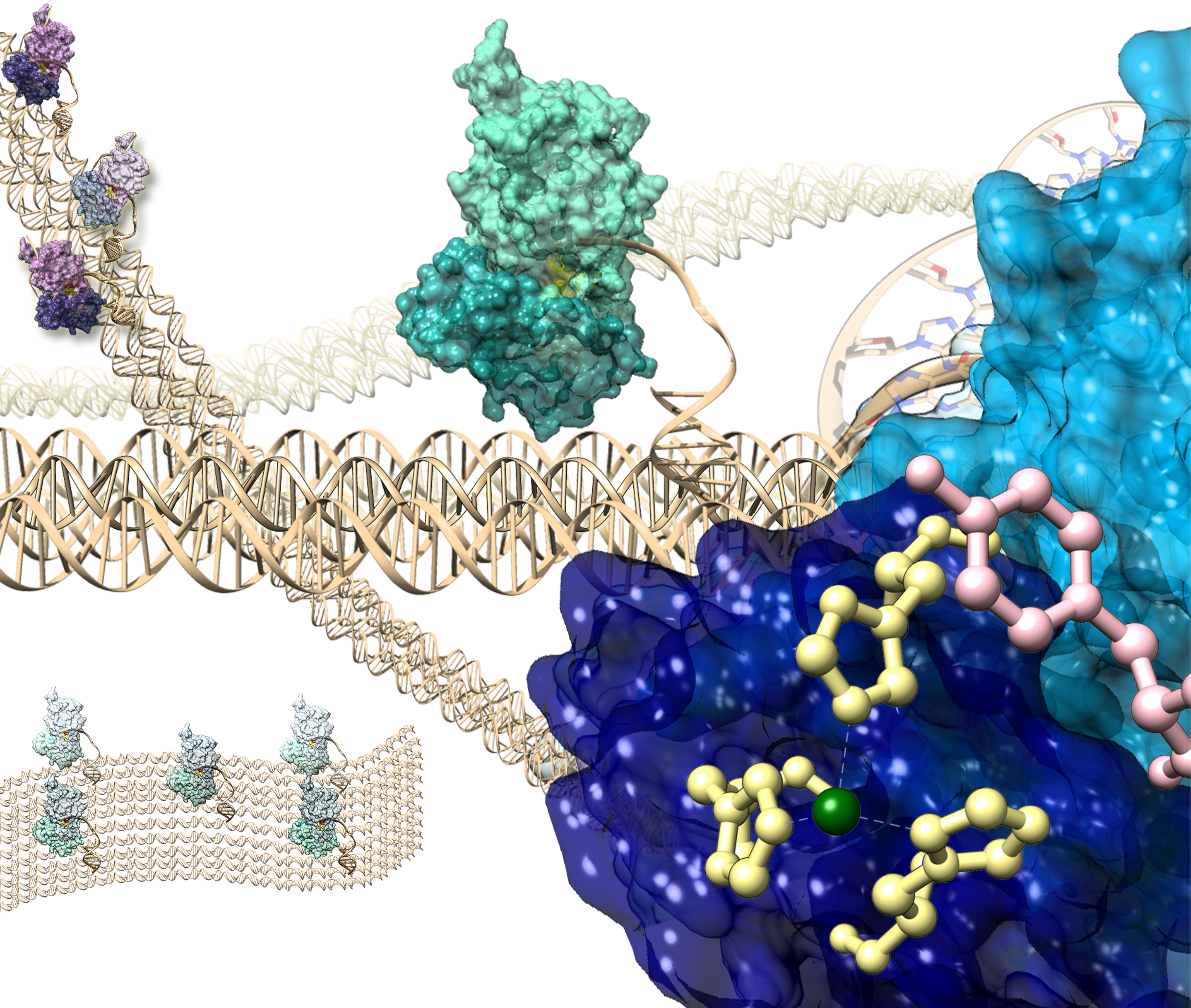
Universidad de Cantabria
Facultad de Medicina
Departamento de Biología Molecular

Sandra Sagredo de Pedro
Santander, 2016

INGENIERÍA DE RELAXASAS PARA EL ENSAMBLAJE DE MÚLTIPLES PROTEÍNAS SOBRE NANOESTRUCTURAS DE ADN



UNIVERSIDAD DE CANTABRIA

FACULTAD DE MEDICINA

DEPARTAMENTO DE BIOLOGÍA MOLECULAR

INGENIERÍA DE RELAXASAS PARA EL ENSAMBLAJE DE MÚLTIPLES PROTEÍNAS SOBRE NANOESTRUCTURAS DE ADN

Sandra Sagredo de Pedro

Memoria presentada para la obtención del grado de Doctor

Santander, 2016

UNIVERSIDAD DE CANTABRIA

FACULTAD DE MEDICINA

DEPARTAMENTO DE BIOLOGÍA MOLECULAR

RELAXASE ENGINEERING FOR MULTIPROTEIN ASSEMBLY ON DNA NANOSTRUCTURES

Sandra Sagredo de Pedro

Memoria presentada para la obtención del grado de Doctor

Santander, 2016

D. **Fernando de la Cruz Calahorra**, Catedrático de Genética, y D. **Gabriel Moncalián Montes**, Profesor Titular de Genética, ambos pertenecientes al Departamento de Biología Molecular de la Universidad de Cantabria,

CERTIFICAN: Que Dña. **Sandra Sagredo de Pedro** ha realizado bajo nuestra dirección el presente trabajo de Tesis Doctoral titulado: **Ingeniería de relaxasas para el ensamblaje de múltiples proteínas sobre nanoestructuras de ADN**.

Consideramos que dicho trabajo se encuentra terminado y reúne los requisitos de originalidad y calidad científica para su presentación como Memoria de Doctorado al objeto de poder optar al grado de Doctor por la Universidad de Cantabria.

Y para que conste y surta los efectos oportunos, expedimos el presente certificado.

Santander, a 10 de Marzo de 2016.

Fdo. Fernando de la Cruz Calahorra Fdo. Gabriel Moncalián Montes

El presente trabajo ha sido realizado en el Departamento de Biología Molecular de la Universidad de Cantabria, bajo la dirección de los profesores Doctores Fernando de la Cruz Calahorra y Gabriel Moncalián Montes.

Parte de los experimentos presentados en la Memoria de Doctorado fueron realizados por la Lda. Sandra Sagredo de Pedro durante varias estancias en el laboratorio del profesor Doctor Friedrich Simmel en la Technische Universität München (TU Munich, Alemania).

This work was conducted in the Department of Molecular Biology, University of Cantabria, under the supervision of professor Dr. Fernando de la Cruz Calahorra and professor Dr. Gabriel Moncalián Montes.

Parts of the experiments presented in this thesis were made by Sandra Sagredo de Pedro during several stays in the laboratory of professor Dr. Friedrich Simmel at the Technische Universität München (TU Munich, Germany).

Esta Tesis ha sido financiada con las siguientes ayudas:

-BACTOCOM (FP7-ICT-2009-4-248919)

-EMBO Short Term Fellowship (ASTF 141-2012)

-COBRA Project (270371)

-PLASMWIRES (FP7-ICT-2013-10-612146)

- Volkswagen Stiftung (Az.86 395-1)

Me gustaría agradecer a mis tutores Gabriel Moncalián y Fernando de la Cruz por darme la oportunidad de conocer el mundo de las bacterias y la ingeniería de proteínas. Gracias porque no sólo he aprendido innumerables conceptos y técnicas con vosotros, sino porque también a nivel personal me habéis dado grandes lecciones durante estos cinco años juntos. Gracias por las sugerencias escribiendo esta tesis y los artículos, pero sobretodo gracias por el apoyo durante estos años. En especial, me gustaría agradecer a Gabriel por su extraordinaria forma de motivar, por su empatía, por enseñarme a encontrar soluciones a cada dificultad en el laboratorio y sobretodo por su amor a la ciencia. A Fernando, te estaré siempre agradecida porque gracias a tus ideas innovadoras y tu experiencia has conseguido formarme como investigadora. A Mati, por su ayuda en el laboratorio y sus ánimos. También quiero agradecer a la gente del laboratorio, del grupo, del IBBTEC e IDIVAL y de la facultad de medicina que me ha ayudado durante estos años.

Estoy muy agradecida a Friedrich Simmel porque gracias a él he descubierto el nanomundo, y también a ver el ADN y las proteínas desde una perspectiva que antes no me había planteado. Siento que soy una privilegiada por haber tenido la oportunidad de haber trabajado en su grupo y haber disfrutado tanto aprendiendo con ellos. Especialmente, gracias a Tobias, Ali y Marisa por su ayuda con los microscopios, y a todos los compañeros del e14.

Reconozco que nada de esto hubiera sido posible si no hubiera tenido el apoyo incondicional de mi familia. Agradezco mucho a mis padres y hermanas por estos años llenos de alegrías y metas cumplidas. No tengo palabras para describir lo mucho que os quiero y lo que os agradezco sentiros tan cerca a pesar de la distancia.

Quiero agradecer a mis amigas de Burgos que sigamos tan juntas como siempre, disfrutando de las increíbles aventuras que nos pasan. También a mis “optiqueros”, mis amigos de León, de Lisboa y de Sevilla, me encanta viajar con vosotros, ¡mis pequeños trotamundos! Gracias a todos vosotros es como que el tiempo no pasa. En especial, quiero agradecer a la gente que he conocido en Cantabria por enseñarme tantos rincones y rutas increíbles, por el cariño y por hacerme sentir "como en casa". También quiero dar las gracias a mis amigos de Múnich, por las cenitas afterwork y las rutas por los Alpes. En general, gracias a todos vosotros, mis familiares y amigos, gracias por tantas visitas, llamadas y viajes juntos, me siento una privilegiada por haber compartido y seguir compartiendo momentos increíbles con vosotros. Y también, gracias a todos los que de algún modo han ayudado a que esta tesis y mi investigación de estos años salieran adelante.

I would like to express my most profound gratitude to my advisors Gabriel Moncalián and Fernando de la Cruz for giving me the chance to know the bacterial world and protein engineering. Thank you because I've learned uncountable concepts and techniques with you, and also, on a personal level, you have given me great lessons during these five years. Your constructive feedback and good advice in writing this thesis has been invaluable. Specially, I would like to thank Gabriel for all encouragement, empathy and for your endless suggestions of what to do when things did not go according to plan in the laboratory. To Fernando for your support, thanks to your innovative ideas and creativity I became a researcher. Mati, thanks for your support during these years. Thanks to the people in the group, IBBTEC, IDIVAL and Medicine Faculty who in some way has helped and supported me during these years.

I'm very grateful to Friedrich Simmel because I've discovered the nanoworld with you and now, I can see the DNA and proteins with a perspective I've couldn't imagine before. I feel very privileged to have been given the opportunity to work in your group. Specially, thanks to Tobias, Ali and Marisa for your support with the microscopes. Thanks to everyone in the e14. I have enjoyed with you a lot!

I recognize that the completion of this thesis would not have been possible without the day-to-day support of my family. I am very thankful to my parents and sisters for their encouragement and motivating me to achieve my goals. There are no words to describe how much I love you and appreciate feeling you so close despite the distance.

I want to thank my friends from Burgos for always being together, enjoying the adventures that happened to us when we were at school and also when we meet in Burgos, Germany and Santander. I also thank the friends I meet in Valladolid, Leon, Lisbon and Seville because I enjoyed a lot travelling with you! It's awesome feeling like nothing has changed despite the distance. Specially, I would like to thank the amazing people I meet in Cantabria. Thanks for showing me incredible places, for your care and for make me feel I'm at home. Thanks to my friends from Munich for our many conversations and trekking in the Alps. I would like to express my most profound gratitude to everyone for so many visits, calls and trips together. Thanks everyone who in some way has helped in the making of this thesis and my research work.

You have to try the impossible to achieve the possible.

Para que pueda surgir lo posible es preciso intentar una y otra vez lo imposible.

— Hermann Hesse

A mis padres y hermanas.

CONTENTS

ABBREVIATIONS

Ǻ	Ǻngstrøm (10^{-10} m)
AFM	atomic force microscopy
ATP	adenosine triphosphate
AP	accessory protein
APS	ammonium peroxodisulfate
AAV	adeno associated virus
bp	base pair
CFP	cerulean fluorescent protein
CNT	carbon nanotube
CRAfT	Cre reporter assay for translocation
CSS	complexation significance score
CuAAC	Copper(I) mediated azido-alkyne-cycloaddition
Da	Dalton (g mol^{-1})
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
DR	direct repeats
dsDNA	double strand DNA
DTT	dithiothreitol
EDTA	ethylenediamine tetraacetic acid
EMSA	electrophoretic gel mobility shift assay
EPL	expresed protein ligation
FRET	Förster Resonance Energy Transfer
g	gram
GFP	green fluorescent protein
HUH	histidine-hydrophobic-histidine
ICE	integrative and conjugative elements
IPTG	isopropyl-β-thiogalactopyranoside
IR	inverted repeats
kb	kilobase
kDa	kilodalton
K_D	dissociation constant
LB	Luria-Broth medium
M	mol l^{-1}
MW	molecular weight

NHS	N-hydroxysuccinimide
Ni^{2+}-NTA	nickel-nitrilotriacetate
NLS	nuclear localization signal
NMR	nuclear magnetic resonance
NP	nanoparticle
nt	nucleotide
OD	optical density
oriT	origin of transfer
PAGE	polyacrylamide gel electrophoresis
PEG	polyethylene glycol
PCR	polymerase chain reaction
POI	protein of interest
QD	Quantum Dot
Rep	RCR-initiator proteins
RMSD	root mean square deviation
RNA	ribonucleotide acid
RCR	rolling circle replication
RT	room temperature
SDS	sodium dodecyl sulphate
ssDNA	single strand DNA
Sulfo-EMCS	N-ε-maleimidocaproyl-oxysulfosuccinimide ester
TAE	Tris Acetate EDTA
TBE	Tris Borate EDTA
TEM	transmission electron microscopy
TEMED	N,N,N’,N’-tetramethylethylendiamine
TS	translocation signals
T4CP	type IV coupling protein
T4SS	type IV secretion system
Y1	single conserved catalytic tyrosine
Y2	more than one conserved catalytic tyrosine
ZFP	zinc finger protein
1D	one-dimensional
2D	two-dimensional
3D	three-dimensional

LIST OF TABLES AND FIGURES

List of Figures

List of Tables

INTRODUCTION

Natural structures made of nucleic acids, such as telomeres, ribozymes, ribosomes and many other systems, have complex 3D shapes that allow them to accomplish complex functions in the cell. Base complementarity promotes the self-assembly of these biological structures. Besides, DNA tendency to self-assemble into double helical structures and its diameter of two nanometers makes it also an optimal material for building artificial nanostructures. The breakthrough of DNA origami in the last decade made DNA-based nanostructures designed in almost any desirable shape. Artificial nanostructures have been used for drug delivery (Lucas et al., 2010; Douglas et al., 2012a; Zhang et al., 2014), immune system modulation (Lucas et al., 2010; Li et al., 2011; Schüller et al., 2011; Liu et al., 2012; Mohri et al., 2012), cell signaling (Lucas et al., 2010; Douglas et al., 2012b; Pedersen et al., 2013; Shaw et al., 2014), biosensing (Ke et al., 2008; Lucas et al., 2010; Tintoré et al., 2013) and for the arrangement of enzymatic cascades (Voigt et al., 2010; Delebecque et al., 2011; Fu et al., 2012). Applications in biomedicine imply the functionalization of the nanostructures with nanoparticles or proteins, but many current methods for bioconjugation involve chemical crosslinking that often compromise the activity of the proteins (Rinker et al., 2008; Saccà et al., 2010; Yang et al., 2015). Novel methods to accomplish protein incorporation to DNA nanostructures are required to solve the disadvantages of the bioconjugative methods developed so far.

This thesis reports an approach for addressing the nanoscale arrangement of proteins on DNA nanostructures based on the biospecificity of relaxases and the capacity of these enzymes to form covalent complexes at targeted sequences on DNA. We investigated the coupling of relaxases to DNA without any chemical modification as a tool for bionanotechnology (Chapter 1). Our goal is to use relaxases as an adaptor scaffold for other proteins of interest with optimized targets (Chapter 2) on DNA nanostructures.

1. DNA Nanotechnology

Nanotechnology is defined by the National Science Foundation as research and technology development at the atomic, molecular or macromolecular levels, in the length scale of approximately 1-100 nanometer range, to provide a fundamental understanding of phenomena and materials at the nanoscale and to create and use structures, devices and systems that have novel properties and functions because of their small and/or intermediate size (http://www.nsf.gov/crssprgm/nano/reports/omb_nifty50.jsp) .

Nanotechnology comprises two strategies: top-down and bottom-up. The bottom-up strategy is based on the self-organization of atomic- and molecular-scale building blocks to create a functional structure, in contrast to the top-down approach, that reduces large pieces of materials by microfabrication methods. Some examples of the bottom-up approach are the carbon nanotubes (CNTs), which consist exclusively of carbon atoms and possess interesting electrical and mechanical properties. Other examples are polymers, whose monomers aggregate with themselves to generate supramolecular structures, such as vesicles or laminar planes.

DNA nanotechnology forms part of the bottom-up branch because it builds complex structures of nanometer size based on DNA hybridization. The founder of DNA nanotechnology is Nadrian Seeman, who in the 80s designed and started to use DNA to make artificial objects (Seeman and Kallenbach, 1983; González-Pérez et al., 2007). Inspired by the Holliday Junction, a recombination intermediate in the cell, he proposed the use of double crossovers to build 2D lattices via sticky ends and also to fold a DNA cube and other nanostructures (Figure 1A-C) (Seeman, 1982; Fu and Seeman, 1993; Nash et al., 2010). Moreover, nanowires have been synthesized by rolling circle amplification (Figure 1D) (Beyer et al., 2005; Nash et al., 2010). A variety of complex substrates such as 3D objects and lattices were also self-assembled with high-yield following Seeman´s strategy (Figure 1E and F) (Winfree et al., 1998; Steinhauer et al., 2009; Lin et al., 2012).

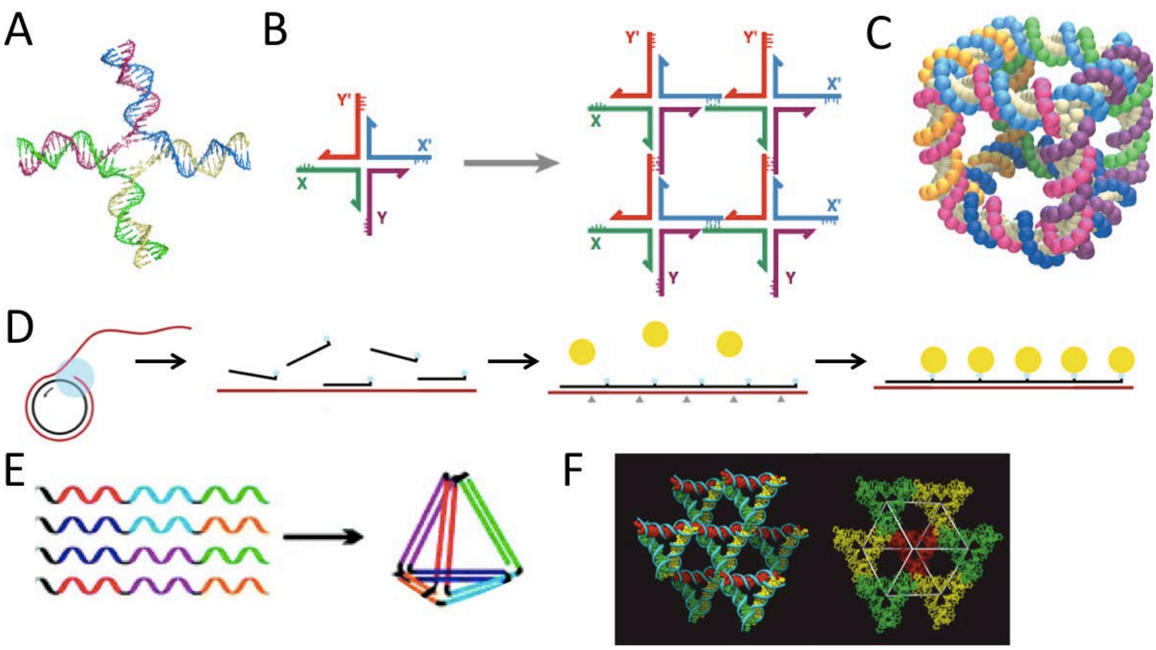


Figure 1. Early constructs made of DNA. (A) Holliday junction (PDB 1KBU). (B) Self-assembly of branched DNA molecules via sticky ends. The resulting structure is a 2D lattice. (C) Seeman's cube. D) Periodic DNA nanowires synthesized by rolling circle amplification. The ssDNA template (in red) hybridizes with oligonucleotides (in black) that allows the functionalization with nanoparticles (yellow). (E) Icosahedric nanostructure generated with four oligonucleotides. (F) 3D lattice formed by triangles. From Beyer et al., 2005; González-Pérez et al., 2007; Seeman, 2010; Li et al., 2011.

In any case, the great breakthrough in this field was achieved in 2006 by Paul Rothemund, who proposed the DNA origami method (Beyer et al., 2005; Rothemund, 2006; Jungmann et al., 2011). It is named after the ancient Japanese art of paper folding, since long single-stranded DNA (ssDNA) molecules from phage M13, “the scaffold”, are folded in any predetermined shape with over 200 short oligonucleotide “staple strands” (Figure 2A). The staple strands are complementary to distinct parts of the scaffold strand and create Holliday junctions that keep the helices together. The availability of commercially synthetized staple strands and the high-yield of the resulting DNA nanostructures enable the assembly of complicated three-dimensional DNA nanostructures (Figure 2B) (Shen et al., 2004; Douglas et al., 2009a).

Different steps are required for the synthesis of a DNA origami nanostructure. After conceiving a target design, the staples are designed and optimized using specific software (Douglas et al., 2009b; Jungmann et al., 2011). The next step is ordering the staple strands and purify the M13 scaffold.

Staples are then pooled to Eppendorf tubes from the 96-well plates, and mixed in a 3-5:1 ratio with the M13 scaffold in presence of magnesium. A long thermal annealing ramp (typically from 95ºC to 4ºC at a constant rate) leads to the self-assembly of the nanostructure. The final step is the evaluation of the folding process with agarose gels and structural features and details with atomic force microscope (AFM) for 1D and 2D structures and transmission electron microscopy (TEM) for 3D nanostructures. (Figure 2C). A detailed description of these microscopy techniques is given in Experimental procedures (4.2 Biophysical methods).

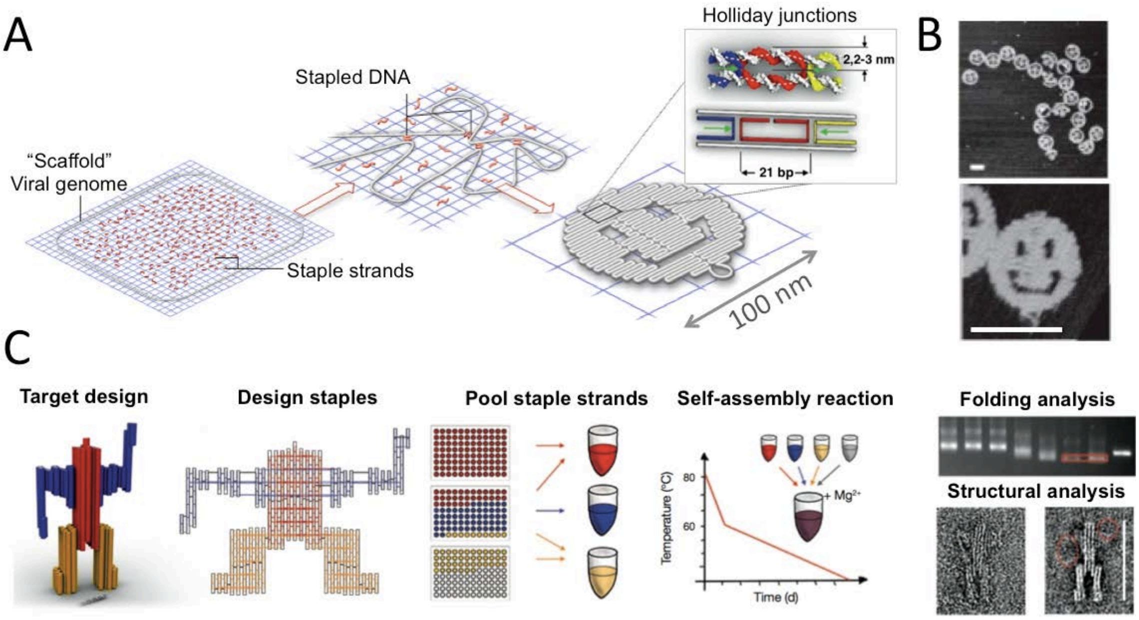


Figure 2. Scaffolded DNA origami. (A) With about 250 specific 'staple' strands of DNA, the 7 Kbp single stranded viral genome of M13 is folded into a 100 nm smiley face. (B) AFM images of the DNA nanostructures. In black is shown the mica, in gray is the DNA. Scale bar 100 nm. (C) Steps of the DNA origami process. See text for details. Adapted from Sanderson, 2010; Castro et al., 2011.

1.1 DNA Origami Based Devices

During the synthesis of a nanostructure, the free-software caDNAno (http://cadnano.org) is used for the general design of the nanostructure and for generating the sequences of the staples (Figure 3) (Douglas et al., 2009b). In addition, the free resource CanDo (http://cando-dna-origami.org) estimates its flexibility and helps in the improvement of the final structure. Each new desired nanostructure requires a new scaffold routing strategy and hence new set of staples.

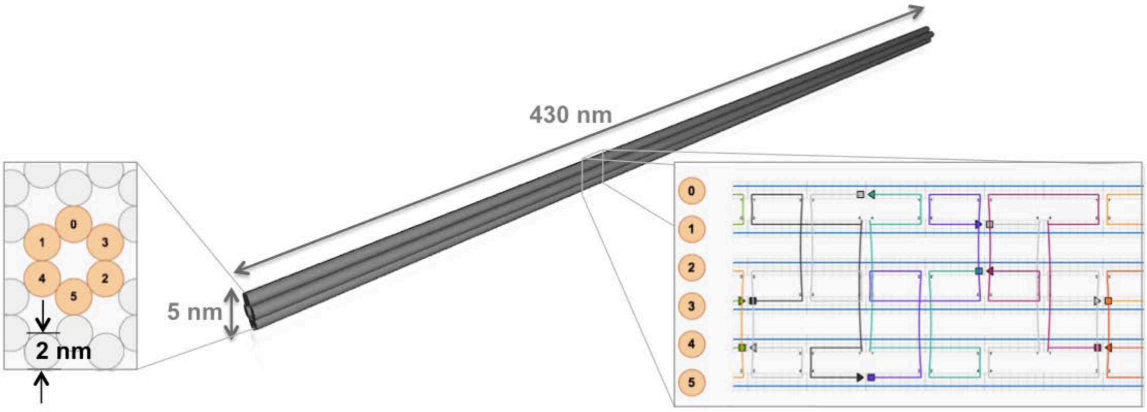


Figure 3. Six helix bundle (6HB) design. caDNAno design software image of the 3D nanostructure. Each cylinder depicts each one of the 6 DNA double helices. Left inset; the honeycomb structure of the 6 helices numbered from 0 to 5 (light orange). Right inset, the M13 viral genome (in blue) is folded with 176 staples (different colours) connected via holliday juntions.

Usually, 2D nanostructures with one layer of DNA are 2 nm wide and 100 nm long, while 3D nanostructures have 50 to 400 nm sides. Since the DNA is negatively charged, repulsion between the double helices could result in an unstable arrangement. A divalent cation, preferentially magnesium, is used for the folding and imaging.

There are several advantages of DNA as a nanobuilding material. DNA can be produced and handle easily, it is stable in a wide range of pHs and temperatures and a variety of chemical modifications can be introduced commercially to the 3’ and 5’ end, into the internucleotide backbone or to the bases. DNA nanostructures are relatively stable against nucleases (Mei et al., 2011; Hahn et al., 2014). Moreover, they can escape from the immune system controls because many of the structures are smaller than 200 nm (Chen et al., 2013). Interestingly, RNA can also be used for building nanostructures. RNA nanostructures can be self-assembled isothermally *in vivo* and serve to arrange recombinant proteins fused to specific RNA binding domains (Sachdeva et al., 2014).

1.2 Functionalization of DNA nanostructures

Most applications of DNA-nanostructures in biomedicine require functionalization with proteins or nanoparticles (NP). These functional molecules can be attached to either specific DNA sequences or functional groups in the staple strands. Moreover, some staple strands can be extended with sequences that hybridize with a DNA probe bound to either proteins or NP. In this section, I will describe in detail these different methods for the functionalization of DNA-nanostructures.

1.2.1 Nucleic acid-mediated nanoparticle assembly

Thiolated DNAs can be attached to the surface of a NP (Kuzyk et al., 2012). DNA-mediated assembly of NPs is an attractive way to organize both metallic and semiconducting NPs into periodic or discrete 1D, 2D and 3D structures (Sharma et al., 2009). Their potential use is related with biolabeling, imaging, diagnostics and photonics. An interesting application uses Quantum Dots, which are fluorescent semiconductor NPs, incorporated in DNA origami shapes to visualize their biodistribution in tumor-bearing mice (Zhang et al., 2014). The preferential accumulation of triangular DNA origamis at the tumor was used for passive drug delivery as will be described in section 1.3.

1.2.2 DNA-Protein crosslinking /Bioconjugation

Mimicking nature, biomolecules can be arranged precisely on DNA nanostructures to study molecular interactions and build drug delivery systems (see also section 1.3). The challenge to create complex arrangements is the developing conjugative methods that couple a protein of interest (POI) specifically to an oligonucleotide. Bioconjugation methods include covalent / non-covalent and site-specific / non-site specific methods (Yang et al., 2015). Their crucial requirement is either the functionalization of both the protein and DNA with chemical tags or the use of recombinant proteins.

1.2.2.1 Non-covalent site-specific conjugation methods

These methods mainly involve commercially available oligonucleotides coupled to chemical groups or small molecules and proteins with nanomolar affinities for their target. We describe five different approaches used to conjugate proteins to DNA nanostructures (Figure 4).

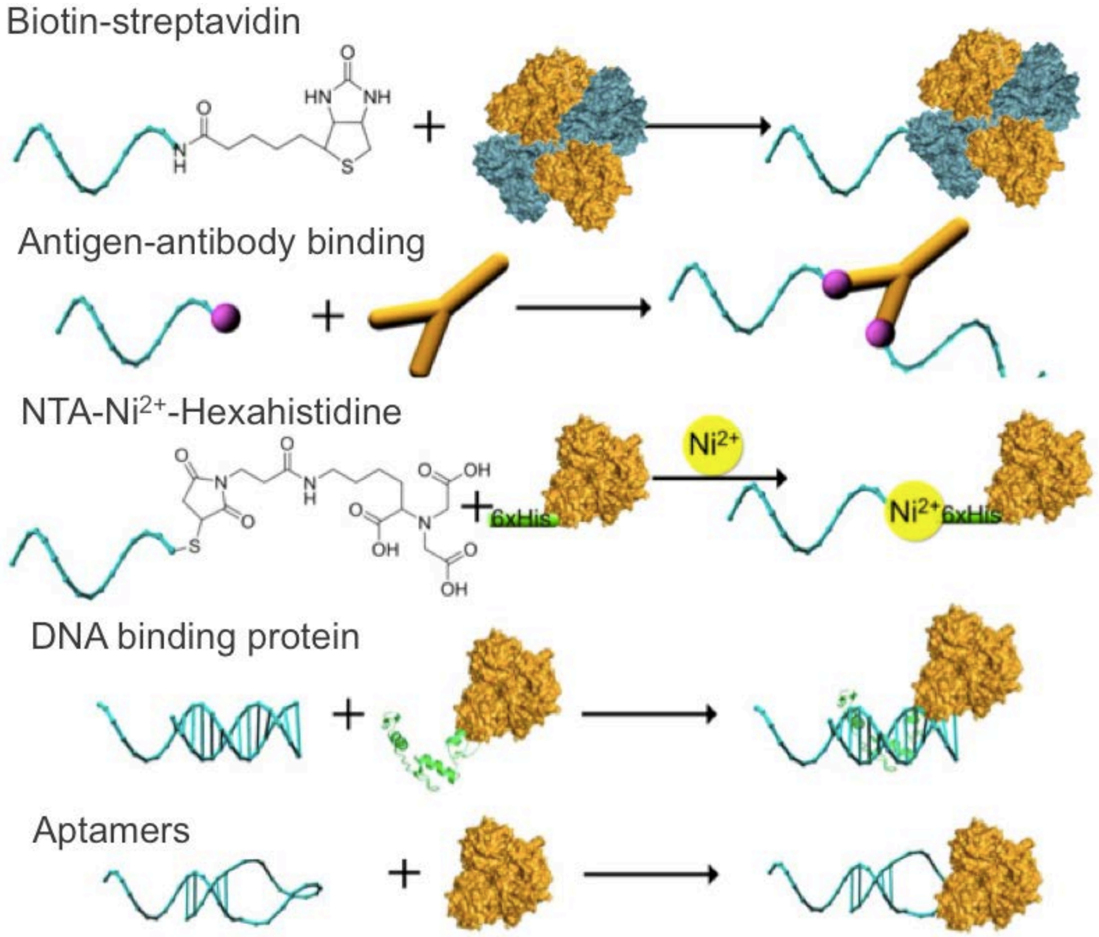


Figure 4. Non-covalent site specific conjugation methods. The target oligonucleotide is depicted as a blue curved line. Streptavidin binds up to four biotinylated oligonucleodies The subunits of the tetrameric streptavidin are coloured in blue and gold. The affinity of the antigen-antibody binding has been used to immobilize IgG on a fluorescein-labelled nanoarray. Fluorescein is depicted as a pink ball. A hexahistidine peptide (6xHis) binds with nitrilotriacetic acid (NTA) ligand through nickel (Ni^{2+}) ion. DNA binding proteins, such as zinc finger proteins (coloured in green), bind to specific sequences of 10 bp. Aptamers are recognized by proteins with high specificy and affinity. Modified from Yang et al., 2015.

Streptavidin

An often-used molecule is streptavidin, which binds biotin with a dissociation constant of K_D=40 fM (Green, 1990), constituting one of the strongest non-covalent bonds known in nature. In order to facilitate binding of streptavidin to DNA nanostructures, biotinylated oligonucleotide staples are incorporated into the structures at the desired locations. Subsequently addition of streptavidin and purification of the nanostructures are required. Its binding yield is about 85%. Kuzyk et al. demonstrated that decoration of DNA nanostructures could be achieved with high yield even when streptavidin proteins were included in the pool of staples and viral DNA during annealing (Kuzyk et al., 2009). The main disadvantage of using streptavidin is the difficulty of controlling the stoichiometry because it is a tetramer that binds up to four biotin molecules (Figure 4). One can be overcome this problem by using monomeric avidin (Yang et al., 2015).

Antibodies

The possibility to attach antibodies that specifically recognize a DNA sequence is difficult to address. Nevertheless, antibodies against different antigens can be used for the immobilization of proteins on DNA nanostructures (Figure 4). Fluorescein-modified DNA strands directed anti-fluorescein monoclonal antibody (IgG) to assemble into 2D arrays (He et al., 2006). These antibody arrays allowed the generation of a well-oriented and high-density platform for proteins arrays.

Affinity tags

A non-covalent attachment of recombinant proteins to DNA was developed based on the Nickel-mediated interaction between nitrilotriacetate (NTA) and a protein carrying a hexa-histidine tag (His-tag). A His-tag at the N or C terminus is routinely incorporated to recombinant proteins because it enables the straightforward purification by using affinity chromatography on a Ni^{2+} NTA-functionalized column. DNA can be functionalized with three adjacent NTA groups for the reversely attachment of His-tag proteins via Ni^{2+} (Goodman et al., 2009). The measured dissociation constant (K_D) of His-tagged protein for NTA-oligonucleotides was 6 nM. The main advantage of this method is that the linkage could be broken controllably by adding either a chelating agent (EDTA) or imidazole.

- **DNA binding proteins**

The group of Morii arranged two well-characterized Zinc-finger proteins (ZFP) on a ten base-pair duplex hold in one cavity of a DNA nanostructure (Nakata et al., 2012). Specific binding of several fused ZFP reached 30-70% and less than 10% of unspecific binding. These results demonstrated an orthogonal binding of ZFP to DNA nanostructures.

- **Aptamers**

Aptamers are single stranded nucleic acids (either DNA or RNA) that fold in a specific molecular shape that is recognized by a given molecular target, which normally does not bind to double helix DNA (Mok and Li, 2008). Aptamers have comparable affinities and specificities than antibodies (Krishnan and Simmel, 2011). They are selected from random pools in an *in vitro* process called SELEX (Systematic evolution of ligands by exponential amplification). Aptamers have been used for detection, immobilization and controlled release of proteins, small molecules or ions (Sun and Zu, 2015). One of the most widely used aptamers targets thrombin, a blood coagulation factor. Two aptamers that fold as G-quadruplex, named TBA A and TBA B, bind to the heparin binding site and the fibrinogen binding site of thrombin, respectively (Lin et al., 2011). Their reported K_D is 0.5 nM and 75-100 nM, respectively.

1.2.2.2 Covalent non-site-specific conjugation methods

Classical conjugation methods that allow the single-site protein modification are difficult because a single cysteine or lysine should be engineered on the surface of the protein. The thiol group of cysteine can form disulfide bonds with 5’-thiol modified oligonucleotides under oxidizing conditions, as well as alkylation with amide reagents, and Michael addition with maleimides (Figure 5) (Sletten and Bertozzi, 2009). Lysine can react with NHS-esters (N-hydroxysuccinimide-esters) to yield amides (Figure 5). Alternatively, heterobispecific crosslinkers like sulfo-EMCS possess two reactive groups at opposite ends: the NHS ester end couples to lysines and the maleimide to the thiolated DNA strand (Hermanson, 2008). The main disadvantages of these methods are the difficult control of the conjugation site, the stoichiometry, and the loss of biological function of the conjugated proteins.

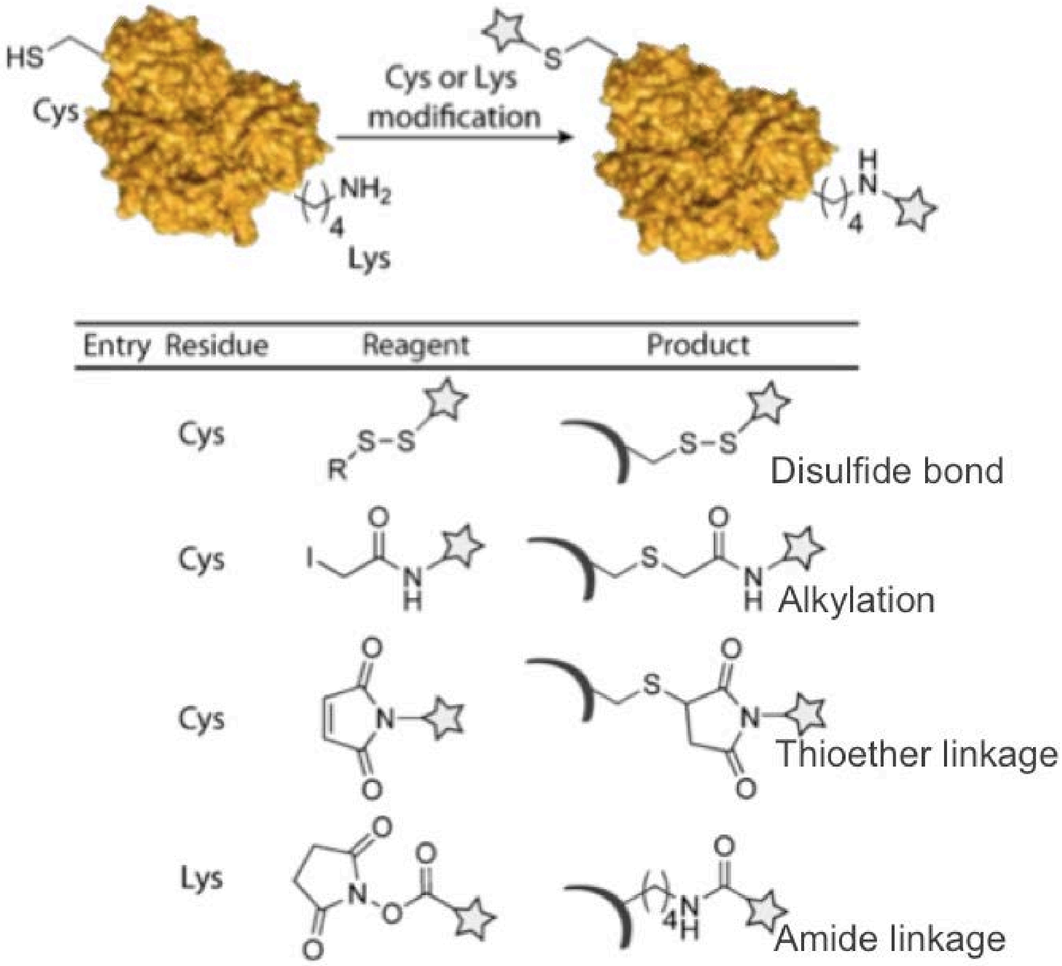


Figure 5. Classic covalent non-site-specific bioconjugation methods. Cys residues can be modified through disulfide exchange, alkylation with iodoacetamide reagents, and Michael addition with maleimides (entries 1–3, respectively). Lys residues can be modified through amide with NHS-activated esters. Adapted from Sletten and Bertozzi, 2009.

1.2.2.3 Covalent site-specific conjugation methods

Additionally, the following covalent conjugation techniques provide site-specific conjugation methods with controlled stoichiometry. The disadvantage is that the structure and thereby the activity of the proteins is compromised by the introduction of non-native chemical tags.

- **Snap-tag, Halo-tag and Clip-tag**

These methods use recombinant fusion-tagged proteins that bind a capture reagent (Figure 6). The human DNA repair protein O_6-alkylguanine-DNA alkyltransferase (hAGT) repairs guanosine residues that are alkylated at the 6-oxo position by transferring the alkyl group to a resident cysteine. The group of Kai Johnson fused hAGT to a protein of interest, and labeled it to DNA via- O^6-benzylguanosine. This method is known as “SNAP tag” (Figure 6) (Keppler et al., 2003). The same group developed the “CLIP tag” with the same enzyme acting on O_2-benzylcytosine derivatives (Gautier et al., 2008). The company Promega developed a “Halo tag”, which consists in a POI fused to a bacterial haloalkane-dehalogenase that has been mutated at the catalytic site (His272Phe) to capture covalently chlorohexane-modified oligonucleotide (Figure 6) (Los et al., 2008).

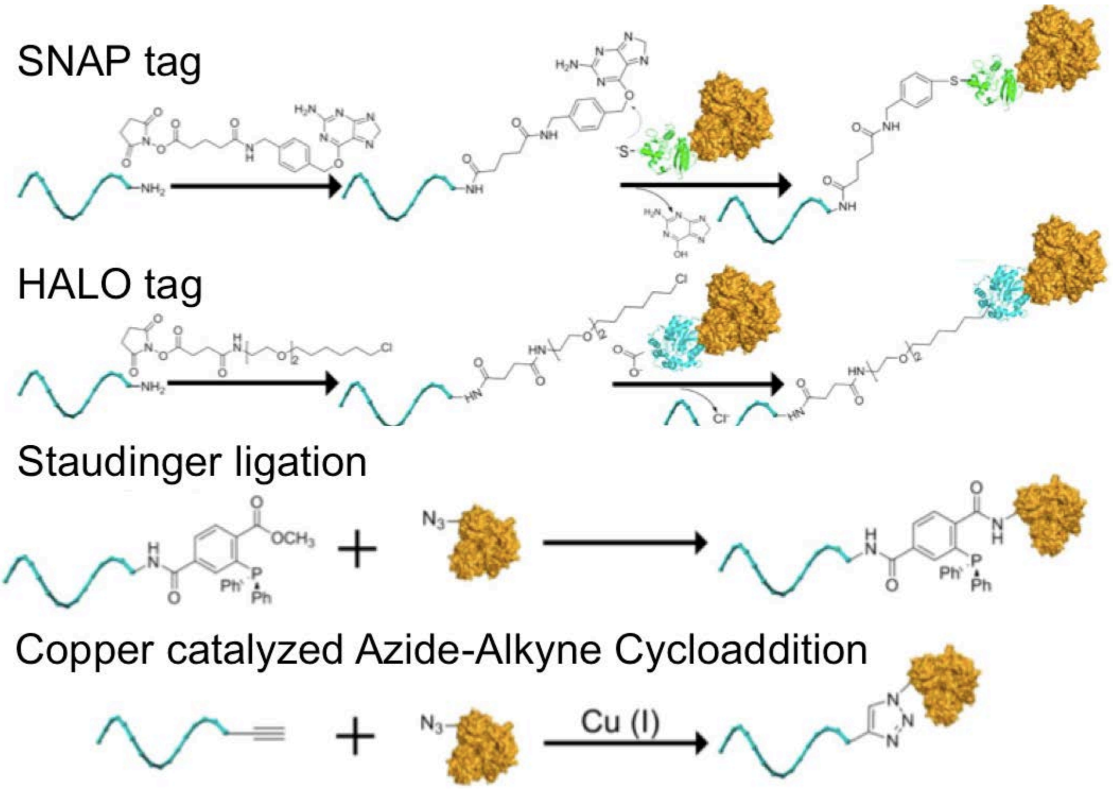


Figure 6. Covalent site-specific conjugation methods. Schemes of protein-DNA conjugation using self-labeling protein tags that lead to covalent bonds DNA-protein when reacting with the enzyme hAGT (SNAP and Halo tag). Schemes of site-specific Azido-protein-DNA conjugation chemistry, including Staudinger ligation, the azido-protein reacts with a phosphine-modified ssDNA; and Copper catalyzed azide-alkyne cycloaddition, the azido-protein reacts with terminal alkyne modified ssDNA. Adapted from Yang et al., 2015.

- ### Staudinger ligation

A useful biorthogonal process to label proteins, DNA, glycan and lipids *in vitro* is the Staudinger ligation, developed by Bertozzi and coworkers (Saxon et al., 2000). Azides and phosphines react to yield amide linkages (Figure 6). However, *in vivo*, this technique has orthogonality problems with various nucleophiles and electrophiles present in complex systems (van Berkel et al., 2011).

- ### "Click" chemistry

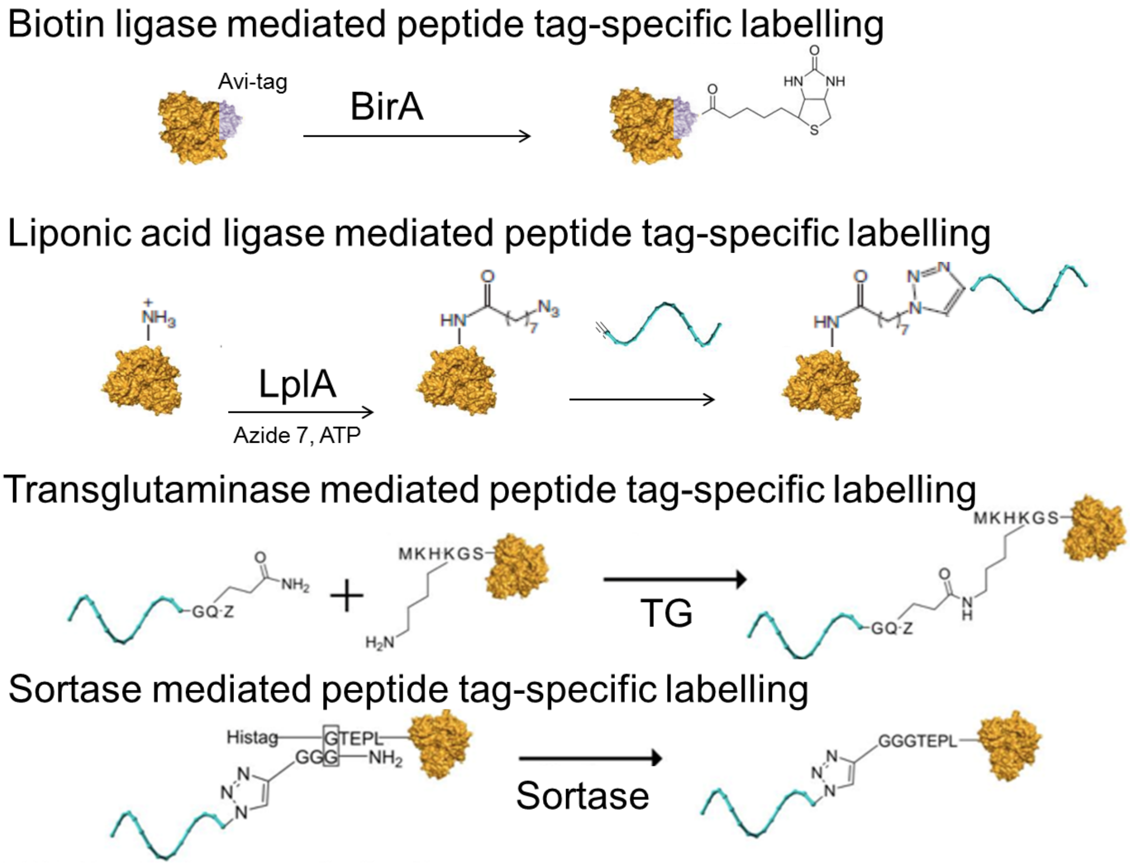
"Click" chemistry efficiently catalyzes covalent attachment of a variety of ligands with selectivity and orthogonality. Click chemistry is almost synonym of copper-catalyzed azide-alkyne cycloaddition in which the azide group of a protein and an alkyne modified oligonucleotide react in the presence of copper to yield a triazole (Figure 6) (McKay and Finn, 2014). Even though this reaction is chemospecific and fast, native biomolecules do not have azides and alkynes, so these groups should be specifically introduced into DNA and proteins. The incorporation of an alkyne in DNA is performed with standard oligonucleotide synthesis. Azido activated esters are used for labeling of peptides and proteins.

- ### Transferases

The specific binding of methyltransferases M.*Hha*I and M.*Msp*I and their capacity to form covalent complexes to 5-fluorocytosine base analogs have been used to attach proteins along a one-dimensional double stranded DNA (dsDNA) fragment (Smith et al., 1997).

- ### Tag-specific labeling

Proteins that modify specifically some peptides with small-molecule cofactors have been used to introduce orthogonal functional groups or moieties. The biotin ligase from *E.coli* biotinylates a lysine residue within a 15-residue acceptor peptide termed the Avi-tag (Sletten and Bertozzi, 2009) (Figure 7). Another method uses lipoic acid ligase (LplA) for site-specific protein labeling. Naturally, LplA couples short-chain azido fatty acids to proteins but it can also ligate an alkyl azide to a lysine side-chain within an engineered LplA acceptor peptide (LAP) (Fernández-Suárez et al, 2007) (Figure 7). Transglutaminases are enzymes that catalyze the conjugation of oligonucleotides modified with a dipeptide substrate (*N*-carbobenzyloxy glutaminyl glycine) to lysine of a recombinant protein fused to a 6-amino acid transglutaminase recognition sequence, called a Q-tag (Tominaga et al., 2007) (Figure 7). This approach has also been used to label engineered Q-tag fused proteins to small probes *in vivo* (Lin and Ting, 2006). The natural function of Sortase A is the conjugation of surface proteins to the bacterial wall of *S. aureus*. Sortase A cleaves a 5 amino acid recognition sequence (LPXTG; X being any amino acid) and forms a new stable peptide bond linking with a glycine on an oligonucleotide (Hagemeyer et al., 2014) (Figure 7).


Biotin ligase mediated peptide tag-specific labelling

Liponic acid ligase mediated peptide tag-specific labelling

Transglutaminase mediated peptide tag-specific labelling

Sortase mediated peptide tag-specific labelling

Figure 7. Tag-specific coupling. Scheme of the enzymes that mediate peptide tag-specific coupling of recombinant proteins. These enzymes recognize a tag sequence in the protein and subsequently transfer biotin or lipoinic acid to it, such as BirA and LplA, respectively, or conjugate a peptide-DNA to the N-terminus of the protein, such as transglutaminase and sortase. Target oligonucleotide is depicted as a blue curved line. Modified from Yang et al., 2015.

• Expressed Protein Ligation

Expressed protein ligation is related to the posttranslational self-splicing of proteins, wherein a domain named intein is able to excise itself and join two exteins through a peptide bond (Takeda et al., 2004). Recombinant proteins with an intein and a chitin-binding domain (CBD) at the C-terminal can be selectively couple to N-terminal cysteine conjugates of nucleic acids (Figure 8) (Lovrinovic and Niemeyer, 2005). When the fusion protein is bound to a chitin-affinity column, low molecular weight thiol compounds are added to activate intein-mediated cleavage. This leads to the formation of a C-terminal thioester that is ligated with the cysteine-nucleic acid conjugate (Saccà and Niemeyer, 2011).

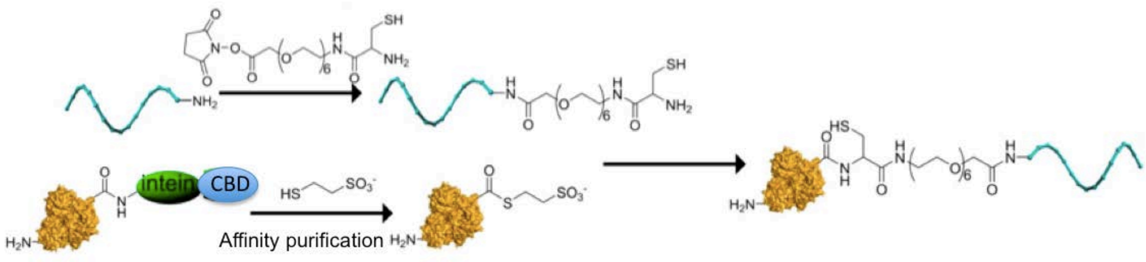


Figure 8. Expressed protein ligation. Scheme of the purification and production of bioconjugates by using an intein protein genetically fused to the protein. Chitin-binding domain (CBD) is used for affinity purification in a chitin column. The intein reacts with thiol compounds to generate thioester and then ligated ot a cysteine modified oligonucleotide. Modified from Yang et al., 2015.

• Enzymatic Ligation

The enzymatic ligation of native DNA to nucleotide triphosphates (NTPs) coupled to proteins and other large macromolecules have been achieved by the use of terminal deoxynucleotidyl transferase (TdT) (Figure 9) (Sørensen et al., 2013). The disadvantage of this strategy is the use of other bioconjugative technique to attach the NTP to the desired macromolecule.

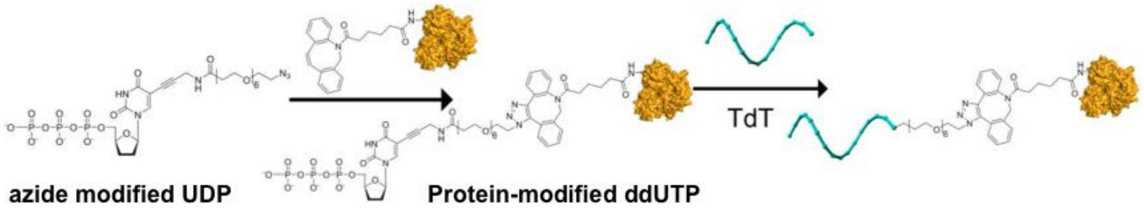


Figure 9. Enzymatic ligation. First, proteins are activated by NTP through a “Click” chemistry reaction. The NTP activated protein is enzymatically ligated to the 3’ end of an oligonucleotide by TdT. Adapted from Yang et al., 2015.

- **HUH-enzymes**

An advanced approach uses HUH endonucleases that naturally cleave the DNA at a specific sequence and form a stable intermediate covalent complex. The protein responsible of the initiation of the replication of phage φX174, gpA, form a covalent phosphotyrosine linkage with oligonucleotides containing its target sequence (Hanai and Wang, 1993). The only requirement of these enzymes is the presence of a divalent metal, usually magnesium. gpA-fused proteins coupled to DNA with a yield of 80-90%, but they were not tested on DNA nanostructures (Mashimo et al., 2012). However, this method has been successfully used to couple an oligonucleotide to antibodies that detect specific antigens (Akter et al., 2012).

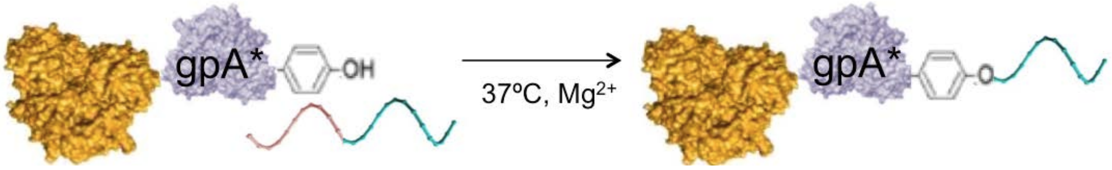


Figure 10. Bioconjugation through gpA. gpA cleaves the recognition sequence (in red), and forms a phosphodiester bond between the 5′-phosphoryl group of the cleavage site and the catalytic tyrosine residue.

1.3 Applications of nanostructures decorated with proteins

In recent years, researches have explored applications of nanostructures related with the creation of synthetic pathways *in vitro* and *in vivo*, structural analysis of proteins, drug delivery and targeted medicine.

1.3.1 DNA as scaffold for enzymatic pathways *in vitro*

Inspired by biological solutions in nature, synthetic enzymatic pathways and nanofactories can be built precisely on DNA nanostructures. The DNA repair activity of the human O^6-alkylguanine-DNA alkyltransferase (hAGT) was visualized using the thrombin aptamer (TBA B) methylated on an origami platform (Tintoré et al., 2013). This methylated aptamer lost the 3D configuration preventing thrombin interaction. The G-quadruplex structure is restored after hAGT incubation and allows thrombin binding to the aptamer (Figure 11A). This single-molecule approach can be used to study the activity of many proteins involved in DNA repair.

In a different example, DNA nanostructures were used as scaffold for coupled reactions. Two enzymes, the glucose oxidase and horseradish peroxidase, were precisely organized at different distances on a rectangular origami. Two different studies concluded that enzyme cascade efficiency was dependent on the interenzyme distance and the diffusion of the substrate (Fu et al., 2012; 2013). Higher efficiency was found with shorter distances because diffusion of the intermediate (H_2O_2) is limited.

Inspired by collagen, a new class of artificial extracellular matrix was constructed from a DNA ribbon surface-functionalized with proteins containing the RGD domain of human fibronectin linked to a monomeric streptavidin domain and a His tag. Within this artificial extracellular matrix, Aldaye et al. studied HeLa cells morphology, cytoskeletal organization and transcription factor localization (Aldaye et al., 2010).

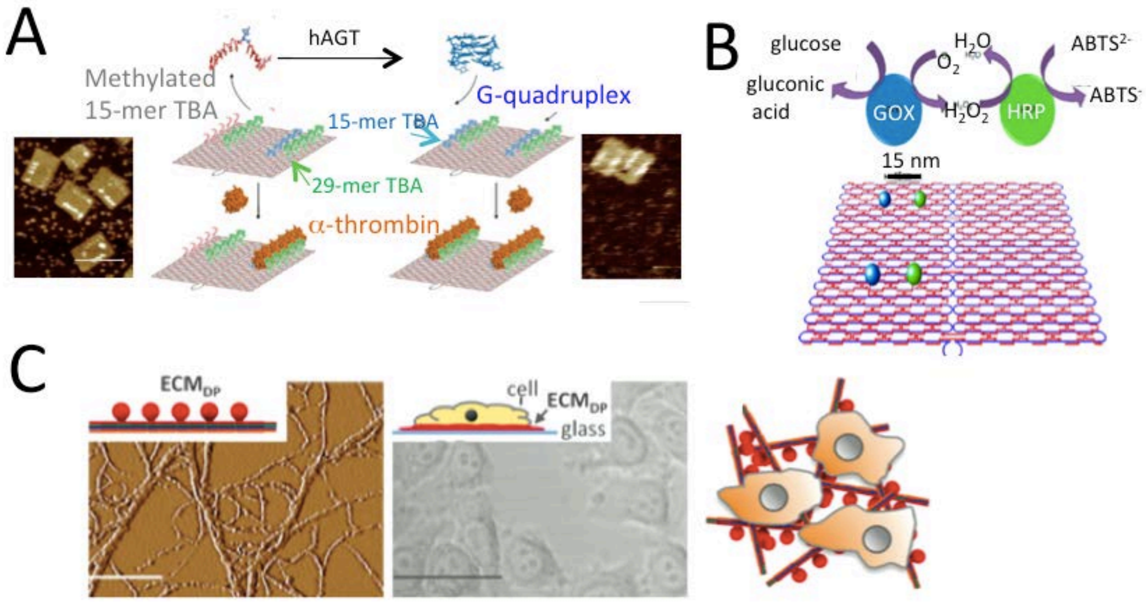


Figure 11. DNA assemblies for study the function of proteins. (A) Representation of the methyl-TBA repair by hAGT. The G-quadruplex configuration of the TBA aptamer is restored after hAGT incubation and leads thrombin binding to TBA-origami. Insets, AFM images of the DNA tiles show

thrombin interaction before (left) and after (right) hAGT incubation. Scale bar 100 nm. (B) Scheme of the reaction of the assembled glucose oxidase/horseradish peroxidase pair on a nanoreactor. The coupled reaction allows the formation of a green product ($ABTS^-$). The distance between these enzymes is 15 nm. (C) DNA/protein based matrix constructed from a DNA-ribbon surface-functionalized with proteins. This matrix promoted cell adhesion and migration of HeLa cells. First frame, AFM characterization of the matrix. Second frame, HeLa cells on matrix-coated glass plates. Scale bar of 15 μm. Third frame, scheme of the attachment. Adapted from Aldaye et al., 2010; Tintoré et al., 2013.

1.3.2 Synthetic cascades *in vivo*

Synthetic multi-enzyme cascade reactions have been built to mimic intracellular biocatalytic processes with the goal of understanding and engineering enzymatic pathways with maximum efficiency and limited cross-talks between signaling pathways. The group of Aldaye engineered synthetic RNA modules that assemble into discrete 1D and 2D scaffolds *in vivo* enabling the arrangement of proteins involved in synthetic cascades (Figure 12A) (Delebecque et al., 2011). These RNA strands possess polymerization domains and two well-known aptamers, PP7 and MS2. PP7 or MS2 aptamer binding proteins were fused to green fluorescent protein (GFP) split into two halves (F_A and F_B). Protein assemblies *in vivo* revealed an increased fluorescence. In a further experiment, hydrogen production was assessed by the assembly of [FeFe]-hydrogenase-PP7 and ferredoxin-MS2 fused proteins (Figure 12B). Hydrogen biosynthesis resulted in an 11- and 48-fold increase on 1D and 2D RNA assemblies, respectively. Therefore, RNA can be used to organize enzymatic pathways into predefined discrete, 1D, and 2D structures *in vivo* to increase output of an enzymatic reaction.

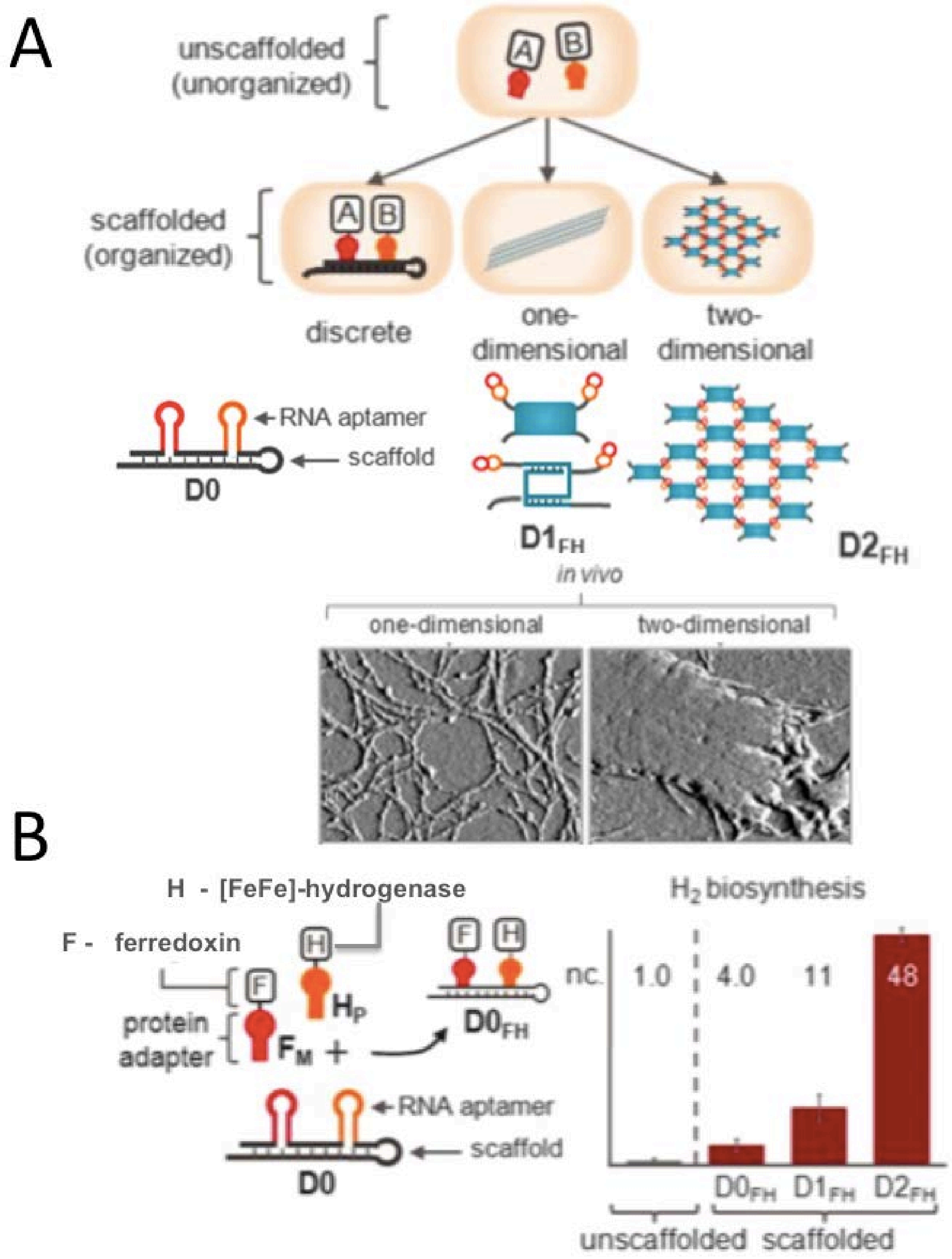


Figure 12. Synthetic cascades *in vivo*. (A) Design of the discrete, 1D, and 2D RNA assemblies to organize proteins (A and B) into cells. RNA folds into a duplex (D0), a ribbon ($D1_{FH}$) or a lattice ($D2_{FH}$) with two aptamers (depicted in red and orange). The aptamers organize protein A and B in these assemblies. AFM images revealed the formation of the 1D and 2D *in vivo* assemblies into thin filaments and compact lattices, respectively. (B) Scheme of the scaffolding hydrogen biosynthesis. Ferredoxin was fused to the MS2 binding protein (F_M) and [FeFe]-hydrogenase was fused to PP7 binding protein (H_P). F_M and H_P bind their respective aptamers and construct a hydrogen-producing pathway. Bar graph showing the hydrogen biosynthesis as a function of scaffold, normalized to unscaffolded cells expressing H_P and F_M. From Delebecque et al., 2011.

1.3.3 Protein function

Synthetic DNA nanostructures have been used to study how motor proteins work together. Motor proteins, such as myosin and kinesin, move along microtubules. Microtubules are thick filaments that contain a “+” end that polymerizes more quickly than the “-” end. Since DNA nanotubes, like 12HB (see Figure 3), mimic the microtubules, dynein and kinesin were coupled to a DNA with SNAP-tag at specific positions of 12HB (Figure 13) (Derr et al., 2012). Super-resolution fluorescence imaging was used to calculate the velocity of one, two, four or seven dynein or kinesin molecules. Dyneins run preferentially to the “-” end, and kinesins to the “+” end. The average speed was 140 nm/s for dynein and 600 nm/s for kinesin. They next mixed these opposite-polarity motors with different ratios (1 dynein:6 kinesin; 2 dynein:5 kinesin, etc). Unexpectedly, all mixed-motor ensembles moved unidirectionally to the “-” end.

Recently, a similar approach has been used to study the velocity of the actin filaments driven by myosin (Debold, 2015). Consistent with a stochastic computer model, an elastic coupling between the motors influenced the velocity of the ensemble.

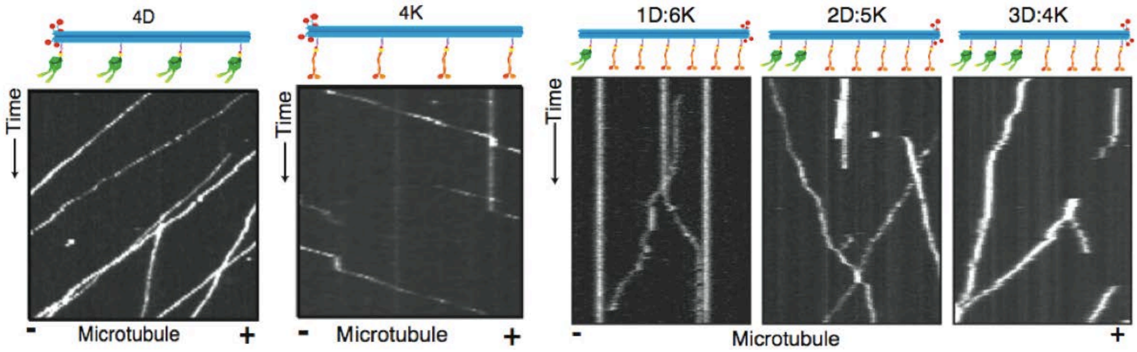


Figure 13. DNA assembled motor proteins. Scheme of the design of a 3D DNA origami (12HB, in blue) covalently attached to fluorophores (5 red dots) and motors: dyneins (D, in green), kinesins (K, in orange). Single-molecule motile properties of 12HB-motor complexes are shown as kymographs. Plus (+) and minus (-) denote microtubule polarity. Scale bars: 1 min (x), 5 µm (y). The 4D ensembles (left panel) moved to the minus end slower than the 4K ensembles (middle panel), which moved to the plus end. When different ratios of dynein and kinesin motors were tested (right panel), the 12HB ensembles were immobile, moving toward the minus end, or moving toward the plus end. From Derr et al., 2012.

1.3.4 Vaccines

The innate immune system recognizes and attacks foreign DNA to prevent infections by virus and bacteria. DNA containing unmethylated cytosine-phosphate-guanine dinucleotides, or CpG motifs, is recognized by the endosomal Toll-like receptor 9 (TLR9), which activates downstream pathways to produce secretion of various pro-inflammatory cytokines (Schüller et al., 2011). The CpG motifs have been included in extensions of 3D DNA tetrahedral and DNA origami nanotubes resulting in enhanced immunostimulatory activity compared with CpG oligonucleotides (Li et al., 2011; Schüller et al., 2011; Mohri et al., 2012). A synthetic vaccine made of 3D DNA origami tetrahedron with CpG motifs and streptavidin as an antigen induced an antibody response against streptavidin (Figure 14) (Liu et al., 2012). Interestingly, DNA nanostructures without CpG motifs did not induce detectable levels of cytokines nor anti-dsDNA antibodies in mice (Schüller et al., 2011; Liu et al., 2012). Therefore, programmable DNA nanostructures provide an excellent platform for construction of vaccines.

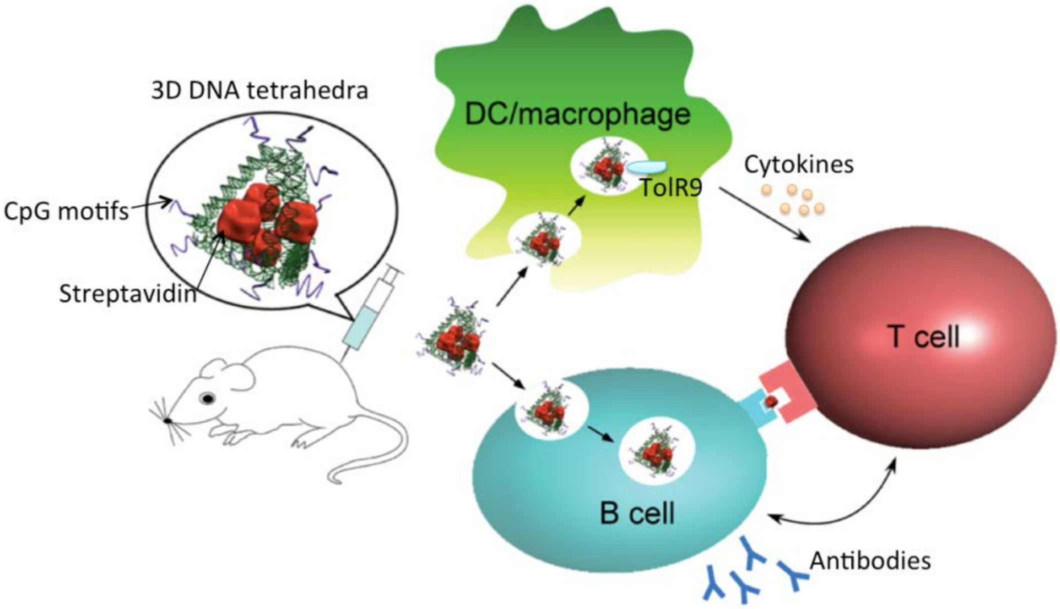


Figure 14. Immunostimulatory effects of DNA nanostructures. Schematic drawing of the DNA tetrahedron vaccine complex (green helices) containing the antigen streptavidin (red) and CpG adjuvant motifs (purple)**.** The injected vaccine complexes bind specifically to B cells and nonspecifically to dendritic cells and macrophages. The complexes are internalized by the three types of antigen-presenting cells, disassembled, and the individual peptide antigens are subsequently presented to T cells to activate B cell response and antibody production. From Liu et al., 2012.

1.3.5 Drug delivery

DNA nanostructures also have a promising future as therapeutic delivery carriers. The strategy of passive targeting accumulates the drug delivery carrier in a non-selective fashion within the tumor because there is a higher permeability and retention effect (Charoenphol and Bermudez, 2014). The anticancer agent doxorubicin (DOX), which intercalates within the double helix, was passively delivered into tumors via DNA nanostructures (Zhao et al., 2012; Choi et al., 2013). The triangle DNA origami was loaded with DOX and the release of the drug was evaluated (Figure 15A) (Zhang et al., 2014). A continuous treatment resulted in a significantly higher ratio of tumor reduction for the DOX/origami than the DOX-treated mice (Figure 15B).

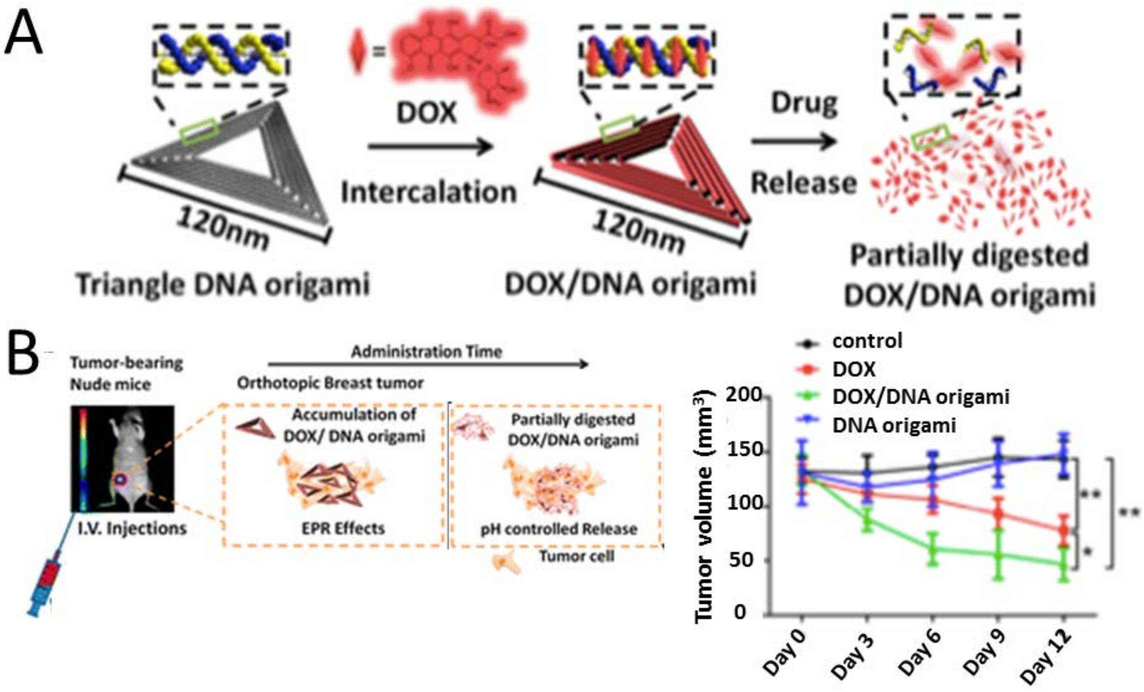


Figure 15. Drug delivery using a passive targeting strategy. (A) The drug DOX intercalates in the DNA on triangular DNA origami. It is released in the tumor, where it is internalized. (B) Antitumor efficacy of drug-loaded DNA origami *in vivo*. DOX/DNA origami complexes were injected to tumor-bearing nude mice. Graph shows the tumor volume evolution with different treatments (control, DOX, DOX/DNA origami, and DNA origami) for 12-days. The tumor volumes were measured using a digital calliper every 3 days. The error bars represent the standard error of the mean of three independent experiments ($*P < 0.05$, $**P < 0.01$). From Zhang et al., 2014.

Nevertheless, this strategy does not have the advantages that an active targeting offers by the specific localization via receptor-ligand interaction. A probed strategy to direct the nanocarrier to target cells uses aptamers incorporated into the “nanorobot” that George Church’s group developed to target leukemic cells (Figure 16A) (Douglas et al., 2012a). The DNA aptamer-based lock mechanism undergoes a drastic reconfiguration in response to binding antigen “keys” that results in an open configuration (Figure 16B). Six different nanorobots with combinations of three aptamers were loaded with fluorescently labeled antibodies against Human Leukocyte Antigen to selectively recognize six different cell lines. Finally, they stimulated the signaling pathway that induces growth arrest in leukemic cells using nanorobots loaded with a combination of antibodies against human CD33 and human CDw328 Fab´ fragments (Figure 16C). The nanorobot induced cell-arrest in a dose-dependent fashion. Therefore, DNA nanostructures can be used to induce a change in cell behavior, but also as a drug payload that can be released to target cells.

Another approach to open and close a 3D DNA box origami structure is in response to DNA keys. DNA keys allow a controlled drug delivery only to cancer cells that express a particular set of miRNAs (Figure 16D) (Andersen et al., 2009).

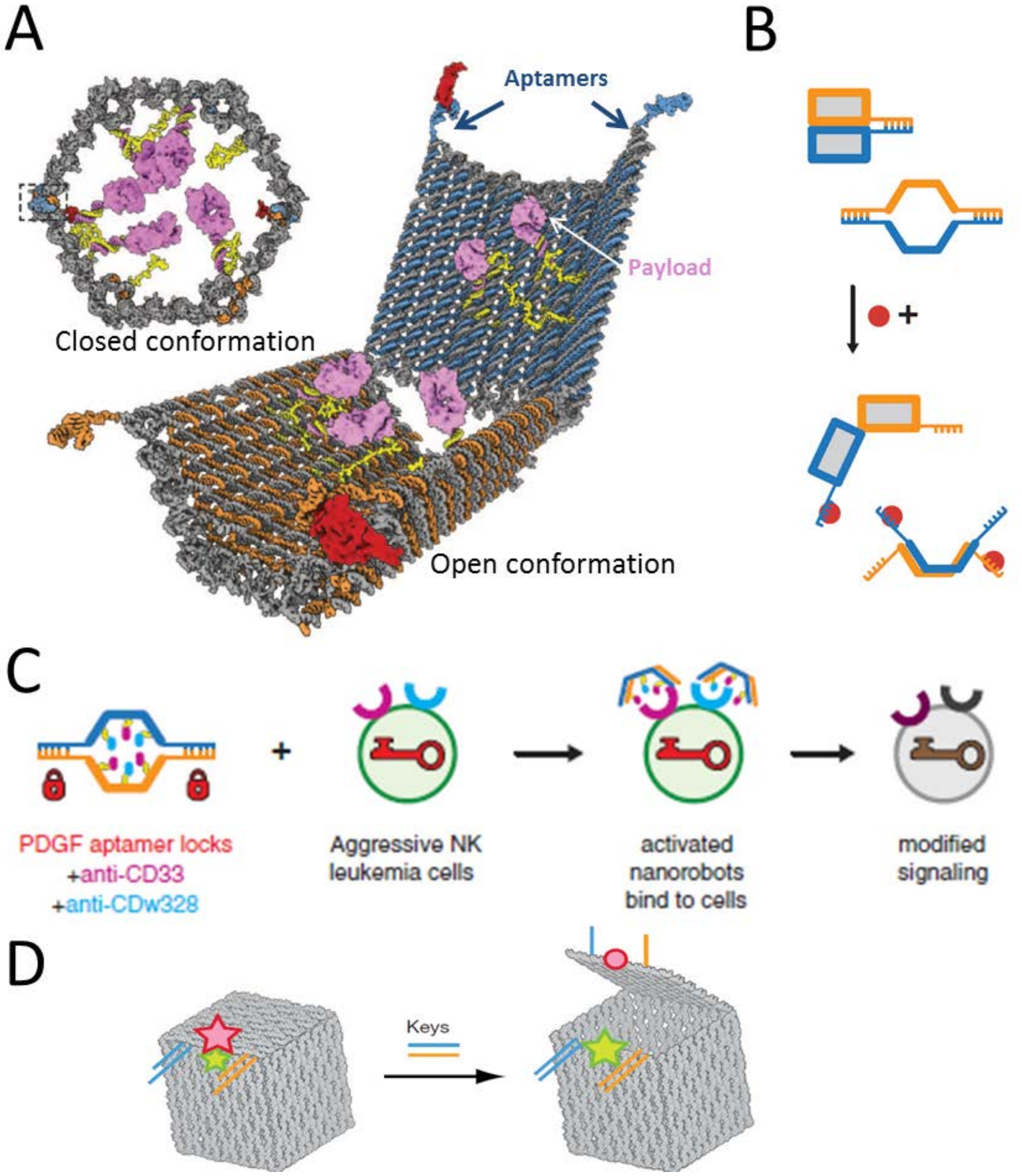


Figure 16. Drug delivery using an active targeting strategy. (A) Schematic view of closed and open states of the nanorobot loaded with a protein payload. Two DNA-aptamer locks fasten the front of the device (boxed in the close conformation). In the internal surface, twelve linker-modified payload attachment sites were arranged, but just four antibodies were attached, on average. (B) Aptamer lock mechanism consists of a DNA aptamer (blue) and a partially complementary strand (orange). The lock can be opened by its antigen key (red). (C) Nanorobots manipulate target NK cell signaling. Aptamer 41t (in red) binds NK cells overexpressing platelet-derived growth factor (PDGF). The activated nanorobots will bind to NK cells expressing CD33 and CDw328 and induce growth arrest. (D) 3D DNA box origami structure activated by a DNA lock. The keys are two oligoncleotides that hybridize with the lock. From Andersen et al., 2009; Douglas et al., 2012a.

1.3.6 Structural analysis

The group of Turberfield developed a new technique for protein structure determination by single-particle cryo-EM based on 2D DNA lattices (Selmi et al., 2011). The lattice consists on one NTA modified oligonucleotide inserted into a hexagonal lattice (Figure 17A). As a consequence, His tagged proteins bind to this nanostructure. The resulting 3D structure can be determined by cryo-EM with high resolution. Another approach uses DNA-nanotubes to enhance the capability of solution NMR in structure determination of α-helical membrane proteins (Figure 17B) (Douglas et al., 2007). These methods validate the potential of DNA nanotechnology to determine high-resolution atomic structures.

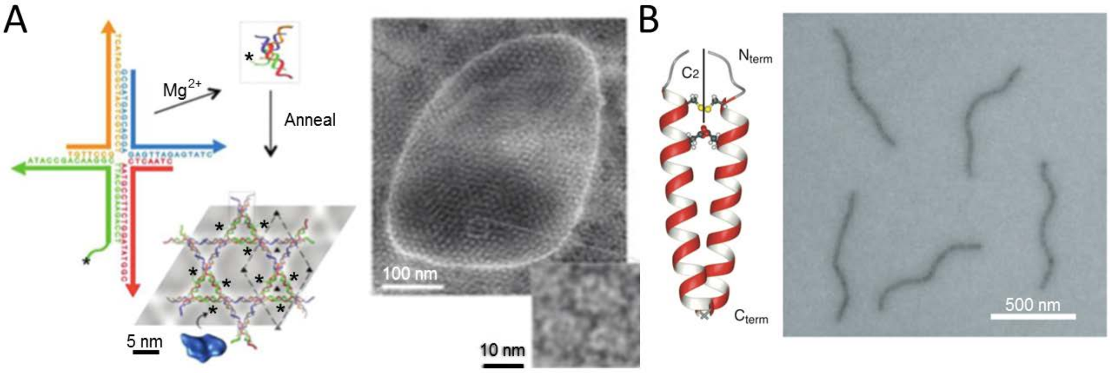


Figure 17. Structural analysis of proteins based on DNA nanostructures. (A) Four oligonucleotides form the Holliday junction that assemble the 2D lattice. * indicates the NTA modification that enables the arrangement of His-tagged proteins (in blue). Negative-stain electron micrographs of the lattice with the small G-protein Gα_{i1} bound to the rat neurotensin receptor type 1 (NTS1). (B) DNA nanotubes were used for NMR structure determination of membrane proteins. Structure of the ζ-ζ TM domain of the T cell receptor. Negative-stain electron micrographs of DNA-nanotube heterodimers. From Douglas et al., 2007; Selmi et al., 2011.

1.4 Challenges

The arrangement of proteins and nanoparticles has originated many applications in biomedicine. The current challenge is to find a good strategy for orthogonal conjugation of proteins to DNA nanostructures. In this thesis we present a site-specific non-chemical strategy for DNA coupling based on a superfamily of DNA modifying proteins: relaxases. The proposed enzyme-based method avoids fussy chemical labeling process, which might provide an ideal candidate for the generation of programmable arrangements of multiple proteins on DNA nanostructures.

2. DNA MODIFYING ENZYMES

In order to functionalize DNA nanostructures one obvious approach is the use of enzymes that naturally modify or process DNA such as DNA endonucleases. In fact HUH endonucleases have already been used for DNA bionanotechnology as described in section 1.2.2.3. These enzymes could be coupled to other proteins or domains as fusion proteins, for some of the applications highlighted above. In this section we cover the mechanism of action of different endonucleases found in nature.

2.1 HUH endonucleases

HUH-endonucleases constitute a protein superfamily that includes relaxases, replicases and transposases, involved in conjugation, rolling-circle replication and transposition, respectively (Chandler et al., 2013). They possess a conserved HUH motif and the Y motif, containing either a single conserved catalytic tyrosine (Y1) or more than one (Y2). They have a common mechanism of action by which they catalyze cleavage and rejoining of ssDNA using an active site Tyr residue to make a transient 5′-phosphotyrosine bond with the DNA substrate (Chandler et al., 2013).

2.1.1 Relaxases

Relaxases are HUH endonucleases involved in DNA processing for bacterial conjugation. They are proteins involved in very specific cleavage and strand transfer reactions. Thus we have chosen this family of proteins for the modification of DNA nanostructures.

2.1.1.1 Bacterial conjugation

In many clinically-relevant bacteria, the virulence genes and the resistance to antibiotics are carried by plasmids. Moreover, the genes responsible for some metabolic pathways in bacteria are also carried in plasmids, giving an adaptive advantage to the bacteria that possess these genes. Some of these plasmids are mobilized among bacteria by conjugation (Frost et al., 2005). Bacterial conjugation implies cell-to-cell contact in order to set up the process where one of the strands of the full-length plasmid DNA is transferred from the donor cell to the recipient cell by a type IV secretion system (T4SS) (Figure 18A). The T4SS is a multiprotein complex that spans the inner and outer bacterial membranes. Conjugative plasmids encode for all the proteins (mobilizable (MOB) machinery and T4SS). Mobilizable plasmids are not self-transmissible because they encode for the minimal MOB machinery (la Cruz et al., 2010). Therefore, they need the assistance of a conjugative plasmid within the same bacterium to transfer successfully through the T4SS.

The mechanism of conjugative DNA transfer can be divided into three different stages (Llosa et al., 2002): (1) assembly of the relaxosome within the origin of transfer (*oriT*), (2) cleavage reaction, (3) substrate recruitment and translocation to the recipient cell. The relaxase, the accessory proteins and IHF bound around the *oriT* assemble the nucleoprotein complex called relaxosome (stage 1, Figure 18: Initiation). This complex melts the dsDNA in order to localize the T-strand (the single strand destined for translocation) within the active centre of the relaxase. Then, the relaxase cleaves the sugar-phosphate backbone of the DNA at a specific position, named *nic* site, establishing a covalent bond with the DNA (stage 2). The coupling protein (T4CP) is recruited and links the relaxosome with the T4SS (Figure 18: Transfer). The relaxase bound to the T-strand is transferred in a 5’-3’ direction through the T4SS. The recircularization of the T strand within the recipient cell is catalyzed by either the same relaxase (relaxases with two catalytic tyrosines, named Y2) or other relaxase monomer (for those relaxases with one catalytic tyrosine or Y1) (Figure 18: Termination) (Carballeira et al., 2014). The complementary strand is simultaneously synthetized within the donor and the recipient cells. At the end of this process, both donor and recipient cells contain the conjugative plasmid.

