







VNIVERSITAT  
DE VALÈNCIA



ibv INSTITUTO DE  
BIOMEDICINA DE  
VALENCIA CSIC

# Structural and functional characterization of IMPACT proteins: a novel nuclease family

Sara Zamora Caballero

DIRECTOR:

Dr. Jerónimo Bravo Sicilia

PROGRAMA DE DOCTORADO: BIOQUÍMICA Y BIOMEDICINA  
Departamento de Bioquímica y Biología Molecular

A dissertation presented to the faculty of biological sciences of the University of  
Valencia in candidacy for the degree of Doctor of Philosophy

2016



D. **Jerónimo Bravo Sicilia**, Doctor en Ciencias por la Universidad politécnica de Cataluña e Investigador Principal en el Consejo Superior de Investigaciones Científicas Instituto de Biomedicina de Valencia,

**INFORMA:**

Que la presente memoria titulada **Structural and function characterization of IMPACT proteins: a novel nuclease family**, realizada por la licenciada en Bioquímica **Sara Zamora Caballero**, bajo su dirección, en el Instituto de Biomedicina de Valencia, corresponde a su tesis doctoral y autoriza mediante este escrito la presentación de la misma para la obtención del grado de Doctor.



VNIVERSITAT  
ID VALÈNCIA

Fdo. Dr. Jerónimo Bravo Sicilia



**A mi familia.**





## **AGRADECIMIENTOS**

Después de tanto tiempo por fin llega el momento ansiado, el momento de daros las gracias a todos los que, de cualquier modo, me habéis ayudado y apoyado durante este proceso.

En primer lugar me gustaría agradecer a Jero la oportunidad que me ha brindado al dejarme formar parte de su grupo de trabajo. Ha sido una experiencia increíble donde he tenido la oportunidad de formarme y crecer a todos los niveles.

En segundo lugar no podría estar más agradecida por los compañeros que he tenido. Gracias a Sandra por ser mi maestra en los comienzos de esta aventura, a Susana por toda su ayuda en cultivos y a Isabel por saber transmitirnos su experiencia y por tener siempre un buen consejo en momentos donde el futuro se vislumbraba oscuro. Gracias a Jessica, por todo su trabajo como técnico de laboratorio y por sus historietas graciosas. Gracias a Nada por tener siempre una sonrisa y un trozo de chocolate que regalar, también por enseñarme a abrir la mente frente a cosas desconocidas. Gracias a Leti por todo el trabajo técnico que ha ido realizando durante los años, por tener el TOC mejor aprovechado del mundo y por sus buenos consejos, siempre con perspectiva. Gracias también a Marcin por prestarse a discutir cualquier cosa, por ser el compañero más musical y por supuesto por convertirse en mi otra mitad dentro del laboratorio. No encuentro palabras para expresar no solo la gratitud infinita que os tengo sino también el cariño que os habéis ganado a la fuerza y si, digo bien, a la fuerza. A fuerza de compartir buenos momentos, a fuerza de sacar ideas de donde no era posible y a fuerza de aportar lo mejor de vosotros para sacar algo positivo de nuestros tormentos. Nos hemos convertido en una familia

un tanto peculiar y si he venido cada día al laboratorio con una sonrisa ha sido principalmente culpa vuestra.

Por otro lado estoy muy contenta de haber contado también con el apoyo del laboratorio de “los Albertos”. Muchas gracias por los sincrotrones compartidos, los seminarios de los lunes, los diferentes puntos de vista, las ideas, palabras de apoyo y comprensión. Sois un grupo excepcional y os deseo lo mejor en vuestras vidas. Particularmente me gustaría destacar a Laura, porque a lo tonto ya llevamos 11 años compartiendo camino y muchas experiencias no habrían sido lo mismo sin ti.

Como no podía ser de otro modo también tengo mucho que agradecer a las estudiantes que han estado a mi cargo. Gracias Lidia, Iolanda y Ana, sois unas chicas estupendas y os puedo asegurar que yo he aprendido más de vosotras que vosotras de mí. Da igual lo que os propongáis, lo conseguiréis seguro y yo siempre voy a estar muy orgullosa de vosotras.

Me gustaría extender mi gratitud a todo el personal del IBV, por el apoyo, el interés, los consejos y las sonrisas. También a la comisión de fiestas, por las cenas navideñas, las celebraciones de halloween etc.

Dejando de lado el ámbito laboral, solo tengo palabras de cariño y agradecimiento para mis “cansinas”. Gracias a Leti y María, porque todo esto empezó en la carrera cuando nos íbamos a la biblioteca a estudiar. Gracias a todas por esas preguntas inocentes sobre ciencia que acaban sacándome una sonrisa, gracias por el apoyo incondicional y constante, sin importar si estaba cerca o lejos y muchísimas gracias por tener fe y confianza en mí incluso cuando yo ya no la tenía. Gracias Puchol por aprenderte de memoria a que me dedico para presumir de mí aunque no termines de entender que es “eso” que hago.

Gracias a Fátima, Bea, Irene, Marina y Amparo por vuestros ánimos. En resumen, gracias a todas por las quedadas, viajes y risas sin fin.

Gracias a SOMRICE, por poner banda sonora a mi tesis y a Juan Carlos por su ayuda y consejos con los trámites de la tesis.

Por último y más importante, gracias a mi familia. Si todo esto ha sido posible es gracias a ellos. Gracias a Carlos, por los viajes, las cenas, conciertos y por todo tipo de distracciones que inventaste para que desconectara, por tu apoyo incondicional en momentos buenos y momentos malos y por creer siempre en mí. Lo que más te quiero agradecer es que fueras capaz de sacar lo positivo de cualquier situación y hacérmelo ver. Gracias Marta por tu apoyo y complicidad. Gracias papás por todos los sacrificios que habéis hecho en vuestra vida para que pudiéramos estudiar y por la ilusión que habéis puesto en este proyecto. Todavía me acuerdo de cuando fui por primera vez al sincrotrón y me dijisteis muy convencidos que iba a volver con la tesis hecha. Este trabajo es gracias, por y para vosotros.

Ya solo queda disculparse si alguien siente que debería estar aquí y no lo he mencionado, tanto el espacio como mi memoria tienen sus límites. Como he dicho con anterioridad, cualquier persona mínimamente involucrada en este proyecto tiene toda mi gratitud.



# INDEX

LIST OF FIGURES .....	5
LIST OF TABLES .....	9
LIST OF ABBREVIATIONS .....	11
Resumen .....	15
Summary.....	27
<b>1. INTRODUCTION</b> .....	<b>33</b>
1.1. Eukaryotic translation initiation: General overview.....	36
1.2. Translation control in eukaryotes.....	38
1.2.1 Regulation of translation initiation.....	38
1.3. General amino acid control pathway .....	40
1.3.1. General amino acid control pathway in yeast.....	40
1.3.2. The general amino acid control in mammals .....	43
1.4. IMPACT gene .....	44
1.5. IMPACT protein .....	45
1.6. Insights into IMPACT proteins function.....	48
1.6.1. IMPACT and GAAC.....	48
1.6.2. IMPACT and neuronal differentiation .....	51
1.6.3. IMPACT and cell cycle.....	51
1.6.4. IMPACT and the immune system .....	52
1.7. Nucleic acid recognition and immune response .....	54
1.8. Objectives.....	57
<b>2. METHODOLOGY</b> .....	<b>59</b>
2.1 Production of IMPACT proteins in <i>E.coli</i> expression system.....	61

2.1.1	Obtainment of expression plasmids .....	61
2.1.1.1	Yeast IMPACT homolog .....	61
2.1.1.2	<i>Chaetomium</i> IMPACT homolog .....	62
2.1.1.3	Human IMPACT.....	65
2.1.2	<i>Expression and purification of the different constructs</i> .....	65
2.2	Differential scanning fluorimetry to assess protein stability.....	67
2.3	Crystallization of <i>Chaetomium</i> IMPACT homolog domains.....	69
2.3.1	Crystallization trials .....	71
2.3.1.1	Crystallization of <i>Chaetomium</i> IMPACT homolog.....	71
2.3.1.2	Crystallization of <i>Chaetomium</i> IMPACT homolog RWD domain .....	72
2.3.1.3	Crystallization of <i>Chaetomium</i> IMPACT homologue ancient domain with DNA72	
2.3.2	<i>Crystal data collection, structure determination, refinement and analysis</i> .....	73
2.3.2.1	<i>RWD domain</i> .....	73
2.3.2.2	<i>Ancient domain</i> .....	73
2.3.2.3	<i>Co-crystallization of the ancient domain with DNA</i> .....	74
2.4	Design and expression of mutants .....	74
2.4.1	<i>Site directed mutagenesis</i> .....	74
2.5	Nucleic acid binding and hydrolysis assays.....	76
2.5.1	Agarose EMSA.....	76
2.5.2	Biolayer interferometry: Dissociation constant.....	76
2.5.3	DNase activity assays .....	77
2.5.4	Metal requirements for DNA cleavage .....	77
2.5.5	Salt requirements for DNA cleavage.....	78
2.5.6	Characterization of the DNA cleavage.....	78
2.5.7	Hyperchromicity assays .....	79
2.5.7.1	<i>Kunitz assay</i> .....	79

2.5.7.2	<i>Substrate inhibition assay</i> .....	79
2.6	Other bioinformatics tools .....	80
2.7	Cells culture and knock down assays.....	81
2.7.1	3T3 cells culture.....	81
2.7.2	siRNA transfection for IMPACT knock down .....	81
2.7.3	Protein extraction from 3T3 cells .....	83
2.7.4	RNA extraction from 3T3 cells .....	84
2.7.5	Quantitative real-time PCR.....	84
<b>3.</b>	<b>RESULTS &amp; DISCUSSION</b> .....	<b>87</b>
3.1	Expression and purification of Yih1 .....	89
3.2	Yih1 crystallization.....	92
3.2.1	Yih1 <sub>FL</sub> Data processing .....	94
3.3	IMPACT expression and purification .....	95
3.4	CIH expression and purification.....	96
3.5	<i>Chaetomium</i> IMPACT is more thermostable than Yih1 or IMPACT.....	99
3.6	Crystallization trials of CIH <sub>FL</sub> yielded ancient domain crystals due to <i>in situ</i> proteolysis .....	100
3.6.1	Data processing and refinement .....	101
3.6.2	<i>In situ</i> CIH <sub>FL</sub> proteolysis due to the presence of PEST regions.....	102
3.7	Ancient domain structural characterization.....	104
3.8	The ancient domain of CIH shows structural similarity with RNase PH domain ...	109
3.9	<i>Chaetomium</i> IMPACT homolog binds nucleic acids .....	111
3.10	Identification of IMPACT proteins as DNases.....	113
3.10.1	Point mutations abolish DNase activity in CIH .....	114
3.10.2	CIH binds both DNA and RNA with similar affinity .....	119
3.10.3	Co-crystallization of the ancient domain with DNA yields crystals with a phosphate bound to the structure .....	120

3.10.3.1	Data collection and processing	121
3.10.3.2	Structural characterization	122
3.11	IMPACT proteins are cytosolic nucleic acid sensors	127
3.12	DNase activity characterization	132
3.12.1	DNase activity in CIH requires $Mg^{2+}$ or $Mn^{2+}$	133
3.12.2	DNase activity is disrupted with high ionic strength	134
3.12.3	IMPACT proteins are endonucleases	134
3.12.4	Kunitz assay	136
3.12.5	CIH DNase activity displays substrate inhibition	138
3.12.6	The RWD domain is dispensable for the DNase activity	138
3.13	Crystallization of the RWD domain from CIH	139
3.13.1	CIH RWD domain data processing and refinement statistics	140
3.14	CIH RWD structural characterization	141
3.14.1.1	Comparison between RWD domains from GCN2 and CIH	144
3.15	Concluding remarks	147
<b>4.</b>	<b>CONCLUSIONS</b>	<b>149</b>
<b>5.</b>	<b>BIBLIOGRAPHY</b>	<b>153</b>
<b>6.</b>	<b>APPENDIX</b>	<b>169</b>
6.1	Non-related publications	171



## LIST OF FIGURES

### Introduction

Figure 1.1: mRNA translation in eukaryotes.....	35
Figure 1.2: Canonical pathway of eukaryotic translation initiation.....	37
Figure 1.3: The eIF2 kinases regulate translation in response to different stress types.....	39
Figure 1.4: Model for allosteric activation of GCN2 KD domain by its interaction with $\Psi$ KD.....	41
Figure 1.5: A model for translational control of yeast GCN4 by phosphorylation of eIF2 $\alpha$ by the protein kinase GCN2.....	42
Figure 1.6: Sequence alignment of the Yih1/IMPACT family of proteins .....	46
Figure 1.7: Proposed models for Yih1 on the ribosome.....	50
Figure 1.8: Activation of IDO expression and enzyme activity.....	53
Figure 1.9: Immune response triggered by the presence of nucleic acids in the cytosol.....	55

### Methodology

Figure 2.1: Scheme showing the constructs pBS4717, 4718, 4719 and 4720 obtained from Dr. Bertrand Séraphin.....	62
Figure 2.2: pOPINF vector map.....	63
Figure 2.3: SYPRO® Orange emission.....	67
Figure 2.4: Scheme of crystallization process.....	70

## **Results and discussion**

Figure 3.1: 10% SDS-PAGE analysis small scale purification.....	89
Figure 3.2: pBS4717 size exclusion chromatography in Superdex 200 26/60.....	90
Figure 3.3: pBS4720 size exclusion chromatography in Superdex 75 16/60.....	91
Figure 3.4: Size exclusion chromatography of IMPACT protein.....	96
Figure 3.5: CIH <sub>FL</sub> size exclusion chromatography in Superdex 200 16/60.....	97
Figure 3.6: Size exclusion elution pattern for different CIH constructs.....	98
Figure 3.7: Differential scanning fluorimetry comparison between Yih1, IMPACT and CIH.....	100
Figure 3.8: Ancient domain crystals from <i>C.terminophilum</i> IMPACT homolog.....	101
Figure 3.9: Disorder probability of CIH structure based on sequence.....	103
Figure 3.10: Ribbon representation of CIH Ancient domain.....	105
Figure 3.11: Ribbon superposition of ancient domains from YigZ and CIH.....	106
Figure 3.12: Sequence alignments among YigZ ancient domain and ancient domain of IMPACT proteins from <i>C.terminophilum</i> , <i>H.sapiens</i> and <i>S.cerevisiae</i> .....	107
Figure 3.13: Surface residue conservation among ancient domains.....	108
Figure 3.14: Three different views of the electrostatic surface potential of the CIH ancient domain.....	109

Figure 3.15: Main chain superposition of RNase PH domain from RRP46 and the ancient domain of CIH.....	110
Figure 3.16: Agarose EMSA with dsDNA and CIH.....	112
Figure 3.17: DNase activity is present in full length Yih1, IMPACT and CIH purifications.....	113
Figure 3.18: Sequence alignment of IMPACT homologs.....	115
Figure 3.19: DNase activity assays for mutants .....	117
Figure 3.20: Position of the mutations within CIH ancient domain structure and electrostatic surface.....	118
Figure 3.21: Representative BI experiments for CIH <sub>R232A</sub> affinity binding calculation to DNA and RNA.....	120
Figure 3.22: Crystal of the ancient domain obtained in presence of dsDNA.....	121
Figure 3.23: Phosphate ion bound to the ancient domain.....	124
Figure 3.24: Superposition of ribbon representations of CIH ancient domain (5fcv) and RNase PH domain from the polyribonucleotide phosphorylase (4aid) superimposition.....	125
Figure 3.25: ssDNA docking over CIH ancient domain.....	126
Figure 3.26: Histidine 235 from a symmetric molecule locates nearby the phosphate ion.....	127
Figure 3.27: IMPACT silencing with siRNA in 3T3 cells.....	128
Figure 3.28: IMPACT protein levels analysis after silencing.....	129

Figure 3.29: Relative amounts of $\beta$ -interferon mRNA levels after IMPACT silencing and ISD or poly I:C treatment.....	130
Figure 3.30: CIH requires magnesium or manganese for its DNase activity .....	133
Figure 3.31: DNase activity with different ionic strength .....	134
Figure 3.32: CIH and IMPACT show endonuclease activity .....	135
Figure 3.33:pBR322 digestion.....	136
Figure 3.34: Standard curve for Kunitz calculation.....	137
Figure 3.35: Substrate inhibition assay.....	138
Figure 3.36: DNase assay with different CIH domains.....	139
Figure 3.37: Crystal of RWD domain from <i>C.terminophilum</i> IMPACT homolog.....	140
Figure 3.38: Ribbon representation of CIH RWD domain.....	141
Figure 3.39: Temperature factors of CIH RWD backbone in worm representation .....	143
Figure 3.40: CIH RWD domain electrostatic surface potential representation .....	144
Figure 3.41: RWD protein sequence alignment.....	145
Figure 3.42: Structural superimposition between RWD domains from CIH and GCN2 proteins.....	146

## LIST OF TABLES

### **Methodology**

Table 2.1: Constructs for Yih1 <sub>FL</sub> , IMPACT and CIH.....	64
Table 2.2: Buffer used for thermofluor assay.....	68
Table 2.3: Oligonucleotides used for site directed mutagenesis.....	75
Table 2.4: Oligonucleotides used for qPCR.....	85

### **Results and discussion**

Table 3.1: Conditions where Yih1 <sub>FL</sub> crystals were obtained.....	93
Table 3.2: Additives present at the improving Yih1 <sub>FL</sub> crystallization conditions and diffracting resolution.....	94
Table 3.3: Data collection and refinement statistics for the ancient domain....	102
Table 3.4: Mutations analyzed for DNase activity.....	116
Table 3.5: Data collection and refinement statistics for the ancient domain bound to a phosphate ion.....	122
Table 3.6: Data collection and refinement statistics for the RWD domain.....	141



## LIST OF ABBREVIATIONS

<b>40S</b>	Small eukaryotic ribosomal subunit
<b>AcNa</b>	Sodium acetate
<b>ATF3</b>	Activating translation factor 3
<b>ATF4</b>	Activating translation factor 4
<b>AU</b>	Arbitrary units
<b>CCP4</b>	Collaborative Computational Project 4
<b>Cdc</b>	Cyclin dependent protein kinase (yeast)
<b>Cdk</b>	Cyclin dependent protein kinase (mammals)
<b>CHOP</b>	CCAAT enhancer binding protein homologous protein
<b>CIH</b>	<i>Chaetomium thermophilum</i> IMPACT homolog
<b>C-terminal, C-term</b>	Carboxy terminal
<b>DMSO</b>	Dimethyl sulfoxide
<b>DNA</b>	Deoxyribonucleic acid
<b>DNase</b>	Deoxyribonuclease
<b>dNTPs</b>	Deoxynucleotides triphosphate
<b>dsDNA</b>	Double stranded DNA
<b>EDTA</b>	Ethylenediaminetetraacetic acid
<b>eIF2</b>	Eukaryotic translation initiator factor 2
<b>eIF2<math>\alpha</math></b>	$\alpha$ subunit of Eukaryotic translation initiator factor 2
<b>EMSA</b>	Electromobility shift assay

<b>GAAC</b>	General amino acid control pathway
<b>GCN1</b>	General control non-derepressible protein 1
<b>GCN2</b>	General control non-derepressible protein 2
<b>GDP</b>	Guanosine diphosphate
<b>GEF</b>	Guanosine exchange factor
<b>GST</b>	Glutation-S-transferase
<b>GTP</b>	Guanosine triphosphate
<b>HisRS</b>	Histidyl-tRNA synthetase like domain
<b>HRI</b>	Heme-regulated inhibitor kinase
<b>IDO</b>	Indoleamine 2,3-dioxygenase
<b>IPTG</b>	Isopropyl $\beta$ -D-1-thiogalactopyranoside
<b>K<sub>D</sub></b>	Dissociation constant
<b>KD</b>	Kinase domain
<b>LB</b>	Lysogeny broth
<b>MMT</b>	L-malic acid: MES: Tris base buffer (1:2:2)
<b>MR</b>	Molecular replacement
<b>mRNA</b>	Messenger RNA
<b>MW</b>	Molecular weight
<b>NCS</b>	Nervous central system
<b>Ni-NTA</b>	Nickel-nitriloacetic acid
<b>NMR</b>	Nuclear magnetic resonance
<b>N-terminal, N-term</b>	Amino terminal



<b>OD</b>	Optical density at 600nm
<b>PCR</b>	Polymerase chain reaction
<b>PDB</b>	Protein data bank
<b>PEK</b>	Pancreatic eIF2 $\alpha$ kinase
<b>PF01205</b>	Protein family 01205
<b>PKR</b>	RNA dependent kinase
<b>qPCR</b>	Quantitative real time PCR
<b>RNA</b>	Ribonucleic acid
<b>RNase</b>	Ribonuclease
<b>RPL39</b>	Large ribosomal protein 39
<b>RPS2</b>	Ribosomal protein 2 from <i>E.coli</i>
<b>RT</b>	Room temperature
<b>RT-PCR</b>	Reverse transcription PCR
<b>SDS-PAGE</b>	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
<b>siRNA</b>	Small interference RNA
<b>SPG</b>	succinic acid: sodium dihydrogen phosphate: glycine buffer (2:7:7)
<b>ssDNA</b>	Single stranded DNA
<b>ssRNA</b>	Single stranded RNA
<b>TLS</b>	Translation/Libration/Screw
<b>Tm</b>	Melting temperature

<b>TOR</b>	Target of rapamycin
<b>tRNA</b>	Transfer RNA
<b>uORF</b>	Upstream open reading frame
<b>UPF0029</b>	Unknown protein family 0029
<b>UV</b>	Ultraviolet
<b>WB</b>	Western blot
<b>YIH1</b>	Yeast IMPACT homolog 1
<b>ΨKD</b>	Pseudo kinase domain

## Resumen

---



Las proteínas IMPACT, objeto de estudio en este trabajo, se encuentran en la mayoría de tejidos y tipos celulares, pero los niveles de expresión más elevados se encuentran en fibroblastos y en células del sistema nervioso central. El término IMPACT viene de la combinación de los términos en inglés “*imprinted*” y “*ancient*”. Por un lado, hoy en día el gen *Impact* es el único gen conocido en el cromosoma 18 que presenta impronta genética en roedores. La impronta genética es un mecanismo de herencia no mendeliano único en mamíferos. En el caso del gen *Impact*, esta impronta exhibe un patrón específico de especie ya que está presente en roedores pero no en el resto de mamíferos. La impronta en roedores se ha relacionado con la ausencia de islas CpG intrónicas en el promotor del gen, ya que el promotor humano constituye una isla CpG que convencionalmente no está metilada. Dado que el gen murino que ha sufrido impronta presenta una mayor expresión en comparación con el gen humano, se ha propuesto un modelo donde la impronta ha evolucionado como una adaptación al aumento de dosis génica en roedores. Por otro lado, las proteínas IMPACT se encuentran organizadas en dos dominios: un dominio amino terminal conocido como RWD y un dominio carboxilo terminal al que se le ha denominado como dominio ancestral debido a que presenta un elevado grado de conservación secuencial entre todos sus homólogos. Sin embargo, actualmente se desconoce si este dominio desempeña alguna función por sí mismo. El área que conecta ambos dominios se prevé que no posea estructura secundaria.

Actualmente hay poca bibliografía relacionada con las proteínas IMPACT, dado que el papel de dichas proteínas dentro de la célula es poco conocido. Hasta la fecha, la única función conocida de las proteínas IMPACT es su rol en el control de la traducción bajo condiciones de estrés, particularmente en ayuno.

La traducción es un proceso crucial para la vida. De hecho, la adaptación a diferentes tipos de estrés que puedan comprometerla marca la diferencia entre la

---

vida y la muerte. En eucariotas el elevado número de factores que intervienen en este proceso permiten que se regule la traducción en base a señales extracelulares. El inicio de la misma es el principal punto de regulación del proceso.

La traducción comienza con el reclutamiento de un ARN de transferencia iniciador ( $\text{tRNA}^{\text{Met}}$ ), capaz de interactuar con el codón de inicio del ARN mensajero (mRNA). Para poder llegar a ese punto, en primer lugar el complejo terciario (TC), que está formado por las tres subunidades del factor de iniciación de la traducción 2 (eIF2) en eucariotas  $\alpha$ ,  $\beta$  y  $\gamma$  unido a GTP y al  $\text{tRNA}^{\text{Met}}$ , se une a la subunidad 40S del ribosoma formando el complejo pre-iniciador 43S. Los factores de traducción 3 y 4 son necesarios para que el complejo 43S pueda unirse al mRNA, ya que reclutan a la cadena de mensajero y relajan la estructura secundaria cerca del extremo 5'. Una vez unido, el complejo 43S rastrea el mensajero en busca del codón de inicio. Cuando el complejo 43S reconoce el codón iniciador, el GTP del TC se hidroliza y se libera eIF2-GDP. Es entonces cuando la subunidad 60S del ribosoma se une dando lugar al complejo iniciador 80S (pág. 39). El reciclaje de eIF2-GDP a eIF2-GTP es un punto crucial si se quiere asegurar unos ratios de inicio de la traducción elevados. Es por ello que existe un factor intercambiador de GDP, conocido como eIF2B, que ayuda a dicho reciclaje.

La fosforilación del factor de iniciación de la traducción 2 (eIF2) en el residuo serina 51 de su subunidad alfa es una de las vías más utilizadas para reprimir la traducción, ya que el factor encargado del reciclaje de eIF2-GDP a eIF2-GTP se une con mayor afinidad a la forma fosforilada de la proteína. De este modo, eIF2B quedaría secuestrado por eIF2 fosforilado y el reciclaje a eIF2-GTP se vería disminuido.

En mamíferos se han descrito 4 quinasas capaces de llevar a cabo dicha fosforilación y cada una de esas quinasas se ve activada ante diferentes situaciones de estrés como infecciones víricas, estrés oxidativo, choque térmico, estrés en el retículo endoplásmico o limitación de nutrientes (pág. 41). La quinasa GCN2 interviene en la regulación en condiciones de ayuno y la respuesta que desencadena se conoce como ruta general de control de aminoácidos.

La ruta general de control de aminoácidos se activa cuando un ARN de transferencia vacío llega al sitio P del ribosoma y es transferido desde el ribosoma hasta el dominio hisRS de GCN2. Este movimiento está mediado por una proteína conocida como GCN1. La unión de GCN2 al ARN de transferencia vacío hace que se produzca un cambio conformacional que desemboca en la activación del dominio quinasa de GCN2, encargado de la fosforilación del eIF2 (pág. 43). Esta fosforilación desencadena la activación de rutas biosintéticas de aminoácidos y en el caso de mamíferos también promueve la expresión de transportadores de aminoácidos que aumentarán el flujo de captación de estos desde el exterior. Las proteínas IMPACT compiten con GCN2 por la unión a GCN1, ya que tanto IMPACT como GCN2 poseen un dominio RWD que es capaz de interactuar con GCN1 en la misma zona. Al evitar la fosforilación de eIF2, IMPACT estaría manteniendo unos niveles continuos de traducción. Dado sus elevados niveles de expresión en el sistema nervioso central, IMPACT podría intervenir en la regulación de la traducción de algunas neuronas específicas que se encuentren en condiciones de escasez de aminoácidos.

Por otro lado, se ha descrito muy recientemente que existe cierta relación entre el ciclo celular y las proteínas IMPACT. Estudios en levadura revelan que, cuando se elimina el homólogo de IMPACT (Yih1), las células tienden a acumularse en las fases G2/M del ciclo celular. Se vio que la ciclina dependiente de quinasas 28 (Cdc28) precipitaba conjuntamente con Yih1 de igual manera que

---

lo hacían sus homólogos en humano IMPACT y CDK1 respectivamente, apuntando a que la interacción se ha conservado durante la evolución. Sin embargo, el rol preciso desempeñado por IMPACT en el ciclo celular se desconoce.

Por último, se ha descrito la relación entre la inducción del enzima indoleamina 2 3-dioxygenasa (IDO) y la sobre expresión de IMPACT debido a la reducción de triptófano que conlleva la actividad de IDO. Dicha enzima se activa tras la exposición de las células a interferón gamma y participa en el catabolismo del triptófano para producir unos compuestos tóxicos conocidos como kineurinas. Teniendo en cuenta el papel de IMPACT en la traducción, podría ayudar a las células a superar dicha situación de escasez de triptófano.

## **Objetivos**

Con estas premisas, se establecieron una serie de objetivos a culminar durante el desarrollo de la tesis:

1. Determinar la estructura tridimensional de la proteína IMPACT.
2. Identificación de nuevas funciones de IMPACT.
3. Integración de las nuevas capacidades de IMPACT en la célula.

## **Resultados**

### Caracterización estructural del dominio ancestral

Dada la experiencia y trayectoria del grupo donde se ha realizado el trabajo, la técnica escogida para la determinación de la estructura tridimensional de la proteína IMPACT fue la cristalografía de proteínas y difracción por rayos X.



En primer lugar, se puso a punto la sobreexpresión y purificación de Yih1 (homólogo de IMPACT en *S.cerevisiae*) e IMPACT (*H.sapiens*) en un sistema de expresión heterólogo como es *Escherichia coli*. A continuación, se procedió a realizar los ensayos de cristalización. Aunque dichos ensayos fueron infructuosos en el caso de la proteína humana, en el caso de Yih1 si se obtuvieron cristales. Dichos cristales, pertenecientes al grupo espacial I222, mostraban una difracción de escasa resolución y baja calidad. Para mejorar la calidad de los cristales obtenidos se utilizaron diversas aproximaciones y aunque se consiguió una mejora en tamaño y apariencia, la calidad y resolución de la difracción no fueron suficientes para poder resolver la estructura.

Dada la reciente tendencia de utilizar organismos termófilos para obtener las estructuras de ciertas proteínas o complejos proteicos de difícil cristalización, se procedió al clonaje, expresión y purificación del homólogo de IMPACT en *Chaetomium thermophilum*, al que se decidió llamarle CIH (del inglés *Chaetomium* IMPACT homolog). Tras analizar que efectivamente tenía una mayor temperatura de desnaturalización en comparación con los homólogos de levadura y humano (pág. 102) y por tanto mayor estabilidad, se procedió con los ensayos de cristalización. Dichos ensayos, se efectuaron utilizando la técnica de difusión de vapor en placas de gota sentada. Se obtuvieron cristales que difractaban a 2.2Å de resolución pero cuyas dimensiones no eran suficientes para albergar una molécula completa de proteína en la unidad asimétrica. Tras resolver el problema de las fases por reemplazo molecular usando el modelo 2cve disponible en el PDB (del inglés Protein Data Bank), se pudo observar que la proteína había sufrido un proceso de proteólisis dentro de la gota de cristalización y que el dominio cristalizado era el conocido como dominio ancestral. Dicho dominio presentaba la topología predicha,  $\beta\alpha\beta\beta\alpha$  (pág. 107). Tras el análisis estructural del mismo se observó, no sin asombro, cierto grado de

---

homología estructural con el dominio RNasa PH (pág.112). Este dominio se encuentra presente en algunas proteínas con capacidad ribonucleolítica, como lo son las proteínas que componen el exosoma bacteriano o en el enzima polinucleotidil fosforilasa.

### Las proteínas IMPACT son enzimas

El hecho de que el dominio ancestral presentase cierta homología estructural con el dominio RNasa PH nos condujo a investigar si CIH tenía la capacidad de unir ADN. Para ello se realizó un retardo en gel de agarosa y se observó que efectivamente, CIH era capaz de unir ADN de doble cadena y que el dominio ancestral por si solo era capaz de unir dicho ácido nucleico. Tras esta comprobación, se prosiguió realizando una pequeña prueba en la que comprobamos que las preparaciones de IMPACT, Yih1 y CIH degradaban ADN lineal de doble cadena (pág. 115). Esta actividad no había sido descrita con anterioridad y de confirmarse asignaría a las proteínas IMPACT la categoría de enzimas.

Para comprobar que la actividad DNasa provenía efectivamente de las proteínas IMPACT, se diseñaron toda una serie de mutantes sobre residuos localizados en zonas potenciales de unión al ADN o de actividad catalítica. Tras la obtención de diferentes mutantes que mostraban nula o reducida actividad DNasa (pág. 118), se procedió a la caracterización de dicha actividad. Durante los ensayos de caracterización, se ha utilizado siempre por cuestiones de mayor estabilidad y solubilidad el homólogo de *C.thermophilum*. Los experimentos realizados nos permitieron concluir que nos encontrábamos ante un enzima con capacidad endolítica (pág. 137), que producía los cortes en las dos cadenas de ADN de manera secuencial (pág. 138) y cuya actividad se veía inhibida en presencia de concentraciones de cloruro sódico igual o superiores a 0,6 M (pág.

136). Además, tras un análisis de hipercromicidad con diferentes concentraciones de ADN, pudimos observar que el enzima presentaba un comportamiento de inhibición por sustrato (pág. 140). Asimismo, una caracterización más exhaustiva de la capacidad de unión de CIH a ácidos nucleicos reveló que la proteína unía con afinidad muy similar ADN de cadena doble, ADN de cadena simple y ARN de cadena simple, sin mostrar especificidad alguna de secuencia. De este modo, se definió a CIH como una proteína con capacidad de unir ácidos nucleicos de manera universal (pág. 122). Por otro lado, un ensayo enzimático utilizando los diferentes dominios de CIH reveló que el dominio RWD no es necesario para llevar a cabo la actividad nucleasa (pág. 141).

Paralelamente, se procedió a intentar cristalizar la proteína en presencia de ADN, utilizando tanto la proteína completa como el dominio ancestral. En este último caso se obtuvieron cristales que difractaron a una resolución de 1.5Å y aunque la estructura cristalizada no se encontraba unida a ADN, si se encontró un fosfato unido en una zona equivalente a la que se observa el fosfato unido en el enzima polinucleótido fosforilasa (PNPasa) y que ha sido descrito como su centro activo.

#### Las proteínas IMPACT se comportan como sensores citosólicos de ácidos nucleicos

Los resultados comentados anteriormente junto con la relación descrita en la bibliografía existente entre IMPACT e IDO, llevaron a plantear la posibilidad de que existiera algún tipo de relación entre IMPACT y el sistema inmune más allá de lo actualmente descrito. Para ello se realizaron ensayos ex vivo con fibroblastos de ratón (línea celular 3T3). En dichos ensayos se transfectaba ARN de interferencia específico para IMPACT o sin diana concreta para las células control, junto con un ADN de doble cadena de 45 pares de bases con una

---

secuencia perteneciente al genoma de *Listeria monocytogenes* (ISD), que estimula la producción de interferón o un análogo de ARN de doble cadena (poli I:C). En estos ensayos se observó que los niveles de interferón  $\beta$  producidos por dichos fibroblastos tras la transfección de ISD o poli I:C se veían alterados en células donde se había silenciado la expresión de IMPACT con respecto a las células control. Además, se observó una respuesta opuesta dependiendo de si se transfectaba ADN de doble cadena (ISD) o ARN de doble cadena (poli I:C). En concreto, tras la transfección con ISD, las células donde se había silenciado la expresión de IMPACT mostraban que la cantidad de ARN mensajero de interferón  $\beta$  caía a unos niveles inferiores a la mitad de los observados en las células control (pág. 132). Para el caso del poli I:C se observó una respuesta totalmente opuesta. En las células donde se había silenciado IMPACT, la transfección con poli I:C produjo un aumento en la cantidad de ARN mensajero de interferón  $\beta$  por encima de tres veces el valor observado en las células control.

### Caracterización estructural del dominio RWD

Por otro lado, teniendo en cuenta que la resolución estructural de la proteína completa fue infructuosa se diseñó una construcción que abarca el dominio RWD de CIH para su caracterización estructural. Los cristales obtenidos difractaron a una resolución de 1.4Å y debido a que el problema de las fases no pudo ser resuelto mediante reemplazo molecular usando un modelo del dominio RWD de GCN2 murino (1ukx) obtenido por resonancia magnética nuclear y disponible en el PDB, se procedió a la obtención de las fases ab initio mediante el uso del programa informático Arcimboldo y con la colaboración de la Dra. Isabel Usón. La estructura del dominio RWD de CIH (pág. 144) presenta una elevadísima similitud estructural con el modelo disponible del dominio RWD murino aunque la secuencia no se encuentre conservada. La combinación de los datos ya publicados sobre la caracterización de la interacción entre GCN1 y Yih1

junto con el análisis del potencial electrostático en superficie de la estructura de CIH, apuntan a que la interacción entre GCN1 e IMPACT o cualquiera de sus homólogos tiene un gran componente electrostático.

### **Conclusiones**

En base al trabajo realizado en esta tesis y a los resultados obtenidos, hemos extraído las siguientes conclusiones:

1. CIH es una proteína con capacidad universal de unir ácidos nucleicos. Además une ADN y ARN con afinidades muy similares y sin aparente especificidad de secuencia.
2. El dominio ancestral de CIH presenta cierto grado de homología estructural con el dominio RNasa PH.
3. El dominio ancestral de CIH une un ion fosfato en la zona equivalente a la que se encuentra el ion unido en el dominio RNasa PH y que ha sido descrito como su centro activo.
4. Las proteínas IMPACT son enzimas. En concreto, son DNAsas con actividad endolítica con un mecanismo enzimático asistido por un nucleófilo donde los iones magnesio o manganeso son esenciales.
5. El corte que produce CIH sobre ADN de doble cadena es secuencial.
6. CIH muestra un valor Kunitz muy bajo que viene dado por las condiciones del ensayo.
7. La actividad DNasa de CIH muestra un patrón de inhibición por sustrato. Entre las concentraciones probadas el óptimo se encontró en

---

100µg/mL. Por otro lado, la actividad DNasa se ve inhibida en presencia de concentraciones de sal por encima de 600mM.

8. Diferentes mutantes de CIH como H203A, R232A, Y253A, C221A, R262G, R271A, D223A/D224A o D223A/E226A muestran actividad nucleasa no detectable frente a ADN.

9. Las proteínas IMPACT se comportan como sensores citosólicos de ácidos nucleicos en fibroblastos, influenciando los niveles de mRNA para INF-β en respuesta a la presencia de ADN de doble cadena o de un análogo de ARN de doble cadena.

10. El dominio RWD no es necesario para la actividad DNasa de CIH.

11. El dominio RWD mantiene una estructura tridimensional muy conservada a pesar de la baja homología de secuencia.

12. Dada la fuerte carga negativa presente en la zona del dominio RWD descrita como necesaria para la interacción con GCN1, la naturaleza de dicha interacción debe ser de carácter electrostático.

## **Summary**

---





IMPACT proteins are known to be present in almost all cell types, but they are highly expressed in fibroblasts and central nervous system. The term IMPACT comes from the fact that the Impact gene in rodents is the only imprinted gene found at chromosome 18 up to date. This imprinting is not found in other mammalian species. The protein is divided in two domains, an RWD domain at the N-terminus and the ancient domain at the C-terminus. Between those domains there is a linker region with no predicted structure. All IMPACT proteins display very high sequence conservation, mostly on its ancient domain. However, there is no known function of the protein related to the ancient domain.

Little is known about the role of IMPACT proteins within the cell. Nowadays, the only well-known function of IMPACT proteins is to assure translation levels upon amino acid starvation conditions. Protein translation is a crucial process for life. Hence, adaptation to stress conditions that compromise translation is a matter of survival. IMPACT proteins are able to compete with GCN2 for the binding to GCN1, since both GCN2 and IMPACT harbour an RWD domain able to interact with GCN1. GCN2 binding to GCN1 is necessary to activate the kinase domain of GCN2. The activation of the kinase domain results in eIF2 phosphorylation at serine 51 on its alpha subunit. Phosphorylation of eIF2 $\alpha$  leads to general translation repression but it activates the expression of some transcription factors involved into cell remediation and amino acid biosynthetic pathways. Since it is overexpressed in central nervous systems it may be involved in the regulation of translation in some specific neuronal cells upon amino acid starvation. On the other hand, it has been described that IDO induction is able to provoke IMPACT overexpression by tryptophan depletion. IDO activity is based on tryptophan catabolism into toxic compounds known as kinenures. Regarding its role in translation regulation, IMPACT proteins would help cells to overcome tryptophan depletion.

---

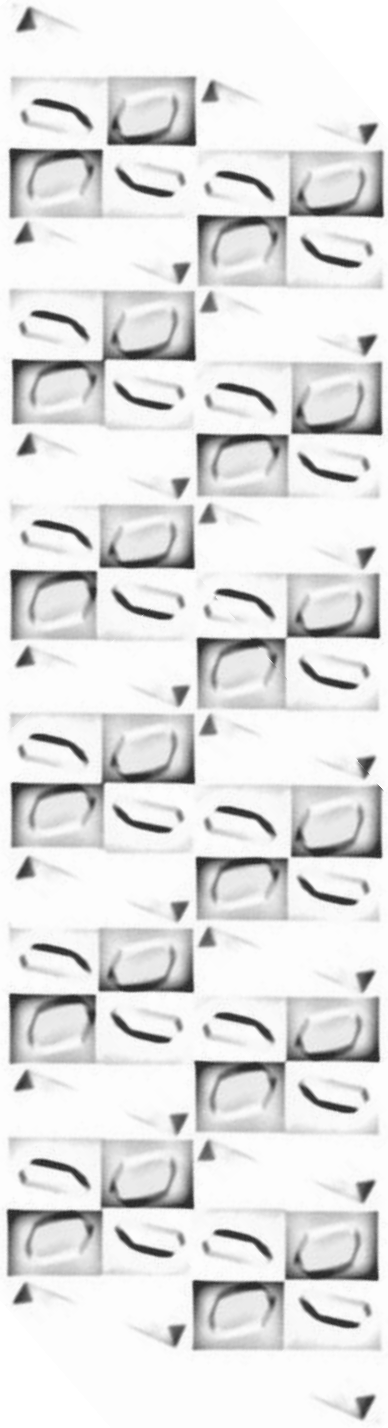
Recent studies are trying to shed light into new possible roles of IMPACT in the cell biology have linked IMPACT proteins with cell cycle, since it has been demonstrated that IMPACT and its yeast counterpart Yih1 are able to interact with CDK1 or *cdc28* respectively, although the particular function of the interaction remains unknown.

In this work we describe for the first time the three dimensional structure of the IMPACT RWD and ancient domains from a thermophilic fungus known as *Chaetomium thermophilum*. We have proved that IMPACT proteins are able to bind both, RNA and DNA *in vitro*. Extensive biochemical analysis from human, *S.cerevisiae* and *C.thermophilum* and detailed structural analysis of the obtained structures lead us to the identification of IMPACT proteins as enzymes that are able to endolytically cleave DNA. Besides, further characterization of the reaction revealed that the cleavage was not simultaneous in both strands and that the enzyme displays substrate inhibition. Given the cytosolic location of the protein, we wonder if IMPACT proteins would be related to immune system. *In vitro* experiments also demonstrated that IMPACT proteins act as cytosolic DNA sensors, since IMPACT knock down in mouse fibroblasts affects to interferon beta mRNA levels after transfection with ISD. In brief, these findings open a new field for the study of the roles of IMPACT proteins within cell biology.

**“La verdadera ciencia enseña, sobre todo, a dudar y ser ignorante.”**

Miguel de Unamuno.



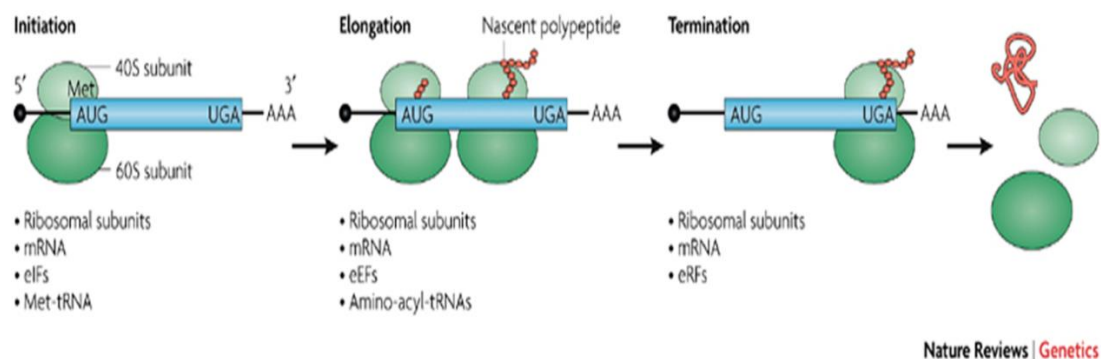


# 1. INTRODUCTION



Inside the cell, deoxyribonucleic acid (DNA) encodes all the functional information needed for life. To execute this information DNA needs to be transcribed into ribonucleic acid (RNA). Some of those RNA's are functional by themselves, but there is a fraction of RNA that needs to be translated and converted into an amino acid sequence. This directionality on information flow has been historically known as the central dogma of molecular biology (CRICK, 1958).

Translation is a complex process essential for life. In order to achieve a reliable translation of messenger RNA (mRNA) to protein, many factors are required. Translation takes place in three different steps: Initiation, elongation and termination (Figure 1.1).



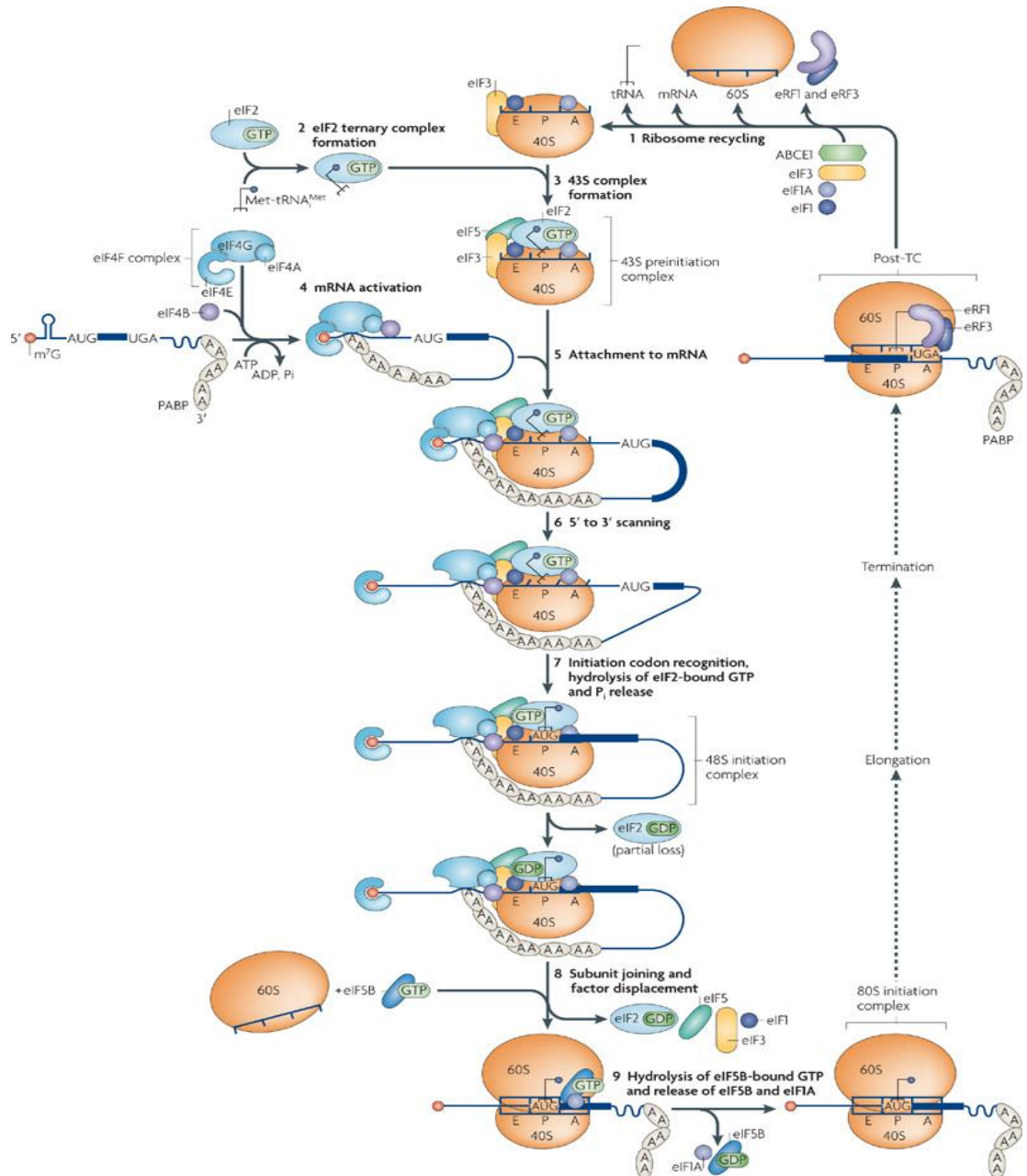
**Figure 1.1. mRNA translation in eukaryotes.** Three stages of mRNA translation in eukaryotes are shown. Open reading frames are depicted in blue. Untranslated regions are shown as black lines being the 5' cap structure drawn as black circle. AAA indicates the poly(A)-tail. Ribosomal subunits (40S and 60S) are depicted in green and the polypeptide in red. Essential components of the translational machinery are indicated for each step. Abbreviations are detailed right after: eEFs, eukaryotic elongation factors; eIFs, eukaryotic initiation factors; eRFs, eukaryotic release factors; Met, methionine (taken from Scheper et al., 2007).

