_pETM-22	T7 expression vector encoding TRX-6xHis-XerT R300Q	Km	pBR324	This study
pMS183Δ	Reporter vector encoding *galK* marker gene	Km	pSC101	Professor Marshall Stark
pAB102	Low copy number reporter plasmid with dif_H-GalK-dif_H cassette	Km	pSC101	This study
pAB106	Low copy number reporter plasmid with TnPZ_LE-GalK-TnPZ_RE cassette	Km	pSC101	This study
pACYC177	Medium-copy number cloning vector	Km/Amp	p15a	NEB
pAB122	Reporter plasmid with dif_H-GalK-dif_H cassette inserted into pACYC177	Km	p15a	This study
pKD13	Template plasmid for amplification of kanamycin resistance cassette	Km	oriR6Kγ	Datsenko and Wanner, 2000
pKD46	Expression vector with phage λ Red genes under araBAD promoter	Amp	oriR101	Datsenko and Wanner, 2000

[a] – Amp – ampicillin; Km – kanamycin.

2.1.5 Oligonucleotides

All oligonucleotides used in this study were purchased from Integrated DNA Technologies (IDT) and were resuspended in distilled water or TE buffer to a final concentration of 1 mM or 100 μM. Oligonucleotides used as primers in polymerase chain reaction (PCR) are listed in Table 2-3. Melting temperature of the primers was estimated using the IDT OligoAnalyzer tool (http://eu.idtdna.com/calc/analyzer). Oligonucleotides used as DNA substrates in the *in vitro* assays and for crystallization of XerH-dif_H complexes are shown in the results sections for individual experiments. Oligonucleotides were purchased with standard desalting and without modifications unless otherwise stated.

Table 2-3: Primers for PCR reactions used in this study.

No	Primer	Sequence	Purpose
1	XerH_26695_NcoI_for	GATCCCATGGTCATGAAACATCCGCTGGAAGAACTG	Cloning of *xerH* into pBAD/MCS
2	XerH_26695_XhoI_rev	GATCCTCGAGTTAGTTTTCTTCCCAAATGCTTGC	Cloning of *xerH* into pBAD/MCS
3	XerT_P12_NcoI_for	GATCCCATGGCTGACTGTAAAATGTCTCGTGTTTC	Cloning of *xerT* into pBAD/MCS
4	XerT_P12_XhoI_rev	GATCCTCGAGTCACTCGCTTTTCTCATTACGCAG	Cloning of *xerT* into pBAD/MCS
5	pETM22-ftsK$_C$A_for	GGAAGTTCTGTTCCAGGGGCCCAAAGATTATGAGCTTCCCACCA	Cloning of *ftsK$_C$-A* into pETM-22
6	pETM22-ftsK$_C$A_rev	GGTGGTGGTGGTGGTGCTCGAGCTAAAAGTTTTGCAAAATCTCTCTG	Cloning of *ftsK$_C$-A* into pETM-22
7	pETM22-ftsK$_C$B_for	GGAAGTTCTGTTCCAGGGGCCCTCTTTAGACGAAAACGAGATTGACC	Cloning of *ftsK$_C$-B* into pETM-22
8	pETM22-ftsK$_C$B_rev	GGTGGTGGTGGTGGTGCTCGAGCTAAAAGTTTTGCAAAATCTCTCTG	Cloning of *ftsK$_C$-B* into pETM-22
9	pETM28-ftsKγ_for	CCAGCAGCAGACCGGTGGATCCCAAGGCGATGACATTTTAGAAAGG	Cloning of *ftsKγ* into pETM-28
10	pETM28-ftsKγ_rev	GGTGGTGGTGGTGGTGCTCGAGCTAAAAGTTTTGCAAAATCTCTCTG	Cloning of *ftsKγ* into pETM-28
11	FtsK_linker_BamHI_for	GATCGGATCCGGCGGCGGCAGCGAAGGCGGCGGCAGCGAAGGCGGCA GCGGCCAAGGCGATGACATTTTAGAAAGG	Amplification of *ftsKγ*, introducing flexible linker upstream of the gene
12	FtsK_linker_XhoI_rev	GATCCTCGAGCTAAAAGTTTTGCAAAATCTCTCTGTTGC	Amplification of *ftsKγ*
13	XerH-FtsKγ_for	GAAGCAGCAAGCATTTGGGAAGAAAACTCCGGCGGCGGCAGCGAA	Cloning of linker-*ftsKγ* into XerH_26695_pETM28
14	XerH-FtsKγ_rev	GGGCTTTGTTAGCAGCCGGATCCTAAAAGTTTTGCAAAATCTCTCTG	Cloning of linker-*ftsKγ* into XerH_26695_pETM28
15	XerH_Y344F_for	CAAGCCTGAATACCAGCCGTATTTTTACCCATTTCGATAAA	Site-directed mutagenesis of XerH Tyr344
16	XerH_Y344F_rev	TTTATCGAAATGGGTAAAAATACGGCTGGTATTCAGGCTTG	Site-directed mutagenesis of XerH Tyr344
17	XerT_Y332F_for	CTGTCGACGAAAATCTTTATCCACACCACACAG	Site-directed mutagenesis of XerT Tyr332
18	XerT_Y332F_rev	CTGTGTGGTGTGGATAAAGATTTTCGTCGACAG	Site-directed mutagenesis of XerT Tyr332
19	XerT_R300Q_for	CTGCACCTGTTCCAGCACAGCTTTGCCAC	Site-directed mutagenesis of XerT Arg300
20	XerT_R300Q_rev	GTGGCAAAGCTGTGCTGGAACAGGTGCAG	Site-directed mutagenesis of XerT Arg300
21	*dif$_H$*_NheI_for	GATCGCTAGCGCAGATTGAAAAGGGGATTG	Cloning of the *dif$_H$* site into pMS183Δ, upstream of *galK*
22	*dif$_H$*_BsrGI_rev	GATCTGTACATCTTTCTTGCGTTCTAAAATTGA	Cloning of the *dif$_H$* site into pMS183Δ, upstream of *galK*
23	*dif$_H$*_XbaI/EcoRI_for	GATCTCTAGAATTCGCAGATTGAAAAGGGGATTG	Cloning of the *dif$_H$* site into pMS183Δ, downstream of *galK*
24	*dif$_H$*_KpnI_rev	GATCGGTACCTCTTTCTTGCGTTCTAAAATTGA	Cloning of the *dif$_H$* site into pMS183Δ, downstream of *galK*
25	TnPZ_L_NheI_for	GATCGCTAGCGATTCACTCTGGCTAGGGCTTG	Cloning of the TnPZ_LE into pMS183Δ, upstream of *galK*
26	TnPZ_L_BsrGI_rev	GATCTGTACATTTTTGGTATAACAACGGAGTAATCA	Cloning of the TnPZ_LE into pMS183Δ, upstream of *galK*
27	TnPZ_R_XbaI/EcoRI_for	GATCTCTAGAATTCCATTGTTCGCATGCATTTCT	Cloning of the TnPZ_RE into pMS183Δ, downstream of *galK*
28	TnPZ_R_KpnI_rev	GATCGGTACCATTTTAAAAAAGCAACTCCA	Cloning of the TnPZ_RE into pMS183Δ, downstream of *galK*
29	*dif$_H$*-GalK-*dif$_H$*_for	CAACTCGGTCGCCGCATACACTGCTAGCGCAGATTGAAAAGG	Cloning of the *dif$_H$*-GalK-*dif$_H$* cassette into pACYC177
30	*dif$_H$*-GalK-*dif$_H$*_rev	TTACCATCTGGCCCCAGTGCTGGGTACCCATTTAGTTATGAAAACTG	Cloning of the *dif$_H$*-GalK-*dif$_H$* cassette into pACYC177
31	EC*dif*_*dif$_H$*-F_for	ATTTAAAAGTTTGAAAAGTGCAGTTTTCATAACTAAATGAGTGTAGG CTGGAGCTGCTTC	Amplification of Km^R cassette from pKD13, introducing *dif$_H$* site upstream of the gene
32	EC*dif*_*dif$_H$*-F_rev	ATTCCGGGGATCCGTCGACCTGCAGTTC	Amplification of Km^R cassette from pKD13
33	EC*dif*_*dif$_H$*-B_for	GTGTAGGCTGGAGCTGCTTCGAAGTTCC	Amplification of Km^R cassette from pKD13
34	EC*dif*_*dif$_H$*-B_rev	ATTTAAAAGTTTGAAAAGTGCAGTTTTCATAACTAAATGAATTCCGG GGATCCGTCGACC	Amplification of Km^R cassette from pKD13, introducing *dif$_H$* site downstream of the gene
35	EC*dif*_*dif$_H$*-F_HR_for	CCTGAACGAATAACGACTACCAGAAAAGCACTTCGCATCACTCAGAC ATGATTTAAAAGTTTGAAAAGTGCAGTTTTCATAAC	Amplification of *dif$_H$*-Km^R cassette, introducing *E. coli* homologous region
36	EC*dif*_*dif$_H$*-F_HR_rev	GCATGACAATCATGACCGCCAACGACTGGATTCTTCCTTGGTTTATA TTATTCCGGGGATCCGTCGACC	Amplification of *dif$_H$*-Km^R cassette, introducing *E. coli* homologous region
37	EC*dif*_*dif$_H$*-B_HR_for	CCTGAACGAATAACGACTACCAGAAAAGCACTTCGCATCACTCAGAC ATGGTGTAGGCTGGAGCTGCTTC	Amplification of Km^R-*dif$_H$* cassette, introducing *E. coli* homologous region
38	EC*dif*_*dif$_H$*-B_HR_rev	GCATGACAATCATGACCGCCAACGACTGGATTCTTCCTTGGTTTATA TTATTTAAAAGTTTGAAAAGTGCAGTTTTCATAAC	Amplification of Km^R-*dif$_H$* cassette, introducing *E. coli* homologous region
39	EC*dif*_seq_for	TAAGTAACAGCACCCTCTGGTG	Amplification and sequencing of *E. coli* ter region
40	EC*dif*_seq_rev	CCGGAAGTGCAGATGCATG	Amplification and sequencing of *E. coli* ter region

2.2 Molecular biology methods

2.2.1 Constructs for protein overexpression

The open reading frame of the *xerH* gene (NCBI accession number HP0675) has been previously identified through *H. pylori* genome sequencing (Alm *et al*., 1999). The full-length gene from *H. pylori* strain 26695 was codon-optimized for expression in *E. coli* and synthesized by Life Technologies in XerH_26695_pMA-T. The gene was then cloned into

expression vector pETM-28 digested with BamHI and XhoI using restriction sites introduced during codon optimization, to give vector XerH_26695_pETM-28 (Figure 2-1 and Table 2-2).

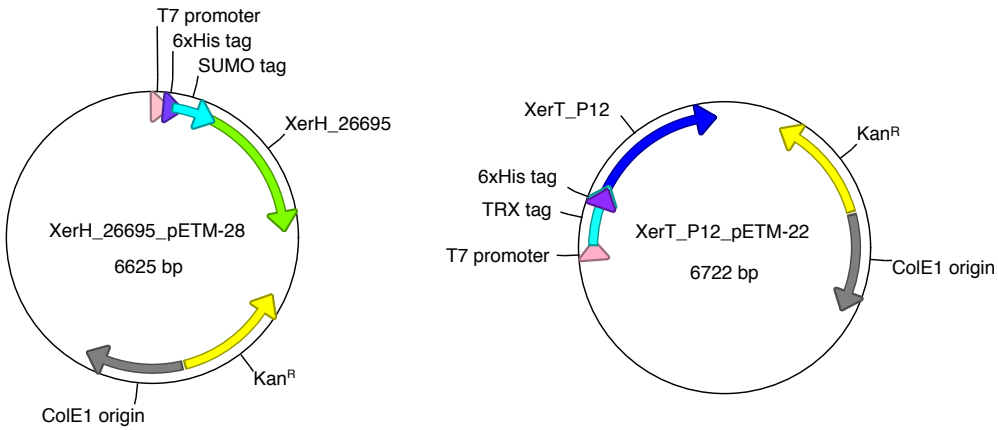


Figure 2-1: Schematic representation of the vectors expressing tagged XerH and XerT under T7 promoter. SUMO – Small Ubiquitin-like Modifier; TRX – thioredoxin. The figure was made in ApE – A plasmid Editor (available at http://biologylabs.utah.edu/jorgensen/wayned/ape/).

The full-length *xerT* gene (NCBI accession number HPP12_0437) from *H. pylori* strain P12 had been codon-optimized for expression in *E. coli* and cloned into expression vectors pETM-22 and pETM-28 to give vectors XerT_P12_pETM-22 (Figure 2-1 and Table 2-2) and XerT_P12_pETM-28 (Table 2-2) prior to this study (Bianca Beusink, personal communication).

PCR-amplified tag-free *xerH* and *xerT* (primers 1/2 and 3/4, Table 2-3) were also cloned into expression vector pBAD/MCS digested with NcoI and XhoI restriction enzymes, to obtain vectors XerH_pBAD/MCS and XerT_pBAD/MCS (Table 2-2).

Two constructs of the C-terminal domain of *H. pylori* P12 FtsK were designed based on secondary structure predictions from PSIPRED (Buchan *et al*., 2013): $FtsK_C$-A including residues 399-863 and $FtsK_C$-B including residues 418-863. The rationale of the design was based on the comparison of *E. coli* FtsK secondary structure elements and *H. pylori* predicted secondary structure, which indicated that *E. coli* $FtsK_C$ (sufficient for supporting XerC/D recombination *in vivo*; Barre *et al*., 2000) corresponds roughly to *H. pylori* FtsK residues 401-863. Considering the secondary structure predictions, residues Lys399 and Ser418 were chosen as N-termini of $FtsK_C$-A and $FtsK_C$-B, respectively, as these

hydrophilic residues were likely to promote solubility of the constructs. Both constructs were cloned into expression vector pETM-22 using restriction-free approach with primers 5/6 and 7/8 (Table 2-3) to give vectors $FtsK_C$-A_P12_pETM-28 and $FtsK_C$-B_P12_pETM-28 (Table 2-2).

The fusion construct of XerH joined to *H. pylori* P12 FtsKγ (See Figure 5-4a) was designed based on the XerD-FtsKγ fusion described in Grainge *et al*., 2011. *H. pylori* FtsKγ construct was designed based on homology of this region to *E. coli* FtsKγ, and the selected construct contained residues 801-863 of *H. pylori* FtsK. The cloning was conducted in three steps. First, the FtsKγ construct was amplified from *H. pylori* P12 genomic DNA and inserted into pETM-28 by restriction-free cloning using primers 9/10 (Table 2-3). Next, a flexible linker (amino acid sequence GGGSEGGGSEGGSG) was introduced into the 5'-end of primer 11, which was used together with primer 12 to amplify the *ftsKγ* gene, introducing the linker at the N-terminus of FtsKγ. The resulting linker-*ftsKγ* PCR product was inserted into pETM-28 using BamHI and XhoI restriction sites. Finally, linker-*ftsKγ* was inserted downstream of the full-length *xerH* gene in XerH_26695_pETM-28 using restriction-free cloning with primers 13/14, to give a continuous *xerH*-linker-*ftsK* open reading frame in the vector XerH-FtsKγ_pETM-28 (Table 2-2).

2.2.2 Restriction enzyme cloning

The general outline of restriction enzyme cloning is shown in Figure 2-2. In brief, an insert fragment (obtained by PCR) and a plasmid were digested with two restriction enzymes, ligated, and transformed into *E. coli*. This approach was applied to cloning of genes into expression vectors when sites for restriction enzymes were available as well as cloning and shuffling various gene cassettes in reporter plasmids. All DNA concentrations were determined by ultraviolet (UV) spectroscopy at 260 nm using a NanoDrop instrument (Thermo Scientific).

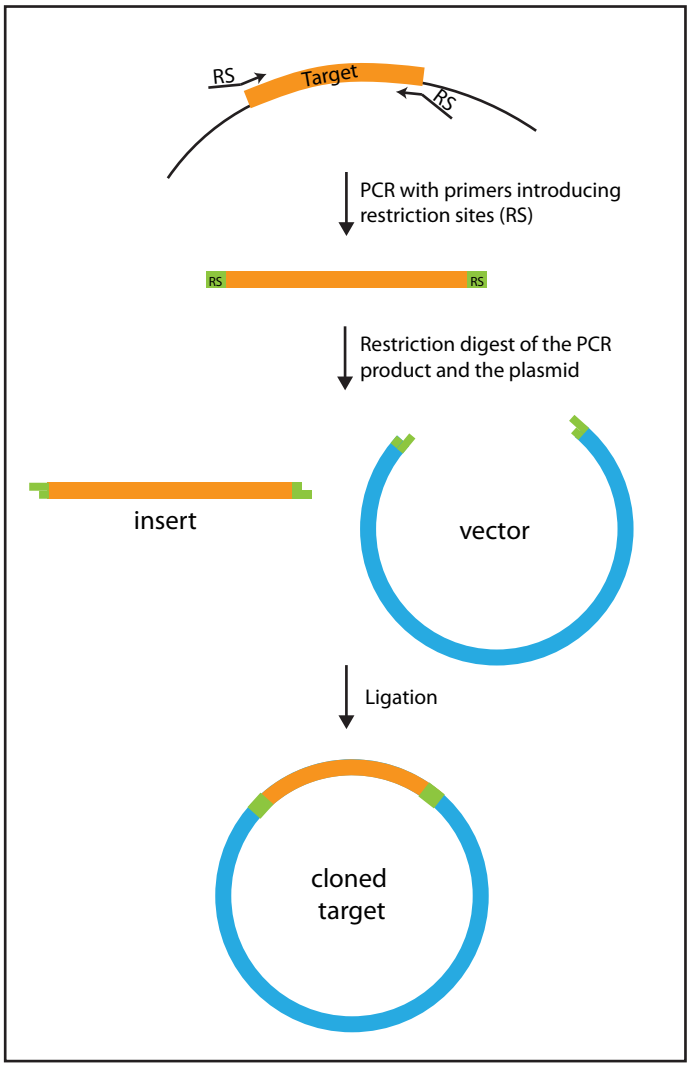


Figure 2-2: General outline of restriction enzyme cloning approach. The target DNA sequence is amplified from genomic or plasmid DNA by PCR using primers introducing 5'-end extensions containing appropriate restriction enzyme sites (RS). The PCR product and a plasmid containing matching RS are subjected to the restriction enzyme digest that results in compatible "sticky ends" on both the insert and the vector. The ends are ligated together, resulting in the final plasmid containing the target DNA. This plasmid can be transformed and propagated in *E. coli*.

2.2.3 Polymerase Chain Reaction (PCR)

PCR amplifications were carried out using primers listed in Table 2-3, with primer stock concentration of 10 µM, resuspended in distilled water and stored at -20 °C when not in use. The PCR mix consisted of 1x Phusion High-Fidelity Buffer, 2 units of Phusion® High-Fidelity DNA Polymerase (both supplied by NEB), 200 µM dNTPs (Bioline), 200 nM of each forward and reverse primer, and 20-200 ng of template DNA (genomic or plasmid) in the total reaction volume of 50 µl. The control reaction contained no template. The thermocycling conditions used are shown in Table 2-4. The reaction products were purified either using QIAquick PCR Purification Kit (QIAGEN) according to manufacturer's instructions, or through agarose gel electrophoresis and gel extraction.

Table 2-4: Thermocycling conditions for PCR with Phusion polymerase.

Step	Temperature	Time	
1. Initial Denaturation	98 °C	60 seconds	
2. Denaturation	98 °C	10 seconds	Steps 2-4 repeated 35 times
3. Annealing	50-60 °C*	15 seconds	
4. Extension	72 °C	30 seconds/kbp	
5. Final extension	72 °C	5 minutes	
6. Hold	4 °C	hold	

* Annealing temperature: lower melting temperature of the primer pair based on estimation by IDT OligoAnalyzer tool.

2.2.4 Restriction enzyme digestion of DNA

Single-strand overhangs were introduced into DNA plasmids and PCR products through restriction enzyme digestion with one or two restriction enzymes. The digestion reaction mix with NEB restriction enzymes (BamHI, BsrGI, EcoRI, KpnI, NcoI, NheI, or XhoI) consisted of 1x NEB buffer (chosen based on the manufacturer's buffer compatibility table) and 20 units of each restriction enzyme together with DNA (plasmid or PCR product) added to the final volume of 20 µl. The reactions were incubated for 1 hour at 37 °C. The reaction products were purified through agarose gel electrophoresis.

2.2.5 Agarose gel electrophoresis and gel extraction

Agarose gels were prepared with 1.2% (w/v) agarose powder dissolved in 1x Tris-acetate-EDTA (TAE) buffer (40 mM Tris, 20 mM acetic acid, and 1 mM EDTA) by heating in a microwave. The gel solution was cooled down to approximately 60 °C followed by addition of ethidium bromide to the final concentration of 1 µg/ml. Next, the gels were poured and allowed to solidify for 30-60 minutes and then transferred to the running chamber containing 1x TAE buffer. The DNA samples to be analysed on the gel were prepared by addition of 5x GelPilot Loading Dye (QIAGEN). The samples were loaded into the wells of the gel alongside a DNA marker (HyperLadder™ 1kb, Bioline). The voltage of 7-10 V per cm was applied for 30-60 minutes. The products of gel electrophoresis were analysed on a UV transilluminator (Alpha Innotech) using AlphaImager® HP software (Fisher Scientific).

When applicable, the bands of interest were excised from the gel with a scalpel and purified using a QIAquick Gel Extraction Kit (QIAGEN) according to the manufacturer's instructions. Purified DNA was stored at -20 °C.

2.2.6 Ligation

Two DNA fragments purified after restriction enzyme digestion containing complementary single-stranded overhangs were ligated to form functional plasmids. The total of 10 µl ligation mix contained 5 µl of 2x Quick Ligation Reaction Buffer, 1 µl of Quick T4 DNA Ligase (both from NEB), and 4 µl of insert and vector DNA fragments at approximately 3:1 molar ratio as determined based on DNA band intensities in agarose gel analysis. The ligation reactions were incubated at room temperature for 1 hour. Ligated plasmid was transformed into *E. coli* XL10-Gold competent cells.

2.2.7 Competent *E. coli* cells

2.2.7.1 Electro-competent cells

Electro-competent *E. coli* strains were prepared as follows: 400 ml of LB with appropriate antibiotics was inoculated with 2 ml of overnight culture of bacteria and grown at 37 °C to OD_{600} of 0.5-0.7. The cultures were chilled on ice for 20 minutes followed by harvesting by centrifugation at 4000 *g* for 15 minutes at 4 °C. The pellet was then washed twice with 400 ml of sterile ice-cold 10% glycerol, followed by one more wash with 20 ml of sterile ice-cold 10% glycerol. After final centrifugation, the pellet was resuspended in 1-2 ml of ice-cold 10% glycerol. 50 µl aliquots were flash-frozen in liquid nitrogen and stored at -80 °C until further use.

2.2.7.2 Chemically competent cells

Chemically competent *E. coli* strains were prepared as follows: 250 ml of SOB medium with 10 mM $MgCl_2$, appropriate antibiotics, and thymine when applicable was inoculated with 5 ml of overnight culture of bacteria and grown at 30 °C to OD_{600} of 0.5-0.6. The cultures were chilled on ice for 10 minutes followed by harvesting by centrifugation at

1100 *g* for 20 minutes at 4 °C. The pellet was then washed once with 80 ml of sterile ice-cold transformation buffer (TB: 10 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 6.7, 15 mM $CaCl_2$, 55 mM $MnCl_2$, 250 mM KCl, pH adjusted to 6.7 before addition of $MnCl_2$) then centrifuged as before. The resulting pellet was resuspended in 20 ml of ice-cold TB buffer. Next, 1.5 ml of dimethyl sulphoxide (DMSO) was added drop-wise to the cells, followed by 15 minutes incubation on ice. 100 µl aliquots were flash-frozen in liquid nitrogen and stored at -80 °C until further use.

2.2.8 Transformation of competent cells

2.2.8.1 Electroporation

Electro-competent cells in 50 µl aliquots were thawed on ice and no more than 2 µl of plasmid DNA was added to the cells. The cells were then transferred into a 0.1 cm electroporation cuvette (Bio-Rad) and electroporated using MicroPulser™ Electroporator (Bio-Rad) with Ec1 setting. 950 µl of LB was added to the electroporated cells immediately and the cells were transferred into an eppendorf tube. Cells were incubated for 1 hour at 37 °C with shaking, and then plated on LB-agar plates containing appropriate antibiotics. Plated cultures were grown for 16 hours at 37 °C.

2.2.8.2 Chemical transformation

Chemically-competent cells in 100 µl aliquots were thawed on ice and no more than 5 µl of plasmid DNA was added to the cells. The cells were then kept on ice for 30 minutes. Heat-shock transformation was performed by placing the cells at 42 °C for 45 seconds followed by placing them back on ice for 2 minutes. 900 µl of LB (containing 0.3 mM thymidine when required) was added to the transformed cells and the cells were incubated for 1 hour at 37 °C with shaking, then plated on LB-agar plates containing appropriate antibiotics and thymidine when required. Plated cultures were grown for 16 hours at 37 °C.

2.2.9 Restriction-free cloning

The general scheme of restriction-free cloning is outlined in Figure 2-3. This approach was employed when restriction sites were not readily available or the insert size was not optimal for PCR amplification in cloning of genes into expression vectors. The primers were designed using the web service at www.rf-cloning.org. All DNA concentrations were determined by UV spectroscopy at 260 nm using a NanoDrop instrument (Thermo Scientific). The first PCR mix contained 1x Phusion Flash High-Fidelity PCR Master Mix (Thermo Scientific), 500 nM of each forward and reverse primer, and 50-100 ng of template DNA in a total volume of 50 µl. The thermocycling conditions are shown in table 2-5. The reaction product, so-called Mega Primer, was analysed on an agarose gel and gel-extracted as described in section 2.2.5. The second PCR mix contained 1x Phusion Flash High-Fidelity PCR Master Mix, 350 ng of Mega Primer, and 50 ng of template DNA in a total volume of 50 µl. Thermocycling conditions of second PCR are shown in Table 2-5. The reaction products were purified using a QIAquick PCR Purification Kit (QIAGEN) according to the manufacturer's instructions and eluted in 17 µl of distilled water. In order to degrade all remaining template plasmid DNA, the sample was incubated with 2 µl of 10x FastDigest Buffer and 2 µl of FastDigest DpnI restriction enzyme (both from Thermo Scientific). The digest was carried out at 37 °C for at least 4 hours. 5 µl of the sample was then transformed into *E. coli* XL10-Gold competent cells.

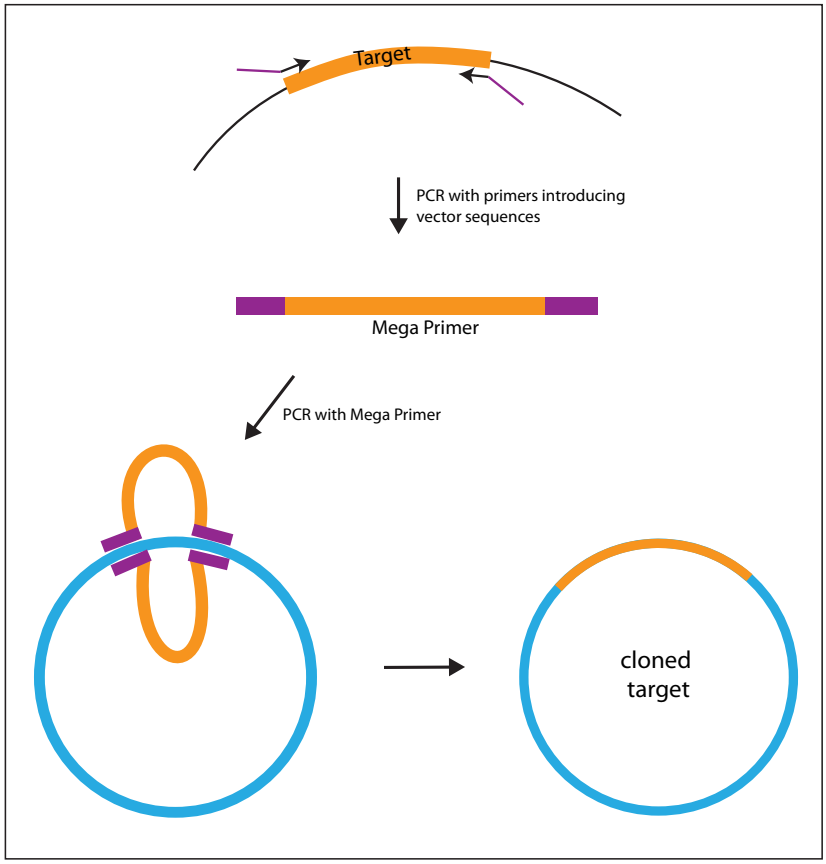


Figure 2-3: General outline of restriction-free cloning. The target DNA sequence is amplified from genomic or plasmid DNA by PCR using primers introducing 5'-end extensions containing sequences complementary to the vector sequences at desired positions. The PCR product called Mega Primer serves as a bidirectional primer for amplification of the vector, introducing the target sequence into the vector. The resulting plasmid, after removal of template vector by DpnI restriction digest, can be transformed and propagated in *E. coli*.

Table 2-5: Thermocycling conditions in restriction-free cloning PCR

Step	Temperature	Time	
PCR1			
1. Initial Denaturation	98 °C	60 seconds	
2. Denaturation	98 °C	10 seconds	Steps 2-4 repeated 35 times
3. Annealing	50-60 °C*	10 seconds	
4. Extension	72 °C	60 seconds/kbp	
5. Final extension	72 °C	10 minutes	
6. Hold	4 °C	hold	
PCR2			
1. Initial Denaturation	98 °C	60 seconds	
2. Denaturation	98 °C	10 seconds	Steps 2-4 repeated 18 times
3. Annealing	60-65 °C*	10 seconds	
4. Extension	72 °C	60 seconds/kbp	
5. Final extension	72 °C	10 minutes	
6. Hold	4 °C	hold	

* Annealing temperature: lower melting temperature of the primer pair based on estimation by IDT OligoAnalyzer tool.

2.2.10 Colony PCR

Single colonies of *E. coli* transformants were analysed for the presence of the desired inserts by colony PCR. The reaction mix consisted of 1x MangoMix™ (Bioline) and 400 nM of each forward and reverse primers (from original cloning PCR). As a template, *E. coli* cells were added by touching a single transformant colony with a sterile tip, then dipping the tip multiple times in the reaction mix. Thermocycling conditions of the colony PCR are shown in Table 2-6. MangoMix™ already contains a loading dye, so the PCR products were directly loaded on agarose gels and analysed as described in section 2.2.5.

Table 2-6: Thermocycling conditions in colony PCR with MangoMix.

Step	Temperature	Time	
1. Initial Denaturation	96 °C	2 minutes	
2. Denaturation	96 °C	15 seconds	Steps 2-4 repeated 30 times
3. Annealing	50-60 °C*	30 seconds	
4. Extension	72 °C	1 minute/kbp	
5. Final extension	72 °C	5 minutes	
6. Hold	4 °C	hold	

* Annealing temperature: lower melting temperature of the primer pair based on estimation by IDT OligoAnalyzer tool.

2.2.11 Plasmid DNA extraction

Transformant colonies were inoculated into 5 ml LB medium containing appropriate antibiotics and thymine when required and grown overnight at 37 °C. Cells were harvested by centrifugation for 5 minutes at 3000 *g* at room temperature and the plasmid DNA was extracted using QIAprep Spin Miniprep Kit (QIAGEN). The samples were eluted in elution buffer provided or in distilled water and stored at -20 °C until further use.

2.2.12 DNA sequencing

All plasmids designed in this study were validated by DNA sequencing performed by GATC Biotech. The sequences were inspected using ApE – A plasmid Editor.

2.2.13 Site-directed mutagenesis

Catalytic tyrosine mutants of XerH and XerT were created through site-directed mutagenesis. Phenylalanine residues were introduced in place of XerH Tyr344 and XerT Tyr332, and a glutamine residue was introduced in place of XerT Arg300 by PCR. The primers used are listed in Table 2-3 (positions 15-20), and the resulting plasmids carrying mutated genes are shown in Table 2-2. The mutations were introduced by PCR where primers contained the desired change flanked by 10-25 nucleotides on each side of the mismatch. The primers were designed so that each half-site flanking the mismatch had a melting temperature of 58 °C with the melting temperature of the full primer close to 68 °C. The primers used were 5' phosphorylated by the manufacturer. The PCR mix consisted of 1x Phusion High-Fidelity Buffer, 2 units of Phusion® High-Fidelity DNA Polymerase (both supplied by NEB), 1x Taq DNA Ligase Reaction Buffer, 40 units Taq DNA ligase (NEB), 200 μM dNTPs (Bioline), 250 nM primer, and 50-200 ng of template plasmid DNA in the total reaction volume of 50 μl. Thermocycling conditions are shown in Table 2-7. A single primer containing the desired mutation was used in each PCR (with two primers used in parallel for each mutagenesis target as shown in Table 2-3), resulting in amplification of a single-stranded plasmid, ligated by Taq DNA ligase in the same reaction. The reaction products were purified using QIAquick PCR Purification Kit (QIAGEN) according to the manufacturer's instructions and eluted in 17 μl of distilled water. In order to degrade all remaining template plasmid DNA, the sample was incubated with 2 μl of 10x FastDigest Buffer and 2 μl of FastDigest DpnI restriction enzyme (both from Thermo Scientific). The digest was carried out at 37 °C for at least 4 hours. 5μl of the sample was then transformed into *E. coli* XL10-Gold competent cells, where the single-stranded plasmid was replicated by the host replication machinery, and plated on LB containing appropriate antibiotics. The outcome of mutagenesis was validated by sequencing of the resulting double-stranded plasmids.

Step	Temperature	Time	
1. Initial Denaturation	95 °C	60 seconds	
2. Denaturation	95 °C	30 seconds	Steps 2-4 repeated 30 times
3. Annealing	55 °C	30 seconds	
4. Extension	65 °C	1 minute/kbp	
5. Final extension	72 °C	5 minutes	
6. Hold	4 °C	hold	

Table 2-7: Thermocycling conditions in site-directed mutagenesis PCR.

2.3 Protein overexpression and purification methods

2.3.1 Computational analysis of proteins

Analysis of protein properties such as molecular weight, theoretical isoelectric point (pI) value, and extinction coefficients was performed using the ProtParam web-based tool, on the ExPASy Bioinformatics Resource Portal (www.expasy.org). To predict secondary structure of the proteins, the PSIPRED Protein Sequence Analysis Workbench (bioinf.cs.ucl.ac.uk/psipred) service was used. Structural models were generated using Phyre2 (Kelley and Sternberg, 2009). Protein and DNA sequence alignments were made using the ClustalW2 service (www.ebi.ac.uk/Tools/msa/clustalw2) and were visualized using Jalview software (Waterhouse *et al*., 2009). The percentage of sequence identity between proteins corresponds to the values computed by ClustalW2. The phylogenetic trees were prepared in MEGA6 (Tamura *et al.,* 2013) using the Neighbour-Joining method (Saitou and Nei, 1987). The protein domain structures were visualized using DOG 1.0 software (Ren *et al*., 2009).

2.3.2 Protein overexpression in *E. coli*

XerH, XerT, $FtsK_C$, and XerH-FtsKγ proteins were overexpressed in *E. coli* strains BL21 (DE3), BL21 (DE3) pLysS, or Rosetta from pET vectors under a T7 promoter. The expression vectors were transformed into chosen *E. coli* strains, and overnight cultures of the transformants were set up in LB medium with appropriate antibiotics. In small-scale test expressions, 100 ml of LB containing appropriate antibiotics was inoculated with 200 µl of the overnight culture and grown at 37 or 42 °C until OD_{600} of 0.8-1.2. Next, the cultures were either transferred into an incubator at 15 °C and cooled down for 1 hour before inducing expression by addition of 0.5 mM of isopropyl β-D-1-thiogalactopyranoside (IPTG), or the cultures were kept at 37 °C and IPTG was added immediately. The proteins were expressed for 3 hours at 37 °C or for 15 hours at 15 °C. The samples collected before induction and after the expression was completed were analysed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) as described in section 2.3.4. The best expression conditions were chosen and the procedure was scaled up. The specific conditions used for large-scale expression of individual proteins are given in Table 2-8. In large-scale expressions, 500 ml of LB medium

containing appropriate antibiotics was inoculated with 1 ml of overnight cultures and grown at 37 or 42 °C until OD_{600} of 0.8-1.2. Next, the cultures were transferred into an incubator at 15 °C and cooled down for 1 hour before inducing expression by addition of 0.5-1 mM of IPTG. The proteins were expressed for 15 hours at 15 °C and then harvested in 1 litre batches by centrifugation at 4000 *g* for 20 minutes at 4 °C. The obtained pellet was then washed once with 50 ml LB and centrifuged at 3000 *g* for 45 minutes at 4 °C. After the final harvest, the pellet was flash-frozen in liquid nitrogen and stored at -80 °C until further use. To assess expression efficiency, samples immediately before addition of IPTG and immediately before culture harvest were collected and kept for analysis by SDS-PAGE as described in section 2.3.4.

Table 2-8: Large-scale expression conditions for proteins overexpressed in this study.

Protein	Vector/tag	*E. coli* strain	Temperature before induction	OD_{600} at induction	IPTG	Temperature of expression
XerH	pETM-28/N-terminal 6xHis-SUMO	Bl21 (DE3)	37°C	1-1.2	0.5 mM	15°C
XerH Y344F	pETM-28/N-terminal 6xHis-SUMO	Bl21 (DE3)	37°C	1-1.2	0.5 mM	15°C
XerT	pETM-22/N-terminal TRX-6xHis	Bl21 (DE3)	42°C	0.8-1	0.5 mM	15°C
XerT Y332F	pETM-22/N-terminal TRX-6xHis	Bl21 (DE3)	42°C	0.8-1	0.5 mM	15°C
XerT R300Q	pETM-22/N-terminal TRX-6xHis	Bl21 (DE3)	42°C	0.8-1	0.5 mM	15°C
$FtsK_C$-A	pETM-22/N-terminal TRX-6xHis	Bl21 (DE3)	37°C	1-1.2	1 mM	15°C
$FtsK_C$-B	pETM-22/N-terminal TRX-6xHis	Bl21 (DE3)	37°C	1-1.2	1 mM	15°C
XerH-FtsKγ	pETM-28/N-terminal 6xHis-SUMO	Bl21 (DE3)	37°C	1	0.5 mM	15°C

The seleno-methionine (Se-Met) derivative of XerH was expressed in methionine-prototrophic strain BL21 (DE3) from plasmid XerH_26695_pETM-28. 5 ml of the overnight culture was washed twice with M9 medium containing 1x M9 salts (Table 2-9), 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.4% (w/v) glucose, 1 µg/ml thiamine, and 1x trace elements solution (Table 2-10). 50 ml of M9 medium was then inoculated with the washed culture (5 ml) and grown overnight at 37 °C. The overnight culture was then added to 1 litre of fresh M9 medium and grown to OD_{600} of 1.2. The cultures were cooled down to 15 °C for 1 hour. Next, essential amino acids were added at the following concentrations: 100 mg/l of lysine, phenylalanine, and threonine and 50 mg/l of isoleucine, leucine, valine, and seleno-methionine. The cultures were then incubated for 20 minutes at 4 °C without shaking, followed by induction of expression with 0.5 mM IPTG. After expression for 15 hours at 15 °C, the cultures were harvested and frozen as described for the other proteins.

Table 2-9: Composition of 10x M9 salts.

Components	Amount per litre
$Na_2HPO_4 \cdot 7H_2O$	70 g
KH_2PO_4	30 g
NaCl	5 g
NH_4Cl	10 g

Table 2-10: Composition of 100x trace elements.

Components	Amount per litre
EDTA	5 g
$FeSO_4 \cdot 7H_2O$	0.85 g
$ZnSO_4 \cdot 7H_2O$	0.173 g
$CuCl_2 \cdot 2H_2O$	0.013 g
$CoCl_2 \cdot 6H_2O$	0.010 g
H_3BO_3	0.010 g
$MnCl_2 \cdot 6H_2O$	0.0016 g
$(NH_4)_6Mo_7O_{24} \cdot 4H_2O$	0.010 g

2.3.3 Protein purification

All proteins described in this study were over-expressed as fusions with an N-terminal 6xHis affinity tag and purified through a 3-step affinity purification scheme, including 1) first purification by nickel affinity chromatography, 2) tag cleavage and removal by second nickel affinity chromatography, and 3) size exclusion chromatography on a gel filtration column. All purification steps were performed using the ÄKTApurifier Protein Purification System (GE Healthcare) and associated equipment, such as sample loops, according to the manufacturer's instructions. The chromatogram was collected at 280 nm (mainly protein absorption) and 260 nm (mainly DNA absorption). The system was operated and the chromatograms collected and analysed using UNICORN Control Software (GE Healthcare).

2.3.3.1 Sample preparation

The harvested expression pellets were thawed on ice and resuspended in 40 ml Lysis buffer (Table 2-11). The resuspended cells were then lysed by sonication using a Branson Sonifier 250 set to 50% duty cycle and 50% output control, in 7 cycles of 30 seconds sonication and 90 seconds rest on ice. A sample of the lysate was collected for analysis by SDS-PAGE as described in section 2.3.4. Next, the lysed cells were ultracentrifuged at 40000 *g* for 45 minutes at 4 °C. The supernatant, corresponding to the soluble fraction of *E. coli* proteins, was transferred into a fresh falcon tube and a sample was collected for analysis by SDS-PAGE. The pellet, corresponding to the insoluble fraction of *E. coli* proteins, was resuspended in 20 ml of 1x phosphate buffered saline (PBS) and a sample was collected for analysis by SDS-PAGE. After SDS-PAGE analysis of the samples confirmed that the protein of interest is present in satisfactory amount in the soluble

fraction, the supernatant was filtered through 0.45 µM disposable filter and loaded onto the column for the first nickel affinity purification.

Table 2-11: Composition of purification and dialysis buffers used in this study.

Buffer	Composition[a]	pH
XerH, XerH Y344F, XerT, XerT Y332F, XerT R300Q, XerH-FtsKγ purification		7,5
Lysis buffer	1x PBS, 1 M NaCl, 5% glycerol, 15 mM imidazole, 0.2 mM TCEP, 1 tablet of cOmplete Protease Inhibitor Cocktail (Roche), 1.5 mM PMSF	7,5
Binding buffer (Buffer A)	1x PBS, 1 M NaCl, 5% glycerol, 15 mM imidazole, 0.2 mM TCEP	7,5
Elution buffer (Buffer B)	1x PBS, 1 M NaCl, 5% glycerol, 1 M imidazole, 0.2 mM TCEP	7,5
Gel filtration buffer (Buffer C)	1x PBS, 1 M NaCl, 5% glycerol, 0.2 mM TCEP	7,5
$FtsK_C$-A, $FtsK_C$-B purification		7,5
Lysis buffer	1x PBS, 500 mM NaCl, 5% glycerol, 15 mM imidazole, 0.2 mM TCEP, 1 tablet of cOmplete Protease Inhibitor Cocktail (Roche), 1.5 mM PMSF	7,5
Binding buffer (Buffer A)	1x PBS, 500 mM NaCl, 5% glycerol, 15 mM imidazole, 0.2 mM TCEP	7,5
Elution buffer (BufferB)	1x PBS, 500 mM NaCl, 5% glycerol, 1 M imidazole, 0.2 mM TCEP	7,5
Gel filtration buffer (Buffer C)	1x PBS, 500 mM NaCl, 5% glycerol, 0.2 mM TCEP	7,5
Dialysis buffers		7,5
HEPES dialysis buffer	50 mM HEPES buffer (pH 7.5), 250 mM NaCl, 5% glycerol, 1 mM DTT	7,5
Acetate dialysis buffer 1	25 mM sodium acetate buffer (pH 5.6), 750 mM NaCl, 5% glycerol, 1 mM DTT	5,5
Acetate dialysis buffer 2	25 mM sodium acetate buffer (pH5.6), 500 mM NaCl, 5% glycerol, 1 mM DTT	5,5
Acetate dialysis buffer 3	25 mM sodium acetate buffer (pH 5.6), 250 mM NaCl, 5% glycerol, 1 mM DTT	5,5

[a] – 1xPBS – 1x phosphate buffered saline (0.2 g KCl, 0.2 g KH_2PO_4, 1.15 g Na_2HPO_4, and 8 g NaCl per litre, prepared by EMBL Media Kitchen Facility); TCEP - tris(2-carboxyethyl)phosphine; PMSF – phenylmethanesulphonylfluoride; HEPES - 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; DTT – dithiothreitol. 1 M HEPES buffer: 238.3 g HEPES per litre, pH adjusted with NaOH. 200 mM sodium acetate buffer: 176 ml 0.2 M acetic acid and 824 ml 0.2 M sodium acetate per litre.

2.3.3.2 First HisTrap purification

For the first nickel affinity chromatography, a 5 ml HisTrap HP column (GE Healthcare) was prepared by prewashing with 5 column volumes (CV) of distilled water followed by 5 CV of buffer A (Binding buffer; Table 2-11), 5 CV of buffer B (Elution buffer; Table 2-11), then again 5 CV of buffer A. The sample was then injected using a superloop into the HisTrap column at a flow rate of 1 ml/min. The column was then washed with buffer A until the absorbance reading reached the baseline. All subsequent steps were carried out at a flow rate of 2 ml/min. In order to remove impurities with weak nickel binding affinity, the column was washed with 95% of buffer A mixed with 5% of buffer B (65 mM imidazole) until the reading reached the baseline again. Next, to elute the 6xHis-tagged protein, a gradient from 5% to 50% buffer B (65-515 mM imidazole) was applied over 10 CV. Finally, the column was washed with 100% buffer B in order to remove all remaining binders. Samples were collected and fractionated throughout the binding, washing, and elution steps, and were subsequently analysed on SDS-PAGE to identify the fractions

containing the protein of interest. The fractions of interest were then combined and the approximate concentration of the protein was measured by UV spectroscopy at 280 nm using a NanoDrop instrument, using the molecular weight and the extinction coefficient calculated by ProtParam. In order to remove the 6xHis-SUMO or TRX-6xHis tags, 5 µg of SenP2 protease or 10 µg of 3C protease per mg of protein, respectively, was added and the sample was transferred into pre-washed Spectra/Por® Dialysis Membrane tubing (molecular weight cut-off of 3.5 kDa, Spectrumlabs). The sample was then dialysed against buffer C (Gel filtration buffer; Table 2-11) overnight at 4 °C in order to remove imidazole and allow for protease cleavage. The samples for SDS-PAGE analysis were collected before addition of the protease and after completion of protease digest to assess efficiency of cleavage.

2.3.3.3 Second HisTrap purification

The dialysed sample was transferred into the superloop and injected into the HisTrap column (prewashed with buffer A) at a flow rate of 1 ml/min. As the protein of interest was no longer 6xHis-tagged, it did not bind to the column efficiently and was collected either in the flow-through or, as in the case of XerH and XerT, eluted with low concentration of imidazole (65 mM). Both SenP2 and 3C proteases contain 6xHis tag and therefore were retained on the column and later eluted with 50% of buffer B. Finally, the column was washed with 100% buffer B in order to remove all the remaining bound proteins. Samples were collected and fractionated throughout all steps, and were subsequently analysed by SDS-PAGE to identify the fractions containing the protein of interest. The fractions of interest were then combined and concentrated using a Vivaspin® Turbo 15 concentrator (molecular weight cut-off of 10 kDa, Sartorius stedim biotech) by centrifugation at 3000 *g* at 4 °C until the volume reached approximately 5 ml.

2.3.3.4 Size exclusion chromatography

The gel filtration column (HighLoad™ 16/60 Superdex™ 200, GE Healthcare) was equilibrated with 1.5 CV distilled water followed by 1 CV buffer C. The sample was loaded into a 5 ml sample loop and injected into the column at 1 ml/min. The column was then run for 1 CV in buffer C, during which the protein sample was separated by size and fractionated. The fractions were analysed by SDS-PAGE and the fractions containing the protein of interest were combined. The combined fractions were then dialysed step-wise

against the Acetate dialysis buffers 1, 2, and 3 or against the HEPES dialysis buffer (Table 2-11), with each dialysis step carried out for at least 4 hours at 4 °C. The dialysed proteins were concentrated to approximately 1 ml using a Vivaspin® Turbo 15 concentrator (molecular weight cut-off of 10 kDa, Sartorius stedim biotech) and kept on ice or frozen in liquid nitrogen and stored at -80 °C until further use.

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

Protein samples were routinely analysed by electrophoresis on 12% SDS-polyacrylamide gels. The gels consisted of a stacking and a resolving layer to ensure good sample separation. The recipes for each layer are shown in Table 2-12. Briefly, the glass plates were assembled in a Mini-PROTEAN® Handcast System (Bio-Rad) and the resolving layer was poured up to 1 cm below the comb depth. 1 ml of isopropanol was added on top to ensure even distribution of the gel. The gel layer was allowed to set for 30 minutes. The isopropanol was removed by decanting and the stacking layer was poured, after which the well comb was inserted. The gel was allowed to polymerize for 1 hour.

Table 2-12: Components of the 12% SDS-polyacrylamide gels.

Component[a]	Amount per 25 ml
5% stacking gel	
4x Stacking buffer (0.5 M Tris-Cl, pH 6.8, and 0.4% (w/v) SDS)	6.25 ml
30% Acrylamide:Bisacrylamide 37.5:1	4.25 ml
distilled water	14.5 ml
10% (w/v) APS	250 µl
TEMED	40 µl
12% resolving gel	
4x Resolving buffer (1.5 M Tris-Cl, pH 8.8 and 0.4% (w/v) SDS)	6.25 ml
30% Acrylamide:Bisacrylamide 37.5:1	10 ml
distilled water	8.75 ml
10% (w/v) APS	250 µl
TEMED	40 µl

[a] – SDS - sodium dodecyl sulphate; APS - ammonium persulphate; TEMED – tetramethylethylenediamine.

Generally, 30 µl of the sample to be analysed on SDS-PAGE was combined with 10 µl of 4x SDS loading buffer (200 mM Tris-Cl, pH 6.8, 24% glycerol, 8% SDS, 0.02% bromophenol blue, and 0.5 M dithiothreitol (DTT)) and 3-10 µl were loaded on the gel alongside the protein marker Mark12™ (Life Technologies). The gel was run at 200 V for 50 minutes.

After the run was finished, the gel was transferred into a plastic box, washed 3 times with 100 ml of distilled water, then 100 ml of Coomassie staining solution containing 0.075% (w/v) G250 Coomassie Brilliant Blue (Thermo Scientific) and 0.1% (v/v) HCl in water was added and the gel was incubated for 1 hour at room temperature with shaking. Finally, the gel was destained with 100 ml distilled water for 1 hour at room temperature with shaking and stored at room temperature until imaging. The products of gel electrophoresis were analysed under white light on a transilluminator (Alpha Innotech) using AlphaImager® HP software (Fisher Scientific).

2.3.5 Mass spectrometry

The efficiency of seleno-methionine incorporation in the XerH derivative (see section 2.3.2) was investigated by mass spectrometry performed by the EMBL Proteomics Core Facility. In brief, the sample after gel filtration was concentrated to >1 mg/ml and submitted to the Core Facility where mass determination of intact protein was performed on a Q-Tof2 iMass Spectrometer (Micromass/Waters). The raw data were deconvoluted using MaxEnt1 software (Micromass/Waters). Based on the single peak with assigned mass, the number of methionine sites occupied by seleno-methionine was estimated.

2.4 Biochemical methods

2.4.1 Annealing of DNA substrates

Short DNA substrates (up to 50 bp) were annealed by mixing equivalent amounts of oligonucleotides (stocks at 400 nM, 100 µM, or 1 mM concentration, resuspended in TE buffer), followed by incubation at 98 °C for 5 minutes and slow cooling (2-3 hours) to room temperature in the switched off heating block. Suicide substrates containing nicks in

both top and bottom strand were further cooled slowly to 4 °C to ensure proper annealing of the 4-bp overlap region.

2.4.2 Electrophoretic mobility shift assay (EMSA)

The electrophoretic mobility shift assay (EMSA) was carried out in 1x Activity buffer (25 mM HEPES, pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$, 5% glycerol, 1 mM DTT, and 1 mM EDTA). The DNA concentration was kept constant (0.5 or 4.5 μM) in all the samples while the concentration of the protein was gradually increased from 0 to 1.3 or 25 μM, as described in the appropriate results section. The reactions were incubated for 1 hour at 30 °C, then briefly centrifuged and loaded directly onto a native gel without addition of a loading dye. Low Molecular Weight DNA Ladder (NEB) was run as a reference along the samples. The components of the 12% native Tris-borate-EDTA (TBE) gel are listed in Table 2-13. The gels were run for 90 minutes at 12.5 V/cm of the gel at room temperature. The gels were stained with 1x SYBR® Gold Nucleic Acid Gel Stain (Life technologies) for 40 minutes and then visualized in the UV transilluminator (Alpha Innotech) using AlphaImager® HP software (Fisher Scientific).

Table 2-13: Composition of the 12% native TBE polyacrylamide gel.

Component[a]	Amount per 16 ml
10x TBE buffer	1.6 ml
30% Acrylamide:Bisacrylamide 37.5:1	6.4 ml
distilled water	8 ml
10% (w/v) APS	160 µl
TEMED	16 µl

[a] – 10x TBE buffer - 10x Tris-borate-EDTA buffer (1 M Tris base, 61.8 g/litre boric acid, 20 mM EDTA); APS - ammonium persulphate; TEMED – tetramethylethylenediamine.

2.4.3 DNA cleavage assay

The cleavage reactions contained 1x Activity buffer (see section 2.4.2), 25 μM protein of interest, and 25 – 33 μM DNA substrate in the total volume of 10 µl. The substrates used in the assay were suicide substrates containing a nick in top, bottom, or both DNA strands.

The control reactions contained no DNA. The reactions were allowed to proceed for 1-4 hours at 37 °C. To stop the reactions, 5 µl of hot 4x SDS loading buffer (see section 2.3.4) and 100 µM DTT were added and the reaction mix was incubated at 98 °C for 10 minutes to denature the protein. The whole sample volume was then loaded on an SDS gel and analysed as described in section 2.3.4. A shift to a higher molecular weight position indicated the presence of protein-DNA covalent intermediate.

2.4.4 Radioactive labelling of DNA substrates

All oligonucleotides used in labelling reactions were purchased unphosphorylated at the 5'-end. [γ-^{32}P]-ATP used for 5'-end labelling of oligonucleotides was purchased from PerkinElmer. The labelling reaction contained 1x T4 Polynucleotide Kinase Reaction Buffer, 10 units of T4 Polynucleotide Kinase (both from NEB), 2 µM oligonucleotide, and 92.5 MBq (2.5 mCi)/ml of [γ-^{32}P]-ATP in 10 µl reaction mix. The reactions were incubated at 37 °C for 1 hour, after which 40 µl of TE buffer was added. The kinase was then heat-inactivated by incubation at 80 °C for 30 minutes. In order to remove all unincorporated [γ-^{32}P]-ATP, the samples were applied to Micro Bio-Spin® Chromatography Columns (Bio-Rad) and purified according to the manufacturer's instructions. The final concentration of labelled oligos was 400 nM. The Oligo Length Standards 10/60 and 20/100 Ladder markers (IDT) were labelled as above. To 40 µl of the marker, 60 µl of distilled water and 100 µl of 2x Formamide loading buffer (1x TBE, 90% formamide, 0.005% xylene cyanol, and 0.005% bromophenol blue) were added and the markers were stored at 4 °C until further use.

2.4.5 DNaseI footprinting

The DNA substrates for DNaseI footprinting reactions were amplified by PCR from genomic DNA of *H. pylori* strains 26695 and P12 using [γ^{32}-P]-ATP 5'-end labelled primers (primers 21-28 as listed in Table 2-3) according to the protocol presented in section 2.2.3. For each reaction, only the forward or reverse primer was labelled, which resulted in amplification of a top- or bottom-strand labelled product, respectively. The PCR products were analysed on a 1.2% agarose gel and visualized briefly using UVGL-58 Handheld UV Lamp (UVP). The bands of interest were excised and the DNA was

extracted from the gel using a QIAquick Gel Extraction Kit (QIAGEN) according to the manufacturer's instructions.

The binding reaction mix contained 1x Activity buffer (described in section 2.4.2), 2 µl of radioactively labelled purified PCR product, and 2 µM of the protein of interest in 10 µl total reaction volume. The control samples contained no protein. The samples were incubated at 30 °C for 1 hour (XerH) or at 37 °C for 3 hours (XerT). Next, 50 µl of Ca^{2+}/Mg^{2+} solution (5 mM $CaCl_2$ and 10 mM $MgCl_2$) was added to provide cations necessary for DNaseI activity. The samples were incubated for 1 minute at room temperature. 3 µl of DNaseI at different dilutions ranging from 10 mg/ml to 0.0625 mg/ml were added to the samples and the samples were incubated for exactly 1 minute at room temperature. After 1 minute, 90 µl of Stop solution (0.2 M NaCl, 30 mM EDTA, and 1% SDS) was added to the sample to inactivate DNaseI. The DNA from the reactions was purified by ethanol precipitation: the sample was mixed with 1/10 sample volume of 3M sodium acetate (pH 5.2), 3 sample volumes of absolute ethanol, and 20 µg of glycogen (Thermo Scientific), mixed, and stored at -20 °C overnight. Next, the sample was centrifuged for 1 hour at 17000 *g* at 4 °C in a benchtop centrifuge. The supernatant was removed and the DNA was allowed to air-dry in the heating block at 37 °C for 10 minutes. The DNA was resuspended in 10 µl of distilled water, 1x Formamide loading buffer (section 2.4.5) was added, and the sample was kept at 4 °C until analysis on the sequencing gel.

The sequencing ladder was prepared using the DNA Cycle Sequencing Kit (Jena Bioscience) according to the manufacturer's instructions, but substituting the fluorescently labelled primer with 0.8 nmol of the [γ-^{32}P]-ATP-labelled primer used previously for DNA substrate amplification. The template used was a PCR product identical to the analysed reaction substrate, previously amplified with unlabelled primers. The resulting four samples containing sequencing ladders made with ddGTP, ddATP, ddTTP, or ddCTP were stored at 4 °C until analysis on the sequencing gel.

The 12% urea-TBE sequencing gel (17 cm x 34 cm) was prepared in a Model SA Sequencing Gel Electrophoresis Apparatus (Biometra) according to the recipe shown in Table 2-14. The gel was pre-run in 1x TBE buffer at 35 W (1 W/cm of the gel) for 45 minutes. The samples were then loaded on the gel in the following amounts: 2 µl of each sequencing ladder, 7 µl of each DNaseI footprint sample, and 3 µl of Oligo Length

Standard. The gel was run for 2 hours at 35 W (1 W/cm of the gel). After the run was finished, the gel was removed from the glass plates and a BAS-SR 2040 Phosphor Screen (FujiFilm) was exposed for 1 hour in a Hypercassette™ (Amersham Life Science) without gel drying. The screen was imaged in a Typhoon FLA 7000 Phosphoimager (GE Healthcare Life Sciences).

Table 2-14: Composition of the 12% sequencing TBE polyacrylamide gel.

Component[a]	Amount per 30 ml
10x TBE buffer	3 ml
40% Acrylamide:Bisacrylamide 19:1	9 ml
Urea	15 g
Distilled water	6 ml
10% (w/v) APS	0.5 ml
TEMED	10 µl

[a] – 10x TBE buffer - 10x Tris-borate-EDTA buffer (1 M Tris base, 1 M boric acid, 20 mM EDTA); APS - ammonium persulphate; TEMED – tetramethylethylenediamine.

2.4.6 DNA cleavage and strand-exchange assay

The DNA cleavage and strand-exchange assay was carried out with [γ-^{32}P]-ATP 5'-end labelled double-stranded DNA corresponding to the left- or right-end TnPZ substrates. The reaction mix contained 1x Activity buffer (as in section 2.4.2), 100 nM [γ-^{32}P]-ATP-labelled nicked suicide substrate, 33 µM unlabelled DNA substrate without a nick (corresponding to the same TnPZ end), and 16 µM of XerT in 15 µl reaction volume. The reactions were incubated for 4 hours at 37 °C. Next, the protein in the reaction was digested with Proteinase K: 15 µl of the cleavage reaction was incubated with 1x Proteinase K buffer (10 mM Tris-HCl, pH 7.5, 5 mM EDTA, 0.5% SDS) and 0.25 µg/µl Proteinase K (Carl Roth) in the total reaction volume of 200 µl at 50 °C for 1 hour. The samples were purified by ethanol precipitation as described in section 2.4.6. 1x Formamide loading buffer was added, after which the DNA was denatured by incubation at 98 °C for 5 minutes, and the sample was kept at 4 °C until analysis on the sequencing gel. 7 µl of each sample along with 3 µl of appropriate size markers were loaded and run on a 12% urea-TBE sequencing gel and the gel was imaged as described in section 2.4.5. The control reactions contained no XerH.

2.4.7 Half-site cleavage assay

The half-site cleavage assay was carried out with [γ-^{32}P]-ATP 5'-end labelled substrates corresponding to either half of the XerT binding site (a TnPZ end). The reaction mix contained 1x Activity buffer (as in section 2.4.2), 100 nM [γ-^{32}P]-ATP-labelled oligonucleotide corresponding to one half-site, 33 μM unlabelled oligonucleotide corresponding to the second half-site, and 16 μM protein of interest in 15 μl volume. The control reactions contained no protein or no unlabelled substrate. The reactions were incubated for 2 hours at 37 °C. Next, the samples were incubated with Proteinase K, and the DNA was ethanol precipitated and analysed on a 12% urea-TBE sequencing gel as described in section 2.4.6.

2.4.8 Holliday junction (HJ) resolution assay

The *dif*$_H$ HJ substrate was prepared by annealing four oligonucleotides (listed in Table 2-15) as described in section 2.4.1, with one of the oligonucleotides [γ-^{32}P]-ATP 5'-end labelled as described in section 2.4.4. The annealed HJ substrate is shown in Figure 2-4. The HJ resolution assay was carried out in 1x Activity buffer (see section 2.4.2) in the presence of 100 nM ^{32}P-labelled HJ substrate and 2 μM or 20 μM XerH. The reactions were incubated at 37 °C for 2 hours. Next, the samples were incubated with Proteinase K, and the DNA was ethanol precipitated and analysed on a 12% urea-TBE sequencing gel as described in section 2.4.6. The control reactions contained no XerH.

Table 2-15: Oligonucleotides used in HJ resolution assay.

Oligonucleotide	Sequence[a]
dif$_H$_HJ_1	CTGCCGTGATCACGCTGAACGCGTTTTAGCTTTAGTTATGAAAACTGCACTTTTCAAACTTTTAAATCGCCTAACGCCTAAAGCGGCCGCCTAGTCC
dif$_H$_HJ_2	GGACTAGGCGGCCGCTTTAGGCGTTAGGCGATTTAAAAGTTTGAAAAGTGCAGTTTTCATAACTAAACGACCGTTGCCGGATCCGCTGC
dif$_H$_HJ_3	GGATTCGAGATCTCAGGATGTGATTTAAAAGTTTGAAAAGTGCAGTTTTCATAACTAAAGCTAAAACGCGTTCAGCGTGATCACGGCAG
dif$_H$_HJ_4	GCAGCGGATCCGGCAACGGTCGTTTAGTTATGAAAACTGCACTTTTCAAACTTTTAAATCACATCCTGAGATCTCGAATCC

[a] – the 31-bp *dif*$_H$ site is marked in red

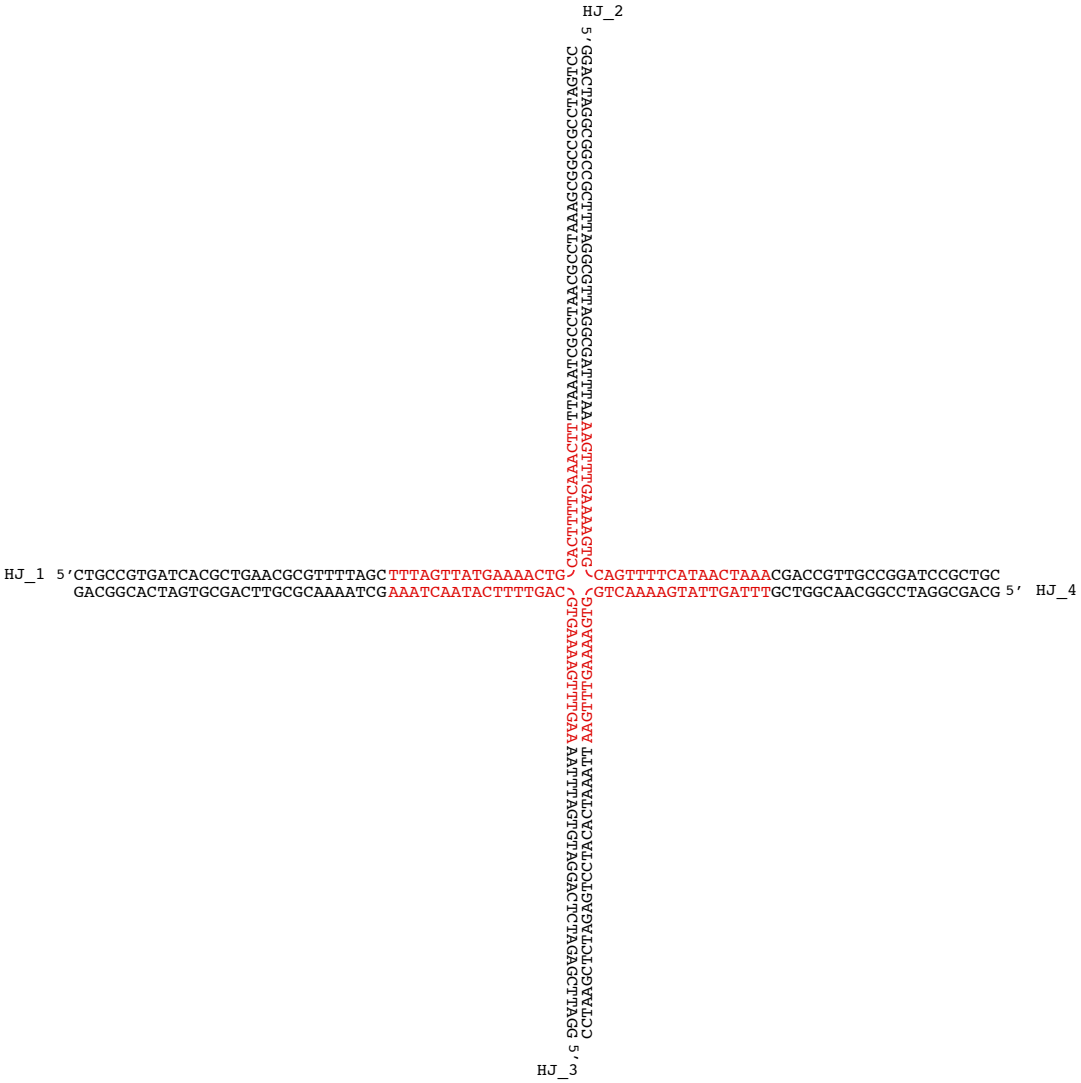


Figure 2-4: Schematic view of the *dif*$_H$-containing HJ substrate used in the HJ resolution assay. The *dif*$_H$ sites are marked in red. The 5' end of each oligonucleotide used is marked.

2.4.9 Pull-down assay with purified proteins

The assay was carried out with purified XerH and FtsK$_C$-A. XerH binds to the nickel column in low salt conditions without the 6xHis tag and can be eluted with imidazole, and therefore was used as "bait" in the assay. In detail, purified XerH and FtsK$_C$-A were dialysed overnight into HEPES dialysis buffer (Table 2-11). 4.6 nmol of XerH was bound to 250 µl of Ni-NTA nickel-charged agarose resin (QIAGEN) prewashed with HEPES dialysis buffer. To one reaction, 5 nmol *dif*$_H$-LP DNA substrate (see section 4.2.1) was added. XerH was incubated with the beads for 2 hours at 4 °C on a rocking platform. Next, the suspension was centrifuged for 1 minute at 1000 *g* at 4 °C in a benchtop centrifuge. To wash any unbound XerH, the resin was washed 3 times with HEPES dialysis buffer. Next, 1.8 nmol FtsK$_C$-A was added to the resin and the suspension was incubated for 1 hour at 4

°C on a rocking platform. The suspension was transferred into a Bio-Spin® Disposable Chromatography Column (Bio-Rad) and the unbound FtsK was allowed to drip through the column. The resin was then washed three times with HEPES dialysis buffer. Finally, the proteins bound to the column were eluted by addition of 500 mM imidazole in HEPES dialysis buffer. Two samples were collected at each stage of binding, washing, and elution and were later analysed by SDS-PAGE as described in section 2.3.4. Control reactions were carried out with one of the proteins absent from the experiment.

2.4.10 Affinity chromatography with *E. coli* lysate

This protocol was adapted from a previously published method (Vollmer *et al.,* 1999).

2.4.10.1 Preparation of *E. coli* protein fractions

2 litres of *E. coli* DS941 were grown at 37 °C to an OD_{600} of 0.5-0.6. The cultures were harvested at 4000 *g* for 20 minutes at 4 °C and resuspended in 20 ml of EF buffer I (Table 2-16). 1 tablet of cOmplete Protease Inhibitor Cocktail (Roche) was added and the cells were lysed using a Branson Sonifier 250 set to 50% duty cycle and 50% output control in 10 cycles of 30 seconds sonication altered with 60 seconds rest on ice. The lysate was then centrifuged to separate the soluble and membrane protein fractions at 40000 *g* for 45 minutes at 4 °C. The supernatant constituted the soluble fraction of *E. coli* proteins and was kept on ice until further use. The pellet, corresponding to the membrane fraction of *E. coli* proteins, was resuspended in 20 ml of EF buffer II (Table 2-16) and membrane proteins were extracted by overnight incubation on a rocking platform at 4 °C. The supernatant obtained after another ultracentrifugation step at 40000 *g* for 45 minutes at 4 °C constituted the solubilized fraction of membrane proteins. Both protein fractions were dialysed against 2 litres of EF dialysis buffer I, II, or III (Table 2-16), depending on the amount of salt to be used in affinity chromatography.

Table 2-16: Composition of buffers used in the affinity chromatography assay.

Buffer[a]	Composition[b]	pH
EF buffer I	10 mM Tris/maleate, 10 mM $MgCl_2$	6.8
EF buffer II	10 mM Tris-maleate, 10 mM $MgCl_2$, 1 M NaCl, 2% Triton X-100	6.8
EF dialysis buffer I	10 mM Tris/maleate, 10 mM $MgCl_2$, 50 mM NaCl	6.8
EF dialysis buffer II	10 mM Tris/maleate, 10 mM $MgCl_2$, 150 mM NaCl	6.8
EF dialysis buffer III	10 mM Tris/maleate, 10 mM $MgCl_2$, 400 mM NaCl	6.8
AC blocking buffer	200 mM Tris/HCl, 10 mM $MgCl_2$ 500 mM NaCl, 10% glycerol and 0.25% Triton X-100	7.4
AC acetate buffer	100 mM sodium acetate buffer, 10 mM $MgCl_2$, 500 mM NaCl, 10% glycerol and 0.25% Triton X-100	4.8
AC wash buffer I	10 mM Tris/maleate, 10 mM $MgCl_2$, 50 mM NaCl and 0.05% Triton X-100	6.8
AC wash buffer II	10 mM Tris/maleate, 10 mM $MgCl_2$, 150 mM NaCl and 0.05% Triton X-100	6.8
AC wash buffer III	10 mM Tris/maleate, 10 mM $MgCl_2$, 400 mM NaCl and 0.05% Triton X-100	6.8
AC elution buffer I	10 mM Tris/maleate, 10 mM $MgCl_2$, 1 M NaCl, 0.05% Triton X-100	6.8
AC elution buffer II	10 mM Tris/maleate, 10 mM $MgCl_2$, 2 M NaCl, 0.05% Triton X-100	6.8

[a] – EF – *E. coli* fractions; AC – affinity chromatography

[b] – 1 M Tris/maleate: 121.1 g Tris base per litre, pH adjusted to 6.8 with maleic acid; 1 M Tris/HCl: 121.1 g Tris base per litre, pH adjusted to 7.5 with HCl; 200 mM sodium acetate buffer: 176 ml 0.2 M acetic acid and 824 ml 0.2 M sodium acetate per litre.

2.4.10.2 Affinity chromatography

0.15 g per experiment of CNBr-activated Sepharose™ 4B beads (GE Healthcare) were activated following the manufacturer's instructions. 5 mg of purified XerH was dialysed overnight against HEPES dialysis buffer (Table 2-11) and coupled to 0.15 g of beads prewashed with the same buffer. The coupling was carried out overnight at 4 °C on a rocking platform. The beads were centrifuged at 3000 *g* for 10 minutes at 4 °C and then washed with 10 ml HEPES dialysis buffer. The remaining coupling sites were blocked with 10 ml AC blocking buffer (Table 2-16) by incubation at 4 °C for 5 hours on a rocking platform. The beads were washed three times with 10 ml AC blocking buffer alternating with 10 ml AC acetate buffer (Table 2-16), and resuspended in EF dialysis buffer I, II, or III (Table 2-16). Finally, dialysed *E. coli* protein fraction (soluble or membrane) was added and the sample was incubated overnight at 4 °C on a rocking platform. The beads were transferred to a Poly-Prep® Chromatography Column (Bio-Rad) and the liquid was allowed to flow through. The column was washed with 20 ml of AC wash buffer I, II or III (Table 2-16) depending on the salt concentration used previously for dialysis of the applied sample. The 500 µl samples for analysis were collected at the beginning and at the end of the wash step. Retained proteins were eluted with 1.2 ml of AC elution buffer I (Table 2-16) for the experiment with 50 mM NaCl or AC elution buffer II (Table 2-16) for the experiments with 150 mM and 400 mM NaCl after 30 minutes incubation in the buffer. The whole eluent was collected and kept at -80 °C before analysis by Mass Spectrometry. The Sepharose™ beads were recharged for the next experiment by alternating washes with

AC elution buffer, AC acetate buffer, and AC wash buffer. As a control, one batch of activated Sepharose™ was treated identically, except that no XerH was coupled to the beads. All samples collected throughout the experiments were analysed by SDS-PAGE as described in section 2.3.4.

2.4.10.3 Mass Spectrometry

The identity of soluble and membrane proteins that bind to XerH was investigated with mass spectrometry performed by EMBL Proteomics Core Facility. In brief, 200 µl of the eluted sample was submitted to the Core Facility where it was digested with trypsin and analysed on a LTQ Orbitrap Velos Pro Mass Spectrometer (Thermo Fisher) with Nano Acquity UPLC System for liquid chromatography (Waters). The raw data was searched with Mascot (v 2.2.07) against the Uniprot_*E.coli* database. Search files were uploaded to and analysed using Scaffold3 (Proteome Software).

2.4.11 Analytical size exclusion chromatography (SEC)

Complexes of XerH and XerT with a variety of DNA substrates were analysed by analytical size exclusion chromatography (SEC) on Superdex 200 3.2/30 gel filtration column (GE Healthcare) using ÄKTAettan and ÄKTAmicro liquid chromatography systems (GE Healthcare) according to the manufacturer's instructions.

The XerH-DNA complexes were prepared as follows: 1 ml of purified XerH after gel filtration (concentration: 10-12 µM) was combined with double-stranded DNA substrate to give desired protein:DNA ratio. The sample was then dialysed step-wise against Acetate dialysis buffers A, B, and C (Table 2-11), with each dialysis step carried out for at least 4 hours at 4 °C. The complexes were then concentrated using Amicon® Ultra – 0.5 ml Centrifugal Filters with a molecular weight cut-off of 10 kDa (Merck Millipore) according to the manufacturer's instructions until the volume was reduced to 100 µl. The samples were then filtered using Centrifugal Filter Units (Millipore).
The XerT-DNA complexes were prepared as follows: Purified XerT after gel filtration was dialysed step-wise against Acetate dialysis buffers A, B, and C (Table 2-11). The protein was then concentrated in Amicon® Ultra – 0.5 ml Centrifugal Filters with molecular weight cut-off of 10 kDa (Merck Millipore) until the concentration reached 30-35 µM.

Next, the protein was combined with double-stranded DNA substrate to give the desired protein:DNA ratio, and incubated on ice for 2 hours. The samples were then filtered using Centrifugal Filter Units (Millipore).

In each case, 10 µl of the complex was loaded onto the gel filtration column prewashed with the Acetate dialysis buffer C and run in this buffer at 0.05 ml/min for 1.5 CV. UV absorbance chromatograms at 280 nm and 260 nm were collected and analysed using UNICORN Control Software (GE Healthcare). In some cases, the samples were additionally analysed by loading on a 12% native TBE gel, prepared and run as described in section 2.4.2.

2.5 X-ray crystallography methods

2.5.1 Preparation of XerH-DNA complexes for crystallization

In order to increase the chance of crystallization, XerH-DNA complexes were prepared as for SEC experiment with at least 95%-pure protein, as determined by SDS-PAGE analysis. XerH-DNA complexes, varying in DNA length and protein:DNA ratio, were analysed by analytical SEC to select the most promising crystallization targets. Only the complexes that showed a homogenous peak in the analytical gel filtration run were used for crystallization trials; these were the complexes with 2.2-fold and 2.5-fold excess of XerH monomer over DNA substrate. The complexes formed with more symmetric DNA substrates (such as palindromes) were favoured as the presence of internal DNA symmetry is thought to facilitate crystallization. The concentration of the complexes was estimated assuming that the protein amount remained the same before and after complex formation and concentration.

2.5.2 Crystallization of XerH-DNA complexes

In order to form crystals of macromolecular complexes such as XerH-dif_H, the macromolecular solution has to be brought to a supersaturated state. In the vapour diffusion method, the supersaturation of the macromolecule is accomplished by increasing the concentration of the macromolecule and the precipitant simultaneously. In detail, a 0.1-

2 µl drop of the macromolecule solution is mixed with a 0.1-2 µl drop of the precipitant solution and the mixture is placed in a sealed environment with a larger (0.5-1 ml) reservoir of the precipitant solution. The concentration of the precipitant in the reservoir is higher than in the drop (since the precipitant concentration in the drop was diluted by addition of the macromolecule solution) and as a result, water vapour will move from the drop to the reservoir, resulting in simultaneous increase in macromolecule and precipitant concentration within the drop. The vapour diffusion method is the most common method for crystallization of macromolecules due to the easy handling, adaptability for large-scale screening of precipitation conditions, and adjustability and good control over protein to precipitant ratio. This method was therefore chosen for crystallization of XerH-DNA complexes.

Since the suitable crystallization conditions (precipitant identity and concentration, pH, buffer conditions, etc.) have to be determined experimentally for each macromolecule, a large number of conditions was screened to find the crystallogenic condition for XerH-DNA complex. The initial crystallization screens were prepared by the EMBL Crystallization Facility and included commercially available crystallization screens The Classics Suite (QIAGEN) and Index (Hampton Research), as well as the JCSG+ screen (Page *et al.*, 2003) prepared in-house. These screens employ a sparse matrix approach in which a wide range of chemically different precipitants and additives is tested in a non-systematic way. This approach allows for many non-redundant conditions to be tested in order to identify compounds that might allow crystallization, but in turn requires further optimization of the crystallization conditions. The screens were prepared in 96-well MRC sitting drop Crystallization Plates™ (Molecular Dimensions) using a Mosquito Crystallization Robot (TTP Labtech). To a 0.1 µl drop of the reservoir precipitant solution, 0.1 µl of Xer-DNA complex solution was added and the plates were sealed. The screens were handled and stored for crystal growth at 6 °C.

The initial condition that yielded small crystals (JCSG+ 65: 0.1 M HEPES, pH7.5, 0.2 M $MgCl_2$, and 30% (w/v) polyethylene glycol (PEG) 400) was further optimized to yield larger crystals. For this, fine screening of varying PEG 400 concentrations against varying pH was manually set up in 24-well crystallization plates (Hampton Research), as shown in Figure 2-5. After 500 µl of specified solution was placed in the reservoir, a 2 µl drop of the mother liquor was set on the glass cover slip and 2 µl of XerH-DNA complex solution was added. The cover slips were placed over the reservoir (so that the drop was hanging

downwards over the reservoir solution) and sealed, allowing for vapour diffusion to occur. The plates were set up and stored at 6 °C. This setup allowed for fine-tuning of parameters such as precipitant amount, pH, protein-DNA complex concentration, DNA length, and precipitant-to-complex ratio in large drops that supported growth of large crystals. Additionally, hanging drop 24-well plates at room temperature and sitting drop 24-well plates at 6 °C were also set up but yielded no crystals.

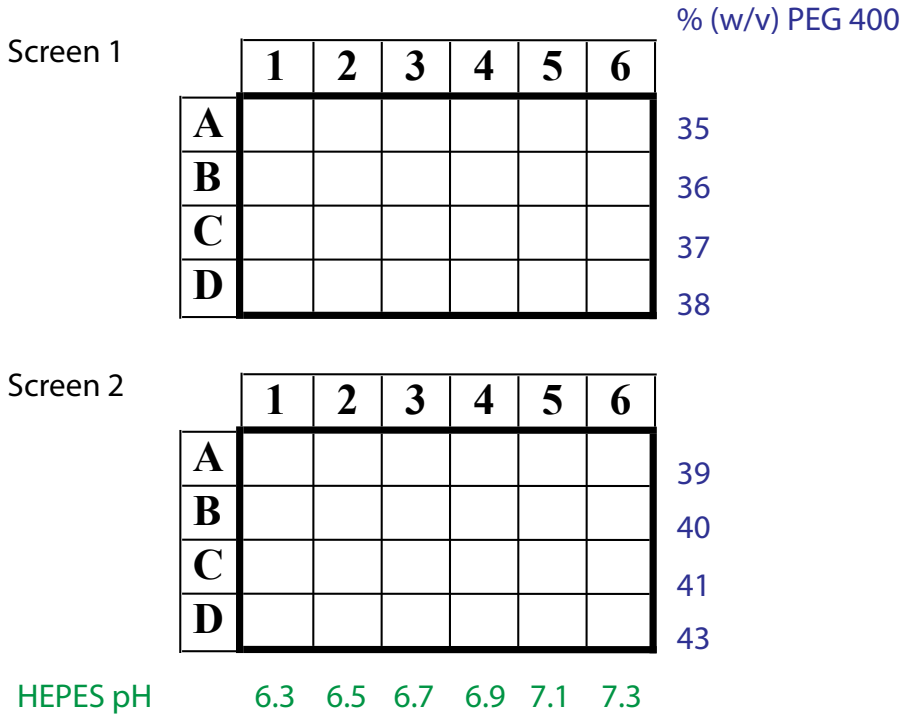


Figure 2-5: Optimization screen of crystallization parameters (precipitant concentration and pH) in 24-well crystallization plates. All wells contained 0.1 M HEPES buffer at the indicated pH, 0.2 M $MgCl_2$, and PEG 400 at the indicated concentrations.

2.5.3 Heavy atom derivative crystals

To solve the phase problem (see section 4.1.1), heavy atom derivatization of XerH crystals was performed by crystal soaking. The mercury derivative crystals were prepared by transferring the crystals to a 2 µl drop of the reservoir solution supplemented with 1 µM p-chloromercuribenzene sulphonate (PCMBS) for 5-30 minutes, and then transferring the crystal back to a fresh drop of reservoir solution to wash off any unbound mercury atoms (so called back-soaking). The crystal was immediately fished and frozen in liquid nitrogen as described in section 2.5.4. The iodine derivative crystals were made by transferring the crystals to a 2 µl drop of the reservoir solution supplemented with 250 mM or 500 mM of

the “Magic Triangle” 5-Amino-2,4,6-triiodoisophthalic acid – I3C (JBS Magic Triangle Phasing Kit, Jena Bioscience) for 5-25 minutes without back-soaking. The crystals were immediately fished and frozen in liquid nitrogen as described in section 2.5.4. In both cases, the crystals were inspected under a microscope during soaking to assess damage to the crystal, and only the crystals not showing visible cracks or signs of dissolving were fished and used for data collection.

Additionally, Se-Met derivative crystals were obtained with purified Se-Met-containing XerH, following the same complex formation and crystallization protocols as for native complex crystallization (see sections 2.5.1 and 2.5.2).

2.5.4 Data collection

Crystals obtained from 24-well plates were fished using 0.05 – 0.4 mm CryoLoops (Hampton Research) and transferred immediately into liquid nitrogen-containing MicroTubes (Hampton Research).

Crystals must normally be transferred into a solution containing a cryoprotecting agent such as glycerol or low molecular weight PEG before freezing in liquid nitrogen, otherwise crystalline ice might destroy the protein by volume expansion, or obscure the diffraction pattern through so called ice-rings. According to McFerrin and Snell, 2002, the condition from which the XerH-DNA crystals were obtained is considered cryo-protected by the high amount of PEG 400, and therefore no additive was required for freezing the crystals. The crystals were transferred to the X-ray source in Cryogenic Dewar Flasks (TED PELLA) under cryogenic conditions.

X-ray diffraction data from native and derivative XerH-DNA crystals were collected by the rotation method on tuneable beamlines ID-23 and ID-29 at the European Synchrotron Radiation Facility (ESRF) in Grenoble, France. The samples were mounted onto the goniometer by an automatic sample changer and rotated as the data were collected. The details of data collection for the best native and derivative datasets are shown in Table 2-17. All datasets were collected following the collection strategy designed by EDNA diffraction characterization and data collection strategy software. EDNA software is implemented in the data collection pipeline at ESRF beamlines. Based on initially

collected 2-4 images, EDNA estimates the collection strategy (number of images, percentage of the beam energy transmission, detector distance) considering the crystal's size, symmetry, and space group, and accounting for predicted radiation damage of the crystal. The derivative datasets were collected at X-ray wavelengths (shown in Table 2-17) that corresponded to the absorption peak of the appropriate heavy atom as determined by an initial fluorescent energy scan. While usually two additional datasets are collected at wavelengths corresponding to the inflection point and a remote position to allow for phasing with the multi-wavelength anomalous diffraction (MAD) approach (see section 4.1.1), in this study the radiation damage to the crystals did not allow for further data collection.

