



HAL
open science

Structural studies of virulence factors from the bacterium *Helicobacter pylori*

Tommaso Tosi

► **To cite this version:**

Tommaso Tosi. Structural studies of virulence factors from the bacterium *Helicobacter pylori*. Biochemistry [q-bio.BM]. European Synchrotron Radiation Facility, 2010. English. NNT: . tel-00521213

HAL Id: tel-00521213

<https://theses.hal.science/tel-00521213>

Submitted on 26 Sep 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Université de Grenoble
Ecole Doctorale Chimie et Sciences du Vivant

THESE

Pour l'obtention du diplôme de
DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Discipline : Biologie Structurale et Nanobiologie
Présentée et soutenue par

TOMMASO TOSI

ÉTUDES STRUCTURALES DE FACTEURS DE VIRULENCE DE LA BACTÉRIE *HELICOBACTER PYLORI*

Jury:

Prof. Patrice GOUET	Rapporteur
Dr. Wolfgang FISCHER	Rapporteur
Dr. Andréa DESSEN	Examineur
Dr. Sean MC SWEENEY	Examineur
Prof. Giuseppe ZANOTTI	Examineur
Prof. Serge PEREZ	Examineur
Dr. Laurent TERRADOT	Examineur

Travaux effectués au Laboratoire de Cristallographie Macromoléculaire,



European Synchrotron Radiation Facility

Abbreviations

“A”

ABC: ATP binding cassette
Abl: Cellular Abelson kinase
Acr: Acriflavin resistance pathway
aIF2: translation initiator factor 2
APC: Antigen presenting cell
AT: Autotransporter
ATM: Ataxia-Telangiectasia mutated kinase
ATP: Adenosine tri-phosphate
Arp2/3: Actin-Related Proteins 2/3
Ars: Acidic response signalling

“B”

BabA: B-antigen binding protein A
BadA: Bartonella adhesin A
BCL-2: B-cell CLL/lymphoma 2
BopE: Burkholderia outer protein E
BSA: Bovine serum albumin

“C”

Cag: Cytotoxin associated gene
CCD: Charged coupled device
CDC: Cholesterol-dependent cytolysin
CDT: Cytotoxic distending toxin
CFP-10: Culture filtrate protein-10
CHAPS: 3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
Cif: Cycle inhibiting factor
CNF1: Cytotoxic necrotizing factor 1
COX-2: Cyclooxygenase 2
Csk: C-terminal Src kinase
CT: Cholera toxin

“D”

DBD: DNA-binding domain
DDM: n-dodecyl beta-D-maltoside
DNA: Deoxyribonucleic acid
Dot: Defective organelle trafficking

“E”

EDTA: Ethylenediaminetetraacetic acid
EPEC: Enteropathogenic Escherichia coli
EHEC: Enterohemorrhagic Escherichia coli
EM: Electron Microscopy
Eps: Extracellular protein secretion system

Erk: Extracellular signal-regulated kinase
ESAT-6: Early secreted antigen target 6
ExoS: Pseudomonas exotoxin S

“F”

FAK: Focal adhesion kinase
Fur: Ferric uptake regulator

“G”

Gab: Grb2-binder adaptor protein
GAP: GTP activating protein
GDP: Guanosine di-phosphate
GEF: Guanine nucleotide exchange factor
Grb2: Growth factor receptor-bound protein 2
GTP: Guanosine tri-phosphate

“H”

Hcp: Hemolysin conjugated protein
Hia: Haemophilus intercellular adhesion protein
Hly: Hemolysin secretion system
Hop: Helicobacter outer membrane protein
Hor: Hop-related protein
HP-NAP: Helicobacter pylori neutrophil-activating protein

“I”

Icm: Intracellular multiplication
IL-1: Interleukin 1
IL-2: Interleukin 2
IL-6: Interleukin 6
IL-8: Interleukin 8
IL-12: Interleukin 12
IL-23: Interleukin 23
IFN- γ : Interferon γ
Imp α 5: Human Importin α 5
IMR: Inner membrane region
IPTG: Isopropyl- β -D-thio-galactopyranoside

“K”

KatAB: Catalase AB

“L”

LPS: Lypopolysaccharide
Le^b: Lewis-b antigen
Le^x: Lewis-X antigen

“M”

Mac: Macrolides resistance pathway
MACPF: Membrane Attack Complex/Perforin
MAD: Multiple anomalous dispersion
MALT: Mucosa-associated lymphoid tissue
MAPK: Mitogen-activated protein kinase
MARK: Mitogen-activated receptor kinase
MDIA: Mammalian Diaphanous protein
MFP: Membrane fusion protein
MHC-II: Major histocompatibility complex II
MIR: Multiple isomorphous replacement
MIRAS: Multiple isomorphous replacement with anomalous scattering
MME: Monomethylether

“N”

NBT/BCIP: Nitro-Blue tetrazolium_5-Bromo-4-Chloro-3-Indolyl-Phosphate
NC: Needle Complex
NCK: native creatine kinase
NFAT: Nuclear factor of activated T-cells
NF- κ B: Nuclear factor κ B
NikR: Nickel response regulator
NTA: Nitrilotriacetic acid

“O”

Oca: Oligomeric coiled-coil Adhesin
OD: Optical density
OG: n-octyl- β -D-glucoside

“P”

PAI: Pathogenicity island
PAK: p21 activated kinase
Peg: Polyethylenglycole
PFT: Pore-forming toxin
pIgR: Polymeric Immunoglobulin Receptor
PI₃: phosphatidil-inositol 1,4,5-triphosphate
PI₂: phosphatidil-inositol 4,5-bisphosphate
PTPN11: Protein tyrosine phosphatase N11

“R”

Rb: Retinoblastoma protein
ROS: Reactive Oxygen Species

“S”

SabA: Sialyl-X antigen binding protein A
SAD: Single anomalous dispersion
Sag: Superantigen

SAM: Sterile alpha motif
Sec: Secretory pathway
SHP-1: Src homology phosphatase-1
SHP-2: Src homology phosphatase-2
SodB: Superoxide Dismutase B
SodC: Superoxide Dismutase C
Sos1: Sons of sevenless 1
SPase I: Signal peptidase I
SppA: Signal peptide peptidase A
Src: Sarcoma associated tyrosine kinase

“T”

T1SS: Type 1 Secretion System
T2SS: Type 2 Secretion System
T3SS: Type 3 Secretion System
T4SS: Type 4 Secretion System
T5SS: Type 5 Secretion System
T6SS: Type 6 Secretion System
T7SS: Type 7 Secretion System
TAE : Tris/Acetate/EDTA
Tarp: Translocated actin-recruiting phosphoprotein
TD: Transcriptional domain
Tip α : Tumor necrosis factor α -inducing protein
Tir: Translocated Intimin receptor
TMH: Transmembrane hairpin region
TNF- α : Tumor Necrosis Factor α
TPA: 12-O-tetradecanoylphorbol-13-acetate
TPS: Two-partners system
Tris: Trishydroxymethylaminomethane

“U”

Ure: urease cluster

“V”

VacA: Vacuolating cytotoxin A
Vav2: vav oncogene 2
VgrG: Valine-glycine repeat protein G

“W”

WASP: Wiskott-Aldritch syndrome protein
WAVE2: WASP-family *protein* member 2
WHO: World health organization

“Y”

Y2H: Yeast two-hybrid assay
YopE : Yersinia outer protein E

Acknowledgements

Before thanking all the people that worked with me and helped me throughout my PhD, I would like to thank the members of the Jury, Wolfgang Fischer, Patrice Gouet, Giuseppe Zanotti and Andréa Dessen, for doing me the great honor to read this manuscript and judge my work.

Very well, where should I start? Got so many people to thank, so many people that helped me a lot throughout my entire PhD...well, I'll try to include them all here, ok? I'll do my best!

First things first, I want to thank my scientific supervisor, Laurent, for many things: first, for following my adventures in this terrific (or terrifying?) world, that is the PhD, with immense attention, seriousness, interest, then for teaching me not to give up if our bloody proteins do not seem to want to crystallize by any means...a good scientist also knows how to pick up a cool backup project that can work! We always do things with style!

Ulrike, I can not forget how much you've helped me with all those CagA fragments! Thanks a lot, for this and all the times you've backed me up in the lab!

Of all the people that I have worked with in the lab, Samira deserves a special acknowledgement! Starting every morning with you yelling "TITIIIIIIIIIIH!" at me has no price! Really, thank you for your sympathy and the laughs we had in the past three years.

I'd like to thank also the other PhD students I've shared the joys and pains of this work with: Sofia, Amandine, Joana, Simone, Rita, Erika, Hayretin. Thanks a lot for your help, your support and your sympathy, that made of our lab (and the one nearby) a really cool place to work!

Then I want to thank Dr Darren J. Hart and Dr. Joanna Timmins for all the useful discussions and suggestions that helped me a lot during my PhD.

Big thanks to Dr. Sean McSweeney, Dr. Gordon Leonard, Dr. Sine Larsen and Prof. Serge Perez for all the help they gave me (with this thesis and with the university :)), their sympathy and support.

Cyril, thanks a lot for your helpful suggestions, contributions and scientific discussions on my projects. We'll see each other soon, so be prepared mwahahahaahah!

Time to thank all the post-docs and scientists I've bothered these years: Olivia, Tobias, Jens, Emanuela, Ganesh, Mats, Stephan, Franck...yep, that should be all of'em, I hope! Thanks a lot guys, you're cool!

A huge THANKS to Karina, for helping me so much in the lab, her support and sympathy.

Thanks also to my Italian colleagues, Dario and Daniele, for the laughs and all the joyful times spent together.

A growling, grunting, pig-squealing THANK YOU to my fellow metalheads, Sylvain and Jean-Laurent! Playing in a band with you is cool! ROOOAAAARRR!!!

A BIG, HUGE, COSMIC thanks to Gianluca and Elisa for the help, outside and inside the ESRF, and for all the wonderful times spent together.

Thanks mum and dad, for being always there when I needed you!

Summary

Introduction

I.1	<i>Helicobacter pylori</i>: general features and epidemiology	3
I.1.1	Adaption to the gastric acidic environment in bacteria	4
I.1.2	Adaption to the gastric acidic environment in <i>H. pylori</i>	5
I.1.2.1	<i>Motility</i>	5
I.1.2.2	<i>Urease</i>	5
I.1.2.3	<i>The amidases AmiE and AmiF and the Arginase RocF</i>	6
I.1.2.4	<i>The two-component system Ars</i>	6
I.1.2.5	<i>NikR and Fur</i>	6
I.2	Bacterial virulence: roles of virulence factors in bacterial survival and host cell colonization	8
I.2.1	Bacterial adhesion to the host cell	8
I.2.2	<i>Helicobacter pylori</i> adhesins	9
I.3	Toxins: complex evolutionary tools to increase bacterial fitness	10
I.3.1	Bacterial toxins targeting the host cell cycle	11
I.3.1.1	<i>The Cytolethal Distending Toxin (CDT)</i>	11
I.3.1.2	<i>The Cycle Inhibiting Factor (Cif)</i>	12
I.3.2	Bacterial proteins targeting Rho GTPases	13
I.3.2.1	<i>Rho GTPases and their role in actin polymerization</i>	13
I.3.2.2	<i>The Cytotoxic Necrotizing factor (CNF) 1</i>	14
I.3.2.3	<i>YopE and ExoS</i>	15
I.3.2.4	<i>BopE</i>	16
I.3.3	Bacterial toxins targeting the host immune system	16
I.3.3.1	<i>The Superantigens (Sags)</i>	16
I.3.3.2	<i>Secreted proteases</i>	17
I.3.3.3	<i>OspB and OspF</i>	18
I.3.4	The Pore-Formin Toxins (PFTs)	18
I.3.5	Phosphorylated bacterial toxins	20
I.3.5.1	<i>The Translocated Intimin Receptor</i>	20
I.3.5.2	<i>The Translocated Actin-recruiting Phosphoprotein (Tarp)</i>	21
I.3.5.3	<i>LspA1 and LspA2</i>	21
I.3.5.4	<i>AnkA</i>	21
I.4	Delivery of toxins towards the host cells: roles of the bacterial secretion systems	22
I.4.1	The general secretory pathway (Sec)	22
I.4.2	The Type 2 Secretion System (T2SS)	23
I.4.3	The Type 1 Secretion System (T1SS)	25
I.4.4	The Type 3 Secretion System (T3SS)	26
I.4.5	The Type 4 Secretion system (T4SS)	28
I.4.5.1	<i>The VirB/D4 T4ASS of Agrobacterium tumefaciens</i>	28
I.4.5.2	<i>The Dot/Icm system from Legionella Pneumophila (T4BSS)</i>	30
I.4.6	The Type 5 Secretion System (T5SS)	31

I.4.7	Type 6 (T6SS) and Type 7 (T7SS) Secretion Systems	32
I.5	Virulence factors of <i>H. pylori</i>	32
I.5.1	The Vacuolating Cytotoxin A (VacA)	33
I.5.2	The Neutrophil-Activating Protein HP-NAP	35
I.5.3	The <i>cag</i> -PAI T4ASS	35
I.5.4	The <i>cytotoxin associate gene A</i> (CagA) protein	39
I.5.4.1	<i>CagA</i> tyrosine phosphorylation at EPIYA motifs	40
I.5.4.2	Activation of SHP-2 phosphatase by phosphorylated CagA	41
I.5.4.3	Other phosphorylated CagA binding partners	42
I.5.4.4	Phosphorylation-independent interactions with host proteins	43
I.5.4.5	<i>CagA</i> correlation with gastric cancer	45
	Overview on the Thesis	46
	<u>Chapter 1: Biochemical and Structural studies of the <i>H. pylori</i> CagA Protein-Protein Interaction Map</u>	
	Résumé	51
1.1	The <i>Helicobacter pylori</i> protein-protein interaction map	53
1.2	The protein interaction map of CagA	54
1.2.1	Hp0496 (HpYbgC)	55
1.2.2	RibA (HP0802)	56
1.2.3	HP0739 (EstV): The first <i>Helicobacter pylori</i> Type V Lipase	57
1.2.4	The high affinity iron transporter FeoB (HP0687)	58
1.2.5	The putative proteins HP0249, HP0383, HP0431, HP1015	59
1.3	Cloning, expression and purification of the interacting partners	60
1.3.1	Cloning of the interacting partners	60
1.3.2	Expression of the target proteins	61
1.4	CagA-target protein interaction studies	63
1.5	Conclusion and discussion	64
	<u>Chapter 2: Expression of <i>Helicobacter pylori</i> CagA soluble domains by library-based construct screening</u>	
	Resumé	69
2.1	Construction of soluble fragments of CagA: overview of the project	70
2.2	Results	73
2.3	Conclusion and discussion	73

Publication: Angelini A., Tosi T., et al., *Expression of Helicobacter pylori CagA soluble domains by library-based construct screening*. FEBS J., 2009 75

Chapter 3: Biochemical and Structural characterization of a 33 kDa C-terminal fragment of CagA

Resumé	89
3.1 Expression and purification of CagA^{C33}	91
3.1.1 Construction of a CagA ^{C33} fragment suitable for biochemical and structural studies	91
3.1.2 Expression tests of His ₆ -CagA ^{C33} with pHAR3011-37ΔBio in various <i>E. coli</i> strains	92
3.1.3 Large scale expression and purification of His ₆ -CagA ^{C33}	93
3.2 Oligomeric state of CagA^{C33}	95
3.2.1 Study of the oligomeric state of His ₆ -CagA ^{C33} by gel-filtration	95
3.2.2 Analysis of His ₆ -CagA ^{C33} multimers by Dynamic Light scattering (DLS) and comparison with the gel-filtration experiment	97
3.2.3 Role of detergents in the stabilization of His ₆ -CagA ^{C33} tetramer	98
3.2.4 Involvement of Cysteine 1164 in His ₆ -CagA ^{C33} multimerization	99
3.3 Electron Microscopy analysis of His₆-CagA^{C33}	100
3.3.1 Sample preparation	101
3.3.2 Image analysis	101
3.3.3 Cross-linking experiments	102
3.4 Interaction of His₆-CagA^{C33} with the <i>H. pylori</i> protein CagF	104
3.4.1 His ₆ -CagA ^{C33} /CagF interaction by pull-down assay	104
3.4.2 His ₆ -CagA ^{C33} /CagF interaction by limited proteolysis	105
3.5 The domains within His₆-CagA^{C33}: identification of CagA^{C11}	107
3.5.1 Cloning, expression and purification of CagA ^{C11}	108
3.5.2 Dimerization of His ₆ -CagA ^{C11}	109
3.6 <i>In vitro</i> phosphorylation of CagA^{C33} and CagA^{C11}	110
3.6.1 Phosphorylation of His ₆ -CagA ^{C33}	110
3.6.2 Phosphorylation of His ₆ -CagA ^{C11}	111
3.7 Interaction studies of His₆-CagA^{C33} with the SH2 domain of the v-Src kinase	111
3.7.1 SH2 domains: general features and functions	112
3.7.2 Cloning, expression and purification of v-Src SH2 domain	113
3.7.3 Pull-down assay between cleaved CagA ^{C33} and His ₆ -SH2	114
3.8 Preliminary biophysical characterization of the His₆-CagA^{C11}/v-SH2 interaction	115
3.8.1 Construction, expression and purification of His ₆ -CagA ^{C11} mutants	115
3.8.2 Phosphorylation of His ₆ -CagA ^{C11} mutants with c-Src and c-Abl	116

3.9	Characterization of His₆-CagA^{C11}/v-SH2 interaction by ITC	117
3.10	Analysis of CagA^{C11}/SH2 interaction by SPR	119
3.11	Interaction study between His₆-CagA^{C11} and the SH2 domains of the human phosphatase SHP-2	121
3.11.1	Cloning, expression and purification of His ₆ -SHP-2, His ₆ -NSH2 and His ₆ -TSH2	122
3.11.2	Gel-filtration chromatography of ΔAΔB/SHP-2, ΔAΔB/NSH2 and ΔAΔB/TSH2 complexes	123
3.11.3	<i>In vitro</i> dephosphorylation of CagA by the phosphatase SHP-2	125
3.12	Crystallization trials with CagA^{C33}, CagA^{C11} and complexes	126
3.13	Conclusion and discussion	128
3.13.1	CagA ^{C33} oligomerization and architecture	128
3.13.2	CagA ^{C33} interaction with CagF	129
3.13.3	CagA ^{C33} motif requirements for phosphorylation	130
3.13.4	CagA interaction with v-SH2	131
3.13.5	Interaction with the SH2 domains of the human SHP-2 phosphatase	131

Chapter 4: Crystal structures of the *H. pylori* Tumor Necrosis Factor α -inducing protein Tip α

Résumé		135
4.1	The Tumor-Necrosis-factor α (TNF-α) inducing protein Tipα: a novel carcinogenic factor of <i>H. pylori</i>	137
4.1.1	The discovery of HP0596 as a TNF- α inducing protein	137
4.1.2	Correlation between Tip α and gastric carcinogenesis	137
4.1.3	General features of Tip α	139
4.2	Cloning, expression and purification of Tipα^{M21}, Tipα^{C25} and Tipα^{C27}	140
4.3	Crystallization of Tipα^{M21}, Tipα^{C25} and Tipα²⁷	141
4.4	Identification of a stable Tipα fragment starting at N34 (Tipα^{N34})	143
4.5	Crystallization of Tipα^{N34}	144
4.6	Structure determination of Tipα^{N34} Form I	145
4.6.1	Expression and purification of the Seleno-Methionine Tip α ^{C27} (SeTip α)	146
4.6.2	Data collection, structure determination and refinement	146
4.6.3	Structure of Tip α ^{N34}	147
4.6.4	Charge distribution on the surface of Tip α ^{N34}	149
4.7	Assay of the DNA-binding activity of Tipα	150
4.8	<i>In vitro</i> phosphorylation of Tipα^{M21} and Tipα^{N34}	151

4.8.1 Preliminary study of the Tip α ^{M21} phosphorylation motifs	153
4.9 Conclusion and discussion	154
4.9.1 Structure and oligomerization of Tip α	154
4.9.2 DNA binding activity	155
Publication: Tosi T., Cioci G., et al., <i>Structures of the Tumor Necrosis Factor α-inducing protein Tipα: a novel virulence factor from <i>H. pylori</i></i> . FEBS L., 2009	157
<u>General Conclusions and perspectives</u>	
CagA protein-protein interaction map	167
Application of the ESPRIT methodology to CagA	168
CagA oligomerization and architecture	168
CagA phosphorylation-dependent interactions	169
Tip α , a novel bacterial toxin	171
<u>Appendix</u>	
A.1: Overview of the vector pET151/D- TOPO	175
A.2: Dynamic Light Scattering (DLS)	178
A.3: Isothermal Titration Calorimetry (ITC)	179
A.4: Surface Plasmon Resonance (SPR)	181
A.5: Overview of protein crystallization, structure determination and refinement	183
A.5.1 Introduction	183
A.5.2 Protein crystallization	183
A.5.3 X-ray diffraction	186
A.5.4 The phase problem in structure solution	188
A.5.4.1 <i>Molecular Replacement</i>	189
A.5.4.2 <i>Isomorphous replacement</i>	190
A.5.4.3 <i>Single and Multiple Wavelength Anomalous Dispersion (SAD/MAD)</i>	191
A.5.5 Phase Improvement	194
A.5.6 Model building and refinement	195
A.5.7 Data collection	195
A.5.8 Macromolecular crystallography and synchrotron radiation	197
<u>Bibliography</u>	199

Introduction

I.1 *Helicobacter pylori*: general features and epidemiology

In 1982, in Perth (Australia), Robin Warren managed to isolate a *Campylobacter*-like bacterium from specimens of gastric biopsies belonging to gastritis-affected patients and named it *Campylobacter pylori* (Warren, 1984). Subsequently, Barry Marshall managed to cultivate it by accidentally leaving Petri dishes in the incubator for five days instead of two, as required by the established protocol for *Campylobacter* cultivation. To differentiate this slow-growing bacterium from the *Campylobacter* genera it was given the name “*Helicobacter pylori*” (Marshall & Warren, 1984).

It was still unknown whether this bacterium was the cause of gastritis. To address that, Barry Marshall and his colleague Arthur Morris voluntarily ingested a very concentrated suspension of *H. pylori*. They both developed signs of acute gastritis and biopsies revealed the presence

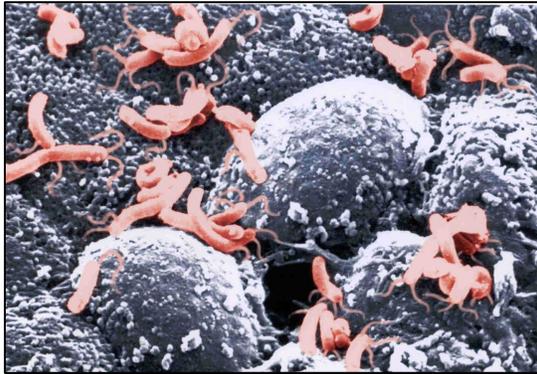


Figure I.1: microphotograph in false colors of numerous *H. pylori* bacilli laying on a gastric epithelium.

Taken from the website of Prof. Steven Blanke, Department of Microbiology, University of Illinois.

(<http://www.life.illinois.edu/blanke/>)

of the bacterium in their stomach. Both the infection and the gastritis disappeared after antibiotic treatment, finally demonstrating that *H. pylori* is the etiologic agent of acute gastritis and this pathology must be considered as an infectious disease. In 2005, Marshall and Warren received the Noble Prize for this discovery (Hellstrom, 2006).

H. pylori (Figure I.1) is spiral-shaped, Gram-negative and microaerophilic. It is fairly small (its dimensions are roughly $3.5 \mu\text{m} \times 0.5 \mu\text{m}$) and is transmitted through the oro-fecal and, possibly, the oral-oral route (Nourai, *et al.*, 2009, Weyermann, *et al.*, 2009). Its presence is widespread in all countries, with a major

incidence in developing countries. It is endemic, in particular, in Japan, Korea and China. It has been reported in literature that almost 60% of the world’s human population is infected by *H. pylori* (Siavoshi, *et al.*, 2005). Most infected people are asymptomatic (70%) or exhibit the so called “benign gastritis” clinical phenotype, characterized by mild pangastritis with little or no damage of parietal cells in the stomach corpus. 15% of infected patients develop a more clinically relevant phenotype characterized by the formation of duodenal and prepyloric

ulcers, high secretion of gastrin, antral gastritis and high acid production. Approximately 1% of patients are affected by a more serious phenotype characterized by multi-focal gastric atrophy, intestinal metaplasia (neoplastic transformation of gastric epithelial cells into a cellular type resembling the intestinal epithelium), achlorhydria and subsequent formation of gastric adenocarcinoma and MALT (Mucosa-Associated Lymphoid Tissue) lymphoma (McNamara & El-Omar, 2008).

Gastric cancer is the second leading cause of cancer death in the world after the Non-Small Cell Lung Carcinoma, with around 800000 deaths per year as reported in February 2009 by the World Health Organization (WHO Fact Sheet n.297, February 2009). Because of the established correlation between *H. pylori* infection and gastric cancer, the WHO has designated it the first bacterial class I carcinogen in 1994 (Eidt & Stolte, 1995, Suerbaum & Michetti, 2002, Sugiyama, 2004).

I.1.1 Adaption to the gastric acidic environment by bacteria

The human stomach is characterized by an acidic pH, ranging between 1.0 during starvation and 5.0 during the early digestive phase (Sachs, 2003). The high amount of HCl secreted by parietal cells (estimated to be around 1.5 L of 100 mM HCl per day) is lethal to most non-acidophylic bacteria that enter the gastric lumen. Nevertheless, there are some neutralophilic bacteria which are able to survive, but not grow, in the human stomach for the time necessary to transit and go to the more neutral environment of the intestine. Their survival depends on an adaption system that keeps their cytoplasmic pH around 5.0. This adaption system (often referred to as “acid resistance/tolerance”) involves several enzymes that prevent the excessive influx of protons that can acidify the cytoplasm and hamper the Redox potential created during the bacterial metabolism. Many acid-tolerant neutralophils use the F_1F_0 ATPase that is activated by high levels of H^+ and exports them from the cell as a consequence of ATP hydrolysis (Sachs, *et al.*, 2005). Other components, the aminoacid decarboxylases, catalyze, in the presence of H^+ ions, the conversion of an aminoacid (namely Aspartate, Histidine and Arginine) into an intermediate that is subsequently exported from the cell. This mechanism, which is finely tuned by a set of genes encoding the decarboxylases and the antiporters, also prevents an excessive increase of acidity inside the bacterial cytoplasm (Kieboom & Abee, 2006, Moreau, 2007). As a further regulation mechanism to prevent alkalinization due to the export of protons, the Na^+/H^+ exchanger drives sodium ions out of the cell in exchange of $2H^+$ (Sachs, *et al.*, 2005).

I.1.2 Adaption to the gastric acidic environment by *H. pylori*

I.1.2.1 Motility

The capability of *H. pylori* to efficiently colonize and infect the gastric tract strongly depends on its motility and adhesion features. The bacterium presents on its surface between 2 and 8 monopolar flagella, that propel the pathogen through the mucus layer in the orogastric route (Ottemann & Lowenthal, 2002). Once inside the gastric lumen, *H. pylori* motility is influenced by pH gradients that naturally occur between the mucus layer and the gastric wall. This pH gradient is sensed by chemoceptors, like TlpB, that induce the bacterium to “swim away” from too highly acidic areas towards a more viable area (Amieva & El-Omar, 2008). It has been reported that if the bacteria are kept artificially at pH 4 they lose their motility in a matter of minutes and mutants for TlpB always swim in acidic areas, losing much of their ability to colonize the gastric cells (Schreiber, *et al.*, 2005, Croxen, *et al.*, 2006). Ultrastructural studies of infected gastric epithelia have shown that approximately 75% of total bacteria moves through the gastric mucus layer, whereas the remaining 25% is found in intimate contact with the cells (Hessey, *et al.*, 1990).

H. pylori acid resistance is unique among bacteria and is often termed as “acidic acclimation”, as its adaption mechanisms not only allow it to survive in the human stomach, but also to replicate (Sachs, *et al.*, 2005). The bacterium uses several enzymes that buffer the periplasm consequently to the entry of protons in cells.

I.1.2.2 Urease

The Urease catalyzes the degradation of imported extracellular urea to ammonia and carbamate, which is further degraded to ammonia and carbonic anhydride (Scott, *et al.*, 2002). The Urease is a Ni²⁺-dependent dodecameric enzyme of 1.1 MDa, made up of six heterodimers of two subunits, UreA and UreB (Figure I.2). The enzyme is present in the Urease gene cluster, which includes also UreI, UreE, UreF, UreG, UreH. The heterodimers UreE/UreG and UreF/UreH mediate the insertion of nickel into the forming Urease dodecamer (Volland, *et al.*, 2003), whereas UreI forms an inner membrane, pH-gated channel whose function is to selectively import urea from the periplasmic space to the cytoplasm (Sachs, *et al.*, 2006).

I.1.2.3 The amidases AmiE and AmiF and the Arginase RocF

In addition to Urease, ammonia can be produced by two other enzymes. Two related enzymes, AmiE and AmiF (Figure I.2), are upregulated in response to low pH and hydrolyze short-chain amides to produce ammonia and the related organic acid. The substrate of AmiE is still unknown, whereas AmiF has a restricted specificity for formamide, which is hydrolyzed to ammonia and formate (Hung, *et al.*, 2007). The third enzyme, RocF, is an arginase that produces urea and L-ornithin from its substrate arginine. RocF is postulated to refurbish the stock of intracellular urea when this compound is not abundant in the stomach (Pflock, *et al.*, 2006).

I.1.2.4 The two-component system Ars

The production of Urease and the arginase RocF is finely tuned and dependent on several factors. The expression of *UreA*, *UreB* and *RocF* genes is regulated under acidic conditions by a two-component system, the Ars (acidic response signaling), composed of the Histidine Kinase ArsS and the response regulator ArsR. ArsS functions as a pH sensor, triggering the autophosphorylation of the Histidine residue at low pH. The phosphate group is then transferred to the residue D52 of the cognate response regulator ArsR, which in turn functions as a transcription regulator of the Urease genes (Pflock, *et al.*, 2006).

I.1.2.5 NikR and Fur

Transcription of *UreA* and *UreB* is also regulated by the transcription factor NikR (Figure I.2). NikR is the structural homolog of the *E. coli* EcNikR (Dian, *et al.*, 2006), which acts as a transcriptional repressor of the *nikABCDE* genes in the presence of an excess of extracellular Nickel (Leitch, *et al.*, 2007). In contrast, *H. pylori* NikR acts both as a repressor and activator of transcription of different genes and, in particular, activates the expression of *UreA*, *UreB*, detoxifying enzymes and outer membrane proteins in a Ni²⁺-dependent manner (Dosanjh & Michel, 2006). Another gene regulated by NikR is *Fur*, that in turn regulates the expression of the amidases AmiE and AmiF (Pflock, *et al.*, 2006). It has also been postulated that an unknown sensor of intracellular ammonia and/or urea might be involved in the regulation of AmiE and AmiF (Bury-Mone, *et al.*, 2004, Pflock, *et al.*, 2006).

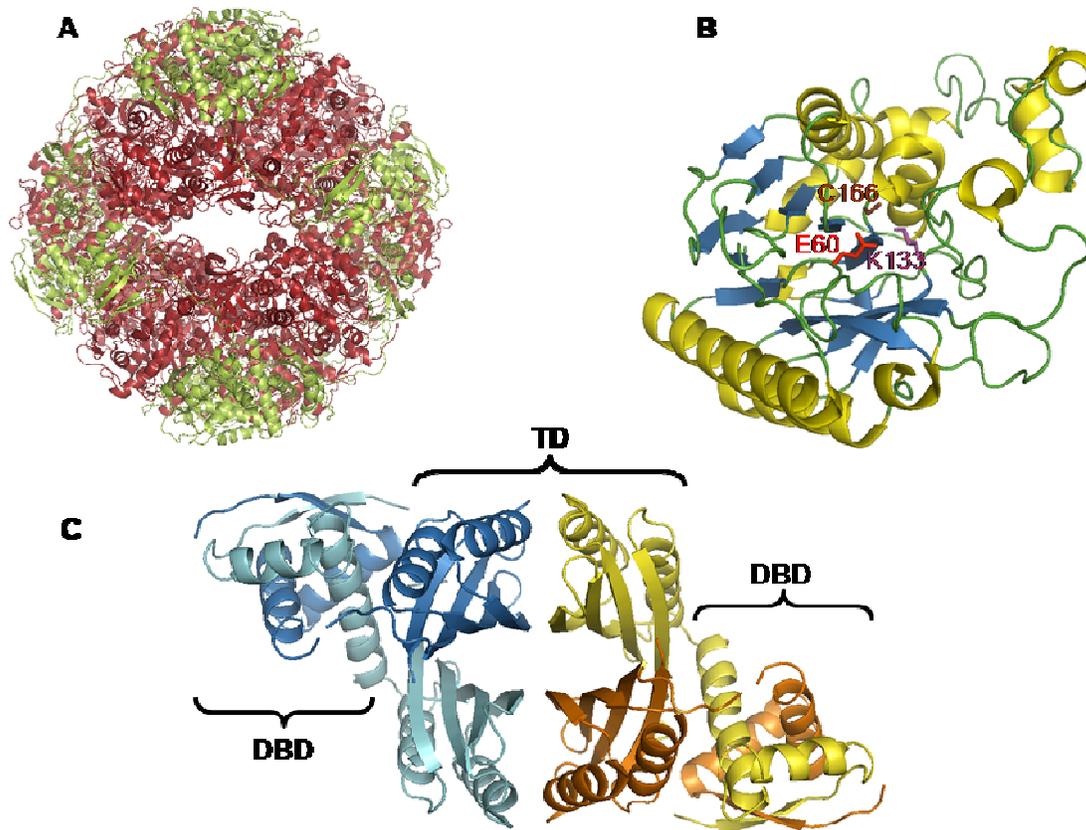


Figure I.2: panel summarizing the known structures of proteins involved in *H. pylori* acid acclimation.

(A) Structure of the dodecameric Urease (accession code 1E9Z). The subunit UreA is colored in green, whilst UreB in red. The assembly of the dodecamer is made by the interaction of four trimers, where each trimer is composed by three UreA/UreB complexes (Ha, *et al.*, 2001).

(B) Structure of the amidase AmiF (accession code 2DYU). Three different β -sheets (blue) are flanked by two long α -helices and 7 shorter ones (yellow). The catalytic triad comprises E60 (red), K133 (magenta) and C166 (brown) (Hung, *et al.*, 2007)

(C) Structure of the apo form of NikR (accession code 2CA9). The structure is a homotetramer (each subunit is colored in blue, cyan, yellow and orange respectively) composed of two different domains. Two DNA binding domains (DBD) composed by a helical bundle made by two subunits, flank the transcriptional domain (TD) (Dian, *et al.*, 2006).

I.2 Bacterial virulence: roles of virulence factors in bacterial survival and host cell colonization

Throughout the evolution process, pathogenic bacteria have established several strategies to adapt, survive and cause damage in their host. It is particularly the case for *H. pylori*. Colonization is a multistep process, with each step requiring key elements, that ultimately leads to infection (Figure I.3). The capability of a given bacteria to infect its host organism will thus depend on the presence of virulence factors that are directly involved in these infection steps. As a consequence, virulence factor functions are extremely diversified. Here, the key steps of infection are reviewed, with particular emphasis on examples of specific bacterial virulence factors that play an essential role in each step.

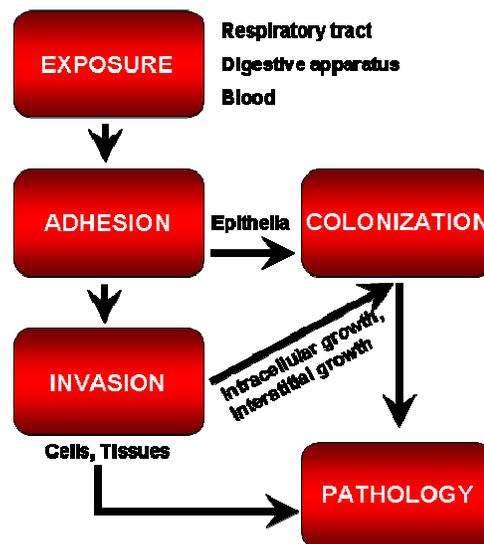


Figure I.3: schematic view of bacterial infection process. Bacteria enter the human body through various pathways, and then adhere to specific tissues. Bacteria then multiply (colonization) or enter the cells (invasion). These steps lead to infection and pathologies.

I.2.1 Bacterial adhesion to the host cell

Bacteria can enter the host through several pathways, like the respiratory tract, the orogastric way or by accessing the blood stream through skin wounds. Once inside, they can stably occupy a niche in the target organism, mainly by adhering to the epithelia in different cellular districts. The adhesion process is thus essential for many bacteria in order to allow

both survival and colonization. Adhesion is mediated by diverse bacterium-specific adhesion molecules, called adhesins, which are located at the bacterial surface and recognize either specific host cellular receptors or proteins of the extracellular matrix. Adhesins can be present on elongated appendices on the surface of the bacterial cells (pili or fimbriae), or be directly tethered to the outer membrane (Soto & Hultgren, 1999). In both cases, these proteins mediate the first steps of the host-microbe interactions.

Bacterial adhesins are very diverse and are generally specific for certain receptors. For instance, the human-infecting Gram-negative pathogen *Bartonella quintana*, which causes a wide spectrum of pathologies in humans, produces the protein BadA (Bartonella adhesin A), a protein of 328 kDa that allows these bacteria to self-aggregate as well as to attach to the extracellular matrix proteins fibronectin and collagen. *Streptococcus pneumoniae* cbpA, a 75 kDa protein, binds to sialic acid present on cytokine-activated epithelial cells and to interact with the polymeric immunoglobulin receptor (pIgR) present on the human mucosa (Murdoch, *et al.*, 2002, Lu, *et al.*, 2003).

I.2.2 Helicobacter pylori adhesins

H. pylori features an arsenal of adhesins that contribute to the colonization, survival and virulence properties of the bacterium. The unusually high number of adhesins involved in *H. pylori* attachment to gastric cells gives proof of the high adaptability of this bacterium to the surrounding environment. *H. pylori* uses different adhesins with different specificities to modulate and adapt its colonization process in response to several host stimuli. *H. pylori* adhesins encompass 33 members divided into two categories based on their amino-acid sequences: the *Helicobacter* Outer membrane Porins (Hops) and the Hop-Related proteins (Hors) (Odenbreit, *et al.*, 2009). Studies on *H. pylori* adhesins have until now focused only on members of the Hops family.

The protein BabA belongs to the Hops family and is probably the most characterized adhesin of *H. pylori*. BabA binds to the Lewis b antigen (Le^b), a di-fucosylated carbohydrate of the blood group antigen ABO (Yamaoka, 2008) and mediates attachment to gastric cells. BabA variants from different clinical isolates have been reported to bind also ALe^b and BLe^b, the antigenic determinants of blood groups A and B, respectively. Most variants bound to all blood groups determinants (A, B and O), therefore strains bearing those variants have been termed “Generalist Strains”. A small subset bind only to Ale^b, therefore they are referred to as “Specialized strains” (Aspholm-Hurtig, *et al.*, 2004).

SabA, also belonging to the Hops family, binds to laminin and Sia-Le^X, the sialylated variant of the Lewis-X antigen. This antigen is commonly up-regulated in gastric cells as a consequence of an inflammatory process. Thus, SabA is proposed to actively participate to the *H. pylori* chronic infection (Aspholm, *et al.*, 2006).

Other reported adhesins are AlpA and AlpB, which are present in asian strains of *H. pylori* and have been identified in patients suffering from gastritis and gastroduodenal ulcers (Odenbreit, *et al.*, 2009). Another adhesin, OipA (Outer Inflammatory Protein A) is associated with the presence of other important virulence factors (Dossumbekova, *et al.*, 2006). Very recently, two other adhesins, HomA and HomB, were identified in strains from patients suffering of gastritis and peptic ulcers, respectively (Oleastro, *et al.*, 2009).

I.3 Toxins: complex evolutionary tools to increase bacterial fitness

Many pathogenic bacteria produce specific proteins dedicated to increase their survival and fitness in the host. Several of these proteins can cause severe damage to host cells and tissues and are therefore termed “toxins”. With the exception of the Lipopolysaccharide (LPS), which is the endotoxin found on the envelope of all Gram-negative bacteria, all bacterial toxins known to date are secreted outside the cell and are therefore termed “Exotoxins”.

Toxins are characterized by a great complexity. Some of them are constituted by a single domain, but many exist as multi-domain proteins, with each domain corresponding to a specific function (such as catalysis, interaction with host proteins or DNA, self oligomerization etc.). For this reason, one toxin can elicit several effects in the target cell, each one contributing to the pathogenesis. Furthermore, with the structural characterization of several bacterial toxins, it is becoming increasingly clear that many of these proteins exert their functions by mimicking host substrates involved in important signaling pathways. This occurs through conserved functional residues or the presence of the same surface properties, like charge and hydrophobicity (Stebbins & Galan, 2001).

The general properties of bacterial toxins are described here according to the host function they target (the cell cycle, the cell cytoskeleton or the immune system functions), the damage they cause on the target cell (for example, cytolysis) or the modifications they undergo in the host cells such as phosphorylation. Selected examples of each of these

categories will be presented in detail, in particular those for which structural information is available and those from *H. pylori*.

I.3.1 Bacterial toxins targeting the host cell cycle

Several bacterial toxins have been found to target and subvert the cellular division process. Since the cell cycle is a critical process for all cells, bacteria producing these toxins might have a selective advantage to survive in the host, by inhibiting the proliferation of immune cells (Nougayrede, *et al.*, 2005). Bacterial toxins modulating the host cell cycle are termed “Cyclomodulins”. The first discovered member of this important class of toxins is the Cytolethal Distending Toxin (CDT), which can be found in many pathogenic bacteria, like *E. coli*, *S. typhimurium* and *S. dysenteriae* (Smith & Bayles, 2006).

I.3.1.1 The Cytolethal Distending Toxin (CDT)

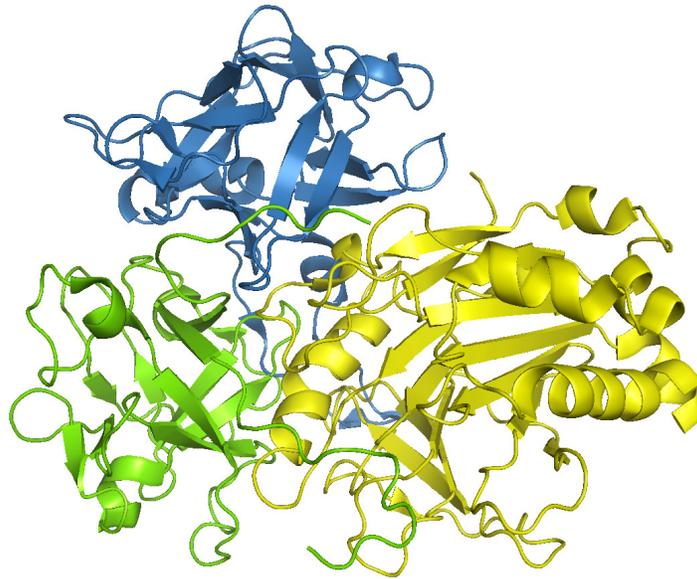


Figure I.4: crystal structure of the tripartite complex of CDT from *Haemophilus ducreyi* (Nesic, *et al.*, 2004) .

Blue: CdtA; yellow: the biologically active subunit CdtB; green: CdtC. CdtA and CdtC are both globular proteins, characterized by a lectin-type fold with 12 β -strands arranged in three symmetric β -sheets. CdtB presents an overall α - β sandwich fold structurally similar to the enzyme DNase I. The tripartite complex buries a total area of 10700 \AA^2 . Accession code: 1SR4.

CDT is composed of three subunits (Figure I.4): CdtA and CdtC are able to form a heterodimer that is required for the delivery of the third subunit, the nuclease CdtB, into the host lymphocytes. Here, CdtB is translocated into the nucleus, where it causes lesions in dsDNA. This genotoxic damage triggers the autophosphorylation of the Ataxia-Telangiectasia Mutated kinase (ATM), which in turn triggers a signaling cascade that arrests the progression of the cell cycle from phase G₂ to M (Ge, *et al.*, 2008). Recently, it has been found that CdtB can also function as a phosphatidil-inositol 1,4,5-triphosphate phosphatase, thus converting PI₃ into PI₂ and causing cell cycle arrest through an unknown mechanism (Shenker, *et al.*, 2007).

I.3.1.2 The Cycle Inhibiting Factor (Cif)

Another cyclomodulin is the Cif toxin, present in the enteropathogenic (EPEC) and enterohemorrhagic (EHEC) strains of *E. coli* and in the human pathogens *Y. pseudotuberculosis* and *Burkholderia pseudomallei* (Nougayrede, *et al.*, 2005, Taieb, *et al.*, 2006, Jubelin, *et al.*, 2009). The gene encoding Cif is found in all EPEC and EHEC strains, and associates with a type III secretion system (T3SSs, see paragraph I.4.4 for details). This protein is a substrate of this T3SS and presents an N-terminal translocation signal (residues 1-20) and an active site (residues 21-282). Its crystal structure (Figure I.5) has revealed that Cif is a member of the Cysteine Protease superfamily that includes cysteine proteases, acetyltransferases and transglutaminases (Hsu, *et al.*, 2008, Jubelin, *et al.*, 2009).

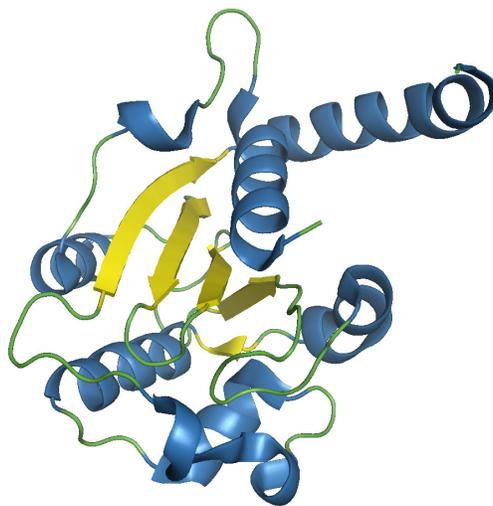


Figure I.5: crystal structure of Cif from *E. coli* (Hsu, *et al.*, 2008).

The structure of Cif display a α - β fold, with a central, four stranded β -sheet (yellow) flanked on both sides by six α -helices (blue). The structure is homologous to the protein AvrPphB from *Pseudomonas syringae*, a member of the Cysteine Protease superfamily that contains cysteine proteases, acetyltransferases, and transglutaminases. Accession code: 3EFY.

In contrast with CDT, Cif does not cause DNA damage, yet the accumulation of the cyclin-dependent kinase (CDK) inhibitor proteins via the inhibition of their degradation by the proteasome. The accumulation of these proteins subsequently causes an irreversible cytopathic effect characterized by cell cycle arrest either in G₁ or in G₂ and the formation of actin stress fibers (Samba-Louaka, *et al.*, 2008).

I.3.2 Bacterial proteins targeting Rho GTPases

I.3.2.1 Rho GTPases and their role in actin polymerization

Rho GTPases are among the most important host targets for many bacterial toxins. They consist of small G-proteins belonging to the Ras superfamily and control a wide range of crucial cellular functions, like cell polarity, proliferation and migration (Etienne-Manneville & Hall, 2002). The best characterized Rho GTPases are Rho, Rac1 and CDC42. They play a very important role in regulating actin polymerization, thus controlling the dynamics of the cytoskeleton (Figure I.6). Their activity is linked to the hydrolysis of GTP. When bound to GTP, Rho GTPases trigger numerous signaling pathways by interacting with several downstream effector proteins, whilst the GTP hydrolysis inactivates the Rho GTPases leaving GDP bound to their catalytic site. GTP hydrolysis is enhanced by several proteins known as GTPase Activating Proteins (GAPs). Rho GTPases can be re-activated by another family of proteins called Guanine-nucleotide Exchange Factors (GEFs) that exchange the bound GDP for GTP (Schwartz, 2004).

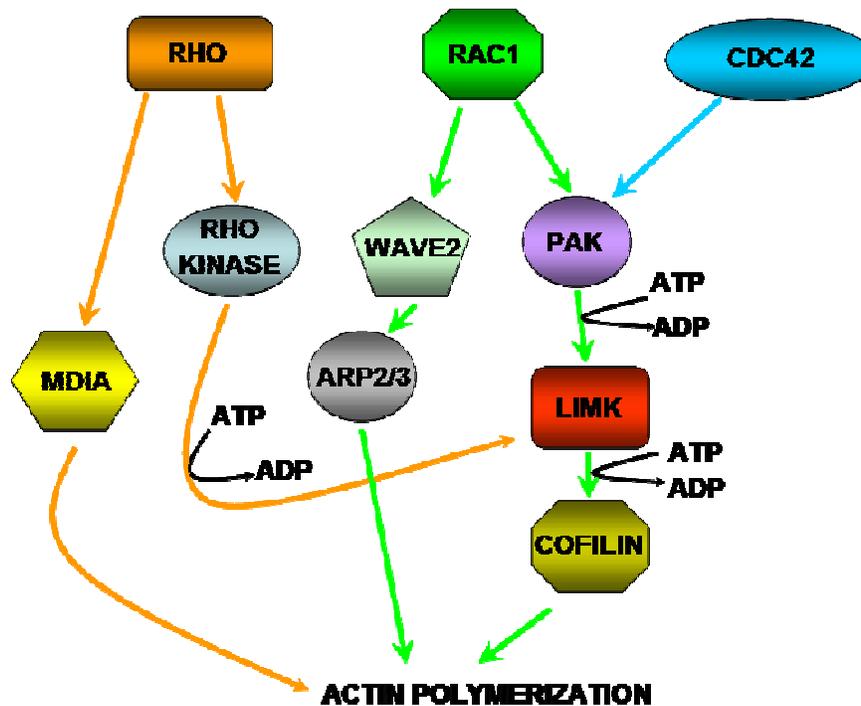


Figure I.6: Rho GTPases trigger several signaling cascades that lead to actin polymerization.

In orange: Rho activates the proteins MDIA (Mammalian Diaphanous protein) and Rho kinase. The latter phosphorylates and activates the Lim Kinase (LIMK), which in turn phosphorylates cofilin, an important scaffold protein that mediates actin polymerization.

In green: Rac activates the proteins WAVE2 and the p21-activated kinase (PAK). Wave2 activates the Arp2/3 complex, which acts as a nucleation site for actin polymerization. PAK in turn phosphorylates and activates the Lim kinase. In blue: CDC42, like Rac1, activates PAK, which leads to the activation of cofilin.

I.3.2.2 The Cytotoxic Necrotizing factor (CNF) 1

Some cyclomodulins promote cell cycle progression instead of arresting it. One of these proteins is CNF1, a 110 kDa protein that is expressed by the uropathogenic (UPEC) strains of *E. coli* (Kouokam, *et al.*, 2006). CNF1 activates the Rho GTPases Rho, Rac and CDC42. The activation process requires the CNF1 dependent deamidation of specific glutamine residues found in the GTP-hydrolyzing domain of Rho proteins, which inhibits GTP hydrolysis and freezes G-proteins in an active state (Flatau, *et al.*, 2000). CNF1 has also a promototic activity, i.e. it promotes the progression of the cell cycle towards the M phase by acting on the cyclin B1 (Falzano, *et al.*, 2006, Giamboi-Miraglia, *et al.*, 2007).

I.3.2.3 YopE and ExoS

Y. pestis effector protein YopE is translocated into host cells through a T3SS and modifies the actin cytoskeleton of the macrophages, therefore inhibiting phagocytosis. This protein acts like a prokaryotic GAP, and increases the GTP hydrolysis activity of Rho GTPase (Hentschke, *et al.*, 2007). Although YopE exerts the same function as eukaryotic regulator proteins, its crystal structure (Figure I.7B) has revealed very little similarities with eukaryotic GAPs (like RICS, figure I.7A) . It was therefore postulated that the molecular surface (which is the “molecular key” by which two proteins interact) might be similar enough to interact with the host effectors through intrinsic charge or hydrophobicity properties. Such phenomenon was named “molecular mimicry” (Stebbins & Galan, 2001).

Other examples of bacterial molecular mimicry are the two effector proteins of *P. aeruginosa*, ExoS and ExoT. In particular, they contain a N-terminal GAP domain, which deactivates Rho GTPases found in macrophages, hereby preventing the phagocytosis of the bacterium (Shames, *et al.*, 2009). The crystal structures of the GAP domains of both YopE and ExoS revealed a striking similarity between these two proteins (Figure I.7 panels B and C), that supports the fact they have the same mode of action (Wurtele, *et al.*, 2001, Evdokimov, *et al.*, 2002).

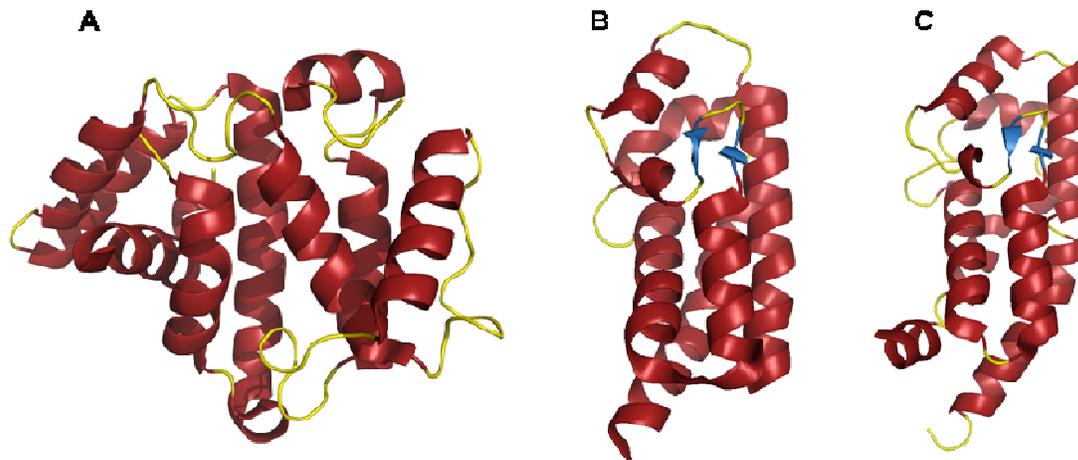


Figure I.7: crystal structures of the GAP domains of the human protein RICS (A), YopE (B) and ExoS (C).

Panel A: the GAP domain of RICS (accession code 3IUG, Nedyalkova *et al.*, not published) presents an overall fold characterized by the presence of 10 α -helices (red) connected by various loops (yellow).

Panel B and C: the crystal structures of YopE (B, accession code 1HY5, Evdokimov *et al.*, 2002) and ExoS (C, accession code 1HE9, Wurtele *et al.*, 2001) are both characterized by an overall very similar

helical fold, characterized by a bundle of 6 α -helices (red) connected by loops (yellow) and by two very short β -strands (blue).

Despite the three proteins all activate the small Rho GTPases Rho1, Rac and CDC42, the fold is similar only between YopE and ExoS, whilst the GAP domain of the human protein RICS strikingly differs.

I.3.2.4 BopE

The *B. pseudomallei* toxin BopE acts as a Guanine-nucleotide Exchange Factor (GEF), activating the Rho GTPases and promoting cytoskeletal rearrangements (Etienne-Manneville & Hall, 2002). BopE is a sequence and functional homolog of the *S. typhimurium* SopE and SopE2 effector proteins, both GEF-like proteins (Upadhyay, *et al.*, 2008). It is postulated that these GEF-like prokaryotic proteins promote the internalization of the bacterium in non-phagocytic cells and therefore enhance its invasion process (Schlumberger & Hardt, 2005, Demali, *et al.*, 2006, Shames, *et al.*, 2009).

I.3.3 Bacterial toxins targeting the host immune system

Establishing a successful strategy to avoid or subvert the immune system is of crucial importance for an efficient bacterial colonization in the host tissues. Several Gram-positive and Gram-negative bacteria produce toxins that specifically target the host immune cells.

I.3.3.1 The Superantigens (Sags)

The Gram-positive pathogens *Streptococcus pyogenes* and *Staphylococcus aureus* produce several exotoxins that specifically bind to MHC-II and T-cell receptors located on the surface of the human Antigen-Presenting Cells (APC). APCs are one of the early mediators of the native immune response. They bind to antigens and present them to the effectors cells such as T-lymphocytes. These exotoxins are termed “Superantigens” (Sags) and can stimulate a massive production of cytokines from the T lymphocytes (Fraser & Proft, 2008). At the early stage of infection, Sags mediate the production of proinflammatory and anti-inflammatory cytokines (like IL-2, IFN- γ), the latter preventing the arrival of macrophages and neutrophils at the site of infection (Fraser & Proft, 2008). The decrease of the local inflammation and of the innate immune response can facilitate the colonization of the host by the bacterium (Vojtov, *et al.*, 2002).

All Sags found in *S. pyogenes* and *S. aureus* present the same overall fold (Figure I.8): a C-terminal β -grasp domain composed of four to five β -strands encircling a conserved α -helix and a N-terminal OB domain (Sundberg, *et al.*, 2002). The OB domain consists of a β -barrel capped by a small α -helix and is also found in the Cholera and Pertussis Toxins (Arcus, *et al.*, 2000, Zhang & Kim, 2000).

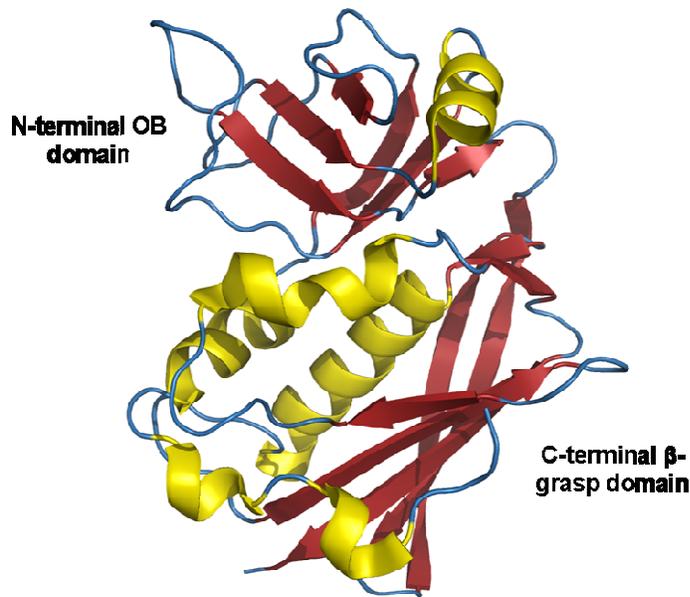


Figure I.8: crystal structure of *S. pyogenes* Superantigen A (Sundberg, *et al.*, 2002).

The structure is divided into two domains. The first is the N-terminal domain (OB domain) characterized by six β -strands (red) organized in two β -sheets connected by several loops (blue) and a small α -helix (yellow). The second domain is the C-terminal β -grasp domain containing 5 antiparallel β -strands flanked by two long and three short α -helices. Accession code 1LOX.

I.3.3.2 Secreted proteases

Several pathogens produce proteases that enable them to survive without being recognized by the host immune system. The bacterium *Pseudomonas aeruginosa*, which causes several opportunistic infections in humans, produces several proteases that directly modulate the host immune response. One of these is the alkaline protease AprA, a 50 kDa protein that targets immune cells (phagocytes, Natural Killer cells, T lymphocytes), degrades several cytokines (IL-1, IL-2, IFN- γ , TNF- α), immunoglobulins and complement proteins (Parmely, *et al.*, 1990, Leidal, *et al.*, 2003, Kida, *et al.*, 2008). Other proteases produced by this bacterium (LasA and LasB) can degrade proteins present in the extracellular matrix

(elastin, collagen) as well as immunoglobulins and complement proteins, enabling the bacterium to disseminate the infection in the occupied niche (Coin, *et al.*, 1997, Cowell, *et al.*, 2003). Once settled in the target tissue, bacteria locally multiply in small areas which are called “infection foci”. However, if the growth is fast enough and not efficiently hampered by the host immune system, they can disseminate and infect other organs through the lymphatic and circulatory system, causing sepsis.

I.3.3.3 OspB and OspF

Other bacterial toxins, like *S. flexneri* OspB and OspF, diminish the host inflammatory response. OspF has been reported to impair the Mitogen Activated Protein Kinase (MAPK)-dependent phosphorylation of the histonic protein H3, which in turn inhibits the production of proinflammatory cytokines (Arbibe, *et al.*, 2007).

OspB functions in synergy with OspF. OspB enters the nucleus where it induces chromatin remodeling through the interaction with the Retinoblastoma (Rb) protein. Furthermore OspB and OspF concerted action reduces the migration of Polymorphonuclear leukocytes towards the site of infection (Zurawski, *et al.*, 2009).

I.3.4 The Pore-Formin Toxins (PFTs)

PFTs represent a large family of toxins that are secreted in a water-soluble form and, upon contact with the host plasmamembrane, oligomerize and form large pores on the cell surface (Rosado, *et al.*, 2008). PFTs are divided into two categories: the Membrane Attack Complex/Perforin superfamily (MACPF) produced by mammals and the bacterial Cholesterol-Dependent Cytolysins (CDC).

CDCs represent a large and widely studied family of bacterial toxins, with more than 700 members found in five different genera of Gram-positive bacteria. CDCs promote cell invasion, release of nutrients and allow the passage of other toxic proteinaceous compounds of bacterial origin. Furthermore, they are characterized by a proinflammatory activity, that promotes tissue damage at the infection foci (Aroian & van der Goot, 2007). These proteins share from 40% to 80% of sequence similarity, suggesting that they have similar modes of action and similar structures (Rossjohn, *et al.*, 2007).

Among CDCs, the Perfringolysin O from *Clostridium perfringens* has been extensively studied from both a structural and a mechanistic point of view. This protein, of

around 100 kDa, is elongated, dimeric, rod-shaped and constituted by four domains (Figure I.9, panel A). The binding of domain 4 to the membrane promotes an allosteric change on the distal domain 3 that rearranges two α -helices near the domains 2 and 3 into two β -hairpin structures, called Transmembrane-Hairpin Regions (TMH). The TMH1 and TMH2 then insert at the membrane to create a multimeric channel of more than 200 Å in diameter (Rossjohn, *et al.*, 2007). In Figure I.9 (panels B and C), electron microscopy structures of the *S. pneumoniae* CDC Pneumolysin pre-pore and pore are depicted. In panels D and E, structural models of the Pneumolysin pre-pore and pore are depicted with the structure of Perfringolysin O fitted inside the EM map (Tilley, *et al.*, 2005).

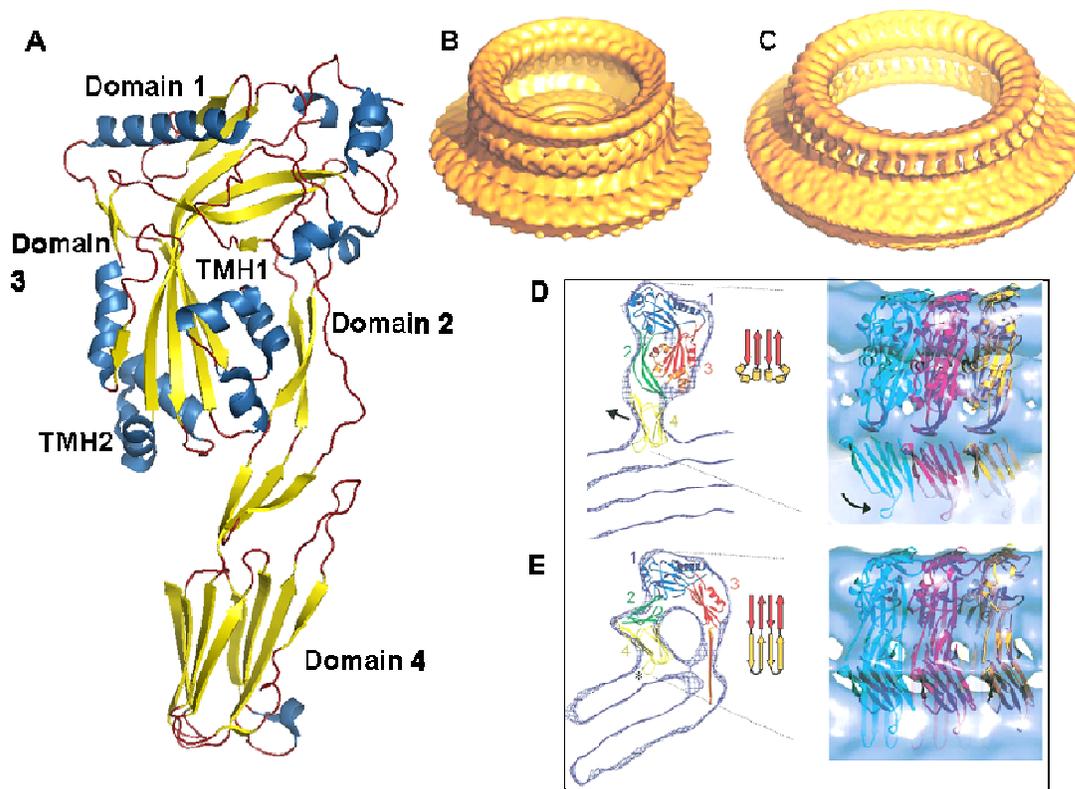


Figure I.9: crystal structure of monomeric *C. perfringens* Perfringolysin O (accession code 1M3I) and Pneumolysin pore model (Tilley, *et al.*, 2005, Rossjohn, *et al.*, 2007).

(A) The Perfringolysin O structure contains four domains. Domain 4 binds to the plasmamembrane of the host cell and triggers a conformational change of domain 3 that collapses downward by around 40 Å. The conformational change expose TMH1 and TMH2, that turn into β -hairpins and change the overall fold of the protein before the creation of the channel. (B) Electron microscopy model of Pneumolysin Pre-pore, after the insertion of Pneumolysin into the membrane. (C) Electron microscopy

model of Pneumolysin pore, after the restructuring of the domains. (D) Fitting of the Perfringolysin O structure into the pre-pore electron density map. The domains of Perfringolysin O are depicted in different colors. By comparison with Perfringolysin O structure it is possible to note that no conformational changes occur in the pre-pore structure. (E) Fitting of the Perfringolysin O structure into the pore electron density map. Here, the domains of Perfringolysin O are heavily restructured, with the TMH1 and TMH2 of domain 3 that fold to a β -hairpin that inserts into the membrane. This pulls the domains 2 and 4 and tilts the protein to form an arc-shaped structure at the membrane, that oligomerizes with other subunits to create the pore.

I.3.5 Phosphorylated bacterial toxins

Within the last decade, some bacterial proteins that are delivered into mammalian host cells have been found to be tyrosine-phosphorylated by host kinases (Backert & Selbach, 2005). Before this intriguing discovery, it was widely accepted that tyrosine phosphorylation was a prerogative of higher organisms only. As a matter of fact, this process is one of the most frequent in eukaryotic signaling pathways and regulates many of the most crucial cellular functions, from immune defense to cell differentiation and morphology (Johnson, 2009).

Toxins that are tyrosine-phosphorylated present similar phosphorylation motifs, consisting of a specific sequence of 5-8 residues. The position -3 from the phosphorylated tyrosine is always a glutamic acid and the position -1 is always leucine or isoleucine. At position +1, there is a preference for small residues (like an alanine) or an acidic residue, like aspartic acid. At position +3 there is a trend for a hydrophobic residue, either a valine or an isoleucine (Backert & Selbach, 2005).

I.3.5.1 The Translocated Intimin Receptor

The first discovered bacterial phosphorylated toxin is the Translocated Intimin Receptor (Tir) from the EPEC and EHEC strains of *E. coli* (Kenny, 1999). Tir, a protein with a molecular weight of around 56 kDa, is translocated into human intestinal cells and spontaneously inserts into the membrane in a hairpin loop topology (Frankel & Phillips, 2008). The loop protrudes from the external face of the plasmamembrane and acts as a receptor for the human Intimin. Bound intimin then triggers the Y474 phosphorylation of Tir C-terminal domain by host Fyn and Abl protein kinases. Phosphorylated Tir then recruits Nck

at the sites of bacterial adhesion (Campellone, *et al.*, 2002, Frankel & Phillips, 2008). Nck in turn triggers a signaling cascade resulting in the actin polymerization and the formation of cellular pedestals which support bacterial adhesion (Bommarius, *et al.*, 2007, Schuller, *et al.*, 2007).

I.3.5.2 The Translocated Actin-recruiting Phosphoprotein (Tarp)

Another tyrosine-phosphorylated protein, the Translocated Actin-recruiting Phosphoprotein (Tarp) from *Chlamidia trachomatis*, targets the cellular cytoskeleton. Tarp is translocated into the host cell cytosol and is immediately tyrosine-phosphorylated by several host kinases, like Src, Yes, Fyn and Abl (Elwell, *et al.*, 2008, Jewett, *et al.*, 2008, Mehlitz, *et al.*, 2008). Phosphorylated Tarp then recruits Sos1 and Vav2 (both GEFs that activate the Rho GTPase Rac) that in turn trigger a signaling cascade that activates Rac, Abi-1, WAVE2 and Arp2/3, thus initiating the actin polymerization (Betts, *et al.*, 2009).

I.3.5.3 LspA1 and LspA2

Some bacterial phosphorylated effectors are crucial to inhibit the host immune response. For instance, two proteins, LspA1 and LspA2, encoded by *Haemophilus ducreyi*, are secreted and then internalized into macrophages. Here, they both are tyrosine-phosphorylated by the host c-Src kinase (Mock, *et al.*, 2005, Deng, *et al.*, 2008). It is postulated that the sustained phosphorylation of both LspA proteins depletes the levels of active c-Src in macrophages, thus inhibiting all c-Src dependent signal pathways and *de facto* inhibiting their phagocytic activity (Deng, *et al.*, 2008). For now, it is not clear whether LspA proteins can have more specific effects on the macrophage signaling systems.

I.3.5.4 AnkA

Recently, AnkA, a protein toxin secreted by the VirB/D4-homologous T4SS of *Anaplasma phagocytophilum*, has been found to be tyrosine-phosphorylated by c-Src and c-Abl in human Neutrophils (JW, *et al.*, 2007, Lin, *et al.*, 2007). Phosphorylated AnkA then binds to and activate the phosphatase SHP-1, triggering a still unknown signal pathway. How AnkA can influence the cellular functions of Neutrophils remains to be unravelled, although it is postulated to have a crucial role in *A. phagocytophilum* infection process, as anti-AnkA

antibodies severely hamper the intracellular cycle of this pathogen (Lin, *et al.*, 2007).

I.4 Delivery of toxins towards the host cells: roles of the bacterial Secretion systems

To exert their effects in the host cell, toxins need to be either uptaken through mechanisms of endocytosis/pinocytosis or delivered directly inside the cytoplasm through macromolecular assemblies known as “Secretion Systems”. Recently, bacterial secretion systems have been subject of great attention due to their predominant role in bacterial virulence. These machineries are involved in several steps of bacterial pathogenesis, like colonization, quorum sensing, invasion and toxins delivery (Gerlach & Hensel, 2007).

Bacterial secretion systems are classified into Sec-dependent and Sec-independent group. Sec dependent systems export toxins that are first accumulated in the periplasmic space via the Sec-pathway (see below).

I.4.1 The general secretory pathway (Sec)

This secretion pathway is conserved among different Gram-negative bacteria and accounts for the majority of protein secretion processes (Oliver & Paetzel, 2008). The Sec translocon is composed of 11 proteins: SecA, SecB, SecD, SecE, SecF, SecG, SecY, YajC, YidC, type I signal peptidase (SPase I), and signal peptide peptidase (SppA). Six of these proteins form two heterotrimeric complexes that localize across the inner membrane, named SecYEG and SecDFYajC. SecYEG forms the inner membrane pore across which the protein substrates are translocated (Figure I.10). The ATPase SecA provides the energy for the efficient translocation of the substrate across SecYEG through a mechanism of membrane insertion-deinsertion, with the support of SecDFYajC (Nouwen & Driessen, 2002).

Secretory proteins are presented to the Sec-pathway as precursors, or pre-proteins, containing a Secretion signal peptide of around 20 residues at the N-terminus. During the precursor synthesis, the molecular chaperone SecB binds to molten globules of the nascent polypeptide, keeping the pre-protein in an unfolded state (Nouwen, *et al.*, 2007). The binding of the SecB-precursor to SecA triggers the ATPasic activity of SecA, promoting the translocation of the precursor across SecYEG (Gerlach & Hensel, 2007). Two signal peptide peptidases, SpaseI and SppA, cleave the N-terminal signal peptide from the preprotein and

remove it when the translocation process has ended (Kim, *et al.*, 2008).

Up to now, crystal structures of several components of the Sec-pathway have been solved.

