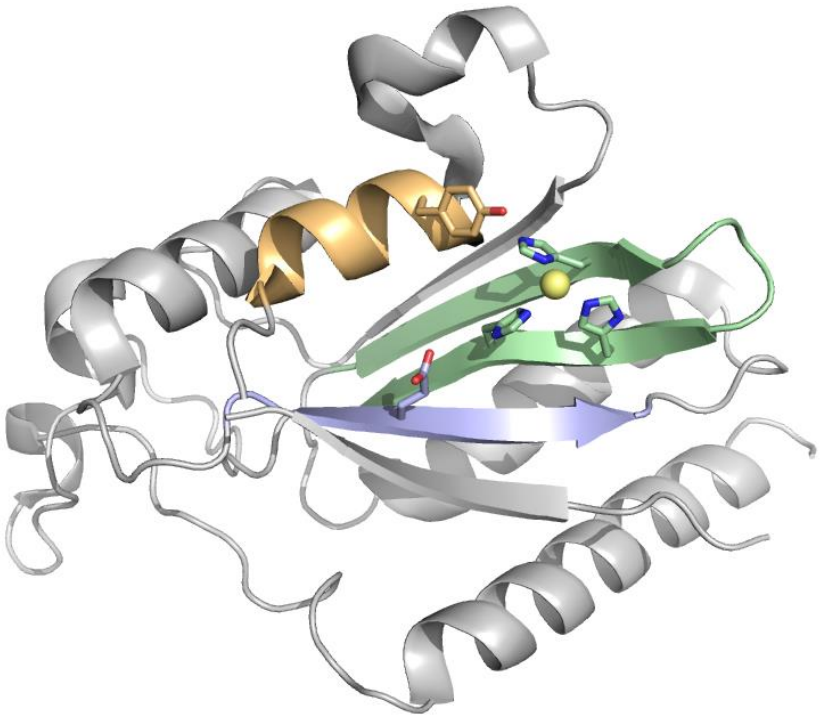
Identification of residues indirectly involved in cation coordination by HUH endonucleases



Lorena González Montes

Master in Molecular Biology and Biomedicine

Department of Molecular biology, UC

Department of Microbiology and Genomics, IBBTEC

June 2015

Supervised by Gabriel Moncalián

ACKNOWLEDGMENTS

First of all, I am especially grateful to Gabriel Moncalián for guiding me on this work, putting faith on me and lot of time and knowledge in helping me with the project. I would also like to thank Sandra Sagredo for introducing me into the relaxases world and for helping me at the first stages of the project, giving me priceless advice. I take this opportunity to express my sincere gratitude to the rest of the Protein Engineering group members, especially Laura Giner, Omar Santín and Mati Cabezas, for their experience and support, and for all the good moments shared at the lab. Finally, I want to thank the Intergenomics and Molecular Motors groups for their helpful discussions. I also acknowledge the European Commission for supporting EvoTAR project to Fernando de la Cruz and the *Ministerio de Cultura y Deporte* for the collaboration grant that I received. And as they all know, I really appreciate the support of my family and closer friends, who always believe in me and make my life easier.

ABSTRACT

Metal ions are necessary in about one third of all proteins, playing both functional and structural roles. HUH endonucleases are metalloproteins containing conserved His-hydrophobic-His residues at the active site. However, the nature of the metallic cofactor is not identical in all members of this family. TrwC, a protein involved in the single strand DNA (ssDNA) processing during bacterial conjugation, is one of the most studied HUH endonucleases at a biochemical and structural level. The physiological cofactors of TrwC relaxase ($TrwC_R$) are Mg^{+2} and Mn^{+2} while other relaxases only use Mn^{+2} as a cofactor. *In vitro* assays demonstrated that cleavage ability of $TrwC_R$T87I is notably reduced in the presence of Mg^{+2}. Moreover, structural analysis showed that in TrwCT87I the orientation of H163 has changed. In order to evaluate the role of the catalytic ion in vivo, conjugation assays were performed. Strikingly, no differences in conjugation frequencies were observed when using TrwC T87I. Thus, TrwC probably has evolved to be active using Mg^{+2} or Mn^{+2} as cofactors to increase the probability to find an appropriate metal in the conjugative environment. These findings open a new path for modification of cofactor specificity and de novo protein design.

INDEX

BACKGROUND

Metalloproteins

Metal ions are necessary in about one third of all proteins, playing both functional and structural roles (Peacock, 2013). The fact that almost half of the known enzymatic reactions involve metals reveals its high impact in the biochemistry of living systems. Indeed, metals are required as cofactors in such important processes as photosynthesis, respiration or N_2 fixation. Thus, metalloproteins are present in all the enzymatic families. It has been estimated that 44% of oxidoreductases, 40% of transferases, 39% of hydrolases, 36% of lyases, 36% of isomerases and 59% of ligases rely on a metal cofactor to be active (Waldron *et al*., 2009).

The shape of the active site is very important in enzymes, and so it is in metalloenzymes. The metal ion is usually located in the active site pocket, where it perfectly fits. In metalloproteins, metals are usually coordinated by nitrogen, oxygen or sulfur atoms that act as donor groups and belong to the amino acid side chains of the residues found at the active site of the protein. Moreover, the donor group can also be provided by the peptide backbone. The most common donor groups are:

- Imidazole (a nitrogen donor atom) found in histidine residues

- Thiolate (sulfur donor atom) coming from cysteine residues

- Carboxylate groups (oxygen donor atom) provided by aspartic acid and glutamic acid

Big effort is being made to generate functional artificial metalloproteins *de novo* able to perform desired functions not necessarily already present in nature (Peacock, 2013). For advanced engineering of metalloenzymes much knowledge is still required. Therefore, a lot of work is being made to understand the organization of native metalloenzymes, to understand structure-function relationship, and to mimic them. With all this knowledge, and different techniques as random evolution of proteins or even non-natural amino acids incorporation, scientist are trying to improve metalloproteins, enhance metal-binding affinity leading to more active proteins, create metalloproteins with new activities, new substrate requirements, etc. (Petrik *et al*., 2014).

HUH endonucleases

The HUH endonuclease superfamily is a group of metalloproteins in charge of breaking and joining single-stranded DNA at different biological processes such as viral DNA replication, conjugative plasmid transfer or different types of DNA transposition. There are two main groups in this superfamily: *Rep* proteins involved in replication, and *Mob* proteins or relaxases implicated in plasmid conjugation (Chandler *et al.*, 2013).

Structurally, what they have in common is an HUH domain where two histidines are separated by a hydrophobic residue and a Y motif, which can contain one or two catalytic tyrosines. The active site is formed by four or five antiparallel β-sheets, where the HUH motif is found, and usually sandwiched between several α-helices, at one of which the catalytic Tyr is placed. In addition to this HUH domain, these endonucleases usually have other associated domains, as primase, helicase or other unknown-function domains (Chandler *et al*., 2013). These additional domains will allow HUH endonucleases to cover a diverse range of functions.

The cleavage and ligation mechanism shared by HUH endonucleases is as follows. It starts with the action of a polar residue in charge of the activation of the hydroxyl group of the catalytic tyrosine by proton abstraction (Asp 85 and Tyr18 in Fig. 1). The next step is the nucleophilic attack of the catalytic tyrosine to a specific phosphate of the DNA strand to be cleaved, creating a 5'phosphotyrosine protein-DNA covalent complex and liberating a 3'OH free end at the cleavage site. This 3' hydroxyl group is then used to prime replication in some cases or to act as a nucleophile for strand transfer in the cases where the 5'phosphotyrosine intermediate needs to be resolved (Guasch *et al*., 2003; Boer *et al*., 2006).

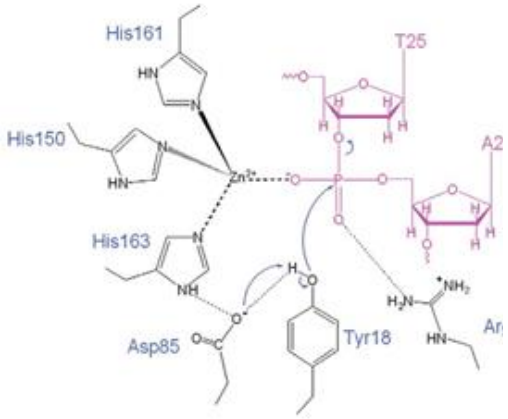


Figure 1. Schematic diagram depicting the DNA cleavage mechanism catalyzed by HUH endonucleases. Image took from Boer *et al.*, 2006.

The active site includes a metal divalent cation which function is thought to be the coordination of one of the oxygen atoms of the phosphate DNA to be attacked, polarizing it and making easier the nucleophilic attack of the hydroxyl group of the tyrosine (Figure 1; Boer *et al*.,

2006). This metal cofactor is coordinated at the active site by three polar residues, two of which are the HUH histidines.

Relaxases

Relaxases are HUH endonucleases of special interest due to their involvement in bacterial conjugation, the main horizontal gene transfer mechanism in prokaryotes. This process is the one responsible of bacterial genome evolution causing big worldwide problems related to antibiotic resistance genes dissemination (de la Cruz and Davies, 2000).

The role of relaxases is thus to transfer a ssDNA plasmid copy from a donor cell to a recipient one. The DNA conjugative transfer system is formed by the relaxosome, responsible of ssDNA cleavage at the *nic* site of the origin of transfer *(oriT)*, the pore of a type IV secretion system (T4SS) or transport channel and the coupling protein that connects both modules and pumps the DNA to the recipient cell. The genes coding for this machinery are found in conjugative or mobilizable plasmids that have a mobilization region (*Mob*) (Carballeira *et al.*, 2014).

The action mechanism of these enzymes starts with the recognition and strand-specific cleavage at the unique site known as the *nic* site of the *oriT* contained on the plasmid (Fig. 2). After the phosphotyrosine bond is created between the cleaved strand and the catalytic Tyr residue, the relaxase drives de ssDNA to the recipient cell through the pore. At the same time, the donor cell starts to synthetize the complementary strand until *nic* site is reconstructed. This *nic* site is then cleaved again by the relaxase in order to liberate a single strand circularized copy of the plasmid into the recipient cell that later on will be replicated there by the recipient cellular machinery (Chandler *et al*., 2013).

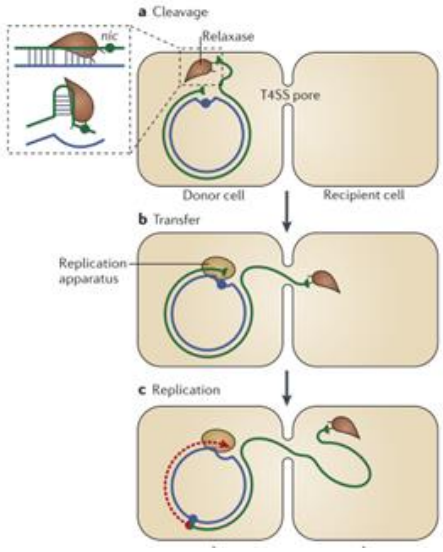


Figure 2. Schematic representation of the first conjugation steps. Image took from (Chandler *et al*., 2013).

In the majority of relaxases, the ability to recognize and bind a specific sequence in the plasmid and transfer it from one cell to another rests in the N-terminal domain, meanwhile it is the C-terminal domain the one which provides variable functions (Guasch *et al.*, 2003). According to the endonuclease domain, relaxases can be classified in six families, four of which (MobF, MobQ, MobP and MobV) have the HUH motif located on a β-sheet and a third polar residue to coordinate the metal ion (Glu, Asp, Gln or more often a His) in a contiguous β-sheet (motif III) and the Y motif placed on a α-helix (motif I). Another conserved motif can also be defined (motif II) with a conserved Asp or Glu needed to activate the hydroxyl group of the catalytic tyrosine by proton abstraction and that is expected to enhance the interaction of the His pocket with the metal ion (Chandler *et al.*, 2013). Motif II is placed in another β-sheet contiguous to the one containing the HUH motif. A generic relaxase structure is shown as cartoon in Fig. 3 with the catalytic motif I tyrosine, the three coordinating residues of motif III and the polar catalytic residue in motif II shown in sticks.