Table 2-17: Data collection strategies for the best obtained native and derivative datasets.

Crystal	Beamline	Wavelength (Å)	f'	f''	Number of images	Oscillation (°)	Transmission (%)	Exposure (s)	Estimated resolution (Å)
XerH-dif_H-LP									
Native	ID-29	0.979	-	-	2160	0.1	4.21	0.037	2.45
Mercury derivative	ID-29	1.00	-11.85	10.52	1240	0.1	4.5	0.037	3.4
Iodine derivative	ID-29	1.77	n/d	n/d	927	0.15	4	0.037	3.4
Se-Met derivative	ID-29	1.03	-7.78	5.47	2280	0.05	3.23	0.037	2.8
XerH-dif_H									
Native	ID-23.2	0.88	-	-	3600	0.1	100	0.1	2.7

n/d – not determined (as the iodine absorption edges do not correspond to wavelengths accessible at ESRF ID-29, a wavelength of 1.77 Å at which anomalous scattering was expected to be sufficient was chosen without performing an energy scan)

2.5.5 Data processing

All datasets were initially processed in XDS, part of the XDS Program Package (Kabsch, 2010), to select the datasets with highest resolution and the best statistics. Briefly, XDS sequentially refines the data collection parameters such as beam divergence and crystal-to-detector distance, finds strong reflections from a number of collected diffraction images, indexes strong reflections and determines crystal lattice parameters, defines what should be considered a diffraction spot and what background in the diffraction images, integrates the intensities of all defined reflections, scales symmetry-related reflections based on the selected crystal lattice parameters, and finally prints out the data statistics. The initially predicted unit cell parameters were refined in an iterative process until convergence. The reflections related by the Friedel's law (see section 4.1.1) were considered to be equivalent only in the case of native datasets. An example of the data statistics as reported in the CORRECT.LP log file is shown in Figure 2-6. The resolution cut-off was defined at this

stage at I/σ(I) (a measure of signal-to-noise ratio of the intensities) of 2 (meaning that the observed signal was twice as strong as the observed noise) and unless otherwise specified only the data above this value were used for further processing. The other factors considered in the assessment of the data quality were: data completeness (the percentage of unique reflections observed in the experiment versus the number of reflections expected to be seen for the given crystal lattice; preferred over 98%), the R-meas value (an indicator of data consistency; preferred as low as possible in high resolution shells and close to 10% for the whole dataset), and CC(1/2) (another indicator of data consistency (Karplus and Diederichs, 2012; Diederichs and Karplus, 2013); preferred value above 60). For the derivative datasets, SigAno values above 1 and AnomalCorr values above 10 indicated presence of the anomalous signal. The derivative datasets with anomalous signal extending to the highest resolution were used for solving the phase problem.

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

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR COMPARED expected	I/SIGMA	R-meas	CC(1/2)	Anomal Corr	SigAno	Nano
10.00	3248	1184	1247	94.9%	2.1%	2.4%	3152	37.86	2.6%	99.9*	92*	3.610	415
6.00	12223	4372	4470	97.8%	3.4%	3.5%	11880	24.09	4.2%	99.8*	75*	2.244	1697
5.00	12034	4123	4161	99.1%	6.2%	6.3%	11813	16.42	7.5%	99.4*	57*	1.524	1733
4.70	5181	1996	2027	98.5%	7.1%	7.0%	4935	14.04	8.8%	99.1*	44*	1.288	734
4.50	4547	1609	1626	99.0%	7.8%	7.8%	4437	13.93	9.5%	99.1*	41*	1.245	659
4.40	2774	970	975	99.5%	8.1%	7.8%	2725	14.23	9.9%	98.8*	34*	1.125	415
4.30	2921	1012	1021	99.1%	9.0%	8.9%	2863	12.53	11.0%	98.9*	34*	1.122	428
4.20	3242	1119	1131	98.9%	9.4%	9.8%	3184	11.59	11.4%	99.0*	32*	1.086	474
4.10	3531	1222	1238	98.7%	10.1%	10.2%	3475	11.97	12.3%	98.7*	27*	1.046	525
4.00	4075	1383	1396	99.1%	12.2%	12.1%	4010	10.08	14.8%	98.3*	23*	1.033	599
3.60	19280	7048	7147	98.6%	19.6%	19.2%	18664	6.83	24.1%	95.7*	18*	0.930	2819
3.40	13546	4851	4957	97.9%	31.5%	31.5%	13222	4.36	38.6%	90.6*	10*	0.844	2008
3.30	8179	2854	2911	98.0%	41.1%	40.8%	8033	3.35	50.2%	87.4*	12*	0.806	1228
3.20	9364	3242	3291	98.5%	52.0%	52.6%	9212	2.60	63.6%	77.1*	13*	0.844	1415
3.10	10842	3738	3804	98.3%	67.5%	67.4%	10656	2.07	82.5%	69.1*	6	0.782	1616
total	114987	40723	41402	98.4%	11.6%	11.6%	112261	10.59	14.2%	99.3*	40*	1.199	16765

Figure 2-6: Example data statistics table from the CORRECT.LP file, output of XDS data processing. The data are statistically analysed by resolution shells. I/σ(I) is a measure of signal-to-noise ratio. R-meas and CC(1/2) are statistical indicators of data consistency. AnomalCorr and SigAno are indicators of anomalous signal strength.

Together, the information from the data statistics table gave an indication for which datasets should be chosen for further data processing, and these datasets were internally scaled in XSCALE from the XDS Program Package, producing a final table of data statistics that are reported in the results sections. The data were cut to the resolution where the observed signal was significant in XSCALE. Since it was shown recently that data even below I/σ(I) of 2 can be useful for later refinement of the model as long as the CC(1/2) values are high, in some cases the data were cut at slightly higher resolution than indicated by the I/σ(I) >2 cut-off (Diederichs and Karplus, 2013). The reflection file was then converted to .mtz format using XDSCONV from the XDS Program Package, and the resulting reflection file was used for structure solution as described below.

2.5.6 Experimental phasing

The phase problem (introduced in section 4.1.1) for XerH-*dif*$_H$-LP structure was solved by single-wavelength anomalous diffraction (SAD) phasing using the Se-Met derivative dataset in autoSHARP (Vonrhein *et al*., 2007). The automatic structure solution pipeline involved heavy atom search by SHELXD (Schneider and Sheldrick, 2002), refinement of the heavy-atom positions and phasing by SHARP (Bricogne *et al*., 2003), and density modification by SOLOMON (Abrahams and Leslie, 1996). Both the Se-Met derivative (3.1 Å) and native (2.4 Å) reflection files were used as input for autoSHARP. In detail, autoSHARP was instructed to find 12 selenium atoms from two XerH monomers present in an asymmetric unit as estimated by calculation of the Matthews coefficient (a coefficient used to estimate the number of macromolecules per asymmetric unit; Matthews 1968) using the MATTHEWS_COEF program from the CCP4 suite. The f' and f'' values of -7.78 and 5.47, respectively, were given in the input of autoSHARP run. After the initial phases were determined by the SAD method, density modification was performed and the phases were extended onto the native dataset.

2.5.7 Model building, refinement and validation

The initial model was automatically built by the phenix.autobuild program of the Phenix suite (Adams *et al*., 2010). The reflection file containing the phases obtained from the autoSHARP run and the sequence of XerH were used as input files. The model was refined during building in 3 cycles, with final R-work of 0.32 and R-free of 0.37. To improve the phases, the density map, and the model, real-space optimization and reciprocal-space refinement rounds were performed iteratively until the final satisfactory model was obtained. The real-space optimization included manual model rebuilding and automated real-space refinement in *Coot* software (Emsley *et al*., 2010). All operations in *Coot* were performed with the difference electron density maps 2Fo-Fc typically contoured at 1.2 sigma and Fo-Fc contoured at 3 sigma. DNA was built manually using DNA from the Cre synaptic complex crystal structure (PDB accession number 1CRX) as a guide. Each round of real-space optimization involved visual inspection of all residues for their fit to the electron density, building new residues into visible electron density, and correction of geometry to improve the protein or DNA geometry. The reciprocal-space refinement was performed using phenix.refine from the Phenix suite. In each round of refinement, XYZ

coordinates, individual B-factors, and the translation/libration/screw (TLS) parameters were refined in 3 consecutive cycles. X-ray/stereochemistry restraint and X-ray/atomic displacement parameters (ADP) restraint weights were optimized and experimental phase restraints were applied in the refinement. In the first round of refinement, an initial rigid-body fitting to the data was also performed before positional refinement of individual coordinates. Additionally, Phenix was allowed to automatically correct N/Q/H errors and to automatically link the tyrosine of XerH to the phosphate of the DNA backbone ("Automatic ligand linking" option). Water molecules were built in at a later stage of refinement in phenix.refine and validated manually in *Coot*.

2.5.8 Molecular replacement

The initial attempts to solve the crystal structure of the XerH-dif_H-LP complex used the molecular replacement method (introduced in section 4.1.1). Here, various tyrosine recombinase structures and structure-derived models (summarized in Table 4-1) were used as search models in Phaser-MR (McCoy *et al.*, 2007), part of the Phenix suite. No correct solutions could be identified, and therefore experimental phasing methods were employed as described in section 2.5.6.

The crystal structure of XerH-dif_H at 3 Å resolution was solved through molecular replacement using one XerH monomer from the XerH-dif_H-LP structure as a search model in Phaser. Phaser was instructed to search for 2 copies of the XerH monomer using the full search method. Phaser found a unique solution with the translation function Z-score (TFZ score) of 9.2 and log-likelihood gain (LLG) score of 1679.8. The TFZ score is a measure to judge the signal-to-noise ratio, and is normally expected to be higher than 7 for the correct molecular replacement solution, while the LLG score measures how well the model agrees with the data and should have a high positive value in the correct solution. Therefore, both values obtained in molecular replacement for the XerH-dif_H structure were consistent with a correct solution being found. Initial refinement in phenix.refine resulted in R-work value of 0.34 and R-free value of 0.41, which were both lower than R values of ≈0.6 expected at random for an incorrect solution. The model was then refined as described in section 2.5.7. As the resolution of 3 Å does not allow one to confidently build water atoms, these were not added to the final model for this structure.

2.5.9 Visualization of the models and XerH-DNA contacts

The obtained models were visualized using the PyMOL Molecular Graphics System (Version 1.5.0.4, Schrödinger, LLC). Superimposition with other models was performed in *Coot* using secondary-structure matching (SSM) superimposition function (Krissinel and Henrick, 2004). The interactions between XerH monomers and DNA were mapped using NUCPLOT v.1.1.4 - a program to generate schematic diagrams of protein-nucleic acid interactions (Luscombe *et al.,* 1997). The interactions were verified by visual inspection and distance measurements in *Coot*. The DNA parameters were calculated using 3DNA: a suite of software programs for the analysis, rebuilding and visualization of 3-Dimensional Nucleic Acid structures (Zheng *et al.,* 2009).

2.6 Microbiology methods

2.6.1 Reporter plasmid constructs

Plasmids used for *in vivo* and *in vitro* recombination assays were constructed with restriction-enzyme and restriction-free cloning as described in sections 2.2.2 to 2.2.9. Low-copy plasmids for the *in vivo* recombination assay with a *galK* marker were based on plasmid pMS183Δ containing *galK* cassette under *argR* promoter. A 100 bp DNA fragment containing the dif_H site was cloned upstream of the *galK* gene using NheI and BsrGI restriction enzymes and downstream of the *galK* gene using EcoRI and KpnI restriction enzymes using primers 21-24 (Table 2-3), resulting in plasmid pAB102 (Table 2-2). The same sets of restriction enzymes were used to introduce the left end of TnPZ upstream and the right end of TnPZ downstream of the *galK* gene, using primers 25-28 (Table 2-3), resulting in plasmid pAB106 (Table 2-2). Additionally, a higher copy number plasmid suitable for recombination in liquid *E. coli* cultures and *in vitro* recombination assays was constructed by amplification of the whole dif_H-*galK*-dif_H cassette and restriction-free cloning of the cassette into pACYC177 plasmid using primers 29 and 30 (Table 2-3), resulting in the plasmid pAB122 (Table 2-2).

2.6.2 Xer recombination with *galK* marker in *E. coli*

XerH and XerT excisive recombination in *E. coli* was assessed by an *in vivo* recombination assay with *galK* marker similar to that described by Arnold *et al.,* 1999. The assay utilizes MacConkey agar medium that can be supplemented with a carbohydrate of choice (here: galactose). Bacteria that can utilize the carbohydrate will produce acid as a metabolic product, which in turn will reduce the pH of the surrounding medium, resulting in a red colour of the colonies. Bacteria that cannot utilize the carbohydrate will instead utilize peptone present in the medium, which in turn leads to formation of ammonia and increase of the pH resulting in a white appearance of the colonies. The MacConkey agar medium can therefore be efficiently used as a colour marker for the presence of the *galK* gene in bacteria. The strains used in this assay, DS941 (wild-type) and DS9041 ($\Delta FtsK_C$), are deficient in GalK production and therefore grow as white colonies on galactose-supplemented MacConkey agar. The reporter plasmids used in this study (pAB102 for XerH and pAB106 for XerT as described in section 2.6.1) carry the *galK* gene, and *E. coli* cells transformed with those plasmids grow red on MacConkey agar. Upon arabinose-induced expression of XerH or XerT from the expression plasmids (XerH_pBAD/MCS or XerT_pBAD/MCS; Table 2-2), the two dif_H sites or the two TnPZ ends, respectively, can be recombined resulting in excision and loss of the GalK marker, leading to appearance of white *E. coli* colonies.

Initially, electro-competent *E. coli* cells were transformed with the expression plasmid (XerH_pBAD/MCS or XerT_pBAD/MCS) as described in section 2.2.8.1, and the transformants were grown overnight in LB at 37 °C. Ca^{2+} chemically-competent cells were prepared from the transformants according to the following protocol: 8 ml of LB containing ampicillin was inoculated with 40 µl of the overnight transformant cultures and grown to OD_{600} of 0.4-0.5. The cells were chilled on ice for 20 minutes. Next, 1.5 ml of the culture was transferred to an eppendorf tube and centrifuged at 10000 *g* for 1 min at 4 °C. The supernatant was removed and the cells were resuspended in 1.5 ml ice-cold 50 mM $CaCl_2$ and mixed by pipetting. The cells were centrifuged as before and the supernatant was removed. Finally, the cells were resuspended in 200 µl of ice-cold $CaCl_2$ and left on ice for at least 1 hour. Such chemically competent cells retained their competence for up to 24 hours when kept on ice but lost their competence upon freezing. The cells were transformed as described in section 2.2.8.2 with the reporter plasmid of choice, and plated on MacConkey agar medium containing 4% (w/v) MacConkey agar base (Difco), 1%

(w/v) galactose, kanamycin, ampicillin, and various amounts (0, 0.02, or 0.05%) arabinose. The resulting double transformants could appear red, white, or pink on MacConkey agar depending on the number of reporter plasmids still carrying the *galK* gene. As controls, untransformed DS941, and DS9041 as well as single transformants with reporter plasmids were plated on MacConkey agar.

The reporter plasmid is not a single-copy plasmid and therefore a single cell could contain both recombined and non-recombined plasmids giving a false-negative phenotype (red or pink colonies). Additionally, using arabinose on MacConkey agar plates to induce XerH expression is expected to interfere with the red/white phenotype. This is because the *araC*-mutation of DS941 (see Table 2-1) resulting in the strain appearing as red colonies on MacConkey agar can be complemented by the *araC* gene carried by the pBAD plasmids. The *araC* gene product is a transcriptional activator of the araBAD genes required for expression of arabinose metabolic genes. The pBAD expression vectors carry the *araC* gene to allow regulation of gene expression from the pBAD promoter with arabinose in any *E. coli* strain. Thus, in the presence of pBAD plasmid DS941 cells are able to utilize arabinose as a carbon source and appear white on MacConkey agar supplemented with arabinose, regardless of XerH recombination at dif_H.

To overcome these problems, DNA was isolated from colonies that had been grown in the presence of arabinose and introduced into DS941 or DS9041 as follows: the transformant plates were scraped with 1 ml of LB and 1 µl of the cell suspension was transferred into 6 ml of fresh LB with kanamycin to select only for the reporter plasmid. The culture was then grown overnight at 37 °C. Plasmid DNA was purified using QIAprep Spin Miniprep Kit (QIAGEN) according to the manufacturer's instructions and subsequently 1 µl of the plasmid DNA was transformed into *E. coli* DS941 or DS9041 and plated on MacConkey agar medium with kanamycin. As the reporter plasmid is a low copy number plasmid and is therefore transformed at low DNA concentrations, each transformant colony is expected to arise from transformation with a single DNA molecule. Thus the ratio of red to white colonies should reflect the ratio of recombined to unrecombined plasmid DNA in the plasmid mixture obtained from the *in vivo* recombination assay. Accordingly, recombination rates were calculated by dividing the number of white colonies by the total number of cells in each experiment.

To further confirm that *galK* was lost in white colonies through site-specific recombination precisely at the expected sites, overnight cultures of white and red colonies were prepared and plasmid DNA was extracted using a QIAprep Spin Miniprep Kit (QIAGEN). The plasmid DNA was analysed on a 0.8% agarose-TBE gel free from ethidium bromide, run overnight in TBE buffer at 3 V/cm of the gel. The gel was then stained by soaking in 4 µg/ml ethidium bromide in TBE buffer for 45 minutes and then destained in distilled water for 30 minutes. The products of gel electrophoresis were analysed on a UV transilluminator (Alpha Innotech) using AlphaImager® HP software (Fisher Scientific). Additionally, the plasmid DNA from both red and white colonies was analysed by sequencing (see section 2.2.12).

2.6.3 Xer recombination assay in liquid culture

XerH recombination at dif_H was also assessed in *E. coli* DS941 (wild-type) and DS9041 ($FtsK_C\Delta$) liquid cultures grown over 80 generations. In detail, electro-competent *E. coli* cells were double-transformed with the expression plasmid (XerH_pBAD/MCS; Table 2-2) and the reporter plasmid (pAB122, Table 2-2) as described in section 2.2.8.1, and plated on LB agar containing kanamycin and ampicillin, as well as 0.2% glucose to repress the arabinose promoter and prevent expression of XerH. Next, overnight cultures in LB supplemented with the antibiotics and 0.2% glucose were prepared from single transformant colonies. 400 ml of fresh LB containing antibiotics and glucose was then inoculated with 2 ml of the overnight culture and the cultures were grown to OD_{600} of 0.6-0.7. A sample (20 ml) was collected for analysis before harvesting the cells by centrifugation at 4000 *g* for 20 minutes at 4 °C. The cells were then resuspended in 400 ml LB containing appropriate antibiotics and 0.05% arabinose to induce XerH expression. The cultures were grown for 24 hours, after which a 5 ml sample for analysis was collected, and fresh LB with antibiotics and arabinose was inoculated with 1 ml of the culture. This was repeated 10 times to give the total generation number of 80, given that growth back to stationary phase after a dilution of 1 in 400 is equivalent to between 8 and 9 cell doublings or generations. As controls, cultures transformed only with the expression plasmid or the reporter plasmid were prepared and treated as described above for double transformants.

The plasmid DNA was extracted from the collected samples immediately after collection using QIAprep Spin Miniprep Kit (QIAGEN). The plasmid DNA was analysed on a 0.8%

agarose-TBE gel free from ethidium bromide, run overnight in TBE buffer at 3 V/cm of the gel. The gel was then stained with 4 μg/ml ethidium bromide in TBE buffer for 45 minutes and destained in distilled water for 30 minutes. The products of gel electrophoresis were analysed on a UV transilluminator (Alpha Innotech) using AlphaImager® HP software (Fisher Scientific).

2.6.4 Replacement of *E. coli dif* with dif_H

In order to replace the *E. coli* chromosomal *dif* site with *H. pylori* dif_H, a previously described method employing the Red recombination system was utilized (Datsenko and Wanner, 2000). Bacteriophage λ Red promotes efficient homologous recombination of transformed linear DNA, which normally gets degraded because of the presence of *E. coli* exonucleases. This allows for efficient gene disruption or replacement through homologous recombination with as little as 36 nt of homology at each end of the linear DNA. In detail, electro-competent *E. coli* DS941 (wild-type) and DS9041 ($\Delta FtsK_C$) cells were transformed with pKD46 plasmid carrying the Red genes. This plasmid is temperature-sensitive, and the solid and liquid cultures where it was to be maintained were grown at 30 °C. After transformation, the cells containing the pKD46 plasmid were grown to OD_{600} of 0.6 in LB with antibiotics and 10 μM arabinose to allow for Red expression. Next, the cells were made electro-competent as described in section 2.2.7.1. In the meantime, linear DNA to be transformed was prepared by two PCR steps: first introducing the dif_H site either upstream, downstream, or on both sides of the kanamycin resistance cassette from a plasmid pKD13 (primers 31-34, Table 2-3) and then adding the homologous sequences at the flanks of the cassette (primers 35-38), as shown in Figure 2-7. The plasmid pKD13 is a *pir*-dependent vector and was propagated in *E. coli pir* strain π1 (Table 2-1). The thermocycling conditions of each PCR are shown in Table 2-18. The PCR products were purified by ethanol precipitation: 1/10 of the reaction volume of 3 M sodium acetate and 3 volumes of ethanol (absolute) were added to the PCR reactions and kept at -20 °C overnight. Next, the samples were centrifuged at 17000 *g* for 40 minutes at 4 °C in a benchtop centrifuge and the supernatant was discarded. The pellets were washed once with 1 ml of 70% ethanol. Finally, the pellets were resuspended in 10 μl of distilled water. 5 μl of each purified PCR product was transformed into electro-competent cells carrying pKD46. The cells were then plated on LB with kanamycin, and incubated at 42 °C overnight. At this temperature, pKD46 does not replicate and is lost from the culture.

Kanamycin selection ensures that only the cells where homologous recombination (and therefore *dif* replacement with dif_H) occurred can grow. To validate the genotype, DNA from the *ter* region of *E. coli* was amplified by colony PCR (section 2.2.10) with primers 39 and 40 (Table 2-3) and sequenced. As a control, a strain with a *dif* deletion (DS941 Δ*dif*) was constructed in a similar way using primers 36 and 37 in a single PCR (PCR2, Table 2-18).

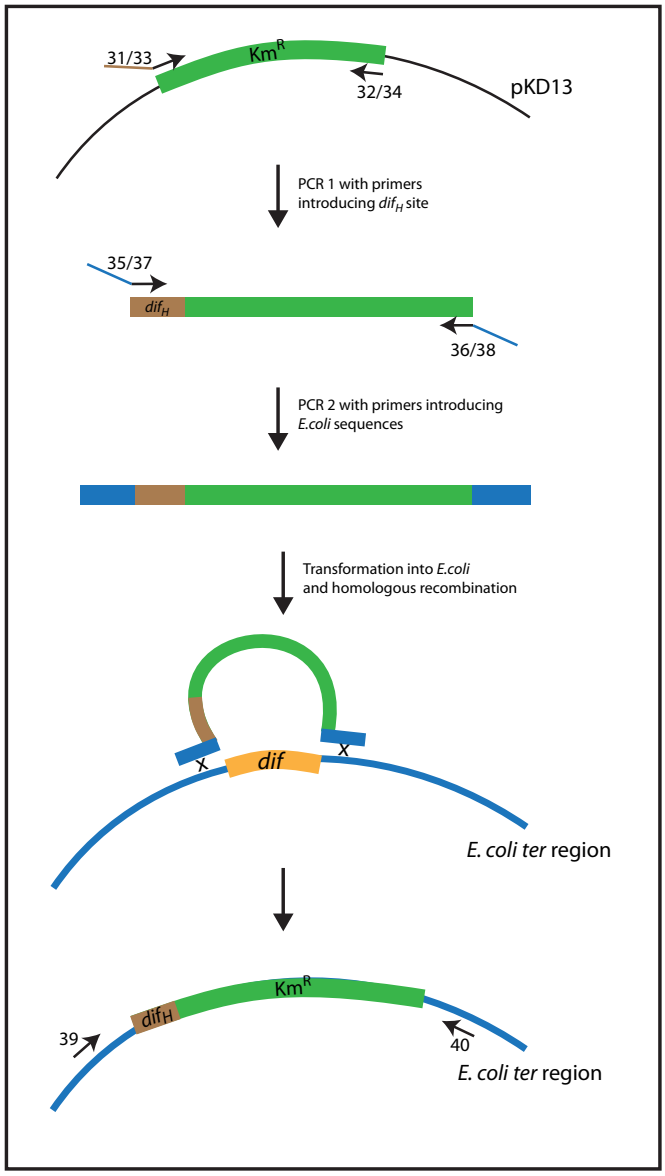


Figure 2-7: Overview of the λ Red recombinase-assisted chromosomal substitutions in *E. coli*. Initially, kanamycin resistance cassette was amplified by PCR introducing 33-bp dif_H in an extended forward and/or reverse primer (here, as an example, only dif_H introduced with the forward primer is shown). Next, the dif_H-Km or dif_H-Km-dif_H cassette was amplified with primers introducing a short (49 or 50 bp) region corresponding to the *E. coli* chromosomal sequences flanking the *dif* site. After transformation into *E. coli*, homologous recombination occurs resulting in incorporation of the dif_H-Km or dif_H-Km-dif_H cassette and deletion of the native *dif* site from *E. coli* genome. The cells carrying the substitution can be identified by plating on LB medium containing kanamycin. The numbers of the primers used (as shown in Table 2-3) are indicated.

Table 2-18: Thermocycling conditions used for preparation of linear DNA for *E. coli* transformation

Step	Temperature	Time	
PCR1			
1. Initial Denaturation	98 °C	60 seconds	
2. Denaturation	98 °C	15 seconds	Steps 2-4 repeated 15 times
3. Annealing	58 °C	15 seconds	
4. Extension	72 °C	60 seconds	
5. Denaturation	98 °C	15 seconds	Steps 5-7 repeated 20 times
6. Annealing	66 °C	15 seconds	
7. Extension	72 °C	60 seconds	
8. Final extension	72 °C	5 minutes	
9. Hold	4 °C	hold	
PCR2			
1. Initial Denaturation	98 °C	60 seconds	
2. Denaturation	98 °C	15 seconds	Steps 2-4 repeated 15 times
3. Annealing	54 °C	15 seconds	
4. Extension	72 °C	60 seconds	
5. Denaturation	98 °C	15 seconds	Steps 5-7 repeated 20 times
6. Annealing	68 °C	15 seconds	
7. Extension	72 °C	60 seconds	
8. Final extension	72 °C	5 minutes	
9. Hold	4 °C	hold	

2.6.5 Assessment of *E. coli* filamentous phenotype

The phenotypes of *E. coli* strains DS941 (wild-type), DS941Δ*dif,* and DS941dif_H were assessed by light microscopy. First, *E. coli* cultures were grown in LB medium with antibiotics overnight from glycerol stocks. Next, 10 ml LB with antibiotics was inoculated with 50 µl of the overnight culture and grown to OD_{600} of 0.5. 1 ml of the culture was then transferred to an eppendorf tube. To fix the cells for imaging, 121 µl of 37% formaldehyde was added for the final concentration of 4%. The cells were incubated for 15 minutes at room temperature on a rocking platform, followed by centrifugation at 10000 *g* in a benchtop centrifuge. Next, the cells were washed three times with 1 ml of 1x PBS buffer and finally resuspended in 200 µl of 1x PBS. The microscopy slides were prepared as follows: 1% agarose gel was poured into a pre-assembled, 1-mm thick Mini-PROTEAN® Handcast System (BioRad). The gel was then removed from between the glass plates and 1.5 cm by 1.5 cm gel squares were cut out. The gel squares were placed on microscopy slides (Thermo scientific) and 3 µl of forlmaldehyde-fixed bacteria were added on top of each square. The gel squares were covered with glass cover slips (Thermo scientific) and imaged immediately with Zeiss CellObserver microscope using phase contrast setting and

100x objective. 30 images were collected for each *E. coli* strain, resulting in at least 200 *E. coli* cells imaged in each case. The images were viewed and analysed using Fiji image processing package (Schindelin *et al.*, 2012).

2.6.6 Chromosomal XerH recombination assay in *E. coli*

Electro-competent DS941dif_H-Km-dif_H and DS9041dif_H-Km-dif_H (prepared as in section 2.2.7.1) were transformed with the XerH expression plasmid (XerH_pBAD/MCS; Table 2-2) as described in section 2.2.8.1, and plated on LB agar containing kanamycin and ampicillin. Next, overnight cultures in LB supplemented with the antibiotics were prepared from single transformant colonies. 10 ml of fresh LB supplemented with 0.2% arabinose was then inoculated with 50 µl of the overnight culture and the cultures were grown at 37 °C until OD_{600} of 1.5. 50 µl of 1:10, 1:100, and 1:1000 dilutions of the cultures were plated on LB-ampicillin plates. 80 of the resulting colonies were restreaked on LB-kanamycin plates to assess the percentage of cells where the dif_H-Km-dif_H cassette was lost. To further confirm the presence or absence of the dif_H-Km-dif_H cassette, colony PCR was performed using primers 39 and 40 (Table 2-3) as described in section 2.2.10.

3. Biochemical characterization of XerH recombinase

3.1 Introduction

3.1.1 Classification of XerH as a tyrosine recombinase

XerH belongs to the large family of tyrosine recombinases. Members of this family can all be characterized by the presence of an approximately 180 amino acids long C-terminal catalytic domain with a very well conserved fold, consisting of nine α helices and a three-stranded β sheet, and an N-terminal domain, which is less conserved (Yang and Mizuuchi, 1997). Within the catalytic domain, five invariant residues are absolutely conserved in the tyrosine recombinase family as part of the recombinase active site: a tyrosine, two arginines, a lysine, and a histidine (Nunes-Duby *et al.,* 1998). Another residue of the catalytic domain required for the active site assembly is invariably a tryptophan or a histidine (Nunes-Duby *et al*., 1998). The presence of the catalytic domain containing the motif RKHR(W/H)Y is therefore a signature of tyrosine recombinases and a criterion for classification of newly discovered proteins to this family of site-specific recombinases.

The alignment of XerH from *H. pylori* strain 26695 with λ integrase and Cre recombinase, the two prototypical members of the tyrosine recombinase family, is shown in Figure 3-1. All the invariant catalytic residues are conserved in XerH: Tyr344 is the catalytic nucleophile; Arg213, Lys239, His309, and Arg312 form the catalytic tetrad required to support the cleavage and ligation reactions by acting as a general acid or base and stabilizing reaction intermediates. His335 of XerH is the more variable catalytic residue, corresponding to a histidine in λ integrase and a tryptophan in Cre. The conservation of all catalytic residues corresponding to the RKHR(W/H)Y motif in the conserved catalytic domain of XerH classifies this protein as a member of the tyrosine recombinase family.

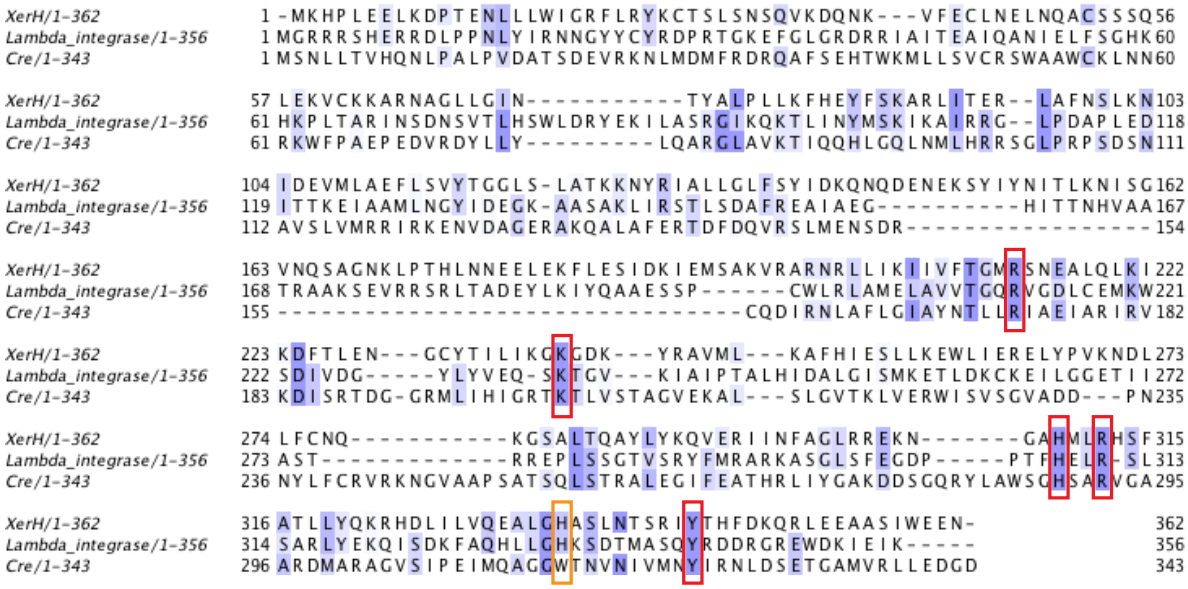


Figure 3-1: Sequence alignment of *H. pylori* XerH, λ integrase, and Cre recombinase. Invariant catalytic residues are marked by red boxes. The less conserved histidine/tryptophan residue is marked by an orange box. The colour of the residues indicates the level of residue conservation: the darker the colour, the more conserved the residue. The alignment was prepared by ClustalW2 (Larkin *et al.,* 2007) using default parameters and visualized using Jalview software (Waterhouse *et al.,* 2007).

3.1.2 Comparison of XerH and other Xer recombinases

Xer recombinases are classified based on their sequence homology, phylogenetic clustering distinct from other tyrosine recombinase types, and their function in resolution of chromosome dimers (Carnoy and Roten, 2009). XerH is phylogenetically related to other Xer recombinases (see Figure 1-8). The alignment of XerH from *H. pylori* strain 26695 with other members of the Xer recombinase family (Figure 3-2) shows sequence identity between XerH and other Xers ranging from 14% to 48%. As for other tyrosine recombinases, the C-terminal catalytic domain of Xers shows much higher sequence conservation than the more variable N-terminal DNA-binding domain, reflecting on the different DNA substrates bound by the proteins. Among investigated Xer proteins (Figure 3-2), XerH is the longest (362 amino acids) and contains an N-terminal extension that is also present in XerH from *C. jejuni*, XerT, and in shorter form in XerS, but is absent from XerC/D and archaeal XerA, suggesting that this extension might constitute a feature of single-recombinase bacterial Xer systems. In addition, *H. pylori* XerH also contains a seven amino acid insertion (residues 93-99) that is a distinct feature not present in any other Xer recombinase.

Figure 3-2: Sequence alignment of *H. pylori* XerH and other Xer recombinases. Invariant catalytic residues are marked by red arrows. The colour of the residues indicates level of residue conservation; the darker the colour, the more conserved the residue. The alignment was prepared by ClustalW2 (Larkin *et al.,* 2007) and visualized using Jalview software (Waterhouse *et al.,* 2007).

3.1.3 The dif_H site

The putative dif_H site was identified in *H. pylori* strain 26695 genome using similarity search tools such as Basic Local Alignment Search Tool (BLAST; Altschul *et al.,* 1990), assuming the following criteria: 1. The site should be somewhat homologous to other known *dif* sites, for example *dif* from *E. coli* or dif_{SL} from *L. lactis*; 2. It should be present on the chromosome near the replication terminus (*ter*); 3. It should be present exactly once per chromosome; and 4. It should be conserved in different strains of the same species (Carnoy and Roten, 2009).

The 40-bp dif_H site thus identified, shown in Figure 3-3, displays features common to all known *dif* sites: it has low G+C content (0.225 compared to an average 0.388 of the *H. pylori* chromosome), it shows significant palindromicity (shown as arrows in Figure 3-3),

and it is located in a non-coding part of the genome (Carnoy and Roten, 2009). Similarly to the XerS/dif_{SL} recombination system, the *xerH* gene and the dif_H site are located close to each other (1981 bp), together forming a distinct genetic module (Le Bourgeois *et al.*, 2007).

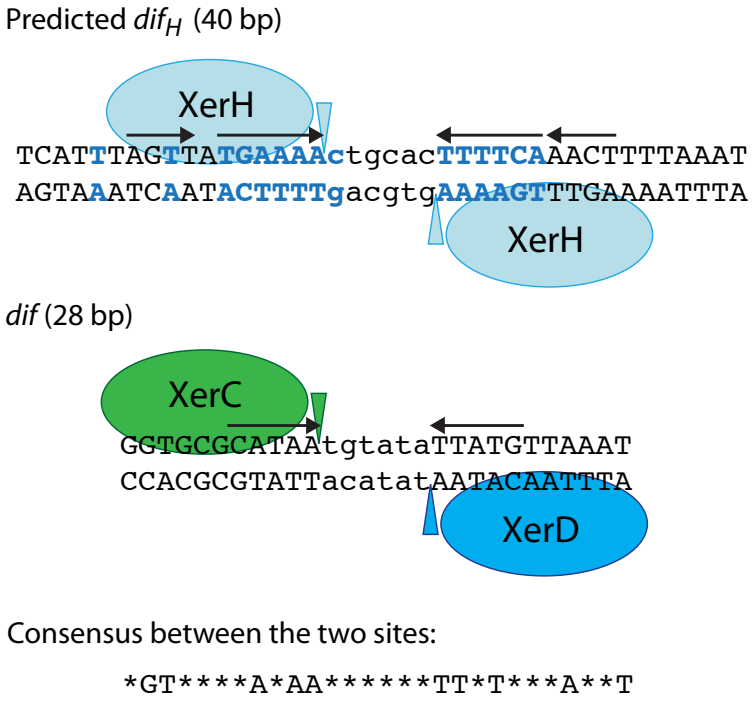


Figure 3-3: The predicted 40-bp *H. pylori* dif_H (top) and its comparison to *dif* from *E. coli* (middle). In both cases, the central region is shown in small letters and the palindromic regions are indicated by arrows. The residues of dif_H invariably found in species related to *H. pylori* are marked in blue. The expected XerH cleavage positions on dif_H are marked by light blue triangles and were predicted based on the known XerC/D cleavage sites on *dif* (green and dark blue triangles), assuming that the cleavage takes place at the boundaries of the palindromic regions and that the cleavage sites flank a 6-bp central region. The consensus sequence between *E. coli dif* and *H. pylori* dif_H is shown below. Stars indicate no sequence conservation between the two sites.

The XerH cleavage sites can be predicted based on a comparison of dif_H and *E. coli dif* (Figure 3-3), assuming that the cleavage takes place at the boundaries of the palindromic regions and that the cleavage sites flank a 6-bp central region. The consensus between the two sites shows that the nucleotides of the palindromic binding arms close to the central region are conserved (Figure 3-3). This observation is in agreement with high sequence conservation in these parts of all studied *dif* sites, and it was proposed that these regions might be crucial for Xer binding and activity (Carnoy and Roten, 2009).

The functionality of the putative 40-bp dif_H site was tested in an *in vivo* assay mimicking chromosome dimer resolution (Debowski *et al.,* 2012). In this assay, two dif_H sites flanking an antibiotic resistance gene cassette were introduced ectopically into the *H. pylori* genome and the excision products were monitored by PCR. The products could be detected only in the presence of the functional *xerH* gene, and no products were detected when a random 40-bp sequence was used instead of the dif_H sites (Debowski *et al.,* 2012).

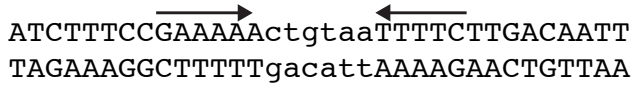
3.1.4 Biochemical studies of single recombinase Xer systems

To date, three single-recombinase Xer systems have been characterized *in vitro*: XerS from *L. lactis* (Nolivos *et al.,* 2010), XerA from *P. abyssi* (Cortez *et al.,* 2010; Serre *et al.,* 2013), and XerH from *C. jejuni* (Leroux *et al.,* 2013).

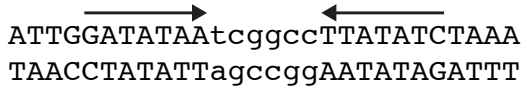
3.1.4.1 XerS

The XerS recombination site, dif_{SL} (Figure 3-4), was identified by comparative genomics studies of *Streptococci* (Le Bourgeois *et al.,* 2007). The site was tested for XerS binding *in vitro*, which revealed that, upon addition of increasing amounts of the protein, distinct higher molecular weight species appear on a native polyacrylamide gel: first, a species corresponding to an XerS monomer binding to the DNA, then a second species corresponding to an XerS dimer binding cooperatively to one dif_{SL} site (Nolivos *et al*., 2010). When oligonucleotides containing only one dif_{SL} half-site were tested for XerS binding, only the substrate containing the left half-site (arm) of dif_{SL} was bound efficiently. For the right arm substrate, only a small amount of XerS monomer-bound dif_{SL} was seen at high protein concentrations. Interestingly, another higher molecular weight complex also appeared in the assay with the right half site of dif_{SL}. However, this product did not correspond in size to the previously observed XerS dimer bound to wild-type dif_{SL}, suggesting that the left arm of dif_{SL} is essential for the assembly of a functional XerS recombination complex (Nolivos *et al*., 2010). Additionally, the ability of XerS to resolve a dif_{SL} HJ substrate was assessed, showing that XerS can catalyse cleavage and strand exchange of both top and bottom strands in the HJ intermediate. Interestingly, a strong asymmetry was also observed here, as the activity on the bottom strand (right arm) was higher than that on the top strand (left arm; Nolivos *et al.,* 2010).

L. lactis dif_{SL} (34 bp)



Archaeal *dif* (28 bp)



C. jejuni dif_H (29 bp)

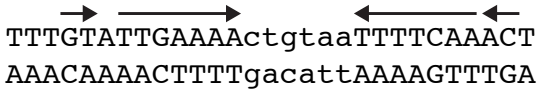


Figure 3-4: *dif* sites of single-recombinase Xer systems from *L. lactis*, the archaeon *P. abyssi*, and *C. jejuni* (Cortez *et al.,* 2010; Leroux *et al.,* 2013; Nolivos *et al.,* 2010). The central regions are shown in small letters. The regions of palindromicity are indicated by arrows.

3.1.4.2 XerA

The search for Xer recombination systems in Archaea revealed a conserved recombinase XerA and a putative chromosome dimer resolution site, *dif* (Cortez *et al.,* 2010). Archaeal *dif* resembles the *dif* site of the XerC/D recombination system more than the *dif* sites of other single-recombinase Xer systems. Binding studies of archaeal *dif* revealed that XerA from *P. abyssi* binds to *dif*, forming sequentially two higher molecular weight complexes (Cortez *et al*., 2010). However, incubation of XerA with a DNA substrate representing the binding site of an unrelated tyrosine recombinase resulted in similar binding. At high temperature (65 °C) that is more physiological for the thermophilic *P. abyssi*, formation of the specific XerA-*dif* complex increased eight-fold compared to the non-specific complex (Cortez *et al.,* 2010). XerA was also active in an *in vitro* cleavage and strand exchange assay with DNA substrates representing *dif* half-sites, which allowed for precise mapping of the recombination sites (Serre *et al*., 2013). Lastly, XerA was able to recombine plasmids *in vitro* (in both resolution and multimerization reactions) without any additional proteins (Cortez *et al.,* 2010), a result that to date could not be accomplished for any other Xer recombinase.

3.1.4.3 XerH from *C. jejuni*

The dif_H site of *C. jejuni* was computationally predicted using similarity search tools (Carnoy and Roten, 2009) and is largely similar to that of *H. pylori*, consistent with the fact that the two XerH proteins are closely related (48.6% sequence identity, Figure 3-2). The binding of *C. jejuni* XerH to its dif_H site was analysed *in vitro* (Leroux *et al.,* 2013). Upon addition of XerH to DNA substrates containing dif_H, two predominant higher molecular weight species appeared, corresponding to an XerH monomer and a dimer bound to dif_H. Despite the similar pattern, binding was much less efficient and less cooperative than that of XerS to dif_{SL} (Nolivos *et al.,* 2010). Additionally, a third, highest molecular weight species was present when a high amount of protein was added. As this band was also present when a non-dif_H control substrate was used, it is most likely a product of non-specific binding (Leroux *et al.,* 2013). Interestingly, binding of XerH to a substrate containing either half-site of dif_H showed similar pattern to that observed with the full-length dif_H, unlike in the XerS/dif_{SL} interaction where the binding of the right half-site substrate was strongly compromised (Leroux *et al*., 2013; Nolivos *et al.,* 2010). This suggests that these two recombinases do not share the same mechanism of DNA binding and might not be functionally homologous (Leroux *et al.,* 2013).

3.1.5 The studies of *H. pylori* XerH to date

Various aspects of *H. pylori* XerH recombination have been investigated in a study by Debowski and colleagues in a set of *in vivo* assays (Debowski *et al*., 2012). The functionality of the dif_H site in XerH recombination was confirmed as discussed in section 3.1.3. It was also shown that certain point mutations in the dif_H DNA sequence (shown in Figure 3-5) abolish XerH recombination *in vivo*. Moreover, the direct role of XerH in dif_H recombination was confirmed by deletions in the *xerH* gene that abolished recombination of dif_H sites. However, it is worth noting that XerH truncation did not cause the filamentous phenotype associated with XerC/D deletions in *E. coli* (Blakely *et al.,* 1993), but showed slower growth, increased DNA content, and increased UV sensitivity compared to the wild-type strain (Debowski *et al.,* 2012). Interestingly, the second Xer recombinase found in some *H. pylori* strains, the TnPZ-associated XerT, is not required for recombination at dif_H. Since typically Xer recombination is regulated by the host factor FtsK, the role of the *H. pylori* FtsK homologue in XerH recombination was also addressed.

A strain that expressed a truncated version of FtsK with disrupted C-terminal helicase domain did not support dif_H recombination. Together, these results indicate that XerH acts as a chromosome dimer recombinase in *H. pylori* by recombining two dif_H sites in an FtsK-dependent fashion.

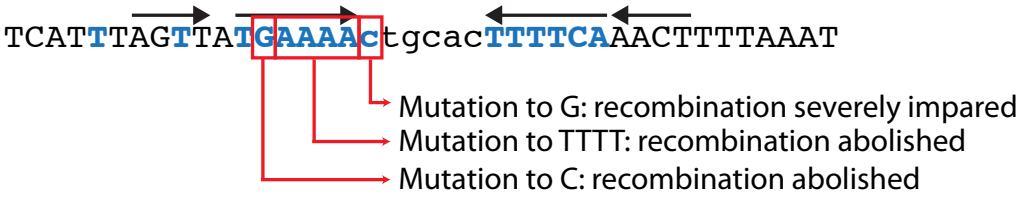


Figure 3-5: Summary of the mutational analysis of dif_H (Debowski *et al*., 2012). The residues of dif_H invariably found in species related to *H. pylori* are marked in blue. The regions of palindromicity are indicated by arrows. The effects of dif_H mutations at positions indicated by the red boxes apply to *in vivo* recombination of two chromosomal dif_H sites assessed by PCR (Debowski *et al*., 2012).

3.1.6 Aims and objectives

Despite several studies on various single Xer recombination systems, the mechanism and regulation of these recombinases remains unclear. For XerH/dif_H in particular, the study by Debowski and colleagues elucidated the role of this recombination system in *H. pylori* (Debowski *et al*., 2012). However, mechanistic questions regarding XerH recombination remained unanswered. The exact requirements of a functional dif_H site in terms of length and sequence were only partially tested; for example, the dif_H substrates used in the study contained a 40-bp predicted dif_H sequence, but the exact length of the dif_H site necessary for XerH recombination is not clear. Additionally, the direct interaction between XerH and dif_H was never assessed, and it was possible that XerH is not the only or not the directly interacting component of the recombination machinery. In this biochemical characterization of XerH, I attempt to address the above questions, keeping in mind the surprising, inconsistent findings in the studies of other single-recombinase Xer/*dif* systems, suggesting that various mechanisms of Xer recombination at *dif* sites are possible (Cortez *et al.,* 2010; Leroux *et al.,* 2013; Nolivos *et al.,* 2010).

3.2 Results

3.2.1 Construct design and cloning of XerH

In order to perform biochemical analysis of XerH/dif_H recombination *in vitro*, purified *H. pylori* XerH protein was required. Due to the pathogenic nature and difficult culturing conditions of *H. pylori*, the pET expression system was employed to produce recombinant XerH in *E. coli* culture. The *xerH* gene (HP0675) from *H. pylori* reference strain 26695 was chosen for biochemical and structural studies as the XerH from this strain has been previously characterized *in vivo* (Debowski *et al*., 2012). Since the codon usage of *E. coli* and *H. pylori* differs (Nakamura *et al.,* 2000), the XerH gene was codon-optimized for expression in *E. coli* (Figure 3-6a).

a

```
ori ATGAAACACCCCCTAGAAGAATTGAAAGACCCTACAGAAAATCTTTTGCTATGGATTGGG CGCTTTTTGCGTTACAAATGCACGAGCCTGTCTAATTCCCAAGTGAAAGACCAAAATAAG 120
opt ATGAAACATCCGCTGGAAGAACTGAAAGATCCGACCGAAAATCTGCTGCTGTGGATTGGT CGTTTTCTGCGTTATAAATGTACCAGCCTGAGCAATAGCCAGGTGAAAGATCAGAATAAA 120
    ******** ** ** ****** ******* ** ** ********  **** ********  ** *** ******* ***** ** ******   ***  *** ******** ** *****

ori GTTTTTGAATGCTTGAACGAACTAAATCAAGCGTGCAGCTCAAGCCAATTAGAAAAAGTT TGTAAAAAAGCGCGTAATGCCGGATTGCTTGGGATCAACACCTACGCATTGCCCCTACTC 240
opt GTGTTTGAATGCCTGAACGAACTGAATCAGGCATGTAGCAGCAGCCAGCTGGAAAAAGTT TGTAAAAAAGCACGTAATGCAGGTCTGCTGGGCATTAATACCTATGCACTGCCGCTGCTG 240
    ** ********* ********** ***** ** ** ***   *****  * ********* ********** ******** **  **** ** ** ** ***** *** **** ** **

ori AAATTTCATGAATACTTCAGCAAAGCTAGACTAATAACTGAAAGGCTTGCTTTTAATTCG CTCAAAAACATTGATGAAGTCATGCTAGCTGAATTTTTGAGCGTTTATACCGGGGGTTTG 360
opt AAATTCCATGAGTATTTTAGCAAAGCACGCCTGATTACCGAACGTCTGGCATTTAATAGC CTGAAAAACATCGATGAAGTTATGCTGGCAGAATTCCTGAGCGTTTATACCGGTGGTCTG 360
    ***** ***** ** ** ********  * ** ** ** ** *** * ** ** ******    ** ******** ******** ***** ** *****  ********** *** **

ori AGTTTAGCCACTAAGAAAAATTACAGGATCGCCCTACTCGGGCTTTTTAGCTATATAGAC AAGCAAAATCAAGATGAAAATGAAAAATCTTATATTTATAATATCACGCTTAAAAACATC 480
opt AGCCTGGCAACCAAAAAAAACTATCGTATTGCACTGCTGGGCCTGTTTAGCTATATTGAT AAACAGAACCAGGATGAAAACGAAAAAAGCTACATCTACAACATCACCCTGAAAAATATC 480
    **  * ** ** ** ***** **  * ** ** ** ** ** ** ** ********** **  ** ** ** ** ******** ******   ** ** ** ** ***** ** ***** ***

ori AGTGGAGTCAATCAAAGCGCAGGCAATAAGCTCCCTACTCATTTAAATAATGAAGAATTA GAAAAATTTTTAGAGAGCATTGATAAAATAGAAATGTCCGCTAAAGTGCGTGCACGAAAC 600
opt AGCGGTGTTAATCAGAGCGCAGGTAATAAACTGCCGACCCATCTGAATAATGAGGAACTG GAAAAATTCCTGGAAAGCATCGATAAAATTGAAATGAGCGCAAAAGTTCGTGCACGTAAT 600
    ** ** ** ***** ******** ***** ** ** ** ** *** * ******** *** *  ********  * ** ***** ******** ******  *** ***** ******** **

ori CGCTTGCTCATTAAAATCATCGTTTTTACAGGCATGCGATCTAATGAAGCCTTGCAGCTT AAAATAAAGGATTTTACTTTAGAAAATGGCTGTTATACGATTTTGATTAAAGGTAAGGGC 720
opt CGTCTGCTGATTAAAATCATTGTGTTTACCGGTATGCGTAGCAATGAAGCACTGCAGCTG AAAATCAAAGATTTCACACTGGAAAATGGCTGCTATACCATTCTGATTAAAGGCAAAGGT 720
    **  **** *********** ** ***** ** *****    ********  *******  ***** ** ***** **  * *********** ***** *** ********** ** **

ori GATAAATACAGAGCGGTGATGCTCAAAGCTTTCCACATTGAGAGCCTTTTGAAAGAATGG CTGATAGAAAGGGAATTGTATCCTGTTAAAAACGATTTATTGTTTTGCAACCAAAAAGGC 840
opt GATAAATATCGTGCCGTTATGCTGAAAGCCTTTCATATTGAAAGCCTGCTGAAAGAATGG CTGATTGAACGTGAACTGTATCCGGTTAAAAACGATCTGCTGTTTTGCAATCAGAAAGGT 840
    ********  * ** ** ***** ***** ** ** ***** *****  *********** ***** *** * *** ******* *********** *  ********** ** *****

ori AGTGCTTTAACGCAAGCTTATTTGTATAAGCAAGTGGAGCGCATCATCAATTTTGCAGGA CTCAGGCGAGAGAAAAATGGGGCGCACATGTTAAGGCATTCTTTTGCGACCTTGCTTTAT 960
opt AGCGCACTGACCCAGGCATATCTGTATAAACAGGTTGAACGCATCATTAACTTTGCCGGT CTGCGTCGTGAAAAAAATGGTGCACATATGCTGCGTCATAGCTTTGCAACCCTGCTGTAT 960
    ** **  * ** ** ** *** ******* ** ** ** ******** ** ***** **  **  * ** ** ******** ** ** *** *  * ***   ***** *** **** ***

ori CAAAAACGCCATGATTTGATTTTAGTTCAAGAAGCTCTAGGGCATGCGAGCTTGAATACC AGTAGGATTTACACGCATTTTGACAAGCAGCGTTTAGAAGAAGCGGCGAGCATTTGGGAA 1080
opt CAGAAACGTCATGATCTGATTCTGGTTCAAGAAGCACTGGGTCACGCAAGCCTGAATACC AGCCGTATTTATACCCATTTCGATAAACAGCGTCTGGAAGAAGCAGCAAGCATTTGGGAA 1080
    ** ***** ****** ***** * ********** ** ** ** ** *** ******** **  * ***** ** ***** ** ** ****** * ******** ** ***********

ori GAAAATTAA 1089
opt GAAAACTAA 1089
    ***** ***
```

b

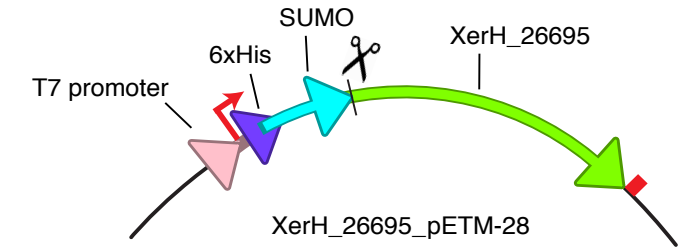


Figure 3-6: XerH construct for protein overexpression in *E. coli*. a) Comparison of *H. pylori xerH* gene (ori) and its derivative codon-optimized for overexpression in *E. coli* (opt). Unchanged nucleotides are marked with stars. b) Schematic representation of the XerH *E. coli* expression vector XerH_26695_pETM-28. The start and end of the open reading frame are marked in red. The scissors mark the encoded SenP2 protease cleavage site.

This codon-optimized gene was then inserted with restriction enzyme cloning into the pETM-28 vector (Figure 3-6b). This vector contains a 6xHis affinity tag and a SUMO solubility tag, both upstream of the multiple cloning site, resulting in the inserted gene being expressed as an N-terminal 6xHis-SUMO fusion. The 6xHis affinity tag is used for purification by nickel-affinity chromatography, and the SUMO tag improves solubility of the construct (Butt *et al*., 2005). The fusion gene is expressed under control of the T7 promoter, which can be induced with addition of IPTG or repressed with addition of glucose. Overexpression under the T7 promoter results in the protein of interest being the most abundant protein in *E. coli* upon induction of expression (Studier and Moffatt, 1986).

3.2.2 Overexpression of XerH in *E. coli*

In order to identify the best conditions for XerH overexpression in *E. coli*, small-scale expressions were carried out in three different *E. coli* strains: BL21 (DE3), BL21 (DE3) pLysS, and Rosetta. The cells carrying the XerH expression plasmid were grown at 37 °C until OD_{600} of 1 then cooled down to 15 °C and induced with IPTG. Samples for analysis by SDS-PAGE were collected immediately before induction and 3 and 15 hours later. SDS-PAGE analysis of the samples (Figure 3-7a) showed lower expression in BL21 (DE3) pLysS and comparably higher expression in both BL21 (DE3) and Rosetta strains. Strain BL21 (DE3) was chosen for further testing in small-scale expression: the expression experiment was repeated with an initial growth temperature of 42 °C (a temperature at which heat shock proteins are expressed, many of which act as chaperones to promote correct protein folding; De Maio, 1999) and of 37 °C (corresponding to the optimal *E. coli* growth temperature) for comparison. The samples for SDS-PAGE analysis were collected before induction with IPTG and after 5 and 15 hours of expression at 15 °C. Analysis of the samples (Figure 3-7b) showed that initial growth at 42 °C did not improve the expression of XerH. Thus, large-scale expressions for protein purification were performed in strain BL21 (DE3) with initial growth at 37 °C, with induction and expression at 15 °C for 15 hours, resulting in a high level of XerH expression (Figure 3-7c). In order to assess the solubility of expressed XerH, the cultures were harvested and the cells were lysed by sonication. The lysate was then ultra-centrifuged to separate the soluble and insoluble fractions. The samples before and after centrifugation were analysed by SDS-PAGE (Figure 3-7c), which showed that XerH is present in large amounts in the soluble fraction of the lysate. Therefore, the overexpression of XerH as a fusion with N-terminal 6xHis-

SUMO tag in the chosen conditions produced large amounts of soluble protein suitable for further purification.

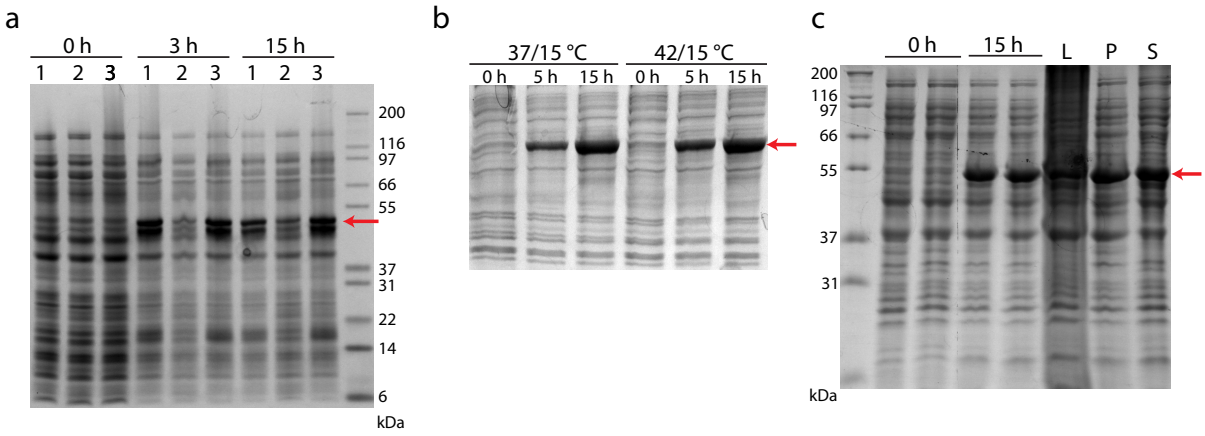


Figure 3-7: Overexpression of XerH in *E. coli* as analysed by SDS-PAGE. The positions of the expressed XerH protein on the gels are indicated by red arrows. a) Small scale test expression of XerH in three different *E. coli* strains: 1- BL21 (DE3); 2- BL21 (DE3) pLysS; 3- Rosetta. The samples were analysed immediately before induction (0 h), then 3 hours (3 h) and 15 hours (15 h) later. b) Small-scale test expression of XerH in *E. coli* BL21 (DE3) with initial growth at 37 °C or 42 °C, analysed before induction (0 h), then 5 hours (5 h) and 15 hours (15 h) later. c) Large scale expression of XerH in *E. coli* BL21 (DE3). The samples were analysed before induction (0 h), 15 hours later (15 h), after culture harvest and lysis by sonication (L), and after ultra-centrifugation that separated the soluble lysate fraction (supernatant – S) from the insoluble lysate fraction (pellet – P).

3.2.3 Computational analysis of XerH protein

Computational analysis of XerH protein performed using the ProtParam web-based tool allowed for calculation of such parameters as molecular weight, theoretical pI value, and UV extinction coefficients of both the 6xHis-SUMO-XerH fusion and the native XerH protein (Table 3-1). These parameters were used to help design purification buffers and purification strategies. For example, the pI value indicates the pH at which the protein of interest carries no electrical charge, and is the least soluble. Therefore, buffer solutions with a pH at least one unit away from the theoretical pI value are generally chosen for purification. Accordingly, for the two proteins 6xHis-SUMO-XerH and XerH, for which the pI values are 9.1 and 9.24, respectively, purification buffers at pH 7.5 were chosen. The high pI indicated that the proteins contain many basic (positively charged) residues, consistent with the proposed DNA binding function of XerH.

Accordingly, phosphate buffers were used for XerH purification to exploit the DNA-mimicking properties of the phosphate groups, which could improve XerH stability and

solubility. The molecular weight and extinction coefficient values were used to estimate the protein concentrations during and after purification.

Table 3-1: Properties of XerH and 6xHis-SUMO-XerH as determined by ProtParam.

Feature	XerH	6xHis-SUMO-XerH
Number of amino acids	362	463
Molecular weight (Da)	41851.4	53266.2
Theoretical pI	9.24	9.1
Molar extinction coefficient ($M^{-1}cm^{-1}$)	39000	40500

Additionally, the secondary structure of XerH was computationally predicted using PSIPRED Protein Sequence Analysis Workbench (Figure 3-8; Buchan *et al*., 2013). This analysis is important for designing protein constructs suitable for successful purification and crystallization: generally proteins with many or long unstructured regions tend to aggregate and are considered less likely to crystallize due to their conformational heterogeneity. Therefore, it is a common practice to remove the unstructured regions from the ends or inter-domain regions of proteins used in purification and crystallization studies. The secondary structure prediction for XerH revealed that the C-terminus of the protein ended with a predicted helix, while the N-terminus contained a short stretch of unstructured amino acids (Figure 3-8). Additionally, the linker between the two XerH domains (marked in blue in Figure 3-8) was predicted to be unstructured but the two domains showed well-defined secondary structure. Together, the lack of long unstructured stretches of amino acids at the ends of the protein suggested that the full-length protein was suitable as the initial target for purification and further studies. In case this full-length construct was unstable or non-crystallisable, removing of the N-terminal unstructured region or shortening the inter-domain linker could be considered as construct improvement strategies.

Finally, to visualize the predicted tertiary structure fold of XerH, a 3-dimensional homology model of XerH was generated using the Phyre2 server (Figure 3-9; Kelley and Sternberg, 2009). Phyre2 predicts and prepares a protein model based on multiple structures of sequence-related proteins available in the PDB database. The homologous protein models found by Phyre2 included mostly available structures from other tyrosine recombinases (discussed in detail in section 4.1.1). The resulting 3D model of XerH shows two distinct domains connected by a linker in a conformation that would allow DNA

binding. The catalytic residues face the predicted DNA-binding cleft, but are not forming a compact active site. Additionally, the protruding C-terminal segment characteristic for tyrosine recombinases is present in the model of XerH and includes a single α helix. The very N-terminus of the DNA binding domain of XerH, for which there is no counterpart in other tyrosine recombinases, in the model protrudes out of the body of the domain, and is likely positioned differently in the true structure of the recombinase.

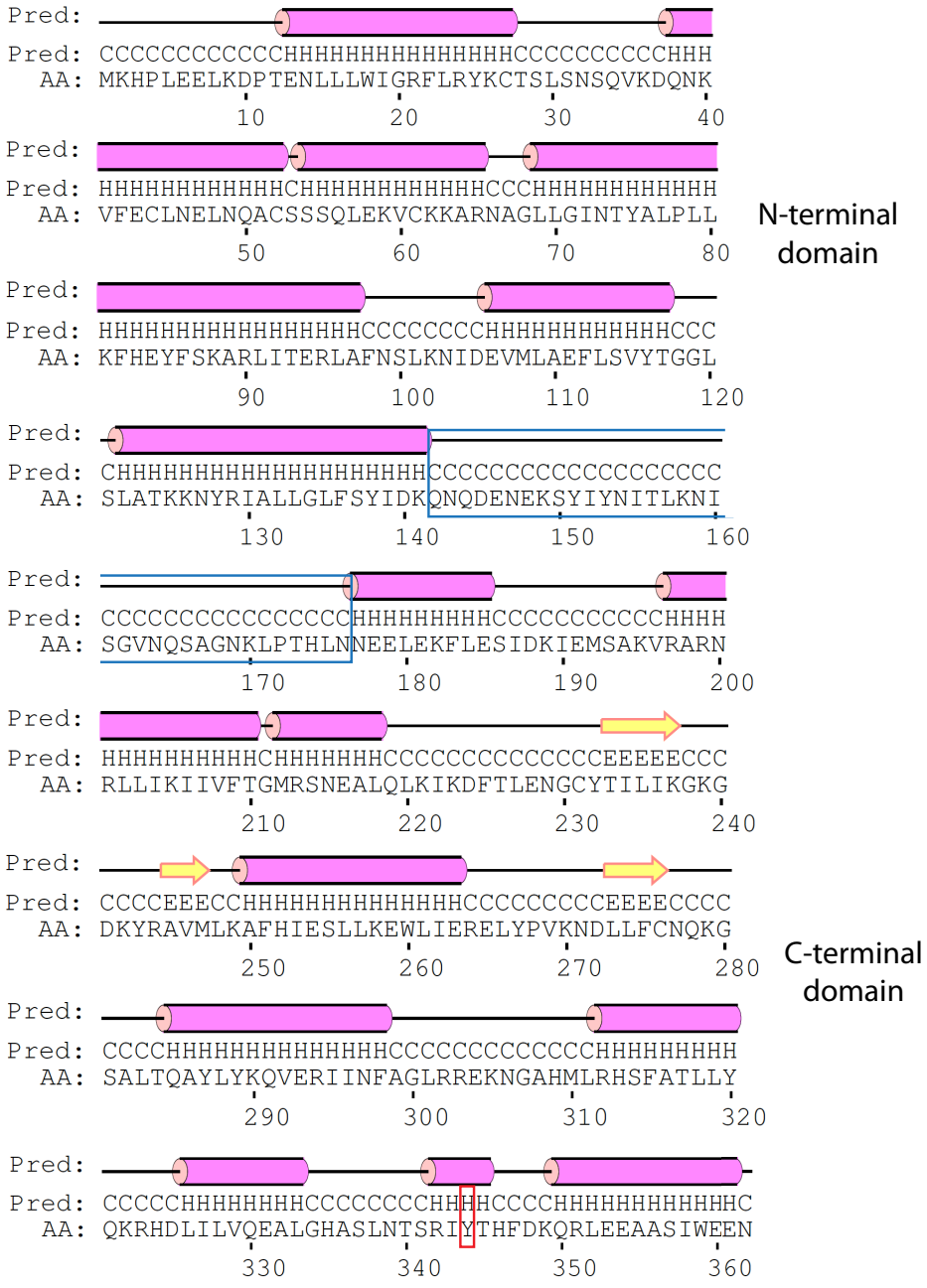


Figure 3-8: Secondary structure prediction for *H. pylori* XerH prepared using PSIPRED Protein Sequence Analysis Workbench (Buchan *et al*., 2013). The amino acid sequence (AA) is accompanied by the secondary structure prediction for each residue (Pred) expressed as a letter (H: α-helix, E: β-strand, and C: coil, loop, or unstructured region) and as a graphical representation (magenta cylinder: α-helix, yellow arrow: β-strand). The catalytic tyrosine residue is marked with a red box. The unstructured linker between the N-terminal and the C-terminal domains is marked with a blue box.

Figure 3-9: Cartoon representation of the XerH structure model prepared using the Phyre2 server (Kelley and Sternberg, 2009). Predicted α-helices are shown as ribbon spirals. The catalytic residues of XerH (Tyr344, Arg213, Lys239, His309, Arg312 and His335) are shown as a stick model and marked in red to highlight their plausible position facing the cleft where DNA could be bound. The N- and C-termini of the protein are marked.

3.2.4 Purification of XerH

XerH was purified to homogeneity by combined affinity and size exclusion chromatography. In detail, the overexpressed 6xHis-SUMO-XerH fusion protein was extracted from the *E. coli* lysate by application to a Ni^{2+}-Sepharose HisTrap column on which only proteins with affinity for nickel ions are retained. Poly-histidine tags such as 6xHis bind strongly to the Ni^{2+}-containing beads, while native proteins show no or little specific binding affinity, allowing for separation of the His-tagged protein from other lysate components by step-wise elution with an imidazole gradient (Hengen, 1995). The elution fractions containing 6xHis-SUMO-XerH were combined and subjected to proteolysis by 6xHis-tagged SenP2, a protease that specifically cleaves the protein main chain at the C-terminus of the 6xHis-SUMO tag. Consequently, the XerH protein was

detached from the 6xHis-SUMO tag, leaving a single amino acid (serine) post-cleavage scar at the N-terminus of XerH. In order to separate XerH, the 6xHis-SUMO tag, and the 6xHis-SenP2 protease, the sample was applied again onto the HisTrap column, to which the untagged XerH binds with weak affinity while the 6xHis-SUMO tag and the 6xHis-SenP2 bind strongly. Step-wise elution allowed for separation of the XerH from other components. XerH-containing fractions were then loaded onto a gel filtration column, which separates the sample components by size (Figure 3-10a). The resulting elution peak corresponded in size to monomeric XerH. Collected peak fractions were analysed by SDS-PAGE (Figure 3-10b) and fractions 4-7 were combined for further studies. From one litre of *E. coli* expression culture, this purification protocol yielded approximately 4 mg of purified XerH of purity (>95%) appropriate for further biochemical and crystallographic studies.

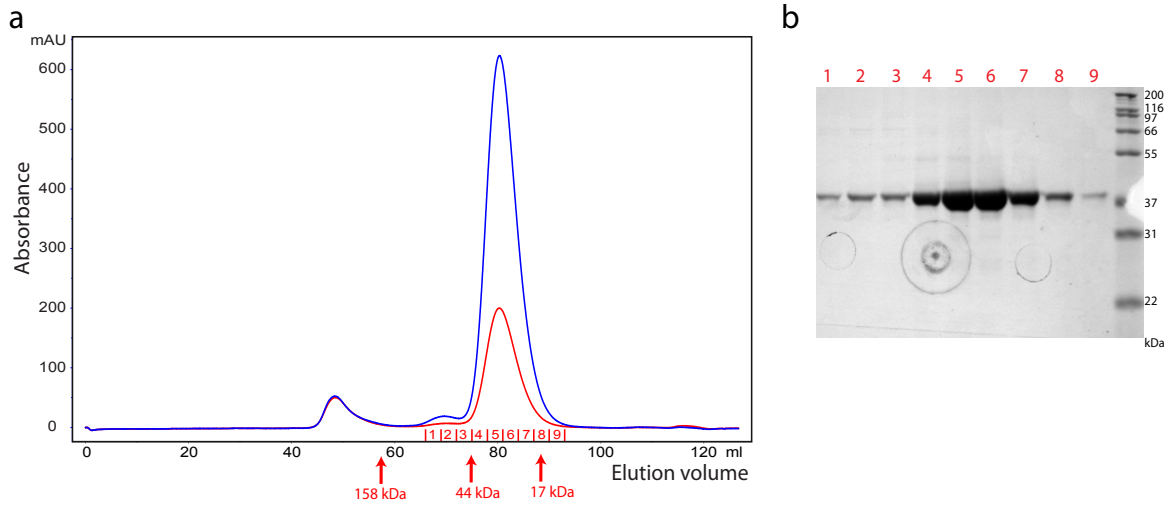


Figure 3-10: Purification of XerH. a) The final step of XerH purification: SEC on a Superdex 16/60 gel filtration column. The chromatogram shows the main elution peak that corresponds in size to a monomer of XerH (42 kDa). UV absorbance at 280 nm is shown in blue, absorbance at 260 nm is shown in red. The fractions collected are shown in red under the elution peak. The elution volumes of molecular weight markers of 17, 44, and 158 kDa on this column are indicated by red arrows. b) SDS-PAGE analysis of the fractions collected from the gel filtration, stained for the presence of protein.

3.2.5 Identification of the XerH-protected DNA recombination site

The 40-bp dif_H site has been shown to be sufficient for XerH recombination (Debowski *et al.*, 2012). However, considering the size of other known *dif* sites (ranging between 28 and 34 bp) it is likely that the minimal dif_H site required for XerH recombination is smaller than 40 bp. To test this, DNaseI footprinting experiment was performed. In this

experiment, a longer (200 bp) oligonucleotide with a centrally located dif_H site was 5'-end labelled with [γ-^{32}P]-ATP on the top strand and incubated with XerH. Next, the XerH-bound DNA was subjected to DNaseI digestion and the samples were analysed on a denaturing sequencing gel alongside a sequencing ladder of the substrate. The results show a region of DNA that has not been digested by DNaseI, corresponding to the site where XerH was bound (Figure 3-11a). This region was digested in the absence of XerH, suggesting that the observed protection pattern is directly caused by XerH binding to the substrate DNA. The sequence of the bound region could be read from the sequencing ladder run alongside the reactions and was found to correspond to a 31-bp region of dif_H and consisting of 14-bp left XerH binding arm, 11-bp right XerH binding arm, and 6-bp central region between the two arms (Figure 3-11b). It is important to note that the DNaseI footprinting assay does not determine the exact protein binding site, but rather the region protected from DNaseI cleavage by the bulk of the protein bound to the DNA, therefore the identified region can be larger or smaller than the actual bound site. Nevertheless, the assay allowed for identification of 31-bp binding site, which is in the range of the sizes of known dif_H sites and likely corresponds to or only differs slightly from the actual minimal dif_H site required for XerH recombination.

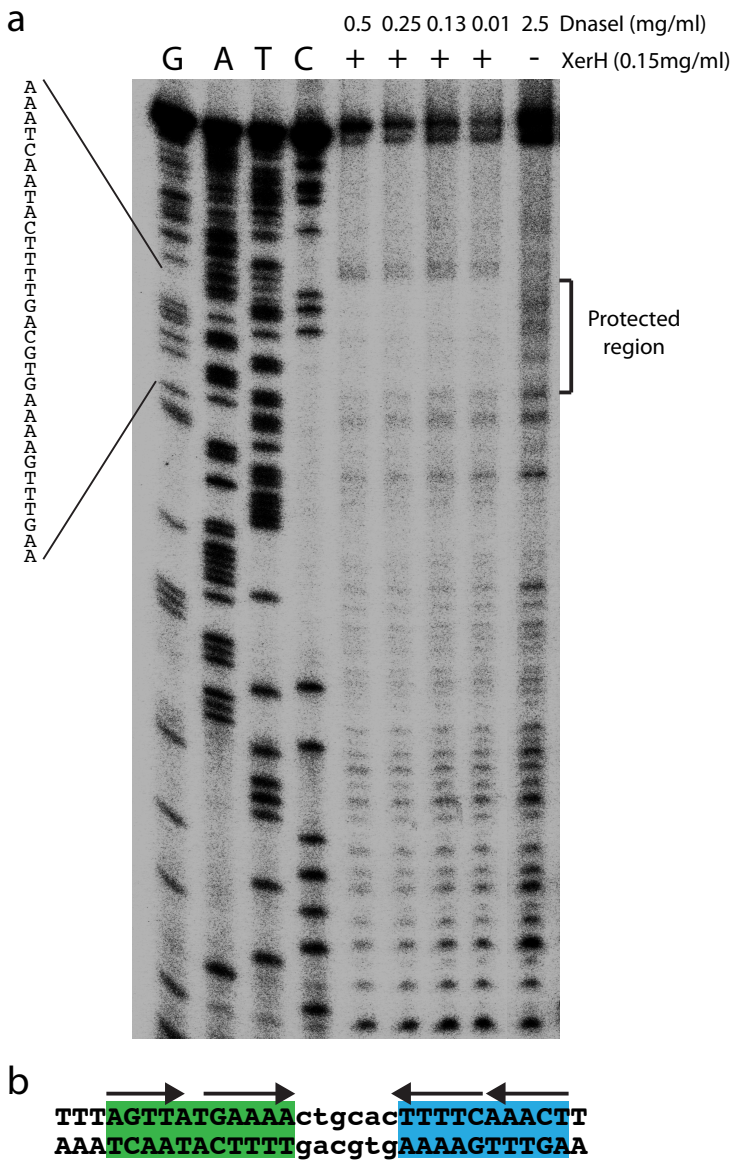


Figure 3-11: DNaseI footprinting assay to map the XerH-protected region of *dif*$_H$. a) *dif*$_H$-containing DNA substrate was incubated with XerH and then digested by the indicated amounts of DNaseI to reveal an XerH-protected and therefore undigested region of the substrate. As a control, a reaction without XerH was set up and run alongside the XerH-containing samples. The sequence of the protected region can be directly identified by comparison to the sequencing ladder run alongside the samples. b) *dif*$_H$ site identified in the DNaseI footprinting experiment. The regions of palindromicity are marked by arrows and coloured boxes (green for the left *dif*$_H$ arm and blue for the right arm).

3.2.6 Binding of XerH to 31-bp *dif*$_H$

In order to confirm that the 31-bp *dif*$_H$ site identified in the DNaseI footprinting assay is a functional XerH recombination site, binding of XerH to the site was tested in an EMSA experiment. A constant amount of 31-bp oligonucleotide containing the identified *dif*$_H$ sequence was incubated with increasing amounts of XerH and the complexes formed were observed on a non-denaturing TBE polyacrylamide gel by staining for the DNA (Figure 3-12). Due to their increased size compared to DNA only, the XerH-DNA complexes are

expected to exhibit lower mobility on a gel and be seen at a higher position. In this experiment, two such larger complexes appeared sequentially: at lower protein concentrations only one complex (lower band) was formed while upon addition of more XerH, the free DNA band disappeared and the second complex (higher band) appeared (Figure 3-12). At the highest concentration of XerH, the higher band was the more prominent of the two complexes. The complex bands observed are likely to correspond to the complex of DNA with one (lower band) or two (higher band) XerH monomers. The binding is very efficient and can be observed even at a 3.3-fold molar excess of DNA over protein.

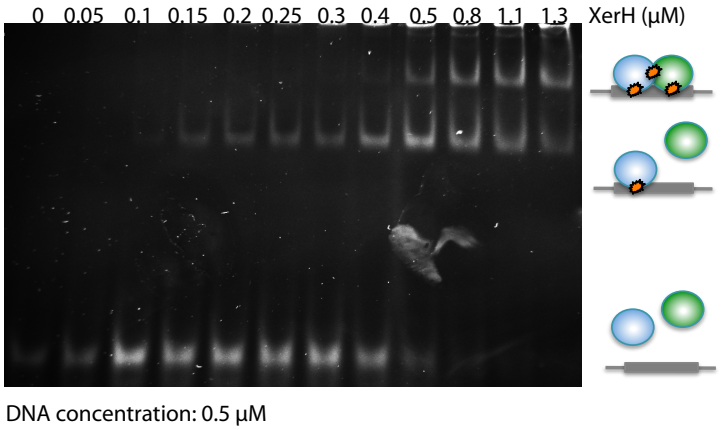


Figure 3-12: Analysis of XerH-dif_H complex formation by EMSA. The bands observed correspond to free DNA, DNA bound by one XerH monomer, or DNA bound by two XerH monomers as indicated by the schematics on the right hand side of the gel.

3.2.7 Analysis of XerH binding to dif_H by analytical SEC

Binding of XerH to the 31-bp dif_H was also analysed by SEC on an analytical gel filtration column. In this assay, the samples of protein only, DNA only, and protein incubated with DNA are run separately on the analytical gel filtration column and the chromatograms from all the runs are overlaid and compared. If a complex between protein and the DNA is formed, the size of the complex will be larger than that of either protein alone or DNA alone and the complex will elute earlier from the column. On the other hand, if a complex is not formed, the components of the sample will elute at volumes corresponding to protein alone and DNA alone. Apart from confirming the presence of stable protein-DNA complexes, this method allows for accurate visualization of DNA or protein excess in the sample. Accordingly, this method was used to optimize the molecular ratios between

complex components to obtain a homogenous sample without DNA or protein excess. Additionally, this method allows for analysis of highly concentrated complexes, which are normally required for crystallization and some biochemical assays.

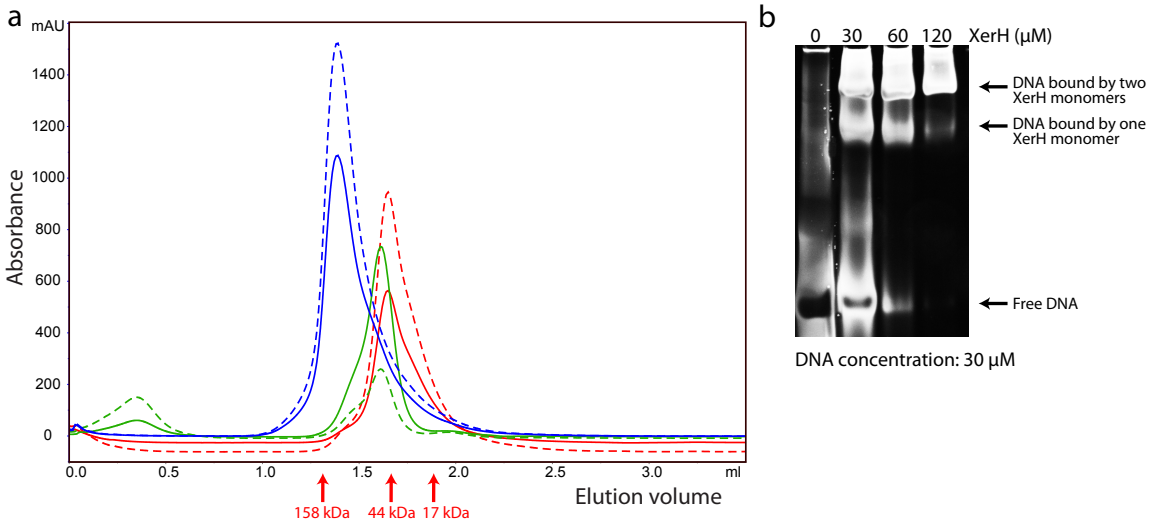


Figure 3-13: Analysis of XerH-dif_H complex formation by analytical SEC on a Superdex 3.2/30 gel filtration column. a) The chromatogram of analytical gel filtration of XerH alone (green), 31-bp dif_H DNA alone (red), and the complex of XerH and 31-bp dif_H DNA (blue). UV absorbance at 280 nm is shown as continuous line, absorbance at 260 nm is shown as a dashed line. The elution volumes of molecular weight markers of 17, 44, and 158 kDa on this column are indicated by red arrows. b) Analysis of XerH-dif_H complexes with different protein:DNA ratios on a native TBE polyacrylamide gel.

By comparison to DNA alone and protein alone controls, it could be clearly seen that the complex between XerH and the 31-bp dif_H site has formed (Figure 3-13a). The estimated complex size based on the elution volume of the complex peak corresponds to a dimer of XerH bound to a single dif_H site (approximately 100 kDa complex). To form the complex shown in Figure 3-13a, an estimated 2-fold molar excess of XerH over DNA was used. The complex peak was not fully homogenous and showed a lower molecular weight shoulder that could correspond to one XerH monomer bound to dif_H. Consistently, analysis of the complexes on a native TBE polyacrylamide gel (Figure 3-13b) revealed that XerH excess higher than 2-fold is required to form a homogenous dimeric complex. This could be caused by some protein precipitation during complex formation resulting in overestimation of protein concentration.

3.2.8 DNA cleavage activity of XerH on suicide substrates

To test the functionality of both the purified XerH and the 31-bp dif_H site, activity was tested in an *in vitro* cleavage assay. Due to the transient nature of the intermediate cleavage product observed previously for other tyrosine recombinases, oligonucleotides that trap the covalent protein-DNA intermediate were designed for this assay (Figure 3-14a). These so-called suicide substrates contain a nick on top, bottom, or both strands of dif_H one nucleotide away from the expected XerH cleavage site. Upon cleavage, a single nucleotide is separated from the DNA strand and diffuses away from the complex. As a result, no free 5' DNA hydroxyl is available for continuing or reversing the reaction and the complex is trapped in a post-cleavage stage. Since the trapped intermediate includes a covalent link between the protein and the DNA, it migrates slower on SDS-PAGE than the protein alone. The presence of the additional higher molecular weight band and its amount depend on the efficiency of XerH-mediated DNA cleavage.

To test cleavage activity, XerH was incubated with a range of suicide substrates (Figure 3-14b) that trapped the cleavage of the top, bottom, or both dif_H strands. The results show that the purified XerH is proficient in cleavage of both dif_H arms (Figure 3-14c) and that the 31-bp dif_H is sufficient to support this cleavage activity (Figure 3-14d). Furthermore, these results show that the cleavage is more efficient on the bottom strand (cleavage performed by the monomer bound to the right dif_H arm) than that of the top strand (cleavage performed by the monomer bound to the left dif_H arm; lanes 1 and 2, Figure 3-14c). Interestingly, in the presence of nicks on both top and bottom strand, the top strand is cleaved with higher efficiency (lane 3, Figure 3-14c).