Not only extrachromosomal plasmid can conjugate. The “integrative and conjugative elements” (ICEs) share the mechanism of self-transmissible plasmids. However, they should excise from the chromosome through the action of a recombinase/excisionase complex and generate a circular intermediate (Alvarez-Martinez and Christie, 2009). In the recipient cell, ICEs reintegrate into the chromosome.

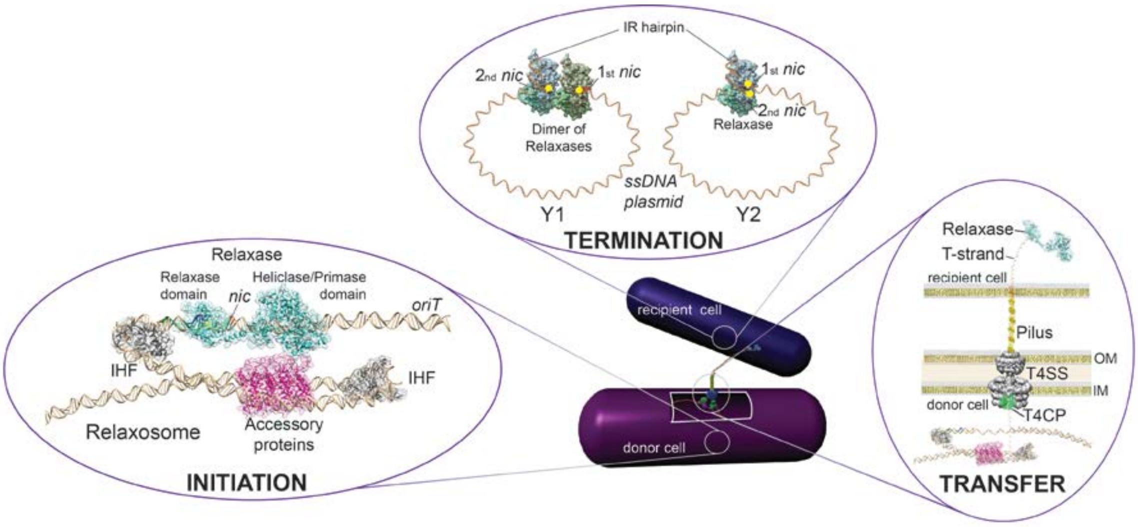


Figure 18. Schematic representation of bacterial conjugation. Molecular model of the Relaxase-DNA transfer through the T4SS during conjugation. See text for details. Adapted from (Moncalián et al., 1999; Cabezon and la Cruz, 2014; Carballeira et al., 2014). The PDBs used are 1OMH-TrwC, 1W36-RecBCD, 1IHF-IHF, 2BA3-NikA, 1GKI-TrwB and EMC2567-T4SS.

2.1.1.2 Origin of Transfer

The origin of transfer (*oriT*) is the DNA sequence required for a DNA molecule to be transmitted by conjugation. The *oriT* is located in an intergenic region that possesses several direct (DR) and inverted repeats (IR) that are binding sites for the relaxase and accessory proteins (RAPs). RAPs direct the binding of the relaxase to the proximal arm of an IR contiguous to the *nic*-site. The distal arm of this IR is not required for initiation, while it is needed for termination because the IR forms a hairpin loop structure at the recipient cell that directs the ligation step of the transferred strand (Pansegrau and Lanka, 1996; Lucas et al., 2010). The differences in length and structure of the *oriTs* of different systems are mainly dependent of the proteins required for *oriT* processing. The *oriT* sequence of RSF1010 plasmid is short with around 40 bp has only one IR, in contrast with R388, F or R46 *oriT*s that are complex and long: 330 bp, 292 bp and 141 bp, respectively (Thompson et al., 1984; Furuya and Komano, 1991; Llosa et al., 1991). Many groups have defined the minimal *oriT* that is sufficient for transfer some plasmids. For example, 38 bp are required for mobilization of plasmid R1162, 41 bp in ColE1 and 44 bp in R64 (Brasch and Meyer, 1987; Furuya and Komano, 2000; Varsaki et al., 2009).

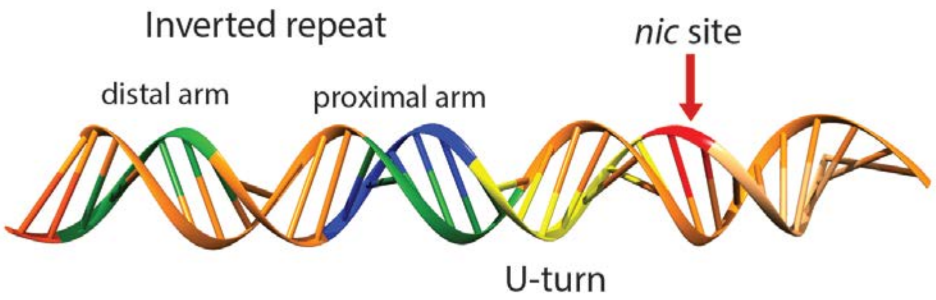


Figure 19. Scheme of the minimal *oriT* of conjugative plasmids. The proximal arm (in blue) of the IR forms a duplex in the plasmid dsDNA, and hybridizes with the distal arm (in green) when the target is ssDNA. The region in yellow forms a U-turn when bound to the relaxase. The scissile phosphate (*nic* site) is show in red.

2.1.1.3 Relaxases

The relaxase is responsible for initiation and termination of DNA processing during plasmid conjugation. Relaxases are ssDNA HUH endonucleases that form a covalent bond between the catalytic tyrosine and the 5’phosphate at the *nic* site during cleavage. These proteins are named relaxases because the nicked supercoiled plasmid is converted to the relaxed open circular form.

Relaxases are similar to rolling-circle replication (RCR) initiator proteins or Rep described in section 2.1.2 (Koonin and Ilyina, 1993). They possess three conserved motifs:

- Motif I: Contains the catalytic tyrosine(s) that cleave and rejoin DNA.
- Motif II: Conserved among MOB families, interacts with DNA.
- Motif III: Contains the HUH motif (Histidine-hydrophobic-Histidine) involved in the coordination of the divalent metal needed for cleavage.

Relaxases display in their structure a circular permutation of the primary sequence with respect to Rep replicases (Chandler et al., 2013). Therefore, HUH (motif III) and Y (motif I) change positions in the primary sequence, but their three dimensional structure is virtually identical (Figure 19).

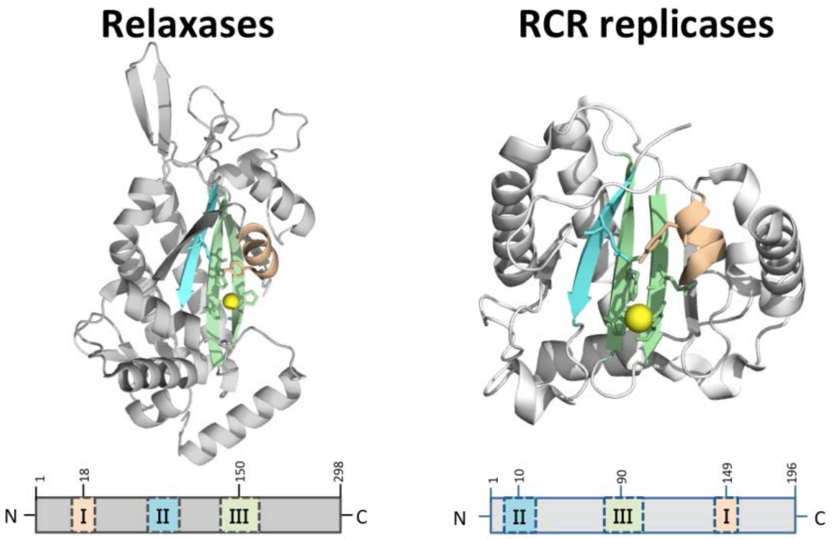


Figure 20. Comparison of the organization of conserved motifs of relaxases and RCR replicases. Both relaxases and RCR replicases (Rep) have a five−stranded β−sheet with a βαββ core that harbours the HUH motif on the central strand (light green). Due to the different domain organization in Rep proteins and relaxases, the position of the motif I, II and III differs. $TrwC_{R388}$ and Rep_{AAV5} are used as models. HUH residues and catalytic tyrosine are shown as sticks. The divalent metal ion is shown as a yellow sphere.

A classification scheme based on the 300 N-terminal amino acids of relaxases resulted in 6 families: MOB_F, MOB_H, MOB_Q, MOB_C, MOB_P and MOB_V (Francia et al., 2004; Garcillan-Barcia and Francia, 2009). Of the four HUH relaxase families, MOB_Q and MOB_P comprise relaxases with a single catalytic Tyr (Y1), whereas MOB_F groups enzymes with two catalytic Tyr (Y2), and MOB_V seems to have a catalytic histidine instead of a Tyr (Figure 21) (Francia et al., 2004). MOB_H and MOB_C have a different architecture: MOB_H family contain a HHH motif, a hydrolase motif and a conserved Y motif but MOB_C contain a conserved potential metal ion-binding D..E..E triad together with the Y motif that does not form a stable DNA-protein intermediate (Núñez and la Cruz, 2001; Chandler et al., 2013; Francia et al., 2013).

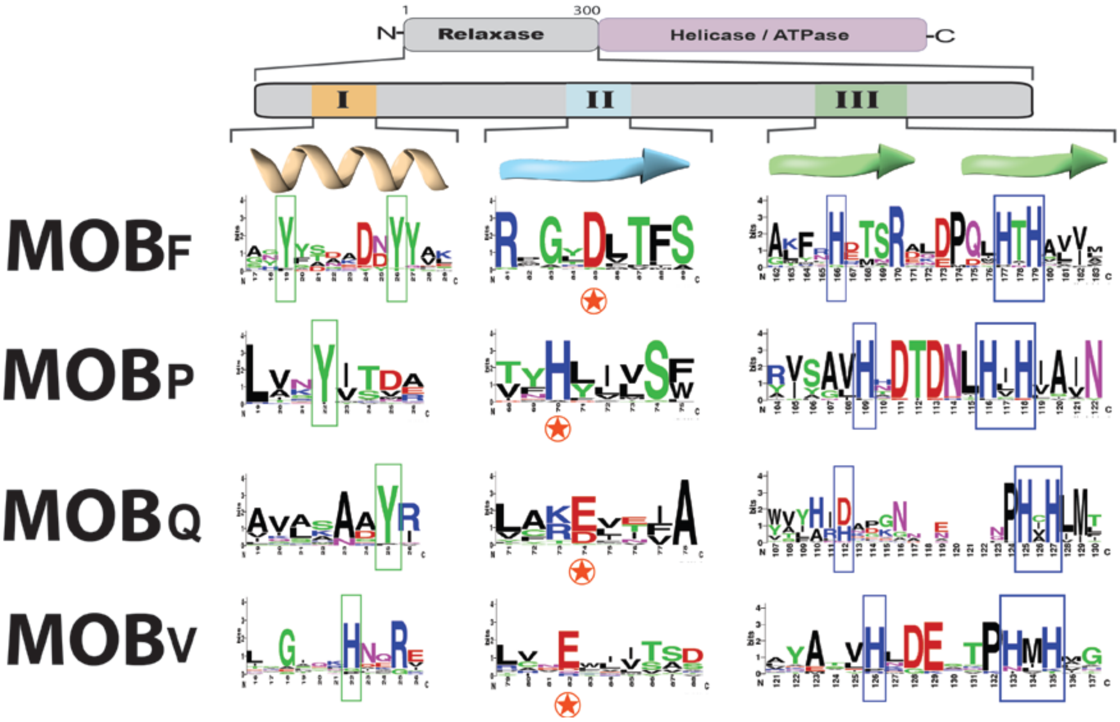


Figure 21. MOB WebLogo of the three conserved motifs of the four HUH-relaxase families. Relaxases are usually multidomain proteins, in which the relaxase domain always occupies the N-terminal position and the C-terminal holds a helicase, ATPase or a transfer related domain. The three conserved motifs of the relaxase domain consist of a α-helix (motif I), a β-sheet (motif II) and two antiparallel β sheets (motif III). The catalytic tyrosine(s)/histidine is highlighted with a green rectangle in motif I. The glutamic, aspartic or histidine residue involved in activation of the hydroxyl group of the catalytic residue is highlighted with a red star in motif II. The histidines H+HUH are highlighted with a blue square in motif III. Logos obtained by using WebLogo.

Relaxases are generally large proteins with two or more domains. The N-terminal displays the endonuclease activity and the C-terminal has a transfer-related function. MOB_Q family generally has a primase while MOB_F possesses a 5´-3´ helicase at the C-terminus. Moreover, domains involved in protein-protein interactions and other functions have been reported (Chandler et al., 2013).

The relaxase is transferred together with the DNA because relaxases possess translocation signals (TS). An assay termed the Cre reporter assay for translocation (CRAfT) confirmed that the relaxases $TraI_F$, $TrwC_{R388}$, $MobA_{RSF1010}$, $VirD2_{At}$ and $TraA_{pAT58}$ have a TS conserved within the MOB_F and MOB_Q families of relaxases (Parker and Meyer, 2007; van Kregten et al., 2009; Lang et al., 2010; Alperi et al., 2013). In this assay, Cre recombinase was fused to peptide fragments carrying suspected secretion signals. Translocation of the fusion protein to recipient cells is monitored by Cre recombination at *loxP* sites engineered into the target cells (Lang et al., 2010). TrwC$_{R388}$ has just one TS, whereas TraI$_F$ and MobA$_{RSF1010}$ have two TS. It has been proposed that the second recognition sequence can distinguish between the protein bound to DNA and the free-DNA protein (Lang et al., 2010). The crystal structure of the TSA domain of the TraI from R1 is part of a vestigial helicase (Redzej et al., 2013).

The relaxase domain of MobA$_{R1162}$ and TrwC$_{R388}$, as well as the entire NikB$_{R64}$ and TraX$_{pAD1}$ exhibited *oriT*-specific recombinase activity (Bhattacharjee et al., 1992; Furuya and Komano, 2003; César et al., 2006). Recombination requires the generation of a single-stranded region around *nic* and the host machinery. Moreover, TrwC$_{R388}$ mediates DNA integration that occurs in the absence of conjugation when the transferred DNA becomes dsDNA (Llosa et al., 1994; Draper et al., 2005). Therefore, it has been proposed TrwC$_{R388}$ as a potential tool for gene therapy (Agúndez et al., 2012).

2.1.1.4 Structure of Relaxases

The three dimensional structures of TrwC$_{R388}$, TraI$_F$, TraI$_{pUC1,}$ MobA$_{R1162,}$ NES$_{pLW1043}$ and MobM$_{pMV158}$ relaxases show high structural but low sequence similarity. Their structure consists of several α-helices that wrap around a five-stranded antiparallel β-sheet core. Motif II and III are located in the β-sheet core, whereas the catalytic tyrosine is placed on a nearby α-helix (Figure 20). The crystal structure of protein-DNA complexes reveals the DNA-binding cleft that directs the *nic* site to the active center. For its similarity in folding with DNA polymerases, MOB$_F$ relaxase resembles an open left hand, further divided into a palm, thumb and fingers domain. The palm is the antiparallel β-sheet core that participates in direct base-specific interactions with the ssDNA. The thumb is the α-helix that carries the catalytic tyrosine. The fingers are the C-terminal α-helical domain that becomes ordered upon binding to DNA (Larkin et al., 2005). Nevertheless, MOB$_Q$ and MOB$_V$ lack the fingers domain (Monzingo et al., 2007; Edwards et al., 2013).

2.1.1.5 Catalytic Mechanism of Relaxases

Relaxases are metal-dependent enzymes in which the two histidines from the HUH motif and usually another histidine from the motif III are involved in metal coordination. Mg^{2+}, Mn^{2+} or Ni^{2+} are considered physiological cofactors *in vivo*. The role of the metal is to interact with phosphate oxygens to orient the scissile phosphate relative to the catalytic tyrosine (Guasch et al., 2003). The hydroxyl group of the catalytic tyrosine acts as the nucleophile that attacks the phosphodiester bond. The transesterification reaction results in a covalent DNA-(5'-phosphotyrosyl)-relaxase intermediate and a free 3'OH (Figure 22). This reaction is isoenergetic and reversible: the energy of the cleaved phosphodiester bond is kept in the form of a phosphotyrosine linkage, which allows subsequent religation of the nicked DNA by the free 3'OH. During the strand-transfer reaction in the recipient cell, either a second tyrosine present in the same molecule (in Y2 relaxases) or in a different molecule (in Y1 relaxases) attacks the second *nic*-site, regenerating the *oriT*. In TrwC$_{R388}$, D85 is essential because the side chain establishes hydrogen bonds with the tyrosine hydroxyl group, activating the Y18 by proton abstraction (Guasch et al., 2003). The metal ion and R185 contribute to stabilize the penta-coordinated intermediate. However, the residue that provides the proton to oxygen O3'(T25) is not known (Boer et al., 2006).

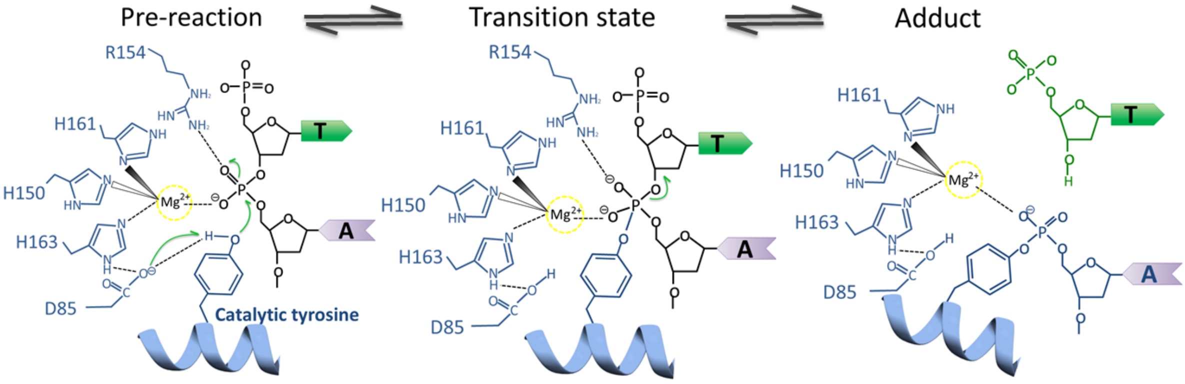


Figure 22. Schematic diagram illustrating the proposed catalytic mechanism for the DNA cleaving reaction of relaxases. TrwC$_{R388}$ active center has the triad of histidines coordinating the divalent cation. The attack of the catalytic tyrosine is promoted by abstraction of the proton. A pentacoordinated phosphorane transition state leads to the formation of a protein-DNA adduct. The relaxase becomes covalently attached to the 5'-phosphate of the DNA (in blue) and the 3'end primes replication of the plasmid during transfer (in green). Adapted from Boer et al., 2006.

2.1.1.6 Phosphorothiolate oligonucleotides

Phosphorothiolate or “suicide substrates” have been employed to separate the DNA cleavage and ligation reactions and study them independently for most of the DNA processing enzymes. Phosphorothiolate oligonucleotides are synthetic DNAs in which sulfur replaces a specific 5′ or 3′ bridging oxygen at the site of strand cleavage and religation. Cleavage of substrates that contain a 5′-bridging phosphorothiolate generates a 3′-terminal –SH group (Figure 23). Sulphur is a soft nucleophile, so attack of the phosphotyrosine linkage is inhibited. Thus, DNA-protein covalent complexes accumulate because religation is blocked. TrwC$_{R388}$ reached 100% of covalent complexes with suicide oligonucleotides (González-Pérez et al., 2007). These oligonucleotides have been used with topoisomerase II to study the critical interactions of the divalent cations and also to resolve the structure of the cleavage complex (Deweese et al., 2008a; Wendorff et al., 2012).

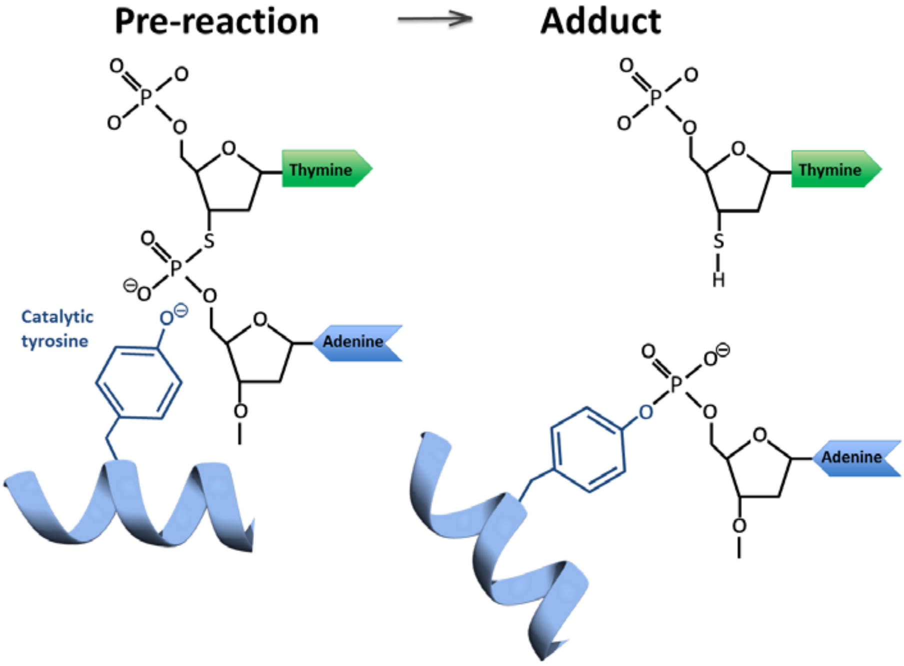


Figure 23. Phosphorothiolate oligonucleotides. The 5’ bridging oxygen has been substituted by sulphur. Upon cleavage, the religation reaction is inhibited.

2.1.1.7 Y2 Relaxases

2.1.1.7.1 MOB_F family

- **TrwC_R388**

TrwC_R388 is a large protein (966 residues) with a relaxase and helicase domain. It is a Y2 enzyme, in which Y18 specifically catalyzes the initial cleavage reaction, while Y26 performs the strand-transfer reaction which ends the transfer process (Grandoso et al., 2000). Cleavage is controlled by TrwC-DNA intermolecular interactions and by intramolecular DNA interactions in the U-turn (Carballeira et al., 2014). The β4-β5 ribbon interacts with the major grove of the dsDNA hairpin, moreover residues along β1, β3 and α6 interact with the ssDNA (Guasch et al., 2003). The K_D of the relaxase domain (N293) for the target oligonucleotide W(35+8) is 34 nM (Lucas et al., 2010). Mutations on the nucleotides that allow formation of the U-turn affect conjugation efficiency (Carballeira et al., 2014). Interestingly, mutations in Lys262 changed the specificity from C24 to G24 at the *nic* site (González-Pérez et al., 2009). TrwC helicase domain is similar to RecD2 from *Deinococcus radiodurans* (Saikrishnan et al., 2008; Chandler et al., 2013). It belongs to superfamily 1B helicases (SF1B) that translocate DNA in 5'-3' direction. TrwC_R388 catalyzes site-specific recombination between two *oriT* sequences and also site-specific integration of the transferred DNA strand into an *oriT* copy present in the recipient cell (César et al., 2006). TrwC is transported to the recipient cell in the absence of DNA transfer (Draper et al., 2005).

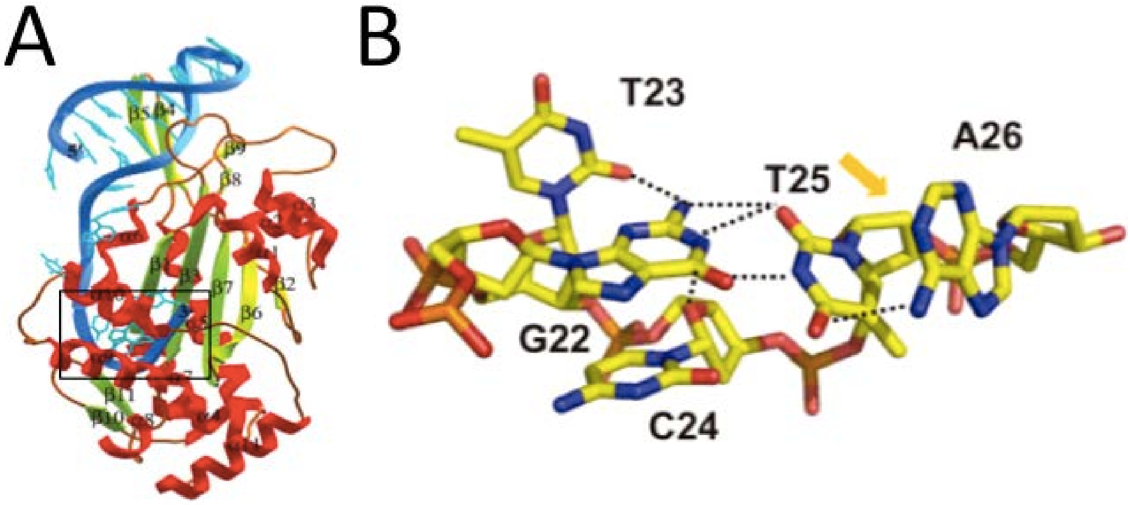


Figure 24. Intermolecular (A) and intramolecular interactions (B) in TrwC$_{R388}$. (A) Structure of the relaxase domain of TrwC$_{R388}$ bound to a 23-nt *oriT*-containing hairpin. The protein is made up of 11 β-strands (yellow), 11 α-helix (red) and 7 turns (orange). (B) Enlargement of the area within the square in (A) that shows the U-shaped turn in the structure of TrwC$_{R388}$:W(25+2) complex. The position of the nic-cleavage site is highlighted by an orange arrow. Taken from Guasch et al., 2003; Carballeira et al., 2014.

• TraI$_F$

TraI$_F$ is a large monomeric protein (1756 residues) with four domains: a 300 residue relaxase domain, followed by a ssDNA-binding domain, a helicase domain and a domain at the C-terminus that contains an essential proline-rich loop and coiled-coils involved in protein-protein contacts (Clark et al., 2014). There is negative cooperativity in ssDNA binding between the relaxase and the ssDNA-binding domain (Sut et al., 2009). This means that protein interaction with the ssDNA adjacent to the *nic* site after cleavage switches to a helicase active state, able to bind and transfer ssDNA. TraI$_F$ is a Y2 relaxase in which Y16 is the catalytic tyrosine, but there is no evidence that Y23, which corresponds to Y26 in TrwC, is also a nucleophile active in transfer (Dostal et al., 2011). It has been suggested that a second molecule of TraI acts in the recipient cell. The relaxase domain, TraI$_{R\_F}$, binds a single-stranded *nic* sequence in a knob-into-hole fashion with a K_D of 0.55 nM (Harley and Schildbach, 2003). Single-base substitutions can reduce binding affinity by between 10- and 10,000-fold.

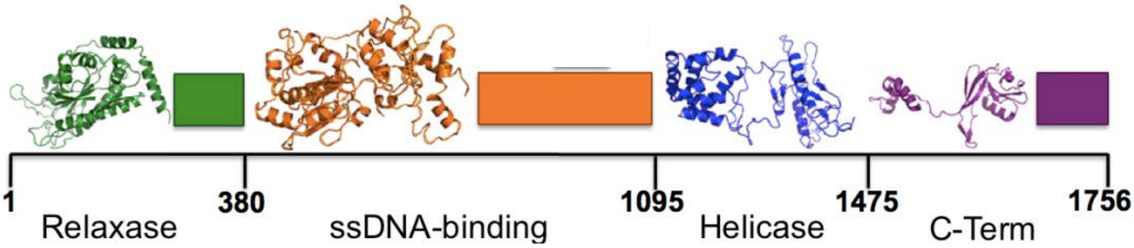


Figure 25. Domain structure of TraI$_F$. TraI consists of four domains: a relaxase domain, ssDNA-binding domain, a helicase domain and a C-terminal domain. Parts of the structure of TraI have been solved (ribbon representation) but other sections lack structural data (depicted by boxes). The model uses PDB: 1P4D, 4LOJ, recD, 3FLD. Modified from Clark et al., 2014.

- **TraI$_{R100}$**

TraI$_{R100}$ is highly similar (91% identity) to TraI$_F$ and its *nic*-site differs only by two nucleotides. The relaxase domain, TraI$_{R\_R100}$, binds *nic*$_{R100}$ oligonucleotide with a K_D= 2.4 nM but *nic*$_F$ with K_D= 3.2 μM (Harley and Schildbach, 2003). Two amino acid substitutions in the DNA binding cleft (R193Q/ Q201R) change TraI$_{R\_R100}$ DNA-binding specificity to that of TraI$_{R\_F}$. But swapping six amino acids in TraI$_{R\_F}$ (M2L / I4F / E153D /Y190L / Q193R / R201Q) resulted in the same binding specificity from TraI$_{R\_R100}$ to *nic*$_{R100}$ (Guja and Schildbach, 2015). These residues are involved in the pocket formation of some knob-into-hole interactions that play a central role in determining the specificity of TraI$_{R\_F}$.

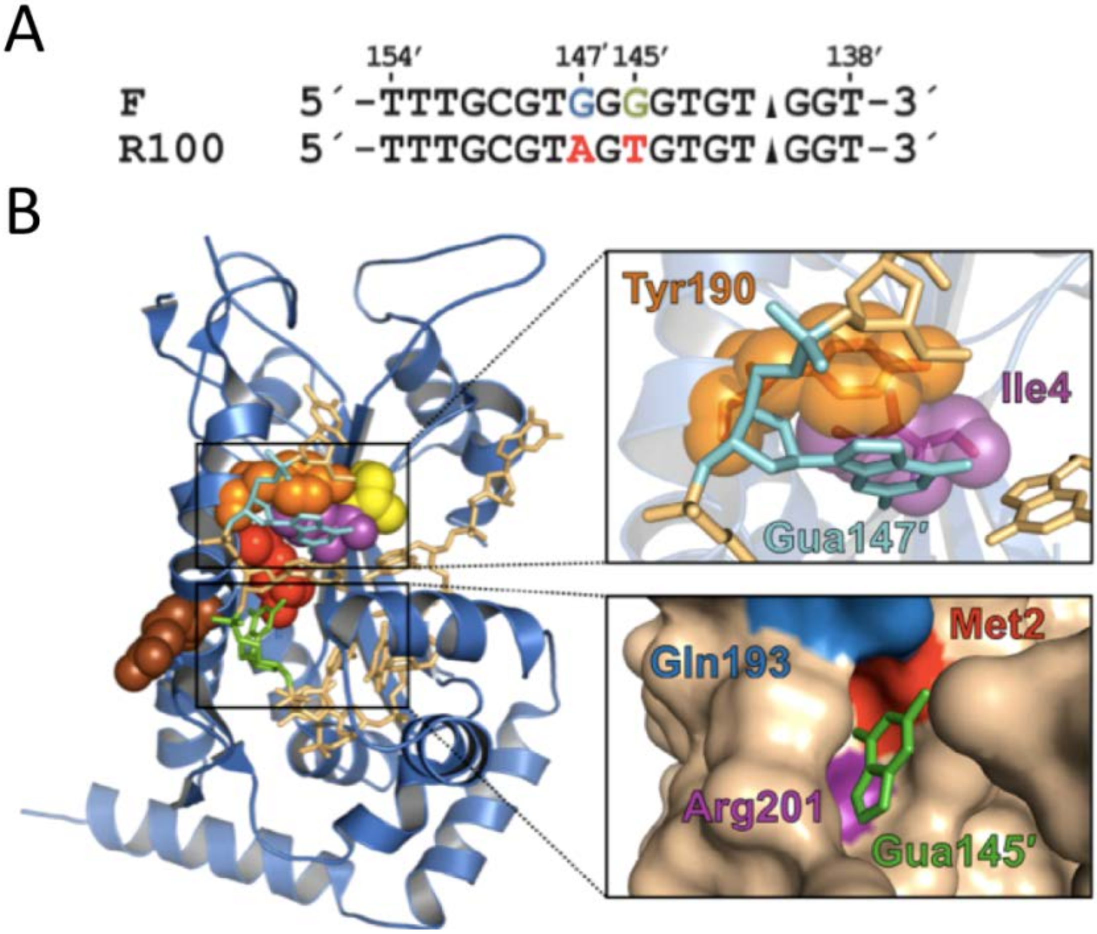


Figure 26. TraI$_{R100}$. (A) Alignment of the DNA *nic* sequences of F and R100. The *nic* site is indicated by an arrowhead. (B) Cartoon representation of TraI$_{R\_F}$ with residues mutated shown as space-filling models. Met2 in red, Ile4 in purple, Ala5 in yellow, Tyr190 in orange, and Arg198 in brown. The *oriT* oligonucleotide is shown in beige, while bases Gua145′ and Gua147′ are shown as green and cyan stick models, respectively. From Guja and Schildbach, 2015.

2.1.1.7.1.4 TraI$_{pCU1}$

TraI$_{pCU1}$ is 1078-residue bifunctional Y2 protein with both relaxase and helicase activities, like TrwC$_{R388}$ and TraI$_{F}$. The entire length shares 99,8% identity with TraI$_{pKM101}$ and it is 43% identical to TrwC (Nash et al., 2010). Motif I has four-tyrosine residues that are structurally mobile and functionally redundant. TraI$_{R\_pCU1}$ binds DNA in a weak sequence-independent manner, with K$_{D}$ from 700 nM to 1,2 μM. It has been suggested that it requires the accessory protein TraK$_{pCU1}$ to initiate sequence-specific DNA binding to *oriT* (Nash et al., 2010). Interestingly, the U-turn bases establish several contacts with the relaxase and regulate cleavage activity.

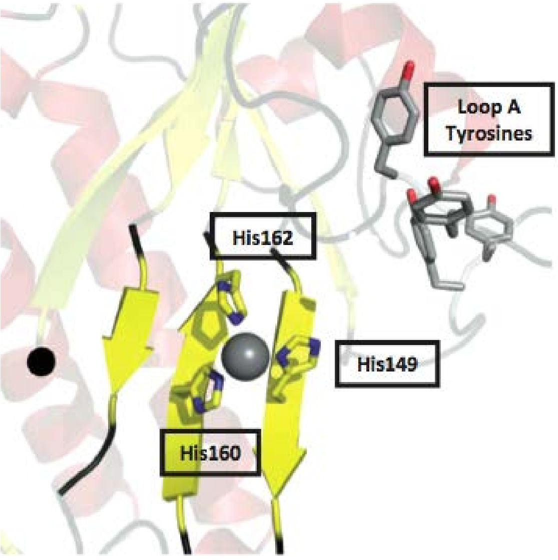


Figure 27. Active site of TraI$_{pCU1}$. The four-tyrosine residues are at positions 18, 19, 26, and 27 (Loop A). The histidine triad of the active site and the tyrosines implicated in DNA nicking are shown as sticks. The manganese ion is shown as a gray sphere. PDB 3L57. From Nash et al., 2010.

2.1.1.8 Y1 Relaxases

2.1.1.8.1 MOB$_{Q}$ family

- **MobA$_{RSF1010-R1162}$**

The mobilizable plasmid R1162 encodes a large relaxase MobA, of 709 residues, virtually identical to RSF1010 relaxase. The first 186 residues encode the Y1 relaxase domain (Bhattacharjee and Meyer, 1993). Then, the termination reaction is carried out by another MobA$_{R1162}$ molecule that enters the recipient as part of the general flux of DNA-free relaxases secreted by the donor during plasmid transfer (Becker and Meyer, 2012). The C-terminal domain is a primase (residues 388–709) required for vegetative replication of the plasmid, which is also translated separately as the small RepB' (Scherzinger et al., 1992; Perwez and Meyer, 1996). Crystallographic and biophysical studies revealed that, apart from the active centre, there were two more Mn^{2+} binding sites of lower affinities that are required for DNA nicking activity (Monzingo et al., 2007). $MobA_{R1162}$ mediates *oriT* recombination of bacteriophage M13 derivatives that contain two directly repeated copies of *oriT* (Bhattacharjee and Meyer, 1993). $MobB_{R1162}$ stimulates R1162 transfer because it interacts with the TS on $MobA_{R1162}$ to be successfully transferred by a number of different T4SS (Parker and Meyer, 2007).

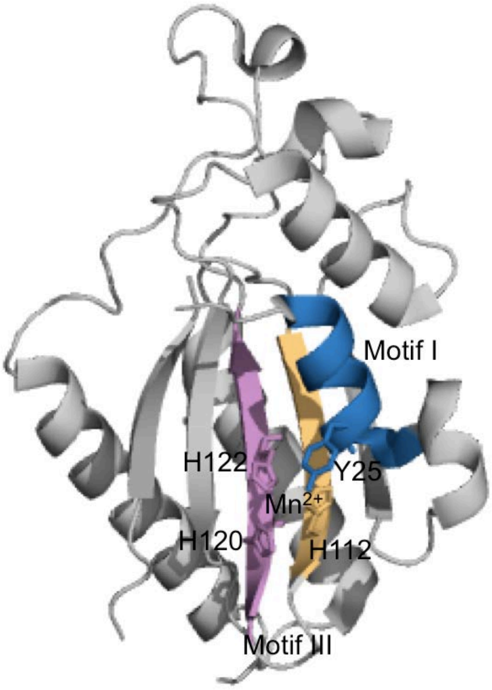


Figure 28. $MobA_{RSF1010}$. Crystal structure of the MobA N-terminal relaxase (1–186) with motifs I and III highlighted in blue, orange and magenta respectively. Catalytic residues are shown as sticks. The active site Mn2+ atom is also shown. PDB 4HT4. From Monzingo et al., 2007.

- **$NES_{pLW1043}$**

The 665-residue NES protein is responsible of the transfer of antibiotic resistance plasmids into strains of methicillin-resistant *S. aureus* (MRSA). Cleavage by this Y1 relaxase is dependent of a number of nucleotides upstream the *nic* site in the target sequence (Edwards et al., 2013). Two essential loops are involved in specific interactions with the DNA hairpin. Deletion of these loops traps the enzyme into a DNA-relaxase complex, but results in unspecific binding and eliminates conjugative DNA transfer. Biochemical and SAXS-based structural data revealed that the C-terminal domain of NES regulates the cleavage-rejoining activity of the full-length relaxase and is required for conjugation.

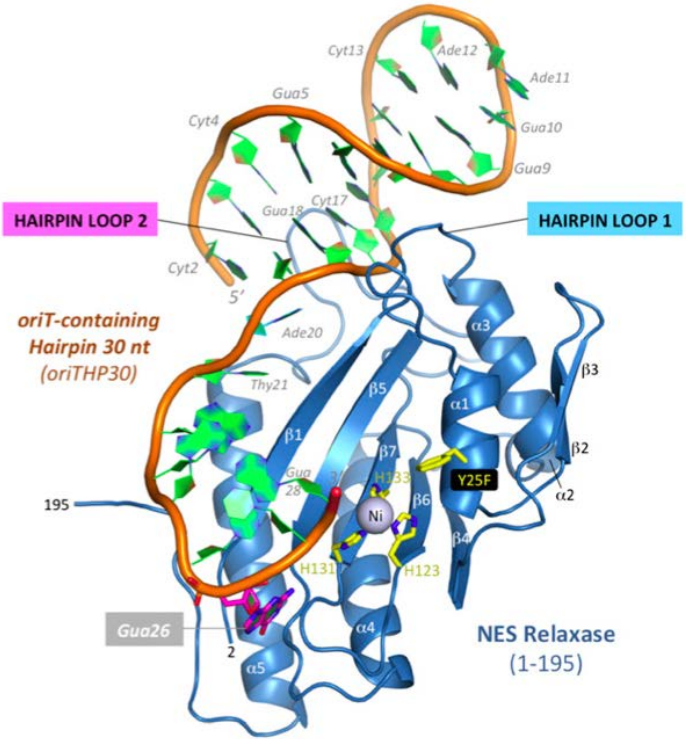


Figure 29. Nicking enzyme in *S.aureus* (NES). A) Crystal structure of the NES N-terminal relaxase (1–195) bound to a 30-nt *oriT*-containing hairpin showing the minor groove-binding hairpin loop 1 and the major groove-binding hairpin loop 2. PDB 4HT4. From Edwards et al., 2013.

- **TraA$_{pIP501}$**

The gram-positive pIP501 plasmid encodes for the relaxase TraA. The full length protein (654 aa) and the N-terminal domain (246 aa) bind a 42-mer containing the IR with a K_D of 55 nM and 26 nM, respectively (Kopec et al., 2005). However the truncated protein is less efficient in relaxing supercoiled DNA that contains the *oriT* sequence. Since the *oriT* overlaps with the promoter of the *tra* operon, the relaxase TraA acts as a repressor of the 15 genes that encodes for the T4SS (Kurenbach et al., 2006). This regulation appears in other G+ plasmids as a mechanism to optimize the energy-consuming process of plasmid transfer.

2.1.1.8.2 MOB$_P$ family

- **TraI$_{RP4}$**

TraI$_{RP4}$ was biochemically characterized in the 90s at Lanka laboratory and it is the relaxase representative of the MOB$_P$ family (Pansegrau et al., 1990). The 732-residue TraI$_{RP4}$ is a Y1 relaxase, in which histidines H70 and H116 and the aspartic D111 are essential residues for cleavage (Pansegrau and Lanka, 1996). Interestingly, the 3' end downstream of the *nic* site and Mg^{2+} are required for the correct organization of the active center of the relaxase. The structure prediction of $TraI_{RP4}$ relaxase domain, obtained by using the software $Phyre^2$ and MobA_R1162 as a template, is shown in Figure 30.

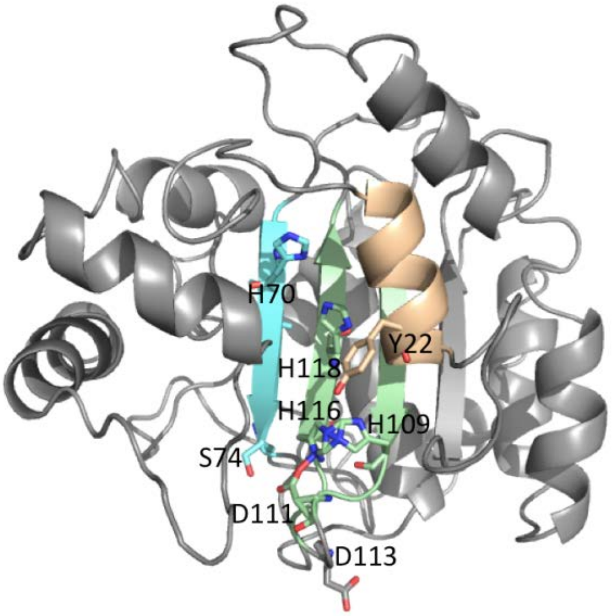


Figure 30. Cartoon model of $TraI_{RP4}$ N-terminal domain. The residues that were mutated in the study carried out by Pansegrau and Lanka, 1996, are shown as sticks and labelled. This residues are involved in substrate binding and catalitic activity of $TraI_{RP4}$. Created with the server $Phyre^2$.

- ## $VirD2_{Ti}$

The Tumor-inducing (Ti)-plasmid from *A. tumefaciens* encodes the VirD2 relaxase (424 residues) that guides the transfer of the oncogenic T-DNA into plant cell (Mysore et al., 1998). The VirD2-T-strand complex targets the nucleus where it is integrated into the genome. VirD2 contains a poorly conserved domain of unknown function (DUF) that contains the TS, whose sequence is not known (van Kregten et al., 2009). The C-terminal has the ω sequence that is important for T-DNA transfer and for integration (Mysore et al., 1998).

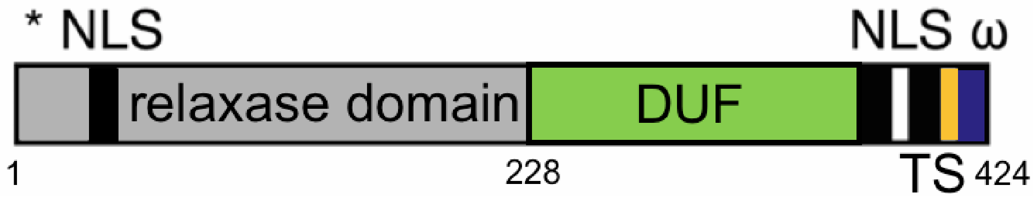


Figure 31. Domains of the VirD2 relaxase. VirD2 consists of three domains: the relaxase domain (in gray), a domain of unknown function (DUF, in green) and a ω domain (in blue). The catalytic tyrosine is indicated by *. NLS, nuclear localization signal, are indicated by black boxes. TS, the translocation signal through the T4SS, is indicated by an orange box. Modified from (van Kregten et al., 2009)

2.1.1.8.3 MOB$_V$ family

- **MobM$_{pMV158}$**

The streptococcal pMV158 plasmid encodes the relaxase MobM (494 aa), the representative relaxase of the MOB$_V$ family (Francia et al., 2004; Garcillan-Barcia and Francia, 2009). A conserved histidine at motif I (HxxR) is thought to carry out the nucleophilic attack to the DNA (Pluta et al., unpublished work). MobM forms stable complexes with the relaxed plasmid molecules(Fernández-López et al., 2013). MobM binds specifically to IR$_3$ during initiation of conjugation and to IR$_2$ in termination in the recipient cell . The K_D for IR$_3$ is 60±7 nM, but decreases to >320 nM when it lacks the distal arm (Lorenzo-Diaz et al., 2011). The C-terminal domain of MobM has a leucine-zipper motif that mediates dimerization, and a coiled-coil region that is involved in cell membrane association (Fernández-López et al., 2013).

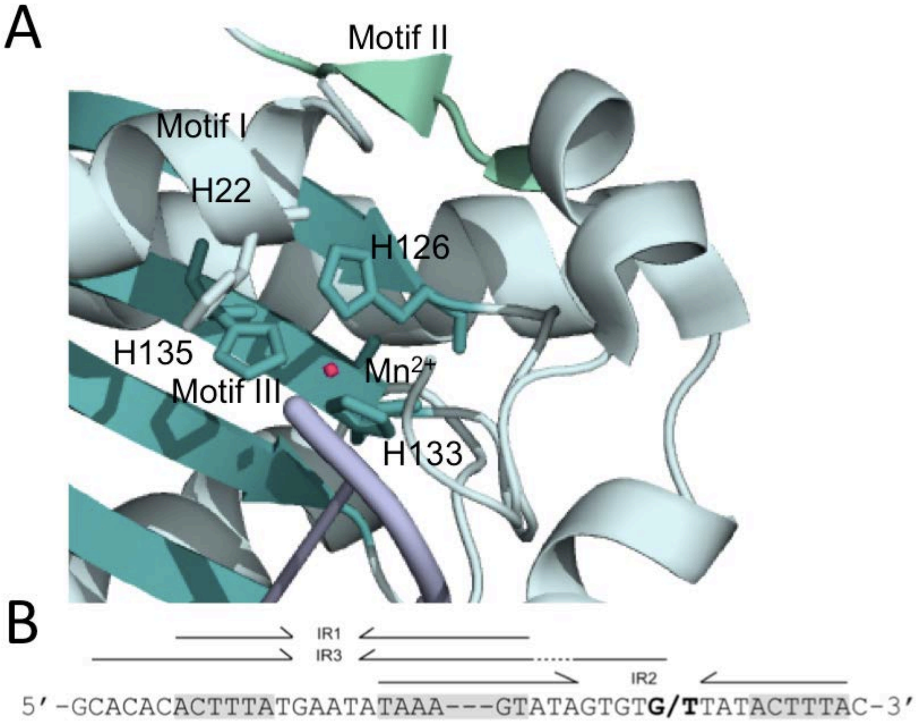


Figure 32. MobM$_{pMV158}$. (A) Active center. The conserved motifs are depicted as ribbon and catalytic residues are shown as sticks. The magnesium ion is shown as a red sphere. PDB: 4LVL. (B) *oriT* sequence of pMV158 with the overlaping inverted repeats highlighted in gray. The *nic* site is indicated with a /. From (Fernández-López et al., 2013).

2.1.2 Rolling circle replicases

Plasmids and many viruses use rolling circle replication (RCR) to generate multiple copies of their circular ssDNA genomes (Ilyina and Koonin, 1992; Khan, 1997). Like in relaxases, the RCR-initiator proteins (Rep) were classified in two superfamilies, Y1 and Y2 Reps depending of the number of catalytic tyrosines (Koonin and Ilyina, 1993; Chandler et al., 2013).

2.1.2.1 Y2 Reps

This group includes the Rep proteins of φX174, P2-related dsDNA bacteriophages and several cyanobacterial and archaebacterial plasmids. Y2 proteins share the cleavage-ligation mechanism proposed for φX174 gpA protein (Koonin and Ilyina, 1993). gpA acts by a flip-flop mechanism in which the two catalytic tyrosines (Y343 and Y347) alternate during the initiation and termination reactions (Hanai and Wang, 1993; Odegrip and Haggård-Ljungquist, 2001). The details of how gpA recognizes and binds the target DNA are unknown (Chandler et al., 2013).

During RCR, Y2 Rep binds to a tandem direct repeat sequence (*dso*) and initiates the replication by cleaving the *nic* site within a hairpin structure that generates a free 3'-OH end (Figure 33) (Khan, 1997; Boer et al., 2009). This leads to a covalent complex between the Rep and the 5'-DNA that will be displaced during replication. After one round of replication, Rep cleaves the newly synthesized *nic* site using the second catalytic tyrosine. The first phosphotyrosine linkage is attacked by the new 3'-OH, generating a free ssDNA circle.

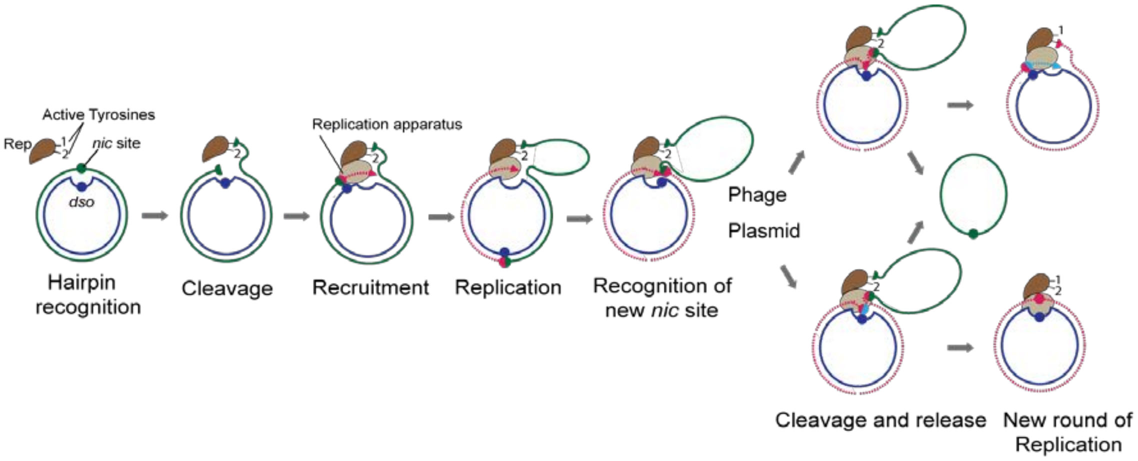


Figure 33. Rolling-circle replication mechanism of Y2. The Rep protein contains two active-site Tyr residues, 1 and 2, and recognizes circular dsDNA (dso). Rep cleaves the DNA at the *nic* site and forms a phosphotyrosine bond with Tyr residue 1. Then, the replication apparatus is recruited and replication is initiated. When the entire circle of DNA has been replicated, Rep and the replication apparatus reach the reconstituted *nic* site. In the case of phages, the *nic* site is cleaved by the Tyr residue 2 to form a new phosphotyrosine. The resulting 3′-OH attacks the original phosphotyrosine bond (with Tyr residue 1) to release a single-stranded circular phage copy. Replication of the Rep-bound circular DNA then continues rounding the circle again. In the case of copy number-regulated plasmids, no second phosphotyrosine bond is formed, and recircularization is addressed through the use of a second nucleophile (such as water) instead of a Tyr. Positive-sense strand in g reen, negative-sense strand in blue, newly synthetized strand in red. Modified from (Chandler et al., 2013)

2.1.2.2 Y1 Reps

This group includes Rep proteins of ssDNA plasmids replicating mainly in gram-positive bacteria, plant geminiviruses and animal parvoviruses (Koonin and Ilyina, 1993). RepA of plasmid pC194 from *S. aureus* contains a conserved tyrosine for initiation and a conserved glutamate residue that catalyze the release of the protein from the intermediate at the termination step (Noirot-Gros et al., 1994). The glutamate E210 directs hydrolytic cleavage of the newly synthesized origin and the resulting 3’-OH group undergoes transesterification with the phosphotyrosine link (Figure 33). Interestingly, replacement of glutamate with tyrosine in RepA, changes the termination reaction so that it becomes coupled to reinitiation, as in φX174 (Noirot-Gros and Ehrlich, 1996). Therefore, in pC194 termination is uncoupled from reinitiation of replication and thus allows regulation of plasmid copy number. In contrast, reinitiation in plasmid pT181 is prevented by inactivation of Rep through the formation of a covalent adduct with a short oligonucleotide (Zhao et al., 1998).

The rolling hairpin replication is a linear adaptation of the RCR used by parvovirus to replicate their linear ssDNA with terminal hairpins (Figure 34) (Hickman et al., 2002). The non-pathogenic human Adeno-associated virus (AAV) contains two inverted terminal repeats (ITR) that are the origins of replication. Within each ITR there are several DR, named the Rep binding sequence (RBS) and a terminal resolution site (trs) where Rep cleaves the stem loop generated by the C-terminal Rep helicase (Hickman et al., 2004). Rep requires a glutamic residue together with the HUH motif to ligate the metal (Figure 34B). In AAV, the residue that serve as a general base to deprotonate the catalytic tyrosine has not been identified but the Nε of Lys214 may serve to deprotonate the hydroxyl group of Tyr210 in the non-structural protein 1 (NS1) from the parvovirus minute virus of mice (MVM) (Figure 34C) (Tewary et al., 2015). This mechanism differs from that based on the acidic base found in TrwC$_{R388}$ and TraI$_F$

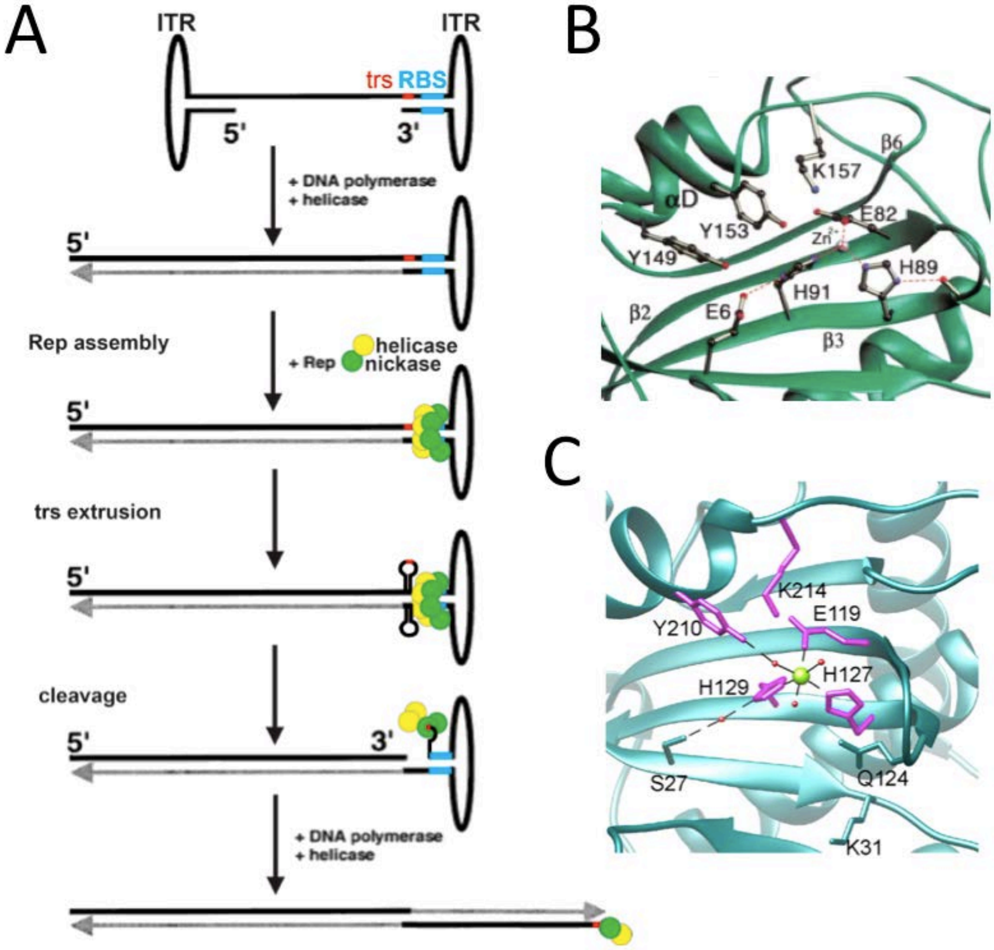


Figure 34. Rolling hairpin replication and Y1 Reps. (A) AAV replication. N-terminal domains of Rep (green spheres) bind to the RBS and leads to the

assembly of the helicase domains (yellow spheres) close to the trs site. The helicase activity generates a stem-loop at the trs (red) that is cleaved by Rep. A phosphotyrosine bond between Rep and the DNA is generated. Replication is initiated using the host replication machinery and the 3′-OH at the trs. (B) The catalytic residues of AAV5 Rep are shown as sticks and Zinc ion is shown in silver. C) Active site of the NS1N Rep. Waters are shown as red spheres and H-bonds as dashed lines. Adapted from (Hickman et al., 2002; 2004; Tewary et al., 2015).

2.1.3 Transposases

DNA transposases are enzymes that catalyze the movement of DNA transposable elements or transposons from one location in the genome to a new site. The transposases TnpA of IS*91*, IS*605*, IS*200*, IS*Hp608*, IS*CR*s and some helitrons belong to the family of HUH endonucleases (2015). ISCRs are Insertion Sequence Common Regions similar to IS*91* that are associated with antibiotic resistance genes. The helitrons are transposons found in eukaryotes. The best understood transposases are the TnpA of IS*605*/IS*200*, which are Y1 enzymes that form dimers (Chandler et al., 2013). The active sites of these enzymes adopt two functionally important conformations: the *trans* configuration, in which each active site is composed of the HUH motif from one monomer and the Tyr residue on an α-helix (αD) from the other monomer, and the *cis* configuration, in which both motifs in an active site are contributed by the same monomer (Figure 35). The model claims that, when two dimers cleave their ssDNA target sequence, there is a conformational change from *trans* to *cis* that allows exchange of the 5′ ends between subunits. The following attack of the two free 3′ ends on the active site phosphotyrosines, leads to the generation of an excised single-stranded circle that is inserted in the new location by a *cis* to *trans* conformational change.

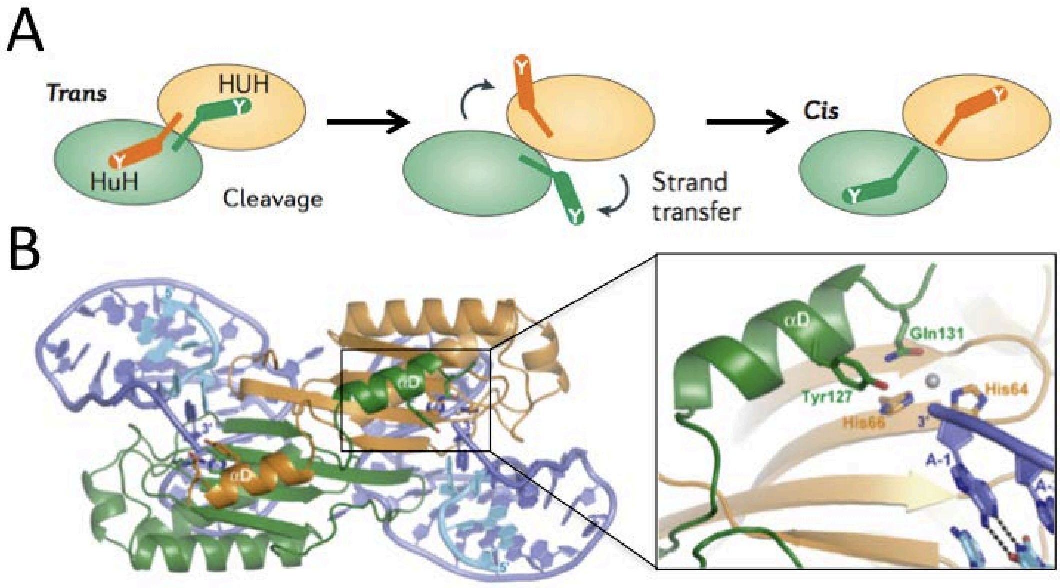


Figure 35. Mechanism of HUH transposition and structure of TnpA of IS*605*/IS*200*. (A) *trans*-*cis* conformational change during the transposition cycle. The active sites adopt a *trans* configuration for binding and cleavage of the left and right ends of the insertion sequence. In the *trans* configuration, the HUH motif of an active site is contributed by one TnpA monomer and the Tyr residue (Y), situated on an α−helix (αD), is contributed by the other. Following DNA cleavage, the two αD−helices rotate to the *cis* configuration. In *cis* configuration, the αD−helices then catalyze the formation of a circular single-stranded transposon and join the cleaved ends of the donor sequence. The configuration is then reset to the *trans* form to end excision and return to a *cis* configuration to integrate the transposon. (B) TnpA *trans* structure. The cartoon representation of the TnpA/RE34 complex crystal structure corresponds to the model in (A). The HUH motif and the catalytic tyrosine are shown as sticks. Inset, active site. Q131 together with the HUH motif coordinates the Mn^{2+} (gray sphere). Adapted from Barabas et al., 2008; Chandler et al., 2013.

2.2 Topoisomerases

DNA topoisomerases resolve tangles, knots and torsional stress generated during replication, transcription and other biological processes. DNA cleavage by topoisomerases results in the formation of a transient covalent bond between the catalytic tyrosine and one of the ends of the broken strand. The monomeric topoisomerases that cleave one strand of the DNA are defined as type I. They are further classified as either type IA topoisomerases if the protein is attached to 5′ phosphate or type IB and IC subfamily members if the protein links to a 3′ phosphate (Vos et al., 2011). In contrast, type II topoisomerases are dimeric enzymes that catalyze ATP-dependent DNA strand passage through the transient double-stranded break they generated on another DNA segment (Huang and Lin, 2015). Both type IIA and IIB subfamilies of topoisomerases form 5′-phosphotyrosyl bonds.

The best-characterized IIA type is the human topoisomerase Iiα (Deweese et al., 2008b; Wendorff et al., 2012; Lee et al., 2013). It has a conserved DXD motif involved in metal binding and a third invariant residue, Glu449, which establishes hydrogen bonds with the 3′ OH of the DNA strand. It has been proposed that strand scission relies on a variation of the canonical “two-metal mechanism” used by DD(E/D) enzymes (see 2.8.4 and Figure 37). In this scheme (Figure 36A), metal A actively assists during DNA cleavage and religation and is needed to stabilize the transition state together with Arg781 (Schmidt et al., 2010). Metal B and His736 play a role in anchoring the DNA at the -1 phosphate position. Interestingly, Type IA topoisomerases do not require a metal ion to cleave the DNA (Changela et al., 2001). Lys-8 and Arg-330, which coordinate the scissile phosphate from opposite sides of the active site, together with the nucleophilic tyrosine are sufficient for cleavage (Figure 36B). Lys-8 plays an important role in catalysis by contributing to the proper positioning of the scissile phosphate or to transition state stabilization as is shown in B.

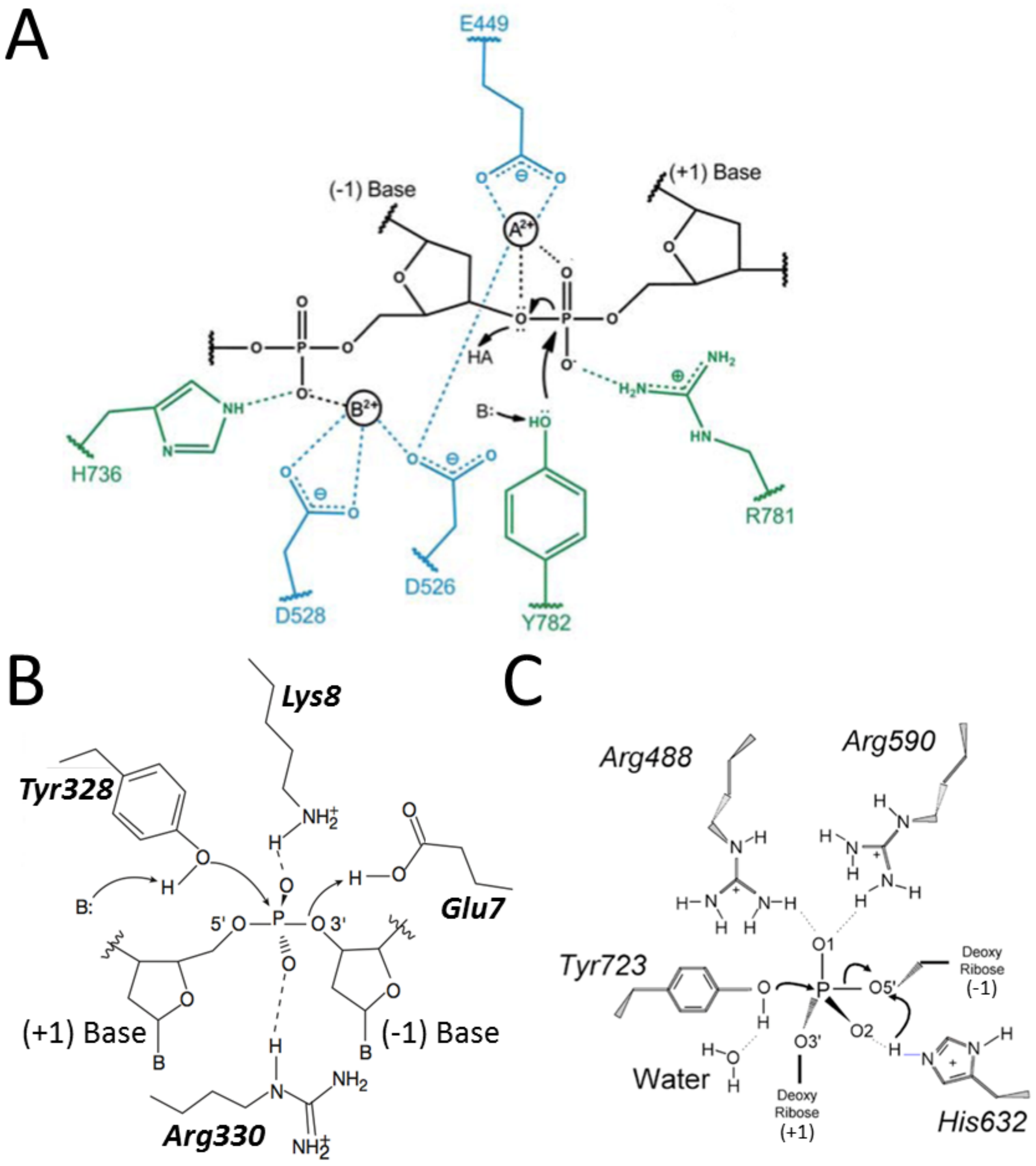


Figure 36. Proposed cleavage mechanism of topisomerases. Schematic diagrams illustrating a possible mechanism of DNA cleavage used by (A) Type II topoisomerases (human IIα). (B) Type IA (*E.coli* Topoisomerase III) . (C) Type IB (human topo I). The general base (B:) abstracts the proton from the tyrosine activating it for nucleophilic attack. The general base (B:) and acid (HA) are unknown, but may be waters. Metal A and B are shown as spheres. Hydrogen bonds are depicted as dotted lines. Adapted from Stewart, 1998; Changela et al., 2001; Schmidt et al., 2010.

The active-site architecture of type IB topoisomerases is evolutionarily related to that of tyrosine recombinases and integrases (Vos et al., 2011). Type IB topoisomerases do not require any metal cofactor for the nucleophilic attack and subsequent covalent attachment of the catalytic tyrosine to the 3’ end of the scissile phosphate (Redinbo et al., 1998). In the crystal structures of human topoisomerase I, the conserved residues Arg488, Arg590 and His632 interact with the nonbridging oxygens and contribute to stabilization of the pentavalent coordination state (Figure 36C) (Stewart, 1998). The general base that deprotonates the hydroxyl of Tyr723 is not known. His632 might function also as a general acid by donating a proton to the 5'-leaving group during the cleavage reaction.

2.3. Recombinases and integrases

A site-specific recombination process is responsible for integration and excision of bacteriophage genomes in and out of their bacterial host chromosomes. The tyrosine recombinase (YR) family includes bacteriophage λ-integrase, bacteriophage P1 Cre recombinase, the Flp recombinase from yeast, the XerC/XerD recombinases from *E. coli* among others (Gibb et al., 2010). Their catalytic tyrosines form a 3'-phosphotyrosine intermediate substrate during cleavage, and the released free 5' hydroxyl DNA ends carry out the intermolecular nucleophilic attack of the partner phosphotyrosine linkages and produces Holliday Junction intermediates (Figure 37A) (Gopaul et al., 1998). A second pair of recombinase units resolves the recombinant products by cleavage and strand exchange of the remaining pair of strands. The YR active site is composed of five conserved residues, described as the "RKHRH" catalytic pentad, and a Glu or Asp, which jointly coordinates and catalyzes the attack by the active site tyrosine (Figure 37B andC) (Huang et al., 2004). One of these Histidines is substituted by a tryptophan in the Cre and Flp recombinases (Gibb et al., 2010). Interestingly, four of the seven active site residues aforementioned are strictly conserved among the topoisomerases IB family.

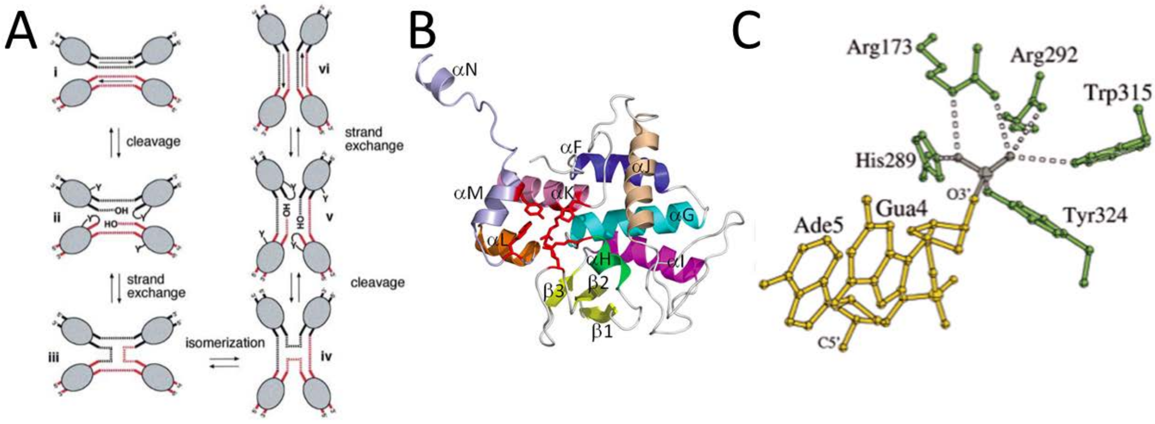


Figure 37. Tyrosine recombinases. (A) Scheme of YR site-specific recombination pathway. Two of the four recombinases in a synaptic tetramer cleave the DNA backbones and form 3'-phosphotyrosine intermediates. The released 5'-OH ends of the cleaved DNA undergo an intermolecular nucleophilic attack of the partner phosphotyrosine linkages to complete the exchange of one pair of strands and form a Holliday junction intermediate. A second round of cleavages and strand exchanges using the remaining two recombinase subunits and the complementary DNA strands gives recombinant products. (B) Ribbon structure of Cre recombinase. The conserved residues "RKHRH" are shown as sticks in red. (C) Cre–*loxA* active site showing the 3'-phosphotyrosine linkage. Hydrogen bonds are depicted as dashed lines. From Gopaul et al., 1998; Gibb et al., 2010; Zhan et al., 2012.

2.4 Protelomerases

Some bacteria possess linear plasmids and chromosomes, like *Borrelia burgdorferi* and one of the two chromosomes of *A. tumefaciens* (Huang et al., 2004). These linear DNA molecules do not have open ends, instead they have hairpin ends that are not vulnerable to exonuclease degradations. Protelomerases use a topoisomerase-IB/YR type mechanism to generate hairpin ends and resolve linear chromosomes during replication. A dimer of protelomerases cleaves and rejoins one strand to its complementary to form hairpin ends (Figure 38A) (Shi et al., 2013). No cofactor such as ATP or divalent cation is required for the reaction. The signature motif "RKHRH" of YR is also found in *A. tumefaciens* C58 TelA, *Borrelia* ResT and plasmid prophague N15 and fKO2 protelomerases (Figure 38B) (Huang et al., 2004). During DNA cleavage, the amino group of the conserved lysine residue might protonate the 5'-OH leaving group (Aihara et al., 2007).

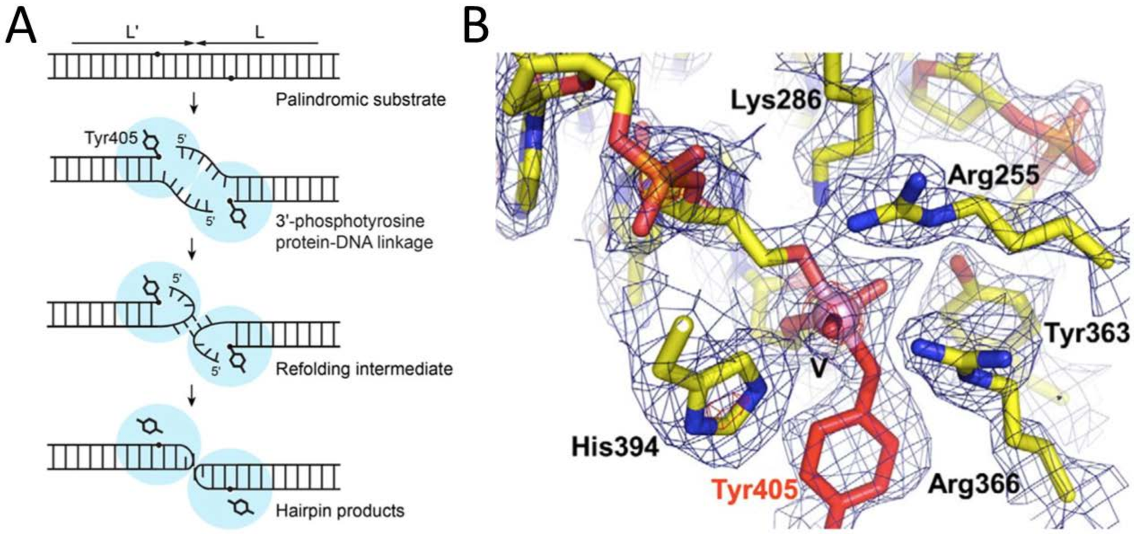


Figure 38. Protelomerases. (A) Model of hairpin end formation of protelomerases. Two molecules of protelomerase (in blight blue) bind to the IR (L’-L). The catalytic Tyr405 cleaves the DNA and forms a 3’-phosphotyrosine-DNA linkage. This is followed by a strand exchange in which the 5’-OH end of each strand is paired with the opposite protein-linked end. Religation results in hairpin formation and the release of the proteins. (B) TelA active site. Orthovanadate (V, magenta sphere) bridges the DNA 5’-OH group and 3’-OH groups, as well as the Tyr405 sidechain, mimicking the pentavalent transition state of the DNA ligation reaction. Taken from Shi et al., 2013.

2.5 Restriction endonucleases

The family of DD(E/D) endonucleases are dimeric metal-dependent proteins that generate double-strand breaks by using a two-metal polymerase/phosphodiesterase reaction (Yang, 2008). These enzymes have three conserved acidic residues that coordinate two metal ions (usually two Mg^{2+} cations) (Figure 39A). One cation activates a water molecule that performs the nucleophilic attack to the scissile phosphate, while the other coordinates a leaving group. DD(E/D) family includes Tn5 transposase and most of the insertion sequence element (IS) transposases, HIV integrase, Piv invertase, RNase H, the 3′-5′ exonuclease of the Klenow fragment of *E. coli* DNA polymerase I and some Holliday junction resolvases such as RuvC (Hickman and Dyda, 2015).

DNA nucleases of the phospholipase D (PLD) superfamily use a different mechanism for the hydrolysis of phosphodiester bonds, a two-step metal-independent scheme (Sasnauskas et al., 2010). All PLD enzymes contain two copies of a conserved sequence motif ‘HXK’. A histidine from one ‘HXK’ motif acts as the nucleophile that attacks the scissile phosphate to create a covalent phosphohistidine intermediate, while the histidine from the second ‘HXK’ motif stabilizes and protonates the hydroxyl of the 3’-leaving group by activation of a water molecule (Figure 39B). The hydrolysis of the covalent enzyme-DNA intermediate is a rapid process and the covalent complex cannot be isolated.

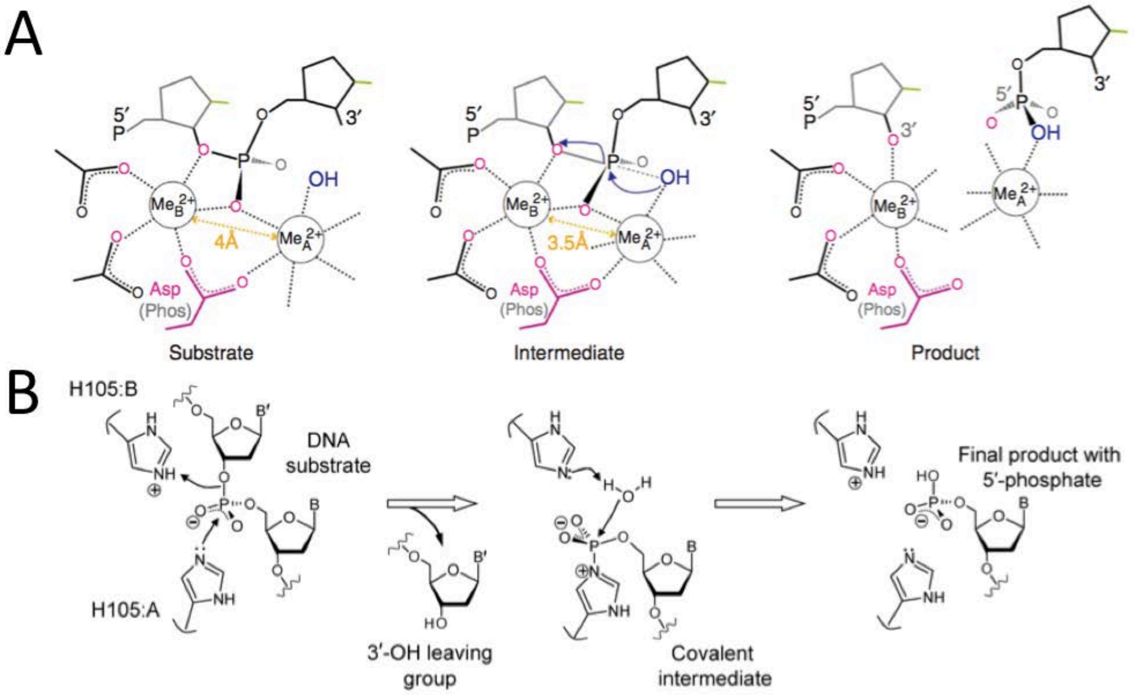


Figure 39. Catalytic mechanism of restriction endonucleases. (A) Two metal ion catalysis by RNaseH. The attacking nucleophile (OH, in blue) attacks the scissile phosphate leading to cleavage of the phosphodiester bond. Metal B interacts with the leaving group from hydrolysis to lower the activation energy of the transition state. Coordination of metal ions A and B is indicated by dashed lines. (B) Reaction mechanism of Phospholipase D superfamily. During the first step of the reaction,BfiI His105 from subunit A (H105:A) attacks the scissile phosphate to generate the covalent intermediate, while His105 from subunit B (H105:B) protonates the 3′-leaving group. During the second step, a water molecule resolves the covalent intermediate releasing the first histidine (H105:A); the second histidine (H105:B) may facilitate this reaction by subtracting a proton from the water molecule. Adapted from Yang, 2008 and Sasnauskas et al., 2010.

2.6 Serine Recombinases

The serine family of recombinases accomplishes strand cleavage, exchange and religation using a synaptic protein tetramer (Ritacco et al., 2013). A conserved serine residue from each bound monomer attacks the DNA backbone to create a covalent DNA-(5'-phosphoseryl)-protein linkage (Li et al., 2005). They make double-strand breaks on both DNA duplexes leaving the free 3'OH groups (Turan and Bode, 2011). The synaptic structures of serine recombinases Gin, γδ resolvase and Sin displayed evidence that one dimer rotates 180º around the exchange interface to perform the strand exchange (Ritacco et al., 2013).

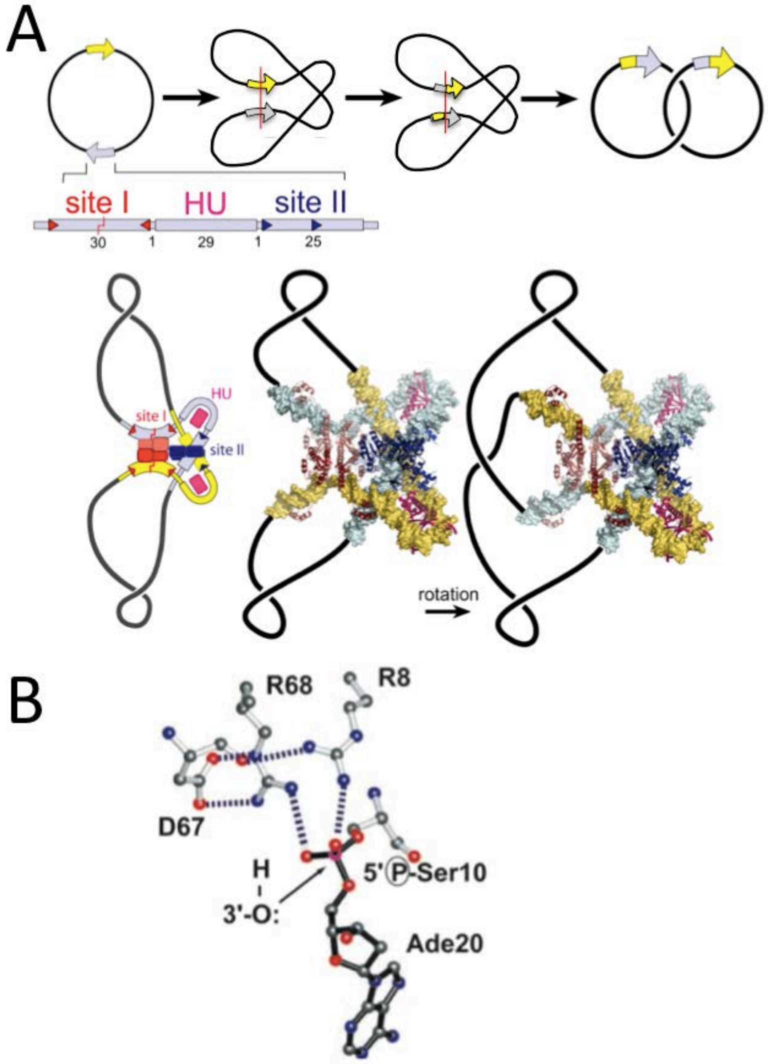


Figure 40. Serine-Recombinases. (A) Structural model of the synaptic complex and rotation model performed by the serine recombinase Sin. The Sin site (gray and yellow arrows) binds two Sin dimers (at sites I and II) and a DNA-bending protein (HU or IHF). Synaptosome formation brings two sites I together. One Sin dimer catalyzes cleavage and rejoining at site I center (red line). A

180° rotation within the site I-bound tetramer creates recombinant products, which are catenated daughter circles. (B) Active site of serine-recombinases. The phosphoserine bond is formed between Ade20 and S10 of γδ resolvase. R8, D67, R68 and the two nonbridging oxygens of the phosphoserine form a hydrogen-bonding network (blue dashed lines). In the religation step, the free 3’OH attacks the DNA strand to be exchanged (black arrow). From Li et al., 2005.

OBJECTIVES

Since the ordered arrangement of proteins on DNA nanostructures is at the source of many applications in biomedicine, we focused in the development of a site-specific, non-chemical strategy for DNA coupling based on relaxase proteins. Considering the previous experience of our group on this type of proteins, we thought that relaxases offer an alternative to existing methods for protein conjugation to DNA nanostructures. The aim of this work was to develop a strategy to bioconjugate HUH relaxases to DNA nanostructures. Particularly, our efforts focused on achieving the orthogonal binding of several relaxases to DNA nanostructures, to allow the programmable arrangement of several fused proteins in close proximity on DNA nanostructures. To reach this main goal, specific objectives were selected:

1. Definition of the best target and its best position on the DNA nanostructures for our model relaxase $\text{TrwC}_{\text{R388}}$.
2. Demonstration that any protein fused to the relaxase domain, e.g. a fluorescent protein-relaxase fusion, could also be attached to specific locations on DNA nanostructures.
3. Displacement of the cleavage-religation equilibrium of HUH relaxases to trap the covalent complex relaxase-DNA by using “suicide substrates” on DNA nanostructures.
4. Evaluation of the specific binding of several relaxases to DNA nanostructures.
5. Development of a method for DNA nanoribbons self-assembly mediated by relaxases.
6. Study of substrates designed to enhance the covalent binding of the relaxase.

EXPERIMENTAL PROCEDURES

3. Materials

3.1 Strains

Table M1. Strain derivatives of *E. coli* used in this thesis

Strain	Genotype	Reference
DH5α	F^- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80d*lacZ*ΔM15 Δ(*lacZYA*- *argF*)U169, hsdR17(rK^- mK^+), λ^-	(Grant et al., 1990)
C41(DE3)	F^- *ompT* $hsdS_B$ (r_B^- m_B^-) *gal dcm* (DE3)	(Miroux and Walker, 1996)
Rosetta	F^- *ompT* $hsdS_B$ (r_B^- m_B^-) *gal dcm* $pRARE^2$ (Cm^R)	(Novy et al, 2001).
UB1637	F^- *recA56 his lys trp rpsL*	(De la Cruz and Grinsted, 1982)

3.2 Plasmids

Table M2. Plasmids used during this thesis.

Plasmid	Description	Phenotype	Size (Kb)	Reference
pET3a	Expression vector	Ap^R Rep(pMB8)	4.6	Novagen
pET29c	Expression vector	Km^R Rep (pMB1)	5.4	Novagen
pSU19	Cloning vector	Cm^R Rep (p15A)	2.3	Bartolome et al. 1992
R388	R388 wild type plasmid	Su^R, Tp^R, Tra^+, IncW	33	((Datta and Hedges, 1972)
pSU2007	R388 derivative Km^R	Km^R Tp^R IncW Tra^+	32	Martínez and de la Cruz, 1988
pSU1186	pUC8:*oriT*(R388)	Ap^R Rep (pMB8)	3.1	(Llosa et al., 1991)
pSU1445	pSU1425(TrwC::Tn5tac1)	Su^R Tp^R Km^R IncW	37.6	(Llosa et al., 1994)
pSU1501	pKK223-3::trwC	Ap^R Rep(pMB8)	7.7	Grandoso et

Plasmid	Description	Phenotype	Size (Kb)	Reference
				al, 1994
pSU1621	pET3a:TrwC (XhoSac)	Ap^R Rep(pMB8)	7.7	(Guasch et al., 2003)
pSU1588	pET3a:TrwC-N293	Ap^R Rep(pMB8)	5.5	Hernando et al., 2000
pSU1614	pET3a:TrwC-N293(Y18F)	Ap^R Rep(pMB8)	5.5	Lucas et al., 2006
pSU1618	pET3a:TrwC-N293(Y26F)	Ap^R Rep(pMB8)	5.5	Lucas et al., 2006
pSU10064	pET3a:MobA-N186_RSF1010	Ap^R Rep(pMB8)	5.2	(Alvarado et al., 2012)
pSU10018	pET3a:TraI-N312_R100	Ap^R Rep(pMB8)	5.5	(Alvarado et al., 2012)
pSU4280	pET3a:TraI-N274_pKM101=R46	Ap^R Rep(pMB8)	5.4	(Alvarado et al., 2012)
pSSP10	pET29c:TraI-N270_ RP4	Km^R Rep (pMB1)	6.2	This work
pSU10032	pET3a:TraI_pOLA52	Ap^R Rep(pMB8)	5.8	(Alvarado et al., 2012)
pYTP4	pSU19: synthetic *oriT*	Cm^R Rep (p15A)	2.6	Yera Tena PhD thesis
pYTP5	pSU19: synthetic *oriT* ∆IR1	Cm^R Rep (p15A)	2.6	Yera Tena PhD thesis
pYTP7	pSU19:*oriT* synthetic ∆IR2	Cm^R Rep (p15A)	2.6	Yera Tena PhD thesis
pSSP01	pSU19:*oriT* synthetic Rep-like	Cm^R Rep (p15A)	2.6	This work
pSSP02	pMAT:*oriT* synthetic Rep-like ∆IR1 (GeneArt)	Ap^R Rep(colE1)	2.5	This work
pSSP03	pSU19:*oriT* synthetic Rep-like strategy 3	Cm^R Rep (p15A)	2.6	This work
pSSP04	pSU19:*oriT* synthetic Reverse like	Cm^R Rep (p15A)	2.6	This work
pSSP05	pMAT:*oriT* synthetic Reverse like ∆IR1(GeneArt)	Ap^R Rep(colE1)	2.5	This work

Plasmid	Description	Phenotype	Size (Kb)	Reference
pSSP06	pSU19:*oriT* Mut Distal Arm IR2	Cm^R Rep (p15A)	2.6	This work
pSSP07	pSU19:*oriT* ∆ Distal Arm IR2	Cm^R Rep (p15A)	2.6	This work
pSSP08	pSU19:minimal Rep-like	Cm^R Rep (p15A)	2.6	This work
pSSP09	pET3a:TrwC N207	Ap^R Rep(pMB8)	5.2	This work
pSUMC48	pET29c:TrwC N293HIS	Km^R Rep (pMB1)	6.3	This work
pSUMC49	pET29c:TrwC N207HIS	Km^R Rep (pMB1)	6.0	This work
pSUMC36	pET29c: $TrwC_{R388}$mGFP	Km^R Rep (pMB1)	6.8	This work
pSUMC37	pET29c: $TrwC_{R388}$mCherry	Km^R Rep (pMB1)	6.8	This work
pSUMC47	pET29c: $TrwC_{R388}$mCherryhis	Km^R Rep (pMB1)	6.8	This work
pSSP11	pET29c: $TrwC_{R388}$CFPhis	Km^R Rep (pMB1)	6.8	This work
pSSP12	pET29c: $MobA_{R1162}$CFPhis	Km^R Rep (pMB1)	6.5	This work
pSSP13	pET29c: MobA $_{R1162}$KATEhis	Km^R Rep (pMB1)	6.5	This work
pSSP14	pET29c: $TraI_{R100}$KATEhis	Km^R Rep (pMB1)	6.8	This work
pSSP15	pET29c: $TraI_{pKM101}$CFPhis	Km^R Rep (pMB1)	6.8	This work
pSSP16	TrwC-N293 (K215A)	Ap^R Rep(pMB8)	5.5	This work
pSSP17	TrwC (K215A)	Ap^R Rep(pMB8)	7.7	This work
pSSP18	TrwC-N293 (K215I)	Ap^R Rep(pMB8)	5.5	This work
pSSP19	TrwC (K215I)	Ap^R Rep(pMB8)	7.7	This work
pSSP20	TrwC-N293 (Q205K)	Ap^R Rep(pMB8)	5.5	This work
pSSP21	TrwC (Q205K)	Ap^R Rep(pMB8)	7.7	This work

Plasmid	Description	Phenotype	Size (Kb)	Reference
pSSP22	TrwC-N293 (Y191K)	Ap^R Rep(pMB8)	5.5	This work
pSSP23	TrwC (Y191K)	Ap^R Rep(pMB8)	7.7	This work
pSSP24	TrwC (Q205K+Y191K)	Ap^R Rep(pMB8)	7.7	This work

3.3 Oligonucleotides

Oligonucleotides were purchased to Sigma-Aldrich (Madrid, Spain), except IRDye labelled oligonucleotides that were obtained from Teknokroma Analítica (Barcelona, Spain). Phosphorothiolate oligonucleotides were obtained from W. M. Keck Oligonucleotide Synthesis Facility (Yale University, USA). The modified phosphoramidite used was 3’-Thio-dT from Berry&Associates (BA0255, supplementary S1). Oligonucleotides used for assembly of DNA nanostructures were purchased from Eurofins MWG (Ebersberg, Germany), but oligonucleotides carrying the relaxases binding site were obtained from Biomers (Ulm, Germany). Oligonucleotides longer than 60 nt were HPLC purified. Oligonucleotide sequences are included in the Supporting Information (Tables S1 to S7).