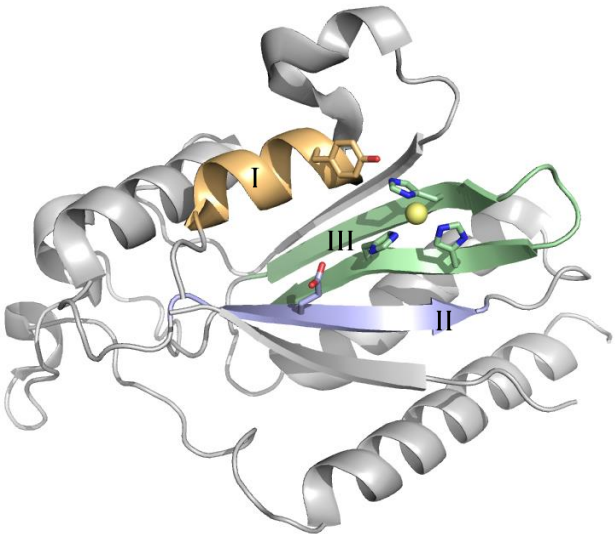


Figure 3. Cartoon representation of generic HUH relaxases colored by conserved structural motives. Roman numbers represent the conserved motifs I, II and III (see text). The main residues involved in the catalytic activity are shown in stick representation. A Mn atom (in yellow) is coordinated by three residues of motif III (PDB:2NS6).

Metal usage by relaxases

As already said, a metallic divalent ion is required for relaxases to act. It is thought to be needed for polarizing and localizing the scissile phosphodiester bond (Boer *et al.*, 2006). It is commonly thought that Mg^{+2} or Mn^{+2} are the physiological cofactors (Chandler *et al.*, 2013).

However, the metal usage is highly dependent on the protein as it can be noticed in Table 1 for some relaxases.

Table 1. Metal specificity of different relaxases.

Relaxase	Plasmid	Metals used	Metals not used	References
TraI	F	Mg^{2+}		Matson *et al*., 1993; Sherman and Matson, 1994
TraI	RP4	Mg^{2+}		Pansegrau *et al*., 1993
TaxC	R6K	Mg^{2+}		Núñez *et al*., 1997
MobA	R1162	Mg^{2+}, Mn^{2+}, Ca^{2+}, Ba^{2+}		Monzingo *et al*., 2006
TraA	pIP501	Mg^{+2}, Mn^{+2}		Kurenbach *et al*., 2002
NES	pLW1043	Ni^{2+}, Mn^{2+}, Co^{2+}, Cu^{2+}	Mg^{2+}	Edwards *et al*., 2013
MbeA	ColE1	Mg^{2+}, Co^{2+}, Ni^{2+}	Mn^{2+}	Varsaki *et al*., 2003
MobM	pMV158	Mg^{2+}, Mn^{2+}, Ca^{2+}	Ba^{2+}, Zn^{2+}	Guzman and Espinosa, 1997
TrwC	R388	Mg^{2+},Mn^{2+}, Ca^{2+},Zn^{2+},Ni^{2+},Cu^{2+}, Co^{2+}		Boer *et al*., 2006

OBJECTIVE

In this work we are interested in studying the role of the metal cation in the ssDNA cleavage process performed by HUH relaxases. Despite all of them having an HUH motif coordinating the metal ion, some of them have preference for specific cofactors. Thus, we would like to study, more specifically, the role of secondary residues in this metal coordination and to try to elucidate their structural involvement in the active site.

MATERIALS AND METHODS

HUH relaxases sequence alignment

The three structural conserved motifs of HUH endonucleases were analyzed by ClustalW alignment (Larkin *et al*., 2007). Relaxase amino acid sequences from the mobilizable plasmids already classified (Garcillán-Barcia *et al*., 2009) in the four different families that show a HUH motif were aligned. The plasmids used for each family (MobF, MobP, MobQ and MobV) are shown below.

Aligned relaxases belong to the following plasmids:

Mob**F:** R388 (**PDB model 1OMH**), pCT14, R46, pBI709, pWWO, pXcB, pAOVO01, pLPP, F, R100, pSLT, pKPN3, pED208, pASA5, pENTE01, pSWIT01, pREB1, pREB6, pREB5, pYJ016, pVT2, pMFLV03, pNG2, pCG4, pBMUL01,pSQ10 .

Mob**V:** pMV158 (**PDB model 4LVL**), pSSU1, pRW35, pPLA4, pLC88, pI4, pUB110, pTB19, pBM300, p1414, pE194, pBBR1, pA, Tn5520, Tn4555, pYHBi1, pSin9.7.

Mob**P:** RP4 (**PDB modelled with RaptorX**), PTFFC2, pRAS3.2, RMS149, Rms149, pRSB101, R64, P9, pSC138, pETEC_73, pSERB1, pET46, PCTX-M3, pTB11, pBS228, pB10, pADP-1, TRAI_R751, pB3, pA81, pCNB, pA1, pBP136, pKJK5, Plasmid2, Plasmid1, PEST4011, Plasmid_59kb, PS19503, PBI1063, PXAG81, Large Plasmid, PTCF14.

Mob**Q:** RSF1010 (**PDB model 2NS6**), PKLC102, pRL12, PTIC58, pTiBo542, pTiSAKURA, pRL7, pRi1724, PRL8, pSMED02, pNGR234a, Plasmid 3, PHCG3, pAt, pSymA, P42D, Plasmid1, PGOX3, pSF118-20, pWCFS103, PMRC01, PIP501, pRE25, PLW1071, pGS18, VRSAp, PSK41, pGO1, pMS260, pIE1130, pVM111, pTcM1, pSJ7.4, pAB6, P11745, PSC101, PAV2.

After alignment, different consensus sequences were represented using WebLogo (Crooks *et al.*, 2004) which is a graphical representation of an amino acid sequence alignment. Each logo consists of stacks of symbols for every sequence position. The overall height of the stack indicates the sequence conservation at that position, and the height of symbols within the stack represents the relative frequency of each amino acid at that position. The number of residues selected for each motif was related to the length of the secondary structure of the PDB from the model relaxase shown in bold letters.

Microorganisms

Escherichia coli DH5α was the strain used for cloning. *E.coli* C41 was chosen for protein overexpression. *E.coli* UB1637 strain was used for *in vivo* conjugations as recipient cell. Their characteristics are summarized in Table 2.

Table 2. E. coli strains used in this Project.

Strain	Genotype	Original strain	Reference
DH5α	F *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 ΔlacU169Φ80d lacZΔM15*)	*E. coli* K-12	Grant *et al*., 1990
C41	F^- *ompT* $hsdS_B$ (r_B^- m_B^-) *gal dcm (DE3)*	*E. coli* B	Miroux and Walker, 1996
UB1637	F^- *recA56 his lys trp rpsL*	*E. coli* K-12	de la Cruz and Grinsted, 1982

Electro-competent cells were prepared by culturing *E.coli* DH5α until a DO_{600} = 0.6 - 0.8 and then placed on ice for 30 minutes. Cultures were then centrifuged at 4,000 g for 15 minutes at 4 ºC and washed twice with *milli-Q* sterile water and a final wash with glycerol 10 % (w/v). Cells were resuspended in 2 % of the initial volume with glycerol 10 % (w/v). Cells were aliquoted and ultra-frozen on dry ice and ethanol and maintained at -80 ºC until use.

Culture medium and growing and storing conditions

The culture medium used for the growth and maintenance of *Escherichia coli* was LB medium (casein peptone 10 g/L, yeast extract 5 g/L and NaCl 10 g/L) as liquid medium and LB with agar 15 g/L *(Scharlau)* in the solid medium. The medium was sterilized by autoclaving at 120 ºC for 20 minutes. It was supplemented with the required selection antibiotic at a final concentration of 100 µg/mL of ampicillin, 50 µg/mL of kanamycin, 300 µg/mL of streptomycin or 20 µg/mL of nalidixic acid. The growing temperature for *E. coli* was 37 ºC. For the growth in liquid medium, the culture was maintained in agitation at 120 rpm overnight (*o/n*). In order to store the different *E.coli* strains, the pellet of a 5 mL *o/n* culture after centrifugation at 4000 rpm for 10 minutes was resuspended in 1.7 mL of a mixture of 50 % peptone 15 % (p/v) and 50 % glycerol (v/v) and stored at -20 ºC. For new cultures, an inoculum of 15 µL of these suspensions or an isolated colony of a previous culture plated and conserved at 4 ºC, was used.

Plasmids and plasmid constructs

The plasmid constructs used in this work are summarized in Table 3. The relaxase domain of the TrwC protein ($TrwC_R$, amino acids 1-293, MW= 32,840.7) with the T87I mutation was obtained by directed mutagenesis from pSU1588 as it is described in the correspondent section. The full length TrwC (amino acids 1-966, MW=107,676.0) mutant was obtained by mutating pSU1621 as described below. For *in vitro* relaxation assays pSU1186 was needed and pSU1445 was used for conjugation experiments.

Table 3. Plasmids used in this work

Plasmid	Description	Resistance	Size (Kb)	Reference
pET3a	Expression vector	Amp	4.6	Rosenberg *et al.*, 1987
pLGM0001	pET3a:: TrwC-N293 T87I	Amp	5.7	This work
pLGM0002	pET3a:: TrwC T87I	Amp	7.7	This work
pSU1186	pUC18::*oriT*(R388)	Amp	3.1	Llosa *et al.*, 1991
pSU1445	R388 $TrwC^-$	Kan	37.6	Llosa *et al.*,1994
pSU1588	pET3a:: TrwC-N293	Amp	5.7	Hernando, 2000
pSU1621	pET3a:: TrwC	Amp	7.7	Guasch *et al.*, 2003

Oligonucleotides

The sequences of the oligonucleotides used in this project for mutagenesis, plasmid sequencing, activity assays or crystallization are shown in Table 4.

Table 4. Oligonucleotides used in this project

Function	Oligonucleotide	Sequence (5'-3')
Mutagenesis	**T87I1**	GAGCGCATCGGCCTTGACCTCA**T**ATTCAGCGCGCCCAAGAGCGTA
	T87I2	TACGCTCTTGGGCGCGCTGAAT**A**TGAGGTCAAGGCCGATGCGCTC
Sequencing	**T7**	TAATACGACTCACTATAGGG
	pT7	GCTAGTTATTGCTCAGCGG
Activity assays	**12+18**	TGCGTATTGTCT/ATAGCCCAGATTTAAGGA
Crystallization	**23+0**	GCACCGAAAGGTGCGTATTGTCT/

Letters in bold indicate the mutation to be introduced at codon 87. The slash indicates the *nic* cleavage site.

DNA molecular biology techniques

Site directed Mutagenesis

$TrwC_R$ mutant was generated from construction pSU1588 by site directed mutagenic method adapting the *QuickChange II XL Site-Directed Mutagenesis Kit* protocol. *KOD Hot Start (Novagen)* polymerase was used to perform the synthesis of the mutated plasmid strands thanks to the T87I1 and T87I2 oligonucleotides carrying the mutations to be introduced. The PCR conditions are described below in Table 5 and 6. For conjugation assays it was necessary to generate the T87I mutant over the full length TrwC protein coding sequence. It was generated from pSU1621 with the same oligonucleotides and method at the PCR conditions described in Tables 5 and 6.

The PCR products mixed with DNA sample application buffer 4 X (bromophenol blue 0.2 %, sucrose 40 % in TBE 0.5 M) were analyzed by electrophoresis in 1% agarose gel with 0.05 µL/mL of *Midori Green Advance DNA stain (Nippon Gnenetics)* in TBE buffer 0.5 X (Tris-HCl 45 mM, boric acid 45 mM, EDTA 0.5 mM at pH 8.2) at 100 V for 30 minutes. The digestion of the original plasmid was done with *DpnI FD (Fermentas)* following the protocol given by the manufacturer (5 minutes 37 ºC + 5 minutes 80 ºC). Next, 60 µL of electrocompetent *E.coli* DH5α cells were transformed by electroporation with 2 µL of DNA (about 100 ng) by giving a 2.5 kV pulse. Cells were plated on solid LB with the selection antibiotic and cultivated at 37 ºC *o/n*.