---

## 1.1. Eukaryotic translation initiation: General overview

The initiation of protein synthesis consists in the recruitment of an initiator transfer RNA (tRNA) complex to the initiation codon of a messenger RNA (mRNA). Eukaryotes have taken advantage of the evolution of novel mRNA structures, to develop new mechanisms for the recruitment of the ribosome to the mRNA (Scheper *et al.*, 2007). In that sense we find that the ternary complex (TC), composed by the three subunits ( $\alpha$ ,  $\beta$  and  $\gamma$ ) of the eukaryotic translation initiation factor 2 (eIF2) complexed with GTP and the initiator tRNA<sup>Met</sup>, binds to the 40S ribosomal subunit forming a pre-initiator complex 43S. Translation factors eIF4 and eIF3 are essential since they recruit the mRNA and relax the secondary mRNA structure nearby the 5' capped end. The 43S complex binds the mRNA close to the 5' capped end, and although it has natural capacity to bind mRNA, the 43S complex needs the activity of eIF4 and eIF3 to do it (Sonenberg and Hinnebusch, 2007). The ribosomal subunit with the ternary complex scans the mRNA looking for the first AUG codon. After reaching this codon, the 43S pre-initiator complex binds the ribosomal 60S subunit, forming the 80S initiation complex. Once the AUG codon has been recognized, the GTP from the TC gets hydrolysed and eIF2-GDP is released. Recycling eIF2-GTP is crucial if rapid translation initiation wants to be assured, so there is a GDP exchange factor known as eIF2B that helps recycling eIF2-GDP to eIF2-GTP, hence allowing the availability of eIF2-GTP for a new 43S ribosome ready to start the translation process (Jackson *et al.*, 2010) (Figure 1.2).





**Figure 1.2. Canonical pathway of eukaryotic translation initiation.** The model for the canonical pathway of eukaryotic translation initiation comprises 9 steps. mRNA has been depicted in blue with 5' capped end and a poly-A tail. Ribosomes are depicted in light orange. All the factors involved have been named at the picture (taken from (Jackson *et al.*, 2010)).

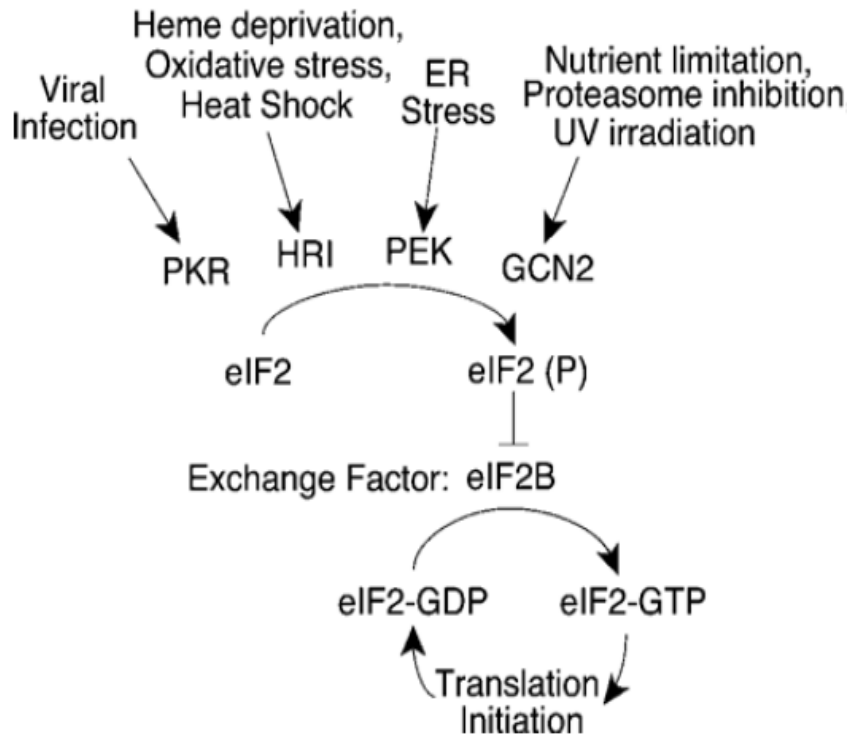
---

## 1.2. Translation control in eukaryotes

The complex composition of the ribosome allows a very tight regulation. The huge amount of factors required during the whole translation process allows translation control depending on extracellular signals that might, for example, phosphorylate some translation factors important for translation regulation (Sonenberg and Hinnebusch, 2009). Translation initiation is the main stage where protein synthesis is regulated (Jackson *et al.*, 2010).

### 1.2.1 Regulation of translation initiation

One of the crucial events in the control of protein synthesis is the eIF2 phosphorylation on serine 51 located in the  $\alpha$  subunit (eIF2 $\alpha$ ). In response to environmental stresses there are several kinases that phosphorylate eIF2, leading to the activation of several pathways that remediate the damage or induce apoptosis. In mammals there are four known kinases that phosphorylate eIF2 at the residue 51 (eIF2 $\alpha$ -P). Each kinase acts in different stress situations. The RNA dependent protein kinase (PKR) is activated by viral infections and participates in the anti-viral defense mediated by interferon. There is another kinase, haeme-regulated inhibitor (HRI) which is activated by heat shock, oxidative stress and haeme deficiency in erythroid tissues. Pancreatic eIF2 $\alpha$  kinase (PEK) is activated in response to misfolded proteins in the endoplasmic reticulum. The last kinase to mention is the general control non-derepressible-2 (GCN2) also known in mammals as eukaryotic translation initiation factor 2 alpha kinase 4 (Eif2 $\alpha$ K4). From now on we will refer to it as GCN2. GCN2 has been identified as the sole eIF2 $\alpha$  kinase found in *Saccharomyces cerevisiae* until now and it is known to be induced by nutrient deprivation, UV irradiation and proteasome inhibition (Wek *et al.*, 2006) (Figure 1.3).



**Figure 1.3. The eIF2 kinases regulate translation in response to different stress types.** The four known kinases in mammals phosphorylate eIF2 at Ser-51 in response to different environmental stresses. The phosphorylation inhibits eIF2B GEF activity reducing translation initiation levels. (Wek *et al.*, 2006)

GCN2 is a protein composed by 1659 amino acids which harbours 4 different domains: an RWD domain which is necessary for protein interaction, a pseudo-kinase domain ( $\Psi$ KD), necessary for the activation of the following kinase domain (KD), responsible for eIF2 phosphorylation in starved cells, a histidyl-tRNA synthetase-like domain (HisRS) and a C-terminal domain involved in ribosome interaction (Figure 3) (Donnelly *et al.*, 2013). GCN2 has been shown to interact not only with translating ribosomes but also with free ribosomal subunits (Ramirez *et al.*, 1991).

---

Now that the kinases involved in eIF2 $\alpha$  phosphorylation have been presented, the next issue to address is how this phosphorylation is affecting the translation initiation process. As mentioned at the initiation step (section 1.1), eIF2-GTP-tRNA<sup>MET</sup> complex is crucial for the initiation to occur and there is one GEF protein that recycles eIF2-GTP from eIF2-GDP. This protein, named eIF2B, is therefore essential for translation initiation. eIF2B also has the ability to bind eIF2 $\alpha$ -P with higher affinity than un-phosphorylated eIF2. In the event of eIF2 phosphorylation, eIF2B will not be able to exchange the guanine in eIF2 $\alpha$ -P and, in addition, any remaining un-phosphorylated eIF2-GDP will not be recycled to eIF2-GTP since eIF2B would be sequestered by eIF2 $\alpha$ -P leading to suppression or acute deceleration of translation initiation.

### 1.3. General amino acid control pathway

The environment where we live changes continuously. Unfortunately nutrient availability is not always assured and an appropriate response to this particular stress condition is essential for survival. There are different pathways that recognize the deficiency on nutrient availability and coordinate a response that implies changes in gene expression in order to alleviate the particular deficiency and be able to survive.

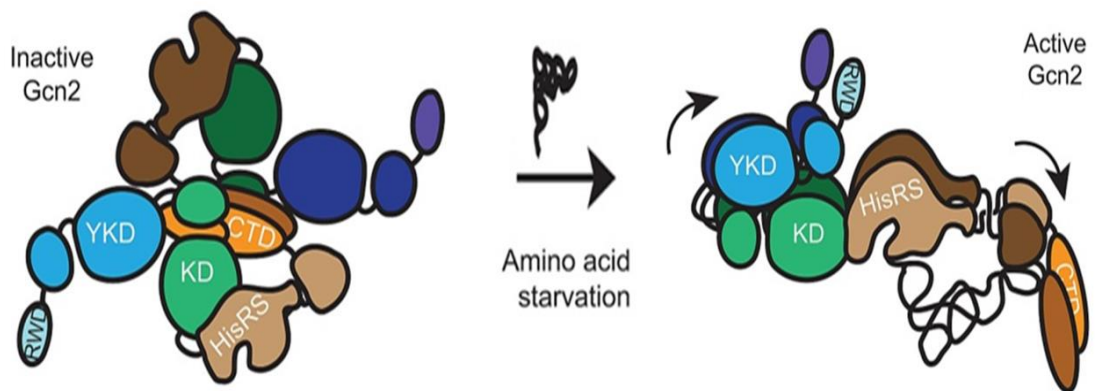
We are particularly interested in the general amino acid control (GAAC) pathway since the protein object of this study is involved in its regulation.

#### 1.3.1. General amino acid control pathway in yeast

Amino acid starvation triggers the accumulation of empty tRNA inside the cell. In this situation, empty tRNAs go inside the translating ribosome, where general control non derepressible protein 1 (GCN1), facilitates the shift of the empty tRNA from the A site of the ribosome to GCN2 hisRS domain (Garcia-Barrio

*et al.*, 2000) (Kubota *et al.*, 2000) (Kubota *et al.*, 2001).

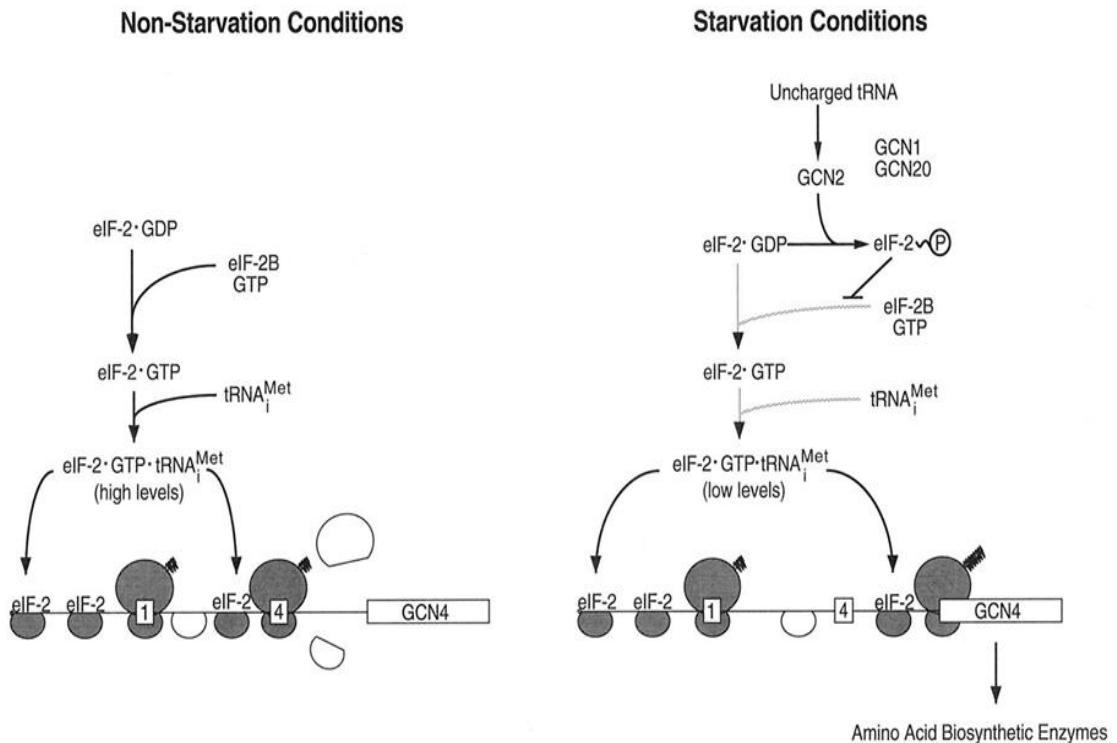
The binding of the tRNA to the HisRS induces conformational change that results in an increased accessibility of the  $\Psi$ KD to sites of allosteric activation present in the KD and thus, the activation of the GCN2 kinase domain which phosphorylates the eIF2 $\alpha$  (Lageix *et al.*, 2014) (Figure 1.4).



**Figure 1.4. Model for allosteric activation of GCN2 KD domain by its interaction with  $\Psi$ KD.** The interaction with empty tRNA provokes a conformational change that activates the GCN2 kinase domain.  $\Psi$ KD has been marked at the figure as YKD. Modified from (Lageix *et al.*, 2014)

As we have explained before, this phosphorylation is inhibiting the recycle of eIF2-GDP to eIF2-GTP and thus, the translation initiation. In this situation, GCN4 protein starts expressing, since the translation of this protein is inversely related to the amount of eIF2. This can be explained because GCN4 mRNA has 4 upstream open reading frames (uORF). The presence of uORF 1 and uORF 4 is necessary for a total prevention of GCN4 translation. When there is plenty of TC available for translation, ribosomes will translate uORF 1 and they will reinitiate translation at uORF 4 and dissociate before reaching AUG from GCN4 and hence preventing its translation. However under starvation conditions, available TC levels are low and the ribosome, which starts scanning after uORF 1 translation,

will pass by uORFs 2, 3 and 4 without translating them and will reinitiate the translation at GCN4 (Ramirez *et al.*, 1991)(Hinnebusch, 1997)(Figure 1.5) (Abastado *et al.*, 1991). Removal of the 4 uORFs leads to high GCN4 levels in starvation and non-starvation conditions (Mueller and Hinnebusch, 1986).



**Figure 1.5. A model for translational control of yeast GCN4 by phosphorylation of eIF2 $\alpha$  by the protein kinase GCN2.** When eIF2-GDP can be rapidly recycled to eIF2-GTP, the fourth ORF can be read and thus the ribosome disassembles before reaching GCN4 ORF (left panel). When low levels of eIF2-GTP are present, ORF4 is not read and ribosome reaches GCN4 ORF ready to initiate translation (right panel). ORF2 and ORF3 are not represented at the figure. Figure from (Hinnebusch, 1997)

GCN4 is a transcriptional activator of amino acid biosynthetic enzymes, and thus its expression activates the amino acid biosynthetic pathways, relieving the lack of amino acids (Ramirez *et al.*, 1991) . GCN4 is also involved in starvation-

induced autophagy (Tallóczy *et al.*, 2002).

### 1.3.2. The general amino acid control in mammals

Mammalian amino acid homeostasis is more complex than yeast homeostasis because there are 8 essential amino acids that cannot be synthesized by the cells. Those amino acids are obtained from the diet, so eIF2 $\alpha$  phosphorylation is not only controlling the activation of amino acid biosynthetic pathways, but is also regulating mammalian target of rapamycin (mTOR) and autophagy (Chen *et al.*, 2014).

GAAC pathway in mammals is also based in empty tRNA binding to GCN2 that leads to eIF2 phosphorylation. This phosphorylation also leads to a general translation repression and an enhanced translation of selected mRNAs. The way GCN4 counterpart, activating transcription factor 4 (ATF4), gets an enhanced expression is very similar to GCN4 overexpression. In this case, ATF4 mRNA has two uORFs and deficiency in TC availability allows the small ribosome subunit to reach scanning the ATF4 starting codon and initiate the translation. ATF4 expression entails the induction of other transcriptional regulators as activating translation factor 3 (ATF3) or CCAAT enhancer-binding protein homologous protein (CHOP). They are able to induce an expression program for genes important to alleviate the starvation, not only because activates amino acid biosynthetic pathways but also because upregulates the expression of amino acid transporters (Wek *et al.*, 2006). This elevated expression leads to an increase in the uptake of amino acids, which ends in mTOR reactivation and autophagy suppression (Chen *et al.*, 2014). On the other hand, CHOP is also able to trigger an apoptotic program (Wek *et al.*, 2006).

---

#### 1.4. IMPACT gene

Genomic imprinting is a non-Mendelian inheritance system unique to mammals where one of the inherited copies of a gene is silenced by epigenetic modifications. Expression of imprinted genes is parent-of-origin specific, meaning that, depending on the gene, always one of the parental alleles will be silenced (Kaneko-Ishino *et al*, 2003; Morison & Reeve, 1998). The imprint is re-established or erased during gametogenesis, depending on the sex of the individual and it is maintained during the offspring's life (Okamura *et al.*, 2004).

Mouse *Impact* has been found to be the only imprinted gene at chromosome 18 until today (Okamura *et al.*, 2000). The imprinting of *Impact* is very interesting for two reasons. One reason is that the vast majority of imprinted genes have been found inside of a cluster of imprinted genes, but *Impact* is surrounded by genes that do not show any imprinting. The other interesting point is that *Impact* exhibits a specie-specific imprinting, while it has been found to be imprinted in rodents, it is not in most vertebrates such as monkey, pig, frog and human (Okamura *et al.*, 2004).

Human IMPACT has been localized in chromosome 18q11.2, a syntenic region to the locus of mouse *Impact*, within the critical region for bipolar affective disorder (Kosaki *et al.*, 2001). The imprinting has been related with the lack of intronic CpG islands in the mouse gene promoter, due to the fact that the human promoter constitutes a conventional non methylated CpG island. On the other hand, there is a methylated CpG island inside the first intron of the mouse gene but not at the human one, which are frequently found nearby imprinted genes. It is noteworthy that imprinted mouse *Impact* shows an enhanced expression when compared with human IMPACT and this is thought to be due to the presence of those methylated CpG islands. Regarding this situation, a model has been proposed where the *Impact* imprinting has evolved as an adaptation for increased gene dosage in rodents (Okamura *et al.*, 2004, 2000).



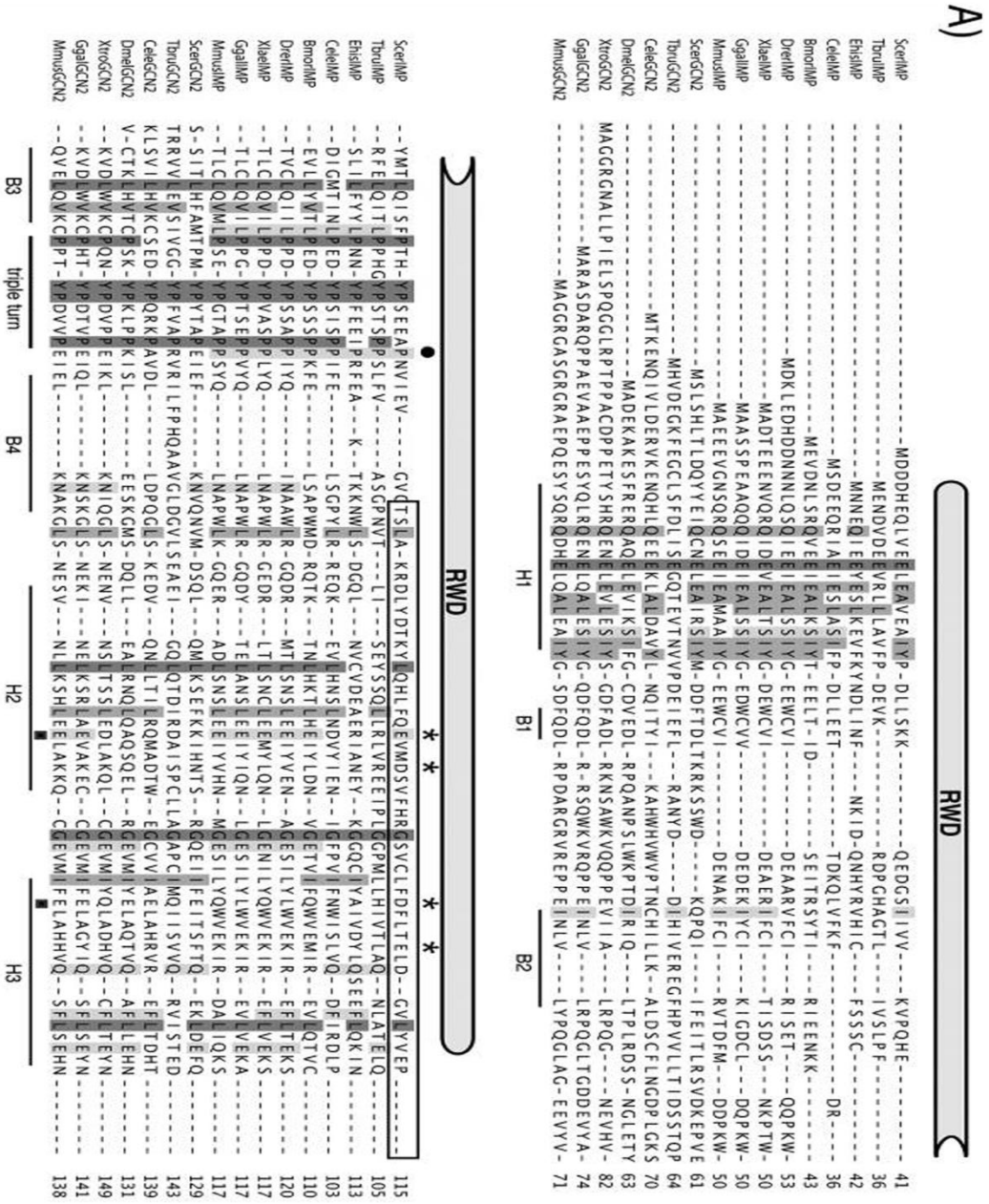
### 1.5. IMPACT protein

The name of the family arises from imprinted and ancient. Impact proteins are known to have two domains separated by a linker region. This family of cytosolic proteins is characterized because it shows high degree of sequence conservation, mostly at the half C-terminal part (Figure 1.6).

The N-terminal domain harbours an RWD domain which plays an essential role for the sole known function of the protein, which we will discuss later. The name of this RWD domain comes from domain found in RING finger proteins, WD-repeat-containing proteins and DEAD-like helicases (Waller et al., 2012). The C-terminal region, known as ancient domain, is the only domain forming part of the Pfam protein family 01205 (PF01205) which is also known as unknown protein family 0029 domain (UPF0029). This last domain has been predicted to have a  $\beta\beta\alpha\beta\beta\alpha$  topology, very similar to the ferredoxins ( $\beta\alpha\beta\beta\alpha\beta$ ), but the particular function of the ancient domain remains unknown (Figure 1.6) (Sattlegger *et al.*, 2011).

In yeast, there is only one isoform of the protein. It corresponds with a full length protein and it has been called yeast IMPACT homologue 1 (Yih1). In other organisms such as humans or *Caenorhabditis elegans*, two isoforms of the protein are found to be produced by alternative splicing. The isoform 1 corresponds with the full length protein and the isoform 2 lacks the RWD domain.

**Figure 1.6. Sequence alignment of the Yih1/IMPACT family of proteins.** Protein IMPACT sequence from organisms of diverse phylogenetic groups have been aligned, highlighting identical residues. Panel A, structural characteristics of GCN2-RWD domain from *Mus musculus* are shown for RWD domain sequence at the bottom of the alignment. At panel B, structural features of UPF0029 domain of YigZ protein from *E. coli* is shown for the UPF0029 sequence at the bottom of the alignment. Adapted from (Sattlegger *et al.*, 2011)



B)

linker region

UPF0029

ScenIMP ----- EETEPYQQSDIPTDFEGWTASDPIIDRGGSTFMAFAAHVYTS EQAFAMLDLKTDRKMKANH-VMS 182  
 TruIMP -----VTR EAQRQAEEQAKHKL EA AAEELIQVTA VEWSDDPIIDRKS K FVAHMARVNS EAGVQEVQHLRQKHIAEATHPTIY 188  
 EhisIMP ----- INIYKDEAEKEIQAKQELISKEE EESVWHGLSVKGIPIFVSNPIIDRKS K FIAFAAPVTNEKEVRIYFEE L R N K K I S I A T H - N I Q 200  
 CehIMP ----- EEVAEP SHVADDDVTPMAEEIDDMNIRGVEVLTDRKSAFQAH LAEVRSKEDVDVYMR I KSNTKICRATH-NIT 176  
 BroIMP ----- KVEKKI E V T E P Y Q V S D L T T L E I N C P E I T H G E I I A D R K S I F Q G H A E V H S I D D V K A V L N K I Q N R K I L N A T H - N M Y 185  
 DrieIMP ----- QHSDGPDCTKVTMEEGGHDCDEDDLPDILSVLKLSQSEQIFSPASODEELPKIGESITDRRSTFQPHLSAENPQVQRV L N K I E N K K I A S A T H - N I Y 220  
 XheIMP ----- QTDPGP S L K S T S E E T D V G D D C E V S T D E L T E S T K N S M I F G M N V S C V E E I P P I V H G E T I S D R S T F Q A H L A P V V S P Y Q V K L I L N K I E N K K I A T A T H - N I Y 217  
 GgalIMP ----- QSSNEPDIKKTTEEVEDECGNDFLDQYQVQEDVEILNYITSESQDEELPMIHGTP L T D R S T F Q A H L A P V V T P R Q V K R V L E K I E N K K I A S A T H - N I Y 219  
 MmuIMP ----- K P T E P D P V K K T E E V E S E E D P I L E H P P N V K T L D L K I S E E T Q P E T E L P P V A G V P I T D R R S T F Q A H V A P V V C P E Q V K I V L A K I E N K K I A S A T H - N I Y 219  
 BraIMP ----- M L H S Y F T V K E A G E H E I V I E K S R F I C H L S R Y S T E Q E A Q E I Q K I K Q H - W N A T H - N C S 55  
 MmaIMP ----- M D E N S Y K T L K S F S K V E L V F K G L F L G Y G P M K S E A K E F I D K I K E I H - N D A T H - N V S 56  
 EcoIMP ----- M E S W L I P A P V T V E E I K K S R E I T M L A H T D G V E A K A F E S V R A E H - P D A R H - H C V 54

UPF0029

ScenIMP AWRIKQDQ--SAATYQDSDDDG--TAAGRMLHLITIMDVNVIYVVARWFGGAHIGPDRFKHINSTAR EAVYRACFDS 258  
 TruIMP AYRFTDAG---VLHADCDNDDG---TGAA SRIMFLLEQQKVDGYVVTWVFGGILGPD RFKHMEVHNILLTP 260  
 EhisIMP AYRL-EING---SIIIGDFDDG---HAA GKQLK SLEQI SATNVCVMVSRWFGGILGPD RFKHIMAAQDAINIMNTSLOTTSTKVKKN 285  
 CehIMP AYRYMTEI NCKP IHHHDCVDDG---FGASSKMELMDRMKADNVKVVVSRWYGGIHLGPD RFKHINLTRQILSDNNYGRPKS 257  
 BroIMP AYRIERKTA K T N V I Q D C D D D G E --- A H A G G R M H L L O V L N Q K N T L V V S R W Y G G V Q L G P D R F R H I N N S R Q V I Q A C L L K K 264  
 DrieIMP AYRIYQCE--KNSVIQDCEDDG---TAGGRLLHLLOLDVRNVLVVS R W Y G G I L G P D R F K H I N N C A R T I L I Q E G Y A D S T E E T S K A G G S K K P K S K - K T K 317  
 XheIMP AYRIYIKE--RNSFIQDCEDDG---TAGGRMLHMQILDARDVMVVS R W Y G G I L G P D R F K H I N N C A R N V L M E H N Y C S T V E E S S K Q T S K S K S S Q R - K 312  
 GgalIMP AYRIYCED--KHTFIQDCEDDG---TAGGRLLHLMQILNVQNVLVVSRWYGGILGPD RFKHINNCARNI LVEVYVH S V E E S S K P A C K N K I R K D M K R T E H 320  
 MmuIMP AYRIYCED--KQTFIQDCEDDG---TAGGRLLHMLELVNKNVWVVS R W Y G G I L G P D R F K H I N N C A R N I L V E K N F N T P D E S T K S L G K - K V Y K D - K K K N D H 318  
 BraIMP AYVI G--EN---DHIQKANDDGEP SGTAGVPMELVKRRRLKDTCAVTVRYFGGILGAGGLIRAYGKSVSEGLN----- 125  
 MmaIMP AYRIINI-EN---NFAMKYVDDGEPQGS S G K P I Y K V I E L K N I Q N T V I V T R Y F G G V K L G V G L K A Y S D T A T E V I N----- 127  
 EcoIMP AWVAGAPDD---SQQLGFSDDGEPAGTAKGPMILAQLMGSGVCEITAVVRYGGILLGTGGLKAYAGGVNQLR----- 126

B3'

B4' ★

H2'

B5' ★

H3'

---

## 1.6. Insights into IMPACT proteins function

Little is known about the function of IMPACT inside the cell. Most of the information available is relating IMPACT protein with the general amino acid control pathway (GAAC), but recent studies are shedding light on new possible roles.

### 1.6.1. IMPACT and GAAC

The study of the relation between IMPACT and the GAAC pathway has been carried out mainly using the yeast homologue Yih1. Below, all the characteristics of Yih1 related to the GAAC pathway are commented.

#### **1.6.1.1 Yih1 binds GCN1**

As mentioned before, Yih1 harbours a RWD domain at its N-terminal part. The RWD domain is enough for GCN1 binding; in fact it has been shown that residues 1-67 do not participate in the interaction, revealing that helices 2 and 3 of the RWD domain are sufficient for GCN1 binding (Sattlegger *et al.*, 2011). The observation of an increased GCN1 binding affinity when Asp90 and Glu87 are mutated to alanine, together with the role of Asp102 and Glu106 for in vivo binding, led to the conclusion that helices 2 and 3 of the RWD domain as direct interactors with GCN1. From GCN1, Arg2259 is essential for Yih1 binding in vivo and in vitro, as it is for GCN2 binding (Sattlegger *et al.*, 2004; Sattlegger & Hinnebusch, 2000). Yih1-GCN1 interaction under stress conditions is very important since it is responsible for inhibition of GCN2 dependent eIF2 $\alpha$  phosphorylation.

#### **1.6.1.2 Yih1 binds actin**

Endogenous Yih1 FLAG-tagged when purified appeared as 1:1 complex with monomeric actin without GCN1 implication (Sattlegger *et al.*, 2004). A

fragment of Yih1 comprised by residues range 68 to 258 showed the strongest interaction with actin, although fragments 2 to 171 or 68 to 171 also succeeded precipitating actin (Sattlegger *et al.*, 2011). Same Yih1 mutations that increased affinity for GCN1 binding, E87A and D90A, also increased actin binding affinity (Sattlegger *et al.*, 2011).

#### **1.6.1.3 Yih1 resides in a complex with ribosomes**

Yih1 (and IMPACT) can associate with ribosomes independently of GCN1. A fragment corresponding to the central part of the protein (residues 68 to 171) is sufficient for polyribosome cosedimentation. This interaction was also present under amino acid starvation conditions (Waller *et al.*, 2012).

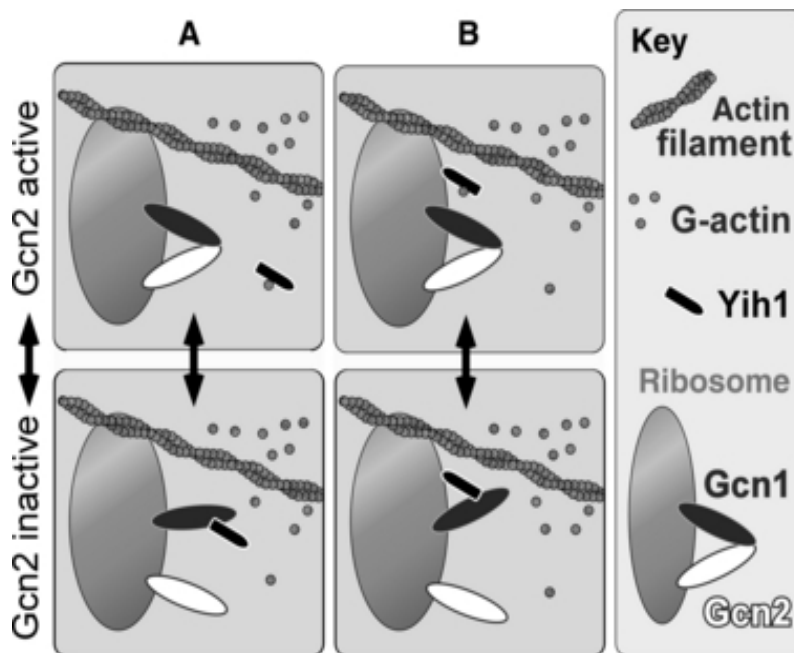
By crosslinking it was determined that Yih1 is interacting with the large ribosomal protein 39 (RPL39) (Waller *et al.*, 2012).

#### **1.6.1.4 GAAC model for GCN2 regulation**

Two possible models for Yih1/IMPACT dependent GCN2 regulation have been proposed regarding all the characteristics commented above (Figure 1.7).

As we have explained before GCN2 binds through its RWD domain to a particular area of GCN1 also recognized by Yih1. Under non stressed conditions, Yih1 would be in an inactive complex with actin. GCN1 and GCN2 stay in a complex with the ribosome, which are known to be bound to the cytoskeleton. When Yih1 is released from monomeric actin, starts competing with GCN2 for GCN1 binding. On the other side, there are evidences for Yih1 binding to ribosome independently of GCN1. This would allow a fastest regulation of GCN2 (Waller *et al.*, 2012) in comparison with the model where the Yih1-actin complex is free in the cytoplasm. Avoiding GCN2-GCN1 interaction, Yih1 is inhibiting the kinase domain of GCN2 and thus, assuring the maintenance of general translation levels. Nevertheless, in normal conditions Yih1 (IMPACT) is complexed with actin,

so the inhibition of the kinase activity of GCN2 has been proposed to occur when maximal protein synthesis is necessary due to any special condition, when there is a high overexpression of Yih1 (IMPACT) or at specific cellular compartment where Yih1 (IMPACT) might be released from its interaction with actin (Sattlegger *et al.*, 2004). Taking into account that the number of Yih1 molecules inside the cells is less than half of GCN1, and that only a small portion of Yih1 will bind ribosomes, only a small population of ribosome-GCN1-GCN2 complexes will be under possible Yih1 control (Waller *et al.*, 2012).



**Figure 1.7. Proposed models for Yih1 on the ribosome.** Ribosomes are known to be attached to the actin cytoskeleton. GCN1 and GCN2 reside on the ribosome. GCN1 direct contact to GCN2 facilitates its activation. Under certain conditions Yih1, usually complexed with globular actin, is released from its interaction and binds GCN1, thus inhibiting GCN2 activity (Panel A). Yih1 remains bound to the ribosome although inhibited by its interaction with actin. When GCN2 has to be inactivated, Yih1 bound to ribosome levels increase allowing faster GCN2 regulation (Panel B) Taken from (Waller *et al.*, 2012).

### 1.6.2. IMPACT and neuronal differentiation

In mice, high levels of IMPACT protein have been found at the central nervous system (NCS) not only in adult individuals but also in embryos. It is worth to note that those levels even increased during brain development. This increased expression was coordinated with higher levels of neuronal differentiation markers and GCN2 expression reduction. Polysome profiling revealed that IMPACT binding to polysomes was highly increased during differentiation (Roffé *et al.*, 2013).

ATF5, a transcriptional factor regulated by the transcriptional activation mediated by ATF4 (Zhou *et al.*, 2008), inhibits neurite development induced by nerve growth factor (Angelastro *et al.*, 2003). Then, regarding that extracellular signals may promote or inhibit neurite outgrowth and that high levels of IMPACT bound to polysomes have been found in differentiating cells, it has been proposed that differentiation of N2a cells (mouse neuroblastoma cells) triggers IMPACT association with translating ribosomes and that the abundance of IMPACT may promote translation by decreasing GCN2 activity in a timely manner to support neurite outgrowth (Roffé *et al.*, 2013).

### 1.6.3. IMPACT and cell cycle

Cyclin dependent Kinases (Cdks) are protein kinases which its activation or inactivation controls the eukaryotic cell division. In yeast there is only one Cdk that regulates the cell cycle progression, Cdc28, the orthologue of mammalian CDK1. Cdk's activity is regulated by phosphorylation and the interaction with negative effectors (Alberghina *et al.*, 2004; Bloom and Cross, 2007).

It has been recently described a relation between cell cycle and Yih1. When Yih1 is deleted from yeast, cells tend to accumulate in the G2/M phases of the cell cycle and this delay in the cycle is not related with GCN1 or GCN2. Cdc28 was found to co-precipitate with Yih1, and IMPACT with CDK1, result that is

---

evidencing the evolutionary conservation of the interaction. This binding has been proven to be independent of GCN1 (Silva *et al.*, 2015).

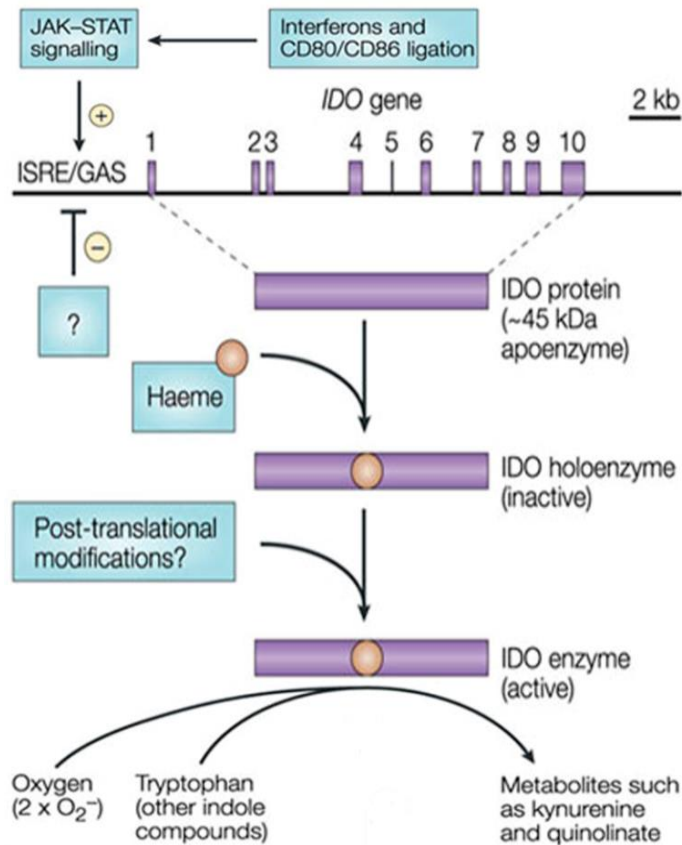
Yih1 preferentially binds active Cdc28 complexes, and the RWD domain has been found to be the responsible for the interaction. Substitution of two residues within the RWD domain, glutamic 87 and aspartic 90 to alanine, is producing a stronger interaction as it was also described for GCN1 and actin (Sattlegger *et al.*, 2011; Silva *et al.*, 2015).

#### 1.6.4. **IMPACT and the immune system**

A direct link between IMPACT and a very well-known immunomodulatory factor, Indoleamine 2,3-dioxygenase (IDO), has been reported (Habibi *et al.*, 2010). IDO is an enzyme involved in the tryptophan catabolism, performing the first and rate-limiting step of the pathway (Takikawa, 2005) and it was firstly identified in rabbit intestine. IDO catalyses tryptophan conversion to N-formyl-kynurenine. This product is further metabolised in different downstream metabolites, depending on the enzymes that are expressed downstream by each cell type. IDO expression can be induced upon cell exposure to interferon gamma (INF- $\gamma$ ) (Taylor and Feng, 1991) and other factors such as transforming growth factor  $\beta$  or interleukin 10 (Powrie and Maloy, 2003) (Figure 1.8).

It has been shown that overexpression of IDO has immunosuppressive effects, and those effects are attributed in part to tryptophan depletion and/or accumulation of kynurenines, which are toxic metabolites obtained from tryptophan catabolism (Fallarino *et al.*, 2006). Some studies point to the signalling pathway triggered by GCN2 as a mechanism that enables T-cell to sense stress conditions generated by IDO (Munn *et al.*, 2005).





**Figure 1.8. Activation of IDO expression and enzyme activity.** IDO inducers activate JAK-STAT signalling pathway that acts on interferon stimulatory response elements (ISREs) and  $\gamma$ -activating sequences (GAS) present on the IDO promoter. Functional activity regulation can occur via undescribed post-translational modifications and also by the presence of inhibitors or by limiting the access to the haeme cofactor. The obtained products vary depending on the cell type. Adapted from (Mellor and Munn, 2004).

It has been shown that GCN2 pathway is selectively activated in immune cells but not in skin cells. Interestingly, this has been related with the non-detectable or low constitutive expression of IMPACT in Jurkat and T cells

---

respectively in contrast with the higher expression levels found in skin cells (Habibi *et al.*, 2010). In stress condition induced by IDO, Jurkat cells initiate GCN2 pathway which induces cell cycle arrest and finally ends in apoptosis. Overexpression of IMPACT in Jurkat cells reduces the apoptosis ratio (Habibi *et al.*, 2010) which is consistent with previous observations where Jurkat cells, due to its lower IMPACT expression levels when compared with primary T cells, showed higher apoptotic tendency relative to those primary T cells (Ghahary *et al.*, 2004).

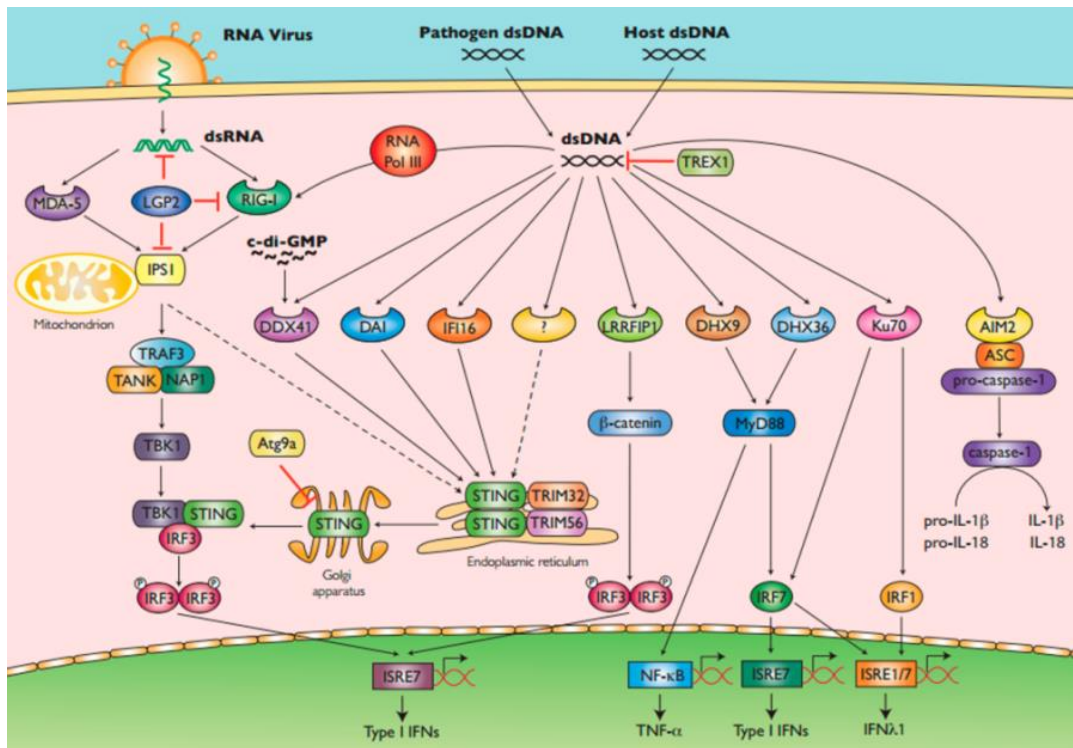
### 1.7. Nucleic acid recognition and immune response

There are different mechanisms that allow cells to respond to infectious agents. Recognition of foreign nucleic acids within the cytoplasm, usually due to a viral infection or intracellular bacteria is one of them (Gehrke *et al.*, 2013). Several cytosolic nucleic acid sensors have been described and they form part of the innate immune response. Usually those cytosolic sensors will trigger, by activating different signalling pathways, the secretion of cytokines and type I interferons (interferons  $\alpha$  and  $\beta$ ) (Atianand and Fitzgerald, 2013; Hornung and Latz, 2010) (Figure 1.9).

Although the cell is able to differentiate between self and non self RNA, it is less specific recognising self DNA (Keating *et al.*, 2011) . Special characteristics of nucleic acids, as unmethylation of C-G dinucleotides, particular secondary structures, location or abundance are specifically contributing to immune recognition. These are critical parameters since inappropriate nucleic acid recognition and activation of the immune response can lead to autoimmune diseases (O'Neill and Bowie, 2010).

Given the increasing number of cytosolic DNA sensors and that they usually recognise the same substrates; we can say that there is high redundancy among them. It is thought that some of the DNA sensors can be cell type specific

and as consequence one can find many different signalling pathways implicated.



**Figure 1.9. Immune response triggered by the presence of nucleic acids in the cytosol.** Cytosolic nucleic acids sensors for each type of nucleic acid and triggered downstream pathways are indicated. Picture obtained from “Cytosolic DNA sensors: a STING in the tail, from invivogen”.



### 1.8. Objectives

IMPACT proteins are highly conserved, but little is known about its role in the cell. Despite the relation found between IMPACT and the immune system or cell cycle, the only known function is related to GCN2 inhibition. The role of IMPACT proteins has been always linked to other proteins, but there are not intrinsic functions for IMPACT proteins yet described.

From the structural point of view, no previous information regarding its tridimensional architecture is available. Although the N-terminal domain has been annotated as an RWD domain, there are no experimental evidences of its structure. The ancient domain, located at protein C-terminus, belongs to the PF01205. This domain shows very high degree of sequence conservation but any particularity about it is known until date.

With these premises, the objectives were:

1. Determine the three dimensional structure of the protein IMPACT using X-ray crystallographic techniques.
2. Identification and characterization of new IMPACT capabilities.
3. Investigate possible functions within the cell for new IMPACT capabilities.