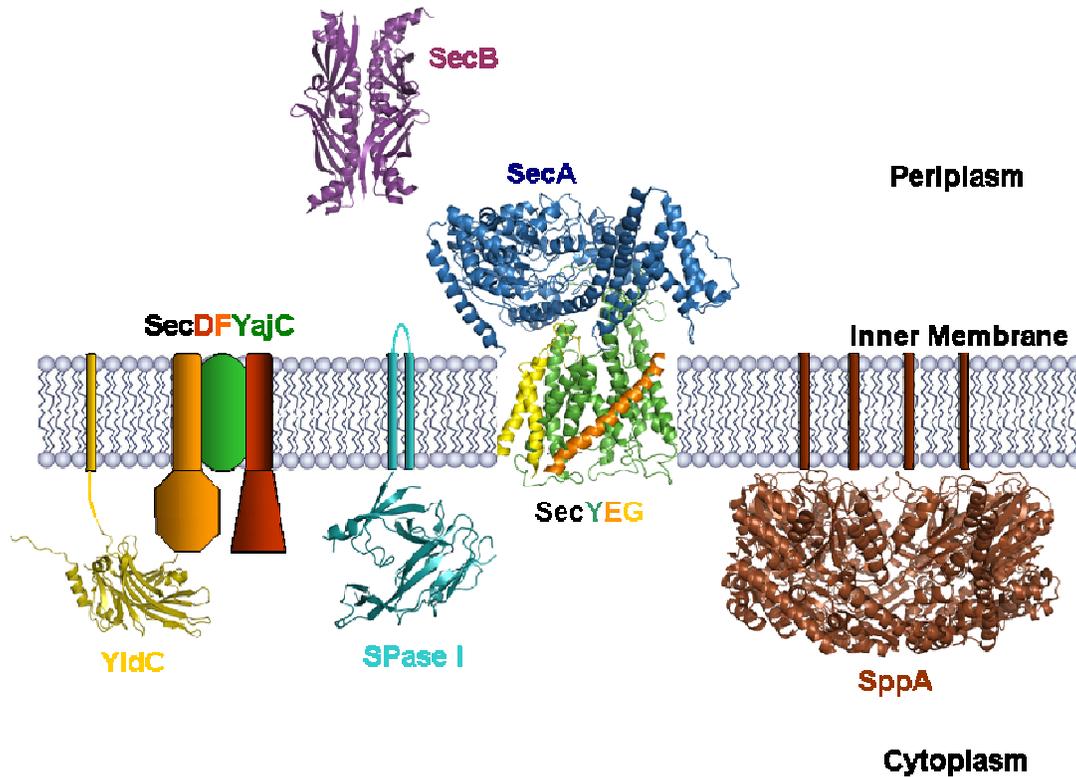


Figure I.10: putative structural model of the bacterial Sec-pathway.

Known structures of Sec-pathway components are depicted, along with their subcellular localization. Accession codes: 1QYN for SecB (Dekker, *et al.*, 2003), 1T7D for Spase I periplasmic domain (Paetzel, *et al.*, 2004), 3BF0 for SppA periplasmic domain (Kim, *et al.*, 2008), 3BLC for YidC periplasmic domain (Oliver & Paetzel, 2008), 3DL8 for the complex SecA-SecYEG (Zimmer, *et al.*, 2008). For unknown structures, a drawing model is reported, according to the available literature (Oliver & Paetzel, 2008).

I.4.2 The Type 2 Secretion System (T2SS)

When proteins are translocated into the periplasm through the Sec-pathway, they are exported out of the bacterial outer membrane by a T2SS. For this reason, T2SSs are considered as the terminal branch of the Sec-pathway (Gerlach & Hensel, 2007) and this is also why we present it before the T1SS. These secretion machineries consist of 12 proteins, denoted from C to O (Peabody, *et al.*, 2003), which are ubiquitously found among the proteobacteria (examples are the human pathogens *Yersinia enterocolitica*, *L. pneumophila*,

P. aeruginosa and the plant pathogens *Erwinia chrysanthemii*, *Xanthomonas axonopodia* and *Xylella fastidiosa*). T2SS proteins assemble to form a macromolecular complex that spans the entire cell envelope. One of the most studied T2SSs is the Eps system (Figure I.11), that belongs to the human pathogen *Vibrio cholerae* and secretes the protein CT (Cholera Toxin) into the human intestinal lumen (Aoki, *et al.*, 2009). Several crystal structures of the *V. cholerae* Eps T2SS components are now available in the Protein Data Bank (PDB). The machinery includes an inner membrane complex formed by EpsF, EpsL, EpsM and the cytoplasmic ATPase EpsE. A particular appendix, the pseudopilin complex, is formed by EpsG proteins and the accessory pseudopilins EpsH, EpsI, EpsJ, EpsK. The assembly of these pseudopilins is mediated by the pre-pilin peptidase EpsO. The system is completed by an outer membrane apparatus formed by the secretin EpsD and the substrate-recognition protein EpsC.

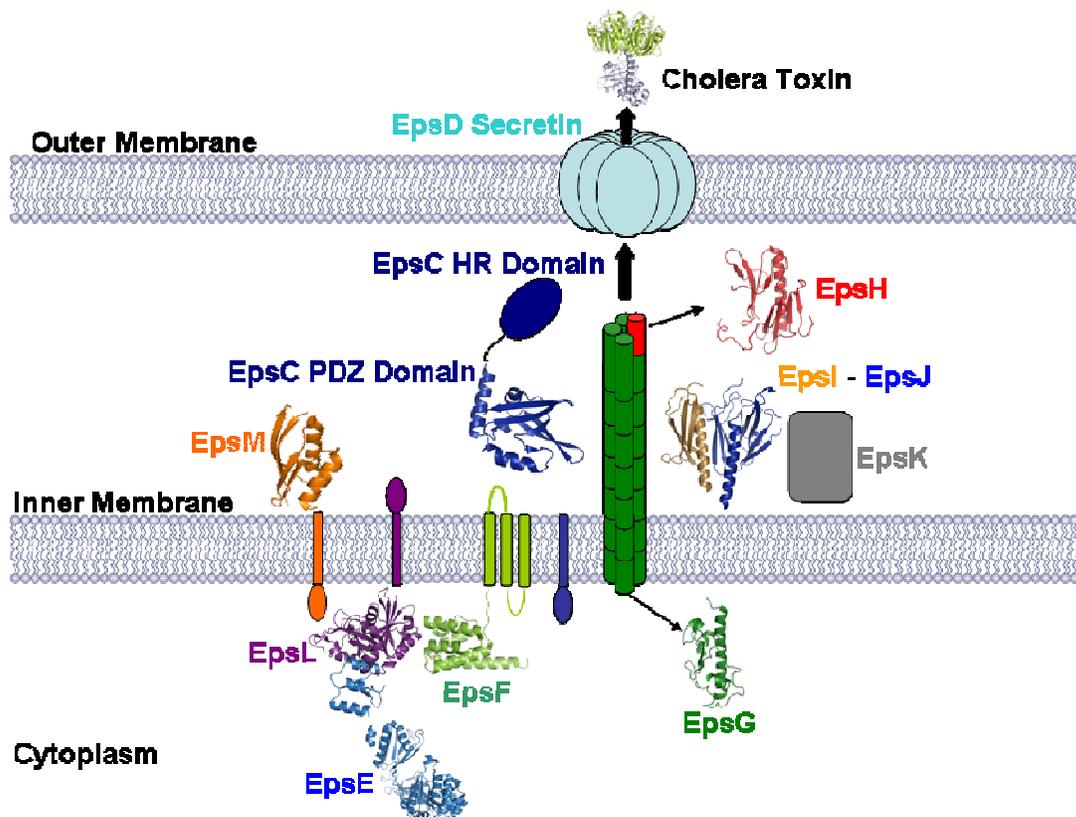


Figure I.11: structure and assembly of *V. cholerae* Eps T2SS.

The inner membrane complex is formed by the ATPase EpsE, which is linked to the inner membrane by the protein EpsL, that stimulates its ATPase activity. EpsM and EpsF interact with EpsL and are essential for T2SS functionality. EpsG is the pseudopilus component and interacts with the other pseudopilins

EpsH, EpsI–EpsJ (that form a tight complex together) and EpsK. EpsC interacts with the secretin EpsD through the HR domain, whilst interacting with the substrate through the PDZ domain.

Accession codes: 1P9R for EpsE (Robien, *et al.*, 2003), 2BH1 for EpsL (Abendroth, *et al.*, 2004), 1UV7 for EpsM (Abendroth, *et al.*, 2004), 3C1Q for EpsF (Abendroth, *et al.*, 2009), 3FU1 for EpsG (Korotkov, *et al.*, 2009), 2QV8 for EpsH (Yanez, *et al.*, 2008), 2RET for the complex EpsI–EpsJ (Yanez, *et al.*, 2008), 2I4S for the EpsC PDZ domain (Korotkov, *et al.*, 2006).

I.4.3 The Type 1 Secretion System (T1SS)

The type 1 secretion system, found in all proteobacteria and in some cyanobacteria, allows the fast, single-step secretion of proteins of various sizes from the cytoplasm directly to the extracellular space (Delepelaire, 2004, Gerlach & Hensel, 2007).

The T1SS machinery consists of a heterotrimeric complex that comprises: i) an inner membrane ABC exporter with a cytoplasmic nucleotide-binding domain; ii) a membrane-fusion protein (MFP) featuring a short cytoplasmic tail, a membrane anchor and a large periplasmic domain; iii) a trimeric outer-membrane protein (OMP), belonging to the TolC superfamily, forming a 140 Å-long outer membrane channel. TolC are common trimeric outer-membrane channels involved in several transport pathways, including the drug efflux pathways. The formation of the apparatus is triggered by the interaction of the C-terminal secretion signal on the secreted protein with the ABC exporter, that provides substrate specificity and energizes the assembly of the machinery through the ATP hydrolysis (Binet & Wandersman, 1995).

In the well studied Hly system from *E. coli*, the T1SS allows for the secretion of the toxin HlyA (also known as α -Hemolysin), a 107 kDa protein that forms pores in the plasmamembrane of target cells (Thanabalu, *et al.*, 1998). The proteins HlyB and HlyD function respectively as the ABC exporter and the MFP (Pimenta, *et al.*, 1999, Benabdelhak, *et al.*, 2005, Holland, *et al.*, 2005). Each subunit of TolC features four outer-membrane β -strands and four long α -helices that protrude towards the periplasmic space.

I.4.4 The Type 3 Secretion System (T3SS)

T3SSs are a very important class of protein-delivery machineries, which are found in at least 25 different species of Gram-negative bacteria (Cornelis, 2006). They are different from the other secretion systems discussed so far, in the sense that their supramolecular architecture

not only spans the entire bacterial cell envelope, it pierces the host cell to translocate effector proteins inside. From their resemblance to syringes, T3SSs are also known as injectisomes (Mueller, *et al.*, 2008). T3SSs consist in a structural core complex made of nine proteins, however, additional proteins participate in the assembly and function of this secretion system, raising the number of T3SS components to 20-25, depending on the bacterium (Cornelis, 2006).

All T3SSs share a similar overall architecture: a cylindrical structure, termed the needle complex (NC) that spans both bacterial membranes. Among them, the Mxi system from *Shigella flexneri*, has been extensively study. This system is used to delivers a set of protein effectors inside the cytoplasm of intestinal cells. These proteins promote the internalization of the bacterium inside the host cell cytoplasm and subvert signaling pathways involved in the innate immune response to escape bacterial killing and support the infection (Phalipon & Sansonetti, 2007, Ogawa, *et al.*, 2008). The structure of the needle complex of the Mxi, has been reconstructed by electron microscopy (Hodgkinson, *et al.*, 2009, Spreter, *et al.*, 2009). It forms three outer membrane rings (OMR1 to 3), a connector ring (Conn) and two inner membrane ring (IMR-1 and -2). Structures of proteins from *E. coli* and *S. typhimurium* T3SS, which are homologue to Mxi components, have been fitted inside the electron density of the reconstructed EM model (Figure I.12). The two membrane complexes are joined together by a central periplasmic rod (inner rod), protruding from an inner membrane socket (Figure I.12). An extracellular needle, that extends the inner rod and allows the translocation of the effector proteins into the target host cell, completes the assembly but its structure has not been solved. Interestingly, many T3SS proteins are sequence homologues of proteins from bacterial flagella. Moreover, electron-microscopy structures of T3SS needle complexes have also highlighted a striking structural resemblance between these two bacterial supramolecular assemblies (Kubori, *et al.*, 1998, Marlovits, *et al.*, 2004), concluding that T3SSs and flagella might be evolutionary related.

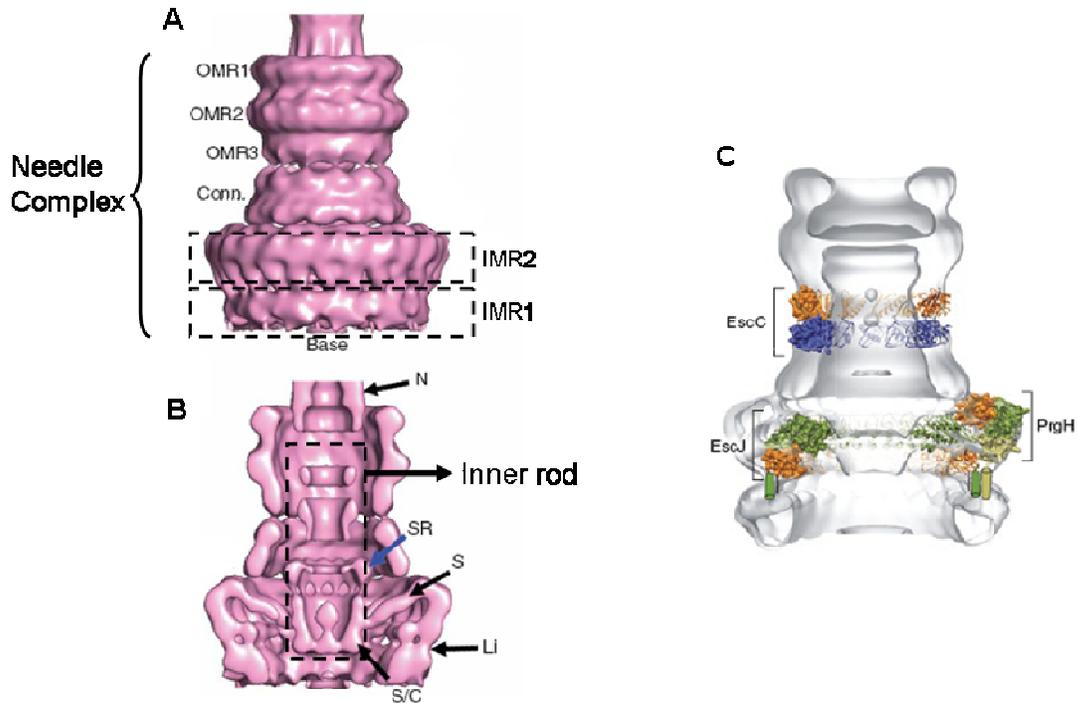


Figure I.12: 3D electron Microscopy reconstruction of *S. flexneri* Mxi T3SS needle complex (hodgkinson et al., 2009) .

(A) The overall architecture of the needle complex is composed of: three outer membrane rings (OMR1, OMR2, OMR3), a connector ring (Conn.) linked to the OMR3 and two inner membrane rings (IMR1, IMR2). OMR1, OMR2, OMR3 and the connector ring are characterized by a 12-fold symmetry, in contrast with the IMRs, which are characterized by a 24-fold symmetry. (B) The internal components are: the needle basis (N) protruding from the OMR1, the socket ring (SR), which is connected with the socket cup (S/C) beneath it. S and Li are two linker regions connecting the socket cup and the underside of the IMR to the base of the needle complex. (C) Three structures of Mxi homologous proteins coming from *E. coli* (EscC, EscJ) and *S. typhimurium* (PrgH) have been fitted into the electron density of the Mxi needle complex model, showing that domains of the inner and outer membrane components are structurally similar (similar domains are colored in orange and green).

I.4.5 The Type 4 Secretion system (T4SS)

T4SSs are unique, multifunctional secretion machineries that are used by both Gram-positive and Gram-negative bacteria to transfer effector proteins and protein-DNA complexes across membranes (Juhas, *et al.*, 2008). They are related to bacterial conjugation systems, which mediate horizontal gene transfer among bacteria of the same strain (Christie, *et al.*,

2005). The best known conjugation system is encoded by the F-plasmid, which mediates the diffusion of antibiotic resistance and virulence genes among bacteria through a conjugation pilus. Based on sequence similarities and functional homologies, these secretion systems are divided into 2 classes, A and B. T4ASSs comprise conjugative systems, like the ones encoded by F- and P-plasmids, and protein or DNA-protein secretion systems. Members of this class have a common architecture, resembling the prototypal VirB/D4 T4SS of the plant bacterium *Agrobacterium tumefaciens* (Lawley, *et al.*, 2003). T4BSSs are considerably different and their protein resembles more those of the Dot/Icm T4SS of *L. pneumophila* (Juhas, *et al.*, 2008).

I.4.5.1 The VirB/D4 T4ASS of *Agrobacterium tumefaciens*

A. tumefaciens is a Gram-negative bacterium which causes tumors in several species of plants. It harbors a 10 kb gene cluster, termed “*virB* operon”, that encodes the archetypal *VirB/D4* T4ASS (Cascales & Christie, 2004). This system consists of 11 VirB proteins (VirB1 to VirB11) and the protein VirD4 (Figure I.13). The assembly of the machinery is driven by two ATPases, VirB4 and VirB11, located at the cytosolic face of the inner membrane. Another ATPase, VirD4, also referred to as the “coupling protein”, recruits the DNA substrate (T-DNA) to the other T4SS components (Atmakuri, *et al.*, 2004, Backert & Meyer, 2006, Guo, *et al.*, 2007).

VirB6, VirB7, VirB8, VirB9 and VirB10 form the membrane-spanning secretion channel. VirB6 is an integral inner membrane protein required for the pilus assembly (Judd, *et al.*, 2005). VirB7 is a lipoprotein that interacts with the putative outer membrane pore protein VirB9, forming a channel complex that stabilizes other T4SS components and regulates pilus synthesis and substrate selection (Jakubowski, *et al.*, 2005, Bayliss, *et al.*, 2007).

VirB8 is conserved in all T4SSs and is considered as a key element for the assembly of the machinery as well as for the substrate translocation (Baron, 2006). It interacts with several T4SS components, like VirB1, VirB4, VirB5, VirB6, VirB9, VirB10, VirB11 and itself. (Judd, *et al.*, 2005, Yuan, *et al.*, 2005). VirB10 is a multi-domain protein, localized in the inner membrane, that interacts with the VirB7-VirB9 channel complex and participates in pilus biogenesis and assembly of the secretion channel (Jakubowski, *et al.*, 2009).

VirB2 and VirB5 are respectively the major and minor components of the so called T-pilus (Aly & Baron, 2007). VirB2 forms the pilus structure while VirB5 localizes at the pilus tip and can function as a bacterial adhesin, mediating contact with host cells (Aly &

Baron, 2007, Backert, *et al.*, 2008). VirB3 role in *A. tumefaciens* T4SS assembly is still not clear. However, it is considered as a minor component of the T-pilus and is present at the bacterial outer membrane (Judd, *et al.*, 2005, Juhas, *et al.*, 2008). VirB1 is a multifunctional protein that has both an enzymatic activity and a role in T4SS assembly. Its N-terminal domain shows a transglycosylase activity, that degrades the peptidoglycan layer to allow the formation of the secretion pore (Hoppner, *et al.*, 2004). VirB1 undergoes C-terminal cleavage and the resulting C-terminal fragment, called VirB1*, is secreted outside the bacterial cell (Llosa, *et al.*, 2000). Recently, VirB1* has been reported to be also essential for the pilus assembly, mediating protein-protein interactions between VirB2 subunits (Zupan, *et al.*, 2007).

Although the T4SS of *A. tumefaciens* is considered the archetypal model for all T4ASSs, its architecture and the stoichiometry of its components are still poorly understood. Recent studies have shed some light on the assembly of homologous T4ASSs from other bacteria. In a recent work by Fronzes and coworkers, the electron microscopy structure of the core complex from the *E. coli* Tra T4SS has been revealed (Fronzes, *et al.*, 2009). This complex, which constitutes the secretion channel, is formed by two membrane-spanning channels composed of VirB7, VirB9 and VirB10. The outer membrane ring (called the O-layer) is formed by the C-terminus of VirB9, the C-terminus of VirB10 and the full-length VirB7, whereas the inner membrane ring (called the I-layer) is made of the N-termini of VirB9 and VirB10. The overall structure presents a 14-fold symmetry and dimensions of 185 Å in height and diameter (Figure I.13).

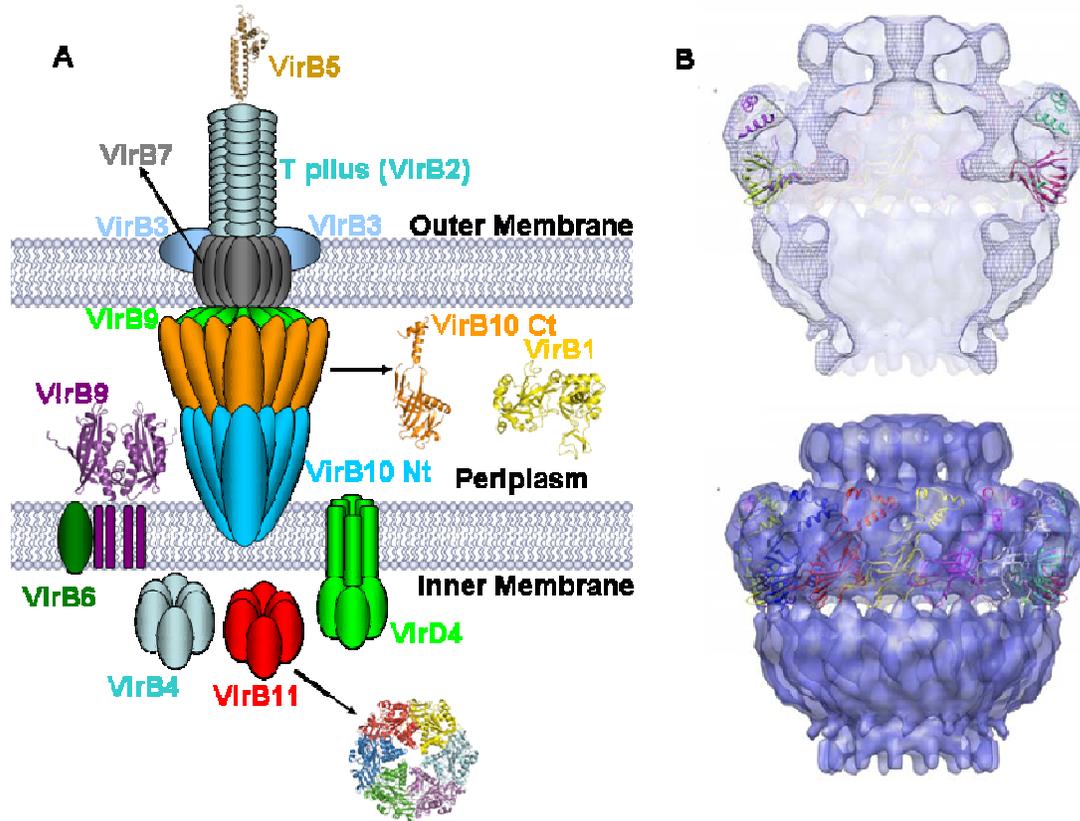


Figure I.13: putative model of the T4ASS

(A) Structures (where available) and subcellular localizations of the T4ASS components are depicted.

Accession codes: 1NLZ for the VirB11 from *Helicobacter pylori* (Savvides, *et al.*, 2003), 2CC3 for *A. tumefaciens* VirB8 (Bailey, *et al.*, 2006), 2BHV for the VirB10 homolog ComB10 from *H. pylori* (Terradot, *et al.*, 2005), 1R8I for the VirB5 homolog TraC from *E. coli* (Yeo, *et al.*, 2003). (B) electron microscopy model of the *Tra* T4ASS, where the structure of the C-terminal periplasmic domain of ComB10 (VirB10 homolog) is fitted into the O-layer electron density map (Fronzes, *et al.*, 2009).

I.4.5.2 The Dot/Icm system from *Legionella Pneumophila* (T4BSS)

L. pneumophila is an intracellular Gram-negative bacterium that infects several species of environmental amoebae and protozoa. It is also capable of infecting humans through inhalation of contaminated water aerosols, causing an atypical pneumonia termed “Legionnaire’s fever” (Philippe, *et al.*, 2006). Once in the lungs, the bacterium is phagocytized by the alveolar macrophages, where it starts to replicate by inhibiting the phago-lysosome fusion process (Harada, *et al.*, 2007). The bacterium uses a peculiar T4SS to survive

in human cells. This system is very different from the T4ASS, as its component proteins have no homologies with the VirB/D4 system or other T4ASSs from other organisms. This system consists of 27 proteins, divided into two categories: Intracellular Multiplication (Icm) and Defective Organelle Trafficking (Dot). This system Dot/Icm system was thus classified as T4BSS (Segal, *et al.*, 1999, Juhas, *et al.*, 2008). Bioinformatic approaches have recently reported more than 100 different proteins as possible substrates for *L. pneumophila* Dot/Icm (Burstein, *et al.*, 2009). *L. pneumophila* Dot/Icm system remains largely unknown and very few of its proteins have been studied.

I.4.6 The Type 5 Secretion System (T5SS)

T5SSs are the simplest and most represented secretion mechanisms in bacteria. They include over 700 virulence factors which are involved in several important mechanisms such as adhesion, invasion, proteolysis, cytotoxicity, antibody resistance and cell-to-cell spreading (Ackermann, *et al.*, 2008). T5SS are divided into three main subfamilies: the Autotransporters (AT), the Two-Partners Systems (TPS) and the Oligomeric Coiled-Coil Adhesins (Oca).

AT proteins mediate their own translocation across the outer membrane without the help of accessory proteins, like ATPases, chaperones or adaptors commonly seen in other secretion systems, hence the term “Autotransporters” (Kajava & Steven, 2006). ATs are multi-domain proteins composed of an N-terminal signal peptide, a passenger domain and a translocator domain. Their delivery into the periplasm is Sec-dependent. The translocator domain of ATs is monomeric (300 residus) and forms a 12-fold β -barrel in the outer membrane. This β -barrel acts as a pore through which the passenger domain is secreted outside the bacterial cell (Desvaux, *et al.*, 2004, Henderson, *et al.*, 2004).

TPSs are slightly different from the ATs, in the sense that the passenger domain and translocator domain are present in different proteins, often referred as TPSA and TPSB (Jacob-Dubuisson, *et al.*, 2004).

Ocas are related to the ATs, but in contrast with them, the translocator domain (70 to 90 residues) of the Ocas oligomerizes in the outer membrane to form a trimeric 12-fold β -barrel (Ackermann, *et al.*, 2008). The translocation process of the passenger domain is still poorly understood, and three putative models have been hypothesized to address this issue. According to the “hairpin” model, the translocator domain first integrates into the outer membrane, then the passenger domain slides through it starting from either the C- or N-terminus (Oomen, *et al.*, 2004, Ackermann, *et al.*, 2008). The second model, or

“multimerization model”, requires the multimerization of the translocator domain to form a pore in the outer membrane (Veiga, *et al.*, 2002). The “OMP85” model requires the action of the outer membrane protein OMP85 to trigger the translocation of the passenger domain (Bernstein, 2007). One of the best structurally characterized T5SSs is the Oca system of the adhesion protein Hia from *Haemophilus influenzae*. This protein mediates bacterial cell contact with the respiratory epithelium (Meng, *et al.*, 2008).

I.4.7 Type 6 (T6SS) and Type 7 (T7SS) Secretion Systems

These two new secretion systems have been recently discovered in several pathogenic bacteria, though their role in pathogenesis is still unknown. T6SSs have been found in bacteria like *V. cholerae*, *Y. pestis*, *S. typhimurium*, *P. aeruginosa* and the plant pathogen *Rhizobium leguminosarum* (Bladergroen, *et al.*, 2003, Mougous, *et al.*, 2006, Pukatzki, *et al.*, 2009). The assembly of the system and the secreted effector proteins are still unknown, but two conserved proteins, Hcp (Hemolysin conjugated protein) and VgrG (Valine-Glycine repeat protein G) have been recently found to be associated with a functional T6SS in many bacteria and are perhaps involved in the assembly of pilus-like structures (Pukatzki, *et al.*, 2009)

T7SSs have been found in bacteria such as *Listeria monocytogenes*, *Staphylococcus aureus* and *Mycobacterium tuberculosis* (Simeone, *et al.*, 2009). The ESX-1 T7SS of *M. tuberculosis* has been partially characterized in recent years and delivers two small effector proteins, ESAT-6 and CFP-10. These proteins form heterodimers that promote a T cell-mediated immune response and cytolysis of macrophages (Gao, *et al.*, 2004).

I.5 Virulence factors of *H. pylori*

H. pylori produces a number of toxins, through which this bacterium is able to subvert the cellular functions of the host. Several of these products have been linked to the most relevant clinical outcomes of *H. pylori* infection, like gastric adenocarcinoma and ulcers (Wen & Moss, 2009). Four *H. pylori* virulence factors have been extensively studied : VacA, HP-NAP, the cag-PAI encoded T4ASS and its secreted protein substrate, CagA.

I.5.1 The Vacuolating Cytotoxin A (VacA)

VacA, the first toxin discovered in *H. pylori*, was found secreted in the culture broth and formed vacuoles in the cytoplasm of mammalian cells (Leunk, *et al.*, 1988). VacA is a protein of around 140 kDa. It contains a Sec-dependent N-terminal translocation signal and a C-terminal autotransporter-like domain, suggesting a secretion mechanism similar to the AT pathways (Letley, *et al.*, 2006). Both the secretion signal and the C-terminal domain are cleaved off during the secretion process, yielding a mature product of around 88 kDa. This protein is cleaved again, yielding two domains of respectively 33 kDa and 55 kDa that interact to form a stable complex (Moll, *et al.*, 1995). This complex, of around 900 kDa, has a flower shape that de-assembles at high or low pH and interacts with the host cell membranes, forming anion-specific channels (Czajkowsky, *et al.*, 2005).

VacA effects are multiple (Figure I.14). The channels it forms cause the leakage of ions, small sugars, urea and weak bases (like bicarbonate ions) that support bacterial growth *in situ* (Tombola, *et al.*, 2001). In the presence of weak bases VacA can be internalized in endosomes where it forms large vacuoles by interacting with the small GTPase Rab7, and subvert the formation of late endosomes and lysosomes (Molinari, *et al.*, 1997). It has also been shown that the formation of vacuoles allows the intracellular survival of *H. pylori* in the gastric epithelium and the establishment of a persistent infection (Terebiznik, *et al.*, 2006).

VacA has other cytotoxic effects, including mitochondrial damage and apoptosis. VacA activates pro-apoptotic proteins via ERK and p38 MAP kinases, causing the release of the Cytochrome C from the mitochondria, activation of the Caspase 3 and cell death (Yamasaki, *et al.*, 2006, Ki, *et al.*, 2008). VacA also influences the immune response in the gastric lumen. It has been shown that VacA inhibits the Nuclear Factor of Activated T-cells (NFAT), inhibiting the production of the proinflammatory cytokine IL-2 and down-regulating the expression of the IL-2 receptor- α (Gebert, *et al.*, 2003). In contrast with this immunosuppressive activity, VacA can also activate the expression of the proinflammatory cytokines IL-6 and TNF- α by mastocytes and induce the expression of the Cyclooxygenase 2 (COX-2) in neutrophils and macrophages (Supajatura, *et al.*, 2002, Hisatsune, *et al.*, 2007). These paradoxical effects might be required to keep gastric cells mildly damaged in order to acquire nutrients and support *H. pylori* growth but at the same time quench the inflammation process to prevent the arrival of too many immune cells at the site of infection.

The structure of the 55 kDa domain of VacA (p55) has been solved (Gangwer, *et al.*, 2007). It consists of an elongated β -helix formed by three parallel β -strands connected by

loops of various lengths (Figure I.14). A similar fold has been previously seen in the passenger domains of the toxins pertactin from *B. pertussis* and the Hemoglobin protease Hbp from *E. coli*. The C-terminus consists of a small domain of mixed α and β structural elements. Docking this structure into a 19 Å cryo-EM map resulted in a model for VacA oligomerization and anionic channel formation (Figure I.14).

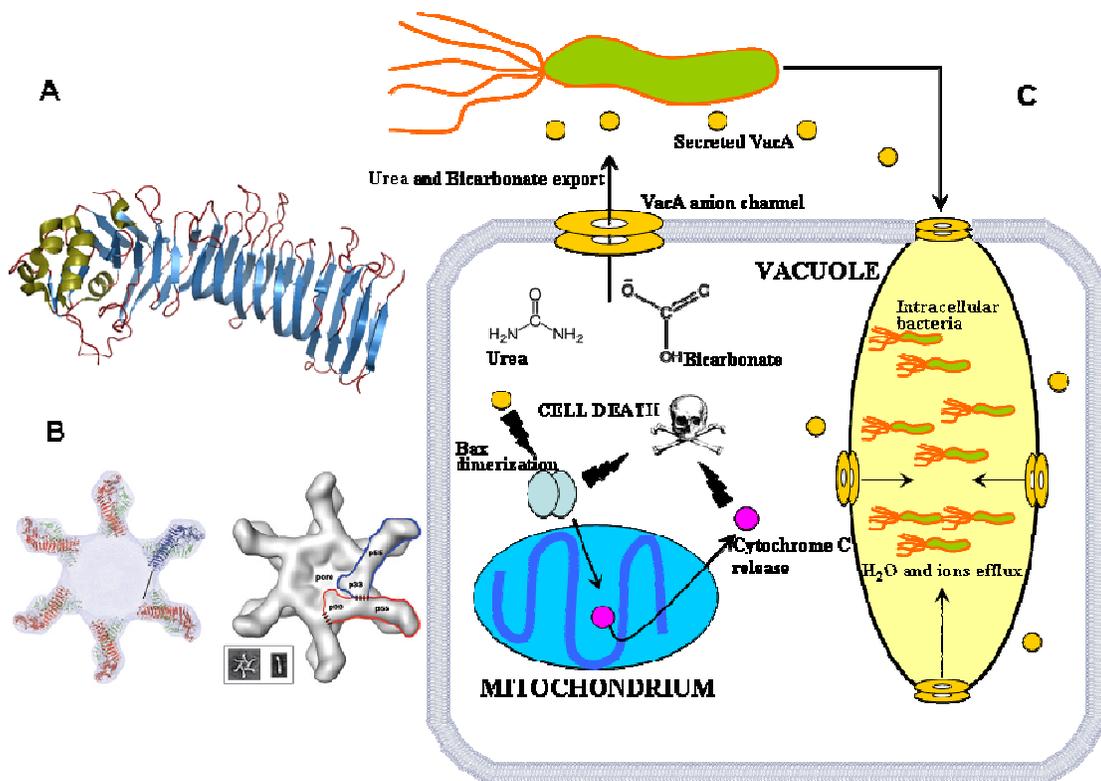


Figure I.14: Structure, oligomerization mechanism and cellular effects of *H. pylori* VacA

(A) Crystal structure of VacA p55 domain. Accession code: 2QV3 (Gangwer, *et al.*, 2007). (B) Double ring, dodecameric oligomer of VacA as seen in electron microscopy. VacA p55 occupies the external part of the ring, in a “flower petals” fashion (Gangwer, *et al.*, 2007). (C) VacA-mediated effects in gastric epithelial cells. VacA is secreted by *H. pylori* and forms a double ring anionic pore in the plasma membrane, that exports urea and bicarbonate ions to support the bacterial growth. Endocytized VacA can form vacuoles through a flux of water and ions in the late endosome, probably allowing the entry of bacteria in the cells. VacA also activates the dimerization of the proapoptotic protein Bax, that causes Cytochrome C release from the mitochondrion and apoptosis.

I.5.2 The Neutrophil-Activating Protein HP-NAP

This protein is a major proinflammatory factor secreted by *H. pylori* (D'Elia, *et al.*, 2007). HP-NAP, when secreted, forms dodecamers of around 150 kDa with a central cavity that is able to store up to 500 atoms of iron. The structure of the monomer, a four helix bundle (Figure I.15), presents some similarities with bacterioferritins and the DNA-binding proteins Dps, though no DNA binding has been demonstrated for HP-NAP (Zanotti, *et al.*, 2002). It is possible that this protein originally functioned as an iron-binding or an iron-regulated protein and then evolved to a immune cell activator to mediate inflammation, tissue damage and easy release of nutrients from the target cells (Dundon, *et al.*, 2001). HP-NAP induces the adhesion of leukocytes to the gastric endothelium through the activation of p38 MAPK and the induction of a high affinity state of local integrins. The protein has also been shown to stimulate the production of Reactive Oxygen Species (ROS), Tissue Factor and Plasminogen Activator Inhibitor 2 by neutrophils, monocytes and macrophages. In addition, it promotes a cytotoxic immune response mediated by the lymphocytes T helper Th1 through the induction of IL-12 and IL-23 in neutrophils and monocytes (Amedei, *et al.*, 2006).

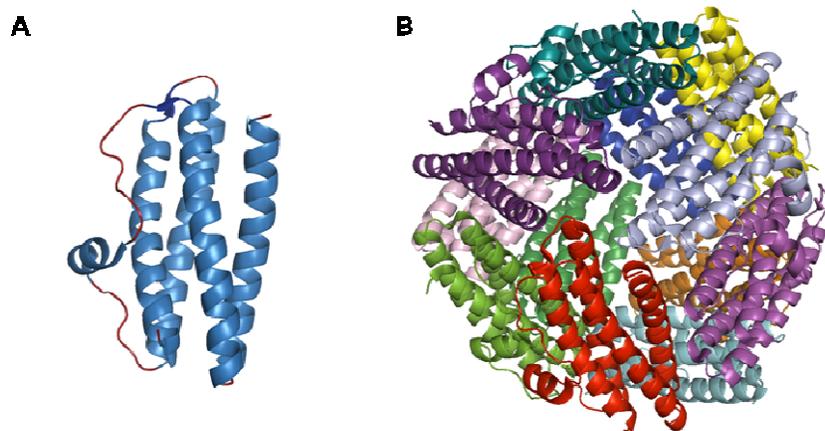


Figure I.15: structure of the Neutrophil Activating Protein (HP-NAP)

(A) Ribbon representation of the HP-NAP monomer. B) Ribbon representation of the dodecamer with each subunit colored differently. Accession code: 1J14 (Zanotti, *et al.*, 2002).

I.5.3 The *cag*-PAI T4ASS

During the early studies of *H. pylori* pathogenesis, the protein CagA was found to strongly stimulate antibody production in the host. Moreover, another protein CagE, was

identified as an homologue of a *B. pertussis* protein (Tsp) involved in toxin secretion (Cover, *et al.*, 1990, Tummuru, *et al.*, 1995). This led to the hypothesis that CagA could be a protein toxin secreted by the bacterium via an unknown secretion system. Further studies confirmed this and identified a 40 kb gene cluster in some *H. pylori* strains, carrying between 27 to 32 ORFs flanked by 31 bp direct repeats suggesting an horizontal gene transfer (Censini, *et al.*, 1996). This region has been named “*cytotoxin-associated genes Pathogenicity Island*” (*cag*-PAI) and was found later to encode for a VirB/D4-like T4ASS necessary to deliver the cytotoxin CagA into gastric cells (Covacci & Rappuoli, 2000). The *cag* system is a particular member of the T4ASS family, as it features two times as many proteins as the prototypical *A. tumefaciens* VirB/D4 system. The assembly and morphology of this apparatus are largely unknown. As listed in Table I.1, Cag β , Cag α , CagT, CagY, CagX, and CagE are sequence and functional homologs respectively to VirD4, VirB11, VirB7, VirB10, VirB9 and VirB4, whereas CagV, CagW, Cagy and CagC are considered to be functional orthologs of respectively VirB8, VirB6, VirB1 and VirB2 (Suerbaum & Josenhans, 2007). A recent study has identified CagL as a possible functional ortholog of VirB5 that interacts with the $\alpha_5\beta_1$ integrin on gastric cells to trigger the injection of CagA (Kwok, *et al.*, 2007)

<i>A. tumefaciens</i> Vir proteins	<i>H. pylori</i> Vir sequence homologs	<i>H. pylori</i> Vir functional orthologs
VirB1	Cagy (HP0523)	
VirB2		CagC (HP0546)
VirB3	N-CagE (HP0544)	
VirB4	C-CagE (HP0544)	
VirB5		CagL (HP0539)
VirB6		CagW (HP0529)
VirB7	CagT (HP0532)	
VirB8		CagV (HP0530)
VirB9	CagX (HP0528)	
VirB10	CagY (HP0527)	
VirB11	Cag α (HP0525)	
VirD4	Cag β (HP0524)	

Table I.1: reported sequence or functional homologies between *H. pylori cag* proteins and *A. tumefaciens* VirB/D4 proteins.

In a recent paper (Kutter, *et al.*, 2008), it was reported that the N-terminal portion of CagE (N-CagE) contains weak similarities with VirB3. It was therefore postulated that CagE might be a fusion protein, containing parts of both VirB3 and VirB4 proteins, respectively at the N-terminus and C-terminus. According to the putative subcellular localizations of the various *cag* proteins and the study of their interactions, an hypothetical structural model (Figure 1.16) of the *cag* T4ASS was proposed (Kutter, *et al.*, 2008).

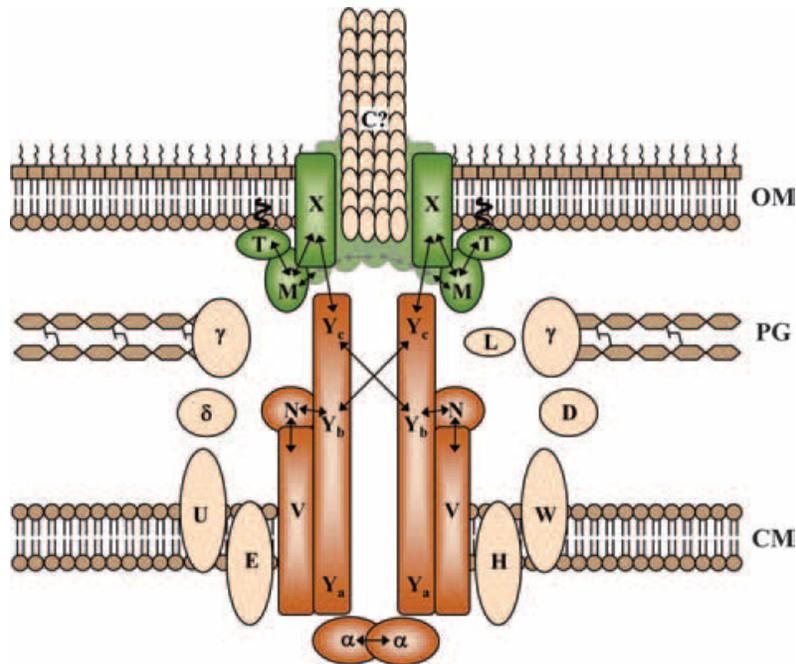


Figure 1.16: putative structural model of the *cag* T4ASS (Kutter, *et al.*, 2008)

Cag proteins are shown at their putative localizations according to bioinformatic and cellular fractionation data. OM: Outer Membrane; PG: peptidoglycan; CM: Cytoplasmic Membrane (Inner membrane). Every component is abbreviated so that Y, for example, is CagY. Direct interactions between proteins, as detected in the work of Kutter *et al.*, 2008, are indicated with arrows.

Structural studies have shed light on some of the *cag* T4ASS components. The structures of Cag α , homologue of VirB11 (Figure I.17), showed that the protein is a dynamic AAA+ ATPase (Yeo, *et al.*, 2000). The structure of CagD showed that the protein is dimeric protein and resembles the SycT chaperone from T3SS. CagD was found to be essential for CagA translocation and partially involved in pilus assembly (Cendron, *et al.*, 2009). The structures of CagS (Cendron, *et al.*, 2007) and CagZ (Cendron, *et al.*, 2004) were also solved but did not provide a clear function for these proteins (Figure I.17). However, CagZ might be involved in CagA translocation probably by interacting with other T4ASS components.

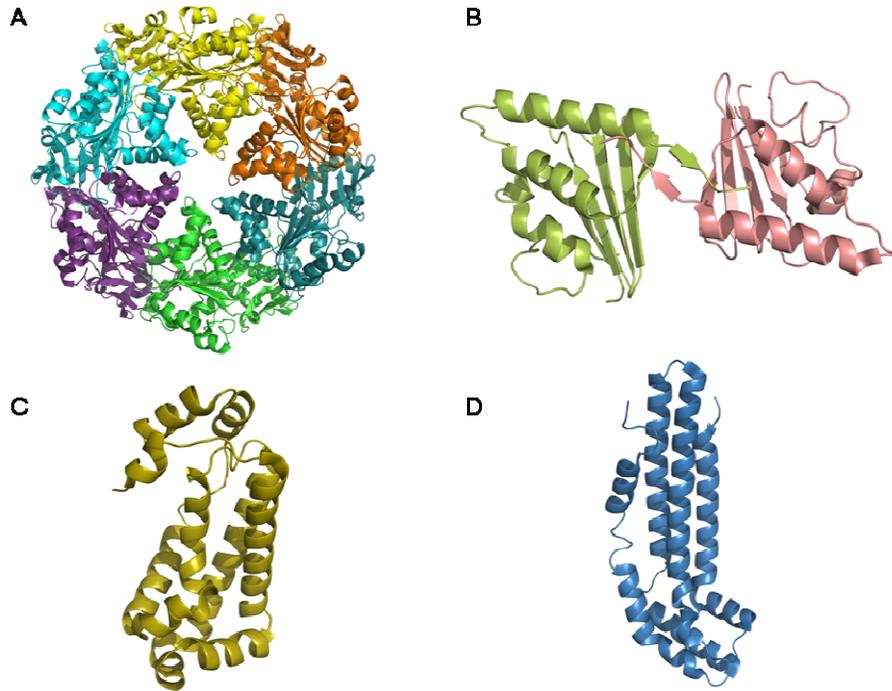


Figure 1.17: known structures of *cag* proteins.

(A): Structure of CagA, an hexameric ATPase homologue of the VirB11 ATPase of *A. tumefaciens* (Yeo, *et al.*, 2000). Accession code 1G6O. (B): The structure of CagD consists of four α -helices flanked on one side by a four-stranded antiparallel β -sheet. A small β -strand located at the C-terminus of the protein is involved in the dimerization (Cendron, *et al.*, 2009). Accession code 3CWX. (C): Structure of CagS, a monomeric, elongated protein composed by ten α -helices (Cendron, *et al.*, 2007). Accession code 2G3V. (D): Structure of CagZ, a monomeric, L-shaped helical protein, consisting of seven α -helices (Cendron, *et al.*, 2004). Accession code 1S2X.

CagA is the only protein known to be secreted by the *cag* T4ASS. However, there is evidence of cellular events that are independent of CagA (Figure I.18). T4ASS components have been found to activate the nuclear factor NF- κ B, AP-1 and the proto-oncogenes c-Fos and c-Jun, triggering the expression of the proinflammatory cytokines IL-8 and β -Defensin 2 (Meyer-ter-Vehn, *et al.*, 2000). Furthermore, the T4ASS activates the Rho GTPases Rac1 and CDC42, resulting in actin reorganization (Churin, *et al.*, 2001). Further studies have demonstrated that the T4ASS delivers peptidoglycan into gastric epithelial cells, which activates the pathogen recognition molecule Nod1 that in turn activates NF- κ B (Viala, *et al.*, 2004). Recently, it has been shown that the T4ASS activates JNK through the integrin β_1 , increasing the motility of gastric cells in a CagA- and peptidoglycan-independent manner

(Snider, *et al.*, 2008).

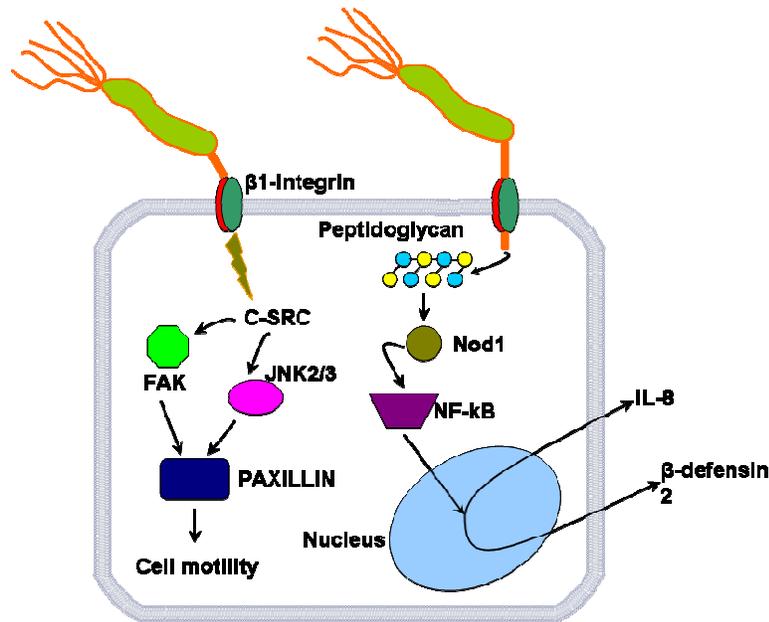


Figure I.18: *cag* T4ASS-dependent effects in gastric epithelial cells.

The T4ASS-pilus interacts with the human β 1-integrin via CagL, located on the pilus tip. The integrin then activates the c-Src kinase that in turn phosphorylates FAK and JNK proteins. These proteins then phosphorylate the protein Paxillin, that in turn initiates a signaling cascade leading to an increase in cell motility. Injected peptidoglycan also triggers a signaling cascade that starts with the activation by Nod1 of the nuclear factor NF- κ B. NF- κ B then migrates to the cell nucleus and induce the transcription of genes encoding the proinflammatory cytokines IL-8 and β -defensin 2.

I.5.4 The cytotoxin associate gene A (CagA) protein

CagA is probably the major toxin produced by *H. pylori* and is the only protein delivered inside the cytoplasm of gastric epithelial cells via the *cag*-T4ASS. CagA is a large protein of variable molecular weight (between 125 and 140kDa) that has no homology to any known protein. CagA secretion also remains poorly understood. Two motifs, located at both the N-and the C-termini of the protein, were proposed to function as translocation signals (Hohlfeld, *et al.*, 2006). The C-terminal translocation signal of CagA was mapped in the last 20 C-terminal amino acids, with several lysine residues. The sequence of the N-terminal translocation signal is within the first 57 N-terminal residues (Hohlfeld, *et al.*, 2006). Adjacent to the C-terminal translocation signal is a sequence of around 100 residues that has been shown to bind to the *cag*-PAI protein CagF. CagF has been found to be necessary for

CagA translocation and to function as a T4ASS chaperone-like protein (Pattis, *et al.*, 2007).

I.5.4.1 CagA tyrosine phosphorylation at EPIYA motifs

Once injected, CagA localizes in the inner leaflet of the plasmamembrane, where it undergoes tyrosine phosphorylation by SFKs (Src Family Kinases, like c-Src and c-Lyn) and the kinase c-Abl (Stein, *et al.*, 2002, Tammer, *et al.*, 2007). The phosphorylation occurs at sites localized at the C-terminal third of CagA and characterized by the presence of Glu-Pro-Ile-Tyr-Ala (EPIYA) motifs. These phosphorylation motifs are present in multiple copies among different *H. pylori* strains, therefore CagA proteins from different strains vary in length and molecular weight, ranging from 127 to 145 kDa and contain between 2 to 6 repetitions of the EPIYA motifs (Hatakeyama, 2003). Because of this variability, the EPIYA motif-containing region is termed the “Carboxy-Terminal Polymorphic Region” (Hatakeyama, 2009). According to the sequence flanking the EPIYA motif, four distinct segments have been identified, namely EPIYA-A, EPIYA-B, EPIYA-C and EPIYA-D (Figure I.19) . In Western *H. pylori* strains (*i.e* strains that are commonly found in clinical isolates from European and American patients), CagA proteins are characterized by the presence of EPIYA-A, EPIYA-B and from 0 to 4 copies of the EPIYA-C, whereas in East Asian strains (commonly found in Chinese, Korean and Japanese patients), the EPIYA-C is substituted by the EPIYA-D, which is always present as a single copy (Hatakeyama, 2003, Hatakeyama, 2009). A panel showing the make up of various CagAs is reported in Figure I.19.

and EPIYA-C motifs and termed Multimerization Domain (Ren, *et al.*, 2006, Kurashima, *et al.*, 2008) (Figure I.19).

Physiologically, SHP-2 is activated by the Grb2-binder (Gab) adaptor protein. For this reason, a first postulated model of CagA mode of action in gastric cells implied the functional mimicry of the eukaryotic Gab by CagA, possibly by structural similarity (Hatakeyama, 2003). Recently, this hypothesis has been confirmed using a *Drosophila melanogaster* model in which CagA has been shown to complement the action of the Gab-homolog DOS (Daughter of Sevenless) in the development of eye photoreceptors in a SHP-2-dependent manner (Botham, *et al.*, 2008).

Activated SHP-2 then activates the Ras/MAPK kinase cascade, resulting in a strong rearrangement of the cellular cytoskeleton and an increased cell motility, also termed cell scattering (Hatakeyama, 2002). CagA also downregulates the Focal Adhesion Kinase (FAK) in a SHP-2-dependent manner. Dephosphorylation of FAK results in a decreased interaction between the epithelial cells and the extracellular matrix. This, in turn, contributes to increase the cell scattering (Tsutsumi, *et al.*, 2006).

Another cellular effect mediated by CagA/SHP-2 interaction is the sustained activation of the Erk/MAPK pathway, resulting in a deregulation of cell-cycle (precisely the transition from phase G1 to phase S) and an increased cellular proliferation (Zhu, *et al.*, 2005).

I.5.4.3 Other phosphorylated CagA binding partners

While most phosphorylated CagA interacts with SHP-2, some also interacts with the C-terminal Src Kinase (Csk) and the adaptor protein Crk, which both contain SH2 domains. The interaction with Csk through the phosphorylated EPIYA-A and EPIYA-B inhibits the activity of the c-Src kinase by triggering the phosphorylation of Y527 located in the regulatory C-terminal domain, decreasing c-Src-dependent CagA phosphorylation through a negative feedback inhibition mechanism (Tsutsumi, *et al.*, 2003). The phosphorylation-dependent activation of Crk in turn induces the signaling pathways SoS1/H-Ras/Raf1 and C3G/Rap1/B-Raf, causing cell dissociation and cell scattering (Suzuki, *et al.*, 2005, Brandt, *et al.*, 2007).

I.5.4.4 Phosphorylation-independent interactions with host proteins

CagA also interacts with host cell proteins in a phosphorylation-independent manner. Binding partners include the tight junction scaffolding protein Zonulin (ZO-1), the cell adhesion protein E-cadherin, the tyrosine kinase c-Met, the E-Cadherin associated protein β -Catenin, the adaptor protein Grb2 and the Par1b/MARK kinase. These interactions result in tight junctions dysfunction, leading to loss of cell polarity and increased cellular proliferation, contributing to cell scattering (Wen & Moss, 2009).

Recently, the interaction between CagA and Par1b/MARK kinase has been characterized. Par1b/MARK is an important regulatory factor of cell polarity that localizes at the basolateral membrane of gastric epithelial cells. CagA multimerization domains were found to interact with the kinase domain of Par1b/MARK, thereby inhibiting the activity of the protein. This causes a wide disorganization of the epithelial monolayer due to the loss of cell-cell contacts and cell-cell proliferation inhibition (Saadat, *et al.*, 2007). Interestingly, the interaction with Par1/MARK seems to facilitate CagA dimerization and to stabilize CagA/SHP-2 interaction (Lu, *et al.*, 2009). CagA multimerization has been reported to occur only in mammalian cells and proven to be very labile and hard to quantify, thus raising the hypothesis that another cellular factor might have been involved in this process (Ren, *et al.*, 2006).

CagA has been shown to exert indirect transcriptional effects related to the production of proinflammatory cytokines. CagA activates NF- κ B and the Nuclear Factor of Activated T-cells (NFAT) through the Ras/Raf/Mek/Erk pathway, which respectively induce the transcription of proinflammatory genes like IL-8 and trigger a signaling cascade resulting in cell growth and differentiation (Yokoyama, *et al.*, 2005).

In conclusion, CagA is a pleiotropic protein, triggering a wide variety of cellular events by both mimicking host substrates (like the adaptor protein Gab) and deregulating several signaling cascades resulting in gastric cells morphological alterations and disruption of the epithelial functions. A summary of all CagA effects is presented in Figure I.20.

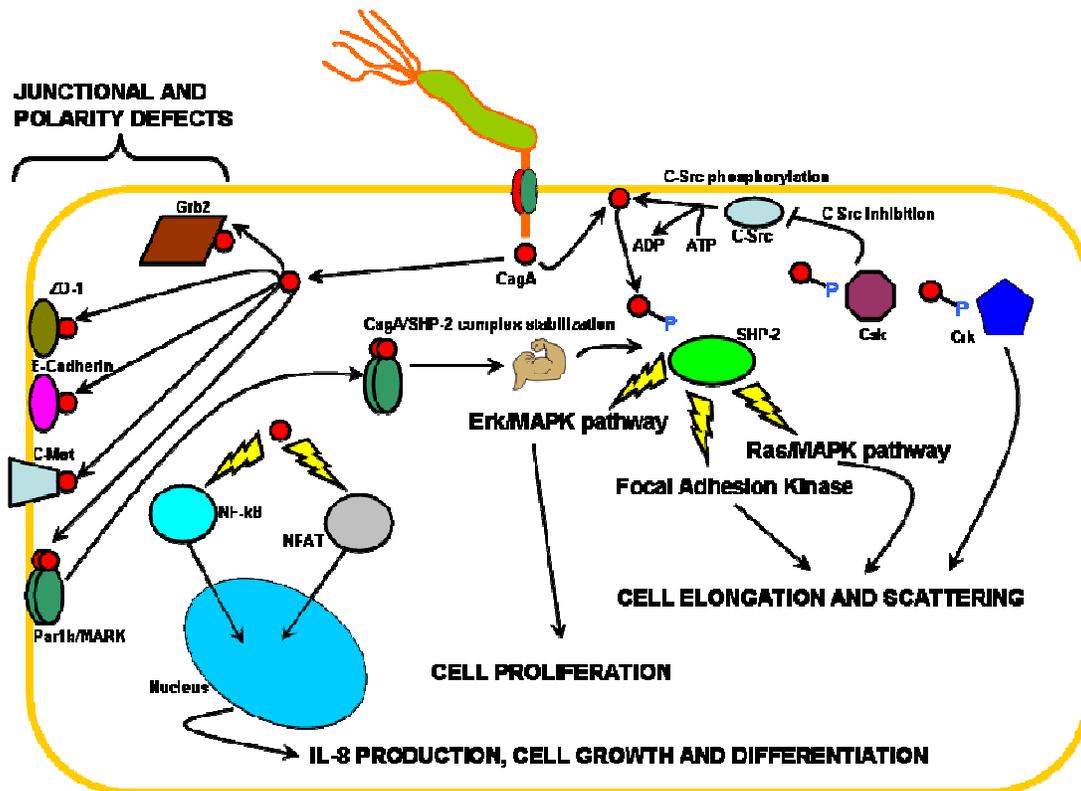


Figure I.20: roles of CagA in hijacking the host cell signal pathways

The *H. pylori* protein CagA (shown in red) has been shown to de-regulate the host signaling systems influencing cell morphology, proliferation and differentiation through both phosphorylation dependent and independent mechanisms. Once phosphorylated by SFKs (like c-Src), CagA activates the phosphatase SHP-2 and Crk, which both trigger several signal transduction pathways involved in cell elongation, proliferation and motility (cell scattering). A negative feedback regulation mechanism of CagA action involves the activation of Csk, which phosphorylates Y527 on the regulatory domain of c-Src and inhibits its kinase activity.

CagA also interacts, in a phosphorylation-independent manner, with proteins involved in cell-cell adhesion and polarity. These interactions destabilize the intercellular tight junctions, resulting in loss of epithelial cells polarity. Furthermore, the interaction with Par1b/MARK destabilizes and detaches Par1b/MARK from the membrane and the complex has been shown to stabilize the interaction between phosphorylated CagA and SHP-2. This mechanism probably involves the passive dimerization of CagA, establishing a stable complex with the dimeric Par1b/MARK.

CagA also induces the transcription of genes involved in proinflammatory cytokines production, cell growth and differentiation by interacting with and activating the nuclear factors NF-κB and NFAT.

I.5.4.5 CagA correlation with gastric cancer

CagA seems to have a direct correlation with gastric cancer. More precisely, chronic infection with CagA-harboring *H. pylori* strains result in progressive histopathological modifications of the epithelial cells of the stomach, resulting in atrophic gastritis, intestinal metaplasia and adenocarcinoma (Hatakeyama, 2009). Recently, a transgenic murine model has demonstrated a direct correlation between CagA and carcinogenesis *in vivo*. Mice constitutively expressing CagA developed gastric epithelial hyperplasia, gastric polyps and adenocarcinomas as well as myeloid leukemia and B-cell lymphomas (Ohnishi, *et al.*, 2008). Furthermore, all these pathogenetical events are dependent on Tyrosine phosphorylation, as mice expressing a phosphorylation-resistant CagA mutant did not develop any pathology.

The complex CagA-SHP-2 seems to play a critical role. Indeed, mutations and or polymorphism in the gene *PTPN11*, encoding SHP-2, result in a sustained dephosphorylation of downstream protein substrates. This deregulated SHP-2 activity increases the risk of gastric atrophy and gastric adenocarcinoma significantly (Hatakeyama, 2003). Therefore, CagA, by aberrantly activating SHP-2, provokes similar effects as SHP-2-deregulating mutations and contributes to gastric cancer establishment over the years. Another potential cancer associated effect of CagA is the deregulation of E-cadherin/ β -catenin signal pathways. These might have a direct role in the establishment of intestinal metaplasia and colorectal carcinogenesis, as this pathway is involved in intestinal cells differentiation (Murata-Kamiya, *et al.*, 2007).

Overview of the Thesis

From the introduction, it is clear that CagA is a major virulence factor of the pathogenesis of *H. pylori* and the establishment of gastric malignancies. Despite years of study, no structural information is available for CagA and few studies have addressed the biochemistry of this intriguing protein and of its numerous interactions with Host-cell proteins. This is mainly due to the difficulty to produce stable recombinant CagA.

A thorough investigation of this protein to overcome CagA production difficulties was initiated. First we have used a protein interaction approach using a fragment available in our laboratory to evaluate the possibility of co-expressing the protein with a partner in order to stabilize and crystallize it (Chapter 1). This approach has proved successful for other insoluble proteins (Hare et al., 2007). The experiments performed to validate protein-protein complexes between CagA and other *H. pylori* proteins are described in this thesis.

To explore further this protein, a novel high-throughput method was applied to generate a library of CagA fragments (Chapter 2). This study aimed at producing pure and stable protein fragments from CagA for further biochemical and structural studies. Using this work, fragments of CagA that could be purified efficiently were defined, in particular a C-terminal fragment of around 33kDa, containing important CagA features.

In Chapter 3, the study of this particular fragment, named CagA^{C33} and obtained from the methodology described in Chapter 2, is presented. This fragment contains numerous functional motifs, including the three tyrosine phosphorylation motifs (the EPIYA motifs), multimerization domains and CagF interaction sequence that are particularly important for CagA functions. Using a panel of Biochemical and Biophysics techniques, CagA^{C33} was characterized and several important new properties were identified. The work described here constitutes the first *in vitro* biochemical analysis ever performed for CagA, and will serve as an important step for structural studies and to study new aspects of CagA properties.

In Chapter 4, the determination of the first structures of Tip α by X-ray crystallography is described. Tip α was produced and crystallized in two different forms (I and II). The structure of Tip α was determined at 1.9Å using the single anomalous dispersion method with a Seleno-Methionine derivative crystal. The structure of Tip α and also additional experiments that were conducted based on this structure are described in this chapter. In the last section (General conclusions and Perspectives) of this manuscript, the results obtained are assembled and used to discuss their implications in *H. pylori* infection process.

Chapter 1

Biochemical and Structural studies of the H. pylori CagA Protein-Protein Interaction Map

Résumé

Les interactions protéine-protéine sont absolument essentielles au fonctionnement des cellules. Ainsi, de nombreuses études visant l'identification des « interactomes » ont été réalisées par des techniques à « haut débit » de biologie moléculaire, notamment celle du double-hybride dans la levure (Y2H).

La première carte d'interactions protéine-protéine bactérienne a ainsi été élaborée chez la bactérie *Helicobacter pylori*, pour laquelle environ 1285 interactions ont été identifiées en utilisant 285 « cadres ouverts de lecture » et une librairie initiale de deux millions de fragments du génome.

Dans ce travail de thèse, Je me suis intéressé aux interactions identifiées pour la seule protéine CagA d'*Helicobacter pylori*. L'objectif était 1) la validation de ces interactions par des méthodologies biochimiques *in vitro*, en utilisant des protéines recombinantes et 2) l'identification de complexes CagA/protéine permettant des études structurales. Sept protéines ont été sélectionnées pour ces études. Parmi celles-ci, celles dont la fonction est connue sont :

(a) RibA, une GTP cyclohydrolase II qui catalyse la conversion du GTP vers un substrat important pour la voie de biosynthèse des flavines;

(b) EstV, une lipase de type V qui joue probablement un rôle dans la colonisation, en dégradant les lipides de la muqueuse de l'estomac et augmentant ainsi l'accessibilité de l'épithélium;

(c) le domaine N-terminal de FeoB (FeoB¹⁻²²⁴), un transporteur de fer impliqué aussi dans le processus de colonisation de l'épithélium gastrique.

Les autres protéines de la carte, HP0249, HP0383, HP1015 et HP0431, n'ont pas de fonction établie. Nous avons cependant identifié des similarités de séquence entre HP0431 et les phosphatases de type 2C (PP2C) par une approche bioinformatique.

Ces protéines ont été exprimées dans *E. coli*. RibA a pu être purifiée en large quantité, alors que EstV et HP0431 ont été purifiées en faible quantité. Les autres protéines n'ont pas été exprimées ou n'étaient pas solubles.

Un fragment de CagA comprenant la partie centrale de cette protéine (résidus 392 à 793), CagA³⁹²⁻⁷³³, a été ensuite utilisé pour étudier les interactions avec RibA, EstV et HP0431 par différentes techniques biochimiques (far-western blot, le pull-down et la gel-filtration analytique). Les résultats montrent que RibA interagit très bien avec CagA³⁹²⁻⁷³³, alors que HP0431 n'interagit pas et qu'EstV interagit très faiblement. Les études

d'interactions par pull-down et gel-filtration analytique suggèrent que ces interactions sont transitoires et/ou pas suffisamment fortes pour former des complexes stables. Les résultats de ces expériences montrent donc que ces interactions, identifiées dans la carte d'interactions protéine-protéine de CagA, ne constituent pas un point de départ solide pour l'obtention d'un complexe entre CagA et un partenaire.

1.1 The *Helicobacter pylori* protein-protein interaction map

In recent years, increasing attention has been directed towards understanding protein-protein interaction networks in living cells. Most vital cellular functions (like signaling, growth and proliferation) rely on protein complexes and the characterization of the interactions made by a protein is a key element for understanding its function. The study of protein complexes has benefited from *in vitro* and *in vivo* high-throughput methods to identify protein-protein interactions and protein complexes at the genome scale. These methods include tandem affinity purification (TAP, (Gunzl & Schimanski, 2009)), native 2D gel electrophoresis (Klepsch, *et al.*, 2008) and yeast two-hybrid assays (Y2H, (Bruckner, *et al.*, 2009)). In particular, important results have been obtained with Y2H and large-scale libraries of Open Reading Frames (ORF) to retrieve possible binding partners (Walhout, *et al.*, 2000).

A bacterial Y2H protein-protein interaction map (PIM) has been established for *H. pylori*, and has resulted in the detection of 1285 interactions using 285 ORF baits and an initial library of 2 million prey (Rain, *et al.*, 2001). This map can be found on the Hybrigenics web server (<http://pim.hybrigenics.com>), which allows the retrieval of the binding partners detected for a given protein. The server reports a schematic view of the interactions network of a query protein with the query protein shown linked with a line to other loci of the putative interaction partners. In addition, based on the boundaries of interacting fragments, selected interacting domains (SIDS) can be identified. Based on the redundancy of the library and literature available, interactions are assigned a Predicted Biological Score (PBS) to evaluate their significance. One possible caveat of Y2H is that it generates both false positive and false negative results. For this reason, direct biochemical and/or structural studies are important to complement and validate these studies.

A biochemical analysis of 17 protein interactions reported in the PIM has been made using 31 *H. pylori* full-length proteins and protein SIDS (Terradot, *et al.*, 2004). In that study, 13 (76%) Y2H-detected interactions were confirmed by tandem affinity chromatography using Histidine and/or MBP-tagged proteins. Furthermore, one of these interactions involving the *cag* T4ASS ATPase VirB11 (HP0525) and a non T4ASS-related protein, HP1451, resulted in the crystallization and structural characterization of this complex (Figure 1.1), showing that the protein HP1451 is a regulator of the ATPase (Hare, *et al.*, 2007). These results suggest that the PIM is indeed a valid starting point for a further biochemical and structural characterization of protein-protein complexes.

In this thesis a similar approach was used to investigate the CagA protein-protein interaction map. The objective was to try to validate biochemically the putative interactions and to find CagA partners suitable for structural studies.

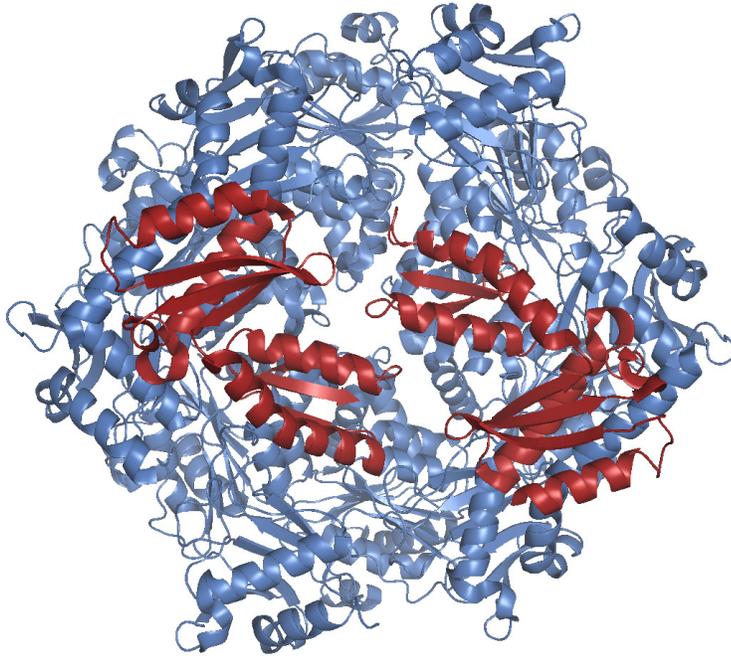


Figure 1.1: crystal structure of HP0525 (a VirB11 homolog of the *cagT4ASS*) in complex with the regulatory protein HP1451 (Hare *et al.*, 2007).

The region 92–264 of HP1451 (in red) interacts with HP0525 (in blue) at the N-terminus. Two HP1451 subunits interact with the ATPase at opposite sides, locking the enzyme in a closed conformation and thus modulating its activity. Accession code 2PT7.

1.2 The protein interaction map of CagA

Following submission of CagA (HP0547) to the Hybrigenics server, a total of 12 proteins of both known and unknown function were shown as putative interaction partners (Figure 1.2). Six of these partners have a reported functional and/or sequence homology with known proteins, whilst the other six have not been characterized. Among all targets, the proteins RibA and HP0496 (HpYbgC) show the strongest interaction (PBS score A, Figure 1.2).

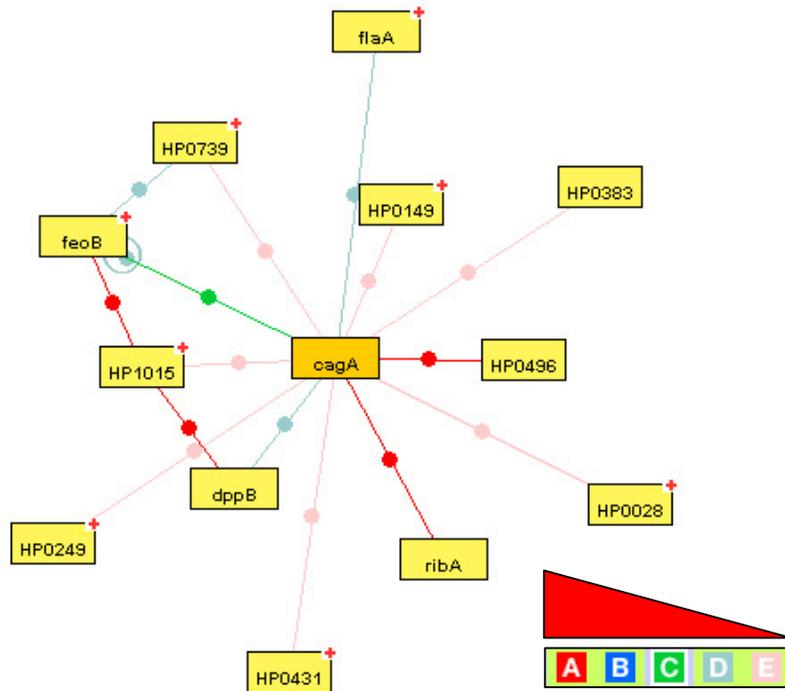


Figure 1.2: CagA protein-protein interaction map.

Putative interactions between CagA and other 12 *H. pylori* proteins have been detected by Y2H and assigned a Predicted Biological Score (PBS) from A (strongest interaction) to E (weakest interaction).

1.2.1 Hp0496 (HpYbgC)

The interaction between a CagA fragment, comprising the amino acids A392-D733 (CagA³⁹²⁻⁷³³), and HP0496 has been already investigated and confirmed by TAP and Far-Western Blot (Figure 1.3, Angelini *et al.*, 2008; Terradot *et al.*, 2004).

