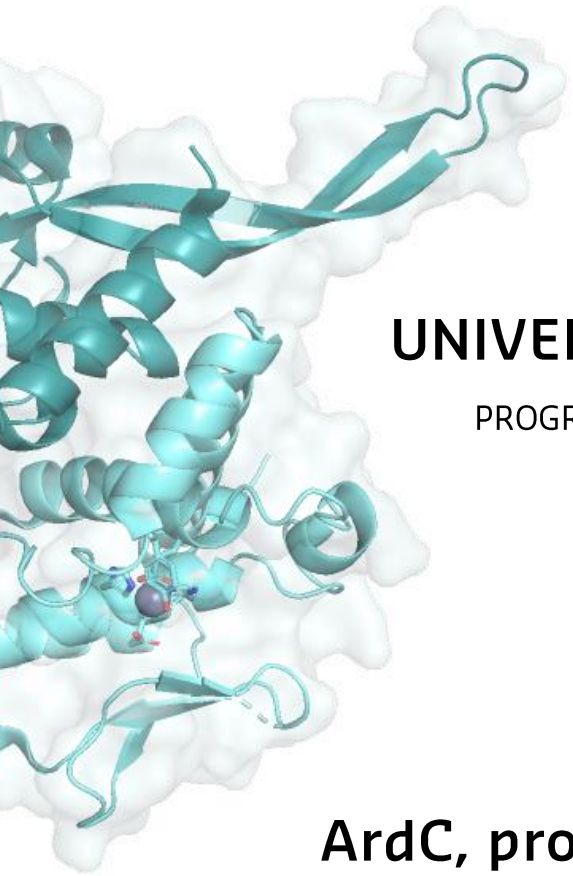


# UNIVERSIDAD DE CANTABRIA

PROGRAMA DE DOCTORADO EN BIOLOGÍA
MOLECULAR Y BIOMEDICINA

Tesis Doctoral

# ArdC, proteína antirestricción que amplía el rango de huésped del plásmido

PhD Thesis

# ArdC, antirestriction protein that broadens plasmid host range

Realizada por: Lorena González Montes

Dirigida por: Fernando de la Cruz y Gabriel Moncalián

Escuela de Doctorado de la Universidad de Cantabria

**Santander, 2019**

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 D. **Lorena González Montes** ha realizado bajo nuestra dirección el presente trabajo de Tesis Doctoral titulado: **La proteína ArdC de unión a ssDNA ayuda a superar los controles de inmigración en la célula receptora ampliando el rango de huésped en la conjugación.**

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 con opción a mención Internacional por la Universidad de Cantabria. Así mismo, Lorena González Montes se encuentra actualmente preparando dos artículos científicos que serán publicados próximamente.

Y para que conste y surta los efectos oportunos, expedimos el presente certificado en Santander, a de 2019.

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 Lorena González Montes durante su estancia en el laboratorio del profesor R.E.W. (Bob) Hancock en la Universidad de British Columbia (UBC Vancouver, Canadá).

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 Lorena González Montes during a short stay in the laboratory of professor R.E.W. (Bob) Hancock at the University of British Columbia (UBC Vancouver, Canada).

Esta investigación ha sido financiada por una beca FPU (Formación de Profesorado Universitario) del Ministerio de Ciencia e Innovación (MCINN): FPU014/06013.

El trabajo en los laboratorios Fernando de la Cruz y Gabriel Moncalián ha sido financiado por el Ministerio de Economía, Industria y Competitividad de España: BFU2014-55534-C2.

La estancia de investigación internacional de tres meses en la Universidad de Bristish Columbia (UBC) se realizó gracias a una ayuda de la Universidad de Cantabria.

Los experimentos estructurales se realizaron en la línea de luz XALOC en el Sincrotrón ALBA con la colaboración del personal de ALBA gracias a las propuestas anuales del Grupo de Asignación de Bloques con grupos de investigación de la Universidad de Salamanca.

Agradecemos al grupo de Victor de Lorenzo por cedernos *P. Putida* y sus mutantes.

This research was supported by an FPU grant (Formación de Profesorado Universitario) from the Spanish Ministry of Science and Innovation (MCINN): FPU014/06013.

The work in Fernando de la Cruz and Gabriel Moncalián laboratories was financed by the Spanish Ministry of Economy, Industry and Competitiveness: BFU2014-55534-C2.

The three-month international research stay at University of Bristish Columbia (UBC) was supported by a fellowship of the University of Cantabria.

Structural experiments were performed at XALOC beamline at ALBA Synchrotron with the collaboration of ALBA staff thanks to annual Block Allocation Group proposals with research groups of the University of Salamanca.

We thank Victor de Lorenzo's group for sharing *P. Putida* and its mutants.

En primer lugar, quiero dar las gracias a Gabriel Moncalián y a Fernando de la Cruz por todo lo que me habéis enseñado mientras me dirigíais y tutorizabais esta Tesis Doctoral. Así mismo al comité de sabios Ruli, Mapi, María Lucas, Marta y Santi por toda su ayuda, sus conocimientos compartidos y su ejemplo. Gracias a mis evaluadoras Asun y Elena por sus críticas constructivas y buenísimas recomendaciones año tras año. Omar y Martín, gracias por vuestra amabilidad para revisar esta tesis.

A Mati, la mami del laboratorio, quiero darla las gracias por toda su experiencia, sus mil y un protocolos para todo, por sus consejos y ánimos, por cuidarnos tan bien y haberme hecho sentir como en casa desde el primer día en el IBBTEC. Gracias a Sandra por enseñarme a dar mis primeros pasos en el mundillo de la ingeniería de proteínas y que la tesis es compatible con muchas otras actividades que se me pudieran pasar por la cabeza. Quiero dar las gracias a Laura por aguantarme día tras día, por sacarme de tantos apuros y sobrevivir a los tantos sustos que le he dado, por nuestros ratos terapéuticos y los viajes tan amenos en coche entre Santander y Bilbao. Al resto de compis del 2.08; Raúl, Martín, Serena, Omar, Andrés, Carla, Jorge, Arantxa, Aurora y toda la gente de prácticas de verano o FPU que, en mayor o menor medida, en lo laboral o personal, me han aportado un pedacito de ellos para realizar esta tesis. Mil gracias al servicio de técnicos del IBBTEC por las tantísimas horas de trabajo que nos ahorráis teniendo todo a punto. De verdad, ole por vuestro trabajo y dedicación.

Por supuesto mil y un gracias a mis amigas del Máster Judit y Yelina, por las conversaciones ultra motivadoras de autobús, por aquel primer helado que tanto unió, por nuestros mega viajes y escapadas más humildes, las siestas y momentos de relax, los bailes que nos hemos echado y cualquier otra actividad que compensase un día regulero o pusiese la guinda a un día maravilloso. A Fer le quiero dar las gracias por las tantas risas que nos hemos echado con absolutamente cualquier tontería, por ser mi compañero de quejas cuando algo no iba bien y por sus incontables abrazos por los pasillos. Aurora, gracias por ese soplo de aire fresco que trajiste este último año al laboratorio y así aportarme la mejor de las compañías.

Pero no todo queda en el IBBTEC, también quiero mostrar mi profundo agradecimiento a Alexandra Elbakyan por hacer posible el libre acceso a la información y facilitarnos la investigación a tantísimas personas.

Volviendo a mis orígenes, tengo que dar las gracias a Ana Sierra, la profesora de instituto que hizo que redescubriera mi pasión por la biología y que me dio fuerzas para reencaminar mis estudios para poder estar hoy orgullosa del desvió tomado a última hora hacia la biotecnología. Y por supuesto ya en la Universidad de Oviedo a mi profesora Pilar de la Peña por haberme hecho descubrir el maravilloso mundo de la biología molecular. Gracias a Iñaqui, Nacho Varela y María Lucas por enseñarme la profesión de profesor de Universidad durante las prácticas impartidas en la Facultad de Medicina en las que, año tras año y grupo tras grupo, nunca perdimos la motivación por enseñar. También quiero agradecerles a Rosa Poo, Cruz Celis, Nieves Pérez, Javier Menéndez, Bea Salas, Félix Sangari, Gabriel Moncalián y Laura Giner todos los momentos y oportunidades que me han dado para descubrir cuánto me apasiona el mundo de la educación y la divulgación científica, y que sin duda me han aportado grandes granitos de motivación, inspiración e ilusión durante esta tesis.

No me puedo olvidar de dar las gracias a Javier León y su equipo por hacer posible mi estancia en Canadá. A Bob Hancock y Susan Former por recibirme en su laboratorio y gran familia en la UBC, haciendo que pasase allí uno de los momentos más emocionantes de esta tesis. A Daniel por toda su motivación y ayuda con mi proyecto. Por todos sus “buena suerte” para cada

experimento. Y a Manjeet, por hacerme hueco en el mejor laboratorio y enseñarme a manejarme por el laboratorio.

Thanks to Bob and Susan, for the wonderful opportunity of joining your big lab family at UBC probably being the most exciting months of this thesis. Thanks to all the people in the lab, specially thanks to Daniel for been such a good mentor and all the “good lucks” for every experiment and to Manjeet for letting me space to be in the best lab I could be and teaching me how to manage in the lab.

A mis amigos de toda la vida, y a los de no hace tanto, los de esos preciosos años de carrera en Asturias, los que he conocido en mis viajes de aventuras, a Félix, mi cubano favorito, y mi gran amiga Mar por todo lo vivido juntos en Canadá y en especial a mis queridos Ru7eros por hacer de mí una mejor persona para poder afrontar estos años con coherencia e ilusión. Gracias a todos los que habéis venido a verme a Cantabria, por cargarme tanto las pilas con esos reencuentros.

Casi, por último, pero no menos importante, quiero dar las gracias al trabajo de los médicos y expertos que he conocido estos años tras los que puedo decir que han valido más que la pena para conocerme y poder hacer frente a mi futuro, ahora sí, con plena salud y fortaleza.

Y ya para acabar, a quienes dedico esta tesis, y las personas más importantes en mi vida: Gracias al apoyo incondicional y paciencia de mis padres, por creer y sentirse tan orgullosos de mí. Doy gracias por el día en que decidí pasar 5 años más de mi vida cerca de ellos. ¡Vosotros sí que sois mis grandes maestros! Gracias Christian, mil y un gracias por acompañarme desde el primer día de esta tesis, escuchándome, apoyándome, aprendiendo y enseñandome, “inspirándome a inspirar al mundo” … ¡Gracias por tu amor incondicional! Mi mayor motivación, quién siempre ha querido lo mejor para mí y se ha preocupado de que disfrutara y aprovechara cada día de esta tesis. Mi mejor ingrediente para el desarrollo de este trabajo y para mi propio desarrollo personal.

*"Estoy entre esos que piensan que la ciencia tiene una gran belleza. Un científico en su laboratorio no es sólo un técnico, sino que también es un niño colocado ante fenómenos naturales que lo impresionan como un cuento de hadas." … "Si veo algo vital a mi alrededor, es precisamente ese espíritu de aventura, que parece indestructible y es similar a la curiosidad."* - Marie Curie

*"I am among those who think that science has great beauty. A scientist in his laboratory is not only a technician: he is also a child placed before natural phenomena which impress him like a fairy tale." ... "If I see anything vital around me, it is precisely that spirit of adventure, which seems indestructible and is akin to curiosity."* - Marie Curie

A mis padres
y a Christian

# Contents

Contents

**1 Introduction ..... 1**

1.1 SSDNA BINDING PROTEINS ..... 3
1.2 HORIZONTAL GENE TRANSFER ..... 3
*1.2.1 Clinical relevance ..... 4*
*1.2.2 Environmental relevance ..... 5*
*1.2.3 Relevance in evolution ..... 6*
1.3 BACTERIAL CONJUGATION ..... 6
*1.3.1 Conjugation process ..... 7*
1.4 R388 PLASMID ..... 8
*1.4.1 R388 regulatory network and fitness cost ..... 9*
*1.4.2 R388 interspecific conjugation ..... 9*
1.5 BACTERIAL STRATEGIES TO CONTROL CONJUGATION ..... 10
*1.5.1 Plasmid barriers ..... 10*
1.5.1.1 Entry exclusion ..... 10
1.5.1.2 Fertility Inhibition ..... 11
*1.5.2 Host barriers ..... 11*
1.5.2.1 SOS response ..... 12
1.5.2.2 CRISPR-Cas systems ..... 13
1.5.2.3 Restriction and Modification (R-M) systems ..... 14
1.6 ANTIRESTRICTION STRATEGIES ..... 16
*1.6.1 Changes in DNA sequence ..... 17*
*1.6.2 Transient restriction sites occlusion ..... 18*
*1.6.3 Sabotage of host R-M activities ..... 18*
*1.6.4 Inhibition of R-M enzymes ..... 18*
1.7 ARDC ..... 19

**2 Aims and scope ..... 21**

**3 Experimental procedure ..... 25**

3.1 MATERIALS ..... 27
*3.1.1. Strains ..... 27*
*3.1.2. Plasmids ..... 28*
*3.1.3. Oligonucleotides ..... 29*
*3.1.4. Culture medium ..... 31*
*3.1.5. Enzymes, antibiotics and other chemical reagents ..... 31*
3.2 METHODS ..... 31
*3.2.1 Microbiological methods ..... 31*
3.2.1.1 Bacterial cultures ..... 31
3.2.1.2 Competent cells and Transformation by electroporation ..... 32
3.2.1.3 Bacterial conjugation experiments ..... 33
3.2.1.4 Growing Curves ..... 33
3.2.1.5 Plasmid Stability ..... 34
3.2.1.6 UV survival assay ..... 34
3.2.1.7 Minimum inhibitory concentration (MIC) ..... 34
3.2.1.8 RNA sequencing for transcriptome analysis ..... 35
*3.2.2 Molecular Biology methods ..... 36*
3.2.2.1 DNA purification and extraction ..... 36
3.2.2.2 DNA electrophoresis in agarose gels ..... 36
3.2.2.3 Plasmid construction by restriction enzymes ..... 37
3.2.2.4 Plasmid construction by Isothermal assembly ..... 37

3.2.2.5 Site-directed in vitro mutagenesis method ..... 38
3.2.2.6 Site-directed MAGE in vivo mutagenesis method ..... 40
3.2.2.7 Deletion by Wanner and Datsenko ..... 41
*3.2.3 Biochemical methods ..... 42*
3.2.3.1 Protein expression and purification ..... 42
3.2.3.2 Protein separation by SDS PAGE ..... 43
3.2.3.3 Determination of protein concentration ..... 43
3.2.3.4 DNA-binding Assays ..... 44
3.2.3.5 DNA binding and protection assays ..... 44
3.2.3.6 Crosslinking with glutaraldehyde ..... 44
3.2.3.7 Proteolytic activity assay ..... 45
3.2.3.8 Thermal Stability Assay based on fluorescence ..... 45
3.2.3.9 Crystallization ..... 46
3.2.3.10 Pull-Down assay ..... 47
3.2.3.11 Mass Spectrometry analysis ..... 47
*3.2.4 Bioinformatic methods ..... 48*
3.2.4.1 Databases ..... 48
3.2.4.2 Study of DNA sequences ..... 48
3.2.4.3 Study of proteins ..... 49

**4 Results ..... 51**

4.1 *ARDC* GENETIC CHARACTERIZATION ..... 53
*4.1.1. ardC shows two possible starting codons ..... 53*
*4.1.2. Analysis of the two putative protein isoforms ..... 55*
*4.1.3. ArdC diverse genetic environment and widely distribution ..... 57*
4.2 ARDC ROLE IN DNA TRANSFER BY CONJUGATION ..... 60
*4.2.1 ArdC effect in Conjugation ..... 60*
4.2.1.1 ArdC is needed to conjugate R388 from E. coli to different species but not the other way around 60
4.2.1.2 ArdC acts in recipient cells ..... 62
4.2.1.3 ArdC_l shows more activity than the shorter isoform ..... 62
*4.2.2 ArdC contributes to plasmid instability ..... 62*
*4.2.3 Modifications in the gene expression profile ..... 63*
4.2.3.1 Differential expression of R388 genes ..... 68
4.2.3.2 Differential expression of donor E. coli genes when the ardC-containing plasmid is transferred 72
4.2.3.3 Differential expression of P. putida genes ..... 74
4.2.3.4 Differential expression of EM422 ..... 77
4.3 ARDC STRUCTURAL CHARACTERIZATION ..... 77
*4.3.1 Overall crystal structure ..... 77*
*4.3.2 Nucleotide excision repair protein as closest ArdC N-terminal domain ..... 84*
*4.3.3 ArdC N-terminal domain ..... 85*
*4.3.4 ArdC C-terminal domain ..... 87*
4.3.4.1 ArdC thermal stability ..... 87
4.3.4.2 Metalloprotease active center ..... 88
*4.3.5 Transcriptional DNA damage repair activator and human genome instability protector protein as ArdC C-terminal closest structural homologues ..... 91*
*4.3.6 Identification of a putative ArdC target by homology ..... 93*
4.4 ARDC FUNCTIONAL CHARACTERIZATION ..... 94
*4.4.1 Analysis of ArdC oligomeric state ..... 94*
*4.4.2 ArdC preferentially binds ssDNA ..... 95*
*4.4.3 ArdC protects ssDNA from degradation by type II endonuclease ..... 95*
*4.4.4 ArdC do not protect DNA from UV damage ..... 96*
*4.4.5 Identification of putative ArdC partners by proteolytic activity ..... 98*
*4.4.6 Identification of putative ArdC partners by pull down assay ..... 100*
*4.4.7 Conjugation ..... 101*

4.4.7.1 ArdC is not needed if recipient hsdRMS is depleted ..... 101
4.4.7.2 ArdC protective activity is not due to metalloprotease activity ..... 104

**5 Discussion ..... 107**

5.1. ArdC improves interspecific bacterial conjugation by acting in recipient cells ..... 109
5.2 ArdC positively contributes to plasmid fitness ..... 109
5.3. Conjugation alters the transcriptome of both donors and recipient cells ..... 110
5.4. ArdC is a ssDNA binding protein ..... 111
5.5 ArdC has a metalloprotease domain that is not be needed for conjugation to *P. putida*. ..... 112
5.6 ArdC antirestriction activity prevents from degradation by *P. putida* Type I R-M system ..... 114

**6 Conclusions ..... 117**

**7 Spanish version ..... 121**

7.1 Introducción ..... 123
7.2 Objetivos ..... 126
7.3 Resultados y discusión de resultados ..... 127
*7.3.1 ArdC mejora la conjugación bacteriana entre especies actuando en las receptoras ..... 127*
*7.3.2 La conjugación altera el transcriptoma de las células donadoras y receptoras ..... 128*
*7.3.3 Caracterización estructural de ArdC ..... 131*
*7.3.4 Caracterización de la actividad de ArdC ..... 135*
7.4 Discusión general ..... 136
7.5 Conclusiones ..... 137

**8 Bibliography ..... 139**

**9 Supplementary matherial ..... 151**

Contents

# Abbreviations

| | |
|---|---|
| Å | Ångström (10-10 m) |
| aa | amino acid |
| AbR | antibiotic resistance |
| ACN | acetonitrile |
| Amp | ampicillin |
| APS | ammonium persulfate |
| Ard | alleviation of restriction of DNA |
| ATP | adenosine triphosphate |
| bp | base pair |
| BREX | bacteriophage exclusion |
| BSA | bovine serum albumin |
| Cas | CRISPR-associated protein |
| CD | circular dichroism |
| Cm | chloramphenicol |
| CRISPR | clustered regularly interspaced short palindromic repeats |
| crRNAs | small CRISPR RNAs |
| Da | Dalton (g mol-1) |
| dATP | deoxyadenosine triphosphate |
| dCTP | deoxycytidine triphosphate |
| dGTP | deoxyguanosine triphosphate |
| DNA | deoxyribonucleic acid |
| dNTP | deoxyribonucleotide triphosphate |
| dsDNA | double strand DNA |
| DTT | dithiothreitol |
| dTTP | deoxythymidine triphosphate |
| DUF | domain of unknown function |
| DYT | double yeast and tryptone |
| EDTA | ethylene diamine tetra acetic acid |
| EEX | entry exclusion |
| EMSA | electrophoretic gel mobility shift assay |
| Enase | endonuclease |
| FA | formic acid |
| g | gram |
| GTP | guanosine triphosphate |
| h | hour |
| HGT | horizontal gene transfer |
| Hsd | host specificity for DNA |
| ICE | integrative conjugative element |
| Inc | incompatibility group |
| IPTG | isopropyl-β-thiogalactopyranoside |
| Kn | kanamycin |

| | |
|---|---|
| LB | Luria-broth medium |
| LC-MS/MS | Liquid chromatography-tandem mass spectrometry |
| LDR | long direct repeat |
| M | mol l-1 |
| m4C | N4-methylcytosine |
| m5C | N5- methylcytosine |
| m6A | N6-methyladenine |
| MA(E) | gluzinzin metalloprotease subclan |
| MAD | multiwavelength anomalous diffraction |
| MAGE | multiplex automated genome engineering |
| MGE | mobile genetic element |
| MHB | Mueller-Hinton broth |
| MIC | minimum inhibitory concentration |
| min | minutes |
| MOB | mobility |
| MPD | 2-methyl-2,4-pentanediol |
| MPD | metalloprotease sub-domain |
| MPF | mating pair formation |
| MPTase | metalloprotease |
| MR | molecular replacement |
| Mtase | methyltransferase |
| MW | molecular weight |
| NER | nucleotide excision repair |
| NMR | nuclear magnetic resonance |
| nt | nucleotide |
| Nx | nalidixic |
| o/n | over night |
| Ocr | overcome classical restriction |
| OD | optical density |
| *oriT* | origin of transfer |
| *oriV* | origin of vegetative replication |
| PAGE | polyacrylamide gel electrophoresis |
| PAM | protospacer adjacent motif |
| PCR | polymerase chain reaction |
| PDB | protein data bank |
| PEG | polyethylene glycol |
| RBA | Ribonuclease A |
| Rcf | relative centrifugal force |
| RCR | rolling circle replication |
| Rep | RCR-initiator proteins |
| Rif | rifampicin |
| RINe | RNA integrity number equivalent |
| R-M | restriction and modification |

| | |
|---|---|
| RMSD | root mean square deviation |
| RNA | Ribonucleic acid |
| RPA | replication protein A |
| RPKM | reads per kilobase and million mapped reads |
| rpm | revolutions per minute |
| RT | room temperature |
| SAD | single anomalous dispersion |
| SAM | S-adenosyl methionine |
| SD | Shine-Dalgarno |
| SD | standard deviation |
| SDS | sodium dodecyl sulphate |
| Sec | seconds |
| Seq | sequencing |
| SFX | surface exclusion |
| ssDNA | single strand DNA |
| T4CP | type IV coupling protein |
| T4SS | type IV secretion system |
| TBE | Tris Borate EDTA |
| Tc | tetracycline |
| TEMED | N,N,N',N'-Tetramethylethylenediamine |
| TM | melting temperature |
| Tmp | trimethoprim |
| UV | ultra violet |
| v/v | volume/volume |
| w/v | weight /volume |
| WT | wild type |
| ZBD | Zn2+- binding sub-domain |
| ε | extinction coefficient value |

# List of Figures and Tables

# List of Figures

Figure 1. Main mechanism of horizontal gene transfer (HGT) involved in the dissemination of resistance genes. ..... 3
Figure 2. Worldwide deaths attributable to $Ab^R$ every year compared to other major causes of death. ..... 4
Figure 3. Timeline of antibiotic deployment and $Ab^R$ appearance. ..... 4
Figure 4 Genetic mechanisms that confer antibiotic resistance to bacteria. ..... 5
Figure 5. Overview of the ecology of antibiotics and $Ab^R$. ..... 6
Figure 6. Schematic representation of the conjugative process. ..... 7
Figure 7. Genetic map of R388 plasmid. ..... 8
Figure 8. R388 plasmid transcriptional regulatory network. ..... 9
Figure 9. Conjugation frequencies of R388 and pIC10 from E. coli to other bacteria. ..... 10
Figure 10. Schematic representation of RecA-LexA SOS response activation. ..... 12
Figure 11. General overview of the Stages of CRISPR-Cas immunization and immunity. ..... 13
Figure 12. Schematic representation of restriction–modification (R-M) systems. ..... 14
Figure 13. Schematic representation of host R-M system and antirestriction strategies. ..... 17
Figure 14. Tertiary structure and surface charges distribution of ArdA dimer, mimicking DNA.. ..... 18
Figure 15. Schematic representation of the final antibiotic concentration in the 96 well MIC plate. ..... 35
Figure 16. Schematic representation of the Isothermal Assembly cloning method. ..... 38
Figure 17. Schematic representation of the Site-directed QuickChange method. ..... 39
Figure 18. Schematic representation of the steps followed for deletion of ardC from R388 by a modification of the Wanner and Datsenko method. ..... 41
Figure 19. Schematic representation of the Thermal stability assay based on fluorescence by Sypro-Orange dye ..... 45
Figure 20. Nucleotide and amino acid sequence for ardC gene. ..... 54
Figure 21. Alignment of the two DNA repeats flanking ardC gene. ..... 54
Figure 22. BLAST nucleotide alignment output for R388 ardC_l sequence. ..... 55
Figure 23. ArdC protein overexpression and purification. ..... 55
Figure 24. ArdC_l protein overexpression and purification. ..... 56
Figure 25. Molecular Weight determination by ESI-MS ..... 57
Figure 26. Effect of ardC and kfrA-orf14 region on plasmid conjugative transfer from E. coli to different bacteria. ..... 60
Figure 27. Effect in the conjugation frequency when complementing pLGM25 with ardC in donors or recipients. Complementation in donors (shown in teal) or in recipients (shown in maroon) with pUCP22, pUCP22::ardC_l or pUCP22::ardC is shown. ..... 62
Figure 28. Generation times of P. putida KT2440 empty or carrying pSU2007, pLGM25 or pIC10 in LB media. ..... 63
Figure 29. Stability of plasmids pSU2007, pLGM25 and pIC10 in P. putida KT2440. ..... 63
Figure 30. Conjugation experiments designed for RNA-Seq analysis. ..... 65
Figure 31. Total RNA integrity quantity and quality extracted from experiments #1, #2 and #3 validated by Agilent RNA ScreenTape assay. ..... 66
Figure 32. Differential expression of genes in experiment pSU2007 vs. pLGM25 to test influence of ardC in the three reference genomes. ..... 68
Figure 33. Genetic map of R388 plasmid.. ..... 70
Figure 34. Expression profile of R388 genes. ..... 71

Figure 35. Processes and functions with differential expression in E. coli during conjugation to P. putida with plasmid having or lacking ardC. ..... 72
Figure 36. Processes and functions with differential expression in E. coli during conjugation of pLGM25 to P.putida. ..... 74
Figure 37. ArdC protein crystals. ..... 77
Figure 38. Overall structure of ArdC. N-terminal domain surface is shown in pink and C-terminal domain surface is shown in blue. ..... 79
Figure 39. ArdC secondary structure representation. ..... 79
Figure 40. Schematic topological representation of ArdC protein. N-terminal domain is shadowed in pink and C-terminal domain in blue. ..... 80
Figure 41. Overall labelled crystal structure of ArdC protein. ..... 81
Figure 42. Structure-based sequence alignment of ArdC from different plasmids and microorganisms. ..... 82
Figure 43. ArdC amino acid conservation.. ..... 83
Figure 44. Electrostatic potential surface of ArdC protein. Negative surface is coloured in red, positive in blue (calculated by APBS tool). ..... 84
Figure 45. ArdC-Rad4 structural alignment. ..... 84
Figure 46. Protein sequence alignment of ArdC-N and Rad4. ..... 85
Figure 47. ArdC-ssDNA crystals. . ..... 85
Figure 48. Structural alignment of ArdC and ArdC bound to DNA. ..... 86
Figure 49. Surface electrostatic potential of two ArdC molecules crystallized with ssDNA The two molecules are oriented in symmetrical positions along DNA axis. ..... 87
Figure 50. Melting curve exemplifying the ThermoFuor assay. ..... 88
Figure 51. $T_M$ of ArdC in the presence of $NiCl_2$, $CoCl_2$ and $MnCl_2$ at different metal concentrations.. ..... 88
Figure 52. ArdC protein crystals for ArdC-metal structures. ..... 89
Figure 53. Metal binding site of ArdC. Apo form is shown in maroon and zinc-bound form in blue. ..... 89
Figure 54. ArdC-Mn asymmetric unit. ..... 89
Figure 55. Metal binding site of ArdC bound to $Mn^{2+}$ at 2.7 Å resolution.. ..... 90
Figure 56. Structural alignment of ArdC unbound (in maroon), bound to Zn (in teal) or Mn (orange). A) Global alignment. ..... 91
Figure 57. Structural alignment of ArdC and SprT.. ..... 92
Figure 58. Protein sequence alignment of ArdC-C, SprT-MPD and IrrE-N. ..... 92
Figure 59. Structural alignment of ArdC and IrrE. ..... 93
Figure 60. S75 gel chromatography of ArdC eluting at a MW between BSA and RBA. ..... 94
Figure 61. Oligomeric state of ArdC analysed by glutaraldehyde crosslinking. ..... 94
Figure 62. Oligomeric state of ArdC in complex with ssDNA analysed by glutaraldehyde crosslinking on a SDS-PAGE gel. ..... 94
Figure 63. ArdC DNA-binding analysed by EMSA. . ..... 95
Figure 64. Agarose gel showing the retardation and protection of long ssDNA (M13mp18) by ArdC. ..... 96
Figure 65. UV sensibility assay.. ..... 97
Figure 66. UV sensibility assay coupled to conjugation. ..... 98
Figure 67. Proteolytic assay of ArdC to HhaI in the presence of different DNA forms and metal cofactors. ..... 99
Figure 68. Proteolytic assay of ArdC and ArdK in buffer containing different metal cofactors in the presence or absence of ssDNA (45 bases oligonucleotide). ..... 99

Figure 69. Proteolytic assay of ArdC to RecA in buffer containing different metal cofactors. ... 99
Figure 70. Pull down assay using ArdC as bait and the whole cell lysate of P. putida as prey. 100
Figure 71. Analysis of pull-down fractions by size exclusion chromatography..... 101
Figure 72. A) Protein from pull down experiment cleaved from gel and sent for mass spectrometry analysis. ..... 101
Figure 73. Effect of ardC on plasmid conjugative transfer from E. coli to P. putida KT2440 mutants at 37 °C. ..... 102
Figure 74. Effect of ardC on plasmid conjugative transfer from E. coli to P. putida KT2440 mutants at 30 °C. ..... 103
Figure 75. Effect of temperature on plasmid conjugative transfer from E. coli BW27783-Nx to E. coli BW27783-Rif. ..... 104
Figure 76. Effect in the conjugation frequency when ArdC E229 residue is mutated to E229A in pSU2007 plasmid..... 104
Figure 77. Effect in the conjugation frequency when ArdC E229 residue is mutated to A at different temperratures. ..... 105

## List of Tables

Table 1. Bacterial strain used in this thesis. ..... 27
Table 2. Plasmids already published and used during this thesis. ..... 28
Table 3. Plasmids constructed during this thesis. ..... 29
Table 4. Oligonucleotides used for the construction of recombinant plasmids. ..... 29
Table 5. Oligonucleotides used for sequencing. ..... 30
Table 6. Oligonucleotides used for colony PCRs. ..... 30
Table 7. Oligonucleotides used for protein binding and cleavage assays. ..... 30
Table 8. Oligonucleotides used for crystallization. ..... 31
Table 9. PCR reaction mixture for site-directed mutagenesis with Vent® polymerase. ..... 39
Table 10. PCR program used for site-directed mutagenesis with Vent® polymerase. ..... 39
Table 11. PCR reaction mixture for site-directed mutagenesis with PfuUltra II Hotstart PCR Master Mix. ..... 40
Table 12. PCR program used for site-directed mutagenesis with PfuUltra II Hotstart PCR Master Mix. ..... 40
Table 13. ArdC BlastP taxonomic analysis. Table was obtained by searching in BlastP for 5000 homologue sequences from the non-redundant protein database (nr) and Blosum 62 matrix. 59
Table 14. Influence of temperature in conjugation frequency per recipients (T/R) from E. coli grown at 37 °C to E. coli grown at 30 °C in 1h of conjugation at 30 °C or 37 °C. ..... 61
Table 15. Influence of ArdC in the conjugation frequency per recipients (T/R) from P. putida to E. coli when conjugating 1h at 30 °C or 37°C. ..... 61
Table 16.Conjugation titration for RNA-seq assay. ..... 64
Table 17. Conjugation results for experiments #1, #2 and #3 further analyzed by RNA-seq. .... 65
Table 18. TruSeq® Stranded mRNA (Strand specific) Illumina sequencing results.. ..... 66
Table 19. Coverage calculation for each sequencing experiment. ..... 66
Table 20. Percentage of reads aligned to the three reference genomes by Bowtie2. ..... 67
Table 21. Expression level of R388 genes in RPKMs from experiment #2 (pSU2007) and #3 (pLGM25) and comparison of expression in RPKMs (pSU2007/pLGM25) for R388 plasmid genes. ..... 69
Table 22. Expression level of E. Coli genes in RPKMs from experiment #1 (BW), #2 (pSU2007) and #3 (pLGM25) and comparison of expression in RPKMs (Fold Change) for the most differentially expressed genes. ..... 72
Table 23. Expression level of E. Coli genes in RPKMs from experiment #1 (BW), #2 (pSU2007) and #3 (pLGM25) and comparison of expression in RPKMs (Fold Change) for the most differentially expressed genes. ..... 73
Table 24. Expression level of P. putida genes in RPKMs from experiment #1 (BW), #2 (pSU2007) and #3 (pLGM25) and comparison of expression in RPKMs (Fold Change) for every gene..... 75
Table 25. Expression level of P. putida genes in RPKMs from experiment #1 (BW), #2 (pSU2007) and #3 (pLGM25) and comparison of expression in RPKMs (Fold Change) for every gene..... 76
Table 26. Data collection and refinement statistics for ArdC structure. Statistics for the highest-resolution shell are shown in parentheses. ..... 78
Table 27. ArdC closest structural homologues obtained by Dali for each structural domain. ... 81
Table 28. Data collection and refinement statistics for ArdC-DNA structure. ..... 85
Table 29. $T_M$ value of ArdC in different solutions. ..... 87
Table 30. Data collection and refinement statistics for ArdC-Zn and ArdC-Mn structures.. ..... 91
Table 31. Effect of ardC on plasmid conjugative transfer between E. coli rK- mK+ BW27783 strain and E. coli rB- mB- C41 strain in both directions carrying pSU2007, pIC10 or pLGM25 in donors.. ..... 103

# 1 Introduction

## 1.1 SsDNA binding proteins

DNA binding proteins recognize and bind DNA to regulate different cellular functions related to the maintenance, replication, or transcription of the DNA. Among them, proteins that bind single strand DNA (ssDNA) are crucial for DNA function in cells of all three domains of life, as well as in mitochondria, phages and viruses (Bochkarev et al., 1997). SsDNA-binding proteins have a broad range of functions and structures to achieve each of those specific activities. In addition, ssDNA recognition and binding can be specific or unspecific of sequence depending on the biological process.

For instance, in eukaryotes, stable ssDNA is present at the ends of the chromosomes and at some promoter regions. And transiently, it is present in processes as telomere synthesis, transcription, and DNA replication and repair where, for example, replication protein A (RPA), Rad51 and Rad52 proteins are crucial proteins involved in the homologous recombination process (Dickey et al., 2013; Sugiyama and Kantake, 2009).

The recognition and processing of ssDNA needs to be carefully managed by the ssDNA binding proteins as an aberrant actuation due to a lacking or mutated ssDNA binding protein can cause DNA damage and cell death (Dickey et al., 2013). For example, mutations in BRCA2 tumour suppressor ssDNA binding protein can lead to breast cancer due to a deficient Rad51-mediated recombination that causes genome instability (Yang et al., 2002). Or producing mutated TREX1 protein, a major 3' --> 5' exonuclease in human cells, can drive to autoimmune diseases due to unnecessary accumulation of DNA (Bailey et al., 2012).

On the other hand, in Prokaryotes, RecA is a multifunctional enzyme that acts in DNA repair by homologous recombination acting as a DNA-dependent ATPase, and in the induction of the SOS response to repair the damaged DNA acting as a protease cofactor of LexA repressor (Selbitschka et al., 1991). As other examples, helicases and SSBs (single-strand DNA binding proteins) are needed in the replication fork for dsDNA unwinding and ssDNA stability and protection during DNA replication.

Other process where ssDNA needs special attention is in DNA transfer. DNA can be transferred vertically through cell division or horizontally, through horizontal gene transfer (HGT) between unrelated cells.

## 1.2 Horizontal gene transfer

Mobile genetic elements (MGEs), which are DNA molecules with the ability to move inside and between cells, facilitate HGT. Transposons move intracellularly between different parts of the genome by recombination. In addition, endosymbiotic gene transfer occurs in eukaryotes when DNA from an endosymbiont or organelle (as a chloroplast or mitochondrion) is introduced into the host genome. On the other hand, intercellular movements happen mainly by transduction, transformation or conjugation.

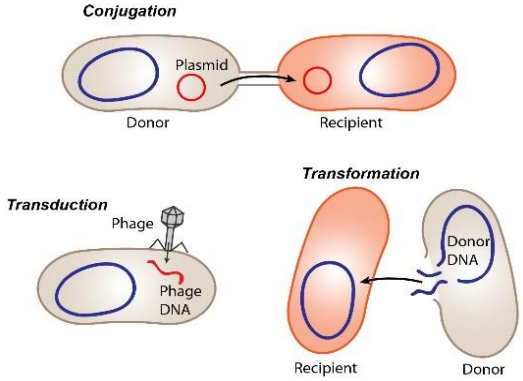



***Figure 1.*** *Main mechanism of horizontal gene transfer (HGT) involved in the dissemination of resistance genes. (Francesca Short, 2018).*

Transduction is the transfer of DNA from one cell to another through bacteriophage infection, it means, a virus that packs host DNA segments and injects them in a new host cell. Transformation implies the free DNA uptake by bacteria in the physiological estate of competence to incorporate DNA from outside released by dead cells. And conjugation requires the machinery to build a direct contact between a donor and a recipient cell (Soucy et al., 2015) (Figure 1).

### 1.2.1 Clinical relevance

Antibiotic resistance in bacteria represents one of the biggest threats that humanity is currently facing, killing more than 700,000 people every year. Moreover, the problem does nothing but increase, as it is predicted to be the greatest worldwide challenge in healthcare by 2050 as it is expected to cause even more deaths than cancer (O'Neill, 2014) (**Figure 2**).

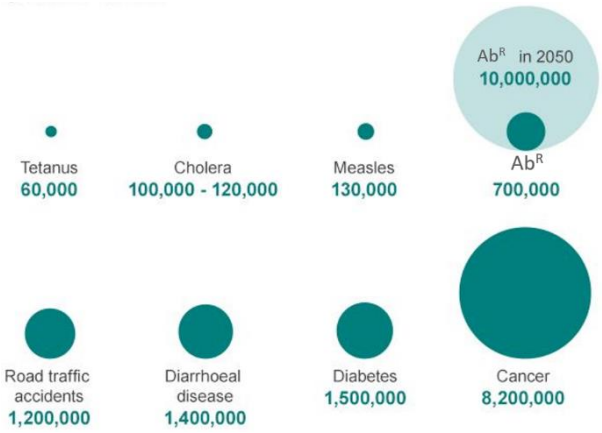



***Figure 2.*** *Worldwide deaths attributable to Ab$^R$ every year compared to other major causes of death. Current deaths are shown in dark teal and estimation for 2050 are shown in light teal. (O'Neill, 2014).*

Antibiotics (or antimicrobial agents) have been saving lives since Alexander Fleming discovered in 1928 the Penicillin (Fleming, 1929). However, soon after antibiotic usage starts, antibiotic resistance (Ab$^R$) for that antibiotic appears (Clatworthy et al., 2007) (Figure 3).

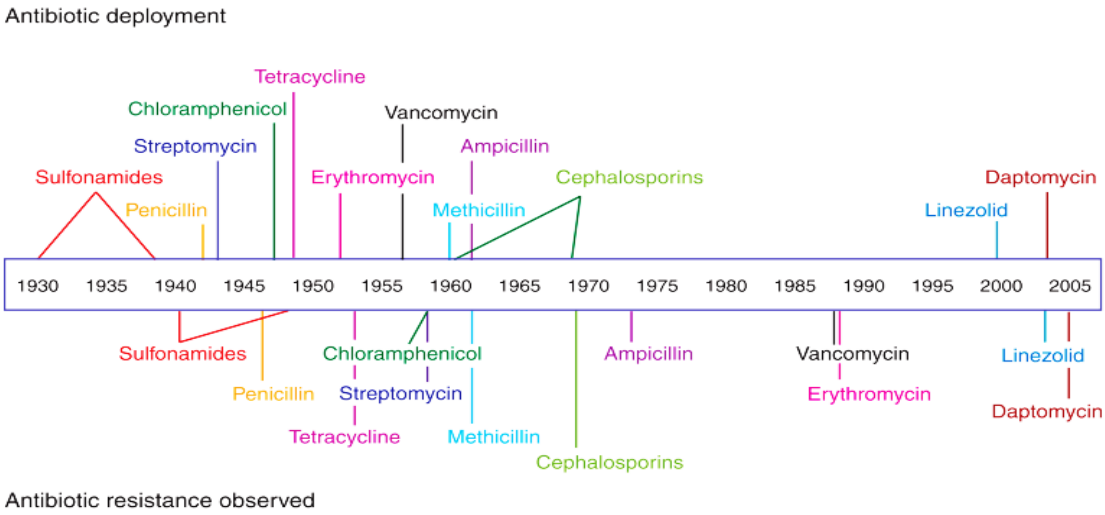



***Figure 3****. Timeline of antibiotic deployment and Ab$^R$ appearance. Above the timeline, it is shown the year at which each antibiotic was deployed and below the timeline the year at which resistance to each antibiotic treatment was first observed. (Clatworthy et al., 2007).*

There are several strategies to resist antimicrobial action (Figure 4) such as i) modifications in the antibiotic target, ii) alternative pathways to bypass the antibiotic blocked step, iii) decreased membrane permeability or iv) increased efflux to pump out the antibiotics and thus reduce the antibiotic concentration inside the cell, v) modification or inactivation of the antibiotic by enzymatic action, or vi) overproduction of antibiotic target (Coates et al., 2002).

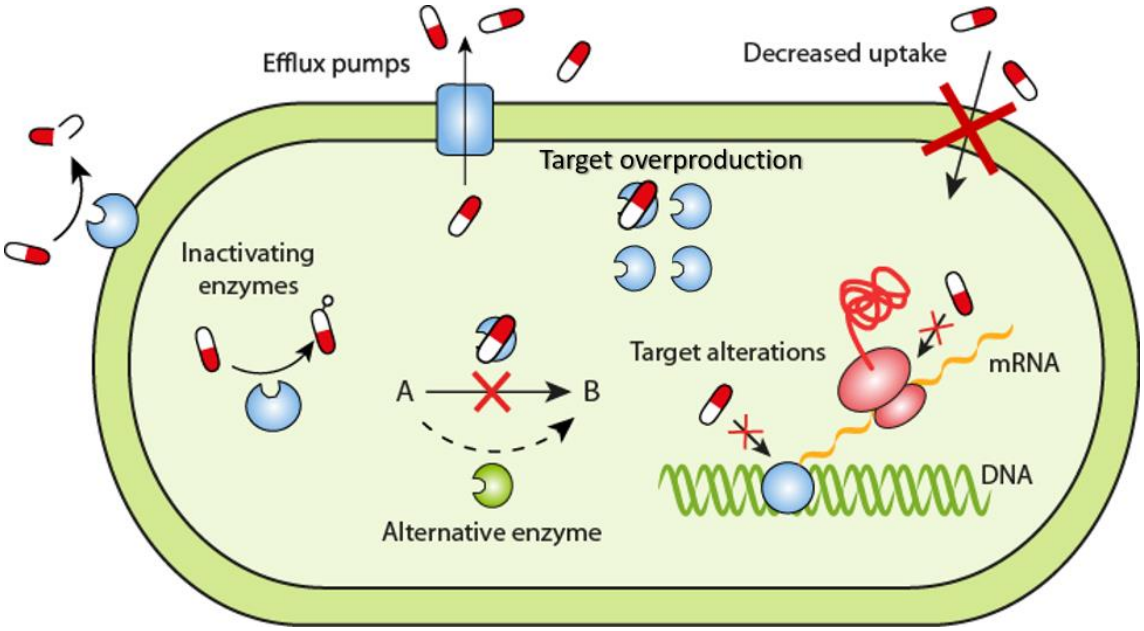



***Figure 4*** *Genetic mechanisms that confer antibiotic resistance to bacteria. Adapted from (Gullberg, 2014).*

The clinical relevance of the HGT process lies on the fast acquisition and dissemination of antibiotic resistance genes ($Ab^R$) involved in any or several of the already mentioned strategies (Figure 4) by horizontal gene transfer between unrelated pathogens. In addition, this phenomenon is coupled to the strong selective pressures set by the abuse and misuse of antibiotics in medicine and animal husbandry (Figure 5). When bacteria have to face with selective pressures as the one exerted by antibiotics, horizontal acquisition of $Ab^R$ allows the diversification of the genomes creating a fast adaptation and thus, increasing the survival opportunities (Andersson and Hughes, 2014; Davison, 1999; Lerminiaux and Cameron, 2019). This alarming situation asks for solutions, and special research attention is required to characterize the HGT mechanisms that contribute to the spread of $Ab^R$ between bacteria.

### 1.2.2 Environmental relevance

HGT-mediated evolution allows the adaptability of bacteria to face environmental challenges imposed by human activities as the already mentioned overuse of antibiotics in medicine, intensive agriculture with increasing use of bactericides to manage plant diseases, antibacterial agents applied in intensive fish farms or general industrial spill over contamination. The evolution and transfer of degradative plasmids is a response to this worrying increase in the presence of xenobiotic pollutants in soil as well as in water (Aminov, 2011; Andersson and Hughes, 2014; Davison, 1999) (Figure 5). There are different types of genes that confer adaptive advantages for survival under certain environmental conditions that could contain antibiotics, xenobiotics, heavy metals, and other compounds. Some of the most interesting cargoes exchanged in environment are transporter-encoding genes and catabolic genes as they allow the use by the organism of new metabolites present in the medium as growth and energy substrates. (Milner et al., 2019; Shintani and Nojiri, 2013).

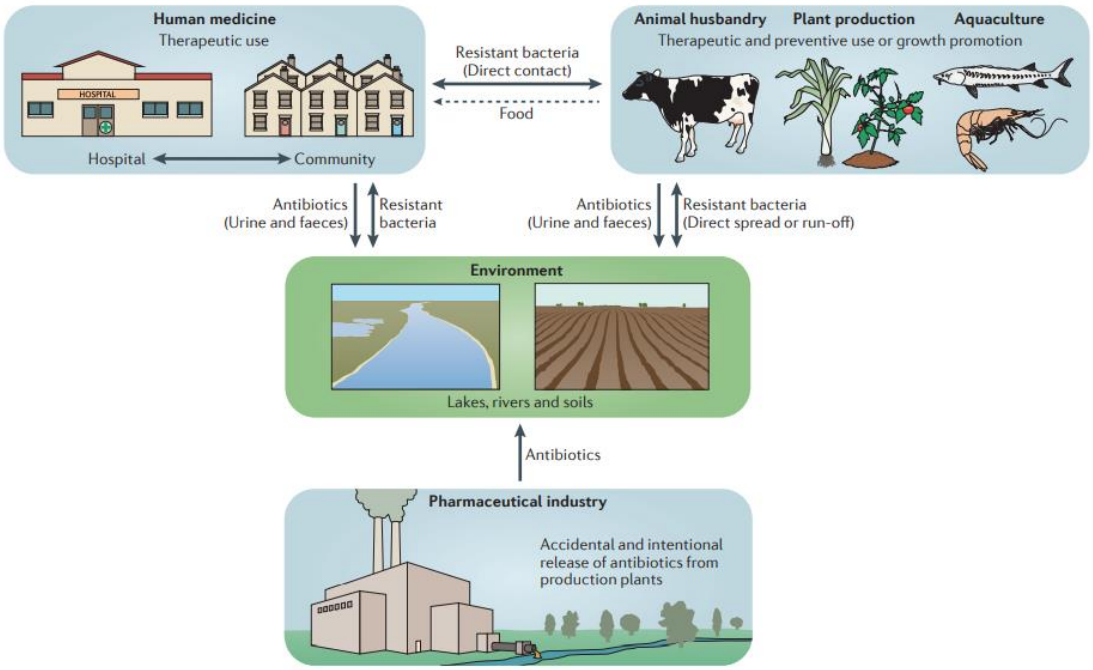



***Figure 5**. Overview of the ecology of antibiotics and $Ab^R$. The image also shows how antibiotics and $Ab^R$ are spread between different environments in a circular way. Special attention is on how antibiotics are released into the environment due to intentional or unintentional spill over exerting a selective pressure on bacteria that will lead to the selection of resistant strains able to transfer their traits by HGT to other bacteria. In addition, these resistant bacteria are capable of being transferred to a different environment where they can disseminate the $Ab^R$. (Andersson & Hughes, 2014).*

### 1.2.3 Relevance in evolution

HGT is a source of new traits by the acquisition of novelties that allow adaptation and evolution of the recipient cell. Fruitful HGT results from the successful transfer of DNA combined with the survival of the transferred genetic material over time. The maintenance of the transferred DNA is associated with positive selection. Thus, genes with useful functions are more prone to be preserved throughout generations than useless ones which are more prone to be removed (Gogarten et al., 2002).

Horizontal gene transfer is an important evolution driving force in Archaea and the microbial world and, with less relevance, in the Eukaryotic kingdom too. Despite being gene exchange easier between closely related organisms, different cases of HGT have been reported among and between the three domains of life (Boto, 2010). However, most of the genes acquired by HGT in eukaryotes come from bacteria, in part due to their bigger metabolic diversity supply (Keeling, 2009). It has been reported that the transfer of microbial genes to eukaryotes comes mainly from endosymbiotic events, having been transferred from mitochondrial and plastid ancestors to the nucleic DNA of their hosts (Keeling and Palmer, 2008).

## 1.3 Bacterial conjugation

Conjugation is the main process that allows the transfer of genes encoded in autonomous plasmids or in integrating conjugative elements (ICEs) integrated into a host genome. Plasmids are long DNA molecules usually between 1,000 and 100,000 bp containing a diverse range of adaptive traits, including genes conferring resistance to antibiotics ($Ab^R$) (Thomas, C. M., and

Summers, 2008). Plasmids, as well as bacteriophages, are between the most primitive life forms (Zavilgelsky, 2000). They are unable to live outside an organism, thus having a parasite live cycle.

Conjugative systems possess two groups of genes: mobility (MOB) genes required for conjugative DNA processing and mating pair formation (MPF) genes required for building the conjugative channel or Type IV secretion system (T4SS) between donor and recipient cells. MOB set of genes include the origin of transfer (*oriT*), which is a small DNA sequence required to start the DNA processing, a relaxase to catalyze the first and last stages of conjugation and a type IV coupling protein (T4CP) to interconnect the DNA processing to the transport channel. Plasmids can be conjugative, mobilizable, and non-mobilizable. They are named conjugative or self-transmissible if they are autonomous to replicate in a cell and able to be transfer between cells by conjugation. These plasmids are usually bigger in size as they have both MOB and MPF sets of genes. In addition, they are usually present in a low copy number. On the other hand, mobilizable plasmids are those which their transfer depends on the help of another conjugative plasmid, as they lack the MPF set of genes. They are usually smaller and are present in a bigger copy number. Plasmids unable to conjugate nor to be mobilized are called non-mobilizable (Smillie et al., 2010).

There are several incompatibility groups of plasmids. It means that bacteria can harbor several plasmids at the same time, but they should be compatible between them, or what is the same, they should belong to different incompatibility groups. Plasmids that share one or several elements of the replication or partitioning functions belong to the same incompatibility group (Garcillán-Barcia et al., 2009). Another aspect to introduce about plasmids is their host range. For instance, plasmids of narrow host range are found in incompatibility groups IncF, IncI and IncH. In contrast, IncP, IncN an IncW plasmids can be transferred to and replicated and maintained in a broad host range (Suzuki et al., 2010).

### 1.3.1 Conjugation process

For bacterial conjugation, the initial requirement is the expression of MPF genes in the donor cell. The T4SS complex is formed by four components: the pilus, an appendage that protrudes from the donor towards the proximal recipients and in some cases to latter on retract to produce a closer cell contact; the central channel complex, the inner membrane platform, and the ATPase, an energy generator required for pilus formation and DNA transport (Low et al., 2014).

Once cell contact has been stablished, MOB genes coordinate to start the DNA processing. Several ssDNA-binding proteins are involved

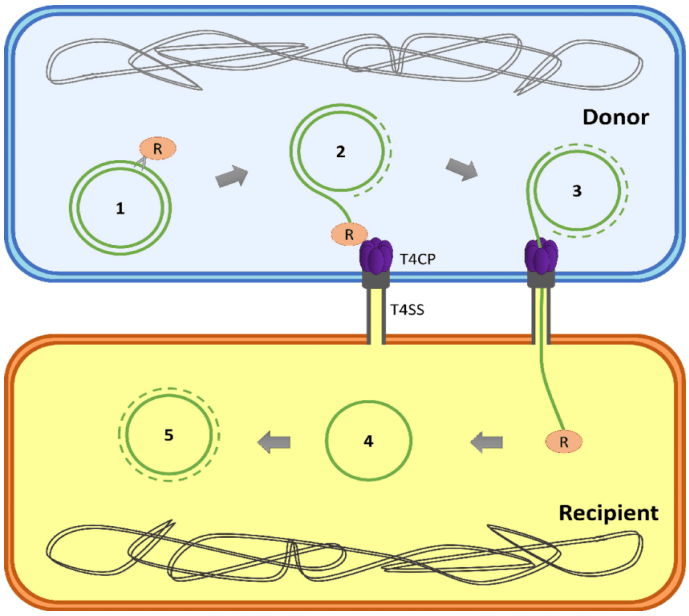



***Figure 6**. Schematic representation of the conjugative process. (Getino and de la Cruz, 2018).*

in processing DNA for conjugation, being the relaxase (R) a crucial one (Figure 6). The relaxase is the protein that in the donor strain cleaves the plasmid DNA at the *nic* site of the origin of transfer (*oriT*) forming a 5' covalent intermediate, called the relaxosome, with the *oriT* and other accessory proteins (step 1). Meanwhile replication starts from the 3' end of the cleaved strand using the uncleaved strand as template, the relaxosome unwinds the DNA and displaces the DNA strand to be transferred, called T-strand, towards the type IV secretion system (T4SS) (step 2). The coupling protein of the T4SS (T4CP) recruits the relaxosome and the transfer of the ssDNA through the channel starts, helped by the ATPase pumping activity of the T4SS (step 3). Once a complete copy of the plasmid ssDNA passes to the recipient cell, the relaxase circularizes the ssDNA molecule (step 4) before it is replicated to form a dsDNA complete copy of the plasmid (step 5) becoming the recipient cell a new donor (Getino and de la Cruz, 2018).

## 1.4 R388 plasmid

R388 plasmid is a prototype of the IncW incompatibility group due to its relatively small size (33,926 bp). IncW plasmids have a small copy number (two or three copies per cell), a wide range of antibiotic resistances, and a broad host range as there are members spread in a good variety of bacteria species. R388 has been found in the α-Proteobacteria *Agrobacterium tumefaciens,* the γ-Proteobacteria *Acinetobacter calcoaceticus, Pseudomonas aeruginosa, P. solanacearum, Salmonella typhimurium, Shigella flexneri, Stenotrophomonas maltophilia* and *E. coli (from where R388 was first isolated by* (Naomi Datta & R. W. Hedges, 1972). Actually, there are other IncW plasmids found in β-Proteobacteria, δ-Proteobacteria and even in Bacteroidetes (Fernández-López et al., 2006). Another feature of IncW is its replication, which starts bidirectionally from the *oriV* or origin of vegetative replication. In addition, IncW plasmids only conjugate on solid surfaces (Bradley et al., 1980) as pili consist of rigid filaments with a variable number depending on the specie.

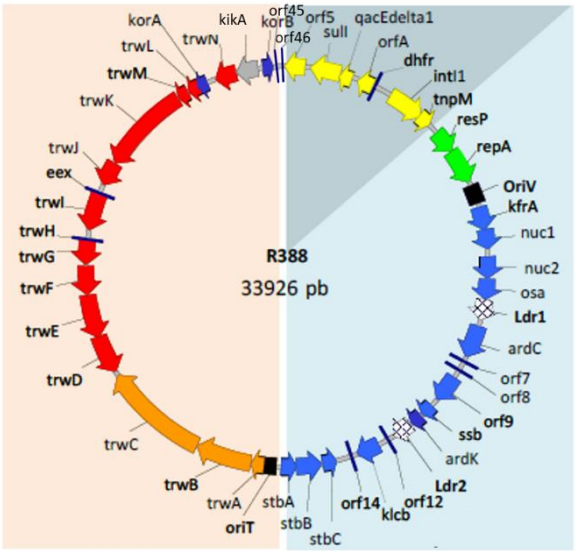



***Figure 7**. Genetic map of R388 plasmid. The figure shows the genetic organization of the plasmid in three sectors, conjugtion shadowed in pink, general maintenance in blue and antibiotic resistance and integration in grey. Adapted from (del Campo, 2016).*

R388 possess two non-coding long direct repeats (LDR1 and LDR2) that can create certain instability. In addition to the *oriV* and *oriT*, R388 has 35 genes annotated, which are divided in functional groups or modules (Figure 7). There is a sector (blue section in Figure 7) with genes for general maintenance (modules of replication, stable inheritance and establishment) where the "accessory" genes are located, being these genes the first ones to enter the recipient cell as they are in the leading region next to the *oriT*. There is second sector for antibiotic resistance and integration (grey section in Figure 7) and a third one for conjugation (modules of DNA transfer replication and mating pore formation) (salmon section in Figure 7) (Fernández-López et al., 2006) (Figure 7).

### 1.2.2 R388 regulatory network and fitness cost

R388 has strong promoters, however, they are strongly repressed. ArdK, KorA, StbA, ResP, KfrA as well as TrwA are the R388 transcriptional regulators that showed repression activity for some of the plasmid promoters of which the last three only repressed their own promoter by negative feedback loops as shown in Figure 8. On the other hand, ArdK controls the maintenance of the plasmid (P*ardC*, P*orf7*, P*ssb,* P*orf12*, and P*orf14)*, KorA is involved in the regulation of the expression of the pilus (P*trwH*, P*korA*, P*kikA* and P*korB*) and StbA essentially regulates the plasmid segregation (P*stbA*, P*trwH*, P*KorA* and P*kikA*) but also represses other promoters already regulated by KorA (P*ardC*, P*orf7*, P*ssb*, P*orf12* and P*orf14*) (Fernandez-Lopez et al., 2014).

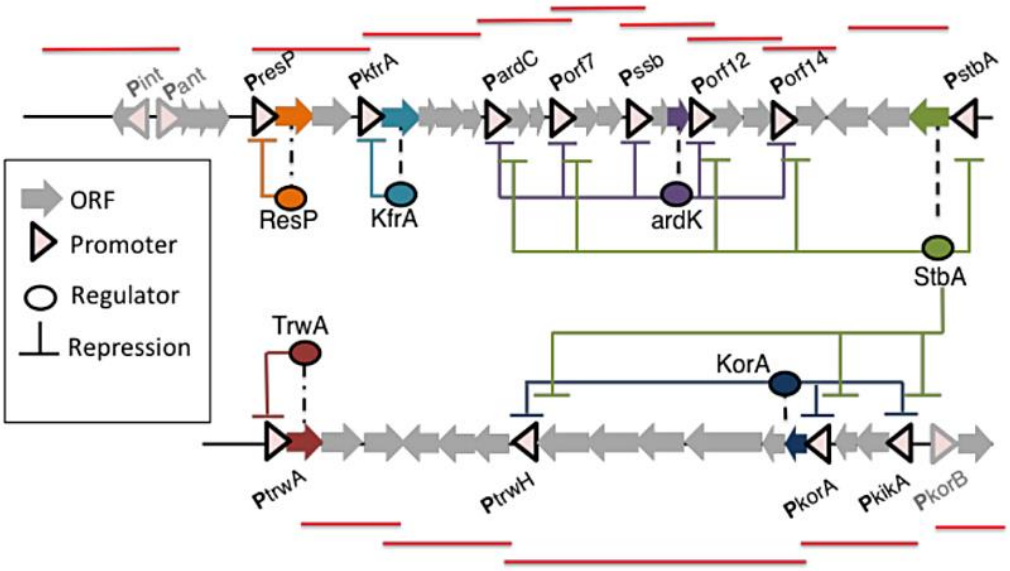



***Figure 8.*** *R388 plasmid transcriptional regulatory network. (Fernandez-Lopez et al., 2014).*

The point of this complex regulatory network is to generate a transcriptional overshoot after plasmid conjugation and the rebooting of the genome afterwards. This way, the plasmid can be highly infective. This network also showed to be independent of environmental changes as quorum signals or pheromones, temperature changes, different culture media, and stressing agents that trigger SOS response as none of the promoters were activated under these conditions (Fernandez-Lopez et al., 2014).

Plasmids involve a metabolic burden for the host fitness (Fernandez-Lopez et al., 2014). Plasmid fitness depends on the maintenance in the host and the ability to be transferred into new hosts. Therefore, the fitness cost of HGT comes from the sum of the cost of the transfer process, the integration of the incoming DNA in the bacterial genome, the cost of replication and the expression of the acquired genes, as well as the effects of the interactions between the acquired genes and the host. This is why it is so important for plasmids genes to be perfectly regulated and not wasting unnecessary resources from the host cell. The overriding relevance of the fitness cost of HGT falls into the opportunities of the incoming DNA to settle in the new host cell (San Millan et al., 2015).

### 1.2.3 R388 interspecific conjugation

Conjugation is a promiscuous process that allows DNA transfer between phylogenetically distant bacteria (Wilkins, 2002). As already mentioned R388 is a broad host range plasmid meaning that

it can be transferred to and maintained in a wide range of microorganisms. Previous results from our group showed the conjugation frequency of R388 from *Escherichia coli* to different bacterial species. They also showed the importance of the *kfrA* to *orf14* region (covering most of the stable inheritance and establishment functional modules) for efficient conjugation towards different bacterial strains as *Pseudomonas putida* or *Agrobacterium tumefaciens* (del Campo, 2016)(Figure 9). Thus, they showed the importance of this plasmid module for the broad host range character of the plasmid.

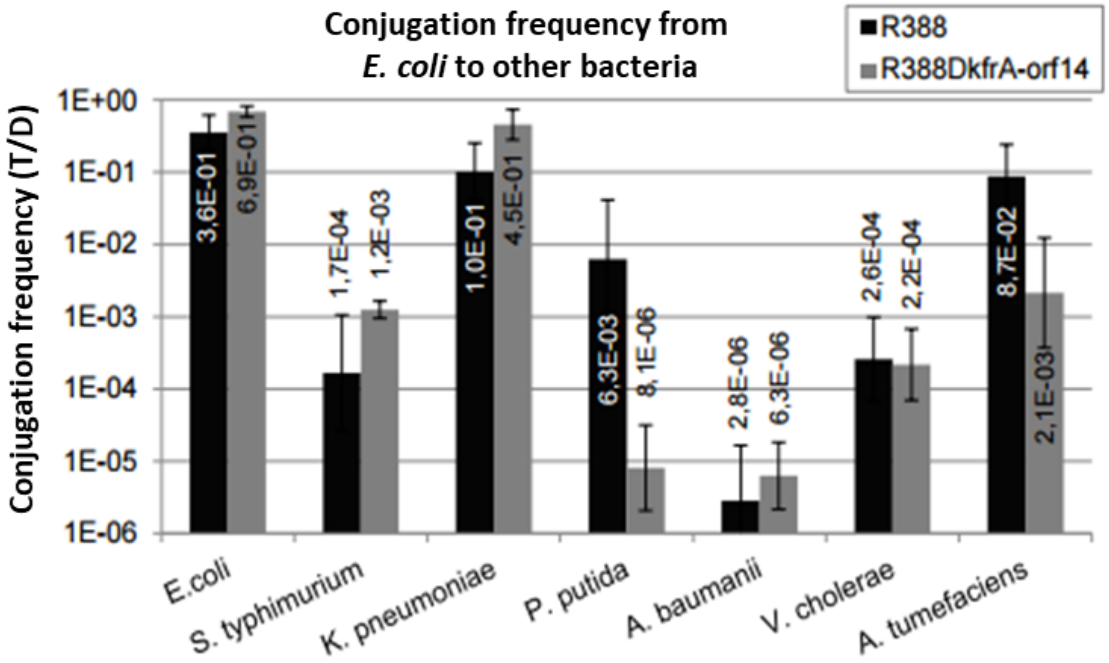



***Figure 9.*** *Conjugation frequencies of R388 and pIC10 from E. coli to other bacteria. Conjugation frequencies for R388 from E. coli to other bacteria are shown in black, and for pIC10, an R388 mutant without the kfrA to orf14 region, are shown in grey. Mean and SD calculated by doing the logarithm of the conjugation frequencies per donor and then the antilogarithms of the mean are shown. Adapted from (del Campo, 2016).*

## 1.3 Bacterial strategies to control conjugation

In the environment there are different factors affecting the conjugation rate, as it can be the temperature, the nature of the liquid or surface of conjugation, the pH, the cell density and composition of the population or the humidity (Aminov, 2011). Nevertheless, there are other natural strategies to control the conjugation based on genetic approaches as the strategies that are codified in the chromosome (host barriers) or plasmid DNA (plasmid barriers).