Table 5. Reaction mix content for mutagenic PCR

	$TrwC_R$ mutagenesis		**full length TrwC mutagenesis**	
Component	**Stock conc.**	**Added volume (µL)**	**Stock conc.**	**Added volume (µL)**
***Mili-Q* water**	-	13.45	-	30.50
Buffer (*Novagen)*	10 X	2.00	10 X	5.00
dNTPs	10 mM	1.00	10 mM	5.00
T87I1	10 µM	1.00	10 µM	1.50
T87I2	10 µM	1.00	10 µM	1.50
DNA solution	50 ng/µL	0.50	4 ng/µL	2.50
$MgSO_4$	50 mM	0.80	25 mM	3.00
Polimerase (*Novagen)*	1 U/µL	0.25	1 U/µL	1.00
Final volume		20.00		50.00

Table 6. PCR programs used for mutagenesis

Step	$TrwC_R$ mutagenesis			full length TrwC mutagenesis		
	Temperature (ºC)	Duration (min)	Cycle Number	Temperature (ºC)	Duration (min)	Cycle Number
Initial Denaturalization	95	3		95	3	
Denaturalization	95	1		95	1	
Annealing	72	4.5	34	72	1	34
Elongation				70	3.5	
Termination	72	5		70	5	

Clone sequencing

Plasmid DNA was extracted from a number of transformed colonies by using the commercial kit *GeneJet plasmid miniprep kit (Thermo Scientific)* from 5 ml LB *o/n* cultures (inoculated with an isolated colony and the correspondent selection antibiotic) centrifuged for 10 minutes at 4,000 rpm. Around 500 ng of plasmid were sent to Macrogen Europe to be sequenced separately with the oligonucleotides T7 or pT7 at a final concentration of 2.5 µM. The results received were assembled and compared with the expected sequence using the *VectorNTI* software.

In vivo plasmid conjugations

Bacterial conjugation was performed between DH5α donor cells containing pSU1445 (R388ΔTrwC; Kan^R) plus the plasmid containing the full length TrwC WT (pSU1621) or TrwC T87I (pLGM0002) coding sequence (Amp^R) and UB1621 *E.coli* strain (Str^R) as receptors as shown in Fig. 4.

Overnight cultures of donors and receptors washed with LB were mixed at a 1:1 ratio (estimated by DO_{600}). Cell mixture was washed again with LB and incubated at 37 ºC for 1 hour over a pre-warmed solid LB agar layer. Conjugation was stopped by cell resuspension in LB. Serial dilutions by triplicate until 10^{-6} dilution were plated on LB agar medium with the relevant antibiotics for donors (Amp and Kan), receptors (Str) or trans conjugants (Kan and Str) selection. Plates were incubated at 37 ºC *o/n* and colonies counted. Conjugation frequencies were calculated by determining the number of trans conjugant colonies per donor colony. Three independent assays were done and the media and standard deviation of the triplicates was calculated.

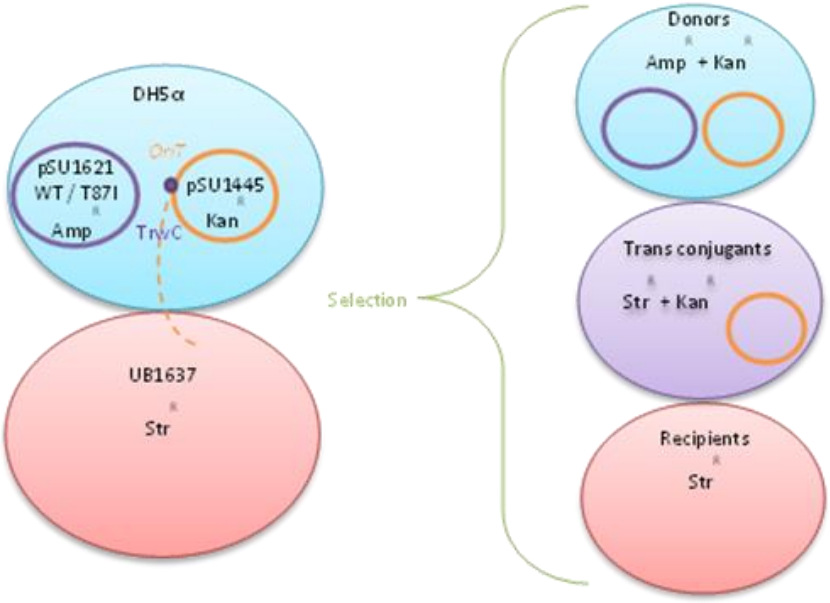


Figure 4. Schematic representation of the conjugation assay. It was performed by the complementation of pSU1445 (R388ΔTrwC) by pSU1621 coding the full length WT TrwC relaxase or pLGM0002 coding the full length T87I TrwC.

Protein overexpression

For protein overexpression, around 200 ng of plasmid DNA were electroporated into 70 µL of electro-competent *E.coli* C41 cells. After 1 hour in 1 mL LB medium at 37 ºC, 100 µL of cells were plated in solid LB medium with the selection antibiotic and incubated at 37 ºC *o/n*. Isolated colonies were inoculated in 25 mL LB medium flasks and incubated at 37 ºC and agitation at 120 rpm *o/n*. From these pre-inoculums, 1 L flasks were inoculated with 1/20 of the volume and were placed to grow at 37 ºC and 120 rpm until reaching a DO_{600} of approximately 0.7 (exponential grow state). After taking a 1 mL sample (t=0), overexpression was induced by the addition of 500 µL of 1 M IPTG to the flask. Following to 3 hours of incubation at the same conditions, another sample was taken (t=3h) and the remaining cells were collected by centrifugation at 4 ºC and 4,000 g for 15 minutes in the centrifuge 5810R *(Eppendorf)* and the rotor A-4-62. The cell pellet was frozen at - 80 ºC.

Protein overexpression was analyzed using the samples collected at different times. Samples were centrifuged for 5 min at 13,000 rpm and the pellet was resuspended in the same volume of protein sample application buffer 2X (400 mM Tris-HCl pH 6.8, 4 % SDS, 4 % β-mercaptoethanol, 30 % glycerol and 0.04 % bromophenol blue). Then, the mixture was boiled for 5

minutes for a subsequent analysis of the protein expression process by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) as it is described later on.

Protein purification

Protein purification was performed in three steps: cation exchange chromatography with phosphocellulose P-11 column for proteins that bind DNA, affinity chromatography with a *HiTrap™ Heparin HP* 5 mL (*GE Healthcare*) column and a molecular exclusion chromatography on a *Superdex™ 75 10/300 GL* (*GE Healthcare)* column to separate proteins by size.

First of all cells obtained in the 1 L culture were thawed at on ice, resuspended in 50 mL of lysis solution (Tris-HCl pH 7.5 100 mM, EDTA 1 mM, NaCl 500 mM and PMSF 1 %) and sonicated in a *Labsonic 2000 (B.Braun)* equipment at 50 % of potency for 3 cycles of 1.5 minutes at intervals of 1 minute on ice. The lysate was ultra-centrifuged at 40,000 rpm for 15 minutes at 4 ºC on a *Sorvall® WX Ultra Centrifuge Series (Thermo Scientific)* equipment. After taking 30 µL of lysate and pellet samples, the supernatant was diluted to 200 mM NaCl Tris-HCl pH 7.5 100 mM, EDTA 1 mM, and PMSF 1 % (Buffer A).

After lysis, ionic interchange chromatography was performed on a P-11 phosphocellulose column by batch separation. The phosphocellulose column was previously activated and washed with buffer A. Protein elution was performed in two fractions, first with elution buffer 1 (Tris-HCl pH 7.5 100 mM, EDTA 1 mM, NaCl 600 mM) and then with elution buffer 2 (Tris-HCl pH 7.5 100 mM, EDTA 1 mM, NaCl 1 M). Samples of 30 µL were taken from the flow through (FL, protein not bound) and from elutions 1 and 2. These samples were mixed with 30µL of protein charge buffer 2X.

The overexpression and purification process was followed by 12 % SDS-PAGE, performed at 180 V for 60 minutes in 1 X buffer (25 mM Tris base, 192 mM glycine and 0.05 % SDS). Gels were then stained on a staining solution (Coomassie brilliant blue R-250 0.1 % (p/v), methanol 50 % (v/v) and glacial acetic acid 10 % (v/v) for 30 minutes. Afterwards, gels were destained with a destaining solution (methanol 40 % (v/v) and glacial acetic acid 10 % (v/v)).

The fraction with the highest amount of protein was eluted at a NaCl concentration of 200 mM. Then, it was loaded onto an affinity chromatographic heparin column previously equilibrated with buffer 2. Elution was done by a lineal gradient between buffer A (Tris-HCl pH 7.5

100 mM, EDTA 1 mM, NaCl 200 mM) and B (the same as A but with NaCl 1 M) at a 4 mL/min flux and collecting 4 mL fractions in the *ÄKTA prime plus (GE Healthcare*) equipment.

After analyzing the collected fractions on a SDS-PAGE gel and being those quantified by spectrometry in *NanoDrop 2000c (Thermo Scientific)*, fractions with the highest amount of protein were pooled and, if needed, concentrated in a 10 KDa centricon concentrator *(Millipore)* by centrifugation at 4 ºC and 4,000 rpm until reaching the desired concentration. At the same time the protein buffer was changed to the desired gel-filtration buffer (Tris-HCl pH 7.5 100 mM, EDTA 1 mM, NaCl 300 mM) for the next purification step.

For the last step, protein separation by size was performed by pumping the protein through a molecular exclusion column, *SuperdexTM 75* previously equilibrated with gel filtration buffer. Separation was performed at a 0.4 mL/min flux and recollecting 0.5 mL fractions. Portions with the purest and highest protein concentration were aliquoted and ultrafrozen to be conserved at -80ºC.

DNA cleavage reactions analyzed by SDS-PAGE

In order to compare the activities of WT and T87I $TrwC_R$ proteins, cleavage of *nic*-containing oligonucleotides assay was performed as described. At those conditions at which the purified protein is able to recognize and cleave the *nic* site of the *oriT*, a DNA-protein covalent complex (Fig. 5) is going to be formed. This covalent complex migrates slower in a SDS-PAGE, allowing the identification of the percentage of protein that is covalently bound to the DNA related to the total protein.

Figure 5. Cartoon representation of $TrwC_R$-18mer complex after protein cleavage.

Reaction conditions for the cleavage assays were $TrwC_R$ WT or $TrwC_R$ T87I at 6 µM and 12+18 oligonucleotide (TGCGTATTGTCT/ATAGCCCAGATTTAAGGA) 15 µM in Tris-HCl pH 7.5 10 mM and 10 µM, 100 µM, 1 mM or 10 mM of metal (Mg^{2+}, Mn^{2+} or Ni^{2+}). No metal or EDTA 5mM were used as negative controls. The mix was incubated at 37ºC for 1 hour and then the cleavage reaction was stopped by the addition of protein loading buffer 2X and analyzed by visualizing the bands in a 12% SDS-PAGE gel after brilliant blue Coomassie staining. The cleavage percentage was calculated with the use of *Quantity one* software by dividing the retarded band intensity of the protein-oligonucleotide complex by the total protein intensity.

DNA relaxation assays

In addition to measuring the DNA cleavage ability of a oligonucleotide containing the *nic* site by active protein quantification, it was measured the ability to cleave one of the strains of a whole plasmid, the pSU1186 one, containing the R388 *oriT*. A schematic representation of the process is shown in Fig. 6.

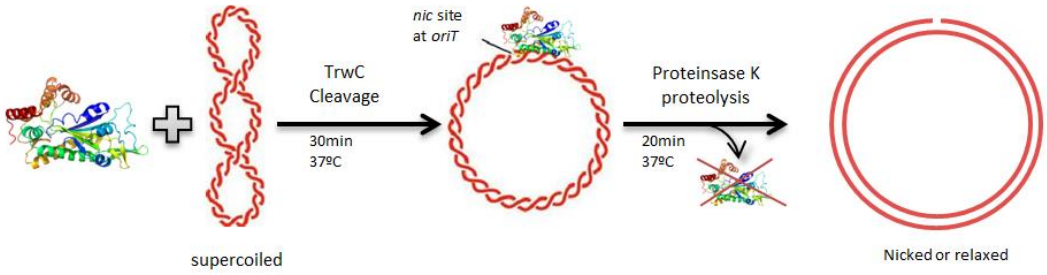


Figura 6. Schematic representation of the supercoiled *nic*-containing plasmid relaxation assay.

