

Université de Montréal

**Analyse biochimique et inhibition de complexes
macromoléculaires dans des cellules humaines et
bactériennes**

par **Flore OUDOUHOU**

Département de biochimie et médecine moléculaire, Faculté de médecine

Université de Montréal

Thèse présentée en vue de l'obtention du grade de

Docteur en Biochimie

Août 2018

© Flore OUDOUHOU, 2018

RÉSUMÉ

Les complexes macromoléculaires sont au cœur des processus fondamentaux de la cellule. L'étude de leur composition, de leur structure et des interactions qui s'exercent entre leurs composants permet de mieux appréhender leurs fonctions ; leurs composants représentant bien souvent des cibles d'intérêt en recherche clinique. Lors de mon doctorat nous avons ainsi mis en place une approche multidisciplinaire afin d'étudier deux complexes macromoléculaires : la machinerie eucaryote de synthèse et d'incorporation des sélénocystéines dans les protéines ; et le système de sécrétion de type 4 d'*Helicobacter pylori*.

La sélénocystéine, 21^e acide aminé du code génétique, a pour particularité de ne pas avoir de codon dédié et d'être encodé par le codon « stop » UGA. Ainsi, le décodage en sélénocystéine requiert une machinerie de traduction spécialisée, qui repose sur des interactions transitoires entre protéines et ARN. Bien que les principaux acteurs de la machinerie soient à présent identifiés, de nombreux détails moléculaires restent à clarifier. Afin de caractériser les interactions, de nature dynamique et transitoire, entre les acteurs de la machinerie connus à ce jour, nous avons tiré avantage de l'essai BRET (*Bioluminescence Resonance Energy Transfer*). Ainsi, nos études d'interaction *in vivo* nous ont permis de définir l'état oligomérique des différents facteurs et d'identifier plusieurs interactions inconnues jusqu'alors. Les sites d'interaction entre plusieurs facteurs de la machinerie ont également été déterminés par des approches *in vitro*. Nous proposons ainsi un modèle plus détaillé des étapes conduisant au décodage de la Sec. De plus, nous proposons le premier modèle structural d'une protéine clé, mais mal connue, de la machinerie (SECp43), qui pourra par la suite être exploité pour élucider l'implication de la protéine dans la machinerie de traduction.

Dans un second temps, nous nous sommes intéressés à l'étude du système de sécrétion de type 4 d'*Helicobacter pylori*. La bactérie *H. pylori* est un pathogène humain dont la présence dans l'organisme accroît le risque de développement de cancer gastrique, le système de sécrétion de type 4, encodé par certaines souches, constituant un facteur aggravant de la pathogénicité. Du fait de l'émergence de souches d'*H. pylori* multirésistantes aux antibiotiques, les traitements actuels pour éradiquer la bactérie deviennent inadaptés. Ainsi, nous proposons d'adopter une stratégie alternative aux antibiotiques. Nous avons donc entrepris de développer et caractériser

des molécules capables d'inhiber la virulence de la bactérie, en ciblant la protéine Cag α , une ATPase du système de sécrétion de type 4. La molécule 1G2 a ainsi été identifiée. Celle-ci inhibe l'activité ATPase de Cag α et diminue de près de 50% la réponse pro-inflammatoire des cellules eucaryotes, induite sous l'effet de l'infection par *H. pylori*. L'efficacité de la molécule a, par la suite, été évaluée sur des souches cliniques d'*H. pylori* résistantes aux antibiotiques, préalablement caractérisées. Nous avons ainsi mis en évidence que la molécule 1G2 était également efficace sur des souches cliniques hautement virulentes, de type occidental et d'Asie de l'Est. La chronicité de l'infection par *H. pylori* pouvant conduire au développement de cancers gastriques, nos résultats suggèrent le fort potentiel de la molécule 1G2 dans le traitement de l'infection. Cette molécule pourra notamment servir de base pour la conception d'autres molécules ayant un effet plus notoire sur la virulence d'*H. pylori*.

Mots-clés : Complexes macromoléculaires, sélénocystéine, sélénoprotéine, interaction protéique, BRET, SEPHS1, SEPHS2, SEPSECS, SECp43, *H. pylori*, SST4, inhibiteurs de virulence, CagL, CagA, Cag α

ABSTRACT

All cellular processes depend on macromolecular complexes. Studying their composition, structure and the interactions between their components is a crucial step toward understanding of their functions. Moreover, their components can be targeted in applied research aimed at drug design. During my Ph.D., we set up a multidisciplinary approach to study two macromolecular complexes: the eukaryotic selenocysteine (Sec) biosynthesis and incorporation machinery and the *H. pylori* type 4 secretion system.

Sec is the 21st amino acid of the genetic code and is encoded by the UGA stop codon. Thus, Sec decoding requires a specialized translation machinery and is driven by transient interactions between RNA and proteins. Although the main factors of the machinery are now identified, many molecular details remain to be clarified. In order to characterize the interactions between all the known factors, we took advantage of the BRET (*Bioluminescence Resonance Energy Transfer*) assay. Our *in vivo* interaction studies allowed us to define the oligomeric state of several factors of the machinery and to identify previously unknown interactions. Sites of interaction between several factors of the machinery were also determined using *in vitro* approaches. Thus, we propose here a more detailed model of the steps leading to the Sec decoding. In addition, we propose the first structural model of a key protein of the machinery (SECp43), that could be later exploited to elucidate the function of the protein in the translation machinery.

Then we focused on the type 4 secretion system of *Helicobacter pylori*. The *H. pylori* bacterium is a human pathogen whose presence increases the risk of gastric carcinogenesis. The type 4 secretion system encoded by some strains constitute an aggravating factor of pathogenicity. Due to the emergence of multidrug-resistant *H. pylori* strains, current treatments to eradicate bacteria become unsuitable. Thus, we propose to adopt an alternative strategy to antibiotics and undertook to develop and characterize molecules capable of inhibiting the virulence of the bacterium. The 1G2 molecule, that targets the Cag α ATPase protein, was identified. The 1G2 molecule inhibits the ATPase activity of Cag α and decreases by almost 50% the proinflammatory response of eukaryotic cells induced by the *H. pylori* infection. The

effectiveness of the molecule was subsequently evaluated on drug-resistant clinical strains of *H. pylori*, genetically and biochemically characterized. We have demonstrated that the 1G2 molecule was also effective on highly virulent clinical strains. As chronicity of *H. pylori* infection may lead to the development of gastric cancers, our results suggest the high potential of the 1G2 molecule in the treatment of *H. pylori* infection. This molecule could serve as a basis for the design of other molecules having a more noticeable effect on the virulence of *H. pylori*.

Keywords : Macromolecular complexes, selenocysteine, selenoprotein, protein interaction, BRET, SEPHS1, SEPHS2, SEPSECS, SECp43, *H. pylori*, T4SS, virulence inhibitors, CagL, CagA, Cag α

TABLE DES MATIÈRES

Résumé	ii
ABSTRACT	iv
Table des matières	vi
Liste des tableaux	xiii
Liste des figures	xiv
Liste des Sigles et Abréviations	xviii
Remerciements	xxiii
Chapitre I : Introduction	1
Avant-propos: Les complexes macromoléculaires au cœur des processus cellulaires	1
I.1. Biosynthèse et incorporation de la sélénocystéine dans les protéines	2
1.1 Le sélénium et ses fonctions biologiques	2
1.1.1 Sélénium, découverte et importance physiologique	2
1.1.2 La sélénocystéine : un avantage catalytique	3
1.1.3 Le sélénoprotéome eucaryote	4
1.1.1.1 Glutathionne peroxydase	4
1.1.1.2 Thiorédoxine réductase	4
1.1.1.3 Iodothyronine déiodinase.....	5
1.1.1.4 Sélénoprotéine P	5
1.1.1.5 Sélénophosphate synthétase 2.....	5
1.1.1.6 Sélénoprotéine N.....	5
1.1.1.7 Autres sélénoprotéines eucaryotes	6
1.2 Synthèse de la sélénocystéine	6
1.2.1 L'ARNt ^{Sec} , un élément clé de la synthèse des sélénoprotéines	7
1.2.2 Synthèse de la Sec sur son ARNt.....	8
1.2.2.1 Le modèle procaryote	8

1.2.2.2	Le modèle eucaryote	9
1.3	Recodage du codon stop UGA en Sec et incorporation dans les protéines	11
1.3.1	Les éléments de structure d'ARN	11
1.3.1.1	L'élément SECIS : un élément déterminant du décodage	11
1.3.1.2	L'élément SRE : un élément de régulation supplémentaire.....	14
1.3.2	Les protéines liant l'élément SECIS	14
1.3.2.1	SBP2 : une protéine polyvalente.....	14
1.3.2.2	EFSec : le facteur d'élongation spécifique des sélénoprotéines	16
1.3.2.3	La protéine ribosomale L30 : concurrente de SBP2 pour la liaison à l'élément SECIS.....	17
1.3.3	SECp43 : formation, stabilité et localisation des complexes supramoléculaires...	17
1.3.4	Modèles de l'incorporation de la Sec chez les eucaryotes.....	18
1.3.5	L'altération des composants de la machinerie de synthèse et de traduction de la Sec affecte l'intégrité de l'organisme	19
1.4.	Régulation de la synthèse des sélénoprotéines.....	21
1.4.1	Assemblage nucléaire : la formation de RNPm permet d'éviter la dégradation par NMD	21
1.4.2	Hierarchisation de la synthèse des sélénoprotéines	22
1.4.2.1	Facteur d'initiation de la traduction eucaryote eIF4a3	22
1.4.2.2	Nucléoline	22
1.4.3	Position du codon UGA et efficacité d'insertion de la Sec	22
I.2.	Le système de sécrétion de type 4 d'<i>Helicobacter pylori</i>	24
2.1.	<i>H. pylori</i> : la bactérie pathogène humaine colonise l'estomac	24
2.1.1	Prévalence géographique et transmission	24
2.1.2	Pathologies gastriques sévères liées à l'infection par <i>H. pylori</i>	25
2.1.2.1	Ulcère gastroduodéal	25
2.1.2.2	Adénocarcinome gastrique.....	25
2.1.2.3	Lymphome MALT.....	26
2.1.3	Persistance d' <i>H. pylori</i> dans l'organisme	26

2.1.4	Éradication d' <i>H. pylori</i> et émergence des résistances	27
2.2	Le CagSST4 : facteur de pathogénicité d'<i>H. pylori</i>	28
2.2.1	Versatilité des SST4 et fonctions	28
2.2.2	Le <i>cagPAI</i> encode un pilus CagSST4 chez <i>H. pylori</i>	29
2.2.3	CagA : un composant énergétique du CagSST4.....	35
2.2.4	Le pilus CagSST4 permet l'adhésion à la cellule hôte et l'injection des effecteurs	36
2.2.4.1	Intégrines : les récepteurs du CagSST4	37
2.2.4.2	CagL : une adhésine du CagSST4	37
2.2.4.3	CagA : la protéine effectrice du CagSST4.....	38
2.2.4.3.1	CagA participe à sa propre translocation dans la cellule hôte	38
2.2.4.3.2	CagA : une protéine polymorphe.....	39
2.2.4.3.3	CagA détourne les voies de signalisation cellulaire de la cellule hôte	40
2.2.4.4	Autres facteurs transloqués par le CagSST4: peptidoglycane, ADN et HBP.	44
2.3	Facteurs de virulence indépendants du CagSST4.....	44
2.3.1	VacA (<i>Vacuolating cytotoxin</i>)	44
2.3.2	OMPs (<i>Outer membrane proteins</i>)	46
2.3.3	Autres facteurs de virulence	46
2.4	Recherche sur les molécules d'anti-virulence, génotypage de souches d'<i>H. pylori</i> et marqueurs de pathogénicité.....	47
2.4.1	Développement des molécules d'anti-virulence comme alternative thérapeutique.....	47
2.4.2	Génotypage des souches <i>H. pylori</i> et marqueurs de pathogénicité.....	48
I.3.	Objectifs de la thèse.....	50
Chapitre II : Résultats	51
II.1	Article 1 - Analyse de nouvelles interactions entre les composants de la voie de biosynthèse de la sélénocystéine SEPHS1, SEPHS2, SEPSECS et SECp43	51
II.1.1	Présentation de l'article 1 et contribution des coauteurs.....	51
II.1.2	Abbreviations	54

II.1.3 Abstract	54
II.1.4 Introduction	55
II.1.5 Materials and Methods	56
II.1.5.1 Construction of expression vectors	56
II.1.5.2 Construction of expression plasmids for protein variants	57
II.1.5.3 Cell culture and transfection	57
II.1.5.4 BRET assay	57
II.1.5.5 Co-immunoprecipitation	58
II.1.5.6 Cloning, expression and purification of SECp43 ¹⁸⁰	59
II.1.5.7 Analysis of SECp43 dimerization by crosslinking	60
II.1.5.8 Small Angle X-ray Scattering data collection and analysis	60
II.1.5.9 Phage display.....	61
II.1.6 Results	62
II.1.6.1 Identification of interactions between components of the Sec machinery via BRET analysis	62
II.1.6.2 Co-immunoprecipitation confirms the interactions between SEPHS1, SEPHS2 and SEPSECS	65
II.1.6.3 Structural analysis of the N-terminal part of SECp43 by SAXS	66
II.1.6.4 Identification of a peptide required for the dimerization of SECp43 ¹⁸⁰	68
II.1.7 Discussion	70
II.1.8 Acknowledgements	72
II.1.9 Supporting information	73
II.1.10 References	80
II.2 Article 2 - Criblage et identification de molécules inhibant l'activité ATPase et la formation d'hexamère de CagA du système de sécrétion de type IV d'<i>H. pylori</i>	84
II.2.1 Présentation de l'article 2 et contribution des coauteurs.....	84
II.2.2 Abstract	87
II.2.3 Introduction	87
II.2.4 Methods.....	89

II.2.4.1 Bacterial strains, cell lines and culture conditions	89
II.2.4.2 Cloning, expression and purification of Cag α	90
II.2.4.3 Analytical gel filtration chromatography	90
II.2.4.4 Enzyme activity assay	91
II.2.4.5 IC ₅₀ determination	91
II.2.4.6 Differential scanning fluorimetry (DSF).....	91
II.2.4.7 Crystallisation and structure determination.....	92
II.2.4.8 Analysis of protein-protein interactions by cross-linking	92
II.2.4.9 Electron microscopy and image processing.....	92
II.2.4.10 Measurement of <i>H. pylori</i> and AGS cell viability	93
II.2.4.11 Assay for monitoring CagA transfer into AGS cells.....	93
II.2.4.12 Assay for IL-8 induction	93
II.2.5 Results and Discussion.....	94
II.2.5.1 Differential scanning fluorimetry to identify Cag α -binding fragments	94
II.2.5.2 Effects of binding fragments on the Cag α ATPase activity	95
II.2.5.3 Binding fragments impact the conformation and dissociate the Cag α hexamer	97
II.2.5.4 X-ray analysis reveals the 1G2 binding site and conformational changes.....	100
II.2.5.5 Molecule 1G2 inhibits the production of interleukin-8 upon binding of <i>H. pylori</i> to AGS cells	102
II.2.6 Acknowledgements	104
II.2.7 Supplementary content.....	106
II.2.8 References	114
II.3 Article 3 - Analyse génétique et biochimique de souches d'<i>H. pylori</i> résistantes aux antibiotiques et inhibition de la virulence par la molécule 1G2	118
II.3.1 Présentation de l'article 3 et contribution des coauteurs.....	118
II.3.2 Abstract	120
II.3.3 Introduction	120

II.3.4 Material and Methods	123
II.3.4.1 Bacterial strains, cell lines and culture conditions.	123
II.3.4.2 Extraction of the bacterial genomic DNA.....	123
II.3.4.3 Assessing the levels of Cag proteins produced in <i>H. pylori</i> strains	123
II.3.4.4 Imaging the AGS cell Hummingbird phenotype by interference contrast microscopy and quantification.....	123
II.3.4.5 Assay for measuring IL-8 secretion by AGS epithelial cells.....	124
II.3.5 Results.....	124
II.3.5.1 Analysis of the <i>cagPAI</i> in <i>H. pylori</i> strains from patient biopsies.....	124
II.3.5.2 Analysis of the variability of the CagA C-terminal domain	125
II.3.5.3 Sequence variations of CagL.....	126
II.3.5.4 Detection of Cag proteins production	127
II.3.5.5 Effect of infection with <i>H. pylori</i> on cytoskeletal changes	128
II.3.5.6 Analysis of IL-8 induction after co-cultivation with AGS cells	130
II.3.5.7 The anti-virulence 1G2 molecule is efficient on the clinical <i>H. pylori</i> strains	130
II.3.5.8 Variation in the Cag α peptide sequence of the 3822 strain reduced 1G2 molecule efficiency.....	131
II.3.6 Discussion	132
II.3.7 Supplementary data.....	134
II.3.8 References	138
Chapitre III : Discussion	0
III.1 Identification de nouvelles interactions dans la machinerie Sec par BRET	0
III.2 Analyse structurale du domaine N-terminal de SECp43 et identification de son site de dimérisation.....	4
III.3 Caractérisation d'inhibiteurs de l'activité ATPase de Cag α d' <i>H. pylori</i> : 1G2, une molécule prometteuse	5
III.4 Rôle présumé de Cag α dans le SST4 d' <i>H. pylori</i>	7
III.5 Intérêt du développement d'inhibiteurs de virulence dans l'éradication d' <i>H. pylori</i> ...	10

III.6 Variabilité génétique et biochimique des isolats <i>H. pylori</i> , conséquences sur la virulence et inhibition de la virulence par 1G2.....	12
III.7 Étude de la localisation de CagL par microscopie à fluorescence.....	14
Conclusion	16
Bibliographie	17

LISTE DES TABLEAUX

Chapitre I - Introduction

Tableau I.1 Sélénoprotéines eucaryotes.. 6

Tableau I.2. Protéines encodées par le *cagPAI* et homologues avec le système VirB/D d'*A. tumefaciens*. 30

Chapitre II - Résultats

Table II.1.1 Protein interactions assessed by BRET and their BRET₅₀..... 64

Table II.1.2. SAXS data collection and scattering-derived parameters..... 67

Suppl. Table SII.1.1. Expression vectors 73

Suppl. Table SII.1.2. Oligonucleotide primers for the construction of expression vectors 73

Table II.2.1: Structures and IC₅₀ of molecules that inhibit the ATPase activity of Cag α 96

Suppl. Table SII.2.1: Data collection and refinement statistics. 106

Table II.3.1. *cagPAI* positive and negative *H. pylori* strains..... 125

Table II.3.2. Characteristics of the CagA isoforms isolated from the *H. pylori* strains 126

Suppl. Table SII.3.1. *H. pylori* strains and resistance to antibiotics 134

Suppl. Table SII.3.2. Primers used for PCR amplification and sequencing 135

Chapitre III - Discussion

Tableau III.1. Interactions identifiées dans cette étude et dans des travaux précédents 2

Tableau III.2. Synthèse des données de caractérisation génétique et biochimique des souches d'*H. pylori*..... 12

LISTE DES FIGURES

Chapitre I - Introduction

Figure I.1. Structures chimiques de la cystéine et de la sélénocystéine	3
Figure I.2. Comparaison de la structure secondaire de l'ARNt canonique à celle des ARNt ^{Sec} bactérien et eucaryote..	7
Figure I.3. Représentation schématique de la biosynthèse bactérienne de la Sec sur l'ARNt ^{[Ser]Sec}	9
Figure I.4. Structures des Sec synthases eucaryote et procaryote en interaction avec l'ARNt ^{[Ser]Sec}	10
Figure I.5. Représentation schématique de la biosynthèse eucaryote de la Sec sur son ARNt ^{[Ser]Sec}	11
Figure I.6. Représentation schématique de la localisation de l'élément SECIS dans l'ARNm de sélénoprotéines eucaryotes et procaryotes.	12
Figure I.7. Représentation schématique des éléments SECIS eucaryotes et procaryotes.....	13
Figure I.8. Représentation schématique des domaines de la protéine SBP2.	15
Figure I.9. Représentation schématique des domaines du facteur d'élongation spécifique eucaryote EFSec.....	16
Figure I.10. Modèles alternatifs d'incorporation de la Sec dans les protéines chez les eucaryotes.	19
Figure I.11. L'altération des composants impliqués dans la synthèse et le décodage de la Sec a des conséquences multiples sur l'organisme humain.	20
Figure I.12. Fréquence de résistance des souches <i>H. pylori</i> aux antibiotiques dans les diverses régions du monde.....	28
Figure I.13. Représentation de l'organisation des gènes <i>cag</i> dans l'îlot de pathogénicité <i>cagPAI</i> d' <i>H. pylori</i>	29
Figure I.14. Analyse du complexe cœur du CagSST4 par microscopie électronique.	31
Figure I.15. Pili observés suite au contact direct d' <i>H. pylori</i> avec les cellules épithéliales gastriques <i>AGS in vitro</i>	32
Figure I.16. Représentation schématique de l'architecture du CagSST4 d' <i>H. pylori</i> encodé par le <i>cagPAI</i>	34

Figure I.17. Représentation de la structure cristallographique de Cag α	36
Figure I.18. Diversité structurale des motifs EPIYA de CagA.....	40
Figure I.19. Schématisation des voies de signalisations cellulaires activées dans la cellule hôte par CagA phosphorylée et non phosphorylée.	42
Figure I.20. Représentation simplifiée des conséquences de l'injection de CagA sur les cellules gastriques.	43
Figure I.21. Organisation de VacA et régions polymorphes.	45

Chapitre II - Résultats

Article 1

Figure II.1.1 Oligomerization profiles of Sec machinery components measured by BRET... 63	63
Figure II.1.2. Interaction analyzes of Sec machinery components by BRET.	64
Figure II.1.3. Analysis of SEPSECS, SEPHS1 and SEPHS2 interactions by co-immunoprecipitation.....	66
Figure II.1.4. Structural modelling of SECp43 ¹⁸⁰ based on SAXS-WAXS data.	68
Figure II.1.5. Phage display to identify the dimerization site of SECp43.....	69
Figure II.1.6. BRET analysis of SECp43 variants.....	70
Suppl. Table SII.1.3. Oligonucleotide primers for directed-site mutagenesis	75
Suppl. Table SII.1.4. Quantitative BRET measurement of SECp43 variants.....	75
Suppl. Fig. SII.1.1. Measurement of the interaction between G β 1 and G γ 1 subunits of the G-protein by BRET shows a strong affinity.	76
Suppl. Fig. SII.1.2. SEPSECS-SEPHS1 interaction BRET profiles..	77
Suppl. Fig. SII.1.3. <i>In vitro</i> dimerization of purified SECp43 in solution.	78
Suppl. Fig. SII.1.4. Structural representation of the residues involved in SECp43 dimerization.	78
Suppl. Fig. SII.1.5. Potential interaction site between SECp43 and SEPSECS identified by phage display..	79

Article 2

Figure II.2.1. Melting temperature of Cag α in the presence of ligands and cofactors..	95
---	----

Figure II.2.2. Enzyme Kinetics of Cag α in the presence of molecule 1G2 and 1G2#4.....	97
Figure II.2.3. Chemical cross-linking using DSS to study the formation of Cag α oligomers in the presence of ligands.....	98
Figure II.2.4. Analytical size exclusion chromatography of Cag α apoprotein and in the presence of ligands.....	99
Figure II.2.5. Electron micrographs of negatively stained Cag α apoprotein after gel filtration..	100
Figure II.2.6. Crystal structure of the Cag α bound to molecule 1G2.....	102
Figure II.2.7. Molecule 1G2 decreases IL-8 induction in co-cultivated AGS cells..	103
Suppl. Figure SII.2.1. Differential scanning fluorimetry screening of fragment libraries to identify Cag α binding molecules.....	107
Suppl. Figure SII.2.2. Structures of 16 Cag α -stabilizing molecules identified by differential scanning fluorimetry.	108
Suppl. Figure SII.2.3. Changes of melting temperatures in the presence ADP and Mg ²⁺ after co-incubation with 16 Cag α -stabilizing molecules..	108
Suppl. Figure SII.2.4. Dose response curves of ATPase activity showing IC ₅₀ values in the presence of four molecules that inhibit the enzyme activity.....	109
Suppl. Figure SII.2.5. Multiple sequence alignment of Cag α /VirB11 homologs.....	110
Suppl. Figure SII.2.6. Crystal structures of the Cag α -1G2 complex compared to previously determined structures of Cag α	111
Suppl. Figure SII.2.7. Effects of molecule 1G2 and its derivates on <i>H. pylori</i> growth.....	111
Suppl. Figure SII.2.8. Effects of molecule 1G2 and its derivatives on the viability of AGS cells.	112
Suppl. Figure SII.2.9. CagA phosphorylation in the presence of molecule 1G2 and its derivates.	112
Suppl. Figure SII.2.10. Overlay of 1G2#4 modeled over the Cag α -1G2 complex..	113
 <i>Article 3</i>	
Figure II.3.1. Alignment of the CagL sequences from the <i>H. pylori</i> isolates.....	127
Figure II.3.2. Comparison of the levels of Cag proteins in <i>H. pylori</i> strains.	128

Figure II.3.3. AGS Hummingbird phenotype induced by <i>H. pylori</i> infection.	129
Figure II.3.4. Interleukin-8 production by AGS cells infected by the <i>H. pylori</i> strains.....	130
Figure II.3.5. Molecule 1G2 decreases interleukin-8 production by AGS cells upon co-cultivation with the <i>H. pylori</i> clinical strains.....	131
Figure II.3.6. Sequence alignment of the Cag α isoforms from 26695 and 3822 strains shows variation in the 1G2 molecule binding site.....	132
Suppl. Figure SII.3.1. PCR amplification of <i>cagA</i> and <i>cagL</i> genes in <i>H. pylori</i> isolates.	135
Suppl. Figure SII.3.2. Alignment of the sequences of CagA comprising the 3 EPIYA motifs.	136
Suppl. Figure SII.3.3. Alignment between the sequences of CagA isoforms from <i>H. pylori</i> 26695 and ATCC 43504 strains.....	136
Suppl. Figure SII.3.4. Comparison of production levels of Cag proteins in the <i>H. pylori</i> strains.	137

Chapitre III - Discussion

Figure III.1. Principe du BRET.	1
Figure III.2. Modèles illustrant les rôles hypothétiques de Cag α et des autres ATPases dans la translocation des effecteurs du CagSST4 à travers la membrane interne.....	9

LISTE DES SIGLES ET ABRÉVIATIONS

ADN : acide désoxyribonucléique

AGS : adenocarcinoma gastric cells

ARN : acide ribonucléique

ARNm : ARN messenger

ARNr : ARN ribosomal

ARNsn : petits ARN nucléaires

ARNt : ARN de transfert

A. tumefaciens : *Agrobacterium tumefaciens*

BabA : blood group antigen-binding adhesin

BRET: bioluminescence resonance energy transfer

cagPAI : *cag* pathogenicity island

CEACAM1 : carcinoembryonic antigen-related cell adhesion molecules

CHUM : centre hospitalier de l'université de Montréal

CIRC/OMS : centre international de recherche sur le cancer/ organisation mondiale de la santé

Cys : cystéine

CM : CagA multimerization

DIO : iodothyronine déiodinase

DSF : differential scanning fluorimetry

dupA : duodenal ulcer-promoting gene A

E. coli : *Escherichia coli*

EJC : exon junction complexes

EMT : epithelial-mesenchymal transition

GFP : green fluorescent protein

GPCR : G protein-coupled receptors

GPx : glutathionne peroxydase

HBP : heptose-1,7-bisphosphate

HEK : human embryonic kidney

H. pylori : *Helicobacter pylori*

IPP: inhibiteur de pompe à protons

LPS : lipopolysaccharide

MALT : mucosa-associated lymphoid tissue

MAP: mitogen-activated protein kinase

Msr : methionine sulfoxide reductase

NapA : neutrophil activating protein A

NES : nuclear export signal

NLS : nuclear localization signal

NMD : nonsense-mediated decay

NOD1 : nucleotide-binding Oligomerization Domain-containing protein 1

OipA : outer membrane inflammatory protein

OMP : outer membrane proteins

PCCA : atrophie cérébrale progressive du cervelet

PI3K : phosphoinositide 3-kinase

PLP: pyridoxal phosphate

PS : phosphatidylserine

PSTK : phosphoséryl-ARNt kinase

RBD : RNA-binding domain

RLuc : *Renilla* luciférase

RMN : résonance magnétique nucléaire

RNPm : ribonucléoprotéine particules messenger

RRM : RNA recognition motif

SabA : sialic acid-binding adhesin

SBP2 : SECIS binding protein 2

Se : sélénium

Sec : sélélocystéine

SECIS : selenocysteine insertion sequence

SecS : Sec synthase

SeIN : séléloprotéine N

SeIP : séléloprotéine P

SeMet : sélélométhionine

SeP : monosélélophosphate

SEPHS : sélélophosphate synthétase

Ser : sérine

SerS : séryl-ARNt-synthétase

SID : selenocysteine insertion domain

SIM: structure illumination microscopy

siRNA : small interfering RNA

SLA/LP : soluble liver antigen/liver-pancreas

SNP: single nucleotide polymorphism

SRE : sec redefinition element

SST4: système de secretion de type 4

TIFA : TRAF-interacting protein with forkhead associated domain

TLR :Toll-like receptors

TNF: tumor necrosis factor

Tr : thiorédoxine reductase

TRAF : TNF-receptor associated factor

TRG: thiorédoxine glutathione réductase

UTR : untranslated region

VIH : virus de l'immunodéficience humaine

À ma sœur, à mes parents.

"Toute science crée une nouvelle ignorance. Tout conscient, un nouvel inconscient."

Henri Michaux (1899-1984)

REMERCIEMENTS

Je tiens tout d'abord à remercier Dr Christian Baron, professeur et vice-doyen à l'Université de Montréal, qui m'a supervisé tout au long de mon doctorat. Je le remercie pour son encadrement, le partage des connaissances, pour ses conseils avisés et pour sa confiance.

Je tiens également à remercier les membres de mon comité de thèse, Dre Pascale Legault et Dr Gerardo Ferbeyre, professeurs à l'Université de Montréal, qui ont suivi mon parcours et les remercie pour leurs précieux conseils et encouragements.

Je remercie également les membres de mon jury pour avoir accepté d'évaluer ma thèse et pour le temps consacré.

Je remercie tous les membres du laboratoire du Dr Baron et tout particulièrement Benoit Bessette, Charline Mary, Aurélien Fouillen et Jonathan Smart, pour leur aide, leur soutien et leur amitié. Je remercie également les membres actuels, Jaafar Amro et Zakaria Jemouai, ainsi que tous les anciens membres du laboratoire, Bastien Casu, Mahzad Sharifahmadian, Rodolfo Frazao, Tarun Arya, Etienne Pageau-Crevier, Arnold Puemi Dopgwa, Ingrid Um Nlend et Mathieu Coincon. Merci aussi aux deux étudiantes, Florence Mailhot-Leonard et récemment Lavisha Parab, que j'ai eu la chance d'encadrer.

Merci également aux membres des autres laboratoires de l'Université de Montréal et notamment à Maxime Lalonde, Hadrien Laprade et Alix Salvail que j'ai eu le plaisir de côtoyer. Je remercie tout particulièrement Pierre Bensidoun pour son soutien, son aide précieuse et son implication.

Finalement, je remercie tous les membres de l'Université que je n'ai pas pu citer (étudiants, professeurs, techniciens, enseignants...), avec qui j'ai eu la chance d'interagir au cours de ces dernières années.

Chapitre I : INTRODUCTION

AVANT-PROPOS :

Les complexes macromoléculaires au cœur des processus cellulaires

Dans la cellule, la majorité des processus biologiques sont assurés par des complexes macromoléculaires. Ceux-ci sont principalement composés de protéines et d'acides nucléiques. L'un des exemples les plus éloquents est certainement celui du ribosome, qui se compose de quatre ARNr et de plus de 80 protéines chez les eucaryotes collaborant pour permettre un processus cellulaire fondamental, la synthèse des protéines. L'étude des complexes macromoléculaires, de leur composition, de leur assemblage, de leur dynamique et de leur fonction est une entreprise autant élémentaire que fascinante. Du fait de leur complexité, l'étude de ces complexes soulève de nombreux défis et questions, et requiert souvent l'utilisation d'approches interdisciplinaires, combinant des approches structurales, biochimiques, biophysiques et protéomiques. Les complexes macromoléculaires constituent, d'autant plus, des cibles privilégiées en recherche clinique, leur compréhension étant une étape clé pour le développement de traitements thérapeutiques qui ciblent les composants des complexes.

Ainsi, au cours de mon doctorat, nous nous sommes intéressés à deux complexes macromoléculaires, l'un intervenant dans la traduction eucaryote et l'autre dans la virulence d'une bactérie pathogène : 1) la machinerie de synthèse et d'incorporation des sélénocystéines chez l'humain, caractérisée par des dynamiques d'interactions complexes entre facteurs protéiques et ARN, dans les différents compartiments de la cellule, qui aboutit à la traduction d'une sélénocystéine à partir d'un codon stop ; 2) le système de sécrétion de type 4 d'*H. pylori*, un complexe macromoléculaire de type seringue qui permet la sécrétion d'effecteurs dans les cellules hôtes et à l'origine de la virulence de la bactérie.

I.1. BIOSYNTHÈSE ET INCORPORATION DE LA SÉLÉNOCYSTÉINE DANS LES PROTÉINES

1.1 Le sélénium et ses fonctions biologiques

1.1.1 Sélénium, découverte et importance physiologique

Le sélénium (Se) est un élément trace découvert en 1817 par le chimiste suédois Jöns Jacob Berzelius. Il fut nommé d'après *Selene*, la déesse grecque de la Lune, en référence à sa précédente découverte, le tellurium, de *Tellus*, Terre en latin. Le sélénium fut longtemps considéré comme une puissante substance toxique dont la dangerosité fut d'abord établie sur le bétail. En 1930, le sélénium fut finalement classé comme micronutriment bénéfique pour la santé et, dès 1970, son activité biologique fut imputée à sa présence dans les protéines. Le sélénium est à présent établi comme étant un oligoélément essentiel pour la santé humaine, nécessaire à la prévention des infections virales, de certains cancers et de l'infertilité masculine et important pour le système immunitaire, le développement et le fonctionnement musculaire [1, 2]. Chez l'Homme, la carence en sélénium constitue notamment un facteur de risque menant à la maladie de Keshan, une cardiomyopathie endémique identifiée dans certaines régions de l'est de la Chine, en raison d'un sol particulièrement pauvre en sélénium [3, 4]. La supplémentation en sélénium permet d'éviter ou de pallier cette affection [5]. De nombreuses études démontrent également l'effet protecteur du sélénium dans des maladies inflammatoires, et dans la progression du VIH chez des patients infectés, ainsi que l'impact significatif de la supplémentation sur la réduction de la mortalité de patients atteints de cancers de la prostate, du côlon ou des poumons [1, 2, 6-10]. Bénéfique à faibles doses, de hautes concentrations en sélénium entraînent, néanmoins, son accumulation dans la cellule, conduisant à un stress oxydatif qui endommage les composants cellulaires [11].

Les fonctions physiologiques qui lui sont imputées, et notamment son activité antioxydante, sont liées à sa présence dans les sélénoprotéines, lesquelles possèdent dans leurs chaînes peptidiques des aminés contenant du sélénium organique tels que la sélénocystéine (Sec) ou la sélénométhionine (SeMet) [12].

1.1.2 La sélénocystéine : un avantage catalytique

Au sein des protéines, le sélénium est majoritairement présent dans la Sec, considérée comme le 21^e acide aminé du code génétique. Les Sec sont retrouvées au site actif des sélénoprotéines où elles jouent un rôle catalytique, et sont généralement impliquées dans des réactions d'oxydoréduction. Analogue de la cystéine (Cys), la structure chimique de la Sec diffère exclusivement par la présence d'un atome de sélénium dans son radical alors que la Cys possède un atome de soufre (Fig. I.1).