### 1.3.1 Plasmid barriers

#### 1.3.1.1 Entry exclusion

Entry exclusion is a mechanism own by all conjugative plasmids by which the host cell becomes to be a bad conjugation recipient, avoiding the entrance of a related plasmid. This feature plays an essential role in plasmid survival and fitness of the host, avoiding wasteful redundancy of plasmids that could generate instability by recombination between them and to prevent

recipient cells to die by lethal zygosis (membrane damage produced by an excess of conjugative cell contacts). Interestingly, there are two types of entry exclusion systems, first studied in F plasmid, and named after its genes *traS* and *traT* (Getino and de la Cruz, 2018).

TraT dependent exclusion is also called surface exclusion (SFX). *traT* gene produces protein S which reduces (about 10 fold) the ability of recipients to form multimeric aggregates on the outer membrane of the host, and thus, preventing the interaction between donor and recipient cell (Garcillán-Barcia and de la Cruz, 2008).

TraS dependent exclusion, also called entry exclusion (EEX) acts by inhibiting DNA transfer (about 100 fold) even if stable mating pairs are formed. This protein is found in the inner membrane and its activity to produce EEX activity is ligated to TraG protein blockade, also present in the inner membrane of the donor (Garcillán-Barcia and de la Cruz, 2008). Most plasmids have a single type of entry exclusion, usually of TraS type. For instance, R388 only has TraS dependent exclusion, encoded by the *eex* gene.

#### 1.3.1.2 Fertility Inhibition

Fertility inhibition systems reduce conjugative transfer of unrelated coexisting plasmids. FinOP system from F-like plasmids is the most well-known type. FinO protein reduces plasmid transfer by increasing the levels of the antisense RNA FinP in donors. FinP RNA precisely downregulates *traJ* mRNA translation, being TraJ the transcriptional activator of the conjugative transfer region(Jerome and Frost, 1999). At the same time, FinO binds FinP and *traJ* mRNA to protect FinP from cleavage by RNase E, thus, increasing FinP levels (Jerome et al., 1999). FinOP system regulates the conjugative transfer rate or epidemic spread in a bacterial population by just allowing a fell cells to be transfer-competent, regulating the conjugation rate and the plasmid burden balance (Frost and Koraimann, 2010).

Other genes have been found in different plasmid groups that perform fertility inhibition activity through different mechanisms. For example, *fiwA* and *fiwB* encoded by the IncP1α RP1 plasmid showed to inhibit the transfer of IncW plasmids R388, pSa or pRA3 reducing for instance R388 conjugation 1 million times in combination by inhibiting plasmid transfer (through *fiwA*) and pilus formation (through *fiwB*) (Fong and Stanisich, 1989).

### 1.3.2 Host barriers

In bacteria, CRISPR-Cas systems and restriction and modification (R-M) systems are the main mechanisms to avoid stable acquisition of foreign DNA detailed in Section 1.3.2.2 and Section 1.3.2.3 respectively. However, other defense systems could be involved in preventing gene acquisition by HGT:

For example, prokaryotic Argonaute-PIWI family of proteins are homologue to the eukaryotic Ago/PIWI nucleases involved in RNA silencing by RNA interference, the main defence system against viruses and transposable elements in eukaryotes (Makarova et al., 2009). Other example is the bacteriophage exclusion (BREX) system, a bacteriophage resistance mechanism that protects bacteria by innate immunity against virulent and temperate phages by replication and integration inhibition (Barrangou and Oost, 2015).

### 1.3.2.1 SOS response

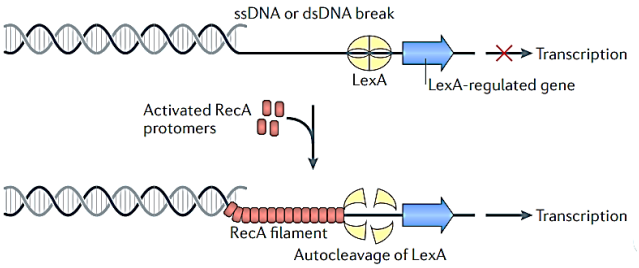



***Figure 10.*** *Schematic representation of RecA-LexA SOS response activation. SOS activation leads to the transcription of the LexA-regulated genes involved in DNA damage repair, recombination, and mutagenesis. (Andersson and Hughes, 2014).*

Another mechanism to control HGT is associated with SOS response. This response is turned on by an abnormal accumulation of ssDNA in the cell. In addition to the ssDNA induction by inner double strand breaks associated with DNA management within the cell, SOS response is induced in a variety of external circumstances as the presence of antibiotics or other environmental stressor factors that produce ssDNA, as UV irradiation, high pressure, gamma radiation, osmotic stress or reactive oxygen species (ROS). A third case to turn on the SOS response is the presence of ssDNA due to DNA uptake either by transformation, transduction or conjugation (Baharoglu and Mazel, 2014).

The mechanism to trigger the SOS response is as follows: RecA is recruited on ssDNA by RecBCD (that forms ssDNA substrate from recognized double-strand DNA breaks) or RecFOR (which recognizes DNA gaps and nicks) presynaptic complexes. The interaction of RecA protein with ssDNA produces a RecA-ssDNA nucleofilament that triggers the inactivation of LexA repressor by self-cleavage. LexA represses an SOS regulon by attaching to LexA box sequences on the promoters of those genes. This LexA auto-proteolysis leads to the derepression of the SOS regulon, inducing numerous genes involved in DNA repair, recombination, and mutagenesis (Baharoglu and Mazel, 2014; Roca and Cox, 1997) (Figure 10). This SOS regulon comprises a different set of genes for different bacteria. However, the ancestral core set of genes was determined to be formed by *recA, uvrA, ruvAB* and r*ecN* (Erill et al., 2007).

SOS has been demonstrated to enhance the horizontal conjugative transfer of antibiotic resistance genes in integrating conjugative elements (ICEs), as it has been observed in *Vibrio cholera* SXT ICE by cleavage of the LexA repressor homolog, SetR, that regulates integrase expression and ICEs propagation (Beaber et al., 2004). Thus, in addition to the DNA damage response function, we can consider SOS response as a bacterial mean to share information (Baharoglu and Mazel, 2014). What is more, another study showed that conjugation induces the SOS response in *E. coli* and *V. cholera* recipient cells (Baharoglu, Bikard, & Mazel, 2010). It is proposed that by this SOS induction, incoming DNA increases the chances of integration into the new host genome by SOS-mediated genomic rearrangements. This is especially interesting for narrow host range plasmids that are not able to replicate or be maintained in some bacterial strains. Even if DNA breaks occur, by the SOS induction associated to conjugation plasmids can induce genome plasticity in the recipient cell (Baharoglu and Mazel, 2014; Baharoglu et al., 2010).

On the other hand, some conjugative plasmids are able to prevent the SOS response through PsiB, a plasmid SOS interference (psi) system. PsiB is produced early during conjugation and inhibits the binding of RecA to the incoming ssDNA in recipient cells avoiding the formation of RecA-ssDNA nucleoprotein filaments and thus, avoiding the activation of the SOS response (Bagdasarian et al., 1992; Petrova et al., 2009). As proposed by (Baharoglu and Mazel, 2014), the motivation for this *psiB* gene to be present in some narrow host-range plasmids could be to prevent from a SOS response in the recipient cells that could lead to unwanted mutations in the

plasmid and a bigger fitness cost. Thus, the SOS response can act as a positive or a negative regulator of the conjugation process (Getino and de la Cruz, 2018).

#### 1.3.2.2 CRISPR-Cas systems

Clustered regularly interspaced short palindromic repeats (CRISPR)-Cas systems provide bacteria and archaea an adaptive immunity against foreign elements. This type of defence system is present in about 40 % of sequenced bacterial genomes and almost all archaeal genomes known (Marraffini and Sontheimer, 2008). CRISPR loci consist of an array of repetitive sequences of 24–47 bp, separated by unique short spacer sequences of similar size of viral or plasmidic origin. Usually these loci are flanked by an operon of CRISPR-associated (*Cas*) protein-coding genes that encode the machinery of the system (Grissa et al., 2007).

The CRISPR-Cas defence mechanism can be divided in two stages: immunization and immunity (Figure 11). In the immunization step, also known as adaptation or spacer acquisition, sequences from the foreign invading genome are captured into the cell and inserted at the beginning of the CRISPR array. This DNA fragments are formed by spontaneous breaks during plasmid or viral entrance. In the immunity step, the spacer is used against the invading genome that carries the complementary sequence for its destruction. For this step spacers need to be transcribed and processed to form small CRISPR RNAs (crRNAs). Afterwards, these crRNAs act as antisense guides for the Cas nucleases that specifically locate and cleave the targeted sequence in the foreign invading genome (Marraffini, 2015).

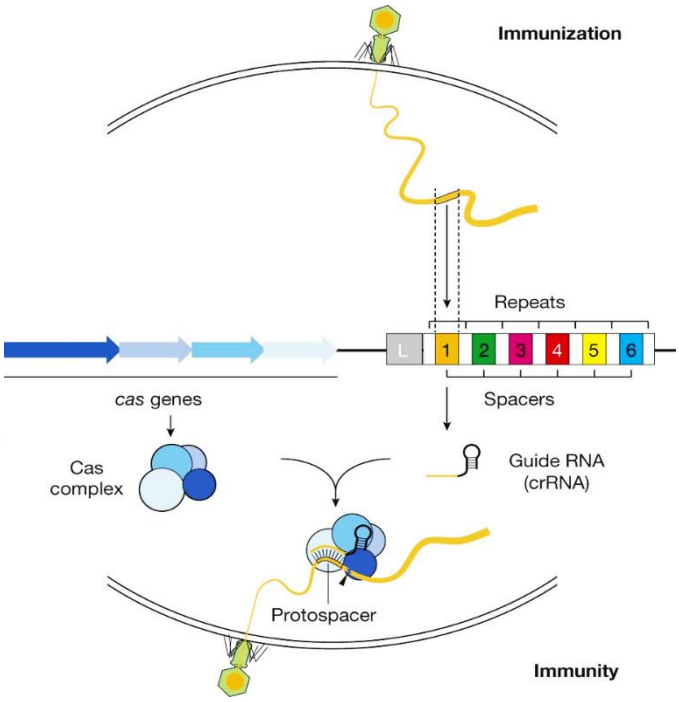



***Figure 11.*** *General overview of the Stages of CRISPR-Cas immunization and immunity. (Marraffini, 2015).*

To distinguish between own short spacer sequences of viral or plasmidic origin already in the bacterial CRISPR array and the invading DNA, CRISPR-Cas systems have different methods to protect themselves. CRISPR-Cas systems are divided in three types based on Cas protein content, targeting requirements and biogenesis of crRNAs. In Type I and Type II, a protospacer adjacent motif (PAM) sequence present in the invading DNA adjacent to the target is required for CRISPR-Cas activity and the lack of this PAM sequence in the CRISPR array confers autoimmunity. Interestingly, some viruses have developed the strategy of incorporating mutations in their PAM sequences to escape from the Type I and Type II CRISPR immunity system. In Type III no PAM sequences are required. Thus, it is through that distinctive base pairing between the crRNA and the CRISPR locus or the invading DNA target that Type III confers autoimmunity (Marraffini, 2015).

Highlighting the importance of CRISPR-Cas system in avoiding the acquisition of plasmids by HGT, a spacer present in the CRISPR loci of *Staphylococcus epidermidis* was discovered which specifically matched a region of the nickase gene of staphylococcal conjugative plasmids, which avoided the transfer of these plasmids (Marraffini and Sontheimer, 2008).

### 1.3.2.3 Restriction and Modification (R-M) systems

Restriction–modification systems let bacteria discern between their self DNA and the incoming DNA invading the cell, in addition to the destruction of this foreign one. They perform this function through two enzymatic activities: a modification methyltransferase that provides protection to the own host DNA and a restriction endonuclease that cleaves the foreign invading DNA (Gormley et al., 2005) (Figure 12).

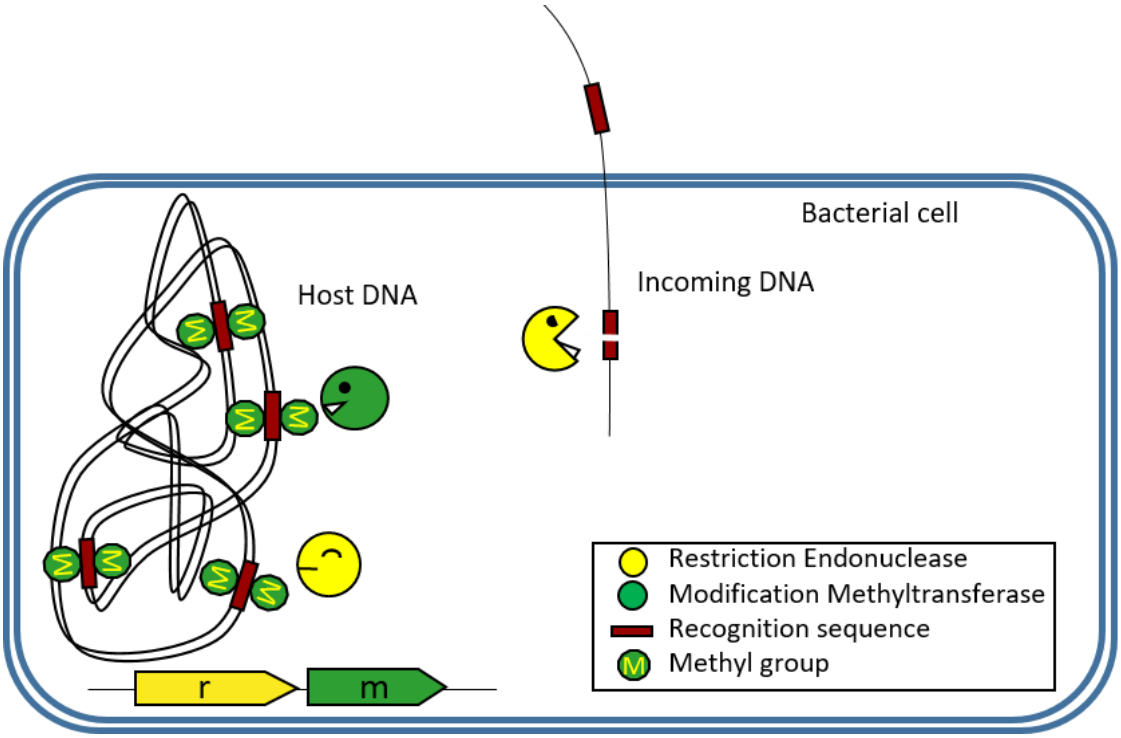



***Figure 12.*** *Schematic representation of restriction–modification (R-M) systems. Restriction of unmethylated incoming DNA invading the cell by a restriction endonuclease and protection of own DNA by methylation marks generated by a modification methyltransferase.*

Nevertheless, a R-M system will not safeguard a bacterial cell from invasion of DNA from an organism carrying the same R-M system (Furuta and Kobayashi, 2013) as the imprinted modification pattern will be the same as the one of the host DNA, and thus, looking like self DNA. The prevention of this situations is the reason why there are so many different R-M systems known up to know, which are collected in REBASE database (Roberts et al., 2015). R-M systems are widely distributed among bacteria and archaea. In fact, about 80 % of the sequenced organisms have more than one R-M system, especially those naturally competent bacteria as *Neisseria gonorrhoeae*, *Helicobacter pylori, Haemophilus influenza* or *Streptococcus pneumonia* as many R-M systems are specific against each individual competitor (Roberts et al., 2015; Furuta & Kobayashi, 2015).

Restriction-modification systems are also able to move through recombination to insert into the genome by themselves or by symbiosis with other mobile elements such as plasmids, phages,

transposons, integrons or conjugative elements/genomic islands involving important genome rearrangements (Furuta and Kobayashi, 2013). However, R-M systems are rare in prophages and almost absent in virus, and only about 10 % of the plasmids encode R-M systems compared to the 69 % of the sequenced chromosomes encoding them (Oliveira et al., 2014).

We can also say that they act as a tightly regulated toxin (restriction enzyme/degradation) antitoxin (modification enzyme/protection) system in mobile genetic elements. If during cell division R-M systems are not efficiently segregated, postsegregational killing of the progeny without the R-M-containing plasmids will occur. The reason why this happens is because the restriction enzyme (toxin) has a higher stability, which will attack the unmethylated host genome of the progeny lacking the methyltransferase modification enzyme (antitoxin). Thus, the R-M system has a stability role in plasmids carrying R-M genes (Mruk and Kobayashi, 2014).

In addition, methyltransferases methylate specific bases generating the following different epigenetic marks: N5- methylcytosine (m5C), N4-methylcytosine (m4C), and N6-methyladenine (m6A). This modifications are maintained after DNA replication, thus, they are in part responsible of the gene expression status of the cells (Furuta and Kobayashi, 2013).

To sum up, in addition to the primitive innate immune system as a defence from the invasion of foreign DNA, R-M systems have been postulated to play different secondary roles as in nutrition by the uptake of deoxyribonucleotides from invading degraded viral genomes or recombination as a way to acquire new genes and thus generate genetic diversity. In addition, they are responsible of the regulation of genomic fluxes and hence the rate of evolution, of the maintenance of epigenetic patterns to conserve transcription status, or to stabilize genomic islands. In general, to increase the relative fitness of the cell in the population under different environmental situations (Vasu and Nagaraja, 2013).

There are four main groups of R-M systems. They are classified according to the basis of enzyme subunit composition, cofactor requirements, DNA specificity characteristics and reaction products detailed in next subsections.

#### *1.3.2.3.1 Type I R-M system*

EcoKI is the prototype of this group. Three *hsd* (host specificity for DNA) genes are involved: *hsdR*, *hsdM* and *hsdS* in two contiguous transcription units, one for *hsdR* and another for *hsdM* and *hsdS* together. Type I R-M system is formed by the combination of three different types of subunits (R, M and S) to form $R_2M_2S$ complexes, however, $M_2S$ ones with only methyltransferase activity also exist (Murray, 2000). Three enzymatic activities are involved: endonuclease, methyltransferase and ATPase (required for restriction) and specific cofactors are needed: adenosine triphosphate (ATP), S-adenosyl methionine (SAM) and $Mg^{2+}$. The recognition site recognized by the S subunit, which contains two DNA binding domains, is a sequence of 13-15 bp, usually asymmetric and bipartite. For example, EcoKI recognizes $\frac{5'\text{AACNNNNNNGTGC}3'}{3'\text{TTGNNNNNNCACG}5'}$ with the highlighted bases being the ones to be N6-methylated. DNA cleavage is at a location away from the specificity site (in some cases several Kbp away) an occurs when DNA translocation collides (Mark R Tock and David TF Dryden, 2005; Wilkins, 2002). If the $R_2M_2S$ complex binds to a recognition sequence that has methyl marks on both strands, the complex does nothing. However, if the complex binds to a sequence only methylated in one strand, the methyltransferase activity is stimulated becoming a sequence methylated in both strands. If none of the strands of the recognition sequence are methylated, the $R_2M_2S$ complex will have a

strong restriction activity and low methyltransferase one, predominating the DNA cleavage. Type I R-M system is also divided in four families or subtypes, A, B, C and D. The genes encoding the IC type systems are generally encoded on plasmids, in contrast with IA, IB and ID, encoded in chromosomes (Gormley et al., 2005; Murray, 2000).

#### *1.3.2.3.2 Type II R-M system*

This system is composed by two independent proteins, a methyltransferase and an endonuclease encoded by two individual genes. However, the restriction enzymes usually associate in homodimers to be active, being each of the subunits in charge of the cleavage of one strand. $Mg^{2+}$ is needed for the nuclease activity and SAM for the methyltransferase one. The recognition site is usually symmetric and 4-8 bp long. For example, the prototype of type II R-M system is EcoRI, which recognition site is $\frac{5'G|AATTC3'}{3'CTTAA|G5'}$ with the highlighted bases being the ones susceptible of N6-methylation. DNA cleavage is at or near the recognition site, as no DNA translocation occurs. The endonuclease can produce blunt ends or overhangs (Gormley et al., 2005; Tock and Dryden, 2005).

#### *1.3.2.3.3 Type III R-M system*

Type III R-M system represent the smallest group out of the four, being EcoP1I and EcoP15I the best studied cases. They are conformed by two subunits, the restriction one called Res and the methylation one called Mod combined in a $Res_2Mod_2$ complex. Mod subunits are the one with DNA binding ability. They recognize asymmetric DNA sequences and methylate an alanine only in one of the DNA strands. For cleavage, two copies of the recognition sequence in inverted orientation are needed. Cleavage occurs at a fixed location 25-27 bp from the recognition sequence and it requires two molecules of the enzyme to bind to two inverted recognition sequences. Through DNA translocation towards the neighbour molecules they will end encountering each other and there, the $Res_2Mod_2$ complex is complete and able to cleave at both strands. Each DNA strand is cleaved by each ResMod complex half. Modification requires SAM and restriction requires $Mg^{2+}$ and ATP (Gormley et al., 2005; Tock and Dryden, 2005).

#### *1.3.2.3.4 Type IV R-M system*

Type IV R-M enzymes require two activities, methyltransferase (MTase) and endonuclease (ENase), in a single polypeptide chain. ENases only cleave DNA recognition sequences that have been modified, for example methylated, hydroxymethylated or glucosyl-hydroxymethylated. Recognition sequences are usually asymmetrical and two separated copies are needed. Cleavage by ENases occur between the two recognition sites. As the other types, they require $Mg^{2+}$ and, in addition, guanosine triphosphate (GTP) is required by the ENase translocation until collision initiates cleavage. McrBC from *E. coli K12* is the best studied one (Lepikhov et al., 2001; Tock and Dryden, 2005).

## 1.4 Antirestriction strategies

There is a coevolutionary arms race between bacteria to avoid entrance of foreign DNA molecules and parasitic DNA molecules as plasmids or bacteriophages to enter a putative host avoiding the restriction by bacterial R-M systems. The antirestriction mechanisms to counteract

R-M systems can be divided in four main types based on: DNA modification, transient blockage of restriction sites, sabotage of host R-M activities, and inhibition of restriction enzymes, schematized in Figure 13 and described below. Interestingly, several strategies can be present in a single mobile genetic element.

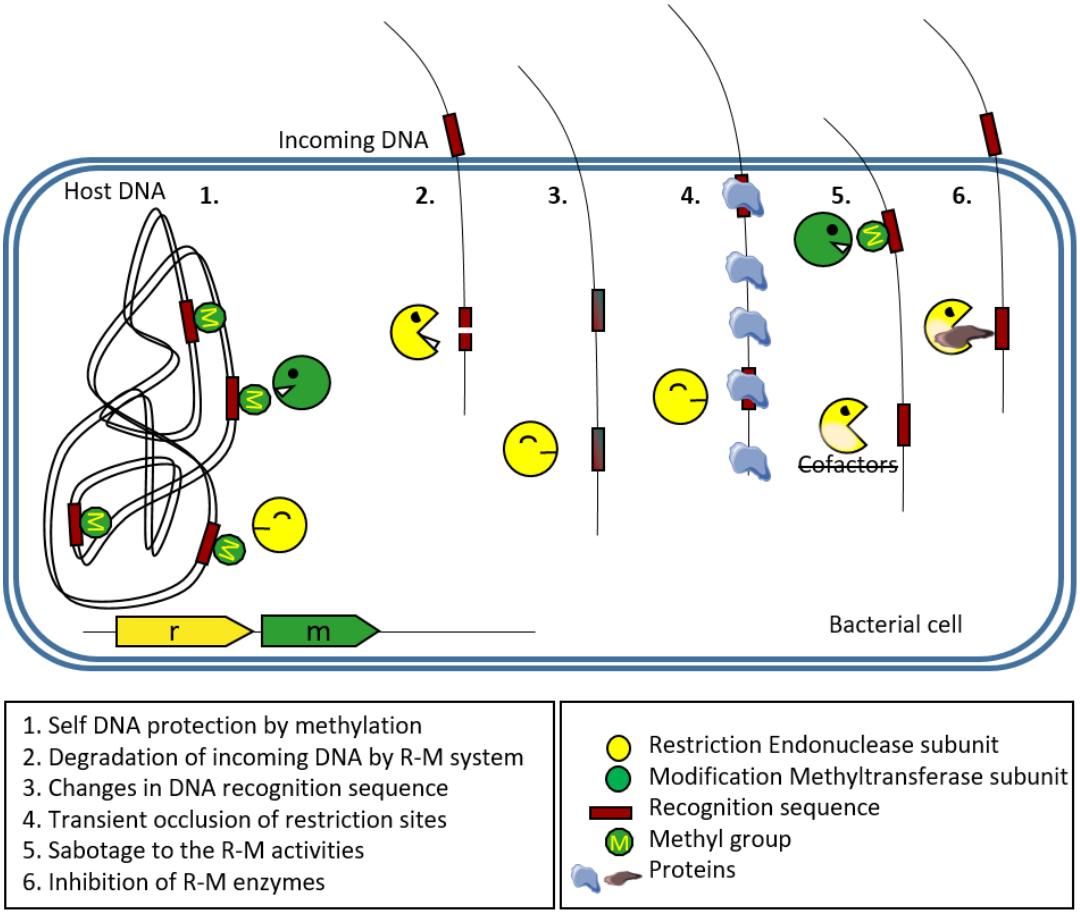



***Figure 13.*** *Schematic representation of host R-M system and antirestriction strategies. Host DNA protection (1) and degradation of foreign incoming DNA (2) by R-M systems and the four main groups of antirestriction strategies known (3-6) to avoid bacterial R-M systems are shown.*

## 1.4.1 Changes in DNA sequence

The simplest way to avoid restriction used by some plasmids and phages, as T7 and T3, is to remove or reduce the number of recognition sequences. But it has some limitations as it is mainly only effective to protect against type II R-M systems and it is more frequent in the genomes of non-temperate bacteriophages (Rusinov et al.). In addition, phage T7 avoids restriction by Type III R-M system by orienting all the recognition sequences in the same direction instead of in the required inverted orientation (Meisel et al., 1992). Other phages integrate rare bases in their DNA to overpass host R-M systems. For example, some phages substitute the thymines by 5-hydroxymethyluracil, others incorporate hydroxymethylcytosine into their DNA. On the other hand, phage Mu changes its recognition sites adenines by N6-(1-acetamido) adenines to avoid Type I and Type II R-M systems. Other phages, as SPb phage codify methyl transferases to protect its DNA by mimicking the host protection marks (Tock and Dryden, 2005).

### 1.4.2 Transient restriction sites occlusion

Transient occlusion of restriction sites occurs via the coinyection of phage- and plasmid-encoded DNA-binding proteins. For instance, DarA and DarB proteins encoded in the phage P1 genome are cotransported with the DNA into the bacterial host coating Type I recognition sites of the entering DNA (Iida et al., 1987).

### 1.4.3 Sabotage of host R-M activities

There are two different strategies known to alter the host R-M system. The first one consist on the host methiltransferase stimulation to modify invading DNA. For example, phage λ codes for a protein called Ral that stimulates the activity of Type I methyltransferases to methylate λ phage incoming DNA (Zabeau et al., 1980). The other alteration method is by sequestering the intracellular cofactors needed by R-M systems. For example, phage T3 encodes a SAM hydrolase that reduces the concentration of SAM as soon as it enters the cell reducing the activity of Type I, type II and Type III R-M systems that require SAM to be active (Studier and Movva, 1976).

### 1.4.4 Inhibition of R-M enzymes

This is the most frequent and studied mechanism out of the four and consist of direct inhibition of restriction endonucleases. For example, Ocr (overcome classical restriction) protein (also known as 0.3 protein) of T3 and T7 phages is soon expressed after entrance to the host cell. Ocr dimer mimics DNA in size, electrical charge and shape of a 24 bp DNA fragment and by binding to type I R-M enzymes blocks their DNA binding sites and thus, inhibits their restriction and modification activities (Atanasiu et al., 2002; Gormley et al., 2005; Tock and Dryden, 2005; Walkinshaw et al., 2002). In a similar DNA mimicking way, Arn (Anti restriction nuclease) protein from T4 phage inhibits type IV R-M system (Ho et al., 2014).

Conjugative plasmids and transposons also produce Ard (alleviation of restriction of DNA) proteins, as plasmidic ArdA and ArdB, which are also soon expressed after entrance by conjugation into the host cell. ArdA is a very acidic protein with a net charge of -22 to -29 negatively amino acids. It acts as a dimer to block the endonuclease as well as the methyltransferase of Type I R-M system by mimicking both in size and shape a bent B-form DNA molecule of 20 bp (Figure 14). It was first discovered in the IncI1 group of plasmids but it has later been detected in IncB, IncFV, IncK, and IncN incompatibility groups of plasmids. ArdA is only active in conjugation and do not protect during bacterial transformation probably because *ardA* is transcribed from a single strand DNA promoter *(ssi3)* form in the leading region of the plasmid T-strand (Walkinshaw et al., 2002; Wilkins, 2002). However, ArdU protein found in pUE10, a cryptic plasmid from *Deinococcus radiodurans*, instead of being a new type of Ard protein, it is a protein homologous to ArdA but was shown to be important for high transformation frequencies. They suggested that it is due to the protection of the incoming transforming DNA from degradation (Meima and Lidstrom, 2000).

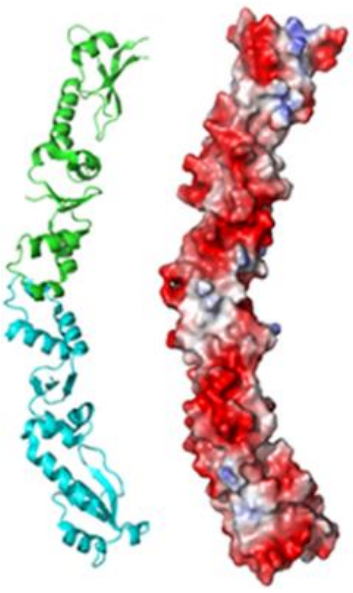


***Figure 14.*** *Tertiary structure and surface charges distribution of ArdA dimer, mimicking DNA. (Wang et al., 2014).*

ArdB first studied in IncN plasmids, and its homologue KlcA found in IncP plasmids, act indirectly specifically inhibiting the endonuclease activity of type I R-M systems. Activity against Type I R-M systems has been observed only *in vivo* and they produce no effect on modification reactions. In addition, weaker protection activity against Type II R-M systems has been noticed. They are slightly acidic proteins (with a net charge of 7 negative amino acids) (Goryanin et al., 2018; Serfiotis-Mitsa et al., 2009; Wilkins, 2002).

ArdD was the first antirestriction protein found in a non-conjugative transposon, specifically in the mercury resistance transposon Tn5053. It has activity against the endonuclease activity of Type I R-M systems. The activity of ArdD is through an SOS-dependent activity reduction of the Type I R-M complex linked to the proteolysis of the endonuclease subunit by the ClpXP protease (Balabanov et al., 2012; Zavilgelsky et al., 2014, 2015).

As it was determined by Belogurov group, all Ard proteins protect at least from Type I R-M systems. In addition, they all contain an "antirestriction motif", composed by 14 conserved amino acids that they proposed to be essential for the antirestriction activity. It is characterized by the presence of negative charges all along the motif separated by hydrophobic amino acids. In fact, a similar amino acid sequence has been found in the HsdS subunit of Type I enzymes, called the Argos repeat. The consensus sequence is "xIx*xxD*LxxxxD" with an x representing lack of consensus and * any polar amino acid (Belogurov and Delver, 1995; Zavilgelsky, 2000). This Argos repeat is known to interact with the hsdM subunit proposing a putative role of antirestriction motif disturbing the integrity of Type I R-M systems by competing with the Argos region of the hsdS subunit (Belogurov and Delver, 1995). In ArdA, which protein structure is solved (PDB: 2W82) the "antirestriction signature" has a structural role in maintaining the fold of ArdA rather than a direct role in inhibiting the R-M system (McMahon et al., 2009). In ArdB, also with a structure well studied (PDB: 2KMG) the conserved "antirestriction signature" does not necessarily imply a functional role (Serfiotis-Mitsa et al., 2009).

## 1.5 ArdC

Ard protein type C, was first studied from the IncW plasmid pSA (a plasmid more than 95 % identical to the R388 DNA sequence) first isolated from *Shigella* by (Watanabe et al., 1968). This protein was first studied by (Belogurov et al., 2000) who delimited ArdC to be 297 amino acids long, and 33.2 KDa. As ArdB, it is only slightly acidic (a net charge of 3 negative amino acids). In addition, they showed ArdC to have an antirestriction function towards type I and II R-M systems and a 14 amino acid "antirestriction signature" similar to the ones in ArdA and ArdB; in this case "LipDfdQS-aayvQ". ArdC shows no significant sequence similarity to ArdA and ArdB except for this signature. ArdC demonstrated to have a moderate protection activity against EcoK restriction system in comparison with ArdA and ArdB. They also observed that ArdC had a high degree of similarity (38 % identity) with the N-terminal region of TraC1 primase from RP4 plasmid.

The primase *traC* gene of RP4 plasmid codifies two products: TraC1 (1061 amino acids) and TraC2 (746 amino acids). Both TraC proteins have primase activity and thus are able to initiate the synthesis of the complementary strand during plasmid replication (Miele et al., 1991). Only TraC1 is known to travel to the recipient cell during conjugation, probably bound to the ssDNA

T-strand that is being transferred (Rees and Wilkins, 1990). TraC2, is in charge of the priming function, as it holds the primase motif EGYATA as part of the primase active center (Strack et al., 1992). Trac2 has not been detected on recipient cells so it may remain in donor cells to initiate the synthesis of the complementary strand (Rees and Wilkins, 1990)s & Wilkins, 1990).

(Belogurov et al., 2000) defined ArdC as a ssDNA binding protein able to protect *in vitro* this single-stranded DNA from type II restriction enzyme HhaI, an endonuclease able to bind both single and double strand DNA. Thus, they determined ArdC to be a protein that, as TraC1, could be transferred during conjugation bound to the plasmid T-strand in a way that will protect the incoming DNA from the host endonucleases, thus, forming part of the "Transient restriction sites occlusion" strategy group unlike the other known Ard proteins known. This antirestriction strategy may be employed by the broad host range conjugative plasmid R388 though the expression of ArdC. Due to its interest in overcoming immigration controls and as so little is known about this antirestriction protein, in this thesis we will focus in the structure and biological role of ArdC.

# 2 Aims and scope

Bacterial conjugation is the main mechanism for antibiotic resistance gene dissemination. Some plasmids are able to confer antibiotic resistances by transfer to a broad range of bacterial strains, becoming a well-known worldwide problem. Therefore, the study of the strategies used by plasmids to be remarkably promiscuous and the mechanisms to escape from recipient R-M systems is essential in the fight against the spread of antibiotic resistance genes. These strategies may be employed by the broad host range conjugative plasmid R388 though the expression of antirestriction proteins such as ArdC.

For this reason, our main objective is the characterization of the role and mechanism of action of the antirestriction protein ArdC through biological, biochemical and structural approaches. To reach this main goal, the specific objectives for this purpose were:

1. Biological characterization of the role of ArdC by conjugation of an R388-derivative plasmid without *ardC* towards different wt and mutant bacterial strains.

2. Analysis of the conjugation process by RNA-seq to identify the differentially expressed genes in the process.

3. Biochemical characterization of ArdC to check ssDNA and antirestriction activities of wt protein.

4. Biological characterization of ArdC mutant to check antirestriction activity.

5. Structural characterization of ArdC by X-ray crystallography.

# 3 Experimental procedure

# 3.1 Materials

## 3.1.1 Strains

***Table 1.*** *Bacterial strain used in this thesis.*

| Strain | Genotype/Relevant characteristics | Reference |
|---|---|---|
| ***Escherichia coli*** | | |
| DH5α | *F- end*A1 *gln*V44 *thi-1 rec*A1 *rel*A1 *gyr*A96 *deoR nupG, Φ80dlacZ*ΔM15 *Δ(lacZYA-argF)*U169, *hsd*R17($r_K^-$ $m_K^+$), $\lambda^-$ | (Grant et al., 1990) |
| BL21 (DE3) | $F^-$ *ompT gal dcm lon hsdS*$_B$($r_B^-m_B^-$) λ(DE3[*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [*malB*$^+$]$_{K\text{-}12}$($\lambda^S$) | (Studier and Moffatt, 1986) |
| C41 (DE3) | F- *ompT hsdSB (rB- mB-) gal dcm* (DE3) | (Miroux and Walker, 1996) |
| β834(DE3) | $F^-$ *ompT hsdS*$_B$($r_B^-$ $m_B^-$) *gal dcm met* (DE3) | (Budisa et al., 1995) |
| TB10 | TB10 is the result of a P1 transduction from DY329 into MG1655. It has a large amount of the λ prophage genome inserted into a biotin operon. The λ red genes α, β and γ are under the control of $cI^{857}$, making it temperature inducible. | (Yu et al., 2000) and (Silver et al., 2017) |
| BW27783 | $F^-$, Δ(*araD-araB*)567, ΔlacZ4787(::rrnB-3), $\lambda^-$, *rph-1*, Δ(*rhaD-rhaB*)568, *hsd*R514 Δ(*araH-araF*)570(::FRT), Δ*araE*p-532::FRT, *φPcp8ara*E535 | (Keasling et al., 2001) |
| BW27783-Nx | Nalidixic resistant spontaneous mutant of BW27783 | (del Campo et al., 2012) |
| BW27783-Rif | Rifampicin resistant spontaneous mutant of BW27783 | (del Campo et al., 2012) |
| EcMR2ΔmutS | MG1655, *lacI*$^-$ *bla, bio*$^-$*, lambda-Red1, mutS*$^-$, $cm^R$ | (Wang et al., 2009) |
| ***Pseudomonas*** | | |
| *Pseudomonas putida KT2440* | Wild-type strain; mt-2 derivative cured of its plasmid (pWW0-) | (Bagdasarian et al., 1981) |
| *Pseudomonas putida* EM178 | KT2440 derivative; Δ*prophage1* Δ*prophage4* Δ*prophage3* Δ*prophage2* | (Martínez-García et al., 2015) |
| *Pseudomonas putida* EM42 | KT2440 derivative; Δ*prophage1* Δ*prophage4* Δ*prophage3* Δ*prophage2* Δ*tn7* Δ*endA-1* Δ*endA-2* Δ*hsdRMS* Δ*flagellum* Δ*tn4652* | (Martínez-García et al., 2014) |
| *Pseudomonas putida* EM422 | KT2440 derivative; Δ*hsdRMS* | From De Lorenzo group |
| KT2440Δ*recA* | KT2440 derivative; Δ*recA* | From De Lorenzo group |
| KT2440 Δ*flagellum* | KT2440 derivative; Δ*flagellum* | From De Lorenzo group |
| KT2440 Δ*endA1* | KT2440 derivative; Δ*endA-1* | From De Lorenzo group |

| Strain | Genotype/Relevant characteristics | Reference |
|---|---|---|
| KT2440 Δ*endA2* | KT2440 derivative; Δ*endA-2* | From De Lorenzo group |
| KT2440 Δ*tn7* | KT2440 derivative; Δ*tn7* | From De Lorenzo group |
| KT2440 Δ*tn4652* pSW | KT2440 derivative; Δ*tn4652 bearing pSW plasmid* | From De Lorenzo group |
| KT2440 Δ*tn4652* | KT2440 Δ*tn4652* derivative cured of pSW plasmid | This work |
| **Agrobacterium** | | |
| *Agrobacterium tumefaciens GMI9023* | C58 derivative cured of its plasmids ($pTi^-$, $pAT^-$) | (Rosenberg and Huguet, 1984) |

### 3.1.2 Plasmids

***Table 2.*** *Plasmids already published and used during this thesis.*

| Plasmid | Description | Phenotype | Size (Kb) | Reference |
|---|---|---|---|---|
| R388 | R388 wild type plasmid | $Su^R$ $Tp^R$; (IncW) | 33.9 | (Datta & Hedges, 1972) |
| pSU2007 | R388 derivative; $Km^R$ cassette insertion | $Su^R$ $Tp^R$ $Km^R$; (IncW) | 32.9 | (Martinez and de la Cruz, 1988) |
| pIC10 | R388 derivative; Δ*kfrA-orf14* ($Km^R$ cassette insertion) | $Tp^R$ $Km^R$; (IncW) | 26.2 | (del Campo, 2016) |
| pET3a | Expression vector | $Ap^R$; Rep (pMB8); Overexpression controlled by T7 promoter | 4.6 | Addgene |
| pET29c | Expression vector | $Km^R$; Rep (pMB1); Overexpression controlled by T7 promoter with a 6-HisTag. | 5.4 | Addgene |
| pET29c:*ardK* | Vector for ArdK_R388 expression | $Km^R$; Rep (pMB1); Overexpression of ArdK controlled by T7 promoter with a 6-HisTag. | 5.7 | Our lab |
| pUCP22 | Shuttle Vector; *Escherichia-Pseudomonas* broad-host-range expression vector | $Ap^R$ $Gm^R$; Plac promoter. | 4.7 | (West et al., 1994) |
| pHERD20T | Shuttle Vector; *Escherichia-Pseudomonas* broad-host-range expression vector | $Cb^R$; $P_{BAD}$ promoter and *araC* regulator. | 5.1 | (Wiegand et al., 2008) |

***Table 3.*** *Plasmids constructed during this thesis.*

| Name | Insert | Promoter/Inducer | Size (Kbp) | Vector | Ab for Selection |
|---|---|---|---|---|---|
| pLGM21 | *ardC* | $P_{T7}$/IPTG | 6.1 | pET29c | $Kn^R$ |
| pLGM25 | $kn^R$ cassette | - | 33.9 | R388Δ*ardC* | $Kn^R$ |
| pLGM28 | *ardC_E229A* | $P_{T7}$/IPTG | 6.1 | pET29c | $Kn^R$ |
| pLGM33 | *ardC_E229A* | - | 32.9 | pSU2007 *(ardC_E229A)* | $Kn^R$ |
| pLGM36 | *ardC_l* | Plac/IPTG | 5.6 | pUCP22 | $Ap^R$ , $Gm^R$ |
| pLGM37 | *ardC_l_E229A* | Plac/IPTG | 5.6 | pUCP22 | $Ap^R$ , $Gm^R$ |
| pLGM38 | *ardC* | Plac/IPTG | 5.6 | pUCP22 | $Ap^R$ , $Gm^R$ |
| pLGM39 | *ardC_l* | $P_{T7}$/IPTG | 6.2 | pET29c | $Kn^R$ |

*ardC_l stands for the long gene version starting 63 nucleotides before de first methionine.*

### 3.1.3 Oligonucleotides

Oligonucleotides were purchased to Sigma-Aldrich (Madrid, Spain), IDT (Leuven, Belgium) or Eurofins (Louisville, USA).

***Table 4.*** *Oligonucleotides used for the construction of recombinant plasmids.*

| Name | Oligonucleotides 5'-3' [a] | Template [b] | Plasmid [c] | Method [d] |
|---|---|---|---|---|
| ArdCNterm | *TAGAAATAATTTTGTTTAACTTTAAG AAGGAGATATACAT*ATGAACGCAAA AACCAAGTTTGAC | R388 | pLGM21 | IA |
| ArdCCterm | *TAGCAGCCGGATCTCAGTGGTGGTG GTGGTGGTGCTCGAG*TGCGGCTTCT TTCCTTTGGA | | | |
| pET29CNdeI | ATGTATATCTCCTTCTTAAAGTTAAA C | pET29C | | |
| pET29CXhoI | CTCGAGCACCACCACC | | | |
| N_Kn_Promot er_Wanner | *AAATCAAAGCAGGCCCGGAAAAGCG CGGAAATGCAAGGGTTAAGCAGTGA* TACAGAGTTCTTGAAGTGGTGGCC | pET29C | pLGM25 | W |
| C_Kn_Wanner | *AACTGATGGCACAAAAAAAATCCCCC GCCGGAGCGGGGGAGGGCAGGTTA* GAAAAACTCATCGAGCATCAAATGA | | | |
| ArdC E229A d | ATTGCCGATTTCTGCAATTAGT**G**CCT CGAAAGCGTAGCTCTTGCG | pLGM21 | pLGM28 | QC |
| ArdC E229A r | CGCAAGAGCTACGCTTTCGAGG**C**AC TAATTGCAGAAATCGGCAAT | | | |
| ArdC_E229A_ MAGE | T*C*GTTTTAGCCGATTCAGCGACCG CAAGAGCTACGCTTTCGAGG**C**ACTA ATTGCAGAAATCGGCAATTGCATGC TTTGCGCAAGCCTTGG | pSU2007 | pLGM33 | M |
| | | pLGM34 +pUCP22 | pLGM36 | RE |

| Name | Oligonucleotides 5'-3' [a] | Template [b] | Plasmid [c] | Method [d] |
|---|---|---|---|---|
| ArdC E229A d | ATTGCCGATTTCTGCAATTAGT**G**CCT CGAAAGCGTAGCTCTTGCG | pLGM36 | pLGM37 | QC |
| ArdC E229A r | CGCAAGAGCTACGCTTTCGAGG**C**AC TAATTGCAGAAATCGGCAAT | | | |
| ArdC_short_fw(Eco) | GAGCTCGAATTCATGAACGCAAAAA CCAAG | pSU2007 | pLGM38 | RE |
| ardC_rev(Hin) | TGCAAGCTTTTATGCGGCTTCTTTCC | | | |
| ArdC_long_Nterm | *TAGAAATAATTTTGTTTAACTTTAAG AAGGAGATATACAT*GTGACCCGGAA CAAAGCGG | R388 | pLGM39 | IA |
| ArdCCterm | *TAGCAGCCGGATCTCAGTGGTGGTG GTGGTGGTGCTCGAG*TGCGGCTTCT TTCCTTTGGA | | | |
| pET29CNdeI | ATGTATATCTCCTTCTTAAAGTTAAA C | pET29c | | |
| pET29CXhoI | CTCGAGCACCACCACC | | | |

*a Oligonucleotide sequences: phosphorothioate (PS) bonds. Underlined: restriction enzyme recognition sequence. Italics: tails for recombineering. Bold: mutagenic introduction. Underlined and italics: linker. b Template used to obtain the fragments of interest. c Name of the recombinant plasmid generated. d Method used: IA: Isothermal Assembly method. RE: Restriction Enzymes method. QC: Quick-Change method. M: MAGE. W: Wanner method.*

***Table 5.*** *Oligonucleotides used for sequencing.*

| Name | Oligonucleotides 5'-3' |
|---|---|
| T7 | TAATACGACTCACTATAGGG |
| pT7 | GCTAGTTATTGCTCAGCGG |
| M13 fw | CGCCAGGGTTTTCCCAGTCACGAC |
| M13 rv | CAGGAAACAGCTATGAC |

***Table 6.*** *Oligonucleotides used for colony PCRs.*

| Name | Oligonucleotides 5'-3' |
|---|---|
| Up | CGCTCCCTTCACTCGGAAATC |
| Down | CGAACGGCCCGGATTGA |
| Middle Up | GGGGATCGCAGTGGTGAGTAAC |
| Middle Down | CTTTTGCCATTCTCACCGGA |

***Table 7.*** *Oligonucleotides used for protein binding and cleavage assays.*

| Name | Oligonucleotides 5'-3' |
|---|---|
| T87I1 (45b) | GAGCGCATCGGCCTTGACCTCATATTCAGCGCGCCCAAGAGCGTA |
| T87I2 (45b) | TACGCTCTTGGGCGCGCTGAATATGAGGTCAAGGCCGATGCGCTC |
| T87I2 (20b) | TACGCTCTTGGGCGCGCTGA |
| T87I2 (25b) | TACGCTCTTGGGCGCGCTGAATATG |
| T87I2 (30b) | TACGCTCTTGGGCGCGCTGAATATGAGGTC |
| T87I2 (35b) | TACGCTCTTGGGCGCGCTGAATATGAGGTCAAGGC |

| | |
|---|---|
| T87I2 (40b) | TACGCTCTTGGGCGCGCTGAATATGAGGTCAAGGCCGATG |
| T87I2 (50b) | TACGCTCTTGGGCGCGCTGAATATGAGGTCAAGGCCGATGCGCTCTTTGC |

***Table 8.*** *Oligonucleotides used for crystallization.*

| Name | Oligonucleotides 5'-3' |
|---|---|
| 5Ts (5b) | TTTTT |
| 8Ts (8b) | TTTTTTTT |
| 17mer (17b) | TGAGGATCCGGCTGCTA |
| 19mer (19b) | AGCCGCCGGGAATGGTCAG |

### 3.1.4 Culture medium

- LB (Luria-Bertani): 10 g of NaCl, 10 g of tryptone, 5 g of yeast extract in deionized H2O to a final volume of 1 litter. Adjust pH to 7.0 before autoclaving.
- LB Agar: Add 15 g of agar per litter of LB media.
- DYT (double yeast tryptone): 10 g yeast extract, 16 g tryptone and 5 g NaCl in deionized $H_2O$ to a final volume of 1 liter. Adjust pH to 7.0 before autoclaving.
- MHB (Mueller Hinton Broth): 2 g of beef extract, 17.5 g of acid hydrolysate of casein, 1.5 g of starch in deionized $H_2O$ to a final volume of 1 litter. Adjust pH to 7.3 before autoclaving.

### 3.1.5 Enzymes, antibiotics and other chemical reagents

FastDigest Restriction Enzymes, HhaI, T4 DNA Ligase, SYBR™ Safe, SYBR® Gold Nucleic Acid Gel Stain and DNA Phusion® High-Fidelity DNA Polymerase were purchased to Thermo Scientific. For Isothermal assembly, T5 exonuclease was from Epicentre and Taq DNA ligase was obtained from NEB, as well as RecA, Vent® polymerase and M13mp18 single-stranded DNA. Imidazole, lysozyme, N,N,N',N'-Tetramethylethylenediamine (TEMED), glutaraldehyde, Bovine serum albumin (BSA), Ribonuclease A (RBA) and all the antibiotics are from Sigma, as well as isopropylthio-β-galactoside (IPTG). SelenoMet™ Medium Base, SelenoMet™ Nutrient Mix and SelenoMethionine Solution were obtained from Molecular Dimensions. SYPRO® Orange is produced by Life technology, Proteinase K used is from Roche and Ammonium Persulfate (APS) and acrylamide from BioRad.

## 3.2 Methods

### 3.2.1 Microbiological methods

#### 3.2.1.1 Bacterial cultures

All strains were grown in Luria–Bertani broth (LB) at 30°C (*Pseudomonas Putida and Agrobacterium Tumefaciens*) or 37 °C *(Escherichia Coli* or *Pseudomonas aeruginosa)* overnight (o/n) with shaking (120 rpm). Plating was done on LB agar (1.5 %).

When needed, the following antibiotics were used at the indicated concentrations: ampicillin, 100 µg/mL; kanamycin, 50 µg/mL; chloramphenicol, 25 µg/mL; gentamycin, 10 µg/mL; streptomycin, 500 µg/mL; rifampicin, 50 µg/mL; carbenicillin, 100 µg/mL; tetracycline, 10 µg/mL and nalidixic acid, 20 µg/mL unless otherwise indicated.

For plasmid curation, four consecutive o/n liquid cultures without antibiotic were performed and plated in an ampicillin 1000 µg/mL and a chloramphenicol 25 µg/mL LB agar plates. Colony growing in chloramphenicol and not in ampicillin was further analysed by antibiogram and glycerol saved.

The bacterial strains were conserved at -20°C or -80 °C in glycerol-peptone (50 % glycerol (v/v) and 0.75 % peptone (w/v)) from cell pellets recovered at stationary phase.

#### 3.2.1.2 Competent cells and Transformation by electroporation

##### *E. coli electrocompetent cells*

Bacteria from a frozen stock were streaked on an LB agar plate and incubated at 37 °C o/n. A saturated culture from a single colony was obtained o/n with the appropriate antibiotic if needed and with shaking at 37 °C. Next day, 50 mL of LB in flask were inoculated with the o/n culture (1:20 dilution) and incubated at 37 °C with shaking until an $OD_{600}$ ~ 0.5-0.7 (approximately 1 h). Then, cells were incubated on ice for 30 min and transferred to sterile falcon tubes previously cooled on ice too. After centrifugation at 4,000 rpm for 10 min at 4 °C, supernatant was decanted, and pellets were resuspended by gently pipetting in 25 mL of ice-cold sterile water. Cells were again recovered by centrifugation at 4,000 rpm for 10 min at 4 °C and washed again in ice cold sterile water. The supernatant was decanted from the cell pellets and the pellets were resuspended in 25 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. 60 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 -80 °C until needed.

For transformation, an aliquot of 60 µL of competent *E. coli* cells was thawed on ice for no more than 15 min. If proceeding from a reaction mixture, DNA samples were previously dialyzed on 0.05 µm Millipore GS filters on a Petri plate with MilliQ water for 30 min to reduce the salt content. 100 ng of DNA were added to the cell suspension and placed in a 0.2 cm electroporation cuvette (Molecular BioProducts). Following electroporation at 2.5 keV and time constant (3-5 ms) in a Micropulser™ electroporator (Bio-Rad), cells were immediately recovered in 1 mL of sterile LB prewarmed at 37 °C. Transformed cells were incubated at 37 °C for 1 h and plated on selective LB agar plates.

##### *P. putida electrocompetent cells*

Bacteria from a frozen stock were streaked on an LB agar plate and incubated o/n at 30 °C. Next day, the following steps for the preparation of competent cells were done at room temperature (RT). Several colonies were scratched from the plate and resuspended in 1mL of sucrose 300 mM. Then, centrifugation at 8,000 rpm was done for 5 min. Supernatant was removed by pipetting in order to remove the unattached cells, mainly cells in clumps that are not interesting for making them electrocompetent. This washing step was repeated three times and after the last centrifugation, cells were resuspended in 200 µL of the sucrose solution. 100 µL aliquots were used for each electroporation, adding 1 µg of DNA to each tube and mixing well. Electroporation was done at 2.5 KeV and time constant (3-5 ms) in a Micropulser™

electroporator (Bio-Rad) adding immediately 1mL of DYT medium at RT and incubating with shaking 2 h at 30 °C. Then, the culture was centrifuged and the supernatant decanted leaving approximately 100 µL for resuspending the pellet and plating in LB agar plates containing the antibiotic to select the cells containing the electroporated plasmid.

#### 3.2.1.3 Bacterial conjugation experiments

Donor and recipient cell cultures were grown o/n at its optimal growing temperature with shaking (120 rpm) in the presence of the selecting antibiotics. After $OD_{600}$ measurement, the needed volumes to have an $OD_{600}$(D:R) = 0.6:0.6 in 1 mL were mixed and washed twice by resuspension in 1 mL LB and centrifugation to remove the antibiotics. After the last centrifugation, the cell mixture was resuspended in 30 µL of LB. The conjugations were done in solid LB agar plates, as described by (Bradley et al., 1980). R388 and its derivatives conjugate better in solid than in liquid. LB agar plates were previously incubated at the mating temperature with a 0.22 µm pore size cellulose acetate filter of 25 mm of diameter (Sartorius Stedim). The 30 µL of conjugation mixture were placed over the filter and kept at the conjugating temperature for the desired time. If not specified, the standard conditions were 1 h at 37 °C. After this time, the filters were removed with sterile tweezers and introduced in 1 mL LB, where the cells were resuspended by vortexing for a few seconds in order to stop the conjugation. This tube was considered dilution zero. 1/10 serial dilutions were done and 10 µL drops were plated in LB agar plates with the appropriate selecting antibiotics for donors, recipients and transconjugants. The plates were incubated o/n. The conjugation frequency was determined dividing either the number of transconjugants per donors or the number of transconjugants per recipients. For conjugations in the presence of complementing plasmids, IPTG was added to the conjugation mixture in order to have a 0.1 mM IPTG final concentration.

Representation of the mean ± SD as well as the comparison of the means between two different conditions by using t-test tool was carried out with GraphPad Prism® (v 7.04) biostatistics software (San Diego, CA). For comparing three groups, we used one-way ANOVA, with Dunnett's test to compares every mean to a control mean.

#### 3.2.1.4 Growing Curves

Bacterial growing curves were obtained using a Victor3TM 1420 Multilabel Counter plate reader spectrophotometer (Perkin Elmer) as plate lector with a 600 nm filter for 96-well microtiter plates. *P. putida KT2440* harbouring pSU2007, pIC10 or pLGM25 plasmids was streaked from a −20 °C freezer stock on selective LB plates with Kn and Cm and incubated at 30 °C o/n. Then, starting from three separate colonies per plasmid, o/n cultures were incubated at 30 °C and 120 rpm in LB with Cm and Kn. The grown cultures were diluted 1:1000 in fresh LB medium, and 150 µL of the dilution were pipetted in a 96-well plate, to be grown in the Victor3TM spectrophotometer at 30 °C and orbital agitation during 24 h. Absorbance was measured every 9.5 min and, to offset evaporation, 5 µL of water were injected every 22.5 min in the wells. The absorbance values obtained were transformed in $OD_{600}$ values using a calibration curve with measures from a conventional spectrometer in cuvettes of l = 1 cm. The generation time in exponential phase was calculated from the $OD_{600}$ data. The ln of $OD_{600}$ values between 0.2 and 0.5 were obtained to calculate the rate of growth (α) by linear regression. The generation time was calculated applying the formula g= ln (2)/$\alpha$. Each experiment was done 10 times for each culture, and the n = 30 data were averaged, and the mean and standard deviation is represented in the graph.

#### 3.2.1.5 Plasmid Stability

Plasmid stability was measured as described by (De Gelder et al., 2007). *P. putida KT2440* harbouring pSU2007, pIC10 or pLGM25 was streaked from a −20 °C freezer stock on selective LB plates with Kn and Cm and incubated at 30 °C o/n. For each strain, stability experiments were performed by triplicate, starting from three separate colonies, which were each inoculated into 5 mL LB with Kn and Cm. After incubation for 24 h at 30 °C with shaking at 120 rpm, cultures were washed to remove the antibiotics by centrifuging 1 mL culture and resuspending the pellet in 1 mL LB. From these cell suspensions, 5 µL were transferred to 5 mL LB. These inoculated cultures constituted time point zero. After diluted and plated onto LB-Cm plates, cultures were incubated for 24 h at 30 °C and 120 rpm. From then on, 5 µL of the full-grown cultures were transferred every 24 h (around 10 generations before reaching saturation) to 5 mL of fresh LB with Cm to select the strain and incubated at 30 °C with shaking at 120 rpm. At the same time, cultures were diluted and plated onto LB-Cm plates. Determination of the fraction of plasmid-free cells in the population was done by replica-picking 100 randomly chosen colonies per culture from the LB-Cm plates onto LB-Kn-Cm and LB-Cm plates. Cells that have lost the plasmid do not grow in the plate with Kn as selecting antibiotic.

#### 3.2.1.6 UV survival assay

UV survival assays were performed in order to determine if ArdC is involved in DNA repair after UV DNA damage. The sensitivity of *P. putida KT2440* empty or carrying PSU2007, pIC10 or pLGM25 were measured by a semi quantitative assay. Overnight grown cultures in LB medium containing selective antibiotic were subcultured until the $OD_{600}$ reached 0.5. The bacterial cultures were serially diluted and plated in drops onto sterile LB agar plates containing selective antibiotic. The plates were exposed to UV light (302 nm) at increasing times (0, 15sec, 30sec and 1min) using a UV transilluminator and incubated at 30 °C o/n protected from light. The UV survival rate between the different strains was analysed.

UV sensibility assays coupled to conjugation were also performed. 300 µL of donors' o/n culture and 300 µL of recipient o/n cultures were mixed and washed twice by centrifugation for 1 min at 13,000 rpm and resuspension with 1 mL LB. After the second wash, cells were resuspended in 90 µL of LB and 30 µL of this resuspension were placed in three different prewarmed plates with filters for 1 h at 37 °C as described in Section 3.2.1.3. Each plate was placed under the UV lamp at time zero or after 30 min of conjugation for 0 sec, 15 sec or 30 sec and the UV survival rate between the different strains was analysed.

#### 3.2.1.7 Minimum inhibitory concentration (MIC)

MIC was calculated with a protocol slightly modified from the standard broth microdilution method (Wiegand et al., 2008). Bacteria were grown o/n on LB agar plates. The next day, bacteria were scratched from agar plate and re-suspended in 1 mL sterile water in an Eppendorf tube. The optical density ($OD_{600}$) of a 1:10 dilution was measured to prepare a bacterial suspension of $OD_{600}$ = 1 (corresponding to approximately $1x10^9$ bacteria / mL) in a new tube with sterile water. The $OD_{600}$ of a 1:10 dilution was measured again in order to be more precise in the preparation of a bacterial suspension of $OD_{600}$ = 0.002 (corresponding to approximately $2x10^6$ bacteria / mL) in a 50 mL falcon tube with MHB medium.

A sterile 96-well microtiter plate (**Figure 15**) was prepared as follows: 180 µL MHB and 20 µL of the antibiotic solution were added to row A. Antibiotics to be tested were prepared at 10 mg/mL, 1 mg/mL and 0.1 mg/mL. 100 µL MHB were added to the rest of the rows. Then, 100 µL from row A were well mixed and transferred to row B and mixed carefully before transferring another 100µL to row C and so on until row H, were we discarded the remaining 100 µL. Finally, 100µL of the bacterial suspension were added to all rows (corresponding to $1x10^6$ bacteria / mL) and the plates were incubated at 30 °C or 37 °C for 16 h.

The MIC was read from the first wells with no growth at all considering the final antibiotic concentration in each well as shown in **Figure 15**. Every antibiotic and strain combination was tested at least three times and the mode obtained is shown.