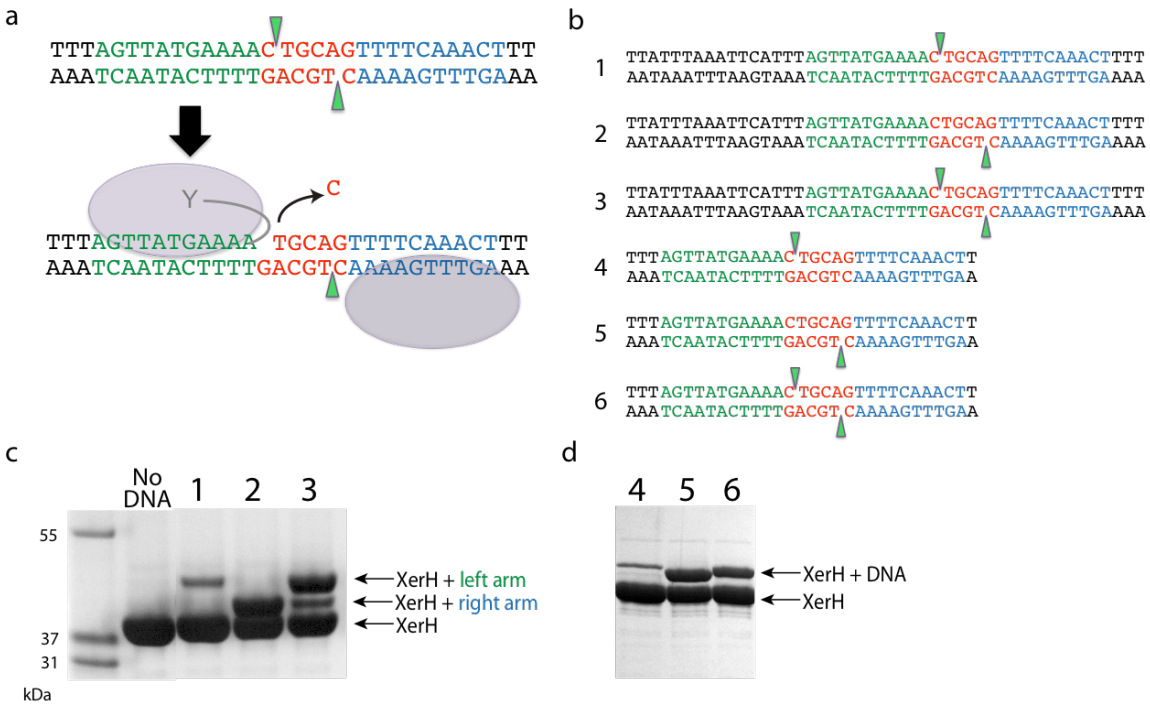


Figure 3-14: XerH cleavage activity *in vitro*. a) The concept of a DNA suicide substrate designed to assess XerH-mediated cleavage by trapping the phosphotyrosyl intermediate. The substrate contains nicks (green triangles) one nucleotide away from the expected XerH cleavage site. Upon cleavage, the released nucleotide diffuses away and the intermediate is trapped as there is no source of 5' hydroxyl available to allow progression of the reaction. b) List of suicide substrates used to assess cleavage activity of purified XerH, containing nicks in top, bottom, or both strands of dif_H (marked with green triangles) designed as described in (a). The left and right arms of dif_H are coloured green and blue, respectively. The central region of dif_H is coloured red. c) SDS-PAGE analysis of cleavage reactions with XerH and dif_H suicide substrates that contained a longer left dif_H arm to distinguish between left and right arm cleavage products. The numbers above the gel correspond to the numbers of the DNA substrates in (b). d) SDS-PAGE analysis of cleavage reactions with XerH and 31-bp dif_H suicide substrates to assess if 31 bp is a sufficient substrate length to support XerH cleavage. The numbers above the gel correspond to the numbers of the DNA substrates in (b).

3.2.9 Cloning, expression, and purification of XerH catalytic mutant Y344F

To ensure that the observed cleavage activity is resulting from the specific catalytic activity of XerH, mutant XerH containing the predicted catalytic tyrosine (Tyr344) mutated to phenylalanine (Y344F) was designed and cloned by a site-directed mutagenesis approach. Expression and solubility of XerH Y344F variant were first tested in small-scale expressions using conditions that previously supported soluble expression of wild-type XerH (Figure 3-15a). Based on the successful results, the same conditions (expression in *E. coli* strain BL21 (DE3) with initial growth at 37 °C and expression at 15 °C for 15 hours) were chosen for large-scale production of XerH Y344F. The mutant protein was then purified according to the same protocol as the wild-type XerH. The purification yielded a homogenous SEC peak corresponding in size to XerH Y344F monomer (Figure 3-15b) and protein purity (>95%) suitable for further biochemical assays as assessed by

analysis on SDS-PAGE (Figure 3-15c). For further analysis, fractions 3-6 (Figure 3-15c) were combined.

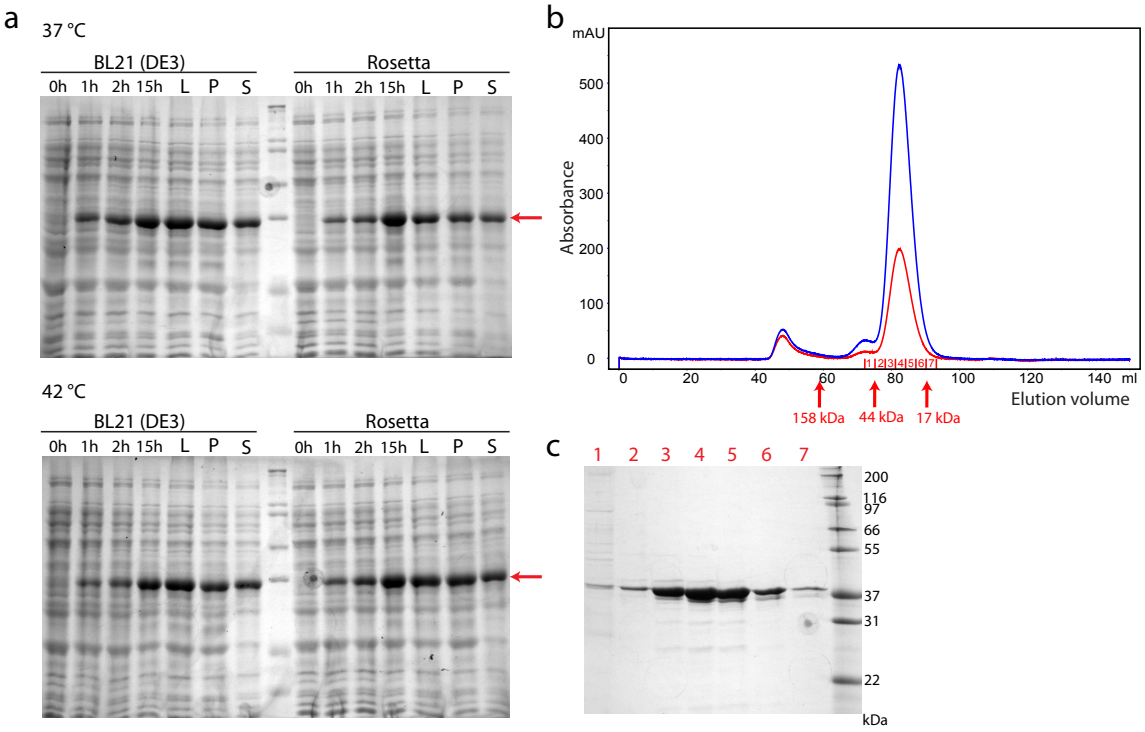


Figure 3-15: Overexpression and purification of XerH Y344F mutant. a) Small scale expression of XerH Y344F in two different *E. coli* strains, BL21 (DE3) and Rosetta (DE3), as analysed on SDS-PAGE. The samples were analysed immediately after induction (0h), then 1 (1h), 2 (2h), and 15 (15h) hours later, after culture harvest and lysis by sonication (L), and after ultra-centrifugation that separated the soluble lysate fraction (supernatant-S) from the insoluble lysate fraction (pellet-P). The positions of the expressed XerH Y344F protein on the gels are indicated by red arrows. b) The final step of XerH Y344F purification: SEC on a Superdex 16/60 gel filtration column. The chromatogram shows the main elution peak that corresponds in size to a monomer of XerH Y344F (42 kDa). UV absorbance at 280 nm is shown in blue, absorbance at 260 nm is shown in red. The fractions collected are shown in red under the elution peak. The elution volumes of molecular weight markers of 17, 44, and 158 kDa on this column are indicated by red arrows. c) SDS-PAGE analysis of the fractions collected from the gel filtration, stained for the presence of protein.

3.2.10 DNA binding and cleavage activity of the XerH Y344F mutant

The DNA binding efficiency of the XerH Y344F mutant was assessed by analytical SEC (Figure 3-16a). The dif_H-XerH Y344F complex was detected when 2-fold protein excess was incubated with the 31-bp dif_H site, with the detected peak corresponding in size to two XerH Y344F monomers bound to one dif_H site. The observed peak resembled that of wild-type XerH, indicating that XerH Y344F binds dif_H with the same stoichiometry and affinity as the wild-type XerH.

Cleavage activity of XerH Y344F was assessed using a dif_H suicide substrate containing a nick on the bottom strand (substrate 2 in Figure 3-14b) using the same conditions as for the wild-type XerH-mediated cleavage assay described in section 3.2.8. Increasing amounts of the dif_H substrate (0-330 pmol) were combined with XerH Y344F and the covalent intermediate formation was assessed by SDS-PAGE (Figure 3-16b). The reactions were also carried out with wild-type XerH as a control. No product could be observed at any DNA concentration when XerH Y344F was used, while the substrate was efficiently cleaved by the wild-type XerH.

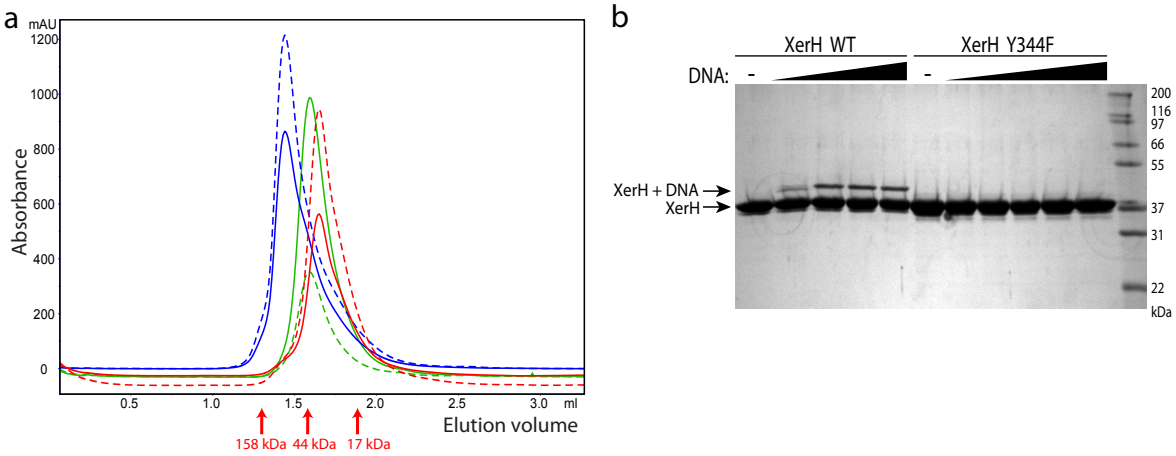


Figure 3-16: Analysis of binding and cleavage activity of XerH Y344F mutant. a) Analysis of XerH Y344F-dif_H complex formation by SEC on a Superdex 3.2/30 gel filtration column. The chromatogram shows analytical gel filtration of XerH Y344F alone (green), 31-bp dif_H DNA alone (red), and the complex of XerH Y344F and 31-bp dif_H DNA (blue). UV absorbance at 280 nm is shown as continuous line, absorbance at 260 nm is shown as a dashed line. The elution volumes of molecular weight markers of 17, 44, and 158 kDa on this column are indicated by red arrows. b) SDS-PAGE analysis of cleavage reactions with XerH Y344F and wild-type XerH, titrated with increasing amounts of 31-bp dif_H suicide substrates.

3.2.11 *In vitro* reconstitution of XerH recombination

Despite many attempts to reconstitute the full XerH recombination pathway from substrates to products *in vitro*, including recombination between two linear, two circular, or one linear and one circular dif_H-containing substrate in conditions supporting efficient DNA cleavage, no recombination products could be observed. As shown in previous experiments, the binding and cleavage of both dif_H half-sites could be reconstituted efficiently; therefore these were likely not the limiting steps in reconstitution of the full pathway *in vitro*. The next expected steps of the reaction pathway are the formation of HJ intermediate by strand exchange between the two DNA recombination partners, isomerization of the HJ, and its resolution into reaction products (see Figure 1-3). The

formation of the HJ is very hard to monitor, as this intermediate is very unstable and quickly resolved back to substrates or forward to products and such intermediates could be not detected in any of the assays performed in this study. To look at the later steps of the recombination pathway, I investigated the ability of purified XerH to resolve the HJ *in vitro*. For this, artificially constructed dif_H HJ with one arm 5'-end labelled with [γ-^{32}P]-ATP (Figure 3-17a) was incubated with XerH in conditions supporting XerH cleavage. The products of the reaction were analysed on a denaturing sequencing TBE gel. Depending on which strands of the HJ were cleaved and rejoined, one or both of the labelled substrates would be converted into products of distinct size. The results show that only the bottom strand of the dif_H HJ substrates was recombined (Figure 3-17b).

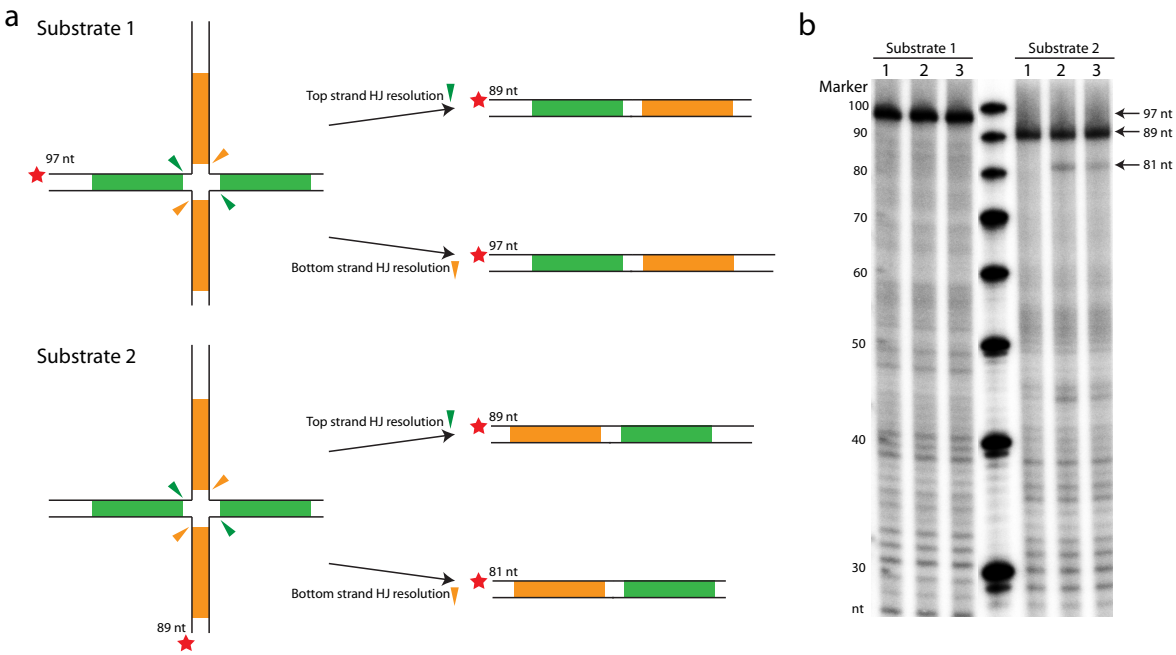


Figure 3-17: HJ resolution assay. a) Schematic representation of dif_H-HJ substrates 5'-end labelled with [γ-^{32}P]-ATP and expected products resulting from cleavage of the top or bottom strand of each substrate by XerH. The left and right binding arms of dif_H are marked in green and orange, respectively. Top and bottom strand cleavage sites are marked by triangles. The red star denotes the oligonucleotide that was labelled and monitored for each substrate. The size of that oligonucleotide in the substrate and expected reaction products are shown in nucleotides (nt). b) Analysis of the reaction products of HJ resolution assay on a sequencing UREA-TBE polyacrylamide gel. Lane 1: no XerH; lane 2: 2 µM XerH; lane 3: 20 µM XerH.

3.2.12 XerH-mediated binding and cleavage of palindromic dif_H substrates

The sequential binding and cleavage of DNA by XerH suggested that the differences between the two binding/cleavage sites of dif_H might be crucial for the mechanism of XerH recombination. Within the region of palindromicity, the two dif_H half-sites differ only by

an insertion of a single basepair present in the left arm (Figure 3-18a). Additionally, the two half sites vary by one basepair within the central region and contain different sequences flanking the palindromic region. To elucidate the functional significance of the differences between the two arms, five palindromic substrates were constructed containing either two left or two right arms in an inverted repeat orientation, with the original or a symmetrized central region between them and the flanking sequences originally from the left or the right arm (Figure 3-18a). The binding of XerH to these substrates was tested by analytical SEC (Figure 3-18b). The results show that at identical XerH:DNA ratios only the left-arm palindrome substrates form stable complexes with XerH, which can be seen as a distinct shifted elution peak relative to DNA only and XerH only controls. For the right-arm palindrome substrates, small amount of a higher molecular weight complex could be detected in the case of dif_H palindrome 1 only, suggesting that binding occurs but is very inefficient. For the remaining right-arm palindrome substrates, no complex formation was observed.

To see if the palindromic substrates can be cleaved by XerH, nicks on one or both strands were introduced into dif_H palindromes 1 and 4 (similarly to the suicide substrates shown in Figure 3-14a) and the substrates were incubated with XerH. The trapped intermediates were detected on SDS-PAGE (Figure 3-18c). A trapped phosphotyrosyl intermediate could only be detected when the dif_H palindrome 4 substrate (left-arm palindrome) was used. As the dif_H palindrome 1 (right-arm palindrome) retains partial ability to bind XerH (Figure 3-18b), the complete lack of cleavage product cannot be attributed to the lack of XerH binding and is a true effect of inability of bound XerH to cleave this substrate.

As the analytical gel filtration assay is disruptive and may not detect weaker complexes, the binding of XerH to various palindromic and half-site substrates was also tested with EMSA. Here, samples containing a constant concentration (4.5 μM) of a 50-bp DNA substrate and increasing concentrations of XerH (0 to 25 μM) were prepared and complex formation was assessed by DNA staining on non-denaturing TBE polyacrylamide gels (Figure 3-19). The native dif_H substrate (Figure 3-19a) showed gradual formation of dif_H complexes first with an XerH monomer and then with an XerH dimer bound. A band of even higher molecular weight could be seen at high XerH concentrations, but this band was also present when XerH was incubated with a non-specific substrate that does not contain XerH binding sites (Figure 3-19b), suggesting that this is a product of non-specific binding caused by high XerH concentration. Incubation of XerH with either a left-arm or a

right-arm palindrome substrate resulted in a comparable binding pattern (Figure 3-19c-d). Interestingly, the palindrome substrates favoured binding of XerH dimers and hardly any single monomer-bound complex could be observed. Additionally, substrates containing only one (left or right) *dif*$_H$ arm, with the adjoining arm replaced with random DNA sequence, were tested in this assay (Figure 3-19e-f). As expected, the majority of the complex formed with these substrates corresponds to an XerH monomer bound to the site. However, despite the lack of a second binding site, a small amount of the XerH dimer-*dif*$_H$ complex could be detected, supporting a cooperative binding model.

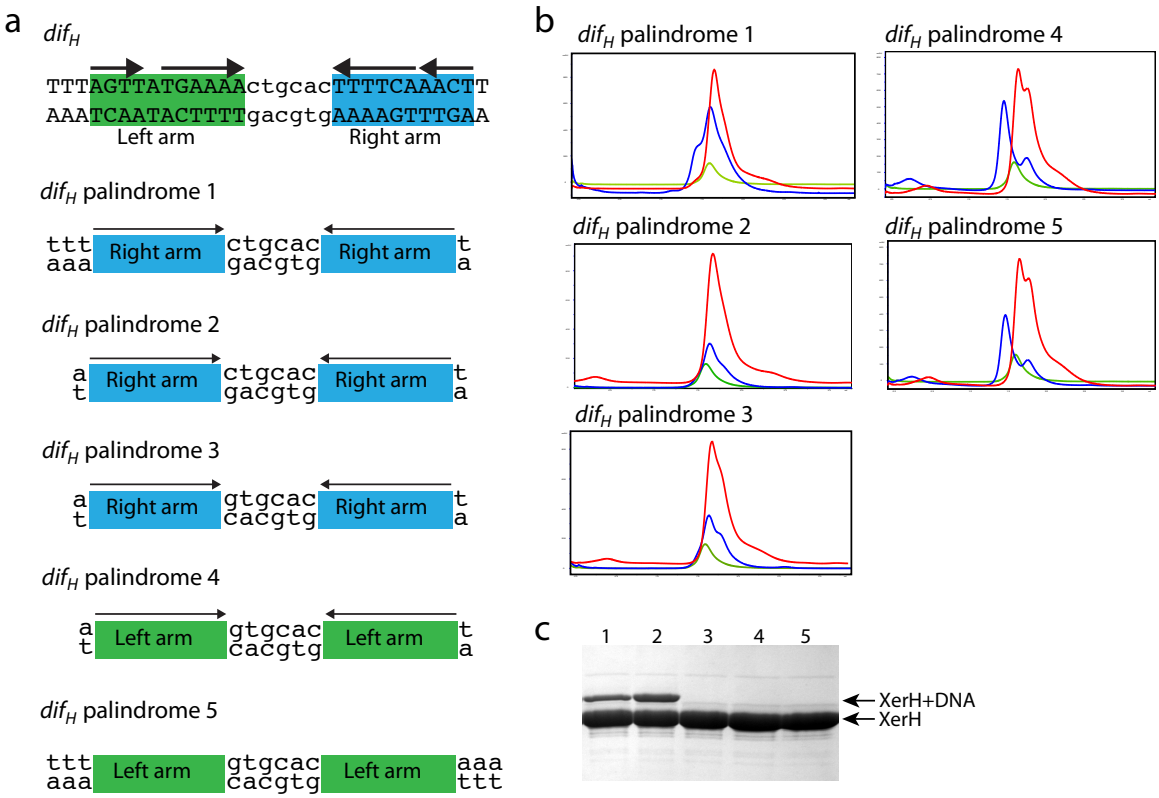


Figure 3-18: XerH binding and cleavage activity on palindromic *dif*$_H$ substrates. a) List of the palindromic *dif*$_H$ substrates used in the study. The left and right XerH-binding arms of *dif*$_H$ are marked in green and blue, respectively. The regions of palindromicity are denoted by arrows. b) Analysis of complex formation between XerH and the palindromic *dif*$_H$ substrates by SEC. The chromatograms show analytical gel filtration of XerH alone (green), *dif*$_H$ substrate alone (red), and the complex of XerH and *dif*$_H$ substrate (blue). Absorbance is plotted against elution volume. Only absorbance at 260 nm is shown. c) SDS-PAGE analysis of cleavage reactions with XerH and left-arm or right-arm palindromic substrates. Lane 1: *dif*$_H$ palindrome 4 (left-arm palindrome) with a nick on one strand; lane 2: *dif*$_H$ palindrome 4 nicked on both strands; lane 3: *dif*$_H$ palindrome 1 (right-arm palindrome) with nick on one strand; lane 4: *dif*$_H$ palindrome 1 nicked on both strands; lane 5: no DNA.

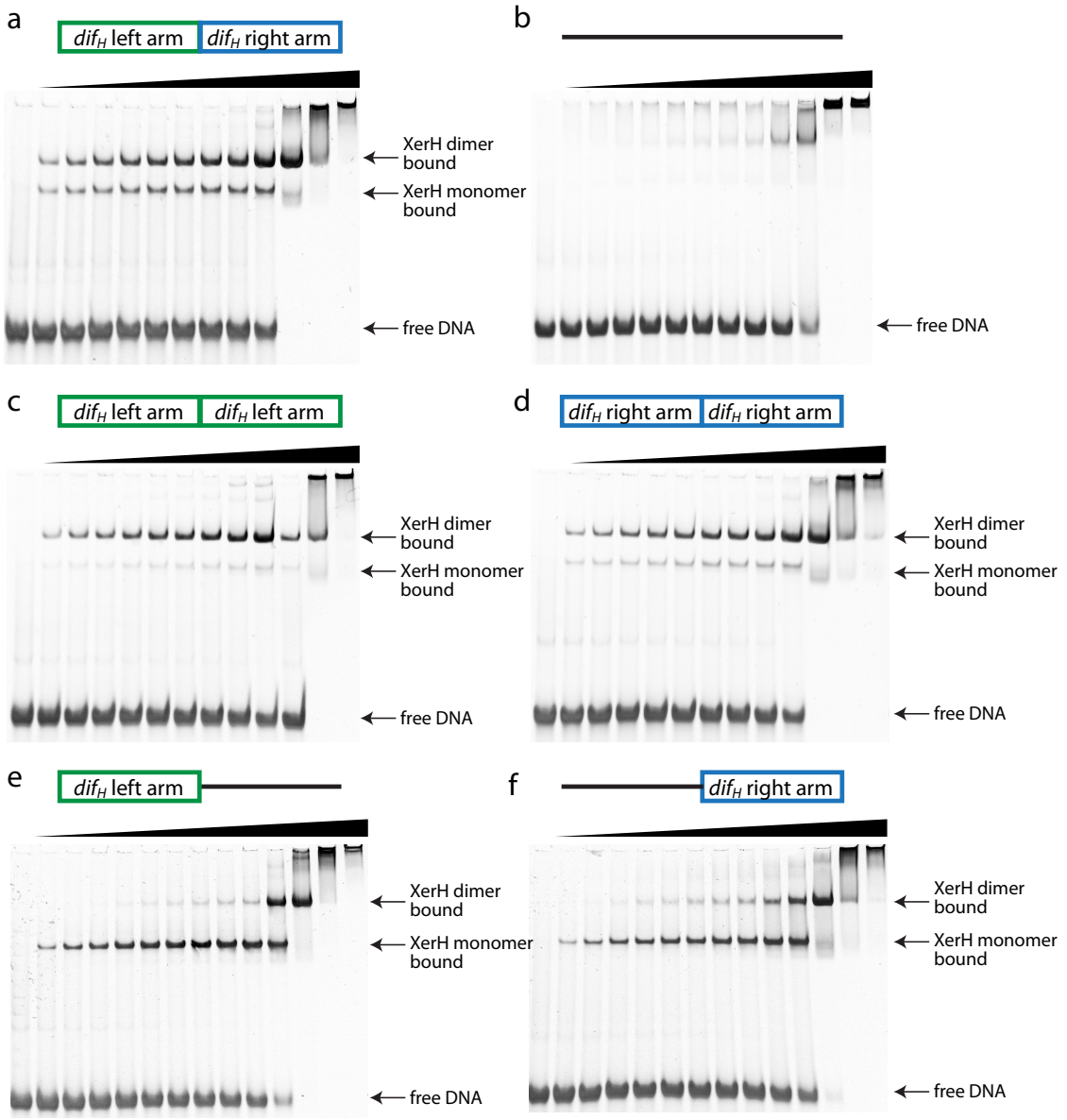


Figure 3-19: XerH binding to various 50-bp dif_H substrates analysed by EMSA on native TBE polyacrylamide gel. The protein concentrations used (from left to right): 0, 0.5, 0.7, 0.9, 1.2, 1.4, 1.6, 1.9, 2.1, 4.2, 8.6, 17.0, and 25.7 µM. a) Native dif_H substrate. b) Random 50-bp DNA with the same G+C content as dif_H. c) Left-arm palindrome substrate. d) Right-arm palindrome substrate. e) Substrate with the left arm of dif_H and a random sequence instead of the right arm. f) Substrate with the right arm of dif_H and a random sequence instead of the left arm.

Together, the results of the experiments with palindromic substrates suggest that both dif_H arms can be bound efficiently (as seen in EMSA), but it seems that the left-arm complexes are more stable (and therefore can be detected in the SEC assay) and the left arm is required for dif_H to be cleaved (as seen in the suicide substrate cleavage assay).

3.2.13 Mutations investigating differences between the two half-sites of *dif*$_H$

The results of the palindrome experiments indicate that the imperfect palindromic regions in the native *dif*$_H$ site have the strongest effect on determining the efficiency of XerH binding and cleavage, while the flanks and the central region have less influence (Figure 3-18). To test the effect of the outer sequences beyond the region of palindromicity more directly, mutational analysis of this region was performed (Figure 3-20a). The results of the DNA cleavage assay revealed that the addition of the left-arm flanking sequence to the right-arm palindrome was not able to rescue XerH-mediated cleavage activity on this arm (Figure 3-20b).

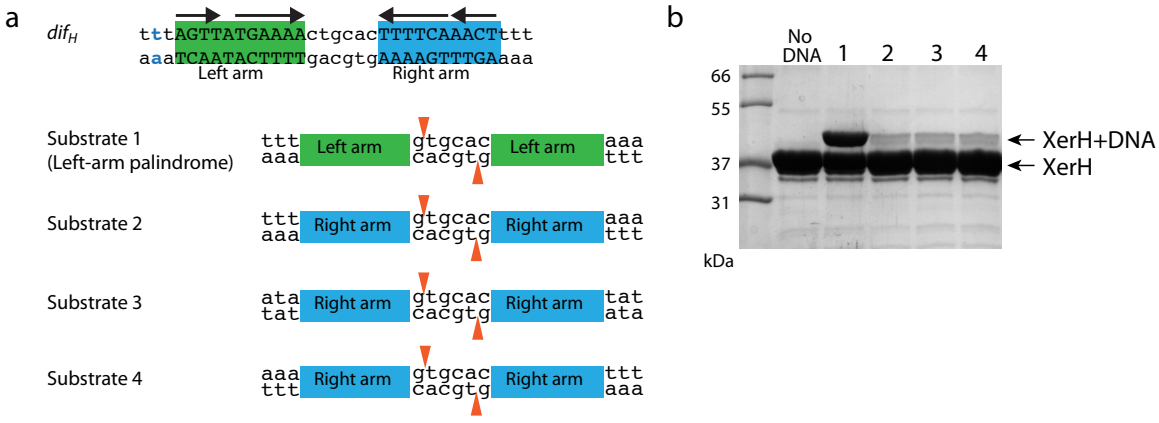


Figure 3-20: Mutational analysis of *dif*$_H$ substrates to analyse the importance of the outer sequences beyond the palindromic region. a) The mutated *dif*$_H$ palindrome substrates used in the experiment. The original *dif*$_H$ site is shown for reference (top), including three basepairs flanking the palindromicity regions of each arm. The T/A basepair marked in blue in the reference sequence is invariant amongst various *dif*$_H$ sites (Debowski *et al*., 2012). b) SDS-PAGE analysis of XerH-mediated cleavage reactions with suicide substrates shown in (a).

The differences in XerH binding and cleavage activity at the two arms of *dif*$_H$ are striking, yet the main functional difference between the two arms is caused by a single-basepair insertion in the palindromic region of the left arm (highlighted in purple in Figure 3-21a). To test the specific effect of this single basepair insertion, left-arm palindrome suicide substrates with the original A-T basepair mutated to other basepairs were designed (Figure 3-21b). Additionally, two substrates containing an additional A-T basepair or no A-T insertion were also tested. The substrates were incubated with XerH and the reactions were analysed by SDS-PAGE (Figure 3-21c). The results show that mutating the A-T basepair to T-A or G-C has no effect on cleavage efficiency and mutation to C-G reduces cleavage activity compared to the original A-T-containing substrate (Lanes 1-4, Figure 3-21c). This

indicates that the sequence of the inserted basepair is not crucial for XerH cleavage and does not explain why the right-arm palindrome lacking this basepair is inactive. On the other hand, addition of another A-T basepair or deletion of the A-T (making the arm sequence equivalent to that of the right arm) results in greatly reduced cleavage activity (Lines 5 and 6 in Figure 3-21c). This suggests that exactly one basepair has to be inserted in the left arm for efficient XerH cleavage, perhaps because the protein-DNA contacts around the insertion site are disrupted when more or less than one basepair is inserted at that position.

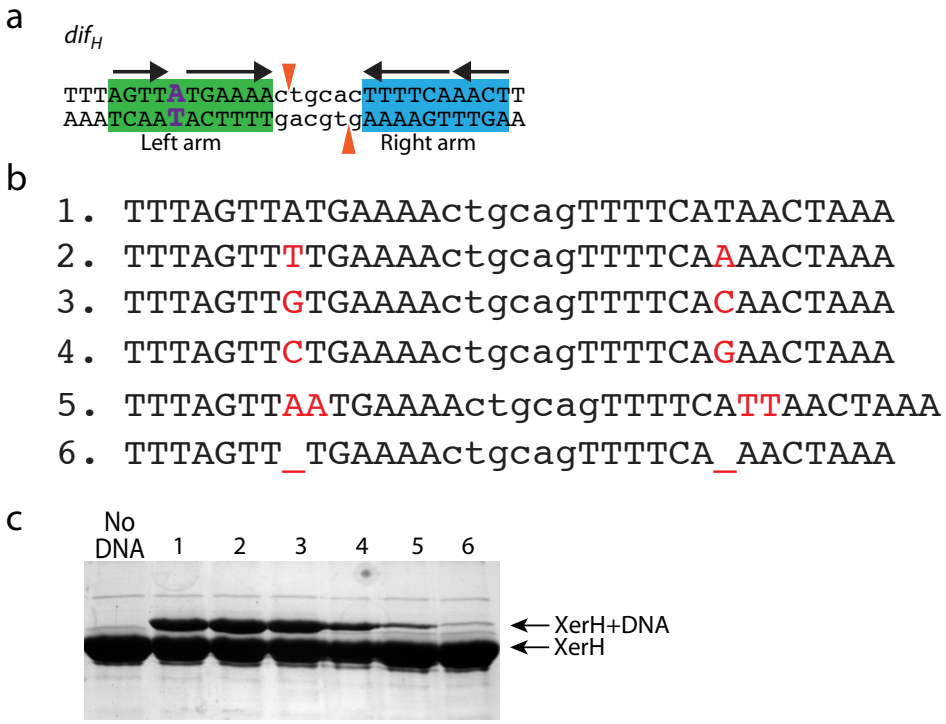


Figure 3-21: Mutational analysis of dif_H to explain the effect of the single-basepair insertion in the left arm. a) dif_H site with the additional nucleotide in the left arm highlighted in purple. The regions of palindromicity are denoted by arrows. Positions of the pre-introduced nicks are indicated by orange triangles. b) List of the mutated dif_H substrates analysed in this study (top strand only shown). The substrates contained double suicide nicks at positions shown in (a). c) SDS-PAGE analysis of the XerH-mediated cleavage of suicide substrates designed based on substrates shown in (b). Numbering of the lanes refers to the substrate numbering in (b).

3.2.14 Broad mutational analysis of dif_H

To further elucidate, which regions of the dif_H site are specifically required for efficient cleavage, and perhaps shed more light on the influence of the A-T insertion of the left dif_H arm on the protein-DNA contacts to the adjoining nucleotides, a broader mutagenic analysis of dif_H was performed. First, the minimal length of dif_H site required for efficient

recombination was established by analysing the XerH-mediated cleavage of progressively shortened suicide substrates by SDS-PAGE (Figure3-22a and b). The analysis revealed that the region of palindromicity is sufficient for efficient cleavage of the left arm (lanes 1-4, Figure 3-22b), while an additional basepair extending from the region of palindromicity is required for efficient cleavage of the right arm (lane 6, Figure 3-22b). Cleavage was also observed when only the palindromic region of the right arm was present in the substrate, however, the cleavage efficiency was reduced approximately three fold compared to the longer sequence (lane 7, Figure 3-22b). Further shortening of either arm resulted in abolished cleavage (lanes 5 and 8, Figure 3-22b). Altogether, these results allowed for definition of the minimal *dif*$_H$ site consisting of the 11-bp left arm, 11-bp right arm, and the 6-bp central region between the arms (Figure 3-22c).

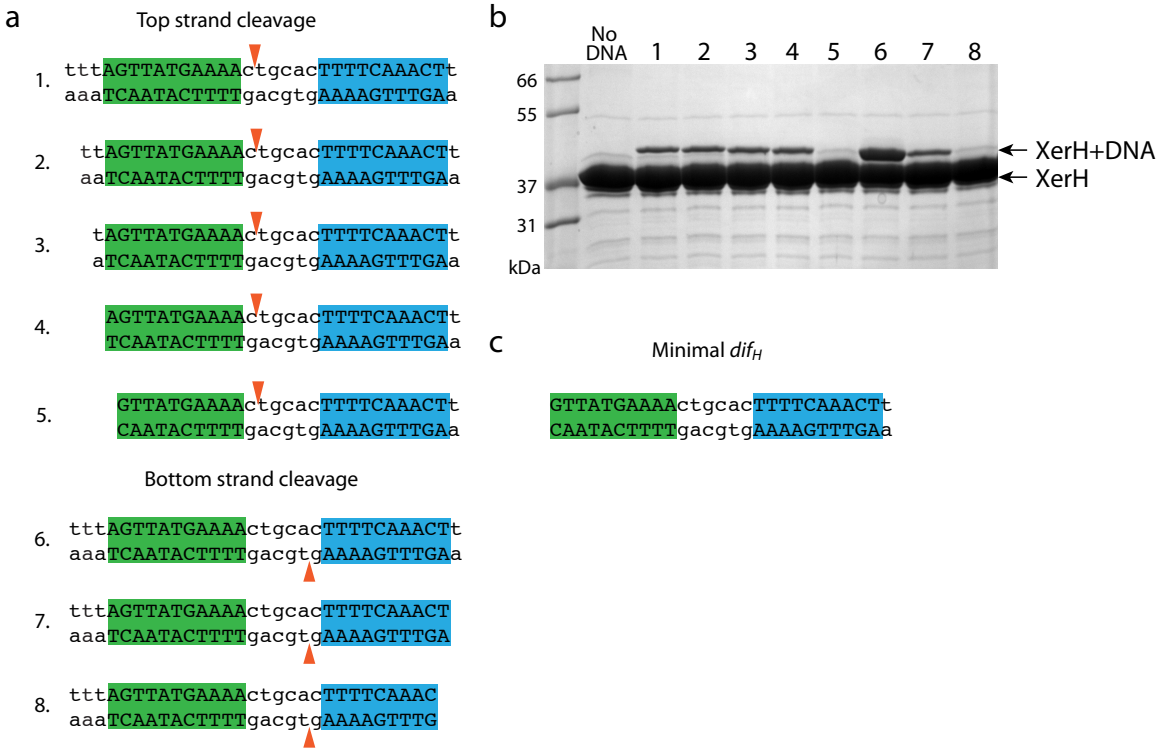


Figure 3-22: Analysis of XerH-mediated cleavage of *dif*$_H$ substrates with progressively shortened binding arms. a) A list of *dif*$_H$ site suicide substrates analysed in the experiment, testing the left arm (top strand) or the right arm (bottom strand) cleavage. Substrates 1 and 6 correspond to the *dif*$_H$ site identified before in the DNaseI footprinting assay and serve as a positive control in this experiment. b) SDS-PAGE analysis of the XerH-mediated cleavage of suicide substrates shown in (a). Numbering of the lanes refers to the substrate numbering in (b). c) The minimal *dif*$_H$ sequence necessary for efficient cleavage of both arms by XerH.

Some insights into the required residues of *dif*$_H$ have been gained previously by Debowski and colleagues (Debowski *et al*., 2012): the invariant nucleotides of the *dif*$_H$ site (Figure 3-23a, nucleotides in blue) were identified based on analysis of 24 *H. pylori*-related bacterial

species and further dif_H mutagenesis revealed that some of these conserved nucleotides are essential for XerH recombination *in vivo* (Figure 3-23a, marked with stars).

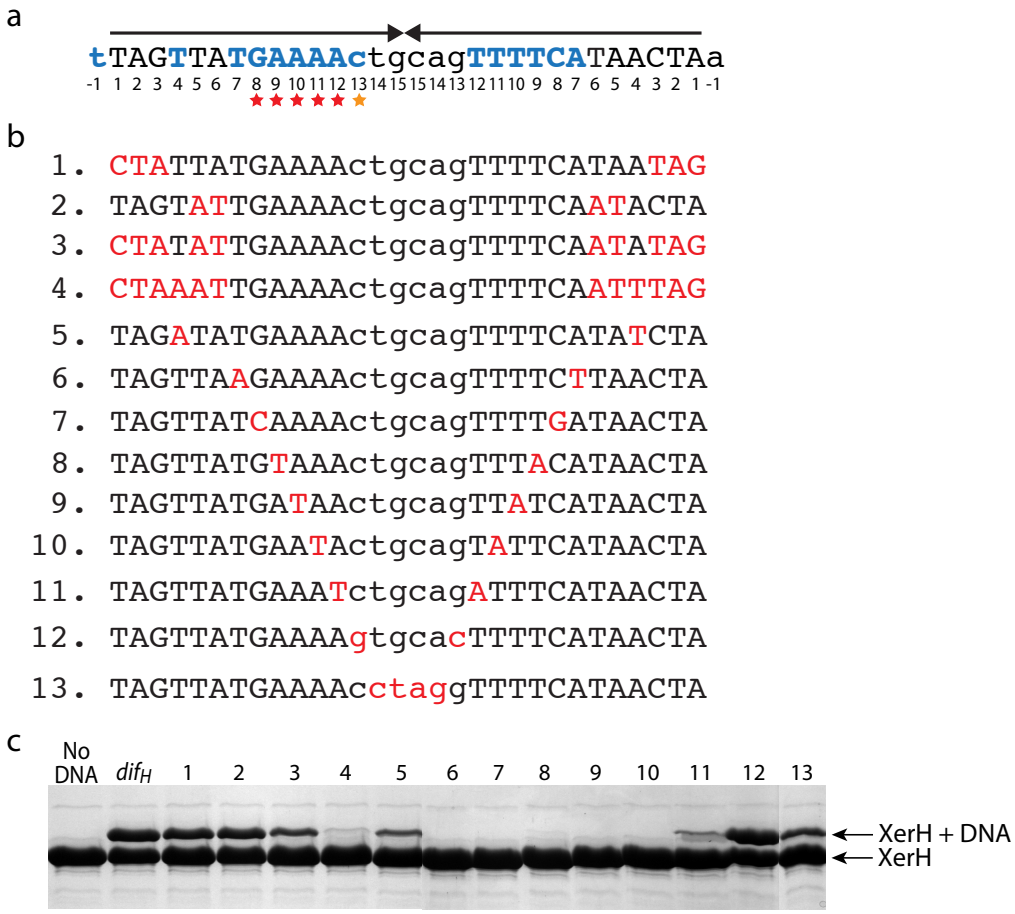


Figure 3-23: Mutational analysis of dif_H. a) Left-arm palindrome dif_H derivative used as a base for the mutagenesis. The top strand only is shown. The invariant dif_H nucleotides are shown in blue. The nucleotides previously mutated resulting in abolished XerH recombination *in vivo* are denoted by red stars, and the nucleotide which when mutated greatly reduced XerH recombination *in vivo* is denoted by an orange star (Debowski *et al*., 2012). The arrows indicate the palindromic region. b) dif_H palindromic substrates used in this study (top strand shown). Nucleotides in red are mutated compared to the original sequence shown in (a). The substrates contained pre-introduced nicks positioned as in previous experiments (see for example Figure 3-21). c) SDS-PAGE analysis of XerH-mediated cleavage of the mutated substrates shown in (b). Numbering of the lanes refers to the substrate numbering in (b).

I extended this analysis by mutating individually all the invariant nucleotides of dif_H as well as the parts that are not conserved in *H. pylori*-related species (Figure 3-23b) and tested the mutation effects *in vitro*. Suicide substrates with the desired mutations were incubated with XerH and analysed on SDS-PAGE. The analysis revealed the following features: 1. Mutations of dif_H positions 7-12 completely abolished or greatly reduced XerH cleavage activity in agreement with the previous study (lanes 6-11, Figure 3-23); 2. Mutation of position 13, which was shown to reduce recombination *in vivo*, had no effect

on XerH cleavage efficiency (lane 12, Figure 3-23); 3. Mutations of positions 1-3 and 5-6 had no or little effect on cleavage activity (lanes 1-3,Figure 3-23); 4. Mutation of position 4 (conserved nucleotide) reduced cleavage activity approximately 3-fold (lane 5, Figure 3-23) but in conjunction with mutations at positions 1-3 and 5-6 it abolished the cleavage altogether (lane 4, Figure 3-23); 5. Mutation of positions 14-15 reduced cleavage activity approximately two-fold (lane 13, Figure 3-23).

3.3 Discussion

3.3.1 Recombinant XerH is proficient in dif_H binding and cleavage

H. pylori XerH was successfully overexpressed in *E. coli* and purified to homogeneity from *E. coli* lysate (Figures 3-7 and 3-10). The purified protein bound efficiently to its predicted DNA site dif_H in two distinct assays: EMSA, which is performed at low DNA and protein concentrations, and analytical SEC, in which the complexes are highly concentrated (Figures 3-12 and 3-13). Additionally, the obtained protein showed cleavage activity on suicide dif_H substrates (Figure 3-14). The substrates trap the covalent intermediate between XerH and dif_H DNA (which is otherwise unstable and is normally immediately converted to further intermediates or back to substrates), allowing for visualization of cleavage. The assay has been previously used in characterization of other tyrosine recombinases (Guo *et al*., 1997; Pargellis *et al*., 1988) as it depends on the formation of a phosphotyrosyl bond between the recombinase and the DNA, which is the characteristic intermediate of tyrosine recombination. Accordingly, the ability of XerH to form products in this assay suggests that XerH employs the same or an equivalent mechanism as other tyrosine recombinases. To further confirm this notion, a mutated XerH variant containing a substitution of the predicted catalytic tyrosine for phenylalanine (Y344F) was tested in the cleavage assay (Figure 3-16b). This mutant did not cleave the dif_H suicide substrate, and since the XerH Y344F mutant was proficient in binding (Figure 3-16a), the observed effect was directly caused by the inability of the mutant to catalyse the cleavage reaction. This indicates that the tyrosine at position 344 is indeed the catalytic tyrosine of XerH required for DNA cleavage, and that the covalent intermediate observed contains the expected phosphotyrosyl bond. The observed conservation of cleavage mechanism is not surprising considering that the catalytic residues of tyrosine recombinases are all present in XerH (Figure 3-1). Similarly, a homology-based XerH

model created with the Phyre2 server was based on structurally characterized tyrosine recombinases, which Phyre2 selected as most closely related to XerH. In this model, XerH assumes a conformation that allows DNA binding, with the predicted catalytic residues assembled into the catalytic centre facing the predicted DNA-binding cleft (Figure 3-9). Together, these findings strongly suggest that the mechanism of XerH resembles that of other tyrosine recombinases. However, the above assays only address the predicted XerH monomer structure, catalytic residue position and importance, and the first steps of the recombination reaction (binding and cleavage), while they do not investigate questions related to the synaptic complex assembly and the later steps of the recombination pathway, which have to be addressed with other assays.

3.3.2 Characterization of the minimal dif_H site required for XerH activity

The 40-bp dif_H site previously identified and tested in *H. pylori* (Debowski *et al*., 2012) appeared to be longer than other known *dif* sites (ranging between 28 and 34 bp; Cortez *et al*., 2010; Leroux *et al*., 2013; Nolivos *et al*., 2010) or even those of other tyrosine recombinases (for example, Cre *loxP* site is 34 bp long; Hoess *et al*., 1982). Indeed, the DNaseI footprinting assay revealed a shorter site of 31 bp protected by XerH binding (Figure 3-11). This length lies within the range of known Xer recombination site lengths and would be sufficient for binding of two XerH monomers. A 31-bp long oligonucleotide representing this site was efficiently bound and cleaved by XerH (Figures 3-12, 3-13, and 3-14d), suggesting that 31-bp dif_H might be sufficient for XerH recombination. Further shortening of the substrate at either end revealed that 28-bp long dif_H site is sufficient for efficient XerH cleavage of both strands (Figure 3-22). However, as *in vitro* reconstitution of XerH recombination with this or longer sites was unsuccessful, it is not clear if this length of dif_H would be sufficient for XerH recombination from substrates to products. On the other hand, examples of other Xer recombinases show that this length of a *dif* site is sufficient also for *in vivo* recombination; for example, the presence of 28 bp of *E. coli dif* site is sufficient for *in vivo* XerC/D recombination, and archaeal XerA recombines *dif* sites of the same lengths (Cortez *et al*., 2010; Leslie and Sherratt, 1995). Also the single-recombinase Xer systems from bacteria include *dif* sites of similar lengths tested in *in vivo* recombination assays: *C. jejuni* 29-bp dif_H and streptococcal 34-bp dif_{SL} are both sufficient for recombination by their respective Xer recombinases XerH and XerS (Le Bourgeois *et al*., 2007; Leroux *et al*., 2013).

3.3.3 Sequential binding and cleavage of *dif$_H$* by XerH

The EMSA experiment (Figure 3-12) revealed an XerH/*dif$_H$* binding pattern, indicating that two monomers of XerH bind sequentially to two *dif$_H$* arms. The binding of the two monomers appeared to be cooperative as the dimeric complex was efficiently formed even at protein to DNA ratios not saturating the monomeric complex. Unfortunately, the cooperativity levels could not be calculated accurately from the EMSA gel and other methods to measure affinity for naked and monomer-bound DNA (such as thermophoresis or isothermal titration calorimetry) were not employed in this study. However, the observation that *dif$_H$* substrates containing only one XerH binding arm were still bound (albeit weakly) by the second XerH monomer (Figure 3-19e and f) supports the notion that the binding of the second XerH monomer is enhanced by interactions with the already *dif$_H$*-bound monomer. This observation is in agreement with previous studies of Xer and other tyrosine recombinases where various degrees of binding cooperativity have been observed (Blakely *et al*., 1993; Leroux *et al*., 2013; Nolivos *et al*., 2010; Ringrose *et al.,* 1998). Since the binding of XerH to the right arm palindrome substrates was greatly reduced or completely abolished in the analytical SEC assay compared to the binding to the left-arm palindrome and to the native *dif$_H$* site, with the latter two showing comparable levels of XerH-*dif$_H$* complex formation (Figures 3-13a and 3-18b), it is reasonable to assume that in the sequential binding of XerH monomers to *dif$_H$* the left arm is bound first and this event promotes binding of the second monomer to the right arm. As shown by the EMSA assay, the right arm of *dif$_H$* can also be bound by XerH in the absence of the left arm (Figure 3-19c and e), but it seems that these complexes are less stable and therefore cannot be detected in an analytical SEC assay, which might disrupt unstable complexes at the applied conditions (Figure 3-18b). Furthermore, since the cleavage assay with palindromic substrates (Figure 3-18c) showed that only left arm palindromic substrates can be cleaved, it is likely that the observed binding to the right arm in the absence of the left arm is not permissive for DNA cleavage, and the left arm binding is a prerequisite for correct, functional assembly of the synaptic complex.

A cleavage assay with suicide *dif$_H$* substrates where the left arm has been extended revealed that the product of cleavage at this arm is much less abundant than that of the right arm when entrapment of single arm cleavage is promoted by a single strand nick in DNA (Figure 3-14). This suggests that cleavage, similarly to binding, might be sequential and the right arm is cleaved first. Taken together with the results of the EMSA and SEC

experiments, this would indicate that the left arm is bound first but the right arm is cleaved first. Such a mechanism would ensure that no cleavage could occur until both monomers of XerH are bound to the site, preventing uncontrolled and aberrant cleavage that could lead to detrimental chromosomal DNA breaks. In fact, such a mechanism has been observed for the XerC/D recombination system, where the XerC monomer binds to *dif* first, but the first cleavage in the recombination pathway is performed by the XerD monomer (Aussel *et al*., 2003).

3.3.4 The role of dif_H asymmetry in ordering XerH binding and cleavage

The two arms of dif_H differ only by a single basepair insertion in the left arm within the palindromic region required for XerH binding, yet the effect on XerH binding and cleavage activity at each arm is striking. Mutational analysis suggests that the effect of the single nucleotide insertion is not exerted by the identity of the inserted basepair, as single-basepair mutations do not have an effect (Figure 3-21). Instead, insertion of exactly one basepair is required to assure correct positioning of the outer sequences further away from the central region for XerH interactions, as indicated by the fact that dif_H site derivatives with no or two inserted basepairs are no longer cleaved by XerH. Two highly conserved basepairs at positions -1 and 4 of the left arm of dif_H (Figure 3-23) could be involved in such contacts since they are invariant amongst bacteria related to *H. pylori* (Debowski *et al*., 2012). Deletion or addition of a basepair at position 6 would reposition the T-A basepair at position 4 and instead G-C basepair would be present at this position, perhaps weakening the interaction between XerH and dif_H. However, the exact nature of such XerH-dif_H interactions could not be derived from the mutational experiments.

In light of other Xer recombinase studies, the high symmetry of dif_H and its strong effect on XerH binding and cleavage order is unusual. The *dif* site of the XerC/D system shows no sequence identity between the outer parts of the two binding arms (see Figure 3-3), consistent with the fact that two distinct proteins bind to the two arms (hence, the asymmetry of that system is not surprising). On the other hand, single recombination systems such as archaeal XerA/*dif* or streptococcal XerS/dif_{SL} employ a single Xer protein, yet the outer sequences of the *dif* binding arms differ from each other (see Figure 3-4). This in turn provides the observed different Xer binding and cleavage affinities for each arm, and likely regulates the recombination pathway (Cortez *et al*., 2010; Nolivos *et al*.,

2010). XerH/dif_H systems from both *H. pylori* and *C. jejuni* differ from the other Xer systems in that the dif_H site is highly symmetric throughout the whole region required for XerH activity, yet the binding and cleavage activities are asymmetric (Leroux *et al*., 2013). This suggests that this Xer recombination system might employ a more subtle mechanism to regulate the order of DNA binding and cleavage through the sequence of the recombination site.

3.3.5 Sequence requirements of a functional dif_H site

To answer the question of which parts of dif_H are required for XerH recombination and to shed light on the mechanism of XerH regulation by the dif_H asymmetry, mutational analysis of dif_H was performed (Figure 3-23). This confirmed that the same residues of dif_H (positions 8-12 in Figure 3-23) are required for XerH cleavage *in vitro* and recombination *in vivo* (Debowski *et al*., 2012). The only exception found is the residue at position 13, mutation of which was found to greatly reduce recombination efficiency (Debowski *et al*., 2012) but did not affect cleavage efficiency *in vitro* (lane 12, Figure 3-23), suggesting that this residue might be important for later strand-exchange steps in the recombination pathway. This would be in accord with results previously observed for other recombinases: for example, the sequence of the central region nucleotide adjacent to the cleavage site was proposed to define the order of cleavage and strand exchange in Cre recombination (Guo *et al.,* 1999). Furthermore, mutations at position 4 of dif_H, a highly conserved dif_H residue, produced surprising results: while mutation of position 4 alone reduced cleavage activity, in conjunction with mutations of non-conserved residues 1-3 and 5-6 this mutation abolished cleavage activity completely (Figure 3-23c). This suggests that the residues surrounding the A-T basepair at position 4 are crucial for providing the right environment for correct interactions with XerH. Together, mutational analysis of dif_H allowed for identification of residues specifically required for XerH cleavage activity *in vitro,* but mechanistic explanation of these findings would require visualization of specific protein-DNA contacts in a crystal structure.

3.3.6 XerH is insufficient for dif_H recombination *in vitro*

Despite many attempts, XerH recombination could not be reconstituted *in vitro* in conditions supporting DNA cleavage. The observed lack of *in vitro* XerH recombination activity could be caused by many factors, including inappropriate buffer conditions or undetectable recombination levels. However, it could also indicate that some other factor required for recombination is missing from the *in vitro* setting.

The results of the HJ resolution experiment (Figure 3-17) show that only the bottom strand of the dif_H (the right arm) is cleaved. Therefore, assuming that the XerH monomer bound to the right arm cleaves first as implied by the cleavage assay (Figure 3-14c), it seems that the HJ intermediate can be converted back to substrates but does not proceed to products (which would require cleavage of the second, top strand). Together, the *in vitro* results suggest that, in given conditions, XerH can perform the first DNA cleavage (of either strand) and strand exchange reactions, likely resulting in formation of the HJ intermediate; however, XerH can only convert the HJ back to substrates (Figure 3-17). As a result, no full recombination reaction can be observed. This is strikingly similar to what was observed for XerC/D recombination assays *in vitro* in the absence of the FtsK protein (Hallet *et al.,* 1999). There, the XerC monomer cleaves the *dif* substrate and forms the HJ intermediate, which can be only resolved back to products. However, in the presence of FtsK the XerD monomer becomes active, cleaves and forms the HJ, which is then resolved by cleavage and strand exchange by the XerC monomer. Thus, it is likely that XerH recombination was never observed *in vitro* because XerH is not sufficient for recombination at dif_H, and an additional factor (FtsK or another protein) might be required for recombination to occur. However, the fact that initial cleavage by both XerH monomers could be seen (Figure 3-14c) suggests that the mechanism of XerH regulation might be different from that of XerC/D; for example, the additional protein could facilitate HJ isomerization to support progress of the reaction towards recombination products.

4. The crystal structure of the XerH-dif_H complex

4.1 Introduction

4.1.1 Principles of biomolecular X-ray crystallography

The structure of molecules such as proteins or nucleic acids cannot be seen directly using a light microscope, and even application of recent advances in the much more powerful electron microscopy does not always allow for unambiguous visualization of fine features such as amino acid side chains or DNA bases. Therefore, a different, indirect visualization method is nowadays the most popular tool for obtaining structural information about biological macromolecules. This method is called X-ray crystallography and is based on diffraction of X-rays by a crystal composed of an array of molecules of interest (summarized in Rhodes, 2006).

In order to visualize an object of interest using electromagnetic radiation, the radiation must interact with the object and should have a wavelength similar to the smallest features to be observed. X-ray radiation fulfils this requirement for macromolecules such as proteins and DNA because it interacts with electrons and its approximately 1 Å wavelength allows visualizing inter-atomic distances of ≈1.2 Å. However, the nature of X-rays does not allow for direct visualization of the scattering object as X-rays cannot be bent by a lens in order to reproduce the image of the object as is possible in the case of visible light in a microscope or eye lenses (Figure 4-1). Instead, the pattern of the diffracted X-rays can be collected on a detector and then mathematically converted back into the image of the scattering object (in this case, the electron density). This means that the computer simulates the effect of a lens in reconstruction of the object, rendering X-ray diffraction an indirect method of object visualization.

X-ray diffraction of a single macromolecule is very weak and cannot be detected on today's detectors. Therefore, diffraction of multiple macromolecules organized in a crystal lattice consisting of millions of repeated identical unit cells is measured in an X-ray diffraction experiment. Molecules ordered in the same conformation and orientation within the crystal diffract X-rays in an identical manner, together producing a stronger and

detectable diffraction pattern. However, the production of high-quality, well-ordered crystals from macromolecules such as proteins, DNA, or complexes of both is often the limiting step in the X-ray crystallography experiment, as such macromolecules normally do not crystallize in nature. In order to form crystals of a macromolecule, the macromolecular solution has to be brought up to supersaturated state in a controlled manner. This could be achieved by increasing the macromolecule concentration, by addition of crystallizing agent (so called precipitant), or by combination of both. When supersaturation occurs, the macromolecule in the solution is forced to leave the solution state and can precipitate, form so-called protein gels, or nucleate and form crystals. In order to identify the right conditions to form crystals rather than another form of precipitate, a broad range of crystallization conditions is usually screened.

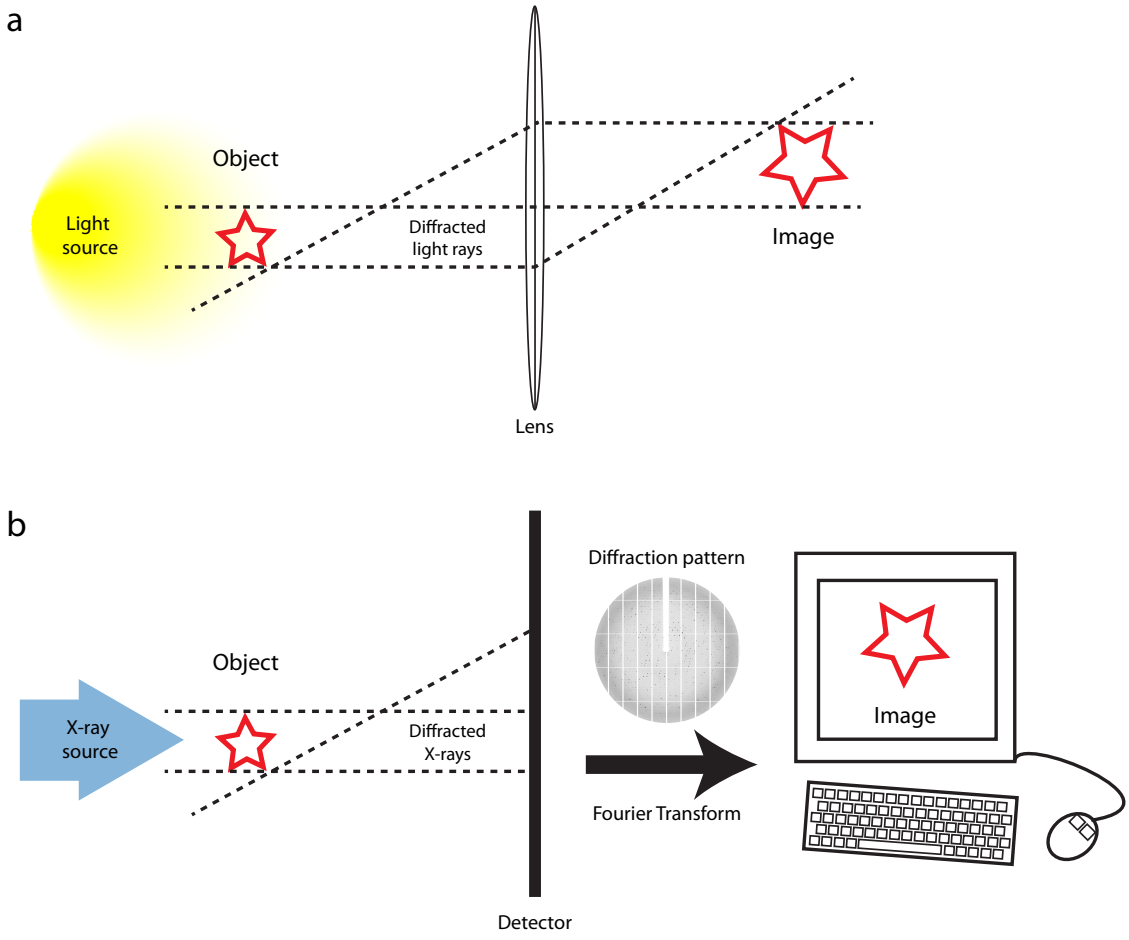


Figure 4-1: Visualization of objects using direct and indirect methods. a) Direct visualization of an object using light rays refocused by a lens, such as the objective lens of a microscope. b) Indirect visualization of an object using X-ray diffraction. The diffracted X-rays cannot be refocused by a lens, but collected diffraction pattern can be converted into an image by Fourier Transform methods.

Once the crystals are obtained, the X-ray diffraction pattern can be collected using high-intensity X-ray sources such as synchrotrons or in-house X-ray generators. The diffraction pattern collected on the detector (Figure 4-2) consists of spots created by exposure to the diffracted X-rays, known as reflections. These reflections result from interference on the crystal lattice and their position on the detector is inversely related to the intermolecular spacing of the repeated unit cells in the real crystal lattice. For each reflection, the position *hkl* (and therefore the direction of diffraction) and the intensity I_{hkl} can be observed directly from the diffraction pattern, and the parameters of the unit cell can be derived directly. What is more, there is a mathematical relationship between the scattering object and its diffraction pattern which is described by the so-called Fourier transform (Fourier, 1822). The Fourier transform is a mathematical operation performed on a diffraction pattern to obtain the original image of the diffracting object, and *vice versa*. Interestingly, one consequence of Fourier transformations is that the reciprocal space is centro-symmetric, indicating that reflections *hkl* and *-h-k-l* from a diffraction pattern are equivalent.

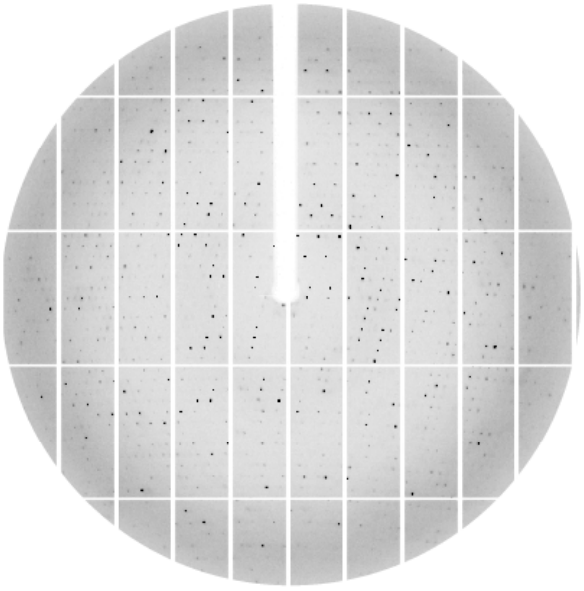


Figure 4-2: Example diffraction pattern collected in an X-ray diffraction experiment

X-rays are complex electromagnetic waves, and as such can be described as a sum of sinusoids of different amplitudes and frequencies. Fourier transform allows deconvolution of any function into the sum of its individual components with their individual amplitudes, frequencies and phase shifts. Each X-ray beam diffracted by the electrons of the object and resulting in reflection spots in the X-ray diffraction experiment corresponds to one term in

the Fourier sum that describes the image (the electron density) of the macromolecule of interest (Equation 4-1). This means that if the information about the amplitude, frequency, and the phase of each individual collected reflection (which together constitute the structure factor function F) is available, the electron density distribution of the diffracting object can be reconstituted by performing an inverse Fourier transform.

$$\rho(x,y,z) = \frac{1}{V} \sum_{h=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} \sum_{l=-\infty}^{+\infty} \mathbf{F}\,(h,k,l) \cdot \exp(-2\pi i\,(hx + ky + lz))$$

Equation 4-1: The electron density equation. ρ (x,y,z) – electron density at real space coordinates x, y, and z; V – the volume of the unit cell; $\mathbf{F}$ (h,k,l) – the structure factor for reciprocal space coordinates h, k, and l.

However, due to the nature of the experiment, the only parameters available for all reflections in the diffraction pattern are the angle of diffraction (θ; described by position *hkl*) and the amplitude (described by intensity I_{hkl}), while the phase information is missing from the obtained diffraction pattern. This results in the so-called crystallographic phase problem, and in order to complete the Fourier summation, phase information has to be obtained separately.

The phases, necessary to solve the structure of a macromolecule analysed in the diffraction experiment, can be derived experimentally (experimental phasing) or from another macromolecular model (molecular replacement).

In the molecular replacement method, a known structure of a related protein is used as a phasing model – the phase values from the structure factors of that protein are used as the initial phases for the new protein structure. The method involves placing the phasing model structure into the unit cell of the crystal and performing rotational and translational searches to match the position of the protein in the new structure. The more similar the model, the easier it is to find the position of the new protein, and successful molecular replacement requires that the two proteins share a similar fold and conformation. Since structural similarity can be roughly predicted based on amino acid sequence identity, a sequence identity of approximately 30% is generally expected for successful molecular replacement. A match is generally assessed by comparing the observed and calculated

structure factor amplitudes or Patterson functions (density functions calculated with |F| only without phases; Patterson, 1935) of the data and the model.

Experimental phasing by isomorphous replacement involves collecting at least two diffraction datasets: one on a crystal of the native protein of interest and one on its heavy atom derivative. The heavy atom derivative crystals can be produced by heavy-atom soaking of already formed crystals, by addition of heavy atoms to the crystallization condition so that they are incorporated into the forming crystal, or by heavy-atom derivatization of the macromolecule, for example by replacing methionines with seleno-methionines in a protein. Heavy atoms such as mercury or iodine contain large number of electrons and therefore diffract X-rays strongly, resulting in increased reflection intensities at specific positions. The differences between the spot intensities in the diffraction patterns of native and derivative datasets can be used to determine the heavy atom substructure using the so-called Patterson map (Patterson, 1935). This means that the structure factors for the heavy atoms, together with their phases, can be calculated, which provides the initial phases for the entire protein structure. Depending on the number of different derivative crystals used, the method is referred to as single isomorphous replacement (SIR) or multiple isomorphous replacement (MIR). The most important requirement for application of this method is that the native and derivative crystals have to be isomorphous, meaning the crystal packing and unit cell parameters should be practically identical, which may not always be achievable.

Experimental phasing can also be achieved by measuring the anomalous scattering of certain heavy atoms. Heavy atoms absorb X-rays at specific wavelengths, resulting in a detectable inequality of the reflection intensities in symmetry-related reflections *hkl* and -*h*-*k*-*l* that are otherwise equivalent (Friedel's law). This phenomenon is known as anomalous scattering or anomalous dispersion of heavy atoms. It occurs because at wavelengths corresponding to the absorption peak for a given element (e.g. the heavy atom), a fraction of the X-ray radiation is absorbed and re-emitted with an altered phase. This disparity can be used to locate the heavy atom positions with Patterson methods, and then to determine the phases of all reflections. Commonly used is the multi-wavelength anomalous dispersion (MAD) phasing that uses multiple datasets collected on the same crystal at wavelengths corresponding to the absorption peak, inflection point, and remote positions on the heavy atom's absorption spectrum. The advantage of this method is that the obtained phase values are less ambiguous than when only one dataset is collected.

However, this method subjects the crystal to long exposure to the X-rays during multiple data collection rounds, which can cause serious radiation damage to the crystal and greatly decrease diffraction quality over time. As a consequence, often only one data set at the absorption peak wavelength is collected and a less accurate, yet still powerful single-wavelength anomalous diffraction (SAD) phasing is performed.

Once the phases are estimated, an initial electron density map can be calculated from the structure factors. Based on the map, the model of the macromolecule is built either automatically (when the resolution and map quality are high) or manually. In both experimental phasing and molecular replacement, the obtained initial phases are only approximate, and need to be further refined in an iterative process that includes manual model building and fitting, and real space and reciprocal space refinement. During model building and refinement, the fit of the model to the experimental data is constantly monitored by the R-factors (Equation 4-2), which measure how well the proposed model explains the observed data by calculating the difference between the observed and calculated structure factors that correspond to the current model. In general practice, two R-factors are used: R-work that considers all the data obtained, and R-free, which uses only a randomly generated subset (usually 5%) of the reflections that were put aside for this purpose before structure solution and are not biased by the rounds of refinement. Good R-factor values for a complete and refined protein structure are expected to be in the range of 10-25%, while a random set of atoms would give an R-factor of 60%. The final model is also validated in terms of protein chemistry and geometry, and if satisfactory, it is deposited in the PDB for public use.

$$R = \frac{\sum \left| \left|\mathbf{F}_{\text{obs}}\right| - \left|\mathbf{F}_{\text{calc}}\right| \right|}{\sum \left|\mathbf{F}_{\text{obs}}\right|}$$

Equation 4-2: The R-factor. Fobs – measured structure factor amplitude; Fcalc – structure factor amplitudes calculated from the model.

The success of a macromolecular X-ray crystallography experiment depends greatly on many factors including but not limited to the quality of the samples, chosen crystallization screening strategy, data collection strategy, and data analysis. The strategies described in

this chapter have been optimized to yield high quality diffraction data through careful preparation and screening of XerH-DNA complexes, thorough screening and optimization of crystallogenic conditions, data collection strategy adjusted to the observed properties of the crystals, and data analysis using a wide variety of available crystallographic software.

4.1.2 Crystal structures of tyrosine recombinases

Much of our knowledge of the mechanism of tyrosine recombination comes from the X-ray crystal structures of these proteins obtained to date. The available structures include structures of DNA-free monomers and dimers, DNA-bound monomers, and fully assembled tetrameric synaptic complexes trapped at various stages of the recombination reaction (summarized in Table 4-1).

Table 4-1: Crystal structures of tyrosine recombinases available in the PDB database

Structure[a]	Resolution (Å)	Recombinase	Reaction intermediate	DNA substrate	Reference
1CRX	2,40	Cre	Post-cleavage synaptic complex	*loxA*	Guo *et al.*, 1997
2CRX	2,70	Cre	HJ	immobile *lox* HJ	Gopaul *et al*., 1998
3CRX	2,50	Cre R173K	HJ	*loxS6* HJ	Gopaul *et al.*, 1998
4CRX	2,20	Cre R173K	Pre-cleavage synaptic complex	*loxS*	Guo *et al.,* 1999
5CRX	2,70	Cre Y324F	Pre-cleavage synaptic complex	*loxS*	Guo *et al.*, 1999
1F44	2,05	Cre Y324F	Trimer on a three-way junction	Y-shaped *lox* junction	Woods *et al.*, 2001
1KBU	2,20	His-Cre	HJ	*loxP* (C5 to G)	Martin *et al.*, 2002
1MA7	2,30	Cre	HJ	*loxAT*	Martin *et al.*, 2003
1NZB	3,10	Cre	Pre-cleavage synaptic complex	*loxP* (phosphothioate)	Ennifar *et al.*, 2003
1OUQ	3,20	Cre	Post-cleavage synaptic complex	*loxP*	Ennifar *et al*., 2003
1PVP	2,35	Cre ALSHG	Post-cleavage synaptic complex	*loxM7*	Baldwin *et al.*, 2003
1PVQ	2,75	Cre LNSGG	Post-cleavage synaptic complex	*loxM7*	Baldwin *et al.*, 2003
1PVR	2,65	Cre LNSGG	Post-cleavage synaptic complex	*loxP*	Baldwin *et al.*, 2003
1XNS	2,80	Cre	HJ with peptide inhibitor	*loxP*	Ghosh *et al.,* 2005
1XO0	2,00	Cre R173K	HJ	*loxP*	Ghosh *et al.*, 2005
2HOF	2,40	Cre K201A	Pre-cleavage synaptic complex	*loxP*	Ghosh *et al*., 2007
2HOI	2,60	Cre K201A	Pre-cleavage synaptic complex	*loxP*	Ghosh *et al.*, 2007
3MGV	2,29	Cre	Transition state mimic	*loxP* (vanadate)	Gibb *et al.*, 2010
1AE9	1,90	λ integrase	Unbound catalytic domain	-	Kwon *et al.*, 1997
1P7D	2,95	λ integrase	Monomer bound to DNA, post-cleavage	*att* half-site	Aihara *et al.*, 2003
1Z1B	3,80	λ integrase	Post-strand exchange synaptic complex	*att COC'*	Biswas *et al.*, 2005
1Z1G	4,40	λ integrase Y342F	HJ	*att COC'* and arm site	Biswas *et al.,* 2005
1Z19	2,80	λ integrase	Post-cleavage synaptic complex	*att COC'* (vanadate)	Biswas *et al.*, 2005
1FLO	2,65	FLP	HJ	*FRT*	Chen *et al,* 2000
1M6X	2,80	FLPe	HJ	*FRT*	Conway *et al.*, 2003
1P4E	2,70	FLPe W330F	HJ	*FRT*	Chen and Rice, 2003
1AIH	2,70	HP1 integrase	Unbound catalytic domain (dimer)	-	Hickman *et al.,* 1997
1A0P	2,50	XerD	Unbound full-length monomer	-	Subramanya *et al.*, 1997
4A8E	2,99	XerA	Unbound full-length dimer	-	Serre *et al*., 2013

[a] – PDB database accession number

4.1.2.1 Cre

Cre is certainly the most extensively characterized tyrosine recombinase with a large number of crystal structures available for the tetrameric synaptic complex trapped at various steps of the recombination reaction: before cleavage (Ennifar *et al*., 2003; Guo *et al.,* 1999), after first strand cleavage (Guo *et al*., 1997), and with a HJ intermediate (Gopaul *et al.,* 1998; Martin *et al.*, 2002).

The first structure of Cre in a tetrameric synaptic complex, solved at 2.4 Å resolution (Figure 4-3a; Guo *et al.,* 1997), was also the first structure of any tyrosine recombinase in complex with DNA and its main findings still apply to all other tyrosine recombinases investigated to date. This structure provided first insights into the assembly of a tetrameric recombination complex by trapping the post-cleavage synaptic complex using a double-nicked palindromic DNA suicide substrate (*loxA*) with symmetrized core region. In the structure, two Cre monomers bound to one *loxA* site are related by two-fold crystallographic symmetry to another pair of *loxA*-bound monomers, together forming the tetrameric synaptic complex. Similar protein-protein contacts can be found between the two monomers bound to the same *loxA* site as between the two monomers bound to different *loxA* sites, resulting in a pseudo-four-fold symmetric arrangement of the complex. The C-terminal helices of each monomer extend into the body of the neighbouring monomer, forming a cyclic arrangement of the recombinase monomers in the synaptic complex (Figure 4-3a).

Each Cre monomer consists of the N-terminal DNA-binding domain (crystallized residues 20-129), which comprises five α-helices connected by short loops, and the C-terminal catalytic domain (residues 132-341), which contains nine α-helices and a three-stranded β-sheet (Figure 4-3b). Despite the approximate four-fold symmetry of the complex, the differences between the two monomers bound to one *loxA* site are clear: the two most C-terminal α-helices αM and αN are shifted in one monomer relative to the other. This positioning of the C-terminus corresponds directly to the catalytic state of each monomer as the catalytic tyrosine (Tyr324) is positioned within this region (between α-helices αM and αN; Figure 4-3b). In one monomer, the tyrosine is attached covalently to the DNA phosphate (see Figure 1-2), and in the other monomer the tyrosine is further away (5.8 Å) from the DNA sugar-phosphate backbone. Since the C-terminal helices are swapped between the monomers in a cyclic manner, this provides a very efficient way of regulating

monomer activity in an alternating fashion and assures that only two monomers in the tetramer are active at the same time (the "half-of-the-sites reactivity" rule).

The catalytic site of each monomer consists of the well-conserved residues of tyrosine recombinase family. In Cre, these are: Tyr324 (the nucleophile), Arg173, His289, Arg292, and Trp315 (all coordinating the scissile phosphate). An additional conserved residue Lys201 located at a turn in the β-sheet contacts the basepair adjacent to the scissile phosphate. All catalytic residues are contributed to the active site in *cis* by one Cre monomer.

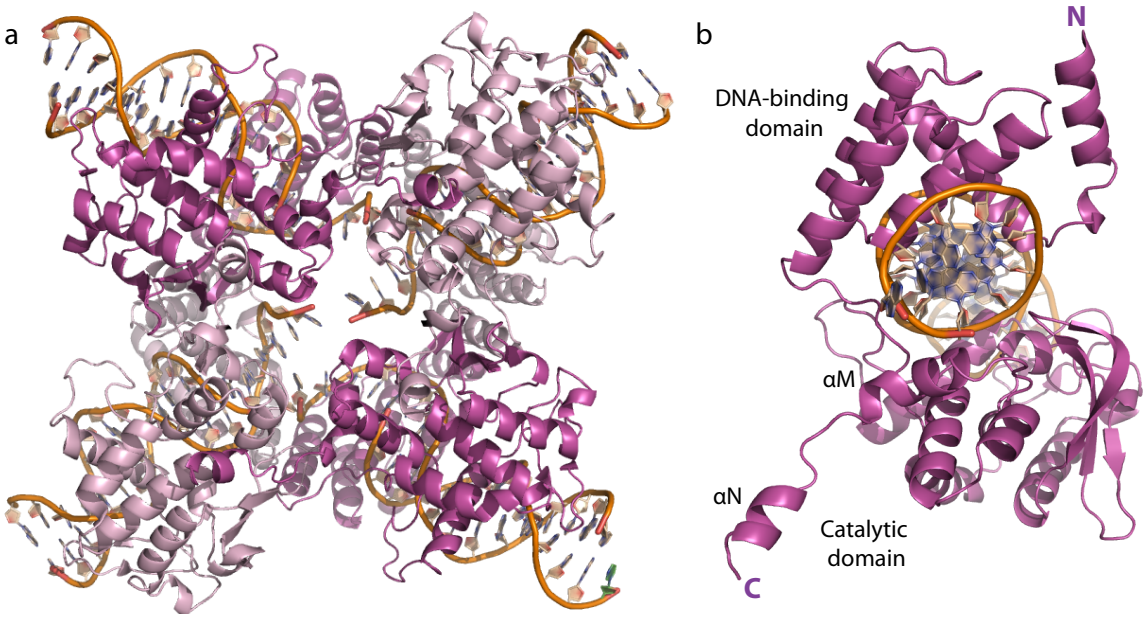


Figure 4-3: The crystal structure of a Cre-*loxA* post-cleavage synaptic complex (PDB accession number 1CRX; Guo *et al.,* 1997). a) Cartoon representation of the tetrameric structure. The two Cre monomers bound to the same *loxA* site (in orange) are shown in two different colours. b) A monomer of Cre forming a C-shaped clamp around DNA.

Each Cre monomer contacts 13 bp of one *loxA* binding arm and 2 bp of the central region. The four innermost nucleotides of the central region are not basepaired, but rather protrude into the centre of the synaptic complex (Figure 4-3a) as if caught in the act of exchanging strands, which is the prerequisite for formation of the HJ intermediate. The protein makes a high number of contacts with both the DNA sugar-phosphate backbone (39 contacts) and the bases, consistent with the site-specific nature of Cre activity. The *loxA* DNA in the complex is bent by 100° via narrowing of the DNA minor groove through protein-DNA

interactions. This bending brings the four DNA arms in a four-fold symmetric arrangement, contributing to the apparent four-fold symmetry of the complex.

The findings from the first Cre synaptic complex structure were complemented by two structures of Cre bound to HJ substrates, one of a Cre catalytic mutant bound to an artificially constructed, immobile HJ (2.7 Å; Figure 4-4a) and one of the wild-type Cre bound to a more flexible HJ formed from nicked DNA substrates by Cre cleavage (2.5 Å; Gopaul *et al*., 1998). Both structures showed nearly identical conformation with pseudo-four-fold symmetry as seen for the post-cleavage synaptic complex (Guo *et al.,* 1997), suggesting that the various intermediates of Cre-*loxP* recombination share similar architecture, which in turn might facilitate progression from one intermediate to another. The similarity of the structures with the immobile and flexible HJ also indicated that no branch migration occurs in the Cre-*loxP* HJ. The main difference between the HJ and the post-cleavage synaptic complex structures is that one pair of the scissile phosphates within the HJ complex has shifted by approximately 1 Å compared to the post-cleavage complex, which removes this phosphate from an ideal hydrogen-bonding position with respect to the catalytic tyrosine. This could be important for the stereochemical switch that inhibits cleavage by the previously active pair of monomers within the HJ intermediate complex. Consequently, it was proposed that the strand cleavage specificity of Cre (and perhaps other tyrosine recombinases) could be coupled to the HJ isomer present and change accordingly with HJ isomerization (Gopaul *et al.,* 1998).

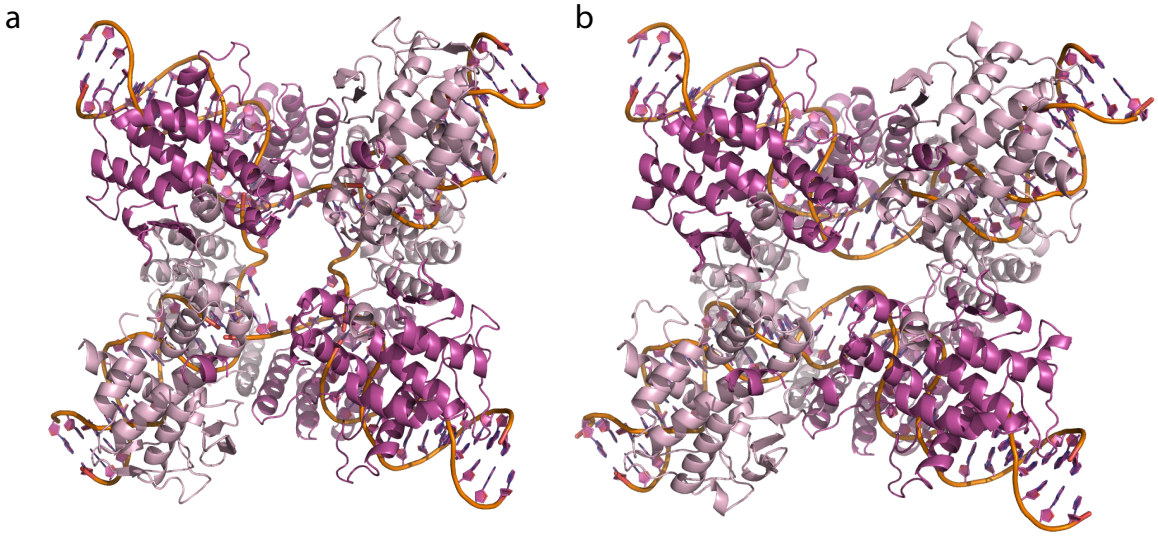


Figure 4-4: Cartoon representation of the crystal structures of tetrameric Cre complexes. The two Cre monomers bound to the same *loxA* site (shown in orange) are shown in two different colours. a) Cre bound to HJ intermediate (PDB accession number 2CRX; Gopaul *et al.,* 1998). b) Cre-*loxA* pre-cleavage synaptic complex (PDB 4CRX; Guo *et al*., 1999)

An additional structure of a Cre HJ intermediate (Martin *et al.,* 2002) containing the native *loxP* site shed light on the mechanism that determines the order of Cre-mediated cleavage events. The structure revealed that the asymmetry of the basepairs adjacent to the cleavage site at each *loxP* arm promotes the first cleavage by the monomer bound to the left arm. Additionally, Lys201 (located on a loop between two β-strands) was also suggested to play an important role in ordering of the cleavage events by specifically recognizing an adenine base adjacent to the left-arm cleavage site. The adenine residue is absent from the right arm, and therefore the contact of Lys201 could trigger the first cleavage specifically at the left arm of *loxP*.