3.4 Enzymes and chemical reagents

Restriction enzymes, T4-ligase and DNA polymerases, as well as the DNA purification kits (GeneJET Plasmid MiniPrep and GeneJET PCR Purification kit) were purchased to Thermo Scientific. Chemical reagents were purchased to Sigma Life Sciences. Antibiotics and IPTG (isopropyl-β-D-thiogalactoside) were obtained from Apollo Scientific.

3.5 Buffers and solutions

Cell Lysis Buffer	Tris 100mM pH 7.5, EDTA 1mM, NaCl 500mM, PMSF 1%
Lysis Dilution Buffer	Tris 100mM pH 7.5,EDTA 1mM, PMSF 1%
Buffer A and Washing Buffer	Tris 100mM pH 7.5, EDTA 1mM, NaCl 200mM

Dilution Buffer	Tris 100mM pH 7.5,EDTA 1mM
Buffer Elution I	Tris 100mM pH 7.5, EDTA 1mM, NaCl 600mM
Buffer Elution II and Buffer B	Tris 100mM pH 7.5, EDTA 1mM, NaCl 1M
Buffer Gel Filtration	Tris 100mM pH 7.5, EDTA 1mM, NaCl 300mM
Buffer C	Tris 10 mM pH 7.5, EDTA 1mM, NaCl 100mM
Buffer D	Tris 100mM pH 6.5, EDTA 1mM, NaCl 1M
Buffer E	100 mM Tris pH 7.5, 300 mM NaCl
Buffer F	100 mM Tris pH 7.5, 300 mM NaCl,500mM imidazole
dNTPs	each 10 mM dGTP, dCTP, dATP, dTTP
DNA Sample Loading Buffer 6x	0.25% bromophenol blue (w/v), 40% sucrose in 1x TBE
SDS Sample Loading Buffer 2x	50 mM Tris-HCl pH 6.8, 4% SDS, 4% glycerol and 0.02% Bromophenol Blue
Coomassie Staining Solution	0.025% Coomassie Blue R-250, 40% methanol, 10% acetic acid
Destaining Solution	10% acetic acid, 40% ethanol
1x SDS-PAGE Buffer	25 mM Tris-HCl pH 8.3, 200 mM glycine, 1% SDS
TAE	40 mM Tris, 0.5 mM EDTA, pH 8.0
TBE	89 mM Tris, 89 mM boric acid, 2 mM EDTA, pH 8.3
10x Folding Buffer 2D	10x TAE with 125 mM $MgCl_2$
10x Folding Buffer 3D	10x TAE , 200 mM $MgCl_2$, 50 mM NaCl
1x Folding Buffer 2D	1x TAE with 12,5 mM $MgCl_2$
1x Folding Buffer 3D	1x TAE , 20 mM $MgCl_2$, 5 mM NaCl
1x Cleaving Buffer	40 mM Tris, 0.5 mM EDTA, 10 mM $MgCl_2$, pH 7.5

4. Methods

4.1 Biological& Biochemical methods

4.1.1 Molecular biology techniques

To assure high specificity and high yield of amplification, *Phusion* and *Vent* DNA polymerases were used because exhibit 3’->5’ proofreading activity. PCR reaction conditions and program were performed with the following protocol. 50 µl sample reactions were prepared, containing 10 pmol/µl of each primer, 50-100 ng of template, 200 µM of dNTPs, 1U of DNA polymerase in 1x polymerase buffer. A iCycler (BioRad) was used with the following program: 1 minute of denaturation at 98ºC, 30 cycles of amplification (including 10 seconds at 98ºC, 15 seconds at a temperature according to TM of primers used, 1.5 minutes at 72ºC (30 seconds per kb) and 10 minutes of elongation at 72ºC.

Vector and insert DNAs were digested with the appropriate enzymes to generate compatible sticky ends. A mixture containing 20 µl DNA in the Buffer recommended by Thermo and the respective restriction enzymes, was incubated at 37°C for 1 to 2 h. Digested vector DNA was purified from the agarose gels with the GeneJET gel extraction kit (Thermo), whereas digested PCR DNA was extracted with the GeneJET PCR purification kit (Thermo).

DNA ligation was performed using a molar ratio of vector to insert DNA of 1:3 and 1U of T4 DNA ligase (Thermo) in a final volume of 20 µl for overnight incubation at room temperature.

Colony PCR was performed to screen bacterial clones with desired plasmid products. Selected colonies of bacteria were picked and lysed in 100 µl of water by boiling the sample 5 min. Samples were chilled on ice and centrifuged 1 min at 13200 rpm. 3-5 µl of the supernatant served as template DNA for PCR.

The sequences of the plasmids were confirmed by sequencing performed by the Macrogen Sequencing Service (Macrogen, Netherlands) and Stabvida Sequencing Service (Stabvida, Portugal).

4.1.2 Plasmid construction

Synthetic *oriT*s were used for the study of relaxation of supercoiled plasmids and bacterial conjugation. The minimal 28 bp Rep-like hairpin was inserted between *BamHI* and *HindIII* sites of vector pSU19. The Rep-like *oriT* plasmid was obtained from GeneArt (Life technologies-Invitrogen). The 190 bp product was amplified by PCR, purified and cloned into an *oriT* synthetic plasmid (pYTP4) between *KpnI* and *SacI* sites (it contains the IR1 sequence before *KpnI* site). Consequently, a plasmid with the complete R388 *oriT* and IR_2 with Rep-like design was generated. The same procedure was done with the Reverse-like sequence. The ∆distal arm plasmid was PCR-amplified with primers containing the deletion (Figure M1.A). The linearized PCR-product was self-ligated using the T4-ligase protocol (Thermo). The distal arm mutated plasmid and the Rep-like strategy 3 plasmid were generated following the same steps.

The N terminal domain of RP4 TraI ($TraI_{RP4}$) (270 residues) was PCR amplified and subcloned in pET3a using *NdeI* and *BamHI* restriction sites, resulting in plasmid pSSP10. Similarly, the first 207 residues from R388 TrwC (TrwCnf), were PCR amplified from pSU1588 and cloned into pET3a between *NdeI* and *BamHI* restriction sites, resulting in pSSP09. Since purification resulted in low yield, we also obtained the His-tagged version of TrwCnf by cloning it into pET29c with the *NdeI* and *XhoI* restriction sites (pSUMC49). Similarly, the $TrwC_{R388}$ His tagged protein was also constructed with the six histidines fused at the C-terminus (pSUMC48).

The gene fragment codifying for $TrwC_{R388}$, $MobA_{R1162}$, $TraI_{pKM101}$ and $TraI_{R100}$ was amplified by PCR from plasmid R388, RSF1010, R100 and pKM101 using oligonucleotides containing *NdeI, BamHI* or *BglII* restrictions sites. The genes coding for monomeric fluorescence proteins, mCFP, mKATE and mCherry, were amplified by PCR from different plasmid constructions using primers containing *HindIII* and *XhoI* or *EcoRI* and *XhoI* restriction sites. These inserts were placed on plasmid pET29c (Novagen) in a two-step ligation (see schemes in supplementary Figure S1). The resulting His-tag fusion proteins have the fluorescent moiety at the C terminal domain, separated from the relaxase domain by a linker of eight to ten residues.

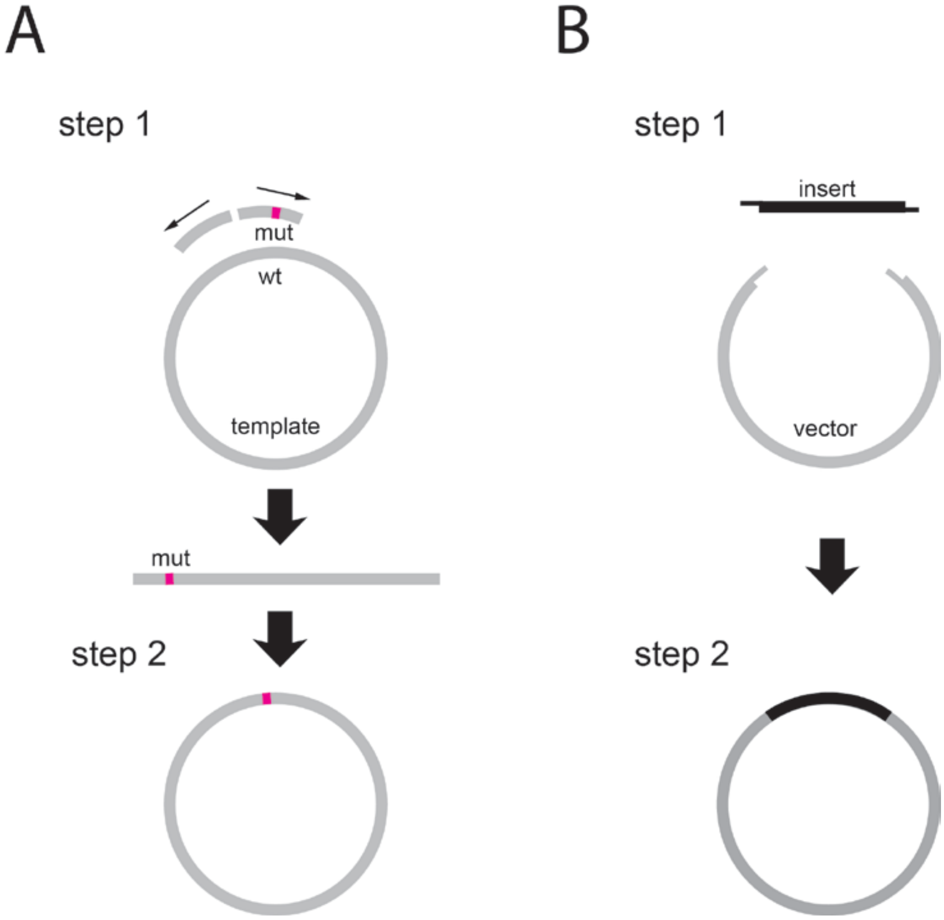


Figure M1. Scheme of the two methodologies used for the construction of synthetic plasmids. (A) Point-mutation, deletions and insertions were included in a given plasmid by PCR-amplifying the plasmid with primers that possessed the mutation. In step 1 in one PCR the full length of the plasmid is amplified with *Phusion* polymerase, because PCR products are blunt ended. In a second step, linearized plasmid is self-circularized. (B) Classical cloning method, where a digested insert and vector ligates in a circular plasmid. The PCR product is digested with the restriction enzymes (step 1) that enable ligation with the digested plasmid (step 2).

4.1.3 Site-directed mutagenesis

TrwC point mutants were constructed by site-directed mutagenesis using the Quick-change method (Agilent) but with KOD Hot Start DNA polymerase (Novagen). Primers were designed to be between 40 and 45 bases in length, with the desired mutation in the middle of the primer and terminated in a C or G base. Forward and reverse primers have complementary sequences.

The following components were used for the PCR reaction mix:

	Volume for one reaction	Volume for four reactions
10x Buffer	2 µl	8 µl
2mM dNTPs	1 µl	4 µl
Primer Forward (10 µM)	1 µl	4 µl
Primer Reverse (10 µM)	1 µl	4 µl
25 mM Mg^{2+}	0,8 µl	3,2 µl
Template DNA (≈ 50 ng/ µl)	0,5 µl	2 µl
KOD Hot start DNA Polymerase	0,25 µl	1 µl
Water	13,45 µl	53,8 µl
Total volume	20 µl	80 µl

The thermocycler program for Quick-change PCR contained the following steps:

Stage	Conditions
Initial denaturation	3 min at 95°C
Denaturation	1 min at 95°C, 40 cycles
Annealing	1 min at 54-56°C
Elongation	0,5-1 min/kbp at 72°C
Final incubation	10 min at 72ºC

After running 4 µl of each PCR product on a 1% agarose gel, PCR products were digested with the restriction enzyme *DpnI* at 37ºC for 2 h to remove any methylated DNA, leaving only the PCR product. 2 µl of PCR product was transformed into electrocompetent DH5α cells and plated onto appropriate LB selective media. Plasmid DNA extracted from the selected clones was sequenced to verify that the selected clones contain only the desired mutation.

4.1.4 DNA electrophoresis in agarose gels

DNA fragments were separated via (1-2%) agarose gel electrophoresis. The gels were prepared by dissolving the agarose in the 0.5x TBE buffer. To visualize the DNA, 5 µl of SyBRSafe® (Thermo) or SafeView^{TM} (Gentaur) was added to the 100 ml (1%) agarose solution. Usually, 20 µl of sample solution mixed with 4 µl of DNA sample-loading buffer was loaded to each well. Electrophoresis was performed at 100 V for 20-60 minutes. The DNA was detected using UV light in a GelDoc and the size of the DNA was determined using standard 100 bp-DNA ladder.

The experiments carried in Munich used an optimized protocol to evaluate the DNA nanostructures. 1% agarose gel was run in 40 mM Tris, 20 mM acetic acid, 1 mM EDTA, 12.5 mM $MgCl_2$ (pH 8.0) at 100V for 2 h. 10 μl of 10 nM sample solution mixed with 2 μl of DNA sample-loading buffer was loaded to each well. Gels were scanned with a Typhoon scanner in 2 channels 483 nm and 532 nm to evaluate the fluorescent proteins. Afterwards, staining with SybrGold^{TM} (Invitrogen) for 10 minutes was evaluated under a GelDoc.

4.1.5 DNA electrophoresis in acrylamide gels

Electrophoretic mobility shift assays were performed on 5% native polyacrylamide gel containing 1x TBE buffer. 20 μl of sample mixed with DNA loading buffer was loaded to each well. Gels were running 15-30 min at 150-200V in 1x TBE buffer.

40% Acrylamide 19:1	1,25 ml
TBE 10x	1 ml
Water	7,75 ml
APS	50 μl
TEMED	20 μl

The electrophoretic separation of urea/heat-denatured nucleic acids in 8 M urea-containing acrylamide gels in 1xTBE buffer was performed for resolving 18-40 mer oligonucleotides. DNA samples were mixed with 1.25x formamide-DNA loading buffer and heated to 95ºC for 2 min. Gels were pre-run in 1xTBE buffer for 30 min at 200V. Before loading the samples, the wells were rinsed until unpolymerized material was removed. Gels were run at 200V for 70 min.

Urea	7,2 g
40 % Acrylamide 29:1	4,5 ml
TBE 10x	1,5 ml
Water	Up to 15 ml
APS	50 μl [a]
TEMED	20 μl [a]

[a] For polymerization of just 10 ml of reaction mixture.

When oligonucleotides were labeled with a fluorescent dye, gels were directly scanned for fluorescence at the wavelength needed. When standard oligonucleotides were used, gels were stained in SyBrGold (Thermo) solution for 10 minutes with shaking. After staining, the excess of dye was removed by washing with buffer. Gels were visualized with a BioRad GelDoc and quantified by Quantity One software (Bio-Rad).

4.1.6 Protein separation by SDS PAGE

Protein electrophoresis was carried on 10% SDS-polyacrylamide gels for $\text{TrwC}_{\text{R388}}$ and the relaxases fused to fluorescent proteins, 12% for the relaxase domains of $\text{TrwC}_{\text{R388}}$, TraI_{RP4}, $\text{TraI}_{\text{R100}}$ and $\text{TraI}_{\text{pKM101}}$, 15% SDS-polyacrylamide gels for the relaxase domain of $\text{MobA}_{\text{RSF1010}}$. Samples were mixed with SDS loading buffer in a 1:1 ratio and heated in boiling water for 3 minutes, before being loaded on the polyacrylamide gel. Low Range standards (Bio-Rad) were used for molecular mass measurement. Electrophoresis was carried out at 180 V for 60-90 minutes in 1x SDS-PAGE buffer. Gels were stained with Coomassie staining solution for unless 30 minutes and destained until background was clear. Gels were acquired in a HP scanner and processed with Quantity One software (Bio-Rad).

4.1.7 Bacterial culture

Escherichia coli strains were grown in Luria–Bertani broth (LB) at 37 °C. Plating was on LB agar. The following antibiotics were used at the indicated concentrations: ampicillin, 100 µg/ml; kanamycin, 50µg/ml; chloramphenicol, 25µg/ml; streptomycin, 500µg/ml and nalidixic acid, 20µg/ml. IPTG (isopropyl-β-D-thiogalactoside) was used at a concentration of 500 µM.

4.1.8 Transformation by electroporation

The following protocol is suitable for preparation of electrocompetent cells from all the strains of *E.coli* and can yield transformation efficiencies ≥ 5×10^8 colonies/µg. Bacteria from a frozen stock were streaked on an LB agar plate, and incubated at 37°C overnight. A single colony was transferred into 10 ml of LB containing antibiotic (except for C41 strain) and grown with shaking at 37ºC overnight. 200 ml of LB medium in a 1l flask were inoculated with overnight culture (1:20 dilution) and incubated at 37ºC with shaking to an OD600 ~ 0.6-0.8 (approximately 2 h). Then, cells were incubated on ice for 30 minutes and transferred to 500ml sterile centrifuge bottles previously incubated on ice. After centrifugation at 4,000 rcf for 10 minutes at 4ºC, supernatant was decanted and pellets were resuspended by gently vortexing, in 200 ml of ice-cold sterile water. Cells were again recovered by centrifugation at 4,000 rcf for 10 minutes. The supernatant was decanted from the cell pellets and the pellets were resuspended in another 200 ml of DWE. After a new centrifugation step, cells were resuspended in 200 ml of ice-cold sterile 10% glycerol. Finally, cells were recovered by centrifugation and the supernatant was decanted, leaving 1 ml to resuspend the cells. 80 µl aliquots were dispensed into sterile Eppendorf tubes and cells were snap-freezed by immersing the tubes in ethanol with dry ice. Competent cells were stored at -80º C until needed.

An aliquot of 80 µl of competent *E. coli* cells was thawed on ice for 15 min. DNA samples were dialyzed on 0.05 µM Millipore GS filters on a Petri plate with MilliQ water for 30 minutes. 200 ng DNA were added to the cell suspension and placed in a 0.2 cm electroporation cuvette Gene Pulser (BioRad Laboratories). Following electroporation at 2.5 kV and time constant (3-5 ms) in a Micropulser^{TM} electroporator (BioRad), cells were immediately recovered in 1 ml of sterile LB. Transformed cells were incubated at 37ºC for 1 to 2 h and plated on selective LB-Agar plates.

4.1.9 Bacterial conjugation experiments

Matings were carried out by the plate-mating procedure (Guasch et al., 2003). Saturated cultures of donor and recipient strains were centrifuged at 4000 rcf for 10 min, and resuspended in 10 ml of LB. After determining the OD at 600 nm of the 10^{-1} dilution, cells were mixed in a 1:1 ratio in 4 ml. Samples were centrifuged and resuspended into 400 µl of LB and 15 µl mated on a LB agar surface of a sterile 24-well plate F24 PS (Corning, USA). After 1 h at 37ºC, mating was stopped adding 1 mL of LB. Drops of 10 µl of the dilutions from 10^{-1} to 10^{-6} of each mating were cultivated at 37ºC overnight. The estimation of donors (eg., Nx^{R}, Cm^{R}, Kn^{R}), recipients (eg., Sm^{R}) and transconjugants (eg., Cm^{R}, Sm^{R}) per milliliter were done by triplicates of each mating performed in three independent days. Conjugation frequencies were expressed as the number of transconjugants per donor cell.

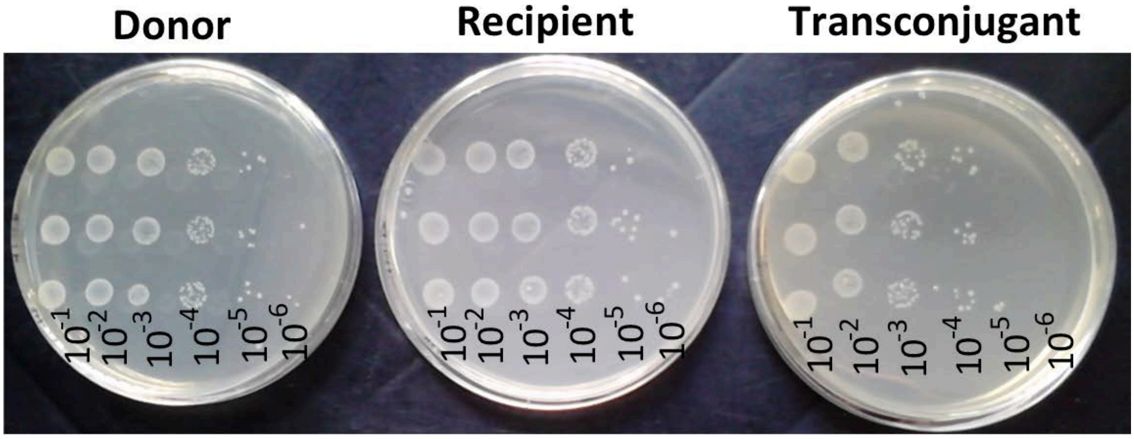


Figure M2. Plate-mating procedure. Drop plate method was used to determine the number of donor, recipient and transconjugant bacteria after 1 h conjugation at 37ºC. A volume of 10μl is deposited as a drop onto the surface of a LB agar plate in selective media. Triplicates of a 10-fold serial dilution were pippeted as shown in this figure.

4.1.10 Protein purification

The protein was overexpressed with 0.5 mM IPTG for 3 h in *Escherichia coli* strain C41. Cells were harvested by centrifugation and stored at -80ºC. Frozen cells were resuspended in 50 ml of Cell Lysis Buffer, and sonicated non–continuously for 10 minutes. The lysate was centrifuged at 40,000g for 15 min at 4ºC. Supernatants were diluted with Lysis Dilution Buffer to 200 mM NaCl before being applied to a P11-phosphocellulose column equilibrated in Buffer A. Proteins were eluted with Elution I and Elution II buffer. Fractions containing the protein were pooled, diluted to 200 mM NaCl, and loaded to either a Hi-Trap SP HP column (GE Healthcare) or Heparine HP column (GE Healthcare). Proteins were eluted with a linear NaCl gradient (200 mM–1000 mM NaCl) with buffer A and B. The fractions containing the protein were loaded onto a gel filtration S75 column 10/300 GL (GE Healthcare) with Gel Filtration buffer. The flow rate was 0.4 ml/min and the protein's absorption was detected at wavelength 280 nm.

Expression of His-tagged fusion proteins was induced with 0.5mM IPTG for 3 h in Rosetta strain (Novagen). Cells were lysed and centrifuged with the same protocol. Supernatants were diluted with Lysis Dilution II buffer to 200 mM NaCl before being applied to a HisTrap HP column (GE Healthcare) equilibrated in Buffer E. Fluorescent proteins were eluted with a linear imidazol gradient (0 mM–500 mM) with buffer E and F. The fractions containing the protein were loaded onto a gel filtration S200 column 10/300 GL (GE Healthcare) with Gel Filtration buffer. The flow rate was 0.4 ml/min and the protein's absorption was detected at wavelength 280 nm.

4.1.11 Determination of protein concentration

The protein solutions were concentrated by ultrafiltration with Amicon Ultra Centrifugal Filter Units (Millipore, USA). Membranes with a pore size at least 2 fold smaller than the molecular mass of the protein were used. The protein solutions were centrifuged at 4,000 rcf at 4°C until the required concentration was reached. Its concentration was estimated in a Nanodrop spectrophotometer by UV absorbance at 280 nm using an extinction coefficient calculated in ProtParam tool from Expasy resource portal (http://web.expasy.org/protparam/). The concentration was calculated as:

$$\text{Concentration } (\mu\text{M}) = \frac{\text{Absorbance } \times 10^{6}}{\text{Abs } 0.1\% \times \text{MW}}$$

4.1.12 ssDNA-protein complex formation

Oligonucleotides (listed in Table S1) were resuspended in miliQ water, heated to 95ºC for 10 min, and then either allowed to cool passively to room temperature or snap cooled on ice. Cleavage reactions were carried out by incubating 6.3 μM TrwC$_{R388}$ with 15 μM of each oligonucleotide at 37ºC for 1 h in 10mM Tris-HCl, pH7.6, 5mM $MgCl_2$, 375 mM NaCl and 15 μM EDTA (Grandoso et al., 2000). Similarly, 7 μM MobA$_{RSF1010}$ was incubated with 15 μM of each substrate in 25 mM Tris-HCl pH 8.0, 5 mM $MgCl_2$, 225 mM NaCl and 15 μM EDTA (Scherzinger et al., 1992). For TraI$_{RP4}$, the optimized reaction mixture was 10 mM Tris pH 7.6, 5 mM $MgCl_2$, 225mM NaCl and 15 μM EDTA (Pansegrau et al., 1993). 1.5 μM of TraI$_{RP4}$ was incubated with 15 μM oligonucleotide in each reaction. 7 μM TraI$_{R100}$ and 5 μM TraI$_{pKM101}$ with 15 μM of each oligonucleotide at 37ºC for 1 h in 10mM Tris-HCl, pH7.5, 5 mM $MgCl_2$

and 15 μM EDTA. Reactions were stopped by adding SDS and boiling the samples. The cleavage activity was checked by the lower mobility of the protein-DNA covalent complexes in SDS-PAGE. The densitogram of each band was obtained with Quantity One (BioRad). DNA nicking activity was expressed as percentage of covalent complexes generated (the intensity of the product band intensity divided by the sum of product and free protein band intensities).

4.1.13 DNA cleavage assays

50 nM IRDye800-labelled oligonucleotide R388 W(18+18) were used to measure equilibrium DNA cleavage via denaturing polyacrylamide gels. Each 20 µL reaction contained 1 µM TrwC$_{R388}$, 50 nM IRDye800-labelled oligonucleotide R388 W(18+18) and 500 nM DNA substrate complementary to a given sequence of the target W(18+18) in 10 mM Tris-HCl, pH7.5, 5mM $MgCl_2$, 100 mM NaCl. After incubation for 1 h at 37ºC, reaction mixtures were digested with 0.6 mg/ml proteinase K and 0.05 % (w/v) SDS for 20 min at 37ºC. Reactions were run through a denaturing urea polyacrylamide gel at 200V for 70 min to separate cleaved product DNA from the substrate. Oligonucleotides were visualized using an Odyssey Infrared Image System (LI-COR Biosciences) and quantified by Quantity One software (Bio-Rad). The percent cleavage product formation was calculated as a percentage of the product band intensity divided by the product plus the substrate band intensities.

4.1.14 Strand transfer reactions

For oligonucleotide strand-transfer reactions, each reaction mixture contained 0.25 μM oligonucleotide (Table S1) with 1 µM of protein TrwC$_{R388}$ in 10 mM Tris- HCl pH 7.6, 5 mM $MgCl_2$, 110 mM NaCl and 15 µM EDTA. After 30 min, 50 nM IRDye800-labelled oligonucleotide R388 W(25+0) was added to the reaction mixture. After incubation for 30 min at 37ºC, reaction mixtures were digested with 0.6 mg/ml proteinase K and 0.05 % (w/v) SDS for 20 min at 37ºC. Oligonucleotide separation was performed by denaturing urea polyacrylamide gel electrophoresis at 200V for 70 min. Images were processed with Odyssey Infrared Image System (LI-COR Biosciences) and quantified by Quantity One software (Bio-Rad). The percent cleavage product formation was calculated as a percentage of the product band intensity divided by the product plus the substrate band intensities.

4.1.15 Electrophoretic mobility shift assay (EMSA)

Binding reactions contained 50 nM IRDye-labeled oligonucleotide and increasing concentrations of TrwC$_{R388}$ (42 nM, 85 nM, 210 nM, 420 nM, 850 nM, 4,2 μM and 8.5 μM) in buffer C. Reaction mixtures were incubated for 30 min at room temperature and loaded onto a 10 % non-denaturing polyacrylamide gel. Analyses were carried out in Odyssey Infrared Image System (LI-COR Biosciences).

4.1.16 Gel filtration of ssDNA substrates

The presence of intramolecular and intermolecular structures in the oligonucleotides library created for the determination of the best substrate for TrwC$_{R388}$ was determined by gel filtration in a Superdex^{TM} 75 PC 3.2/30 (GE Healthcare). To this end, two microliters of each oligonucleotide (100 μM) was dissolved in eighteen microliters of the buffer A and injected in the column loop of 25 μL. The carried buffer was buffer A. When we analyzed the protein, twenty microliters of the stock solution of 42 nM of TrwC$_{R388}$ was injected alone. The complexes of TrwC$_{R388}$ with oligonucleotide were loaded after 30 minutes incubation at room temperature, with an excess of oligonucleotide (ratio 1:1.5) to ensure the maximum complex formation.

Due to study the new balance achieved with Rep-like substrates, IRDye labeled oligonucleotide R388 H14+14 was incubated with TrwC$_{R388}$. After 1 h of incubation with 10 mM Mg_2^+, the mixture was analyzed jointly by gel filtration and native electrophoresis. After high-resolution gel filtration column chromatography of the binding mixture, fractions were loaded into a 5% non-denaturing acrylamide gel, and the fluorescent label of the oligonucleotide was detected in Odyssey infrared scanner.

4.1.17 Crystallization of relaxase-DNA complexes

Sitting-drop diffusion method was the preferred method for crystallization of relaxase-DNA complexes. Crystallization screening was carried out in 96-well crystallization plates (CrystalEX microplate) by the sitting-drop method and optimization was set up into a 24-well crystallization plates (Cryschem plate) from Hampton Research (CA,USA). Either the Crystal Screen™ reagent Kits from Hampton Research (CA, USA) or the JCSG-plus HT-96 from Molecular dimensions (Suffolk, UK) were used as a first screening. The sitting drop technique utilizes diffusion and evaporation of water between a small droplet containing protein, buffer and precipitant, and a reservoir (well), containing a solution with similar buffer and precipitant, but at higher concentrations with respect to the droplet. The drops were set up with 50 μl reservoir-solution volume and a drop size of 1.0 μl (0.5 μl protein solution and 0.5 μl reservoir solution) in a 96-well plate whereas 500 μl of reservoir volume and a variable drop size (2-3 μl, with different stoichiometry of complexes and reservoir solution) were used in the 24-well plates. Wells were sealed using a sealing tape (Crystal Clear High Performance, Henkel Consumer Adhesives). Water molecules migrate from the protein droplet through the vapor space into the more concentrated reservoir. Since the volume of the protein droplet decreases, the concentration of protein increases and protein crystals can be form.

- **Crystallization of TrwC$_{R388}$ with different DNA substrates**

The oligonucleotides (R388 W(16+0)) and (R388 W(7::)) hybridized as R388_W(7::16+0), mimicking the DNA structure at the donor cell. The protein TrwC$_{R388}$ was incubated with R388_W(7::16+0) at room temperature for 1 h in 500 mM NaCl, 20 mM Tris HCl pH 7.5, 0.5 mM EDTA. Similarly, phosphorothiolate oligonucleotide (R388 W(16s2)) and standard (R388 W(7::)) were incubated overnight with TrwC$_{R388}$ in 500 mM NaCl, 20 mM Tris HCl pH 7.5, 0.1 mM $MgCl_2$, resulting in complex TrwC$_{R388}$–R388_W(7::16s2). An excess of DNA (molar ratio 1.5:1) was used to ensure that all the protein was complexed. The complexes were concentrated using Amicon ultra-10 (Millipore) to 10-12 mg/ml. Crystals were grown at 22ºC by mixing of 2 μl TrwC$_{R388}$–

R388_W(7::16+0) complex with 1 μl reservoir solution A (2 M ammonium phosphate, 0.1 M Tris-HCl pH 8.5).

The screening of complex TrwC$_{R388}$–R388_(W7::16s2) revealed crystals in reservoir solution B (20% (w/v) polyethylene glycol 8000, 0.1 M phosphate/citrate pH 4.2, 0.2 M Sodium Chloride). Plates were obtained at 22ºC by mixing 1.5 μl TrwC$_{R388}$–R388_W(7::16s2) complex with 1.5 μl reservoir solution B, and also when the mutant TrwC$_{R388}$(Y26F) was used. Moreover, some crystals of TrwC$_{R388}$–R388_W(7::16s2) and TrwC$_{R388}$(Y26F)–R388_W(7::16s2) grew in reservoir solution C (20 % (v/v) polyethylene glycol 3000, 0.2 M ammonium chloride). These conditions were further optimized by varying the precipitant concentration and pH.

- **Crystallization of other proteins**

Some relaxases bound to their cognate DNA substrate were assayed for crystallization. Complexes of TraI$_{RP4}$ with W(25+1), TaxC$_{pOLA52}$ with W(25+1), MobA$_{RSF1010}$ with W(23+1) and the relaxase of Anaeromyxobacter with W(23+0) were screened for the optimal conditions.

- **X-Ray data collection and structure determination**

A cryoprotectant solution (20% (v/v) ethylene glycol) was added to the droplet with crystals to minimize the ice formation during cooling. Protein crystals were mounted in fiber loop at room temperature and flash-cooled under the nitrogen flux. Measurements were done at 105K at the ALBA light source (Barcelona, Spain), at beamline XALOC. Complete data sets were collected using 0.5º oscillations and a total of 180º were collected. Data were collected in Δφ= 0.5º steps over 180° from a single frozen crystal. Diffraction was processed using iMosflm and Scala as part of the CCP4 package. Molecular replacement solutions in MolRep (CCP4) were obtained with the structure of TrwC$_{R388}$ bound to the 23-mer *nic*-site as a search model (PDB ID: 1OMH). Refinement was performed in Phenix and modeling in COOT (Crystallographic Object-Oriented Toolkit).

4.1.18 Design of 2D and 3D DNA nanostructures

The free software cadnano (http://cadnano.org/) was used for designing the DNA nanostructures. Using this tool, we define the shape of the desired nanostructure and the scaffold length, and it generates a list of the necessary sequences of the staple strands. The 3D conformational analysis of the DNA nanostructures was simulated with CanDo (http://cando-dna-origami.org/). DNA origami structures used in this thesis were long, rod-like six-helix bundles (6HB) or flat, twist-corrected rectangular origami sheets (tcRO), with sizes of 430nm x 5nm and 90nm x 60nm, respectively. tcRO and 6HB were designed and optimized at Dr. Simmel lab. Afterwards, extension of some of these staples with the relaxase target was carried *in silico*. A selected staple of the original design (in green in Figure M3) was split into two separate halves at 8 nt after the Holliday junction, in order to have the target projecting perpendicular to the base of the nanostructure. The 3' end of the first half (in green) was extended with a random sequence of 3 to 15 nt. The 5' of the second half (in pink) was extended with the complementary sequence of the random sequence extended before, and the target sequence of the relaxase we want to attach. Upon folding, the random sequence will fold a double-stranded linker that allows the target be accessible for the relaxase (in purple).

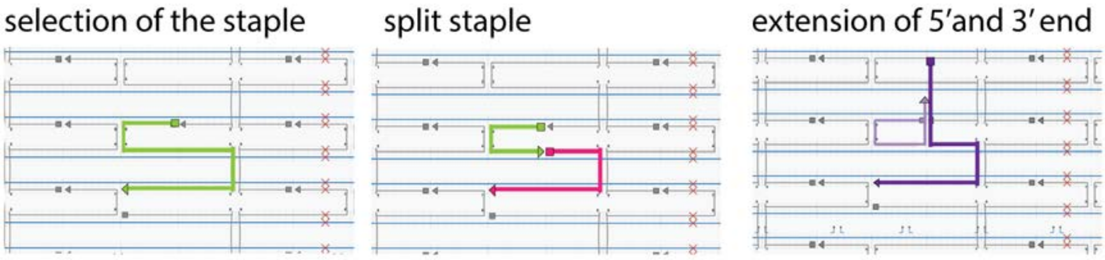


Figure M3. Target design for tcRO. A selected staple of the original design (left) was split into two fragments 8 nt after the Holliday junction (center). At this position, the 3' and 5' ends were extended with a random sequence of 3 to 15 bp. The 5' of the second half (in purple) hold a single stranded region with the target sequence of the desired relaxase (right). Square depicts the 5' end, whereas the triangle the 3' end.

Each 6HB was equipped along its length with specific binding positions for the relaxases TrwC$_{\text{R388}}$, MobA$_{\text{RSF1010}}$, TraI$_{\text{R100}}$ and TraI$_{\text{pKM101}}$. Each binding position (Figure M4) contains five binding sites for the specific relaxase, since it has been shown that the binding yield usually increases with multiple binding sites in close proximity. Each binding position of the tcRO has either one, two or three binding sites for each relaxase (Figure M4).

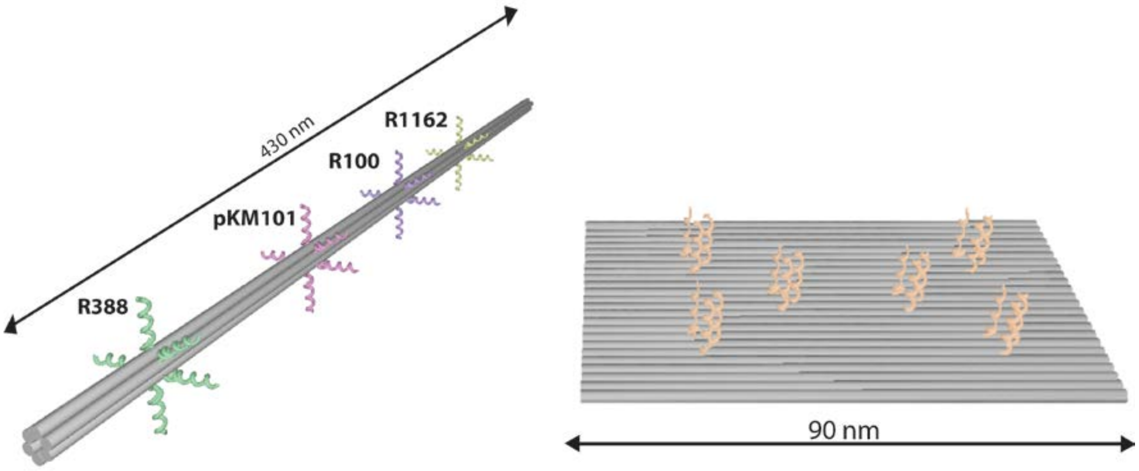


Figure M4. DNA nanostructures used to study the coupling of four relaxases. Left, six-helix bundle (6HB). Right, twist-corrected rectangular origami (tcRO). Each cylinder represents one DNA duplex. The colored extended strands hold the target of relaxases at different positions. The cylinders (duplex DNA) are connected to one another through Holliday junctions every seven (6HB) or eight (tcRO) bp in each cylinder. The dimensions are also shown.

4.1.19 Assembly of DNA nanostructures

More than 200 short "staple" oligomers (Eurofins MWG, Germany) were used to fold a 7249 nucleotide (nt) long single-stranded "scaffold" strand from phage M13mp18 into a tcRO. 3-fold excess of standard staples and a 5-fold excess of extended staples was used for the functionalization with the desired targets. As folding buffer, we used 1x Folding Buffer 2D. A temperature ramp from 70ºC to 20ºC for 1 h was used for tcRO. For purification of the folded structures we used the PEG precipitation protocol from (Stahl et al., 2014). Afterwards the concentration was adjusted to either 45 or 100 nM in 1x Folding Buffer 2D.

For the six helix bundle (6HB), a 7560 nt strand from M13mp18 was used as scaffold. 50 nM of the scaffold were incubated with 3-fold molar excess of standard staples and 5-fold excess of extended staples. As folding buffer we used 1x Folding Buffer 2D. The 6HB mixture was heated up to 80ºC and slowly annealed from 80ºC to 20ºC over 24 h. For purification of the folded structures we used the PEG precipitation protocol from (Stahl et al., 2014). The final concentration was adjusted to 100 nM in 1x Folding Buffer 2D.

Scaffold (100 nM)	25 µl
10x Folding Buffer	5 µl
Standard Staples	13,3 µl
Special Staples	6,7 µl
Total volume	50 µl

Nanostructures were mixed with 2.0 µM of relaxase in 1x folding buffer. If streptavidin was used the sample concentration was 0.1 µM. After 24-h incubation with relaxases, the excess of protein was removed with Amicon centrifugal filters (100 kDa cut-off) for the rectangles and an adjusted PEG precipitation protocol from (Stahl et al., 2014) for 6HB. The initial experiments to check the best linker length were performed similarly, but the initial concentration of nanostructures was 40 nM.

4.1.20 dsNanodevice

TrwC$_{R388}$ directed assembly of nanoribbons was done at two concentrations (1 µM and 10 µM) at room temperature. To generate the intramolecular duplex at the inverted repeat region of the TrwC target of both edges-oligonucleotides, 100 µM oligonucleotides were heated at 95ºC and quickly cooled in ice. The stoichiometry of the mixture of the four oligonucleotides was 1:1:1:1 to ensure the assembly of the monomer, and hybridized at 37ºC during 1 h. The nanodevices were mixed with 3µM and 10µM of TrwC$_{R388}$mCherry, respectively. Polymerization was checked after 1 h, overnight, one day and one week incubation.

4.2 Biophysical methods

4.2.1 Atomic force microscopy (AFM) imaging

An AFM is a type of scanning probe microscope (SPM) with high-resolution for imaging matter at the nanoscale. It was invented in 1986 by Binning et al. at IBM research and it became a versatile tool for nanotechnology. AFM allows the imaging of DNA and RNA, proteins and DNA-protein complexes or any biological sample either on air or in solution in the absence of staining or labeling. It is a “near-field microscope” because it measures interaction forces between atoms within the sample and atoms within the tip of the AFM, in contrast with the “far-field” approaches based on optical or electron microscopy. The sample is probed by a needle-like structure, the sharp cantilever tip. The cantilever oscillates above the sample at its resonance frequency and provides a topographical image of the surface. During scanning, the cantilever is displaced through the x, y and z via a piezoelectric system. The sharpness of the cantilever tip results to be the limiting factor for imaging small objects. AFM uses an optical system for detection, in which a laser spot reflects from the top surface of the cantilever to a photodiode detector. When the tip of the cantilever approximates to the sample, the laser spot deflects. The positional change of the reflected laser beam on the detector is translated into height information for AFM imaging or an interaction force between tip and sample for AFM force spectroscopy.

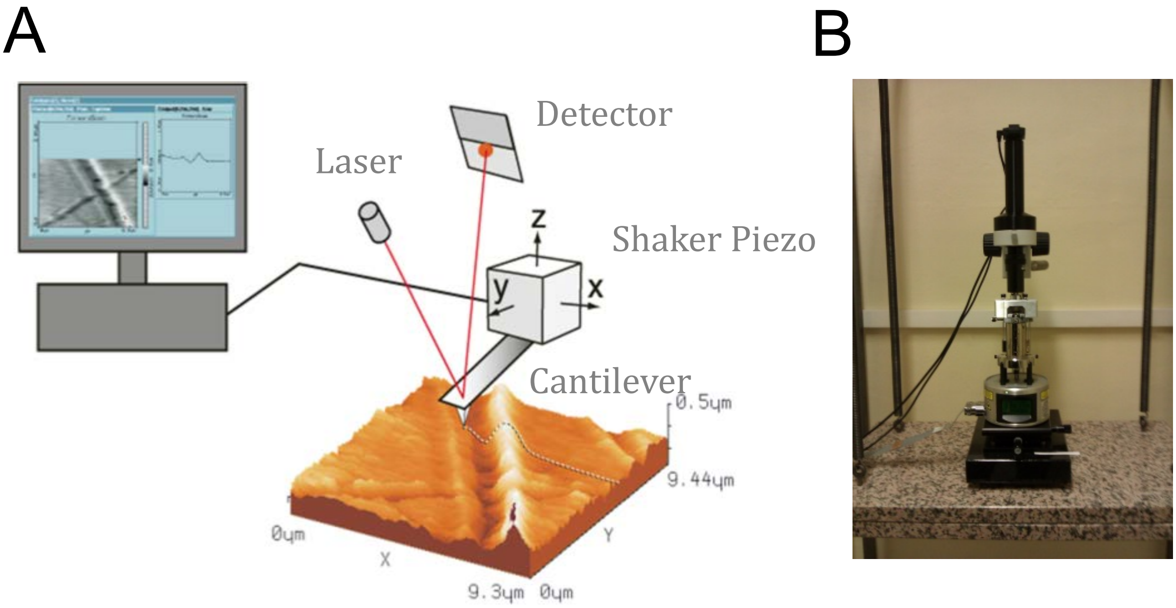


Figure M5. Scheme of the AFM. (A) Schematic representation of the AFM components. See text for details. (B) The SPM Multimode (Veeco) used for acquiring some of the images of this thesis.

AFM images were obtained using an SPM Multimode with Digital Instruments Nanoscope V controller (Veeco) with SNL-10 silicon nitride cantilever chip (Veeco Probes), and an Asylum Research Cypher Scanning Probe Microscope with Olympus micro-cantilevers BL-AC40TS-C2 in tapping mode under liquid conditions. The samples were prepared by adding five microliter of diluted nanostructures (2 nM) onto the surface of a freshly cleaved mica chip and left for approximately 2 minutes to allow adsorption. Imaging was performed after adding fifty microliters of 1x folding buffer onto the mica surface.

Imaging of dsDNA nanoribbons was performed taping imaging mode in air. 30 μl of folding buffer was spotted on freshly cleaved mica and left to be absorbed by the surface. The surface was then air dried under a nitrogen flux. Then, 5 μl of a diluted sample (3 to 5 nM) was spotted on the mica and left to adhere for 5 minutes. Afterwards, it was also dried with nitrogen flux.

4.2.2 Transmission electron microscopy

For the analysis of 3D nanostructures, we used transmission electron microscopy (TEM). Under the electron microscope, biological samples are bombarded by a stream of electrons in high vacuum. The image is magnified and detected by a CCD camera.

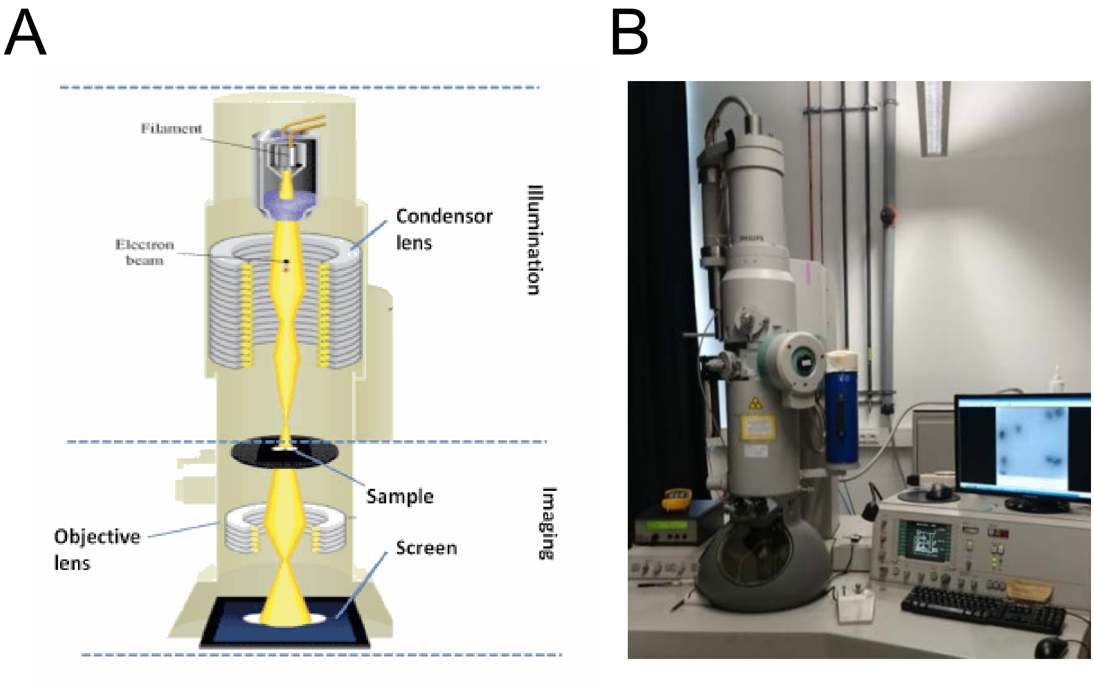


Figure M6. Transmission electron microscopy. (A) Schematic diagram of TEM. Electrons are generated in a tungsten filament, and are condensed to illuminate the sample. The resulting images are visualized with a CCD camera in the Philips CM100 TEM show in (B).

A critical step before sample preparation is to glow-discharge the grids to ensure that the sample coats the grid smoothly (Figure M7). Glow-discharging/plasma-cleaning removes any residual hydrocarbons and makes the grids hydrophilic. Formvar-supported, carbon-coated Cu400 TEM grids were glow-discharged for 25 s at approximately 20 W. Afterwards, 5 µl of reaction products with a final concentration of 3 nM were adsorbed on glow-discharged grids (Science Services, Munich, Germany) for 30 s. The excess liquid was soaked up with filter paper, and the sample was negatively stained using 25 µl of a 2% aqueous uranyl formate (UFO) solution with 25 mM sodium hydroxide for 40-45s. Grids were dried and stored at RT until imaging.

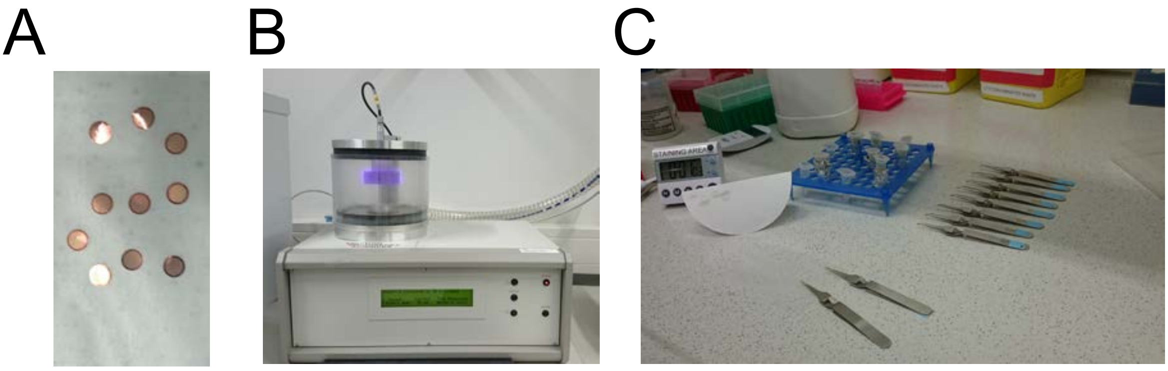


Figure M7. TEM sample preparation. (A) Carbon-coated TEM grids before being (B) plasma-cleaned in a glow-discharge apparatus. (C) Samples during negative staining.

Imaging was performed using a Philips CM100 transmission electron microscope at 100 kV at the TU Munich (Figure M6B). An AMT 4 Megapixel CCD camera was used for acquiring images. Imaging was performed at x28500 and x39000 magnification. Processing was carried out using the plugin “tools for microscope” for Java-based software ImageJ.

4.2.3 Fluorescence microscopy imaging

For DNA-PAINT imaging, samples were prepared by mixing 70 nM 6HB, 1.6 µM biotinylated staples with 2 µM TrwC$_{R388}$mCherry and 2 µM MobA$_{R1162}$mCFP. After overnight incubation, samples were purified and diluted. A piece of coverslip and a glass slide were sandwiched together by two strips of double-side tapes to form a flow chamber with inner volume of 15 µl. First, 20 µl of biotin-labeled bovine albumin (1 mg/mL, dissolved in Buffer A (10 mM Tris-HCl, 100 mM NaCl, 0.05% Tween-20, pH 7.5) was flowed into the chamber and incubated for 2 min. The chamber was then washed using 20 µl of Buffer A twice. 20 µl of streptavidin (0.5 mg/mL, dissolved in Buffer A) was then flown through the chamber and allowed to bind for 2 min. After washing twice with 20 µl of Buffer B (5 mM Tris-HCl, 10 mM MgCl$_2$, 1 mM EDTA, 0.05% Tween-20 pH 8), 20 µl of biotin-labeled nanotubes with relaxases bound (100 pM and 50 pM in Buffer B) were finally flown into the chamber and incubated for 2 min. The chamber was washed using 20 µl of buffer B three times, sealed with nail polish and mounted on the sample stage of a microscope with the coverslip facing the objective.

Fluorescence imaging was carried out on an inverted microscope Olympus IX-71 (Olympus Corporation, Japan) applying an objective-type total internal reflection fluorescence (TIRF) configuration with an oil-immersion objective (UPlanSApo 100X, NA 1.40 Oil, Olympus). Samples were imaged using microscopic chamber slides (LabTek, Nunc). Additional lenses were used to achieve a final imaging magnification of ~160 fold, corresponding to a pixel size of 145 nm. Three lasers (λ=650, λ=561 nm, λ=488 nm) were used for TIRF excitation. Fluorescence light was spectrally filtered with an emission filter (HQ 700/75 M or BrightLine HC 582/75, AHF Analysentechnik) and imaged on an EMCCD camera (Andor Ixon DU-897). The laser intensity was set as low as possible to avoid photobleaching while being sufficiently high to obtain a reasonable signal-to-background ratio, which was in the range of 4–8 mW. 5000 frames were recorded at a frame rate of 50 Hz for each color channel sequentially. Exposure times were 1 s for 488 nm excitation and 1s for 561 nm excitations. To obtain reliable data, usually 5 positions of the sample chamber were analyzed. Images were processed using ImageJ.

CHAPTER 1

Applications of Relaxases in DNA-Nanotechnology

RESULTS

Chapter 1: Applications of relaxases in DNA-Nanotechnology

One of the most important areas in the field of DNA nanotechnology is the search for proteins that can easily be assembled within DNA nanostructures in a specific and programmable manner. In this thesis, we have developed the initial stages towards an experimental demonstration that HUH relaxases are a promising family of proteins for the bottom-up nanoassembly of DNA nanoarchitectures. Relaxases possess some of the features that DNA nanotechnology is searching for: (1) HUH relaxases recognize a short DNA sequence of about 14 nucleotides with nanomolar affinity. (2) Relaxases form a covalent bond with the DNA after cleaving the sequence they recognize. The covalent bond is a stable phosphotyrosine linkage between the catalytic tyrosine and the 5´phosphate of the cleaved ssDNA. (3) The lifetime of the covalent adduct intermediate is long. The half-life for TrwC$_{R388}$ complexed with oligonucleotide 25+8 has been determined to be 11 h (Guasch et al., 2003). (4) Some relaxases have a thermally stable core, which makes them non-denaturable in an ample range of temperatures (Arrondo et al., 2003; Lorenzo-Diaz et al., 2011). (5) The pH range of activity of nucleophilic attack to the DNA is also broad. For instance, TrwC$_{R388}$, our model relaxase, is active between pH 4.6 and 8.6 (data not shown) and TraI$_F$ over a pH range from 5.5 to 9.7 (Stern et al., 2004). (6) Relaxases are bifunctional proteins, having a N-terminal relaxase domain composed of roughly 300 residues fused to a C-terminal domain with various biochemical activities (helicase, primase, etc). The relaxase domain is always monomeric in solution. Therefore, it can be fused with heterologous proteins without losing activity. As a proof of concept, we addressed the purification of different relaxases fused to fluorescent proteins and tested their organization on DNA nanoarrays.

This chapter begins with the basic theory of binding equilibrium of proteins to its ligands. We then show the results of TrwC$_{R388,}$ arrangement on planar nanostructures. Next, we describe the orthogonal binding of several relaxases to 2D and 3D DNA nanostructures. Finally, we analyzed the polymerization of DNA devices performed by relaxases.

5. Assembly of Relaxases on DNA nanostructures

The question at the beginning of this project was: can relaxases be a tool for bioconjugation of any protein of interest to DNA nanostructures? Our first task was the design of targets on different DNA nanostructures. Both rectangular planar and tubular nanostructures were chosen for the assembly of relaxases (Figure R1). We modified some staple oligonucleotides at the positions we wanted the target to protrude. The desired staple strands were extended with the relaxase target sequence at the 5' end (see Experimental procedures) because relaxases remain covalently attached to the 5'-end of the *nic* site after cleavage. The nomenclature for the relaxase targets uses two numbers and a "plus" symbol, e.g. 14+5. The first number indicates the number of nucleotides before the *nic* site, represented by the symbol "+".This DNA segment contains either the complete IR (usually if n>16) or the distal arm and the U-turn (if n<15) (Figure 19). The number after the plus sign indicates the number of nucleotides of the DNA segment that contains the "staple", which remains bound to the relaxase after cleavage.

We studied the arrangement of the relaxase domains of TraI$_{R100}$, TraI$_{pKM101}$, MobA$_{R1162}$ and TrwC$_{R388}$ on DNA nanostructures. They belong to the HUH endonuclease family being the HUH motif located in an antiparallel five strand β-helix core and the catalytic tyrosine is placed in an α-helix close to the N-terminal end of the relaxase (see Figure 20). TrwC$_{R388}$, TraI$_{pKM101}$ and TraI$_{R100}$ belong to the MOB$_F$ family of relaxases and they contain two catalytic tyrosines (Figure R1A). MobA$_{R1162}$ belongs to the MOB$_Q$ family and contains one catalytic tyrosine. These relaxases bind specifically to the *nic*-region of the plasmid that encodes it and catalyze a site-specific cleavage reaction (Figure R1B-C). The relaxase-DNA covalent complex migrates slowly in a SDS-PAGE gel than the free or unreacted relaxase (Figure R1C). The cleavage yield of these relaxases can be found in supplementary material (Table S15 and S16).

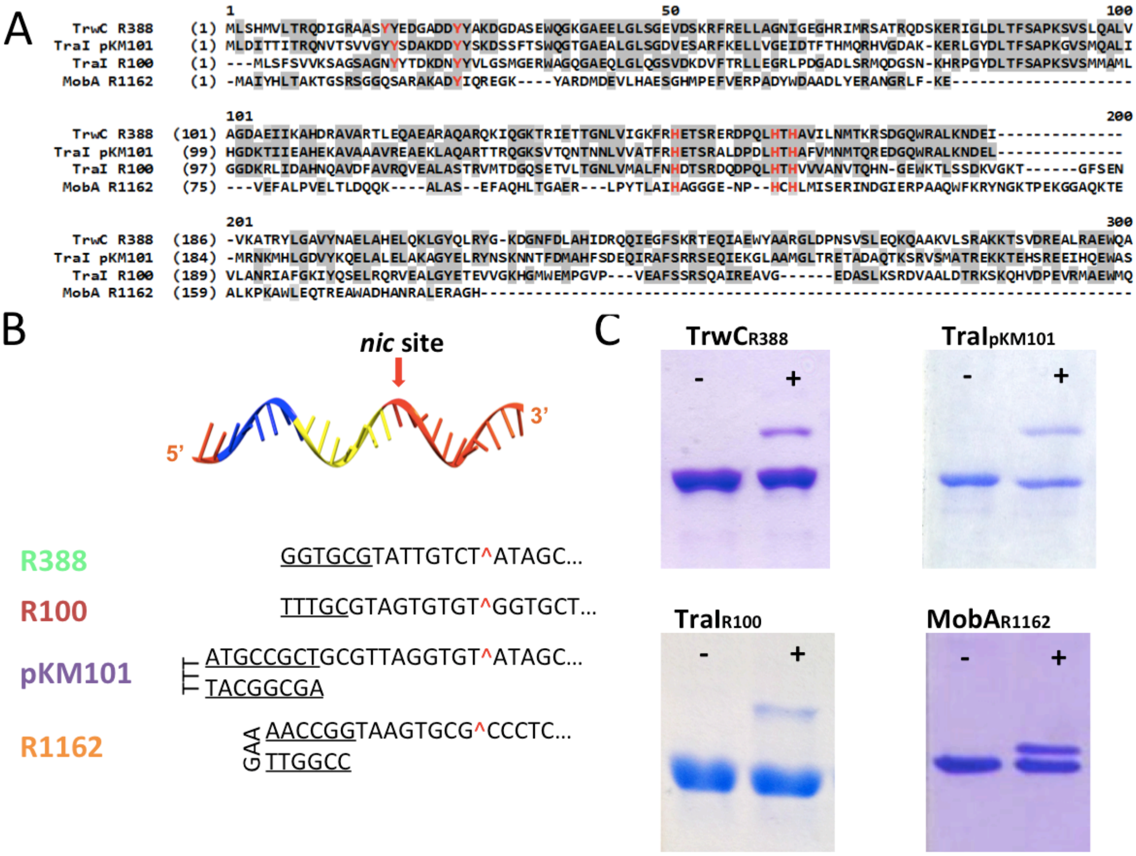


Figure R1. Characterization of relaxase proteins and their targets. (A) Amino acid sequences of relaxases used in this thesis. Red letters highlight the residues that are involved in the catalytic mechanism: the catalytic tyrosine(s) (Y) at the N-terminal region form a covalent linkage to the 5'-phosphate of the *nic* site after cleavage. Since these relaxase proteins are HUH endonucleases, they contain a conserved triad of histidines H+HUH involved in metal coordination. Identical residues are shadowed. (B) Recognition sequences of these relaxases. R388 W(14+5), R100 W(14+6), pKM101 W(30+5), R1162 W(23+5). Underlined nucleotides are IRs. The *nic* site is depicted with ^. (C) SDS-PAGE shows the covalent complexes obtained when relaxases are incubated with their target DNAs. Relaxases cleave the phosphodiester bond at the *nic* site and remain covalently bound to the DNA upstream of the *nic* site by a phosphotyrosine linkage. The free protein is shown in lane "-". The covalent complex relaxase-DNA is the second band with lower mobility (lane "+").

The knowledge of relaxases gathered during the last decades allows the prediction that relaxase attachment yields follow a two-stage model. First, relaxases approximate the surface of the origamis, binding the targets they specifically recognize. Second, the relaxase cleaves the target in a Mg^{2+}-dependent manner. The first process (binding) could be assumed irreversible, considering its nanomolar affinity, while the second stage (cleavage) implies a balance. The target is supposed to be continuously cleaved and religated, re-stablishing the original substrate (Pansegrau and Lanka, 1996). After treatment with proteinase K and SDS, 10-30% of the relaxase molecules remain covalently bound to the DNA. These results are in accordance with the percentage of covalent complexes analyzed by SDS-PAGE gels. Therefore, we assumed that when a protein is bound to its target, 10-30% of the relaxases are covalently bound to the origami. Indeed, there is no way to distinguish between covalent and non-covalent bound relaxases on nanostructures. The approaches to displace the balance to the formation of a covalent complex relaxase-DNA include “suicide substrates” that will be discussed in section 5.2.3 and 5.3.4.

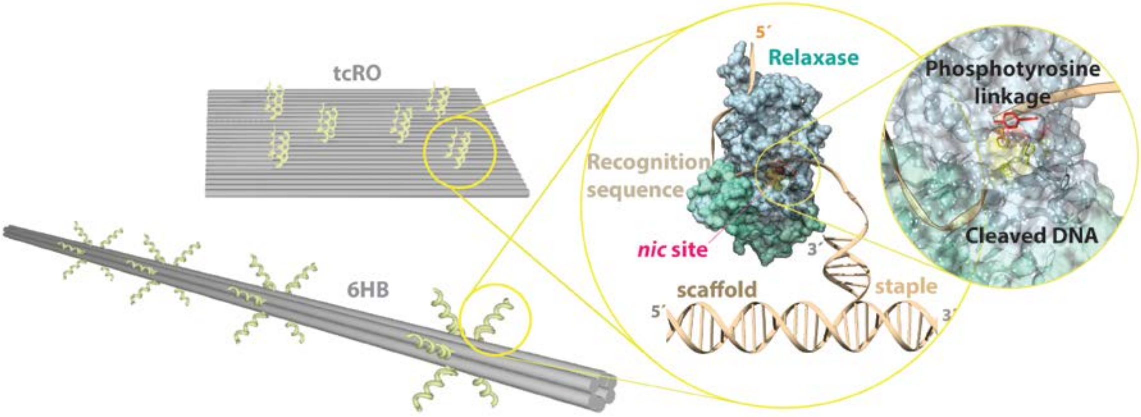


Figure R2. Scheme of relaxase attachment to DNA nanostructures. A planar rectangular or a tubular nanostructure, named tcRO and 6HB respectively, were the platforms for attachment of relaxases. The extended staples containing the relaxase recognition sequence, are designed to protrude out of the base of the nanostructure, stabilized by a double stranded linker. After binding, the relaxase locates the *nic* site at the active center of the relaxase (in yellow). Upon cleavage, the catalytic tyrosine (in red) forms the phosphotyrosine linkage with the 5-phosphate of the *nic* site.

The platforms for attachment of relaxases were a planar rectangular or a tubular nanostructure, named tcRO and 6HB respectively (Figure R2). All experiments were performed with 12.5 mM Mg^{2+}. Mg^{2+} is required for both the nucleophilic attack of the relaxase to the DNA that enables the formation of the covalent linkage with the DNA and the correct assembly of the origami. We utilize the N-terminal domain of relaxases $TrwC_{R388}$, $TraI_{R100}$, $MobA_{R1162}$ and $TraI_{pKM101}$ fused to monomeric fluorescent proteins (mCFP, MCherry and mKATE) (purification of relaxases is shown in supplementary figures S2 and S3). As a negative control, we incubated these relaxases with DNA origamis without the relaxase recognition sequence. These controls revealed that relaxases do not bind to these DNA nanostructures neither when there is no target protruding nor when there is a misfolded region on the origami (supplementary figures S6 and S7).

5.1 Modelling the Binding Equilibrium of relaxase TrwC$_{R388}$ on DNA Origamis

We expect a concentration-dependent equilibrium between relaxases and nanostructures. To better predict the binding efficiency of relaxases to several targets protruding from a nanostructure, we modeled the binding equilibrium with the law of mass action. The fraction of complexes obtained would be defined by the dissociation constant, as shown below.

$$\underset{Protein}{P} \quad + \quad \underset{Ligand}{L} \quad \rightleftarrows \quad \underset{Complex}{C}$$

The dissociation constant is defined as $K_D = \frac{[P][L]}{[C]}$

Considering the initial concentrations of ligand (l_0) and protein (p_0) as known:

$$l_0 = l + c \qquad p_0 = p + c$$

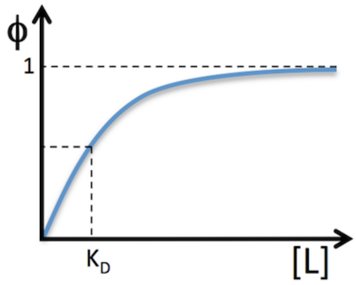
Then, the constant can be defined as: $K_D = \frac{p\,l}{c} = \frac{(p_0 - c)(l_0 - c)}{c}$

The solution is: $c^2 - (p_0 + l_0 + K_D)c + p_0 l_0 = 0$

$$c = \frac{1}{2}\{p_0 + l_0 + K_D - \sqrt{(p_0 + l_0 + K_D)^2 - 4p_0 l_0\}}$$

The fraction of complexes can be calculated as:

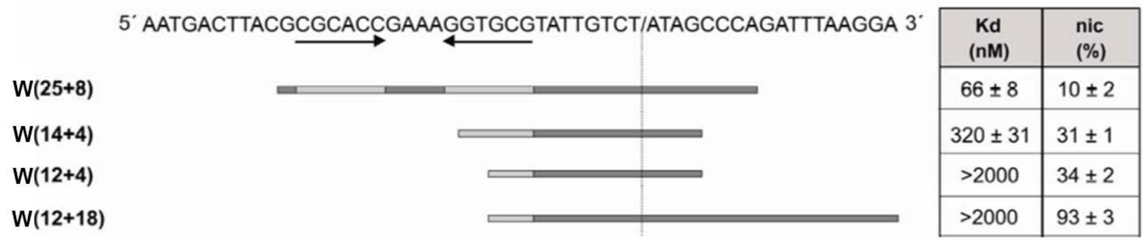
$$\phi = \frac{c}{p_0} = \frac{c}{p+c} = \frac{1}{1+ {}^{p}/_{c}} = \frac{1}{1+ {}^{K_D}/_{[L]}} = \frac{[L]}{[L]+ K_D}$$



$[L] << K_D \qquad \phi \approx {}^{[L]}/_{K_D}$

$[L] >> K_D \qquad \phi \approx 1$

Friedrich Simmel wrote a Matlab script in which the K_D for the target chosen, and the initial concentrations of relaxase and nanostructures determine the probability that a given number of targets are occupied (see supplementary methods). Due to their intrinsic attachment to the DNA, specificity and stability, we have studied if relaxase $TrwC_{R388}$ from plasmid R388 could be used for protein decoration on DNA nanostructures. It has been proven that $TrwC_{R388}$ is covalently bound to the ssDNA downstream from the *nic* site with different efficiency depending on the 5'- and 3'-end length of the target chosen. Previous studies have reported that 3- to 9-fold higher proportion of $TrwC_{R388}$-covalent complex is found with targets W(14+4) or W(12+18), relative to W(25+8) target, even though the latter has a significantly lower K_D (Figure R3) (Lucas et al., 2010).



	Kd (nM)	nic (%)
W(25+8)	66 ± 8	10 ± 2
W(14+4)	320 ± 31	31 ± 1
W(12+4)	>2000	34 ± 2
W(12+18)	>2000	93 ± 3

Figure R3. TrwC-mediated cleavage of oligonucleotides embracing the R388 *nic* site. The figure shows the dissociation constant (Kd) and percentage cleavage (nic) of the oligonucleotides represented below the sequence by horizontal lines. The DNA sequence of the R388 *nic* site is shown at the top. The inverted repeat IR_2 is symbolized by horizontal arrows below the DNA sequence. The *nic* site is represented by a slash in the sequence. Adapted from (Lucas et al., 2010).

Considering the target W(14+5) that shows a K_D= 320 nM (Lucas et al., 2010), and a fixed concentration of nanostructures of 20 nM, the theoretical concentration of relaxase to reach 90% binding efficient to one target per nanostructure is 2 μM (Figure R4, red squares). Similar results were obtained when the concentration of nanostructures was fixed to 50 nM and 100 nM (data not shown). In the case of 2 μM relaxase concentration, the probability stays constant even if the target concentration would drop to 20 nM (Figure R4, blue circles). Due to this result we used nanostructure concentrations between 20 nM and 100 nM and a fixed relaxase concentration of 2 μM.

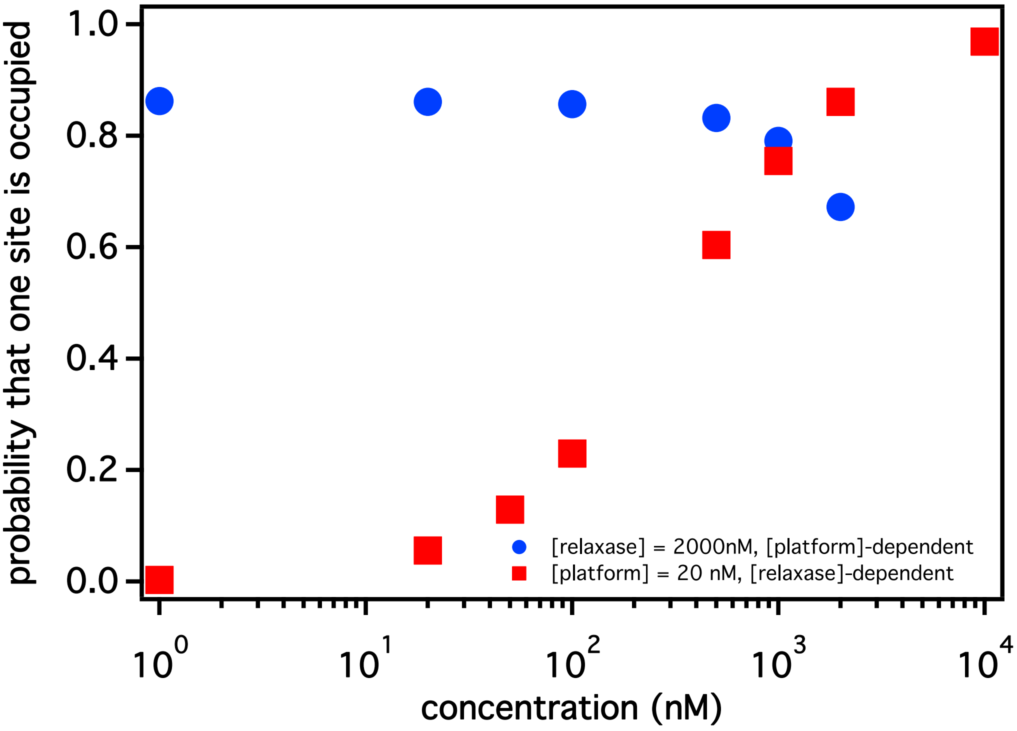


Figure R4. Probability that a single site is occupied when the concentration of nanostructures (red squares) or the concentration of relaxase TrwC$_{R388}$ is fixed (blue circular marker). The considered target is W(14+5) which shows K_D=320 nM (Lucas et al., 2010).

Our initial approach used W(12+18) as the relaxase target, because the reported percentage of covalent complexes was 93%. However, the model predicts low efficiency of binding because its K_D is >2000 nM. The higher percentage of relaxase-origami complexes with less initial concentration is expected for target W(24+5) and fewer for W(14+5). Therefore, R388-targets used are either W(14+5) or W(24+5), shortened to 14+5 and 24+5 respectively.

Furthermore, using Simmel script, we analyzed the theoretical yield when several targets are distributed in a nanostructure, the probability that up to 12 binding sites are occupied is shown in the following graph.

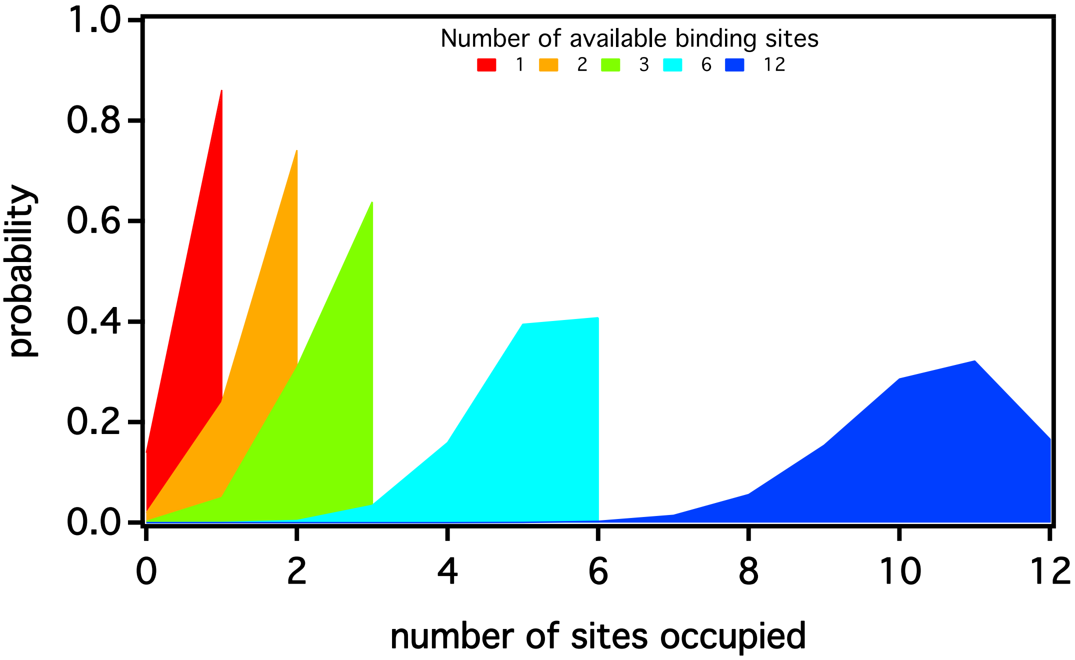


Figure R5. Probability of binding site occupation in origami structures containing multiple relaxases bindig sites (with the target 14+5). The figure shows the probability of occupation of each binding site, when the origami structures were incubated with 2 μM of $TrwC_{R388}$.

5.2 Decoration of Rectangular Tiles with $TrwC_{R388}$

5.2.1 Effect of the Linker Length

To investigate the assembly of relaxase $TrwC_{R388}$ to DNA nanostructures, we assembled a twist-corrected rectangular origami (tcRO) with several 24+5 target-staples perpendicular to the origami in just one face, as depicted in Figure R6A. Since ssDNA adopts a random 3D conformation, some neighbor staples of the target staples were extended to allow the generation of a double stranded linker (ds-linker) on the base of the target. A detailed description of the design of the relaxase target-staples with ds-linker is in experimental procedures, (section 4.1.18). The linker has two advantages. First, it becomes a rigid base for the target. Second, it limits the chance that the target flips to the opposite surface of the origami. Seven R388 targets were designed to protrude from the origami surface. 24+5-staples were placed at 9,6 nm from each other in the central line of the origami tile. A ds-linker that varies from 3 (staple 1, named #1) to 15 bp (staple 7, named #7) in steps of two base pairs was included (Figure R6A). In order to avoid any unwanted interaction of the linkers random sequences were used. The seven target nanoarray was incubated with $\text{TrwC}_{\text{R388}}$mCherry together with streptavidin for 3 h before filtering. AFM images revealed relaxase-origami interaction in every origami (Figure R6C).

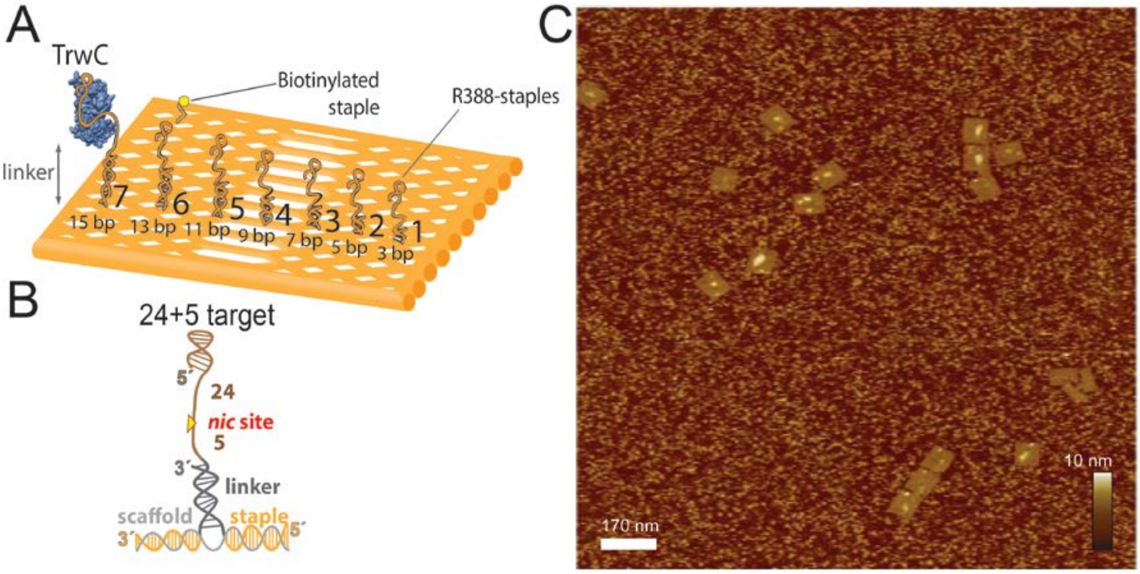
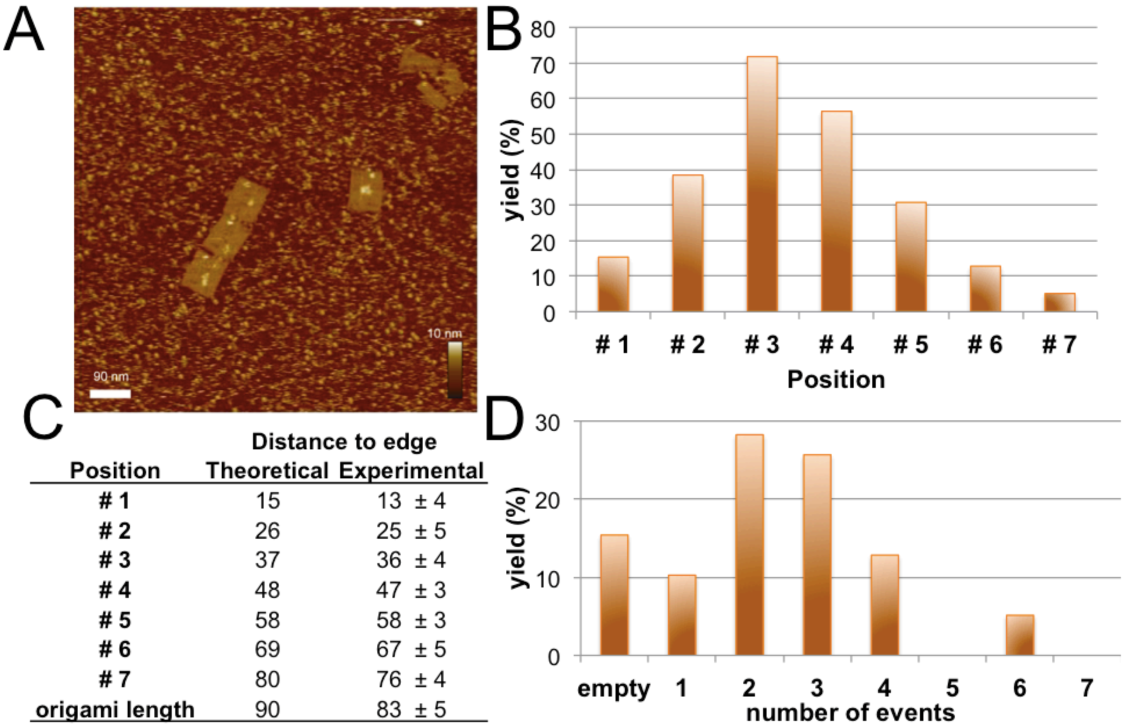


Figure R6. Effect of the linker length of a W24+5 target-staple nanoarray. (A) Scheme of the organization of the nanoarray. The tcRO platform is shown in orange, with seven ds-linkers of varying length added to a series of seven W24+5 staples. Linkers vary in length from 3 bp (staple 1) to 15 bp (staple 7) in steps of two bp. All seven staples were included in the tcRO central line. In the upper left corner, an additional staple strand modified with biotin, was included to provide a marker postion to identify the other staples, when streptavidin was added (see supplementary Figure S8 for further details). (B) Scheme of a 24+5-staple with 11 bp linker included. (C) AFM image of the nanoarray incubated with $\text{TrwC}_{\text{R388}}$mCherry and streptavidin. The bright spots disclose the proteins bound to the nanoarray. The mica surface is shown as dark brown color. Scale bar 170 nm.

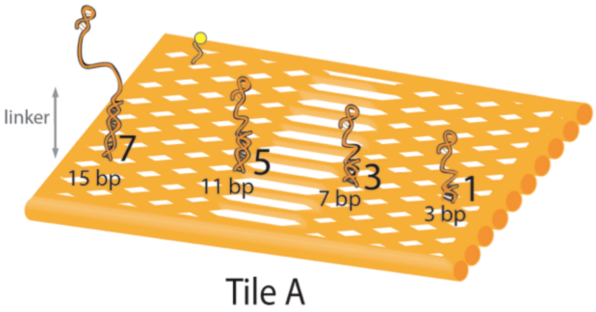
Relaxases were arranged preferably in centrally located targets, those containing medium size linkers (7 to 11 bp) (Figure R7A and B). Some target positions appeared barely occupied such as #1(3bp) or #7(15bp). The distances were in accordance to the theoretical position (Figure R7C). Nevertheless, seven bright spots were not observed in any origami (Figure R7D). This reflects that relaxases seem to have a preference for central targets.



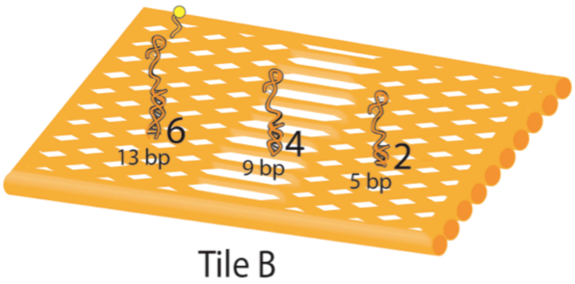
Position	Distance to edge Theoretical	Experimental
# 1	15	13 ± 4
# 2	26	25 ± 5
# 3	37	36 ± 4
# 4	48	47 ± 3
# 5	58	58 ± 3
# 6	69	67 ± 5
# 7	80	76 ± 4
origami length	90	83 ± 5

Figure R7. Analysis of the events and binding yields of TrwC$_{R388}$mCherry to the tcRO with different linker lengths. See Figure R5 for details of tcRO construction and supplementary Figure S9 for details of the analysis. (A) Zoom of the previous AFM image. The bright spots disclose the proteins bound to the nanoarray: streptavidin on the left corner, the relaxases at the central line. Scale bar 90 nm. (B) Distribution of the relaxases on the tcRO. Most origami particles show relaxases on cental binding positions (#3 and #4) and only a small fraction on positions #1 and #7. (C) Average distance between the center of protein and the right edge on tcRO. The table shows the theoretical and experimental distance measurement of each binding position. (C) Binding yield of TrwC$_{R388}$mCherry to tcRO equipped with seven binding sites. Total of 40 origami particles were analyzed.

To explore if site specificity was caused by steric interactions between targets, "odd" and "even" targets were incubated separately. For this purpose, we built origamis named "tile A", with four target-staples separated by approximately 20 nm containing 3bp, 7bp, 11 bp and 15 bp linkers, respectively, and "tile B" origamis, with three target-staples separated by the same distance and containing 5pb, 9bp and 13 bp linkers. Both tiles contained the biotinylated staple at the bottom-right corner defining the position of the staple with the tallest linker (15 bp in A and 13 bp in B). From 50 "Tile A" origamis analyzed, 70% of them possessed a relaxase bound to the 11 bp linker. On the other hand, 55% of the "tile B" origamis contained TrwC$_{R388}$mCherry bound to the 9 bp linker, and the other positions were occupied equally (from 29 tiles) (see Figure R8).



Tile A	Cases	Percentage
# 1 (3pb linker)	1	2,0
# 3 (7pb linker)	13	26,0
# 5 (11pb linker)	35	70,0
# 7 (15pb linker)	1	2,0
Sum	50	



Tile B	Cases	Percentage
# 2 (5pb linker)	7	24,1
# 4 (9pb linker)	16	55,2
# 6 (13pb linker)	6	20,7
Sum	29	

Figure R8. Scheme and quantification of the binding location of TrwC$_{R388}$mCherry on Tiles A and B. Positions were calculated by measuring the distance of the bright spot to the right edge and compared with the theoretical distance. The average error for the positions analyzed was +/- 4.9 nm in tile A, and 5,4 nm in case of Tile B. This deviation is in accordance with a flexible linker. The biotinylated staple is depicted with a yellow circle.

The properly arranged TrwC$_{R388}$s where located at 50 and 60 nm from the right border, that corresponds to nine (tile B, # 4) and eleven (tile A, # 5) bp linker, respectively. These results either mean that short and long linkers are unstable for proper binding of TrwC$_{R388}$ or that TrwC$_{R388}$ binds preferentially to central positions rather than those at the boundaries. In order to distinguish between these possibilities, we performed single-molecule experiments varying the length of the linker in any given position.