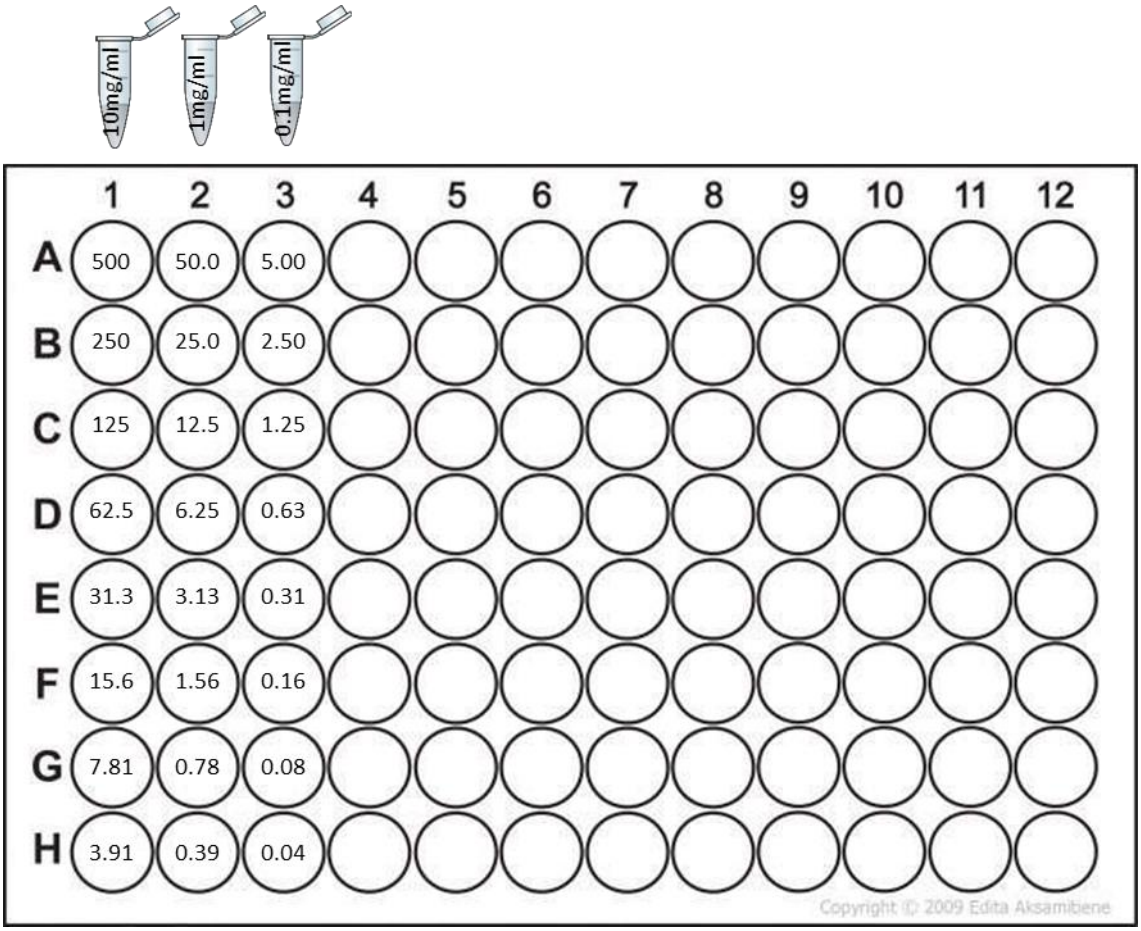



***Figure 15.*** *Schematic representation of the final antibiotic concentration in the 96 well MIC plate. Final concentrations in plate are in µg/mL starting from three antibiotic solutions with different concentrations shown in the top.*

### 3.2.1.8 RNA sequencing for transcriptome analysis

Conjugation experiments were carried out by the plate-mating procedure as described in (Llosa, Bolland, & de la Cruz, 1991) for 30 min at 37 °C and stopped with 1 mL LB. The conjugation protocol was optimized for this experiment as follows. A ratio of five donor cells per recipient was chosen in order to make sure that all the recipient cells could be in contact with a donor to start the conjugation process. The traditional protocol of conjugation was shortened to 30 min in an attempt to obtain the RNA synthesized in the recipient cell during the conjugation process.

Harvested cultures were treated with two volumes of RNAprotect® Bacteria Reagent (Qiagen). Cells were centrifuged, snap-frozen, and stored at -80 °C. Cells were lysed with 5 µg lysozyme

(Sigma) and 50 ng proteinase k (Roche). After cell lysis, total RNA was extracted with RNeasy® Mini Kit (Qiagen) and treated with RNase-free DNase (Qiagen) in column for DNA removal. Ambion® TURBO DNA-*free*™ DNase Treatment was also applied for a better DNA removal. In all cases, the manufacturer protocol was followed. RNA integrity and quality were validated by the Agilent RNA ScreenTape assay. The RNA integrity number equivalent ($RIN^e$) was assured to be above 8 to use the isolated RNA in the RNA-seq experiment.

Transcriptome libraries were prepared by Macrogen (Seoul, Korea) with the help of Ribo-Zero rRNA Removal Kit and TruSeq® Stranded mRNA sample preparation kit (Illumina) by following the Low Sample LS protocol. Libraries were sequenced by Macrogen on the Illumina HiSeq 4000 platform. The transcriptome libraries were paired-end sequenced with 100-bp reads.

Raw reads in FASTQ format were quality analysed with FastQC. For mapping the reads, sequences of R388 (NCBI Accession number NC_028464.1), *Escherichia coli str. K-12 substr. MG1655* (U00096.3) and *Pseudomonas putida KT2440* (AE015451.2) were used as genome template. The alignment of reads was done by each side independently with Bowtie2 software. Artemis program was used to visualize the alignment and do the RPKM (reads per kilobase and million mapped reads) calculations. RPKMs allow comparison of transcript levels between and within samples as it normalizes the RNA length and total read number.

$$RPKM = \#Mapped\ reads * \frac{1000 bases \cdot 10^6}{length\ of\ transcript * total\ number\ of\ maapped\ reads}$$

Genes with less than 10 RPKMs in all experimental conditions were removed from the analysis. DAVID online tool v6.8 was used to test for gene ontology enrichment among the list of differentially expressed genes in an attempt to do a functional classification.

## 3.2.2 Molecular Biology methods

### 3.2.2.1 DNA purification and extraction

For the extraction and purification of DNA, the following commercial kits were used: GeneJET Plasmid Miniprep Kit (Thermo Fisher Scientific) for *E. coli* plasmid extractions and QIAprep Spin Miniprep Kit (Qiagen) for *P. putida* plasmid extractions. GeneJET PCR Purification Kit (Thermo Fisher Scientific) was used for purification of PCR products and GeneJET Gel Extraction Kit (Thermo Fisher Scientific) for purification of DNA fragments separated in agarose gels. In all cases, the manufacturer protocol was followed.

### 3.2.2.2 DNA electrophoresis in agarose gels

DNA fragments were separated by electrophoresis in 1 % (w/v) agarose gels. Gels were prepared by dissolving the agarose in 0.5x TBE buffer (40 mM Tris-HCl, pH 8.3/ 45 mM boric acid/1 mM EDTA). To visualize the DNA, 5 µL of SYBR™ Safe were added per 100 mL of the agarose solution. Samples were mixed with 6x DNA sample-loading buffer (0.25 % bromophenol blue (w/v), 40 % sucrose in 0.5x TBE). Electrophoresis was performed at 120 V for 30 min in 0.5x TBE buffer. The DNA was detected using UV light in a GelDoc (Bio-Rad) equipment and the size of the DNA was determined using Thermo Scientific™ GeneRuler™ 100 bp or 1 kb DNA Ladders.

### 3.2.2.3 Plasmid construction by restriction enzymes

Phusion® DNA polymerase was used to amplify the inserts as it exhibits a 3'->5' proofreading activity assuring high specificity and yield of amplification. The oligonucleotides were designed with tails containing the recognition site for the desired restriction enzyme and some extra base pairs as indicated by NEB webpage (https://international.neb.com/tools-and-resources/usage-guidelines/cleavage-close-to-the-end-of-dna-fragments) needed for an efficient cleavage. C1000 Touch™ Thermal Cycler (BioRad) was used for all the PCR reactions following the enzyme's manufacturer protocol and the Tm calculated by NEB webpage (https://tmcalculator.neb.com/#!/main). PCR products were run on a 1 % agarose gel and extracted by the already mentioned GeneJET Gel Extraction kit eluting in 16 µL of MilliQ® water.

Thermo Scientific FastDigest Restriction Enzymes were used following the manufacturer indications for generating compatible sticky ends in both plasmid and insert. Usually, a mixture of 16 µL of DNA+ 2 µL buffer +1 µL of each enzyme was incubated at 37 °C for 1 h. Digested vector and insert were purified with PCR DNA kit eluting in 20 µL of MilliQ® water.

Usually, DNA ligation was performed with 100 ng of vector and using a molar ratio of vector to insert DNA of 1:1 and 1:3 with 5U of T4 DNA ligase for 1 h at room temperature following the manufacturer protocol. 20 µL of the constructed plasmid were electroporated into DH5α-T1$^{R}$ cells as described in Section 3.2.1.2 .

Colony PCR was performed with a conventional Taq polymerase to screen bacterial colonies with the desired plasmid product. For *E. coli*, a certain number of colonies were picked and lysed in 50 µL of sterile water by boiling the sample for 5 min. For *P. putida*, a certain number of colonies were picked and lysed in 100 µL of sterile water. Samples were heated for 10 min at 90 °C under shaking at 1,000 rpm to inactivate de nucleases and lyse the cells. Both *E. coli* and *P. putida* samples were chilled on ice for another 5 min and centrifuged 3 min at 13,000 rpm. 2 µL of the *E. coli* supernatant were used as template DNA for a conventional PCR with Taq polymerase, and 5 µL in the case of *P. putida* samples. Three or four positive colonies were grown o/n. A plasmid extraction was done with the appropriate kit, followed by a checking digestion with the same enzymes used for its creation. If the desired bands were observed, the sequences of the plasmids were confirmed by DNA sequencing.

### 3.2.2.4 Plasmid construction by Isothermal assembly

Isothermal assembly (Gibson et al., 2009) is a one reaction cloning method which is based in the homology between the ends of the fragments to be assembled. Fragments of interest were obtained by PCR with the high fidelity Phusion® polymerase and oligonucleotides containing a 50 bases homology sequence as a tail (**Figure 16**). PCR products were run on a gel and extracted. Then, the PCR products were digested for 5 min at 37 °C with 1 µL of DpnI FD restriction enzyme to eliminate possible template DNA.

The isothermal assembly reaction is a three in one reaction done at 50 °C with the following steps (**Figure 16**):

- T5 exonuclease removes 5' nucleotides from both strains of the fragments until one point when the enzyme is inactivated due to the temperature.

- Cohesive endings from fragments join by homology and Phusion® DNA polymerase fills with the template of the other strand.
- Taq DNA ligase joins the ends.

The 5x isothermal assembly reaction buffer needed to complete the three steps was prepared as follows: 3 mL 1M Tris-HCl pH 7.5, 150 µL 2 M MgCl2, 60 µL 100 mM dGTP, 60 µL 100 mM dCTP, 60 µL 100 mM dTTP, 60 µL 100 mM dATP, 300 µL 1 M DTT, 1.5 g PEG-8000, 300 µL 100 mM NAD and Milli-Q® water up to 6ml. 320 µL aliquots of this 5x buffer were prepared and frozen at -20°C. For preparing the reaction mix, the following volumes of enzymes were added to one of these 320 µL 5x reaction buffer aliquots: 1.2 µL T5 Exonuclease, 20 µL of Phusion® High-Fidelity DNA polymerase, 160 µL of Taq DNA ligase and 700 µL Milli-Q® water. 15 µL aliquots were prepared in PCR tubes and stored at -20°C. For each isothermal assembly reaction, 5 µL containing 100 ng of vector and the corresponding amount of insert for a molar ratio of 1:1, 1:3 or 1:5 were added to each 15 µL reaction aliquot. The reaction mixtures were incubated for 60 min at 50 °C. Later, samples were dialyzed for 30 min and electroporated in DH5α-T1$^{R}$ cells.

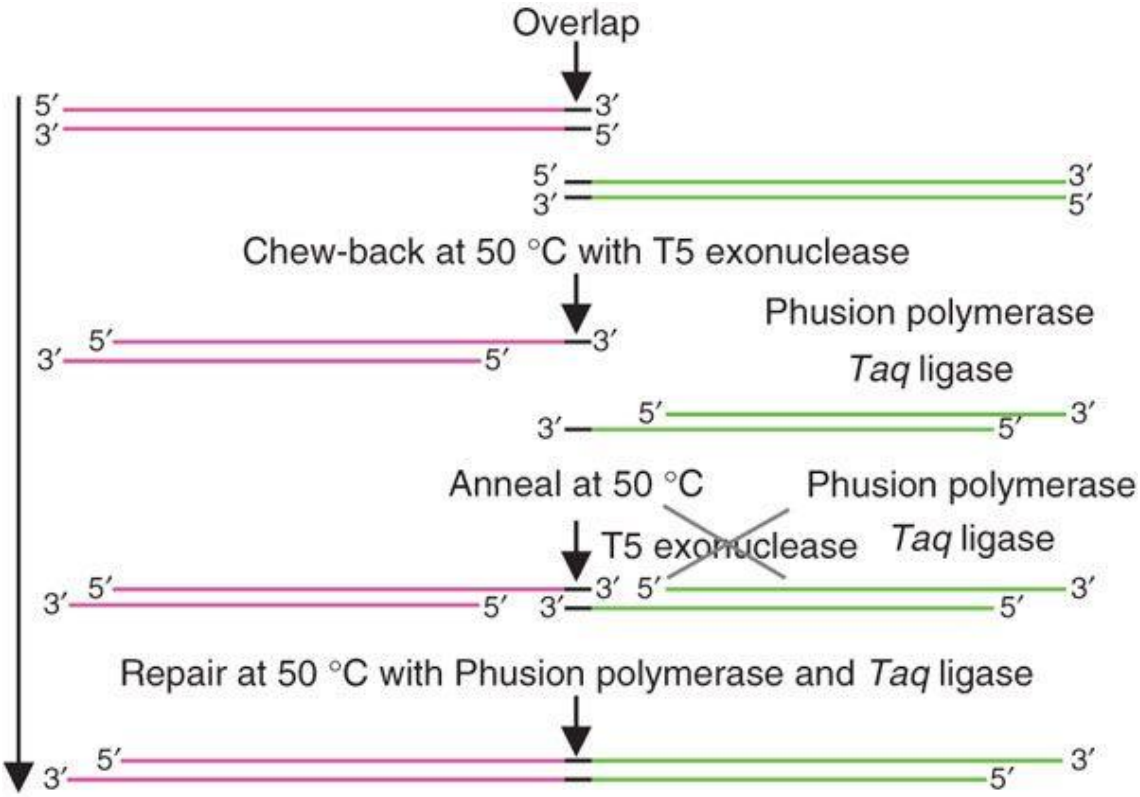



***Figure 16.*** *Schematic representation of the Isothermal Assembly cloning method. Image shows the different steps and enzymes involved in joining two DNA segments. (Gibson et al., 2009).*

### 3.2.2.5 Site-directed in vitro mutagenesis method

Point mutants in expression vectors were done by site-directed mutagenesis using an adaptation of the QuikChange II Site-Directed Mutagenesis Kit (Agilent Technologies) with Vent® DNA polymerase (**Figure 17**).

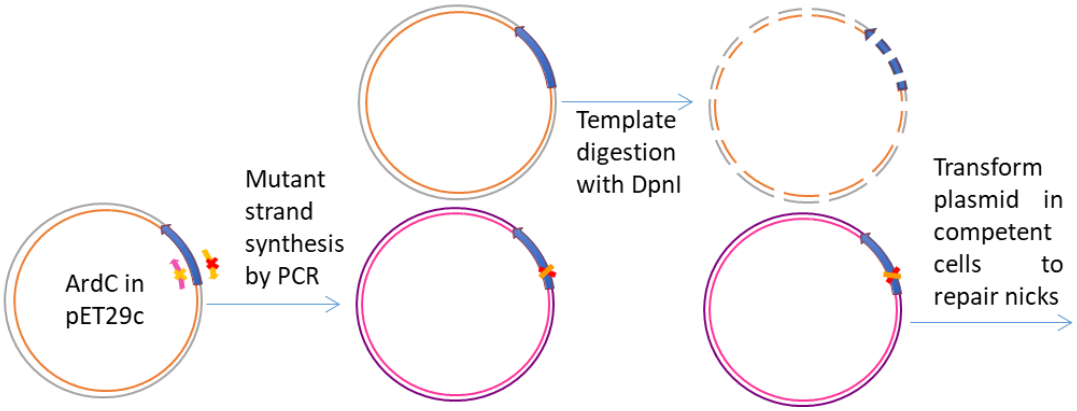



***Figure 17.*** *Schematic representation of the Site-directed QuickChange method.*

Complementary primers were designed to be 45 bases in length, with the desired mutation in the middle of the primers. The mutant strand synthesis reaction was done with the PCR conditions described below in **Table 9** and **Table 10**.

***Table 9.*** *PCR reaction mixture for site-directed mutagenesis with Vent® polymerase.*

| Component | Volume |
|---|---|
| 10x Thermo Pol buffer | 5 µL |
| 10mM dNTPs | 1 µL |
| 10uM forward primer | 0.5 µL |
| 10uM reverse primer | 0.5 µL |
| Template DNA | 100 ng |
| Vent® DNA polymerase | 0.5 µL |
| $H_2O$ | Up to 50 µL |

***Table 10****. PCR program used for site-directed mutagenesis with Vent® polymerase.*

| **Time** | **Temperature (°C)** | **Cycles** |
|---|---|---|
| 3 min | 95 | 1 |
| 30 sec | 95 | 30 |
| 30 sec | 68-72 gradient | |
| 1 min/Kb | 72 | |
| 5 min | 72 | 1 |
| ∞ | 4 | 1 |

10 µL of the products were analysed by electrophoresis in 1 % (w/v) agarose TBE gel. PCR products were treated with 1 µL of FD DpnI at 37 °C for 15 min to remove the template methylated DNA, leaving only the PCR product. FD DpnI was heat inactivated by incubating 20 min at 80 °C. Samples were desalted by membrane dialysis for 30 min. 2 µL of PCR product were transformed into electrocompetent DH5α-T1$^R$ cells and plated onto appropriate LB selective media for an o/n incubation at 37 °C. Plasmid DNA was extracted from some colonies and sequenced to verify that the selected clones contained the expected mutation.

As the method was not as efficient as expected, another protocol was used for doing the quick-change mutation to construct pLGM37. It was done with PfuUltra II Hotstart PCR Master Mix

(Agilent Technologies) with the mix shown in Table 11 and following the PCR program shown in Table 12.

***Table 11.*** *PCR reaction mixture for site-directed mutagenesis with PfuUltra II Hotstart PCR Master Mix.*

| Component | Volume |
|---|---|
| PfuUltra II Hotstart PCR Master Mix | 25 µL |
| DNA ( 100 ng/ µL) | 1 µL |
| P1 | 1 µL |
| P2 | 1 µL |
| $H_2O$ | 22 µL |

***Table 12.*** *PCR program used for site-directed mutagenesis with PfuUltra II Hotstart PCR Master Mix.*

| **Time** | **Temperature (ºC)** | **Cycles** |
|---|---|---|
| 1 min | 95 | 1 |
| 50 sec | 95 | 1 |
| 50 sec | 60 | X18 |
| 6 min | 68 | |
| 7 min | 68 | 1 |
| ∞ | 4 | 1 |

### 3.2.2.6 Site-directed MAGE in vivo mutagenesis method

In order to perform point mutations in bigger plasmids as R388 or its derivatives, we used the non-automated version of the MAGE (Multiplex Automated Genome Engineering) method described by (Wang et al., 2009).

A 90 bases oligonucleotide containing the mutation in the middle and two phosphorothioate (PS) bonds in the 5' end was designed. The phosphorothioate bond substitutes a sulphur atom for a non-bridging oxygen in the phosphate backbone of the primer. This modification in the internucleotide bond make the primer resistant to exonuclease degradation.

EcMR2ΔmutS *E. coli* strain was used. These cells were cultured at 30 °C when recombination was not needed. pSU2007 plasmid was introduced by conjugation into this strain and transconjugants ($Kn^R$ and $Rif^R$) were grown at 30 °C o/n. Next day, a 1/40 dilution was done in LB and cells were grown until an $OD_{600}$ of 0.5. Then culture was incubated at 42 °C for 15 min with shaking to induce the recombineering system and after this time cells were made electrocompetent as described in Section 3.2.1.2 except for the last wash, when cells were resuspended in 50 µL of a 1 µM oligonucleotide suspension so the mixture is ready for electroporation. Once recovered in 1 mL LB, cultures were grown at 30 °C. After 2 h, 50 µL were plated in Kn Rif plates and the rest of the volume was grown o/n labelled as cycle #1 until next day when the protocol was repeated. When needed, stocks were saved at -80 °C in 25 % glycerol for further analysis. After 10 cycles, plasmid extraction from some colonies was done and the PCR product obtained with Phusion® polymerase and oligonucleotides "Up" and "Down", was sent to sequence with "Down" oligonucleotide. The colony that gave an overlapping pick at the sequencing pane in the mutagenic position was further analysed by reelectroporating in DH5a-$T1^R$ and sequencing the PCR fragments from some of the colonies until a clean mutant pick was obtained for the desired position.

### 3.2.2.7 Deletion by Wanner and Datsenko

The R388Δ*ardC* plasmid (pLGM25) was constructed by a modification of the Wanner and Datsenko method (Datsenko and Wanner, 2000) which is based on the phage λ Red recombination system. The gene disruption strategy consists in the homologous recombination between the plasmid and a PCR product with homologous regions to the plasmid as represented in **Figure 18**.

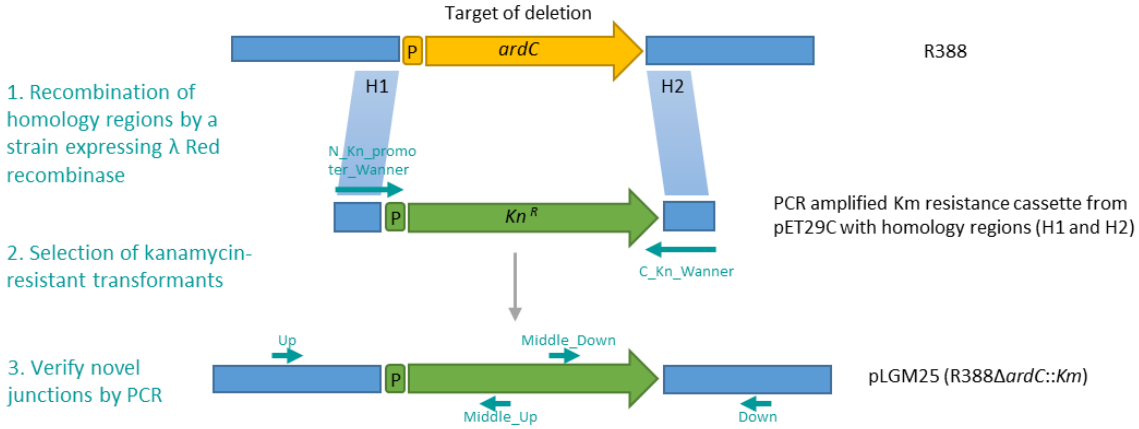



***Figure 18.*** *Schematic representation of the steps followed for deletion of ardC from R388 by a modification of the Wanner and Datsenko method.*

"N_Kn_promoter_Wanner" and "C_Kn_Wanner" oligonucleotides were designed to amplify by PCR the Kanamycin resistance gene with its promoter from pET29C by adding to both sides 50 nt of homology to the DNA flanking *ardC* cassette. Amplification was done with Vent® polymerase for 1 min and 20 sec of extension and an annealing temperature gradient of 58 °C - 70 °C following the Vent® manufacturer protocol. The 1227 bp DNA fragment was extracted from gel, treated with FD DpnI restriction enzyme, and dialyzed against water for 30 min.

The cell strain used for recombination was *E. coli* TB10. It is a strain that contains all the genes needed for λ RED recombination under the control of a temperature sensitive repressor in a way that when cells are exposed to 42 °C, recombineering genes are expressed as well as a nuclease inhibitor *recBCD* that prevents the degradation of linear DNA. These cells need to be cultured at 30 °C when recombination is not needed. R388 plasmid was introduced by conjugation to this strain and transconjugants ($Tc^R$ and $Tmp^R$) were grown at 30 °C o/n. Next day, a 1/70 dilution was done in LB and cells were grown until an $OD_{600}$ of 0.5. Then, the culture was incubated at 42 °C for 15 min with shaking to induce the recombineering system. After this time, cells were made electrocompetent as already described.

For transformation, 100 ng of the PCR product were used. Electroporation was performed as described in Section 3.2.1.2, but after transformation, cells were let at 30 °C for 3 h. Then, cells were plated in Tc and Kn at half the normal antibiotic concentration. After one day, colonies were restricken in a plate with the normal antibiotic concentration. Three colony checking PCRs were done with Taq polymerase with oligonucleotides "Up", "Down", "Middle_Up" and "Middle_Down" to verify the substitution. Plasmid DNA from some of the positive colonies was extracted and introduced in DH5α-$T1^R$ by electroporation in order to make sure that only one type of plasmid (mutated or WT) entered each cell. Another checking colony PCR was done for the selected mutated colonies with oligonucleotides "ArdC-Cterm" and "ArdC-Nterm" to make sure that they do not amplify any fragment and thus, confirm that we had the mutant plasmid isolated.

## 3.2.3 Biochemical methods

### 3.2.3.1 Protein expression and purification

For the overexpression of ArdC, a starter 50 mL culture of a *BL21 (DE3)-T1$^R$ E. coli* strain carrying pLGM21 plasmid was grown o/n at 37 °C with shaking (120 rpm) in LB medium supplemented with Kn. A 1/20 dilution was done in 1 L flasks containing fresh LB with Kn. In order to carry out the induction, when the culture reached an $OD_{600}$ = 0.5-0.6, isopropyl β-D-thiogalactoside (IPTG) was added to the culture to a final concentration of 0.5 mM. The overexpression was done o/n at 18 °C. To check the protein expression levels, an aliquot of 1 mL of the cell culture was harvested by centrifugation at 13,000 rpm for 5 min in an Eppendorf centrifuge 5415R. Then, pellet was frozen at -20 °C for posterior electrophoresis analysis.

For protein purification, induced cell cultures were harvested by centrifugation at 5,000 rpm at 4 °C for 15 min in a JA10 rotor (Beckman Coulter, USA) using a Beckman coulter Avanti J-30I centrifuge. Pellets were frozen at -80 °C, thawed and resuspended in 50 mL of buffer A (100 mM Tris-HCl pH 7.5, 20mM imidazole, 500 mM NaCl) supplemented with PMSF 1% (protease inhibitor phenylmethylsulfonyl fluoride). Then, they were sonicated in a *Labsonic 2000 (B. Braun)* equipment at 50 % of potency for 3 cycles of 1.5 min at intervals of 1 min on ice. The cell lysate was ultra-centrifuged at 40,000 rpm for 15 min at 4 °C on a *Sorvall® WX Ultra Centrifuge Series (Thermo Scientific)* equipment. To check the solubility of the protein, 30 µL of lysate and pellet samples were taken for further analysis by electrophoresis analysis.

The supernatant that contains the soluble protein was loaded onto a 5 mL nickel column HisTrap™ HP (GE Healthcare) previously equilibrated with buffer A. Proteins bound to the column were eluted by a lineal gradient between buffer A and B (100 mM Tris-HCl pH 7.5, 500 mM imidazole, 300 mM NaCl) at a 4 mL/min flux collecting 4 mL fractions in the *ÄKTA prime plus (GE Healthcare*) equipment.

ArdC containing fractions were pooled and diluted to a final NaCl concentration of 200 mM. Then, the sample was loaded onto an affinity chromatographic HiTrap® Heparin HP (GE Healthcare) column, especially designed for high-resolution purification of DNA-binding proteins. The column was previously equilibrated with buffer C (100 mM Tris-HCl pH 7.5, 200 mM NaCl). Elution of bound proteins was done by a lineal gradient between buffer C and D (100 mM Tris-HCl pH 7.5, 1 M NaCl) at a flow rate of 4 mL/min.

An additional step of size exclusion chromatography was done to separate proteins by size and shape when we needed a higher purity, and for having an approximation of the size and the oligomeric state of the protein. The fractions containing the protein were loaded in 500 µL aliquots onto a gel filtration Superdex™ S75 column 10/300 GL (GE Healthcare) with gel filtration buffer E (100mM Tris-HCl pH 7.5, 1 mM EDTA, 300 mM NaCl). The flow rate was 0.4 mL/min and the protein absorption was detected at wavelength 280 nm. In order to crystallize ArdC with a metal cofactor, all the lysis and purification steps were done as described but with 0.1 mM $NiCl_2$, 1 mM $CoCl_2$ or 1 mM $MnCl_2$ in buffers instead of EDTA. ArdC E229A was purified using pLGM28 plasmid following exactly the same protocol.

Samples of each elution fraction and purification step were analysed by SDS-PAGE. Fractions with the purest and highest protein concentration were aliquoted and, if needed, ultra-frozen with ethanol and dry ice to be conserved at -80 °C.

Preparation of selenomethionine (SeMet)-labelled ArdC was also carried out as described above but using strain *E. coli B834 (DE3)* and minimal medium (SelenoMet™ Medium Base + SelenoMet™ Nutrient Mix) supplemented with SelenoMethionine Solution (Molecular Dimensions) as indicated by the manufacturer.

ArdK protein was overexpressed and purified as ArdC with the following differences: The overexpression was done in strain C41(DE3) containing pET29c::*ardK* plasmid for 3 h at 37 °C. Cell pellets were resuspended in 50 mL of lysis buffer containing 100 mM Tris-HCl pH 7.5 and 1 M NaCl supplemented with PMSF 1%. For the purification step through the HisTrap™ HP (GE Healthcare) column, buffer A contained 100 mM Tris-HCl pH 7.5, 20 mM imidazole, 300 mM NaCl).

#### 3.2.3.2 Protein separation by SDS PAGE

Protein overexpression and purification process was analysed using the samples collected at different times. Pellets were resuspended in the same volume of protein loading buffer 2x (400 mM Tris-HCl pH 6.8, 4 % SDS, 30 % glycerol and 0.04 % bromophenol blue) and liquid samples were mixed in a 1:1 volume ratio with this buffer. Then, samples were boiled for 5 min before loading on a gel for a Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis (SDS-PAGE). 5 µL of PageRuler™ Plus Prestained Protein Ladder (Thermo Scientific) were used as protein standards. Electrophoresis was carried out at 180 V for 60 min in 1x SDS-PAGE buffer (25 mM Tris base, 192 mM glycine and 0.1 % SDS). Gels were then stained with Coomassie staining solution (0.1 % (p/v) Coomassie brilliant blue R-250, 50 % (v/v) methanol and 10 % (v/v) glacial acetic acid) for at least 30 min and destained with destaining solution (40 % (v/v) methanol and 10 % (v/v) glacial acetic acid) until background was clear. Gel images were acquired in an HP scanner.

#### 3.2.3.3 Determination of protein concentration

When needed, protein solutions were concentrated by ultrafiltration with Amicon Ultra Centrifugal Filter Units (Millipore, USA) or, for smaller volumes (up to 500 µL), with Vivaspin® 500 centrifugal filter units (Sartorius, Germany). Membranes with a pore size of at least 2-fold smaller than the molecular mass of the protein were used in order to be sure that we were not losing protein. The protein solutions were centrifuged at 4,000 rpm at 4 °C until the required volume was reached. Protein concentration measurement in solution was carried out by two different spectroscopic methods:

- Protein concentration was normally measured in a Nanodrop 2000c equipment (Thermo) spectrophotometer by UV absorbance at 280 nm using the molecular weight (MW) and the extinction coefficient value (ε) calculated by ProtParam tool from Expasy resource portal (http://web.expasy.org/protparam/).
- Bradford assay was used for protein concentration measurements, when bound to DNA. It is a colorimetric technique based on the ability of Coomassie Blue dye to change colour according to different protein concentrations. This effect can be observed by measuring the absorbance at 595 nm. A standard curve was made using Bovine Serum Albumin (BSA, Sigma-Aldrich) (range 0 – 15 µg). 20 µL protein samples were mixed with

1 mL of Quick Start™ Bradford 1x Dye Reagent (Bio-Rad). The reactions were allowed to proceed for 5 min at room temperature (RT) before measuring the absorbance at 595 nm on a Nanodrop 2000c spectrophotometer. Protein concentration was calculated according to the standard curve or with the naked eye for an inaccurate but faster measurement.

#### 3.2.3.4 DNA-binding Assays

We performed electrophoretic mobility gel assasy (EMSA) under non-denaturing conditions for short DNA sequences (<45 bases long). Increasing concentrations of ArdC were mixed with 45mer ssDNA oligonucleotide T87I2 or dsDNA oligonucleotides T87I1 and T87I2 (boiled together and cooled down slowly) at 0.3 µM final concentration in EMSA buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM EDTA, 1 mM DTT, 1 mM PMSF) in a 10 µL reaction mix. Reaction mixture was incubated for 10 min at RT. Afterwards, 2.5 µL of 6x DNA loading buffer (bromophenol blue 0.25 % (w/v), glycerol 40 % (v/v) in TBE 0.5x) were added to each reaction. Samples were loaded onto a 10 % acrylamide gel and electrophoresed in cold TBE 1x for 40 min at 180 V. Gels were stained for 30 min with SYBER® gold before visualization under UV light in a GelDoc system. 10 % polyacrylamide gels for separation of fragments 30 to 1000 bases long were prepared by mixing 1.5 mL TBE 10x, 3.75 mL acrylamide 19:1 40 %, 9,65 mL $H_2O$, 75 µL APS and 30 µL TEMED.

#### 3.2.3.5 DNA binding and protection assays

The assay was done under non-denaturing conditions to see DNA binding and retardation in parallel to under denaturing and proteolytic conditions:

M13mp18 ssDNA (7.2 Kb, 5.5 nM final concentration) was incubated with increasing concentrations of ArdC for 10 min at RT in a total volume of 20 µL binding buffer: 10 mM Tris-HCl, 10 mM NaCl and 10 mM $MgCl_2$. Then, 7 U of HhaI were added and incubated for 20 min at 37 °C. Afterwards, 2.5 µL of DNA loading buffer were added to 10 µL of the sample. Samples were subjected to electrophoresis on a 1 % agarose gel with SYBER safe for 30 min at 120 V. 1.5 µL of proteinase K at 20 mg/mL and 1 µL SDS 10 % were added to the remaining 10 µL of the sample and the mixture was incubated for another 20 min at 37 °C. Reactions were mixed with 2.5 µL DNA loading buffer and electrophoresed in a 1 % agarose gel with SYBER safe for 30 min at 120V.

#### 3.2.3.6 Crosslinking with glutaraldehyde

For evaluating the oligomeric state of ArdC, glutaraldehyde cross linker was used at increasing concentrations. Glutaraldehyde forms covalent bonds between two proteins by reacting with primary amine groups. 5 µL of ArdC at 13.2 µM in phosphate buffer (20 mM sodium phosphate pH7, 100 mM NaCl, 1 mM DTT and 5 % glycerol; buffer free from amines that could interfere) were incubated for 1 h at 37 °C with 5 µL of glutaraldehyde at increasing concentrations from 1 µM to 100 mM. Reaction was stopped by the addition of 4 µL of 0.5 M Tris-HCl pH 8. 14 µL of protein loading buffer 2x supplemented with 4 % β-mercaptoethanol were added, samples were boiled for 5 min and an electrophoresis was performed for 1 h at 180 V on a 12 % SDS-PAGE gel.

12% SDS-PAGE gel was done in two parts. For the lower or separating gel we mixed: 4.33 mL $H_2O$, 2.5 mL 1.5 M Tris-HCl pH 8.8, 3 mL acrylamide 29:1 40 %, 0.1 mL SDS 10 %, 0.05 mL APS 10 % and 0.02 mL TEMED. For the upper or stacking gel we mixed 2.8 mL $H_2O$, 1.3 mL 0.5 M Tris-

HCl pH 6.8, 0.85 mL acrylamide 29:1 40%, 0.05 mL SDS 10 %, 0.028 mL APS 10 % and 0.008 mL TEMED.

### 3.2.3.7 Proteolytic activity assay

ArdC proteolytic activity was analysed using a modification of the method described by (Ludanyi et al., 2014) for the study of IrrE metalloprotease.

ArdC at a final concentration of 3.45 µM in 20 µL was incubated in buffer P (10 mM Tris-HCl pH 7.5, 10 mM NaCl and optionally 10 mM of $CoCl_2$) and 45 mer ssDNA T87I2 or 45mer dsDNA T87I1/T87I2 with in a 1.5: 1 DNA: Protein molar ratio for 10 min at RT. Then 60 U of HhaI were added and incubated for 20 min at 37 °C. Reactions were stop by the addition of 20 µL of protein loading buffer 2x and boiled for 5 min. A 12 % SDS-PAGE was performed for 60 min at 180 V.

Similar experiments were performed to check if ArdC cleaved ArdK or RecA. 19 µM ArdC was mixed with 1.5:1 ssDNA and 1:1 ArdK, while 7 µM ArdC was mixed with 1.5:1 ssDNA and 1:1 RecA. The preincubation was also done for RecA and ssDNA before the addition of ArdC. Divalent cations or EDTA were added when indicated at 10 mM final concentration. 15 % SDS-PAGE was performed in both assays. The recipe of the gel was the same as described in the section before adjusting the acrylamide concentration.

### 3.2.3.8 Thermal Stability Assay based on fluorescence

ThermoFluor is a thermal stability assay based on fluorescence that serves for evaluating cofactor effects on protein stability and that can be used as an alternative to circular dichroism (CD). The increase in fluorescence intensity of SYPRO® Orange (Invitrogen) non-polar dye, due to protein unfolding as temperature is raised, can be measured. SYPRO® orange fluorescence is quenched when it is in solution, but not once it binds to the hydrophobic core of the unfolded protein (Figure 19).

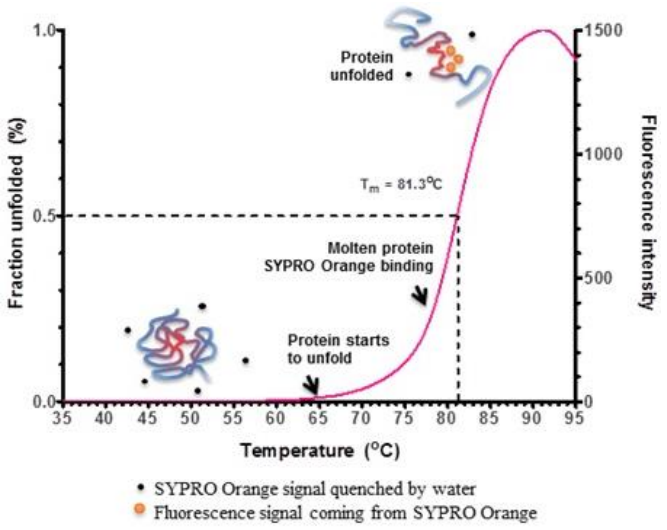



***Figure 19.*** *Schematic representation of the Thermal stability assay based on fluorescence by Sypro-Orange dye. In a Florescence vs. Temperature plot, the midpoint of the protein unfolding transition is the $T_M$ (melting temperature). Source: moleculardimensions.com.*

20 µL samples containing the protein of interest at 2.5 µM in 100 mM Tris-HCl pH 7.5 and 500 mM NaCl buffer containing 1mM of EDTA or the metal to be analysed or 7.5 µM DNA and the SYPRO® Orange dye at a 2x final concentration were evaluated in a StepOnePlus™ Real-Time PCR System (Thermo Fisher). For measuring SYPRO® Orange (excitation: 470 nm/ Emission: 570 nm), filter for NED™ dye was used, with excitation at 546 nm and emission of 575 nm. Temperature was raised from 25 °C to 85 °C at 0.5 °C per minute, measuring the fluorescence every 0.5 °C. Melting temperature ($T_M$) was determined as the maximum of the fluorescence versus temperature variation (dF/dT). The experiments were done by duplicate.

### 3.2.3.9 Crystallization

For protein crystallization, high purity protein samples (at least 90 % pure) were obtained by a third purification step by gel filtration as already described. The buffer of the protein samples was changed to crystallization buffer (20 mM Tris-HCl, 50 mM NaCl and 1 mM EDTA) with the aid of a centricon. In order to find initial crystallization conditions, the protein of interest was screened using Hampton Research Crystal Screen containing the 96 most common reagents that have produced crystals with other proteins and which cover a large range of precipitants, pH and different compounds. The solutions were pipetted into the reservoirs of 96 wells sitting drop plates in 50 µL aliquots. A 1 µL drop of the protein solution was then placed in the well of each chamber and mixed with 1 µL of precipitant solution from the reservoir. Plates were sealed with Crystal Clear Tape (Henkel Duck) and placed in an incubator at 22 °C until crystals were formed by vapour diffusion method. Once crystals were formed in any of these conditions, the precipitant concentration and the pH of the buffer were optimized in 24 well sitting drop plates mixing 1.5 µL of protein with 1.5 µL of precipitant solution.

In an attempt to crystallize ArdC bound to ssDNA, four oligonucleotides were tested “5Ts”, “8Ts”, “17mer” and “19mer” of 5, 8, 17 and 19 bases respectively. First, oligonucleotides were boiled and cooled down fast, then, a molar ratio of 1.5: 1 or 3: 1 DNA: protein was mixed and incubated for 30 min on ice before concentration on centricon until 15- 20 mg/ mL protein concentration.

For crystallizing ArdC bound to its metal cofactor, all the purification steps were done in the presence of metal salts as described in Section 3.2.3.1. The crystallization buffers contained 20 mM Tris-HCl and 150 mM NaCl supplemented with 0.1 mM $NiCl_2$, 0.5 mM $MnCl_2$ or 1 mM $CoCl_2$).

For the X-ray diffraction experiment, first, crystals need to be rescued from the crystallization plate with a nylon, fibre or plastic loop and immersed in the cryoprotectant solution, which is the same as the crystallization solution but with a cryoprotectant, to minimize the ice formation during cooling. This process was done at room temperature and then flash-frozen by immersion in liquid nitrogen if crystals were sent to synchrotron for a remote experiment, or flash-frozen directly under the nitrogen flux if the experiment was done at synchrotron. After that, crystals are mounted in the goniometer and ready to be diffracted by an X-ray beam.

All X-ray data were collected at beamline XALOC at the ALBA Synchrotron Radiation Facility (Barcelona, Spain) with a Dectris PILATUS3 6M Pixel detector. Data were collected at 105 K and 12,66 KeV by rotating the single frozen crystals in $\Delta\phi$= 0.25º steps through $180^{o}$-$360^{o}$. For the single SeMet crystal, data was collected at 0.9793Å, the wavelength corresponding to the heavy atom absorption maximum.

Diffraction images were processed using iMosflm (Battye et al., 2011) and Scala (Evans and IUCr, 2006) as part of the CCP4 package (Winn et al., 2011). For solving the phase problem, single anomalous dispersion (SAD) method was used thanks to the selenium introduced in the protein. For improving the resolution, data took from native protein crystals was used, solving the phase problem by molecular replacement (MR) with the structure obtained by SAD and using the program MolRep (CCP4). Refinement of the initial model was performed through several cycles by Phenix refine (Afonine P.V.; Grosse-Kunstleve R.W; Adams P.D, 2005) until appropriate R factors are reached. Final manual modelling was done in COOT (Crystallographic Object-Oriented Toolkit, (Emsley and Cowtan, 2004)). For the ArdC-Zn, ArdC-Mn and ArdC-ssDNA structures, MR was also used.

The atomic coordinates for ArdC structure (metal-free H3) have been deposited in the Protein Data Bank (accession code: 6I89).

### 3.2.3.10 Pull-Down assay

In order to find a putative ArdC target, we carried out a pull down assay to study the unknown proteins interacting with the predicted non-proteolytic ArdC mutant E229A.

2 L of *P. putida KT2440* grown o/n at 30 °C were centrifuged at 5,000 rpm at 4 °C for 15 min in a JA10 rotor (Beckman Coulter, USA) for Beckman coulter Avanti J-30I centrifuge. Supernatants were removed and pellet was frozen at -80 °C. Cells were then thawed and resuspended in 50 mL of lysis buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, and 20 mM imidazole). Then, cells were disrupted twice in a French press (Constant Systems LTD) at 4 °C and 23,000 psi and washed with 50 mL of lysis buffer. Lysed culture was ultra-centrifuged at 40,000 rpm for 15 min at 4 °C on a *Sorvall® WX Ultra Centrifuge Series (Thermo Scientific)* equipment. This lysate was divided in two fractions of 50 mL for performing two assays (I and II):

**Assay I:** Pull-Down with a previous removal of unspecific binding proteins:

The 50 mL of *P. putida* lysate were loaded with a peristaltic pump onto a 1 mL HisTrap™ HP (GE Healthcare), previously equilibrated with lysis buffer. The flow though was collected for further use.

4 mL of ArdC_E229A at 1 mg/mL in lysis buffer were loaded onto the 1 mL HisTrap™ HP, previously equilibrated with lysis buffer. Then, the FT from the previous step was also loaded. Finally, a coelution was performed with a lineal gradient between lysis buffer and buffer B (100 mM Tris-HCl pH 7.5, 500 mM imidazole, 300 mM NaCl) in the *ÄKTA prime plus (GE Healthcare*) equipment at 1 mL/min.0.5 mL fractions were collected.

**Assay II:** Direct Pull-Down

50 mL of *P .putida* lysate were mixed with 4 mL of ArdC_E229A in lysis buffer and incubated for 1 h 45 min at RT and agitation. This mixture was then loaded onto a 1 mL HisTrap™ HP (GE Healthcare) and eluted as described in assay I.

The proteins that co-eluted with ArdC were further analysed by gel filtration in the Superdex™ S200 column 10/300 GL (GE Healthcare) with buffer 150 mM NaCl and 50 mM Tris-HCl pH 7.5. Fractions were collected and analysed by SDS-PAGE. Samples containing ArdC and an additional protein were concentrated on a centricon and electrophoresed again by SDS-PAGE. The band corresponding to the putative ArdC-bound protein was excised from the gel and sent to Mass Spectrometry service (UPV, Leioa) for identification.

### 3.2.3.11 Mass Spectrometry analysis

To identify the putative ArdC protease target obtained by the pull-down assay, we did a protein identification assay by Liquid chromatography-tandem mass spectrometry (LC-MS/MS). Selected protein band was subjected to in-gel tryptic digestion according to (Shevchenko et al., 1996), with minor modifications. Gel pieces were swollen in digestion buffer containing 50 mM $NH_4HCO_3$ and 12.5 ng/µL proteomics grade trypsin (Roche, Basel, Switzerland), and the digestion processed at 37 °C overnight. The supernatant was recovered and peptides were extracted twice: first, with 25 mM $NH_4HCO_3$ and acetonitrile (ACN), and then with 0.1% (v/v) trifluoroacetic

acid and ACN. The recovered supernatants and extracted peptides were pooled, dried in a SpeedVac (ThermoElectron, Waltham, MA) dissolved in 10 µL of 0.1 % (v/v) formic acid (FA) and sonicated for 5 min. LC-MS/MS spectra were obtained using a SYNAPT HDMS mass spectrometer (Waters, Milford, MA) interfaced with a nanoAcquity UPLC System (Waters). An aliquot (8 µL) of each sample was loaded onto a Symmetry 300 C18, 180 µm x 20 mm precolumn (Waters) and washed with 0.1 % (v/v) FA for 3 min at a flow rate of 5 µL/min. The precolumn was connected to a BEH130 C18, 75 µm × 200 mm, 1.7 µm (Waters), equilibrated in 3 % (v/v) ACN and 0.1 % (v/v) FA. Peptides were eluted with a 30 min linear gradient of 3−60 % (v/v) ACN directly onto a homemade nano-electrospray capillary tip. Capillary voltage was set to 3,500 V and data-dependent MS/ MS acquisitions performed on precursors with charge states of 2, 3, or 4 over a survey m/z range of 350−1990. Raw files were processed with VEMS (Matthiesen et al., 2005) and searched against the NCBI non-redundant (nr) database restricted to Proteobacteria (version 20171205, 49911253 sequences) using the online MASCOT server (Matrix Science Ltd., London; http://www.matrixscience.com). Protein identification was carried out by adopting the carbamidomethylation of Cys as fixed modification and the oxidation of Met as variable modification. Up to one missed cleavage site was allowed, and values of 50 ppm and 0.1 Da were set for peptide and fragment mass tolerances, respectively.

Protein intact mass determination of ArdC and ArdC_I was carried out as follows: 40 µg of each protein sample were desalted using C4 and C18 Micro SpinColumn™ (Harvard Apparatus). Samples were dried in a Speed Vac (Thermo Scientific) and resuspended in 25 µL of 50 % acetonitrile, 0.25 % formic acid. Each protein was directly injected into a SYNAPT HDMS mass spectrometer (Waters) and MS spectra were manually acquired in the m/z range 500-1700. Protein intact mass was determined by MaxEnt1 software (Waters). Default deconvolution parameters were used. Mass ranges were selected based on available protein sequence information and software was set to iterate to convergence. Experimentally obtained masses were matched to protein amino acid sequences using the BioLynx tool embedded in MassLynx 4.1 software (Waters).

Mass spectrometry analysis were performed in the Proteomics Core Facility-SGIKER (member of ProteoRed-ISCIII) at the University of the Basque Country, UPV/EHU.

## 3.2.4 Bioinformatic methods

### 3.2.4.1 Databases

Nucleotide and protein sequences were obtained from the National Centre for Biotechnology Information (NCBI). Protein structures were obtained from the Protein Data Bank (PDB).

### 3.2.4.2 Study of DNA sequences

Vector NTI 10.3.0 (ThermoFisher Scientific) software was used for the visualization of DNA sequences, primer design and *in silico* construction of plasmids. It was also used for sequence alignment after sequencing.

NEB Tm calculator was used (https://tmcalculator.neb.com/) for calculating the melting temperature of primers. Nucleic Acid Package: NUPACK (Caltech) was used to predict the secondary structure of DNA sequences.

Blast (*National Center for Biotechnology Information,* NCBI) online server was used for finding DNA homology sequences from a query.

We used KEGG (Kyoto Encyclopedia of Genes and Genomes) for the visualization and the study of the genetic environment of our genes of interest.

SignalP was used to predict the presence and location of signal peptide cleavage sites in amino acid sequences.

BPROM (Softberry) was used for the prediction of promoter sequences.

ExPASy translate tool was used for translating DNA sequences into amino acid code.

#### 3.2.4.3 Study of proteins

For the protein alignment, BlastP online search (*National Center for Biotechnology Information,* NCBI) was used followed by an alignment using ClustalOmega. BlastP was also used for studying the protein distribution in nature.

Pymol software (Delano 2002) was used for the visualization of protein structures and image generation.

The ProtParam tool (Expasy) was used to obtain theoretical parameters of the protein such as molecular mass or extinction coefficient.

For structure predictions by homology modelling and finding structural homologues of the protein of interest, we used Swiss-Model (University of Basel, (Waterhouse et al., 2018)) and Phyre2 server (Imperial College of London, (Kelley L. et al., 2015)).

APBS Pymol tool was used for calculating the surface electrostatic potential of the protein with the aid of PDB2PQR Server.

We used Stride and ESPript3 web servers for obtaining the secondary structure representation of proteins.

PISA server was used to predict the multimeric state of a protein according to the interactions calculated from a pdb file.

PDBsum server was used for obtaining a topological representation of a protein from its structure.

Consurf server was chosen for studying the evolutionary conservation of amino acid positions in a protein.

# 4 Results

## 4.1 *ArdC* genetic characterization

### 4.1.1. ardC shows two possible starting codons

*ardC* in R388 plasmid has the same coding sequence as *ardC* of pSA plasmid, already described by (Belogurov et al., 2000). Belogurov et al. stated that *ardC* gene starts at a methionine codifying ATG codon. However, an alternative Valine starting codon (GTG) was found 63 nt upstream (implying 21 extra amino acids). This last annotation was done by Hisashi Anbutsu, from the National Institute of Advanced Industrial Science and Technology, Institute for Biological Resources and Functions in Japan. "NCBI Reference Sequence: NC_028464.1; protein_id="YP_009182134.1" and in "NCBI Reference Sequence: BR000038.1; protein_id="FAA00054.1" (Swift et al., 1981) (Figure 20).

GUG is the second most common starting codon in prokaryote genes present in about 14 %, while AUG in 83 % (Hecht et al., 2017). On average, genes starting by AUG are expressed up to higher levels than those genes that start by GUG. However, this data is dependent on the organism and its GC content. The Shine-Dalgarno (SD) sequence or ribosome-binding site (RBS) is an important translation initiation signal that also determines the efficiency of initiation. Weaker starting codons, as it is GUG, tend to be preceded by stronger (more conserved) SD sequences to compensate by having a stronger translation initiation signal (Belinky et al., 2017). However, It has also been proposed that the role of these weaker alternative starting codons in combination with weak SD sequences camouflaged in secondary structures is to allow for a tightly regulated expression as it is found in the translation initiation region of the ColIb-P9 plasmid *rep* gene (Asano, 2014).

We could not find the RBS upstream of the GUG starting codon, neither hidden by a secondary structure. Thus, we expect *ardC* to start at the methionine AUG starting codon. However, and although the preceding 21 amino acids were not predicted to be part of a signaling peptide either (according to SignalP software), some experiments were done with the coding sequence starting from this GUG codon and are indicated as *ardC_l* standing for long *ardC* sequence.

```
ggtt aag cag tga aaa aaa agt gtt gac acc tat tga caa ccg ccc tgt ttt gca tca taa

tag cat cat ggt tgc atc act ggt gca acg cag gtg acc cgg aac aaa gcg gcg gca acc
                                            V   T   R   N   K   A   A   A   T
gct ccg ggt ttt gga act ctc gca gga gtc gaa aca atg aac gca aaa acc aag ttt gac
 A   P   G   F   G   T   L   A   G   V   E   T   M   N   A   K   T   K   F   D
ctt tac caa cac gtt acc gac cgc atc att gcc agc att gag gca ggc acc ccg gca tgg
 L   Y   Q   H   V   T   D   R   I   I   A   S   I   E   A   G   T   P   A   W
cgc aag ccg tgg act ggt gaa gcc gcc aca atg caa atg ccg ctg cgc tcc aat ggc gaa
 R   K   P   W   T   G   E   A   A   T   M   Q   M   P   L   R   S   N   G   E
gcc tac cgg ggc atc aat gtt gtg atg ctt tgg ctt acc gcc gcc gaa aaa ggc tac cgc
 A   Y   R   G   I   N   V   V   M   L   W   L   T   A   A   E   K   G   Y   R
agc gcc tat tgg ttc acc tat cgc caa gct aaa gaa ctt ggg ggg cag gta cgc aaa ggc
 S   A   Y   W   F   T   Y   R   Q   A   K   E   L   G   G   Q   V   R   K   G
gag aaa ggc tca acc gtt gtt aaa ttc ggc acc atc gag cgc gag gac gag caa acc ggc
 E   K   G   S   T   V   V   K   F   G   T   I   E   R   E   D   E   Q   T   G
gaa gaa aag aaa att ccc tat ttg aag ggt tac acc gtt ttc aac gcc gac cag atc gac
 E   E   K   K   I   P   Y   L   K   G   Y   T   V   F   N   A   D   Q   I   D
ggc ttg ccc gag cag tac cac gcc gca ccg gca gaa gcc gcc cgc gat ctt ggc acc gcc
 G   L   P   E   Q   Y   H   A   A   P   A   E   A   A   R   D   L   G   T   A
gcc gat ccc gag ctt gac gcc ttt ttt gcc gcg acc ggc gca gac att cgc acc agc agc
 A   D   P   E   L   D   A   F   F   A   A   T   G   A   D   I   R   T   S   S
gaa ccc cgc gcc tac tac aac ccg acc ggc gac tat atc cac atg ccg ccg att gcg acc
 E   P   R   A   Y   Y   N   P   T   G   D   Y   I   H   M   P   P   I   A   T
ttt cac agc gcc gca ggc tat tac gcc acg ctg gcc cat gag gcg acc cac tgg aca ggc
 F   H   S   A   A   G   Y   Y   A   T   L   A   H   E   A   T   H   W   T   G
cac aaa tcg cgc ctt gat cgt ttt agc cga ttc agc gac cgc aag agc tac gct ttc gag
 H   K   S   R   L   D   R   F   S   R   F   S   D   R   K   S   Y   A   F   E
gaa cta att gca gaa atc ggc aat tgc atg ctt tgc gca agc ctt ggc ttg ata ccc gat
 E   L   I   A   E   I   G   N   C   M   L   C   A   S   L   G   L   I   P   D
ttt gac caa tcc gcc gca tac gtt caa agc tgg ctg cgt gct ttg aag gac gac aag cgg
 F   D   Q   S   A   A   Y   V   Q   S   W   L   R   A   L   K   D   D   K   R
ctg att ttc aaa gcc gcc acc gag gca cag aaa gcc gcc gac ctc ctg caa gag aac gcg
 L   I   F   K   A   A   T   E   A   Q   K   A   A   D   L   L   Q   E   N   A
gcc aac ttc caa agg aaa gaa gcc gca taa cct gcc ctc ccc cgc tcc ggc ggg gga ttt
 A   N   F   Q   R   K   E   A   A   -
ttt ttg tgc cat cag tta caa aat aaa gtg ttg aca gct att gac aaa acc ccg gtt ttg

cat cat aat tgc atc atg gtt gca tca tta aac
```

***Figure 20.*** *Nucleotide and amino acid sequence for ardC gene. Alternative initiation codons of ArdC are underlined with the most probable open reading frame highlighted in pink. -35 box is shown underlined in green, -10 box is underlined in orange and RBS underlined in purple. Terminator sequence is also underlined, with the inverted repeats in blue. Flanking 69 bp repeats are highlighted in light grey with the HincII sites highlighted in dark grey.*

In addition, *ardC* in R388 is also flanked by two repeats, as in plasmid pSA, that could confer instability due to homologous recombination events. Two copies of a 79 bp region containing the promoter and a HincII restriction site (85 % consensus positions) flank *ardC* in R388 (Figure 20 and Figure 21).

```
Start | 1| AGCAGTGAAAAAA--AAGTGTTGACAACCTATTGACAACCGCCCTGTTTTGCATCATAATAGCATCATGGTTGCATCACT
Term  | 1| ATCAGTTACAAAATAAAGTGTTGACAGCTATTGACAAAACCCCGGTTTTGCATCATAATTGCATCATGGTTGCATCATT
```

***Figure 21.*** *Alignment of the two DNA repeats flanking ardC gene. The nucleotides present in both repeats are shown in red highlighted in yellow. HincII sequence is framed.*

To be noted, there are only seven other *ardC* genes that share these initial 63 nucleotides in common with R388 *ardC_I* (100% identity) that could also start at the GUG alternative initial codon (Figure 22). This conserved initial region only appears in IncW plasmids: P2055-IMP, pMTY10660, pMAK3, pSA, R7K, pIE321 and pHH2-227 (IncW/IncP-6 hybrid plasmid) indicating a probable recent origin in common.

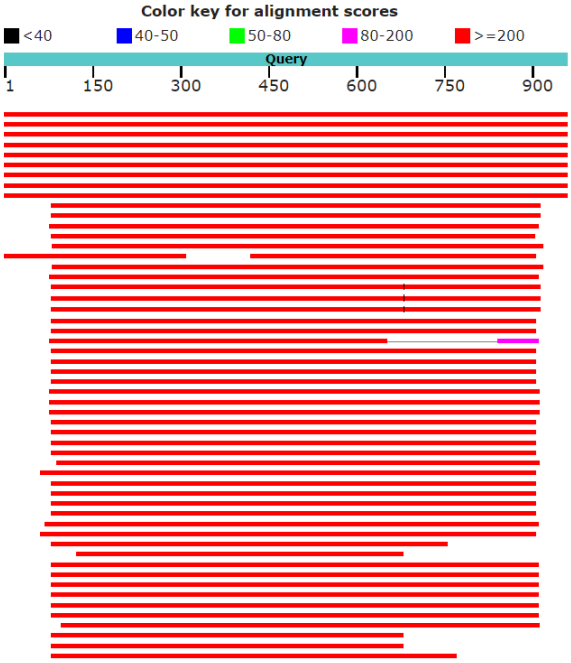



***Figure 22.*** *BLAST nucleotide alignment output for R388 ardC_I sequence.*

## 4.1.2. Analysis of the two putative protein isoforms

Returning to the question posed in Section 4.1.1 that was left open about which was the true starting codon, we wanted to check it now at a protein level. Sometimes, a gene shows the ability to produce several protein isoforms. Usually they perform the same or similar biological functions but sometimes one isoform can have an enhanced function.

In order to make sure which is the effective starting codon, both *ardC* and *ardC_I* coding sequences were cloned in pet29C under the strong T7 promoter. Expected proteins of 297 amino acids and 318 amino acids respectively, both with an LEHHHHHH tail, were purified as shown in Figure 23 and **Figure 24**.

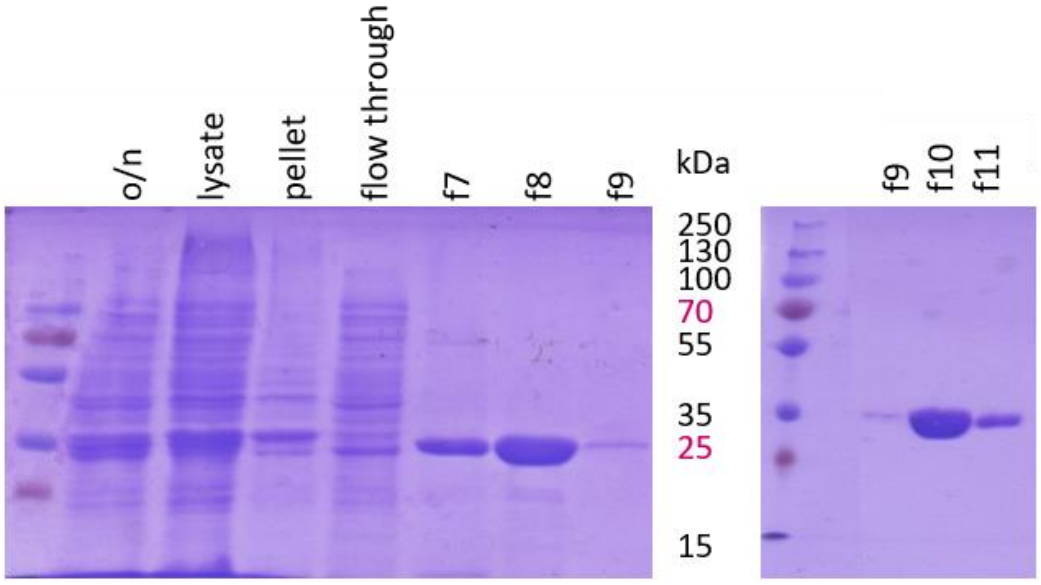



***Figure 23.*** *ArdC protein overexpression and purification. SDS-Page gel showing the overexpression of ArdC, lysis and purification fractions from two different purification steps through HisTrap (left) and HiTrap Heparin (right) HP columns. Page Ruler protein ladder sizes are indicated in between.*

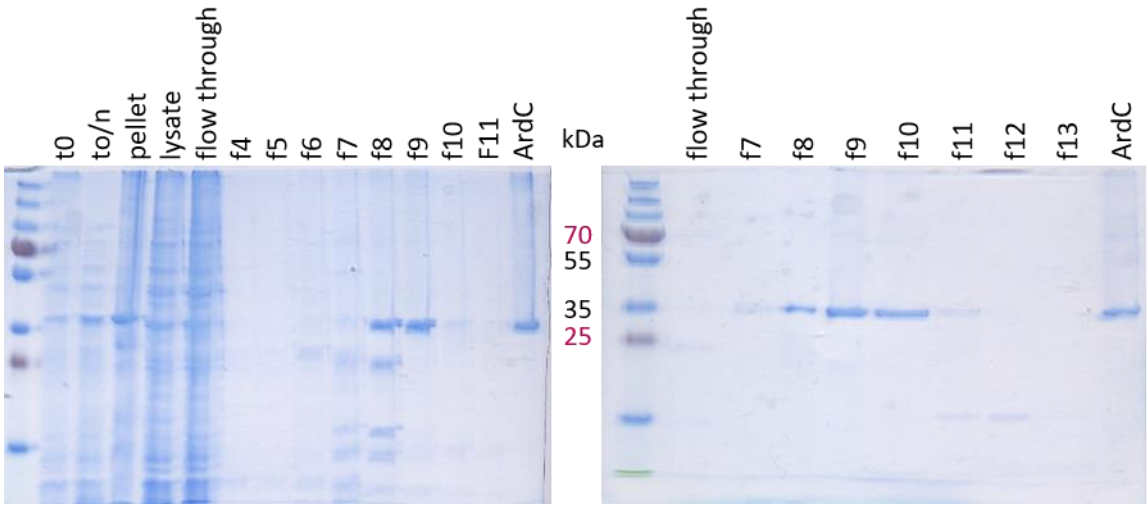



***Figure 24.*** *ArdC_l protein overexpression and purification. SDS-Page gel showing the overexpression or ArdC_l, lysis and purification fractions from two different purification steps through HisTrap (left) and HiTrap Heparin (right) HP columns. Pure ArdC (f11) after HiTrap Heparin is loaded as a size control. Some PageRuler protein ladder sizes are indicated in between.*

As the molecular weight of both isoforms is similar (ArdC_l theoretical MW of 36281.7 Da and ArdC theoretical MW of 34267.5 Da) and thus, difficult to separate by SDS-PAGE, we decided to analyze the sizes of the two purified proteins by mass spectrometry (UPV service, Bilbao). Interestingly, both samples showed to have the same MW of 34137 (Figure 25), corresponding to the size of the shortest version without the initial methionine (theoretical MW= 34136.3 Da, +0.7 Da error). With this experiment, we confirm that the produced ArdC protein is the short version, lacking the initial 21 amino acids predicted.