Reaction mixtures contained 10 nM pSU1186 of supercoiled plasmid DNA (Llosa, Bolland, and de la Cruz, 1991), and 500 nM $TrwC_R$ (WT or T87I) in 10 mM Tris-HCl pH 7.5 and 5 mM EDTA or 0, 10 µM, 100 µM or 1000µM $MgCl_2$ / $MnCl_2$. After 30 minutes of incubation at 37 ºC proteinase K and SDS were added to 1 mg/ml and 0.5 % (w/v) respectively. Samples were incubated for a further 20 minutes at 37 ºC. For visualizing the different DNA states, reactions products were loaded onto a 0.8 % (w/v) agarose gel and electrophoresed at 120V for 60 min in 0.5 x Tris-borate-EDTA buffer. The gel was further poststained with *Midori Green Advanced DNA Stain* and bands were visualized under UV light in a *Gel Doc system (BioRad)* and quantified using the *Quantity One software (BioRad)*. Different DNA migration was observed according to its state as schematically shown in Fig. 7.

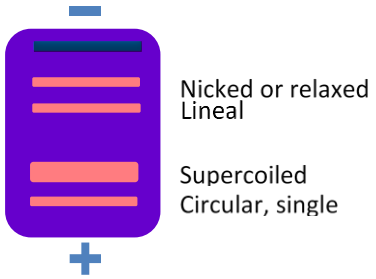


Figure 7. Differential DNA migration according to its state at an agarose gel.

Complex formation, Crystallization and X-ray data collection and processing

For the structural analysis of $TrwC_R$ bound to 23+0 DNA, $TrwC_R$ WT or $TrwC_R$ T87I (at 10 mg/mL in 300 mM NaCl, 100 mM Tris-HCl pH 7.5 and 1 mM ethylenediaminetetraacetic acid

(EDTA)) and 23+0 oligonucleotide (in Tris-HCl pH 7.5 20 mM, EDTA 1 mM, NaCl 500 mM heated at 95 ºC for 5 minutes and cooled at room temperature for 2 hours) were mixed at a 1:1.5 Protein:DNA molar ratio. Crystals (space group P65) were grown with sitting-drop vapor diffusion at 22ºC by mixing of 2 µL of $TrwC_R$-23+0 complex with 1µL of reservoir solution. A battery of 96 conditions was used *JCSG-plusTM HT-96 (Molecular dimensions)* for a first crystalizing screening. Once crystals were formed, those conditions were used for optimization in larger reservoirs until bigger crystals were formed.

Prior to data collection, crystals were treated by transferring them to cryoprotectant solution (20 % (v/v) ethylene glycol, 1.6 M ammonium phosphate, 80 mM Tris-HCl pH 8.5). Crystal mounted on a cryoloop were soaked in 100 mM metal solution and frozen at - 68ºC. Diffraction data was obtained at ALBA Synchrotron beamline XALOC and processed using MOSFLM and SCALA as part of the CCP4 package (Leslie, 2006; Winn *et al.*, 2011). Molecular replacement solutions in MolRep (CCP4) were obtained to solve the phase problem. For solving $TrwC_R$ bond to 23+0 oligonucleotide, $TrwC_R$ bound to a 25mer oligonucleotide already solved (Guasch et al., 2003) was used as a search model (PDB code 1OMH) and for T87I mutant, $TrwC_R$-23+0 complex was used as model. Refinement was performed in Phenix (Adams *et al.*, 2010) as well as the final density map superposition. Modelling step was done in COOT (Emsley *et al*., 2010).

RESULTS

Histidine-positioning residues are conserved within each relaxase family

In order to have a glimpse of the main structural features of HUH relaxases, an alignment of relaxases was done as described in Materials and Methods.

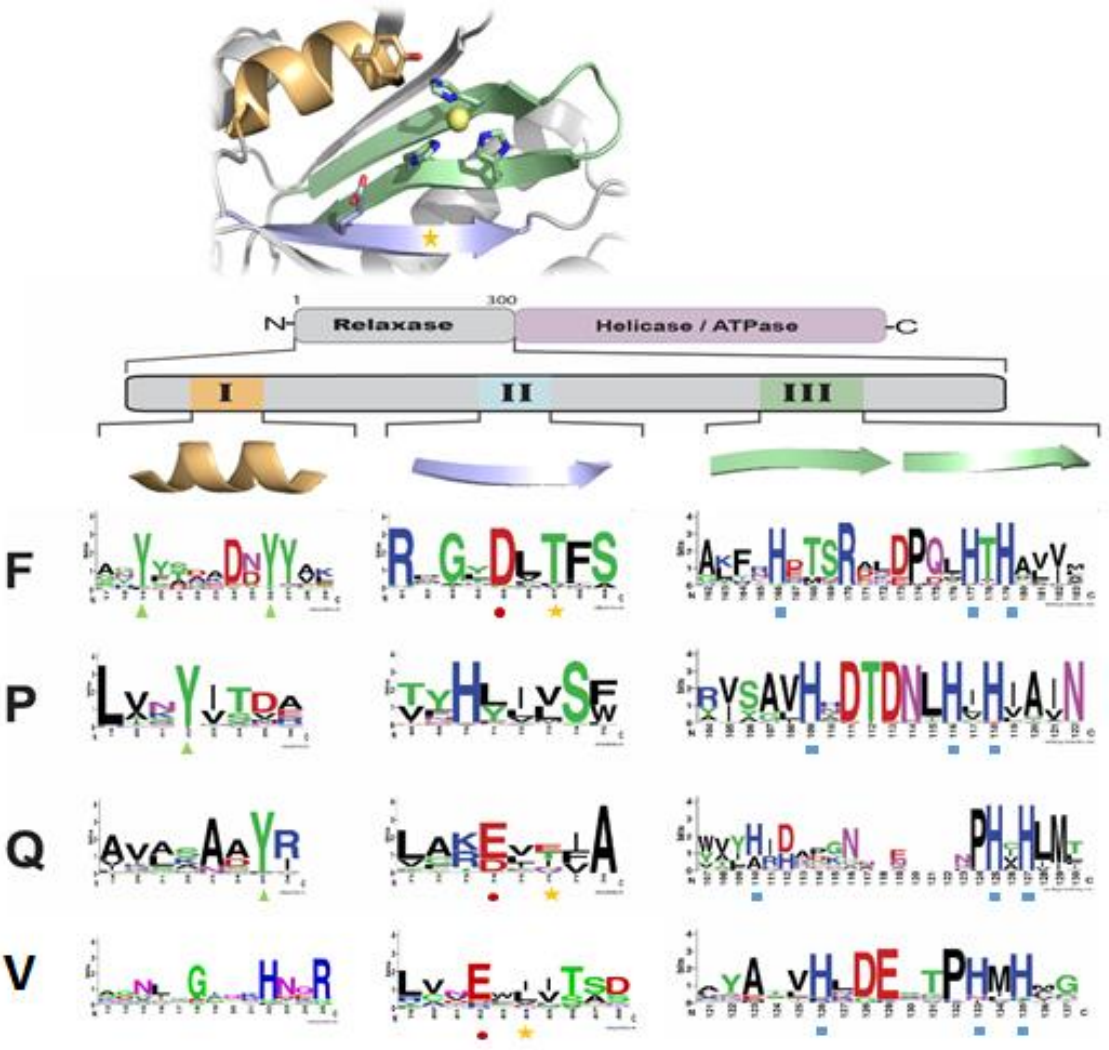


Figure 8. Description of the sequence and secondary structure of conserved motifs I, II and III of HUH relaxases. Image of generic HUH relaxase active site colored by conserved structural motives (PDB:2NS6). Mn atom is shown in yellow. The first 300 residues of the N-terminal domain of relaxases hold motif I in an alpha-helix, motif II in a beta-sheet and the three conserved histidines (H+HUH) of the motif III in two beta-strands. Sequence logo of the secondary structures of MOBF (F), MOBP (P), MOBQ (Q) and MOBV (V) family is shown as output by WEBLOGO. Each residue position contains stacks of letters depending on the relative frequency from a multiple sequence alignment using ClustalW. The number of residues of each motif is related to the length of the secondary structure of the model relaxase for each group. Green triangle = catalytic Tyr; red circle = catalytic acidic residue, yellow star = conserved residue with putative implication in metal coordination; blue square = residues directly implicated in metal coordination. For proteins aligned see Material and Methods section.

As it is observed in the logo representative of the relaxase alignment shown in Fig. 8, all relaxase families show the three histidines involved in cation coordination at motif III (shown as blue squares); the HUH ones and a third residue, usually another His that coordinate the metal. At motif I, the catalytic tyrosines are marked as green triangles. At motif II the catalytic acidic residue is pointed as a red circle and another residue at a position feasible to be involved in catalytic mechanism is shown as a yellow star.

Some differences respect the general structure described in introduction were observed between each relaxase family. These structural changes could have an important implication in active site mechanism and metal coordination.

MobV, although having the HUH motif, seems to have a Histidine or Asparagine instead of a Tyrosine for the nucleophilic attack. So, their action mechanism and maybe metal coordination way could be slightly different.

MobP motif II is structurally very different to the other three families' motif II. It does not has an acidic residue to help the activation of the catalytic Tyr. Its function is thought to be different, to recognize and bind the *nic* site at *oriT* but not to be involved in the catalytic cleaving and joining activity (Pansegrau *et al.,* 1993; Pansegrau and Lanka, 1996). Thus, we do not expect to find at this motif any residue indirectly involved in metal coordination.

MobQ together with MobF seems to be the family that better fits the general HUH endonucleases description, showing at least a catalytic Tyr, an acidic residue at motif II and the HUH motif. It also shows a combination of polar and apolar residues at position shown with a yellow star in Fig. 8, so it is a good family to hypothesize about metal coordination in HUH endonucleases.

The structure of two relaxases of the MobQ family are already solved, MobA (Monzingo, *et al.*, 2007) and NES (Edwards *et al*., 2013). It is reported that MobA shows *in vitro* activity with Mg^{2+} and Mn^{2+} meanwhile NES is only active in the presence of Mn^{2+} (Table 7). We have observed that this difference in metal requirements could be due to the amino acids less conserved inside each family that are present in this motif II at position marked as a yellow star in Table 7 and Fig. 8 that could also be indirectly implicated in metal coordination by hydrogen-bond interactions with some of the triad histidines.

Table 7. Comparison of three well studied relaxases.

Relaxase	Plasmid	Relaxase family	Acidic residue ●	Residue ★	Metals used	PDB structure	Reference
MobA	R1162	Q	E74	E76	Mg^{+2} Mn^{+2}	2NS6	Monzingo *et al*., 2007
NES	pLW1043	Q	E86	I88	Mn^{+2}	4HT4	Edwards *et al*., 2013
TrwC	R388	F	D87	T87	Mg^{+2} Mn^{+2}	1OMH	Boer *et al.*, 2006

Yellow star= amino acid at motif II at hydrogen bond distance from a catalytic His.

TrwC is a well-studied relaxase at both biochemical and structural level. It has its transferase activity at the N-terminal domain (amino acids 1 to 293) and a helicase activity at the C-terminal domain. TrwC relaxase domain is made of 11 β-strand, 11 α-helices and 7 turns. But are 5 stranded β-sheets and two α-helices the ones that form the core structure of the protein where the active site is located. As it is shown in Fig. 9, the active site of TrwC is formed by a catalytic tyrosine (Tyr18) and a histidine triad (His150, His161 and His163) where the metal ion is tetrahedrically coordinated. Asp85 is the acidic catalytic residue and Thr87 is the residue at a hydrogen bond distance from His163. TrwC has shown activity with Mg^{2+} and Mn^{2+} and has a polar amino acid interacting with a His (also shown in Table 7) as it is the case of MobA. Due to its extensive knowledge, TrwC could be a good model to prove if these putative residues involved in metal coordination at motif II are effectively implicated in metal coordination and specificity.

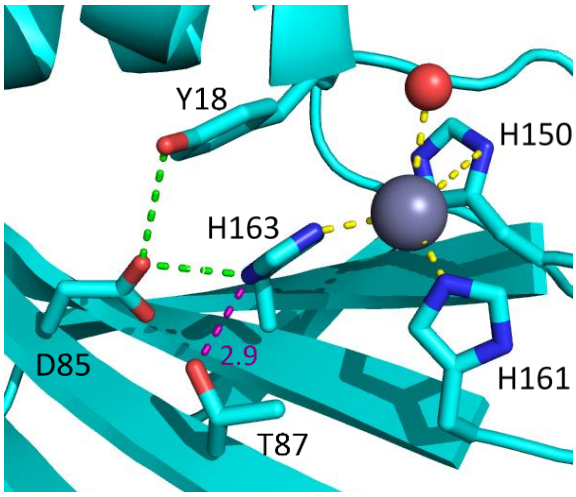


Figure 9. $TrwC_R$ active site showing the metal (grey sphere) coordination (yellow dashed lines) between the three histidines and a water molecule (red sphere). The hydrogen bonds formed with D85 and T87 are shown in green and purple dashed lines respectively (PDB:1QX0).