5.2.2 Single Molecule Experiments for Studying the Best Linker Length

A tcRO holding one target-staple at position #2 was analyzed. We tried different single- and double-stranded linker lengths (ss-linker and ds-linker) to determine if the distance between the relaxase target and the base of the staple implies a change of the binding yield. Moreover, our goal was also the characterization of the targets 24+5 and 14+5 (Figure R9). In order to make a comparison of the binding of the relaxase to its target staples with the binding of streptavidin to this biotin-conjugated staple (the most-broad used protein in nanotechnology), a staple located at the left corner of the nanostructure was modified with biotin (Figure R9A).

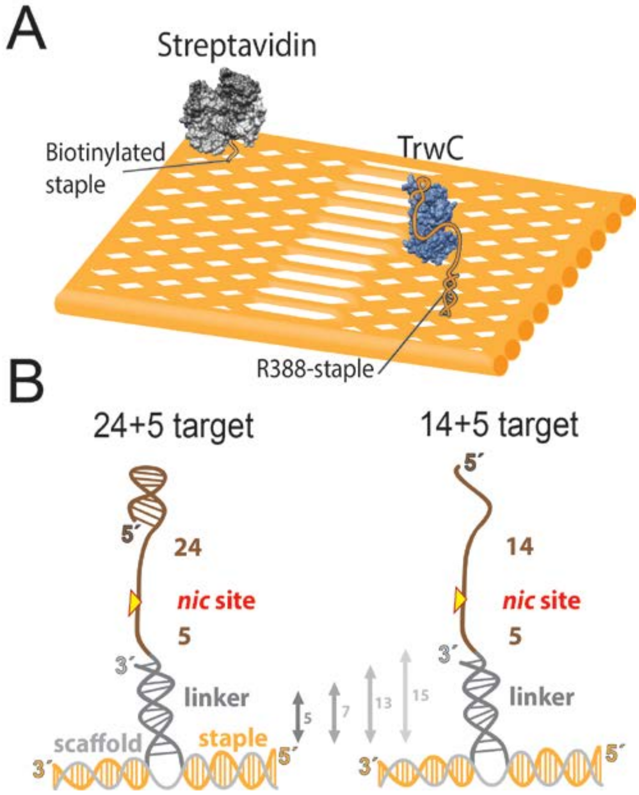


Figure R9. Architecture of the target-staples for TrwC$_{R388}$ binding. (A) Disposition of R388 and biotinylated staples in the tcRO nanostructure. TrwC$_{R388}$ (in blue) bind its R388-target staple on the rectangular nanoarray, and streptavidin (in gray) the biotinylated staple. (B) Different target designs tested for TrwC$_{R388}$ binding. The target is depicted in brown.Target 24+5 possesses the complete IR that enables a lower K$_D$ for TrwC$_{R388}$ binding. Without IR, target 14+5, the K$_D$ is higher and resulted in an increased cleavage. ss-linker (not depicted) and ds-linker lengths (in gray) with up to fifteen base pairs were tested for both targets.

Since the percentage of covalent complexes is proportional to the K$_D$ of the target (see section 5.1), we reasoned that extended incubation times could increase the binding yield for target 14+5. The nanoarray was incubated overnight with TrwC$_{R388}$mCherry together with streptavidin as described in experimental procedures. AFM images showed bright spots in central positions of tcROs indicating TrwC$_{R388}$arrangement at the correct position (Figure R10).

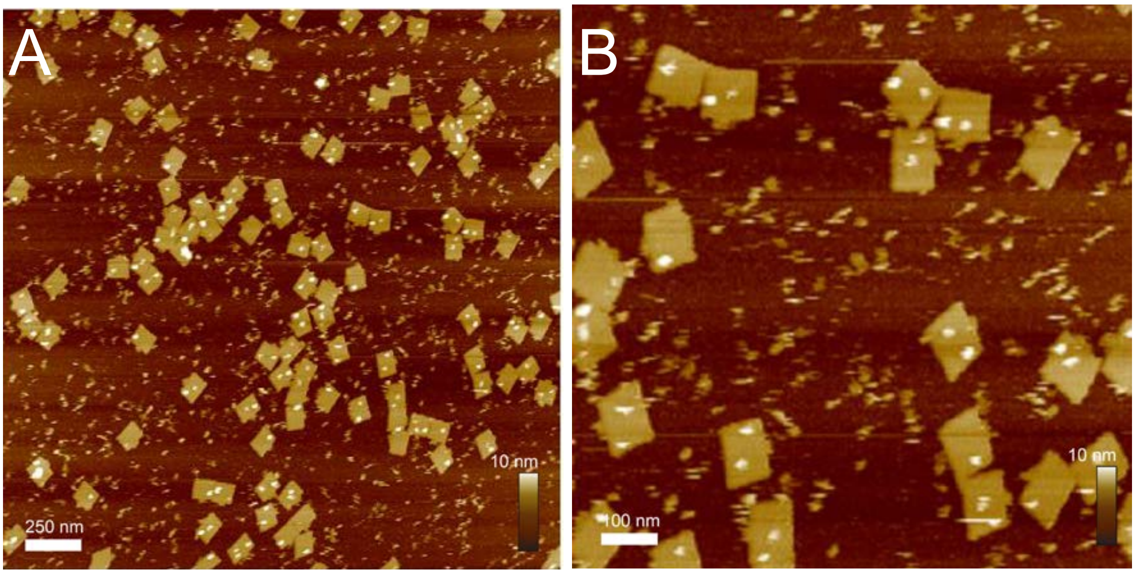


Figure R10. AFM image of tcRO with 14+5 target and a linker of 5 pb incubated with TrwC$_{R388}$mCherry and streptavidin. (A) Nanostructures with both proteins bound. (B) Zoom image of (A). See Figure R8 for details of tcRO design.

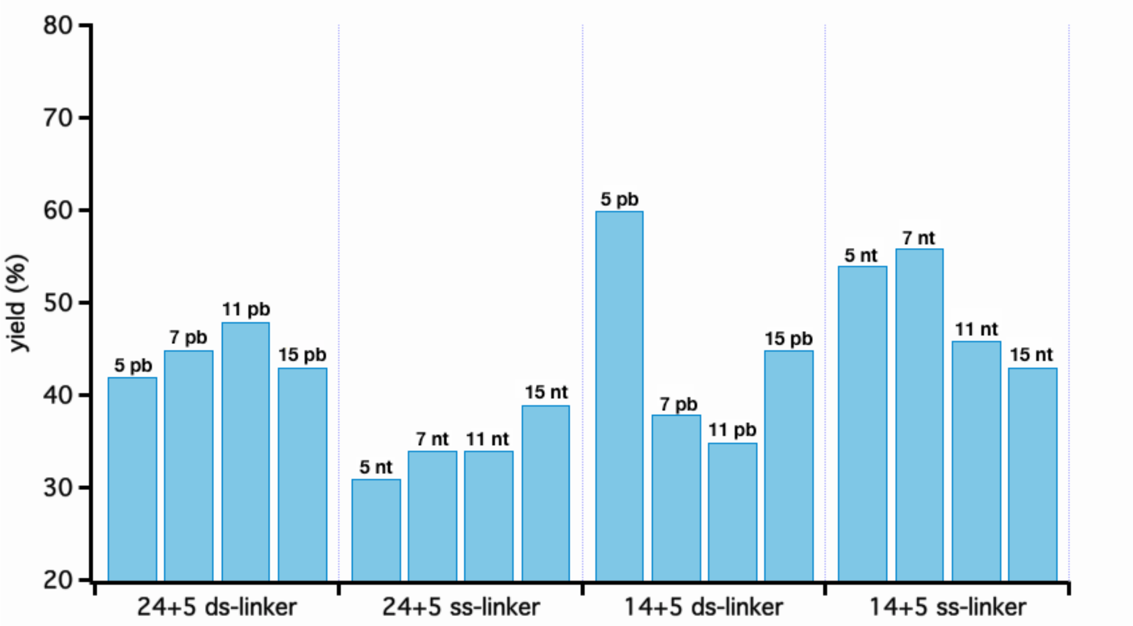


Figure R11. Yield of the different binding sites analysed. The chart shows the resulting binding yield of TrwC$_{R388}$mCherry for various target designs, namely 24+5 and 14+5 with ss-linker and ds-linker each and different linker lengths. In total between 600 and 1900 origami structures were counted for the different samples. TrwC$_{R388}$ reached 45% (blue bars) on average and streptavidin reached 90% binding yield (see supplementary Figure S12 and Table S8 for detailed analysis).

Exhaustive analysis of more than 700 tcROs of each design was carried out (Figure R11 and supplementary S12). Approximately 45% of the tcROs possess both proteins at the correct localization. $TrwC_{R388}$ was found in 30-60 % of the nanostructures while 67-92% of the nanostructures were decorated with streptavidin. The latter result is in accordance with other studies were just one biotinylated staple was used (Voigt et al., 2010). The length of the ds-linker of the 24+5-staples does not influence the binding yield of the nanostructures. A slight decrease of the percentage is noticed when the ss-linker was used. The results with the ds-linker of 14+5-staples revealed a variable yield up to 60% while ss-linkers reach percentages of 43 to 56%. Altogether, a decrease in the efficiency is shown when the linker length increased to fifteen nucleotides, except for 25+5 with ss-DNA linker. Indeed, 14+5-staples with 5 bp of linker length give the best yield, 60% of nanostructures decorated with $TrwC_{R388}$. Taken these results together, $TrwC_{R388}$ seem to have similar binding yield when a short linker is used (5 bp) at the rim (# 2) than when a long one (9-11 bp) is in the central positions (# 4 and # 5).

5.2.3 Domino Tiles for Studying Position Dependency

The interaction between relaxase $TrwC_{R388}$ and its target on a DNA origami is likely to be position dependent. In order to have a better understanding of $TrwC_{R388}$ arrangement and study the effectiveness of binding in different locations, a new tile (called domino) was designed. The “domino tile” is a six-dot nanostructure with 14+5-staples in four corners and two central positions (schematized in Figure R12) and either two or three target-staples per position. Every target had a ds-linker of five bp.

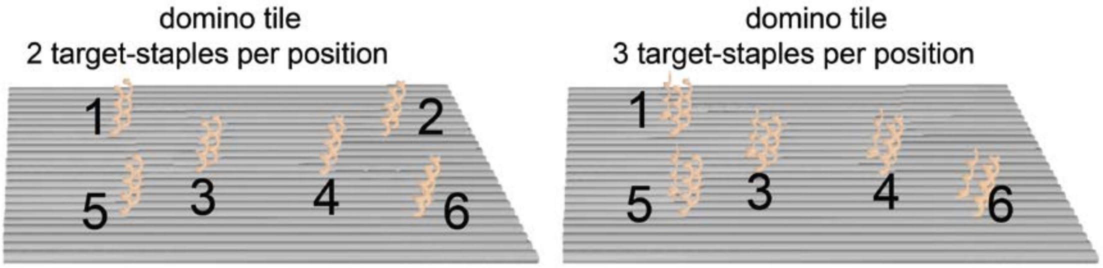
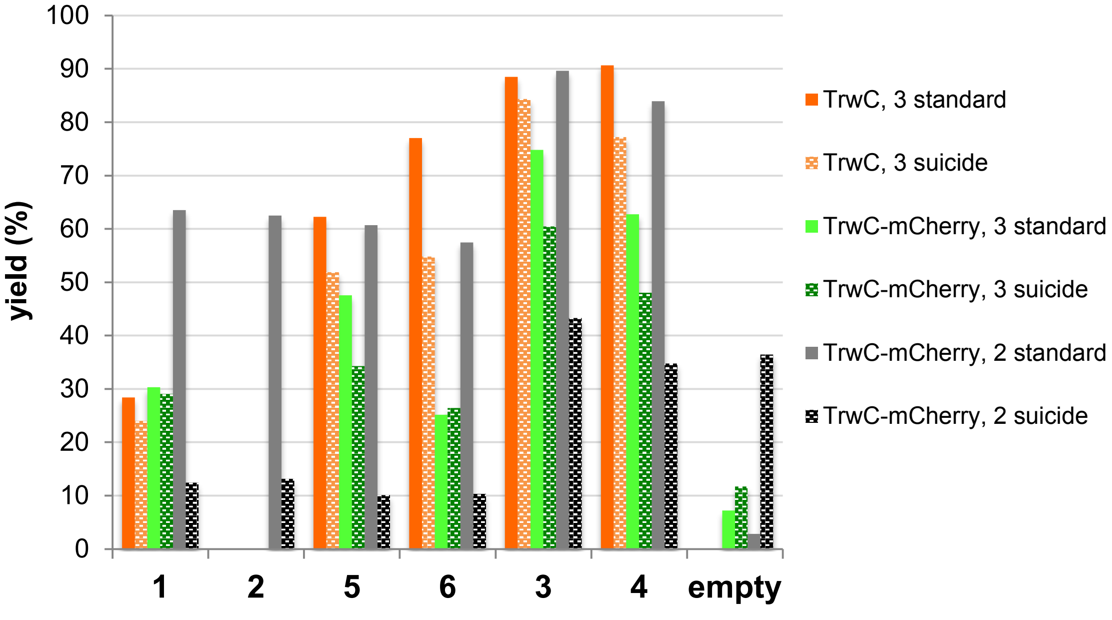


Figure R12. Illustration of the domino tile. 3D scheme of the domino tiles eqquiped with two target-staples per position (left) and three targets-staples per position (right) each. The boundary positions are numbered 1, 2, 5 and 6, whereas the central positions are numbered 3 and 4. Positon 2 is missing in the domino tile with 3 target-staples. Targets are design to protrude from the same face of the nanostructure.

It has been reported that binding yield improves when several targets are in the proximity (Rinker et al., 2008; Saccà et al., 2010). In our preliminary results (data not shown), just one binding site per position revealed poor effectiveness. Thus, domino tiles carrying either two or three target-staples per position were assembled. We distinguished these two domino tiles by assembling an asymmetric domino tile for the three target-staples per position that lacked position 2 and some staples from this corner (Figure R12).

It was described previously that suicide oligonucleotides allowed the formation of 100% covalent complexes (González-Pérez et al., 2007). Suicide oligonucleotides contain a phosphorothiolate linkage at the *nic* site in a way that the 3´-SH end generated after cleavage reaction cannot catalyze the ligation reaction. Thus, modified staples containing a phosphothiolate linkage at the *nic* site were obtained from Yale University to improve the yield of TrwC$_{R388}$ binding to nanostructures (supplementary Table S2).

Both standard and suicide target-staples nanostructures were incubated overnight with TrwC$_{R388}$ and TrwC$_{R388}$mCherry. After analyzing more than 200 nanostructures, we observed that central positions showed higher yields (Figure R13 and supplementary figures S13 and S14). In general, binding yields varied considerably from experiment to experiment (± 10%). Overall, from 40 to 90% of the nanostructures had the central positions occupied, while the boundary positions revealed poorer yields (10 to 77%). Thus, these studies reveal a preferential binding of relaxase TrwC$_{R388}$ in the central positions of a rectangular nanostructure (Figure R14). TrwC$_{R388}$ showed a slightly better yield than TrwC$_{R388}$mCherry, when domino tiles with three target-staples were analyzed (for instance, position 3 reach 89% vs. 75% with standard targets, and 84% vs. 60 with suicide staples for TrwC$_{R388}$ and TrwC$_{R388}$mCherry, respectively). On the other hand, there is no significant improvement with suicide oligonucleotides. Indeed, the worst yield was shown when target-staples were made of suicide oligonucleotides because many nanostructures were empty in this case.



position on domino tile

Figure R13. Binding yields of relaxase fusions on domino tiles with two or three binding sites per position. The efficiency of the fluorescent relaxase TrwC$_{R388}$mCherry and the native TrwC$_{R388}$ protein was analysed. Notice that position two is missing in the design of 3bs (see text and Figure R11 for details). See also Table S9 and S10 for statistical details

Within the range of this variation, we found no clear dependence of the binding yield on the number of binding sites per binding position. However, it seems that the best overall yield for relaxase fusions was reached on domino tiles with two standard target-staples per position (Figure R14, gray bars).

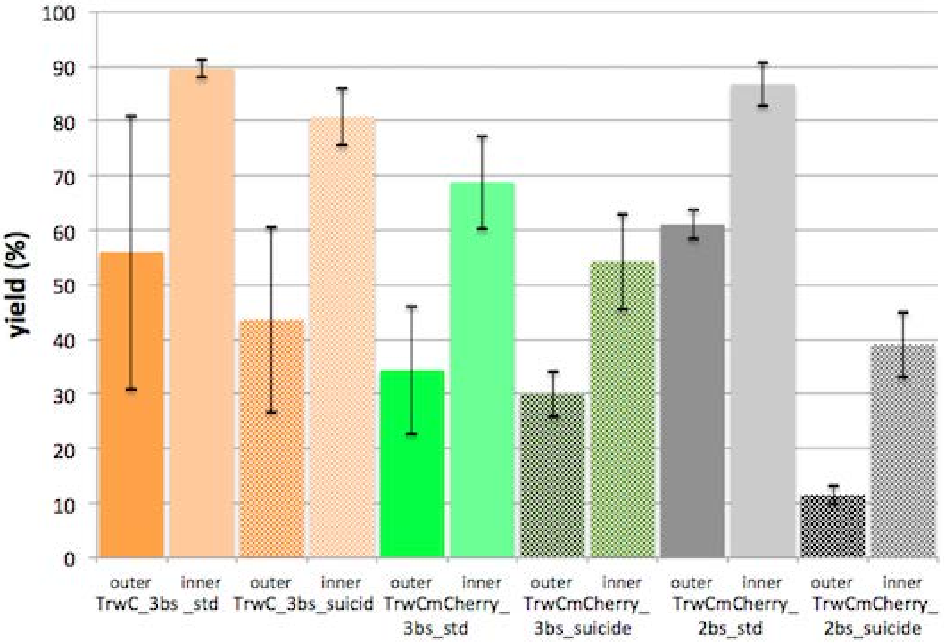


Figure R14. Position dependence of the binding sites of relaxases $TrwC_{R388}$ and $TrwC_{R388}mCherry$ within domino tiles. Outer means positions 1,2,5 and 6, whereas inner means positions 3 and 4. See Figure R12 for details. Binding yield to tiles with 2 (2bs) or 3 (3bs) binding sites per position has been calculated. Standar (std) and suicide (suicid) target-staples are represented. The error bars are calculated as the error of the data form the chart above.

5.3 Orthogonal Binding of Different Relaxases on DNA Origamis

The ability of relaxases to naturally generate a covalent linkage with DNA can be applied *in vitro* to enable nanoscale assembly of fused relaxases on DNA nanostructures. Our previous results have demonstrated that relaxase $TrwC_{R388}$ is able to bind to planar nanostructures in a sequence-specific manner. However, the functionality of the specific DNA attachment would have to be expanded with other relaxases that recognize different sequences. In chemistry, the notion of orthogonality refers to the possibility that several components react together independently and without interference from the other. Hence, our goal is to attach some relaxases orthogonally to certain positions of a nanostructure.

5.3.1 Decoration of nanotubes with several relaxases

With the goal of coupling several relaxases to DNA nanostructures, a nanotube of six-helix bundle (6HB) was used to precisely tether four relaxases along their length. The axial dimension of the nanostructure is 430 nm and it has a diameter of 5 nm. In order to increase the local concentration of binding sites, each binding position was equipped with five binding sites each for the specific relaxase. The relative positions for each target are depicted in Figure R15.

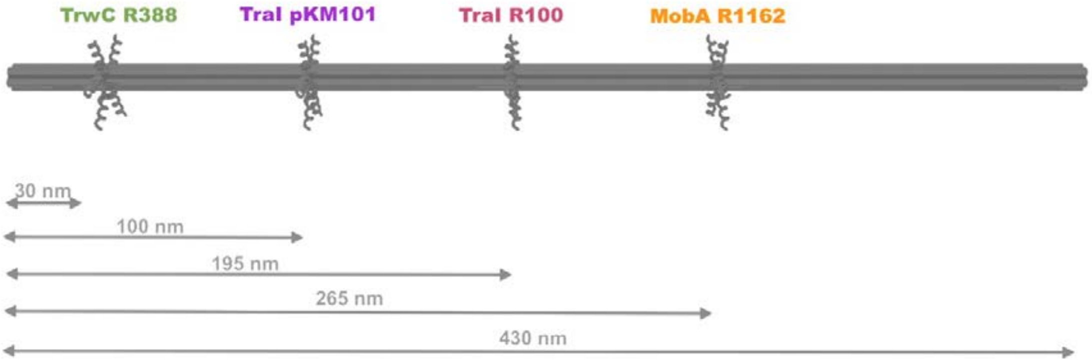


Figure R15. Target design of the nanotube. (A) The 6HB nanotube consists of six DNA double helices flanking an inner cavity whose diameter is similar to that of a double helix. Five targets per position come out with the binding site of relaxases TrwC$_{R388}$, TraI$_{pKM101}$, TraI$_{R100}$ and MobA$_{R1162}$ at different binding positions. The names of the relaxases that recognize these targets are written close to their binding position. The exact positions of each target-staple are given as the distance from one ending of the 6HB.

Each relaxase was incubated for 24 h with 6HB nanotubes containing all four binding positions. After incubation and purification, the number of occupied binding positions was determined from TEM images (Figure R16). A given protein was considered bound when it was clearly observed on the nanostructure and the measured distance corresponded with the theoretical distance. Table R1 collects the determined distances of four fusion proteins bound to 6HB relative to one endpoint.

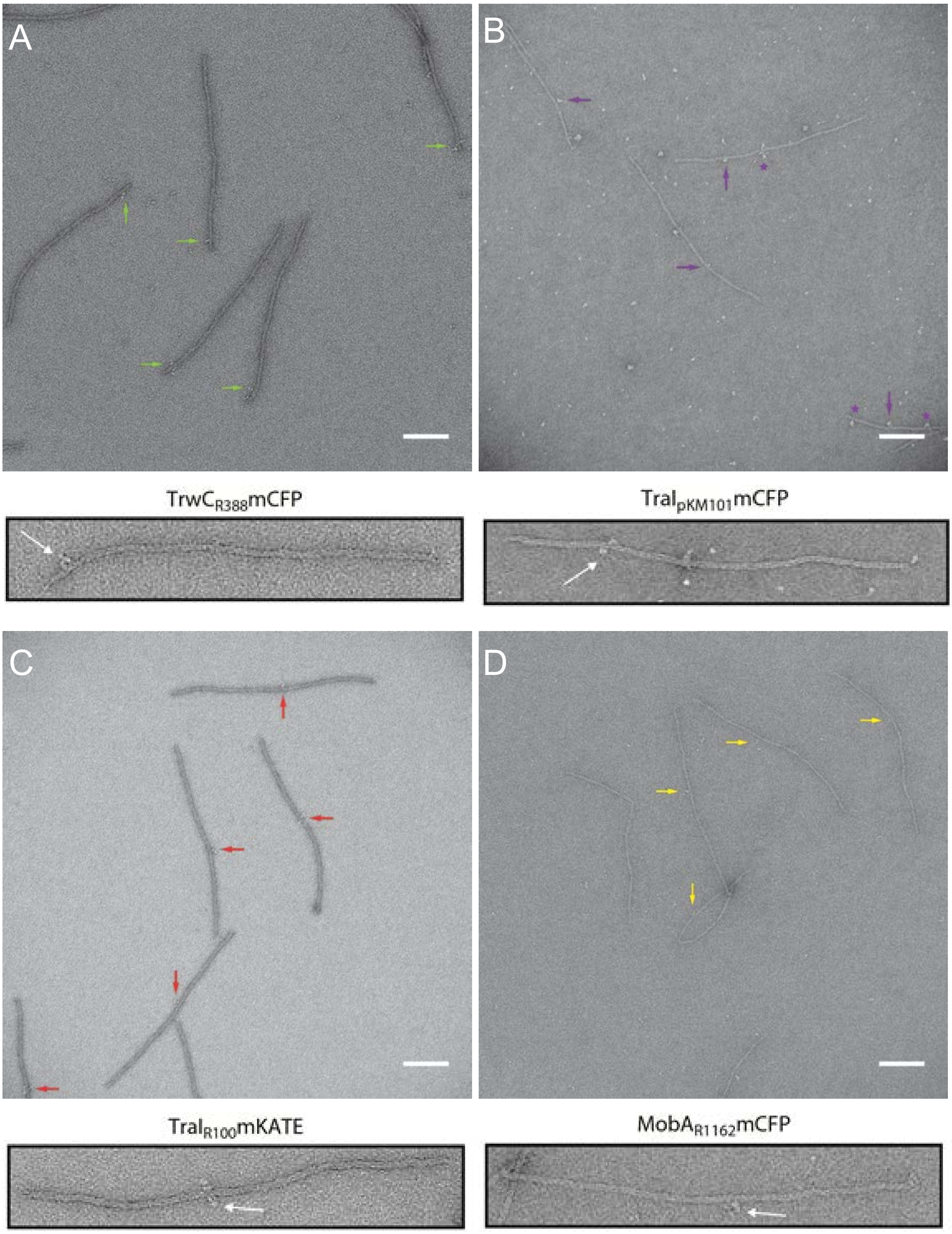


Figure R16. TEM micrographs of the relaxases TrwC$_{R388}$mCFP, TraI$_{R100}$mKATE and MobA$_{R1162}$mCFP bound to 6HB. (A) TrwC$_{R388}$CFP; (B) TraI$_{pKM101}$CFP; (C) TraI$_{R100}$KATE; (D) MobA$_{R1162}$CFP. The colored arrows show the protein bound to the correct position. The stars show the relaxase bound non-specifically to other position. Scale bars of 100 nm. Insets show 6HB with the relaxase bound.

Table R1. Binding yields and determined distances of four fusion proteins bound to 6HB. The determined distances are average values relative to one endpoint. Distances with * corresponds to the distances from the protein to the other end-point. The given uncertainty is the standard deviation. Corresponding TEM images are shown from Figure R16.

	Number of nanoestruc with protein bound	Number of empty nanostructures	YIELD	6 HB Lenght	Lenght to edge	Theoretical length
TrwC R388	102	38	72,9%	**425,2** ± 6,1	**29,0** ± 3,3	30 nm
TraI pKM101	26	5	83,9%	**430,0** ± 3,6	**106,5** ± 8,3	100 nm
TraI R100	150	11	93,2%	**430,5** ± 7,1	**199,4** ± 9,1 ***241,7** ± 6,1	195 nm *235 nm
MobA R1162	44	86	33,9%	**426,1** ± 6,9	**255,2** ± 20,4 ***164,0** ± 10,1	265 nm *165 nm

The relaxase TraI$_{R100}$KATE showed the best yield, 93,2%. TraI$_{pKM101}$CFP and TrwC$_{R388}$CFP reached a percentage of 83,9 and 72,9 respectively. MobA$_{R1162}$CFP showed the worst coupling efficiency, with only 33,9% of proteins bound to its binding site (Table R1).

Since the 6HB contained all four binding positions, quantitative values for the specificity of each relaxase to its target could be measured. TraI$_{R100}$mKATE also showed highly selective binding abilities, with only 1,2% non-specific binding to R1162. TrwC$_{R388}$mCFP and MobA$_{R1162}$CFP revealed sequence-specific binding, with non-specific binding yields ranging from 0,8% to 7.1%. Despite its high yield for its specific target, TraI$_{pKM101}$CFP shows significant non-specific binding to R100 (77,4%), R1162 (12,9%) and R388 (29%) targets (Figure R17). pKM101 and R100 targets show 55,6% sequence identity. On the other hand, TraI$_{pKM101}$ shares 99,8% identity with TraI$_{pCU1}$, which has been biochemically and structurally characterized (Nash et al., 2010). TraI$_{pCU1}$ binds DNA in a weak sequence-independent manner. Its reported K_D for oligonucleotides containing the *nic*-region range from 840 nM to 960 nM (Nash et al., 2010). It has been suggested that TraI$_{pCU1}$ requires the accessory protein TraK$_{pKM101}$ to initiate sequence-specific DNA binding to *oriT*. Therefore, TraI$_{pKM101}$ does not show orthogonality as the other proteins do.

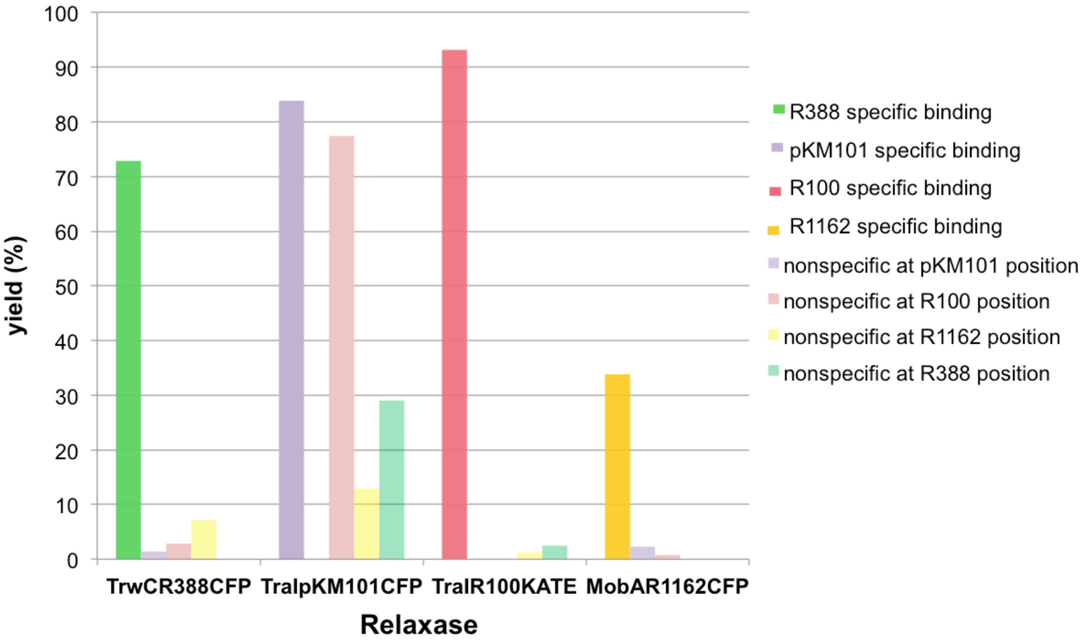


Figure R17. Total binding yields of relaxases to 6HB and their orthogonality. The binding yield of the relaxase to its specific target is shown in dark colours. The percentage of unspecific binding to other targets is shown with bars in lighter colours. See also Table S12 for statistical details.

Our efforts focused on achieving the orthogonal binding of several relaxases to DNA nanostructures, to allow the programmable arrangement of several fused proteins on DNA nanostructures. We studied their arrangement on 6HB using a super high-resolution fluorescence microscopy technique named DNA-PAINT (points accumulation for imaging in nanoscale topography) (Lin et al., 2012; Steinhauer et al., 2009). This technique is used in the field of DNA nanotechnology to image fluorescent dyes with <300 nm resolution. TrwC$_{R388}$mCherry and MobA$_{R1162}$CFP were used because, when assembled, the distance between them is 235 nm (see Figure R15). The nanotube had also some biotinylated staples that allowed its immobilization on streptavidin-coated coverslips before imaging. Fluorescence emission was detected using total internal reflection (TIR) microscopy. Fluorescent proteins exhibit very poor photostability and they were photobleached during imaging. Unfortunately, we could not reconstruct the high-resolution images. Consequently, 6HB was incubated with all the relaxases for 24h, purified and imaged under TEM. Even though we could not discriminate if the binding was or not orthogonal, we showed nanotubes decorated with the desired combination of relaxases (Figure R18A and B).

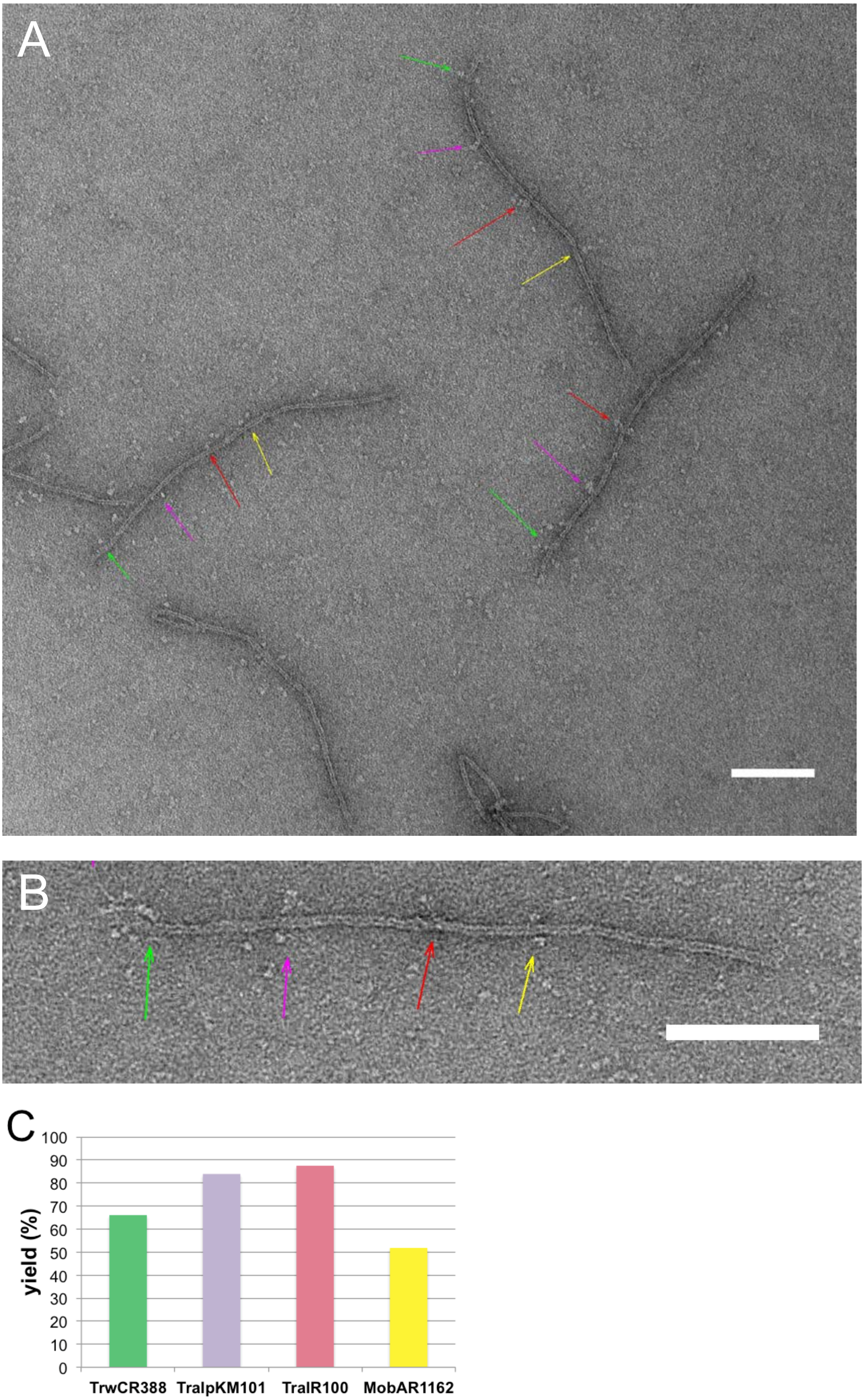

A
B
C
yield (%)
100
90
80
70
60
50
40
30
20
10
0
TrwCR388
TraIpKM101
TraIR100
MobAR1162

Figure R18. Nanotubes incubated with four relaxases. (A) TEM micrograph of 6HB incubated with $TraI_{R100}$mKATE, $TraI_{pKM101}$mCFP, $TrwC_{R388}$mCFP and $MobA_{R1162}$mCFP. Proteins are indicated with arrows: $TrwC_{R388}$ (green), $TraI_{pKM101}$ (purple), $TraI_{R100}$ (red) and $MobA_{R1162}$ (yellow). For a description of the 6HB design see Figure R15. (B) Expanded view of a TEM micrograph showing the nanotube decorated with four relaxases. Same colour code. Scale bars of 100 nm. (C) Binding yield of the 6HB incubated with four relaxases.

The results of the analysis of more than 50 nanostructures are shown in Figure R18C. The resulting binding yields per position were 88% for $TraI_{R100}$mKATE, 84% for $TraI_{pKM101}$mCFP, 66% for $TrwC_{R388}$mCFP and 52% for $MobA_{R1162}$mCFP. Overall, binding yields were high and were in accordance with the results obtained when the nanotube was incubated with each protein (see Table R1).

5.3.2 Decoration of rectangular sheets with several relaxases

In order to analyze the binding ability of the aforementioned relaxases to decorate 2D nanostructures, we designed a tcRO with binding sites for them. Two relaxase targets per position with linkers of 5 bp protrude from the base of the origami (Figure R19). On the upper left corner were R100 target staples (position 1). The upper right corner contained pKM101 target staples (position 2). The lower left corner had R388 target staples (position 5) and the middle position was occupied with R1162 target staples (position 4).

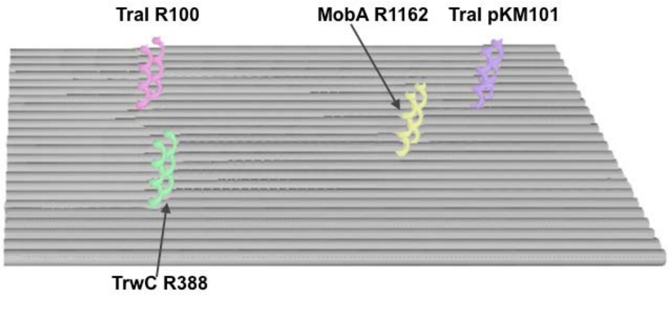


Figure R19. illustration of the tcRO with four binding positions for relaxases $TraI_{R100}$, $TraI_{pKM101}$, $MobA_{R1162}$ and $TrwC_{R388}$. The rectangle was equipped with two target-staples for each relaxase with a linker of 5 bp.

We then compared the binding of the orthogonal relaxases $TraI_{R100}$mKATE and $TrwC_{R388}$mCFP onto the origami sheets with positions 1 and 5. Both positions are equivalent; they are at the rim of the tcRO, next to the corners (Figure R19). Each protein was incubated for 24h with the tcRO, purified and imaged under the AFM (Figure R20). The binding yield of Tral$_{R100}$mKATE was found to be 70%, and 48% for TrwC$_{R388}$mCFP. Non-specific binding to the unwanted position was 1% for Tral$_{R100}$mKATE or 3% for TrwC$_{R388}$mCFP. TrwC$_{R388}$ binding yield is in accordance with our previous results (see Figure R10 and R12).

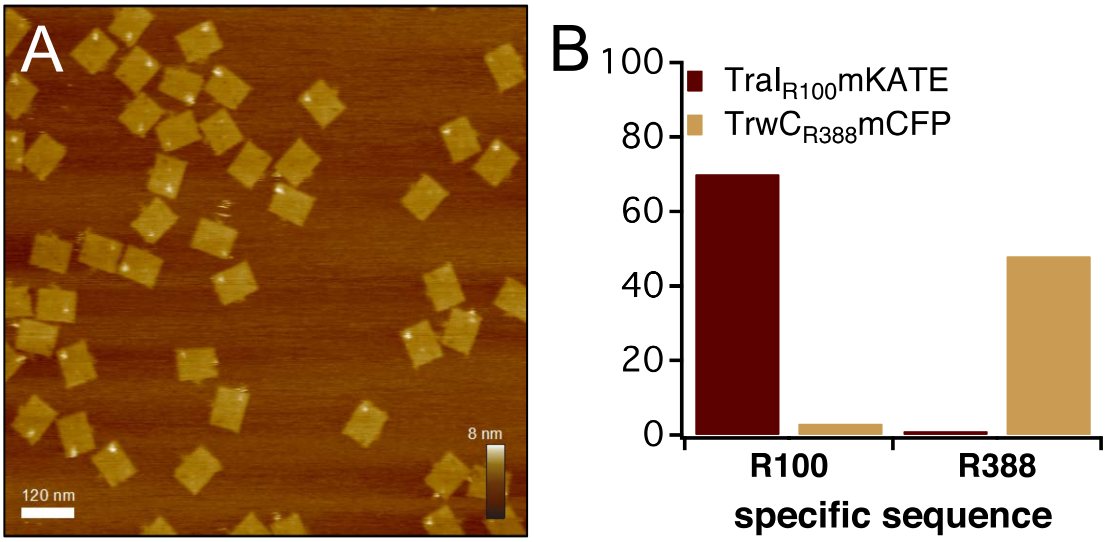


Figure R20. Specific binding of Tral$_{R100}$KATE and TrwC$_{R388}$CFP to a planar nanostructure. (A) AFM image of the nanostructures decorated with Tral$_{R100}$KATE at position 1. The brighter dots are the proteins that are localized preferably in the upper left corner. Notice that some nanostructures are upside-down. (B) Binding yields of TrwC$_{R388}$mCFP and Tral$_{R100}$mKATE to available sites on the tcRO. The relaxases bind preferably to their specifc target. See also supplementary Figure S15 and Table S11 for statistical details.

We were wondering if rectangular nanostructures could be decorated with four relaxases, as we previously did with nanotubes. Planar nanostructures equipped with binding sites for each relaxase were incubated with relaxases for 24h (Figure R19). The number of occupied binding positions was determined from AFM images (Figure R21A and B). Most of the nanostructures held either 1 or 2 proteins (Figure R21C). The binding yield of Tral$_{R100}$ and TrwC$_{R388}$ was 74% and 37,5%, respectively. These results are in accordance with our previous yields of these proteins on tcROs (see Figure R20). Surprisingly, the yield of Tral$_{pKM101}$ and MobA$_{R1162}$ was less than 20%. Compared to 6HB, the binding yields on tcRO decreased significantly.

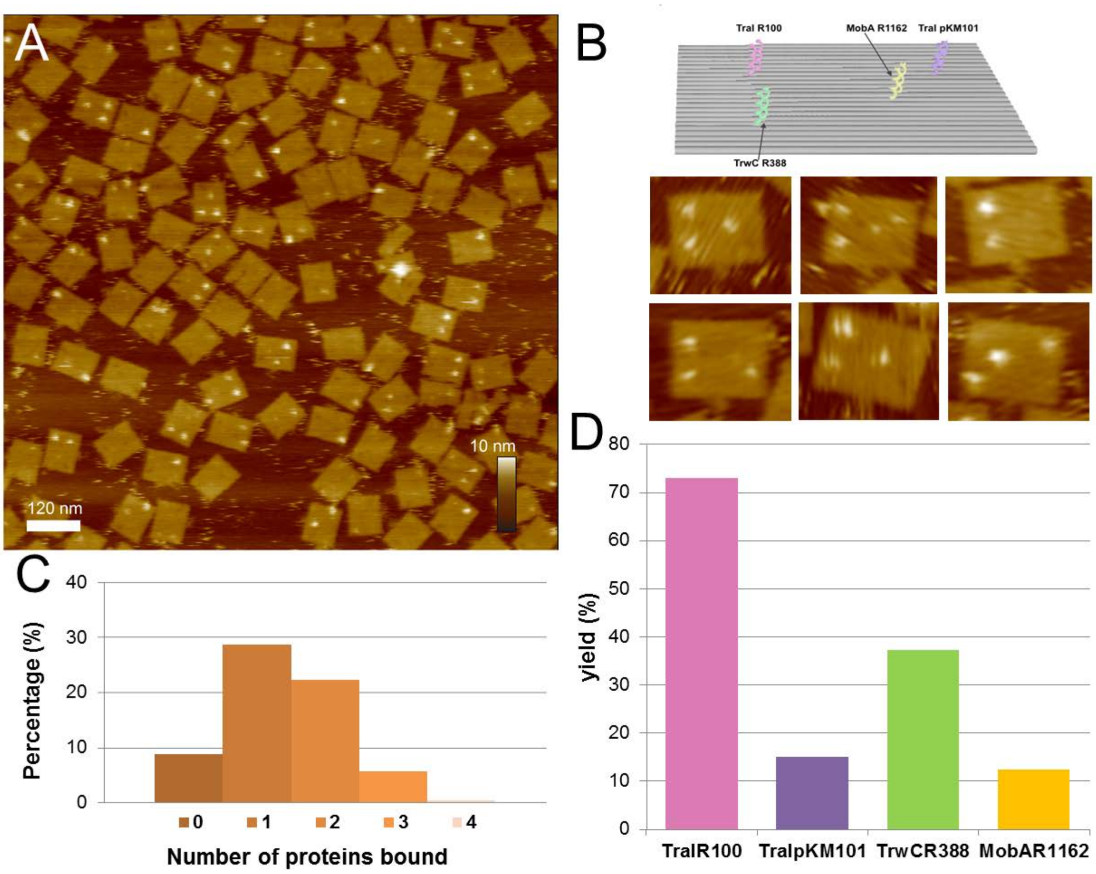


Figure R21. Binding yields of Tral$_{R100}$, Tralp$_{KM101}$, TrwC$_{R388}$ and MobA$_{R1162}$ on tcRO. (A) AFM image of the nanostructures decorated with all the relaxases. (B) Scheme of the relaxases binding positions carrying two standard binding sites at each position on the tcRO. Insets, AFM images of tcRO with several relaxases arranged. (C) Percentage of sites occupied in the nanostructure. (D) Binding yield of all the relaxases to available sites on the tcRO.

5.3.3 Effect of the Incubation Time

We examined the ability of relaxases to bind 6HB for short incubation times by using a nanotube with R388 and R1162 binding sites (Figure R22A). After incubation and purification, the number of occupied binding positions was determined from TEM images (Figure R22B.) When the 6HB was incubated with only one relaxase for 1h, the binding yield of TrwC$_{R388}$mCFP and MobA$_{R1162}$mCFP was similar (53 and 49 % respectively). But the 24h-incubation time experiment resulted in 68% TrwC$_{R388}$CFP binding yield. MobA$_{R1162}$CFP binding yield decreased to 26%. However, when both proteins were incubated together, the yield was constant over time (Figure R22C). Since these proteins show high specificity, we could attribute these differences to small changes in the sample preparation.

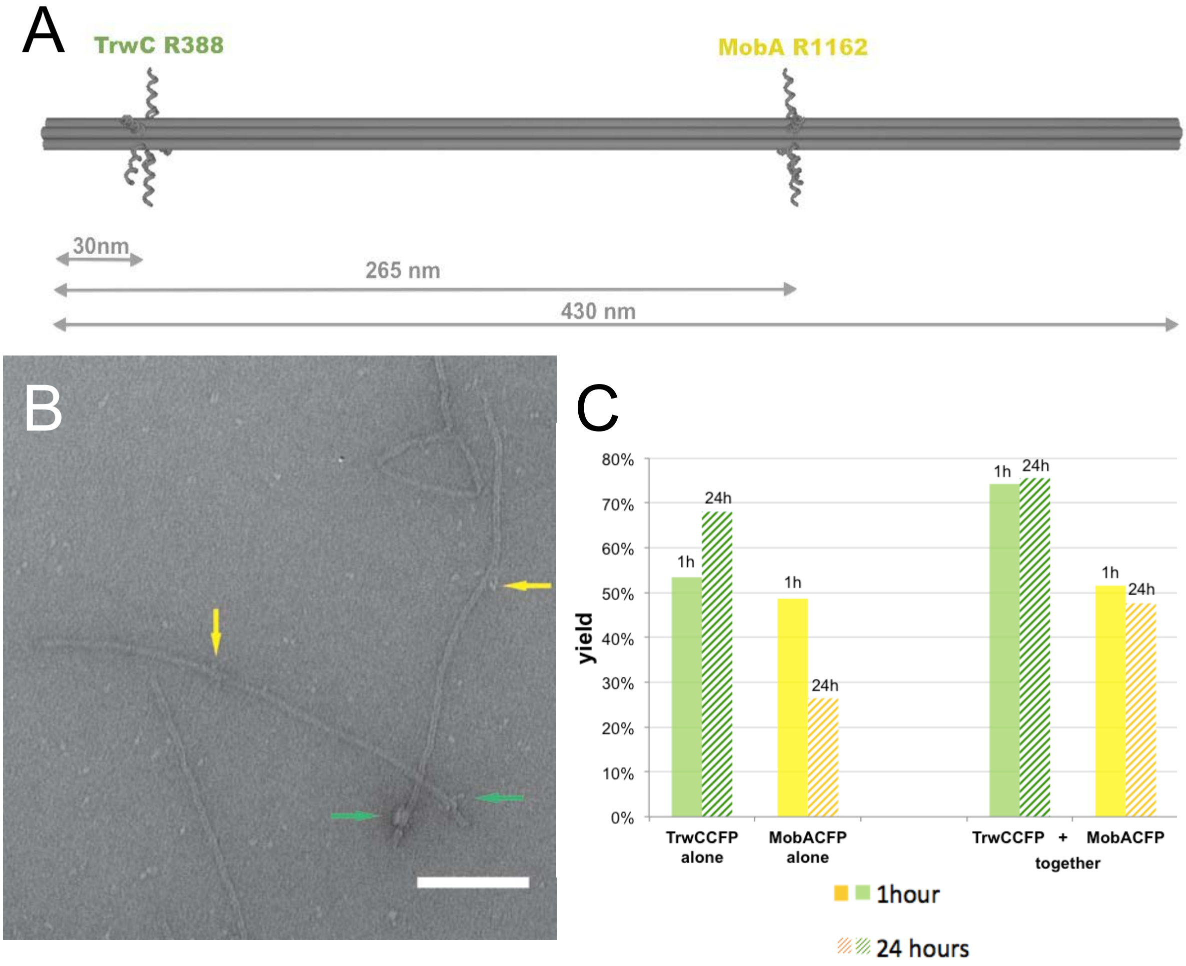


Figure R22. Effect of incubation time on the binding of TrwC$_{R388}$mCFP and MobA$_{R1162}$mCFP to 6HB. (A) Six-helix bundle design with binding positions for TrwC$_{R388}$ and MobA$_{R1162}$. Each position contains five extended staples with the respective target. (B) TEM micrograph with both relaxases bound to 6HB after 24h incubation. Green arrows denote TrwC$_{R388}$ and yellow arrows MobA$_{R11162}$. Scale bar 100 nm. (C) Binding yield of TrwC$_{R388}$mCFP and MobA$_{R1162}$mCFP to 6HB for incubation times of 1h and 24h. On the left side of the chart, 6HB were incubated with only one relaxase. On the right side of the chart both relaxases were incubated in one pot. The yield was determined from TEM images. See also Table S13 for statistical details.

5.3.4 Effect of the Suicide Staples on the Nanotube

It was described previously that suicide target-staples were obtained to improve TrwC$_{R388}$ binding yield to tcRO because TrwC$_{R388}$-DNA covalent complexes accumulated when suicide oligonucleotides were used (González-Pérez et al., 2007). Previous studies with the four relaxases showed that only TrwC$_{R388}$ and TraI$_{pKM101}$ cleaved suicide oligonucleotides (supplementary Figure S18). Thus, 6HB with three suicide target-staples for TrwC$_{R388}$ and TraI$_{pKM101}$ binding sites were assembled (Figure R23A).

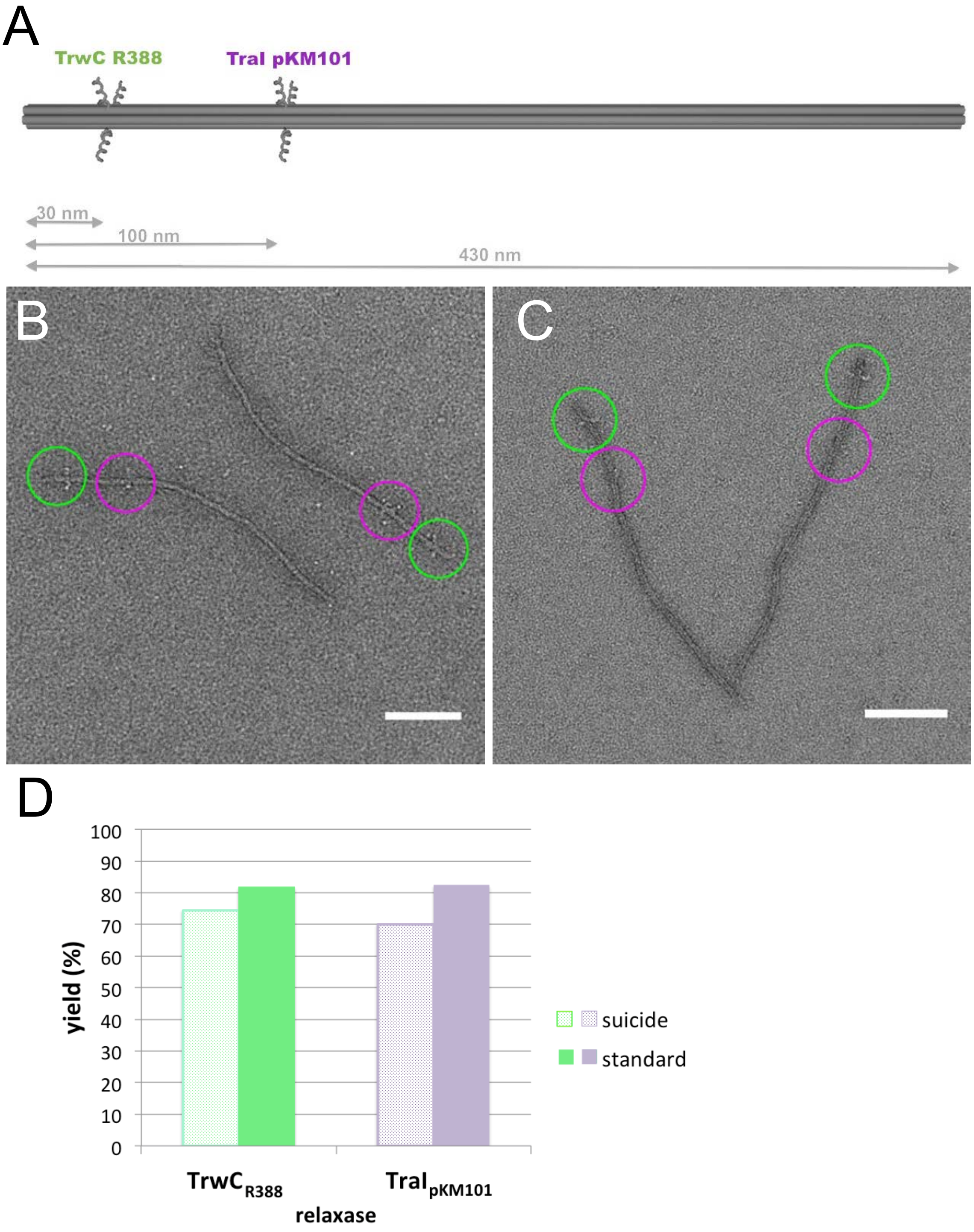


Figure R23. Arrangement of TrwC$_{R388}$ and TraI$_{pKM101}$ on DNA nanotubes holding suicide target staples. (A) Scheme of the 6HB assembled with suicide target-staples. Every binding site possesses three targets projecting. (B) TEM micrograph showing the 6HB with standard and (C) suicide targets bound to relaxases. Green circles indicate TrwC$_{R388}$CFP molecules and purple circles show TraI$_{pKM101}$CFP molecules. (D) Graph representing the binding yield of TrwC$_{R388}$CFP and TraI$_{pKM101}$CFP to 6HB with three binding sites per position. See also Table S14 for statistical details.

After incubation for 24h and purification, the number of occupied binding positions was determined from TEM images (Figure R23B and C). The binding yield for TrwC$_{R388}$mCFP and TraI$_{pKM101}$mCFP to 6HB with suicide target-staples per position was compared to the same nanostructures assembled with standard target-staples (Figure R23D). The resulting binding yields to 6HB with suicide and standard targets were 70% vs. 89% for TraI$_{pKM101}$mCFP and 74% vs. 88% for TrwC$_{R388}$mCFP, respectively.

Suicide staples were also evaluated on rectangular nanostructures. tcRO containing two targets per position were incubated with TrwC$_{R388}$CFP and TraI$_{pKM101}$CFP (Figure R24A). After 24h, samples were purified and analyzed under the AFM microscope. Surprisingly, most of the nanostructures were empty in the case of suicide oligonucleotides. Even though 5-fold excess target-staples were used in each case, the phosphorothiolate oligonucleotides were not efficiently assembled on their binding positions. Even thought the absolute binding yields were different, the analysis revealed similar relative yield of both proteins to their target (Figure R24B).

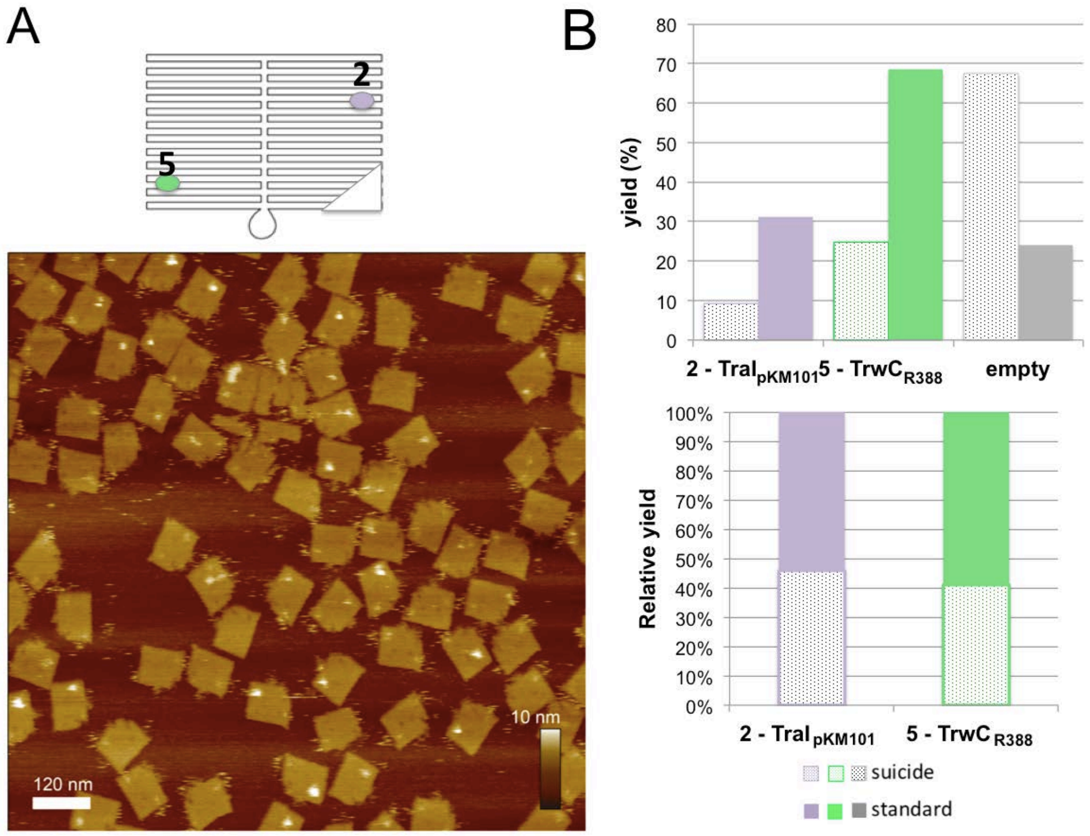


Figure R24. Arrangement of TrwC$_{R388}$ and TraI$_{pKM101}$ on rectangular nanostructures holding suicide target staples. (A) Scheme of the tcRO assembled with suicide oligonucleotides, and AFM image of the tcRO incubated with TrwC$_{R388}$mCFP and TraI$_{pKM101}$mCFP. The yield was compared to the same nanostructures assembled with standard oligonucleotides. (B) Graph representing the absolute and relative binding yield of TrwC$_{R388}$CFP and TraI$_{pKM101}$CFP to tcRO eqqiped with suicide and standard target-staples.

Unexpectedly, the binding yield did not improve with suicide oligonucleotides. The strong coupling theoretically expected for a covalent bond is thus not reflected in the actual binding yields.

In summary, the approaches used to study the orthogonal binding of proteins to nanostructures have revealed that two relaxases, TraI$_{R100}$ and TrwC$_{R388}$, seem to bind more efficiently to its binding sites on nanotubes and tcROs than the relaxases MobA$_{R1162}$ and TraI$_{pKM101}$. Depending on the origami geometry, the specific relaxase, and also the utilization of redundant binding sites in close proximity, in some cases, binding yields as high as 90% could be achieved.

5.4 Polymerization of DNA Nanostructures by Relaxases

We were also interested in developing an approach in which relaxases assist on the assembly of DNA nanoribbons. Since relaxases recognize a specific sequence of DNA and are able to ligate DNA that contains its target, we hypothesized that relaxases could be used as a novel method for the polymerization of DNA nanoribbons. So far, the methods explored to direct the assembly of DNA nanoribbons implied rolling circle amplification (RCA) and base stacking (Beyer et al., 2005; Jungmann et al., 2011) (see Figure 1D). Thus, our next goal was the fabrication of extended nanoribbon structures based on DNA for being polymerized by relaxases. Our approach has the advantage that the assembly of the nanoribbon is coupled with the periodical arrangement of proteins to it.

5.4.1 dsDNA Nanodevices Polymerization

A linear nucleic acid nanodevice was designed to be polymerized by the binding and catalytic reaction of $\mathrm{TrwC}_{\mathrm{R388}}$. Four single stranded oligonucleotides were the component molecules of the nanodevice. After assembly, the device possesses a central stable double-stranded region (75 bp) and a single stranded segment in both edges holding $\mathrm{TrwC}_{\mathrm{R388}}$ targets 24+8 and 24+0 (Figure R25A). The length of the central region of each subunit is 25 nm (75bp*0,34nm/bp) (Figure R25B). As we have seen previously, both targets 24+8 and 24+0 possess the IR_2 inverted repeat, that folds as a 6 base pair duplex, and a single stranded region surrounding the *nic* site. There is no complementarity between both edges, so base stacking does not allow the growth of the nanoribbon. Therefore, polymerization can be addressed only via the strand transfer reaction catalyzed by relaxase $\mathrm{TrwC}_{\mathrm{R388}}$. Moreover, this thesis proves that the reverse target R(8+24) serves as an improved substrate for $\mathrm{TrwC}_{\mathrm{R388}}$ cleavage and strand transfer reaction (see section 6.3 for further details of this substrate). Therefore, a device with the 8+24 at the + end was also tested.

In order to induce flexibility in the device the bacterial integration host factor (IHF) target has been incorporated in the central region. This bacterial histone-like protein can induce an extreme bending of DNA. Our aim was the generation of modifiable nanodevices that allowed the incorporation of other proteins contributing to the conformational modification of the duplex structure (Shen et al., 2004). Although, we did not add IHF proteins so far, they could be added to these nanoribbons in a future.

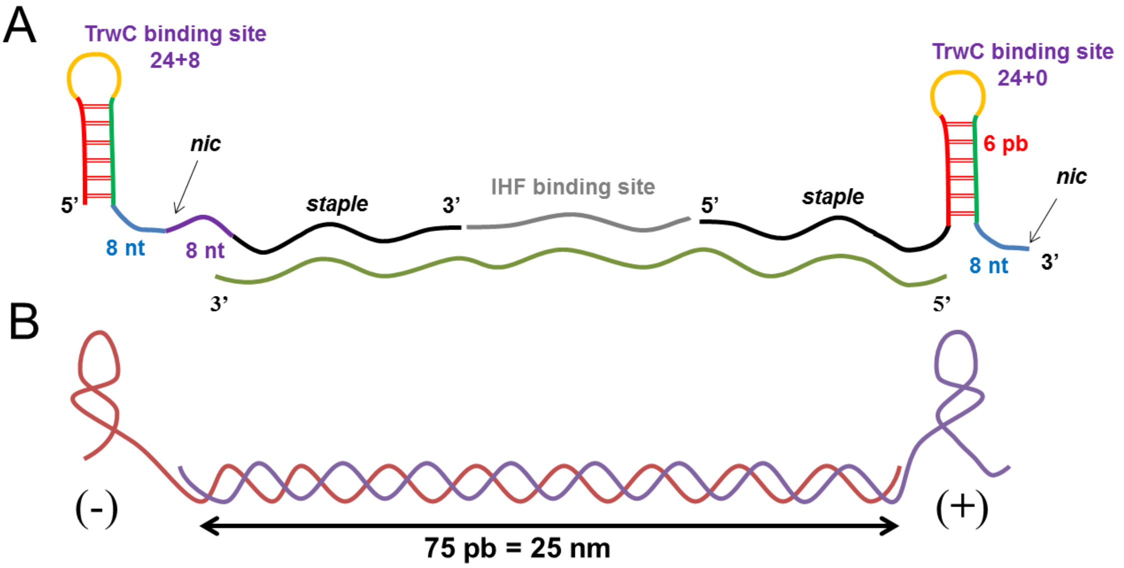


Figure R25. Schematic representation and design of the DNA nanodevice used for TrwC-directed nanoassembly. (A) Four oligonucleotides are the components of the designed nanodevice. TrwC target is located in both edges. The 5' end contains the wild type TrwC$_{R388}$ binding site 24+8, and the 3' end the non-cleavable 24+0 target. The distal arm of the inverted repeat is shown in red, the proximal arm in orange, and the U turn is in blue. Upon cleavage, TrwC$_{R388}$ is covalently bound to the ssDNA downstream from the *nic* site (purple region). (B) Single and double stranded representation of the nanodevice. The distance between the targets is shown. Considering that the distance between 10 bp is 3,4 nm, the length is 75x3,4/10= 25.5 nm. The symbols (-) and (+) represent the minus end and plus end, respectively.

TrwC-directed assembly of nanoribbons was performed at two concentrations (1 µM and 10 µM) at room temperature (22ºC) in 1xTAE pH 7.6 with 12.5 mM magnesium. Diluted samples were loaded onto a 15% non-denaturing acrylamide gel as described in materials and methods. SYBR Gold staining allowed the visualization of DNA afterwards (Figure R26A, upper gel). Native gels showed a fluorescent signal according to the presence of TrwC$_{R388}$mCherry bound to the DNA (middle gel in Figure R26A). The polymerization of the 8+24-devices increased after 2 weeks of polymerization (compare lane 2 with lane 3), whereas 24+8-devices seem to form polymers more efficiently after 1 week (lane 5). These polymers are stable after 2 weeks (lane 6). Actually, few dimers were observed with shorter incubation times (data not shown). After protein digestion (lanes 8 to 14), nanoribbons migrated depending on their molecular weight. A ladder appeared in both samples, 24+8- and 8+24-devices (lanes 8-9 and 11-12) because the nanodevices formed a dimer, a trimer, a tetramer, etc. TrwC$_{R388}$mCherry promotes the formation of longer polymers with the target 8+24, though. The analysis of the shifted bands revealed that the longest polymer migrates as a 375 bp duplex, which means that five monomers of 8+24-device have been combined (Figure R26B). In the case of the 24+8-device the longest polymer was a tetramer. It is noticeable that the band of the monomeric nanoribbons disappeared when larger oligomeric nanoribbons were forming.

Regarding the fluorescent signal, two bands of fluorescent protein were shown after one week of incubation (lane 2 and 5), which means that two species are mainly present in the sample. After two weeks of incubation at room temperature, the band was lost. Probably, because most of the proteins were degraded. A intense band appears on lane 8 and 11 seems after proteinase K treatment. Similar bands are shown in lanes 9 and 12 with lower intensity. The presence of these bands can be explained as digested fragments of the fluorescent moiety that are in the sample. The band with high mobility that appears in lanes 1,4,7,10 and 15 is related to the blue loading buffer used in this samples for visualization the migration on the gel.

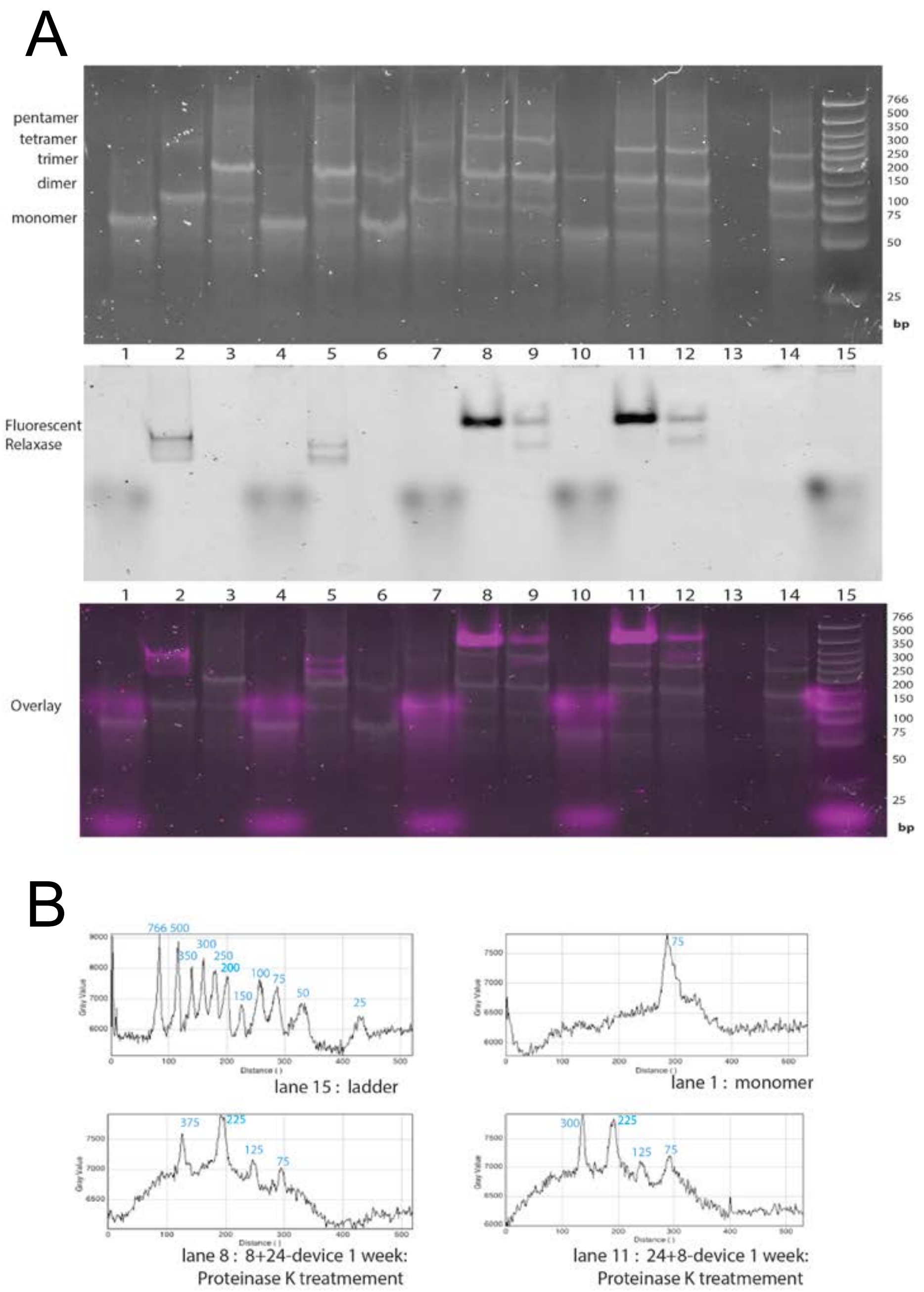


Figure R26. Polymerization of wt and reverse nanodevices. (A) Non-denaturing acrylamide gel that shows the process of polymerization of the nanodevices after one or two weeks of incubation. 1µM of the nanodevice with the reverse R(8+24) (lanes 1 and 7) or the wt W(24+8) target sequences (lanes 4 and 10) was incubated for either 1 or 2 weeks with 3µM TrwC$_{R388m}$Cherry (lanes 2-3, and 5-6, respectivelly). 1µM and 10µM reactions incubated during one week were treated with proteinase K, 8+24 (lanes 8, 9) and 24+8 (lanes 11, 12), respectivelly. Lane 13: empty; Lane 14: 1µM W(24+8) after 2 weeks with TrwC$_{R388m}$Cherry and treated with Proteinase K; Lane 15: Low molecular weight DNA ladder NEB. Upper gel is after SyBrGold stain while middle gel shows the fluorescence of TrwC$_{R388}$mCherry. Bottom gel is an overlay of both gels. (B) Profile of the intensities of sections of lane 1, 8, 11 and 15 obtained with ImageJ. The label indicates the length in base pairs of the bands detected.

Samples were diluted to 2 nanomolar and analyzed in air with AFM. The samples contained magnesium that allows the DNA to interact with the mica surface by non-specific ionic charge attraction. Linear superstructures deposited on the mica are the programmed nanoassemblies of DNA and proteins shaping a nanoribbon (Figure R27A). Interestingly, the negative controls of DNA monomers alone and free proteins, showed a random distribution of them on the mica (Figure R27B).

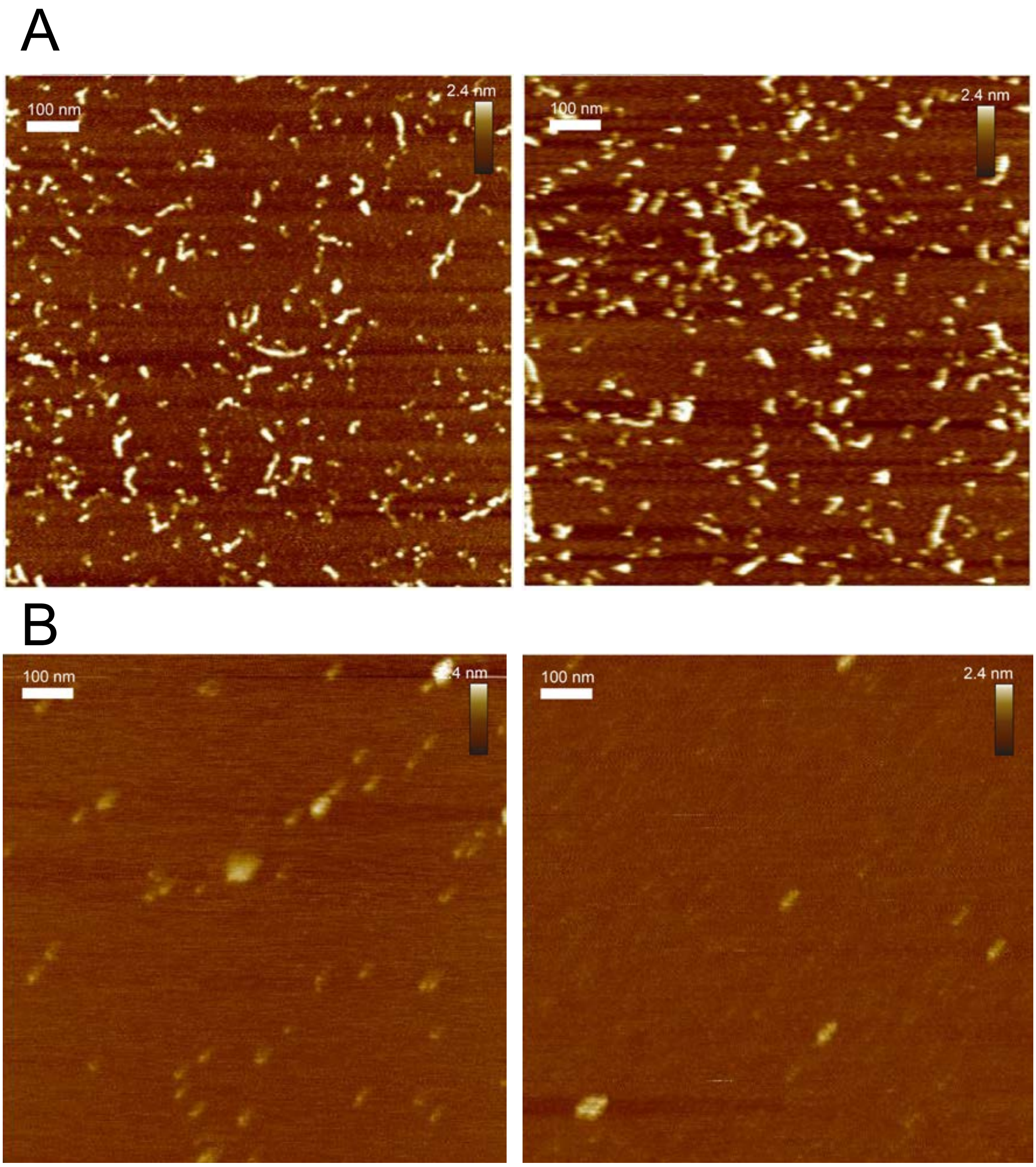


Figure R27. AFM images of ds-nanodevices. (A) 24+8 (left) and 8+24 (right) - nanodevices incubated for one week at room temperature. A necklace pattern was obtained in the samples when they were incubated with TrwC$_{R388}$mCherry. The necklace pattern corresponds to the relaxase bound the nanodevice shaping a nanoribbon. (B) Negative controls. Relaxases (left) and monomers of DNA (right) alone. The proteins are observed as white balls in the AFM images and the dsDNA as narrow lines. Random dsDNA monomers appear when bare DNA is deposited on mica. However, when they are incubated with relaxase, the protein molecules appeared periodically on the mica surface.

The height, diameter and longitude were measured from multiple cross-sections of the AFM images (Figure R28). We determined the length of the polymers by measuring the centre-to-centre distance between proteins of the nanoribbon in AFM images. Images correspond to the fused protein TrwC$_{R388}$mCherry, because it resulted in increased visibility in AFM images. However, the cantilever tip shrinks the biological sample when is tapping the sample, and the measured height is less than the theoretical height. As it is shown in Figure R29, the height of the circular dots was 1- 1.5 nm and 5 nm diameter, which corresponds to the protein, whereas the protein is approximately a sphere of 4 nm. The dsDNA was 0.5 nm high but the real diameter of dsDNA is 2nm. The distance between the central points of each protein was approximately 25 nm, which corresponds with the predicted length of the nanodevice (see Figure R28).

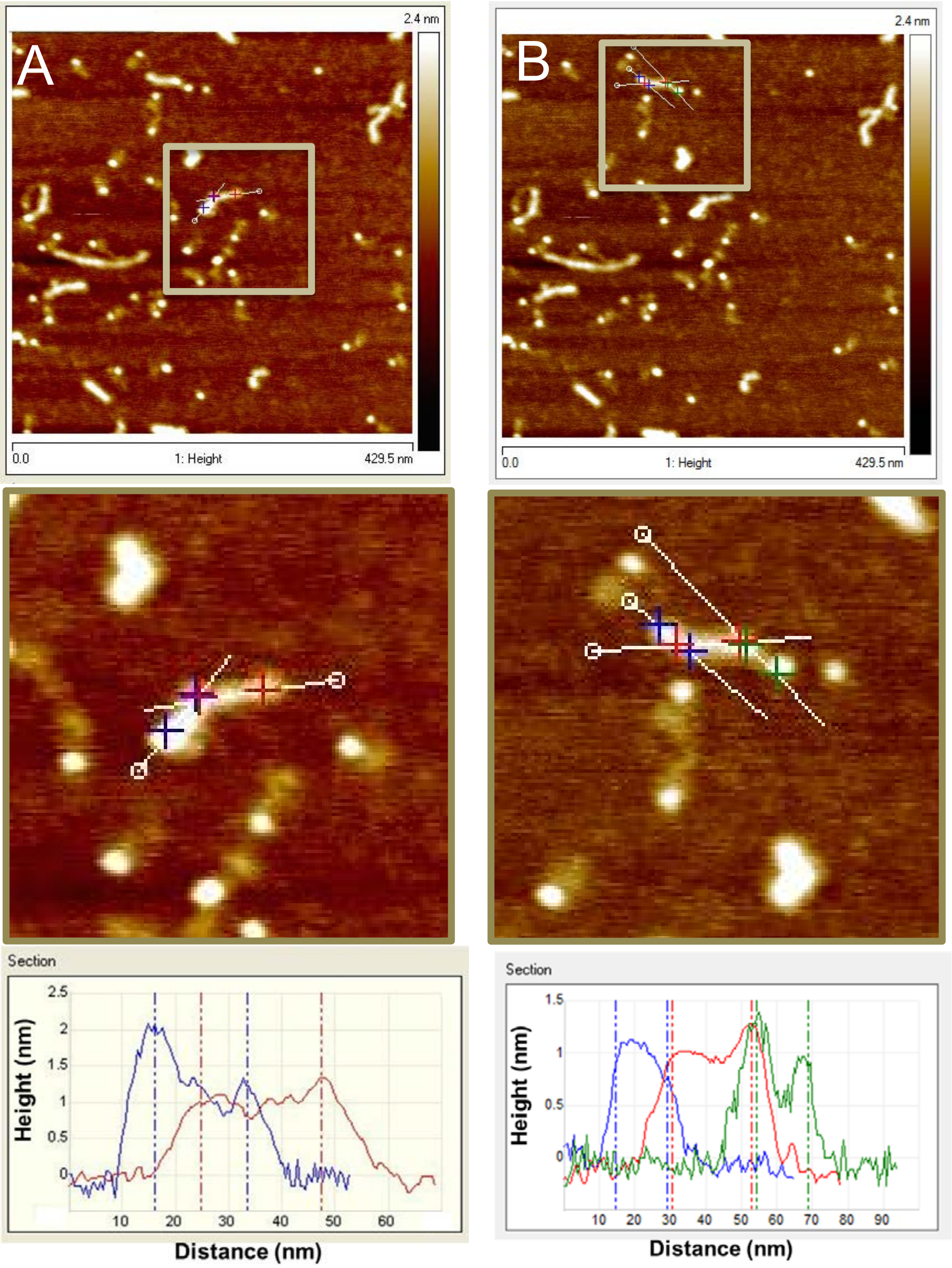


Figure R28. AFM cross-section and height profile of linear relaxase-dsDNA nanoribbons. Inset, zoom of the nanodevice. Height profiles along the lines indicated bellow each image. (A) The connection is an L-shape, which corresponds to three relaxases bound to two monomers. The distance between two connected proteins is 25 nm approximately (blue and red crosses), which is in accordance with the theoretical lenght. (B) Inset, zoom of the nanodevice. In the section diagram the three lines corresponds with the monomers connected by four relaxases. The total lenght of the device is 70 nm. The distance between them is 25 nm approximately. TrwC$_{R388}$mCherry has a diameter of 5 nm (white spots) and an average height of 1.25 nm. dsDNA is 0.5-0.75 nm high.

The distance between peaks along each single monomer of approximately a total of one hundred nanoribbons was analyzed (Table R2). The length distribution of the nanoribbons is also shown. The distance between adjacent proteins in the nanoribbon was 20±5 nm, which matches with the predicted distance. The yield decreased exponentially as a function of the ribbon length. Overall, 50% of the nanodevices were dimers, but only 7% of them formed pentamers.

Table R2. Quantification of relaxase-dsDNA nanoribbons. The distances between peaks and total length of the nanoribbons were measured from height images and corresponding cross-sectional profiles of AFM images. The average values were obtained from 96 8+24-polymers and 111 24+8-polymers. See supplementary figures S19 and S20.

	Polymers	Number of dots	Events	Yield	Average length
8+24-device	dimer	3	52	54,2%	43
	trimer	4	26	27,1%	62
	tetramer	5	10	10,4%	97
	pentamer	6	7	7,3%	110
	hexamer	7	1	1,0%	116
	Total		96		
24+8-device	dimer	3	58	52,3%	42
	trimer	4	35	31,5%	60
	tetramer	5	11	9,9%	86
	pentamer	6	7	6,3%	104
	hexamer	7	0	0,0%	-
	total		111		

We expected circularization of the nanodevice in the case of the attachment between +end and – end of the same molecule. However, none of the imaged analyzed revealed the presence of nanocircles.

5.4.2 Modeling Polymerization of DNA Devices

Asymmetric edges were used in the nanodevice design to direct the assembly of the monomers by the relaxase. We can consider a plus (+) end and a minus (–) end by similarity to the polarity of filaments during polymerization. The wt target 24+8 is positioned at the + end, necessary to carry out the ligation to the - end, where 24+0 is positioned. TrwC$_{R388}$ would connect the DNA nanoribbons by a three-step reaction (Figure R29). The first step involves the binding of TrwC$_{R388}$ to the + end. The second step corresponds with the activation, when TrwC$_{R388}$ generates a covalent complex with the 24+8 oligomer, schematized as TrwC-8-monomer (activated monomer). Subsequently, polymerization can be performed when the – end is recognized by the activated TrwC$_{R388}$, the 24-mer of the 24+8 +end is interchanged with the 24+0 - end and a new substrate 24+8 is generated (this interchange is known as the strand transfer reaction). TrwC$_{R388}$ remains bound to the target connecting two monomers and a resultant one-dimensional nanoribbon is assembled. The scheme bellow shows each step of the polymerization (Figure R29A) and the expected polymer after incubation (Figure R29B).

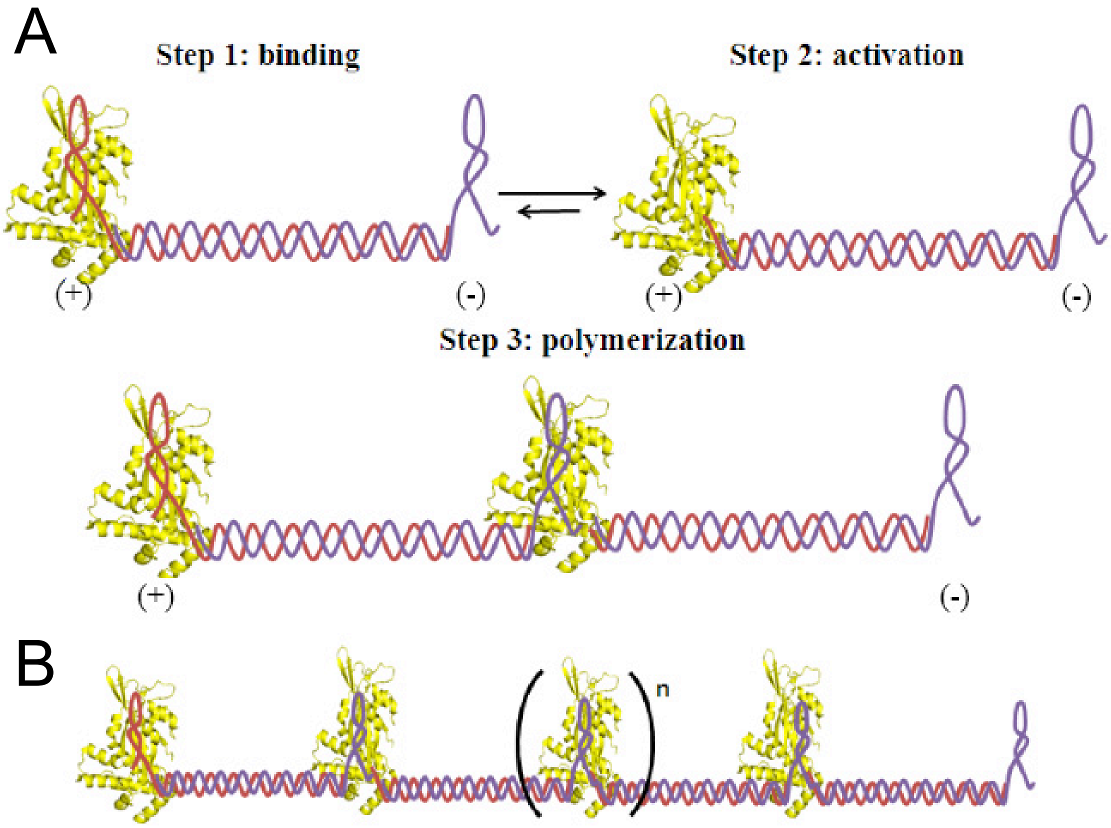


Figure R29. Scheme representing each step during the nanoassembly of dsDNA directed by TrwC$_{R388}$. (A) The relaxase TrwC$_{R388}$ carries out the polymerization in three steps: binding, activation and polymerization (for details see the text). During activation the relaxase is covalently bound to the DNA. (B) The expected nanoribbon generated after the nanoassembly of DNA by TrwC$_{R388}$. The activated -end is connecting new nanostructures, where n represents the number of units in the central region.

5.4.3 ssDNA Nanodevices Polymerization

In a further approach, we also investigated the polymerization of ssDNA oligonucleotides via TrwC$_{R388}$. In this experiment, one single stranded DNA oligonucleotide contained TrwC$_{R388}$ target at both ends in a similar way to the aforementioned device (Figure R30A). Now the 5'-end contained the target 12+6, whereas the target at the 3' end was 24+0. TrwC$_{R388}$ can cleave the target at the 5' end and remain covalently bound to the monomer, until it catalyzes the strand transfer reaction to either other monomer or to the fluorescent target F24+19. The polymerization reaction was evaluated after incubation for 1h at 37ºC. After proteinase K and SDS treatment, monomers and dimers appeared in a denaturing gel (Figure R30B). However, we were not able to discriminate longer polymers, because the intensity of the bands with lower mobility was exponentially decreasing. We neither investigated the polymerization in longer periods nor imaged it under the microscope. Altogether, our results demonstrate that TrwC$_{R388}$ was able to polymerize ssDNA as well as dsDNA devices.