HP0496 has been crystallized and its crystal structure elucidated (Figure 1.3, panel C). HP0496 has been shown to be a homolog of the YbgC Thioesterase family (Angelini, *et al.*, 2008). This protein is associated with the Tol-Pal system, which is important in bacteria in the transport of macromolecules across the membranes, in the maintenance of membrane integrity and also in the cell division process (Cascales, *et al.*, 2002, Lazzaroni, *et al.*, 2002, Gerding, *et al.*, 2007). The biological relevance of this interaction is still unknown, and no other interaction targets of YbgC have been reported that may help to clarify it.

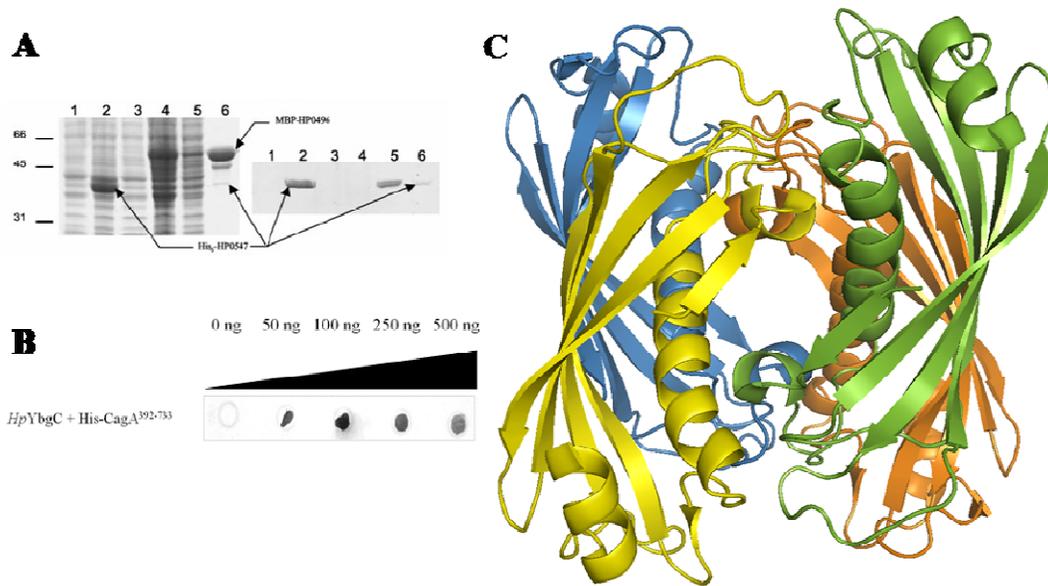


Figure 1.3: Structure of HpYbgC (HP0496) and its interaction with CagA³⁹²⁻⁷³³ (Angelini *et al.*, 2008; Terradot *et al.*, 2004).

Panel A: SDS-PAGE gel showing the co-purification of HP0496 (HpYbgC) and CagA³⁹²⁻⁷³³ in tandem affinity purification. Panel B: far-western blot confirming the interaction between HpYbgC (blotted on the membrane) and increasing amounts (from 0 to 500 ng) of His-CagA³⁹²⁻⁷³³. Dark spots indicate the presence of an interaction between the two proteins. Panel C: crystal structure of HpYbgC. The protein shows a tetrameric arrangement which is typical of the YbgC Thioesterases family. The monomer is characterized by a hot-dog fold, where a long α -helix is surrounded by a four-stranded antiparallel β -sheet. Accession code: 2PZH.

1.2.2 RibA (HP0802)

RibA was identified after the observation that various strains of *H. pylori* could lyse human erythrocytes when grown on blood agar, and that this phenotype was mediated by RibA (Bereswill, *et al.*, 1998). RibA is a GTP-Cyclohydrolase II, an enzyme that catalyzes the conversion of GTP into DARP (2,5-diamino-6-ribosylamino-4-pyrimidinone-5'-phosphate), a precursor of flavin biosynthesis in many bacteria, yeasts, fungi and plants. The *H. pylori* enzyme shows a very high degree of homology (50% identity, 70% similarity) with other RibA proteins and it has been shown biochemically that *H. pylori* RibA is able to complement RibA deletion mutants of *E. coli*, thus confirming that these enzymes have the same function (Bereswill, *et al.*, 1998).

Only the *E. coli* RibA protein has been crystallized and its structure elucidated (Ren, *et al.*, 2005). The protein showed a novel fold characterized by a central core of seven antiparallel β -sheets surrounded four long and three short α -helices interconnected by loops (Figure 1.4).

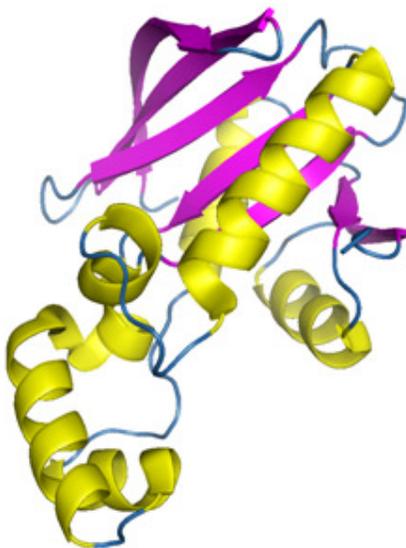


Figure 1.4: crystal structure of *E. coli* RibA (Ren, *et al.*, 2005).

The protein is rather elongated and characterized by a novel fold. Seven antiparallel β -sheets (purple) are surrounded by four long and three short α -helices (yellow). The loops connecting the various structural elements are shown in blue. Accession code: 2BZ1.

1.2.3 HP0739 (EstV): The first *Helicobacter pylori* Type V Lipase

Recently HP0739 was identified as a Type V lipase, named EstV, the first discovered in *H. pylori* and in all ϵ -proteobacteria (Ruiz, *et al.*, 2007). Bacterial lipases are divided into eight different families, based on conserved sequence motifs and their enzymatic properties (Arpigny & Jaeger, 1999). Recently, they have been extensively studied, due to their potential biotechnological application and their role in bacterial virulence (Jaeger & Eggert, 2002). *H. pylori* EstV has been reported to have a role in the onset of gastric ulcers, degrading the lipids in the stomach mucus and thereby making the gastric epithelium more accessible for colonization (Ruiz, *et al.*, 2007). Furthermore, the activity of EstV is inhibited by the anti-ulcer drugs β -aescin and glycyrrhizic acid. For these reasons, this protein might be considered as a new type of virulence factor of *H. pylori* (Ruiz, *et al.*, 2007), although more studies are needed to confirm this.

1.2.4 The high affinity iron transporter FeoB (HP0687)

HP0687 is an homologue of the ferrous iron transporter FeoB in *E. coli* and plays an essential role in iron transport in *H. pylori*. Furthermore, it has been suggested that FeoB is also involved in gastric colonization because, in murine models, FeoB-deficient strains of *H. pylori* fail to colonize the gastric mucosa (Velayudhan, *et al.*, 2000). Interestingly, sequence alignments show that the N-terminus of *E. coli* and *H. pylori* FeoB proteins are conserved and hydrophilic, and show some features common to G-proteins. In particular, submitting the amino-acid sequence of *H. pylori* FeoB to the topology prediction server TMHMM (Krogh, *et al.*, 2001) suggested that the residues from 1 to 224 are predicted to be intracellular (Figure 1.5).

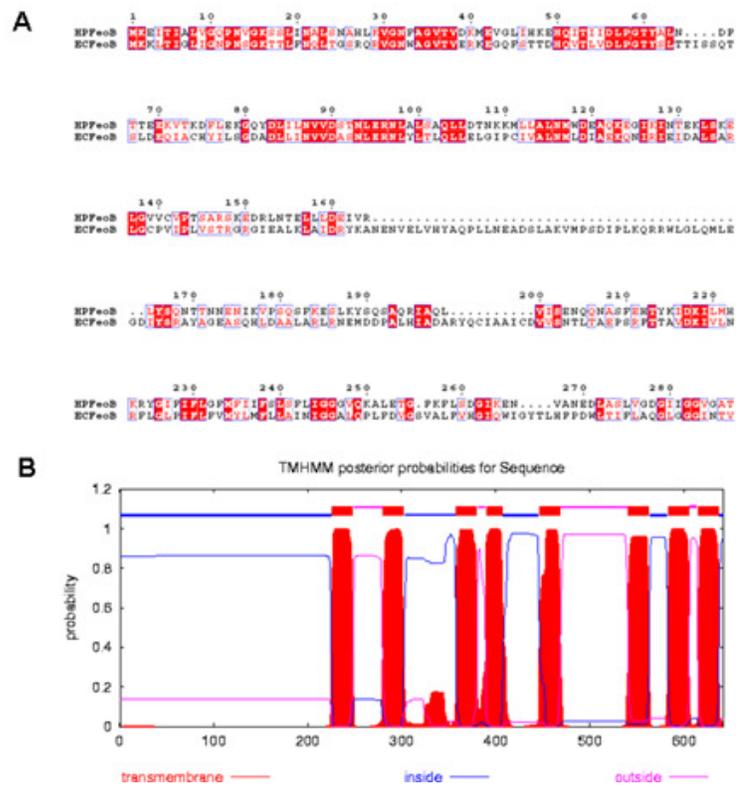


Figure 1.5: sequence alignment of *H. pylori* and *E. coli* FeoB proteins (A) and topology prediction of *H. pylori* FeoB (B).

(A) Sequence alignment of FeoB from *H. pylori* and *E. coli*. The sequence alignment shows that the N-terminus of both proteins is conserved, with a total 32% of identity and 60% of similarity. (B) Topology prediction of *H. pylori* feoB performed using the TMHMM server, showing the presence of 9 transmembrane helices, the first one starting after the residue 224. The sequence comprising residues 1 to 224 is predicted to be inside the cell.

1.2.5 The putative proteins HP0249, HP0383, HP0431, HP1015

There is no information available in the literature concerning the function of the putative CagA interacting partners HP0249, HP0383, HP0431, HP1015. Submitting the primary sequences of HP0249, HP0383 and HP1015 to the server protein-BLAST (<http://blast.ncbi.nlm.nih.gov>) does not report a significant sequence similarity with other proteins coming from organisms different than *H. pylori*, showing that these proteins are only present in *H. pylori* and closely related species.

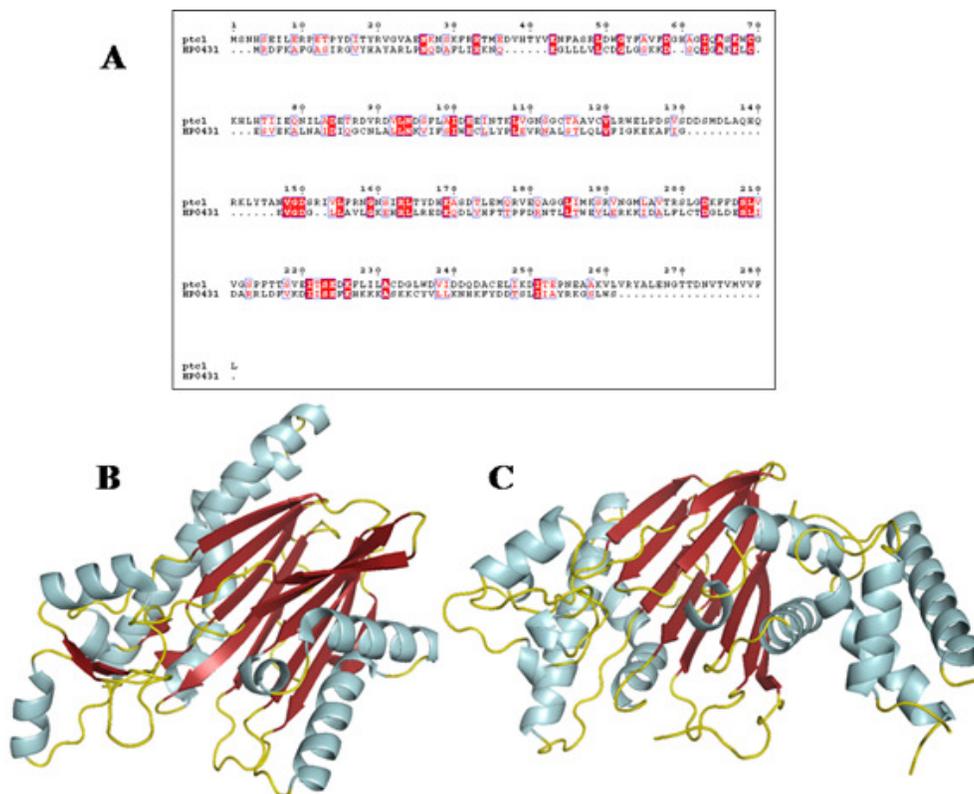


Figure 1.6: sequence alignment of HP0431 and Ptc1 and crystal structures of PP2C phosphatases from *Toxoplasma gondii* and *Homo sapiens*.

(A): the alignment shows that HP0431 and the yeast PP2C Ptc1 share a significant sequence homology (15% identity, 35% similarity). (B) and (C): the structures of PP2Cs from *T. gondii* (B) and *H. sapiens* (C) are characterized by a double layer of antiparallel β -sheets flanked by two helical bundles. The double β -layer is the only part of the protein whose fold is similar in both structures.

Accession codes: 3D8K for the PP2C from *T. Gondii* (Damodharan et al., to be published), 1A6Q for the *H. sapiens* PP2C (Das, *et al.*, 1996).

However, an amino acid sequence alignment using the algorithm blastp in the server protein-BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) reports a significant similarity between HP0431 and the Yeast protein Ptc1 (15% identity, 35% similarity), a member of the Protein Phosphatase 2C (PP2C) Superfamily Figure 1.6). PP2Cs belong to an evolutionarily conserved group of monomeric Serine/Treonine phosphatases, whose members have been best characterized in *S. cerevisiae*. The first member of these phosphatases, Ptc1, functions in yeast as a stress response protein and is involved in several regulatory pathways, mediating cell wall integrity, tRNA splicing, pH and heavy metal tolerance (Gonzalez, *et al.*, 2006, Gonzalez, *et al.*, 2009).

The role of HP0431 is completely unknown. However, recent reports have demonstrated the presence of PP2C-like proteins in many bacteria, all involved in stress response pathways (Halbedel, *et al.*, 2006, Osaki, *et al.*, 2009, Pane-Farre, *et al.*, 2009). Thus, it is possible that HP0431 might also function as a stress response regulator, involved in a still uncharacterized signaling pathway.

Two crystal structures of PP2C family members are available, one from *Homo sapiens* and one from *Toxoplasma gondii*. The two structures are shown in Figure 1.6. An alignment performed with the sequences of the three protein did not report any significant similarity, showing that the proteins do not have any conserved functional motif (data not shown).

1.3 Cloning, expression and purification of the putative interacting partners

1.3.1 Cloning of the interacting partners

The genes encoding the putative interacting partners of CagA were amplified by PCR using the genomic DNA of *H. pylori* strain 26695 as a template. The lengths of the genes, the sizes of the proteins and the primers used for PCR amplification are shown in Figure 1.7A. The amplified constructs were isolated by band excision from an agarose gel and purified with the “Qiaquick Gel Extraction Kit”, according to manufacturer procedure (Qiagen). The genes were then cloned into the expression vector pET151/D- TOPO (see Appendix A.1) following the manufacturer’s procedure (Invitrogen). *E. coli* strains 10G Duos (Lucigen) were transformed with the ligation products by heat shock at 42°C for 40 seconds, plated on solid agar supplemented with 100 µg/ml of ampicillin and grown overnight at 37°C. Single

colonies have been amplified in 10 ml of liquid LB medium supplemented with 100 µg/ml of ampicillin overnight at 37°C. The plasmids were purified using the “Qiaprep Spin Miniprep” kit, according to the manufacturer’s instructions (Qiagen).

The insertion of the genes in the expression vectors was confirmed by digesting 100 ng of plasmids with the restriction enzymes *NcoI* and *SacI* (Fermentas) for 4 hours at 37°C and loading the digestion products on an agarose gel. Positive clones were checked by sequencing of the genes (Genecore, EMBL of Heidelberg, Germany).

1.3.2 Expression of the target proteins

After transformation of *E. coli* strains BL21 (DE3), the proteins were expressed in 1 litre of Luria Bertani (LB) broth supplemented with 100 µg/ml of ampicillin at 37°C, after induction with 1 mM of IPTG. Cells were centrifuged at 7000g for 20 minutes, resuspended in 50 mM Tris pH 8, 200 mM NaCl, 5% glycerol, lysed using a cell disrupter (Constant System) and then centrifuged at 40000 g for 20 minutes. The soluble fraction was loaded onto gravity columns filled with 2 ml of Ni-NTA resin to perform Nickel affinity chromatography purification. Washes were performed first with buffer A (50 ml of 50 mM Tris pH 8, 200 mM NaCl, 5% glycerol) and then with 10 ml of buffer A supplemented with 50 mM imidazole. Elution of the proteins was performed with buffer A supplemented with 500 mM imidazole. The presence of the proteins in induced cells and purified fractions has been monitored by western blot analysis using primary anti-His antibodies grown in mouse (Sigma) and secondary antibodies conjugated to the alkaline phosphatase (Sigma). The presence of secondary antibodies was detected by colorimetry using 1 to 4 ml of the alkaline phosphatase NBT/BCIP as substrate.

The results obtained are presented in Figure 1.7. RibA, HP0383, HP0431, EstV and FeoB¹⁻²²⁴ were successfully expressed in *E. coli* cells. HP0249 and HP1015 could not be detected by western blot analysis of induced cells and thus were not expressed. RibA, EstV and HP0431 were also found in the purified fraction from Ni-NTA chromatography (Figure 1.7). HP0383 and FeoB were not detected by western blot in the purified fraction and are thus not soluble or did not bind to the affinity column.

Target	Gene length	Protein Mw	Primers	Expressed	Purified
RibA	579 bp	21.7 kDa	Fw: CACCATGAAACGATTAGAAGTTT CTAACC Rv: TCACAATAAAATGCCCCATTTTGAG	YES	+++
EstV	726 bp	27.5 kDa	Fw: CACCATGCCCCAAACGCCAGTATCCG Rv: CTAAGACTTTGCATGATGATAATTC	YES	+
FeoB ¹⁻²²⁴	672 bp	22.5 kDa	Fw: CACCATGAAAGAAATCACTATGCC CTTG Rv: TTACTTGTGCATTAAAATCTTATC	YES	NO
HP0249	540 bp	20.3 kDa	Fw: CACCATGGCATCTCTTGCCTTTATC Rv: TTATAAACAAAACAAAGAACTTGAT AAG	NO	NO
HP0383	525 bp	19.8 kDa	Fw: CACCATGCCCCATTCTTTTAAAAAG CG Rv: TTAITGGATTTTATAAAGCCACCTAG	NO	NO
HP0431	687 bp	26 kDa	Fw: CACCATGAGAGATTTTAAAGCGTTTG GG Rv: CTAACCTCCATAAACTACCTTTTC	YES	+
HP1015	507 bp	18.7 kDa	Fw: CACCGTCTTTTTATTGTTTCAAGCGG Rv: TCATGCTTTCAAAAAGACTTGTG	NO	NO

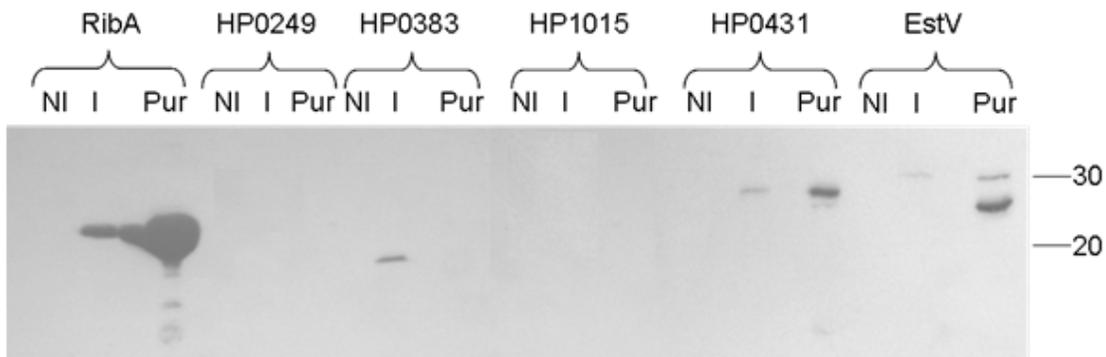


Figure 1.7: Summary of the cloning, expression and purification of CagA putative interacting partners. (A): table summarizing targets' gene length, molecular weight, primers used for amplification of the constructs and purification of the proteins. (B): western blot analyses of the samples corresponding to the non-induced bacteria (NI), IPTG induced bacteria (I) and purified fractions (Pur) for all interacting partners except FeoB¹⁻²²⁴.

1.4 CagA-target protein interaction studies

At the start of this project, the only fragment of CagA that could be expressed and purified was CagA³⁹²⁻⁷³³ (Terradot *et al.*, 2004). This construct was therefore used to test CagA interactions with purified RibA, HP0431 and EstV using a far western blot technique.

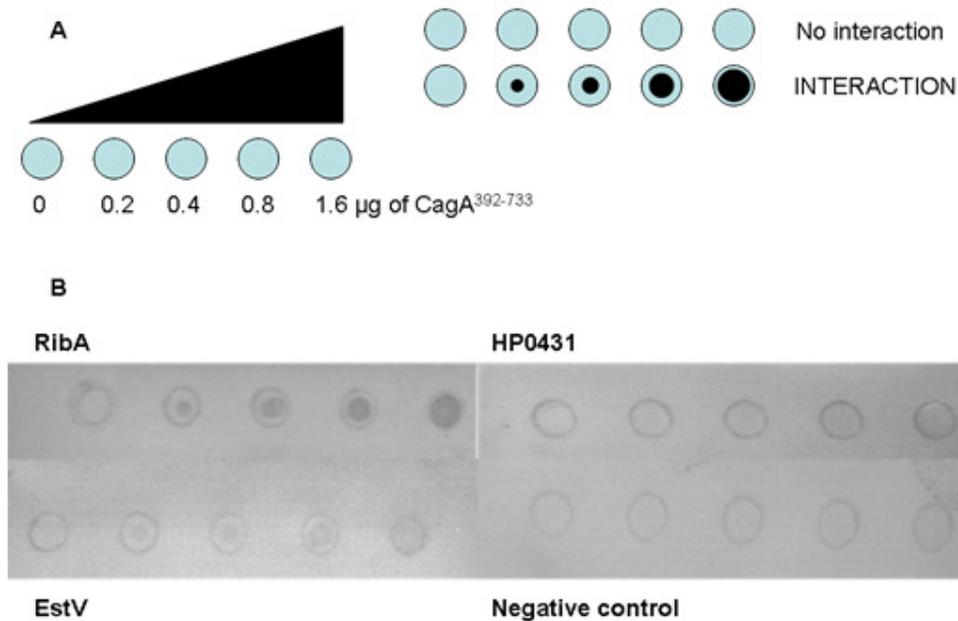


Figure 1.8: far western blot assay of potential interactions of RibA, HP0431 and EstV with CagA³⁹²⁻⁷³³.

(A): schematic overview of the experimental setup: from 0 to 1.6 µg of CagA³⁹²⁻⁷³³ have been blotted onto a nitrocellulose membrane and incubated with the interaction targets. After detection of bound proteins with antibodies, a dark spot corresponds to an interaction between CagA³⁹²⁻⁷³³ and the interacting partner. (B): results of RibA, HP0431 and EstV interaction assay. A negative control with no CagA³⁹²⁻⁷³³ blotted on the membrane has been performed to confirm the assay's specificity.

The following protocol was used with the three purified putative interacting partners: EstV, HP0431 and RibA. 0.2, 0.4, 0.8 and 1.6 µg of previously purified CagA³⁹²⁻⁷³³ were blotted onto a nitrocellulose membrane (Biorad). The membranes were then saturated by incubation into a Phosphate Buffer Saline pH 7.5 (PBS) buffer supplemented with 5% BSA. After three washes with PBS buffer, the membranes were incubated for 5 hours at 4°C with 5 ml of PBS containing the His-V5-tagged interacting partner at 0.3 mg/ml. The membranes were then washed twice with phosphate buffer and incubated with a primary antibody directed against the V5 epitope (Sigma), located upstream the His-tag on the pET151/D-

TOPO plasmid (see Appendix A.1). Detection of the anti-V5 antibodies was performed with a secondary antibody conjugated with alkaline phosphatase, followed by colorimetry with the substrate NBT/BCIP. The results, presented in Figure 1.8, show a clear interaction for RibA, while a weak interaction is detected for EstV. HP0431 does not appear to interact with CagA³⁹²⁻⁷³³.

1.5 Conclusion and discussion

Most of the *H. pylori* CagA interacting partners from the protein interaction map examined here have shown poor expression levels and only three of them have been purified in amounts sufficient for preliminary interaction assays (Figure 1.7). To test the interactions, the fragment CagA³⁹²⁻⁷³³ (Terradot *et al.*, 2004) was used. Of the three purified targets, only RibA and EstV showed a positive interaction with CagA³⁹²⁻⁷³³ as shown by the far western blot experiments.

CagA interaction with RibA

Our study shows that RibA interacts with CagA³⁹²⁻⁷³³ in far western blot. However, other techniques used (analytical gel-filtration chromatography and co-purification) have failed to give any confirm this interaction (data not shown). Far-western blot analysis is a very sensitive assay that can detect the formation of weak complexes and it can thus be concluded that the interaction between CagA and RibA is too weak ($K_d \geq \mu\text{M}$) to be detected using low-sensitivity techniques. This suggests that the affinity of RibA for CagA³⁹²⁻⁷³³ is rather low. As we used just a rather small CagA fragment, encompassing residues 392 to 733, it is possible that another CagA region might also be involved in the binding and in the stabilization of this interaction. However, at present, the impossibility of purifying a recombinant, full-length CagA makes this hypothesis hard to verify.

The physiological relevance of the CagA-RibA interaction is questionable. RibA is a metabolic enzyme and its activity is specific to GTP (Ren *et al.*, 2005). There is no evidence that leads one to think that RibA might use GTP or any other substrate to directly modify CagA. There are no reports in literature that show any post-translational modification of the CagA protein inside *H. pylori*. Furthermore, an interaction between CagA and the RibA substrate, GTP, has never been reported, so it is rather difficult to propose an hypothesis on

the function of the CagA/RibA interaction. For now, it is impossible for to conclude whether this interaction between CagA and RibA, first assessed in the Y2H protein-protein interaction map and then partially supported by our work, is physiological or artifactual.

CagA interaction with EstV

Similar conclusions to those concerning RibA can be drawn for EstV. The results shown here indicate that this protein interacts very weakly with CagA³⁹²⁻⁷³³. However, the domain of CagA reported to interact with EstV is comprised between the residues 851 and 1105 (Rain *et al.*, 2001), which is located at the C-terminus of CagA and its not included in the construct used here. For this reason, it is possible that the weak interaction we have assessed for EstV might be an artifact. As with RibA, the physiological relevance of the putative interaction between CagA and EstV is unknown. It is possible to speculate that, as CagA C-terminal portion has been reported to interact with the membrane of human gastric cell (Higashi, *et al.*, 2005), EstV and CagA might have lipid-binding domains that can promote their reciprocal interaction in *H. pylori* cells.

As seen from this study it was difficult to validate the interaction detected in the PIM since, at the time this study was carried out, only a small fragment of CagA (CagA³⁹²⁻⁷³³) could be purified and used to assay potential interactions. Clearly, the limit in this approach was the availability of purified CagA protein, in particular of region of interest such as the C-terminal portion (interacting with EstV). As will be seen in Chapters 2 and 3 of this work, several other fragments of CagA have now been purified. Future work, especially concerning CagA-EstV interactions, should centre on the new fragments purified.

Chapter 2

Expression of Helicobacter pylori

CagA soluble domains by library-

based construct screening

Résumé

La protéine CagA est la toxine majeure de *H. pylori*. Malgré la grande quantité de travaux décrivant les fonctions de cette protéine *in vivo*, la difficulté à exprimer cette protéine par voie recombinante a empêché de véritables études biochimiques et/ou structurales. D'autre part, la séquence de CagA ne montre pas de similarité avec des protéines connues, rendant également difficile l'identification *in silico* de domaines pouvant être exprimés et purifiés séparément.

Nous avons donc décidé d'utiliser une méthodologie expérimentale à haut débit qui a permis de fragmenter le gène afin de créer une librairie de clones. Cette librairie a ensuite été testée pour l'expression protéique dans *E. coli*. Cette méthodologie, appelée ESPRIT (Expression of Soluble Proteins through Random Incremental Truncation), a été développée à l'European Molecular Biology Laboratory'' (EMBL) de Grenoble, sous la supervision du Dr. Darren Hart.

Nous avons opté pour la troncation systématique du gène *cagA* à partir de l'extrémité 5', afin d'obtenir des fragments de CagA contenant le domaine C-terminal. Nous avons assemblé une librairie de 18432 fragments, parmi lesquels 40 ont été utilisés pour des tests d'expression à grande échelle. Deux de ces fragments, de poids moléculaires de 29 et 33 kDa, ont montré une solubilité satisfaisante et ont pu être purifiés. Les fragments plus larges étaient caractérisés par une dégradation importante de l'extrémité C-terminale durant l'expression dans les bactéries. Ces résultats nous ont permis de proposer un site commun de protéolyse de ces fragments, localisé au niveau d'une séquence de cinq asparagines (aminoacides 887-992). Pour valider cette hypothèse, deux fragments ont été clonés, exprimés, et purifiés sans la région C-terminale. Ainsi, ces fragments, dont les poids moléculaires sont respectivement de 48 et 53 kDa, n'ont pas montré de dégradation et ont pu être purifiés. CagA peut donc être décrite comme une protéine contenant deux domaines principaux, un domaine N-terminal d'environ 100kDa (CagA^{N100}) suivi d'un domaine C-terminal de 33kDa (CagAC33), tous deux reliés par une séquence de cinq asparagines. Les résultats de cette partie de la thèse ont été publiés dans le journal ''*Febs Journals*'' ([Angelini, Tosi et al., 2009](#)) (Co-premiers auteurs).

2.1 Construction of soluble fragments of CagA: overview of the project

Due to the role of CagA in gastric pathogenesis and its correlation with the onset of gastric cancer, this protein can be considered as one of the most important virulence factors expressed by *H. pylori*. Despite the availability of literature concerning the characterization of the mechanisms of action of CagA *in vivo*, its biochemistry is poorly understood, mainly because of the instability of the recombinant protein in *E. coli* cells (Pattis, *et al.*, 2007). This intrinsic resistance of CagA to *in vitro* characterization has resulted in the absence of crystal or solution structures of either the full-length protein or of any of its domains. Moreover, submitting CagA sequence to homology prediction servers gives no significant results. Therefore it has been difficult to predict the domain boundaries of this protein through direct sequence alignments and/or to manually design CagA constructs to be tested for expression, purification and structural characterization.

The utilization of high-throughput methods in structural genomics projects has become routinely applied to clone, express and purify a wide range of targets to be tested for crystallization (Burley, 2000). While these robotic techniques increase the chances of rapidly obtaining soluble proteins suitable for crystallization and structure determination, they also result in the loss of many interesting targets due to problems such as low expression yield, low purifiability, degradation, insolubility.

To maximise the chances of the successful production of stable, soluble proteins, structural genomics techniques are often used in parallel on several homologous targets from different organisms. The high-throughput methodologies developed for structural genomics have also been applied to single proteins. In particular, combinatorial library-based strategies, such as point mutations or construction of deletion fragments, have been successfully applied to artificially generate diversity in a single target protein and have been used, for example, to mutate an industrial enzyme to be more active, to generate a high-affinity binding protein/peptide for a given target (e.g. a receptor or drug target), or even to find a physiologically relevant protein domain to be crystallized (Bornscheuer & Pohl, 2001, Tarendeau, *et al.*, 2007).

A library-based screening technology recently developed in the laboratory of Dr. Darren Hart (EMBL, Grenoble) was used to identify soluble and potentially crystallizable CagA domains. This strategy was designed to keep the C-terminal part of CagA as the main

target, since it contains numerous features of the protein, including phosphorylation motifs, multimerization domain, CagF binding sequence and translocation signal (See Introduction).

In this method, the “Expression of Soluble Proteins by Random Incremental Truncation” (ESPRIT, (Tarendeau, *et al.*, 2007), the enzymes Exonuclease III and Mung Bean nuclease were used to truncate the *CagA* gene from the 5' extremity base by base, resulting in the production of a randomized library of 5'-truncated constructs. These constructs are then tested for expression and solubility in *E. coli*. A detailed explanation of the methodology is reported in Figure 2.1.

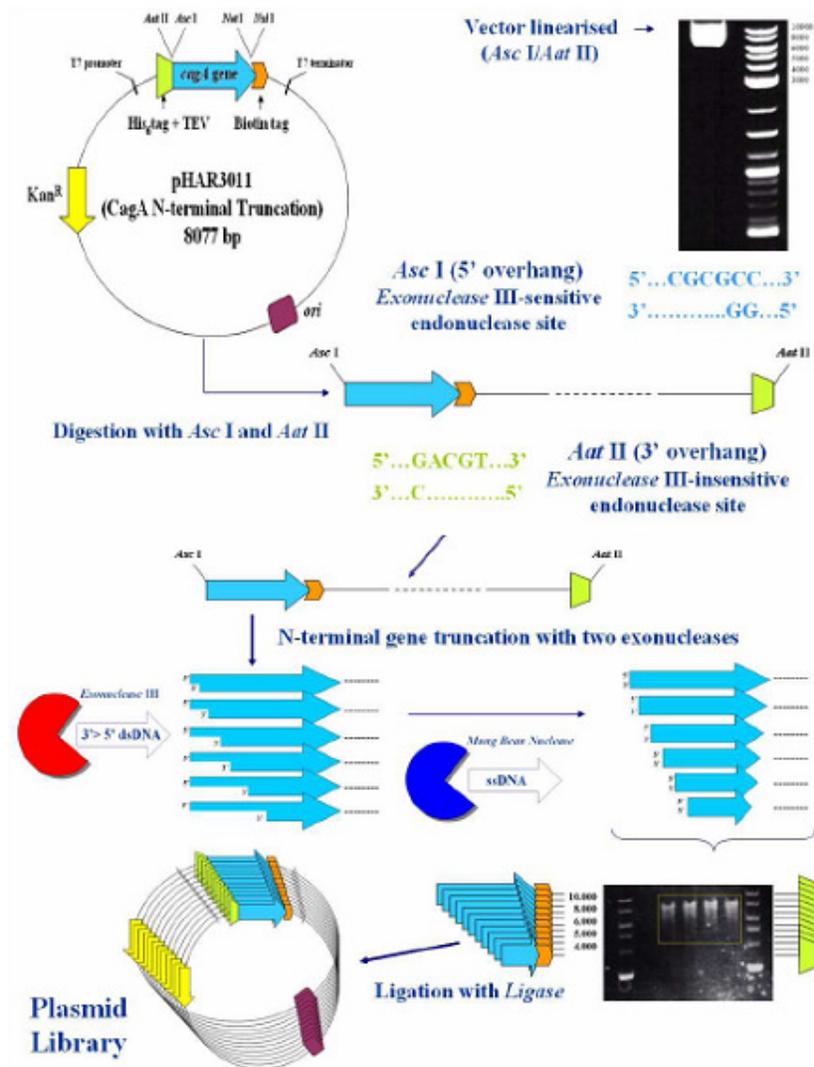


Figure 2.1: progressive deletion of pHAR3011 vector carrying *CagA* gene previously linearized with *Asc I* and *Aat II* restriction enzyme and digested with Exonuclease III/Mung Bean Nuclease.

CagA gene is cloned in frame with a N-terminal histidine tag and a C-terminal biotin tag. Digestion of this gene with *Asc I* and *Aat II* results in the generation of a 5' blunt end which is sensitive to the

Exonuclease III. The enzyme digests base by base the gene in the direction 3'–5', producing a 5' single strand overhang which is subsequently polished with the Mung Bean nuclease. The double digestion of the gene results in the generation of a library of linearized plasmids containing 5'-truncated constructs, which are subsequently ligated and purified on agarose gel. Picture taken from "Structural Characterization of Cag proteins from the Pathogenicity Island of *Helicobacter pylori*" (Alessandro Angelini, PhD Thesis defended at the University of Padua, 2008).

The efficient screening of such fragments requires readouts that allow for the discrimination of constructs that, despite being expressed, are not soluble. To achieve this, *CagA* was cloned in frame with a 5' histidine tag and a 3' biotin acceptor peptide (Figure 2.1). Subsequently to the expression of the truncated constructs, the histidine tag can be detected using specific antibodies only when this is accessible and exposed to the solvent, therefore functioning as a indicator for expression and purification, whilst the C-terminal biotin acceptor peptide functions as a readout for the solubility. The function of the biotin acceptor peptide is explained in detail in Figure 2.2.

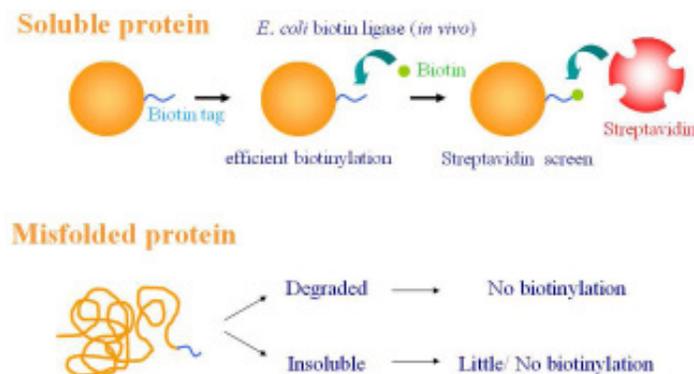


Figure 2.2: principles of utilization of a biotinylated acceptor peptide as a readout for solubility and folding. The function of the biotin acceptor peptide is to detect folded and soluble proteins: if the proteins fulfill these requirements, the biotin acceptor peptide will be accessible to the cytoplasmic enzyme Biotin Ligase, that will biotinilate the peptide. The biotinilated peptide can then be detected with the marked streptavidin. If the protein is misfolded, it will be insoluble or degraded in the cells, thus the biotin acceptor peptide will be inaccessible to the Biotin Ligase or degraded.

2.2 Results

A total 18432 clones have been tested for expression, solubility and purifiability, resulting in a panel of 40 constructs that have been furtherly investigated with large-scale experiments. Two C-terminal protein fragments of 25 and 33 kDa have been purified near to homogeneity. Five large fragments of molecular weight up to 100 kDa were proteolysed in the cells resulting in the loss of a ~30 kDa C-terminal fragment for each of them. A possible cleavage site is constituted by a stretch of five asparagines (N885-N889) located upstream the EPIYA-A motif. Subcloning two of these larger fragments without the C-terminal portion resulted in more stable products that could be purified to near homogeneity. All these results have been published in *FEBS JOURNAL* in February 2009 (Angelini, *et al.*, 2009).

2.3 Conclusion and discussion

In this study, the ESPRIT methodology was applied to bacterial protein for the first time. The screening of soluble and purifiable fragments of the *H. pylori* CagA protein resulted in the identification of: a) two purifiable C-terminal soluble constructs of respectively 25 and 33 kDa and b) five large fragments of respectively 81, 87, 95, 106 and 107 kDa, that degraded in bacterial cells.

We can conclude that CagA can be described as containing two major domains an N-terminal domain of ~100kDa (CagA^{N100}) and a C-terminal one of 33kDa (CagA^{C33}). This finding strongly support previous observations reporting that CagA was cleaved *in vivo* and a C-terminal fragment of around 30-35 kDa was detected in mammalian cells (Moese, *et al.*, 2001). These results provide a clear definition of this domain and suggest that a stretch of asparagines might be the cleavable linker. Still, the screening of the library did not identify other domain boundaries within the C-terminal or in the N-terminal portion of the protein.

A domain boundary is usually revealed by the ESPRIT method when a set of soluble and purifiable fragments comprises a well defined region of the protein. In our case, the identified CagA fragments span the region 225-1186 of the protein (Supplementary material, Figure S1). This suggests that no clear domain boundaries of CagA were identified from the ESPRIT a part from the CagA^{C33} domain.

A possible way to retrieve other CagA domain boundaries would be to truncate the CagA gene from both ends simultaneously (5' and 3'), and perform a so called "window

truncation”. This procedure decreases the number of soluble fragments output and conversely increases the number of fragments corresponding precisely to the regions where the domains are located (Figure 2.3).

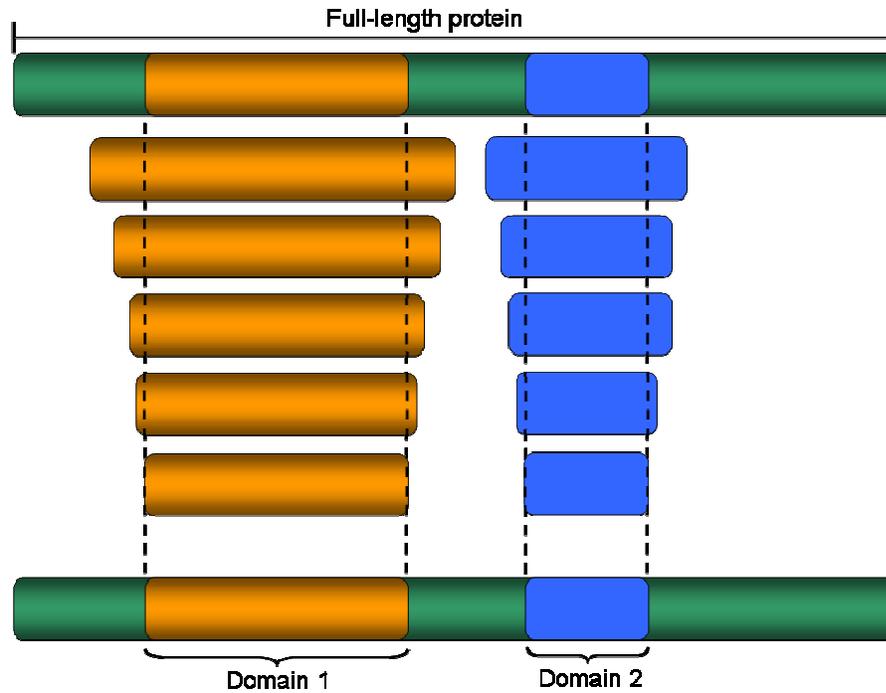


Figure 2.3: example of identification of domain boundaries from the ESPRIT methodology

A hypothetical protein, composed of two domains (domain 1 and domain 2) is submitted to the ESPRIT methodology. When random truncations are performed from both the N and C-terminal part of the protein, only two sets of soluble constructs are retrieved, spanning the regions comprised in domain 1 and domain 2.



Expression of *Helicobacter pylori* CagA domains by library-based construct screening

Alessandro Angelini^{1,2,*}, Tommaso Tosi^{1,*}, Philippe Mas³, Samira Acajjaoui¹, Giuseppe Zanotti², Laurent Terradot¹ and Darren J. Hart³

1 ESRF, Grenoble, France

2 Department of Chemistry, Institute of Biomolecular Chemistry, CNR, University of Padua, Italy

3 European Molecular Biology Laboratory, Grenoble Outstation, France

Keywords

CagA; *Helicobacter pylori*; protein expression; type IV secretion system; virulence factor

Correspondence

D. J. Hart, European Molecular Biology Laboratory, Grenoble Outstation, 6 rue Jules Horowitz, BP181, 38042 Grenoble Cedex 9, France

Fax: +33 476 20 71 99

Tel: +33 476 20 77 68

E-mail: hart@embl.fr

L. Terradot, ESRF, 6 Rue Jules Horowitz, BP220, 38043 Grenoble Cedex 9, France

Fax: +33 476 20 94 00

Tel: +33 476 20 94 54

E-mail: terradot@esrf.fr

*These authors contributed equally to this work

(Received 22 October 2008, revised 1 December 2008, accepted 3 December 2008)

doi:10.1111/j.1742-4658.2008.06826.x

Helicobacter pylori is a human pathogen that infects more than half of the human population [1,2]. It is considered to be the main cause of most gastric pathologies, including chronic gastritis, peptic ulcer, gastric adenocarcinoma, and lymphoma of the mucosa-associated lymphoid tissue [3]. Infection with *H. pylori* strains carrying the *cytotoxin-associated gene* pathogenicity island (*cag*-PAI) leads to a higher risk of gastric cancer [4,5]. This gene cluster of around 40 kb

Highly pathogenic strains of *Helicobacter pylori* use a type IV secretion system to inject the CagA protein into human gastric cells. There, CagA associates with the inner side of the membrane and is tyrosine-phosphorylated at EPIYA motifs by host kinases. The phosphorylation triggers a series of interactions between CagA and human proteins that result in a dramatic change of cellular morphology. Structural and functional analyses of the protein have proved difficult, due to the proteolytically sensitive nature of the recombinant protein. To circumvent these difficulties, we applied ESPRIT, a library-based construct screening method, to generate a comprehensive set of 5'-randomly deleted gene fragments. Screening of 18 432 constructs for soluble expression resulted in a panel of 40 clones, which were further investigated by large-scale purification. Two constructs of approximately 25 and 33 kDa were particularly soluble and were purified to near homogeneity. CagA fragments larger than 40 kDa were prone to heavy proteolysis at the C-terminus, with a favoured cleavage site near the first EPIYA motif. Thus, these well-expressed recombinant constructs isolated are likely to be similar to those observed following natural proteolysis in human cells, and open the way for structural and functional studies requiring large amounts of purified material.

codes for 29 proteins involved in the assembly of a type IV secretion system (T4SS) [6,7]. T4SSs are used by many bacteria for genetic conjugation, high-frequency recombination, and delivery of effector macromolecules [8]. The *cag*-encoded T4SS forms a long appendage that is necessary to deliver the effector protein, CagA, into gastric epithelial cells [9,10].

Once injected inside target cells, CagA associates with the inner side of the cytoplasmic membrane and

Abbreviations

cag-PAI, *cytotoxin-associated gene* pathogenicity island; IPTG, isopropyl thio- β -D-galactoside; T4SS, type IV secretion system.

is tyrosine-phosphorylated by host kinases, including c-Src, Lyn, Fyn, Yes and c-Abl [11]. Phosphorylated CagA then binds several SH2 domain-containing host proteins involved in signalling pathways, including the SHP-2 phosphatase, Csk tyrosine kinase, and the Crk adapter protein [12–14]. These events ultimately result in cellular morphological changes known as the ‘hummingbird phenotype’, and enhance cellular motility, resulting in cell scattering [11,15–17]. Phosphorylated CagA has also been found to interact with PAR1/MARK kinase, promoting both cell polarity defects and the hummingbird phenotype through the inhibition of PAR1 phosphorylation [18]. Several CagA-mediated effects are independent of its phosphorylation status. CagA is known to interact with the scaffold protein ZO-1 and also the tight junctional adhesion protein JAM, leading to a redistribution of tight junction proteins and the alteration of apical–junctional complex function [19,20]. An increase in cell motility is furthermore induced by the intracellular interaction of CagA with the scatter factor receptor c-Met, which deregulates the c-Met receptor pathway and thereby induces cell invasion [21,22]. The C-terminal portion of CagA has also been found to interact with Grb2, triggering a Ras-dependent signalling pathway that results in increased cell scattering and proliferation [23].

CagA effects exerted *in vivo* are mediated by protein–protein interactions with host cell components. However, the biochemistry of CagA and of its interactions remains largely elusive, because the full-length protein is unstable in recombinant expression systems and rapidly degraded [24]. The purification of a CagA fragment (from residues 392–733) has been previously reported, and this was shown to bind *HpYbgC* *in vitro* [25,26], but the biological relevance of this interaction has yet to be elucidated. The N-terminal 57 residues interact with ZO-1 and JAM [19] and contain a sequence important for translocation by the T4SS [27]. The C-terminus of CagA contains the tyrosine phosphorylation sites, which are major determinants of its mode of action. One phosphorylation site comprises a stretch of 34 residues including the five amino acid EPIYA motif. EPIYA motifs can be found multiple times in the C-terminal region of the protein, depending on the bacterial strain [11]. Whereas the first two motifs (EPIYA-A and EPIYA-B) are conserved among all strains, Western isotypes possess 0–4 copies of the EPIYA-C motif, thereby explaining the variation in molecular mass observed in different CagA proteins. Furthermore, the Eastern CagA isotypes do not carry an EPIYA-C motif, but a different sequence, EPIYA–D [28]. The EPIYA-C and EPIYA-D motifs are the

major, if not the only, phosphorylation sites in CagA [29], and sequences within the EPIYA motifs also mediate the attachment to the membrane [30]. In addition, the multimerization of CagA mediated by EPIYA motifs seems to be a prerequisite for the interaction with the SHP-2 phosphatase. Finally, the CagA C-terminus mediates translocation inside human gastric cells. A lysine-rich motif located within the last 20 residues is necessary, together with the CagA N-terminus, for the translocation of the protein [27]. Furthermore, CagF, another component of the *cag*-PAI, is known to interact with a stretch of 100 amino acids located at the CagA C-terminus, just upstream of the C-terminal motif necessary for translocation [24,31]. Finally, the pilus-associated protein CagL is required to facilitate the receptor interaction prior to injection of CagA [32].

Although these data clearly define the C-terminal region as being of great functional importance, information about the boundaries of domains involved in protein–protein interactions and their *in vitro* biochemical and structural characterization is limited. In order to investigate the biochemistry and structure of CagA, we sought to identify soluble fragments in the C-terminal region of CagA for use in subsequent studies. As CagA has no significant sequence homology with other proteins, it was impossible to identify putative domains through analysis of multiple sequence alignments. We therefore used a recently developed high-throughput screen for soluble constructs, ESPRIT (expression of soluble proteins by random incremental truncation) [33–35], in which a comprehensive 18 432 clone random library of 5'-deletion constructs was synthesized and screened for soluble expression. A panel of C-terminal fragments was isolated that were well expressed in *Escherichia coli* and purifiable by affinity chromatography. Endogenous proteolysis of several of these fragments led to further refinement of domain boundaries. The resulting constructs will be of significant use in biochemical and structural studies on the role of this important virulence factor.

Results

Library construction

An exonuclease III and mung bean nuclease truncation protocol was performed on a precloned *cagA* gene to generate a comprehensive random library of constructs in which most (if not all) possible fusion points of a cleavable N-terminal hexahistidine tag were present. One-third of the constructs were in-frame with this sequence, resulting in T7 promoter-driven expression

constructs. All constructs were fused in-frame with a C-terminal biotin acceptor peptide used as a marker for protein solubility and stability during colony screening. The library was divided into three insert size ranges (500–1500, 1500–2500, and 2500–3500 bp) by size selection on agarose gel to reduce the dominance of small expression constructs. The resulting sublibraries were handled separately thereafter. Colony PCR and DNA sequencing of unselected clones indicated an approximately even distribution of construct sizes, with no obvious bias. The frequencies of amplified DNA inserts were 40% (13/32), 56% (18/32) and 47% (15/32), representing minimum insert efficiencies for the three sublibraries. The pooled sublibraries were used to transform BL21-CodonPlus(DE3)-RIL, and 18 432 colonies (6144 for each sublibrary) were picked into 384-well plates. The minimum insert efficiency combined with the length of the gene (3561 bp) predicts an approximately threefold oversampling of constructs.

Assessment of clones for soluble purifiable protein expression

A three-step solubility screening process was employed (Fig. 1A): (a) robotically arraying clones on nitrocellulose membranes; (b) growing colonies; and (c) inducing protein expression by shifting membranes to agar plates containing isopropyl thio- β -D-galactoside (IPTG). Putative soluble expression clones were isolated from each sublibrary by analysing membranes hybridized simultaneously with Alexa488 streptavidin and a monoclonal anti-hexahistidine tag with associated Alexa532 mouse secondary antibody (Fig. 1B). Values of signal intensities for both N-terminal and C-terminal tags were extracted from arrays into MICROSOFT EXCEL, and an initial filtering was applied to eliminate all clones with no detectable hexahistidine tag signals.

In total, 9635 colonies showed some signal for the N-terminal hexahistidine tag and were subsequently ranked for streptavidin fluorescence, resulting in 453 clones with significant signals for both tags. In this way, clones exhibiting only one terminus due to internal translational initiation, premature translational termination or proteolysis were removed, generating a higher-quality subset for subsequent liquid expression testing and purification trials. Indeed, we observed a high number of clones in medium and large insert sublibraries exhibiting strong C-terminal biotinylation signals, but a complete absence of detectable hexahistidine tag. This effect was not found in the small sublibrary, and is suggestive of an internal translational

start site common to all fragments larger than 1500 bp; these constructs were eliminated, due to the absence of hexahistidine tag.

Ninety-six clones comprising the 32 highest expressers from each sublibrary were assessed for insert size by PCR and simultaneously for expression of Ni²⁺-nitrilotriacetic acid purifiable protein from 4 mL cultures. Proteins eluted in the imidazole-containing buffer were visualized by SDS/PAGE and western blot performed with both fluorescent streptavidin and antibody against hexahistidine tag (Fig. 1C).

Analysis of expression clones

Of the 96 constructs tested, three short constructs [20.3 (not shown), 29.2 and 29.7 kDa] and six medium constructs (36.4, 37.4, 41.6, 42.1, 42.2 and 44.5 kDa) were well expressed, and purified as a single band (Fig. 1C). Sizes of the proteins were predicted from DNA sequencing of inserts and confirmed by SDS/PAGE (including 5 kDa of tags). Nineteen other constructs were purified, but displayed weak bands by fluorescent western assay, suggesting that levels of soluble expression were very low. Nineteen further constructs between 45 and 60 kDa exhibited significant degradation. Owing to the weak nature of many of the bands, western blot analysis was also performed using streptavidin and antibody against hexahistidine to confirm the SDS/PAGE results. The remaining 49 constructs showed no significant protein expression in these small-scale experiments, and so they were not studied further, due to lack of expression (false positives) or because of the requirement for inconvenient scale-up. Further sequencing identified the expression-compatible N-termini and revealed a total of 40 unique sequences that were in-frame with the hexahistidine tag. Significant heterogeneity at the tag fusion position was observed (Fig. S1).

Scale-up expression and purification

One litre scale protein expression trials were performed on all unique clones exhibiting detectable protein expression in small-scale experiments, and Ni²⁺-nitrilotriacetic acid purification was performed using 1 mL His-trap columns. Two proteins of 25 and 33 kDa were purified as a single band in SDS/PAGE (Fig. 2A; clones 8 and 37) at high yield and quality. Five larger constructs (Fig. 2A; clones 94, 90, 70, 83 and 88) revealed proteins of 81, 87, 95, 106 and 107 kDa. They were also purified, but a second band of lower molecular mass was detectable in the SDS/PAGE gel. Western blot analysis of these five

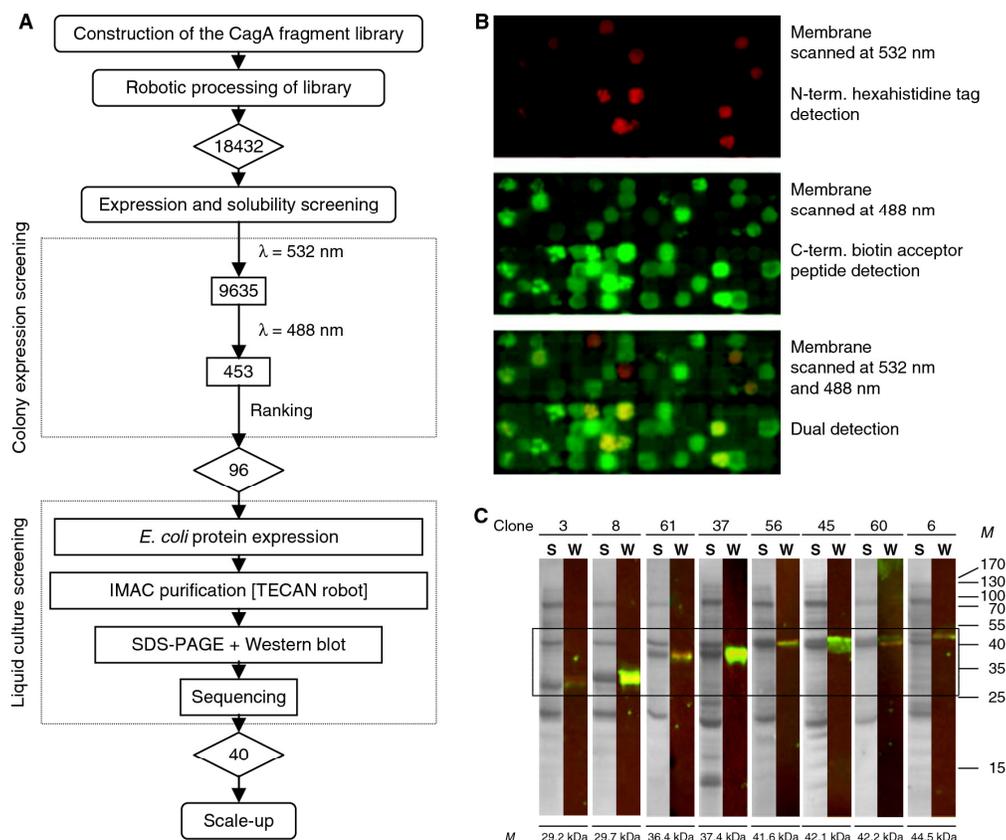


Fig. 1. Screening for protein expression and solubility from a CagA random truncation library. (A) Flow chart for improving CagA protein expression and solubility. A library of 18 342 clones of 5'-random *cagA* truncations was robotically picked and gridded onto nitrocellulose agar to grow colony arrays in which protein expression was induced and assessed by intensity of N-terminal hexahistidine tag and C-terminal biotin acceptor peptide signals. The 96 best ranked clones were purified from small-scale expression cultures by Ni^{2+} -nitrilotriacetic acid purification. From SDS/PAGE analysis, 40 constructs were selected for larger-scale purification. (B) Colony-based solubility screening of the construct library. Example of detection of N-terminal hexahistidine (red, top panel), C-terminal biotinylation (green, middle panel), and merged image (lower panel). (C) Testing of putative soluble CagA fragments by purification on Ni^{2+} -nitrilotriacetic acid agarose from 4 mL expression cultures. Results are presented for eight of the 96 clones tested showing successful outcomes. For each clone, the Coomassie blue-stained SDS polyacrylamide gel (S) is displayed alongside fluorescent western blot analysis (W) of the same sample. An overlay of the hybridizations with antibody and streptavidin [as in (B)] is presented, showing the presence of both genetically encoded protein termini, indicating proteolytically stable fragments. Protein molecular mass markers are indicated (M).

fragments with antibodies against hexahistidine tag indicated that the subfragments correspond to truncated proteins lacking a C-terminal portion of about 33 kDa (Fig. 2A). This repeated pattern suggested that these constructs were all expressed and soluble but contained a C-terminal domain that was particularly prone to proteolysis in *E. coli* cells. From the size of the fragments and the CagA sequence, we hypothesized that the C-terminal domain cleaved in these constructs started at a stretch of five asparagines, corresponding to residues 885–889 (Fig. 2B,C). We thus designed two new constructs, 90 Δ C and 94 Δ C,

with the same N-terminal residues as clones 90 and 94, respectively, but ending at these five asparagines (Fig. 2B). The expression and purification of these new constructs resulted in stable purifiable domains exhibiting little degradation and located internally within the CagA sequence (Fig. 2B).

Discussion

The pathogen *H. pylori* uses a T4SS to inject the CagA pathogenicity factor into human gastric cells, where it perturbs host signalling pathways to provide

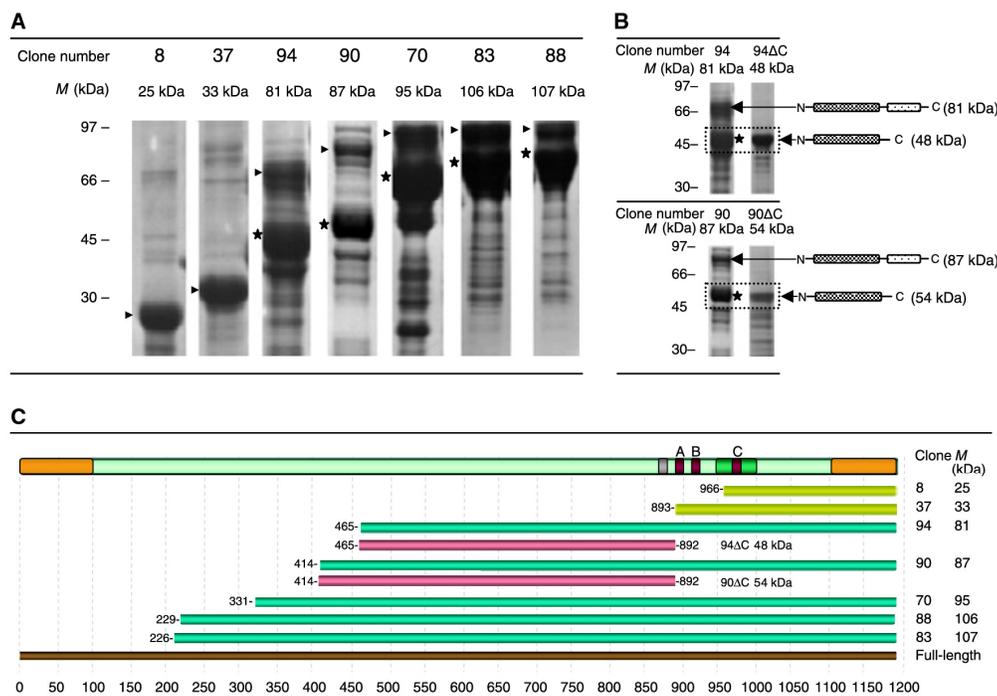


Fig. 2. Expression of soluble C-terminal CagA fragments. (A) Seven CagA protein fragments (clones: 8, 37, 94, 90, 70, 88 and 83) of different sizes (25, 33, 87, 81, 95, 106 and 107 kDa, respectively) were expressed at 1 L scale with an N-terminal hexahistidine tag and purified by Ni^{2+} -nitrilotriacetic acid affinity chromatography. All purified parental fragments are indicated with an arrow. From left: clones 8 and 37 (25 and 33 kDa, respectively) show relatively high proteolytic stability, whereas five other constructs (clones 94, 90, 70, 83 and 88) clearly show the presence of a cleaved 'subfragment' (indicated by an asterisk). (B) Clones 94ΔC and 90ΔC, subcloned from clones 94 and 90 with deletion of a C-terminal fragment (around 33 kDa), express proteolytically stable, purifiable protein fragments as assessed by SDS/PAGE. (C) Summary diagram of CagA constructs yielding purifiable fragments. Above: a schematic view of the CagA protein sequence. N-terminal and C-terminal translocation signals are orange, the stretch of five asparagines (885–889) is grey, EPIYA motifs are dark red, and multimerization motifs are dark green. Below: the positions of the purifiable fragments are indicated by solid bars. Yellow bars indicate C-terminal constructs exhibiting little proteolytic degradation. Blue bars are constructs yielding mixtures of both unproteolysed and proteolysed fragments, the latter corresponding to loss of the C-terminal region. Pink bars represent clones that were subsequently constructed by deleting the region corresponding to the longer stable C-terminal construct (clone 37) shown in yellow.

a local environment that is more suitable for the survival of the pathogen. Biochemical studies have associated specific functions with some protein regions [11]. However, the large size of the CagA protein and its weak homology with other eukaryotic and prokaryotic proteins of known structure and function make it difficult to predict regions that constitute well-folded domains through the use of multiple sequence alignments. This has prevented both structural and most *in vitro* functional studies on this important virulence factor.

To circumvent these problems, we have applied ESPRIT, a random construct library screening approach, to identify C-terminal regions of CagA that can be overexpressed in a soluble form in *E. coli* and

purified with a yield high enough for structural studies and high protein consumption *in vitro* analytical methods. All possible 5'-unidirectional truncations of the target gene were generated and tested for soluble expression using high-throughput screening robotics [33]. The usefulness of this method for empirical dissection of large, crystallographically intractable proteins into manageable domains has been demonstrated recently for influenza proteins [33–35], and it is used here on a bacterial pathogen for the first time.

In this study, CagA reveals itself to be a challenging protein for recombinant expression, due to a propensity for degradation. This may be ascribed to the high proportion of basic residues (15% value overall), together with predicted unstructured regions.

Systematic truncation analysis, as performed here, permits all structural N-termini to be tested as well as the effects of nonstructural factors, such as translation-inhibiting mRNA secondary structure and compatibility with *E. coli* cellular proteolysis mechanisms.

Two types of CagA protein constructs were identified during the screening process: first, two C-terminal fragments, 25 and 33 kDa, that are well expressed and easily purifiable (Fig. 2C); and second, longer constructs of 40–100 kDa that were partially proteolysed by the cell. Analysis of the degradation products revealed a region upstream of the C-terminal fragment that can be expressed as a stable entity once subcloned. Thus, we have been able to identify two adjacent domains, one directly and the second indirectly.

These results are in partial agreement with *in vivo* and *in vitro* experiments [24,31,36–38], as our larger purified protein samples between 40 and 100 kDa showed a similar degradation pattern to that previously observed. These earlier experiments had already suggested that full-length CagA was a fragile protein that was proteolytically sensitive, breaking into subunits at defined positions. In particular, CagA cleaved to yield a C-terminal fragment of about 35–40 kDa in human cells [39], as we have observed in *E. coli*. However, no further optimization of these constructs into high-level expression constructs useful for subsequent studies has been reported.

The 33 kDa stable fragment from clone 37 corresponds approximately to the C-terminus of CagA, starting at the first EPIYA motif. The region preceding this fragment (around residues 877–918) is predicted to be highly disordered according to order/disorder predictors; moreover, a long loop is predicted in position 882–916 (Fig. 2). It may also be of significance that a stretch of five asparagines lies just before it, perhaps constituting a natural linker that is cleaved inside human cells. Interestingly, the C-terminal part of CagA contains the domain responsible for interacting with CagF, a putative chaperone of CagA [24,31]. The binding of CagF may be important in stabilizing the entire protein and in preventing proteolysis before its injection.

The soluble, purifiable constructs of the CagA protein identified here through application of ESPRIT have not been described previously. We believe that large-scale expression and purification of these domains will enable progress towards a definition of the CagA structure and will aid functional interaction studies of CagA with CagF and host cell components.

Experimental procedures

Amplification and subcloning of the CagA gene

The gene *hp0547* coding for CagA was amplified by PCR from genomic DNA (*H. pylori* strain 26695) with primers *hp0547for1* (5'-CACCA TGA CT AACGA AACTA TTG AT C-3') and *hp0547rev1* (5'-TTAAG ATTTT TGGAA ACCAC CTTTT G-3') and cloned into pET151/D topo (Invitrogen, Carlsbad, CA, USA). Internal *NsiI* restriction sites were silenced (ATGCAT to ATGCGT) by Quikchange mutagenesis (Stratagene, La Jolla, CA, USA), and the mutated gene used as a template to again amplify *hp0547* by PCR, adding a 5'-*AscI* and a 3'-*NsiI* site, by using the flanking primers *hp0547for2* (5'-GATCC TAGGG CGCGC CACTA ACGAA ACTAT TGATC AAACA AGAAC AC CAG-3') and *hp0547rev2* (5'-CTAGG ATCAT GCATT AG ATT TTTGG AAACC ACCTT TTGTA TTAAC ATT-3'). The PCR product was then digested with *AscI* and *NsiI*, and inserted into a pET9a-derived vector, pTAR010 [34], designed for the 5'-truncation of the gene, in-frame with a TEV-cleavable N-terminal hexahistidine tag (MGHHHH HHDYDIPTTENLYFQG) and a C-terminal biotin acceptor peptide (SNNGGGGLNDIFEAQKIEWHE) to generate pHAR3011-CagA.

Construction of the 5'-CagA deletion library

Plasmid was purified from a 200 mL (LB medium, 50 $\mu\text{g}\cdot\text{mL}^{-1}$ kanamycin) saturated culture of *E. coli* 10G Duos (Lucigen, Middleton, WI, USA) (pHAR3011-CagA). After overnight growth at 37 °C, the cells were harvested, and lysed by alkaline lysis treatment, and the DNA was extracted using phenol/chloroform/isoamyl alcohol (25 : 24 : 1) followed by isopropanol precipitation. The plasmid DNA was then further purified using a midiprep kit (Qiagen, Valencia, CA, USA) to remove contaminants.

Plasmid (10 μg) was digested with *AscI* and *AatII*, yielding an *AscI* 5'-end (sensitive to exonuclease III) and an *AatII* 3'-end (insensitive). For the exonuclease III truncation reaction, 4 μg of digested plasmid was diluted in 130 μL of reaction buffer [1 \times buffer 1 from New England Biolabs (Ipswich, MA, USA) supplemented with 30 mM NaCl] and 400 units of exonuclease III (New England Biolabs), and incubated at 22 °C. To ensure even fragment distribution, 2 μL aliquots were taken every 2 min over 2 h and immediately added to an ice-cold tube containing 200 μL of 3 M NaCl. The quenched reaction was denatured at 70 °C for 20 min, and DNA was purified using a Nucleospin Extract II kit (Macherey-Nagel, Düren, Germany). The remaining 5'-overhang was removed by incubation with 5 units of mung bean nuclease (New England Biolabs) at 30 °C for 30 min, and then repurified with the Nucleospin Extract II kit. The library DNA (40 μL from column elution) was then incubated with *Pfu* native polymerase

(Stratagene) in a total volume of 50 μL ($1\times$ *Pfu* polymerase native buffer, 2.5 mM dNTPs, and 1 unit of enzyme) at 72 °C for 20 min to generate blunt end fragments. The reaction was loaded onto a 0.5% agarose gel, and plasmids with inserts in the ranges 500–1500, 1500–2500 and 2500–3500 bp were excised from the gel and purified using the QiaexII gel extraction kit (Qiagen), generating three sublibraries. The linear DNA fragments of the three sublibraries were then recircularized using T4 DNA ligase (Roche, Mannheim, Germany). One Shot Omnimax 2 T1 cells (Invitrogen) were transformed with 2 μL of ligation reaction, and then recovered at 37 °C for 1 h and plated on 22-cm-square LB agar plates with 50 $\mu\text{g}\cdot\text{mL}^{-1}$ kanamycin. Approximately 19 000 colonies were scraped from the plate, and resuspended in NaCl/P_i; plasmid DNA was then purified with a midiprep kit (Qiagen). The supercoiled library DNA was then used to transform *E. coli* BL21-Codon-Plus(DE3)-RIL (Stratagene) for protein expression testing.

Robotic processing of the library

The bacteria were plated on 22 cm LB agar plates supplemented with kanamycin and chloramphenicol at 50 $\mu\text{g}\cdot\text{mL}^{-1}$, and incubated overnight at 37 °C. Approximately 20 000 colonies were picked into 384-well plates containing 70 μL of TB medium (kanamycin and chloramphenicol at 50 $\mu\text{g}\cdot\text{mL}^{-1}$) per well using a colony-picking robot (KBiosystems, Basildon, UK). Plates were shaken at 30 °C overnight, and then arrayed onto 22 cm nitrocellulose membranes on LB agar plates supplemented with antibiotics as above. Plates were incubated at 25 °C until small colonies appeared, and then the membrane was moved to similar LB plates additionally supplemented with 1 mM IPTG and 50 μM biotin to induce protein expression at 30 °C for a further 4 h.

Identification of putative soluble protein-expressing clones

Colonies were lysed *in situ* on the nitrocellulose membrane by placing the membrane on Whatman filter paper soaked in lysis solution (500 mM NaOH, 1.5 M NaCl) for 10 min, followed by a neutralization solution (1.5 M Tris, pH 7.5, 1.5 M NaCl; 2 \times 10 min). Membranes were washed in 2 \times SSC buffer for 15 min, the cellular debris was removed by gentle scraping of the membrane under liquid, and then the membranes were incubated overnight at 4 °C in Superblock (Pierce, Woburn, MA, USA). Membranes were washed with PBS-T buffer (NaCl/P_i with 0.2% Tween-20), and then incubated for 1 h at 4 °C with monoclonal antibody against hexahistidine (Roche; 1 : 3000 in PBS-T buffer). After being washed in PBS-T buffer, the membranes were further incubated for 1 h at 4 °C with a secondary mouse antibody–Alexa532 conjugate (Invitrogen) and streptavidin–Alexa488 (Invitrogen) at dilutions of 1 : 1000 and 1 : 5000, respectively. The membranes were washed with

PBS-T buffer, and then with distilled water, and visualized with a Typhoon 9400 fluorescence imager (GE Healthcare, Piscataway, NJ, USA), using lasers at 488 and 532 nm to quantify hexahistidine tag and biotin acceptor peptide signal intensities, respectively.