$TrwC_R$ T87I is not able to cleave *nic*-containing DNA in the presence of Mg^{+2}

As it has been described, an essential role of $TrwC_R$ T87 was expected in the nature of the cations used by TrwC. In order to prove the effect of having a polar (Thr) or apolar (Ile) residue at this position, pLGM0001 plasmid for the expression and purification of $TrwC_R$ with the mutation T87I was constructed (Fig. 10). Pure $TrwC_R$ WT and $TrwC_R$ T87I proteins were obtained after the expression and purification procedure described in Materials and Methods. $TrwC_R$ T87I behavior during the purification process was similar to $TrwC_R$ WT. Chromatogram showing the final purification step of both proteins is shown in Fig. 11.

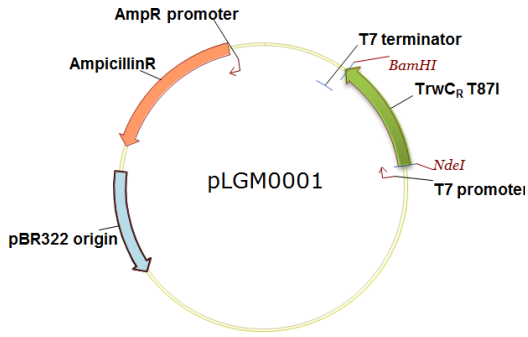


Figure 10. Schematic representation of the plasmid construct pLGM0001 created by site directed mutagenesis from pSU1588.

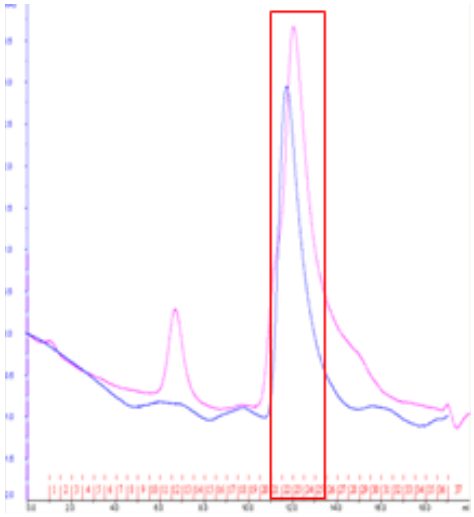


Figure 11. WT and T87I $TrwC_R$ protein purification by molecular exclusion chromatography (Absorbance 280 nm (mAU) *vs.* Elution volume (mL)). $TrwC_R$ WT chromatogram is shown in pink and $TrwC_R$ T87I is shown in blue. Pure $TrwC_R$ protein was obtained in fractions 21 to 25.

$TrwC_R$ WT and $TrwC_R$ T87I cleavage ability was evaluated against oligonucleotide 12+18 (TGCGTATTGTCT/ATAGCCCAGATTTAAGGA) in the presence of Mg^{2+}, Mn^{2+} and Ni^{2+} cofactors at different concentrations in the mix reaction. Negative controls with the metal chelation agent EDTA and without metal addition were also included. In those conditions at which the relaxase is able to bind and cleave the *nic* site of the *oriT*, a DNA-protein complex is formed that migrates slower on SDS-PAGE. As observed in Fig. 12, $TrwC_R$ WT protein is equally active in the presence of 10 μM Mg^{2+} or Mn^{2+}. On the other hand, $TrwC_R$ T87I is more active with Mn^{2+}. As

expected, $TrwC_R$ is not active in the absence of any metal. It can also be observed that $TrwC_R$ T87I is not active in nickel containing buffer. In Fig. 13 the percentage of cleavage (covalent complex per total protein) is represented for several concentrations, at which the same overall behavior was observed. An important decrease in cleavage activity is observed at high Ni^{2+} concentration.

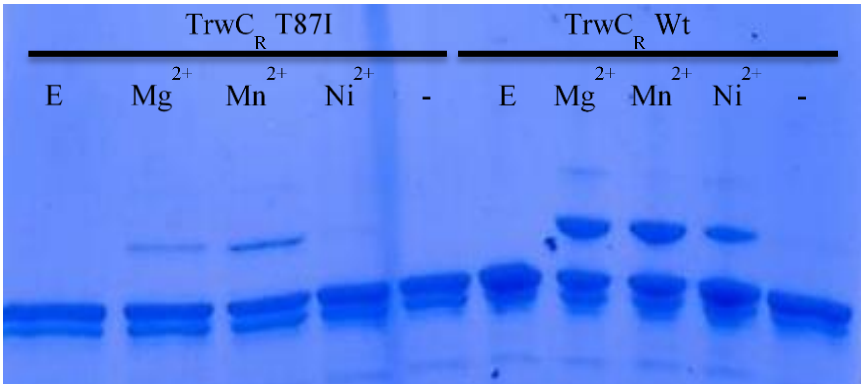


Figure 12. Differential *nic*-containing oligonucleotide cleavage yield by $TrwC_R$ WT *vs.* $TrwC_R$ T87I in the presence of different cofactors shown on a Coomasie-stained 12% SDS-PAGE gel. Protein (6 µM) and 12+18 oligonucleotide (15 µM) were incubated at 37 ºC for 1 hour in the presence of different cations at a 10 µM concentration, without metal or with EDTA in Tris-HCl pH 7.5 10 mM. Lane 1: $TrwC_R$(T87I) with EDTA 5mM (E). Lane 2: $TrwC_R$(T87I) with 10µM Mg^{2+}. Lane 3: $TrwC_R$(T87I) with 10 µM Mn^{2+}. Lane 4: $TrwC_R$(T87I) with 10 µM Ni^{2+} . Lane 5: $TrwC_R$(T87I) without adding any metal (-). Lane 6: $TrwC_R$ (WT) with EDTA 5mM. Lane 7: $TrwC_R$ (WT) with 10 µM Mg^{2+}. Lane 8: $TrwC_R$ (WT) with 10 µM Mn^{2+}. Lane 9: $TrwC_R$ (WT) with 10 µM Ni^{2+}. Lane 10: $TrwC_R$ (WT) without adding any metal.

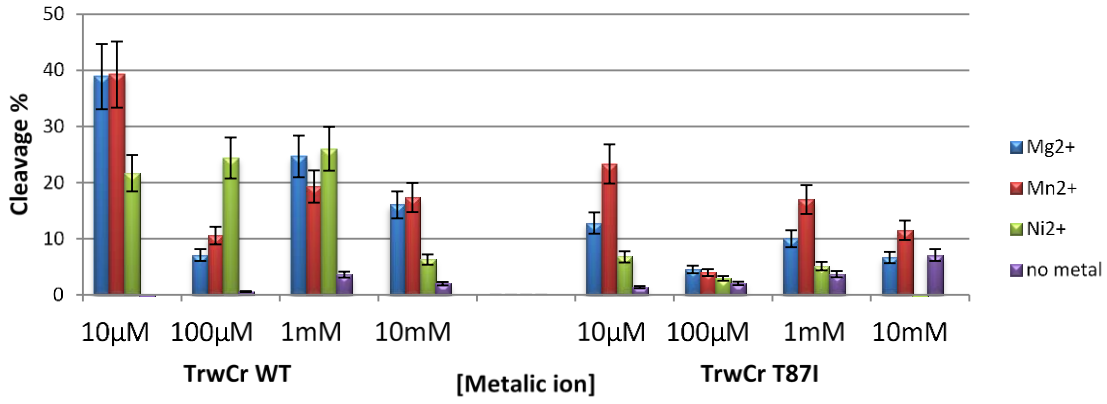


Figure 13. Effect of different cations in the oligonucleotide cleavage by $TrwC_R$ WT and $TrwC_R$ T87I at different cofactor concentrations. The assays were performed by incubating for 1 hour 12+18 oligonucleotide (TGCGTATTGTCT/ATAGCCCAGATTTAAGGA) 15 µM and protein 6 µM at 37 ºC in Tris-HCl pH 7.5 10 mM and metals at 0 (no metal), 10 µM, 100 µM, 1 mM or 10 mM. Mean and standard deviation of three independent experiments are represented in the graph.

The ability to cleave and relax a supercoiled plasmid (sc) containing the *nic* site has been analyzed too. This experiment is more similar to the *in vivo* conditions than oligonucleotide cleavage. In this assay it was analyzed how $TrwC_R$ cleaves one of the sc DNA strands and after proteinase K treatment to digest the protein, how the plasmid is unwound. This nicked and relaxed plasmid has a lower electrophoretic mobility than in a supercoiled state, and this behavior can be seen in an agarose gel. We observed that the amount of relaxed DNA by $TrwC_R$ WT is similar in the presence of Mg^{2+} than with Mn^{2+}. By contrast, $TrwC_R$ T87I mutant cleaves sc DNA with a higher efficiency when Mn^{2+} is present as cofactor (Fig. 14).

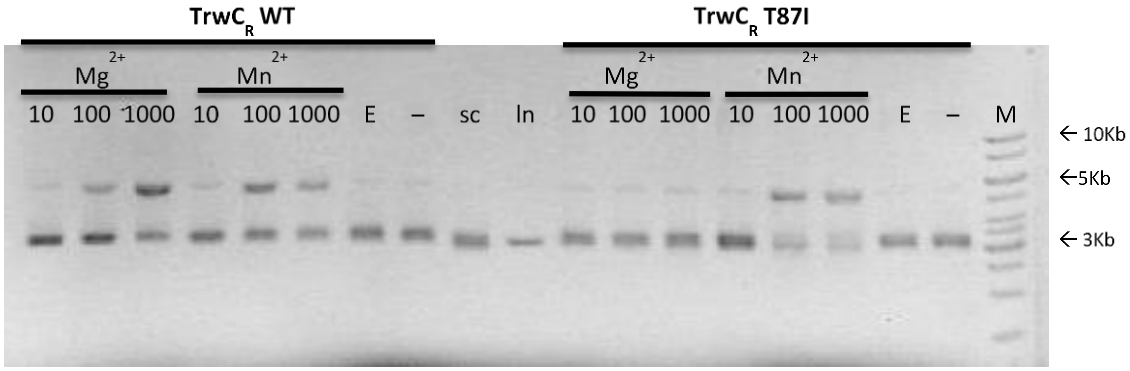


Figure 14. Effect of different metallic ions in the cleavage of a supercoiled (sc) plasmid containing the *oriT* by $TrwC_R$ WT or T87I mutant at different cofactor concentrations. The relaxation assays were carried out by incubating for 30 minutes at 37 ºC pSU1186 plasmid 10 nM with $TrwC_R$ 500 nM in Tris-HCl pH 7.5 10 mM buffer with 0 (-), 10 µM, 100 µM, 1000 µM or in the presence of EDTA (E) chelator 5 mM. After the digestion of the protein with SDS and proteinase K the relaxed (upper band) and supercoiled form (lower band) of the DNA were resolved in an agarose gel of 0.8 %. Supercoiled (sc) and linear (ln) pSU1186 without protein and *GeneRuler*^{TM} 1 Kb DNA ladder (M) were also loaded in the gel.

TrwC T87I is perfectly able to efficiently transfer DNA in conjugation.

The ability of TrwC WT *vs.* TrwC T87I to mobilize a plasmid *in vivo* was analyzed as well. A plasmid coding the TrwC WT (pSU1621) or T87I full length protein (pLGM0001) was introduced in *E.coli* DH5α cells containing pSU1445 plasmid (coding all the conjugative components except TrwC) to act as donor strains. *E. coli* UB1637 was used as receptor strain for conjugations. pSU1621 coding TrwC WT is able to promote pSU1445 plasmid transfer and thus, it was used as positive control. On the other side pET3a plasmid is not able to complement pSU1445 plasmid (conjugation frequency lower than 10^{-6} trans conjugants/donor) and it was used as negative conjugative control. Despite being a highly variable experiment, the conjugation frequencies

obtained at each independent assay performed by triplicate showed no difference between WT and T87I protein as can be observed in Fig. 15.