Further structures of Cre pre-cleavage synaptic complexes, containing a symmetrized *loxS* site and a Cre mutant R173K (2.2 Å) or Y324F (2.7 Å) confirmed that the overall pseudo-four-fold architecture is established already before the first strand cleavage (Figure 4-4b; Guo *et al.*, 1999). Mutating of the conserved arginine (R173K) was used as an alternative for the catalytic tyrosine mutation (Y324F), because the latter could disrupt the hydrogen bonds between the catalytic tyrosine and the DNA resulting in disrupted docking of the C-terminal α-helices of Cre into the active site. The *loxS* DNA in the structure is sharply bent (102°) and the bend is a result of a single asymmetrically placed kink within the central cross-over region. Apart from that kink, *loxS* adopts almost ideal B-DNA structure, suggesting that the protein-DNA interactions at the *loxS* binding arms are only stabilizing the bend induced by the DNA kink. The energy stored in the DNA bend was proposed to promote strand exchange events, and the position of the kink was suggested to determine the order of cleavage. Within the central region, it could be seen that the crossing DNA strand is free of any interactions with Cre monomers, while the continuous strand was stabilized in place by protein-phosphate contacts, which is consistent with one of the strands being positioned for strand exchange.

Two structures of a Cre-*loxP* synaptic complex before (2.8 Å) and after cleavage (2.9 Å) with the full tetrameric complex in one asymmetric unit showed that the observed two-fold symmetry of the complex with antiparallel *loxP* orientation is a true feature of Cre recombination and not an artefact of crystal packing (Ennifar *et al.,* 2003). In addition, these structures revealed a different DNA kink than previously observed, highlighting the importance of the central region sequence (unchanged from the wild-type *loxP* in these structures). Nevertheless, the observed kink position near the cleavage site still supported

the notion that the position of the kink defines the order of DNA cleavage and strand exchange.

Later structures of Cre (summarized in Table 4-1) further contributed to the understanding of Cre recombination mechanism. For example, a structure of trimeric Cre bound to a three-way DNA junction, together with biochemical data showed that Cre can efficiently recombine such substrates, revealing the plasticity of protein-protein interactions within the Cre recombination complex (Woods *et al.*, 2001); a structure of two Cre mutants bound to the *loxP* and mutated *loxM7* substrates showed that the Cre-*lox* interactions can be modulated by creating matching mutations in the recombinase and its DNA substrate (Baldwin *et al.*, 2003); the structure of Cre bound to the HJ in the presence of a hexapeptide inhibitor WKHYNY indicated how such an inhibitor can block tyrosine recombination (Ghosh *et al.,* 2005); a structure of Cre-*loxP*-vanadate transition state mimic allowed for better understanding of the role of catalytic residues in stabilizing the transition states in the recombination reaction (Gibb *et al.*, 2010).

Together, the numerous crystal structures of the Cre recombinase bound to a variety of recombination substrates and intermediates made Cre the model member of the tyrosine recombinase family. The dogmas of tyrosine recombination such as site-specific binding, half-of-the-sites reactivity, and the requirement for homology within the central region that were derived from these structures are presumed to hold true for most members of the tyrosine recombinase family.

4.1.2.2 λ integrase

The first structure of λ integrase (1.9 Å) contained only 170 residues of its catalytic domain (Figure 4-5a), yet as it was the first published structure of any tyrosine recombinase it provided useful insights into its mechanism (Kwon *et al.,* 1997). The structure revealed that the C-terminal domain consists of seven α-helices and seven β-strands and showed how the invariant residues of tyrosine recombinases (Arg-His-Arg) cluster together at the proposed DNA interaction face of the protein. However, in this structure the catalytic tyrosine residue was 20 Å away from this active site assembly, which presented a question of whether λ integrase cleaves the DNA in *cis* (with all catalytic residues contributed by one monomer) or in *trans* (with the catalytic tyrosine donated by one monomer and the remaining catalytic residues provided by the other).

In order to solve the structure of the λ integrase bound to its recombination site *att*, unique half-site suicide substrates were used. This resulted in a crystal structure of one λ integrase subunit bound to one arm of the *att* site solved at 2.95 Å resolution (Figure 4-5b; Aihara *et al*., 2003). The complex had to be trapped this way as the binding of λ integrase to its core site is weak and the intermediates formed with the longer, full-site suicide substrates (see Figure 3-14) were immediately hydrolysed back to substrates. The hairpin turn in the DNA prevented two λ integrase monomers from forming a dimer, resulting in this unusual monomeric structure.

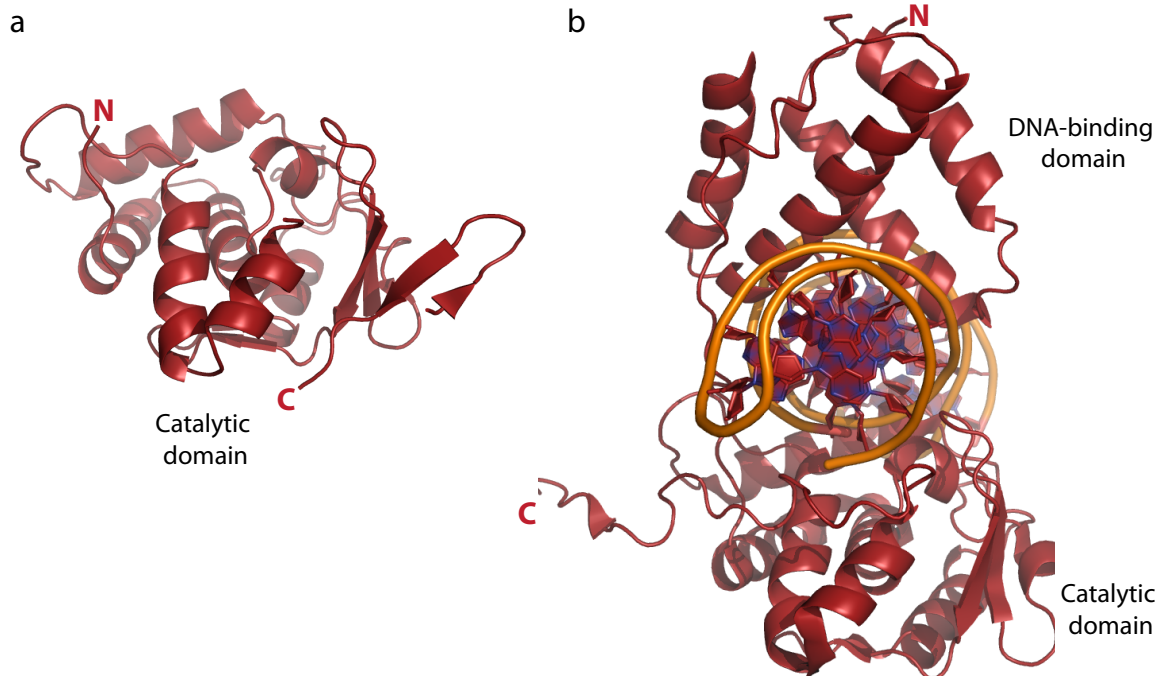


Figure 4-5: The cartoon representations of the first crystal structures of λ integrase. a) The structure of the isolated catalytic domain of λ integrase (PDB accession number 1AE9; Kwon *et al.,* 1997). b) Structure of a λ integrase monomer bound to an *att* half-site suicide substrate (PDB 1P7D; Aihara *et al*., 2003). The DNA is shown in orange.

This structure of λ integrase included residues 75-356, comprising the core binding and catalytic domains, which corresponds to the full-length Cre recombinase. The N-terminal, arm-binding domain was not included in the crystallization construct here. Similarly to Cre, λ integrase forms a C-shaped clamp around the DNA. The core-binding domain consists of two pairs of antiparallel α-helices that form a four-helix bundle inserted deep into the major groove of DNA. The catalytic domain contacts both the major and minor grooves on the other side of the DNA. The DNA is smoothly bent around the point where the catalytic domain is inserted into the major groove of the DNA.

Interestingly, in comparison to the DNA-free catalytic domain structure (Kwon *et al.,* 1997), the catalytic tyrosine Tyr342 has moved approximately 20 Å into the active site, which in this structure is fully assembled. This is a result of repacking of the final C-terminal residues 331-356, suggesting that large conformational changes in this region are associated with DNA binding. In contrast to Cre, in which this segment consists of two α-helices, in λ integrase this segment is unstructured. Eight terminal residues (349-356) extend from the body of the protein into the solvent. It was proposed that, similarly to Cre, this C-terminal segment might be protruding into the body of the neighbouring monomer and this exchange of C-terminal tails might be important for regulation of DNA cleavage. Due to lack of a multimeric assembly, it was again not possible to conclude from this structure whether λ integrase cleaves DNA in *cis* or in *trans*.

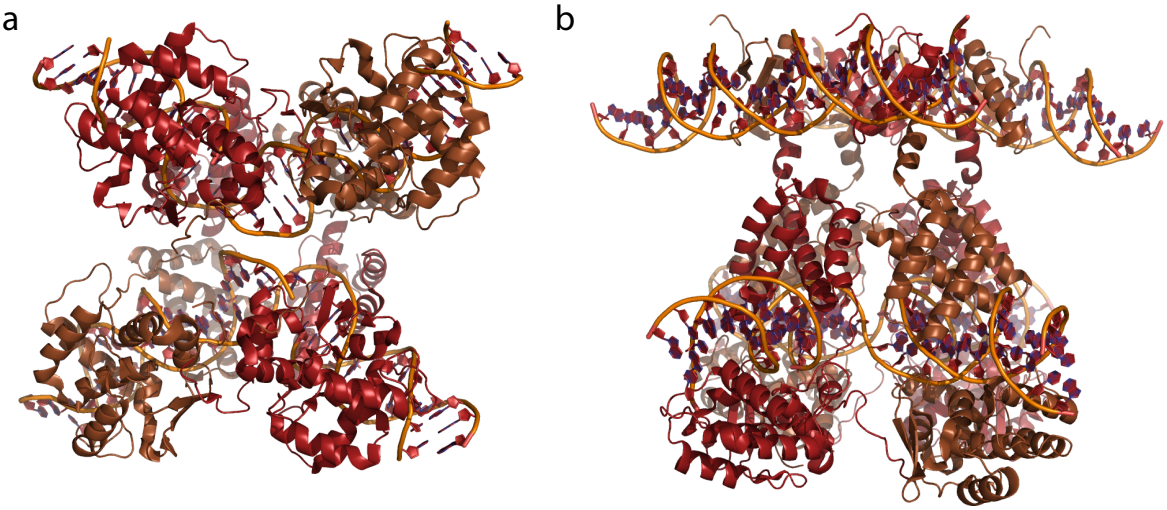


Figure 4-6: Cartoon representation of the crystal structures of the λ integrase synaptic complex. The two λ integrase monomers bound to the same *attP* site (shown in orange) are shown in two different colours. a) Tetrameric view of the structure of the λ integrase post-cleavage synaptic complex (PDB accession number 1Z19; Biswas *et al.,* 2005). b) The structure of the λ integrase-*attP* post-strand exchange synaptic complex with the N-terminal arm-binding domain bound to accessory arm DNA sequences (PDB 1Z1B; Biswas *et al*., 2005).

The most recent set of λ integrase-*att* complex structures (Figure 4-6) includes structures of the tetrameric post-cleavage synaptic complex trapped in a complex with a vanadate derivative of the *attP* site (2.8 Å), post-strand exchange synaptic complex (3.8 Å), and the HJ intermediate (4.4 Å; Biswas *et al.*, 2005). In the first two structures, two monomers bound to one *attP* site are related by two-fold crystallographic symmetry to another pair of monomers bound to another *attP*. The third structure reveals a tetramer bound to the HJ intermediate. As for Cre, in all structures the tetramers form a cyclic arrangement, but here

the swapped C-terminal segment forms a β-strand rather than an α-helix. Unlike in the Cre structures, no pseudo-four-fold symmetry is observed in the post-cleavage synaptic complex structure, but rather the two pairs of monomers bound to one *attP* site are in a two-fold symmetric, skewed conformation with respect to each other. This is not the case for the structures of the λ integrase post-strand exchange complex and the HJ intermediate, which do show pseudo-four-fold symmetry.

In all of the structures, the core-binding and catalytic domains of λ integrase form a clamp around the core *attP* site DNA. The extended C-terminal segment (residues 350-356) packs in *trans* against the neighbouring subunit in a cyclic arrangement as seen for Cre. The skewed packing of the protein subunits bound to the same *attP* site in the post-cleavage complex results in two different subunit conformations, containing active or inactive catalytic sites. The catalytic sites resemble those of other tyrosine recombinases and comprise the nucleophilic Tyr342 and the catalytic pentad of Arg212, Lys235, His308, Arg311, and His333. In these structures, it can be seen unambiguously that in λ integrase recombination all catalytic residues are donated in *cis* by one monomer.

The DNA in the post-cleavage synaptic complex is bent asymmetrically in the central region, which is caused by a kink in the non-cleaved DNA strand 2 bp from the cleavage site. Here, two adenine residues are unstacked with a 90° tilt, resulting in the kink and leading to an overall 109° bend in the DNA. This sharp kink was proposed to facilitate strand exchange and prevent strand re-ligation back to substrates. Interestingly, in the post-strand exchange synaptic complex (a step before strand ligation and formation of the HJ) the kink has shifted to a more central location 4 bp from the cleavage site, resulting in an overall bend angle of 83° and a pseudo-four-fold symmetric arrangement of this complex.

The post-strand exchange synaptic complex and the HJ intermediate structures contain the full-length protein, including the very N-terminal arm-binding domain. In these structures, the three protein domains form three distinct layers, with the core-binding domain bound to the core *attP* site and the arm-binding domain bound to short *attP* arm sequences. It was proposed that the binding of the *attP* arm is required for isomerization of the HJ and therefore for completion of the recombination reaction (Biswas *et al.,* 2005).

The three structures of the λ integrase synaptic complex revealed three distinct snapshots of the recombination pathway and together explained the mechanism that allows

progression of the reaction from substrates to products in λ integrase recombination: the binding to the arm sites drives the first strand exchange and leads to isomerization of the HJ into the conformation in which the second pair of strand cleavage and exchanges can be performed.

4.1.2.3 Flp

The eukaryotic recombinase Flp differs from the bacterial tyrosine recombinases, which was emphasised by the crystal structures of Flp bound to the *FRT* HJ intermediate (Chen *et al.,* 2000; Conway *et al.,* 2003). The first structure (2.65 Å resolution; Figure 4-7a) was trapped using symmetrized *FRT* suicide substrates with a nick one nucleotide away from the cleavage site; however, unlike in the Cre and λ integrase structures where the post-cleavage synaptic complex was trapped, here the reaction had proceeded through the strand exchange step to form an unligated HJ intermediate (Chen *et al.,* 2000). The structure contains the whole tetramer in an asymmetric unit, with a roughly square planar conformation of the tetramer. The synaptic complex of Flp appears to be more flexible than that of Cre, which could explain the observation that Flp can efficiently recombine *FRT* sites with variable (7-9 bp) central regions (Senecoff and Cox, 1986).

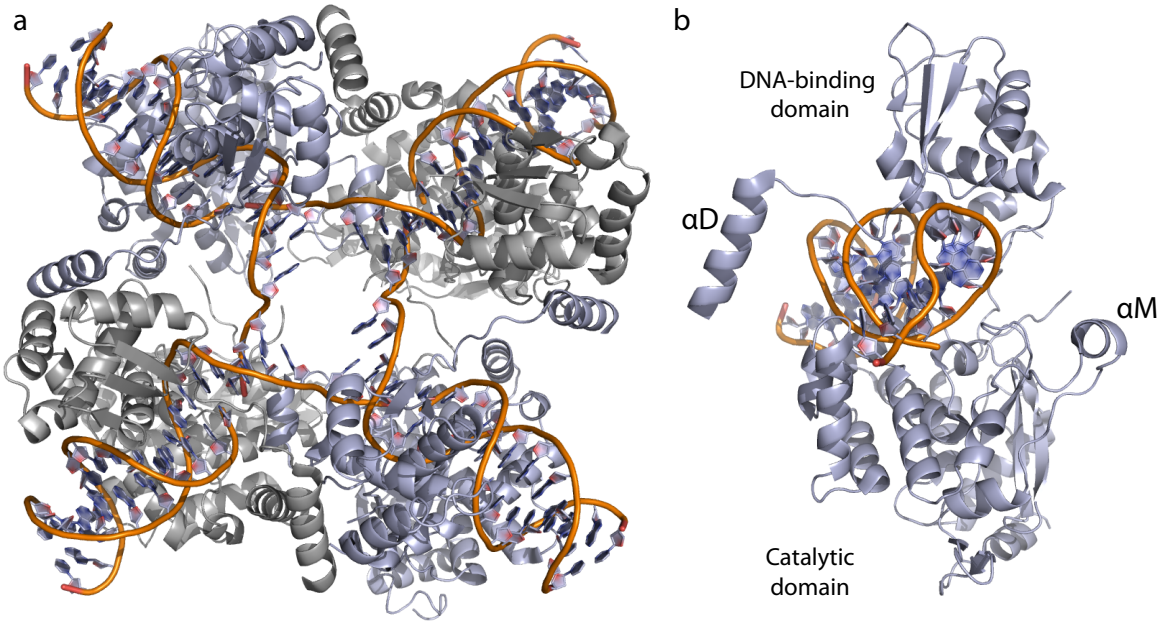


Figure 4-7: The cartoon representation of the crystal structure of a Flp-*FRT* HJ intermediate (PDB accession number 1FLO; Chen *et al.,* 2000). The DNA is shown in orange. a) Tetrameric view of the structure. Flp monomers bound to the same *FRT* site are shown in two different colours. b) A monomer of Flp forming a C-shaped clamp around DNA. The two α-helices (αD and αM) packing in *trans* against the neighbouring monomer are marked.

Each monomer is composed of two compact domains and two protruding segments comprising helices αD and αM that pack in *trans* arrangement against the two neighbouring monomers (Figure 4-7b). Helix αM contains the catalytic tyrosine Tyr343, which is therefore donated in *trans* to the active site of the neighbouring monomer. Helix αM is connected to the helix αN, which returns the polypeptide chain to the original monomer. The C-terminal catalytic domain is structurally homologous to those of Cre and λ integrase. However, Flp contains additional C-terminal sequences not found in other tyrosine recombinases, and its N-terminal domain is structurally unrelated to other tyrosine recombinases.

As for Cre and λ integrase, two distinct monomer conformations were observed in the structure of Flp: the active conformation, where the Tyr343 is in line for attack of the scissile phosphate, and the inactive conformation, where the Tyr343 is placed 10 Å away from the cleavage site. The remaining residues of the catalytic site, Arg191, K223, His305, Arg308, and Trp330 show similar conformations in both monomers. Additionally, it was proposed that the fairly well conserved residue His345 plays a role in orienting the catalytic tyrosine.

Flp interacts with the 13-bp arm of the *FRT* site mostly through contacts to the DNA sugar-phosphate backbone, with only a small number of base-specific interactions. An ATT sequence within the arm displays an unusually narrow minor groove which was proposed to promote indirect sequence readout by Flp. Flp interacts only non-specifically with the central region of *FRT*, and apart from the central adenine, all nucleotides of the 7-nt central region are basepaired. The DNA contacts are similar in all four Flp monomers and introduce a 24° bend to each binding arm. The major bend is contained within the central region, creating a pseudo-four-fold symmetric complex.

The second structure of Flp synaptic complex (3 Å) consisted of the same DNA substrate as in the first structure, but in complex with a hyperactive mutant of Flp, termed Flpe (Conway *et al.*, 2003). This structure explored the role of the two interfaces formed between Flp monomers: interface I between the monomers bound to the same *FRT* site, and interface II between the monomers bound to the different *FRT* sites. Interface I is compact and inflexible, and its presence is a prerequisite for activity. The arrangement of the monomers at interface II is more flexible with the monomers spaced further apart. Consequently, the protruding helix αM does not reach into the active site of the

neighbouring monomer, rendering the monomer inactive. Additionally, it was shown that despite the fact that only interface I is required for Flp cleavage activity, interface II must be present and the interfaces have to be circularly permuted to allow for the assembly of the tetrameric synaptic complex, otherwise the four monomers cannot be accommodated due to clashes. This provides an elegant way to regulate the order of cleavage activity that is similar compared to Cre.

Taken together, the structural studies of Flp recombinase showed that Flp shares the general complex assembly and mechanism of recombination with other tyrosine recombinases, employing a similar helix-swapping strategy to ensure the half-of-the-sites reactivity. The pseudo-four-fold symmetry of the HJ intermediate suggests that this might be the favourable conformation for efficient strand exchange. The main difference from other tyrosine recombinases is that Flp cleaves DNA with catalytic site assembly in *trans*, but the biological significance of this mechanism could not be determined from the structural studies of Flp.

4.1.2.4 HP1 integrase

The structure of the C-terminal catalytic domain (residues 165-337) of another phage recombinase, the HP1 integrase, has been solved at 2.7 Å resolution (Figure 4-8, Hickman *et al.,* 1997). The structure consists of four monomers per asymmetric unit, two of which share a more extended dimer interface in which the C-terminal tail (residues Ser321-Gln337) extends away from the body of one monomer and docks into a cleft of the second monomer. These interactions are similar to the helix swapping in Cre, but they are reciprocal within the dimer and therefore could not accommodate the same cyclic, tetrameric arrangement. Such a dimeric arrangement could be physiologically relevant, but it could also be a crystallographic artefact that does not occur in the normal recombination pathway. Additionally, the structure showed that all catalytic residues assembled in a compact active site around a sulphate ion (mimicking phosphate of the DNA backbone) came from one monomer, suggesting *cis* cleavage by the HP1 integrase.

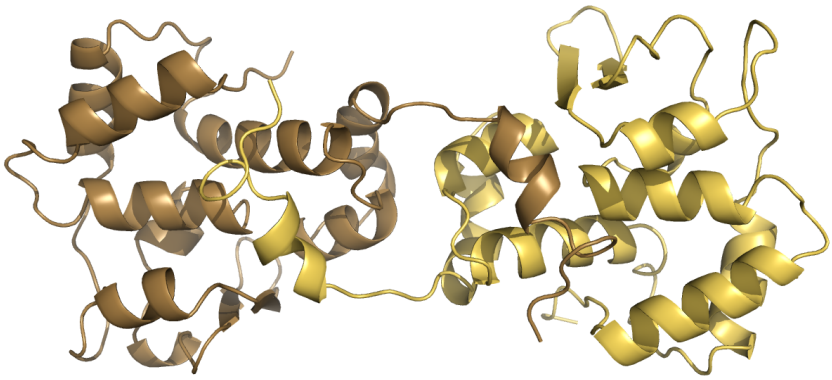


Figure 4-8: The cartoon representation of the structure of the isolated catalytic domain of HP1 integrase (PDB accession number 1AIH; Hickman *et al.,* 1997). The two monomers, related to each other by non-crystallographic symmetry, are shown in two different colours.

4.1.3 Crystal structures of Xer recombinases

To date, two crystal structures of Xer recombinases have been determined: one of the full-length XerD from the *E. coli* XerC/D recombination system (Subramanya *et al.,* 1997) and one of the full-length archaeal XerA from *P. abyssi* (Serre *et al.,* 2013). Both structures have contributed to confirming that Xer recombinases share a common fold and are likely to employ mechanisms equivalent to those of other members of the tyrosine recombinase family.

4.1.3.1 XerD

The crystal structure of XerD (2.5 Å resolution; Figure 4-9) revealed a single monomer of XerD showing 271 out of total 298 amino acid residues built in the final structure model (Subramanya *et al.,* 1997). The N-terminal DNA-binding domain (residues 1-107) consists of four α-helices arranged in two parallel helix hairpins with a 90° angle between them. The C-terminal catalytic domain (residues 108-298) is mainly α-helical with a three-stranded β-sheet, and is structurally homologous to the catalytic domains of Cre, λ integrase, and HP1 integrase.

The very C-terminal segment of the protein (residues 271-298) forms a turn followed by a long α-helix, αN, which contains the catalytic tyrosine Tyr279. The positions of catalytic residues (Arg148, Arg247, and Tyr279) and the electrostatic surface potential calculations revealed a potential DNA binding site on the catalytic domain. However, the observed

compact positioning of the two XerD domains would result in a clash if DNA was modelled into this structure, suggesting that large domain rearrangements are required for DNA binding by XerD. In the model of the isolated catalytic domain bound to DNA, the catalytic residues Arg148 and Arg247 are assembled around the scissile phosphate and only simple side chain rotation is required to position Tyr279 for the nucleophilic attack. The structure suggested that all catalytic residues are provided in *cis* in XerD recombination. Additionally, two lysine residues (Lys172 and Lys175) positioned at the turn of the β2-β3 hairpin were implicated to contact the DNA sugar-phosphate backbone near the scissile phosphate. This region is structurally conserved in tyrosine recombinases and was also proposed to contribute to the regulation of Cre (Martin *et al*., 2002).

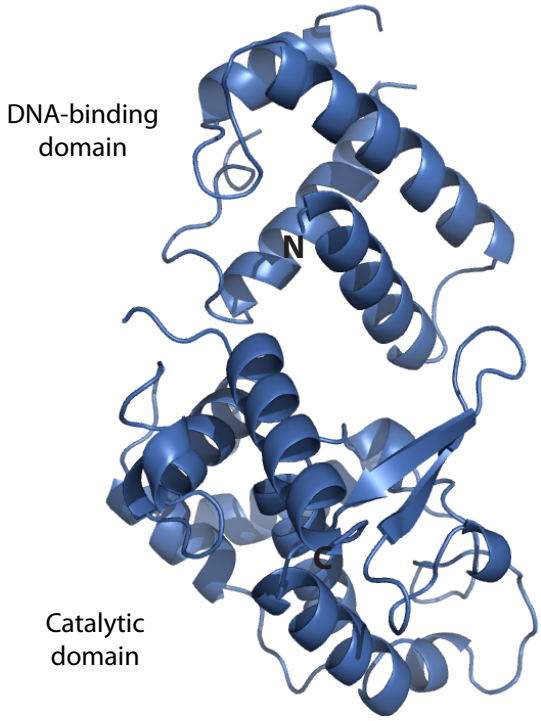


Figure 4-9: Cartoon representation of the crystal structure of the XerD monomer (PDB accession number 1A0P; Subramanya *et al.,* 1997).

4.1.3.2 XerA

The crystal structure of XerA recombinase from *P. abyssi* solved at 3.0 Å resolution (Figure 4-10) revealed a single monomer of XerA in the asymmetric unit (Serre *et al.,* 2013). Inspection of the crystal packing contacts showed a large interface between two monomers related by crystallographic symmetry, which could represent a true dimer

interface as XerA dimers could be also shown to exist in solution by the small-angle X-ray scattering (SAXS) method (Serre *et al.*, 2013).

Each monomer consists of two well-separated domains: the N-terminal, DNA-binding domain, slightly resembling the core-binding domain of λ integrase, and the C-terminal, catalytic domain. The domains are arranged in an open conformation as opposed to the structure of XerD where the two domains are tightly connected (Subramanya *et al.,* 1997). The N-terminal domain of XerA consists of two hairpins composed of two antiparallel α-helices each, while the C-terminal domain is mainly α-helical apart from a three-stranded β-sheet, as seen for other tyrosine recombinases. Within the proposed dimer, the two catalytic domains are related to each other by 180° as opposed to 90° observed for Cre recombinase (Guo *et al.,* 1997). Additionally, the C-terminal segment containing α-helix αN protrudes away from the body of the protein into a cleft of the neighbouring monomer. As in the case of HP1, this arrangement does not allow for formation of a cyclic tetrameric synaptic complex.

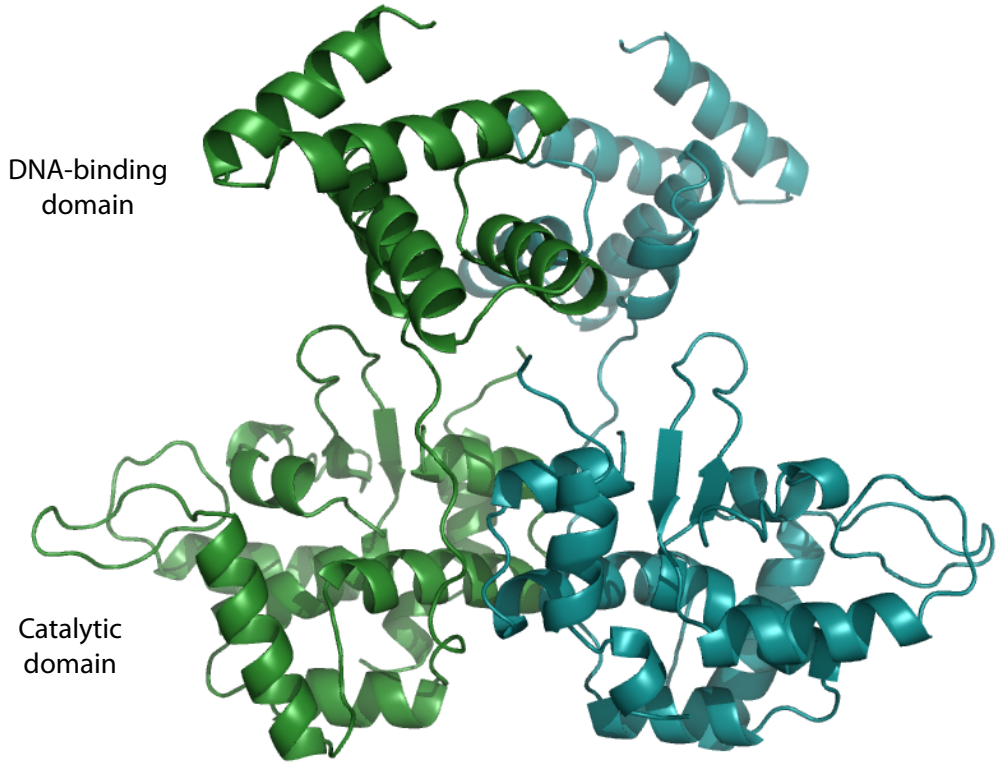


Figure 4-10: Cartoon representation of the crystal structure of XerA (PDB accession number 4A8E; Serre *et al.,* 2013). The two monomers, related to each other by crystallographic symmetry, are shown in two different colours.

The catalytic residues (Arg135, His226) form hydrogen bonds with a sulphate ion, which likely mimics the phosphate of the DNA backbone. However, the catalytic tyrosine Tyr261 is far away from this active site, although a simple rotation of the helix αM containing the Tyr261 was proposed to be sufficient for placing Tyr261 in the catalytic site. Because the helix αM carrying the Tyr261 was detached from the main body of the protein due to missing electron density, it was not clear from this structure whether the tyrosine is donated in *cis* or in *trans*; however, the complementary SAXS and mutational analysis support the idea that cleavage occurs in *cis* conformation.

4.1.4 Structural questions concerning Xer recombination

While both structures of Xer recombinases XerD and XerA contributed to the understanding of Xer recombination, at the same time many questions about the structure and workings of the Xer synaptic complex remained unanswered. This is mainly caused by the fact that both structures represent a DNA-free Xer recombinase and could not address questions regarding binding and recombination of the DNA substrates.

A DNA substrate could not be modelled directly into either Xer recombinase structure. The XerD structure displayed a closed conformation, with the positively charged cleft of the catalytic domain that could comprise the DNA binding interface blocked by the proximity of the N-terminal domain (Subramanya *et al*., 1997). On the other hand, while the XerA structure displayed a more open conformation with the two domains loosely connected, still no DNA would fit into the inter-domain cleft without protein-DNA clashes. This suggests that DNA binding requires large conformational changes in both cases (and likely in the case of other Xer recombinases) but the exact nature of such changes and the organization of the Xer synaptic complex is so far not known. While for both XerD and XerA the proposed cleavage mechanism included assembly of the catalytic site in *cis*, it is not impossible that the large conformational changes required for DNA binding could remodel the catalytic site and change the nature of the cleavage mechanism.

Finally, the mechanism of half-of-the-sites reactivity, which in other tyrosine recombinases such as Cre, λ integrase, and Flp, has been exerted by the cyclic exchange of the C-terminal segments within the synaptic complex, could not be seen in the available

structures of XerD and XerA, raising a possibility that this subfamily of tyrosine recombinases might employ a different mechanism of recombination regulation.

4.1.5 Aims and objectives

A great amount of knowledge about tyrosine recombination has been acquired from the crystal structures of recombination complexes of various recombinases bound to their specific DNA substrates. At the same time, no such complex structures are available for the Xer recombinases. The two protein-only structures of XerD and XerA contributed to understanding of Xer recombination, but left many questions unanswered regarding the functional recombination complex assembly. To fill this knowledge gap, the objective of this study is to crystallize XerH recombinase (the subject of this work) in complex with its DNA substrate, dif_H, taking advantage of what has been learned from previous tyrosine recombinase structures: namely that symmetrized DNA substrates support crystallization of synaptic complexes, and that suicide substrates are an efficient way to trap post-cleavage intermediates of the recombination reaction. From the obtained structure, I would like to elucidate the mechanism of XerH recombination, putting it in the context of mechanistic insights obtained for other Xer and tyrosine recombinases, and in the context of the biochemical data discussed in Chapter 3.

4.2 Results

4.2.1 Crystallization of the XerH-dif_H-LP complex

Biochemical experiments showed that XerH binds and cleaves left-arm dif_H palindrome (dif_H-LP) substrates with similar efficiency as the native dif_H substrates (see Figure 3-18). It is also known that symmetric DNA substrates facilitate crystallization and higher resolution diffraction compared to asymmetric substrates (as seen for example in structural studies of Cre), likely because the sample is more homogenous and the symmetry supports creation of repetitive lattices and crystal packing. dif_H-LP was therefore chosen as a substrate for crystallization of XerH-DNA complexes. The initial suicide substrate used in crystallization trials was based on the footprinted region of the left dif_H arm (see Figure 3-11) and is shown in Figure 4-11a. After optimization of the protein:DNA ratios required

for a homogenous complex formation in an analytical SEC assay (Figure 4-11b), the crystallization sample contained a 2.2-fold molar excess of XerH over DNA. This sample eluted from the analytical gel filtration column as a sharp peak corresponding to the XerH dimer bound to *dif*$_H$-LP, with a small shoulder likely corresponding to the XerH tetramer bound to two *dif*$_H$-LP sites.

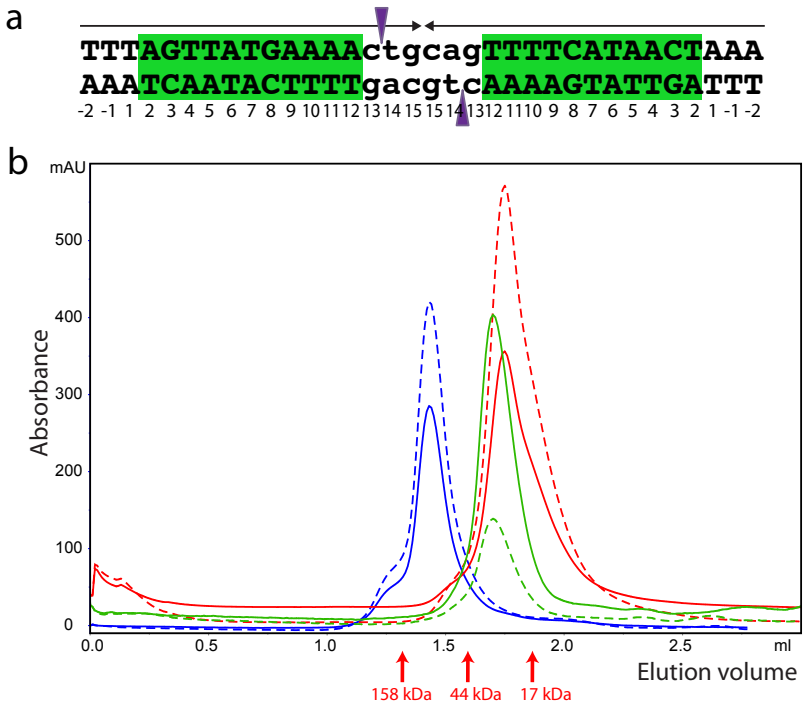


Figure 4-11: XerH-*difH*-LP complex used for crystallization. a) Palindromic DNA suicide substrate used. The two XerH-binding arms are shown in green. The pre-introduced nicks are indicated by purple triangles. b) Analysis of XerH-*dif*$_H$-LP complex formation by SEC on a Superdex 3.2/30 column. The overlaid chromatograms show analytical gel filtration of XerH alone (green), *dif*$_H$-LP DNA alone (red), and the complex of XerH and *dif*$_H$-LP DNA (blue). UV absorbance at 280 nm is shown as a continuous line, absorbance at 260 nm is shown as a dashed line. The elution volumes of molecular weight markers of 17, 44, and 158 kDa on this column are indicated by red arrows.

Initial screening of potential crystallogenic conditions with crystallization screens The Classics Suite (QIAGEN), Index (Hampton Research), and JCSG+ screen (Page *et al.*, 2003) resulted in a single condition (JCSG+, reagent number 65) yielding small, elongated crystals (Figure 4-12a). The identified condition contained 0.1 M HEPES at pH 7.5, 0.2 M $MgCl_2$, and 30% (w/v) PEG 400 as a precipitant. In order to reproduce and improve the crystals, the crystallization conditions were refined by screening around the initial crystallogenic condition in larger drops, while at the same time DNA substrates of different lengths (Figure 4-12b) were tested. The shortest oligonucleotide tested (number 3

in Figure 4-12b) yielded many large crystals with characteristic elongated, butterfly-like split morphology (Figure 4-12c) in all conditions containing 0.2 M $MgCl_2$, 35-38% of PEG 400, and 0.1 M HEPES at pH 6.3-7.3. The crystals were fished from the drops, frozen in liquid nitrogen and transported to the ESRF synchrotron, where the X-ray diffraction experiment was performed.

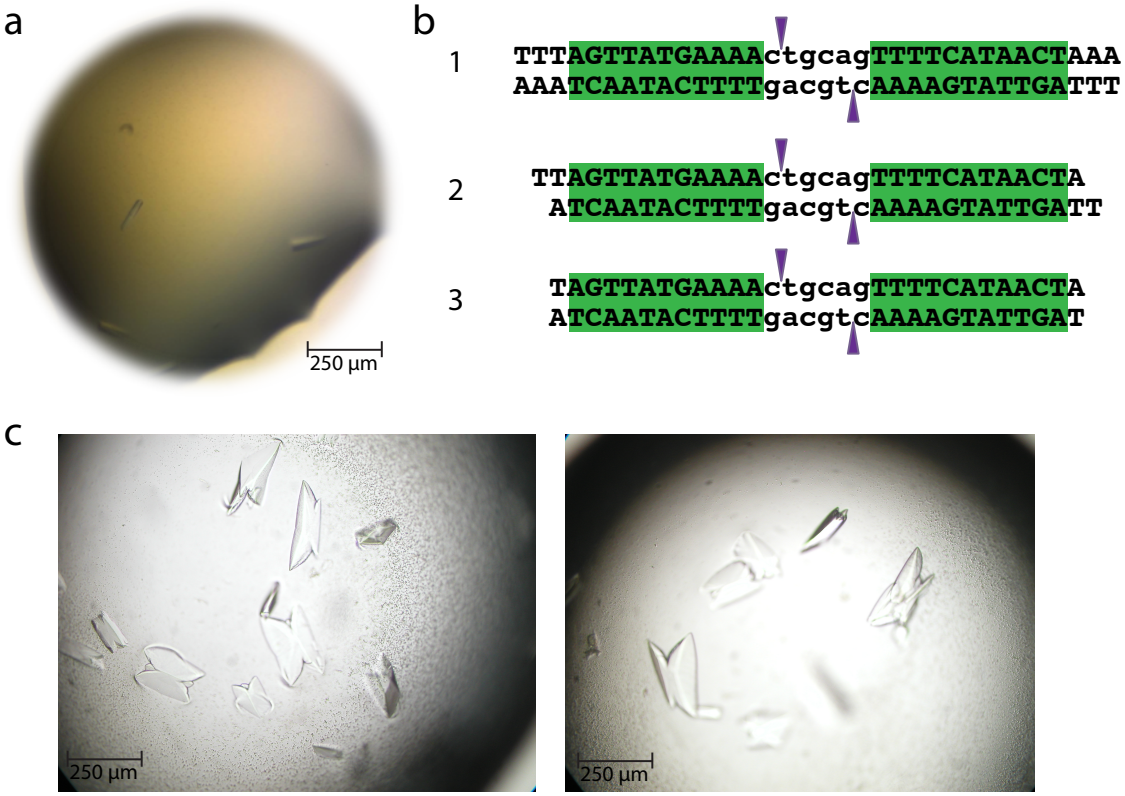


Figure 4-12: Crystallization of XerH-dif_H-LP complexes. a) The initial crystals obtained in a large-scale crystallization screen. b) Suicide-substrate nucleotides of various lengths used for optimization of the initial crystals of XerH-dif_H-LP complex. c) Improved XerH-dif_H-LP crystals obtained after optimization of crystallization conditions.

4.2.2 Data collection and processing

In the X-ray diffraction experiment, the data were collected according to the strategy calculated by the EDNA software (data collection software integrated at the beamlines of ESRF) based on initial 2-4 diffraction images collected at four different crystal rotation positions (*phi* angles) to sample the diffraction potential of the crystal. One crystal (formed in a condition containing 0.1 M HEPES, pH 6.7, 0.2 M $MgCl_2$, and 36% PEG 400) showed

better diffraction than others, estimated by EDNA at 2.55 Å resolution based on the initial two diffraction images (Figure 4-13). According to the EDNA strategy, 2160 diffraction images with 0.1° crystal oscillation angle were collected, resulting in diffraction data from a total of 216° of crystal rotation. The dataset was processed in XDS, resulting in complete data to 2.4 Å resolution (Figure 4-14). XDS processing revealed that the crystals were formed in I222 space group with the cell dimensions of 86.38, 115.22, and 235.2 Å. Data statistics are shown in Figure 4-14. According to the customary $I/\sigma(I) \geq 2$ criterion, the resolution of this dataset should be cut at 2.5 Å resolution. However, a recent study (Karplus and Diederichs, 2012) showed that resolution shells that do not meet this criterion but show strong CC(1/2) values (see section 2.5.5) might hold useful data. Therefore, given the high CC(1/2) value (66.7%, compared to the accepted cut-off of 15%) and high completeness (99.7%) of the 2.4-2.5 Å resolution shell, these data were also included in further processing. The overall completeness of the dataset was 99.9%, with an 8.06-fold redundancy as calculated from the number of unique versus all observed reflections. The R-factors (R-FACTOR, also called R-merge, and R-meas in Figure 4-14) indicate the internal consistency of the data (for example, the degree of similarity between symmetry-equivalent reflections) and are preferred to be as low as possible in high-resolution shells and lower than 10% for the entire dataset and in the lowest resolution shell. Here, both R-factors are below 2.6% in the lowest resolution shell, but rise as the resolution increases to give values over 130% in the highest resolution shell (2.4-2.5 Å). The rather high R-values at high resolution could result from a slight anisotropy of the dataset observed during collection, meaning that diffraction in some directions was stronger than in other directions, which could be caused by the elongated shape of the crystals. However, the overall R-merge of the dataset was estimated at 9.7% and R-meas at 10.3%, which indicates a data set of good quality. As expected, no anomalous signal could be seen in this dataset, as indicated by the low AnomalCorr and SigAno values (Figure 4-14).

The processed reflection intensities obtained as a result of processing in the XDS suite were used as an input for molecular replacement trials in Phaser (McCoy *et al*., 2007) in order to solve the structure of the XerH-dif_H-LP complex. Calculation of the Matthews coefficient (Matthews, 1968) using the CCP4 suite indicated that there are two monomers of XerH each bound to one dif_H-LP half-site per asymmetric unit, and therefore two molecules of XerH were searched for in molecular replacement trials. Various available structures of tyrosine recombinases were used as search models, but no convincing solution could be found in any case, as indicated by low LLG scores and TFZ scores < 8

(Table 4-2; section 2.5.8). This is likely due to the low sequence identity between XerH and the model structures used (highest sequence identity shared with XerD equals 24%) and many differences in the secondary structure of the N-terminal domains within this subset.

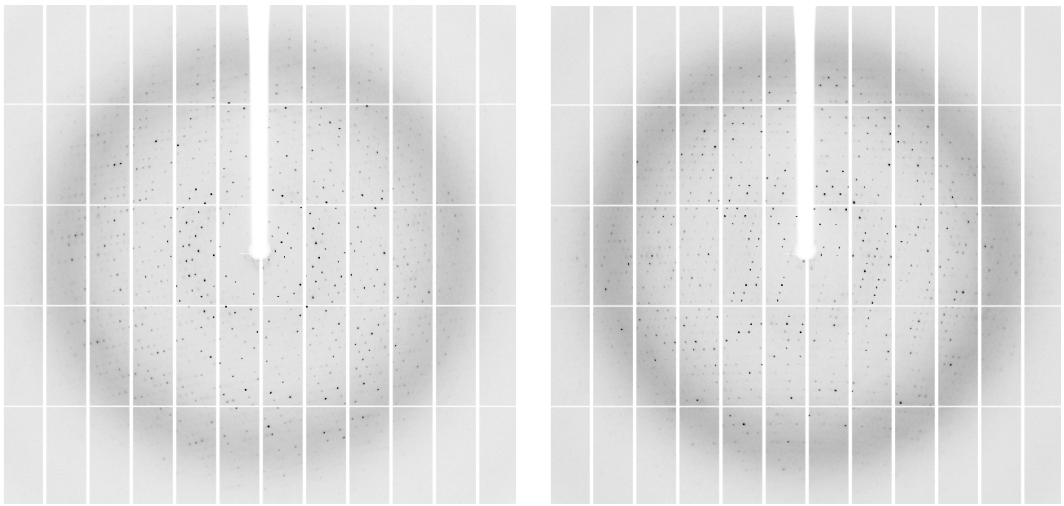


Figure 4-13: Diffraction images used for initial characterization of the XerH-dif_H-LP crystal by the EDNA software. The *phi* angles shown below the images indicate the rotation angle of the crystal during data collection. The edge of the detector corresponds to 2.4 Å resolution.

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

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR expected	COMPARED	I/SIGMA	R-meas	CC(1/2)	Anomal Corr	SigAno	Nano
10.00	4908	720	735	98.0%	2.4%	2.8%	4901	60.98	2.6%	99.9*	0	0.815	488
6.00	18130	2413	2415	99.9%	3.0%	3.3%	18119	49.65	3.2%	100.0*	-1	0.785	1967
4.00	56503	7142	7148	99.9%	4.6%	4.7%	56489	35.95	4.9%	99.9*	-4	0.758	6270
3.60	30714	3717	3719	99.9%	7.3%	7.4%	30712	23.73	7.8%	99.8*	-3	0.770	3364
3.40	20983	2572	2573	100.0%	10.7%	10.6%	20978	17.20	11.4%	99.5*	-3	0.792	2316
3.20	24715	3201	3204	99.9%	15.8%	15.9%	24699	11.66	16.9%	99.0*	-2	0.790	2892
3.10	16085	1973	1973	100.0%	22.0%	21.9%	16083	9.06	23.5%	97.8*	0	0.805	1807
3.00	18212	2195	2195	100.0%	29.0%	29.0%	18209	6.96	30.9%	96.6*	0	0.800	2024
2.90	21339	2542	2542	100.0%	35.6%	35.6%	21336	5.67	37.9%	95.7*	3	0.791	2343
2.80	24309	2870	2870	100.0%	47.5%	47.4%	24307	4.29	50.6%	93.9*	0	0.771	2657
2.70	25076	3299	3305	99.8%	61.3%	60.6%	25060	3.25	65.8%	87.6*	2	0.773	2953
2.60	31254	3868	3870	99.9%	74.2%	73.7%	31249	2.77	79.2%	85.9*	1	0.728	3575
2.50	37295	4493	4494	100.0%	99.2%	98.7%	37291	2.06	105.7%	77.2*	0	0.711	4179
2.40	43159	5233	5250	99.7%	131.4%	130.3%	43144	1.56	140.1%	66.7*	2	0.713	4809
total	372682	46238	46293	99.9%	9.7%	9.8%	372577	14.90	10.3%	99.9*	0	0.760	41644

Figure 4-14: Data statistics for the XerH-dif_H-LP native dataset reported in the CORRECT.LP file, an output of XDS data processing. The data were statistically analysed by resolution shells. I/sigma is a measure of signal-to-noise ratio. R-factor, R-meas and CC(1/2) are statistical indicators of internal data consistency. AnomalCorr and SigAno are indicators of anomalous signal strength.

Table 4-2: Molecular replacement search models and solution statistics obtained from Phaser runs.

Search model[a]	Solutions found	Top LLG score[b]	Top TFZ score[c]
XerD *(10AP)*	104	53	5.6
XerD *(10AP)* trimmed model	59	53	4.5
XerD *(10AP)* poly-alanine model	31	54	5.2
XerD *(10AP)* poly-serine model	60	61	5
XerD *(10AP)* poly-serine loopless model	53	59	5.5
Cre *(1CRX)* dimer with DNA	0	-	-
λ integrase *(1P7D)* monomer with DNA	10	272	5
Ensemble *(1MA7, 1CRX, 1P7D, 1A0P, 2A3V, 4A8E,1Z1B)*	8	131	5.5
XerH homology model generated by Phyre	93	-96	5.9

[a] – PDB accession numbers are shown in brackets.

[b] – LLG score of a correct molecular replacement solution usually shows high, positive value

[c] – TFZ score of a correct molecular replacement solution usually shows value above 8.

4.2.3 Heavy atom derivatization of XerH-*difH*-LP crystals

Since molecular replacement trials were unsuccessful, experimental phasing approaches were employed in order to solve the phase problem and obtain the structure. Initially, the already available native crystals were soaked with heavy atom compounds, mercury-containing PCMBS and iodine-containing "Magic Triangle" I3C, in order to solve the structure with the isomorphous replacement method. However, the diffraction resolution obtained from the derivatized crystals was reduced to 4.9 Å for I3C and 3.6 Å for PCMBS, with weak anomalous signal extending only to 6-7 Å. Therefore, these datasets were not suitable for experimental phasing.

4.2.4 Crystallization, data collection, and processing of Se-Met XerH crystals

To incorporate selenium heavy atoms by replacing methionines with Se-Met, XerH was overexpressed in *E. coli* BL21 (DE3) in a minimal medium containing Se-Met. The expressed protein showed good solubility (Figure 4-15a) and was purified to homogeneity using the same purification protocols as used for purification of the native XerH (Figure 4-15b). In order to confirm that Se-Met was incorporated into expressed XerH, the obtained protein was analysed by mass spectrometry (Figure 4-15c). The most prominent peak corresponded to the mass of 42223 Da, which compared to the calculated mass of native XerH (41939 Da) gave a difference of 284 Da. This difference suggested that exactly six sulphur atoms (32 Da each) were substituted by six selenium atoms (79 Da each), which corresponds to the total number of methionine residues in XerH, suggesting full

incorporation of Se-Met into the purified protein. Se-Met XerH was then used to form complexes with dif_H-LP according to the protocols used previously for the native protein, and the complex was crystallized in the 24-well plates with the refinement screen used for the native complex. This yielded crystals that were slightly smaller than the native crystals, but exhibited the same elongated morphology (Figure 4-16a).

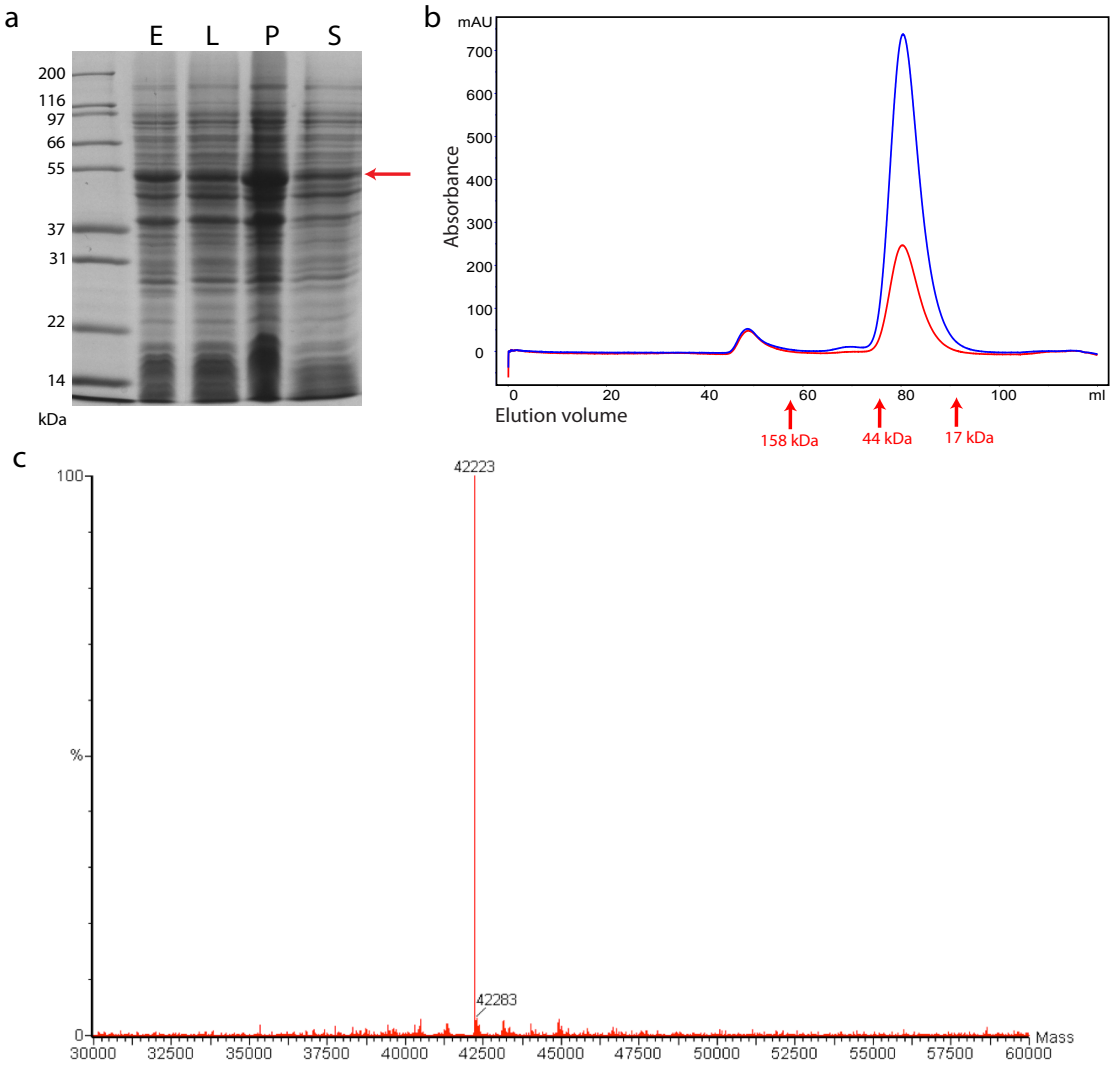


Figure 4-15: Overexpression, purification, and mass spectrometry analysis of the Se-Met XerH derivative a) Expression and solubility of Se-Met XerH analysed on SDS-PAGE. Samples from the expression culture were analysed 15 hours after induction with IPTG (E), after culture harvest and lysis by sonication (L), and after ultra-centrifugation that separated the soluble lysate fraction (supernatant-S) from the insoluble lysate fraction (pellet-P). The position of the expressed XerH protein on the gel is indicated by a red arrow. b) The final step of Se-Met XerH purification: size exclusion chromatography on a gel filtration column. The chromatogram shows the main elution peak that corresponds in size to a monomer of Se-Met XerH (42 kDa). UV absorbance at 280 nm is shown in blue, absorbance at 260 nm is shown in red. The elution volumes of molecular weight markers of 17, 44, and 158 kDa on this column are indicated by red arrows c) Mass spectrum of intact Se-Met XerH obtained by analysis in Q-Tof2 iMass Spectrometer.

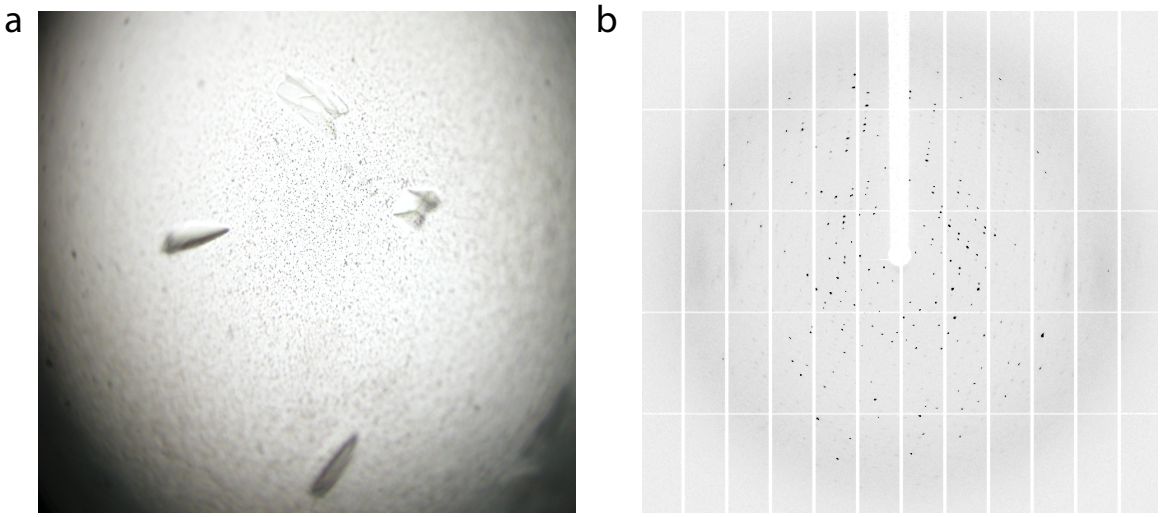


Figure 4-16: Crystallization and data collection of the Se-Met derivative crystals of XerH in complex with dif_H-LP substrate. a) Se-Met XerH crystals. b) Example diffraction pattern obtained from the crystals shown in (a).

The X-ray diffraction data from the selenium derivative crystals were collected at a wavelength of 12.66 KeV (0.979 Å) corresponding to the absorption peak for selenium atoms, according to a data collection strategy designed by EDNA. Unfortunately, due to radiation-induced crystal damage, no further datasets could be collected from the same crystals at wavelengths corresponding to the inflection point and a remote position on the absorption spectrum of selenium, and therefore structure solution by the MAD approach was not possible. An example diffraction pattern from a derivative dataset extending to 3.1 Å resolution, the highest resolution obtained from these crystals, is shown in Figure 4-16b. 2280 diffraction images with 0.05° oscillation were collected, resulting in recorded diffraction data from a total of 114° of crystal rotation.

The selenium derivative dataset was processed in the XDS suite, with final dataset statistics shown in Figure 4-17. The data were cut at 3.1 Å resolution according to the $I/\sigma(I) \geq 2$ cut-off criterion. The dataset showed 98.4% overall completeness with 2.8-fold redundancy as calculated from the number of all observed versus unique reflections. The R-factors in the lowest resolution shell were comparable to those of the native dataset (below 2.6%) and were again relatively high in the highest resolution shell (over 67%), albeit lower than in the native dataset. The overall R-factor of the dataset was estimated at 14.2% and R-meas at 11.6%. Importantly, strong anomalous signal could be seen in this dataset, as indicated by high AnomalCorr and SigAno values, and extended up to 4 Å resolution as judged by the SigAno > 1 criterion. This dataset was therefore suitable for attempting structure solution with anomalous phasing.

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

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR COMPARED expected	I/SIGMA	R-meas	CC(1/2)	Anomal Corr	SigAno	Nano
10.00	3248	1184	1247	94.9%	2.1%	2.4%	3152	37.86	2.6%	99.9*	92*	3.610	415
6.00	12223	4372	4470	97.8%	3.4%	3.5%	11880	24.09	4.2%	99.8*	75*	2.244	1697
5.00	12034	4123	4161	99.1%	6.2%	6.3%	11813	16.42	7.5%	99.4*	57*	1.524	1733
4.70	5181	1996	2027	98.5%	7.1%	7.0%	4935	14.04	8.8%	99.1*	44*	1.288	734
4.50	4547	1609	1626	99.0%	7.8%	7.8%	4437	13.93	9.5%	99.1*	41*	1.245	659
4.40	2774	970	975	99.5%	8.1%	7.8%	2725	14.23	9.9%	98.8*	34*	1.125	415
4.30	2921	1012	1021	99.1%	9.0%	8.9%	2863	12.53	11.0%	98.9*	34*	1.122	428
4.20	3242	1119	1131	98.9%	9.4%	9.8%	3184	11.59	11.4%	99.0*	32*	1.086	474
4.10	3531	1222	1238	98.7%	10.1%	10.2%	3475	11.97	12.3%	98.7*	27*	1.046	525
4.00	4075	1383	1396	99.1%	12.2%	12.1%	4010	10.08	14.8%	98.3*	23*	1.033	599
3.60	19280	7048	7147	98.6%	19.6%	19.2%	18664	6.83	24.1%	95.7*	18*	0.930	2819
3.40	13546	4851	4957	97.9%	31.5%	31.5%	13222	4.36	38.6%	90.6*	10*	0.844	2008
3.30	8179	2854	2911	98.0%	41.1%	40.8%	8033	3.35	50.2%	87.4*	12*	0.806	1228
3.20	9364	3242	3291	98.5%	52.0%	52.6%	9212	2.60	63.6%	77.1*	13*	0.844	1415
3.10	10842	3738	3804	98.3%	67.5%	67.4%	10656	2.07	82.5%	69.1*	6	0.782	1616
total	114987	40723	41402	98.4%	11.6%	11.6%	112261	10.59	14.2%	99.3*	40*	1.199	16765

Figure 4-17: Data statistics table of the XerH-*difH*-LP Se-Met derivative dataset reported in the CORRECT.LP file, output of XDS data processing. The data were statistically analysed by resolution shells. I/sigma is a measure of signal-to-noise ratio. R-factor, R-meas and CC(1/2) are statistical indicators of internal data consistency. AnomalCorr and SigAno are indicators of anomalous signal strength.

4.2.5 Structure solution

The structure of the XerH-dif_H-LP complex was solved using autoSHARP (Vonrhein *et al.,* 2007), which employs an automatic structure solution pipeline. Two datasets were submitted for the single isomorphous replacement with anomalous signal (SIRAS) phasing pipeline: the native dataset extending with 2.4 Å resolution, and the selenium derivative dataset with 3.1 Å resolution. The pipeline attempted to solve the structure with both single isomorphous replacement (SIR) and SAD approaches. The SAD approach resulted in the better outcome, with CC all/weak values (a statistical measure of phase quality) for the initial phasing (based on the anomalous difference Fourier map for two identified heavy atom sites) of 51.82/34.19%, where values above 30% indicate a good SAD solution. The Patterson figure of merit (PATFOM) equalled 11.78, additionally indicating a good SAD solution. The SAD approach was then used further for locating all twelve heavy atom sites in the asymmetric unit (Figure 4-18a). In the final phasing statistics, the figure of merit, a statistical tool for assessing the phase error, showed values above 0.25 (indicating good phases useful for structure solution) at resolutions below 4 Å, which is consistent with the observed anomalous signal extending to 4 Å (Figure 4-17). The obtained phases were then extended onto the native dataset in the density modification step. There, the solvent content within the unit cell has been calculated at 53.9% and phases improved. The initial electron density map (Figure 4-18b), calculated with the resulting phases, showed clear features of a typical protein/DNA chain and was used in the subsequent steps for model building and refinement.

4.2.6 Model building and refinement

A model of two XerH monomers was automatically built into the initial electron density using phenix.autobuild. The obtained model contained no DNA chains but rather polypeptide chains were mistakenly built into the density that clearly corresponded to DNA. These polypeptide chains were manually removed and correctly placed into the protein electron density so that the DNA chains could be built. For building of the DNA, DNA chains from the Cre post-cleavage synaptic complex (PDB: 1CRX; Guo *et al.,* 1997) were placed as an initial model that was then corrected to match the dif_H sequence and further extended by *de-novo* nucleotide building. The model and the electron density map were then improved in 60 cycles of consecutive restrained real-space (manual in *Coot*; Emsley *et al.,* 2010) and reciprocal (automatic in Phenix; Adams *et al.*, 2010) refinement runs. This procedure allowed building of 704 out of 726 amino acids of both XerH monomers and 59 out of 60 DNA bases in the asymmetric unit. The final model statistics are shown in Table 4-3. The obtained R-factor values (R-work of 0.19 and R-free of 0.22) indicate good correlation between the model and the observed data without over-fitting. The stereochemistry of the model was assessed by the fit to the Ramachandran plot, which showed that 97% of non-glycine protein residues lie within the favoured Ramachandran region, with only one outlier in weak density. This, together with the low root mean square (R.m.s) deviations (to be compared with expected values of 0.02 Å for bond length and 2° for angles), indicated good protein stereochemistry. The improvement of the electron density could be seen by comparison of the initial map (Figure 4-19a) with the final map (Figure 4-19b), overlaid with the final model shown in both cases.

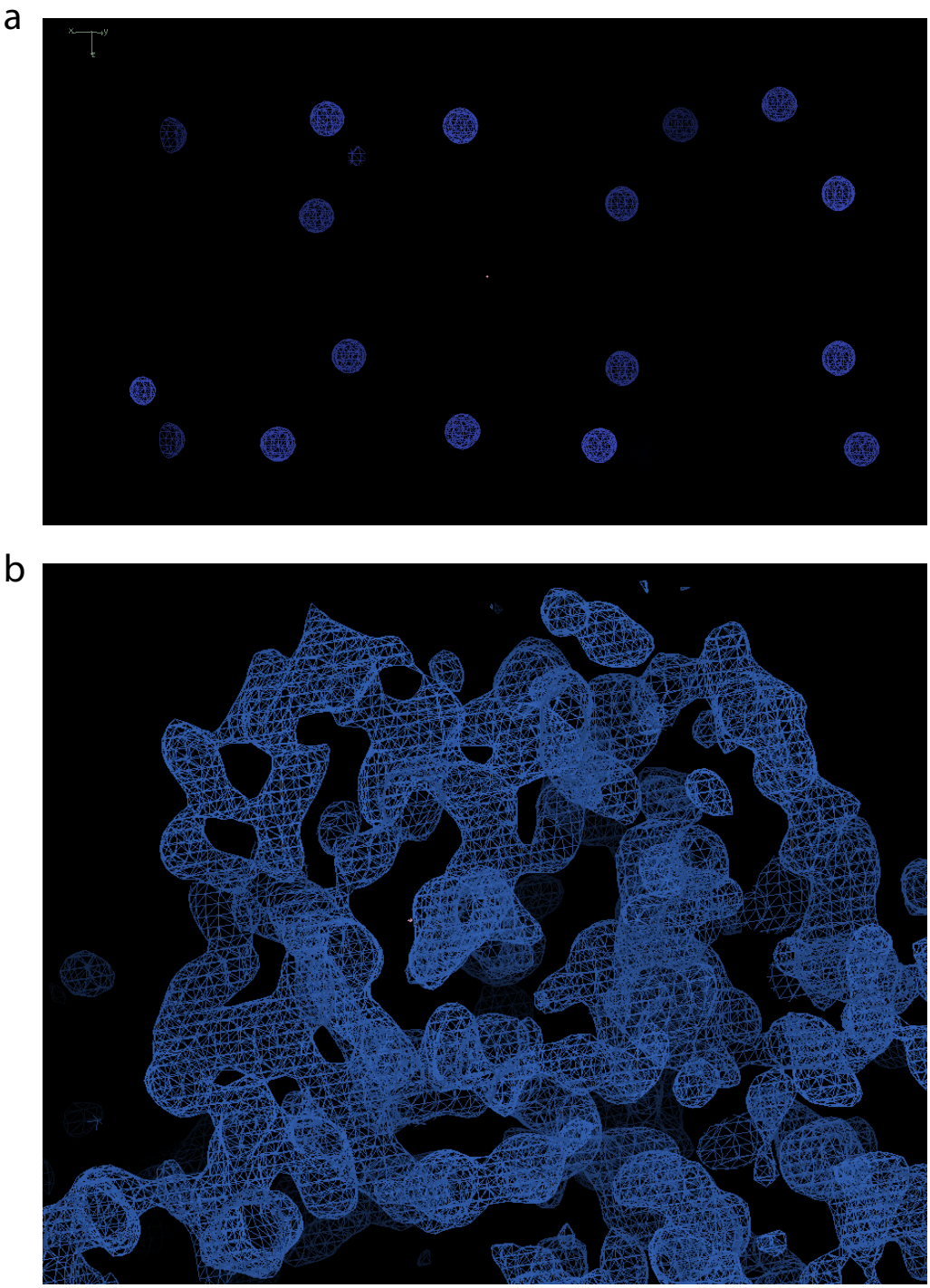
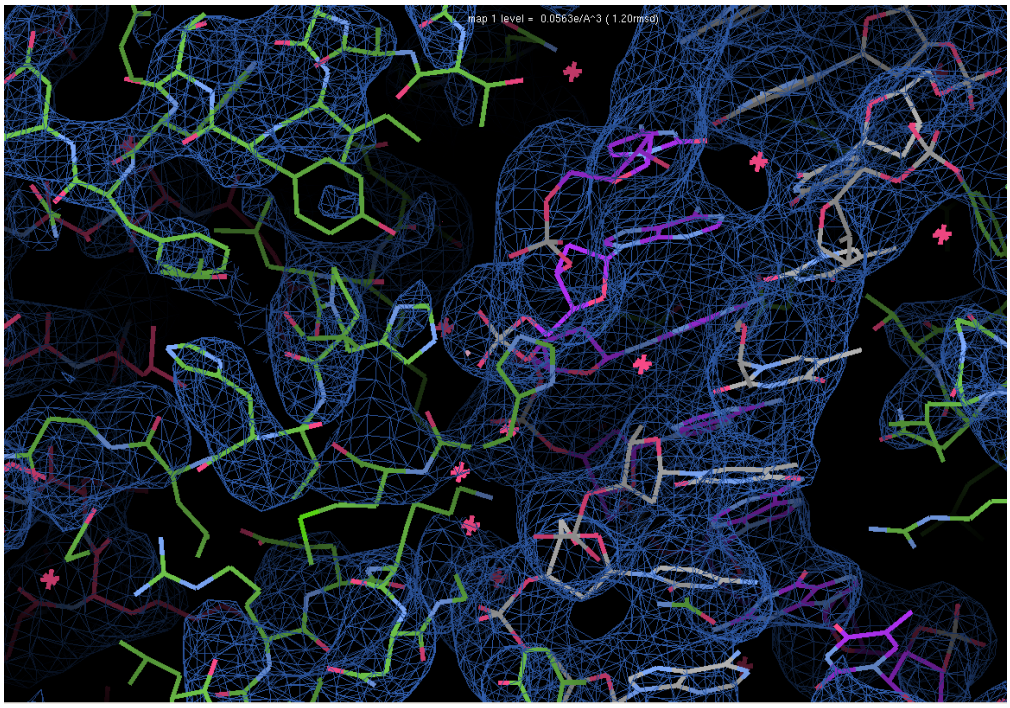


Figure 4-18: Electron density maps obtained from SAD phasing, contoured at 1.2 sigma levels. a) Anomalous difference electron density map of the located selenium atoms. b) Initial electron density map (2Fo-Fc) of the XerH-dif_H-LP complex resulting from SAD phasing.

a



b

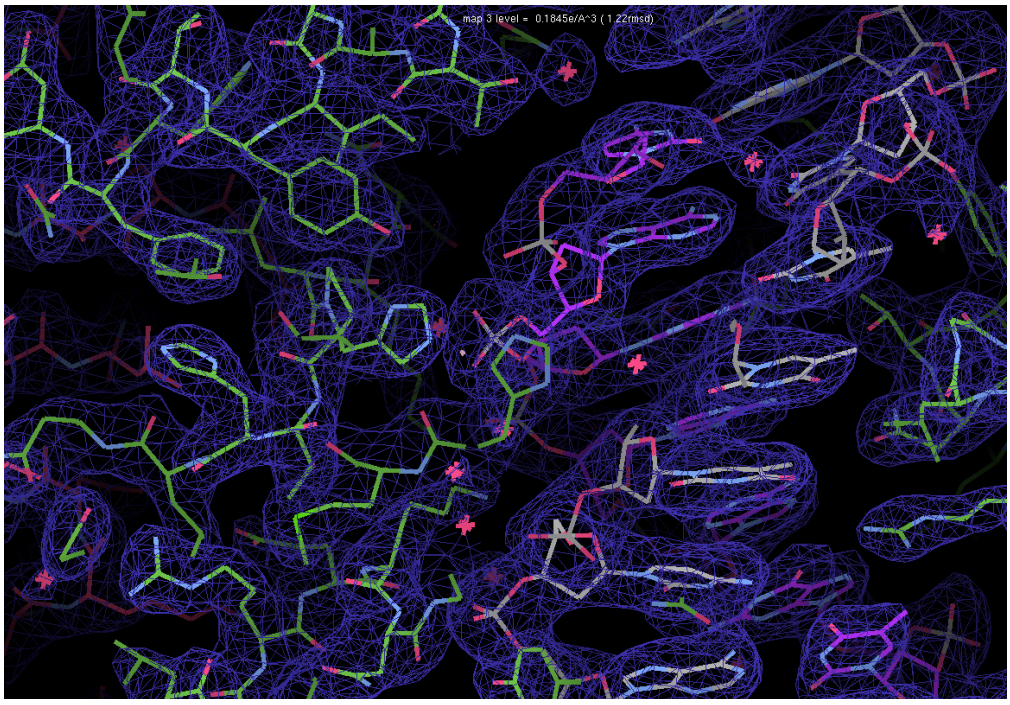


Figure 4-19: Electron density maps (2Fo-Fc) of the XerH-dif_H-LP complex contoured at 1.2 sigma levels, with the final model of the complex shown. a) Initial electron density map calculated with the initial phases from SAD phasing. b) Final electron density map at the end of the refinement.

Table 4-3:Crystallographic "Table 1": Data collection and refinement statistics for the native and derivative datasets.

	Native	Selenium derivative
Data collection		
Space group	I 2 2 2	I 2 2 2
Cell dimensions		
a, *b*, *c* (Å)	86.38, 115.22, 235.2	86.29, 115.54, 236.05
a, b, g (°)	90, 90, 90	90, 90, 90
Wavelength (Å)	0.97908	0.97908
Resolution (Å)	46.96 - 2.4 (2.486 - 2.4)	47.04-3.1 (3.2-3.1)
R-merge	0.096 (1.348)	0.116 (0.675)
R-meas	0.103	0.142
I / sigma (I)	14.90 (1.51)	10.59 (2.07)
CC1/2	0.999 (0.648)	0.993 (0.691)
Completeness (%)	99.90 (99.65)	98.4 (98.3)
Redundancy	8.1 (8.2)	2.8 (2.9)
Refinement		
Resolution (Å)	46.96 - 2.4	
No. reflections	46238	
R_{work} / R_{free}	0.1929 / 0.2234	
No. atoms		
Macromolecule	6841	
Water	97	
B-factors		
Protein	72.1	
Water	55.2	
R.m.s deviations		
Bond lengths (Å)	0.003	
Bond angles (°)	0.62	
Ramachandran favored (%)	97	
Ramachandran outliers (%)	0.29	

R.m.s. – root mean square

4.2.7 Overview of the structure of XerH bound to dif_H-LP

The obtained structure revealed four XerH monomers bound to two pre-nicked dif_H-LP sites in a tetrameric synaptic complex (Figure 4-20). Two monomers bound to one dif_H-LP site are related to the other two monomers bound to the second dif_H-LP site by two-fold rotational crystallographic symmetry. The protein construct used for crystallization consisted of 363 amino acid residues, which includes the N-terminal serine that was present as a result of SenP2 cleavage to remove the 6xHis-SUMO tags. Of those 363 residues, 352 residues of each monomer could be seen in the refined model. The missing residues belong to flexible loops and the inter-domain linker.

Within the tetrameric synaptic complex, the XerH monomers are arranged into a cyclic assembly with C-terminal helices protruding into the body of the adjoining monomer, as

seen for Cre (Guo *et al.,* 1997). The monomers interact with each other through both N-terminal and C-terminal domains, and each monomer interacts extensively with one half-site of the dif_H DNA.

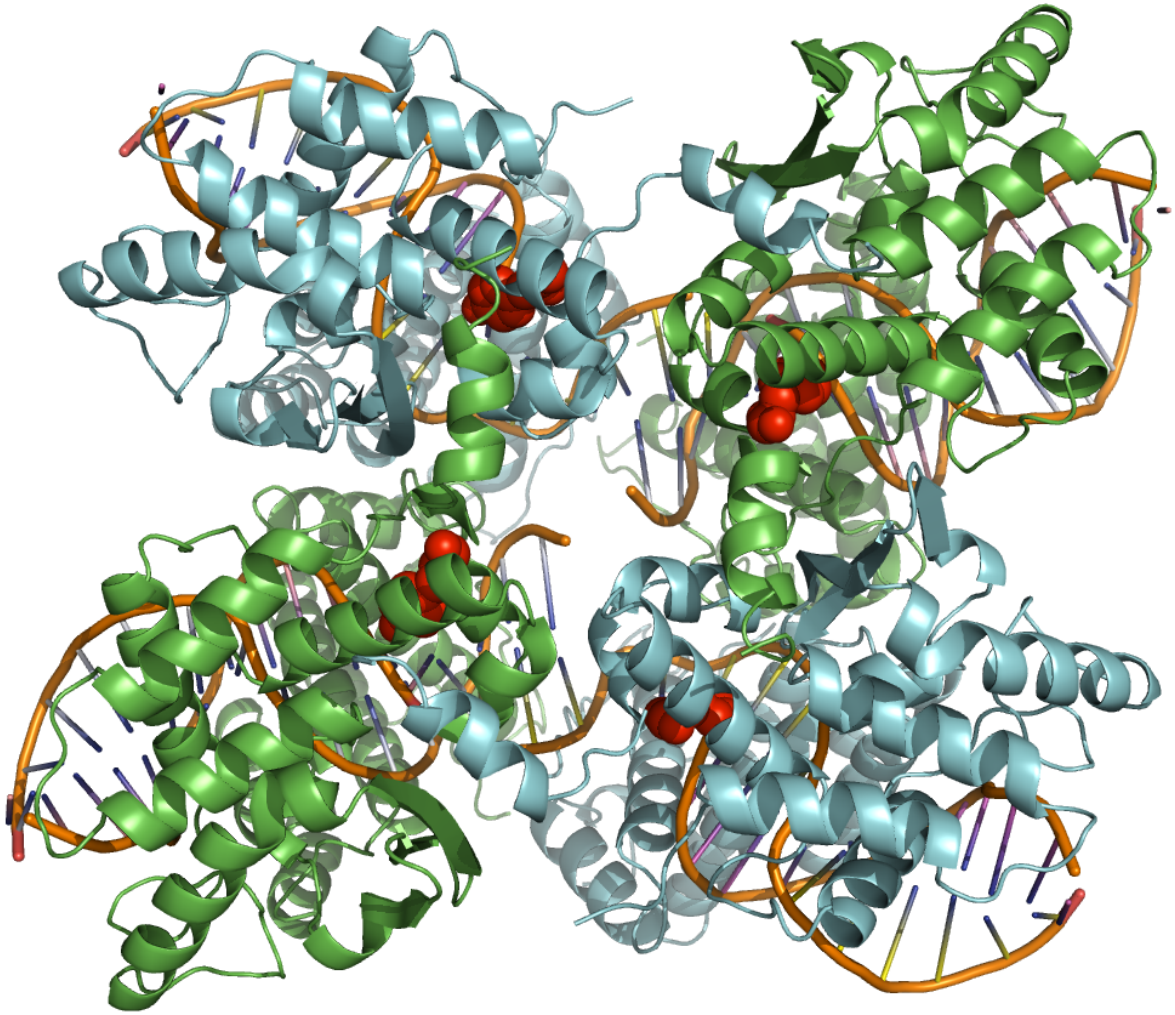


Figure 4-20: Overview of the XerH-dif_H-LP synaptic complex structure, shown in cartoon representation. The two monomers bound to the same dif_H-LP site (shown in orange) are shown in two different colours (green and cyan). The catalytic tyrosine residues (Tyr344) are shown as red spheres.

4.2.8 XerH monomer

Each monomer comprises two distinct domains connected by a flexible linker (Figure 4-21). The N-terminal domain (residues 1-162) is equivalent to the DNA-binding domain of Cre and the core-binding domain of λ Int (Aihara *et al*., 2003; Guo *et al.*, 1997). In XerH, this domain consists of six antiparallel α-helices (αA to αG, Figure 4-21). Comparison of XerH and Cre monomers shows that helices corresponding to XerH αA and αB are not

present in the Cre structure. Additionally, while the Cre DNA-binding domain contains a C-terminal helix immediately adjacent to the inter-domain linker, in XerH that region is disordered. XerD contains only four helices corresponding to XerH αA, αC, αF, and αG. These observations are consistent with the sequence divergence between N-terminal domains of XerH and various tyrosine recombinases (See Figure 3-1). The C-terminal domain (residues 178-362) is mostly helical (αH-αR) and contains a β-sheet of three antiparallel β-strands on one face of the domain (Figure 4-21). C-terminal helices αP and αR protrude from the body of the domain into a cleft of the adjacent monomer. Together, the two domains form a tight C-shaped clamp around the DNA (Figure 4-21), and the position of the DNA within this clamp agrees with the previous surface charge analysis of XerD structure (Subramanya *et al.*, 1997). The structure of the XerH monomer showed R.m.s. deviation of 2.95 Å from the XerH model predicted by Phyre2 (see Figure 3-9) over 254 Cα carbons.

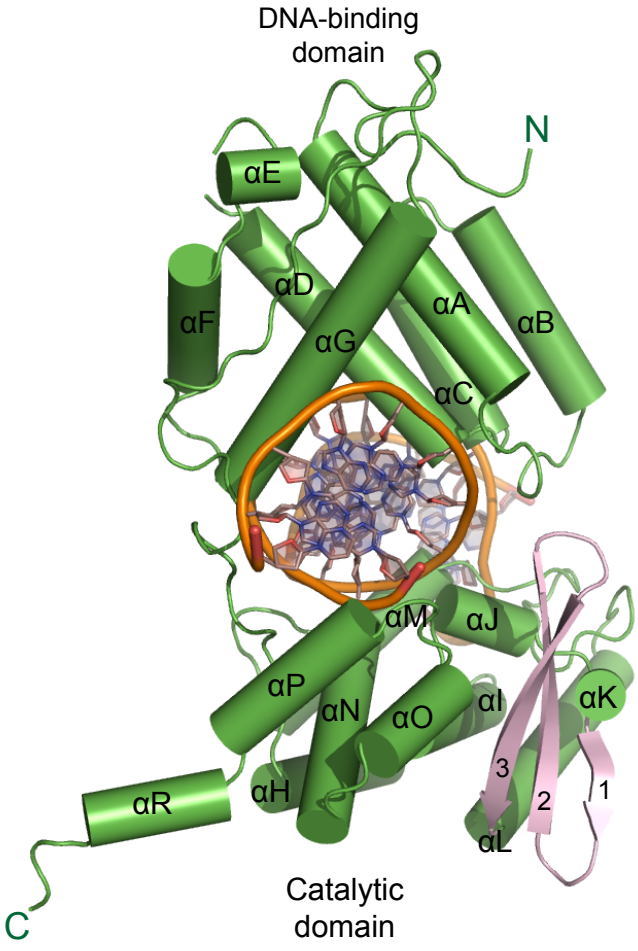


Figure 4-21: Cartoon representation of the structure of the inactive XerH monomer (green) bound to DNA (orange). α-helices are shown as cylinders and their order is indicated by letters. Strands of the β-sheet (pink) are shown as arrows and numbered.

4.2.9 Modelling of DNA binding events in XerH recombination

The comparison of XerH and XerD monomer structures apart from the differences between the structures of the N-terminal domain revealed also a large change in the arrangement of the two domains with respect to each other. In the XerH structure, the two domains form a C-shaped clamp around the DNA, while the two domains of XerD are placed close together with no space that could accommodate the DNA available between them (Figure 4-22). Additionally, the C-terminal helices (αP and αR in XerH) in the two monomers are located differently: in the XerH structure, the helices protrude and dock into a cleft of the neighbouring monomer, while in XerD the helices fold back into an equivalent cleft of their own monomer. This suggests that large rearrangements might be required during DNA binding and synaptic complex formation in Xer recombination.

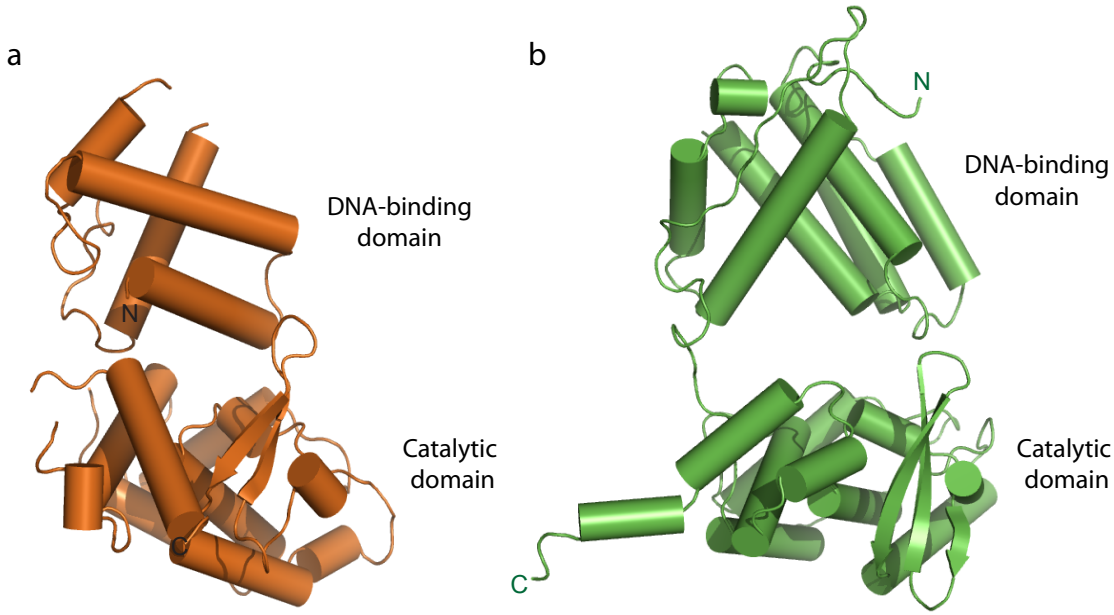


Figure 4-22: Comparison of the XerD and XerH monomer structures, shown in a cartoon representation with α-helices shown as cylinders and β-strands shown as arrows. a) Structure of XerD (Subramanya *et al.*, 1997). b) Structure of XerH.