High-throughput expression and purification of CagA fragments

Clones with clearly visible hexahistidine tag signals were ranked according to their biotinylation signal intensity, and the most intense 96 clones were analysed for insert size by PCR direct from culture using standard T7 forward and T7 reverse primers. Small-scale protein expression testing of positive clones was performed at 4-mL scale in 24-well plates with TB medium (kanamycin and chloramphenicol, 50 $\mu\text{g}\cdot\text{mL}^{-1}$). Cells were grown at 37 °C with shaking at 250 r.p.m. to an attenuation at 600 nm (D_{600}) of 0.6, and then induced with 1 mM IPTG and grown overnight at 25 °C with shaking at 250 r.p.m. Cells were pelleted by centrifugation (2250 *g* for 20 min), and then resuspended in 4 mL of sphaeroplast buffer (20 mM Tris, pH 8, 250 mM NaCl, 20% sucrose) with 1 $\text{mg}\cdot\text{mL}^{-1}$ lysozyme). The resulting sphaeroplasts were pelleted by centrifugation, and the supernatant was discarded and resuspended in 700 μL of sphaeroplast lysis buffer (10 mM Tris, pH 7.5, 0.5% Brij-58, DNaseI, and Roche protease inhibitors mix). After 20 min at 4 °C, the lysate was centrifuged (2250 *g* for 20 min), and the hexahistidine-tagged proteins were purified from the supernatants with a liquid handling robot (Tecan) in 96-well format, using filter plates charged manually with 60 μL of Ni²⁺-nitrilotriacetic acid resin (Sigma, St Louis, MO, USA). Wash buffer (50 mM phosphate buffer, pH 7, 300 mM NaCl, 5 mM imidazole) and elution buffer (the same, with 300 mM imidazole) were used. Samples from the elution step were analysed by SDS/PAGE, and the sizes of fragments were compared with the sizes of DNA bands obtained by PCR screen. The DNA sequences of the selected clones were determined.

Scale-up expression trials and purification of selected clones

Plasmid DNA from the selected clones was used to transform *E. coli* BL21(DE3) pLysS (Invitrogen). Expression trials were performed in 1 L of TB medium (kanamycin and chloramphenicol, 50 $\mu\text{g}\cdot\text{mL}^{-1}$). Cultures were grown to D_{600} 0.6 at 37 °C; the temperature was reduced to 20 °C, and cultures were induced with 1 mM IPTG for 4 h. Cells were harvested by centrifugation (7500 *g* for 20 min), and resuspended in 20 mL of lysis buffer (50 mM Tris, pH 8, 150 mM NaCl, 5% glycerol, 0.2% Triton X-100). Lysis was performed by sonication or by cell disrupter, and the clarified supernatant was loaded on a 1 mL HiTrap affinity column (GE Healthcare). All purifications were performed

on an AKTA Prime system (Amersham Biosciences, Piscataway, NJ, USA), using a gradient of imidazole for the elution (buffers used were: 50 mM Tris, pH 8, 150 mM NaCl, 5% glycerol; and 50 mM Tris, pH 8, 150 mM NaCl, 5% glycerol, 500 mM imidazole). Samples were analysed by SDS/PAGE and western blot, performed using monoclonal antibodies against hexahistidine tag.

The internally located clones 94ΔC and 90ΔC were cloned with primers starting at residues Glu465 (5'-CACCG AGTTT AATAA TGGGG ATTTG AGC-3') and Lys414 (5'-CACCA AATTA GACAA CTTGA GCGAG AAAG-3') and reverse primer 5'-GTTTT TGAGT CCATT ATTATT CTAA TTG-3', with Lys892 being substituted by a stop codon in the latter. The purified PCR products were cloned into pET151/D-topo vector (Invitrogen) according to the instructions supplied. Protein expression and purification was carried out as above, using BL21(DE3) strains Codon-Plus-RIL (Stratagene) or pLysS (Invitrogen).

Acknowledgements

We acknowledge Partnership for Structural Biology for an integrated structural biology environment. A. Angelini was supported by grants from Ing. Aldo Gini Foundation and the Italian Ministries of University and Research (MIUR). This work was partly funded by the ESRF 'In House' Research program.

References

- Covacci A, Telford JL, Del Giudice G, Parsonnet J & Rappuoli R (1999) *Helicobacter pylori* virulence and genetic geography. *Science* **284**, 1328–1333.
- Montecucco C & Rappuoli R (2001) Living dangerously: how *Helicobacter pylori* survives in the human stomach. *Nat Rev Mol Cell Biol* **2**, 457–466.
- Franco AT, Johnston E, Krishna U, Yamaoka Y, Israel DA, Nagy TA, Wroblewski LE, Piazuolo MB, Correa P & Peek RM Jr (2008) Regulation of gastric carcinogenesis by *Helicobacter pylori* virulence factors. *Cancer Res* **68**, 379–387.
- Uemura N, Okamoto S, Yamamoto S, Matsumura N, Yamaguchi S, Yamakido M, Taniyama K, Sasaki N & Schlemper RJ (2001) *Helicobacter pylori* infection and the development of gastric cancer. *N Engl J Med* **345**, 784–789.
- Peek RM Jr & Blaser MJ (2002) *Helicobacter pylori* and gastrointestinal tract adenocarcinomas. *Nat Rev Cancer* **2**, 28–37.
- Censini S, Lange C, Xiang Z, Crabtree JE, Ghiara P, Borodovsky M, Rappuoli R & Covacci A (1996) *cag*, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. *Proc Natl Acad Sci USA* **93**, 14648–14653.
- Akopyants NS, Clifton SW, Kersulyte D, Crabtree JE, Youree BE, Reece CA, Bukanov NO, Drazek ES, Roe BA & Berg DE (1998) Analyses of the *cag* pathogenicity island of *Helicobacter pylori*. *Mol Microbiol* **28**, 37–53.
- Christie PJ, Atmakuri K, Krishnamoorthy V, Jakubowski S & Cascales E (2005) Biogenesis, architecture, and function of bacterial type IV secretion systems. *Annu Rev Microbiol* **59**, 451–485.
- Backert S, Ziska E, Brinkmann V, Zimny-Arndt U, Fauconnier A, Jungblut PR, Naumann M & Meyer TF (2000) Translocation of the *Helicobacter pylori* CagA protein in gastric epithelial cells by a type IV secretion apparatus. *Cell Microbiol* **2**, 155–164.
- Odenbreit S, Puls J, Sedlmaier B, Gerland E, Fischer W & Haas R (2000) Translocation of *Helicobacter pylori* CagA into gastric epithelial cells by type IV secretion. *Science* **287**, 1497–1500.
- Hatakeyama M (2008) SagA of CagA in *Helicobacter pylori* pathogenesis. *Curr Opin Microbiol* **11**, 30–37.
- Tsutsumi R, Higashi H, Higuchi M, Okada M & Hatakeyama M (2003) Attenuation of *Helicobacter pylori* CagA × SHP-2 signaling by interaction between CagA and C-terminal Src kinase. *J Biol Chem* **278**, 3664–3670.
- Suzuki M, Mimuro H, Suzuki T, Park M, Yamamoto T & Sasakawa C (2005) Interaction of CagA with Crk plays an important role in *Helicobacter pylori*-induced loss of gastric epithelial cell adhesion. *J Exp Med* **202**, 1235–1247.
- Brandt S, Shafikhani S, Balachandran P, Jin S, Hartig R, König W, Engel J & Backert S (2007) Use of a novel coinfection system reveals a role for Rac1, H-Ras, and CrkII phosphorylation in *Helicobacter pylori*-induced host cell actin cytoskeletal rearrangements. *FEMS Immunol Med Microbiol* **50**, 190–205.
- Tammer I, Brandt S, Hartig R, König W & Backert S (2007) Activation of Abl by *Helicobacter pylori*: a novel kinase for CagA and crucial mediator of host cell scattering. *Gastroenterology* **132**, 1309–1319.
- Stein M, Bagnoli F, Halenbeck R, Rappuoli R, Fantl WJ & Covacci A (2002) c-Src/Lyn kinases activate *Helicobacter pylori* CagA through tyrosine phosphorylation of the EPIYA motifs. *Mol Microbiol* **43**, 971–980.
- Selbach M, Moese S, Hauck CR, Meyer TF & Backert S (2002) Src is the kinase of the *Helicobacter pylori* CagA protein in vitro and in vivo. *J Biol Chem* **277**, 6775–6778.
- Saadat I, Higashi H, Obuse C, Umeda M, Murata-Kamiya N, Saito Y, Lu H, Ohnishi N, Azuma T, Suzuki A *et al.* (2007) *Helicobacter pylori* CagA targets PAR1/MARK kinase to disrupt epithelial cell polarity. *Nature* **447**, 330–333.
- Amieva MR, Vogelmann R, Covacci A, Tompkins LS, Nelson WJ & Falkow S (2003) Disruption of the epithelial apical-junctional complex by *Helicobacter pylori* CagA. *Science* **300**, 1430–1434.

- 20 Bagnoli F, Buti L, Tompkins L, Covacci A & Amieva MR (2005) *Helicobacter pylori* CagA induces a transition from polarized to invasive phenotypes in MDCK cells. *Proc Natl Acad Sci USA* **102**, 16339–16344.
- 21 Churin Y, Al-Ghoul L, Kepp O, Meyer TF, Birchmeier W & Naumann M (2003) *Helicobacter pylori* CagA protein targets the c-Met receptor and enhances the motogenic response. *J Cell Biol* **161**, 249–255.
- 22 Oliveira MJ, Costa AC, Costa AM, Henriques L, Suriano G, Atherton JC, Machado JC, Carneiro F, Seruca R, Mareel M *et al.* (2006) *Helicobacter pylori* induces gastric epithelial cell invasion in a c-Met and type IV secretion system-dependent manner. *J Biol Chem* **281**, 34888–34896.
- 23 Mimuro H, Suzuki T, Tanaka J, Asahi M, Haas R & Sasakawa C (2002) Grb2 is a key mediator of *Helicobacter pylori* CagA protein activities. *Mol Cell* **10**, 745–755.
- 24 Pattis I, Weiss E, Laugks R, Haas R & Fischer W (2007) The *Helicobacter pylori* CagF protein is a type IV secretion chaperone-like molecule that binds close to the C-terminal secretion signal of the CagA effector protein. *Microbiology* **153**, 2896–2909.
- 25 Terradot L, Durnell N, Li M, Ory J, Labigne A, Legrain P, Collard F & Waksman G (2004) Biochemical characterization of protein complexes from the *Helicobacter pylori* protein interaction map: strategies for complex formation and evidence for novel interactions within type IV secretion systems. *Mol Cell Proteomics* **3**, 809–819.
- 26 Angelini A, Cendron L, Goncalves S, Zanotti G & Terradot L (2008) Structural and enzymatic characterization of HP0496, a YbgC thioesterase from *Helicobacter pylori*. *Proteins* **72**, 1212–1221.
- 27 Hohlfeld S, Pattis I, Puls J, Plano GV, Haas R & Fischer W (2006) A C-terminal translocation signal is necessary, but not sufficient for type IV secretion of the *Helicobacter pylori* CagA protein. *Mol Microbiol* **59**, 1624–1637.
- 28 Hatakeyama M (2004) Oncogenic mechanisms of the *Helicobacter pylori* CagA protein. *Nat Rev Cancer* **4**, 688–694.
- 29 Higashi H, Tsutsumi R, Fujita A, Yamazaki S, Asaka M, Azuma T & Hatakeyama M (2002) Biological activity of the *Helicobacter pylori* virulence factor CagA is determined by variation in the tyrosine phosphorylation sites. *Proc Natl Acad Sci USA* **99**, 14428–14433.
- 30 Higashi H, Yokoyama K, Fujii Y, Ren S, Yuasa H, Saadat I, Murata-Kamiya N, Azuma T & Hatakeyama M (2005) EPIYA motif is a membrane-targeting signal of *Helicobacter pylori* virulence factor CagA in mammalian cells. *J Biol Chem* **280**, 23130–23137.
- 31 Couturier MR, Tasca E, Montecucco C & Stein M (2006) Interaction with CagF is required for translocation of CagA into the host via the *Helicobacter pylori* type IV secretion system. *Infect Immun* **74**, 273–281.
- 32 Kwok T, Zabler D, Urman S, Rohde M, Hartig R, Wessler S, Misselwitz R, Berger J, Sewald N, Konig W *et al.* (2007) *Helicobacter* exploits integrin for type IV secretion and kinase activation. *Nature* **449**, 862–866.
- 33 Tarendeau F, Boudet J, Guilligay D, Mas PJ, Bougault CM, Boulo S, Baudin F, Ruigrok RW, Daigle N, Ellenberg J *et al.* (2007) Structure and nuclear import function of the C-terminal domain of influenza virus polymerase PB2 subunit. *Nat Struct Mol Biol* **14**, 229–233.
- 34 Tarendeau F, Crepin T, Guilligay D, Ruigrok RW, Cusack S & Hart DJ (2008) Host determinant residue lysine 627 lies on the surface of a discrete, folded domain of influenza virus polymerase PB2 subunit. *PLoS Pathog* **4**, e1000136.
- 35 Guilligay D, Tarendeau F, Resa-Infante P, Coloma R, Crepin T, Sehr P, Lewis J, Ruigrok RW, Ortin J, Hart DJ *et al.* (2008) The structural basis for cap binding by influenza virus polymerase subunit PB2. *Nat Struct Mol Biol* **15**, 500–506.
- 36 Xiang Z, Censini S, Bayeli PF, Telford JL, Figura N, Rappuoli R & Covacci A (1995) Analysis of expression of CagA and VacA virulence factors in 43 strains of *Helicobacter pylori* reveals that clinical isolates can be divided into two major types and that CagA is not necessary for expression of the vacuolating cytotoxin. *Infect Immun* **63**, 94–98.
- 37 Backert S, Muller EC, Jungblut PR & Meyer TF (2001) Tyrosine phosphorylation patterns and size modification of the *Helicobacter pylori* CagA protein after translocation into gastric epithelial cells. *Proteomics* **1**, 608–617.
- 38 Odenbreit S, Gebert B, Puls J, Fischer W & Haas R (2001) Interaction of *Helicobacter pylori* with professional phagocytes: role of the cag pathogenicity island and translocation, phosphorylation and processing of CagA. *Cell Microbiol* **3**, 21–31.
- 39 Moese S, Selbach M, Zimny-Arndt U, Jungblut PR, Meyer TF & Backert S (2001) Identification of a tyrosine-phosphorylated 35 kDa carboxy-terminal fragment (p35CagA) of the *Helicobacter pylori* CagA protein in phagocytic cells: processing or breakage? *Proteomics* **1**, 618–629.

Supporting information

The following supplementary material is available:

Fig. S1. Alignment of construct sequences selected for soluble protein production after the small-scale Ni²⁺-nitrilotriacetic acid purification step.

This supplementary material can be found in the online version of this article.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

Supplementary material

(<http://www3.interscience.wiley.com/journal/121628547/supinfo>)

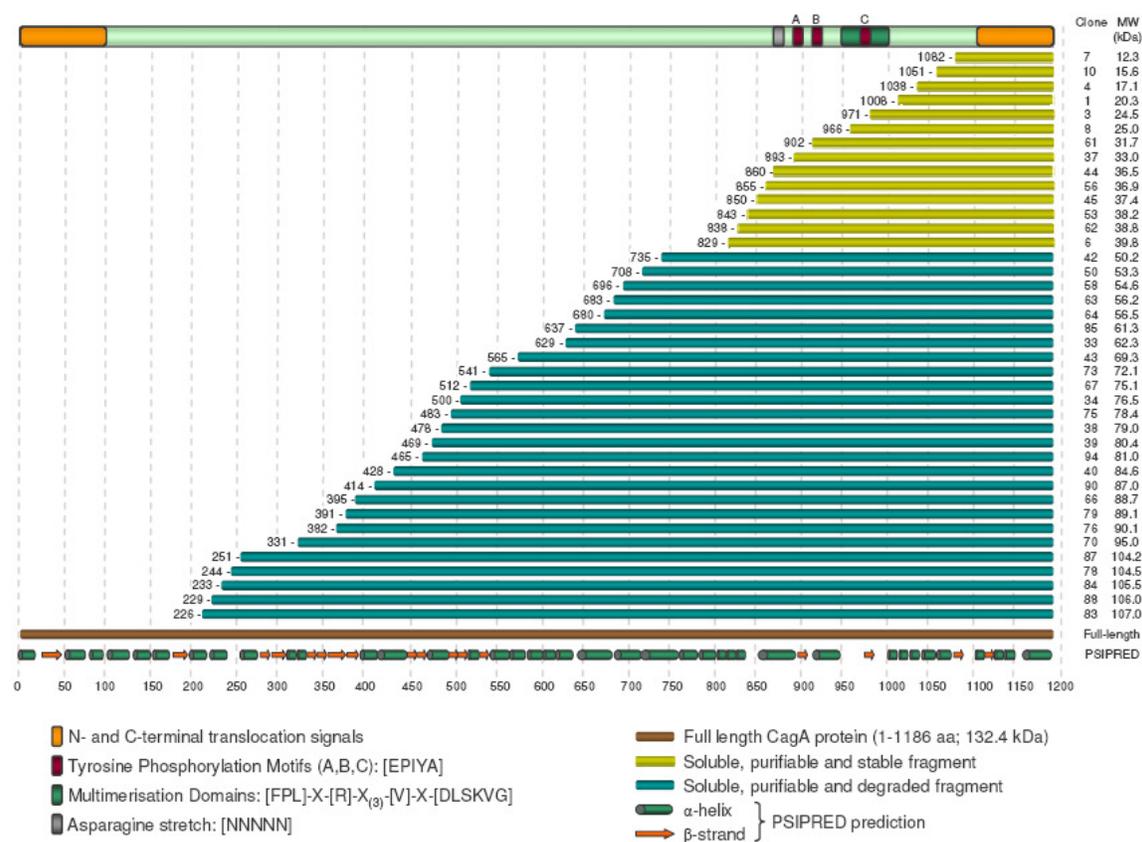


Figure S1. Alignment of construct sequences selected for soluble protein production after small-scale Ni²⁺-NTA purification step. Proteins were visible either directly by SDS-PAGE or following dual detection western-blot. Yellow lines represent constructs purified in a relatively undegraded form while blue indicates the purified longer constructs exhibiting *in vivo* proteolysis, frequently corresponding to loss of a C-terminal region. This construct panel permitted selection of clones for one litre scale expression trials. The CagA linear motif map is presented above with a secondary structure prediction below performed using PSIPRED [McGuffin LJ, Bryson K & Jones DT (2000) The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404–405]. The molecular weight (MW) calculated from sequencing of the genetic insert of each selected clone corresponds to the CagA fragment without additional tags.

Chapter 3

Biochemical and Structural characterization of a 33 kDa C- terminal fragment of CagA

Résumé

Le travail de thèse s'est ensuite concentré sur le fragment C-terminal d'environ 33 kD obtenu par la méthode ESPRIT, CagA^{C33}. Ce fragment est particulièrement intéressant car il contient les motifs (EPIYA) de phosphorylation, les domaines de multimérisation et la séquence nécessaire à l'interaction avec CagF. L'étude structurale et biochimique de ce domaine de CagA a été donc entrepris en trois étapes :

- a) la définition de l'état oligomérique *in vitro* du fragment CagA^{C33} par différentes approches (gel-filtration, dispersion dynamique de la lumière, microscopie électronique)
- b) l'analyse préliminaire de l'interaction entre CagA^{C33} et la protéine CagF
- c) l'analyse par approches biochimiques et biophysiques des interactions entre CagA phosphorylée et les domaines SH2 de la kinase v-Src ou de la phosphatase SHP-2.

a) Les résultats présentés ici montrent que CagA^{C33} adopte plusieurs formes oligomériques en solution et en microscopie électronique. Entre ces formes multimériques, le tétramère semble la forme la plus représentée. Autres résultats suggèrent que cette multimérisation est rendue possible par deux mécanismes. Tout d'abord, une homodimérisation covalente de CagA^{C33} est réalisée par des ponts disulfures entre les cystéines C1168. De plus, la protéine peut former des tétramères par un second mode de dimerisation qui impliquerait vraisemblablement les domaines de multimérisation connus.

b) CagA^{C33} a été ensuite utilisée dans des expériences de protéolyse limitée pour étudier *in vitro* l'interaction avec la protéine CagF. Les résultats, bien que préliminaires, indiquent que les deux protéines interagissent et que cette interaction conduit à des effets différents sur chaque protéine. En présence de CagF, CagA^{C33} devient plus sensible à l'action des protéases, alors que CagF devient moins sensible. De plus l'action des protéases sur CagA produit deux fragments séparés de 11 (CagA^{C11}) et 22kDa (CagA^{C22}) que ont été identifiés. Ceci suggère que l'interaction induit un changement de conformation de CagA^{C33} qui pourrait avoir des implications dans sa translocation par le T4SS d'*H. pylori*.

c) Enfin, il est montré ici que CagA^{C33} est phosphorylé (P-CagA^{C33}) *in vitro* par deux kinases humaines recombinantes, c-Src et c-Abl. Une fois phosphorylé, P-CagA^{C33} interagit avec le domaine SH2 purifié de la kinase virale v-Src (homologue de c-Src). Les interactions de CagA^{C11} avec différents domaines SH2 ont ensuite été étudiées par une approche biophysique. Des mutants de CagA^{C11} ont été construits pour évaluer le rôle de chacun des motifs EPIYA (-A, -B et -C) dans l'interaction avec v-SH2. Le mutant contenant uniquement EPIYA-C phosphorylé (P-CagA^{C11}ΔAΔB) possède la plus haute affinité pour v-SH2. Enfin, les interactions de P-CagA^{C11}ΔAΔB avec la phosphatase SHP-2 ont été étudiées. Cette dernière contient deux domaines SH2 en tandem en amont du domaine phosphatase. Ces résultats indiquent que seul un fragment contenant uniquement les deux domaines SH2 est capable de former un complexe stable avec P-CagA^{C11}ΔAΔB alors que la phosphatase SHP-2 entière ne forme pas un complexe avec ce mutant. Cependant, nos résultats indiquent aussi que SHP-2 est capable de déphosphoryler P-CagA^{C11}. Ceci pourrait donc expliquer l'absence d'interaction entre les deux protéines. Les études présentées dans ce chapitre constituent une première étape importante vers la caractérisation structurale de CagA.

3.1 Expression and purification of CagA^{C33}

As reported in Chapter 2, a 33 kDa C-terminal fragment of CagA (CagA^{C33}) could be expressed and purified following the ESPRIT analysis of the CagA gene. CagA^{C33} encompasses residues N893 to S1186 of the CagA amino acid sequence. Interestingly, a very similar fragment has been reported to be present in mammalian cells and to be phosphorylated by host kinases following CagA translocation (Moese et al., 2001). CagA^{C33} (Figure 3.1) contains the three EPIYA motifs of CagA and one of the sequences necessary for its secretion. This fragment has therefore been characterised from both a biochemical and structural point of view.

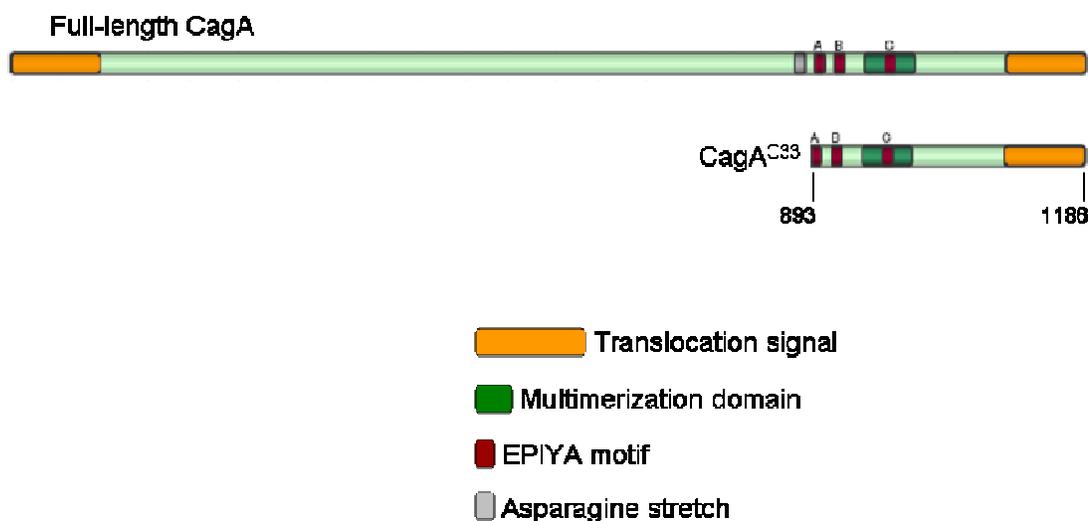


Figure 3.1: CagA functional motifs found in the fragment CagA^{C33}. CagA is characterized by two translocation signals (orange boxes) located at the N-terminus and C-terminus. Both are necessary for translocation. The EPIYA motifs (EPIYA-A, EPIYA-B and EPIYA-C in western strains of *H. pylori*) are tyrosine-phosphorylated by host kinases. Downstream the EPIYA-B and EPIYA-C motifs, two multimerization domains of sixteen residues each are reported to mediate the multimerization of CagA *in vivo*. The asparagine stretch, located upstream CaC³³ is indicated by a grey box.

3.1.1 Construction of a CagA^{C33} fragment suitable for biochemical and structural studies

Clone 37 of the ESPRIT protocol (see Chapter 2) produces a protein, CagA^{C33}, with an additional, non-cleavable C-terminal biotin tag. This tag might affect the biochemical properties of the protein and could also impair crystallization. This tag was therefore removed

by introducing a stop codon (TAA), in position 4703-4705 of the DNA construct pHAR3011-37, upstream of its sequence (nucleotides 4709-4774). Site-directed mutagenesis was carried out by PCR, using the QiaQuick site-directed mutagenesis kit (Qiagen), according to manufacturer's protocol. 50 ng of the plasmid pHAR3011-37 were used as a template and 100 ng of the following mutagenesis primers were added to the reaction: UL89stopfw: 5'-ccatccgtttcagggtcagcTaAgcatcgaacaatgggagtg; UL89stoprv: 5'-ccactcccattgttcgatgcTtAgctgaccctgaaacggatgg (the mutations introduced are indicated in uppercase). The PCR reaction was carried out performing 18 cycles of denaturation (95°C for 30 seconds), primer annealing (55 °C for 30 seconds) and elongation (68°C for 7 minutes). The PCR mix was then digested for 2 hours at 37°C with the restriction enzyme DpnI (1 µl at 5 U/µl), which is specific for methylated DNA (template). *E. coli* DNA is methylated by the enzyme Dam methylase, which transfers a methyl group from 6-adenosylmethionine to the position N6 of adenines located in GATC sequences. As the plasmid pHAR3011-37 was isolated from *E. coli* cells, its GATC sequences will be methylated (Barras & Marinus, 1989). The PCR-synthesized mutated plasmid prepared as described above will not bear any methyl group.

After digestion, *E. coli* strains 10G Duos competent cells were transformed by heat shock with 2 µl of the reaction mixture and plated onto solid agar medium supplemented with 50 µg/ml of kanamycin and grown overnight at 37°C. The plasmids were then isolated by miniprep from the colonies. The DNA sequences of the CagA fragments were determined (Genecore, EMBL Heidelberg, Germany) and the presence of the stop codon was verified. The resulting expression vector was named pHAR3011-37ΔBio and encodes a fusion protein between a TEV-cleavable N-terminal His₆-tag vector and residues N893 to S1186 of CagA. The produced protein will be named His₆-CagA^{C33}.

3.1.2 Expression tests of His₆-CagA^{C33} with pHAR3011-37ΔBio in various *E. coli* strains

To determine the most favorable conditions for protein expression, preliminary expression tests of His₆-CagA^{C33} were performed in 3 ml cultures of various strains of *E. coli*. Protein expression was induced at a cell OD₆₀₀=0.6, with 1 mM IPTG at 37°C and 20°C under thorough agitation. Samples were collected every hour and total cell extracts were loaded on SDS-PAGE gels. A sample was also collected after overnight expression at 20°C. The presence of the protein was detected on SDS-PAGE gels (Figure 3.2).

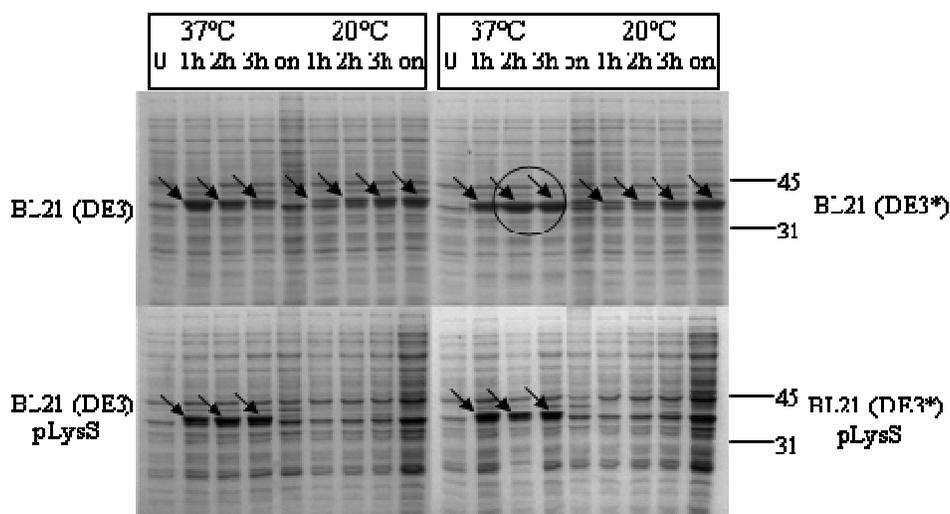


Figure 3.2: small-scale expression test of His₆-CagA^{C33}.

His₆-CagA^{C33} was expressed in *E. coli* strains BL21 (DE3), BL21 (DE3*), BL21 (DE3) pLysS and BL21 (DE3*) pLysS. The expression was monitored at two different temperatures (37°C and 20°C) and at different times (1 to 3 hours after the induction, and overnight). Samples of crude cell extracts were loaded onto SDS-PAGE gels stained with Coomassie blue after the run. The stained SDS-PAGE gels show the overexpression of a ~37 kDa product, corresponding to the His₆-CagA^{C33} construct.

The results presented in Figure 3.2 show a better expression of His₆-CagA^{C33} at 37°C compared to 20°C for the *E. coli* strains BL21 (DE3) and BL21 (DE3*). BL21 (DE3) pLysS and BL21 (DE3*) pLysS strains did not overexpress the protein at 20°C.

3.1.3 Large scale expression and purification of His₆-CagA^{C33}

Large scale expression of His₆-CagA^{C33} was carried out in four litres of LB medium supplemented with 50 µg/ml of kanamycin, inoculated with the glycerol stock of *E. coli* strain BL21 (DE3*). Protein expression was induced at OD₆₀₀=0.6 with 1 mM IPTG for 3 hours at 37°C. After the expression, the cells were centrifuged at ~7000 g for 20 minutes and resuspended in 50 mM Tris pH 8, 200 mM NaCl, 5% glycerol, 0.5% Triton X-100.

The bacterial suspension was subsequently lysed using a French Press (1000 Psi) and centrifuged at ~40000 g for 20 minutes. The soluble fraction was loaded onto a 5 ml HiTrap Ni-NTA column (GE Healthcare), pre-equilibrated with buffer A (50 mM Tris pH 8, 200 mM NaCl, 5% glycerol), using an AKTA Prime system (Amersham Biosciences). Elution was performed using a linear gradient of imidazole (from 0 to 500 mM) in buffer A. His₆-CagA^{C33}

eluted at an imidazole concentration of 350 mM.

Fractions were collected and the total protein amount was estimated using the Bradford colorimetric method (Noble & Bailey, 2009): 1 μ l of protein fraction was added to 999 μ l of 20% diluted Bradford Reagent (Biorad). From the absorbance measured at 595 nm we estimated the protein concentration (in mg/ml) using a calibration curve made with BSA protein solutions concentrated from 0.5 mg/ml to 12 mg/ml. The total amount of purified His₆-CagA^{C33} was estimated to be 30 mg (yield of 7.5 mg/litre of culture, purity ~90%).

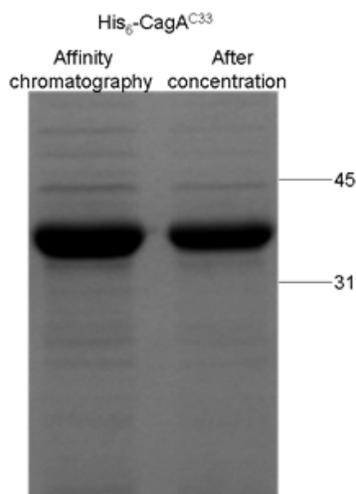


Figure 3.3: SDS-PAGE gel showing the purity grade of His₆-CagA^{C33}

Left side: protein sample after affinity purification. Right side: protein sample after concentration in 4 mM CHAPS detergent. In both cases, the purity is approximately 90%.

The protein was then concentrated by ultrafiltration using a 10 kDa-cutoff centricon column (Millipore). Heavy precipitation and aggregation were observed during the concentration step. This resulted in a significant decrease of the concentration speed and in the presence of flocculated particles. This problem was overcome by adding a detergent (4mM CHAPS) to the protein solution. In the presence of detergent, His₆-CagA^{C33} was concentrated to 4 mg/ml and only slight precipitation was observed. The protein purity at the end of this step was ~90% as estimated from SDS-PAGE electrophoresis (Figure 3.3).

3.2 Oligomeric state of CagA^{C33}

3.2.1 Study of the oligomeric state of His₆-CagA^{C33} by gel-filtration

To increase His₆-CagA^{C33} purity, a gel-filtration chromatography was performed. His₆-CagA^{C33} (500 µl of protein solution at 4 mg/ml) was loaded on a Superdex 200 10/300 GL (GE Healthcare), using 50 mM Tris pH 8, 200 mM NaCl, 4 mM CHAPS as a buffer. The gel-filtration run was carried out at a flow rate of 0.5 ml/min. The resulting chromatogram is reported in Figure 3.4.

As can be seen, His₆-CagA^{C33} elutes in three different peaks, characterized by retention volumes of 9 ml, 11.8 ml and 13.9 ml respectively. The first peak (9ml) corresponds to a macromolecule of very high molecular weight (>440 kDa) possibly aggregated, as its retention volume is very close to the void volume of the column (that is 8.5 ml). The second (11.8 ml) and third (13.9 ml) peaks correspond to 140 kDa and 75 kDa oligomers, respectively, according to the calibration curve (Figure 3.3, panel B). The fractions corresponding to the three peaks were loaded onto a SDS-PAGE gel (Figure 3.4). The gel shows the presence of His₆-CagA^{C33} in every peak. CagA therefore forms multimers of different molecular weights in solution.

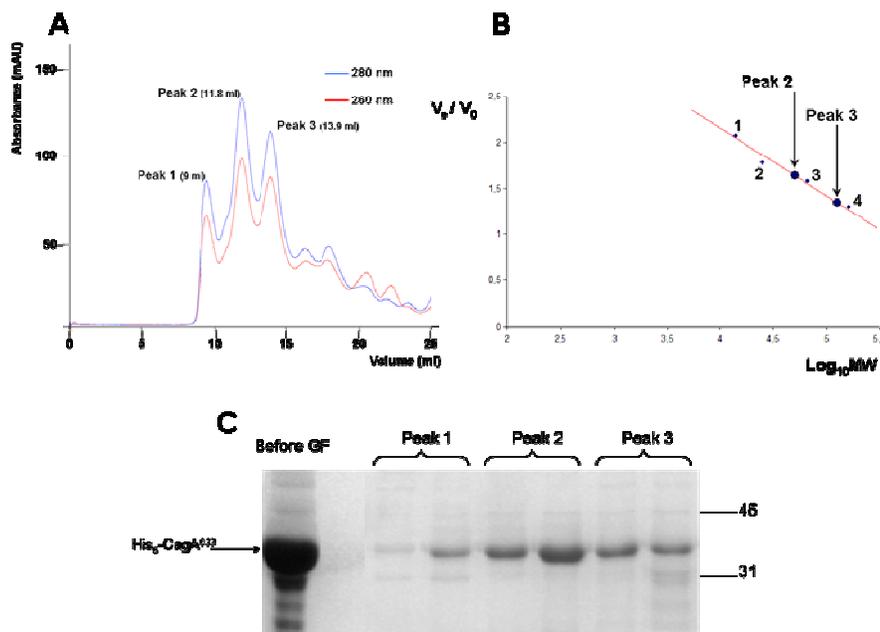


Figure 3.4: Gel-filtration chromatography of His₆-CagA^{C33} and analysis of the molecular weight of the eluted species.

(A): His₆-CagA^{C33} elutes in three different peaks at retention volumes of 9, 11.8 and 13.9 ml, respectively. (B): calibration curve of the Superdex 200 10/300 GL. The Log₁₀ of the molecular weight (Log₁₀MW) was plotted against V_e/V_0 , where V_e is the retention volume of the eluted macromolecule, and V_0 is the void volume of the column (8.5 ml). Four proteins of known molecular weight were loaded onto the column and used as markers: 1-Ribonuclease A (13.7 kDa, $V_e/V_0=2.1$); 2-Chymotrypsinogen A (25 kDa, $V_e/V_0=1.8$); 3-Albumin (67 kDa, $V_e/V_0=1.57$); 4-Aldolase (158 kDa, $V_e/V_0=1.3$). Their Log₁₀MW and V_e/V_0 were plotted to draw the calibration curve. According to this curve, the retention volumes of peak 2, 11.8 ml ($V_e/V_0=1.39$) and peak 3, 13.9 ($V_e/V_0=1.63$) in the gel-filtration (panel A) correspond to a molecular weight of respectively 140 and 75 kDa. (C): SDS-PAGE gel of different CagA multimers eluted in gel-filtration chromatography. The band corresponding to His₆-CagA^{C33} was detected in samples from all three peaks of the gel-filtration.

3.2.2 Analysis of His₆-CagA^{C33} multimers by Dynamic Light scattering (DLS) and comparison with the gel-filtration experiment

To further evaluate the molecular weight of the His₆-CagA^{C33} multimers observed, these were analyzed using DLS (for an overview of this technique, see Appendix, section A.2). 50 μ l of His₆-CagA^{C33} eluted at peak 2 and peak 3 from gel-filtration (both at a concentration of 0.5 mg/ml) were analyzed at 20°C in a Zetasizer Nano DLS machine (Malvern Instruments). The dimensions of the particles were calculated. The molecular weight of His₆-CagA^{C33} particles was estimated assuming that such species are globular and not elongated. DLS profiles of the His₆-CagA^{C33} species eluting in peak 2 and peak 3 are shown in Figure 3.5.

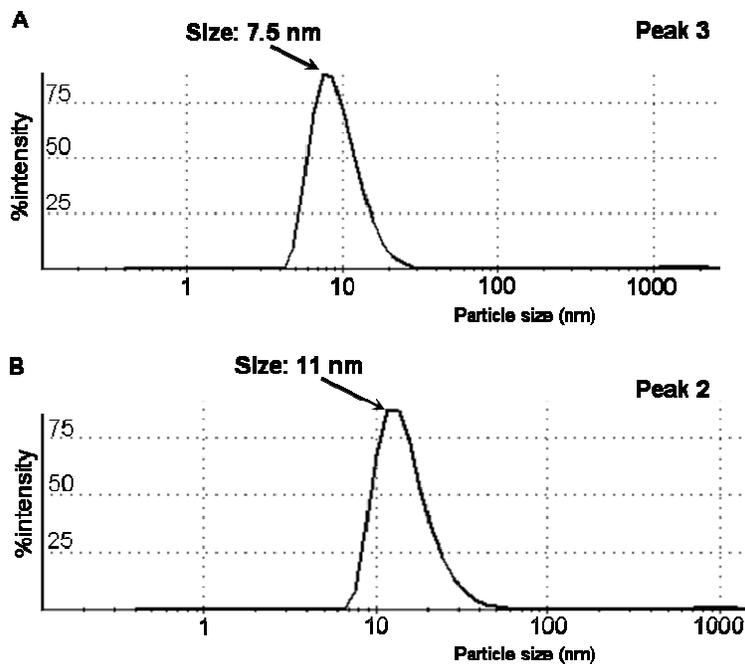


Figure 3.5: DLS measurement of different His₆-CagA^{C33} multimers eluted in gel-filtration chromatography.

The calculated particle sizes are plotted against the intensity (%) of the light scattered by the particle. (A): peak of scattering intensity corresponding to the His₆-CagA^{C33} from peak 2. The calculated particle size is 7.5 nm. (B): peak of scattering intensity corresponding to the His₆-CagA^{C33} from peak 3. The calculated particle size is 11 nm.

From the particles sizes, the estimated molecular weights are: 70 kDa for the 7.5 nm particle, 160 kDa for the 11 nm particle. As the molecular weight of His₆-CagA^{C33} monomer is 36.4 kDa, it is possible to propose that the 7.5 nm particle observed in peak 3 is consistent with a dimer of His₆-CagA^{C33} (MW: 72.8 kDa), whilst the 11 nm particle observed in peak 2 could correspond to a His₆-CagA^{C33} tetramer (MW: 146 kDa).

3.2.3 Role of detergents in the stabilization of His₆-CagA^{C33} tetramer

The role of different detergents in the stabilization of the multimers was then examined. His₆-CagA^{C33} was purified in the presence of two mild saccharidic detergents, n-octyl-β-D-glucoside (OG) and n-Dodecyl-β-D-maltoside (DDM), at concentrations of 20 mM and 0.5 mM respectively. The gel-filtration experiments were then repeated according to the protocol reported in paragraph 3.2.1 (Figure 3.6).

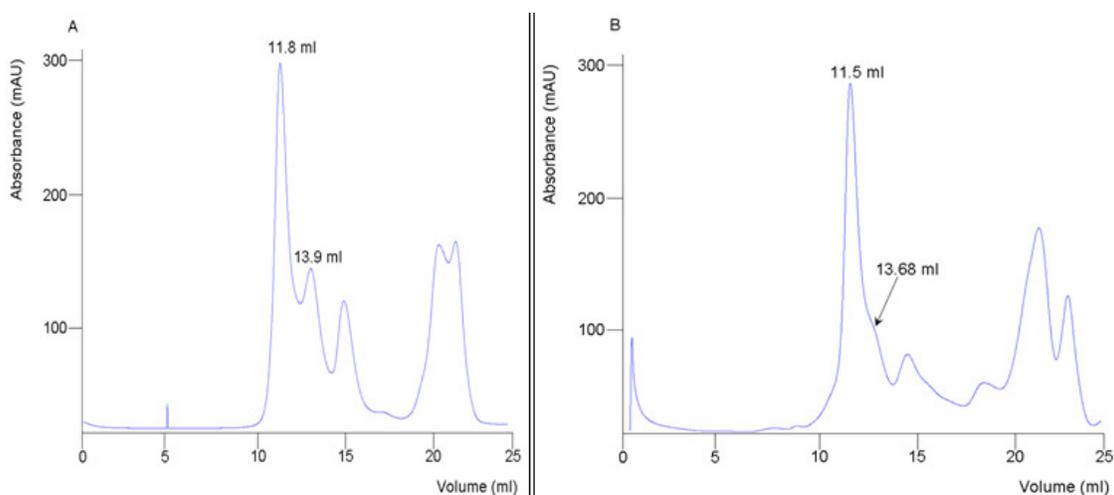


Figure 3.6: gel-filtrations of His₆-CagA^{C33} in the presence of the detergents OG and DDM.

(A): gel-filtration of His₆-CagA^{C33} in the presence of OG. There is no variation of the retention volumes of His₆-CagA^{C33} tetramer (11.8 ml) and dimer (13.9 ml). (B): gel-filtration of His₆-CagA^{C33} in the presence of DDM. A slight variation of the retention volume of His₆-CagA^{C33} tetramer and dimer (11.5 and 13.68 ml, respectively) was observed, but not correlated with a molecular weight variation of such species according to the calibration curve.

In contrast with the gel-filtration experiments reported in Figure 3.4, the presence of the aggregates that previously eluted at a retention volume of 9 ml was not observed. In the presence of both OG and DDM His₆-CagA^{C33} elutes in only two peaks corresponding, as confirmed by further DLS measurements, to His₆-CagA^{C33} tetramers and dimers.

In the presence of both OG and DDM, a significant change in the relative abundance of the peak corresponding to the His₆-CagA^{C33} tetramer was observed. In the presence of these two detergents, this is the preponderant His₆-CagA^{C33} species in solution.

3.2.4 Involvement of Cysteine 1164 in His₆-CagA^{C33} multimerization

The amino acid sequence of His₆-CagA^{C33} contains a cysteine at position 1164. It is possible that this cysteine may participate in a C₁₁₆₄-C₁₁₆₄ disulphide bond between two His₆-CagA^{C33} subunits. To test this hypothesis, the gel-filtration fractions corresponding to the tetramer and dimer peaks were loaded onto a SDS-PAGE gel without any reducing agents. If this hypothesis is correct a ~70 kDa band, corresponding to a His₆-CagA^{C33} covalent dimer should be observed. The resulting gel (Figure 3.7) shows a band corresponding to a dimer of (molecular weight ~70 kDa) in the absence of reducing agent. It can thus be concluded that, when its thiol group is not reduced, C1164 is involved in the formation of His₆-CagA^{C33} covalent dimers.

To confirm the role of C1164 in His₆-CagA^{C33} tetramerization, site directed mutagenesis was performed to change this residue to a serine. The mutagenesis was performed using the same protocol (Section 3.1.1) used to introduce a stop codon inside the *Nsi*I site of pHAR3011-37, but with the following primers: 37CS fw: 5'-tctacaggatattac**A**gcttggcgagggg; 37CS rv: 5'-tcctcgccaagc**T**gtaatacctgtag (bold capital letters indicate the mutated base). The new construct, CagA^{C33}C1164S, was sequenced to confirm the presence of the mutation. Expression and affinity purification were carried out according to the protocol previously established for His₆-CagA^{C33} and described in Section 3.1.3. The protein oligomerisation state was analyzed by gel-filtration in the presence of 0.5 mM DDM, applying the same protocol used for His₆-CagA^{C33}. The result of the gel filtration experiments on CagA^{C33}C1164S are shown in Figure 3.8.

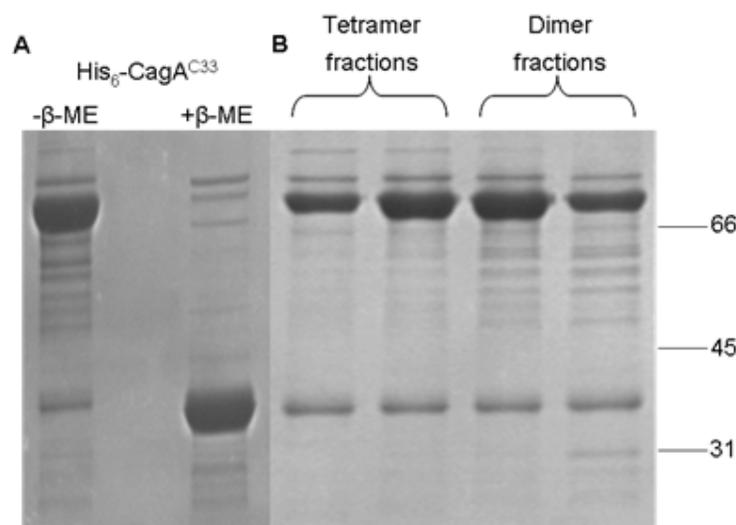


Figure 3.7: SDS-PAGE gels of His₆-CagA^{C33} in the presence and absence of reducing agents.

(A): SDS-PAGE gel of His₆-CagA^{C33} in the absence (-β-ME) and presence (+β-ME) of β-Mercaptoethanol in the sample buffer. (B): SDS-PAGE of gel-filtration fractions corresponding to His₆-CagA^{C33} tetramer and dimer. The run has been performed without β-Mercaptoethanol in the sample buffer.

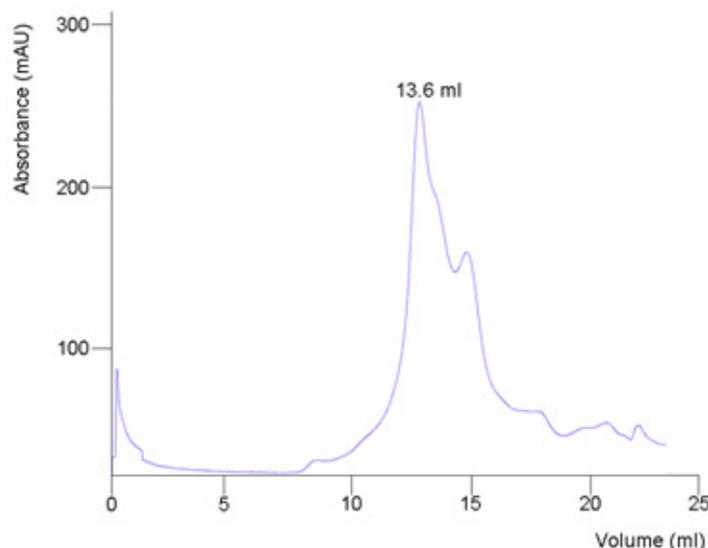


Figure 3.8: gel-filtration of His₆-CagA^{C33}C1164S

In gel filtration experiments the mutated protein elutes at a retention volume of 13.6 ml.

The gel-filtration chromatography of His₆-CagA^{C33}C1164S shows that the protein elutes at a retention volume of 13.6 ml and that His₆-CagA^{C33}C1164S is a dimer. As no peak corresponding to His₆-CagA^{C33}C1164S tetramer (elution volume ~11.8 ml) was observed, it can also be concluded that His₆-CagA^{C33}C1164S does not form tetramers.

These experiments indicate that, in solution, His₆-CagA^{C33} forms dimers by the formation of a disulfide bond between two His₆-CagA^{C33} C1164 residues. More over, this dimer oligomerisation is extended by the formation of dimer of dimers.

3.3 Electron Microscopy analysis of His₆-CagA^{C33}

His₆-CagA^{C33} was then analyzed by Electron Microscopy, in collaboration of Dr. Irina Gutsche (UVHCI, Grenoble). Two samples of His₆-CagA^{C33} corresponding to the peak 2 of the gel filtration were used at a concentration of 1 mg/ml and in the presence of either 20 mM OG (OG-His₆-CagA^{C33}) or 0.5 mM DDM (DDM-His₆-CagA^{C33}).

3.3.1 Sample preparation

The protein solution was applied to the clear side of carbon on a carbon-mica interface and stained with 2 % (w/v) sodium silicotungstate. Images were recorded under low-dose conditions with a JEOL 1200 EX II microscope at 100 kV and a calibrated magnification of 39,750 (based on the helical pitch of tobacco mosaic virus). Selected negatives were then digitized on a Zeiss scanner (Photoscan TD) at a step size of 14 micrometer giving a pixel size of 3.5 Å at the specimen level. Using EMAN image processing software, 13036 subframes containing individual CagA particles were selected interactively from several micrographs. The translationally centered data set was subjected to multivariate statistical classification with Imagic software package.

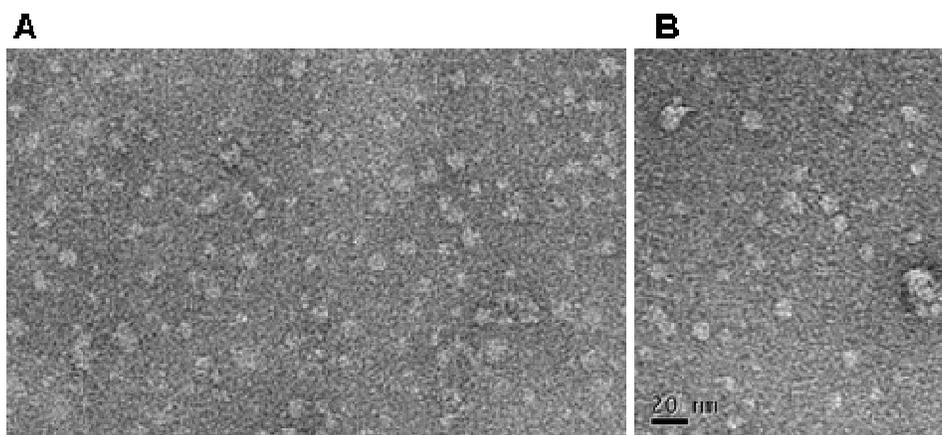


Figure 3.9: Electron micrographs of OG-His₆-CagA^{C33} and DDM-His₆-CagA^{C33}.

(A): OG-His₆-CagA^{C33} (B): DDM-His₆-CagA^{C33}.

3.3.2 Image analysis

Electron micrographs from OG-His₆-CagA^{C33} and DDM-His₆-CagA^{C33} preparations are shown in Figure 3.9. Both show the presence of defined particles of various sizes. The particles in the micrographs corresponding to OG-His₆-CagA^{C33} seem more homogeneous than those obtained with DDM-His₆-CagA^{C33}. The latter contains different particles, with dimensions ranging from ~10 nm to more than 20 nm and characterized by an amorphous shape.

A good classification and analysis of the shapes that the protein assumes in solution was possible for the OG-His₆-CagA^{C33} only. This sample is nevertheless heterogeneous and characterized by the presence of several type of multimers. The different multimers tend to

form, in all types of particles, a fairly symmetric shape, characterized in some cases by a sort of cavity in the centre. Interestingly we could discriminate between particles with apparent trimeric, tetrameric and hexameric symmetries (Figure 3.10).

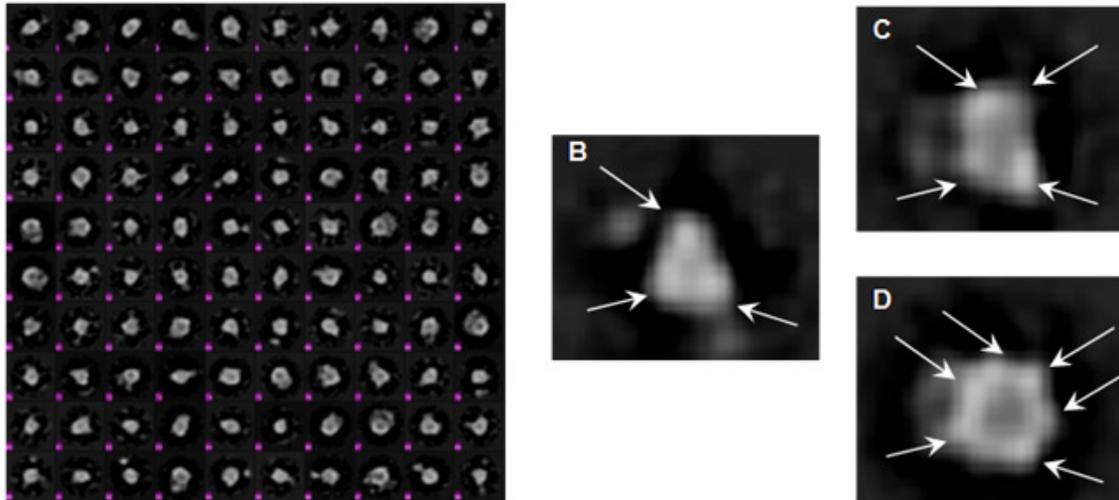


Figure 3.10: observed shapes of OG-His₆-CagA^{C33} particles

(A): Ensemble of isolated observed particles. (B): particle characterized by a presumably trimeric arrangement. (C): tetrameric particle. (D): hexameric particle.

3.3.3 Cross-linking experiments

Due to the heterogeneity of the samples, it was decided to try to stabilize them by crosslinking both OG-His₆-CagA^{C33} and DDM-His₆-CagA^{C33} with glutaraldehyde. 1 µl of a 2.5% glutaraldehyde stock was added to a 50 µl solution of purified OG-His₆-CagA^{C33} and DDM-His₆-CagA^{C33} at a concentration of 1 mg/ml (50 µg of protein). 10 µl of solution were taken every 5 minutes to analyze the crosslinking reaction in a gradient SDS-PAGE gel (Figure 3.11).

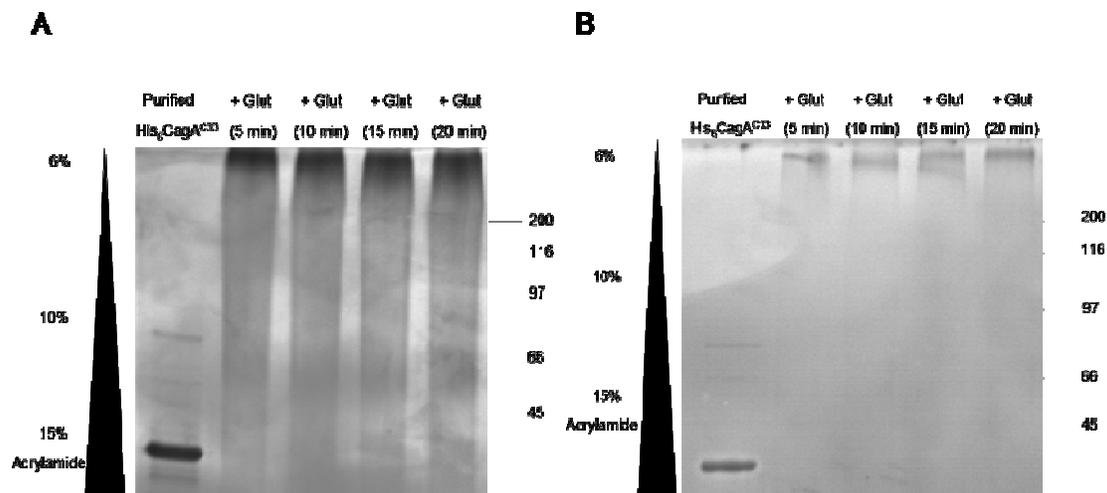


Figure 3.11: crosslinking of His₆-CagA^{C33} with 0.05% glutaraldehyde.

(A): crosslinking of OG-His₆-CagA^{C33}. (B): crosslinking of DDM-His₆-CagA^{C33}. 10 μ l of reaction were taken every 5 minutes after the addition of glutaraldehyde. The samples were subsequently loaded onto a gradient SDS-PAGE gel with acrylamide phases of 6%, 10% and 12% (from top).

The crosslinking experiment shows that both OG-His₆-CagA^{C33} and DDM-His₆-CagA^{C33} form, after 5 minutes, a high molecular weight oligomer (>200 kDa). The oligomeric sample from the DDM-His₆-CagA^{C33} was subsequently analyzed by EM. However, as shown in Figure 3.12, the crosslinked DDM-His₆-CagA^{C33} sample proved to be highly aggregated and heterogeneous and unsuitable for any further analyses. As the sample from OG-His₆-CagA^{C33} presents the same level of oligomerization as DDM-His₆-CagA^{C33} (see Figure 3.11), the EM analysis of OG-His₆-CagA^{C33} oligomers was not performed.

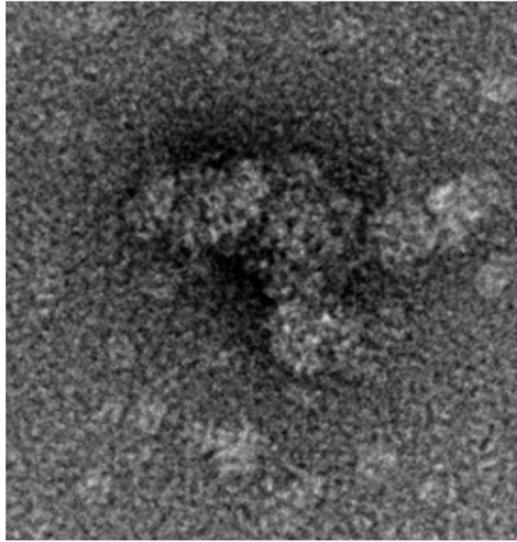


Figure 3.12: electron microscopy screenshot of DDM-His₆-CagA^{C33} crosslinked with glutaraldehyde.

3.4 Interaction of His₆-CagA^{C33} with the *H. pylori* protein CagF

An important CagA interacting partner is the *H. pylori* protein CagF, previously proposed to bind to the C-terminus of CagA and to function as a chaperone before its secretion (Pattis et al., 2007). His₆-CagA^{C33} and purified CagF (provided by Prof. Zanotti's Group, University of Padova, Italy) were therefore used to study the interaction.

3.4.1 His₆-CagA^{C33}/CagF interaction by pull-down assay

A pull-down assay with His₆-CagA^{C33} and CagF (with no tags) was set-up. 400 µg of CagF and 200 µg of His₆-CagA^{C33} were mixed in buffer A (50 mM Tris pH 8, 100 mM NaCl) at a total volume of 500 µl. The mixture was loaded onto a Spin-Trap column and centrifuged at 100g. The column was washed three times with 600 µl of buffer A and 3 times with 10% buffer B (50 mM Tris pH 8, 100 mM NaCl, 50 mM imidazole) to elute the unbound protein. Elution was done with 50 mM Tris pH 8, 100 mM NaCl, 500 mM imidazole. The same procedure was performed with CagF alone, as a negative control (Figure 3.13).

A SDS gel shows that CagF binds aspecifically to the Ni-NTA resin without any tag, and is just partially washed off the resin with buffers A and B, both in the presence and absence of His₆-CagA^{C33}. Therefore a pull-down assay was not appropriate to test this interaction.

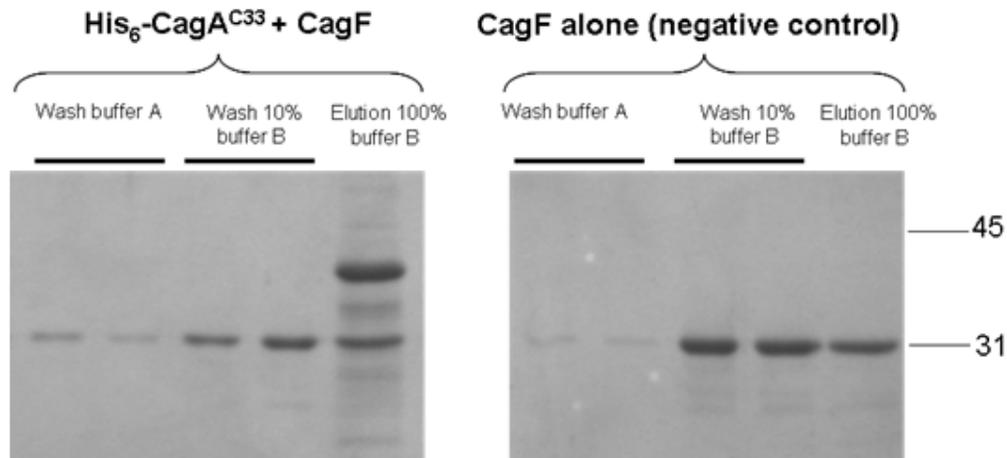


Figure 3.13: pull-down interaction study between His₆-CagA^{C33} and CagF.

Left: the protein CagF is present in all samples from two washes performed with buffer A and 10% buffer B, and then in the elution with 500 mM imidazole, together with His₆-CagA^{C33}. Right: CagF is also present in two washes performed with buffer A and 10% buffer B, and then in the elution with 500 mM imidazole.

3.4.2 His₆-CagA^{C33}/CagF interaction by limited proteolysis

Limited trypsin proteolysis was then used to assess whether the interaction between His₆-CagA^{C33} and CagF might protect one, or both, of the proteins, from proteolysis. 50 µg of His₆-CagA^{C33}, 50 µg of CagF (as controls) and a mixture of 50 µg of His₆-CagA^{C33} and CagF (100 µg in total) were used. Each sample was incubated with trypsin, (protease/protein ratio of 1/1000) on ice for 1 hour and proteolysis was monitored on a 12% SDS-PAGE gel (Figure 3.14).

The experiment revealed several interesting features. First, His₆-CagA^{C33} is proteolyzed into two fragments of around 11 and 20 kDa that resist further degradation. Second, when His₆-CagA^{C33} is in the presence of CagF it is degraded faster, compared to the protein alone. In contrast, in the His₆-CagA^{C33} - CagF mixture, CagF is still degraded but the reaction seems slower. This was confirmed by using different proteases, subtilysin and thermolysin, which gave similar results (Figure 3.15).

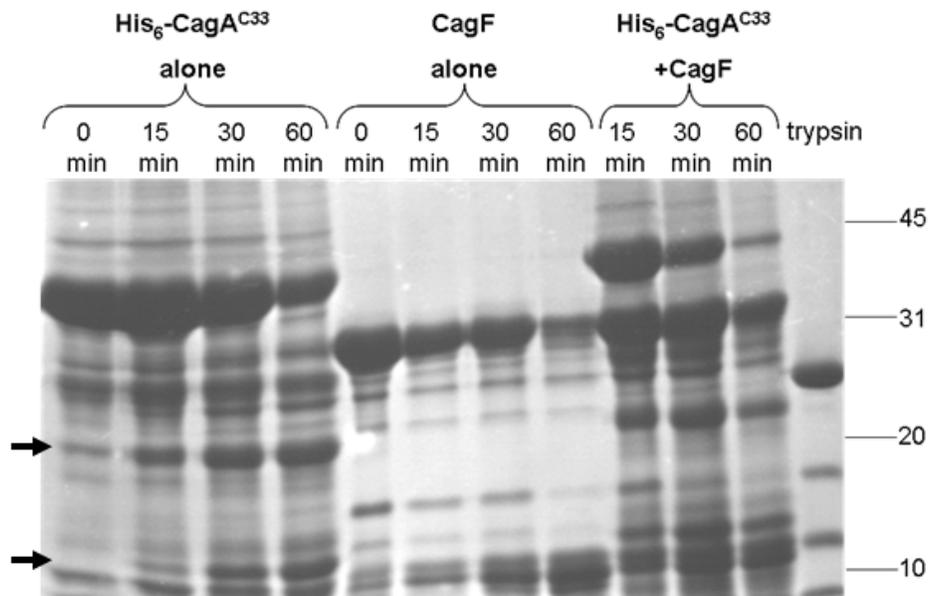


Figure 3.14: limited trypsin proteolysis experiment of His₆-CagA^{C33}/CagF putative complex.

Samples at 5 minutes, 30 minutes and 60 minutes of each reaction (His₆-CagA^{C33} alone, CagF alone and His₆-CagA^{C33}+CagF) have been taken and loaded onto a SDS-PAGE gel. His₆-CagA^{C33} is degraded into two smaller fragments of around 20 and 11 kDa. The arrows indicate His₆-CagA^{C33} degradation fragments resulting from the proteolysis.

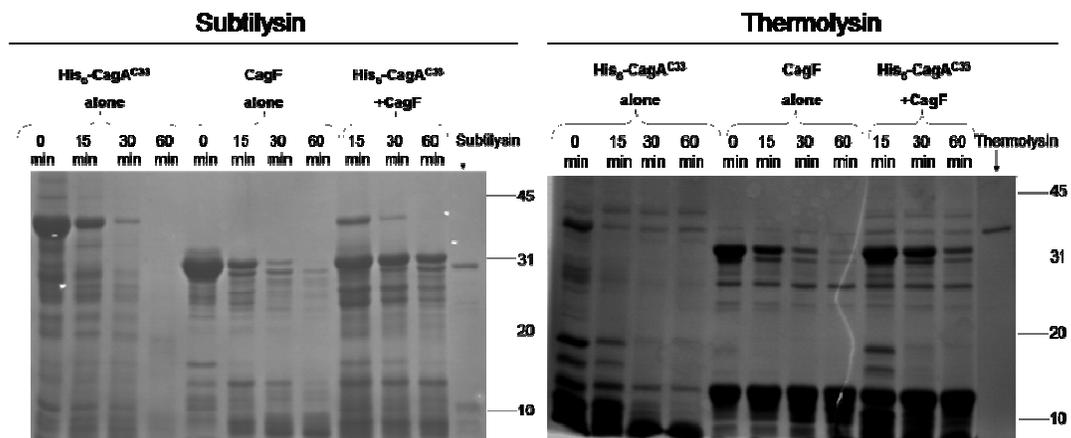


Figure 3.15: limited proteolysis experiment of His₆-CagA^{C33}/CagF putative complex with subtilysin and thermolysin.

Left: limited proteolysis of His₆-CagA^{C33} alone, CagF alone and His₆-CagA^{C33}/CagF with subtilysin. Right: limited proteolysis of His₆-CagA^{C33} alone, CagF alone and His₆-CagA^{C33}/CagF with thermolysin. In both cases, the limited proteolysis was performed by following the protocol described when using trypsin.

Therefore, based on the evidence of these proteolysis experiments, His₆-CagA^{C33} and CagF seem to interact and the interaction has different consequences on the proteins: protection of CagF from proteolysis and an increased sensitivity of His₆-CagA^{C33} to degradation. This suggests that the interaction mediates conformational changes where CagF is protected from the protease while CagA becomes more accessible.

3.5 The domains within His₆-CagA^{C33}: identification of CagA^{C11}

The fact that His₆-CagA^{C33} is proteolyzed by trypsin into two fragments of ~11 and ~20 kDa that resist degradation suggests that His₆-CagA^{C33} might be composed of two different domains.

To investigate this, the trypsin limited proteolysis of His₆-CagA^{C33} was repeated and the samples were blotted onto a PVDF membrane by western-blot. The bands corresponding to the 20 kDa fragment were cut and sent to the Institut de Biologie Structurale (IBS) for N-terminal sequencing by Edman degradation (Tschesche & Dietl, 1975). The sequencing results show that this fragment starts at S999 (sequence SRNQE), immediately after the second multimerization domain and has a molecular weight of 22 kDa. this fragment will be referred to as CagA^{C22}.

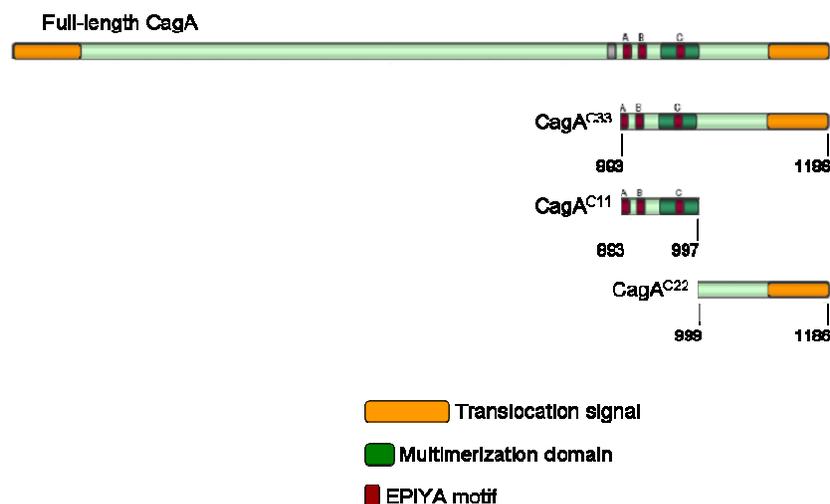


Figure 3.16: Domain boundaries of CagA^{C33}.

Fragments CagA^{C22} and CagA^{C11} as deduced from N-terminal sequencing are indicated. Functional domains have been described in Fig. 3.1.

As the sum of the two degradation fragments yields the exact molecular weight of His₆-CagA^{C33} (33 kDa), we conclude that the 11 kDa fragment (named CagA^{C11}) is comprised of the residues N893 and G997, with the trypsin cut site at R998. An overview of the domains is reported in Figure 3.16.

3.5.1 Cloning, expression and purification of CagA^{C11}

As CagA^{C11} contains the three EPIYA motifs found in CagA^{C33}, a new construct was designed to express CagA^{C11}. The corresponding DNA sequence was amplified by PCR using the following primers: CagA^{C11} fw: 5'-caccaacagcacagaacccattatg; CagA^{C11} rv: 5'-ttaccctacctactgagatc. Subsequently, the PCR product was cloned into pET151/D- TOPO. To check the insert, the plasmidic DNA was amplified in *E. coli* 10G Duos cells, isolated by miniprep and sequenced.

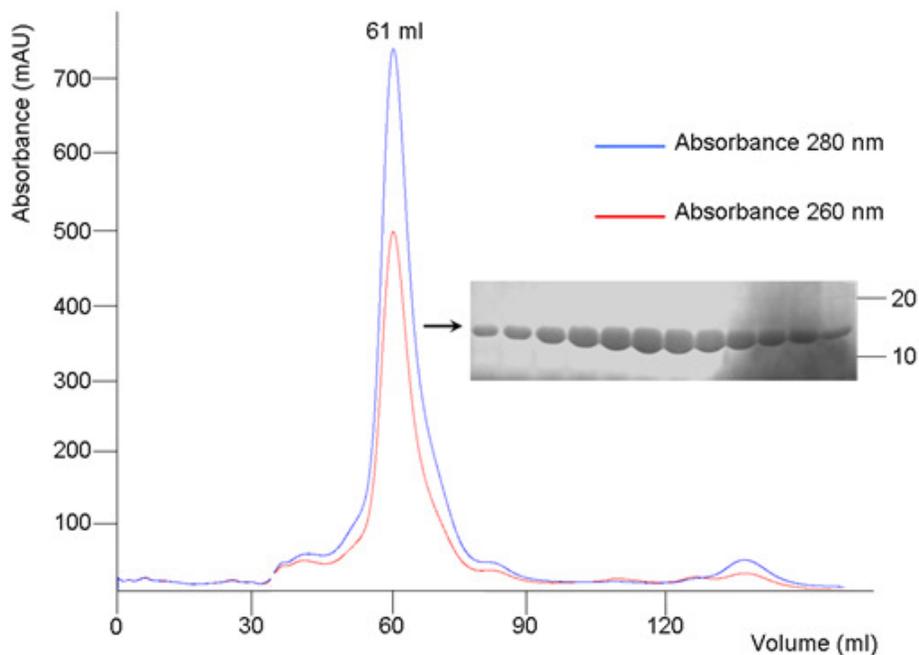


Figure 3.17: gel-filtration chromatography of His₆-CagA^{C11}.

His₆-CagA^{C11} eluted as a single peak (61 ml) on a Sephacryl PrepGrade s100. Fractions from the peak were loaded on a SDS_PAGE gel (right) to ascertain the protein purity.