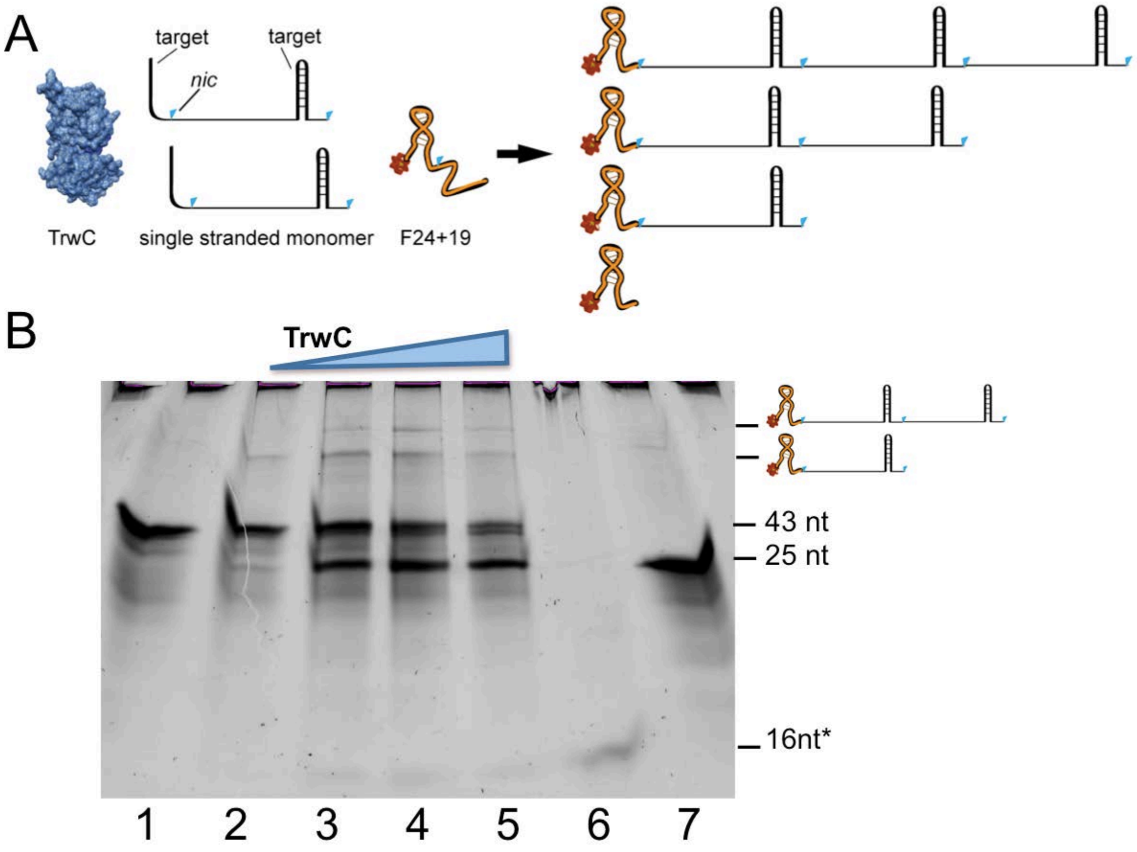


Figure R30. ssDNA polymerization. (A) Scheme of the ssDNA monomer and the polymerization reaction that TrwC$_{R388}$ carries out. The monomer's length is 58nt. It posses R388 targets at both ends: 12+6 positioned at the - end and 24+0 at the +end. F24+19 is a fluorescent target that enables the detection of the polymers in a denaturing gel. (B) Denaturing gel showing the polymerization reaction. Lane 1, F24+19; Lanes 2-5, increasing concentrations of TrwC$_{R388}$ incubated with the ssDNA monomers and F24+19; Lanes 6 and 7, fluorescent oligonucleotides of 16 nt and 25 nt, respectively. *the oligonucleotide of 16nt seems to be degraded degraded. The fluorescent target F24+19 is cleaved by TrwC$_{R388}$ and a 24-mer fluorescent produc appears in the samples, as well as fluorescent polymers generated via strand-transfer reaction of the ssDNA monomers and the fluorescent probe.

5.2.4 Polymerization of Origami Tiles by TrwC$_{R388}$

Several approaches for DNA origami polymerization have been investigated so far. Simmel et al developed three methods for sequence specific polymerization of tcRO (Jungmann et al., 2011). Either one (first approach) or three (second approach) single-stranded extension strands were attached to the edges of tcRO monomers (Figure R31). Only short origami oligomers were formed after incubation for 14 days because polymerization is sensitive to the stoichiometry between linker strands and origami structures. The most successful approach (third one) connects two origami with 24 "bridging strands". Several micrometer long origami nanoribbons were obtained with this method. In contrast with the other approaches, polymerization strands can be added in excess, because there is no a kinetic trap involved.

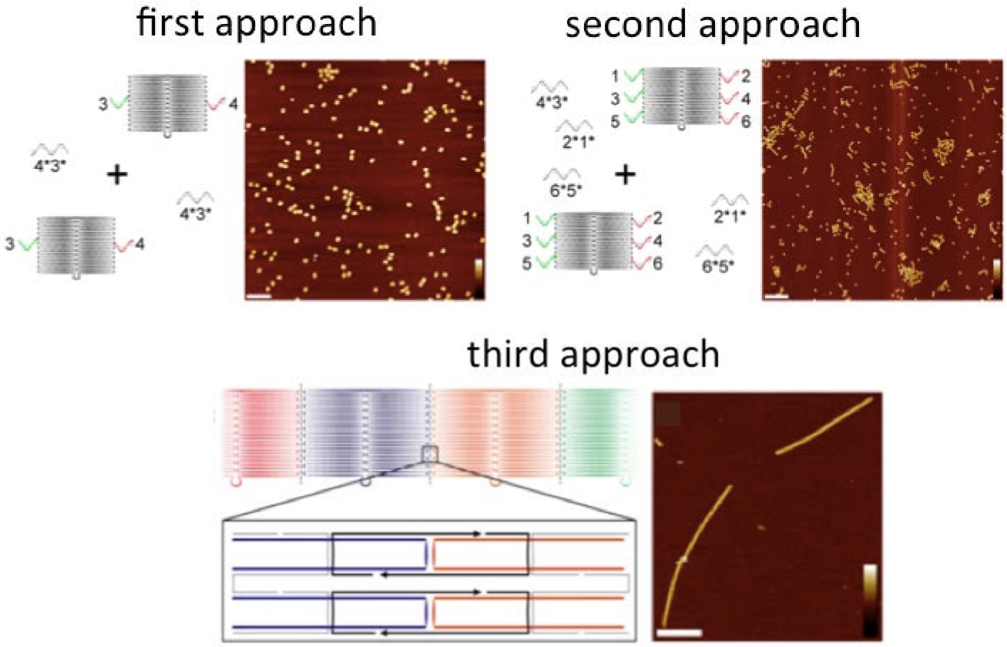


Figure R31. Polymerization methods for the assembly of nanoribbons from rectangular DNA origami structures developed by (Jungmann et al., 2011). First approach: DNA origami monomers containing non-complementary single-stranded extensions (green (3) and red (4)) are joined together using one 20 bases long linker strand (gray (3∗ 4∗)). The AFM image shows the resulting structures with only a small fraction of linked monomers. Second approach: Same principle as before, but origami monomers are linked at three positions using three different extension/linker sequences. The AFM image reveals longer ribbons compared to the first approach. Thrid approach: DNA origami monomers are joined by 24 linkers or "bridging strands". Each bridging strand (black arrow) contains 24 nt complementary to the scaffold sequence on one edge of the origami rectangle and 8 nt complementary to the scaffold on the opposing edge. This method produces specifically linked origami polymers. Length scale: 800 nm, height scale: 8 nm, polymerization time: 14 days.

Here, we investigated a relaxase-directed polymerization method of tcRO, as we have previously shown with dsDNA and ssDNA devices. We used four single-stranded extension strands with the relaxase target in two positions of the edges of an RRO monomer (Figure R32). The cleavable sequence, 24+8, was aligned with the 24+0 ligation-target sequence (Figure R33).

Figure R32. Illustration of the DNA origami monomers containing four *nic* sequence-containing staples. Each monomer is eqquiped with four *nic* sequence-containing staples located in opposite edges of the tcRO. The cleavable strand (24+8, in orange) from each edge is joined to the target 24+0 (in green) from the opposite edge of another monomer via the strand transfer reaction catalyzed by $\text{TrwC}_{\text{R388}}$.

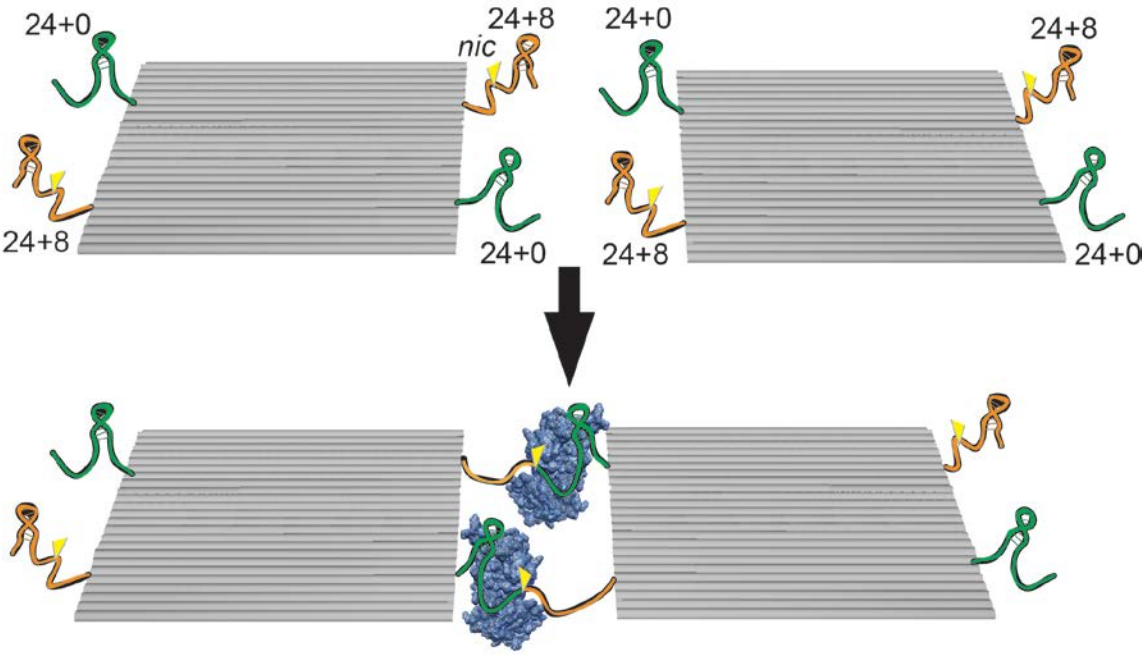


Figure R33. Schematic representation of the TrwC-directed polymerization of DNA origami tiles. Each DNA origami (gray sheets) is connected by the strand transfer reaction with the next tile via $\text{TrwC}_{\text{R388}}$. Afterwards, the relaxase (in blue) is bound to the linked extended staples in between tiles because the target 24+8 is regenerated.

We analyzed if this new protein-directed polymerization generates long nanoribbons. To this end, origami monomers and $\text{TrwC}_{\text{R388}}$ were incubated at room temperature for 1h, 2h, overnight and overweekend. After purification, samples were analyzed by AFM. Experiments were carried in 1xTAE pH 7.6 with 12.5 mM magnesium. Sodium chloride was not added, because some clustering appeared in samples with 200 mM NaCl in the buffer (data not shown). It seems that Na^+ generates agglomeration of DNA origamis, because monovalent positive ions affect the interaction of the nanostructures with the mica surface (Aghebat Rafat et al., 2014). Successful polymerization was achieved when samples were incubated at 37ªC overnight (Figure R34A-C). Interestingly, trimers appeared when samples were incubated for just 2 h (Figure R34D). Most of the oligomers were dimers, trimers and tetramers. Longer polymers were not found. Polymerization was directed by $\text{TrwC}_{\text{R388}}$ because when the same samples were incubated at 4ºC and when samples lack $\text{TrwC}_{\text{R388}}$ just monomers appeared (Figure R34E-F). We decreased the temperature to 4ºC because it has been reported that relaxases are sensitive to temperature (Stern et al., 2004 and oral communication from GM).

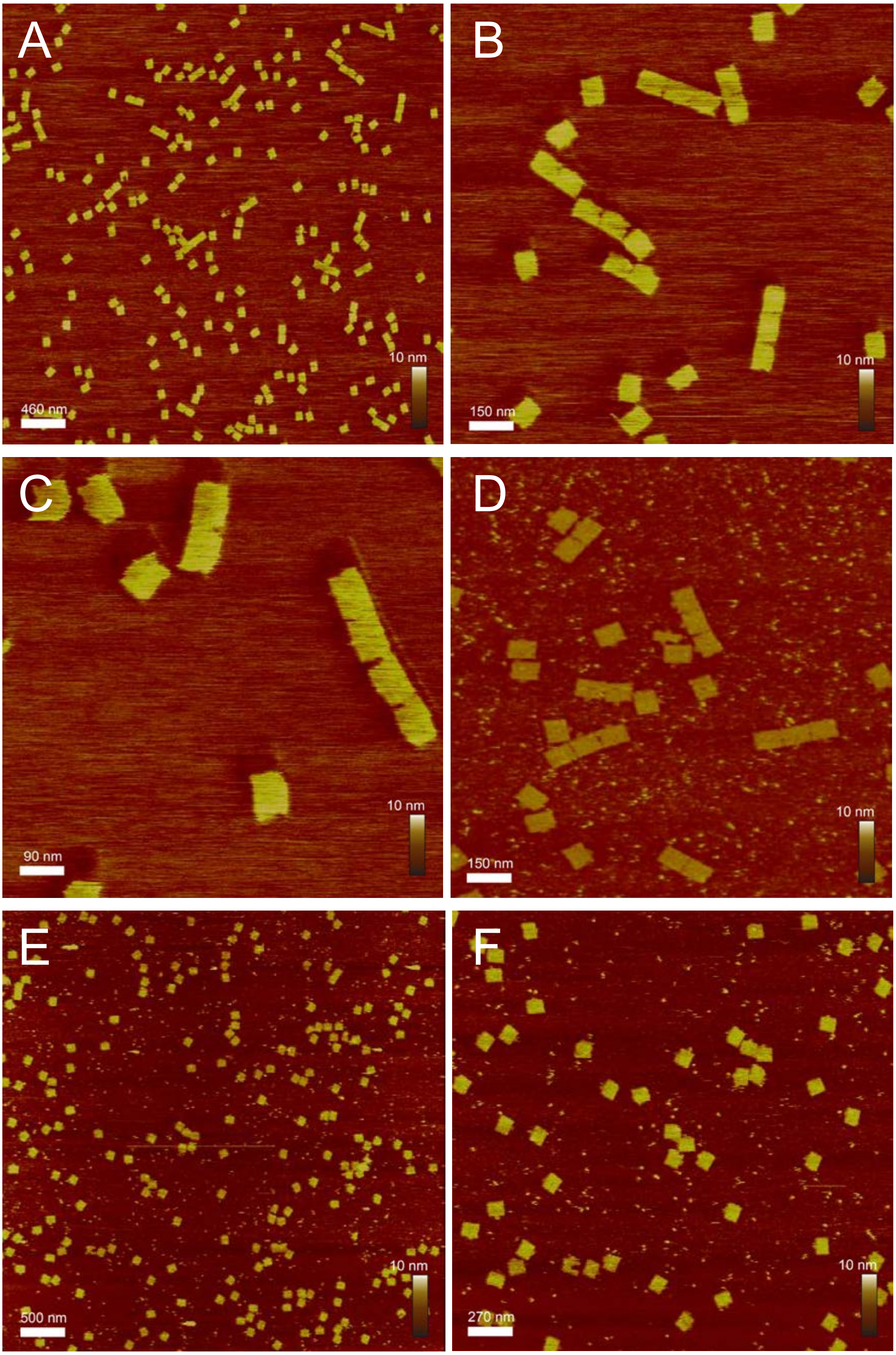


Figure R34. AFM images of origami nanoribbons assembled by the relaxase TrwC$_{R388}$mCherry. (A) TrwC-origami incubation overnight demonstrated several oligomers, most of them dimers, trimers. (B) Zoom image of (A). (C) Tetramer and dimer. (D) Polymers obtained after incubation of tcROs with TrwC_{R388}mCherry for 2 h. (E) AFM image of TrwC-origami incubation at 4ºC. (F) Zoom image of (E) showing monomers.

To summarize, TrwC_{R388} could be used as a useful tool for DNA origami polymerization. In an ideal case, these origami monomers could be assembled in larger two or three-dimensional molecular nanostructures by using several relaxases. On the other hand, this method cannot compete with the methods developed so far. For instance, with the third approach developed in the group of Friedrich Simmel (see Figure R31). But to our knowledge, this is the first approach that describes polymerization of DNA nanodevices carried out by enzymes. Our results shed light on the applications that relaxases offer in the field of DNA nanotechnology.

DISCUSSION

In this thesis, we have demonstrated the use of relaxases for site-specific covalent conjugation of proteins to single stranded DNA extensions on DNA origamis. This is potentially an interesting alternative to the current methods for immobilization of DNA labeled proteins, because it does not require the initial step of conjugation of DNA to the protein by chemical methods; attachment of nanostructures with relaxases is performed in a one-step reaction. Other advantage is that stoichiometry is controlled while using a chemical approach uncontrolled multiple DNA-labels are attached to a single protein. This variability problem also affects the use of streptavidin because the streptavidin tetramer can attach one, two, three and up to four biotinylated DNA strands. Moreover, the key advantage of the relaxase method is that it can be used for orthogonal labeling of different sites at the origami, since relaxases from different bacteria and with unique recognition sequences are available.

We note that there are still unanswered questions about the binding process of relaxase proteins to DNA nanostructures, but these problems are in common to DNA-binding proteins in general. We demonstrated that relaxases are competitive with other binding strategies and result in effective binding yields in the range between 70% and 90%. The K_D in the nanomolar range of $TrwC_{R388}$ to its recognition sequence should translate into an equilibrium binding yield of 86% for a single binding site and even 99.7% for three binding sites under our incubation conditions, where $TrwC_{R388}$ and tcRO concentrations were ≈ 2 µM and 100 nM, respectively (see section 5.1 for details). Overall, the binding yields obtained by AFM were significantly lower than expected from the law of mass action. As AFM imaging is performed with purified samples under buffer (i.e., without relaxases), however, re-equilibration of the sample will result in unbinding of the proteins from the target and thus a reduced apparent binding yield. Our findings are in line with results for other protein-modification approaches for DNA origami. For example streptavidin, which binds to biotin with a dissociation constant of K_D=40 fM (Green, 1990), experimentally showed a binding yield to nanostructures of only about 85% (Voigt et al., 2010) although the theoretical value was close to 100%. The other covalent coupling methods presented in the introduction also yielded around 84-90% experimental binding (Voigt et al., 2010, Sørensen et al., 2013).

Our efforts to covalently bind relaxases to DNA origamis were successful with at least three (TraI$_{R100}$, TrwC$_{R388}$ and MobA$_{R1162}$) of the four investigated relaxases showed good orthogonal binding performance. Although the relaxase TrwC$_{R388}$ showed a preference to bind to central positions on rectangular origamis.

One feature that can affect protein binding is the electric charge of the nanostructure. Since DNA is electronegative, planar nanostructures would have a higher relative negative charge in the central positions than in the boundary (Figure D1). Similarly, the group of Hao Yan found a position dependency of the hybridization of mRNA in solution to DNA probes on rectangular planar nanostructures (Ke et al., 2008). The optimal position of the probes for detecting mRNA was at the edge, whereas the probes at the middle of the edge give the worst hybridization efficiency. They suggested that electric-charge repulsions between the target and the underlying tile, as well as repulsions between probes, were stronger in the middle of the tile than at the edge of the tile. In our opinion, these factors influenced also our results in a different way. We guess that the preference in binding to the central region of the tcRO is favored by some of the positive residues that are in the surface of the relaxases (Figure D1). The electrostatic potential of the relaxases appears to have a positive potential throughout its binding cleft, the region that interacts with their target DNA, and an overall neutral net charge. On the other hand, negative charge of the nanotubular structure would be more homogeneous, and position does not play an important role in binding efficiency.

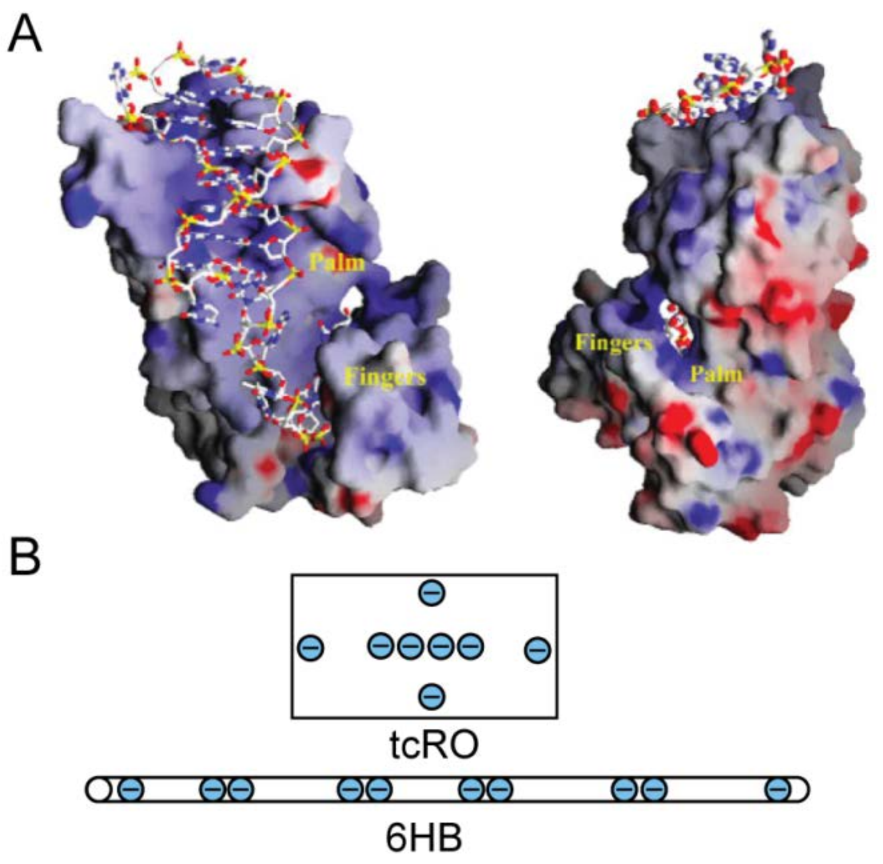


Figure D1. Electrostatic potential of the surface of the relaxase TrwC$_{R388}$ and the two nanostructures. (A) Two views, related by 90º rotation of the TrwC$_{R388}$ electrostatic potential surface, contoured from -20 (red) to 20 (blue) kTe-1. The DNA is shown as sticks. Taken from (Guasch et al., 2003). (B) Schematic representation of the negative potential landscape of the tcRO and the 6HB.

In an attempt to rationalize our observations, we used simple theoretical considerations and made a model of the dynamics of binding. Attachment of relaxases to DNA nanostructures can be modeled in the context of diffusion of proteins in a randomly distributed medium of either rectangular or rod-like nanostructures (Figure D2). Diffusion may be affected by the geometry of the binding site or the overall shape and flexibility of the underlying origami structure. Since the dissociation constant of relaxases to their target DNA is in the nanomolar range, the concentration of proteins and nanostructures may affect the efficiency: the number of collisions with other molecules determines the binding yield (see section 5.1).

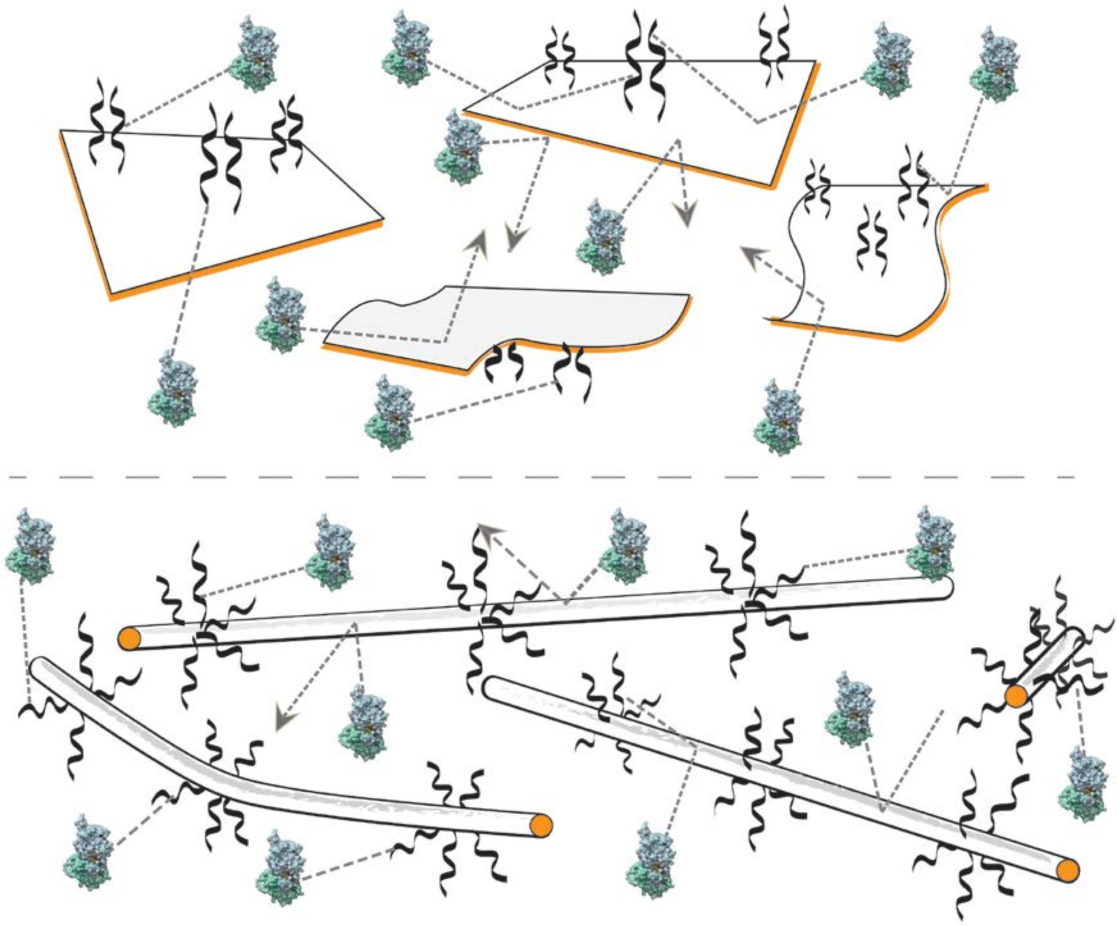


Figure D2. Factors that influence binding efficiency of relaxases to DNA nanostructures. Interactions between tcROs (upper image) and 6HB (lower image) with relaxases (depicted in blue) can result in modified binding preferences. The shape of the nanostructure can affect the binding of relaxases, possibly by affecting the target accessibility. One face of the tcRO hold the relaxase targets, but they are accessible from all sides on the 6HB. The number of binding sites can also affect its binding probability. The concentration of proteins and the restricted diffusion are other parameters that affect binding. Black lines represent the relaxase binding sites. Dash lines represents the collisions that occur during diffusion.

In general, our experimental results with several relaxases showed lower binding yield on the tcRO than on the 6HB. In our design, planar nanostructures only have binding sites on one face (Figure D2). Thus, 50% of the collisions would be unsuccessful in attaching a protein to the tcRO. On the other hand, tubular nanostructures with several binding sites along its diameter have more probability that any of these collisions results in binding because binding sites were accessible from all sides. The other issue is the number of binding sites and the proximity among them. We have shown that three targets protruding in a region of 5 nm in a tcRO have lower yield than two binding sites when fused proteins are used. Steric hindrance can explain these observations.

In this thesis, we have demonstrated that the relaxases are promising reagents for the sequence-addressable modification of DNA nanostructures with proteins. Relaxases $\mathrm{TraI}_{\mathrm{R100}}$, $\mathrm{TrwC}_{\mathrm{R388}}$ and $\mathrm{MobA}_{\mathrm{R1162}}$ showed highly selective binding abilities on DNA origami structures. Binding yields ranged between 40%-50% for a single binding site, but reached up to 90% when several binding sites were on the proximity. Due to their specificity for a single-stranded DNA target, their orthogonality and binding properties, relaxases are a uniquely useful addition to the toolbox available for protein modification of DNA nanostructures.

CHAPTER 2

Improvement of relaxase substrates

RESULTS

Chapter 2: Improvement of relaxases substrates

6. Improvement of the cleavage reaction of relaxases

For further applications in the field of bionanotechnology, approaches to displace the balance toward the formation of a covalent complex relaxase-DNA are required. The goal in this chapter is the use of computational and structural tools to rationally design novel substrates for the improvement of the cleavage reaction of several relaxases. We boosted the relaxase-DNA complexes by changing the layout of the DNA substrate. Moreover, we have also performed crystallographic studies that revealed novel features of a possible dimer interface and the structure of a covalent complex with DNA.

6.1 TrwC$_{R388}$ recognizes and cleaves substrates designed based on rolling circle replicases

Location of the *nic* site in their cognate substrate is structurally different in replicases, relaxases and other HUH endonucleases (Chandler et al., 2013). Despite the fact that HUH relaxases and replicases both recognize a hairpin-like structure, conjugative *nic* sites are located 3´ to the hairpin (Figure R35A) while replicase origins are located in the loop within the hairpin region (Figure R35 B). The wt conformation recognized by relaxases possesses an inverted repeat (IR) that can be divided into a <u>d</u>istal arm (D) and a <u>p</u>roximal arm (P) with respect to the <u>*n*</u>*ic* site (N). Moreover, there are eight nucleotides between P and N that are bent with a U-shape (U). The <u>s</u>equence (S) downstream N is tethered to the relaxase after cleavage (Figure R35C). In order to improve the relaxase *nic*-cleavage reaction, redesign of hairpin substrates was carried out by reconverting relaxase *nic* substrates into replicase like (Rep-like) substrates (Figure R35B). To create the Rep-like substrate, the 5´end of the distal sequence (D) was attached to the 3´end of the ssDNA region (S) after the *nic* site, thus creating a novel loop region (U+S) containing the *nic* site (N) and a dsDNA stem of 6 bp. The length of the loop (U+S) was tuned by including different number of nucleotides in the S region before the proximal arm (P) (see

Table Supplementary S1). Oligonucleotides with redesigned secondary structures were used to analyze $\text{TrwC}_{\text{R388}}$ relaxase activity *in vitro*.

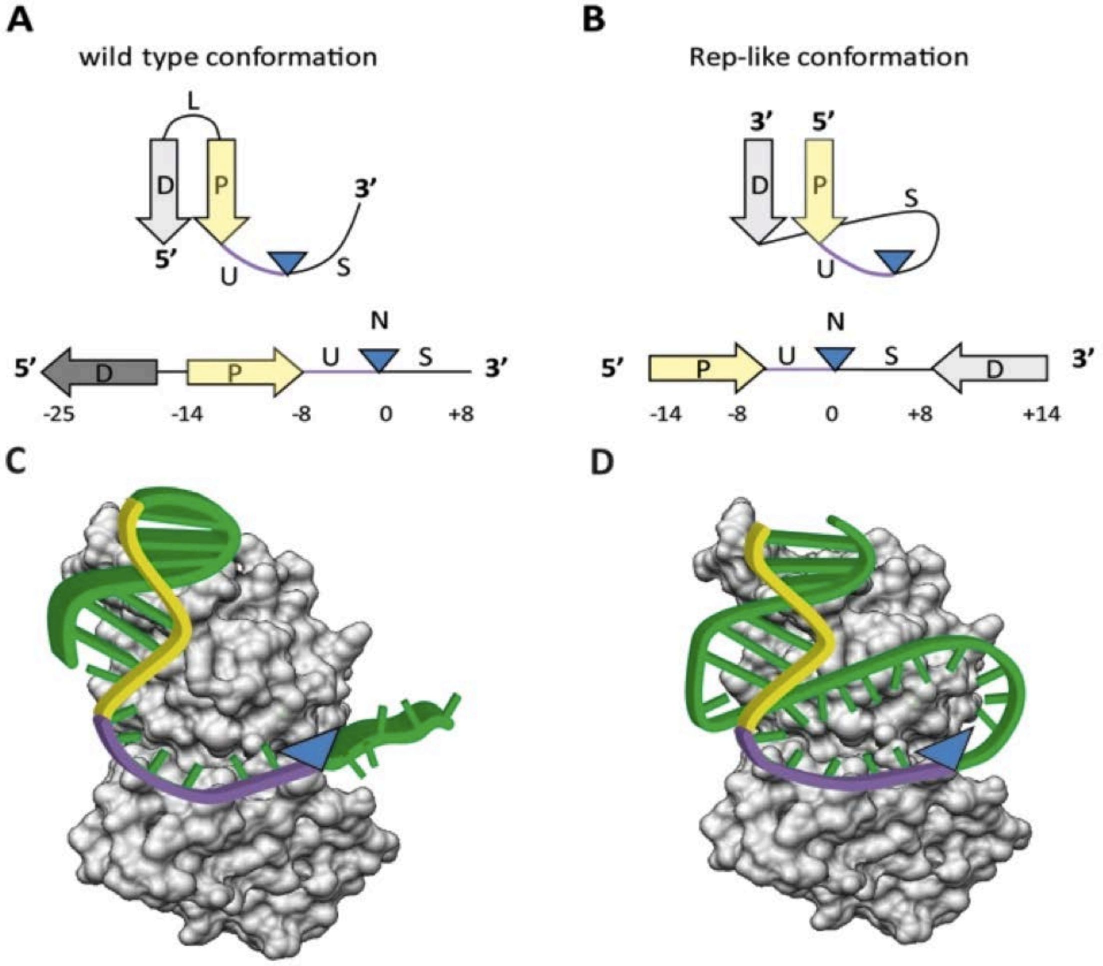


Figure R35. Scheme of relaxase and replicase DNA targets. (A) The relaxase DNA substrate contains an IR, defined by a Distal arm (D) and a Proximal Arm (P), that shape a hairpin structure. The *nic* site (N) is located between a U-turn (U) sequence and a ssDNA strand (S) that is tethered to the relaxase after cleavage. (B) The DNA substrate cleaved by replicases has the Distal arm (D) located downstream from *nic* (N), which allows stem-loop formation. Rep-like substrates for relaxases were designed by displacing the D sequence to the 3' end of S of their original wt substrate. (C) Scheme depicting the cleavage reaction of the wt substrate. Upon binding, relaxase bends its target in order to localize the *nic* site (N) within its active center. In presence of a divalent cation, the relaxase cleaves *nic* (blue arrowhead) and remains covalently bound to the 5'-phosphate of S, downstream from N. (D) Scheme depicting the cleavage reaction of the Rep-like substrate. The stem is bound by the relaxase, and the loop is located within the DNA binding cleft. Thus, the relaxase cleaves the scissile nucleotide within a stable cruciform (blue arrowhead). After cleavage the relaxase will be tethered to S and D. P is shown in yellow and U in purple. Blue triangles show the position of the *nic* site.

The effect of the DNA substrate length and secondary structure on relaxase cleavage was investigated through *in vitro nic*-cleavage reactions. This reaction generates a protein-DNA covalent complex that can be quantified by its lower mobility using SDS-PAGE as described in Experimental procedures.

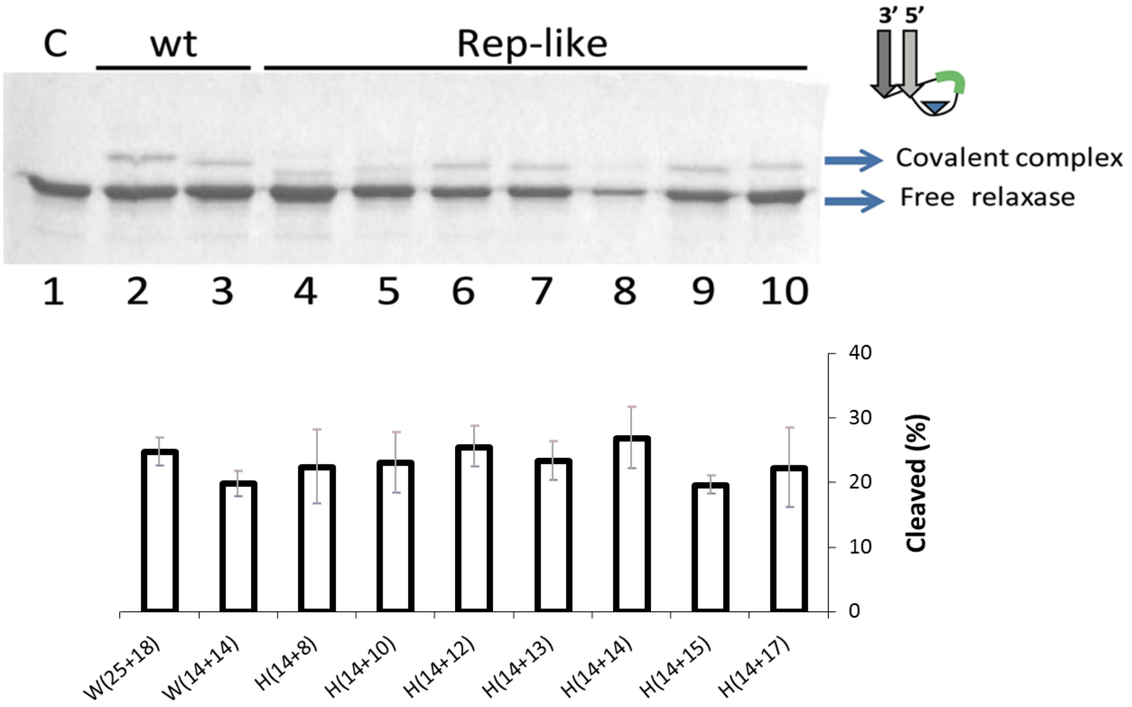


Figure R36. Interaction of $\text{TrwC}_{\text{R388}}$ with Rep-like substrates. (A) SDS-PAGE of oligonucleotides with R388wt or Rep-Like structures, when incubated with $\text{TrwC}_{\text{R388}}$. 6 µM $\text{TrwC}_{\text{R388}}$ was incubated with 15 µM of different oligonucleotides. The reaction products were separated by electrophoresis in 12% SDS-PAGE gels. Lane 1, no oligonucleotide; Lanes 2 and 3, R388wt oligonucleotides W(25+18) and W(14+14), respectively; in subsequent lanes, $\text{TrwC}_{\text{R388}}$ was incubated with Rep-like oligonucleotides. The S length of Rep-like substrates (in green), varies from two to eleven nucleotides. Lane 4, H(14+8), Lane 5, H(14+10), Lane 6, H(14+12); Lane 7, H(14+13), Lane 8, H(14+14); Lane 9, H(14+15); Lane 10, H(14+17). In the center chart, percentage of bound complexes was calculated in three separate experiments such as that shown in (A).

$\text{TrwC}_{\text{R388}}$ was able to cleave and remain covalently bound to oligonucleotides containing IR_2-*nic* (W(25+18)) or P-*nic* (W(14+14)), (Figure R36, Lanes 2 and 3). Interestingly, $\text{TrwC}_{\text{R388}}$ could also cleave a Rep-like oligonucleotide containing a loop with only 2 nucleotides in the S region

(H(14+8)) (Figure R36, lane 4). $\text{TrwC}_{\text{R388}}$ cleavage activity on H(14+8) was similar to that on W oligonucleotides. Incubation of $\text{TrwC}_{\text{R388}}$ with Rep-like hairpins containing longer loops, such as H(14+10), H(14+12) or H(14+13), resulted in a band with reduced mobility, as expected by the formation of a $\text{TrwC}_{\text{R388}}$ covalent complex with a decamer (Lane 5), dodecamer (Lane 6) and tridecamer (Lane 7), respectively. Similar results were obtained with substrates H(14+14), H(14+15) or H(14+17) (Lanes 8, 9 and 10, respectively).

$\text{TrwC}_{\text{R388}}$ protein was previously shown to cleave oligonucleotides containing the R388 *nic* site and to perform an *in vitro* strand transfer reaction to a second oligonucleotide also containing the *nic* site (Grandoso et al., 2000; González-Pérez et al., 2007). Since $\text{TrwC}_{\text{R388}}$ cleaves Rep-like oligonucleotides and remains covalently attached to the oligomer downstream *nic* (Figure R35D), it could carry out the strand transfer reaction. When a labeled wt oligonucleotide acceptor F(25+0) was provided and incubated with $\text{TrwC}_{\text{R388}}$ and oligonucleotide H(14+8), 29 % of the labeled oligonucleotide shifted to a position corresponding to a F(25+8) oligonucleotide (Figure R37). Compared to the transfer reaction using wt oligonucleotides as donors, this transfer efficiency is similar to W(14+14) (31 %) and lower than W(25+18) (41%). Longer Rep-like oligonucleotides, such as H(14+10) and H(14+14), generated around 17% product. Overall, all Rep-like substrates were efficiently transferred to receptor oligonucleotides (see Figure R37).

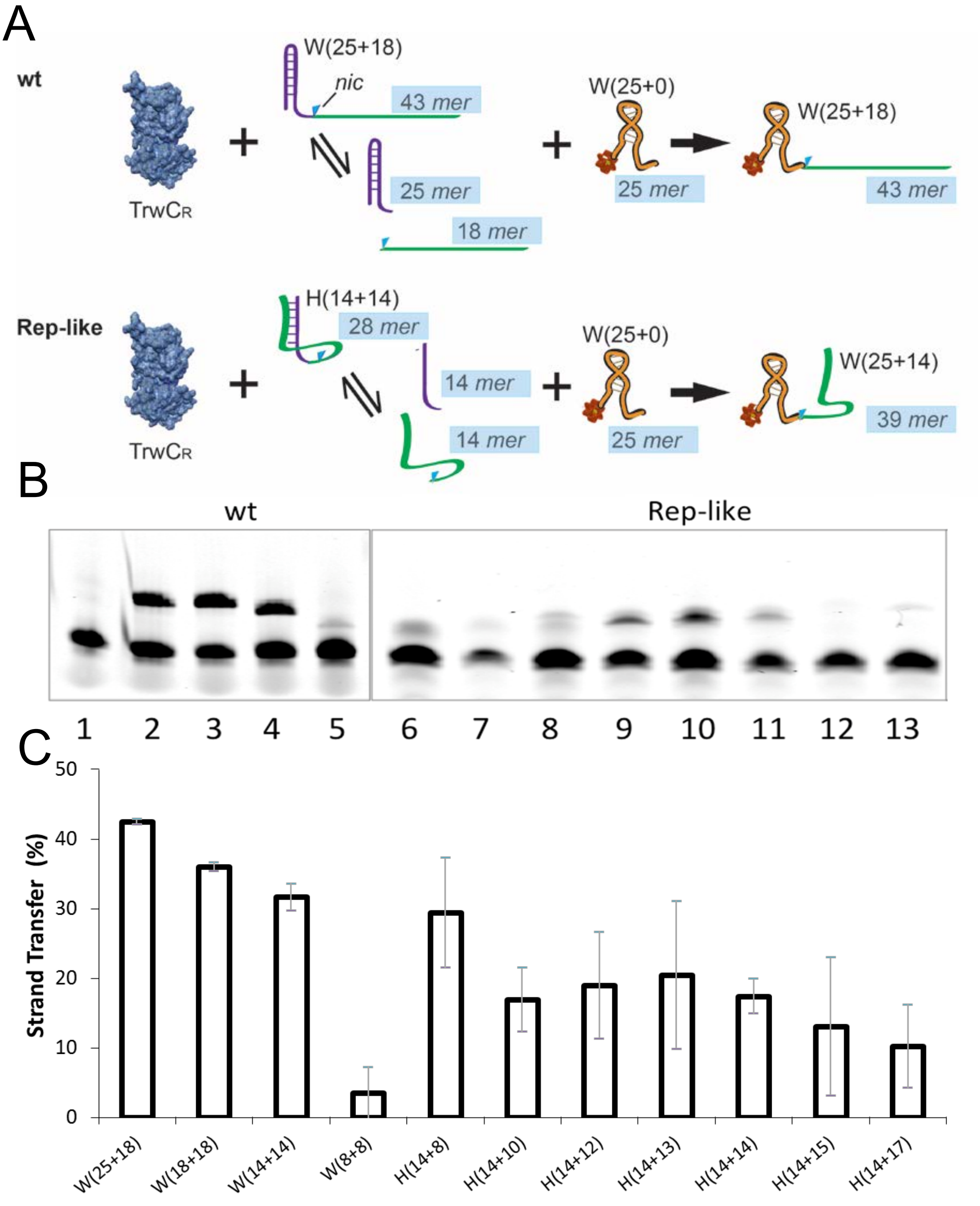


Figure R37. Strand transfer reaction catalyzed by TrwC$_{R388}$ with wt and Rep-like substrates. (A) Schematic representation of the strand transfer reaction. TrwC$_{R388}$ cleaves oligonucleotides containing the R388 *nic*-site and performs the strand transfer reaction to a labeled oligonucleotide W(25+0). In green are shown the oligonucleotides that remain covalently attached to TrwC$_{R388}$. The gels show the appearance of a labeled transfer product. (B) Denaturing gels showing the strand transfer reaction catalyzed by TrwC$_{R388}$ with wt and Rep-like substrates. 5'IRDye-labelled W(25+0) oligonucleotide is shown in control lanes 1 and 7. A 25+x labeled oligonucleotide was obtained after strand transfer reaction with the wild type oligonucleotides W(25+18), W(18+18), W(14+14) and W(8+8) in lanes 2, 3, 4 and 5, respectively. Strand transfer reaction to W(25+0) was also analyzed for Rep-like H14+x oligonucleotides. lane 6, H(14+8); lane 8, H(14+10); lane 9, H(14+12); lane 10, H(14+13); lane 11, H(14+14); lane 12, H(14+15) and lane 13, H(14+17). (C) Bar graph with the quantification of cleaved substrates were calculated in three separate experiments such as that shown in (B). Data shows mean±s.d.

We thought that the distance (S) from *nic* (N) to the proximal arm (P) would be critical for cleavage, since only certain loop lengths would allow the *nic* site to reach the active site of the protein. As similar *nic*-cleavage efficiencies were observed regardless of loop length (Figure R36), we wondered if TrwC$_{R388}$ could melt the hairpin and stabilize a "linear" DNA rather than cleave the stem-loop conformation. In order to study how TrwC$_{R388}$ interacts with Rep-like conformation substrates, increasing concentrations of TrwC$_{R388}$ were incubated with 50 nM of a green-fluorescent labeled Rep-like oligonucleotide H(14+14) or with a red-fluorescent labeled wt oligonucleotide W(25+8) (Figure R38).

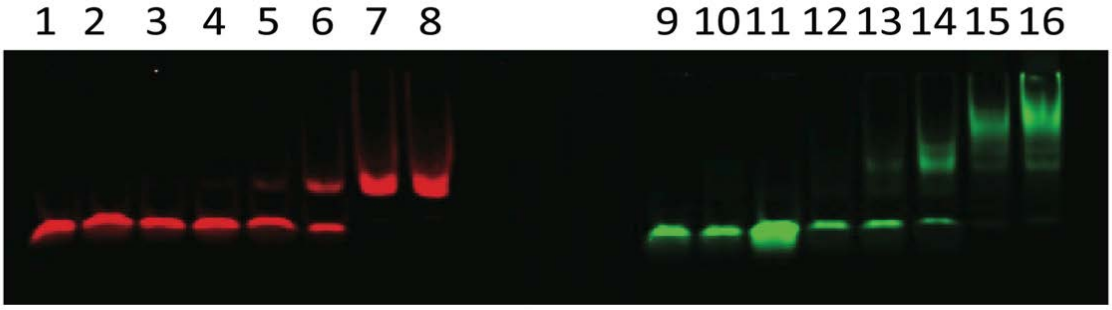


Figure R38. EMSA showing the interaction of TrwC$_{R388}$ to wt and Rep-like substrates. Increasing amounts of TrwC$_{R388}$ were incubated with wt oligonucleotide W(25+8) (red shift, lanes 1 to 8) or Rep-like hairpin H(14+14) (green shift, lanes 9-16). Lanes 1 and 9, no protein added; 2 and 10, 42 nM of TrwC$_{R388}$; 3 and 11, 85 nM; 4 and 12, 210 nM; 5 and 13, 420 nM; 6 and 14,850 nM; 7 and 15, 4,2 µM, 8 and 16, 8,5 µM.

Although the EC$_{50}$ (half maximal effective concentration) was the same with both oligonucleotides, only one shifted band was observed with W(25+8), while several shifted bands appeared with H(14+14). Both unbound oligonucleotides produced a single band (Figure R38, lane 1 and 9, respectively). Thus, upon TrwC$_{R388}$ binding, W(25+8) seems to form a unique complex, while different complexes are formed with H(14+14).

The interaction profile of Rep-like oligonucleotides incubated with $TrwC_{R388}$ was further analyzed by gel filtration. Consistent with the EMSA result, several peaks with higher molecular weight than the predicted peak for 1:1 complexes appeared when either Rep-like H(14+14) and H(16+16) (S=8), H(14+15) (S=9) or H(14+17) (S=11) oligonucleotides were incubated with $TrwC_{R388}$ (Figure R39). In order to stabilize the stem-loop structure of Rep-like substrates, longer IRs were used (H(23+23), H(23+26), H(24+24), H(24+27), H(24+31)). Now, a sole band of protein-DNA was shown in chromatography. Moreover, SDS-PAGE gels revealed that the covalent complexes are still being formed with longer stem substrates (Figure Supporting S23).

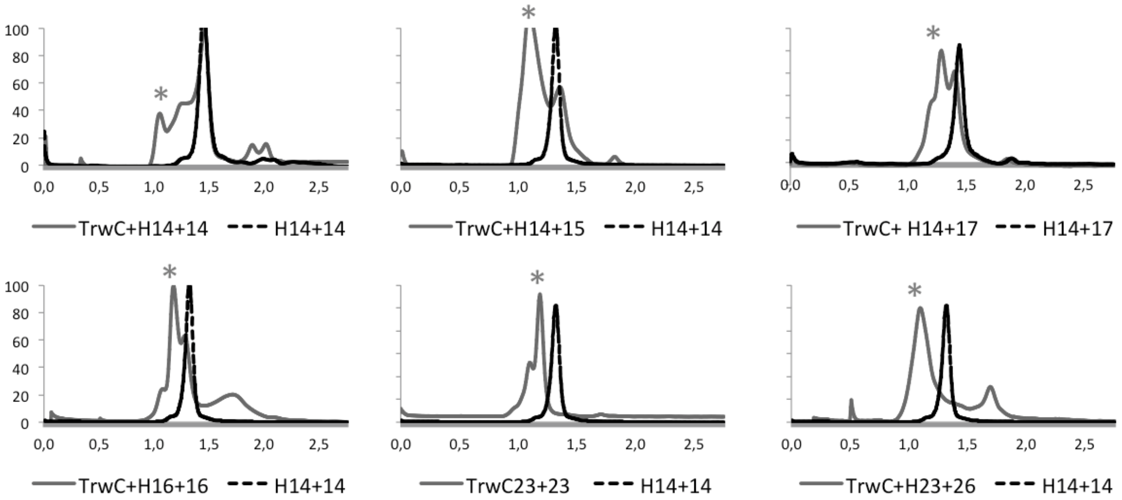


Figure R39. Chromatograms of Rep-like oligonucleotides with or without $TrwC_{R388}$. 6.3 µM $TrwC_{R388}$ was incubated during 1 h in presence of EDTA with a 1.5:1 molar excess of each oligonucleotide. Then 20 µl of the samples were injected in a S75 column using the ETHAM system (GE Biosciences). Free oligonucleotides at the same concentration were also injected separately for comparison. Chromatograms of oligonucleotides are shown as dashed lines while $TrwC_{R388}$ with oligonucleotides are shown as continuous lines. Several shifted peaks and broaden peaks are observed in the cases of short hairpins H(14+14), H(14+15) and H(14+17). $TrwC_{R388}$ interacts with oligonucleotides with longer stems H(16+16), H(23+23), and H(23+26) in a similar fashion. Oligonucleotides tested were Rep-like oligonucleotides with D=P=6 and different loop lengths; H(14+14) S=8 , H(14+15) S=9 and H(14+17), Rep-like oligonucleotides with S=11 and different stem lengths: H(16+16) D=P=8 S=8, H(23+23) D=P=15 S=8, and H(23+26) D=P=15 S=11. Absorbance at 260 nm was used during chromatography to determine the presence of DNA. All the chromatograms are normalized and shifted peaks are shown by stars.

Fractions separated by gel filtration chromatography of a sample of TrwC$_{R388}$ incubated with IRDye H(14+14) oligonucleotide in presence of Mg^{2+} were simultaneously analyzed on an acrylamide native gel. Similarly, several bands appeared in the peak with higher molecular weight (peak 4, Figure R40). Peak 3 corresponded to the 1:1 complexes, whereas peak 4 could be the 2:2 complex. It is likely that TrwC$_{R388}$ could melt the hairpin and stabilize a “linear” DNA in a way that the distal and proximal arms of the IRs of different DNA molecules could hybridize, creating these higher molecular height complexes.

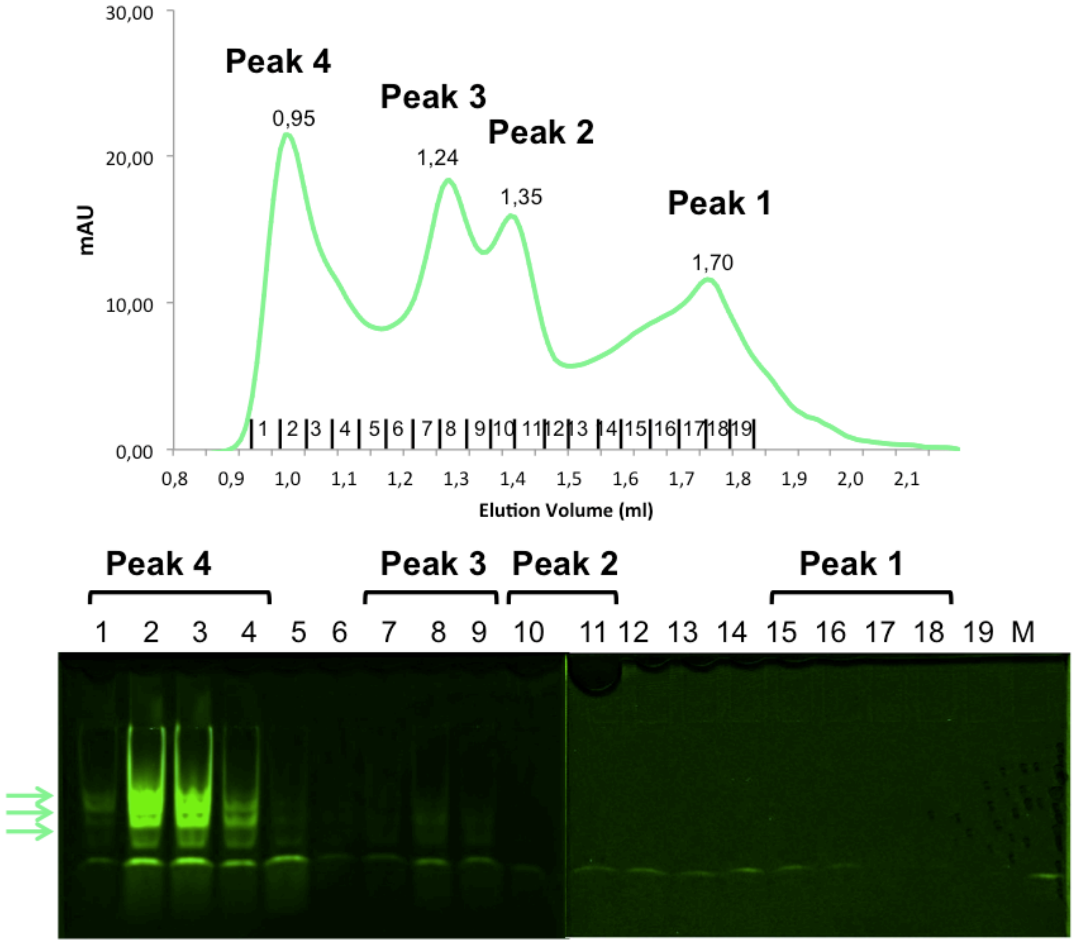


Figure R40. Chromatogram and acrylamide native gel analysis of TrwC$_{R388}$ with fluorescent oligonucleotide H14+14. 6.3 µM TrwC$_{R388}$ was incubated during 1 h at RT in presence of 10 mM MgCl$_2$ with a 2:1 molar excess of IRDye H(14+14) oligonucleotide. Then 20 µl of the sample was injected in a S75 column using the ETHAM system (GE Biosciences). The chromatogram shows the obtained four A$_{260}$ peaks with the Ve of each peak. After gel filtration column chromatography, 20 µl of 19 fractions collected within these four peaks were loaded into an acrylamide native gel. Lane M contains free oligo. Shifted oligonucleotides are shown by arrowheads. Gels were scanned in Odyssey (LI-COR).

6.2 Rep-like substrates are cleaved by single-Y relaxases

The ssDNA U-turn observed in Y2 relaxases TrwC$_{R388}$- and TraI$_{F}$ -DNA complex structures was also observed in the structure of Y1 relaxase NES (Boer et al., 2006; Lujan et al., 2007; Edwards et al., 2013). However, relevant differences in U-turn formation have been described (Carballeira et al., 2014). In order to check if Y1 relaxases could also catalyze cleavage of Rep-like substrates, we analyzed these substrates for the best known Y1-relaxases, MobA$_{RSF1010}$ and TraI$_{RP4}$. The MobA$_{RSF1010}$ binding site was identified as a 10 bp IR that forms a hairpin structure downstream *nic* (Monzingo et al., 2007). Mozingo et al. modeled the relaxase MobA$_{RSF1010}$ complexed with a 33-mer oligonucleotide and compared this model with TrwC$_{R388}$ bound to the 23-mer (Monzingo et al., 2007) (Figure R41A). According to this comparison, only nucleotides at the base of the 10 bp-hairpin interact with MobA$_{RSF1010}$. To study the influence of IR length, we designed a shorter wt substrate, containing a 6 bp IR WQ(23+7) (Figure R41B).

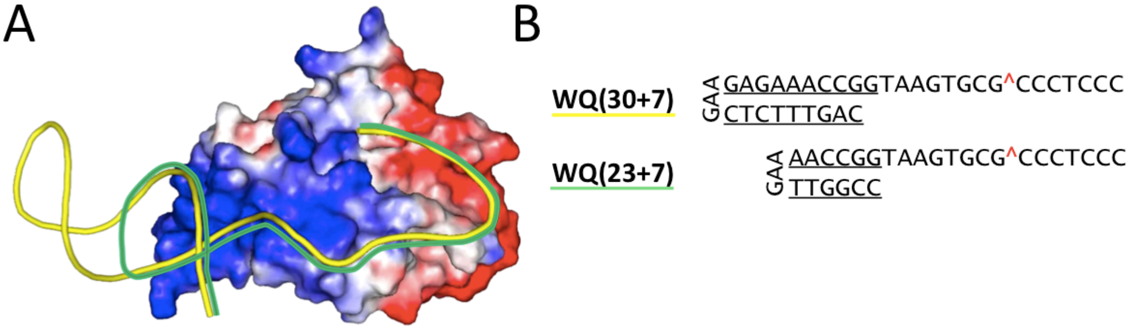


Figure R41. Model of MobA$_{RSF1010}$-DNA complex. (A) The surface electrostatic potential of the protein structure is show. DNA is depicted as a ribbon. WQ(30+7) substrate is shown in yellow, whereas the green ribbon depicts the shorter wt substrate WQ(23+7). Modified from Monzingo et al., 2007. (B) Sequence of the wt substrate WQ(30+7) and the truncated version WQ(23+7) that lacks the nucleotides of the top of the hairpin.

Analysis by SDS-PAGE of the covalent complexes obtained after incubation with the protein showed that either the long wt substrate WQ(30+7) (Figure R41 Lane 2) or the short wt substrate WQ(23+7) (Figure R41 Lane 3) generated 40% covalent complexes. These results confirm that just the nucleotides at the base of the stem make specific interactions with relaxase MobA$_{RSF1010}$.

Rep-like substrates of RSF1010 *nic-site* sequence were designed as for R388 *nic* (Figure R35). Three different loop lengths (S) were tested, while maintaining the first 18 nucleotides invariant (P-U-N). The three different oligonucleotides tested, HQ(18+16), HQ(18+19) or HQ(18+22), showed reduced activity compared with wt Q oligonucleotides. MobA$_{RSF1010}$ cleaved the Rep-like substrate HQ(18+16) to 13% (Figure R41 Lane 4). HQ(18+19) (Lane 5) behaved similarly, producing 10 % covalent complexes. In turn, HQ(18+22) (Lane 6) generated 27% covalent complexes (Table Supplementary S15). These results suggest that Rep-like substrates also allow the correct location of the *nic* site in the catalytic center of the relaxase MobA$_{RSF1010}$.

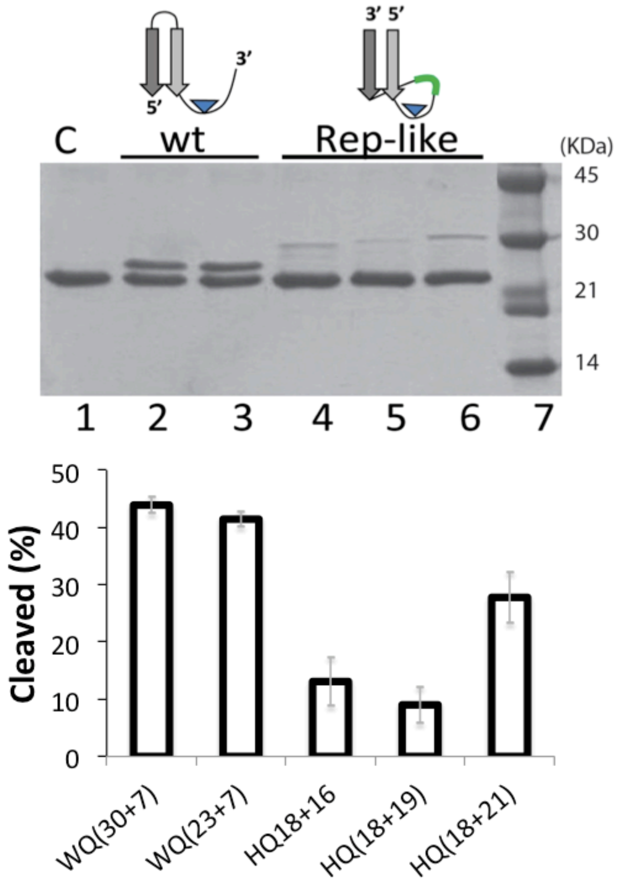


Figure R42. SDS-PAGE of MobA$_{RSF1010}$ with its targets. 7 µM MobA$_{RSF1010}$ was incubated with 15 µM of different oligonucleotides. Lane 1, MobA$_{RSF1010}$. Lanes 2 and 3, wt oligonucleotide WQ(30+7) and a substrate that lacks the upper-hairpin nucleotides of the IR WQ(23+7). Lanes 4, 5 and 6, Rep-like oligonucleotides HQ(16+16),HQ(16+19) and HQ(16+22), respectively. Lane 7, molecular weight ladder. Bar graph with the quantification of covalent complexes. Data showed mean±s.d. of three independent experiments.

Another widely characterized Y1 relaxase is TraI$_{RP4}$ (Pansegrau et al., 1993; 1994; Pansegrau and Lanka, 1996). The DNA substrate routinely used in TraI$_{RP4}$ cleavage assays was the 21-mer oligonucleotide WP(15+6). None of the previous assays had used any substrate carrying an IR downstream from *nic*. In our study, a substrate with a 6-bp hairpin conformation WP(24+8) was designed. As shown in Figure R43, WP(24+8) was efficiently cleaved by TraI$_{RP4}$. Covalent complex formation by the wt substrate was slightly higher than when WP(24+8) was used (Lane 3).

We then used a rational approach to design DNA stem-loops that could be cleaved by TraI$_{RP4}$. Rep-like oligonucleotides containing the minimal RP4 *nic* sequence were developed as we did before for R388 (P-U-N-L-D, see Figure R34). These substrates were incubated with TraI$_{RP4}$, and the resulting complexes where analyzed by SDS-PAGE. Neither Rep-like oligonucleotide HP(14+14) (Figure R43 Lane 4) nor HP(14+21) were cleaved, even though saturating concentrations of oligonucleotides where used (data not shown).

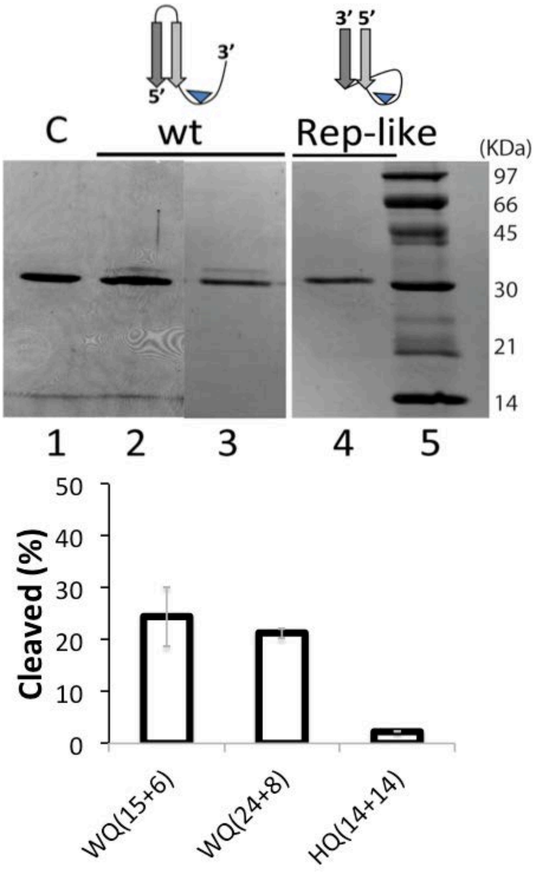


Figure R43. SDS-PAGE of TraI$_{RP4}$ with its targets. 1.5 µM TraI$_{RP4}$ was incubated with 15 µM of different oligonucleotides. Lane 1, TraI$_{RP4}$; lane 2, wt substrate WP(15+6); lane 3, wt substrate WP(24+8); lane 4, Rep-like substrate HP(14+14) and lane 5, molecular weight ladder. Bar graph with the quantification of covalent complexes is shown below the SDS-PAGE gel. Data showed mean±s.d. of three independent experiments.

6.3 Improved scissile substrates were obtained by shuffling the *nic* sequence

In TrwC$_{R388}$ wt substrate W(25+8), the stem loop is non-covalently bound by the relaxase after cleavage, while the 8-mer downstream *nic* is covalently attached to the catalytic tyrosine. There is a cleavage-ligation equilibrium with wt oligonucleotides, because the 25-mer oligonucleotide remains in the protein DNA binding domain after cleavage and, therefore, the 8-mer can be easily religated (Figure R35C).

We thought that, by linking the 3´ end of *nic* (U-N-S) to the 5´end of the hairpin, the 5´side of *nic* (the single strand U sequence upstream the *nic* site) would lose binding affinity (Figure R44A). As a consequence, the resulting 5´ssDNA could be released from the relaxase, avoiding religation and displacing the reaction equilibrium towards the formation of covalent complexes (Figure R44B). Under this rationale, novel substrates were designed by shuffling the wt sequence in the order U-N-S-D-P. These synthetic oligonucleotides were called “reverse substrates” because the hairpin is located downstream from the *nic* site.

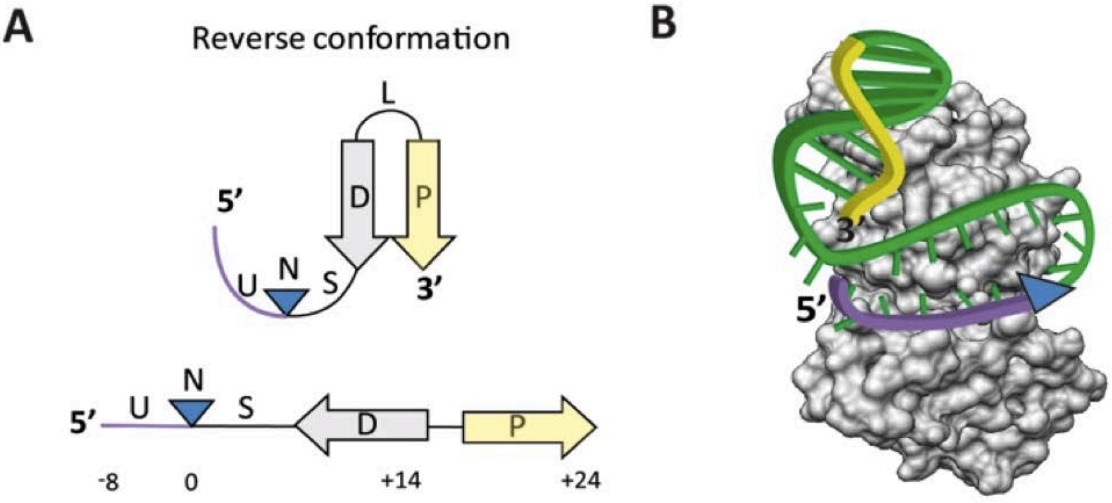


Figure R44. Design of Reverse substrates. (A) Reverse substrates were designed by swapping the 5’ region of the *nic* site to the 3’ end. This designed DNA substrate possesses the complete inverted repeat (D-P) at the 5’ end of the *nic* site (N). Either the U or the S lengths were tuned to allow the correct location of the hairpin within the relaxase binding domain. (B) Scheme depicting the cleavage reaction of the reverse substrate. Relaxase binding to the reverse substrate allows both the hairpin and the single strand U-turn localize at the DNA binding cleft. This way the cleavage reaction forms a covalent complex of the relaxase with the region downstream of the *nic* site (blue arrowhead). Now the 5´side of *nic* do not contain the IR avoiding the re-ligation reaction.

R388 reverse oligonucleotides were designed using either the whole U ssDNA (8 nt before *nic*) or a fragment (up to seven nt), which would be liberated after cleavage. Besides, S length varied from 8 to 11 nucleotides, to allow the correct disposition of the hairpin into the relaxase. As shown in the SDS-PAGE gel of Figure R45, the relaxase band shifted when incubated with R(8+24), R(7+27) or R(8+27) reverse oligonucleotides. In fact, the percentage of protein-bound DNA boosted with reverse oligonucleotides R(8+27) (Lane 4) and R(8+24) (Lane 5) to almost 50% when U had 8 nt. It is noteworthy that the covalent complex decreased to 18% when the U length was 7 nt (lane 6). Moreover, the covalent product was not detected when the U-region was shorter, such as when oligonucleotides R(4+27) (Lane 6), R(1+27) (Lane 7) or R(0+27) (Lane 8) were used. The different yields of covalent complexes revealed that U length is a determinant factor for the reaction, while S length had no significant influence.

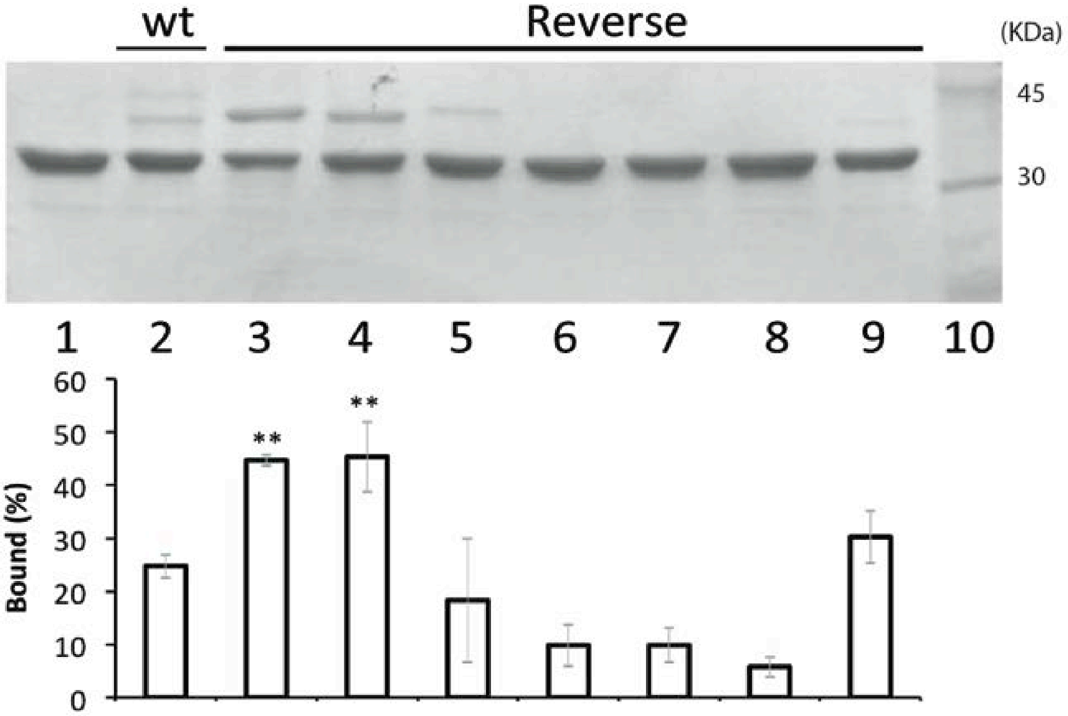


Figure R45. Interaction of TrwC$_{R388}$ with Reverse substrates. 12% SDS-PAGE of reverse oligonucleotides, when incubated with TrwC$_{R388}$. 6 μM

TrwC$_{R388}$ was incubated with 15 μM of different reverse oligonucleotides. Lane 1, no oligonucleotide; Lanes 2, R388wt oligonucleotide W(25+18). Lanes 3 and 4, Reverse substrates R(8+27) and R(8+24), both with U=8 nt and S=11nt or S=8 nt respectively. Lane 5, R(7+27) U=7; Lane 6, R(4+27) U=4; Lane 7, R(1+27) U=1 and Lane 8, R(0+27) U=0. Lane 9, R(8+14), U=8 P=0. Lane 10, Molecular weight marker. Graph quantifying the percentages of covalent complexes is shown below the SDS-PAGE gel. Data show mean±s.d. of three independent experiments. Two asterisks indicate P-value<0.05 by two-sided student's t-text.

The complexes of TrwC$_{R388}$ with reverse oligonucleotides were further analyzed with gel filtration S75 column using the ETHAM system (GE Biosciences). The complexes eluted as a single sharp peak as is shown in Figure R46 A 1:1 stoichiometry was assumed for comparison with the protein-wt substrate chromatogram.

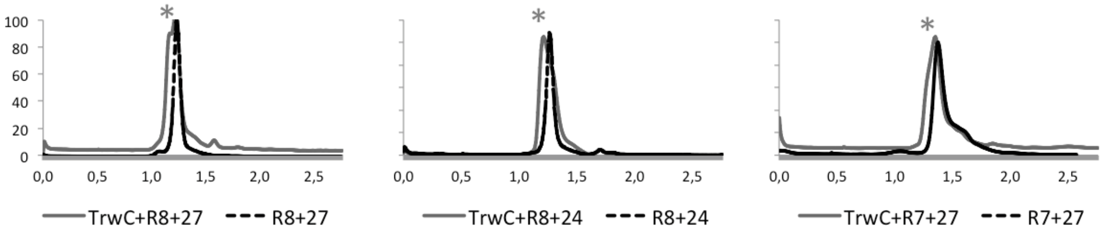


Figure R46. Chromatograms of reverse oligonucleotides with or without TrwC$_{R388}$. 6.3 μM TrwC$_{R388}$ was incubated during 1 h in presence of EDTA with a 1.5:1 molar excess of each oligonucleotide. Free oligonucleotides at the same concentration were also injected separately for comparison. Chromatograms of oligonucleotides are shown as dashed lines while TrwC$_{R388}$ with oligonucleotides are shown as continuous lines. Reverse oligonucleotides R(8+27), R(8+24), R(7+27) show a slight shift when TrwC$_{R388}$ is bound to it, similar to the one obtained with wt oligonucleotides (data not shown). R(8+27) D=P=6 U=8 S= 11, R(8+24) D=P=6 U=8 S= 8 and R(7+27)with D=P=6 U= 7 S= 11. Absorbance at 260 nm was used during chromatography to determine the presence of DNA. All the chromatograms are normalized and shifted peaks are shown by stars.

DNA processing of reverse substrates was also analyzed by the strand transfer reaction. When a labeled-wild type oligonucleotide acceptor (F25+0) was provided, only 4% of this oligonucleotide shifted to a position corresponding to a F25+27 oligonucleotide when incubated with TrwC$_{R388}$ and R8+27 while the rate of strand transfer when using WT 25+8 oligonucleotide is around 43% (Table R3).

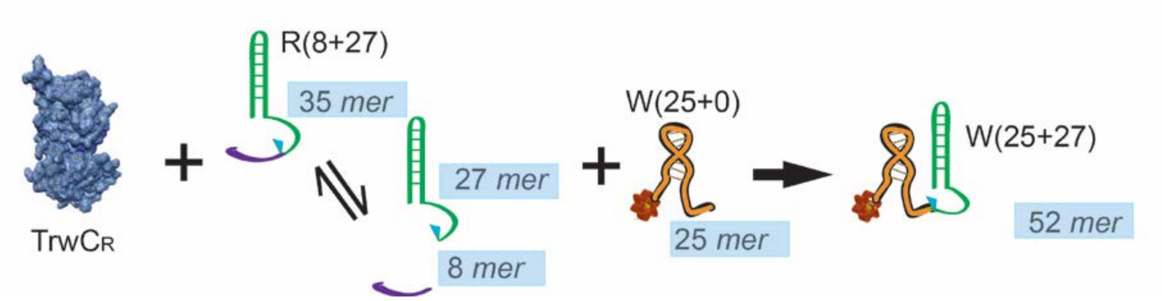


Figure R47. Schematic representation of the strand transfer reaction with reverse substrates. TrwC$_{R388}$ cleaves reverse oligonucleotides and performs the strand transfer reaction to a wt labeled oligonucleotide W(25+0). The oligonucleotides that remain covalently attached to TrwC$_{R388}$ are shown in green. Denaturing gels show the appearance of a labeled transfer product.

<u>Table R3.</u> Strand transfer reaction catalyzed by TrwC$_{R388}$ of oligonucleotides with reverse layout of the W(25+0) oligonucleotide.

Nicking substrate[a]	Transfer activity %[b]
R388 W(25+18)	43 ±1
R388 R(8+27)	4 ± 2
R388 R(8+24)	2 ± 2
R388 R(7+27)	3 ± 3
R388 R(8+14)	2 ± 3

[a] Substrates sequences are in Table S1 [b] The strand transfer activity of the relaxase was measured by the formation of fluorescent W(25+x) oligonucleotides in denaturing PAGE. Values are the average of three experiments.

Regarding Y1 relaxases, reverse substrate RP(8+24) was recognized and cleaved by TraI$_{RP4}$ (Figure R48A, Lane 2). The amount of covalent complex obtained with RP(8+24) was comparable to the amount obtained with wt WP(24+8) oligonucleotide. Interestingly, MobA$_{RSF1010}$ could not cleave similarly designed oligonucleotides RQ(8+28), RQ(8+34) or RQ(8+40) (Figure R48B Lane 2, 3 and 4, respectively).

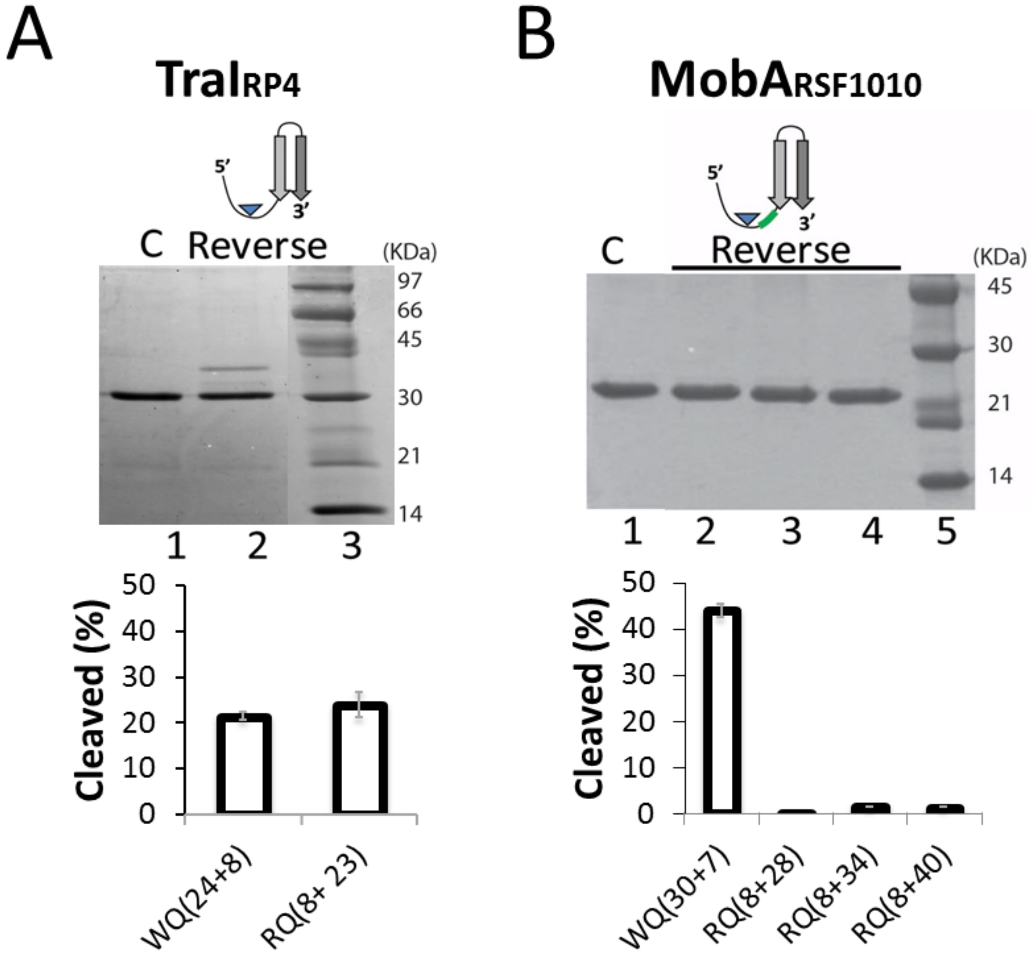


Figure R48. Reverse substrates of model single-Y relaxases. (A) SDS-PAGE of $\mathrm{TraI_{RP4}}$ with reverse target. Lane 1, $\mathrm{TraI_{RP4}}$; lane 2, reverse substrate RP(8+24) and lane 3, molecular weight ladder. (B) SDS-PAGE of $\mathrm{MobA_{RSF1010}}$ with reverse targets. Lane 1, $\mathrm{MobA_{RSF1010}}$. Lanes 2, 3 and 4 reverse substrates RQ(8+28), RQ(8+34) and RQ(8+40). Lane 5, molecular weight ladder. Bar graphs with the quantification of covalent complexes are shown below the SDS-PAGE gels. Data showed mean±s.d. of three independent experiments.

6.4 Plasmids carrying novel *nic*-site conformations show diminished conjugation rates

The stable interaction of $\mathrm{TrwC_{R388}}$ with the novel substrates supports the notion that both Rep-like and reverse oligonucleotides could modify the *in vivo* efficiency of relaxases. The importance of the DNA stem loop structure for efficient relaxase-mediated DNA processing *in vivo* was investigated by using plasmids containing synthetic *oriT*s. Synthetic R388 *oriT*s were designed by substitution of the wt *nic*-cleavage site ($\mathrm{IR_2}$+*nic*) by either H(14+14) Rep-like or R(8+24) Reverse-like sequences (Figures R49 and Supplementary S24).

OriT-containing plasmids (Cm^R) can be mobilized to a recipient cell (strain UB1637; Sm^R) in the presence of the helper plasmid pSU2007 (Km^R). Mobilization frequencies of the synthetic *oriT*-containing plasmids were measured as the number of mobilized plasmids (Cm^RSm^R) per donor cell (Cm^RNx^R) (Figure R48). The Rep-like *oriT* plasmid showed 10^4 fold lower mobilization than the wt o*riT* plasmid. By substitution of the Rep-like site for the wt *nic*-site, not only the relaxase recognition site was changed, but also the distance between the TrwC and TrwA/IHF binding sites. TrwA binding to *oriT* was shown to be essential for proper conjugation (Moncalián and la Cruz, 2004). In the Rep-like *nic*-site, TrwC and TrwA binding sites were 10 nucleotides closer. Thus, we also analyzed the mobilization frequency of a synthetic plasmid that lacks the distal arm (ΔDistal Arm *oriT*). Interestingly, the conjugation rate of ΔDistal Arm *oriT* plasmid dramatically dropped below detection levels (<10^{-6} transconjugants/donor).

We also checked the conjugation ability of a plasmid containing just the Rep-like *nic* site without any binding sites for IHF and TrwA, named minimal Rep-like. Mating carried out with minimal Rep-like *oriT* did not generate transconjugants. Similarly, reverse-like plasmids, which have the *nic* site placed upstream IR_2, also resulted in a substantial reduction of plasmid conjugation.

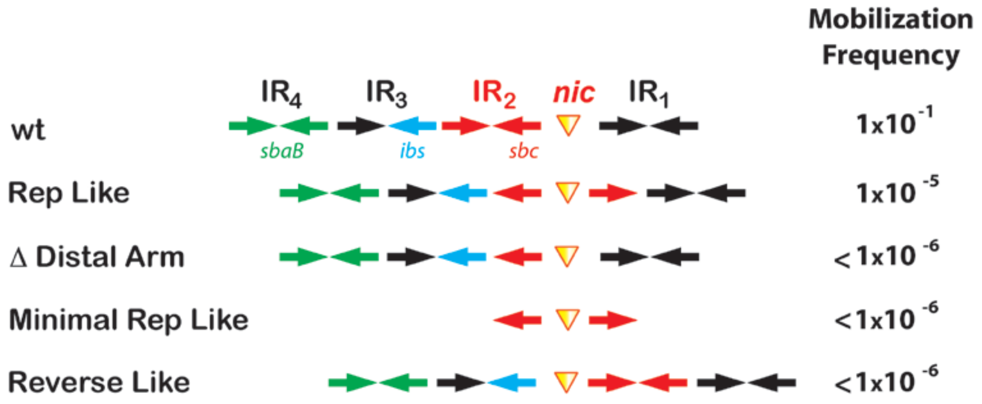
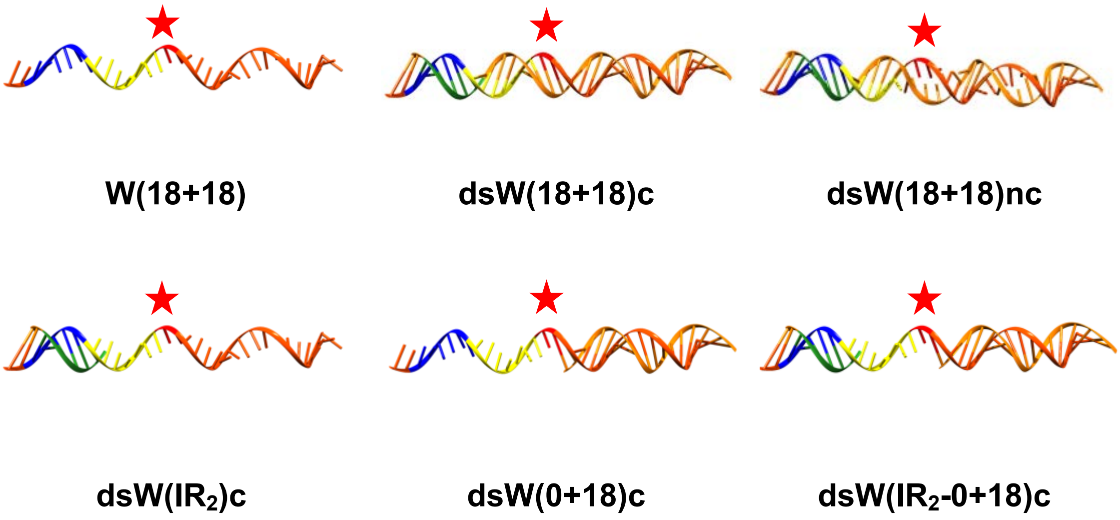


Figure R49. Mobilization frequencies of synthetic *oriT*-plasmids. Mobilization frequencies obtained using pSU2007 as helper plasmid are expressed as number of transconjugants per donor cell. Values are the average of three experiments. The relative position of the secondary structure elements of the assayed synthetic *oriT*s is shown. R388 wt *oriT* has four IR that are binding sites for the relaxase and nicking-accessory proteins TrwA and IHF. TrwC binds IR_2 and the *nic* site (*sbc*, red), TrwA binds IR_4 (*sbaB*, green) while IHF binds IR_3 (*ibs*, blue). The distance between secondary structure elements is not shown at scale. See Figure supplementary S25 for details.

6.5 Optimization of the DNA cleavage reaction by using dsDNA around the *nic* site

To date, although TrwC$_{R388}$ recognizes and cleaves dsDNA that is partially melted around the *nic* site, there are no studies of the influence of the presence of dsDNA around the *nic* site. Several substrates that hybridize with the target W(18+18) were used to study the feasibility to carry out the cleavage reaction when dsDNA is in the boundary of the *nic* site. The different substrates are schematized in Figure R50.


W(18+18) **dsW(18+18)c** **dsW(18+18)nc**

dsW(IR$_2$)c **dsW(0+18)c** **dsW(IR$_2$-0+18)c**

Figure R50. Schematic representation of the substrates used for the determination of the effect of the double strand around the *nic* site. The "complementary" strand forms a duplex with the target W(18+18) in the dsW(18+18)c substrate. However, in dsW(18+18)nc, the "non-complementary" strand generates a bubble of 18 nucleotides around the *nic*-containing central part of the oligonucleotide W(18+18). The dsW(IR$_2$)c substrate contains duplex sequence at the inverted repeat (IR$_2$ Distal Arm). In dsW(0+18)c substrate, the "2nd half *nic* 0" oligonucleotide is complementary to the region upstream of the *nic* site, where 0 means that the duplex starts right at the *nic* site. The dsW(IR$_2$-0+18)c substrate contains duplex sequence at the IR$_2$ and the region upstream of the *nic* site. The proximal arm of the IR is blue, the distal arm is colored green. Yellow indicates the U-turn. The *nic* site is depicted in red and with a red star. Sequences are listed in Table Supplementary S1.