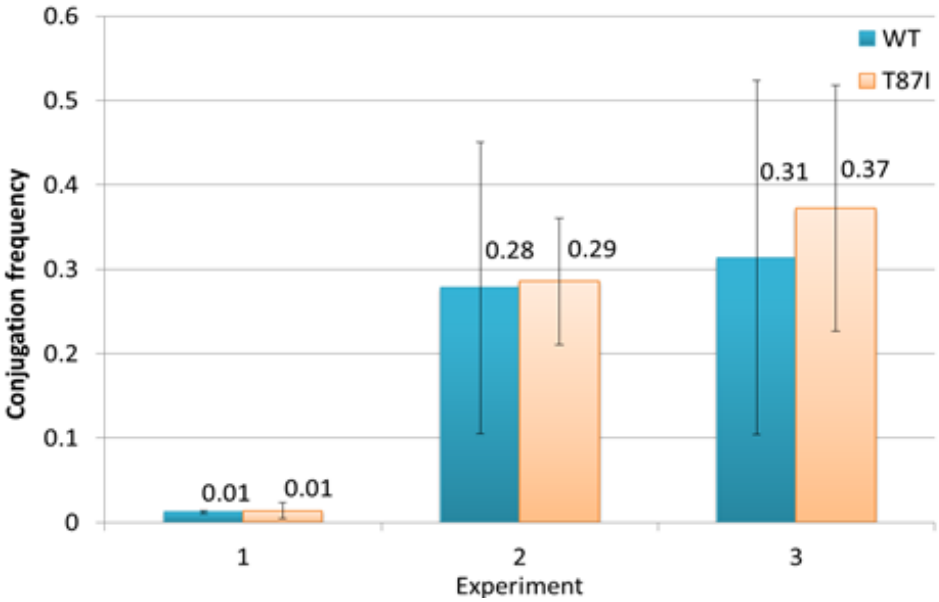


Figure 15. Comparison of the conjugative capacity of TrwC WT *vs*. TrwC T87I. Mean of conjugation frequency (trans conjugants / donor) and standard deviation of three independent experiments done by triplicate are shown in the graph.

$TrwC_R$ T87I structure shows a slight change in the H163 orientation

In order to analyze the effect of T87I mutation on TrwC structure and metal coordination, we have solved the structure of $TrwC_R$ T87I bound to a 23+0 oligonucleotide in the presence of different metal ions. Several crystals (Fig. 16) were obtained of the $TrwC_R$-23+0 complex as described in Materials and Methods. They were obtained in the conditions with precipitant solutions containing Tris-HCl pH=8.5 0.05 M, NH_4PO_4 2 M and Mn^{2+} at 5mM. Crystallographic data from this crystals collected at synchrotron is shown in Table 8.

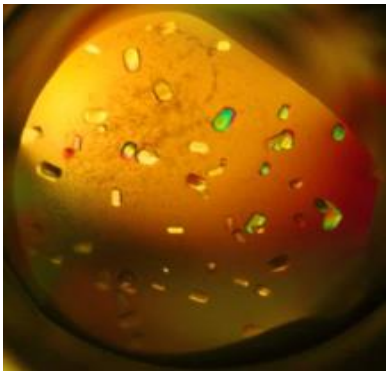


Figure 16. $TrwC_R$ T87I – 23+0 crystals obtained by sitting drop method at a 1:2 protein: 23+0 oligonucleotide molar ratio and at 22 ºC. Precipitant: $(NH_4)_3PO_4$ 2M, Tris-HCl pH=8.5 0.05 M and Mn^{2+} at 5 mM.

Table 8. Crystallographic data collection and analysis

	TrwC$_R$T87I 23+0	**TrwC$_R$WT 23+0**
Space group	$P6_5$	$P6_5$
Unit cell dimensions:		
a	149.3	149.3
b	149.3	149.3
c	78.3	78.3
Data range (last shell)	20-1.6 (1.69-1.6)	37-2.2 (2.32-2.2)
Observations (unique)	2435404 (131025)	956138 (50565)
Completeness (%) (last shell)	99.9 (100.0)	100.0 (100.0)
R_{sym}[a] (last shell)	0.063 (0.373)	0.099 (0.376)
Non-hydrogen atoms (solvent molecules)	5459 (856)	5969 (531)
R_{cryst}[b] (last shell)	0.192 (0.212)	0.223 (0.233)
R_{free}[c] (last shell)	0.170 (0.172)	0.207 (0.208)
r.m.s.[d] bond length (A°)	0.009	0.008
r.m.s.[d] bond angles (deg.)	1.246	1.043

[a] R_{sym} is the unweighted R value on I between symmetry mates.

[b] $R_{cryst}=\Sigma_{hkl}((F_{obs}(hkl))-(F_{calc}(hkl))/\Sigma_{hkl}(F_{obs}(hkl))$

[c] R_{free} is the cross-validation R factor for 5% of reflections against which the model was not refined.

[d] r.m.s. is the root mean square.

From this data we could solve the structure of both WT and T87I $TrwC_R$ proteins. In Fig. 17 both models are superposed so the structural change can be easily observed. Both structures almost perfectly match except at the mutated position 87 and at residue 163. It is highlited how hystidine 163 Nδ1 of WT $TrwC_R$ (shown in blue) is forming an hydrogen bond with T87, pushing the imidazole ring to be rotated to face the threonin residue. When in $TrwC_R$ T87I mutant protein (shown in orange) an apolar isoleucine is placed at position 87 the hydrogen bond can not be formed and the hystidine 163 imidazole ring is rotated about 30 º. The electronic density map of these two residues is shown in Fig. 18, where the same mutational effect can be observed.

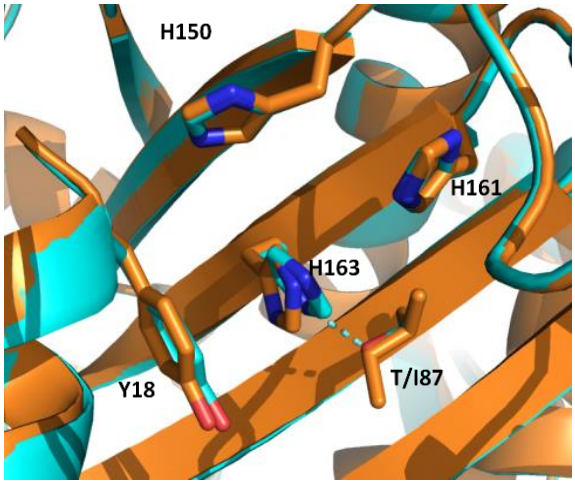


Figure 17. Superposed models of $TrwC_R$ active site showing the mutation and its consequence on His163 orientation. $TrwC_R$ WT in shown in blue and $TrwC_R$ T87I is shown in orange. Hydrogen bond in $TrwC_R$ WT is shown in blue.

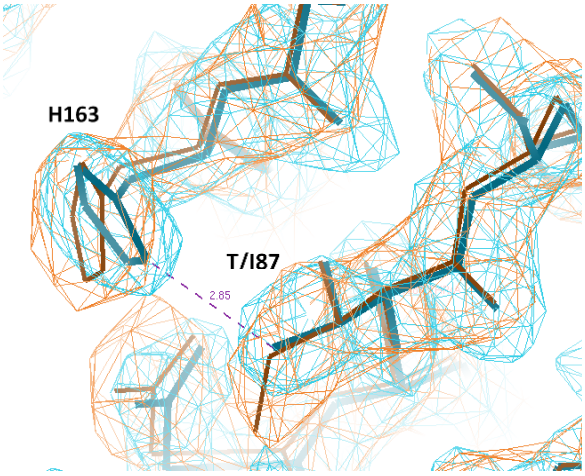


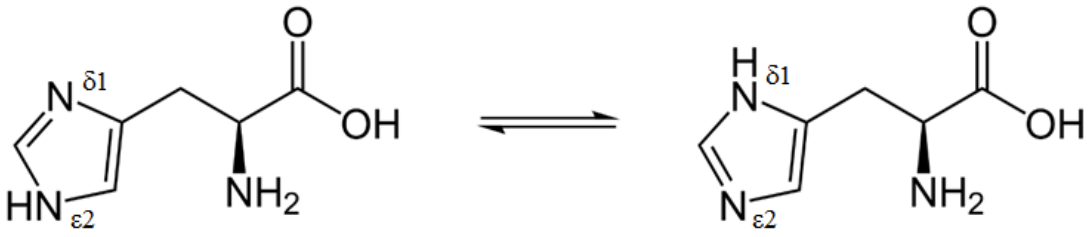
Figure 18. Electronic density map of residues H163 and T/I87 of $TrwC_R$ WT (blue; 2.2 Å) and $TrwC_R$ T87I (orange; 1.6 Å). Hydrogen bond distance between His163 and Thr87 in WT protein is shown in purple.

DISCUSSION

Role of the residues surrounding the catalytic histidines in the conformation of the active site

In this work we have studied the role of residues surrounding the catalytic histidines in HUH endonucleases. Coordination through either of the two imidazole ring nitrogen atoms in histidines (Nδ1 or Nε2, see Fig. 19) is possible in metalloenzymes and thus in HUH endonucleases. Mixed coordination is usually found in active sites, like in TrwC where the divalent cation is tetrahedrically coordinated between a water molecule and the Nδ1 of H150, the Nδ1 of H163 and the Nε2 of H161. This is the coordination geometry previously reported for TrwC, independently of the metal bound. However, Mg^{2+} and Mn^{2+} usually present an octahedral coordination sphere, being TrwC an exception inside the Mg^{2+} -binding proteins group (Boer *et al*., 2006).

Other studies state that when bound to a metal ion, the Nδl-protonated and Nε2 ligand-bound are the predominant forms as they are energetically more stable due to steric factors (Chakrabarti, 1990). The orientation of histidine residues is usually stabilized through hydrogen bonding in a number of metalloenzymes, and this bond affects the strength of the metal-ligand bond (Chakrabarti, 1990) as it is in the case of the TrwC H163. When this bond is formed, the tautomeric equilibrium is displaced towards the more stable Nε2 ligand atom form and thus, Mg^{2+} metals with lower binding affinities can be used too. The two tautomeric forms of histidine amino acid are shown in Fig. 19. However, when H163 Nδ1 cannot form a stable hydrogen bond, then the affinity of H163 Nε2 for Mg^{2+} is reduced.



Nε2-H tautomer Nδ1-H tautomer

Figure 19. Tautomeric forms of histidine

Taking this into account we propose that depending on the polar or apolar character of the amino acids close to the catalytic histidines, the tautomeric form of the histidines should change and thus, hydrogen bonds are formed or not, facilitating or making more difficult the metal coordination. This hypothesis has been corroborated in the crystal structure of $TrwC_R$T87I bound to 23+0. This differential charge distribution in the histidines could be associated to differential metal specificity.

The nature of residues around the catalytic histidines is important to define the metal bound.

Atomic sizes, ligand affinities, coordination geometries, redox states, and metal concentrations availability are some of the characteristics that impose the identity of the bound metals (Macomber and Hausinger, 2011). Proteins have naturally evolved to use the metal ions available at their environment. Nevertheless, some divalent metals tend to bind proteins more easily than others, following the order of stability described by Irving-Williams. Cooper and zinc form the tightest complexes followed by nickel and cobalt, then iron and manganese and finally calcium and magnesium forming the weakest ion-protein complexes (Waldron *et al*., 2009). We have observed that not all HUH endonucleases follow this general rule, but the majority of them have a higher preference for Mn^{2+} than for Mg^{2+}. In fact, previous work showed that TrwC relaxase has a binding affinity constant 10,000 times higher for Mn^{2+} than for Mg^{2+} ($TrwC_R$ K_d for Mg^{2+} =0.27± 0.32 mM and $TrwC_R$ K_d for Mn^{2+} = 9.7± 0.3 nM) (Lucas, M., 2006). Despite this fact, magnesium is the second most commonly used metal as cofactor (9% of the total amount of proteins) meanwhile manganese dependent proteins represent 6% of the enzymes (Waldron *et al*., 2009). This is expected to be due to the metal availability in the medium. Magnesium is a bulk element that appears to be at millimolar concentration inside cells (Foster *et al.*, 2014) meanwhile cytosolic concentration of manganese is much lower, being a trace element ($[Mg^{2+}]$=10 mM *vs.* $[Mn^{2+}]$= 10 µM) (Wieghardt, 1989).