His₆-CagA^{C11} was expressed at 30°C in two litres of LB medium inoculated with transformed *E. coli* strain BL21 (DE3), inducing the cells with 1 mM IPTG after reaching

O.D.=0.6. Cells were pelleted, resuspended in 30 mM Tris pH 8, 200 mM NaCl, 5% glycerol and lysed with a cell disrupter. After centrifuging the lysate at ~40000 g, the soluble fraction was loaded into a 5 ml HiTrap Ni-NTA column (GE Healthcare) with an AKTA Prime system. The column was previously equilibrated with 30 mM Tris pH 8, 200 mM NaCl, 5% glycerol. Elution of the protein was performed with a linear gradient of imidazole (from 0 to 500 mM). His₆-CagA^{C11} proved to be very soluble yielding 100 mg of protein per litre processed.

During the concentration process some protein precipitation was observed. The buffer was changed for 30 mM HEPES pH 7.5, 200 mM NaCl, 5% glycerol and a gel-filtration was performed using a HiPrep 16/60 s100 preparative column (GE Healthcare). The protein was purified to near homogeneity (Figure 3.17).

3.5.2 Dimerization of His₆-CagA^{C11}

An analytical gel filtration on CagA^{C11} was performed on a Superdex 200 10/300 GL and showed that the calculated molecular weight of CagA^{C11} was 18 kDa (Data not shown). This suggests that CagA^{C11} is in equilibrium between dimer (22 kDa) and monomer (11 kDa). To investigate whether the His₆-CagA^{C11} can dimerize or multimerize, a cross-linking experiment was performed by using the protocol described for His₆-CagA^{C33} (paragraph 3.3.3). The results are shown in Figure 3.18.

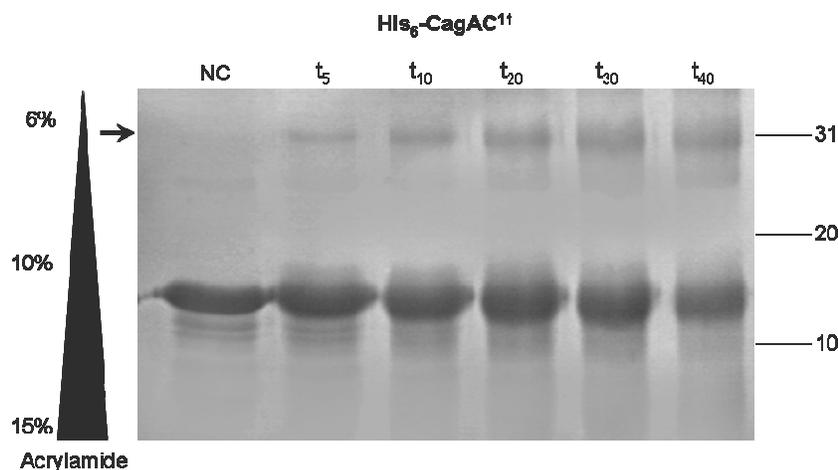


Figure 3.18: cross-linking experiment of His₆-CagA^{C11}

To evaluate the oligomerization state of His₆-CagA^{C11}, the protein was mixed with glutaraldehyde at a final concentration of 0.05%. Samples were collected after 5, 10, 20, 30 and 40 minutes and loaded

onto a gradient SDS-PAGE gel (the concentrations of acrylamide in the gel were 6%, 10% and 15% from top).

The results clearly show that His₆-CagA^{C11} is cross-linked and forms a species characterized by a molecular weight of around 30 kDa (marked in Figure with an arrow). As the molecular weight of His₆-CagA^{C11} is around 15 kDa, the cross-linked band visible in Figure 3.18 is a dimer of His₆-CagA^{C11}.

3.6 *In vitro* phosphorylation of CagA^{C33} and CagA^{C11}

3.6.1 Phosphorylation of His₆-CagA^{C33}

One major aspect of the mode of action of CagA is that it is phosphorylated in human cells on tyrosine residues in EPIYA motifs. To test the functionality of His₆-CagA^{C33}, its *in vitro* phosphorylation by two human kinases, c-Src and c-Abl, was tested.

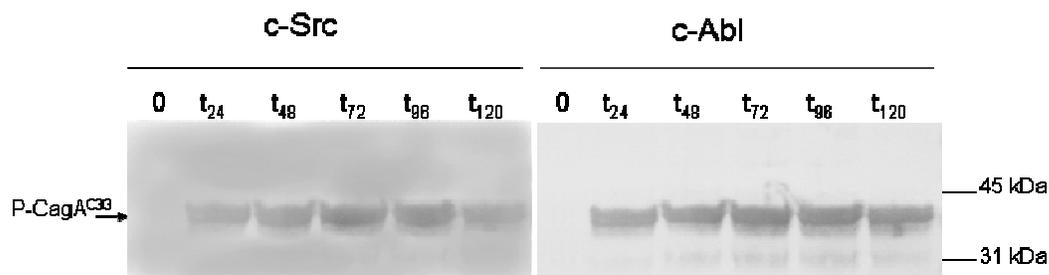


Figure 3.20 Phosphorylation assay of purified CagA^{C33} using c-Src kinase (left) or c-Abl kinase (right). Western blot analysis of phosphorylated CagA. After 24 minutes (t₂₄), CagA^{C33} is phosphorylated by both kinases and the intensity of the band does not increase with time (t₄₈ to t₁₂₀).

A phosphorylation assay was set-up in which 200 µg of purified His₆-CagA^{C33} at 1 mg/ml were diluted with the c-Src (125 mM Hepes pH 7.5, 7.5 µM MgCl₂, 7.5 µM MnCl₂, 2.5 mM DTT) or c-Abl phosphorylation buffer (30 mM Tris pH 7.5, 0.01% NP-40, 10 mM MgCl₂) and with 5 mM ATP in a total volume of 200 µl. 2 µg of c-Src or c-Abl kinase (ProQinase GmbH) at 0.25 mg/ml were then added to the solution. The reactions were carried out for 2 hours at 30°C, taking samples every 24 minutes. To detect the presence of phosphorylated His₆-CagA^{C33}, primary mouse anti-phosphotyrosine antibodies (Sigma-Aldrich) in a dilution of 1:2000 (Figure 3.20) were used, followed by incubation with

secondary anti-mouse antibodies directed against the fragment antigen binding (Fab) of the primary ones. Staining was performed using NBT/BCIP afterwards.

The Western blot analyses show that His₆-CagA^{C33} is phosphorylated by both c-Src and c-Abl. The intensity of the phosphorylated His₆-CagA^{C33} bands on the Western blots does not change with time suggesting that the phosphorylation of His₆-CagA^{C33} has reached its maximum after 24 minutes under these experimental conditions. His₆-CagA^{C33} is thus functional as it can be phosphorylated efficiently by the recombinant human kinases c-Src and c-Abl.

3.6.2 Phosphorylation of His₆-CagA^{C11}

To confirm the functionality of His₆-CagA^{C11} a phosphorylation assay was carried out with the c-Src kinase and the same protocol as for His₆-CagA^{C33}. Figure 3.21 shows that His₆-CagA^{C11} can be phosphorylated by c-Src and, like His₆-CagA^{C33}, this fragment is fully functional in terms of phosphorylation.

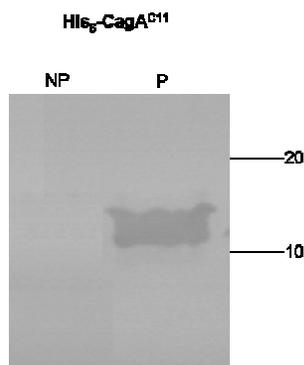


Figure 3.21: Western blot of phosphorylated His₆-CagA^{C11}. NP: non phosphorylated His₆-CagA^{C11}; P: phosphorylated His₆-CagA^{C11}

3.7 Interaction studies of His₆-CagA^{C33} with the SH2 domain of the v-Src kinase

As previously reported (see Introduction), phosphorylated CagA binds to, and activates, SHP-2, triggering signaling cascades that lead to cell scattering and elongation. SHP-2 is characterized by two SH2 domains located at the N-terminus of the protein and termed N-SH2 and C-SH2 (Figure 3.22). N-SH2 is involved in the activation of the

phosphatase when bound to a phosphorylated target, whilst C-SH2 is involved in the binding but not activation (Hof et al., 1998).

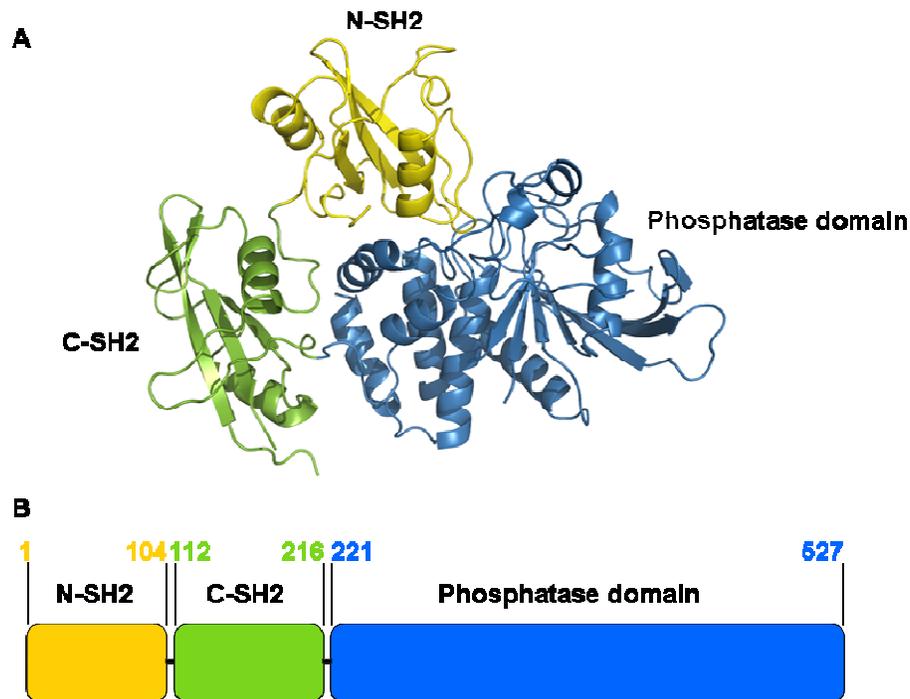


Figure 3.22: SHP-2 phosphatase domain boundaries.

(A): crystal structure of human SHP-2 (PDB accession code: 2SHP). The N-SH2 and C-SH2 domains are characterized by the same fold: a three stranded antiparallel β -sheet flanked by two short α -helices, all connected by several loop regions. (B): domain boundaries of SHP-2.

As His₆-CagA^{C33} is phosphorylated by the human c-Src kinase *in vitro*, its ability to bind a SH2 domain was tested. The SH2 domain tested was from the oncogenic v-Src kinase (homologue to the human c-Src kinase) has been kindly provided by the Group of Prof. Gabriel Waksman.

3.7.1 SH2 domains: general features and functions

The Src Homology region 2 (SH2) domains are interaction modules that recognize and bind mainly tyrosine-phosphorylated substrates. They are common in signaling proteins, where they couple activated protein tyrosine kinases (PTKs) to several signal transduction pathways (Pawson, *et al.*, 2001). Up to now, 120 SH2 domains have been identified in 110

human proteins, of which 10, like the SHP-2 phosphatase shown in Figure 3.21 contain dual domains (Liu, *et al.*, 2006).

SH2 domains comprise approximately 100 residues, which are structurally arranged to form a α - β fold where a β -strand composed by three long β -sheets is flanked by two α -helices, located at both the N-terminus and C-terminus of the domains. The central β -sheet separates a conserved phosphotyrosine binding pocket from a more variable binding surface that interacts with residues located downstream the phosphotyrosine (Liu, 2006). Most commonly, the mechanism of binding, as analyzed from crystal structures of phosphopeptide/SH2 complexes involves a conserved arginine residue, located in the β E strand (see General Conclusions and Perspectives, figure C.1, panel C) that is critical for the interaction with the phosphotyrosine (Waksman, *et al.*, 1993, Liu, *et al.*, 2006). The phosphopeptide crosses the central β -sheet and exposes at least three residues downstream the phosphotyrosine to the variable binding surface.

3.7.2 Cloning, expression and purification of v-Src SH2 domain

This SH2 domain was cloned into a Pet3a plasmid (Waksman *et al.*, 1992), a plasmid that does not include a histidine tag in frame with the cloned gene. As the purification of such construct would have required longer times and purification steps, it was amplified by PCR and cloned into the plasmid pET151/D-TOPO. The amplification of the gene was done using pET3a-SH2 as a template and with the primers: fw: 5'-CACCCAGGCTGAAGAGTGGTACTTTGG; rv: 5'-TCACTTGGACGTGGGGCAGACG.

Cloning into pET151/D- TOPO was carried out according to the manufacturer's protocol (Invitrogen). The plasmid was then amplified and isolated by miniprep from transformed *E. coli* strain 10G Duo (Lucigen) and used to transform *E. coli* BL21 (DE3) cells. Expression of His₆-SH2 was performed at 37°C in 2 litres of liquid LB medium supplemented with 100 μ g/ml of ampicillin. Cells were induced at O.D.= 0.6 with 1 mM final concentration of IPTG. Cells were harvested by centrifugation, resuspended in 50 mM tris pH 8, 200 mM NaCl, 5% glycerol and lysed by sonication. The soluble fraction was recovered by centrifugation and injected into a 5 ml HiTrap Ni-NTA column (GE Healthcare) with an Akta Prime System (Amersham), to purify the protein by affinity chromatography. His₆-SH2 was eluted with 300 mM imidazole. Samples from protein fractions were loaded onto a SDS-PAGE gel. His₆-SH2 was purified almost to homogeneity (98%, Figure 3.23) and at high yields (50 mg per litre of culture).

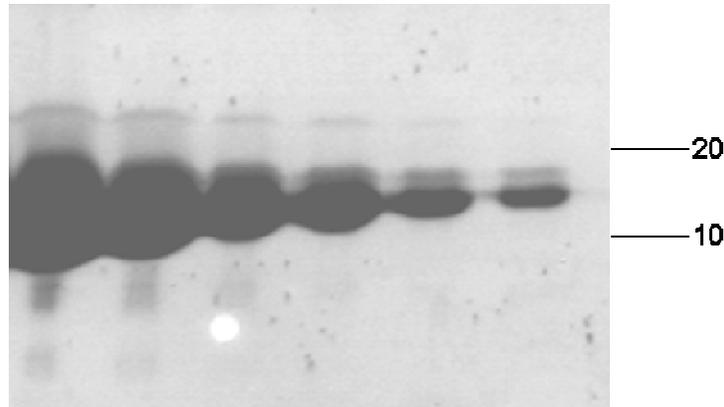


Figure 3.23: affinity purification of His₆-SH2 from the v-Src kinase.

3.7.3 Pull-down assay between cleaved CagA^{C33} and His₆-SH2

To test the interaction between phosphorylated CagA^{C33} (P-CagA^{C33}) and His₆-SH2, a pull-down assay was performed.

The His₆-tag from His₆-CagA^{C33} was first removed by incubating overnight and at 4°C 10 mg of the protein with 500 µg of Tobacco Etch Virus (TEV) protease, in the presence of 0.5 mM EDTA. After the incubation, the EDTA was removed by ultrafiltration and the cleaved protein separated from the uncleaved one using Ni-NTA column. CagA^{C33} was then phosphorylated using the protocol reported in paragraph 3.6.1 to produce P-CagA^{C33}. A Pull-down assay was carried out on a small Spin-Trap column (GE Healthcare) mixing 400 µg of P-CagA^{C33} and 200 µg of previously purified, His₆-SH2. The protocol previously applied for the His₆-CagA^{C33}/CagF pull-down assay was used (paragraph 3.4.1).

Samples from washes and elutions were loaded on a SDS-PAGE gel (Figure 3.24). These showed that unbound P-CagA^{C33} is completely removed after the three washes with buffer A. P-CagA^{C33} is co-eluted when His₆-SH2 is eluted with the buffer containing 500 mM imidazole. As P-CagA^{C33} does not bind to the Ni-NTA resin without the His₆-tag (data not shown), it is clear that *in vitro*, P-CagA^{C33} interacts with His₆-SH2.

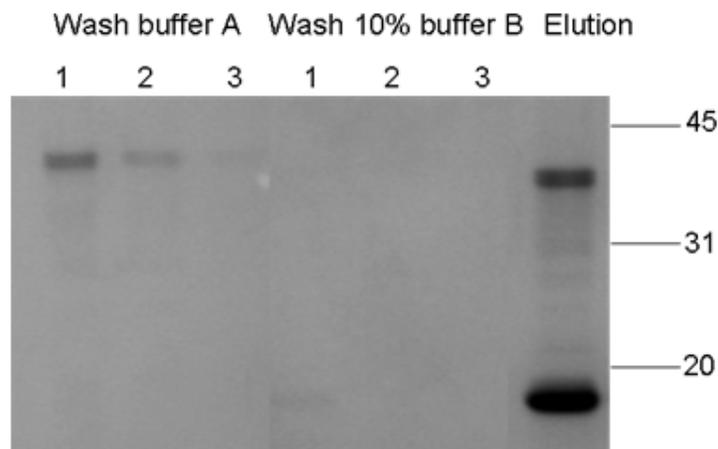


Figure 3.24: pull-down assay between P-CagA^{C33} and His₆-SH2

Three washes (600 μ l each) of buffer A (50 mM Tris pH 8, 100 mM NaCl) and 10% buffer B (50 mM Tris pH 8, 100 mM NaCl, 50 mM imidazole) were performed to wash the unbound CagA^{C33}. Elution was done with 50 mM Tris pH 8, 100 mM NaCl, 500 mM imidazole.

3.8 Preliminary biophysical characterization of the His₆-CagA^{C11}/v- SH2 interaction

Following the successful observation of the interaction between P-CagA^{C33} and His₆-SH2, it was decided to try to further characterize the features of this interaction. However, since His₆-CagA^{C33} forms various types of oligomer (sections 3.2, 3.3), His₆-CagA^{C11} was used because of its higher solubility, stability and homogeneity in gel filtration. The aim was to determine more precisely the contribution of each of the EPIYA motif to SH2 binding. To achieve this, His₆-CagA^{C11} mutants in which the tyrosines of the different EPIYA motifs were mutated to a phenylalanine were designed. To get quantitative information on the binding, Isothermal Titration Calorimetry (ITC) and Surface Plasmon Resonance (SPR) were used.

3.8.1 Construction, expression and purification of His₆-CagA^{C11} mutants

The His₆-CagA^{C11} mutants characterized are: Δ A Δ B (EPIYA-A and EPIYA-B mutated to EPIFA); Δ B Δ C (EPIYA-B and EPIYA-C mutated to EPIFA); Δ A Δ C (EPIYA-A and EPIYA-C mutated to EPIFA); Δ A (EPIYA-A mutated to EPIFA). A schematic representation of these mutants is depicted in Figure 3.25.

The mutants were constructed by PCR using the Qiaquick multi-site directed mutagenesis kit, according to manufacturer (Qiagen). 50 ng of His₆-CagA^{C11} plasmidic DNA were mixed with 100 ng of the following primers: ΔA fw: 5'-acagaaccatt**T**tgctaaagttaataaaaag; ΔB fw: 5'-cctgaagaaccatt**T**tactcaagttgc; ΔC fw: 5'-gccctgaaccatt**T**cgctacgattg (the bold capital letter indicates the mutation introduced). Mutagenesis was performed by PCR following the same protocol reported in paragraph 3.1.1. All mutant plasmids were purified by miniprep and then sequenced to confirm the presence of the desired mutation. All the mutants were expressed and purified according to the protocol established for the wild-type His₆-CagA^{C11}.

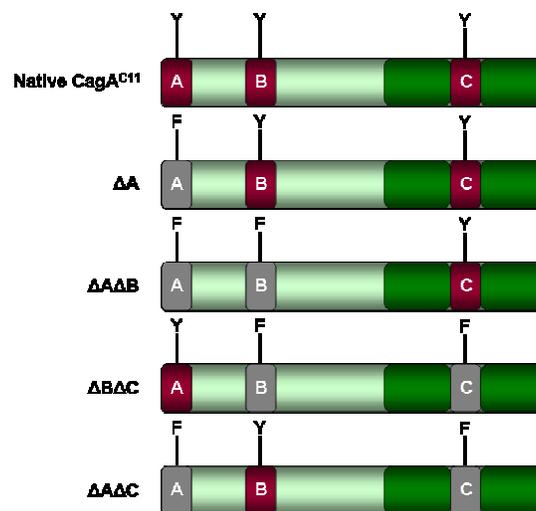


Figure 3.25: overview of CagA^{C11} mutants.

EPIYA motifs (A, B, C) are depicted with brown rectangles. Mutated EPIYA motifs (EPIFA) are depicted with grey rectangles. Dark green rectangles indicates the position of the two multimerization domains. For an overview of CagA^{C11}, see paragraph 3.5

3.8.2 Phosphorylation of His₆-CagA^{C11} mutants with c-Src and c-Abl

Phosphorylation of His₆-CagA^{C11} mutants by c-Src and c-Abl kinases was performed following the protocol used for His₆-CagA^{C33} (paragraph 3.6). The results are shown in Figure 3.26. All mutants are phosphorylated by c-Src, except ΔBΔC. As this mutant bears just the EPIYA-A motif, we can conclude that this motif is not phosphorylated by the human c-Src kinase. In contrast all three mutants are phosphorylated by c-Abl. These results suggest that

the EPIYA-A, EPIYA-B and EPIYA-C motifs can be phosphorylated by the human c-Abl kinase whilst only the EPIYA-B and EPIYA-C motifs can be phosphorylated by the human c-Src kinase.

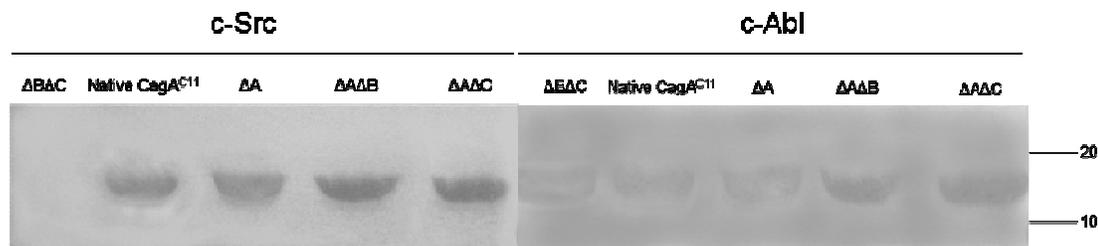


Figure 3.26: western blot of His₆-CagA^{C11} mutants phosphorylated by c-Src (left) and c-Abl (right).

3.9 Characterization of His₆-CagA^{C11}/v-SH2 interaction by ITC

The interaction between His₆-CagA^{C11} and v-SH2 was measured by Isothermal Titration Calorimetry (ITC), a technique that measures the binding between two species by quantifying the thermal variations in the system in response to the association reaction. For a full overview of the technique, see Appendix, A.3.

The strength of the interaction between the mutants ΔBΔC, ΔAΔC, ΔAΔB, wild-type His₆-CagA^{C11} and v-SH2 was measured. His₆-CagA^{C11} and its mutants were previously phosphorylated with the c-Src kinase and then placed in the ITC sample cell at a 10 μM concentration, whilst v-SH2 domain was placed in the microsyringe at a concentration of 150 μM (both measured using the Bradford method). Before the ITC experiments, P-CagA^{C11} constructs and v-SH2 were dialyzed O/N at 4°C against 1 litre of 30 mM Hepes pH 7.5, 100 mM NaCl, using a 3.5 kDa cutoff semi-permeable membrane.

The experiments were carried out at 25°C, with 10 μl injections of v-SH2 into the sample cell under continuous stirring, and 300 seconds between each injection. The ITC profiles for each P-CagA^{C11} construct (wild-type and mutants) are shown in Figure 3.27.

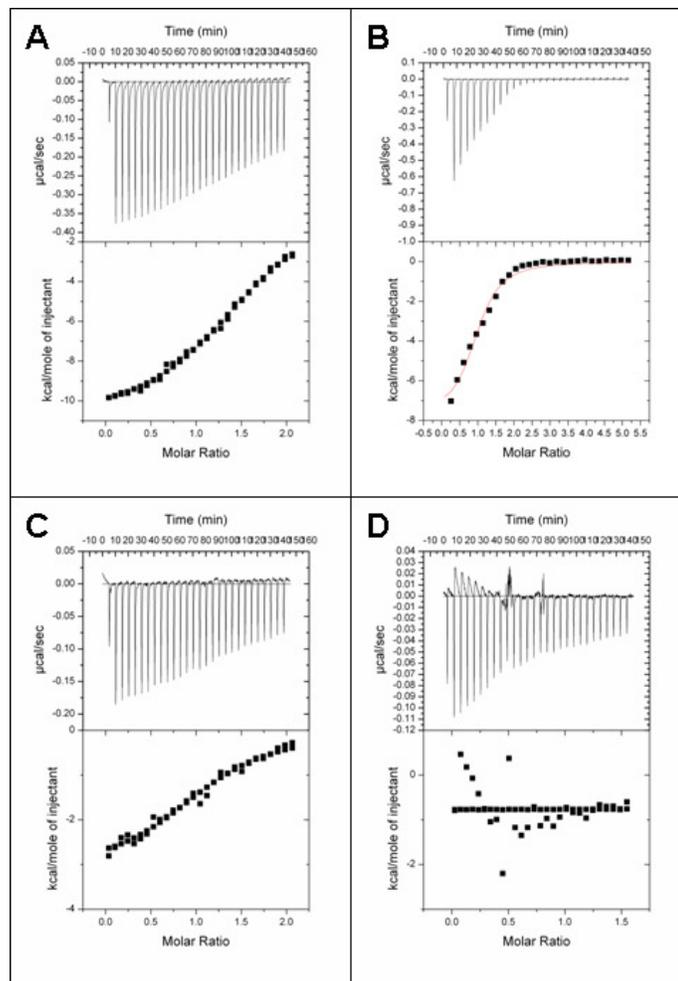


Figure 3.27: ITC profiles of His₆-CagA^{C11} and His₆-CagA^{C11} mutants interactions with His₆-SH2.

The ITC profile is divided in two panels: the upper one reports the microcalories released in the cell at each injection, i.e. at each interaction between the injected His₆-SH2 and His₆-CagA^{C11} construct in the cell. The lower ones report the estimated stoichiometric ratio between the two associated species, according to the kcal/mole of injected v-SH2. (A): His₆-CagA^{C11} wild-type fragment + v-SH2. (B): ΔAΔB + v-SH2. (C): ΔAΔC + v-SH2. (D): non-phosphorylated ΔBΔC + v-SH2.

The experiments show different profiles for each construct tested. Wild-type P-CagA^{C11} is characterized by a weak binding that does not reach saturation, i.e, the complex formation is still observed at the end of the run, suggesting that the protein in the cell is still available for interaction. A similar profile is obtained for the ΔAΔC mutant.

The profile of the mutant ΔAΔB (panel B) is completely different: the ITC shows a higher binding affinity for His₆-SH2 and the interaction reaches saturation after 13 injections. Calculations of the binding constant and the stoichiometry result in values of 1.6 μM and

0.98:1 (very close to a 1:1 ratio, as calculated using the ITC software Origin 70 and considering one-site binding as a binding model) respectively.

The profile of the mutant $\Delta B\Delta C$ (which can be considered as a negative control, as the EPIYA-A motif is not phosphorylated by c-Src) is different from any of the other reported ITC results. The baseline is unstable and there is no correlation between the kcal/mole of v-SH2 injected and the stoichiometry of the complex. Actually, the kcal/mole values seem to keep constant in a value very close to -1 for the entire duration of the experiment. There thus appears to be no specific binding of $\Delta B\Delta C$ to v-SH2. The heat peaks reported in the upper panel can be explained by bulk effect (weak structural recognition of motifs by v-SH2) or aggregation/precipitation of the proteins in the cell. This result is consistent with the fact that the EPIYA-A is not phosphorylated by c-Src and therefore does not bind to v-SH2.

It can be concluded from these experiments that the phosphorylated EPIYA-C motif is a high-affinity binding site for the v-SH2 domain, while the phosphorylated EPIYA-B site is a low-affinity binding site. From the comparison of the ITC profiles reported in panels A and B we can also conclude that the phosphorylated EPIYA-B interferes with the v-SH2 binding to the phosphorylated EPIYA-C.

3.10 Analysis of CagA^{C11}/SH2 interaction by SPR

To investigate further the binding of P-CagA^{C11} to His₆-SH2, a Surface Plasmon resonance (SPR) experiment was carried out. Briefly, SPR analyzes and quantifies the complex formation between two entities according to the change of the measured optical refraction index of the species immobilized on a support chip. For a detailed explanation of the SPR, see Appendix, section A.4.

As for the ITC experiments, we analyzed P-CagA^{C11} and the mutants $\Delta A\Delta B$, $\Delta B\Delta C$ and $\Delta A\Delta C$. In these experiments, the His₆-tag from all CagA^{C11} proteins was cleaved using TEV (section 3.7.3). A Ni-NTA chip (Biacore) was used to immobilize His₆-SH2 from v-Src (v-SH2). The chip was first loaded with 500 μ M of NiCl₂ at a flow rate of 10 μ l/min and the v-SH2 domain was immobilized in flow cell 1 (F₁) at a concentration of 200 nM at the same flow rate. Equilibration of the chip and protein injections were performed with 10 mM Hepes pH 7.5, 150 mM NaCl, 3 mM EDTA and 0.05% Tween20 as a running buffer.

Five injections at different dilutions of P-CagA^{C11} and the related phosphorylated mutants ($\Delta A\Delta B$, $\Delta B\Delta C$ and $\Delta A\Delta C$) were performed at a flow rate of 70 μ l/min in flow cells 1

and 0 (F_1 and F_0). F_0 was chosen as the reference cell, to monitor the aspecific binding of P-CagA^{C11}. At the end of each run, the chip was regenerated with 10 mM Hepes pH 7.5, 150 mM NaCl, 0.36 M EDTA and 0.02% Tween 20 and recharged with NiCl₂ and His₆-SH2. The sensorgrams related to each P-CagA^{C11} construct are reported in Figure 3.28.

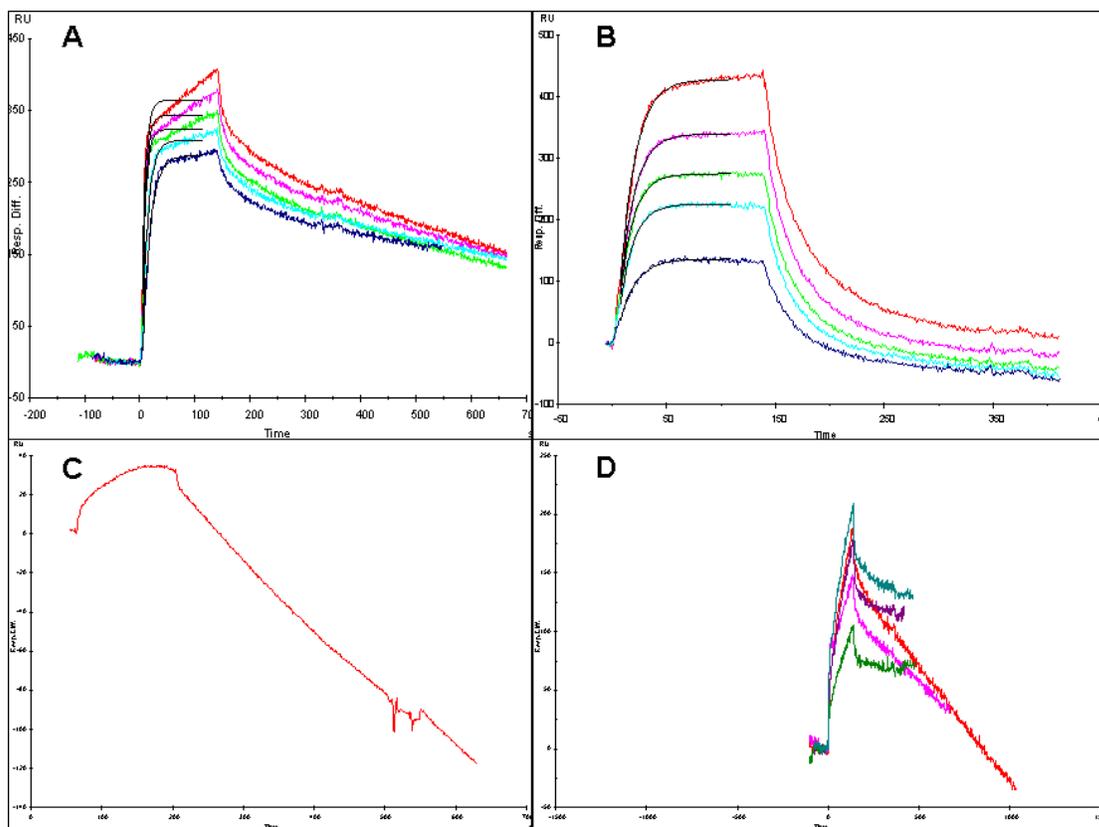


Figure 3.28: SPR sensorgrams of P-CagA^{C11} or P-CagA^{C11} mutants interactions with v-SH2.

The sensorgrams report the difference of signals detected in the sample cell (F_1) and the reference cell (F_0). The curves were created and analyzed using the BIAEvaluation software (Biacore). (A): wild-type P-CagA^{C11} (concentrations: 134 nM, 81 nM, 47 nM, 36 nM, 25 nM). (B): $\Delta A\Delta B$ (concentrations: 144 nM, 112 nM, 94 nM, 62 nM, 40 nM). (C): $\Delta A\Delta C$ (concentration 100 nM, not further analyzed as the complex is stripped away from the chip). (D): $\Delta B\Delta C$ (concentrations: 160 nM, 143 nM, 117 nM, 80 nM, 53 nM).

The binding profile of the P-CagA^{C11} is indicates a fast association to His₆-SH2, but not a fast saturation. P-CagA^{C11} seems to continuously accumulate on the chip, when v-SH2 is immobilized. After injection, the sensorgram for P-CagA^{C11} shows a first phase of short and fast detachment of the complex, then a second phase characterized by a very slow detachment of P-CagA^{C11} from v-SH2. It can thus be inferred that phosphorylated EPIYA-B and EPIYA-

C possibly interfere with each other. From the SPR experiment it seems that the interaction P-CagA^{C11}/ v-SH2 is stronger than $\Delta A\Delta B$ / v-SH2. These results are in contrast with those obtained with ITC. Unfortunately, it was not possible to integrate the data to estimate the k_D (equilibrium constant).

The binding profile of the mutant $\Delta A\Delta B$ is partially consistent with the result obtained with the ITC. SPR shows binding of the $\Delta A\Delta B$ mutant to v-SH2 and saturation. However, the binding seems transient, as $\Delta A\Delta B$ mutant detaches rapidly from the immobilized v-SH2. From the SPR results, it can be concluded that the $\Delta A\Delta B$ / v-SH2 interaction is characterized by both a fast binding and a fast detachment, that makes the complex labile. The calculated k_D from the integration of the five sensorgrams is $52 \text{ nM} \pm 40$.

For the $\Delta A\Delta C$ mutant, the complex proved very unstable on the Ni-NTA chip and the sensorgrams obtained could not be used to analyze the binding.

The mutant $\Delta B\Delta C$ is characterized by a very weak binding, that could be due to aggregation or weak structural recognition of $\Delta B\Delta C$ motifs by v-SH2. These results are in agreement with the ITC observations (section 3.9).

To conclude, the results obtained by both ITC and SPR are only partially consistent with each other but both experiments indicate that the phosphorylated $\Delta A\Delta B$ mutant (bearing just the EPIYA-C) is the one that binds to v-SH2 with the highest affinity. Both phosphorylated EPIYA-B and EPIYA-C motifs bind to v-SH2 and interfere with the binding of each of the motifs separately. However, it is difficult to say whether the interference between these two motifs leads to a weaker or stronger complex between P-CagA^{C11} and v-SH2, as ITC and SPR experiments produced conflicting results.

3.11 Interaction study between His₆-CagA^{C11} and the SH2 domains of the human phosphatase SHP-2

Due to the contrasting results obtained from the ITC and SPR experiments concerning the binding of P-CagA^{C11} constructs to v-SH2, it was decided to use a different SH2 domain, from the human phosphatase SHP-2, to try to characterize the CagA^{C11} / SH2 interaction.

The cDNA of the gene *PTPN11* encoding SHP-2 (Genecopoeia) was purchased. Constructs of the full length SHP-2 protein (residues 1-539, 60 kDa), the N-terminal SH2 domain NSH2 (residues 1-108, 12 kDa) and the tandem SH2 domain TSH2 (residues 1-220, 24 kDa) were designed and subcloned into the pET151/D expression vector. Subsequently,

the characterization of the binding between the phosphorylated $\Delta A\Delta B$ mutant and the three proteins SHP-2, NSH2 and TSH2, was performed.

3.11.1 Cloning, expression and purification of His₆-SHP-2, His₆-NSH2 and His₆-TSH2

DNA fragments encoding for SHP-2, NSH2 and TSH2 were amplified by PCR from the *PTPN11* cDNA (Table 3.1) and cloned into the plasmid pET151/D- TOPO. Plasmids were purified by miniprep from *E. coli* strain 10G Duo 10 ml cultures transformed with the three ligation reactions.

Target	Gene length	Protein Mw	Primers	Purified
SHP-2	1587 bp	60.7 kDa	Fw:CACCATGACATCGCGGAGATGGTTTC Rv:TTAGCGCTGTAGTGTTCATATAATG	+++
NSH2	224 bp	12.1 kDa	Fw:CACCATGACATCGCGGAGATGGTTTC Rv:TTAGGTAGGATCTGCACAGTTCAG	+++
TSH2	672 bp	24.7 kDa	Fw:CACCATGACATCGCGGAGATGGTTTC Rv:TTAACGAGTCGTGTTAAGGGGCTG	+++

Table 3.1: summary of SHP-2, NSH2 and TSH2 gene length, molecular weight and primers used for amplification of the construct and purification of the proteins.

The His₆-tagged proteins were expressed in 2 L cultures of *E. coli* strain BL21 at 37°C. Expression was induced at an O.D. of 0.6 with 1 mM IPTG. Cells were harvested by centrifugation at ~7000 g for 20 minutes, resuspended in 50 mM Hepes pH 7.5, 200 mM NaCl and lysed with a cell disrupter. The soluble fraction was obtained by centrifugation at 40000 g for 20 minutes and loaded on a 5 ml HiTrap Ni-NTA column (GE Healthcare) equilibrated with 50 mM Hepes pH 7.5, 200 mM NaCl, using an Akta Prime (Amersham). His₆-SHP-2 was eluted with 250 mM imidazole, His₆-NSH2 and His₆-TSH2 with 166 mM imidazole (Figure 3.29).

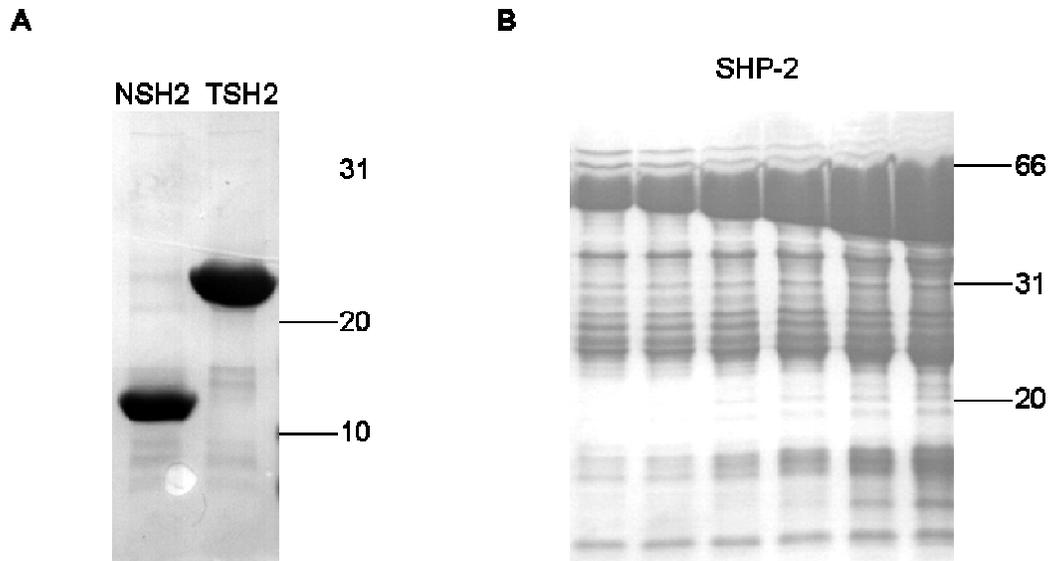


Figure 3.29: affinity purification of SHP-2, TSH2 and SHP-2.

(A): SDS-PAGE gel with purified NSH2 and TSH2. (B) affinity chromatography of the human SHP-2 phosphatase.

3.11.2 Gel-filtration chromatography of $\Delta\Delta\Delta\text{B}/\text{SHP-2}$, $\Delta\Delta\Delta\text{B}/\text{NSH2}$ and $\Delta\Delta\Delta\text{B}/\text{TSH2}$ complexes

The His₆-CagA^{C11} mutant $\Delta\Delta\Delta\text{B}$ was phosphorylated with c-Src kinase following the protocol previously described (Section 3.6). Subsequently, 200 μg of phosphorylated $\Delta\Delta\Delta\text{B}$ were mixed, separately, with 400 μg of His₆-SHP-2, His₆-NSH2, His₆-TSH2. All the mixtures were diluted with 30 mM Hepes pH 7.5, 100 mM NaCl and incubated on ice for 1 hour.

The solutions containing $\Delta\Delta\Delta\text{B}/\text{His}_6\text{-SHP-2}$, $\Delta\Delta\Delta\text{B}/\text{His}_6\text{-NSH2}$ and $\Delta\Delta\Delta\text{B}/\text{His}_6\text{-TSH2}$ were loaded separately in a Superdex 200 5/150 (3 ml bed volume) to perform an analytical gel-filtration. The chromatograms, together with the calibration curve, are shown in Figure 3.30.

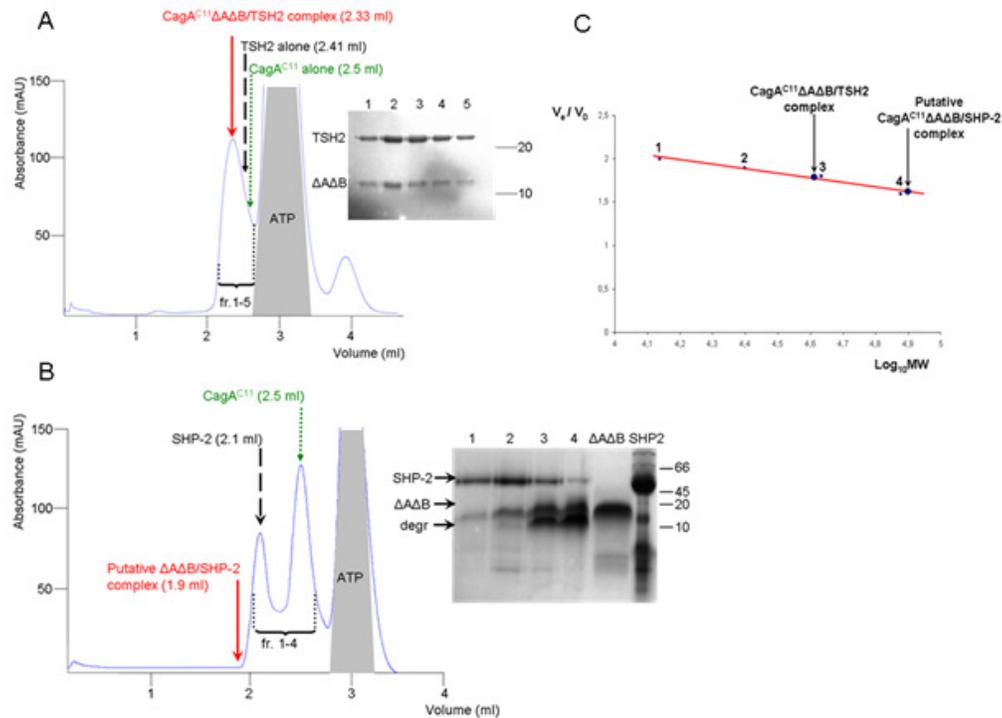


Figure 3.30: gel-filtration chromatography of $\Delta A\Delta B$ in the presence of His₆-SHP-2, His₆-NSH2 and His₆-TSH2.

For clarity, the peak corresponding to residual ATP (from the phosphorylation buffer) has been shadowed in grey in the chromatograms. (A): chromatogram of $\Delta A\Delta B$ incubated with TSH2. The red arrow indicates the peak of the eluted CagA^{C11}/TSH2 complex characterized by a retention volume of 2.33 ml. Fractions of the peak (from 1 to 5) were loaded onto a SDS-PAGE gel (shown on the right), where both proteins are visible. Retention volumes and positions in the chromatogram of TSH2 alone (2.41 ml) and $\Delta A\Delta B$ alone (2.5 ml) are indicated by black and green dashed arrows, respectively. (B): chromatogram of $\Delta A\Delta B$ in the presence of SHP-2. Fractions from 1 to 4 corresponding to the two eluted peaks of SHP-2 and $\Delta A\Delta B$ (retention volumes 2.1 ml and 2.5 ml respectively) were loaded onto a SDS-PAGE gel (shown on the right). The position of the expected $\Delta A\Delta B$ /SHP-2 complex is indicated by a red arrow on the chromatogram together with the expected retention volume. (C): calibration curve of the column Superdex 200 5/150. The Log_{10} of the molecular weight (Log_{10}MW) of Ribonuclease A (1, 13.7 kDa), Chymotrypsinogen A (2, 25,0 kDa), Ovalbumin (3, 43,0 kDa) and albumin (4, 67 kDa) was plotted against the V_e/V_0 , where V_e is the retention volume of the protein and V_0 is the void volume of the column (1.3 ml). The point corresponding to $\Delta A\Delta B$ /His₆-TSH2 (2,33 ml, $V_e/V_0=1.79$) is indicated with an arrow and corresponds to a molecular weight of ~43 kDa. The position of the putative $\Delta A\Delta B$ /SHP-2 complex (not obtained), as calculated from the calibration curve (molecular weight ~75 kDa, $V_e/V_0=1.6$), is also indicated with an arrow.

Gel filtration experiments and SDS-PAGE analysis show that CagA^{C11}ΔAΔB and TSH2 co-eluted as a single peak at a retention volume of 2.33 ml. In different experiments it was found that CagA^{C11}ΔAΔB eluted at a retention volume of 2.5 ml (in agreement with a monomer of 14 kDa) and His₆-TSH2 eluted at a retention volume of 2.41 ml (in agreement with a monomer of 27 kDa). These experiments suggests that ΔAΔB and TSH2 form a complex in gel-filtration.

In contrast, ΔAΔB and SHP-2 eluted as separated peaks, characterized by retention volumes of 2.5 ml and 2.1 ml, respectively. According to the calibration curve, a putative ΔAΔB/SHP-2 complex (~75 kDa) would have eluted at a retention volume of around 1.9 ml. Such a peak is not observed, suggesting that, under the conditions used, ΔAΔB and SHP-2 do not form a stable complex. Similar experiments conducted with ΔAΔB and NSH2 did not show the formation of a complex. A summary of the results of these gel filtration experiments is reported in Table 3.2.

Protein	Elution volume	Protein Mw	Complex formation	complex retention volume
SHP-2	1587 bp	60.7 kDa	NO	1.9 ml (estimated)
NSH2	224 bp	12.1 kDa	NO	2.4 ml (estimated)
TSH2	672 bp	24.7 kDa	YES	2.33 ml (observed)

Table 3.2: summary of the results obtained from the gel-filtration study of ΔAΔB/SHP-2 interaction.

3.11.3 *In vitro* dephosphorylation of CagA by the phosphatase SHP-2

As no stable complex formation between phosphorylated ΔAΔB and SHP-2 was observed in gel filtration experiments it was decided to investigate whether SHP-2 could dephosphorylate CagA and, therefore, prevent the binding of CagA^{C11} to the SH2 domain(s) of SHP-2.

100 μg of both P-CagA^{C33} and P-CagA^{C11} were mixed with the purified His₆-SHP-2 phosphatase in ratio of 1:1. The mixture was incubated on ice for one hour. As a control the P-CagA proteins were separately kept on ice for one hour and overnight were also tested in

order to verify that CagA phosphorylation was stable. Subsequently, the presence of P-CagA^{C33} and P-CagA^{C11} was detected by western blot analyses (Figure 3.31).

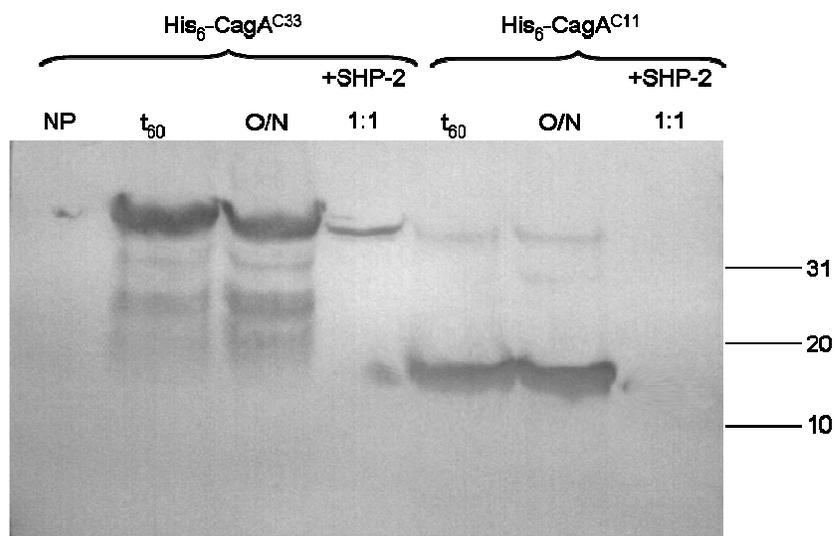


Figure 3.31: Western blot analyses of P-CagAs when mixed with the phosphatase SHP-2.

Both P-CagA^{C33} and P-CagA^{C11} were mixed with the SHP-2 phosphatase and incubated on ice for one hour. Samples of non-phosphorylated proteins (NP), phosphorylated proteins after one hour (t60), phosphorylated proteins after overnight incubation at 4°C and phosphorylated protein mixed with the phosphatase at a 1:1 ratio were loaded on a SDS-PAGE gel, then transferred onto a nitrocellulose membrane for detecting the amount of phosphorylated protein by Western blot with anti-phosphotyrosine antibodies.

After incubation with the phosphatase SHP-2, the intensity of the band due to P-CagA^{C33} noticeably decreased, whilst that for P-CagA^{C11} completely disappeared. As the amount of detected P-CagA decreases when the proteins are incubated with SHP-2, we can conclude that SHP-2 is active towards CagA and is able to dephosphorylate it *in vitro*. This also could explain why a Δ AAB/SHP-2 complex was not co-eluted in gel-filtration, as an unphosphorylated CagA protein can not bind to the SH2 domain(s) of the SHP-2 phosphatase.

3.12 Crystallization trials with CagA^{C33}, CagA^{C11} and complexes

CagA^{C33}, CagA^{C11} and their complexes with SH2 domains were subjected to extensive crystallization tests. Several of these targets were sent to the EMBL High-Throughput

Crystallization Robot (HTXlab), where more than 500 conditions were tested at 20°C and 4°C. For each protein and complex tested, manual screening was also performed. A summary of the crystallization trials is presented in Table 3.3.

No hits were obtained for any of these proteins and protein complexes, except for $\Delta\Delta\Delta\text{B}/\text{v-SH2}$, where small spherulites were detected in the following condition: 100 mM Hepes pH 7.5, 20% Peg8000 (Wizard 1, Emeral Biosciences, condition 21). Unfortunately, when dissolved spherulites were loaded onto a SDS-PAGE gel we observed the presence of one single band, corresponding to the v-SH2 alone. It can therefore be concluded that the spherulites contained only the v-SH2 protein.

Target	Conc. (mg/ml)	Robotic screen	Manual screens	Hits
CagA^{C33}	6	YES	CS1, CS2, NX, PI1, PI2, WZ1, WZ2, MF	NO
P-CagA^{C33}/v-SH2	5	NO	CS1, CS2, NX, PI1, PI2, WZ1, WZ2, MF	NO
CagA^{C11}	20 - 10 - 5	YES	CS1, CS2, NX, WZ1, WZ2	NO
P-CagA^{C11}/v-SH2	5	NO	CS1, CS2, NX, WZ1, WZ2	NO
$\Delta\Delta\Delta\text{B}/\text{v-SH2}$	5	NO	CS1, CS2, NX, WZ1, WZ2	Spherulites of v-SH2
$\Delta\Delta\Delta\text{B}/\text{TSH2}$	5	YES	CS1, CS2, NX, WZ1, WZ2	NO

Table 3.3: crystallization trials of CagA C-terminal constructs and their complexes with SH2 domains.

The proteins, at different concentrations, were subject to manual crystallization screens and/or robotic screen at the EMBL HTXlab. Manual screens used: Hampton's Crystal Screen 1 (CS1), Crystal Screen 2 (CS2), Natrix (NX), Peg-Ion 1 (PI1), Peg-Ion 2 (PI2), MembFac (MF); Emerals Biosciences' Wizard 1 (WZ1), Wizard 2 (WZ2). No hits were retrieved, except for some v-SH2 spherulites using the kit Wizard 1, condition 21.

3.13 Conclusion and discussion

3.13.1 CagA^{C33} oligomerization and architecture

CagA^{C33} is a 33 kDa C-terminal domain of CagA. It contains the two multimerization domains and the three EPIYA tyrosine phosphorylation motifs, which are correlated with CagA function and effects in human gastric cells (see Introduction, paragraphs I.6.4.1, I.6.4.2). The aim was to crystallize this construct and obtain structural information on CagA. Unfortunately, the protein proved to be unstable in all tested crystallization conditions and no crystals could be obtained. Various techniques were therefore used to investigate CagA^{C33} biochemical properties and gain insight into its mode of action. A limited proteolysis study showed that CagA^{C33} is formed by two subdomains of 11 kDa (CagA^{C11}) and 22 kDa (CagA^{C22}) (Figure. 3.16). This is also supported by the finding that CagA^{C11} can be expressed and purified separately.

CagA^{C33} can form oligomers in solution and each of the subdomains seems to play a role in this feature. The results obtained here show that, in solution, CagA^{C33} forms dimers via a disulfide bond between two C1164 residues (from the CagA^{C22}), that can be described as the first dimerization mode. The second, non-covalent dimerization mode involves either one of the two domains composing CagA^{C33}, or both. The CagA^{C11} domain can form dimers, but in solution the equilibrium between monomer and dimer is markedly shifted towards the monomer. For this reason, CagA^{C11} can be involved in the non-covalent dimerization of CagA^{C33}, but it is not sufficient to achieve a complete dimeric status of the protein in solution. CagA^{C22} may possibly participate and stabilize this oligomerization process. However, we conclude from our results that the two dimerization modes (covalent and non-covalent) complement each other to yield a “dimer of dimers”.

Finally, the electron microscopy study shows that the protein forms regular particles, with some in agreement with a tetrameric architecture. However, lower (trimers) and higher order particles (hexamers) can also be formed. Therefore, CagA^{C33} could form dimers/trimer of dimers in solution. A model summarizing CagA^{C33} possible multimerization modes is shown in Figure 3.19.

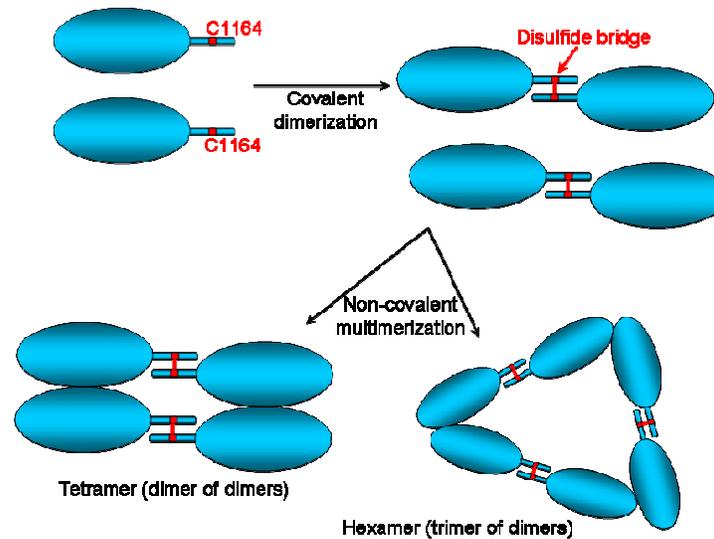


Figure 3.19: putative mechanism of CagA^{C33} multimerization.

Two CagA^{C33} subunits form a covalent dimer through a disulfide bridge mediated by C1164. Subsequently, two or three covalent dimers can interact non-covalently to form respectively a tetramer (dimer of dimers) and/or a hexamer (trimer of dimers).

3.13.2 CagA^{C33} interaction with CagF

Previous studies showed that the CagF protein encoded by the *cag*-PAI is required for CagA translocation into host cells. Since CagF interacts with a C-terminal region (residues 1019-1123) of CagA (Pattis et al., 2007), the interaction between CagA^{C33} and CagF protein was studied here. The experiments described here confirmed that CagA^{C33} and CagF interact. Moreover, limited proteolysis experiments showed that the interaction affects each protein differently: CagA^{C33} is degraded by proteases much faster in the presence of CagF while CagF is degraded slower. These results suggest that CagF interacts with CagA^{C33} and also partially unfolds it.

The conformational change of CagA^{C33} inferred may possibly have a role in CagA translocation across the T4SS apparatus, perhaps by triggering the interaction with other components of the *cag* T4ASS. In T3SSs, the unfolding of the protein substrates by either chaperones and/or the ATPases allows the transport of the substrate across the machinery. An hypothesis that can be made is that CagA^{C33} can be first partially unfolded by the chaperone CagF, in order to interact with the coupling protein homologue Cag β , which was also found essential for CagA translocation (Fischer, *et al.*, 2001). This interaction can then allow the

passage of CagA across the machinery and into the pilus. This hypothesis should be investigated *in vitro* by assessing the formation of a triple partner interaction (CagF-CagA-Cag β) either by co-expression and co-purification of the proteins or by gel-filtration of the separately purified targets.

3.13.3 CagA^{C33} motif requirements for phosphorylation

It is shown here that both CagA^{C33} and CagA^{C11} can be phosphorylated by the human kinases c-Src and c-Abl, confirming that both fragments are functional and that CagA^{C11} is the minimum functional domain that can be phosphorylated by human kinases.

We also show by mutagenesis studies of CagA^{C11} that the EPIYA-A site can be phosphorylated only by the c-Abl, whilst the EPIYA-B and EPIYA-C sites can be phosphorylated by both c-Src and c-Abl kinases. This contrasts with what was reported in literature, where it was shown that all the EPIYA sites can be phosphorylated by both c-Src and c-Abl independently (Tammer et al., 2007).

The results presented here suggest that different EPIYA motifs are not equivalent *in vitro* and that there could be some structural or sequence basis that influences the recognition by c-Src and c-Abl. The sequence of the EPIYA-A motif is immediately preceded by the residues Asparagine-Serine-Threonine (NST), whilst the EPIYA-B and EPIYA-C sites are preceded by the residues Aspartate-Serine-Proline (ASP). It is possible that the sequence ASP might be a consensus sequence for the c-Src kinase, allowing the enzyme to phosphorylate the EPIYA-B and EPIYA-C sites only. Future studies should investigate by mutagenesis the putative role of this sequence.

These results might also have implications for the mode of action of CagA. It was reported in literature (Tsutsumi *et al.*, 2003) that the phosphorylated EPIYA-A and/or EPIYA-B motifs bind to and activate the Csk kinase to inhibit the activity of the c-Src kinase. Moreover, phosphorylated CagA also interacts with the scaffold Crk proteins (Suzuki *et al.*, 2005), but the exact role of the EPIYA motif(s) involved remain uncharacterized. It is possible to hypothesize that the phosphorylated EPIYA-A can specifically bind to Crk and the phosphorylated EPIYA-B can bind to Csk. One way to investigate this would be to assess by SPR whether the mutant $\Delta B\Delta C$, when phosphorylated by c-Abl, can bind to Crk or Csk.

3.13.4 CagA interaction with v-SH2

This study shows that P-CagA^{C33} and P-CagA^{C11} can bind to the v-Src SH2 domain (v-SH2) *in vitro*. ITC and SPR experiments showed that P-CagA^{C11} binds to v-SH2 with a low affinity. In addition, similar experiments performed with CagA^{C11} mutants show that, in fact, the protein bearing only the EPIYA-C motif displays the highest affinity for v-SH2, with a Kd in the nM and μ M ranges from SPR and ITC, respectively. A Protein bearing the EPIYA-B motif was estimated, only by ITC, to have a lower affinity than wild-type CagA^{C11}. The presence of two phosphorylation motifs (as in P-Cag^{C11}) does not increase the affinity of P-CagA for v-SH2 compared with P-CagA with EPIYA-C only. Therefore, these results suggest that v-SH2 has the highest affinity for EPIYA-C, but this affinity is different if EPIYA-B is also phosphorylated. Therefore, as previously suggested by *in vivo* studies, the different EPIYA motif seem to have different affinities for the SH2 domains. Interestingly, CagA containing several copies of EPIYA-C associates with the most severe strains of *H pylori* (see Introduction). Therefore, the EPIYA-C high affinity for SH2 domain we show might explain why this motif was considered as the most important for CagA function *in vivo* (Naito, *et al.*, 2006).

3.13.5 Interaction with the SH2 domains of the human SHP-2 phosphatase

It has been shown here that a P-CagA^{C11} mutant containing only EPIYA-C binds strongly to the TSH2 domain of the SHP-2 phosphatase, whilst it does not bind to the NSH2 alone. This suggests that the CSH2 domain present in the TSH2 is required in the interaction with the phosphorylated CagA but also that both domains might be necessary. Surprisingly, the full length SHP-2 protein (fSHP2), did not form a stable complex with P-CagA^{C11}. However, the results presented here also show that the SHP-2 phosphatase can dephosphorylate CagA *in vitro*. This demonstrates that SHP-2 phosphatase is active and shows for the first time that P-CagA can be one of its substrates. This also can explain why a stable a complex can not be obtained with the fSHP-2, since dephosphorylated CagA can no longer bind to SH2 domain. Although more studies will be required to confirm this, it seems that either CagA can be dephosphorylated by SHP-2 also *in vivo* or that the N-terminal portion of CagA can operate as a regulatory element that inhibits SHP-2 activity.

Chapter 4

Crystal structures of the H. pylori

Tumor Necrosis Factor α -inducing

protein Tip α

Résumé du chapitre 4

Ici, la caractérisation biochimique et structurale d'un nouveau facteur de virulence d'*H. pylori*, Tip α est présentée. Cette protéine est sécrétée par la bactérie dans l'environnement gastrique et internalisée par les cellules de l'épithélium de l'estomac. Deux cystéines, en position 25 et 27 (C25, C27), forment des ponts disulfures entre deux sous unités. Cette forme dimérique de Tip α constitue la forme active de cette protéine. Dans les cellules gastriques, Tip α semble pénétrer dans le noyau, au sein duquel elle induit la transcription de gènes impliquée dans l'expression de plusieurs cytokines et chemokines cellulaires, notamment le « tumor necrosis factor α (TNF- α) ».

Tip α a été cristallisée et sa structure tridimensionnelle résolue en utilisant un fragment de la protéine (Tip α^{N34}), sous deux formes cristallines obtenues à pH 4 (forme 1, résolution 1.8 Å) et à pH 8.5 (forme 2, résolution 1.7 Å). Les phases ont été déterminées par la méthode SAD (Single Anomalous Dispersion) sur des cristaux de Tip α^{N34} contenant des Sélénométhionines (SeMet).

Le monomère de Tip α^{N34} présente un repliement original α - β dans chacune des formes cristallines (1 et 2) dans lequel nous avons identifié trois domaines principaux. Une extension flexible N-terminale (N-terminal Flexible Extension, NFE) incluant une courte hélice 3_{10} , un second domaine similaire à la structure de la dodecine (Dodecin-like Domain, DoD) et un troisième domaine similaire aux « Sterile Alpha Motifs » (SAM), qui sont des domaines retrouvés fréquemment dans les protéines eucaryotes.

Si les structures des monomères des formes 1 et 2 de Tip α^{N34} sont très similaires, leurs structures quaternaires ne le sont pas. La forme 1 (obtenue à pH4) présente un dimère allongé, tête-tête, dans lequel les contacts et surfaces d'interactions sont limités. La forme cristalline 2 de Tip α^{N34} (obtenue à pH 8.5) forme un différent dimère dans lequel deux sous-unités interagissent fortement par trois régions différentes de la protéine.

L'analyse de l'état oligomérique de Tip α^{N34} par gel filtration a montré qu'à pH 4, la protéine est un monomère, alors qu'elle forme un dimère à pH8.5. Ceci suggère que le dimère observé dans la forme 1 est sans doute un artefact de cristallisation.

Enfin, les résultats obtenus montrent que Tip α est sous forme dimérique, en absence des C25 et C27 et que les ponts disulfures sont nécessaires à la formation du dimère à pH4. C'est proposé donc que ces ponts disulfures jouent un rôle important lors de la sécrétion de

Tip α dans l'estomac à pH acide en maintenant deux sous-unités ensemble et facilitant ainsi leur intégration dans la cellule.

Il est enfin montré que la protéine est phosphorylée par la kinase c-Src *in vitro*. Ce résultat ouvre de nouvelles perspectives dans l'étude de Tip α qui sont discutées en fin de chapitre. La caractérisation structurale de Tip α ^{N34} a été publiée dans le journal "*Febs Letters*", et a été choisie comme couverture (Tosi *et al.*, 2009).

4.1 The Tumor-Necrosis-factor α (TNF- α) inducing protein Tip α : a novel carcinogenic factor of *H. pylori*

4.1.1 The discovery of HP0596 as a TNF- α inducing protein

As previously reported (see section I.5), infection with *H. pylori* causes the production of proinflammatory cytokines in the gastric environment. Sustained inflammation and the action of the oncoprotein CagA in gastric cells can lead to gastric adenocarcinoma and MALT lymphoma.

Tumor-Necrosis-factor α (TNF- α) is an important cytokine that regulates immune cells and stimulates acute inflammation (Locksley, *et al.*, 2001). TNF- α was also found to promote tumorigenesis in mice (Suganuma, *et al.*, 2002), therefore recent studies have focused on the identification of factors from *H. pylori* inducing the expression of TNF- α in gastric cells. This led to the discovery of three proteins, belonging to different *H. pylori* strains, that strongly induced TNF- α in murine cells: HP-MP1 (strain SR7791), jhp0543 (strain J99) and HP0596, from the strain 26695 (Suganuma, *et al.*, 2005, Suganuma, *et al.*, 2006). These TNF- α inducing proteins (Tip α) share ~95% sequence identity and, due to their activity, are considered the prototypes of the PRK12303 superfamily of TNF- α inducing proteins. They were also found to be highly immunogenic, as specific antibodies were detected in the sera of *H. pylori* infected patients (Yoshida, *et al.*, 1999, Godlewska, *et al.*, 2008).

4.1.2 Correlation between Tip α and gastric carcinogenesis

Tip α was found to induce a number of chemokines and cytokines which may contribute in sustaining epithelial cell inflammation. In particular, Tip α induces TNF- α expression in mammalian cells (Suganuma, *et al.*, 2005).

TNF- α is an important cytokine, that can promote carcinogenesis (Suganuma, *et al.*, 1999). In addition, higher levels of secreted Tip α were found in the supernatants of cultured *H. pylori* strains from gastric cancer patients, compared to cultured strains from patients affected by gastritis, suggesting an role of Tip α in the onset of gastric cancer. It was recently shown that, in addition to these proinflammatory molecules, Tip α induces the expression of the oncoprotein BCL-2, whilst decreasing the expression of p53 (Cheng, *et al.*, 2008). P53 is a

crucial human DNA-binding protein that promotes the apoptosis of tumoral cells and therefore exerts a protective function against the establishment of cancer (Efeyan & Serrano, 2007). In contrast, BCL-2 can act as an anti-apoptotic agent and de-regulation of its activity has been correlated with the onset of several types of tumors (Frenzel, *et al.*, 2009). The combination of the down-regulation of p53 and the up-regulation of BCL-2 caused by Tip α could thus increase the transformation of damaged gastric cells into cancerous cells. These findings suggest that Tip α is a new carcinogenic factor of *H. pylori*, that could contribute to the onset and progression of gastric cancer through several different mechanisms (Figure 4.1).

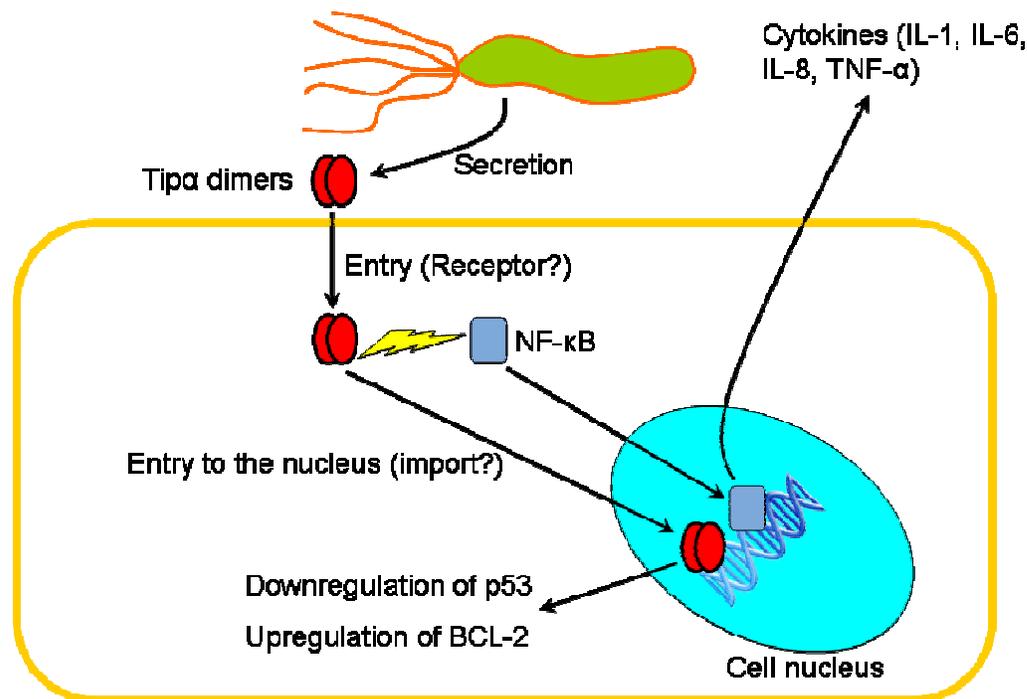


Figure 4.1: schematic view of the proposed mode of action of Tip α and its involvement in gastric carcinogenesis

Tip α is secreted by *H. pylori* as a dimer and enters the gastric cells of the human stomach through an unknown mechanism (it was hypothesized that a receptor might be involved, Suganuma *et al.*, 2008). Once inside the cell, Tip α activates the nuclear factor NF- κ B, that subsequently enters the nucleus and promotes the transcription of cytokine genes involved in sustained inflammation (IL-1, IL-6, IL-8 and TNF- α). Tip α was also reported to enter the nucleus and promote the deregulation of p53 and BCL-2, a process which may be involved in carcinogenesis.

4.1.3 General features of Tip α

The *HP0596* gene product was named “Tumor-Necrosis Factor α -inducing protein”, abbreviated as Tip α . This protein has a molecular weight of 21.8 kDa (192 residues) and presents a Sec-dependent signal peptide at the N-terminus, comprising the first 20 residues (Figure 5.1). Two cysteine residues, C25 and C27, were found to be involved in the homodimerization of the protein by forming two inter-subunits disulfide bridges (Suganuma, *et al.*, 2006). These two residues are conserved in all Tip α proteins (Figure 4.2).

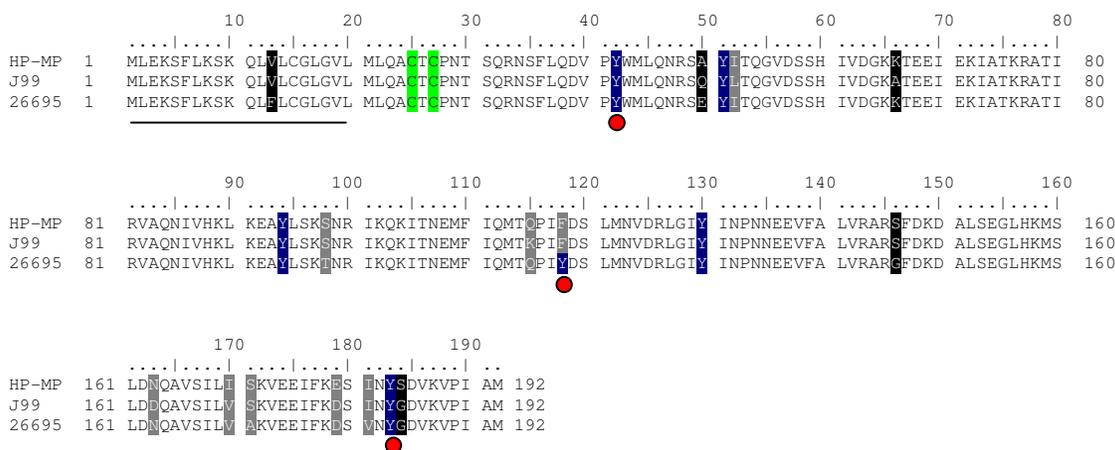


Figure 4.2: sequence comparison of Tip α sequences.