None of the experiments performed can make us think that it is the annotation of NCBI the real coding sequence. What is more, NCBI notifies that the reference sequence has not yet been subjected to final NCBI review. However, to make sure, a STOP codon could be introduced between both codons (with special care of RBS) to confirm our results and discard a longer isoform.

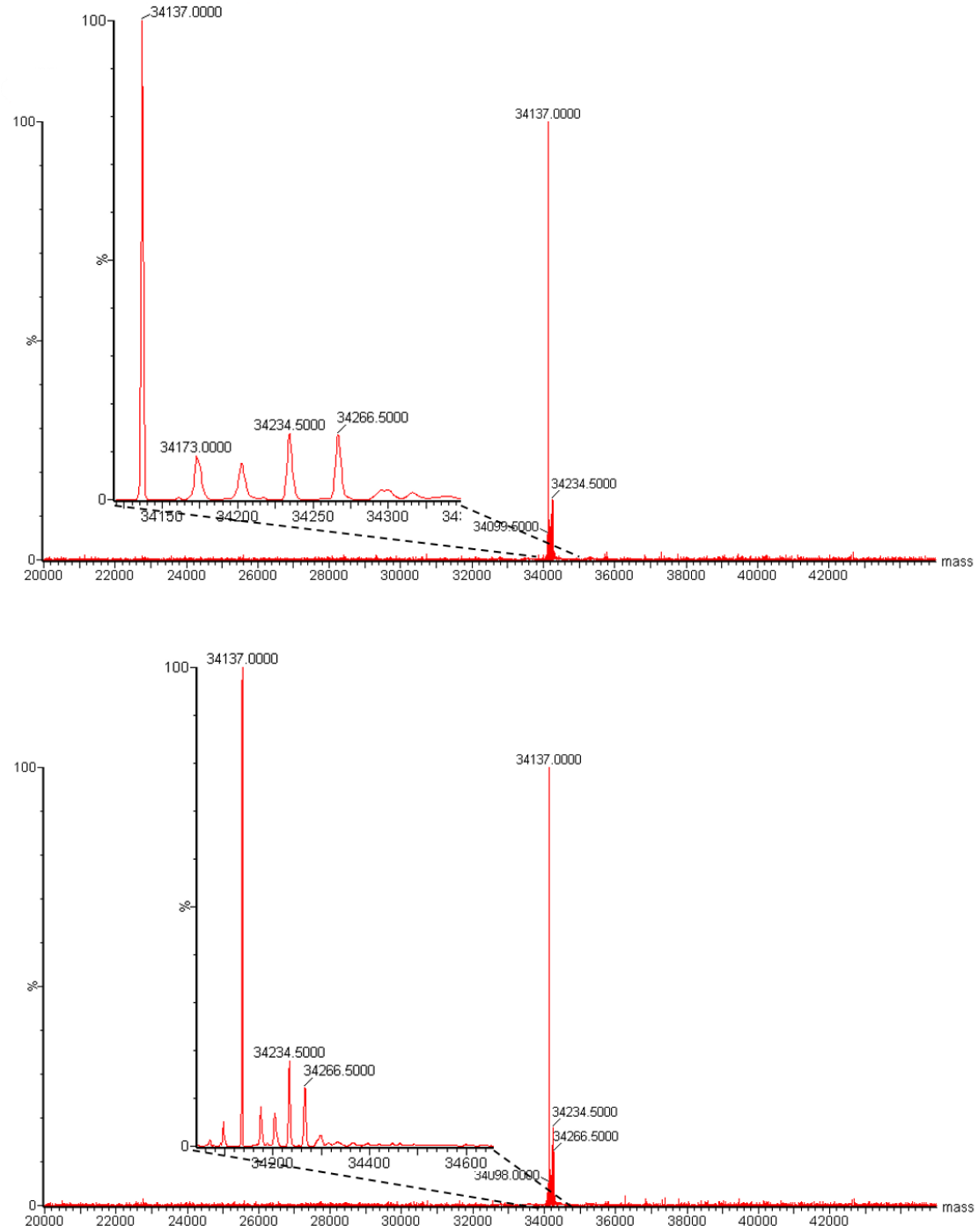



***Figure 25.*** *Molecular Weight determination by ESI-MS. Proteins were analysed between 20000 Da and 45000 Da and a closer range view of size peaks is also shown for A) ArdC protein and B) ArdC_I protein.*

### 4.1.3 ArdC diverse genetic environment and widely distribution

According to Pfam database (a large collection of protein families), the first part of ArdC protein is a Domain of unknown function (DUF1738). Only in 8 % of the sequences where this DUF1738 has been found, it is present alone in the polypeptide chain. On the other hand, in 85 % of the sequences it is found together with a Metallopeptidase superfamily domain (PF18818). Only in 4 % of the sequences, DUF1738 is found together with PF18818 and other extra domains.

If we do a study of the whole protein in an attempt to discover the precise function of ArdC, we did an analysis of the genetic environment of *ardC* genes with the help of Artemis and KEGG

(Kyoto Encyclopedia of Genes and Genomes). We found that *ardC* is usually surrounded by genes that codify for proteins which functions are: confer resistance to metals, arsenic, and other toxic compounds, DNA repair, integrases, exo- and endo- nucleases, toxin and antitoxin systems, transcriptional regulators, and proteins involved in conjugation and chromosome partition. Thus, *ardC* can be found in a broad range of genetic environments and not categorized to a particular stage.

In order to have a global view of where is ArdC found, we performed a BlastP taxonomy analysis with 5000 sequences using Blosum 62 matrix (Table 13). The sequences belonged to 3110 different organisms, all of them belonging to cellular organisms except for 10 viruses; the *Caulobacter crescentus* (α-Proteobacterium) bacteriophages (Gill et al., 2012) with about 48 % identity and a 94 % query cover and other 6 unclassified Phicbkvirus.

There is 1 homologue in the Archaea domain, 6 in Eukaryota domain and the rest are found in Bacteria. Exemplifying the Eukaryotic members, there is a protein homologue in *Thecamonas trahens*, a Zooflagellate, which is defined as a single-stranded DNA-specific exonuclease. *T. trahens* ArdC homologue contains a ssDNA binding domain and a protease domain like R388 ArdC (with a global identity of 47 % and a query cover of 93 % with R388 ArdC), but in addition, it is fused to an extra ssDNA exonuclease domain. On the other hand, *Ostreococcus tauri* is a Viridiplantae unicellular alga that also possess an ArdC homologue with an identity of 44 % and a query cover of 93 % at a protein level, without extra domains.

If we focus our attention in the bacterial homologues, 96 % of the organisms are classified as Proteobacteria but there are also a minority of Terrabacteria, Planctobacteria (PVC group), Sphingobacteria (FCB group), Acidobacteria, Nitrospira, Elusimicrobia and other unclassified and uncultured bacteria. Within Proteobacteria, most of them are present in α-Proteobacteria (61 %; half of them being Rhizobiales) and γ-Proteobacteria (37 %; mainly Enterobacterales as Salmonella, Escherichia or Klebsiella) with a few homologues in β-Proteobacteria (2 %; mainly Burkholderiales). There are also a minority of homologs in δ- and ε-Proteobacteria in addition to other Oligoflexia, Acidithiobacillales, Mariprofundus and other unclassified Proteobacteria members (Table 13).

Plasmids containing ArdC homologs are present in α-Proteobacteria as pTiBo5 plasmid "AAZ50566.1" and pAtF4 plasmid "KJX90203.1" from *A. tumefaciens,* in β-Proteobacteria as byi_1p plasmid "AET95037.1" from *Burkholderia sp. YI23*, in γ-Proteobacteria as pPHDP60 plasmid "AGE91731.1" found in *Photobacterium damselae* or even in Planctomycetes as PALBO2 plasmid "APW64319.1" from *Paludisphaera borealis*.

There are no ArdC homologues neither in *E. coli* K-12 nor in *P. putida KT2440, or Agrobacterium tumefaciens C58*. However, there are homologs in other *P. putida* strains, codified in the chromosome as in *P. putida H8234* (37 % identity and 93 % query cover) or in plasmids as in pND6-2 isolated from *P. putida ND6* (39 % identity and 92 % query cover). The ArdC protein with a sequence more similar to the R388 ArdC in *Pseudomonas Spp.* has been found in *Pseudomonas Stutzeris* with a 49 % identity and a 94 % query cover.

***Table 13.*** *ArdC BlastP taxonomic analysis. Table was obtained by searching in BlastP for 5000 homologue sequences from the non-redundant protein database (nr) and Blosum 62 matrix.*

| Taxonomy | Number of Organisms |
|---|---|
| ⊟root | 3110 |
| . ⊟cellular organisms | 3099 |
| . . ⊟Bacteria | 3092 |
| . . . ⊟Proteobacteria | 2978 |
| . . . . ⊟Gammaproteobacteria | 1089 |
| . . . . . ⊞Enterobacterales | 862 |
| . . . . . ⊞Alteromonadales | 19 |
| . . . . . ⊞Sinobacteraceae | 2 |
| . . . . . ⊞Cellvibrionales | 4 |
| . . . . . ⊞unclassified Gammaproteobacteria | 5 |
| . . . . . ⊞Oceanospirillales | 12 |
| . . . . . ⊞Pseudomonadales | 4 |
| . . . . . ⊞Vibrionaceae | 66 |
| . . . . . ⊞Chromatiales | 15 |
| . . . . . ⊞Methylococcales | 3 |
| . . . . . ⊞Pasteurellales | 93 |
| . . . . . Thiothrix lacustris | 1 |
| . . . . . ⊞Dyella | 2 |
| . . . . ⊟Alphaproteobacteria | 1814 |
| . . . . . ⊞Rhizobiales | 914 |
| . . . . . ⊞Rhodobacterales | 316 |
| . . . . . ⊞Sphingomonadales | 385 |
| . . . . . ⊞unclassified Alphaproteobacteria | 27 |
| . . . . . ⊞Rhodospirillales | 99 |
| . . . . . ⊞Rhodothalassium | 2 |
| . . . . . Sneathiella sp. | 1 |
| . . . . . ⊞Caulobacterales | 62 |
| . . . . . ⊞Parvularculaceae | 6 |
| . . . . . Minwuia thermotolerans | 1 |

| Taxonomy | Number of Organisms |
|---|---|
| . . . . ⊟Betaproteobacteria | 61 |
| . . . . . ⊞Burkholderiales | 46 |
| . . . . . ⊞Nitrosomonadales | 12 |
| . . . . . Candidatus Accumulibacter sp. SK-01 | 1 |
| . . . . . ⊞Thauera | 2 |
| . . . . ⊞unclassified Proteobacteria (miscellaneous) | 4 |
| . . . . Bdellovibrio sp. 28-41-41 | 1 |
| . . . . ⊞delta/epsilon subdivisions | 4 |
| . . . . ⊞Acidithiobacillales | 2 |
| . . . . ⊞Mariprofundus | 2 |
| . . . uncultured bacterium | 1 |
| . . . ⊞Terrabacteria group | 14 |
| . . . ⊞PVC group | 21 |
| . . . ⊞FCB group | 52 |
| . . . ⊞Acidobacteria | 10 |
| . . . ⊞unclassified Bacteria | 13 |
| . . . ⊞Nitrospira | 2 |
| . . . Elusimicrobia bacterium RBG_16_66_12 | 1 |
| . . ⊟Eukaryota | 6 |
| . . . ⊞Bilateria | 3 |
| . . . ⊞Viridiplantae | 2 |
| . . . Thecamonas trahens ATCC 50062 | 1 |
| . . Thermoplasmata archaeon | 1 |
| . IncW plasmid pSa | 1 |
| . ⊟Phicbkvirus | 10 |
| . . ⊞unclassified Phicbkvirus | 6 |
| . . Caulobacter virus Karma | 1 |
| . . Caulobacter virus Magneto | 1 |
| . . Caulobacter virus Swift | 1 |
| . . Caulobacter virus phiCbK | 1 |

We can conclude that ArdC is a widely distributed protein that may have jump from plasmids to the chromosomes of diverse species to reach such characteristic of broad distribution. In addition, it is important to mention, as described by (Iyer et al., 2017), that the two domains of ArdC (ssDNA-binding and metalloprotease) are the two most common ones in polyvalent proteins found in bacteriophages and conjugative elements, being most of the times joined together. What is more, it is very common to find them together forming part of longer proteins with additional domains to perform more complex functions.

## 4.2 ArdC role in DNA transfer by conjugation

### 4.2.1 ArdC effect in Conjugation

#### 4.2.1.1 ArdC is needed to conjugate R388 from E. coli to different species but not the other way around

According to previous interspecies conjugation results of pSU2007 and pIC10 (a R388 derivative without the stability and maintenance gene region) (Figure 9) (del Campo, 2016), pIC10 is not efficiently transferred from *E. coli* to *P. putida* or *A. tumefaciens* as pSU2007. In the stability and maintenance gene region deleted in pIC10 there are 13 genes, being *ardC* one of them. To determine the role of ArdC in the transfer of R388-derived plasmids to different bacteria, we have constructed the plasmid pLGM25 (a R388 derivative without *ardC* gene) (see Section 3.2.2.7). This plasmid was introduced into *E. coli BW27783-Nx*$^R$ strain and then conjugated to *E. coli BW27783-Rf*$^R$, *P. putida KT2440* or *A. timefaciens GMI9023* strains.

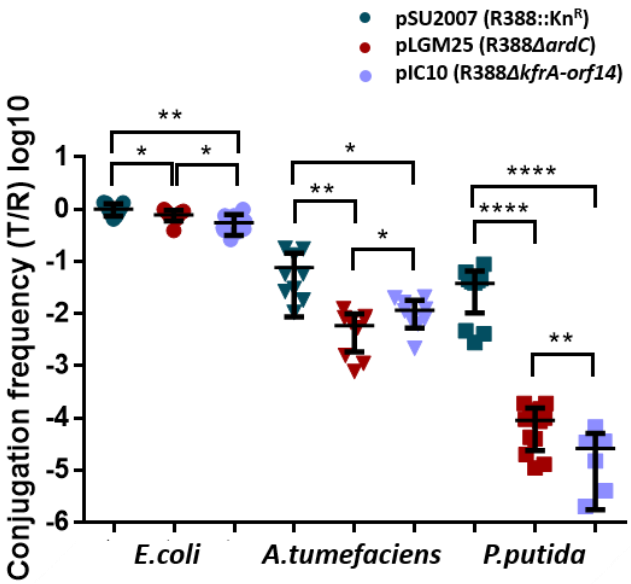



***Figure 26.*** *Effect of ardC and kfrA-orf14 region on plasmid conjugative transfer from E. coli to different bacteria. The conjugation frequencies per recipient (T/R) from E. coli BW27783-Nx into E. coli BW27783-Rif (1h 37 ºC), into A. tumefaciens GMI9023 (1h 30 ºC) and into P. putida KT2440 (1h 37 ºC) are shown. Horizontal and vertical bars represent mean ± SD obtained for each dataset of n=6-12 (t-test: * p < 0.1, ** p < 0.01, *** p < 0.001, **** p <0.0001).*

Conjugations in 1 h at 30 °C from *E. coli* to *P. putida* gave conjugation frequencies close to our detection limit and with bigger variances, thus, we decided to do the experiments in 1 h but at 37 °C as done by (Gemperlein et al., 2016). We observed that the absence of *ardC* in the conjugative plasmid pLGM25 reduced substantially the conjugation ability of the plasmid from *E. coli* to *A. tumefaciens* and even more notably to *P. putida*, but not from *E. coli* to *E. coli* (Figure 26). The conjugation frequency was also reduced in pIC10 when transferred from *E. coli* to *A. tumefacens* or *P. putida*, as already observed. Thus, the results previously described for pIC10 could be explained to a large extent by the *ardC* absence.

Interestingly, pSU2007 transconjugant colonies were smaller than the pLGM25 or pIC10 transconjugant colonies.

To discard that the differences in the conjugation frequencies are due to the temperature, we conjugated at 30 °C and 37 °C from *E. coli* grown at 37 °C to *E. coli* grown at 30 °C. Same transfer frequencies than when growth and conjugation were performed at 37 °C were obtained.

***Table 14.*** *Influence of temperature in conjugation frequency per recipients (T/R) from E. coli grown at 37 °C to E. coli grown at 30 °C in 1h of conjugation at 30 °C or 37 °C. The mean and SD of 9 replicas is shown (1way ANOVA respect pSU2007 experiment at each temperature).*

| f (T/R) | 30 °C | 37 °C |
|---|---|---|
| **pSU2007** | 2.4E-01 ± 1.0E-01 | 7.1E-01 ± 3.5E-01 |
| **pLGM25** | 8.9E-01 ± 4.3E-02 (n.s.) | 7.3E-01 ± 4.3E-01 (n.s.) |
| **pIC10** | 7.1E-01 ± 4.5E-02 (n.s.) | 6.3E-01 ± 7.2E-01 (n.s.) |

Based on the differences observed when conjugating pSU2007 or pLGM25 from *E. coli* to *P. putida* we tried to test if the same differences were observed using *P. putida* containing pSU2007, pLGM25 or pIC10 as donor strain and *E. coli* as recipient strain. However, differences in conjugation frequency were not observed when conjugating pSU2007 or pLGM25 at 30 °C nor at 37 °C (Table 15). Small differences, not attributable to *ardC* were observed for pIC10. Thus, *ardC* is not needed when using *E. coli* as recipient strain regardless the donor strain. The size of the pSU2007 *E. coli* transconjugant colonies was again smaller than the regular size of *E. coli* colonies.

***Table 15****. Influence of ArdC in the conjugation frequency per recipients (T/R) from P. putida to E. coli when conjugating 1h at 30 °C or 37°C. The mean and SD of 8-15 replicas is shown (1way ANOVA respect pSU2007 experiment at each temperature).*

| f (T/R) | 30 °C | 37 °C |
|---|---|---|
| **pSU2007** | 8.6E-03 ± 1.9E-03 | 3.5E-02 ± 1.6E-02 |
| **pLGM25** | 6.5E-03 ± 3.6E-03 (n.s.) | 9.7E-02 ± 5.2E-02 (n.s.) |
| **pIC10** | 7.5E-03 ± 9.0E-04 (n.s.) | 2.4E-01 ± 8.3E-02 (*) |

### 4.2.1.2 ArdC acts in recipient cells

Based on the differences found in Figure 26, we tried to complement the conjugations from *E. coli* to *P. putida* by overexpressing ArdC or ArdC_l cloned in pUCP22 in donor *E. coli BW27783-Nx*$^{R}$ carrying the conjugative plasmid pLGM25 and in recipient *P. putida KT2440* cells. In Figure 27, we observe how neither *ardC* nor *ardC_l* improve the conjugation frequency when overexpressing them in the donor strain after 1h of conjugation at 37 °C with 0.1 mM IPTG. Overexpression of *ardC and ardC_l* in recipient cells involves a significant increase in conjugation frequency, implying the recovery up to pSU2007 conjugation frequency levels. This means that ArdC is acting in recipient cells. This activity seems to be due to the expression of ArdC in the recipient cells and not by the expression in donor cells. Thus, ArdC is not able to travel to recipients through the conjugation pore to act there as proposed by (Belogurov et al., 2000).

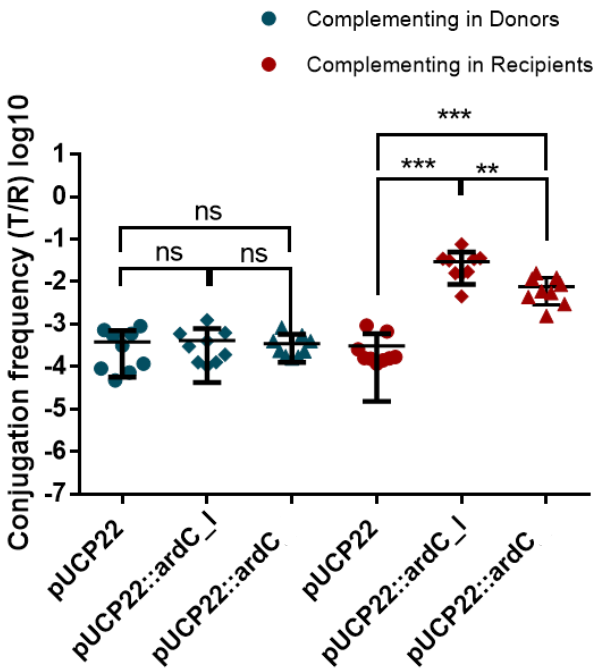



***Figure 27.*** *Effect in the conjugation frequency when complementing pLGM25 with ardC in donors or recipients. Complementation in donors (shown in teal) or in recipients (shown in maroon) with pUCP22, pUCP22::ardC_l or pUCP22::ardC is shown. Conjugation was done for 1h at 37 °C with 0.1 mM IPTG added to the mating mixture. Horizontal and vertical bars represent the mean ± SD obtained for each dataset of n=9 (t-test: ** p < 0.01, *** p < 0.001).*

### 4.2.1.3 ArdC_l shows more activity than the shorter isoform

Returning to the question previously posed in Section 4.1.1 about the real initiation codon of ArdC, we checked if *ardC_l* cloned with the 63 additional nucleotides was equally good at complementing in recipient cells (Figure 27). We showed that *ardC* is able to complement pLGM25 plasmid in recipients but not to the same stent as *ardC_l* does (rising the pSU2007 conjugation frequency levels). This means that there are significant differences in expression. Maybe this region could be important for the stability of the mRNA.

## 4.2.2 ArdC contributes to plasmid instability

Due to the observation that pSU2007_*P. putida* transconjugant colonies are smaller than pIC10 or pLGM25 transconjugants, we decided to test if ArdC is implicated in the fitness of the cell, checking the generation time (**Figure 28**) and the stability of the plasmid (**Figure 29**) in comparison with the transconjugant cell colonies bearing pIC10 or pLGM25.

The generation time (g) is the time taken by bacteria to double in number. We have observed that g is reduced when *ardC* is not present in the plasmid (pLGM25 and pIC10) respect the growth rate of the plasmid-free *P. putida KT2440* or carrying pSU2007 plasmid with *ardC* (**Figure 28**). This agrees with the fastest growth or bigger sizes of the pLGM25 and pIC10 transconjugant

colonies after conjugation. However, the smaller size of pSU2007 transconjugants respect the other two types of transconjugant colonies cannot be explained by the 7-8 minutes of difference in the generation time. Thus, we decided to test the stability of the plasmids. The loss of pSU2007, pLGM25 or pIC10 plasmids from *P. putida KT2440* was checked by plating the strains in plates with antibiotic selecting for the strain and plates with the additional antibiotic selecting for the plasmid after growing the cultures o/n just with the antibiotic that selects the strain as described in Section 3.2.1.5.

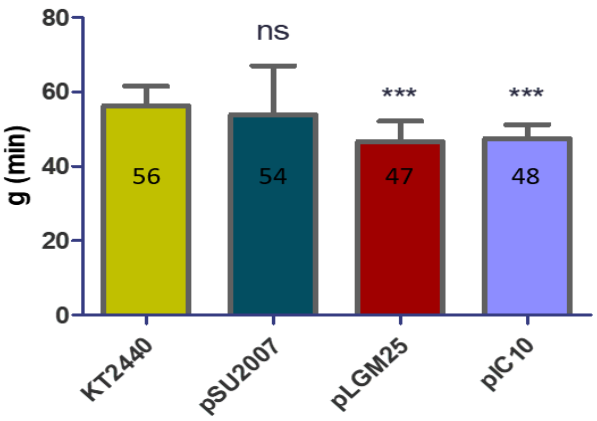



***Figure 28.*** *Generation times of P. putida KT2440 empty or carrying pSU2007, pLGM25 or pIC10 in LB media. Horizontal and vertical bars represent the mean ± SD of n=30 observations, (1-way ANOVA: *** p < 0.001).*

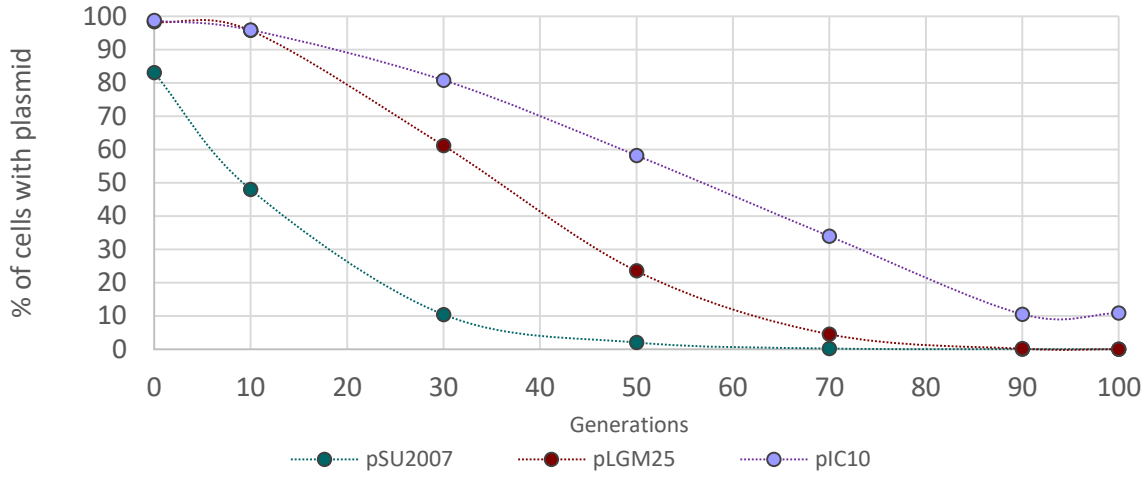



***Figure 29.*** *Stability of plasmids pSU2007, pLGM25 and pIC10 in P. putida KT2440. The datum points are averages of the results from 5 to 6 independent experiments.*

We observed that after 10 generations, around 50 % of the pSU2007 transconjugants had already lost the plasmid. Meanwhile, about 35 and 55 generations had to happen to reach such plasmid lost levels for pLGM25 and pIC10 *P. putida* transconjugants respectively (**Figure 29**). These results are in agreement with the results of the generation times shown in **Figure 28**, and with the smaller size of pSU2007 transconjugant colonies. As expected, *ardC* seems to be involved in plasmid instability in *P. putida*. Plasmid loss may not simply be a matter of pSU2007 bigger plasmid size, as it may seem when compared with the much higher stability of pIC10.

### 4.2.3 Modifications in the gene expression profile

The aim of this experiment was to try to find a transcriptional regulator that could be the putative target of ArdC and to understand the role of *ardC* in interspecies conjugation given the observances described in Section 4.2.1.1. Thus, our goal in this experiment was to analyze the

changes in gene expression when an a*rdC*-containing plasmid is introduced in *P. putida* by conjugation. Gene expression was analyzed by RNA-seq as described in materials and methods.

To analyze what is happening in the recipient cells upon conjugation, we have to be able to differentiate RNA from the recipient cell and from the donor cells. This is something we cannot do in *E. coli* to *E. coli* conjugation, but it could be feasible in interspecies conjugation assays, although we had to titer the donor to recipient ration. A high donor (*E. coli*) to recipient (*P. putida*) ratio ensures that all the recipient cells are surrounded by donors, and thus, accessible to receive the plasmid, but then the recipient RNA was going to be very diluted. On the other side, if only a small amount of recipient cells receive the plasmid we are not going to be able to observe the expression changes. Thus, first of all we have titrated the conjugation assay with different donor to recipient ratios to find the minimal ratio with a significant conjugation frequency (Table 16). A ratio of approximately 5:1 gave a conjugation frequency of about 0.1 for pSU2007 meaning that, in theory, we would have around 5 times more *E. coli* RNA than *P. putida* RNA and that about 10 % of the *P. putida* cells will show the changes associated to the plasmid acquisition.

***Table 16**. Conjugation titration for RNA-seq assay. Conjugation was performed from E. coli BW27783-Nx bearing plasmid or not towards P. putida KT2440 recipients for 37 °C and 30 min.*

| Donor strain | Donors (D) | Recipients (R) | Transconjugants (T) | D : R | Freq (T/D) | Freq (T/R) |
|---|---|---|---|---|---|---|
| ***BW*** | 1.3E+06 | 2.0E+05 | 0.0E+00 | 6.50 | 0.0E+00 | 0.0E+00 |
| | 8.0E+05 | 2.0E+05 | 0.0E+00 | 4.00 | 0.0E+00 | 0.0E+00 |
| | 1.2E+07 | 2.8E+05 | 0.0E+00 | 42.8 | 0.0E+00 | 0.0E+00 |
| ***BW*+ pSU2007** | 1.4E+06 | 2.5E+05 | 2.9E+04 | 5.60 | 2.0E-02 | 1.2E-01 |
| | 1.4E+06 | 3.0E+05 | 2.8E+03 | 4.67 | 2.0E-03 | 9.3E-03 |
| | 7.0E+06 | 2.0E+05 | 2.0E+05 | 35.0 | 2.8E-02 | 1.0E+00 |
| ***BW*+ pLGM25** | 1.7E+06 | 2.2E+05 | 1.0E+01 | 7.73 | 5.9E-06 | 4.6E-05 |
| | 1.2E+06 | 4.0E+05 | 1.0E+01 | 3.00 | 8.3E-06 | 2.5E-05 |
| | 1.5E+07 | 3.1E+05 | 4.0E+01 | 48.4 | 2.7E-06 | 1.3E-04 |

We first designed three different “conjugation” assays, Figure 30. Conjugation was performed from *E. coli BW27783* bearing no plasmid (#1, no conjugation), pSU2007 (#2) or pLGM25 (#3) to P. putida KT2440 for 30 min at 37 °C.

The results of the conjugations are shown in Table 17 where we can observe that a ratio of around 5 donor cells per recipient showed a notable conjugation frequency difference between having *ardC* (#2) and lacking *ardC* (#3) in the conjugative plasmid. Experiment #1 is the negative control, and thus, no conjugation is observed due to the lack of a conjugative plasmid.

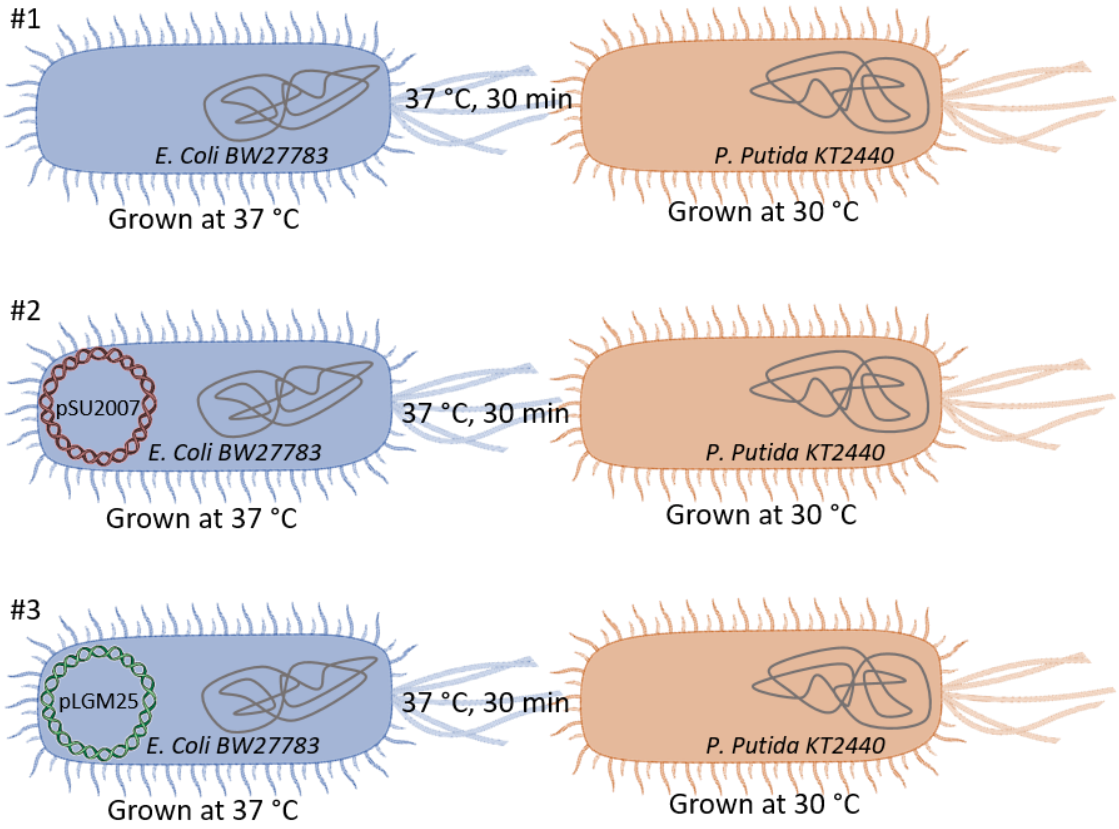
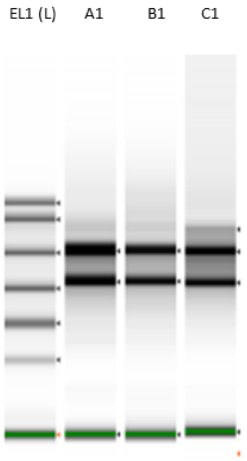



***Figure 30.*** *Conjugation experiments designed for RNA-Seq analysis. Conjugation was done in filter plates from E. coli BW27783 bearing no plasmid (#1, no conjugation), pSU2007 (#2) or pLGM25 (#3) to P. putida KT2440 for 30 min at 37 °C.*

***Table 17.*** *Conjugation results for experiments #1, #2 and #3 further analyzed by RNA-seq. The experiments are shown in figure 30, where donor cells are E. coli BW27783 ($Nx^R$) containing no plasmid, pSU2007 or pLGM25 and the recipient cells are P. putida KT2440. The day after the conjugation for 30 min at 37 °C, colonies from selecting plates for Donors (Don.), Recipients (Rec.) and Transconjugants (Trans.) were counted. The conjugation frequencies per donor (Freq. T/D) and per recipient (Freq. T/R) are also shown.*

| Exp. | Donor strain | Recipient strain | Donors | Recipients | Transc. | D : R | Freq. (T/D) | Freq. (T/R) |
|---|---|---|---|---|---|---|---|---|
| (#1) | *BW27783* | *KT2440* | 1.3E+06 | 2.0E+05 | 0.0E+00 | 6.5 | 0.0E+00 | 0.0E+00 |
| (#2) | pSU2007 in *BW27783* | *KT2440* | 1.4E+06 | 2.5E+05 | 2.9E+04 | 5.6 | 2.1E-02 | 1.2E-01 |
| (#3) | pLGM25 in *BW27783* | *KT2440* | 1.2E+06 | 4.0E+05 | 1.0E+01 | 3.0 | 8.3E-06 | 2.5E-05 |
| | Antibiotics for selection | | KnNx | ApCm | KnCm | | | |

Total RNA was extracted and evaluated (Figure 31) assuring an $RIN^e$ value above 8 before sending it for analysis as it is the threshold recommended to assure a good enough quality for transcriptome analysis. The $RIN^e$ is an algorithm based on an electropherogram for assigning integrity values from 1 to 10 to RNA measurements being 10 the least degraded. In Table 18 we can see the sequencing statistics. The current study was limited by the conjugation frequency, as not all the cells are conjugating or are being conjugate (Table 17). This fact implies that we needed a high sequencing coverage as there would be an important amount of noise coming from non-conjugating cells, and thus, the observance of lower expression increment or decrement values.




| Well | RIN$^e$ | 23S/16S (Area) | Conc. [ng/uL] | Sample Description |
|---|---|---|---|---|
| EL1 | - | - | 97.5 | Electronic Ladder |
| A1 | 8.9 | 1.3 | 196 | BW (#1) |
| B1 | 8.6 | 1.1 | 136 | pSU2007 (#2) |
| C1 | 8.8 | 1.5 | 131 | pLGM25 (#3) |

***Figure 31.*** *Total RNA integrity quantity and quality extracted from experiments #1, #2 and #3 validated by Agilent RNA ScreenTape assay. The RNA integrity number equivalent (RIN$^e$) was assured to be above 8 for the three samples.*

***Table 18****. TruSeq® Stranded mRNA (Strand specific) Illumina sequencing results. The total number of bases, total reads (sum of read 1 and read 2), GC (%) and AT (%) content, Q20 (%) and Q30 (%) (phred quality score over 20 or 30 respectively) calculated for the three samples.*

| Sample ID | Total read bases (bp) | Total reads | GC (%) | AT (%) | Q20 (%) | Q30 (%) |
|---|---|---|---|---|---|---|
| BW (#1) | 22,467,804,106 | 222,453,506 | 54.59 | 45.41 | 97.96 | 94.54 |
| pSU2007 (#2) | 19,110,456,642 | 189,212,442 | 54.55 | 45.45 | 97.95 | 94.45 |
| pLGM25 (#3) | 23,255,871,352 | 230,256,152 | 56.05 | 43.95 | 97.9 | 94.33 |

Coverage (C) is the number of times a genome has been sequenced (the depth of sequencing) and is calculated as C=LN/G where L is the length of the reads (bp), N stands for the number of reads and G indicates the genome size (bp). Given the following genome sizes:

- *Escherichia coli* plasmid R388 (NC_028464.1) → 33,913 bp
- *Escherichia coli K12 MG1655* (NC_000913.3) → 4,641,652 bp
- *Pseudomonas putida KT2440* (NC_002947.4) → 6,181,873 bp

We could calculate the coverage for each of the three sequencing experiment as shown in Table 19 giving the fact that, for each experiment, a combination of 2 or 3 genomes is present.

***Table 19.*** *Coverage calculation for each sequencing experiment.*

| Sample ID | Read size (bp) (L) | Total reads (N) | Comb. genome size (bp) (G) | Coverage (C=L*N/G) |
|---|---|---|---|---|
| BW | 100 | 222,453,506 | 10,823,525 | 2055x |
| pSU2007 | 100 | 189,212,442 | 10,857,438 | 1743x |
| pLGM25 | 100 | 230,256,152 | 10,857,438 | 2121x |

We run a FastQ program to check the quality of the reads received after sequencing, giving us good quality scores (data not shown) for all 6 files (the three samples sequenced from both sides) so we decided not to trimmer the reads. We have then aligned using Bowtie2 the reads

obtained in the three experiments to each of the three reference genomes: *E. coli*, *R388* and *P. putida*. As shown in Table 20, around 50 % of the RNA mapped to *E. coli* and 50 % to *P. putida*, meaning that RNA expression is higher in *P. putida* as we had 5 times more *E. coli* cells than *P. putida* being the genome size of *P. putida* only 1.33 times larger than the *E. coli* genome.

***Table 20.*** *Percentage of reads aligned to the three reference genomes by Bowtie2.*

| Sequenced sample | total reads | % mapped to *E. coli* | % mapped to *P. Putida* | % mapped to R388 |
|---|---|---|---|---|
| BW_1 | 111,226,753 | 52.81 | 46.70 | 0.00 |
| BW_2 | 111,226,753 | 52.74 | 45.92 | 0.00 |
| pSU2007_1 | 94,606,221 | 51.84 | 43.72 | 1.73 |
| pSU2007_2 | 94,606,221 | 51.43 | 43.18 | 1.84 |
| pLGM25_1 | 115,128,076 | 35.35 | 62.63 | 1.03 |
| pLGM25_2 | 115,128,076 | 35.40 | 61.76 | 1.03 |

With the RPKMs calculated by Artemis, we could plot the overall distribution of the differentially expressed genes of each of the genomes when *ardC* is present or absent in the plasmid (**Figure 32**). We have considered an increment or reduction of 2 fold as significative. When the *ardC*-containing plasmid (pSU2007) is transfer to the recipient *P. putida* strain most of the R388 genes are overexpressed respect the experiment where the plasmid without *ardC* (pLGM25) cannot be transferred (**Figure 32**A).

In the donor *E. coli,* we observe a higher number of genes upregulated (149) than downregulated (12) with pSU2007 respect with pLG25. However, *E. coli* in experiments #2 bearing pSU2007 and #3 with pLGM25 shows considerable more downregulated genes (52 and 211 respectively) than upregulated (18 and 27 respectively) in comparison with the empty cell. In the case of *P. putida,* more genes are upregulated (52) than downregulated (20). However, experiments #2 with pSU2007 and #3 bearing pLGM25 show more downregulated genes (63 and 51 respectively) than upregulated (41 and 22 respectively) respect the empty cell. In sections 4.2.3.1, 4.2.3.2 and 4.2.3.3 we analyze in more detail the expression changes for each of the genomes.

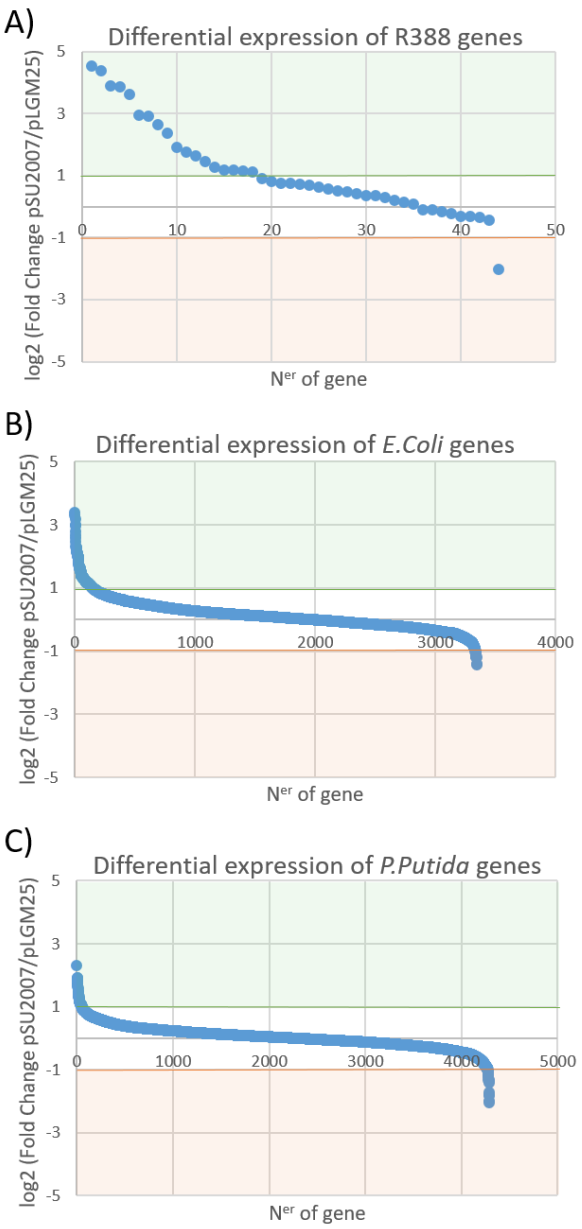



***Figure 32.*** *Differential expression of genes in experiment pSU2007 vs. pLGM25 to test influence of ardC in the three reference genomes. Upregulated genes with a RPKM fold change >2 are in the green zone, and downregulated genes with RPKM fold change <2 are in the salmon zone.*

### 4.2.3.1 Differential expression of R388 genes

The conjugation conditions studied here induced extensive transcriptional changes in R388 plasmid genes. We can observe in **Table 21** that when *ardC* is present (pSU2007, experiment #2), and thus, conjugation happens, genes involved in conjugation and pilus formation (shown in salmon as in **Figure 33**) are mainly overexpressed. On the other hand, those genes involved in replication as *repA,* as well as *dhfr* (that we expected not to change for being out of the control of a negative regulator), are barely altered with fold changes of around 1. There is no transcriptional overshoot in pLGM25, indicating that it is not conjugating well, as previously pointed in Table 11 and in accordance what it was described by (Fernandez-Lopez et al., 2014), that if no conjugation occurs, al the R388 promoters are strongly repressed.

***Table 21.*** *Expression level of R388 genes in RPKMs from experiment #2 (pSU2007) and #3 (pLGM25) and comparison of expression in RPKMs (pSU2007/pLGM25) for R388 plasmid genes. Genes are colored according to Figure 33 and list is ordered from highest to lowest according to the RPKMs (pSU2007/pLGM25) column and colored from greenish to reddish. Genes not present in pSU2007 or pLGM25 plasmid are not shown.*

| | Experiment | | Fold Change | Information | |
|---|---|---|---|---|---|
| Locus_tag | pSU2007 | pLGM25 | **pSU2007/ pLGM25** | Gene | Protein Function |
| R388_0003 | 23018 | 985 | 23.38 | *trwN* | Lysozyme-like domain |
| R388_0002 | 24093 | 1160 | 20.77 | *KikA* | Causes reversible growth inhibition |
| R388_0022 | 39522 | 2674 | 14.78 | *orf14* | Putative cold shock DNA-binding domain-like |
| R388_0024 | 40586 | 2782 | 14.59 | *orf12* | Putative type I R-M system methyltransferase subunit |
| R388_0023 | 33625 | 2762 | 12.17 | *klcB* | Derepresses KorA-regulated operons |
| R388_0046 | 238133 | 30898 | 7.71 | *orf45* | |
| R388_0047 | 77021 | 10165 | 7.58 | *orf46* | |
| R388_0004 | 86511 | 13806 | 6.27 | *korA* | Putative transcriptional repressor |
| R388_0005 | 87087 | 16717 | 5.21 | *trwL* | TrbC/VIRB2 family |
| R388_0006 | 61669 | 16422 | 3.76 | *trwM* | Type IV secretory pathway. VirB3-like protein |
| R388_0012 | 18435 | 5437 | 3.39 | *trwG* | VirB8 protein |
| R388_0018 | 8153 | 2604 | 3.13 | *trwA* | DNA binding/nic-cleavage accessory protein |
| R388_0011 | 28951 | 10600 | 2.73 | *trwH* | Prokaryotic membrane lipoprotein lipid attachment site |
| R388_0028 | 15085 | 6244 | 2.42 | *ssb* | |
| R388_0007 | 22440 | 9952 | 2.25 | *trwK* | Type IV secretion/conjugal transfer ATPase. VirB4 family |
| R388_0013 | 9409 | 4178 | 2.25 | *trwF* | VirB9/CagX/TrbG. a component of the type IV secretion system |
| R388_0027 | 65575 | 29438 | 2.23 | *ardK* | Putative transcriptional repressor |
| R388_0029 | 34340 | 15686 | 2.19 | *orf9* | |
| R388_0026 | 103954 | 55001 | 1.89 | *ldr2* | |
| R388_0017 | 9785 | 5591 | 1.75 | *trwB* | Type IV secretion-system coupling protein DNA-binding domain |
| R388_0020 | 86196 | 50701 | 1.70 | *stbaB* | |
| R388_0019 | 60846 | 35816 | 1.70 | *stbaA* | |
| R388_0009 | 18994 | 11420 | 1.66 | *eex* | Prokaryotic membrane lipoprotein lipid attachment site |
| R388_0008 | 17433 | 10810 | 1.61 | *trwJ* | VirB5 protein family |
| R388_0014 | 5651 | 3632 | 1.56 | *trwE* | Type IV secretion system protein VirB10 |
| R388_0015 | 5831 | 3961 | 1.47 | *trwD* | ATP hydrolase |
| R388_0040 | 34281 | 24217 | 1.42 | *intI1* | Integrase/recombinase |
| R388_0001 | 8003 | 5810 | 1.38 | *korB* | Domain in histone-like proteins of HNS family |
| R388_0021 | 62284 | 46346 | 1.34 | *stbC* | Putative ribonucleotide reductase-like. ferritin-like domain |
| R388_0016 | 8208 | 6411 | 1.28 | *trwC* | DNA helicase/Relaxase |
| R388_0010 | 7092 | 5576 | 1.27 | *trwI* | TrbL/VirB6 plasmid conjugal transfer protein |
| R388_0041 | 1365553 | 1118654 | 1.22 | *dhfr* | Dihydrofolate reductase. confers resistance to trimethoprim |
| R388_0032 | 15406 | 13371 | 1.15 | *ldr1* | |
| R388_0025 | 40748 | 37204 | 1.10 | *ldr2* | |
| R388_0039 | 1390 | 1303 | 1.07 | *tnpM* | |
| R388_0036 | 10110 | 10803 | 0.94 | *kfrA* | Plasmid replication region DNA-binding N-term |
| R388_0038 | 7958 | 8558 | 0.93 | *resP* | Putative resolvase |
| R388_0034 | 7687 | 8633 | 0.89 | *nuc2* | Putative signal transduction protein |
| R388_0035 | 7247 | 8560 | 0.85 | *nuc1* | DNAse |
| R388_0030 | 13688 | 16891 | 0.81 | *orf7-8* | |
| R388_0037 | 6114 | 7568 | 0.81 | *repA* | DNA replicase |
| R388_0033 | 4695 | 5957 | 0.79 | *osa* | Fertility inhibition factor |
| R388_0045 | 32661 | 44680 | 0.73 | *orf5* | Acetyltransferase |
| R388_0042 | 129423 | 531166 | 0.24 | *orfA* | Putative glutamine gamma-glutamyltransferase |

$Kn^R$ cassette of pSU2007 is inserted instead of *sul1 (*which protein product is implicated in the resistance to trimethoprim*), qacEdelta1* (coding for a multidrug efflux protein*),* and part of *orfA* (coding for a putative glutamine gamma-glutamyltransferase) and *orf5* (coding for a putative acetyltransferase)(**Figure 33**). In pLGM25, $Kn^R$ cassette is inserted instead of *ardC*, thus, no expression is observed. These substitutions are clearly seen in the expression profile of pSU2007 or pLGM25 (Figure 34). It is also observed that the *oriT* is an untranscribed region and that expression levels decrease as long as we move further in the operons from the promoter region. Alignment of reads from experiment #2 over pSU2007 and reads from experiment #3 over pLGM25 plasmid instead of R388 gave the same overall fold change results, with just smaller rearrangements and a 7.9-fold change overexpression of $Kn^R$ gene in pSU2007 respect pLGM25 experiment probably due to the origin of the $Kn^R$ cassette.

*trwN* and *kikA,* under the PtrwN promoter (repressed by both KorA and StbA) and *klcb* and *orf12* under de Porf12 and orf14 under Porf14 promoter (all repressed by both ArdK and StbA) (Fernandez-Lopez et al., 2014) are barely expressed in pLGM25 experiment #3 in comparison with pSU2007 experiment #2. Curiously, both overexpressed regions lie in opposed positions of the plasmid.

TrwN is a protein which closest homologue is VirB1 protein of pXAC64 plasmid (46% identity). VirB1-like proteins are thought to act as lytic transglycosylases, facilitating the assembly of the type IV secretion systems through localized lysis of the peptidoglycan (Hö et al., 2005).

KikA has been proved to be responsible of a marked reduction of viability of *Klebsiella oxytoca*, but not *E. coli* transconjugants after receiving an IncN plasmid. kikA, in collaboration with other two ORFs is known to cause reversible growth inhibition. (Holčík and Iyer, 1996) suggests that these genes are involved in the regulation of plasmid replication or conjugation.

KlcB is a transcriptional regulator. (Bhattacharyya and Figurski, 2001) found that the induction of the *klcB* gene is toxic to *E. coli* host cells that carry an IncP plasmid. They described that KlcB has a V-L-P domain in common with KorA by which both proteins interact causing KorA release, and thus, the derepression of KorA-regulated genes. In the case of R388, this derepression by KorA will shoot up the expression of the genes under PkikA, PkorA and PtrwH as it can be examined in Figure 34 and Table 21.

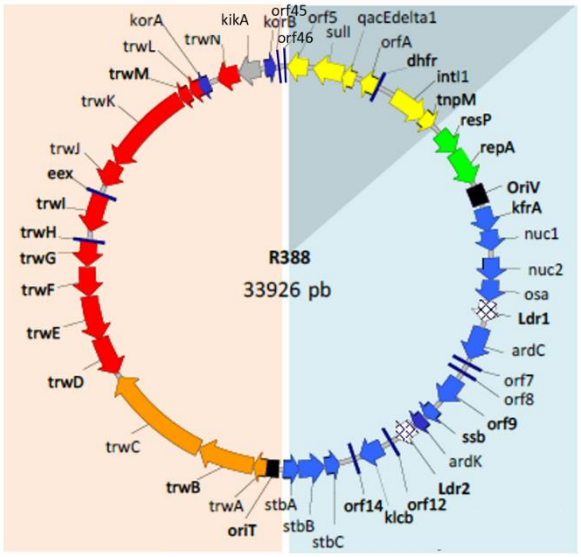



***Figure 33.*** *Genetic map of R388 plasmid. The figure shows the genetic organization of the plasmid, coloured by functional modules: conjugation region in light orange, maintenance, stability and replication region in blue, integration and antibiotic resistance in grey. Adapted from (Fernandez-Lopez, et al., 2006).*

Next to *klcB* is *orf12*, which product is predicted to be a putative type I R-M system methiltransferase subunit. On the other hand, *orf14* has similarity with a putative cold shock DNA-binding domain. These proteins seem to be involved in self-DNA protection and stress.

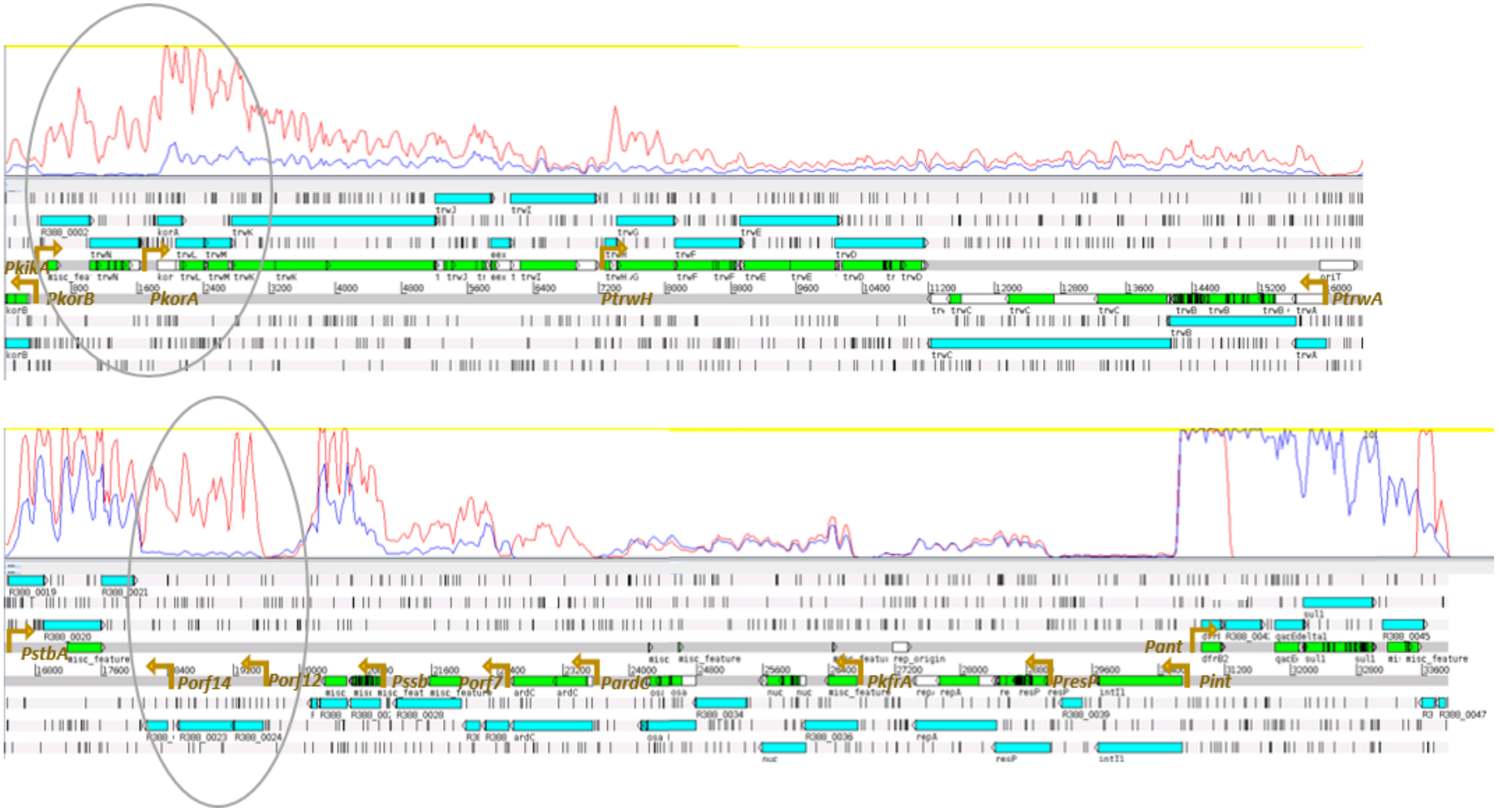



***Figure 34.*** *Expression profile of R388 genes. Experiment #2 (pSU2007 is in red) and #3 (pLGM25 is in blue) showing the number of reads (or coverage) by region. Promoter regions are indicated in yellow. The areas with the biggest expression differences are circled. Image created with Artemis.*

### 4.2.3.2 Differential expression of donor E. coli genes when the ardC-containing plasmid is transferred

In the *E. coli* genome analysis, there are few differentially downregulated genes when we compare *E. coli* with pSU2007 (experiment #2) respect *E. coli* with pLGM25 (experiment #3) in the presence of *P. putida*, as detailed in Table 22. Similar genes were downregulated when *E. coli* with pSU2007 (experiment #2) was compared with empty *E. coli* (experiment #1).

***Table 22.*** *Expression level of E. Coli genes in RPKMs from experiment #1 (BW), #2 (pSU2007) and #3 (pLGM25) and comparison of expression in RPKMs (Fold Change) for the most differentially expressed genes. Gene list is ordered from lowest to highest according to the Fold Change (pSU2007/pLGM25) column and colored from reddish to greener.*

| | Experiment | | | Fold Change | | | Information | |
|---|---|---|---|---|---|---|---|---|
| | BW | pSU2007 | pLGM25 | pSU2007/ BW | pLGM25/ BW | **pSU2007/ pLGM25** | Gene | Protein Function |
| b2576 | 201 | 90 | 243 | 0.45 | 1.21 | 0.37 | *srmB* | ATP-dependent RNA helicase |
| b4012 | 40 | 14 | 38 | 0.36 | 0.95 | 0.38 | *yjaB* | GNAT-family putative N-acetyltransferase |
| b0208 | 21 | 8 | 19 | 0.39 | 0.89 | 0.44 | *yafC* | LysR family putative transcriptional regulator |
| b1750 | 16 | 10 | 23 | 0.64 | 1.46 | 0.44 | *ydjX* | TVP38/TMEM64 family inner membrane protein |
| b0798 | 46 | 27 | 62 | 0.59 | 1.33 | 0.44 | *ybiA* | DUF1768 family protein |
| b0218 | 12 | 10 | 22 | 0.80 | 1.80 | 0.44 | *yafU* | |
| b0992 | 35 | 24 | 53 | 0.69 | 1.54 | 0.45 | *yccM* | putative 4Fe-4S membrane protein |
| b1546 | 23 | 19 | 39 | 0.81 | 1.69 | 0.48 | *tfaQ* | Qin prophage |
| b4204 | 79 | 60 | 124 | 0.76 | 1.57 | 0.48 | *yjfZ* | uncharacterized protein |
| b1374 | 114 | 95 | 193 | 0.83 | 1.69 | 0.49 | *pinR* | Rac prophage |
| b2297 | 423 | 242 | 492 | 0.57 | 1.16 | 0.49 | *pta* | phosphate acetyltransferase |
| b1545 | 109 | 91 | 184 | 0.83 | 1.68 | 0.50 | *pinQ* | Qin prophage |

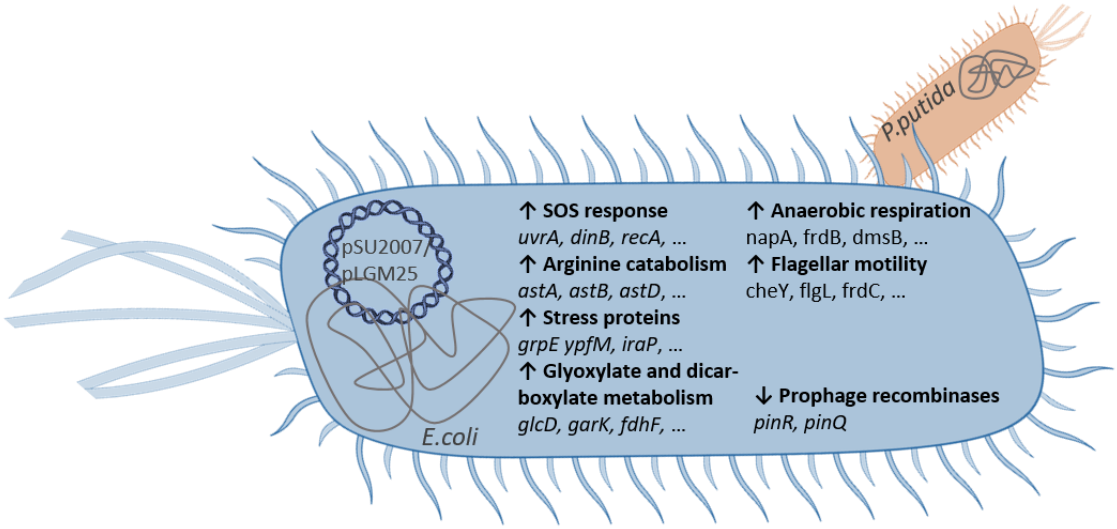



***Figure 35.*** *Processes and functions with differential expression in E. coli during conjugation to P. putida with plasmid having or lacking ardC. Overview as observed by RNA sequencing on a transcriptional level clustered by DAVID 6.7. Clustering annotation for differentially expressed genes in the presence of pSU2007 (#2) respect with pLGM25 plasmid (#3) is shown ordered from higher to lower enrichment score.*

***Table 23.*** *Expression level of E. coli genes in RPKMs from experiment #1 (BW), #2 (pSU2007) and #3 (pLGM25) and comparison of expression in RPKMs (Fold Change) for the most differentially expressed genes. Gene list is ordered from highest to lowest according to the Fold Change (pSU2007/pLGM25) column and coloured from greener to reddish. For complete list, see supplementary material. Gene names involved in the SOS signalling pathway are shown in bold.*

| | Experiment | | | Fold Change | | | Information | |
|---|---|---|---|---|---|---|---|---|
| | BW | pSU2007 | pLGM25 | pSU2007/ BW | pLGM25/ BW | **pSU2007/ pLGM25** | Gene | Protein function |
| b1389 | 10 | 19 | 2 | 1.86 | 0.18 | 10.49 | *paaB* | putative ring 1 |
| b1388 | 7 | 13 | 1 | 2.01 | 0.20 | 9.83 | *paaA* | ring 1 |
| b1391 | 9 | 13 | 1 | 1.54 | 0.17 | 9.11 | *paaD* | ring 1 |
| b1390 | 9 | 14 | 2 | 1.50 | 0.19 | 8.03 | *paaC* | ring 1 |
| b4618 | 1627 | 2354 | 298 | 1.45 | 0.18 | 7.90 | *tisB* | toxic membrane persister formation peptide |
| b2699 | 1014 | 1525 | 221 | 1.50 | 0.22 | 6.91 | ***recA*** | DNA recombination and repair protein |
| b3117 | 133 | 97 | 15 | 0.73 | 0.11 | 6.42 | *tdcB* | L-threonine dehydratase |
| b3113 | 135 | 92 | 15 | 0.68 | 0.11 | 6.04 | *tdcF* | putative reactive intermediate deaminase |
| b1882 | 11 | 14 | 3 | 1.26 | 0.23 | 5.51 | *cheY* | chemotaxis regulator transmitting signal to flagellar motor component |
| b3114 | 130 | 92 | 19 | 0.71 | 0.14 | 4.96 | *tdcE* | pyruvate formate-lyase 4/2-ketobutyrate formate-lyase |
| b2616 | 125 | 171 | 35 | 1.37 | 0.28 | 4.89 | ***recN*** | recombination and repair protein |
| b3116 | 173 | 125 | 27 | 0.72 | 0.15 | 4.70 | *tdcC* | L-threonine/L-serine transporter |
| b1848 | 730 | 1008 | 222 | 1.38 | 0.30 | 4.54 | ***yebG*** | DNA damage-inducible protein regulated by LexA |
| b2457 | 9 | 12 | 3 | 1.35 | 0.30 | 4.49 | *eutM* | ethanolamine utilization protein |
| b4471 | 76 | 48 | 11 | 0.63 | 0.14 | 4.44 | *tdcG* | L-serine dehydratase 3 |
| b3115 | 160 | 114 | 26 | 0.71 | 0.16 | 4.35 | *tdcD* | propionate kinase/acetate kinase C |
| b1184 | 13 | 24 | 5 | 1.85 | 0.43 | 4.30 | ***umuC*** | translesion error-prone DNA polymerase V subunit |
| b2698 | 99 | 108 | 26 | 1.09 | 0.26 | 4.22 | ***recX*** | regulatory protein for RecA |
| b1847 | 608 | 745 | 177 | 1.23 | 0.29 | 4.22 | *yebF* | extracellular Colicin M immunity family protein |
| b4606 | 115 | 96 | 23 | 0.84 | 0.20 | 4.16 | *ypfM* | stress-induced small enterobacterial protein |
| b0958 | 406 | 502 | 122 | 1.24 | 0.30 | 4.13 | ***sulA*** | SOS cell division inhibitor |
| b4408 | 800 | 1238 | 306 | 1.55 | 0.38 | 4.05 | *csrB* | CsrA-binding sRNA |
| b1846 | 163 | 224 | 58 | 1.38 | 0.35 | 3.88 | *yebE* | DUF533 family inner membrane protein |
| b4062 | 98 | 143 | 39 | 1.45 | 0.40 | 3.64 | *soxS* | superoxide response regulon transcriptional activator |
| b1183 | 24 | 33 | 9 | 1.35 | 0.37 | 3.62 | ***umuD*** | translesion error-prone DNA polymerase V subunit |
| b3157 | 68 | 64 | 19 | 0.94 | 0.28 | 3.39 | *yhbT* | SCP-2 sterol transfer family protein |
| b1396 | 13 | 13 | 4 | 1.04 | 0.31 | 3.34 | *paaI* | hydroxyphenylacetyl-CoA thioesterase |
| b2977 | 72 | 110 | 34 | 1.53 | 0.48 | 3.22 | *glcG* | DUF336 family protein |
| b2979 | 42 | 60 | 19 | 1.42 | 0.45 | 3.19 | *glcD* | glycolate oxidase subunit |
| b3118 | 27 | 24 | 8 | 0.88 | 0.28 | 3.15 | *tdcA* | tdc operon transcriptional activator |
| b1415 | 137 | 153 | 49 | 1.12 | 0.36 | 3.13 | *aldA* | aldehyde dehydrogenase A |
| b3565 | 8 | 10 | 3 | 1.32 | 0.42 | 3.12 | *xylA* | D-xylose isomerase |
| b1283 | 129 | 195 | 63 | 1.51 | 0.49 | 3.10 | *osmB* | osmotically and stress inducible lipoprotein |

However, when we arrange the *E. coli* genes from higher to lower fold change value of pSU2007/pLGM25 (**Table 23**), we observe a big number of differentially upregulated genes. Between those, we observe several SOS genes being differentially upregulated. The main affected cellular functions or pathways, in addition to SOS response, are depicted in Figure 35 according to DAVID classification. Curiously, these genes are downregulated if we compare the expression of BW genes in experiment 3 (BW with pLGM25) vs. experiment 1 (BW). It seems that when *E. coli* has the plasmid that lacks *ardC* (pLGM25), it downregulates the expression of a big number of genes and pathways. Between those, we found flagellar motility (*flgG, flgB, fliM, fliN, ...*), arginine catabolism (*astA, astB, astC…*) and other stress proteins (*sbmC, bsmA, ypfM*), sulfur metabolism (*dmsA, dmsB, dmsC*), glyoxylate and dicarboxylate metabolism (*acnB, aldA, glcF…*) as well as genes involved in the SOS response (*yebG, recA, recN…*) between other routes less affected. In the presence of the plasmid lacking *ardC*, only 15 genes are upregulated, being cell redox homeostasis (*ahpC, ahpF, grxA* and *trxC*) the most affected pathway (Figure 36).