To date, the protein rearrangements during DNA binding of tyrosine recombinases have been described in detail with regard to catalytic domain rearrangements only, based on comparison between the DNA-bound and DNA-free structures obtained for the C-terminal domain of the λ integrase (Aihara *et al.,* 2003; Kwon *et al.,* 1997). However, the effect of DNA binding on the arrangement of the DNA-binding and catalytic domains remained

unaddressed due to lack of structural information for any full-length recombinase in both DNA-bound and unbound state. Therefore, a model of XerH in its DNA-free form was prepared based on the previously described crystal structure of DNA-free XerD (Figure 4-23a). The major difference between the DNA-free model and the DNA-bound structure is that without DNA the N-terminal and the C-terminal domains do not form a clamp that could accommodate the dif_H DNA; instead, the two domains are contacting each other. However, simple rotation of the N-terminal domain along the flexible linker joining the two domains allows opening of the DNA clamp and potential accommodation of the dif_H DNA (Figure 4-23b). When the open-clamp model and the XerH-DNA complex structure are overlaid, it can be seen that the dif_H-LP arm DNA fits into the cleft of the model without clashes (Figure 4-23c).

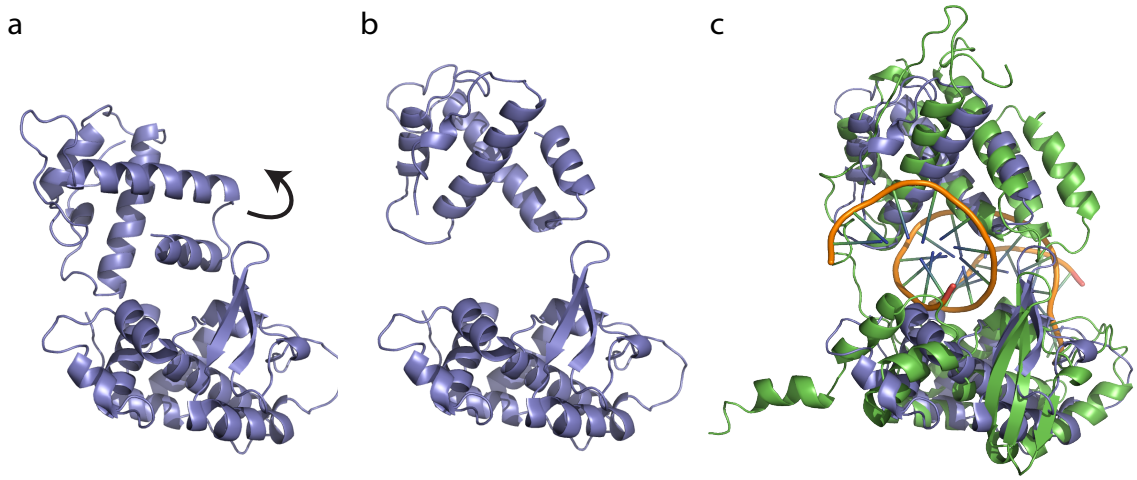


Figure 4-23: Modelling the DNA-binding events in XerH recombination. a) XerH model prepared using SwissModel based on the DNA-free XerD structure (Subramanya *et al.*, 1997). The model does not allow for placing of the DNA between the two domains due to protein-DNA clashes. The rotation required to open the clamp is indicated by an arrow. b) A simple rotation of the N-terminal domain as indicated in (a) brings the domains in a conformation allowing DNA placing within the clamp. c) Superposition of the XerH model with rotated N-terminal domain (blue) to the XerH monomer as observed in the XerH-dif_H-LP structure (green). The DNA from the XerH-dif_H-LP structure fits into the model shown in (b) without clashes.

Similarly to the λ integrase and Cre structures with DNA (Aihara *et al.,* 2003; Guo *et al.*, 1997), the C-terminal helix of the DNA-bound XerH protrudes into the neighbouring monomer to form a circular assembly in the synaptic complex. In the DNA-free XerH model, this extended helix is folded back into a cleft on its own monomer, with the catalytic site rearranged and the catalytic tyrosine buried in the body of the protein (Figure 4-23b). This suggests that in XerH recombination, extension of the C-terminal helix is a

crucial component of protein multimerization and synaptic complex assembly, as for other members of the tyrosine recombinase family such as Cre or λ integrase.

4.2.10 Two distinct monomer conformations

The two XerH monomers bound to the same *dif*$_H$-LP site show two distinct conformations (Figure 4-24). Consequently, within the tetrameric synaptic complex two non-adjoining monomers exhibit the same monomer conformation due to the two-fold crystallographic symmetry. The overall structure of the two different monomers in the asymmetric unit is similar, with an R.m.s. deviation of 0.782Å over the Cα carbons of 324 residues. The main differences lie in the arrangement of the two most C-terminal helices αP and αR, which define the assembly of the active site, and in the position of the turn between β-strands β2 and β3 of the antiparallel β-sheet (residues 232-250), in particular in the position of residue Lys239 (Figure 4-25).

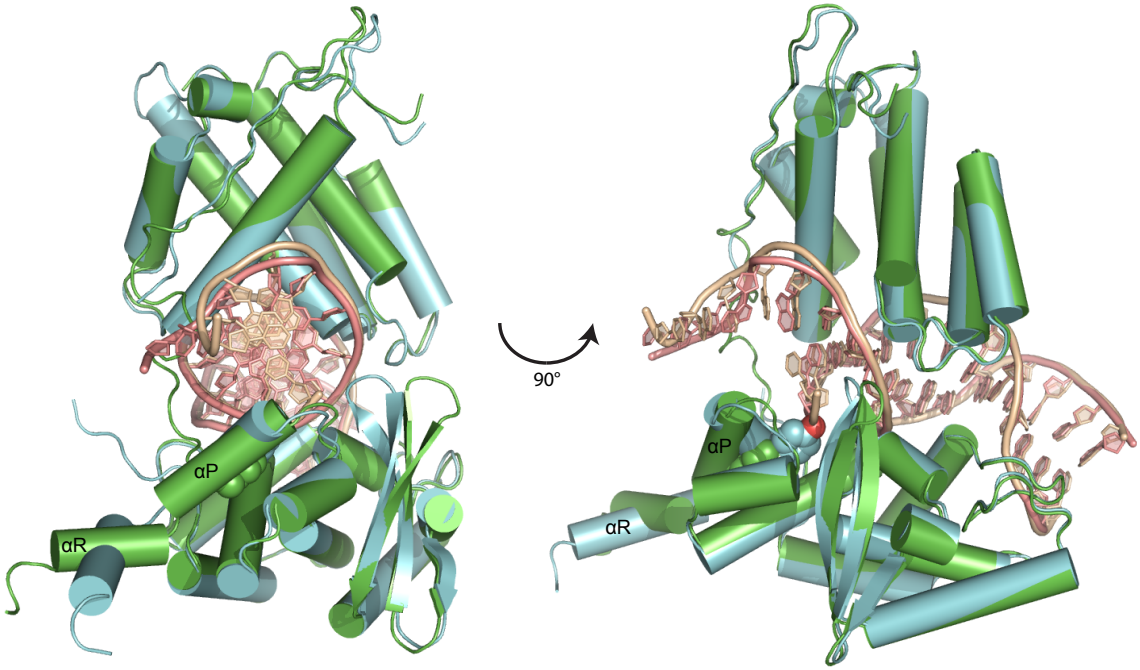


Figure 4-24: Two XerH monomer conformations within the XerH-*dif*$_H$-LP synaptic complex, shown in a cartoon representation with α-helices shown as cylinders and β-strands shown as arrows. The DNA chains associated with the cyan (active) monomer are shown in pink; The DNA chains associated with the green (inactive) monomer are shown in beige. The catalytic tyrosines of both monomers are shown as spheres with atomic colouring.

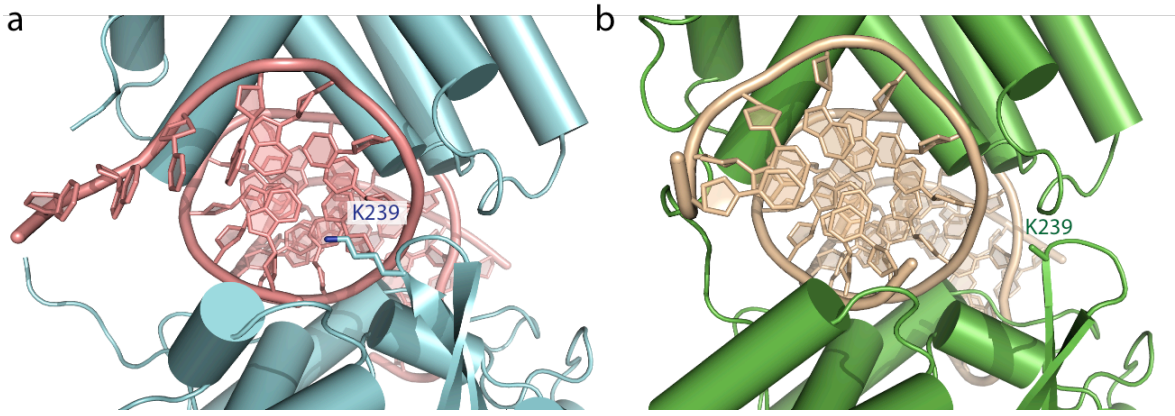


Figure 4-25: Enlarged view of the two XerH monomer conformations in the XerH-*dif*$_H$-LP synaptic complex, shown in a cartoon representation with α-helices shown as cylinders and β-strands shown as arrows. a) Active monomer. Lys239 (shown as a stick model with atomic colouring) makes direct contact to a DNA base (pink). b) Inactive monomer. The β2-β3 loop is more distant from the DNA (beige) and the Lys239 side chain is disordered (shown partially as a stick model) and far from the DNA bases.

The active sites of both monomers consist of the highly conserved tyrosine recombinase residues: the catalytic tyrosine Tyr344, two arginine residues Arg213 and Arg312, and two histidine residues His309 and His335, together forming the catalytic pocket around the scissile phosphate (Figure 4-26). Another conserved residue Lys239 is seen in one of the monomers where it contacts the base adjacent to the scissile phosphate (Figure 4-25a). Together, these active sites residues seen constitute the motif RKHRHY, the signature of tyrosine recombination.

However, the catalytic residues are arranged differently in the two monomers bound to the same *dif*$_H$-LP site. In the active monomer (Figure 4-26a), Tyr344 is joined with a covalent bond to the phosphate of cytosine C13 (bond length 1.56 Å). The sugar and the base moieties of C13 have diffused away and no corresponding electron density could be seen in the structure. Arg213, His309, Arg312, and His335 make hydrogen bonds with the non-bridging oxygen atoms of the tyrosine-bound phosphate group. Lys239 of this monomer contacts the base of A12, the nucleotide adjacent to the cleaved-off C13 (Figure 4-25a). In the inactive monomer (Figure 4-26b), Tyr344 does not make any contacts to the DNA and is rather distant from the phosphorus atom of C13 (5.5Å) that is intact and visible. The catalytic site is not fully assembled in this monomer: Arg213, His309, and Arg312 are still contacting the phosphate but His335 has moved away to 5.5 Å from the oxygen atoms of the phosphate. The density of the Lys239 side chain in this monomer could not be located (Figure 4-25b). The main differences between the two conformations can be attributed to the fact that the catalytic Tyr344 is located on helix αP, which is directly attached to helix αR swapped between the adjacent monomers (Figure 4-24). When comparing the active

and the inactive monomers, it is clear that the conformation of these two most C-terminal helices changes according to the catalytic state of the monomer, critically positioning the tyrosine in an active or inactive position.

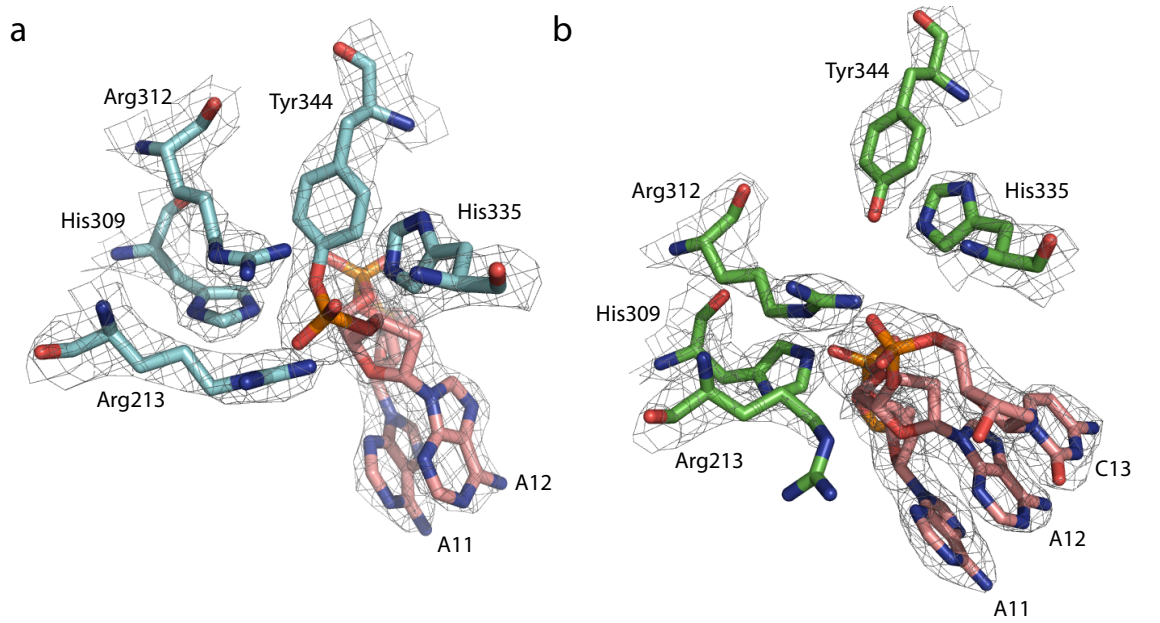


Figure 4-26: The two different active site conformations of XerH monomers within the XerH-dif_H-LP synaptic complex. Catalytic residues are shown with stick representation in atomic colouring (carbon – cyan or green, oxygen – red, nitrogen –blue phosphorus – orange). Grey mesh shows the 2Fo-Fc electron density map for the indicated residues contoured at 1.5 sigma level. The DNA is shown in pink with atomic colouring a) The active site of the active monomer. Tyr344 is covalently joined to the phosphate of C13 that diffused away after DNA cleavage. b) The active site of the inactive monomer. Tyr344 is 5.5 Å from the phosphate of C13, which can still be seen as no DNA cleavage took place.

Additionally, the position of the turn between β-strands β2 and β3 of the antiparallel β-sheet is strikingly different in the two monomers: in the active monomer, this element is closely associated with the body of the protein and contacts the DNA, while in the inactive monomer it has flipped away from the body of the protein and the DNA (Figure 4-24). This positions the catalytic residue Lys239 (present in the loop region) in the active site of the active monomer, where it interacts with the base adjacent to the cleavage position, while in the inactive monomer this residue is far away from the active site (Figure 4-25). Although the Lys239 side chain is disordered in the structure of the non-cleaving monomer, the large distance from the DNA ensures that this residue cannot take part in the interactions observed in the active monomer.

The two distinct arrangements of the catalytic domains of adjacent monomers are consistent with the half-of-the-sites reactivity mechanism of tyrosine recombination, where

in order to assure concerted cleavage and strand exchange reactions only two monomers in the tetramer are active at any given time.

4.2.11 Conformation of the dif_H-LP DNA in the structure

The two dif_H-LP DNA sites in the synaptic complex form a two-fold symmetric antiparallel assembly. Unlike in the analogous structure of the Cre synaptic complex (Guo *et al.,* 1997), the DNA does not show a pseudo-four-fold symmetric arrangement that would resemble a square-planar HJ, but rather a more stacked conformation with an angle of 120° between the two dif_H arms, arrangement that is more similar to that of λ integrase (Biswas *et al.*, 2005). The central region of dif_H is fully base-paired, apart from G13 that normally forms a basepair with C13, which diffused away after cleavage by XerH (Figure 4-27). The electron density of the residues of the central regions is well defined, suggesting low DNA flexibility. However, the central region is substantially distorted from the ideal B-form DNA: bases G13 and A14 are unstacked, with a 90° tilt, introducing a DNA kink and resulting in an asymmetric bending of the dif_H-LP site (Figure 4-27).

Due to the DNA bend, the last two nucleotides of the protruding strand of the non-cleaved dif_H arm are extending into the centre of the synaptic complex (Figure 4-27). The two 5' hydroxyl groups of these protruding strands form a hydrogen bond (2.27 Å) in the centre of the synaptic complex. This positioning is stabilized by the inter-domain linker of the inactive XerH monomer (green protein and beige DNA in Figures 4-24 and 4-25). This part of the linker is disordered in the active monomer and could not be seen in our structure, yet from the position of the more C-terminal, visible part of the linker it is likely that the linker assumes an altogether different conformation (cyan in Figure 4-25). However, the functional significance of the observed hydrogen bonding is not clear, especially given that in the real, un-nicked recombination substrates these strands would be one nucleotide longer.

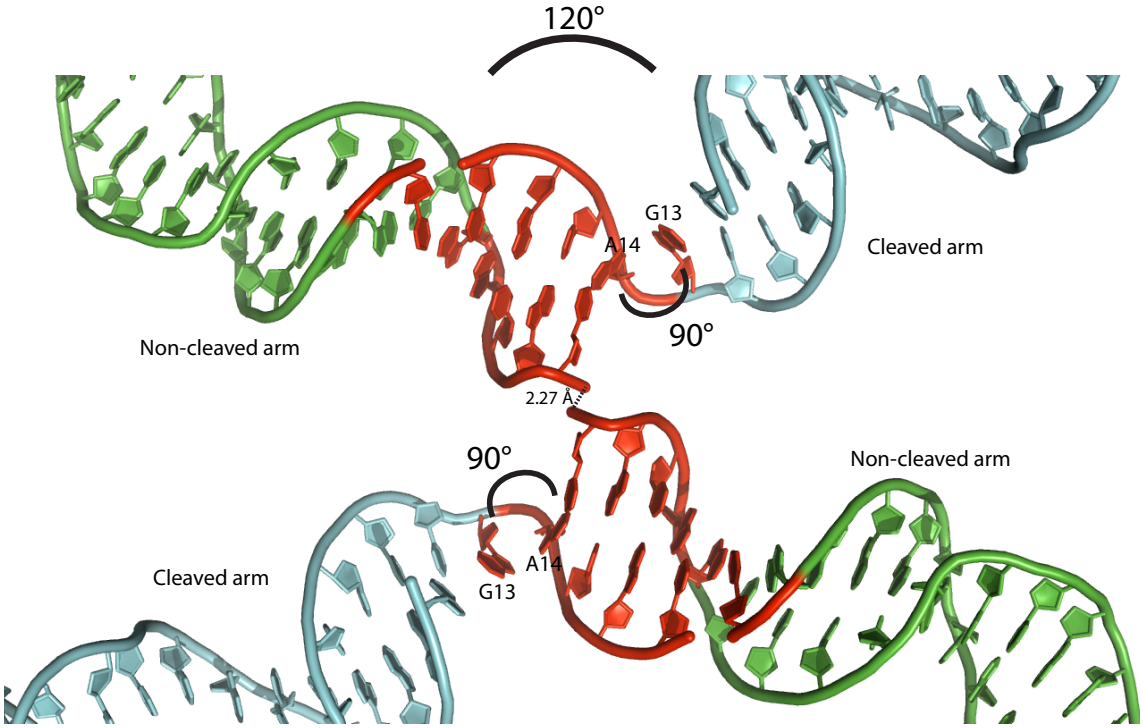


Figure 4-27: The *dif*$_H$ DNA conformation in the XerH-*dif*$_H$-LP synaptic complex. The central region is shown in red. The two binding arms for the active and inactive XerH monomers are shown in cyan and green, respectively. The 90° tilt between bases G13 and A14 that results in DNA bending is indicated. The hydrogen bond length between the two DNA substrates is shown as a dashed line and the bond distance is indicated.

4.2.12 Sequence-specific and non-specific interactions of XerH and *dif*$_H$-LP

XerH interacts with DNA through contacts of both DNA-binding and catalytic domains. The main contacts of the DNA-binding domain are through four helices (αA-αD) forming a helix bundle that is inserted into the major grove of the DNA (Figure 4-21). Most residues interacting with DNA are placed on helices αC and αD, while αA and αB together with the separate αF contribute mainly to protein-protein interactions in the tetramer. αG is inserted into the major groove opposite to the insertion of αA-αD, together forming a tweezer-like structure holding the DNA. The interactions of the C-terminal domain with the DNA are more delocalized. Interestingly, helix αM is inserted into the major groove between basepairs at positions 4-6 and this insertion is stabilized by the interactions between the helix and the sugar-phosphate backbones of adjacent basepairs. Helix αM insertion causes slight narrowing of the major groove (17 Å as opposed to the 22 Å typical for B-DNA), which adds to the DNA bending introduced by a kink at the G13-A14 step (section 4.2.11). It seems that the extensive interactions between XerH and *dif*$_H$-LP stabilize the DNA bend introduced by the kink, but apart from the contribution of helix

αM, XerH-DNA interactions do not induce any further DNA bending as the DNA surrounding the kink resembles closely the ideal, straight B-DNA form (Figure 4-20 and 4-27).

Each XerH monomer in the synaptic complex contacts 11 bp of one of the four dif_H-LP arms through extensive interactions with the sugar-phosphate backbone as well as through direct base readout (Figure 4-28, base pairs 2-11). These two 11-bp arms, together with the 6-bp central region, constitute the full 28-bp dif_H site, consistent with biochemical analysis shown in Chapter 3 and identical in size to the *dif* site of XerC/D (Leslie and Sherratt, 1995).

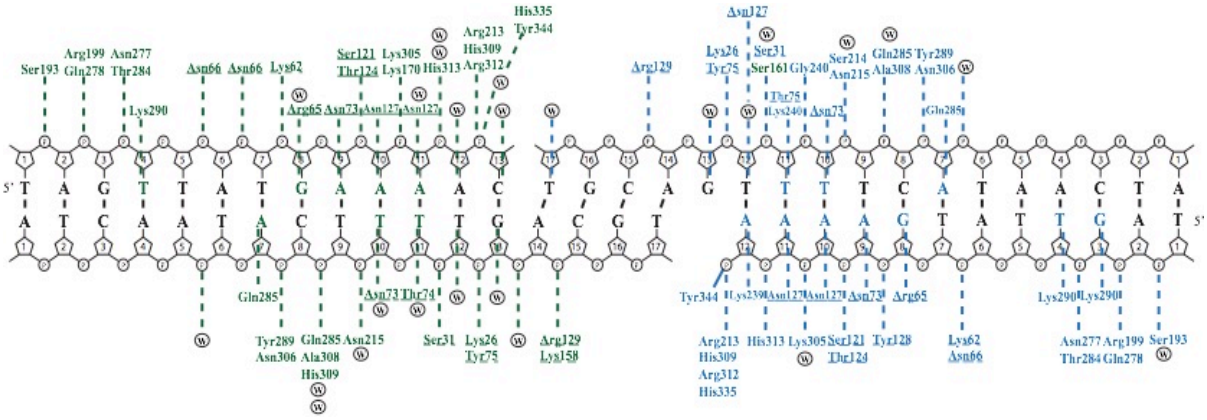


Figure 4-28: Interactions between XerH and the *difH*-LP DNA. The residues of the active monomer are shown in blue and those of the inactive monomer are shown in green. The residues of the N-terminal domain are underlined. Hydrogen bonds (< 3.5 Å) are shown as dashed lines. The phosphotyrosyl bond is shown as a solid line. Bases directly contacted by XerH are coloured green or blue. Interacting water molecules are shown as encircled W. The diagram is based on analysis of protein-DNA interactions performed using NUCPLOT (Luscombe *et al.,* 1997).

Each dif_H half-site is contacted only by the monomer that forms the clamp around it, with the exception of Ser161 of the inactive monomers that contacts the phosphate of thymine T11 of the cleaved arm of the partner dif_H site (Figure 4-29). The 6 bases closest to the central region (guanine G8 and adenines A9 to A12 on the cleaved strand, adenine A7 and thymines T10 and T11 of the non-cleaved strand, Figure 4-28) are contacted directly by the residues of the N-terminal domain (Arg65, Asn73, Thr74, Asn127) and the catalytic domain (Lys239 and Gln285). Adenine bases A10 and A12 are only contacted by the residues of the active monomer. Additionally, basepairs 7-12 are contacted by both XerH domains through non-specific interactions with the sugar-phosphate backbone. These

protein-DNA contacts differ between the active and the inactive monomers emphasizing their different catalytic states. The outer 5 bp are contacted by the catalytic domain of XerH mostly non-specifically through sugar-phosphate backbone interactions with one exception of thymine T4 that is hydrogen-bonded to Lys290. The above observations are consistent with the *in vitro* biochemical data (see Chapter 3) and with the previous study showing that mutating T4 as well as A9-A12 abolished XerH recombination *in vivo* (Debowski *et al*. 2012).

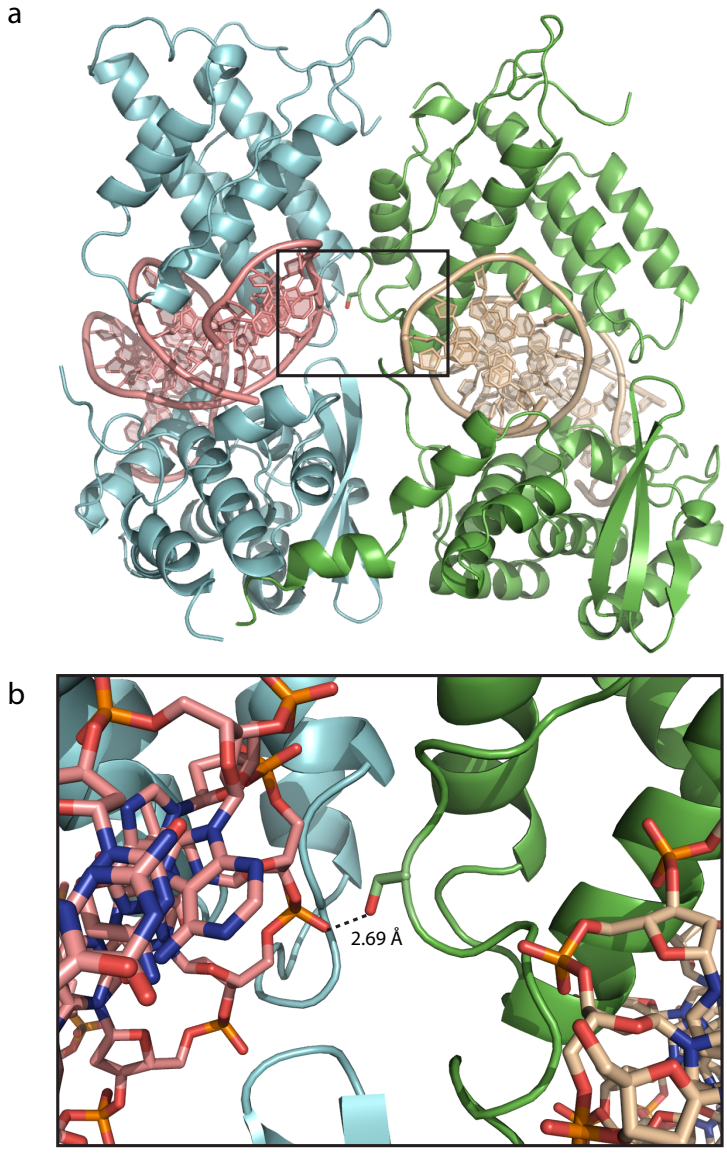


Figure 4-29: Contacts across the synapsis interface in the XerH-dif_H-LP synaptic complex are stabilized by a serine residue Ser161. a) Two XerH monomers bound to two dif_H sites. The active monomer is shown in cyan with bound DNA in pink, the inactive monomer is shown in green with bound DNA in beige. Ser161 is shown as a stick model in atomic colouring. b) Zoomed-in view of the Ser161-DNA interactions. The DNA nucleotides are shown as stick model in atomic colouring. The hydrogen bond between Ser161 and a non-bridging oxygen of the DNA phosphate backbone is shown as a dashed line, with the bond length indicated.

4.2.13 Crystallization of XerH bound to native dif_H

In order to assess if the XerH-dif_H-LP structure does not represent an artefact in some respects due to the use of two identical half-sites, a complex of XerH with native dif_H was prepared using similar double suicide substrates as for XerH-dif_H-LP (see section 4.2.1). This complex crystallized and the structure was solved with molecular replacement at 3.0 Å resolution. However, after structure refinement it became apparent that the resulting structure corresponded to the XerH-dif_H-LP structure. This is likely a consequence of the nature of the suicide oligonucleotides used, which allows annealing of the left arm substrate with the right arm substrate to form the native dif_H, but also permits annealing of the left arm substrate to another left arm substrate to form palindromic dif_H-LP. As a result, the XerH-dif_H-LP complex was formed and only this complex crystallized.

4.3 Discussion

4.3.1 XerH structure in context of known tyrosine recombinase structures

The crystal structure of XerH in a post-cleavage synaptic complex with the dif_H-LP site presented in this work reveals the architecture of the site-specific recombinase XerH, a member of the large family of tyrosine recombinases, of which a number of members has been structurally characterized. Of the available structures, three correspond to a similar, post-cleavage intermediate of the recombination pathway: two structures of Cre bound to a symmetrized *loxP* substrate called *loxA* (Guo *et al.,* 1997) and to a native *loxP* substrate (Ennifar *et al.,* 2003), and a structure of λ integrase bound to an *attP* site (Biswas *et al.,* 2005). In all these structures, the otherwise unstable post-cleavage pre-strand exchange reaction intermediate has been trapped by the use of nicked suicide substrates analogous to the ones used here in the crystallization of the XerH-dif_H-LP synaptic complex (Figure 4-11a). The early work to obtain crystals of recombinase-DNA complexes took advantage of the symmetric oligomeric nature of such complexes, and introduced additional symmetry into the DNA substrates used for complex formation (Guo *et al.,* 1997). This allowed for crystallization of the Cre synaptic complex, which for the first time presented structural insights into the mechanism of tyrosine recombination and provided a paradigm for studying tyrosine recombination. It is likely that the use of a similarly designed symmetrized dif_H site, dif_H-LP, was the prerequisite for crystallization of the XerH-DNA

complex. This notion is supported by the fact that even in attempts to crystallize the complex of XerH with a native dif_H site, only the XerH-dif_H-LP complex crystallized (Section 4.2.13). However, it remains a possibility that the complex crystallized with the dif_H-LP substrate is different from the biologically relevant XerH-dif_H complex. On the other hand, the biochemical activity of dif_H-LP and the observed asymmetry of the XerH-dif_H-LP complex despite the absence of asymmetric DNA, are reassuring that the complex crystallized in this study represents a functional assembly similar to that of a complex containing the native dif_H site. In addition, the fact that the Cre complex structures with symmetrized and native *loxP* sites displayed the same conformation supports this notion.

The XerH-dif_H-LP structure shows that the two domains of XerH are assembled into a C-shaped clamp around the DNA. This assembly is consistent in all structures of tyrosine recombinases bound to DNA. However, the N-terminal domain of XerH differs substantially from that of other tyrosine recombinases: although mainly α-helical, it shows different helix number and arrangement from those of Cre, XerD, and other tyrosine recombinases. Nevertheless, the function of this domain resembles that of the equivalent domains of other tyrosine recombinases: it provides most of the residues that directly recognize DNA bases (Figure 4-28) and contributes to protein-protein interactions in the tetramer through its α-helices αB and αF. The large variability between the N-terminal domains of tyrosine recombinases reflects the large variety of their DNA substrates and their cross-incompatibility, highlighting how the same biochemical mechanism can be preserved in a large family of enzymes in spite of low sequence homology and diverse biological functions.

The C-terminal catalytic domain of XerH is structurally homologous to those of other tyrosine recombinases. Its conserved fold contains the catalytic core consisting of a nucleophilic tyrosine, two arginines, a lysine, a histidine, and an additional variable residue, usually histidine or tryptophan, together constituting the RKHR(H/W)Y motif (see Figure 1-2). The catalytic domain of XerH contains all the residues of the strictly conserved catalytic pentad (Tyr344, Arg213, Lys239, His309, and Arg312), as well as the less conserved residue His336, indicating that XerH is a typical member of the tyrosine recombinase family with respect to the assembly of the active site. The contacts of Arg213, His309, Arg312, and His336 to the non-bridging oxygens of the scissile phosphate are also conserved in XerH compared to Cre, λ integrase, and FLP, and therefore it is likely that the role of these residues in catalysis is also conserved (Aihara *et al*., 2003; Chen *et al*., 2000;

Guo *et al*., 1997). Similarly, the interaction between Lys239 and the DNA base adjacent to the scissile phosphate seen in XerH has been previously observed in the Cre, λ integrase, and FLP structures (Aihara *et al*., 2003; Chen *et al*., 2000; Guo *et al*., 1997). The assembly of the active site of XerH clearly shows that all the residues are provided by one monomer in *cis*.

Together, the structure of XerH is similar to the analogous published structures of Cre and λ integrase, two tyrosine recombinases only distantly related to XerH. This suggests that XerH (and likely other Xer recombinases) are canonical members of the tyrosine recombinase family, and emphasizes that the tyrosine recombinase family, despite sequence diversity and multiple biological roles, is a family of proteins with a well-conserved secondary structure fold, nucleoprotein complex assembly, and recombination mechanism.

4.3.2 XerH structure in context of XerD and XerA structures

To date, there is no published structure of an Xer recombinase in complex with DNA, which means that the structure of XerH-dif_H-LP might be the first such crystal structure solved. The two previously published structures of Xer proteins XerD and XerA (Serre *et al*., 2013; Subramanya *et al*., 1997) contained DNA-free protein monomers corresponding to the protein before the assembly of the recombination complex. Neither of the structures resembled the Xer monomers in the XerH-dif_H-LP synaptic complex. The structure of XerD shows a conformation that could be described as a closed C-shaped clamp, and which does not allow for placing the DNA in the catalytic site without protein-DNA clashes (Figure 4-9). The C-terminal segments that provide an allosteric regulatory mechanism in other tyrosine recombinases are folded back into the body of the protein and do not reach out to interact with a neighbouring monomer. This closed, self-contained arrangement could correspond to the free, monomeric form of the recombinase before assembly of the synaptic complex, providing an elegant mechanism to prevent cleavage of random DNA sequences. XerA shows a more open C-clamp conformation (Figure 4-10), yet still DNA could not be placed in this clamp without protein-DNA clashes, suggesting that this structure is also equivalent to an inactive conformation before synaptic complex assembly. Based on the arrangement in the crystal, two XerA monomers were proposed to form a dimer related through two-fold symmetry; however, the suggested interface is not

consistent with the formation of a cyclic tetrameric synaptic complex and is therefore likely to be caused by a crystallographic artefact. Alternatively, it is also possible that archaeal Xer recombinases employ a distinct recombination mechanism that utilizes an active protein dimer.

In contrast to previous Xer structures, the XerH-*dif*$_H$-LP structure shows a tetrameric DNA-bound synaptic complex, and thus represents an Xer recombination intermediate that was expected to exist based on biochemical studies, but until now was never directly observed.

4.3.3 Large conformational changes involved in DNA binding by XerH

From the analysis of the previous XerD and XerA structures it was clear that large rearrangements of the domains are necessary to accommodate DNA strands in the positively charged cleft containing the catalytic residues. Comparison of the DNA-bound XerH structure with an XerH model based on the DNA-free XerD structure showed how such rearrangement could proceed through a simple rotation of the N-terminal domain along the flexible inter-domain linker (Figure 4-23). With the DNA bound by the inter-domain clamp, the C-terminal helices αP and αR are free to protrude out of the body of the protein and prepare the active site for catalysis as well as stabilize the interactions with the neighbouring monomers necessary for the observed cooperative binding to DNA (see Figure 3-12).

Interestingly, this type of comparison between DNA-bound and DNA-free conformations of a single protein was not possible until now because a pair of full-length DNA-bound and DNA-free recombinase structures has not been solved for any of the tyrosine recombinases. In fact, the only available structures of full-length DNA-free tyrosine recombinases are those of XerD and XerA. These, together with the structure of the XerH-*dif*$_H$-LP synaptic complex, now allowed visualization of the putative changes required to accommodate DNA in the C-shaped clamp (Figure 4-23). Since the DNA-free model was not experimentally determined, but rather modelled based on XerD, which shares only 24% sequence identity with XerH, it is still possible that unbound XerH does not exhibit the same conformation as the XerD monomer or that the two proteins employ a different mechanism of DNA binding. However, the shared fold and the simplicity of the transition

from DNA-free to DNA-bound form (simple rotation of the N-terminal domain) suggest that the proposed mechanism might be biologically relevant for both XerH and XerD.

4.3.4 XerH-*dif$_H$* interactions within the synaptic complex

XerH interacts with the *dif$_H$*-LP substrate extensively both through specific contacts with the bases and non-specific interactions with the sugar-phosphate backbone (Figure 4-28). The specific interactions observed (11 at the cleaved *dif$_H$* arm and 8 at the non-cleaved arm) confer sequence specificity of the protein, which ensures that only true *dif$_H$* sites are recombined by XerH. Additional, non-specific interactions with the sugar-phosphate backbone (28 at the cleaved *dif$_H$* arm, 25 at the non-cleaved arm, and 3 in the central region; 56 contacts in total) are likely to stabilize the binding of XerH throughout the recombination reaction. It is worth noting that the number of XerH-*dif$_H$*-LP contacts is very high: for example, in the equivalent structure of Cre, 39 interactions between the two Cre monomers and the *loxA* site were reported as a strikingly high number (Guo *et al.,* 1997). As another example, a monomer of λ integrase makes 30 interactions with the sugar-phosphate backbone, but only five specific interactions with the DNA bases (Aihara *et al.,* 2003). Interestingly, apart from the direct contacts between XerH and the DNA, many interactions are mediated by water molecules from the solvent. Such high numbers of specific and non-specific interactions and tightly bound solvent are characteristic for site-specific tyrosine recombinases and seem to be preserved in the Xer family.

While the two arms of *dif$_H$*-LP are extensively contacted by XerH monomers, the contacts in the central region are rather sparse. There is no direct interaction between XerH and the DNA bases in this region, and only three contacts to the DNA sugar-phosphate backbone, all to the phosphate of adenine 14 of both strands (Figure 4-28). Such sparse contacts within the central region have been previously seen in Cre structures; however, in that case a bias towards a smaller number of protein-DNA contacts seemed to be observed on the cleaved strand overhang, which is transferred across the synaptic complex in the strand exchange reaction, while the other strand was stabilized in place by more numerous contacts with the recombinase (Gopaul *et al.,* 1998; Guo *et al.,* 1997, 1999). For XerH, no such bias could be observed.

Each of the monomers binds exclusively to one half-site of *dif*$_H$-LP with the exception of a single interaction: serine Ser161 of the inactive monomer contacts the non-bridging oxygen of a phosphate group belonging to the partner *dif*$_H$ site, within the half-site associated with the active monomer (Figure 4-29). Such contacts to the partner recombination site have not been reported for any other tyrosine recombinase, and the serine residue is not conserved in Xer recombinases (see Figure 3-2). Therefore, while this interaction might stabilize the synaptic complex assembly, its physiological relevance remains unclear.

Together, the protein-DNA contacts within the XerH-*dif*$_H$-LP synaptic complex follow a similar pattern as seen for other tyrosine recombinases, including high numbers of both specific and non-specific contacts and the importance of tightly bound solvent, yet the interactions display some differences such as very sparse contacts within the central region and the unique protein-DNA contact across the synapsis.

4.3.5 Allosteric control of XerH activity

The structure of the XerH-*dif*$_H$-LP complex revealed that XerH employs a similar mechanism to ensure half-of-the-sites reactivity as other tyrosine recombinases. Within the synaptic complex, the catalytic domains of XerH monomers contact each other in a circular arrangement mainly through their protruding C-terminal helices αP-αR (Figures 4-20 and 4-21), similarly to the arrangement previously observed in the Cre and λ integrase tetrameric assemblies (Biswas *et al*., 2005; Guo *et al.*, 1997). Therefore, based on the structure of XerH synaptic complex, it is likely that C-terminal helix swapping is also a signature of Xer recombinase synapsis and allows regulation of protein activity also in this enzyme family.

Comparison of the two monomers within the asymmetric unit revealed two distinct conformations, one corresponding to the active monomer and the other to the inactive one. Such comparison allows for visualization of what conformational changes would be required for transition between the two states, required by the mechanism of half-of-the-sites reactivity. Like all other tyrosine recombinases, the key change seems to involve the very C-terminal segment that shows distinct conformations in the two monomers (Figure 4-24). In XerH this segment consists of two C-terminal helices αP and αR (the latter

protruding into the body of the neighbouring monomer). The positioning of these helices has a direct impact on the assembly of the active site as the catalytic tyrosine is placed at the end of helix αP. The C-terminal helix αR of the inactive monomer faces more into the body of the neighbouring (active) monomer, which might push its helix αP towards the DNA and therefore position the catalytic tyrosine ideally for cleavage. However, from the structure of XerH-*dif*$_H$-LP it is not clear how exactly the transition between the two catalytic states of the monomers is brought about.

Another difference between the two monomers is the position of a loop between β-strands β2 and β3 of the catalytic domain (Figures 4-24 and 4-25). In the active monomer, this loop is well ordered and is in close contact with the body of the protein and the DNA, while in the inactive monomer the loop is positioned further away and some of the amino acid side chains cannot be seen in the electron density map, suggesting higher flexibility of this region. This loop carries a conserved lysine residue Lys239 that in the cleaving monomer makes a direct contact with a base adjacent to the scissile phosphate. Such an interaction has also been observed in the Cre HJ structure, where it was proposed that the interaction of this lysine with the DNA is sufficient to reposition the β2-β3 loop (Martin *et al.,* 2002). Based on this observation, the authors concluded that Cre distinguishes between the contacted bases in the left and the right arms of *loxP*, ordering the catalytic lysine close to the active site for one arm but not the other (Martin *et al.*, 2002). However, in *dif*$_H$ the corresponding positions are equivalent (adenine) and therefore this interaction is not likely to contribute greatly to half-site differentiation. Nevertheless, the repositioning of the β2-β3 loop could have important implications for determining the overall protein conformation and therefore for defining the order of cleavage in XerH recombination.

4.3.6 *dif*$_H$ DNA bending in the synaptic complex

The DNA in the XerH-*dif*$_H$-LP synaptic complex is asymmetrically bent by a sharp kink caused by a 90° tilt at a single base step adjacent to the XerH cleavage site (Figure 4-27). Such a sharp bend at an equivalent position has been observed before in the Cre pre-cleavage synaptic complex (Ennifar *et al*., 2003), suggesting that it is not an artefact of using pre-nicked substrates or a result of cleavage, but rather an important and conserved regulatory element of tyrosine recombination. The molecular mechanism introducing the DNA bend could not be derived from the XerH-*dif*$_H$-LP structure as there are no protein-

DNA interactions in the vicinity of the bend, and the unstacked bases are exposed to the solvent. The vast interactions between XerH and the DNA clearly stabilize the central kink and also contribute slightly to further DNA bending by narrowing the major groove on each dif_H arm by insertion of the α-helix αM.

Previously described structures of tyrosine recombinase post-cleavage synaptic complexes, namely those of Cre and λ integrase, present two different conformations of DNA within the synaptic complex (Biswas *et al*., 2005; Guo *et al.,* 1997). The synaptic complex of Cre, a recombinase that does not require any additional factors for successful recombination, adopts a pseudo-fourfold symmetric conformation with the DNA bent to 100°. Following cleavage, this intermediate can easily proceed into the HJ intermediate without any large conformational changes: in fact, only the protruding DNA overhangs have to change their position to complete the ligation and form the HJ (Gopaul *et al.*, 1998). On the other hand, λ integrase adopts a different synaptic complex conformation with the DNA bent to 110° in a two-fold symmetric, rhombic arrangement (Biswas *et al*., 2005). Interestingly, strand exchange (and therefore HJ formation) in λ integrase is facilitated by binding of accessory DNA sequences by its N-terminal arm-binding domain, which helps remodelling the complex for HJ formation (Biswas *et al*., 2005; Radman-Livaja *et al*., 2003). The XerH-dif_H-LP synaptic complex adopts a similar rhombic arrangement to the λ integrase complex (see the arrangement of catalytic tyrosines in Figure 4-20). The DNA is bent only by 60° and the central region is fully base-paired, suggesting that a remodelling of the complex as well as breaking of the base-paring in the central region might be required for the formation of the HJ intermediate. This could indicate that accessory factors might be required for strand exchange and further HJ isomerization and resolution in XerH recombination, which is supported by the failure to observe recombination *in vitro* despite efficient first-strand cleavage. However, further studies, and preferably further structures of various recombination intermediates of XerH are required to support this notion.

4.3.7 Ordered XerH cleavage of dif_H

One of the puzzling aspects of XerH recombination and other systems employing a single recombinase is how the correct order of binding and cleavage events can be assured. Unlike in the case of the highly asymmetric *dif* site of XerC/D, both dif_H arms are almost identical apart from a single basepair insertion at position 6 of the left arm (see Figure 3-

23a). This high symmetry of dif_H reflects the need for two identical monomers to bind the site both specifically and efficiently. Despite this, the results of biochemical experiments imply that the dif_H asymmetry caused by the single-basepair insertion defines order of recombination events. However, inspection of the crystal structure of XerH-dif_H-LP did not reveal any specific contacts with this basepair (Figure 4-28). This agrees with the biochemical experiments indicating that a basepair of any identity at this position promotes efficient cleavage as long as exactly one basepair is inserted (See Figure 3-21). It is possible that the structure of the DNA in this region adopts a specific conformation, which XerH could recognize through indirect readout: A/T rich sequences (here: TTAT) have been shown to cause effects such as narrowing of the minor groove and increased bendability of DNA (Shatzky-Schwartz *et al.,* 1997). Unfortunately, a structure of XerH in complex with native dif_H, which could shed light on such specific differences between the two arms of dif_H, could not be obtained in this study.

Despite the symmetrized DNA substrates used in this study, the obtained structure of the XerH-dif_H-LP shows asymmetric features, with the most striking being the two different conformations of the XerH monomers. An additional asymmetric feature is the sharp bend in the DNA, which is caused by the 90° tilt between the adenine 14 and the unpaired guanine 13 of the top (non-cleaved) strand of the right dif_H arm (Figures 4-27 and 4-28). A similar bend was observed in Cre structures and was proposed to define the order of cleavage within the synaptic complex by promoting cleavage of the arm adjacent to the bend (Ennifar *et al.*, 2003). However, the exact role of the bend is not clear from any of the Cre structures and it could not be elucidated from the XerH-dif_H-LP structure. It is also worth noting that in the first structure of Cre where such bend was initially observed (Martin *et al.,* 2002), the bend was positioned on the opposite side of the central region and it appeared to promote cleavage of the non-adjacent *loxP* arm. However, this structure was solved with a symmetrized *loxA* substrate, which masked the differences in sequence at the bend position present in the native *loxP* site. The next Cre structure solved with native *loxP* showed that the sequence of the central region was in fact the determinant of the position of the bend (Ennifar *et al.,* 2003). Since the native dif_H also contains two different basepairs at the two outermost positions of the central region, the position of the bend could be defined by the identity of the basepairs just like in Cre recombination. In this case, the use of symmetrized dif_H-LP substrate might have affected the positioning of the bend, similarly to what was seen for Cre recombinase. To further explore this, obtaining a structure of XerH bound to the native dif_H site would be beneficial.

4.3.8 Implications of the XerH-dif_H-LP structure

The structure of XerH in a synaptic complex with dif_H-LP carries implications for further research of both XerH recombination in particular and Xer recombination in general. First of all, *H. pylori* as a cause of gastric cancer presents an important target for design of anti-microbial agents. Phylogenetic studies showed that XerH recombinase found in *Helicobacter* and *Campylobacter* genera is only distantly related to other bacterial genera, providing an interesting genus-specific target (Carnoy and Roten, 2009). The indication that XerH requires additional factors for recombination similar to XerC/D regulation by FtsK suggests that allosteric interactions might be involved; those could be targeted using a large peptide screen based on the information available from our structure. Additionally, studies in *H. pylori* have been limited by the lack of genetic tools in this gram-positive bacterium, and XerH has been recently introduced as a potential genetic tool for *H. pylori* markerless gene deletions (Debowski *et al.,* 2012a). The structure of XerH bound to DNA and biochemical characterization of dif_H mutations could help in the construction of a designer recombinase that could be used on a larger variety of sites to allow for site-specific gene deletion and replacement in *H. pylori*.

The XerC/D system of *E. coli* has been extensively characterized biochemically on a variety of sites (chromosomal *dif*, plasmid *cer* and *psi*, CTXΦ *att* sites), yet no high-resolution structural data revealing Xer-DNA interactions were obtained to further explain its mechanism. The XerH-dif_H-LP structure reveals the molecular basis of Xer-*dif* interactions, confirms that Xer family follows the dogmas of tyrosine recombination (tetrameric cyclic assembly, protruding C-terminal helix, active site assembly), and explains how the structural assembly of the synaptic complex necessitates the help of accessory factors for Xer recombination. Moreover, the structure of XerH-dif_H-LP could allow confident modelling of the XerC/D synaptic complex, which in turn might help to address the remaining questions about the mechanism of Xer recombination.

5. Regulation of XerH recombination

5.1 Introduction

5.1.1 The role of FtsK in single-recombinase Xer systems

The role of *E. coli* FtsK in XerC/D recombination has been described in detail in section 1.7. Briefly, FtsK regulates XerC/D by two different mechanisms: it directly activates XerD, rendering it competent for cleavage, and it assures the correct topology of recombination substrates and products. For activation of the XerD subunit *in vivo* and *in vitro*, only the γ subdomain (FtsKγ; Figure 5-1) of the FtsK ATP-helicase domain ($FtsK_C$) is required. However, the products of recombination in the presence of FtsKγ but in the absence of the remaining parts of $FtsK_C$ exhibit complicated topology, including catenanes and knots (Grainge *et al.,* 2011). In order to generate free circular products, the full functional hexameric helicase domain, consisting of $FtsK_C$ α, β, and γ subdomains and an additional stretch of 50 amino acids from a flexible inter-domain linker, is required.

Homologues of FtsK can be found in most bacterial species carrying Xer genes, including the ones that feature single Xer recombinase systems (Recchia and Sherratt, 1999). The role of FtsK in Xer recombination has been investigated for several single-recombinase systems, including streptococcal XerS/dif_{SL} (Le Bourgeois *et al.,* 2007; Nolivos *et al.,* 2010), archaeal XerA/*dif* (Cortez *et al.*, 2010), and both *C. jejuni* and *H. pylori* XerH/dif_H (Debowski *et al.,* 2012; Leroux *et al*., 2013). This showed that each of those systems seems to depend to a different degree on FtsK for successful recombination, ranging from no requirement for FtsK to a strict requirement for species-specific interaction with FtsK.

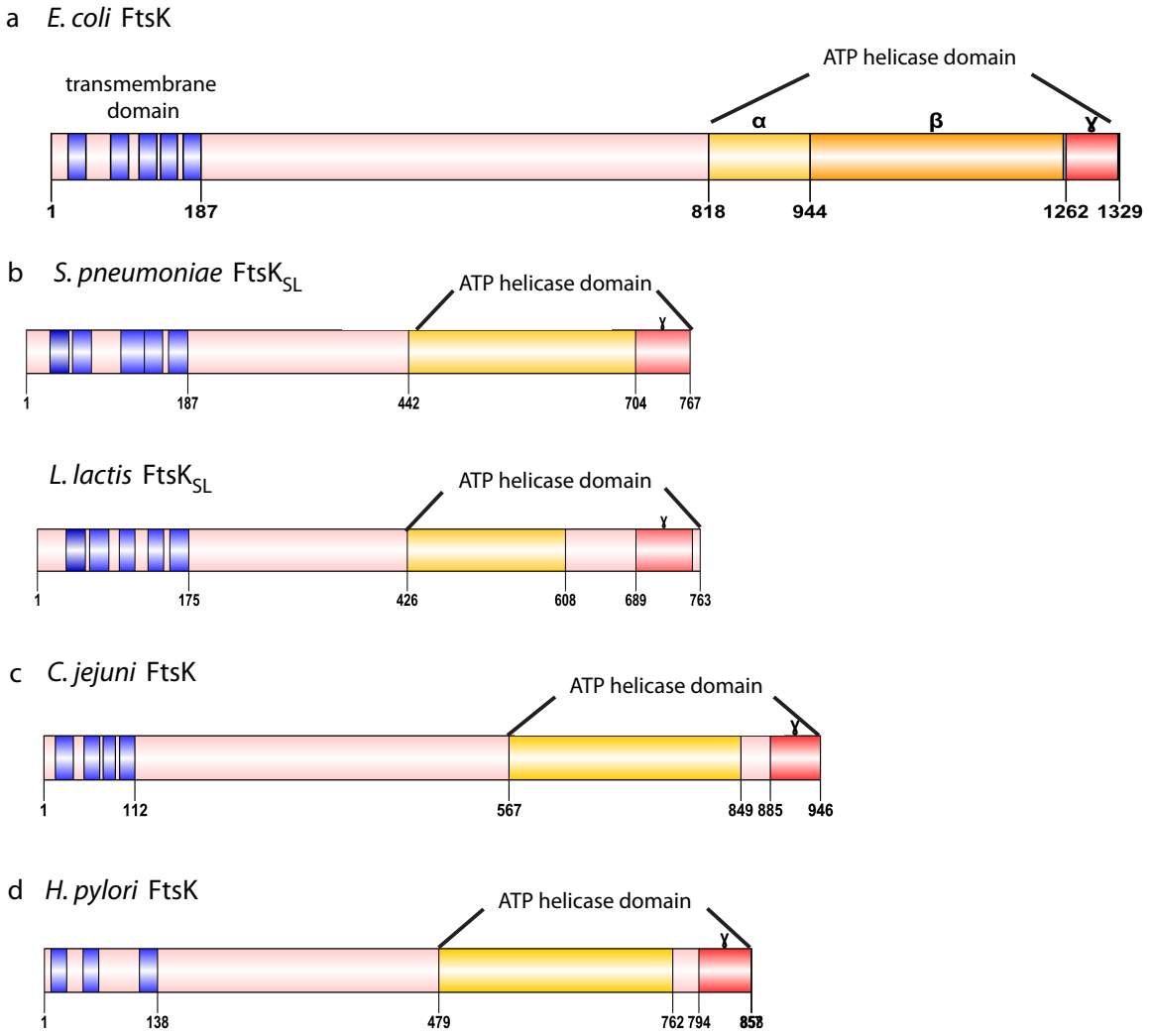


Figure 5-1: Domain representation of FtsK proteins from various bacterial species. The transmembrane helices within the N-terminal transmembrane domain are shown in blue. The ATP helicase domain is shown in yellow, and the conserved γ domain is shown in red. a) *E. coli* FtsK. The C-terminal ATP helicase domain consists of three subdomains α (yellow), β (orange), and γ (red). b) FtsK$_{SL}$ proteins from *S. pneumoniae* and *L. lactis*, bacteria featuring the XerS recombination system. c) *C. jejuni* FtsK. d) *H. pylori* FtsK.

5.1.1.1 XerS

Bacteria carrying the XerS recombination system (streptococci and lactococci) also carry a homologue of FtsK, termed FtsK$_{SL}$ (Figure 5-1b). The size of the protein ranges between 758 and 816 amino acids in streptococcal genomes (Le Bourgeois *et al.*, 2007). Similarly to *E. coli* FtsK, FtsK$_{SL}$ contains an N-terminal domain with five transmembrane helices and a conserved C-terminal DNA helicase domain. FtsK$_{SL}$ localizes to the division septum through its N-terminal domain, but (unlike in *E. coli*) this domain is not essential for viability (Le Bourgeois *et al.,* 2007).

Recombination experiments in *S. pneumoniae*, where the integration of a plasmid carrying the dif_{SL} site into the chromosome was monitored, showed that XerS recombination efficiency decreases dramatically in the absence of streptococcal $FtsK_C$ to approximately 1% of the recombination efficiency in the wild-type strain (Le Bourgeois *et al.*, 2007). This showed that streptococcal $FtsK_C$ is an essential part of the XerS/dif_{SL} recombination machinery.

Replacement of *E. coli* chromosomal *dif* site with a dif_{SL}-Km-dif_{SL} cassette and expression of XerS from a helper plasmid resulted in an observed XerS-mediated cassette excision frequency of 10% per cell per generation (Le Bourgeois *et al.,* 2007), a rate similar to that of XerC/D recombination in *E. coli* (Perals *et al.,* 2000). However, in the absence of $FtsK_C$, XerS recombination rates decreased from 10% to 0.1% per cell per generation (Le Bourgeois *et al*., 2007), showing that in *E. coli*, *E. coli* $FtsK_C$ is required for efficient XerS recombination, functionally replacing the native $FtsK_{SL}$. It is worth noting that in XerC/D recombination in *E. coli*, no recombination can be seen in the absence of $FtsK_C$, while XerS recombination yields 0.1% of recombination per cell per generation, making the dependence of XerS on FtsK less stringent than that of XerC/D.

The observations regarding the requirement for FtsK in the *S. pneumoniae* XerS/dif_{SL} system were also confirmed for *L. lactis* XerS in experiments in *E. coli* where dif_{SL} replaced the native *dif* site (Nolivos *et al.,* 2010). Interestingly, the reduced rate of XerS recombination in the $FtsK_C$-deficient *E. coli* strain could not be restored by ectopic expression of *E. coli* $FtsK_C$, suggesting that cellular localization of $FtsK_C$ to the division septum conferred by the N-terminal transmembrane domain is required for XerS-mediated chromosome dimer resolution. At the same time, $FtsK_C$ expressed from a plasmid was sufficient to support excision of the dif_{SL}-lacI-dif_{SL} cassette replacing *E. coli dif*, showing that septum-independent $FtsK_C$ is sufficient for XerS recombination, but not for chromosome dimer resolution (Le Bourgeois *et al.,* 2007). A similar pattern has been observed for XerC/D recombination (Capiaux *et al.*, 2002).

Investigation of the specific role of FtsKγ in *L. lactis* XerS recombination revealed that deletion of *E. coli* FtsKγ had the same effect on XerS recombination as deletion of the complete *E. coli* $FtsK_C$, suggesting that this subdomain is essential for XerS recombination in *E. coli* (Nolivos *et al.,* 2010). However, rescue experiments revealed that *E. coli* FtsKγ is not sufficient to support XerS/dif_{SL} recombination in $FtsK_C$-deficient *E. coli* strain

whereas it activates XerC/D recombination, and *L. lactis* FtsKγ is not sufficient for activation of either XerS or XerC/D recombination (Nolivos *et al.,* 2010). This implies that while $FtsK_{SL}γ$ is essential for XerS recombination, the direct activation of XerS by $FtsK_{SL}γ$ is different than that of *E. coli* XerD by FtsKγ, and the details of XerS recombination activation by $FtsK_{SL}γ$ remain unclear. Interestingly, recognition of KOPS (see section 1.7) and translocation of FtsK along the DNA towards the recombination complex were also shown to be essential for XerS recombination in *E. coli* (Nolivos *et al.,* 2010), which confirms that the dual nature of FtsK function previously described for XerC/D is conserved in XerS recombination.

It is also worth noting that no XerS recombination could be observed *in vitro* in the presence of purified XerS and $FtsK_{SL}$ in conditions supporting XerC/D recombination (Nolivos *et al*., 2010).

5.1.1.2 XerA

While most archaeal genomes encode XerA recombinase that resolves chromosomal dimers, no FtsK homologues could be identified in Archaea (Serre and Duguet, 2003). Additionally, the XerA protein of *P. abyssi* was shown to recombine plasmids *in vitro* in the absence of any protein partners (Cortez *et al.*, 2010). It is therefore likely that XerA recombination is not regulated by FtsK. However, the *in vitro* assay showed much higher levels of plasmid multimerization than resolution into monomers, suggesting that an additional mechanism might help to ensure topological resolution of chromosome dimers. This mechanism could work similarly to regulation by FtsK. In fact, a related DNA translocase HerA was suggested as a potential protein that could regulate XerA recombination (Iyer *et al.,* 2004). In agreement with this notion, polarized DNA sequences that could serve as a guide for the DNA helicase were identified (Cortez *et al.,* 2010). However, the role of HerA in coupling XerA recombination to cell division is yet to be confirmed.

5.1.1.3 *C. jejuni* XerH

C. jejuni carries a homologue of FtsK consisting of 946 amino acids with a typical FtsK domain architecture (transmembrane N-terminal domain and helicase C-terminal domain; Figure 5-1c). However, the role of *C. jejuni* FtsK in XerH recombination has not been

tested. Instead, *C. jejuni* XerH recombination was assessed in *E. coli* wild-type and $FtsK_C$-deficient strains (Leroux *et al.,* 2013). In this experiment, a *dif*$_H$-Km-*dif*$_H$ cassette was introduced into a plasmid so that upon recombination the kanamycin resistance cassette is lost. This plasmid, together with an XerH expression plasmid, was introduced into wild-type and $FtsK_C$-deficient *E. coli* strains. Recombination was observed in both strains, but was very inefficient and only seen after many generations of growth (Leroux *et al.,* 2013). The observed activity could be due to genuinely inefficient FtsK-independent XerH activity or the inability of *E. coli* FtsK to substitute for the functional *C. jejuni* FtsK, and the biological relevance of this activity remains to be tested.

As in the case of XerS, no recombination could be seen *in vitro* when using purified *C. jejuni* XerH, suggesting that partner proteins might be required for XerH recombination.

5.1.2 The role of FtsK in *H. pylori* XerH recombination

H. pylori carries an FtsK homologue (HP1090 in strain 26695) closely related to that of *C. jejuni* (43% sequence identity). This 858-amino acid protein (Figure 5-1d) consists of a 138-amino acid N-terminal transmembrane domain with at least three transmembrane helices, a 341-amino acid flexible interdomain linker, and a 379-amino acid C-terminal domain featuring the ATP-helicase fold. The C-terminal domain is well conserved and shows 47% amino acid sequence identity with *E. coli* $FtsK_C$, with the FtsKγ subdomain sequence identity of 38.8% as determined using ClustalW2 alignment tool (Larkin *et al.,* 2007).

The study of *H. pylori* XerH (Debowski *et al.,* 2012) investigated the role of FtsK in XerH recombination in *H. pylori*: 404 C-terminal amino acids of FtsK were deleted to generate a truncated version of the protein, missing the FtsKγ domain. Recombination efficiency was assessed by monitoring excision of a *dif*$_H$-Cm-*dif*$_H$ cassette from an ectopic location on the *H. pylori* chromosome. No *dif*$_H$ recombination could be detected in the strain harbouring the FtsK truncation (Debowski *et al.,* 2012). It was therefore concluded that FtsK is essential for recombination of *dif*$_H$ sites. Similarly to Archaea, a putative motif that could serve as an equivalent of FtsK-guiding *E. coli* KOPS in *H. pylori* has been identified (Debowski *et al.,* 2012; Hendrickson and Lawrence, 2006).

5.1.3 Species-specificity of FtsK-Xer interactions

In *E. coli*, the interaction interface between XerD and FtsK has been mapped (Yates *et al*., 2006). It appears that residues 282-292 of the XerD catalytic domain containing an amino acid motif RQxxQQ are essential for interaction with FtsKγ, while on FtsK the residues 1277-1282 that form an amino acid motif TEKRKA are required (Figure 5-2). Due to these highly localized interactions it was proposed that interactions between Xer and FtsK proteins might be species-specific in nature. This is supported by the fact that FtsK does not interact with XerC, and that homologues of XerD and FtsK from other species do not contain the above motifs and therefore the same interaction would not be possible (Figure 5-2).

The species-specificity of the Xer-FtsK interaction was first addressed in a study of XerC/D systems from *E. coli* and *H. influenzae* (Yates *et al.,* 2003). The study showed that *E. coli* FtsK did not support *H. influenzae* XerC/D recombination *in vivo* and *vice versa*. However, an *E. coli* FtsK chimeric protein containing *H. influenzae* FtsKγ could activate *H. influenzae* XerC/D recombination, and in a reciprocal test *H. influenzae* FtsK chimeric protein containing *E. coli* FtsKγ could activate *E. coli* XerC/D recombination (Yates *et al.,* 2003). This indicated that the species-specificity lies within the γ subdomain of FtsK.

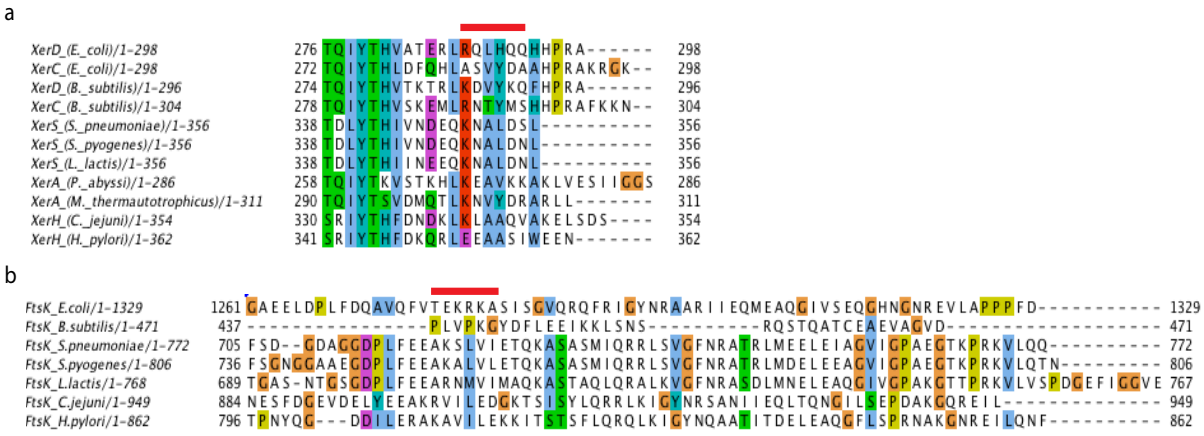


Figure 5-2: Sequence alignments of the C-termini of Xer (a) and FtsK (b) proteins from various bacterial and archaeal species. The alignments were prepared by ClustalW2 (Larkin *et al.,* 2007) and visualized using Jalview software (Waterhouse *et al.,* 2007). The residues are coloured according to the ClustalX scheme in Jalview. The regions of *E. coli* XerD and FtsK implicated in XerD-FtsK interactions are indicated by red bars above the alignments.

The studies of the XerS/dif_{SL} system also contributed to the understanding of Xer-FtsK interactions. The recombination assays with chromosomal dif_{SL}-Km-dif_{SL} performed in *E. coli* wild-type and $FtsK_C$-deficient strains showed that FtsK of *E. coli* supports XerS recombination despite the lack of a recognition motif equivalent to that of XerD (Le Bourgeois *et al.,* 2007). This would suggest that the $FtsK_C$-XerS interaction requires no species specificity. Similarly, further studies showed that the $FtsK_C$ domain from either *E. coli* or *L. lactis* supports XerC/D/*dif* and XerS/dif_{SL} recombination (Nolivos *et al.,* 2010). However, the FtsKγ subdomain of *E. coli* only supported XerC/D recombination, while the FtsKγ subdomain of *L. lactis* did not support any of the recombination systems. This suggests that, despite high sequence conservation (41%) between the two FtsKγ subdomains, *L. lactis* FtsKγ activates XerS recombination by a different mechanism. This further implies that the species-specificity of Xer-FtsK interactions might not only require matching Xer-FtsK recognition motifs, and altogether different mechanisms of activation might also be involved.

5.1.4 Aims and objectives

Reconstitution of the complete XerH recombination pathway *in vitro* has been so far unsuccessful, likely because (similarly to other Xer proteins) XerH requires additional regulatory factors not present in the *in vitro* reactions. Consequently, and considering that various studies show that different Xer systems depend on FtsK to different extents, it would be interesting to assess the role of FtsK in *H. pylori* XerH recombination. To do so, I aimed to investigate the direct interactions between purified *H. pylori* XerH and FtsK proteins as well as the impact of FtsK on the outcome of *in vitro* cleavage and recombination assays. Furthermore, I set out to study XerH recombination *in vivo* in *E. coli* to further address the questions regarding the role of FtsK in XerH recombination of plasmid- and chromosome-borne dif_H sites and the species-specificity of the XerH recombination machinery. Together, this study will help us to understand the regulation and requirements for successful XerH chromosome dimer resolution and will therefore contribute to further elucidation of the XerH recombination mechanism, derived so far from structural and biochemical analysis.

5.2 Results

5.2.1 Overexpression and purification of *H. pylori* $FtsK_C$ constructs

To investigate the role of FtsK in XerH recombination in *in vitro* experiments, *H. pylori* $FtsK_C$ constructs were designed for overexpression in *E. coli* and purification. FtsK proteins from the two *H. pylori* strains used in this study (26695 and P12) show 96% sequence identity, with over 99% sequence identity in the conserved helicase domain, and therefore are likely interchangeable. Here, FtsK from strain P12 was chosen for further studies as this strain carries both XerH and XerT recombinases, and therefore one protein could be used in studies of both systems. The sequence alignment of *H. pylori* and *E. coli* FtsK proteins revealed that the *E. coli* $FtsK_C$ construct sufficient to support XerC/D recombination *in vitro* and *in vivo* (Barre *et al.,* 2000) corresponds to residues 401-863 of *H. pylori* FtsK. Accordingly, two constructs $FtsK_C$-A and $FtsK_C$-B have been designed (Figure 5-3), both of which include the bulk of the predicted $FtsK_C$, but have different N-terminus designed based on secondary structure predictions. $FtsK_C$-A (399-863) starts close to the predicted $FtsK_C$ start, while $FtsK_C$-B (418-863) starts at the beginning of a predicted α-helix to eliminate the disordered N-terminal region (Figure 5-3). Each construct was introduced into the pETM-22 expression vector by restriction-free cloning. This vector contains the N-terminal TRX solubility and 6xHis affinity tags. The TRX is a small *E. coli* protein shown to improve solubility of overexpressed protein constructs when attached to the N-terminus of a protein of interest (LaVallie *et al.,* 2000).

$FtsK_C$-A and $FtsK_C$-B were overexpressed in *E. coli* BL21 by growth of expression cultures at 37 °C until OD_{600} of 1.2, followed by induction with 1 mM IPTG and expression at 15 °C for 15 hours. This initial protocol yielded satisfactory amounts of $FtsK_C$ (Figure 5-4a) and was not optimized any further.

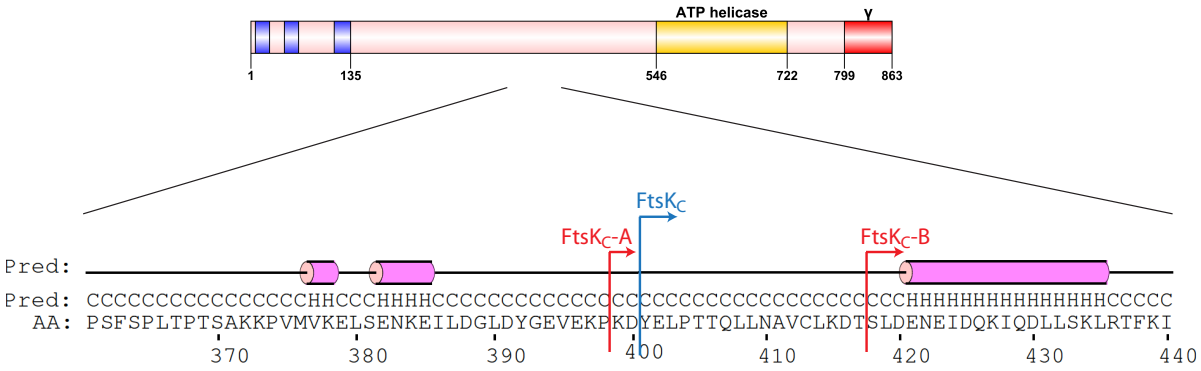


Figure 5-3: *H. pylori* FtsK$_C$ construct design. Top: the domain representation of *H. pylori* P12 FtsK, coloured as in Figure 5-1. Bottom: zoomed-in view of residues 361-440 of *H. pylori* P12 FtsK with secondary structure predictions prepared using PSIPRED Protein Sequence Analysis Workbench (Buchan *et al.,* 2013). The amino acid sequence (AA) is accompanied by the secondary structure prediction for each residue (Pred) expressed as a letter (H: α-helix and C: coil, loop, or unstructured region) and as a graphical representation (magenta cylinder: α-helix). The start of the FtsK$_C$ region based on alignment with *E. coli* FtsK is indicated in blue. The start points of the designed constructs FtsK$_C$-A and FtsK$_C$-B are shown in red.

Both constructs were purified to homogeneity by combined affinity and size exclusion chromatography purification. In detail, the overexpressed TRX-6xHis-FtsK$_C$ constructs were extracted from the soluble fraction of *E. coli* lysate by application to the HisTrap column. The elution fractions containing TRX-6xHis-FtsK$_C$-A or TRX-6xHis-FtsK$_C$-B were combined and subjected to proteolytic cleavage by 6xHis-tagged 3C protease. The 3C protease cleavage site is located downstream from the TRX-6xHis tag so that 3C cleavage can liberate the target protein constructs. The resulting tag-free FtsK$_C$-A or FtsK$_C$-B was applied again onto the HisTrap column and the protein of interest was collected in the flow through. FtsK$_C$-containing fractions were then loaded onto a SEC column. The FtsK$_C$-B protein eluted at the void volume of the column, indicating that this protein is aggregated in the conditions used (Figure 5-4b, bottom panel). The FtsK$_C$-A protein eluted from the SEC column at a volume corresponding to a monomer of FtsK$_C$-A with only a small amount of aggregated protein (Figure 5-4b, top panel). For FtsK$_C$-A, the peak fractions were analysed by SDS-PAGE (Figure 5-4c) and fractions 1-4 were combined. The analysis showed that the obtained FtsK$_C$-A was pure apart from a small amount of a contaminant, likely a degradation product of FtsK$_C$-A, rendering the purified protein suitable for further biochemical studies. The purification protocol yielded approximately 2 mg of purified FtsK$_C$-A from one litre of the expression culture.

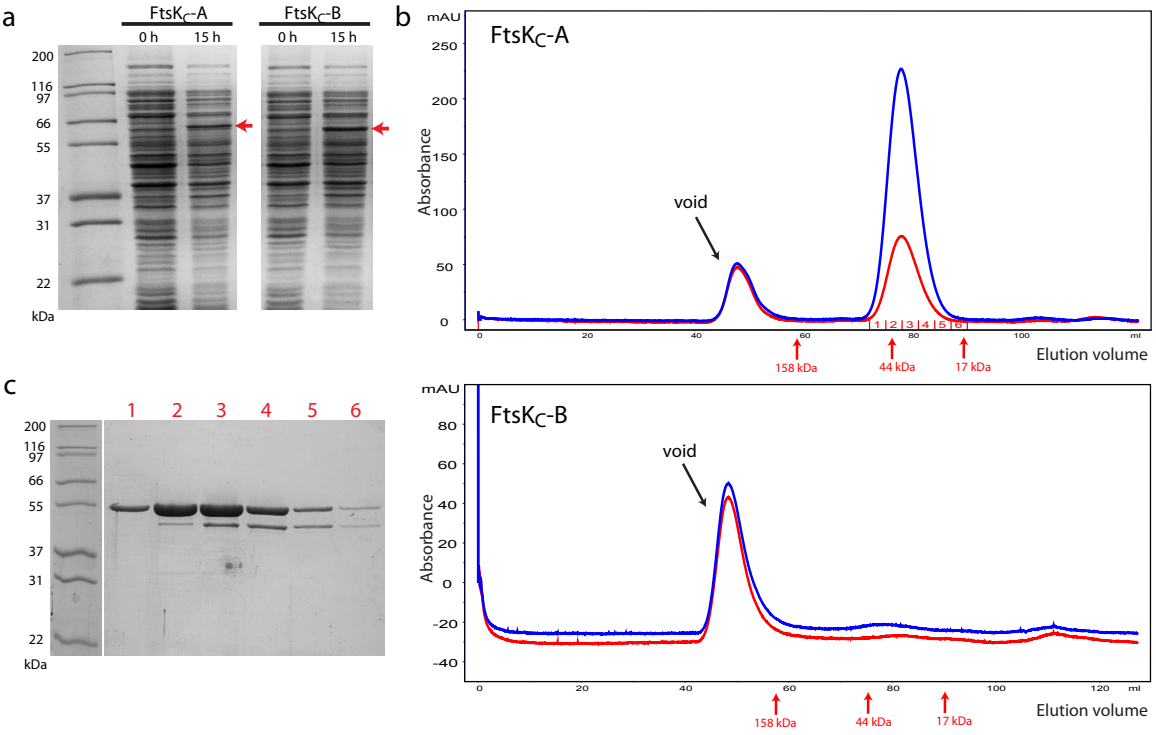


Figure 5-4: Overexpression and purification of $FtsK_C$ constructs. a) Overexpression of $FtsK_C$-A and $FtsK_C$-B in *E. coli* strain BL21 (DE3), as analysed by SDS-PAGE. The samples were analysed immediately after induction (0 h) and 15 hours later (15 h). The positions of the expressed $FtsK_C$ proteins on the gel are indicated by red arrows. b) The final step of $FtsK_C$ purification: SEC on a Superdex 16/60 gel filtration column. The chromatograms show elution peaks that correspond to protein aggregates (marked as void) or a monomer of $FtsK_C$-A (54 kDa). UV absorbance at 280 nm is shown in blue, absorbance at 260 nm is shown in red. The elution volumes of molecular weight markers of 17, 44, and 158 kDa on this column are indicated by red arrows. The fractions collected are marked in red under the elution peak. c) SDS-PAGE analysis of the fractions collected from the gel filtration, stained with Coomassie for the presence of protein.

5.2.2 *In vitro* XerH recombination with purified $FtsK_C$-A

Initial attempts to reconstitute the full XerH recombination process *in vitro* were unsuccessful in the presence of XerH only, even in conditions supporting XerH cleavage and HJ resolution (see section 3.2.11). This might be because additional factors are required for reconstitution of XerH recombination *in vitro*, as seen for *E. coli* XerC/D recombination (Aussel *et al.,* 2002). To test this, an XerH recombination assay with dif_H-containing linear and plasmid DNA substrates was performed in the presence of the purified $FtsK_C$-A and ATP. Still no recombination was observed in any of the conditions tested, including conditions that supported cleavage and HJ resolution by XerH and conditions that supported XerC/D recombination *in vitro*.

5.2.3 Pull-down assay with purified XerH and $FtsK_C$-A

Since no dif_H recombination could be observed *in vitro* in the presence of XerH and $FtsK_C$, it was interesting to see if the two proteins interact with each other at all. To do that, purified XerH, which shows affinity for nickel in low-salt conditions, was bound to Ni-NTA agarose beads in HEPES buffer containing 250 mM NaCl, as described in section 2.4.9. After a series of washes to remove unbound XerH, purified $FtsK_C$-A (which does not bind nickel by itself) was added to the beads and the suspension was incubated for 1 hour on a rocking platform. If XerH and $FtsK_C$-A directly interacted with each other in these conditions, some of the $FtsK_C$-A protein would be retained on the beads through interaction with XerH. Unbound $FtsK_C$-A was removed by a series of washes. Finally, bound proteins were eluted from the resin by addition of imidazole. Samples for analysis by SDS-PAGE were collected at every step and the results are shown in Figure 5-5. The control with XerH only showed that XerH in fact bound to the nickel beads and eluted upon addition of imidazole (Figure 5-5a), while the control with $FtsK_C$-A only showed no binding of the protein to the resin (Figure 5-5b). In the presence of both proteins, the elution contained only XerH, indicating that $FtsK_C$-A was not retained on the beads, probably because it did not directly interact with XerH in these conditions (Figure 5-5c). This result did not change when a DNA fragment containing the dif_H site was also included in the experiment (Figure 5-5d).

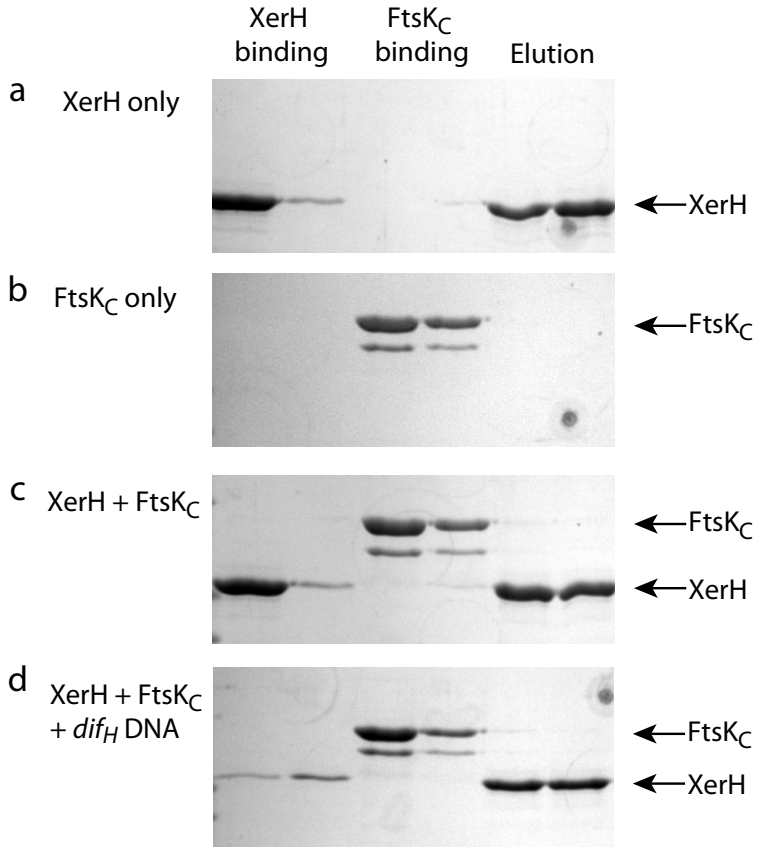


Figure 5-5: SDS-PAGE analysis of a pull-down assay with purified XerH and FtsK$_C$-A proteins. The samples were analysed after XerH was incubated with the nickel beads (XerH binding), after addition of FtsK$_C$-A to the XerH-bound beads (FtsK$_C$ binding), and after elution of the bound proteins (Elution). Two samples collected consecutively in each step were analysed. (a) and (b) The control experiments where only XerH (a) or FtsK$_C$-A (b) was applied to the nickel resin. (c) Experiment with both XerH and FtsK$_C$-A added to the resin. (d) Experiment with both XerH and FtsK$_C$-A and a *dif$_H$*-containing DNA fragment added to the resin.

5.2.4 Overexpression and purification of the XerH-FtsKγ fusion protein

Since no XerH recombination in the presence of FtsK$_C$-A and no direct interaction between XerH and FtsK$_C$-A could be observed, another FtsK construct was designed. For successful activation of XerD monomer and therefore recombination *in vitro,* only the most C-terminal γ subdomain of *E. coli* FtsK$_C$ is required (Grainge *et al*., 2011). This subdomain of the ATP helicase domain is moderately conserved amongst bacteria employing different Xer recombination systems (Figure 5-2b) and it is this subdomain that was shown to confer species-specificity of Xer-FtsK interactions (Yates *et al*., 2006). To investigate if *H. pylori* FtsKγ could support XerH recombination *in vitro*, a fusion protein containing the full-length XerH and C-terminally attached FtsKγ (residues 801-863 of FtsK, Figure 5-6) separated by a glycine-rich flexible linker was designed based on a previously described, recombination-proficient XerD-FtsKγ fusion (Figure 5-7a; Grainge

et al., 2011). This construct was introduced into the pETM-28 vector that was previously used for overexpression of XerH alone.

Given the similarity between the XerH and XerH-FtsKγ proteins, the fusion protein was overexpressed in *E. coli* and purified following the same protocols that were previously used for overexpression and purification of XerH. Overexpression in *E. coli* strain BL21 (DE3) yielded large amounts of protein (Figure 5-7b), which was then affinity-purified on a HisTrap column. The 6xHis-SUMO tag was removed by SenP2 cleavage and a second purification on a HisTrap column. The tag-free XerH-FtsKγ was applied to a SEC column, from which the protein eluted in the main peak corresponding in size to an XerH-FtsKγ monomer (50 kDa; Figure 5-7c). Collected fractions were analysed by SDS-PAGE (Figure 5-7d) and fractions 4-9 were combined for further biochemical studies. This protocol yielded 5 mg of purified protein from one litre of the expression culture.