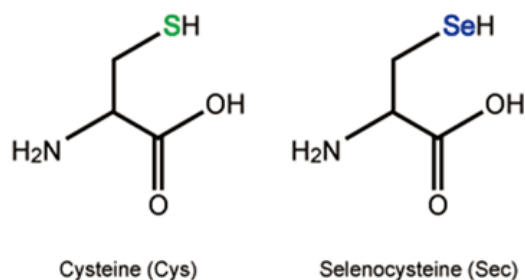


Figure I.1. Structures chimiques de la cystéine et de la sélénocystéine

Cette différence a des conséquences sur les propriétés chimiques et physiques de cet acide aminé. La Sec, dont le pK_a est inférieur à celui de la Cys (Sec : pK_a=5,2 ; Cys : pK_a=8,5), est déprotonée à pH physiologique. De ce fait, Sec présente une nucléophilie et une réactivité accrue par rapport à la Cys. Ainsi, les homologues des sélénoenzymes contenant une Cys à la place de la Sec sont généralement des catalyseurs plus faibles, le changement de Sec en Cys résultant en une diminution de l'activité catalytique de 10 à 100 fois [12-14]. De façon intéressante, des études ont mis en évidence qu'en cas de carence en sélénium, la machinerie de biosynthèse de la Sec pouvait être détournée pour produire un Cys-ARNt et ainsi remplacer la Sec d'origine par une Cys [15], produisant une enzyme dont l'activité catalytique, pourtant plus faible, permet de compenser l'absence de l'homologue sélénoenzyme. Ce processus n'est cependant pas à généraliser et serait spécifique de certaines sélénoprotéines seulement. Ainsi, il est proposé que Sec, dont la synthèse est coûteuse d'un point de vue énergétique, serait conservée dans certains organismes car catalytiquement supérieure à la Cys.

1.1.3 Le sélénoprotéome eucaryote

Bien que retrouvé dans les trois domaines du vivant (bactéries, archaea et eucaryotes) et observé dans certains virus, le sélénoprotéome des organismes varie grandement d'une espèce à l'autre, oscillant d'une sélénoprotéine pour le nématode *Caenorhabditis elegans* à 59 pour l'algue *Aureococcus anophagefferens* [16, 17]. Alors que 25 sélénoprotéines sont dénombrées chez les mammifères [18], les plantes supérieures, champignons, et certaines espèces animales en sont dépourvus. Les analyses comparatives de sélénoprotéomes provenant de divers organismes montrent, en effet, que dans plusieurs groupes d'organismes terrestres, dont les mammifères, le nombre de sélénoprotéines est réduit ou nul, certaines ayant été remplacées par des homologues contenant une Cys au niveau du site actif. Au contraire, la plupart des organismes aquatiques présentent un grand nombre de sélénoprotéines, ce qui suggère une pression environnementale importante sur l'évolution des sélénoprotéomes [19, 20].

Retrouvées au site actif des sélénoprotéines, les Sec sont impliquées dans les réactions d'oxydoréduction. Les sélénoprotéines détiennent ainsi une activité d'oxydoréductase et interviennent, entre autres, dans la défense antioxydante, l'homéostasie redox ou encore la signalisation redox (Tableau I.1).

1.1.1.1 Glutathionne peroxydase

GPx1 est la sélénoprotéine la plus abondante chez les mammifères, elle fut d'ailleurs la première sélénoprotéine identifiée [21]. Les GPx sont connues pour leur rôle d'antioxydants. Elles induisent notamment la réduction du peroxyde d'hydrogène et des hydroperoxydes organiques, protégeant les cellules contre les dommages oxydatifs. Chez l'Homme, sept GPx sont connues, cinq sont des sélénoenzymes et deux sont des variants contenant une Cys au site catalytique. Ces enzymes diffèrent dans leurs distributions tissulaires et leur spécificité de substrats.

1.1.1.2 Thiorédoxine réductase

Les thiorédoxine réductases sont des sélénoprotéines qui catalysent la réduction de la thiorédoxine de façon NADPH-dépendante ; elles agissent alors comme des régulateurs de son activité métabolique. Les thiorédoxine réductases semblent jouer un rôle important dans la prévention de certaines formes de cancer [22].

1.1.1.3 Iodothyronine déiodinase

La famille des iodothyronine déiodinases (DIO) est constituée de trois enzymes (DIO1, DIO2, DIO3) qui catalysent l'activation (DIO1 et DIO2) et l'inactivation (DIO3) des hormones thyroïdiennes T4 (thyroxine), T3 (triiodothyronine) et rT3 (reverse T3). Le rôle de ces sélénoprotéines est fondamental dans le maintien de l'homéostasie des hormones thyroïdiennes et donc pour le développement de l'organisme. En effet, les hormones thyroïdiennes sont impliquées dans de nombreux processus métaboliques, tels que la thermogénèse, la croissance et l'audition, et sont indispensables pour le développement normal du cerveau fœtal. Elles contrôlent l'expression des gènes dans tous les tissus de vertébrés via des activateurs de transcription de la famille des récepteurs de l'hormone thyroïdienne [22].

1.1.1.4 Sélénoprotéine P

La fonction principale de la sélénoprotéine P (SelP) consiste au transport et à la livraison du sélénium dans l'ensemble des tissus. Elle contient à elle seule 60% du sélénium du plasma humain, incorporé dans ses dix Sec. Un défaut dans les niveaux de SelP affecte le cerveau et l'appareil reproductif masculin, et dans une moindre mesure, le bon fonctionnement du cœur et des reins [22].

1.1.1.5 Sélénophosphate synthétase 2

La sélénophosphate synthétase 2 (SEPHS2) est la sélénoprotéine responsable de la génération du sélénophosphate chez l'humain, le donneur actif de sélénium. Elle est ainsi impliquée dans la biosynthèse de la Sec (*cf.* section 1.2.2.2).

1.1.1.6 Sélénoprotéine N

La sélénoprotéine N (SelN) est la seule sélénoprotéine à avoir été directement reliée à des maladies génétiques. Des mutations dans le gène *SEPNI* codant pour SelN sont associées à diverses myopathies (référées comme *SEPNI-related myopathies*) [23]. Ces mutations, localisées dans la région 3' non codante (3'UTR, 3' *UnTranslated Region*) de l'ARNm de SelN, au niveau d'un élément régulateur de la synthèse des sélénoprotéines appelé SECIS (pour *SElenoCysteine Insertion Sequence*) (*cf.* section 1.3.1.1), sont à l'origine d'une réduction du

niveau de production de la protéine SelN dans les cellules. La fonction exacte de SelN dans le métabolisme reste néanmoins méconnue [22].

1.1.1.7 Autres sélénoprotéines eucaryotes

Les autres sélénoprotéines identifiées chez les eucaryotes sont illustrées dans le Tableau I.1.

Tableau I.1 Sélénoprotéines eucaryotes. Inspiré de Rederstorff *et al.*, 2006 [24].

Fonctions		Représentation du nombre et de la localisation des Sec dans les sélénoprotéines
DOI1	Maturation ou catabolisme des hormones thyroïdiennes (T4, T3 et rT3)	
DOI2		
DOI3		
GPx1	Protection antioxydante	
GPx2		
GPx3	Maintenance du statut redox cellulaire	
GPx4	Détoxification des lipides hydroperoxydes	
GPx6	Protection antioxydante	
TR1	Métabolisme de la thiorédoxine, défense antioxydante, régulation redox, signalisation cellulaire	
TR2		
TR3		
SEPHS2	Synthèse de Sec	
Sep15	Apoptose et médiation des effets chimiopréventifs du Se	
SelN	Associée aux maladies musculaires	
SelW	Protection antioxydante	
SelP	Transport du Se plasmatique, défense antioxydante	
SelH	Inconnue	
SelI	Inconnue	
SelJ	Inconnue (poisson et oursin)	
SelK	Inconnue	
SelL	Inconnue	
SelM	Inconnue	
SelO	Inconnue	
SelS	Inconnue, balance redox	
SelT	Rôle dans la régulation de l'homéostasie du Ca ²⁺ et de la sécrétion neuroendocrinienne	
SelU	Inconnue (poisson, oiseau, algues)	
SelV	Inconnue	
MsrA	Réduction des résidus Met oxydés (<i>C. reinhardtii</i>)	
MsrB	Réduction des résidus Met oxydés	

1.2 Synthèse de la sélénocystéine

Du fait de la dégénérescence du code génétique, la terminaison de la traduction est encodée par les codons UGA, UAA et UAG. Cependant, dans le génome mitochondrial, UGA code également pour le tryptophane [25], tandis que dans le génome nucléaire le codon UGA code

pour la Sec. Ainsi la Sec est reconnue comme le 21^e acide aminé du code génétique [26]. *GPXI* fut le premier gène de mammifère identifié à contenir un codon UGA correspondant à une Sec dans la séquence protéique. Il fut ensuite démontré que la Sec était incorporée dans la protéine GPx1 par une molécule d'ARNt portant un anti-codon complémentaire à UGA [27].

1.2.1 L'ARNt^{Sec}, un élément clé de la synthèse des sélénoprotéines

La biosynthèse de la Sec est un processus unique qui se produit de façon co-translationnelle. Elle prend place directement sur un ARNt spécifique de la Sec, l'ARNt^{[Ser]^{Sec}} (aussi désigné Sec-ARNt), qui est préalablement chargé par une sérine, suite à la conversion de la sérine en Sec [26, 28]. L'ARNt^{[Ser]^{Sec}} est donc décrit comme une molécule clé de la biosynthèse des sélénoprotéines [28, 29]. L'ARNt^{[Ser]^{Sec}} possède un anticodon complémentaire au codon UGA et s'associe avec plusieurs facteurs requis pour la synthèse et l'incorporation de la Sec. L'ARNt^{[Ser]^{Sec}} détient des caractéristiques uniques qui le distinguent des autres ARNt, en termes de longueur, d'aminocyclation, de transcription et de transport. Il s'agit du plus long ARNt identifié jusqu'à ce jour, avec une longueur de 90 nucléotides chez les eucaryotes [30]. Ses bras accepteurs et variables sont également particulièrement longs (Fig. I.2). Ses caractéristiques empêcheraient sa liaison au facteur d'élongation canonique eEF1A et une entrée prématurée au site A du ribosome.

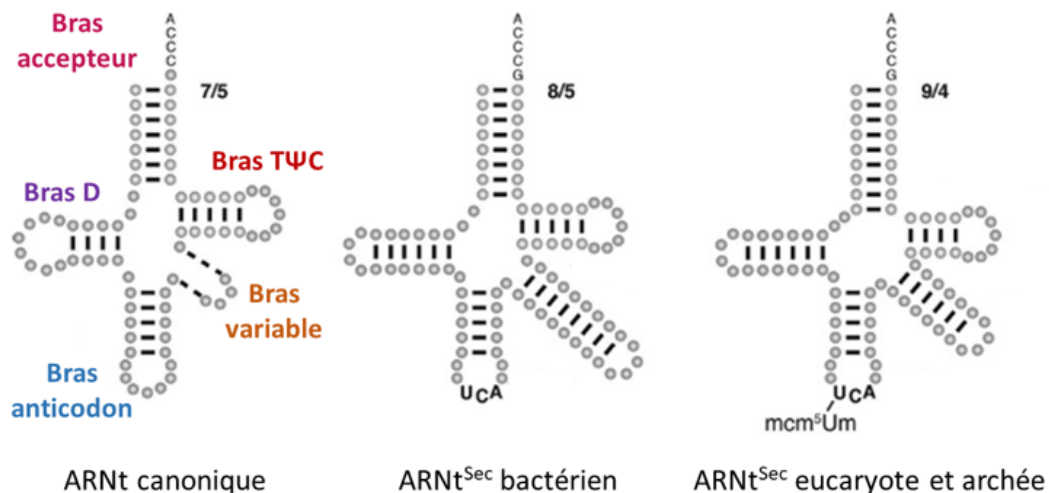


Figure I.2. Comparaison de la structure secondaire de l'ARNt canonique à celle des ARNt^{Sec} bactérien et eucaryote. Adapté de Allamang *et al.*, 2006 [31].

Les génomes des mammifères ne comportent qu'une seule copie du gène *trsp* codant pour l'ARNt^{[Ser]Sec}, dont la délétion est létale au stade embryonnaire chez la souris [32]. Deux isoformes de l'ARNt^{[Ser]Sec} sont produites : une forme non méthylée (mcm⁵U ; 5-méthylcarboxyméthyluridine) et une forme méthylée (mcm⁵Um ; 5-méthylcarboxyméthyluridine-2'-méthylribose) sur l'uracile en position 34 [33, 34]. La distribution et les quantités relatives de chaque isoforme varient en fonction du tissu ou du type cellulaire et de la disponibilité en sélénium [34]. De plus, de façon intéressante, les sélénoprotéines constitutives (ou *housekeeping*) telles que les thiorédoxine réductases TR1 et TR3 sont synthétisées par l'isoforme non méthylée tandis que les sélénoprotéines liées au stress, comme les glutathione peroxydases GPx1 et GPx3 sont synthétisées par l'isoforme méthylée [35]. Néanmoins, certaines sélénoprotéines (GPx4 et Selp) sont synthétisées par les deux isoformes [36]. La méthylase responsable de la méthylation de l'isoforme mcm⁵Um n'a pas encore été identifiée, mais des données suggèrent que la protéine SECp43 serait impliquée dans le processus (*cf.* section 1.3.3).

1.2.2 Synthèse de la Sec sur son ARNt

1.2.2.1 Le modèle procaryote

La première étape de la synthèse de la Sec consiste au chargement d'une sérine sur l'ARNt^{[Ser]Sec}, par une réaction d'aminocyclation catalysée par une séryl-ARNt-synthétase conventionnelle. Le sélénophosphate (SeP), forme active du sélénium, est produit à partir de sélényde et d'ATP suite à une réaction catalysée par une sélénophosphate synthétase unique (SelD chez *E. coli*) [37]. La partie séryl du Ser-ARNt^{[Ser]Sec} est alors convertie en sélénocysteyl-ARNt^{[Ser]Sec} (Sec-ARNt^{[Ser]Sec}) par une Sec synthase (SelA chez *E. coli*), une enzyme pyridoxal-phosphate (PLP)-dépendante, qui incorpore le sélénophosphate, suite à l'élimination de la partie hydroxyle de la sérine pour former un aminoacrylyle intermédiaire. La biosynthèse de Sec fut d'abord établie chez *E. coli*, au début des années 1990 [38] (Fig. I.3).

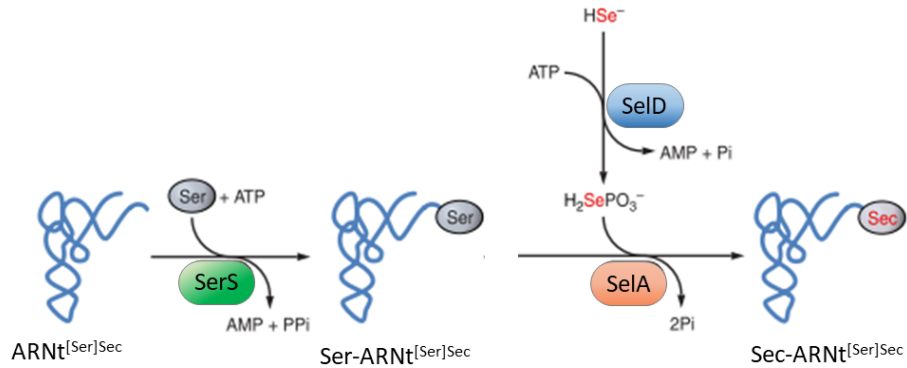


Figure I.3. Représentation schématique de la biosynthèse bactérienne de la Sec sur l'ARNt^[Ser]Sec. SerS : séryl-ARNt-synthétase, SelA : Sec synthase, SelD : sélénophosphate synthétase, H₂SePO₃⁻ : monosélénophosphate

1.2.2.2 Le modèle eucaryote

Il existe des différences fondamentales entre le mécanisme de biosynthèse de Sec des procaryotes, des eucaryotes et des archées. Tout d'abord, en se basant sur les similarités de séquence, il ne fut pas possible d'identifier d'homologue de la Sec synthase bactérienne SelA chez les eucaryotes. En revanche, d'autres facteurs protéiques jouant un rôle dans la biosynthèse de Sec ont été caractérisés chez les mammifères. Chez des patients atteints d'hépatite chronique auto-immune, la protéine alors désignée SLA/LP (pour *Soluble Liver Antigen/Liver-Pancreas*) fut isolée liée à l'ARNt^[Ser]Sec [39]. Elle fut ensuite caractérisée comme faisant partie d'une super-famille de transférases PLP-dépendantes [40], et proposée comme jouant le rôle de la Sec synthase chez les mammifères [31, 40]. Finalement, son rôle fut confirmé lorsqu'elle fut identifiée en complexe avec d'autres facteurs protéiques impliqués dans la synthèse et l'incorporation de la Sec dans les protéines [41, 42]. À présent désignée comme SEPSECS, sa structure cristalline en complexe avec l'ARNt^[Ser]Sec révèle que deux molécules d'ARNt^[Ser]Sec lient quatre sous-unités de la Sec synthase chez l'Homme (Fig. I.4 A et B) [43], tandis que chez la bactérie un homodécamère de SelA lie dix molécules d'ARNt^[Ser]Sec (Fig. I.4C) [44].

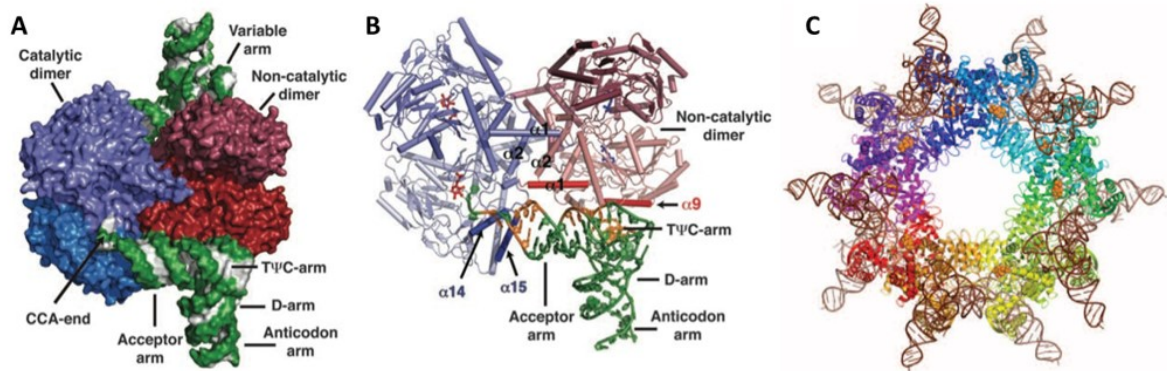


Figure I.4. Structures des Sec synthases eucaryote et procaryote en interaction avec l'ARNt^{Ser}Sec. (A et B) Représentation du complexe SEPSECS humain avec l'ARNt^{Ser}Sec. Deux sous-unités de SEPSECS lient un ARNt^{Ser}Sec selon Palioura *et al.*, 2009 [43]. (C) Représentation du complexe Sela-ARNt^{Ser}Sec. Dix sous-unités de Sela d'*A. aeolicus* lient dix ARNt^{Ser}Sec de *T. tengcongensis* d'après Itoh *et al.*, 2013 [44].

Le second facteur identifié fut la phosphoséryl-ARNt kinase (PSTK) qui phosphoryle spécifiquement la fraction séryl du Ser-ARNt^{Ser}Sec [45, 46]. À la différence des procaryotes, où Ser-ARNt^{Ser}Sec est directement converti en Sec-ARNt^{Ser}Sec, une étape catalytique supplémentaire est requise chez les eucaryotes et les archées. Le Ser-ARNt^{Ser}Sec est d'abord phosphorylé par la PSTK conduisant à la synthèse d'O-phosphoséryl-ARNt^{Ser}Sec [46], qui servira de substrat à la Sec synthase SEPSECS.

Finalement, deux protéines homologues à la sélénophosphate synthétase bactérienne SelD furent identifiées chez les mammifères : SEPHS1 [47, 48] et SEPHS2 [49]. De façon intéressante, seule SEPHS2, elle-même une sélénoprotéine, est capable de synthétiser le sélénophosphate *in vitro* [50]. De plus, tandis que SEPHS2 complémente un mutant déficient en SelD chez *E. coli*, la protéine SEPHS1 en est incapable, sauf dans le cas d'un milieu supplémenté en L-Sec [51]. Ces résultats suggèrent ainsi que SEPHS2 serait nécessaire à la synthèse *de novo* du sélénophosphate tandis que SEPHS1 jouerait un rôle dans le recyclage de la Sec à travers un système de récupération du Se. Le schéma actuel de la synthèse de la Sec eucaryote est résumé dans la figure I.5.

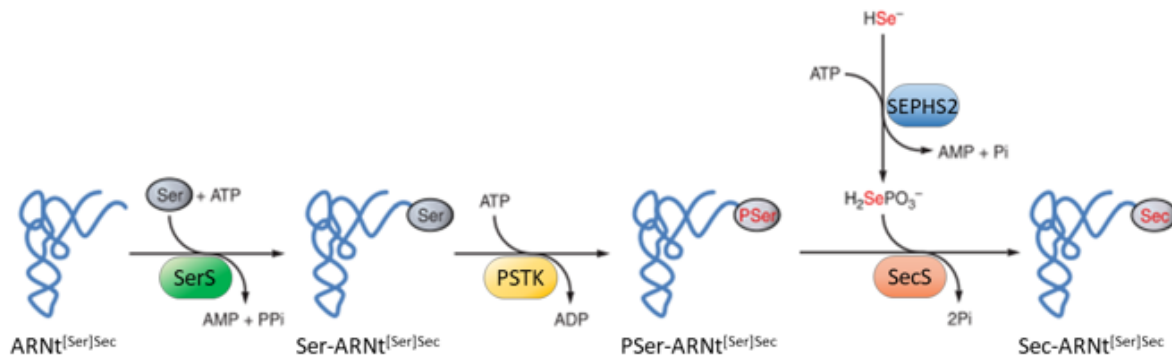


Figure I.5. Représentation schématique de la biosynthèse eucaryote de la Sec sur son ARNt^{[Ser]Sec}. SerS : séryl-ARNt-synthétase, PSTK : phosphoséryl-ARNt kinase, SecS : Sec synthase, SEPHS2 : sélénophosphate synthétase 2, H₂SePO₃⁻ : monosélénophosphate Adapté de Labunskyy *et al.*, 2014 [52].

1.3 Recodage du codon stop UGA en Sec et incorporation dans les protéines

La traduction en Sec est signalée par un codon UGA dans les ARNm des sélénoprotéines mais nécessite la présence d'une machinerie dédiée à son décodage et à son incorporation dans les protéines. Cette machinerie complexe fait intervenir le Sec-ARNt^{[Ser]Sec}, des facteurs protéiques agissant en *trans* ainsi que des éléments agissant en *cis*. Lorsque le ribosome rencontre un codon UGA, la machinerie de décodage agit de concert avec les facteurs impliqués dans la voie canonique pour prévenir une terminaison prématurée de la traduction et permettre le décodage en Sec.

1.3.1 Les éléments de structure d'ARN

1.3.1.1 L'élément SECIS : un élément déterminant du décodage

Le déterminant universel du décodage en Sec est l'élément SECIS (*SElenoCysteine Insertion Sequence*), une structure secondaire de l'ARN d'une centaine de nucléotides et de type tige-boucle, présente sur tous les ARNm de sélénoprotéines. Chez les eucaryotes, l'élément SECIS est localisé dans la région non traduite 3' (3'UTR) des ARNm [53] (Fig. I.6). Une distance minimale de 50 nucléotides entre le codon UGA et l'élément SECIS est d'ailleurs requise pour garantir l'efficacité du recodage ; aucune distance maximale n'a été reportée [54]. Chez les

bactéries, le SECIS est retrouvé dans la région codante du gène, directement en aval du codon UGA codant pour la Sec [55] (Fig. I.6).

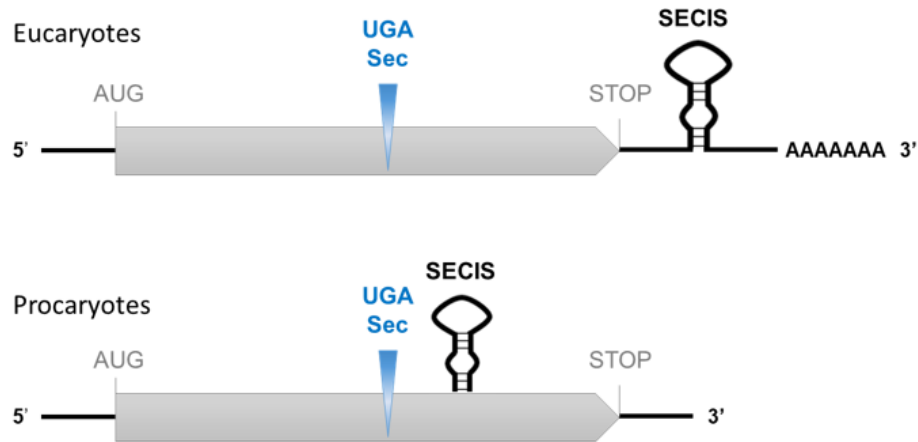


Figure I.6. Représentation schématique de la localisation de l'élément SECIS dans l'ARNm de sélénoprotéines eucaryotes et procaryotes.

Les éléments SECIS présentent peu de similarités de séquence mais leurs structures secondaires sont, quant à elles, hautement conservées et contiennent des séquences consensus indispensables à l'incorporation des Sec. L'élément SECIS est formé de deux hélices (I et II) séparées par une boucle interne, d'un « cœur SECIS » (SECIS *core*) qui contient un quartet GA d'appariement de bases de type non-Watson Crick (dit tandem GA) à la base de l'hélice II et d'une boucle apicale qui contient un motif AAR conservé (Fig. I.7) [56]. Il existe deux types de motif SECIS, la différence entre les deux formes réside dans la présence d'une boucle interne supplémentaire dans le type II [57] (Fig. I.7).

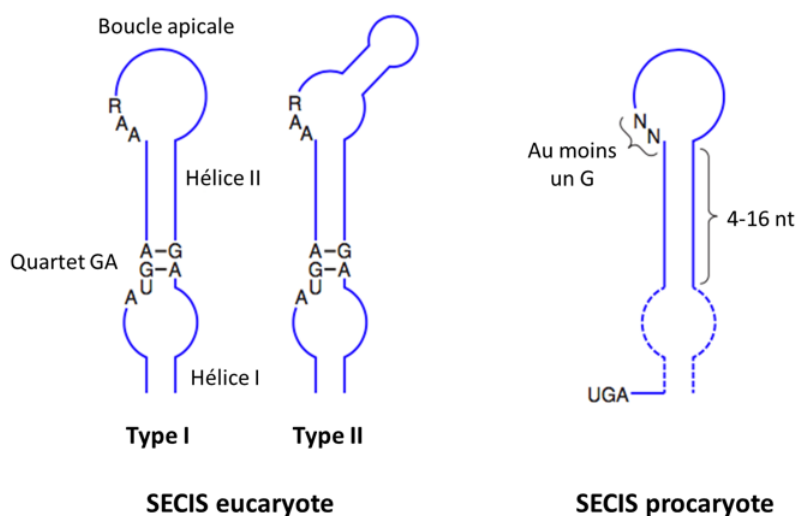


Figure I.7. Représentation schématique des éléments SECIS eucaryotes et procaryotes.

Adapté de Labunskyy *et al.*, 2014 [52].

Tandis que le type II semble prépondérant, l'ARNm encodant une sélénoprotéine peut détenir un élément SECIS de type I ou II, dépendamment de l'espèce ; c'est le cas par exemple de l'ARNm de SelM qui comporte des SECIS de types différents chez les mammifères et le poisson-zèbre [58]. Le quartet GA constitue l'élément fonctionnel principal du SECIS, nécessaire à l'interaction des protéines liant le SECIS (*cf.* section 1.3.2) [59]. Ce motif présente des propriétés caractéristiques des motifs *kink-turn*, capables d'adopter des conformations distinctes en solution, ouverte ou fermée (*kink-turn*), et de lier les protéines de la famille L7Ae, qui interagissent spécifiquement avec les structures hélicoïdales d'ARN. Tandis que cette flexibilité de conformation n'a jamais été montrée pour l'élément SECIS, il est toutefois suggéré que le quartet GA adopte un repliement fermé en présence de $MgCl_2$ [56]. Initialement considéré comme une plateforme permettant le recrutement de facteurs protéiques nécessaires au recodage en Sec, l'élément SECIS contrôle et régule la traduction des ARNm de sélénoprotéines en assurant l'efficacité d'insertion des Sec [60]. Ainsi une délétion ou une mutation dans le « cœur SECIS » ou dans le motif AAR nuit au décodage en Sec [61-63]. Bien que requis pour l'incorporation de la Sec dans les protéines, la fonction exacte du motif AAR n'a pas été déterminée [63]. De façon intéressante, un élément SECIS unique est suffisant pour dicter l'incorporation de l'ensemble des Sec codées dans une sélénoprotéine, exception faite de la

sélénoprotéine SelP qui contient deux élément SECIS dans sa région 3'UTR, nécessaire pour le décodage de multiples codons UGA codant pour des Sec (de 10 à 17 dépendamment de l'espèce) [64, 65].

1.3.1.2 L'élément SRE : un élément de régulation supplémentaire

L'élément SRE (*Sec Redefinition Element*) est l'un des derniers facteurs impliqués dans le décodage en Sec à avoir été identifié [66]. Cette structure conservée d'ARN de 16 nucléotides, de type tige-boucle, fut prédite par des approches bio-informatiques dans la sélénoprotéine SelN, puis dans SEPHS2, SelH, SelO et SelT. Il fut également montré de manière expérimentale que l'élément SRE modulait de façon positive l'efficacité d'insertion des Sec dans les protéines [66]. Contrairement à l'élément SECIS, l'élément SRE est localisé dans la région codante de l'ARNm des sélénoprotéines, six nucléotides en aval du codon UGA. Compte tenu du fait qu'une mutation faux-sens dans l'élément SRE induit une diminution significative de l'efficacité d'insertion de la Sec, il est proposé que l'élément SRE permette de contrôler le décodage en Sec de certaines sélénoprotéines, tel qu'il fut montré pour SelP [66]. Des études supplémentaires sont toutefois nécessaires pour résoudre son rôle biologique.

Outre les éléments d'ARN, de nombreux facteurs protéiques interviennent dans le mécanisme de décodage et d'incorporation de la Sec.

1.3.2 Les protéines liant l'élément SECIS

1.3.2.1 SBP2 : une protéine polyvalente

La protéine SBP2 (*SECIS Binding Protein 2*) joue un rôle central dans l'incorporation de la Sec chez les mammifères. Présente chez tous les eucaryotes, elle n'a jusqu'à présent été caractérisée que chez l'Homme et le rat [67-69]. SBP2 est le facteur de la machinerie de recodage le plus extensivement étudié. SBP2 est connue pour lier l'élément SECIS, interagir avec le ribosome et le facteur d'élongation spécifique EFSec. Elle joue aussi un rôle dans le transport nucléocytoplasmique et dans l'assemblage de particules de ribonucléoprotéines messagères (RNPM - constituées de l'ARNm et de protéines associées). Ainsi, SBP2 est essentielle à l'incorporation de la Sec dans les protéines [70]. En effet, des mutations dans le gène codant pour SBP2 résultent en une diminution significative de l'incorporation de Sec dans les protéines

in vivo [68, 71] et l'inactivation du gène est létale au stade embryonnaire [72]. D'autre part, une mutation dans le gène codant pour SBP2, à l'origine d'un défaut de synthèse de la DIO2, a été identifiée chez des patients atteints d'hyperthyroïdie [73, 74]. Les différents domaines de la protéine SBP2 sont illustrés figure I.8.

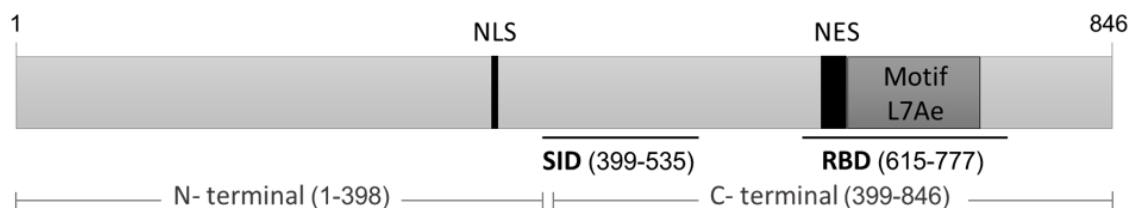


Figure I.8. Représentation schématique des domaines de la protéine SBP2. RBD : *RNA-Binding Domain*; SID : *Selenocysteine Insertion Domain*; NLS : *Nuclear Localization Signal*; NES : *Nuclear Export Signal*. Adapté de Donovan *et al.*, 2010 [75].

Peu conservé, le domaine N-terminal de SBP2 n'est pas essentiel pour l'incorporation de la Sec [68]. Il est d'ailleurs absent dans certaines espèces telles que les insectes [76]. Des signaux de localisation nucléaire (NLS) et des signaux d'export nucléaire (NES) ont néanmoins été prédits dans son domaine N-terminal, expliquant potentiellement la capacité de SBP2 à effectuer des trajets nucléocytoplasmiques, et, ainsi, à prévenir la terminaison prématurée de la traduction en formant des complexes RNPm (*cf.* section 1.4.1) [42, 77]. Le domaine C-terminal est, quant à lui, suffisant pour l'incorporation de la Sec *in vitro*, et contient les domaines de liaison au ribosome et à l'élément SECIS (Fig. I.8) [68]. Ce domaine peut être, lui aussi, divisé en deux : le domaine RBD (*RNA-Binding Domain*) et le domaine SID (*Selenocysteine Insertion Domain*) (Fig. I.8) [78]. Le RBD contient un motif L7Ae de liaison à l'ARN, également retrouvé dans plusieurs protéines connues pour interagir spécifiquement avec des structures hélicoïdales de l'ARN (dites *kink-turn*). Ce motif L7Ae paraît essentiel pour la liaison au SECIS et au ribosome [79]. Le RBD est également retrouvé interagissant avec le facteur d'élongation spécifique EFSec, en présence de l'élément SECIS [78, 80]. Il est proposé que le domaine SID améliorerait les interactions avec les partenaires de SBP2 [78]. Alors que des propriétés et fonctions distinctes leurs soient imputées, ces domaines sont complémentaires et la présence des domaines RBD et SID est essentielle pour garantir l'efficacité d'insertion de la Sec dans les protéines.

Chez les vertébrés, la protéine SBP2L (*SBP2-like*), un paralogue de SBP2, a également été identifiée. Malgré la similarité de séquence avec SBP2, SBP2L ne supporte pas l'incorporation de la Sec chez les mammifères et sa fonction reste inconnue [81]. Il est cependant intéressant de noter que certains invertébrés ne comportent que l'homologue SBP2L lequel assure l'incorporation de la Sec [82].

1.3.2.2 EFSec : le facteur d'élongation spécifique des sélénoprotéines

Chez les bactéries, SelB est le facteur d'élongation spécialisé dans l'incorporation des Sec dans les protéines. Le domaine N-terminal de SelB présente une forte similarité de séquence et une homologie fonctionnelle avec EF-Tu, le facteur d'élongation général des procaryotes. Tandis que SelB lie spécifiquement Sec-ARNt^{[Ser]^{Sec}}, EF-Tu en est incapable [83]. De plus, SelB lie l'élément SECIS. EFSec est l'homologue de SelB chez les mammifères [84, 85]. Le facteur d'élongation EFSec peut être délimité en quatre domaines : les domaines I, II et III qui présentent de fortes similarités avec le facteur d'élongation eucaryote EF1-A et un domaine IV unique, en C-terminal (Fig. I.9). Tout comme EF1-A, EFSec détient une activité GTPase. En revanche, EFSec possède une forte spécificité pour le Sec-ARNt^{[Ser]^{Sec}} tandis qu'il ne peut se lier ni au O-phosphoséryl-ARNt^{[Ser]^{Sec}} ni à d'autres ARNt aminoacylés [46, 85]. Contrairement à SelB, EFSec ne possède pas de domaine de liaison à l'élément SECIS. Cependant, EFSec peut interagir avec SBP2 et former un complexe, dépendant de la présence de l'élément SECIS. La formation du complexe est, de plus, favorisée par la présence du Sec-ARNt^{[Ser]^{Sec}}. Ainsi, le domaine IV de EFSec semble important pour l'interaction avec SBP2 et Sec-ARNt^{[Ser]^{Sec}} [86]. EFSec serait également capable de transloquer entre nucléoplasme et cytoplasme, seul ou en complexe avec SBP2, puisque des signaux de localisation nucléaire et des signaux d'export nucléaire ont été prédits dans sa séquence.

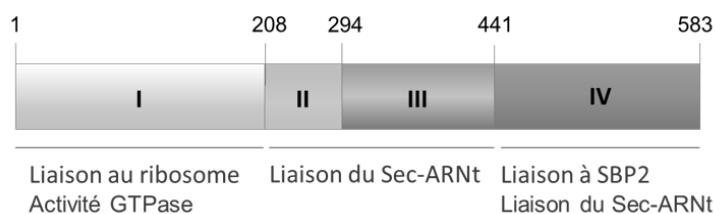


Figure I.9. Représentation schématique des domaines du facteur d'élongation spécifique eucaryote EFSec. Les domaines I-IV sont représentés. Les fonctions connues ou putatives sont indiquées. Adapté de Hatfield *et al.*, 2012 [2].