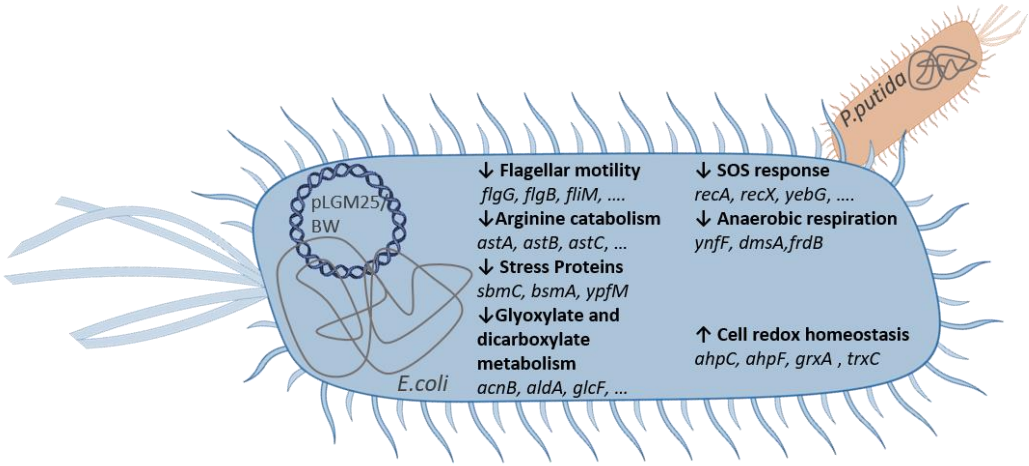



***Figure 36.*** *Processes and functions with differential expression in E. coli during conjugation of pLGM25 to P. putida. Overview as observed by RNA sequencing on a transcriptional level clustered by DAVID 6.7. Clustering annotation for differentially expressed genes in the presence of pLGM25 (#3) respect the empty cell (#1) is shown ordered from higher to lower enrichment score.*

The fold change for pSU2007/BW shows only 4 non-related upregulated genes and several downregulated, being flagellar motility (*flgG, flgB, fliM, fliN…*) the main downregulated pathway. As pSU2007/BW ratio do not vary for the SOS genes, there is not a stress response triggered in donors coupled to conjugation. *E. coli* cells carrying pLGM25 in experiment 3 try to transfer the plasmid to *P. putida* without success. These *E. coli* (pLGM25) donors could be doing a bigger metabolic effort to conjugate and as long as they are not able, they see themselves losing their integrity with a general altered expression profile. As there are no *eex* (exclusion system) signal, donors keep trying to conjugate unsuccessfully.

#### 4.2.3.3 Differential expression of P. putida genes

Lastly, in Table 24 we have ordered the recipient *P. putida* gene list from lower to higher fold change (pSU2007/pGM25P). The downregulated genes are involved in the amino acid and carbohydrate transport and metabolism (as *fruB, fruK, gnu-k…*). When we listed the genes in the opposite order (

Table 25 ), we observe an increment in the expression levels of several SOS genes when conjugating pSU2007 respect pLGM25P (or empty BW experiment). DNA damage and repair is the only functional group of upregulated genes (*recA, recN, lexA-I, endX, recX, dinB, yebG,* …). There are also some upregulated genes involved in stress response (as the mayor cold shock protein coding *cspA-I* gene), amino acid metabolism (*spuC-I, ldH*) and cell wall biogenesis (*ddlA*).

No SOS induction in pLGM25 *vs.* BW (values close to 1) may indicate that conjugation is not happening well in experiment #3 (as shown in Table 17) or that it is the consequence of ArdC not being present. However, the induction of the DNA damage repair pathway in recipients that happens in pSU2007 experiment #2 could be the consequence of the higher conjugation frequencies observed in accordance with (Baharoglu and Mazel, 2014) proposal postulating that it is conjugation which induces SOS pathways in recipients..

***Table 24.*** *Expression level of P. putida genes in RPKMs from experiment #1 (BW), #2 (pSU2007) and #3 (pLGM25) and comparison of expression in RPKMs (Fold Change) for every gene. Gene list is ordered from lowest to highest according to the RPKMs (pSU2007/pLGM25) column and colored from reddish to greener.*

| | Experiment | | | Fold Change | | | Information | |
|---|---|---|---|---|---|---|---|---|
| Locus_ tag | BW | pSU2007 | pLGM25 | pSU2007 /BW | pLGM25 /BW | **pSU2007/ pLGM25** | Gene | Protein Function |
| PP_3418 | 20 | 14 | 59 | 0,74 | 3,04 | 0,24 | | HP |
| PP_3417 | 9 | 7 | 27 | 0,72 | 2,86 | 0,25 | *gntT* | D-gluconate transporter |
| PP_4547 | 43 | 11 | 40 | 0,27 | 0,93 | 0,29 | | glutamine synthetase |
| PP_0794 | 27 | 25 | 86 | 0,92 | 3,19 | 0,29 | *fruK* | 1-phosphofructokinase monomer |
| PP_0795 | 22 | 21 | 68 | 0,93 | 3,05 | 0,31 | *fruA* | PTS fructose transporter subunit IIBC |
| PP_2441 | 652 | 334 | 886 | 0,51 | 1,36 | 0,38 | | HP |
| PP_4548 | 22 | 10 | 25 | 0,46 | 1,18 | 0,39 | | oxidoreductase |
| PP_0793 | 29 | 29 | 73 | 0,98 | 2,48 | 0,40 | *fruB* | PTS fructose transporter subunit EI/HPr/IIA |
| PP_0579 | 32 | 8 | 21 | 0,26 | 0,65 | 0,40 | | HP |
| PP_3148 | 19 | 5 | 13 | 0,28 | 0,66 | 0,42 | | glutamine synthetase |
| PP_3416 | 35 | 35 | 80 | 1,00 | 2,27 | 0,44 | *gnuK* | D-gluconate kinase |
| PP_3444 | 8 | 7 | 15 | 0,78 | 1,71 | 0,45 | | glyoxalase family protein |
| PP_0578 | 38 | 10 | 23 | 0,27 | 0,59 | 0,46 | | putative Methyltransferase |
| PP_2457 | 6 | 8 | 17 | 1,19 | 2,58 | 0,46 | *rbsR* | DNA-binding transcriptional repressor |
| PP_1006 | 80 | 44 | 91 | 0,55 | 1,15 | 0,48 | | heme receptor |
| PP_1368 | 29 | 15 | 30 | 0,50 | 1,03 | 0,48 | | glycosyltransferase domain-containing protein |
| PP_0596 | 96 | 84 | 173 | 0,87 | 1,80 | 0,49 | | omega-amino acid--pyruvate aminotransferase |
| PP_4070 | 545 | 252 | 506 | 0,46 | 0,93 | 0,50 | | HP |
| PP_0872 | 297 | 177 | 353 | 0,60 | 1,19 | 0,50 | *prfC* | peptide chain release factor 3 |
| PP_3767 | 180 | 59 | 117 | 0,32 | 0,65 | 0,50 | | HP |

***Table 25.*** *Expression level of P. putida genes in RPKMs from experiment #1 (BW), #2 (pSU2007) and #3 (pLGM25) and comparison of expression in RPKMs (Fold Change) for every gene. Gene list is ordered from highest to lowest according to the RPKMs (pSU2007/pLGM25) column and colored from greener to reddish. Gene names involved in the SOS signaling pathway are shown in bold.*

| | Experiment | | | Fold Change | | | Information | |
|---|---|---|---|---|---|---|---|---|
| Locus_tag | BW | pSU2007 | pLGM25 | pSU2007/BW | pLGM25/BW | **pSU2007/pLGM25** | Gene | Protein Function |
| PP_5744 | 18 | 76 | 16 | 4,27 | 0,87 | 4,91 | ***hp*** | HP LexA binding site predicted |
| PP_3773 | 7 | 51 | 13 | 7,44 | 1,96 | 3,80 | | HP |
| PP_4616 | 63 | 298 | 79 | 4,69 | 1,25 | 3,75 | ***yebG*** | LexA-dependent |
| PP_2177 | 9 | 29 | 8 | 3,26 | 0,89 | 3,68 | | transcriptional regulator |
| PP_4729 | 20 | 76 | 21 | 3,89 | 1,09 | 3,57 | ***recN*** | DNA repair/recombination protein |
| PP_4617 | 20 | 108 | 31 | 5,30 | 1,50 | 3,53 | *ldh* | leucine dehydrogenase |
| PP_5580 | 42 | 33 | 10 | 0,80 | 0,23 | 3,45 | | HP |
| PP_2109 | 30 | 125 | 36 | 4,22 | 1,23 | 3,43 | ***hp*** | HP LexA binding site predicted |
| PP_2143 | 58 | 206 | 62 | 3,56 | 1,07 | 3,31 | ***lexA-I*** | transcriptional repressor |
| PP_3694 | 9 | 13 | 4 | 1,41 | 0,44 | 3,21 | | HP |
| PP_3850 | 5 | 15 | 5 | 3,30 | 1,05 | 3,15 | | HP |
| PP_2924 | 9 | 23 | 8 | 2,68 | 0,90 | 2,98 | ***hp*** | HP LexA binding site predicted |
| PP_5579 | 29 | 24 | 9 | 0,82 | 0,29 | 2,79 | | HP |
| PP_4346 | 25 | 25 | 9 | 1,01 | 0,37 | 2,76 | *ddlA* | D-alanine--D-alanine ligase A |
| PP_1521 | 101 | 107 | 40 | 1,06 | 0,39 | 2,71 | | HP |
| PP_5091 | 125 | 141 | 52 | 1,12 | 0,41 | 2,71 | | membrane protein |
| PP_2451 | 45 | 122 | 47 | 2,72 | 1,06 | 2,57 | *endX* | extracellular DNA endonuclease |
| PP_1522 | 335 | 339 | 134 | 1,01 | 0,40 | 2,53 | *cspA-I* | major cold shock protein |
| PP_3109 | 14 | 9 | 4 | 0,68 | 0,27 | 2,52 | | HP |
| PP_5637 | 9 | 20 | 8 | 2,31 | 0,93 | 2,49 | | HP |
| PP_5487 | 178 | 211 | 85 | 1,19 | 0,48 | 2,49 | ***hp*** | HP LexA binding site predicted |
| PP_1203 | 21 | 57 | 24 | 2,73 | 1,15 | 2,37 | ***dinB*** | DNA polymerase IV |
| PP_1630 | 217 | 712 | 306 | 3,28 | 1,41 | 2,33 | ***recX*** | regulatory protein RecX |
| PP_3089 | 3218 | 4942 | 2135 | 1,54 | 0,66 | 2,32 | *hcp1* | Hcp1 |
| PP_3901 | 22 | 23 | 10 | 1,06 | 0,46 | 2,30 | | HP |
| PP_1629 | 462 | 1654 | 720 | 3,58 | 1,56 | 2,30 | ***recA*** | recombinase RecA |
| PP_1625 | 878 | 925 | 404 | 1,05 | 0,46 | 2,29 | *fdxA* | ferredoxin 1 |
| PP_2839 | 26 | 87 | 38 | 3,37 | 1,49 | 2,26 | ***hp*** | HP LexA binding site predicted |
| PP_2180 | 15 | 20 | 9 | 1,35 | 0,61 | 2,23 | *spuC-I* | polyamine: pyruvate transaminase |
| PP_4350 | 8 | 12 | 6 | 1,61 | 0,72 | 2,23 | | class V NifS/IscS family aminotransferase |
| PP_5694 | 784 | 851 | 391 | 1,09 | 0,50 | 2,18 | | HP |
| PP_0641 | 245 | 481 | 221 | 1,96 | 0,91 | 2,17 | | HP |
| PP_4349 | 16 | 23 | 11 | 1,47 | 0,68 | 2,17 | | HP |
| PP_2838 | 47 | 144 | 66 | 3,06 | 1,42 | 2,16 | | HP |
| PP_5464 | 24 | 38 | 17 | 1,59 | 0,74 | 2,15 | | HP |
| PP_2840 | 11 | 33 | 16 | 2,93 | 1,39 | 2,11 | | membrane protein |
| PP_3592 | 9 | 17 | 8 | 1,85 | 0,88 | 2,10 | | RpiR family transcriptional regulator |
| PP_1631 | 24 | 82 | 40 | 3,46 | 1,69 | 2,05 | | HP |
| PP_3117 | 62 | 159 | 78 | 2,58 | 1,27 | 2,03 | ***3117*** | DNA lesion error-prone processing protein |
| PP_4856 | 16 | 23 | 11 | 1,47 | 0,73 | 2,01 | | Dps family ferritin |
| PP_5554 | 68 | 39 | 19 | 0,57 | 0,28 | 2,00 | | HP |

To sum up, we have observed that when *ardC* is present in the plasmid (exp. #2), conjugation events occur and a SOS response is triggered in recipient cells. However, when *ardC* is not

present in the plasmid (exp. #3), conjugation events do not occur so often and SOS response is downregulated in these donor cells.

#### 4.2.3.4 Differential expression of EM422

After the results observed in Section 4.4.7.1 when mating from *E. coli* to *P. putida EM422* (*KT2440*ΔhsdRMS) we decided to analyze if there were expression differences in our RNA-seq results for the tree genes deleted in EM422 that could give us a clue about the target of ArdC protease. However, none of the *hsdRMS* genes showed a differential expression between the three experiments. This result indicates that ArdC does not affect the transcriptional regulation of *hsdRMS*. We can conclude that the observed transcriptional effects are probably due to the mating process itself.

## 4.3 ArdC structural characterization

### 4.3.1 Overall crystal structure

R388 ArdC crystal structure was solved at a resolution of 2.00 Å by molecular replacement (MR) using a selenomethionine (Se-Met) derivative protein structure solved by single anomalous dispersion (SAD).

ArdC-SeMet crystals were obtained at 20 mg/mL with precipitant solution containing 0.1 M HEPES pH 7.5; 10 % w/v polyethylene glycol 6,000 and 5 % v/v (+/-)-2-methyl-2,4-pentanediol and were cryoprotected with 10 % 2-methyl-2 4-pentanediol (Figure 37 A).

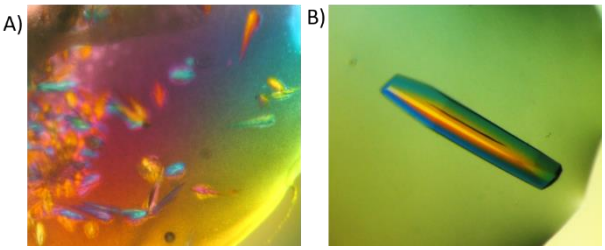



***Figure 37.*** *ArdC protein crystals. A) ArdC- SeMet derivative and B) Native ArdC protein crystallized at G6 Crystal Screen HT condition.*

ArdC native crystals were obtained from ArdC protein solution at 20 mg/mL with the precipitant solution containing 0.1 M HEPES pH 7.5; 10 % w/v polyethylene glycol 6,000 and 5 % v/v (+/-)-2-methyl-2,4-pentanediol; cryoprotected with 20 % 2-methyl-2 4-pentanediol (Figure 37 B).

Experimental data of the X-Ray diffraction solution from the crystals obtained (Figure 37) are shown in Table 26.

***Table 26.*** *Data collection and refinement statistics for ArdC structure. Statistics for the highest-resolution shell are shown in parentheses.*

| | ArdC |
|---|---|
| **Wavelength** | 0.9792 |
| **Resolution range** | 39.49-2.0 (2.072 -2.0) |
| **Space group** | R 3 :H |
| **Unit cell** | 136.798 136.798 51.7013 90 90 120 |
| **Total reflections** | 537291 (33696) |
| **Unique reflections** | 24366 (2408) |
| **Multiplicity** | 22.1 (13.9) |
| **Completeness (%)** | 99.81 (98.49) |
| **Mean I/sigma(I)** | 37.15 (3.14) |
| **Wilson B-factor** | 31.87 |
| **R-merge** | 0.7636 (1.274) |
| **R-meas** | 0.7814 (1.326) |
| **CC1/2** | 0.773 (0.474) |
| **CC*** | 0.934 (0.802) |
| **Reflections used in refinement** | 24339 (2406) |
| **Reflections used for R-free** | 1230 (111) |
| **R-work** | 0.1726 (0.2154) |
| **R-free** | 0.1976 (0.2410) |
| **CC(work)** | 0.864 (0.814) |
| **CC(free)** | 0.817 (0.837) |
| **Number of non-hydrogen atoms** | 2412 |
| **Protein residues** | 276 |
| **RMS(bonds)** | 0.008 |
| **RMS(angles)** | 1.16 |
| **Ramachandran favored (%)** | 97.76 |
| **Ramachandran allowed (%)** | 1.49 |
| **Ramachandran outliers (%)** | 0.75 |
| **Rotamer outliers (%)** | 0.00 |
| **Clashscore** | 3.46 |
| **Average B-factor** | 36.39 |

The crystal belongs to the trigonal space group H3, with unit-cell parameters a=b= 136.8 Å, c = 51.7 Å, and contains one molecule in the asymmetric unit. ArdC, as shown in Figure 38 is composed of two structural domains: an N-terminal domain (residues 1-134) and a C-terminal domain (residues 151-297) joined by a long and flexible loop (135-150).

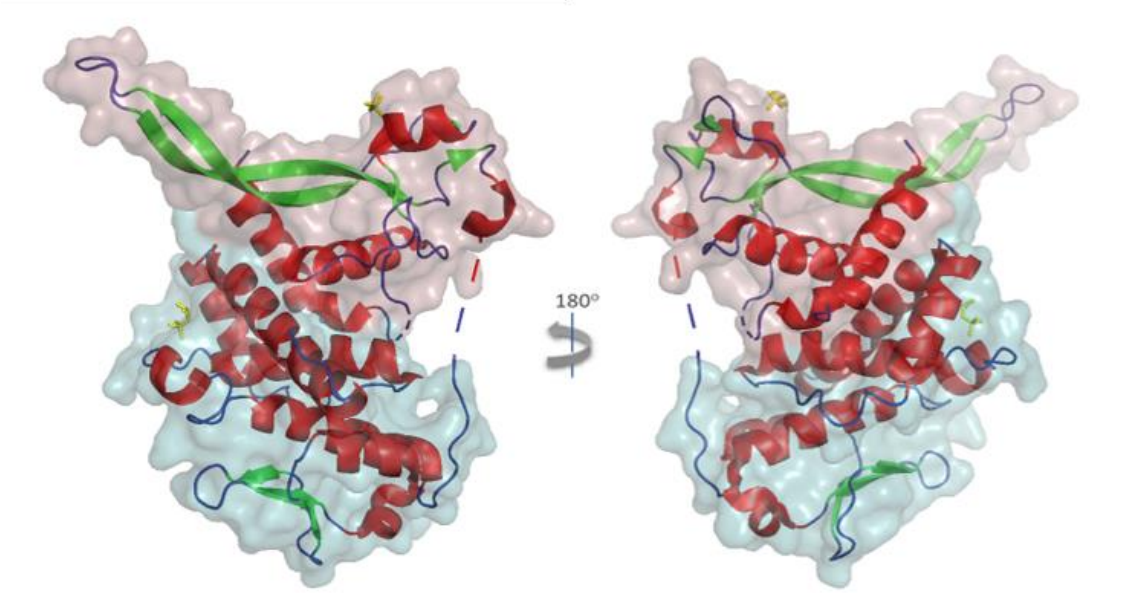



***Figure 38.*** *Overall structure of ArdC. N-terminal domain surface is shown in pink and C-terminal domain surface is shown in blue. Cartoon representation is also shown, with α-helices in red and β-strands in green. The disordered loop joining both domains is schematized by a dotted line. The two MPD (2-methyl-2,4-pentanediol) molecules from crystallizing buffer are shown in yellow.*

In the native structure we could not observe electron density for residues 136-141 connecting both domains. Moreover, electron density is not observed for the end terminal residues 1-6 and 294-297 nor the flexible small loop residues 33-39. The N-terminal domain is composed of three α-helices (α1-α3), a three-stranded β-sheet (β1, β3 and β4) that supports a long and protuberant β-hairpin (β3-β4) interacting by the β4 with the small β1, a smaller two-stranded antiparallel β-sheet formed by β2 and β5, as well as other smaller secondary structures shown in Figure 39. The C-terminal domain is composed of six α-helices (α4-α9) and three short stranded antiparallel β-sheets (β6-β8) as shown in Figure 38, Figure 39 and Figure 40. For a clearer labelling, the protein structure is colored from N-terminal to C-terminal and tagged in **Figure 41**. The overall dimensions of the structure are approximately 45 Å x 60 Å x 70 Å.

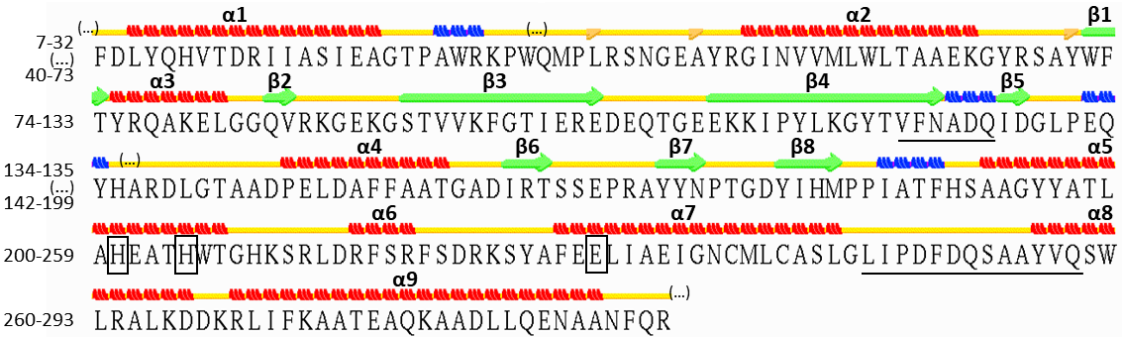



***Figure 39.*** *ArdC secondary structure representation. Helices are shown in red labelled from α1 to α9. Beta-strands are shown in green labelled from β1 to β8. Turns and coils are shown in yellow. $3_{10}$ helices are shown in blue. Isolated β bridges are shown in orange. Ellipsis represent non-solved structure. The residues coordinating the metal are framed and the "squiggle" signature proposed by* (Krishnan et al., 2018) *and the "Antirestriction" signature defined by* (Belogurov et al., 2000) *are underlined. Image obtained from Stride web server.*

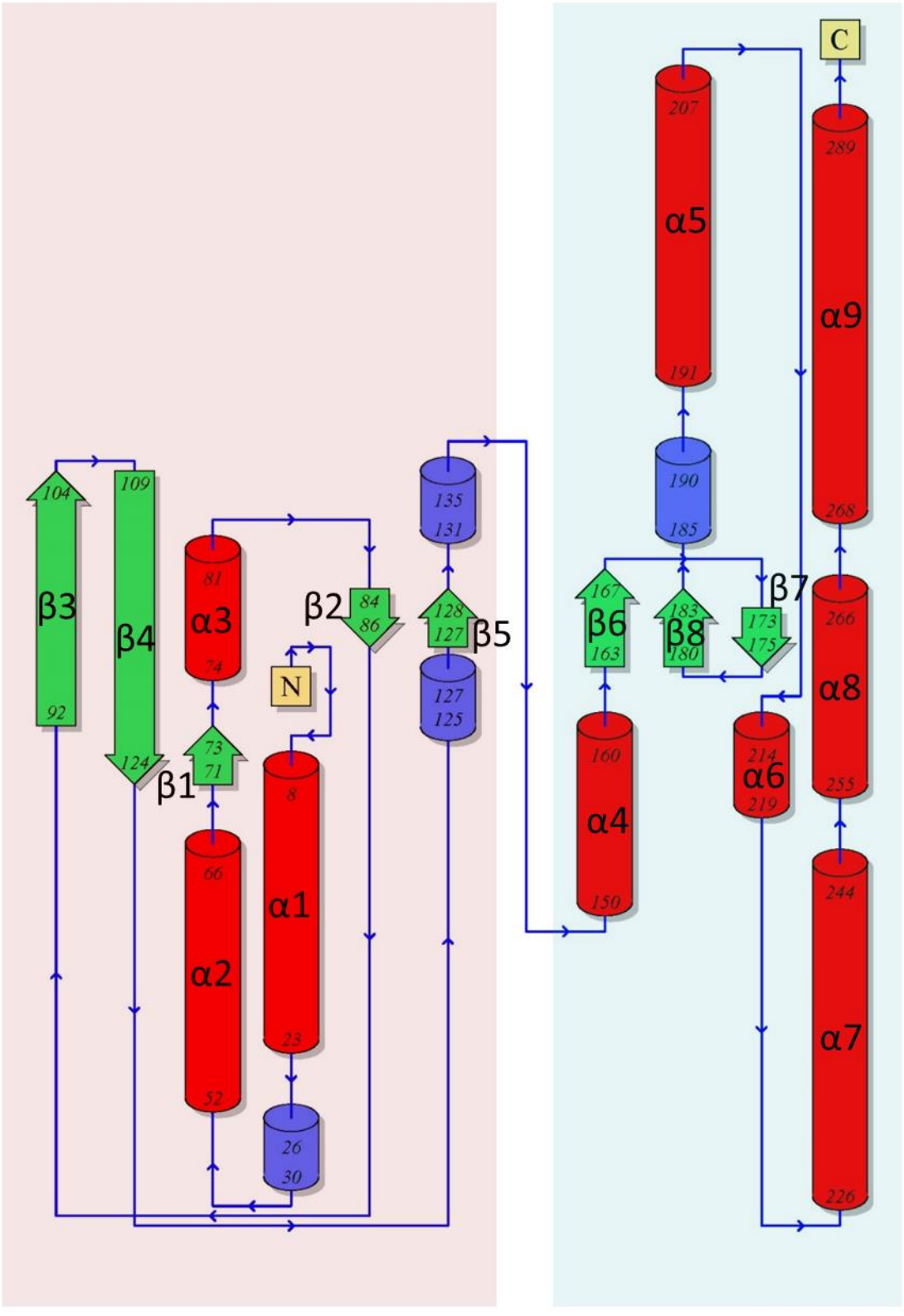



***Figure 40.*** *Schematic topological representation of ArdC protein. N-terminal domain is shadowed in pink and C-terminal domain in blue. Structural scaffold is colored as in figure 39: α-helixes are shown in red labelled from α1 to α9. β-strands are shown in green labelled from β1 to β8. $3_{10}$ helices are shown in blue. Image was modified from the one generated by PDBsum server.*

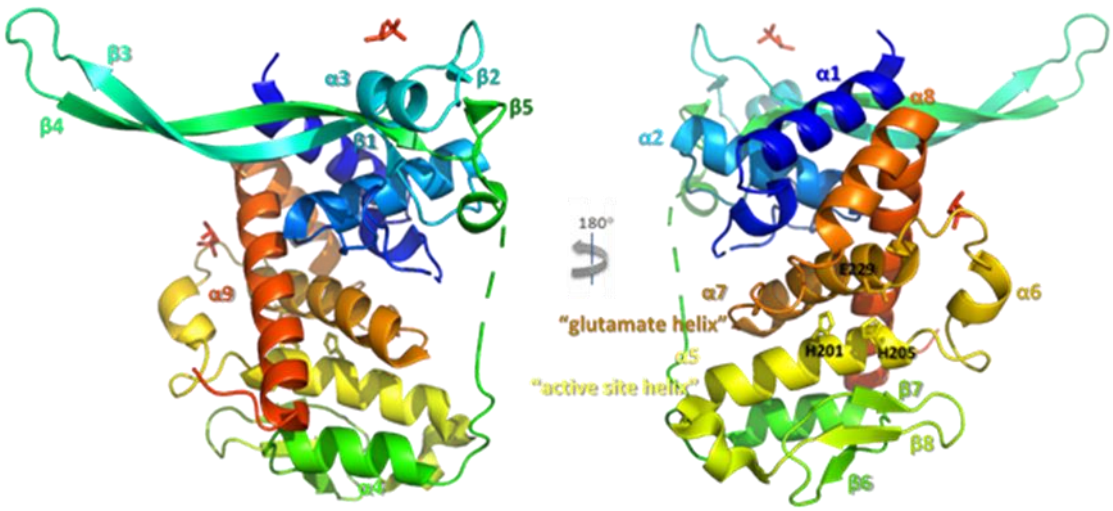



***Figure 41.*** *Overall labelled crystal structure of ArdC protein. Cartoon representation of ArdC tertiary structure coloured from bluish (N-terminal) to reddish (C-terminal). The disordered loop is schematized by a dotted line. Residues composing the metal-binding site are labelled and in ball-and-stick representation as well as the two MPD (2-methyl-2,4-pentanediol) molecules shown in red.*

ArdC solved structure was compared by structural domains to structures deposited in the Protein Data Bank (PDB) using the Dali server to find the closest structural homologues. Significant similarities have a *Z*-score above 2. The most significate results are shown in Table 27.

*Table 27. ArdC closest structural homologues obtained by Dali for each structural domain.*

| | Z-score | % id | PDB | Protein |
|---|---|---|---|---|
| **N-terminal domain** | 4.5 | 9 | 2QSG | Rad4 DNA repair protein |
| **C-terminal domain** | 6.4 | 12 | 6MDW | Spartan metalloprotein |
| | 6.2 | 16 | 3DTE | IrrE metalloprotease |
| | 5.8 | 6 | 1ADU | Adenovirus ssDNA binding protein |

Using the N-terminal domain, the DNA repair protein Rad4 was identified as the best and unique match to this domain (2QSG, Z score = 4.9), thus, we will define it for now as a DNA binding domain. The human metalloprotease Spartan protein (6MDW; Z score = 7.0), the regulator metalloprotease IrrE involved in DNA repair (3DTI; Z score = 6.1) and the adenovirus ssDNA binding protein (1ADU; Z score = 5.8), were identified as the best matches when the C-terminal domain of ArdC was submitted to the Dali server. Due to this similarities, we will define it for the time being as a metalloprotease domain. A detailed description of each domain is provided below.

A protein alignment shows us the most conserved amino acids of the protein, found along the whole length of ArdC but specially in the C-terminal half of the protein sequence (Figure 42).

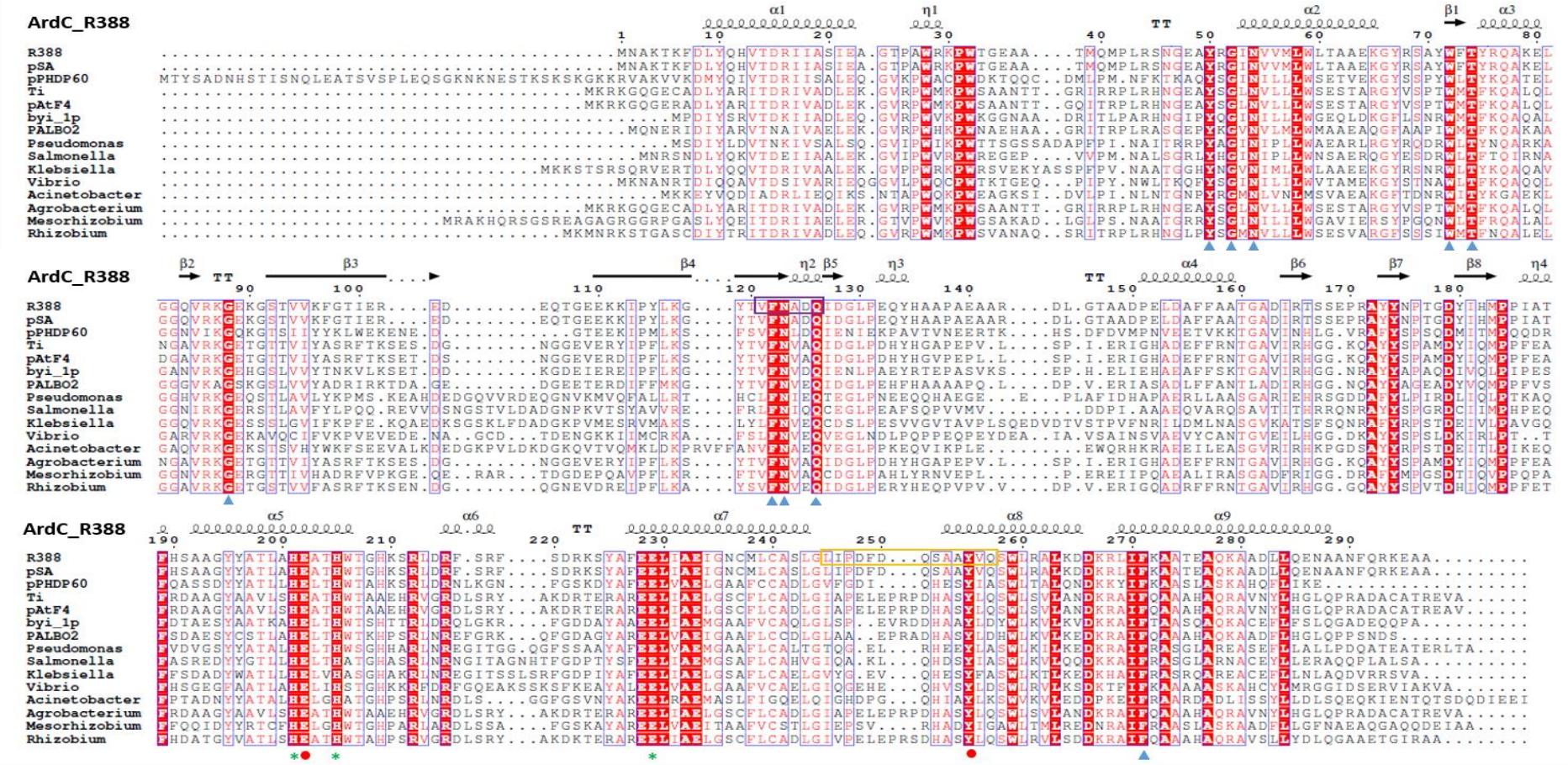



***Figure 42.*** *Structure-based sequence alignment of ArdC from different plasmids and microorganisms. The secondary structural elements of ArdC are shown at top. The conserved residues are highlighted in red. The metal binding residues are marked with a green asterisk. Polar and aromatic amino acids in disposition to bind DNA are marked with blue triangles. Other residues putatively involved in the catalysis are shown with red circles. The "squiggle" motif is framed in purple and the putative "antirestriction" motif is framed in yellow. GenBank accession numbers of the sequences used for the alignment: R388 plasmid "YP_009182134.1" , pSA plasmid "AAD52160.1", pPHDP60 plasmid "AGE91731.1" from Photobacterium damselae, pTiBo5 plasmid " AAZ50566.1" and pAtF4 plasmid "KJX90203.1" from A.tumefaciens, byi_1p plasmid "AET95037.1" from Burkholderia sp. YI23, PALBO2 plasmid "APW64319.1" from Paludisphaera borealis, Pseudomonas Putida "BAW24104.1", Salmonella enterica "WP_017441175.1", Klebsiella pneumoniae "WP_004151764.1", Vibrio cholerae "WP_042988667.1", Acinetobacter baumannii "WP_064534766.1", Agrobacterium tumefaciens "WP_012478119.1", Mesorhizobium loti "WP_027033346.1" and Rhizobium "WP_071835905.1". Image generated by ESPript3 with 6I89.pdb ArdC structure.*

Between those conserved amino acids, we found some residues probably important for structure stability as Q126, other polar and aromatic amino acids needed for binding to DNA as W72, and those forming part of the metalloprotease active center as H201. Interestingly, the motif LIpDfdQS-aayvQ similar to the "antirestriction" motif conserved for all other known Ard proteins (Belogurov et al., 2000) does not seem to be conserved in the ArdC family except for its Y255 (**Figure 42**). In addition, we found a conserved motif, that represents a crossover of the polypeptide chain going from β4 to β5 and from β2 to β3 (Figure 40, Figure 42). In the β4 to β5 region, there is a $3_{10}$ helix that helps to create a big twist of the chain known as a "squiggle" motif. This motif is formed by a hhsxxQ sequence being h hydrophobic residues (the first one usually aliphatic and the second aromatic), s a small residue, x any residues and a conserved final glutamine being VFNADQ the amino acids in ArdC. It has been proposed by (Krishnan et al., 2018) that this squiggle motif may be responsible of a highly flexible region that could facilitate recognition of DNA sequences by generating conformational changes.

The degree of evolutionarily conservation of an amino acid position is strongly related to its structural and functional importance. If we focus our attention in the position of these conserved amino acids in the protein structure (Figure 43) we can see how those are mainly near the metalloprotease active center. Interestingly, the N-terminal domain is less conserved, and specially, the sequence of the protruding β-hairpin is not well conserved.

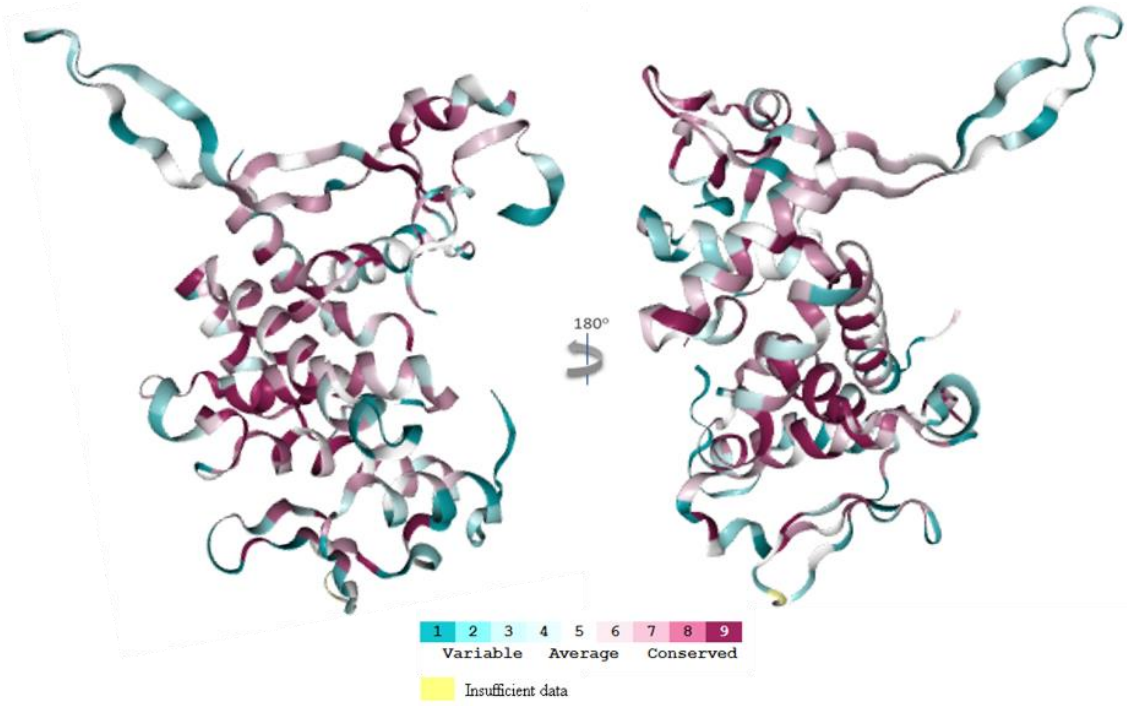



***Figure 43.*** *ArdC amino acid conservation. Evolutionary conservation of amino acid positions in ArdC protein based on the phylogenetic relations between homologous sequences. Image obtained by ConSurf server over 6I89.pdb. Alignment was build using ClustalW. Homologs were collected from UniProt database with HMMER search algorithm picking 150 representative samples with a sequence identity with the query between 35% and 70%. Conservation scores were calculated by Bayesian method and colored according to the scale of colors shown in the bottom from blueish to purplish for the most variable to the most conserved amino acids.*

The surface electrostatic map (Figure 44) reveals a positively charged grove in the region of the N-terminal domain adjacent to the C-terminal domain suggesting a possible DNA binding groove between both structural domains. This could be the reason why when using ArdC C-terminal domain for Dali, in addition to metalloproteases, we also recover the Adenovirus ssDNA binding protein. In the opposite face of the protein, the catalytic pocket is negatively charged in disposition to bind a divalent cation.

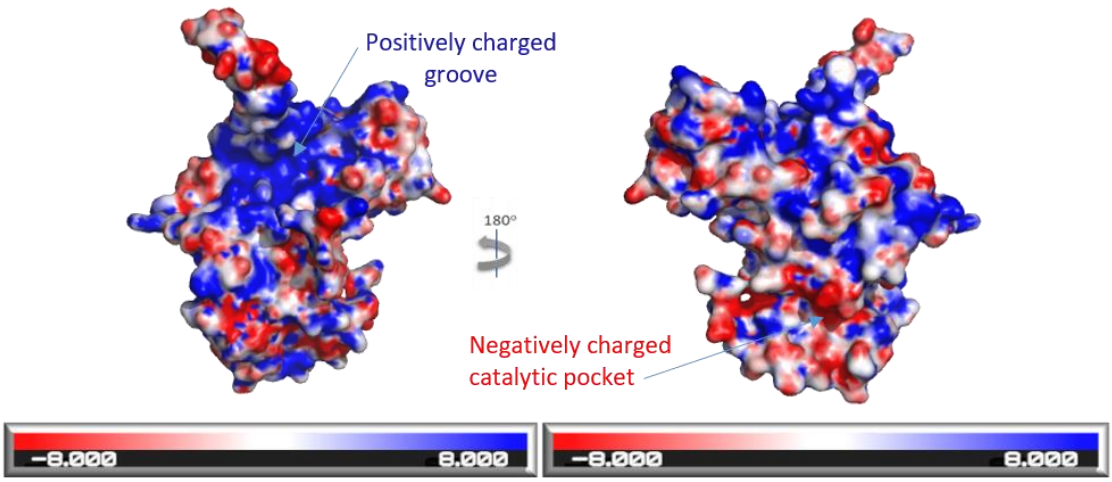



***Figure 44.*** *Electrostatic potential surface of ArdC protein. Negative surface is colored in red, positive in blue (calculated by APBS tool).*

### 4.3.2 Nucleotide excision repair protein as closest ArdC N-terminal domain

Rad4 is a component of the nucleotide excision repair (NER) pathway in yeast that acts in the removal of bulky and DNA helix distortions as thymine dimers caused by UV. Rad4 is homolog to the human protein XPC. Rad4 is composed by an inactive transglutaminase TGL fold domain and three different BHD domains to display a bigger DNA binding surface (Min and Pavletich, 2007). ArdC N-terminal is more similar at a sequence level to the second BHD domain of Rad4 (BHD2), however BHD2 is considerable smaller (about 50 amino acids long compared to the 134 residues of ArdC-N) as it lacks some ArdC-N structural features, as the starter ArdC α1, α2, and final $3_{10}$ motifs. In addition, BHD2 motifs are shorter, specially its protuberant β-hairpin (Figure 45).

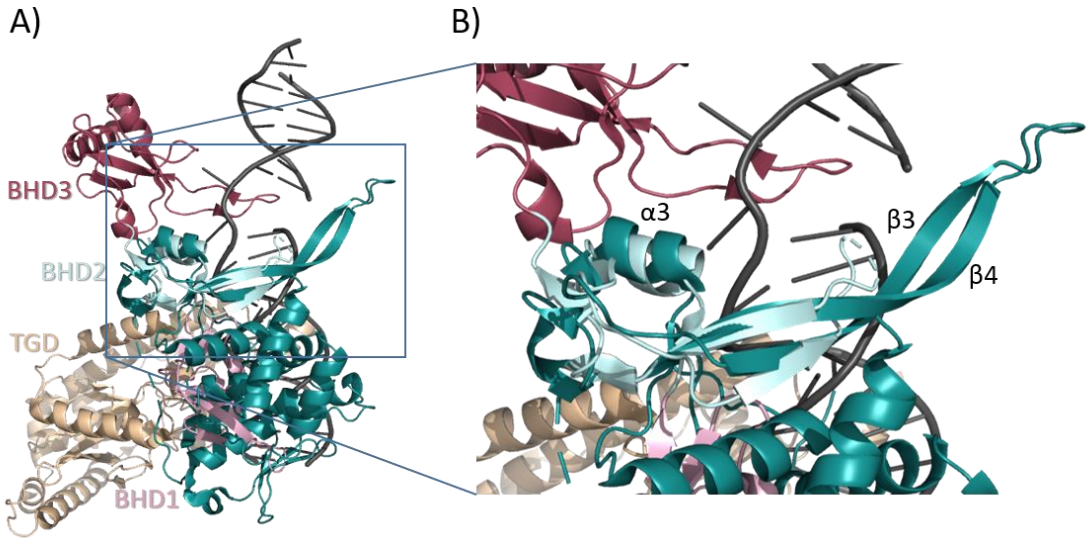



***Figure 45.*** *ArdC-Rad4 structural alignment. Superposition of ArdC structure (in teal, 2.0 Å) with its closest structural homologue Rad4 protein bound to UV-damaged DNA (2QSG; 3,1 Å) Rad4 TGD domain is shown in wheat, BHD1 in pink, BHD2 domain in light blue, BHD3 in raspberry. For clarity, Rad23 has been removed from the picture. B) A detailed view of ArdC N-terminal superposed to Rad4 BHD2 domain.*

Rad4 senses the damage and inserts the loop at the tip of the long β-hairpin through the DNA duplex making the damaged base pairs to flip out of the helix to expose them to other NER

enzymes (Min & Pavletich, 2007, Krishnan et al., 2018). Rad4 binds phosphate, ribose and base groups of the undamaged strand. However, ArdC-N β-hairpin is much longer. In our case, ArdC has a very electronegative loop in the tip of the β-hairpin (Figure 44), predicting electrostatic repulsion with the backbone phosphates of the DNA. The alignment in Figure 46 shows the amino acid conservation between Rad4 (BHD2) and ArdC N-terminal. Interestingly conserved amino acid Q77 marked with an * in Figure 46 is in disposition to interact as Rad4 Q495 with the DNA.

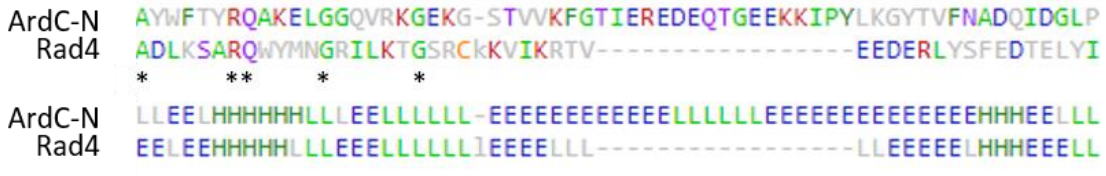



***Figure 46.*** *Protein sequence alignment of ArdC-N and Rad4. In the second line, the secondary structure assignments (H stands for helix, E for strand, and L for coil) is shown. The most frequent amino acid type is colored for each position. Conserved positions are marked with an *.*

## 4.3.3 ArdC N-terminal domain

To further identify the DNA binding region within ArdC, we obtained protein crystals with 5Ts oligonucleotide (5´TTTTT3´) at Hampton screening HT conditions A9 (0,2 M ammonium acetate, 0,1 M sodium citrate tribasic dehydrate pH 5,6 and 30 % w/v polyethylene glycol 4,000) and D10 (0,2 M calcium acetate hydrate, 0,1 M sodium cacodylate trihydrate pH 6,5 and 18 % w/v polyethylene glycol 8,000). The last ones were obtained at 15 mg/mL and were cryoprotected with an additional 20 % ethylene glycol (Figure 47). Experimental data of the X-Ray diffraction solution from the D10 crystal obtained (Figure 47) is shown in **Table 28**.

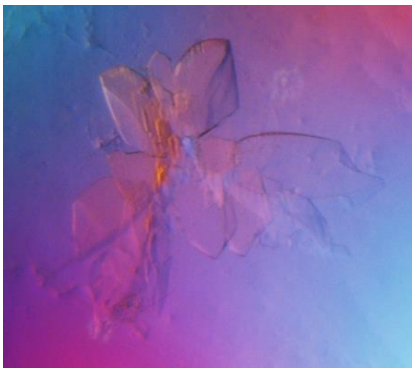


***Figure 47.*** *ArdC-ssDNA crystals. Crystals were obtained at D10 Crystal Screen HT condition.*

***Table 28****. Data collection and refinement statistics for ArdC-DNA structure. Statistics for the highest-resolution shell are shown in parentheses.*

| | ArdC-DNA |
|---|---|
| **Wavelength** | |
| **Resolution range** | 51.78-1.8 (1.864-1.8) |
| **Space group** | C 2 2 21 |
| **Unit cell** | 67.9797 83.3698 103.57 90 90 90 |
| **Total reflections** | 1296961 (125319) |
| **Unique reflections** | 27617 (2704) |
| **Multiplicity** | 47.0 (46.3) |
| **Completeness (%)** | 1.00 (1.00) |
| **Mean I/sigma(I)** | 100.66 (4.73) |
| **Wilson B-factor** | 20.78 |
| **R-merge** | 0.7386 (1.77) |
| **R-meas** | 0.7471 (1.791) |
| **CC1/2** | 0.937 (0.722) |
| **CC*** | 0.984 (0.916) |

| | ArdC-DNA |
|---|---|
| **Reflections used in refinement** | 27587 (2701) |
| **Reflections used for R-free** | 1265 (125) |
| **R-work** | 0.1882 (0.2139) |
| **R-free** | 0.2282 (0.2746) |
| **CC(work)** | 0.851 (0.781) |
| **CC(free)** | 0.841 (0.672) |
| **Number of non-hydrogen atoms** | 2411 |
| **Macromolecules** | 2132 |
| **Protein residues** | 267 |
| **RMS(bonds)** | 0.006 |
| **RMS(angles)** | 1.09 |
| **Ramachandran favored (%)** | 97 |
| **Ramachandran allowed (%)** | 1.5 |
| **Ramachandran outliers (%)** | 1.2 |
| **Rotamer outliers (%)** | 0.48 |
| **Clashscore** | 4.07 |
| **Average B-factor** | 26.48 |

The Cα alignment of ArdC-DNA structure (1.8 Å) with the native ArdC structure (2.0 Å) gives an r.m.s.d. value of 0.81 Å (on 254 aligned residues) with a displacement of the protuberant β-hairpin towards the α3 when bound to DNA (**Figure 48**).

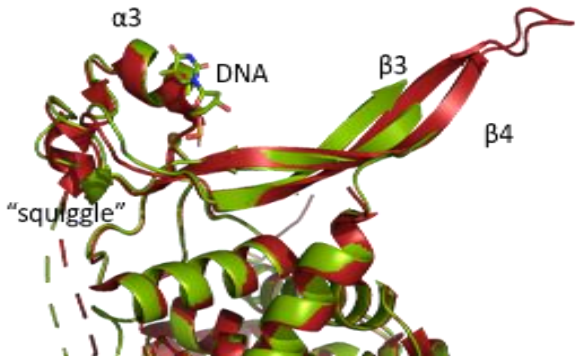



***Figure 48.*** *Structural alignment of ArdC and ArdC bound to DNA. Only part of the ArdC-N is shown for clarity. ArdC is in maroon and ArdC bound to DNA (dT) is shown in olive green.*

Even though we could not solve the structure at a good resolution due to the poor diffraction of the crystal grown in A9 condition, we observed for both A9 and D10 conditions a different space group packaging C2221, and electronic density facing the aromatic ring of Y75 residue. Although we were only able to locate one thymine base we observe that it could be the thymine aromatic ring interacting with the tyrosine amino acid ring through DNA - protein, face to face, π - π stacking interactions. At this position of the homologue proteins (see **Figure 42**), we only find tyrosine or phenylalanine amino acids, both aromatic amino acids able to bind DNA trough this kind of nucleobase – amino acid π-interactions (Wilson et al., 2014) (Figure 49). In addition, a clear electropositive groove is generated between two molecules crystallized in contiguous asymmetric units indicating a putative ssDNA binding region between two molecules as it can be observed in Figure 49.

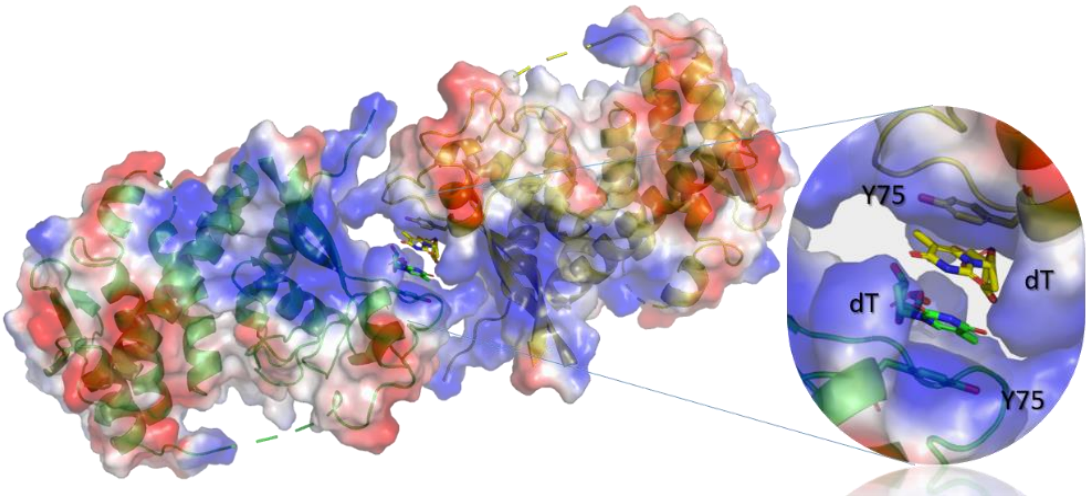



***Figure 49.*** *Surface electrostatic potential of two ArdC molecules crystallized with ssDNA The two molecules are oriented in symmetrical positions along DNA axis. Electropositive area is colored in blue and electronegative area in red. A closer view shows the stacking between the aromatic ring of the thymine and the tyrosine Y75 between both molecules (one shown in yellow and another in green).*

## 4.3.4 ArdC C-terminal domain

### 4.3.4.1 ArdC thermal stability

Knowing that ArdC is a ssDNA binding protein and predictors indicated a putative metallodomain, we decided to study the stability of ArdC in the presence of different ssDNA oligonucleotides and metal cofactors. These studies could provide us information about the metal or DNA requirements for ArdC activity, in addition to helping us to find conditions that stabilize ArdC to increase the crystallization opportunities. Protein stability was measure by the ThermoFluor assay as described in materials and methods. ThermoFluor is a thermal stability assay based on fluorescence measurements as protein is being unfolded by increments in temperature. The melting temperature ($T_M$) for different metals and DNAs calculated using SYPRO® Orange dye are shown in **Table 29**.

***Table 29.*** *$T_M$ value of ArdC in different solutions. $T_M$ values are for protein in 100 mM Tris-HCl pH 7.5, 500 mM NaCl alone or plus EDTA, different metals, ssDNA oligonucleotides or dsDNA fragments. $T_M$ was calculated by fluorescence measurements using SYPRO® Orange dye.*

| Condition | $T_M$ (°C) |
|---|---|
| - | 56 |
| 1 mM **EDTA** | 55 |
| 1 mM **$NiCl_2$** | **62** |
| 1 mM **$CaCl_2$** | 55 |
| 1 mM **$MgCl_2$** | 55 |
| 1 mM **$MnCl_2$** | **60** |
| 1 mM **$ZnCl_2$** | 56 |
| 1 mM **$CuCl_2$** | **59** |
| 1 mM **$CoCl_2$** | **69** |
| 1 mM **$FeCl_3$** | 55 |
| 7.5 µM **ssDNA** (8,23,45,57 nt) | 55-56 |
| 7.5 µM **dsDNA** (45 bp) | 55 |

Increased $T_M$ values of ArdC were obtained in the presence of $Ni^{2+}$, $Mn^{2+}$, $Cu^{2+}$ and $Co^{2+}$. ArdC is thus more stable in the presence of any of these four metals as more temperature is needed to have half of the protein denatured. For cobalt, an especially slower denaturalization process was observed, indicated by a melting curve with a less inclined slope that gave a distribution with a flat maximum in the first derivative of 69 °C (**Figure 50**). In the rest of the conditions with other metals, the $T_M$ profile is similar to the one with the protein alone or with EDTA ($T_M$ between 55 °C and 56 °C). DNA in single-strand form of different lengths or in double-strand form did not increased the $T_M$ in the conditions assayed without metal in the buffer.

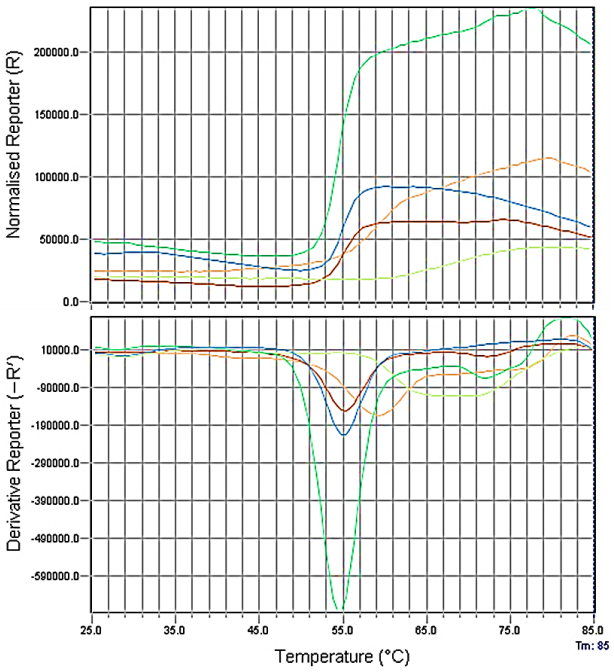



***Figure 50.*** *Melting curve exemplifying the ThermoFuor assay. A) Melting curve showing fluorescence at increasing temperatures. B) First derivative of the fluorescence (dF/dT) which maximum indicates the $T_M$. Green: ssDNA 45 nt; Blue: dsDNA 45 bp; Red: $FeCl_3$; Pink: $CuCl_2$ and Yellow: $CoCl_2$.*

According to the results obtained, we decided to coexpress ArdC for crystallization with 0.1 mM $NiCl_2$, 1 mM $CoCl_2$ or 1 mM $MnCl_2$ in all purifying buffers instead of EDTA as described in Section 3.2.3.1. Concentrations were chosen based on the titrations shown in **Figure 51** picking the highest metal concentration before stability of ArdC started to decrease.

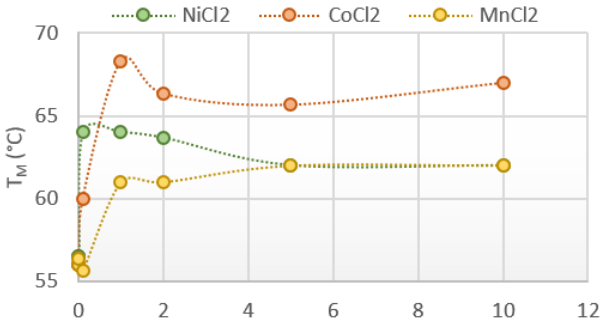



***Figure 51.*** *$T_M$ of ArdC in the presence of $NiCl_2$, $CoCl_2$ and $MnCl_2$ at different metal concentrations. Concentrations assayed are 0, 0.01, 0.1, 1, 2, 5 and 10 mM.*

### 4.3.4.2 Metalloprotease active center

ArdC, due to its C-terminal domain, is a member of the MEROPS peptidase database of the clan MA (Rawlings et al., 2014) identified by the large motif Xaa-Xbb-Xcc-His-Glu-Xbb-Xbb-His-Xbb-Xdd in which Xaa is an hydrophobic or Thr amino acid, Xbb are uncharged, Xcc is any residue except Pro, and Xdd is an hydrophobic amino acid (Jongeneel et al., 1989, Barrett & Rawlings, 1995). Thermolysin is the best characterized member of this clan (Matthews, 1988). However, the sequence for ArdC is TLAHEATHWT which fits in the definition except for the last Thr residue, not being a Xdd hydrophobic amino acid, but it is still classified as a metalloprotease. ArdC metalloprotease active site contains a zincin metal binding motif (HEXXH) of type gluzinzin, classified as MA(E) as the zinc ion is coordinated by the two histidines of the HEXXH consensus sequence through their Nε atoms located on the "active site helix" (α5) and a glutamic acid of the (E,H)XX(A,F,T,S,G) motif located in the contiguous α-helix or "glutamate helix" (α7) (Cerda-Costa and Gomis-Ruth, 2013). Both catalytic helices are joined by a long loop containing another

small α-helix (α6) that partially covers the entrance to the catalytic center and we think it could be involved in the target recognition or selection by untwisting the turn found between α5 and α6. See **Figure 41**.