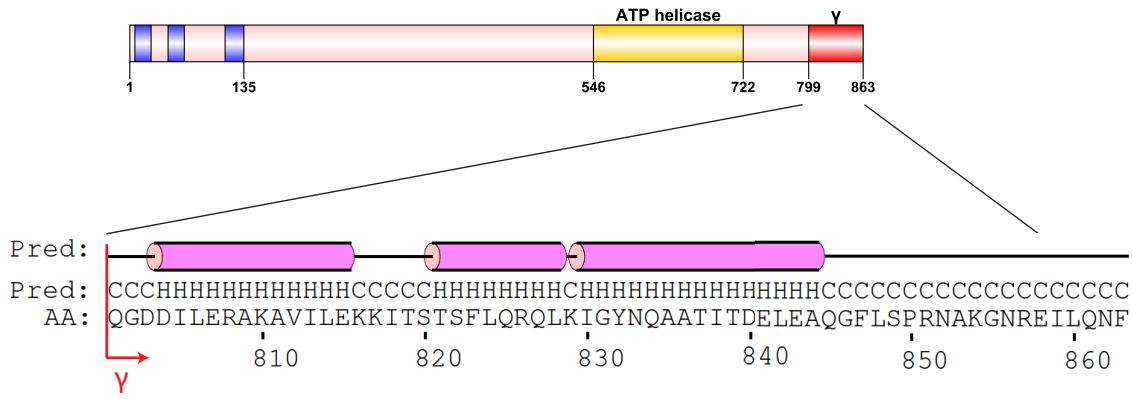


Figure 5-6: The design of *H. pylori* FtsKγ construct. Top: schematic domain representation of *H. pylori* P12 FtsK, coloured as in Figure 5-1. Bottom: zoomed-in view of residues 801-863 of *H. pylori* P12 FtsK with secondary structure predictions prepared using PSIPRED Protein Sequence Analysis Workbench (Buchan *et al.*, 2013). The amino acid sequence (AA) is accompanied by the secondary structure prediction for each residue (Pred) expressed as a letter (H: α-helix and C: coil, loop, or unstructured region) and as a graphical representation (magenta cylinder: α-helix). The start of the FtsKγ region used to produce the XerH-FtsKγ fusion protein is marked in red.

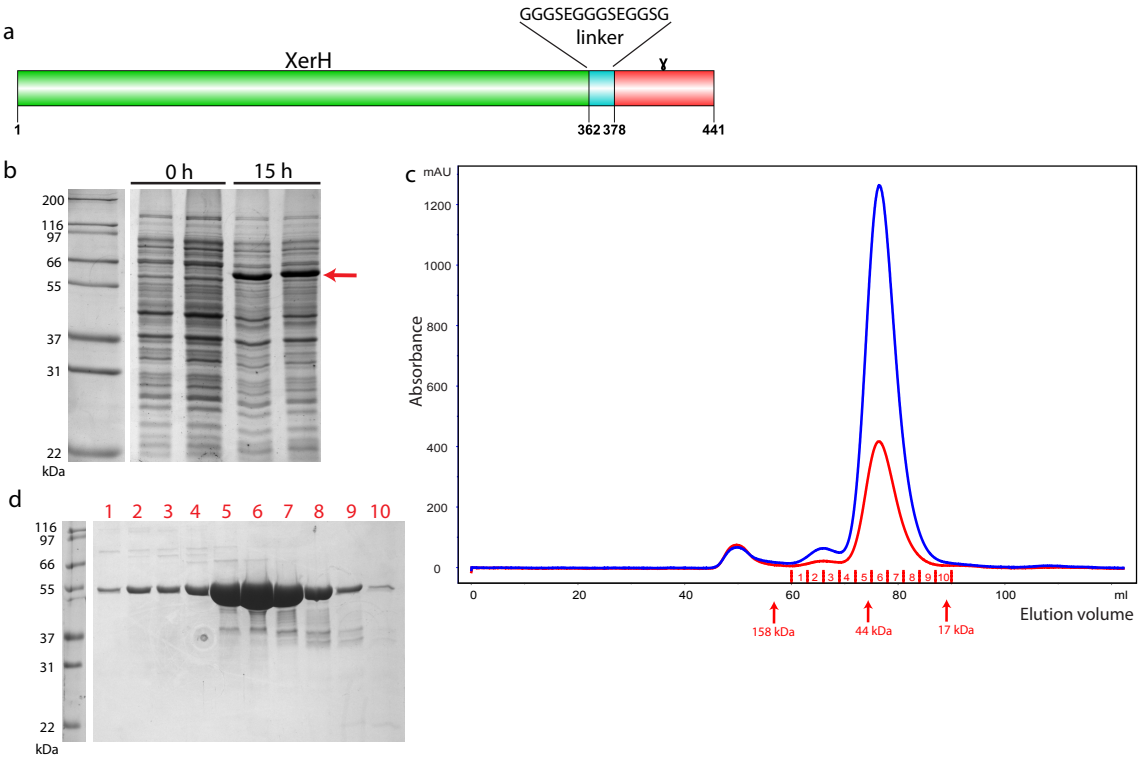


Figure 5-7: Overexpression and purification of the XerH-FtsKγ fusion protein. a) Domain representation of the fusion protein. b) Overexpression of XerH-FtsKγ in *E. coli* strain BL21 (DE3), as analysed by SDS-PAGE. The samples were analysed immediately after induction (0 h) and 15 hours later (15 h). The position of the expressed XerH-FtsKγ proteins on the gel is indicated by a red arrow. b) The final step of XerH-FtsKγ purification: SEC on a Superdex 16/60 gel filtration column. The chromatogram shows a main elution peak corresponding to a monomer of XerH-FtsKγ (50 kDa). UV absorbance at 280 nm is shown in blue, absorbance at 260 nm is shown in red. The elution volumes of molecular weight markers of 17, 44, and 158 kDa on this column are indicated by red arrows. The fractions collected are shown in red under the elution peak. c) SDS-PAGE analysis of the fractions collected from the gel filtration, stained with Coomassie for the presence of protein.

5.2.5 Recombination activity of the XerH-FtsKγ fusion protein

The biochemical activity of the XerH-FtsKγ fusion protein was assessed *in vitro*. Initially, the protein was used in the DNA cleavage assay as described for wild-type XerH (see Figure 3-14b and c) in order to see if the added C-terminal linker and FtsKγ did not interfere with XerH activity. Additionally, this assay was expected to show whether FtsKγ fused to XerH can promote the activity switch of XerH monomers as observed in XerC/D recombination, where in the absence of FtsK the XerC subunit cleaves the DNA first, whereas in the presence of FtsK the XerD subunit is activated and performs the first DNA cleavage (Aussel *et al.,* 2002). In the cleavage assay with pre-nicked suicide oligos (shown in Figure 5-8a), cleavage mediated by XerH-FtsKγ followed the same pattern as that of wild-type XerH: the substrate with a bottom strand nick (cleaved by the monomer bound to the right half-site of dif_H; lane 2 in Figure 5-8b) was cleaved much more efficiently than

the substrate with a top strand nick (cleaved by the monomer bound to the left half-site; lane 1 in Figure 5-8b). A substrate with a double nick was cleaved by both monomers with a similar efficiency (lane 3 in Figure 5-8b). An additional, higher molecular weight band could be seen when these substrates were used, possibly corresponding to XerH-FtsKγ attached to DNA that was not properly denatured. These results suggest that the FtsKγ subdomain fused to XerH did not induce the activity switch of XerH monomers, which would be seen as higher levels of cleavage in the presence of DNA substrates with top strand nick compared to the DNA substrates with bottom strand nick.

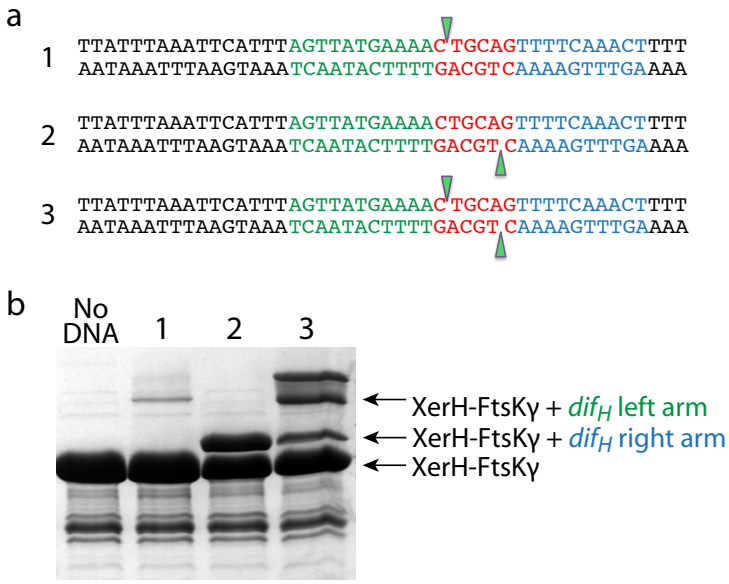


Figure 5-8: *In vitro* cleavage activity of the XerH-FtsKγ fusion protein. a) Suicide substrates used to assess cleavage activity of purified XerH-FtsKγ, containing nicks in top, bottom, or both strands of dif_H (marked with green triangles), and a longer left dif_H arm to distinguish between the left and right arm cleavage products. The left and right arms of dif_H are coloured green and blue, respectively. The central region of dif_H is coloured red. b) SDS-PAGE analysis of the cleavage reactions with XerH-FtsKγ and dif_H suicide substrates. The numbers above the gel correspond to the number of DNA substrate shown in (a).

Furthermore, linear and circular DNA substrates containing dif_H site were used in an *in vitro* recombination assay with the XerH-FtsKγ fusion protein in conditions supporting cleavage of suicide substrates. However, similarly to all previous attempts, no recombination product could be detected in any of the conditions tested.

5.2.6 XerH plasmid recombination with *galK* marker in *E. coli*

To further investigate the regulation and requirements for XerH recombination, an *in vivo* assay based on *galK* selection on MacConkey agar was performed in *E. coli* (Arnold *et al.,* 1999). The design of the assay is shown in Figure 5-9a. In detail, *E. coli galK*-deficient strain DS941 was transformed with a reporter plasmid pAB102 (see Table 2-2) carrying two dif_H sites in a direct repeat orientation flanking a functional *galK* gene. Upon expression of XerH from a helper plasmid XerH_pBAD/MCS (see Table 2-2) under the control of an arabinose promoter, the two dif_H sites on the reporter vector can get recombined resulting in excision of a non-replicative circular DNA carrying the *galK* gene. This vector, and therefore the *galK* gene, is lost from daughter cells during cell division. The presence or absence of the functional *galK* gene is then assessed by plating the bacteria on MacConkey agar supplemented with galactose. Bacteria that carry *galK* can utilize galactose and will produce acid as a metabolic product, which in turn will reduce the pH of the surrounding medium and result in red colour of colonies on the MacConkey agar plates. Bacteria that cannot utilize galactose will instead utilize peptone present in the medium, which in turn leads to formation of ammonia and increase of the pH, resulting in white appearance of the colonies on the MacConkey agar (Figure 5-9a).

The assay was performed as described in section 2.6.2 in the presence or absence of XerH expression plasmid and appearance of white colonies was assessed after retransformation of the reporter plasmid and plating on MacConkey agar supplemented with kanamycin. White colonies were detected in the assay only in the presence of XerH, suggesting that XerH-mediated recombination took place (Figure 5-9b). To validate that, the reporter plasmid DNA from the single red and white colonies was analysed on an agarose gel (Figure 5-9c). The analysis of the reporter plasmid from the red colonies showed a band corresponding in size to the supercoiled non-recombined reporter plasmid (5042 bp), while the reporter plasmid from the white colonies was shorter (3197 bp) and represented the recombined plasmid missing the *galK* gene. In order to confirm that the loss of *galK* gene was in fact a result of XerH site-specific recombination (and not, for example, of homologous recombination) the plasmid DNA from red and white colonies was sequenced. The results were consistent with XerH recombination: white colonies showed a single dif_H site flanked by the sequences that in the non-recombined reporter plasmid flanked the dif_H-*galK*-dif_H cassette. Together, these results show that XerH can recombine plasmid-borne

dif$_H$ sites in *E. coli*, indicating that no *H. pylori*-specific factors are required for this recombination.

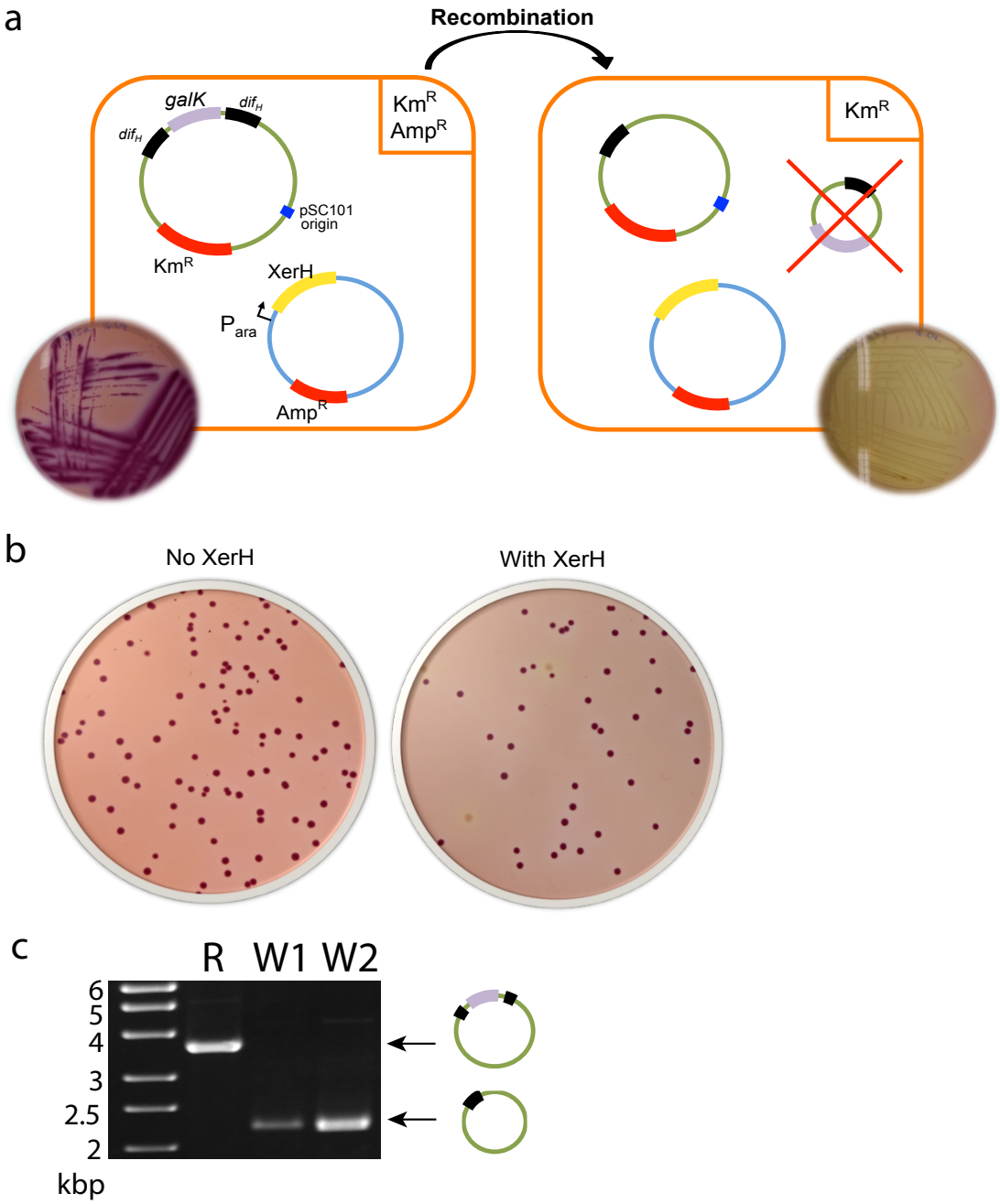


Figure 5-9: *galK*-MacConkey XerH recombination assay in *E. coli*. a) The design of the recombination assay. *galK*-deficient *E. coli* strain is transformed with the reporter plasmid (containing two directly repeated *dif*$_H$ sites flanking a functional *galK* gene) and with the XerH expression plasmid. Upon XerH expression, the two *dif*$_H$ sites are recombined and a small, non-replicative DNA circle containing the *galK* gene is excised and lost from the cell population. Cells carrying the *galK* gene (no XerH recombination) and those without the *galK* gene (post XerH recombination) can be distinguished by plating on MacConkey agar plates, where the two kinds of cells grow red or white colonies, respectively. b) Photographic images of MacConkey agar plates, showing the results of the recombination assay in *E. coli* cells carrying the *dif*$_H$-*galK*-*dif*$_H$ reporter plasmid in the presence or absence of XerH expression plasmid (as indicated above each plate). c) Analysis of plasmid DNA extracted from red (R) and white (W) colonies observed in the recombination assay. The supercoiled plasmid DNA was analysed on a 0.8% agarose gel. The expected positions of the parent and recombination product plasmids are marked next to the gel.

5.2.7 XerH recombination in FtsK$_C$-deficient strain of *E. coli*

XerH-mediated plasmid recombination worked efficiently in *E. coli*, suggesting that if any accessory factor is required for XerH recombination of plasmid-borne *dif$_H$* sites, it is present in the *E. coli* strain DS941. In order to test whether *E. coli* FtsK is required for XerH plasmid recombination, the recombination assay with *galK* marker was performed in *E. coli* strains DS941 (wild-type) and DS9041, a derivative of DS941 that carries an FtsK$_C$ deletion and was previously used to show that FtsK$_C$ is required for XerC/D recombination (Recchia *et al.,* 1999). In both strains, white colonies were detected at similar frequencies, suggesting similar recombination rates (Figure 5-10a and b). The rates of recombination above 10% observed even without arabinose induction suggest that recombination is very efficient in both wild-type and mutant *E. coli* strains. The reporter plasmid DNA from the red and white colonies obtained was again analysed on an agarose gel, which showed that red colonies carry intact reporter plasmid, while white colonies carry a smaller plasmid that has lost the *galK* gene as a result of XerH recombination (Figure 5-10c). This was also confirmed by sequencing of the reporter plasmid from white colonies.

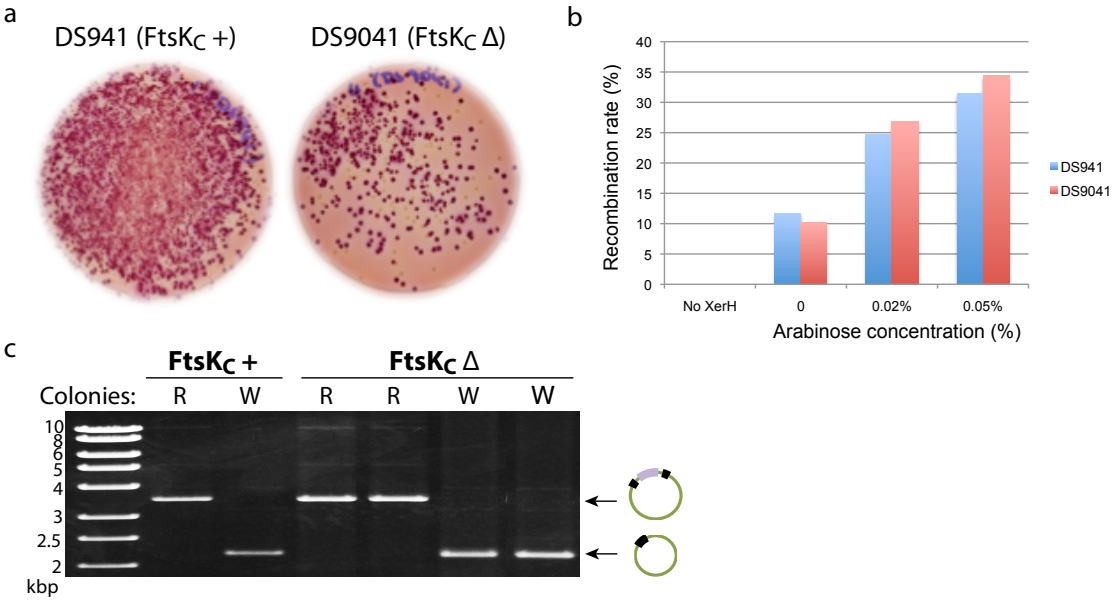


Figure 5-10: *galK*-MacConkey XerH recombination assay in wild-type (DS941) and FtsK$_C$-deficient (DS9041) *E. coli* strains. a) Appearance of the MacConkey agar plates after XerH recombination assay. Recombination was allowed to take place in either DS941 or DS9041 carrying the *galK* reporter plasmid and the arabinose inducible XerH expression plasmid in the presence of different concentrations of arabinose. DNA was isolated and retransformed into DS941 or DS9041 and the cells were plated on MacConkey galactose plates. The plates shown are from the experiment with 0.02% arabinose. b) Recombination efficiency in DS941 and DS9041 *E. coli* strains, calculated by dividing the number of white colonies by the total number of colonies obtained on MacConkey agar plates such as those shown in (a). c) Analysis of reporter plasmid DNA extracted from red (R) and white (W) colonies obtained in the recombination assay. The supercoiled plasmid DNA was analysed on a 0.8% agarose gel. The expected positions of the parent and recombination product plasmids are marked next to the gel.

5.2.8 Recombination assay in liquid culture

The *galK*-MacConkey assay is a very powerful method to detect site-specific recombination and quantify recombination rates. However, the recombination events cannot be observed in a time course or over longer growth periods (up to 80 generations). The wild-type and $FtsK_C$-deficient *E. coli* strains appeared to support the same levels of XerH recombination, but it was also possible that the differences between the two strains could only be seen after longer *E. coli* growth than that in the *galK*-MacConkey assay. Therefore, to complement the results of the *galK*-MacConkey assay and check whether the absence of $FtsK_C$ has an effect on XerH recombination over longer growth periods, a recombination assay in liquid culture was performed. A reporter plasmid pAB122 (see Table 2-2) containing two dif_H sites in direct repeat orientation separated by approximately 1000 bp of DNA was co-transformed with the arabinose-driven XerH expression plasmid XerH_pBAD/MCS (see Table 2-2) into *E. coli* strains DS941 and DS9041. Double transformants were then grown in serial cultures supplemented with arabinose for approximately 80 generations as described in section 2.6.3. Samples were collected every 8 generations, plasmid DNA was extracted and analysed on an agarose gel. Figure 5-11a shows the agarose gel analysis of selected samples. In both DS941 ($FtsK_C$ +) and DS9041 ($FtsK_C$ Δ) strains, similar levels of the recombination product could be detected at all times, including low levels at the beginning of the assay suggesting that a small amount of recombination occurs even before arabinose induction (in agreement with the observations from the *galK* recombination assay, Figure 5-9b). In both strains, higher molecular weight products (likely plasmid catenanes and multimers) appeared as the assay proceeded. These were more prominent in the DS9041 ($FtsK_C$ Δ) strain, but most of them are likely unrelated to XerH recombination as they were also formed in the absence of the reporter plasmid or the expression plasmid (Figure 5-11b). Together, these results suggest that plasmid recombination in *E. coli* liquid culture is equally efficient in the presence and absence of *E. coli* $FtsK_C$.

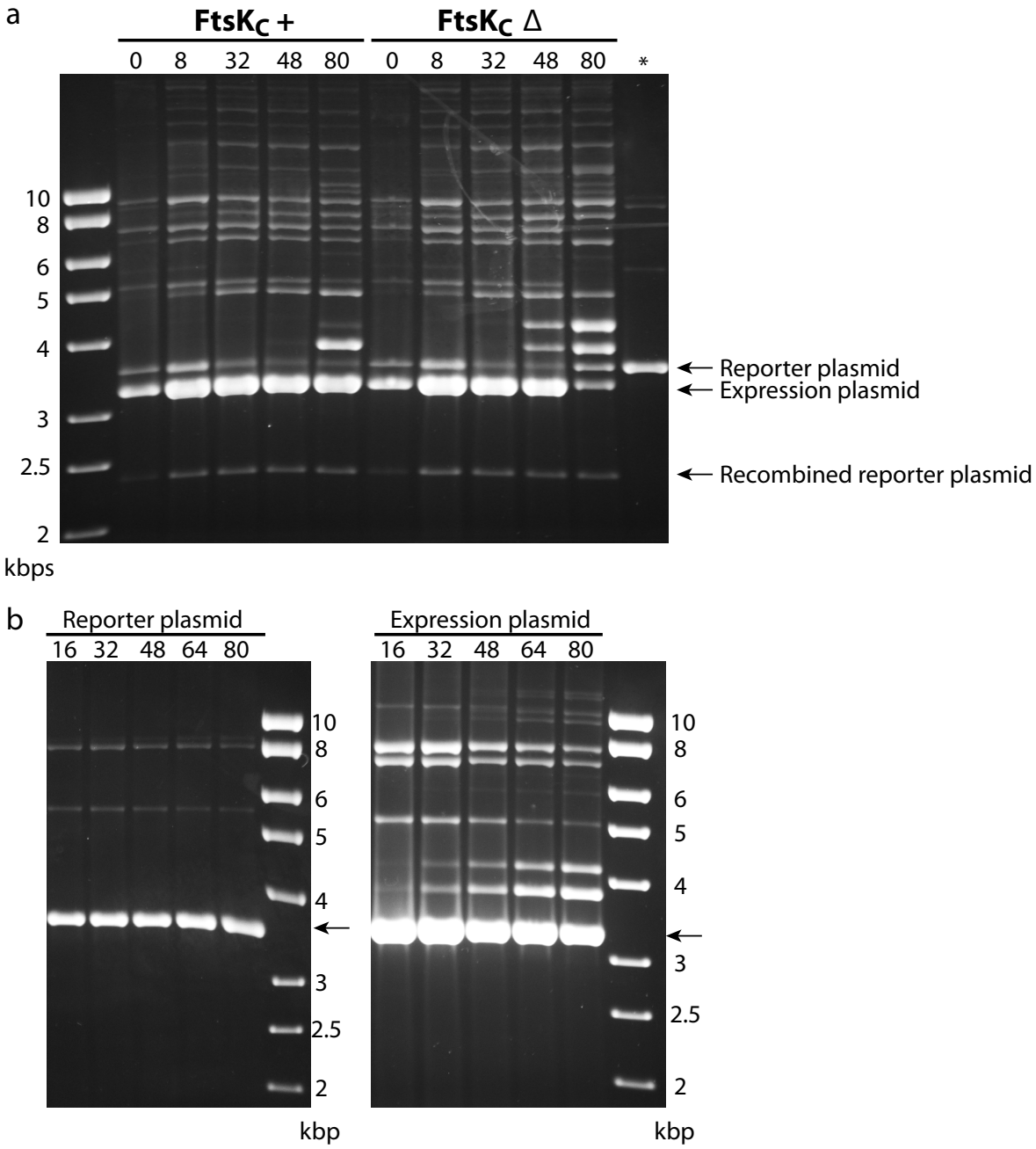


Figure 5-11: XerH recombination assay in *E. coli* liquid culture. a) Analysis of recombination products from *E. coli* strains DS941 and DS9041 on a 0.8% agarose gel. The supercoiled reporter and expression plasmids are marked together with the recombination product (recombined reporter plasmid). The time point at which each sample was collected is shown above the gel as a number of *E. coli* generations, assuming growth to saturation constitutes approximately 8 generations. The unmarked higher molecular weight bands likely correspond to catenated and multimerized plasmids. (*) denotes the reporter plasmid control. b) and c) Control assays performed in *E. coli* strain DS9041in the same conditions as (a) but with reporter plasmid (b) or XerH expression plasmid (c) only. Respective plasmid bands are marked by arrows. The unmarked higher molecular weight bands likely correspond to multimers and catenanes of the respective plasmids.

5.2.9 Affinity chromatography of XerH and *E. coli* lysate

The *E. coli* DS941 background is sufficient to support XerH recombination of plasmid-borne dif_H sites, even in the absence of $FtsK_C$, yet no recombination could be reconstituted

with XerH and $FtsK_C$ constructs *in vitro*. This suggests that an additional factor might be required for XerH recombination and that this factor is present in *E. coli* DS941. To identify potential XerH regulators, an affinity chromatography assay using XerH as bait for *E. coli* DS941 proteins was performed. In this assay, purified XerH was covalently bound to CNBr-activated Sepharose beads and then incubated with the soluble or membrane fraction of *E. coli* lysate at 50, 150, and 400 mM NaCl as described in section 2.4.10. Unbound proteins were then washed from the beads and proteins interacting with XerH were retained. Subsequently, the *E. coli* proteins interacting with XerH were eluted with a high-salt buffer and analysed by mass spectrometry to reveal their identity. The same steps were applied to beads with bound XerH ("sample") and to XerH-free beads ("control"). A total of 1261 proteins were initially identified in all the experiments, with the specific numbers for each experiment shown in Table 5-1.

Table 5-1: Results of mass spectrometry identification of *E. coli* proteins interacting with XerH in the affinity chromatography experiment.

Experiment	Number of identified proteins[a]	Number of proteins used for further analysis[b]
Soluble 50 mM NaCl	624	292
Soluble 150 mM NaCl	884	160
Soluble 400 mM NaCl	750	227
Membrane 50 mM NaCl	739	419
Membrane 150 mM NaCl	469	124
Membrane 400 mM NaCl	584	200

Soluble – soluble fraction of *E. coli* lysate; membrane – lipidic fraction of *E. coli* lysate.

[a] – Total number of unique proteins identified by mass spectrometry in each affinity chromatography experiment, including proteins identified in the sample and the control.

[b] – A subset of identified proteins used for further analysis, selected based on one of the following criteria: (a) in cases where the protein was absent from the control, the read count in the sample was higher than 5 spectral counts or (b) in cases when the protein was found in both the control and the sample, at least 5-fold enrichment was observed in the sample.

Of all the identified proteins, only a subset was considered for further analysis (Table 5-1) based on the following criteria: (a) in cases where the protein was absent from the control, the read count in the sample was higher than 5 spectral counts (*i.e.* the peptides assigned to the protein were seen at least 5 times in the mass spectrometry analysis) and (b) in cases where the protein was found in both the control and the sample, at least 5-fold enrichment was observed in the sample. From this subset, proteins showing consistent enrichment in the sample across all or most experimental conditions were analysed in terms of their function as candidate XerH regulators. For example, protein MukF fulfilled the criterion (a) in all three experiments with the membrane fractions of *E. coli* lysate, and showed

enrichment in the sample compared to the control in the remaining experiments with soluble fractions (Table 5-2), and was therefore selected for further analysis in terms of function. Generally, proteins that are involved in gene transcription, protein synthesis, and metabolic functions were excluded from further analysis as unlikely to participate in Xer recombination. The remaining identified proteins (Table 5-2) included eight proteins involved in *E. coli* genome maintenance, two proteins of unrelated function, and four proteins of unknown function. Interestingly, of the *E. coli* Xer recombinases only XerD fulfilled the chosen selection criteria (Table 5-2). Moreover, consistently with the previous results, FtsK did not appear to directly interact with XerH (Table 5-2). In addition to a biologically relevant function, the potential XerH regulator should have a homologue in *H. pylori*. Therefore, such homologues were searched for using the *H. pylori* genome browsers and in the literature, with the outcome shown in Table 5-2. In summary, the affinity chromatography experiment revealed a number of interesting candidates that could be involved in XerH recombination in *E. coli* and in *H. pylori*, the functional significance of which will be interesting to test in future follow up experiments.

Table 5-2: *E. coli* proteins selected as potential activators of XerH recombination.

***E. coli* protein interacting with XerH**	**Total spectral counts (s/c)[a]**						**Function**	***H. pylori* homologue**
	Soluble fraction			**Membrane fraction**				
	50	**150**	**400**	**50**	**150**	**400**		
Chromosome dimer resolution proteins								
XerD	7/4	12/0	4/0	10/0	4/0	0/0	Resolution of chromosome dimers	yes
XerC[b]	9/0	0/0	0/0	0/0	0/0	0/0	Resolution of chromosome dimers	yes
FtsK[b]	4/0	2/0	5/0	3/0	0/0	0/0	DNA translocase activating XerC/D recombination	yes
Genome maintenance proteins								
HsdR	28/0	7/3	26/0	4/0	0/0	0/0	Restriction/modification system	yes
MukE	11/0	13/0	6/0	6/0	0/0	0/0	Chromosome partitioning protein	no
MukF	22/4	30/9	32/7	22/0	10/0	6/0	Chromosome partitioning protein	no
DnaB	29/8	24/3	39/10	24/3	15/0	35/0	Replicative helicase	yes
ObgE	45/5	45/7	10/0	31/6	14/5	14/0	GTPase required for chromosomal segregation	yes
ParC	76/16	51/9	24/10	45/8	16/7	12/0	DNA topoisomerase IV subunit A	no
TopA	117/10	62/15	38/16	48/10	16/4	7/0	DNA topoisomerase	yes
RuvB	0/0	3/0	12/0	0/0	0/0	7/0	HJ ATP-dependent DNA helicase	yes
Other function								
YjeE (TsaE)	2/0	6/2	10/0	6/0	5/0	11/0	ATPase in TSA interaction network	no
RdgC	22/0	14/8	4/0	13/0	4/0	7/0	RecA regulator	no
Unknown function								
YggL	14/0	20/13	22/0	23/0	20/0	27/8	Unknown	?
YgiQ	2/0	5/0	15/0	3/0	0/0	0/0	Unknown	?
YedD	0/0	0/0	0/0	13/0	10/0	12/0	Unknown	?
YfjD	0/0	0/0	4/0	5/0	2/0	9/0	Unknown	?

[a] – The values of 50, 150, and 400 indicate the amount of NaCl (mM) used in each affinity chromatography experiment. The two numeric values in these columns (s/c) correspond to total spectral counts in the sample (s) and the control (c) experiments.

[b] – XerC and FtsK spectral counts suggest no interactions with XerH and the two proteins are shown for comparison only.

5.2.10 Chromosomal XerH recombination in *E. coli*

Previous results showed that XerH can recombine plasmid-borne dif_H sites in *E. coli*, suggesting that XerH regulation is not species-specific and all regulatory factors are present in *E. coli*. However, recombination of plasmid-borne dif_H sites might require different, less stringent regulation and have a different temporal and spatial context than chromosome dimer resolution. Therefore, XerH recombination activity was also tested on a dif_H site replacing the native *dif* site on the *E. coli* chromosome. To do this, a gene cassette containing a 33-bp dif_H site and a kanamycin resistance gene was introduced into the *E. coli* chromosome by homologous recombination using the Red recombination system (Datsenko and Wanner, 2000) so that the dif_H-Km cassette replaced the native *dif* site (Figure 5-12a; section 2.6.4). The obtained strain DS941dif_H (see Table 2-1) with or without an XerH expression plasmid were analysed under the microscope to assess their phenotype in terms of filament formation (Figure 5-12b). As a control for the Xer-deficiency phenotype, a strain with deleted *dif* (DS941*Δdif*, see Table 2-1) was analysed. The results shown in Figure 5-12b reveal that expression of XerH could not rescue the filamentous phenotype, suggesting that chromosome dimer resolution at dif_H did not occur in *E. coli* in the presence of XerH.

Furthermore, to directly assess the efficiency of XerH recombination on chromosome-borne dif_H sites, a dif_H-Km-dif_H cassette with directly repeated dif_H sites flanking a kanamycin resistance gene was introduced into the *E. coli* DS941 and DS9041 genomes, where it replaced the *E. coli dif* site (Figure 5-13a). The strains DS941dif_H-Km-dif_H and DS9041dif_H-Km-dif_H were transformed with an XerH expression plasmid XerH_pBAD/MCS (see Table 2-2) and the transformant colonies were grown in LB without antibiotic selection in the presence or absence of arabinose until OD_{600} of 1.5. Next, dilutions of the cultures were plated on LB agar containing ampicillin to maintain the expression plasmid. 80 of the obtained colonies from each experiment (both strains grown initially with or without arabinose) were restreaked on kanamycin plates to assess in what fraction of the cells the dif_H-Km-dif_H cassette was lost as a result of XerH recombination. All restreaked colonies grew on kanamycin plates, indicating that no recombination between the two dif_H sites occurred in either DS941 or DS9041 *E. coli* background. This was confirmed by analysis of some of the colonies by colony PCR, which showed that the dif_H-Km-dif_H cassette was retained in all cases (Figure 5-13a). Altogether, these results

indicate that no XerH recombination occurs in *E. coli* on chromosome borne-dif_H sites replacing the native *dif*.

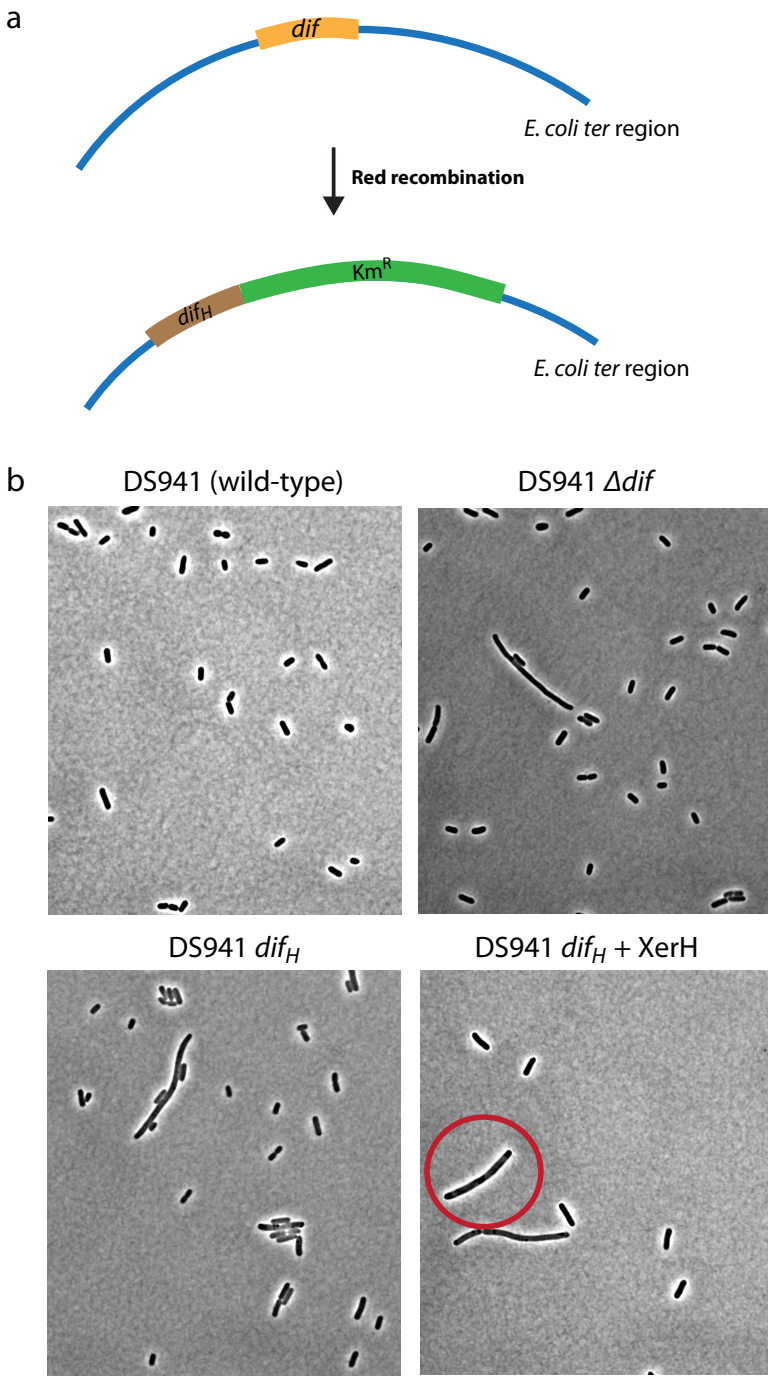


Figure 5-12: Phenotypic analysis of *E. coli* strains carrying dif_H replacing the native *dif* site. a) Schematic view of the native *E. coli dif* site in the chromosome *ter* region and the same region after replacement of *dif* with a dif_H-Km cassette. b) Phenotypic analysis of the derivatives of *E. coli* DS941 strain: wild-type DS941, DS941Δ*dif* (*dif* replaced by Km cassette), DS941dif_H (with *dif* replaced by dif_H-Km cassette), and the same strain in the presence of XerH expression plasmid (grown in the presence of arabinose to induce XerH expression). Representative microscopic images are shown. An example multicellular filament is circled in red. The images were taken using a Zeiss CellObserver microscope with phase contrast and 1000x magnification.

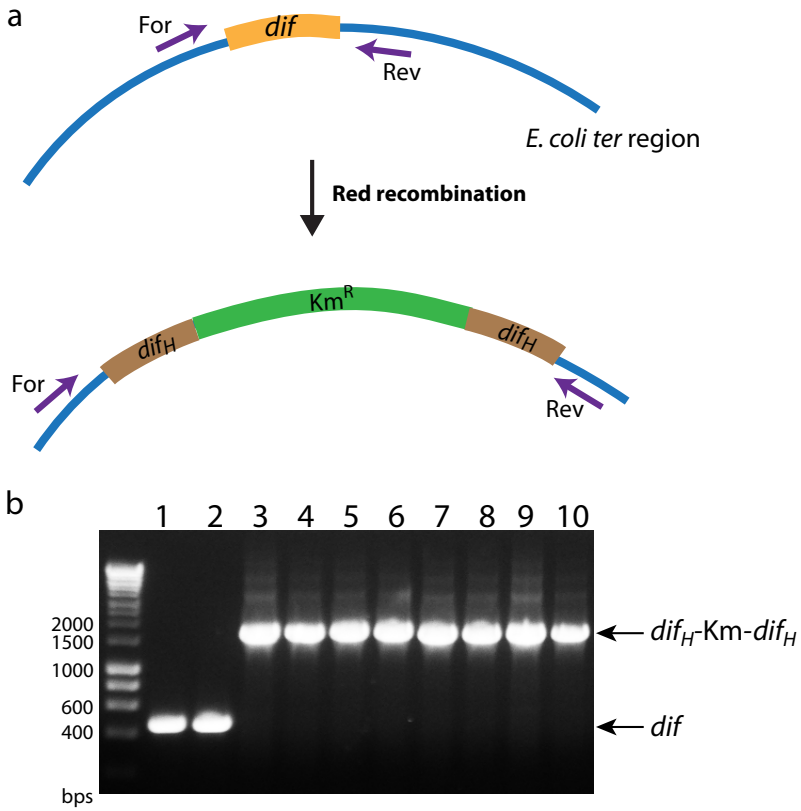


Figure 5-13: XerH recombination assay with a chromosomal dif_H-Km-dif_H cassette. a) Schematic view of the native *E. coli dif* site in the chromosome *ter* region and the same region after replacement of *dif* with a dif_H-Km-dif_H cassette. The forward (For) and reverse (Rev) primers used for colony PCR (see Table 2-3) are shown as purple arrows. b) Analysis of colony PCR products obtained using primers shown in (a) following an XerH recombination assay (1.2% agarose gel). Lane 1: DS941 control; lane 2: DS9041 control; lane 3: DS941dif_H-Km-dif_H control; lanes 4-6: DS941dif_H-Km-dif_H with XerH expression plasmid after recombination assay; lane 7: DS9041dif_H-Km-dif_H control; lanes 8-10: DS9041dif_H-Km-dif_H with XerH expression plasmid after recombination assay.

5.3 Discussion

5.3.1 Hints for external regulation of XerH recombination

All bacterial Xer recombinases studied to date require additional regulatory factors either for direct activation of one of the Xer monomers or for assuring the correct topology of recombination substrates and products. Such regulation assures that Xer recombination happens in a well-defined temporal and spatial context, which in turn guarantees that only the recombination sites present on the dimeric chromosome are recombined. This prevents formation of unwanted chromosome dimers instead of their resolution. Since XerH was shown to resolve chromosome dimers in *H. pylori* (Debowski *et al.,* 2012), it is likely that XerH recombination is also coupled to cell division by an additional protein. This notion is

supported by the observed lack of XerH recombination *in vitro* in the absence of additional proteins (see section 3.2.11), although this failure to reconstitute XerH recombination might also be caused by unfavourable assay conditions chosen. At the same time, the resolution of a HJ intermediate by XerH could be reconstituted *in vitro*, suggesting that the *in vitro* conditions used support XerH-mediated cleavage and strand exchange activity. However, the HJ resolution assay resulted only in conversion of HJ into substrates and not to products (see Figure 3-17), suggesting that one of the conformations of the synaptic complex (where the right arm-bound XerH monomer cleaves the DNA) is more favourable than the other, and remodelling of the complex is required for progression of the reaction. As seen in the structure of the XerH post-cleavage synaptic complex (Figure 4-20), such remodelling (isomerization) of the synaptic complex might be energetically challenging due to the lack of pseudo four-fold symmetry and therefore might require additional factors (discussed in section 4.3.6). Based on these data, it is feasible to infer that an additional factor activating and regulating XerH recombination is required for full recombination.

5.3.2 Lack of interactions between XerH and FtsK *in vitro*

Based on analogy to the XerC/D recombination system, it is plausible that *H. pylori* FtsK could be a critical regulatory factor required for XerH recombination. Accordingly, two distinct constructs of *H. pylori* $FtsK_C$ were designed, overexpressed in *E. coli*, and purified to homogeneity in order to test if *H. pylori* $FtsK_C$ can activate XerH recombination *in vitro*. Both constructs were designed based on sequence alignments with *E. coli* FtsK. The first construct, $FtsK_C$-A, corresponded to *E. coli* $FtsK_C$ and contained the whole predicted ATP helicase domain (Figure 5-3). Equivalent constructs from various species were shown to support XerC/D recombination *in vitro* and *in vivo* (Barre *et al.,* 2000) as well as XerS recombination *in vivo* (Le Bourgeois *et al.,* 2007). The second construct was a fusion protein XerH-FtsKγ, where FtsKγ subdomain was attached by a flexible linker to XerH to help their interaction (Figures 5-6 and 5-7a). FtsKγ is the most C-terminal subdomain of the FtsK ATP helicase domain, and was shown to be the part of FtsK required for direct interaction with XerD in *E. coli* (Grainge *et al*., 2011). The *E. coli* XerD-FtsKγ fusion protein was shown to be proficient in *dif* recombination *in vitro* and *in vivo* (Grainge *et al*., 2011).

Interestingly, neither of the purified *H. pylori* FtsK constructs supported XerH recombination *in vitro* in the conditions tested, suggesting that FtsK is not the regulatory

protein required for XerH recombination. However, technical limitations could also contribute to the observed lack of recombination. For example, it is possible that the purified proteins were not folded properly and therefore were not active in the assays; however, the fact that two very different constructs resulted in the same observed outcome suggests it is likely not caused by a protein folding problem. It is also possible that the constructs used did not include a part of FtsK that is necessary for XerH recombination. The $FtsK_C$-A construct was designed carefully based on the *E. coli* $FtsK_C$ construct, but due to a considerable sequence divergence between the proteins the selected construct might still lack sequences essential for XerH recombination. Nevertheless, the $FtsK_C$-A construct included the entire predicted ATP helicase domain, and therefore, if *H. pylori* FtsK activates XerH by a similar mechanism to that observed for XerC/D recombination, this domain should have been sufficient for activation of XerH recombination. Previous studies with XerS showed that $FtsK_{SL}\gamma$ is not sufficient for XerS recombination, but a construct equivalent to *H. pylori* $FtsK_C$-A was sufficient for XerS recombination in *E. coli*. However, it is worth noting that XerS recombination, as for *H. pylori* XerH, could not be reconstituted *in vitro* even in the presence of the $FtsK_C$ construct supporting recombination *in vivo* (Nolivos *et al.,* 2010).

Consistently with the lack of FtsK activity *in vitro*, no direct interaction between purified XerH and FtsK could be detected in the pull-down assay (Figure 5-5). It is possible that the assay conditions did not permit the interaction or that the method used to detect the interaction was not sensitive enough. However, a similar study of XerC/D-FtsK interactions revealed that $FtsK_C$ interacts with XerD but not with XerC (Yates *et al.*, 2006), so the results of the experiment presented here might in fact represent true lack of interactions between *H. pylori* XerH and $FtsK_C$.

Together, the results of the *in vitro* assays suggest that some factor other than FtsK could be required for activation of XerH recombination, but it is also possible that the FtsK constructs designed or the assay conditions chosen were not permissive for successful XerH-FtsK recombination.

5.3.3 XerH is proficient in plasmid recombination in *E. coli*

H. pylori XerH efficiently recombined plasmids containing two directly repeated dif_H sites in two independent recombination assays performed in *E. coli*: the *galK*-MacConkey assay (Figure 5-9) and the liquid culture assay over 200 generations (Figure 5-11). Apart from XerH, no other genes from *H. pylori* were supplied in this assay, indicating that XerH recombination of plasmid-borne dif_H sites does not require any *H. pylori*-specific regulatory factors. Furthermore, if a regulatory factor is required for XerH recombination, it must be present in *E. coli* strain DS941, suggesting that *E. coli* carries a homologue of the native *H. pylori* XerH activator, or an unrelated protein can function as an XerH activator in *E. coli*.

However, it is possible that the observed recombination of plasmid-borne dif_H sites might not be caused by the complete XerH recombination reaction: for example, XerH could form HJ intermediates that could then be resolved by the host DNA-repair machinery as seen for XerC/D recombination during phage CTX integration (Val *et al.,* 2005), meaning that XerH recombination might only appear to be fully functional in *E. coli*, while it is only half functional as seen *in vitro.* However, the fact that high levels of recombination were observed and that those levels were closely related to the level of XerH expression (Figure 5-9b) suggests that the observed plasmid recombination is likely caused by XerH enzymatic activity alone.

Additionally, the assays were performed on plasmid-borne dif_H sites and not at the replication terminus region of the chromosome, placing these sites in a different spatial and regulatory context than that of their native chromosomal position. Such differences in dif_H site location could obscure the native regulation of XerH. For example, while XerH recombination could be activated on plasmid-borne dif_H sites, perhaps the spatial and temporal aspect of regulation does not apply to these sites. Such dual nature of Xer regulation has been described for the XerC/D recombination system, where isolated FtsKγ subdomain can directly activate XerD allowing XerC/D recombination, while the full $FtsK_C$ domain is required for assuring correct topology of the substrates and products and correct timing of the reaction, and hence the spatial and temporal context of recombination.

Of other studied single-recombinase Xer systems, only XerH from *C. jejuni* has been shown to recombine two recombination sites present on the same plasmid in *E. coli*

without the need for *H. pylori*-specific factors (Leroux *et al.,* 2013). No such recombination could be detected for the XerS/dif_{SL} system (Le Bourgeois *et al.,* 2007), and the archaeal XerA recombination system has not been tested in *E. coli.* This suggests that XerH proteins could employ a unique regulatory mechanism not shared by other Xer systems.

5.3.4 *E. coli* $FtsK_C$ is not required for XerH plasmid recombination in *E. coli*

Both plasmid-based XerH recombination assays were performed also in an $FtsK_C$-deficient *E. coli* background, resulting in XerH recombination efficiency comparable in both wild-type and $FtsK_C$-deficient *E. coli* strains (Figures 5-10 and 5-11). The only difference observed between the two strains was that in the recombination assay in liquid culture, more catenated and multimerized plasmids were observed in the $FtsK_C$-deficient strain; however, this phenotype does not seem to be directly related to XerH recombination as it was also observed in the absence of the dif_H-carrying plasmid (Figure 5-11b). It seems therefore that *E. coli* $FtsK_C$ is not required for XerH recombination of plasmid-borne dif_H sites.

However, as mentioned in the previous section, the observed excision of dif_H-*galK*-dif_H cassette could be a result of partial XerH recombination complemented by HJ resolution by host machinery, which would be independent of $FtsK_C$. At the same time, it is also possible that the FtsK role in XerH recombination does not include direct activation of XerH monomers, but instead it has only a regulatory role in terms of spatial and temporal context. If so, the effect of $FtsK_C$ deletion in the plasmid-based recombination assay would not be seen as the recombination sites are placed out of a normal FtsK activity context, where FtsK is guided to *dif* recombination sites by chromosomal KOPS DNA (Sivanathan *et al.,* 2006).

Finally, the observed XerH recombination could be truly independent of FtsK. In support of this are similar assays with *C. jejuni* XerH that showed recombination of plasmid-borne dif_H sites in *E. coli* in the absence of $FtsK_C$ (although in both wild-type and $FtsK_C$-deficient *E. coli* strains recombination was very inefficient; Leroux *et al.,* 2013). Additionally, it is worth noting that canonical KOPS could not be identified on the *H. pylori* chromosome

(Debowski *et al*., 2012), supporting the notion that the *H. pylori* FtsK might play no role or a different role in XerH recombination.

5.3.5 Searching for XerH regulators in *E. coli*

If a regulatory factor is required for XerH recombination to proceed from substrates to products and XerH recombination is efficient in *E. coli* also in the absence of $FtsK_C$, it is possible that another factor fulfils this role an has functional homologues in both *E. coli* and *H. pylori*. It would not be the first instance that Xer recombination is not regulated by FtsK alone: *Bacillus subtilis* encodes two FtsK-related proteins, SpoIIIE and SftA, both of which are involved in chromosome dimer resolution although neither is essential for this process (Kaimer *et al.,* 2011). To search for a potential regulatory factor, XerH was used as "bait" in an affinity chromatography assay with *E. coli* lysate. From this, a number of candidate proteins was identified and sorted based on the protein function or the phenotypes of deletion mutants (Table 5-2).

Most of the identified proteins play a role in genome maintenance. For example, two ATP helicases were identified: DnaB, a replicative helicase that opens the replication fork during DNA replication, and RuvB, a helicase that mediates HJ migration in homologous recombination. As FtsK is also an ATP helicase, one of those proteins could additionally function as an analogue of *E. coli* FtsK in XerH recombination, especially since both helicases have homologues in *H. pylori*. Additionally, two topoisomerases (ParC and TopA) were identified, one of which could couple XerH recombination to chromosome decatenation, which was proposed to be a second function of XerH in *H. pylori* (Debowski *et al.,* 2012). Of the two proteins, only TopA has an identified homologue in *H. pylori*. Yet another interesting protein identified is ObgE: a GTPase required for chromosome segregation. ObgE mutations in *E. coli* result in elongated cells that do not divide, with the daughter chromosomes that do not partition to the daughter cells (Kobayashi *et al.*, 2001). Interestingly, ObgE mutants are also impaired in plasmid replication, and *H. pylori* encodes a homologue of ObgE. Two other proteins required for chromosome partitioning, MukE and MukF, were also identified in this study; however, there does not seem to be a homologue of any of them in *H. pylori*. One more candidate protein involved in genome maintenance is HsdR, an endonuclease of the host restriction machinery with a homologue in *H. pylori*, although how it could be involved in XerH regulation is not clear. Two

additional proteins that are functionally unrelated to chromosome maintenance were also selected: TsaE, an ATPase that was selected as a candidate because its mutations result in elongated cells with unusual distribution of DNA (Handford *et al*., 2009), and RdgC, a regulator of RecA that is essential for *E. coli* growth in recombination-deficient *recA*- strains (Ryder *et al.,* 1996). Homologues of neither protein could be found in *H. pylori*. Finally, four proteins of unknown function were also identified. All above proteins are only proposed candidates at present and their potential role in XerH recombination will have to be studied further in *E. coli* and *H. pylori* deletion strains, which might prove challenging as mutations of many of these proteins are not viable.

5.3.6 XerH does not recombine chromosome-borne dif_H sites in *E. coli*

In order to see if *H. pylori* XerH can also recombine chromosome-borne dif_H sites in *E. coli*, two experiments with dif_H cassettes replacing *E. coli dif* were designed. First, a dif_H-Km cassette was used to assess XerH-mediated chromosome dimer resolution by observing the cell growth phenotype. Second, a dif_H-Km-dif_H cassette was used to directly detect excisive recombination between the two dif_H sites. In both cases, there was no evidence for XerH recombination: the cells displayed the Xer-deficiency phenotype and no excision products could be detected in the respective experiments.

It is possible that, unlike plasmid-borne sites, dif_H site(s) replacing the *dif* site in the *E. coli ter* region are under tight spatial and temporal control that prohibits access of any regulatory factors required for XerH recombination. The *ter* region in bacteria is a place of action of many proteins, it is highly polarized, and it is subjected to tight spatial control (for example, it is the last part of the genome that gets segregated to the daughter cells). Accordingly, when the *E. coli dif* site was replaced by *loxP*, the recombination site of Cre recombinase, Cre recombination became dependent on the presence of functional $FtsK_C$ despite the fact that Cre does not require any additional factors for successful recombination of *loxP* sites (Leslie and Sherratt, 1995). In this context of tight regulation, an XerH activator that could otherwise promote recombination might not have access to the dif_H site in *E. coli,* where FtsK slides towards the *ter* region using polarized KOPS, possibly preventing the protein that activates XerH recombination from reaching dif_H sites, for example by displacing it from the DNA. This would not affect recombination of plasmid DNA, as there is no evidence that FtsK interacts with plasmid-borne sites. In *H.*

pylori, canonical KOPS are absent and FtsK regulation of this region might be less stringent or differently organized, allowing the required factor to reach the dif_H site. However, this hypothesis requires further experimental confirmation, including identification of the required XerH regulatory factor.

The role of *H. pylori* FtsK in XerH recombination has been addressed *in vivo* (Debowski *et al.,* 2012). The results showed that an ectopically placed chromosomal dif_H-Cm-dif_H cassette is not recombined by XerH in the absence of $FtsK_C$ in *H. pylori*. These results are in disagreement with this study of XerH recombination in *E. coli,* since the observed recombination efficiencies are comparable between the wild-type and FtsK-deficient *E. coli* strains. The reason for this disparity could be that XerH might have two separate and differently regulated functions: the first being chromosome dimer resolution and the second chromosome decatenation. The second role has been proposed as *H. pylori* does not carry a homologue of topoisomerase IV, an enzyme responsible for decatenation in *E. coli* (Debowski *et al*., 2012). This would mean that the reconstituted XerH recombination of plasmid-borne dif_H sites in *E. coli* corresponds to an FtsK-independent decatenation pathway, likely activated by another host factor (for example, one of the candidate factors identified in the affinity chromatography assay). At the same time, this would also mean that chromosomal dif_H recombination depends on *H. pylori* FtsK, which agrees with the fact that chromosomal XerH recombination could not be reconstituted in *E. coli,* while recombination in *H. pylori* was abolished in the $FtsK_C$-deficient strain. This hypothesis is therefore in agreement with the previously published experiments and results shown in this study. Nevertheless, this hypothesis requires further confirmation by recombination experiments in *H. pylori* and in *E. coli,* for example by reconstituting recombination of chromosome-borne dif_H sites in *E. coli* in the presence of *H. pylori* FtsK.

6. Study of XerT conjugative transposase

6.1 Introduction

6.1.1 Identification and classification of Plasticity Zones (PZs)

While *H. pylori* infections affect approximately one half of the world's population, only about 10% of infections lead to serious health problems ranging from gastritis, through gastric and duodenal ulcers, to gastric cancer (Matysiak-Budnik and Megraud, 1997). This variety of possible clinical outcomes of *H. pylori* infections raised the question of variability between *H. pylori* strains and isolates, which was addressed in the initial genome sequencing study of two *H. pylori* strains, 26695 and J99 (Alm *et al.*, 1999). The study revealed that the overall genomic organization of the two strains is similar and only 6-7% of the genes identified are strain-specific. Of those, almost half (46 and 48% in 26695 and J99, respectively) were found on a single hyper-variable region termed Plasticity Zone (PZ). This DNA region displays lower G+C content than the overall G+C content of the *H. pylori* genome (34-35% in PZs as opposed to overall 39%) and carries insertion sequences and a number of virulence genes, suggesting that PZs might represent pathogenicity islands. The low G+C content of the PZs further implied that they might have been acquired from other species, but since they were found at the same genomic location in both *H. pylori* strains, there was no evidence of their mobilization.

A wide search for PZs in additional *H. pylori* strains revealed that these elements are present in many but not all *H. pylori* strains (Kersulyte *et al.,* 2009). Identified PZs could be classified into three main types (type 1, type 1b, and type 2, Figure 6-1) based on their length, content and arrangement of the open reading frames (ORFs), and the sequences of these ORFs (Kersulyte *et al.,* 2009). Additionally, many strains contained incomplete remnants of PZs, or hybrid PZs containing parts of different PZs fused to each other.

PZs are large genetic islands that carry many ORFs. For example, the first identified type 1 PZ from strain Shi470 is 39 kbp long and encodes 37 ORFs, while the first identified type 2 PZ from strain PeCan18B is 52 kbp long and carries 36 ORFs. The only known type 1b element was found in strain P12, it is 41 kbp long and contains 35 ORFs. Most ORFs

encoded by the PZs are expected to contain functional genes, as they are highly conserved amongst different PZs. The functions of some of the encoded genes could be predicted, while the function of many others remains unknown since they do not share significant sequence homology with any proteins in the available databases.

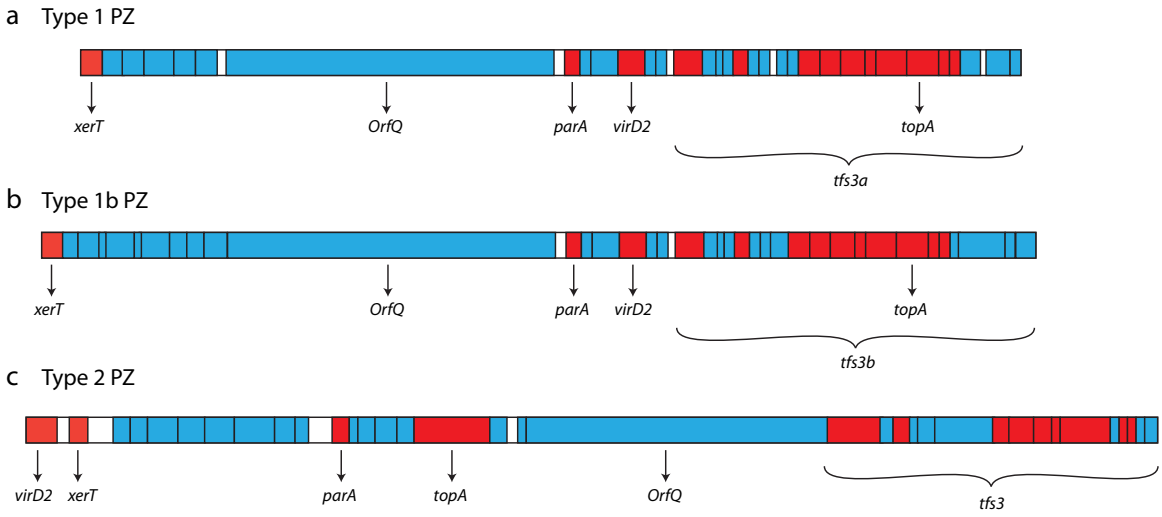


Figure 6-1: Schematic representation of the genetic organization of the three identified types of *H. pylori* PZs. Blue and red boxes denote ORFs within PZs, with the red boxes marking ORFs expected to be involved in mobilization and transfer of PZ based on their homology to known genes from other bacterial species. The potential functions of proteins indicated by arrows are discussed in section 6.1.3. a) Type 1 PZ from *H. pylori* strain Shi470. b) Type 1b PZ from *H. pylori* strain P12. Type 1b shares 95% sequence identity with type 1 PZ in the region between *OrfQ* and the end of the *tfs3* region, while the remaining regions differ significantly. c) Type 2 PZ from *H. pylori* strain Shi170. (Based on Kersulyte *et al.,* 2009)

6.1.2 Phenotypic effects of PZ insertion

PZs carry a number of virulence genes and a large ORF *OrfQ* of unknown function that is also found in other pathogenic bacteria, suggesting that the presence of a PZ in the *H. pylori* genome might contribute to its ability to colonize and persist in the stomach. To test this, a study of mouse stomach colonization was performed with various *H. pylori* strains (Kersulyte *et al.,* 2009). The study revealed that the presence of PZ is not required for successful colonization, but some strains with deleted PZ showed reduced fitness and decreased competitive ability compared to the equivalent strain containing the PZ (Kersulyte *et al.,* 2009). These findings suggested that the presence of a PZ might be beneficial for the *H. pylori* host. However, the extent of the advantage and the exact role of PZs in stomach colonization require further studies.

6.1.3 TnPZ: a conjugative transposon

A number of observations suggested that PZs might in fact be transposable elements and they were therefore renamed as transposable plasticity zones (TnPZs). First, the elements were found at various genomic locations in different strains, clearly indicating their mobile nature. Second, the elements were invariably flanked by a direct repeat of a heptanucleotide 5'AAGAATG, with the same sequence present only once at the corresponding empty sites in other *H. pylori* strains (Figure 6-2; Kersulyte *et al.,* 2009). Such target site duplication is a hallmark of integration of some DNA transposons and is frequently used as a criterion in genomic searches of transposons. Third, the two TnPZ ends showed high DNA sequence conservation over at least 80 bp and contained 13-bp terminal inverted repeats that could constitute transposase binding sites. However, despite many indications that TnPZs are mobile, the transposition of any TnPZ could not be seen directly (Kersulyte *et al.,* 2009).

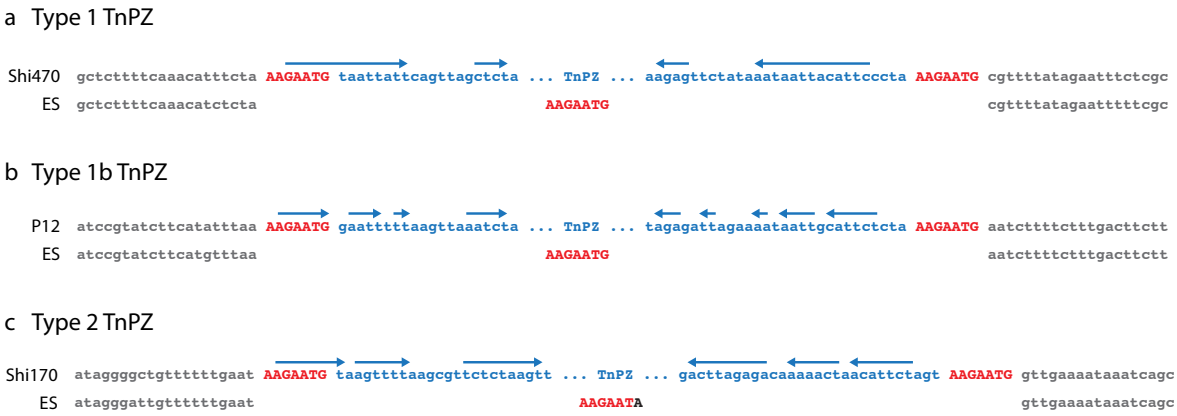


Figure 6-2: DNA sequences at TnPZ ends and corresponding empty sites (ES). The TnPZ sequences are shown in blue and the flanking *H. pylori* sequences are shown in grey. The conserved heptanucleotide 5'AAGAATG is shown in red. The arrows indicate inverted repeat regions. Since TnPZ insertion has never been directly observed, the corresponding ES come from *H. pylori* strains that showed the highest sequence homology in this region; therefore, the sequences may differ in some places from the transposon flanking regions. (Based on Kersulyte *et al.,* 2009)

Furthermore, the gene content of TnPZ showed similarities with a particular class of transposons, the conjugative transposons, again indicating the transposable nature of TnPZ. For example, all identified TnPZs carry a number of conserved ORFs that could be involved in conjugative transfer of TnPZ (Figure 6-1; Kersulyte *et al.*, 2009). A large (approximately 16 kbp) gene cluster contained genes for a predicted type IV secretion system that was termed *tfs3* (Figure 6-1). Since type IV secretion systems are involved in

bacterial conjugation (Christie *et al.,* 2005), *tfs3* could act in conjugation of TnPZ, although the mechanism of such transfer is not clear. Each TnPZ also encodes a DNA topoisomerase TopA that could assure the correct topology of TnPZ, a DNA relaxase VirD2 that could nick the transposon DNA at the origin of transfer, and a chromosome partitioning protein ParA that could be involved in positioning of the type IV secretion system on the cell surface and in transferring the transposon DNA nicked by VirD2. Together, these proteins could form a system that allows conjugal transfer of an excised circular transposon (which is a characteristic intermediate of conjugative transposition) from a donor into a recipient cell without additional host factors. In fact, fragments of TnPZ were shown to transfer by conjugation between *H. pylori* strains, although it is not clear if the TnPZ conjugation machinery was directly involved (Fischer *et al.,* 2010). Moreover, a putative origin of transfer *oriT* has been identified on a TnPZ, further supporting the idea that TnPZs belong to the extended family of conjugative transposable elements (Grove *et al.,* 2013).

All TnPZs also contain a large ORF of more than 2800 codons termed *OrfQ* (Kersulyte *et al.,* 2009). This ORF likely corresponds to a protein-encoding gene since a conserved domain search revealed DNA methylase and helicase domains (Kersulyte *et al.,* 2009). Proteins displaying similar domain architectures could be found in other pathogenic bacteria.

Interestingly, no canonical transposase gene could be found in any of the discovered TnPZs. Instead, all of them encode a member of the tyrosine recombinase family called XerT, which shares approximately 28% sequence identity with both XerC and XerD from *E. coli* and was proposed to act as the transposase (Kersulyte *et al.,* 2009).

6.1.4 Mechanism of conjugative transposition

Conjugative transposons are widely spread, promiscuous chromosomal mobile elements that can move from a donor to a recipient bacterial cell through conjugation. They usually encode a set of essential genes to enable excision from the chromosome, conjugational transfer, and integration of the element, as well as a set of cargo genes, including virulence or antibiotic resistance genes. The elements are therefore large, with sizes ranging from 18 kbp to over 100 kbp. The DNA cleavage and ligation reactions required for the excision

and integration of most conjugative transposons are catalysed by an integrase protein, which is usually a member of the tyrosine recombinase family, and likely follows a recombination mechanism that resembles that of the λ integrase (Scott and Churchward, 1995).

The mechanism of conjugative transposition is poorly understood. The current model (shown in Figure 6-3) is based on the studies of Tn916 transposon from *Streptococcus faecalis* (Salyers *et al.,* 1995) and it proposes a mechanism that proceeds through two steps of recombination, one for excision and one for integration, both of which follow analogous chemistry to canonical tyrosine recombination. First, for excision the integrase protein binds and cleaves the top and bottom strands of each transposon end with a six-nucleotide stagger, likely via formation of a covalent phosphotyrosyl protein-DNA intermediate, followed by formation of a double-stranded circular transposon junction through strand exchange and staggered-end ligation (Caparon and Scott, 1989). The cleavage of one DNA strand occurs exactly at the transposon-flank boundary of one transposon end and 6 nt into the flanking DNA at the other transposon end. This pattern is reversed on the other DNA strand, which results in a cleavage with a stagger on both transposon ends, and since the two overhangs are not identical, formation of a 6 bp heteroduplex region in the circular transposon. This region is referred to as a "coupling sequence" in conjugative transposition and likely corresponds to the central region between the two recombinase binding sites in canonical tyrosine recombination. One strand of the excised circular transposon is then transferred into the recipient cell via an unknown mechanism, likely involving the conjugation machinery encoded by the transposon itself. In the recipient cell, the second transposon strand is synthesised by the host replication machinery before the transposon is integrated into the recipient genome. Conjugative transposons show variable specificity regarding the integration site: for example, Tn916 preferentially uses A+T rich target sequences, while other conjugative transposons tend to integrate to more specific target sites. For integration, it was proposed that the transposon-end junction and the target site are brought together and cleaved with a six-nucleotide stagger by the integrase in a similar way as for excision, and strand exchange and religation reactions result in the integration of the transposon DNA (Figure 6-3; Salyers *et al.,* 1995). Mechanistically, integration is an analogous recombination reaction to excision, except with different substrates and products. It is likely that the single-stranded copy of the circular transposon left in the donor cell can be replicated and also reinsert into a target site in the donor genome. However, in some cases, only one copy of a conjugative transposon can be present in a

single genome and transposition within a single host is not observed (reviewed in Burrus *et al*., 2002), although the exact function and mechanism of such self-exclusion are not clear.

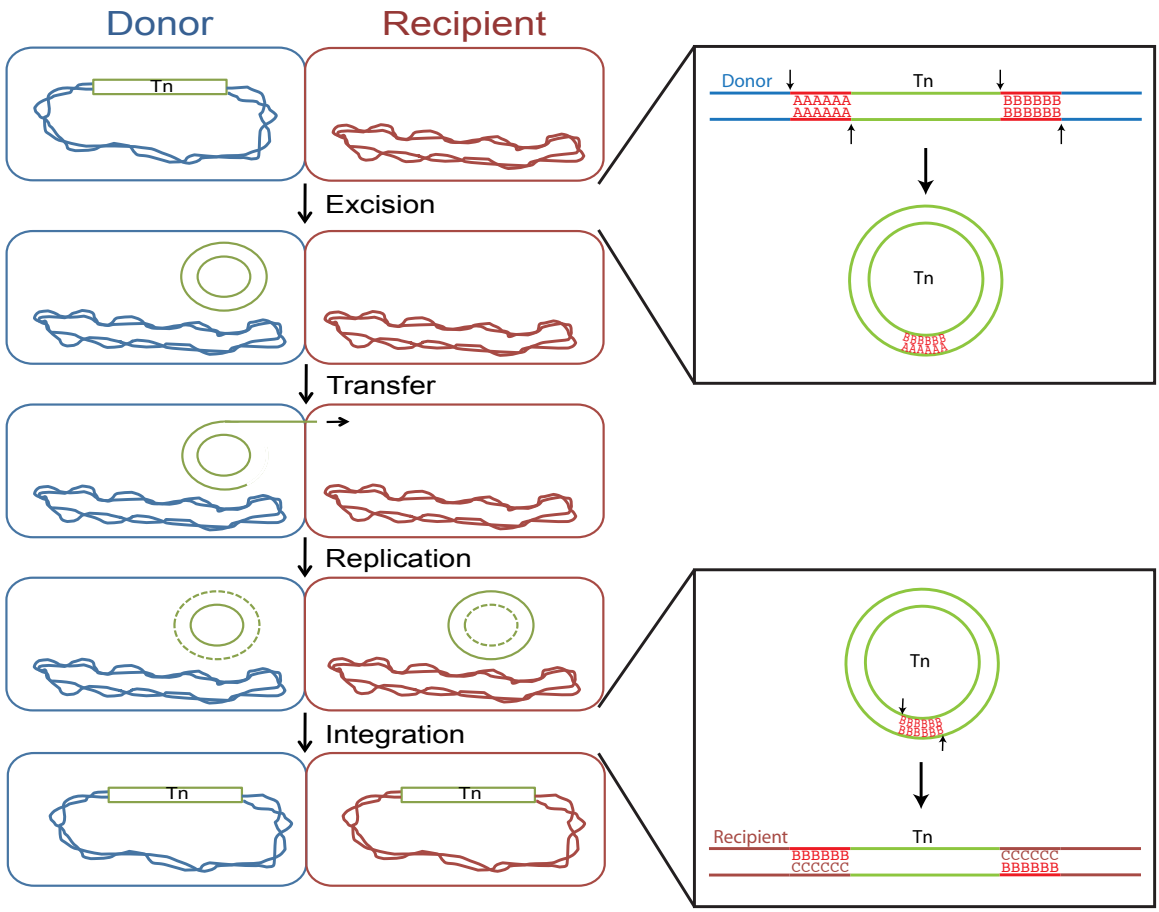


Figure 6-3: Proposed mechanism of conjugative transposition based on the studies of Tn916. The transposon-encoded integrase brings together and recombines the two transposon ends by cleavage at the coupling sequences (shown in red) with a six-nucleotide stagger. This results in excision of a circular transposon containing a six-nucleotide mismatch within the coupling sequence. One strand of the circular transposon is transferred into the recipient cell through bacterial conjugation. Both single-stranded circles are then replicated by host replication machinery. The double-stranded circular transposon, specifically its transposon end-junction region, and a target site in the recipient genome are brought together by the integrase and recombined in a reaction analogous to the transposon excision. This results in the integration of the transposon into the recipient genome, so that it is flanked by two regions of DNA heteroduplex. Regions of heteroduplex are resolved by the host replication machinery. Reintegration can also occur in the donor cell. (Based on Salyers *et al.,* 1995)

The lack of homology between the central sequences of the two recombined sites (either the two transposon ends in excision, or the circular transposon and the target insertion site in integration) is a hallmark of conjugative transposition. It is not clear how the integrase can support such a reaction, considering that all canonical tyrosine recombinases require absolute sequence homology between the central regions of the two recombined sites for

successful recombination. This paradox remains the most puzzling aspect of the mechanism of conjugative transposition.

6.1.5 Examples of conjugative transposons and their integrases

Most of the conjugative transposons identified to date encode a tyrosine recombinase that acts as a transposon integrase (or a transposase). The majority of these integrases are closely related to that of the Tn916 element from *S. faecalis*, which in turn resembles λ integrase. Additionally, unrelated recombinases have been found to mobilize the CTnDOT element from Bacteroides species. Many other conjugative transposons and their specific integrases have been identified, although in many cases further validation of the putative elements is required (Burrus *et al.,* 2002).

6.1.5.1 Tn916

Tn916 is by far the best studied conjugative transposon and is often used as the model for conjugative transposition. Tn916 was first identified in *S. faecalis*, and is the smallest known conjugative transposon with a size of 18.4 kbp (Flannagan *et al.,* 1994). It confers resistance to tetracycline and can spread between gram-positive and from gram-positive to gram-negative bacteria, showing a very broad host range (Bertram *et al.,* 1991). More than one copy of Tn916 can be found in bacterial genomes, suggesting that the presence of one copy of Tn916 does not prevent acquisition of another (Norgren and Scott, 1991).

Apart from the integrase protein, an additional protein required for Tn916 excision, Xis, is encoded by the element (Marra and Scott, 1999). No host factors are known to be required for Tn916 excision. As shown in Figure 6-3, in Tn916 transposition the DNA is cut with a six-nucleotide stagger, with some variability observed in the position of cleavage at the right transposon end (Caparon and Scott, 1989). The Tn916 integrase shows a domain architecture similar to that of λ integrase, with two DNA-binding domains interacting with the core binding site at the transposon ends, as well as with additional arm sites within the transposon (Lu and Churchward, 1994). The protein does not seem to have a strict binding specificity as various sequences can be used as insertion targets, although sequences with A- and T-tracts resembling transposon ends are preferred (Scott *et al.,* 1994).

There is a number of conjugative transposons closely related to Tn916, for example Tn1545 from *S. pneumoniae*, which is larger than Tn916 and additionally carries kanamycin and erythromycin resistance genes. Another medically relevant conjugative transposon from this family is Tn1549 from *Enterococcus faecium*, which carries a vancomycin resistance gene and was implicated as one of the major factors in the spread of vancomycin resistance amongst bacteria (Garnier *et al.,* 2000).

6.1.5.2 CTnDOT

Gram-negative Bacteroides species carry large conjugative transposons of at least 50 kbp, with the best studied example being CTnDOT (Smith *et al.*, 1998). CTnDOT carries a tetracycline resistance determinant *tetQ,* and its conjugative transposition is induced by tetracycline (Stevens *et al.,* 1993). Amongst other genes, CTnDOT encodes a tyrosine recombinase IntDOT and an additional small basic protein that (unlike in Tn916 transposition) is not required for CTnDOT excision (Cheng *et al.,* 2000). Additionally, IntDOT contains the arm-binding domain and therefore resembles λ integrase more closely than Cre recombinase. The main features of the mechanism proposed for CTnDOT conjugative transposition are similar to those of Tn916, with the most notable difference being the target specificity of IntDOT: the insertion requires a GTTnnTTGC sequence to be present at the target DNA site (Cheng *et al.,* 2000).

6.1.6 Classification of XerT as a potential transposase

Of all the ORFs carried on *H. pylori* TnPZs, the *xerT* gene is the most promising candidate for encoding a protein responsible for mobilization of TnPZ as it is a member of the tyrosine recombinase family. Consequently, XerT was proposed to act as the TnPZ transposase (Kersulyte *et al.,* 2009). To date, the role of XerT in mobilization of TnPZ was tested in only one *H. pylori* strain, P12, where the role of XerT from type 1b TnPZ was assessed (Fischer *et al.,* 2010). An experiment with XerT-deficient *H. pylori* strain showed that TnPZ genes were transferred from a donor to a recipient strain genome only in the presence of XerT. Additionally, the study showed that the expected circular transposon was absent from an XerT-deficient strain, while it could be detected in the isogenic XerT-containing strain (Fischer *et al.,* 2010). Together, these first insights into TnPZ

mobilization suggested that XerT from P12 is competent in excision and circularization of TnPZ, and that functional XerT is required for TnPZ mobilization in *H. pylori*.

6.1.7 Phylogeny and alignment of XerT within the Xer family

XerT has been classified as an Xer-type tyrosine recombinase due to its significant sequence homology with other Xer recombinases and the preservation of all catalytic residues required for tyrosine recombination (see Figure 3-2). XerT from *H. pylori* strain P12 shares 20% sequence identity with XerC, 18% with XerD, and 25% with XerH from the same *H. pylori* strain, as determined by ClustalW alignment (Larkin *et al.,* 2007). Accordingly, the phylogenetic analysis of XerT proteins and other Xer recombinases revealed that XerT is more closely related to XerH than to other Xers, suggesting a common (although distant) evolutionary origin (Figure 6-4). However, the analysis also showed that XerTs from different TnPZ types cluster together and form a separate phylogenetic branch only distantly related to other Xers, including XerHs from their own host *H. Pylori* strain. Therefore, it seems that XerH and XerT proteins evolved separately to carry out two different functions in *H. pylori*, which involves recombining distinct DNA sites with no cross-reactivity, while retaining a similar protein fold and likely similar catalytic mechanism.

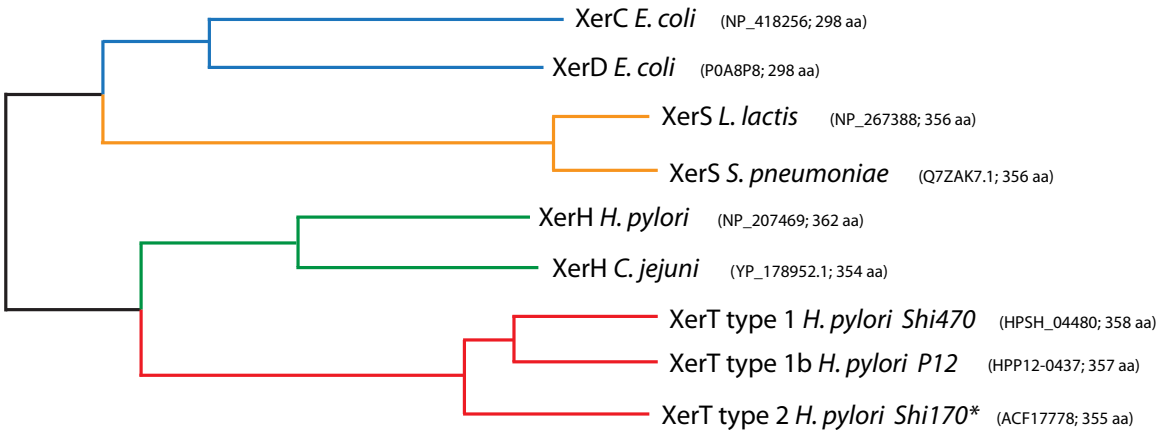


Figure 6-4: Phylogenetic tree showing the evolutionary relationship of XerT to other bacterial Xer recombinases. The NCBI accession number and the size of the protein in amino acids (aa) are shown in brackets. The XerT protein denoted by a star was found to have a mutation resulting in protein truncation; the protein sequence used in this analysis corresponds to the reconstituted full-length protein. The phylogenetic tree was prepared in MEGA6 (Tamura *et al.,* 2013) using the Neighbour-Joining method (Saitou and Nei, 1987). Phylogenetic branches are marked with different colours. The length of each branch corresponds to the phylogenetic distance.

6.1.8 Aims and objectives

Despite the fact that conjugative transposons are one of the major factors in the spread of antibiotic resistance amongst bacteria, and therefore are directly involved in the alarming spread of multidrug-resistant bacteria amongst humans, very little is known about their mechanism of transfer. The study of TnPZ, a putative conjugative transposon, presents a great opportunity to contribute to our knowledge of conjugative transposition because the proposed transposase of TnPZ, XerT, belongs to the well-studied family of Xer tyrosine recombinases, with a canonical Xer recombinase, XerH, also present in the same *H. pylori* genome. This provides technical advantages and allows for direct comparison of the two systems, helping to draw solid conclusions about TnPZ mechanism. Investigating this system, I particularly aimed to understand the role of XerT in mobilization of TnPZ and to define the requirements for successful mobilization of TnPZ in terms of protein factors and DNA sequences required. To achieve this, a combination of biochemical and microbiological approaches has been employed, taking advantage of the methods and protocols established in the study of the XerH recombinase.

6.2 Results

6.2.1 Construct design and cloning of XerT

In order to investigate the interactions between XerT and TnPZ transposon ends *in vitro*, constructs of XerT from *H. pylori* strain P12 were designed. The *H. pylori* strain P12 was chosen from the list of TnPZ-containing strains because the type 1b TnPZ from this strain was shown to excise from *H. pylori* genome in the presence of XerT, confirming that this TnPZ and its recombinase are functional (Fischer *et al.,* 2010). As in the case of XerH, XerT constructs were designed for overexpression in *E. coli* using the pET expression system. The XerT gene (HHP12-0437) was codon-optimized for expression in *E. coli* (Figure 6-5a) and inserted into vectors pETM-22 and pETM-28 by restriction enzyme cloning (Figure 6-5b). The pETM-22 vector contains the TRX solubility and the 6xHis affinity tags upstream of the gene of interest, resulting in the protein of interest being expressed as an N-terminal TRX-6xHis fusion. The pETM-28 vector contains the 6xHis affinity and the SUMO solubility tags upstream of the multiple cloning site, resulting in the protein being expressed as an N-terminal 6xHis-SUMO fusion. The two separate

constructs were made to increase the chance of successful overexpression of large quantities of soluble XerT protein.