## **2. METHODOLOGY**





## 2.1 Production of IMPACT proteins in *E.coli* expression system

### 2.1.1 Obtainment of expression plasmids

Two different methods were used to clone all constructs:

**In-Fusion:** Commercial Kit from Clontech Laboratories. This method was used with pOPINF vector. It was used accordingly to the manufacturer's instructions. In brief, an insert of the desired construct was generated carrying two different overhangs. Each specific overhang in the insert is able to recombine with one end of the linearized pOPINF, so the insert can be introduced into the plasmid making it circular again. Since the recombination is sequence dependent, the proper direction of the cloning is assured.

**Sequence and ligase independent cloning (SLIC):** The SLIC method was used to clone proteins into pET28-NKI/LIC 6His/3C (NKI) (kind gift from Dr. Tassos Perrakis) and GKI (generated in the lab) plasmids. GKI conserves the same backbone as NKI but the histidine tag has been changed for a glutathione-S-transferase (GST) tag. For the reaction, two microliters of insert were mixed with two microliters of linearized plasmid and 1 $\mu$ L of T4 DNA polymerase buffer (Fermentas). After the addition of 0.2 $\mu$ L of T4 DNA polymerase, the sample was incubated for 1 minute at room temperature. Afterwards it was transformed as described below.

#### 2.1.1.1 Yeast IMPACT homolog

Three different orthologues have been expressed, IMPACT human protein, Yeast Impact Homolog (Yih1) and *Chaetomium thermophilum* Impact Homolog (CIH). Plasmids containing yeast Yih1 and GCN1 constructs (Figure 2.1) were obtained from Dr. Bertrand Séraphin (Illkirch, Strasbourg) as part of the European

Union 6<sup>th</sup> Framework program 3D repertoire (contract LSHG-CT-2005-512028).



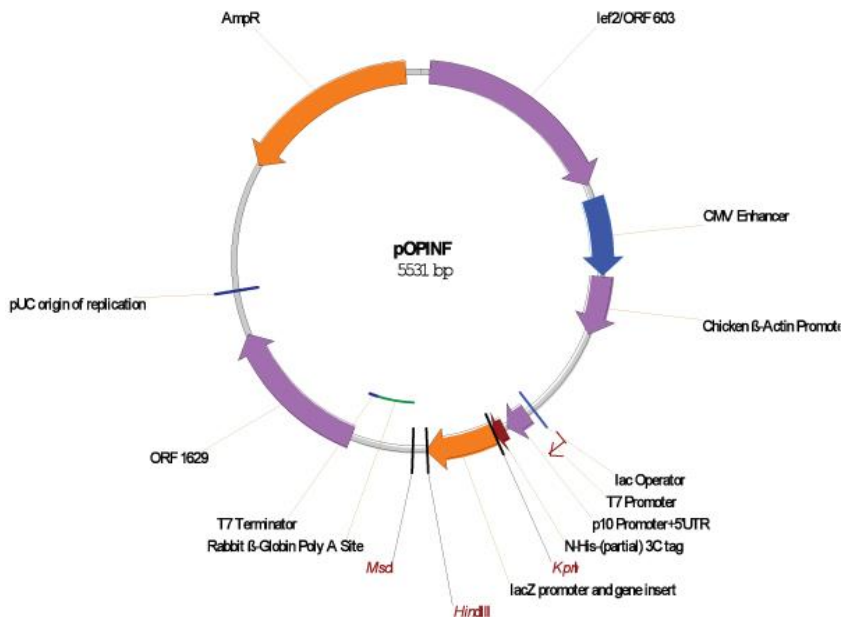
**Figure 2.1. Scheme showing the constructs obtained from Dr. Bertrand Séraphin.** The black bar shows the position of the 6 histidine tag at the N-terminal part of Yih1. Numbers between parentheses are pointing the residue range of the respective domain. The plasmid scaffold for these constructs is pET24d+. Figures are not drawn to scale.

Since GCN1 expression was not detected in none of the four plasmids from Bertrand Séraphin, a construct carrying only Yih1<sub>FL</sub> was made amplifying the open reading frame (ORF) from pBS4717 with primers carrying the overhangs necessary for pOPINF cloning (Table 2.1).

### **2.1.1.2 *Chaetomium* IMPACT homolog**

The plasmid pOPINF (Figure 2.2) containing CIH full length (CIH<sub>FL</sub>) was obtained by cloning the IMPACT gene from a *Chaetomium thermophilum* gene library generated in the laboratory. The DNA that was used as insert was amplified by the polymerase chain reaction (PCR) technique using an Eppendorf thermal cycler. The reaction consisted in 1 unit of KAPA HiFi DNA polymerase,

(Kapa Biosystems) in the commercial buffer (KAPA HiFi fidelity buffer), 0.3 mM dNTP mix, 0.3  $\mu$ M oligonucleotide primer synthesized by Sigma (Table 2.1) , DMSO and around 50 ng DNA template in 25 $\mu$ L final volume. The PCR reaction was carried out as given: 95°C for 3 minutes followed by 30 cycles of amplification (denaturation step of 98°C for 20 seconds, annealing for 20 seconds at 55°C followed by elongation step at 72°C calculated for 1Kb amplified for 30 seconds of reaction). The PCR product was confirmed by 1% agarose gel and cleaned with a DNA purification kit (GE healthcare). After extracting the gene from the library, a new PCR was performed as described above with primers that were incorporating an overhang necessary for the cloning with pOPINF vector (table 2.1).



**Figure 2.2. pOPINF vector map.** Resistance, origin of replication and position of cleavage site for some restriction enzymes are shown. Map obtained from: [www.addgene.org](http://www.addgene.org)

Different truncations of the protein were made on the CIH sequence and

cloned into different plasmids. The different constructs and the oligonucleotide primers used to generate them are shown in table 2.1.

Completed reactions were transformed into 100µL of chemically competent *E. coli* DH5α cells by incubating for 20 minutes at 4°C and then heat shocking at 42°C for 1 minute 30 seconds. Afterwards the mix was kept in ice for 5 minutes prior to incubation at 37°C for 1 hour in 300µL of Lysogeny Broth (LB) without antibiotics. Finally all the cells were spread in an LB plate with the selected antibiotic and incubated overnight at 37°C. Selected colonies were grown in 5mL of LB with proper antibiotic at 37°C and the plasmid DNA was extracted with Qiagen miniprep spin column kit (Qiagen). The constructs were confirmed by sequencing.

**Table 2.1. Constructs for Yih1, IMPACT and CIH .** Column named Res does refer to vector resistance. Primers used for cloning are shown and marked with F for forward and R for reverse. Overhangs necessary for cloning are highlighted in grey.

Construct	Vector	Boundaries	Res	Primers
<b>Yih1</b>	pOPINF	1-258	Amp	F-AAGTTCTGTTTCAGGGCCCCGATGGATGACGATCACGAACAGTTGGTCG R-ATGGTCTAGAAAAGCTTTACGAGTCGAAGCCGCGCTGACAAC
<b>IMPACT</b>	pOPINF	1-320	Amp	F-AAGTTCTGTTTCAGGGACCCGGTATGGCTGAGGGGGACGCAGGGAG R-CGAGGAGAAGCCCGGTTAATGTTTCATTCTCTTCTTGTCTTTCTTAC TTTTTGTCTTTCCAAAAGC
<b>CIH</b>	pOPINF	1-292	Amp	F-AAGTTCTGTTTCAGGGCCCCGATGTCAGAAGCGCTGCTCG R-ATGGTCTAGAAAAGCTTTATTTTCTCCCTTCTTCTTC
<b>CIH<sub>1-111</sub></b>	NKI	1-111	Kan	F-CAGGGACCCGGTATGTCAGAAGCGCTGCTCGATGAAATAGAAGCCATC R-CGAGGAGAAGCCCGGTTATTTACGGCAGCCAGGAGTTCCTGTACC
<b>CIH<sub>111-292</sub></b>	NKI	111-292	Kan	F-CAGGGACCCGGTAACCGTGATGGCCGGCACAC R-CGAGGAGAAGCCCGGTTATTTTCTCCCTTCTTCTTCTTCTCCCCGC
<b>CIH<sub>1-160</sub></b>	NKI	1-160	Kan	F-CAGGGACCCGGTATGTCAGAAGCGCTGCTCG R-CGAGGAGAAGCCCGGTTACAGAGTCCACGGTGGAGGGCTAGATA
<b>CIH<sub>152-292</sub></b>	GKI	152-292	Kan	F-CAGGGACCCGGTCCCTCCACCGTGGACTCTGTC R-CGAGGAGAAGCCCGGTTATTTTCTCCCTTCTTCTTCTTCTCCCCGC

Confirmed constructs and mutants were transformed in 50 $\mu$ L *E.coli* (DE3) BL21 codon plus RIPL (BL21C+ RIPL) (Agilent technologies). This strain is used for protein expression. The transformation by heat shock was performed as follows: heat shock at 42°C for 25 seconds, 5 minutes incubation on ice followed by the addition of 300 $\mu$ L of LB and the incubation for 1 hour at 37°C. For obtaining single colonies, 100 $\mu$ L of these transformed cells were spread in an LB plate with the selected antibiotic for the vector resistance and chloramphenicol for the RIPL cells resistance, and kept it overnight at 37°C. A mixture of 750 $\mu$ L of a single colony grown overnight in LB plus the proper antibiotics and 250 $\mu$ L of 100% glycerol were kept in a -80°C freezer as a stock for starting new cultures.

### **2.1.1.3 Human IMPACT**

In order to prepare a vector containing the human IMPACT open reading frame, it was PCR-amplified from the MegaMan human transcriptome library (Agilent technologies) and cloned into pOPINF vector as described above. Primers for cloning are shown at Table 2.1.

### ***2.1.2 Expression and purification of the different constructs***

Adequate expression of the proteins was first assured by small scale expression in 20mL cultures (LB media containing the selected antibiotic) induced with 1mM IPTG final concentration, followed by batch purification using affinity Ni-NTA resin (Agarose Bead Technologies, ABT) and confirmed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE).

Large scale expression of the recombinant proteins, if not mentioned otherwise, was carried out in LB with IPTG induction. A pre-culture was grown in 50mL of LB containing appropriate antibiotic, starting from the glycerol stock, at 37°C overnight in a shaking incubator at 190 r.p.m. The next day the pre-culture

---

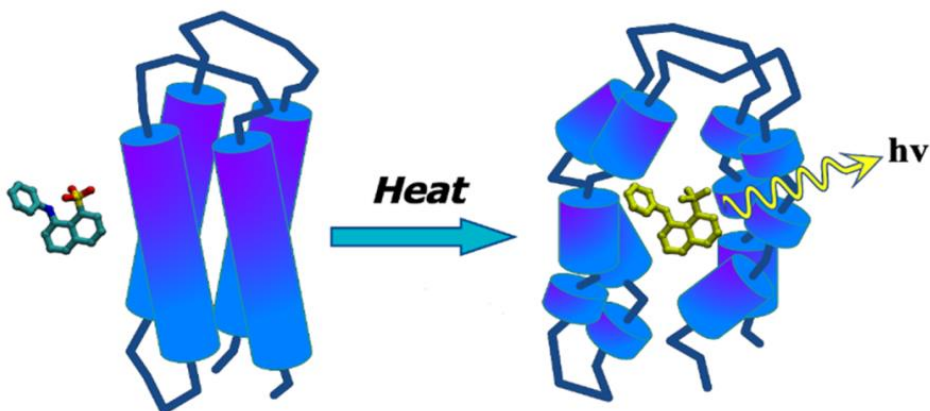
was inoculated into 950mL LB containing appropriate antibiotic inside a 2 liter conical flask and grown until  $O.D_{600nm}$  between 0.5-0.7. When reaching this point, a final concentration of 1mM IPTG was added to the media and the temperature decreased to 20°C. The culture was left incubating overnight at 190 r.p.m. The next day the cells were harvested by centrifugation at 4547g for 30min in a Beckman Coulter J6-HC centrifuge using a JS-4.2A rotor at 4°C. The pellet was washed with 35mL PBS 1x and centrifuged again in an Eppendorf Centrifuge 5810R at 3220g for 25min at 4°C. The supernatant was discarded and the pellet frozen with liquid nitrogen and stored in a -80°C freezer until use (Baneyx, 1999; Makrides, 1996).

Frozen pellets, originated from 1 liter culture, were thawed on ice and resuspended in 30mL lysis buffer (1% Triton X-100, 500 mM NaCl, 5% Glycerol, 80 mM Tris pH 7.5 and 1 mM  $\beta$ -mercapto-ethanol) plus one tablet of complete™ EDTA free protease inhibitor cocktail (Roche). Cells were sonicated using a vibracell 75042 sonicator (Bioblock scientific) for 15 minutes on ice using 37% amplitude with 1 second on 1 second off cycles. Afterwards the sonicated fraction was centrifuged for 40 minutes at 17000 r.p.m in a Sorval centrifuge using ss-34 rotor in order to separate cell debris. The supernatant was filtered through a 0.4  $\mu$ m syringe filter prior to the first step of purification: affinity chromatography. For purifying proteins with a 6-histidine tag (all pOPINF and NKI constructs) from the lysate extract, an affinity chromatography step using 5 mL Ni-HisTrap column (GE Healthcare) was performed. The individual tagged proteins were isolated since the 6-histidines tag binds to the nickel ions in the column resin. Proteins of interest were eluted by applying step-wise imidazole gradient. Imidazole competes with histidines for the binding sites of nickel allowing the elution of the tagged proteins. Besides, the use of a step-wise gradient facilitates the separation between proteins harboring histidines on their surface which might be retained due to unspecific binding, and the protein of interest.

All the fractions collected were analyzed by SDS polyacrylamide gel electrophoresis (SDS-PAGE), and those containing the desired protein were further purified by size exclusion chromatography. The type of column used for size exclusion chromatography was depending on the size of each construct. Taking into account the molecular weight of the protein, Superdex 200 26/60 or Superdex 75 16/60 (GE Healthcare) was used. For each protein, the different peaks eluted in size exclusion were analyzed by SDS-PAGE and homogenous pure protein was concentrated using centrifugal filter units (Millipore) with adequate pore size for each protein and used for crystallization trials.

## 2.2 Differential scanning fluorimetry to assess protein stability

Fluorescence based thermal shift assay is also known as ThermoFluor assay. It uses a hydrophobic fluorophore, in this case SYPRO® Orange (Sigma) to monitor protein denaturation (Reinhard *et al.*, 2013) (Figure 2.3).



**Figure 2.3. SYPRO® Orange emission.** With progressive protein denaturation hydrophobic areas are exposed. Then SYPRO® Orange dye binds to the hydrophobic core of the protein and the fluorescent emission starts. (<http://www.beta-sheet.org>).

Thermofluor was performed using 198 $\mu$ L of protein samples at 1mg/mL in buffer 50mM Hepes pH 7.5 plus 2 $\mu$ L of SYPRO<sup>®</sup> Orange (final concentration 5x) in a 96 well Micro-Amp Fast Optical reaction plate. The plates were aliquoted with 38 $\mu$ L of 48 different buffers (table 2.2), so each condition could be measured by duplicate using the 96 wells.

**Table 2.2. Buffers used for thermofluor assay.** Variables at each condition are shown in red.

	1	2	3	4	5	6	7	8	9	10	11	12
<b>A</b>	150mM NaCl	150mM NaCl	150mM NaCl	150mM NaCl	150mM NaCl	150mM NaCl	150mM NaCl	150mM NaCl	150mM NaCl	150mM NaCl	150mM NaCl	150mM NaCl
	100mM AcNa pH 4.5	100mM Bis Tris pH 5.5	100mM citrate pH 5.6	100mM cacodilate pH 6.5	100mM MES pH 6.5	100mM Hepes pH 7	100mM Hepes pH 7.5	100mM Phosphate pH 7.5	100mM Tris pH 8	100mM Tris pH 8.5	100mM Bicine pH 9	100mM CHES pH 9.5
<b>B</b>	150mM NaCl	150mM NaCl	150mM NaCl	150mM NaCl	150mM NaCl	150mM NaCl	150mM NaCl	150mM NaCl	150mM NaCl	150mM NaCl	150mM NaCl	150mM NaCl
	100mM SPG 4.5	100mM SPG 5.5	100mM SPG 6.5	100mM SPG 7.5	100mM SPG 8	100mM SPG 9	100mM MMT4.5	100mM MMT5.5	100mM MMT pH6.5	100mM MMT pH 7.5	100mM MMT 8	100mM MMT 9
<b>C</b>	Hepes pH 7.5	100mM NaCl Hepes pH 7.5	250mM NaCl Hepes pH 7.5	500mM NaCl Hepes pH 7.5	150mM NaCl Hepes pH7.5 5% Glycerol	150mM NaCl Hepes pH7.5 10% Glycerol	150mM NaCl Hepes pH 7.5 15% Glycerol	150mM NaCl Hepes pH 7.5 1mM MgCl <sub>2</sub>	150mM NaC Hepes pH 7.5 1mM CaCl <sub>2</sub>	150mM NaCl Hepes pH 7.5 0.1mM ZnCl <sub>2</sub> 0.1mM CoCl <sub>2</sub>	150mM NaCl Hepes pH 7.5 1mM CdCl <sub>2</sub> 1mM MnCl <sub>2</sub>	150mM NaCl Hepes pH 7.5 1mM DTT
	150mM NaCl 100mM Hepes pH7.5	100mM Hepes pH7.5 150mM LiCl <sub>2</sub>	100mM Hepes pH7.5 150mM KCl	100mM Hepes pH7.5 150mM NaF	100mM Hepes pH7.5 150mM SO <sub>4</sub> NH <sub>4</sub>	100mM Hepes pH7.5 150mM NO <sub>3</sub> NH <sub>4</sub>	150mM NaCl 100mM Hepes pH 7.5 100mM Sucrose	150mM NaCl 100mM Hepes pH 7.5 300mM Sucrose	150mM NaCl 100mM Hepes pH 7.5 Urea 0.5M	150mM NaCl 100mM Hepes pH 7.5 L-Arg 50mM L-Glu 50mM	150mM NaCl 100mM Hepes pH 7.5 Betaine 100mM	150mM NaCl 100mM Hepes pH 7.5 1% Triton X-100
<b>D</b>	150mM NaCl 100mM Hepes pH7.5 1mM EDTA	100mM Hepes pH7.5 150mM LiCl <sub>2</sub>	100mM Hepes pH7.5 150mM KCl	100mM Hepes pH7.5 150mM NaF	100mM Hepes pH7.5 150mM SO <sub>4</sub> NH <sub>4</sub>	100mM Hepes pH7.5 150mM NO <sub>3</sub> NH <sub>4</sub>	150mM NaCl 100mM Hepes pH 7.5 100mM Sucrose	150mM NaCl 100mM Hepes pH 7.5 300mM Sucrose	150mM NaCl 100mM Hepes pH 7.5 Urea 0.5M	150mM NaCl 100mM Hepes pH 7.5 L-Arg 50mM L-Glu 50mM	150mM NaCl 100mM Hepes pH 7.5 Betaine 100mM	150mM NaCl 100mM Hepes pH 7.5 1% Triton X-100

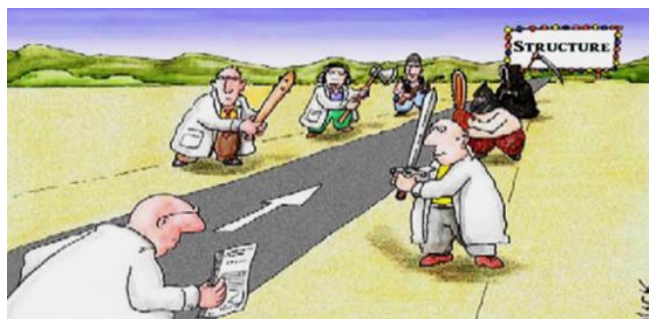


After the addition of 2 $\mu$ L of the protein-dye mix to each well, 20 $\mu$ L were pipetted out to perform the replicates. Plates were sealed with Micro-Amp optical sealing tape (Applied Biosystems), spin down for 1minute at 1000 r.p.m and heated from 20°C to 85°C in increments of 1°C in a T7500 Fast Real-Time PCR System (Applied Biosystems). The wavelength for excitation and emission were 490nm and 550nm respectively. Results were analyzed and visualized graphically in GraphPad Prism 4 software (GraphPad Software Inc.).

This technique is a good choice for testing protein stability. However it has some limitations, as for example, proteins showing hydrophobic patches on their surface will bind the hydrophobic dye and the emission will start when the protein is not denatured. These aspects must be taken into account when analyzing graphical results.

### 2.3 Crystallization of *Chaetomium* IMPACT homolog domains

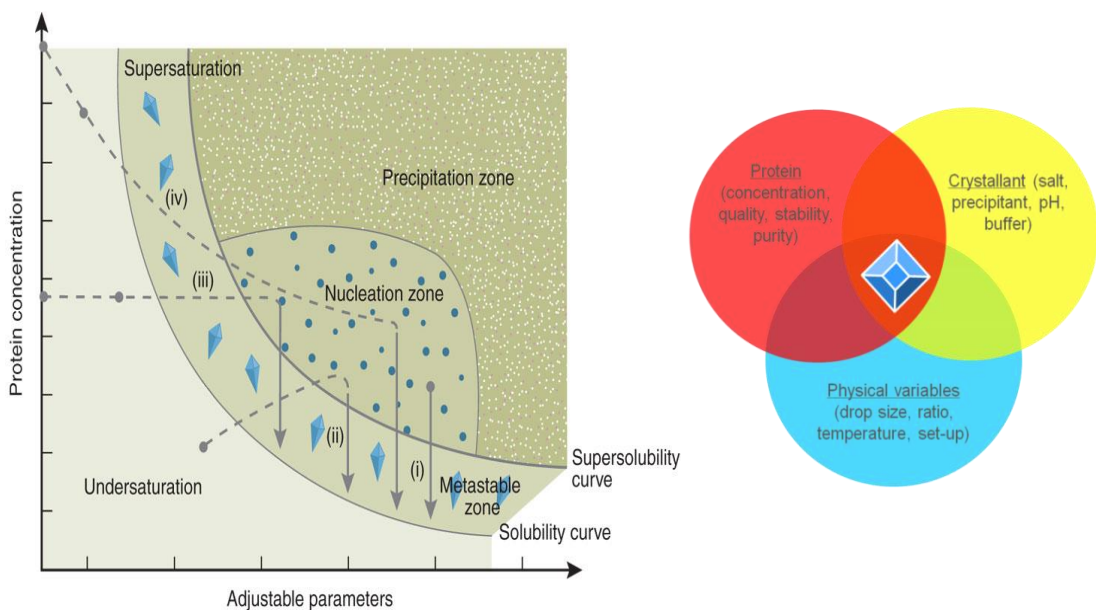
Protein crystallization is, usually, the limiting step in a crystallographer's life. The conditions where a given protein would crystallize cannot be predicted.



Modified from [http:// www.dissertationwritingservicebro.com](http://www.dissertationwritingservicebro.com)

Crystallization occurs when the protein precipitates in an ordered manner. There are several methods to obtain protein crystals, but at the institute the most

extended is vapor diffusion, which is perhaps the most convenient for screening. The vapor diffusion technique allows protein concentration over time. Protein solution is mixed with the reservoir solution, which usually contains a crystallant that decreases protein solubility, and by vapor diffusion the water from the mixed drop migrates to the reservoir in the well, leading to a slow concentration of the protein and precipitating agent. There are many variables influencing protein crystallization that can be changed when seeking for a proper crystallization condition (Figure 2.4). There are also different techniques that one can try when crystals need to be improved, as for example random micro-seeding or macro-seeding (Chayen and Saridakis, 2008; St John *et al.*, 2008) .



**Figure 2.4. Scheme of crystallization process.** In order to obtain crystals, the protein solution should favor nucleation. Once crystal nuclei are formed, protein concentration decreases due to crystal growth, reaching and ideal metastable area where crystal will remain (Left panel). Several parameters can be adjusted to move from under-saturated conditions to more optimal conditions for crystal growth (Right panel) (<http://crystal.csiro.au/User-Guide/Crystallisation>) (Chayen and Saridakis, 2008)

### 2.3.1 Crystallization trials

The crystallization service from the institute was in charge of dispensing both, the purified protein and the mother liquor for each condition. Devices used by the service in order to set up crystallization screenings and optimization of crystal growth conditions were Janus automated workstation (Perkin Elmer), that aliquots the reservoir condition into every well of the crystallization plates, and honey bee X<sup>8</sup> (Cartesian dispensing systems) which dispenses the reservoir and the protein drop into the well suitable for sitting drop. Typical volumes for protein drops were varying from 0.3µL to 0.5µL range and were mixed in 1:1 ratio with the reservoir solution.

Sparse matrix screenings like INDEX (Hampton Research) or JBS I and II (Jena Bioscience), grid screenings as JCSG+ (Molecular Dimensions) or screenings with a combination grid, sparse matrix and incomplete factorial screening like INDEX (Hampton research) were used for crystallization trials. Screenings with alternative precipitants as MIDAS (Molecular Dimensions) were also tested.

#### **2.3.1.1 Crystallization of *Chaetomium* IMPACT homolog**

JBS I and JBS II, Wizard (Emerald Biosystems), Index, JCSG + and MIDAS screenings were firstly tested in order to obtain full length CIH crystals. Poorly diffracting large crystals were obtained in JBS I F4 condition (15 % w/v PEG 6000, 50 mM Potassium Chloride and 10 mM Magnesium Chloride). In order to avoid crystal damage during cryo-cooling, crystals were separated into smaller pieces for a faster cryo-cooling. For cryoprotection, crystals were taken out of the original drop by using a nylon loop (Hampton Research) and placed in a drop with Paratone<sup>®</sup> (Hampton Research) and moved until all the solvent was removed. Afterwards, they were cryo-cooled in liquid nitrogen using the same nylon loop

---

and kept in a Dewar (Jencons) upon arrival to ALBA Synchrotron.

### **2.3.1.2 Crystallization of *Chaetomium* IMPACT homolog RWD domain**

Different commercially available crystallization screens were set up in order to obtain crystals. CIH residues 1 to 111 (RWD domain) crystals were obtained in 1.26 M Sodium phosphate monobasic monohydrate, 0.14 M Potassium phosphate dibasic, pH 5.6, from INDEX screening (Hampton research). Crystals were cryoprotected with Paratone® (Hampton Research) before freezing in liquid nitrogen.

### **2.3.1.3 Crystallization of *Chaetomium* IMPACT homologue ancient domain with DNA**

After checking the ability of CIH for binding nucleic acids, crystal screenings were set up with full length CIH and CIH<sub>152-292</sub> with different length of dsDNA with randomly chosen sequence. Some crystals were obtained with Wizard and Index screenings in the presence of 20bp dsDNA (ggatgaagatgaataactcg - cgagttattcatcttcatcc). Crystal optimization was performed by modification of the initial crystallization conditions.

Co-crystallization trials of CIH<sub>152-292</sub> with 20bp dsDNA yielded plate-like crystals (grown in 18% PEG 3500, Bis-tris propane 0.1M pH 5.5 and 0.2M NaCl) that diffracted up to 1.5 Å. Data reduction was carried out using iMOSFLM and SCALA (Battye *et al.*, 2011; Evans, 2011). Phases were obtained by molecular replacement using Phaser from phenix (Adams *et al.*, 2010; McCoy *et al.*, 2007).

### **2.3.2 *Crystal data collection, structure determination, refinement and analysis***

A complete dataset for every type of crystal was collected at XALOC beamline, ALBA synchrotron. This beamline has a Pilatus 6M-dectris detector. The photon energy range goes from 5 to 22 Kiloelectron-Volts (keV). The diffractometer, where the crystal is going to be rotated for data collection is a MD2M device from Bruker. All the structure refinement carried out was cross-validated using  $R_{\text{WORK}}$  and  $R_{\text{FREE}}$  values (Brünger, 1992; Kleywegt and Brünger, 1996; Tickle *et al.*, 2000).

#### **2.3.2.1 RWD domain**

CIH<sub>1-111</sub> crystals diffracted up to 1.4 Å. Data reduction was carried out using XDS (Kabsch, 2010). *Ab initio* phasing was successfully performed with Arcimboldo (Sammito *et al.*, 2014) and the model was built inside the electron density using Autobuild as implemented in Phenix (Adams *et al.*, 2010). Initial refinement with rigid body was performed and final refinement was carried out with Phenix refine, using anisotropy refinement in the last cycles until  $R_{\text{FREE}}$  and  $R_{\text{WORK}}$  values were 0.2078 and 0.1674 respectively (table 3.4). The automatic refinement was performed in combination with manual model building using Coot (Emsley and Cowtan, 2004; Emsley *et al.*, 2010).

#### **2.3.2.2 Ancient domain**

Crystals were tested for diffraction at XALOC beamline, taking 4 images separated by 90° for indexing. The best crystal was diffracting up to 2Å, so the detector distance was adjusted to the resolution and a complete dataset was collected. Diffraction data integration was performed with iMOSFLM (Battye *et*

---

*al.*, 2011; Powell *et al.*, 2013). Data merging and scaling was carried out with Aimless (Evans and Murshudov, 2013; Potterton *et al.*, 2003). Phases were obtained by molecular replacement, using MRage (Adams *et al.*, 2010) from phenix, which using the sequence of CIH protein, found a model from the protein data bank corresponding with a highly conserved hypothetical protein from *Thermus thermophilus* (2cve) valid for molecular replacement. Automatic refinement using Phenix refine was combined with manual cycles of model building using Coot. Translation/Libration/Screw (TLS) refinement was carried out in the last refinement cycles. Refinement was carried out until  $R_{\text{WORK}}$  and  $R_{\text{FREE}}$  values were 19.5 and 23.6 respectively.

### **2.3.2.3 Co-crystallization of the ancient domain with DNA**

Ancient domain co-crystallization with DNA yielded crystals that diffracted up to 1.5Å resolution. A complete dataset was collected by taking 800 images using 0.25 degrees oscillation per frame, thus collecting 200 degrees in one single crystal. Data integration, merging and scaling was performed as described at section 2.3.2.2. Phases were obtained by molecular replacement using as a template our previous model of the ancient domain (5hcs). Refinement of the model was done with Phenix refine, performing rigid body refinement during the first refinement cycles and TLS refinement at the last cycles. Final  $R_{\text{WORK}}$  and  $R_{\text{FREE}}$  values were 14.45 and 17.20 respectively.

## **2.4 Design and expression of mutants**

### **2.4.1 Site directed mutagenesis**

Mutagenesis of full length CIH was accomplished using a modified site directed mutagenesis protocol. In brief, a PCR was carried out with

oligonucleotides designed with the mutation of interest (Table 2.3). All the steps remain as described above (section 2.1) except for the number of cycles (decreased to 16) and the elongation time, adjusted to the insert and plasmid size (6.4 Kb approximately). Afterwards, the PCR product was digested for 1 hour at 37°C with DPN I (New England Biolabs), an enzyme that digests parental DNA by recognizing methylated DNA. After parental DNA digestion, the resulting product was transformed in *E.coli* DH5 $\alpha$  as described before. Plasmid DNA amplification and extraction was performed as already mentioned and the mutations were confirmed by sequencing.

**Table 2.3. Oligonucleotides used for directed mutagenesis.** Forward and reverse oligonucleotides start with an F or R respectively.

Mutation	Oligonucleotides
<b>H203A</b>	F- GCGAGCCGCAACGGCAAACATGACGGCC R- GGCCGTCATGTTTGCCGTTGCGGCTCGC
<b>C221A</b>	F- CAGTTTCCAAGATGCAGATGATGATGGC R- GCCATCATCATCTGCATCTTGAAACTG
<b>D223A/D224A</b>	F- CCAAGATTGTGATGCTGATGGCGCGACGGCAGCTGG R- CCAGCTGCCGTCGCGCCATCAGCATCACAATCTTGG
<b>D223A/E226A</b>	F- TGTGATGATGCTGATGGCGCGACGGCAGCTGG R- CCAGCTGCCGTCGCGCCATCAGCATCATCACA
<b>R232A</b>	F- GCAGCTGGAGGCGCATTACTTCATCTC R- GAGATGAAGTAATGCGCCTCCAGCTGC
<b>Y253A</b>	F- GTGGTGTACGATGGGCCGGCGGTCAAGCTG R- CAGCTTGACGCCCGGCCCATCGTGACACCAC
<b>R261E/R262E</b>	F- CTGGGACCTGAGGAGTTCACGCTC R- GAGCGTGAATCCTCAGGTCCCAG
<b>R262G</b>	F- GGGACCTAGGGGTTTCACGCTCATC R- GATGAGCGTGAAACCCCTAGGTCCC
<b>R271A</b>	F- CAACCAAGTCGCCGCCGACGCATTTCGTTT R- GAACGAATGCGTCGGCGGCGACTTGTTG

---

## 2.5 Nucleic acid binding and hydrolysis assays

### 2.5.1 Agarose EMSA

In order to test the capability of *Chaetomium* IMPACT homolog to bind DNA, an electrophoretic mobility shift assay (EMSA) in low melting agarose (Conda) gel with unlabelled DNA was performed (Hellman and Fried, 2007). A mixture of 50ng of random DNA (control insert of Infusion kit from clontech, 2 kilobase long) with increasing amounts of protein (6, 16, 32, 38, 57 and 64µg respectively) were incubated for five minutes at 4°C in binding buffer (50mM Tris pH 7.5, 150mM NaCl, 1mM β-mercaptoethanol and 5mM MgCl<sub>2</sub>) and run at 8°C in Tris EDTA acetic acid pH 8 (TAE) buffer 1% agarose gel containing GelRed (Biotium) for DNA visualization. The EMSA performed with the RWD domain and the ancient domain was performed as described above but using a single protein concentration of 1mM for each construct.

### 2.5.2 Biolayer interferometry: Dissociation constant

Dissociation constant (KD) was determined accurately by BioLayer Interferometry using BLitZ system (Forte Bio) (Sultana and Lee, 2015). A sample containing 40µg/ml of 20 nucleotide-long 5'-biotinylated single stranded DNA, or hybridized with complementary 20 nucleotide-long (for double stranded DNA) from Sigma-Aldrich was immobilized on Streptavidin biosensors (Forte Bio) previously hydrated with sample buffer (50mM Hepes pH 7.5; 150mM NaCl; 5mM glycerol and 2mM β-mercaptoethanol). Increasing amounts of CIH<sub>R232A</sub> (0µM, 0.125µM, 0.25µM, 0.5µM, 0.75µM, 1µM, 2µM and 7µM) were used in association and dissociation steps. Curve fitting of each triplicate and KD calculation were carried out with BLitZ Pro 1.2 software. The same procedure was followed for the characterization of the RNA dissociation constant, immobilizing



on Streptavidin sensors (Forte Bio) a sample containing 50µg/ml of 15 nucleotide-long 5'-biotinylated polyU from Sigma-Aldrich.

### **2.5.3 DNase activity assays**

Different constructs and mutants of CIH were tested for DNase activity. DNase activity assays were performed in reaction buffer 50mM Tris pH 7.5, 2mM MgCl<sub>2</sub> 150mM NaCl. Protein (1µL of a 1mM dilution) was incubated with DNA (900 nucleotides, 100ng) for 1 hour at 40°C adding 5µL of reaction buffer. Afterwards, 1% agarose gel was run in order to monitor DNA degradation. In the case of the different CIH constructs (section 3.12.6), the pH of reaction buffer was adjusted to 8.5.

Several mutants were designed in order to confirm that the DNase activity was originated from our protein of interest, CIH, and was not due to contaminants from the purification. The absence of activity was checked by running an agarose gel after DNA treatment for every mutant.

### **2.5.4 Metal requirements for DNA cleavage**

In order to eliminate any trace metal originated during the purification, protein was diluted in buffer (50mM Hepes pH 7.5, 150mM NaCl, 2.5mM EDTA and 10% glycerol) to a final concentration of 1mM and afterwards diluted in the same buffer without EDTA and the specific metal (5mM) to a final concentration of 0.3mM. Then the reaction was performed as follows. A mix was prepared with 5µL of reaction buffer (Hepes 7.5, 150mM NaCl, 10% Glycerol and 5mM of the correspondent metal) and dsDNA (100ng final concentration, 800 base pairs long) obtained from PCR of an unrelated gene and purified with E.Z.N.A PCR cycle pure (OMEGA). Only 1µL of the 0.3mM protein solution was added to the respective tubes containing the same metal and the reaction was incubated at 42°C for

---

30minutes. DNA degradation was monitored by running the digestion in an agarose gel (1%). A control with the protein in the buffer solution with 1mM EDTA (final concentration) was performed for each metal tested.

### **2.5.5 Salt requirements for DNA cleavage**

In order to figure out how ionic strength was affecting nuclease activity of CIH, five reaction buffers containing different salt concentrations were tested. The reaction buffer composition was 50mM Hepes pH 7.5, 3mM MgCl<sub>2</sub>, 5% Glycerol and 0, 0.15, 0.3, 0.6 and 1mM NaCl respectively. The reaction mixture (50 ng dsDNA +2µg CIH + 7µL of reaction buffer) was incubated at 40°C for 1 hour. After incubation time, DNA cleavage was monitored for every buffer running the digested samples on 1% agarose gel with GelRed, in order to stain dsDNA and be able to follow degradation. The DNA used, a fragment of 1.1 Kb, was obtained by PCR using as template the ORF of an unrelated gene.

### **2.5.6 Characterization of the DNA cleavage**

Covalently closed circular DNA (cccDNA) is known to present different conformations: supercoiled, relaxed nicked circular and linear. We took advantage of the supercoiled pBR322 to study the CIH mechanism of DNA cleaving. DNA amplification was performed by transformation of pBR322 into DH5α and plasmid extraction using Qiagen miniprep kit following manufacturer's instructions. After isolation of the plasmid, the supercoiled DNA was purified by cutting the band from an agarose gel and further extracted with NucleoSpin® Gel and PCR Clean-up (Macherey-Nagel). DNase activity assay was performed as explained before (section 2.5.3), by incubating 1µg of CIH with the supercoiled DNA for 3 and 60 minutes at 37 °C. The input control was incubated for 60 minutes in reaction buffer at 37 °C with no protein.

## 2.5.7 Hyperchromicity assays

Hyperchromicity assays take advantage of differential absorption of single nucleotides, oligonucleotides or highly polymerized DNA at 260nm to monitor nucleic acid degradation. This technique is not only adequate for exonucleases, but it also permits monitor nucleic acid degradation by endonucleases (KUNITZ, 1950).

The increase in the absorbance at 260nm due to DNA cleavage, was monitored in all the assays using a UV-Visible Spectrophotometer (Ultrospec 3000, Pharmacia Biotech®), and absorbance was measured every 15 seconds until absorbance increment at 260nm was zero. Synthetic Quartz glass cuvettes of 0.5mL volume (Hellma®) with 10mm light path were used for every assay.

### 2.5.7.1 Kunitz assay

One Kunitz unit is defined as the amount of protein necessary to produce an increment of 0.001 units of absorbance at 260nm in one minute, carried out with 1mg/mL of salmon sperm DNA (boehringer bioproducts) in sodium acetate pH 5 buffer and performed at 25°C. In order to calculate the Kunitz present in one milligram of CIH protein, a standard curve was obtained by measuring initial 260nm absorbance increments obtained in 1mL of the DNA solution at 25°C after the addition of 0, 0.3, 1 or 3 DNase I Kunitzs. This standard was used to obtain the Kunitz from the increments obtained from the assays carried out with 120 µg and 244 µg of CIH in the conditions described above.

### 2.5.7.2 Substrate inhibition assay

A fixed concentration of CIH (3.18µM) and different salmon sperm DNA (boehringer bioproducts) concentrations (0, 50, 60, 75, 100, 150 and 200µg/mL)

---

in reaction buffer (50mM Hepes pH 7.5, 150mM NaCl and 5mM MgCl<sub>2</sub>) were used as a substrate for the hyperchromicity assay (KUNITZ, 1950). Reaction was carried out in 1mL quartz cuvette with 600μL of the DNA solution and the absorbance at 260nm was measured immediately after addition of CIH. Due to stability reasons, the assay was carried out only with *C.thermophilum* IMPACT homolog, at 40°C, since it is a thermophilic organism that is able to live at temperatures up to 60°C.

## 2.6 Other bioinformatics tools

Different softwares were used to structurally characterize the protein. Electronic surface potential was calculated using APBS module from Chimera. Surface conservation was rendered with Chimera (Pettersen *et al.*, 2004) upon addition of all reviewed IMPACT sequence and YigZ alignment. Alignment was carried out with multalin web server (Corpet, 1988) and residues were coloured using clustalX (Thompson *et al.*, 2002) as implemented in Jalview (Waterhouse *et al.*, 2009) . Nucleic acid (ssDNA) binding docking was performed using HADDOCK web server (de Vries *et al.*, 2010). Structure comparison and superimposition was done using the MatchMaker module from Chimera (Pettersen *et al.*, 2004) . Pisa web server (Krissinel and Henrick, 2007) was used to assess biological assembly of every domain in solution.

Search for structural homologs was performed using the webserver DALI (Holm and Rosenström, 2010) and prediction of protein disordered regions was carried out with the online server PrDOS (Ishida and Kinoshita, 2007)

## 2.7 .Cells culture and knock down assays

### 2.7.1 3T3 cells culture

Cells were a kind gift from Dr. Susana Masiá. 3T3 (mouse fibroblast) cells were cultured in DMEM with glutamine (Sigma), 10% of fetal bovine serum (FBS) and with Penicillin (100 units/mL) and streptomycin (100 µg/mL) as antibiotics. When 80% of confluency was reached, cells were split in two plates until the amount of cells was enough for the desired experiment (approximately  $1.8 \times 10^6$ ).

### 2.7.2 siRNA transfection for IMPACT knock down

In order to obtain a good gene silencing for IMPACT protein, 3T3 cells were transfected with 10, 25 and 40nM of a commercial pool of 4 siRNA targeting murine IMPACT (Dharmacon) (Agrawal *et al.*, 2003; Fire *et al.*, 1998). A control with the same concentrations of non-targeting siRNA (Dharmacon) was performed in order to compare mRNA levels. The previous day of the transfection, 200.000 cells were plated in each well of a 6- wells plate (P6).

Transfection was carried out as follows. A mixture of the siRNA with OPTIMEM (Gibco) was prepared.

10nM IMPACT: 1µL of siRNA IMPACT (20µM stock) + 199µL OPTIMEM

20nM IMPACT: 2µL of siRNA IMPACT (20µM stock) + 198µL OPTIMEM

40nM IMPACT: 4µL of siRNA IMPACT (20µM stock) + 196µL OPTIMEM

10nM Control: 1µL of non-targeting siRNA (20µM stock) + 199µL OPTIMEM

20nM Control: 2µL of non-targeting siRNA (20µM stock) + 198µL OPTIMEM

40nM Control: 4µL of non-targeting siRNA (20µM stock) +196µL OPTIMEM

---

Afterwards we prepared 6 tubes containing 5 $\mu$ L of dharmafect 4 (Dharmacon<sup>®</sup>) with 195  $\mu$ L OPTIMEM each. After 5 minutes of incubation at room temperature (RT) we added the tubes containing the siRNA in the tubes containing the dharmafect and let incubate for 20 minutes at RT. Meanwhile, we changed the media of the 6 wells for OPTIMEM (1.6 mL). After incubation time, we added the 400 $\mu$ L of mixture to the corresponding wells and let the cells incubate during 6 hours. After the incubation period, the media was replaced by DMEM with glutamine with 10% FBS and penicillin/streptomycin. Cells were harvested for mRNA or protein analysis after 48 and 72 hours respectively. RNA extraction was performed immediately and RNA was kept at -80 $^{\circ}$ C until retro-transcription to cDNA was performed. Retro-transcription was performed using high capacity cDNA reverse transcription kit from applied Biosystems following manufacturer's instructions. Cell pellet for protein extracts were kept at -20 $^{\circ}$ C and protein extract was prepared immediately before western blot was performed.

The same procedure was followed for co-transfecting IMPACT siRNA with interferon stimulatory DNA (ISD) or poly I:C. ISD is a non-CpG 45 base pairs dsDNA oligonucleotide from *Listeria monocytogenes* genome and poly I:C is a double stranded RNA analogue. For ISD co-transfection with IMPACT siRNA, 40nM IMPACT siRNA or non-targeting siRNA for control cells was mixed with 4 $\mu$ L of 25 $\mu$ M ISD solution and 192 $\mu$ L of OPTIMEM. Tubes for poly I:C IMPACT siRNA co-transfection were prepared by mixing 4 $\mu$ g of poly I:C (from 1 mg/mL stock) with 40nM IMPACT siRNA or non-targeting siRNA for control cells and 192 $\mu$ L of OPTIMEM. Tubes containing dharmafect were prepared as described above. All the tubes were incubated for 5 minutes separately and afterwards solution containing siRNA and nucleic acid was added to the respectively tube containing dharmafect and left incubating for 20 minutes at room temperature (RT). The procedure continues as explained above for siRNA transfection.

### 2.7.3 Protein extraction from 3T3 cells

Protein extraction was performed 72 hours after transfection of 20nM and 40nM siRNA IMPACT or non-targeting siRNA in order to find concentration of siRNA suitable for expression silencing. In brief,  $3 \times 10^5$  cells were washed with phosphate saline buffer (PBS) 1X and resuspended with RipA buffer (10mM sodium phosphate, 150mM sodium chloride, 0.5% sodium deoxycolate and 1% NP-40) with protease and phosphatase inhibitors (Roche). Afterwards, the suspension was vortexed for 30 seconds and incubated on ice for 20 minutes. After that period of time Bradford was performed to measure protein concentration and equal amounts of proteins (usually about 20 $\mu$ g) were loaded onto an SDS-PAGE and western blot (WB) was performed with an IMPACT antibody from Santa Cruz Biotechnology (ref: sc-84851) using nitrocellulose membrane and a semidry device from BioRad. Transference was carried out during 1 hour and 15 minutes with a constant voltage of 25mV. Once transference was done, the membrane was blocked by incubating with TBST and 0.5% powder milk during 2 hours. This step was followed by incubation with the primary antibody, goat anti-murine IMPACT (Santa Cruz), performed overnight at 4°C. The next morning the membrane was washed three times with 0.5% milk in TBST for 10 minutes each wash step and then incubated with the secondary antibody anti-goat from abcam® (ref: ab6741) conjugated with horseradish peroxidase for 1 hour. The WB was revealed using Pierce® ECL Western Blotting Substrate (Thermo Scientific). After revealing the WB, the antibody was stripped out of the membrane to perform a loading control with a rabbit anti-mouse actin antibody from abcam® (ref: 15263). In this case, a secondary antibody anti-rabbit fused to horseradish peroxidase from abcam® (ref: 6728) was used to reveal the WB.

---

#### **2.7.4 RNA extraction from 3T3 cells**

After washing each well on a P6 plate with PBS, cells were resuspended with 500  $\mu\text{L}$  of Trizol<sup>®</sup> (Chadderton *et al.*, 1997), and transferred into a 1.5mL tube. 100  $\mu\text{L}$  of chloroform was added to each tube. After vortexing 15 seconds, the tubes were incubated 5 minutes at room temperature and then, centrifuged at 12000g for 15 minutes at 4°C. We transferred the aqueous phase into a new tube carefully, without disturbing the white phase where proteins accumulate. Isopropanol (0.3 mL) was added to the aqueous phase in order to precipitate the RNA, and after vortexing 5 seconds we left the tube incubating for 10 minutes at RT. The mixture was centrifuged at 12000g for 10 minutes at 4°C. The RNA precipitates as a transparent gel. The RNA pellet was washed once with ethanol 75% (0.5 mL) and centrifuged at 7500g for 5 minutes at 4°C. The ethanol was pipetted out and the tube was left open until the RNA pellet was dry. Then, the RNA was resuspended with 30  $\mu\text{L}$  of RNase free water. The RNA extraction was always checked in an agarose gel and the concentration measured with a Nanodrop ND-1000 spectrophotometer (Nanodrop<sup>™</sup>).