ClustalW alignment of Tip α sequences from *H. pylori* strains SR7791 (HP-MP), J99 (jhp0543) and 26695 (hp0596). Residues not conserved are indicated by a black box and similar residues are shaded in grey. The signal peptide is indicated by a bar. Cysteines 25 and 27 are colored in green and tyrosines in blue. Red dots point to tyrosine mutants generated in this study.

The protein is secreted after cleavage of its N-terminal signal peptide and was found in culture supernatants *in vitro* and also inside human gastric cells (Suganuma, *et al.*, 2005). In which conditions Tip α is secreted by *H. pylori*, and how it enters the gastric cells is unknown.

In gastric cells, Tip α was shown to induce the expression of many chemokines (molecules that mediate the recruitment of immune cells at the inflammation sites) and cytokines (especially IL-1, IL-6, IL-8 and TNF- α) via the activation of the nuclear factor NF- κ B. Tip α was also shown to be localized in the cell nucleus and to have direct DNA-binding activity *in vitro* (Kuzuhara, *et al.*, 2007, Suganuma, *et al.*, 2008). It was also shown that the homodimer is the active form of this protein. A non-dimeric, deletion variant of Tip α , lacking the cysteines C25 and C27 (del-Tip α), is greatly impaired in the induction of cytokine production and DNA binding (Kuzuhara, *et al.*, 2007). The DNA-binding activity of this

protein will be discussed in detail later in this Chapter.

4.2 Cloning, expression and purification of Tip α^{M21} , Tip α^{C25} and Tip α^{C27}

DNA fragments of the gene *hp0596* starting at residues M21 (the first residue after the N-terminal signal peptide), C25 and C27 were amplified by PCR and cloned into the plasmid pET151/D- TOPO (Figure 4.3B). *E. coli* BL21 (DE3) cells were transformed with the three constructs and grown in liquid LB medium supplemented with 100 μ g of ampicillin. Expression of the three constructs was carried out at 30°C for 4 hours under thorough agitation, after inducing the cells at O.D.=0.6 with 1 mM IPTG. Cells were harvested by centrifugation (~7000 g, 20 minutes) and resuspended in 30 mM tris pH 8, 1 M NaCl, prior to lysis by cell disrupter. The soluble fraction was separated from the insoluble pellet by centrifugation (40000 g, 20 minutes) and loaded onto a 5 ml HiTrap Ni-NTA column (GE Healthcare) for affinity chromatography. The column was previously equilibrated with buffer A (30 mM tris pH 8, 200 mM NaCl, 5% glycerol). The proteins were eluted with 250 mM imidazole (final yield: ~50 mg/L each). A summary of the cloning, expression and purification of these constructs is reported in Figure 4.3.

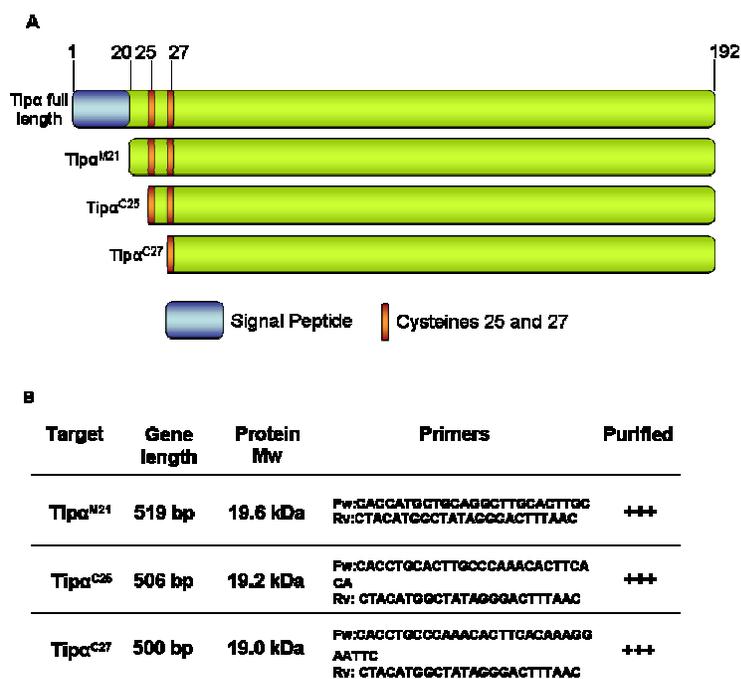


Figure 4.3: summary of the cloning, expression and purification of Tip α^{M21} , Tip α^{C25} and Tip α^{C27} .

(A): overview of the constructs used in this Chapter. (B): summary of the constructs, with reported gene length, molecular weight, primers used and purification yield.

Proteins were subjected to His₆-tag cleavage by TEV protease overnight at 4°C, under dialysis against 2L of 30 mM Tris pH 8, 200 mM NaCl, 5% glycerol to remove imidazole and EDTA. The cleaved proteins were subsequently purified by affinity chromatography and gel filtration chromatography in a preparative grade Sephacryl S100 16/60 column (GE Healthcare, 125 ml bed volume). After gel filtration the proteins were almost 99% pure. As an example, the results of the gel-filtration purification of Tipα^{M21} are shown in Figure 4.4.

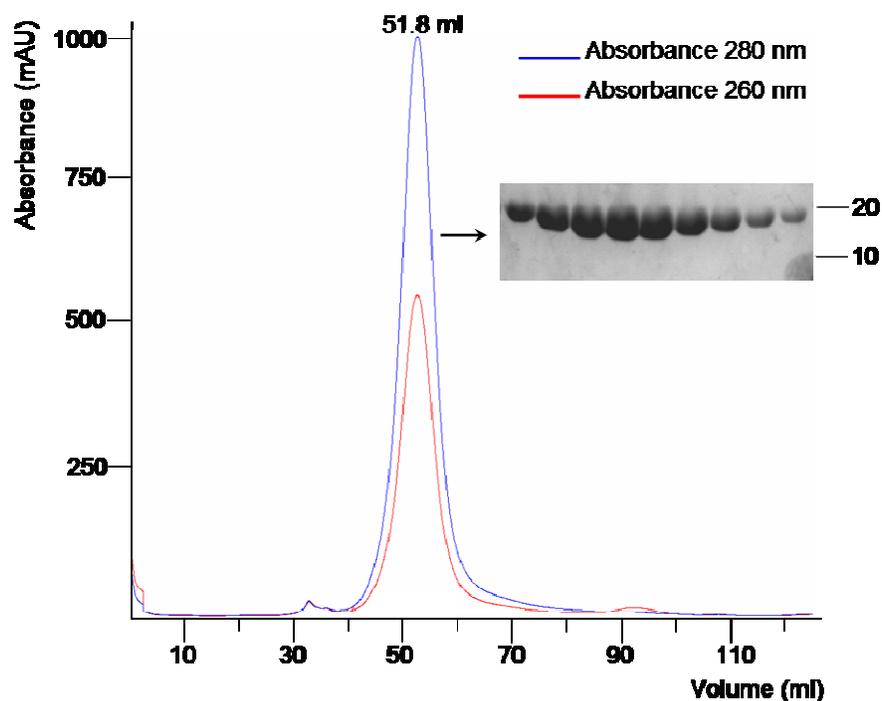


Figure 4.4: gel-filtration chromatography of Tipα^{M21}

Tipα^{M21} eluted at a retention volume of 51.8 ml and fractions were collected. SDS-PAGE analysis shows the purity of the protein in the collected fractions.

4.3 Crystallization of Tipα^{M21}, Tipα^{C25} and Tipα²⁷

Samples of Tipα proteins were sent to the High-throughput crystallization robot (HTXlab, EMBL of Grenoble) at concentrations of 20, 10 and 5 mg/ml. Crystals were obtained with all proteins: Tipα^{M21} crystallized in the presence of 100 mM Bis-Tris pH 5.5, 15% Peg3350, whilst both Tipα^{C25} and Tipα^{C27} crystallized in the presence of 100 mM citric

acid pH 4.0 and 15% Peg5000 MME. All crystals were tested for reproducibility by performing manual screens around the robot crystallization conditions in 24-wells Linbro plates (Hampton Research). Crystallization conditions were prepared manually from stock solutions of buffers and precipitants.

A

Protein	Crystals	Crystal condition	Reproduced	Picture sample
Tipα ^{M21}	YES	100 mM Bis-Tris pH 5.0, 15% Peg3350	YES	
Tipα ^{C25}	YES	100 mM Citrate pH 4.0, 15% Peg5000 MME	15 mM ATP required	
Tipα ^{C27}	YES	100 mM Citrate pH 4.0, 15% Peg5000 MME	15 mM ATP required	

B

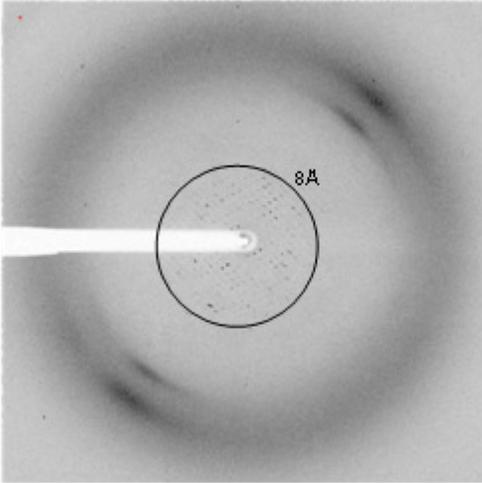


Figure 4.5: summary of the crystallization of Tipα^{M21}, Tipα^{C25} and Tipα^{C27}.

A: crystals obtained with Tipα proteins. The crystallization conditions and the reproducibility of the crystals are summarized, together with a picture of some of the crystals obtained for each construct. B: diffraction image obtained with Tipα^{M21} crystal. The maximum resolution (indicated with a circle) was ~8Å.

Tipα^{M21} crystals were reproduced using 100 mM Bis-Tris (pH 5.0 – 6.0) and 10% - 20% of different PEGs (2000, 3350, 4000). Tipα^{C25} and Tipα^{C27} crystals were difficult to reproduce in the same conditions. It was observed, from manual screens, that such crystals

were reproducible only in the presence of high concentrations of ATP (12-15 mM) in the crystallization drop. Subsequent to the addition of 15 mM ATP, crystals were obtained using 100 mM citrate (pH 3.5-4.5) and 10% - 25% of different PEGs (1500, 2000, 3350, 4000, 5000, 6000, 8000). A summary of the crystallization tests of Tip α ^{M21}, Tip α ^{C25}, Tip α ^{C27} is reported in Figure 4.5.

Crystals of Tip α ^{M21} were frozen in a 0.1 mm loop in the presence of 15% glycerol as a cryoprotectant and subsequently tested for diffraction at ID23-2 of the ESRF. Crystals of Tip α ^{C25} and Tip α ^{C27} were fished, frozen and tested under the same conditions of Tip α ^{M21}, at the beamline ID29 of the ESRF.

The two 0.5° oscillation images (one of the images used is shown in Figure 4.5) taken at phi angles of 0° and 90° from Tip α ^{M21} crystals were analyzed using the program MOSFLM (Powell, 1999) in an attempt to derive unit cell dimensions and Bravais lattice type. Unfortunately, the low resolution (~8 Å) diffraction patterns observed did not allow the successful determination of unit cell dimensions and a complete dataset was not collected. The analyses of the diffraction images of crystals of both Tip α ^{C25} and Tip α ^{C27} also failed to allow unit cell determination (data not shown).

4.4 Identification of a stable Tip α fragment starting at N34 (Tip α ^{N34})

To try to increase the quality of the crystals obtained, a limited proteolysis of Tip α ^{M21} with Trypsin was performed, to remove putative flexible regions of the protein that could impair crystal packing.

Purified Tip α ^{M21} (100 µg) at a concentration of 1mg/ml were incubated on ice with 100 ng of Trypsin (2 µl of a 50 ng/µl stock solution) for 4 hours. Samples were taken every hour and loaded onto a SDS-PAGE gel. The results are shown in Figure 4.6. Tip α ^{M21} is cleaved to produce a defined, smaller fragment (~18 kDa) that appears to be stable. This fragment was blotted onto a PVDF membrane and sent to the Institut de Biologie Structurale (IBS) Mass spectrometry platform for protein N-terminal sequencing. The results showed that this fragment starts at Asparagine 34 (N34) and was therefore named Tip α ^{N34}.

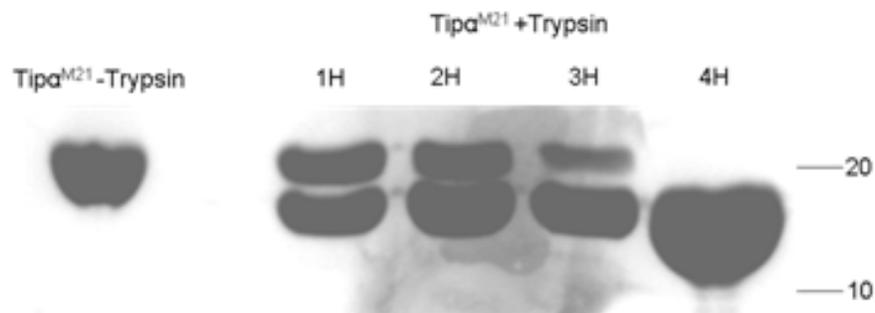


Figure 4.6: Tip α^{M21} limited proteolysis

Tip α^{M21} was digested for 4 hours on ice with trypsin protease, 1000 fold diluted. Samples were taken every hour and loaded onto a SDS-PAGE gel to monitor the digestion.

Due to the absence of other cleavage products in our limited proteolysis experiment, it can be concluded that R33 is the only trypsin cleavage site in Tip α^{M21} .

4.5 Crystallization of Tip α^{N34}

Tip α^{N34} yielded by limited proteolysis was used to perform manual crystallization screens, due to by its high stability and the fact that 1000-fold diluted trypsin is a negligible quantity in a crystallization drop.

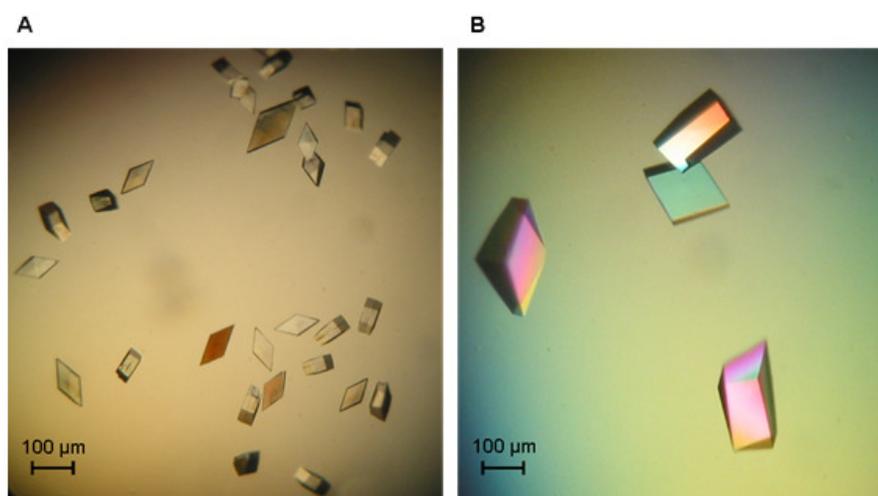


Figure 4.7: crystals of Tip α^{N34} .

(A): crystals grown in 100 mM citric acid pH 4, 14% Peg7000, 15 mM ATP (Form I). (B): crystals grown in 100 mM tris pH 8.5, 24% Peg4000, 50 mM MgCl₂ (Form II).

Initial hits were obtained with the crystallization conditions of Tip α^{C25} and Tip α^{C27} (100 mM citric acid pH 4, 10-25% of different Pegs, 15 mM ATP). Tip α^{N34} crystallized in two different forms: Form I (100 mM citric acid pH 4, 10-25% of different Pegs, 15 mM ATP) and Form II (100 mM Tris pH 8.5, 24% Peg4000, 200 mM MgCl₂). Optimization of the latter condition, by screening different concentrations of PEG 4000 and MgCl₂, led to the formation of crystals with maximum dimension of 200-300 μ m in around 7 days. Pictures of the crystals are shown in Figure 4.7.

4.6 Structure determination of Tip α^{N34} Form I

4.6.1 Expression and purification of the Seleno-Methionine Tip α^{C27} (SeTip α)

A 10 ml of LB preculture of BL21 DE3 cells harbouring the pET151/D-Tip α^{C27} plasmid were inoculated into M9 medium and incubated at 37°C. At O.D 0.6, the following amino acids were added as powder into the culture to inhibit the methionine biosynthesis: 50 mg of Threonine, Lysine, Phenilalanine, Leucine and 25 mg of Isoleucine and Valine. Subsequently, 30 mg of Seleno-methionine was added for incorporation in expressed proteins. After 15 minutes of incubation at 37°C, the expression was induced with 1 mM IPTG and by incubating the bacteria overnight at 20°C. SeTip α was purified following the same protocole reported in the section 4.2 except that the NaCl concentration was raised to 500 mM during the gel filtration step. SeTip α was subjected to limited proteolysis, following the same protocol used for the native Tip α^{M21} . SeTip α^{N34} crystallized in form I (see section 4.5).

4.6.2 Data collection, structure determination and refinement

Crystal form I was used to determined the structure of Tip α^{N34} . Native and Se-Met crystals diffracted to 1.8 Å and 2.0 Å, respectively and belonged to the space group P2₁ with almost identical unit cell parameters and 2 molecules per asymmetric unit. For structure determination, a highly redundant dataset was collected at the peak of the Selenium absorption K-edge (~0.9795 Å) on SeMet Tip α^{N34} crystals (Table 4.1). Diffraction images were integrated with MOSFLM (Powell, 1999), the resulting intensities scaled and merged with SCALA and structure factor amplitudes and anomalous differences derived using TRUNCATE (Evans, 2006). SHELXD (Schneider & Sheldrick, 2002) was used to find

selenium sites using the anomalous differences observed. The positions of 10 sites out of a possible 12 were found (the amino acid sequence of Tip α ^{N34} contains 6 methionines). These, coupled with the observed structure factors and anomalous differences, were used in SHARP (de La Fortelle & Bricogne, 1997, Dauter, *et al.*, 2002) to solve the structure of Tip α ^{N34} by the single anomalous dispersion method (SAD, see appendix, paragraph A.5.4.3) followed by automatic density modification. This procedure yielded an excellent electron density map in which 240 residues from two chains were automatically built and docked using ARP/WARP (Perrakis, *et al.*, 1999) (Figure 4.8) The model was completed manually using COOT (Emsley & Cowtan, 2004) and then submitted to rigid body and restrained refinement against the structure factor data from native Tip α ^{N34} crystals with REFMAC5 (Murshudov, *et al.*, 1999), alternated with manual rebuilding with COOT.

	SeMet-Tip α ^{N34} Form I	Tip α ^{N34} Form I	Tip α ^{N34} Form II
Data collection statistics			
Unit Cell Parameters (Å)	a= 31.64 b=102.21 c=47.63 β = 94.40	a=31.36 b=101.99 c=47.26 β = 94.38	a=48.88 b=67.26 c=97.47
Beamline	ID14-4	ID14-4	ID29
Spacegroup	P2 ₁	P2 ₁	P2 ₁ 2 ₁ 2 ₁
Wavelength (Å)	0.979	0.939	0.991
Resolution limit (Å)	43.0-2.0	50.9-1.6	27.7-1.7
Total observations	151032 (22287)	93609 (10799)	128160 (
Unique reflections	20350 (2969)	22966 (3030)	36864 (6168)
Completeness (%)	99.9 (99.9)	98.4 (90.4)	99.3 (99.6)
Multiplicity	7.4 (7.6)	4.1 (3.6)	3.6 (3.6)
Rmerge (%)	7.4 (40.3)	6.0 (32.5)	4.0 (40.7)
Wilson B-factor (Å ²)	24.4	26.2	26.4
Phasing statistics			
Se sites	10		
FOM (centric/acentric)	0.13/0.41		
Anomalous phasing power	1.74		
Refinement statistics			
R _{work}		19.6 (27.7)	13.2 (29.3)
R _{free}		23.1 (37.7)	21.9 (34.7)
RMS bonds		0.015	0.017
RMS angles		1.482°	1.51°

Table 4.1: Data collection, phasing and refinement statistics.

This table summarizes the main statistics of Tip α ^{N34} crystallographic data collection, phasing and structure refinement. Values in brackets refer to the outer resolution shell (maximum resolution).

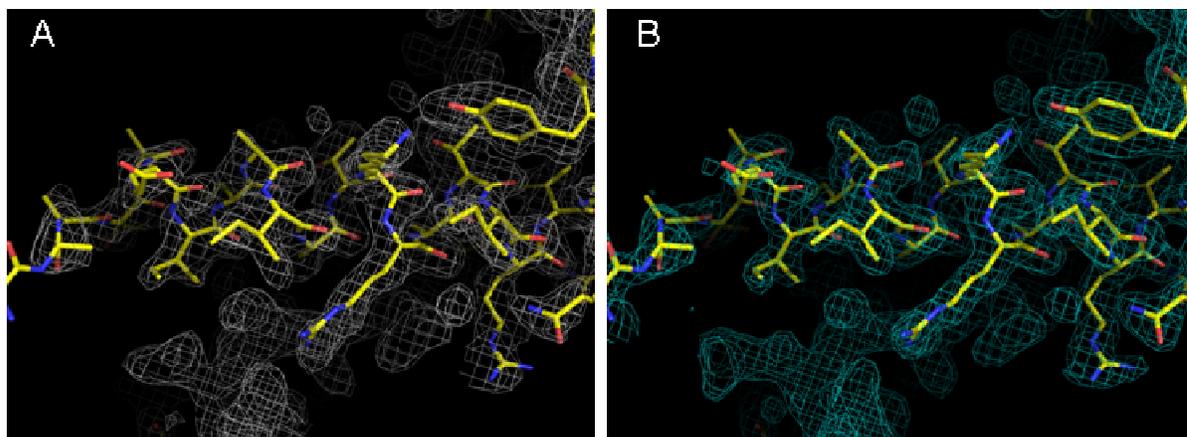


Figure 4.8: detailed views of Tip α^{N34} experimental and refined electron density maps.

(A): An example of the electron density map obtained after density modification of SHARP experimental phases at 2.0 Å. (B): Final 2Fo-Fc electron density map obtained after refinement at 1.9 Å. The final refined Tip α model is displayed as ball and stick representation with carbon colored in yellow, oxygen in red and nitrogen in blue. The electron density maps are contoured at 1 x r.m.s. level..

Crystal form II diffracted at 1.7Å resolution and belonged to the P2₁2₁2₁ space group and also contained two molecules per asymmetric unit. Crystal form II structure was solved by molecular replacement using PHASER (McCoy, *et al.*, 2007) using the structure of the monomer crystal form I as a search template. COOT was used for model rebuilding and the structure was refined using PHENIX (Adams, *et al.*, 2002). All datasets were integrated with MOSFLM (Leslie, 1992) and the data scaled and merged with SCALA (Collaborative Computational Project-4, 1994) (Table 4.1).

4.6.3. Structure of Tip α^{N34}

Tip α^{N34} adopts a novel fold, characterized by an elongated α - β sandwich completed by a N-terminal short 3₁₀ helix (Figure 4.9 panels A and B). The protein is composed by three different domains (Figure 4.8C): a N-terminal flexible extension (NFE, residues N34 to I52); a Dodecin-like domain (DoD), constituted by the three β -sheets and the N-terminal half of the long α 1 helix and similar to the protein Dodecin (accession code 1MOG); a helical bundle (HB), similar to three types of helical domains such as the N-terminal domain of the Nab2 protein (accession code 2V75), the Sterile Alpha Motif (SAM) domain, and the translation initiation factor 2-alpha (accession code 1YZ6).

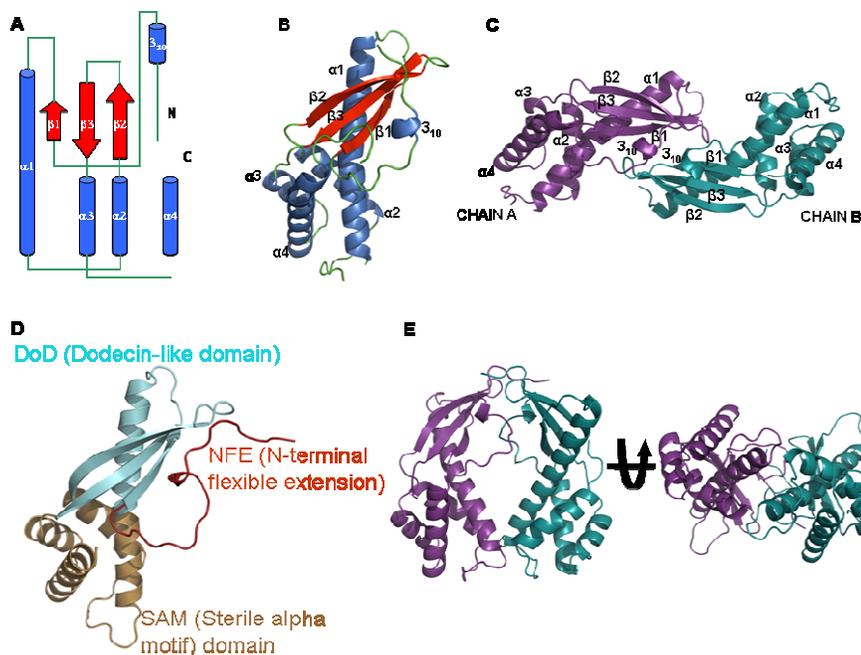


Figure 4.9: cartoon representation of Tip α^{N34} structures.

(A): Topology diagram of Tip α^{N34} . (B): Front view of Tip α^{N34} monomer (form I). α -helices are colored in blue, β -sheets in red, loops in green. Secondary structure elements are numbered according to the topology diagram. (C): dimer of Tip α^{N34} in form I crystals. (D) domain organization of Tip α^{N34} : The N-terminal flexible extension (NFE) is colored in green, the Dodecin-like domain (DoD) is colored in cyan, the Sterile Alpha Motif (SAM) domain is colored in brown. (E): front and bottom view of Tip α^{N34} dimer (Form II). Chain A is colored in violet, Chain B in aqua.

In crystal form 1, Tip α^{N34} forms a head-to head dimer related by a pseudo two-fold axis. Structure at form II was also found as a symmetrical dimer in the crystal, with a pseudo-twofold axis going from the N-terminus of the protein to the bottom (Figure 4.9, panel D). Three areas of each monomer were found to reciprocally interact by mainly hydrogen bonds. In total, 18 interactions were found to be potentially linking the two monomers together in solution.

To investigate this, we performed analytical gel-filtration experiments of Tip α^{M21} and Tip α^{N34} at pH 4.0 and 8.5. The results show that Tip α^{M21} elutes as a dimer at both pHs, whilst Tip α^{N34} elutes as a dimer at pH 8.5 and as a monomer at pH 4.0. The SDS-PAGE gel of the eluted proteins at both pH 4.0 and 8.5 shows that, in the absence of reducing agents, Tip α^{M21} forms dimers of ~38 kDa, whilst Tip α^{N34} migrates only as a monomer of ~18 kDa.

This indicates that C25 and C27 are dispensable for the dimer formation at physiological pH, whilst they are required at pH 4.0. This suggests a role of C25 and C27 as

elements to keep the two monomers together after Tip α secretion in the acidic gastric environment.

All these results were published in *FEBS letters* in May 2009 (Tosi, *et al.*, 2009). The paper is included in this manuscript and gives a fuller account of the structure of Tip α^{N34} .

4.6.4 Charge distribution on the surface of Tip α^{N34}

In the literature (Kuzuhara *et al.*, 2007b) the dimeric Tip α was reputed to bind DNA *in vitro*, the charge distribution on the surface of Tip α^{N34} dimer crystal structure obtained at pH 8.5 (form II) was therefore examined in order to identify regions where DNA might bind to Tip α . The analysis was performed by using the Apbs plugin for Pymol (Baker, *et al.*, 2001).

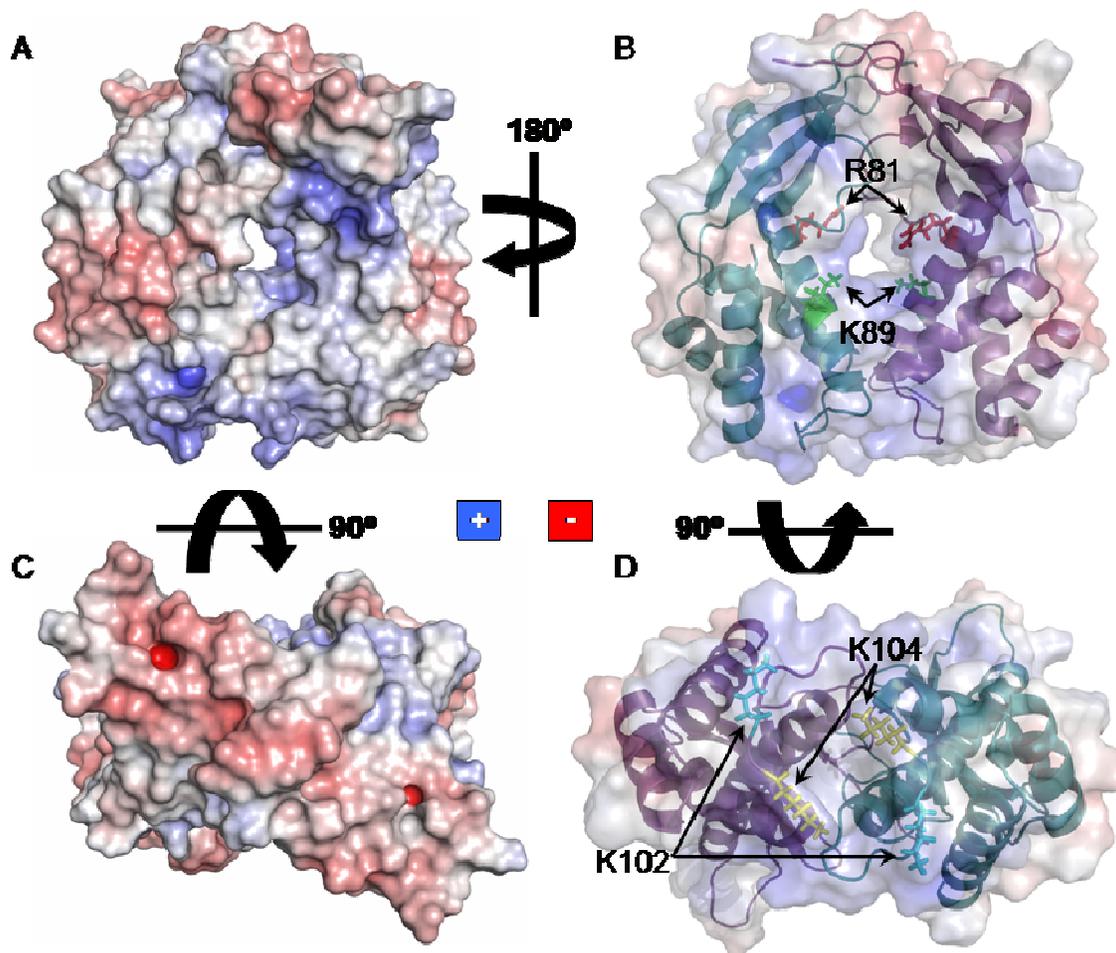


Figure 4.10: surface charge distribution on Tip α^{N34} dimer.

Positively and negatively charged regions are colored in blue and red, respectively. Neutral (hydrophobic) regions are colored in white. (A): front view. (B): back view. The potential surface is shown in transparency. R81 and K89 are shown in stick representation and their positions indicated

with arrows. (C): top view of the dimer. (D): view of the bottom of the dimer. The potential surface is shown in transparency, K102 and K104 are shown in stick representation and their positions indicated with arrows.

As shown in Figure 4.10 the surface of Tip α ^{N34} dimers is highly polar. In particular, the back and bottom surfaces of the dimer are characterized by large positively charged regions that could be involved in the binding of the negatively charged sugar-phosphate backbone of DNA.

4.7 Assay of the DNA-binding activity of Tip α

To investigate the DNA-binding activity of Tip α ^{M21}, band-shift assays were performed with poly-dT₁₅ and poly-poly-(dGdC)₁₄. According to the results reported in literature, Tip α can bind single-stranded and double-stranded oligonucleotides if they are at least 9 bp long (Kuzuhara, *et al.*, 2007).

Tip α ^{M21} at different concentrations (2 μ M, 5 μ M, 10 μ M, 20 μ M) was mixed with a 10 μ M concentration of poly-dT₁₅ and poly-dGdC₁₄. The poly-dGdC₁₄ was previously denatured at 95°C for 10 minutes, then slowly annealed at room temperature for 1 hour. Protein/DNA mixtures were incubated at room temperature for 1 hour, mixed with Loading Buffer (30 mM tris pH 8, 50 mM NaCl) and then loaded onto a 1% agarose gel in Tris/Acetate/EDTA (TAE) buffer pH 8.0 supplemented with 10000-fold diluted SYBR Safe (Invitrogen). The run was performed at 50V for 30 minutes, then the gel was irradiated with UV rays at 260 nm to visualize the DNA bands. If the protein bound to DNA, the band corresponding to the DNA oligonucleotide will shift. As a positive control we used, the DNA binding protein UvrD from *Deinococcus radiodurans*. Bovine Serum Albumin (BSA) was used as a negative control. The results are reported in Figure 4.11..

The results obtained show no DNA mobility shift using poly-dT₁₅. The same experiment was repeated using poly-(dGdC)₁₄ but, again, no DNA binding was observed (data not shown). The same experimental protocol was repeated with shorter oligonucleotides (poly-dT₁₀, poly-dT₁₂, poly-(dGdC)₁₀, poly-(dGdC)₁₂) and in the presence of 1 mM MgCl₂, but no DNA binding was observed. Under the experimental conditions used, Tip α ^{M21} does not bind to any of the oligonucleotide tested.

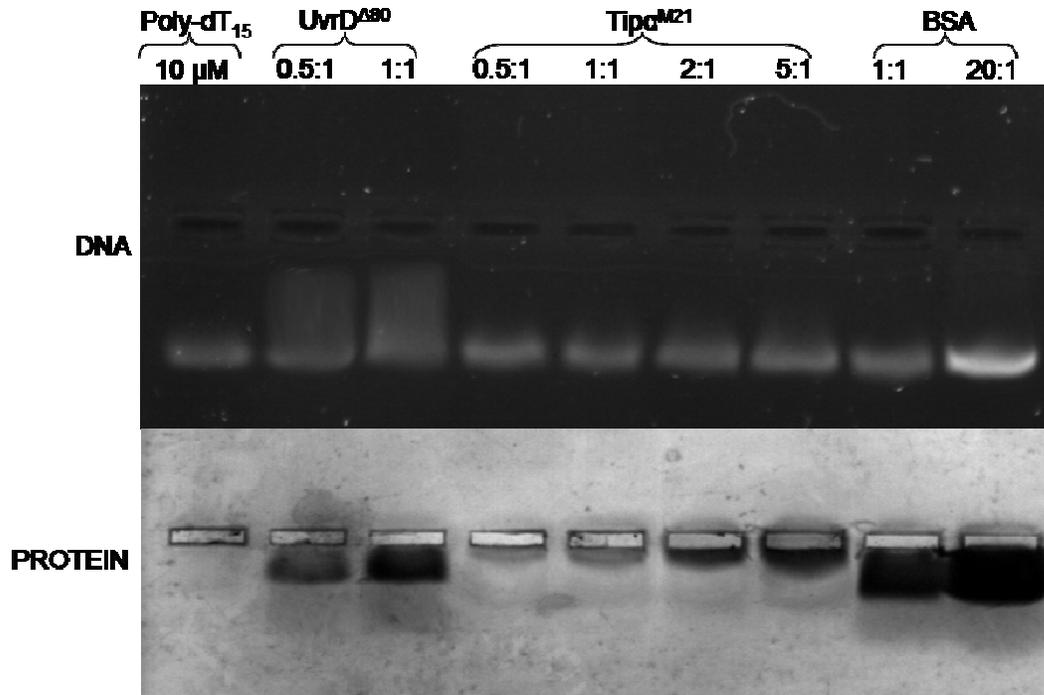


Figure 4.11: DNA mobility shift assay of Tip α^{M21} .

Tip α^{M21} was mixed with poly-dT₁₅ at protein:DNA ratios of 1:2, 1:1, 2:1, 5:1. As a positive control, we used the DNA-repairing enzyme UvrD^{Δ80} from the radioresistant bacterium *Deinococcus radiodurans* (kindly provided by Dr. Joanna Timmins, ESRF Structural Biology Group), at a protein:DNA concentration of 0.5:1 and 1:1, whilst as a negative control, Bovine Serum Albumin (BSA) was used at a protein:DNA ratio of 1:1 and 20:1. Upper panel: DNA staining with Sybr Gold. Lower panel: protein staining with Coomassie Blue.

4.8 *In vitro* phosphorylation of Tip α^{M21} and Tip α^{N34}

Analysis of the crystal structure of Tip α^{N34} showed that six tyrosine residues (Y42, Y51, Y94, Y118, Y130, Y183) are relatively exposed to the solvent (Figure 4.12). One of these residues, Y118, is located in a motif (QPIYD) whose sequence resembles the EPIYA tyrosine-phosphorylation motif of CagA (paragraph I.5.4.1). The *in vitro* phosphorylation of Tip α^{M21} and Tip α^{N34} was therefore tested.

To obtain homogeneous Tip α^{N34} protein and remove the proteolysis step, a new construct allowing the production of Tip α^{N34} was designed. Tip α^{N34} was amplified by PCR using the following primers: N34 fw: 5'-caccaattcttcttgcagaatgtgcc; N34 rv: ctacatggctatagggactttaac. The construct was subsequently cloned into the plasmid pET151/D-

TOPO. Clones were amplified and purified by miniprep and sent to Genecore (EMBL of Heidelberg, Germany) for sequencing. Protein expression and purification was performed according to the protocol established for Tip α^{M21} , Tip α^{C25} and Tip α^{C27} (section 4.2).

The *in vitro* phosphorylation of Tip α^{M21} and Tip α^{N34} with the human recombinant kinases c-Src and c-Abl was tested, applying the same protocol already used for the *in vitro* His₆-CagA^{C33} phosphorylation (paragraph 3.6). The results are shown in Figure 4.12.

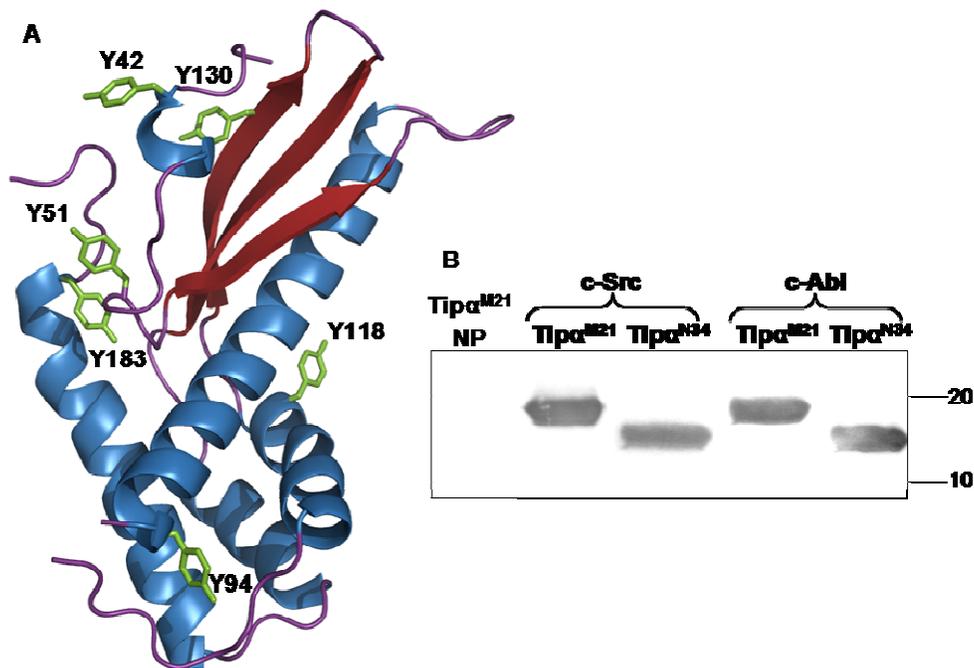


Figure 4.12: *in vitro* phosphorylation of Tip α^{M21} and Tip α^{N34} with the kinases c-Src and c-Abl

(A): cartoon representation of Tip α^{N34} monomer. The exposed tyrosine residues (Y42, Y51, Y94, Y118, Y130, Y183) are depicted in green. (B): western blot of non-phosphorylated Tip α^{M21} (Tip α^{M21} NP) and phosphorylated Tip α^{M21} and Tip α^{N34} . Both c-Src and c-Abl kinases were used.

As shown in Figure 4.12, both Tip α^{M21} and Tip α^{N34} can be tyrosine-phosphorylated *in vitro* by both c-Src and c-Abl. Therefore, at least one tyrosine residue of Tip α can be phosphorylated by both human c-Src and c-Abl kinases *in vitro*.

4.8.1 Preliminary study of the Tip α ^{M21} phosphorylation motifs

To gain further insight into which of the tyrosine residues of Tip α might be phosphorylated a search for putative phosphorylation motifs was carried out using a prediction server of kinase phosphorylation sites (<http://www.hprd.org>, (Prasad, *et al.*, 2009)). This server predicted three phosphorylation sites for c-Src kinase in Tip α sequence at residues Y42, Y118 and Y183 (Table 4.2). Three Tip α ^{M21} mutants, Tip α ^{M21}Y42A, Tip α ^{M21}Y118A and Tip α ^{M21}Y183A were therefore generated.

Position in query protein	Sequence in query protein	Corresponding motif described in the literature (phosphorylated residues in bold)	Features of motif described in the literature
39 - 42	DVPY	[E/D]XX pY	ALK kinase
40 - 45	VPYWML	[I/V/L/S]X pYXX [L/I]	Src family kinase
42 - 45	YWML	pYXX [L/I/V]	JAK2 kinase
49 - 52	SEYI	X[E/D] pYX	EGFR kinase
49 - 52	SEYI	X[E/D] pY [I/L/V]	EGFR kinase
118 - 119	YD	pY [A/G/S/T/E/D]	Src kinase
118 - 121	YDSL	pYXX [L/I/V]	JAK2 kinase
118 - 121	YDSL	pY [E/D]X[I/L/V/M]	ALK kinase
183 - 184	YG	pY [A/G/S/T/E/D]	Src kinase
183 - 186	YGDV	pYXX [L/I/V]	JAK2 kinase

Table 4.2: Tyrosine phosphorylation sites by various kinases predicted in of Tip α sequence

The mutagenesis was performed by the same protocol described for the mutagenesis of His₆-CagA^{C33} (paragraph 3.1.1), using the following primers: Y42A: 5'-cttgcaagatgtgcct**G**Cttggatgtgcaaaatc; Y118A: 5'-tgacacagcccatt**G**Ctgacagcttgatgaatg; Y183A: 5'-aaagattctgcaat**G**Ccggagatgttaaagtcc (mutated bases are indicated in bold and uppercase). Attempts to generate a triple mutant of Tip α ^{M21} (Y42A, Y118A, Y183A) failed despite several attempts.

All the mutants were confirmed by sequencing. Subsequently, the constructs were expressed in *E. coli* strains BL21 (DE3) and purified by affinity chromatography following the protocol already used to express and purify Tip α ^{M21} (section 4.2). The purified proteins were then used in a phosphorylation assay with c-Src and the presence of phosphorylated tyrosines detected by western blot. The results are shown in Figure 4.13.

The results show that Tip α ^{M21}Y42A, Tip α ^{M21}Y118A and Tip α ^{M21}Y183A are all phosphorylated by c-Src. Therefore, none of these Tip α tyrosine sites is a unique phosphorylation site for c-Src. Therefore it cannot be ruled out that i) more than one these

three sites is phosphorylated and/or ii) that additional phosphorylation sites not identified by the server exist in the Tip α sequence. An evaluation of the accessibility of the remaining tyrosines of Tip α suggests that a possible candidate for phosphorylation is Y51, as it is located in a solvent-accessible loop and is exposed on the surface of the protein, in contact with the solvent. More complete mutagenesis experiments will be necessary to determine the number and position(s) of the phosphorylation sites in Tip α .

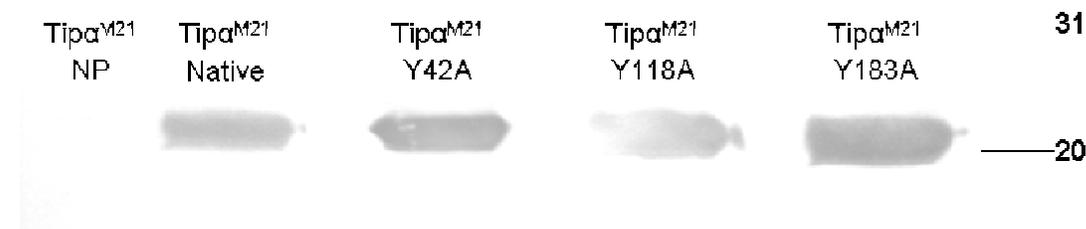


Figure 4.13: Phosphorylation of Tip α ^{M21} mutants.

Non-phosphorylated (NP) and phosphorylated Tip α ^{M21} native and mutated proteins were loaded onto a SDS-PAGE gel and subsequently blotted onto a nitrocellulose membrane by electrotransfer. Detection of phosphorylated tyrosines was performed using anti-phosphotyrosine antibodies.

4.9 Conclusion and discussion

4.9.1 Structure and oligomerization of Tip α

In this Chapter, the structure of Tip α is described in two crystal forms. The structure of Tip α shows a novel fold: a α - β sandwich featuring three distinct regions: a short N-terminal flexible extension, a Dodecin-like domain (DoD) and a helical bundle structurally similar to PWI fold-containing proteins. This organization does not resemble any known bacterial virulence factors and therefore makes of Tip α structure a model for this family of proteins. Each crystal form of Tip α ^{N34} presents a different dimeric form. Crystal form I obtained at pH 4.0 showed an elongated dimer with few electrostatic contacts whilst crystal form II dimer, obtained at pH 8.5 showed extensive interactions between two subunits. Using gel-filtration performed at the pH found in the crystals, we show that Tip α ^{N34} is a monomer at pH4 and a dimer at pH 8.5.

However, the full length Tip α (Tip α ^{M21}) is a dimer at both pHs. Residues cysteines 25 and 27, present in Tip α ^{M21}, were found to form disulfide bridges (Suganuma, *et al.*, 2006)

and the results obtained confirm this and suggest that these disulfide bridges could maintain two Tip α subunit together when secreted in the gastric environment, at an acidic pH that would disrupt otherwise the interactions mediating the dimer (as seen in gel filtration chromatography).

It is possible therefore to propose that Tip α dimer (maintained by the disulfide bridges) opens at low pH and this opening can perhaps favor the interaction with a cellular receptor and/or the endocytosis inside the cell (Figure 4.14).

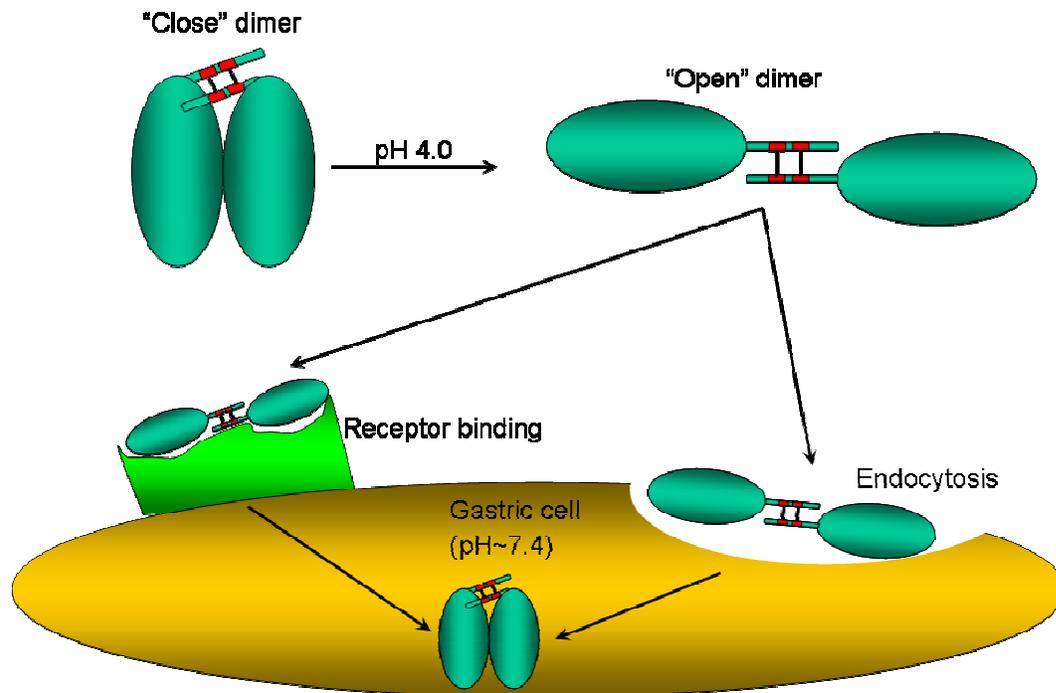


Figure 4.14: hypothetical mechanism of Tip α internalization in gastric cells.

The dimer, once in the acidic gastric environment, "opens" but the two subunits are held together by the cysteines in position 25 and 27 (in red). Once in the "open dimer" conformation, Tip α either binds to a specific unknown receptor placed on the surface of the gastric cells, or triggers its internalization through endocytosis. Once inside the cell, at a pH of around 7.5, the protein can adopt a "close dimer" form.

4.9.2 DNA binding activity

As reported in section 4.7, Tip α^{M21} was shown to bind to ssDNA and ds DNA *in vitro* with micromolar affinity (Kuzuhara, *et al.*, 2007). In a recent paper, Jang *et al.* reported a DNA mobility shift assay where a very weak band shift was obtained using a protein:DNA

ratio of 20:1 (Jang, *et al.*, 2009) and a DNA of 545 bp. It was also reported that a double lysine mutant, K102A/K104A, result in the loss of DNA binding activity (Figure 4.15).

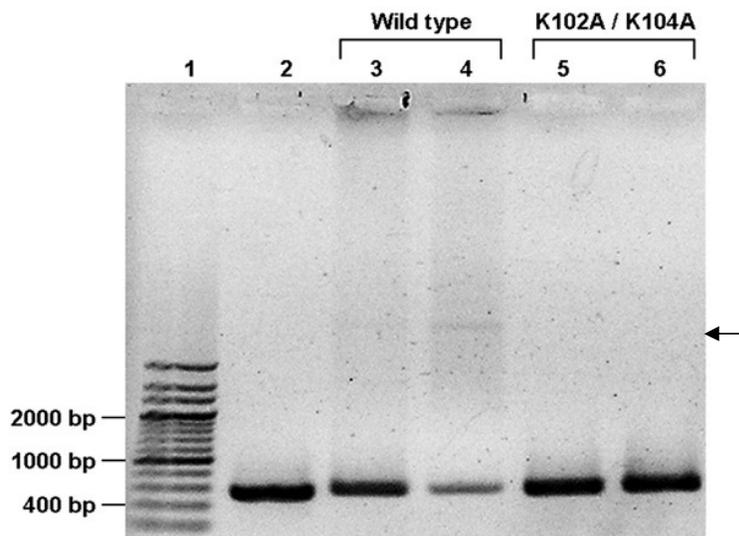


Figure 4.15: DNA gel-mobility shift assay performed by Jang and coworkers (Jang, *et al.*, 2009).

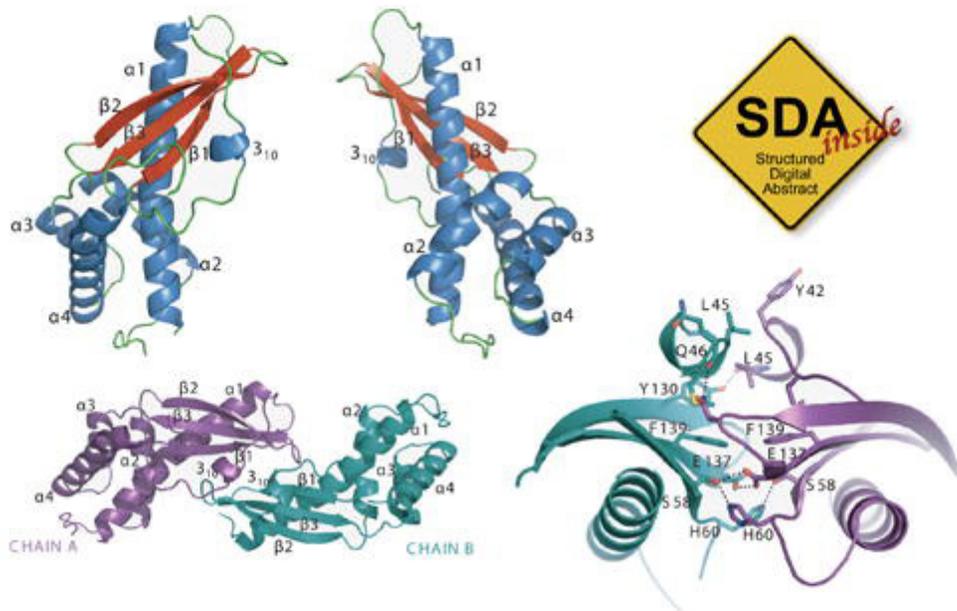
In lane 3 and 4, Tip α^{M21} was added to a DNA PCR product of 545 bp respectively at a 10:1 and 20:1 molar ratio (for Tip α^{M21} dimer to dsDNA respectively). At both ratios, a very weak shift of the DNA band is observed suggesting the formation of a Tip α^{M21} /DNA complex (indicated by an arrow). In lanes 5 and 6, the same experimental setup was repeated with a Tip α^{M21} double mutant (K102A/K104A) and the DNA binding activity is lost.

In this study, no DNA-binding activity for Tip α could be detected at protein:DNA ratios of up to 5:1. This result is in contradiction with that obtained by Suganuma *et al* (2007) and Jang *et al.*, (2009), however the concentrations and protein/DNA ratio used in Jang *et al.*, (2009) experiments are very high. At these concentrations, non specific DNA binding may occur. The Tip α^{N34} dimer structure is indeed characterized by an important distribution of positive charges on one face of the dimer and on the bottom part of the helical bundle domain. At high protein concentrations, like those used in the experiments of Jang *et al.*, the Tip α charge distribution could promote aspecific binding to DNA. The mutations of K102 and K104, both located at the bottom part of the helical bundle domain, would just decrease the distribution of positive charges in this region and, thus, reduce the binding of Tip α to DNA.

FEBS *Letters*

*The journal for rapid publication
of short reports in molecular biosciences*

- ▶ TDP-43 regulates synaptic terminals and locomotive behaviour
 - ▶ *N*-glycosylation of the human solute carrier PAT1
 - ▶ Structure of the TNF α inducing protein Tip α



Published by Elsevier on behalf of the Federation of European Biochemical Societies



Submit to:
<http://www.ees.elsevier.com/febsletters>

ISSN 0014 5793
Volume 583 Number 10 19 May 2009



Structures of the tumor necrosis factor α inducing protein Tip α : A novel virulence factor from *Helicobacter pylori*

Tommaso Tosi¹, Gianluca Cioci¹, Karina Jouravleva, Cyril Dian, Laurent Terradot*

European Synchrotron Radiation Facility, MX Group, BP 220, F-38043 Grenoble Cedex 9, France

ARTICLE INFO

Article history:

Received 16 March 2009
Revised 8 April 2009
Accepted 20 April 2009
Available online 4 May 2009

Edited by Hans Eklund

Keywords:

Bacterial toxin
X-ray crystallography
Nucleus import
Inflammatory response
Stomach cancer
Helicobacter pylori

ABSTRACT

Helicobacter pylori secretes a unique virulence factor, Tip α , that enters gastric cells and both stimulates the production of the TNF- α and activates the NF- κ B pathway. The structures of a truncated version of Tip α (Tip α N34) in two crystal forms are presented here. Tip α adopts a novel $\beta_1\alpha_1\alpha_2\beta_2\beta_3\alpha_3\alpha_4$ topology that can be described as a combination of three domains. A first region consists in a short flexible extension, a second displays a dodecin-like fold and a third is a helical bundle domain similar to the sterile alpha motif (SAM). Analysis of the oligomerisation states of Tip α N34 in the crystals and in solution suggests that the disulfide bridges could hold together Tip α monomers during their secretion in the gastric environment.

Structured summary:

MINT-7033680: TIP alpha (uniprotkb:B2UTN0) and TIP alpha (uniprotkb:B2UTN0) bind (MI:0407) by cosedimentation (MI:0027)

© 2009 Published by Elsevier B.V. on behalf of the Federation of European Biochemical Societies.

1. Introduction

Helicobacter pylori is a Gram-negative bacterium that colonises the human stomach. *H. pylori* is reported as the major cause of several gastric pathologies such as gastritis, peptic ulceration, stomach adenocarcinoma and mucosa-associated lymphoid tissue (MALT) lymphomas [1]. Virulence factors produced by *H. pylori* have a pivotal role in chronic inflammation and increase the risk of cancer [2,3]. The contribution of major virulence factors produced by *H. pylori* to cancer risks has been characterised such as the Cytotoxin associated gene A protein (CagA), the Vacuolating toxin A (VacA) and the BabA adhesin (reviewed in [2,4]). A highly immunogenic *H. pylori* protein, Tip α (for Tumor necrosis factor α – TNF- α – inducing protein) was identified as a new virulence factor activating NF- κ B and inducing the expression of TNF- α , a pro-inflammatory cytokine, produced in response to *H. pylori* infection [5–7]. Tip α also enhances the expression of different chemokines, upregulates Bcl-2 gene and downregulates p53 gene, thereby contribut-

ing to cancer establishment [8,9]. Tip α is present in all *H. pylori* genomes sequenced and has been detected in numerous clinical isolates from patients affected by gastritis, gastroduodenal ulcers and gastric cancers [6]. The mode of action of Tip α remains elusive but the protein has been shown to enter gastric cells and possibly penetrate the nucleus [10]. Tip α sequence displays no similarity to any known protein but a conserved motif has been proposed between Tip α and penicillin binding protein (PBP) suggesting that the protein might adopt a similar fold [11]. The protein is secreted in the extracellular medium [5,10,12,13] and forms homodimers of 38 kDa held together by inter-subunits disulfide bridges involving C25 and C27 [6]. *In vitro*, recombinant Tip α can bind both ss- and dsDNA [14]. A fragment of Tip α starting at residue 28 (delTip α) showed a considerably reduced ability to induce TNF- α gene expression, NF- κ B activation [6] and to enter gastric cells [10] suggesting an important role for the disulfide bridges.

We describe here the crystal structures of a truncated form of Tip α (Tip α N34) in two different crystal forms obtained at pH 4 and pH 8.5. These structures show that Tip α adopts a novel fold, consisting of an α - β sandwich combined with a helical bundle domain. Structural comparison and analysis provides insight into regions that may be involved in protein–protein interactions within the gastric cells. Gel filtration analysis of the oligomeric state of Tip α and Tip α N34 at different pHs suggests a role for the disulfide bridges during toxin secretion.

Abbreviations: MALT, mucosa-associated lymphoid tissue; PBP, penicillin binding protein; TEV, tobacco etch virus; SAD, single anomalous dispersion

* Corresponding author.

E-mail address: laurent.terradot@esrf.fr (L. Terradot).

¹ These authors contributed equally to this work.

2. Materials and methods

A DNA fragment corresponding to the secreted portion of Tip α (starting at M21) was amplified from *H. pylori* genomic DNA and cloned into pET151/D- topo vector (pET-Tip α). The protein was expressed in *Escherichia coli* and purified with a His-trap column (GE Healthcare) equilibrated with buffer A (30 mM Tris pH 8, 200 mM NaCl, 5% glycerol). Tip α was eluted with a linear gradient of imidazole from 0 to 500 mM in Buffer A. After tobacco etch virus (TEV) cleavage, the protein was further purified by Nickel affinity and size-exclusion chromatography. Seleno-methionine incorporated Tip α (SeTip α) was purified as Tip α except for NaCl concentration, raised to 500 mM in all buffers. Tip α N34 and SeTip α N34 were obtained by incubating respectively Tip α and SeTip α with trypsin at a protease/protein ratio of 1:1000 (w/w) on ice for 1 h. The vapour diffusion method was used for crystallisation. Form I crystals were grown by mixing equal amounts of protein (6 mg/ml) and reservoir solution containing 100 mM Na Citrate pH 4, 15 mM ATP and from 22% to 28% of Polyethylene glycol (PEG) 4000. Crystals of SeTip α N34 grew in 100 mM Na Citrate pH 4, 14% PEG6000 and 15 mM ATP. Crystal form II was obtained by mixing 2 μ l of Tip α N34 with 2 μ l of reservoir solution (100 mM Tris pH 8.5, 26% PEG4000, 50 mM MgCl₂). Data were collected at ESRF beamlines (Table 1). Native and Se-Met crystal form I diffracted to 1.9 Å and 2.0 Å, respectively, and belonged to the space group *P*2₁ with two molecules per asymmetric unit (Table 1). A highly redundant dataset was collected at the Se edge with SeTip α N34 crystals. Se sites were used to obtain experimental phases using the single anomalous

dispersion (SAD) method followed by density modification using SHARP [15]. ARP-WARP [16] built an initial model that was completed manually and refined against the native data using REFMAC [17]. Crystal form II diffracted at 1.7 Å and belonged to the *P*2₁2₁ space group with two molecules per asymmetric unit (Table 1). Crystal form II structure was solved by molecular replacement using PHASER [18] and the structure of form I as a search probe. The coordinates of Tip α crystal forms I and II structures have been deposited in the PDB, entry codes 2wcc and 2wcr, respectively. Analytical gel filtration experiments were performed by injecting 200 μ g of Tip α or Tip α N34 on a Superdex 75 10/300 column equilibrated in buffer A or in citrate buffer (50 mM citric acid pH 4, 200 mM NaCl). Details of the methods can be found in Supplementary data.

3. Results and discussion

3.1. Crystal structure of Tip α N34 form I

The structure of Tip α N34 monomer consists of an elongated, bean-shaped α - β sandwich combined to a helical bundle domain (Fig. 1A and B). It shows an original $\beta_1\alpha_1\alpha_2\beta_2\beta_3\alpha_3\alpha_4$ topology completed by a short 3_{10} helix at the N-terminal portion (Fig. 1A and B). The antiparallel β -sheet formed by β_1 , β_2 and β_3 wraps around the N-terminal half of α_1 . The C-terminal portion forms a four helix bundle with α_2 , α_3 and α_4 , at the top of which lies a long loop connecting α_1 - α_2 (Fig. 1B). Differences between the two subunits reside essentially at the N-terminal region, which appears more flexible with L37 to I52 displaying high B-factor values. This extension together with the bottom part of the β -sheet also mediates inter subunit contacts via hydrophobic interactions between V40, L45, Y42 and F139 from chain A with V40, M44, L45, F139 from chain B (Fig. 1C). Those contacts are reinforced by a network of hydrogen bonds between S58, H60, E137 from both subunits resulting in a head-to-head dimer related by a pseudo twofold axis (Fig. 1C and D). However, this dimer interface, burying 814 Å², is centred on E137-E137 interactions, contacts that would likely be unfavoured in solution at a physiological pH.

3.2. Domain organisation and structural similarity with known proteins

A DALI [19] search identified more than 500 proteins with Z-scores lower than 4.3 but none of these structures were larger than 100 residues and no homologue was identified for the entire Tip α N34 structure. Structural homology analysis divides Tip α N34 in three parts: a N-terminal loop, consisting of residues N34 to I52, showing no similarities to existing structures. A second domain, consisting of β -strands β_1 , β_2 , β_3 and the N-terminal half of α_1 , is similar to dodecin proteins, in particular to *H. salinarum* dodecin (pdb 1mog, rmsd 2.2 Å, 59 residues). Dodecins are small proteins that oligomerise in dodecamers and bind to FAD [20]. Tip α N34 superimposes well with dodecin monomer (Fig. 2A) but the FAD-binding residues (W36, E45 and Q55) are not conserved. The third domain consists of the helical bundle, similar to more than 100 proteins with best scores falling into three categories: (i) the nuclear abundant RNA-binding Nab2 protein (pdb 2v75, rmsd 2.7 Å, 70 residues), (ii) proteins containing the sterile alpha motif (SAM) domain and (iii) the translation initiation factor 2 alpha, aIF2 (pdb 1yz6, rmsd 4.8 Å, 90 residues) (Fig. 2B). The helical bundles of Tip α and Nab2 are similar to the PWI fold, a domain found in several nucleic acid binding proteins [21]. However, the motif involved in DNA binding in PWI-containing proteins is not conserved in Tip α nor in Nab2 [21]. More than 20 proteins containing a SAM domain were also identified as putative homologues, includ-

Table 1

Data collection, phasing and refinement statistics.

	Tip- α N34 Se-Met pH 4	Tip- α N34 pH 4	Tip- α N34 pH 8.5
Data collection statistics			
Beamline	ID14-4	ID14-4	ID29
Unit cell (Å)			
a	31.64	31.36	48.88
b	102.21	101.99	67.26
c	47.53	47.25	97.47
β	94.40	94.38	
Space group	<i>P</i> 2 ₁	<i>P</i> 2 ₁	<i>P</i> 2 ₁ 2 ₁ 2 ₁
Wavelength (Å)	0.979	0.939	0.981
Resolution limit (Å)	43.0–2.0	50.9–1.9	27.7–1.7
Total observations	151 032	93 609	128 160
	(22 287) [*]	(10 799)	(18 556)
Unique reflections	20 350 (2989)	22 965 (3030)	35 854 (5158)
Completeness	99.9 (99.9)	98.4 (90.4)	99.3 (99.6)
Multiplicity	7.4 (7.5)	4.1(3.6)	3.6 (3.6)
<i>I</i> / (σ <i>I</i>)	7.5 (1.8)	7.6 (2.2)	10.8 (1.9)
R _{merge} (%) ^b	7.4 (40.3)	6.0 (32.5)	4.0 (40.7)
Wilson B-factor (Å ²)	24.4	25.2	26.4
Phasing statistics			
Se sites	10		
FOM (centric/acentric)	0.13/0.41		
Anomalous phasing power	1.74		
Refinement statistics			
R _{cryst}		19.6 (27.7)	18.2 (29.3)
R _{free}		23.1 (37.7)	21.9 (34.7)
RMS bonds		0.015	0.017
RMS angles		1.482	1.51
Ramachandran outliers		Ile52B	Ser49A
B-factors			
Overall		28.0	17.0
Protein atoms		27.6	15.6
Water atoms		35.0	26.3

$$R_{\text{cryst}} = \frac{\sum ||F_{\text{obs}} - F_{\text{calc}}||}{\sum ||F_{\text{obs}}||}$$

^{*} Values in parentheses refer to the highest resolution shell.

$$R_{\text{merge}} = \frac{\sum |I - \langle I \rangle|}{\sum I}$$

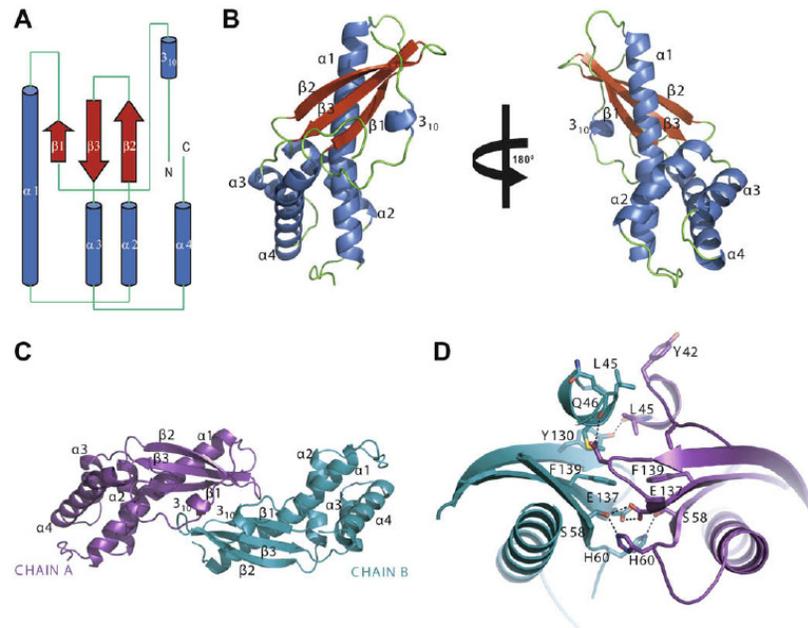


Fig. 1. Structure of Tip α N34. (A) Topology diagram of Tip α N34. (B) Two different views of Tip α N34 structure with α -helices in blue, β -strands in red and loops in green. (C) Ribbon representation of the dimer with chain A and B colored in violet and blue, respectively. (D) Head-to-head dimer interface depicted in ribbon with side chains indicated as sticks (with carbon atoms colored depending of the subunit, nitrogen in blue, oxygen in red and sulfur in yellow).

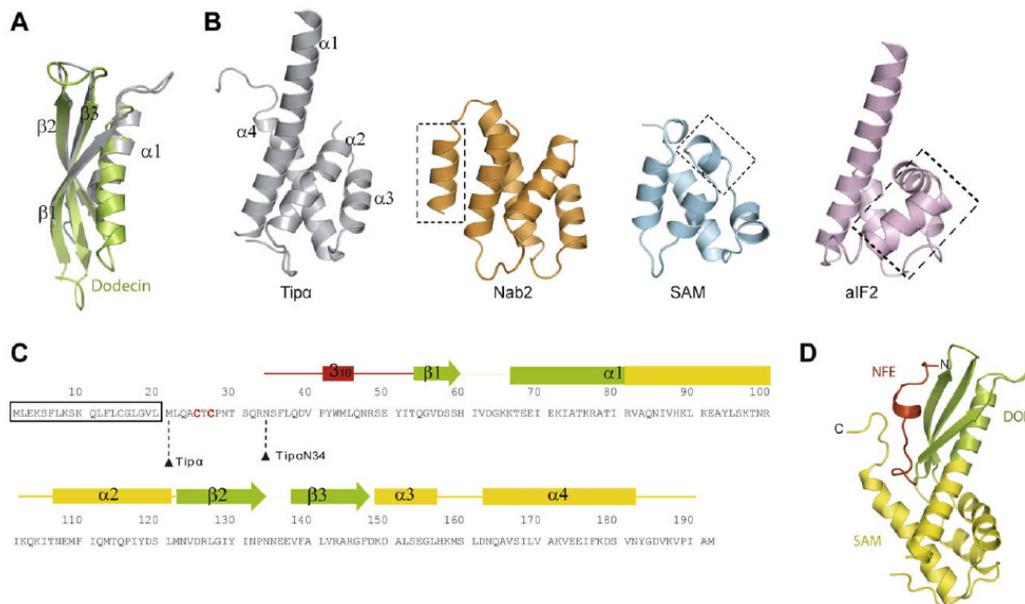


Fig. 2. Domain organisation of Tip α structure. (A) Structural superimposition of Tip α N34 DOD (grey) with dodecin (green). (B) Comparison of the Tip α SAM domain (grey) with the structure of Nab2 (orange), SAM domain from polyhomeotic-proximal chromatin protein (blue) and aIF2 (pink). Areas of main divergence are indicated by dashed boxes. (C) Sequence of Tip α with secondary structures indicated above and colored according to their respective domains (NFE, red; DOD, green and SAM, yellow). The signal peptide is indicated by a black box, C25 and C27 are colored in red. Starting residues for Tip α and Tip α N34 are indicated. (D) Structure of Tip α N34 colored according to domain boundaries.