Concerning to activity of TrwC relaxase in the presence of nickel, two curious things where noticed. First of all, $TrwC_R$ T87I showed a reduced activity with this metal in comparison with the activity observed with $TrwC_R$ WT. In addition, a drastic decrease in activity was observed when nickel concentration was increased from 1 mM to 10 mM. Nickel has been defined as a toxic element for living beings. It has been proposed in microorganism to negatively act by replacing the correct cofactor used by some metalloenzymes, by binding to catalytic residues in proteins that do not rely on a metal ion, by allosterically inhibiting the catalytic site of some enzymes or even by

affecting proteins and DNA due to oxidative stress (Macomber *et al.*, 2011). For example, nickel showed to have negative effects on magnesium-dependent DNA polymerases such us *E. coli* DNA polymerase I or the polymerases of the bacteriophages T4 and T7, with an activity reduction of about 50 % (Snow *et al.*, 1993) .Some of this reasons could explain our unexpected results with this metal.

These results for *in-vitro nic*-containing oligonucleotides cleavage do not agree with the ones obtained in (Boer *et al*., 2006) where big differences were found in $TrwC_R$ activity at 10 µM Mn^{2+} / Mg^{2+} containing buffers, and the drastic activity decrease in the presence of Ni^{2+} at high concentrations was not observed. But this could be due to the technique used; the cleaved oligonucleotides percentage was measured instead of the Protein-DNA covalent complex formation that we have analyzed.

Despite the highly variability of the experiment it is curious to find that both WT and mutant TrwC proteins are able to trigger conjugation at the same efficiency after the in vitro observations. However, this findings agree with the results obtained for T87A mutant (Guasch *et al*., 2003). It seems that the overall metallic content of the medium used for the assay was enough for TrwC to carry out the DNA transfer function independently of the amino acid present at position 87. Probably, due to the higher K_d for Mn^{2+}, TrwC uses Mn^{2+} up to the moment the medium runs out of this metal, and then starts to be active thanks to the Mg^{2+} that is found in higher amounts.

CONCLUSIONS AND FUTURE RESEARCH

- It has been determined the importance of threonine 87 of R388 TrwC relaxase in the histidine pocket formation. We have proved that this charged amino acid not directly interacting with the metal cofactor is involved in the metal coordination and specificity through hydrogen bond formation.

- In order to make general this statement, it should be proven if by mutating the equivalent hydrophobic amino acid for a polar residue on a HUH endonuclease only able to bind Mn^{2+}, the protein recovers the ability to use Mg^{2+} too.

- These findings open a new path for protein engineering to modify cofactor specificity and are also valuable for the new trend of *de novo* design of metalloproteins with new activities not found in nature for different biotechnological purposes.

REFERENCES

Adams, P., Afonine, P. , Bunkóczi, G., Chen, V., Davis, I., Echols, N., Headd, J., Hung, L., Kapral, G., Grosse-Kuntleve, R., McCoy, A., Moriarty, N., Oeffner, R., Read, R., Richardson, D., Richardson, J., Terwilliger T., Zwart, P. (2010). PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D: Biological Crystallography*, *66*, 213–221.

Boer, R., Russi, S., Guasch, A., Lucas, M., Blanco, A. G., Pérez-Luque, R., Coll, M., and de la Cruz, 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.

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

Chakrabarti, P. (1990). Geometry of interaction of metal ions with histidine residues in protein structures. *Protein Engineering*, *4*, 57–63.

Chandler, M., de la Cruz, F., Dyda, F., Hickman, A., Moncalian, G., and Ton-Hoang, B. (2013). Breaking and joining single-stranded DNA: the HUH endonuclease superfamily. *Nature Reviews. Microbiology*, *11*, 525–538.

Crooks, G., Hon, G., Chandonia, J., and Brenner, S. (2004). WebLogo: a sequence logo generator. *Genome Research*, *14*, 1188–1190.

De la Cruz, F., and Davies, J. (2000). Horizontal gene transfer and the origin of species: lessons from bacteria. *Trends in Microbiology*, *8*, 128–133.

De la Cruz, F., and Grinsted, J. (1982). Genetic and molecular characterization of Tn21, a multiple resistance transposon from R100.1. *Journal of Bacteriology*, *151*, 222–228.

Edwards, J. , Betts, L., Frazier, M. , Pollet, R., Kwong, S., Walton, W., Ballentine, W., Huang, J., Habibi, S., del Campo, M., Meier, J., Dervan, P., Firth, N., and Redinbo, M. (2013). Molecular

basis of antibiotic multiresistance transfer in Staphylococcus aureus. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 2804–2809.

Eide, D., Clark, S., Nair, T. , Gehl, M., Gribskov, M., Guerinot, M., Lou, M., and Harper, J. (2005). Characterization of the yeast ionome: a genome-wide analysis of nutrient mineral and trace element homeostasis in Saccharomyces cerevisiae. *Genome Biology*, *6*, R77.

Emsley, P., Lohkamp, B., Scott, W., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallographica Section D: Biological Crystallography*, *66*, 486–501.

Foster, A., Osman, D., and Robinson, N. (2014). Metal preferences and metallation. *The Journal of Biological Chemistry*, *289*, 28095–28103.

Garcillán-Barcia, M., Francia, M., and De La Cruz, F. The diversity of conjugative relaxases and its application in plasmid classification. *FEMS Microbiology Reviews*, *33*, 657–687.

Grant, S., Jessee, J., Bloom, F., and Hanahan, D. (1990). Differential plasmid rescue from transgenic mouse DNAs into Escherichia coli methylation-restriction mutants. *Proceedings of the National Academy of Sciences*, *87*, 4645–4649.

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

Guzman, L., and Espinosa, M. (1997). The mobilization protein, MobM, of the streptococcal plasmid pMV158 specifically cleaves supercoiled DNA at the plasmid oriT. *Journal of Molecular Biology*, *266*, 688-702.

Hernando, M. (2000) Bioquímica de las reacciones de transferencia de cadenas de DNA catalizadas por la proteína TrwC. Thesis/Dissertation. Universidad de Cantabria, Spain.

Larkin, M., Blackshields, G., Brown, N., Chenna, R., McGettigan, P., McWilliam, H., Valentin, F., Wallace, I., Wilm, A., Lopez, R., Thompson, J., Gibson, T., and Higgins, D. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, *23*, 2947–2948.

Leslie, A. (2006). The integration of macromolecular diffraction data. In *Acta Crystallographica Section D: Biological Crystallography*, 62, 48–57.

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

Llosa, M., Bolland, S., and de la Cruz, F. (1994). Genetic organization of the conjugal DNA processing region of the IncW plasmid R388. *Journal of Molecular Biology*, *235*, 448–464.

Lucas, M. (2006). Relación estructura-función de la relaxasa del plásmido R388 TrwC. Thesis/Dissertation. Universidad de Cantabria, Spain.

Macomber, L., and Hausinger, R. P. (2011). Mechanisms of nickel toxicity in microorganisms. *Metallomics*, *3*, 1153-1162.

Miroux, B., and Walker, J. (1996). Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *Journal of Molecular Biology*, *260*, 289–298.

Monzingo, A. , Ozburn, A., Xia, S., Meyer, R., and Robertus, J. (2007). The structure of the minimal relaxase domain of MobA at 2.1 Å resolution. *Journal of Molecular Biology*, *366*, 165–178.

Nuñez, B., Avila, P., and de la Cruz, F. (1997). Genes involved in conjugative DNA processing of plasmid R6K. *Molecular Microbiology, 24*, 1157-1168.

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 (tral) of IncPalpha plasmid RP4. *Journal of Biological Chemistry*, *22*, 13068-13076.

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. *Proceedings of the National Academy of Sciences of the United States of America*, *90*, 2925–2929.

Peacock, A. (2013). Incorporating metals into de novo proteins. *Current Opinion in Chemical Biology*, *17*, 934–939.

Petrik, I., Liu, J., and Lu, Y. (2014). Metalloenzyme design and engineering through strategic modifications of native protein scaffolds. *Current Opinion in Chemical Biology*, *19*, 67–75.

Rosenberg, A., Lade, B., Chui, D., Lin, S., Dunn, J., and Studier, F. (1987). Vectors for selective expression of cloned DNAs by T7 RNA polymerase. *Gene*, *56*, 125–135.

Snow, E., Xu, L., and Kinney, P. (1993). Effects of nickel ions on polymerase activity and fidelity during DNA replication in vitro. *Chemico-Biological Interactions*, *88*, 155–173.

Waldron, K., Rutherford, J., Ford, D., and Robinson, N. (2009). Metalloproteins and metal sensing. *Nature*, *460*, 823–830.

Wieghardt, K. (1989). The Active Sites in Manganese-Containing Metalloproteins and Inorganic Model Complexes. *Angewandte Chemie International Edition in English*, *28*, 1153–1172.

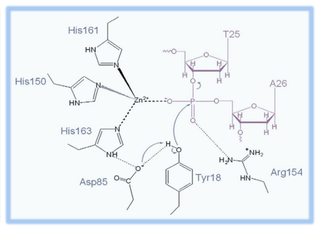
Winn, M., Ballard, C., Cowtan, K., Dodson, E., Emsley, P., Evans, P., Keegan, R., Krissinel, E., Leslie, A., McCoy, A., McNicholas, S., Murshudov, G., Pannu,N., Potterton, E., Powell, H., Read, R., Vagin, A. and Wilson, K. (2011). Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D: Biological Crystallography*, *67*, 235-242.

Identification of residues indirectly involved in cation coordination by HUH endonucleases

Lorena González-Montes and Gabriel Moncalián
Molecular Biology Department and
Institute of Biomedicine and Biotechnology of Cantabria, University of Cantabria, Spain

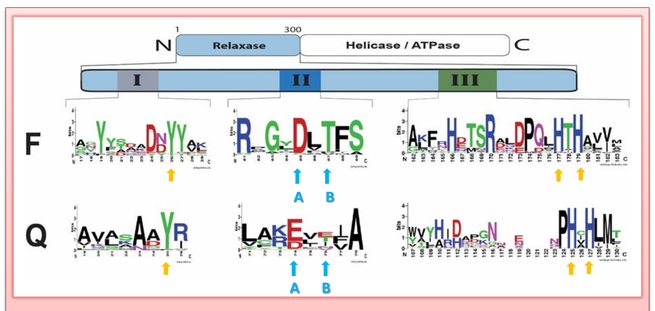
Background

HUH endonucleases rely on a divalent metal ion to perform different site-specific DNA processing reactions in biological processes like plasmid replication, transposition or bacterial conjugation. They all have an HUH motif (U=hydrophobic residue) and a Y motif (one or two catalytic Tyr). A catalytic Tyr creates a covalent 5'-phosphotyrosine intermediate and a free 3'OH at the cleavage site. This 3'-OH primes replication or acts as nucleophile for strand transfer. The metal is coordinated by the two HUH His and a third polar residue (Glu, Asp, His or Gln) being Mg^{+2} and Mn^{+2} the physiological cofactors. TrwC is one of the most studied HUH endonucleases at a biochemical and structural level. Its function is to transfer a single-stranded DNA (ssDNA) plasmid copy from one cell to another at conjugation process by nicking at the *nic* site of the plasmid origin of transfer *(oriT),* guiding the copy to the recipient cell and catalyzing there the recircularization of the transferred ssDNA plasmid.

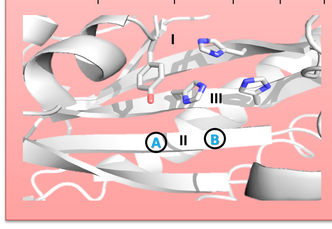


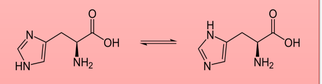
Hypothesis

All HUH endonucleases superfamily members have conserved HUH and Y catalytic domains (orange arrows). However, they have different metal specificities. We propose that metal affinity depends on the different character of the residues surrounding the amino acids directly involved in metal coordination at the active site, which are conserved within each relaxase family (blue arrows). Polar residues are found in the relaxases binding Mg^{+2} or Mn^{+2} such as TrwC and MobA. Hydrophobic residues are present in relaxases only binding Mn^{+2} such as NES. These secondary residues could modify cation specificity by affecting histidine tautomerization. In this work we have changed the protein metal specificity by mutating one of these polar amino acids (T87I in TrwC) by a non polar residue.