The efficiency of the cleavage reaction was analyzed by SDS-PAGE as described in Experimental procedures (Figure R51A). The relaxase was not able to cleave linear dsDNA containing *nic* (lane 2), as it was reported elsewhere (González-Pérez et al., 2007). But when the duplex has several mismatches (dsW(18+18)nc), TrwC_{R388} was able to form covalent complexes (lane 3) with similar yield than the single stranded substrate (lane 1). Strikingly, the results revealed that dsW(0+18)c produced a 4-fold increase in the amount of relaxase-DNA covalent complexes (lanes 5 and 6). The percentage of covalent complexes was constant over time when TrwC_{R388} was incubated with dsW(0+18)c (Figure Supplementary S25). The duplex at the IR_2 slightly affects the yield (lanes 4 and 6).

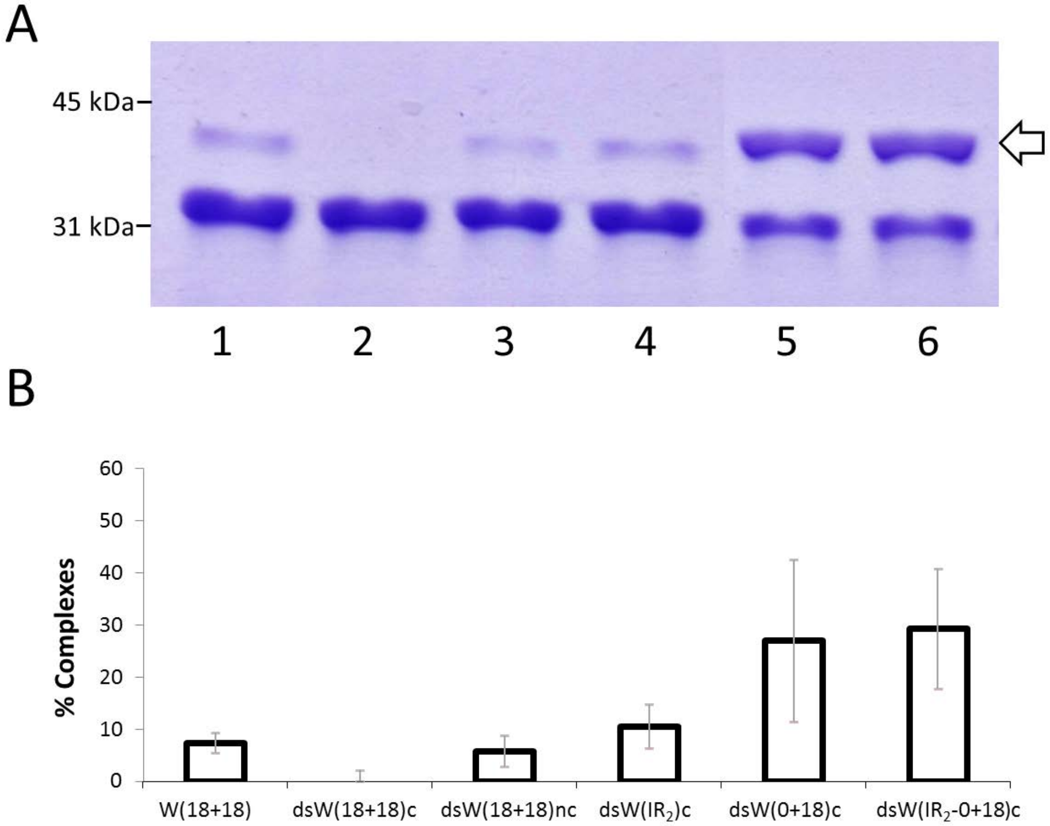


Figure R51. DNA cleavage reaction of TrwC_{R388} by using substrates with double stranded DNA around the *nic* site. (A) SDS-PAGE of the covalent complex formation of TrwC_{R388} to substrates with double strand at the boundary of the *nic* site. See Figure R50 for further details of the substrates. Lane 1, TrwC_{R388} incubated with W(18+18); Lane 2, TrwC_{R388} with dsW(18+18)c; Lane 3, TrwC_{R388} with dsW(18+18)nc; Lane 4, TrwC_{R388} with dsW(IR_2)c; Lane 5, TrwC_{R388} with dsW(0+18)c; Lane 6, TrwC_{R388} with dsW(IR_2-0+18)c. The arrow indicates the covalent complex TrwC_{R388}-DNA. (B) Bar graphs with the quantification of covalent complexes. Data showed mean±s.d. of three independent experiments.

To confirm that the substrate with double stranded DNA around the *nic* site is indeed cleaved more efficiently than the single stranded substrate, electrophoresis in a denaturing urea-polyacrylamide gel was performed. 500 nM of TrwC$_{R388}$ was incubated with 50 nM of 5'-labelled oligonucleotide W(18+18) annealed to different substrates complementary to the *nic* site region. An excess of the second oligonucleotide has been used to ensure that all the nic-containing DNA strand was forming a duplex DNA. After 1 h at 37ºC, samples were treated with proteinase K and SDS to release the products (Figure R52A). As shown in Figure R52B, dsW(0+18)c substrate exhibited the highest cleavage efficiency. The percentage of cleaved substrates was in accordance with the percentage of covalent complexes shown previously.

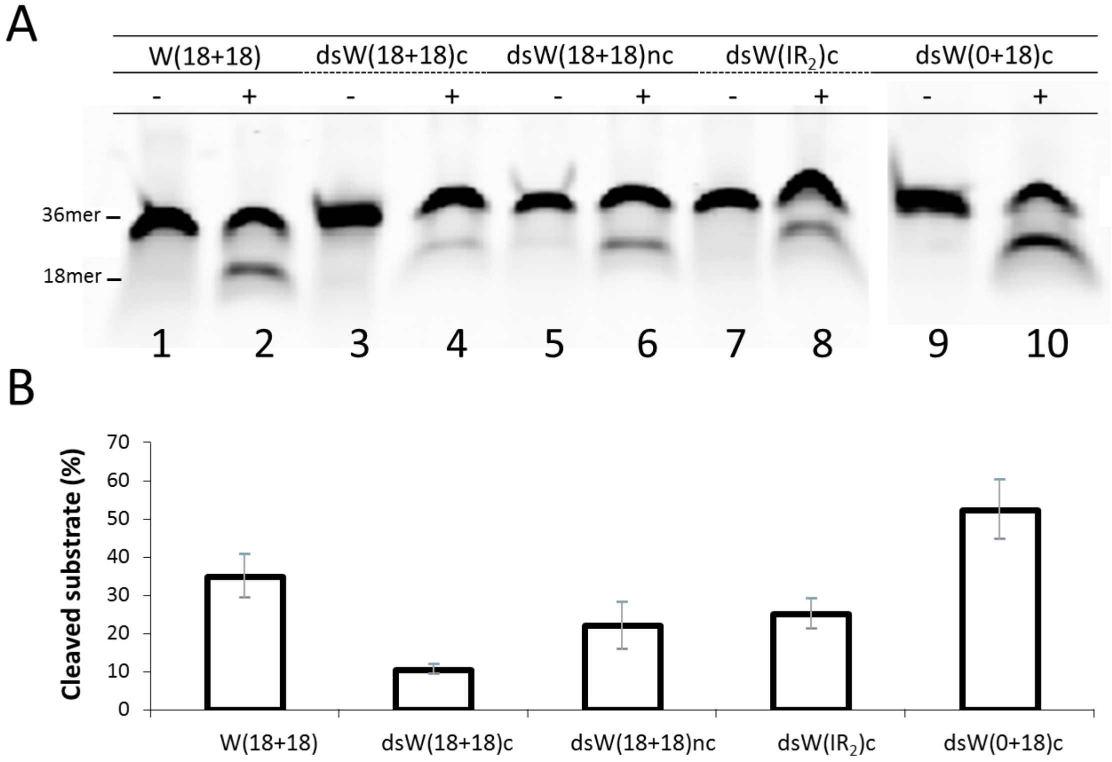


Figure R52. Cleavage reaction of TrwC$_{R388}$ on oligonucleotides containing double strand at the *nic* site. A) Denaturing gel showing the DNA cleavage assay of TrwC$_{R388}$ to 5'IRDye-labelled W(18+18) oligonucleotide annealed with its unlabeled, complementary oligonucleotide. See Figure R50 for further details. 5'IRDye-labelled W(18+18) oligonucleotide is shown in control lanes 1, 3, 5, 7 and 9 (denoted with a minus "-" symbol). After incubation for 1 h (Lanes with a plus "+" symbol) and protein digestion, a cleaved substrate 18-mer appeared. Lane 1 and 2, the single stranded substrate, W(18+18), without (-) and with (+) incubation with TrwC$_{R388}$, respectively. Lane 3, dsW(18+18)c incubated with TrwC$_{R388,}$ lane 4. The substrate dsW(18+18)nc, without (-) and with (+) TrwC$_{R388}$, lane 5 and 6, respectively. dsW(IR$_2$)c without (-) and with (+) TrwC$_{R388}$, lane 7 and 8. The substrate dsW(0+18)c (lane 9) was incubated with TrwC$_{R388}$, lane 10. B) Bar graphs with the quantification of cleaved substrates were calculated in three separate experiments such as that shown in (A). Data shows mean±s.d.

Interestingly, our results shown in Figures R51 and R52 indicate that the percentage of cleaved substrates increases when the duplex region begins at the *nic* site (2^{nd} Half *nic* 0). Although the relaxase can cleave partially melted duplexes, such as the dsW(18+18)nc that possesses 18 mismatches in the central region, it cannot cleave linear duplexes. Moreover, the inverted repeat enhances the religation reaction (see results of dsW(IR_2)c), as it has been reported elsewhere (Lucas et al., 2010).

In order to better understand how the relaxase was achieving this new balance, the mutants $TrwC_{R388}$(Y18F) and $TrwC_{R388}$(Y26F), with one catalytic tyrosine (Y26) and (Y18) respectively, were tested. $TrwC_{R388}$(Y18F) generated the same amount of covalent complexes after incubation with oligonucleotides with double strand at the *nic* site (Figure R53, lane 3) than with the W(18+18) oligonucleotide (lane 2). In contrast to $TrwC_{R388}$(Y26F) and $TrwC_{R388}$, that produced a higher amount of the product when the *nic* site was double stranded (lane 5 and 7). Taken together, these findings indicate that the catalytic Y18 allows the accumulation of the covalent complexes when the substrate is double stranded at the *nic* site.

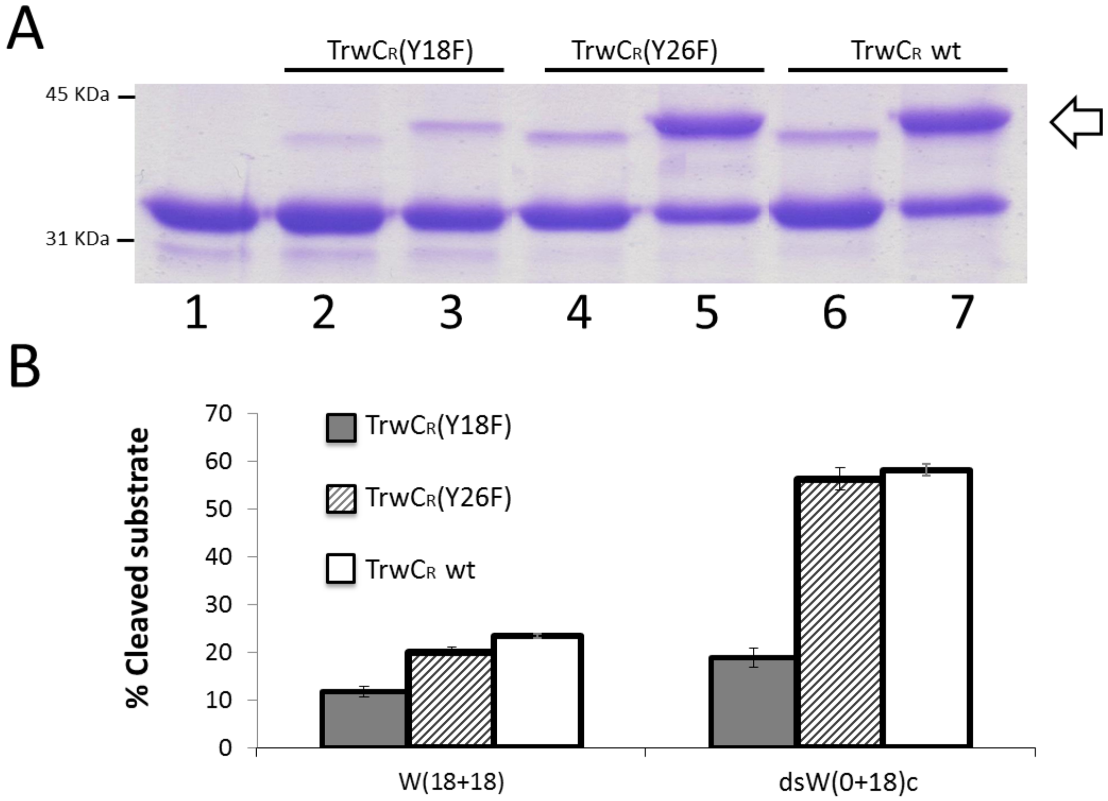


Figure R53. Cleavage efficiency of TrwC$_{R388}$ mutants to oligonucleotides with double strand at the *nic* site. (A) SDS-PAGE of oligonucleotide W(18+18) or dsW(0+18)c substrate, when incubated with TrwC$_{R388}$(Y18F), TrwC$_{R388}$(Y26F) or TrwC$_{R388}$ wt. 7 µM TrwC$_{R388}$ was incubated with 15 µM of different oligonucleotides. The reaction products were separated by electrophoresis in 12% SDS-PAGE gels. Lane 1, TrwC $_{R388}$(Y18F); Lane 2, TrwC$_{R388}$(Y18F) incubated with W(18+18); Lane 3, TrwC $_{R388}$(Y18F) incubated with dsW(0+18)c; Lane 5, TrwC$_{R388}$(Y26F) incubated with W(18+18); Lane 6, TrwC$_{R388}$(Y26F) incubated with dsW(0+18)c; Lane 6, TrwC$_{R388}$ incubated with W(18+18); Lane 7, TrwC$_{R388}$ incubated with dsW(0+18)c. (B) Bar graphs with the quantification of covalent complexes. Data showed mean±s.d. of three independent experiments.

Based in this rationale, we thought that we could also increase the covalent complexes formation of the well-known relaxase MobA$_{RSF1010.}$ The cleavage reaction of MobA$_{RSF1010}$ to wild type target showed a 35% of covalent complexes when the substrate contains the complete IR (WQ(30+7), Figure R54, lane 2) or at least contains the base of the stem (WQ(23+8), lane 3) (see Figure R41 for details). However, the yield did not improve when the targets were incubated with the complementary strand to the region upstream from the *nic* site (lanes 4 and 5 for WQ(30+7) and WQ(23+7) respectively).

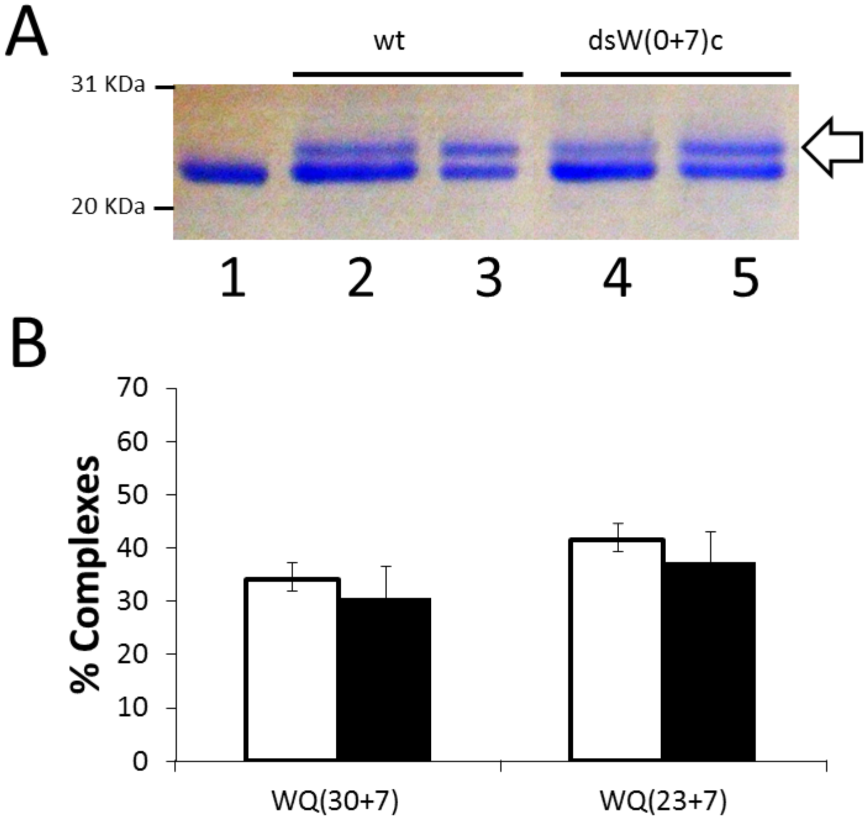


Figure R54. MobA$_{RSF1010}$ cleavage of substrate with double stranded DNA at the *nic* site. (A) SDS-PAGE of MobA$_{RSF1010}$ cleavage reactions. Lane 1, MobA$_{RSF1010}$. Lane 2, MobA$_{RSF1010}$ incubated with W30+7; Lane 3, MobA$_{RSF1010}$ incubated with W23+7; Lane 4, MobA$_{RSF1010}$ incubated with dsW0+7 (W(30+7) duplexed with oligonucleotide 2nd Half *nic* 0); Lane 5, MobA$_{RSF1010}$ incubated with dsW0+7 (W(23+7) duplexed with oligonucleotide 2nd Half *nic* 0). The arrow indicates the covalent complexes protein-DNA. (B) Bar graphs with the quantification of covalent complexes. White bars indicate the linear single stranded target. Black bars represent the double stranded substrate. Data showed mean±s.d. of three independent experiments.

7. Crystallography of TrwC$_{R388}$

7.1 Complexes of TrwC$_{R388}$ with a blunt-ended DNA substrate

It was observed by protein crystallography that TrwC$_{R388}$ is able to recognize a hairpin-forming oligonucleotide containing the *nic*-cleavage site (Guasch et al., 2003). This hairpin DNA W(25+0) mimics the ssDNA recognized by TrwC$_{R388}$ in the recipient cell to perform the final strand transfer reaction. Now, we have used a novel substrate W(7::16+0) that better mimics the DNA in the initial stage to reveal the details of how the relaxase TrwC$_{R388}$ starts the conjugation in the host cell. Two oligomers were used to generate the blunt-ended DNA substrate. The long one W(16+0) possesses the sequence of the IR$_2$ proximal arm and the U-turn, while the shorter one W(7::) contains the sequence of the IR$_2$ distal arm. Both oligonucleotides hybridize and generate a seven bp-duplex (Figure R55).

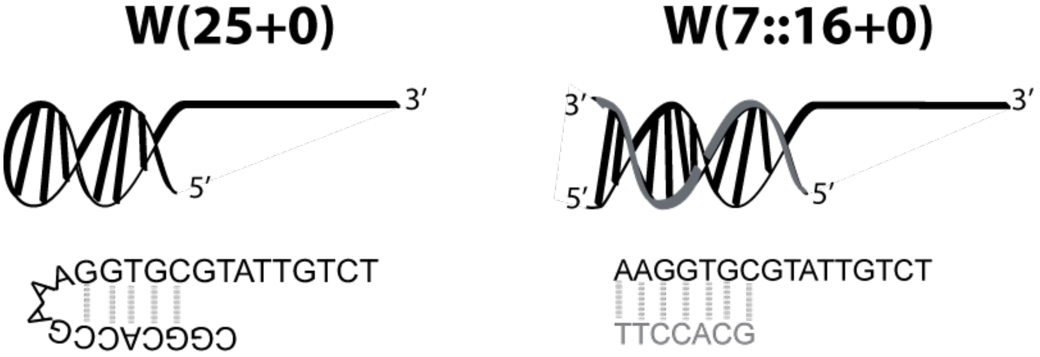


Figure R55. Schematic representation of the DNA sequences used to generate a complex with TrwC$_{R388}$. The *nic*-containing DNA substrate is depicted in black, the short complementary DNA W(7::) in gray. (W(7::16+0) lacks the loop region present in W(25+0).

The relaxase TrwC$_{R388}$ was used to form the complexes to be crystallized. Even though most of the TrwC$_{R388}$ structures deposited in the PDB were solved with the selenomethionine protein, our crystals have been obtained with the native relaxase. Crystals grew with the precipitant solution A (2 M Ammonium Phosphate and 0.1 M Tris-HCl pH 8.5) by the sitting drop method at 22ºC. The shape of the crystals of TrwC-W(7::16+0) were plates (Figure R56A) while TrwC-23+0 crystals grew as hexagonal prisms (Figure R56B).

X-ray crystal diffraction was carried out at ALBA synchrotron (Barcelona, Spain). XALOC beamline was used to determine the 2.5 Angstroms (Å) resolution crystal structure of TrwC-W(7::16+0) complex. The space group was C1$_{21}$ while the space group previously found in other TrwC$_{R388}$ crystals was P6$_5$ or P6$_1$22. The unit cell size and angles are shown in Table R4. Diffraction data was collected and analyzed as described in experimental procedures. The map obtained by molecular replacement was used to build a model of the relaxase TrwC$_{R388}$ with the blunt-ended substrate. However some residues were not ordered and their positions were not calculated.

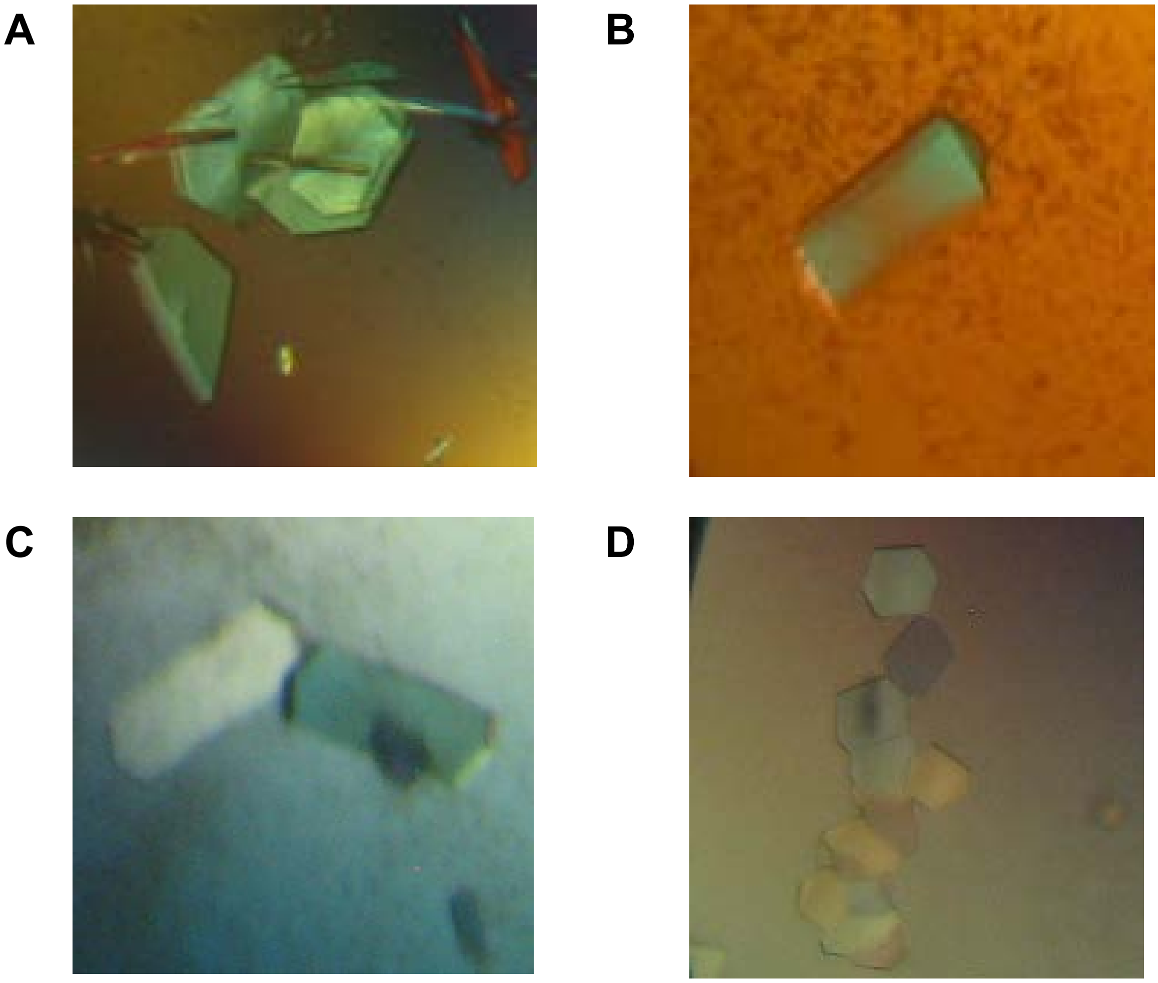


Figure R56. Crystals obtained with TrwC$_{R388}$. (A) The TrwC$_{R388}$-W(7::16+0) complex in solution A. (B) TrwC$_{R388}$-23+0 complex in solution A . (C) TrwC$_{R388}$–(W7::16s+2) in solution B. (D) TrwC$_{R388}$(Y18F)–R388_(W7::16s+2) in solution B.

The high-resolution structure reveals an unreported antiparallel TrwC$_{R388}$-DNA dimer in the asymmetric unit. The DNA in the crystal acquires a disposition similar to a Holliday junction, but lacking one of the strands where the double stranded stems are aligned along one of the axis, while the *nic* sites are disposed perpendicular to this axis (Figure R57). The structure of each of the two components of the dimer is identical to the deposited structures (Guasch et al., 2003; Boer et al., 2006). Interestingly, the two TrwC$_{R388}$ monomers are interacting by the back of the palms (α-helices α-4, α-5, α-6, α-7), whereas the fingers domains are pointing up and down respectively (Figure R57). Some intermolecular contacts with the DNA bound in the opposite monomer are also observed. The single stranded DNA from G17 to T20 runs parallel to the same sequence in the other monomer. The Thymine 20 is pointing out towards the Thymine 20 of the other monomer. Bases T21 to C24 shapes a U-turn within a pocket of the protein. The T25 is positioned within the active centre of the relaxase.

Table R4. Experimental data of the X-Ray diffraction of complex TrwC_{R388}-W(7::16+0).

Resolution (Å):	2.50 (2.5-67.48)
Space Group Name:	$C_{1\ 2\ 1}$
Wavelength (Å):	0.9795
Measured Reflections	25691
Resolution Range for refinement	2.5-67.48
Crystallographic R-factor:	0.246
R-Free:	0.298
Number of atoms	
Protein	15375
DNA	1293
Solvent	6
Heteroatoms	2
Completeness (%)	98.81
B-value	41,03
r.m.s. deviations from target values	
Bonds (Å):	0,011
Angles (°):	1,669
Unit Cell:	
Length [Å]	Angles [°]
a = 147.29	α = 90.00
b = 67.33	β = 113.61
c = 87.47	γ = 90.00

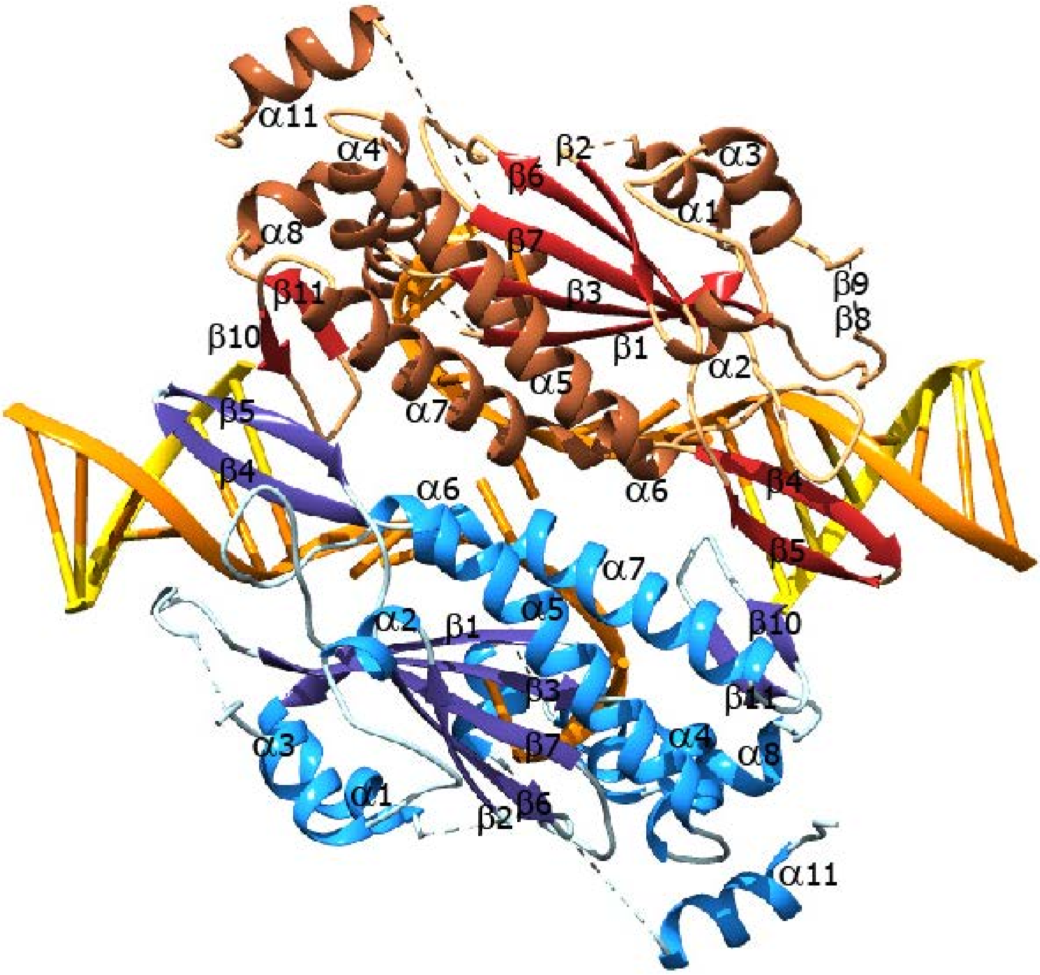


Figure R57. Ribbon structure of the dimer TrwC-W(7::16+0). The upper monomer is shown in reddish colors, and the monomer bellow in bluish colors. The *nic*-containing DNA strand is shown in orange, and in yellow the IR_2 distal strand. α-helices (α-1 to α-11), β-sheets (β-1 to β-11) and turns are shown in different tones. Dashes lines that connect the secondary structures mean that these residues were no modeled.

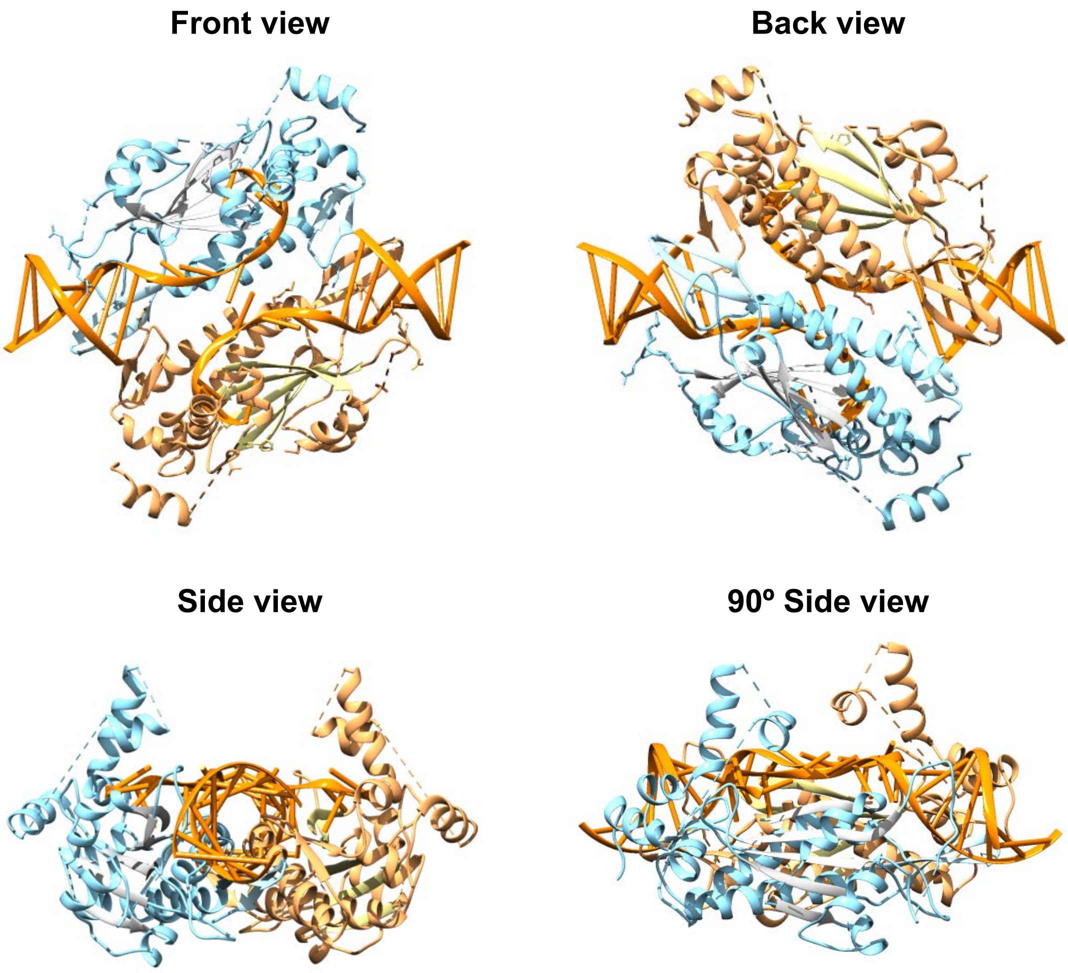


Figure R58. Structure of the relaxase TrwC$_{R388}$ that dimerizes with the substrate mimicking the origin of transfer. Ribbon diagram of the dimeric structure of TrwC-W(7::16+0). The DNA is colored in orange.

The β-turn between β-10 and β-11 connects with the base of the inverted repeat of the opposite monomer. The lysine K215 in TrwC$_{R388}$ chain A establishes contacts with the guanine 1 of chain B right at the base of duplex DNA of the opposite monomer. This lysine K215 is conserved in other Y2 relaxases (Figure R61). Moreover, structural alignment with the program Pymol of the structure of MOB$_F$ relaxases deposited in the PDB revealed that the conserved lysines are pointing out at the same direction (Figure R59). The lysines K220 of TraI$_F$ (PDB 2A0I), and K215 of TraI$_{pCU1}$ (PDB 3L57) are in a β-turn between two small beta sheets β10-β11. The apo protein TraI$_F$ (PDB 1P4D) has also the lysine K220 localized at this β-turn, but lacks some of the α-helix that shapes the fingers domain (data not shown). These data suggest that the lysine K215 stabilizes the melted DNA strands for efficient cleavage and the state captured in the crystal structure snap shot may accurately represent a relaxase dimer during the initiation of conjugation. Interestingly, the main feature of this β-turn is that it shapes the binding cleft for ssDNA. In $\mathrm{TrwC_{R388}}$, residue N218 within this β-turn establishes H-bonds with T21, promoting the bending of the DNA to shape the characteristic U-turn.

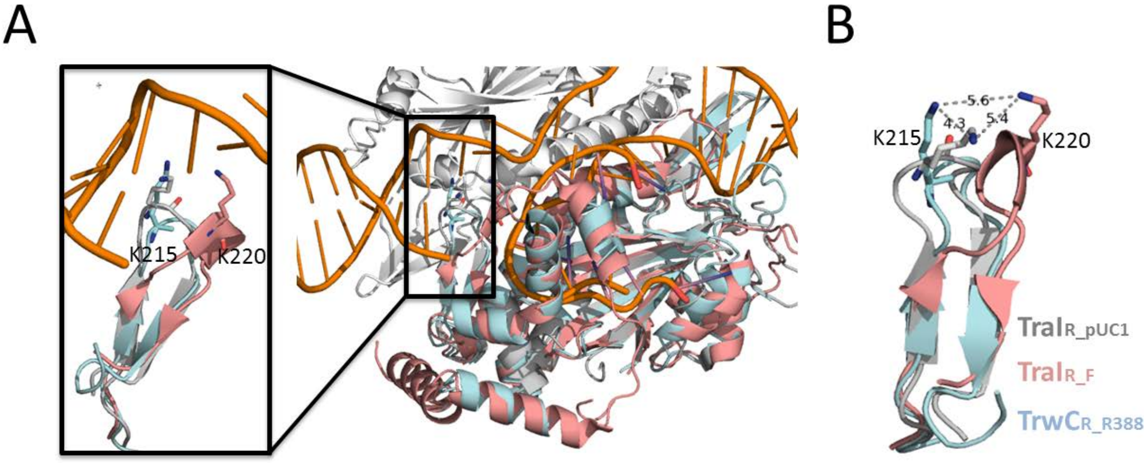


Figure R59. Ribbon representation of the conserved lysines K215 (shown in sticks) at the β-turn at the fingers domain of relaxases from $\mathrm{MOB_F}$ family. A) Localization of the β10–β11 turn in the overall alignment of three relaxase structures. The inset shows the conserved lysine as sticks interacting with the duplex DNA of the juxtaposed monomer in $\mathrm{TrwC_{R388}}$. The lysines seem to direct the Nε to the base of the dsDNA (colored in orange). The distances between the Nε of the three conserved lysines is shown in (B). $\mathrm{TraI_F}$ (depicted in salmon), $\mathrm{TraI_{pCU1}}$ (gray color) and the novel structure of $\mathrm{TrwC_{R388}}$-W(7:16+0) (colored in light blue, the opposite monomer is shown in white).

Residues at the α-6 and α-7 helices also establish contacts with some residues of the other monomer (Figure R60). Q205A has an H-bond with the R136B at β5 and vice versa (Figure R60B, left panel). Moreover, it seems that tyrosine Y191 of one $\mathrm{TrwC_{R388}}$ monomer interacts with the corresponding Y191 of the other $\mathrm{TrwC_{R388}}$ monomer by stacking their aromatic rings, and establishes two hydrogen bonds with the arginine R124 at β4 (Figure R60B, central panel). Two residues from the β5 sheet, K134 and I137, established hydrogen bonds with residues, L211 and Y123, of the β10 sheet of the other monomer (Figure R60B, right panel). They could play an important role in stabilization of the dimer.

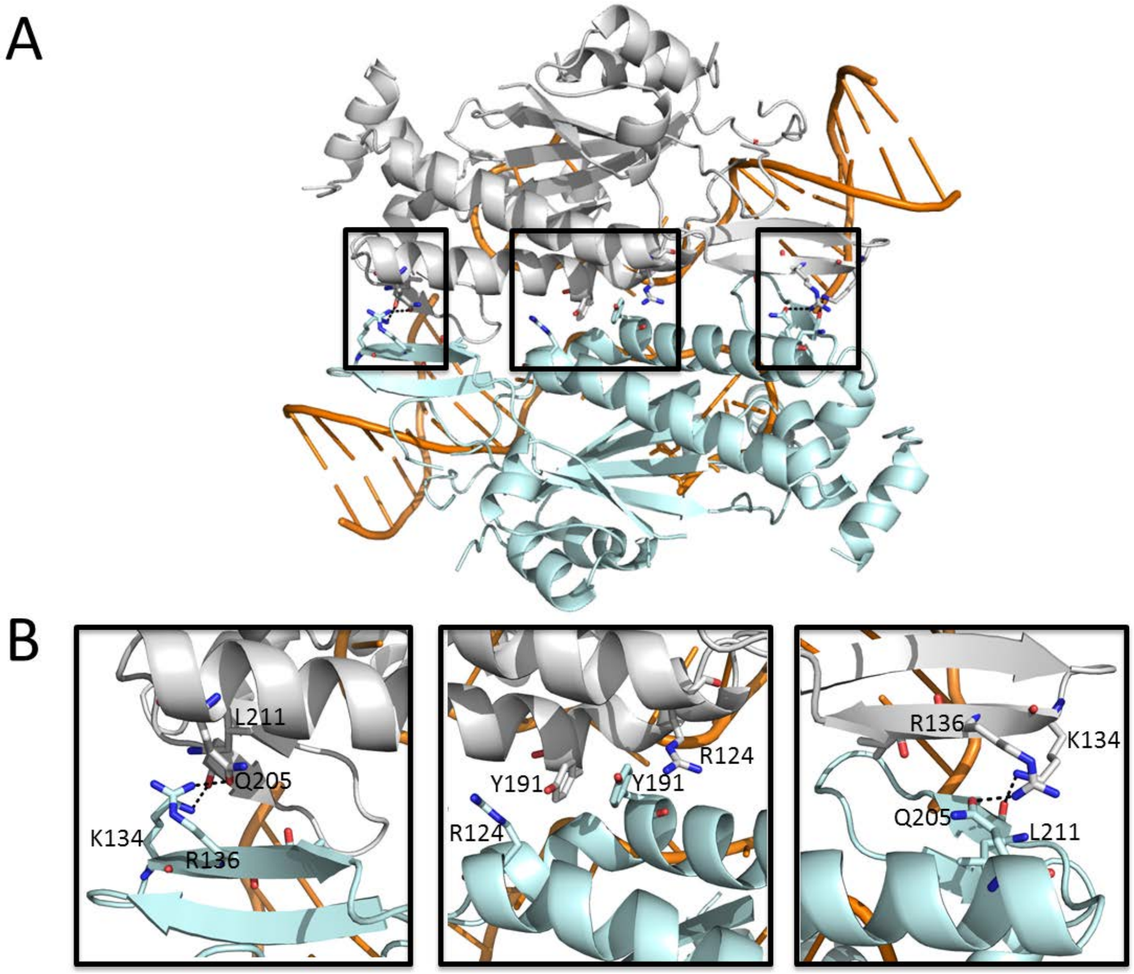


Figure R60. Residues from α-6 and α-7 helices and β-5 and β-10 sheets build the dimer interface. A) Cartoon representation of the dimer. B) Insets of the residues involved in the polar contacts. Dashed lines are the hydrogen bonds.

However, an analysis of the residues involved in $\text{TrwC}_{\text{R388}}$ dimerization across the MOB_{F1} family shows that these residues are not conserved (Figure R61).

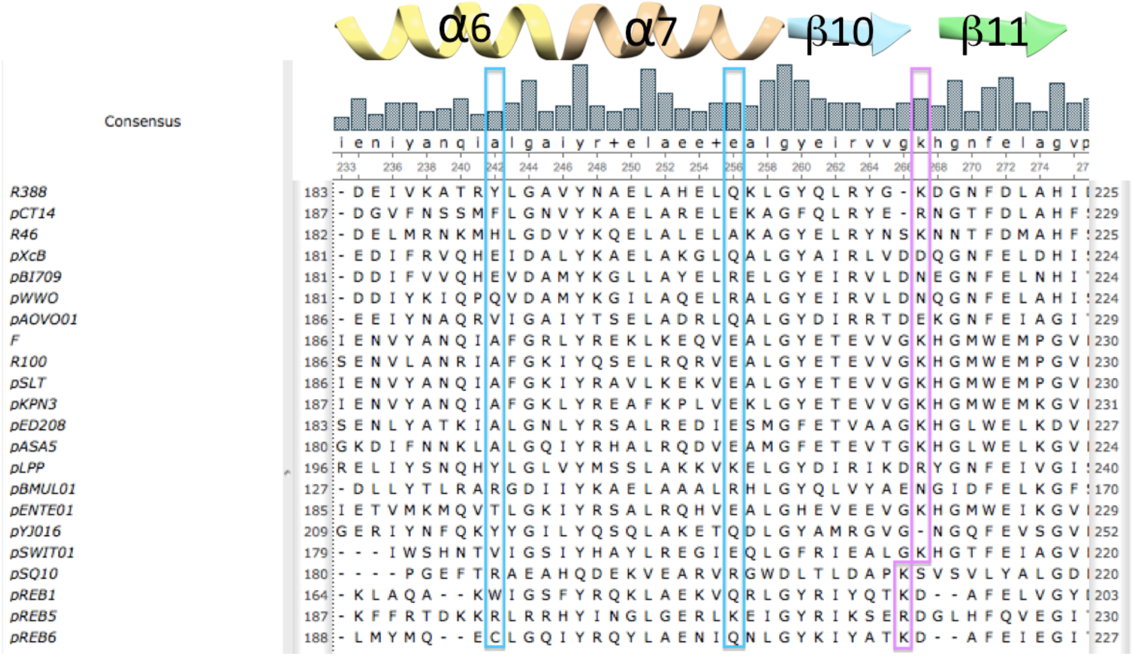


Figure R61. Alignment of the MOB$_{F1}$ family of relaxases. The region shows the secondary structures that seem to play a role in dimerization. In blue rectangles are shown the residues Y191 at α6 and Q205 at α7. The conservation of K215 is highlighted with a pink rectangle. The height of the gray bars displays an overall low consensus score. Alignment was generated using ClustalW.

7.2 Mutational studies of residues involved in the dimer formation

To date, TrwC$_{R388}$-W(7::16+0) is the first structure that mimics the substrate at the initiation of the conjugation at the donor cell. The dimer arrangement is perfectly compatible with the partial melting of the DNA and the stabilization of the ssDNA target within the active center of the other monomer. In order to study the role of these residues at the interface, point mutation TrwC proteins encoded in a non-mobilizable plasmid pET3a were electroporated to *E.coli* DH5α containing a R388 derivative plasmid lacking *trwC* (pSU1445). If some of these point mutations result in a significant loss of the conjugation rate of pSU1445 plasmid, it can be assumed that dimerization is a key step during initiation of the conjugation. Thus, the capability of point mutations of full length TrwC to transfer the R388 plasmid was tested (Table R5).

Table R5. Conjugation frequency of TrwC mutants.

Donor cell	Transfer Frequency[a]
TrwC	$3{,}7*10^{-1} \pm 1{,}5*10^{-1}$
TrwC(Y191K)	$8{,}1*10^{-1} \pm 0.6*10^{-1}$
TrwC(Q205Y)	$5{,}4*10^{-1} \pm 2{,}8*10^{-1}$
TrwC(Y191K/Q205K)	$2{,}0*10^{-1} \pm 1{,}4*10^{-1}$
TrwC(K215A)	$4{,}3*10^{-1} \pm 3{,}6*10^{-1}$
pET3a	-

[a] *E.coli* containing the TrwC mutants in a pET3a plasmid (Ap^R) and the plasmid pSU1445 (Kn^R) as donors. pSU1445 has a Tn5 insertion in the TrwC gene that converts it in non-self transmissible. Complementation with multicopy plasmid pET3a that encoded the full-length relaxase TrwC (pSU1621) made this plasmid mobilizable at the same rate that wild type R388. TrwC mutants were assayed for its mobilization capacity to the recipient cell UB1637 (Sm^R). Upon conjugation the plasmid pSU1445 is transferred to the UB1637 recipient cell. Mating was performed as described in experimental procedures.

TrwC mutants carried out DNA processing and transfer with the same efficiency than wild type TrwC. Thus, it seems that the selected residues are involved in crystal packing, but do not have an evident role in TrwC biological function. In fact, the $TrwC_{R388}$-W(7::16+0) interface was further analyzed with the web-server PDBePISA (www.ebi.pisa), bioinformatic tool for the characterization of the interaction between macromolecules. PDBePISA results showed that there are nine hydrogen bonds at the interface (Table supplementary S17), being 1066.1 A^2 the buried surface in the dimer. This surface is associated with moderate binding affinities. The interface was scored with complexation significance score (CSS) of 0.00, which means that the interface is not predicted to play any role in complex formation. Hence, the interactions on the dimer can be a result of crystal packing, instead of having any biological role *in vivo*.

7.3 Protein crystallography studies of TrwC$_{\text{R388}}$ covalently bound to DNA

Although the structures of several relaxase-DNA complexes have been solved, none of them showed the result of the cleavage reaction produced by the catalytic tyrosine. Thus, with the goal to crystallize the relaxase attached to the cleaved DNA, phosphorothiolate or "suicide" oligonucleotides with the blunt-ended DNA substrate were used to form complexes with the protein as described in experimental procedures. An overnight incubation of TrwC$_{\text{R388}}$ with the W(16s+2) substrate in presence of magnesium leads to the formation of 100% of covalent complexes (Figure R62).

A

AAGGTGCGTATTGTCT s^AT

B

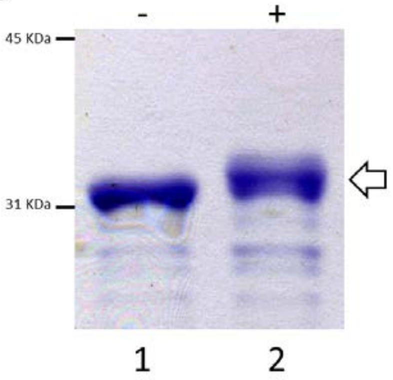


Figure R62. Formation of DNA–TrwC$_{\text{R388}}$ covalent complexes. (A) Oligonucleotide sequence of the phosphorothiolate substrate W(16s+2). The position of the phosphorothiolate modification is shown by a lower case blue s letter within the sequence. The *nic* site is represented by symbol ^. The inverted repeat sequence is underlined. See Figure 32 for details. (B) SDS–PAGE shows the cleavage of the suicide substrate by TrwC$_{\text{R388}}$. Lane 1, 7 µM TrwC $_{\text{R388}}$ with no oligonucleotide added. 7 µM TrwC $_{\text{R388}}$ incubated with 15 µM oligonucleotide W(16s+2) (lane 2) for 24 h. The position of the TrwC$_{\text{R388}}$-AT covalent complex is shown by a white arrow.

Previously to crystallization the distal arm substrate W(7::) was added to stabilize the complex. Crystals of the complexes TrwC$_{\text{R388}}$-W(7::16s+2) grew in solution B (20% (w/v) polyethylene glycol 8000, 0.1 M phosphate/citrate pH 4.2, 0.2 M Sodium Chloride). Plates were obtained with TrwC$_{\text{R388}}$–W(7::16s+2) complex, and also when the mutant TrwC$_{\text{R388}}$(Y26F) was used in the complex (see Figure R56C and D). Smaller crystals with the selenomethionine substituted protein TrwC$_{R388}$ grew also in solution B (data not shown). However, no crystal formation was observed when these conditions were used for the complex TrwC$_{R388}$ with phosphorothiolate oligonucleotide R388-W(25s+2) or wt W(23+0).

The diffraction of the protein crystals was done with the X-ray source XALOC in Barcelona. Data was processed with the CCP4 suite of crystallography programs. Crystals diffracted to 2.5 Å resolution and adopted space group C_{121} (Table R6). The structure was determined by molecular replacement using TrwC-W(25+0) complex (PDB 1OMH) as the search model. The asymmetric unit contains two proteins, as in the crystal obtained with the standard oligonucleotides (Figure R58).

Even though, TrwC$_{R388}$(Y26F) and selenomethionine substituted TrwC$_{R388}$ –W(7::16s+2) complexes have similar unit-cell parameters, TrwC$_{R388}$–W(7::16s+2) complex resulted in a different unit cell. Changes in the crystal unit cell suggest conformational changes produced by the nucleophilic attack of the catalytic tyrosines to DNA. The crystal structure of TrwC$_{R388}$(Y26F) – W(7::16s+2) and selenomethionine resulted in the same 3D structure as that of the wild-type TrwC$_{R388}$-W(7::16). Least-squares Cα superposition of the three structures reveals this similarity and yields an overall root-mean-square deviation (rmsd) value of 0.39 to 0.63 Å for 423 atom pairs (Table R6). It is then logical to assume that the backbone conformation of TrwC$_{R388}$ does not change extensively upon DNA cleavage.

Table R6. Experimental data of the X-Ray diffraction of wt TrwC$_{R388}$, selenomethionine variant and TrwC$_{R388}$(Y26F) complexed with W(7::16s+2).

	TrwC$_{R388}$	SeMet TrwC$_{R388}$	TrwC$_{R388}$(Y26F)
Resolution (Å):	2.5	2.5	2.5
Space Group Name:	C 1 2 1	C 1 2 1	C 1 2 1
Wavelength (Å):	0.9795	0.9793	0.9795
Measured Reflections:	27286	27801	27911
Resolution Range for refinement:	2.5-79.91	2.5-43.467	2.5-80.76
Reflections:	25509	27541	27666
Crystallographic R-factor:	0.261	0.222	0.320
R-Free:	0.258	0.269	0.395
Number of atoms:	5090	5359	3977
Protein:	3980	4385	3023
DNA:	940	961	943
Solvent:	5	8	5
Heteroatoms:	19	10	11
Completeness (%):	98.42	99.08	99.15
B-value:	46.95	44.89	41.23
r.m.s. deviations from target values:			
Bonds (Å):	0.013	0.011	0.010
Angles (º):	1.714	1.377	1.305
Unit Cell:			
Length [Å]:			
a =	147.22	147.26	147.67
b =	67.38	67.28	67.98
c =	87.26	89.00	80.75
Angles [°]:			
α =	90	90	90
β =	113.69	113.81	90.17
γ =	90	90	90

The loop residues containing the catalytic tyrosine Y26 were partially disordered in all three structures and were partially visible in the density map; residues 12-32 (wt), 20-29 (SeMet) and residues 20-32 (Y26F). The refined structure of TrwC$_{R388}$–W(7::16s+2) complex did not include the C-terminal 250-293 residues of the fingers domain because they were most likely disordered. Our efforts to complete the model of the tyrosines-loop and the fingers domain by placing Cα atoms within the density map manually were unsuccessful. In addition, we were not able to model neither the tyrosines covalently bound to the DNA nor the two nucleotides that remain covalently bound to the catalytic tyrosine in all the structures.

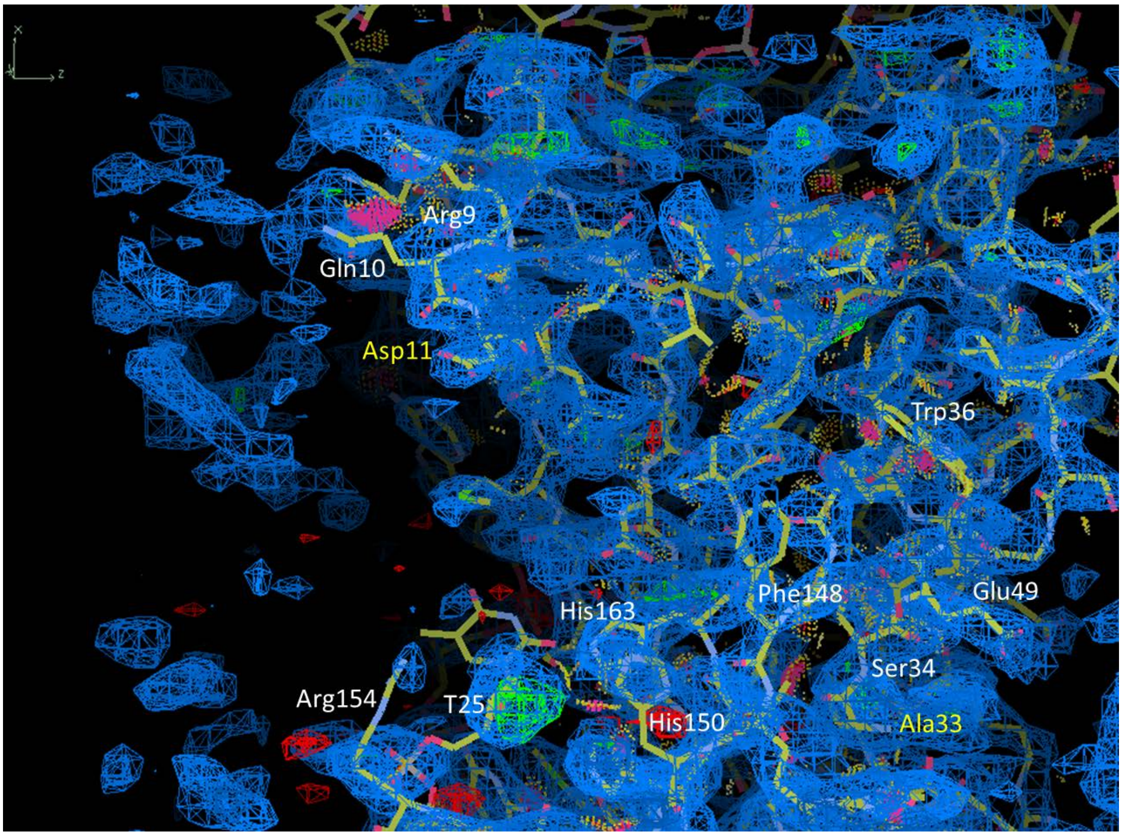


Figure R64. The atomic model based on the experimental electron density from TrwC$_{R388}$-W(7::16s+2). The mesh in blue surrounds the regions with atoms modeled in the structure. The mesh in red surrounds regions with high densities of electrons, which do not fit with the atoms in the model. Red regions depict regions in which the model is present where there is no data. The residues between the labelled residues in yellow (Asp11 and Ala33) are missing. We could not connect regions of the electron-density map (referred to as blobs, in blue, at the left) to build the model of the loop containing the catalytic tyrosines.

With the goal of optimization of the structures, we constructed a model by superposition using the native structure of TrwC as template. When the complete model was constructed (residues 1-293), a single round of atom placement optimization was done manually to ensure that all atoms fit within the density. Then, the complete model of the structure was refined using *phenix.refine*. Finally, the B-factor or temperature factor was used to describe coordinate errors. Figure R65 shows 3D graphs of the atomic B-factors from the proteins resolved. As is clear from Figure R65, the magnitude of B-factors is greater for the TrwC$_{R388}$–W(7::16s+2) than the other structures. The red colors in the loop of the catalytic tyrosines (residues 18-30) and the fingers domain (residues 230-293) indicate that these regions were not accurately modeled. Then, the difficulties in modeling the protein-DNA covalent complexes are therefore probably caused by its high conformational flexibility.

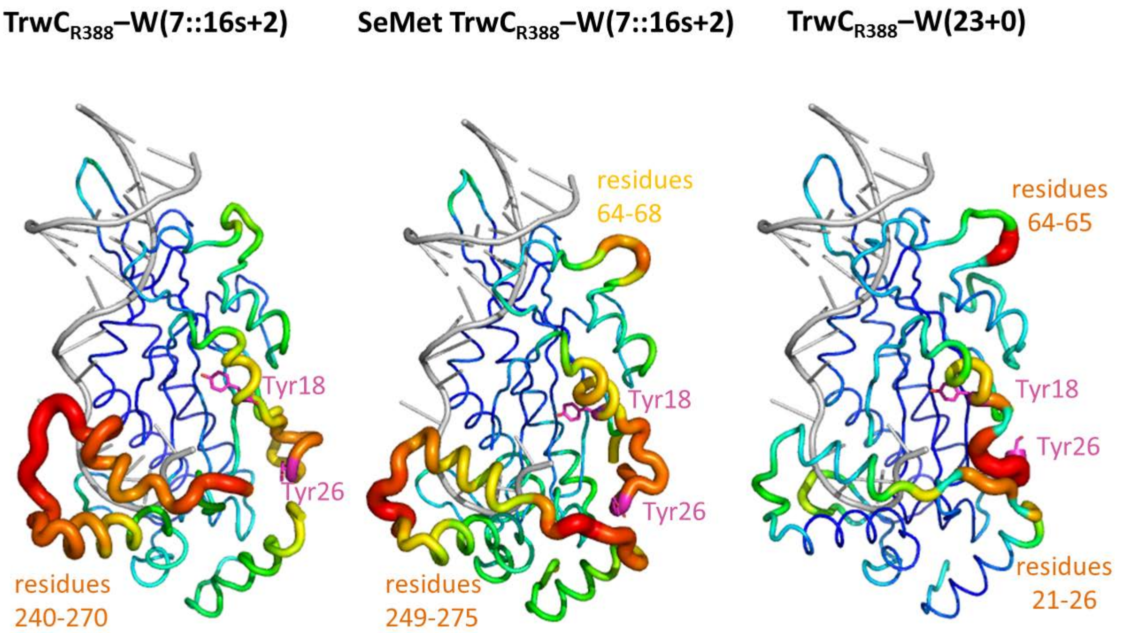

TrwC$_{R388}$–W(7::16s+2) **SeMet TrwC$_{R388}$–W(7::16s+2)** **TrwC$_{R388}$–W(23+0)**

Figure R65. B-factor presentations of the three crystal structures of TrwC$_{R388}$ complexed with DNA. The B-factor indicates the dynamic mobilities of different resolved parts within the structure. The thicker the lines and the warmer the color, the higher is the disorder. The tyrosines-loop and the helices of the fingers domain have the highest B-factors in all three structures. Maximum B-value was 133.2 Å^2, 122.5 Å^2 and 58.9 Å^2, respectively. The DNA is represented in gray.

Interestingly, only small rearrangements in $\text{TrwC}_{\text{R388}}$ were needed to accommodate the cleaved DNA. Our results imply that the catalytic tyrosine Y18 when covalently bound to the DNA moves away from the active center. This rearrangement could allow the movement of the catalytic tyrosine Y26 to attack the *nic* site. We will continue working in the improvement of the structure of the covalent complex $\text{TrwC}_{\text{R388}}$-DNA.

DISCUSSION

8. Substrates designed to enhance the covalent binding of the relaxase.

Relaxases are very well tuned proteins that have been improved through natural evolution to process plasmid DNA for bacterial conjugation in a very precise manner. However, this specialization is a problem when using these enzymes for *in vitro* biochemical applications. In order to widen the range of substrates effectively processed by relaxases, we have used different approaches, which results are discussed in this section.

8.1 Rep-like substrates

It is known that the cleavage efficiently of replicases is higher than that of relaxases because the trigger of bacterial conjugation is more complex than the initiation of RCR itself. Thus, to improve relaxase activity we have designed Rep-like substrates for relaxases with secondary structures mimicking the stem-loop recognized and cleaved by rolling circle replicases (Figure R35B). We observed that $TrwC_{R388}$ cleaves Rep-like *nic*-sites with a yield similar to wt substrates. Although the strand-transfer reaction also proved to be efficient, $TrwC_{R388}$ produced different complexes when acting on Rep-like oligonucleotides with shorter loops (S region between 2 and 11 nucleotides) according to EMSA experiments (Figure R37 and R38). We think $TrwC_{R388}$ is melting the stem loop of the Rep-like substrates in order to position the *nic* site close to the catalytic tyrosine, which facilitates the generation of intermolecular complexes. Indeed, we prevented product polydispersion by designing substrates with a longer IR (15 bp in spite of either 6 or 8 bp) (Figure R39). Thus, when the hairpin is longer, $TrwC_{R388}$ does not melt the DNA to cleave the *nic* site. However, SDS-PAGE gels showed that the formation of $TrwC_{R388}$ covalent complexes is slightly lower with H(23+23), H(23+26), H(24+24) and H(24+27) substrates (Figure supplementary S22). This could mean that $TrwC_{R388}$ either religates easily or cleaves less efficiently substrates with larger IRs.

Similar Rep-like substrates were designed and assayed for Y1 relaxases MobA$_{RSF1010}$ and TraI$_{RP4}$. MobA$_{Rsf1010}$ efficiently cleaved the corresponding Rep-like substrates (Figure R42) but TraI$_{RP4}$ did not cleave them (Figure R43). Relaxase activity with Rep-like substrates was even lower than with their cognate substrates. Relaxases establish unique contacts to the U-shaped turn and *nic*-region, including an interaction with the IR, that facilitate religation of the 3' terminal end of the *nic* to the phosphotyrosine intermediate (Guasch et al., 2003; Larkin et al., 2005; Edwards et al., 2013). Moreover, in Rep-substrates the DNA fragment covalently attached to the relaxase forms a duplex with region that is non-covalently bound to the relaxase favoring the religation reaction. It seems that Rep proteins cleave more efficiently oligonucleotides mimicking their cognate substrates *in vitro* because Rep-binding sequence does not include the *nic* region, limiting the religation of the cleaved strand (Campos-Olivas et al., 2002; Khan, 2005; Vega-Rocha et al., 2007; Boer et al., 2009).

8.2 Reverse substrates

Y2 Relaxases are known to have a fingers domain, absent in replicases, that tightly bind the U-turn and helps in the religation process (Figure R35). To avoid this situation we designed novel reverse substrates where the 3´side of *nic* contains now the entire IR (Figure R44). In this substrate contacts of the 5´side of *nic* with the relaxase are reduced and the religation reaction is diminished. Even though the balance was displaced to DNA-protein complex formation, 100% yield could not be achieved, probably because the relaxase has residual affinity for the U-turn 5´side of the cleaved reverse oligonucleotide. TraI$_{RP4}$ also cleaved the reverse substrate RP(8+24) although, unexpectedly, the yield of covalent complex formation with the novel substrates did not improve (Figure R48). On the other hand, MobA$_{RSF1010}$ did not cleave reverse-oligonucleotides. This protein lacks the fingers domain at the C-terminal end that shapes a cleft that stabilizes the U-turn in the active center in MOB$_F$ relaxases (Figure R48). Probably this feature allows MobA$_{RSF1010}$ to form 40% covalent complexes with wt oligonucleotides WQ(30+7) and WQ(23+7) (Figure R42) because religation is limited. We hypothesize that this feature could restrict the cleavage of reverse substrates. Overall, the cleavage-ligation reaction is tightly in balance for each relaxase, and not all relaxases can improve the yield of covalent complexes.

8.3 dsDNA-*nic* substrates

TrwC$_{R388}$ needs partially melted DNA to cleave the *nic* site, but the influence of double strand downstream of the *nic* site on the cleavage-ligation equilibrium has not been studied before. To directly address the question of whether the presence of double stranded DNA at the *nic* site have any influence in the relaxase reaction, several oligonucleotides that hybridizes with the target W(18+18) were tested. Noticeably, the covalent product boosted when the region upstream of *nic* site was double stranded (dsW(0+18)c) (Figure R51). These results revealed that the covalent complex formation equilibrium was significantly displaced to the product. This increase in covalent complexes can either result from increased rates of DNA cleavage by TrwC$_{R388}$ or decreased rates of religation. It seems that the dsW(0+18)c substrate favors Y18 cleavage but not Y26 cleavage (Figure R53). Some rearrangements on the protein-DNA complex with substrates with double stranded DNA right at the *nic* site could facilitate the nucleophilic attack of Y18. Other explanation is a reduced religation reaction performed by Y18 when it is bound to dsW(0+18)c leads to an accumulation of the covalent complexes.

Interestingly, the covalent complex quantification revealed that the oligonucleotide "2nd half *nic* 0" does not alter the equilibrium in MobA$_{RSF1010}$ (Figure R54). The cleavage-religation balance reaches approximately 50% of covalent complexes for this Y1 relaxase. These results suggest that the religation of the cleaved DNA by the catalytic tyrosine is not altered when there is a duplex at the boundary of the *nic* site. The absence of the fingers domain in Y1 relaxases could allow the presence of dsDNA in the *nic* site without affecting the religation reaction.

8.4 Novel *nic* substrates are not functional *in vivo*

Once demonstrated that *in vitro* reactions were enhanced by novel relaxase substrates, we analyzed their behavior in whole cell. With the goal of improving the efficiency of relaxases within living cells, synthetic plasmids with the relaxase target converted to either Rep-like sequence H(14+14) or Reverse R(8+24) were tested. Conjugation rates critically fell for all synthetic plasmids tested (Figure R49). Rep proteins can cleave dsDNA plasmids in the absence of RAPs because the *nic* site is at the loop of a cruciform IR. However, relaxases need RAPs to locally alter DNA supercoiling in a way the relaxase can melt the DNA to display the *nic*-containg ssDNA (Moncalián et al., 1999; Yoshida et al., 2008; Varsaki et al., 2009). In the conjugation assay of Rep-like and reverse *nic* containing plasmids, we changed the relative orientation of *nic* with respect to TrwA (*sbaB*) and IHF (*ibs*) binding sites. Moreover, ssDNA hairpin formation by the modified IR region is different in recipient cells and could affect the termination reaction. Further studies will focus on the directed evolution of these synthetic *ori*Ts to obtain transferable plasmids in the absence of RAPs.

Previous work on TrwC found alternative functional *nic*-site sequences by partially randomized mutagenesis. Carballeira et al. identified the crucial nucleotides involved in target selection. Here, we show that these targets could be expanded by the use of permuted sites. With these novel substrates, the *nic*-sequence conformation resembles its 3D structural shape when located in the structural context of the *oriT*. Our finding broadens the number of potential targets for HUH relaxases, thus increasing the possibility of using relaxases for biotechnological applications. Future work on protein engineering of relaxases could improve recognition and processing of alternative DNA sequences.

9. The structure of $\text{TrwC}_{\text{R388}}$ bound to DNA

Coupling of conformational changes in a general tyrosine-containing loop to covalent catalysis of phosphoryl transfer is one of several mechanistic features shared by the transposases, topoisomerases, recombinases and restriction endonucleases (Krogh and Shuman, 2000). For example, the active sites of transposases adopt two functionally important conformations: the *trans* configuration, in which each active site is composed of the HUH motif from one monomer and the Tyr residue on an α-helix (αD) from the other monomer, and the *cis* configuration, in which both motifs in an active site are contributed by the same monomer (Figure 35). In the *trans* configuration, the transposase binds and cleaves the insertion sequence. In *cis* configuration, the αD-helices catalyze the formation of a circular single-stranded transposon and join the cleaved ends of the donor sequence (Barabas et al., 2008). Most of these enzymes acts as dimers, however, the role of dimerization of relaxases has not been proved so far. Even though the presence of only one active tyrosine in MOB_Q and other Y1 relaxases indicates that dimerization is required to recircularize the plasmid in the recipient cell (Carballeira et al., 2014).

Interestingly, only small rearrangements in $TrwC_{R388}$ were needed to accommodate the cleaved DNA. Changes in the crystal unit cell suggest conformational changes produced by the nucleophilic attack of the catalytic tyrosines to DNA. Our results imply that the catalytic tyrosine Y18 when covalently bound to the DNA moves away from the active center. This rearrangement could allow the attack of the catalytic tyrosine Y26 to the *nic* site. The solved structure shows also conformational flexibility in the fingers domain when covalently bound to DNA. This result could mean that small conformational changes are needed to rearrange the cleaved DNA end in Y2 relaxases. Now, we are working with other substrates that can help us in the localization of the cleaved DNA and its modeling of Y1 and Y2 relaxases. The resolution of these structures will help us to determine the rearrangements of the protein when it performs the catalytic attack to DNA and it could help us to find inhibitors for the process that will reduce the antibiotic-resistance spread.

GENERAL DISCUSSION

In this work we have showed different approaches to improve the binding yield of relaxases to DNA *in vitro* with the goal of using relaxases as a novel tool for site-specific modification of DNA origami nanostructures. Our results demonstrated that relaxases possess highly selective binding abilities to their wt targets on DNA nanostructures with different architectures. However, the application of these proteins in bionanotechnology requires an improvement of the percentage of relaxase-DNA covalent complexes. To this end, we designed substrates with secondary structures mimicking the stem-loop recognized and cleaved by RC replicases, substrates with the stem-loop in a reverse orientation and also substrates with dsDNA around the *nic* site. Although the covalent complex formation was improved, it should also be achievable to further enhance the reaction by modification of the relaxase itself.

Being TrwC a potential tool for human genome editing (Agundez et al., 2010), the finding of new cleavable substrates by rational design described in this thesis could have an interesting set of applications related with biomedicine. By widening the number of sequences that TrwC and other relaxases can recognize, the potential targets for genome integration and targeted modification of the human genome could be greatly increased (González-Prieto et al., 2013).

This thesis also showed that relaxases can be easily fused to other proteins of interest and become a whole new class of orthogonal, sequence-specific protein linkers for DNA nanotechnology. One of the most interesting possibilities for relaxase modifications would be their application in nanomedicine. For instance, nanostructures could encapsulate relaxases fused to several proteins that, when open, modify the fate of target cells, such as the nanorobot developed by (Douglas et al., 2012). Other applications could be the *in situ* attachment to DNA nanostructures generated within cells. A practical application is related with the generation of nanobiofactories described by (Simmel, 2012), in which relaxases allows the programmable and accuracy attachment of proteins that participates in a biochemical pathway in close proximity. These nanoarrays could be used for the generation of artificial molecular assemblies, for the modification of drugs or for biofuel synthesis.

Thus, this thesis has settled the basis for the biotechnological use of relaxases in state of the art applications, such as bionanotechnology or gene editing.

CONCLUSIONS

1. Recombinant fused relaxases can efficiently bind planar nanostructures using two targets in close proximity with 5 bp rigid linkers whereas they efficiently couple to nanotubes with three to five binding sites per position. Short incubation times, such as 1 hour, results in an efficient decoration of DNA nanostructures.
2. Different relaxases are efficiently organized on rectangular sheets and tubular nanostructures. The relaxase $\text{TrwC}_{\text{R388}}$ is assembled on rectangular nanostructures with position preference for central positions. The specific binding ability and high affinity of $\text{TrwC}_{\text{R388}}$, $\text{TraI}_{\text{R100}}$, $\text{MobA}_{\text{R1162}}$ results on an orthogonal decoration of DNA nanostructures. However, $\text{TraI}_{\text{pKM101}}$ shows significant non-specific binding to other relaxase targets.
3. The binding yield to DNA nanostructures was not improved using phosphorothiolate “suicide” target-staples.
4. The relaxase $\text{TrwC}_{\text{R388}}$ directs the assembly of DNA ssDNA, dsDNA and origami tiles creating nanoribbons. The relaxase keeps bound to the nanostructures when it has performed the DNA polymerization.
5. Relaxases $\text{TrwC}_{\text{R388}}$ and $\text{MobA}_{\text{RSF1010}}$ are able to cleave substrates with replicase *nic* site layout but TraI_{RP4} did not cleave Rep-like oligonucleotides. However, *oriT* layouts with either the Rep-like or Reverse design at the *nic*-region are not functional *in vivo*.
6. The percentage of covalent complexes improved by changing the layout of the cognate substrate. Reverse layout and double strand at the *nic* site, leads to an overall of 50% of $\text{TrwC}_{\text{R388}}$-DNA complexes. On the other hand, the DNA cleavage-ligation equilibrium of the relaxase $\text{MobA}_{\text{RSF1010}}$ and TraI_{RP4} was not shifted toward the cleaved state by using reverse oligonucleotides.
7. The structure of $\text{TrwC}_{\text{R388}}$ covalently bound to DNA suggests a rearrangement of the loop that connects the two catalytic tyrosines.

SUMMARY IN SPANISH

INTRODUCCIÓN

La tendencia a autoensamblarse por la complementariedad de bases y su diámetro de 2 nanómetros convierten a los ácidos nucleicos en un material óptimo para la construcción de nanoestructuras artificiales. El gran avance de los origamis de ADN en la última década, ha conseguido construir nanoestructuras de ADN con cualquier forma imaginable para usarse en la administración de fármacos a células concretas (Douglas et al., 2012; Zhang et al., 2014), modulación del sistema inmune (Li et al., 2011; Liu et al., 2012), como biosensores (Tintoré et al., 2013) y como plataforma para crear cascadas enzimáticas (Fu et al., 2013; Yang et al., 2015).

El método de los origamis de ADN lleva el nombre de papiroflexia en japonés porque largas moléculas de cadena sencilla de ADN de fago M13, el "andamio", se pliegan en cualquier forma predeterminada con más de 200 oligonucleótidos llamados "grapas" (Figura E1A) (Rothemund, 2006). Las grapas son complementarias a distintas partes del genoma de M13 y crean “uniones de tipo Holliday” que pliegan el ADN del fago y mantienen las dobles hélices juntas. El proceso de elaboración de un origami comienza con el diseño de las grapas para el diseño requerido. Posteriormente, las grapas suministradas por una empresa de síntesis de oligonucleótidos, se mezclan con el ADN de M13 en un ratio 3-5:1 en presencia de magnesio (Figura E1B). La reacción de autoensamblado se hace en un termociclador con un gradiente de temperatura durante 1-48 horas. Por último, se analizan en el microscopio de fuerza atómica (MFA) o en el microscopio electrónico de transmisión (MET). Gracias a este método se pueden ensamblar con alto rendimiento nanoestructuras de ADN con cualquier geometría tridimensional (Figura E1C).

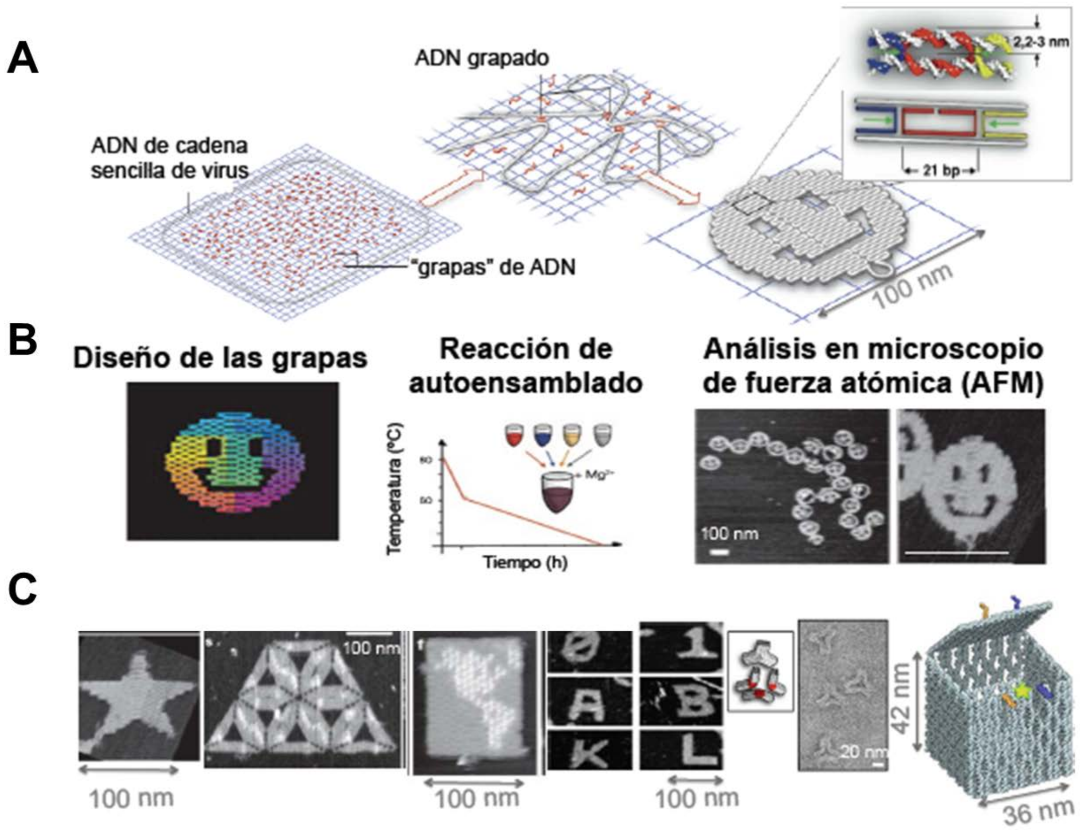


Figura E1. Origamis de ADN. (A) Las 7 kilobases del genoma viral de M13 se pliegan con cerca de 250 “grapas” de ADN formando una “cara sonriente” de un diámetro de 100 nm. (B) Etapas del proceso de elaboración de un origami de ADN. Tras el diseño de las grapas, se mezclan con el ADN del fago M13. El ensamblado se genera con un gradiente de temepratura en un termociclador. Posteriormente, se analizan en un microscopio. (C) Ejemplos de nanoestructuras de ADN de dos y tres dimensiones. Adaptado de (Rothemund, 2006; Andersen et al., 2009; Sanderson, 2010; Castro et al., 2011).

Las aplicaciones en biomedicina que hemos comentado implican la funcionalización de estas nanoestructuras con nanopartículas y proteínas. Para unir proteínas covalentemente al ADN (denominado también bioconjugar) se usan métodos de conjugación covalentes y no covalentes, y dentro de ellos, sitio-específicos y no sitio-específicos (Saccà and Niemeyer, 2011; Yang et al., 2015). Estos métodos modifican químicamente las proteínas, comprometiendo su actividad, o bien suponen el uso de proteínas recombinantes. Actualmente se requieren nuevos métodos para conseguir bioconjugar proteínas al ADN.

En esta tesis, hemos explorado la capacidad de las relaxasas para unirse de forma programada sobre nanoestructuras hechas de ADN ya que estas enzimas son capaces de formar un enlace covalente fosfotirosina con secuencias específicas de ADN.

Las relaxasas son proteínas de unión al ADN que participan en la transferencia de plásmidos por conjugación entre las bacterias. La conjugación bacteriana implica el contacto célula a célula con el fin de iniciar el proceso donde una de las cadenas del plásmido se transfiere desde la célula donadora a la célula receptora por un sistema de secreción de tipo IV (T4SS) (Figura E2) (la Cruz et al., 2010). El mecanismo de transferencia de ADN se puede dividir en tres etapas diferentes (Llosa and la Cruz, 2005):

(1) Ensamblaje del relaxosoma en el origen de transferencia (*oriT*). La relaxasa, las proteínas accesorias e IHF unidos alrededor de la *oriT* forman el complejo nucleoproteico llamado relaxosoma. El relaxosoma abre el ADN de doble cadena con el fin de localizar la hebra destinada a la translocación dentro del centro activo de la relaxasa.

(2) Reacción de corte. La relaxasa corta el ADN en el enlace fosfodiester del ADN de una posición específica, llamada sitio *nic*, estableciendo un enlace covalente con el ADN.

(3) Reclutamiento del sustrato y translocación a la célula receptora. La proteína de acoplamiento (T4CP) es entonces reclutada y vincula la relaxosoma con el T4SS. El T4SS es un complejo multiproteico que se extiende por las membranas bacterianas interior y exterior de la célula donadora. A continuación, se transfiere la relaxasa unida a una hebra del plásmido en dirección 5'-3 a través de la T4SS. En la célula receptora, se cataliza la recircularización el plásmido por la misma relaxasa (en relaxasas con dos tirosinas catalíticas, llamadas Y2) u otro monómero de relaxasa (para aquellas relaxasas con una sola tirosina catalítica o Y1). La síntesis de la cadena se procesa simultáneamente dentro de la donadora y la célula receptora. Al final de este proceso, la célula donadora y receptora contienen el plásmido conjugativo.

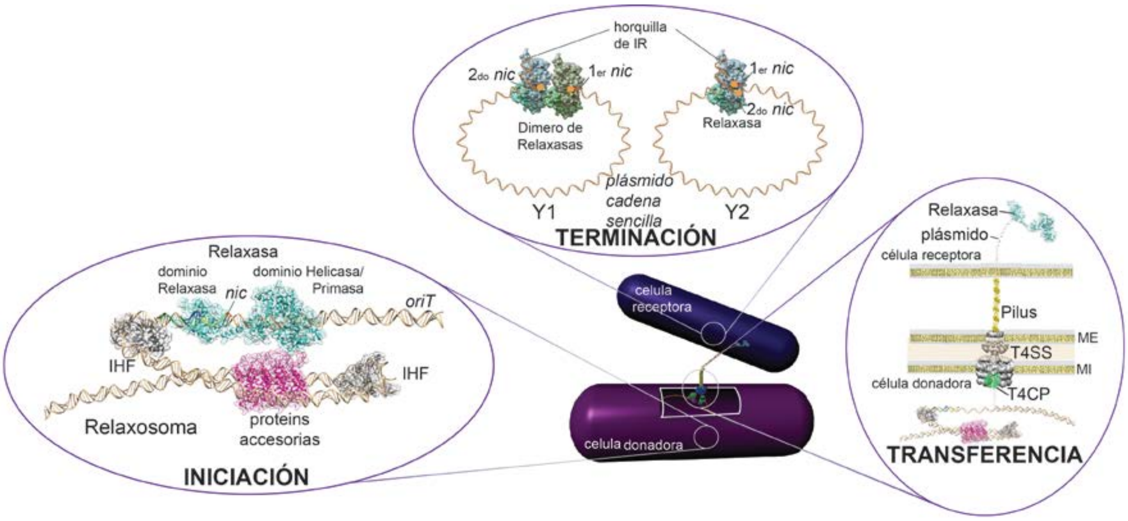


Figura E2. Conjugación bacteriana. Las fases de iniciación, transferencia y terminación están representadas con modelos moleculares. ME, membrana externa; MI, membrana interna. Modelos modificados de (Moncalián et al., 1999; Cabezon and la Cruz, 2014; Carballeira et al., 2014).

Las relaxasas poseen tres motivos conservados, como las proteínas iniciadoras de replicación por círculo rodante (Koonin and Ilyina, 1993):

- Motivo I: Contiene la(s) tirosina(s) catalítica(s) que cortan y ligan el ADN.
- Motivo II: Interactúa con el ADN.
- Motivo III: Posee el motivo HUH (histidina-residuo hidrofóbico-histidina) que participa en la coordinación del metal divalente.

El papel del metal es interactuar con oxígenos del fosfato para orientarle con relación a la tirosina catalítica. La reacción de corte es reversible e isoenergética: la energía del enlace fosfodiéster se mantiene en forma de una unión fosfotirosina, lo que permite la posterior religación del ADN cortado por el 3'OH libre (Guasch et al., 2003). Esta religación hace que la formación de complejos covalentes *in vitro* sea muy poco eficiente. "Sustratos suicidas" se han empleado para aumentar la eficiencia de corte y estudiar de manera independiente las reacciones de corte y religación del ADN en la mayoría de las enzimas de procesamiento de ADN. En estos oligonucleótidos un azufre reemplaza el oxígeno 3’ del sitio *nic*, de forma que tras el corte se genera un grupo 3’-SH que no puede atacar a la fosfotirosina. Por lo tanto, los complejos covalentes de proteína-ADN se acumulan. Con oligonucleótidos suicidas, la relaxasa TrwC de R388 alcanzó el 100% de los complejos covalentes (González-Pérez et al., 2007). Oligonucleótidos suicidas se han utilizado con la topoisomerasa II para resolver la estructura del complejo de corte (Wendorff et al., 2012).

OBJETIVOS

Como el ensamblaje de proteínas en nanoestructuras de ADN ha demostrado tener muchas aplicaciones en biomedicina, nos hemos centrado en el desarrollo de una nueva estrategia de bioconjugación basada en las proteínas relaxasa. El objetivo principal de esta tesis es estudiar si las relaxasas pueden servir para bioconjugar otras proteínas sobre nanoestructuras de ADN. Para este objetivo hemos diseñado dianas de las relaxasas que sobresalen de ciertas nanoestructuras de ADN y que permiten la distribución de relaxasas a lo largo de la nanoestructura. Para alcanzar el objetivo principal se especificaron los siguientes objetivos específicos:

1. Definición del mejor diseño del oligonucleótido diana y su mejor posición en las nanoestructuras de ADN para nuestra relaxasa modelo, TrwC del plásmido R388.

2. Demostración de que cualquier proteína fusionada al dominio relaxasa, por ejemplo, un relaxasa fusionada a un dominio fluorescente, también puede unirse a lugares específicos en nanoestructuras de ADN.

3. Desplazamiento del equilibrio de corte-religación de las relaxasas mediante el uso de "sustratos suicidas" en nanoestructuras de ADN para aumentar el porcentaje de complejos covalentes relaxasa-DNA.

4. Evaluación de la unión específica de varias relaxasas a nanoestructuras de ADN.

5. Estudio de nuevos substratos de ADN que permitirían el aumento de la unión covalente de las relaxasas.

6. Proporción de un mecanismo para la polimerización estable de moléculas de ADN de cadena doble y sencilla o nanoestructuras de ADN que sea llevado a cabo por relaxasas.

RESULTADOS Y DISCUSIÓN

Capítulo 1. Unión de relaxasas sobre nanoestructuras de ADN

En esta tesis, hemos estudiado la unión de las relaxasas TrwC del plásmido R388, TraI de R100, TraI de pKM101 y MobA de R1162 (a las que denominamos: $\text{TrwC}_{\text{R388}}$, $\text{TraI}_{\text{R100}}$, $\text{TraI}_{\text{pKM101}}$ y $\text{MobA}_{\text{R1162}}$, respectivamente) sobre origamis de ADN. Para ello usamos unos origamis rectangulares planos y unos nanotubos hechos con un manojo de seis hélices de ADN. Las dimensiones de los rectángulos son 90x60x2 nm^3 y la longitud de los nanotubos es 430 nm y 5 nm de diámetro. Encontramos que la unión de estas proteínas es altamente específica con rendimientos de unión del 50% a un único sitio de unión, y que se podría lograr hasta un 90% de unión utilizando sitios de unión redundantes en estrecha proximidad (Figuras E3 y E4). Estos porcentajes son comparables a los de otras estrategias de funcionalización. También encontramos que los rendimientos de unión son diferentes para las dos estructuras de origami y también dependen de la posición de los sitios de unión. El rendimiento de unión de nanoestructuras de ADN no se pudo mejorar usando substratos "suicidas" para las relaxasas $\text{TrwC}_{\text{R388}}$ y $\text{TraI}_{\text{pKM101}}$. Sin embargo se encontró un porcentaje similar de proteínas sobre nanotubos cuando se evaluaron tras 1 hora y 24 horas de incubación. Dos de las cuatro relaxasas investigadas, $\text{TraI}_{\text{R100}}$ y $\text{TrwC}_{\text{R388}}$, mostraron un rendimiento de unión muy bueno con alta especificad y compatibilidad, lo que conocemos como ortogonalidad. Debido a su especificidad para dianas de ADN de cadena sencilla, su baja constante de afinidad y su buena ortogonalidad, las relaxasas nos permiten la funcionalización de origamis de ADN. Por ello, proponemos el uso de las relaxasas como un adaptador para unir otras proteínas de interés sobre nanoestructuras de ADN.