6.2.2 Overexpression of XerT in *E. coli*

The overexpression of the two XerT constructs, XerT_pETM-22 and XerT_pETM-28, was carried out in *E. coli* strain BL21 (DE3), which in this study consistently showed the best expression of various *H. pylori* proteins. In the initial small-scale test expressions, the expression cultures were grown to OD_{600} of 1.2 at 37 °C or 42 °C, followed by induction with IPTG, and protein expression at 37 °C for three hours or at 15 °C for 15 hours. Samples were collected before addition of IPTG and at harvest time, and were analysed by SDS-PAGE (Figure 6-6a, lanes 0h, 3h, and 15h). In all cases the expression levels were high enough for protein extraction. Additionally, the solubility of the expressed protein was tested: the expression cultures were harvested and the cells were lysed by sonication. The lysate was then ultra-centrifuged to separate the soluble and insoluble fractions. The samples before and after centrifugation were analysed by SDS-PAGE (Figure 6-6a, lanes L, P, and S). This revealed that the protein expressed at 37 °C for three hours from both pETM-22 and pETM-28 was insoluble, while soluble XerT could be obtained upon expression at 15 °C for 15 hours. The protein yield was considerably higher when the overexpression cultures were grown initially at 42 °C. At this temperature, heat shock proteins are expressed, some of which could act as chaperones to promote correct protein folding and soluble protein expression (De Maio, 1999).

a

```
ori ATGAGTGATTGCAAAATGAGTAGGGTTAGTAGGGAATTGTTTAATAATATTAAGTCATTT TTACACCATAAGCTTAAAACAATGATACGCATTCAAAGCGTGAATGATTTACAATTGGTC 120
opt ATGGCTGACTGTAAAATGTCTCGTGTTTCCCGTGAGCTGTTCAATAATATCAAATCGTTC CTGCACCACAAACTGAAAACCATGATTCGTATCCAGAGCGTGAATGATCTGCAACTGGTC 120
    ***  *** ** ******  * * ***    * **  **** ******** ** ** **   * ***** ** ** ***** ***** ** ** ** ************ * *** *****

ori TTAAAATGGCAGGATAGAGTTTTAGAGTGCCAAAGTCTTATCGCATTGAAAGAACTCAAT CGCAAGCTTTATAATCAAGGTATTAGGCATACGATTATGATGCAAGGCTTGTTTTTGTTC 240
opt CTGAAATGGCAAGACCGTGTACTGGAGTGTCAGTCCCTGATTGCCCTGAAAGAACTGAAC CGTAAACTGTATAACCAGGGTATCCGTCATACCATTATGATGCAAGGCCTGTTTCTGTTT 240
     * ******** **  * **  * ***** **     ** ** **  ********** **  ** ** ** ***** ** *****  * ***** *************** ***** ****

ori TTTGAATATTTTGATAATAGGATTAAGCTCAAATCCTTGCACAATCTAGCAGAAGAACAA GTTATAGACTTCTTATTTGGATTAGCCAAAAACAGGAAGCCAAGCTCTATGGCTAAATAT 360
opt TTCGAGTATTTCGACAACCGTATCAAACTGAAATCCCTGCACAATCTGGCAGAAGAGCAA GTCATCGACTTTCTGTTCGGTCTGGCAAAAAACCGTAAACCGAGCAGCATGGCCAAATAT 360
    ** ** ***** ** **  * ** ** ** ****** ********** ******** *** ** ** *****  * ** **  * ** ****** * ** ** ***   ***** ******

ori GTGATGTATTTAAGACAATTCTTTGATTATTTAGACAAGAAAAGAAACTATAGCTTTGAT TTTGAACTTAAAAATCTATCGTTTGCCAAGAAAGAGACGCATTTGCCTAAACATTTGAAT 480
opt GTGATGTATCTGCGCCAATTCTTCGATTATCTGGACAAAAAACGTAACTATTCCTTCGAC TTCGAGCTGAAAAATCTGAGCTTCGCCAAAAAAGAGACTCACCTGCCGAAACATCTGAAT 480
    ********* *  * ******** ****** * ***** *** * ******  *** **  ** ** ** ********    ** ***** ******** **  **** ****** *****

ori AAAAATGATTTTAAGGCTTTTATCCAAGCTCTTCTAAAATACCATCCCAAAACCAGTTTT GAAAAACGCAATCAGTGTATTTTACTGCTTATCGTATTAGGGGGATTAAGGAAGTTTGAA 600
opt AAAAACGACTTCAAAGCCTTCATCCAGGCACTGCTGAAATATCACCCGAAAACGAGCTTC GAGAAACGTAATCAGTGTATCCTGCTGCTGATTGTTCTGGGTGGTCTGCGCAAATTCGAG 600
    ***** ** ** ** ** ** ***** ** ** ** ***** ** ** ***** ** **  ** ***** ***********  * ***** ** **  * ** **  *  * ** ** **

ori GCCTTAGATTTGGAGCTAAAAAACATTGCTTTAGAAAATAACCACTACCGCCTTTTAATC AAAGGAAAGAACAACAAAGAGAGATACGCTTATATTGAAAAAGAGTTTTTGCAAATTCCT 720
opt GCACTGGATCTGGAACTGAAAAACATTGCCCTGGAGAACAACCATTATCGTCTGCTGATC AAAGGCAAAAACAACAAAGAGCGCTATGCCTATATCGAGAAAGAGTTCCTGCAAATTCCG 720
    **  * *** **** ** **********  * ** ** ***** ** ** **  * *** ***** ** ************ * ** ** ***** ** ********  **********

ori TTGAATGCATGGCTAAGCGATACTAAACGATTAAAGAGTTTTAAGGGGCGTTTTGTATTT AAGAAAGCAAAAAATAACACCACACAAAAAACTTGCTCTTTAAAGGGCTTTATCGCTAAA 840
opt CTGAATGCTTGGCTGAGTGACACCAAACGCCTGAAATCCTTCAAAGGCCGCTTCGTGTTC AAAAAAGCAAAAAACAACACGACTCAGAAAACCTGTTCGCTGAAAGGGTTTATCGCCAAA 840
     ******* ***** ** ** ** *****  * **    ** ** ** ** ** ** **  ** ************ ***** ** ** ***** ** **  * ** ** ******** ***

ori ATTTTTAAGCTCTCTAATATTGATGTAAAATCTTATGGATTAGGTTTGCATCTTTTTAGG CATAGTTTTGCCACTTTTATTTATGATGAAACACAGGATTTGGTTTTAACTTCAAGAGCG 960
opt ATCTTCAAACTGAGCAACATCGACGTGAAATCTTATGGTCTGGGTCTGCACCTGTTCCGT CACAGCTTTGCCACCTTCATTTATGACGAAACCCAGGACCTGGTTCTGACATCTCGTGCT 960
    ** ** ** **    ** ** ** ** ***********  * *** **** ** **  *  ** ** ******** ** ******** ***** *****  ***** * ** **  * **

ori TTAGGGCATAGTTCTTTGCTTTCTACTAAAATCTATATCCATACCACTCAAGAACACAAT AAAAAAGTGGCTCTTGTGTTAGGGGGGTTATTAAGAAATGAAAAAAGCGAGTAG 1074
opt CTGGGACATTCTTCTCTGCTGTCGACGAAAATCTATATCCACACCACACAGGAGCATAAC AAAAAAGTTGCTCTGGTGCTGGGAGGACTGCTGCGTAATGAGAAAAGCGAGTGA 1074
     * ** ***  **** **** ** ** ************** ***** ** ** ** **  ******** ***** *** * ** **  *  *  * ***** **********
```

b

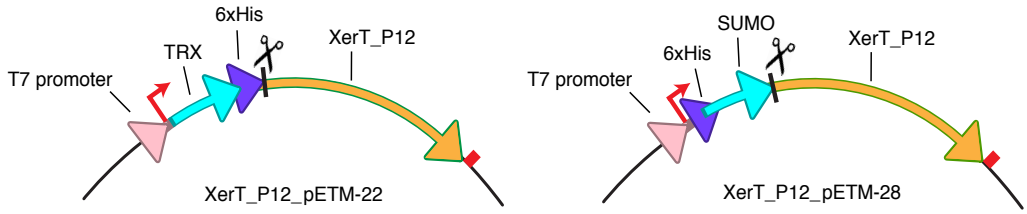


Figure 6-5: XerT constructs used for protein overexpression in *E. coli*. a) Comparison of *H. pylori* P12 *xerT* gene (ori) and its derivative codon-optimized for overexpression in *E. coli* (opt). Unchanged nucleotides are marked with stars. The second amino acid of XerT (Ser) has been mutated to alanine to introduce an NcoI restriction site. b) Schematic representation of the XerT *E. coli* expression vectors XerT_P12_pETM-22 and XerT_P12_pETM-28. The start and end of the open reading frames are marked in red. The scissors represent the protease (SenP2 for pETM-28 or 3C for pETM-22) cleavage site.

The initial small-scale expressions were scaled up for the XerT_pETM-22 vector, which provided a better yield of soluble protein. The final large-scale expressions were performed in *E. coli* strain BL21 (DE3) with initial growth at 42 °C, with expression at 15 °C for 15 hours, resulting in XerT overexpression at sufficient levels for further protein extraction and purification (Figure 6-6b).

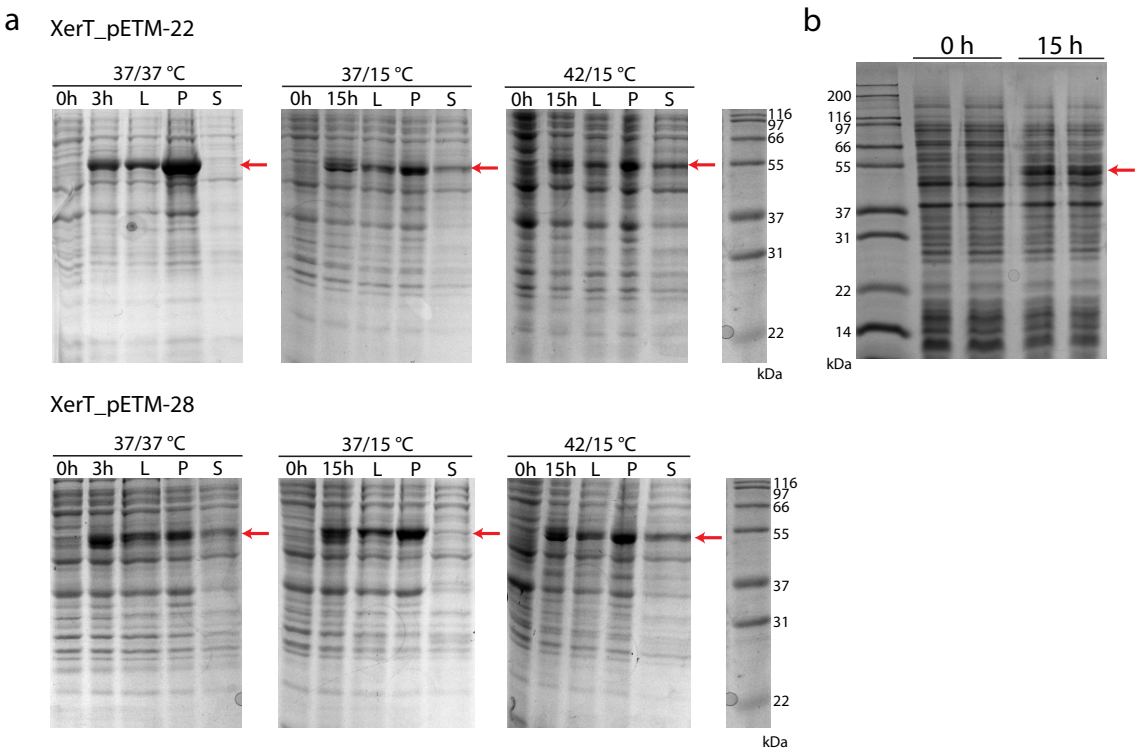


Figure 6-6: Overexpression of XerT in *E. coli* as analysed by SDS-PAGE. The positions of the expressed XerT proteins on the gels are indicated by red arrows. a) Small-scale test expression of XerT from two different vectors, pETM-22 and pETM-28, in *E. coli* strain BL21 (DE3) grown in various conditions. The growth temperatures are indicated above the gel, with the first temperature corresponding to the growth before induction, and the second to the growth after induction with IPTG. The samples were analysed immediately after induction (0h), 3 hours (3h) or 15 hours later (15h), after culture harvest and lysis by sonication (L), and after ultra-centrifugation that separated the soluble lysate fraction (supernatant-S) from the insoluble lysate fraction (pellet-P). b) Large scale expression of XerT from pETM-22 in *E. coli* BL21 (DE3). The samples were analysed immediately after induction (0h) and 15 (15h) hours later.

6.2.3 Computational analysis of XerT

The protein parameters (for example, the molecular weight, pI, and extinction coefficient) required for designing purification buffers and strategies were obtained by computational analysis of the XerT amino acid sequence using the ProtParam web-based tool (Table 6-1). Since both tagged and tag-free constructs of XerT showed high pI values (9.6 and 9.9, respectively), purification buffers at pH 7.5 were chosen. The high pI indicated that the proteins contain many basic residues consistent with the proposed DNA-binding function of XerT, and phosphate buffers were chosen for purification to mimic the DNA phosphate groups and therefore help to stabilize the protein. Molecular weight and extinction coefficient values were used to estimate the concentrations of the protein during and after purification.

Table 6-1: Properties of XerT and TRX-6xHis-XerT as determined by ProtParam.

Feature	XerT	TRX-6xHis-XerT
Number of amino acids	357	490
Molecular weight (Da)	42115.4	56355.7
Theoretical pI	9.9	9.56
Molar extinction coefficient ($M^{-1}cm^{-1}$)	30500	44500

As in the case of XerH, the secondary structure prediction of XerT, prepared using PSIPRED Protein Sequence Analysis Workbench (Buchan *et al*., 2013), suggested that the full-length construct of XerT is feasible for purification since no long unstructured regions could be found in the protein (Figure 6-7). The prediction showed great similarity to the observed secondary structure of XerH (see Figure 3-8), highlighting that these two proteins are structurally homologous. XerT was predicted to consist of two mostly α-helical domains (with only two β-strands assigned within the C-terminal domain), connected by a flexible linker shorter than that of XerH.

To visualize the predicted fold of XerT, a 3-dimensional homology model was prepared using the Phyre2 server (Figure 6-8; Kelley and Sternberg, 2009). As in the case of XerH, the related proteins found by Phyre2 and used for homology modelling included mostly available structures of other tyrosine recombinases. Accordingly, the final model is consistent with the overall domain architecture of tyrosine recombinases: two distinct domains are connected by a linker in a conformation that allows DNA binding. However, the protruding C-terminal segment characteristic of tyrosine recombinases in a synaptic complex has not been built into this model. Most of the catalytic residues (in red in Figure 6-8) appear to be facing towards the predicted DNA-binding site, but the catalytic tyrosine Tyr332 faces away from the DNA cleft. This, together with the failure to model the C-terminus, suggests that the functional arrangement of this region is likely to be different than the one modelled. Finally, the most N-terminal helix of XerT, for which there is no counterpart in other tyrosine recombinases, is sticking out of the body of the domain, and it is likely positioned differently in the true structure of XerT.

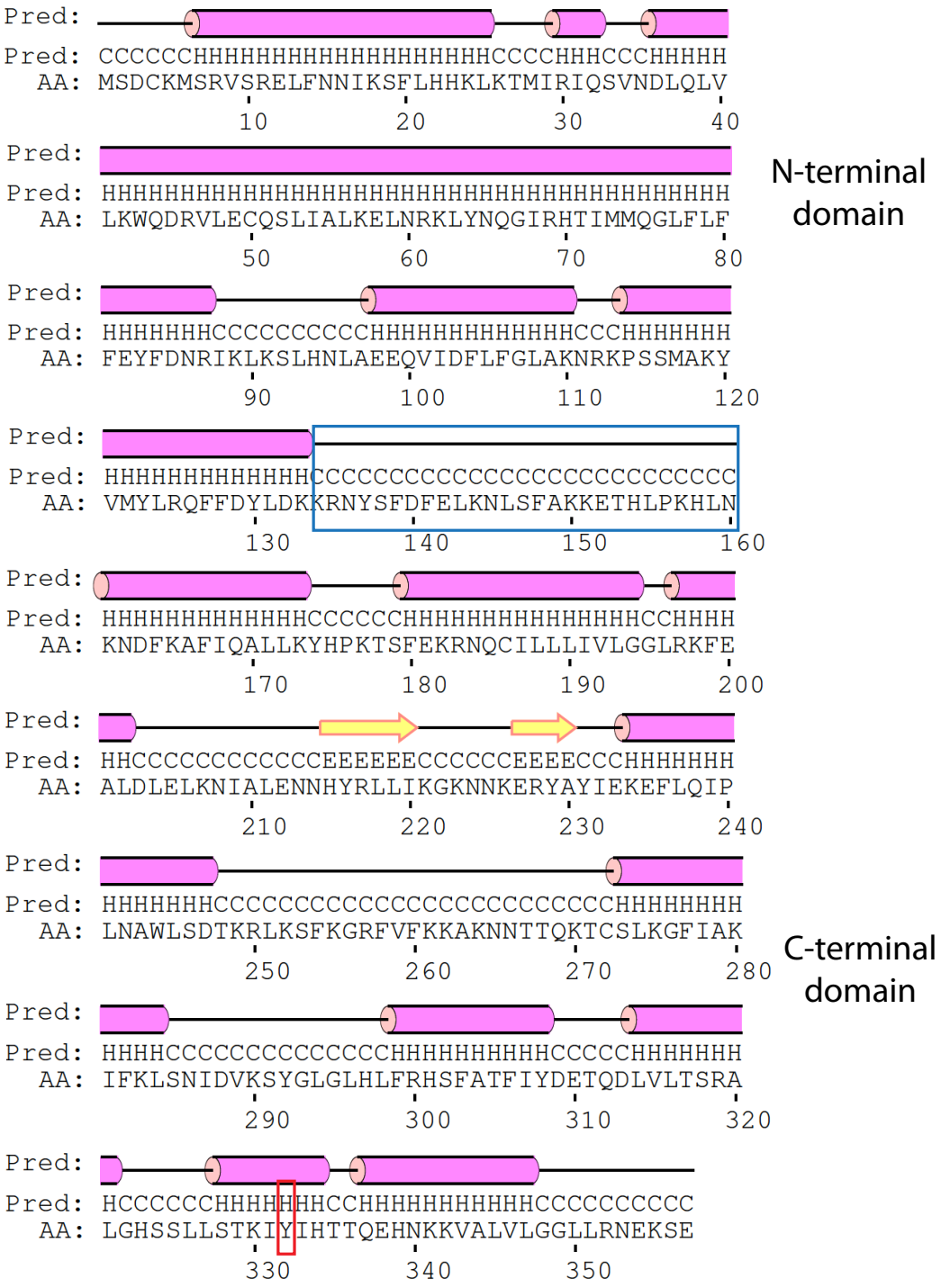


Figure 6-7: Secondary structure prediction for *H. pylori* XerT prepared using PSIPRED Protein Sequence Analysis Workbench (Buchan *et al*., 2013). The amino acid sequence (AA) is accompanied by the secondary structure prediction for each residue (Pred) expressed as a letter (H: α-helix, E: extended β-strand, and C: coil, loop, or unstructured region) and as a graphical representation (magenta cylinder: α-helix, yellow arrow: β-strand). The catalytic tyrosine residue is marked with a red box. The unstructured linker between the N-terminal and the C-terminal domains is marked with a blue box.

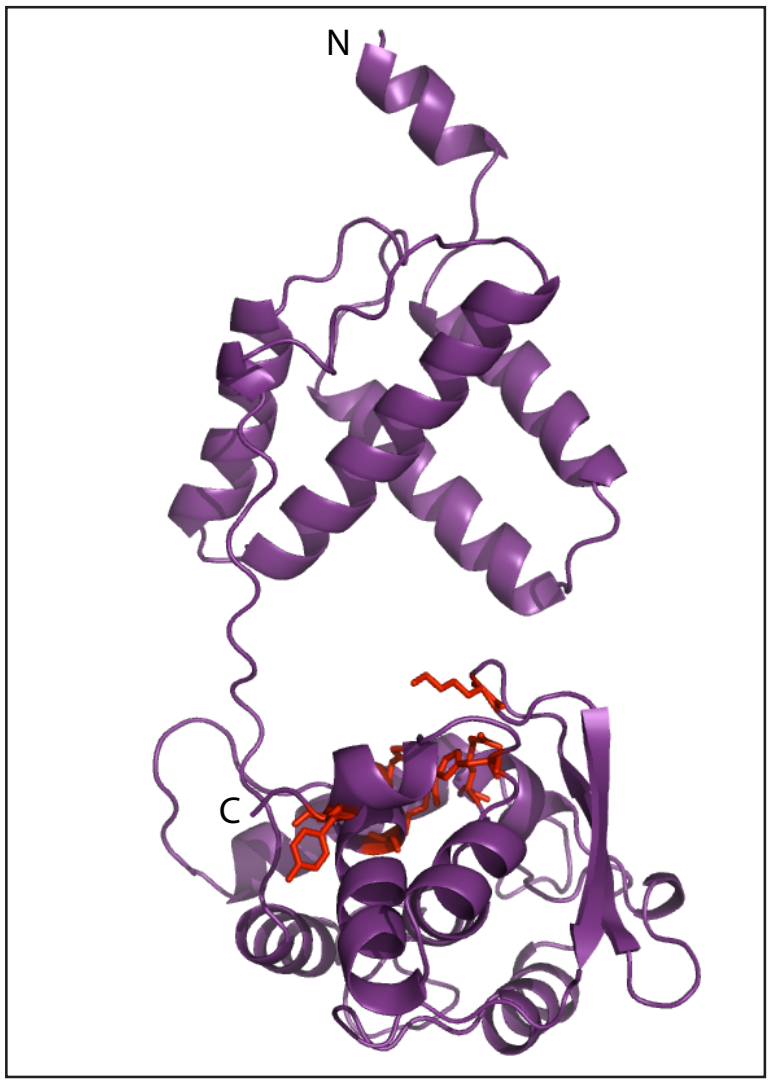


Figure 6-8: Cartoon representation of the XerT homology model prepared using Phyre2 server (Kelley and Sternberg, 2009). Predicted α-helices are shown as ribbon spirals, and β-sheets as ribbon arrows. The catalytic residues of XerT (Tyr332, Arg197, Lys223, His297, Arg300 and His 323) are shown as sticks and marked in red to highlight their plausible positions facing the cleft where DNA could be bound. The N- and C-termini of the protein are marked.

6.2.4 Purification of XerT

XerT overexpressed from XerT_pETM-22 was purified to homogeneity by combined affinity and size exclusion chromatography purification. The overexpressed TRX-6xHis-XerT construct was extracted from the soluble fraction of *E. coli* lysate by application to a HisTrap column, to which the construct binds specifically through its 6xHis affinity tag. The fractions eluting from the column in the presence of imidazole were analysed by SDS-PAGE and the fractions containing TRX-6xHis-XerT were combined. The protein was then subjected to proteolytic cleavage by 6xHis-tagged 3C protease, resulting in a cleavage of the TRX-6xHis tag from the XerT protein. A second purification on the HisTrap column

allowed for separation of tag-free XerT from the TRX-6xHis tag and the 6xHis-tagged 3C protease. The fractions containing XerT were again combined, concentrated, and loaded onto a gel filtration column. The obtained elution profile (Figure 6-9a) indicated that XerT elutes as a monomer from the gel filtration column. The peak fractions were analysed by SDS-PAGE (Figure 6-9b), confirming that the obtained protein was >95% pure and therefore suitable for further studies. Combined fractions 1-5 yielded approximately 2 mg of purified XerT from one litre of expression culture.

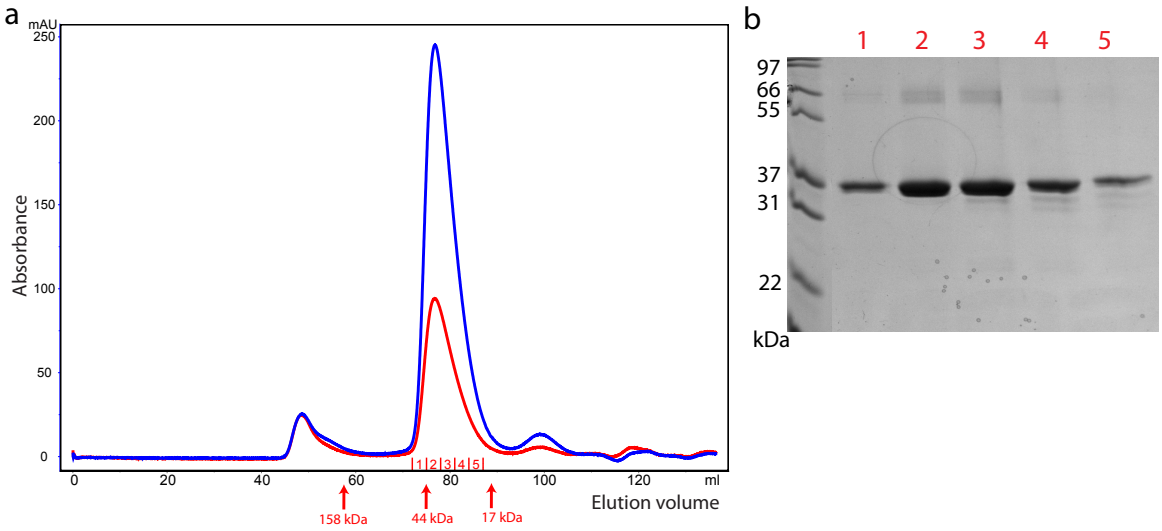


Figure 6-9: Purification of XerT. a) The final step of XerT purification: SEC on a Superdex 16/60 gel filtration column. The chromatogram shows the main elution peak that corresponds in size to a monomer of XerT (42 kDa). UV absorbance at 280 nm is shown in blue, absorbance at 260 nm is shown in red. The elution volumes of molecular weight markers of 17, 44, and 158 kDa on this column are indicated by red arrows. The fractions collected are shown in red under the elution peak. b) SDS-PAGE analysis of the fractions collected from the gel filtration, stained with Coomassie for the presence of protein.

6.2.5 Purification of XerT catalytic mutants Y332F and R300Q

In order to assess the functionality of XerT in *in vitro* cleavage assays with TnPZ ends, two catalytically dead XerT mutants were prepared to serve as negative controls. In the first mutant, the catalytic tyrosine (Tyr332) has been mutated to a phenylalanine (Y332F), while in the second mutant one of the catalytic arginines (Arg300) has been mutated to a glutamine (R300Q). A mutation of this arginine was shown to abolish catalytic activity of other tyrosine recombinases such as Cre, while it maintains the proper assembly of the active site unlike the tyrosine mutations (Guo *et al.,* 1999). Each mutation was introduced into the XerT_pETM-22 construct by site-directed mutagenesis (section 2.2.13). The behaviour of mutants was expected to be similar to that of wild-type XerT, and therefore

the same overexpression and purification protocols were used for the mutant proteins as for wild-type XerT.

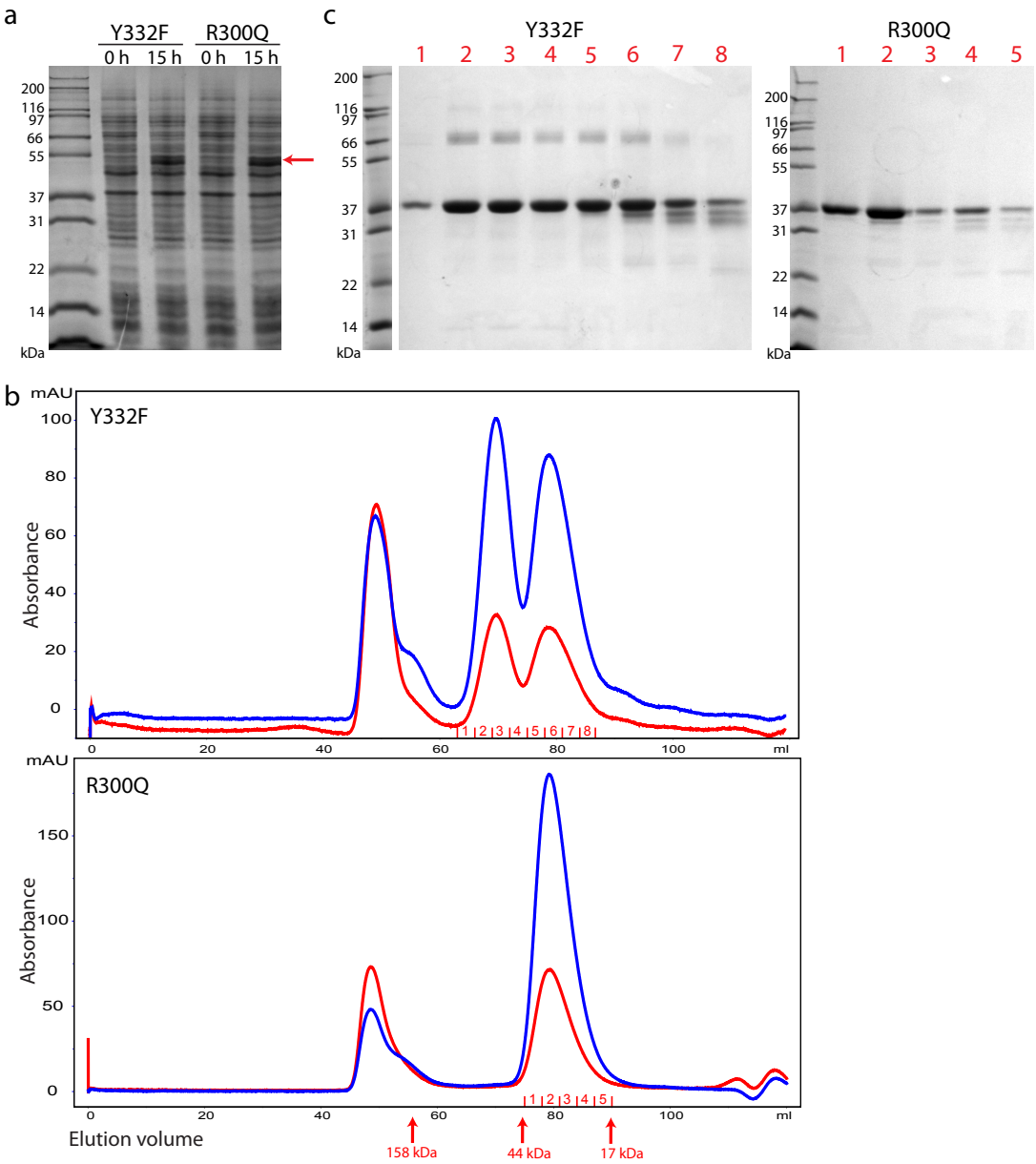


Figure 6-10: Overexpression and purification of XerT mutants Y332F and R300Q. a) Large scale expression of XerT mutants from pETM-22 in *E. coli* BL21 (DE3). The samples were analysed immediately after induction (0h) and 15 (15h) hours later. The positions of the expressed mutant XerT proteins on the gels are indicated by a red arrow. b) The final step of XerT mutant purification: SEC on a Superdex 16/60 gel filtration column. The chromatogram shows the main elution peaks that correspond in elution volume to a monomer (42 kDa) or a dimer (84 kDa) of XerT. UV absorbance at 280 nm is shown in blue, absorbance at 260 nm is shown in red. The elution volumes of molecular weight markers of 17, 44, and 158 kDa on this column are indicated by red arrows. The fractions collected are shown in red under the elution peak. c) SDS-PAGE analysis of the fractions collected from the gel filtration, stained with Coomassie for the presence of protein.

Overexpression in *E. coli* BL21 (DE3) with initial growth at 42 °C, followed by induction with IPTG and growth at 15 °C for 15 hours, resulted in a good amount of protein expression (Figure 6-10a). The proteins were purified using affinity and size exclusion chromatography following purification protocols as for wild-type XerT. In the final step of gel filtration, a single peak corresponding to an XerT monomer was seen for the R300Q mutant (Figure 6-10b). Interestingly, for the Y332F mutant two major peaks were seen: one corresponding to a monomer, and an additional peak corresponding in elution volume to an XerT dimer (Figure 6-10b). Analysis of the peak fractions by SDS-PAGE revealed that both peaks contained XerT (Figure 6-10c), implying that the observed additional peak corresponds to the dimer of XerT Y332F. After analysis by SDS-PAGE, fractions 1-5 for each XerT Y332F and R300Q were combined and used for further biochemical studies. This purification protocol yielded 1.5 mg of each mutant XerT protein from one litre of expression culture.

6.2.6 Identification of XerT binding sites at TnPZ ends

Previous studies of TnPZ elements did not investigate whether XerT directly interacts with TnPZ ends and, if so, what sequences are bound by XerT. To address these questions, a DNaseI footprinting experiment was performed. In this experiment, 200-bp oligonucleotide substrates containing approximately 100 bp of the TnPZ transposon end and 100 bp of the transposon flank sequences were 5'-end labelled with [γ-^{32}P]-ATP and incubated with XerT. XerT-bound DNA was subjected to DNaseI digestion and the samples were analysed on a denaturing sequencing gel alongside a sequencing ladder. A region of DNA that was not digested by DNaseI as a result of protection by bound XerT could be identified for substrates containing each TnPZ end (Figure 6-11a). Both protected regions were digested in the absence of XerT, suggesting that the observed pattern is directly caused by XerT binding to the substrate DNA, which confirms the direct interaction between XerT and TnPZ ends. The sequences of the bound regions were mapped by the sequencing ladders run alongside the reactions, and are shown in Figure 6-11b. The length of the identified binding sites was 34 bp for the left end (LE) of the transposon and 31 bp for the right end (RE), consistent with binding of two XerT monomers at each site, as seen for XerH and other tyrosine recombinases. In both cases, the XerT-protected region included the heptanucleotide 5'AAGAATG that marks the transposon ends located at the centre. The central position of the heptanucleotide further

suggested that this sequence might act as the central region between the two XerT binding sites as previously proposed (Kersulyte *et al.,* 2009), although the region was not placed exactly in the centre but rather shifted by 2-3 nt to the right. Importantly, the results suggest that the transposon-flanking DNA is also contacted by XerT.

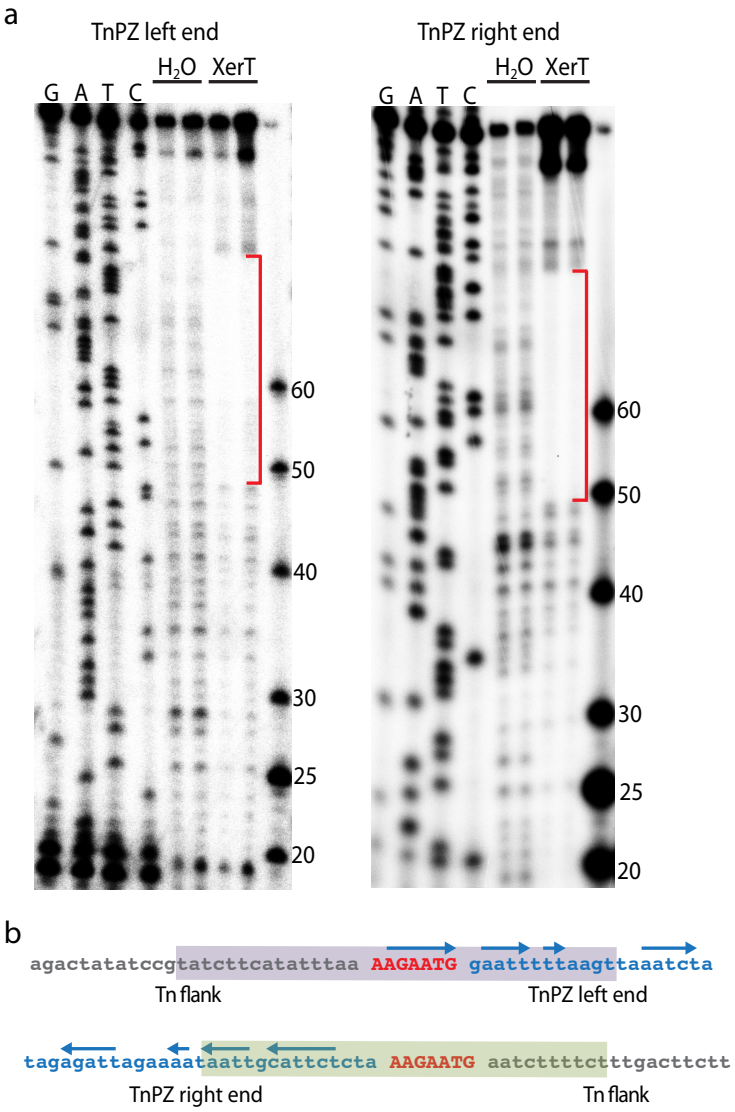


Figure 6-11: DNaseI footprinting assay to map the XerT-protected regions at TnPZ ends. a) A DNA substrate containing the left TnPZ end with corresponding flanking DNA (left panel) or the right TnPZ end with its flanking DNA (right panel) was incubated with XerT and then digested with DNaseI to reveal the XerT-protected (and therefore undigested) region of each substrate. As a control, reaction without XerT ("H_2O") was set up and run alongside the reactions ("XerT"). The two samples for each "XerT" and "H_2O" contained two different concentrations of DNaseI: 120 (left) and 95 (right) µg/ml. The sequence of the protected region (marked by red clamps) can be directly identified by comparison with the sequencing ladder run alongside the samples. The sizes of molecular weight markers (in nucleotides) are shown next to each gel. b) The XerT-protected regions at TnPZ ends identified in the DNaseI footprinting experiment. The protected regions are marked with a lilac box for the left TnPZ and a green box for the right end. The transposon sequences are shown in blue and the flanking *H. pylori* sequences are shown in grey letters. The conserved heptanucleotide 5'AAGAATG is shown in red. The arrows above the sequences indicate the inverted repeat regions at the two transposon ends.

A closer inspection of the two XerT-bound regions revealed their internal symmetry, which includes 18 nt at the RE, but only 8 nt at the LE (Figure 6-12). In summary, the DNaseI footprinting experiment confirmed direct XerT-TnPZ interactions and identified the XerT-bound DNA sequences within the transposon end region.

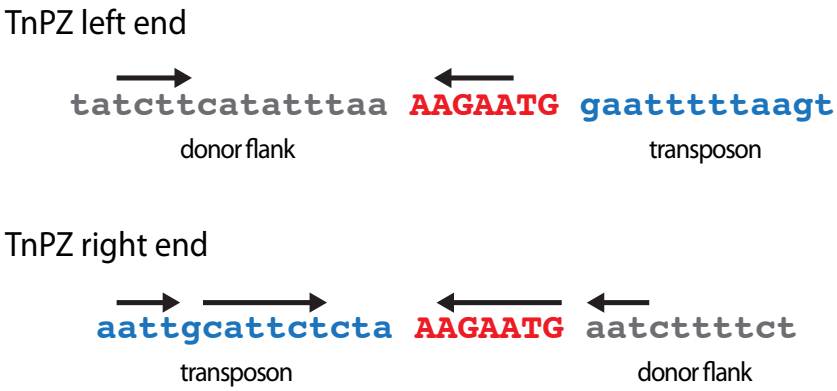


Figure 6-12: Internal symmetry of the XerT-protected TnPZ end regions. The transposon sequences are shown in blue and the flanking *H. pylori* sequences are shown in grey. The conserved heptanucleotide 5'AAGAATG is shown in red. The arrows indicate the inverted repeat regions within each XerT recombination site.

6.2.7 Detection of XerT-TnPZ end complexes

Since an XerT-protected region could be detected in the DNaseI footprinting assay, it was reasonable to assume that XerT and TnPZ directly interact *in vitro*. In order to directly detect and further analyse these interactions, EMSA experiments were performed, but no complexes could be detected on polyacrylamide gels prepared in a variety of buffers, including TBE and Tris-glycine buffers commonly used for native PAGE. Upon addition of XerT, the free DNA bands disappeared and shifted into the gel wells, suggesting that complexes were formed, but could not enter the gel, which could be caused for example by unfavourable gel pH or buffer conditions that cause precipitation of the complex.

As the XerT-TnPZ complexes could not be detected in EMSA, the complexes were also analysed by SEC on an analytical gel filtration column. For this, samples of XerT alone, 50 bp TnPZ-end DNA alone, and XerT incubated with TnPZ-end DNA were run separately on the analytical gel filtration column and the chromatograms from all the runs were compared (Figure 6-13).

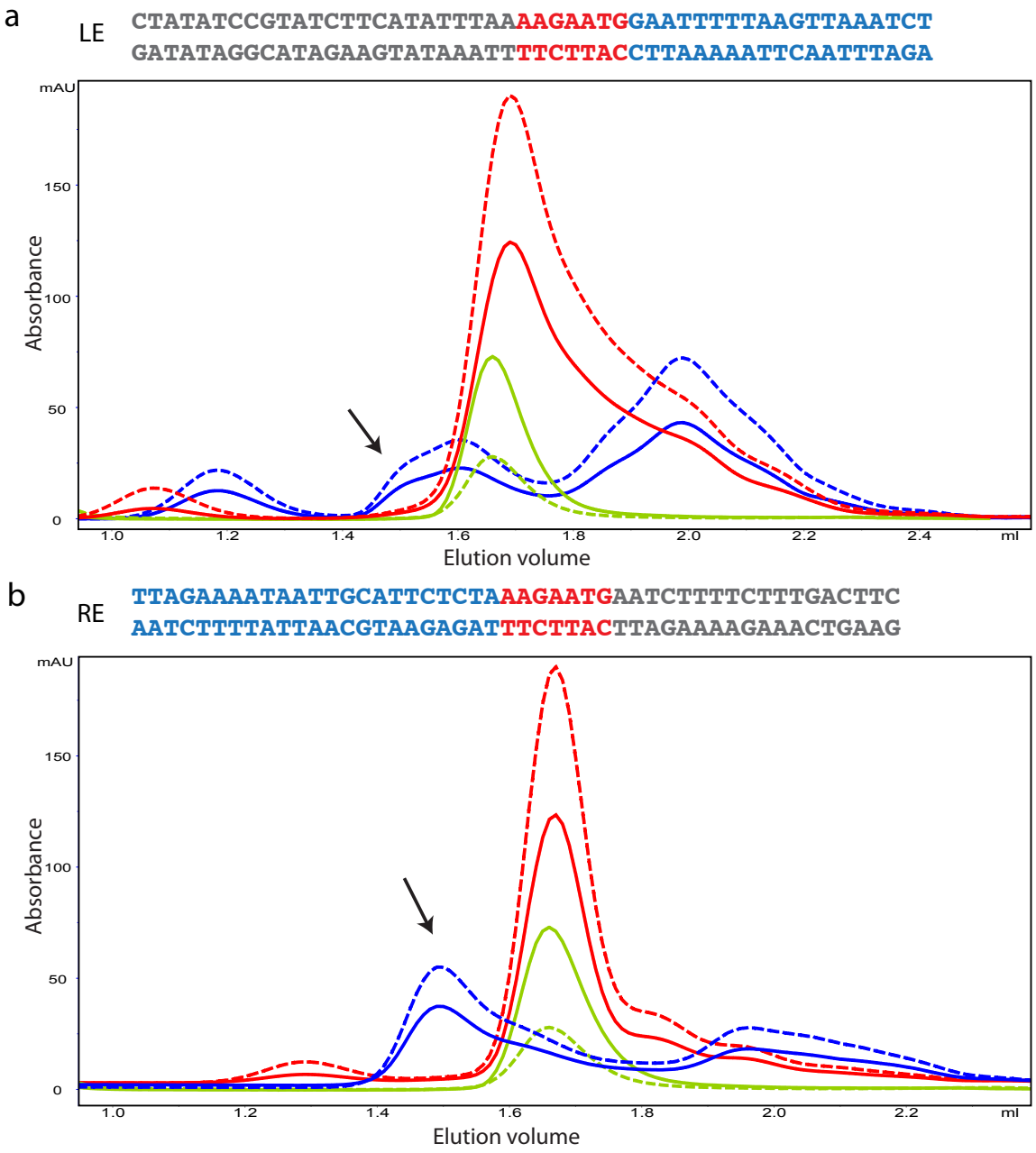


Figure 6-13: Analysis of XerT-TnPZ complex formation by analytical SEC on a Superdex 3.2/30 gel filtration column. The overlaid chromatograms include elution profiles of XerT alone (green), a 50-bp TnPZ end DNA alone (red), and the complex of XerT and 50-bp TnPZ end DNA (blue). UV absorbance at 280 nm is shown as continuous line, absorbance at 260 nm is shown as a dashed line. Arrows indicate elution peaks of XerT-TnPZ complexes. The TnPZ oligonucleotides used for the formation of the complexes, shown above the chromatograms, include the TnPZ end (blue), the conserved heptanucleotide 5'AAGAATG (red), and the transposon flanking sequences (grey). a) Analysis of the XerT-TnPZ LE complex. b) Analysis of the XerT-TnPZ RE complex.

XerT binding to both TnPZ ends was analysed in this assay, with 2.5-fold molar excess of XerT over DNA used in each case. Both chromatograms showing the XerT-TnPZ complex elution contained peaks of lower height than expected based on the amount of sample prepared, which suggested that a large fraction of the complex precipitated and was lost

upon sample filtration before application to the gel filtration column. Nevertheless, in both cases elution peaks corresponding to higher molecular weight complexes could be detected (marked with arrows in Figure 6-13). Based on their elution volume, these likely correspond to a dimer of XerT bound to the TnPZ end. Additionally, peaks corresponding to excess DNA were also present in each case, although it seems that the observed DNA species had lower molecular weight than that of double-stranded DNA, likely corresponding to single-stranded or hairpin DNA substrates. Interestingly, these experiments also revealed that the LE of TnPZ was bound less efficiently than the RE, with large excess of unbound DNA (Figure 6-13a). The more prominent RE complexes showed a major peak corresponding to the XerT dimer bound to the transposon end. In summary, the analytical gel filtration assay allowed for direct detection of XerT-TnPZ complexes, but also showed that such complexes are rather unstable and prone to precipitation. Thus, the complexes formed were not suitable for crystallization trials and require further optimization.

6.2.8 DNA cleavage activity of XerT

With the XerT binding sites at the ends of TnPZ identified and the binding reconstituted, it was interesting to test if XerT can cleave the bound DNA substrates. Based on analogy with XerH and other tyrosine recombinases, it was expected that the cleavage activity would be difficult to detect with native DNA substrates, and therefore pre-nicked suicide substrates were employed (for an explanation on the principles of suicide substrate design, see Figure 3-14a). For successful entrapment of covalent complexes between the tyrosine recombinase and the DNA substrate, the nick has to be placed 1-3 nt downstream from the cleavage site. Therefore, efficient design of suicide substrates requires proper knowledge about the cleavage site position. For XerT, the positions of the cleavage sites were not known, but the heptanucleotide 5'AAGAATG at the transposon end was previously proposed to act as the central region of the recombination site (Kersulyte *et al.,* 2009), suggesting that the cleavage could take place at the boundaries of this heptanucleotide. Accordingly, nicks were introduced one nt away from the heptanucleotide boundaries at the RE of TnPZ (Figure 6-14a). Two different nicks were introduced into each strand: one that would trap the expected 3'-phosphotyrosyl intermediate (Substrates 2 (top strand) and 5 (bottom strand) in Figure 6-14a) and one that would trap a 5'-phosphotyrosyl intermediate, which was not expected but was tested since the polarity of XerT cleavage

was not known (Substrates 3 (top strand) and 4 (bottom strand) in Figure 6-14a). Figure 6-14 shows the substrates and cleavage analysis for the RE of TnPZ only. Cleavage experiments with the TnPZ LE (not shown) followed the same cleavage pattern, although with slightly lower efficiency.

The covalent complexes formed between the RE native and suicide substrates and XerT were detected by SDS-PAGE (Figure 6-14b). A very small amount of covalent complex could be detected when an unnicked substrate was used (lane 1, Figure 6-14b). The same or lower amount of cleavage product could be detected for the substrates that were intended to trap the 3' phosphotyrosyl intermediate (lanes 2 and 5, Figure 6-14b). Surprisingly, a high amount of cleavage product could be seen when the substrate containing the nick on the left side of the heptanucleotide in the bottom strand was used (lane 4, Figure 6-14b). This could suggest that the nature of XerT cleavage is different than that of other tyrosine recombinases and involves a 5'- rather than 3'-phosphotyrosyl intermediate. However, this is very unlikely due to the high conservation of the canonical tyrosine recombinase active site in XerT, and also because hardly any cleavage product could be seen when a corresponding position on the top strand was nicked (lane 3, Figure 6-14b). Therefore, it is far more likely that the true cleavage site was not where it was predicted to be, and the nick on Substrate 4 was placed 1-3 nt downstream from the true cleavage site somewhere inside the heptanucleotide, allowing detection of the 3'-phosphotyrosyl cleavage product (Figure 6-14c).

In order to confirm that the observed covalent intermediate was a true product of a tyrosine recombinase cleavage that is dependent on the presence of canonical catalytic residues, Substrate 4 (Figure 6-14a) was incubated with the XerT mutants Y332F and R300Q in conditions supporting wild-type XerT cleavage. In both cases, no cleavage products could be observed, suggesting that both conserved residues Tyr332 and Arg300 are necessary for the observed XerT-mediated cleavage (Figure 6-14d).

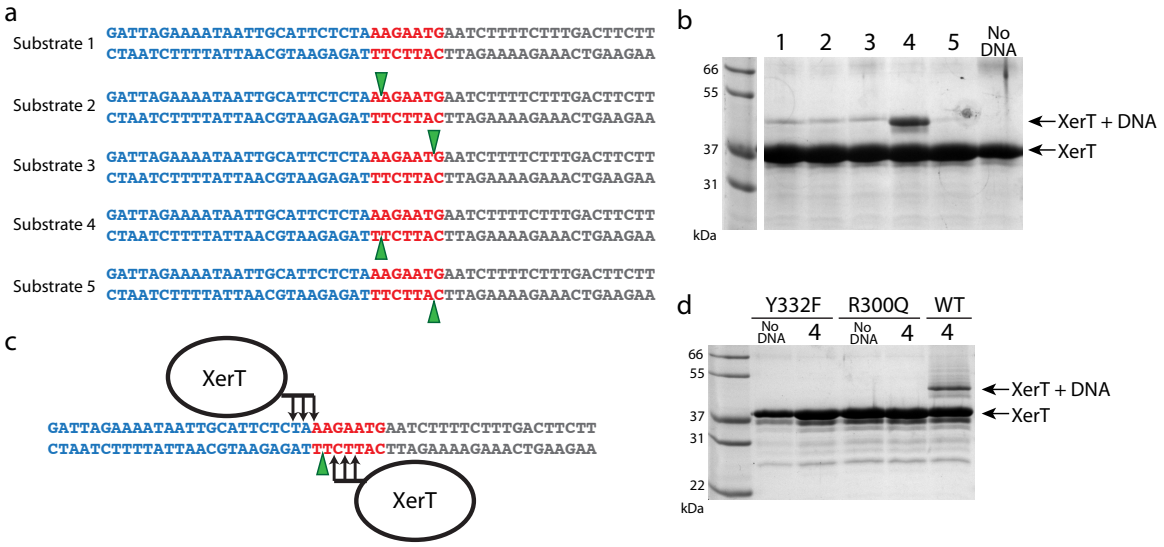


Figure 6-14: XerT-mediated cleavage of TnPZ suicide substrates *in vitro*. a) The TnPZ-RE substrates used to assess cleavage activity of purified XerT. Substrate 1 is the native unnicked substrate, while Substrates 2-5 are suicide substrates containing nicks in top or bottom strands (marked with green triangles) 1 nt away from the end of the conserved heptanucleotide (shown in red). The transposon sequences are shown in blue and the flanking *H. pylori* sequences are shown in grey. b) SDS-PAGE analysis of DNA cleavage reactions with XerT and TnPZ-RE suicide substrates shown in (a). The numbers above the gel correspond to the numbers of the DNA substrates shown in (a). (c) Possible XerT cleavage positions, inferred from the results of DNA cleavage experiment shown in (b). d) SDS-PAGE analysis of DNA cleavage reactions with XerT wild-type (WT), Y332F, and R300Q, and the TnPZ-RE suicide Substrate 4 shown in (a).

To further confirm that the observed cleavage products represent the true intermediates of an XerT-mediated recombination reaction, their ability to proceed into strand exchange products was investigated. The most efficient TnPZ RE substrate with the nick in the bottom strand (Substrate 4 in Figure 6-14a) was 5'-end labelled with [γ-^{32}P]-ATP and incubated with XerT and an excess of various unlabelled, un-nicked substrates (Figure 6-15a). After the reaction was completed, XerT was removed by digestion with Proteinase K and the DNA was purified by ethanol precipitation. The resulting samples were analysed on a 12% Urea-TBE sequencing gel (Figure 6-15b). A number of strand-exchange products of various sizes could be observed in the presence of unlabelled DNA substrates, suggesting that the cleaved intermediate indeed proceeds into strand exchange products. Interestingly, no such products could be observed without addition of unlabelled DNA, and this observation was consistent across many experiments with XerT and various ^{32}P-labelled substrates. Furthermore, a cleavage product with smaller size than the substrate could also not be detected directly, while in many experiments (including those shown in lanes 1 and 3 in Figure 6-15b) a band corresponding to slightly larger size than that of the substrate was found upon incubation with XerT. This band could correspond to the cleavage product, in which the XerT bound through the phosphotyrosyl bond was not

completely removed by Proteinase K treatment, leaving a small number of amino acids still attached to the DNA.

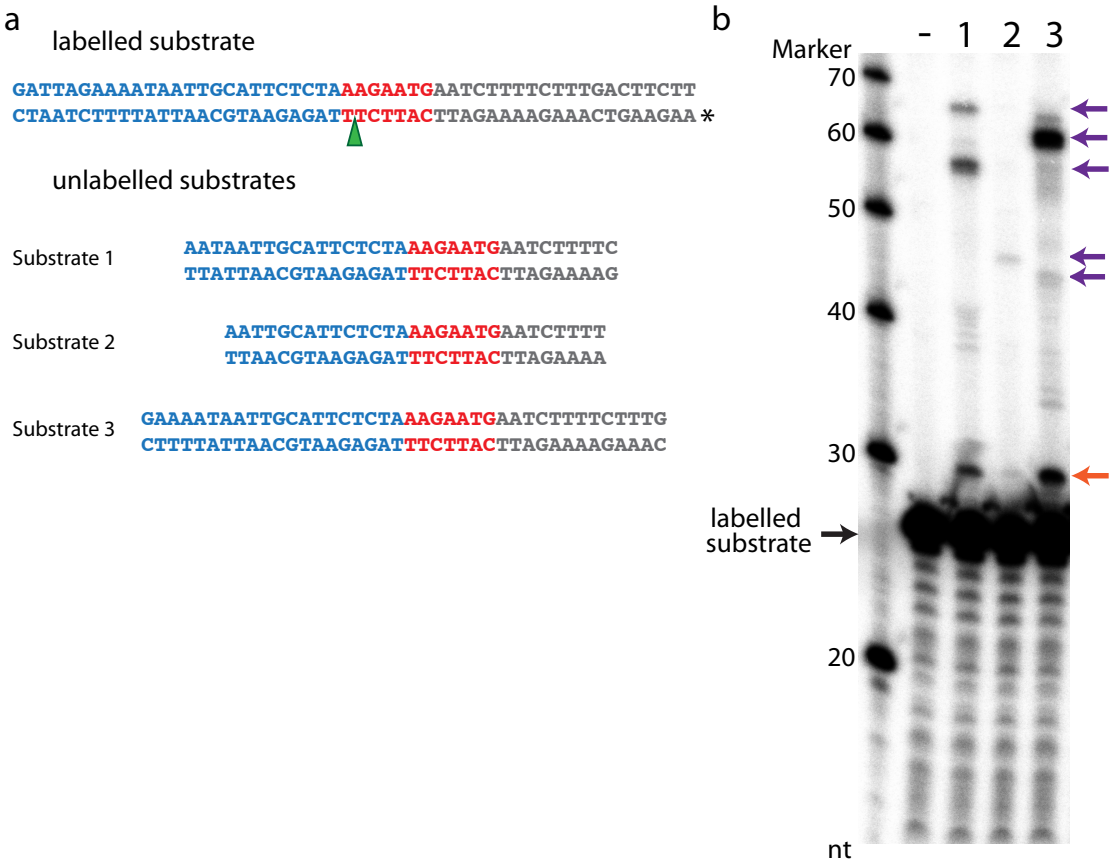


Figure 6-15: Analysis of strand-exchange reactions catalysed by XerT. a) The TnPZ-RE substrates used to assess strand-exchange between a phosphotyrosyl intermediate formed upon cleavage of a labelled suicide substrate (^{32}P label denoted by a star) and un-nicked RE substrates of various lengths. The transposon sequences are shown in blue and the flanking *H. pylori* sequences are shown in grey. The conserved heptanucleotide 5'AAGAATG is shown in red. b) Analysis of the strand-exchange reaction on a 12% Urea-TBE sequencing polyacrylamide gel. The numbers above the gel indicate the unlabelled substrate used in the reaction according to the numbering in (a). (–) denotes no unlabelled substrate used. The arrows mark the putative products of the cleavage (orange) and strand-exchange (purple) reactions.

In summary, the results of the cleavage assays show that XerT is proficient in cleavage of both TnPZ ends and the cleavage products likely represent a functional intermediate of the recombination reaction, although the cleavage sites are likely not located at the flanks of the heptanucleotide 5'AAGAATG as previously predicted.

6.2.9 Mapping of the XerT cleavage sites

Since it became apparent that the XerT cleavage sites might be located at different positions than previously suggested (Kersulyte *et al.,* 2009), a mapping experiment was performed to identify the positions of XerT cleavage on TnPZ ends. This experiment was based on the SDS-PAGE cleavage assay with suicide substrates, taking advantage of the fact that for proteins that form a 3' linkage to the DNA, covalent complexes can only be detected if the nick of the suicide substrate is introduced 1-3 nt downstream from the protein cleavage site (Figure 6-16a). Therefore, by placing the nick at different positions in the suspected cleavage region, the cleavage site can be identified. The observed bottom strand (Figure 6-14b) and the inferred top strand (Figure 6-14c) cleavage positions were used to identify regions potentially containing the XerT cleavage sites for both left and right TnPZ ends, and the suicide substrates mapping these regions were designed accordingly (Figure 6-16b). Analysis of the cleavage products after incubation of each suicide substrate with XerT allowed for identification of the cleavage sites (Figure 6-16c). For example, four suicide substrates with differently positioned nicks were investigated for the cleavage of the bottom strand of LE (lanes 5-8, Figure 6-16c). The first two substrates (lanes 5-6) show cleavage product formation, suggesting that the XerT cleavage site is not placed there but further upstream. However, no product can be detected with the next substrate (lane 7) and the subsequent substrate with a nick further upstream (lane 8), suggesting that the XerT cleavage site is located at the position where the nick in substrate 7 was placed. All other cleavage sites were similarly identified (Figure 6-16c), with the top strand cleavage of the left TnPZ end being less efficient and the determination of its position more ambiguous than in other cases. When the newly identified cleavage sites were drawn onto the TnPZ end sequences (Figure 6-16d), it could be seen that the cleavage sites on each TnPZ end are spaced by a 6-bp central region differing only by one basepair between the two TnPZ ends. Interestingly, the centre of the central region in the RE was at the same position as the centre of symmetry previously shown for the protected region of the RE (Figure 6-12), further supporting the notion that the cleavage sites were mapped correctly.

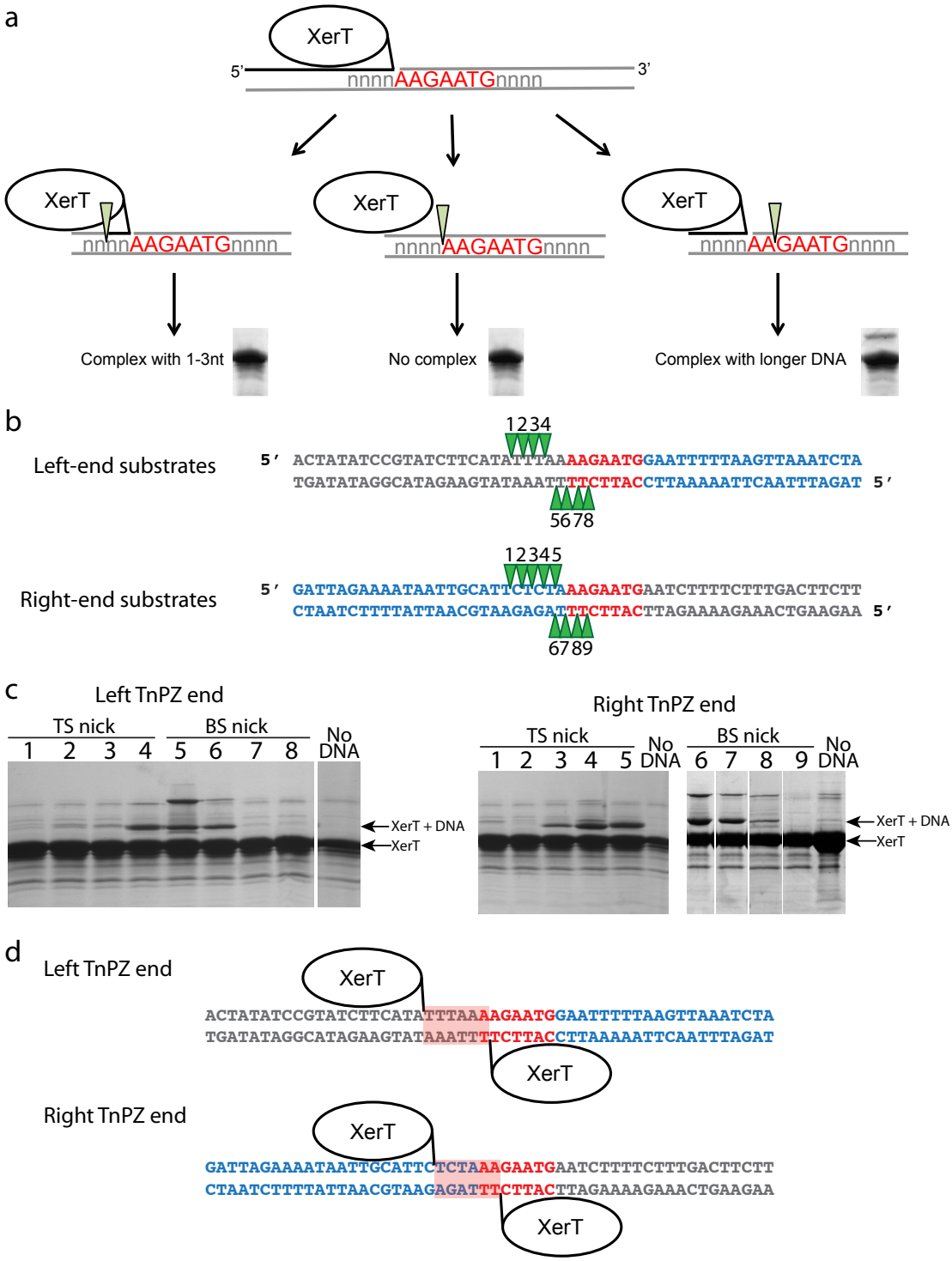


Figure 6-16: Mapping of the XerT cleavage site at TnPZ ends. a) The design of the assay. The DNA is schematically shown in grey or in black for the fragment that is covalently attached to XerT after cleavage. The conserved heptanucleotide sequence is shown in red. b) TnPZ substrates used to identify XerT cleavage sites. The green triangles denote the position of the pre-introduced nick in each substrate, numbered above the triangle. The transposon sequences are shown in blue and the flanking *H. pylori* sequences are shown in grey. The conserved heptanucleotide is shown in red. c) SDS-PAGE analysis of the cleavage products obtained in the cleavage site mapping experiment. The lanes comprised the reactions with substrates containing a nick in top (TS) or bottom (BS) strand of DNA, and are numbered according to (b). (d) XerT cleavage positions identified in the assay. The DNA sequences are coloured as in (b). The pink box denotes the central region between the two arms of each recombination site.

6.2.10 Confirmation of XerT cleavage sites with a half-site cleavage assay

The initial identification of XerT cleavage sites on TnPZ ends required further confirmation; for that, a half-site cleavage assay was performed. The assay has been previously used for identification of cleavage sites of various tyrosine recombinases, including XerA recombinase (Serre *et al.,* 2013). In this assay, substrates corresponding to one half of the recombination site (and therefore to the binding site for one XerT monomer) were 5'-end labelled with [γ-^{32}P]-ATP and incubated with XerT and an excess of unlabelled substrate corresponding to the second half of the recombination site (Figure 6-17a). After cleavage of the labelled half-site by XerT, the 5' hydroxyl group of the unlabelled half-site attacks the phosphotyrosyl bond in a reaction that mimics strand exchange and results in the formation of longer, XerT-free DNA products, which can be detected on the urea-TBE sequencing gel (Figure 6-18b). Since the size of both oligonucleotides is known, the cleavage position can be estimated based on the size of the resulting strand exchange product. In the example of TnPZ LE, labelling of the bottom strand of the right arm (20 nt) and reaction in the presence of the left arm (21 nt) results in formation of a strand-exchange product (38 nt) that is three nucleotides shorter than the sum of the two substrates (LE, lane 6, Figure 6-17b) indicating that three nucleotides have been cleaved off the bottom strand of the labelled right arm. As an additional confirmation, the reaction was also performed in the presence of the unlabelled right arm substrate, and the results confirm that three nucleotides have been cleaved off (LE, lane 5, Figure 6-17b). No strand exchange products could be observed in the absence of unlabelled substrates (LE, lane 4, Figure 6-17b), consistent with previous experiments. The assay confirmed the cleavage sites of three out of four TnPZ binding arms tested: the cleavage position of the left arm of the LE could not be determined as no cleavage or strand exchange products could be observed (LE, lanes 1-3, Figure 6-17b). This shows that the cleavage of this arm is very inefficient and agrees with the previous results of the SDS-PAGE cleavage assay (Figure 6-16c). Additionally, the left arm of the RE showed detectable strand exchange product only in the presence of the same unlabelled half-site (RE, lanes 1-3, Figure 6-17b), but that was sufficient for confirmation of the cleavage site on this arm. The lack of or very inefficient DNA cleavage and strand exchange observed when labelled LE and RE left arm substrates were used might suggest that these sequences do not support XerT cleavage and recombination in absence of the corresponding right arm on the same DNA.

In summary, the results of the half-site cleavage assay confirmed the positions of XerT cleavage on TnPZ ends and showed that the bottom strand (right arm) cleavage of both TnPZ ends is more efficient than that of the top strand (left arm).

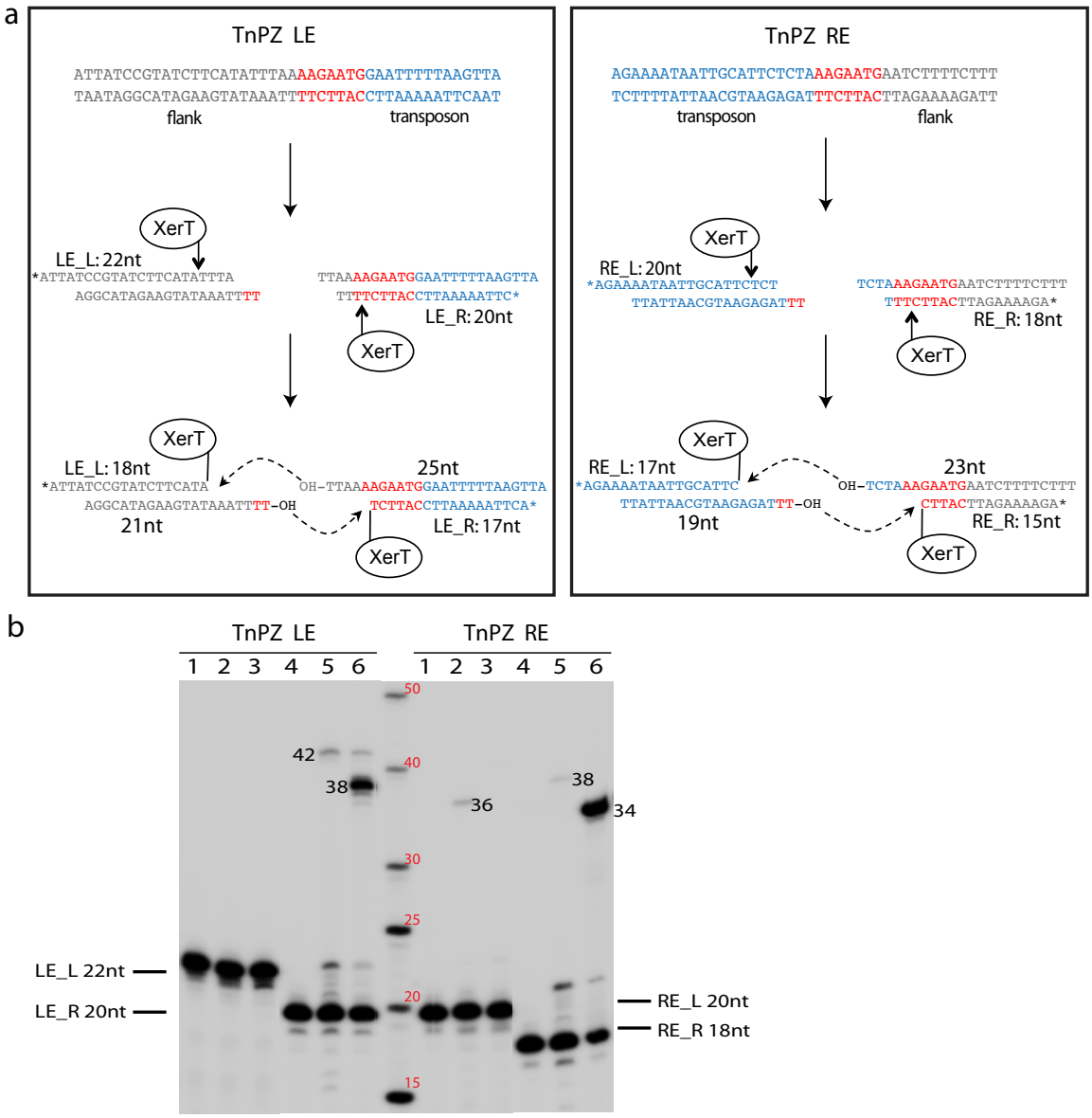


Figure 6-17: Confirmation of the XerT cleavage sites at TnPZ ends by the half-site cleavage assay. a) Design of the half-site cleavage assay for TnPZ LE and RE. The positions of the ^{32}P label are indicated by stars. The transposon sequences are shown in blue and the flanking *H. pylori* sequences are shown in grey. The conserved heptanucleotide 5'AAGAATG is shown in red. The size of each strand in each substrate is indicated. b) Analysis of the half-site cleavage and strand-exchange reaction products on a 12% Urea-TBE sequencing polyacrylamide gel. Lanes 1 and 4: reactions with no unlabelled substrates; lane 2: reaction with unlabelled left-arm substrate of the corresponding transposon end, identical to the labelled substrate; lane 3: reaction with corresponding unlabelled right-arm substrate, complementary to the labelled substrate; lane 5: reaction with unlabelled right-arm substrate, identical to the labelled substrate; lane 6: reaction with unlabelled left-arm substrate, complementary to the labelled substrate. The numbers in black next to the bands on the gel correspond to the sizes of the obtained products (in nucleotides). The numbers in red correspond to the sizes of the DNA marker (in nucleotides). The expected reaction products are missing from lanes 2 and 3 for TnPZ LE and lane 3 for TnPZ RE, likely because the substrates used do not support efficient cleavage and/or strand exchange.

6.2.11 Mutational analysis of TnPZ ends

In order to further investigate the requirements for XerT-mediated binding and cleavage at the TnPZ ends, a mutational analysis of the sequences bound by XerT was performed. The cleavage assays (described in sections 6.2.9 and 6.2.10) showed that the heptanucleotide 5'AAGAATG present at both TnPZ ends does not constitute the central region between the two cleavage sites, but could instead act as a necessary binding site for one monomer of XerT. With this in mind, a range of mutated recombination sites was created and tested in the SDS-PAGE cleavage assay in order to identify sequences necessary for the XerT-TnPZ interaction and perhaps aid creation of symmetrized substrates that would result in more stable XerT-TnPZ complexes suitable for crystallization. A selection of mutated substrates is shown in Figure 6-18a for the right TnPZ end and in Figure 6-19a for the left end.

If XerT recombination follows a similar binding and cleavage pattern as observed for XerH, where one arm is bound more efficiently while the other is cleaved with much higher efficiency, it would suggest that the left arms of the transposon recombination sites might be the arms that are more efficiently bound (since the right arms (bottom strand) are cleaved with higher efficiency). Accordingly, it could be beneficial to symmetrize the recombination site so that it resembles the left-arm palindrome. Since the RE recombination site already shows considerable symmetry (Figure 6-12), this transposon end was used to design fully symmetric substrates. The first efforts to symmetrize the RE tested mutations within the central region, in which the top strand sequence TCTAAA was mutated to palindromic TTTAAA or TCTAGA (Substrates 1 and 2, Figure 6-18a). Cleavage products were observed in both cases (Figure 6-18b, lanes 1 and 2), although the cleavage efficiency was reduced when TCTAGA sequence was used, suggesting that the adenine in the original sequence might be important for XerT recombination (which could be expected since the adenine is a part of the absolutely conserved heptanucleotide). The next substrates tested were left-arm palindromes with unchanged or changed central region (Substrates 3 and 4, Figure 6-18a). These substrates were cleaved with higher efficiency than the wild-type recombination site as long as the central region was unchanged (Figure 6-18b, lanes 3 and 4).

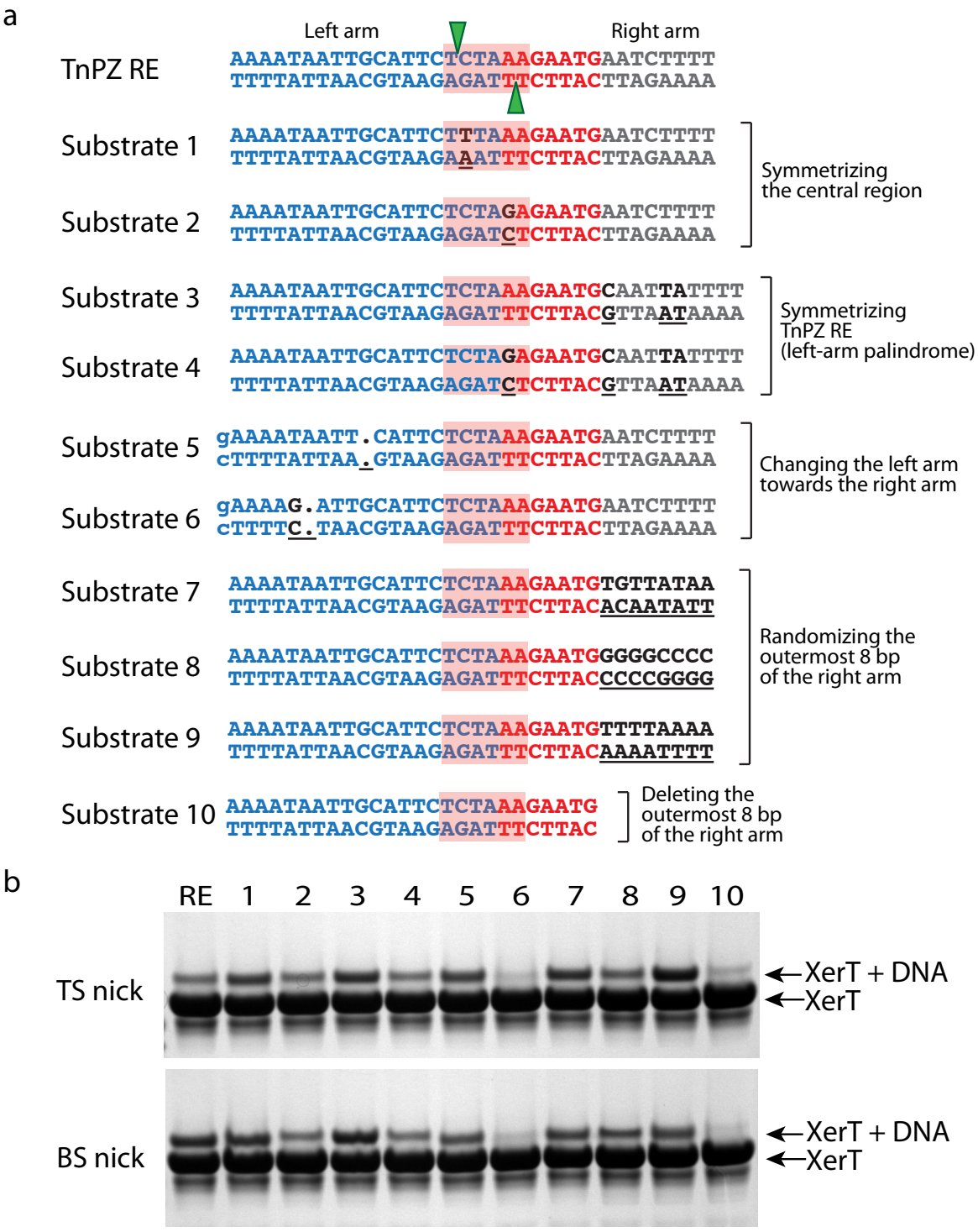


Figure 6-18: Mutational analysis of TnPZ-RE recombination substrates. a) The mutated TnPZ-RE substrates used in the experiment. The wild-type RE sequence is shown on top for reference. The substrates test the impact of mutating various parts of the left and the right arm of RE on cleavage by XerT, and the basis for different XerT cleavage efficiency on the left arm (top strand) and the right arm (bottom strand) of RE. The transposon sequences are shown in blue and the flanking *H. pylori* sequences are shown in grey. The conserved heptanucleotide 5'AAGAATG is shown in red. The central region of the recombination site is marked by an orange box. The introduced mutations are shown in black and underlined, and the positions of the top and bottom strand nicks (shown for the original RE sequence only) are indicated by triangles. b) SDS-PAGE analysis of XerT-mediated cleavage of the suicide substrates containing a nick in the top (TS) or the bottom (BS) strand. The numbers above the gel correspond to the substrate number in (a). The positions of the nicks are shown in (a).

To find out, what part of the left arm determines the efficiency of cleavage, the sequences that differ between the two arms were mutated in the left arm towards the sequences in the right arm (Substrates 5 and 6, Figure 6-18a). Removing the G-C basepair from the left arm did not affect the cleavage (Figure 6-18b, lane 5), while mutating the TA-AT basepairs further upstream to a single G-C basepair abolished cleavage of both strands (Figure 6-18b, lane 6). This suggests that the TA-AT basepair at this position of the left arm is required for cleavage, perhaps by being required for efficient binding of the first XerT monomer.

To assess the importance of the specific sequence outside of the conserved heptanucleotide 5'AAGAATG in the right arm of RE, the outermost 8 bp of the right arm were mutated to unrelated sequences or removed (Substrates 7-10, Figure 6-18a). The XerT cleavage activity on substrates containing changed outermost bp was similar to that observed on the wild-type RE, but removal of the 8 bp resulted in abolished cleavage (Figure 6-18b, lanes 7-10), suggesting that the recombinase-DNA interactions within this region of the right arm of RE are not sequence-specific.

A mutational analysis of the TnPZ LE was also performed (Figure 6-19). Since the cleavage of the LE (in particular of its left arm) was less efficient than that of the RE, the LE sequences were changed so as to resemble the RE sequences to help derive the identity of nucleotides necessary for efficient cleavage of observed for the RE (Substrates 1-3, Figure 6-19a). The change of LE central region sequences (TTTAAA on the top strand) to those of the RE central region (TCTAAA) did not affect the cleavage efficiency (Figure 6-19b, lane 1). Changing the sequences within the right arm of LE towards those of right arm of RE improved cleavage of the top strand (left arm), but did not affect cleavage activity of the bottom strand (right arm). In addition, new bottom strand cleavage products appeared when these substrates were used (Figure 6-19b, lanes 2 and 3), perhaps representing aberrant, alternative cleavage positions.

Since the left arm sequences of the LE are corresponding to the donor flank, these sequences will differ as the transposon is inserted into new locations. Accordingly, the importance of the identity of these sequences for XerT cleavage was investigated by mutating the 20 bp of the left arm of LE to random sequences with variable G+C content (Substrates 4-6, Figure 6-19a). The results showed similar or slightly reduced cleavage

efficiency of both strands for all tested substrates, regardless of the G+C content of the random sequences (Figure 6-19b, lanes 4-6).

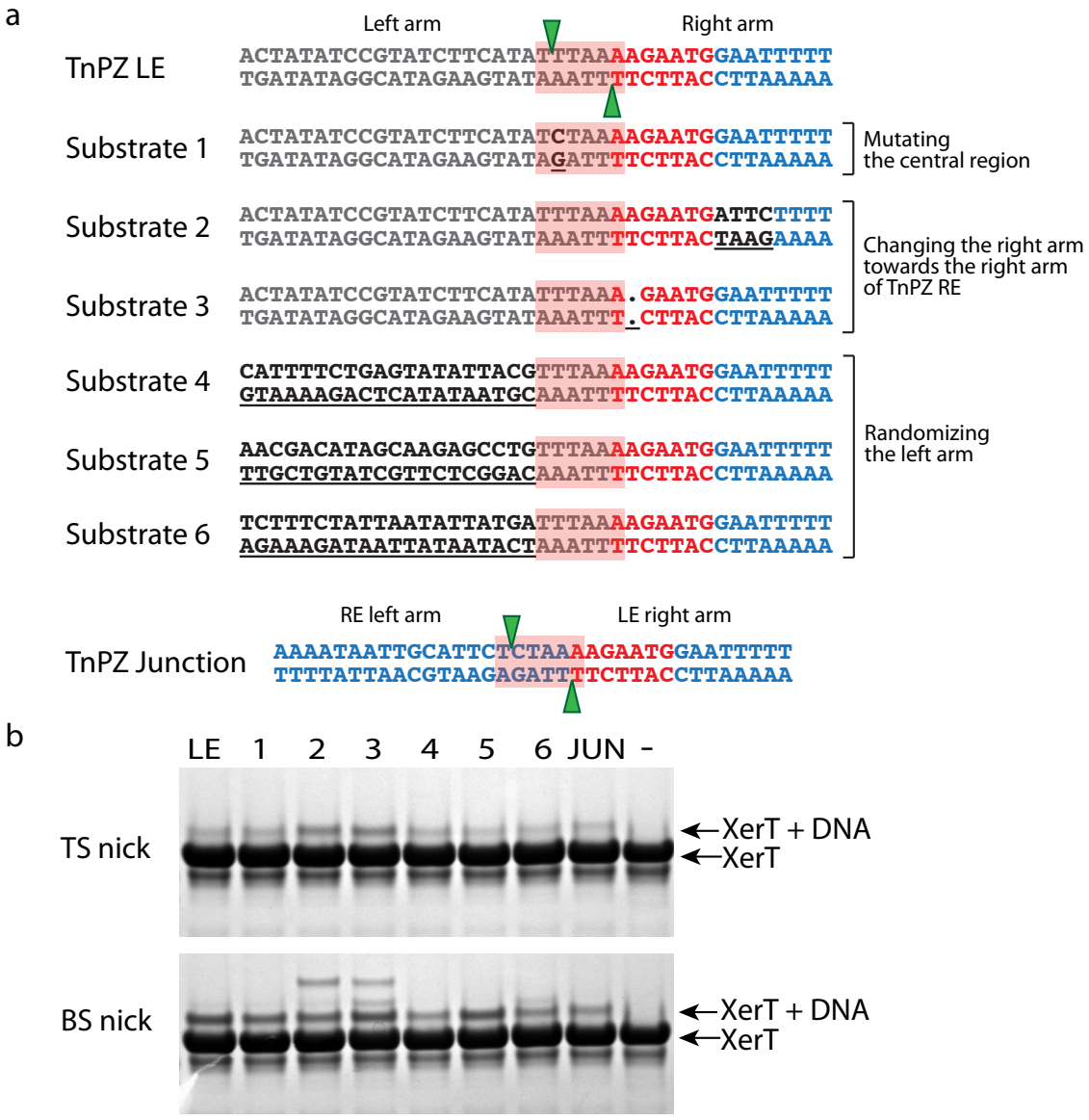


Figure 6-19: Mutational analysis of TnPZ-LE recombination substrates. a) The mutated TnPZ-LE substrates used in the experiment. The wild-type LE sequence is shown on top for reference. The substrates test the impact of mutating various parts of the left and the right arm of LE on cleavage by XerT, and the basis for different XerT cleavage efficiency at the LE and RE. Additionally, a substrate resembling the transposon end junction in the excised circular TnPZ transposon (TnPZ Junction) was also analysed. The transposon sequences are shown in blue and the flanking *H. pylori* sequences are shown in grey. The central region of the recombination site is marked by an orange box. The introduced mutations are shown in black and underlined, and the positions of the top and bottom strand nicks (shown for the original LE sequence only) are indicated by triangles. b) SDS-PAGE analysis of XerT-mediated cleavage of the suicide substrates containing a nick in the top (TS) or the bottom (BS) strand. The numbers above the gel correspond to the substrate number in (a). The position of the nicks is shown in (a). JUN denotes the TnPZ Junction DNA substrate. (–) denotes a control with no DNA.

Finally, a substrate containing both TnPZ ends in an arrangement that could correspond to the transposon junction in the excised circular TnPZ transposon was designed. In this substrate, the left arm of RE and the right arm of LE, both corresponding to the sequences encoded by TnPZ, were joined together with the central region in the middle of the substrate (Junction, Figure 6-19b). The cleavage efficiency of this substrate was even lower than that of the LE of TnPZ (JUN, Figure 6-19b).

With correctly mapped cleavage sites, the efficiency of XerT-mediated cleavage of the two DNA strands at each transposon end could be directly compared. Both RE and LE are cleaved more efficiently on the bottom strand than on the top strand (RE in Figure 6-18b and LE in Figure 6-19b), suggesting that XerT-mediated cleavage at TnPZ ends is orchestrated and might follow a specific cleavage order, probably starting on the bottom strand.

In summary, the mutational analysis of TnPZ ends provided first insights into sequence requirements for a functional XerT recombination site and showed that XerT binding and cleavage events at the TnPZ ends might be ordered and orchestrated as expected for a tyrosine recombinase.

6.2.12 Mini-transposon excision in *E. coli*

In order to find out whether the cleavage by XerT is sufficient for excision of TnPZ or whether additional factors encoded by TnPZ are required, the *galK*-MacConkey recombination assay was performed (see section 5.2.6; Arnold *et al.,* 1999). This recombination assay can function as a transposon excision assay if the two transposon ends are placed to flank the *galK* gene. The design of the assay is shown in Figure 6-20a. In detail, the *galK*-deficient *E. coli* strain DS941 was transformed with the reporter plasmid pAB106 (see Table 2-2) carrying two transposon ends (containing approximately 100 bp of the transposon end and 100 bp of the transposon flank) flanking the *galK* gene. Upon expression of XerT from a helper plasmid XerT_pBAD/MCS (see Table 2-2) under control of the arabinose promoter, the two transposon ends on the reporter vector can get recombined resulting in excision of a non-replicative circular transposon junction carrying the *galK* gene. After excision, the remaining vector backbone will contain the transposon flank sequences with the centrally located heptanucleotide 5'AAGAATG, a product that

corresponds to an empty site. If excision and subsequent loss of the mini-transposon carrying *galK* occurs, the cells can be detected on MacConkey agar by change of colony colour from red to white.