#### **2.7.5 Quantitative real-time PCR**

After RNA extraction, reverse transcription PCR (RT-PCR) was performed using a high capacity cDNA Reverse Transcription Kit (Applied Biosystems) following manufacturer's instructions. Once cDNA was obtained, 7500 fast quantitative real-time PCR (qPCR) (Applied Biosystems) was used to quantify cDNA levels corresponding to mRNA from beta and gamma interferons (Wong and Medrano, 2005). The mix for the reaction was prepared by adding 8 $\mu\text{L}$  of 1:10 diluted cDNA, 1 $\mu\text{L}$  of each oligo at 2 $\mu\text{M}$  and 10 $\mu\text{L}$  of Power SYBR green (Applied Biosystems). The reaction was carried out with a hold step at 95°C for 10 minutes followed by 40 cycles of 95°C during 15 seconds and 60°C for one minute. The



software v.2.0.4 of the 7500 fast quantitative real-time PCR (Applied Biosystems) was used to do comparative analysis following the  $2^{-\Delta\Delta Ct}$  method (Livak and Schmittgen, 2001). Graphical display and analysis was carried out with 7500 software v.2.0.4 (Applied Biosystems). Differences between control and experimental groups were evaluated with Student's *t*-test.

The sequences of the oligonucleotides used for mRNA levels quantitation are available at table 2.4.+

**Table 2.4 Oligonucleotides used for qPCR.** The correspondent sequence for each primer is shown.

Primer	Sequence
Actin mouse F	5'-ATCATGTTTGAGACCTCAACACCCC-3'
Actin mouse R	5'-CACGATTTCCCTCTCAGCTGTGG-3'
Impact mouse F	5'-GAACGCTCCCTGGCTG-3'
Impact mouse R	5'-TCTTCGGACTCAACCTCAAC-3'
Interferon $\beta$ mouse F	5'-ATGCAGAAGAGTTACTACTGCC-3'
Interferon $\beta$ mouse R	5'-AGTTGAGGACATCTCCCAC-3'



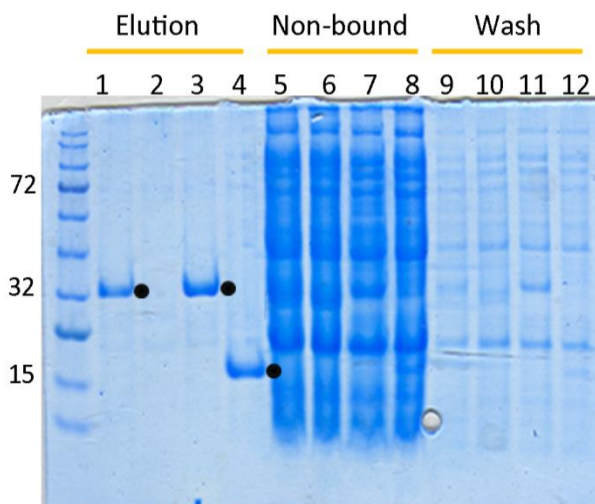


### **3. RESULTS & DISCUSSION**



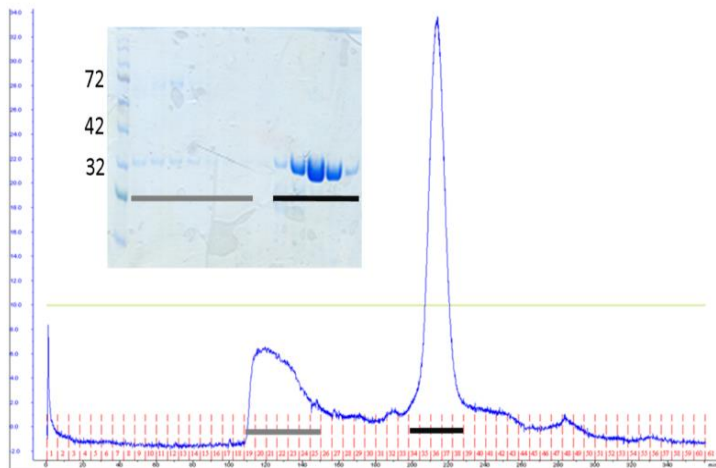
### 3.1 Expression and purification of Yih1

Yih1 was successfully expressed and purified from three of the four constructs from Dr. Bertrand Séraphin: pBS4717, pBS4719 and pBS4720. Small scale purification revealed that the plasmid pBS4718 was not expressing any detectable Yih1. However, GCN1 expression was not detected in any of the plasmids available (Figure 3.1).



**Figure 3.1. 10% SDS-PAGE analysis of small scale purification.** Lines 1 to 4 were loaded with the Nickel-beads elution fraction from bacteria expressing pBS4717, pBS4718, pBS4719 and pBS4720 constructs respectively. The non-bound fraction and the wash fraction from the Nickel-beads, follow the same loading order as lines 1 to 4. Black dots indicate the presence of Yih1 soluble protein. Numbers at MW line indicate the weight in kDa for reference bands.

Since plasmids pBS4718 and pBS4720 carry the same construct boundaries of Yih1 but different boundaries for GCN1 fragment, we do not discard the possibility that small expression levels of GCN1 in pBS4720 construct not detected by SDS-PAGE could be helping in Yih1-RWD domain expression. Afterwards, we decided to prepare a construct carrying only the full length sequence for Yih1 in order to be completely sure that we were having Yih1 proteins with no GCN1 traces. Yih1 was expressed with IPTG induction method. Soluble fraction of the protein was purified in two steps: Affinity chromatography and size exclusion chromatography. This last step was used to achieve structurally homogeneous, crystallization grade, pure protein.

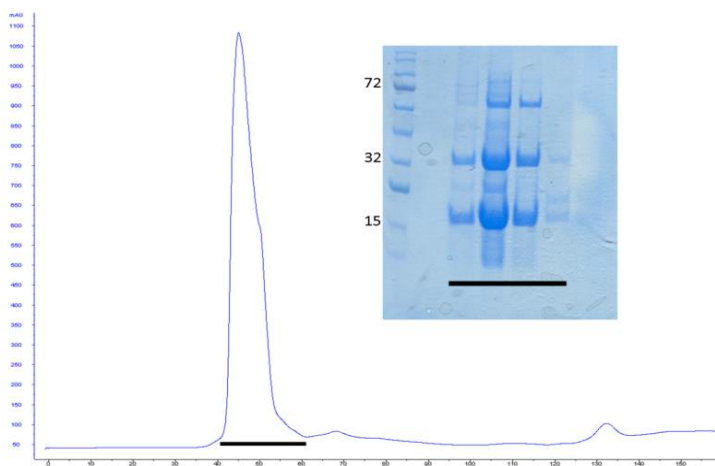


**Figure 3.2.** Yih1<sub>FL</sub> (pBS4717 construct) size exclusion chromatography in Superdex 200 26/60. Fractions marked with a colored bar have been marked with the same color in the gel. Numbers at MW line in the gel indicate the weight in kDa for reference bands. The peak marked with a black bar has an elution volume of 211mL and the one marked with light grey bar 131.7mL.

Protein obtained from the expression of plasmid pBS4719 in *E.coli* BL21C+ RIPL showed exactly the same elution pattern as Yih1<sub>FL</sub> obtained from bacteria expressing pBS4717. This result is not surprising since both plasmids carry the same Yih1 boundaries. Both Yih1 full length (Yih1<sub>FL</sub>), showed two peaks at the size exclusion chromatography (Figure 3.2), which is not surprising regarding that both plasmids carry the same Yih1 boundaries. One peak was corresponding to a monomer, with an elution volume that was varying from 211 to 209 mL, which corresponds approximately to a 35 kDa size. A wide peak was always found to elute at volumes corresponding to very high molecular weight (centred at 130mL which corresponds to a molecular weight of approximately 500 kDa). SDS-PAGE analysis revealed the presence of Yih1<sub>FL</sub> and some faint bands with higher molecular weight. Yih1<sub>FL</sub> has been demonstrated to be present in polyribosomes (Waller *et al.*, 2012) and thus we cannot discard the idea that the peak corresponding to a 500 kDa could be a complex formed by some polyribosome interactors that remained bound to Yih1 until this last step. The purification procedure followed for Yih1<sub>FL</sub> obtained from bacteria containing pBS4717 or pBS4719 plasmids was appropriate for the given protein, since the obtained

amount and purity degree were suitable for crystallization and subsequent assays. The same strategy was applied to Yih1<sub>FL</sub> pOPINF constructs with successful results and the protein showed the same behaviour in size exclusion chromatography as Yih1<sub>FL</sub> protein obtained from the expression of pBS4717 or pBS4719 plasmids in *E.coli* BL21C+ RIPL.

On the other hand, size exclusion chromatography of Yih1 RWD domain expressed from *E.coli* BL21C+ RIPL transformed with pBS4720, eluted in size exclusion with a volume much higher than the expected for a monomer.



**Figure 3.3.** *Yih1-RWD domain (pBS4720 construct) size exclusion chromatography in Superdex 75 16/60.* Fractions marked with a colored bar have been marked with the same color in the gel. Numbers at MW line indicate the weight in kDa for reference bands. The elution volume of the peak is 48 mL.

The SDS-PAGE analysis of the eluted peak (Figure 3.3) showed the presence of bands corresponding to proteins with higher molecular weight. In order to discard abnormal migration of the RWD domain of Yih1 on the SDS-PAGE, mass spectrometry analysis was carried out for the three strongest bands (migrating as 15, 32 and 72 kDa). The analysis confirmed that the band at 15 kDa was Yih1-RWD domain, but the upper bands were contaminants from *E.coli*. It is worth to mention that the band with a molecular weight of 32 kDa corresponded to the ribosomal protein S2 (RPS2). This finding is very interesting since it has been published that Yih1 resides in a complex with ribosomes. The band

---

observed approximately at 72 KDa corresponds to a protein from *E.coli* known as ARNA. Taking into account that the elution volume of the peak in the Superdex 75 16/60 corresponds to an approximated molecular weight of 80 KDa, it is more likely that ARNA protein in that peak is in a monomeric conformation that co-eluted with a complex composed by the RPL2 and the RWD domain from Yih1.

As mentioned in the introduction (1.5.1.3) a fragment corresponding to the central part of the protein (residues 68 to 171) is sufficient for detecting Yih1 together with polysomes. The fact that Yih1-RWD domain co-purifies with RPS2 is suggesting that this domain might be also involved in ribosome binding. In order to prove so, RNase treatment was performed during Yih1-RWD purification. After size exclusion, RPS2 was still co-eluting with Yih1-RWD, confirming that the interaction was not mediated by RNA. RPS2 has found to be exposed to the solvent in the 30S ribosome (Boehringer *et al.*, 2012), but when the binding between both subunits of the ribosome takes place, RPS2 remains in an inner position. This result could suggest that Yih1 binding to polysomes might be performed involving different proteins and is not depending only on RPL39 (Waller *et al.*, 2012), but RPS2 location within the mature ribosome would hinder the interaction with Yih1.

### **3.2 Yih1 crystallization**

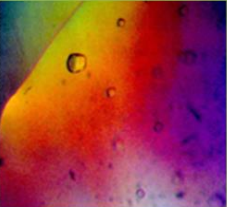

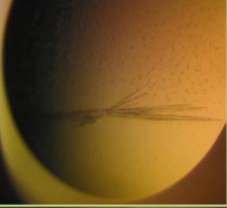


To shed light into new possible functions of IMPACT and its homologs, it was necessary to perform a structural study of the protein. Given the group expertise and the resources available at the institute, X-ray crystallography was the technique of choice for developing this study. The first crystallization attempts were performed with the yeast IMPACT homolog (Yih1).

Purified Yih1<sub>FL</sub> from *E.coli* BL21C+ RIPL transformed with plasmids pBS4717, pBS4719 or Yih1<sub>FL</sub> pOPINF was used to screen for setting up crystallization conditions of the full length protein. Tiny crystals of Yih1<sub>FL</sub> obtained



from *E.coli* BL21 C+ RIPL transformed with pBS4717 appeared in some of the INDEX (A6, A4 y C5) and JBScreen Classic HTS II (A2) (Table 3.1).

**Table 3.1. Conditions where Yih1<sub>FL</sub> crystals were obtained.** Right images correspond to a picture of the obtained crystals illustrating morphology.

Screening	Conditions	
Index A6	0.1M Tris pH 8.5 2M Ammonium sulphate	
Index A4	0.1M Bis-Tris pH 6.5 2M Ammonium sulphate	
Index C5	60% v/v Tacsimate™ pH 7	
JBS HTS II	1M Ammonium sulphate 100mM Sodium Acetate pH 4.6	
Amplification	0.1M Tris pH 8.5 2M Ammonium sulphate 5% Glycerol	

Those crystals were diffracting up to 8Å, so a screening around the initial condition was performed in order to improve the resolution of the crystal diffraction. Larger Yih1 crystals were obtained at 0.1M Tris pH 8.5, 2M ammonium sulphate and 5% glycerol with 30mg/mL of Yih1<sub>FL</sub> at a 1:1 ratio by mixing 0.5µL of protein solution and 0.5 µL of the crystallant condition. Those crystals diffracted up to 3.8 Å. After this result, in order to continue improving resolution, crystals were grown in the presence of a set of additives (Table 3.2), but no improvement on the resolution was achieved.

**Table 3.2. Additives present at the improving Yih1<sub>FL</sub> crystallization condition and diffracting resolution.** Final concentration and type of molecule have been indicated. The mother liquor, common to all additives was 0.1M Tris pH8, 2M ammonium sulphate and 5% glycerol.

Additive	Type	Resolution
0.1M Ectoine	Zwitterion	5.5Å
1M Glycine	Zwitterion	7Å
1M Pyridine	Non-ionic	20Å
0.1M Phenol	Non-ionic	No diffraction
0.1M Betaine monohydrate	Zwitterion	20Å
0.1M Taurine	Zwitterion	No diffraction

### 3.2.1 Yih1<sub>FL</sub> Data processing

As mentioned before, the best diffracting resolution obtained for Yih1 crystals was 3.8 Å. Data merging and scaling carried out with SCALA (Evans, 2011) showed that, due to radiation damage the final resolution of the dataset was 5.3

Å. Crystal belonged to space group I222 and cell dimensions were 110.2, 134.8 and 173.8 with three angles of 90 degrees, enough to accommodate 4 full length Yih1 molecules into the asymmetric unit with 51% of solvent (Matthews coefficient 2.52).

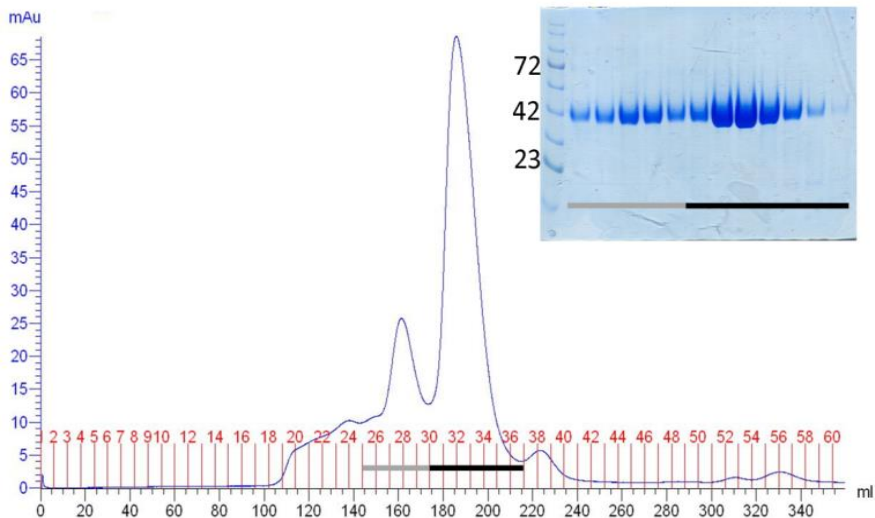
Any further attempts to increase resolution were unsuccessful.

### **3.3 IMPACT expression and purification**

Regarding the limited understanding of the molecular mechanisms involving IMPACT proteins and given the high conservation degree among all the homologs, we considered that overexpression and purification of human IMPACT would be very interesting not only for structural purposes, but also to prove possible new protein features. Protein IMPACT with a 6 histidines tag at the N-terminal was expressed with IPTG induction method as explained in materials and methods. Soluble protein was firstly purified by affinity chromatography and after concentration of the fractions of interest; a size exclusion step was performed in order to obtain a homogenous sample (Figure 3.4).

The obtained elution pattern shows two main populations with different oligomeric states. The peak with an elution volume of 161mL corresponds to a tetrameric conformation whereas the peak eluting at 186mL corresponds to a dimeric conformation.

Purification strategy was the appropriate since enough amounts of soluble protein were obtained. Unfortunately, crystallization attempts performed with all the resources available at the institute did not succeed.

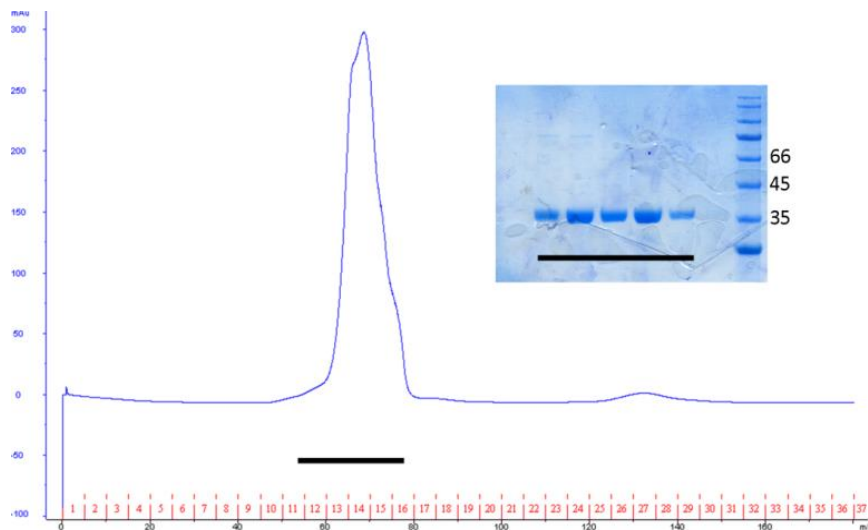


**Figure 3.4.** Size exclusion chromatography of IMPACT protein. Elution pattern from Superdex 200 26/60 is shown. First eluting peak, marked with a grey bar is centred at 161mL and the second peak has its maximum value at 186mL. Correspondent fractions have been marked with the same colour in the 10% SDS-PAGE. Numbers at MW line indicate the weight in KDa reference bands.

### 3.4 CIH expression and purification

The same strategy as the one used for Yih1 and IMPACT has been used for purification of *C.thermophilum* IMPACT homolog (CIH). Different CIH constructs have been successfully expressed and purified (see table 2.1). Except CIH<sub>152-292</sub>, the rest of the constructs were purified using HisTrap column followed by size exclusion chromatography.

CIH full length (CIH<sub>FL</sub>) expressed very well in *E.coli*. Due to the high amount of protein obtained from the pellet of 1 litre culture, (approximately 60mg) the protein was eluted in two peaks in size exclusion (Figure 3.5) corresponding to dimeric and monomeric forms. When lower amounts of protein were injected in size exclusion, the elution pattern was consistent with a unique monomeric form.

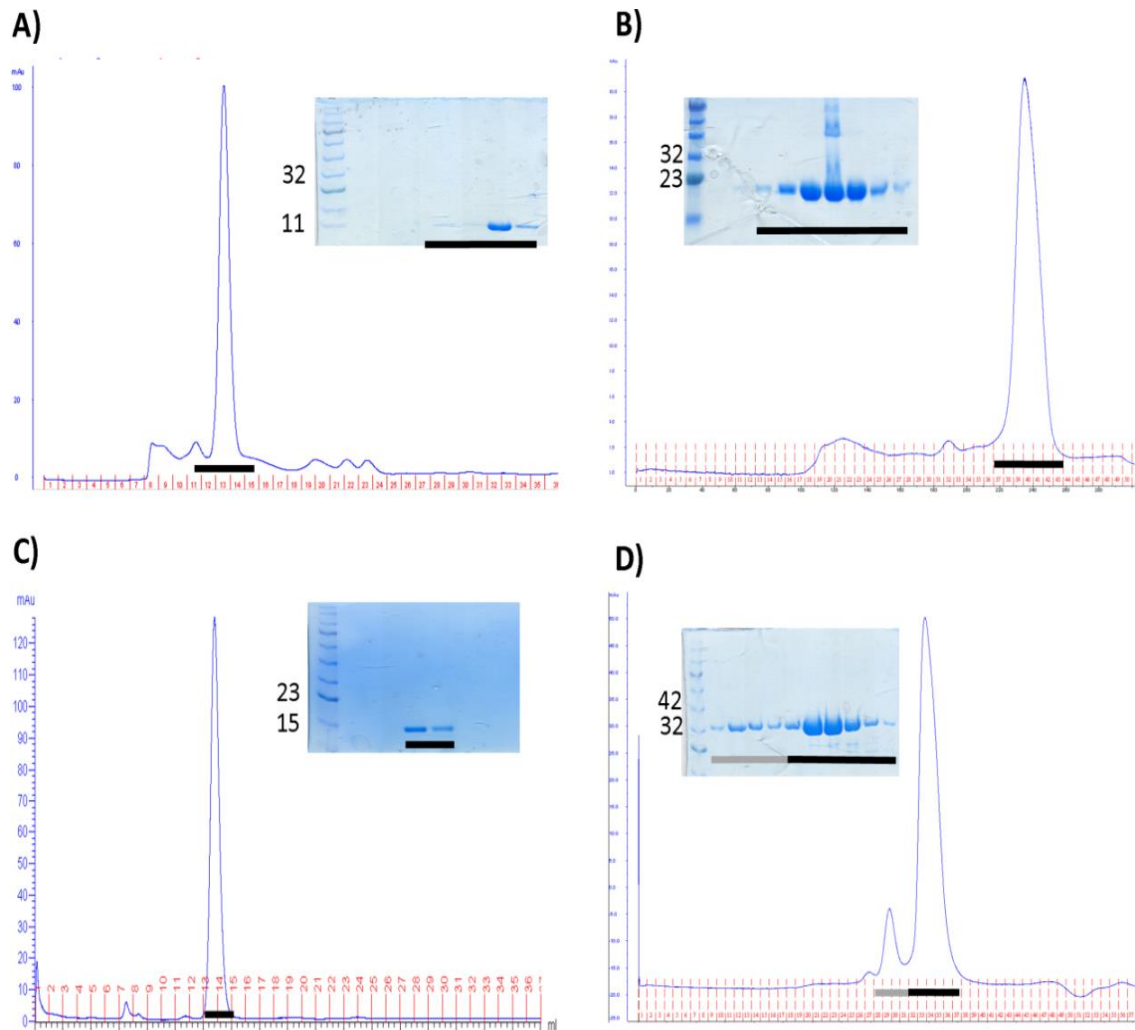


**Figure 3.5** CIH<sub>FL</sub> size exclusion chromatography in Superdex S200 16/60. Fractions marked with a colored bar have been marked with the same color at the gel. Numbers next to MW line indicate the weight in KDa for reference bands. The elution volume of each shoulder of the peak is 77 and 82mL, which corresponds with a dimer and monomer respectively.

The 10% SDS-PAGE is clearly showing the presence of the two populations, since it can be observed how the protein concentration increases and decreases in the chromatogram before increasing again to previous levels. This pattern was not observed when the amount of protein injected in size exclusion was reduced (data not shown) and only a peak corresponding to a monomeric conformation was appearing.

Purification of a construct lacking the amino terminal domain (CIH<sub>111-292</sub>) showed two peaks in size exclusion chromatography. The peak marked with the black bar elutes with an approximated size of a dimer (two molecules of 20 KDa), while the peak marked with a light grey bar corresponds to approximately 190 KDa which corresponds with an oligomer around 10 molecules (Figure 3.6, panel A). The ancient domain (CIH<sub>152-292</sub>) after GST tag removal was injected in size

exclusion and showed one elution peak corresponding to a monomer (Figure 3.6, panel C).



**Figure 3.6** Size exclusion elution pattern for different CIH constructs. A) CIH<sub>1-111</sub> size exclusion chromatography in Superdex 200 26/60, B) CIH<sub>1-160</sub> size exclusion chromatography in S200 26/60, C) CIH<sub>152-292</sub> size exclusion pattern in S75 16/60 and D) CIH<sub>111-292</sub> size exclusion chromatography in S200 26/60. Fractions marked with a colored bar have been marked with the same color at the gel. Numbers next to MW line indicate the weight in kDa for reference bands.

The two peaks observed in CIH<sub>111-292</sub> chromatogram correspond with a elution pattern for a dimer and a monomer ( 137mL and 158mL respectively). The rest of the constructs produced for CIH were behaving in solution as expected, with elution volumes for size exclusion chromatography that were corresponding to monomers.

In general, selected buffers for purification were appropriate for IMPACT, Yih1 and CIH proteins. Size exclusion elution patterns for the three IMPACT counterparts show a clear dimerization tendency of full length proteins.

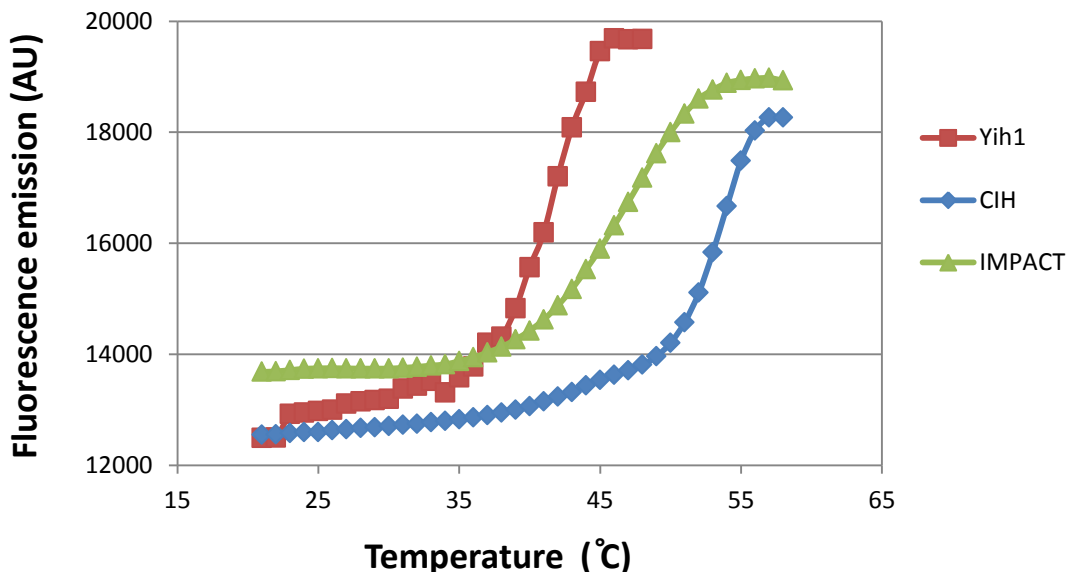
### **3.5 *Chaetomium* IMPACT is more thermostable than Yih1 or IMPACT**

*Chaetomium thermophilum* is a thermophilic fungus of the phylum Ascomycota that can live with temperatures up to 60°C. Due to its higher thermostability it was a good candidate for protein crystallization since the stability of the protein has been related with the success in crystallization and diffraction (Reinhard *et al.*, 2013). The stability of Yih1, IMPACT and *Chaetomium* IMPACT homolog (CIH) in different buffer solutions was measured with differential scanning fluorimetry (Figure 3.7) (Boivin *et al.*, 2013).

As expected, CIH showed higher melting temperature compared to Yih1 or IMPACT. Precisely, CIH presented a melting temperature of 53°C, which is approximately 12°C and 7°C higher than melting temperatures observed for Yih1 (41.5°C) and IMPACT (46°C )respectively. Although IMPACT seems to be more stable than Yih1, IMPACT is less soluble than Yih1. This fact could be the reason why IMPACT does not easily crystallize.

Protein stability and solubility has been directly related with success in crystallization and good diffraction quality (Ericsson *et al.*, 2006) . Since we could not obtain crystals for IMPACT, Yih1 showed poor diffraction quality and, in general, the structure is more conserved than sequence across species, we considered CIH as a good IMPACT homolog for obtaining the three dimensional

structure due to its higher *in vitro* stability and solubility (up to 230mg/mL) .



**Figure 3.7. Differential scanning fluorimetry comparison between Yih1, IMPACT and CIH.** IMPACT denaturation process in 50mM HEPES pH 7.5, 150mM NaCl, 5% Glycerol and 1mM  $\beta$ -mercaptoethanol can be followed by the increase of fluorescence emission. Temperatures in Celsius degrees are indicated in the X axis.

### 3.6 Crystallization trials of CIH<sub>FL</sub> yielded ancient domain crystals due to *in situ* proteolysis

Many crystallization trials were setup for the full-length CIH protein. Initial crystals were obtained using the vapour diffusion method at 21°C. Crystals were obtained using 60mg/mL of more than 95% pure protein (Figure 3.8). Crystal dimensions were approximately 40x20x5  $\mu$ m (long:wide:deep). These crystals were diffracting up to 2Å in XALOC beamline (ALBA synchrotron).





**Figure 3.8: Ancient domain crystals from *C.thermophilum* IMPACT homolog.** CIH<sub>152-292</sub> crystals in JBS II (Jena Bioscience) (15 % w/v 2-propanol, 100 mM Hepes sodium salt pH 7.5 and 200 mM Magnesium Chloride).

### 3.6.1 Data processing and refinement

Complete dataset was collected at XALOC beamline, ALBA synchrotron. The crystals belonged to C2221 space group. Cell parameters (Table 3.3) were not large enough to accommodate one single molecule of the full length protein hence, a proteolytic event was suspected to occur.

Molecular replacement performed with MRage as implemented in Phenix, gave as a result one molecule of CIH C-terminal domain within the asymmetric unit. The crystallized part corresponded to the whole ancient domain, where visible residues after refinement were in the range 152 to 292. Proteolysis inside the drop was thus confirmed.

The diffraction obtained from the crystal was showing certain anisotropy degree, where axis h and l were showing a maximum resolution of 2.03Å whereas the k axis was showing a maximum resolution of 2.5Å. Refinement was carried out at 2.03Å resolution until the model, reached  $R_{\text{WORK}}$  and  $R_{\text{FREE}}$  values of 19.5/23.6 respectively. The model has 99% of residues within the favourable Ramachandran regions and no Ramachandran outliers. One rotamer outlier, correspondent to valine 178, is present in the model. The protein data bank code correspondent to this model is 5hcs. All the refinement statistics can be found at table 3.3.

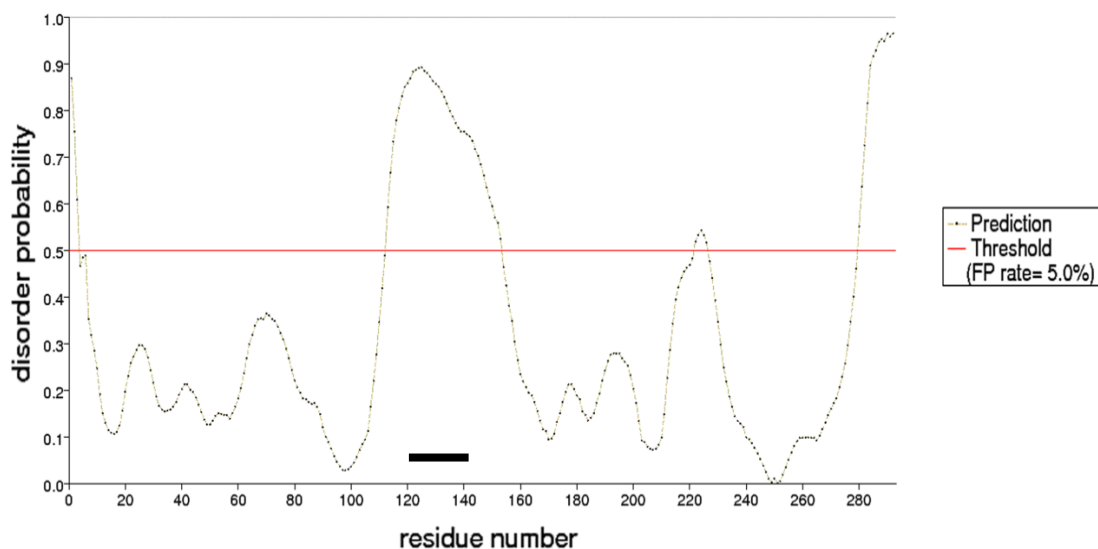
**Table 3.3 Data collection and refinement statistics for the ancient domain.** Statistics for the highest resolution shell are shown in parentheses.

<b>Wavelength (Å)</b>	0.9795
<b>Resolution range (Å)</b>	33.2 - 2.026 (2.098 - 2.026)
<b>Space group</b>	C2 2 2 1
<b>Unit cell</b>	71.75 97.36 45.4 90 90 90
<b>Total reflections</b>	930519 (86216)
<b>Unique reflections</b>	10676 (1056)
<b>Multiplicity</b>	8.5 (7.7)
<b>Completeness (%)</b>	99.9 (99.9)
<b>Mean I/sigma (I)</b>	11.51 (2.86)
<b>Wilson B-factor</b>	39.62
<b>R-merge</b>	0.078 (0.414)
<b>R-meas<sup>a</sup></b>	0.086 (0.477)
<b>CC1/2</b>	0.995(0.978)
<b>CC*</b>	0.987 (0.995)
<b>R-work<sup>b</sup></b>	0.1952 (0.3096)
<b>R-free<sup>c</sup></b>	0.2357 (0.4381)
<b>Number of non-hydrogen atoms</b>	1004
<b>Macromolecules</b>	979
<b>Water</b>	25
<b>Protein residues</b>	127
<b>RMSD<sup>d</sup> (bonds)</b>	0.007
<b>RMSD<sup>d</sup> (angles)</b>	1.04
<b>Ramachandran favored (%)</b>	99
<b>Ramachandran outliers (%)</b>	0
<b>Clashscore</b>	4.65
<b>Average B-factor</b>	61.10
<b>Macromolecules</b>	61.10
<b>Solvent</b>	60.10

<sup>a</sup>  $R_{\text{meas}} = \{\sum_{hkl} [N/(N-1)]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle|\} / \sum_{hkl} \sum_i I_i(hkl)$ , where  $I_i(hkl)$  are the observed intensities,  $\langle I(hkl) \rangle$  are the average intensities and N is the multiplicity of reflection  $hkl$ . <sup>b</sup>  $R_{\text{work}} = \sum_{hkl} \{|F_{\text{obs}}(hkl) - [F_{\text{calc}}(hkl)]\} / \sum_{hkl} [F_{\text{obs}}(hkl)]$ , where  $F_{\text{obs}}(hkl)$  and  $F_{\text{calc}}(hkl)$  are the structure factors observed and calculated, respectively. <sup>c</sup>  $R_{\text{free}}$  corresponds to Rfactor calculated using 5% of the total reflections selected randomly and excluded during refinement. <sup>d</sup> RMSD is the root mean square deviation

### 3.6.2 *In situ* CIH<sub>FL</sub> proteolysis due to the presence of PEST regions

We looked for protease susceptibility sequences as candidate targets and two PEST regions were found. The first PEST motif is composed by 19 amino acids between amino acid 121 and 141 and the second is found between residues 145 and 168. The major part of the area for both PEST motifs is predicted to be unstructured and thus, accessible to proteases (Ishida and Kinoshita, 2007) (Figure 3.9). This finding is very common since PEST motifs are known to display intrinsic unstructuredness (Singh *et al.*, 2006).



**Figure 3.9. Disorder probability of CIH structure based on sequence.** Residues above the threshold, marked as a red line, are considered to be part of an unstructured region. The black bar indicates the position of the PEST sequence among CIH residues. Prediction performed with PrDOS (Ishida and Kinoshita, 2007).

These PEST sequences are also present in human IMPACT and its yeast counterpart Yih1 and they are located in an equivalent area: the linker region. PEST sequences are known to be involved not only in fast protein turnover and

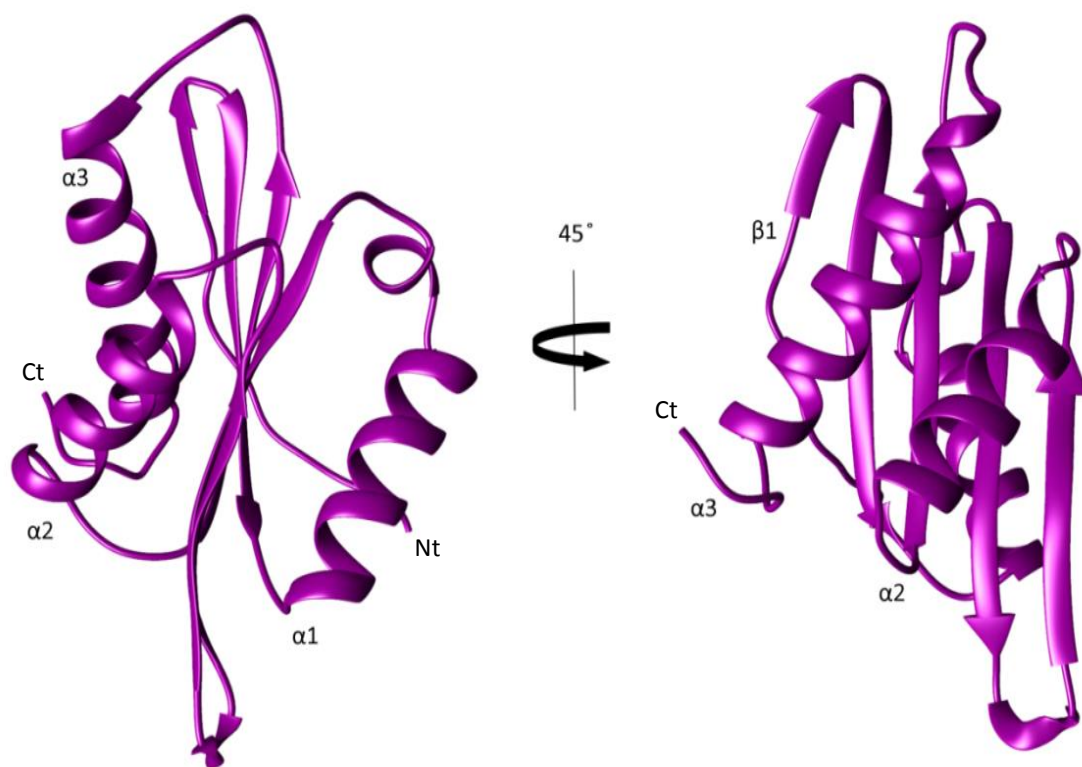
---

regulation, but also in several cellular processes as phosphorylation or protein-protein interaction (Sandhu and Dash, 2006). Both PEST regions are located within the fragment necessary for actin binding and polyribosome co-sedimentation. This observation suggests that the PEST sequence could be involved in fast downregulation of IMPACT related functions such as the reported translational control.

### **3.7 Ancient domain structural characterization**

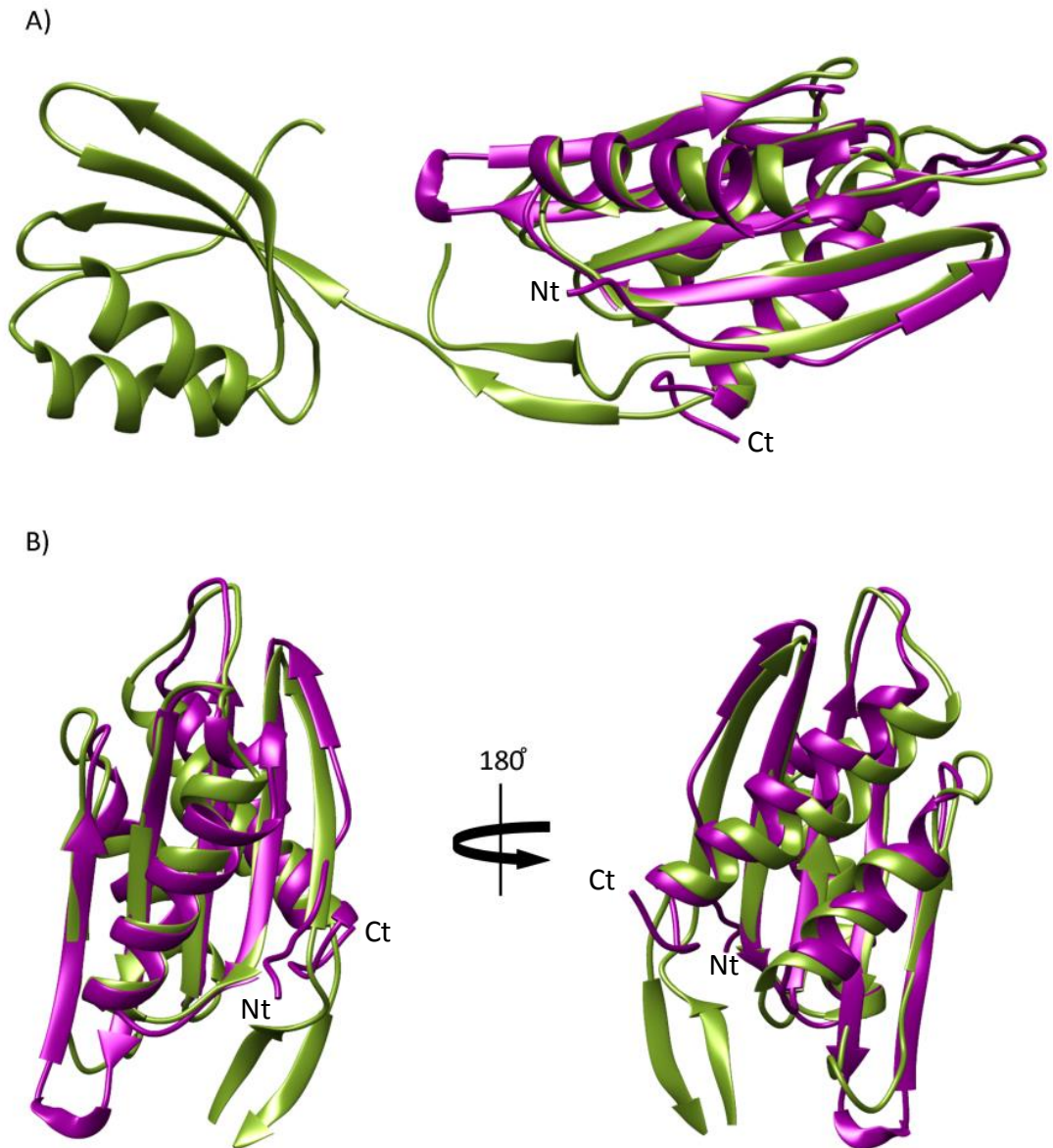
The ancient domain is the sole domain composing the unknown protein family 0029 (UPF0029) also known as protein family 01205 (Pfam: <http://pfam.xfam.org/>). This domain is found only in IMPACT proteins and YigZ, a protein present in bacteria. This last protein is considered part of the IMPACT protein family. There is a structure available for YigZ from *E.coli* at the protein data bank (1vi7) (Park et al., 2004), so the obtained structure for IMPACT ancient domain is very useful to corroborate structural homology and that they do belong to the same family indeed.

Ancient domain from *C.thermophilum* IMPACT homolog displays a  $\beta\beta\alpha\beta\beta\alpha$  topology as predicted. This topology is very similar to the one found in ferredoxins, however the three dimensional structural organization of both motifs is completely different. Three dimensional structure of CIH shows a  $\beta$ -sheet composed of 5 anti-parallel strands that is surrounded by two  $\alpha$  helices in one side and another helix on the opposite side (Figure 3.10).

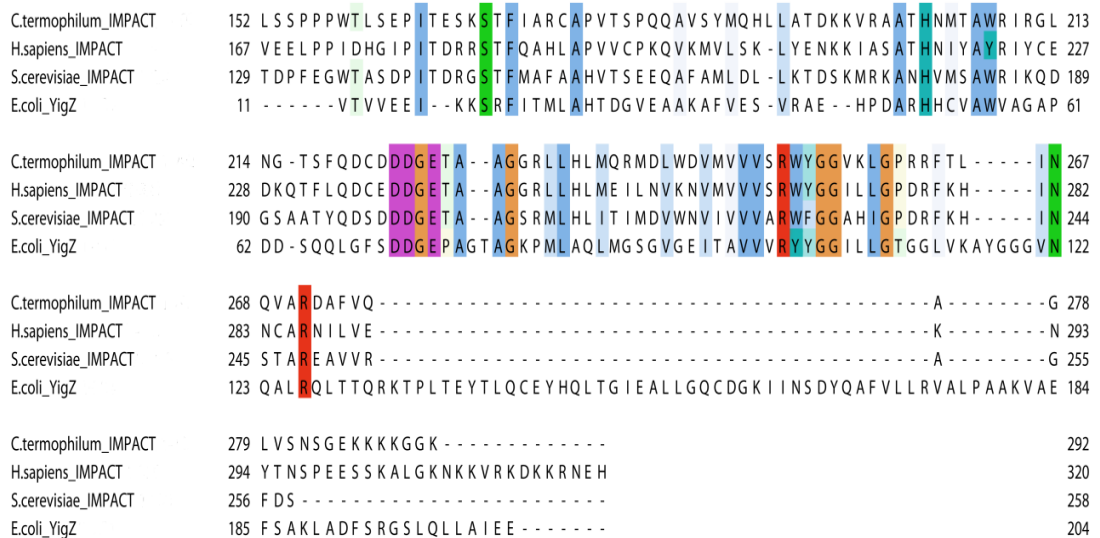


**Figure 3.10. Ribbon representation of CIH Ancient domain.** Two different views are shown for the ancient domain. Degrees and direction of the rotation between both representations are indicated. Helices and the first  $\beta$ -strand have been numbered according to the appearance in the sequence. N-terminal and C-terminal regions are indicated with Nt and Ct respectively.

This organization, namely an antiparallel  $\beta$ -sheet surrounded by  $\alpha$ -helices, represents a common fold among proteins. The structure for CIH ancient domain matches almost perfectly with the ancient domain from YigZ (Park *et al.*, 2004) (RMSD 0.965 for 73 residues) (Figure 3.11). This result is not surprising regarding the high degree of sequence conservation found among ancient domains (Figure 3.12).