Although being ArdC a metalloprotein, neither the Se-Met derivative nor the WT protein structure solved were observed coordinating the metal cofactor. To assure the ability and mode of ArdC to bind metals, a WT protein crystal obtained at 15 mg/mL in condition E3 (25 % v/v ethylene glycol) was soaked in a 10 mM $ZnSO_4$ solution for approximately 1 minute while cryoprotecting with an additional 10 % ethylene glycol before freezing for data collection at synchrotron (Figure 52A) (Table 30). Electronic density was found in the expected metal binding site. The side chain of E229 was differentially oriented in the presence or absence of metal (), indicating a correct metal binding disposition. In Figure 53 we can observe H201, H205, E229 tetrahedrally coordinating the metal, in addition to E202 and Y225 catalytic residues. The overall structure organization of the bound and unbound to zinc structure was not altered. The Cα alignment gives an r.m.s.d. value of 0.73 Å (on 274 aligned residues and 100 % sequence identity).

By the strategy of co-purifying ArdC with $MnCl_2$ in all buffers, we crystallized ArdC at 12 mg/mL bound to $Mn^{2+}$ in the same condition as for soaking with $Zn^{2+}$ (25 % v/v ethylene glycol). Crystals were cryoprotected with an additional 15 % glycerol (Figure 52B) (Table 30). In this case, the structure was solved at P32 with 8 subunits per asymmetric unit in a 4 dimers-like disposition (**Figure 54**).

Structure was solved at a higher resolution, 2.7 Å (Figure 55). We can observe how $Mn^{2+}$ is tetrahedrally coordinated by H201, H205, E229 and a $H_2O$ molecule. H205 is oriented towards the metal by the interaction with the well conserved E228 (**Figure 42**) through the other nitrogen atom. The E202 glutamic acid of the HEXXH motif orients and acts as a catalytic base for the activation of the water molecule that coordinates the metal. The $H_2O$ molecule

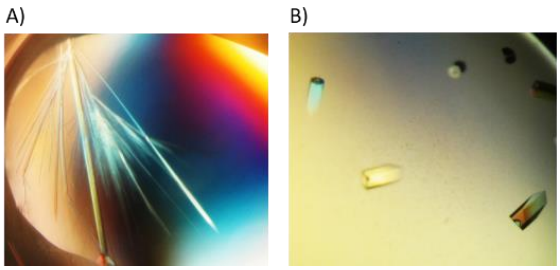



***Figure 52.*** *ArdC protein crystals for ArdC-metal structures. A) Native protein crystals used for soaking with 10 mM $ZnSO_4$. B) ArdC crystals of protein co-purified with $MnCl_2$.*

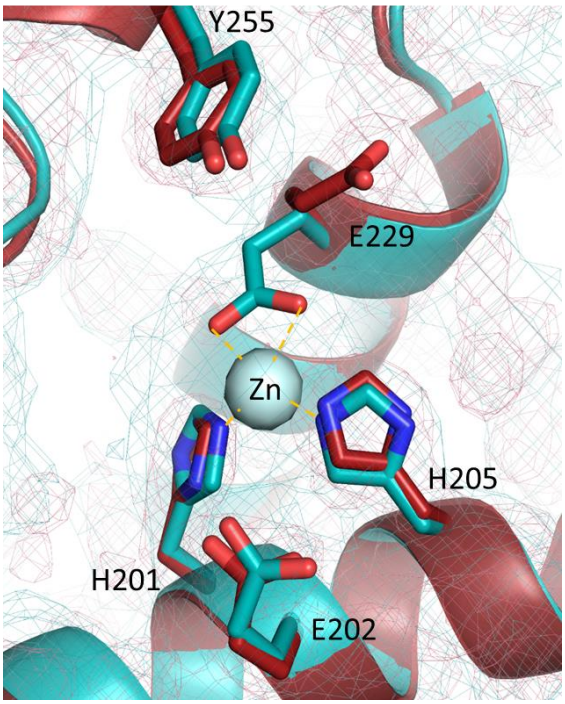



***Figure 53.*** *Metal binding site of ArdC. Apo form is shown in maroon and zinc-bound form in blue. Electron density map is also shown.*

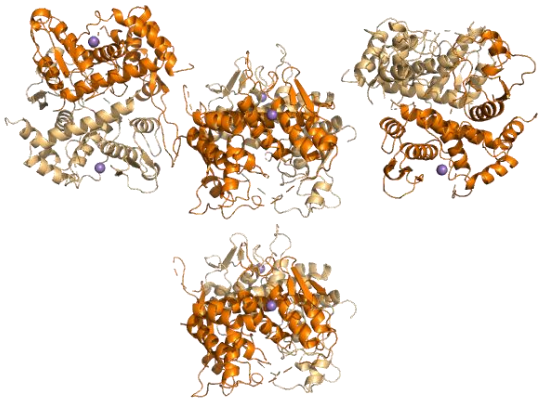


***Figure 54.*** *ArdC-Mn asymmetric unit. Structure shows 8 molecules per unit cell in a 4 dimer-like disposition.*

acts as a Lewis acid to allow the nucleophilic attack (Cerda-Costa and Gomis-Ruth, 2013). Y225 stabilizes by a hydrogen bond the polypeptide chain to be cleaved (Matthews, 1988). The overall sequence of the catalytic site of ArdC is thus $H^{201}EATH^{205}$-$X_{23}$- $E^{229}LIA$.

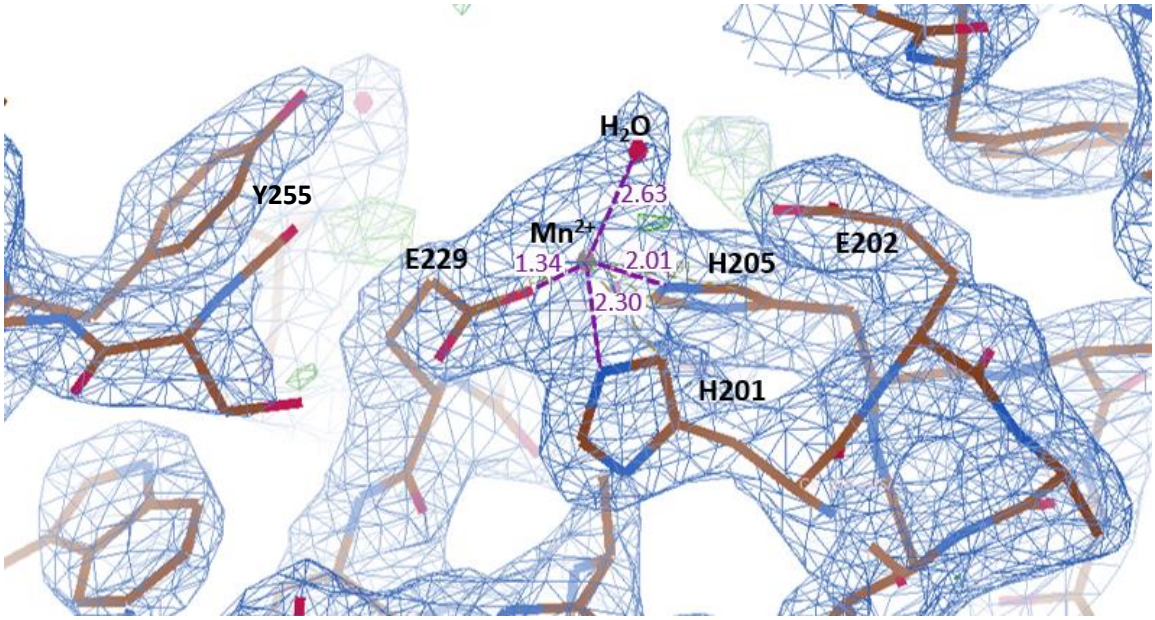



***Figure 55**. Metal binding site of ArdC bound to $Mn^{2+}$ at 2.7 Å resolution. Residues and molecules involved in metal coordination (distance in Å shown in purple) or activity are labelled.*

The Cα alignment gives an r.m.s.d. value of 0.60 Å (on 243 aligned residues and 98 % sequence identity). The structure with $Mn^{2+}$ displays a difference in the join between α5 and α6. ArdC-Mn do not have an α4 and the disposition of the loop could be affecting the access to the active center.  The other difference with the Apo and bound to Zn structure is the orientation of the tip of the β-hairpin (

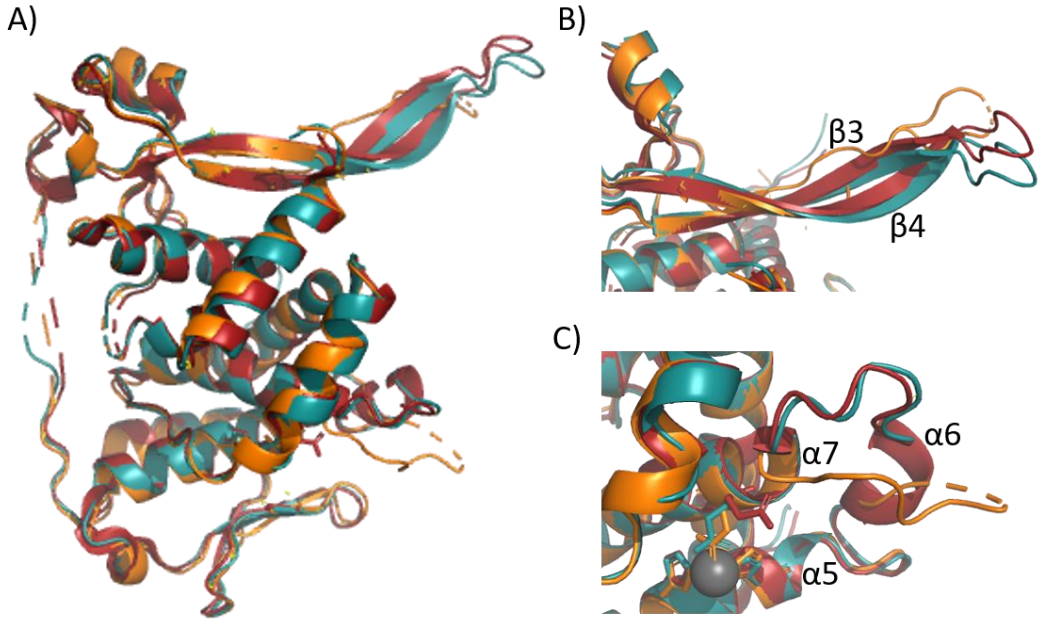



***Figure 56.** Structural alignment of ArdC unbound (in maroon), bound to Zn (in teal) or Mn (orange). A) Global alignment. B) Closer view of the differences in the β3 to β4 motifs. C) Closer view of the differences in the α5 to α7 region.*

).

***Table 30.*** Data collection and refinement statistics for ArdC-Zn and ArdC-Mn structures. Statistics for the highest-resolution shell are shown in parentheses.

| | ArdC-Zn | ArdC-Mn |
|---|---|---|
| **Wavelength** | 0.9792 | 0.9792 |
| **Resolution range** | 33.33-3.15 (3.263-3.15) | 54.82-2.7 (2.797-2.7) |
| **Space group** | R 3 :H | P 32 |
| **Unit cell** | 133.329 133.329 56.9315 90 90 120 | 116.499 116.499 162.123 90 90 120 |
| **Total reflections** | 338565 (25689) | 1158240 (114678) |
| **Unique reflections** | 6522 (583) | 66345 (6488) |
| **Multiplicity** | 51.9 (40.0) | 17.5 (17.4) |
| **Completeness (%)** | 96.00 (100.00) | 96.65 (96.95) |
| **Mean I/sigma(I)** | 44.82 (1.75) | 25.69 (2.57) |
| **Wilson B-factor** | 92.89 | 47.81 |
| **R-merge** | 0.7812 (3.213) | 0.6992 (1.47) |
| **R-meas** | 0.7891 (3.255) | 0.7198 (1.514) |
| **CC1/2** | 0.946 (0.638) | 0.691 (0.572) |
| **CC*** | 0.986 (0.883) | 0.904 (0.853) |
| **Reflections used in refinement** | 6283 (583) | 65280 (6491) |
| **Reflections used for R-free** | 305 (36) | 3008 (320) |
| **R-work** | 0.2043 (0.3065) | 0.2562 (0.3153) |
| **R-free** | 0.2369 (0.3788) | 0.3244 (0.3632) |
| **CC(work)** | 0.780 (0.431) | 0.848 (0.640) |
| **CC(free)** | 0.705 (0.539) | 0.819 (0.491) |
| **Number of non-hydrogen atoms** | 2226 | 16236 |
| **Protein residues** | 279 | 2046 |
| **RMS(bonds)** | 0.010 | 0.012 |
| **RMS(angles)** | 1.30 | 1.58 |
| **Ramachandran favored (%)** | 84 | 85.05 |
| **Ramachandran allowed (%)** | 14 | 11.34 |
| **Ramachandran outliers (%)** | 2.2 | 3.61 |
| **Rotamer outliers (%)** | 5.9 | 0.97 |
| **Clashscore** | 18.52 | 16.30 |
| **Average B-factor** | 82.68 | 51.60 |

### 4.3.5 Transcriptional DNA damage repair activator and human genome instability protector protein as ArdC C-terminal closest structural homologues

The analysis with DALI server for ArdC C-terminal (amino acids 151-297) domain to identify the closest structural homologues revealed a considerable structural similarity to other metalloproteases (Table 27).

On one hand, ArdC-C showed similarity with human Spartan protein. Spartan is a protein that cleaves DNA-proteins crosslinks in order to preserve genome stability. These crosslinks are generated by UV light, reactive aldehydes or ionizing radiation or by a stable blockage of enzyme-DNA covalent intermediates (Li et al., 2019). ArdC shares structural features with the

SprT N-terminal domain of Spartan protein which is composed by a $Zn^{2+}$- binding sub-domain (ZBD) that binds ssDNA and a metalloprotease sub-domain (MPD) stimulated by ssDNA. Amino acids 151-235 of ArdC showed a 12 % sequence identity and 57 % coverage with SprT MPD (residues 43-145). See Figure 58. Superposition of the C-terminal domain Cα of ArdC-Zn (aa 151-235) to the MPD of SprT (aa 43-145) gave an alignment r.m.s.d. value of 2.11 Å (on 64 aligned residues) (Figure 57). MPD shares active center structure with C-terminal domain of ArdC except for that MPD uses a third histidine (His 130) instead of a glutamic acid for metal coordination. The 3 β-strands perfectly align in both structures, however, ArdC α7-helix is displaced toward the metal in comparison with the equivalent α-helix in MPD. This MPD α-helix, is the last one of the sub-domain before the polypeptide chain continues towards the ZBD sub-domain required for DNA binding. To be mentioned, between the two catalytic α5 and α7-helixes, SprT domain has a smaller loop than ArdC.

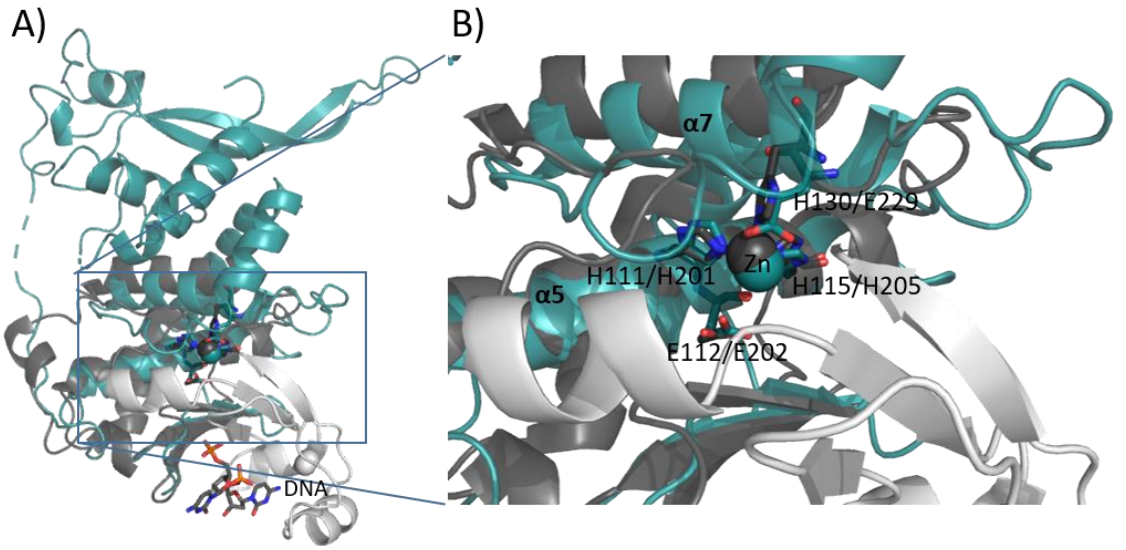



***Figure 57.*** *Structural alignment of ArdC and SprT. A) Superposition of ArdC-Zn structure (in teal, 3.15 Å) with its closest structural homologue Spartan SprT domain (6MDX; in grey, 1.55 Å). Zn2+-binding sub-domain (ZBD) is shown in light grey and metalloprotease sub-domain (MPD) is shown in dark grey. B) A detailed view of the metalloprotease active center with the residues involved in catalysis in sticks numbered as (MPD/ArdC-C).*

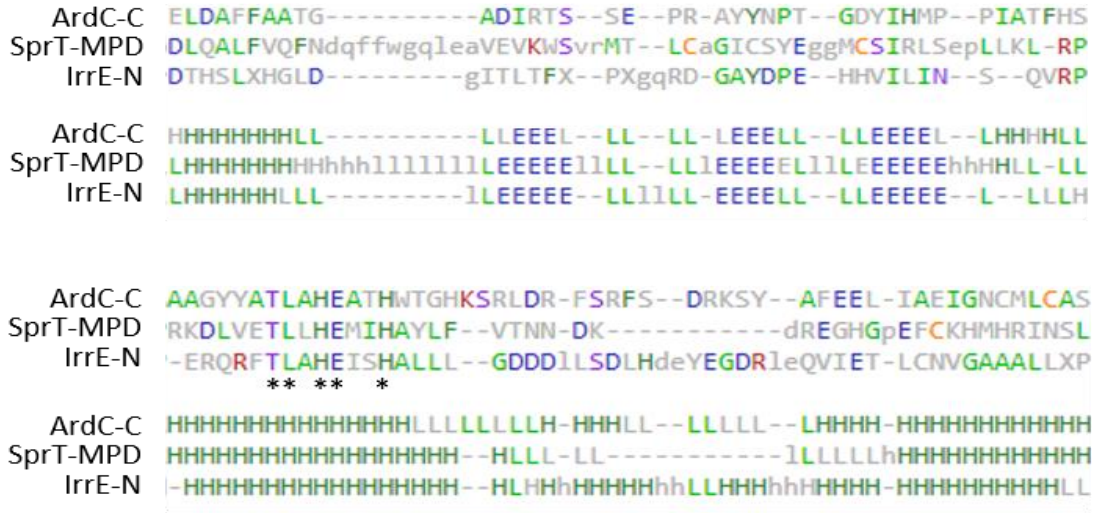



***Figure 58****. Protein sequence alignment of ArdC-C, SprT-MPD and IrrE-N. In the first three lines, the sequence alignment is shown, in the second three lines, the secondary structure assignments (H stands for helix, E for strand, and L for coil) are shown. The most frequent amino acid type is coloured for each position. Conserved positions are marked with an *.*

On the other hand, ArdC-C showed similarity to the proteolytic and part of the HTH domains of PprI protein, also named IrrE from *Deinococcus radiodurans* (PDB: 3DTE). IrrE protects *D. radiodurans* from UV radiation DNA damage by the proteolysis of a transcriptional regulator involved in SOS response. Amino acids 154-277 of ArdC showed an 18 % identity and 75 % coverage with residues 34-150 of IrrE/PprI. See **Figure 58**. Superposition of the C-terminal domain Cα of Zn-ArdC (aa 154-277) to the N-terminal domain Cα of Zn- Irre/PprI (aa 34-150) gave an alignment r.m.s.d. value of 2.46 Å (on 79 aligned residues) which is acceptable given the high resolution of both structures (Figure 59). PprI is a formed by three domains, a zinc peptidase-like domain, a dsDNA-binding helix-turn-helix motif and a GAF-like sensor domain. The N-terminal zinc peptidase-like domain shares structure with C-terminal domain of ArdC except for that Irre lacks the α8 and the $3_{10}$ helix between the β-strands and the α-helixes. On the other hand, Irre has an α-helix instead of the loop that partially covers the entrance to the ArdC active center. And, interestingly, the disposition of the first motif in Irre is opposed in direction, as Irre N-term α-helix aligns with the last C-term α9-helix of ArdC. IrrE polypeptide chain then follows towards ArdC α4 and both chains continue in the same direction. By the cleavage of the transcriptional regulator DdrO in the presence of $Mn^{2+}$, PprI/Irre plays a central regulatory role in the DNA protection or repair pathways in response to radiation (Wang et al., 2015).

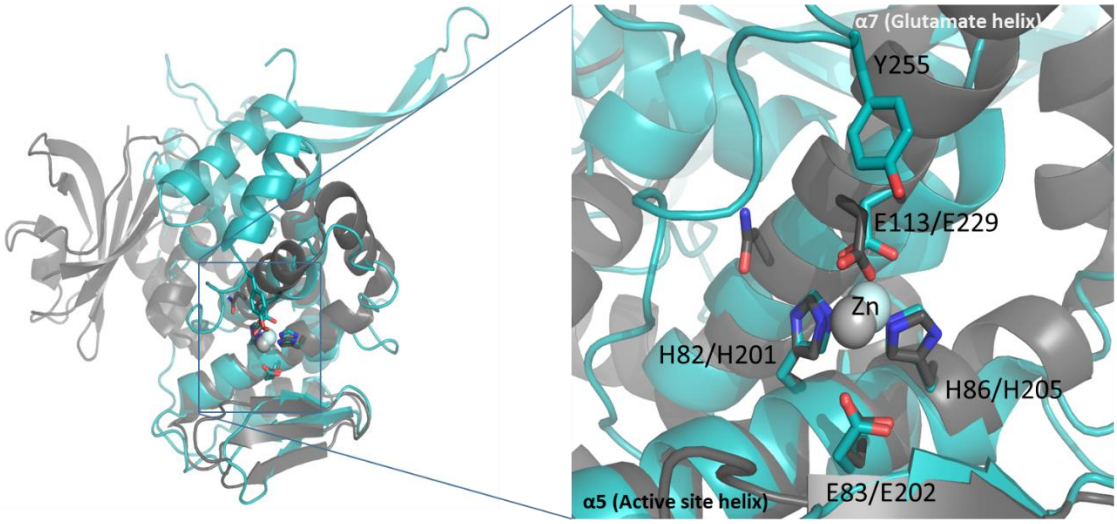



***Figure 59.*** *Structural alignment of ArdC and IrrE. A) Superposition of ArdC-Zn structure (in teal, 3.15 Å) with its close structural homologue Irre-Zn protein from Deinococcus radiodurans (3DTI; in grey, 3,5 Å). B) A detailed view of the active center with the residues involved in catalysis in sticks numbered as (IrrE/ArdC). Superposition generated by Phenix. superpose_maps tool.*

### 4.3.6 Identification of a putative ArdC target by homology

A BLASTP search was done to identify DdrO (from *Deinococcus radiodurans)* homologs in R388 putative receptor strains. DdrO is a proteins transcriptional repressor of the helix-turn-helix XRE-family. The closest homologue found in *P. putida KT2440* was PP_2868, identified as DNA-binding transcriptional repressor PuuR, a member of the Cro/CI family transcriptional regulators (33 % identity and 44 % query cover). The second closest homologue in *P. putida KT2440* is PP_2177 (27 % identity and 69 % query cover). It is predicted to be a helix-turn-helix XRE family transcriptional regulator. There are no closely related transcriptional regulators (low % of identity) of DdrO in *P. Putida*, thus, ArdC is probably having a different mode of action than IrrE.

## 4.4 ArdC functional characterization

### 4.4.1 Analysis of ArdC oligomeric state

S75 gel-filtration chromatography showed that ArdC-his eluted at a calculated molecular mass of 47.2 KDa. ArdC-his theoretical mass is 34.3 KDa corresponding the elution to about 1.3 molecules (Figure 60). Further analysis of ArdC oligomeric state was done by glutaraldehyde crosslinking as described in materials and methods. A single band corresponding to around 35 KDa was observed without glutaraldehyde, but also when increasing the glutaraldehyde concentration (**Figure 61**). Only at very high glutaraldehyde concentrations, bigger oligomeric structures seem to be observed, but none of them prevails. This result agrees with the one obtained with ArdC crystal structure using the PISA server, which expects ArdC to be a monomer, as no quaternary functional structure is predicted.

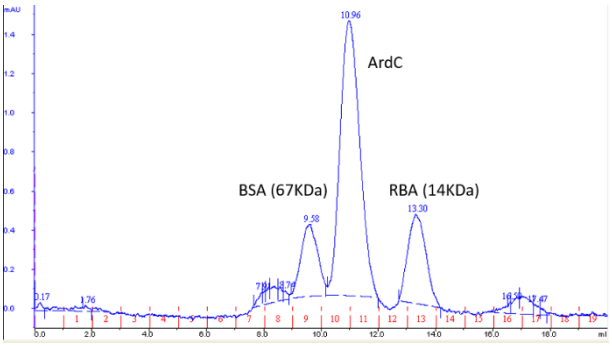



*Figure 60. S75 gel chromatography of ArdC eluting at a MW between BSA and RBA.*

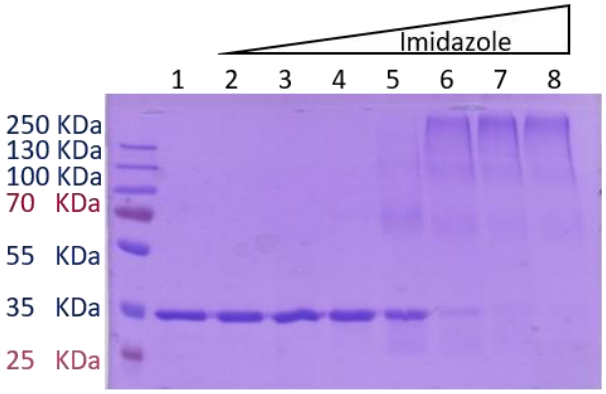



***Figure 61.*** *Oligomeric state of ArdC analysed by glutaraldehyde crosslinking. Native protein cross-linked with increased concentrations of glutaraldehyde on a SDS-PAGE gel. Imidazol concentration by lane: 1: 0 µM; 2: 0.5 µM; 3: 5 µM; 4: 50 µM; 5: 500 µM; 6: 5 mM; 7: 50 mM; 8: 500 mM.*

In order to check if a multimer is formed when bound to DNA, a glutaraldehyde crosslinking assay was also done in the presence of ssDNA at different protein: ssDNA ratios (Figure 62A and B). 45 nt oligonucleotide "T87I2" was boiled for 5 min and cooled down fast in phosphate buffer and incubated with ArdC for 10 min at RT before the addition of glutaraldehyde. We neither observed any oligomerization of ArdC bound to ssDNA (apart from the faint bands at very high glutaraldehyde concentrations).

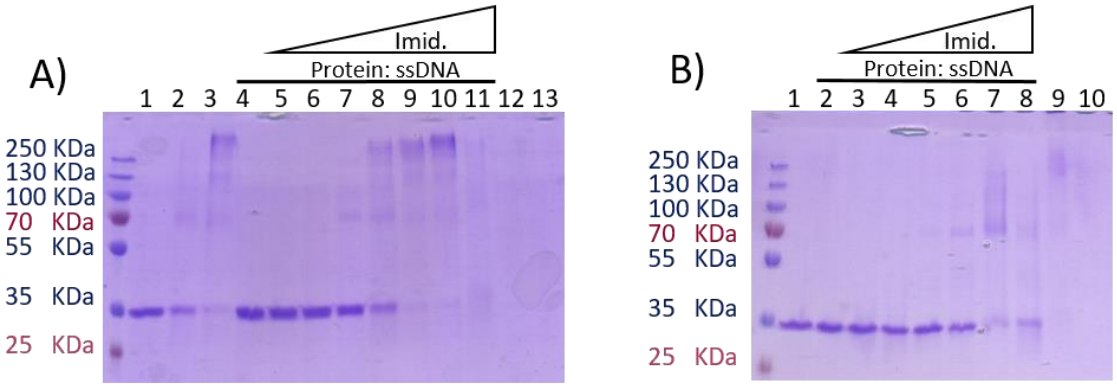



*Figure 62. Oligomeric state of ArdC in complex with ssDNA analyzed by glutaraldehyde crosslinking on a SDS-PAGE gel. A) Protein: ssDNA complex in a 10:1 ratio. Lane 1: ArdC; Lane 2: ArdC in 500 µM imidazole; Lane 3: ArdC in 5 mM Imidazole; Lane 4-11: ArdC + 45 nt oligonucleotide "T87I2" at increasing imidazole concentrations: 0, 0.5 µM, 5 µM,*

*50 µM, 500 µM, 5 mM, 50 mM and 500 mM; Lane 12: 500mM imidazole; Lane 13: 45 nt oligonucleotide "T87I2" in 500mM imidazole. B) Protein: ssDNA complex in a 1:1 ratio. Lane 1: ArdC; Lane 2-8: ArdC + 45 nt oligonucleotide "T87I2" at increasing imidazole concentrations: 0.5 µM, 5 µM, 50 µM, 500 µM, 5 mM, 50 mM and 500 mM; Lane 9: 500mM imidazole; Lane 10: 45 nt oligonucleotide "T87I2" in 500 mM imidazole.*

### 4.4.2 ArdC preferentially binds ssDNA

In order to check the DNA binding ability of ArdC, previously analyzed by (Belogurov et al., 2000), we performed electrophoretic mobility gel assays (EMSAs) under non-denaturing conditions for short DNA in a native polyacrylamide gel. **Figure 63** shows that ssDNA is preferentially bound and retarded over dsDNA. Even though the 45 bp dsDNA complex is not completely formed (**Figure 63** lane 6); we can observe that dsDNA is retarded as well but with less affinity than ssDNA. Retardation was also observed with other ssDNA oligonucleotides (20 nt, 25 nt, 30 nt, 35 nt, 40 nt and 50 nt) but binding was weaker for shorter oligonucleotides probably because less protein molecules could be bound to each DNA molecule (Supplementary Figure 2). Binding is not dependent on the DNA sequence, as random oligonucleotides were chosen for the assay.

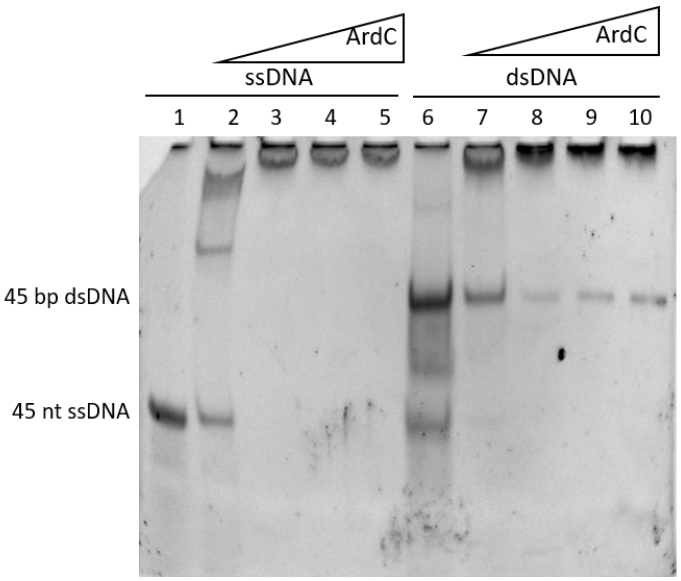



***Figure 63.*** *ArdC DNA-binding analysed by EMSA. 10 % polyacrylamide gel showing the retardation of short ssDNA (oligonucleotide of 45 bases) or short dsDNA (two oligonucleotides of 45 bases) by ArdC under non-denaturing conditions. Lane 1: ssDNA; Lane2-5: ssDNA (0.3 µM) + ArdC at increasing concentrations (2.35 µM (2), 4.70 µM (3), 7.05 µM (4) and 9.4 µM (5)); Lane 6: dsDNA (0.3 µM); Lane 7-10: dsDNA+ ArdC at increasing concentrations (2.35 µM (7), 4.70 µM (8), 7.05 µM (9) and 9.4 µM (10)).*

### 4.4.3 ArdC protects ssDNA from degradation by type II endonuclease

(Belogurov et al., 2000) showed that ArdC protects ssDNA but not dsDNA from degradation by HhaI, a type II endonuclease able to digest both ssDNA and dsDNA. They did it with pBluescriptII that has 24 HhaI recognition sites instead of oligonucleotides. To check the ability of ArdC to bind long DNA, we used ssDNA from M13mp18 (7.2 kb), which has 26 HhaI restriction sites, in the retardation and protection assay. The vast majority of the molecules of M13mp18 ssDNA are circular, although some of them are present in the linear form, as determined by agarose gel electrophoresis.

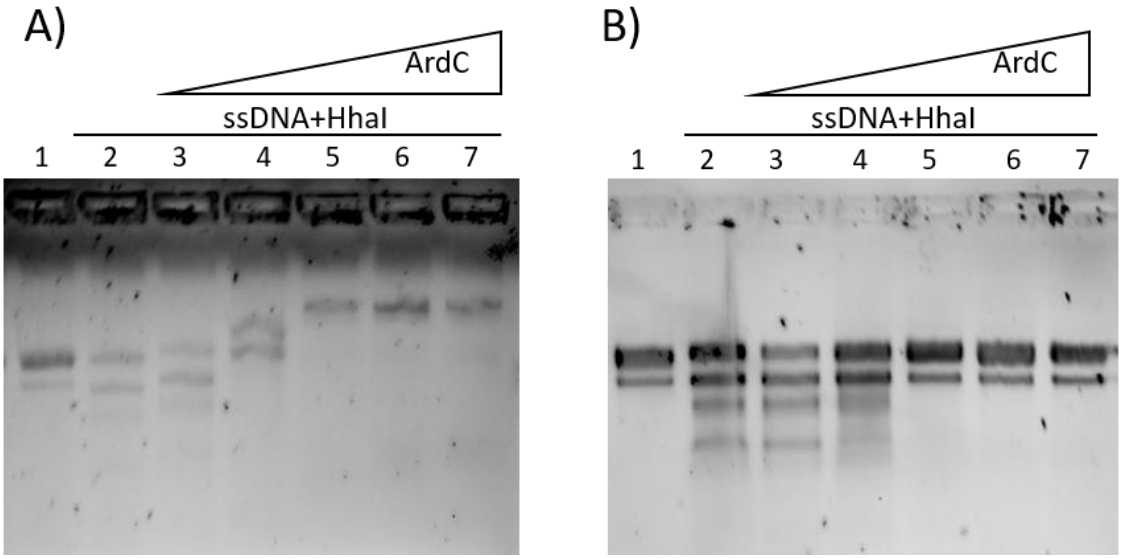



***Figure 64.*** *Agarose gel showing the retardation and protection of long ssDNA (M13mp18) by ArdC. A) ssDNA retardation under non-denaturing conditions. B) ssDNA protection from HhaI proteolysis under denaturing conditions (proteinase K and SDS added). Lane 1: ssDNA (5.5 nM); Lane 2: ssDNA and HhaI (7 U), Lane 3-7: ssDNA and HhaI at increasing concentrations of ArdC (0.95 µM (3), 1.9 µM (4), 3.8 µM (5), 5.7 µM (6) and 7.8 µM (7)).*

In Figure 64A we can observe how M13mp18 ssDNA (upper band in circular shape and lower band in linear form) is cleaved by HhaI (lane 2). When ArdC is added to the reaction mixture, ssDNA is bound and retarded by ArdC in lanes 5 to 7 in spite of HhaI being present. Aliquots of these samples were also treated with proteinase K and SDS before loading to the gel to degrade the proteins and better observe the state of the ssDNA. We can observe how ssDNA is cleaved by HhaI in lane 2, being protected from HhaI digestion in lanes 5 to 7 where higher amounts of ArdC were added (Figure 64B). A high molar ratio of Protein: DNA (700 to 1400 in lanes 5 to 7) is needed to protect ssDNA from HhaI. This may indicate that ssDNA needs to be well covered by ArdC in order to prevent the access of HhaI to the DNA.

### 4.4.4 ArdC do not protect DNA from UV damage

Irre, Rad4 and Spartan proteins are all involved in the repair of DNA lesions caused by UV light. UV sensibility assays were performed in order to determine if ArdC could be able to protect from DNA damage by UV light. First, we tested the effect by introducing pSU2007, pIC10 or pLGM25 plasmids on *P. putida KT2440* and expose these strains to the UV light for different times as detailed in materials and methods. Although it was a semi quantitative analysis, similar numbers of cells survived independently of the absence or presence of *ardC* in the strain (Figure 65). ArdC does not seem to be involved in SOS repair when DNA is found in single strand shape after DNA damage by UV.

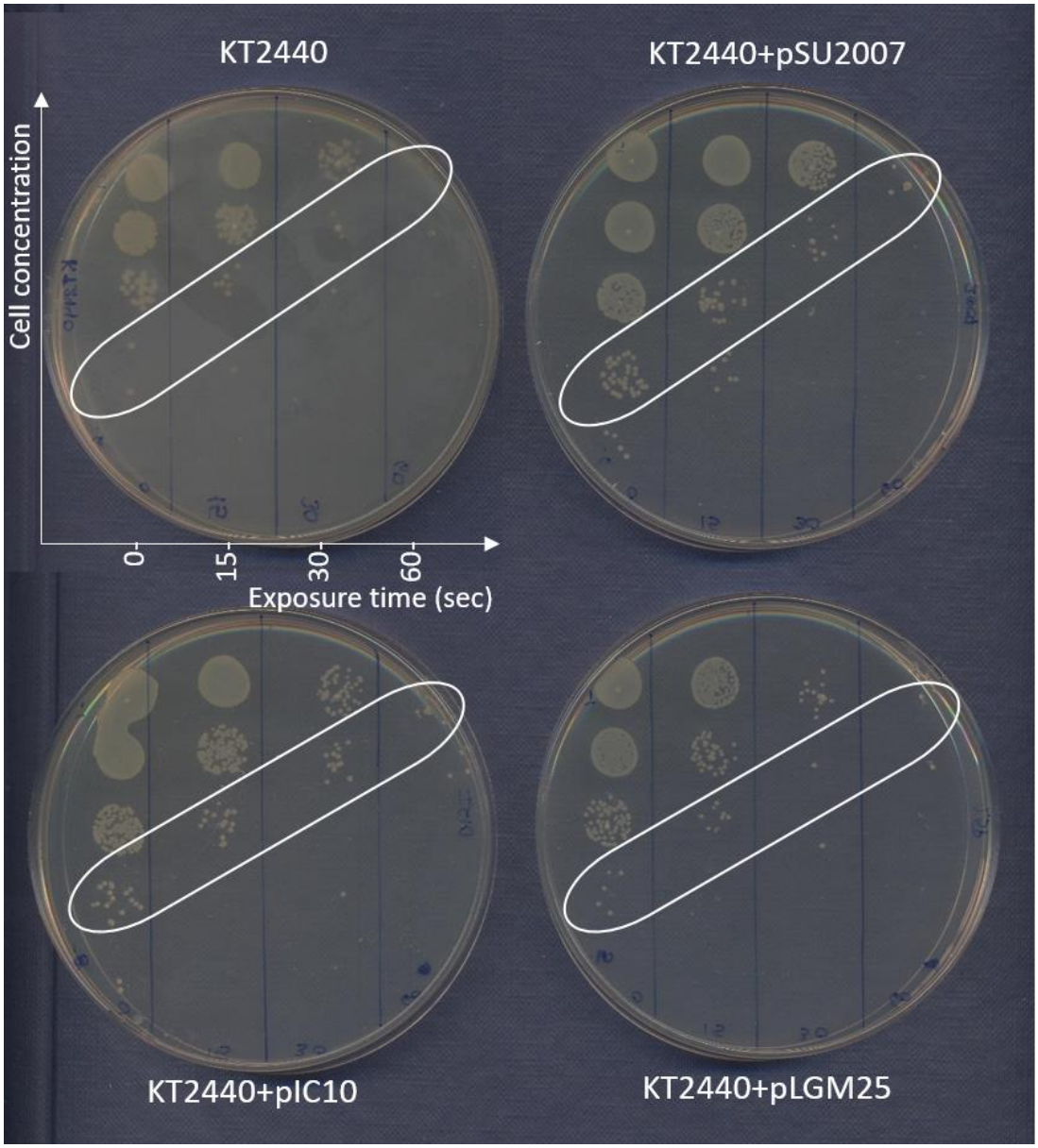



***Figure 65**. UV sensibility assay. The sensitivity of P. Putida KT2440 empty or carrying PSU2007, pIC10 or pLGM25 was analyzed at different cell concentrations after the exposure to 0, 15, 30 or 60 sec of UV (302 nm) and o/n incubation at 30 ºC protected from light.*

Then, we decided to couple the UV sensibility assay to conjugation, to check if ArdC is able to protect the incoming ssDNA from UV DNA damage. We exposed the mating mixture of donors and recipients to UV at time 0 (no conjugation) and after 30 min of conjugation as described in materials and methods.

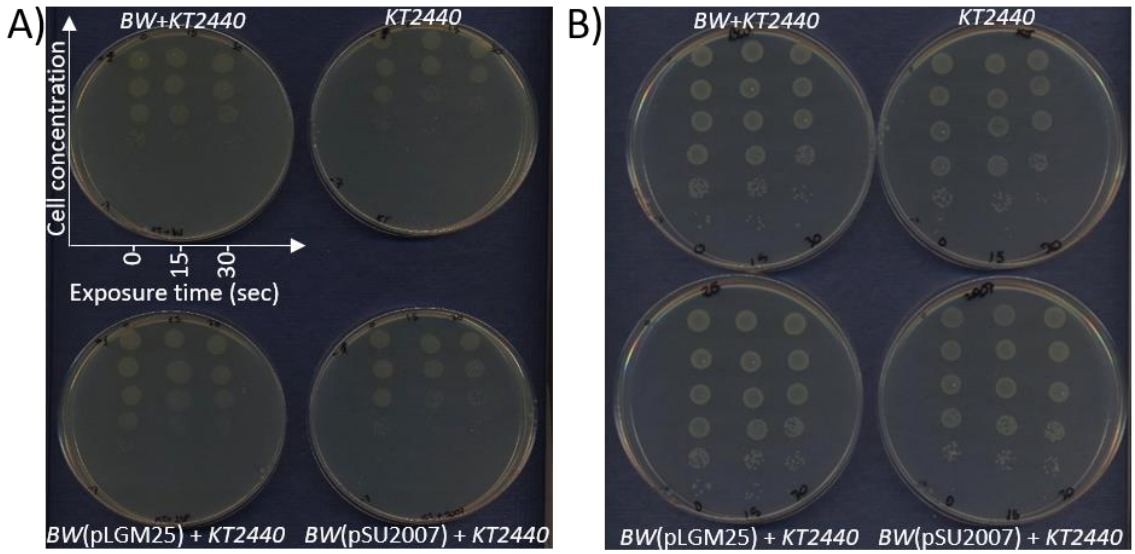



***Figure 66.*** *UV sensibility assay coupled to conjugation. A) UV applied at t=0. B) UV applied at t=30 min. Filter plates with the conjugation mixture were exposed to UV while conjugation for 1 h at 37 ºC was being performed. Plates from left to right and from first row to second row: Empty E. coli BW27783 + P. putida KT2440, Empty P. putida KT2440, E. coli BW27783 + pLGM25 to P. putida KT2440 and E. coli BW27783 + pSU2007 to P. putida KT2440. The first column of drops was exposed to 0 sec of UV, the second line to 15 sec of UV and he third column to 30 sec of UV. From top to bottom, serial dilutions from -1 to -7 were plated.*

Again, no differences in survival rates were observed in *P. putida* alone, *P. putida* with *E. coli*, *P. putida* with *E. coli* + pLGM25 or *P. putida* with *E. coli* + pSU2007 treated with UV light at the moment of mixing donors and recipient cells (Figure 66A) or 30 minutes after (Figure 66B). Thus, ArdC does not seem to be involved in the SOS repair pathway when DNA is found in single strand coupled to conjugation either.

### 4.4.5 Identification of putative ArdC partners by proteolytic activity

As preliminary putative targets of ArdC protease domain we thought about ArdK, for being one of its transcriptional regulators, RecA, for being a DNA binding protein and due to the similarity with Irre-DdrO system, and HhaI based on (Belogurov et al., 2000) results.

First, we assayed the interaction of ArdC with HhaI in the presence of 45 nt/bp ssDNA and dsDNA, as HhaI is able to bind both DNA forms. We tested $Co^{2+}$ as putative cofactor according to the stabilization of ArdC previously determined (Section 4.3.4.1). However, no degradation of HhaI by ArdC was observed (Figure 67 A). Then, we tried another assay with long M13 ssDNA and $MgCl_2$ at the conditions we had seen long ssDNA binding by ArdC and protection from degradation by HhaI (Figure 64) in the presence of $Mg^{2+}$. Again no proteolysis of HhaI was observed at the conditions tested (Figure 67 B).

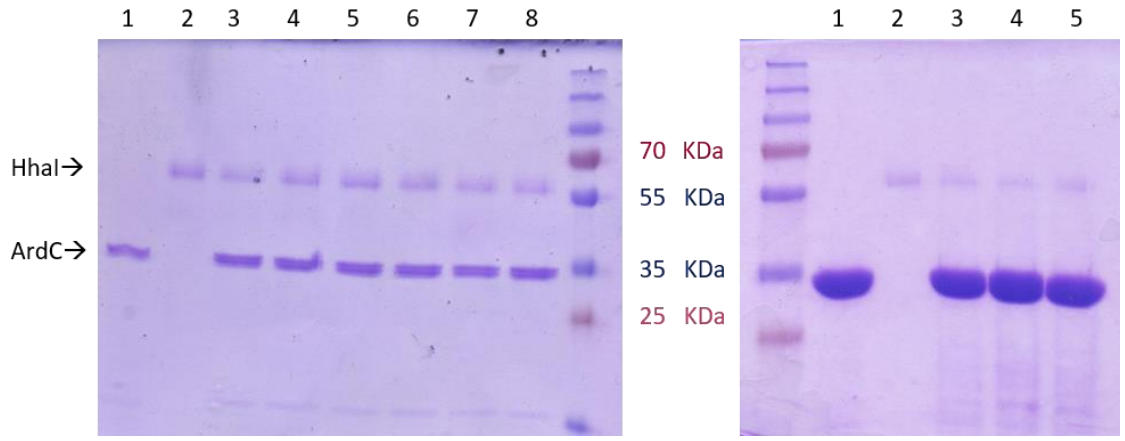



***Figure 67.*** *Proteolytic assay of ArdC to HhaI in the presence of different DNA forms and metal cofactors. A) Preincubation of ArdC with short ssDNA or dsDNA. Lane 1: ArdC; Lane 2: HhaI; Lane 3-4: HhaI and ArdC bound to ssDNA (45 nt) without (3) or with $CoCl_2$ (4); Lane 4-5: HhaI and ArdC bound to dsDNA (45 bp) without (5) or with $CoCl_2$ (6); Lane 7-8: HhaI and ArdC without (7) or with $CoCl_2$ (8). B) Preincubation of ArdC with long M13 ssDNA. Lane 1: ArdC (8 µM) in the presence of $MgCl_2$; Lane 2: HhaI (70 U) in the presence of $MgCl_2$; Lane 3: ArdC and HhaI in the presence of $MgCl_2$; Lane 4-5: ArdC and HhaI in the presence of long ssDNA (M13, 27.5 nM) with $MgCl_2$ (4) or EDTA (5).*

ArdC and other genes involved in the establishment of R388 in the receptor cell are regulated by the transcriptional repressor ArdK. We thought that maybe ArdK could be the transcriptional regulator cleaved by ArdC. Thus, we have purified ArdK as described in M&M and performed a proteolytic activity assay in the presence of ArdK with or without ssDNA (**Figure 68**). We have also tried different putative metal cofactors such as $Ni^{2+}$, $Co^{2+}$ or $Mn^{2+}$. Nevertheless, we did not observe ArdK proteolysis in any case.

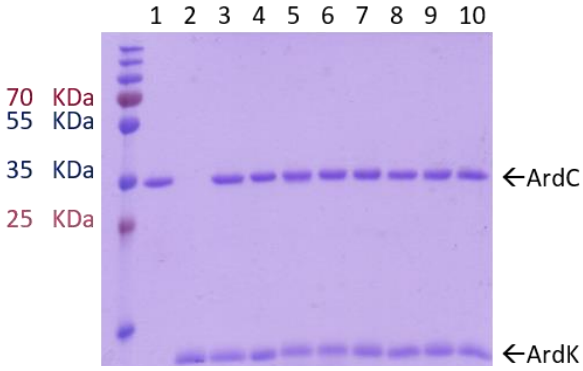



***Figure 68.*** *Proteolytic assay of ArdC and ArdK in buffer containing different metal cofactors in the presence or absence of ssDNA (45 bases oligonucleotide). Lane 1: ArdC; Lane 2: ArdK, Lane 3-6: ArdC and ArdK in the absence of metal (3) and in the presence of $NiCl_2$ (4), $MnCl_2$ (5) or $CoCl_2$ (6); Lane 7-10: ArdC and ArdK in the presence of ssDNA in the absence of metal (7) and in the presence of $NiCl_2$ (8), $MnCl_2$ (9) or $CoCl_2$ (10).*

Finally, we also tested RecA as target. RecA is the ssDNA binding protein responsible for

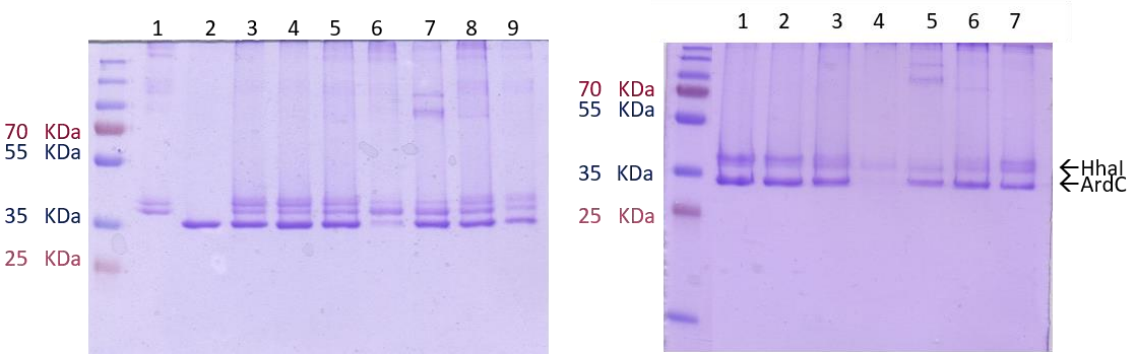



***Figure 69****. Proteolytic assay of ArdC to RecA in buffer containing different metal cofactors. A) Preincubation of RecA with ssDNA(45bases). Lane 1: RecA; Lane 2: ArdC, Lane 3: RecA and ArdC; Lane 4 – 9: ArdC and RecA with ssDNA in the presence of EDTA (4), $MgCl_2$ and $MnCl_2$ (5), $MgCl_2$, $MnCl_2$ and $CoCl_2$ (6), $MgCl_2$, $MnCl_2$ and $CuCl_2$ (7), $MgCl_2$, $MnCl_2$ and $NiCl_2$ (8), $MgCl_2$, $MnCl_2$ and $ZnCl_2$ (9). B) Preincubation of ArdC with ssDNA (45bases). Lane 1: RecA and ArdC; Lane 2-7: RecA and ArdC with ssDNA in the presence of EDTA (2), $MgCl_2$ and $MnCl_2$ (3), $MgCl_2$, $MnCl_2$ and $CoCl_2$ (4), $MgCl_2$, $MnCl_2$ and $CuCl_2$ (5), $MgCl_2$, $MnCl_2$ and $NiCl_2$ (6), $MgCl_2$, $MnCl_2$ and $ZnCl_2$ (7).*

DNA repair and SOS response in bacteria. For that, ArdC or RecA were preincubated with DNA in the presence of a broader range of metal cofactors (Figure 69). We could observe the disappearance of ArdC protein bands in the presence of $Co^{2+}$, but later on, we could associate this observance to the low pH of the buffer, which was degrading the protein.

In none of the cases and conditions assayed, we could observe a clear proteolytic cleavage, discarding for the moment ArdK, RecA and HhaI from being the target of ArdC.

### 4.4.6 Identification of putative ArdC partners by pull down assay

The pull-down assay is a biochemical technique used to detect interactions between two proteins. We have used this assay in order to identify a putative ArdC partner. As the target of a protease is freed after cleavage avoiding its identification by this kind of assays, we have purified ArdC E229A with a mutation in the conserved glutamic acid essential for the activity of HEXXH metalloproteases.

We first removed the proteins of the *P. putida KT2440* lysate that non-specifically bind to the HisTrap column and with the flow through we performed a pull down assay where ArdC eluted with no proteins bound. In order to check if the putative ArdC partner had been discarded between the non-specifically binding proteins, we performed another pull down assay with all the lysate without the washing step. After the pull down assay using ArdC as bait and the whole cell lysate of *P. putida* as prey, several bands were observed in the SDS-PAGE gel coeluting with ArdC (**Figure 70**).

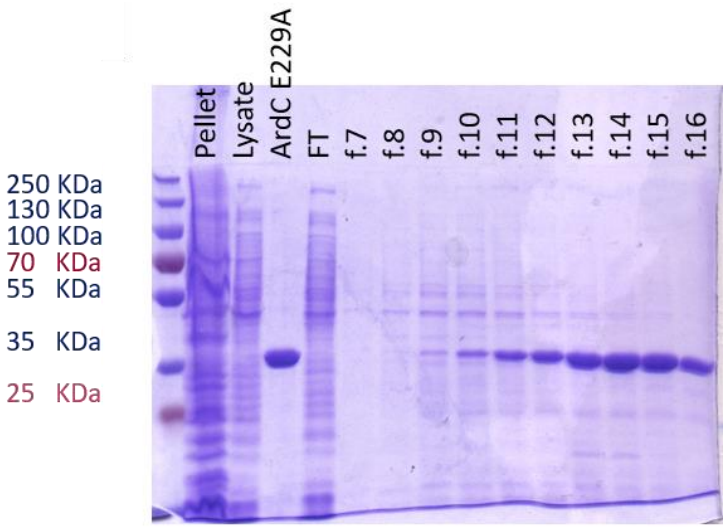



***Figure 70.*** *Pull down assay using ArdC as bait and the whole cell lysate of P. putida as prey. SDS-PAGE gel showing the bands of proteins coeluting with ArdC on a HisTrap column.*

In order to select the protein that eluted bound to ArdC, we performed an S200 size exclusion chromatography with fractions f.13 and f.14 and loaded on a SDS-PAGE gel the elution fractions were one major band is observed (Figure 71). Although this band was also obtained when the *P. putida* cell lysate was loaded onto the HisTrap column without ArdC (data not shown), we sent the band to mass spectrometry analysis (Figure 72).

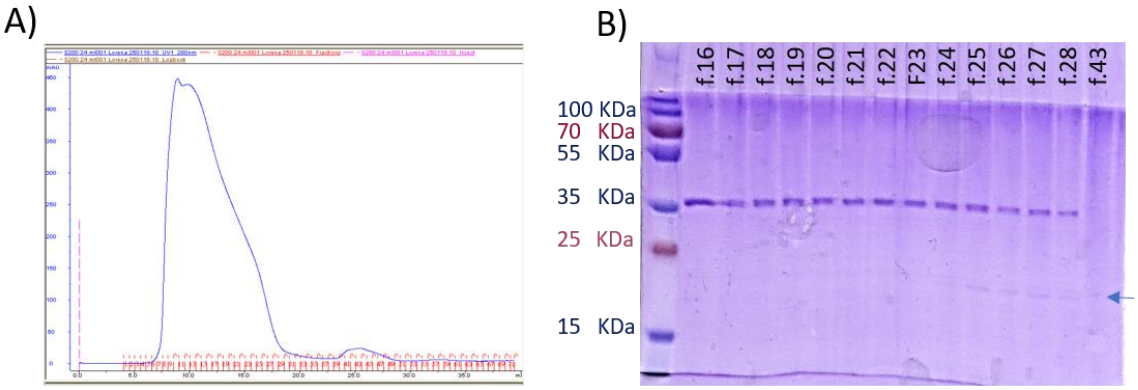



***Figure 71**. Analysis of pull-down fractions by size exclusion chromatography. A) Chromatogram of f.13 and 1.14 through a S200 size exclusion column. B) Fractions of the chromatography loaded on a SDS-PAGE gel. The rest of the fractions were also run but no other bands apart from ArdCE229A appeared (data not shown). Arrow shows the protein band of the potential ArdCE229A partner.*

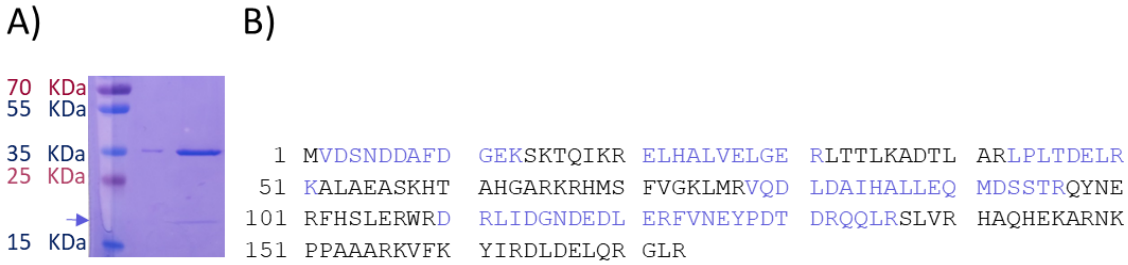



***Figure 72.** A) Protein from pull down experiment cleaved from gel and sent for mass spectrometry analysis. B) MASCOT search result for the peptides obtained after digestion with trypsin (that cuts C-term side of KR unless next residue is P) in SwissProt database for Pseudomonas putida KT2440. Matched peptides with pp_0941 protein are shown in purple. Protein sequence coverage of 45%. pp_0941 has 173 amino acids and a molecular weight of 20238 Da.*

It was found to be protein pp_0941 (with a coverage of 45 %), a protein with unknown function similar to the 50s ribosome subunit associated protein YjgA (Jiang et al., 2006). Unfortunately, we could not relate this protein with the predicted function of ArdC, and thus, this approach was not successful in providing us information about the ArdC target.

## 4.4.7 Conjugation

### 4.4.7.1 ArdC is not needed if recipient hsdRMS is depleted

In an attempt to have a clue about the target of the described ArdC metalloprotease domain we have used different *P. putida* mutant strains as recipients. Irre protects *Deinococcus radiodurans* from radiation by stimulation of *recA* transcription. Thus, first we tried to conjugate to *P. putida KT2440ΔrecA* to check the role of RecA in ArdC mediated conjugation. As shown in Figure 73, pLGM25 was transfer from *E. coli* to *KT2440ΔrecA* with a similar efficiency than to *KT2440* indicating that ArdC activity is not dependent on RecA. Then, we tried to conjugate to a derivative strain of *KT2440*, called *EM42*, which was constructed by deleting several genes (Δmix: *Δprophage1 Δprophage4 Δprophage3 Δprophage2 ΔTn7 ΔendA-1 ΔendA-2 ΔhsdRMS Δflagellum ΔTn4652*) that could harass the heterologous gene expression (because their association to genetic instability or attributed to the unfruitful usage of metabolic resources). Prophages are parasitic sequences inserted in the genome that make cells more susceptible to DNA damage and if active can cause cell lysis (Martínez-García et al., 2015). Transposons are

mobile DNA sequences that can move along the genome inserting themselves in a random target region and are known to be activated under stressful conditions (Kivistik et al., 2007; Peters, 2014). EndA-1 and EndA-2 are two type I deoxyribonucleases that degrade non-specifically dsDNA restricting the entrance of plasmids (Dubnau, 1999; Martínez-García et al., 2014). HsdRMS is a type I R-M system that protects the cell against invading DNA (See Section 1.3.2.3.1). Finally, the removal of the flagellum has been shown to increase the tolerance to stressful environmental conditions.

We thought that any of these genes removed from *P. putida* EM42 could avoid the establishment of the acquired plasmids in our conjugation system. Surprisingly, the conjugation frequencies towards this strain were not affected by the *ardC* deletion indicating us that the products of one or more of the deleted genes in the strain could be involved in the ArdC activity. We performed the conjugation experiment to strains containing single deletions of each of the genes deleted in EM42 and we could observe that *hsdRMS* was the main responsible for the effect observed in *EM42* (Figure 73). Due to the drop in the conjugation frequency of pSU2007 control plasmid towards the *P. putida* mutants lacking the transposons, we cannot quantify the effect of the ardC absence in these two strains. In all cases, the pSU2007 transconjugant colonies were smaller.

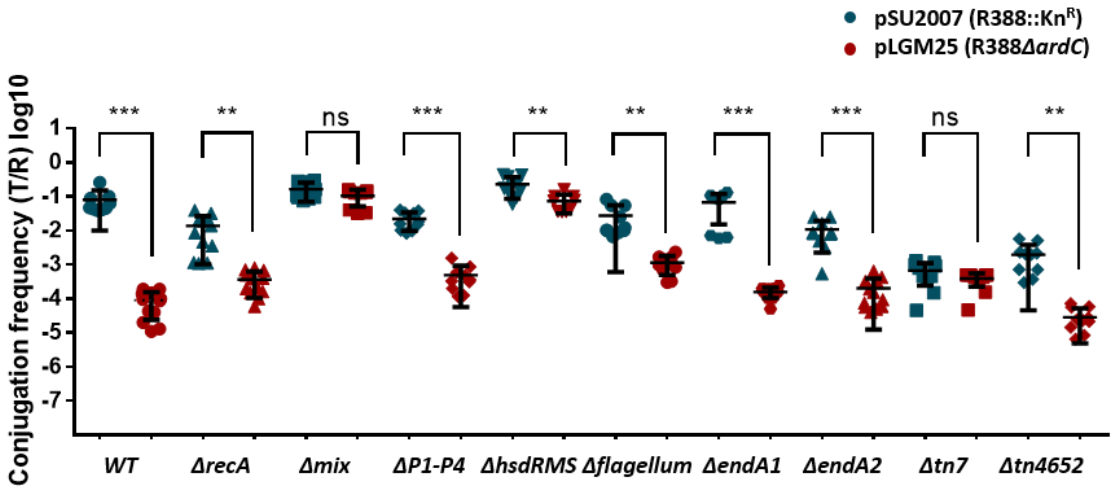



***Figure 73.*** *Effect of ardC on plasmid conjugative transfer from E. coli to P. putida KT2440 mutants at 37 °C. The conjugation frequencies per recipient (T/R) into the WT strain or into different mutants of P. putida KT2440. Δmix=EM42 (Δprophage1 Δprophage4 Δprophage3 Δprophage2 ΔTn7 ΔendA-1 ΔendA-2 ΔhsdRMS Δflagellum ΔTn4652). Conjugations were done for 1h at 37 °C. Horizontal and vertical bars represent the mean ± SD obtained for each dataset of n=8-12 (t-test: ** p < 0.01, *** p < 0.001).*

HsdRMS activity is reported to be inactivated by high temperatures. We wanted to be assured that the previous results obtained were not influenced by the conjugation temperature, thus, we repeated the conjugation experiment at 30 °C to *KT2440 WT, KT2440 EM42 (Δmix)* and *KT2440ΔhsdRMS* observing the same results in terms of conjugation differences (Figure 74). However, conjugation at 37 °C is slightly higher with or without *ardC* probably due to a partial inactivation of the *hsdRMS* system. We can conclude that the hsdRMS system is involved in the activity of ArdC.