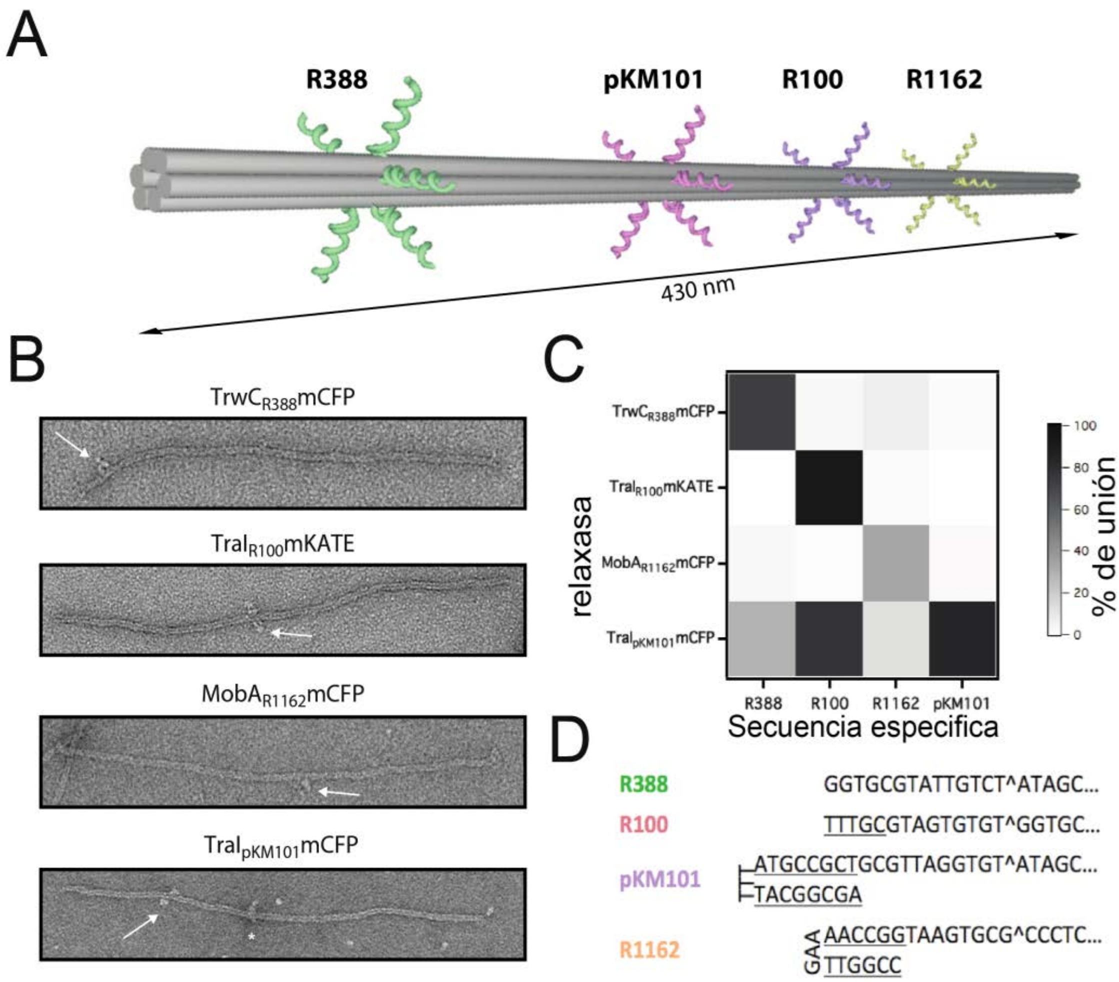


Figura E3. Nanotubos decorados con cuatro relaxasas. (A) Esquema del nanotubo que tiene secuencias diana para las relaxasas de plásmidos R388, pKM101,R100 y R1162. El nanotubo esta formado por seis dobles helices ensambladas con mas de 250 grapas. Cada cilindro representa una doble helice de ADN. Cinco sitos de unión con la secuencia específica de cada relaxasa sobresalen en cada posición. (B) Micrografías tomadas con microscopio electrónico de transmisión (MET). Las relaxasas específicamente unidas están señaladas por una flecha, una estrella indica unión inespecífica. (C) Rendimiento de unión. (D) Secuencias de ADN que cada relaxasa reconoce. El subrayado indica la secuencia que forma la repetición invertida. El sitio *nic* esta representado por el símbolo ^.

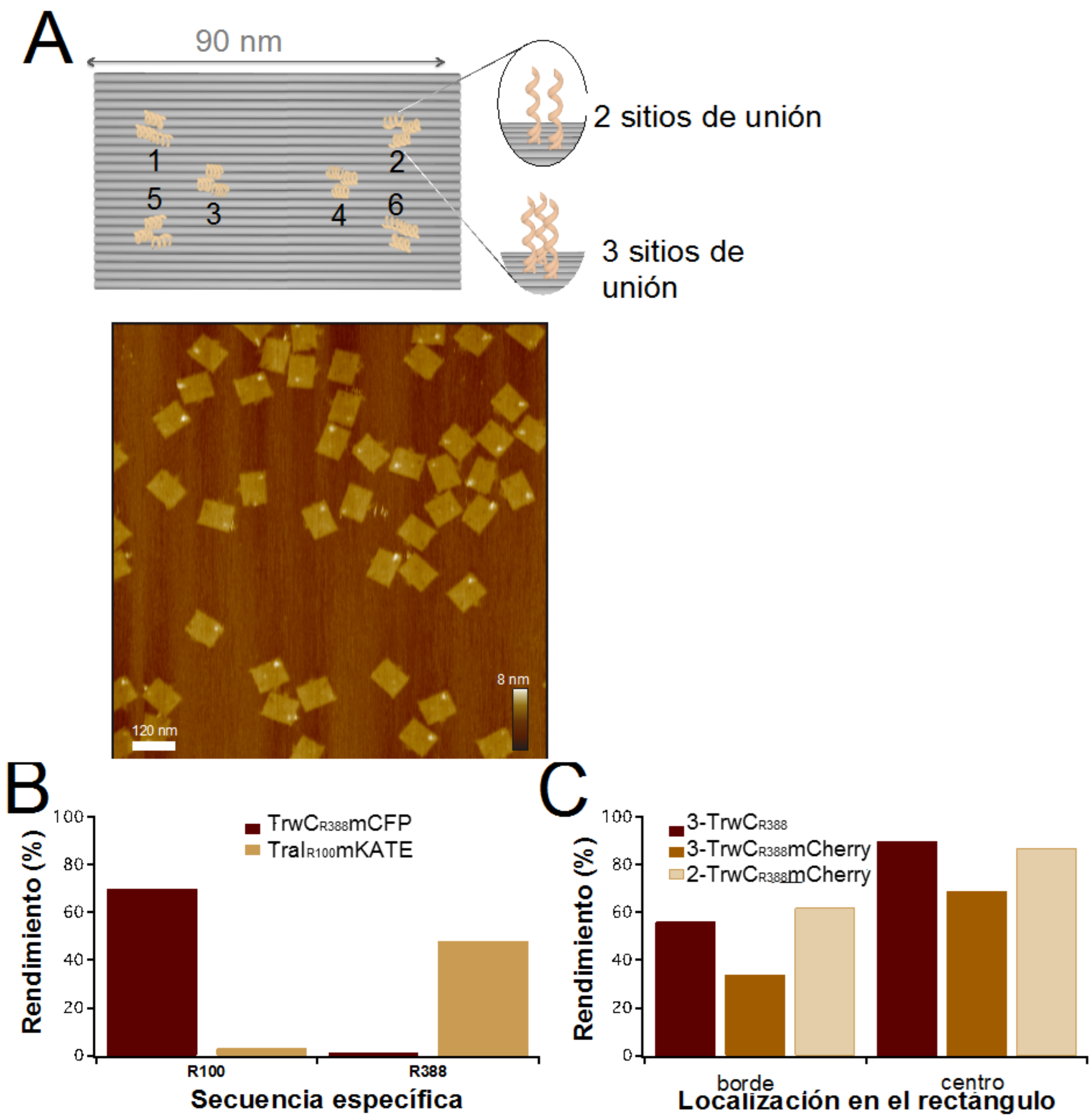


Figura E4. Origamis rectangulares decorados con relaxasas. (A) Esquema de las dianas en rectángulo. Cada cilindro gris representa una doble helice de ADN. Cada posición, numerada del uno al seis, contiene dos o tres sitos de unión con la secuencia específica de cada relaxasa sobresalen en cada posición. Micrografía tomada con microscopio de fuerza atómica (MFA) de la relaxasa TraI$_{R100}$ reconociendo su diana en posición 1. (B) Rendimiento de unión especifica de las relaxasas TraI$_{R100}$ y TrwC$_{R388}$ en rectángulo. Tras una incubación de 24 h, las nanoestructuras se purificaron, y se analizazaron en el MFA. (C) Rendimiento de unión de TrwC$_{R388}$ sobre rectángulo esquematizado en (A) con dos o tres sitios de unión en cada una de las 6 posiciones para la relaxasa TrwC$_{R388}$.

También investigamos las relaxasas como una nueva herramienta de polimerización en bionanotecnología. Nuestro objetivo era unir de forma covalente secuencias cortas de ADN sencillo, doble o incluso origamis de ADN. Nanoestructuras lineares con la secuencia de reconocimiento de TrwC$_{R388}$ en ambos extremos se pueden llegar a polimerizar por la reacción de transferencia de cadenas que lleva a cabo la relaxasa (Figura E5A). Las imágenes de MFA muestran que se consiguieron formar cadenas lineales de ADN y proteína TrwC$_{R388}$ que no aparecían cuando el ADN o la relaxasa se incubaban por separado (Figura E5B). La mayoría de estas cadenas eran dímeros, trímeros y tetrámeros (Tabla E1). Creemos que la relaxasa puede cortar de nuevo la diana reestablecida entre los monómeros de ADN, limitando la síntesis de largas cadenas. Aunque este método no puede competir con las actuales estrategias de polimerización de nanoestructuras, es la primera vez que se una enzima dirige la polimerización de estos ensamblajes de escala nanométrica.

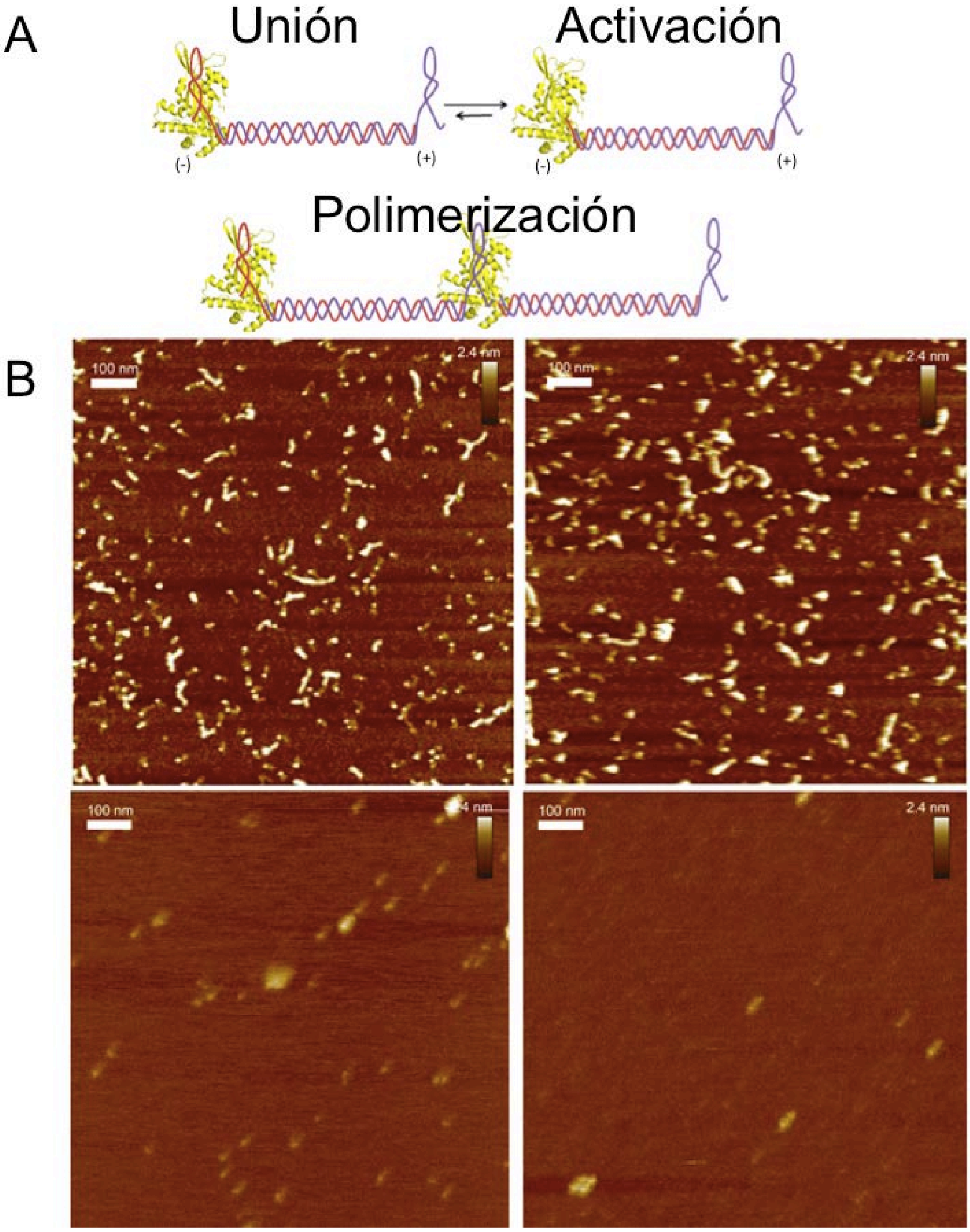


Figura E5. Polimerización de nanostructuras de ADN de doble cadena por la relaxasa TrwC. (A) Mecanismo de polimerización basado en la trasferencia de cadenas del extremo (-) al (+) mediada por la relaxasa TrwC$_{R388}$ (en amarillo). Tras la unión a la diana 24+8 (extremo -), la relaxasa corta el ADN en el sitio *nic* y se queda unida covalentemente al ADN. En este estado de activación, si la relaxasa reconoce la diana 24+0 del extremo +, va a llevar a cabo la reaccion de transferencia de cadenas, y consiguiente polimerización en forma de nanocable. (B) Imágenes de MFA de polímeros con diana 24+8 (arriba a la izquierda) y 8+24 (arriba a la derecha) incubados durante una semana a temperatura ambiente con TrwC$_{R388}$mCherry. Los controles negativos de relaxasas (abajo a la izquierda) y monómeros de ADN (abajo a la derecha) muestran que estas estructuras lineales no aparecen.

Tabla E1. Cuantificación de los polímeros de relaxasa con nanostructuras lineares de doble cadena. Las distancias entre picos de proteinas y la longitud total de los nanocables se midieron de secciones de las imagenes de altura de AFM. La media de los valores obtenidos de 96 polimeros con la diana R(8+24) y de 111 polimeros con la diana 24+8.

	Polímeros	Número de puntos	Eventos	Rendimiento	Longitud media
polímeros con diana 8+24-	dímero	3	52	54,2%	43
	trímero	4	26	27,1%	62
	tetrámero	5	10	10,4%	97
	pentámero	6	7	7,3%	110
	hexámero	7	1	1,0%	116
	Total		96		
polímeros con diana 24+8	dímero	3	58	52,3%	42
	trímero	4	35	31,5%	60
	tetrámero	5	11	9,9%	66
	pentámero	6	7	6,3%	104
	hexámero	7	0	0,0%	-
	total		111		

Capítulo 2. Mejora de la reacción de corte de las relaxasas

Nuestro objetivo también ha sido la mejora de la reacción de corte de las relaxasas porque en la reacción original se llega a un equilibrio dinámico de corte y religación que supone que solo un 10-30% de las proteínas están unidas covalentemente al ADN. Para ello modificamos la secuencia de reconocimiento de las relaxasas TrwC$_{R388}$, TraI$_{RP4}$ y MobA$_{R1162}$ para ver si eran capaces de cortar el ADN cuando este tiene la diana de corte con la conformación de horquilla que reconocen las replicasas, con una conformación en la que se han permutado el sitio *nic* y el IR (conformación inversa), e incluso con doble cadena después del sitio *nic* (Figura E6). Encontramos que el porcentaje de complejos covalentes llegaba hasta el 50% con TrwC$_{R388}$ incubada con oligonucleótidos con conformación inversa. También la presencia de doble cadena en el sitio *nic*, llegaba a ese porcentaje de complejos porque la reacción de religación se veía desfavorecida. Sin embargo, las otras relaxasas no incrementaban el porcentaje de corte de los nuevos substratos, aunque conseguían cortarlos. Los plásmidos sintéticos con las configuraciones replicasa e inversa en el sitio *nic* no eran eficientemente movilizados por TrwC$_{R388}$ (Figura E7). Pensamos que esta disminución de frecuencia puede ser debido a que también hemos cambiado la distancia relativa de las proteínas que forman el relaxosoma, las proteínas IHF y TrwA. Además, la nueva conformación de la repetición invertida puede afectar a la reacción de terminación. Nuestros próximos estudios se centraran en la generación de *oriT*s sintéticos que pueden ser transmisibles sin proteínas accesorias.

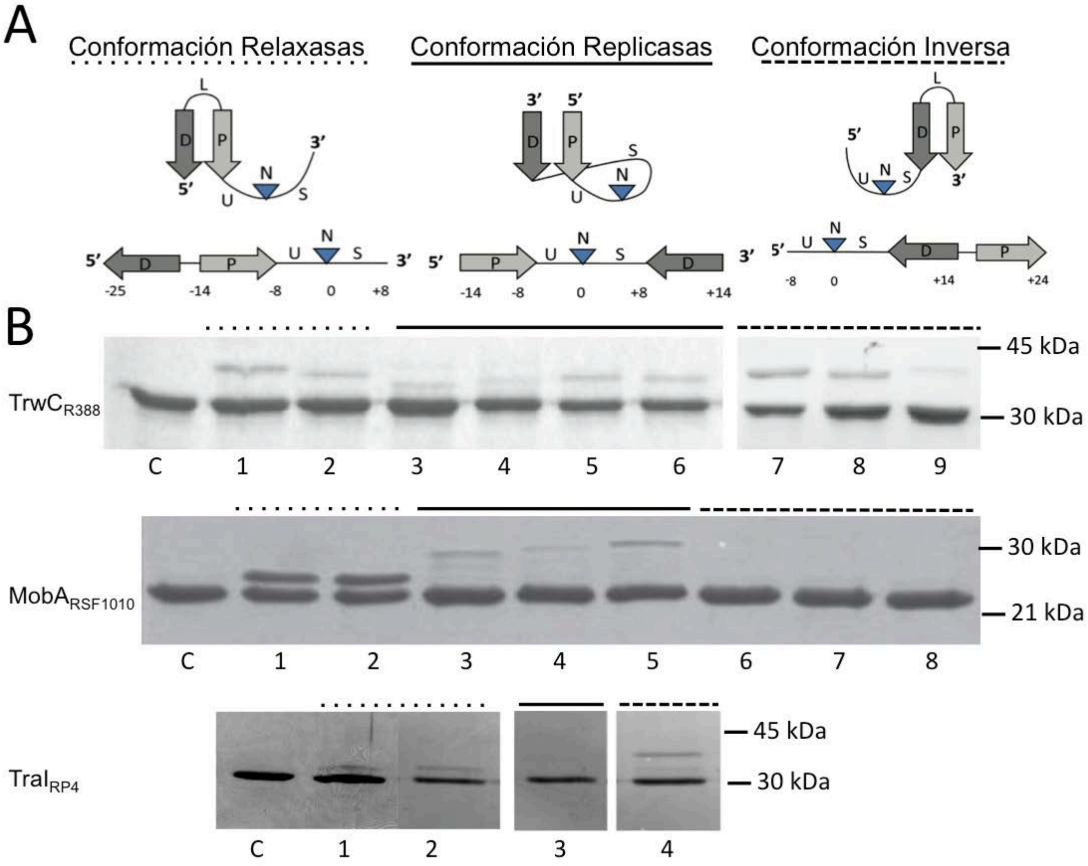


Figura E6. Reaciones de corte de oligonucleótidos con diferentes conformaciones. (A) Esquema de las diferentes conformaciones de los oligonucleotidos con la secuencia de reconocimiento de las relaxasas analizadas. D, brazo distal; L, lazo; P, brazo proximal; U, giro con forma de U;

N, sitio nic; S, sequencia variable. (B) Geles de SDS-PAGE en el que se ve el corte de las diferentes relaxasas estudiadas. 7μM de TrwC$_{R388}$, se incubaron con 15 μM de los siguientes oligonucleótidos: las calles 1 y 2, conformación relaxasa W(25+28) y W(14+14); conformación replicasa en calles 3 a 6 con oligonucleótidos H(14+8), H(14+10), H(14+12) y H(14+13), respectivamente; y inversos en calles 7 a 9, R(8+27), R(8+24) y R(7+26). 7 μM MobA$_{RSF1010}$ se incubaron con 15 μM de los oligonucleotidos tipo relaxasa W(30+7) y W(23+7), calles 1 y 2; replicasa HQ(16+16), HQ(16+19), HQ(16+22), calles 4 a 6; e inversos RQ(8+28); RQ(8+34), RQ(8+40), calles 7 a 9. 1,5 μM de TraI$_{RP4}$ se incubaron con siguientes oligonucleótidos: conformación relaxasa W(15+6) y W(24+8) en las primeras dos calles, de tipo replicasa HP(14+14) en calle 3 y tipo inverso RP(8+24) en calle 4. La banda que aparece con menor movilidad en el gel es el complejo covalente ADN-relaxasa.

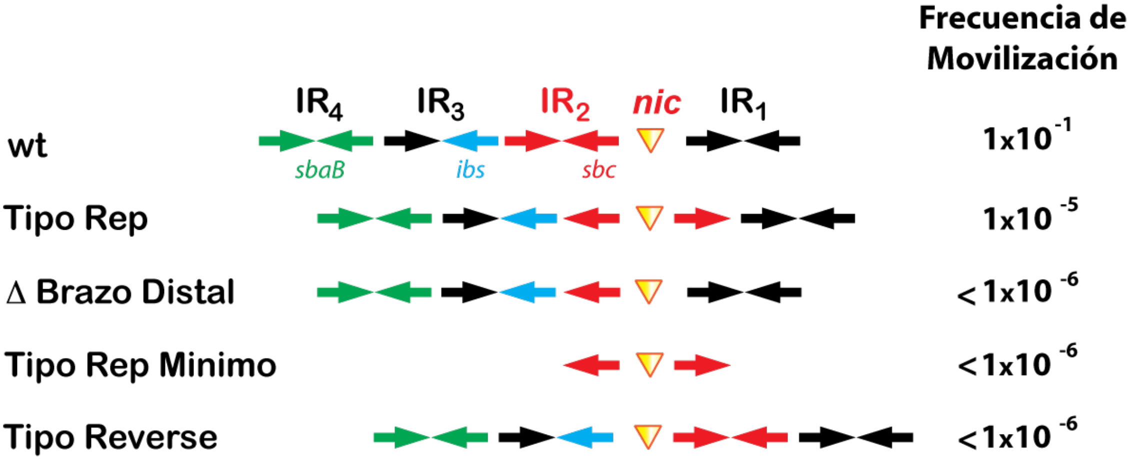


Figura E7. Movilización de plásmidos sintéticos con el *oriT* de R388 con nuevas conformaciones. Los ensayos de movilización se llevaron a cabo con células donadoras que contienen estos plásmidos sintéticos (Cm^R) junto al plásmido ayudante de R388 (pSU2007, Km^R) y a células receptoras UB1637 (Sm^R) durante 1h a 37ºC. Las frecuencias de movilización fueron determinadas calculando el número de transconjugantes (Cm^{R}Sm^R) por donador (Cm^{R}Nx^R).

Durante esta tesis, también hemos llevado a cabo estudios estructurales de la relaxasa TrwC$_{R388}$ con oligonucleótidos que imitan el ADN en la etapa inicial de conjugación y también oligonucleótidos “suicidas” que permiten la formación de 100% de complejos covalentes. En ambos, la celda unidad estaba formada por dímeros de relaxasa. El análisis de los residuos que parecían estar involucrados en la interacción de dímeros de proteína mostró que estos residuos no tenían un papel biológico. Por primera vez hemos conseguido resolver la estructura de una relaxasa unida covalentemente al ADN. En los complejos con oligos “suicidas” se ve una reordenación del loop que lleva las tirosinas catalíticas. El refinamiento de la estructura no nos deja ver claramente la posición de los nucleótidos cortados, pero seguimos trabajando para conseguir una mejor estructura de estos complejos.

CONCLUSIONES

1. Las relaxasas fluorescentes pueden eficientemente ensamblarse en una nanoestructura plana cuando los sitios de unión se organizan en estrecha proximidad con una base rígida de 5 pares de bases seguida de una región de cadena sencilla con la secuencia diana de la relaxasa. Tiempos de incubación cortos, tales como 1 hora, dan como resultado una decoración igual de eficiente que las incubaciones durante 24 horas.

2. Diferentes relaxasas se organizan de manera eficiente sobre origamis rectangulares y nanotubulares. La relaxasa $\text{TrwC}_{\text{R388}}$ se ensambla en nanoestructuras rectangulares con preferencia de posición por las posiciones centrales respecto a las que están cerca del borde. $\text{TrwC}_{\text{R388}}$, $\text{TraI}_{\text{R100}}$, $\text{MobA}_{\text{R1162}}$ tienen la capacidad de unión ortogonal, es decir, específica y de alta afinidad. Sin embargo, $\text{TraI}_{\text{pKM101}}$ muestra una significativa unión no específica a sitios de unión de otras relaxasas.

3. El rendimiento de unión de nanoestructuras de ADN no se ha podido mejorar mediante substratos "suicidas" para las relaxasas $\text{TrwC}_{\text{R388}}$ y $\text{TraI}_{\text{pKM101}}$.

4. La relaxasa $\text{TrwC}_{\text{R388}}$ dirige el ensamblaje en forma de “collar de perlas” de ADN monocatenario y de doble cadena, pero también de “nanolazos” de origamis rectangulares. Tras la polimerización, la relaxasa sigue unida a las nanoestructuras.

5. Las relaxasas $\text{TrwC}_{\text{R388}}$ y $\text{MobA}_{\text{RSF1010}}$ son capaces de cortar sustratos con diseño del sitio *nic* tipo replicasa, pero TraI_{RP4} no es capaz de cortar este tipo de oligonucleótidos. Sin embargo, los plásmidos con *oriT* con diseño en la región del sitio *nic* tipo replicasa o inverso no son funcionales *in vivo*.

6. El porcentaje de complejos covalentes ha mejorado cuando hemos cambiando el diseño del sustrato a una disposición inversa o con presencia de doble cadena en el sitio *nic*. Estos diseños conducen a un total de 50% de

complejos TrwC$_{R388}$-ADN. Por otro lado, el equilibrio de corte y religación de las relaxasas MobA$_{RSF1010}$ y TraI$_{RP4}$ no se desplazó hacia la formación de complejos covalentes mediante el uso de oligonucleótidos con disposición inversa.

7. La estructura de TrwC$_{R388}$ unida covalentemente al ADN revela un reordenamiento del lazo móvil que conecta los dos tirosinas catalíticas Y18 e Y26.

BIBLIOGRAFÍA

Andersen, E.S., Dong, M., Nielsen, M.M., Jahn, K., Subramani, R., Mamdouh, W., Golas, M.M., Sander, B., Stark, H., Cristiano L. P. Oliveira, et al. (2009). Self-assembly of a nanoscale DNA box with a controllable lid. Nature *459*, 73–76.

Cabezon, E., and la Cruz, de, F. (2014). Towards an integrated model of bacterial conjugation. FEMS Microbiology Reviews *39*, n/a–n/a.

Carballeira, J.D., González-Pérez, B., Moncalián, G., and la Cruz, de, F. (2014). A high security double lock and key mechanism in HUH relaxases controls oriT-processing for plasmid conjugation. Nucleic Acids Research *42*, 10632–10643.

Castro, C.E., Kilchherr, F., Kim, D.-N., Shiao, E.L., Wauer, T., Wortmann, P., Bathe, M., and Dietz, H. (2011). A primer to scaffolded DNA origami. Nat Meth *8*, 221–229.

Douglas, S.M., Bachelet, I., and Church, G.M. (2012). A logic-gated nanorobot for targeted transport of molecular payloads. Science *335*, 831–834.

Fu, Y., Zeng, D., Chao, J., Jin, Y., Zhang, Z., Liu, H., Li, D., Ma, H., Huang, Q., Gothelf, K.V., et al. (2013). Single-step rapid assembly of DNA origami nanostructures for addressable nanoscale bioreactors. J. Am. Chem. Soc. *135*, 696–702.

González-Pérez, B., Lucas, M., Cooke, L.A., Vyle, J.S., la Cruz, de, F., and Moncalián, G. (2007). Analysis of DNA processing reactions in bacterial conjugation by using suicide oligonucleotides. The EMBO Journal *26*, 3847–3857.

Guasch, A., Lucas, M., Moncalián, G., Cabezas, M., Pérez-Luque, R., Gomis-Rüth, F.X., la Cruz, de, F., and Coll, M. (2003). Recognition and processing of the origin of transfer DNA by conjugative relaxase TrwC. Nat Struct Biol *10*, 1002–1010.

Koonin, E.V., and Ilyina, T.V. (1993). Computer-assisted dissection of rolling circle DNA replication. BioSystems *30*, 241–268.

la Cruz, de, F., Frost, L.S., Meyer, R.J., and Zechner, E.L. (2010). Conjugative DNA metabolism in Gram-negative bacteria. FEMS Microbiology Reviews *34*, 18–40.

Li, J., Pei, H., Zhu, B., Liang, L., Wei, M., He, Y., Chen, N., Li, D., Huang, Q., and Fan, C. (2011). Self-assembled multivalent DNA nanostructures for noninvasive intracellular delivery of immunostimulatory CpG oligonucleotides. ACS Nano *5*, 8783–8789.

Liu, X., Xu, Y., Yu, T., Clifford, C., Liu, Y., Yan, H., and Chang, Y. (2012). A DNA nanostructure platform for directed assembly of synthetic vaccines. Nano

Lett. *12*, 4254–4259.

Llosa, M., and la Cruz, de, F. (2005). Bacterial conjugation: a potential tool for genomic engineering. Research in Microbiology *156*, 1–6.

Moncalián, G., Valle, M., Valpuesta, J.M., and la Cruz, de, F. (1999). IHF protein inhibits cleavage but not assembly of plasmid R388 relaxosomes. Mol Microbiol *31*, 1643–1652.

Rothemund, P.W.K. (2006). Folding DNA to create nanoscale shapes and patterns. Nature *440*, 297–302.

Saccà, B., and Niemeyer, C.M. (2011). Functionalization of DNA nanostructures with proteins. Chem. Soc. Rev. *40*, 5910–5921.

Sanderson, K. (2010). Bioengineering: What to make with DNA origami. Nature *464*, 158–159.

Tintoré, M., Gállego, I., Manning, B., Eritja, R., and Fàbrega, C. (2013). DNA origami as a DNA repair nanosensor at the single-molecule level. Angew. Chem. Int. Ed. Engl. *52*, 7747–7750.

Wendorff, T.J., Schmidt, B.H., Heslop, P., Austin, C.A., and Berger, J.M. (2012). The Structure of DNA-Bound Human Topoisomerase II Alpha: Conformational Mechanisms for Coordinating Inter-Subunit Interactions with DNA Cleavage. Journal of Molecular Biology *424*, 109–124.

Yang, Y.R., Liu, Y., and Yan, H. (2015). DNA Nanostructures as Programmable Biomolecular Scaffolds. Bioconjugate Chem.

Zhang, Q., Jiang, Q., Li, N., Dai, L., Liu, Q., Song, L., Wang, J., Li, Y., Tian, J., Ding, B., et al. (2014). DNA origami as an in vivo drug delivery vehicle for cancer therapy. ACS Nano *8*, 6633–6643.

REFERENCES

Agundez, L., Machon, C., Cesar, C.E., Rosa-Garrido, M., Delgado, M.D., and Llosa, M. (2010). Nuclear Targeting of a Bacterial Integrase That Mediates Site-Specific Recombination between Bacterial and Human Target Sequences. Applied and Environmental Microbiology *77*, 201–210.

Agúndez, L., González-Prieto, C., Machón, C., and Llosa, M. (2012). Site-Specific Integration of Foreign DNA into Minimal Bacterial and Human Target Sequences Mediated by a Conjugative Relaxase. PLoS ONE *7*, e31047.

Akter, F., Mie, M., Grimm, S., Nygren, P.-Å., and Kobatake, E. (2012). Detection of antigens using a protein-DNA chimera developed by enzymatic covalent bonding with phiX gene A*. Anal. Chem. *84*, 5040–5046.

Aldaye, F.A., Senapedis, W.T., Silver, P.A., and Way, J.C. (2010). A structurally tunable DNA-based extracellular matrix. J. Am. Chem. Soc. *132*, 14727–14729.

Alperi, A., Larrea, D., Fernández-González, E., Dehio, C., Zechner, E.L., and Llosa, M. (2013). A translocation motif in relaxase TrwC specifically affects recruitment by its conjugative type IV secretion system. Journal of Bacteriology *195*, 4999–5006.

Alvarado, A., Garcillán-Barcia, M.P., and la Cruz, de, F. (2012). A Degenerate Primer MOB Typing (DPMT) Method to Classify Gamma-Proteobacterial Plasmids in Clinical and Environmental Settings. PLoS ONE *7*, e40438.

Alvarez-Martinez, C.E., and Christie, P.J. (2009). Biological Diversity of Prokaryotic Type IV Secretion Systems. Microbiology and Molecular Biology Reviews *73*, 775–808.

Andersen, E.S., Dong, M., Nielsen, M.M., Jahn, K., Subramani, R., Mamdouh, W., Golas, M.M., Sander, B., Stark, H., Cristiano L. P. Oliveira, et al. (2009). Self-assembly of a nanoscale DNA box with a controllable lid. Nature *459*, 73–76.

Barabas, O., Ronning, D.R., Guynet, C., Hickman, A.B., Ton-Hoang, B., Chandler, M., and Dyda, F. (2008). Mechanism of IS200/IS605 Family DNA Transposases: Activation and Transposon-Directed Target Site Selection. Cell *132*, 208–220.

Becker, E.C., and Meyer, R. (2012). Origin and fate of the 3' ends of single-stranded DNA generated by conjugal transfer of plasmid R1162. Journal of Bacteriology *194*, 5368–5376.

Beyer, S., Nickels, P., and Simmel, F.C. (2005). Periodic DNA Nanotemplates Synthesized by Rolling Circle Amplification. Nano Lett. *5*, 719–722.

Bhattacharjee, M., Rao, X.M., and Meyer, R.J. (1992). Role of the origin of transfer in termination of strand transfer during bacterial conjugation. Journal of Bacteriology *174*, 6659–6665.

Bhattacharjee, M.K., and Meyer, R.J. (1993). Specific binding of MobA, a plasmid-encoded protein involved in the initiation and termination of conjugal DNA transfer, to single-stranded oriT DNA. Nucleic Acids Research *21*, 4563–4568.

Boer, D.R., Ruíz-Masó, J.A., López-Blanco, J.R., Blanco, A.G., Vives-Llàcer, M., Chacón, P., Usón, I., Gomis-Rüth, F.X., Espinosa, M., Llorca, O., et al. (2009). Plasmid replication initiator RepB forms a hexamer reminiscent of ring helicases and has mobile nuclease domains. The EMBO Journal *28*, 1666–1678.

Boer, R., Russi, S., Guasch, A., Lucas, M., Blanco, A.G., Pérez-Luque, R., Coll, M., and la Cruz, de, F. (2006). Unveiling the Molecular Mechanism of a Conjugative Relaxase: The Structure of TrwC Complexed with a 27-mer DNA Comprising the Recognition Hairpin and the Cleavage Site. Journal of Molecular Biology *358*, 857–869.

Brasch, M.A., and Meyer, R.J. (1987). A 38 base-pair segment of DNA is required in cis for conjugative mobilization of broad host-range plasmid R1162. Journal of Molecular Biology *198*, 361–369.

Cabezon, E., and la Cruz, de, F. (2014). Towards an integrated model of bacterial conjugation. FEMS Microbiology Reviews *39*, n/a–n/a.

Carballeira, J.D., González-Pérez, B., Moncalián, G., and la Cruz, de, F. (2014). A high security double lock and key mechanism in HUH relaxases controls oriT-processing for plasmid conjugation. Nucleic Acids Research *42*, 10632–10643.

Castro, C.E., Kilchherr, F., Kim, D.-N., Shiao, E.L., Wauer, T., Wortmann, P., Bathe, M., and Dietz, H. (2011). A primer to scaffolded DNA origami. Nat Meth *8*, 221–229.

César, C.E., Machón, C., la Cruz, de, F., and Llosa, M. (2006). A new domain of conjugative relaxase TrwC responsible for efficient oriT-specific recombination on minimal target sequences. Mol Microbiol *62*, 984–996.

Chandler, M., la Cruz, de, F., Dyda, F., Hickman, A.B., Moncalián, G., and Ton-Hoang, B. (2013a). Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. Nat Rev Micro *11*, 525–538.

Changela, A., DiGate, R.J., and Mondragón, A. (2001). Crystal structure of a complex of a type IA DNA topoisomerase with a single-stranded DNA molecule. Nature *411*, 1077–1081.

Charoenphol, P., and Bermudez, H. (2014). Design and application of multifunctional DNA nanocarriers for therapeutic delivery. Acta Biomaterialia *10*, 1683–1691.

Chen, H., Zeng, Y., Liu, W., Zhao, S., Wu, J., and Du, Y. (2013). Multifaceted applications of nanomaterials in cell engineering and therapy. Biotechnology Advances *31*, 638–653.

Choi, K.-M., Kwon, I.C., and Ahn, H.J. (2013). Self-assembled amphiphilic DNA-cholesterol/DNA-peptide hybrid duplexes with liposome-like structure for doxorubicin delivery. Biomaterials *34*, 4183–4190.

Clark, N.J., Raththagala, M., Wright, N.T., Buenger, E.A., Schildbach, J.F., Krueger, S., and Curtis, J.E. (2014). Structures of TraI in solution. J Mol Model *20*, 2308.

Datta, N., and Hedges, R.W. (1972). Trimethoprim resistance conferred by W plasmids in Enterobacteriaceae. J. Gen. Microbiol. *72*, 349–355.

Debold, E.P. (2015). Molecular motor teamwork. Nature Nanotech *10*, 656–657.

Delebecque, C.J., Lindner, A.B., Silver, P.A., and Aldaye, F.A. (2011). Organization of Intracellular Reactions with Rationally Designed RNA Assemblies. Science *333*, 470–474.

Derr, N.D., Goodman, B.S., Jungmann, R., Leschziner, A.E., Shih, W.M., and Reck-Peterson, S.L. (2012). Tug-of-war in motor protein ensembles revealed with a programmable DNA origami scaffold. Science *338*, 662–665.

Deweese, J.E., Burgin, A.B., and Osheroff, N. (2008a). Human topoisomerase II uses a two-metal-ion mechanism for DNA cleavage. Nucleic Acids Research *36*, 4883–4893.

Deweese, J.E., Burgin, A.B., and Osheroff, N. (2008b). Using 3′-Bridging Phosphorothiolates To Isolate the Forward DNA Cleavage Reaction of Human Topoisomerase IIα †. Biochemistry *47*, 4129–4140.

Dostal, L., Shao, S., and Schildbach, J.F. (2011). Tracking F plasmid TraI relaxase processing reactions provides insight into F plasmid transfer. Nucleic Acids Research *39*, 2658–2670.

Douglas, S.M., Bachelet, I., and Church, G.M. (2012a). A logic-gated nanorobot for targeted transport of molecular payloads. Science *335*, 831–834.

Douglas, S.M., Chou, J.J., and Shih, W.M. (2007). DNA-nanotube-induced alignment of membrane proteins for NMR structure determination. Proc. Natl. Acad. Sci. U.S.a. *104*, 6644–6648.

Douglas, S.M., Dietz, H., Liedl, T., Högberg, B., Graf, F., and Shih, W.M. (2009a). Self-assembly of DNA into nanoscale three-dimensional shapes. Nature *459*, 414–418.

Douglas, S.M., Marblestone, A.H., Teerapittayanon, S., Vazquez, A., Church, G.M., and Shih, W.M. (2009b). Rapid prototyping of 3D DNA-origami shapes with caDNAno. Nucleic Acids Research *37*, 5001–5006.

Draper, O., César, C.E., Machón, C., la Cruz, de, F., and Llosa, M. (2005). Site-specific recombinase and integrase activities of a conjugative relaxase in recipient cells. Proc. Natl. Acad. Sci. U.S.a. *102*, 16385–16390.

Edwards, J.S., Betts, L., Frazier, M.L., Pollet, R.M., Kwong, S.M., Walton, W.G., Ballentine, W.K., Huang, J.J., Habibi, S., Del Campo, M., et al. (2013). Molecular basis of antibiotic multiresistance transfer in Staphylococcus aureus. Proc. Natl. Acad. Sci. U.S.a. *110*, 2804–2809.

Fernández-López, C., Lorenzo-Díaz, F., Pérez-Luque, R., Rodríguez-González, L., Boer, R., Lurz, R., Bravo, A., Coll, M., and Espinosa, M. (2013). Nicking activity of the pMV158 MobM relaxase on cognate and heterologous origins of transfer. Plasmid *70*, 120–130.

Fradkov, A.F., Chen, Y., Ding, L., Barsova, E.V., Matz, M.V., and Lukyanov, S.A. (2000). Novel fluorescent protein from Discosoma coral and its mutants possesses a unique far-red fluorescence. FEBS Letters *479*, 127–130.

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

Frost, L.S., Leplae, R., Summers, A.O., and Toussaint, A. (2005). Mobile genetic elements: the agents of open source evolution. Nat Rev Micro *3*, 722–732.

Fu, J., Liu, M., Liu, Y., Woodbury, N.W., and Yan, H. (2012). Interenzyme Substrate Diffusion for an Enzyme Cascade Organized on Spatially Addressable DNA Nanostructures. J. Am. Chem. Soc. *134*, 5516–5519.

Fu, T.J., and Seeman, N.C. (1993). DNA double-crossover molecules. Biochemistry *32*, 3211–3220.

Fu, Y., Zeng, D., Chao, J., Jin, Y., Zhang, Z., Liu, H., Li, D., Ma, H., Huang, Q., Gothelf, K.V., et al. (2013). Single-step rapid assembly of DNA origami nanostructures for addressable nanoscale bioreactors. J. Am. Chem. Soc. *135*, 696–702.

Furuya, N., and Komano, T. (1991). Determination of the nick site at oriT of IncI1 plasmid R64: global similarity of oriT structures of IncI1 and IncP plasmids. Journal of Bacteriology *173*, 6612–6617.

Furuya, N., and Komano, T. (2000). Initiation and Termination of DNA Transfer during Conjugation of IncI1 Plasmid R64: Roles of Two Sets of Inverted Repeat Sequences within oriT in Termination of R64 Transfer. Journal of Bacteriology *182*, 3191–3196.

Furuya, N., and Komano, T. (2003). NikAB- or NikB-Dependent Intracellular Recombination between Tandemly Repeated oriT Sequences of Plasmid R64 in Plasmid or Single-Stranded Phage Vectors. Journal of Bacteriology *185*, 3871–3877.

Garcillan-Barcia, M.P., and Francia, M.V. (2009). The diversity of conjugative relaxases and its application in plasmid classification. FEMS Microbiology .

Gautier, A., Juillerat, A., Heinis, C., Corrêa, I.R., Jr., Kindermann, M., Beaufils, F., and Johnsson, K. (2008). An Engineered Protein Tag for Multiprotein Labeling in Living Cells. Chemistry & Biology *15*, 128–136.

Gibb, B., Gupta, K., Ghosh, K., Sharp, R., Chen, J., and Van Duyne, G.D. (2010). Requirements for catalysis in the Cre recombinase active site. Nucleic Acids Research *38*, 5817–5832.

González-Pérez, B., Carballeira, J.D., Moncalián, G., and la Cruz, de, F. (2009). Changing the recognition site of a conjugative relaxase by rational design. Biotechnol. J. *4*, 554–557.

González-Pérez, B., Lucas, M., Cooke, L.A., Vyle, J.S., la Cruz, de, F., and Moncalián, G. (2007). Analysis of DNA processing reactions in bacterial conjugation by using suicide oligonucleotides. The EMBO Journal *26*, 3847–3857.

González-Prieto, C., Agúndez, L., Linden, R.M., and Llosa, M. (2013). HUH site-specific recombinases for targeted modification of the human genome. Trends in

Biotechnology *31*, 305–312.

Goodman, R.P., Erben, C.M., Malo, J., Ho, W.M., McKee, M.L., Kapanidis, A.N., and Turberfield, A.J. (2009). A facile method for reversibly linking a recombinant protein to DNA. Chem. Eur. J. of Chem. Bio. *10*, 1551–1557.

Gopaul, D.N., Guo, F., and Van Duyne, G.D. (1998). Structure of the Holliday junction intermediate in Cre-loxP site-specific recombination. The EMBO Journal *17*, 4175–4187.

Grandoso, G., Avila, P., Cayón, A., Hernando, M.A., Llosa, M., and la Cruz, de, F. (2000). Two active-site tyrosyl residues of protein TrwC act sequentially at the origin of transfer during plasmid R388 conjugation. Journal of Molecular Biology *295*, 1163–1172.

Green, N.M. (1990). Avidin and streptavidin. Meth. Enzymol. *184*, 51–67.

Guasch, A., Lucas, M., Moncalián, G., Cabezas, M., Pérez-Luque, R., Gomis-Rüth, F.X., la Cruz, de, F., and Coll, M. (2003). Recognition and processing of the origin of transfer DNA by conjugative relaxase TrwC. Nat Struct Biol *10*, 1002–1010.

Guja, K.E., and Schildbach, J.F. (2015). Completing the specificity swap: Single-stranded DNA recognition by F and R100 TraI relaxase domains. Plasmid 1–7.

Hagemeyer, C.E., Alt, K., Johnston, A.P.R., Such, G.K., Ta, H.T., Leung, M.K.M., Prabhu, S., Wang, X., Caruso, F., and Peter, K. (2014). Particle generation, functionalization and sortase A–mediated modification with targeting of single-chain antibodies for diagnostic and therapeutic use. Nat Protoc *10*, 90–105.

Hahn, J., Wickham, S.F.J., Shih, W.M., and Perrault, S.D. (2014). Addressing the instability of DNA nanostructures in tissue culture. ACS Nano *8*, 8765–8775.

Hanai, R., and Wang, J.C. (1993). The mechanism of sequence-specific DNA cleavage and strand transfer by phi X174 gene A* protein. J. Biol. Chem. *268*, 23830–23836.

Harley, M.J., and Schildbach, J.F. (2003). Swapping single-stranded DNA sequence specificities of relaxases from conjugative plasmids F and R100. Proc. Natl. Acad. Sci. U.S.a. *100*, 11243–11248.

He, Y., Tian, Y., Ribbe, A.E., and Mao, C. (2006). Antibody nanoarrays with a pitch of approximately 20 nanometers. J. Am. Chem. Soc. *128*, 12664–12665.

Heim, R., Prasher, D.C., and Tsien, R.Y. (1994). Wavelength mutations and posttranslational autoxidation of green fluorescent protein. Proc. Natl. Acad. Sci. U.S.a. *91*, 12501–12504.

Hermanson, G.T. (2008). Bioconjugate TEchniques (2nd) Second Edition (Elsevier).

Hickman, A.B., and Dyda, F. (2015). Mechanisms of DNA Transposition. Microbiol Spectr *3*, MDNA3–M2014.

Hickman, A.B., Ronning, D.R., Kotin, R.M., and Dyda, F. (2002). Structural unity among viral origin binding proteins: crystal structure of the nuclease domain of adeno-associated virus Rep. Molecular Cell *10*, 327–337.

Hickman, A.B., Ronning, D.R., Perez, Z.N., Kotin, R.M., and Dyda, F. (2004). The nuclease domain of adeno-associated virus rep coordinates replication initiation using two distinct DNA recognition interfaces. Molecular Cell.

Huang, N.-L., and Lin, J.-H. (2015). Recovery of the poisoned topoisomerase II for DNA religation: coordinated motion of the cleavage core revealed with the microsecond atomistic simulation. Nucleic Acids Research *43*, 6772–6786.

Huang, W.M., Joss, L., Hsieh, T., and Casjens, S. (2004). Protelomerase Uses a Topoisomerase IB/Y-Recombinase Type Mechanism to Generate DNA Hairpin Ends. Journal of Molecular Biology *337*, 77–92.

Ilyina, T.V., and Koonin, E.V. (1992). Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. Nucleic Acids Research *20*, 3279–3285.

Ke, Y., Lindsay, S., Chang, Y., Liu, Y., and Yan, H. (2008). Self-Assembled Water-Soluble Nucleic Acid Probe Tiles for Label-Free RNA Hybridization Assays. Science *319*, 180–183.

Keppler, A., Gendreizig, S., Gronemeyer, T., Pick, H., Vogel, H., and Johnsson, K. (2003). A general method for the covalent labeling of fusion proteins with small molecules in vivo. Nat. Biotechnol. *21*, 86–89.

Khan, S.A. (1997). Rolling-circle replication of bacterial plasmids. Microbiology and Molecular Biology Reviews *61*, 442–455.

Khan, S.A. (2002). DNA-protein interactions during the initiation and termination of plasmid pT181 rolling-circle replication. Prog Nucleic Acid Res Mol Biol *75*, 113–137.

Khan, S.A. (2005). Plasmid rolling-circle replication: highlights of two decades of research. Plasmid *53*, 126–136.

Koonin, E.V., and Ilyina, T.V. (1993). Computer-assisted dissection of rolling circle DNA replication. BioSystems *30*, 241–268.

Kopec, J., Bergmann, A., Fritz, G., Grohmann, E., and Keller, W. (2005). TraA and its N-terminal relaxase domain of the Gram-positive plasmid pIP501 show specific oriT binding and behave as dimers in solution. Biochem J *387*, 401–409.

Krishnan, Y., and Simmel, F.C. (2011). Nucleic Acid Based Molecular Devices. Angew. Chem. Int. Ed. *50*, 3124–3156.

Kurenbach, B., Kopec, J., Mägdefrau, M., Andreas, K., Keller, W., Bohn, C., Abajy, M.Y., and Grohmann, E. (2006). The TraA relaxase autoregulates the putative type IV secretion-like system encoded by the broad-host-range Streptococcus agalactiae plasmid pIP501. Microbiology (Reading, Engl.) *152*, 637–645.

Kuzyk, A., Laitinen, K.T., and Törmä, P. (2009). DNA origami as a nanoscale template for protein assembly. Nanotechnology *20*, 235305.

la Cruz, de, F., Frost, L.S., Meyer, R.J., and Zechner, E.L. (2010). Conjugative DNA metabolism in Gram-negative bacteria. FEMS Microbiology Reviews *34*, 18–40.

Lang, S., Gruber, K., Mihajlovic, S., Arnold, R., Gruber, C.J., Steinlechner, S., Jehl, M.-A., Rattei, T., Fröhlich, K.-U., and Zechner, E.L. (2010). Molecular recognition determinants for type IV secretion of diverse families of conjugative relaxases. Mol Microbiol *78*, 1539–1555.

Larkin, C., Datta, S., Harley, M.J., Anderson, B.J., Ebie, A., Hargreaves, V., and Schildbach, J.F. (2005). Inter- and Intramolecular Determinants of the Specificity of Single-Stranded DNA Binding and Cleavage by the F Factor Relaxase. Structure *13*, 1533–1544.

Lee, I., Dong, K.C., and Berger, J.M. (2013). The role of DNA bending in type IIA topoisomerase function. Nucleic Acids Research *41*, 5444–5456.

Li, J., Pei, H., Zhu, B., Liang, L., Wei, M., He, Y., Chen, N., Li, D., Huang, Q., and Fan, C. (2011). Self-assembled multivalent DNA nanostructures for noninvasive intracellular delivery of immunostimulatory CpG oligonucleotides. ACS Nano *5*, 8783–8789.

Lin, P.-H., Chen, R.-H., Lee, C.-H., Chang, Y., Chen, C.-S., and Chen, W.-Y. (2011). Studies of the binding mechanism between aptamers and thrombin by circular dichroism, surface plasmon resonance and isothermal titration calorimetry. Colloids and Surfaces B: Biointerfaces *88*, 552–558.

Liu, X., Xu, Y., Yu, T., Clifford, C., Liu, Y., Yan, H., and Chang, Y. (2012). A DNA nanostructure platform for directed assembly of synthetic vaccines. Nano Lett. *12*, 4254–4259.

Llosa, M., Bolland, S., and la Cruz, de, F. (1991). Structural and functional analysis of the origin of conjugal transfer of the broad-host-range IncW plasmid R388 and comparison with the related IncN plasmid R46. Mol. Gen. Genet. *226*, 473–483.

Llosa, M., Gomis-Rüth, F.X., Coll, M., and la Cruz Fd, de, F. (2002). Bacterial conjugation: a two-step mechanism for DNA transport. Mol Microbiol *45*, 1–8.

Llosa, M., Grandoso, G., and la Cruz, de, F. (1994). Nicking activity of TrwC directed against the origin of transfer of the IncW plasmid R388. Journal of Molecular Biology *246*, 54–62.

Lorenzo-Diaz, F., Dostal, L., Coll, M., Schildbach, J.F., Menendez, M., and Espinosa, M. (2011). The MobM relaxase domain of plasmid pMV158: thermal stability and activity upon Mn2+ and specific DNA binding. Nucleic Acids Research *39*, 4315–4329.

Los, G.V., Encell, L.P., McDougall, M.G., Hartzell, D.D., Karassina, N., Zimprich, C., Wood, M.G., Learish, R., Ohana, R.F., Urh, M., et al. (2008). HaloTag: a novel protein labeling technology for cell imaging and protein analysis. ACS Chem. Biol. *3*, 373–382.

Lovrinovic, M., and Niemeyer, C.M. (2005). Rapid synthesis of DNA-cysteine conjugates for expressed protein ligation. Biochemical and Biophysical Research Communications *335*, 943–948.

Lucas, M., Gonzalez-Perez, B., Cabezas, M., Moncalian, G., Rivas, G., and la Cruz, de, F. (2010). Relaxase DNA Binding and Cleavage Are Two Distinguishable Steps in Conjugative DNA Processing That Involve Different Sequence Elements of the nic Site. Journal of Biological Chemistry *285*, 8918–8926.

Lujan, S.A., Guogas, L.M., Ragonese, H., Matson, S.W., and Redinbo, M.R. (2007). Disrupting antibiotic resistance propagation by inhibiting the conjugative DNA relaxase. Proc. Natl. Acad. Sci. U.S.a. *104*, 12282–12287.

Lyubchenko, Y.L., and Shlyakhtenko, L.S. (2009). AFM for analysis of structure and dynamics of DNA and protein–DNA complexes. Methods *47*, 206–213.

Malo, G.D., Pouwels, L.J., Wang, M., Weichsel, A., Montfort, W.R., Rizzo, M.A., Piston, D.W., and Wachter, R.M. (2007). X-ray structure of Cerulean GFP: a tryptophan-based chromophore useful for fluorescence lifetime imaging. Biochemistry *46*, 9865–9873.

Mashimo, Y., Maeda, H., Mie, M., and Kobatake, E. (2012). Construction of semisynthetic DNA-protein conjugates with Phi X174 Gene-A* protein. Bioconjugate Chem. *23*, 1349–1355.

McKay, C.S., and Finn, M.G. (2014). Click Chemistry in Complex Mixtures: Bioorthogonal Bioconjugation. Chemistry & Biology *21*, 1075–1101.

Mei, Q., Wei, X., Su, F., Liu, Y., Youngbull, C., Johnson, R., Lindsay, S., Yan, H., and Meldrum, D. (2011). Stability of DNA Origami Nanoarrays in Cell Lysate. Nano Lett. *11*, 1477–1482.

Mizuno, H., Mal, T.K., Tong, K.I., Ando, R., Furuta, T., Ikura, M., and Miyawaki, A. (2003). Photo-induced peptide cleavage in the green-to-red conversion of a fluorescent protein. Molecular Cell *12*, 1051–1058.

Mohri, K., Nishikawa, M., Takahashi, N., Shiomi, T., Matsuoka, N., Ogawa, K., Endo, M., Hidaka, K., Sugiyama, H., Takahashi, Y., et al. (2012). Design and development of nanosized DNA assemblies in polypod-like structures as efficient vehicles for immunostimulatory CpG motifs to immune cells. ACS Nano *6*, 5931–5940.

Mok, W., and Li, Y. (2008). Recent Progress in Nucleic Acid Aptamer-Based Biosensors and Bioassays. Sensors *8*, 7050–7084.

Moncalián, G., and la Cruz, de, F. (2004). DNA binding properties of protein TrwA, a possible structural variant of the Arc repressor superfamily. Biochimica Et Biophysica Acta (BBA) - Proteins and Proteomics *1701*, 15–23.

Moncalián, G., Valle, M., Valpuesta, J.M., and la Cruz, de, F. (1999). IHF protein inhibits cleavage but not assembly of plasmid R388 relaxosomes. Mol Microbiol *31*, 1643–1652.

Monzingo, A.F., Ozburn, A., Xia, S., Meyer, R.J., and Robertus, J.D. (2007). The

Structure of the Minimal Relaxase Domain of MobA at 2.1 Å Resolution. Journal of Molecular Biology *366*, 165–178.

Mysore, K.S., Bassuner, B., Deng, X.B., Darbinian, N.S., Motchoulski, A., Ream, W., and Gelvin, S.B. (1998). Role of the Agrobacterium tumefaciens VirD2 protein in T-DNA transfer and integration. Mol. Plant Microbe Interact. *11*, 668–683.

Nakata, E., Liew, F.F., Uwatoko, C., Kiyonaka, S., Mori, Y., Katsuda, Y., Endo, M., Sugiyama, H., and Morii, T. (2012). Zinc-Finger Proteins for Site-Specific Protein Positioning on DNA-Origami Structures. Angew. Chem. Int. Ed. *51*, 2421–2424.

Nash, R.P., Habibi, S., Cheng, Y., Lujan, S.A., and Redinbo, M.R. (2010). The mechanism and control of DNA transfer by the conjugative relaxase of resistance plasmid pCU1. Nucleic Acids Research *38*, 5929–5943.

Noirot-Gros, M.F., and Ehrlich, S.D. (1996). Change of a catalytic reaction carried out by a DNA replication protein. Science *274*, 777–780.

Noirot-Gros, M.F., Bidnenko, V., and Ehrlich, S.D. (1994). Active site of the replication protein of the rolling circle plasmid pC194. The EMBO Journal *13*, 4412–4420.

Odegrip, R., and Haggård-Ljungquist, E. (2001). The two active-site tyrosine residues of the A protein play non-equivalent roles during initiation of rolling circle replication of bacteriophage P2. Journal of Molecular Biology *308*, 147–163.

Pansegrau, W., and Lanka, E. (1996). Mechanisms of initiation and termination reactions in conjugative DNA processing. Independence of tight substrate binding and catalytic activity of relaxase (TraI) of IncPalpha plasmid RP4. J. Biol. Chem. *271*, 13068–13076.

Pansegrau, W., Balzer, D., Kruft, V., Lurz, R., and Lanka, E. (1990). In vitro assembly of relaxosomes at the transfer origin of plasmid RP4. Proc. Natl. Acad. Sci. U.S.a. *87*, 6555–6559.

Pansegrau, W., Schröder, W., and Lanka, E. (1993). Relaxase (TraI) of IncP alpha plasmid RP4 catalyzes a site-specific cleaving-joining reaction of single-stranded DNA. Proc. Natl. Acad. Sci. U.S.a. *90*, 2925–2929.

Pansegrau, W., Schröder, W., and Lanka, E. (1994). Concerted action of three distinct domains in the DNA cleaving-joining reaction catalyzed by relaxase (TraI) of conjugative plasmid RP4. J. Biol. Chem. *269*, 2782–2789.

Parker, C., and Meyer, R.J. (2007). The R1162 relaxase/primase contains two, type IV transport signals that require the small plasmid protein MobB. Mol Microbiol *66*, 252–261.

Pedersen, R.O., Loboa, E.G., and LaBean, T.H. (2013). Sensitization of transforming growth factor-β signaling by multiple peptides patterned on DNA nanostructures. Biomacromolecules *14*, 4157–4160.

Perwez, T., and Meyer, R. (1996). MobB protein stimulates nicking at the R1162 origin of transfer by increasing the proportion of complexed plasmid DNA. Journal of Bacteriology *178*, 5762–5767.

Pletnev, S., Shcherbo, D., Chudakov, D.M., Pletneva, N., Merzlyak, E.M., Wlodawer, A., Dauter, Z., and Pletnev, V. (2008). A Crystallographic Study of Bright Far-Red Fluorescent Protein mKate Reveals pH-induced cis-trans Isomerization of the Chromophore. Journal of Biological Chemistry *283*, 28980–28987.

Redinbo, M.R., Stewart, L., Kuhn, P., Champoux, J.J., and Hol, W.G. (1998). Crystal structures of human topoisomerase I in covalent and noncovalent complexes with DNA. Science *279*, 1504–1513.

Redzej, A., Ilangovan, A., Lang, S., Gruber, C.J., Topf, M., Zangger, K., Zechner, E.L., and Waksman, G. (2013). Structure of a translocation signal domain mediating conjugative transfer by type IV secretion systems. Mol Microbiol *89*, 324–333.

Ritacco, C.J., Kamtekar, S., Wang, J., and Steitz, T.A. (2013). Crystal structure of an intermediate of rotating dimers within the synaptic tetramer of the G-segment invertase. Nucleic Acids Research *41*, 2673–2682.

Rothemund, P.W.K. (2006). Folding DNA to create nanoscale shapes and patterns. Nature *440*, 297–302.

Saccà, B., and Niemeyer, C.M. (2011). Functionalization of DNA nanostructures with proteins. Chem. Soc. Rev. *40*, 5910–5921.

Sachdeva, G., Garg, A., Godding, D., Way, J.C., and Silver, P.A. (2014). In vivo co-localization of enzymes on RNA scaffolds increases metabolic production in a geometrically dependent manner. Nucleic Acids Research *42*, 9493–9503.

Saikrishnan, K., Griffiths, S.P., Cook, N., Court, R., and Wigley, D.B. (2008). DNA binding to RecD: role of the 1B domain in SF1B helicase activity. The EMBO Journal *27*, 2222–2229.

Sanderson, K. (2010). Bioengineering: What to make with DNA origami. Nature *464*, 158–159.

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

Schmidt, B.H., Burgin, A.B., Deweese, J.E., Osheroff, N., and Berger, J.M. (2010). A novel and unified two-metal mechanism for DNA cleavage by type II and IA topoisomerases. Nature *465*, 641–644.

Schüller, V.J., Heidegger, S., Sandholzer, N., Nickels, P.C., Suhartha, N.A., Endres, S., Bourquin, C., and Liedl, T. (2011). Cellular immunostimulation by CpG-sequence-coated DNA origami structures. ACS Nano *5*, 9696–9702.

Seeman, N.C. (1982). Nucleic acid junctions and lattices. J Theor Biol *99*, 237–247.

Seeman, N.C. (2010). Structural DNA Nanotechnology: Growing Along with Nano Letters. Nano Lett. *10*, 1971–1978.

Seeman, N.C., and Kallenbach, N.R. (1983). Design of immobile nucleic acid junctions. Biophysical Journal *44*, 201–209.

Selmi, D.N., Adamson, R.J., Attrill, H., Goddard, A.D., Gilbert, R.J.C., Watts, A., and Turberfield, A.J. (2011). DNA-templated protein arrays for single-molecule imaging. Nano Lett. *11*, 657–660.

Sharma, J., Chhabra, R., Cheng, A., Brownell, J., Liu, Y., and Yan, H. (2009). Control of Self-Assembly of DNA Tubules Through Integration of Gold Nanoparticles. Science *323*, 112–116.

Shaw, A., Lundin, V., Petrova, E., s, F.F.O.R.O., Benson, E., Al-Amin, A., Herland, A., Blokzijl, A., gberg, B.O.R.H.O., and Teixeira, A.I. (2014). Spatial control of membrane receptor function using ligand nanocalipers. Nat Meth *11*, 841–846.

Shu, X., Shaner, N.C., Yarbrough, C.A., Tsien, R.Y., and Remington, S.J. (2006). Novel Chromophores and Buried Charges Control Color in mFruits †,‡. Biochemistry *45*, 9639–9647.

Simmel, F.C. (2012). DNA-based assembly lines and nanofactories. Current Opinion in Biotechnology *23*, 516–521.

Sletten, E.M., and Bertozzi, C.R. (2009). Bioorthogonal Chemistry: Fishing for Selectivity in a Sea of Functionality. Angew. Chem. Int. Ed. *48*, 6974–6998.

Smith, S.S., Niu, L., Baker, D.J., Wendel, J.A., Kane, S.E., and Joy, D.S. (1997). Nucleoprotein-based nanoscale assembly. Proc. Natl. Acad. Sci. U.S.a. *94*, 2162–2167.

Stahl, E., Martin, T.G., Praetorius, F., and Dietz, H. (2014). Facile and scalable preparation of pure and dense DNA origami solutions. Angew. Chem. Int. Ed. *53*, 12735–12740.

Stewart, L. (1998). A Model for the Mechanism of Human Topoisomerase I. Science *279*, 1534–1541.

Sun, H., and Zu, Y. (2015). Aptamers and Their Applications in Nanomedicine. Small *11*, 2352–2364.

Sut, M.V., Mihajlovic, S., Lang, S., Gruber, C.J., and Zechner, E.L. (2009). Protein and DNA Effectors Control the TraI Conjugative Helicase of Plasmid R1. Journal of Bacteriology *191*, 6888–6899.

Sørensen, R.S., Okholm, A.H., Schaffert, D., Kodal, A.L.B., Gothelf, K.V., and Kjems, J. (2013). Enzymatic ligation of large biomolecules to DNA. ACS Nano *7*, 8098–8104.

Tessmer, I., Kaur, P., Lin, J., and Wang, H. (2012). Investigating bioconjugation by atomic force microscopy. J Nanobiotechnology *11*, 25–25.

Tewary, S.K., Liang, L., Lin, Z., Lynn, A., Cotmore, S.F., Tattersall, P., Zhao, H., and Tang, L. (2015). Structures of minute virus of mice replication initiator protein N-terminal domain: Insights into DNA nicking and origin binding. Virology *476*, 61–71.

Thompson, R., Taylor, L., Kelly, K., Everett, R., and Willetts, N. (1984). The F plasmid origin of transfer: DNA sequence of wild-type and mutant origins and location of origin-specific nicks. The EMBO Journal *3*, 1175–1180.

Tintoré, M., Gállego, I., Manning, B., Eritja, R., and Fàbrega, C. (2013). DNA origami as a DNA repair nanosensor at the single-molecule level. Angew. Chem. Int. Ed. Engl. *52*, 7747–7750.

Tominaga, J., Kemori, Y., Tanaka, Y., Maruyama, T., Kamiya, N., and Goto, M. (2007). An enzymatic method for site-specific labeling of recombinant proteins with oligonucleotides. Chem. Commun. 401–403.

Turan, S., and Bode, J. (2011). Site-specific recombinases: from tag-and-target- to tag-and-exchange-based genomic modifications. The FASEB Journal *25*, 4088–4107.

van Berkel, S.S., van Eldijk, M.B., and van Hest, J.C.M. (2011). Staudinger ligation as a method for bioconjugation. Angew. Chem. Int. Ed. Engl. *50*, 8806–8827.

van Kregten, M., Lindhout, B.I., Hooykaas, P.J.J., and van der Zaal, B.J. (2009). Agrobacterium-mediated T-DNA transfer and integration by minimal VirD2 consisting of the relaxase domain and a type IV secretion system translocation signal. Mol. Plant Microbe Interact. *22*, 1356–1365.

Varsaki, A., Moncalián, G., Garcillán-Barcia, M.D.P., Drainas, C., and la Cruz, de, F. (2009). Analysis of ColE1 MbeC unveils an extended ribbon-helix-helix family of nicking accessory proteins. Journal of Bacteriology *191*, 1446–1455.

Voigt, N.V., Tørring, T., Rotaru, A., Jacobsen, M.F., Ravnsbæk, J.B., Subramani, R., Mamdouh, W., Kjems, J., Mokhir, A., Besenbacher, F., et al. (2010). Single-molecule chemical reactions on DNA origami. Nature Nanotech *5*, 200–203.

Vos, S.M., Tretter, E.M., Schmidt, B.H., and Berger, J.M. (2011). All tangled up: how cells direct,manage and exploit topoisomerasefunction. Nature Publishing Group *12*, 827–841.

Wendorff, T.J., Schmidt, B.H., Heslop, P., Austin, C.A., and Berger, J.M. (2012). The Structure of DNA-Bound Human Topoisomerase II Alpha: Conformational Mechanisms for Coordinating Inter-Subunit Interactions with DNA Cleavage. Journal of Molecular Biology *424*, 109–124.

Winfree, E., Liu, F., Wenzler, L.A., and Seeman, N.C. (1998). Design and self-assembly of two-dimensional DNA crystals. Nature *394*, 539–544.

Yang, W. (2008). An equivalent metal ion in one- and two-metal-ion catalysis. Nat Struct Mol Biol *15*, 1228–1231.

Yang, Y.R., Liu, Y., and Yan, H. (2015). DNA Nanostructures as Programmable Biomolecular Scaffolds. Bioconjugate Chem.

Zhan, Z., Ouyang, S., Liang, W., Zhang, Z., Liu, Z.J., and Huang, L. (2012). research papers. Acta Cryst (2012). D68, 659-670 [Doi:10.1107/S0907444912007202] 1–12.

Zhang, Q., Jiang, Q., Li, N., Dai, L., Liu, Q., Song, L., Wang, J., Li, Y., Tian, J., Ding, B., et al. (2014). DNA origami as an in vivo drug delivery vehicle for cancer therapy. ACS Nano *8*, 6633–6643.

Zhao, A.C., Ansari, R.A., Schmidt, M.C., and Khan, S.A. (1998). An oligonucleotide inhibits oligomerization of a rolling circle initiator protein at the pT181 origin of replication. J. Biol. Chem. *273*, 16082–16089.

Zhao, Y.-X., Shaw, A., Zeng, X., Benson, E., Nyström, A.M., and Högberg, B. (2012). DNA Origami Delivery System for Cancer Therapy with Tunable Release Properties. ACS Nano *6*, 8684–8691.

PUBLICATIONS

DNA Nanotechnology

International Edition: DOI: 10.1002/anie.201510313
German Edition: DOI: 10.1002/ange.201510313

Orthogonal Protein Assembly on DNA Nanostructures Using Relaxases

Sandra Sagredo, Tobias Pirzer, Ali Aghebat Rafat, Marisa A. Goetzfried, Gabriel Moncalian, Friedrich C. Simmel, and Fernando de la Cruz**

Abstract: *DNA-binding proteins are promising reagents for the sequence-specific modification of DNA-based nanostructures. Here, we investigate the utility of a series of relaxase proteins—TrwC, TraI, and MobA—for nanofunctionalization. Relaxases are involved in the conjugative transfer of plasmids between bacteria, and bind to their DNA target sites via a covalent phosphotyrosine linkage. We study the binding of the relaxases to two standard DNA origami structures—rodlike six-helix bundles and flat rectangular origami sheets. We find highly orthogonal binding of the proteins with binding yields of 40–50 % per binding site, which is comparable to other functionalization methods. The yields differ for the two origami structures and also depend on the position of the binding sites. Due to their specificity for a single-stranded DNA target, their orthogonality, and their binding properties, relaxases are a uniquely useful addition to the toolbox available for the modification of DNA nanostructures with proteins.*

Recent advances in DNA nanotechnology have resulted in methodologies that enable the generation of molecular objects made from DNA with almost any user-defined shape,[1] and these structures also facilitate the arrangement of functional molecular components or nanoparticles (e.g., proteins) into three-dimensional configurations with nanometer precision. One of the most popular methods is the DNA origami technique, which utilizes hundreds of short DNA oligomers, called “staples” to fold a long DNA scaffold into a desired shape.[1a,c]

A variety of approaches were previously utilized for the functionalization of DNA nanostructures with proteins. An often used modification is streptavidin, which binds to biotin with a dissociation constant of $K_D = 40$ fM,[2] constituting one of the strongest noncovalent bonds known in nature. In order to facilitate the binding of streptavidin to DNA nanostructures, biotinylated oligonucleotide staples are incorporated into the structures at the desired locations. The effective yield for streptavidin-binding on origami structures is roughly 85 % per binding site.[3] Naturally occurring DNA-binding proteins typically show dissociation constants in the nanomolar range and thus could also be used for modification of DNA nanostructures. A good example are zinc finger proteins, which bind to DNA in the presence of Zn^{2+} and result in effective binding yields of 30–70 % for DNA nanostructures.[4] Other proteins that have been used in DNA origami assemblies are DNA–antibody conjugates[5] and engineered proteins containing His-,[6] Snap-, and Halo-tags.[7] Other covalent conjugation techniques such as coupling by transglutaminase[8] or methyltransferases,[9] expressed protein ligation, and enzymatic ligation[10] also provide site-specific conjugation methods with controlled stoichiometry. Alternatively, chemical crosslinkers containing maleimide or *N*-hydroxysuccinimide groups are extensively used for nonspecific covalent coupling (for a more complete overview see Ref. [11]). The disadvantage of most of the aforementioned methods is that they are laborious and that the structure and thereby the activity of the proteins is compromised by the introduction of nonnative chemical tags.

In this work we present a novel strategy for DNA coupling based on relaxase proteins. Relaxases are involved in DNA processing for bacterial conjugation,[12] the horizontal transfer of genomic material among bacteria. They are single-strand (ss) DNA cleaving endonucleases, which naturally remain covalently bound to DNA after cleavage. Thus, relaxases offer a potentially superior alternative to existing methods for protein conjugation to DNA nanostructures, due to their sequence specificity, orthogonality and also their natural capacity to form covalent bonds with standard oligonucleotides. Relaxases belong to the HUH protein superfamily,[13] and the three-dimensional structure of some relaxases is known.[14] Their overall shape resembles a left hand, where the β-sheet “palm” holds the HUH motif (histidine–hydrophobic amino acid–histidine) responsible for metal coordination, and the α-helix “thumb” contains the catalytic tyrosine. These conserved residues form the active center of each relaxase and allow cleavage of a phosphodiester bond located at a specific plasmid DNA sequence called the *nic* site. Each plasmid contains a particular relaxase, recognizing a different

---

[*] S. Sagredo, Dr. G. Moncalian, Prof. Dr. F. de la Cruz
Departamento de Biología Molecular e Instituto de Biomedicina y Biotecnología de Cantabria, Universidad de Cantabria-Consejo Superior de Investigaciones Científicas-SODERCAN
Albert Einstein 22, 39011 Santander (Spain)
E-mail: delacruz@unican.es

Dr. T. Pirzer, A. Aghebat Rafat, M. A. Goetzfried,
Prof. Dr. F. C. Simmel
Physik-Department E14 and ZNN/WSI, TU Munich
Am Coulombwall 4a, 85748 Garching (Germany)
E-mail: simmel@tum.de

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under http://dx.doi.org/10.1002/anie.201510313.

nic site, so the number of relaxase/*nic* site pairs is enormous. The cleavage reaction also occurs in vitro on ssDNA containing the *nic* region.[15] The *nic* region contains an inverted repeat (IR) and an AT-rich region downstream from *nic*. Upon binding ssDNA in vitro, the relaxase bends its substrate into a U shape and fits the *nic* sequence into its active center; here, the tyrosine generates a covalent 5′-phosphotyrosine intermediate and a free 3′-OH end (cf. Figure 1). In this state, cleavage and religation of substrate are in fact in a dynamic equilibrium.[15b,16] Each relaxase recognizes its own *nic* region with a K_D between 1 nM and 400 nM, depending on the individual protein and the target sequence.

In the following, we utilize the N-terminal domain of relaxases TrwC from conjugative plasmid R388 (denoted as TrwC$_{R388}$), TraI from R100 (TraI$_{R100}$), MobA from R1162 (MobA$_{R1162}$), and TraI from pKM101 (TraI$_{pKM101}$). All relaxase domains are small monomeric proteins (20–30 kDa) with low K_D[15a,17] that can be engineered and fused to other proteins such as fluorescent proteins (Figures S6, S7, and S9). TraI$_{R100}$ and TrwC$_{R388}$ are able to cleave proficiently the *nic* region lacking the distal arm of the IR. The reported K_d values of single-stranded R100 and R388 *nic* regions are 2.4 nM[17a] and 320 nM,[17b] respectively. It was shown that MobA$_{R1162}$ effectively binds targets with the entire IR, but binding was compromised after deletion of the distal arm.[15a] The relaxase TraI$_{pKM101}$ had not been characterized yet, but we expected a similar behavior as it has 51 % and 37.7 % sequence identity with TrwC$_{R388}$ and TraI$_{R100}$, respectively. Moreover, the pKM101 target is 64.7 % identical to the R100 target. Comparison of targets with and without a complete IR showed that in the pKM101 target a complete IR is also needed (cf. Figure S8).

DNA origami structures used in this work were long, rodlike six-helix bundles (6HB) or flat, twist-corrected rectangular origami sheets (tcRO), with dimensions of 430 nm × 5 nm and 90 nm × 60 nm, respectively. The binding sites with the target sequence were designed with a double-stranded stem as a rigid linker in close proximity to the nanostructure followed by a single-stranded region with the *nic* site (cf. Figures 1 A and S10 for further details). Along its length, each 6HB is equipped with specific binding positions for the aforementioned four relaxases. In order to increase the local concentration of binding sites, each binding *position* (see Figure 1 B and the Supporting Information) was equipped with five binding *sites* each for the specific relaxase. Due to the small size of the relaxase domains, we used fusion proteins of relaxases with fluorescent proteins mKATE or mCFP (45–62 kDa). These fused proteins resulted in increased visibility in TEM images. Each relaxase was incubated for 24 h with 6HB nanotubes containing all four binding positions. After incubation and purification, the number of occupied binding positions was determined from TEM images (Figure 2). The resulting binding yields *per position* were 93 % for TraI$_{R100m}$mKATE, 84 % for TraI$_{pKM101}$mCFP, 73 % for TrwC$_{R388}$mCFP, and 34 % for MobA$_{R1162}$mCFP.

Since the 6HB contained all four binding positions, quantitative values for the specificity of each relaxase to its target could be measured (see Figure 2). Proteins TraI$_{R100}$mKATE, TrwC$_{R388}$mCFP, and MobA$_{R1162}$mCFP show highly selective binding abilities, with unspecific binding yields ranging from 0 % to 7.1 % (cf. Figure 2). Despite its high yield for its specific target, TraI$_{pKM101}$mCFP shows significant nonspecific binding and therefore was not used simultaneously with the other relaxases in the following.

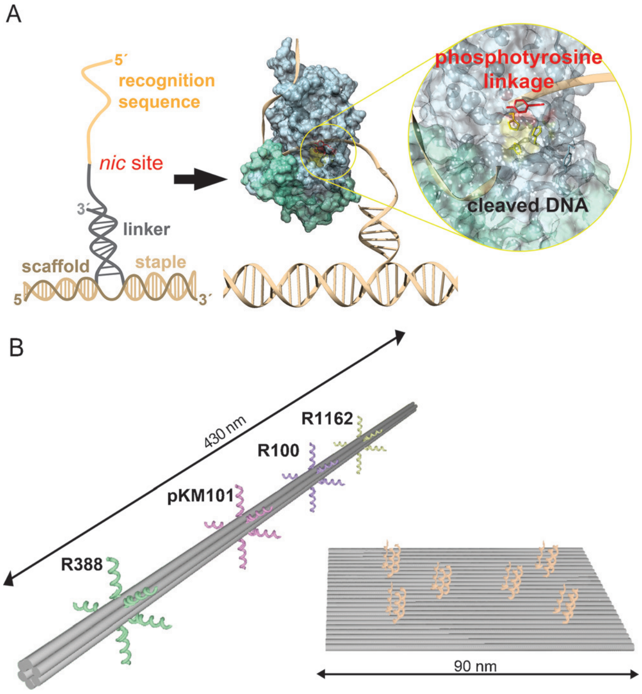


Figure 1. Strategy for the coupling of relaxases onto DNA nanostructures. A) Extended staples contain the recognition sequence of a relaxase and several base pairs that allow the target to protrude from the nanostructure. Binding of the relaxase to its cognate site involves bending of the ssDNA to fit the *nic* site into the relaxase active center. Inset: the catalytic tyrosine carries out a transesterification reaction that links it covalently to the 5′-phosphate between the scissile nucleotides. As a result, the relaxase becomes covalently bound to the extended staple. B) DNA nanostructures used in this approach to study the coupling of four relaxases.

In further experiments we used the twist-corrected rectangular origami sheet tcRO, for which we designed six binding positions with up to three redundant binding sites each (cf. Figure 3 A). We first compared the binding of the orthogonal relaxases TraI$_{R100}$mKATE and TrwC$_{R388}$mCFP onto the origami sheets. To test their binding yields, only position 1 for TraI$_{R100}$mKATE and position 5 for TrwC$_{R388}$mCFP, both equivalently at the rim of the tcRO next to the corners (cf. Figure 3), were used. Each protein was incubated for 24 h with the tcRO, each sheet containing two binding sites per position for each protein, and purified afterwards (see the Methods section in the Supporting Information for details). For visualization of the bound proteins we used atomic force microscopy (AFM) imaging (Figure 3). The binding yield of TraI$_{R100}$mKATE was found to be 70 %, and 48 % for TrwC$_{R388}$mCFP. Nonspecific binding to the unwanted position was 1 % for TraI$_{R100}$mKATE and 3 % for TrwC$_{R388}$mCFP.

We then checked for a potential position dependence of relaxase binding. To this end, the tcRO was equipped with all

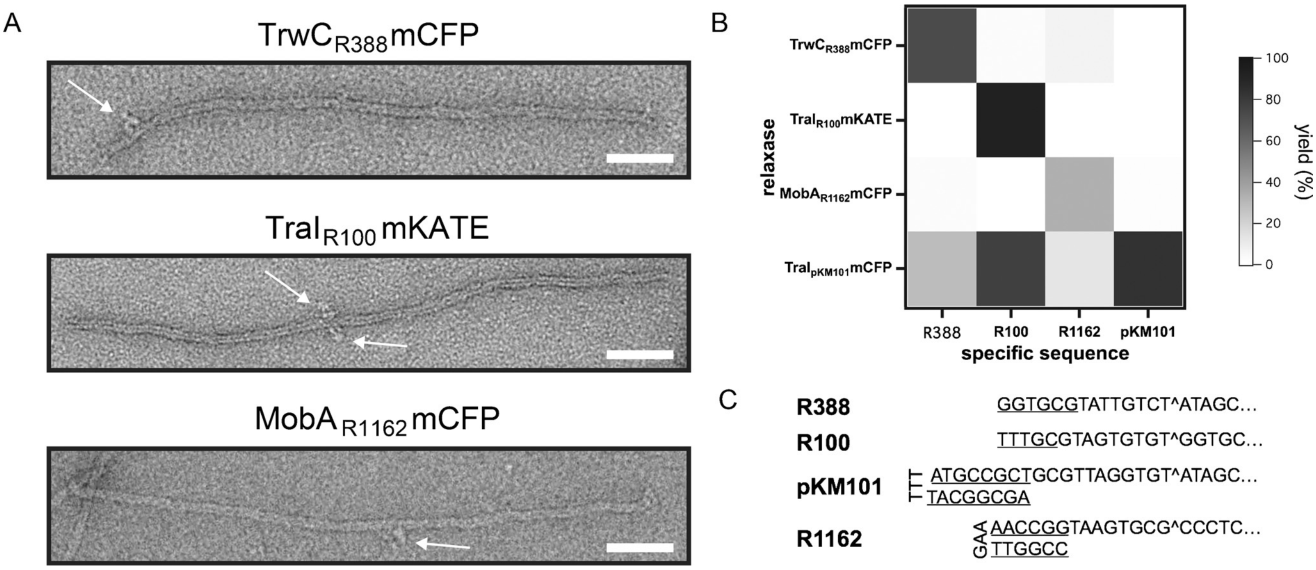


Figure 2. Orthogonal binding of relaxases to six-helix bundles (6HB). A) TEM images of the relaxases TrwC$_{R388}$mCFP, TraI$_{R100}$mKATE, and MobA$_{R1162}$mCFP bound to 6HB. Scale bar: 50 nm. B) Graphical representation of total binding yields and their orthogonality. For details on the statistical analysis see the Supporting Information. C) Recognition sequences of the relaxases used in this study. The *nic* site is depicted with ^ and the proximal arm of the inverted repeat is underlined.

designed positions containing only R388 binding sites. From these experiments, the binding yield at the tcRO rim (positions 1, 2, 5, and 6 in Figure 3 A) was found to range between 40 % and 60 %, whereas it appeared to be significantly higher (between 60 % and 90 %) in the tcRO center (positions 3 and 4, cf. Figure 3 C). In general, binding yields varied considerably from experiment to experiment (± 10 %). Surprisingly, within the range of this variation, we found no clear dependence of the binding yield on the number of binding sites per binding position. When only a single binding site for R388 was used, binding yields were as high as for multiple binding sites (between 40–50 %, see the Supporting Information).

Overall, the binding yields obtained by AFM were significantly lower than that expected based on the law of mass action (see the Supporting Information for details). For instance, in previous experiments with single-stranded targets containing the *nic* site, the K_D for the binding of TrwC$_{R388}$ to its recognition sequence was determined to be 320 nM.[17b] This should translate into an *equilibrium* binding yield of 86 % for a single binding site and even 99.7 % for three binding sites under our incubation conditions, where TrwC$_{R388}$ and tcRO concentrations were ca. 2 μM and 100 nM, respectively. As AFM imaging is performed with purified samples in buffer (i.e., without relaxases), however, re-equilibration of the sample will result in unbinding of the proteins from the target and thus a reduced apparent binding yield. This is essentially unavoidable for relatively high K_D values in the nanomolar range.

Since the covalent phosphotyrosine linkage of a relaxase to its *nic* site is in dynamic equilibrium with the unbound state, we also used phosphorothioate suicide oligonucleotides to shift the equilibrium towards the covalent bond (and thus achieve a much lower effective K_D). In suicide oligonucleotides, the 3′ oxygen is replaced by sulfur at the scissile phosphodiester bond, preventing the religation reaction.[18] Surprisingly, the binding yield did not improve with these oligonucleotides (see the Supporting Information for details). The strong coupling theoretically expected for a covalent bond is thus not reflected in the actual binding yields.

Our findings are in line with results for other protein-modification approaches for DNA origami. Streptavidin binding to biotin showed only about 85 %[3] binding yield (with a theoretical value of close to 100 %) and also for covalent coupling methods only 84–90 % experimental binding yields were obtained.[3] Apart from the re-equilibration issue mentioned above, binding yields on origami structures may be reduced by a variety of other factors. Some target sites may be inaccessible or simply absent (e.g., because of missing staples), and diffusion to the target may be affected by the geometry of the binding site or the overall shape and flexibility of the underlying origami structure. For instance, in our experiments with six-helix bundles, the binding sites were accessible from all sides, and we typically observed higher binding yields than for the tcRO structures.

In summary, we have introduced DNA-binding relaxase proteins as promising reagents for the sequence-addressable modification of DNA nanostructures with proteins. Binding yields on DNA origami structures ranged between 40–50 % for a single binding site, and were thus comparable to those of other modification strategies. Depending on the origami geometry, the specific relaxase, and also the utilization of redundant binding sites in close proximity, in some cases apparent binding yields as high as 90 % could be achieved. Upon further optimization it should be possible to fully exploit the covalent phosphotyrosine linkage between the relaxases and the DNA target site, and thus achieve even higher binding yields. At least two of the four investigated

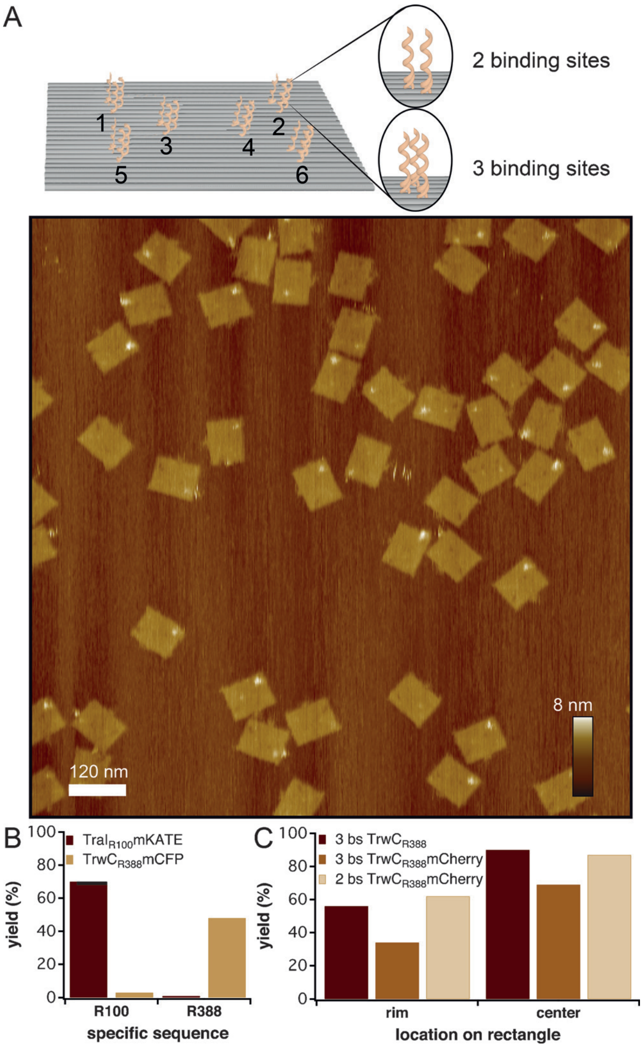


Figure 3. Position-dependent binding of relaxases to the rectangular DNA sheet tcRO. A) Top: Illustration of the six binding positions with two or three binding sites each. Bottom: AFM image of the nanostructures decorated with TraIR100mKATE at position 1. B) Binding yields of TrwCR388mCFP and TraIR100mKATE to available sites. The relaxases bind preferably to their specifc target. C) Binding yield of TrwCR388 to tcRO equipped with two and three targets per position. Binding positions at the rim of the tcRO (1, 2, 5, and 6) show lower yields than in the center (3 and 4). For details on the statistical analysis see the Supporting Information.

relaxases (TraI from plasmid R100 and TrwC from R388) showed very good orthogonal binding performance. Relaxases can be easily engineered and fused to other proteins of choice. As many relaxases from different sources and with unique recognition sequences are available, they potentially can be developed into a whole new class of orthogonal, sequence-selective protein linkers for DNA nanotechnology. Of particular interest could be their application for the biological generation of DNA–protein hybrid nanostructures and for the modification of DNA nanostructures in vivo.