1.3.2.3 La protéine ribosomale L30 : concurrente de SBP2 pour la liaison à l'élément SECIS

La protéine ribosomale L30 est une composante de la grande sous-unité ribosomique (60S) chez les eucaryotes. Une petite fraction libre de la protéine (non associée au ribosome) existe également dans la cellule, régulant l'épissage de son propre pré-ARNm ainsi que son niveau d'expression [87]. Il a été montré que la protéine L30 pouvait, elle aussi, lier l'élément SECIS par l'intermédiaire de son motif L7Ae de liaison à l'ARN [88]. Bien que son rôle dans l'incorporation de la Sec ne soit pas clair, il est intéressant de noter que L30 semble lier l'élément SECIS à un site chevauchant celui de SBP2. De plus, les études *in vitro* montrent que SBP2 lie préférentiellement la forme « ouverte » de l'élément SECIS, tandis que L30 lie la forme fermée, *kink-turn* [88]. Par ailleurs, dans des conditions de concentrations élevées en Mg^{2+} , favorisant une conformation *kink-turn* de l'élément SECIS, l'interaction L30-SECIS est améliorée et L30 entre en compétition avec SBP2 pour l'élément SECIS. Ces découvertes suggèrent un modèle dans lequel SBP2 lierait initialement l'élément SECIS puis, lors du décodage du codon UGA en Sec, serait déplacée par la protéine L30, laquelle serait responsable du recrutement du complexe au ribosome [88]. Ce modèle nécessite toutefois d'être confirmé expérimentalement.

1.3.3 SECp43 : formation, stabilité et localisation des complexes supramoléculaires

La protéine SECp43 fut identifiée associée à l'ARNt^{[Ser]^{Sec}}, suite à une réaction de précipitation à partir de lysats de cellules de mammifères [89]. SECp43 constitue l'une des rares protéines contenant des motifs de liaison des ribonucléoprotéines (RNP) - habituellement associés à la reconnaissance des pré-ARNm, ARNm, ARNr et ARNsn - à être capable de lier un ARNt. Cependant, la preuve d'une liaison directe avec le Sec-ARNt^{[Ser]^{Sec}} n'a pas pu être démontrée *in vitro*. Principalement présente dans le noyau, SECp43 interagit avec le complexe Sec-ARNt^{[Ser]^{Sec}}-EFSec *in vitro* et *in vivo* ainsi qu'avec les protéines SEPSECS et SEPHS1 *in vivo* [41, 42]. Lorsque co-exprimées dans les cellules, SECp43 semble responsable du transport de SEPSECS depuis le cytoplasme vers le compartiment nucléaire et sa séquestration dans le noyau [41]. Il fut aussi montré que SECp43 joue un rôle dans la formation et/ou la stabilisation du complexe Sec-ARNt^{[Ser]^{Sec}}-EFSec-SBP2. De plus, SECp43 favorise la formation et la localisation subcellulaire du complexe SEPHS1-SEPSECS-SECp43 dans le noyau [42]. Finalement, la déplétion de SECp43 par des petits ARN interférents (siRNA) inhibe la

production de l'isoforme méthylée mcm^5Um de l'ARNt^{[Ser]^{Sec}}, notamment requise pour la synthèse des sélénoprotéines liées au stress [41]. Conjointement, ces données suggèrent que la protéine SECp43 est nécessaire à la machinerie de biosynthèse de la Sec et à son incorporation dans les protéines. SECp43 semble promouvoir et stabiliser les interactions entre plusieurs facteurs impliqués dans la synthèse des sélénoprotéines, et supporter le transport nucléocytoplasmique des complexes supramoléculaires.

1.3.4 Modèles de l'incorporation de la Sec chez les eucaryotes

Bien que les acteurs principaux de l'incorporation de la Sec dans les sélénoprotéines soient à présent identifiés, les événements moléculaires qui conduisent au recodage en Sec sont incertains. Plusieurs modèles mettant en scène les facteurs essentiels à l'incorporation de la Sec sont ainsi proposés.

Dans le premier modèle, SBP2 est pré-associée au ribosome. Une fois le ribosome arrêté au niveau du codon UGA, l'élément SECIS, dans une conformation ouverte, interagit avec SBP2. EFSec en complexe avec le Sec-ARNt^{[Ser]^{Sec}} interagit alors avec SBP2, ce qui permet de recruter le Sec-ARNt^{[Ser]^{Sec}} au site A du ribosome. Puis L30 entre en compétition avec SBP2 et la décroche du SECIS [78, 90]. Ce modèle se base notamment sur des données préliminaires montrant que SBP2 ne peut pas interagir simultanément avec l'élément SECIS et avec le ribosome (Fig. I.10A).

Dans le second modèle, SBP2 est liée à l'élément SECIS et recrute le complexe EFSec-Sec-ARNt^{[Ser]^{Sec}} directement en 3'UTR de l'ARNm, avant le décodage du codon UGA. Puis L30, préalablement associée au ribosome décroche SBP2 de son interaction avec l'élément SECIS. Ainsi, le complexe SBP2-EFSec-Sec-ARNt^{[Ser]^{Sec}} est livré au site A du ribosome pour permettre le décodage en Sec [88] (Fig. I.10B).

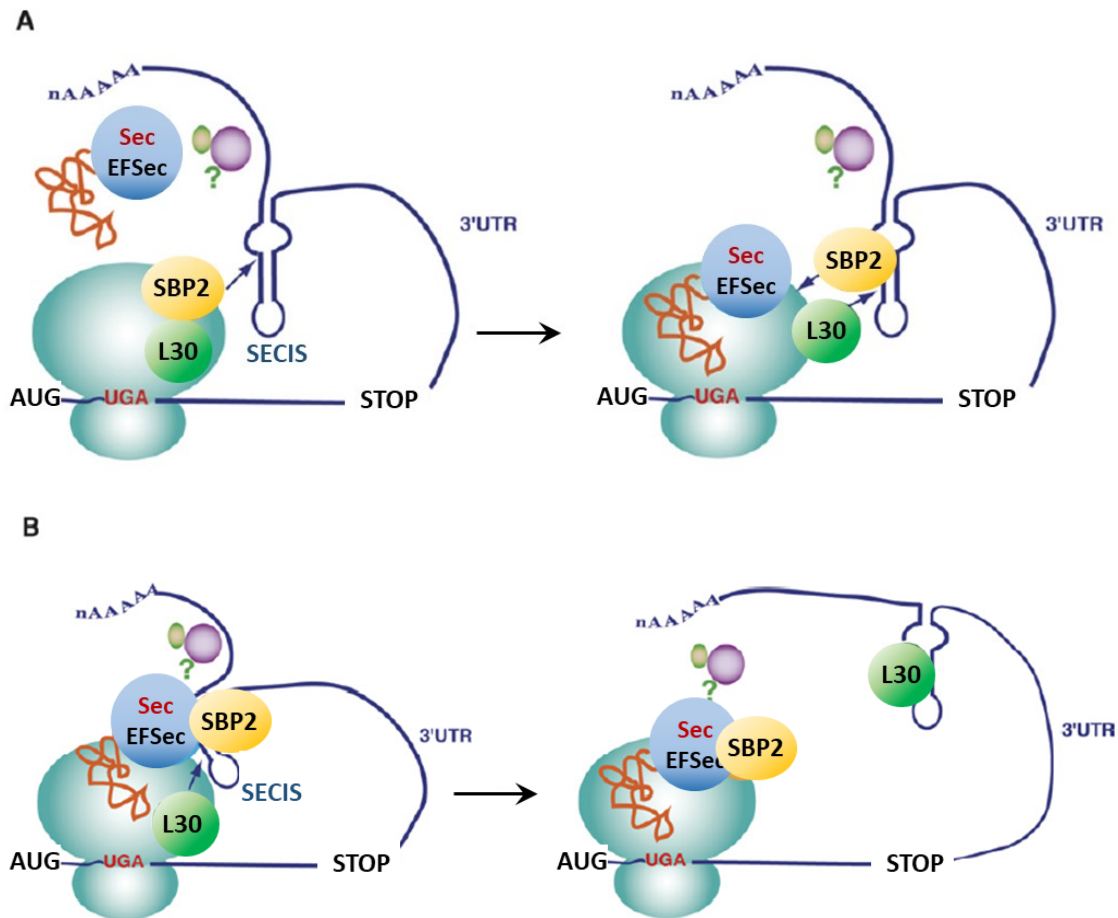


Figure I.10. Modèles alternatifs d'incorporation de la Sec dans les protéines chez les eucaryotes. (A) Dans le modèle 1, SBP2 est liée au ribosome, puis interagit avec l'élément SECIS et le complexe Sec-ARNt^{Sec}-EFSec pour livrer le complexe au site A du ribosome. L30 déplace alors SBP2 de sa liaison au SECIS. (B) Dans le modèle 2, SBP2 liée au SECIS recrute le complexe Sec-ARNt^{Sec}-EFSec. L30 liée au ribosome déplace SBP2. Adapté de Allamang *et al.*, 2006 [31].

1.3.5 L'altération des composants de la machinerie de synthèse et de traduction de la Sec affecte l'intégrité de l'organisme

De nombreux troubles physiologiques et des maladies variées ont été attribués à une carence en sélénium ou à des mutations dans les gènes codant les sélénoprotéines (*cf.* section 1.1). En revanche, très peu de mutations dans des gènes encodant les composants responsables de la synthèse et de l'incorporation des Sec furent identifiées. Cette observation n'est pas surprenante compte tenu de l'importance de la voie de synthèse et d'incorporation des Sec pour la

préservation de l'intégrité des organismes. De telles mutations seraient délétères pour l'organisme, à des stades précoces du développement. C'est d'ailleurs ce qui est observé dans le cas de la délétion du gène *trsp* codant pour l'ARNt^{[Ser]Sec}, létale au stade embryonnaire chez la souris [32]. De ce fait, seules les mutations n'abolissant pas totalement la machinerie de synthèse et d'incorporation des Sec ont pu être décrites. Ainsi, diverses mutations dans le gène codant SBP2 ont été mises en évidence, conduisant à des troubles multisystémiques plus ou moins sévères [91]. Des mutations à l'origine d'une atrophie cérébrale progressive du cervelet (PCCA) ont également été identifiées dans le gène encodant SEPSECS [92]. Les individus atteints par cette pathologie n'atteignent généralement pas l'âge adulte. Finalement des mutations dans les éléments SECIS et SRE sont responsables du développement de myopathies diverses (*SEPN1-related myopathies*) [93, 94], comme observé dans le cas de SelN (*cf.* section 1.1.1.6). Les conséquences de ces différentes mutations sur le phénotype sont résumées dans la figure I.11.

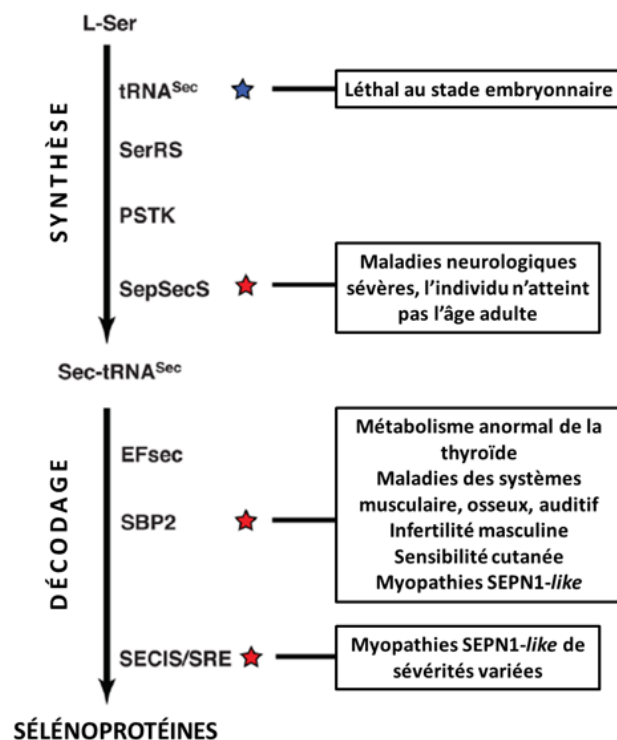


Figure I.11. L'altération des composants impliqués dans la synthèse et le décodage de la Sec a des conséquences multiples sur l'organisme humain. Les mutations observées chez des individus sont indiquées par une étoile rouge. Le mutant *knockout* de l'ARNt^{[Ser]Sec} est signalé par une étoile bleue. Adapté de Schmidt *et al.*, 2012 [91].

1.4. Régulation de la synthèse des sélénoprotéines

Sous forme libre, ou incorporé dans les protéines, le taux de sélénium présent dans les cellules est finement régulé. Les niveaux de synthèse des sélénoprotéines dépendent, par ailleurs, de la disponibilité en sélénium, et une hiérarchisation dans la production des sélénoprotéines est induite en cas de déficit en sélénium. Ainsi, le sélénium est un régulateur central dans la production des sélénoprotéines, agissant à différents niveaux tels que la stabilité des ARNm de sélénoprotéines, augmentant leur susceptibilité à la voie NMD (*Nonsense-Mediated Decay*) [95] ; la régulation de l'efficacité du décodage du codon UGA en Sec [96, 97] ; les taux de Sec-ARNt^{[Ser]^{Sec}} et de ses isoformes non méthylées ou méthylées, responsables respectivement de la synthèse des sélénoprotéines *housekeeping* ou liées au stress (*cf.* section 1.2.1) [98, 99]. Parmi les autres facteurs proposés pour réguler la synthèse des sélénoprotéines, nous pouvons citer la formation de complexes RNPM ; la séquestration nucléaire de SBP2 en contexte de stress oxydatif [71] ; l'interaction de eIF4a3 et de la nucléoline avec l'élément SECIS [100, 101] ; et la position du codon UGA par rapport à l'élément SECIS [102]. Certains de ces facteurs sont discutés dans les sections suivantes.

1.4.1 Assemblage nucléaire : la formation de RNPM permet d'éviter la dégradation par NMD

La voie de dégradation des ARNm non-sens (NMD, *Nonsense-Mediated Decay*) est un mécanisme impliqué dans l'élimination des ARNm présentant un codon de terminaison prématuré, situé à environ 50 nucléotides en amont d'un complexe EJC (*Exon Junction Complexes*) [103]. La NMD est un mécanisme de contrôle-qualité des ARNm essentiel chez les eucaryotes. Les ARNm de sélénoprotéines, qui contiennent des codons UGA à des positions variables dans leur cadre de lecture, sont des cibles de la dégradation par NMD. En effet, des études suggèrent que 14 des 25 ARNm de sélénoprotéines humaines sont sensibles à la NMD, en condition de carence en sélénium, notamment [104]. Or, l'assemblage dans le noyau de l'ARNm des sélénoprotéines avec des composants de la machinerie d'incorporation de la Sec permettrait d'éviter la dégradation par NMD. Tandis que plusieurs facteurs nécessaires au décodage de la Sec présentent des signaux potentiels de localisation ou d'export nucléaire (SBP2, EFSec, SECp43, SEPSECS), la formation de tels complexes dans le noyau a pu être

observée par des essais de co-immunoprécipitation et par microscopie à fluorescence. Des études suggèrent également que la formation de RNPm, contenant des composants de la machinerie d'incorporation de la Sec, avant l'export de l'ARNm de la sélénoprotéine, contribuerait à améliorer l'efficacité de synthèse des sélénoprotéines de façon significative [42, 71].

1.4.2 Hiérarchisation de la synthèse des sélénoprotéines

1.4.2.1 Facteur d'initiation de la traduction eucaryote eIF4a3

Le facteur d'initiation de la traduction eucaryote eIF4a3 est connu pour son implication dans l'épissage des pré-ARNm dans le noyau et sa participation à la NMD par assemblage des EJC [105]. En plus de sa fonction canonique, eIF4a3 sert de régulateur négatif dans la voie d'incorporation des Sec [100]. En effet, eIF4a3 interagit avec les éléments SECIS des sélénoprotéines liées au stress, à un site chevauchant le site de liaison de SBP2. D'autre part, le niveau de eIF4a3 dans la cellule est considérablement accru lors d'un déficit en sélénium. Ceci suggère que dans le cas d'une disponibilité limitée en sélénium, eIF4a3 se lierait sélectivement aux éléments SECIS des ARNm de sélénoprotéines non-essentiels, prévenant la liaison de la protéine SBP2, ce qui empêcherait la synthèse de ces sélénoprotéines [100].

1.4.2.2 Nucléoline

La nucléoline est une protéine abondante localisée dans le nucléole et impliquée dans la synthèse de l'ARNr et dans la biogenèse des ribosomes. Bien qu'elle ait été identifiée en liaison avec l'élément SECIS des sélénoprotéines [101], son rôle exact dans la voie de régulation de la synthèse des sélénoprotéines n'est pas clair. En effet, les données sont contradictoires montrant, dans une étude, qu'elle agirait comme régulateur positif de la traduction des sélénoprotéines essentielles [106], et, dans une autre, qu'elle lierait les éléments SECIS des ARNm de sélénoprotéines sans discrimination, contredisant son implication dans la hiérarchisation de la synthèse des sélénoprotéines [107].

1.4.3 Position du codon UGA et efficacité d'insertion de la Sec

Initialement il était proposé que la Sec était insérée dans les sélénoprotéines eucaryotes dès lors que l'élément SECIS était présent en 3'UTR de l'ARNm et situé à une distance minimale de 50

nucléotides du codon UGA. Cependant, des études récentes ont révélé que l'efficacité d'insertion dépendait également de la position originale du codon UGA dans la séquence. En effet, la modification, par mutagenèse, de la position du codon UGA en diverses positions de la séquence codante conduit à une réduction, voire à l'abolition, de l'insertion de la Sec dans la protéine [102]. De façon intéressante, cette régulation est sélénoprotéine-dépendante. Dans les sélénoprotéines TR1 et TGR l'insertion de la Sec est limitée à sa position naturelle tandis que cette règle ne s'applique pas pour TR3. Bien que le mécanisme ne soit pas clairement élucidé, il est proposé que la structure spécifique du SECIS (de type I ou II) et son affinité pour les protéines de liaison au SECIS (SBP2, eIF4a3, nucléoline) pourraient expliquer cette régulation [102].

Ainsi, la machinerie de biosynthèse et d'incorporation des Sec dans les protéines est une machinerie macromoléculaire complexe, hautement dynamique et finement régulée. Lors de ce processus cellulaire, l'ARN, au cœur de la machinerie, les facteurs protéiques et le ribosome s'associent et collaborent pour permettre le décodage de la Sec à partir d'un codon stop UGA et son insertion dans les chaînes polypeptidiques en cours de synthèse. Cette machinerie de traduction spécialisée repose sur une série d'interactions transitoires, prenant place entre les différents compartiments cellulaires. La nature dynamique de la machinerie soulève de nombreuses questions, tandis que de nombreux détails mécanistiques et moléculaires sont encore à clarifier. La machinerie permet la synthèse des sélénoprotéines, qui interviennent dans diverses voies métaboliques et dont l'intégrité est essentielle pour assurer le bon fonctionnement de l'organisme.

Au cours de mon doctorat, nous nous sommes également intéressés à un autre complexe macromoléculaire, lui d'origine procaryote, couramment présent dans les bactéries pathogènes humaines. Le système de sécrétion de type 4 est, en effet, retrouvé dans la bactérie *Helicobacter pylori*, où il participe à la virulence.

I.2. LE SYSTÈME DE SÉCRÉTION DE TYPE 4 D'*HELICOBACTER PYLORI*

Helicobacter pylori est une bactérie pathogène humaine découverte au début des années 1980 par Barry Marshall et Robin Warren. Elle fut isolée à partir de muqueuses gastriques provenant de patients atteints de gastrite ou d'ulcère gastroduodéal. Dans la majorité des cas, l'infection par *H. pylori* est asymptomatique ; cependant, 10-15% des personnes infectées développent un ulcère gastroduodéal et 1-2% un cancer gastrique. La sévérité de l'infection dépend de plusieurs facteurs, dont le génotype de la souche colonisatrice, des réponses immunitaires de l'hôte, ainsi que des facteurs environnementaux. L'émergence rapide et l'augmentation des souches *H. pylori* multirésistantes posent un problème majeur dans le traitement et l'éradication de la bactérie. En 1994, *H. pylori* fut la première bactérie à être reconnue comme carcinogène de type I par le CIRC/OMS (Centre International de Recherche sur le Cancer/ Organisation Mondiale de la Santé). Depuis 2017, *H. pylori* est classée dans la liste de l'OMS des 12 « agents pathogènes prioritaires » résistants aux antibiotiques pour lesquelles un réel effort de recherche est nécessaire pour le développement de nouveaux traitements.

2.1. *H. pylori* : la bactérie pathogène humaine colonise l'estomac

2.1.1 Prévalence géographique et transmission

La bactérie *H. pylori* est une bactérie à Gram négative qui colonise généralement les muqueuses de l'estomac à l'enfance et persiste dans l'organisme, tout au long de la vie, en l'absence d'un traitement d'éradication efficace. Il est actuellement estimé qu'*H. pylori* colonise près de 50% de la population mondiale, avec une incidence allant de 30% dans les pays développés de l'Ouest (Canada, États-Unis) à 90% dans certains pays d'Afrique et d'Amérique latine, la prévalence d'*H. pylori* diminuant avec l'urbanisation. De façon générale, les taux de cancers gastriques concordent avec les taux d'infection par *H. pylori* ; en Asie, la forte prévalence de cancers gastriques reflète les taux élevés d'infection par la bactérie. Cependant, cette relation de causalité n'est pas universelle ; en Afrique, par exemple, le taux d'infection par *H. pylori* est élevé, mais la fréquence de cancers gastriques est extrêmement faible [108]. De plus, le

mécanisme de transmission reste incertain. Il est proposé que la transmission se produirait par contact direct, requérant un contact étroit entre individus (*p. ex.* transmission intrafamiliale), mais une voie de transmission indirecte par le biais de la nourriture, des eaux contaminées ou des animaux a aussi été reportée [109].

2.1.2 Pathologies gastriques sévères liées à l'infection par *H. pylori*

Tandis que la colonisation par *H. pylori* conduit systématiquement à une inflammation locale de la muqueuse gastrique, seule une minorité d'individus développent des signes cliniques apparents. La sévérité de ceux-ci dépend du génotype bactérien, des réponses immunitaires et du patrimoine génétique de l'hôte et de facteurs environnementaux, et peuvent conduire à une gastrite, un ulcère voire au cancer gastrique.

2.1.2.1 Ulcère gastroduodénal

L'ulcère gastroduodénal consiste en une rupture des muqueuses duodénales ou gastriques et est associé à une mortalité importante. Ce type d'ulcère peut s'accompagner de complications telles que des risques d'hémorragie ou de perforation. Des lésions précancéreuses ou un adénocarcinome gastrique peuvent également survenir [110]. Dans 95% des cas pour l'ulcère duodénal et 70% des cas pour l'ulcère gastrique, *H. pylori* est l'agent causal de la maladie [111]. L'éradication d'*H. pylori* permet la guérison de l'ulcère et prévient sa récurrence [112].

2.1.2.2 Adénocarcinome gastrique

Avec environ 100 000 nouveaux cas par année, le cancer gastrique constitue le 5^e type de cancer le plus répandu dans le monde, et se classe comme la 3^e cause de mortalité par cancer, du fait d'un diagnostic généralement tardif [113]. Chez un individu infecté par *H. pylori*, le risque de développer un cancer gastrique atteint de 1-2% [110, 114]. En effet, l'infection par *H. pylori*, qui conduit à une gastrite chronique, entraîne, après plusieurs décennies, l'atrophie de la glande gastrique, puis une métaplasie et dysplasie intestinale, et conduit finalement au développement d'un adénocarcinome. Bien que l'éradication d'*H. pylori* réduise l'incidence de la gastrite atrophique, le risque de développer un cancer gastrique n'est pas diminué, à moins que l'infection soit traitée de façon précoce [115]. Ainsi, *H. pylori* serait responsable de l'apparition

de 89% des cas des cancers gastriques non-cardia (dits distaux ou hors jonction œso-gastrique) ; classant la bactérie dans la liste des agents carcinogènes humains [116].

2.1.2.3 Lymphome MALT

Le lymphome MALT (*Mucosa Associated Lymphoma Tissue*) est une affection rare attribuée à l'infection par *H. pylori* dans 90% des cas. Elle consiste en l'apparition de tissus lymphoïdes dans la muqueuse gastrique, qui en est normalement dépourvue [117]. Ces tissus lymphoïdes induisent une prolifération incontrôlée de cellules lymphocytes B qui, dans certains cas, se traduit par l'apparition d'un lymphome gastrique [118]. L'éradication d'*H. pylori*, permet une rémission complète. Une corrélation entre les facteurs de virulence d'*H. pylori* et le risque de développement du lymphome MALT n'a pas été mise en évidence ; en revanche, dans ce cas, les facteurs génétiques de l'hôte semblent jouer un rôle important.

2.1.3 Persistance d'*H. pylori* dans l'organisme

H. pylori évolue dans un environnement hostile, acide et visqueux. Cependant, *H. pylori* n'est pas une bactérie acidophile au sens propre du terme puisqu'elle peut survivre pour une courte période à un pH 2 mais évolue préférentiellement à des pH 5-6. Or, le pH de l'estomac est généralement compris entre pH 1-2. *H. pylori* a donc su mettre en place divers mécanismes pour s'adapter et coloniser cet environnement. Pour contourner l'acidité de l'estomac, *H. pylori* sécrète de grandes quantités d'uréases, 15% de son protéome y est, en effet, consacré [110]. Le pH alcalin de l'ammonium, généré suite à l'hydrolyse de l'urée, produite par les cellules épithéliales, neutralise le milieu qui atteint un pH 6 [119]. L'anhydrase α -carbonique, produite par *H. pylori*, conduit également à une alcalinisation du milieu [120]. Ces deux mécanismes sont régulés par le système à deux composants ArsRS, un système de signalisation activé par des pH acides qui régule, d'autant plus, l'amidase AmiE, la formamidase AmiF, et l'arginase RocF favorisant ainsi la tolérance à l'acidité [120-122].

Pour persister dans l'estomac, *H. pylori* a également dû développer des stratégies pour s'adapter à la viscosité de la muqueuse gastrique. Tandis que l'activité uréase de la bactérie réduit le pH du milieu, diminuant du même fait la viscosité du mucus gastrique, la forme hélicoïdale de la bactérie et sa motilité, liée à la présence de flagelles, permettent à *H. pylori* de pénétrer rapidement en profondeur dans la couche de mucus recouvrant l'épithélium de l'estomac. *H.*

pylori évite ainsi les conditions très acides de la lumière gastrique, mais aussi d'être expulsée du fait du renouvellement gastrique régulier (vidange gastrique) [110, 123].

Finalement, l'interaction d'*H. pylori* avec l'épithélium gastrique stimule l'expression de cytokines pro-inflammatoires, de chimiokines et de peptides antimicrobiens. Malgré cette puissante activité inflammatoire et bactéricide, *H. pylori* met en œuvre de multiples stratégies pour échapper à l'immunité et favoriser l'immunosuppression. En minimisant les défenses immunitaires de l'hôte, la bactérie peut maintenir une colonisation persistante et permanente. L'injection de la cytotoxine CagA (dont la fonction est détaillée dans la section 2.2.3.3) dans les cellules épithéliales gastriques, induit une réponse inflammatoire qui, à long terme, peut s'avérer néfaste pour la persistance de la bactérie dans l'organisme. Pour limiter cette réponse, la bactérie a mis en place un mécanisme, activé par l'intermédiaire de la toxine VacA (dont la fonction est discutée section 2.3.1), sécrétée dans le milieu extracellulaire, qui stimule la dégradation de CagA par autophagie [124]. VacA agit également comme un puissant immunosuppresseur, modulant la réponse immunitaire adaptative, en affectant la prolifération des cellules T et en favorisant la différenciation des cellules dendritiques tolérogènes [125, 126]. Parmi les autres stratégies mises en place par *H. pylori*, la modulation de l'expression et de la structure de molécules localisées à sa surface (flagelline, lipopolysaccharide) permet d'éviter ou d'atténuer la reconnaissance par les TLR (*Toll-like receptors*), récepteurs du système immunitaire impliqués dans la reconnaissance des pathogènes et leur adressage à la voie d'endocytose [127].

2.1.4 Éradication d'*H. pylori* et émergence des résistances

L'efficacité du traitement d'éradication d'*H. pylori* est grandement affectée par la localisation de la bactérie dans l'organisme. En effet, l'environnement visqueux et le pH acide dans l'estomac rendent une monothérapie inefficace. Le traitement standard d'éradication d'*H. pylori* consiste en une trithérapie, composée de deux antibiotiques (généralement la clarithromycine et l'amoxicilline ou le métronidazole) et d'un inhibiteur de pompe à protons (IPP) pour diminuer la production d'acide gastrique. Ce traitement s'est avéré efficace pour traiter les ulcères existants et prévenir leur récurrence [128, 129]. Néanmoins, la résistance à la clarithromycine, à

l'amoxicilline et au métronidazole, dont la prévalence varie considérablement à travers le monde (Fig. I.12), est un problème croissant qui entraîne l'échec des traitements d'éradication.

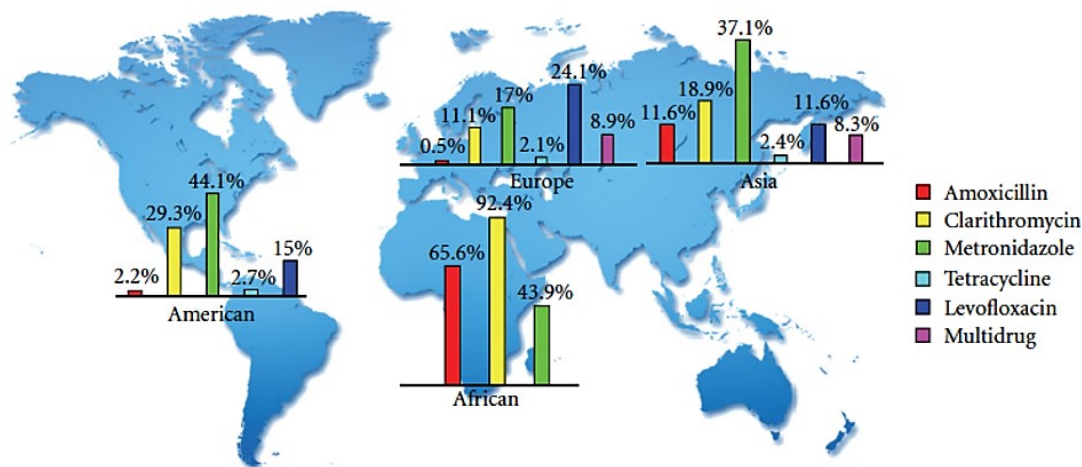


Figure I.12. Fréquence de résistance des souches *H. pylori* aux antibiotiques dans les diverses régions du monde. Depuis Wu et al., 2012 [130].

Outre l'émergence de souches multirésistantes, la pathogénicité d'*H. pylori* dépend principalement du génotype de la souche, celles encodant pour un pilus CagSST4 s'avérant nettement plus virulentes.

2.2 Le CagSST4 : facteur de pathogénicité d'*H. pylori*

2.2.1 Versatilité des SST4 et fonctions

Actuellement neuf systèmes de sécrétion ont été décrits dont six chez les bactéries à Gram négatif [131]. Parmi ceux-ci on distingue les SST4, des dispositifs macromoléculaires polyvalents, utilisés par les bactéries pour transporter diverses macromolécules à travers leurs membranes [132, 133]. Les SST4 peuvent être subdivisés en trois catégories selon leurs fonctions. Le système de conjugaison, dont celui de la bactérie *Agrobacterium tumefaciens* est l'archétype, est utilisé pour le transfert d'ADN entre cellules, ce qui permet une adaptation à l'environnement et la propagation des gènes de résistance aux antibiotiques [134, 135]. Le second type de système, le système de transformation, permet l'absorption et/ou la libération d'ADN depuis et vers le milieu extracellulaire, et est représenté par le système GGI de *Neisseria*

gonorrhoeae et par le système *comB* d'*H. pylori* [136, 137]. Le dernier type, présent chez de nombreuses bactéries pathogènes telles que *Bordetella pertusis*, *Legionella pneumophila*, *Brucella spp.*, *Bartonella spp* et *H. pylori*, permet le transfert de protéines effectrices dans le cytoplasme des cellules eucaryotes [138]. Dans la bactérie, *H. pylori* quatre SST4 ont été identifiés : le SST4 encodé par le *cagPAI* qui permet l'injection de l'oncoprotéine CagA dans les cellules épithéliales gastriques, le système *comB* qui assure les processus de transformation d'ADN, et les SST4 encodés par les opérons *tfs3* et *tfs4* qui joueraient un rôle dans la virulence et la conjugaison respectivement. Dans les prochaines sections, seul le SST4 encodé par l'îlot génomique *cagPAI*, désigné CagSST4 dans ce mémoire, sera abordé plus en détail.

2.2.2 Le *cagPAI* encode un pilus CagSST4 chez *H. pylori*

Le *cagPAI* est un îlot génomique d'environ 40 kb qui contient de 27-30 gènes, selon les souches d'*H. pylori*, organisés en plusieurs opérons (Fig. I.13). Les souches d'*H. pylori* qui portent le *cagPAI* sont dites *cag*-positives ou de type 1 (en opposition aux souches *cag*-négatives ou de type 2). Le *cagPAI* code pour un système de sécrétion de type 4 (SST4) et pour la cytotoxine CagA [139]. De nombreuses études ont mis en évidence que les souches d'*H. pylori* *cag*-positives étaient associées aux formes les plus sévères des maladies gastriques du fait notamment de la translocation de l'oncoprotéine CagA dans les cellules gastriques [140-142].



Figure I.13. Représentation de l'organisation des gènes *cag* dans l'îlot de pathogénicité *cagPAI* d'*H. pylori*. Les gènes *cag* présentant des similarités de séquences avec le SST4 *vir* d'*A. tumefaciens* sont indiqués. Adapté de Backert *et al.*, 2017 [143].

Comparativement aux SST4 de conjugaison (*ex.* SST4 d'*A. tumefaciens* ou des plasmides R388 et pKM101 d'*E. coli*), relativement peu d'informations structurales sont disponibles concernant le CagSST4. Parmi les protéines Cag encodées par l'îlot génomique *cagPAI*, 17 sont essentielles pour la formation d'un CagSST4 fonctionnel [144]. Parmi celles-ci, neuf sont uniques à *H. pylori* et ne présentent pas de similarité de séquence avec d'autres protéines bactériennes. Les

protéines encodées par le *cagPAI* et leurs homologues proposées avec le système VirB/D d'*A. tumefaciens* sont illustrées dans le Tableau I.2

Tableau I.2. Protéines encodées par le *cagPAI* et homologues avec le système VirB/D d'*A. tumefaciens*. Adapté de Backert et *al.*, 2017 [143].