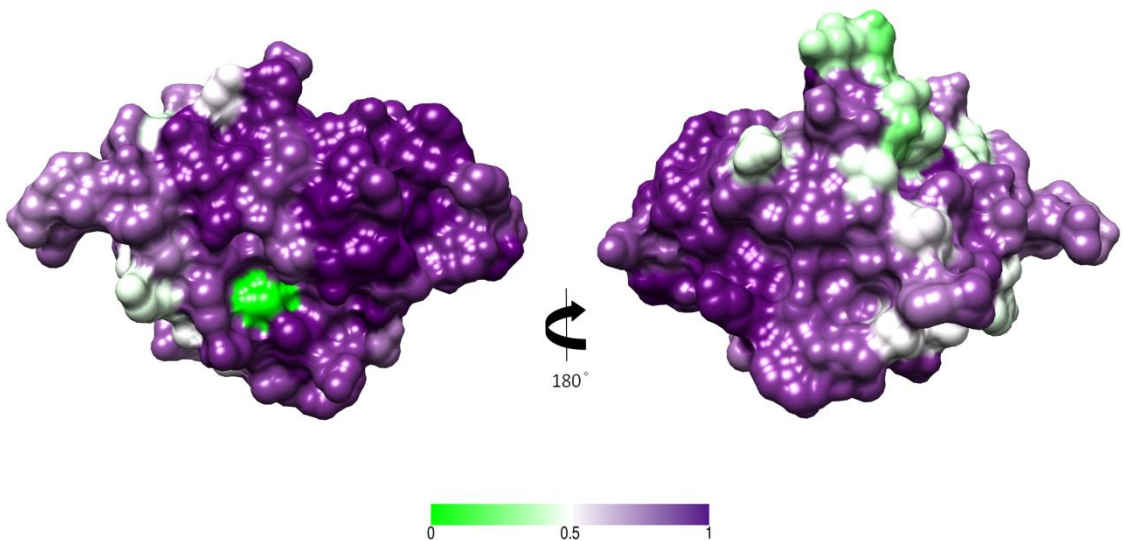


**Figure 3.11. Ribbon superposition of ancient domains from YigZ and CIH.** Ribbon superposition of the ancient domain of CIH, rendered in magenta, with YigZ full length structure depicted in olive green (Panel A). Two views of the structural superposition of both ancient domains rotated 180° (Panel B). N-terminal and C-terminal regions are indicated with Nt and Ct respectively.



**Figure 3.12** Sequence alignments among YigZ ancient domain and the ancient domain of IMPACT proteins from *C.terminophilum*, *H.sapiens* and *S.cerevisiae*. Residues have been coloured by conservation using clustalX as implemented in Jalview.

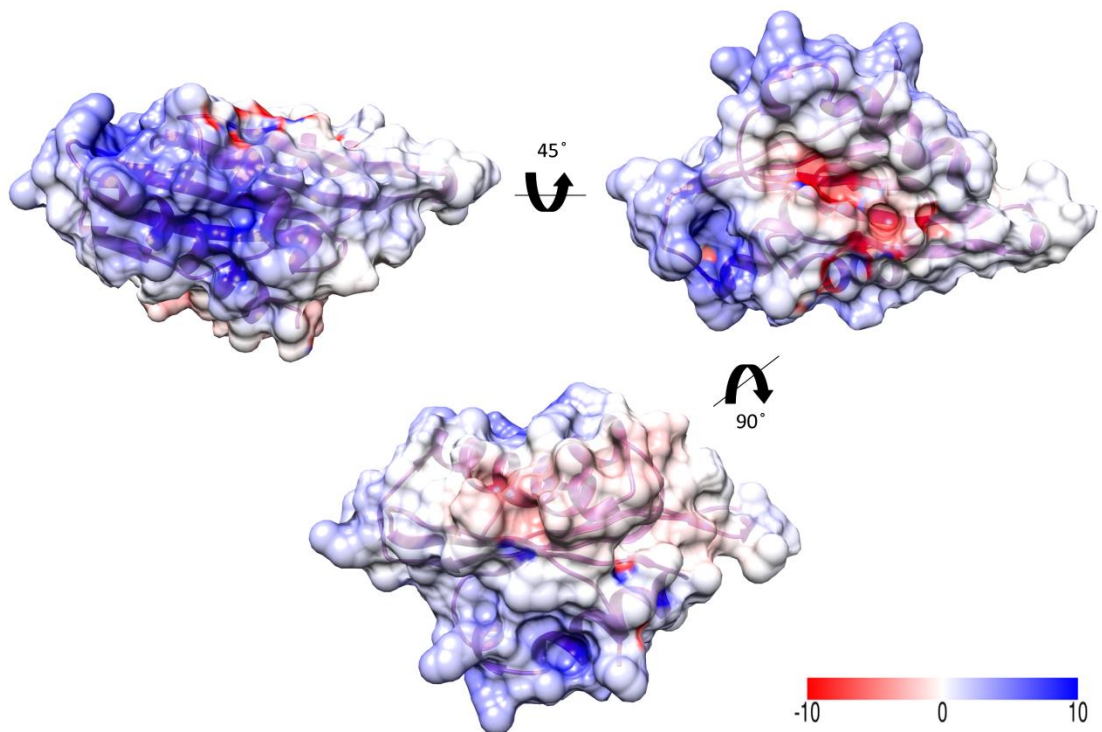
Residue conservation becomes evident in a surface representation rendered depending on the aligned sequences among ancient domains from all reviewed IMPACT proteins at Uniprot, and the only reviewed sequence at Uniprot for YigZ, which belongs to the protein from *E.coli* (Figure 3.13). The conservation shows the peculiarity of being much higher in one half of the domain structure. Taking into account that in some organisms as humans, there is an isoform generated by alternative splicing that lacks the RWD domain (Gerhard *et al.*, 2004) and the strong sequence conservation found in the ancient domain, it turns out to be interesting that no known function developed by the ancient domain has yet been reported.



**Figure 3.13. Surface residue conservation among ancient domains.** Surface residue conservation representation among IMPACT homologs and the ancient domain of YigZ. Conservation is rendered with green and purple. Contour levels are in the range from 0 (not conserved) to 1 (totally conserved) respectively.

Concerning to the electrostatic potential on the surface (Figure 3.14), it is remarkable to say that the two charged patches present at the surface, one small negative area and other bigger positively charged patch, match with the most conserved areas. A particular function for these charged areas might be selecting its conservation degree. It is worth to note that in both charged patches there is a crevice. Those cavities might be indicating two possibilities. Either the charged patches are needed for the interaction with the RWD domain if that interaction exists, or those areas are potential binding surfaces for unknown ligands and finding those ligands might give a clue about the particular function of the ancient domain if any.



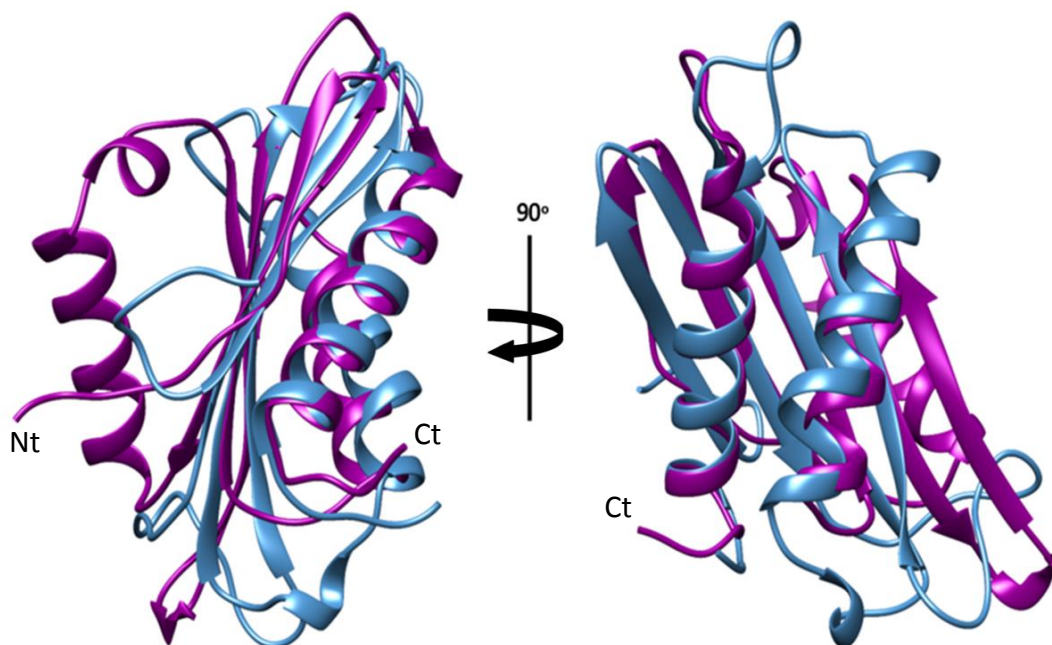


**Figure 3.14. Three different views for the electrostatic surface potential of the ancient domain.** Positively charged areas have been depicted in blue, and negatively charged areas in red. Neutral charged zones have been coloured with white. Contour levels, are in range from -10 to 10 kT/e and have been calculated at 300K.

### 3.8 The ancient domain of CIH shows structural similarity with RNase PH domain

As we have shown before, the sequence and structure of the ancient domain is highly conserved but no function for this domain has been described so far. The three dimensional structure of a protein is almost in all cases directly related with its function. We speculated that, regarding the presence in some organisms such as *C.elegans* or *H.sapiens sapiens* of an isoform of the IMPACT protein composed only by the ancient domain and produced by alternative splicing, it should be playing a role of enough relevance for this alternative

splicing to occur.



**Figure 3.15. Main chain superposition of RNase PH domain from RRP46 and the ancient domain of CIH.** Ribbon representation for both structures. RNase PH domain from RRP46 (4ifd) is depicted in blue and CIH ancient domain in magenta. N-terminal and C-terminal regions are indicated with Nt and Ct respectively.

To shed light into possible functions for the ancient domain, a search performed to find structural homologs was performed using DALI web server (Holm and Rosenström, 2010b). The top two scores in the structural comparison were related to two proteins containing the UPF0029 motif, but interestingly, the third and fourth proteins with an RMSD value of 2.9 for 87 residues and 2.8 for 90 were a polyribonucleotide phosphorilase (4aid) and the exosome complex component RRP46 respectively (4ifd, chain D). In both cases the similarity was related to the RNA nuclease (RNase) PH domain that both proteins harbour (Figure 3.15).

RNase PH in bacteria is a phosphorolytic 3'. Although it uses inorganic phosphate as the nucleophile, the phosphorolytic activity is  $Mg^{2+}$  dependent

(Deutscher and Li, 2001) . Bacterial polynucleotide phosphorylase (PNPase) is the paralog of RNase PH. PNPase is a homotrimeric enzyme that catalyses either RNA phosphorolysis or polymerization of nucleoside diphosphates (Carzaniga *et al.*, 2014) . RRP46 is a protein that forms part of the exosome. The exosome is a multisubunit RNA-processing complex found in archaea and eukaryotes whose core is composed by six RNase PH subunits, architecture that is homologous to the bacterial PNPase. RNase PH domain has been demonstrated to be the catalytic domain for RNA degradation by the exosome in prokaryotes. In *E.coli*, RNase PH has been demonstrated to be involved in rRNA degradation upon starvation of the cultures (Jain, 2012). In eukaryotes, there is only one protein catalytically active called DIS3 and the rest of proteins carrying RNase PH domain are used to guide the RNA molecule to the exosome centre (Ng *et al.*, 2010).

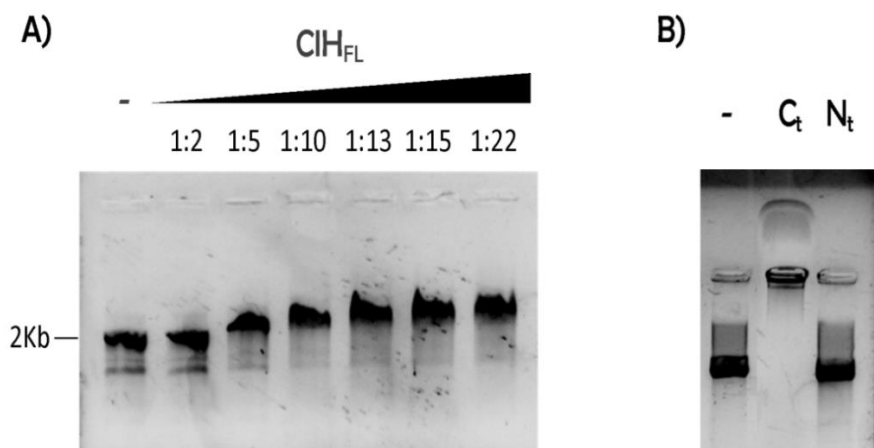
Neither this structural similarity nor an enzymatic activity for any IMPACT homolog has been reported. Usually, structure is directly related to protein function and the similarity found is giving a wide range of potential functions for *Chaetomium* IMPACT protein and given the very high sequence conservation of the domain, probably to all IMPACT proteins. Regarding the structural similarity and the electronic potential distribution on the surface we considered relevant to firstly focus on further investigation about whether CIH, Yih1 or/and IMPACT have any nuclease activity or nucleic acid binding ability.

### **3.9 *Chaetomium* IMPACT homolog binds nucleic acids**

Due to high structural similarity of the ancient domain with the RNase PH domain, an electrophoretic mobility shift assay (EMSA) on agarose was carried out in order to test if CIH<sub>FL</sub> was able to bind nucleic acids. Since DNA has more stability than RNA, we decided to first test the ability to bind double stranded DNA (dsDNA) (Figure 3.16, panel A). The DNA used for the assay, as explained in section 2.5.1, was the insert control used for in-fusion kit, with non-available

sequence.

The obtained result clearly shows that CIH is binding dsDNA. The resulting migration indicates that there is likely to be more than one DNA binding site, since increasing amounts of protein (from 1:2 to 1:21.5 molar ratios) over the equimolarity were augmenting the band shift. Nevertheless, more tests have been performed to prove unspecificity of the binding. Those experiments will be discussed later.



**Figure 3.16. Agarose EMSA with dsDNA and CIH.** Two kilobase (Kb) length dsDNA was incubated with increasing amounts of CIH<sub>FL</sub> protein. The first line, on the left, corresponds to the migration of the 2Kb dsDNA in the absence of protein (Panel A). EMSA performed with the CIH ancient domain (C<sub>t</sub>) or the RWD domain (N<sub>t</sub>) (Panel B). In both gels, the first line corresponds to the input DNA (2Kb) in absence of any CIH.

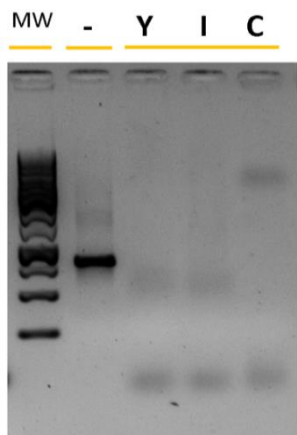
The ability of binding DNA is an uncharacterised new feature of the CIH<sub>FL</sub>. The next step was to clarify whether the DNA binding was carried out through the ancient domain, the RWD domain or whether both domains were necessary to bind DNA (Figure 3.16, panel B). Abnormal migration of the DNA is observed in the presence of the C-terminus of CIH, which corresponds to the ancient domain. This migration is due to a very high isoelectric point of the domain (pI=9.76). In

contrast, DNA in presence of the RWD domain does not show shift in the migration. This result indicates that for CIH, the RWD domain appears to be dispensable for binding DNA whereas the sole ancient domain is enough for DNA binding. RNA binding ability will be discussed below (section 3.10.2).

This function might explain why the ancient domain has been highly conserved during evolution. The particular utility of the DNA binding ability will be further investigated.

### 3.10 Identification of IMPACT proteins as DNases

Following the observation that CIH has the ability to bind DNA, we further investigated the possibility of a catalytic activity on the IMPACT family. Indeed, we actually demonstrate the presence of a DNase activity in full length IMPACT, Yih1 and CIH preparations indicated by the disappearance of the DNA band in the presence of the purified proteins (Figure 3.17). This result suggests that there is an evolutionarily conserved DNase activity within the IMPACT family. The next step was to assure that DNase activity was indeed originated from the IMPACT proteins and not a contamination product of the preparation. Due to stability and solubility reasons the characterization of the DNase activity and mutants lacking nuclease activity have been performed using CIH.



**Figure 3.17. DNase activity is present in full length Yih1, IMPACT and CIH purifications.** First line, marked as MW corresponds to the DNA ladder. Second line (-) is the DNA used as input. Lanes Y, I and C correspond with digestions performed using full length Yih1, IMPACT and CIH proteins respectively.

---

### 3.10.1 Point mutations abolish DNase activity in CIH

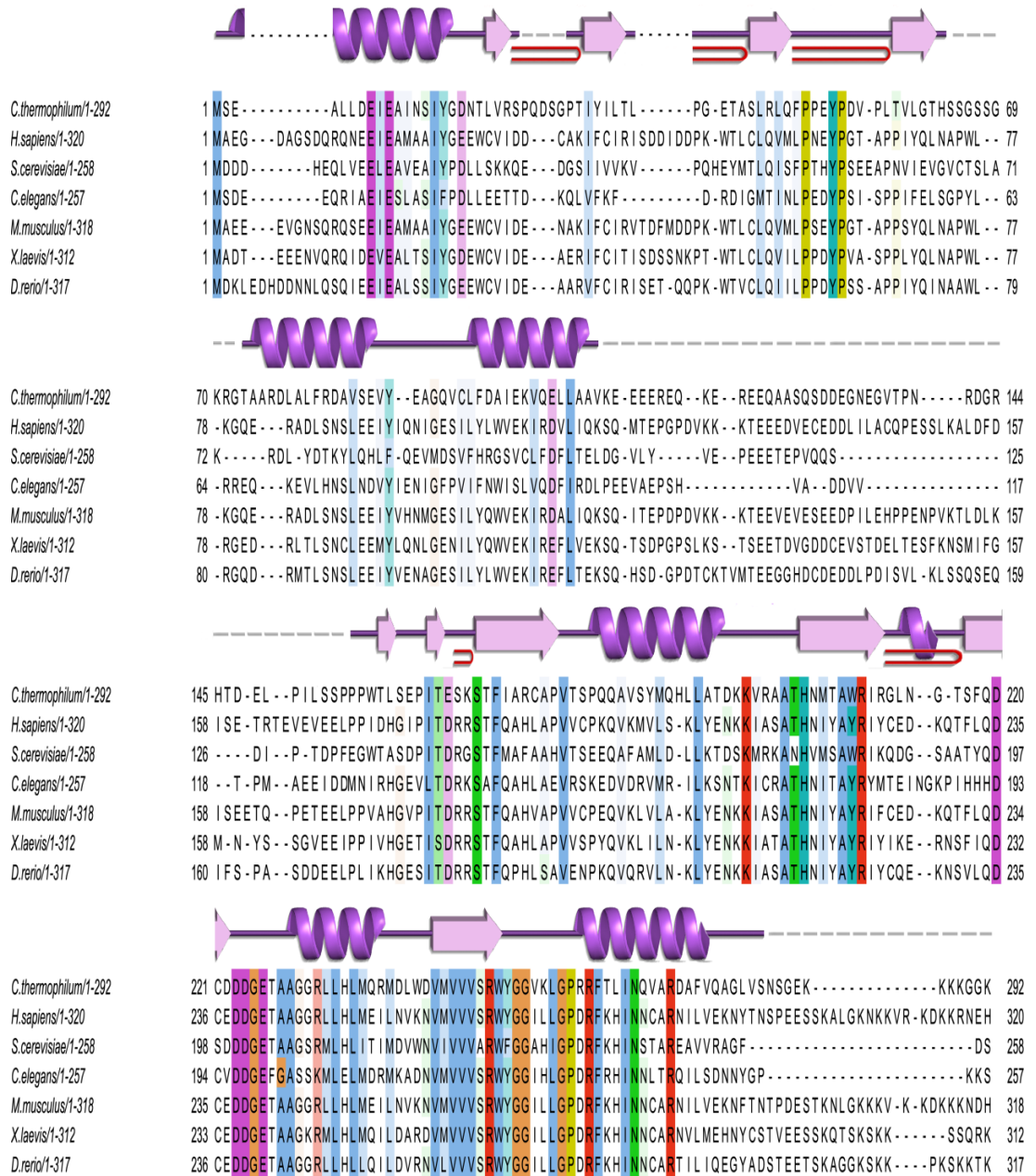
To prove that the DNase activity belongs to IMPACT proteins and it is not due to any contamination of the protein preparation, several mutants were designed. Residues were mutated regarding mainly its conservation degree.

Using an alignment of some reviewed IMPACT sequences, residues were coloured using clustalX as implemented in Jalview following sequence conservation (Figure 3.18). As shown, many residues are conserved, so it was very difficult to select among all of them. It was necessary to enclose a smaller area to look for candidates. Hence we delimited the search using the following criteria:

On one hand there was one patch positively charged at the ancient domain and it was located in a very well conserved area, as it has been shown before.

On the other hand, there are some defined sequence motifs that are found in the active site of the different nuclease families. The most wide extended is the DEDD motif, but other variants of this family such as DEDDy or DEDDh can be found, where D refers to amino acid aspartic, E is a glutamic residue, y is tyrosine and h is histidine. There are other sequence motifs as DERK, DHH or PD(D/E)XK, this last motif is also known as DEK. DEK motif represents the major restriction enzyme nuclease fold. All the motifs mentioned above can be found in the different types of nucleases, including RNases (Yang, 2011).

One variation of the motifs listed above that we were able to find in the sequence is DDDE (residues 222, 223, 224 and 226), where the motif is located in a loop next to a positively charged pocket. That pocket corresponds to a highly conserved area where D224 and E226 residues are very close to a histidine, H203, what from our point of view, was increasing the probabilities of these residues to be involved in the nuclease reaction.



**Figure 3.18. Sequence alignment of IMPACT homologs.** Structural motifs related to CIH sequence are shown above the sequence. Unknown structure has been depicted as grey bars and gaps in sequence alignment with black dots. Red loops represent  $\beta$ -hairpins in the structure. Sequence conservation has been rendered with clustalX.

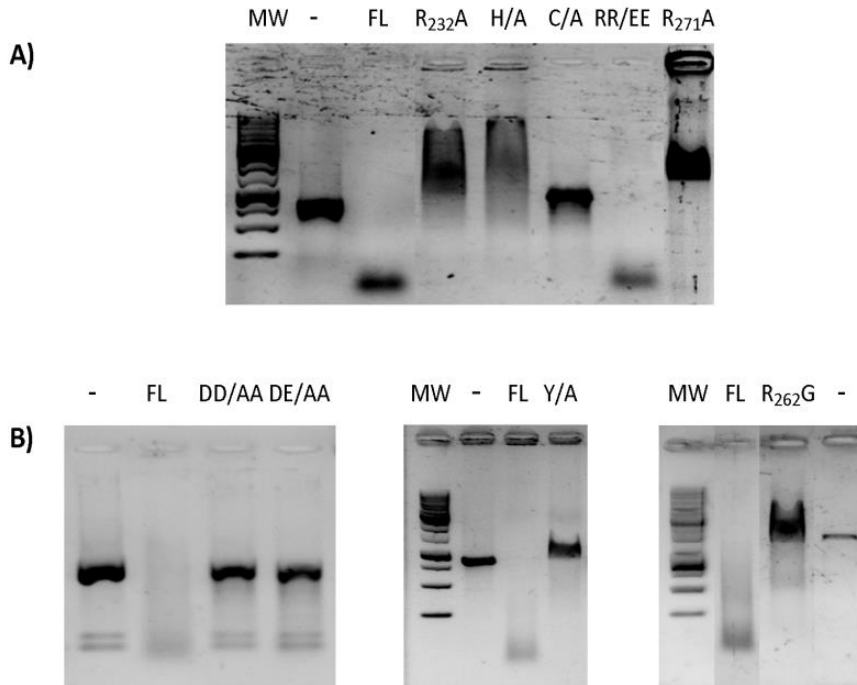
*A priori* the possibility for the RWD domain to be the catalytic domain by itself without involving the ancient domain was excluded, since RWD domain have been more extensively studied and characterised than the ancient domain and nuclease activity has never been reported. Besides, negatively charged surface is not optimal for nucleic acid binding (section 3.14) and little sequence conservation among RWD domains from IMPACT proteins was suggesting that the presence of an enzymatic activity conserved during evolution was very difficult to sustain. Moreover, this was consistent in our experiments with the fact that the RWD domain did not produce any band shift when an agarose EMSA was performed (figure 3.16). The high structure similarity of the ancient domain with the RNase PH domain showed very obviously to us that the catalytic properties of CIH were not related to the RWD domain; nevertheless, cooperation between both domains was not discarded.

Hence, with all the information available from the literature and the structure, several mutants were generated and tested (table 3.4). Some of them were still exhibiting nuclease activity, but there were mutations abolishing it, confirming that DNase activity is present in CIH proteins (Figure 3.19).

**Table 3.4. Mutations analysed for DNase activity.** One letter code residue followed by the position on the sequence and the amino acid used in the mutation are indicated.

Mutation	DNase activity?
H203A	NO
C221A	NO
D223A/D224A	NO
D223A/E226A	NO
R232A	NO
Y253A	NO
R261E/R262E	YES
R262G	REDUCED
R271A	NO



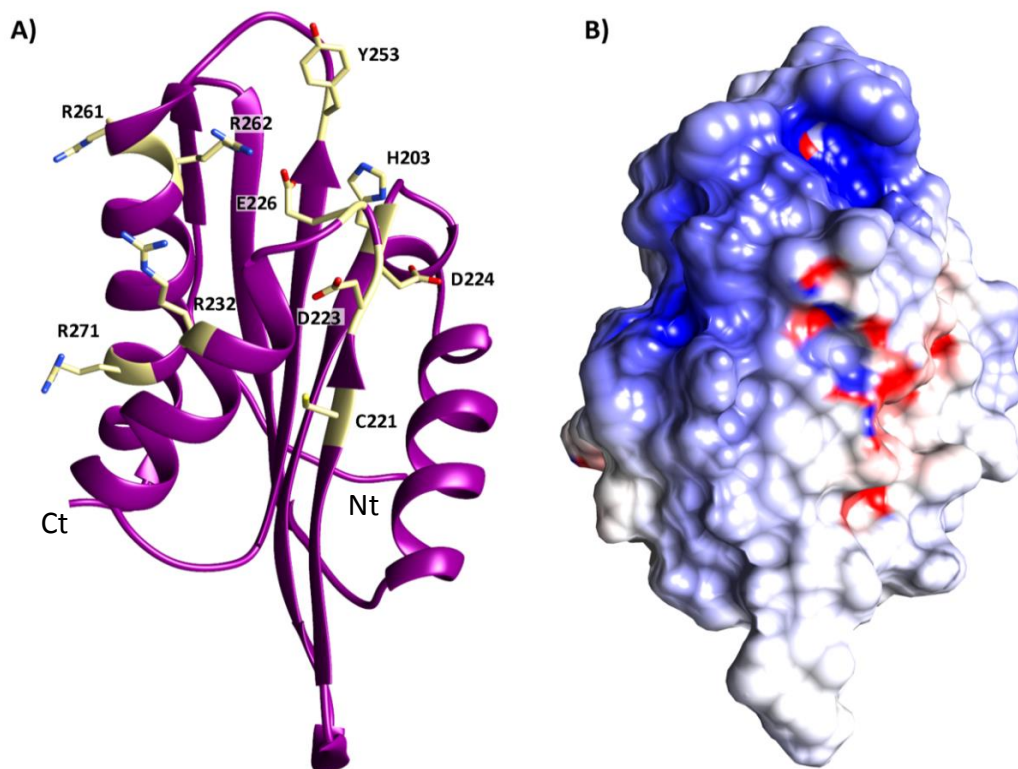


**Figure 3.19. DNase activity assays for mutants.** CIH mutants R232A, H203A (H/A), C221A (C/A) R261E/R262E (RR/EE) and R271A activities have been tested following this order (Panel A). Three different gels with activity assays for CIH mutants D223A/D224A (DD/AA), D223A/E226A (DE/AA), Y253A (Y/A) and R262G are shown. FL refers to CIH<sub>FL</sub> (positive control) and – to the input DNA.

As shown (Figure 3.19, panels A and B), several point mutants displayed no detectable DNase activity as R232A, H203A, Y253A C221A. The three first clearly display a band shift in the agarose gel, so the DNA binding affinity has not been affected by the mutation of R232, H203 or Y253. However, although the C221A mutant does not seem to degrade DNA, it does not display a clear band shift. The mutant R262G showed reduced activity when compared with the wild type protein, since a faint smear can be observed at lower molecular weights than the input DNA. The R271A mutant does not seem to display detectable activity, and the fraction of DNA precipitated in the well could be pointing to a reduced solubility of the protein that precipitated in complex with DNA. Double

mutants do not show detectable activity.

Mutated residues are located nearby or within positively charged areas (Figure 3.20). Given the electro-positive surface extension it is unlikely that a point mutation would prevent DNA binding to the ancient domain. The effect on the DNase activity might therefore be related to the role of the mutated residues in the DNA proper orientation for catalysis or directly involved in the nuclease reaction itself.



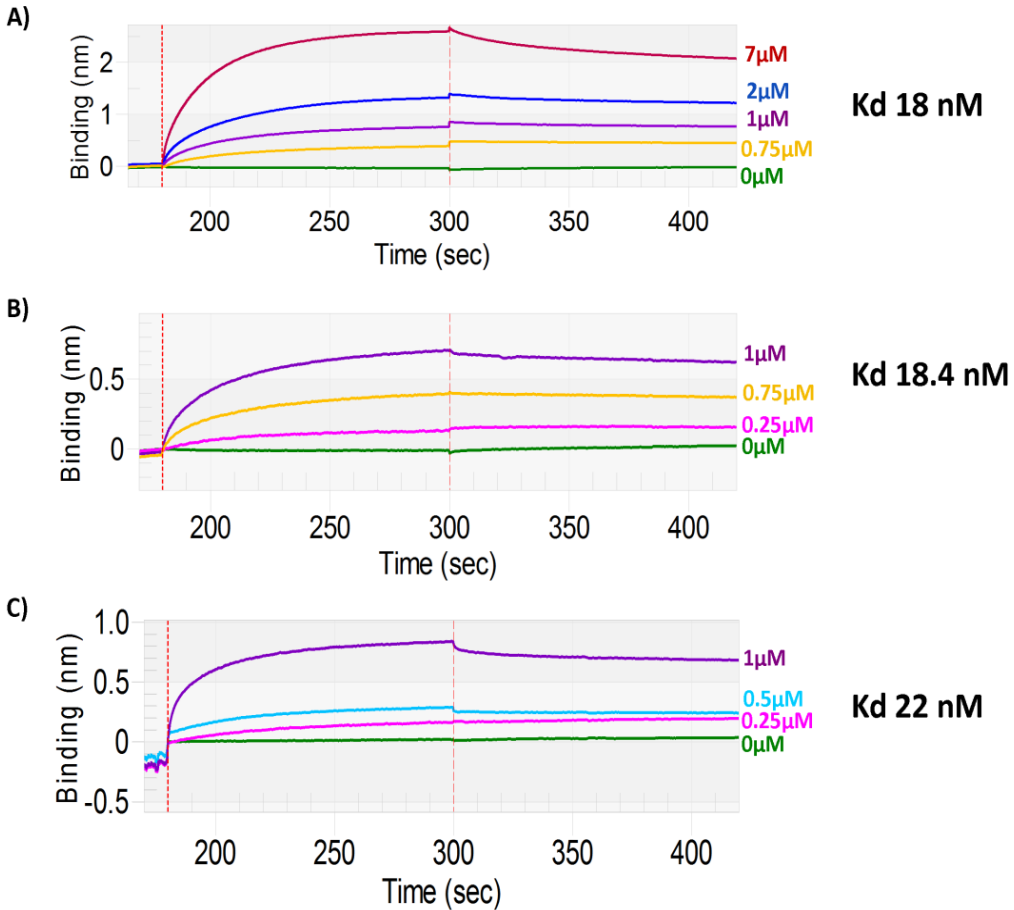
**Figure 3.20. Position of the mutations within the CIH ancient domain structure and electrostatic surface.** Ribbon representation of CIH ancient domain. Shown residues rendered in pale yellow correspond with mutated residues tested for DNase activity (Panel A). Surface representation of CIH ancient domain with the same orientation as in panel A. Electrostatic surface potential, calculated at 300K, has been rendered with red (negative) and blue (positive) colours, using contour levels in the range from -10 to 10 kT/e respectively (panel B). N-terminal and C-terminal regions are indicated with Nt and Ct respectively.

The result here provided evidences, for the first time, that IMPACT proteins have an intrinsic enzymatic activity that involves the ancient domain. Until now the only characterized function for IMPACT protein inside the cell was related to the interaction with another partner, GCN1. With no doubt this result is laying the foundations for future studies that will shed light on the role of IMPACT within the cell.

### **3.10.2 CIH binds both DNA and RNA with similar affinity**

To accurately characterise the binding affinity of CIH to nucleic acids, biolayer interferometry (BI) assays were carried out using BLitZ. The inactive mutant CIH<sub>R232A</sub> was used to measure binding affinities for three different types of nucleic acids: ssDNA, dsDNA and ssRNA. (Figure 3.21 panels A, B and C respectively).

After the performance of three independent experiments for each nucleic acid type, no significant differences have been found in the binding affinities for dsDNA, ssDNA or ssRNA. This lack of substrate specificity reveals IMPACT proteins as universal nucleic acid binding proteins. Yih1 has been found to co-sediment with polyribosomes. As it has been already mentioned, using crosslinking and pull-down techniques, it was determined that Yih1 was interacting with RPL39 and RPS22, two ribosomal proteins (Waller *et al.*, 2012) . Taking into account the new characterized ability of CIH, this result has to be handled carefully, since at the published article the authors did not eliminate RNA from the sample before the crosslinking or the pulldown, and the interaction found could be the result of finding both proteins bound to RNA.



**Figure 3.21. Representative BI experiments for CIH<sub>R232A</sub> binding affinity calculation to DNA and RNA.** Association and dissociation have been monitored by BI. Binding for dsDNA (A) ssDNA (B) and ssRNA (C) have been tested. Protein concentration tested on every curve has been indicated. Numbers on the right side of the binding curves correspond to the dissociation constant calculated from 3 independent experiments.

### 3.10.3 Co-crystallization of the ancient domain with DNA yields crystals with a phosphate bound to the structure

For further extend the understanding of CIH nucleic acid binding, we

decided to set up co-crystallization trials with the ancient domain of CIH (CIH<sub>152-292</sub>) or CIH<sub>R232A</sub>, with different lengths of dsDNA oligonucleotides. This co-crystallization was performed in the presence of MgCl<sub>2</sub>. Crystals were obtained in 18% polyethylene glycol 3500, 0.1M Bis Tris pH 5 and 0.2M sodium chloride in the presence of 10 base pairs long dsDNA (Figure 3.22)



**Figure 3.22. Crystal of the ancient domain obtained in presence of dsDNA.** Co-crystallization of Ancient domain with dsDNA (10 base pairs long) obtained in a 96 well sitting drop plate.

### 3.10.3.1 Data collection and processing

A complete dataset was collected at XALOC beamline, ALBA synchrotron. Crystals belonged to C2 space group and diffracted up to 1.4Å resolution. Data reduction was performed using SCALA (Evans, 2011) and molecular replacement was carried out using our previous model of the ancient domain (PDB code 5hcs).

Refinement was carried out using Phenix.refine (Adams *et al.*, 2010) combined with manual model building using Coot (Emsley and Cowtan, 2004; Emsley *et al.*, 2010). Anisotropic refinement was carried out at final refinement cycles. The structure of the ancient domain with a phosphate ion has been solved at 1.5Å resolution. All the refinement statistics can be found at Table 3.5. Electron density maps obtained after structure refinement showed that the DNA was not present in the crystal, however a phosphate was found to be bound to the structure. The refined model showed no Ramachandran or rotamer outliers and the PDB code assigned after deposition was 5fcv.

**Table 3.5. Data collection and refinement statistics for the ancient domain bound to a phosphate ion.** Statistics for the highest resolution shell are shown in parentheses

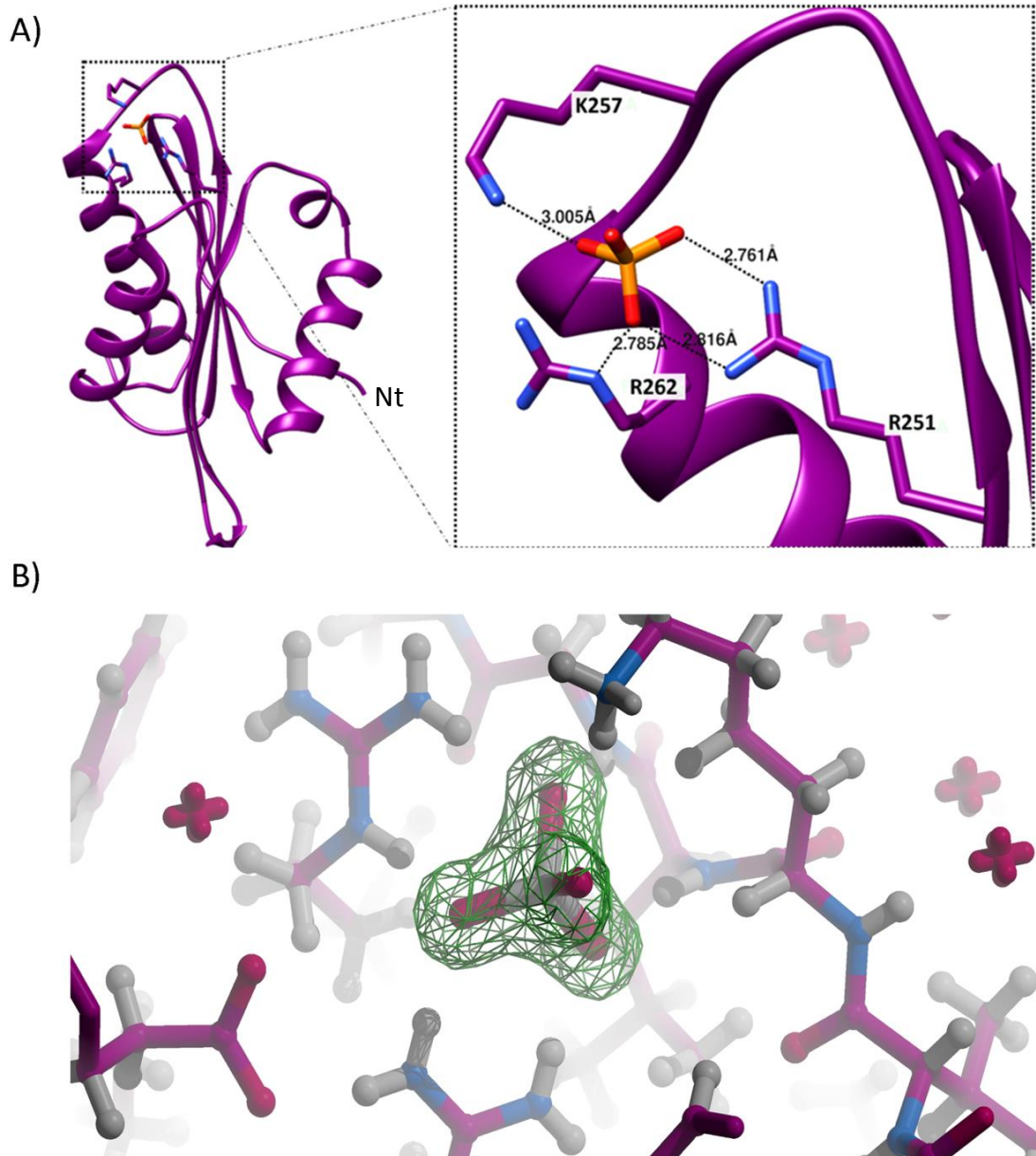
<b>Wavelength (Å)</b>	1.072
<b>Resolution range (Å)</b>	38.02 - 1.5 (1.554 - 1.5)
<b>Space group</b>	C1 2 1
<b>Unit cell</b>	81.32 37.60 51.01 90 117.41 90
<b>Total reflections</b>	80797 (8064)
<b>Unique reflections</b>	21880 (2139)
<b>Multiplicity</b>	3.7 (3.8)
<b>Completeness (%)</b>	98.90 (97.89)
<b>Mean I/sigma (I)</b>	12.19 (1.96)
<b>Wilson B-factor</b>	15.23
<b>R-merge</b>	0.066 (0.67)
<b>R-meas<sup>a</sup></b>	0.077 (0.736)
<b>CC1/2</b>	0.998 (0.778)
<b>CC*</b>	1 (0.936)
<b>R-work<sup>b</sup></b>	0.1445 (0.2356)
<b>R-free<sup>c</sup></b>	0.1720 (0.2796)
<b>Number of non-hydrogen atoms</b>	1264
<b>Macromolecules</b>	1039
<b>Ligands</b>	25
<b>Water</b>	200
<b>Protein residues</b>	130
<b>RMSD<sup>d</sup> (bonds)</b>	0.006
<b>RMSD<sup>d</sup> (angles)</b>	1.00
<b>Ramachandran favored (%)</b>	99
<b>Ramachandran outliers (%)</b>	0
<b>Clashscore</b>	0.47
<b>Average B-factor</b>	20.70
<b>Macromolecules</b>	17.70
<b>Ligands</b>	42.70
<b>Solvent</b>	33.50

<sup>a</sup>  $R_{\text{meas}} = \left\{ \sum_{hkl} [N/(N-1)]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| \right\} / \sum_{hkl} \sum_i I_i(hkl)$ , where  $I_i(hkl)$  are the observed intensities,  $\langle I(hkl) \rangle$  are the average intensities and N is the multiplicity of reflection  $hkl$ . <sup>b</sup>  $R_{\text{work}} = \sum_{hkl} \{ |F_{\text{obs}}(hkl) - F_{\text{calc}}(hkl)| \} / \sum_{hkl} [F_{\text{obs}}(hkl)]$ , where  $F_{\text{obs}}(hkl)$  and  $F_{\text{calc}}(hkl)$  are the structure factors observed and calculated, respectively. <sup>c</sup>  $R_{\text{free}}$  corresponds to Rfactor calculated using 5% of the total reflections selected randomly and excluded during refinement. <sup>d</sup> RMSD is the root mean square deviation.

### 3.10.3.2 Structural characterization

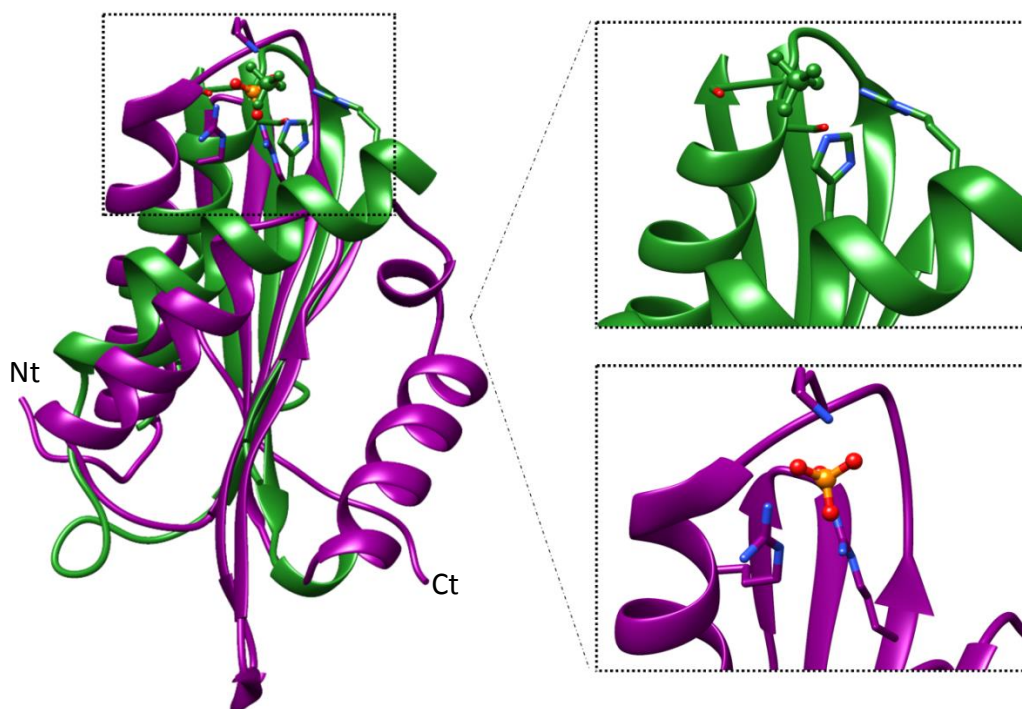
Ancient domain crystals showed no visible DNA, but a phosphate ion was found in the structure, located in the same pocket where the DDE motif and H203 are present. Forming part of this cavity there are several residues as arginine 251 (R251), arginine 262 (R262), tyrosine 253 (Y253) and lysine 257 (K257), whose side chains are facing to the inner part of the hollow. Other residues on the vicinity of the cavity are arginine 232 (R232), phenylalanine 263 (F263). No conformational changes in the domain have been observed upon phosphate binding. Phosphate ion is located on a positively charged pocket on the surface of the domain (Figure 3.23) and in the previous obtained structure (5hcs) at 2 Å; a water molecule was occupying that position. It is interacting mainly with three residues on the cavity: arginine 251, lysine 257 and arginine 262. It is worth to note that some point mutations of the residues located in the vicinity of the phosphate (Y253, R262, H203 and R232) abolish DNase activity. In particular, R262 establishes direct interaction with the phosphate ion.

Phosphate binding area is highly conserved among homologs. When the RNase PH domain from 4aid is superimposed with CIH, both phosphates occupy similar positions (Figure 3.24). This phosphate in RNase PH domain is bound at the active site of the molecule. The fact that the structure of the domain is similar to a domain present in phosphorolytic enzymes (RNase PH and PNPase) and that the phosphate binds in an equivalent area might be suggesting that it also has a phosphorolytic mechanism. In any case, we have trapped a substrate related phosphate, evidencing the similarity of the area with the active site of the PNPase or RNase PH.



**Figure 3.23. Phosphate ion bound to the ancient domain.** Ribbon representation of the ancient domain. Residues involved in phosphate contacts are shown. Interactions are marked with a discontinuous black line and distances are indicated for each interaction (Panel A). Model inside  $F_o - F_c$  map (green), calculated in the absence of the phosphate ion is contoured at  $6\sigma$  (Panel B). N-terminal region is indicated with Nt.

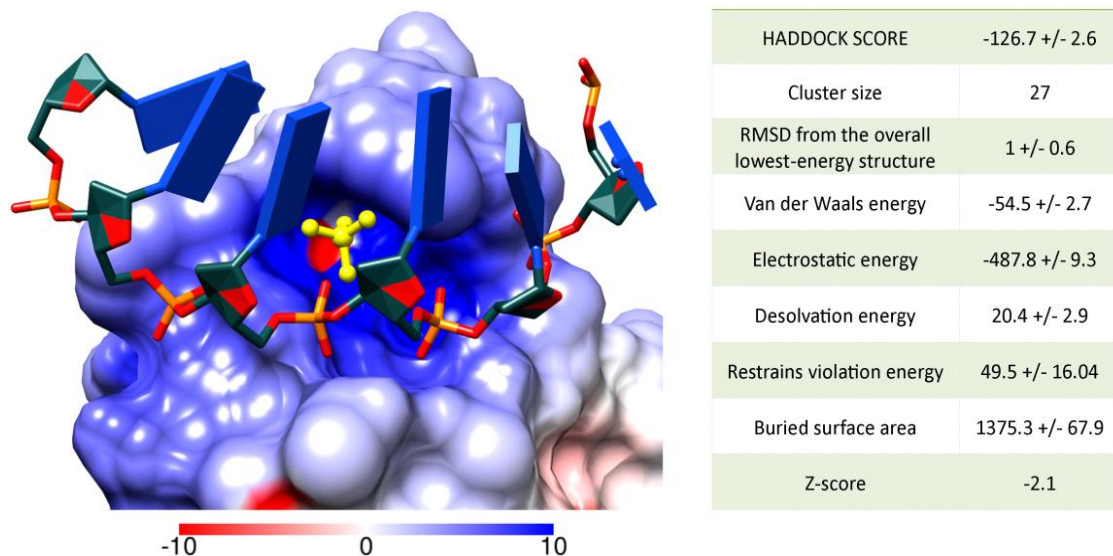




**Figure 3.24. Ribbon representation of CIH ancient domain (5fcv) and RNase PH domain from a polyribonucleotide phosphorylase (4aid) superposition.** RNase PH domain and its bound phosphate have been depicted in green, whereas the ancient domain has been depicted in purple. Phosphate bound to the ancient domain has been rendered in red and gold. On the right side of the picture, zoomed phosphate binding areas and the residues involved in phosphate coordination are shown. N-terminal and C-terminal regions are indicated with Nt and Ct respectively.