Acknowledgements

We gratefully acknowledge financial support through the Volkswagen Foundation (grant 86 395-1), the European Commission (FP7 grant no. 248919 (BACTOCOM)), the DFG (SFB 1032 TP A2 and Nanosystems Initiative Munich), and EMBOEMBO. We thank F. Praetorius for providing the M13 origami scaffold and A. Kuzyk for initial work on this project.

Keywords: bacterial conjugation · bioconjugation · DNA nanotechnology · DNA origami · relaxase

How to cite: *Angew. Chem. Int. Ed.* **2016**, *55*, 4348–4352
Angew. Chem. **2016**, *128*, 4421–4425

---

[1] a) P. W. K. Rothemund, *Nature* **2006**, *440*, 297–302; b) S. M. Douglas, H. Dietz, T. Liedl, B. Högberg, F. Graf, W. M. Shih, *Nature* **2009**, *459*, 414–418; c) B. Kick, F. Praetorius, H. Dietz, D. Weuster-Botz, *Nano Lett.* **2015**, *15*, 4672–4676.

[2] N. M. Green, *Methods Enzymol.* **1990**, *184*, 51–67.

[3] N. V. Voigt, T. Tørring, A. Rotaru, M. F. Jacobsen, J. B. Ravnsbæk, R. Subramani, W. Mamdouh, J. Kjems, A. Mokhir, F. Besenbacher, K. V. Gothelf, *Nat. Nanotechnol.* **2010**, *5*, 200–203.

[4] E. Nakata, F. F. Liew, C. Uwatoko, S. Kiyonaka, Y. Mori, Y. Katsuda, M. Endo, H. Sugiyama, T. Morii, *Angew. Chem. Int. Ed.* **2012**, *51*, 2421–2424; *Angew. Chem.* **2012**, *124*, 2471–2474.

[5] a) T. Yamazaki, J. G. Heddle, A. Kuzuya, M. Komiyama, *Nanoscale* **2014**, *6*, 9122–9126; b) A. Shaw, E. Benson, B. Högberg, *ACS Nano* **2015**, *9*, 4968–4975.

[6] a) R. P. Goodman, C. M. Erben, J. Malo, W. M. Ho, M. L. McKee, A. N. Kapanidis, A. J. Turberfield, *ChemBioChem* **2009**, *10*, 1551–1557; b) W. Shen, H. Zhong, D. Neff, M. L. Norton, *J. Am. Chem. Soc.* **2009**, *131*, 6660–6661.

[7] B. Saccà, R. Meyer, M. Erkelenz, K. Kiko, A. Arndt, H. Schroeder, K. S. Rabe, C. M. Niemeyer, *Angew. Chem. Int. Ed.* **2010**, *49*, 9378–9383; *Angew. Chem.* **2010**, *122*, 9568–9573.

[8] J. Tominaga, Y. Kemori, Y. Tanaka, T. Maruyama, N. Kamiya, M. Goto, *Chem. Commun.* **2007**, 401–403.

[9] V. G. Metelev, E. A. Kubareva, O. V. Vorob'eva, A. S. Romanenkov, T. S. Oretskaya, *FEBS Lett.* **2003**, *538*, 48–52.

[10] a) S. Takeda, S. Tsukiji, T. Nagamune, *Bioorg. Med. Chem. Lett.* **2004**, *14*, 2407–2410; b) R. S. Sørensen, A. H. Okholm, D. Schaffert, A. L. B. Kodal, K. V. Gothelf, J. Kjems, *ACS Nano* **2013**, *7*, 8098–8104.

[11] a) Y. R. Yang, Y. Liu, H. Yan, *Bioconjugate Chem.* **2015**, *26*, 1381–1395; b) B. Saccà, C. M. Niemeyer, *Chem. Soc. Rev.* **2011**, *40*, 5910–5921.

[12] F. de la Cruz, L. S. Frost, R. J. Meyer, E. L. Zechner, *FEMS Microbiol. Rev.* **2010**, *34*, 18–40.

[13] M. Chandler, F. de la Cruz, F. Dyda, A. B. Hickman, G. Moncalián, B. Ton-Hoang, *Nat. Rev. Microbiol.* **2013**, *11*, 525–538.

[14] a) A. Guasch, M. Lucas, G. Moncalián, M. Cabezas, R. Pérez-Luque, F. X. Gomis-Rüth, F. de la Cruz, M. Coll, *Nat. Struct. Biol.* **2003**, *10*, 1002–1010; b) S. Datta, C. Larkin, J. F. Schildbach, *Structure* **2003**, *11*, 1369–1379; c) A. F. Monzingo, A. Ozburn, S. Xia, R. J. Meyer, J. D. Robertus, *J. Mol. Biol.* **2007**, *366*, 165–178; d) J. S. Edwards, L. Betts, M. L. Frazier, R. M. Pollet, S. M. Kwong, W. G. Walton, W. K. Ballentine, J. J. Huang, S. Habibi, M. Del Campo, J. L. Meier, P. B. Dervan, N. Firth, M. R. Redinbo, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 2804–2809; e) R. P. Nash, S. Habibi, Y. Cheng, S. A. Lujan, M. R. Redinbo, *Nucleic Acids Res.* **2010**, *38*, 5929–5943.

[15] a) M. K. Bhattacharjee, R. J. Meyer, *Nucleic Acids Res.* **1993**, *21*, 4563 – 4568; b) M. Llosa, G. Grandoso, F. de la Cruz, *J. Mol. Biol.* **1995**, *246*, 54 – 62.
[16] W. Pansegrau, E. Lanka, *J. Biol. Chem.* **1996**, *271*, 13068 – 13076.
[17] a) M. J. Harley, J. F. Schildbach, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 11243 – 11248; b) M. Lucas, B. Gonzalez-Perez, M. Cabezas, G. Moncalian, G. Rivas, F. de la Cruz, *J. Biol. Chem.* **2010**, *285*, 8918 – 8926.
[18] B. González-Pérez, M. Lucas, L. A. Cooke, J. S. Vyle, F. de la Cruz, G. Moncalián, *EMBO J.* **2007**, *26*, 3847 – 3857.

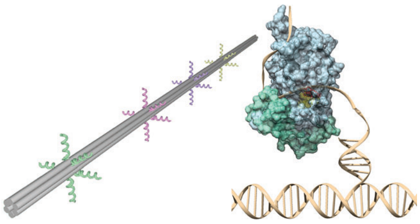
Received: November 5, 2015
Revised: December 17, 2015

Communications

DNA Nanotechnology

S. Sagredo, T. Pirzer, A. Aghebat Rafat,
M. A. Goetzfried, G. Moncalian,
F. C. Simmel,*
F. de la Cruz* ______ ■■■■–■■■■

Orthogonal Protein Assembly on DNA Nanostructures Using Relaxases



In nature relaxase proteins are involved in the conjugative transfer of plasmids between bacteria. Here it is shown that their specificity for a single-stranded DNA target, their good orthogonality, and their binding properties can also be utilized for the functionalization of DNA nanostructures with proteins.

PLOS | ONE

RESEARCH ARTICLE

Design of Novel Relaxase Substrates Based on Rolling Circle Replicases for Bioconjugation to DNA Nanostructures

Sandra Sagredo, Fernando de la Cruz*, Gabriel Moncalián*

Departamento de Biología Molecular e Instituto de Biomedicina y Biotecnología de Cantabria, Universidad de Cantabria-Consejo Superior de Investigaciones Científicas-SODERCAN, C/ Albert Einstein 22, 39011, Santander, Spain

* delacruz@unican.es (FC); moncalig@unican.es (GM)

Abstract

During bacterial conjugation and rolling circle replication, HUH endonucleases, respectively known as relaxases and replicases, form a covalent bond with ssDNA when they cleave their target sequence (*nic* site). Both protein families show structural similarity but limited amino acid identity. Moreover, the organization of the inverted repeat (IR) and the loop that shape the *nic* site differs in both proteins. Arguably, replicases cleave their target site more efficiently, while relaxases exert more biochemical control over the process. Here we show that engineering a relaxase target by mimicking the replicase target, results in enhanced formation of protein-DNA covalent complexes. Three widely different relaxases, which belong to MOB_F, MOB_Q and MOB_P families, can properly cleave DNA sequences with permuted target sequences. Collaterally, the secondary structure that the permuted targets acquired within a supercoiled plasmid DNA resulted in poor conjugation frequencies underlying the importance of relaxase accessory proteins in conjugative DNA processing. Our results reveal that relaxase and replicase targets can be interchangeable *in vitro*. The new Rep substrates provide new bioconjugation tools for the design of sophisticated DNA-protein nanostructures.

Citation: Sagredo S, de la Cruz F, Moncalián G (2016) Design of Novel Relaxase Substrates Based on Rolling Circle Replicases for Bioconjugation to DNA Nanostructures. PLoS ONE 11(3): e0152666. doi:10.1371/journal.pone.0152666

Editor: Eric Cascales, Centre National de la Recherche Scientifique, Aix-Marseille Université, FRANCE

Received: November 24, 2015

Accepted: March 17, 2016

Published: March 30, 2016

Data Availability Statement: All relevant data are within the paper and its Supporting Information file.

Funding: This work was financed by grants BFU2014-55534-C2-1-P from the Spanish Ministry of Economy and Competitiveness and 612146/FP7-ICT-2013 and 282004/FP7-HEALTH.2011.2.3.1-2 from the European Union Seventh Framework Programme to FC and grant BFU2014-55534-C2-2-P from the Spanish Ministry of Economy and Competitiveness to GM. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Introduction

HUH proteins are enzymes widespread in all three domains of life, where they process DNA during initiation of rolling circle replication (RCR) of certain phages and eukaryotic viruses, conjugative transfer of plasmid between cells and transposition of insertion sequences and helitrons [1]. These enzymes contain an HUH motif, in which two conserved histidines (H) involved in metal coordination are separated by a hydrophobic residue (U). HUH endonucleases also contain a Y motif, which contains the tyrosine(s) involved in the nucleophilic attack to cleave and rejoin the target single strand DNA (ssDNA). One attractive feature of HUH proteins is that the nucleophilic attack on specific ssDNA sequences results in stable protein-DNA covalent linkages. Thus, HUH proteins provide a tool for site specific bioconjugation of proteins to ssDNA, for instance to DNA origami nanostructures, where ssDNA is folded into a

Competing Interests: The authors have declared that no competing interests exist.

desired shape with the aid of hundreds of oligonucleotides named “staples” [2]. Relaxases such as TrwC of plasmid R388 or TraI of plasmid F have been proven to form protein-DNA conjugates efficiently [3]. Using this approach, DNA can be conjugated to any desired protein fused to either the N-terminus or the C-terminus of recombinant relaxases, without losing activity. Moreover, HUH proteins exhibit site-specific recombination and have been used as an efficient strategy for genome editing. The replicase of Adeno-associate virus (AAV) and relaxase TrwC from plasmid R388 catalyze site-specific DNA integration into human genomes [4] and thus constitute potential new tools for genome editing.

Relaxases and rolling-circle replicases belong to the HUH-endonuclease superfamily [1]. Replicases initiate replication of a large number of plasmids and viruses to generate new copies of their circular genomes [5]. Relaxases catalyze the transfer of one DNA strand of the plasmid genome to the recipient cell though a Type IV Secretion System during plasmid conjugation [6]. Both HUH endonucleases recognize their target ssDNA sites (called *nic* sites) with nanomolar specificity and form a protein-DNA covalent bond in presence of divalent cations. The linkage is a phosphotyrosyl intermediate between the catalytic tyrosine and the 5’ phosphate of the *nic* site. Moreover, they both carry out a second nucleophilic attack to the newly synthesized ssDNA strand, which results in recircularization of one unit of the ssDNA genome. Relaxases and replicases display structural similarity [1] but also outstanding differences. Structural data of relaxases and replicases revealed a similar fold (S1 Fig). In replicases, the HUH motif is located in a central five-stranded antiparallel β-sheet [1]. The catalytic tyrosine is placed in an α-helix close to the C-terminal end of the replicase core. Relaxases can be understood as having suffered a circular permutation of the primary sequence with respect to replicases [1]. Thus, the catalytic tyrosine in relaxases is located at the N-terminus, while the HUH motif resides closer to the C-terminus (S1 Fig). Replicases extrude the dsDNA, allowing the formation of a cruciform structure and a stable *nic*-containing ssDNA loop [7,8]. On the other side, the accepted model of conjugation claims that the cruciform structure is not formed to attain the *nic*-cleavage reaction [9]. The relaxase binds the proximal arm of an inverted repeat (IR) and locates the *nic* site as ssDNA within the active center of the protein (Fig 1). Nevertheless, the IR forms a hairpin structure on ssDNA after conjugative replication (in the recipient cell), allowing the relaxase-catalized strand-transfer reaction that leads to circularization of the ssDNA. Most relaxases do not cleave *nic* efficiently unless relaxase-accessory proteins (RAPs) allow the generation of ssDNA. On the other hand, RAPs are not required for replicase activity. Thus, in order to facilitate the relaxase *nic*-cleavage reaction, we relocated the *nic* site within the loop of the IR, mimicking a replicase substrate. This substrate was further improved by exchanging the relative position of IR and *nic* sites. The results obtained provide a strategy for the design of more efficient substrates to be used *in vitro* for relaxase bionanotechnological reactions.

Results

TrwC$_R$ cleaves substrates containing the *nic* site within the hairpin loop

Location of the *nic* site in their cognate substrate is structurally different in replicases, relaxases and other HUH endonucleases [1]. Despite the fact that HUH relaxases and replicases both recognize a hairpin-like structure, conjugative *nic* sites are located 3′ to the hairpin (Fig 1A) while replicase origins are located in the loop within the hairpin region (Fig 1B). The wt conformation recognized by relaxases possesses an inverted repeat (IR) that can be divided into a distal arm (D) and a proximal arm (P) with respect to the *nic* site (N). Moreover, there are eight nucleotides between P and N that are bent with a U-shape (U). The sequence (S) downstream N is tethered to the relaxase after cleavage (Fig 1C). In order to improve the relaxase

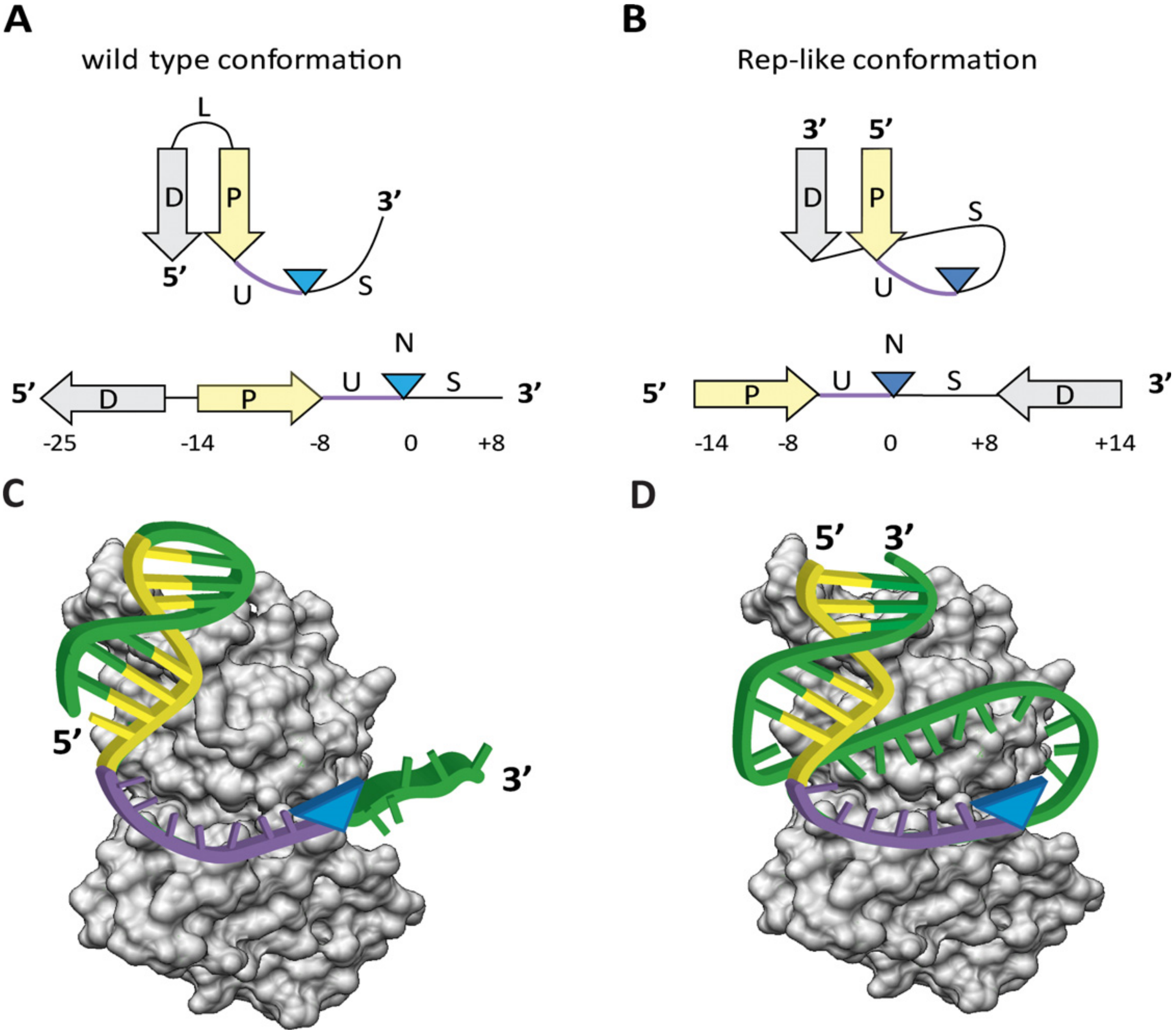


Fig 1. Scheme of relaxase and replicase DNA targets. (A) The relaxase DNA substrate contains an IR, defined by a Distal arm (D) and a Proximal Arm (P), that shape a hairpin structure. The *nic* site (N) is located between a U-turn (U) sequence and a ssDNA strand (S) that is tethered to the relaxase after cleavage. (B) The DNA substrate cleaved by replicases has the Distal arm (D) located downstream from *nic* (N), which allows stem-loop formation. Rep-like substrates for relaxases were designed by displacing the D sequence to the 3' end of S of their original wt substrate. (C) Scheme depicting the cleavage reaction of the wt substrate. Upon binding, relaxase bends its target in order to localize the *nic* site (N) within its active center. In presence of a divalent cation, the relaxase cleaves *nic* (blue arrowhead) and remains covalently bound to the 5'-phosphate of S, downstream from N. (D) Scheme depicting the cleavage reaction of the Rep-like substrate. The stem is bound by the relaxase, and the loop is located within the DNA binding cleft. Thus, the relaxase cleaves the scissile nucleotide within a stable cruciform (blue arrowhead). After cleavage the relaxase will be tethered to S and D. P is shown in yellow and U in purple. Blue triangles show the position of the *nic* site.

doi:10.1371/journal.pone.0152666.g001

nic-cleavage reaction, redesign of hairpin substrates was carried out by reconverting relaxase *nic* substrates into replicase like (Rep-like) substrates (Fig 1B). To create the Rep-like substrate, the 5´end of the distal sequence (D) was attached to the 3´end of the ssDNA region (S) after

Table 1. Cleavage activity of TrwC$_R$ on different oligonucleotides.

Nicking substrate[a]	*DNA sequence*[b]	*Complex Formation %*[c]
R388 W(25+18)	GCGCACCGAAAGGTGCGTATTGTCT/ATAGCCCAGATTTAAGGA	**24.7** ± 2.1
R388 W(14+14)	GGTGCGTATTGTCT/ATAGCCCAGATTTA	**19.8** ± 1.9
R388 H(14+8)	GGTGCGTATTGTCT/ATCGCACC	**22.4** ± 5.7
R388 H(14+10)	GGTGCGTATTGTCT/ATAGCGCACC	**23.0** ± 4.7
R388 H(14+12)	GGTGCGTATTGTCT/ATAGCGCGCACC	**25.5** ± 3.1
R388 H(14+13)	GGTGCGTATTGTCT/ATAGCCGCGCACC	**23.3** ± 2.9
R388 H(14+14)	GGTGCGTATTGTCT/ATAGCCCGCGCACC	**26.8** ± 4.7
R388 H(14+15)	GGTGCGTATTGTCT/ATAGCCCAGCGCACC	**19.6** ± 1.4
R388 H(14+17)	GGTGCGTATTGTCT/ATAGCCCAGATCGCACC	**22.3** ± 6.1
R388 H(16+16)	CCGGTGCGTATTGTCT/ATAGCCCGCGCACCGG	**-**
R388 H(23+23)	AACCGGCTAGGTGCGTATTGTCT/ATAGCCCACGCACCTAGCCGGTT	**11,2**
R388 H(23+26)	AACCGGCTAGGTGCGTATTGTCT/ATAGCCCAGATCGCACCTAGCCGGTT	**13,5**
R388 H(24+24)	CCCAATGCGCGGTGCGTATTGTCT/ATAGCCCACGCACCGCGCATTGGG	**12,4**
R388 H(24+27)	CCCAATGCGCGGTGCGTATTGTCT/ATAGCCCAGATCGCACCGCGCATTGGG	**12,0**
R388H(24+31)	CCCAATGCGCGGTGCGTATTGTCT/ATAGCCCAGATCCACCGCACCGCGCATTGGG	**-**
R388 R(8+14)	TATTGTCT/ATAGCCCACGCACC	**30.2** ± 4.9
R388 R(0+27)	ATAGCCCAGATCGCACCGAAAGGTGCG	**5.8** ± 1.9
R388 R(1+27)	T/ATAGCCCAGATCGCACCGAAAGGTGCG	**9.9** ± 3.2
R388 R(4+27)	GTCT/ATAGCCCAGATCGCACCGAAAGGTGCG	**9.9** ± 3.9
R388 R(7+27)	ATTGTCT/ATAGCCCAGATCGCACCGAAAGGTGCG	**18.3** ± 11.6
R388 R(8+24)	TATTGTCT/ATAGCCCACGCACCGAAAGGTGCG	**45.3** ± 6.6
R388 R(8+27)	TATTGTCT/ATAGCCCAGATCGCACCGAAAGGTGCG	**44.7** ± 1.0
Rsf1010 WQ(30+7)	CAGTTTCTCGAAGAGAAACCGGTAAGTGCG/CCCTCCC	**43.9** ± 1.4
Rsf1010 WQ(23+7)	CCGGTTGAAAACCGGTAAGTGCG/CCCTCCCC	**41.4** ± 1.3
Rsf1010 HQ(18+16)	GAGAAACCGGTAAGTGCG/CCCTCCCCAGTTTCTC	**13.1** ± 4.2
Rsf1010 HQ(18+19)	GAGAAACCGGTAAGTGCG/CCCTCCCGATCAGTTTCTC	**9.0** ± 3.1
Rsf1010 HQ(18+22)	GAGAAACCGGTAAGTGCG/CCCTCCCTAGCCCCAGTTTCTC	**27.8** ± 4.5
Rsf1010 RQ(8+28)	TAAGTGCG/CCCTCCCCAGTTTCTCGAAGAGAAACCG	**-**
Rsf1010 RQ(8+34)	TAAGTGCG/CCCTCCCAGCCCCCAGTTTCTCGAAGAGAAACCG	**1.4**
Rsf1010 RQ(8+40)	TAAGTGCG/CCCTCCCAGCTGAATGTTCGAGTTTCTCGAAGAGAAACCG	**-**
RP4 WP(24+8)	GTGAAGGAAACTTCACCTATCCTG/CCCGGCTG	**21.3** ± 0.9
RP4 WP(15+6)	ACTTCACCTATCCTG/CCCGGC	**24.4** ± 5.7
RP4 HP(14+14)	CTTCACCTATCCTG / CCCGGCTGGTGAAG	**-**
RP4 HP(14+21)	CTTCACCTATCCTG / CCCGGCTGTACCTACGTGAAG	**-**
RP4 RP(8+24)	CTATCCTG/CCCGGCTGGTGAAGGAAACTTCAC	**23.7** ± 2.7

[a] Nicking substrate names contain the plasmid name followed by either W (wild type) H (hairpin) or R (reverse), family group (Q for MOB_Q and P for MOB_P, omitted in R388 MOB_F) and the sequence length. + represents the *nic* site, and the number after it, the oligonucleotide length which remains bound to the relaxase after cleavage.

[b] The slash in the DNA sequence depicts the *nic* site. The underlined sequences highlight the inverted repeats.

[c] The cleavage activity of the relaxase was measured by the formation of relaxase-DNA covalent complexes in SDS-PAGE. Values are the average of three experiments, but R388 H(23+23),H(23+26), H(24+24), H(24+27) and RSF1010 RQ(8+34) that were not replicated.

doi:10.1371/journal.pone.0152666.t001

the *nic* site, thus creating a novel loop region (U+S) containing the *nic* site (N) and a dsDNA stem of 6 bp. The length of the loop (U+S) was tuned by including different number of nucleotides in the S region before the proximal arm (P) (see Table 1). Oligonucleotides with

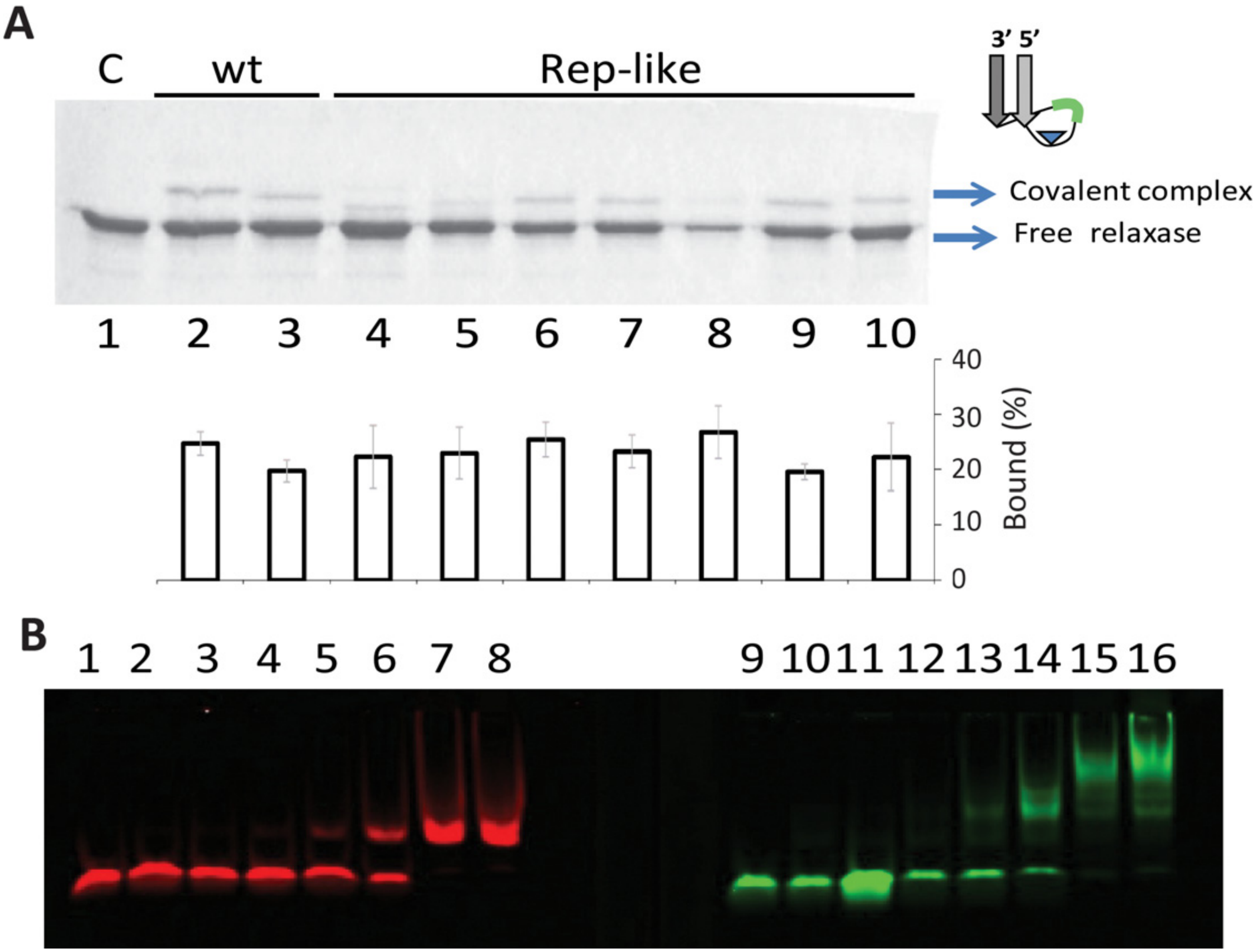


Fig 2. Interaction of TrwC_R with Rep-like substrates. (A) SDS-PAGE of oligonucleotides with R388wt or Rep-Like structures, when incubated with TrwC_R. 6 μM TrwC_R was incubated with 15 μM of different oligonucleotides. The reaction products were separated by electrophoresis in 12% SDS-PAGE gels. Lane 1, no oligonucleotide; Lanes 2 and 3, R388wt oligonucleotides W(25+18) and W(14+14), respectively; in subsequent lanes, TrwC_R was incubated with Rep-like oligonucleotides. The S length of Rep-like substrates (in green), varies from two to eleven nucleotides. Lane 4, H(14+8), Lane 5, H(14+10), Lane 6, H(14+12); Lane 7, H(14+13), Lane 8, H(14+14); Lane 9, H(14+15); Lane 10, H(14+17). In the center chart, percentage of bound complexes were calculated in three separate experiments such as that shown in (A). (B) Increasing amounts of TrwC_R were incubated with wt oligonucleotide W(25+8) (red shift, lanes 1 to 8) or Rep-like hairpin H(14+14) (green shift, lanes 9–16). Lanes 1 and 9, no protein added; 2 and 10, 42 nM of TrwC_R; 3 and 11, 85 nM; 4 and 12, 210 nM; 5 and 13, 420 nM; 6 and 14,850 nM; 7 and 15, 4,2 μM, 8 and 16, 8,5 μM.

doi:10.1371/journal.pone.0152666.g002

redesigned secondary structures were used to analyze *in vitro* the activity of the N-terminal relaxase domain of TrwC (residues 1 to 293) called TrwC_R hereafter.

The effect of the DNA substrate length and secondary structure on relaxase cleavage was investigated through *in vitro nic*-cleavage reactions. This reaction generates a protein-DNA covalent complex that can be quantified by its lower mobility using SDS-PAGE as described in Materials and Methods. TrwC_R was able to cleave and remain covalently bound to oligonucleotides containing IR_2-*nic* (W(25+18)) or P-*nic* (W(14+14)), (Fig 2A, Lanes 2 and 3). As previously described, the entire IR_2 increases the binding affinity of TrwC_R but the distal arm is not required for efficient cleavage (9). Interestingly, TrwC_R could also cleave a Rep-like

Table 2. Strand transfer reaction catalyzed by TrwC$_R$ of oligonucleotides with the Rep-like and Reverse-like layout to the W(25+0) oligonucleotide.

Nicking substrate[a]	*Transfer activity %*[b]
R388 H(14+8)	**29** ± 8
R388 H(14+10)	**17** ± 5
R388 H(14+12)	**19** ± 7
R388 H(14+13)	**20** ± 10
R388 H(14+14)	**17** ± 2
R388 H(14+15)	**13** ± 10
R388 H(14+17)	**10** ± 6
R388 R(8+14)	**2** ± 3
R388 R(7+27)	**3** ± 3
R388 R(8+24)	**2** ± 2
R388 R(8+27)	**4** ± 2
R388 W(14+14)	**31** ± 2
R388 W(25+18)	**43** ± 1

[a] Substrates are named as in Table 1.

[b] The strand transfer activity of the relaxase was measured by the formation of fluorescent W(25+x) oligonucleotides in denaturing PAGE. Values are the average of three experiments.

doi:10.1371/journal.pone.0152666.t002

oligonucleotide containing a loop with only 2 nucleotides in the S region (H(14+8)) (Fig 2, lane 4). TrwC$_R$ cleavage activity on H(14+8) was similar to that on W oligonucleotides. Incubation of TrwC$_R$ with Rep-like hairpins containing longer loops, such as H(14+10), H(14+12) or H (14+13), resulted in a band with reduced mobility, as expected by the formation of a TrwC$_R$ covalent complex with a decamer (Lane 5), dodecamer (Lane 6) and tridecamer (Lane 7), respectively. Similar results were obtained with substrates H(14+14), H(14+15) or H(14+17) (Lanes 8, 9 and 10, respectively). Labeled Rep-like oligonucleotide H14+14 was also found to be effectively cleaved by TrwC$_R$ using TBE-urea gel electrophoresis analysis (S2 Fig).

TrwC$_R$ protein was previously shown to cleave oligonucleotides containing the R388 *nic* site and to perform an *in vitro* strand transfer reaction to a second oligonucleotide also containing the *nic* site [10,11]. Since TrwC$_R$ cleaves Rep-like oligonucleotides and remains covalently attached to the oligomer downstream *nic* (Fig 1D), it could carry out the strand transfer reaction. When a labeled wt oligonucleotide acceptor F(25+0) was provided and incubated with TrwC$_R$ and oligonucleotide H(14+8), 29% of the labeled oligonucleotide shifted to a position corresponding to a F(25+8) oligonucleotide (Table 2 and S3 Fig). Compared to the transfer reaction using wt oligonucleotides as donors, this transfer efficiency is similar to W(14+14) (31%) and lower than W(25+18) (41%). Longer Rep-like oligonucleotides, such as H(14+10) and H(14+14), generated around 17% product. Overall, all Rep-like substrates were efficiently transferred to receptor oligonucleotides (see Table 2).

We thought that the distance (S) from *nic* (N) to the proximal arm (P) would be critical for cleavage, since only certain loop lengths would allow the *nic* site to reach the active site of the protein. As similar *nic*-cleavage efficiencies were observed regardless of loop length (Fig 2A), gel shift assays were performed to study how TrwC$_R$ interacts with Rep-like conformation substrates. Increasing concentrations of TrwC$_R$ were incubated with 50 nM of a green-fluorescent labeled Rep-like oligonucleotide H(14+14) or with a red-fluorescent labeled wt oligonucleotide W(25+8). Although the EC$_{50}$ (half maximal effective concentration) was the same with both oligonucleotides, only one shifted band was observed with W(25+8), while several shifted bands appeared with H(14+14). Both unbound oligonucleotides produced a single band

(Fig 2B, lane 1 and 9, respectively). Thus, upon TrwC$_R$ binding, W(25+8) seems to form a unique complex, while different complexes are formed with H(14+14).

The interaction profile of Rep-like oligonucleotides incubated with TrwC$_R$ was further analyzed by gel filtration. Consistent with the EMSA result, several peaks with higher molecular weight than the predicted peak for 1:1 complexes appeared when either Rep-like H(14+14) and H(16+16) (S = 8), H(14+15) (S = 9) or H(14+17) (S = 11) oligonucleotides were incubated with TrwC$_R$ (S4 and S5 Figs). It is likely that TrwC$_R$ could melt the hairpin and stabilize a "linear" DNA in a way that the distal and proximal arms of the IRs of different DNA molecules could interact, creating these higher molecular weight complexes.

In order to stabilize the stem-loop structure of Rep-like substrates to avoid the formation of intermolecular complexes, longer IRs were used (H(23+23), H(23+26), H(24+24), H(24+27), H(24+31)). Now, a sole band of protein-DNA was shown in chromatography, while SDS-PAGE gels revealed that the covalent complexes are still being formed with longer stem substrates (S6 Fig).

Rep-like substrates are cleaved by single-Y relaxases

Relaxases contain either one (Y1) or two (Y2) catalytic tyrosines [1,12]. The ssDNA U-turn observed in Y2 relaxase domains TrwC$_R$- and TraI$_R$ -DNA complex structures was also observed in the structure of Y1 relaxase NES [13–15]. However, relevant differences in U-turn formation have been described [12]. In order to check if Y1 relaxases could also catalyze cleavage of Rep-like substrates, we analyzed these substrates for the best known Y1-relaxases, MobA_RSF1010 and TraI_RP4 (S7 Fig). The MobA_RSF1010 binding site was identified as a 10 bp IR that forms a hairpin structure upstream from *nic* [16]. Mozingo et al. modeled the relaxase minMobA (amino acids 1 to 186, named MobA$_R$ in this study) complexed with a 33-mer oligonucleotide and compared this model with TrwC$_R$ bound to the 23-mer [16]. According to this comparison, only nucleotides at the base of the 10 bp-hairpin interact with MobA$_R$. To study the influence of IR length, we designed a shorter wt substrate, containing a 6 bp IR WQ(23+7) (Fig 3A and Table 2). Analysis by SDS-PAGE of the covalent complexes obtained after incubation with the protein showed that either the long wt substrate WQ(30+7) (Fig 3A Lane 2) or the short wt substrate WQ(23+7) (Fig 3A Lane 3) generated 40% covalent complexes. These results confirm that just the nucleotides at the base of the stem make specific interactions with relaxase MobA$_R$.

Rep-like substrates of RSF1010 *nic-site* sequence were designed as for R388 *nic* (Fig 1). Three different loop lengths (S) were tested, while maintaining the first 18 nucleotides invariant (P-U-N). The three different oligonucleotides tested, HQ(18+16), HQ(18+19) or HQ(18+21), showed reduced activity compared with wt Q oligonucleotides. MobA$_R$ cleaved the Rep-like substrate HQ(18+16) to 13% (Fig 3A Lane 4). HQ(18+19) (Lane 5) behaved similarly, producing 10% covalent complexes. In turn, HQ(18+21) (Lane 6) generated 20% covalent complexes (Table 2). These results suggest that Rep-like substrates also allow the correct location of the *nic* site in the catalytic center of the relaxase MobA$_R$.

Another widely characterized Y1 relaxase is TraI_RP4 [17–19]. The DNA substrate routinely used in TraI cleavage assays was the 21-mer oligonucleotide WP(15+6). None of the previous assays had used any substrate carrying an IR upstream from *nic*. In our study, a substrate with a 6-bp hairpin conformation WP(24+8) was designed. As shown in Fig 3B, lane 3 WP(24+8) was efficiently cleaved by a protein containing the 270 N-terminal residues of RP4 TraI (TraI$_R$). Covalent complex formation by the wt substrate was slightly lower than when WP(15+6) was used (Lane 2 and Table 2).

We then used a rational approach to design DNA stem-loops that could be cleaved by TraI$_R$. Rep-like oligonucleotides containing the minimal RP4 *nic* sequence were developed as we did before for R388 (P-U-N-S-D, see Fig 1). These substrates were incubated with TraI$_R$,

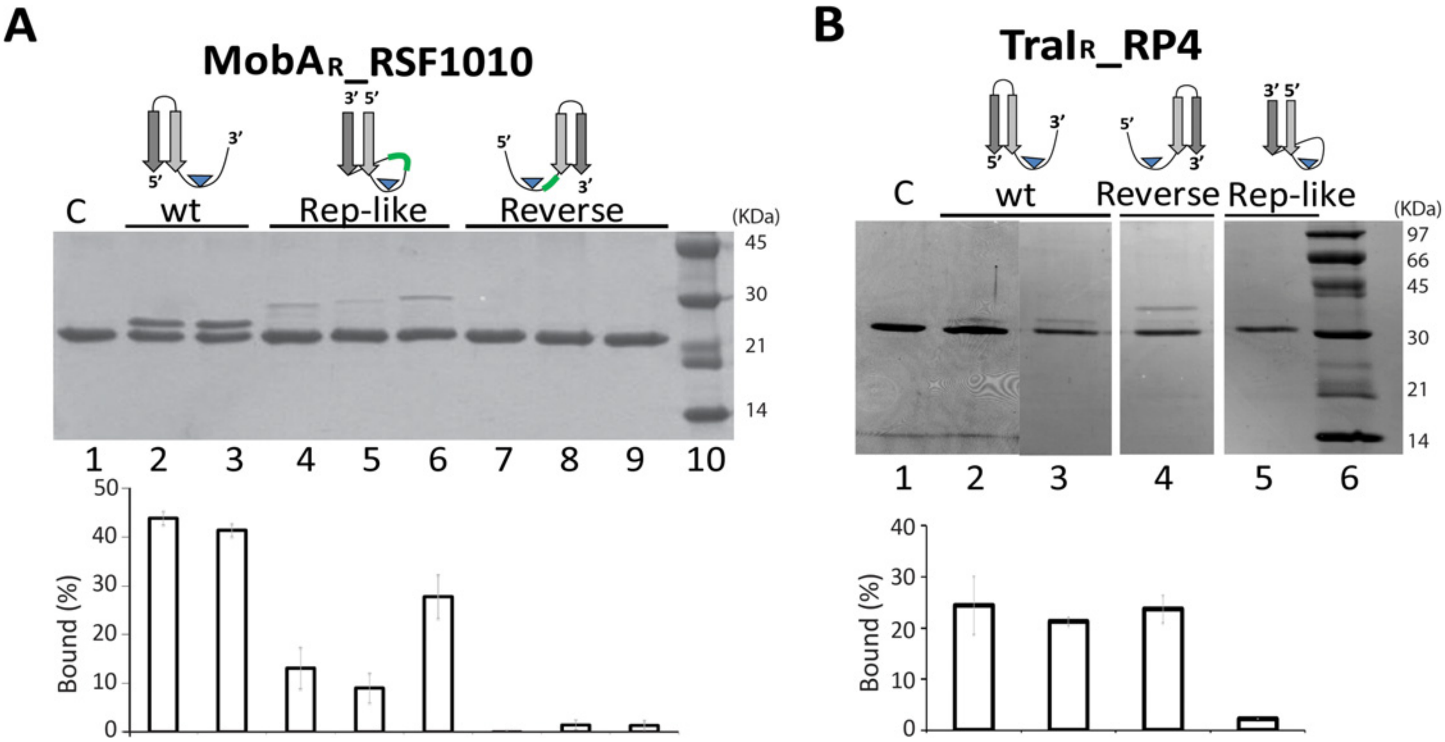


Fig 3. Novel designs for DNA substrates of model single-Y relaxases. (A) SDS-PAGE of MobA$_R$_RSF1010 with its targets. 7 μM MobA$_R$ was incubated with 15 μM of different oligonucleotides. Lane 1, MobA$_R$. Lanes 2 and 3, wt oligonucleotide WQ(30+7) and a substrate that lacks the upper-hairpin nucleotides of the IR WQ(23+7). Lanes 4, 5 and 6, Rep-like oligonucleotides HQ(16+16),HQ(16+19) and HQ(16+22), respectively. Lanes 7, 8 and 9 reverse substrates RQ(8+28), RQ(8+34) and RQ(8+40). Lane 10, molecular weight ladder. (B) SDS-PAGE of TraI$_R$_RP4 with its targets. 1.5 μM TraI$_R$ was incubated with 15 μM of different oligonucleotides. Lane 1, TraI$_R$; lane 2, wt substrate WP(15+6); lane 3, wt substrate WP(24+8); lane 4, reverse substrate RP(8+24) and lane 5, Rep-like substrate HP(14+14). Molecular weight ladder is shown on Lane 6. Bar graphs with the quantification of covalent complexes are shown below the SDS-PAGE gels. Data showed mean±s.d. of three independent experiments.

doi:10.1371/journal.pone.0152666.g003

and the resulting complexes where analyzed by SDS-PAGE. Neither Rep-like oligonucleotide HP(14+14) (Fig 3B Lane 5) nor HP(14+21) were cleaved, even though saturating concentrations of oligonucleotides where used (data not shown).

Improved scissile substrates were obtained by permutating the *nic* sequence

In TrwC$_R$ wt substrate W(25+8), the stem loop is non-covalently bound by the relaxase after cleavage, while the 8-mer downstream *nic* is covalently attached to the catalytic tyrosine. There is a cleavage-ligation equilibrium with wt oligonucleotides, because the 25-mer oligonucleotide remains in the protein DNA binding domain after cleavage and, therefore, the 8-mer can be easily religated (Fig 1C). We thought that, by linking the 3′ end of *nic* (U-N-S) to the 5′end of the hairpin, the 5′side of *nic* (the single strand U sequence upstream the *nic* site) would lose binding affinity (Fig 4A). As a consequence, the resulting 5′ssDNA could be released from the relaxase, avoiding religation and displacing the reaction equilibrium towards the formation of covalent complexes (Fig 4B). Under this rationale, novel substrates were designed by shuffling the wt sequence in the order U-N-S-D-P. These synthetic oligonucleotides were called "reverse substrates" because the hairpin is located downstream from the *nic* site.

R388 reverse oligonucleotides were designed using either the whole U ssDNA (8 nt before *nic*) or a fragment (up to seven nt), which would be liberated after cleavage. Besides, S length varied from 8 to 11 nucleotides, to allow the correct disposition of the hairpin into the relaxase.

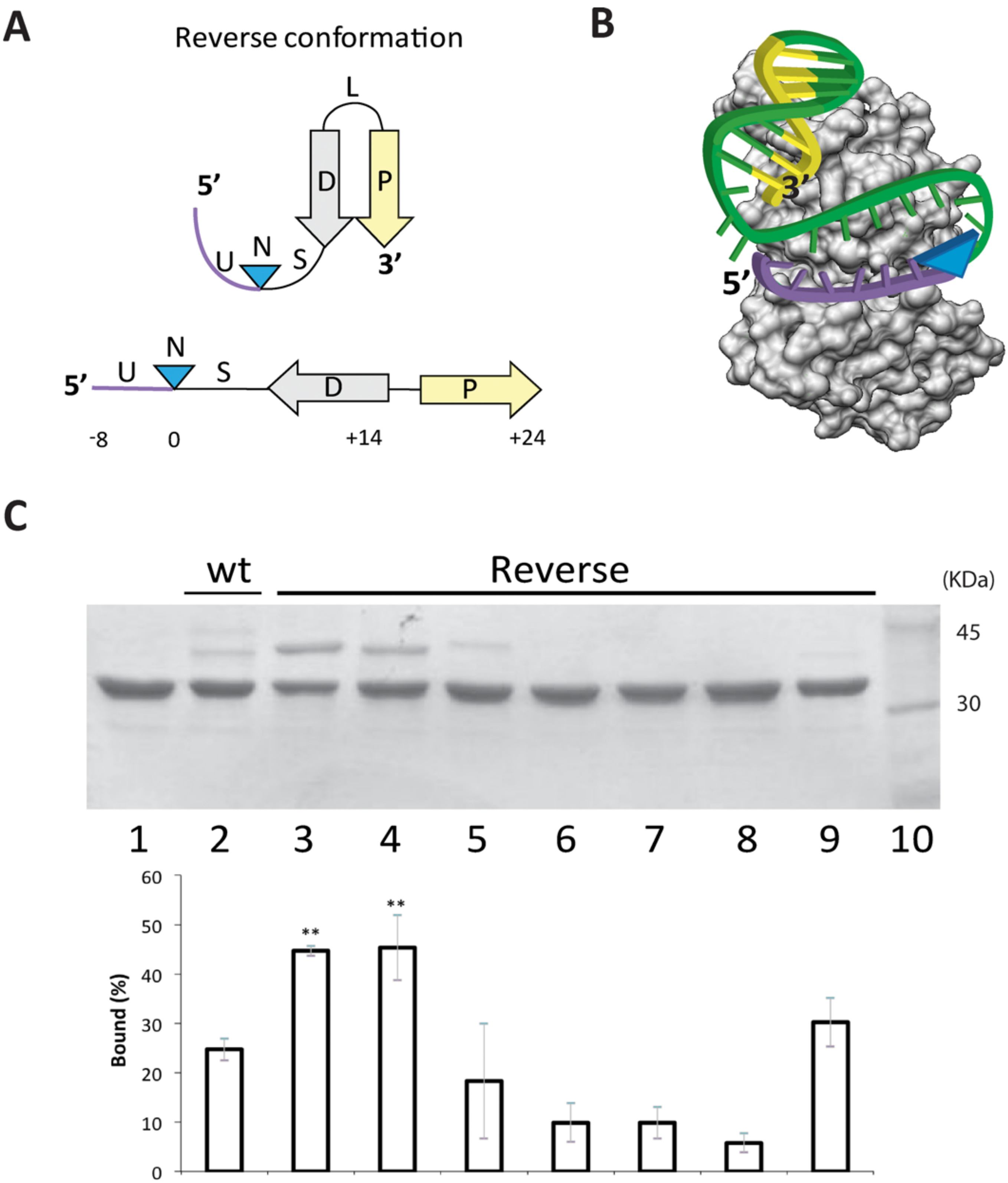


Fig 4. Interaction of TrwC$_R$ with Reverse substrates. (A) Reverse substrates were designed by swapping the 5’ region of the *nic* site to the 3’ end. This designed DNA substrate possesses the complete inverted repeat (D-P) at the 3’ end of the *nic* site (N). Either the U or the S lengths were tuned to allow the correct location of the hairpin within the relaxase binding domain. (B) Scheme depicting the cleavage reaction of the reverse substrate. Relaxase binding to

the reverse substrate allows both the hairpin and the single strand U-turn localize at the DNA binding cleft. This way the cleavage reaction forms a covalent complex of the relaxase with the region downstream of the *nic* site (blue arrowhead). Now the 5´side of *nic* do not contain the IR avoiding the re-ligation reaction. Color code is the same than Fig 1. (C) 12% SDS-PAGE of reverse oligonucleotides, when incubated with $TrwC_R$. 6 μM $TrwC_R$ was incubated with 15 μM of different reverse oligonucleotides. Lane 1, no oligonucleotide; Lanes 2, R388wt oligonucleotide W(25+18). Lanes 3 and 4, Reverse substrates R(8+27) and R(8+24), both with U = 8 nt and S = 11nt or S = 8 nt respectively. Lane 5, R(7+27) U = 7; Lane 6, R(4+27) U = 4; Lane 7, R(1+27) U = 1 and Lane 8, R(0+27) U = 0. Lane 9, R(8+14), U = 8 P = 0. Lane 10, Molecular weight marker. Graph quantifying the percentages of covalent complexes is shown below the SDS-PAGE gel. Data show mean±s.d. of three independent experiments. Two asterisks indicate P-value<0.05 by two-sided student's t-text.

doi:10.1371/journal.pone.0152666.g004

As shown in the SDS-PAGE gel of Fig 4C, the relaxase band shifted when incubated with R(8+24), R(7+27) or R(8+27) reverse oligonucleotides. In fact, the percentage of protein-bound DNA boosted with reverse oligonucleotides R(8+27) (Lane 3) and R(8+24) (Lane 4) to almost 50% when U had 8 nt. It is noteworthy that the covalent complex decreased to 18% when the U length was 7 nt (lane 5). Moreover, the covalent product was not detected when the U-region was shorter, such as when oligonucleotides R(4+27) (Lane 6), R(1+27) (Lane 7) or R(0+27) (Lane 8) were used. The different yields of covalent complexes revealed that U length is a determinant factor for the reaction, while S length had no significant influence.

Regarding Y1 relaxases, reverse substrate RP(8+24) was recognized and cleaved by $TraI_R$ _RP4 (Fig 3B Lane 4). The amount of covalent complex obtained with RP(8+24) was comparable to the amount obtained with wt WP(24+8) oligonucleotide (Table 1). Interestingly, $MobA_R$ could not cleave similarly designed oligonucleotides RQ(8+28), RQ(8+34) or RQ(8+40) (Fig 3A Lane 7, 8 and 9, respectively).

Plasmids carrying novel *nic*-site conformations show diminished conjugation rates

The stable interaction of $TrwC_R$ with the novel substrates supports the notion that both Rep-like and reverse oligonucleotides could modify the *in vivo* efficiency of relaxases. The importance of the DNA stem loop structure for efficient relaxase-mediated DNA processing *in vivo* was investigated by using plasmids containing synthetic *oriT*s. Synthetic R388 *oriT*s were designed by substitution of the wt *nic*-cleavage site (IR_2+*nic*) by either H(14+14) Rep-like or R(8+24) Reverse-like sequences (Fig 5 and S8 Fig).

OriT-containing plasmids (Cm^R) can be mobilized to a recipient cell (strain UB1637; Sm^R) in the presence of the helper plasmid pSU2007 (Km^R). Mobilization frequencies of the synthetic *oriT*-containing plasmids were measured as the number of mobilized plasmids (Cm^RSm^R) per donor cell (Cm^RNx^R) (Fig 5). The Rep-like *oriT* plasmid showed 10^4 fold lower mobilization than the wt *oriT* plasmid. By substitution of the Rep-like site for the wt *nic*-site, not only the relaxase recognition site was changed, but also the distance between the TrwC and TrwA/IHF binding sites. TrwA binding to *oriT* was shown to be essential for proper conjugation [20]. In the Rep-like *nic*-site, TrwC and TrwA binding sites were 10 nucleotides closer. Thus, we also analyzed the mobilization frequency of a synthetic plasmid that lacks the distal arm (ΔDistal Arm *oriT*). Interestingly, the conjugation rate of ΔDistal Arm *oriT* plasmid dramatically dropped below detection levels ($<10^{-6}$ transconjugants/donor).

We also checked the conjugation ability of a plasmid containing just the Rep-like *nic* site without any binding sites for IHF and TrwA, named minimal Rep-like. Mating carried out with minimal Rep-like *oriT* did not generate transconjugants. Similarly, reverse-like plasmids, which have the *nic* site placed upstream IR_2, also resulted in a substantial reduction of plasmid conjugation.

Discussion

In this work, we analyzed if relaxases would recognize and cleave substrates with a replicase *nic* site layout. Thus, we designed Rep-like substrates with secondary structures mimicking the

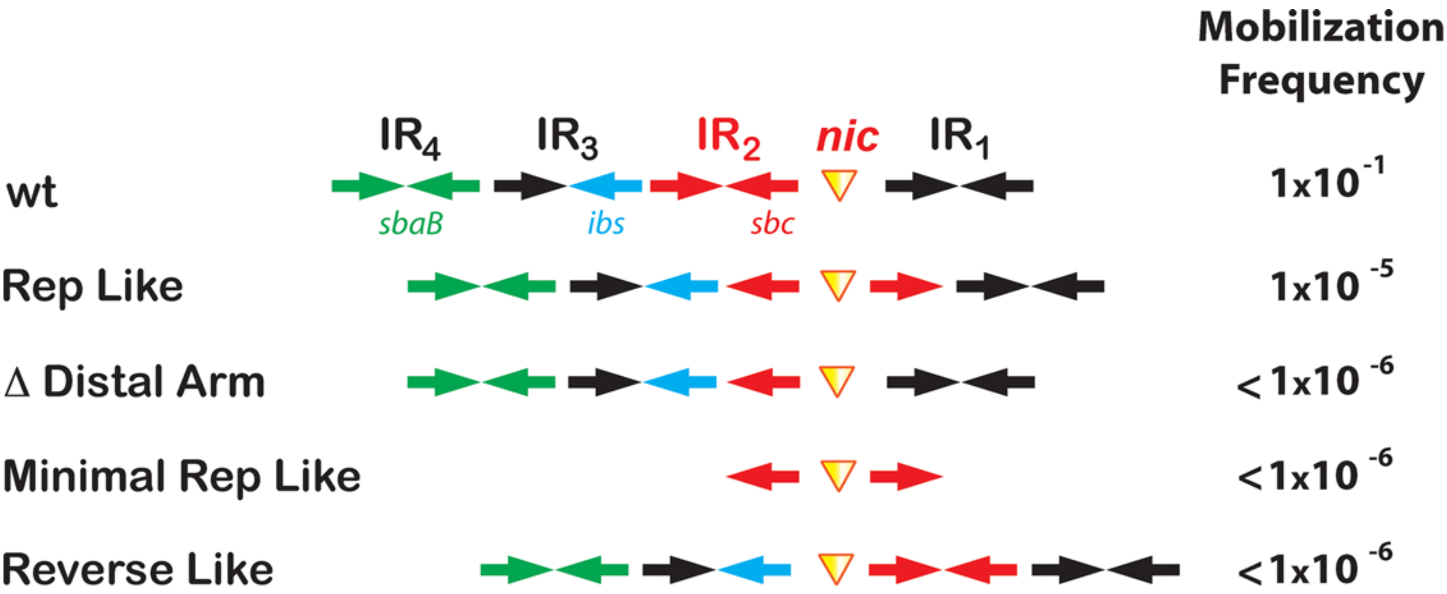


Fig 5. Mobilization frequencies of synthetic *ori*T-plasmids. Mobilization frequencies obtained using pSU2007 as helper plasmid are expressed as number of transconjugants per donor cell. Values are the average of three experiments. The relative position of the secondary structure elements of the assayed synthetic *oriT*s is shown. R388 wt *oriT* has four IR that are binding sites for the relaxase and nicking-accessory proteins TrwA and IHF. TrwC binds IR_2 and the *nic* site (*sbc*, red), TrwA binds IR_4 (*sbaB*, green) while IHF binds IR_3 (*ibs*, blue). The distance between secondary structure elements is not shown at scale.

doi:10.1371/journal.pone.0152666.g005

stem-loop recognized and cleaved by rolling circle replicases (Fig 1B). We observed that $TrwC_R$, the relaxase of plasmid R388, cleaved Rep-like *nic*- sites with a yield similar to wt substrates. Similar Rep-like substrates were designed and assayed for Y1 relaxases $MobA_R$_RSF1010 and $TraI_R$_RP4. $MobA_R$ efficiently cleaved the corresponding Rep-like substrates (Fig 3A) without melting the DNA to cleave the *nic* site. However, despite having found an effective wt hairpin substrate that is cleaved by $TraI_R$_RP4, this relaxase did not cleave Rep-like oligonucleotides (Fig 3B).

Although Rep proteins effectively cleave oligonucleotides mimicking their cognate substrates *in vitro* [21–23], in general relaxase activity with Rep-like substrates was lower than with their cognate substrates. The most likely explanation is that in Rep-like substrates the D and P segments of the IR are located downstream and upstream from N, respectively. Thus, now the relaxase tightly binds both sides of the *nic* site favoring the religation process (Fig 1D). To avoid this situation, we designed novel reverse substrates where the entire IR is now located downstream from N (Fig 4B). Using this substrate, relaxase contacts 5´ of *nic* are reduced and the religation reaction is diminished. Even though the balance was displaced to DNA-protein complex formation, 100% yield could not be achieved, probably because the relaxase has residual affinity for the U-turn 5´side of the cleaved reverse oligonucleotide. $TraI_R$ also cleaved the reverse substrate RP(8+24) although, unexpectedly, the yield of covalent complex formation with the novel substrates did not improve (Fig 3B). On the other hand, $MobA_R$ did not cleave reverse-oligonucleotides. This protein lacks the fingers domain at the C-terminal end. The fingers domain shapes a cleft that stabilizes the U-turn in the active center in MOB_F relaxases (S1 Fig). Probably this feature allows $MobA_R$ to form 40% covalent complexes with wt oligonucleotides WQ(30+7) and WQ(23+7) (Fig 3A), because religation is limited. We hypothesize that this feature could restrict the cleavage of reverse substrates. Overall, the cleavage-ligation reaction is tightly in balance for each relaxase, and not all relaxases can improve the yield of covalent complexes.

Once demonstrated that *in vitro* reactions were enhanced by novel relaxase substrates, we analyzed their behavior in whole cell experiments. With the goal of improving the efficiency of relaxases within living cells, synthetic plasmids in which the relaxase target was converted to either Rep-like sequence H(14+14) or Reverse R(8+24) were tested. Conjugation rates critically fell for all synthetic plasmids tested (Fig 5). Rep proteins can cleave dsDNA plasmids in the absence of RAPs because the *nic* site is at the loop of a cruciform IR [24]. However, most relaxases need RAPs to locally alter DNA supercoiling in a way the relaxase can melt the DNA to display the *nic*-containg ssDNA [25–27]. In the conjugation assay of Rep-like and reverse *nic* containing plasmids, we changed the relative orientation of *nic* with respect to TrwA (*sbaB*) and IHF (*ibs*) binding sites. Moreover, ssDNA hairpin formation by the modified IR region is different in recipient cells and could affect the termination reaction. It has been described for relaxases as well as replicases that the sequence requirements for the termination are different than the requirements for initiation [7,9]. Further studies will focus on the directed evolution of these synthetic *ori*Ts to obtain transferable plasmids in the absence of RAPs.

The significance of the *in vitro* results derives from the finding of new cleavable substrates for HUH proteins by rational design. Results show that relaxase $TrwC_R$ acting on reverse *nic* sites is more efficient in cleavage than when acting on its cognate sites. Relaxases are bifunctional proteins, having an N-terminal relaxase domain composed of roughly 300 residues fused to a C-terminal domain with various biochemical activities (helicase, primase, etc). The relaxase domain can be fused with heterologous proteins (e.g., fluorescent proteins) without losing activity. Since relaxases (and relaxase proteins fused to fluorescent proteins) attach covalently to a specific sequence of DNA, even on DNA nanostructures [3], our results are promising for nanobiotechnology applications using DNA nanostructures, where efficient methods for bioconjugation are needed. In an ideal case, the ability of relaxases to form covalent bonds with DNA nanostructures could be used for creating nanofactories. The relaxase domains would act as linkers between catalytic proteins (e.g., proteins involved in a given biochemical pathway) and the DNA nanostructures [28]. Relaxases can be tethered to specific locations in 2D and 3D DNA nanostructures so that any protein of interest can be placed in a precise and programmable mode [29]. As a second set of applications, TrwC is a potential tool for human genome editing [30]. By widening the number of sequences that TrwC and other relaxases can recognize, the potential targets for genome integration and targeted modification of the human genome could be greatly increased [4]. The present results have to be considered together with previous data where crucial nucleotides involved in target selection were identified [12]. Here we show that these targets could be expanded by the use of permuted sites. Overall, our results put relaxases among the most versatile tools for bioconjugation to DNA nanostructures.

Materials and Methods

Protein purification

The N-terminal relaxase domain of TrwC ($TrwC_R$, amino acid residues 1 to 293) was purified as previously described [14] with minor modifications. $TrwC_R$ expression from plasmid pSU1588 (pET3a:$trwc_R$) in *Escherichia coli* strain C41 was induced with 0.5mM IPTG for 3 h. Cells were harvested by centrifugation and stored at -80°C. Frozen cells were thawed at 37°C, resuspended in a solution containing Tris 100 mM (pH = 7.6), NaCl 500 mM, EDTA 1mM, PMSF 1% and lysed by 10 min. sonication in pulses. The lysate was centrifuged at 40,000g for 15 min at 4°C. Supernatants were applied to a P11-phosphocellulose column equilibrated in buffer A (50 mM Tris–HCl pH 7.6, 0.1 mM EDTA) containing 200 mM NaCl and eluted with buffer A with 600 mM NaCl. Fractions containing $TrwC_R$ were pooled, diluted to 200 mM NaCl, loaded onto a Hi-Trap SP HP column (GE Healthcare), and eluted with a linear NaCl gradient (200 mM–1000 mM NaCl) in buffer A.

The N terminal fragment of RSF1010 *mobA*, which codes for the first 186 residues of MobA (called minMobA by Mozigo et al.[16] and here named $\mathrm{MobA_R}$), was cloned in pET3a expression vector between *Nde*I and *Bam*HI restricition sites (pSU10064). Expression of $\mathrm{MobA_R}$ was achieved using *E.coli* C41 strain. After induction for 3 h with IPTG, cells were harvested and stored at -80°C. Purification was performed according to the procedure described for $\mathrm{TrwC_R}$.

The N terminal domain encoding the 270 residues of RP4 TraI ($\mathrm{TraI_R}$) was PCR amplified from plasmid RP4 and subcloned in pET3a using *Nde*I and *Bam*HI restriction sites, resulting in plasmid pSSP10. C41 strain carrying pSSP10 was induced for 3 h with IPTG. Cells were collected by centrifugation and stored at -80°C. RP4 $\mathrm{TraI_R}$ was purified as described above for $\mathrm{TrwC_R}$, except the Hi-Trap SP HP column was substituted by a Hi-Trap Heparin column, from which $\mathrm{TraI_R}$ was eluted with a linear 200–1000 mM NaCl gradient.

Gel filtration of all relaxases was carried out in a Superdex75 column 10/300 GL (GE Healthcare) equilibrated in 100 mM Tris–HCl (pH 7.6), 500 mM NaCl, 0.1 mM EDTA. All three relaxases eluted as a monomers of about 30 kDa for $\mathrm{TrwC_R}$, 30 kDa for $\mathrm{TraI_R}$ and 25 kDa for $\mathrm{MobA_R}$ (see S7 Fig). Protein concentration was estimated by UV absorbance at 280 nm in a Nanodrop spectrophotometer, using extinction coefficients of 35410 M^{-1} $cm^{-}1$ (Abs 0.1% of 1.078) for $\mathrm{TrwC_R}$, 32430 M^{-1} cm^{-1} (Abs 0.1% of 1.551) for $\mathrm{MobA_R}$ and 20190 M^{-1} cm^{-1} (Abs 0.1% of 0.636) for $\mathrm{TraI_R}$. After Superdex75 column chromatography, relaxase containing fractions were pooled and stored at -80C for further use.

ssDNA- Protein complex formation

Unlabeled DNA oligonucleotides were obtained from Sigma-Aldrich (St. Louis, MO). Their sequences are listed in Table 1. Oligonucleotides were resuspended in miliQ water at 100 μM, heated to 95°C for 10 min, and then either allowed to cool passively to room temperature or snap cooled on ice.

Cleavage reactions were carried out by incubating 6.3 μM $\mathrm{TrwC_R}$ with 15 μM of each oligonucleotide at 37°C for 1 h in 10mM Tris-HCl, pH7.6, 5mM $MgCl_2$, 375 mM NaCl and 15 μM EDTA [11]. Similarly, $\mathrm{MobA_R}$ (7 μM) was incubated with 15 μM of each substrate in 25 mM Tris-HCl pH 8.0, 5 mM $MgCl_2$, 225 mM NaCl and 15 μM EDTA [31]. For $\mathrm{TraI_R}$, the optimized reaction mixture was 10 mM Tris pH 7.6, 5 mM $MgCl_2$, 225mM NaCl and 15 μM EDTA [17]. 1.5 μM of $\mathrm{TraI_R}$ was incubated with 15 μM oligonucleotide in each reaction. Reactions were stopped by adding SDS and boiling the samples. The cleavage activity was checked by the lower mobility of the protein-DNA covalent complexes in SDS-PAGE. There was always a 2:1 molar excess of oligonucleotides to guarantee that all the protein is in complex with the oligonucleotide. DNA nicking activity was expressed as percentage of covalent complexes generated (the intensity of the product band divided by the sum of product and free protein band intensities). Each data point represents the average of three reactions. Data were processed using Quantity One (Biorad, Hercules, CA).

IRDye-labelled oligonucleotide cleavage assays

Cleavage reaction mixtures (20 μL) contained 50 nM IRDye-labelled oligonucleotide W(25+8) or H(14+14), 1 μM $\mathrm{TrwC_R}$ in 10 mM Tris-HCl, pH7.5, 5 mM $MgCl_2$, 100 mM NaCl. After incubation for 1 h at 37°C, reaction mixtures were digested with 0.6 mg/ml proteinase K and 0.05% (w/v) SDS for 20 min at 37°C. Reactions were run through a denaturing TBE-urea polyacrylamide gel at 200V for 70 min to separate cleaved product DNA from the substrate. Oligonucleotides were visualized using an Odyssey Infrared Image System (LI-COR Biosciences).

Strand-transfer assays

For oligonucleotide strand-transfer reactions, mixtures contained 0.25 μM of Rep-like oligonucleotides (Table 1) and1 μM of protein TrwC$_R$ in 10 mM Tris- HCl pH 7.6, 5 mM MgCl$_2$, 110 mM NaCl and 15 μM EDTA. After 30 min, 50 nM IRDye800-labelled oligonucleotide R388 W(25+0) (5´- CGCACCGAAAGGTGCGTATTGTCT) was added to the reaction mixture. After incubation for 30 min at 37°C, reaction mixtures were digested with 0.6 mg/ml proteinase K and 0.05% (w/v) SDS for 20 min at 37°C. Oligonucleotide separation was performed by electrophoresis in a denaturing 18% Acrylamide, 8 M Urea gel at 200V for 70 min. Images were processed with Odyssey Infrared Image System (LI-COR Biosciences) and quantified by Quantity One software (Biorad).

Electrophoretic Mobility Shift Assay (EMSA)

TrwC$_R$ binding to IRDye800-labelled oligonucleotide R388 W(25+8) (CGCACCGAAAGGTGCGTATTGTCT/ATATTGTCT), or IRDye700-labelled R388 H(14+14) (GGTGCGTATTGTCT/ATAGCCCGCGCACC) was analyzed by EMSA. Binding reactions contained 50 nM IRDye-labeled oligonucleotide and increasing concentrations of TrwC$_R$ in buffer (10 mM Tris-HCl pH 7.6, 100 mM NaCl, 1 mM EDTA). Reaction mixtures were incubated for 30 min at room temperature and loaded onto a 10% non-denaturing polyacrylamide gel. After electrophoresis at 200 V during 15 min, images were processed and quantified using Odyssey Infrared Image System (LI-COR Biosciences).

Intra- and intermolecular structures of *nic*-site oligonucleotides

Analysis of intra- and intermolecular structures in the oligonucleotides library created for the determination of the best substrate for TrwC$_R$ was determined by gel filtration in a Superdex75 PC 3.2/30 (GE Healthcare). 2.0 μl of each oligonucleotide (100 μM) were diluted in 18 μl of buffer solution (100 mM Tris pH 7.6, 200 mM NaCl, 10 mM EDTA) and injected in the column loop of 25 μl. The elution buffer was 100 mM Tris 7,6, 200 mM NaCl, 1 mM EDTA. When we analyzed the protein, 20 μl of stock solution (42 nM of TrwC$_R$) was injected alone. The complexes of TrwC$_R$ with oligonucleotides were loaded after 30 min incubation at room temperature, with an excess of oligonucleotide (ratio 1:1.5) to ensure maximum complex formation.

To study the new balance achieved with Rep-like substrates, IRDye700 labeled oligonucleotide R388 H(14+14) mixed with TrwC$_R$ in presence of 10 mM Mg$_2^+$ during 1 h was analyzed by gel filtration and native electrophoresis. After high resolution gel filtration column chromatography of the binding mixture, 10 μl of the fractions were loaded into a 5% non-denaturing acrylamide gel, and the fluorescent label of the oligonucleotide was detected with Odyssey infrared scanner.

Conjugation experiments

The Rep-like and Reverse-like *oriT* plasmids were design and ordered to GeneArt (Life technologies-Invitrogen). The ΔDistal Arm plasmid resulted from pSU1186 PCR-amplified with primers containing the deletion of the distal arm of IR$_2$. The linearized plasmid was self-circularized with T4 ligase. For minimal Rep-like plasmid, two oligonucleotides with the 28 base pairs Rep-like sequence between *Bam*HI and *Hind*III sites were cloned in the pSU19 digested vector.

Matings were carried out by the plate-mating procedure described in [32]. DH5α (Nx^R) cells containing the helper plasmid pSU2007 (a Km^R derivative of R388) [9,33] and the mobilizable synthetic *oriT*-containing plasmids to be checked (Cm^R) were mated with strain UB1637 (Sm^R) as a recipient strain. Saturated cultures of donor and recipient strains were mixed in a 1:1 ratio and mated on a LB agar surface. After 1 h at 37°C, serial dilutions were plated. Mobilization frequencies were expressed as number of Cm^R transconjugants (Cm^RSm^R) per donor cell.

Supporting Information

S1 Fig. Comparison of the 3D structure of relaxases and replicases. A) Organization of the catalytic centre in Relaxases ($TrwC_R$ from R388 plasmid bound to its target (PDB 1OMH)) and RCR replicases (AAV with RBE (PDB 1UUT)). Relaxases and RC-Rep suffered a circular permutation in the primary sequence that localizes the catalytic tyrosine at the N-terminal in the relaxases, but close to the C-terminal in Rep proteins. Relaxases recognize the *nic* site (red triangle) 5´ to an inverted repeat, while Reps cleave a *nic* site within a stem-loop. B) Ribbon structure of relaxases (left) and RC-Reps (right). $TraI_R$ from plasmid pCU1 (PDB 3L57) and $TraI_R$ from plasmid F (PDB 2AOI) belong to MOB_F family of relaxases whereas NES from pLW1043 (PDB 4HT4) and minMobA frm R1162 (PDB 2NS6) belong to MOB_Q family of relaxases. RepB, the RC initiators of plasmid pMV158 (PDB 3DKX) and Rep from geminivirus (PDB 1L5I) also have a similar folding. All these proteins possess a core of five antiparallel β-strand, where the HUH motif is located at the third β-strand, near the α-helix in which the catalytic tyrosine(s) is held.
(PDF)

S2 Fig. Rep-like substrate cleavage by $TrwC_R$. Denaturing TBE-Urea gels showing the products of cleavage of wt and Rep-like substrates by $TrwC_R$. Wt 5'IRDye700-labelled W(25+8) oligonucleotide is shown in lanes 1 and 2, whereas Rep-like 5'IRDye800-labelled H(14+14) oligonucleotide is shown in lanes 3 and 4. Oligonucleotides were incubated with $TrwC_R$, treated with proteinase K and SDS, as described in Material and methods. A 25-mer (lane 2) and a 14-mer (lane 4) oligonucleotide appeared in samples treated with $TrwC_R$.—represents the control DNA substrate (lanes 1 and 3).
(PDF)

S3 Fig. Strand transfer reaction catalyzed by $TrwC_R$ with wt, Rep-like and reverse substrates. (A) Schematic representation of the strand transfer reaction. $TrwC_R$ cleaves oligonucleotides containing the R388 *nic*-site and performs the strand transfer reaction to a labeled oligonucleotide W(25+0). In green are shown the oligonucleotides that remain covalently attached to $TrwC_R$. The gels show the appearance of a labeled transfer product. (B) Denaturing gels showing the strand transfer reaction catalyzed by $TrwC_R$ with wt, Rep-like and reverse substrates. 5'IRDye-labelled W(25+0) oligonucleotide is shown in control lanes 1 and 7. A 25+x labeled oligonucleotide was obtained after strand transfer reaction with the wild type oligonucleotides W(25+18), W(18+18), W(14+14) and W(8+8) in lanes 2, 3, 4 and 5, respectively. Strand transfer reaction to W(25+0) was also analyzed for Rep-like H14+x oligonucleotides. Lane 6, H(14+8); lane 8, H(14+10); lane 9, H(14+12); lane 10, H(14+13); lane 11, H(14+14); lane 12, H(14+15) and lane 13, H(14+17). Reverse substrates were also analyzed for strand transfer in lane 14, R(7+27); lane 15, R(8+24); lane 16, R(8+27) and lane 17, R(8+14).
(PDF)

S4 Fig. Chromatograms of *nic*-containing oligonucleotides with or without $TrwC_R$. 6.3 μM $TrwC_R$ was incubated during one hour in presence of EDTA with a 1.5:1 molar excess of each oligonucleotide. Then 20 μl of the samples were injected in a S75 column using the ETHAM system (GE Biosciences). Free oligonucleotides at the same concentration were also injected

separately for comparison. Chromatograms of oligonucleotides are shown as dashed lines while $TrwC_R$ with oligonucleotides are shown as continuous lines. Several shifted peaks and broaden peaks are observed in the cases of short hairpins H(14+14), H(14+15) and H(14+17). $TrwC_R$ interacts with oligonucleotides with longer stems H(16+16), H(23+23), and H(23+26) in a similar fashion. Reverse oligonucleotides R(8+27), R(8+24), R(7+27) show a slight shift when $TrwC_R$ is bound to it, similar to the one obtained with wt oligonucleotides (data not shown). Oligonucleotides tested were Rep-like oligonucleotides with D = P = 6 and different loop lengths; H(14+14) S = 8, H(14+15) S = 9 and H(14+17), Rep-like oligonucleotides with S = 11 and different stem lengths: H(16+16) D = P = 8 S = 8, H(23+23) D = P = 15 S = 8, and H(23+26) D = P = 15 S = 11 and Reverse oligonucleotides; R(8+27) D = P = 6 U = 8 S = 11, R(8+24) D = P = 6 U = 8 S = 8 and R(7+27) with D = P = 6 U = 7 S = 11. Absorbance at 260 nm was used during chromatography to determine the presence of DNA. All the chromatograms are normalized and shifted peaks are shown by stars.
(PDF)

S5 Fig. Chromatogram and acrylamide native gel analysis of $TrwC_R$ with fluorescent oligonucleotide H14+14. 6.3 μM $TrwC_R$ was incubated during one hour at RT in presence of 10 mM $MgCl_2$ with a 2:1 molar excess of IRDye H(14+14) oligonucleotide. Then 20 μl of the sample was injected in a S75 column using the ETHAM system (GE Biosciences). The chromatogram shows the obtained four A_{260} peaks with the Ve of each peak. After gel filtration column chromatography, 20 μl of 19 fractions collected within these four peaks were loaded into an acrylamide native gel. Lane M contains free oligo. Shifted oligonucleotides are shown by arrowheads. Gels were scanned in Odyssey (LI-COR).
(PDF)

S6 Fig. SDS-PAGE gel of $TrwC_R$ cleavage reactions on Rep-like substrates with short (L = 6) and long (L = 15/16) loops. 6.3 μM $TrwC_R$ was incubated with 15 μM of different oligonucleotides. Lane 1, no oligonucleotide; lane 2, H(14+14) P = D = 6 S = 8; lane 3, H(14+17) P = D = 6 S = 11; lane 4, H(23+23) P = D = 15 S = 8, lane 5, H(23+26) P = D = 15 S = 11, lane 6 H(24+24) P = D = 16 S = 8; lane 7, H(24+27) P = D = 16 S = 11. Lane 8, SDS-low range molecular ladder.
(PDF)

S7 Fig. Comparison of relaxases $TrwC_R$, $TraI_R$ and $MobA_R$ used in this study. (A) Location of the catalytic tyrosine and HUH motif in relaxases and replicases. The alignment of Y and HUH motifs in the model relaxases used in this study is shown. (B) Three dimensional structure of HUH relaxases $TrwC_R$, $TraI_R$ and $MobA_R$. 3D structures of $TrwC_R$ and $MobA_R$ were determined by x-ray crystallography (PDBs 1OMH and 2NS6) while $TraI_R$ was modelled with RaptorX (raptorx.uchicago.edu). HUH relaxases have the catalytic tyrosine within an α-helix (depicted in blue in the ribbon structures) and the motif H+HUH located in two juxtaposed β-sheets (depicted in light pink and wheat respectively). (C) SDS PAGE gels showing the protein purity after HPSP (Lane1, $TrwC_R$ and $TraI_R$) or Heparin (Lane1, $MobA_R$) column chromatography and the protein purity after S75 gel filtration column chromatography (lanes 2). M, standards of the Low Range Protein Ladder (BioRad). Overlay of S75 chomatograms of the gel filtration molecular weight markers Bovine-Serum-Albumin (BSA, 67 kDa) and Ribonuclease A (RBA, 13,7 kDa) (blue); $TrwC_R$ (green); $TraI_R$ (yellow) and $MobA_R$ (orange). All the relaxases elute as monomers with an apparent calculated molecular weight of 30 kDa for $TrwC_R$ and $TraI_R$, and 21 kDa for $MobA_R$.
(PDF)

S8 Fig. Alignment of the synthetic *oriT* based on Rep and reverse-like *nic* sites. The four inverted repeats (IR_1 to IR_4) are highlighted and their sequences underlined. New IR_2s obtained in Rep-like and reverse plasmids are shown by orange arrows. The *nic* site is indicated with a yellow triangle. Stars show point mutations included in some of the R388 synthetic plasmids to create *BglII*, *BamHI*, *KpnI* and *PstI* restriction sites.
(PDF)

Acknowledgments

We are grateful to Matilde Cabezas for technical assistance. This work was financed by grants FU2014-55534-C2-1-P from the Spanish Ministry of Economy and Competitiveness and 612146/FP7-ICT-2013 and 282004/FP7-HEALTH.2011.2.3.1–2 from the European Union Seventh Framework Programme to F.C. and grant BFU2014-55534-C2-2-P from the Spanish Ministry of Economy and Competitiveness to G.M.

Author Contributions

Conceived and designed the experiments: FC GM. Performed the experiments: SS. Analyzed the data: SS FC GM. Wrote the paper: SS FC GM.

References

1. Chandler M, la Cruz de F, Dyda F, Hickman AB, Moncalián G, Ton-Hoang B (2013) Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. Nature Publishing Group 11: 525–538. doi: 10.1038/nrmicro3067
2. Rothemund PWK (2006) Folding DNA to create nanoscale shapes and patterns. Nature 440: 297–302. doi: 10.1038/nature04586 PMID: 16541064
3. Sagredo S, Pirzer T, Aghebat Rafat A, Goetzfried MA, Moncalián G, Simmel FC, et al. (2016) Orthogonal Protein Assembly on DNA Nanostructures Using Relaxases. Angew Chem Int Ed Engl. doi: 10.1002/anie.201510313
4. González-Prieto C, Agúndez L, Linden RM, Llosa M (2013) HUH site-specific recombinases for targeted modification of the human genome. Trends in Biotechnology 31: 305–312. doi: 10.1016/j.tibtech.2013.02.002 PMID: 23545167
5. Ilyina TV, Koonin EV (1992) Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. Nucleic Acids Research 20: 3279–3285. doi: 10.1093/nar/20.13.3279 PMID: 1630899
6. Llosa M, Gomis-Rüth FX, Coll M, la Cruz Fd de F (2002) Bacterial conjugation: a two-step mechanism for DNA transport. Mol Microbiol 45: 1–8. PMID: 12100543
7. Zhao AC, Khan SA (1997) Sequence requirements for the termination of rolling-circle replication of plasmid pT181. Mol Microbiol 24: 535–544. PMID: 9179847
8. Khan SA (1997) Rolling-circle replication of bacterial plasmids. Microbiology and Molecular Biology Reviews 61: 442–455. PMID: 9409148
9. Lucas M, Gonzalez-Perez B, Cabezas M, Moncalian G, Rivas G, la Cruz de F (2010) Relaxase DNA Binding and Cleavage Are Two Distinguishable Steps in Conjugative DNA Processing That Involve Different Sequence Elements of the nic Site. Journal of Biological Chemistry 285: 8918–8926. doi: 10.1074/jbc.M109.057539 PMID: 20061574
10. González-Pérez B, Lucas M, Cooke LA, Vyle JS, la Cruz de F, Moncalián G (2007) Analysis of DNA processing reactions in bacterial conjugation by using suicide oligonucleotides. The EMBO Journal 26: 3847–3857. doi: 10.1038/sj.emboj.7601806 PMID: 17660746
11. Grandoso G, Avila P, Cayón A, Hernando MA, Llosa M, la Cruz de F (2000) Two active-site tyrosyl residues of protein TrwC act sequentially at the origin of transfer during plasmid R388 conjugation. Journal of Molecular Biology 295: 1163–1172. doi: 10.1006/jmbi.1999.3425 PMID: 10653694
12. Carballeira JD, González-Pérez B, Moncalián G, la Cruz de F (2014) A high security double lock and key mechanism in HUH relaxases controls oriT-processing for plasmid conjugation. Nucleic Acids Research 42: 10632–10643. doi: 10.1093/nar/gku741 PMID: 25123661

13. Edwards JS, Betts L, Frazier ML, Pollet RM, Kwong SM, Walton WG, et al. (2013) Molecular basis of antibiotic multiresistance transfer in Staphylococcus aureus. Proc Natl Acad Sci USA 110: 2804–2809. doi: 10.1073/pnas.1219701110 PMID: 23359708

14. Boer R, Russi S, Guasch A, Lucas M, Blanco AG, Pérez-Luque R, et al. (2006) Unveiling the Molecular Mechanism of a Conjugative Relaxase: The Structure of TrwC Complexed with a 27-mer DNA Comprising the Recognition Hairpin and the Cleavage Site. Journal of Molecular Biology 358: 857–869. doi: 10.1016/j.jmb.2006.02.018 PMID: 16540117

15. Lujan SA, Guogas LM, Ragonese H, Matson SW, Redinbo MR (2007) Disrupting antibiotic resistance propagation by inhibiting the conjugative DNA relaxase. Proc Natl Acad Sci USA 104: 12282–12287. doi: 10.1073/pnas.0702760104 PMID: 17630285

16. Monzingo AF, Ozburn A, Xia S, Meyer RJ, Robertus JD (2007) The Structure of the Minimal Relaxase Domain of MobA at 2.1 Å Resolution. Journal of Molecular Biology 366: 165–178. doi: 10.1016/j.jmb.2006.11.031 PMID: 17157875

17. Pansegrau W, Schröder W, Lanka E (1993) Relaxase (TraI) of IncP alpha plasmid RP4 catalyzes a site-specific cleaving-joining reaction of single-stranded DNA. Proc Natl Acad Sci USA 90: 2925–2929. PMID: 8385350

18. Pansegrau W, Schröder W, Lanka E (1994) Concerted action of three distinct domains in the DNA cleaving-joining reaction catalyzed by relaxase (TraI) of conjugative plasmid RP4. J Biol Chem 269: 2782–2789. PMID: 8300611

19. Pansegrau W, Lanka E (1996) Mechanisms of initiation and termination reactions in conjugative DNA processing. Independence of tight substrate binding and catalytic activity of relaxase (TraI) of IncPalpha plasmid RP4. J Biol Chem 271: 13068–13076. PMID: 8662726

20. Moncalián G, la Cruz de F (2004) DNA binding properties of protein TrwA, a possible structural variant of the Arc repressor superfamily. Biochimica et Biophysica Acta (BBA)—Proteins and Proteomics 1701: 15–23. doi: 10.1016/j.bbapap.2004.05.009

21. Vega-Rocha S, Byeon I-JL, Gronenborn B, Gronenborn AM, Campos-Olivas R (2007) Solution Structure, Divalent Metal and DNA Binding of the Endonuclease Domain from the Replication Initiation Protein from Porcine Circovirus 2. Journal of Molecular Biology 367: 473–487. doi: 10.1016/j.jmb.2007.01.002 PMID: 17275023

22. Campos-Olivas R, Louis JM, Clerot D, Gronenborn B, Gronenborn AM (2002) The structure of a replication initiator unites diverse aspects of nucleic acid metabolism. Proc Natl Acad Sci USA 99: 10310–10315. doi: 10.1073/pnas.152342699 PMID: 12130667

23. Boer DR, Ruíz-Masó JA, López-Blanco JR, Blanco AG, Vives-Llàcer M, Chacón P, et al. (2009) Plasmid replication initiator RepB forms a hexamer reminiscent of ring helicases and has mobile nuclease domains. The EMBO Journal 28: 1666–1678. doi: 10.1038/emboj.2009.125 PMID: 19440202

24. Khan SA (2005) Plasmid rolling-circle replication: highlights of two decades of research. Plasmid 53: 126–136. doi: 10.1016/j.plasmid.2004.12.008 PMID: 15737400

25. Yoshida H, Furuya N, Lin Y-J, Güntert P, Komano T, Kainosho M (2008) Structural Basis of the Role of the NikA Ribbon-Helix-Helix Domain in Initiating Bacterial Conjugation. Journal of Molecular Biology 384: 690–701. doi: 10.1016/j.jmb.2008.09.067 PMID: 18929573

26. Varsaki A, Moncalián G, Garcillán-Barcia MDP, Drainas C, la Cruz de F (2009) Analysis of ColE1 MbeC unveils an extended ribbon-helix-helix family of nicking accessory proteins. Journal of Bacteriology 191: 1446–1455. doi: 10.1128/JB.01342-08 PMID: 19114496

27. Moncalián G, Valle M, Valpuesta JM, la Cruz de F (1999) IHF protein inhibits cleavage but not assembly of plasmid R388 relaxosomes. Mol Microbiol 31: 1643–1652. doi: 10.1046/j.1365-2958.1999.01288.x PMID: 10209739

28. Simmel FC (2012) DNA-based assembly lines and nanofactories. Current Opinion in Biotechnology 23: 516–521. doi: 10.1016/j.copbio.2011.12.024 PMID: 22237015

29. Yang YR, Liu Y, Yan H (2015) DNA Nanostructures as Programmable Biomolecular Scaffolds. Bioconjugate Chem 26: 1381–1395. doi: 10.1021/acs.bioconjchem.5b00194

30. Agundez L, Machon C, Cesar CE, Rosa-Garrido M, Delgado MD, Llosa M (2010) Nuclear Targeting of a Bacterial Integrase That Mediates Site-Specific Recombination between Bacterial and Human Target Sequences. Applied and Environmental Microbiology 77: 201–210. doi: 10.1128/AEM.01371-10 PMID: 21037296

31. Scherzinger E, Lurz R, Otto S, Dobrinski B (1992) In vitro cleavage of double- and single-stranded DNA by plasmid RSF1010-encoded mobilization proteins. Nucleic Acids Research 20: 41–48. PMID: 1738602

32. Llosa M, Bolland S, la Cruz de F (1991) Structural and functional analysis of the origin of conjugal transfer of the broad-host-range IncW plasmid R388 and comparison with the related IncN plasmid R46. Mol Gen Genet 226: 473–483. PMID: 2038309

33. Bolland S, Llosa M, Avila P, la Cruz de F (1990) General organization of the conjugal transfer genes of the IncW plasmid R388 and interactions between R388 and IncN and IncP plasmids. Journal of Bacteriology 172: 5795–5802. PMID: 2170327

All the figures and tables that appear in this thesis have been reprinted with the permission from the publisher in both print and electronic format.