Gène	Protéine	Localisation cellulaire	Fonction proposée	Homologues dans le système VirB/D
<i>hp0520</i>	Cagζ/Cag1	MI		
<i>hp0521</i>	Cagε/Cag2	C		
<i>hp0522</i>	Cagδ/Cag3	ME	Facteur du complexe cœur	
<i>hp0523</i>	Cagγ/Cag4	P	PG hydrolase	VirB1
<i>hp0524</i>	Cagβ/Cag5	MI	NTPase, recrutement du substrat	VirD4
<i>hp0525</i>	Cagα	C, MI	NTPase	VirB11
<i>hp0526</i>	CagZ	C, MI	Stabilisation de Cagβ	
<i>hp0527</i>	CagY	MI, ME, S	Facteur du complexe cœur, interaction avec les intégrines	VirB10
<i>hp0528</i>	CagX	MI, ME, S	Facteur du complexe cœur	VirB9
<i>hp0529</i>	CagW	MI	Facteur associé au complexe cœur	VirB6
<i>hp0530</i>	CagV	MI	Facteur associé au complexe cœur	VirB8
<i>hp0531</i>	CagU	MI	Facteur associé au complexe cœur	
<i>hp0532</i>	CagT	ME, S	Facteur du complexe cœur, lipoprotéine de la ME	VirB7
<i>hp0534</i>	CagS	C		
<i>hp0535</i>	CagQ	MI		
<i>hp0536</i>	CagP	MI		
<i>hp0537</i>	CagM	ME	Facteur du complexe cœur	
<i>hp0538</i>	CagN	P, MI	Facteur associé au complexe cœur	
<i>hp0539</i>	CagL	P, S	Biogénèse du pilus, ciblage des intégrines	VirB5-like
<i>hp0540</i>	CagI	P, S	Biogénèse du pilus, ciblage des intégrines	
<i>hp0541</i>	CagH	MI	Biogénèse du pilus, facteur associé au complexe cœur	
<i>hp0542</i>	CagG	P	Facteur accessoire	
<i>hp0543</i>	CagF	C, MI	Chaperone de CagA	
<i>hp0544</i>	CagE	MI	NTPase	VirB3/B4
<i>hp0545</i>	CagD	ME, P, S	Facteur accessoire	
<i>hp0546</i>	CagC	MI, ME, S	Sous-unité du pilus	VirB2
NA	CagB	C		
<i>hp0547</i>	CagA	C, S	Effecteur transloqué	

C : cytoplasme; ME; membrane externe; MI : membrane interne; NA : non annotée; P; périplasm; PG : peptidoglycane; S : exposé à la surface

Le CagSST4 d'*H. pylori* est un complexe macromoléculaire qui s'étend entre les deux membranes de la bactérie à Gram négatif et comprend trois sous-unités interconnectées : le complexe « cœur » est ancré dans la membrane externe et forme un pore qui s'étend entre les deux membranes ; le complexe ancré dans la membrane interne avec une partie cytoplasmique ; et le pilus. L'analyse récente du complexe « cœur » du CagSST4 par microscopie électronique a mis en évidence une architecture très différente de celle des autres SST4 [145]. Le complexe « cœur » se compose d'un anneau central de 19 nm constitué des protéines CagM, CagX et CagY, et d'un anneau externe d'un diamètre de 41 nm constitué de Cag3 et CagT [145] (Fig. I.14). Nos données suggèrent également que la protéine Cag δ serait localisée à la périphérie du complexe « cœur » [146].

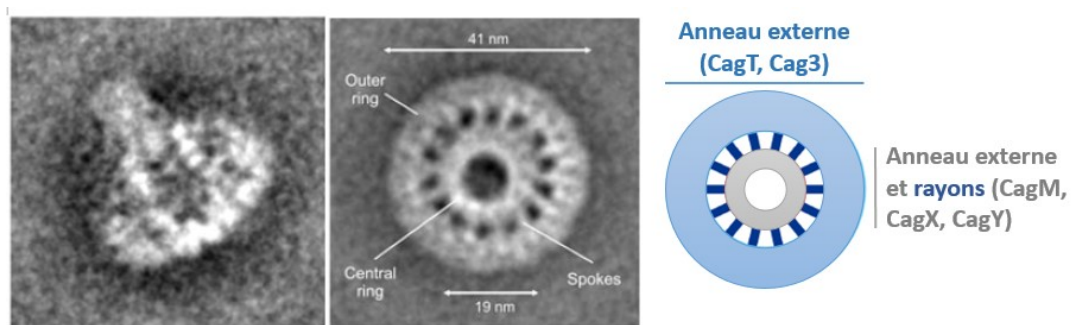


Figure I.14. Analyse du complexe cœur du CagSST4 par microscopie électronique. Vue de face et de côté du complexe cœur et représentation schématique. Les tailles des anneaux internes et externes sont indiquées sur la figure, ainsi que les protéines Cag qui les composent. Adapté de Frick-Cheng *et al.*, 2016 [145].

Le *cagPAI* code également pour le complexe de la membrane interne. Basé sur les similarités de séquence des protéines Cag avec celles du plasmide de conjugaison R388, dont la structure a été résolue récemment [147], il est proposé que le complexe de la membrane interne soit composé des protéines CagE, CagW et CagV [143]. Finalement, le *cagPAI* encode pour le pilus CagSST4. La biogenèse des pili du CagSST4 est induite à la suite du contact avec les cellules épithéliales gastriques, provoquant la formation de pili sur le pourtour de la bactérie [148] (Fig. I.15 A et B).

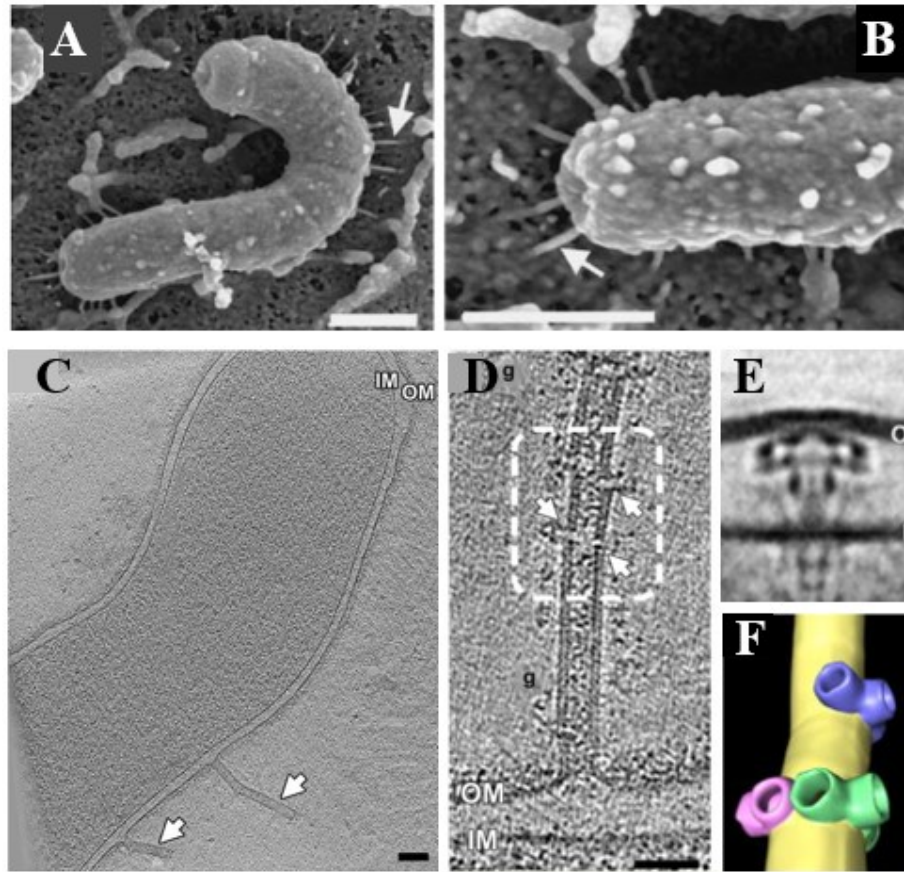


Figure I.15. Pili observés suite au contact direct d’*H. pylori* avec les cellules épithéliales gastriques AGS *in vitro*. (A-B) Images acquises par microscopie électronique à balayage à effet de champ (FESEM) haute résolution. Barre : 0,5 μm . Depuis Rhode *et al.* 2003 [148]. (C) Tomographie d’une coupe d’*H. pylori* mettant en évidence des structures de type tubes (pili) à la surface de la bactérie (flèches). Barre : 100 nm. (D) Structure tubulaire (pilus) présentant des pores latéraux (flèches). Barre : 50 nm. (E) Sous-tomogramme moyen de la structure du *cagSST4* identifié entre les membranes d’*H. pylori*. (C-D) Images acquises par cryotomographie électronique. (F) Représentation 3D de la région encadrée en D, représentant des pores latéraux. Depuis Chang *et al.* 2018 [149].

Les pili d’*H. pylori* ont récemment été observés par cryotomographie électronique (Fig. I.15 C et D) [149]. Induit par contact direct avec les cellules gastriques, ces pili présentent une longueur médiane de 193 nm. De façon intéressante, une structure qui correspondrait au complexe du CagSST4 fut également observée entre les membranes de la bactérie (Fig. I.15 E). Néanmoins, cette structure ne fut pas visualisée directement associée aux pili. De plus, des pores latéraux

furent observés sur le pourtour des pili. D'une largeur de 10-12 nm, ces pores paraissent suffisamment larges pour le transport de CagA repliée; cependant, leur fonction n'est pas élucidée [149].

Alors que l'architecture du pilus d'*H. pylori* paraît plus complexe que dans d'autres SST4, la composition du pilus n'a pas pu être déterminée avec exactitude du fait de données contradictoires. En effet, la protéine CagC fut initialement proposée pour être la piline majeure en raison de sa similarité de séquence avec la piline VirB2, identifiée dans plusieurs SST4, et de sa localisation à la surface de la bactérie [150]. Or, il fut ensuite observé que la biogenèse des pili n'était pas altérée dans un mutant $\Delta cagC$ [151], rendant la fonction proposée de CagC controversée. Le pilus semble également comprendre CagL qui présente des similarités de séquence avec VirB5, la piline mineure de plusieurs SST4. La localisation de CagL à la surface du pilus a été observée par immunofluorescence et la délétion du gène empêche la formation de pili [152]. Il est également proposé que les protéines CagH, CagI, CagY et CagX seraient associées au pilus [145, 148, 153]. Finalement, trois ATPases cytoplasmiques associées au complexe de la membrane interne, Cag α , Cag β et CagE, permettent d'énergiser le système afin de promouvoir l'assemblage du pilus et la sécrétion de CagA. Il est également proposé que Cag β soit un élément déterminant pour la reconnaissance de l'effecteur CagA [154]. Les autres protéines Cag ne sont pas indispensables à la formation du pilus CagSST4 mais certaines sont requises pour la translocation de CagA [144]. CagF est une protéine chaperonne qui s'associe à CagA pour permettre sa sécrétion [155], CagY est une peptidoglycane hydrolase qui faciliterait l'assemblage du complexe [156] et plusieurs protéines Cag semblent être des protéines accessoires. L'architecture du CagSST4 telle que postulée actuellement est schématisée dans la figure I.16.

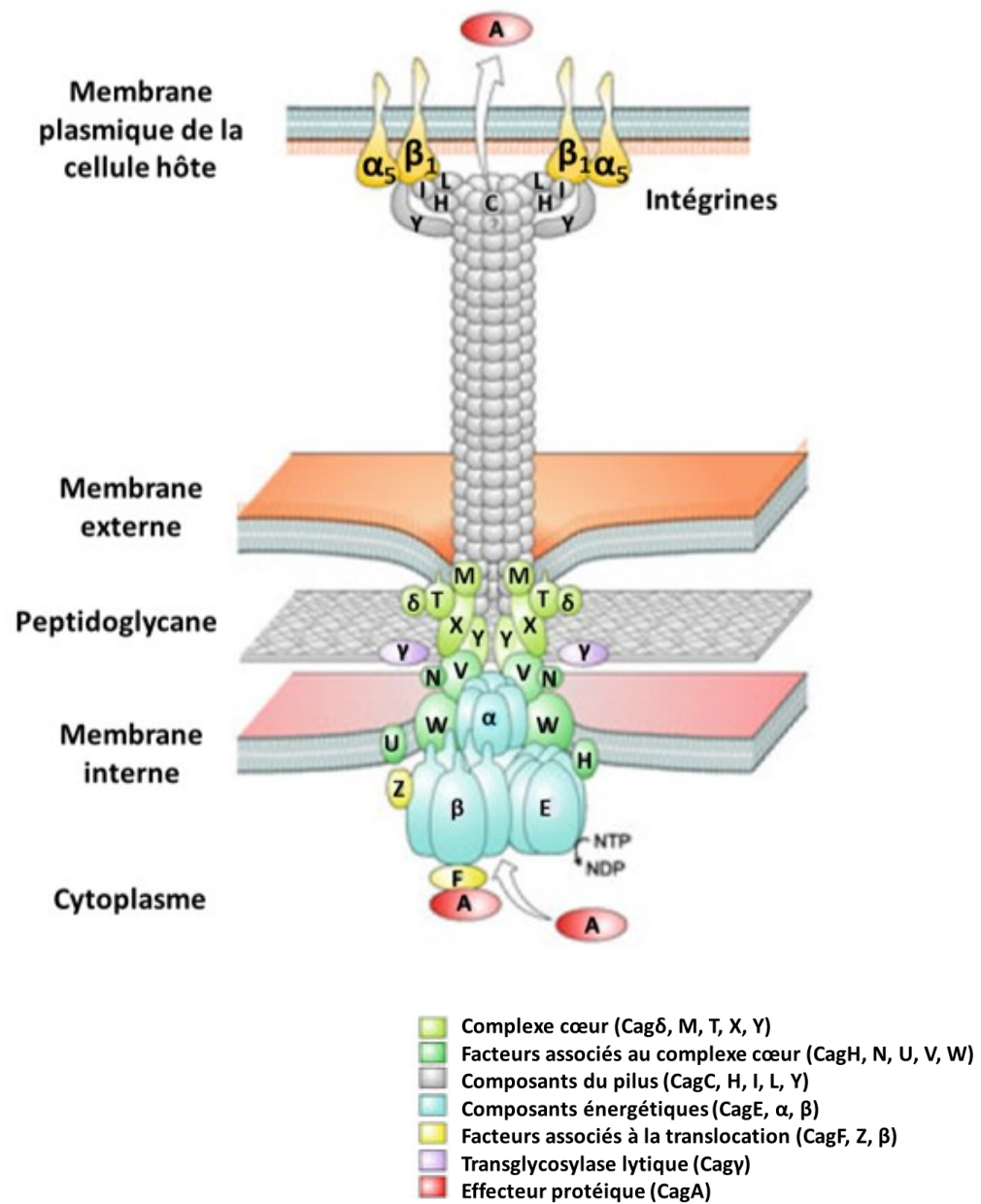


Figure I.16. Représentation schématique de l'architecture du CagSST4 d'*H. pylori* encodé par le *cagPAI*. Adapté de Backert *et al.*, 2017 [143].

Ensemble ces données suggèrent que le CagSST4 est unique, cette notion étant d'autant plus confortée par le fait que le CagSST4 est le seul SST4 connu à ce jour capable de transloquer, à la fois, un substrat protéique (CagA) [139], un substrat non protéique (peptidoglycane) [157], un substrat ADN [158] et un substrat lipidique (HBP : heptose-1,7-bisphosphate) [159].

2.2.3 Cag α : un composant énergétique du CagSST4

La fonctionnalité du CagSST4 d'*H. pylori* est assurée par les ATPases CagE, Cag α et Cag β . Celles-ci permettent de fournir l'énergie nécessaire à l'assemblage des composants du CagSST4 et/ou à la translocation des substrats dans la cellule hôte. Tandis que Cag β est responsable du recrutement de CagA, les ATPases CagE et Cag α sont requises pour son injection dans la cellule hôte [144]. En effet, la délétion du gène *cag α* , inhibe la translocation de CagA et l'induction d'une réponse pro-inflammatoire dans les cellules eucaryotes [144].

Cag α est l'homologue fonctionnelle de VirB11, dans le système VirB/D d'*A. tumefaciens*. Localisée du côté cytoplasmique et associée à la membrane interne, elle appartient à la famille des ATPases AAA+ de trafic. Cag α est une protéine hexamérique en forme de double anneau (Fig. I.17). Les domaines C-terminaux (CTD) de la protéine s'associent pour former un anneau hexamérique fermé monté sur anneau hexamérique ouvert et dynamique, formé par les domaines N-terminaux (NTD). Les domaines NTD et CTD sont reliés par un court « linker » flexible (Fig. I.17). Ainsi Cag α se caractérise par une structure en forme de dôme, ouverte à l'extrémité NTD et fermée à l'extrémité CTD. Les structures cristallines de Cag α , exempte de nucléotide (forme *apo-Cag α*) et complexée aux nucléotides (ADP et ATP γ S) ont été résolues [160, 161]. Ces structures révèlent que sous l'effet de la liaison d'ATP, un changement dans la conformation du NTD s'opère et conduit à la formation d'un anneau NTD fermé. Cag α adopte alors une conformation hexamérique en double anneau, parfaitement symétrique et rigide, jusqu'à l'hydrolyse des molécules d'ATP et la libération des nucléotides. Les changements dynamiques dans la conformation de Cag α sont ainsi régulés par les étapes de liaison/libération des nucléotides. Bien que les structures des homologues de Cag α présentent des différences (ex : VirB11 d'*A. tumefaciens* et de *Brucella suis*), cette dynamique structurale est également observée.

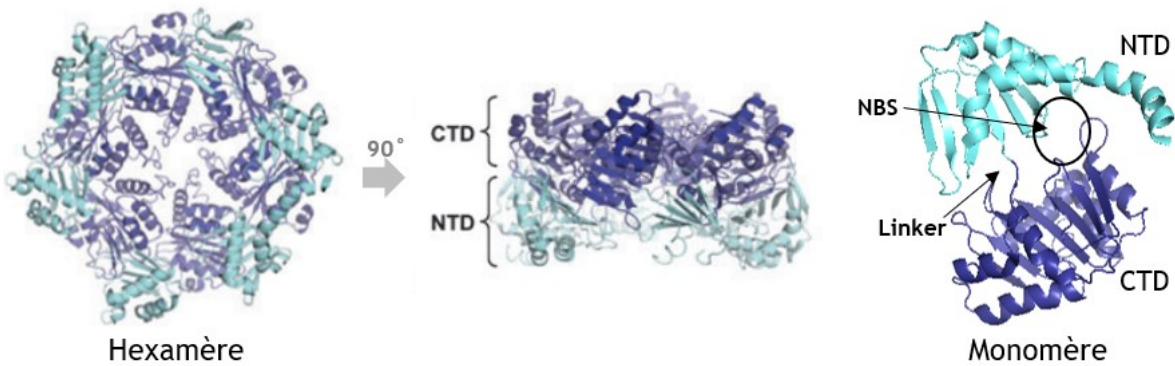


Figure I.17. Représentation de la structure cristallographique de Cag α . La forme hexamérique en double anneau ainsi que le monomère de Cag α , composé des domaines N-terminaux (NTD) et C-terminaux (CTD), sont illustrés. Le site de liaison (linker) entre le NTD et le CTD est indiqué. Le site d'interaction des nucléotides (NBS) est également représenté. Adapté de Terradot et *al.*, 2011 [162].

2.2.4 Le pilus CagSST4 permet l'adhésion à la cellule hôte et l'injection des effecteurs

Le *cagPAI* est un segment d'ADN de 40 kb qui code pour un SST4, dont la présence accroît considérablement la pathogénicité des souches *H. pylori* qui le possèdent. Ainsi les personnes infectées par une souche *H. pylori* *cag*-positive développent un ulcère gastroduodéal dans 10-15% des cas tandis qu'un adénocarcinome gastrique est développé dans 1-2% des cas. Ceci représente un risque 2,2 fois plus élevé que chez des personnes infectées par une souche *cag*-négative et six fois plus élevé que chez des personnes non infectées par la bactérie [163]. *H. pylori* exploite le pilus du CagSST4 pour transloquer différentes molécules effectrices dans les cellules épithéliales gastriques dont l'oncoprotéine CagA, du peptidoglycane (PG) bactérien, de l'ADN et l'HBP, un précurseur de la biosynthèse des lipopolysaccharides (LPS) [139, 157-159]. La translocation de ces effecteurs, et de CagA principalement, qui repose, entre autres, sur l'interaction de CagL avec les intégrines des cellules épithéliales gastriques, détourne la signalisation intracellulaire de la cellule hôte.

2.2.4.1 Intégrines : les récepteurs du CagSST4

La biogenèse des pili du CagSST4 est induite par un contact direct avec les cellules épithéliales gastriques [148]. Ceci suggère notamment la présence de récepteurs du CagSST4 à la surface de la cellule hôte, permettant la reconnaissance et l'adhésion de la bactérie. L'étape d'adhésion est essentielle pour la translocation de CagA et des autres effecteurs CagSST4-dépendant. Le mécanisme d'interaction le mieux caractérisé entre le pilus CagSST4 et les cellules épithéliales gastriques est celui impliquant des intégrines.

Les intégrines sont essentielles pour les processus d'adhésion cellulaire ainsi que pour l'interaction avec la matrice extracellulaire [164]. Ces protéines appartiennent à une famille de récepteurs transmembranaires hétérodimériques constitués des sous-unités α et β . Les deux sous-unités forment un récepteur flexible constitué d'une partie extracellulaire, d'une courte région transmembranaire et d'une partie cytoplasmique. Les intégrines reconnaissent généralement les motifs RGD (Arg-Gly-Asp) de leurs ligands. Suite à l'interaction avec le ligand, un changement conformationnel s'opère et des cascades de signalisation intracellulaire sont activées. Les intégrines sont la cible de nombreux pathogènes qui les utilisent comme récepteurs pour promouvoir l'adhésion, l'invasion de la cellule hôte et l'injection de facteurs de virulence [165, 166]. Il a été montré que, chez *H. pylori*, quatre protéines Cag étaient capables d'interagir avec les intégrines de la cellule hôte, à savoir CagA, CagL, CagI et CagY [152, 167]. Le peu d'informations structurales disponibles concernant CagI et CagY ne permet pas d'expliquer précisément l'incidence de ces interactions. En revanche, l'interaction de CagL, et dans une moindre mesure de CagA, avec les intégrines a été plus largement étudiée.

2.2.4.2 CagL : une adhésine du CagSST4

CagL est essentielle pour la biogenèse des pili, requise pour l'injection de la cytotoxine CagA dans la cellule hôte et pour l'induction de la réponse pro-inflammatoire menant à la production d'interleukines [144, 152, 153]. En plus de présenter une similarité de séquence avec la piline mineure VirB5, elle a été détectée à la surface des pili par des analyses d'immunofluorescence et de microscopie électronique [152]. Son interaction avec les intégrines de la cellule hôte eucaryote a également été démontrée de façon expérimentale [152, 168, 169]. Ainsi, CagL semble jouer un rôle d'adhésine entre le pilus de la bactérie et les cellules gastriques.

L'interaction de CagL avec les intégrines se fait généralement par l'intermédiaire de son motif conservé RGD qui paraît important, bien que dispensable pour l'interaction avec les intégrines, comme l'atteste des expériences de mutagenèse [152]. De façon intéressante, la résolution de structures cristallographiques de CagL dans différentes conditions expérimentales [170-173] a révélé que le domaine RGD était enfoui dans la structure de CagL à pH acide et que la neutralisation du pH induisait un changement conformationnel de la structure de la protéine, rendant le domaine RGD accessible [171, 172]. Les motifs LXXL (Leu/Met-X-X-Leu/Iso) et FEANE (Phe-Glu-Ala-Asn-Glu) qui avoisinent le motif RGD, semblent également améliorer l'efficacité d'interaction avec les intégrines [169, 174]. La mise en évidence de l'interaction de CagL avec l'intégrine $\alpha 5\beta 1$ a permis de démontrer, du même fait, l'importance de l'interaction entre le CagSST4 et les intégrines pour l'injection de CagA [152]. Le domaine RGD de CagL est également impliqué dans la reconnaissance des intégrines $\alpha V\beta 6$, $\alpha V\beta 3$ et $\alpha V\beta 8$, [168, 169], tandis que CagL interagit avec l'intégrine $\alpha V\beta 5$ de manière RGD-indépendante [168].

Tandis que l'interaction de CagL avec les intégrines des cellules épithéliales gastriques est essentielle pour la translocation de CagA, son interaction avec l'intégrine $\alpha 5\beta 1$ active également plusieurs kinases intracellulaires ainsi que le facteur de transcription NF- κ B, conduisant notamment à une réponse pro-inflammatoire et à la production d'interleukine-8 dans la cellule hôte [175, 176]. L'interaction avec l'intégrine $\alpha V\beta 5$ conduit, quant à elle, à une surexpression de la gastrine, une hormone responsable de la régulation de la sécrétion d'acide gastrique [168].

2.2.4.3 CagA : la protéine effectrice du CagSST4

2.2.4.3.1 CagA participe à sa propre translocation dans la cellule hôte

Bien qu'une grande quantité de protéines CagA soit synthétisée dans *H. pylori*, seules 20% des protéines produites sont injectées dans les cellules gastriques ce qui suggère que, outre sa fonction de cytotoxine, CagA serait impliquée dans d'autres processus métaboliques [177]. La protéine CagA a été localisée au bout du pilus CagSST4. Il fut également montré que le domaine N-terminal de CagA interagissait avec les intégrines $\alpha 5\beta 1$ et $\alpha V\beta 3$ avec une très forte affinité [167]. D'un point de vue mécanistique, l'interaction de CagA avec les intégrines est intéressante car elle suggère que la protéine participerait activement à sa propre translocation dans les cellules épithéliales gastriques [178]. De plus, sa localisation au bout du pilus lui permet de lier

la phosphatidylsérine (PS), un composant de la double couche lipidique membranaire de la cellule hôte qui devient exposé du côté extracellulaire lors de l'infection par *H. pylori*, favorisant sa propre internalisation dans la cellule hôte [179]. Néanmoins, la capacité de CagA à interagir avec des récepteurs de la cellule hôte ne suffit pas à permettre sa translocation, ce qui souligne l'importance des autres composants du CagSST4 dans le processus.

2.2.4.3.2 CagA : une protéine polymorphe

CagA est une protéine de 120-145 kDa (dépendamment de la souche d'*H. pylori*) encodée par le *cagPAI*, qui ne présente de similarité de séquence avec aucune autre protéine. Décrite comme une oncoprotéine, elle est le seul substrat protéique connu du CagSST4.

Une fois injectée dans les cellules hôtes, une partie des protéines CagA sont phosphorylées sur les résidus tyrosines (CagA^{PY}), au niveau du motif répété EPIYA (Glu-Pro-Ile-Tyr-Ala), par des kinases de l'hôte [180]. De façon intéressante, quatre motifs EPIYA ont été décrits, EPIYA-A, -B, -C et -D, chaque motif constitué d'une séquence conservée [181]. Les motifs EPIYA-A et EPIYA-B sont présents dans toutes les souches d'*H. pylori*, indépendamment de l'origine géographique. Tandis que le motif EPIYA-C est retrouvé, en nombre variable (d'une à six répétitions), dans les isoformes CagA des pays occidentaux, le motif EPIYA-D prédomine dans les isoformes des pays d'Asie [182, 183] (Fig. I.18). Les motifs EPIYA sont des déterminants clés pour l'interaction de CagA avec les protéines de la cellule hôte. L'affinité de CagA pour certaines de ces protéines est dépendante du motif présent (EPIYA-C ou-D). Il a également été observé qu'un grand nombre de répétitions du motif EPIYA-C ou la présence d'un motif EPIYA-D étaient associés à des risques élevés de développement de cancers gastriques [184].

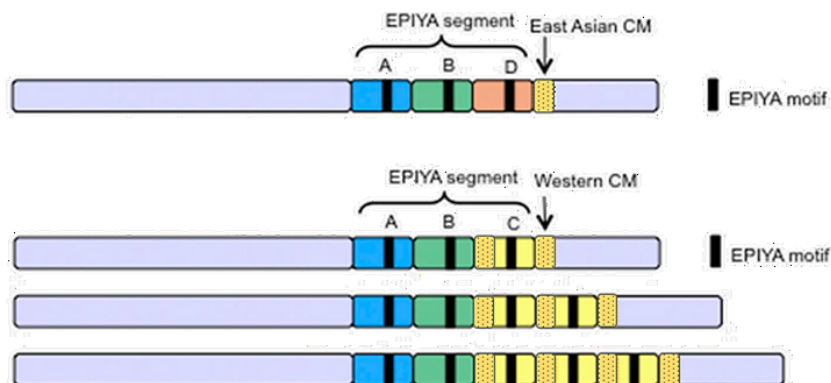


Figure I.18. Diversité structurale des motifs EPIYA de CagA. Les souches *H. pylori* isolées dans les pays d'Asie contiennent généralement les motifs EPIYA-A/B/D alors que celles issues des pays occidentaux contiennent les motifs EPIYA-A/B/C. Le motif EPIYA-C est répétable. Chaque motif contient un site de phosphorylation sur une tyrosine unique schématisé en noir. Le domaine CM représente un domaine de multimérisation de CagA dont la séquence varie également en fonction de l'origine géographique. Les souches d'Asie ne possèdent qu'un seul motif CM immédiatement en aval du segment EPIYA-D, tandis que celles d'Occident possèdent au moins deux motifs CM, un dans le segment EPIYA-C et l'autre en aval du dernier segment EPIYA-C. Depuis Hatekamaya *et al.*, 2014 [141].

La région C-terminale de CagA contient également le motif CM (*CagA multimerization*), un motif de 16 résidus impliqué dans la dimérisation de la protéine [185]. Alors que le motif CM est hautement conservé, cinq résidus varient entre les isoformes CagA d'Asie et d'Occident. Tandis que dans les souches *H. pylori* d'Asie le motif CM est localisé directement en aval du motif EPIYA-D, dans les souches d'Occident le nombre de motifs CM est variable puisque chaque segment EPIYA-C présente un motif en son sein en plus du motif situé en aval du dernier motif EPIYA-C (Fig. I.18). Le motif CM est également ciblé par des protéines de la cellule hôte.

2.2.4.3.3 CagA détourne les voies de signalisation cellulaire de la cellule hôte

Une fois transloquée dans les cellules gastriques, CagA interagit avec plus de 25 protéines de la cellule hôte, dépendamment de son état de phosphorylation, déclenchant des cascades de signalisation variées. Ces partenaires d'interaction ont été largement étudiés et sont abordés en détail dans de nombreuses revues [186-190]. Les kinases eucaryotes responsables de la phosphorylation des motifs EPIYA appartiennent à la famille des kinases Src et Abl. Tandis

qu'Ab1 peut phosphoryler chacun des motifs EPIYA, Src ne phosphoryle que les motifs EPIYA-C/D [191]. De façon intéressante, les deux kinases sont considérées comme des acteurs clés dans l'oncogenèse.

La tyrosine-phosphatase SHP2 fut le premier partenaire d'interaction de CagA phosphorylée identifié. Son affinité d'interaction avec CagA est dépendante de la présence d'un motif EPIYA-C/D [192]. Tout comme SHP1, un autre partenaire d'interaction de CagA, elle est impliquée dans les voies de signalisation d'une variété de facteurs de croissance et de cytokines. CagA phosphorylée lie également la kinase PI3 impliquée dans la mobilité cellulaire et la régulation du cycle cellulaire. Par l'intermédiaire des motifs EPIYA phosphorylés, CagA interagit avec les protéines adaptatrices de la famille Crk et de la famille Grb, ainsi qu'avec la protéine Ras-GAP, qui interviennent en amont des voies de signalisation oncogénique, et notamment dans l'activation de la voie des kinases MAP. Or, chacune de ces protéines contient le domaine SH2 (*Src Homology 2*) qui permet une reconnaissance des phosphotyrosines chez les eucaryotes. Ainsi, CagA phosphorylée semble imiter les protéines eucaryotes endogènes, également phosphorylées sur des résidus tyrosine, et détourner la signalisation cellulaire de l'hôte [193]. La cascade de signalisation induite par CagA phosphorylée est illustrée dans la figure I.19.

D'autre part, CagA interagit avec des protéines de la cellule hôte indépendamment de son état de phosphorylation. Parmi ses partenaires d'interaction, on dénombre la protéine adaptatrice GRB2, des protéines de jonction cellulaire (ZO1, JAM, E-cadhérine), divers facteurs de transcription impliqués dans la prolifération cellulaire et l'inflammation (β -caténine, Smad3), le suppresseur de tumeur RUNX3 et le facteur d'apoptose ASPP2. De plus CagA lie la kinase PAR1, impliquée dans la polarité cellulaire, et la protéine c-Met qui intervient dans la prolifération et la morphologie cellulaire, par l'intermédiaire du motif CM. La cascade de signalisation induite par CagA non phosphorylée est représentée dans la figure I.19.

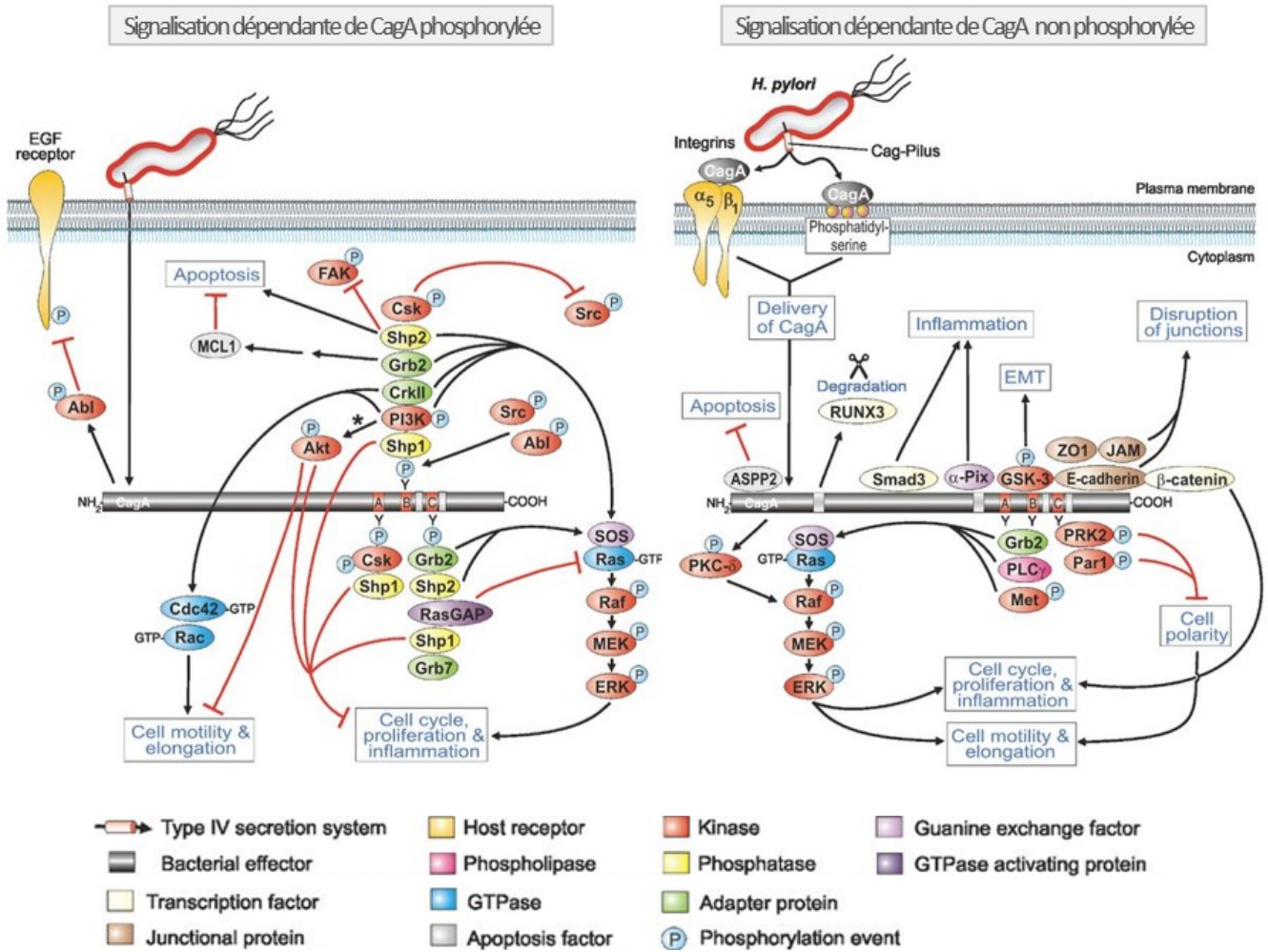


Figure I.19. Schématisation des voies de signalisations cellulaires activées dans la cellule hôte par CagA phosphorylée et non phosphorylée. Depuis Backert *et al.*, 2016 [189].

Ces interactions ont des conséquences multiples sur les cellules hôtes. CagA induit notamment la perte de l'adhésion cellulaire, stimule la mobilité cellulaire, le réarrangement du cytosquelette et l'élongation cellulaire. Ces événements conduisent à une modification de la morphologie cellulaire qui devient fusiforme; ce phénotype est appelé « *Hummingbird* » (colibri). D'autre part, l'association de CagA avec les protéines de jonction provoque l'altération des jonctions cellulaires, or celles-ci sont essentielles pour préserver l'intégrité de l'épithélium gastrique. L'interaction avec PAR1 favorise également la perte de la polarité cellulaire. De plus, CagA induit un dérèglement du cycle cellulaire, la prolifération cellulaire et l'inhibition de l'apoptose. Le facteur de transcription NK- κ B est, lui, activé par une cascade de signalisation et stimule la

libération de chimiokines et d'interleukines, dont l'interleukine 8, un facteur déterminant de l'inflammation chronique. CagA stimule également la β -caténine de manière aberrante, un facteur de transcription connu pour induire l'expression de gènes associés aux cancers [194] et module l'activité des suppresseurs de tumeurs [195]. De plus, CagA induit un phénotype semblable au phénotype EMT (*Epithelial-Mesenchymal Transition*), qui constitue une étape précoce dans la formation de métastases. Il est ainsi proposé que, conjointement, ces réponses provoquent des lésions pré-malignes dans l'estomac, pouvant aboutir à la formation de cancers gastriques. Le rôle de CagA dans l'oncogenèse a d'ailleurs été confirmé à la suite de l'infection de gerbilles de Mongolie par une souche d'*H. pylori* *cagA*⁺ qui, après seulement 16 semaines, développent une succession de lésions aboutissant au développement d'adénocarcinomes gastriques invasifs ; ces effets n'ont pas été observés en réponse à l'infection par un *mutant H. pylori* isogénique *cagA*⁻ [196]. Les conséquences de la signalisation induite par CagA sont schématisées de façon simplifiée dans la figure I.20.