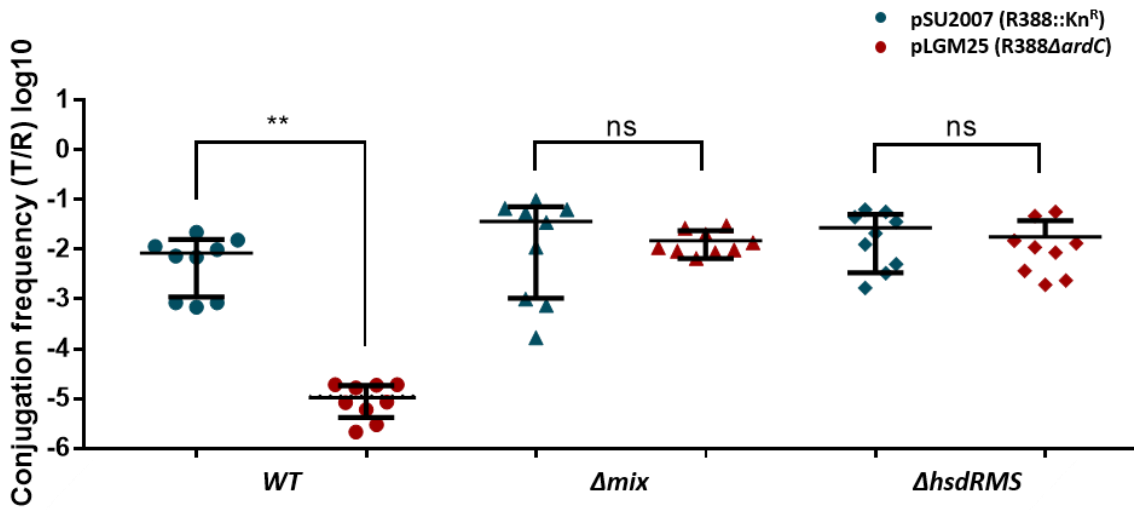



***Figure 74.*** *Effect of ardC on plasmid conjugative transfer from E. coli to P. putida KT2440 mutants at 30 °C. The conjugation frequencies per recipient (T/R) into the WT strain or into different mutants of P. putida KT2440 are shown. Conjugation was done for 1h at 30 °C. Horizontal and vertical bars represent the mean ± SD obtained for each dataset of n=9 (t-test: ** p < 0.01).*

In order to test if this drop in the conjugation frequencies when *ardC* is absent was due to differences in the restriction-modification system of the recipient cells, whichever the strain, we tested to conjugate from *E. coli BW27783 (*rK- mK+) to *C41*, an rB- mB- *E. coli* strain. Although the experiment was only performed once, no differences were observed in the frequency of conjugation when conjugating pSU2007, pIC10 or pLGM25 from *E. coli BW27783* to natural antibiotic resistant *E. coli C41* or *vice* versa (). This indicates us that the *E. coli* R-M system is not the target or in relation with ArdC activity classifying its function in interspecific conjugation.

***Table 31****. Effect of ardC on plasmid conjugative transfer between E. coli rK- mK+ BW27783 strain and E. coli rB- mB- C41 strain in both directions carrying pSU2007, pIC10 or pLGM25 in donors. Conjugations were done for 1 h at 37 °C. The conjugation frequencies per recipient (T/R) are shown for n=1.*

| | | f(T/R) | |
|---|---|---|---|
| From\To | | C41 | BW27783 |
| PSU2007 | BW27783 | 6,67E-01 | - |
| pIC10 | BW27783 | 4,00E-01 | - |
| pLGM25 | BW27783 | 9,09E-01 | - |
| PSU2007 | C41 | - | 3,74E-01 |
| pIC10 | C41 | - | 1,13E-01 |
| pLGM25 | C41 | - | 1,58E-01 |

As we observed an approximate 10-fold increase in all the conjugation frequencies for all conjugation experiments performed from *E. coli* to *P. putida*, when conjugating at 37 °C instead of at 30 °C, the optimal growing temperature of *P. putida*, we wanted to know if the same effect could be observed in conjugation from *E. coli* to *E. coli* at 42 °C. However, in this case we did not observe an increase in the conjugation frequency (**Figure 75**). Thus, as heat-shock response is not involved in *E. coli* to *E. coli* conjugation, as it is a characteristic of *P. putida* and not of R388-derivative plasmids.

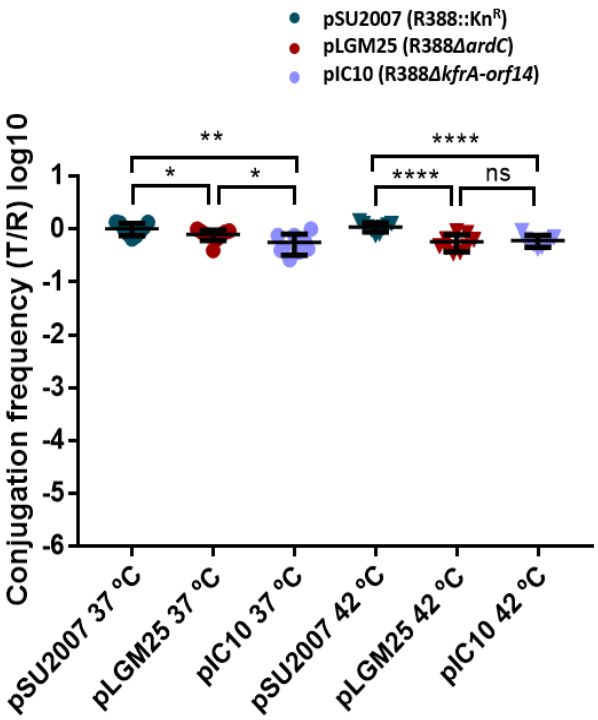



***Figure 75**. Effect of temperature on plasmid conjugative transfer from E. coli BW27783-Nx to E. coli BW27783-Rif. The conjugation frequencies per recipient (T/R) are shown. Horizontal and vertical bars represent mean ± SD obtained for each dataset of n=9 (t-test: * p < 0.1, ** p < 0.01, ****

### 4.4.7.2 ArdC protective activity is not due to metalloprotease activity

In order to check if the differences in conjugation were due to the protease activity of ArdC, we mutated the glutamic acid of the active site E229A from pSU2007 that is expected to deactivate the proteolytic center. Surprisingly, non-significative differences were observed (Figure 76).

In order to double check this previous result, we also tested the effect of this E229A mutation in ArdC_I protein (position 229 respect the first methionine amino acid, defined as position 1) overexpressed in pUCP22 (pLGM37) (Figure 77). Plasmid with mutant expressed in *P. putida KT2440* recipients was tested when conjugating pLGM25 plasmid from *E. coli* to *P. putida KT2440* as previously done in Section 4.2.1.2 at 37 °C and 30 °C.

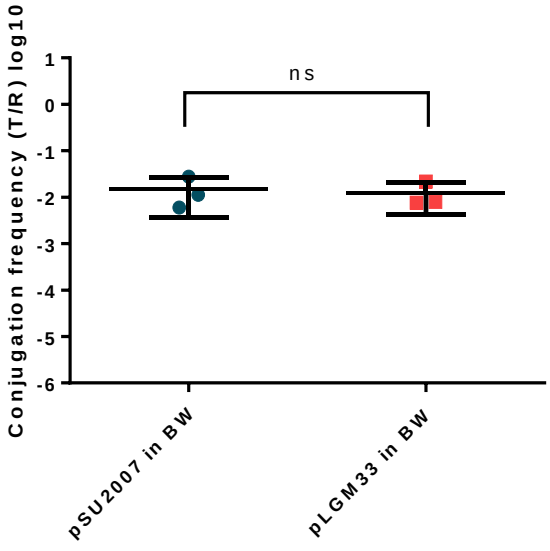



***Figure 76.** Effect in the conjugation frequency when ArdC E229 residue is mutated to E229A in pSU2007 plasmid. Conjugation was done for 1h at 37 °C. Horizontal and vertical bars represent the mean ± SD of n=3 observations.*

ArdC E229A expression in recipient cells was able to rescue the conjugation to *P. putida* as efficiently as ArdC wt. This could be due to the protease not being inactivated just with this mutation, as partially happens with metalloprotease Zmp1 (Schacherl et al., 2015) that requires a second active center residue to be altered or, more probably, this could mean that the differences in conjugation are due to the DNA protection exerted by ArdC binding to ssDNA. Temperature was not affecting the target of ArdC, as for wt and mutant ArdC protein the conjugation frequency order of magnitude was the same, and again approximately 10 fold higher at 37 °C.

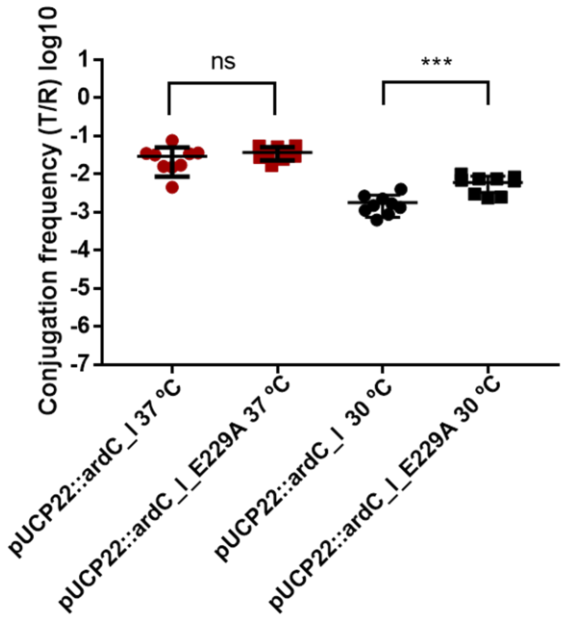



***Figure 77.*** *Effect in the conjugation frequency when ArdC E229 residue is mutated to A at different temperatures. Complementation of pLGM25 was done in recipients with pUCP22::ardC_I (shown as circles) or pUCP22:: ardC_I_E229A (shown as squares). Conjugation was done for 1h at 37 °C (in maroon) and at 30 °C (in black) with 0.1 mM IPTG added to the mating mixture. Horizontal and vertical bars represent the mean ± SD obtained for each dataset of n=9.*

# 3. Discussion

## 5.1 ArdC improves interspecific bacterial conjugation by acting in recipient cells

As proposed by (Belogurov et al., 2000), ArdC could play a key role in plasmid promiscuity due to the similarity with the TraC1 primase. They mentioned that IncW plasmids do not have a primase gene and that ArdC could do this function of 'plasmid accommodation' into the recipient cell once the T-strand is transfer, increasing the probabilities of a successful adaptation. We agree with this benefit conferred by ArdC to the plasmid as we have observed that ArdC is responsible of a broader host range, helping at least in the conjugation towards *A. tumefaciens* and *P. putida*. However, they propose that this benefit is conferred by the co-injection of ArdC protein coating the T-strand from donor to recipient and preventing from recipient R-M defense systems. However, we have not observed any ArdC transfer through the conjugative channel as when protein was overexpressed in donors to complement pLGM25 (R388Δ*ardC*) no recovery of pSU2007 (R388:: $Kn^R$) phenotype was observed. Indeed, Belogurov et al. contradicted themselves also suggesting that ArdC is not been transferred during conjugation due to some negative results they obtained. They tried to compare the conjugation rate of F self-transmissible plasmid, which lacks *ardC* and the conjugation frequency of pSA which code for *ardC* from a r- m- *E. coli C* strain to a r+ *E. coli* strain and they did not detect any difference when overexpressing *ardC* in trans in a multicopy plasmid (Belogurov et al., 2000). In accordance with these results of Belogurov et al., *ardC* from R388 has shown to positively contribute to the conjugation efficiency from *E. coli* to *P. putida* and to *A. tumefaciens* but not between *E. coli* strains (Figure 26), indicating a role in interspecific conjugation.

## 3.2. ArdC positively contributes to plasmid fitness

As shown in previous results of our group (del Campo, 2016) the establishment and maintenance module of R388 plasmid is important for conjugation from *E .coli* to different species. Although R388 is known to be highly stable in *E. coli* (del Campo, 2016), on the contrary, this plasmid is greatly unstable in *P. putida* (Figure 29). R388 derivate lacking kfrA-orf14 region (pIC10) showed higher stability in *P. putida* than pSU2007 and we have shown that ArdC is partly responsible for this characteristic as pLGM25 showed to be more stable than pSU2007. Plasmid fitness refers to the maintenance of the plasmid within a host and the ability to transfer into new hosts. Both functions are encoded and determined by the plasmid genome (Fernandez-Lopez et al., 2014). Our interspecific conjugation results supports the idea that plasmid fitness is host dependent (Fernandez-Lopez et al., 2014). Although *ardC* seems to increase the host range of R388 plasmid in terms of conjugation ability, it is not contributing to the plasmid maintenance in the host under non selective pressure. *ardC* could be unbalancing the plasmid fitness towards a more successful conjugation transfer in an attempt to reach as many different hosts as possible before taking care of the plasmid maintenance task. Selective pressure and time will end stabilizing the plasmid in the population by compensatory mutations as observed for pNUK73 plasmid (San Millan et al., 2015). This group studied by RNA-seq the effect of horizontal gene transfer in *P. aeruginosa*. Small plasmid pNUK73 altered the transcriptional gene expression of *P. aeruginosa* in a big stent, causing an important fitness cost to the host bacteria. pNUK73 cost was found to be due to the high expression of *rep*, the gene codifying for the plasmid replication protein, which overshooted the SOS response between other expression changes in the host. It was not until 300 generations, that the host compensated the cost of carrying the plasmid by compensatory mutations in a putative helicase and two putative serine/threonine kinases codified in the host chromosome. These mutations lead to a reduction in the expression of the plasmid replication protein gene *rep* that at the same time reduced the SOS response increasing the plasmid fitness in the host (San Millan et al., 2015).

Plasmidic antirestriction strategies seem to be an important feature of plasmid biology as entry exclusion or fertility inhibition described in Section 1.3.1. Following the open path incited by (Garcillán-Barcia and de la Cruz, 2008) for plasmid gene mining, we have investigated the role of *ardC* in conjugation and our results could suggest for *ardC* to be a gene contributing to plasmid fitness. Therefore, we would like to incorporate the *ardC* information obtained in this thesis to the enormous wealth of the plasmid genetic pool.

## 3.3. Conjugation alters the transcriptome of both donors and recipient cells

In an attempt to study the functional role of *ardC* and how it contributes to plasmid fitness, we have developed an interspecific RNA-seq experiment of conjugation between *E. coli* and *P. putida* to analyze the changes in gene expression when an *ardC*-containing or *ardC*-lacking plasmid is introduced in *P. putida* by conjugation. To the best of our knowledge, this is the first study to analyze the conjugation process itself. The experimental design had to overcome several difficulties as how to discern between donor and recipient transcriptome or the low rate of transconjugants in the population that was going to reduce by at least 10 times the real transcriptional changes, masking small fold-changes.

Nevertheless, we have observed that when *ardC* is present in the plasmid, conjugation events occur and a high expression of R388 genes occur, especially those involved in the conjugation process as previously shown through RT-qPCR studies by (Fernandez-Lopez et al., 2014). In addition, an intense SOS response is triggered in recipient cells. However, when *ardC* is removed from the plasmid, conjugation events almost do not occur and SOS response is downregulated in donor cells. Due to the big differences in conjugation frequency between the a*rdC*-containing and the *ardC*-lacking plasmid, we consider that the transcriptomic changes observed can be attributed to the conjugation process itself. These results are in line with those of (Baharoglu et al., 2010). Mazel group showed that conjugation induces the bacterial SOS response by fusing *lacZ* to a gene of the SOS regulon of *E. coli* (*sfiA*) and another of *Vibrio cholera* (*recN*). They demonstrated by β-gal induction tests that during conjugation, the presence of an abnormal amount of plasmidic ssDNA in the recipient cell induces the SOS stress response of DNA repair and recombination (Baharoglu et al., 2010). SOS induction due to HGT also leads to genetic rearrangements. This way, the incoming DNA, by inducing the SOS response, increases the possibilities to integrate into the host genome (Baharoglu and Mazel, 2014).

Other publications support the evidence of plasmids causing transcriptomic changes in the host cell. As already mentioned, pNUK73 showed to cause an important SOS response overshoot in its *P. aeruginosa* host (San Millan et al., 2015). Other previous studies evaluating the transcriptional effect of horizontal gene transfer of a plasmid (pCAR1) to *P. putida KT2440* by microarray only observed subtle changes in the chromosomal transcriptome of *KT2440*. pCAR1, an IncP-7 degradative plasmid, induced *parI*, a gene that encodes a protein with a ParA-like AtPase fused to a HTH Xre-type DNA binding motif. This ParA homologue is thought to be required for the correct chromosomes segregation upon cell division (Miyakoshi et al., 2007). The transcriptome of *KT2440* bearing pWW0 (the catabolic IncP-9 plasmid present in the original *P. putida mt-2* strain) was also analyzed by microarray by (Domínguez-Cuevas et al., 2006). pWWO plasmid encodes genes for the TOL pathway that allows the cell to metabolize toluene and other aromatic compounds that at the same time are toxic for cells. They showed that the

transcriptional machinery was mainly focused in enduring general stress instead of in the metabolic pathways to degrade the aromatic compounds (Domínguez-Cuevas et al., 2006).

It is well known that bacteria adapt to stressful circumstances by gene expression modulation to overcome each situation. And to do that there are different ways to control gene expression, for example, by the activation of regulators or through small regulatory RNAs. *P. putida* is a laboratory model for environmental bacteria. It is a robust strain towards stress. However, it is known that each stress type shoots a unique transcriptional response in *P. putida* (Bojanovič et al., 2017). The global transcriptional response of *P. putida* to different stress conditions, as osmotic, oxidative and imipenem antibiotic was well studied by Long, K. S. group. They found 194 differentially expressed genes in common for the three types of stress. Almost half of them are of unknown function, others are involved in general stress response, efflux pumps and other transporters or redox enzymes for energy production (Bojanovič et al., 2017). In this study they observed a strong SOS pathway activation for both oxidative and osmotic stress similar to the one we have observed in our conjugation transcriptional results.

When we conjugated our R388 derivative plasmids to *P. putida KT2440ΔrecA i*n an attempt to check a putative implication of RecA in *ardC*-mediated conjugation and showed that pLGM25 was transferred from *E. coli* to *KT2440ΔrecA* with a similar efficiency than to *KT2440,* we concluded that ArdC activity is not dependent on RecA. In accordance with this results, we think that the high activation of the SOS pathway in WT recipients observed by RNA-seq is a consequence of conjugation and not to the presence of ArdC.

We have to point out that our RNA-seq experimental design has some limitations. Samples containing only donors and only recipients would be needed as negative controls too. In addition, doing the analysis by triplicate is also the optimum to extract reliable results. Despite this, we believe our work is an interesting preliminary study to understand the biology of interspecies conjugation. We have gain a better understanding about interspecific conjugation and its associated SOS response.

## 3.4. ArdC is a ssDNA binding protein

ArdC was first defined by Belogurov et al. to be a ssDNA binding protein. To check this, we performed EMSA assays observing that ArdC preferentially bind ssDNA but also dsDNA. In addition, ArdC structure has been solved *de novo* by X-ray crystallography. The results of this study show that ArdC is formed by two domains; a ssDNA binding domain for which only a similar structure is available in the PDB database (Rad4, 2QSG) (Min and Pavletich, 2007), and a more common zinzin metalloprotease (MPTase) domain. ArdC-N domain has no close relationship with any other protein fold, thus having a quite distinctive structure.

MPTase domain and ArdC-N domains are the two most common domains found in polyvalent proteins, proteins with 2 to 15 domains of disparate activities found in viruses and plasmids to overcome biological conflicts with potential hosts. Interestingly these two domains are usually found together, along or in larger polypeptides able to perform a diverse range of activities (Iyer et al., 2017). Thus, ArdC is the simplest and most common core representative of polyvalent proteins.

Due to the fact that most of the times ArdC-N is found in the leader N-terminal region of polyvalent proteins, it has been proposed that ArdC-N could have a domain-coupling role meanwhile the polypeptide chain is delivered into the host (Iyer et al., 2017).

ArdC-N closest homologue Rad4 recognizes DNA duplex distortions and with the long β-hairpin flips-out the damaged base pairs in DNA lesions. ArdC has a very electronegative loop in the tip of the β-hairpin (Figure 44), predicting electrostatic repulsion with the backbone phosphates of the DNA. However, we have seen that the tip of the long protruding β-hairpin of ArdC do not show to be very conserved (Figure 43) indicating that it may not have a crucial function in ArdC.

We have also observed the "squiggle" motif defined by (Krishnan et al., 2018) also observed in Rad4 BHD domains. They propose that this motif may be responsible of a highly flexible region that could facilitate conformational changes during DNA sequence recognition. However, we did not observe big structural rearrangement in this region except for the long β-hairpin itself in the ArdC structure bound to ssDNA (Figure 48).

In addition, Rad4 BHD1 and TGD domains bind to undamaged dsDNA, however, BHD2 and BHD3 bind together to the DNA region with the lesion but does not interact with the damaged nucleotides, they bind to the flipped-out nucleotides of the undamaged strand. (Min and Pavletich, 2007). In a similar way, and due to the observance that ArdC can bind ssDNA (preferentially) as well as dsDNA as observed by EMSA, we suggest that ArdC could be binding to unpaired nucleotides in the interfaces between ssDNA and dsDNA during plasmid replication. However, further work is required to test this hypothesis.

We find surprising to discover by Dali that ArdC-N closest structural homologue was in the eukaryotic kingdom. Rad4 DNA binding domains were acquired by eukaryotes and now form part of a DNA repair protein of the nucleotide excision repair (NER) pathway. In addition, a homologue of ArdC-N has also been found in kinetoplastids (Krishnan et al., 2018). *Trypanosoma* Tc38 protein functions as a ssDNA binding protein at the origin of replication of kinetoplastid DNA playing an important role in replication and maintenance of the DNA. Both Tc38 and Rad4 have several copies of the ArdC-N domain to interact with a larger DNA region. As postulated by (Krishnan et al., 2018), in both cases ArdC-N domains were probably acquired through horizontal gene transfer through genetic mobile elements in bacterial endosymbionts confirming the widely distribution of *ardC*.

## 3.5. ArdC has a metalloprotease domain that is not be needed for conjugation to *P. putida*.

In regards to the C-terminal domain, once we realized by X-ray crystallography that ArdC contains a MPTase domain and that apparently there is no metal bound to the active site in the crystallization conditions, we checked the thermal stability of the protein in the presence of different metals which could give us a clue about the metal usage. The most stabilizing cation was cobalt. Cobalt has been found in other proteins with the same structural characteristics, as in the YfcM hydroxylase (PDB: 3WTR), however we could not crystallize ArdC bound to this metal. We obtained the structure of ArdC bound to $Zn^{2+}$ only by soaking native protein crystals obtained in a specific condition and not by soaking crystals obtained in the condition used for SAD structural solution. IrrE homologue was crystallized with $Zn^{2+}$ too. However, they found that IrrE proteolytic activity requires either $Zn^{2+}$, $Mn^{2+}$ or $Fe^{2+}$ (Blanchard et al., 2017). We also obtained at the same condition as for ArdC-Zn structure the structure of ArdC bound to $Mn^{2+}$ but only after forcing the presence of the metal by co-purifying ArdC with $Mn^{2+}$ in the purification buffers and not by soaking. Most metalloproteases contain one or two $Zn^{2+}$ ions. However, it is known that zinc metalloproteases maintain the catalytic activity with $Co^{2+}$ and $Mn^{2+}$ too due to the flexibility of these three geometrical metal coordination (Fukasawa et al.,

2011). As we have not found a proteolytic activity to perform assays, we could not determine the biological metal needs of ArdC.

ArdC, due to the structure of its catalytic active center is defined as a gluzinzin, a metalloprotease with a glutamic acid (E) coordinating a $Zn^{2+}$ cation. There are members of the gluzinzin metalloprotease MA(E) subclan synthetized as proenzymes that require the removal of the N-terminal propeptide to become active as the elastase of *P. aeruginosa* (Kessler and Safrin, 1994). The thermolysin family has a large number of zinc metalloproteases in the subclan MA(E) and the autoproteolytic mechanism was well studied by (Gao et al., 2010). DNA in different forms also activate the autocleavage of ArdC-C Spartan homologue protein. As suggested by Iyer group, the MPTase domain could be released by autoproteolysis or serve to convert the protein into an active product once inside the host (Iyer et al., 2017). Nevertheless, we have not observed self-proteolytic cleavage in ArdC with any of the tested conditions.

In Spartan homologue protein, ZBD sub-domain restricts the substrate access to the active center. In our case, we propose that it is the highly flexible loop between α5 and α7 the one that could be doing a similar function in ArdC as it does the corresponding loop in Zmp1 metalloprotease, which suffers large motions after substrate recognition (Schacherl et al., 2015).

According to our results, ArdC is not a protein that protects from DNA damage by UV light as their homologues IrrE, Rad4 or Spartan. However, the sensitivity to other DNA damage agents as mitomycin C in the absence or presence of ArdC would be interesting to discard a DNA damage repairing role in ArdC.

We have seen by crosslinking assays that ArdC does not form multimers alone or bound to ssDNA. In addition, PISA analysis neither indicated a predicted stable quaternary structure. Even with the dimers-like from the unit cell obtained for the structure of ArdC bound to $Mn^{2+}$ PISA did not predict to be a complex in solution. Spartan protein also crystallized in a dimer disposition with a 2-fold symmetry axis along ssDNA. They proposed that this disposition where two Spartan molecules bind simultaneously to the same DNA molecule could be the functional active unit of the protease. Further crosslinking assays, thermal stability or gel filtration chromatography should be done for ArdC in the presence of $Mn^{2+}$ to make sure that ArdC do not form an active multimer in solution.

Mutations in the glutamic acid that coordinate the metal cofactor generally lead to inactivation of the protease as in IrrE (Ludanyi et al., 2014) and YfcM hydroxylase ((Kobayashi et al., 2014), LF protein , astacin and aminopeptidase A (Schacherl et al., 2015). An exception was found in Zmp1 metalloprotease which active site mutant (E143A) maintains notable residual activity (18 % of proteolytic activity) and a second residue had to be mutated (catalytic Tyr) for a complete inactivation (Schacherl et al., 2015). We only mutated one residue in ArdC; E229A. We could mutate a second residue of the active center, nevertheless, we have not observed nor a slightly influence in the conjugation experiment from *E. coli* to *P. putida*. On the other hand, we are aware that trying to find the metalloprotease target by doing proteolytic assays against the specific putative targets selected was highly improbable. However, we were neither able to find an ArdC target by pull down assay. We are conscious that this approach only allows to detect binding partners present at high concentrations and tightly joined to the pray. Thus, after all the efforts to find ArdC metalloprotease activity, we consider that the metalloprotease domain of ArdC could be either inactive or simply be prepared to be active in other conditions or situations.

## 3.6. ArdC antirestriction activity prevents from degradation by *P. putida* Type I R-M system

Despite the fact that (Belogurov et al., 2000) defined the motif "LIpDfdQS-aayvQ" to be similar to other "antirestriction" signatures conserved for all known Ard proteins and to be the one responsible for antirestriction activity, we have observed that this region does not appear to be conserved in the ArdC family except for its Y255 (Figure 42). By similarity with other metalloproteases, this tyrosine, which is oriented towards the catalytic metalloprotease active center, may be involved in target stabilization while cleavage. As we have not observed any protease activity during our experiments, we postulate that the antirestriction activity of ArdC is not because of this "antirestriction" signature neither by the metalloprotease activity itself.

On the other hand, we neither think that ArdC has a specific activity against Type II R-M system. We believe that the *in vitro* protection that Belogurov group showed against HhaI restriction enzyme is a mere artefact of a massive protein binding to the ssDNA as much more protein than DNA was used for the assays, and recognition sites could be completely occluded. We first thought that this observance could be due to HhaI degradation by ArdC proteolytic domain. However, we could not observe any proteolysis of this restriction enzyme.

However, we are in good agreement with earlier findings about the ArdC activity against Type I R-M systems shown by Belogurov group. By conjugation to *P. putida KT2440 mutants*, we observed that ardC was not needed if the Type IA R-M system of *P. putida* was depleted. It is worth mentioning that KT2440 was first described as a plasmid free spontaneous restriction-deficient (*hsdR1*) derivative of *P. putida* mt-2 by (Bagdasarian et al., 1981). They showed that KT2440 was a good recipient for RSF1010 plasmid in transformation experiments and they assigned this characteristic to a defective restriction system against incoming DNA (Regenhardt et al., 2002). Due to this defect in its R-M system, KT2440 is thought to be a good host able to easily recruit plasmids (Regenhardt et al., 2002). However, in our RNA-seq experiments we have observed transcription for the three *hsdRMS* genes (PP4740-PP4742) indicating a possibly functionality as it was previously suggested by (Martínez-García et al., 2014).

We have not figured out yet which is the mechanism of action by which ArdC protects from *P. putida* Type I R-M system. The results with ArdC E229A mutant indicating a probably inactive protease in wt protein discard an IrrE-like indirect mode of action by cleavage of a transcriptional factor that for example could regulate the *hsdRMS* operon. In terms of time, and knowing that ArdC is produced once in recipient cells, it seems more probable to be a direct interaction by inhibiting R or S subunits activity for example by blockage or by dismantling the hsdRMS complex structure as ArdA does.

Knowing that ArdC protects from degradation by *P. putida* hsdRMS, and following this line of thoughts, we can suggest that ArdC is less active against *A. tu*mefaciens hsdRMS and is not needed at all against *E. coli* Type I R-M system according to our conjugation results (Figure 26). Thus, we could consider *A. tumefaciens* being in an "intermediate" position between *P. putida* and *E. coli*. We checked with Phyre2 software the similarities between the three genes of the three species and found that *E. coli* hsdM has a 68 % identity and 99 % query cover and *A. tumefaciens* a 60 % identity and 99 % query cover with hsdM of *P. putida*. *E. coli* hsdR has a 62 % identity and 99 % query cover and *A. tumefaciens* a 30 % identity and 67 % query cover with hsdR of *P. putida*. Finally, *E. coli* hsdS has a 45 % identity and 96 % query cover and *A. tumefaciens* a 36 % identity and 96 % query cover with hsdS of *P. putida*. These results do not support our

idea and are not enough to help us identify ArdC target by protein conservancy. Thus, further research is needed to gain a better understanding about ArdC mode of action.

An interesting observation was that conjugation frequency from *E. coli* to *P. putida* at 37 °C increased about 10 times in comparison with the conjugation at 30 °C. We first thought it was due to a partial inactivation of the R-M system, however, the same outcome was observed when conjugating to KT2440ΔhsdRMS strain. This characteristic could be explained by a heat-shock or SOS response upregulation in response to high temperatures in recipients, as the one observed by conjugation events, that could increase the frequency of DNA transfer. However, we did not observe the same phenomenon in conjugation between *E. coli* strains. R388 is a non-thermosensitive plasmid (Fernandez-Lopez et al., 2014). However, a recent publication (Hashimoto et al., 2019) showed that cell-to-cell transformation in *E. coli* is promoted at high temperatures. Thus, our observances of conjugation being facilitated by high temperatures seem to be due to a characteristic of *P. putida KT2440* that needs further research.

Another important question is when ArdC is expressed and when it starts to protect in the recipient cell. Known Type I R-M systems target double stranded DNA, thus, it is probably once the plasmid is replicated in the recipient and is in dsDNA form when ArdC is produced and begins to protects from *P. putida* hsdRMS and not before as ColIb-P9 ArdA protein, which is produced soon in the cell from a ssDNA promoter (*ssi*)(Zavilgelsky, 2000).

As stated, the picture is still incomplete, and further research is needed to fully understand ArdC mode of action. Nevertheless, with this work, we have increase the knowledge about antirestriction strategies employed by mobile genetic elements to increase their fitness and be able to conquer new host to spread $Ab^R$ genes and other traits between unrelated bacteria.

# 6 Conclusions

1. ArdC is only needed for interspecific conjugation between certain strains. Thus, the role of ArdC might be the plasmid host range expansion.

2. ArdC is needed in *P. putida* recipient cells for an efficient R388 plasmid conjugation.

3. ArdC is not needed for conjugation to *P. putida* recipient cells lacking the Type I R-M system. Therefore, ArdC seems to be involved in avoiding recipients' immigration control.

4. R388 ArdC possesses two domains: a ssDNA binding domain and a metalloprotease domain.

5. ArdC metalloprotease activity is not needed for conjugation to *P. putida*.

6. SOS response is highly activated in recipients during R388 transfer from *E. coli* to *P. putida*.

# 7 Spanish version

## 7.1 Introducción

Las proteínas de unión a ADN son necesarias para numerosas funciones celulares como el mantenimiento, la replicación y la transcripción del DNA. Dentro de ellas, se encuentran las proteínas de unión a ADN de cadena sencilla (ssDNA). Estas proteínas de unión a ssDNA tienen diversas estructuras para llevar a cabo diferentes actividades en la célula como estabilizar el ssDNA, la reparación del daño en el ADN, o la transferencia génica horizontal (HGT).

HGT es el mecanismo por el cual elementos genéticos móviles (MGEs), como el ADN plasmídico, se transfieren a una célula sin relación con el organismo que los poseían inicialmente. Los principales mecanismos de HGT son la transformación (entrada de ADN desde el exterior celular), la transducción (transferencia de ADN de una célula a otra a través de fagos) y la conjugación (que requiere un contacto directo entre dos células) (Soucy et al., 2015).

La relevancia de estos procesos de HGT recae en que es el principal mecanismo por el que se transfieren genes que confieren resistencia a antibióticos ($Ab^R$). Este hecho supone un gran reto para la humanidad, ya que los antibióticos dejan de ser efectivos para tratar infecciones bacterianas debido a la adquisición por parte de las bacterias de estos mecanismos de defensa (Clatworthy et al., 2007).

Los plásmidos conjugativos son moléculas grandes de DNA con capacidad autónoma para replicarse en una célula y ser transferidos entre células mediante conjugación. Estos plásmidos contienen numerosos genes que confieren adaptación a diferentes circunstancias, como genes que confieren resistencia a antibióticos ($Ab^R$), genes necesarios para la movilización del plásmido y construcción del canal conjugativo entre la célula receptora y la donadora, un origen de transferencia (*oriT*) y un origen de replicación (*oriV*) además de los genes que codifican para la maquinaria replicativa del plásmido (Smillie et al., 2010).

A continuación, se muestra un esquema del proceso conjugativo (Figura 1).

1) Lo primero es el contacto entre una bacteria donadora y una receptora y que se produzca el canal conjugativo o sistema de secreción de tipo IV (T4SS).
2) A continuación, una proteína relaxasa (R) corta una de las hebras del plásmido por el *oriT*, y unido al DNA de cadena sencilla lo conduce hacia el T4SS mientras que el plásmido empieza a replicarse para dejar una copia de doble cadena en la célula donadora.
3) La proteína acopladora (T4CP) junto al T4SS empieza a bombear el

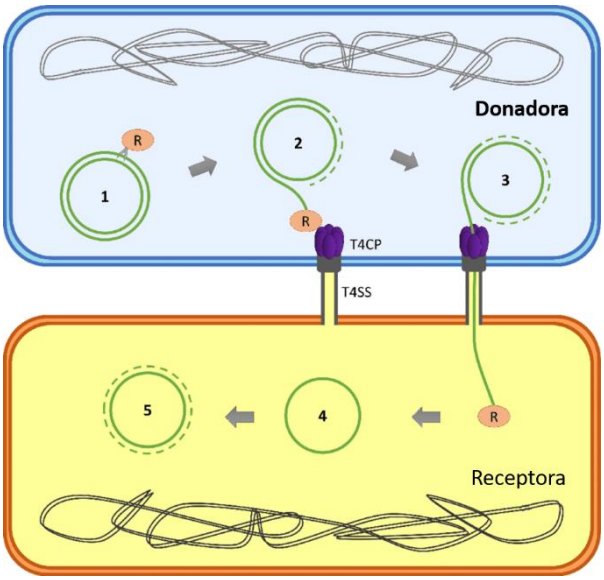



***Figura 1.*** *Representación esquemática del proceso conjugativo. Modificado de (Getino and de la Cruz, 2018).*

ADN de cadena sencilla hacia la célula receptora.

4) Una vez en la célula receptora, la relaxasa recircularíza el plásmido.
5) Esta copia de cadena sencilla es replicada para formar el plásmido de cadena doble, copia del que había en la célula donadora.

En la Figura 2 se muestran los diferentes bloques funcionales de genes en los que está compuesto el plásmido R388, plásmido conjugativo modelo en nuestro laboratorio. En azul se muestran los genes implicados en la estabilidad y mantenimiento del plásmido. En esta región se encuentran los llamados genes "accesorios" pues no tienen una función vital para la célula pero que confieren ciertas ventajas selectivas bajo distintas circunstancias y que son los que primero entran en la célula receptora durante la conjugación por estar precedidos por el *oriT*.

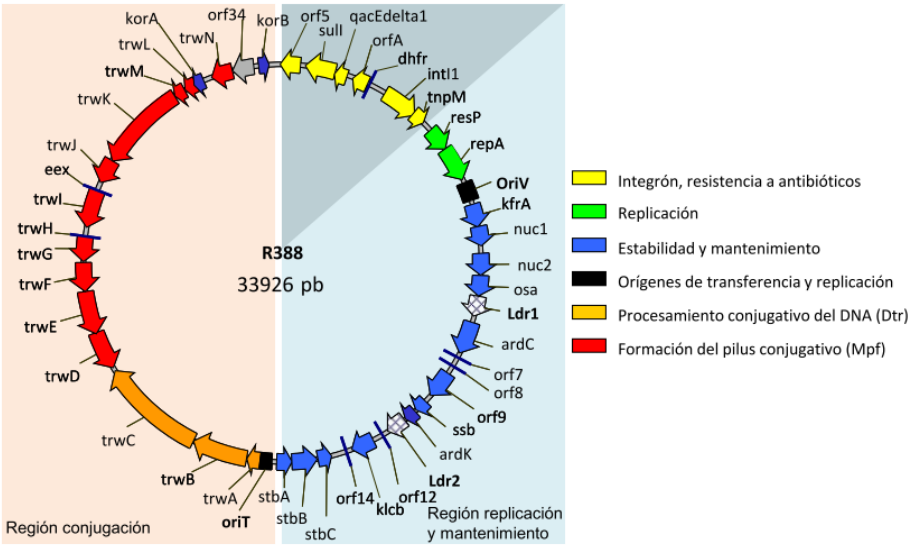



***Figura 2.*** *Mapa génico del plásmido R388. (del Campo, 2016).*

R388 es el plásmido prototipo del grupo de incompatibilidad IncW, y no es estable en una célula en la que haya otros plásmidos con este mismo grupo de incompatibilidad (Garcillán-Barcia et al., 2009). Los plásmidos IncW se caracterizan por estar presentes en un bajo número de copias por célula y poseer un amplio número de genes de resistencia a antibióticos. Además, son plásmidos de amplio rango de huésped, es decir, que son capaces de replicarse y permanecer en un gran número de huéspedes. Por ejemplo, R388 se ha encontrado en α-Proteobacteria como *Agrobacterium tumefaciens*, γ-Proteobacteria como *Acinetobacter calcoaceticus, Pseudomonas aeruginosa, P. solanacearum*, *Salmonella typhimurium, Shigella flexneri, Stenotrophomonas maltophilia* o *E. coli* (de donde fue inicialmente aislado) (Naomi Datta & R. W. Hedges, 1972). R388 cuenta con promotores fuertes pero intensamente reprimidos de forma que solo se sobreexpesan cuando ocurre la conjugación y seguidamente se vuelven a reprimir haciendo al plásmido más eficaz en su capacidad de conjugación (Fernandez-Lopez et al., 2014). Al mismo tiempo, de esta forma también se facilita la convivencia del plásmido con su célula hospedadora, causando un menor gasto de recursos (San Millan et al., 2015).

La promiscuidad o capacidad de conjugación de R388 hacia distintas células receptoras ya había sido estudiada en nuestro laboratorio por del Campo, 2016. Se vió que la región de estabilidad y mantenimiento de R388 (de *kfrA* a *orf14* ) era importante para la conjugación a otras bacterias como *Pseudomonas putida* o *Agrobacterium tumefaciens* y su mantenimiento pero no entre *E. coli* (del Campo, 2016)(Figura 3).

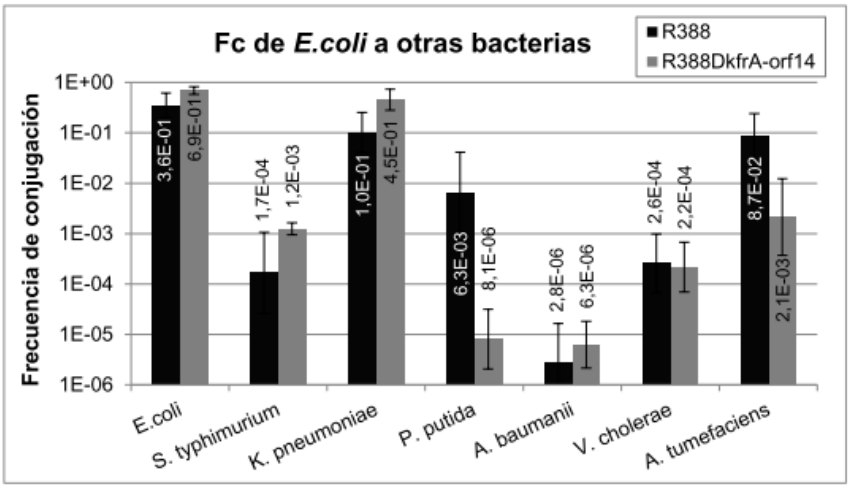



***Figura 4.*** *Frecuencia de Conjugación de R388 y pIC10 (R388ΔkfrA-orf14) desde E. coli hacia otras bacterias (del Campo, 2016).*

Los plásmidos, a la hora de ser transferidos a una célula receptora durante el proceso de conjugación, tienen que pasar una especie de "control interno de inmigración" llamado sistema de restricción-modificación (R-M), que decide si permite la entrada de ADN exógeno o no. Estos sistemas de R-M marcan con unos patrones de metilación el ADN propio (gracias a una metilasa, M) y van a actuar degradando o no el ADN entrante (gracias a una enzima de restricción, R) de acuerdo a la amenaza o familiaridad del ADN según la presencia o ausencia de estas marcas epigenéticas. Hay distintos tipos de sistemas R-M (I, II, III y IV) (Ver Tock and Dryden, 2005).Para superar estas medidas de control de entrada, algunos plásmidos y bacteriófagos han desarrollado mecanismos anti restricción. En la Figura 4 se muestra cómo la célula, gracias a la metilasa, protege su propio ADN marcándolo (1), sin embargo, la enzima de restricción degrada el ADN entrante que no está marcado como propio (2). También se muestran diferentes estrategias anti restricción: cambios en la secuencia de reconocimiento (3), protección de los sitios de corte (4), abolir las funciones de la metilasa o enzima de restricción (5) o inhibición y bloqueo de las enzimas del sistema R-M (6) ) (Tock and Dryden, 2005).

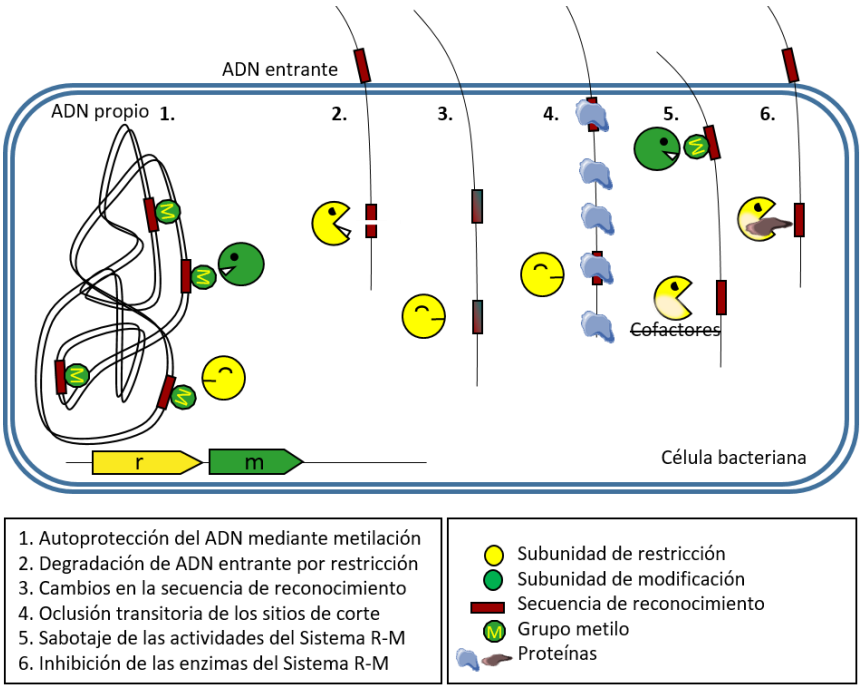



***Figura 3.*** *Representación esquemática del sistema R-M de defensa de la célula y diferentes estrategias de anti restricción.*

El plásmido R388 contiene un gen llamado *ardC* (alivio de la restricción del ADN de tipo C), que posee un 100 % de homología con *ardC* del plásmido pSA aislado de *Shigella* y descrito por Belogurov et al., 2000. Ellos describieron cómo esta proteína ArdC es capaz de unir ADN de cadena sencilla y en qué medida tiene una alta homología con la primasa TraC1 de RP4. Es por todo ello que predijeron una posible transferencia activa de esta proteína con el ADN hacia la célula receptora para protegerla así de la degradación por parte de las enzimas de restricción durante la conjugación (Belogurov et al., 2000; Fernandez-López et al., 2006).

El gen *ardC* se encuentra en la región de R388 con genes principalmente descritos con función de establecimiento y que resultó prescindible a la hora de conjugar de *E. coli* a *E. coli* en condiciones de laboratorio aunque no de *E. coli* a *P. putida* (del Campo, 2016). Esto nos indujo a pensar que la función de ArdC sea la del establecimiento de estos plásmidos de amplio rango de huésped en células receptoras con sistemas de R-M diferentes a los de *E. coli*.

## 7.2 Objetivos

La conjugación bacteriana es el principal mecanismo para la diseminación de genes de resistencia a antibióticos. Algunos plásmidos pueden conferir resistencias a los antibióticos mediante la transferencia a una amplia gama de cepas bacterianas, convirtiéndose en un problema a nivel mundial. Por lo tanto, el estudio de las estrategias utilizadas por los plásmidos para ser altamente promiscuos y los mecanismos para escapar de los sistemas R-M de las células receptoras es esencial en la lucha contra la propagación de los genes de resistencia a los antibióticos. Estas estrategias pueden ser empleadas por el plásmido conjugativo R388 a través de la expresión de proteínas anti restricción como ArdC.

Por este motivo, nuestro principal objetivo es la caracterización del papel y el mecanismo de acción de la proteína anti restricción ArdC a través de un enfoque biológico, bioquímico y estructural. Para alcanzar este objetivo principal, los objetivos específicos para este propósito fueron:

1. Caracterización biológica del papel de ArdC mediante conjugación de un plásmido derivado de R388 sin ardC hacia diferentes cepas bacterianas wt y mutantes.

2. Análisis del proceso de conjugación mediante RNA-seq para identificar los genes expresados diferencialmente en el proceso.

3. Caracterización bioquímica de ArdC para verificar las actividades de ssDNA y anti restricción de la proteína wt.

4. Caracterización biológica del mutante ArdC para verificar la actividad anti restricción.

5. Caracterización estructural de ArdC por cristalografía de rayos X.

## 7.3 Resultados y discusión de resultados

### 7.3.1 ArdC mejora la conjugación bacteriana entre especies actuando en las células receptoras

rdC es una proteína codificada por un gen que se encuentra en una zona con función de estabilidad y mantenimiento en el plásmido R388, definida como "no esencial", y que ya fue delecionada en el pasado (pIC10: R388Δ*kfrA-orf14) (del Campo, 2016).* Se vio que esta región de establecimiento era necesaria para conjugar desde *E. coli* hacia *P. putida* y mejoraba la frecuencia de conjugación hacia *A. tumefaciens* pero no hacia otra *E. coli* (del Campo, 2016). En esta tesis, hemos construido mediante el método Wanner (Datsenko and Wanner, 2000) el plásmido pLGM25 (R388Δ*ardC*) para hacer ensayos de conjugación y ver si el efecto observado por del Campo se debe principalmente a que se delecionó el gen *ardC*. En efecto, *in vivo*, *ardC* no parece ser imprescindible a la hora de transferir R388 de *E. coli* a *E. coli,* pero sí de *E. coli* a *A. tumefaciens* y especialmente de *E. coli* a *P. putida* en las condiciones de laboratorio en las que hemos llevado a cabo los ensayos de conjugación (Figura 5). La ausencia de *ardC* reduce la frecuencia de conjugación hacia *P. putida* más de dos órdenes de magnitud, luego esta proteína está facilitando de algún modo la entrada del plásmido en estas células receptoras y, por tanto, podemos decir que aumentando el rango de hospedador. Este clon nos ha permitido encontrar condiciones en las que *ardC* sí es un gen necesario de R388 ya que aporta una ventaja importante en el contexto de la conjugación interespecífica. Tras estos resultados, decidimos hacer ensayos de

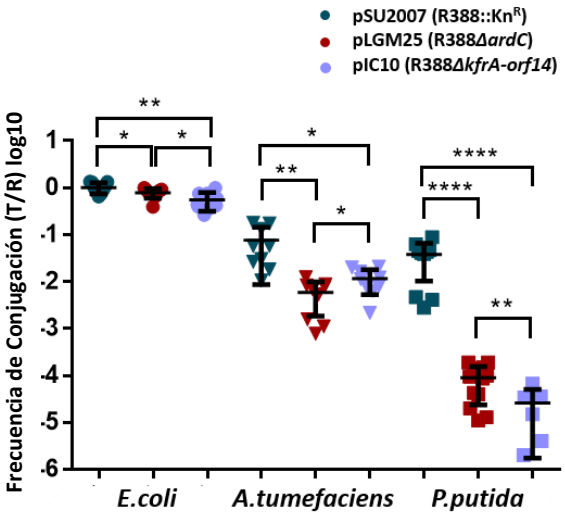



***Figura 5.*** *Frecuencias de conjugación de pSU2007 (derivado de R388 Kn$^R$), pLGM25 (mutante que carece de ardC) y pIC10 (mutante que carece de la región comprendida entre kfrA y orf14) desde E. coli a E. coli (1 h 37 °C), A. tumefaciens (1h 30 °C) y P. putida (1h 37 °C). Se muestran las frecuencias de conjugación por receptor (T/R). Las barras horizontales y verticales representan la media ± SD obtenida para cada conjunto de datos de n = 6-12 (ANOVA de 1 vía: ** p <0.01, *** p <0.001).*

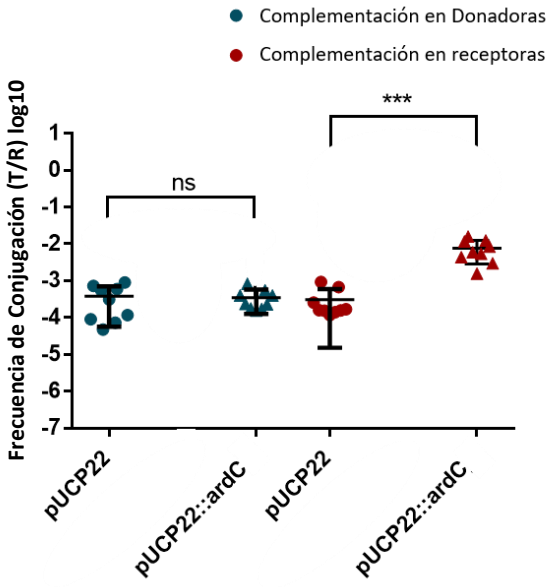



***Figura 6.*** *Efecto en la frecuencia de conjugación cuando se complementa pLGM25 en donadores (se muestra en azulado) o en receptoras (se muestra en granate) con pUCP22 o pUCP22:: ardC. La conjugación se realizó durante 1 hora a 37 °C con IPTG 0,1 mM . Las barras horizontales y verticales representan la media ± SD obtenida para cada conjunto de datos de n = 9 (*** p <0.001).*

complementación de pLGM25 tanto en donadoras como en receptoras para ver en qué células actúa ArdC (Figura 6).

Como se aprecia en la Figura 6, solo sobreexpresando *ardC* en las células receptoras se complementa pLGM25 y se recuperan los valores de frecuencia de conjugación del plásmido pSU2007 descartando la hipótesis de *(Belogurov et al., 2000).* Estos investigadores postularon que la alta similitud de ArdC con el dominio N-terminal de la primasa TraC1 del plásmido puede implicar una transferencia de la proteína unida al ADN de cadena sencilla de donadoras a receptoras durante la conjugación bacteriana (Belogurov et al., 2000). Sin embargo, creemos que es en las células receptoras, y una vez que el plásmido se encuentra en forma de cadena doble cuando ArdC se empieza a producir y a ejercer su actividad.

### 7.3.2 La conjugación altera el transcriptoma de las células donadoras y receptoras

Hemos llevado a cabo estudios de transcriptómica comparada mediante RNA-seq para intentar entender el papel de *ardC* en la conjugación interespecífica de *E. coli* a *P. putida.* Para ello hemos realizado dos experimentos de conjugación; uno con el plásmido pSU2007 (R388::$Kn^R$), otro con pLGM25 y un control negativo sin plásmidos (Tabla 1).

***Tabla 1.*** *Los resultados de la conjugación para los experimentos # 1, # 2 y # 3, donde las células donadoras fueron E. coli BW27783 ($Nx^R$) que no contienen plásmido, pSU2007 o pLGM25 y las células receptoras son P. putida KT2440. La conjugación se llevó a cabo durante 30 minutos a 37 °C. Se muestran las frecuencias de conjugación por receptor (Frec. T / R).*

| Exp. | Don. | Rec. | D:R | Frec. (T/R) |
|---|---|---|---|---|
| (# 1) | *BW27783* | *KT2440* | 6.5 | 0.0E+00 |
| (# 2) | pSU2007 in *BW27783* | *KT2440* | 5.6 | 1.2E-01 |
| (# 3) | pLGM25 in *BW27783* | *KT2440* | 3.0 | 2.5E-05 |

Mediante el análisis transcripcional de donadoras, receptoras y niveles de expresión de los genes de los plásmidos, hemos localizado qué rutas o qué genes están siendo sobreexpresados o reprimidos en comparación con el transcriptoma de referencia. Hemos visto que el plásmido pSU2007 del experimento # 2 en relación con la expresión de los genes de pLGM25 del experimento # 3, sobreexpresa significativamente muchos de los genes implicados en conjugación, especialmente *trwN* y *kikA* (Tabla 2). Vemos en la Tabla 1 que en el experimento # 2 hay una frecuencia de conjugación de 0,1 mientras en el experimento # 3 prácticamente no hay conjugación. Así, estamos observando los genes que aumentan su expresión al pasar el plásmido de las células donadoras a las receptoras. Este resultado indica que en el experimento # 2 están ocurriendo eventos de conjugación y en el # 3 no, tal y como cuantificamos en la Tabla 1.

***Tabla 2.*** *Nivel de expresión de los genes R388 en RPKMs del experimento # 2 (pSU2007) y # 3 (pLGM25) y la relación de expresión en RPKMs (pSU2007 / pLGM25) para los genes diferencialmente sobre expresados. La lista se ordena de mayor a menor según la columna RPKMs (pSU2007 / pLGM25) y se colorea de verde a rojizo. No se muestran los genes no presentes en el plásmido pSU2007 o pLGM25.*

| | Experimento | | Relación | Información | |
|---|---|---|---|---|---|
| Código | pSU2007 | pLGM25 | **pSU2007/ pLGM25** | Gen | Función de la proteína |
| R388_0003 | 23018 | 985 | 23.38 | *trwN* | Dominio de tipo lisozima |
| R388_0002 | 24093 | 1160 | 20.77 | *KikA* | Causa inhibición reversible del crecimiento |
| R388_0022 | 39522 | 2674 | 14.78 | *orf14* | Dominio putativo de unión a DNA de respuesta a cambios de temperatura |
| R388_0024 | 40586 | 2782 | 14.59 | *orf12* | Subunidad metiltransferasa putativa de sistemas R-M de tipo I |
| R388_0023 | 33625 | 2762 | 12.17 | *klcB* | Des reprime los operones regulados por KorA |
| R388_0046 | 238133 | 30898 | 7.71 | *orf45* | |
| R388_0047 | 77021 | 10165 | 7.58 | *orf46* | |
| R388_0004 | 86511 | 13806 | 6.27 | *korA* | Represor transcripcional putativo |
| R388_0005 | 87087 | 16717 | 5.21 | *trwL* | Proteína de la familia de TrbC/VIRB2 |
| R388_0006 | 61669 | 16422 | 3.76 | *trwM* | Proteína tipo VirB3 del Sistema de secreción tipo IV (SST4) |
| R388_0012 | 18435 | 5437 | 3.39 | *trwG* | Proteína VirB8 |
| R388_0018 | 8153 | 2604 | 3.13 | *trwA* | Proteína accesoria de unión a DNA |
| R388_0011 | 28951 | 10600 | 2.73 | *trwH* | lipoproteína de membrana |
| R388_0028 | 15085 | 6244 | 2.42 | *ssb* | |
| R388_0007 | 22440 | 9952 | 2.25 | *trwK* | ATPasa del SST4 |
| R388_0013 | 9409 | 4178 | 2.25 | *trwF* | VirB9/CagX/TrbG. Componente del SST4 |
| R388_0027 | 65575 | 29438 | 2.23 | *ardK* | Represor transcripcional putativo |
| R388_0029 | 34340 | 15686 | 2.19 | *orf9* | |

En cuanto a los niveles de expresión de los genes de *E. coli*, vimos que cuando el plásmido sin *ardC* pLGM25 está en *E. coli* BW27783, se desregulan los niveles de expresión de las rutas de movilidad del flagelo, catabolismo de la arginina entre otras proteínas de respuesta al estrés, el metabolismo del glioxilato y dicarboxilato, respuesta a SOS y respiración anaerobia y aumentan los niveles de expresión de la homeostasis redox de la célula (Figura 8). Este descontrol de la expresión no se observa cuando pSU2007 está en donadoras pudiendo indicar un sobresfuerzo en BW27783 (pLGM25) para intentar conjugar sin éxito.

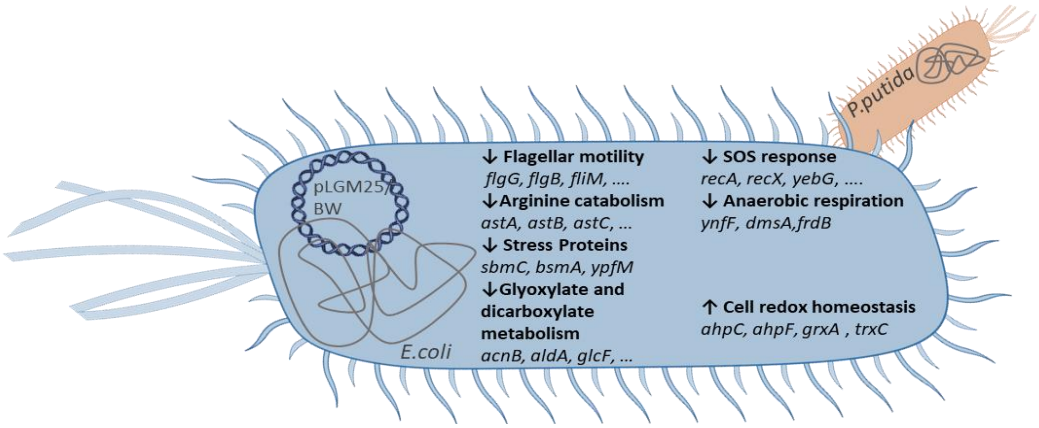



***Figura 5.*** *Visión general de los procesos y funciones con expresión diferencial en E. coli durante la conjugación hacia P. putida según lo observado por la secuenciación de ARN a nivel transcripcional agrupado por DAVID 6.7. Se muestra ordenado de mayor a menor puntuación de enriquecimiento el agrupamiento para los genes expresados diferencialmente en presencia de pLGM25 (# 3) respecto a la célula vacía (# 1).*

En *P. putida* observamos que cuando entra pSU2007, los genes cromosómicos que se sobreexpresan principalmente forman parte de la respuesta SOS (Tabla 3). Dada la gran diferencia en la frecuencia de conjugación del experimento #2 y #3, no podemos saber con seguridad si esta activación es debida a que la conjugación ocurre más o menos eficientemente, o a la presencia o ausencia de *ardC* en el plásmido.

***Tabla 3.*** *Nivel de expresión de los genes de P. putida en RPKM del experimento # 1 (BW), # 2 (pSU2007) y # 3 (pLGM25) y relación de la expresión en RPKM para cada gen. La lista de genes se ordena de mayor a menor según la columna RPKM (pSU2007 / pLGM25) y se colorea de verde a rojizo. Los nombres de los genes involucrados en la ruta de señalización SOS se muestran en negrita. PH: proteína hipotética.*

| | Experimento | | | Relación | | | Información | |
|---|---|---|---|---|---|---|---|---|
| Código | BW | pSU2007 | pLGM25 | pSU2007 /BW | pLGM25 /BW | **pSU2007/ pLGM25** | Gen | Función de la proteína |
| PP_5744 | 18 | 76 | 16 | 4,27 | 0,87 | 4,91 | ***PH*** | PH con sitio de unión LexA predicho |
| PP_3773 | 7 | 51 | 13 | 7,44 | 1,96 | 3,80 | | PH |
| PP_4616 | 63 | 298 | 79 | 4,69 | 1,25 | 3,75 | ***yebG*** | Proteína dependiente de LexA |
| PP_2177 | 9 | 29 | 8 | 3,26 | 0,89 | 3,68 | | Regulador transcripcional |
| PP_4729 | 20 | 76 | 21 | 3,89 | 1,09 | 3,57 | ***recN*** | Proteína de reparación y recombinación del ADN |
| PP_4617 | 20 | 108 | 31 | 5,30 | 1,50 | 3,53 | *ldh* | leucina deshidrogenasa |
| PP_5580 | 42 | 33 | 10 | 0,80 | 0,23 | 3,45 | | PH |
| PP_2109 | 30 | 125 | 36 | 4,22 | 1,23 | 3,43 | ***PH*** | PH con sitio de unión LexA predicho |
| PP_2143 | 58 | 206 | 62 | 3,56 | 1,07 | 3,31 | ***lexA-I*** | Represor transcripcional |
| PP_3694 | 9 | 13 | 4 | 1,41 | 0,44 | 3,21 | | PH |
| PP_3850 | 5 | 15 | 5 | 3,30 | 1,05 | 3,15 | | PH |
| PP_2924 | 9 | 23 | 8 | 2,68 | 0,90 | 2,98 | ***PH*** | PH con sitio de unión LexA predicho |
| PP_5579 | 29 | 24 | 9 | 0,82 | 0,29 | 2,79 | | PH |
| PP_4346 | 25 | 25 | 9 | 1,01 | 0,37 | 2,76 | *ddlA* | D-alanina--D-alanina ligasa A |
| PP_1521 | 101 | 107 | 40 | 1,06 | 0,39 | 2,71 | | PH |

| | Experimento | | | Relación | | | Información | |
|---|---|---|---|---|---|---|---|---|
| Código | BW | pSU2007 | pLGM25 | pSU2007 /BW | pLGM25 /BW | **pSU2007/ pLGM25** | Gen | Función de la proteína |
| PP_5091 | 125 | 141 | 52 | 1,12 | 0,41 | 2,71 | | Proteína de membrana |
| PP_2451 | 45 | 122 | 47 | 2,72 | 1,06 | 2,57 | *endX* | Endonucleasa de AND extracelular |
| PP_1522 | 335 | 339 | 134 | 1,01 | 0,40 | 2,53 | *cspA-I* | proteína de respuesta a shock por temperatura |
| PP_3109 | 14 | 9 | 4 | 0,68 | 0,27 | 2,52 | | PH |
| PP_5637 | 9 | 20 | 8 | 2,31 | 0,93 | 2,49 | | PH |
| PP_5487 | 178 | 211 | 85 | 1,19 | 0,48 | 2,49 | ***PH*** | PH con sitio de unión LexA predicho |
| PP_1203 | 21 | 57 | 24 | 2,73 | 1,15 | 2,37 | ***dinB*** | ADN polimerasa IV |
| PP_1630 | 217 | 712 | 306 | 3,28 | 1,41 | 2,33 | ***recX*** | Proteína regladora |
| PP_3089 | 3218 | 4942 | 2135 | 1,54 | 0,66 | 2,32 | *hcp1* | Hcp1 |
| PP_3901 | 22 | 23 | 10 | 1,06 | 0,46 | 2,30 | | PH |
| PP_1629 | 462 | 1654 | 720 | 3,58 | 1,56 | 2,30 | ***recA*** | recombinasa RecA |
| PP_1625 | 878 | 925 | 404 | 1,05 | 0,46 | 2,29 | *fdxA* | ferredoxina 1 |
| PP_2839 | 26 | 87 | 38 | 3,37 | 1,49 | 2,26 | ***PH*** | PH con sitio de unión LexA predicho |
| PP_2180 | 15 | 20 | 9 | 1,35 | 0,61 | 2,23 | *spuC-I* | poliamina: piruvato transaminasa |
| PP_4350 | 8 | 12 | 6 | 1,61 | 0,72 | 2,23 | | aminotransferasa de tipo V de la familia de NifS/IscS |
| PP_5694 | 784 | 851 | 391 | 1,09 | 0,50 | 2,18 | | PH |
| PP_0641 | 245 | 481 | 221 | 1,96 | 0,91 | 2,17 | | PH |
| PP_4349 | 16 | 23 | 11 | 1,47 | 0,68 | 2,17 | | PH |
| PP_2838 | 47 | 144 | 66 | 3,06 | 1,42 | 2,16 | | PH |
| PP_5464 | 24 | 38 | 17 | 1,59 | 0,74 | 2,15 | | PH |
| PP_2840 | 11 | 33 | 16 | 2,93 | 1,39 | 2,11 | | Proteína de membrana |

En resumen, hemos observado que cuando *ardC* está presente en el plásmido (exp. # 2), se producen eventos de conjugación que generan una respuesta SOS en las células receptoras. Sin embargo, cuando *ardC* no está presente en el plásmido (exp. # 3), la conjugación no es efectiva y la respuesta SOS está reprimida en estas células donadoras.