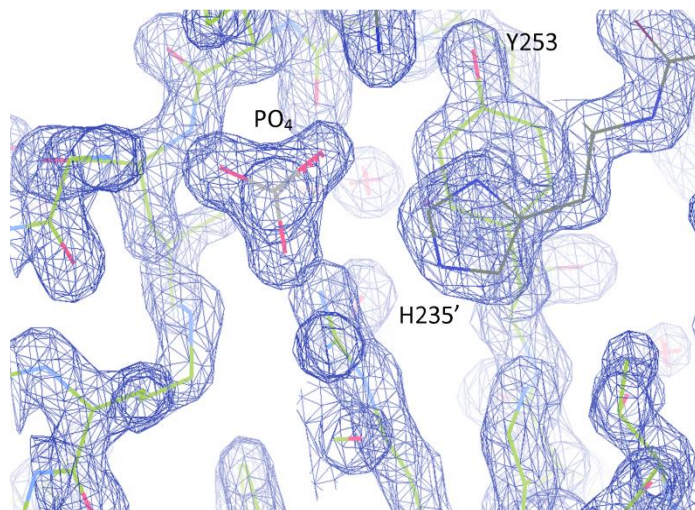
The phosphate present in the structure could be produced by DNA degradation inside the drop or it could be from cellular origin during overexpression, since no phosphate containing buffers have been used during the purification. HADDOCK docking assays performed with ssDNA and the ancient domain confirmed that, the positively charged groove where the phosphate is present in the structure is compatible with the binding of the nucleic acid (Figure 3.25). Besides, the point that mainly the phosphate backbone from the nucleic

acid is establishing the interaction with the surface would be consistent with a no sequence specific DNA binding.



**Figure 3.25. ssDNA docking over CIH ancient domain.** The phosphate found in the structure has been depicted in yellow with a ball and stick representation. Six nucleotides long ssDNA has been coloured using the heteroatom colour pattern from chimera. Docking parameters for the selected cluster are shown on the right. Electrostatic surface potential, calculated at 300K, has been rendered using red (negative) and blue (positive) colours. Contour levels are in the range from -10 to 10 kT/e respectively.

When looking closer to the surrounding phosphate binding area, we find a histidine from a neighbouring molecule very close to the phosphate. The combination on the histidine ring and the phosphate might be mimicking the position of a nitrogenous base from a potential nucleic acid substrate (Figure 3.26). The tyrosine that is located behind the histidine 235 from the symmetric molecule (named H235' at the figure) could be involved in the positioning of the nucleic acid in the groove due to stacking interactions with the DNA bases.



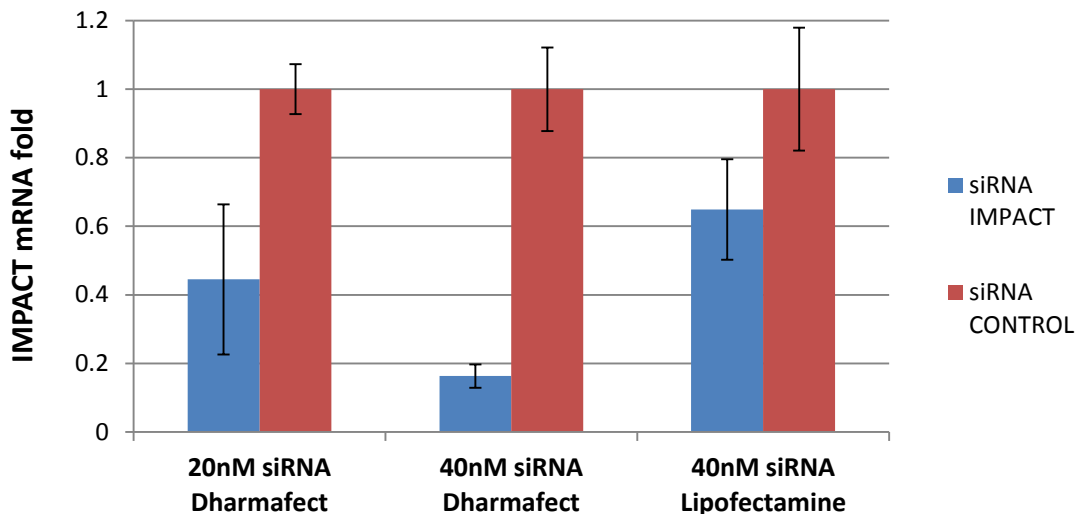
**Figure 3.26. Histidine 235 from a symmetric molecule locates nearby the phosphate ion.** The symmetric molecule and the phosphate have been rendered with grey backbone. Residues from the main molecule are coloured in light green. The histidine 235 from the symmetric molecule is indicated in the figure as H235'. The phosphate ion and residue Y253 are also indicated.

### 3.11 IMPACT proteins are cytosolic nucleic acid sensors

Cytoplasm is a compartment of the cell that is usually free of DNA. The presence of self or foreign DNA in the cytosol, usually as consequence of an infection or tissue damage, triggers a response that involves the production of type I interferons and the proteolytic activation of caspases (Hornung and Latz, 2010). There are many types of proteins that are able to sense and signal the presence of cytosolic DNA or dsRNA (Keating *et al.*, 2011) Some nucleic acid cytosolic sensors are nucleases as TREX1 (Atianand and Fitzgerald, 2013; Gehrke *et al.*, 2013) . In that sense, since we have demonstrated that IMPACT proteins bind with the same affinity DNA and RNA, we wanted to test whether IMPACT proteins were related to innate immune response inside the cell being able to signal or not the presence of such nucleic acids.

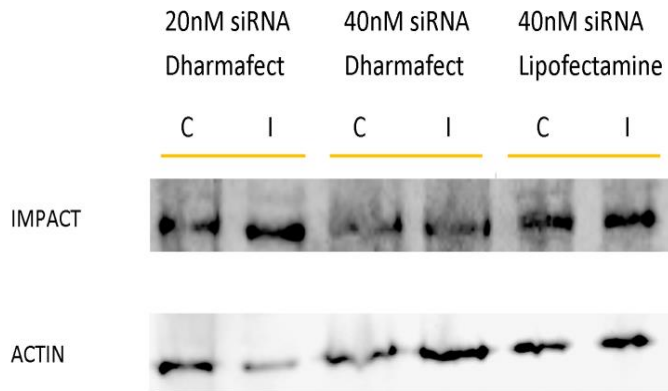
An *ex vivo* assay with 3T3 cells and small interference (siRNA) was performed in order to optimize the concentration of siRNA necessary for silencing the *Impact* gene. The siRNA used for IMPACT silencing was a commercial

pool composed by 4 different siRNAs. Normalization of mRNA levels obtained by real-time PCR was carried out using actin mRNA (Figure 3.27).



**Figure 3.27. IMPACT silencing with siRNA in 3T3 cells.** Silencing after transfection with dharmafect was tested for 20nM and 40nM of siRNA. For lipofectamine only 40nM siRNA was tested. Blue bars indicate IMPACT mRNA fold levels after transfection with siRNA IMPACT normalized against the IMPACT mRNA levels obtained with siRNA control (red bars).

The very poor quality of the only commercially available antibody for murine IMPACT protein was making very difficult to achieve a good western blot where the silencing was obvious. Regarding the loading control (actin), for the case of 40nM siRNA transfected with dharmafect, the amount of the protein loaded for the siRNA IMPACT sample is approximately three times the amount loaded for siRNA CONTROL whereas the IMPACT protein detected by the antibody is similar in both cases. We can say that we have obtained a decreased expression of the protein when compared with the control (Figure 3.28). Transfection of 20nM IMPACT siRNA with dharmafect or 40nM IMPACT siRNA with lipofectamine did not silenced IMPACT.

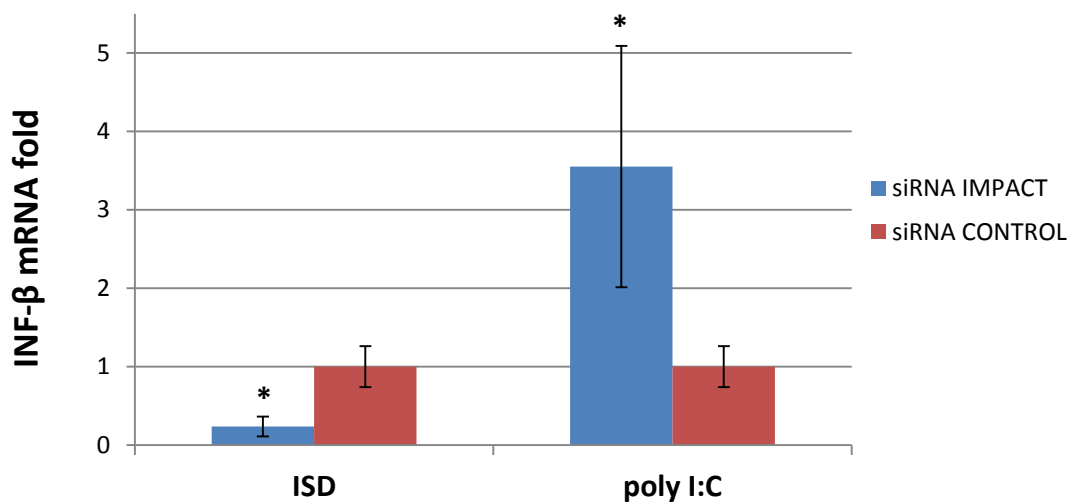


**Figure 3.28. IMPACT protein levels analysis after silencing.** siRNA concentration and reactive used for transfection are indicated above each pair of samples. Lines labelled as C indicate samples from cells transfected with siRNA CONTROL and lines I correspond to cell extracts treated with siRNA IMPACT.

Once the interference conditions were established, we proceeded to test whether IMPACT proteins were related to the innate immune system by analysing the correlation of INF- $\beta$  levels with IMPACT. Fibroblasts are known to produce mainly INF- $\beta$  as response to an infection. ISD and poly I:C have been widely used to induce interferon production in these cells. ISD, from interferon stimulatory DNA, is a dsDNA non CpG oligonucleotide 45 base pairs long from the genome of *Listeria monocytogenes*, and poly I:C is a dsRNA analogue. We have measured mRNA levels for INF- $\beta$  after transfection of ISD or poly I:C in control cells and cells with the Impact gene partially silenced by siRNA. Surprisingly, two different responses were found depending on the nature of the nucleic acid co-transfected with the siRNA. When ISD and IMPACT siRNA were co-transfected, mRNA levels for INF- $\beta$  appeared to decrease approximately 5 times in comparison with cells transfected with ISD and non-targeting siRNA (figure 3.29). On the other hand, when IMPACT gene is silenced and there is poly I:C in the cytoplasm, mRNA levels for INF- $\beta$  increased more than 3 times in comparison with control cells (Figure 3.29). Normalization of mRNA levels was also carried out using actin mRNA.

These results are revealing that IMPACT proteins have effects on INF- $\beta$

levels and suggest a function related to innate immune response. Regarding its ability to unspecifically bind nucleic acids, the obtained responses of the treatment with ISD or poly I:C in knock down or wild type cells are pointing IMPACT proteins as cytosolic nucleic acid sensors. The specific mechanism used by IMPACT proteins to trigger these responses remains unclear.



**Figure 3.29. Relative amounts of  $\beta$ -interferon mRNA levels after IMPACT silencing and ISD or poly I:C treatment.** INF- $\beta$  mRNA levels normalized against control cell mRNA levels (red bars) and IMPACT knock down (blue bars) cells after ISD (n=3) or poly I: C (n=3) transfection are shown. Asterisk,  $P < 0.015$  when compared with siRNA CONTROL.

As it has been mentioned in the introduction (1.6.4), a relationship between the immune suppressive effects of Indoleamine 2,3-dioxygenase (IDO) dependent tryptophan depletion and the over expression of IMPACT exists. Tryptophan depletion by IDO activates the GCN2 kinase pathway that results in translation inhibition, proliferative arrest and finally apoptosis (Lee *et al.*, 2002). Cells, like fibroblasts, that display IDO induced IMPACT overexpression are able to overcome tryptophan depletion and to avoid apoptosis. However, some immune

cells like Jurkat or B cells do not survive GCN2 kinase pathway activation since they show very low or non-detectable levels of IMPACT and therefore fail to overcome tryptophan depletion (Habibi *et al.*, 2010; Poormasjedi-Meibod *et al.*, 2013).

IDO activity is not the only immunoregulatory mechanism available and although it has many functions in the immune system, not all of them are clearly beneficial to the host. It has been demonstrated that tryptophan depletion induced by IDO can inhibit pathogen replication *in vitro* in the case of auxotroph organisms and some virus (Adams *et al.*, 2004; Bodaghi *et al.*, 1999; Gupta *et al.*, 1994; Pfefferkorn, 1984). However, the biological relevance *in vivo* for infection control is not well understood. Besides the ability of control over some infections, IDO induced expression leads to the suppression of T-cell response by the adaptive immune system either by the production of toxic metabolites or by tryptophan depletion. With the results obtained, we can hypothesize that, despite interferon- $\gamma$  induced overexpression of IDO, and subsequent suppression of T-cell response, compensating effects might occur in IMPACT expressing cells. Over expression of IMPACT could degrade DNA inside the cytosol and/or rapidly signal the infection to neighbouring cells by INF- $\beta$  production. Besides, interferon  $\beta$  is known to increase survival of B and T cells and decreases the threshold for the activation of B cells (Braun *et al.*, 2002), fact that could somehow compensate death by tryptophan depletion. Hence, IMPACT overexpression could be helping not only to overcome tryptophan depletion but also to have the cell prepared for detecting, signalling and fighting the infection. However, this proposed model would be valid only for infections involving DNA from the pathogen, since the observed response for the dsRNA analogue is the opposite.

In relation to the observed response obtained with poly I:C, since IMPACT proteins are involved in translation and can be found in polyribosomes, it is reasonable to think that IMPACT would be silencing immune response against

---

RNA molecules. Due to its role in translation control it will be in contact with different types of RNA molecules and it would be reasonable to think that IMPACT proteins would not be involved in immune response against cytosolic RNA. However, it has been described that cells are able to discriminate between self and non self RNA (Keating et al., 2011), so no immune response from self RNA would be expected to be triggered by IMPACT proteins. Although we have not been able to unequivocally confirm IMPACT RNase activity, it cannot be discarded. IMPACT RNase activity could provide an explanation for the response obtained in these assays, since some cytosolic exoribonucleases that suppress interferon production against some virus, like Trex-1 does with HIV, have been reported (Yan *et al.*, 2010). Nevertheless the reason why IMPACT is silencing immune response to foreign RNA is not obvious and more experiments would need to be performed in order to understand it.

These results confirm that the IMPACT protein behave as cytosolic nucleic acid sensor in murine fibroblasts, although the mechanisms they use to trigger or silence innate immune response, depending on the nature of the nucleic acid, remain unknown. The response pattern in other cell types may not be conserved, since cytoplasmic nucleic acid sensors have been demonstrated to be in many cases cell type specific. On the other side, if the nucleic acid sensor role inside the cell is related with its nuclease activity remains an open question.

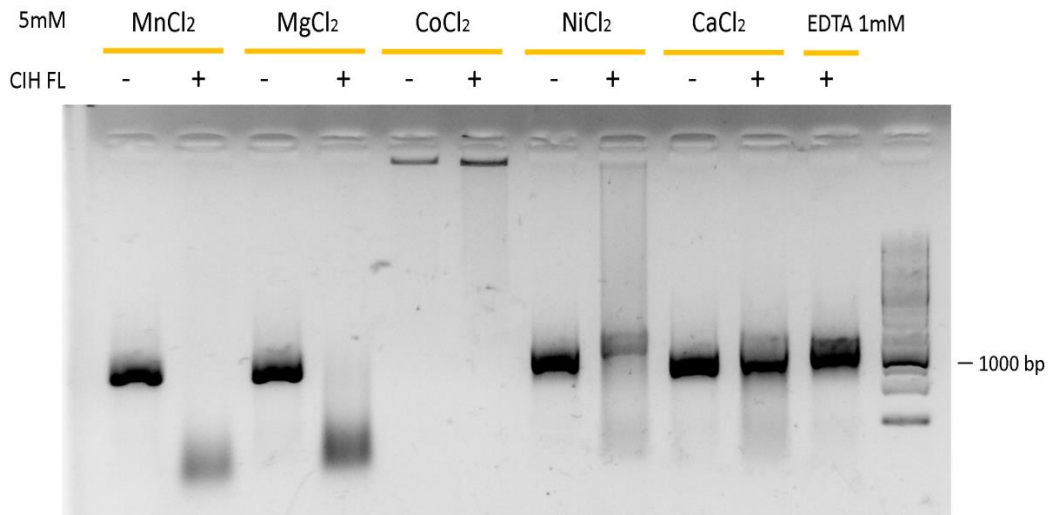
### **3.12 DNase activity characterization**

Nucleases can be classified in many ways, depending on the parameter we are looking at. Type of cleavage, metal ion activity dependence and sequence motifs are characteristics frequently used for their classification.



### 3.12.1 DNase activity in CIH requires $Mg^{2+}$ or $Mn^{2+}$

The first question to answer was if there was any specific metal requirement. DNase assay in presence of different metal cations revealed that either magnesium or manganese were essential for the reaction to occur (Figure 3.30). This result is not surprising since most nucleases show requirement for magnesium or manganese. In contrast, calcium which is necessary for DNase I activity and stability (Guérout *et al.*, 2010) does not seem to be required for CIH DNase activity, and with Nickel the protein displays very little DNase activity. In the particular case of cobalt, it seems to precipitate the DNA. CIH in presence of 1mM EDTA and no metal did not show DNase activity, reinforcing the idea of an ion assisted hydrolytic mechanism.

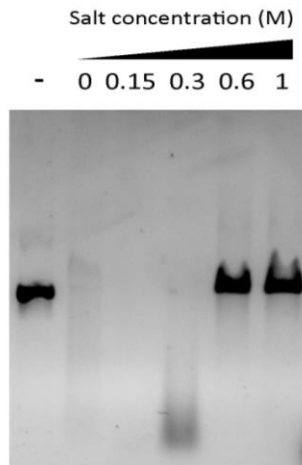


**Figure 3.30. CIH requires magnesium or manganese for its DNase activity.** 1% Agarose gel showing DNase activity on 100ng of 1100bp long dsDNA. Five different reactions have been performed using different salts in order to assess metal requirements. Type of salt or EDTA are indicated. CIH<sub>FL</sub> presence (+) or absence (-) in the sample has been marked. Enzymatic activity is monitored by disappearance of the control band (lane -) after incubation of the reaction mix during 40 minutes at 40°C.

Magnesium is an ion highly available inside the cell (between 2 and 3mM free  $Mg^{2+}$  available). Reviews about nuclease activities and their requirement point that the suitability of  $Mg^{2+}$  is due to its small ionic radius, which makes  $Mg^{2+}$  in some cases easy to incorporate inside the protein. But  $Mg^{2+}$  has high tendency to hydration, being most of the times coordinated by 6 or 4 water molecules, so its ionic radius increases drastically (Cowan, 2002; Yang, 2011). Manganese, the other suitable nucleophile for CIH, is usually presenting a coordination that involves 4 water molecules. This fact is leading us to the idea that the area binding the nucleophile necessary for the reaction must be an accessible groove large enough to accommodate such coordination.

### 3.12.2 DNase activity is disrupted with high ionic strength

To shed light on DNase optimum conditions, buffers with different NaCl concentration, and therefore different ionic strength were tested.



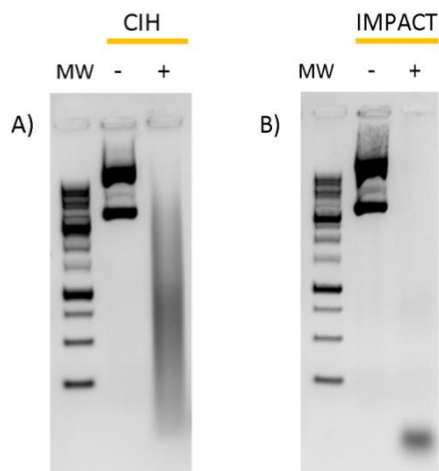
**Figure 3.31. DNase activity with different ionic strength.** 1% Agarose gel showing DNase activity on 100ng of 800bp long dsDNA. The first line corresponds to the dsDNA input (-). The same amount of CIH<sub>FL</sub> (15 $\mu$ g) was added to all the samples except the input. Numbers above each line refer to NaCl concentration present in the reaction buffer. Enzymatic activity is monitored by disappearance of the control band (lane 1) after incubation of the reaction mix during 40 minutes at 40°C.

As shown (Figure 3.31) high salt concentrations are avoiding DNA hydrolysis. A range of different salt concentrations between 0 and 300 mM showed nuclease activity. These results are not surprising since those concentrations are somehow in the physiological range, confirming that DNase

activity would be feasible inside the cell. High ionic strength is probably disrupting the interaction between the DNA substrate and the protein.

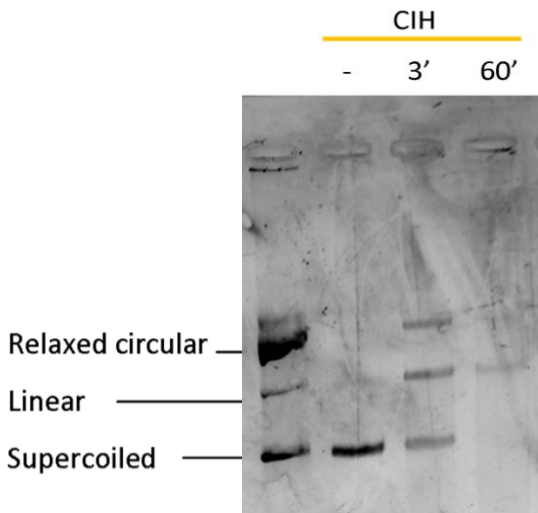
### 3.12.3 IMPACT proteins are endonucleases

It was necessary to clarify whether IMPACT proteins are endonucleases or exonucleases. Previous experiments related to DNase activity of IMPACT proteins were always performed with linear dsDNA. In order to figure out if the IMPACT family of proteins was showing endonuclease activity, a DNase treatment with covalently closed circular DNA (cccDNA) was carried out to test if CIH and IMPACT were able to degrade it (Figure 3.32). The obtained result clearly shows that IMPACT proteins are endonucleases.



**Figure 3.32. CIH and IMPACT show endonuclease activity.** 1% Agarose gel showing DNase activity on 150ng of plasmidic DNA. CIH is able to degrade cccDNA (Panel A). The same test was performed with human IMPACT protein, and the endonucleolytic activity is found to be conserved (Panel B). Enzymatic activity is monitored by disappearance of the control bands (lane 2) after incubation of the reaction mix during 40 minutes at 30 and 40°C for IMPACT and CIH respectively.

Taking advantage of pBR322 supercoiled form, an assay to elucidate how CIH cuts dsDNA was performed. Since supercoiled, relaxed circular and linear DNA present different migration properties on agarose gel, simultaneous CIH cleaving of both chains or sequential single chain cleavage could be monitored (Figure 3.33). This assay is giving us precious information necessary to figure out what is the catalytic mechanism of the cleavage.



**Figure 3.33. pBR322 digestion.** 1% Agarose gel showing DNase activity on of cccDNA. First line shows the 3 different conformation of the pBR322 plasmid. Gel purified supercoiled isoform used as input is shown at line 2. Lines 3 and 4 show digestion products after 3 and 60 minutes respectively performed at 40°C. Line 2 (-) was incubated for 60 minutes in reaction buffer with no protein at 37 °C.

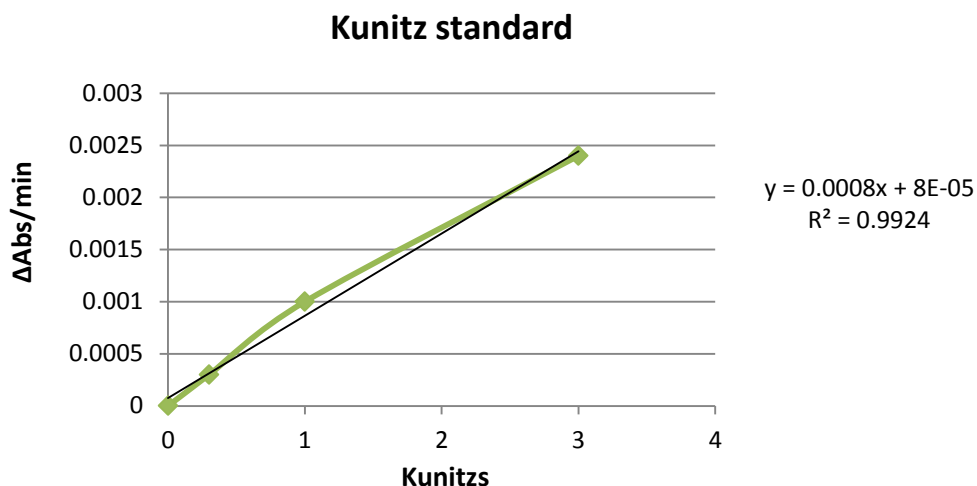
The ability to cut both strands simultaneously in the same site or to produce single strand cuts is a criterion used to classify endonucleases. The appearance of two upper bands after 3 minutes of incubation with purified supercoiled DNA (cccDNA) is clearly revealing that the cut at both strands is sequential and not simultaneous

### 3.12.4 Kunitz assay

Although nowadays there are different methods to detect DNase activity, hyperchromicity assay is the most widely used. One Kunitz assay is defined as the amount of enzyme necessary to produce an increase of the absorbance at 260nm of 0.001 in one minute using a 1mg/mL of salmon sperm DNA solution in sodium acetate pH 5 buffer and performed at 25°C.

In order to calculate how many Kunitz were present in one milligram of CIH, we first performed a standard curve with known DNaseI Kunits (Figure 3.34). Two different concentrations of CIH were measured. For 122µg of CIH<sub>FL</sub> the obtained ΔAbs/min is 0.0003 and for 244µg ΔAbs/min is 0.0006. When we extrapolate the ΔAbs/min and calculate the value for 1mg of CIH, the obtained

value was 2.68 Kunitz/mg. This is an extraordinary low Kunitz value, since one can find in the literature values ranging from 56000 Kunitz/mg for DNase I to 660 Kunitz/mg LS-DNase (Baron *et al.*, 1998).

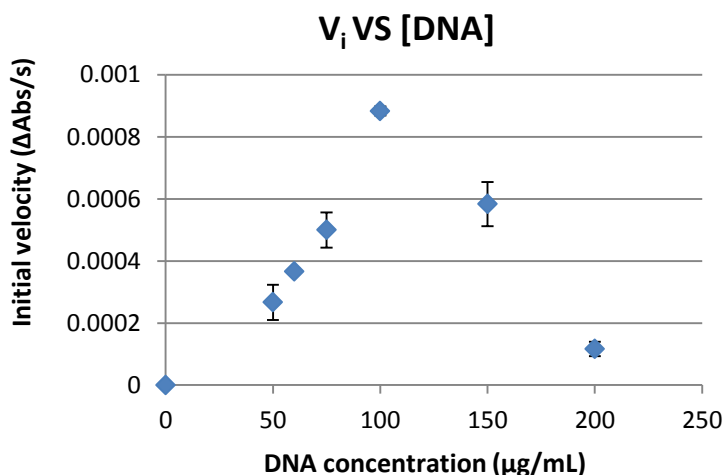


**Figure 3.34. Standard curve for Kunitz calculation.** The  $\Delta\text{Abs}/\text{min}$  has been monitored for each DNase I Kunitz measurement. The equation used to extrapolate  $\Delta\text{Abs}/\text{min}$  into Kunitz is indicated. Enzymatic reaction was followed by increase of absorbance at 260 nm.

It could be that the low Kunitz obtained might be related to the relative low temperature that was carried out since *C. termophilum* is a thermophilic organism. Other possibility is that the pH of the reaction was not the optimum for the protein, since theoretical pI for the protein is 5.05 and protein solubility is reduced when the pH is similar to the pI of the protein. In fact, small precipitation was observed when the protein was added to the cuvette to start the reaction. This obtained result further encouraged us to characterise the particularities of the enzyme.

### 3.12.5 CIH DNase activity displays substrate inhibition

Initial velocities,  $\Delta\text{Absorbance}/\text{second}$  ( $\Delta\text{Abs}/\text{sec}$ ), for the reaction were measured with different substrate concentrations. The units had to remain in absorbance units because the size of the obtained product is not exactly known. When represented, the obtained data (Figure 3.35) clearly showed a decrease of the initial velocity when DNA concentration was above  $100\mu\text{g}/\text{mL}$ .



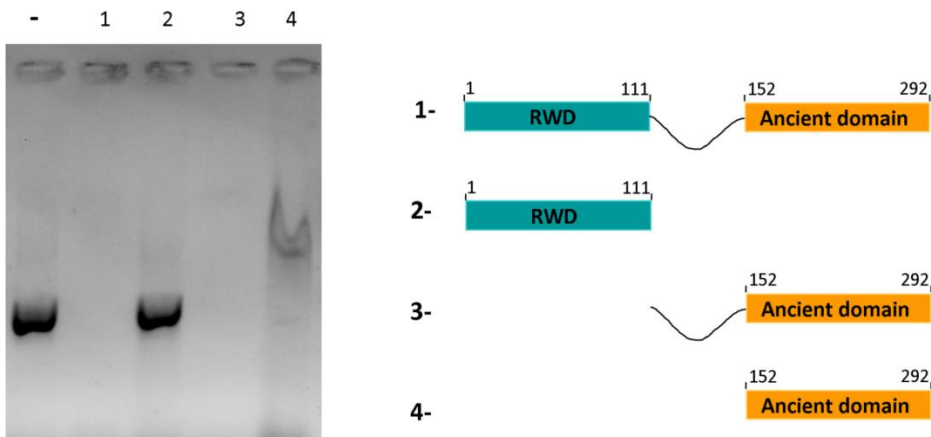
**Figure 3.35. Substrate inhibition assay.** Average initial velocity of the reaction from two independent experiments (represented as  $\Delta$  Absorbance at 260nm/second) has been represented for each DNA concentration tested. Standard deviations for each value are represented as black bars.

Among the concentrations tested,  $100\mu\text{g}/\text{mL}$  showed optimum values whereas increasing the concentration to  $200\mu\text{g}/\text{mL}$  was decreasing the initial velocity under values obtained with  $50\mu\text{g}/\text{mL}$ . This pattern is consistent with an enzymatic substrate inhibition profile, and gives an explanation to the very low Kunitz value obtained, since the assay is carried out at a DNA concentration of  $1\text{mg}/\text{mL}$ .

### 3.12.6 The RWD domain is dispensable for the DNase activity

Although the RWD domain did not show the ability of binding DNA, the cooperation between both domains or the implication of the linker in the nuclease activity cannot be discarded.

In order to figure out what is the role of each domain on the enzymatic activity, a DNase assay was carried out using different truncations of the protein (Figure 3.36).



**Figure 3.36. DNase assay with different CIH domains.** The first line (-) corresponds with the linear dsDNA input used for the assay. A scheme of the domains used for each assay (1 to 4) is represented on the right side of the figure. After 1 hour of reaction, samples were boiled before were run in 1% agarose gel to monitor DNA degradation.

As it is shown, the elimination of the RWD from the full length protein is not affecting the DNase activity. However, the sole ancient domain appears to have a strong decrease on its activity. Since removal of the linker region seems to affect DNase activity, it must be somehow implicated in the reaction, either in the nucleophile binding or in the proper positioning of the DNA for the reaction.

### 3.13 Crystallization of the RWD domain from CIH

Since crystallization trials with CIH<sub>FL</sub> were giving crystals only after protein degradation, we decided to set up crystallization screenings with the N-terminal domain of CIH (RWD domain). Large single crystals (Figure 3.37) were obtained after the first set of screenings using 47mg/mL of pure protein (CIH<sub>1-111</sub>) and using the vapour diffusion technique at 21°C. Crystals appeared in 1.26 M Sodium phosphate monobasic monohydrate, 0.14 M Potassium phosphate dibasic, pH 5.6.



**Figure 3.37.** Crystal of RWD domain from *C.thermophilum* IMPACT homolog. Crystal grown in 96-well sitting drop plate at 21°C.

#### 3.13.1 CIH RWD domain data processing and refinement statistics

RWD crystals were diffracting up to 1.37Å. Crystals belonged to P2<sub>1</sub> 2<sub>1</sub> 2<sub>1</sub> space group. In order to solve the phase problem, and despite there was not any available model in the protein data bank with high sequence identity to CIH RWD domain, we tried molecular replacement with the RWD domain of murine GCN2 (Nameki *et al.*, 2004) structure that was solved by Nuclear Magnetic Resonance (NMR) and was available at the PDB (1UKX). Unfortunately, this molecular replacement did not give any confident result.

Since the dataset of the diffracting crystal was collected at 1.37Å with a low resolution limit of 37.94 Å, the dataset was suitable for Ab initio phasing, that was carried out in collaboration with Dr. Isabel Usón IBMB-CSIC, Barcelona using Arcimboldo software (Sammuto *et al.*, 2014).



**Table 3.6. Data collection and refinement statistics for RWD domain (Sedo).** Statistics for the highest resolution shell are shown in parentheses

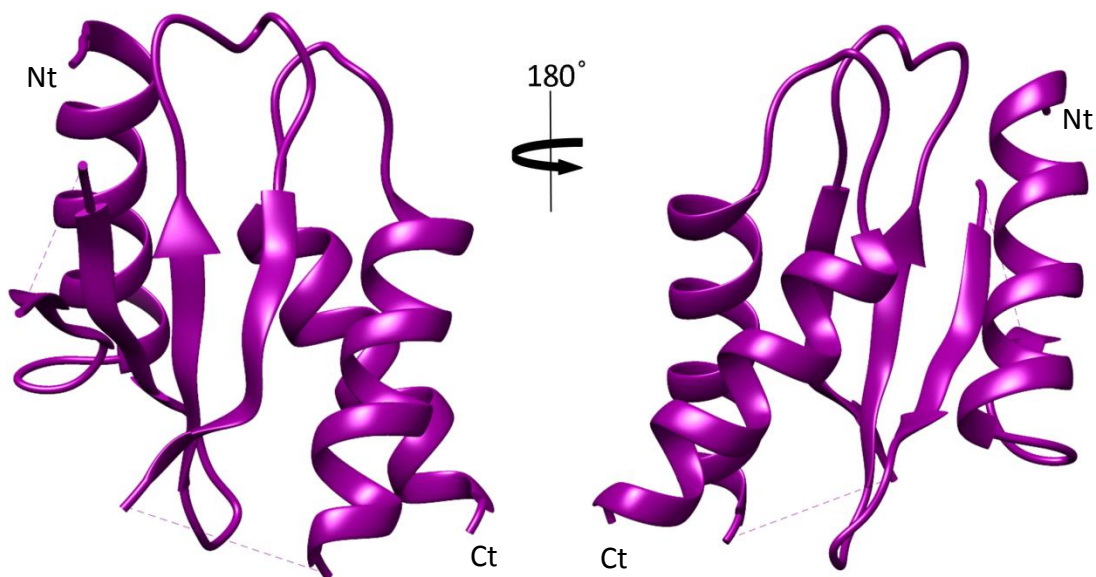
<b>Wavelength (Å)</b>	0.9795
<b>Resolution range (Å)</b>	35.77 - 1.4 (1.45 - 1.4)
<b>Space group</b>	P 21 21 21
<b>Unit cell</b>	37.89 48.36 53.15 90 117.41 90
<b>Total reflections</b>	278273 (27299)
<b>Unique reflections</b>	19815 (1937)
<b>Multiplicity</b>	14 (14.1)
<b>Completeness (%)</b>	99.96 (99.95)
<b>Mean I/sigma (I)</b>	28.57 (5.19)
<b>Wilson B-factor</b>	18.86
<b>R-merge</b>	0.049 (0.54)
<b>R-meas<sup>a</sup></b>	0.051 (0.506)
<b>CC1/2</b>	0.999 (0.964)
<b>CC*</b>	1 (0.991)
<b>R-work<sup>b</sup></b>	0.1674 (0.1598)
<b>R-free<sup>c</sup></b>	0.2078 (0.2251)
<b>Number of non-hydrogen atoms</b>	791
<b>Macromolecules</b>	744
<b>Water</b>	47
<b>Protein residues</b>	97
<b>RMSD<sup>d</sup> (bonds)</b>	0.008
<b>RMSD<sup>d</sup> (angles)</b>	1.17
<b>Ramachandran favored (%)</b>	100
<b>Ramachandran outliers (%)</b>	0
<b>Clashscore</b>	0
<b>Average B-factor</b>	30.80
<b>Macromolecules</b>	30.70
<b>Solvent</b>	32.30

<sup>a</sup>  $R_{\text{meas}} = \left\{ \sum_{hkl} [N/(N-1)]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| \right\} / \sum_{hkl} \sum_i I_i(hkl)$ , where  $I_i(hkl)$  are the observed intensities,  $\langle I(hkl) \rangle$  are the average intensities and  $N$  is the multiplicity of reflection  $hkl$ . <sup>b</sup>  $R_{\text{work}} = \sum_{hkl} \{ [F_{\text{obs}}(hkl)] - [F_{\text{calc}}(hkl)] \} / \sum_{hkl} [F_{\text{obs}}(hkl)]$ , where  $F_{\text{obs}}(hkl)$  and  $F_{\text{calc}}(hkl)$  are the structure factors observed and calculated, respectively. <sup>c</sup>  $R_{\text{free}}$  corresponds to Rfactor calculated using 5% of the total reflections selected randomly and excluded during refinement. <sup>d</sup> RMSD is the root mean square deviation.

The obtained density showed one molecule per asymmetric unit. Refinement was carried out until  $R_{\text{WORK}}$  and  $R_{\text{FREE}}$  parameters were 16.74 and 20.78 respectively. Refinement statistics are shown in Table 3.6. PDB code assigned 5edo.

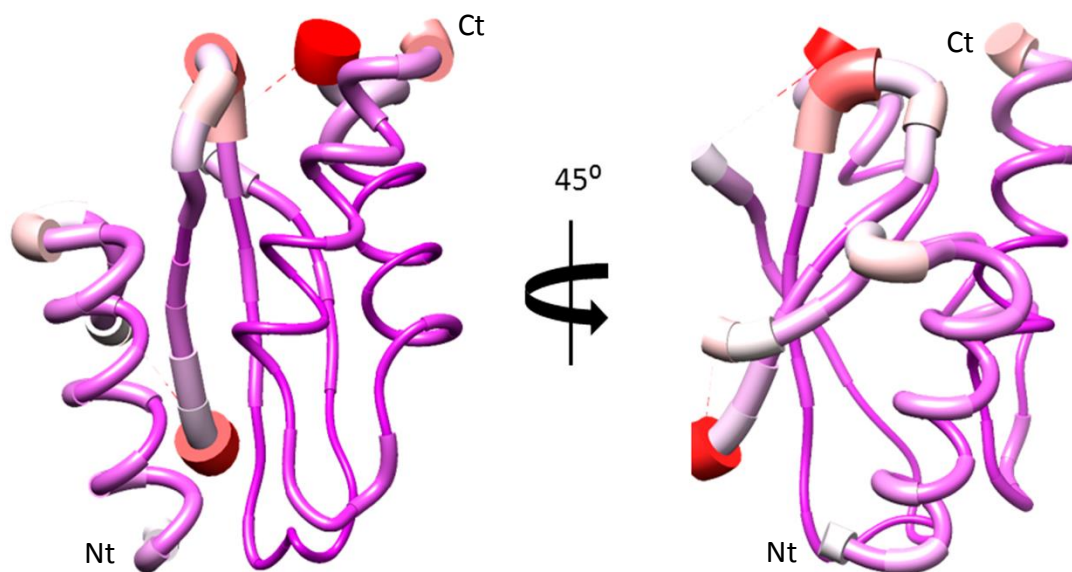
### 3.14 CIH RWD structural characterization

The structure of the RWD CIH domain (residues 1 to 111) exhibits a topology  $\alpha\beta\beta\beta\alpha$ , where the four anti-parallel  $\beta$ -strands are composing a  $\beta$ -sheet that has one  $\alpha$  helix in one side and the other two at the other side (Figure 3.38). The helix  $\alpha_3$  contacts with  $\alpha_1$ , forming a globular structure. One side of the  $\beta$ -sheet stays hidden by the three  $\alpha$  helices, thus any possible interactions would not involve this area unless a drastic conformational change takes place.



**Figure 3.38. Ribbon representation of the CIH RWD domain.** Non visible parts of the model have been represented as a discontinuous line. N-terminal and C-terminal regions are indicated with Nt and Ct respectively.

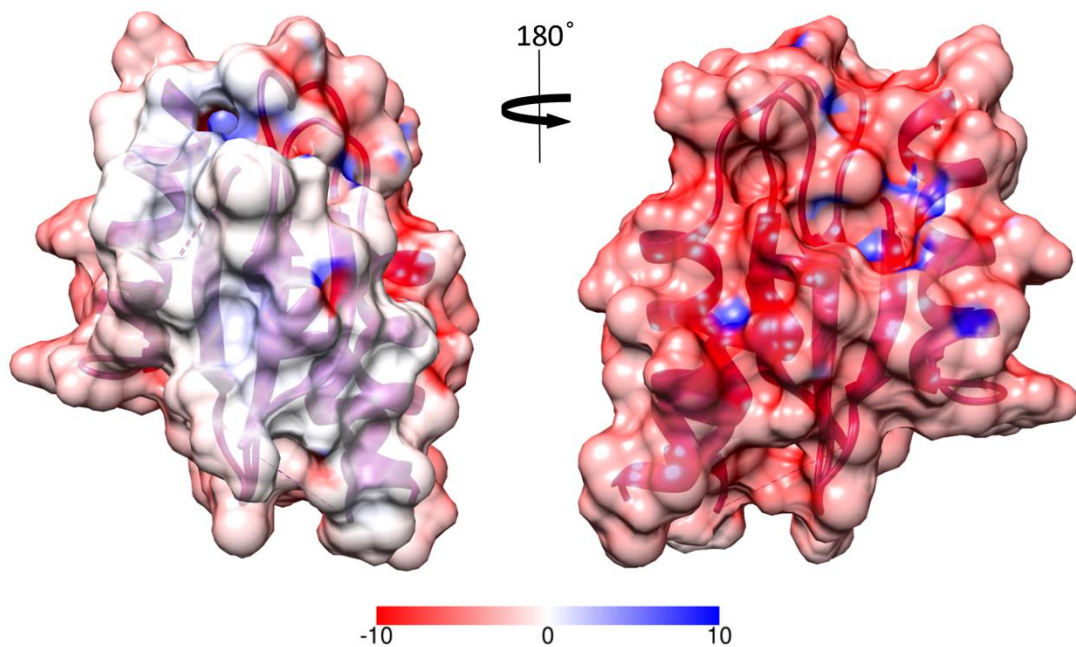
Analysis of temperature factors in the structure show very low values except for some residues very close to non-visible loops and for the  $\beta$ -turn connecting  $\beta$ -strand 2 with the  $\beta$ -strand 3, which suggest that a non-induced conformational change involving the movement of the  $\alpha$ -helices is not likely to take place (Figure 3.39).



**Figure 3.39. Temperature factors of CIH RWD backbone in worm representation.** Radii and colour are proportional to the average temperature factor of each residue. Residues with high temperature factors are depicted in red and residues with low temperature factors are rendered in purple. Radii is increased for residues with high average temperature factors. . N-terminal and C-terminal regions are indicated with Nt and Ct respectively.

The electrostatic surface potential shows two different distributions. There is a negatively charged area corresponding to the surface of the alpha helices, and a neutral area with small positive patches distributed all over the beta-sheet (Figure 3.40). The negatively charged area corresponds to the helices that have

been previously described as essential for GCN1 binding. Thus, this finding suggests that GCN1-IMPACT interaction might have a strong electrostatic component.



**Figure 3.40. CIH RWD domain electrostatic surface potential representation.** RWD surface coloured depending on its electronic potential. The electronic potential has been rendered in red and blue for negatively and positively charged areas respectively. White colour shows areas with neutral charge. Contour levels, indicated in the colour key, are in range from -10 to 10 kT/e and have been calculated at 300K. Both views are rotated 180° between them.

### 3.14.1 Comparison between RWD domains from GCN2 and CIH

Sequence alignment of RWD domains from GCN2 and IMPACT reveals low residue conservation, displaying 13.7% of identity between murine RWD domain from GCN2 and RWD domain from *C.thermophilum* sequences (calculated using SIAS tool: <http://imed.med.ucm.es/Tools/sias.html>). Alignment between RWD domain from GCN2 and IMPACT from different species shows that conserved

residues are grouped towards both terminal ends of the domain (Figure 3.41).

```

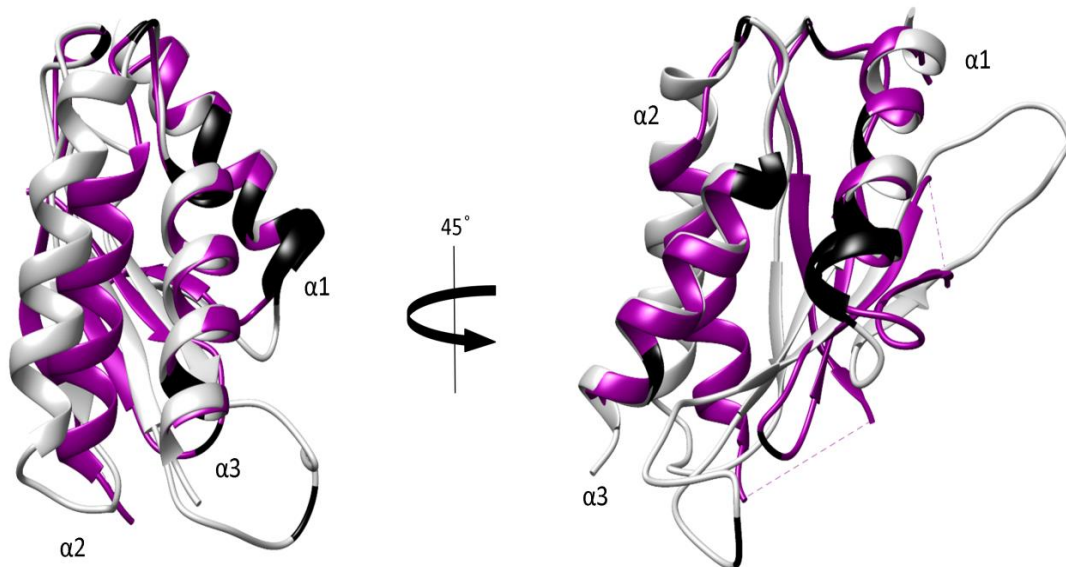
CHAETOMIUM_IMPACT/1-111  1  -----MS E A L D E I E A I N S I Y G D N T L V R S P Q D S G -----P T I Y I L T L P G E T A S L 44
HUMAN_IMPACT/1-116      1  -----M A E G D A G S D Q R Q N E E I E A M A A I Y G E E W C V I D D C A K -----I F C I R I S D D I D D P K W T 51
MOUSE_IMPACT/1-116     1  -----M A E E E V G N S Q R Q S E E I E A M A A I Y G E E W C V I D E N A K -----I F C I R V T D F M D D P K W T 51
YIH1/1-114             1  -----M D D D H E Q L V E E L E A V E A I Y P D - - L L S K K Q E D G -----S I I V V K V P Q H E Y M T L 45
HUMAN_GCN2/1-137       1  M A G G R G A P G R G R D E P P E S Y P Q R Q D H E L Q A L E A I Y G A D F Q D L R P D A C G P V K E P P E I N L V L Y P Q G L T G E E V Y V 71
MOUSE_GCN2/1-137       1  M A G G R G A S G R G R A E P Q E S Y S Q R Q D H E L Q A L E A I Y G S D F Q D L R P D A R G R V R E P P E I N L V L Y P Q G L A G E E V Y V 71
YEAST_GCN2/1-128       1  -----M S L S H L T L D Q Y Y E I Q C N E L E A I R S I Y M D D F T D L T K R K S S W D K Q P - Q I I F E I T L R S V D K E P V E S 62

CHAETOMIUM_IMPACT/1-111 45  R L Q F P P E Y P D V P L - T V L G T H S G S S G K R G T A A R D L A - - L F R D A V S E V Y E A G Q V C L F D A I E K V Q E L L A A V K 111
HUMAN_IMPACT/1-116     52  - L C L Q V M L P N E Y P G T A P P I Y Q L N A P W L K G Q E R A D L S N S L E E I Y I Q N - - - I G E S I L Y L W V E K I R D V L I Q K - 116
MOUSE_IMPACT/1-116     52  - L C L Q V M L P S E Y P G T A P P S Y Q L N A P W L K G Q E R A D L S N S L E E I Y V H N - - - M G E S I L Y Q W V E K I R D A L I Q K - 116
YIH1/1-114            46  Q I S F P T H Y P S E E A P N V I E V G V C T S L A K R D L Y D T K Y L Q H L F Q E V M D S V F H R G S V C L F D F L T E L D G V L Y V E - 114
HUMAN_GCN2/1-137       72  K V D L R V K C P P T Y P D V V P E I E L K N A K G L S N E S V N L L K S R L E E L A K K H - - - C G E V M I F E L A H V Q S F L S E H - 137
MOUSE_GCN2/1-137       72  Q V E L Q V K C P P T Y P D V V P E I E L K N A K G L S N E S V N L L K S H L E E L A K K Q - - - C G E V M I F E L A H H V Q S F L S E H - 137
YEAST_GCN2/1-128       63  S I T L H F A M T P M Y P Y T A P E I E F K N V Q N V M D S Q L Q M L K S E F K K I H N T S - - - R G Q E I I F E I T S F T Q E K L D E F - 128

```

**Figure 3.41. RWD protein sequence alignment.** Sequence for RWD domains of IMPACT proteins from *C.terminophilum*, *H.sapiens* and *M.musculus* are aligned with RWD domain sequence from GCN2 of *H.sapiens*, *M.musculus* and *S.cerevisiae*. Residue conservation has been coloured with ClustalX. Protein and source organism are indicated for each sequence.