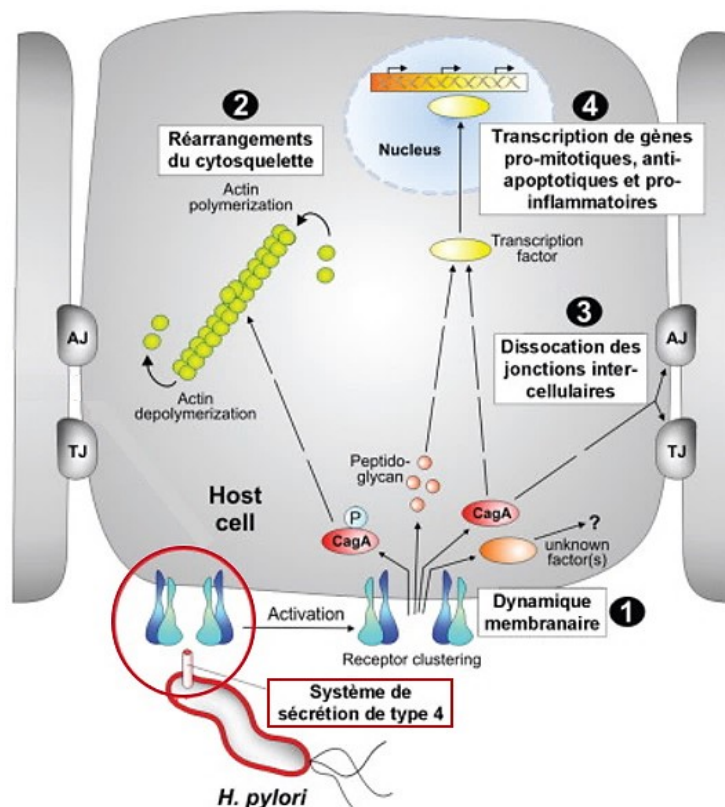


Figure I.20. Représentation simplifiée des conséquences de l'injection de CagA sur les cellules gastriques. Depuis Varona *et al.* [197].

2.2.4.4 Autres facteurs transloqués par le CagSST4: peptidoglycane, ADN et HBP

En plus de l'oncoprotéine CagA, *H. pylori* exploite le pilus du CagSST4 pour transloquer du peptidoglycane bactérien, de l'ADN et l'HBP [139, 157-159]. La délivrance du peptidoglycane bactérien par le CagSST4 déclenche la voie NOD1, qui, tout comme les TLR, assure la reconnaissance des bactéries pathogènes par la cellule eucaryote. L'activation de NOD1 induit la réponse pro-inflammatoire par activation du facteur de transcription NF- κ B, ayant pour effet de faciliter la survie bactérienne et la réplication dans l'hôte [157]. Il a également été observé que le CagSST4 permettait la translocation d'ADN génomique dans les cellules hôtes [158]. L'ADN transloqué active le récepteur TLR9, connu pour moduler à la fois la signalisation pro- et anti-inflammatoire dans les cellules épithéliales intestinales. Dans les cellules épithéliales gastriques sa fonction n'est pas élucidée, mais il est proposé que l'activation de TLR9 par *H. pylori* permettrait de réguler la réponse inflammatoire et ainsi de promouvoir la persistance de la bactérie dans la niche gastrique [158]. Finalement, le dérivé bactérien HBP, un précurseur métabolique dans la biosynthèse des LPS, semble lui aussi être injecté dans les cellules par l'intermédiaire du CagSST4. L'HBP active la voie TIFA (*TRAF-Interacting protein with a Forkhead-Associated domain*) qui conduit également à une réponse pro-inflammatoire NF- κ B-dépendante [159].

Hormis les effecteurs dépendants du CagSST4, plusieurs facteurs produits par *H. pylori* et impliqués dans sa virulence ont été identifiés, les principaux étant VacA et les OMPs.

2.3 Facteurs de virulence indépendants du CagSST4

2.3.1 VacA (*Vacuolating cytotoxin*)

VacA est une toxine sécrétée par *H. pylori*, dont le nom est dérivé de sa capacité à induire la formation de vacuoles dans les cellules gastriques *in vitro* [198]. Les effets de VacA sur les cellules épithéliales gastriques sont variés et comprennent la perturbation de la voie d'endocytose, l'altération des mitochondries, la dépolarisation du potentiel de membrane plasmique et son changement de perméabilité, l'efflux de divers ions (dont le chlorure, le bicarbonate et l'urée), l'activation des kinases MAP, la modulation de l'autophagie, et l'apoptose [199-201]. VacA peut inhiber la fonction et la prolifération d'une variété de cellules

immunitaires, y compris les cellules T, les cellules B, les éosinophiles, les macrophages, les cellules dendritiques et les neutrophiles [200, 202]. Le gène *vacA* est présent dans la plupart des souches d'*H. pylori* et est hautement polymorphe. Il comprend deux variants alléliques alternatifs pour les régions signal (s1/s2), intermédiaire (i1/i2) et moyenne (m1/m2) [203] (Fig. I.21).

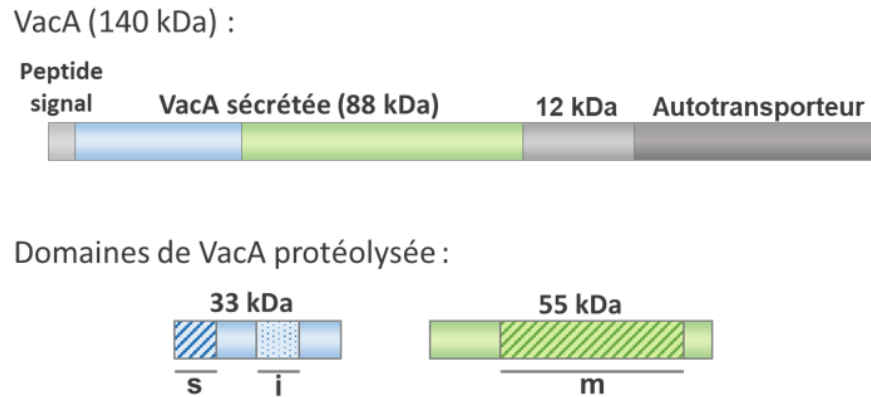


Figure I.21. Organisation de VacA et régions polymorphes. La protéine VacA de 140 kDa comprend un peptide signal, un domaine de 88 kDa sécrété, un peptide de 12 kDa de fonction inconnue, et un domaine C-terminal autotransporteur de type feuillet β . La région de VacA sécrétée subit un clivage protéolytique qui conduit à l'obtention de deux domaines, p33 (bleu) et p55 (vert). Les trois régions polymorphes (s, i et m) sont identifiées. Adapté de Foegeding *et al.*, 2016 [204].

Tandis que la région m est impliquée dans l'interaction avec les cellules épithéliales, le variant m1 permettant de lier une plus large gamme de types cellulaires, les variants s1 et i1 présentent une activité de vacuolisation amplifiée par rapport aux variants s2 et i2 [203, 205]. Ces variants sont d'ailleurs associés à un risque accru de développement d'ulcère gastroduodénal, d'atrophie et d'adénocarcinome gastrique; néanmoins les niveaux d'expression de *vacA* sont également importants. De façon intéressante, la majorité des souches d'*H. pylori* *cag*-positives possèdent les allèles s1/i1 de *vacA*. La délétion de *vacA* n'a pas d'effet sur la capacité de la souche à transloquer CagA ou sur l'induction d'une réponse pro-inflammatoire, et la délétion de *cagA* n'altère pas non plus l'activité de vacuolisation de VacA [206]. Toutefois les expériences de mutagenèse ont révélé qu'un mécanisme de régulation s'opérait entre CagA et VacA, les deux

protéines atténuant leurs effets réciproques sur les cellules épithéliales, dans le but d'éviter de causer des dommages excessifs aux cellules et de favoriser la persistance de l'infection [206].

2.3.2 OMPs (*Outer membrane proteins*)

H. pylori contient plus de 30 gènes codant pour des protéines de la membrane externe (OMPs), dont plusieurs sont associées à la virulence, parmi lesquelles BabA, SabA, OipA, AlpA/B et HopQ. La protéine BabA est l'adhésine la mieux caractérisée dont le récepteur cellulaire est l'antigène fucosylé Lewis^b (Le^b), exprimé à la surface des cellules épithéliales gastriques [207]. Bien qu'une occurrence entre la présence de l'allèle *babA2* et le développement d'ulcère gastroduodéal ait été montrée, les données sont controversées. L'allèle *babA2* semble néanmoins stimuler la colonisation d'*H. pylori* et l'inflammation [208]. SabA est une adhésine qui reconnaît une forme sialylée des antigènes du système Lewis X. Elle est essentielle pour l'adhésion d'*H. pylori* aux cellules épithéliales gastriques et pour la persistance de l'infection, et constitue l'adhésine prédominante dans les cas de gastrite chronique [209]. La protéine OipA induit la production d'interleukine-8 et stimule le réarrangement du cytosquelette, elle est aussi associée à un risque accru de développement d'ulcère duodéal et de cancer gastrique [210]. Les adhésines AlpA et AlpB sont, quant à elles, importantes pour l'adhésion aux cellules gastriques [211]. Finalement, l'adhésine HopQ se lie aux CEACAM1, une classe de glycoprotéines apparentées à la superfamille des immunoglobulines, exprimées à la surface des cellules épithéliales humaines. De façon intéressante, la délétion de *hopQ* inhibe la translocation de CagA et l'induction de l'interleukine-8 [212, 213]. Ainsi HopQ est la première protéine non encodée par le *cagPAI* qui fut montrée comme étant essentielle à la fonctionnalité du CagSST4.

2.3.3 Autres facteurs de virulence

D'autres facteurs ont été associés à la virulence d'*H. pylori*. Parmi ceux-ci, le gène *dupA* et l'allèle *iceA*, tous deux associés à une production accrue d'interleukine-8 et à la présence d'ulcères gastroduodéaux [214, 215]. Toutefois les fonctions exactes de ces deux protéines sont méconnues. D'autre part, la protéine NapA produite par *H. pylori* recrute les neutrophiles et stimule la production des cytokines pro-inflammatoires et des chimiokines. La délétion de *napA* stimule également la production d'espèces réactives de l'oxygène (ROS) et inhibe

l'adhérence d'*H. pylori* à la cellule hôte. NapA pourrait donc jouer un rôle central dans la pathogénicité d'*H. pylori* et favoriserait l'établissement d'une inflammation chronique [216].

2.4 Recherche sur les molécules d'anti-virulence, génotypage de souches d'*H. pylori* et marqueurs de pathogénicité

2.4.1 Développement des molécules d'anti-virulence comme alternative thérapeutique

Du fait de la propagation de souches bactériennes multirésistantes aux antibiotiques, le développement de stratégies alternatives aux antibiotiques devient urgent. C'est dans ce contexte qu'a émergé la recherche sur les molécules d'anti-virulence, qui continue de se développer. Les molécules d'anti-virulence sont conçues pour désarmer les pathogènes en neutralisant leurs facteurs de virulence, sans toutefois les éliminer. La pression de sélection induite sur les bactéries, et à l'origine du développement de résistance, est ainsi réduite, et confère un fort potentiel thérapeutique à ces molécules [217-220].

La recherche sur les molécules d'anti-virulence a été initiée sur le système de sécrétion de type III (SST3), un facteur de virulence répandu chez des pathogènes de plantes et d'animaux [221, 222]. Les premières études publiées ciblaient, en effet, les SST3 des bactéries *Yersinia pseudotuberculosis* et *Escherichia coli* entéro-pathogénique (EPEC) [223, 224]. Depuis, plusieurs molécules inhibant la virulence des bactéries ont été approuvées par la FDA (Food and Drug Administration) et de nombreuses molécules sont en phases d'essais cliniques [225].

En revanche, peu d'études ciblant le SST4 sont disponibles. Notre groupe a, néanmoins, reporté la découverte d'inhibiteurs de virulence de *Brucella suis*, ciblant la protéine VirB8, un facteur d'assemblage du SST4 [226, 227], ainsi que son homologue dans le plasmide de conjugaison pKM101 [228]. Des molécules inhibitrices de la virulence du pathogène *H. pylori* ont également été identifiées récemment, dans trois études dont deux ciblaient l'ATPase Cag α [229-231]. Tandis que la première étude met en évidence la non-spécificité des molécules inhibitrices envers la bactérie *H. pylori*, le mode d'action des molécules n'a pas été clairement caractérisé

dans les deux autres. Ainsi, un effort de recherche supplémentaire est nécessaire pour identifier des molécules ayant le potentiel d'inhiber la virulence d'*H. pylori* de façon spécifique.

2.4.2 Génotypage des souches *H. pylori* et marqueurs de pathogénicité

Les bactéries *H. pylori* présentent une grande diversité génétique, du fait de taux élevés de mutations et de la fréquence des recombinaisons entre souches [232, 233]. De plus, les souches d'*H. pylori* sont subdivisées en sous-populations distinctes, dont la diversité génétique est étroitement liée à la localisation géographique et reflète des migrations humaines ancestrales [234]. La distribution géographique globale des populations d'*H. pylori*, basée sur l'analyse de sept gènes de ménages de la bactérie, est à présent bien établie. Cependant, peu d'informations sont disponibles concernant la variabilité au sein des facteurs de virulence, tels que le *cagPAI*, ou l'association avec des pathologies gastriques sévères. Plusieurs séquences complètes de l'îlot *cagPAI*, provenant notamment de sous-populations d'*H. pylori* d'Asie, d'Europe et d'Afrique sont, néanmoins, accessibles dans les banques de données. L'analyse de ces séquences a, entre autres, mis en évidence que près d'un tiers des protéines encodées par le *cagPAI* présentaient des variabilités entre sous-populations [235].

Depuis le début des années 2010, plusieurs groupes tentent d'identifier des variations génétiques dans la bactérie *H. pylori*, notamment dans les protéines encodées par le *cagPAI*, qui pourraient influencer la sévérité de l'infection. Il est ainsi proposé que la variation CagL-Glu59 soit associée à un risque accru de développement de cancers gastriques [236, 237]. Néanmoins, seul l'impact de CagA sur l'oncogenèse a été confirmée [196]; et la présence d'isoforme CagA de type Asie de l'Est constitue un risque accru de carcinogenèse gastrique [238]. Ainsi, actuellement, seule CagA peut être considérée comme un marqueur de pathogénicité pour l'infection par *H. pylori*.

La bactérie *H. pylori* utilise différentes stratégies qui lui ont permis de s'adapter aux conditions hostiles de l'estomac humain et de perdurer dans l'organisme. A long terme, la chronicité de l'infection peut induire le développement de pathologies gastriques sévères, dont l'adénocarcinome gastrique, 3^e cause mondiale de mortalité par cancer. Tandis que la sévérité de l'infection dépend largement de la présence du *cagPAI* qui encode pour un système de

sécrétion de type 4, la variabilité génétique observée entre les souches d'*H. pylori* pourrait également être en cause. Néanmoins, CagA constitue le seul marqueur de pathogénicité identifié à ce jour. Du fait de l'émergence des souches multi-résistantes aux antibiotiques, les stratégies d'éradication de la bactérie, employées jusqu'alors, deviennent inefficaces. Le développement de stratégies alternatives pour le traitement de l'infection, telles que le développement de molécules d'anti-virulence, devient donc nécessaire et constituera un objectif à long terme de nos travaux.

I.3. OBJECTIFS DE LA THÈSE

Ce projet a été élaboré autour de l'étude de deux complexes macromoléculaires.

La première partie du projet a consisté en l'étude de la machinerie de synthèse et d'incorporation des sélénocystéines dans les cellules humaines. Tandis que les composants clés de la machinerie sont à présent définis, les détails moléculaires des interactions, la séquence d'assemblage et la structure de certains composants doivent encore être clarifiés. Cette machinerie reposant notamment sur une dynamique d'interactions transitoires entre de multiples facteurs protéiques, nous proposons que toutes les interactions n'aient pas été élucidées. Ainsi notre premier objectif fut de mettre au point une approche *in vivo* permettant l'étude des interactions entre l'ensemble des composants de la machinerie. Nous nous sommes, par la suite, intéressés à la protéine SECp43, une protéine centrale de la machinerie pour laquelle aucune donnée structurale n'était disponible. Dans ce contexte, nos objectifs furent de résoudre la structure tridimensionnelle de la protéine et d'identifier les résidus impliqués dans sa dimérisation.

La seconde partie du projet a été consacrée à l'étude du système de sécrétion de type 4 d'*H. pylori*. Du fait de l'émergence des souches d'*H. pylori* résistantes aux antibiotiques, nous suggérons, comme stratégie alternative, le développement d'inhibiteurs de virulence. Ainsi, nous proposons que le ciblage des protéines du système de sécrétion de type 4 et, plus spécifiquement, de l'ATPase Cag α , par des petites molécules inhibitrices, réduirait la pathogénicité et la virulence d'*H. pylori*. Après avoir identifié et caractérisé des molécules liant Cag α *in vitro*, notre objectif fut d'évaluer les effets des molécules identifiées sur la fonctionnalité du SST4 d'*H. pylori* *in vivo*.

Finalement, nous avons entrepris d'analyser la variabilité de souches d'*H. pylori* résistantes aux antibiotiques, isolées au Québec, à partir de biopsies de patients. Notre objectif fut de caractériser ces souches, au niveau génétique, en se focalisant sur les gènes du *cagPAI*, d'établir les profils de virulence des souches *in vivo*, puis d'évaluer l'efficacité de l'inhibiteur de Cag α sur ces souches.

Chapitre II : RÉSULTATS

II.1 Analyse de nouvelles interactions entre les composants de la voie de biosynthèse de la sélénocystéine SEPHS1, SEPHS2, SEPSECS et SECp43

II.1.1 Présentation de l'article 1 et contribution des coauteurs

Le premier objectif de ce projet a consisté à optimiser une technique permettant l'étude des interactions entre les composants de la machinerie de synthèse et d'incorporation de la Sec *in vivo*. Notre but fut notamment d'identifier de nouvelles interactions afin d'améliorer la compréhension du système. Ainsi le BRET fut utilisé pour étudier les interactions *in vivo*, et celles-ci furent confirmées par co-immunoprécipitation. Dans un second temps, nous avons eu pour objectif de résoudre la structure de SECp43. Après avoir purifié la protéine, une analyse de SAXS a permis de définir une structure globale de sa région N-terminale. Le BRET ayant mis en évidence que SECp43 dimérisait, nous avons par la suite vérifié que la protéine purifiée dimérisait également *in vitro*, par des essais de *crosslinking* et par gel filtration. Les sites de dimérisation potentiels de SECp43 ont été identifiés par *phage display* et l'effet de la substitution en alanine des résidus de cette région a été analysé par BRET.

Dans le cadre de ce projet, j'ai construit l'ensemble des vecteurs d'expression requis pour les analyses d'interaction et ai réalisé les expériences de culture cellulaire. J'ai mis au point et optimisé la technique du BRET au sein de notre laboratoire et ai effectué l'ensemble des expériences de BRET. J'ai également confirmé les interactions par co-immunoprécipitation. De plus, j'ai produit et purifié la protéine SECp43, réalisé les expériences de *crosslinking* et de *phage display* et j'ai participé aux premières étapes de validation des données de SAXS. Finalement, j'ai participé à l'écriture du manuscrit dans son ensemble.

Bastien Casu a collecté et analysé les données de SAXS. Il a réalisé les approches bio-informatiques et a généré les modèles structuraux de SECp43 relatifs aux données de SAXS. Finalement, il a contribué à l'écriture du manuscrit. Arnold Steve Dopgwa Puemi a mis au point les conditions d'expression et de purification de SECp43.

Dr Sygusch a participé à la conception et à la supervision des expériences de SAXS. Il a également participé à l'analyse des données de SAXS et à la révision du manuscrit.

Toutes les expériences ont été faites sous la supervision du Dr Baron qui a participé à la conception des projets, à l'analyse des résultats, à la rédaction du manuscrit et au processus de révision.

L'article a été publié dans la revue *Biochemistry* en Avril 2017.

Article 1

Analysis of novel interactions between components of the selenocysteine biosynthesis pathway SEPHS1, SEPHS2, SEPSECS and SECp43

Flore Oudouhou, Bastien Casu, Arnold Steve Dopgwa Puemi, Jurgen Sygusch and Christian Baron

Department of Biochemistry and Molecular Medicine, Université de Montreal, Faculty of Medicine, Montreal, QC, Canada

Keywords: Selenocysteine, selenophosphate synthetase, SECp43, bioluminescence resonance energy transfer (BRET), small angle X-ray scattering (SAXS)

Biochemistry. 2017 May 2;56(17):2261-2270. doi: 10.1021/acs.biochem.6b01116. Epub 2017 Apr 20.

II.1.2 Abbreviations

BRET, bioluminescence resonance energy transfer; BSA, bovine serum albumin; DMEM, Dulbeccos modified eagle medium; DSS, disuccinimidyl suberate; Fw, forward; GFP, green fluorescent protein; HEK, human embryonic kidney; IP, immunoprecipitation; IPTG, isopropyl β -D-1-thiogalactopyranoside; kPsi, kilopounds per square inch; LB, lysogeny broth (Luria-Bertani); mAU, milliabsorbance units; Ni-NTA, nickel nitrilotriacetic acid; NMR, nuclear magnetic resonance; PBS, phosphate buffered saline; PCR, polymerase chain reaction; Rg, radius of gyration; RLuc, *Renilla* luciferase; RRM, RNA recognition motif; Rv, reverse; SAXS, small angle X-ray scattering; Sec, selenocysteine; SECIS, selenocysteine insertion sequence; SeP, monoselenophosphate; SEPHS; selenophosphate synthetase; SEPSECS, selenocysteine synthase; TEV, tobacco etch virus; WAXS, wide angle X-ray scattering WB, western blot.

II.1.3 Abstract

In mammalian cells, the incorporation of the 21st amino acid selenocysteine into proteins is guided by the Sec machinery. The function of this protein complex requires several protein-protein and protein-RNA interactions, leading to the incorporation of selenocysteine at UGA codons. It is guided by stemloop structures localized in the 3' untranslated regions of the selenoprotein-encoding genes. Here, we conducted a global analysis of interactions between the Sec biosynthesis and incorporation components using a bioluminescence resonance energy transfer assay in mammalian cells that showed that selenocysteine synthase (SEPSECS), SECp43, and selenophosphate synthetases SEPHS1 and SEPHS2 form oligomers in eukaryotic cells. We also showed that SEPHS2 interacts with SEPSECS and SEPHS1; these interactions were confirmed by co-immunoprecipitation. To further analyze the interactions of SECp43, the protein was expressed in *Escherichia coli* and small angle X-ray scattering analysis revealed that it is a globular protein comprising two RNA-binding domains. Using phage display, we identified potential interaction sites and highlighted two residues (K166 and R169) required for its dimerization. The SECp43 structural model presented here constitutes the basis of future exploration of the protein-protein interactions among early components of the selenocysteine biosynthesis and incorporation pathway.

II.1.4 Introduction

Selenium is an essential trace element for human health. It is co-translationally incorporated into selenoproteins as 21st amino acid, selenocysteine (Sec) [1]. This process requires the re-definition of the UGA codon from a stop codon to a codon incorporating Sec. Sec biosynthesis and incorporation are achieved via unique machinery that depends on several trans- and cis-acting factors and on a series of protein-protein and protein-RNA interactions [2]. One of the many remarkable aspects of Sec biosynthesis is that it is the only known amino acid in eukaryotes whose biosynthesis occurs on a specialized tRNA, the tRNA^{[Ser]Sec} [3, 4]. Sec biosynthesis initiates with the attachment of serine to the tRNA^{[Ser]Sec} by the conventional seryl-tRNA synthetase to yield Ser-tRNA^{[Ser](Sec)} [4, 5]. The seryl residue is then phosphorylated by phosphoseryl-tRNA kinase [6]. O-phosphoseryl-tRNA^{[Ser]Sec} is a substrate for selenocysteine synthase (SEPSECS), which accepts the active selenium donor monoselenophosphate (SeP) generated by selenophosphate synthetase SEPHS2 (a Sec-carrying selenoprotein itself), and converts the phosphoryl moiety to yield selenocysteyl-tRNA^{[Ser]Sec} [7]. Sec co-translational incorporation at UGA codons requires the presence of the Sec insertion sequence (SECIS) element, a stem-loop RNA structure present in the 3'untranslated region of eukaryotic and archaeal selenoprotein mRNAs [8-10]. The SECIS element has been shown to be a platform that recruits factors involved in Sec incorporation. The SECIS binding protein 2 (SBP2), which also interacts with the ribosome, binds the SECIS element with high affinity and specificity [11-13]. SBP2 also recruits the specific eukaryotic elongation factor eEFSec to selenoprotein mRNAs [10]. Together those factors allow an efficient recoding of UGA to Sec. The ribosomal protein L30 (RPL30) is the second SECIS-binding protein that has been characterized. RPL30 and SBP2 bind to the SECIS element at overlapping sites, and might actually compete for SECIS binding, followed by Sec incorporation into the growing polypeptide chain [14]. SECp43 is another factor involved in Sec biosynthesis and incorporation, but its exact function is currently unknown. SECp43 precipitates Sec-tRNA^{[Ser]Sec} from mammalian cell lysates [15] and interacts with a 48-kDa protein in HeLa cell extracts, corresponding to the size of SEPSECS [7, 16]. SECp43 may be required for methylation of Sec-tRNA^{[Ser]Sec} prior to the synthesis of some stress-related selenoproteins [17]. SECp43 may also play a role in promoting Sec machinery

complex assembly, notably by mediating the association between eEFSec and SBP2 as well as SEPSECS and SEPHS1 [18].

Interestingly, in eukaryotes, two selenophosphate synthetases exist (SEPHS1 and SEPHS2) that have significant sequence similarity with the bacterial selenophosphate synthetase SelD [19-21]. However, the role of SEPHS1 is currently unknown. Previous work showed that only the SEPHS2Cys variant of SEPHS2 (Sec changed to Cys at position 60) complemented SelD deficiency in *Escherichia coli* [22] and SEPHS2 could generate SeP, the active selenium donor, *in vitro* [7]. In contrast, complementation of the *selD* mutant by SEPHS1 occurred only when *E. coli* was cultivated in medium supplemented with selenite [23]. Furthermore, knockdown of SEPHS1 in NIH3T3 cells had no effect on selenoprotein synthesis whereas knockdown of SEPHS2 disrupted it [7]. These results suggest that SEPHS2 is responsible for *de novo* synthesis of SeP, while SEPHS1 may have a role in recycling Sec via a selenium salvage system.

Although most of the key factors of the Sec machinery have been identified, several mechanistic aspects of Sec biosynthesis and incorporation remain unclear. Also, the sequence and relative strengths of interactions could not necessarily be identified using approaches such as co-immunoprecipitation. We here analyzed the interactions between components of the Sec biosynthesis pathway using an *in vivo* interaction assay in mammalian cells. We confirmed many previously known interactions and report in addition that components from the early steps of the Sec biosynthesis pathway interact (SEPSECS and SEPHS2 as well as SEPHS1 and SEPHS2). We also showed that SECp43, which interacts with SEPSECS and SEPHS1, is a globular protein that forms oligomers *in vivo*.

II.1.5 Materials and Methods

II.1.5.1 Construction of expression vectors

The origins of the DNA constructs are shown in Supplementary Table SII.1.1. For the BRET experiments, proteins of interest were fused with genes encoding either the donor (*RLuc2*) or the acceptor (*GFP2*) in *N*- and *C*-terminus. The DNA coding sequences were amplified by PCR (primers shown in Supplementary Table SII.1.2) and were subcloned into pcDNA3.1 vector using NotI and ClaI for the constructs carrying the reporter in *C*-terminus, or BspEI and XbaI

for those with the reporter in *N*-terminus. In the case of EFSec, cDNA was subcloned into pcDNA3.1 vector using BamHI and NotI or XhoI and XbaI restriction enzymes for the constructs with the reporter in *C*-terminus or in *N*-terminus, respectively. To create RLuc2- and GFP2- tagged proteins, PCR fragments encoding full-length *Renilla* luciferase or GFP2 sequences from the pRLuc2 (C1 or N2) or pGFP2 (C1 or N2) vectors were subcloned into the pcDNA3.1 vectors encoding the proteins of interest. All constructs were validated by DNA sequencing.

II.1.5.2 Construction of expression plasmids for protein variants

The Sec TGA codon in the gene encoding SEPHS2 was mutated to a Cys TGT codon using PCR-based mutagenesis. To generate the SECp43 variants (L163A, G164A, S165A, K166A, P167A, V168A and R169A), each residue of the LGSKPVR peptide was substituted for Ala by proceeding to a site directed mutagenesis on SECp43 encoding sequence. Mutagenesis were performed using the KOD Hot Start DNA polymerase kit (Novagen). Primers are shown in Supplementary Table SII.1.3. All constructs were validated by sequencing.

II.1.5.3 Cell culture and transfection

Human embryonic kidney 293T cells (HEK293T) were maintained in Dulbecco's Modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (Wisent Inc.) in a 5% CO₂ humidified incubator at 37 °C. Cells were seeded at $2,5 \times 10^5$ cells/well density in 6-well plates for BRET experiments and 3×10^6 cells per 100-mm dish for co-immunoprecipitation assay. Transient transfections were performed on the following day with a constant amount of transfected DNA (1 µg/well total plasmid for 6-well plates and 10 µg for 100-mm dishes), using FuGENE® 6 Transfection Reagent (3 µl/ µg of DNA) (Roche).

II.1.5.4 BRET assay

For BRET saturation assays, a constant amount of RLuc2 construct (50 ng for interaction studies and 10 ng for SECp43 dimerization analysis) was transfected alone or with increasing quantities of GFP2 construct (from 0 to 1.5 µg) in 6-well plates. The total amount of DNA transfected in each well was completed to 1.5 µg with pcDNA3.1 empty vector DNA. 48 h post transfection, cells were washed and resuspended in Tyrode's buffer (10 mM HEPES, 137 mM NaCl, 2.68

mM KCl, 0.42 mM NaH₂PO₄, 1.7 mM MgCl₂, 11.9 mM NaHCO₃ and 5 mM glucose), then 100 000 cells were distributed in 96-well white plates (Greiner bio-one). Measurements were performed in an Envision 2104 Multilabel Reader (Perkin Elmer). First, *h*-coelenterazine (Nanolight Technology) was added at 5 mM final concentration, and bioluminescence resulting from *Rluc*-mediated coelenterazine degradation was recorded at 410 ± 40 nm. GFP2 fluorescence was determined using an excitation filter at 410 nm and an emission filter at 515 nm filter ± 15 nm. Total fluorescence and luminescence were used as relative measures of total expression of the acceptor and donor proteins, respectively.

After the addition of coelenterazine 400a (Nanolight Technology) at 5 mM final concentration, the resonance energy transferred was recorded with the Envision Reader, allowing the sequential integration of the signals detected in the 410 ± 40 nm and 515 ± 15 nm windows for luciferase and GFP light emissions respectively. The BRET signal was determined by calculating the ratio of the light intensity emitted by the receptor-GFP over the light intensity emitted by the receptor-*RLuc*. The values were corrected by subtracting the background BRET signal detected when the receptor-*RLuc* construct was expressed alone.

II.1.5.5 Co-immunoprecipitation

The co-immunoprecipitation assays were performed using the Pierce™ Co-Immunoprecipitation Kit, following the procedure provided by the manufacturer (Pierce Biotechnology). 50 µl of the resin was added to a Pierce Spin Column and washed twice with the coupling buffer. Then 10-75 µg of affinity purified antibody (anti-*Renilla* luciferase antibody, MAB4400 – Millipore; anti-GFP - Roche) were prepared in coupling buffer and added to the column. Sodium cyanoborohydride solution was added and the resin was incubated on a rotator at room temperature for 120 min. The resin was washed twice with coupling buffer and quenching buffer was added. A new incubation for 15 min was performed after adding sodium cyanoborohydride solution. The antibody-coupled resin was washed twice with coupling buffer and six times with the wash solution. Mammalian cells were then lysed following the lysis protocol for adherent cells. They were first washed in Modified Dulbecco's PBS, then incubated in ice-cold IP lysis/wash buffer on ice for 5 minutes with periodic mixing. Protein concentration

of the lysates was determined and 1 mg of lysate was pre-cleared with the control agarose resin for 60 min.

Once the antibody-coupled resins were washed with IP lysis/wash buffer, pre-cleared sample and controls were added to the appropriate resin. The co-IP assays were performed overnight at 4°C. After the incubation, the resins were washed three times in IP lysis/wash buffer. The elution was done in two steps using the elution buffer and eluates were analyzed by SDS-PAGE and Western blot using anti-*Renilla* luciferase and anti-GFP antibodies.

II.1.5.6 Cloning, expression and purification of SECp43¹⁸⁰

The DNA encoding full-length SECp43 was amplified from pUC19-SECp43 (D. Su and D. Soll, Yale University) and subcloned into the PstI/HindIII sites of pETDuet-1 vector (Novagen), containing a His-tag coding sequence; a tobacco etch virus (TEV) protease cleavage sequence was added by PCR. The pETDuet-1-SECp43¹⁸⁰ construct contains a stop codon inserted by directed mutagenesis after amino acid 180. pETDuet-1-SECp43¹⁸⁰ and pRARE (Novagen), required to express human tRNAs for codons not well translated in *E. coli*, were co-transformed into BL21(DE3) cells and grown under selection in ampicillin (100 µg/mL).

Protein expression was conducted by auto-induction at 25°C for 72 h, in 8ZYM- 4x LAC media [24]. Cells were harvested by centrifugation, resuspended in binding buffer (20 mM NaHPO₄ pH 7.4, 500 mM NaCl, 40 mM imidazole) (7.5 mL of binding buffer per 1 g of pellet), supplemented with Complete Protease Inhibitor Tablet (Roche) and DNase (1 mg per 1 g of pellet). Cells were disrupted in French Press at 27 kPsi and lysates clarified by centrifugation at 38,500 g for 30 minutes (4°C) followed by filtration of the supernatant through a 0.45 µm membrane. The supernatant was loaded onto a Ni-NTA column (HisTrap TM HP, GE Healthcare). Elution of the bound fraction was performed using a linear gradient of elution buffer (20 mM NaP pH 7.4, 500 mM NaCl, 500 mM imidazole) followed by TEV protease cleavage of the N-terminal His-tag at 20°C for 16 h as previously described [25]. The buffer of the reaction mixture was dialyzed against the binding buffer (50 mM NaP pH 7.4, 300 mM NaCl, 40 mM imidazole) and loaded again onto a Ni-NTA column. The flow through fractions were then concentrated to 5 ml using a Vivaspin 6 dialysis membrane (10 kDa MWCO, GE Healthcare) and further purified by size-exclusion chromatography (Superdex TM 75 10/300

GL; GE Healthcare), previously equilibrated against a mobile phase buffer (20 mM Tris pH 7.5, 200 mM NaCl). Fractions containing SECp43¹⁸⁰ were pooled.

II.1.5.7 Analysis of SECp43 dimerization by crosslinking

Chemical crosslinking with disuccinimidyl suberate (DSS) was performed as described previously [26]. Briefly, purified His₆-tagged SECp43 at 350 ng/μl were mixed with different concentrations (0 to 200 μM) of the cross-linking agent disuccinimidyl suberate (10 mM stock in Me₂SO, Pierce Biotechnology). The samples were incubated for 1 h at 22°C, followed by analysis by SDS-PAGE and western blot, using anti-His antibody (Abgent).