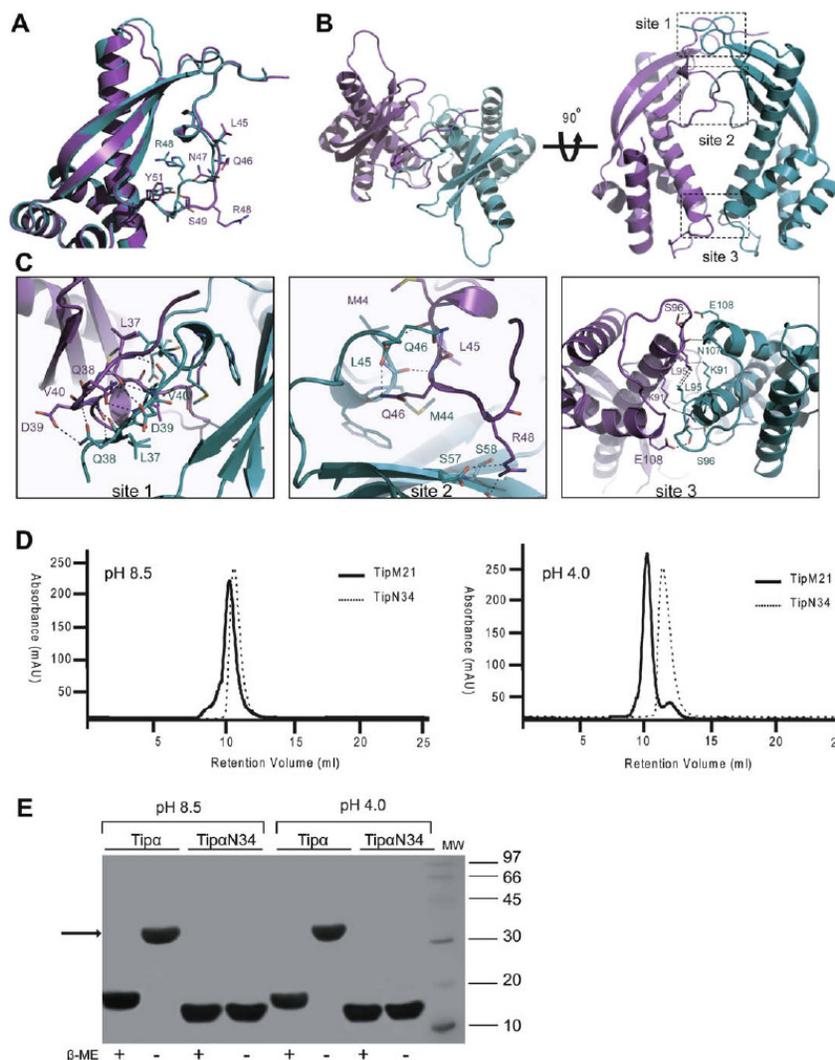


Fig. 3. Crystal form II of Tip α N34. (A) Superimposition of chain A and chain B showing the different conformations of NFE loop. (B) Two views of the dimer of Tip α N34 at pH 8.5 with chain A and B colored in violet and blue, respectively. Dotted boxes points to the three contact sites between subunits. (C) Close-up view of Tip α N34 dimer interaction sites indicated in (A). Side chains of the residues involved are displayed in sticks. (D) Gel filtration analysis of Tip α and Tip α N34 at pH 4 and pH 8.5. (E) SDS-PAGE analysis of the gel filtration experiments at pH 8.5 and pH 4 in the presence or absence of 5 mM β -mercapto-ethanol (β -ME). The arrow indicates Tip α dimer and the molecular weight markers (MW) sizes are indicated on the right in kDa.

ing polyhomeotic-proximal chromatin, sex comb, aveugle and ephA4 receptor tyrosine kinase. SAM domains are short domains found in a wide variety of eukaryotic proteins and in few bacterial ones [21,22]. The structural similarity of Tip α to SAM domain raises interesting questions since these domains can mediate protein–protein, protein–DNA and protein–RNA interactions [23]. Tip α N34 also resembles domain II of aIF2 (archaea) and eIF2 (eukaryotes) but with a high rmsd and a good fit only with α 1 (Fig. 2D). Thus, the structure of Tip α N34 can be considered as a novel fold containing three regions: a N-terminal flexible extension (NFE), a dodecin-like domain (DOD) and a SAM-like domain (SAM) (Fig. 2C and D).

3.3. Crystal structure of form II and dimer organisation

The structure of crystal form II monomer is essentially identical to crystal form I. Major differences are located in the NFE region that undergoes an important conformational change that swaps the orientation of the N-terminus. Strand β_1 is rotated by 45° and both β_2 and β_3 are shifted in the structure at pH 4 compared to the crystal form I. Noteworthy, this extension adopts very different conformations in the two chains (Fig. 3A). In chain A, residues from this loop have their side chains oriented towards the solvent, disconnecting the loop from the rest of the molecule and forming a small solvent channel (Fig. 3A). In chain B, the same side chains,

in particular R48, point towards the concave face of the protein which changes the orientation of the loop and closes the solvent channel. The two molecules in the asymmetric unit are arranged in a different dimer than form I (Figs. 1D and 3B). This arrangement also generates a pseudo twofold symmetry axis that goes from the distal N-terminus to the helical bundle burying a total of 1358 Å² (Fig. 3B). The two subunits are held together by numerous electrostatic and few hydrophobic interactions in three sites (Fig. 3C). At site 1 the NFEs interact with each other. L37, Q38, D39 and V40 of chain A interact with V40, D39, Q38 and L37 of Chain B, respectively, to form a highly connected structure (Fig. 3C). At site 2, the NFEs interact with each other and the NFE from chain A also contacts β1 of chain B. NFEs interactions include hydrogen bonds between M44 and Q46 and L45 and Q46. In addition, Q46A side chain interacts also with W43 of chain B. On one side only, the side chain of R38 from chain A interacts with D57 from chain B (Fig. 3C). Site 3 comprises interactions between K91 and E92, L95 and N107, S96 and E108, and L95 side chains from both chains.

3.4. Oligomeric state of Tipα

To determine the oligomerisation states of Tipα and TipαN34 in solution, we performed analytical gel filtration experiments (Fig. 3D). At pH 8.5, both Tipα and TipαN34 eluted at a retention volume of 10.9 ml, corresponding to a molecular weight of around 30 kDa, in agreement with a theoretical Tipα homodimer. To mimic the form I crystallisation conditions, we performed analytical gel filtrations at pH 4 and found that TipαN34 eluted at 12 ml, corresponding to a monomer. In contrast, Tipα still eluted as a dimer at around 10.9 ml (Fig. 3D). SDS-PAGE experiments on protein samples corresponding to the elution peak of the two gel filtration experiments, confirm that, in absence of reducing agent, Tipα forms dimers but not TipαN34 (Fig. 3E). This strongly suggests that crystal form I dimer is likely a crystallisation artefact and that TipαN34 dimer observed in crystal from II at pH 8.5 is more likely to be relevant in solution.

4. Conclusions

Tipα is a unique virulence factor, secreted by all *H. pylori* strains, capable of penetrating gastric cells [10]. Our study reveals that this protein adopts an original fold with no resemblance to known PBPs as previously suggested. Tipα structure is a combination of three interconnected regions. These domains point towards putative protein–protein or protein–nucleic acid interaction sites that could play important roles in Tipα function. Finally, C25 and C27 were suggested to be crucial for dimer formation and Tipα activity [10,14]. We have shown that the two cysteines C25 and C27 are not necessary for Tipα to form a dimer at pH 8.5. However, at pH 4, their presence seems to be required in solution to hold the two subunits together. It is possible that *in vivo* these disulfide bridges could be necessary for keeping Tipα in its dimeric form when released into the acidic extracellular medium, i.e., the gastric mucosa.

Acknowledgements

We wish to thank members of the MX group at ESRF for help during data collection and Joanna Timmins for critical comments

on the manuscript. This work was funded by the ESRF “In House Research” program.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.febslet.2009.04.033.

References

- [1] IARC (1994) Schistosomes, liver flukes and *Helicobacter pylori*. IARC Working Group on the evaluation of carcinogenic risks to humans. Lyon, 7–14 June 1994. IARC Monogr. Eval. Carcinog. Risks Hum. 61, 1–241.
- [2] Franco, A.T. et al. (2008) Regulation of gastric carcinogenesis by *Helicobacter pylori* virulence factors. *Cancer Res.* 68, 379–387.
- [3] Ferreira, A.C., Isomoto, H., Moriyama, M., Fujioka, T., Machado, J.C. and Yamaoka, Y. (2008) *Helicobacter* and gastric malignancies. *Helicobacter* 13 (Suppl. 1), 28–34.
- [4] McNamara, D. and El-Omar, E. (2008) *Helicobacter pylori* infection and the pathogenesis of gastric cancer: a paradigm for host–bacterial interactions. *Dig. Liver Dis.* 40, 504–509.
- [5] Yoshida, M. et al. (1999) Cloning and characterization of a novel membrane-associated antigenic protein of *Helicobacter pylori*. *Infect. Immun.* 67, 286–293.
- [6] Suganuma, M., Kurusu, M., Suzuki, K., Nishizono, A., Murakami, K., Fujioka, T. and Fujiki, H. (2005) New tumor necrosis factor-α-inducing protein released from *Helicobacter pylori* for gastric cancer progression. *J. Cancer Res. Clin. Oncol.* 131, 305–313.
- [7] Suganuma, M., Kurusu, M., Okabe, S., Sueoka, N., Yoshida, M., Wakatsuki, Y. and Fujiki, H. (2001) *Helicobacter pylori* membrane protein 1: a new carcinogenic factor of *Helicobacter pylori*. *Cancer Res.* 61, 6356–6359.
- [8] Cheng, P., Shi, R.H., Zhang, H.J., Yu, L.Z., Zhang, G.X. and Hao, B. (2008) Effects of tumor necrosis factor-α inducing protein-α secreted by *Helicobacter pylori* on human gastric epithelial cells. *Zhonghua Yi Xue Za Zhi* 88, 1528–1532.
- [9] Kuzuhara, T., Suganuma, M., Kurusu, M. and Fujiki, H. (2007) *Helicobacter pylori*-secreting protein Tipα is a potent inducer of chemokine gene expressions in stomach cancer cells. *J. Cancer Res. Clin. Oncol.* 133, 287–296.
- [10] Suganuma, M. et al. (2008) TNF-α-inducing protein, a carcinogenic factor secreted from *H. pylori*, enters gastric cancer cells. *Int. J. Cancer* 123, 117–122.
- [11] Kuzuhara, T., Suganuma, M., Tsuge, H. and Fujiki, H. (2005) Presence of a motif conserved between *Helicobacter pylori* TNF-α inducing protein (Tipα) and penicillin-binding proteins. *Biol. Pharm. Bull.* 28, 2133–2137.
- [12] Godlewska, R., Pawlowski, M., Dzwonek, A., Mikula, M., Ostrowski, J., Drela, N. and Jaguszyn-Krynicka, E.K. (2008) Tip-α (hp0596 gene product) is a highly immunogenic *Helicobacter pylori* protein involved in colonization of mouse gastric mucosa. *Curr. Microbiol.* 56, 279–286.
- [13] Suganuma, M., Kuzuhara, T., Yamaguchi, K. and Fujiki, H. (2006) Carcinogenic role of tumor necrosis factor-α inducing protein of *Helicobacter pylori* in human stomach. *J. Biochem. Mol. Biol.* 39, 1–8.
- [14] Kuzuhara, T., Suganuma, M., Oka, K. and Fujiki, H. (2007) DNA-binding activity of TNF-α inducing protein from *Helicobacter pylori*. *Biochem. Biophys. Res. Commun.* 362, 805–810.
- [15] de La Fortelle, E. and Bricogne, G. (1997) Maximum-likelihood heavy-atom parameter refinement in the MIR and MAD methods.
- [16] Perrakis, A., Morris, R. and Lamzin, V.S. (1999) Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* 6, 458–463.
- [17] Murshudov, G.N., Vagin, A.A., Lebedev, A., Wilson, K.S. and Dodson, E.J. (1999) Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr. D Biol. Crystallogr.* 55, 247–255.
- [18] 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.
- [19] Holm, L. and Sander, C. (1998) Dictionary of recurrent domains in protein structures. *Proteins* 33, 88–96.
- [20] Grininger, M., Seiler, F., Zeth, K. and Oesterheld, D. (2006) Dodecin sequesters FAD in closed conformation from the aqueous solution. *J. Mol. Biol.* 364, 561–566.
- [21] Grant, R.P. et al. (2008) Structure of the N-terminal Mlp1-binding domain of the *Saccharomyces cerevisiae* mRNA-binding protein, Nab2. *J. Mol. Biol.* 376, 1048–1059.
- [22] Ponting, C.P. (1995) SAM: a novel motif in yeast sterile and *Drosophila* polyhomeotic proteins. *Protein Sci.* 4, 1928–1930.
- [23] Kim, C.A. and Bowie, J.U. (2003) SAM domains: uniform structure, diversity of function. *Trends Biochem. Sci.* 28, 625–628.

*General Conclusions and
Perspectives*

H. pylori is a major human pathogen that infects half of the world population. It thrives in the human stomach and is associated with virtually all gastric diseases, including ulcers and gastric adenocarcinoma. Two virulence factors are major determinants of *H. pylori* pathogenicity. The CagA cytotoxin has been widely studied and found to trigger numerous cellular events in gastric cells, potentially leading to gastric cancer. Tipα is a recently discovered *H. pylori* protein presenting original features, that also associates with gastric cancers. To gain insight the structure/function of these proteins, a panel of biochemical and biophysical techniques was used to study these two important proteins.

CagA protein-protein interaction map

The *H. pylori* protein-protein interactions maps (PIMs) is a valid starting tool to dissect the numerous interactions in which a given protein is involved. In the case of CagA, only two interactions could be confirmed, those between a domain of CagA (CagA³⁹²⁻⁷³³) and two *H. pylori* proteins, RibA and EstV. RibA is an enzyme and its activity is specific towards GTP only (Ren, *et al.*, 2005). EstV is a short-chain specific lipase that could have a role in *H. pylori* colonization (Ruiz, *et al.*, 2007), but no specific association between EstV and CagA has been reported. For now, it is difficult to conclude on the validity and thus the relevance of the interactions between CagA and RibA and CagA and EstV. This is in part because the interaction of CagA³⁹²⁻⁷³³ with RibA seems transient and the one CagA³⁹²⁻⁷³³ of EstV seems very weak. On the other hand, the unavailability of other CagA fragments to be tested prevented the study of these interactions in more detail and to provision of a firm conclusion.

This problem now is partially solved by the identification of several soluble CagA fragments through the ESPRIT methodology (Chapter 2). These new fragments should help to study the interactions of CagA with RibA and EstV *in vitro*. Moreover, the biological relevance should also be analysed by, for instance, determining if these proteins interact in *H. pylori* cells. It would also be interesting to determine if the deletions of the genes encoding RibA, EstV and other proteins found in the PIM of CagA influence the translocation process of the toxin.

Application of the ESPRIT methodology to CagA

This part of the thesis describes both the first application of this High-Throughput methodology for a bacterial protein and the first systematic identification and analysis of CagA soluble and purifiable fragments. The findings presented in this thesis suggest that CagA is made of two main “maxi-domains”, separated by a linker of five asparagines that is cleaved in the *E. coli* cells. The proteolysis of CagA in the human cell, which was previously demonstrated, could serve a purpose. Indeed, this reaction produces two domains that could, in the human cell, interact with different host proteins *in vivo* via different mechanisms of action. The N-terminal part could interact with the tight junction proteins from the membrane (Zo-1 and JAM, Amieva et al., 2003) or with other proteins (Grb2, E-cadherin, β -catenin, c-Met) that are involved in loss of cell adhesion (Wen and Moss, 2009). The C-terminal part can interact with signalling proteins (PAR1b/MARK, Csk, Crk, SHP2), hijacking the corresponding pathways and contributing to cell proliferation and scattering (Hatakeyama, 2009).

None of the CagA fragments could be crystallized. It is possible that the fragments that were identified contain very flexible parts, connecting small domains one to another, that may have hampered the crystallization. Furthermore, as CagA binds to numerous partners (as widely discussed in the introduction of this thesis), it is possible that the fragments we identified can crystallize only in the presence of a partner (bacterial or human) that help the protein to fold properly or stabilize it. Future works should “refine” the CagA constructs reported here to identify other subdomains and test their interactions with human and bacterial protein partners. Furthermore, a protocol should be established to try to identify stable CagA fragments comprised in the N-terminal part of the protein. In this case, the usage of a bidirectional truncation of CagA (as suggested in chapter 2) can prove useful, as well as the limited proteolysis of the degradation fragments resulting from the expression of the full-length CagA.

The works reported here pave the way to identify CagA/protein complexes, that can subsequently be used to pursue structural investigations of the CagA protein.

CagA oligomerization and architecture

As reported in Chapter 3, CagA^{C33} is composed by two domains, CagA^{C11} and

CagA^{C22}. CagA^{C33} can multimerize via two different mechanisms: one mediated by a disulfide bridge between two C1164 residues and one non-covalent, that likely involves CagA^{C11} and CagA^{C22}. This suggests that oligomerization of CagA can take place without additional partners, in contrast with what has been suggested in other studies. Although more work will be necessary to define the exact mode(s) of oligomerization, our findings suggest that CagA can form oligomers and even particles inside the cell.

That CagA^{C11} on its own can be phosphorylated suggests that CagA^{C33} can also be divided into two main functional domains: 1) a multifunctional effector domain (CagA^{C11}), that is involved in phosphorylation-dependent and phosphorylation-independent interactions with host proteins, and 2) a translocation domain (CagA^{C22}), responsible for CagA delivery inside host cells and containing the CagF-binding sequence and the C-terminal translocation signal.

However, it is questionable whether CagA C1164 mediated dimerization is functionally relevant. C1164 is, in fact, not completely conserved and the only strains of *H. pylori* carrying a CagA with a cysteine residue are 26695 (C1168), G27 (C1230), P12 (C1192) and HPAG-1 (C1178). It is interesting to note that, in these CagA proteins, only one cysteine is present in the entire sequence and is found always at the extremity of the C-terminus (in the case of CagA from the strain G27, C1230 is the very last residue). We can hypothesize that these cysteines might generally be exposed and could mediate the covalent dimerization of CagA. It is rather unlikely that this type of dimerization could take place in the cytoplasmic space of *H. pylori*, as Gram-negative bacteria are characterized by a reducing environment in the cytoplasm. However, when CagA is translocated in the gastric cells it is possible that it can form a covalent dimer. It would be interesting to investigate, in the future, whether this covalent dimer can influence somehow the effects exerted by CagA in gastric cells.

CagA phosphorylation-dependent interactions

The analysis of P-CagA/SH2 interaction constitutes the first attempt to investigate *in vitro* how CagA binds to and activates SH2-containing proteins, a determinant aspect of CagA function. We show that the phosphorylated EPIYA-C site is the motif binding with the

highest affinity to v-SH2. V-SH2 was selected because of its extensive characterization in scientific literature (Waksman, *et al.*, 2004). Furthermore, as shown in Figure C.1, the structures of v-SH2 and of the N-terminal SH2 domain (NSH2) of the SHP-2 phosphatase are similar and the two critical binding residues (R32 and H58) are conserved in all the aligned SH2 domains. It thus makes of v-SH2 an excellent model to study P-CagA and SH2 interactions.

As this domain is extremely similar to the SH2 domain from c-Src (2 amino acid differences), It is also possible to hypothesize that P-CagA might bind to SH2 from c-Src. A possibility that has been mentioned but never demonstrated: P-CagA was shown to inhibit the phosphorylation activity of c-Src *in vitro* (Selbach, *et al.*, 2003) but no mechanism was proposed for this inhibition. The work presented in this thesis suggests that P-CagA could inhibit c-Src activity by first binding to its SH2 domain (c-SH2). However, this binding must be accompanied by another mechanism since, binding to c-Src SH2 domain generally activates this kinase. Further work will be necessary to determine if P-CagA binding to c-SH2 is sufficient to inactivate c-Src.

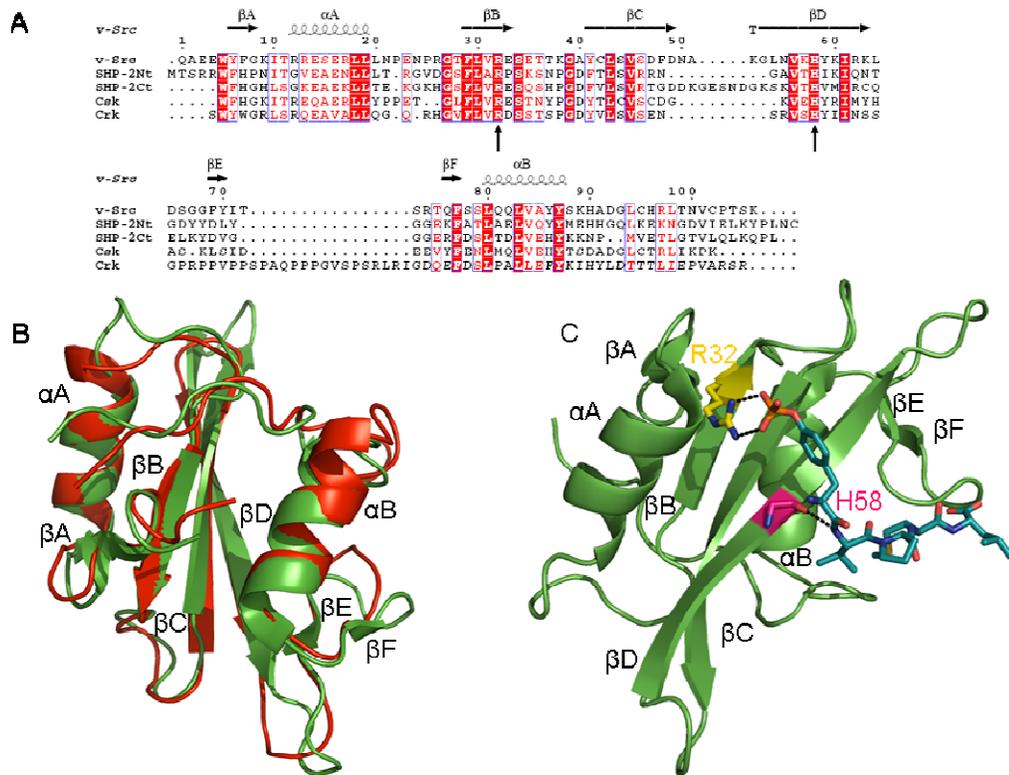


Figure C.1: sequence and structure alignment between v-SH2 and the SH2 domains of other phosphorylated CagA interaction partners.

(A): Sequence alignment of different SH2 domains. V-Src: SH2 domain of v-Src; SHP-2Nt: NSH2 domain; SHP-2Ct: CSH2 (immediately downstream the NSH2 domain); Csk: SH2 domain of Csk; Crk: SH2 domain of Crk. Conserved binding residues (R32 and H58) are indicated with an arrow (B): Alignment of the crystal structures of v-SH2 (red) and NSH2 of SHP-2 (green). Secondary structure elements are also indicated (C): interactions between R32 (yellow) and H58 (magenta) with a phospho-pentapeptide co-crystallized with v-SH2.

It has been demonstrated here that c-Abl can phosphorylate the three EPIYA motifs while c-Src phosphorylates only EPIYA -B and EPIYA-C. This specificity is interesting because the EPIYA motifs are not equivalent. It was found, using phosphorylated peptides corresponding to the sequences of the EPIYA motifs, that P-EPIYA-A can bind to Csk and the phosphatase SHP-1, P-EPIYA-B to Grb2, Csk, SHP-1 and EPIYA-C to SHP-2, SHP-1 and Grb2. (Selbach, *et al.*, 2009). Therefore by phosphorylating a subset of the EPIYA motifs, the respective kinases might activate the binding of CagA to specific partners. This could modulate CagA effects and offer a means of regulation of CagA activities, once inside the human cell. It will therefore be interesting in future works to use the different CagA constructs and mutants described here to see if the various P-EPIYAs affect differently each CagA host targets, such as s-Src, Csk, Crk, SHP-1 and SHP-2 *in vitro*.

Tipa, a novel bacterial toxin

The structure of Tipa^{N34} was solved and described in this thesis. The analysis performed showed that its structure is unique and summarise the main structural features of the full length protein. This work opens new and interesting perspectives in the study of Tipa. First, the results presented will be useful for the design of new experiments. The experiments described here clearly show that Tipa can be tyrosine-phosphorylated by both c-Src and c-Abl *in vitro* even though a more complete study will be necessary to fully characterize Tipa phosphorylation. The phosphorylation of Tipa is interesting because it suggests new mechanisms of action of this protein, also consistent with the mode of action of the other cancer-associated toxin of *H. pylori*, CagA. Further studies should therefore address the question of Tipa phosphorylation *in vivo* and if this plays a role in its function(s) in gastric cells.

The import of Tipa in the cell nucleus is also an very intriguing aspect of Tipa

function. In a recent study, it was reported that Tip α can bind to the human nucleolin (Tsuge, *et al.*, 2009). Nucleolin is an abundant multifunctional eukaryotic phosphoprotein involved in several critical cellular processes, like DNA or RNA binding, ribosome biogenesis, signal transduction, cell growth and differentiation, transcriptional repression and more (Tuteja & Tuteja, 1998). Further work will have to investigate whether the nuclear import of Tip α ^{M21} can be mediated by this protein and/or by other nucleolin-interacting cellular partners. Nucleolin can both bind to phosphorylated proteins and trigger their phosphorylation by human kinases (Di Segni, *et al.*, 2008). Therefore, it could be interesting to examine the role(s) of Tip α phosphorylation in Tip α /nucleolin interaction.

Appendix

A.1 Overview of the vector pET151/D- TOPO

pET151/D- TOPO is a pET derived plasmid, containing a strong T7 promoter modified to contain a *lac* operator sequence to allow the IPTG-driven expression of the recombinant protein.

In this vector, ligation occurs through the action of the enzyme Topoisomerase I from a *Vaccinia* virus, which cleaves the phosphodiester bond present after a consensus sequence 5'-CCCTT in a duplex DNA strand. After the cleavage, tyrosine 274 of Topoisomerase I covalently binds with its side chain hydroxyl group to the 3'-phosphate group remaining on the cleaved strand, generating a DNA/protein complex. The phosphotyrosil bond can be cleaved by the 5'-OH group on the DNA strand, reversing the reaction and releasing the enzyme.

To clone a gene inside the pET151/D- TOPO vector, the forward primer must contain a 4 base overhang (CACC) before the first codon of the PCR product. The vector contains a single-stranded GTGG overhang on one strand, that anneals with the CACC on the forward primer, stabilizing the PCR product in the correct orientation. The phosphotyrosil bond of the Topoisomerase I linked on the opposite strand is attacked by the 5'-OH group of the CACC sequence, releasing the enzyme and creating the phosphodiester bond with the sequence CCCTT (Figure A.1).

The gene is cloned in frame with a 5' DNA sequence encoding a N-terminal His₆-tag, followed by a V5 epitope and a Tobacco Etch Virus (TEV) protease cleavage consensus allowing for an easy removal of the His₆-tag from the expressed protein with a recombinant TEV protease (Figure A.2)

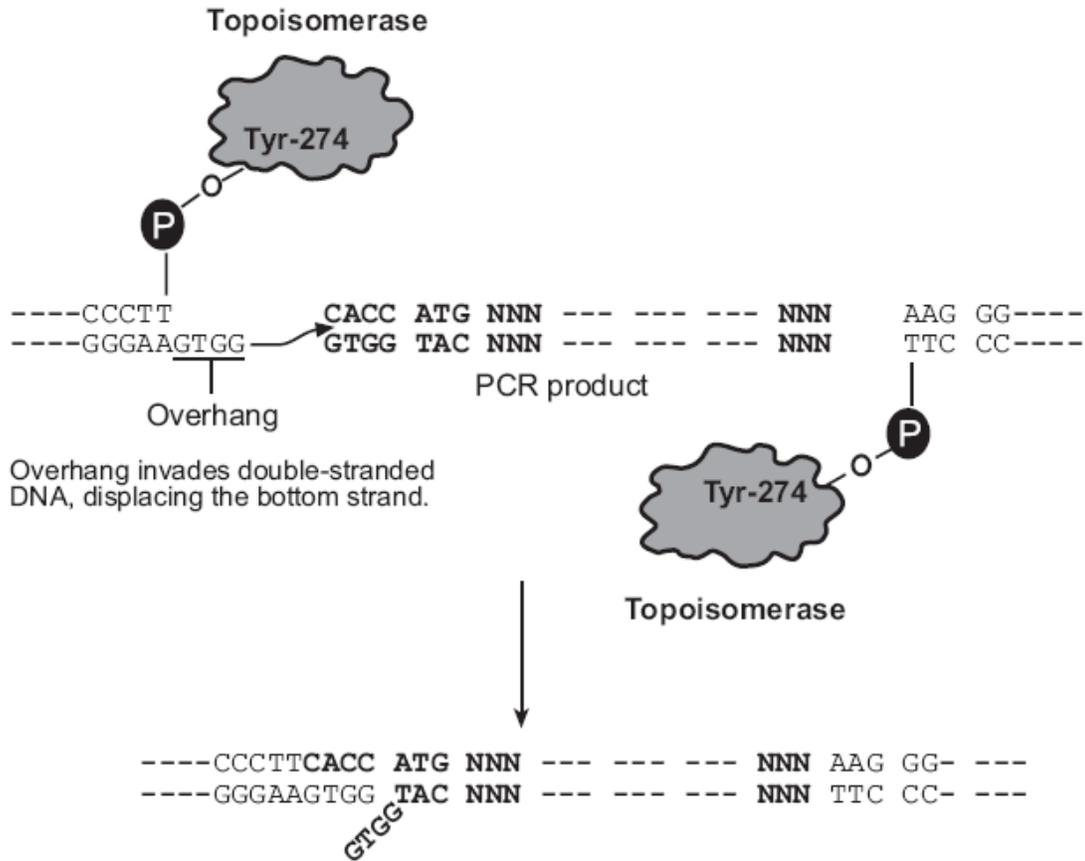


Figure A.1: cloning of a PCR product in the pET151/D- TOPO vector.

The Topoisomerase 1 is attached to the 3'-OH at the end of the CCCTT sequence on one strand of the vector. On the opposite strand, the GTGG overhang anneals with the CACC sequence present at the 5' end of the gene, stabilizing the PCR product in the correct orientation. The Topoisomerase 1 is then displaced by the attack of the CACC 5'-OH to the 3' phosphate group, creating the phosphodiester bond.

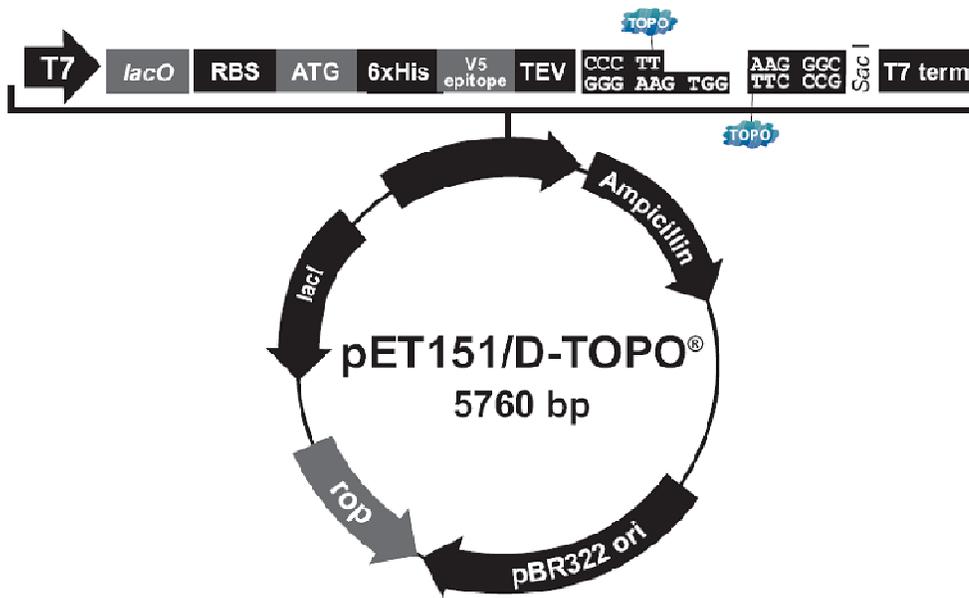


Figure A.2: Schematic view of the pET151/D- vector.

The PCR product is cloned in frame with a hexa-histidine tag, a V5 epitope and a TEV cleavage site. The V5 epitope is used to detect the His₆-tag with specific anti-V5 antibodies. T7: T7 promoter; *lacO*: *lac* operator for IPTG-induced protein expression; RBS: ribosome binding site; ATG: methionine codon to start the expression; 6xHis: hexa-histidine tag; V5 epitope: V5 epitope; TEV: Tobacco Etch Virus protease cleavage site; TOPO: Topoisomerase 1; SacI: restriction sequence for the enzyme SacI. T7 term: T7 terminator for detaching the ribosome after protein expression.

A.2: Dynamic Light Scattering (DLS)

When the diameter (d) of a particle is substantially smaller than the wavelength (λ) of an incident laser light ($d \leq \lambda/10$), laser light will be scattered isotropically by the particle in all directions. The intensity of the scattered light is proportional, according to the Rayleigh approximation, to the particle diameter:

$$I \propto d^6$$

The Rayleigh approximation (which is the basis of the “Rayleigh Scattering”) is the principle used in DLS to measure the diameter of a particle (such as a protein) in solution. A DLS instrument measures the intensity of the light scattered by the sample and correlates this with particle size.

As the particles in solution are subject to Brownian motion (caused by the solvent molecules hitting the particles), the intensity of the light scattered by the sample fluctuates with time. The DLS instrument measures the intensity of the scattered light at very small time intervals (usually in the μs range) and constructs a “correlation function” ($G(\tau)$) to extrapolate the hydrodynamic diameter of the particle:

$$G(\tau) = A[1 + B^{(-2\Gamma\tau)}]$$

where τ is the time interval, A is the baseline and B the intercept of the correlation function (i.e. the intensity of the scattered light plotted versus time), Γ is another function, described as

$$\Gamma = Dq^2$$

where D is the translational diffusion coefficient (the “speed” of the Brownian motion) and

$$q = \frac{4\pi n}{\lambda_0} \sin \frac{\theta}{2}$$

where n is the refractive index of the solvent, λ_0 is the wavelength of the laser light and θ is the scattering angle.

The hydrodynamic diameter of the particle is extrapolated, knowing D , from the formula

$$d(H) = \frac{kT}{3\pi\eta D}$$

Where $d(H)$ is the hydrodynamic diameter, k is the Boltzmann constant ($1.38 \cdot 10^{-23} \text{ JK}^{-1}$), T is the temperature, D is the translational diffusion coefficient and η is the viscosity coefficient of the medium.

The correlation function is then used by DLS instrument (such as DLSNano, Malvern

Instruments, which is the one used in this work) to plot the relative intensity of the light scattered by the particle(s) with their size (in nm). A narrow peak gives a precise estimation of the size (i.e. the hydrodynamic diameter) of the particle. The measurement of the translational diffusion coefficient is influenced by many factors, like the temperature, the viscosity of the solvent and its ionic strength.

This information was taken from the user guide “Dynamic Light Scattering: an introduction in 30 seconds” by Malvern Instruments (www.malvern.com). A schematic view of the DLS process is depicted in figure A.3.

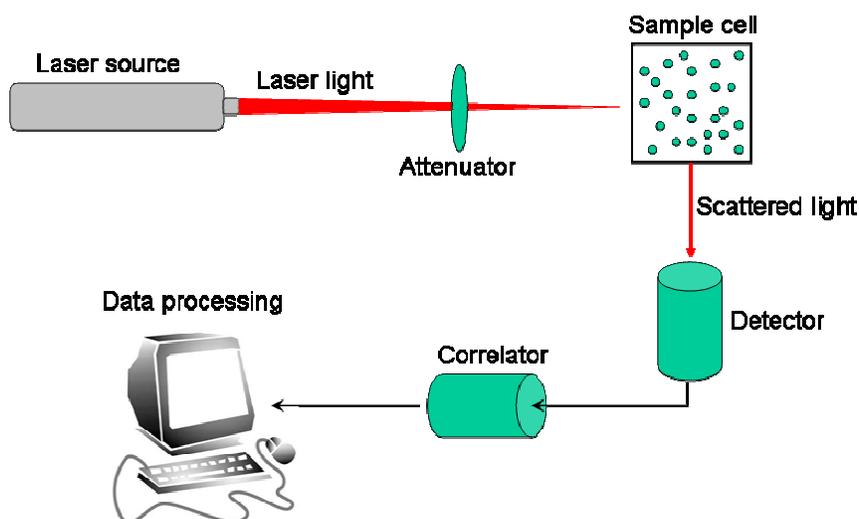


Figure A.3: schematic view of Dynamic Light Scattering.

Laser light hits the particles in the sample cell and is scattered in all directions. A detector is placed in such a way that it detects the light scattered by the sample with an angle of 90° . The intensity of the scattered light is then used by the correlator to construct a correlation function that is processed and analyzed with a computer software.

A.3 Isothermal Titration Calorimetry (ITC)

Isothermal Titration calorimetry (Pierce et al., 1999) is a widely used technique to test, measure and quantify chemical or enzymatic reactions, as well as protein-protein, protein-DNA and protein/ligand interactions. It is based on the thermodynamic changes gone when two entities bind reciprocally. The thermodynamics of association are related to the stoichiometry of the interaction (n), the association constant (K_a), the free binding energy variation (ΔG_b), the binding enthalpy (ΔH_b), the binding entropy (ΔS_b) and the heat capacity

of binding (ΔC_p). The parameters which are experimentally determined are K_a , n and ΔH_b , whilst ΔG_b and ΔS_b are extrapolated from the association constant K_a , with the following formulas:

$$\Delta G_b = -RT \ln K_a \quad \text{and} \quad \Delta G_b = \Delta H_b - T\Delta S_b$$

Where R is the gas constant ($8.314 \text{ JK}^{-1}\text{mol}^{-1}$) and T is the absolute temperature of the system.

The typical ITC instrument consists of two identical cells, made with a highly efficient thermal conducting material (Hastelloy or gold) surrounded by an adiabatic jacket. The macromolecule is placed in the sample cell, whilst water or buffer is placed in the other one, functioning as the reference cell. The temperature in the system is maintained constant (according to the value set up by the user) by two heaters located on both cells, which are activated or de-activated in response to the temperature variation during the experiment.

The ligand is placed inside a microsyringe which performs several injections at a volume and time whose values are set up by the user. Continuous stirring of the syringe maintains the homogeneity of the analyte-ligand solution in the sample cell constant and favours the interaction between the two species.

During the first injections, all or most of the injected ligand is bound to the analyte and this results in large heat variation signals. If the reaction is exothermic, the heat in the sample cell will rise and the connected heater will turn off to keep the temperature constant. As the instrument measures the supplied power difference between the two cell heaters with time, an exothermic reaction is characterized by a negative heat peak, as the sample cell heater is off during that time. The opposite occurs during an endothermic reaction, which corresponds to a positive heat peak measured by the instrument, as the heater will turn on, providing heat to keep the temperature constant. An example of ITC curve is reported in Figure A.4.

As the heat absorbed or produced during the experiment depends on the fraction of bound ligand, it is extremely important to measure with very high precision the concentration of both the analyte and the ligand. Furthermore, it is preferable that both species be solubilized in the same buffer, to avoid dilution effects that could hinder the measurement of the signal.

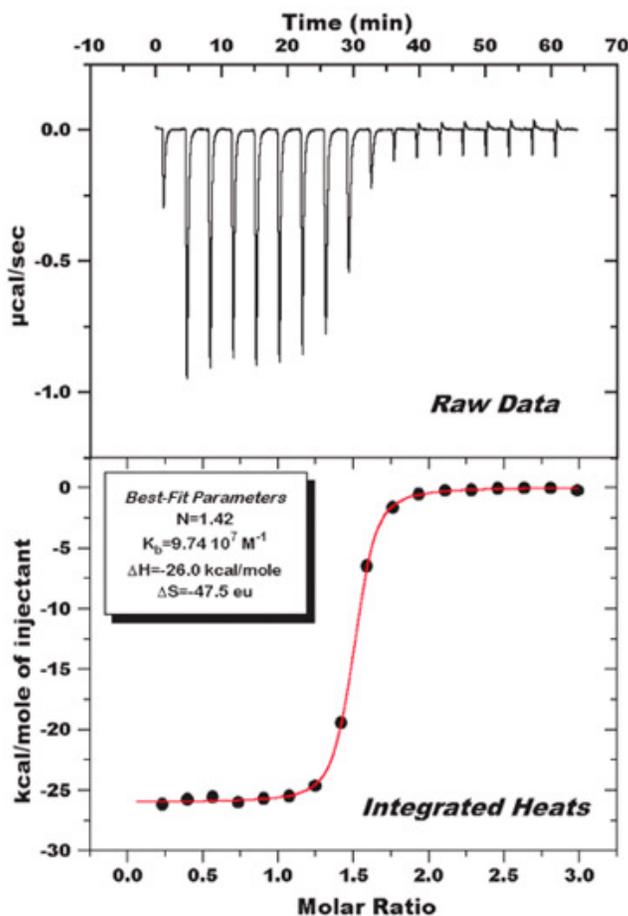


Figure A.4: example of ITC curve.

In the upper panel, the heat variations between the two cells (sample and reference) at different injections and time are depicted as narrow, negative peaks, meaning that the interaction reaction is exothermic. After a certain injection, no more peaks are detected, meaning that all the analyte in the cell is saturated by the injected ligand. In the lower panel, the measured heats (kcal/mole of ligand) are plotted against the molar ratio, i.e. the stoichiometry of interaction. ΔH is measured experimentally, whilst the inflection point of the curve gives the estimated stoichiometry, N . ΔS is derived from ΔH with the formula previously reported.

This data processing is required to obtain the association constant K_a , which is calculated as $1/K_D$. Taken from the Microcal VP-ITC user guide.

A.4 Surface Plasmon Resonance (SPR)

This technique (Torreri, *et al.*, 2005) has been established in 1990 to analyze biomolecular interactions and, since then, has been successfully applied to characterize

protein-protein, protein-antibody, protein-DNA and protein-ligand interactions.

The principle is based on the detection of changes in the refractive index in the proximity (up to 300 nm) of a sensor surface when this is hit by a polarized light at a defined angle. The sensor surface (chip) consists of a small flow cell of 20-60 nL of volume, where a liquid phase (water or buffer) passes under continuous flow (1 to 100 $\mu\text{l}/\text{min}$). One of the interacting molecules, namely the ligand, is injected into the flow cell and is immobilized on the surface of the chip. The second interacting molecule, the analyte, is then injected onto the ligand and, if an interaction occurs, the accumulation of the two molecules at the surface results in a change of the refractive index near the surface of the chip, that is detected and expressed as “Resonance Units” (RUs).

The chip is made of a solid support covered with a thin gold layer. The gold layer is then covered by a matrix that immobilizes the ligand. The matrix used depends on the type of chip and its desired application. One of the most used chip is the CM5, where the gold surface is covered by a layer of CarboxyMethylDextrane that covalently immobilizes the ligand in its native state via its free amine, thiol, aldehyde or carboxyl groups.

Another very used support is the Ni-NTA chip, where a matrix of Nickel-chelating nitrilotriacetic acid is immobilized on the gold layer for the coupling of His-tagged proteins. The real-time progress of an interaction is monitored by the instrument as a sensorgram. An example of sensorgram is reported in Figure A.5.

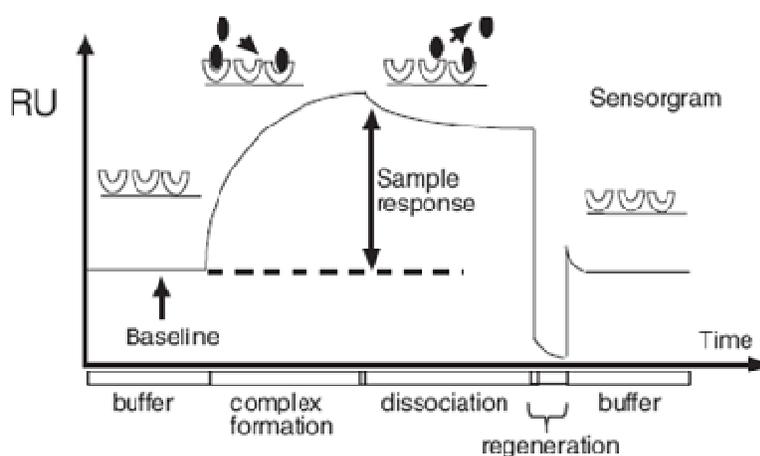


Figure A.5: example of SPR sensorgram (Torreri et al., 2005).

The baseline refers to the signal given by the buffer when the sensor chip is equilibrated. Complex formation is represented by a signal increase when the analyte is injected. When the injection is stopped, the ligand is replaced by a continuous buffer flow and the signal decreases, meaning that the

bound analyte dissociates from the ligand with time. The chip is then regenerated by stripping away all the complex, returning to the baseline levels

A.5 Overview of protein crystallization, structure determination and refinement

A.5.1 Introduction

It is widely accepted in Structural Biology and related fields that the most important feature of a protein is its biological function. Such functions are extremely diverse and comprise catalysis, signal transduction, transport of molecules, architecture of biological systems (muscles, tissues, bones etc...), regulation of other proteins and many more. As it is generally accepted that the biological function of a protein is related to its three-dimensional structure (i.e. its “shape”), it is of the essence to characterize such structures in order to know how different proteins carry out their different physiological roles.

X-ray crystallography is currently the most powerful and effective tool for the elucidation of the structures of biological macromolecules. The first X-ray protein structure determinations were carried in 1956-1959 by Max Perutz and John Kendrew, who solved the crystal structure of the sheep Hemoglobin and horse Myoglobin. They were both awarded the Noble Prize for this outstanding discovery in 1962 (Strandberg, 2009). The field they pioneered, protein crystallography, should be considered as an ensemble of techniques that allow the structural biologist to obtain crystals from purified proteins and, through the X-ray diffraction analysis of such crystals, to solve the three-dimensional structure of the target protein. Since the mid-1980s protein crystallography has evolved rapidly thanks to the advent of powerful computers, advanced computer programs for structure determination and analysis, the increasing use of synchrotron radiation in X-ray diffraction experiments and the availability of robots to rapidly screen and optimise thousands of crystallization conditions. Here, the various steps involved in protein crystallography, from crystallization to structure solving and analysis, will be briefly described.

A.5.2 Protein crystallization

Large biological macromolecules are very often rather difficult to crystallize due to their size and intrinsic flexibility. It is therefore important to screen as many conditions as possible in order to find the one most suitable to provide the formation of protein crystals.

Figure A.6 shows a phase diagram of the process of crystallisation which can be defined as the passage of molecules from a solution state to a solid, highly ordered state. To achieve this, the solution needs first to reach a supersaturation state, in which some of the solubilized macromolecule can pass from solution to a crystalline phase, without denaturation (Figure A.6).

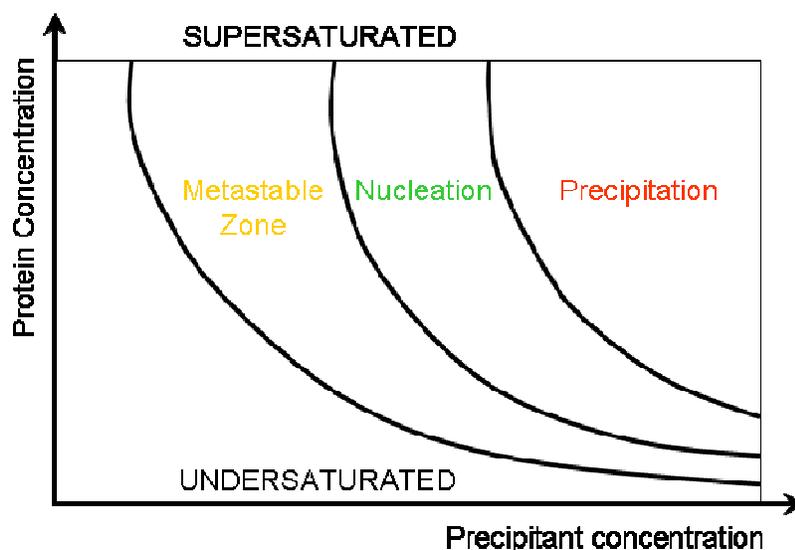


Figure A.6: crystallization phase diagram

A crystallisation phase diagram plotting the concentration of the protein in solution vs. the concentration of the precipitating agent. Under supersaturation conditions three main phases occur: in the metastable phase, nucleation does not occur but growth of existing nuclei will take place; in the nucleation phase the supersaturation process reaches its optimum, and protein crystals may appear and grow; in the precipitation zone, the concentrations of either the protein or the precipitant (or both) are too high and amorphous aggregation/precipitation usually occurs.

The development of a proto-crystalline solid phase is termed as “nucleation” and such process gives rise to discrete and ordered precipitates known as “nuclei”. Crystals can grow over time through the precipitation of more macromolecules onto the pre-formed nuclei. When crystallising a protein, reaching the nucleation state is challenging and depends on the equilibrium between several factors. Among these factors, the protein concentration in

solution plays a very important role, as it influences directly the supersaturation process and nucleation. Other factors can influence the precipitation and nucleation of a protein:

- The presence of precipitating agents such as polymers (Polyethylenglicole, Polyvinilpyrrolidone) or glycerol, which sequester the water that solubilizes the protein, pushing the equilibrium towards the supersaturation state;
- buffering agents, which provide a pH in solution that influences the net charge of the protein and therefore its solubility;
- ionic agents (salts), which have the effect of both sequestering water molecules (and therefore acting as precipitating agents) or neutralizing charges on the protein surface, therefore increasing the protein contacts that may contribute to nucleation;
- the temperature, which directly influences the solubility of the protein and therefore the supersaturation state;
- ligands (nucleotides, substrates, cofactors etc.) that bind to the protein and stabilize its folding, thus increasing its “order” and decreasing its flexibility.

The combination of precipitating agents, salts, buffers and ligands in a single solution is termed “precipitation solution” or, more optimistically, “crystallization condition”.

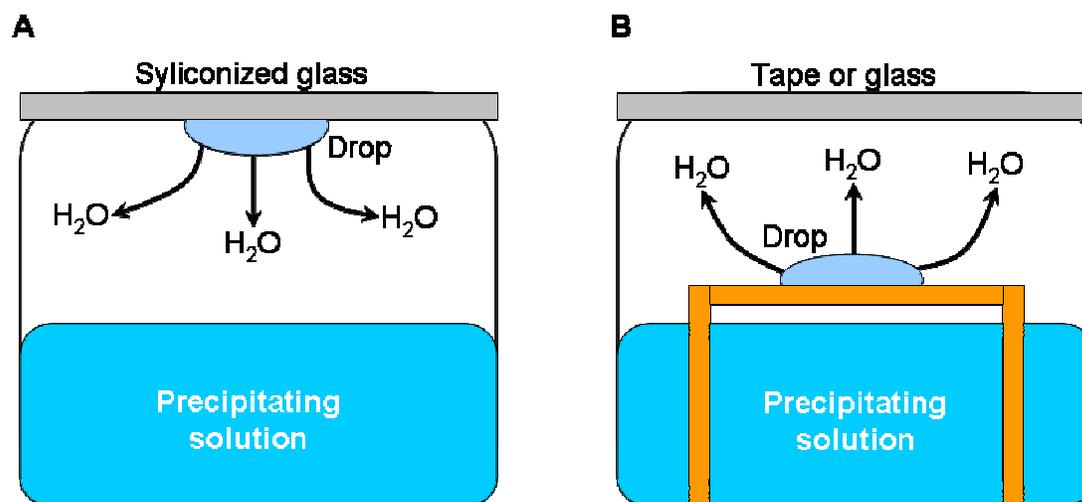


Figure A.7: Schematics of the hanging drop and sitting drop crystallisation techniques. (A): hanging drop. The crystallization drop is placed onto a siliconized glass slide, oriented towards the precipitating solution in the well. The glass and well are then sealed with paraffin or silicone. Water diffuses from the drop with time and concentrates the protein. (B): sitting drop. The crystallization drop is placed onto a bridge inside the well, filled with the precipitating solution. As for hanging drops,

water diffuses out of the drop thus increasing protein concentration. The system is sealed with tape or with a glass.

In protein crystallography several techniques are used to reach the supersaturation/nucleation phase and grow protein crystals. The most commonly used is the “vapor diffusion method”, where a few microliters of a protein solution at a defined concentration are mixed with an equal volume of precipitation solution (a mix termed “crystallization drop”) and placed in an isolated system. In this system, a larger volume of pure precipitation solution (0.5-1 ml) causes the diffusion of the water from the crystallization drop, which subsequently causes an increase of the protein concentration that can lead to the supersaturation/nucleation phase. This water diffusion method is used in two similar techniques, called “hanging drop” and “sitting drop” (Figure A.7). In the hanging drop method the crystallization drop is deposited on a siliconized glass slide, which is used to cover a 2 ml well filled with 0.5-1 ml of the pure precipitation solution. The well and glass are sealed with paraffin or silicone to prevent the drop drying out. In the sitting drop method, the crystallization drop is placed onto a support (“bridge”) inside the well. 0.5-1 ml of precipitation solution is added to the well which is then sealed with tape or glass.

A.5.3 X-ray diffraction

A crystal is defined as a highly ordered solid characterized by a regular lattice, described by the repetition in the three dimensions of an element called “unit cell”. The unit cell is described by three vectors, a , b and c , which are oriented one to another by three angles, α , β and γ .

When a crystal is irradiated with X-rays, the electrons inside the lattice interact with the incoming radiation and scatter it with a reflection angle 2θ , a phenomenon which is termed “scattering”. The mathematical and geometrical basis of X-ray diffraction from crystals was described in the beginning of the 20th century by Max Theodor Felix von Laue, William Lawrence Bragg and Paul Peter Ewald. Laue won the Nobel Prize in 1914 for discovery the phenomenon of X-ray diffraction, whilst Bragg formulated the first mathematical law (Bragg’s law) that describes the diffraction of X-rays by crystals according to the reflection angle (2θ), the distance between the planes of the lattice (d) and the wavelength of the incoming radiation (λ):

$$n\lambda = 2d\sin\theta$$

Ewald described a mathematical model of diffraction by a single crystal in terms of a reciprocal lattice (i.e. the distance between planes is $1/d$ rather than d). In Ewald's model (Figure A.8), the incident X-ray hits the crystal and is scattered with an angle 2θ . The amplitudes of both the incident and the scattered beams are the same ($1/\lambda$). The difference between the vectors generated by the incident beam and the scattered beam describes the scattering vector, whose amplitude, by the Bragg's law, is $1/d$. As the incident beams can be scattered in every direction in the three dimensions, the tips of the vectors of all scattered rays lay on the surface of a sphere (the "Ewald sphere") of radius $1/\lambda$. The origin of the reciprocal lattice (coinciding with the origin of the scattering vector) lays on the point where the incident X-ray beam would exit the sphere if not scattered.

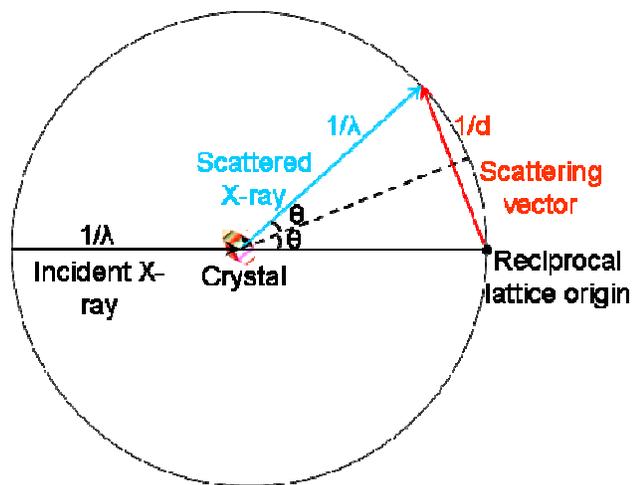


Figure A.8: the Ewald's sphere.

The crystal is placed in the centre of the sphere and diffracts the incident x-rays with an angle 2θ (scattered x-ray is shown in blue). The reciprocal space vector is colored in red and is described by the difference between the vector of the incident ray and the vector of the scattered ray. The origin of the scattering vector lays on the point of the sphere where the incident x-ray would exit the sphere unscattered. The Bragg's planes can diffract only if they are perpendicular to the direction of the scattered beam, so it is important to rotate the crystal.

From this construction, it is possible to understand that if a set of Bragg's planes is oriented perpendicularly to the scattered X-ray, the scattering vector will end on the surface of

the Ewald's sphere, whilst if the Bragg's planes are not in the right orientation, such vector will not be on the Ewald's sphere. Hence, it is necessary, in order to have diffraction from all Bragg's planes in the crystal, to rotate the crystal (and conversely the reciprocal space) to put all the planes in the right orientation to diffract.

A.5.4 The phase problem in structure solution

Every reflection from a plane hkl in a crystal is characterized by a parameter, called the "Structure Factor", F_{hkl} . To calculate the electron density distribution in the unit cell of a crystal and hence 'solve' the structure all that needs to be done is to carry out the following calculation:

$$\rho(x, y, z) = 1/V \sum_{hkl} F_{hkl} \exp(-2\pi i hx + ky + lz) = 1/V \sum_{hkl} |F_{hkl}| \exp(i\alpha_{hkl}) \exp(-2\pi i hx + ky + lz)$$

As can be seen from the equation above, the structure factor F_{hkl} consists of two components, an amplitude $|F_{hkl}|$ and a phase α_{hkl} . The structure factor is therefore a complex number, composed by a real component (the amplitude) and an imaginary component (the phase). In a diffraction experiment, the intensity I_{hkl} for a given reflection is related to the number of electrons on the plane from which the reflection occurs. The amplitudes of the structure factors can be derived from the diffraction experiment, as they are proportional to the square root of the intensities of the reflections measured on the detector:

$$|F_{hkl}| \propto \sqrt{I_{hkl}}$$

To retrieve the structure factors, correction factors are applied to the intensities, which are reported in the following formula:

$$I_{hkl} = |F_{hkl}|^2 LP \times A$$

Where LP is a combined geometry and polarization factor that depends on the type of experiment, A is an absorption correction

However, the phases for the structure factors cannot be measured directly and, in protein crystallography, we must use one of several methods used to obtain such phases in order to be able to calculate the electron density distribution:

A.5.4.1 Molecular Replacement

This method derives new phase information from already known phases, i.e. those of an already solved structure. This method is very useful when the target protein is structurally similar (homologous) to another protein whose structure is known. Usually, the molecular replacement method is successful when the sequence identity between the two proteins is around 25% and the root mean square deviation (rmsd) of the C α positions between the known and the unknown structure is less than 2.0 Å (Taylor, 2003).

Molecular Replacement (MR) is based on the use of the the Patterson function (P) which produces a map of interatomic vectors (uvw). Two structurally similar macromolecules have similar Patterson functions.

$$P_{(u,v,w)} = \frac{1}{V_c} \sum_h \sum_k \sum_l |F_{hkl}|^2 \exp(-2\pi i(hu + kv + lw))$$

To obtain the new structure from a pre-existing model, it is necessary to have two Patterson functions: one calculated (from the atomic positions of the structural model) and one observed (from the diffraction data of the crystallized new protein). To solve a structure using MR: the model patterson is first rotated and superposed on the observed patterson to obtain the correct orientation of the molecule. This correct orientation is that which produces the best superposition between model and observed patterson functions. The correctly oriented patterson is then translated relative to the origin of the new structure's unit cell to correctly position the model structure in the new unit cell. The MR method is exemplified in Figure A.9:

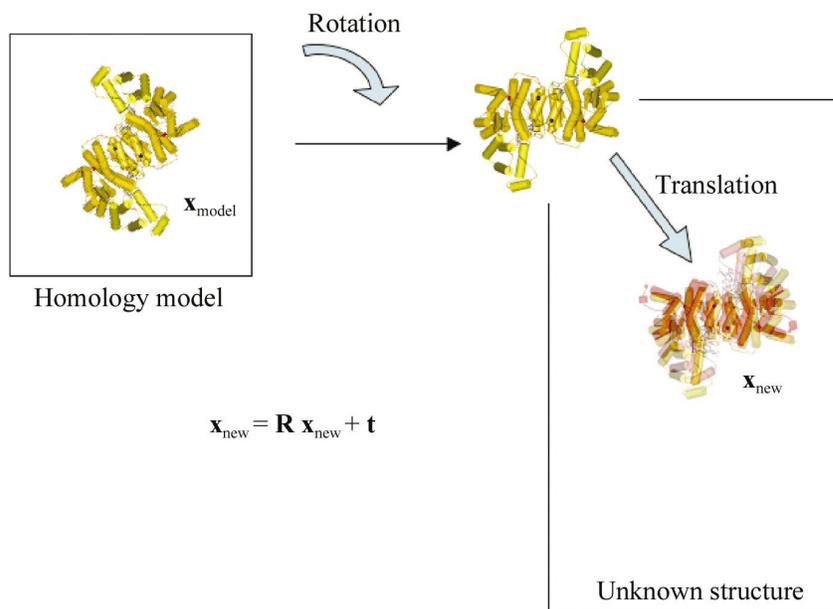


Figure A.9: Schematic view of Molecular Replacement

The structure of a homology model (x_{model}) is used as a template to solve the structure of a new, uncharacterized protein (x_{new}). According to the Patterson method, the model is first rotated and correctly oriented in the unit cell of the new protein, then translated to fit the similar structural element according to the origin of the unit cell. \mathbf{R} and \mathbf{t} are respectively the rotation and the translation required to correctly position the model in the new unit cell.

Molecular replacement is performed using a set of programs, like AmoRe (Trapani & Navaza, 2008), MolRep (Lebedev, *et al.*, 2008) and Phaser (McCoy, *et al.*, 2007).

A.5.4.2 Isomorphous replacement

With the isomorphous replacement, a native, preformed crystal is “modified” by diffusing a compound containing a heavy atom (Mercury, Platinum, Cadmium, Lead, Gold etc.) inside (Taylor, 2003). This process is often called “soaking”, as the crystal is literally placed inside a crystallization solution with a given concentration (usually from 0.1 to 10 mM) of the heavy atom.

An isomorphous replacement experiment consists in collecting a full dataset of the native crystals, then a full dataset of the same crystals soaked in heavy atoms. The structure factors amplitudes of the heavy atoms ($|F_H|$) are calculated from the approximated isomorphous

difference between the amplitudes of the reflections measured for the derivative crystal ($|F_{PH}|$) and the amplitudes of the reflections measured for the native crystal ($|F_P|$):

$$|F_H| \cong |F_{PH}| - |F_P| .$$

This is best represented in the following Argand diagram (Figure A.10):

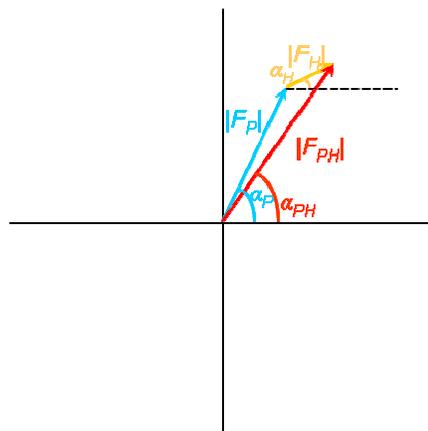


Figure A.10: Argand diagram of IR.

$|F_P|$ is the amplitude of the reflection measured for the native protein, $|F_H|$ the amplitude measured for the heavy atom and $|F_{PH}|$ the amplitude measured for the derivative crystal.

The positions of the heavy atoms are then found using the Patterson function.

The phases of the native protein can be calculated from the following formula:

$$\alpha_P = \alpha_H + \cos^{-1} [(F_{PH}^2 - F_P^2 - F_H^2)/2F_P F_H]$$

where α_P and α_H are the phases of respectively the native protein and the heavy atom and F_P , F_H and F_{PH} are the amplitudes of the reflections of the protein, the heavy atom and the isomorphous derivative, respectively. As the function \cos^{-1} gives two solutions (angles θ and $-\theta$), the phases derived for the native protein are characterized by an ambiguity, which is “corrected” by using more than one type of heavy atom to derivatize the crystal (a method called Multiple Isomorphous Replacement).

A.5.4.3 Single and Multiple Wavelength Anomalous Dispersion (SAD/MAD)

These methods rely on the property of certain atoms to give an “anomalous scattering” when irradiated with X-rays at a given wavelength. The scattering factor f is composed by three different entities: a normal scattering term, f_0 dependent on the diffraction angle (2θ), an two terms dependent on the wavelength: $f'(\lambda)$, or dispersive term, and $f''(\lambda)$, or absorption term.

$$f = f_0(\sin \theta / \lambda) + f'(\lambda) + if''(\lambda)$$

The dispersive term is in phase with f_0 . The absorption term, though, is an imaginary component and its phase is advanced of 90° compared to the normal scattering term. This causes a breakdown in Friedel’s law with the result that

$$F_{hkl} \neq F_{-h-k-l}$$

Or, as exemplified by the Argand diagram (Figure A.11):

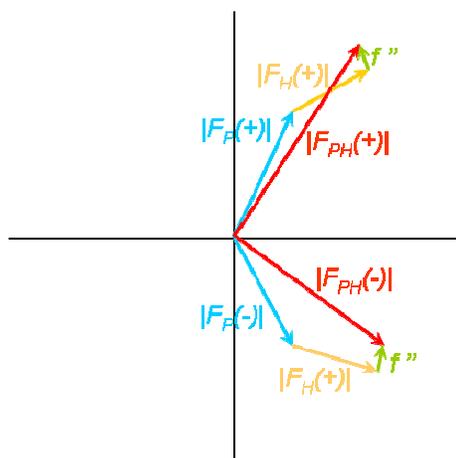


Figure A.11: Argand diagram for SAD.

As in Figure 4.5, $|F_P|$ is the amplitude of the reflection measured for the native protein, $|F_H|$ the amplitude measured for the heavy atom and $|F_{PH}|$ the amplitude measured for the derivative crystal. In

these diagrams, the amplitudes related to the Friedel pairs (\mp) are shown. When an anomalous scatterer is present, the amplitudes of the reflections of the derivative crystal $|F_{PH}|$ do not respect the Friedel law and $|F_{PH(+)}| \neq |F_{PH(-)}|$.

This generates an anomalous difference,

$$\Delta|F|_{ano} = [|F_{PH(+)}| - |F_{PH(-)}|] \frac{f'}{2f''}$$

that can be measured and used to retrieve the positions of the anomalous scatterers and their phases, which are subsequently used to approximate the phases of all the other atoms in the crystal.

One of the most used anomalous scatterers in Macromolecular Crystallography is Selenium, which is characterized by an absorption edge at a wavelength of 0.9795 Å (Figure A.12). Such wavelength is routinely accessible in many crystallographic beamlines and this, combined with the possibility to derivatize a recombinant protein by substituting the sulfur atoms of its methionines with a Seleno-Methionine, makes the selenium a “user-friendly” atom to do phasing with (Pappa, *et al.*, 1996).

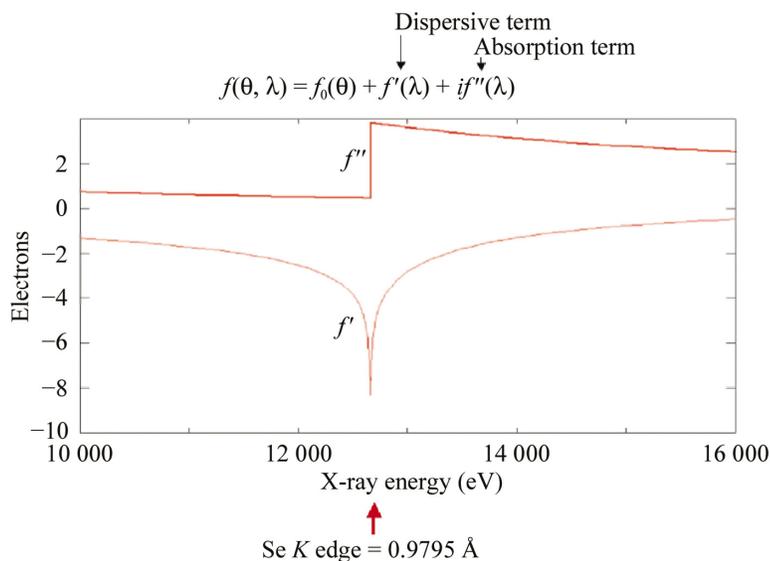


Figure A.12: variation in anomalous scattering around the K-edge of Selenium.

In the SAD method, we measure the anomalous signal at one wavelength only, that is the one corresponding to the peak of the absorption edge. In the MAD method, to maximize the absorption and dispersive effects we measure diffraction data at three wavelengths, corresponding to the peak (maximum f''), the inflection point (maximum f') and the remote

point. This is equivalent to having two derivative datasets and one native to solve the phase problem (Figure A.13):

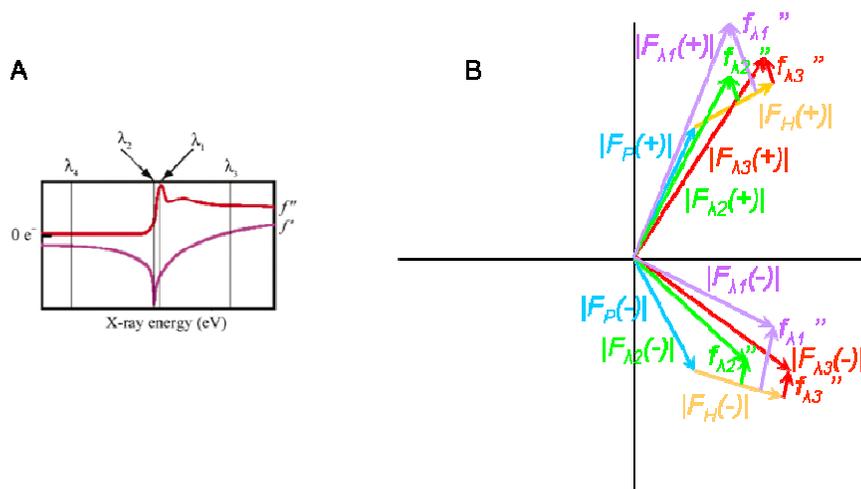


Figure A.13: MAD phasing setup.

(A): Absorption edge of an anomalous scatterer. λ_1 corresponds to the absorption edge (0.9795 Å for Selenium); λ_2 corresponds to the inflection point; λ_3 and λ_4 are the remote wavelengths. (B): Argand diagram of the amplitudes of the reflections measured at the three wavelengths (λ_1 , λ_2 and λ_3).

Several programs are used to perform phasing by SAD or MAD, like ShelXD/ShelXE (Sheldrick, 2008), Solve/Resolve (Terwilliger, 2004) and Sharp (Vonrhein, *et al.*, 2007).

A.5.5 Phase Improvement

As the phases, from both the anomalous scattering and isomorphous replacement methods are determined from experimental data, they are effected by experimental error and are often not accurate enough to yield a good quality electron density map. Therefore, such phases need to be “refined” and improved by a process called “density modification”.

Density modification uses three different approaches:

- a) solvent flattening, where the negative values of the calculated electron density are removed and the electron density value corresponding to the solvent regions are normalized to a value of to 0.33 electrons per cubic Angstrom ($0.33 \text{ e } \text{Å}^{-3}$);

- b) histogram matching, where the calculated electron density is “altered” to match an ideal distribution of electron density, which is dependent on the resolution and temperature factors (B), but not on the structure itself;
- c) non-crystallographic symmetry, which imposes the equivalence of electron density maps of the same molecules present in multiple copies inside the asymmetric unit.

Density modification is a cyclic process, where the modified electron density maps are transformed to give new phases, the new phases are recombined with the experimental phases to limit bias to calculate a new electron density map. The process is reiterated up to convergence.

Several programs are used to perform density modification: DM (Cowtan & Zhang, 1999), Resolve (Terwilliger, 2002), CNS ((Brunger, *et al.*, 1998).

A.5.6 Model building and refinement

Once an electron density map is obtained, it is possible to create a “model” of the three-dimensional structure of the target protein. This is performed by fitting each single residue (if possible) of the protein inside the electron density map. Programs that allow the automatic creation of such models are Arp/Warp (Langer, *et al.*, 2008) or Phenix (Zwart, *et al.*, 2008).

After a first model is built, it must be “refined”, which means to minimize the difference between the measured amplitudes F_o and the calculated amplitudes F_c of the reflections ($F_o - F_c$). In the most used model refinement programs, the “maximum likelihood” approach is routinely applied. This method is based on the assumption that the model must be as much consistent as possible with the experimental observations. Such consistency is the function of the probability that the observations that we can make in the experiment are the same we could make given the model. Therefore, refinement programs iteratively introduce systematic variations in the model to make such observations more probable, and therefore increase the likelihood of the model with each step. Variations introduced comprise the bond distances between atoms, the torsion and bond angles, the chiral volumes, the distances of non-bonded atoms (like water molecules coordinated by certain residues).

Several programs are used for structural refinement. The most common ones are Refmac5 (Murshudov, *et al.*, 1997) or Phenix (Zwart, *et al.*, 2008). To visualize and correct

the structural model of the protein, Coot (Emsley & Cowtan, 2004) is one of the most used and “user-friendly” programs.

A.5.7 Data collection

To obtain a diffraction dataset from a protein crystal, it has to be treated and manipulated in such a way that it won't be destroyed in the process.

An important step is to freeze the crystal in liquid nitrogen (-173°C) in the presence of a cryo-protectant, which is generally made up of the crystallization condition supplemented with glycerol, ethylene glycole, small molecular weight Polyethylenglycols (MW 200, 400) or salts at high concentrations ($>1\text{M}$ malonic acid, ammonium sulphate etc.). Such a freezing process inhibits the intrinsic decay of the crystal when it is not in its mother liquor (i.e. the crystallization condition where the crystal appeared and grew) and attenuates the effects of radiation damage (Garman & Owen, 2007).

Frozen crystals are placed onto a support, in the presence of a stream of liquid nitrogen, which is called “goniometer head”. This support rotates the crystal around an axis (Φ) which is perpendicular to the incident X-rays beam. Diffracted beams are measured on a detector placed in their path. Several detectors can be used (ImagePlate, Charged Coupled Device, Pixel detectors). A typical data collection setup is represented in Figure A.14.

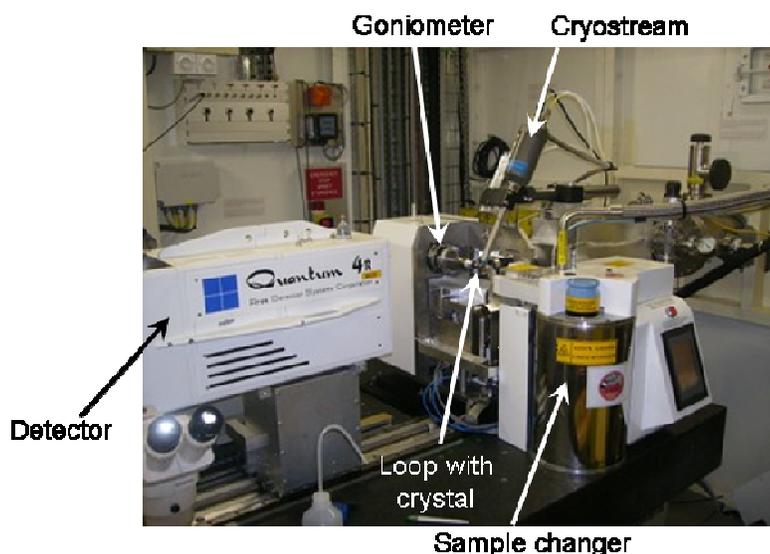


Figure A.14: crystal diffraction experiment at a ESRF beamline

This Figure represents the various instruments at the macromolecular crystallography beamline ID14-2 at the ESRF.