However, this low conservation degree does not prevent similar overall structures for RWD domains, as it is observed when RWD-GCN2 alpha carbons from mouse and RWD-CIH structures are superimposed (Figure 3.41). This result is showing that the folding can be conserved despite the poor conservation degree of the sequence. Structural alignment between RWD domains from GCN2 and CIH reveal that some of the conserved residues might be related to structure conservation since they are involved in contacts between two helices (Figure 3.42).



**Figure 3.42. Structural superimposition between RWD domains from CIH and GCN2 proteins.** Two different views of RWD domain from CIH (5edo) and murine GCN2 (1ukx) depicted in purple and light grey respectively. Identical residues in both RWD domains are coloured in black in both structures.

High structural similarity observed between both models makes difficult to understand why the molecular replacement did not yield a better result. Prior to molecular replacement all unstructured regions were removed from the model. The change in the angle of helix 2 is the most significant difference and might be responsible for the lack of solution in the molecular replacement. It is also worth mentioning in this sense that NMR coordinates in general do not usually provide good starting models for molecular replacement phasing.

As commented in the introduction (1.5.1.1), helices 2 and 3 of the Yih1 RWD domain have been proposed as direct interactors with GCN1 and actin since Asp90 and Glu87 mutations to alanine increased Yih1 binding affinity for both proteins and Asp102 and Glu106 are required for in vivo binding (Sattlegger *et al.*, 2011). However, conservation of these specific residues is very poor among both

IMPACT and GCN2 proteins (Figure 3.41). Although glutamic 106 in Yih1 is conserved among IMPACT proteins it is not conserved among GCN2 proteins. Aspartic 102 is not conserved in GCN2 or IMPACT proteins and glutamic 87 is the residue displaying higher conservation degree. It is worth to note that residue arginine 2259 in GCN1, is essential for the binding of both Yih1 and GCN2, but it has been pointed that GCN1 binding affinity for Yih1 is much lower than GCN1 binding affinity for GCN2.

A reason for the conservation of Yih1 glutamic acids 106 and 87 among IMPACT sequences could be a GCN1 non-related function, but the truth is that although it has been demonstrated that residues E87 and E106 affect to Yih1 cdc28 binding and that this binding is GCN1 independent, the same residues affect the binding of Yih1 to GCN1 or actin.

Mutation on Yih1 E87 and D90 to alanine increased the binding affinity for GCN1 and actin. Those residues, E87 and D90, that apparently decrease the binding affinity for Yih1 interactors, might be important for the regulation of Yih1 activity in translation control.

### 3.15 Concluding remarks

*Chaetomium thermophilum* has been essential in this research work. Its higher stability has allowed the success of the structural characterization. Structural information has been essential to extend the functional knowledge of IMPACT proteins. This work is a clear example about how a given function is going to be always related to the three dimensional structure of the protein, and even more important, how structural information can help to discover totally unexpected functional characteristics of a given protein.

We have discovered IMPACT proteins as DNases, and although ribonuclease activity has not been confirmed it has not been discarded either since CIH is able to bind with similar affinity dsDNA, ssDNA and ssRNA. The

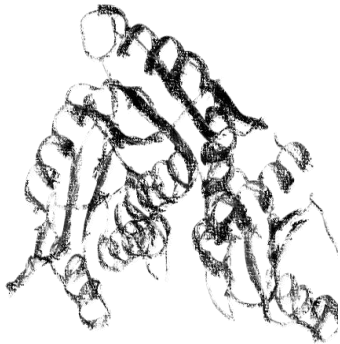
---

DNase activity has been shown to be metal dependent and to display substrate inhibition. Besides, the distortion of INF- $\beta$  mRNA levels in cells with the IMPACT expression silenced is suggesting another function for the IMPACT family. IMPACT proteins behave as cytosolic nucleic acid sensors that affect immune response and the different response for DNA or RNA might be related to other binding proteins or to its DNase activity.

On the other side, structural characterization of RWD domain shows high degree of structural conformation related to very poor sequence conservation and remarks the electrostatic nature of the RWD – GCN1 interaction.

This work lays the foundations for future detailed studies related to the role and importance of IMPACT proteins within the cell and reasserts the importance of the structural information for a better understanding of protein functions.





## 4. CONCLUSIONS



1. CIH is a universal nucleic acid binding protein. It binds DNA and RNA with similar affinity and no sequence specificity.
2. The ancient domain has certain structural similarity with RNase PH domain.
3. The ancient domain binds a phosphate ion in an equivalent area of the RNase PH domain and that has been described as its active site.
4. IMPACT proteins are enzymes. In particular, IMPACT proteins are DNases with endolytic activity that display a nucleophile assisted mechanism where magnesium or manganese are essential.
5. The cleavage produced on dsDNA is sequential.
6. CIH displays a very low Kunitz value given by the conditions of the assay.
7. CIH DNase activity displays substrate inhibition with an optimal DNA concentration (among tested) of 100 $\mu$ g/mL. On the other side, the activity gets disrupted with salt concentration above 600mM.
8. Different CIH mutants as H203A, R232A, Y253A, C221A, R271A, D223A/D224A o D223A/E226A show not detectable nuclease activity against DNA.
9. IMPACT proteins behave as cytosolic nucleic acid sensors in mouse fibroblasts, affecting INF- $\beta$  mRNA levels in response to the presence of dsDNA or dsRNA analogue in the cytosol.
10. RWD domain is dispensable for the DNase activity.
11. RWD domain maintains a conserved three dimensional structure despite

---

the low sequence homology.

12. Regarding the RWD surface electrostatic potential at the area described as necessary for GCN1 binding, the interaction between the CIH RWD domain with GCN1 might have a strong electrostatic nature.

## **5. BIBLIOGRAPHY**

---



- Abastado, J.P., Miller, P.F., Jackson, B.M., Hinnebusch, A.G., 1991. Suppression of ribosomal reinitiation at upstream open reading frames in amino acid-starved cells forms the basis for GCN4 translational control. *Molecular and cellular biology* 11, 486–96.
- Adams, O., Besken, K., Oberdorfer, C., MacKenzie, C.R., Takikawa, O., Daubener, W., 2004. Role of Indoleamine-2,3-Dioxygenase in Alpha/Beta and Gamma Interferon-Mediated Antiviral Effects against Herpes Simplex Virus Infections. *Journal of Virology* 78, 2632–2636. doi:10.1128/JVI.78.5.2632-2636.2004
- Adams, P.D., Afonine, P. V, Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., Zwart, P.H., 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography* 66, 213–21. doi:10.1107/S0907444909052925
- Agrawal, N., Dasaradhi, P.V.N., Mohmmed, A., Malhotra, P., Bhatnagar, R.K., Mukherjee, S.K., 2003. RNA Interference: Biology, Mechanism, and Applications. *Microbiology and Molecular Biology Reviews* 67, 657–685. doi:10.1128/MMBR.67.4.657-685.2003
- Alberghina, L., Rossi, R.L., Querin, L., Wanke, V., Vanoni, M., 2004. A cell sizer network involving Cln3 and Far1 controls entrance into S phase in the mitotic cycle of budding yeast. *The Journal of cell biology* 167, 433–43. doi:10.1083/jcb.200405102
- Angelastro, J.M., Ignatova, T.N., Kukekov, V.G., Steindler, D.A., Stengren, G.B., Mendelsohn, C., Greene, L.A., 2003. Regulated expression of ATF5 is required for the progression of neural progenitor cells to neurons. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 23, 4590–600.
- Atianand, M.K., Fitzgerald, K.A., 2013. Molecular basis of DNA recognition in the immune system. *Journal of immunology (Baltimore, Md. : 1950)* 190, 1911–8. doi:10.4049/jimmunol.1203162

- 
- Baneyx, F., 1999. Recombinant protein expression in *Escherichia coli*. *Current opinion in biotechnology* 10, 411–21.
- Baron, W.F., Pan, C.Q., Spencer, S.A., Ryan, A.M., Lazarus, R.A., Baker, K.P., 1998. Cloning and characterization of an actin-resistant DNase I-like endonuclease secreted by macrophages. *Gene* 215, 291–301.
- Battye, T.G.G., Kontogiannis, L., Johnson, O., Powell, H.R., Leslie, A.G.W., 2011. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta crystallographica. Section D, Biological crystallography* 67, 271–81. doi:10.1107/S0907444910048675
- Bloom, J., Cross, F.R., 2007. Multiple levels of cyclin specificity in cell-cycle control. *Nature reviews. Molecular cell biology* 8, 149–60. doi:10.1038/nrm2105
- Bodaghi, B., Goureau, O., Zipeto, D., Laurent, L., Virelizier, J.L., Michelson, S., 1999. Role of IFN-gamma-induced indoleamine 2,3 dioxygenase and inducible nitric oxide synthase in the replication of human cytomegalovirus in retinal pigment epithelial cells. *Journal of immunology (Baltimore, Md. : 1950)* 162, 957–64.
- Boehringer, D., O'Farrell, H.C., Rife, J.P., Ban, N., 2012. Structural insights into methyltransferase KsgA function in 30S ribosomal subunit biogenesis. *The Journal of biological chemistry* 287, 10453–9. doi:10.1074/jbc.M111.318121
- Boivin, S., Kozak, S., Meijers, R., 2013. Optimization of protein purification and characterization using Thermofluor screens. *Protein expression and purification* 91, 192–206. doi:10.1016/j.pep.2013.08.002
- Braun, D., Caramalho, I., Demengeot, J., 2002. IFN-alpha/beta enhances BCR-dependent B cell responses. *International immunology* 14, 411–9.
- Brünger, A.T., 1992. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* 355, 472–475. doi:10.1038/355472a0
- Carzaniga, T., Mazzantini, E., Nardini, M., Regonesi, M.E., Greco, C., Briani, F., De Gioia, L., Dehò, G., Tortora, P., 2014. A conserved loop in polynucleotide



- phosphorylase (PNPase) essential for both RNA and ADP/phosphate binding. *Biochimie* 97, 49–59. doi:10.1016/j.biochi.2013.09.018
- Chadderton, T., Wilson, C., Bewick, M., Glück, S., 1997. Evaluation of three rapid RNA extraction reagents: relevance for use in RT-PCR's and measurement of low level gene expression in clinical samples. *Cellular and molecular biology (Noisy-le-Grand, France)* 43, 1227–34.
- Chayen, N.E., Saridakis, E., 2008. Protein crystallization: from purified protein to diffraction-quality crystal. *Nature methods* 5, 147–53. doi:10.1038/nmeth.f.203
- Chen, R., Zou, Y., Mao, D., Sun, D., Gao, G., Shi, J., Liu, X., Zhu, C., Yang, M., Ye, W., Hao, Q., Li, R., Yu, L., 2014. The general amino acid control pathway regulates mTOR and autophagy during serum/glutamine starvation. *The Journal of cell biology* 206, 173–82. doi:10.1083/jcb.201403009
- Corpet, F., 1988. Multiple sequence alignment with hierarchical clustering. *Nucleic acids research* 16, 10881–90.
- Cowan, J.A., 2002. Structural and catalytic chemistry of magnesium-dependent enzymes. *Biometals : an international journal on the role of metal ions in biology, biochemistry, and medicine* 15, 225–35.
- CRICK, F.H., 1958. On protein synthesis. *Symposia of the Society for Experimental Biology* 12, 138–63.
- De Vries, S.J., van Dijk, M., Bonvin, A.M.J.J., 2010. The HADDOCK web server for data-driven biomolecular docking. *Nature protocols* 5, 883–97. doi:10.1038/nprot.2010.32
- Deutscher, M.P., Li, Z., 2001. Exoribonucleases and their multiple roles in RNA metabolism. *Progress in nucleic acid research and molecular biology* 66, 67–105.
- Donnelly, N., Gorman, A.M., Gupta, S., Samali, A., 2013. The eIF2 $\alpha$  kinases: their structures and functions. *Cellular and Molecular Life Sciences* 70, 3493–3511. doi:10.1007/s00018-012-1252-6

- 
- Emsley, P., Cowtan, K., 2004. Coot: model-building tools for molecular graphics. *Acta crystallographica. Section D, Biological crystallography* 60, 2126–32. doi:10.1107/S0907444904019158
- Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography* 66, 486–501. doi:10.1107/S0907444910007493
- Ericsson, U.B., Hallberg, B.M., Detitta, G.T., Dekker, N., Nordlund, P., 2006. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Analytical biochemistry* 357, 289–98. doi:10.1016/j.ab.2006.07.027
- Evans, P.R., 2011. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta crystallographica. Section D, Biological crystallography* 67, 282–92. doi:10.1107/S090744491003982X
- Evans, P.R., Murshudov, G.N., 2013. How good are my data and what is the resolution? *Acta crystallographica. Section D, Biological crystallography* 69, 1204–14. doi:10.1107/S0907444913000061
- Fallarino, F., Grohmann, U., You, S., McGrath, B.C., Cavener, D.R., Vacca, C., Orabona, C., Bianchi, R., Belladonna, M.L., Volpi, C., Santamaria, P., Fioretti, M.C., Puccetti, P., 2006. The combined effects of tryptophan starvation and tryptophan catabolites down-regulate T cell receptor zeta-chain and induce a regulatory phenotype in naive T cells. *Journal of immunology (Baltimore, Md. : 1950)* 176, 6752–61.
- Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., Mello, C.C., 1998. Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391, 806–11. doi:10.1038/35888
- Garcia-Barrio, M., Dong, J., Ufano, S., Hinnebusch, A.G., 2000. Association of GCN1-GCN20 regulatory complex with the N-terminus of eIF2alpha kinase GCN2 is required for GCN2 activation. *The EMBO journal* 19, 1887–99. doi:10.1093/emboj/19.8.1887

- Gehrke, N., Mertens, C., Zillinger, T., Wenzel, J., Bald, T., Zahn, S., Tüting, T., Hartmann, G., Barchet, W., 2013. Oxidative damage of DNA confers resistance to cytosolic nuclease TREX1 degradation and potentiates STING-dependent immune sensing. *Immunity* 39, 482–95. doi:10.1016/j.immuni.2013.08.004
- Gerhard, D.S., Wagner, L., Feingold, E.A., Shenmen, C.M., Grouse, L.H., Schuler, G., Klein, S.L., Old, S., Rasooly, R., Good, P., Guyer, M., Peck, A.M., Derge, J.G., Lipman, D., Collins, F.S., Jang, W., Sherry, S., Feolo, M., Misquitta, L., Lee, E., Rotmistrovsky, K., Greenhut, S.F., Schaefer, C.F., Buetow, K., Bonner, T.I., Haussler, D., Kent, J., Kiekhuis, M., Furey, T., Brent, M., Prange, C., Schreiber, K., Shapiro, N., Bhat, N.K., Hopkins, R.F., Hsie, F., Driscoll, T., Soares, M.B., Casavant, T.L., Scheetz, T.E., Brown-stein, M.J., Usdin, T.B., Toshiyuki, S., Carninci, P., Piao, Y., Dudekula, D.B., Ko, M.S.H., Kawakami, K., Suzuki, Y., Sugano, S., Gruber, C.E., Smith, M.R., Simmons, B., Moore, T., Waterman, R., Johnson, S.L., Ruan, Y., Wei, C.L., Mathavan, S., Gunaratne, P.H., Wu, J., Garcia, A.M., Hulyk, S.W., Fuh, E., Yuan, Y., Sneed, A., Kowis, C., Hodgson, A., Muzny, D.M., McPherson, J., Gibbs, R.A., Fahey, J., Helton, E., Ketteman, M., Madan, A., Rodrigues, S., Sanchez, A., Whiting, M., Madari, A., Young, A.C., Wetherby, K.D., Granite, S.J., Kwong, P.N., Brinkley, C.P., Pearson, R.L., Bouffard, G.G., Blakesly, R.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butterfield, Y.S.N., Griffith, M., Griffith, O.L., Krzywinski, M.I., Liao, N., Morin, R., Morrin, R., Palmquist, D., Petrescu, A.S., Skalska, U., Smailus, D.E., Stott, J.M., Schnerch, A., Schein, J.E., Jones, S.J.M., Holt, R.A., Baross, A., Marra, M.A., Clifton, S., Makowski, K.A., Bosak, S., Malek, J., 2004. The status, quality, and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC). *Genome research* 14, 2121–7. doi:10.1101/gr.2596504
- Ghahary, A., Li, Y., Tredget, E.E., Kilani, R.T., Iwashina, T., Karami, A., Lin, X., 2004. Expression of indoleamine 2,3-dioxygenase in dermal fibroblasts functions as a local immunosuppressive factor. *The Journal of investigative dermatology* 122, 953–64. doi:10.1111/j.0022-202X.2004.22409.x
- Guérout, M., Picot, D., Abi-Ghanem, J., Hartmann, B., Baaden, M., 2010. How cations can assist DNase I in DNA binding and hydrolysis. *PLoS computational biology* 6, e1001000. doi:10.1371/journal.pcbi.1001000

- 
- Gupta, S.L., Carlin, J.M., Pyati, P., Dai, W., Pfefferkorn, E.R., Murphy, M.J., 1994. Antiparasitic and antiproliferative effects of indoleamine 2,3-dioxygenase enzyme expression in human fibroblasts. *Infection and immunity* 62, 2277–84.
- Habibi, D., Jalili, R.B., Forouzandeh, F., Ong, C.J., Ghahary, A., 2010. High expression of IMPACT protein promotes resistance to indoleamine 2,3-dioxygenase-induced cell death. *Journal of Cellular Physiology* 225, 196–205. doi:10.1002/jcp.22220
- Hellman, L.M., Fried, M.G., 2007. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nature protocols* 2, 1849–61. doi:10.1038/nprot.2007.249
- Hinnebusch, A.G., 1997. Translational regulation of yeast GCN4. A window on factors that control initiator-trna binding to the ribosome. *The Journal of biological chemistry* 272, 21661–4.
- Holm, L., Rosenström, P., 2010a. Dali server: conservation mapping in 3D. *Nucleic acids research* 38, W545–9. doi:10.1093/nar/gkq366
- Holm, L., Rosenström, P., 2010b. Dali server: conservation mapping in 3D. *Nucleic acids research* 38, W545–9. doi:10.1093/nar/gkq366
- Hornung, V., Latz, E., 2010. Intracellular DNA recognition. *Nature reviews. Immunology* 10, 123–30. doi:10.1038/nri2690
- Ishida, T., Kinoshita, K., 2007. PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic acids research* 35, W460–4. doi:10.1093/nar/gkm363
- Jackson, R.J., Hellen, C.U.T., Pestova, T. V., 2010. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature reviews. Molecular cell biology* 11, 113–27. doi:10.1038/nrm2838
- Jain, C., 2012. Novel role for RNase PH in the degradation of structured RNA. *Journal of bacteriology* 194, 3883–90. doi:10.1128/JB.06554-11

- Kabsch, W., 2010. XDS. *Acta crystallographica. Section D, Biological crystallography* 66, 125–32. doi:10.1107/S0907444909047337
- Kaneko-Ishino, T., Kohda, T., Ishino, F., 2003. The regulation and biological significance of genomic imprinting in mammals. *Journal of biochemistry* 133, 699–711.
- Keating, S.E., Baran, M., Bowie, A.G., 2011. Cytosolic DNA sensors regulating type I interferon induction. *Trends in immunology* 32, 574–81. doi:10.1016/j.it.2011.08.004
- Kleywegt, G.J., Brünger, A.T., 1996. Checking your imagination: applications of the free R value. *Structure (London, England : 1993)* 4, 897–904.
- Kosaki, K., Suzuki, T., Kosaki, R., Yoshihashi, H., Itoh, M., Goto, Y., Matsuo, N., 2001. Human homolog of the mouse imprinted gene *Impact* resides at the pericentric region of chromosome 18 within the critical region for bipolar affective disorder. *Molecular psychiatry* 6, 87–91.
- Krissinel, E., Henrick, K., 2007. Inference of macromolecular assemblies from crystalline state. *Journal of molecular biology* 372, 774–97. doi:10.1016/j.jmb.2007.05.022
- Kubota, H., Ota, K., Sakaki, Y., Ito, T., 2001. Budding yeast GCN1 binds the GI domain to activate the eIF2 $\alpha$  kinase GCN2. *The Journal of biological chemistry* 276, 17591–6. doi:10.1074/jbc.M011793200
- Kubota, H., Sakaki, Y., Ito, T., 2000. GI domain-mediated association of the eukaryotic initiation factor 2 $\alpha$  kinase GCN2 with its activator GCN1 is required for general amino acid control in budding yeast. *The Journal of biological chemistry* 275, 20243–6. doi:10.1074/jbc.C000262200
- KUNITZ, M., 1950. Crystalline desoxyribonuclease; isolation and general properties; spectrophotometric method for the measurement of desoxyribonuclease activity. *The Journal of general physiology* 33, 349–62.

- 
- Lageix, S., Rothenburg, S., Dever, T.E., Hinnebusch, A.G., 2014. Enhanced interaction between pseudokinase and kinase domains in Gcn2 stimulates eIF2 $\alpha$  phosphorylation in starved cells. *PLoS genetics* 10, e1004326. doi:10.1371/journal.pgen.1004326
- Lee, G.K., Park, H.J., Macleod, M., Chandler, P., Munn, D.H., Mellor, A.L., 2002. Tryptophan deprivation sensitizes activated T cells to apoptosis prior to cell division. *Immunology* 107, 452–60.
- Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2<sup>-</sup>( $\Delta\Delta C_T$ ) Method. *Methods (San Diego, Calif.)* 25, 402–8. doi:10.1006/meth.2001.1262
- Makrides, S.C., 1996. Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiological reviews* 60, 512–38.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., 2007. Phaser crystallographic software. *Journal of applied crystallography* 40, 658–674. doi:10.1107/S0021889807021206
- Mellor, A.L., Munn, D.H., 2004. IDO expression by dendritic cells: tolerance and tryptophan catabolism. *Nature reviews. Immunology* 4, 762–74. doi:10.1038/nri1457
- Morison, I.M., Reeve, A.E., 1998. A catalogue of imprinted genes and parent-of-origin effects in humans and animals. *Human molecular genetics* 7, 1599–609.
- Mueller, P.P., Hinnebusch, A.G., 1986. Multiple upstream AUG codons mediate translational control of GCN4. *Cell* 45, 201–7.
- Munn, D.H., Sharma, M.D., Baban, B., Harding, H.P., Zhang, Y., Ron, D., Mellor, A.L., 2005. GCN2 kinase in T cells mediates proliferative arrest and anergy induction in response to indoleamine 2,3-dioxygenase. *Immunity* 22, 633–42. doi:10.1016/j.immuni.2005.03.013
- Nameki, N., Yoneyama, M., Koshiba, S., Tochio, N., Inoue, M., Seki, E., Matsuda, T., Tomo, Y., Harada, T., Saito, K., Kobayashi, N., Yabuki, T., Aoki, M., Nunokawa, E.,

- Matsuda, N., Sakagami, N., Terada, T., Shirouzu, M., Yoshida, M., Hirota, H., Osanai, T., Tanaka, A., Arakawa, T., Carninci, P., Kawai, J., Hayashizaki, Y., Kinoshita, K., Güntert, P., Kigawa, T., Yokoyama, S., 2004. Solution structure of the RWD domain of the mouse GCN2 protein. *Protein science : a publication of the Protein Society* 13, 2089–100. doi:10.1110/ps.04751804
- Ng, C.L., Waterman, D.G., Antson, A.A., Ortiz-Lombardía, M., 2010. Structure of the *Methanothermobacter thermautotrophicus* exosome RNase PH ring. *Acta crystallographica. Section D, Biological crystallography* 66, 522–8. doi:10.1107/S0907444910002908
- O'Neill, L.A.J., Bowie, A.G., 2010. Sensing and signaling in antiviral innate immunity. *Current biology : CB* 20, R328–33. doi:10.1016/j.cub.2010.01.044
- Okamura, K., Hagiwara-takeuchi, Y., Li, T., Vu, T.H., Hirai, M., Hattori, M., Sakaki, Y., Hoffman, A.R., Ito, T., 2000. Comparative Genome Analysis of the Mouse Imprinted Gene *Impact* and Its Nonimprinted Human Homolog *IMPACT* : Toward the Structural Basis for Species-Specific Imprinting 1878–1889. doi:10.1101/gr.139200.7
- Okamura, K., Yamada, Y., Sakaki, Y., Ito, T., 2004. An evolutionary scenario for genomic imprinting of *Impact* lying between nonimprinted neighbors. *DNA research : an international journal for rapid publication of reports on genes and genomes* 11, 381–90.
- Park, F., Gajiwala, K., Eroshkina, G., Furlong, E., He, D., Batiyenko, Y., Romero, R., Christopher, J., Badger, J., Hendle, J., Lin, J., Peat, T., Buchanan, S., 2004. Crystal structure of YIGZ, a conserved hypothetical protein from *Escherichia coli* k12 with a novel fold. *Proteins* 55, 775–7. doi:10.1002/prot.20087
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of computational chemistry* 25, 1605–12. doi:10.1002/jcc.20084
- Pfefferkorn, E.R., 1984. Interferon gamma blocks the growth of *Toxoplasma gondii* in human fibroblasts by inducing the host cells to degrade tryptophan.

---

Proceedings of the National Academy of Sciences of the United States of America 81, 908–12.

- Poormasjedi-Meibod, M.-S., Jalili, R.B., Hosseini-Tabatabaei, A., Hartwell, R., Ghahary, A., 2013. Immuno-regulatory function of indoleamine 2,3 dioxygenase through modulation of innate immune responses. *PloS one* 8, e71044. doi:10.1371/journal.pone.0071044
- Potterton, E., Briggs, P., Turkenburg, M., Dodson, E., 2003. A graphical user interface to the CCP4 program suite. *Acta crystallographica. Section D, Biological crystallography* 59, 1131–7.
- Powell, H.R., Johnson, O., Leslie, A.G.W., 2013. Autoindexing diffraction images with iMosflm. *Acta crystallographica. Section D, Biological crystallography* 69, 1195–203. doi:10.1107/S0907444912048524
- Powrie, F., Maloy, K.J., 2003. Immunology. Regulating the regulators. *Science (New York, N.Y.)* 299, 1030–1. doi:10.1126/science.1082031
- Ramirez, M., Wek, R.C., Hinnebusch, A.G., 1991. Ribosome association of GCN2 protein kinase, a translational activator of the GCN4 gene of *Saccharomyces cerevisiae*. *Molecular and cellular biology* 11, 3027–36.
- Reinhard, L., Mayerhofer, H., Geerlof, A., Mueller-Dieckmann, J., Weiss, M.S., 2013. Optimization of protein buffer cocktails using Thermofluor. *Acta crystallographica. Section F, Structural biology and crystallization communications* 69, 209–14. doi:10.1107/S1744309112051858
- Roffé, M., Hajj, G.N.M., Azevedo, H.F., Alves, V.S., Castilho, B.A., 2013. IMPACT is a developmentally regulated protein in neurons that opposes the eukaryotic initiation factor 2 $\alpha$  kinase GCN2 in the modulation of neurite outgrowth. *The Journal of biological chemistry* 288, 10860–9. doi:10.1074/jbc.M113.461970
- Sammito, M., Meindl, K., de Ilarduya, I.M., Millán, C., Artola-Recolons, C., Hermoso, J.A., Usón, I., 2014. Structure solution with ARCIMBOLDO using fragments derived from distant homology models. *The FEBS journal* 281, 4029–45. doi:10.1111/febs.12897



- Sandhu, K.S., Dash, D., 2006. Conformational flexibility may explain multiple cellular roles of PEST motifs. *Proteins* 63, 727–32. doi:10.1002/prot.20918
- Sattlegger, E., Barbosa, J. a R.G., Moraes, M.C.S., Martins, R.M., Hinnebusch, A.G., Castilho, B. a, 2011. Gcn1 and actin binding to Yih1: implications for activation of the eIF2 kinase GCN2. *The Journal of biological chemistry* 286, 10341–55. doi:10.1074/jbc.M110.171587
- Sattlegger, E., Hinnebusch, A.G., 2000. Separate domains in GCN1 for binding protein kinase GCN2 and ribosomes are required for GCN2 activation in amino acid-starved cells. *The EMBO journal* 19, 6622–33. doi:10.1093/emboj/19.23.6622
- Sattlegger, E., Swanson, M.J., Ashcraft, E. a, Jennings, J.L., Fekete, R. a, Link, A.J., Hinnebusch, A.G., 2004. YIH1 is an actin-binding protein that inhibits protein kinase GCN2 and impairs general amino acid control when overexpressed. *The Journal of biological chemistry* 279, 29952–62. doi:10.1074/jbc.M404009200
- Scheper, G.C., van der Knaap, M.S., Proud, C.G., 2007. Translation matters: protein synthesis defects in inherited disease. *Nature reviews. Genetics* 8, 711–23. doi:10.1038/nrg2142
- Silva, R.C., Dautel, M., Di Genova, B.M., Amberg, D.C., Castilho, B.A., Sattlegger, E., 2015. The Gcn2 Regulator Yih1 Interacts with the Cyclin Dependent Kinase Cdc28 and Promotes Cell Cycle Progression through G2/M in Budding Yeast. *PloS one* 10, e0131070. doi:10.1371/journal.pone.0131070
- Singh, G.P., Ganapathi, M., Sandhu, K.S., Dash, D., 2006. Intrinsic unstructuredness and abundance of PEST motifs in eukaryotic proteomes. *Proteins* 62, 309–15. doi:10.1002/prot.20746
- Sonenberg, N., Hinnebusch, A.G., 2009. Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell* 136, 731–45. doi:10.1016/j.cell.2009.01.042

- 
- Sonenberg, N., Hinnebusch, A.G., 2007. New Modes of Translational Control in Development, Behavior, and Disease. *Molecular Cell* 28, 721–729. doi:10.1016/j.molcel.2007.11.018
- St John, F.J., Feng, B., Pozharski, E., 2008. The role of bias in crystallization conditions in automated microseeding. *Acta crystallographica. Section D, Biological crystallography* 64, 1222–7. doi:10.1107/S0907444908031302
- Sultana, A., Lee, J.E., 2015. Measuring protein-protein and protein-nucleic Acid interactions by biolayer interferometry. *Current protocols in protein science / editorial board, John E. Coligan ... [et al.]* 79, 19.25.1–19.25.26. doi:10.1002/0471140864.ps1925s79
- Takikawa, O., 2005. Biochemical and medical aspects of the indoleamine 2,3-dioxygenase-initiated L-tryptophan metabolism. *Biochemical and biophysical research communications* 338, 12–9. doi:10.1016/j.bbrc.2005.09.032
- Tallóczy, Z., Jiang, W., Virgin, H.W., Leib, D.A., Scheuner, D., Kaufman, R.J., Eskelinen, E.-L., Levine, B., 2002. Regulation of starvation- and virus-induced autophagy by the eIF2alpha kinase signaling pathway. *Proceedings of the National Academy of Sciences of the United States of America* 99, 190–5. doi:10.1073/pnas.012485299
- Taylor, M.W., Feng, G.S., 1991. Relationship between interferon-gamma, indoleamine 2,3-dioxygenase, and tryptophan catabolism. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 5, 2516–22.
- Thompson, J.D., Gibson, T.J., Higgins, D.G., 2002. Multiple sequence alignment using ClustalW and ClustalX. *Current protocols in bioinformatics / editorial board, Andreas D. Baxevanis ... [et al.]* Chapter 2, Unit 2.3. doi:10.1002/0471250953.bi0203s00
- Tickle, I.J., Laskowski, R.A., Moss, D.S., 2000. Rfree and the rfree ratio. II. Calculation Of the expected values and variances of cross-validation statistics in macromolecular least-squares refinement. *Acta crystallographica. Section D, Biological crystallography* 56, 442–50.

- Waller, T., Lee, S.J., Sattlegger, E., 2012. Evidence that Yih1 resides in a complex with ribosomes. *The FEBS journal* 279, 1761–76. doi:10.1111/j.1742-4658.2012.08553.x
- Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., Barton, G.J., 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics (Oxford, England)* 25, 1189–91. doi:10.1093/bioinformatics/btp033
- Wek, R.C., Jiang, H.-Y., Anthony, T.G., 2006. Coping with stress: eIF2 kinases and translational control. *Biochemical Society transactions* 34, 7–11. doi:10.1042/BST20060007
- Wong, M.L., Medrano, J.F., 2005. Real-time PCR for mRNA quantitation. *BioTechniques* 39, 75–85.
- Yan, N., Regalado-Magdos, A.D., Stiggelbout, B., Lee-Kirsch, M.A., Lieberman, J., 2010. The cytosolic exonuclease TREX1 inhibits the innate immune response to human immunodeficiency virus type 1. *Nature immunology* 11, 1005–13. doi:10.1038/ni.1941
- Yang, W., 2011. Nucleases: diversity of structure, function and mechanism. *Quarterly reviews of biophysics* 44, 1–93. doi:10.1017/S0033583510000181
- Zhou, D., Palam, L.R., Jiang, L., Narasimhan, J., Staschke, K.A., Wek, R.C., 2008. Phosphorylation of eIF2 directs ATF5 translational control in response to diverse stress conditions. *The Journal of biological chemistry* 283, 7064–73. doi:10.1074/jbc.M708530200



## **6. APPENDIX**

---





**Table 1**  
Macromolecule-production information.

Source organism	Human
DNA source	cDNA library
Forward primer	GCGCATATGGAGATGGGACAGCGGATTC
Reverse primer	CGCCTCGAGTTAGTCATAGCCGTCGAGATA
Expression vector	pET-28a
Expression host	<i>E. coli</i> BL21 (DE3) CodonPlus RIPL
Complete amino-acid sequence of the construct produced	MGSSHHHHHHSSGLEVLFGQPHMEMGRRIQLLELR-NRTPFDVKELVLDNRSRNEGKLEGLTDFEEL-EFLSTINVLGTSIANLPKLNKLELSDNRV-SGGLEVLAEKCPNLTHLNSGNKIKDLSTIEP-LKKLENLKSLDLNFCVETNLNDYRENVFKLLP-QLTYLDGYD

several LRR domain crystal structures exhibit their ligand-binding domain in this concave surface.

Our high-resolution human PP32A LRR domain structure (PDB entry 4xos) shows some differences compared with previously reported structures (Huyton & Wolberger, 2007), including some modifications at the end of helix  $\alpha 1$  and the loop connecting it to the  $\beta 1$  strand. This change is involved in the generation of a totally different crystal packing. A chloride ion is located in the vicinity, interacting with residues towards the end of the  $\alpha 1$ - $\beta 1$  turn and the initial  $\beta 1$  residues.

## 2. Materials and methods

### 2.1. Macromolecule production

A fragment of the gene encoding human PP32A (protein residues 1–149) was cloned into a modified version of the pET-28a vector (Table 1). This construct was expressed in *Escherichia coli* BL21 (DE3) CodonPlus RIPL cells (Invitrogen) for 3 h at 303 K following induction with 1 mM IPTG when the cell density had reached a value of between 0.6 and 0.8. The cells were harvested, washed with PBS and stored at  $-80^{\circ}\text{C}$  until use.

For purification, the cells were sonicated with a lysis buffer consisting of 50 mM Tris-HCl pH 7.5, 0.1% Triton X-100, 500 mM NaCl, 10% glycerol, 1 mM  $\beta$ -mercaptoethanol (BME). To remove cell debris, the lysate was centrifuged prior to loading onto a 5 ml chelating nickel HisTrap HP column (GE Healthcare). The column was washed with six column volumes of buffer A (50 mM Tris-HCl pH 7.5, 350 mM NaCl, 5% glycerol, 2 mM BME) and the protein was eluted using a gradient to buffer B (50 mM Tris-HCl pH 7.5, 350 mM NaCl, 5% glycerol, 500 mM imidazole, 2 mM BME). These buffers were selected from Thermofluor stability assay results (Boivin *et al.*, 2013; Reinhard *et al.*, 2013). Elution fractions were analyzed by SDS-PAGE and the fractions containing PP32 were treated overnight at  $4^{\circ}\text{C}$  with recombinant human rhinovirus 3C protease fused to glutathione S-transferase. After digestion, the sample was loaded onto a Sepharose column covalently linked to glutathione in order to remove the protease. The sample was concentrated and injected onto a Superdex 200 16/60 size-exclusion column (GE Healthcare) equilibrated with 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM BME.

**Table 2**  
Crystallization conditions.

Method	Sitting-drop vapour diffusion
Plate type	96-well plate
Temperature (K)	294
Protein concentration ( $\mu\text{M}$ )	300
Buffer composition of protein solution	20 mM Tris-HCl pH 7.5, 150 mM NaCl, 1 mM BME
Composition of reservoir solution	20% PEG 3000, 0.1 M HEPES pH 7.5, 0.2 M NaCl
Volume and ratio of drop	0.4 $\mu\text{l}$ (1:1)
Volume of reservoir ( $\mu\text{l}$ )	0.4

**Table 3**  
Data collection and processing.

Values in parentheses are for the outer shell.

Diffraction source	ALBA synchrotron
Wavelength ( $\text{\AA}$ )	0.9795
Temperature (K)	100
Detector	Pilatus 6M
Crystal-to-detector distance (mm)	258.67
Rotation range per image ( $^{\circ}$ )	0.1
Total rotation range ( $^{\circ}$ )	120
Exposure time per image (s)	0.1
Space group	$P12_11$
$a, b, c$ ( $\text{\AA}$ )	48.04, 50.88, 63.02
$\alpha, \beta, \gamma$ ( $^{\circ}$ )	90, 110.19, 90
Mosaicity ( $^{\circ}$ )	0.865
Resolution range ( $\text{\AA}$ )	45.09–1.56 (1.615–1.559)
Total No. of reflections	87304 (7888)
No. of unique reflections	36978 (3587)
Completeness $\dagger$ (%)	90.50 (85.63)
Multiplicity	2.3 (2.2)
$\langle I(\sigma(I)) \rangle \ddagger$	12.14 (1.77)
$R_{\text{meas}} \S$	0.05
Overall $B$ factor from Wilson plot ( $\text{\AA}^2$ )	17.07

$\dagger$  The crystal orientation and morphology did not allow complete data collection.  $\ddagger$  The mean  $\langle I(\sigma(I)) \rangle$  in the outer shell is 2.0 at 1.6  $\text{\AA}$  resolution.  $\S$   $R_{\text{meas}} = \sum_{hkl} |I(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} I(hkl)$ , where  $I(hkl)$  are the observed intensities,  $\langle I(hkl) \rangle$  is the average intensity and  $N(hkl)$  is the multiplicity of reflection  $hkl$ .

### 2.2. Crystallization

Plate-like crystals with approximate dimensions of  $0.2 \times 0.6 \times 0.7$  mm were grown by mixing 0.4  $\mu\text{l}$  protein solution at 300  $\mu\text{M}$  and 0.4  $\mu\text{l}$  reservoir solution (20% PEG 3000, 0.1 M HEPES pH 7.5, 0.2 M NaCl) at 294 K using the sitting-drop vapour-diffusion method (Table 2). Crystals appeared within one week.

### 2.3. Data collection and processing

One- to two-week-old crystals were flash-cryocooled with liquid nitrogen without the use of any additional cryoprotectant. Diffraction data were collected from the crystals at 100 K on the XALOC beamline at the ALBA synchrotron, Barcelona, Spain. A complete data set was collected to 1.56  $\text{\AA}$  resolution from a single crystal. Data were reduced using *MOSFLM* (Battye *et al.*, 2011) and *SCALA* (Evans, 2006) from *CCP4* (Winn *et al.*, 2011).

### 2.4. Structure solution and refinement

Phases were determined by molecular replacement with *Phaser* (McCoy *et al.*, 2007), using the structure of the LRR



## research communications

domain from PP32 (PDB entry 2je0; Huyton & Wolberger, 2007) as a model. The solution, consisting of two molecules in the asymmetric unit, was first refined as a rigid body and then by restrained refinement using *phenix.refine* (Afonine *et al.*, 2012) and manual model building using *Coot* (Emsley *et al.*, 2010). We also performed a final refinement cycle with TLS (translation/libration/screw) refinement (Schomaker & Trueblood, 1968). *phenix.validate* was used for model validation (Adams *et al.*, 2010).

Interchain interactions were analyzed with *CONTACT/ACT* (Kabsch & Sander, 1983) as implemented in *CCP4* (Winn *et al.*, 2011). Only high-probability electrostatic interactions have been considered for comparison. Model images were generated with *Chimera* (Pettersen *et al.*, 2004)

## 3. Results and discussion

The crystals belonged to space group  $P2_1$  (Table 3) and contained two molecules in the asymmetric unit, with a solvent content of 42%, which is around 20% lower than those of the two previously reported structures. The higher resolution structure improves the global validation metrics, and the side-chain outliers decrease from 9.1% in the 2.4 Å resolution model to 1.1% in our 1.56 Å resolution structure. The final model included 149 residues of PP32A and three further amino acids at the N-terminus resulting from the cloning and protease-cleavage strategies. The 1.56 Å resolution structure has been refined to a final  $R_{\text{work}}$  and  $R_{\text{free}}$  of 0.1706 and 0.2048, respectively (Table 4). No PP32A residues in the asymmetric unit were found in the Ramachandran disallowed regions. The better resolution of our crystal allowed us to identify different rotamers for residues Leu22, Asn59, Arg75, Lys99, Thr105, Thr126, Tyr144 and Tyr148 compared with the molecules in

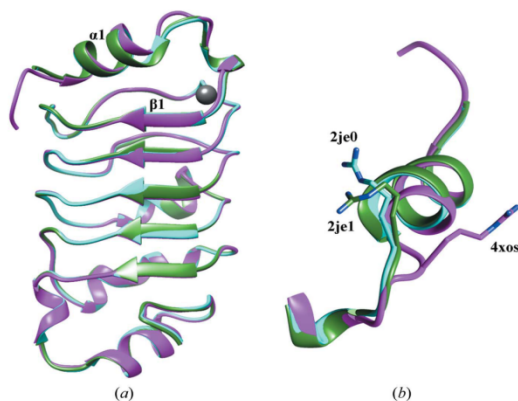


Figure 1  
(a) Superposition of a ribbon representation of PDB entries 4xos (magenta), 2je0 (cyan) and 2je1 (green) showing a shift of the  $\alpha 1$ - $\beta 1$  turn and the chloride position in 4xos (grey). (b) Comparison of the Arg12 side-chain distribution in the shifted area of PDB entries 4xos, 2je0 and 2je1 using the same colour scheme.

Table 4  
Structure refinement.

Values in parentheses are for the outer shell.

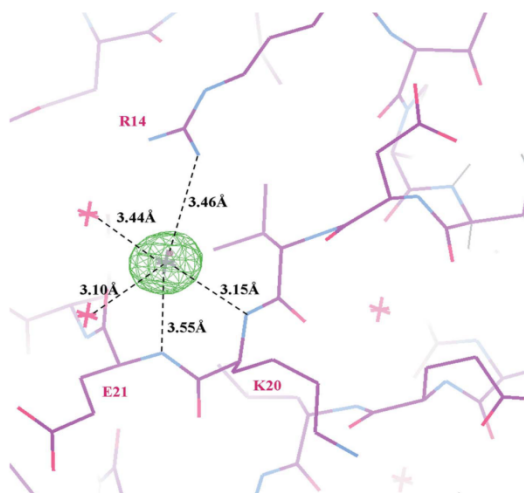
Resolution range (Å)	45.09–1.559
Completeness (%)	90.50 (85.63)
$\sigma$ Cutoff	1.63
No. of reflections, working set	36969
No. of reflections, test set	1851
Final $R_{\text{work}}^{\dagger}$	0.1603 (0.2514)
Final $R_{\text{free}}^{\ddagger}$	0.1929 (0.2808)
Cruickshank DPI§	0.13
No. of non-H atoms	
Protein	2456
Ion	2
Ligand	12
Water	163
Total	2633
R.m.s. deviations	
Bonds (Å)	0.007
Angles (°)	1.07
Average $B$ factors (Å <sup>2</sup> )	
Overall	28.0
Protein	25.6
Ion	27.0
Ligand	33.2
Water	30.3
Ramachandran plot	
Favoured regions (%)	97
Additionally allowed (%)	3
Outliers (%)	0

$\dagger R_{\text{work}} = \sum_{\text{obs}} |F_{\text{obs}}| - |F_{\text{calc}}| / \sum_{\text{obs}} |F_{\text{obs}}|$ , where  $F_{\text{obs}}$  and  $F_{\text{calc}}$  are the observed and calculated structure factors, respectively.  $\ddagger R_{\text{free}}$  is the same as  $R_{\text{work}}$  but was calculated using 5% of the total reflections that were selected randomly and excluded during refinement.  $\S$  Diffraction-component precision indicator (Cruickshank, 1999).

PDB entries 2je0 and 2je1. A double conformation has also been detected for Arg28.