II.1.5.8 Small Angle X-ray Scattering data collection and analysis

Solutions of SECp43¹⁸⁰ were prepared in 200 mM NaCl, 20 mM Tris, pH 7.5 at protein concentrations of 8.5, 5.0, 2.5, and 1.0 mg/ml, respectively. Scattering data were measured at beamline F2 of the Biological Small-Angle X-Ray Solution Scattering (BioSAXS) at Macromolecular Diffraction at the Cornell High Energy Synchrotron Source (MacCHESS), Ithaca, USA. Data were collected on a dual Pilatus 100K-S SAXS/WAXS detector at a wavelength of 1.3 Å, covering a scattering vector range ($q = 4\pi\sin\theta/\lambda$) from 0.008 to 0.8 Å⁻¹. Twenty frames were recorded for each concentration, with an exposure time of 2 s for each frame, with buffer scans collected before each experiment. Data were analyzed, integrated and averaged with RAW software [27]. Buffer blanks were averaged and subtracted from the data. Each of the twenty exposures was inspected visually in RAW software [27] and only the initial data frames showing no evidence for radiation damage or aggregation in terms of individual linear Guinier fit plots were averaged using the RAW software [27]. Guinier analysis and radius of gyration (R_g) estimation were performed in Primus [28] and confirmed by automatic analysis using AutoRG [28]. The largest dimension of the molecule, D_{max} , and the Porod volume were calculated using GNOM [29]. The pair distribution function $P(R)$ and forward scattering $I(0)$ were computed with AutoGNOM [30] and compared with those determined in Primusqt [28]. To merge all the SAXS data sets together, for the four concentrations of SECp43¹⁸⁰ used, a command-line program named ALMERGE was used [31]. For the merging of the data collected at different SECp43¹⁸⁰ concentration, an angular range of $q_{min} = 0.0256 \text{ \AA}^{-1}$ to $q_{max} = 0.2965 \text{ \AA}^{-1}$ was chosen by AutoGNOM [30]. GASBOR [32] was used to reconstruct an *ab initio* protein

structure by a chain-like ensemble of dummy residues. Rigid body modeling was performed with CORAL [33] using the NMR solution structures of RRM1 (PDB accession code: 2DIV) and RRM2 (PDB accession code: 2DHG) domains of the TRNA selenocysteine associated protein. Residues present in the NMR structure but not present in our expression constructs were deleted for consistency and the 7 amino acid linker (residues 87-93) was modeled. Data were cut off at $q_{\max} = 0.38 \text{ \AA}^{-1}$. Final models were docked in the *ab initio* shapes using SUPCOMB [34]. Structural figures were prepared using PyMOL (*The PyMOL Molecular Graphics System, Version 1.3r1*, Schrodinger, LLC).

II.1.5.9 Phage display

The phage display experiments were realized following the Ph.D.TM Phage Display Libraries Instruction Manuel (NEB) as described [35, 36]. The microtiter plate (Thermo Scientific) was coated with 100 $\mu\text{g}/\text{mL}$ of purified SECp43¹⁸⁰ in 0.1 M NaHCO₃ pH 8.6 at 4 °C for 16 h, followed by incubation for 2 h at 4 °C with blocking buffer (0.1 M NaHCO₃, pH 8.6; 5 mg/ml BSA; 0.02% NaN₃). Phage libraries (Ph.D.-12 or Ph.D.-C7C) were diluted to 4×10^{10} pfu/ml from the original library (New England Biolabs, NEB) and added to the plate. Non-binding phages were discarded by 10-times washing with TBS-T (TBS + 0.1% [v/v] Tween-80). SECp43¹⁸⁰-binding phages were eluted using 0.2 M glycine-HCl (pH 2.2) in 1 mg/mL BSA, and neutralized with 1 M Tris-HCl (pH 9.1). The eluate phages were tittered onto LB/IPTG/Xgal plates, amplified and precipitated overnight at 4°C using 1/6 volume of PEG/NaCl (PEG-8000, 2.5 M NaCl). The concentration of the phages was determined and the process was repeated for two additional panning rounds using the obtained eluate. The final eluate was plated onto a LB/IPTG/Xgal plate to yield approximately 100 individual clones. Twenty clones from the plate were chosen at random and the displayed sequence was submitted to DNA sequencing with the -96 gIII sequencing primer (5-CCCTCATAGTTATTAGCGTAACG-3). Unique peptides were aligned to SECp43 sequence using MatchScan program (P.D. Pawelek, Concordia University, Montreal) that allows calculation of peptide similarity using a wide range of residue windows and scores [37].

II.1.6 Results

II.1.6.1 Identification of interactions between components of the Sec machinery via BRET analysis

We conducted a global analysis of interactions using quantitative BRET (bioluminescence resonance energy transfer) analysis in HEK293T tissue culture cells. The Sec biosynthesis and incorporation components (SEPHS1, SEPHS2, SEPSECS, SECp43, eEFSec, RPL30 and SBP2) were expressed as fusions to the luminescence donor protein *RLuc2* (*Renilla* luciferase) and to the acceptor protein GFP2 (a variant of the green fluorescent protein GFP). Since SEPHS2 is a selenoprotein we used a variant containing Cys instead of Sec (Se60C) to avoid possible limitations of its expression, due to limited capacity of the Sec biosynthesis and incorporation machinery in tissue culture cells [21]. Measurement of luminescence after addition of the luciferase substrate (coelenterazine) and GFP fluorescence signals showed that the fusion proteins were expressed in HEK293T cells, with the exception of eEFSec and SBP2 that were not further pursued (data not shown). For the BRET saturation assay, cells were transfected with a constant amount of *RLuc2* fusion-encoding expression plasmids and increasing amounts of plasmid encoding potential interaction partners fused to GFP2. After the addition of coelenterazine, luminescence emitted by the *RLuc2* donor resulted in the excitation of the GFP2 acceptor, leading to an emission of fluorescence, only if the two molecules were in close proximity (1-10 nm) [38, 39]. Then the BRET signal (Y-axis), corresponding to the ratio between the fluorescence signal of the GFP2-fusion protein (acceptor) on the luminescence of the *RLuc2*-fusion protein (donor) emitted after addition of the *RLuc* substrate, were plotted as a function of the [receptor-GFP2]/[donor-*RLuc2*] ratio. In this titration assay, as the amount of BRET acceptor increases, the BRET signal increases as a hyperbolic function until reaching a saturation, indicating a specific interaction. At the opposite, a linear relationship indicates a non-specific interaction. To validate our BRET system we first tested the interaction between two subunits (G β 1 and G γ 1) of the heterotrimeric guanine nucleotide-binding protein (G-protein) which was extensively characterized [40-42] (Supplementary Fig. SII.1.1 A). To quantify the relative affinity of the protein-protein interaction, BRET signals (Y-axis) were also plotted as a function of the relative proteins expression (X-axis; here defined as intensity ratio), by measuring the total GFP2 fluorescence and *RLuc2* luminescence in absence of resonance

energy transfer, before addition of the *RLuc* substrate (Supplementary Fig. SII.1.1 B). While the saturation level ($BRET_{max}$) corresponds to the saturation of all BRET donor molecules, the $BRET_{50}$ represents the amount of acceptor required to reach the half of $BRET_{max}$ for a given amount of donor and reflects the relative affinity of the interaction [38, 43-46]. These analyses revealed that several components of the early steps of the selenocysteine biosynthesis pathway (SEPHS1, SEPHS2, SEPSECS and SECp43) form multimers (Fig. II.1.1 A-D; Table II.1.1). In contrast, multimerization was not observed for RPL30 (Fig. II.1.1 E).

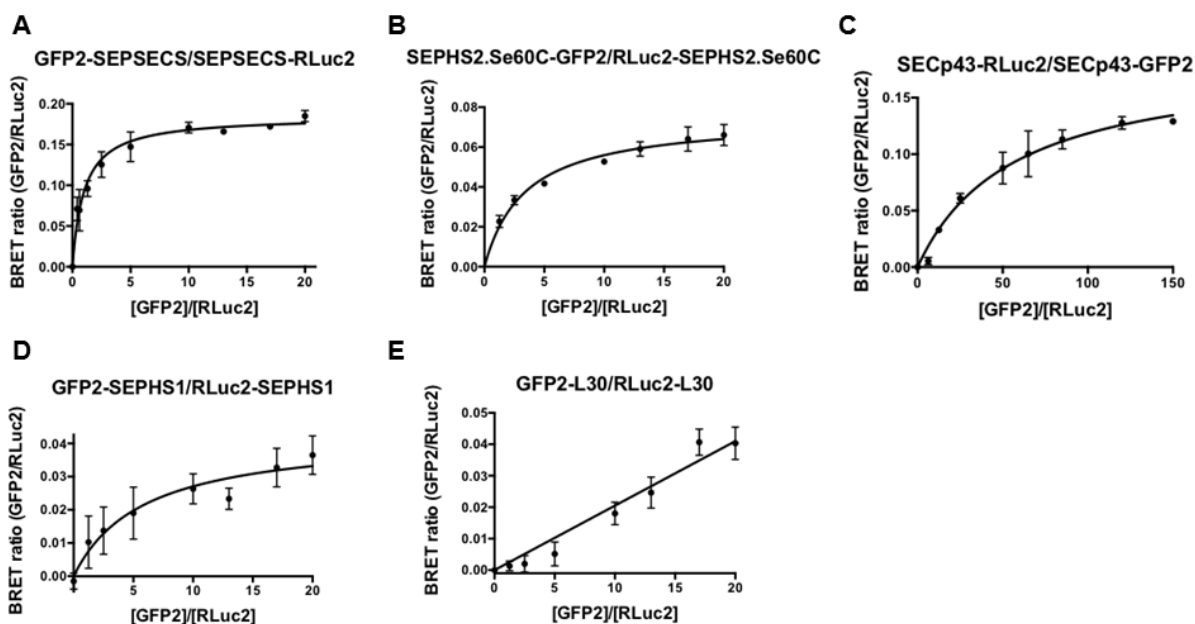


Figure II.1.1 Oligomerization profiles of Sec machinery components measured by BRET.

Components of the Sec machinery were expressed as fusions to *RLuc2* or GFP2 reporters. BRET titration experiments were performed with a fixed amount of *RLuc2* donor construct and increasing amounts of GFP2 acceptor constructs and expressed in HEK293T cells. BRET ratios (515 nm/410 nm) (Y-axis), measured after addition of the luciferase substrate, were plotted against the [receptor-GFP2]/[donor-*RLuc2*] ratio (X-axis). Results were analysed and fitted using non-linear regression assuming a model with one site binding (GraphPad Prism). Titration curves for A) multimerization of SEPSECS (n=4); B) multimerization of SEPHS2 (n=3); C) multimerization of SECp43 (n=3); D) multimerization of SEPHS1 (n=4); E) monomer of RPL30 (n=3). Data are means \pm SD of at least three independent titration experiments.

Table II.1.1 Protein interactions assessed by BRET and their BRET₅₀

	SEPSECS	SEPHS1	SEPHS2	SECp43	RPL30
SEPSECS	0.011 ± 0.005				
SEPHS1	+/-	0.039 ± 0.028			
SEPHS2	0.215 ± 0.127	0.018 ± 0.005	0.017 ± 0.002		
SECp43	0.029 ± 0.011	0.583 ± 0.072	-	0.159 ± 0.024	
RPL30	-	-	-	-	-

+/- : Ambiguous profile of interaction

- : No interaction measured

Next, we tested pairwise interactions between all these proteins. The saturation curves and BRET₅₀ values indicated specific interactions between several components: SEPSECS and SEPHS2 (Fig. II.1.2 A), SEPHS1 and SEPHS2 (Fig. II.1.2 B), as well as for SECp43 and SEPSECS and SEPHS1 (Table II.1.1). Specific interactions were not detected in the case of all the other combinations that were tested (Fig. II.1.2 C; Table II.1.1 and Supplementary Fig. SII.1.2). Consistent with our results, some of those interactions were shown in previous studies [10, 16, 18, 47-50]. Also, while these interactions mainly occur into the cytoplasm, no interaction was observed with RPL30, an abundant cytoplasmic protein, suggesting the specificity of our BRET experiments.

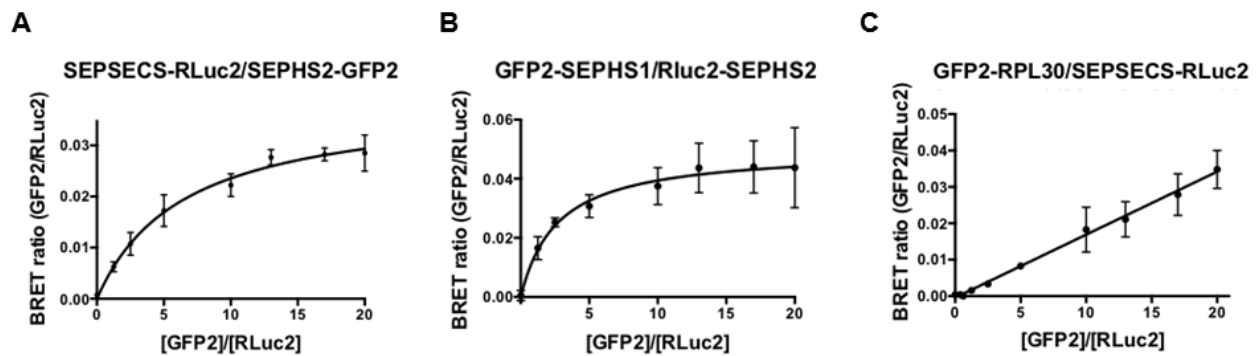


Figure II.1.2. Interaction analyzes of Sec machinery components by BRET. Experiments were performed and analyzed as described in Fig. II.1.1. Titration curves for A) SEPSECS-SEPHS2 interaction; B) SEPHS1-SEPHS2 interaction; C) RPL30-SEPSECS interaction. Data are means ± SD of four independent titration experiments.

II.1.6.2 Co-immunoprecipitation confirms the interactions between SEPHS1, SEPHS2 and SEPSECS

Some of the observed interactions were known from previous work [10, 16, 18, 47-50], but the SEPHS1-SEPHS2 and SEPSECS-SEPHS2 interactions were not described before. Thus, we next examined them by co-immunoprecipitation. To this effect, the SEPHS1, SEPHS2 and SEPSECS expression constructs were transfected into HEK293T cells, followed by cell lysis and pull-down with GFP2- and *RLuc2*-specific antibodies. This approach showed that in cells expressing SEPSECS-*RLuc2* or SEPHS2-GFP2 the antibodies pulled down proteins of the expected size, and there were no non-specific cross-reactions between antibodies (Fig. II.1.3 A, left panel). When SEPSECS-*RLuc2* and SEPHS2-GFP2 were coexpressed, the *RLuc2*-specific antibody co-precipitated SEPSECS-*RLuc2* and SEPHS2-GFP2 (Fig. II.1.3 A, right panel), thus confirming the SEPSECS-SEPHS2 interaction. Similarly, in SEPHS2-*RLuc2* or SEPHS1-GFP2 expressing cells, the antibodies pulled down proteins of the expected size and the antisera did not detect non-specific antibodies cross-reactions (Fig. II.1.3 B, left panel). Interestingly, several bands were observed when SEPHS1-GFP2 were pull-down that may correspond to three of the splice variants of SEPHS1 protein [51]. When SEPHS2-*RLuc2* and SEPHS1-GFP2 were co-expressed, the GFP2-specific antibody co-precipitated SEPHS2-*RLuc2* and SEPHS1-GFP2 (Fig. II.1.3 B, right panel). Similar results were observed when the immunoprecipitation was performed with the *RLuc2*-specific antibody which co-precipitated SEPHS2-*RLuc2* and SEPHS1-GFP2 (Fig. II.1.3 B, right panel), confirming the SEPHS1-SEPHS2 interaction. Surprisingly, in this case, only one band was observed for SEPHS1-GFP2, that may indicate that only one splice variant of SEPHS1 co-eluted with SEPHS2.

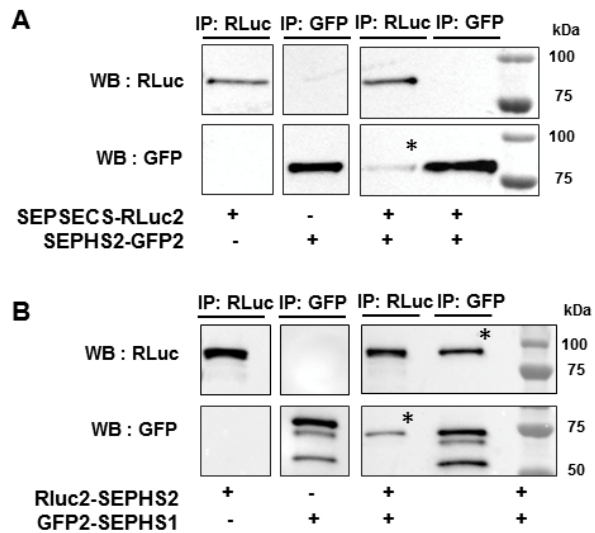


Figure II.1.3. Analysis of SEPSECS, SEPHS1 and SEPHS2 interactions by co-immunoprecipitation. A) SEPSECS-RLuc2 (92 kDa) and SEPHS2-GFP2 (74 kDa) were expressed alone or co-expressed in HEK293T cells. Immunoprecipitations (IP) were performed with anti-GFP or anti-Rluc antibodies, separated by SDS-PAGE and analyzed by western blot (WB) with anti-GFP or anti-Rluc antibodies. B) Rluc2-SEPHS2 (83 kDa) and GFP2-SEPHS1 (70 kDa) were expressed alone or co-expressed in HEK293T cells. Immunoprecipitations (IP) were performed with anti-GFP or anti-Rluc antibodies, separated by SDS-PAGE and analyzed by western blot (WB) with anti-GFP or anti-Rluc antibodies. Molecular masses (MW) of reference proteins are shown on the right. Asterisk (*) indicate the bands that show significant difference compared with the co-IP control.

II.1.6.3 Structural analysis of the N-terminal part of SECp43 by SAXS

The role of SECp43 in the Sec biosynthesis pathway is currently unknown. We and others showed that it interacts with several other Sec machinery components [7, 15-18]. The interactions of this protein are likely important for selenocysteine metabolism. Therefore, we further characterized its structure and interactions. An N-terminally His-tagged version of full-length SECp43 was overexpressed and purified from *E. coli*, but the size of the purified protein was smaller than expected. Mass spectrometric analysis revealed that the protein was truncated and the C-terminal region, predicted to be unstructured (analysis with RONN [52]), was proteolytically removed in the bacteria (data not shown). Based on this information, we

introduced a stop codon after codon 180 terminating the protein before the unstructured region and the resulting SECp43¹⁸⁰ protein was stably expressed and purified from *E. coli*. Since our efforts to crystallize the protein were not successful, we performed analysis by small and wide angle X-ray scattering (SAXS-WAXS) to obtain insights into its overall structure. This analysis led to a model for a globular protein (Fig. II.1.4 A and Table II.1.2).

Table II.1.2. SAXS data collection and scattering-derived parameters

Sample	SECp43
<i>Data collection parameters</i>	
Beamline	F2
Beam geometry (μm^2)	250 x 250
Wavelength (\AA)	1.3
q range (\AA^{-1})	0.008 - 0.8
Exposure time per frame (s) ^a	2
Concentration range (mg/mL)	1.0 - 8.5
Temperature ($^{\circ}\text{C}$)	23
<i>Structural parameters</i>	
$I(0)$ (from Guinier) ($\pm\text{SE}$)	11.51 ± 0.05
R_g (\AA) (from Guinier) ($\pm\text{SE}$)	22.26 ± 0.09
D_{max} (\AA) [$\pm 10\%$ (estimated range)]	73 ± 7
χ^2 (SASREF)	1.083
Porod volume [from $P(r)$] (10^3\AA^3)	23.9
<i>Software employed</i>	
Primary data reduction	RAW
Data processing	PRIMUS
Modeling	CORAL
^a 20 frames were measured for each sample.	

Searches in the protein database (PDB) revealed that NMR structures had already been determined for the two RNA recognition motifs (RRM) of the protein (Fig. II.1.4 B). We fitted these domains into the SAXS-WAXS envelope and modelled the seven interdomain residues not present in the NMR structures, leading to a model of the SECp43¹⁸⁰ structure shown in Fig. II.1.4 C. The theoretical scattering profile of the rigid body model fits well to the experimental SAXS-WAXS data (Fig. II.1.4 D), with a χ^2 of 1.083, suggesting this model can serve as basis for further analyses of SECp43 protein interactions.

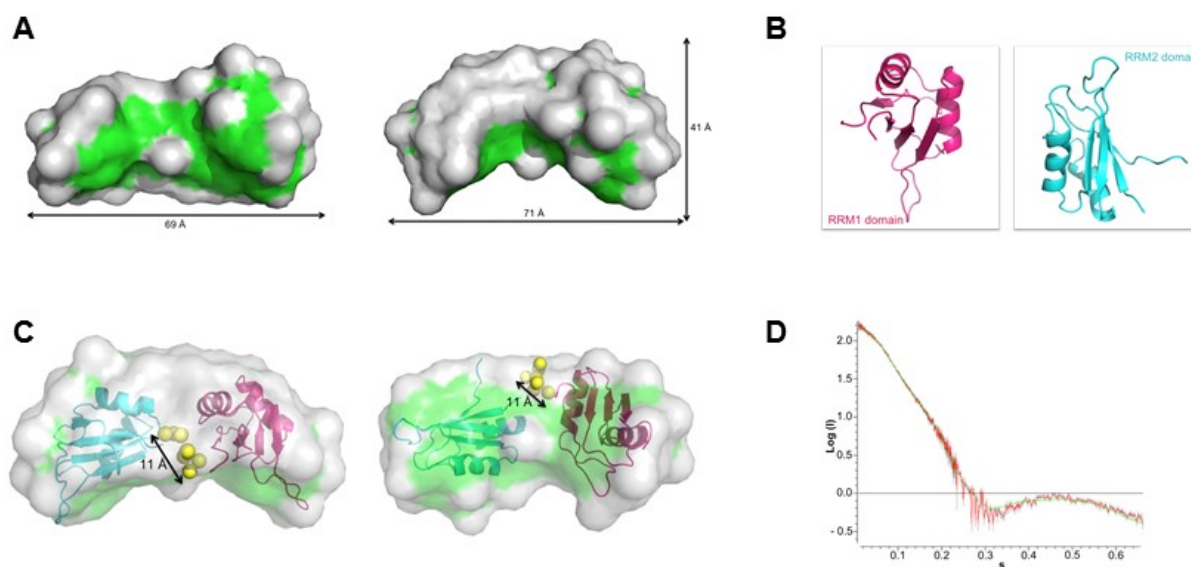


Figure II.1.4. Structural modelling of SECp43¹⁸⁰ based on SAXS-WAXS data. A) Averaged molecular envelope for SECp43¹⁸⁰. The approximate envelope dimensions (in Å) are illustrated. The two views are related by 90° rotation. B) NMR structures of the individual domains of SECp43, RRM1 (PDB: 2DIV) and RRM2 (PDB: 2DHG) used in rigid body modeling. RRM1 is coloured in magenta and RRM2 in blue. C) CORAL [53] rigid body model of SECp43¹⁸⁰ (ribbons, coloured as in B) with the best statistical fit to the experimental data (plotted in D); overlaid is the averaged molecular envelope. Seven inter-domain dummy residues introduced between the C-terminus of RRM1 and the N-terminus of RRM2, in the optimal conformation chosen by CORAL, are depicted as yellow spheres. The distance between the two subunits is 11 Å. D) Fit of the theoretical scattering profile for the rigid body model (green line) with the experimental SAXS-WAXS data (red line) (logarithmic).

II.1.6.4 Identification of a peptide required for the dimerization of SECp43¹⁸⁰

BRET analysis showed that SECp43 may form dimers or higher molecular mass multimers *in vivo*. These results were confirmed by size-exclusion chromatography and crosslinking experiments that both showed that purified SECp43 exists in monomeric and dimeric forms (Supplementary Fig. SII.1.3). We next conducted phage display analysis in order to identify the potential dimerization site of the protein. Purified SECp43¹⁸⁰ was used as bait for the phage display experiments. Two commercially available libraries displaying random peptides on the surface of phage M13 (Ph.D.-12 and Ph.D.-C7C) were tested. We conducted three binding and

amplification rounds in order to specifically enrich binding peptides exposed to the phage surface. Sequencing of 118 individual phages allowed us to identify 30 different peptides (Fig. II.1.5 A). Pairwise alignments of the peptides to the amino acid sequence of SECp43¹⁸⁰ revealed a consensus sequence between residues 163 and 169 (LGSKPVR), having a high likelihood of being part of an interacting region (Fig. II.1.5 B and C).

A		B	
PhD.C7C Library	PhD.12 Library		
HMQGRST	DDGAYTHRSNLI	(158) QGAV LGSKP VRLSVAIPKASRVK (182)	
HGGQTVA	GIWYRDVVQQWP	VPSKP GLG	
NAGHLSQ	SPGNQTYSSQVR	SPGNTY SSQVR YK PVRVH	
TTQMGYS	GHGMLILSPNPT	YK PVRVH	
PEWFRWH	GSMAGPENHRAI	SL FSKNY SMAGPENHRAI	
NSSASKN	EPAILTDAMPFN		
GWKHEQT	YAPDLLPNSTWT		
VPSKPGL	SHVDSASLRYWR		
NSHRHGA	SGVYKVAYDWQH		
WWNNFKH	GLHTSATNLYLH		
FPTGTYW	VHWDFRQWWQPS		
ALHSGQK			
TNASQSY			
NWMINKE			
RDLSLHS			
RTXNGTR			
FSAVVEP			
YKPVRVH			
SLFSKNY			

C		
Amino acid position	Residue	MATCH score
163	L	31.45
164	G	41.92
165	S	53.83
166	K	51.37
167	P	43.38
168	V	33.42
169	R	27.12

Figure II.1.5. Phage display to identify the dimerization site of SECp43. A) Peptides identified by phage display using the PhD-C7C and PhD-12 libraries. Phage display experiments were performed on purified SECp43¹⁸⁰. B) Potential dimerization site on SECp43 determined by alignment of the 30 peptides. Aligning amino acids are shown in bold print. The potential dimerization site is identified by asterisks. C) MATCH score for indicating probability to be part of the interacting peptide for each residue [37].

To assess whether these residues contribute to SECp43 dimerization, we conducted site-directed mutagenesis to individually change them to Ala. Quantitative BRET analysis performed on the variants revealed that changes K166A and P167A had strong negative effects on the dimerization, as shown by linear titration curves (Fig. II.1.6A). The other changes (L163A, G164A, S165A, V168A and R169A) had varying effects on dimerization. Many of the variant showed a decrease in the BRET_{max}, associated to a shift toward higher BRET₅₀ value that indicate a diminution in the affinity. In contrast, the profile saturation curve observed in the case of SECp43^{S165A} was very similar to the wild type (Fig. II.1.6B and Supplementary Table SII.1.4). These results suggest that we identified a region of the SECp43 protein involved in its

dimerization. This site is predicted to localize on the surface of the RRM2 domain of our SECp43 model (Supplementary Fig. SII.1.4).

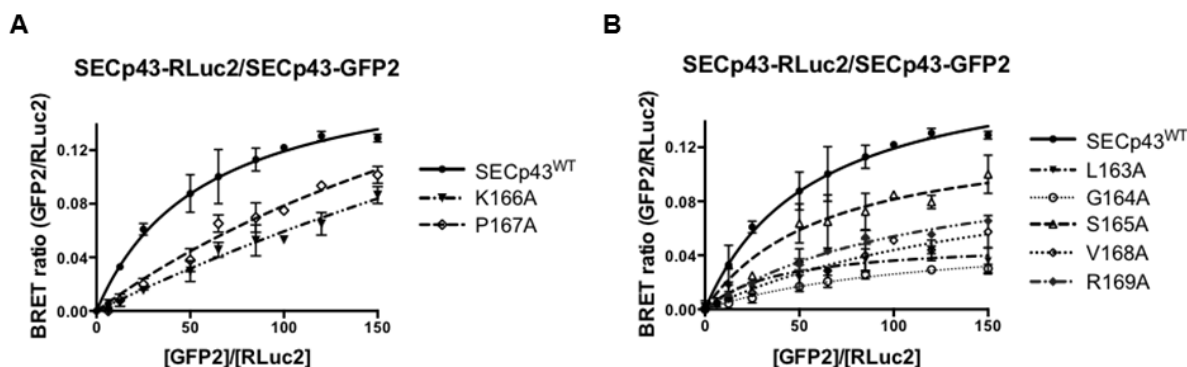


Figure II.1.6. BRET analysis of SECp43 variants. Experiments were performed and analyzed as described in Fig. 2. A) Substitution of K166 and P167 residues disrupts SECp43 dimerization. B) Dimerization profiles of the other SECp43 variants. Data are means \pm SD of three independent titration experiments.

The 30 identified peptides (Fig. II.1.5A) were also aligned on the amino acid sequence of SEPSECS leading to the identification of a region, between residues 120 and 125 (GVHTVA), that may correspond to an interacting site between SECp43 and SEPSECS (Supplementary Fig. SII.1.5).

II.1.7 Discussion

Here, we conducted a global analysis of protein-protein interactions between components of the Sec biosynthesis and incorporation components in mammalian cells. Using quantitative BRET titration analysis, we confirmed several interactions that had already been shown using other approaches [10, 16, 18, 47-50], thus validating our experimental system. The novel results of our analysis suggest that several components of the Sec machinery form multimers (SEPHS1, SEPHS2, SEPSECS and SECp43). We also identified novel interactions between several components of the early steps of Sec biosynthesis (SEPHS1-SEPHS2 and SEPSECS-SEPHS2). In agreement with the existence of a subcomplex of the Sec machinery, we monitored interactions between SEPHS1, SEPHS2, SEPSECS and SECp43, all involved in the early stages of the pathway. Also, the interaction shown between SEPSECS and SEPHS2 is consistent with

current thinking given that, in eukaryotes, SEPHS2 produces the selenium donor SeP used by SEPSECS to synthesize the Sec-tRNA^{[Ser]Sec}. In contrast, SEPHS1 was demonstrated to interact with SEPSECS *in vitro* and *in vivo* [18], but knockdown of SEPHS1 in NIH3T3 cells had no effect on selenoprotein synthesis [7]. Thus, SEPHS1 has been proposed to be responsible for Sec recycling by a selenium salvage system [23] or to have a more specialized function in selenium metabolism which does not involve SeP synthesis [54]. Five alternative splice variants of SEPHS1 have been identified in humans with unique subcellular location and expression patterns [51]. Interestingly, here we showed that SEPHS2 seems to interact with only one of the SEPHS1 splice variants, suggesting that this particular variant may play a role in delivering selenium to SEPHS2 or that it may have a structural role for assembly of the Sec biosynthesis complex.

SECp43 is another Sec machinery component whose function is unknown. The fact it undergoes multiple protein-protein interactions (*e.g.* with SEPSECS, eEFSec, SBP2, SEPHS1, tRNA^{[Ser]Sec}) that it plays a central role for Sec metabolism. We conducted SAXS-WAXS analysis of the *N*-terminal domain (SECp43¹⁸⁰) and the resulting model constitutes the first structural information on this protein. Using phage display followed by quantitative BRET assays, we identified a peptide that is likely involved in SECp43 dimerization and predicted to localize on the surface of the SAXS-WAXS model. Similar analyses will be conducted in future to map the interactions with other proteins that bind SECp43. As for many mammalian proteins, its interactions may be mediated by unstructured regions. Such a domain was predicted to comprise the *C*-terminus, whereas the *N*-terminus, containing two RRM domains, could readily be overexpressed and purified.

The BRET analysis presented here identified novel interactions and suggests the existence of a subcomplex comprising components of the Sec biosynthesis machinery (SEPHS1, SEPHS2, SEPSECS and SECp43). This subcomplex may be linked by SECp43 to the components mediating Sec incorporation at the ribosome (eEFSec, SBP2, RPL30). The *in vivo* analysis of protein-protein interactions by BRET assays provides novel opportunities to characterize binary interactions between proteins in a semi-quantitative fashion. It constitutes an important advantage in comparison with previous studies using qualitative approaches such as immunoprecipitation. BRET analysis also allows more complex analysis involving multiple

interaction partners. In future, we intend to conduct such studies to gain information on higher order complexes between components of the Sec biosynthesis and incorporation pathway. The quantifiable nature of the BRET assay will add information on the dynamics of protein interactions *in vivo*, enabling us to analyze transient as well as more stable structural interactions.

II.1.8 Acknowledgements

This work was supported by NSERC Discovery grants to C.B. and J.S., by the NSERC-CREATE training program on the Cellular Dynamics of Macromolecular Complexes (CDMC), the Canada Foundation for Innovation (CFI) and the Fonds de recherche du Quebec-Sante (FRQ-S). The authors would like to thank Dan Su and Dieter Dr. Soll (Yale University, New Haven, USA) for the SECp43 expression plasmid and Richard Gillian (CHESS, Cornell, USA) for assistance with the SAXS measurements. Synchrotron SAXS data were collected at MacCHESS beamline F2. CHESS is supported by the NSF & NIH/NIGMS via NSF award DMR-1332208, and the MacCHESS resource is supported by NIH/NIGMS award GM-103485.