Relaxase	Plasmid	Family	(A)	(B)	Metals used	PDB	Reference
MobA	R1162	Q	E74	E76	Mg+2/Mn+2	2NS6	(Monzingo *et al.*, 2007)
NES	pLW1043	Q	E86	I88	Mn+2	4HT4	(Edwards *et al.*, 2013)
TrwC	R388	F	D87	T87	Mg+2/Mn+2	1OMH	(Boer *et al.*, 2006)

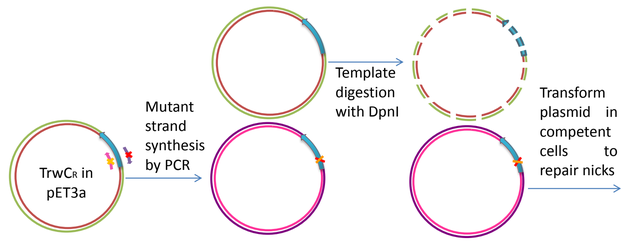




Histidine tautomerization

Methodology

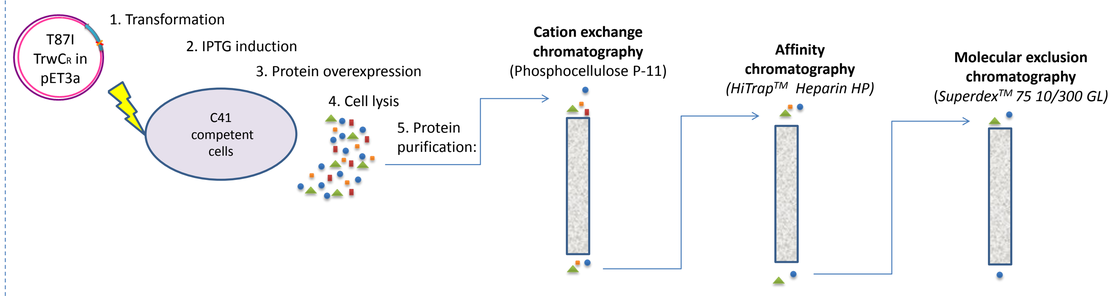
Site-directed mutagenesis



Thr codon changed to Ile at position 87 of $TrwC_R$ coding sequence

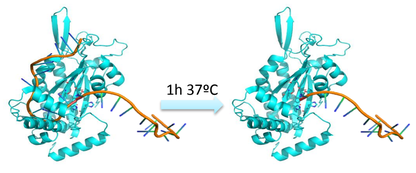
3'-ctcgcgtagccggaactggagtAtaagtcgcgcgggttctcgcat-5'
…cagcaaagagcgcatcggccttgacctcacattcagcgcgcccaagagcgtatcgcttc…
…gtcgtttctcgcgtagccggaactggagtgtaagtcgcgcgggttctcgcatagcgaag…
5'-gagcgcatcggccttgacctcaTattcagcgcgcccaagagcgta-3'

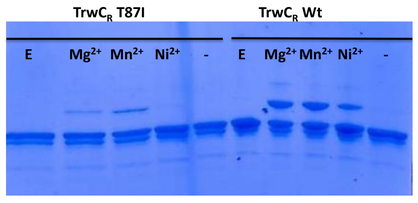
Protein overexpression and purification



Results and Discussion

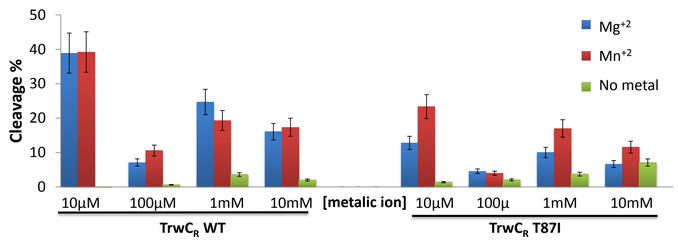
Mg^{+2} does not allow TrwC T87I cleavage of *nic*-containing oligonucleotides





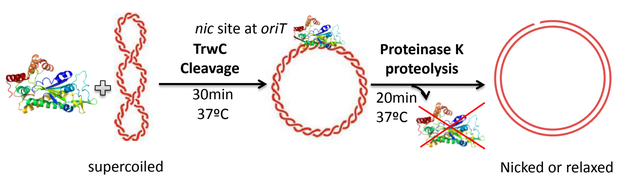
Different 12+18 oligonucleotide cleavage yield by $TrwC_R$ WT *vs.* $TrwC_R$ T87I with different metal cofactors at 10µM, EDTA (E) or without adding any metal (-).

Reaction
12+18 oligonucleotide (TGCGTATTGTCT/ATAGCCCAGATTTAAGGA) 15 µM + $TrwC_R$ WT or $TrwC_R$ T87I at 6 µM in Tris-HCl pH 7.5 10mM and EDTA (5mM) or metals at 0 (-), 10µM, 100µM, 1mM or 10mM. Complex formation was observed by differential migration at SDS-PAGE 12% gel ($TrwC_R$-DNA complex : upper band).

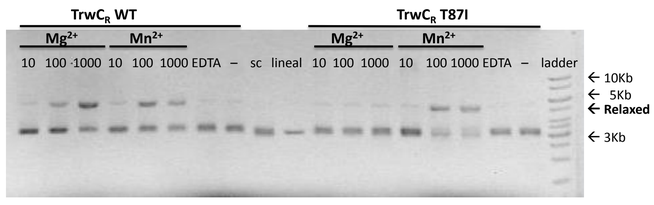


Effect of different cations in the oligonucleotide cleavage by $TrwC_R$ WT and $TrwC_R$ T87I at different cofactor concentrations. Mean and standard deviation of three independent experiments.

Mg^{+2} does not allow $TrwC_R$ T87I relaxation of *nic*-containing sc plasmids



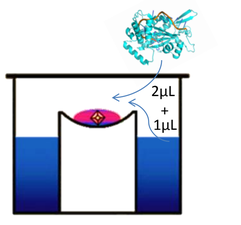
Reaction
$TrwC_R$ 500nM and pSU1186 supercoiled plasmid 10nM in Tris-HCl pH 7.5 10 mM and metals at 0 (-), 10 µM, 100 µM, 1000 µM or EDTA 5 mM.



Effect of different metal ions in the cleavage of a supercoiled (sc) plasmid containing the *oriT* (lower band) by $TrwC_R$ WT or $TrwC_R$ T87I mutant at different cofactor concentrations (expressed in µM).

While $TrwC_R$ WT is active both with Mg^{+2} and with Mn^{+2} as cofactors, TrwC T87I works *in vitro* significantly better in the presence of Mn^{+2}.

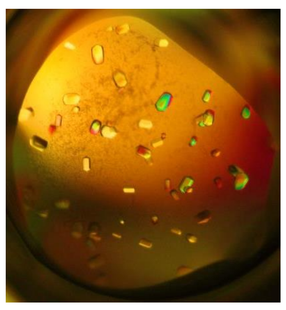
$TrwC_R$ T87I structure shows a slight change in the H163 orientation



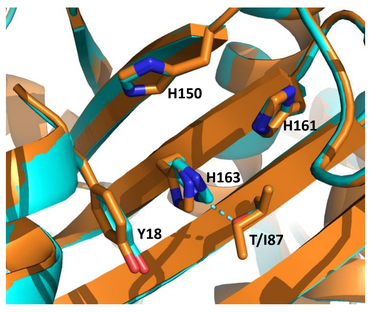
$TrwC_R$ WT or $TrwC_R$ T87I at a 1 : 1.5 protein : 23+0 oligonucleotide molar ratio were incubated with precipitant solution until rising equilibrium at 22ºC.

Crystals were soaked in 100mM Mn^{+2} solution and frozen for X ray diffraction and data collection at synchrotron.

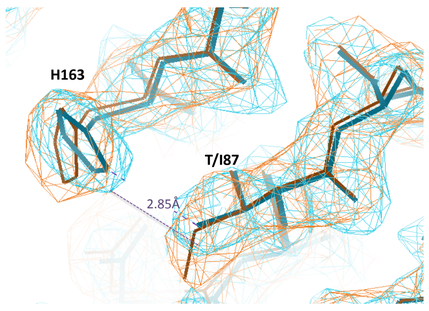
It is proposed that T87 by hydrogen bond formation plays an important role in the orientation of one of the histidines (His163) that coordinate the divalent cation.



$TrwC_R$ T87I – 23+0 crystals. Precipitant: Phosphate Ammonium 2M, TrisHCl pH=8.5 0.05M and Mn^{+2} 5mM.
Space group $P6_5$; a=b=149Å; c=78Å.

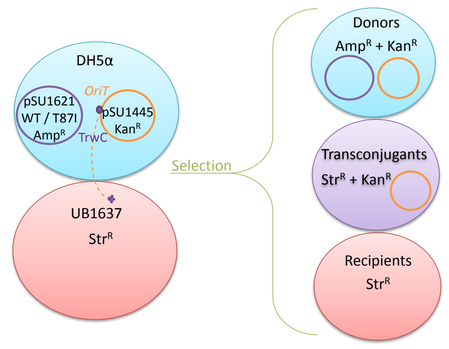


Superposed models of $TrwC_R$ active site showing the mutation and its consequence on His163 orientation. $TrwC_R$ WT is shown in blue and $TrwC_R$ T87I in orange. Hydrogen bond shown in dashed line.

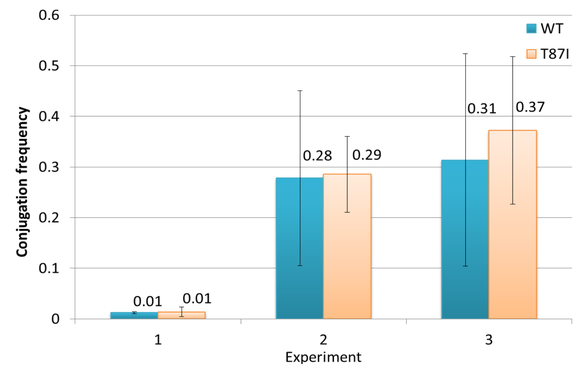


Electronic density map of residues H163 and T/I87 of $TrwC_R$ WT (blue) and T87I mutant (orange). Resolution: 2.2 Å and 1.6 Å, respectively. Hydrogen bond distance between His163 and Thr87 in WT protein is shown in purple.

T87I full length TrwC is able to efficiently transfer DNA *in vivo.*



Donor and receptor cells at stationary state were mixed at 1:1 proportion and plated on a LB plate for 1hour at 37ºC.



The experiment was done by triplicate three independent days. Conjugation frequency=transconjugants/donor colony. Cells containing pSU1445 complemented with pET3a showed a conjugation frequency lower than 10^{-6}.

$TrwC_R$ threonine 87 is not essential *in vivo*. TrwC T87I is able to drive plasmid DNA transfer at the same efficiency than TrwC WT. Thus, TrwC T87I is active using the Mn^{+2} at the low concentration present in the culture media, and TrwC could have evolved to be active using any of the available metals (mainly Mg^{+2} or Mn^{+2}).

Conclusions and Future Research

We have proved that a charged amino acid not directly interacting with the metal cofactor is involved in the orientation of the catalytic histidines in TrwC. This residue is therefore indirectly involved in the metal coordination and specificity. Thus, we think that by mutating the equivalent hydrophobic amino acid for a polar residue on an HUH endonuclease only able to bind Mn^{+2}, the protein could recover the ability to use Mg^{+2} too. But this hypothesis has still to be proven.
These findings open a new path to protein engineering for cofactor specificity modification and *de novo* protein design for different purposes.

References

- Boer, R., Russi, S., Guasch, A., Lucas, M., Blanco, A. G., Pérez-Luque, R., Coll, M., and de la Cruz, F. (2006). *Journal of Molecular Biology*, *358*, 857–869.
- Carballeira, J. D., González-Pérez, B., Moncalián, G., and de la Cruz, F. (2014). *Nucleic Acids Research*, *42*, 10632–10643.
- Chandler, M., de la Cruz, F., Dyda, F., Hickman, A. B., Moncalián, G., and Ton-Hoang, B. (2013). *Nature Reviews. Microbiology*, *11*, 525–538.
- Guasch, A., Lucas, M., Moncalián, G., Cabezas, M., Pérez-Luque, R., Gomis-Rüth, F. X., de la Cruz, F., and Coll, M. (2003). *Nature Structural Biology*, *10*, 1002–1010.

Acknowledgments

We thank Sandra Sagredo and Fernando de la Cruz for sharing their knowledge and advice. We also thank the European Commission for supporting EvoTAR project to FdC and the *Ministerio de Cultura y Deporte* for the collaboration grant given to LG-M.