Superimposition of the  $C^{\alpha}$  atoms among monomers from PDB entries 4xos, 2je0 and 2je1 shows a root-mean-square deviation (r.m.s.d.) of around 0.5 Å, indicating high similarity to the previous structures deposited in the PDB (Huyton & Wolberger, 2007). The main differences in the  $C^{\alpha}$  trace are located in the turn between helix  $\alpha 1$  and strand  $\beta 1$  at the N-terminus (Fig. 1a). This difference is maintained in all of the molecules in the asymmetric unit from previous structures. In particular, Arg12 adopts a totally different conformation (Fig. 1b) in both molecules in the asymmetric unit, establishing a double salt bridge with Asp73 from a neighbouring molecule. As a consequence, both Leu11 and Asn13 also occupy different positions in our model with respect to the previously reported structures. The C-terminal part of the  $\alpha 1$  helix is also displaced. This N-terminal modified region therefore shows a certain degree of conformational flexibility when comparing the different LRR domain structures, and is in agreement with the higher temperature factors observed in the  $\alpha 1$ - $\beta 1$  loop. This finding might be functionally relevant since residues Thr15 and Ser17 have been shown to undergo cell cycle-regulated phosphorylation (Dephoure *et al.*, 2008).

Our structure shows the binding of a chloride ion towards the end of the  $\alpha 1$ - $\beta 1$  turn. We assigned the chloride taking into account the ions that were present during purification and crystallization. The chloride interacts with the  $N^{\eta}$  atom of Arg14, the main-chain N atoms of Lys20 and Glu21 and two water molecules (Fig. 2). This interacting region immediately



**Figure 2**  
Model inside a  $F_o - F_c$  map (green) contoured at the  $6\sigma$  level. The map was calculated in the absence of chloride and the two interacting water molecules. Interacting distances of the chloride ion (grey) with N atoms belonging to Arg14, Lys20 and Glu21 and with water molecules are shown.

precedes the first LRR motif of the protein (Fig. 1a) in the vicinity of the largest conformational change between the previously reported structures and our model. The chloride ion site would be occupied by the side chain of Glu21 from neighbouring molecules in the asymmetric units of both PDB entries 2je0 and 2je1, making the presence of this ion incompatible with the previously reported crystallographic space groups.

Despite belonging to different space groups, the structures with PDB codes 2je0 and 2je1 reveal related crystal packing (Supplementary Table S1). This degree of conservation is lost when compared with our PP32A structure since most of the contacts are not maintained. The new packing observed is related to the conformational change of the turn connecting the first  $\alpha$ -helix to the first  $\beta$ -strand (Supplementary Table S1). PISA analysis (Krissinel & Henrick, 2007) revealed no change in the protein assembly, which remains as a monomer. This is in agreement with the behaviour of the PP32A LRR domain in size-exclusion chromatography. We can therefore conclude that the differences observed in crystal packing, the new conformation of the  $\alpha 1$ - $\beta 1$  turn and the presence of the chloride ion do not have any influence on the oligomerization state of the LRR domain.

### Acknowledgements

The research received funding from the European Community's Seventh Framework Program (FP7/2007-2013) under BioStruct-X (grant agreement No. 283570). The experiments were performed on the XALOC beamline at the ALBA Synchrotron Light Facility with the collaboration of the ALBA staff. This work was supported by Ministerio de Economía y Competitividad (SAF2012-31405 and PROMETEO/2012/061 project to JB), Generalitat Valenciana, Spain. SZC received a fellowship from Ministerio de Economía y Competitividad BES2010-030116.

### References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.  
 Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* **D68**, 352–367.  
 Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* **D67**, 271–281.  
 Bella, J., Hindle, K. L., McEwan, P. A. & Lovell, S. C. (2008). *Cell. Mol. Life Sci.* **65**, 2307–2333.  
 Beresford, P. J., Zhang, D., Oh, D. Y., Fan, Z., Greer, E. L., Russo, M. L., Jaju, M. & Lieberman, J. (2001). *J. Biol. Chem.* **276**, 43285–43293.  
 Boivin, S., Kozak, S. & Meijers, R. (2013). *Protein Expr. Purif.* **91**, 192–206.  
 Cruickshank, D. W. J. (1999). *Acta Cryst.* **D55**, 583–601.  
 Dephoure, N., Zhou, C., Villén, J., Beausoleil, S. A., Bakalarski, C. E., Elledge, S. J. & Gygi, S. P. (2008). *Proc. Natl Acad. Sci. USA*, **105**, 10762–10767.  
 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.  
 Evans, P. (2006). *Acta Cryst.* **D62**, 72–82.  
 Huyton, T. & Wolberger, C. (2007). *Protein Sci.* **16**, 1308–1315.  
 Kabsch, W. & Sander, C. (1983). *Biopolymers*, **22**, 2577–2637.  
 Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.  
 Li, M., Makkinje, A. & Damuni, Z. (1996). *Biochemistry*, **35**, 6998–7002.  
 Lieberman, J. (2010). *Immunol. Rev.* **235**, 93–104.  
 Mazroui, R., Di Marco, S., Clair, E., von Roretz, C., Tenenbaum, S. A., Keene, J. D., Saleh, M. & Gallouzi, I.-E. (2008). *J. Cell Biol.* **180**, 113–127.  
 McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.  
 Muto, S., Senda, M., Akai, Y., Sato, L., Suzuki, T., Nagai, R., Senda, T. & Horikoshi, M. (2007). *Proc. Natl Acad. Sci. USA*, **104**, 4285–4290.  
 Pan, W., da Graca, L. S., Shao, Y., Yin, Q., Wu, H. & Jiang, X. (2009). *J. Biol. Chem.* **284**, 6946–6954.  
 Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605–1612.  
 Reinhard, L., Mayerhofer, H., Geerlof, A., Mueller-Dieckmann, J. & Weiss, M. S. (2013). *Acta Cryst.* **F69**, 209–214.  
 Schomaker, V. & Trueblood, K. N. (1968). *Acta Cryst.* **B24**, 63–76.  
 Seo, S., Macfarlan, T., McNamara, P., Hong, R., Mukai, Y., Heo, S. & Chakravarti, D. (2002). *J. Biol. Chem.* **277**, 14005–14010.  
 Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.



## Quaternary structure of *Dioclea grandiflora* lectin assessed by equilibrium sedimentation and crystallographic analysis of recombinant mutants



Sara Zamora-Caballero<sup>a</sup>, Alicia Pérez<sup>b</sup>, Libia Sanz<sup>b</sup>, Jerónimo Bravo<sup>a,\*</sup>, Juan J. Calvete<sup>b,\*</sup>

<sup>a</sup> Unidad de Transducción de Señales, Instituto de Biomedicina de Valencia, CSIC, Spain

<sup>b</sup> Laboratorio de Venómica Estructural y Funcional, Instituto de Biomedicina de Valencia, CSIC, Spain

### ARTICLE INFO

#### Article history:

Received 29 June 2015

Revised 14 July 2015

Accepted 15 July 2015

Available online 27 July 2015

Edited by Miguel De la Rosa

#### Keywords:

Recombinant Diocleinae lectin

Dimer–tetramer equilibrium

Site-directed mutagenesis

Analytical ultracentrifugation

X-ray crystal structure

### ABSTRACT

The structural basis of the pH dependency of the dimer–tetramer transition exhibited by Brinda's type II *Diocleinae* lectins was investigated by equilibrium sedimentation and X-ray crystal structure determination of recombinant wild-type and site-directed single and double mutants of the pH-stable tetrameric *Dioclea grandiflora* lectin (r- $\alpha$ DGL). Releasing the peripheral site interdimeric contact between R60 and D78 rendered a mutant displaying dimer–tetramer equilibrium in the pH range equivalent to  $pK_a \pm 1$  of the  $\gamma$ -COOH. Mutation of both histidines 51 and 131, but not the mutation of each His separately, abolished the formation of the *Diocleinae* canonical tetramer in the pH range 2.5–8.5. The X-ray structure of the double mutant r- $\alpha$ DGL H51G/H131N suggests that H131 plays a crucial role in networking loop 114–125 residues from all four subunits at the central cavity of the tetrameric lectin, and that H51 maintains the central cavity loops in a proper spatial orientation to make H131-mediated interdimer contacts.

© 2015 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

### 1. Introduction

Ubiquitous in all the kingdoms of life, lectins are proteins of non-immune origin, that recognize specific sugar structures attached to soluble and integral cell membrane glycoconjugates [12]. Compared to proteins and nucleic acids, carbohydrates are unbeatable in information-coding potential. Due to their ability to recognize specific glycodes, lectins play pivotal roles in carbohydrate-mediated biological processes, such as communication between cells, tumor metastasis, innate immunity, sperm-egg recognition, infections by parasites, plant protection against phytopathogens and phyttopredators, etc. [11]. Lectins are not only valuable research tools for the molecular understanding of the sugar code [13,2] but also attractive molecules for targeting drugs to specific destinations [18].

Lectins isolated from the seeds of genera *Dioclea*, *Cratylia*, and *Canavalia* of the *Diocleinae* subtribe of the *Papilionoideae* subfamily of *Leguminosae* display the back-to-back arrangement of two dimers built by the side-by-side, antiparallel alignment of the subunit six-stranded back  $\beta$ -sheets [20], first reported in the X-ray structure of *Canavalia ensiformis*' seed lectin, concanavalin A [25]. Type II legume lectin tetramers (*sensu* [6]) exhibit pH-dependent dimer–tetramer equilibrium [17,28,29]. A  $25 \times 8 \text{ \AA}^2$  water-filled central cavity formed upon tetramerization has been reported to bind non-polar molecules, such as plant cytokinins and  $\beta$ -indoleacetic acid [8,26]. These molecules act as plant growth hormones and thus *Diocleinae* lectin dimer–tetramer transition could serve as a cargo mechanism for these hormones between different parts of the plant. Reported pH values in the xylem–lumen space of higher plants range from 3.5 to 8.3, depending on plant species, plant parts, experimental conditions and method used [33], and the pH values of the two main cellular compartments, the cytoplasm and the vacuole, are, respectively, slightly alkaline (7.4–7.5) and markedly acid (4.5–6.0) [16]. However, the precise distribution of the physiological pH in the different tissues, cells, subcellular compartments, and fluids of the plant is needed to support the “cargo hypothesis”. Nonetheless, knowledge of the structural determinants that modulate the quaternary assembly of *Diocleinae* lectins would facilitate the long-sought biotechnological

**Author contributions:** JJC and LS conceived and supervised the study and designed experiments; AP and LS produced the recombinant wild-type and site-directed lectins; AP performed crystallization experiments; JB and SZ-C collected and analyzed the crystallographic data; JJC, JB and SZ-C wrote the manuscript; all authors made manuscript revisions.

\* Corresponding authors at: Instituto de Biomedicina de Valencia, CSIC, Jaime Roig 11, 46010 Valencia, Spain. Fax: +34 96 369 0800.

E-mail addresses: [jbravo@ibv.csic.es](mailto:jbravo@ibv.csic.es) (J. Bravo), [jcalvete@ibv.csic.es](mailto:jcalvete@ibv.csic.es) (J.J. Calvete).

<http://dx.doi.org/10.1016/j.febslet.2015.07.020>

0014-5793/© 2015 Federation of European Biochemical Societies. Published by Elsevier B.V. All rights reserved.

use of these lectins as transport proteins to release its cargo to their specific (natural or engineered) glycan ligand-bearing destinations (Vasta and Ahmed [31]).

Structural comparisons of the crystal structures of pH-independent tetrameric lectins (*D. grandiflora* (1DGL, [27], and *Dioclea violacea* (2GDF, [14]; 3AX4, [4] lectins) and homologous lectins that exhibit pH-dependent dimer–tetramer transitions, *D. guianensis* lectin (Dguial) (1H9P, 1H9W; [32] and *Cratylia floribunda* lectin (CFL, 2D3R, 2D3P; [15], have revealed that the dimer–tetramer equilibrium depends on interdimeric contacts between regions involving homologous residues 53–78 of opposing monomers (A and D and B and C) at the periphery of the dimers, and the loops formed by residues 117–123 of each monomer facing the central cavity of the tetramer (Fig. 1). More recently, the development of a system for the expression of recombinant DGL (P08902) and Dguial (A9J248), structurally and functionally indistinguishable from their natural versions [23], has opened the door to a more detailed analysis of the interactions that govern the tetrameric association of *Diocleinae* lectins. Thus, conversion of the pH-stable tetrameric r-DGL into a structure exhibiting pH-dependent dimer–tetramer transition was accomplished through three simultaneous mutations, E123A/H131N/K132/Q, that abolished the interdimeric contacts networking the central

cavity loops [23]. Here, we report the distinct structural impact of single and double mutations affecting residues H51, R60, and H131, assessed by equilibrium sedimentation and crystallographic analysis.

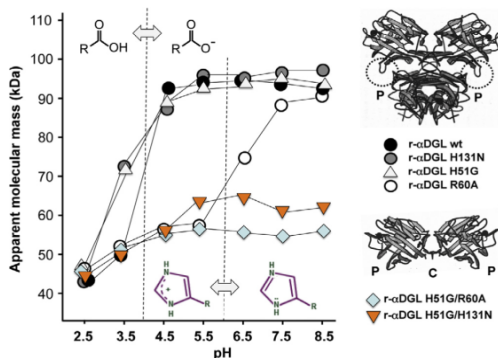
## 2. Materials and methods

### 2.1. Generation, expression, and purification of recombinant lectins

Construction of a synthetic gene encoding the wild-type full-length  $\alpha$ -chain of *D. grandiflora* lectin [P08902] (r- $\alpha$ -DGL) was performed as described previously [23]. Generation of single (H51A, H51G, R60A) and double (H51A/R60A, H51G–H131N) mutants was done using the QuickChange® site-directed mutagenesis kit (Stratagene) as described [23]. Briefly, pET32a plasmids containing the template sequences (Table 1) flanked by NcoI and NotI restriction sites were PCR-amplified [initial denaturation at 94 °C for 2 min, followed by 16 cycles of denaturation (30 s at 94 °C), annealing (60 s at 55 °C), and extension (13 min at 68 °C), and a final extension for 10 min at 68 °C] using the forward (F) and reverse (R) primers listed in Table 1, designed for generating the site-directed mutants. Amplified products were sequenced (using an Applied Biosystems model 377 DNA sequencer) to confirm the correctness of the mutants.

For expression of recombinant wild-type and mutated r- $\alpha$ -DGL, *Escherichia coli* BL21(DE3) cells (Novagen) were transformed by electroporation with the plasmid DNAs. Positive *E. coli* BL21 (DE3) clones, shown by PCR to contain the wild-type and the mutated r- $\alpha$ -chain-thioredoxin-His<sub>6</sub> fusion constructs, were grown overnight at 37 °C in LB medium containing 100  $\mu$ g/mL of ampicillin. For inducing the expression of the recombinant fusion proteins, the cell cultures were diluted 1:10 (v/v) with medium, isopropyl- $\beta$ -D-thiogalactosidase (IPTG) was added to a final concentration of 1 mM, and the cell suspensions were incubated for another 24 h at 14 °C. Thereafter, the cells were pelleted by centrifugation, resuspended in the same volume of 20 mM sodium phosphate, 150 mM NaCl, pH 7.4, washed three times with the same buffer, and resuspended in 100 mL/L of initial cell culture of 20 mM sodium phosphate, 250 mM NaCl, 10 mM imidazole, pH 7.4. The cells were lysed by sonication (15 cycles of 15 s sonication followed by 1 min resting) in an ice bath. The lysates were centrifuged at 10000  $\times$ g for 30 min at 4 °C, and the soluble and the insoluble fractions were analyzed by SDS–15% polyacrylamide gel electrophoresis.

The  $\alpha$ -chain-thioredoxin-His<sub>6</sub> fusion wild-type and mutated r-DGL proteins were purified from the soluble fraction of positive *E. coli* BL21 (DE3) clones by affinity chromatography using an ÄKTA Basic chromatographer equipped with a 5-mL HisTrap HP column (Amersham Biosciences) equilibrated in 20 mM sodium phosphate, 250 mM NaCl, 10 mM imidazole, pH 7.4. After the absorbance at 280 nm of the flowthrough fraction reached



**Fig. 1.** Determination by equilibrium sedimentation analytical centrifugation of the apparent molecular masses of the recombinant lectins as a function of the pH of the solution. Apparent molecular masses ( $M_{app}$ ) between a pure dimer (*D*, 51 kDa) and a pure tetramer (*T*, 102 kDa) correspond to  $M_{app} = (\%D \times 51 + \%T \times 102)$  kDa. The location of the peripheral (P) and central (C) interdimeric regions are indicated in the structures of the canonical dimer and tetramer of r- $\alpha$ -DGL [1DGL, 2JDZ]. Vertical broken lines indicate the pKa3 (3.9) of aspartic acid ( $\beta$ -COOH) and the pKa (6.0) of the imidazole group of histidine.

**Table 1**

Forward (F) and reverse (R) primers used for generating the site-directed mutants of the  $\alpha$ -chains of the seed lectins of *Dioclea grandiflora* (r- $\alpha$ -DGL). Codons producing the mutations are in boldface and underlined.

Mutant	Template used	Primers (5'–3')
r- $\alpha$ -DGL H131N	r- $\alpha$ -DGL wt	F: TCA CTC CAT TTC AGC TTC <b>AAC</b> AAA TTT AGC CAA AAC CCA AAG G R: GG GTT TTG GCT AAA TTT <b>GTT</b> GAA GCT GAA ATG GAG TGA ATT TT
r- $\alpha$ -DGL H51G	r- $\alpha$ -DGL wt	F: GGG AAG GTA GGA ACT GTT <b>GGC</b> ATA AGC TAC AAC TCT GTC GC R: GC GAC AGA GTT GTA GCT TAT <b>GCC</b> AAC AGT TCC TAC CTT CCC
r- $\alpha$ -DGL R60A	r- $\alpha$ -DGL wt	F: GC TAC AAC TCT GTC GCT AAG <b>GCA</b> CTA AGT GCT GTT GTT TCT TAT TCT GG R: CC AGA ATA AGA AAC AAG AGC ACT TAG <b>TGC</b> CTT AGC GAC AGA GTT GTA CC
r- $\alpha$ -DGL H51G/R60A	r- $\alpha$ -DGL H51G	F: GC TAC AAC TCT GTC GCT AAG <b>GCA</b> CTA AGT GCT GTT GTT TCT TAT TCT GG R: CC AGA ATA AGA AAC AAG AGC ACT TAG <b>TGC</b> CTT AGC GAC AGA GTT GTA CC
r- $\alpha$ -DGL H51G/H131N	r- $\alpha$ -DGL H131N	F: GGG AAG GTA GGA ACT GTT <b>GGC</b> ATA AGC TAC AAC TCT GTC CC R: GC GAC AGA GTT GTA GCT TAT <b>GCC</b> AAC AGT TCC TAC CTT CCC

baseline, the bound material was eluted at a flow rate of 1.5 mL/min with a linear gradient of 10–500 mM imidazole for 60 min. The purified protein fractions (checked by SDS–PAGE) were pooled, dialyzed against 50 mM Tris/HCl, pH 7.4, and digested with 0.5 units of enterokinase (Invitrogen) per mg of recombinant protein. The reaction mixture was freed from enterokinase by chromatography on a 0.5-mL column of agarose-trypsin inhibitor (Sigma) equilibrated and eluted with 50 mM Tris–HCl, pH 7.4. The r- $\alpha$ -chain was separated from thioredoxin-His6 by affinity chromatography on a Sephadex G-75 column (2.0  $\times$  7.0 cm) equilibrated in 50 mM Tris–HCl, pH 7.4 containing 150 mM NaCl, 1 mM CaCl<sub>2</sub> and 1 mM MnCl<sub>2</sub>. After exhaustive washing with equilibration buffer, the bound material (carbohydrate-binding r- $\alpha$ -chain) was eluted with equilibration buffer containing 200 mM  $\alpha$ -D-glucose, dialyzed for 30 min against 0.1 M acetic acid, overnight against Milli-Q water, and lyophilized. The purity of the isolated proteins was assessed by SDS–PAGE.

## 2.2. Analytical ultracentrifugation

The apparent molecular masses of the recombinant lectins in solutions of different pH were determined at the Analytical ultracentrifugation Facility (CIB-CSIC, Madrid, Spain) by equilibrium sedimentation at 20 °C using a Beckman XL-A centrifuge with UV absorption scanner optics using an AN-50 Ti 8-hole rotor and charcoal filled eppon 6-channel centrepieces. This setup allows the simultaneous analysis of 21 different samples. The lectins were dissolved at 0.6–1.0 mg/mL in the following buffers containing 1 mM CaCl<sub>2</sub>, 1 mM MgCl<sub>2</sub> and 0.1 M NaCl: 20 mM Tris–HCl (for pH 7.5 and 8.5), 20 mM sodium citrate (for pHs 2.5, 3.5, and 4.5), and 20 mM MES (pH 5.5 or 6.5). Molar masses were determined by sedimentation-diffusion equilibrium experiments using short (approx. 3 mm) sedimentation columns. To avoid differences in apparent molecular masses due to rotor speed-dependent weighting of apparent molecular masses, all experiments were carried out at the same speed (15,000 rpm). Sedimentation equilibrium was assumed to be attained when the measured concentration profile remained unchanged for at least 12 h. The equilibrium concentration gradient for a single species is described by equation:

$$\ln c(r) - \ln c(m) = [M(1 - \bar{v}\rho)/2RT]\omega^2 \times (r^2 - m^2)$$

where  $c(r)$  and  $c(m)$  are the concentrations at radius  $r$  and at the meniscus (radius  $m$ ), respectively;  $M$  is the molar mass of the solute;  $\bar{v}$ , partial specific volume of the solute (assumed to be  $7.35 \times 10^{-3}$  kg);  $\omega$ , angular speed of the rotor (radians  $\times$  s<sup>-1</sup>), and  $R$  and  $T$ , are the gas constant (8.314 J mol<sup>-1</sup> K<sup>-1</sup>) and the temperature (in K), respectively. Apparent molar masses were determined by fitting this function to the measured radial distribution of the concentration gradient at equilibrium using the program EQASSOC [22] provided by the manufacturer. Blank buffer absorption was determined after overspeeding (50,000 rpm) to sediment all material to the bottom of the cell.

## 2.3. Crystallization, data collection and model building

Crystals of r- $\alpha$ DGL H51G/H131N were grown at 22 °C by vapour diffusion in hanging drops composed of equal volumes of protein solution [10–15 mg/mL in 10 mM HEPES pH 7.5, containing 2 mM Cl<sub>2</sub>Ca, 2 mM Cl<sub>2</sub>Mn, and 3 mM X-Man (from a 12 mM stock solution of 5-bromo-4-chloro-3-indolyl- $\alpha$ -D-mannose in DMSO), and of the reservoir solution (100 mM HEPES pH 8.0, 14% polyethylene glycol 400 used as the precipitant. Crystals were flash cryocooled in liquid nitrogen using Paratone® (Hampton Research) as cryoprotectant, and diffracted at 100 K using the

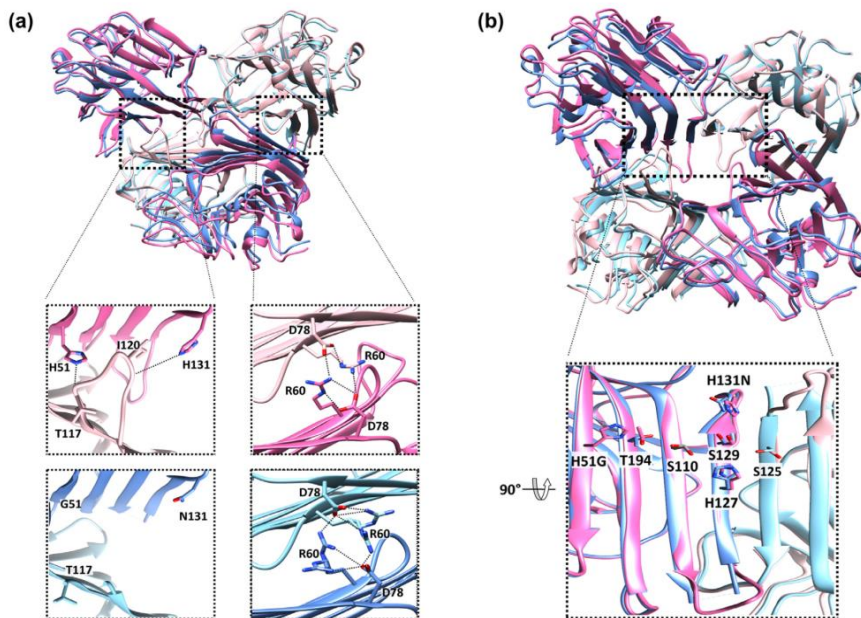
XALOC beamline at ALBA synchrotron (Barcelona, Spain). Data were processed with IMOSFLM [3], and scaled with SCALA [10]. Severe anisotropy effects on the structure refinement could be limited after ellipsoidal truncation, provided by the Diffraction Anisotropy Server [30], and anisotropic refinement. The file obtained from the server with the recommended resolution yielded better maps than any untrimmed combination and was therefore used for subsequent refinement steps. Phases were determined by molecular replacement with PHASER [21], using the structure of the wild-type lectin (1DGL) as model. The solution, consisting of one monomeric subunit in the asymmetric unit, was initially refined with rigid body. Cycles of restrained refinement were performed with PHENIX REFINER [1], combined with manual model building in COOT [9]. A final TLS (translation/libration/screw) refinement cycle using automatic group assignment [24] was also performed. Structure validation was performed with PHENIX validation tools. The atomic co-ordinates of r- $\alpha$ DGL H51G/H131N are accessible from the PDB. Database ID: 4Z8B.

## 3. Results and discussion

### 3.1. Effects of single and double mutations on the quaternary structure of *D. grandiflora* lectin assessed by analytical ultracentrifugation

Sedimentation equilibrium analytical ultracentrifugation was used for measuring the effect of single and double mutations on the distribution of molecular masses of the recombinant wild-type and mutant DGL species in solution of pH in the range 2.5–8.5. The single mutation H131N, that abolished the inter-dimeric interactions at the central cavity between H131 and I120 in an opposite monomer [32], transformed the pH-stable tetrameric wild-type r- $\alpha$ DGL into a structure exhibiting limited dimer–tetramer transition (in the range of pH 3.5–4.5) (Fig. 1). The pH-dependence of this transition correlates with the weakening of the R60–D78 interaction due to the protonation of both the  $\gamma$ -COOH of the aspartic acid (nominal pKa = 3.9) and the guanidinium side chain of the arginine (pKa = 12.5). Lending support to this assumption, *D. guianensis* lectin (1H9P), a Diocleinae lectin that possesses N at position 131, and whose tetrameric structure is not stabilized by R60–D78 inter-dimeric contacts (Fig. 2), exhibits pH-dependent dimer–tetramer association equilibrium over the pH range 3.5–6.5. To experimentally check the hypothesis that releasing the inter-dimeric hydrogen bonding contacts between R60 and D78 of an opposed monomer at the peripheral site may represent the key to generating an r- $\alpha$ DGL mutant displaying dimer–tetramer equilibrium in a wider pH range than the pKa  $\pm$  1 of the  $\gamma$ -COOH, R60 was mutated to alanine. The R60A mutation rendered a lectin species displaying dimer–tetramer equilibrium over a more extended pH range (Fig. 1); at pH  $\leq$  5.5 r- $\alpha$ DGL R60A behaved essentially as a dimer, which gradually associate into tetramers over the pH range 5.5–8.5. In an acid environment the imidazole ring (pKa = 6.0) is protonated, whereas when the pH increases, at nearly 8 the protonation of the imidazole ring is lost. Thus, the finding that abrogating R60 contacts leads to increasing dissociation of the r- $\alpha$ DGL R60A tetramer as the pH decreases from 8.5 to the pKa of histidine (6.0), indicated that, in this mutant, the dimer–tetramer transition was mostly governed by the degree of protonation of a histidine residue.

In addition to the “central” loop contacts mediated by H131, other interdimer interactions are established between H51 of monomers A and B and residues at positions 116 and 117 of monomers D and C, respectively, and between homologous regions encompassing residues 53–78 of monomers A and D and B and C at the periphery of the dimers. Position 131 is occupied by an



**Fig. 2.** Ribbon representation of  $r$ - $\alpha$ -DGL (1DGL) (pink) and  $r$ - $\alpha$ -DGL H51G/H131N (4Z8B) (blue). Panel a, superposition of the symmetry-generated tetramer of both molecules. Left close-up views highlight intermolecular interactions in 1DGL orchestrated by histidine residues H51 and H131 (central cavity) and the absence of these intermolecular contacts in the DGL H51G/H131N double mutant. The close-up views at the right show the peripheral site intermolecular contacts mediated by R60 and D78 from opposite monomers. Panel b, Comparison of side-chain rearrangements for neighboring monomers within 4Z8B crystallographic symmetry-generated tetramer (depicted in blue) compared to the 1DGL tetrameric structure (rendered in pink).

asparagine (N) in lectins exhibiting pH-dependent dimer–tetramer transition, whereas those lectins that behave as pH-independent tetrameric proteins contain histidine (H). On the other hand, H51 is absolutely conserved among Diocleinae lectins. (BLAST analysis; [7]). The interactions of H51 play a role in maintaining the conformation of the central cavity loop [15]. Single mutant  $r$ - $\alpha$ -DGL H51G was indistinguishable from the  $r$ - $\alpha$ -DGL H51N lectin in terms of the pH-dependence dimer–tetramer transition (Fig. 1), suggesting that these two histidine residues may individually play equivalent roles in the maintenance of the tetrameric structure of DGL.

Results from single  $r$ - $\alpha$ -DGL mutants clearly pointed to distinct roles for residues participating directly (H131) or indirectly (H51) in intermolecular contacts within the central cavity and the peripheral interacting region (R60) in the formation and pH dependence of the canonical type II legume lectin tetramer. Double mutants  $r$ - $\alpha$ -DGL H51G/R60A and  $r$ - $\alpha$ -DGL H51G/H131N were then generated to investigate possible cooperative effects. The results from analytical ultracentrifugation displayed in Fig. 1 clearly showed that the structural consequence of both double mutations was essentially the same: preventing the mutated dimeric lectins from associating into dimer-of-dimer structures. The effect of the concerted H51G/R60A mutations may indicate that the simultaneous impairment of contacts at the central cavity and the peripheral intermolecular regions generated dimers unable to form stable tetramers throughout the pH range studied and at the protein concentration used in the analytical ultracentrifugation experiments. In addition, the effect of the double H51/H131 mutation indicated that peripheral site interactions established by R60 are, by themselves, insufficient to maintain the structural integrity of DGL's

dimer of dimer association. The structural ground underlying the structural effects of the double histidine mutation was investigated by X-ray crystallographic analysis of  $r$ - $\alpha$ -DGL H51G/H131N.

### 3.2. Crystal structure of $r$ - $\alpha$ -DGL H51G/H131N

The crystal structure of  $r$ - $\alpha$ -DGL H51G/H131N was solved at 1.95 Å resolution (Table 2). The crystal was composed by one protein molecule in the asymmetric unit that forms a tetramer by crystal symmetry operators. Despite crystallized  $r$ - $\alpha$ -DGL H51G/H131N maintained the same overall oligomeric arrangement observed in the wild type, differences between the two tetramers were apparent (Fig. 2). Thus, when monomers A of both structures were superposed, there was a slight rotation of monomer B within the wild type  $r$ - $\alpha$ -DGL dimer, also affecting the overall tetramer oligomer (Fig. 2a and b, upper panels). The mutations did not alter the architecture of the primary monosaccharide recognition site or the interactions established between the protein and the  $\text{Ca}^{2+}$  and  $\text{Mn}^{2+}$  ions. Furthermore, the double mutation H51G/H131N did not affect the peripheral site interactions (Fig. 2b, lower panel), supporting our interpretation that the dimer–tetramer association is mostly governed by the degree of protonation of a histidine residue. However, both mutations caused a rearrangement of the side chains in the central  $\beta$ -sheet of the dimer interface with respect the wild type structure (Fig. 2b). Moreover, the double histidine mutation abolished critical 114–124 loop stabilizing interactions. As a result, in the mutant structure residues 117 to 122 are not visible in the electron density map (Fig. 2a).

**Table 2**

Data collection and refinement statistics for r- $\alpha$ DGL H51G/H131N. Statistics for the highest-resolution shell are shown in parentheses.

Wavelength (Å)	0.9795
Resolution range (Å)	45.67–1.95 (2.02–1.95)
Space group	I 2 2 2
Unit cell	62.35 67.08 108.62 90 90 90
Total reflections	223419 (23274)
Unique reflections	12687 (1679)
Multiplicity	13.2 (13.9)
Completeness (%) <sup>a</sup>	74.76 (33.1)
Mean I/sigma (I)	16.04 (5.99)
Wilson B-factor	25.71
R <sub>merge</sub> <sup>b</sup>	0.09003 (0.4526)
R <sub>meas</sub> <sup>b</sup>	0.09405
R <sub>work</sub> <sup>c</sup>	0.2014 (0.2267)
R <sub>free</sub> <sup>d</sup>	0.2433 (0.1404)
Number of non-hydrogen atoms	1893
Macromolecules	1782
Ligands	41
Water	70
Protein residues	231
RMSD <sup>e</sup> (bonds)	0.003
RMSD <sup>e</sup> (angles)	0.77
Ramachandran favored (%)	94
Ramachandran outliers (%)	0
Clashscore	1.12
Average B-factor	42.60
Macromolecules	42.10
Ligands	69.30
Solvent	40.70

<sup>a</sup> Completeness is 50.4% at 2.05 Å resolution.

<sup>b</sup>  $R_{meas} = \frac{1}{N} \sum_{hkl} [N(N-1)]^{1/2} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$ , where  $I_i(hkl)$  are the observed intensities,  $\langle I(hkl) \rangle$  are the average intensities and  $N$  is the multiplicity of reflection  $hkl$ .

<sup>c</sup>  $R_{work} = \frac{\sum_{hkl} \{|F_{obs}(hkl)| - |F_{calc}(hkl)|\}}{\sum_{hkl} |F_{obs}(hkl)|}$ , where  $F_{obs}(hkl)$  and  $F_{calc}(hkl)$  are the structure factors observed and calculated, respectively.

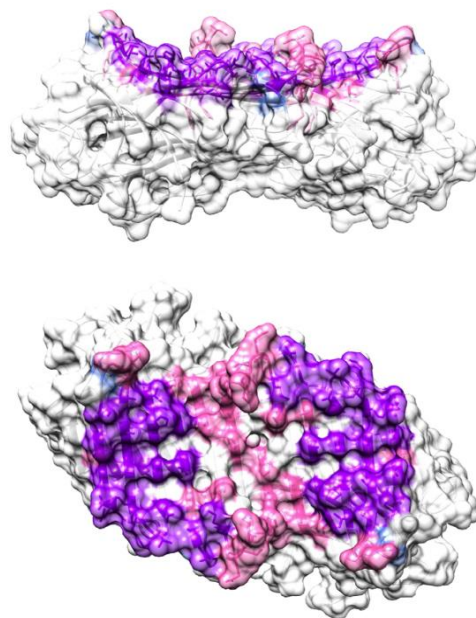
<sup>d</sup>  $R_{free}$  corresponds to Rfactor calculated using 5% of the total reflections selected randomly and excluded during refinement.

<sup>e</sup> RMSD is the root mean square deviation.

Comparison of the interactions between the *bona fide* r- $\alpha$ DGL in-solution tetrameric structure (1DGL) and the crystallographic symmetry-generated r- $\alpha$ DGL H51G/H131N tetramer (4Z8B) revealed a decreased number of interactions in the central area of the double His mutant lectin (Fig. 3). Furthermore, analysis of the protein assembly stability with PISA ([http://www.ebi.ac.uk/msd-srv/prot\\_int/pistart.html](http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html)) [19], suggested a greatly reduced complexation significance score (CSS) for crystalline r- $\alpha$ DGL H51G/H131N double mutant (0.011) compared to the wild-type r- $\alpha$ DGL tetramer (1DGL) (0.336). The calculated  $\Delta G^{diss}$  for these structures were 19.3 and  $-1.2$  kcal/mol, respectively, and the calculated interdimer surface buried in r- $\alpha$ DGL and r- $\alpha$ DGL H51G/H131N crystallographic tetramers was also significantly different: 13660 Å<sup>2</sup> and 10360 Å<sup>2</sup>, respectively. In agreement with Bouckaert et al. [5], who reported that in crystals grown at pH 5, where the lectin is a dimer in solution, ConA dimers associated into a tetramer-like structure that resembled the native tetrameric assembly of ConA, but with a drastically smaller number of interdimeric interactions, Table 3 clearly shows that compared to r- $\alpha$ DGL, the double His mutant exhibits drastically reduced buried surface and number of residues involved in tetrameric assembling contacts. As a whole, the analytical centrifugation, crystallographic, and computational data consistently support the view that although the 'weakened' tetrameric assembly generated in the r- $\alpha$ DGL H51G/H131N crystal may represent the most favorable association generated in the crystallization process, r- $\alpha$ DGL H51G/H131N might be considered a stable dimeric structure both in the crystal lattice and in solution in the pH range 2.5–8.5.

#### 4. Concluding remarks

More than three decades ago, Huet and Claverie [17] studied the kinetics of the reversible dimer–tetramer equilibrium of Con A, and Senear and Teller [28] reported the effects of temperature and pH by analytical centrifugation, concluding that the self-association of the lectin could be better described as a dimer to tetramer transition across the pH range 5.5–7.5. From the ionization thermodynamic parameters, the authors concluded that tetramerization was entropically driven and that “the ionization state of a histidine side chain on each subunit, either His51 or His121, governed the dimer–tetramer equilibrium of the lectin” [28]. Continuing our research on the structural basis of the pH-dependent dimer–tetramer transition of Diocleinae lectins [32,15,23], we have now conducted equilibrium sedimentation studies on single and double histidine mutants and solved the crystal structure of recombinant DGL H51G/H131N mutant. The data shown here confirm the importance of histidine residues 51 and 131 in the pH-dependence dimer–tetramer transition. The finding that only mutation of both histidines, but not the mutation of each His separately, abolished the formation of the Diocleinae canonical dimer of dimer association lends support to a structural model in which H131 plays a crucial role in networking loop 114–125 residues from all four subunits at the central cavity of the tetramer, whereas H51 contributes to stabilizing the canonical tetrameric structure by maintaining the central cavity loops in a proper spatial



**Fig. 3.** Surface representation of the 1DGL dimer indicating the aminoacids involved in the tetrameric ensemble. Residues involved in interactions common for both wild-type and double mutant crystal structures (1DGL and 4Z8B, respectively) are shown in purple. Interacting residues on 1DGL non-participating in the double mutant symmetry-generated tetramer are depicted in pink. The minor blue surface indicates new contacts established in the 4Z8B crystallographic tetramer not present in the 1DGL structure.

**Table 3**  
Buried surface (Å<sup>2</sup>) for residues involved in tetrameric assembling contacts.

	4Z8B	1DGL
THR49	5.52	36.98
GLY51	3.68	
HIS51		78.91
SER53	25.9	27.32
ASN55	32.72	37.25
VAL57	89.57	84.69
ALA58	55.93	59.79
LYS59	4.6	0.92
ARG60	50.45	84.15
SER62	33.97	37.52
ALA63	2.94	0.76
VAL64	50.51	40.16
SER66	23.91	29.58
TYR67		7.67
SER68		25.21
GLY69		1.89
SER70	2.33	
SER72	17.32	8.94
THR74	28.08	19.34
VAL75		0.14
SER76	12.61	9.56
ASP78	3.93	10.55
SER108		15.16
TRP109		0.65
SER110		26.61
LYS114		11.7
LYS116	32.43	37.21
THR117		9.73
ASN118		88.8
SER119		32.97
ILE120		129.33
ALA121		2.68
HIS127		1.09
SER129		2.7
PHE130		1.23
HIS131		32.12
ALA186	0.25	
VAL187	79.17	57.05
VAL188	11.21	24.77
ASP192	1.23	25.29
THR194	24.06	31.22
THR196		9.34

orientation to make H131-mediated interdimer contacts. Our results also indicate that interactions at the peripheral interdimeric contact sites, orchestrated by residues N55, R60, and D78 from opposite monomers, may play only a secondary role in conferring pH dependency to the canonical cross-wise Brinda's type-II legume lectin tetramer.

#### Acknowledgments

This study was supported by grants BFU2010-17373 from the Ministerio de Ciencia e Innovación, Madrid (JJC); BFU2013-42833-P (JJC) and SAF2012-31405 (JB) from the Ministerio de Economía y Competitividad, Madrid, Spain; and PROMETEO/2010/005 (JJC) and PROMETEO/2012/061 (JB) from the Generalitat Valenciana. SZC received a FPI (BES2010-030116) fellowship from Ministerio de Ciencia e Innovación, Madrid Spain. X-ray data collection experiments were performed at XALOC beamline at ALBA Synchrotron with the collaboration of ALBA staff.

#### References

- [1] Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J.,

- Moriarty, T.C., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C. and Zwart, P.H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* 66, 213–221.
- [2] Andre, S., Renaudet, O., Bossu, I., Dumy, P. and Gabius, H.J. (2011) Cyclic neoglycodeceptides: how to increase their inhibitory activity and selectivity on lectin/toxin binding to a glycoprotein and cells. *J. Peptide Sci.* 17, 427–437.
- [3] Batty, 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. D Biol. Crystallogr.* 67, 271–281.
- [4] Bezerra, M.J.B., Rodrigues, N.V.F.C., Pires, A.F., Bezerra, G.A., Nobre, C.B., Alencar, K.L.L., Soares, P.M.G., Nascimento, K.S., Nagano, C.S., Martins, J.L., Gruber, K., Sampaio, A.H., Delatorre, P., Rocha, B.A.M., Assrey, A.M.S. and Cavada, B.S. (2013) Crystal structure of *Dioclea violacea* lectin and a comparative study of vasorelaxant properties with *Dioclea rostrata* lectin. *Int. J. Biochem. Cell Biol.* 45, 807–815.
- [5] Bouckaert, J., Loris, R., Wyns, L. (1996) Change of the crystal space group and of the oligomeric structure of concanavalin A as a function of metal ion binding and pH. In: Van Driessche, E., Rougé, P., Beeckmans, S., Bög-Hansen, T.C. (eds), *Lectins. Biology, Biochemistry, Clinical Biochemistry*, vol. 11, Textop, Hellerup, Denmark, 50–60.
- [6] Brinda, K.V., Mitra, N., Suroliya, A. and Vishveshwara, S. (2004) Determinants of quaternary association in legume lectins. *Protein Sci.* 13, 1735–1749.
- [7] Calvete, J.J., Thole, H.H., Raida, M., Urbanke, C., Romero, A., Grangeiro, T.B., Ramos, M.V., Almeida da Rocha, I.M., Guimarães, F.N. and Cavada, B.S. (1999) Molecular characterization and crystallization of *Diocleinae* lectins. *Biochim. Biophys. Acta* 1430, 367–375.
- [8] Edelman, G.M. and Wang, J.L. (1978) Binding and functional properties of concanavalin A and its derivatives. III. Interactions with indoleacetic acid and other hydrophobic ligands. *J. Biol. Chem.* 253, 3016–3022.
- [9] Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60, 2126–2132.
- [10] Evans, P.R. (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. D Biol. Crystallogr.* 67, 282–292.
- [11] Gabius, H.-J. and Gabius, S., Eds., (1997). *Glycoscience. Status and Perspectives*, Chapman & Hall, Weinheim (Germany).
- [12] Gabius, H.-J., Siebert, H.C., André, S., Jiménez-Barbero, J. and Rüdiger, H. (2004) Chemical biology of the sugar code. *ChemBioChem* 5, 740–764.
- [13] Gabius, H.J., Andre, S., Jimenez-Barbero, J., Romero, A. and Solis, D. (2011) From lectin structure to functional glycomics: principles of the sugar code. *Trends Biochem. Sci.* 36, 298–313.
- [14] Gallego del Sol, F., Vania, M., Nagano, C.S., Moreno, F.B.M.B., Sampaio, A.H., Grangeiro, T.B., Cavada, B.S., Calvete, J.J., 2004. Methods in proteome and protein analysis. In: Kamp, R.M., Calvete, J.J., Choli-Papadopoulou, T. (eds.), Springer-Verlag, Berlin, Heidelberg, 81–91.
- [15] Gallego del Sol, F., Cavada, B.S. and Calvete, J.J. (2007) Crystal structures of *Cratylia floribunda* seed lectin at acidic and basic pHs. Insights into the structural basis of the pH-dependent dimer–tetramer transition. *J. Struct. Biol.* 158, 1–9.
- [16] Gout, E., Bligny, R. and Douce, R. (1992) Regulation of intracellular pH values in higher plant cells. Carbon-13 and phosphorus-31 nuclear magnetic resonance studies. *J. Biol. Chem.* 267, 13903–13909.
- [17] Huet, M. and Claverie, J.M. (1978) Sedimentation studies of the reversible dimer–tetramer transition kinetics of concanavalin A. *Biochemistry* 17, 236–241.
- [18] Kamiya, Y., Yagi-Utsumi, M., Yagi, H. and Kato, K. (2011) Structural and molecular basis of carbohydrate–protein interaction systems as potential therapeutic targets. *Curr. Pharm. Des.* 17, 1672–1684.
- [19] Krissinel, E. and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 372, 774–797.
- [20] Loris, R. (2002) Principles of structures of animal and plant lectins. *Biochim. Biophys. Acta* 1572, 198–208.
- [21] McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C. and Read, R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* 40, 658–674.
- [22] Minton, A.P., 1994. Modern analytical ultracentrifugation. In: Schuster, M., Laue, T.M. (eds.), Birkhauser, Boston (USA), 81–93.
- [23] Nagano, C.S., Calvete, J.J., Barettoni, D., Pérez, A., Cavada, B.S. and Sanz, L. (2008) Insights into the structural basis of the pH-dependent dimer–tetramer equilibrium through crystallographic analysis of recombinant *Diocleinae* lectins. *Biochem. J.* 409, 417–428.
- [24] Painter, J. and Merritt, E.A. (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D Biol. Crystallogr.* 62, 439–450.
- [25] Reece, G.N., Becker, J.W. and Edelman, G.M. (1975) The covalent and three-dimensional structure of concanavalin A. IV. Atomic coordinates, hydrogen bonding, and quaternary structure. *J. Biol. Chem.* 250, 1525–1547.
- [26] Roberts, D.D. and Goldstein, I.J. (1983) Adenine binding sites of the lectin from lima beans (*Phaseolus lunatus*). *J. Biol. Chem.* 258, 13820–13824.
- [27] Rozwarski, D.A., Swami, B.M., Brewer, C.F. and Sacchettini, J.C. (1998) Crystal structure of the lectin from *Dioclea grandiflora* complexed with core trimannoside of asparagine-linked carbohydrates. *J. Biol. Chem.* 273, 32818–32825.



- [28] Senechal, D.F. and Teller, D.C. (1981) Thermodynamics of concanavalin A dimer-tetramer self-association: sedimentation equilibrium studies. *Biochemistry* 20, 3076–3083.
- [29] Senechal, D.F. and Teller, D.C. (1981) Effects of saccharide and salt binding on dimer-tetramer equilibrium of concanavalin A. *Biochemistry* 20, 3083–3091.
- [30] Strong, M., Sawaya, M.R., Wang, S., Phillips, M., Cascio, D. and Eisenberg, D. (2006) Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *PNAS USA* 103, 8060–8065.
- [31] Vasta, G.R. and Ahmed, H. (2008) *Animal Lectins: A Functional View*. CRC Press, Boca Raton, FL.
- [32] Wah, D.A., Romero, A., Gallego del Sol, F., Cavada, B.S., Ramos, M.V., Grangeiro, T.B., Sampaio, A.H. and Calvete, J.J. (2001) Crystal structure of native and Cd/Cd-substituted *Dioclea guianensis* seed lectin. A novel manganese-binding site and structural basis of dimer-tetramer association. *J. Mol. Biol.* 310, 885–894.
- [33] Yu, Q., Tang, C. and Kuo, J. (2000) A critical review on methods to measure apoplastic pH in plants. *Plant Soil* 219, 29–40.