II.1.9 Supporting information

Supplementary Table SII.1.1. Expression vectors

Plasmids	Source
pDNR-Dual-RPL30 (HscD00001761)	DNASU (Tempe, AZ)
pANT7-cGST-eEFSec (HscD00077644)	DNASU (Tempe, AZ)
pANT7-cGST-SEPHS1 (HscD00077688)	DNASU (Tempe, AZ)
pDNR-Dual-SEPHS2 (HscD00021572)	DNASU (Tempe, AZ)
pDONR201-SBP2 (HscD00081427)	DNASU (Tempe, AZ)
pOTB7-SEPHS2mRNA (HscD00322296)	Plasmid Harvard Database (Cambridge, MA)
pTT5-SECp43	D. Söll (Yale University)
pANT7-SEPSECS	D. Söll (Yale University)
pcDNA3.1-Empty Vector	S. Michnick (Université de Montréal)
pcDNA3.1-GFP10-G γ 1	S. Michnick (Université de Montréal)
pcDNA3.1-RLuc2-G β 1	S. Michnick (Université de Montréal)
pGFP2-C1	M. Aubry (Université de Montréal)
pGFP2-N2	M. Aubry (Université de Montréal)
pRLuc2-C1	M. Aubry (Université de Montréal)
pRLuc2-N2	M. Aubry (Université de Montréal)

Supplementary Table SII.1.2. Oligonucleotide primers for the construction of expression vectors

Mutation	Sequence
RLuc Fw (NotI)	5' TCAGATCCGCGGCCGCCACCATGACCAGCAAGG 3'
RLuc Rv (ClaI)	5' GCCGAGATCGATCTGCTCGTTCTTCAGCACTCTCTCC 3'
RLuc Fw (BspEI)	5' GCTTCCGGAACCATGACCAGCAAGGTGTACGACCC 3'
RLuc Fw (XhoI)	5' GCTCTCGAGACCATGACCAGCAAGGTGTACGA 3'
RLuc Rv (XbaI)	5' GATCTAGATCGAATTCTTACTGCTCGTTCTTCA 3'
GFP2 Fw (NotI)	5' ATAAGAATGCGGCCGCACCATGGTGAGCAAGGGCGA 3'
GFP2 Rv (ClaI)	5' CCATCGATCTTGTACAGCTCGTCCATGCCGAGAGTG 3'
GFP2 Fw (BspEI)	5' GTCCGGAACCATGGTGAGCAAGGGCG 3'
GFP2 Fw (XhoI)	5' CCGCTCGAGACCATGGTGAGCAAGGGC 3'

GFP2 Rv (XbaI)	5' GATCTAGAGCTTTACTTGTACAGCTCGTCCATGCC 3'
SEPSECS Fw (NotI)	5' ATAAGAATGCGGCCGCACCATGAACCGCGAGAGCTTCGC 3'
SEPSECS Rv (ClaI)	5' CCATCGATTGAAGAAGCATCCTGGTATGTGTCAA 3'
SEPSECS Fw (BspEI)	5' CCTCCGGAAACCGCGAGAGCTTCGCGG 3'
SEPSECS Rv (XbaI)	5' CTAGTCTAGATTATGAAGAAGCATCCTGGTATGTGTCAA 3'
SECp43 Fw (NotI)	5' ATAAGAATGCGGCCGCACCATGGCGGCCAGCCTGTGGATGGGC 3'
SECp43 Rv (ClaI)	5' CCATCGATCATCATGGCAGGGATCTCTGAAGA 3'
SECp43 Fw (BspEI)	5' CCTCCGGAATGGCGGCCAGCCTGTGG 3'
SECp43 Rv (XbaI)	5' CTAGTCTAGATTACATCATGGCAGGGATCTC 3'
SEPHS1 Fw (NotI)	5' ATAAGAATGCGGCCGCACCATGTCTACGCGGGAGTCCTTTAAC 3'
SEPHS1 Rv (ClaI)	5' CCATCGATCAAAGAGGTGGCCCCGGG 3'
SEPHS1 Fw (BspEI)	5' CCTCCGGATCTACGCGGGAGTCCTTTAACC 3'
SEPHS1 Fw (XbaI)	5' CTAGTCTAGATTACAAAGAGGTGGCCCCG 3'
SEPHS2 Fw (NotI)	5' AAGGAAAAAAGCGGCCGCACCATGGCGGAAGCCTCGGCG 3'
SEPHS2 Rv (ClaI)	5' CCATCGATCTGCGAGCTAGGCTCAGAGGAGGCAT 3'
SEPHS2 Fw (BspEI)	5' CCTCCGGAGCGGAAGCCTCGGCGACG 3'
SEPHS2 Fw (XbaI)	5' CTAGTCTAGACTACGAGCTAGGCTCAGAGGAGGCATTTG 3'
RPL30 Fw (NotI)	5' ATAAGAATGCGGCCGCACCATGGTGGCCGCAAAGAAGA 3'
RPL30 Rv (ClaI)	5' CCATCGATCTTTTACCAGTCTGTTCTGGCA 3'
RPL30 Fw (BspEI)	5' CCTCCGGAGTGGCCGCAAAGAAGACG 3'
RPL30 Fw (XbaI)	5' CTAGTCTAGATTACTTTTTACCAGTCTGTTCTGG 3'
SBP2 Fw (NotI)	5' ATAAGAATGCGGCCGCACCATGGTTAGAGTCCTCAGAAGC 3'
SBP2 Rv (ClaI)	5' CCATCGATTAAATTCAAATTCATCATTTGAG 3'
SBP2 Fw (BspEI)	5' CCTCCGGAGTTAGAGTCCTCAGAAGCATGT 3'
SBP2 Rv (XbaI)	5' CTAGTCTAGATCAATTCAAATTCATCATTTGAG 3'
EFSec Fw (BamHI)	5' CGGGATCCACCATGGCATCCTGCTCCACCG 3'
EFSec Rv (NotI)	5' TTTTCCTTTTGCGGCCGCTCAGGGAGACTGAACCATGCG 3'
EFSec Fw (XhoI)	5' GCTCTCGAGACCATGGCATCCTGCTCCACCGCC 3'
EFSec Rv (XbaI)	5' CTAGTCTAGACTCAGGGAGACTGAACCATGCG 3'

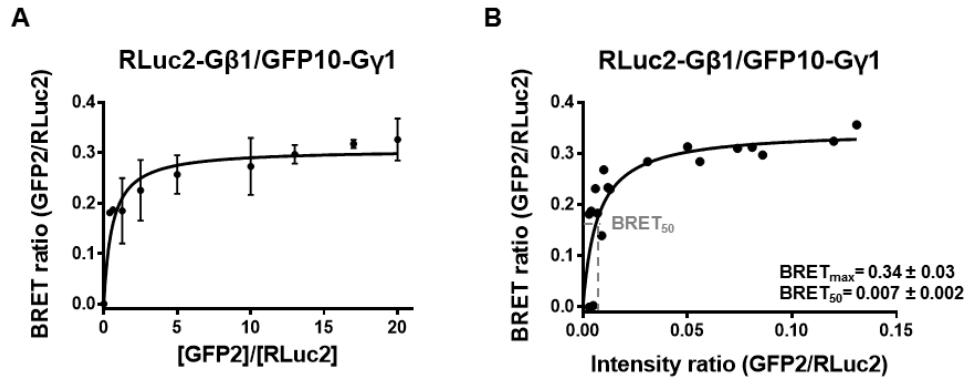
Supplementary Table SII.1.3. Oligonucleotide primers for directed-site mutagenesis

Mutation	Sequence
SEPHS2Cys Fw	5' CGGGACCTTGCAGCCACAGCCCTTCA 3'
SEPHS2Cys Rv	5' TCCGGCATGAAGGGCTGTGGCTGCAA 3'
L163A Fw	5' CAGGGAGCAGTGGGAGCGGGGTCTAAGCCTGTG 3'
L163A Rv	5' CACAGGCTTAGACCCCGCTCCCCTGCTCCCTG 3'
G164A Fw	5' GGAGCAGTGGGACTGGCGTCTAAGCCTGTGCGG 3'
G164A Rv	5' CCGCACAGGCTTAGACGCCAGTCCCCTGCTCC 3'
S165A Fw	5' GCAGTGGGACTGGGGGCTAAGCCTGTGCGGCTG 3'
S165A Rv	5' CAGCCGCACAGGCTTAGCCCCAGTCCCCTGCTGC 3'
K166A Fw	5' GTGGGACTGGGGTCTGCGCCTGTGCGGCTGAGC 3'
K166A Rv	5' GCTCAGCCGCACAGGCGCAGACCCAGTCCCAC 3'
P167A Fw	5' GGACTGGGGTCTAAGGCTGTGCGGCTGAGCGTG 3'
P167A Rv	5' CACGCTCAGCCGCACAGCCTTAGACCCAGTCC 3'
V168A Fw	5' CTGGGGTCTAAGCCTGCGCGGCTGAGCGTGGCA 3'
V168A Rv	5' TGCCACGCTCAGCCGCGCAGGCTTAGACCCAG 3'
R169A Fw	5' GGGTCTAAGCCTGTGGCGCTGAGCGTGGCAATC 3'
R169A Rv	5' GATTGCCACGCTCAGCGCCACAGGCTTAGACCC 3'

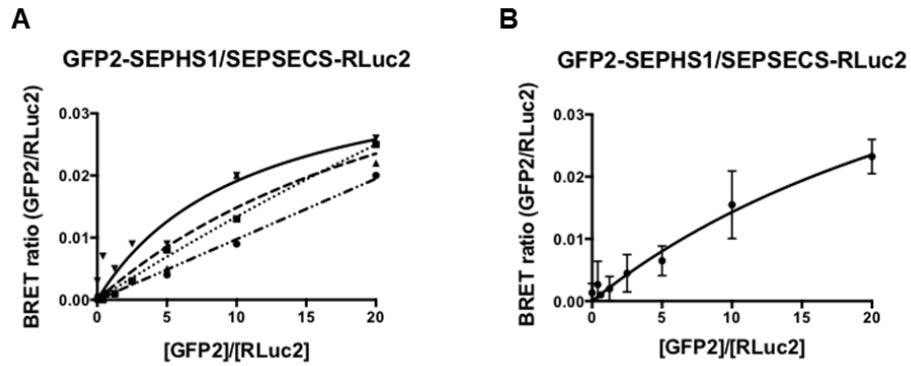
Supplementary Table SII.1.4. Quantitative BRET measurement of SECp43 variants

SECp43 variant	BRET_{max}^(a)	BRET₅₀^(b)
SECp43 _{WT}	0.19 ± 0.02	0.16 ± 0.02
L163A	0.08 ± 0.02	1.24 ± 0.80
G164A	0.05 ± 0.02	0.60 ± 0.80
S165A	0.17 ± 0.04	0.18 ± 0.11
K166A	-	-
P167A	-	-
V168A	0.10 ± 0.04	1.87 ± 0.56
R169A	0.22 ± 0.2	0.48 ± 0.56

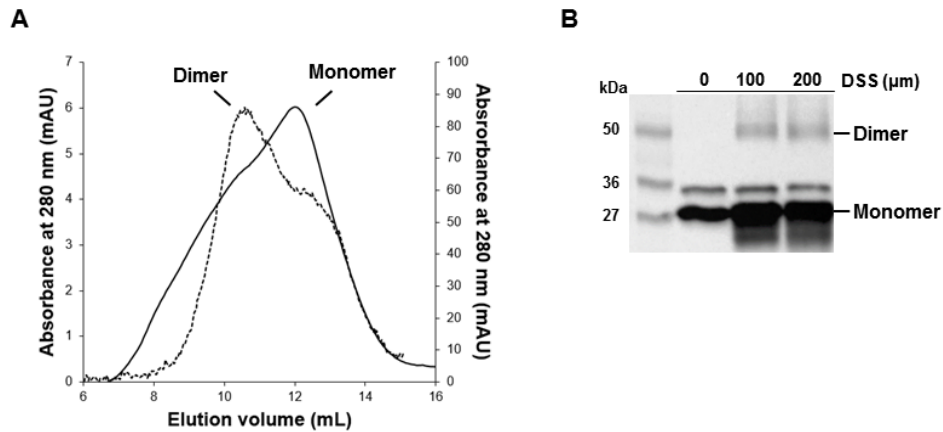
(a and b) These values are presented for information only. For the accuracy of the BRET_{max} determination and the precision of BRET₅₀, the curves must reach a saturation. Following the substitution, the saturation profiles of the variants changed and BRET₅₀ could not be assessed precisely.



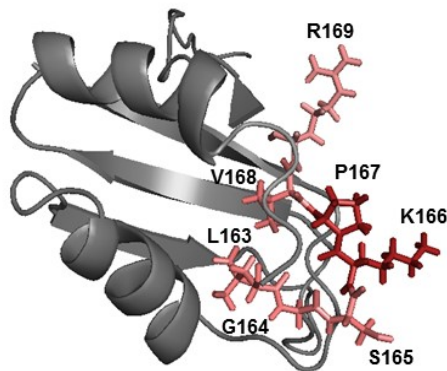
Supplementary Fig. SII.1.1. Measurement of the interaction between Gβ1 and Gγ1 subunits of the G-protein by BRET shows a strong affinity. The two subunits were expressed as fusions to RLuc2 or GFP2 reporters. BRET titration experiments were performed with a fixed amount of RLuc2 donor construct and increasing amounts of GFP2 acceptor constructs and expressed in HEK293T cells. (A) BRET ratio (Y-axis) corresponds to the ratio between the receptor-GFP2 fluorescence (515 nm) and the donor-RLuc2 luminescence (410 nm) emitted after addition of the luciferase substrate. The [receptor-GFP2]/[donor-RLuc2] ratio were plotted on X-axis. Data are means ± SD of two independent titration experiments, analyzed and fitted using non-linear regression assuming a model with one site binding (GraphPad Prism). (B) Total fluorescence of GFP10-Gγ1 and total luminescence of RLuc2-Gβ1 were measured without resonance energy transfer, before the addition of RLuc substrate, then expressed as intensity ratio (GFP2/RLuc2) and plotted on X-axis. BRET ratios were plotted on Y-axis. The results of two independent experiments were plotted on the same graph. Results were analyzed and fitted using non-linear regression assuming a model with one site binding (GraphPad Prism). The dashed line (grey) represents a schematic evaluation of the BRET₅₀ value that corresponds to the acceptor/donor ratio at half-maximal BRET value and indicates the relative strength of the dimerization.



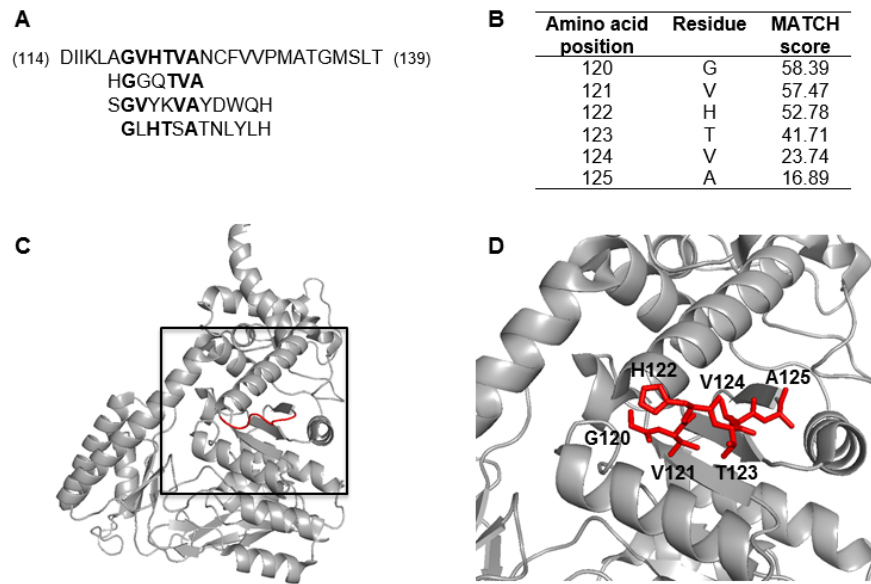
Supplementary Fig. SII.1.2. SEPSECS-SEPHS1 interaction BRET profiles. SEPSECS and SEPHS1 were expressed as fusions to RLuc2 or GFP2 reporters respectively. BRET titration experiments were performed with a fixed amount of RLuc2 donor construct and increasing amounts of GFP2 acceptor constructs and expressed in HEK293T cells. BRET ratios (515 nm/410 nm) (Y-axis), measured after addition of the luciferase substrate, were plotted against [GFP2-SEPHS1]/ [SEPSECS-RLuc2] ratio (X-axis). Results were analysed and fitted using non-linear regression assuming a model with one site binding (GraphPad Prism). A) Profiles of four independent titration assay leading to different BRET profiles. B) Mean and SD of the four BRET titration experiments. The interaction did not led to saturation curves nor a linear relationship and was designated as "ambiguous" in Table II.1.1.



Supplementary Fig. SII.1.3. *In vitro* dimerization of purified SECp43 in solution. (A) Two-step purification of SECp43 dimeric form by size-exclusion chromatography. 0,5 mg of purified SECp43¹⁸⁰ were first injected on a size-exclusion column (full line; right Y axis), leading to a mixture of monomeric and dimeric form of SECp43. The fraction eluted around 10,5 ml was re-injected into the column (dashed line; left Y axis), leading to a majority of dimeric form. (B) Effect of DSS in dimerization of SECp43. His₆-SECp43¹⁸⁰ was incubated with increasing concentrations of DSS (0 to 200 µM) and separated by electrophoresis on SDS-PAGE. SECp43 was detected using anti-His antibody in an immunoblot. The positions of monomeric (25 kDa) and dimeric (50 kDa) proteins are highlighted.



Supplementary Fig. SII.1.4. Structural representation of the residues involved in SECp43 dimerization. The LGSKPVR peptide on the RRM2 domain of SECp43 (grey). Residues essential for dimerization of SECp43 dimerization are shown in red (K166 and P167), the other residues of the LGSKPVR peptide are represented in pink.



Supplementary Fig. SII.1.5. Potential interaction site between SECp43 and SEPSECS identified by phage display. A) Potential interaction site on SEPSECS sequence determined by alignment of the 30 peptides identified during phage display experiments and illustrated in Fig. II.1.5A. Aligning amino acids are shown in bold print. B) MATCH score for indicating probability to be part of the interacting peptide for each residue [37]. C) Structural representation of the localization of the potential interacting site (red) on a SEPSECS monomer (PDB: 3HL2). D) Enlarged view of the residues potentially involved in SECp43-SEPSECS interaction (red).

II.1.10 References

1. Bock, A., Forchhammer, K., Heider, J., Leinfelder, W., Sawers, G., Veprek, B., and Zinoni, F. (1991) Selenocysteine: the 21st amino acid, *Mol Microbiol* 5, 515-520.
2. Squires, J. E., and Berry, M. J. (2008) Eukaryotic selenoprotein synthesis: mechanistic insight incorporating new factors and new functions for old factors, *IUBMB life* 60, 232-235.
3. Hatfield, D. L., Choi, I. S., Ohama, T., Jung, J. E., and Diamond, A. M. (1994) Selenocysteine tRNA isoacceptors as central components in selenoprotein biosynthesis in eukaryotes, In *Selenium in Biology and Human Health* (Burk, R. F., Ed.), pp 25–44, Springer New York.
4. Lee, B. J., Worland, P. J., Davis, J. N., Stadtman, T. C., and Hatfield, D. L. (1989) Identification of a selenocysteyl-tRNA(Ser) in mammalian cells that recognizes the nonsense codon, UGA, *The Journal of biological chemistry* 264, 9724-9727.
5. Leinfelder, W., Stadtman, T. C., and Bock, A. (1989) Occurrence in vivo of selenocysteyl-tRNA(SERUCA) in Escherichia coli. Effect of sel mutations, *The Journal of biological chemistry* 264, 9720-9723.
6. Carlson, B. A., Xu, X. M., Kryukov, G. V., Rao, M., Berry, M. J., Gladyshev, V. N., and Hatfield, D. L. (2004) Identification and characterization of phosphoseryl-tRNA[Ser]Sec kinase, *Proceedings of the National Academy of Sciences of the United States of America* 101, 12848-12853.
7. Xu, X. M., Carlson, B. A., Mix, H., Zhang, Y., Saira, K., Glass, R. S., Berry, M. J., Gladyshev, V.N., and Hatfield, D. L. (2007) Biosynthesis of selenocysteine on its tRNA in eukaryotes, *PLoS biology* 5, e4.
8. Berry, M. J., Banu, L., Chen, Y. Y., Mandel, S. J., Kieffer, J. D., Harney, J. W., and Larsen, P. R. (1991) Recognition of UGA as a selenocysteine codon in type I deiodinase requires sequences in the 3' untranslated region, *Nature* 353, 273-276.
9. Low, S. C., and Berry, M. J. (1996) Knowing when not to stop: selenocysteine incorporation in eukaryotes, *Trends Biochem Sci* 21, 203-208.
10. Tujebajeva, R. M., Copeland, P. R., Xu, X. M., Carlson, B. A., Harney, J. W., Driscoll, D. M., Hatfield, D. L., and Berry, M. J. (2000) Decoding apparatus for eukaryotic selenocysteine insertion, *EMBO reports* 1, 158-163.
11. Copeland, P. R., Fletcher, J. E., Carlson, B. A., Hatfield, D. L., and Driscoll, D. M. (2000) A novel RNA binding protein, SBP2, is required for the translation of mammalian selenoprotein mRNAs, *The EMBO journal* 19, 306-314.
12. Copeland, P. R., Stepanik, V. A., and Driscoll, D. M. (2001) Insight into mammalian selenocysteine insertion: domain structure and ribosome binding properties of Sec insertion sequence binding protein 2, *Molecular and cellular biology* 21, 1491-1498.

13. Low, S. C., Grundner-Culemann, E., Harney, J. W., and Berry, M. J. (2000) SECIS-SBP2 interactions dictate selenocysteine incorporation efficiency and selenoprotein hierarchy, *The EMBO journal* 19, 6882-6890.
14. Chavatte, L., Brown, B. A., and Driscoll, D. M. (2005) Ribosomal protein L30 is a component of the UGA-selenocysteine recoding machinery in eukaryotes, *Nature structural & molecular biology* 12, 408-416.
15. Ding, F., and Grabowski, P. J. (1999) Identification of a protein component of a mammalian tRNA(Sec) complex implicated in the decoding of UGA as selenocysteine, *RNA* 5, 1561-1569.
16. Xu, X. M., Mix, H., Carlson, B. A., Grabowski, P. J., Gladyshev, V. N., Berry, M. J., and Hatfield, D. L. (2005) Evidence for direct roles of two additional factors, SECp43 and soluble liver antigen, in the selenoprotein synthesis machinery, *The Journal of biological chemistry* 280, 41568-41575.
17. Yoo, M. H., Xu, X. M., Carlson, B. A., Gladyshev, V. N., and Hatfield, D. L. (2006) Thioredoxin reductase 1 deficiency reverses tumor phenotype and tumorigenicity of lung carcinoma cells, *The Journal of biological chemistry* 281, 13005-13008.
18. Small-Howard, A., Morozova, N., Stoytcheva, Z., Forry, E. P., Mansell, J. B., Harney, J. W., Carlson, B. A., Xu, X. M., Hatfield, D. L., and Berry, M. J. (2006) Supramolecular complexes mediate selenocysteine incorporation *in vivo*, *Molecular and cellular biology* 26, 2337-2346.
19. Low, S. C., Harney, J. W., and Berry, M. J. (1995) Cloning and functional characterization of human selenophosphate synthetase, an essential component of selenoprotein synthesis, *The Journal of biological chemistry* 270, 21659-21664.
20. Kim, I. Y., and Stadtman, T. C. (1995) Selenophosphate synthetase: detection in extracts of rat tissues by immunoblot assay and partial purification of the enzyme from the archaean *Methanococcus vannielii*, *Proceedings of the National Academy of Sciences of the United States of America* 92, 7710-7713.
21. Guimaraes, M. J., Peterson, D., Vicari, A., Cocks, B. G., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., Ferrick, D. A., Kastelein, R. A., Bazan, J. F., and Zlotnik, A. (1996) Identification of a novel selD homolog from eukaryotes, bacteria, and archaea: is there an autoregulatory mechanism in selenocysteine metabolism?, *Proceedings of the National Academy of Sciences of the United States of America* 93, 15086-15091.
22. Kim, T. S., Yu, M. H., Chung, Y. W., Kim, J., Choi, E. J., Ahn, K., and Kim, I. Y. (1999) Fetal mouse selenophosphate synthetase 2 (SPS2): biological activities of mutant forms in *Escherichia coli*, *Mol Cells* 9, 422-428.
23. Tamura, T., Yamamoto, S., Takahata, M., Sakaguchi, H., Tanaka, H., Stadtman, T. C., and Inagaki, K. (2004) Selenophosphate synthetase genes from lung adenocarcinoma cells: Sps1 for recycling L-selenocysteine and Sps2 for selenite assimilation, *Proceedings of the National Academy of Sciences of the United States of America* 101, 16162-16167.

24. Studier, F. W. (2005) Protein production by auto-induction in high density shaking cultures, *Protein Expr Purif* 41, 207-234.
25. Smith, M. A., Coincon, M., Paschos, A., Jolicoeur, B., Lavallee, P., Sygusch, J., and Baron, C. (2012) Identification of the Binding Site of Brucella VirB8 Interaction Inhibitors, *Chem Biol* 19, 1041-1048.
26. Yuan, Q., Carle, A., Gao, C., Sivanesan, D., Aly, K. A., Hoppner, C., Krall, L., Domke, N., and Baron, C. (2005) Identification of the VirB4-VirB8-VirB5-VirB2 pilus assembly sequence of type IV secretion systems, *The Journal of biological chemistry* 280, 26349-26359.
27. Nielsen, S. S., Toft, K. N., Snakenborg, D., Jeppesen, M. G., Jacobsen, J. K., Vestergaard, B., Kutter, J. P., and Arleth, L. (2009) BioXTAS RAW, a software program for high-throughput automated small-angle X-ray scattering data reduction and preliminary analysis, *J Appl Crystallogr* 42, 959-964.
28. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., and Svergun, D. I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis, *J Appl Crystallogr* 36, 1277-1282.
29. Svergun, D. I. (1992) Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria, *J Appl Crystallogr* 25, 495-503.
30. Petoukhov, M. V., Konarev, P. V., Kikhney, A. G., and Svergun, D. I. (2007) ATSAS 2.1 - towards automated and web-supported small-angle scattering data analysis, *J Appl Crystallogr* 40, S223-S228.
31. Franke, D., Kikhney, A. G., and Svergun, D. I. (2012) Automated acquisition and analysis of small angle X-ray scattering data, *Nucl Instrum Meth A* 689, 52-59.
32. Svergun, D. I., Petoukhov, M. V., and Koch, M. H. J. (2001) Determination of domain structure of proteins from X-ray solution scattering, *Biophysical journal* 80, 2946-2953.
33. Petoukhov, M. V., and Svergun, D. I. (2005) Global rigid body modeling of macromolecular complexes against small-angle scattering data, *Biophysical journal* 89, 1237-1250.
34. Kozin, M. B., and Svergun, D. I. (2001) Automated matching of high- and low-resolution structural models, *J Appl Crystallogr* 34, 33-41.
35. Villamil Giraldo, A. M., Sivanesan, D., Carle, A., Paschos, A., Smith, M. A., Plesa, M., Coulton, J., and Baron, C. (2012) Type IV secretion system core component VirB8 from Brucella binds to the globular domain of VirB5 and to a periplasmic domain of VirB6, *Biochemistry* 51, 3881-3890.
36. Villamil Giraldo, A. M., Mary, C., Sivanesan, D., and Baron, C. (2015) VirB6 and VirB10 from the Brucella type IV secretion system interact via the N-terminal periplasmic domain of VirB6, *FEBS letters* 589, 1883-1889.
37. Mandava, S., Makowski, L., Devarapalli, S., Uzubell, J., and Rodi, D. J. (2004) RELIC—a bioinformatics server for combinatorial peptide analysis and identification of protein-ligand interaction sites, *Proteomics* 4, 1439-1460.

38. Mercier, J. F., Salahpour, A., Angers, S., Breit, A., and Bouvier, M. (2002) Quantitative assessment of beta 1- and beta 2-adrenergic receptor homo- and heterodimerization by bioluminescence resonance energy transfer, *The Journal of biological chemistry* 277, 44925-44931.
39. Lavoie, H., Thevakumaran, N., Gavory, G., Li, J. J., Padeganeh, A., Guiral, S., Duchaine, J., Mao, D. Y., Bouvier, M., Sicheri, F., and Therrien, M. (2013) Inhibitors that stabilize a closed RAF kinase domain conformation induce dimerization, *Nat Chem Biol* 9, 428-436.
40. Schmidt, C. J., Thomas, T. C., Levine, M. A., and Neer, E. J. (1992) Specificity of G protein beta and gamma subunit interactions, *The Journal of biological chemistry* 267, 13807-13810.
41. Morris, A. J., and Malbon, C. C. (1999) Physiological regulation of G protein-linked signaling, *Physiological reviews* 79, 1373-1430.
42. Gilman, A. G. (1987) G proteins: transducers of receptor-generated signals, *Annu Rev Biochem* 56, 615-649.
43. Ayoub, M. A., Couturier, C., Lucas-Meunier, E., Angers, S., Fossier, P., Bouvier, M., and Jockers, R. (2002) Monitoring of ligand-independent dimerization and ligand-induced conformational changes of melatonin receptors in living cells by bioluminescence resonance energy transfer, *The Journal of biological chemistry* 277, 21522-21528.
44. Pflieger, K. D., and Eidne, K. A. (2006) Illuminating insights into protein-protein interactions using bioluminescence resonance energy transfer (BRET), *Nat Methods* 3, 165-174.
45. Issad, T., and Jockers, R. (2006) Bioluminescence resonance energy transfer to monitor protein-protein interactions, *Methods Mol Biol* 332, 195-209.
46. Szalai, B., Hoffmann, P., Prokop, S., Erdelyi, L., Varnai, P., and Hunyady, L. (2014) Improved methodical approach for quantitative BRET analysis of G Protein Coupled Receptor dimerization, *PLoS one* 9, e109503.
47. Palioura, S., Sherrer, R. L., Steitz, T. A., Soll, D., and Simonovic, M. (2009) The human SepSecStRNasec complex reveals the mechanism of selenocysteine formation, *Science* 325, 321-325.
48. Wang, K. T., Wang, J., Li, L. F., and Su, X. D. (2009) Crystal structures of catalytic intermediates of human selenophosphate synthetase 1, *Journal of molecular biology* 390, 747-759.

II.2 Criblage et identification de molécules inhibant l'activité ATPase et la formation d'hexamère de Cag α du système de sécrétion de type IV d'*H. pylori*

II.2.1 Présentation de l'article 2 et contribution des coauteurs

L'objectif de ce projet fut, tout d'abord, d'identifier des molécules inhibitrices de l'activité de Cag α , une ATPase du SST4 d'*H. pylori*. L'étude a initié par le criblage, par DSF, d'une banque disponible de petites molécules capables de lier la protéine Cag α . L'effet *in vitro* sur l'activité ATPase de Cag α de 16 molécules préalablement identifiées fut ensuite mesuré par essai enzymatique. Les expériences de gel filtration, de *crosslinking* et de microscopie électronique ont ensuite mis en évidence que certaines des molécules, dont la molécule 1G2, inhibaient la formation d'hexamères de Cag α . Puis, la protéine Cag α fut co-cristallisée en présence de la molécule 1G2. La toxicité des molécules fut, ensuite, analysée *in vivo*. Finalement, à la suite de l'infection des cellules AGS par *H. pylori* en présence de la molécule 1G2, il fut observé que 1G2 réduisait de 50% la production d'interleukine-8, mais il ne fut pas observé d'effet d'inhibition sur la translocation de CagA.

Tarun Arya a produit et purifié la protéine Cag α . En collaboration avec Benoit Bessette, il a procédé au criblage de la banque de petites molécules par DSF. Il a ensuite réalisé les essais enzymatiques permettant la mesure de l'activité ATPase de Cag α , les expériences de *crosslinking* et la microscopie électronique. Après avoir obtenu le cristal de Cag α complexée à la molécule 1G2, il a travaillé en collaboration avec Bastien Casu et le Dr Sygusch pour analyser les données de cristallographie et déterminer un modèle structural. Finalement, Tarun Arya a rédigé le manuscrit en grande partie.

Dans le cadre de cet article, j'ai réalisé les études menant à la caractérisation des molécules *in vivo*, et nécessitant la manipulation des cellules AGS et de la bactérie *H. pylori*. J'ai donc réalisé les expériences de culture bactérienne et cellulaire. J'ai déterminé la toxicité des inhibiteurs sur les cellules eucaryotes et procaryotes. Puis j'ai optimisé les essais pour permettre d'établir l'effet des molécules sur la translocation de la protéine CagA et sur la sécrétion d'interleukine-8. Finalement, j'ai participé à la rédaction et à la correction du manuscrit.

Toutes les expériences ont été faites sous la supervision du Dr Baron qui a participé à la conception des projets, à l'analyse des résultats, à la rédaction du manuscrit et au processus de révision.

Cet article a été soumis pour publication à la revue *ACS Chemical Biology* en Juillet 2018. Il est en cours de révision.

Article 2

Fragment-based screening identifies inhibitors of the ATPase activity and of hexamer formation of Cag α from the *Helicobacter pylori* type IV secretion system

Tarun Arya, Flore Oudouhou, Bastien Casu, Benoit Bessette, Jurgen Sygusch
and Christian Baron*

Department of Biochemistry and Molecular Medicine, Faculty of Medicine,
Université de Montréal, Québec, Canada

***Corresponding Author:** E-mail: christian.baron@umontreal.ca

Keywords: Type IV secretion, fragment-based screening, enzyme inhibitors, drug design, stomach cancer, ulcer disease, anti-virulence drug, antimicrobial resistance

II.2.2 Abstract

Type IV secretion systems are membrane-bound multiprotein complexes that mediate the translocation of macromolecules across the bacterial cell envelope. In *Helicobacter pylori* a type IV secretion system is encoded by the *cag* pathogenicity island that encodes 27 Cag proteins and most of these are essential for bacterial virulence. We here present our work on the identification and characterization of inhibitors of Cag α , a hexameric ATPase and member of the family of VirB11-like proteins that is essential for translocation of the CagA cytotoxin into mammalian cells. We conducted fragment-based screening using a differential scanning fluorimetry assay and identified 16 molecules that stabilize the protein during thermal denaturation suggesting that they bind Cag α . Several of these molecules affect binding of ADP and four of them inhibit the ATPase enzyme activity of Cag α . Analysis of enzyme kinetics suggests that their mode of action is non-competitive, suggesting that they do not bind to the ATPase active site. Cross-linking analysis suggests that the active molecules change the conformation of the protein and gel filtration and transmission electron microscopy show that molecule 1G2 dissociates the Cag α hexamer. Analysis by X-ray crystallography reveals that molecule 1G2 binds at the interface between Cag α subunits. Addition of the molecule 1G2 inhibits the induction of interleukin-8 production in gastric cancer cells after co-incubation with *H. pylori* suggesting that it inhibits Cag α in vivo. Our results reveal a novel mechanism for the inhibition of the ATPase activity of VirB11-like proteins and the identified molecules have potential for the development into anti-virulence drugs.

II.2.3 Introduction

Helicobacter pylori is a widespread pathogenic bacterium that lives in the stomach of over half of the world's population [1]. The infection with virulent strains causes inflammatory reactions, gastritis, peptic ulcers and it is one of the principal causes of stomach cancer in humans [2-3]. Antibiotic treatments using combination therapies of three or four drugs have generally been successful, but eradication therapy is becoming increasingly difficult due to rising resistance against many antimicrobial agents, such as clarithromycin and metronidazole [4]. Novel treatment options are therefore urgently needed and targeting bacterial virulence factors to

attenuate the inflammation is a strategy that could complement or even replace currently used eradication treatments.

Type IV secretion systems (T4SS) mediate the transfer of virulence factors across the cell envelope of many bacterial pathogens as well as the exchange of plasmids contributing to the spread of antibiotic resistance genes [5-6]. *H. pylori* strains encode T4SSs that mediate the uptake of DNA as well as bacterial virulence like the *cag* pathogenicity island (*cag*-PAI)-encoded T4SS comprising 27 components of which most are essential for bacterial virulence [7-10]. The *cag*-PAI is required for the transfer of the CagA cytotoxin into mammalian cells where it is phosphorylated by Src kinase at tyrosine residues and its interactions with mammalian proteins such as SHP-2 and Grb-2 lead to rearrangements of the cytoskeleton and to proinflammatory reactions [11]. The *cag*-PAI-encoded T4SS is also a conduit for bacterial murein and for the small molecule metabolite heptulose-1,7-bisphosphate triggering signalling cascades via Nod-1 and TIFA, respectively, that contribute to the proinflammatory response [12-13].

The *H. pylori* *cag*-PAI encodes 27 proteins including homologs of all 12 components of the most studied model T4SS from *Agrobacterium tumefaciens* [9]. These conserved proteins are critical for secretion system function and they are either part of surface-exposed pili of the periplasmic T4SS core complex or they energize T4SS assembly or substrate translocation. We here focus on the Cag α (HP0525) protein that is a member of the VirB11 family of ATPases present in all T4SSs. Electron microscopic (EM) analyses and X-ray crystallography have shown that the overall structures of VirB11-like proteins from different organisms are very similar comprising homo-hexameric rings [14-15]. The monomeric subunit consists of an N-terminal domain (NTD) and a C-terminal domain (CTD) that are linked via a short linker region comprising the nucleotide binding site. The X-ray structures of Cag α apoprotein [16], as well as of its complexes with ADP [17] and with the inhibitor ATP γ S [16] have been solved. These studies revealed that the CTD forms a ‘six clawed grapple’ mounted onto the NTD, forming a hexameric ring and a dome-like chamber that is closed at one end and opened at the other [17]. Glycerol gradient centrifugation showed a large conformational change of VirB11 homologs from plasmid RP4 (TrbB) and *H. pylori* upon binding to ATP, underlining the dynamic nature of the protein [16, 18]. The other available X-ray structure from *Brucella suis* VirB11 differs

from Cag α by a domain swap of the large linker region between NTD and CTD [19], but the overall structure is very similar.

Since T4SS are important for bacterial virulence they are very interesting targets for the development of drugs that disarm but do not kill bacterial pathogens [20-21]. In our previous work, we have identified inhibitors of the dimerization of VirB8-like proteins from *B. suis* and plasmid pKM101 using the bacterial two-hybrid system and fragment-based screening approaches and we identified molecules that reduce T4SS function [22-25]. Other groups have identified peptidomimetic inhibitors of the *H. pylori* T4SS, but the targets of these molecules are not known [26]. Certain unsaturated fatty acids inhibit bacterial conjugation and the ATPase activity of the VirB11 homolog TrwD from plasmid R388, but there is no high-resolution structural information available on their binding site [27-28]. High-throughput small molecules screening and chemical synthesis led to the identification of inhibitors of the ATPase activity of Cag α that likely bind at the ATPase active site, but structural information on their binding site is not available [29-30]. Whereas the isolation of competitive inhibitors of the ATPase activity of VirB11 homologs is interesting, there are concerns about the specificity of these molecules since they may also inhibit other ATPases in bacteria or in mammalian cells.

To identify novel chemical entities that inhibit Cag α we here present an unbiased approach that does not specifically target its ATPase activity. To this effect, we carried out fragment based-screening using differential scanning fluorimetry (DSF) to identify molecules that bind and stabilize Cag α [31]. Four of the molecules inhibit the Cag α ATPase activity and the most potent molecule impacts the conformation of the protein and dissociates the hexamer. X-ray crystallography reveals that this molecule causes conformational changes, that it binds at the interface between Cag α monomers and it inhibits the production of interleukin-8 upon interaction between *H. pylori* and mammalian cells.

II.2.4 Methods

II.2.4.1 Bacterial strains, cell lines and culture conditions.

H. pylori strains 26695 and Δ cagV (*hp0530*) mutant have been described [35] and were cultivated on Columbia agar base (BD) containing 10% (v/v) defibrinated horse blood (Wisent

Inc.), vancomycin (10 µg/ml) and amphotericin B (10 µg/ml). Chloramphenicol (34 µg/ml) was added in case of the *ΔcagV* strain to select for the cam gene cassette used to disrupt the gene. For liquid culture, brain heart infusion (BHI) media (Oxoid) were supplemented with 8% fetal bovine serum (FBS) and appropriate antibiotics. Bacteria were cultivated at 37°C, under microaerophilic conditions (5% oxygen, 10% CO₂). AGS cells were grown at 37°C in F12K media (Wisent Inc.) with 10% (v/v) FBS (Wisent Inc.) in a 5% CO₂ containing atmosphere.