The large unit cells associated with crystals of proteins means that reciprocal lattice points, and thus the reflections they produce, are close to each other. To avoid the overlap of reflections on the detector all the data needed for structure solution cannot be collected in one single rotation of the crystal. therefore, data collection is subdivided into a series of rotations known as oscillations.

It is important, before a data collection, to find the crystal Bravais Lattice and cell parameters. This process is called indexing During autoindexing, several solutions of different lattices are given, with a score that represents the “probability” that a particular lattice is the correct one.

For autoindexing, a few frames are collected at different Φ angles and the resulting images are analyzed with different programs, such as Mosflm (Powell, 1999), Denzo (Rossmann & van Beek, 1999) or XDS (Diederichs & Karplus, 1997).

These programs can also estimate a strategy to optimize the data collection, in order to obtain the maximum completeness of data with the minimum number of collected frames. Once a full dataset is obtained, the intensities of the reflections are integrated (by Mosflm or XDS), scaled and merged to give averaged intensities using different programs, like Scala (Evans, 2006). Scala is a very useful program to monitor several parameters which are indicative of the quality of the collected data, such as the merging R-factor (R_{merge}), which is the sum of the differences of all measurements from the average value of the measurement divided by the sum of all measurements, and the $I/\sigma I$, which is the error dependent on the experimental measurement of the intensities of the reflections. When processing a SAD/MAD dataset, Scala also gives the opportunity to keep the anomalous intensities I_{hkl} and I_{h-k-l} separate, for a better scaling of the data. The program Truncate (French and Wilson, 1978), is used to derive structure factor amplitudes from the reflection intensities and (when appropriate) derive anomalous differences.

A.5.8 Macromolecular crystallography and synchrotron radiation

Synchrotron radiation is produced when electrons are forced to travel at relativistic speed (close to the speed of light) in a curved path, such as a synchrotron. This particular radiation was observed for the first time at the General Electric Research Laboratory synchrotron in Schenectady, New York, in 1946.

As the synchrotrons were not optimized, at that time, for synchrotron radiation, this phenomenon was seen at first as a loss of the energy used to perform high-energy physics studies. With time, the use of synchrotron radiation gained interest, and dedicated synchrotrons (2nd generation sources) were built and optimized to use synchrotron radiation in biology, chemistry and physics.

Synchrotron radiation is characterized by the following properties:

- a) high brilliance, collimation and brightness;
- b) high flux;
- c) possibility to tune it over a wide wavelength range (from X-rays to infra-red);

These properties are particularly useful in the field of macromolecular crystallography (Halliwell, 1992). Protein crystals are, in fact, weak x-rays scatterers, as they are composed mainly by light atoms (carbon, nitrogen, oxygen, hydrogen, sulfur) and they are rich in aqueous solvent, that furtherly reduces the scattering. In a 2nd generation synchrotron with low intensity and brilliance, relatively long exposure times are required to collect data. The increased brilliance, intensity and collimation of 3rd generation synchrotron sources such as the ESRF allow to reduce the time of exposure of the crystals, generally increasing the resolution of a dataset. The high collimation of the beam also allows the user to obtain diffraction datasets from microcrystals (<50 μm) which are difficult to collect using the less collimated beams available at 2nd generation sources. However, the increased intensity of 3rd generation sources means that radiation damage is an increasing problem in macromolecular crystallography.

This information was extracted from the book “Macromolecular Crystallography with Synchrotron Radiation”, by John R. Halliwell (1992).

Bibliography

- Abendroth J, Bagdasarian M, Sandkvist M & Hol WG (2004) The structure of the cytoplasmic domain of EpsL, an inner membrane component of the type II secretion system of *Vibrio cholerae*: an unusual member of the actin-like ATPase superfamily. *J Mol Biol* **344**: 619-633.
- Abendroth J, Rice AE, McLuskey K, Bagdasarian M & Hol WG (2004) The crystal structure of the periplasmic domain of the type II secretion system protein EpsM from *Vibrio cholerae*: the simplest version of the ferredoxin fold. *J Mol Biol* **338**: 585-596.
- Abendroth J, Mitchell DD, Korotkov KV, Johnson TL, Kreger A, Sandkvist M & Hol WG (2009) The three-dimensional structure of the cytoplasmic domains of EpsF from the type 2 secretion system of *Vibrio cholerae*. *J Struct Biol* **166**: 303-315.
- Ackermann N, Tiller M, Anding G, Roggenkamp A & Heesemann J (2008) Contribution of trimeric autotransporter C-terminal domains of oligomeric coiled-coil adhesin (Oca) family members YadA, UspA1, EibA, and Hia to translocation of the YadA passenger domain and virulence of *Yersinia enterocolitica*. *J Bacteriol* **190**: 5031-5043.
- Adams PD, Grosse-Kunstleve RW, Hung LW, *et al.* (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr D Biol Crystallogr*. **58**: 1948-1954. Epub 2002 Oct 19.
- Aly KA & Baron C (2007) The VirB5 protein localizes to the T-pilus tips in *Agrobacterium tumefaciens*. *Microbiology* **153**: 3766-3775.
- Amedei A, Cappon A, Codolo G, *et al.* (2006) The neutrophil-activating protein of *Helicobacter pylori* promotes Th1 immune responses. *J Clin Invest* **116**: 1092-1101.
- Amieva MR & El-Omar EM (2008) Host-bacterial interactions in *Helicobacter pylori* infection. *Gastroenterology* **134**: 306-323.
- Angelini A, Cendron L, Goncalves S, Zanotti G & Terradot L (2008) Structural and enzymatic characterization of HP0496, a YbgC thioesterase from *Helicobacter pylori*. *Proteins* **72**: 1212-1221.
- Angelini A, Tosi T, Mas P, Acajjaoui S, Zanotti G, Terradot L & Hart DJ (2009) Expression of *Helicobacter pylori* CagA domains by library-based construct screening. *FEBS J* **276**: 816-824.
- Aoki H, Wu H, Nakano T, *et al.* (2009) Nanotransportation system for cholera toxin in *Vibrio cholerae* 01. *Med Mol Morphol* **42**: 40-46.
- Arbibe L, Kim DW, Batsche E, *et al.* (2007) An injected bacterial effector targets chromatin access for transcription factor NF-kappaB to alter transcription of host genes involved in immune responses. *Nat Immunol* **8**: 47-56.

- Arcus VL, Proft T, Sigrell JA, Baker HM, Fraser JD & Baker EN (2000) Conservation and variation in superantigen structure and activity highlighted by the three-dimensional structures of two new superantigens from *Streptococcus pyogenes*. *J Mol Biol* **299**: 157-168.
- Aroian R & van der Goot FG (2007) Pore-forming toxins and cellular non-immune defenses (CNIDs). *Curr Opin Microbiol* **10**: 57-61.
- Arpigny JL & Jaeger KE (1999) Bacterial lipolytic enzymes: classification and properties. *Biochem J* **343 Pt 1**: 177-183.
- Aspholm-Hurtig M, Dailide G, Lahmann M, *et al.* (2004) Functional adaptation of BabA, the *H. pylori* ABO blood group antigen binding adhesin. *Science* **305**: 519-522.
- Aspholm M, Olfat FO, Norden J, *et al.* (2006) SabA is the *H. pylori* hemagglutinin and is polymorphic in binding to sialylated glycans. *PLoS Pathog* **2**: e110.
- Atmakuri K, Cascales E & Christie PJ (2004) Energetic components VirD4, VirB11 and VirB4 mediate early DNA transfer reactions required for bacterial type IV secretion. *Mol Microbiol* **54**: 1199-1211.
- Backert S & Selbach M (2005) Tyrosine-phosphorylated bacterial effector proteins: the enemies within. *Trends Microbiol* **13**: 476-484.
- Backert S & Meyer TF (2006) Type IV secretion systems and their effectors in bacterial pathogenesis. *Curr Opin Microbiol* **9**: 207-217.
- Backert S, Fronzes R & Waksman G (2008) VirB2 and VirB5 proteins: specialized adhesins in bacterial type-IV secretion systems? *Trends Microbiol* **16**: 409-413.
- Bailey S, Ward D, Middleton R, Grossmann JG & Zambryski PC (2006) *Agrobacterium tumefaciens* VirB8 structure reveals potential protein-protein interaction sites. *Proc Natl Acad Sci U S A* **103**: 2582-2587.
- Baker NA, Sept D, Joseph S, Holst MJ & McCammon JA (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A* **98**: 10037-10041.
- Baron C (2006) VirB8: a conserved type IV secretion system assembly factor and drug target. *Biochem Cell Biol* **84**: 890-899.
- Barras F & Marinus MG (1989) The great GATC: DNA methylation in *E. coli*. *Trends Genet* **5**: 139-143.
- Bayliss R, Harris R, Coutte L, *et al.* (2007) NMR structure of a complex between the VirB9/VirB7 interaction domains of the pKM101 type IV secretion system. *Proc Natl Acad Sci U S A* **104**: 1673-1678.
- Benabdelhak H, Schmitt L, Horn C, Jumel K, Blight MA & Holland IB (2005) Positive co-operative activity and dimerization of the isolated ABC ATPase domain of HlyB from *Escherichia coli*. *Biochem J* **386**: 489-495.

- Bereswill S, Fassbinder F, Volzing C, Covacci A, Haas R & Kist M (1998) Hemolytic properties and riboflavin synthesis of *Helicobacter pylori*: cloning and functional characterization of the *ribA* gene encoding GTP-cyclohydrolase II that confers hemolytic activity to *Escherichia coli*. *Med Microbiol Immunol* **186**: 177-187.
- Bernstein HD (2007) Are bacterial 'autotransporters' really transporters? *Trends Microbiol* **15**: 441-447.
- Betts HJ, Wolf K & Fields KA (2009) Effector protein modulation of host cells: examples in the *Chlamydia spp.* arsenal. *Curr Opin Microbiol* **12**: 81-87.
- Binet R & Wandersman C (1995) Protein secretion by hybrid bacterial ABC-transporters: specific functions of the membrane ATPase and the membrane fusion protein. *EMBO J* **14**: 2298-2306.
- Bladergroen MR, Badelt K & Spaank HP (2003) Infection-blocking genes of a symbiotic *Rhizobium leguminosarum* strain that are involved in temperature-dependent protein secretion. *Mol Plant Microbe Interact* **16**: 53-64.
- Bommarius B, Maxwell D, Swimm A, Leung S, Corbett A, Bornmann W & Kalman D (2007) Enteropathogenic *Escherichia coli* Tir is an SH2/3 ligand that recruits and activates tyrosine kinases required for pedestal formation. *Mol Microbiol* **63**: 1748-1768.
- Bornscheuer UT & Pohl M (2001) Improved biocatalysts by directed evolution and rational protein design. *Curr Opin Chem Biol* **5**: 137-143.
- Botham CM, Wandler AM & Guillemin K (2008) A transgenic *Drosophila* model demonstrates that the *Helicobacter pylori* CagA protein functions as a eukaryotic Gab adaptor. *PLoS Pathog* **4**: e1000064.
- Brandt S, Shafikhani S, Balachandran P, *et al.* (2007) Use of a novel coinfection system reveals a role for Rac1, H-Ras, and CrkII phosphorylation in *Helicobacter pylori*-induced host cell actin cytoskeletal rearrangements. *FEMS Immunol Med Microbiol* **50**: 190-205.
- Bruckner A, Polge C, Lentze N, Auerbach D & Schlattner U (2009) Yeast two-hybrid, a powerful tool for systems biology. *Int J Mol Sci* **10**: 2763-2788.
- Brunger AT, Adams PD, Clore GM, *et al.* (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **54**: 905-921.
- Burley SK (2000) An overview of structural genomics. *Nat Struct Biol* **7 Suppl**: 932-934.
- Burstein D, Zusman T, Degtyar E, Viner R, Segal G & Pupko T (2009) Genome-scale identification of *Legionella pneumophila* effectors using a machine learning approach. *PLoS Pathog* **5**: e1000508.
- Bury-Mone S, Thiberge JM, Contreras M, Maitournam A, Labigne A & De Reuse H (2004) Responsiveness to acidity via metal ion regulators mediates virulence in the gastric pathogen *Helicobacter pylori*. *Mol Microbiol* **53**: 623-638.

- Campellone KG, Giese A, Tipper DJ & Leong JM (2002) A tyrosine-phosphorylated 12-amino-acid sequence of enteropathogenic *Escherichia coli* Tir binds the host adaptor protein Nck and is required for Nck localization to actin pedestals. *Mol Microbiol* **43**: 1227-1241.
- Cascales E & Christie PJ (2004) Definition of a bacterial type IV secretion pathway for a DNA substrate. *Science* **304**: 1170-1173.
- Cascales E, Bernadac A, Gavioli M, Lazzaroni JC & Lloubes R (2002) Pal lipoprotein of *Escherichia coli* plays a major role in outer membrane integrity. *J Bacteriol* **184**: 754-759.
- Cendron L, Seydel A, Angelini A, Battistutta R & Zanotti G (2004) Crystal structure of CagZ, a protein from the *Helicobacter pylori* pathogenicity island that encodes for a type IV secretion system. *J Mol Biol* **340**: 881-889.
- Cendron L, Couturier M, Angelini A, Barison N, Stein M & Zanotti G (2009) The *Helicobacter pylori* CagD (HP0545, Cag24) protein is essential for CagA translocation and maximal induction of interleukin-8 secretion. *J Mol Biol* **386**: 204-217.
- Cendron L, Tasca E, Seraglio T, *et al.* (2007) The crystal structure of CagS from the *Helicobacter pylori* pathogenicity island. *Proteins* **69**: 440-443.
- Censini S, Lange C, Xiang Z, *et al.* (1996) cag, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. *Proc Natl Acad Sci U S A* **93**: 14648-14653.
- Cheng P, Shi RH, Zhang HJ, Yu LZ, Zhang GX & Hao B (2008) [Effects of tumor necrosis factor-alpha inducing protein-alpha secreted by *Helicobacter pylori* on human gastric epithelial cells]. *Zhonghua Yi Xue Za Zhi* **88**: 1528-1532.
- Christie PJ, Atmakuri K, Krishnamoorthy V, Jakubowski S & Cascales E (2005) Biogenesis, architecture, and function of bacterial type IV secretion systems. *Annu Rev Microbiol* **59**: 451-485.
- Churin Y, Kardalidou E, Meyer TF & Naumann M (2001) Pathogenicity island-dependent activation of Rho GTPases Rac1 and Cdc42 in *Helicobacter pylori* infection. *Mol Microbiol* **40**: 815-823.
- Coin D, Louis D, Bernillon J, Guinand M & Wallach J (1997) LasA, alkaline protease and elastase in clinical strains of *Pseudomonas aeruginosa*: quantification by immunochemical methods. *FEMS Immunol Med Microbiol* **18**: 175-184.
- Collaborative Computational Project-4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr.* **D50**: 760-763.
- Cornelis GR (2006) The type III secretion injectisome. *Nat Rev Microbiol* **4**: 811-825.
- Covacci A & Rappuoli R (2000) Tyrosine-phosphorylated bacterial proteins: Trojan horses for the host cell. *J Exp Med* **191**: 587-592.

- Cover TL, Dooley CP & Blaser MJ (1990) Characterization of and human serologic response to proteins in *Helicobacter pylori* broth culture supernatants with vacuolizing cytotoxin activity. *Infect Immun* **58**: 603-610.
- Cowell BA, Twining SS, Hobden JA, Kwong MS & Fleiszig SM (2003) Mutation of lasA and lasB reduces *Pseudomonas aeruginosa* invasion of epithelial cells. *Microbiology* **149**: 2291-2299.
- Cowtan KD & Zhang KY (1999) Density modification for macromolecular phase improvement. *Prog Biophys Mol Biol* **72**: 245-270.
- Croxen MA, Sisson G, Melano R & Hoffman PS (2006) The *Helicobacter pylori* chemotaxis receptor TlpB (HP0103) is required for pH taxis and for colonization of the gastric mucosa. *J Bacteriol* **188**: 2656-2665.
- Czajkowsky DM, Iwamoto H, Szabo G, Cover TL & Shao Z (2005) Mimicry of a host anion channel by a *Helicobacter pylori* pore-forming toxin. *Biophys J* **89**: 3093-3101.
- D'Elis MM, Montecucco C & de Bernard M (2007) VacA and HP-NAP, Ying and Yang of *Helicobacter pylori*-associated gastric inflammation. *Clin Chim Acta* **381**: 32-38.
- Das AK, Helps NR, Cohen PT & Barford D (1996) Crystal structure of the protein serine/threonine phosphatase 2C at 2.0 Å resolution. *EMBO J* **15**: 6798-6809.
- Dauter Z, Dauter M & Dodson E (2002) Jolly SAD. *Acta Crystallogr D Biol Crystallogr* **58**: 494-506.
- de La Fortelle E & Bricogne G (1997) *Maximum-likelihood heavy-atom parameter refinement in the MIR and MAD methods.*
- Dekker C, de Kruijff B & Gros P (2003) Crystal structure of SecB from *Escherichia coli*. *J Struct Biol* **144**: 313-319.
- Delepelaire P (2004) Type I secretion in gram-negative bacteria. *Biochim Biophys Acta* **1694**: 149-161.
- Demali KA, Jue AL & Burridge K (2006) IpaA targets beta1 integrins and rho to promote actin cytoskeleton rearrangements necessary for *Shigella* entry. *J Biol Chem* **281**: 39534-39541.
- Deng K, Mock JR, Greenberg S, van Oers NS & Hansen EJ (2008) *Haemophilus ducreyi* LspA proteins are tyrosine phosphorylated by macrophage-encoded protein tyrosine kinases. *Infect Immun* **76**: 4692-4702.
- Desvaux M, Parham NJ & Henderson IR (2004) The autotransporter secretion system. *Res Microbiol* **155**: 53-60.
- Di Segni A, Farin K & Pinkas-Kramarski R (2008) Identification of nucleolin as new ErbB receptors- interacting protein. *PLoS One* **3**: e2310.

- Dian C, Schauer K, Kapp U, McSweeney SM, Labigne A & Terradot L (2006) Structural basis of the nickel response in *Helicobacter pylori*: crystal structures of HpNikR in Apo and nickel-bound states. *J Mol Biol* **361**: 715-730.
- Diederichs K & Karplus PA (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat Struct Biol* **4**: 269-275.
- Dosanjh NS & Michel SL (2006) Microbial nickel metalloregulation: NikRs for nickel ions. *Curr Opin Chem Biol* **10**: 123-130.
- Dossumbekova A, Prinz C, Mages J, et al. (2006) *Helicobacter pylori* HopH (OipA) and bacterial pathogenicity: genetic and functional genomic analysis of hopH gene polymorphisms. *J Infect Dis* **194**: 1346-1355.
- Dundon WG, Polenghi A, Del Giudice G, Rappuoli R & Montecucco C (2001) Neutrophil-activating protein (HP-NAP) versus ferritin (Pfr): comparison of synthesis in *Helicobacter pylori*. *FEMS Microbiol Lett* **199**: 143-149.
- Efeyan A & Serrano M (2007) p53: guardian of the genome and policeman of the oncogenes. *Cell Cycle* **6**: 1006-1010.
- Eidt S & Stolte M (1995) The significance of *Helicobacter pylori* in relation to gastric cancer and lymphoma. *Eur J Gastroenterol Hepatol* **7**: 318-321.
- Elwell CA, Ceesay A, Kim JH, Kalman D & Engel JN (2008) RNA interference screen identifies Abl kinase and PDGFR signaling in *Chlamydia trachomatis* entry. *PLoS Pathog* **4**: e1000021.
- Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr.* **60**: 2126-2132. Epub 2004 Nov 2126.
- Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**: 2126-2132.
- Etienne-Manneville S & Hall A (2002) Rho GTPases in cell biology. *Nature* **420**: 629-635.
- Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* **62**: 72-82.
- Evdokimov AG, Tropea JE, Routzahn KM & Waugh DS (2002) Crystal structure of the *Yersinia pestis* GTPase activator YopE. *Protein Sci* **11**: 401-408.
- Falzano L, Filippini P, Travaglione S, Miraglia AG, Fabbri A & Fiorentini C (2006) *Escherichia coli* cytotoxic necrotizing factor 1 blocks cell cycle G2/M transition in uroepithelial cells. *Infect Immun* **74**: 3765-3772.
- Fischer W, Puls J, Buhrdorf R, Gebert B, Odenbreit S & Haas R (2001) Systematic mutagenesis of the *Helicobacter pylori* cag pathogenicity island: essential genes for CagA translocation in host cells and induction of interleukin-8. *Mol Microbiol* **42**: 1337-1348.

- Flatau G, Landraud L, Boquet P, Bruzzone M & Munro P (2000) Deamidation of RhoA glutamine 63 by the *Escherichia coli* CNF1 toxin requires a short sequence of the GTPase switch 2 domain. *Biochem Biophys Res Commun* **267**: 588-592.
- Frankel G & Phillips AD (2008) Attaching effacing *Escherichia coli* and paradigms of Tir-triggered actin polymerization: getting off the pedestal. *Cell Microbiol* **10**: 549-556.
- Fraser JD & Proft T (2008) The bacterial superantigen and superantigen-like proteins. *Immunol Rev* **225**: 226-243.
- Frenzel A, Grespi F, Chmielewski W & Villunger A (2009) Bcl2 family proteins in carcinogenesis and the treatment of cancer. *Apoptosis* **14**: 584-596.
- Fronzes R, Schafer E, Wang L, Saibil HR, Orlova EV & Waksman G (2009) Structure of a type IV secretion system core complex. *Science* **323**: 266-268.
- Gangwer KA, Mushrush DJ, Stauff DL, Spiller B, McClain MS, Cover TL & Lacy DB (2007) Crystal structure of the *Helicobacter pylori* vacuolating toxin p55 domain. *Proc Natl Acad Sci U S A* **104**: 16293-16298.
- Gao LY, Guo S, McLaughlin B, Morisaki H, Engel JN & Brown EJ (2004) A mycobacterial virulence gene cluster extending RD1 is required for cytolysis, bacterial spreading and ESAT-6 secretion. *Mol Microbiol* **53**: 1677-1693.
- Garman E & Owen RL (2007) Cryocrystallography of macromolecules: practice and optimization. *Methods Mol Biol* **364**: 1-18.
- Ge Z, Schauer DB & Fox JG (2008) In vivo virulence properties of bacterial cytolethal-distending toxin. *Cell Microbiol* **10**: 1599-1607.
- Gebert B, Fischer W, Weiss E, Hoffmann R & Haas R (2003) *Helicobacter pylori* vacuolating cytotoxin inhibits T lymphocyte activation. *Science* **301**: 1099-1102.
- Gerding MA, Ogata Y, Pecora ND, Niki H & de Boer PA (2007) The trans-envelope Tol-Pal complex is part of the cell division machinery and required for proper outer-membrane invagination during cell constriction in *E. coli*. *Mol Microbiol* **63**: 1008-1025.
- Gerlach RG & Hensel M (2007) Protein secretion systems and adhesins: the molecular armory of Gram-negative pathogens. *Int J Med Microbiol* **297**: 401-415.
- Giamboi-Miraglia A, Travaglione S, Filippini P, Fabbri A, Fiorentini C & Falzano L (2007) A multinucleating *Escherichia coli* cytotoxin perturbs cell cycle in cultured epithelial cells. *Toxicol In Vitro* **21**: 235-239.
- Godlewska R, Pawlowski M, Dzwonek A, Mikula M, Ostrowski J, Drela N & Jagusztyn-Krynicka EK (2008) Tip-alpha (hp0596 gene product) is a highly immunogenic *Helicobacter pylori* protein involved in colonization of mouse gastric mucosa. *Curr Microbiol* **56**: 279-286.

- Gonzalez A, Ruiz A, Casamayor A & Arino J (2009) Normal function of the yeast TOR pathway requires the type 2C protein phosphatase Ptc1. *Mol Cell Biol* **29**: 2876-2888.
- Gonzalez A, Ruiz A, Serrano R, Arino J & Casamayor A (2006) Transcriptional profiling of the protein phosphatase 2C family in yeast provides insights into the unique functional roles of Ptc1. *J Biol Chem* **281**: 35057-35069.
- Gunzl A & Schimanski B (2009) Tandem affinity purification of proteins. *Curr Protoc Protein Sci* **Chapter 19**: Unit 19 19.
- Guo M, Jin S, Sun D, Hew CL & Pan SQ (2007) Recruitment of conjugative DNA transfer substrate to *Agrobacterium* type IV secretion apparatus. *Proc Natl Acad Sci U S A* **104**: 20019-20024.
- Ha NC, Oh ST, Sung JY, Cha KA, Lee MH & Oh BH (2001) Supramolecular assembly and acid resistance of *Helicobacter pylori* urease. *Nat Struct Biol* **8**: 505-509.
- Halbedel S, Busse J, Schmidl SR & Stulke J (2006) Regulatory protein phosphorylation in *Mycoplasma pneumoniae*. A PP2C-type phosphatase serves to dephosphorylate HPr(Ser-P). *J Biol Chem* **281**: 26253-26259.
- Harada T, Miyake M & Imai Y (2007) Evasion of *Legionella pneumophila* from the bactericidal system by reactive oxygen species (ROS) in macrophages. *Microbiol Immunol* **51**: 1161-1170.
- Hare S, Fischer W, Williams R, Terradot L, Bayliss R, Haas R & Waksman G (2007) Identification, structure and mode of action of a new regulator of the *Helicobacter pylori* HP0525 ATPase. *EMBO J* **26**: 4926-4934.
- Hatakeyama M (2002) Deregulation of SHP-2 tyrosine phosphatase by the *Helicobacter pylori* virulence factor CagA. *Keio J Med* **51 Suppl 2**: 26-32.
- Hatakeyama M (2003) *Helicobacter pylori* CagA--a potential bacterial oncoprotein that functionally mimics the mammalian Gab family of adaptor proteins. *Microbes Infect* **5**: 143-150.
- Hatakeyama M (2003) *Helicobacter pylori* CagA as a potential bacterial oncoprotein in gastric carcinogenesis. *Pathol Biol (Paris)* **51**: 393-394.
- Hatakeyama M (2009) *Helicobacter pylori* and gastric carcinogenesis. *J Gastroenterol* **44**: 239-248.
- Hellstrom PM (2006) This year's Nobel Prize to gastroenterology: Robin Warren and Barry Marshall awarded for their discovery of *Helicobacter pylori* as pathogen in the gastrointestinal tract. *World J Gastroenterol* **12**: 3126-3127.
- Henderson IR, Navarro-Garcia F, Desvaux M, Fernandez RC & Ala'Aldeen D (2004) Type V protein secretion pathway: the autotransporter story. *Microbiol Mol Biol Rev* **68**: 692-744.

- Hentschke M, Trulzsch K, Heesemann J, Aepfelbacher M & Ruckdeschel K (2007) Serogroup-related escape of *Yersinia enterocolitica* YopE from degradation by the ubiquitin-proteasome pathway. *Infect Immun* **75**: 4423-4431.
- Hessey SJ, Spencer J, Wyatt JI, Sobala G, Rathbone BJ, Axon AT & Dixon MF (1990) Bacterial adhesion and disease activity in *Helicobacter* associated chronic gastritis. *Gut* **31**: 134-138.
- Higashi H, Tsutsumi R, Muto S, Sugiyama T, Azuma T, Asaka M & Hatakeyama M (2002) SHP-2 tyrosine phosphatase as an intracellular target of *Helicobacter pylori* CagA protein. *Science* **295**: 683-686.
- Higashi H, Yokoyama K, Fujii Y, *et al.* (2005) EPIYA motif is a membrane-targeting signal of *Helicobacter pylori* virulence factor CagA in mammalian cells. *J Biol Chem* **280**: 23130-23137.
- Hisatsune J, Yamasaki E, Nakayama M, *et al.* (2007) *Helicobacter pylori* VacA enhances prostaglandin E2 production through induction of cyclooxygenase 2 expression via a p38 mitogen-activated protein kinase/activating transcription factor 2 cascade in AZ-521 cells. *Infect Immun* **75**: 4472-4481.
- Hodgkinson JL, Horsley A, Stabat D, *et al.* (2009) Three-dimensional reconstruction of the *Shigella* T3SS transmembrane regions reveals 12-fold symmetry and novel features throughout. *Nat Struct Mol Biol* **16**: 477-485.
- Hohlfeld S, Pattis I, Puls J, Plano GV, Haas R & Fischer W (2006) A C-terminal translocation signal is necessary, but not sufficient for type IV secretion of the *Helicobacter pylori* CagA protein. *Mol Microbiol* **59**: 1624-1637.
- Holland IB, Schmitt L & Young J (2005) Type 1 protein secretion in bacteria, the ABC-transporter dependent pathway (review). *Mol Membr Biol* **22**: 29-39.
- Hoppner C, Liu Z, Domke N, Binns AN & Baron C (2004) VirB1 orthologs from *Brucella suis* and pKM101 complement defects of the lytic transglycosylase required for efficient type IV secretion from *Agrobacterium tumefaciens*. *J Bacteriol* **186**: 1415-1422.
- Hsu Y, Jubelin G, Taieb F, Nougayrede JP, Oswald E & Stebbins CE (2008) Structure of the cyclomodulin Cif from pathogenic *Escherichia coli*. *J Mol Biol* **384**: 465-477.
- Hung CL, Liu JH, Chiu WC, Huang SW, Hwang JK & Wang WC (2007) Crystal structure of *Helicobacter pylori* formamidase AmiF reveals a cysteine-glutamate-lysine catalytic triad. *J Biol Chem* **282**: 12220-12229.
- Jacob-Dubuisson F, Fernandez R & Coutte L (2004) Protein secretion through autotransporter and two-partner pathways. *Biochim Biophys Acta* **1694**: 235-257.
- Jaeger KE & Eggert T (2002) Lipases for biotechnology. *Curr Opin Biotechnol* **13**: 390-397.
- Jakubowski SJ, Cascales E, Krishnamoorthy V & Christie PJ (2005) *Agrobacterium tumefaciens* VirB9, an outer-membrane-associated component of a type IV

- secretion system, regulates substrate selection and T-pilus biogenesis. *J Bacteriol* **187**: 3486-3495.
- Jakubowski SJ, Kerr JE, Garza I, Krishnamoorthy V, Bayliss R, Waksman G & Christie PJ (2009) *Agrobacterium* VirB10 domain requirements for type IV secretion and T pilus biogenesis. *Mol Microbiol* **71**: 779-794.
- Jang JY, Yoon HJ, Yoon JY, *et al.* (2009) Crystal structure of the TNF-alpha-Inducing protein (Tipalpha) from *Helicobacter pylori*: Insights into Its DNA-binding activity. *J Mol Biol* **392**: 191-197.
- Jewett TJ, Dooley CA, Mead DJ & Hackstadt T (2008) *Chlamydia trachomatis* tarp is phosphorylated by src family tyrosine kinases. *Biochem Biophys Res Commun* **371**: 339-344.
- Johnson LN (2009) The regulation of protein phosphorylation. *Biochem Soc Trans* **37**: 627-641.
- Jubelin G, Chavez CV, Taieb F, *et al.* (2009) Cycle inhibiting factors (CIFs) are a growing family of functional cyclomodulins present in invertebrate and mammal bacterial pathogens. *PLoS One* **4**: e4855.
- Judd PK, Kumar RB & Das A (2005) Spatial location and requirements for the assembly of the *Agrobacterium tumefaciens* type IV secretion apparatus. *Proc Natl Acad Sci U S A* **102**: 11498-11503.
- Judd PK, Mahli D & Das A (2005) Molecular characterization of the *Agrobacterium tumefaciens* DNA transfer protein VirB6. *Microbiology* **151**: 3483-3492.
- Juhas M, Crook DW & Hood DW (2008) Type IV secretion systems: tools of bacterial horizontal gene transfer and virulence. *Cell Microbiol* **10**: 2377-2386.
- JW IJ, Carlson AC & Kennedy EL (2007) *Anaplasma phagocytophilum* AnkaA is tyrosine-phosphorylated at EPIYA motifs and recruits SHP-1 during early infection. *Cell Microbiol* **9**: 1284-1296.
- Kajava AV & Steven AC (2006) The turn of the screw: variations of the abundant beta-solenoid motif in passenger domains of Type V secretory proteins. *J Struct Biol* **155**: 306-315.
- Kenny B (1999) Phosphorylation of tyrosine 474 of the enteropathogenic *Escherichia coli* (EPEC) Tir receptor molecule is essential for actin nucleating activity and is preceded by additional host modifications. *Mol Microbiol* **31**: 1229-1241.
- Ki MR, Lee HR, Goo MJ, *et al.* (2008) Differential regulation of ERK1/2 and p38 MAP kinases in VacA-induced apoptosis of gastric epithelial cells. *Am J Physiol Gastrointest Liver Physiol* **294**: G635-647.
- Kida Y, Higashimoto Y, Inoue H, Shimizu T & Kuwano K (2008) A novel secreted protease from *Pseudomonas aeruginosa* activates NF-kappaB through protease-activated receptors. *Cell Microbiol* **10**: 1491-1504.

- Kieboom J & Abee T (2006) Arginine-dependent acid resistance in *Salmonella enterica* serovar *Typhimurium*. *J Bacteriol* **188**: 5650-5653.
- Kim AC, Oliver DC & Paetzel M (2008) Crystal structure of a bacterial signal Peptide peptidase. *J Mol Biol* **376**: 352-366.
- Klepsch M, Schlegel S, Wickstrom D, *et al.* (2008) Immobilization of the first dimension in 2D blue native/SDS-PAGE allows the relative quantification of membrane proteomes. *Methods* **46**: 48-53.
- Korotkov KV, Krumm B, Bagdasarian M & Hol WG (2006) Structural and functional studies of EpsC, a crucial component of the type 2 secretion system from *Vibrio cholerae*. *J Mol Biol* **363**: 311-321.
- Korotkov KV, Gray MD, Kreger A, Turley S, Sandkvist M & Hol WG (2009) Calcium is essential for the major pseudopilin in the type 2 secretion system. *J Biol Chem* **284**: 25466-25470.
- Kouokam JC, Wai SN, Fallman M, Dobrindt U, Hacker J & Uhlin BE (2006) Active cytotoxic necrotizing factor 1 associated with outer membrane vesicles from uropathogenic *Escherichia coli*. *Infect Immun* **74**: 2022-2030.
- Krogh A, Larsson B, von Heijne G & Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**: 567-580.
- Kubori T, Matsushima Y, Nakamura D, *et al.* (1998) Supramolecular structure of the *Salmonella typhimurium* type III protein secretion system. *Science* **280**: 602-605.
- Kurashima Y, Murata-Kamiya N, Kikuchi K, Higashi H, Azuma T, Kondo S & Hatakeyama M (2008) Dereglulation of beta-catenin signal by *Helicobacter pylori* CagA requires the CagA-multimerization sequence. *Int J Cancer* **122**: 823-831.
- Kutter S, Buhrdorf R, Haas J, Schneider-Brachert W, Haas R & Fischer W (2008) Protein subassemblies of the *Helicobacter pylori* Cag type IV secretion system revealed by localization and interaction studies. *J Bacteriol* **190**: 2161-2171.
- Kuzuhara T, Suganuma M, Oka K & Fujiki H (2007) DNA-binding activity of TNF-alpha inducing protein from *Helicobacter pylori*. *Biochem Biophys Res Commun* **362**: 805-810.
- Kuzuhara T, Suganuma M, Kurusu M & Fujiki H (2007) *Helicobacter pylori*-secreting protein Tipalpha is a potent inducer of chemokine gene expressions in stomach cancer cells. *J Cancer Res Clin Oncol* **133**: 287-296.
- Kwok T, Zabler D, Urman S, *et al.* (2007) *Helicobacter* exploits integrin for type IV secretion and kinase activation. *Nature* **449**: 862-866.
- Langer G, Cohen SX, Lamzin VS & Perrakis A (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc* **3**: 1171-1179.

- Lawley TD, Klimke WA, Gubbins MJ & Frost LS (2003) F factor conjugation is a true type IV secretion system. *FEMS Microbiol Lett* **224**: 1-15.
- Lazzaroni JC, Dubuisson JF & Vianney A (2002) The Tol proteins of *Escherichia coli* and their involvement in the translocation of group A colicins. *Biochimie* **84**: 391-397.
- Lebedev AA, Vagin AA & Murshudov GN (2008) Model preparation in MOLREP and examples of model improvement using X-ray data. *Acta Crystallogr D Biol Crystallogr* **64**: 33-39.
- Leidal KG, Munson KL, Johnson MC & Denning GM (2003) Metalloproteases from *Pseudomonas aeruginosa* degrade human RANTES, MCP-1, and ENA-78. *J Interferon Cytokine Res* **23**: 307-318.
- Leitch S, Bradley MJ, Rowe JL, Chivers PT & Maroney MJ (2007) Nickel-specific response in the transcriptional regulator, *Escherichia coli* NikR. *J Am Chem Soc* **129**: 5085-5095.
- Leslie AGW (1992) Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 and ESF-EACMB Newsletter on Protein Crystallography No26 Daresbury Laboratory, Warrington, UK*.
- Letley DP, Rhead JL, Bishop K & Atherton JC (2006) Paired cysteine residues are required for high levels of the *Helicobacter pylori* autotransporter VacA. *Microbiology* **152**: 1319-1325.
- Leunk RD, Johnson PT, David BC, Kraft WG & Morgan DR (1988) Cytotoxic activity in broth-culture filtrates of *Campylobacter pylori*. *J Med Microbiol* **26**: 93-99.
- Lin M, den Dulk-Ras A, Hooykaas PJ & Rikihisa Y (2007) *Anaplasma phagocytophilum* AnkA secreted by type IV secretion system is tyrosine phosphorylated by Abl-1 to facilitate infection. *Cell Microbiol* **9**: 2644-2657.
- Liu BA, Jablonowski K, Raina M, Arce M, Pawson T & Nash PD (2006) The human and mouse complement of SH2 domain proteins-establishing the boundaries of phosphotyrosine signaling. *Mol Cell* **22**: 851-868.
- Llosa M, Zupan J, Baron C & Zambryski P (2000) The N- and C-terminal portions of the *Agrobacterium* VirB1 protein independently enhance tumorigenesis. *J Bacteriol* **182**: 3437-3445.
- Locksley RM, Killeen N & Lenardo MJ (2001) The TNF and TNF receptor superfamilies: integrating mammalian biology. *Cell* **104**: 487-501.
- Lu H, Murata-Kamiya N, Saito Y & Hatakeyama M (2009) Role of Partitioning-defective 1/Microtubule Affinity-regulating Kinases in the Morphogenetic Activity of *Helicobacter pylori* CagA. *J Biol Chem* **284**: 23024-23036.
- Lu L, Lamm ME, Li H, Corthesy B & Zhang JR (2003) The human polymeric immunoglobulin receptor binds to *Streptococcus pneumoniae* via domains 3 and 4. *J Biol Chem* **278**: 48178-48187.

- Marlovits TC, Kubori T, Sukhan A, Thomas DR, Galan JE & Unger VM (2004) Structural insights into the assembly of the type III secretion needle complex. *Science* **306**: 1040-1042.
- Marshall BJ & Warren JR (1984) Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *Lancet* **1**: 1311-1315.
- McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC & Read RJ (2007) Phaser crystallographic software. *Journal of Applied Crystallography* **40**: 658-674.
- McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC & Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**: 658-674.
- McNamara D & El-Omar E (2008) *Helicobacter pylori* infection and the pathogenesis of gastric cancer: a paradigm for host-bacterial interactions. *Dig Liver Dis* **40**: 504-509.
- Mehlitz A, Banhart S, Hess S, Selbach M & Meyer TF (2008) Complex kinase requirements for *Chlamydia trachomatis* Tarp phosphorylation. *FEMS Microbiol Lett* **289**: 233-240.
- Meng G, St Geme JW, 3rd & Waksman G (2008) Repetitive architecture of the *Haemophilus influenzae* Hia trimeric autotransporter. *J Mol Biol* **384**: 824-836.
- Meyer-ter-Vehn T, Covacci A, Kist M & Pahl HL (2000) *Helicobacter pylori* activates mitogen-activated protein kinase cascades and induces expression of the proto-oncogenes c-fos and c-jun. *J Biol Chem* **275**: 16064-16072.
- Mock JR, Vakevainen M, Deng K, *et al.* (2005) *Haemophilus ducreyi* targets Src family protein tyrosine kinases to inhibit phagocytic signaling. *Infect Immun* **73**: 7808-7816.
- Moese S, Selbach M, Zimny-Arndt U, Jungblut PR, Meyer TF & Backert S (2001) Identification of a tyrosine-phosphorylated 35 kDa carboxy-terminal fragment (p35CagA) of the *Helicobacter pylori* CagA protein in phagocytic cells: processing or breakage? *Proteomics* **1**: 618-629.
- Molinari M, Galli C, Norais N, Telford JL, Rappuoli R, Luzio JP & Montecucco C (1997) Vacuoles induced by *Helicobacter pylori* toxin contain both late endosomal and lysosomal markers. *J Biol Chem* **272**: 25339-25344.
- Moll G, Papini E, Colonna R, Burrone D, Telford J, Rappuoli R & Montecucco C (1995) Lipid interaction of the 37-kDa and 58-kDa fragments of the *Helicobacter pylori* cytotoxin. *Eur J Biochem* **234**: 947-952.
- Moreau PL (2007) The lysine decarboxylase CadA protects *Escherichia coli* starved of phosphate against fermentation acids. *J Bacteriol* **189**: 2249-2261.
- Mougous JD, Cuff ME, Raunser S, *et al.* (2006) A virulence locus of *Pseudomonas aeruginosa* encodes a protein secretion apparatus. *Science* **312**: 1526-1530.

- Mueller CA, Broz P & Cornelis GR (2008) The type III secretion system tip complex and translocon. *Mol Microbiol* **68**: 1085-1095.
- Murata-Kamiya N, Kurashima Y, Teishikata Y, *et al.* (2007) *Helicobacter pylori* CagA interacts with E-cadherin and deregulates the beta-catenin signal that promotes intestinal transdifferentiation in gastric epithelial cells. *Oncogene* **26**: 4617-4626.
- Murdoch C, Read RC, Zhang Q & Finn A (2002) Choline-binding protein A of *Streptococcus pneumoniae* elicits chemokine production and expression of intercellular adhesion molecule 1 (CD54) by human alveolar epithelial cells. *J Infect Dis* **186**: 1253-1260.
- Murshudov GN, Vagin AA & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**: 240-255.
- Murshudov GN, Vagin AA, Lebedev A, Wilson KS & Dodson EJ (1999) Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr. D Biol. Crystallogr.* **55**: 247-255.
- Naito M, Yamazaki T, Tsutsumi R, *et al.* (2006) Influence of EPIYA-repeat polymorphism on the phosphorylation-dependent biological activity of *Helicobacter pylori* CagA. *Gastroenterology* **130**: 1181-1190.
- Nesic D, Hsu Y & Stebbins CE (2004) Assembly and function of a bacterial genotoxin. *Nature* **429**: 429-433.
- Noble JE & Bailey MJ (2009) Quantitation of protein. *Methods Enzymol* **463**: 73-95.
- Nougayrede JP, Taieb F, De Rycke J & Oswald E (2005) Cyclomodulins: bacterial effectors that modulate the eukaryotic cell cycle. *Trends Microbiol* **13**: 103-110.
- Nourai M, Latifi-Navid S, Rezvan H, *et al.* (2009) Childhood hygienic practice and family education status determine the prevalence of *Helicobacter pylori* infection in Iran. *Helicobacter* **14**: 40-46.
- Nouwen N & Driessen AJ (2002) SecDFyajC forms a heterotetrameric complex with YidC. *Mol Microbiol* **44**: 1397-1405.
- Nouwen N, Berrelkamp G & Driessen AJ (2007) Bacterial sec-translocase unfolds and translocates a class of folded protein domains. *J Mol Biol* **372**: 422-433.
- Odenbreit S, Swoboda K, Barwig I, Ruhl S, Boren T, Koletzko S & Haas R (2009) Outer membrane protein expression profile in *Helicobacter pylori* clinical isolates. *Infect Immun* **77**: 3782-3790.
- Ogawa M, Handa Y, Ashida H, Suzuki M & Sasakawa C (2008) The versatility of *Shigella* effectors. *Nat Rev Microbiol* **6**: 11-16.
- Ohnishi N, Yuasa H, Tanaka S, *et al.* (2008) Transgenic expression of *Helicobacter pylori* CagA induces gastrointestinal and hematopoietic neoplasms in mouse. *Proc Natl Acad Sci U S A* **105**: 1003-1008.

- Oleastro M, Cordeiro R, Yamaoka Y, Queiroz D, Megraud F, Monteiro L & Menard A (2009) Disease association with two *Helicobacter pylori* duplicate outer membrane protein genes, homB and homA. *Gut Pathog* **1**: 12.
- Oliver DC & Paetzel M (2008) Crystal structure of the major periplasmic domain of the bacterial membrane protein assembly facilitator YidC. *J Biol Chem* **283**: 5208-5216.
- Oomen CJ, van Ulsen P, van Gelder P, Feijen M, Tommassen J & Gros P (2004) Structure of the translocator domain of a bacterial autotransporter. *EMBO J* **23**: 1257-1266.
- Osaki M, Arcondeguy T, Bastide A, Touriol C, Prats H & Trombe MC (2009) The StkP/PhpP signaling couple in *Streptococcus pneumoniae*: cellular organization and physiological characterization. *J Bacteriol* **191**: 4943-4950.
- Ottmann KM & Lowenthal AC (2002) *Helicobacter pylori* uses motility for initial colonization and to attain robust infection. *Infect Immun* **70**: 1984-1990.
- Paetzel M, Goodall JJ, Kania M, Dalbey RE & Page MG (2004) Crystallographic and biophysical analysis of a bacterial signal peptidase in complex with a lipopeptide-based inhibitor. *J Biol Chem* **279**: 30781-30790.
- Pane-Farre J, Jonas B, Hardwick SW, Gronau K, Lewis RJ, Hecker M & Engelmann S (2009) Role of RsbU in controlling SigB activity in *Staphylococcus aureus* following alkaline stress. *J Bacteriol* **191**: 2561-2573.
- Pappa HS, Stewart AE & McDonald NQ (1996) Incorporating anomalous scattering centres into macromolecules. *Curr Opin Struct Biol* **6**: 611-616.
- Parmely M, Gale A, Clabaugh M, Horvat R & Zhou WW (1990) Proteolytic inactivation of cytokines by *Pseudomonas aeruginosa*. *Infect Immun* **58**: 3009-3014.
- Pattis I, Weiss E, Laugks R, Haas R & Fischer W (2007) The *Helicobacter pylori* CagF protein is a type IV secretion chaperone-like molecule that binds close to the C-terminal secretion signal of the CagA effector protein. *Microbiology* **153**: 2896-2909.
- Pawson T, Gish GD & Nash P (2001) SH2 domains, interaction modules and cellular wiring. *Trends Cell Biol* **11**: 504-511.
- Peabody CR, Chung YJ, Yen MR, Vidal-Ingigliardi D, Pugsley AP & Saier MH, Jr. (2003) Type II protein secretion and its relationship to bacterial type IV pili and archaeal flagella. *Microbiology* **149**: 3051-3072.
- Perrakis A, Morris R & Lamzin VS (1999) Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6**: 458-463.
- Pflock M, Kennard S, Finsterer N & Beier D (2006) Acid-responsive gene regulation in the human pathogen *Helicobacter pylori*. *J Biotechnol* **126**: 52-60.

- Pflock M, Finsterer N, Joseph B, Mollenkopf H, Meyer TF & Beier D (2006) Characterization of the ArsRS regulon of *Helicobacter pylori*, involved in acid adaptation. *J Bacteriol* **188**: 3449-3462.
- Phalipon A & Sansonetti PJ (2007) *Shigella's* ways of manipulating the host intestinal innate and adaptive immune system: a tool box for survival? *Immunol Cell Biol* **85**: 119-129.
- Philippe C, Blech MF & Hartemann P (2006) [Intra-amoebal development of *Legionella pneumophila* and the potential role of amoebae in the transmission of Legionnaires' disease]. *Med Mal Infect* **36**: 196-200.
- Pimenta AL, Young J, Holland IB & Blight MA (1999) Antibody analysis of the localisation, expression and stability of HlyD, the MFP component of the *E. coli* haemolysin translocator. *Mol Gen Genet* **261**: 122-132.
- Powell HR (1999) The Rossmann Fourier autoindexing algorithm in MOSFLM. *Acta Crystallogr D Biol Crystallogr* **55**: 1690-1695.
- Prasad TS, Kandasamy K & Pandey A (2009) Human Protein Reference Database and Human Proteinpedia as discovery tools for systems biology. *Methods Mol Biol* **577**: 67-79.
- Pukatzki S, McAuley SB & Miyata ST (2009) The type VI secretion system: translocation of effectors and effector-domains. *Curr Opin Microbiol* **12**: 11-17.
- Rain JC, Selig L, De Reuse H, *et al.* (2001) The protein-protein interaction map of *Helicobacter pylori*. *Nature* **409**: 211-215.
- Ren J, Kotaka M, Lockyer M, Lamb HK, Hawkins AR & Stammers DK (2005) GTP cyclohydrolase II structure and mechanism. *J Biol Chem* **280**: 36912-36919.
- Ren S, Higashi H, Lu H, Azuma T & Hatakeyama M (2006) Structural basis and functional consequence of *Helicobacter pylori* CagA multimerization in cells. *J Biol Chem* **281**: 32344-32352.
- Robien MA, Krumm BE, Sandkvist M & Hol WG (2003) Crystal structure of the extracellular protein secretion NTPase EpsE of *Vibrio cholerae*. *J Mol Biol* **333**: 657-674.
- Rosado CJ, Kondos S, Bull TE, *et al.* (2008) The MACPF/CDC family of pore-forming toxins. *Cell Microbiol* **10**: 1765-1774.
- Rossjohn J, Polekhina G, Feil SC, Morton CJ, Tweten RK & Parker MW (2007) Structures of perfringolysin O suggest a pathway for activation of cholesterol-dependent cytolysins. *J Mol Biol* **367**: 1227-1236.
- Rossmann MG & van Beek CG (1999) Data processing. *Acta Crystallogr D Biol Crystallogr* **55**: 1631-1640.
- Ruiz C, Falcocchio S, Pastor FI, Saso L & Diaz P (2007) *Helicobacter pylori* EstV: identification, cloning, and characterization of the first lipase isolated from an epsilon-proteobacterium. *Appl Environ Microbiol* **73**: 2423-2431.

- Saadat I, Higashi H, Obuse C, *et al.* (2007) *Helicobacter pylori* CagA targets PAR1/MARK kinase to disrupt epithelial cell polarity. *Nature* **447**: 330-333.
- Sachs G (2003) Physiology of the parietal cell and therapeutic implications. *Pharmacotherapy* **23**: 68S-73S.
- Sachs G, Kraut JA, Wen Y, Feng J & Scott DR (2006) Urea transport in bacteria: acid acclimation by gastric *Helicobacter spp.* *J Membr Biol* **212**: 71-82.
- Sachs G, Weeks DL, Wen Y, Marcus EA, Scott DR & Melchers K (2005) Acid acclimation by *Helicobacter pylori*. *Physiology (Bethesda)* **20**: 429-438.
- Samba-Louaka A, Nougayrede JP, Watrin C, Jubelin G, Oswald E & Taieb F (2008) Bacterial cyclomodulin Cif blocks the host cell cycle by stabilizing the cyclin-dependent kinase inhibitors p21 and p27. *Cell Microbiol* **10**: 2496-2508.
- Savvides SN, Yeo HJ, Beck MR, *et al.* (2003) VirB11 ATPases are dynamic hexameric assemblies: new insights into bacterial type IV secretion. *EMBO J* **22**: 1969-1980.
- Schlumberger MC & Hardt WD (2005) Triggered phagocytosis by *Salmonella*: bacterial molecular mimicry of RhoGTPase activation/deactivation. *Curr Top Microbiol Immunol* **291**: 29-42.
- Schneider TR & Sheldrick GM (2002) Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr* **58**: 1772-1779.
- Schreiber S, Bucker R, Groll C, *et al.* (2005) Rapid loss of motility of *Helicobacter pylori* in the gastric lumen in vivo. *Infect Immun* **73**: 1584-1589.
- Schuller S, Chong Y, Lewin J, Kenny B, Frankel G & Phillips AD (2007) Tir phosphorylation and Nck/N-WASP recruitment by enteropathogenic and enterohaemorrhagic *Escherichia coli* during ex vivo colonization of human intestinal mucosa is different to cell culture models. *Cell Microbiol* **9**: 1352-1364.
- Schwartz M (2004) Rho signalling at a glance. *J Cell Sci* **117**: 5457-5458.
- Scott DR, Marcus EA, Weeks DL & Sachs G (2002) Mechanisms of acid resistance due to the urease system of *Helicobacter pylori*. *Gastroenterology* **123**: 187-195.
- Segal G, Russo JJ & Shuman HA (1999) Relationships between a new type IV secretion system and the icm/dot virulence system of *Legionella pneumophila*. *Mol Microbiol* **34**: 799-809.
- Selbach M, Moese S, Hurwitz R, Hauck CR, Meyer TF & Backert S (2003) The *Helicobacter pylori* CagA protein induces cortactin dephosphorylation and actin rearrangement by c-Src inactivation. *EMBO J* **22**: 515-528.
- Selbach M, Paul FE, Brandt S, *et al.* (2009) Host cell interactome of tyrosine-phosphorylated bacterial proteins. *Cell Host Microbe* **5**: 397-403.
- Shames SR, Auweter SD & Finlay BB (2009) Co-evolution and exploitation of host cell signaling pathways by bacterial pathogens. *Int J Biochem Cell Biol* **41**: 380-389.

- Sheldrick GM (2008) A short history of SHELX. *Acta Crystallogr A* **64**: 112-122.
- Shenker BJ, Dlakic M, Walker LP, Besack D, Jaffe E, LaBelle E & Boesze-Battaglia K (2007) A novel mode of action for a microbial-derived immunotoxin: the cytolethal distending toxin subunit B exhibits phosphatidylinositol 3,4,5-triphosphate phosphatase activity. *J Immunol* **178**: 5099-5108.
- Siavoshi F, Malekzadeh R, Daneshmand M & Ashktorab H (2005) *Helicobacter pylori* endemic and gastric disease. *Dig Dis Sci* **50**: 2075-2080.
- Simeone R, Bottai D & Brosch R (2009) ESX/type VII secretion systems and their role in host-pathogen interaction. *Curr Opin Microbiol* **12**: 4-10.
- Smith JL & Bayles DO (2006) The contribution of cytolethal distending toxin to bacterial pathogenesis. *Crit Rev Microbiol* **32**: 227-248.
- Snider JL, Allison C, Bellaire BH, Ferrero RL & Cardelli JA (2008) The beta1 integrin activates JNK independent of CagA, and JNK activation is required for *Helicobacter pylori* CagA+-induced motility of gastric cancer cells. *J Biol Chem* **283**: 13952-13963.
- Soto GE & Hultgren SJ (1999) Bacterial adhesins: common themes and variations in architecture and assembly. *J Bacteriol* **181**: 1059-1071.
- Spreter T, Yip CK, Sanowar S, *et al.* (2009) A conserved structural motif mediates formation of the periplasmic rings in the type III secretion system. *Nat Struct Mol Biol* **16**: 468-476.
- Stebbins CE & Galan JE (2001) Structural mimicry in bacterial virulence. *Nature* **412**: 701-705.
- Stein M, Bagnoli F, Halenbeck R, Rappuoli R, Fantl WJ & Covacci A (2002) c-Src/Lyn kinases activate *Helicobacter pylori* CagA through tyrosine phosphorylation of the EPIYA motifs. *Mol Microbiol* **43**: 971-980.
- Strandberg B (2009) Chapter 1: building the ground for the first two protein structures: myoglobin and haemoglobin. *J Mol Biol* **392**: 2-10.
- Suerbaum S & Michetti P (2002) *Helicobacter pylori* infection. *N Engl J Med* **347**: 1175-1186.
- Suerbaum S & Josenhans C (2007) *Helicobacter pylori* evolution and phenotypic diversification in a changing host. *Nat Rev Microbiol* **5**: 441-452.
- Suganuma M, Kuzuhara T, Yamaguchi K & Fujiki H (2006) Carcinogenic role of tumor necrosis factor-alpha inducing protein of *Helicobacter pylori* in human stomach. *J Biochem Mol Biol* **39**: 1-8.
- Suganuma M, Okabe S, Marino MW, Sakai A, Sueoka E & Fujiki H (1999) Essential role of tumor necrosis factor alpha (TNF-alpha) in tumor promotion as revealed by TNF-alpha-deficient mice. *Cancer Res* **59**: 4516-4518.

- Suganuma M, Kurusu M, Suzuki K, Nishizono A, Murakami K, Fujioka T & Fujiki H (2005) New tumor necrosis factor-alpha-inducing protein released from *Helicobacter pylori* for gastric cancer progression. *J Cancer Res Clin Oncol* **131**: 305-313.
- Suganuma M, Okabe S, Kurusu M, *et al.* (2002) Discrete roles of cytokines, TNF-alpha, IL-1, IL-6 in tumor promotion and cell transformation. *Int J Oncol* **20**: 131-136.
- Suganuma M, Yamaguchi K, Ono Y, *et al.* (2008) TNF-alpha-inducing protein, a carcinogenic factor secreted from *H. pylori*, enters gastric cancer cells. *Int J Cancer* **123**: 117-122.
- Sugiyama T (2004) Development of gastric cancer associated with *Helicobacter pylori* infection. *Cancer Chemother Pharmacol* **54 Suppl 1**: S12-20.
- Sundberg EJ, Li H, Llera AS, *et al.* (2002) Structures of two streptococcal superantigens bound to TCR beta chains reveal diversity in the architecture of T cell signaling complexes. *Structure* **10**: 687-699.
- Supajatura V, Ushio H, Wada A, *et al.* (2002) Cutting edge: VacA, a vacuolating cytotoxin of *Helicobacter pylori*, directly activates mast cells for migration and production of proinflammatory cytokines. *J Immunol* **168**: 2603-2607.
- Suzuki M, Mimuro H, Suzuki T, Park M, Yamamoto T & Sasakawa C (2005) Interaction of CagA with Crk plays an important role in *Helicobacter pylori*-induced loss of gastric epithelial cell adhesion. *J Exp Med* **202**: 1235-1247.
- Taieb F, Nougayrede JP, Watrin C, Samba-Louaka A & Oswald E (2006) *Escherichia coli* cyclomodulin Cif induces G2 arrest of the host cell cycle without activation of the DNA-damage checkpoint-signalling pathway. *Cell Microbiol* **8**: 1910-1921.
- Tammer I, Brandt S, Hartig R, Konig W & Backert S (2007) Activation of Abl by *Helicobacter pylori*: a novel kinase for CagA and crucial mediator of host cell scattering. *Gastroenterology* **132**: 1309-1319.
- Tarendeau F, Boudet J, Guilligay D, *et al.* (2007) Structure and nuclear import function of the C-terminal domain of influenza virus polymerase PB2 subunit. *Nat Struct Mol Biol* **14**: 229-233.
- Taylor G (2003) The phase problem. *Acta Crystallogr D Biol Crystallogr* **59**: 1881-1890.
- Terebiznik MR, Vazquez CL, Torbicki K, *et al.* (2006) *Helicobacter pylori* VacA toxin promotes bacterial intracellular survival in gastric epithelial cells. *Infect Immun* **74**: 6599-6614.
- Terradot L, Bayliss R, Oomen C, Leonard GA, Baron C & Waksman G (2005) Structures of two core subunits of the bacterial type IV secretion system, VirB8 from *Brucella suis* and ComB10 from *Helicobacter pylori*. *Proc Natl Acad Sci U S A* **102**: 4596-4601.
- Terradot L, Durnell N, Li M, *et al.* (2004) Biochemical characterization of protein complexes from the *Helicobacter pylori* protein interaction map: strategies for complex

- formation and evidence for novel interactions within type IV secretion systems. *Mol Cell Proteomics* **3**: 809-819.
- Terwilliger T (2004) SOLVE and RESOLVE: automated structure solution, density modification and model building. *J Synchrotron Radiat* **11**: 49-52.
- Terwilliger TC (2002) Automated structure solution, density modification and model building. *Acta Crystallogr D Biol Crystallogr* **58**: 1937-1940.
- Thanabalu T, Koronakis E, Hughes C & Koronakis V (1998) Substrate-induced assembly of a contiguous channel for protein export from *E.coli*: reversible bridging of an inner-membrane translocase to an outer membrane exit pore. *EMBO J* **17**: 6487-6496.
- Tilley SJ, Orlova EV, Gilbert RJ, Andrew PW & Saibil HR (2005) Structural basis of pore formation by the bacterial toxin pneumolysin. *Cell* **121**: 247-256.
- Tombola F, Morbiato L, Del Giudice G, Rappuoli R, Zoratti M & Papini E (2001) The *Helicobacter pylori* VacA toxin is a urea permease that promotes urea diffusion across epithelia. *J Clin Invest* **108**: 929-937.
- Torreri P, Ceccarini M, Macioce P & Petrucci TC (2005) Biomolecular interactions by Surface Plasmon Resonance technology. *Ann Ist Super Sanita* **41**: 437-441.
- Tosi T, Cioci G, Jouravleva K, Dian C & Terradot L (2009) Structures of the tumor necrosis factor alpha inducing protein Tipalpha: a novel virulence factor from *Helicobacter pylori*. *FEBS Lett* **583**: 1581-1585.
- Trapani S & Navaza J (2008) AMoRe: classical and modern. *Acta Crystallogr D Biol Crystallogr* **64**: 11-16.
- Tschesche H & Dietl T (1975) The amino-acid sequence of isoinhibitor K from snails (*Helix pomatia*). A sequence determination by automated Edman degradation and mass-spectral identification of the phenylthiohydantoin. *Eur J Biochem* **58**: 439-451.
- Tsuge H, Tsurumura T, Utsunomiya H, *et al.* (2009) Structural basis for the *Helicobacter pylori*-carcinogenic TNF-alpha-inducing protein. *Biochem Biophys Res Commun* **388**: 193-198.
- Tsutsumi R, Higashi H, Higuchi M, Okada M & Hatakeyama M (2003) Attenuation of *Helicobacter pylori* CagA x SHP-2 signaling by interaction between CagA and C-terminal Src kinase. *J Biol Chem* **278**: 3664-3670.
- Tsutsumi R, Takahashi A, Azuma T, Higashi H & Hatakeyama M (2006) Focal adhesion kinase is a substrate and downstream effector of SHP-2 complexed with *Helicobacter pylori* CagA. *Mol Cell Biol* **26**: 261-276.
- Tummuru MK, Sharma SA & Blaser MJ (1995) *Helicobacter pylori* picB, a homologue of the *Bordetella pertussis* toxin secretion protein, is required for induction of IL-8 in gastric epithelial cells. *Mol Microbiol* **18**: 867-876.
- Tuteja R & Tuteja N (1998) Nucleolin: a multifunctional major nucleolar phosphoprotein. *Crit Rev Biochem Mol Biol* **33**: 407-436.

- Upadhyay A, Wu HL, Williams C, Field T, Galyov EE, van den Elsen JM & Bagby S (2008) The guanine-nucleotide-exchange factor BopE from *Burkholderia pseudomallei* adopts a compact version of the Salmonella SopE/SopE2 fold and undergoes a closed-to-open conformational change upon interaction with Cdc42. *Biochem J* **411**: 485-493.
- Veiga E, Sugawara E, Nikaido H, de Lorenzo V & Fernandez LA (2002) Export of autotransported proteins proceeds through an oligomeric ring shaped by C-terminal domains. *EMBO J* **21**: 2122-2131.
- Velayudhan J, Hughes NJ, McColm AA, Bagshaw J, Clayton CL, Andrews SC & Kelly DJ (2000) Iron acquisition and virulence in *Helicobacter pylori*: a major role for FeoB, a high-affinity ferrous iron transporter. *Mol Microbiol* **37**: 274-286.
- Viala J, Chaput C, Boneca IG, *et al.* (2004) Nod1 responds to peptidoglycan delivered by the *Helicobacter pylori* cag pathogenicity island. *Nat Immunol* **5**: 1166-1174.
- Vojtov N, Ross HF & Novick RP (2002) Global repression of exotoxin synthesis by staphylococcal superantigens. *Proc Natl Acad Sci U S A* **99**: 10102-10107.
- Voland P, Weeks DL, Marcus EA, Prinz C, Sachs G & Scott D (2003) Interactions among the seven *Helicobacter pylori* proteins encoded by the urease gene cluster. *Am J Physiol Gastrointest Liver Physiol* **284**: G96-G106.
- Vonrhein C, Blanc E, Roversi P & Bricogne G (2007) Automated structure solution with autoSHARP. *Methods Mol Biol* **364**: 215-230.
- Waksman G, Kumaran S & Lubman O (2004) SH2 domains: role, structure and implications for molecular medicine. *Expert Rev Mol Med* **6**: 1-18.
- Waksman G, Shoelson SE, Pant N, Cowburn D & Kuriyan J (1993) Binding of a high affinity phosphotyrosyl peptide to the Src SH2 domain: crystal structures of the complexed and peptide-free forms. *Cell* **72**: 779-790.
- Walhout AJ, Boulton SJ & Vidal M (2000) Yeast two-hybrid systems and protein interaction mapping projects for yeast and worm. *Yeast* **17**: 88-94.
- Warren JR (1984) Spiral bacteria of the gastric antrum. *Med J Aust* **141**: 477-478.
- Wen S & Moss SF (2009) *Helicobacter pylori* virulence factors in gastric carcinogenesis. *Cancer Lett* **282**: 1-8.
- Weyermann M, Rothenbacher D & Brenner H (2009) Acquisition of *Helicobacter pylori* infection in early childhood: independent contributions of infected mothers, fathers, and siblings. *Am J Gastroenterol* **104**: 182-189.
- Wurtele M, Renault L, Barbieri JT, Wittinghofer A & Wolf E (2001) Structure of the ExoS GTPase activating domain. *FEBS Lett* **491**: 26-29.
- Yamaoka Y (2008) Roles of *Helicobacter pylori* BabA in gastroduodenal pathogenesis. *World J Gastroenterol* **14**: 4265-4272.

- Yamasaki E, Wada A, Kumatori A, *et al.* (2006) *Helicobacter pylori* vacuolating cytotoxin induces activation of the proapoptotic proteins Bax and Bak, leading to cytochrome c release and cell death, independent of vacuolation. *J Biol Chem* **281**: 11250-11259.
- Yanez ME, Korotkov KV, Abendroth J & Hol WG (2008) The crystal structure of a binary complex of two pseudopilins: EpsI and EpsJ from the type 2 secretion system of *Vibrio vulnificus*. *J Mol Biol* **375**: 471-486.
- Yanez ME, Korotkov KV, Abendroth J & Hol WG (2008) Structure of the minor pseudopilin EpsH from the Type 2 secretion system of *Vibrio cholerae*. *J Mol Biol* **377**: 91-103.
- Yeo HJ, Savvides SN, Herr AB, Lanka E & Waksman G (2000) Crystal structure of the hexameric traffic ATPase of the *Helicobacter pylori* type IV secretion system. *Mol Cell* **6**: 1461-1472.
- Yeo HJ, Yuan Q, Beck MR, Baron C & Waksman G (2003) Structural and functional characterization of the VirB5 protein from the type IV secretion system encoded by the conjugative plasmid pKM101. *Proc Natl Acad Sci U S A* **100**: 15947-15952.
- Yokoyama K, Higashi H, Ishikawa S, *et al.* (2005) Functional antagonism between *Helicobacter pylori* CagA and vacuolating toxin VacA in control of the NFAT signaling pathway in gastric epithelial cells. *Proc Natl Acad Sci U S A* **102**: 9661-9666.
- Yoshida M, Wakatsuki Y, Kobayashi Y, *et al.* (1999) Cloning and characterization of a novel membrane-associated antigenic protein of *Helicobacter pylori*. *Infect Immun* **67**: 286-293.
- Yuan Q, Carle A, Gao C, *et al.* (2005) Identification of the VirB4-VirB8-VirB5-VirB2 pilus assembly sequence of type IV secretion systems. *J Biol Chem* **280**: 26349-26359.
- Zanotti G, Papinutto E, Dundon W, *et al.* (2002) Structure of the neutrophil-activating protein from *Helicobacter pylori*. *J Mol Biol* **323**: 125-130.
- Zhang C & Kim SH (2000) The anatomy of protein beta-sheet topology. *J Mol Biol* **299**: 1075-1089.
- Zhu Y, Zhong X, Zheng S, Du Q & Xu W (2005) Transformed immortalized gastric epithelial cells by virulence factor CagA of *Helicobacter pylori* through Erk mitogen-activated protein kinase pathway. *Oncogene* **24**: 3886-3895.
- Zimmer J, Nam Y & Rapoport TA (2008) Structure of a complex of the ATPase SecA and the protein-translocation channel. *Nature* **455**: 936-943.
- Zupan J, Hackworth CA, Aguilar J, Ward D & Zambryski P (2007) VirB1* promotes T-pilus formation in the vir-Type IV secretion system of *Agrobacterium tumefaciens*. *J Bacteriol* **189**: 6551-6563.

Zurawski DV, Mumy KL, Faherty CS, McCormick BA & Maurelli AT (2009) *Shigella flexneri* type III secretion system effectors OspB and OspF target the nucleus to downregulate the host inflammatory response via interactions with retinoblastoma protein. *Mol Microbiol* **71**: 350-368.

Zwart PH, Afonine PV, Grosse-Kunstleve RW, *et al.* (2008) Automated structure solution with the PHENIX suite. *Methods Mol Biol* **426**: 419-435.

Résumé

Helicobacter pylori est une bactérie qui infecte l'estomac de la moitié de la population mondiale et est impliquée dans la plupart des maladies gastriques, dont les ulcères et le cancer de l'estomac. La bactérie produit deux toxines, CagA et Tipα, qui sont associées avec le développement du cancer. CagA est injectée dans les cellules et interagit avec de nombreuses protéines des voies de signalisation cellulaire. Tipα est sécrétée et internalisée dans les cellules gastriques où elle induit la production de cytokines pro-inflammatoires. Dans ce travail de thèse, nous avons tout d'abord étudié les interactions entre un domaine central de CagA et des protéines de *H. pylori*. Nous avons ensuite identifié de nouveaux fragments solubles de CagA par une méthodologie à "haut-débit". L'un d'eux, correspondant à l'extrémité C-terminale de la protéine de 33kDa, CagA^{C33}, a été caractérisé par différentes techniques de biochimie et de biophysique. CagA^{C33} forme des dimères de dimères en solution et des particules en microscopie électronique. De plus, CagA^{C33} peut être phosphorylé *in vitro* par les kinases c-Src et c-Abl mais l'efficacité de la phosphorylation dépend de la kinase utilisée. L'étude de l'interaction entre la phosphatase SHP-2 et CagA^{C33} montre que CagA est déphosphorylée *in vitro*.

Par ailleurs, les structures de deux formes cristallines de Tipα ont été déterminées par cristallographie aux rayons X. La structure du monomère de Tipα adopte un nouveau repliement, et la protéine forme des dimères différents dans les deux formes cristallines. L'étude de la protéine en solution indique qu'un des dimères est sans doute favorisé et suggère que les ponts disulfures identifiés en N-terminal ont un rôle durant la sécrétion de la protéine.

Mots-clés: *Helicobacter pylori*, carcinogène, inflammation, phosphorylation, toxines, virulence, cristallographie.

Abstract

H. pylori is a bacterium that infects the stomach of half of the world population. It is involved in most gastric diseases, in particular ulcers and the stomach cancer. The bacterium produces two toxins, CagA and Tipα, that associate with gastric cancer development. CagA is injected inside the cells and interacts with several proteins from signaling pathways. Tipα is secreted by *H. pylori* and internalized by the gastric cells, where it induces the production of proinflammatory cytokines. In this thesis, we have first studied the interactions between a central domain of CagA and proteins from *H. pylori*. We have then used a novel High-Throughput methodology to determine new soluble CagA fragments. One of them, corresponding to a C-terminal domain of 33kDa (CagA^{C33}), has been characterized by using biochemical and biophysical techniques. CagA^{C33} forms a dimer of dimers in solution and particles in electron microscopy. Moreover, CagA^{C33} can be phosphorylated *in vitro* by the c-Src and c-Abl kinases, but the phosphorylation efficiency depends on the kinase. The study of the interaction between P-CagA and the SHP-2 phosphatase shows that CagA is dephosphorylated *in vitro*.

We have also solved the structures of two crystal forms of Tipα by X-ray crystallography. The structure of Tipα monomer adopts a novel fold and the protein forms different dimers in the two crystal forms. The study of Tipα in solution indicates that one of the crystallographic dimers is likely favored and suggests that the disulfide bridges identified in the N-terminal portion of the protein have a role during the secretion.

Keywords : *Helicobacter pylori*, carcinogenesis, inflammation, phosphorylation, toxins, virulence, crystallography.