### 7.3.3 Caracterización estructural de ArdC

Mediante cristalografía de rayos X, hemos resuelto la estructura de ArdC (PDB: 6I89, 2.00 Å) (Figura 8) y hemos visto que está compuesta por dos dominios estructurales. Hemos encontrado por homología con otras proteínas con estructura resuelta que ArdC posee un dominio de unión al ADN de cadena sencilla, tal y como vio *(Belogurov et al., 2000)* por ensayos bioquímicos, y un dominio metaloproteasa (Tabla 4).

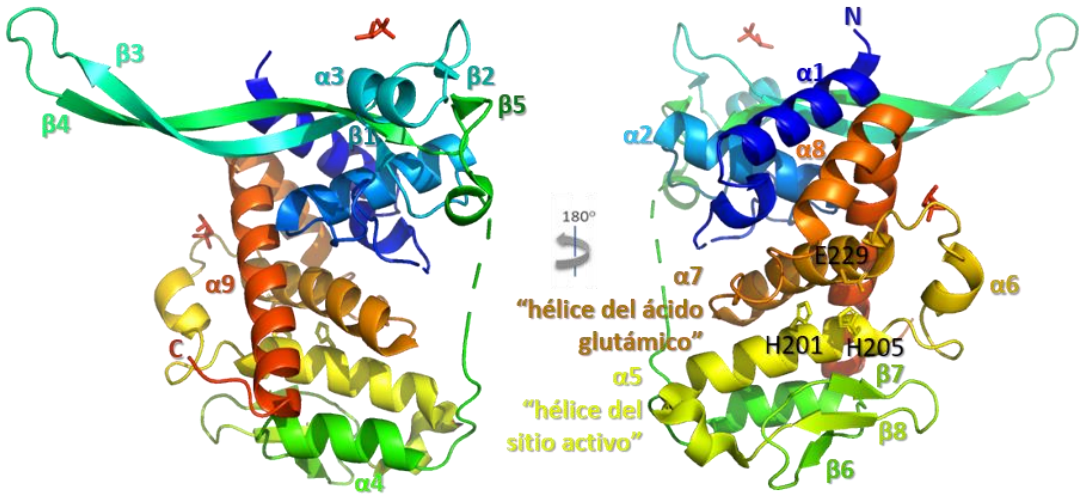



***Figura 6**. Estructura terciaria de ArdC. Se muestra en colores fríos el dominio de unión a ADN y en colores más cálidos el dominio metaloproteasa. Se muestran etiquetados los residuos del bolsillo catalítico que coordinan el cofactor metálico y el extremo N y C de la proteína.*

***Tabla 4.** Homólogos estructurales más cercanos de ArdC obtenidos con Dali para cada dominio estructural.*

| | Z-score | % id | PDB | Protein |
|---|---|---|---|---|
| **N-terminal domain** | 4.5 | 9 | 2QSG | Proteína reparadora de DNA Rad4 |
| **C-terminal domain** | 6.4 | 12 | 6MDW | Metaloproteasa Spartan |
| | 6.2 | 16 | 3DTE | Metaloproteasa IrrE |

El mapa de superficie de cargas electrostáticas (Figura 9) revela un surco cargado electropositivamente entre el dominio N-terminal y el C-terminal donde posiblemente se una el DNA. De hecho, hemos visto densidad electrónica en este surco cuando cristalizamos ArdC unido a un oligonucleótido formado por 5 timinas (1.80 Å).

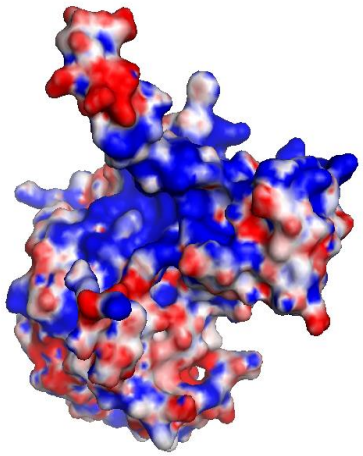


***Figura 9.** Potencial electrostático de superficie de ArdC. El potencial de la superficie accesible para el solvente se calculó y coloreó en el rango (−8 a +8 kT/e) a un pH=7 con ayuda de APBS en PyMOL*

También hemos comprobado por ensayos de retardo en gel que ArdC *in vitro* une ADN de cadena sencilla tanto de cadena corta (45nt) como larga (ADN de M13 de unas 7kb) con mucha mayor afinidad que ADN de cadena doble, tal y como había observado previamente el grupo de Belogurov.

En cuanto al homólogo más cercano del dominio N-terminal vemos que es Rad4, una proteína que contiene tres dominios similares a ArdC-N (BHD1, BHD2 y BHD3) siendo el del medio el más parecido a ArdC-N (Figura 10). Rad4 es un miembro de la ruta NER de reparación y escisión de nucleótidos. Insertando la horquilla de láminas β a través del ADN dañado, saca hacia fuera las bases dañadas a reparar por otras proteínas NER (Min & Pavletich, 2007, Krishnan et al., 2018).

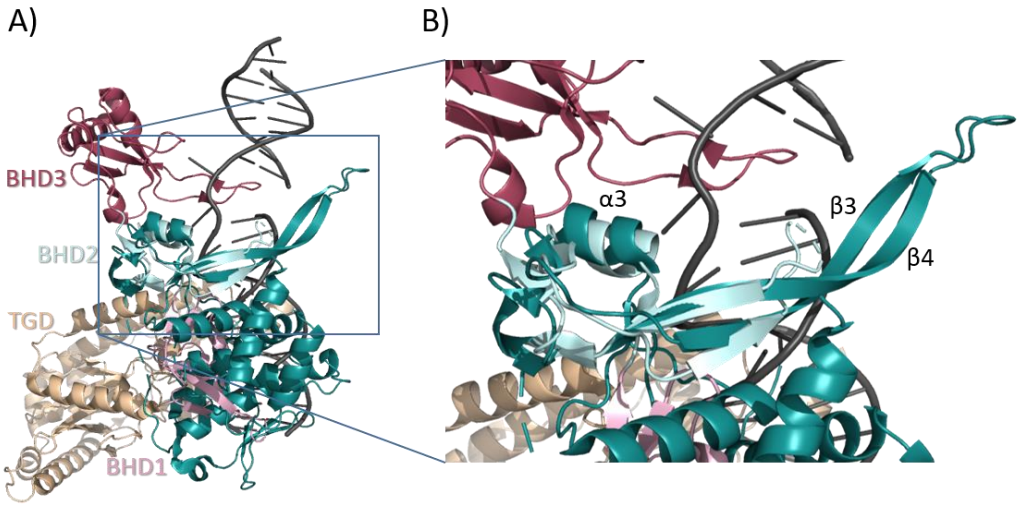



***Figura 7.*** *Alineamiento estructural de ArdC y Rad4. A) ArdC se muestra en verde azulado, y los cuatro dominios de Rad4 unido a ADN dañado por UV (2QSG) se encuentran etiquetados. B) Vista más cercana del dominio ArdC-N superpuesto a Rad4_BHD2.*

También hemos encontrado que ArdC posee un dominio metaloproteasa cuyo centro activo está compuesto por dos histidinas (H201, H205) y un ácido glutámico (E229) catalogándose por tanto dentro del grupo de las gluzincinas cuyo centro activo está compuesto por un motivo de unión a metal HEXXH en la "hélice del sitio activo" y un ácido glutámico (E) en la "hélice del ácido glutámico" (Cerdà-Costa and Gomis-Rüth, 2014). También hemos resuelto la estructura de ArdC unida a zinc (3.15 Å) y a manganeso (2.70 Å). En la Figura 11 se muestra el centro activo de ArdC unido a $Mn^{2+}$ en la que se observa una unión tetraédrica del metal.

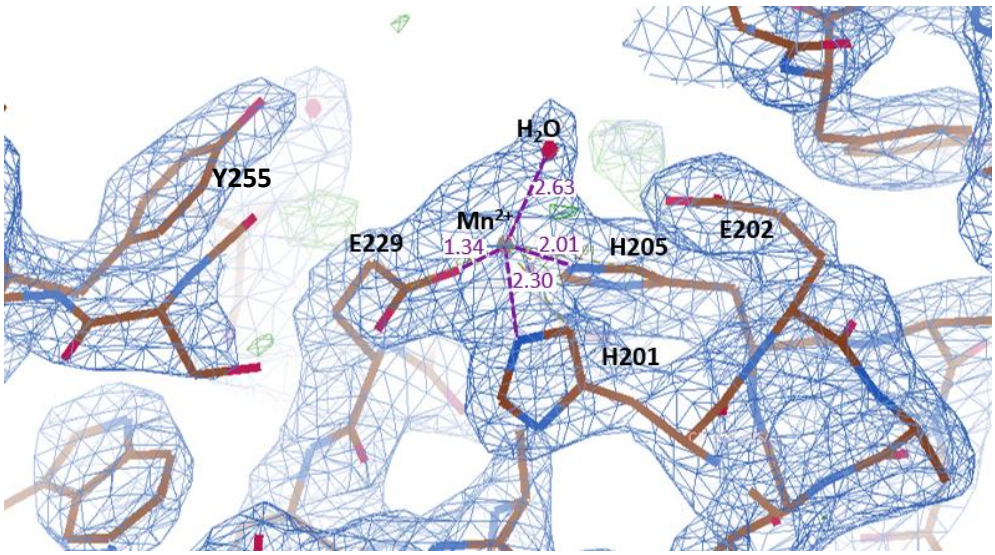



***Figura 8****. Centro activo de ArdC unido a $Mn^{2+}$ a una resolución de 2.7 Å. Los residuos y las moléculas involucradas en la coordinación del metal (distancia en Å mostrada en púrpura) o actividad proteolítica están marcados.*

Los homólogos estructurales más cercanos a ArdC-C son el subdominio metaloproteasa (MPD) del dominio SprT de la proteína Spartan e IrrE. La función de Spartan es cortar enlaces cruzados

DNA-proteína causados por ejemplo por la luz UV, para preservar la integridad del genoma. En la Figura 12 se muestra el alineamiento estructural de SprT con ArdC.

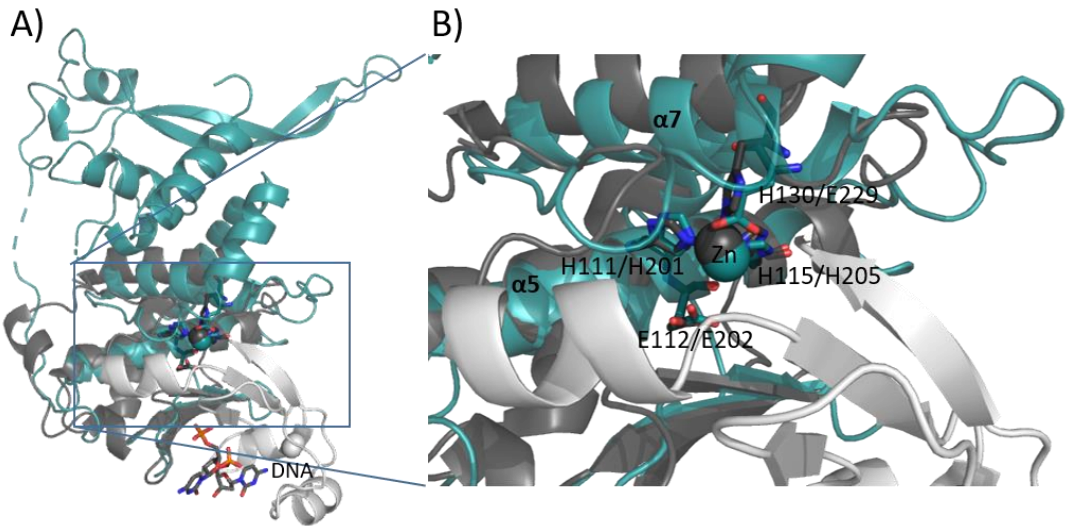



***Figura 9.*** *Alineamiento estructural de ArdC y SprT. A) Superposición de la estructura ArdC-Zn (en verde azulado, 3,15 Å) con su homólogo estructural más cercano Spartan SprT (6MDX; en gris, 1,55 Å). El subdominio de unión a Zn2 + (ZBD) se muestra en gris claro y el subdominio de metaloproteasa (MPD) se muestra en gris oscuro. B) Una vista detallada del centro activo de la metaloproteasa con los residuos involucrados en la catálisis en palos numerados como (MPD / ArdC-C).*

Por otro lado, IrrE (también llamada PprI) es la proteína de *Deinococcus deserti* responsable de la tolerancia a la radiación. Su mecanismo de acción está bien descrito, proteolizando específicamente a DdrO, un represor transcripcional que al ser cortado dispara la expresión de los genes DDR (respuesta al daño en el ADN). Se trata de una ruta de respuesta al daño en el ADN parecida al sistema LexA-RecA de respuesta SOS (Ludanyi et al., 2014; Vujičić-Žagar et al., 2009; Wang et al., 2015). En la Figura 13 se muestra el alineamiento estructural de IrrE con ArdC.

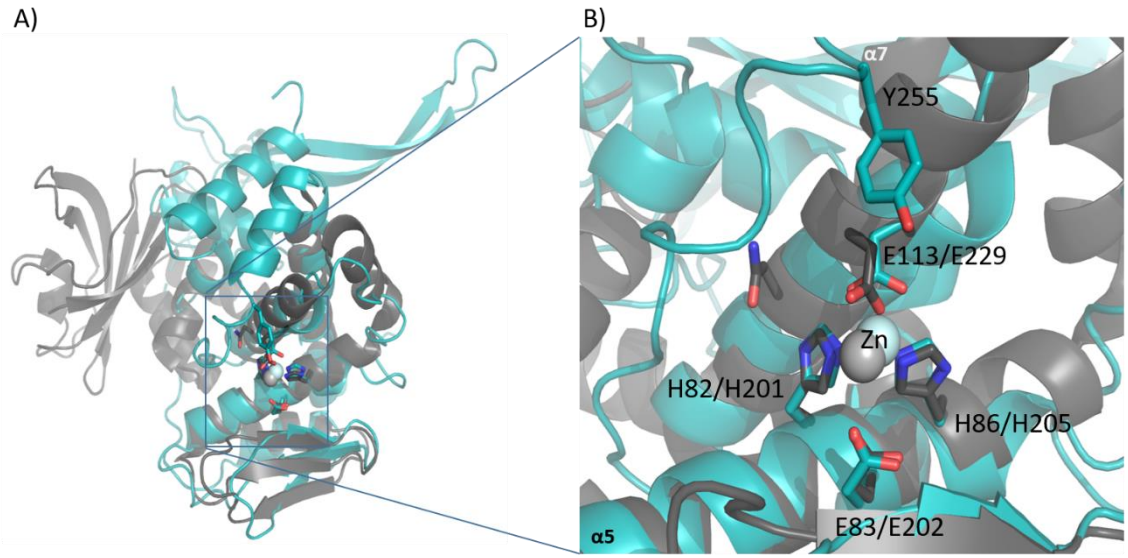



***Figura 10****. Alineamiento estructural de ArdC e IrrE. A) Superposición de la estructura ArdC-Zn (en verde azulado, 3,15 Å) con su homólogo estructural Irre-Zn de Deinococcus radiodurans (3DTI; en gris, 3,5 Å). B) Una vista detallada del centro activo con los residuos involucrados en la catálisis en barras numeradas como (IrrE / ArdC).*

Dado que tanto Rad4, Spartan e IrrE tienen funciones relacionadas con reparación del daño en el ADN causado por agentes como la luz UV, hicimos ensayos de reparación de daños de cadena sencilla en el ADN causado por luz UV acompañados o no de fenómenos conjugativos. Sin embargo, no observamos ningún comportamiento diferencial estando o no presente *ardC*, por lo que hemos descartado su implicación en señales de reparación SOS asociados a daños en el DNA.

### 7.3.4 Caracterización de la actividad de ArdC

Para encontrar la diana del dominio metaloproteasa de ArdC, realizamos ensayos de "pull-down" que sirven para la detección de proteínas que interaccionan con nuestra proteína de interés unida a una columna de purificación. Creamos el mutante ArdC E229A para evitar el corte de la proteína diana, manteniendo la capacidad de unión. Sin embargo, no conseguimos aislar la diana por este método.

En otro intento de descubrir cuál pudiera ser la diana de actuación de ArdC, repetimos el experimento de conjugación de *E. coli* hacia diferentes mutantes de *P. putida* (Figura 14). En primer lugar observamos que RecA no parece estar implicada en el mecanismo de acción de ArdC, ya que la conjugación de pLGM25 a P. putida Δ*recA* solo funciona en presencia de ArdC. A continuación, elegimos una cepa receptora especialmente diseñada para la expresión heteróloga de genes, EM42 (Δmix: *P. putida Δprophage1 Δprophage4 Δprophage3 Δprophage2 ΔTn7 ΔendA-1 ΔendA-2 ΔhsdRMS Δflagellum ΔTn4652*) (Martínez-García et al., 2014) y vimos que ArdC no era necesaria para conjugar eficientemente a esta cepa. Por tanto, procedimos a evaluar cada una de sus mutaciones individualmente y vimos que para conjugar hacia la cepa sin sistema de restricción modificación (*P. putida* Δ*hsdRMS*) no se precisa *ardC* en el plásmido.

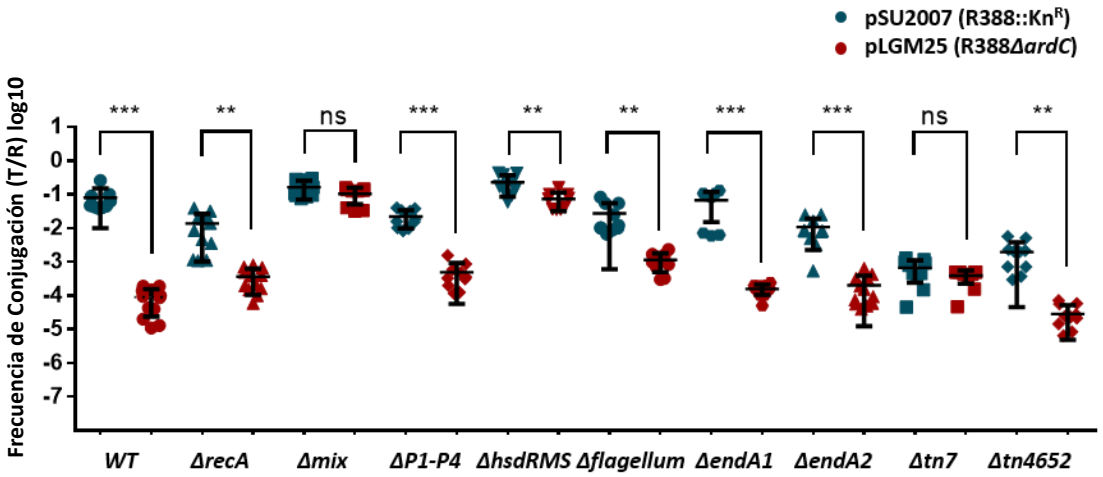



***Figura 11.*** *Efecto de ardC en la transferencia conjugativa de plásmidos de E. coli a diferentes mutantes de P. putida KT2440. Se muestran las frecuencias de conjugación por receptor (T / R). Δmix = EM42 (Δprophage1 Δprophage4 Δprophage3 Δprophage2 ΔTn7 ΔendA-1 ΔendA-2 ΔhsdRMS Δflagellum ΔTn4652). Las conjugaciones se realizaron durante 1 hora a 37ºC. Las barras horizontales y verticales representan la media ± SD obtenida para cada conjunto de datos de n = 8-12 (prueba t: ** p <0.01, *** p <0.001).*

Una de nuestras hipótesis iniciales era que la activación del dominio proteasa ocurre al estar unido al ADN de cadena sencilla entrante durante la conjugación, de tal forma que sea capaz de degradar las enzimas de restricción que intenten atacar al ADN en el proceso de conjugación. O

bien, que su acción proteasa esté implicada en el desbloqueo de la expresión de genes implicados en respuesta SOS al verse amenazada por el ADN entrante tal y como se ha observado en la proteína IrrE para la reparación de daños en el ADN causados por la luz UV (Baharoglu et al., 2010; Petrova et al., 2010). Para comprobar si las diferencias en conjugación hacia *P. putida* observadas son debidas a la actividad metaloproteasa de ArdC, decidimos probar a conjugar con ArdC mutante (E229A), una mutación que en todas las proteínas homologas inactiva la actividad proteasa. Sin embargo, no observamos ninguna diferencia respecto a conjugar o complementar con la proteína WT. Así, la actividad metaloproteasa no es necesaria para el efecto de protección del ADN durante la conjugación en las condiciones ensayadas (Figura 15) descartando ambas hipótesis.

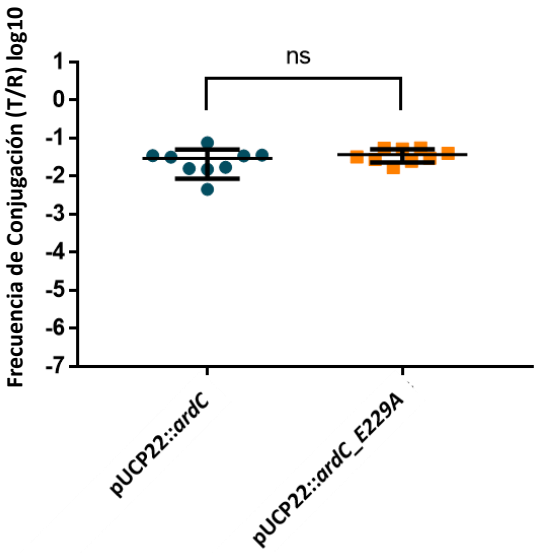



***Figura 12.*** *Efecto en la frecuencia de conjugación cuando el residuo E229 de ArdC se muta a alanina. La complementación de pLGM25 se realizó en receptores con pUCP22::ardC o pUCP22::ardC_E229A. La conjugación se realizó durante 1 hora a 37 ° C con IPTG 0,1 mM. Las barras horizontales y verticales representan la media ± SD obtenida para cada conjunto de datos de n = 9.*

## 7.4 Discusión general

ArdC es una proteína que facilita la conjugación entre bacterias de especies diferentes, por tanto, podemos decir que amplía el rango de hospedador del plásmido. Por este hecho, ArdC contribuye a la adaptabilidad (o "fitness") del plásmido en distintos ambientes. Además, hemos visto que actúa en las células receptoras, al contrario de lo que se pensaba hasta ahora (Belogurov et al., 2000). Por otro lado, hemos visto por RNA-seq que la conjugación genera en las células receptoras la activación de la respuesta SOS tal y como se había visto anteriormente (Baharoglu and Mazel, 2014). ArdC es una proteína de unión a ADN, como fue inicialmente descrita por (Belogurov et al., 2000). Además, contiene un dominio metaloproteasa cuya actividad no es necesaria para la actividad protectora en conjugación hacia *P. putida*. Finalmente, hemos comprobado la actividad anti restricción de ArdC al comprobar que la proteína no es necesaria para conjugar hacia receptoras que carecen de sistemas de restricción-modificación.

## 7.5 Conclusiones

1. ArdC solo es necesaria para la conjugación interespecífica entre ciertas cepas. Por lo tanto, el papel de ArdC podría ser la expansión del rango del huésped del plásmido.

2. Se necesita ArdC en las células receptoras de *P. putida* para una conjugación eficaz del plásmido R388 hacia esta cepa.

3. No se necesita ArdC para la conjugación hacia células receptoras de *P. putida* que carecen del sistema R-M Tipo I. Por lo tanto, ArdC parece estar involucrada en evitar el control de inmigración de las células receptoras.

4. ArdC de R388 posee dos dominios: un dominio de unión a ADN de cadena sencilla y un dominio metaloproteasa.

5. La actividad de la metaloproteasa de ArdC no es necesaria para la conjugación hacia *P. putida.*

6. La respuesta SOS está altamente activada en las células receptoras durante la transferencia de R388 de *E. coli* a *P. putida*.

# 8 Bibliography

Afonine P.V.; Grosse-Kunstleve R.W; Adams P.D (2005). The Phenix refinement framework. CCP4 Newsl. Protein Crystallogr. 42, Contrib. 8.

Aminov, R.I. (2011). Horizontal Gene Exchange in Environmental Microbiota. Front. Microbiol. *2: 158*, 1–19.

Andersson, D.I., and Hughes, D. (2014). Microbiological effects of sublethal levels of antibiotics. Nat. Rev. Microbiol. *12*, 465–478.

Asano, K. (2014). Why is start codon selection so precise in eukaryotes? Translation *2: e28387*, 1–15.

Atanasiu, C., Su, T.-J., Sturrock, S.S., and Dryden, D.T.F. (2002). Interaction of the ocr gene 0.3 protein of bacteriophage T7 with EcoKI restriction/modification enzyme. Nucleic Acids Res.

Bagdasarian, M., Lurz, R., Rückert, B., Franklin, F.C., Bagdasarian, M.M., Frey, J., and Timmis, K.N. (1981). Specific-purpose plasmid cloning vectors. II. Broad host range, high copy number, RSF1010-derived vectors, and a host-vector system for gene cloning in Pseudomonas. Gene *16*, 237–247.

Bagdasarian, M., Bailone, A., Angulo, J.F., Scholz, P., Bagdasarian, M., and Devoret, R. (1992). PsiB, and anti-SOS protein, is transiently expressed by the F sex factor during its transmission to an Escherichia coli K-12 recipient. Mol. Microbiol. *6*, 885–893.

Baharoglu, Z., and Mazel, D. (2014). SOS, the formidable strategy of bacteria against aggressions. FEMS Microbiol Rev. *38*, 1126–1145.

Baharoglu, Z., Bikard, D., and Mazel, D. (2010). Conjugative DNA transfer induces the bacterial SOS response and promotes antibiotic resistance development through integron activation. PLoS Genet. *6: e100116*, 1–10.

Bailey, S.L., Harvey, S., Perrino, F.W., and Hollis, T. (2012). Defects in DNA degradation revealed in crystal structures of TREX1 exonuclease mutations linked to autoimmune disease. DNA Repair (Amst). *11*, 65–73.

Balabanov, V.P., Kotova, V.Y., Kholodii, G.Y., Mindlin, S.Z., and Zavilgelsky, G.B. (2012). ArdD: A novel gene, ardD, determines antirestriction activity of the non-conjugative transposon Tn5053 and is located antisense within the tniA gene. FEMS Microbiol. Lett. *337*, 55–60.

Barrangou, R., and Oost, J. van der (2015). Bacteriophage exclusion, a new defense system. EMBO J. *34*, 134–135.

Barrett, A.J., and Rawlings, N.D. (1995). Families and clans of serine peptidases. Arch. Biochem. Biophys. *318*, 247–250.

Battye, T.G.G., Kontogiannis, L., Johnson, O., Powell, H.R., and Leslie, A.G.W. (2011). iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM. Acta Crystallogr. Sect. D Biol. Crystallogr. *67*, 271–281.

Beaber, J.W., Hochhut, B., and Waldor, M.K. (2004). SOS response promotes horizontal dissemination of antibiotic resistance genes. Nature *427*, 72–74.

Belinky, F., Rogozin, I.B., and Koonin, E. V. (2017). Selection on start codons in prokaryotes and potential compensatory nucleotide substitutions. Sci. Rep. *7: 12422*, 1–10.

Belogurov, A.A., and Delver, E.P. (1995). A motif conserved among the type I restriction-modification enzymes and antirestriction proteins: a possible basis for mechanism of action of plasmid-encoded antirestriction functions. Nucleic Acids Res. *23*, 785–787.

Belogurov, A.A., Delver, E.P., Agafonova, O. V, Belogurova, N.G., Lee, L.Y., and Kado, C.I. (2000). Antirestriction protein Ard (Type C) encoded by IncW plasmid pSa has a high similarity to the "protein transport" domain of TraC1 primase of promiscuous plasmid RP4. J. Mol. Biol. *296*, 969–977.

Bhattacharyya, A., and Figurski, D.H. (2001). A small protein-protein interaction domain common to KlcB and global regulators KorA and TrbA of promiscuous IncP plasmids. J. Mol. Biol. *310*, 51–67.

Blanchard, L., Guérin, P., Roche, D., Cruveiller, S., Pignol, D., Vallenet, D., Armengaud, J., and de Groot, A. (2017). Conservation and diversity of the IrrE/DdrO-controlled radiation response in radiation-resistant Deinococcus bacteria. Microbiologyopen *6: e00477*, 1–14.

Bochkarev, A., Pfuetzner, R.A., Edwards, A.M., and Frappier, L. (1997). Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. Nature *385*, 176–181.

Bojanovič, K., D'Arrigo, I., and Long, K.S. (2017). Global Transcriptional Responses to Osmotic, Oxidative, and Imipenem Stress Conditions in Pseudomonas putida. Appl Env. Microbiol *83 :e03236*, 1–18.

Boto, L. (2010). Horizontal gene transfer in evolution: Facts and challenges. Proc. R. Soc. B Biol. Sci. *277*, 819–827.

Bradley, D.E., Taylor, D.E., and Cohen, D.R. (1980). Specification of surface mating systems among conjugative drug resistance plasmids in Escherichia coli K-12. J. Bacteriol. *143*, 1466–1470.

Budisa, N., Steipe, B., Demange, P., Eckerskorn, C., Kellermann, J., and Huber, R. (1995). High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in Escherichia coli. Eur. J. Biochem. *230*, 788–796.

del Campo, I. (2016). Estudio de la red de regulación global de R388. University of Cantabria.

del Campo, I., Ruiz, R., Cuevas, A., Revilla, C., Vielva, L., and de la Cruz, F. (2012). Determination of conjugation rates on solid surfaces. Plasmid *67*, 174–182.

Cerda-Costa, N., and Gomis-Ruth, F.X. (2013). Architecture and function of metallopeptidase catalytic domains. Protein Sci. *23*, 123–144.

Cerdà-Costa, N., and Gomis-Rüth, F.X. (2014). Architecture and function of metallopeptidase catalytic domains. Protein Sci. *23*, 123–144.

Clatworthy, A.E., Pierson, E., and Hung, D.T. (2007). Targeting virulence: A new paradigm for antimicrobial therapy. Nat. Chem. Biol. *3*, 541–548.

Coates, A., Hu, Y., Bax, R., and Page, C. (2002). The future challenges facing the development of new antimicrobial drugs. Nat. Rev. Drug Discov. *1*, 895–910.

Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc. Natl. Acad. Sci. *97*, 6640–6645.

Davison, J. (1999). Genetic Exchange between Bacteria in the Environment. Plasmid *42*, 73–91.

Dickey, T.H., Altschuler, S.E., and Wuttke, D.S. (2013). Single-stranded DNA-binding proteins: multiple domains for multiple functions. Structure *21*, 1074–1084.

Domínguez-Cuevas, P., González-Pastor, J.-E., Marqués, S., Ramos, J.-L., and de Lorenzo, V. (2006). Transcriptional tradeoff between metabolic and stress-response programs in

Pseudomonas putida KT2440 cells exposed to toluene. J. Biol. Chem. *281*, 11981–11991.

Dubnau, D. (1999). DNA Uptake in Bacteria. Annu. Rev. Microbiol. *53*, 217–244.

Emsley, P., and Cowtan, K. (2004). *Coot* : model-building tools for molecular graphics. Acta Crystallogr. Sect. D Biol. Crystallogr. *60*, 2126–2132.

Erill, I., Campoy, S., and Barbé, J. (2007). Aeons of distress: An evolutionary perspective on the bacterial SOS response. FEMS Microbiol. Rev. *31*, 637–656.

Evans, P., and IUCr (2006). Scaling and assessment of data quality. Acta Crystallogr. Sect. D Biol. Crystallogr. *62*, 72–82.

Fernandez-Lopez, R., del Campo, I., Revilla, C., Cuevas, A., and de la Cruz, F. (2014). Negative Feedback and Transcriptional Overshooting in a Regulatory Network for Horizontal Gene Transfer. PLoS Genet. *10: e10041*, 1–15.

Fernández-López, R., Pilar Garcillán-Barcia, M., Revilla, C., Lázaro, M., Vielva, L., and De La Cruz, F. (2006). Dynamics of the IncW genetic backbone imply general trends in conjugative plasmid evolution. FEMS Microbiol. Rev. *30*, 942–966.

Fleming, A. (1929). On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to their Use in the Isolation of B. influenzæ. Br J Exp Pathol. *10*, 226–236.

Fong, S.T., and Stanisich, V.A. (1989). Location and Characterization of Two Functions on RP1 That Inhibit the Fertility of the IncW Plasmid R388. Microbiology *135*, 499–502.

Frost, L.S., and Koraimann, G. (2010). Regulation of bacterial conjugation: Balancing opportunity with adversity. Future Microbiol. *5*, 1057–1071.

Fukasawa, K.M., Hata, T., Ono, Y., and Hirose, J. (2011). Metal Preferences of Zinc-Binding Motif on Metalloproteases. SAGE-Hindawi Access to Res. J. Amin. Acids *2011: 5748*, 1–7.

Furuta, Y., and Kobayashi, I. (2013). Restriction - Modification Systems as Mobile Epigenetic Elements. Madame Curie Biosci. Database [Internet]. 1–22.

Gao, X., Wang, J., Yu, D.-Q., Bian, F., Xie, B.-B., Chen, X.-L., Zhou, B.-C., Lai, L.-H., Wang, Z.-X., Wu, J.-W., et al. (2010). Structural basis for the autoprocessing of zinc metalloproteases in the thermolysin family. Proc. Natl. Acad. Sci. U. S. A. *107*, 17569–17574.

Garcillán-Barcia, M.P., and de la Cruz, F. (2008). Why is entry exclusion an essential feature of conjugative plasmids? Plasmid *60*, 1–18.

Garcillán-Barcia, M.P., Francia, M.V., and De La Cruz, F. (2009). The diversity of conjugative relaxases and its application in plasmid classification. FEMS Microbiol. Rev. *33*, 657–687.

De Gelder, L., Ponciano, J.M., Joyce, P., and Top, E.M. (2007). Stability of a promiscuous plasmid in different hosts: no guarantee for a long-term relationship. Microbiology *153*, 452–463.

Gemperlein, K., Zipf, G., Bernauer, H.S., Müller, R., and Wenzel, S.C. (2016). Metabolic engineering of Pseudomonas putida for production of docosahexaenoic acid based on a myxobacterial PUFA synthase. Metab. Eng. *33*, 98–108.

Getino, M., and de la Cruz, F. (2018). Natural and Artificial Strategies To Control the Conjugative Transmission of Plasmids. Microbiol. Spectr. *6: MTBP-00*, 1–25.

Gibson, D.G., Young, L., Chuang, R., Venter, J.C., Hutchison, C. a, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods *6*, 343–345.

Gill, J.J., Berry, J.D., Russell, W.K., Lessor, L., Escobar-Garcia, D.A., Hernandez, D., Kane, A., Keene, J., Maddox, M., Martin, R., et al. (2012). The Caulobacter crescentus phage phiCbK: Genomics of a canonical phage. BMC Genomics *13: 542*, 1–20.

Gogarten, J.P., Doolittle, W.F., and Lawrence, J.G. (2002). Prokaryotic Evolution in Light of Gene Transfer. Mol. Biol. Evol. *19*, 2226–2238.

Gormley, N.A., Watson, M.A., and Halford, S.E. (2005). Bacterial Restriction-Modification Systems.

Goryanin, I.I., Kudryavtseva, A.A., Balabanov, V.P., Biryukova, V.S., Manukhov, I. V, and Zavilgelsky, G.B. (2018). Antirestriction activities of KlcA (RP4) and ArdB (R64) proteins. FEMS Microbiol. Lett. *365: fny22*, 1–6.

Grant, S.G., Jessee, J., Bloom, F.R., and Hanahan, D. (1990). Differential plasmid rescue from transgenic mouse DNAs into Escherichia coli methylation-restriction mutants. Proc. Natl. Acad. Sci. *87*, 4645–4649.

Grissa, I., Vergnaud, G., and Pourcel, C. (2007). The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinformatics *8*, 1–10.

Gullberg, E. (2014). Selection of Resistance at very low Antibiotic Concentrations. Uppsala University.

Hashimoto, M., Hasegawa, H., and Maeda, S. (2019). High temperatures promote cell-to-cell plasmid transformation in Escherichia coli. Biochem. Biophys. Res. Commun. *515*, 196–200.

Hecht, A., Glasgow, J., Jaschke, P.R., Bawazer, L.A., Munson, M.S., Cochran, J.R., Endy, D., and Salit, M. (2017). Measurements of translation initiation from all 64 codons in E. coli. Nucleic Acids Res. *45*, 3615–3626.

Ho, C.-H., Wang, H.-C., Ko, T.-P., Chang, Y.-C., and Wang, A.H.-J. (2014). The T4 Phage DNA Mimic Protein Arn Inhibits the DNA Binding Activity of the Bacterial Histone-like Protein H-NS. J. Biol. Chem. *289*, 27046–27054.

Hö, C., Carle, A., Sivanesan, D., Hoeppner, S., and Baron, C. (2005). The putative lytic transglycosylase VirB1 from Brucella suis interacts with the type IV secretion system core components VirB8, VirB9 and VirB11. Microbiology *151,* 3469–3482.

Holčík, M., and Iyer, V.N. (1996). Structure and Mode of Action ofkikA,a Genetic Region Lethal toKlebsiella oxytocaand Associated with Conjugative Antibiotic-Resistance Plasmids of the IncN Group. Plasmid *35*, 189–203.

Iida, S., Streiff, M.B., Bickle, T.A., and Arber, W. (1987). Two DNA antirestriction systems of bacteriophage P1, darA, and darB: characterization of darA- phages. Virology *157*, 156–166.

Iyer, L.M., Burroughs, A.M., Anand, S., De Souza, R.F., and Aravind, L. (2017). Polyvalent Proteins, a Pervasive Theme in the Intergenomic Biological Conflicts of Bacteriophages and Conjugative Elements. J. Bacteriol. *199: e0024*, 1–31.

Jerome, L.J., and Frost, L.S. (1999). *In Vitro* Analysis of the Interaction between the FinO Protein and FinP Antisense RNA of F-like Conjugative Plasmids. J. Biol. Chem. *274*, 10356–10362.

Jerome, L.J., van Biesen, T., and Frost, L.S. (1999). Degradation of FinP antisense RNA from F-like plasmids: the RNA-binding protein, FinO, protects FinP from ribonuclease E. J. Mol. Biol. *285*, 1457–1473.

Jiang, M., Datta, K., Walker, A., Strahler, J., Bagamasbad, P., Andrews, P.C., and Maddock, J.R.

(2006). The Escherichia coli GTPase CgtAE is involved in late steps of large ribosome assembly. J. Bacteriol. *188*, 6757–6770.

Keasling, J.D., Wanner, B.L., Skaug, T., Datsenko, K.A., and Khlebnikov, A. (2001). Homogeneous expression of the PBAD promoter in Escherichia coli by constitutive expression of the low-affinity high-capacity AraE transporter. Microbiology *147*, 3241–3247.

Keeling, P.J. (2009). Functional and ecological impacts of horizontal gene transfer in eukaryotes. Curr. Opin. Genet. Dev. *19*, 613–619.

Keeling, P.J., and Palmer, J.D. (2008). Horizontal gene transfer in eukaryotic evolution. Nat. Rev. Genet. *9*, 605–618.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. *10*, 845–858.

Kessler, E., and Safrin, M. (1994). The propeptide of Pseudomonas aeruginosa elastase acts an elastase inhibitor. J. Biol. Chem. *269*, 22726–22731.

Kivistik, P.A., Kivisaar, M., and Hõrak, R. (2007). Target site selection of Pseudomonas putida transposon Tn4652. J. Bacteriol. *189*, 3918–3921.

Kobayashi, K., Katz, A., Rajkovic, A., Ishii, R., Branson, O.E., Freitas, M.A., Ishitani, R., Ibba, M., and Nureki, O. (2014). The non-canonical hydroxylase structure of YfcM reveals a metal ion-coordination motif required for EF-P hydroxylation. Nucleic Acids Res. *42*, 12295–12305.

Krishnan, A., Burroughs, A.M., Iyer, L.M., and Aravind, L. (2018). Unexpected Evolution of Lesion-Recognition Modules in Eukaryotic NER and Kinetoplast DNA Dynamics Proteins from Bacterial Mobile Elements. *iScience 9*, 192–208.

Lepikhov, K., Tchernov, A., Zheleznaja, L., Matvienko, N., Walter, J., and Trautner, T.A. (2001). Characterization of the type IV restriction modification system BspLU11III from Bacillus sp. LU11. Nucleic Acids Res. *29*, 4691–4698.

Lerminiaux, N.A., and Cameron, A.D.S. (2019). Horizontal transfer of antibiotic resistance genes in clinical environments. Can. J. Microbiol. *65*, 34–44.

Li, F., Raczynska, J.E., Chen, Z., and Yu, H. (2019). Structural Insight into DNA-Dependent Activation of Human Metalloprotease Spartan. Cell Rep. *26*, 3336–3346.e4.

Low, H.H., Gubellini, F., Rivera-Calzada, A., Braun, N., Connery, S., Dujeancourt, A., Lu, F., Redzej, A., Fronzes, R., Orlova, E. V., et al. (2014). Structure of a type IV secretion system. Nature *508*, 550–553.

Ludanyi, M., Blanchard, L., Dulermo, R., Brandelet, G., Bellanger, L., Pignol, D., Lemaire, D., and de Groot, A. (2014). Radiation response in Deinococcus deserti: IrrE is a metalloprotease that cleaves repressor protein DdrO. Mol. Microbiol. *94*, 434–449.

Makarova, K.S., Wolf, Y.I., van der Oost, J., and Koonin, E. V (2009). Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. Biol. Direct *4:29*, 1–15.

Mark R Tock and David TF Dryden (2005). The biology of restriction and anti-restriction. Curr. Opin. Microbiol. *8*, 466–472.

Marraffini, L.A. (2015). CRISPR-Cas immunity in prokaryotes. Nature *526*, 55–61.

Marraffini, L.A., and Sontheimer, E.J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science *322*, 1843–1845.

Martínez-García, E., Nikel, P.I., Aparicio, T., and De Lorenzo, V. (2014). Pseudomonas 2.0: genetic upgrading of P. putida KT2440 as an enhanced host for heterologous gene expression. *13:159*, 1–15.

Martínez-García, E., Jatsenko, T., Kivisaar, M., and de Lorenzo, V. (2015). Freeing Pseudomonas putida KT2440 of its proviral load strengthens endurance to environmental stresses. Environ. Microbiol. *17*, 76–90.

Martinez, E., and de la Cruz, F. (1988). Transposon Tn21 encodes a RecA-independent site-specific integration system. Mol Gen Genet *211*, 320–325.

Matthews, B.W. (1988). Structural Basis of the Action of Thermolysin and Related Zinc Peptidases Inhibitor Binding. Acc. Chem. Res *21*, 333–340.

Matthiesen, R., Trelle, M.B., Højrup, P., Bunkenborg, J., and Jensen, O.N. (2005). VEMS 3.0: Algorithms and computational tools for tandem mass spectrometry based identification of post-translational modifications in proteins. J. Proteome Res. *4*, 2338–2347.

McMahon, S.A., Roberts, G.A., Johnson, K.A., Cooper, L.P., Liu, H., White, J.H., Carter, L.G., Sanghvi, B., Oke, M., Walkinshaw, M.D., et al. (2009). Extensive DNA mimicry by the ArdA anti-restriction protein and its role in the spread of antibiotic resistance. Nucleic Acids Res. *37*, 4887–4897.

Meima, R., and Lidstrom, M.E. (2000). Characterization of the minimal replicon of a cryptic Deinococcus radiodurans SARK plasmid and development of versatile Escherichia coli-D. radiodurans shuttle vectors. Appl. Environ. Microbiol. *66*, 3856–3867.

Meisel, A., Bickle, T.A., Kriiger, D.H., and Schroeder, C. (1992). Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage. Nature *355*, 467–469.

Miele, L., Strack, B., Kruft, V., and Lanka, E. (1991). Gene organization and nucleotide sequence of the primase region of IncP plasmids RP4 and R751. DNA Seq. *2*, 145–162.

Milner, D.S., Attah, V., Cook, E., Maguire, F., Savory, F.R., Morrison, M., Müller, C.A., Foster, P.G., Talbot, N.J., Leonard, G., et al. (2019). Environment-dependent fitness gains can be driven by horizontal gene transfer of transporter-encoding genes. Proc. Natl. Acad. Sci. *116*, 5613–5622.

Min, J.-H., and Pavletich, N.P. (2007). Recognition of DNA damage by the Rad4 nucleotide excision repair protein. Nature *449*, 570–575.

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

Miyakoshi, M., Shintani, M., Terabayashi, T., Kai, S., Yamane, H., and Nojiri, H. (2007). Transcriptome analysis of Pseudomonas putida KT2440 harboring the completely sequenced IncP-7 plasmid pCAR1. J. Bacteriol. *189*, 6849–6860.

Mruk, I., and Kobayashi, I. (2014). To be or not to be: Regulation of restriction-modification systems and other toxin-antitoxin systems. Nucleic Acids Res. *42*, 70–86.

Murray, N.E. (2000). Type I Restriction Systems: Sophisticated Molecular Machines (a Legacy of Bertani and Weigle).

Naomi Datta & R. W. Hedges (1972). Trimethoprim Resistance Conferred by W Plasmids in Enterobacteriaceae. J. Gen. Microbiol. *72*, 349–355.

O'Neill, J. (2014). Antimicrobial Resistance: Tackling a Crisis for the Health and Wealth of

Nations. Rev. Antimicrob. Resist. Tackling Drug-Resistant Infect. Glob. *4*, 1–16.

Oliveira, P.H., Touchon, M., and Rocha, E.P.C. (2014). The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. Nucleic Acids Res. *42*, 10618–10631.

Peters, J.E. (2014). Tn7. Microbiol. Spectr. *2: MDNA3-*, 1–20.

Petrova, V., Chitteni-Pattu, S., Drees, J.C., Inman, R.B., and Cox, M.M. (2009). An SOS inhibitor that binds to free RecA protein: the PsiB protein. Mol. Cell *36*, 121–130.

Petrova, V., Satyshur, K.A., George, N.P., McCaslin, D., Cox, M.M., and Keck, J.L. (2010). X-ray crystal structure of the bacterial conjugation factor PsiB, a negative regulator of RecA. J. Biol. Chem. *285*, 30615–30621.

Rawlings, N.D., Waller, M., Barrett, A.J., and Bateman, A. (2014). MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res. *42*, 503–509.

Rees, C.E.D., and Wilkins, B.M. (1990). Protein transfer into the recipient cell during bacterial conjugation: studies with F and RP4. Mol. Microbiol. *4*, 1199–1205.

Regenhardt, D., Heuer, H., Heim, S., Fernandez, D.U., Strömpl, C., Moore, E.R.B., and Timmis, K.N. (2002). Pedigree and taxonomic credentials of Pseudomonas putida strain KT2440. Environ. Microbiol. *4*, 912–915.

Roberts, R.J., Vincze, T., Posfai, J., and Macelis, D. (2015). REBASE--a database for DNA restriction and modification: enzymes, genes and genomes. Nucleic Acids Res. *43*, 298–999.

Roca, A.I., and Cox, M.M. (1997). RecA Protein: Structure, Function, and Role in Recombinational DNA Repair. Prog. Nucleic Acid Res. Mol. Biol. *56*, 129–223.

Rosenberg, C., and Huguet, T. (1984). The pAtC58 plasmid of Agrobacterium tumefaciens is not essential for tumour induction. MGG Mol. Gen. Genet. *196*, 533–536.

Rusinov, I.S., Ershova, A.S., Karyagina, A.S., Spirin, S.A., and Alexeevski, A. V Avoidance of recognition sites of restriction-modification systems is a widespread but not universal anti-restriction strategy of prokaryotic viruses.

San Millan, A., Toll-Riera, M., Qi, Q., and Maclean, C. (2015). Interactions between horizontally acquired genes create a fitness cost in Pseudomonas aeruginosa. Nat. Commun. *6:6845*, 1–8.

Schacherl, M., Pichlo, C., Neundorf, I., and Baumann, U. (2015). Structural Basis of Proline-Proline Peptide Bond Specificity of the Metalloprotease Zmp1 Implicated in Motility of Clostridium difficile. Struct. Des. *23*, 1632–1642.

Selbitschka, W., Arnold, W., Priefer, U.B., Rottschäfer, T., Schmidt, M., Simon, R., and Pühler, A. (1991). Characterization of recA genes and recA mutants of Rhizobium meliloti and Rhizobium leguminosarum biovar viciae. Mol. Gen. Genet. *229*, 86–95.

Serfiotis-Mitsa, D., Herbert, A.P., Roberts, G.A., Soares, D.C., White, J.H., Blakely, G.W., Uhrín, D., and Dryden, D.T.F. (2009). The structure of the KlcA and ArdB proteins reveals a novel fold and antirestriction activity against type I DNA restriction systems in vivo but not in vitro. Nucleic Acids Res. *38*, 1723–1737.

Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996). Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. Anal. Chem. *68*, 850–858.

Shintani, M., and Nojiri, H. (2013). Mobile Genetic Elements (MGEs) carrying catabolic genes. In Management of Microbial Resources in the Environment, (Dordrecht: Springer Netherlands), pp.

167–214.

Silver, P.A., Stirling, F., Bitzan, L., Way, J., Oliver, J.W.K., Redfield, E., and O'Keefe, S. (2017). Rational Design of Evolutionarily Stable Microbial Kill Switches. Mol. Cell *68: .e3*, 686–697.

Smillie, C., Garcillan-Barcia, M.P., Francia, M. V., Rocha, E.P.C., and de la Cruz, F. (2010). Mobility of Plasmids. Microbiol. Mol. Biol. Rev. *74*, 434–452.

Soucy, S.M., Huang, J., and Gogarten, J.P. (2015). Horizontal gene transfer: Building the web of life. Nat. Rev. Genet. *16*, 472–482.

Strack, B., Lessl, M., Calendar, R., and Lanka, E. (1992). A common sequence motif, -E-G-Y-A-T-A-, identified within the primase domains of plasmid-encoded I- and P-type DNA primases and the a protein of the Escherichia coli satellite phage P4. J. Biol. Chem. *267*, 13062–13072.

Studier, F.W., and Moffatt, B.A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. *189*, 113–130.

Studier, F.W., and Movva, N.R. (1976). SAMase gene of bacteriophage T3 is responsible for overcoming host restriction. J. Virol. *19*, 136–145.

Sugiyama, T., and Kantake, N. (2009). Dynamic Regulatory Interactions of Rad51, Rad52, and Replication Protein-A in Recombination Intermediates. J. Mol. Biol. *390*, 45–55.

Suzuki, H., Yano, H., Brown, C.J., and Top, E.M. (2010). Predicting plasmid promiscuity based on genomic signature. J. Bacteriol. *192*, 6045–6055.

Swift, G., McCarthy, B.J., and Heffron, F. (1981). DNA sequence of a plasmid-encoded dihydrofolate reductase. Mol. Gen. Genet. *181*, 441–447.

Thomas, C. M., and Summers, D. (2008). Bacterial Plasmids. In: Encyclopedia of Life Sciences (ELS) (John Wiley & Sons, Ltd: Chichester.).

Tock, M.R., and Dryden, D.T.F. (2005). The biology of restriction and anti-restriction. Curr. Opin. Microbiol. *8*, 466–472.

Vasu, K., and Nagaraja, V. (2013). Diverse Functions of Restriction-Modification Systems in Addition to Cellular Defense. Microbiol. Mol. Biol. Rev. *77*, 53–72.

Vujičić-Žagar, A., Dulermo, R., Le Gorrec, M., Vannier, F., Servant, P., Sommer, S., de Groot, A., and Serre, L. (2009). Crystal Structure of the IrrE Protein, a Central Regulator of DNA Damage Repair in Deinococcaceae. J. Mol. Biol. *386*, 704–716.

Walkinshaw, M.D., Taylor, P., Sturrock, S.S., Atanasiu, C., Berge, T., Henderson, R.M., Edwardson, J.M., and Dryden, D.T.F. (2002). Structure of Ocr from bacteriophage T7, a protein that mimics b-form DNA. Mol. Cell *9*, 187–194.

Wang, H.-C., Ho, C.-H., Hsu, K.-C., Yang, J.-M., and Wang, A.H.-J. (2014). DNA Mimic Proteins: Functions, Structures, and Bioinformatic Analysis. Biochemistry *53*, 2865–2874.

Wang, H.H., Isaacs, F.J., Carr, P.A., Sun, Z.Z., Xu, G., Forest, C.R., and Church, G.M. (2009). Programming cells by multiplex genome engineering and accelerated evolution. Nature *460*, 894–898.

Wang, Y., Xu, Q., Lu, H., Lin, L., Wang, L., Xu, H., Cui, X., Zhang, H., Li, T., and Hua, Y. (2015). Protease activity of PprI facilitates DNA damage response: Mn(2+)-dependence and substrate sequence-specificity of the proteolytic reaction. PLoS One *10: e01220*, 1–17.

Watanabe, T., Furuse, C., and Sakaizumi, S. (1968). Transduction of Various R Factors by Phage

P1 in Escherichia coli and by Phage P22 in Salmonella typhimurium.

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., et al. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. *46*, 296–303.

West, S.E.H., Schweizer, H.P., Dall, C., Sample, A.K., and Runyen-Janecky, L.J. (1994). Construction of improved Escherichia-Pseudomonas shuttle vectors derived from pUC18/19 and sequence of the region required for their replication in Pseudomonas aeruginosa. Gene *148*, 81–86.

Wiegand, I., Hilpert, K., and Hancock, R.E.W. (2008). Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. Nat. Protoc. *3*, 163–175.

Wilkins, B.M. (2002). Plasmid promiscuity: meeting the challenge of DNA immigration control. Environ. Microbiol. *4*, 495–500.

Wilson, K.A., Kellie, J.L., and Wetmore, S.D. (2014). DNA–protein π-interactions in nature: abundance, structure, composition and strength of contacts between aromatic amino acids and DNA nucleobases or deoxyribose sugar. Nucleic Acids Res. *42*, 6726–6741.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr. Sect. D Biol. Crystallogr. *67*, 235–242.

Yang, H., Jeffrey, P.D., Miller, J., Kinnucan, E., Sun, Y., Thoma, N.H., Zheng, N., Chen, P.-L., Lee, W.-H., and Pavletich, N.P. (2002). BRCA2 function in DNA binding and recombination from a BRCA2-DSS1-ssDNA structure. Science *297*, 1837–1848.

Yu, D., Ellis, H.M., Lee, E.C., Jenkins, N.A., Copeland, N.G., and Court, D.L. (2000). An efficient recombination system for chromosome engineering in Escherichia coli. Proc. Natl. Acad. Sci. U. S. A. *97*, 5978–5983.

Zabeau, M., Friedman, S., Van Montagu, M., and Schell, J. (1980). The ral gene of phage λ - I. Identification of a non-essential gene that modulates restriction and modification in E. coli. MGG Mol. Gen. Genet. *179*, 63–73.

Zavilgelsky, G.B. (2000). Antirestriction. Mol. Biol. *34*, 724–732.

Zavilgelsky, G.B., Kotova, V.Y., Melkina, O.E., and Pustovoit, K.S. (2014). Antirestriction activity of the mercury resistance nonconjugative transposon Tn5053 is controlled by the protease ClpXP. Russ. J. Genet. *50*, 910–915.

Zavilgelsky, G.B., Kotova, V.Y., Melkina, O.E., Balabanov, V.P., and Mindlin, S.Z. (2015). Proteolytic control of the antirestriction activity of Tn21, Tn5053, Tn5045, Tn501, and Tn402 non-conjugative transposons. Mol. Biol. *49*, 295–302.

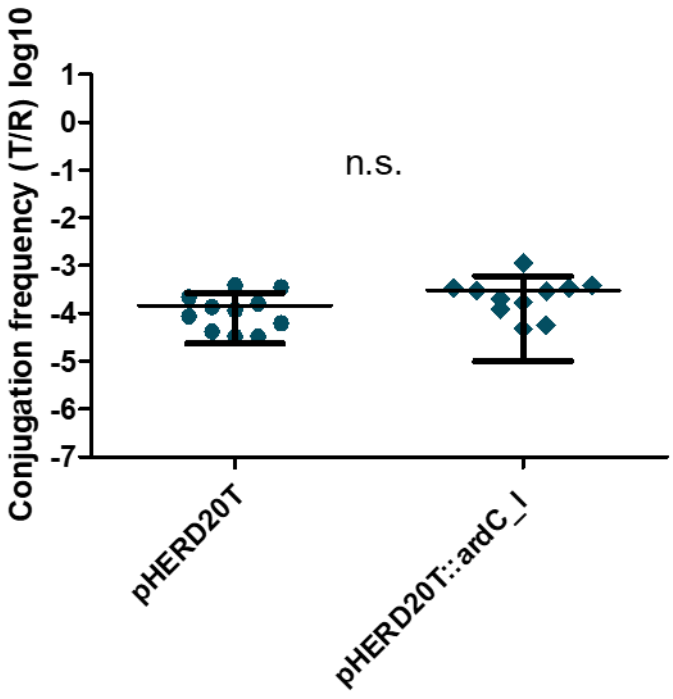
# 9 Supplementary matherial

For selecting donors, recipients and transconjugants in the conjugation experiments, different antibiotic concentrations were used bearing in mind the minimal inhibitory concentrations of each strain used.

*Supplementary Table 1. Minimum inhibitory concentration (MIC) in µg/mL for E. coli and P. putida cultured o/n at 37 °C or 30 °C in MHB medium with different antibiotics. Gm: gentamycin, Cb: carbenicillin, Tc: tetracycline, Kn: kanamycin, Cip: ciprofloxacin, MitC: mitomycin C, Amp: ampicillin, Cm: chloramphenicol, Tmp: trimethoprim, Nx: nalidixic acid. The mode of at least 3 assays for each antibiotic is shown.*

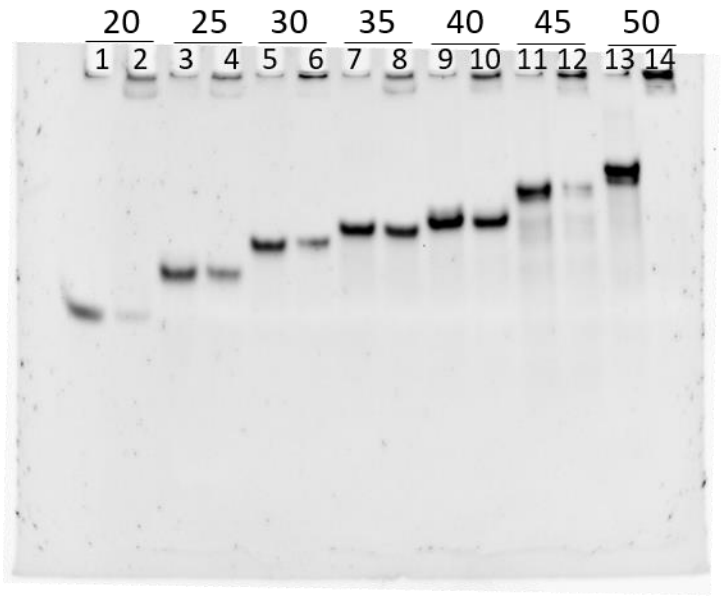
| | ***E. coli*** | | | ***P. putida*** | | |
|---|---|---|---|---|---|---|
| | BW | BW+pSU2007 | BW+pLGM25 | KT2440 | EM42 | EM422 |
| **Gm** | 1.6-2.5 | 2.5-3.12 | 3.12 | 1.2-1.6 | 0.6 | 1.6 |
| **Cb** | 12.5-15.6 | 12.5 | 6.25 | >500 | 500 | >500 |
| **Tc** | 0.6 | 0.3 | 0.3 | 1.6 | 0.6 | 7.8 |
| **Kn** | 6.25 | >500 | >500 | 2.5 | 3.12 | 0.8 |
| **Cip** | 0.3 | 0.3 | 0.6 | 0.08 | 0.16 | 0.6 |
| **MitC** | 1.2 | 0.8-1.2 | 1.2 | 6.25 | 6.25 | 12.5 |
| **Amp** | 25-31.5 | 31.25 | 25-31.5 | 500 | 62.5 | 500 |
| **Cm** | 5-6.25 | 5-6.25 | 5-6.25 | 125 | 62.5 | 250 |
| **Tmp** | 0.3 | >500 | >500 | >500 | >500 | >500 |
| **Nx** | 250 | 250 | 250 | 31.25 | 31.25-50 | 50 |

The results obtained for ardC complementation in donor cells (Figure 27) were checked by overexpression in a different vector obtaining the same results.




*Supplementary Figure 1. Effect in the conjugation frequency when complementing pLGM25 with ardC in donors. Complementation in donors with pHERD20T or pHERD20T::ardC_I. Conjugation was done for 1h at 37 °C with a final arabinose concentration of 1 % added to the mating mixture. Horizontal and vertical bars represent the mean ± SD obtained for each dataset of n=11.*

In order to check the DNA binding ability of ArdC, we performed electrophoretic mobility gel assays (EMSAs) under non-denaturing conditions for short DNA molecules of different lenght in a native polyacrylamide gel. Shorter and longer oligonucleotides were better retarded.




*Supplementary Figure 2. ArdC ssDNA-binding analysed by EMSA. 20 % polyacrylamide gel showing the retardation of oligonucleotide of the sizes shown in first line called T87I 20 nt, 25 nt, 30 nt, 35 nt, 40 nt, 45 nt and 50 nt by ArdC under non-denaturing conditions. Odd lanes (numbers in second line) are without ArdC and even lanes are with ArdC. 1 µL of ArdC at 28 µM or buffer were incubated with 2 µL of ssDNA at 2 µM and 7 µL of buffer. Gel was run 1 h 30 min at 180 V and stained with Syber gold.*