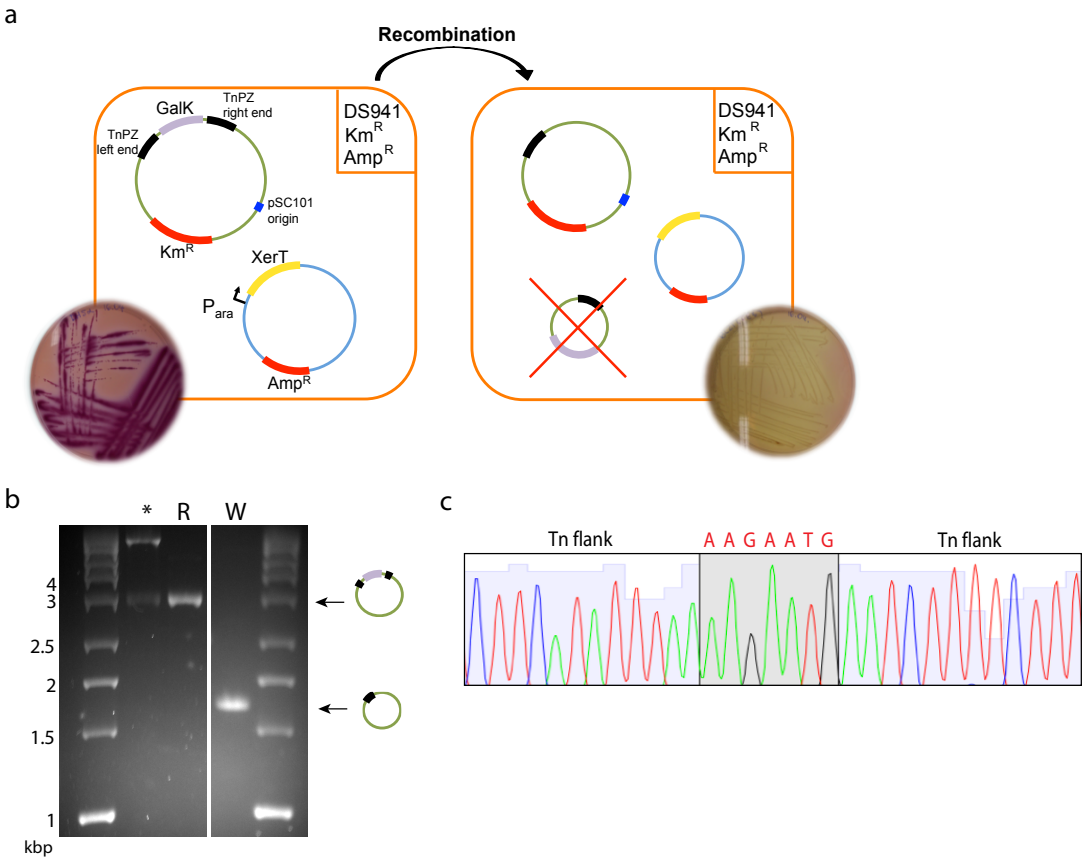


Figure 6-20: *galK*-MacConkey TnPZ mini-transposon excision assay in *E. coli*. a) The design of the recombination assay. An *E. coli galK*-deficient strain is transformed with the reporter plasmid containing the two TnPZ transposon ends flanking a functional *galK* gene and with the XerT expression plasmid. Upon XerT expression, the two transposon ends are recombined and a non-replicative mini-transposon junction containing the *galK* gene is excised and lost from the cell population. Cells carrying the *galK* gene (no XerT recombination) and those without the *galK* gene (post-transposon excision) can be distinguished by plating on MacConkey agar plates, where the two kinds of cells grow as red or white colonies, respectively. b) Analysis of plasmid DNA extracted from the red (R) and white (W) colonies after the TnPZ excision assay. The supercoiled plasmid DNA was analysed on a 0.8% agarose gel. The star denotes the reporter plasmid control. c) Sequencing results of the post-excision empty site on the reporter plasmid. The nucleotides in the chromatogram are coloured as follows: adenine – green; guanine – black; thymidine – red; and cytosine – blue.

The excision assay was performed in DS941 cells as described in section 2.6.2. Briefly, after initial transformation of the expression and reporter plasmids, the cells were plated on MacConkey agar plates supplemented with 0.02% arabinose. The DNA from the obtained colonies was retransformed into DS941 cells and plated on MacConkey agar plates supplemented with galactose, where both white (after excision) and red (no excision)

colonies could be seen. The plasmid DNA from both white and red colonies thus obtained was analysed on an agarose gel (Figure 6-20b). The analysis of the reporter plasmid from the red colonies showed a band corresponding in size to the supercoiled non-recombined reporter plasmid (5204 bp), while the reporter plasmid from the white colonies was shorter (3289 bp) and corresponded to the recombined reporter plasmid missing the *galK* gene. The excision of the TnPZ mini-transposon was confirmed by sequencing of the reporter plasmid from the white colonies, which revealed that the transposon was cleanly excised, leaving behind the heptanucleotide sequence and the transposon-flank sequences as expected (Figure 6-20c). Since the TnPZ mini-transposon used in this experiment did not contain any of the ORFs carried on the full-length TnPZ, these results show that XerT is the only protein encoded on TnPZ that is required for its excision in *E. coli.* Furthermore, these results show that no host-specific *H. pylori* factors are required for TnPZ excision since excision could be reconstituted in the *E. coli* host. No white colonies could be detected in the absence of the XerT expression plasmid.

6.3 Discussion

6.3.1 Recombinant XerT binds and cleaves TnPZ ends

H. pylori XerT was successfully overexpressed in *E. coli* and purified to homogeneity (Figures 6-6 and 6-9). The purified protein was shown to bind to substrates containing TnPZ ends and part of their flanking sequences in a DNaseI footprinting assay (Figure 6-11) and in a SEC assay on an analytical gel filtration column (Figure 6-13). Furthermore, XerT was shown to cleave DNA at both TnPZ ends in an SDS-PAGE cleavage assay (Figures 6-14 and 6-16). The cleavage likely represented the true intermediate of XerT recombination pathway as the cleaved products were proficient in a strand exchange reaction (Figure 6-15). Together, these results show that XerT is likely acting as the transposase of TnPZ since cleavage of the two transposon ends will lead to excision of the circular transposon junction, a TnPZ transposition intermediate previously detected *in vivo* (Fisher *et al.,* 2010) and corresponding to the first step in the conjugative transposition pathway (see Figures 1-1c and 6-3). Interestingly, based on the cleavage and strand exchange assay, XerT seems to follow the canonical biochemistry of tyrosine recombination, with cleavage dependent on the presence of the conserved catalytic residues, Tyr332 and Arg300 (Figure 6-14d), and with 3'-phosphotyrosyl intermediate

formed upon cleavage. The length of the identified XerT binding site (31-34 bp; Figure 6-11) and the observed sizes of XerT-TnPZ complexes (Figure 3-12) suggest that two monomers of XerT bind to each transposon end. However, due to lack of success in performing an EMSA experiment, it could not be unambiguously determined if XerT binding at TnPZ ends follows the canonical pattern of ordered, cooperative binding of the two recombinase monomers as seen for XerH (see Figure 3-12). Moreover, the binding and cleavage experiments investigated only the initial steps of XerT recombination and it is possible that the further steps of strand exchange, HJ formation (if HJ is an intermediate of XerT recombination at all), and HJ resolution follow a different mechanism from that of canonical tyrosine recombinases. To elucidate these steps, further studies of the recombination pathway are required.

6.3.2 Positions of XerT-mediated cleavage at TnPZ ends

The positions of XerT-mediated cleavage at TnPZ ends were identified in two independent experiments: the SDS-PAGE cleavage assay (Figure 6-16) and the half-site cleavage and strand exchange assay (Figure 6-17). The cleavage sites on the LE appeared to be shifted by one nucleotide compared to the cleavage sites on the RE, if the positions of the conserved heptanucleotide 5'AAGAATG motifs are used as reference points (Figure 6-16d). However, it has to be noted that the experimental results were not fully conclusive for the LE, and the XerT cleavage position at its left arm had to be partially derived from similarities between the two TnPZ ends (for example, considering that the size of the central region is likely the same at both ends). Therefore, it remains possible that imprecise cleavage positions were derived for LE due to low cleavage efficiency, and further assays should be carried out for final confirmation of the XerT cleavage sites.

The two identified cleavage sites (one on the top strand and one on the bottom strand) of each TnPZ end flank a central region of 6 bp (Figure 6-16d). The size of the central region is consistent with that observed for all Xer recombinases investigated so far, including XerH. However, the central region does not lie within the invariable region of the heptanucleotide 5'AAGAATG as previously proposed (Kersulyte *et al*., 2009). Instead, only the most 5' AA dinucleotide of this heptanucleotide is located within the central region, while the remaining five nucleotides lie within the right arm of the recombination site (Figure 6-16d). This surprising result suggests that the heptanucleotide acts partially as

a binding site for one XerT monomer, which could explain why the presence of this sequence is absolutely conserved in all identified TnPZ elements (Kersulyte *et al.,* 2009).

The identified 6-bp central region between the two binding arms differs only in the identity of one basepair between the two ends of TnPZ from *H. pylori* P12, and the identity of this basepair does not affect cleavage efficiency (Figure 6-18b). However, it is not clear whether this difference between the two ends is important for further steps of the reaction, for example for the strand exchange step. It is an established dogma of tyrosine recombination that the central regions of two DNA recombination partners must share the same sequence (reviewed in Rajeev *et al.,* 2009). However, conjugative transposons employing tyrosine recombinases fundamentally disobey this rule, as one of the central regions corresponds to the ever-changing transposon flank and therefore is only occasionally homologous to the transposon-borne central region (Caparon and Scott, 1989). The XerT/TnPZ recombination system appears to be somewhere in between since a single-nucleotide mismatch in the central region is tolerated (in fact, other TnPZ types show up to two mismatched base-pairs; Figure 6-2), yet some specificity in the central region seems to be required, as most of the sequence is conserved between the two ends.

6.3.3 Ordered binding and cleavage of the TnPZ ends by XerT

The observed efficiencies of XerT-mediated cleavage on the two DNA strands of each TnPZ end differed greatly (RE in Figure 6-18b and LE in 6-19b), with the bottom strand of the pre-nicked substrates cleaved much more efficiently than the top strand in both cases. Similar results were observed in the SDS-PAGE cleavage assay monitoring XerH-mediated cleavage of dif_H site (see Figure 3-14c). This assay, together with the results of the EMSA and SEC assays with various dif_H-derived substrates (see Figures 3-12, 3-18, and 3-19), allowed the conclusion to be drawn that the XerH-mediated binding and cleavage events at dif_H are ordered, with the left arm of dif_H bound first and the right arm cleaved first. It is possible that similar ordering of the binding and cleavage events occurs at TnPZ ends. In support of this notion is the aforementioned difference in cleavage efficiencies between the two strands (arms) at each TnPZ end. By drawing an analogy to the XerH mechanism, this would mean that the left arm of each TnPZ end is bound by the first XerT monomer, the right arm is bound cooperatively by the second monomer, and the right arm (bottom strand) is cleaved first. This hypothesis is supported by the mutational

analysis of the RE: the left arm palindrome substrate (consisting of two left arms in inverted repeat orientation) shows higher XerT cleavage activity, likely resulting from higher binding efficiency of the left arm compared to the right arm (Figure 6-18b, lane 3). Similarly, the mutation of the left arm that changes basepairs AT-TA to a single G-C basepair, the latter natively present on the right arm, abolishes cleavage at both arms (Figure 6-18b, lane 6), suggesting that the AT-TA basepairs might be essential for the good initial binding of the left arm by XerT. However, this hypothesis cannot easily explain the binding and cleavage events at the LE. At this TnPZ end, the left arm consists of the donor flank DNA that differs as the transposon is inserted at different positions in the genome, and accordingly can be mutated to random sequences without loss of activity (Figure 6-19b, lanes 4-6), suggesting that in the case of the LE the binding and cleavage events are ordered differently. Furthermore, complete conclusions about the DNA binding and cleavage order cannot be derived based on the SDS-PAGE cleavage assay alone: a binding assay is also required. To visualize binding events on various TnPZ-derived substrates upon titration with XerT, and EMSA assay is ideally suited, but these experiments were so far unsuccessful and should be further optimized.

6.3.4 XerT binding site requirements

Mutational analysis of the TnPZ ends provided the first insights into the sequence requirements for efficient XerT-mediated cleavage, and presumably binding, of the recombination sites. Since the heptanucleotide 5'AAGAATG is invariably conserved amongst different TnPZ elements and it does not lie entirely within the central region of the XerT recombination site, it is likely that instead it constitutes an essential part of the XerT binding site.

The RE of TnPZ shows a high level of symmetry (Figure 6-12) and upon closer inspection it becomes apparent that five nucleotides 5'GAATG of the conserved heptanucleotide present in the right arm are repeated in an inverted orientation on the left arm. This implies that each of these arms contains a necessary binding site for one XerT monomer. The predicted different binding efficiency at the two arms could therefore depend on the outer arm sequences, and (as discussed in the previous section) it seems that a TA dinucleotide present only in the left arm contributes to the more efficient binding of this arm. Similar observations were made for the XerH/dif_H system, where a single basepair insertion in the

left arm of dif_H contributes to more efficient binding of this arm of dif_H (discussed in section 3.3.4). It seems that no specific recognition of the outer sequences at the right arm of RE is required, since this region can be mutated without loss of cleavage activity (Figure 6-18b, lanes 7-10). This is particularly interesting considering that many tyrosine recombinases form site-specific interactions with their DNA substrates in the outer sequences. For example, the differences between the outer sequences contribute to differential binding of XerC and XerD to *dif* arms. On the other hand, XerH makes only one such contact in the outer sequences of the binding site. Contacting sequence-specifically only the five innermost basepairs of each recombination half-site could be particularly beneficial for the transposase XerT, as it could provide a mechanism that assures flexibility in transposon target site selection (discussed later).

The binding of XerT to the LE sequences remains puzzling as only one full 5'GAATG binding site can be seen at this TnPZ end and, if the results of cleavage mapping were correctly interpreted, this binding site is shifted one nucleotide away from the central region (Figure 6-16d). Corresponding positions on the left arm (belonging to the transposon flank) do not show any sequence conservation and can be in fact mutated to sequences with variable G+C content without a strong effect on XerT cleavage efficiency (Figure 6-19b, lanes 4-6). How the two monomers of a site-specific recombinase can efficiently bind two sites of which only one contains specific binding sequences remains unclear. It is possible that DNA-free XerT forms an unstable dimer, which due to its instability appears as a monomer during SEC runs. This idea is supported by observation of dimers during purification of the XerT Y332F variant (Figure 6-10b). The dimer could then bind the LE using only one conserved binding site on the right arm. Alternatively, binding of individual XerT monomers to the LE with very strong cooperativity could also explain the observed lack of one binding site: here, the first XerT monomer could bind to the right arm containing the conserved binding site, while the second monomer would bind any sequence through cooperative binding, mainly via protein-protein interactions. Both notions are supported by the fact that no cleavage products could be observed for the left arm of LE in the half-site cleavage assay, where the left arm was detached from the right arm of LE (Figure 6-17). However, both these ideas require additional investigation, for example by a functional EMSA experiment.

In summary, the investigation of the XerT binding site requirements shows that the XerT recombination sites at the two TnPZ ends altogether contain only three conserved XerT

binding sites (GAATG), yet are expected to be bound by four XerT monomers. The studies of XerT binding to TnPZ have to be further complemented by cleavage assays with more mutated TnPZ substrates (for example ones with the conserved binding site mutated) and by analysis of binding events, for example by an EMSA approach.

6.3.5 XerT is the only TnPZ-encoded gene required for TnPZ excision

The *galK*-MacConkey recombination assay was used as a tool to establish what is required for TnPZ excision. The experiment showed that two DNA fragments of 200 bp each containing DNA sequences corresponding to the TnPZ end and the adjacent flanking DNA are sufficient for TnPZ excision in the presence of ectopically expressed XerT (Figure 6-20). This mini-transposon (LE-*galK*-RE) did not contain any ORFs present on TnPZ, which indicates that XerT is not only proficient in TnPZ excision *in vivo*, but is in fact the only protein encoded on TnPZ that is required for excision of this element. XerT might even be sufficient for TnPZ excision altogether, since no other factors are required for cleavage of the transposon ends *in vitro*; however, this suggestion would need to be confirmed by *in vitro* reconstitution of the full recombination reaction. Additionally, the observed XerT-mediated recombination in *E. coli* suggests that TnPZ excision does not require any *H. pylori*-specific host factors, in agreement with the fact that conjugative transposons can usually transfer between unrelated species (Bertram *et al.,* 1991). However, from this assay it is not clear if any TnPZ-encoded or *H. pylori*-specific factors are required for the later steps of conjugative transposition including conjugal transfer (for which all essential genes seem to be encoded by TnPZ itself) and integration into the recipient genome. For comparison, Tn916 requires an additional, small protein called Xis for efficient excision, while this protein is dispensable for the integration step, suggesting that these two reactions might show different requirements (Marra and Scott, 1999). Therefore, further integration assays addressing protein and DNA sequence requirements are necessary. Finally, it is worth remembering that XerT is a member of the Xer family of tyrosine recombinases, and most of the members of this family require an additional factor (usually FtsK) for activation and spatial and temporal regulation. Whether such regulation also takes place in XerT recombination remains to be elucidated.

6.3.6 Target specificity of TnPZ

All TnPZ elements identified to date are flanked by two direct repeats of the heptanucleotide 5'AAGAATG, and one copy of this sequence is present in the corresponding empty sites (Figure 6-1; Kersulyte *et al.,* 2009). This suggests that TnPZ elements are preferentially inserted at the 5'AAGAATG sequence, and it seems that no other sequences are tolerated. Such strict requirement for the presence of this heptanucleotide can be explained by the results presented in this study, which suggest that five nucleotides 5'GAATG of the conserved heptanucleotide act as a necessary binding site for XerT. Accordingly, if a recombination site (including an integration site on target DNA) is to be bound and recombined by XerT, this sequence must be present. Of the remaining two nucleotides (AA) that form part of the central region, mutation of the first adenine slightly reduces cleavage by XerT, indicating the potential importance of this residue in XerT recombination. Such requirement for specific sequences within the central region has been previously shown for the non-canonical conjugative transposon CTnDOT, where the two basepairs of the central region close to the left IntDOT binding site are strictly conserved (Cheng *et al.,* 2000).

The seven-basepair target specificity of TnPZ is unusual in the world of transposons. Most DNA transposons show no or very limited target specificity, for example transposon Tn3 prefers AT-rich sequences but shows no sequence specificity (Tu and Cohen, 1980), Tn5 prefers G-C basepairs at the insertion site (Lodge *et al.,* 1988), while the preferred Mu insertion site is a CGG triplet (Manna *et al.,* 2005). The conjugative transposons are mostly similar: for example, conjugative transposons of the Tn916 family require AT-rich sequences for insertion but no sequence consensus could be identified at the insertion sites. Unusual in its target specificity is CTnDOT, which inserts preferentially into target sites with the consensus sequence GTTnnTTGC (Cheng *et al.,* 2000). Considering that most known conjugative transposons employ site-specific recombinases for their mobilization, it is interesting to see how these different recombinases diverged from the paradigm of strict site-specificity in their mechanism of integration.

While it seems that XerT requires specific sequences for integration of TnPZ, it is worth noting that active integration of a TnPZ element has never been directly observed, and the target requirements for TnPZ insertion should be further confirmed in *in vitro* and *in vivo* experiments.

6.3.7 Model of TnPZ excision and integration

TnPZ was proposed to be a conjugative transposon mobilized by a tyrosine recombinase XerT (Kersulyte *et al.,* 2009). Based on my obtained results investigating the mechanism of XerT recombinase and the role of TnPZ sequences in XerT-TnPZ recombination, a first model of TnPZ conjugative transposition can be proposed (Figure 6-21). As shown by the DNaseI footprinting and cleavage assays (Figures 6-11 and 6-16), XerT recombination sites consist of two XerT binding arms separated by a six-nucleotide central region. Each recombination site includes a half-site that is a part of the transposon and a half-site that is part of the donor flank (shown in green and blue, respectively, in Figure 6-21). In accordance with the results of the cleavage assays, five conserved nucleotides 5'GAATG constitute an essential binding site for XerT (shown in red in Figure 6-21). The RE contains two such conserved binding sites spaced symmetrically on each side of the central region (Figure 6-12), while LE contains only one 5'GAATG sequence within the transposon half-site.

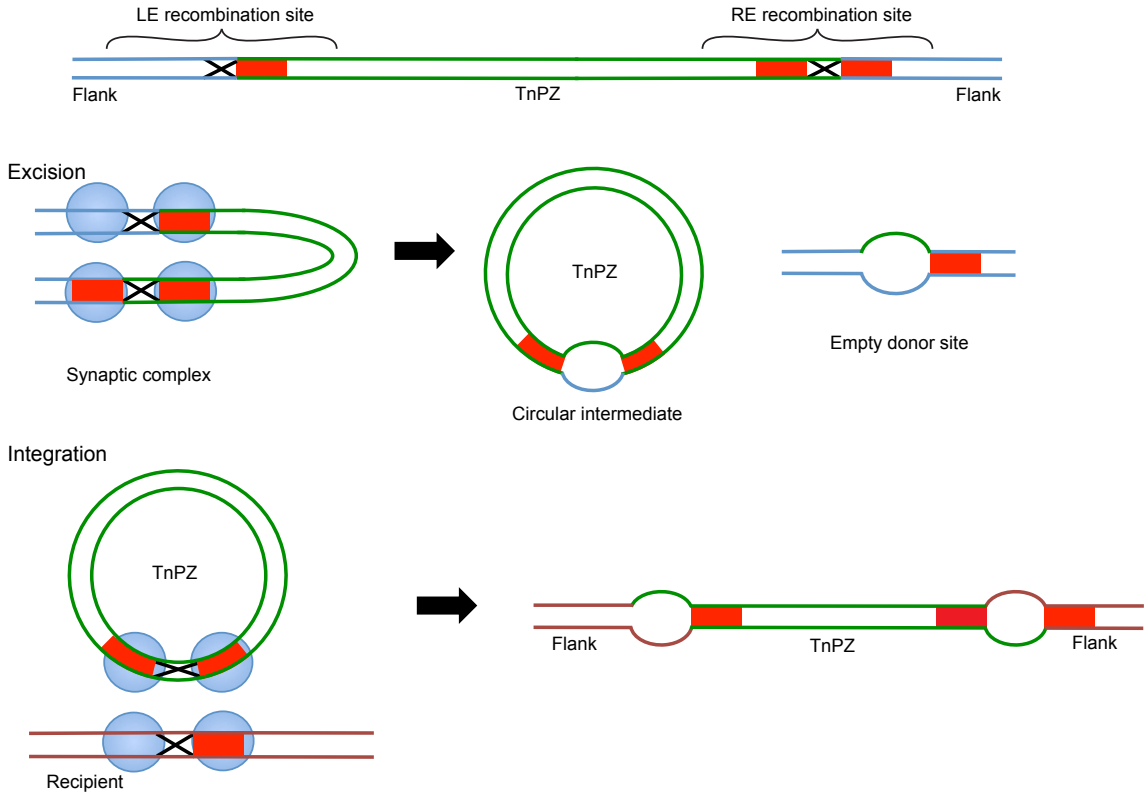


Figure 6-21: Proposed model of TnPZ excision and integration. The transposon DNA is schematically shown in green, the flanking donor sequences are shown in blue, and the flanking recipient (target) sequences are shown in brown. The conserved XerT binding sequence 5'GAATG is represented by a red box. The black cross represents the central region of the recombination site. Mismatches between the two DNA strands are shown as bubbles. XerT monomers are shown as blue spheres.

In the first excision step of transposition, the two TnPZ ends are brought together in an anti-parallel arrangement to form the XerT-TnPZ synaptic complex where the two DNA sites can be recombined. It seems that only three conserved XerT binding sites are required for assembly of the expected tetrameric synaptic complex. Upon excision, the circular TnPZ transposon is released from the donor site, and the donor flanks are resealed. The excised circular TnPZ contains two transposon-borne conserved binding sites flanking the central region that now contains a 1-2 bp mismatch. The sealed donor site contains the single conserved binding site derived from the donor flank near the right transposon end. The excision of TnPZ also leaves a 1-2 bp mismatch in the donor. All mismatches are presumably repaired by the host replication machinery. Following TnPZ excision, one strand of the circular transposon is likely transferred into the recipient cell through conjugation, followed by the replication of the second TnPZ strand. However, these steps of the pathway are not clear and were not addressed in this study. In the final integration step of TnPZ transposition, three conserved binding sites must be brought together to enable formation of a productive recombination complex. This means that the target site within the recipient genome must contain the conserved five nucleotides required for XerT binding, resulting in the observed target specificity of the transposon. The additional two nucleotides of the conserved heptanucleotide that lie within the central region are also required for XerT recombination (although their exact function is unclear) resulting in the full TnPZ target sequence 5'AAGAATG. Since in the so far identified TnPZ sequences the 6-bp central region differs only by one or two basepairs between the two transposon ends (Figure 6-2; Kersulyte *et al.,* 2009), it is likely that these sequences might also contribute to the choice of TnPZ target site. Assembly of the synaptic complex including the circular transposon and the target site leads to recombination, resulting in the insertion of TnPZ into the recipient genome. The inserted transposon will then be flanked by the central regions containing 1-2 bp mismatches that are resolved during genome replication by the host machinery.

Altogether, this proposed model of XerT excision and integration agrees with previously published observations and with the results of this study. At the same time, this model has to be further validated in *in vitro* and *in vivo* experiments since it challenges the dogmas of tyrosine recombination: the necessity of site-specific recognition of the recombining DNA sites by each recombinase subunit in the synaptic complex and the requirement for full sequence homology between the central regions of the recombining sites.

7. General discussion and conclusions

7.1 The study of XerH/*dif*$_H$ recombination and its implications

The work presented in this thesis investigated various aspects of *H. pylori* XerH recombination at *dif*$_H$ sites. XerH resolves chromosome dimers that arise frequently during genome replication, and was also proposed to play a role in decatenation of the *H. pylori* chromosome. Consistently, XerH is required for successful chromosome segregation in this bacterium (Debowski *et al.,* 2012). Interestingly, while Xer recombination systems in many bacteria require the concerted action of two Xer recombinases, *H. pylori* (along with other *Helicobacter*, *Streptococcus*, and *Lactococcus* species) employs a single Xer recombinase to perform chromosome dimer resolution and ensure faithful chromosome segregation, but the mechanism of single-recombinase Xer systems is only partially understood. Therefore, the essential function and the unusual mechanism of XerH recombination in *H. pylori* make XerH an interesting research target that could enable the design of novel antibacterial agents against this deadly pathogen.

To understand the specific mechanism of XerH/*dif*$_H$ recombination, this study focused on addressing questions concerning the peculiar features of Xer systems that employ a single recombinase protein, especially how such systems can be regulated in terms of binding and cleavage order, and what is required for the recombination reaction to be completed. In detail, I wanted to find out what are the exact requirements for a functional *dif*$_H$ recombination site, how XerH interacts with *dif*$_H$, and what is the organization of the XerH-*dif*$_H$ assembly, as well as what is the role of FtsK in XerH recombination and what other factors might be required for this process. Some of these questions apply also more generally to other Xer systems, where the lack of structural information on the protein-DNA recombination complex assembly has so far prevented full characterization of the recombination process.

By combination of biochemical, structural, and microbiology assays, this study showed that the XerH/*dif*$_H$ recombination system follows a mechanism that includes several features of canonical tyrosine recombination, such as sequential binding of four recombinase monomers to two recombination sites forming a catalytically active synaptic

complex, DNA cleavage with formation of the 3'-phosphotyrosyl intermediate, and allosteric regulation of protein activity via a C-terminal protein segment resulting in the half-of-the-sites reactivity (Figures 3-12 and 3-14). However, the analysis also revealed several mechanistic features unique to XerH recombination, for example a different angle of DNA substrate bending (Figure 4-27), specific interactions between XerH and dif_H arms (Figure 4-28), and protein-DNA contacts stabilizing the synaptic complex assembly (Figure 4-29). Moreover, the results showed how single recombinase-mediated binding and cleavage events could be orchestrated by the asymmetry of the recombination substrate (discussed in section 3.3.4). Finally, the study of XerH regulation by host factors revealed that XerH is regulated differently from *E. coli* XerC/D or streptococcal XerS, with FtsK not being required for direct activation of XerH recombination (Figure 5-9), while still having an impact on chromosome dimer resolution (Figures 5-12 and 5-13), perhaps together with additional, so far unknown *H. pylori* regulatory factors.

By combining all the results obtained in this study, a first model of XerH recombination can be presented (Figure 7-1). According to this model, initially two XerH monomers bind to one dif_H site in a sequential, cooperative manner, with the first monomer binding to the left arm (which contains the additional basepair absent from the right arm). Two bound dif_H sites are brought together into a tetrameric synaptic complex (Figure 7-1a). Within this complex, the two XerH subunits bound to the right dif_H arms perform concerted cleavage of bottom strands, resulting in a formation of a covalent intermediate containing 3'-phosphotyrosyl linkage (Figure 7-1b). A subsequent strand exchange step likely leads to formation of the HJ intermediate (Figure 7-1c). At this point, the HJ intermediate can be converted back to substrates (as seen in the HJ resolution assay, Figure 3-17), or the junction can isomerize so that the second pair of XerH monomers become active and can cleave the DNA, leading to formation of recombined products (which could not be seen in this study and is shown in grey in Figure 7-1). It is likely that a so far unidentified host factor (shown as a red sphere in Figure 7-1) is required to bring about a complex conformation that would allow HJ isomerization or, alternatively, to change the complex conformation before the first strand cleavage and exchange; however, it appears that the previously implicated FtsK is not that factor.

The crystal structure of the XerH-dif_H-LP synaptic complex (Figure 4-20) provides also general implications for Xer recombination. It confirms that Xer recombinases employ an equivalent synaptic complex assembly and the same chemical reaction as other tyrosine

recombinases. Based on the structure, a model of domain rearrangements occurring in tyrosine recombinases upon DNA-binding could also be proposed (Figure 4-23). The structure is likely to be useful for future modelling of other Xer recombinases and their recombination assemblies.

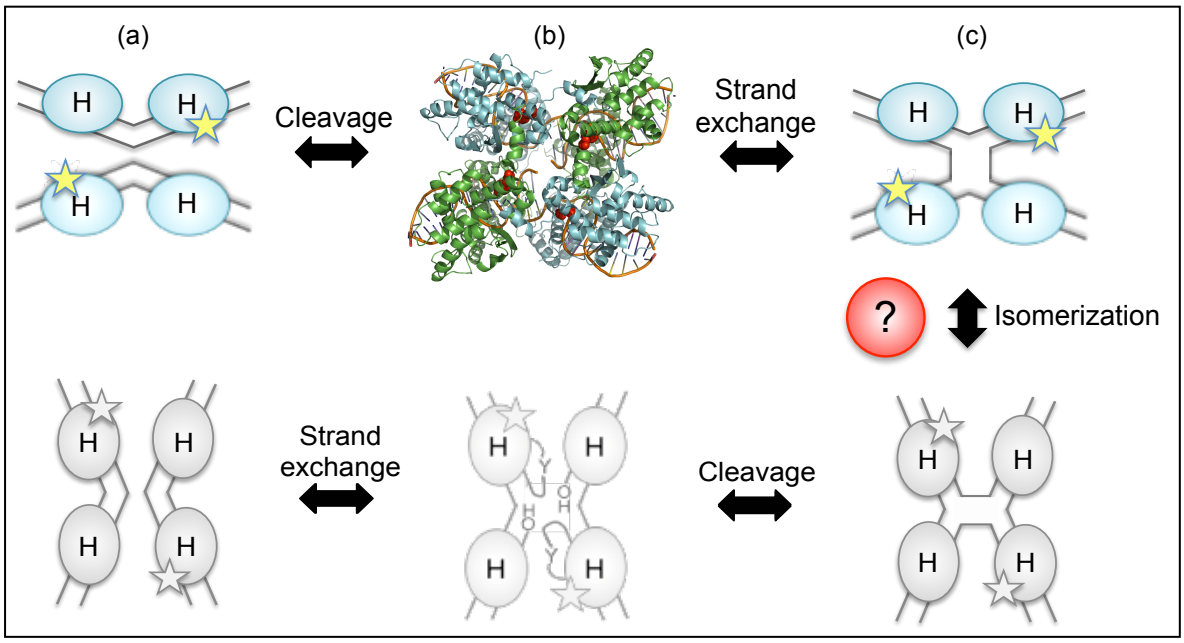


Figure 7-1: The proposed model of XerH recombination. Four monomers of XerH (light blue spheres) bind two dif_H recombination sites (black) in a specific order (left arm first, then right arm) to form a functional synaptic complex (a). Two non-adjoining XerH subunits bound to the right arms of dif_H (marked with stars) cleave the bottom strand of the DNA, resulting in formation of a 3'-phosphotyrosyl intermediate (b), here represented by the obtained XerH-dif_H-LP crystal structure. The strand exchange step likely results in formation of a HJ intermediate (c). HJ isomerization is likely regulated by a so far unspecified factor (shown as red sphere). The further steps of the recombination pathway leading to resolution of the HJ intermediate could not be reconstituted in this study and are shown in grey.

In summary, this study provided insights into various aspects of XerH recombination, contributed to the better general understanding of Xer recombination through presentation of the first structure of an Xer-DNA synaptic complex, and to the better understanding of the single-recombinase Xer systems by showing how the binding and cleavage events of recombination can be orchestrated in the presence of only one recombinase.

7.2 The study of XerT/TnPZ recombination and its implications

This PhD project also addressed the role of the second Xer recombinase found in some *H. pylori* strains, XerT, in mobilization of a putative conjugative transposon TnPZ.

Conjugative transposition has been linked to the lateral transfer of antibiotic resistance and virulence genes in bacteria, in some genera playing a major role in the spread of antibiotic resistance and creation of multiple-drug resistant strains (Clewell, 1986). Despite its medical importance, the mechanism of conjugative transposition is poorly understood due to the lack of structural and biochemical characterization. Therefore, the study of XerT presented an opportunity to improve the understanding of conjugative transposition.

Using biochemical and microbiology methods, the study confirmed the role of XerT as the transposase acting to mobilize TnPZ, by showing that XerT can bind and cleave TnPZ ends *in vitro* (Figures 6-11, 6-12, 6-13, and 6-14) and mobilize a TnPZ mini-transposon in *E. coli* (Figure 6-20). The results suggested that no TnPZ-encoded proteins other than XerT are required for mobilization of TnPZ. Moreover, a range of *in vitro* assays allowed for determination of the exact positions of XerT-mediated binding and cleavage at TnPZ ends (Figures 6-11, 6-16, and 6-17). Furthermore, biochemical characterization of XerT-mediated cleavage events at TnPZ ends revealed a number of important features of the recombination sites and showed that XerT has a limited site-specificity, which increases the chances of finding a suitable integration target site in the host genome for the excised circular transposon (Figures 6-18 and 6-19). The biochemical characterization presented in this study could prove useful for designing symmetrized substrates that could facilitate formation of stable XerT-TnPZ complexes suitable for crystallization trials, which in turn could allow structural characterization of an XerT-TnPZ synaptic complex. Finally, based on the performed experiments and previously published studies of TnPZ, a first testable model of XerT-dependent TnPZ excision and insertion was proposed (Figure 6-21). The model states that three conserved binding sites are required for binding of four XerT monomers in the synaptic complex (with the fourth monomer bound to non-specific site) during excision and integration. It also shows how the conserved sites are maintained by selection of specific integration sites, and shows that the resulting integration products agree with the previously published work.

The obtained results and the proposed model suggest that XerT recombination at TnPZ ends challenges the paradigms of tyrosine recombination, which state that the central regions of the two recombination sites must be homologous and that the four recombinases forming the synaptic complex must recognize four binding sites containing recombinase-specific sequences. In turn, XerT recombination at the TnPZ ends features only limited homology between the central regions of the two recombination sites, and the sites

together comprise only three conserved XerT binding sites. At the same time, the XerT/TnPZ recombination system shows higher similarity to canonical tyrosine recombination rather than other conjugative transposons (for example, those of the Tn916 family) because it still requires seven specific basepairs at the target integration site and has a central region with partial homology between the recombination sites that seems to be conserved in various TnPZ elements (Figure 6.2). The XerT/TnPZ system could therefore provide an example of an intermediate state between canonical tyrosine recombination and homology- and sequence-independent conjugative transposition.

To summarize, the study of XerT recombination provides first insights into the unique mobilization mechanism of TnPZ, a genetic island carrying almost 50% of *H. pylori* strain-specific genes and likely contributing to the fitness of the bacterium. At the same time, this work improves our understanding of conjugative transposition, a poorly understood phenomenon of great medical relevance.

7.3 *H. pylori* Xer recombinases: parallel between Xer recombination and conjugative transposition

The genomes of many *H. pylori* strains carry two tyrosine recombinases of the Xer family that carry out two distinct functions in this bacterium. XerH, encoded in all *H. pylori* strains, resolves chromosome dimers that arise during genome replication (Debowski *et al*., 2012), while XerT, found in some strains only, mobilizes a conjugative transposon that carries a majority of the *H. pylori* strain-specific genes (Fischer *et al*., 2010). Phylogenetic analysis of XerT in the context of other Xer recombinases showed that the XerT proteins are most closely related to the XerH clade (Figure 6-4), suggesting that the proteins share a common predecessor, or that one of these proteins has evolved from the other. Since Xer recombinases are present in most bacteria, it is likely that the mobile genetic element (TnPZ) hijacked the chromosome-borne recombinase to promote its own mobilization. On the other hand, it is also possible that the recombinase from an ancient mobile genetic element has been acquired and utilized by the bacterial host for genome maintenance. Further phylogenetic analysis is required to assess the evolutionary relationship between XerH and XerT of *H. pylori*.

To date, apart from *H. pylori*, no other bacterium carrying an essential canonical tyrosine recombinase as well as a closely related protein acting as a conjugative transposase has

been identified. Consequently, the Xer recombinases of *H. pylori* present a great opportunity for studying such a set of proteins (a recombinase and a transposase) in parallel, offering both technical and fundamental advantages particularly for the study of the conjugative transposition mechanism. First, from a technical point of view the canonical tyrosine recombinases are well characterized and many XerH assays in this study could be easily established based on previously published work on other tyrosine recombinases. The assays could then be applied for the study of XerT, using the XerH results both as a starting point and a control, allowing unusual findings concerning XerT to be easily distinguished as technical artefacts or true deviations from the classical recombination mechanism. Second, from a mechanistic point of view the direct comparison of these two closely related proteins provides exceptional insights into their mechanisms and the principles of their distinct functions.

On one hand, the results presented in this study reveal shared features of XerH recombination and XerT transposition. For example, both proteins appear to form dimers on their respective recombination sites, and both use a catalytic tyrosine residue for DNA cleavage resulting in formation of a 3'-phosphotyrosyl intermediate and a 5' DNA overhang. The recombination sites of both proteins show bias towards more efficient binding and cleavage of one arm by the recombinase, and in both cases this bias seems to be brought about by one or two critical basepairs present in the outer part of one binding arm. On the other hand, despite their presumed evolutionary relationship, fundamental differences can be seen between the recombination mechanisms of XerH and XerT. Most strikingly, the sequences of the recombination sites of the two systems (the dif_H site and the TnPZ ends) differ to such an extent that each protein can only interact with its own recombination site and no cross-reactivity between the proteins and the sites has been observed (Debowski *et al.,* 2012; Fischer *et al.,* 2010). Furthermore, XerT recombination seems to require only three conserved and sequence-specific binding sites within the tetrameric recombination complex, while XerH recombination likely requires four sites to be present (as judged by the fact that substrates with only one dif_H arm bind only one XerH monomer). Finally, the central region sequences differ between the two TnPZ ends (as derived from the cleavage site mapping experiments), suggesting that there is no central region homology between the two recombination sites in the synaptic complex, while XerH recombines two identical dif_H sites with identical central regions.

In summary, the presented parallel between the two recombination systems allows for direct comparison of the obtained results, enabling deeper characterization of the mechanism and requirements associated with the function of each recombinase. At the same time, such a parallel helps the more general understanding of how two proteins evolve to perform two distinct functions while preserving a similar fold and fundamental chemical reaction mechanism.

7.4 Future directions

The presented study provides a solid basis and further resources (such as the XerH-*dif*$_H$ structure and the biochemical and microbiology assays) for further studies of both XerH-*dif*$_H$ and XerT-TnPZ recombination systems. Moreover, the results shown in this work introduce new questions that can be further addressed with combination of structural and molecular biology approaches.

The crystal structure of the XerH-*dif*$_H$-LP synaptic complex presented in this study showed that a post-cleavage 3'-phosphotyrosyl intermediate is a signature of XerH recombination, similarly to other tyrosine recombinases (Figures 4-20 and 4-26). Accordingly, it would be interesting to know whether other expected intermediates of the generally established mechanism of tyrosine recombination (such as the HJ intermediate) are also shared features of XerH recombination. This could be achieved by entrapment and crystallization of XerH-*dif*$_H$ complexes at other steps of the recombination pathway. Such subsequent intermediates could be easily designed based on crystallographic studies of Cre, λ integrase, and Flp. Since it was possible to obtain a crystal structure of the XerH-*dif*$_H$ post-cleavage synaptic complex, it is likely that further complexes would also crystallize, especially if symmetrized DNA substrates (such as those used in this study) were used. In addition to confirming the conservation of further recombination intermediates, such crystal structures could additionally shed light on the so-far elusive process of HJ isomerization.

Based on the results gathered to date, it is apparent that the asymmetry of the recombination sites and the synaptic complex assembly plays an important role in regulating the process of XerH recombination. Accordingly, it would be invaluable to solve a crystal structure of XerH in complex with the native, non-symmetrized *dif*$_H$ site.

This could be achieved by constructing suicide substrates that do not allow annealing of a palindromic substrate but strictly form only one type of DNA substrate, the native dif_H site. Such a structure could help to explain the molecular details of how a single basepair insertion in one of the dif_H arms can determine the order of XerH recombination events.

The study of the regulation of XerH/dif_H recombination shed new light on the role of FtsK and implicated other *H. pylori* host factors in this process, presenting many new questions and hypotheses that need to be further addressed. First, the specific role of *H. pylori* FtsK in XerH recombination should be investigated in further assays in *E. coli* and in *H. pylori*, since so far the results obtained in both organisms are not consistent. For example, an excision assay with the dif_H-Km-dif_H cassette inserted ectopically into the chromosomes of wild-type and $FtsK_C$-deficient *E. coli* strains could answer whether the positioning of the dif_H-Km-dif_H cassette in the *ter* region was detrimental to XerH recombination in this study, and whether dif_H recombination on other parts of the chromosome would depend on the presence of $FtsK_C$. Similarly, excision of such a cassette placed at *ter* or ectopically could be monitored in the presence of a helper plasmid expressing various *H. pylori* FtsK constructs to test the species-specificity of potential XerH-FtsK interactions. Second, the possibility that other *H. pylori* factors might be required for XerH recombination could be tested by screening through the list of potential candidates compiled in this study. The role of the candidates in XerH recombination could be verified by mutational analysis in *E. coli* and in *H. pylori*, for example by assessing XerH plasmid recombination efficiency in *E. coli* strains lacking the *E. coli* homologues of the potential regulatory factors, or by phenotypic screening in *H. pylori*.

The study of the TnPZ conjugative transposon and its recombinase XerT has so far only begun to address the mechanism of TnPZ mobilization. The proposed first model should be validated and further refined in *in vitro* and *in vivo* experiments, for example by detecting the predicted DNA heteroduplex regions that were proposed to form upon TnPZ excision, or by detailed characterization of XerT binding events at TnPZ ends *in vitro*. Furthermore, the mutational analysis presented in this study also requires additional TnPZ-end variants to be tested in order to derive the full picture of XerT binding site requirements. To facilitate such studies, a functional EMSA should also be established. Furthermore, the efforts towards preparation of stable XerT-TnPZ end complexes should be continued with the perspective of crystallizing the XerT-TnPZ synaptic complex for detailed structural analysis. Finally, this study has not addressed the potential regulation or

activation of XerT by host factors, an aspect of TnPZ mobilization that would also be interesting to address.

7.5 Conclusions

This thesis describes the results obtained during my PhD work on two Xer recombinases from *H. pylori*. The first part of the project focused on the XerH recombinase, a canonical Xer recombinase from *H. pylori*. XerH resolves chromosome dimers and is therefore necessary for *H. pylori* growth, similarly to its homologues across the bacterial and archaeal kingdoms. Therefore, XerH presents an interesting research objective, which in the future could contribute to the development of novel antibacterial agents. In this study, the crystal structure of the XerH recombinase from *H. pylori* in a post-cleavage synaptic complex with its DNA substrate, dif_H-LP, was determined. To our knowledge, this is the first structure of any Xer recombinase in complex with DNA, and thus provides novel insights not only into the mechanism of XerH recombination, but also of Xer recombination in general. Additionally, the role of dif_H site asymmetry in ordering of the XerH-mediated binding and cleavage events was addressed and first insights into XerH regulation by *H. pylori* host factors were gained. The combined structural, biochemical, and microbiology approaches allowed for a broad characterization of the recombination events at the dif_H site.

The second part of my PhD project focused on the TnPZ conjugative transposon from the same species. The mechanism of conjugative transposition is poorly understood due to the lack of structural and biochemical data, despite the fact that this process contributes to the spread of antibiotic resistance amongst bacteria and therefore poses a great medical threat. The study of TnPZ began to characterize the mechanism of mobilization of this conjugative transposon. It was shown that XerT supports excision of a TnPZ mini-transposon in *E. coli* in the absence of other TnPZ-encoded genes, suggesting that no TnPZ- or *H. pylori*-specific factors are required for TnPZ excision from the donor genome. Additionally, XerT binding and cleavage at TnPZ ends have been characterized *in vitro* and the first model of TnPZ excision and integration was proposed. These results provide important insights into the mechanism of conjugative transposition and may have implications for fighting the spread of antibiotic resistance.

References

ABRAHAMS, J. P. & LESLIE, A. G. 1996. Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr D Biol Crystallogr,* 52**,** 30-42.

ABREMSKI, K. & HOESS, R. 1984. Bacteriophage P1 site-specific recombination. Purification and properties of the Cre recombinase protein. *J Biol Chem,* 259**,** 1509-14.

ABREMSKI, K., HOESS, R. & STERNBERG, N. 1983. Studies on the properties of P1 site-specific recombination: evidence for topologically unlinked products following recombination. *Cell,* 32**,** 1301-11.

ADAMS, P. D., AFONINE, P. V., BUNKOCZI, G., CHEN, V. B., DAVIS, I. W., ECHOLS, N., HEADD, J. J., HUNG, L. W., KAPRAL, G. J., GROSSE-KUNSTLEVE, R. W., MCCOY, A. J., MORIARTY, N. W., OEFFNER, R., READ, R. J., RICHARDSON, D. C., RICHARDSON, J. S., TERWILLIGER, T. C. & ZWART, P. H. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr,* 66**,** 213-21.

AIHARA, H., KWON, H. J., NUNES-DUBY, S. E., LANDY, A. & ELLENBERGER, T. 2003. A conformational switch controls the DNA cleavage activity of lambda integrase. *Mol Cell,* 12**,** 187-98.

AKOPYANTS, N. S., EATON, K. A. & BERG, D. E. 1995. Adaptive mutation and cocolonization during *Helicobacter pylori* infection of gnotobiotic piglets. *Infect Immun,* 63**,** 116-21.

ALM, R. A., LING, L. S., MOIR, D. T., KING, B. L., BROWN, E. D., DOIG, P. C., SMITH, D. R., NOONAN, B., GUILD, B. C., DEJONGE, B. L., CARMEL, G., TUMMINO, P. J., CARUSO, A., URIA-NICKELSEN, M., MILLS, D. M., IVES, C., GIBSON, R., MERBERG, D., MILLS, S. D., JIANG, Q., TAYLOR, D. E., VOVIS, G. F. & TRUST, T. J. 1999. Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature,* 397**,** 176-80.

ALTSCHUL, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic local alignment search tool. *J Mol Biol,* 215**,** 403-10.

ARNOLD, P. H., BLAKE, D. G., GRINDLEY, N. D., BOOCOCK, M. R. & STARK, W. M. 1999. Mutants of Tn3 resolvase which do not require accessory binding sites for recombination activity. *Embo j,* 18**,** 1407-14.

AUSSEL, L., BARRE, F. X., AROYO, M., STASIAK, A., STASIAK, A. Z. & SHERRATT, D. 2002. FtsK Is a DNA motor protein that activates chromosome dimer resolution by switching the catalytic state of the XerC and XerD recombinases. *Cell,* 108**,** 195-205.

BALDWIN, E. P., MARTIN, S. S., ABEL, J., GELATO, K. A., KIM, H., SCHULTZ, P. G. & SANTORO, S. W. 2003. A specificity switch in selected Cre recombinase variants is mediated by macromolecular plasticity and water. *Chem Biol,* 10**,** 1085-94.

BALTRUS, D. A., AMIEVA, M. R., COVACCI, A., LOWE, T. M., MERRELL, D. S., OTTEMANN, K. M., STEIN, M., SALAMA, N. R. & GUILLEMIN, K. 2009. The complete genome sequence of *Helicobacter pylori* strain G27. *J Bacteriol,* 191**,** 447-8.

BARRE, F. X., AROYO, M., COLLOMS, S. D., HELFRICH, A., CORNET, F. & SHERRATT, D. J. 2000. FtsK functions in the processing of a Holliday junction intermediate during bacterial chromosome segregation. *Genes Dev,* 14**,** 2976-88.

BERTRAM, J., STRATZ, M. & DURRE, P. 1991. Natural transfer of conjugative transposon Tn916 between gram-positive and gram-negative bacteria. *J Bacteriol,* 173**,** 443-8.

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**,** 1059-66.

BLAKELY, G., COLLOMS, S., MAY, G., BURKE, M. & SHERRATT, D. 1991. *Escherichia coli* XerC recombinase is required for chromosomal segregation at cell division. *New Biol,* 3**,** 789-98.

BLAKELY, G., MAY, G., MCCULLOCH, R., ARCISZEWSKA, L. K., BURKE, M., LOVETT, S. T. & SHERRATT, D. J. 1993. Two related recombinases are required for site-specific recombination at *dif* and *cer* in *E. coli* K12. *Cell,* 75**,** 351-61.

BLAKELY, G. W., DAVIDSON, A. O. & SHERRATT, D. J. 1997. Binding and cleavage of nicked substrates by site-specific recombinases XerC and XerD. *J Mol Biol,* 265**,** 30-9.

BRICOGNE, G., VONRHEIN, C., FLENSBURG, C., SCHILTZ, M. & PACIOREK, W. 2003. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr D Biol Crystallogr,* 59**,** 2023-30.

BROWN, L. M. 2000. *Helicobacter pylori*: epidemiology and routes of transmission. *Epidemiol Rev,* 22**,** 283-97.

BUCHAN, D. W., MINNECI, F., NUGENT, T. C., BRYSON, K. & JONES, D. T. 2013. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res,* 41**,** W349-57.

BURGIN, A. B., JR. & NASH, H. A. 1995. Suicide substrates reveal properties of the homology-dependent steps during integrative recombination of bacteriophage lambda. *Curr Biol,* 5**,** 1312-21.

BURRUS, V., PAVLOVIC, G., DECARIS, B. & GUEDON, G. 2002. Conjugative transposons: the tip of the iceberg. *Mol Microbiol,* 46**,** 601-10.

BUTT, T. R., EDAVETTAL, S. C., HALL, J. P. & MATTERN, M. R. 2005. SUMO fusion technology for difficult-to-express proteins. *Protein Expr Purif,* 43**,** 1-9.

CAPARON, M. G. & SCOTT, J. R. 1989. Excision and insertion of the conjugative transposon Tn916 involves a novel recombination mechanism. *Cell,* 59**,** 1027-34.

CAPIAUX, H., LESTERLIN, C., PERALS, K., LOUARN, J. M. & CORNET, F. 2002. A dual role for the FtsK protein in *Escherichia coli* chromosome segregation. *EMBO Rep,* 3**,** 532-6.

CARNOY, C. & ROTEN, C. A. 2009. The *dif*/Xer recombination systems in proteobacteria. *PLoS One,* 4**,** e6531.

CHALKER, A. F., LUPAS, A., INGRAHAM, K., SO, C. Y., LUNSFORD, R. D., LI, T., BRYANT, A., HOLMES, D. J., MARRA, A., PEARSON, S. C., RAY, J., BURNHAM, M. K., PALMER, L. M., BISWAS, S. & ZALACAIN, M. 2000. Genetic characterization of gram-positive homologs of the XerCD site-specific recombinases. *J Mol Microbiol Biotechnol,* 2**,** 225-33.

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. *Mol Cell,* 6**,** 885-97.

CHEN, Y. & RICE, P. A. 2003. The role of the conserved Trp330 in Flp-mediated recombination. Functional and structural analysis. *J Biol Chem,* 278**,** 24800-7.

CHENG, C., KUSSIE, P., PAVLETICH, N. & SHUMAN, S. 1998. Conservation of structure and mechanism between eukaryotic topoisomerase I and site-specific recombinases. *Cell,* 92**,** 841-50.

CHENG, Q., PASZKIET, B. J., SHOEMAKER, N. B., GARDNER, J. F. & SALYERS, A. A. 2000. Integration and excision of a Bacteroides conjugative transposon, CTnDOT. *J Bacteriol,* 182**,** 4035-43.

CHRISTIE, P. J., ATMAKURI, K., KRISHNAMOORTHY, V., JAKUBOWSKI, S. & CASCALES, E. 2005. Biogenesis, architecture, and function of bacterial type IV secretion systems. *Annu Rev Microbiol,* 59**,** 451-85.

COLLOMS, S. D., BATH, J. & SHERRATT, D. J. 1997. Topological selectivity in Xer site-specific recombination. *Cell,* 88**,** 855-64.

COLLOMS, S. D., SYKORA, P., SZATMARI, G. & SHERRATT, D. J. 1990. Recombination at ColE1 *cer* requires the *Escherichia coli* xerC gene product, a member of the lambda integrase family of site-specific recombinases. *J Bacteriol,* 172**,** 6973-80.

CONWAY, A. B., CHEN, Y. & RICE, P. A. 2003. Structural plasticity of the Flp-Holliday junction complex. *J Mol Biol,* 326**,** 425-34.

CORNET, F., MORTIER, I., PATTE, J. & LOUARN, J. M. 1994. Plasmid pSC101 harbors a recombination site, *psi*, which is able to resolve plasmid multimers and to substitute for the analogous chromosomal *Escherichia coli* site *dif*. *J Bacteriol,* 176**,** 3188-95.

CORTEZ, D., QUEVILLON-CHERUEL, S., GRIBALDO, S., DESNOUES, N., SEZONOV, G., FORTERRE, P. & SERRE, M. C. 2010. Evidence for a Xer/dif system for chromosome resolution in Archaea. *PLoS Genet,* 6**,** e1001166.

CRICK, F. 1958. On Protein Synthesis. *The Symposia of the Society for Experimental Biology,* 12**,** 138-163.

CURCIO, M. J. & DERBYSHIRE, K. M. 2003. The outs and ins of transposition: from mu to kangaroo. *Nat Rev Mol Cell Biol,* 4**,** 865-77.

DAS, B., MARTINEZ, E., MIDONET, C. & BARRE, F. X. 2013. Integrative mobile elements exploiting Xer recombination. *Trends Microbiol,* 21**,** 23-30.

DATSENKO, K. A. & WANNER, B. L. 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A,* 97**,** 6640-5.

DAVIS, A. J. & CHEN, D. J. 2013. DNA double strand break repair via non-homologous end-joining. *Transl Cancer Res,* 2**,** 130-143.

DEBOWSKI, A. W., CARNOY, C., VERBRUGGHE, P., NILSSON, H. O., GAUNTLETT, J. C., FULURIJA, A., CAMILLERI, T., BERG, D. E., MARSHALL, B. J. & BENGHEZAL, M. 2012. Xer recombinase and genome integrity in *Helicobacter pylori*, a pathogen without topoisomerase IV. *PLoS One,* 7**,** e33310.

DEBOWSKI, A. W., GAUNTLETT, J. C., LI, H., LIAO, T., SEHNAL, M., NILSSON, H. O., MARSHALL, B. J. & BENGHEZAL, M. 2012a. Xer-cise in *Helicobacter pylori*: one-step transformation for the construction of markerless gene deletions. *Helicobacter,* 17**,** 435-43.

DE MAIO, A. 1999. Heat shock proteins: facts, thoughts, and dreams. *Shock,* 11**,** 1-12.

DEMARRE, G., GUEROUT, A. M., MATSUMOTO-MASHIMO, C., ROWE-MAGNUS, D. A., MARLIERE, P. & MAZEL, D. 2005. A new family of mobilizable suicide plasmids based on broad host range R388 plasmid (IncW) and RP4 plasmid (IncPalpha) conjugative machineries and their cognate *Escherichia coli* host strains. *Res Microbiol,* 156**,** 245-55.

DIEDERICHS, K. & KARPLUS, P. A. 2013. Better models by discarding data? *Acta Crystallogr D Biol Crystallogr,* 69**,** 1215-22.

DIXON, J. E., SHAIKH, A. C. & SADOWSKI, P. D. 1995. The Flp recombinase cleaves Holliday junctions in trans. *Mol Microbiol,* 18**,** 449-58.

EIDT, S. & STOLTE, M. 1995. The significance of *Helicobacter pylori* in relation to gastric cancer and lymphoma. *Eur J Gastroenterol Hepatol,* 7**,** 318-21.

EMSLEY, P., LOHKAMP, B., SCOTT, W. G. & COWTAN, K. 2010. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr,* 66**,** 486-501.

ENNIFAR, E., MEYER, J. E., BUCHHOLZ, F., STEWART, A. F. & SUCK, D. 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 Res,* 31**,** 5449-60.

ESLICK, G. D., LIM, L. L., BYLES, J. E., XIA, H. H. & TALLEY, N. J. 1999. Association of *Helicobacter pylori* infection with gastric carcinoma: a meta-analysis. *Am J Gastroenterol,* 94**,** 2373-9.

ESPOSITO, D. & SCOCCA, J. J. 1997. The integrase family of tyrosine recombinases: evolution of a conserved active site domain. *Nucleic Acids Res,* 25**,** 3605-14.

FERREIRA, H., BUTLER-COLE, B., BURGIN, A., BAKER, R., SHERRATT, D. J. & ARCISZEWSKA, L. K. 2003. Functional analysis of the C-terminal domains of the site-specific recombinases XerC and XerD. *J Mol Biol,* 330**,** 15-27.

FISCHER, W., WINDHAGER, L., ROHRER, S., ZEILLER, M., KARNHOLZ, A., HOFFMANN, R., ZIMMER, R. & HAAS, R. 2010. Strain-specific genes of *Helicobacter pylori:* genome evolution driven by a novel type IV secretion system and genomic island transfer. *Nucleic Acids Res,* 38**,** 6089-101.

FISHER, R. A. 1930. *The Genetical Theory of Natural Selection,* Oxford, Clarendon Press.

FLANNAGAN, S. E., ZITZOW, L. A., SU, Y. A. & CLEWELL, D. B. 1994. Nucleotide sequence of the 18-kb conjugative transposon Tn916 from *Enterococcus faecalis*. *Plasmid,* 32**,** 350-4.

FOURIER, J. B. J. 1822. *Théorie Analytique de la Chaleur,* Paris, Chez Firmin Didot, père et fils.

FRANZ, B. & LANDY, A. 1995. The Holliday junction intermediates of lambda integrative and excisive recombination respond differently to the bending proteins integration host factor and excisionase. *Embo j,* 14**,** 397-406.

GARNIER, F., TAOURIT, S., GLASER, P., COURVALIN, P. & GALIMAND, M. 2000. Characterization of transposon Tn1549, conferring VanB-type resistance in Enterococcus spp. *Microbiology,* 146 (Pt 6)**,** 1481-9.

GHOSH, K., GUO, F. & VAN DUYNE, G. D. 2007. Synapsis of *loxP* sites by Cre recombinase. *J Biol Chem,* 282**,** 24004-16.

GHOSH, K., LAU, C. K., GUO, F., SEGALL, A. M. & VAN DUYNE, G. D. 2005. Peptide trapping of the Holliday junction intermediate in Cre-*loxP* site-specific recombination. *J Biol Chem,* 280**,** 8290-9.

GHOSH, K., LAU, C. K., GUPTA, K. & VAN DUYNE, G. D. 2005a. Preferential synapsis of *loxP* sites drives ordered strand exchange in Cre-*loxP* site-specific recombination. *Nat Chem Biol,* 1**,** 275-82.

GIBB, B., GUPTA, K., GHOSH, K., SHARP, R., CHEN, J. & VAN DUYNE, G. D. 2010. Requirements for catalysis in the Cre recombinase active site. *Nucleic Acids Res,* 38**,** 5817-32.

GOPAUL, D. N. & VAN -DUYNE, G. D. 1999. Structure and mechanism in site-specific recombination. *Curr Opin Struct Biol,* 9**,** 14-20.

GOPAUL, D. N., GUO, F. & VAN DUYNE, G. D. 1998. Structure of the Holliday junction intermediate in Cre-*loxP* site-specific recombination. *Embo j,* 17**,** 4175-87.

GRAHAM, J. E., SIVANATHAN, V., SHERRATT, D. J. & ARCISZEWSKA, L. K. 2010. FtsK translocation on DNA stops at XerCD-*dif*. *Nucleic Acids Res,* 38**,** 72-81.

GRAINGE, I., LESTERLIN, C. & SHERRATT, D. J. 2011. Activation of XerCD-dif recombination by the FtsK DNA translocase. *Nucleic Acids Res,* 39**,** 5140-8.

GRINDLEY, N. D., WHITESON, K. L. & RICE, P. A. 2006. Mechanisms of site-specific recombination. *Annu Rev Biochem,* 75**,** 567-605.

GROVE, J. I., ALANDIYJANY, M. N. & DELAHAY, R. M. 2013. Site-specific relaxase activity of a VirD2-like protein encoded within the tfs4 genomic island of *Helicobacter pylori. J Biol Chem,* 288**,** 26385-96.

GUO, F., GOPAUL, D. N. & VAN DUYNE, G. D. 1997. Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature,* 389**,** 40-6.

GUO, F., GOPAUL, D. N. & VAN DUYNE, G. D. 1999. Asymmetric DNA bending in the Cre-*loxP* site-specific recombination synapse. *Proc Natl Acad Sci U S A,* 96**,** 7143-8.

HALLET, B., ARCISZEWSKA, L. K. & SHERRATT, D. J. 1999. Reciprocal control of catalysis by the tyrosine recombinases XerC and XerD: an enzymatic switch in site-specific recombination. *Mol Cell,* 4**,** 949-59.

HANDFORD, J. I., IZE, B., BUCHANAN, G., BUTLAND, G. P., GREENBLATT, J., EMILI, A. & PALMER, T. 2009. Conserved network of proteins essential for bacterial viability. *J Bacteriol,* 191**,** 4732-49.

HENDRICKS, E. C., SZERLONG, H., HILL, T. & KUEMPEL, P. 2000. Cell division, guillotining of dimer chromosomes and SOS induction in resolution mutants (*dif,* xerC and xerD) of *Escherichia coli*. *Mol Microbiol,* 36**,** 973-81.

HENDRICKSON, H. & LAWRENCE, J. G. 2006. Selection for chromosome architecture in bacteria. *J Mol Evol,* 62**,** 615-29.

HENGEN, P. 1995. Purification of His-Tag fusion proteins from *Escherichia coli*. *Trends Biochem Sci,* 20**,** 285-6.

HICKMAN, A. B., WANINGER, S., SCOCCA, J. J. & DYDA, F. 1997. Molecular organization in site-specific recombination: the catalytic domain of bacteriophage HP1 integrase at 2.7 A resolution. *Cell,* 89**,** 227-37.

HOESS, R. H., ZIESE, M. & STERNBERG, N. 1982. P1 site-specific recombination: nucleotide sequence of the recombining sites. *Proc Natl Acad Sci U S A,* 79**,** 3398-402.

HOLLIDAY, R. 1964. A mechanism for gene conversion in fungi. *Genetics Research,* 89**,** 285-304.

IYER, L. M., MAKAROVA, K. S., KOONIN, E. V. & ARAVIND, L. 2004. Comparative genomics of the FtsK-HerA superfamily of pumping ATPases: implications for the origins of chromosome segregation, cell division and viral capsid packaging. *Nucleic Acids Res,* 32**,** 5260-79.

KABSCH, W. 2010. XDS. *Acta Crystallogr D Biol Crystallogr,* 66**,** 125-32.

KAIMER, C., SCHENK, K. & GRAUMANN, P. L. 2011. Two DNA translocases synergistically affect chromosome dimer resolution in *Bacillus subtilis*. *J Bacteriol,* 193**,** 1334-40.

KARPLUS, P. A. & DIEDERICHS, K. 2012. Linking crystallographic model and data quality. *Science,* 336**,** 1030-3.

KELLEY, L. A. & STERNBERG, M. J. 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc,* 4**,** 363-71.

KERSULYTE, D., LEE, W., SUBRAMANIAM, D., ANANT, S., HERRERA, P., CABRERA, L., BALQUI, J., BARABAS, O., KALIA, A., GILMAN, R. H. & BERG, D. E. 2009. *Helicobacter Pylori's* plasticity zones are novel transposable elements. *PLoS One,* 4**,** e6859.

KERSULYTE, D., VELAPATINO, B., MUKHOPADHYAY, A. K., CAHUAYME, L., BUSSALLEU, A., COMBE, J., GILMAN, R. H. & BERG, D. E. 2003. Cluster of type IV secretion genes in *Helicobacter pylori*'s plasticity zone. *J Bacteriol,* 185**,** 3764-72.

KOBAYASHI, G., MORIYA, S. & WADA, C. 2001. Deficiency of essential GTP-binding protein ObgE in *Escherichia coli* inhibits chromosome partition. *Mol Microbiol,* 41**,** 1037-51.

KRISSINEL, E. & HENRICK, K. 2004. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr,* 60**,** 2256-68.

KROGH, B. O. & SHUMAN, S. 2000. Catalytic mechanism of DNA topoisomerase IB. *Mol Cell,* 5**,** 1035-41.

KROGH, B. O. & SHUMAN, S. 2002. Proton relay mechanism of general acid catalysis by DNA topoisomerase IB. *J Biol Chem,* 277**,** 5711-4.

KUEMPEL, P. L., HENSON, J. M., DIRCKS, L., TECKLENBURG, M. & LIM, D. F. 1991. *dif,* a *recA*-independent recombination site in the terminus region of the chromosome of *Escherichia coli*. *New Biol,* 3**,** 799-811.

KWON, H. J., TIRUMALAI, R., LANDY, A. & ELLENBERGER, T. 1997. Flexibility in DNA recombination: structure of the lambda integrase catalytic core. *Science,* 276**,** 126-31.

LANDY, A. 1989. Dynamic, structural, and regulatory aspects of lambda site-specific recombination. *Annu Rev Biochem,* 58**,** 913-49.

LARKIN, M. A., BLACKSHIELDS, G., BROWN, N. P., CHENNA, R., MCGETTIGAN, P. A., MCWILLIAM, H., VALENTIN, F., WALLACE, I. M., WILM, A., LOPEZ, R., THOMPSON, J. D., GIBSON, T. J. & HIGGINS, D. G. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics,* 23**,** 2947-8.

LAVALLIE, E. R., LU, Z., DIBLASIO-SMITH, E. A., COLLINS-RACIE, L. A. & MCCOY, J. M. 2000. Thioredoxin as a fusion partner for production of soluble recombinant proteins in *Escherichia coli. Methods Enzymol,* 326**,** 322-40.

LE BOURGEOIS, P., BUGAREL, M., CAMPO, N., DAVERAN-MINGOT, M. L., LABONTE, J., LANFRANCHI, D., LAUTIER, T., PAGES, C. & RITZENTHALER, P. 2007. The unconventional Xer recombination machinery of Streptococci/Lactococci. *PLoS Genet,* 3**,** e117.

LEROUX, M., REZOUG, Z. & SZATMARI, G. 2013. The Xer/*dif* site-specific recombination system of *Campylobacter jejuni*. *Mol Genet Genomics,* 288**,** 495-502.

LESLIE, N. R. & SHERRATT, D. J. 1995. Site-specific recombination in the replication terminus region of *Escherichia coli*: functional replacement of *dif*. *Embo j,* 14**,** 1561-70.

LESTERLIN, C., BARRE, F. X. & CORNET, F. 2004. Genetic recombination and the cell cycle: what we have learned from chromosome dimers. *Mol Microbiol,* 54**,** 1151-60.

LEVY, O., PTACIN, J. L., PEASE, P. J., GORE, J., EISEN, M. B., BUSTAMANTE, C. & COZZARELLI, N. R. 2005. Identification of oligonucleotide sequences that direct the movement of the *Escherichia coli* FtsK translocase. *Proc Natl Acad Sci U S A,* 102**,** 17618-23.

LI, X. & HEYER, W. D. 2008. Homologous recombination in DNA repair and DNA damage tolerance. *Cell Res,* 18**,** 99-113.

LODGE, J. K., WESTON-HAFER, K. & BERG, D. E. 1988. Transposon Tn5 target specificity: preference for insertion at G/C pairs. *Genetics,* 120**,** 645-50.

LU, F. & CHURCHWARD, G. 1994. Conjugative transposition: Tn916 integrase contains two independent DNA binding domains that recognize different DNA sequences. *Embo j,* 13**,** 1541-8.

LUSCOMBE, N. M., LASKOWSKI, R. A. & THORNTON, J. M. 1997. NUCPLOT: a program to generate schematic diagrams of protein-nucleic acid interactions. *Nucleic Acids Res,* 25**,** 4940-5.

MANNA, D., DENG, S., BREIER, A. M. & HIGGINS, N. P. 2005. Bacteriophage Mu targets the trinucleotide sequence CGG. *J Bacteriol,* 187**,** 3586-8.

MARRA, D. & SCOTT, J. R. 1999. Regulation of excision of the conjugative transposon Tn916. *Mol Microbiol,* 31**,** 609-21.

MARSHALL, B. J. & WARREN, J. R. 1984. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *Lancet,* 1**,** 1311-5.

MARTIN, S. S., PULIDO, E., CHU, V. C., LECHNER, T. S. & BALDWIN, E. P. 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. *J Mol Biol,* 319**,** 107-27.

MATTHEWS, B. W. 1968. Solvent content of protein crystals. *J Mol Biol,* 33**,** 491-7.

MATYSIAK-BUDNIK, T. & MEGRAUD, F. 1997. Epidemiology of *Helicobacter pylori* infection with special reference to professional risk. *J Physiol Pharmacol,* 48 Suppl 4**,** 3-17.

MCCOY, A. J., GROSSE-KUNSTLEVE, R. W., ADAMS, P. D., WINN, M. D., STORONI, L. C. & READ, R. J. 2007. Phaser crystallographic software. *J Appl Crystallogr,* 40**,** 658-674.

NAGARAJAN, R., KWON, K., NAWROT, B., STEC, W. J. & STIVERS, J. T. 2005. Catalytic phosphoryl interactions of topoisomerase IB. *Biochemistry,* 44**,** 11476-85.

NAGY, A. 2000. Cre recombinase: the universal reagent for genome tailoring. *Genesis,* 26**,** 99-109.

NAKAMURA, Y., GOJOBORI, T. & IKEMURA, T. 2000. Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res,* 28**,** 292.

NOLIVOS, S., PAGES, C., ROUSSEAU, P., LE BOURGEOIS, P. & CORNET, F. 2010. Are two better than one? Analysis of an FtsK/Xer recombination system that uses a single recombinase. *Nucleic Acids Res,* 38**,** 6477-89.

NORGREN, M. & SCOTT, J. R. 1991. The presence of conjugative transposon Tn916 in the recipient strain does not impede transfer of a second copy of the element. *J Bacteriol,* 173**,** 319-24.

NUNES-DUBY, S. E., AZARO, M. A. & LANDY, A. 1995. Swapping DNA strands and sensing homology without branch migration in lambda site-specific recombination. *Curr Biol,* 5**,** 139-48.

NUNES-DUBY, 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 Res,* 26**,** 391-406.

OH, J. D., KLING-BACKHED, H., GIANNAKIS, M., XU, J., FULTON, R. S., FULTON, L. A., CORDUM, H. S., WANG, C., ELLIOTT, G., EDWARDS, J., MARDIS, E. R., ENGSTRAND, L. G. & GORDON, J. I. 2006. The complete genome sequence of a chronic atrophic gastritis *Helicobacter pylori* strain: evolution during disease progression. *Proc Natl Acad Sci U S A,* 103**,** 9999-10004.

PAGE, R., GRZECHNIK, S. K., CANAVES, J. M., SPRAGGON, G., KREUSCH, A., KUHN, P., STEVENS, R. C. & LESLEY, S. A. 2003. Shotgun crystallization strategy for structural genomics: an optimized two-tiered crystallization screen against the *Thermotoga maritima* proteome. *Acta Crystallogr D Biol Crystallogr,* 59**,** 1028-37.

PARGELLIS, C. A., NUNES-DUBY, S. E., DE VARGAS, L. M. & LANDY, A. 1988. Suicide recombination substrates yield covalent lambda integrase-DNA complexes and lead to identification of the active site tyrosine. *J Biol Chem,* 263**,** 7678-85.

PATTERSON, A. L. 1935. A direct method for the determination of the components of interatomic distances in crystals. *Z. Kristallogr.,* 90**,** 517-542.

PERALS, K., CORNET, F., MERLET, Y., DELON, I. & LOUARN, J. M. 2000. Functional polarization of the *Escherichia coli* chromosome terminus: the *dif* site acts in chromosome dimer resolution only when located between long stretches of opposite polarity. *Mol Microbiol,* 36**,** 33-43.

RADMAN-LIVAJA, M., SHAW, C., AZARO, M., BISWAS, T., ELLENBERGER, T. & LANDY, A. 2003. Arm sequences contribute to the architecture and catalytic function of a lambda integrase-Holliday junction complex. *Mol Cell,* 11**,** 783-94.

RAJEEV, L., MALANOWSKA, K. & GARDNER, J. F. 2009. Challenging a paradigm: the role of DNA homology in tyrosine recombinase reactions. *Microbiol Mol Biol Rev,* 73**,** 300-9.

RECCHIA, G. D., AROYO, M., WOLF, D., BLAKELY, G. & SHERRATT, D. J. 1999. FtsK-dependent and -independent pathways of Xer site-specific recombination. *Embo j,* 18**,** 5724-34.

RECCHIA, G. D. & SHERRATT, D. J. 1999. Conservation of xer site-specific recombination genes in bacteria. *Mol Microbiol,* 34**,** 1146-8.

REED, R. R. & GRINDLEY, N. D. 1981. Transposon-mediated site-specific recombination *in vitro*: DNA cleavage and protein-DNA linkage at the recombination site. *Cell,* 25**,** 721-8.

REN, J., WEN, L., GAO, X., JIN, C., XUE, Y. & YAO, X. 2009. DOG 1.0: illustrator of protein domain structures. *Cell Res,* 19**,** 271-3.

RHODES, G. 2006. Chapter 2 - An Overview of Protein Crystallography. *In:* RHODES, G. (ed.) *Crystallography Made Crystal Clear (Third Edition)*. Burlington: Academic Press.

RINGROSE, L., LOUNNAS, V., EHRLICH, L., BUCHHOLZ, F., WADE, R. & STEWART, A. F. 1998. Comparative kinetic analysis of FLP and cre recombinases: mathematical models for DNA binding and recombination. *J Mol Biol,* 284**,** 363-84.

RYDER, L., SHARPLES, G. J. & LLOYD, R. G. 1996. Recombination-dependent growth in exonuclease-depleted recBC sbcBC strains of *Escherichia coli* K-12. *Genetics,* 143**,** 1101-14.

SAITOU, N. & NEI, M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol,* 4**,** 406-25.

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. *Microbiol Rev,* 59**,** 579-90.

SCHINDELIN, J., ARGANDA-CARRERAS, I., FRISE, E., KAYNIG, V., LONGAIR, M., PIETZSCH, T., PREIBISCH, S., RUEDEN, C., SAALFELD, S., SCHMID, B., TINEVEZ, J. Y., WHITE, D. J., HARTENSTEIN, V., ELICEIRI, K., TOMANCAK, P. & CARDONA, A. 2012. Fiji: an open-source platform for biological-image analysis. *Nat Methods,* 9**,** 676-82.

SCHNEIDER, T. R. & SHELDRICK, G. M. 2002. Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr,* 58**,** 1772-9.

SCHWEDE, T., KOPP, J., GUEX, N. & PEITSCH, M. C. 2003. SWISS-MODEL: An automated protein homology-modeling server. *Nucleic Acids Res,* 31**,** 3381-5.

SCIOCHETTI, S. A., PIGGOT, P. J. & BLAKELY, G. W. 2001. Identification and characterization of the *dif* site from *Bacillus subtilis*. *J Bacteriol,* 183**,** 1058-68.

SCOTT, J. R., BRINGEL, F., MARRA, D., VAN ALSTINE, G. & RUDY, C. K. 1994. Conjugative transposition of Tn916: preferred targets and evidence for conjugative transfer of a single strand and for a double-stranded circular intermediate. *Mol Microbiol,* 11**,** 1099-108.

SCOTT, J. R. & CHURCHWARD, G. G. 1995. Conjugative transposition. *Annu Rev Microbiol,* 49**,** 367-97.

SEGAL, D. J. & MECKLER, J. F. (2013). Genome engineering at the dawn of the golden age. *Annu Rev Genomics Hum Genet,* 14, 135-158.

SEIGNEUR, M., BIDNENKO, V., EHRLICH, S. D. & MICHEL, B. 1998. RuvAB acts at arrested replication forks. *Cell,* 95**,** 419-30.

SENECOFF, J. F. & COX, M. M. 1986. Directionality in FLP protein-promoted site-specific recombination is mediated by DNA-DNA pairing. *J Biol Chem,* 261**,** 7380-6.

SERRE, M. C., EL ARNAOUT, T., BROOKS, M. A., DURAND, D., LISBOA, J., LAZAR, N., RAYNAL, B., VAN TILBEURGH, H. & QUEVILLON-CHERUEL, S. 2013. The carboxy-terminal alphaN helix of the archaeal XerA tyrosine recombinase is a molecular switch to control site-specific recombination. *PLoS One,* 8**,** e63010.

SHATZKY-SCHWARTZ, M., ARBUCKLE, N. D., EISENSTEIN, M., RABINOVICH, D., BAREKET-SAMISH, A., HARAN, T. E., LUISI, B. F. & SHAKKED, Z. 1997. X-ray and solution studies of DNA oligomers and implications for the structural basis of A-tract-dependent curvature. *J Mol Biol,* 267**,** 595-623.

SIVANATHAN, V., EMERSON, J. E., PAGES, C., CORNET, F., SHERRATT, D. J. & ARCISZEWSKA, L. K. 2009. KOPS-guided DNA translocation by FtsK safeguards *Escherichia coli* chromosome segregation. *Mol Microbiol,* 71**,** 1031-42.

SMITH, C. J., TRIBBLE, G. D. & BAYLEY, D. P. 1998. Genetic elements of Bacteroides species: a moving story. *Plasmid,* 40**,** 12-29.

SMITH, M. C. & THORPE, H. M. 2002. Diversity in the serine recombinases. *Mol Microbiol,* 44**,** 299-307.

SOSKINE, M. & TAWFIK, D. S. 2010. Mutational effects and the evolution of new protein functions. *Nat Rev Genet,* 11**,** 572-82.

STARK, W. M., SHERRATT, D. J. & BOOCOCK, M. R. 1989. Site-specific recombination by Tn3 resolvase: topological changes in the forward and reverse reactions. *Cell,* 58**,** 779-90.

STEINER, W., LIU, G., DONACHIE, W. D. & KUEMPEL, P. 1999. The cytoplasmic domain of FtsK protein is required for resolution of chromosome dimers. *Mol Microbiol,* 31**,** 579-83.

STEINER, W. W. & KUEMPEL, P. L. 1998. Sister chromatid exchange frequencies in *Escherichia coli* analyzed by recombination at the *dif* resolvase site. *J Bacteriol,* 180**,** 6269-75.

STEVENS, A. M., SHOEMAKER, N. B., LI, L. Y. & SALYERS, A. A. 1993. Tetracycline regulation of genes on Bacteroides conjugative transposons. *J Bacteriol,* 175**,** 6134-41.

STIVERS, J. T., JAGADEESH, G. J., NAWROT, B., STEC, W. J. & SHUMAN, S. 2000. Stereochemical outcome and kinetic effects of Rp- and Sp-phosphorothioate substitutions at the cleavage site of vaccinia type I DNA topoisomerase. *Biochemistry,* 39**,** 5561-72.

STUDIER, F. W. & MOFFATT, B. A. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol,* 189**,** 113-30.

SUBRAMANYA, H. S., ARCISZEWSKA, L. K., BAKER, R. A., BIRD, L. E., SHERRATT, D. J. & WIGLEY, D. B. 1997. Crystal structure of the site-specific recombinase, XerD. *Embo j,* 16**,** 5178-87.

SUMMERS, D. K. & SHERRATT, D. J. 1984. Multimerization of high copy number plasmids causes instability: CoIE1 encodes a determinant essential for plasmid monomerization and stability. *Cell,* 36**,** 1097-103.

TAMURA, K., STECHER, G., PETERSON, D., FILIPSKI, A. & KUMAR, S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol,* 30**,** 2725-9.

TIAN, L., CLAEBOE, C. D., HECHT, S. M. & SHUMAN, S. 2005. Mechanistic plasticity of DNA topoisomerase IB: phosphate electrostatics dictate the need for a catalytic arginine. *Structure,* 13**,** 513-20.

TOMB, J. F., WHITE, O., KERLAVAGE, A. R., CLAYTON, R. A., SUTTON, G. G., FLEISCHMANN, R. D., KETCHUM, K. A., KLENK, H. P., GILL, S., DOUGHERTY, B. A., NELSON, K., QUACKENBUSH, J., ZHOU, L., KIRKNESS, E. F., PETERSON, S., LOFTUS, B., RICHARDSON, D., DODSON, R., KHALAK, H. G., GLODEK, A., MCKENNEY, K., FITZEGERALD, L. M., LEE, N., ADAMS, M. D., HICKEY, E. K., BERG, D. E., GOCAYNE, J. D., UTTERBACK, T. R., PETERSON, J. D., KELLEY, J. M., COTTON, M. D., WEIDMAN, J. M., FUJII, C., BOWMAN, C., WATTHEY, L., WALLIN, E., HAYES, W. S., BORODOVSKY, M., KARP, P. D., SMITH, H. O., FRASER, C. M. & VENTER, J. C. 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature,* 388**,** 539-47.

TRUST, T. J., ALM, R. A. & PAPPO, J. 2001. *Helicobacter pylori*: today's treatment, and possible future treatment. *Eur J Surg Suppl***,** 82-8.

TU, C. P. & COHEN, S. N. 1980. Effect of DNA sequences adjacent to the termini of Tn3 on sequential translocation. *Mol Gen Genet,* 177**,** 597-601.

VAL, M. E., BOUVIER, M., CAMPOS, J., SHERRATT, D., CORNET, F., MAZEL, D. & BARRE, F. X. 2005. The single-stranded genome of phage CTX is the form used for integration into the genome of *Vibrio cholerae*. *Mol Cell,* 19**,** 559-66.

VOLKERT, F. C. & BROACH, J. R. 1986. Site-specific recombination promotes plasmid amplification in yeast. *Cell,* 46**,** 541-50.

VOLLMER, W., VON RECHENBERG, M. & HÖLTJE, J. V. 1999. Demonstration of molecular interactions between the murein polymerase PBP1B, the lytic transglycosylase MltA, and the scaffolding protein MipA of *Escherichia coli*. J Biol Chem 274:6726-34.

VONRHEIN, C., BLANC, E., ROVERSI, P. & BRICOGNE, G. 2007. Automated structure solution with autoSHARP. *Methods Mol Biol,* 364**,** 215-30.

VOZIYANOV, Y., LEE, J., WHANG, I., LEE, J. & JAYARAM, M. 1996. Analyses of the first chemical step in Flp site-specific recombination: Synapsis may not be a pre-requisite for strand cleavage. *J Mol Biol,* 256**,** 720-35.

WANG, L. & LUTKENHAUS, J. 1998. FtsK is an essential cell division protein that is localized to the septum and induced as part of the SOS response. *Mol Microbiol,* 29**,** 731-40.

WATERHOUSE, A. M., PROCTER, J. B., MARTIN, D. M., CLAMP, M. & BARTON, G. J. 2009. Jalview Version 2 - a multiple sequence alignment editor and analysis workbench. *Bioinformatics,* 25**,** 1189-91.

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. *J Mol Biol,* 313**,** 49-69.

YANG, W. & MIZUUCHI, K. 1997. Site-specific recombination in plane view. *Structure,* 5**,** 1401-6.

YATES, J., AROYO, M., SHERRATT, D. J. & BARRE, F. X. 2003. Species specificity in the activation of Xer recombination at *dif* by FtsK. *Mol Microbiol,* 49**,** 241-9.

YATES, J., ZHEKOV, I., BAKER, R., EKLUND, B., SHERRATT, D. J. & ARCISZEWSKA, L. K. 2006. Dissection of a functional interaction between the DNA translocase, FtsK, and the XerD recombinase. *Mol Microbiol,* 59**,** 1754-66.

ZHENG, G., LU, X. J. & OLSON, W. K. 2009. Web 3DNA - a web server for the analysis, reconstruction, and visualization of three-dimensional nucleic-acid structures. *Nucleic Acids Res,* 37**,** W240-6.