II.2.4.2 Cloning, expression and purification of Cagα

The Cagα encoding gene from *H. pylori* 26695 (ATCC) was PCR-amplified from genomic DNA with primers (forward, 5'-TAGCGAATTCGGTACCATGACTGAAGACAGATTGAGTGCA-3' and reverse, 5'-CGATGAATTCCTCGAGCTACCTGTGTGTTTGATATAAAATTC-3'). The PCR product was ligated in between restriction enzymes *NheI* and *XhoI*, into expression vector pET28a. Expression was conducted in *E. coli* BL21 (DE3) cultivated in two liters of LB-medium at 37 °C at 220 rpm, protein production was induced at OD₆₀₀ of 1.0 with 1 mM isopropylthio-β-galactoside (IPTG), followed by further incubation for 16 h at 25°C. For purification, the cell pellet was suspended in binding buffer (50 mM HEPES, 500 mM NaCl, 20 mM imidazole, pH 7.5, 10% glycerol, 0.1% triton, plus two tablets of EDTA-free protease inhibitor cocktail (Roche)) and lysed using a cell disrupter (Constant Systems Inc.) at 27 kPsi, followed by centrifugation at 15,000 rpm at 4°C to reduce cell debris. The supernatant was loaded onto a His-trap Ni-NTA column (GE Healthcare), and eluted using a linear 50 ml gradient of 40-500 mM imidazole in binding buffer. Proteins were then dialysed (25 mM sodium phosphate, 125 mM NaCl, 5 mM DTT, pH 7.4) and subjected to Size exclusion chromatography using a Superdex-200 column (GE Healthcare) with buffer 25 mM HEPES pH7.5 and 100 mM NaCl and peak fractions were analyzed by SDS-PAGE. The fractions containing Cagα hexamers were pooled and concentrated to 6 mg/ml for crystallographic studies.

II.2.4.3 Analytical gel filtration chromatography

Purified protein was further characterized by analytical gel filtration (Superdex 200) in 25 mM HEPES, pH 7.5 and 50 mM NaCl (pH 7.5). The column volume was 3ml and the protein was injected at a flow rate of 0.5 ml/min. To study the effects of ATP-γ-S and of 1G2, 35 µg of Cagα

was pre-incubated with 2 mM of the molecules for 30 min, followed by analytical size exclusion analysis.

II.2.4.4 Enzyme activity assay

The ATPase activity was quantified using a malachite green binding assay [36]. The 100 μ L reaction mixtures contained 25 mM HEPES (pH 7.5), 100 mM NaCl, 60 nM of enzyme and 200 μ M of $MgCl_2$ with different concentrations of ATP (0 μ M – 320 μ M) to determine kinetic parameters. The reaction mixtures were incubated for 30 min at 30°C and then 40 μ L of malachite green assay mixture was added. The formation of the blue phosphomolybdate-malachite green complex was in linear relation to the amount of released inorganic phosphate and measured at 610 nm. To study the mechanism of inhibition, the concentrations of inhibitors were varied between 0 and 500 μ M with different concentrations of ATP (0-40 μ M). Initial velocity data were fit using nonlinear regression analysis to each of the equations describing partial and full models of competitive, uncompetitive, non-competitive, and mixed inhibition using the Enzyme Kinetics Module of SigmaPlot (SigmaPlot version 11.0 software). On the basis of the analysis of fits through “goodness-of-fit” statistics, the full non-competitive inhibition model was determined with the equation $v = V_{max}/[(1 + [I]/K_i) \times (1 + K_m/[S])]$, where $[S] = [ATP]$, $[I] = [1G2]$.

II.2.4.5 IC₅₀ determination

IC₅₀ values were determined by incubating different concentrations of molecules (10 - 1,000 μ M; from stocks of 200 mM) with enzyme in 25 mM HEPES (pH 7.5) and 100 mM NaCl. Mixtures were incubated with inhibitors for 15 min, followed by addition of ATP and incubation for 30 min at 37 °C. The reactions were stopped by addition of 40 μ l malachite green solution and the inorganic phosphate released was determined at 610 nm. Data were plotted as 1/rate versus inhibitor concentration for each substrate concentration and a linear fit was calculated by non-linear regression using SigmaPlot (version 11.0).

II.2.4.6 Differential scanning fluorimetry (DSF)

A fragment library of 505 molecules was used as in our previous work [33]. The reaction mixture contains 5 μ M of Cag α , 10x concentration of SYPRO Orange (from 5000x stock

solution (ThermoFisher)) in 50 mM HEPES (pH 7.5), 100 mM NaCl and 5% final concentration of DMSO. The fragments and nucleotides were added to final concentrations of 5 mM, and the fluorescence was monitored over 20–95 °C with a LightCycler 480 instrument (Roche).

II.2.4.7 Crystallisation and structure determination

Initial crystallization conditions were established using the MCSG screen from (ANATRACE, USA) using 6 mg/ml of Cag α and 1 mM of 1G2 (1:10 ratio). Final crystals were grown at room temperature using the hanging drop vapour diffusion method in 100 mM Bis-Tris (pH 6.5) and 2 M ammonium sulfate. Drops containing 2 μ l of protein-inhibitor-mixture (1:10 ratio) and 2 μ l of reservoir solution were incubated for 2 weeks. Hexagonally-shaped crystals appeared after 7-10 days. The crystals were cryo-protected in 100 mM Tris-HCl buffer (pH 8.5), 2 M ammonium sulfate and 25% glycerol, flash frozen in liquid nitrogen and the data were collected at microfocus beamline F1 at the Cornell High Energy Synchrotron Source (CHESS). The intensity data was processed using the HKL2000 [37] program in p6522 space group (Supplementary table II.2.1). The structure was solved by molecular replacement using the coordinates of PDB ID: 1G6O as search model. Refinement and modeling were performed using REFMAC and Coot [38-39]. Final graphical figures and tables were generated using the Pymol-integrated Phenix software suite [40].

II.2.4.8 Analysis of protein-protein interactions by cross-linking

Chemical cross-linking with disuccinimidyl suberate (DSS; Pierce) was performed as described [41]. 100 nM of Cag α in 50 mM HEPES (pH 7.5) and 100 mM NaCl were first incubated with cofactors (MgCl₂, ADP) or inhibitors (ATP- γ -S, 1G2) for 30 min, followed by crosslinking with DSS (0 - 50 μ M) for 1 h, and reactions were stopped by mixing with an equal volume of 2 x Laemmli buffer. The formation of cross-linking products was analyzed by SDS-PAGE and western blotting using His-tag specific antiserum and ImageLab 4.0 software (Bio-Rad).

II.2.4.9 Electron microscopy and image processing

Carbon-coated grids were negatively glow-discharged at 15 mA and 0.4 mBar for 30 sec. 5 μ l of purified protein at a concentration of 2 ng/ μ l was spotted onto the grids for 60 sec and blotted using grade 1 Whatman filter paper, followed by staining with freshly prepared 1.5% uranyl

formate solution for 60 sec and drying. The samples were imaged at a magnification of 49,000-fold (pixel size: 2.2 Å/pixel) with a defocus of -2.5 µm using a FEI Tecnai T12 electron microscope (FEMR facility at McGill University). Transmission Electron Microscope (TEM) equipped with a Tungsten filament and operated at 120 kV equipped with a 4k x 4k CCD camera (Gatan Ultrascan 4000 CCD camera system model 895). Subsequently, the images were processed using ImageJ.

II.2.4.10 Measurement of *H. pylori* and AGS cell viability

AGS cell viability was monitored using Cell Proliferation Reagent WST-1 (Sigma). To evaluate the sensitivity of *H. pylori* to 1G2 and its derivatives, freshly harvested bacteria were spread on a 150-mm agar plate. Increasing concentrations of compounds (50-500 µM) were spotted onto Whatman paper disks and growth was observed after 72 h incubation at 37°C under microaerophilic conditions and compared to antibiotics (50-250 µM).

II.2.4.11 Assay for monitoring CagA transfer into AGS cells

Preceding the infection, an overnight culture of *H. pylori* was pre-incubated with 1G2 and its derivatives for 30 min. AGS cells at 6×10^5 cells/well density in 6-well plates were infected with the pre-treated cultures of *H. pylori* for 3-6 h at a multiplicity of infection of 100:1. Cells were washed twice with PBS, harvested and lysed at 4°C in RIPA buffer (150 mM NaCl, 50 mM Tris/ HCl, pH 8, 1% NP-40, 2 mM Na₃VO₄, supplemented with Complete Protease Inhibitor Tablet (Roche). After 15 min of centrifugation at 16,000 g, lysates were separated by SDS-PAGE, followed by western blotting with mouse polyclonal antiserum raised against CagA (Abcam), anti-phosphotyrosine (PY99; Santa Cruz Biotechnology) and anti-β-actin (C4, Santa Cruz Biotechnology).

II.2.4.12 Assay for IL-8 induction

Preceding the infection, an overnight culture of *H. pylori* was pre-incubated with 1G2 and its derivatives for 30 min. AGS cells at 6×10^5 cells/well density in 6-well plates were infected with the pre-treated cultures of *H. pylori* at a multiplicity of infection of 100:1. After 24 h incubation under microaerophilic conditions, supernatants were sampled and centrifuged (15,000 g), before

freezing at -80°C . The level of IL-8 in cell culture supernatants was determined by using a commercially available human IL-8 ELISA kit (Invitrogen).

II.2.5 Results and Discussion

VirB11 like proteins (*Cag α* in *Helicobacter pylori*) are key components of the T4SS and play a crucial role in energizing the T4SS for substrate secretion. To identify novel chemical entities against *Cag α* , we have developed a fragment-based screening approach to identify inhibitors of protein-protein interactions. We have previously used this approach to identify inhibitors of the VirB8 homolog TraE from the plasmid pKM101 conjugation system. We identified molecules that target a known inhibitor binding site on VirB8-like proteins, and we also identified a new binding site showing the potential for the discovery of bioactive molecules and of novel inhibitor target sites [23-24, 32].

II.2.5.1 Differential scanning fluorimetry to identify *Cag α* -binding fragments

We conducted a fragment-based screening approach by using a DSF assay. This assay measures binding of molecules to proteins by changes of the thermal melting profile in the presence of the fluorescent dye Sypro Orange [31]. We validated this assay by testing binding to previously characterized ligands that influence the conformation of *Cag α* , such as MgCl_2 , ADP and the non-hydrolysable substrate analog ATP- γ -S. Addition of the nucleotide ligand MgCl_2 increases the melting temperature from 37°C to 42°C , but in the presence of MgCl_2 and ADP or ATP- γ -S, strong increases of the melting temperature to 55°C and 60°C were observed, respectively (Figure II.2.1). The optimized assay conditions were used to screen a library of 505 fragments [24, 33] (Supplementary Figure II.2.1) and 16 molecules (Supplementary Figure II.2.2) were identified that reproducibly increase the melting temperature of *Cag α* by 1°C to 4°C , which is the typical range observed for binding fragments (Supplementary Figure II.2.1). Interestingly, incubation of many of these fragments in the presence of MgCl_2 and ADP reduces the melting temperature when compared to MgCl_2 and ADP alone, suggesting that they impact the conformation of *Cag α* in a way that changes binding of the other ligands (Supplementary Figure II.2.3).

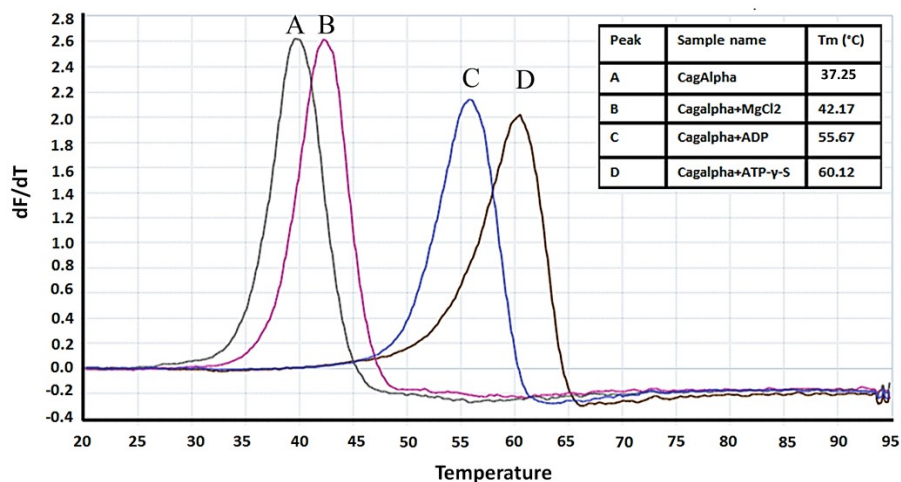
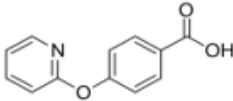
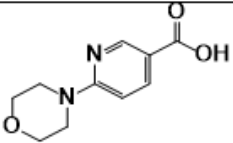
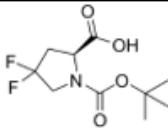
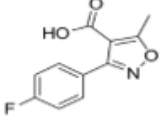
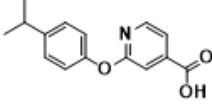
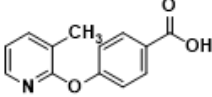
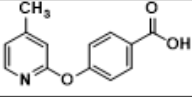
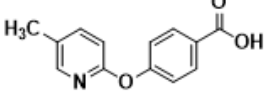


Figure II.2.1. Melting temperature of Cag α in the presence of ligands and cofactors. Melting curves for Cag α were determined using differential scanning fluorimetry (DSF). (A) Cag α apoprotein (green), (B) Cag α and metal cofactor MgCl₂ (pink), (C) Cag α with ADP (blue) and (D) Cag α with ATP- γ -S (black). Table in the upper right corner shows melting temperatures.

II.2.5.2 Effects of binding fragments on the Cag α ATPase activity

We used a Malachite green assay to measure the release of inorganic phosphate from ATP to assess whether the 16 binding fragments impact the enzymatic activity of Cag α . Four of the molecules reduce the ATPase activity and the IC₅₀ values range between 196.2 μ M for molecule 1G2 (Figure II.2.2a) and 4.77 mM in case of molecule 2A5 (Table II.2.1 and Supplementary Figure II.2.4). We used the most potent molecule 1G2 as starting point for a limited structure-activity relationship analysis using six commercially available analogs (Table II.2.1). Two of these molecules (1G2#5 and 1G2#6) do not inhibit the ATPase activity, three of them have higher IC₅₀ values than 1G2 (1G2#1, #2 and #3), but molecule 1G2#4 has a lower IC₅₀ value of 81.9 μ M (Table II.2.1 and Figure II.2.2b). Finally, we tested the mechanism of inhibition by varying the inhibitor concentrations (0 to 500 μ M) and the ATP concentrations (0 to 80 μ M) and fitting of the initial velocity data using nonlinear regression shows that only the V_{max} was affected, whereas the K_m -values remain constant (Figure II.2.2 c, d). These results suggest that we have identified molecules that inhibit the Cag α ATPase activity indirectly via a novel allosteric mechanism. Enzyme kinetic analyses showed that the mechanism of inhibition by molecules 1G2 and 1G2#4 is non-competitive.

Table II.2.1: Structures and IC₅₀ of molecules that inhibit the ATPase activity of Cagα.

Name	Structure	IC ₅₀ values
1G2		196.2 (± 0.026) μM
1G6		0.987 (±0.076) mM
2A5		4.77 (±0.053) mM
1F12		1.85(± 0.068) mM
1G2#1		547.4 (±0.59) μM
1G2#2		619.6 (±0.95) μM
1G2#3		479.6 (±0.46) μM
1G2#4		81.9 (±0.6) μM

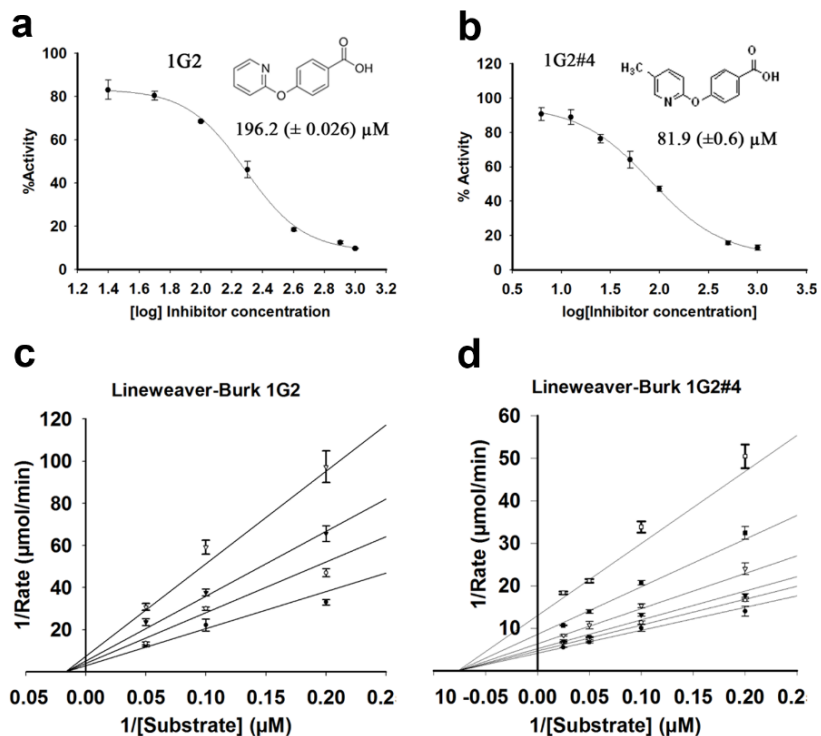


Figure II.2.2. Enzyme Kinetics of Cag α in the presence of molecule 1G2 and 1G2#4. (a, c) Dose response curves of ATPase activity showing IC₅₀ values in the presence of 1G2 and its derivative 1G2#4. (b, d) Lineweaver-Burke plot of Cag α ATPase activity in the presence of 1G2 and 1G2#4. The data were globally fit to a model of non-competitive inhibition. Concentrations varied from 0 to 500 μ M of inhibitors in the presence of 2 mM of MgCl₂.

II.2.5.3 Binding fragments impact the conformation and dissociate the Cag α hexamer

Binding of fragments may impact the conformation and the homo-multimerization of Cag α and we used the homo-bifunctional cross-linking agent disuccinimidyl-suberate (DSS) to obtain insights into the multimerization of the protein. As expected, incubation of Cag α with increasing concentrations of DSS (0-20 μ M), followed by SDS-PAGE and western blot analysis, leads to the successive formation of higher molecular mass forms, which is consistent with the formation of a hexamer (Figure II.2.3a). The cross-linking pattern is similar in the presence of MgCl₂ (Figure II.2.3b), increased amounts of higher molecular mass products are observed in the presence of ADP/MgCl₂ (Figure II.2.3c), but in the presence of ATP- γ -S/MgCl₂ (Figure II.2.3d), and of molecule 1G2 (Figure II.2.3e) a reduced amount of higher molecular mass complexes is observed. Crosslinking assays provide primarily qualitative information and the data suggest

significant changes of the conformation and/or multimerization of Cag α in the presence of ATP- γ -S and 1G2.

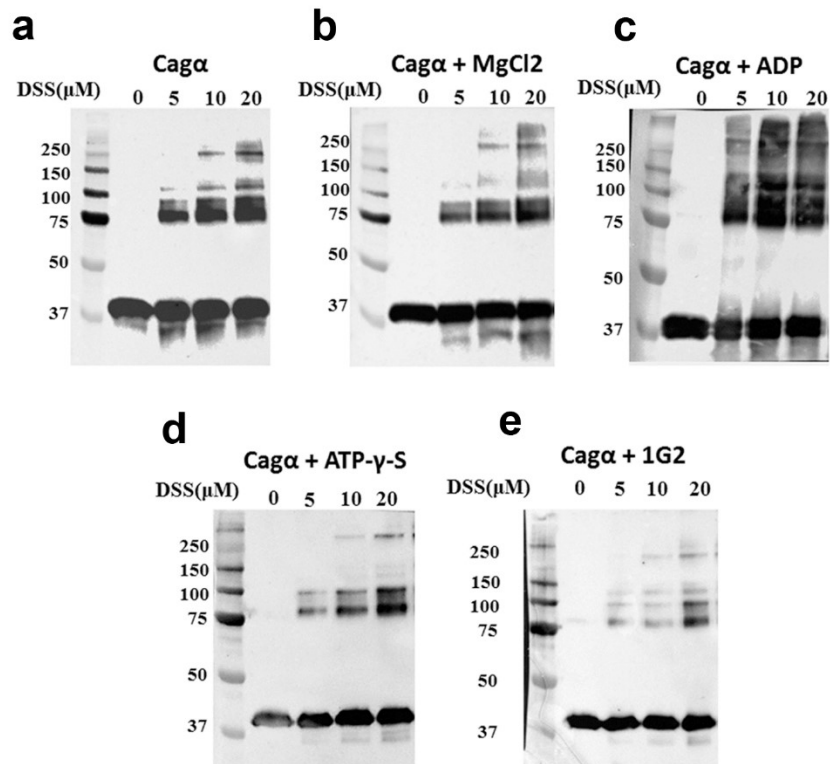


Figure II.2.3. Chemical cross-linking using DSS to study the formation of Cag α oligomers in the presence of ligands. a) Cag α apo protein; b) Cag α with MgCl₂; c) Cag α with ADP and MgCl₂; d) Cag α with ATP- γ -S and MgCl₂; e) Cag α with 1G2 and MgCl₂. The concentrations of DSS varied between 0 and 50 μ M leading to formation of oligomers (indicated by arrows), detection by SDS-PAGE and western blotting using His-tag specific antibodies.

We have performed gel filtration and EM analysis to corroborate our crosslinking experiment result. With gel filtration, Cag α protein homogeneously elutes as a single peak with an elution volume corresponding to a molecular mass of 244 kDa, which is consistent with the formation of a hexamer (Figure II.2.4). The same elution volume was observed in the presence of ATP- γ -S. Interestingly, when Cag α was pre-incubated with molecule 1G2 we observed the elution of two peaks (peak A and peak B in Figure II.2.4) with elution volumes corresponding to apparent molecular masses of 175 kDa and 54 kDa, respectively. These results suggest that incubation with molecule 1G2 dissociates the Cag α hexamer into lower molecular mass species.

Peak	Sample	Elution Volume (ml)	MW(kDa)
Red	Cag α	1.74	244.1
Blue	Cag α +ATP- γ -S	1.74	244.1
Green A	Cag α +1G2	1.83	175.5
Green B	Cag α +1G2	2.14	54

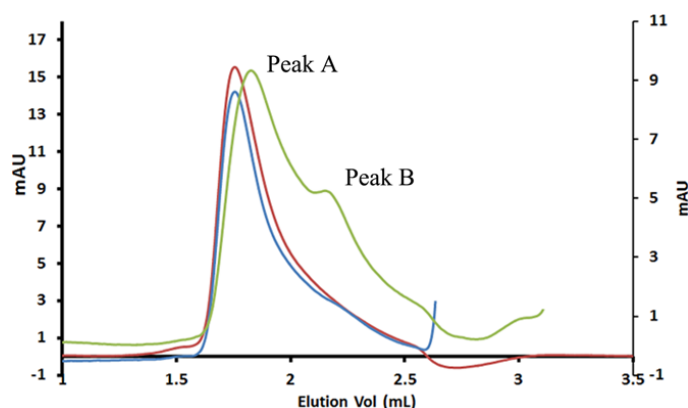


Figure II.2.4. Analytical size exclusion chromatography of Cag α apoprotein and in the presence of ligands. Proteins were separated by gel filtration over a Superdex 200 column. Cag α apoprotein elutes as a hexamer (red curve), elution of Cag α -ATP- γ -S (blue curve) and of two lower molecular mass peaks (A and B) after incubation of Cag α with 1G2 (in green). The molecular masses characterized according to the elution volume are summarized in the table above the graph.

Analysis by negative staining electron microscopy reveals hexamers in the absence of 1G2 (Figure II.2.5a, inset shows a particle of 12 nm diameter), some large rings and lower molecular mass species in the presence of 1G2 in peak A (Figure II.2.5b, inset shows a particle of 7 nm diameter that may be a tetramer) and exclusively lower mass species in the presence of 1G2 in peak B (Figure II.2.5c) confirming this interpretation. Interestingly, gel filtration and EM analysis revealed that binding to 1G2 successively dissociates the Cag α -hexamer. Previously described Cag α inhibitors were believed to bind to the active site and the known competitive inhibitor ATP- γ -S does not dissociate the hexamer. The mechanism of inhibition identified here is therefore novel and it would be interesting to assess the molecular basis of dissociation using approaches that are more sensitive to conformational changes such as NMR or high-resolution cryo-electron microscopy.

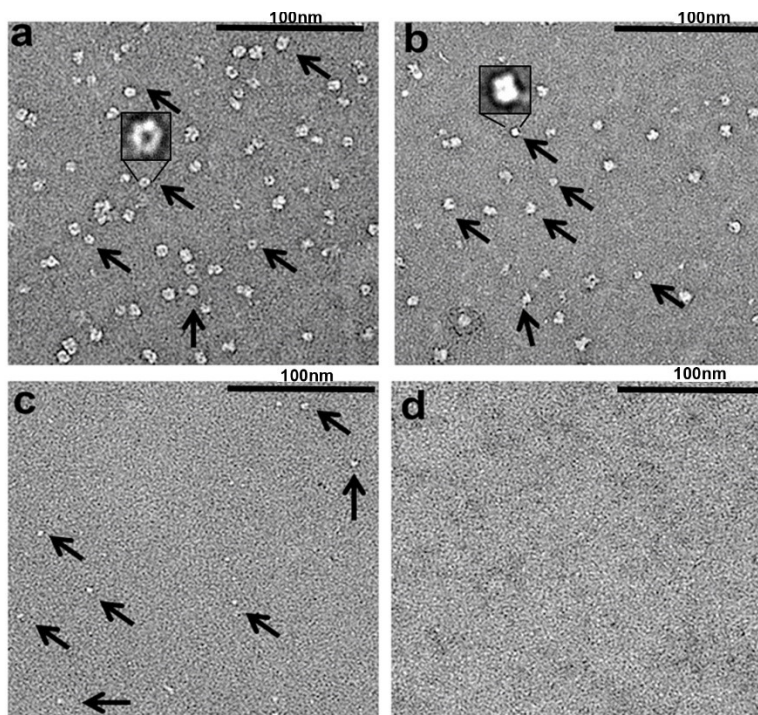


Figure II.2.5. Electron micrographs of negatively stained Cag α apoprotein after gel filtration. Analysis by transmission electron microscopy and negative staining of a) Cag α apoprotein shows a hexameric ring-like structure, insert shows a typical particle; b) peak A of Cag α incubated with 1G2 after elution from the gel-filtration, insert shows a typical smaller particle and c) peak B. d) negative control grid. Arrows show the differently sized complexes and size bars indicate the dimensions.

II.2.5.4 X-ray analysis reveals the 1G2 binding site and conformational changes

To gain further insights into the mechanism of inhibition, we solved the X-ray structure of the Cag α -1G2 complex in the P6322 crystal space group with two molecules in the asymmetric unit (Figure II.2.6a) to a resolution of 2.9 Å (Supplementary table II.2.1). The structure was solved by molecular replacement using ADP bound Cag α (PDB code: 1G6O) as a search model. The overall structure of 1G2-bound Cag α is similar to that of the Cag α -ADP complex, but there are differences of interactions at the protein interface and we identified the electron density of molecule 1G2 sandwiched between two Cag α molecules (Figure II.2.6a). The monomer structure of the Cag α -1G2 complex displays both NTD and CTD with nine α -helices labeled as α 1 to α 9 and 13 β -strands labeled as β 1 to β 13 (Figure II.2.6b). A structural overview from the

top of the NTD reveals that 1G2 interacts with the NTD of both protein subunits (Figure II.2.6c). The 1G2 binding site is distinct from the active site to which ADP and the substrate analog ATP- γ -S bind [16-17]. Molecule 1G2 binds to a hydrophobic pocket created by the interaction between the NTDs of two Cag α subunits and amino acids F68 and F39 make hydrophobic contacts with the two phenyl rings of the inhibitor. R73 and D69 are the amino acids involved in forming a polar contact with 1G2. R73 interacts with the pyridine group of the second phenyl ring via a hydrogen bond and the carboxylic group of 1G2 interacts with the backbone NH group of D69 forming a potential hydrogen bond (Figure II.2.6d). Sequence alignments with other Cag α /VirB11 homologs show that this site is not conserved suggesting that the molecule likely is a specific inhibitor (Supplementary Figure II.2.5). Structural alignment of the Cag α -1G2 complex with Cag α -ADP (PDB code: 1G6O) shows an overall similar structure (RMSD 0.6 Å), but we observe shifts of the β 6 and β 7 sheets and in the linker region between NTD and CTD (Supplementary Figure II.2.6). Alignment of Cag α -1G2 with Cag α apoprotein (PDB code: 1NLZ) reveals slight conformational differences in both the CTD and NTD (RMSD 0.9 Å). The α 8 and α 9 helical region of the CTD as well as the α 1 region of the NTD display changes showing that binding to molecule 1G2 impacts the conformation of the protein (Supplementary Figure II.2.6). Overall, the results indicate that molecule 1G2 binds at the interface between Cag α protein that multimerize via the NTD and this is consistent with the results of enzyme kinetics. However, high-resolution structural information on TrwD and on the potential binding site is not available and this site is distinct from the 1G2 binding site we observe by X-ray crystallography. Analysis of the Cag α -1G2 X-ray structure revealed subtle conformational changes in different parts of the protein including the active site as compared to the apoprotein and its complex with ADP. This may explain the effect of 1G2 binding on the enzymatic activity of Cag α . In contrast, high-resolution structural information on the Cag α homolog TrwD is not available and the potential binding site identified by docking into a model of TraD is distinct from the 1G2 binding site we observe by X-ray crystallography.

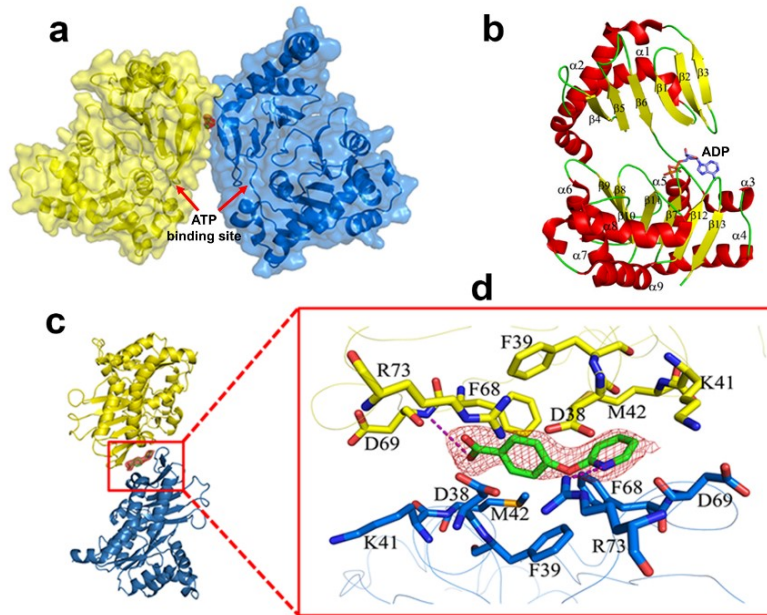


Figure II.2.6. Crystal structure of the Cag α bound to molecule 1G2. a) Cartoon representation of the crystal structure of Cag α crystallized as two molecules in the asymmetric unit. Red map in middle of two subunits represents molecule 1G2 and arrows indicate the ATP/ADP binding site. b) Representation of the monomeric subunit of Cag α in ribbon form: α helices, β strands and loops are represented in yellow, red and green, respectively. The nine helices are labeled as $\alpha 1$ to $\alpha 9$ and the β -strands are labeled as $\beta 1$ to $\beta 13$ and the ADP binding site is indicated. c) Side view of the interaction of two subunits of protein with 1G2 in the middle represented as green stick and red map. d) Enlarged view of 1G2 binding at the interface between two protein subunits. The $2F_o - F_c$ electron density map of 1G2 was contoured at 1.5σ .

II.2.5.5 Molecule 1G2 inhibits the production of interleukin-8 upon binding of *H. pylori* to AGS cells

Finally, we assessed whether molecule 1G2 or its derivatives impact the functionality of the T4SS *in vivo*. To this effect, we tested their impact on the interaction of *H. pylori* strain 26695 with gastric adenocarcinoma (AGS) cells. First, we tested their toxicity and found that molecule 1G2 and derivatives 1G2#1 to #6 have no negative effect on the growth of *H. pylori* on solid agar media at concentrations up to 500 μM (Supplementary Figure II.2.7). Similarly, most molecules do not have negative impact on the viability of AGS cells at concentrations up to 500 μM , showing that they are not toxic (Supplementary Figure II.2.8). We then tested the effects of

these molecules in two commonly used assays for downstream effects of T4SS function: IL-8 production and CagA phosphorylation. When we tested the effects of these molecules at 200 μ M concentration on the production of IL-8 produced by AGS cells upon co-cultivation with *H. pylori*, 1G2 significantly reduces the production of this proinflammatory cytokinin to about 50% of the control (Figure II.2.7). In contrast, derivatives 1G2#1 to #6 have no effect on IL-8 production. Similarly, none of the molecules reduces the tyrosine phosphorylation of translocated CagA, which is generally used as an alternative assay to measure T4SS function (Supplementary Figure II.2.9).

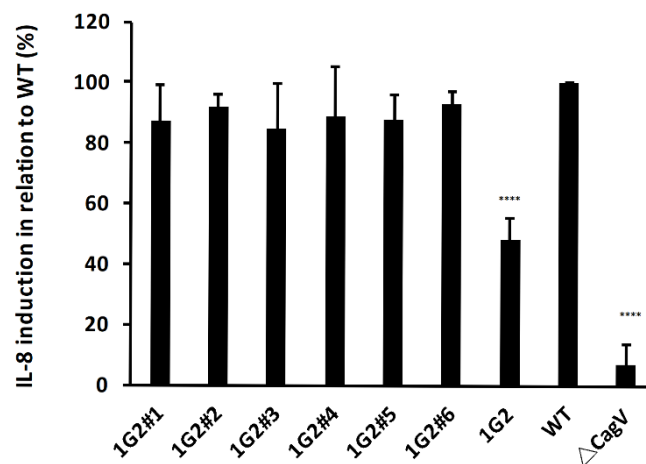


Figure II.2.7. Molecule 1G2 decreases IL-8 induction in co-cultivated AGS cells. *H. pylori* 26695 without and after pre-incubation with 1G2 and its derivatives for 40 min. AGS cells were then co-cultured with *H. pylori* overnight and IL-8 induction was measured by ELISA. The induction of IL-8 by the wild type was calculated as 100% (WT), induction of IL-8 by the Δ cagV strain was used as negative control. The data represent the results from three experiments.

It was therefore somewhat unexpected that molecule 1G2 inhibited IL-8 production to 50% of the control values, but we did not observe an effect on CagA phosphorylation. This observation may be due to the partial inhibition of T4SS function by 1G2 that is more readily quantifiable in the IL-8 production assay as compared to the CagA phosphorylation assay. Enzyme kinetic analysis showed that molecule 1G2#4 was significantly more potent than 1G2 *in vitro*, but repeated crystallisation trials of CagA with 1G2#4 were unsuccessful. Soaking of 1G2#4 into CagA crystals resulted in crystal cracking, suggesting a conformational change in CagA and/or

dissociation of the Cag α quaternary structure. This may be due to the higher potency of molecule 1G2#4 that could be explained by additional hydrophobic contacts of its additional methyl group with amino acid K41 (Supplementary Figure II.2.10). However, this molecule had no effect in the *in vivo* assays, which may be due to its higher hydrophobicity impacting solubility and penetration into cells.

In conclusion, we have here analyzed six derivatives of molecule 1G2 that were commercially available and in future work we will conduct a structure-based structure-activity relationship analysis to synthesize more potent molecules that efficiently penetrate into cells. This approach will enable us to assess whether inhibition of Cag α leads to differential effects on the translocation of effectors CagA, HBP and murein that might also explain the differential effect of molecule 1G2 on IL-8 production and CagA phosphorylation. Potent inhibitors of Cag α could be developed into anti-virulence drugs that are alternative or complementary treatments to currently used triple or quadruple therapy. It would also be interesting to test the specificity of these molecules to assess whether they are narrow or broad-spectrum inhibitors that also impact other T4SS, e.g. bacterial conjugation systems [28, 34].

II.2.6 Acknowledgements

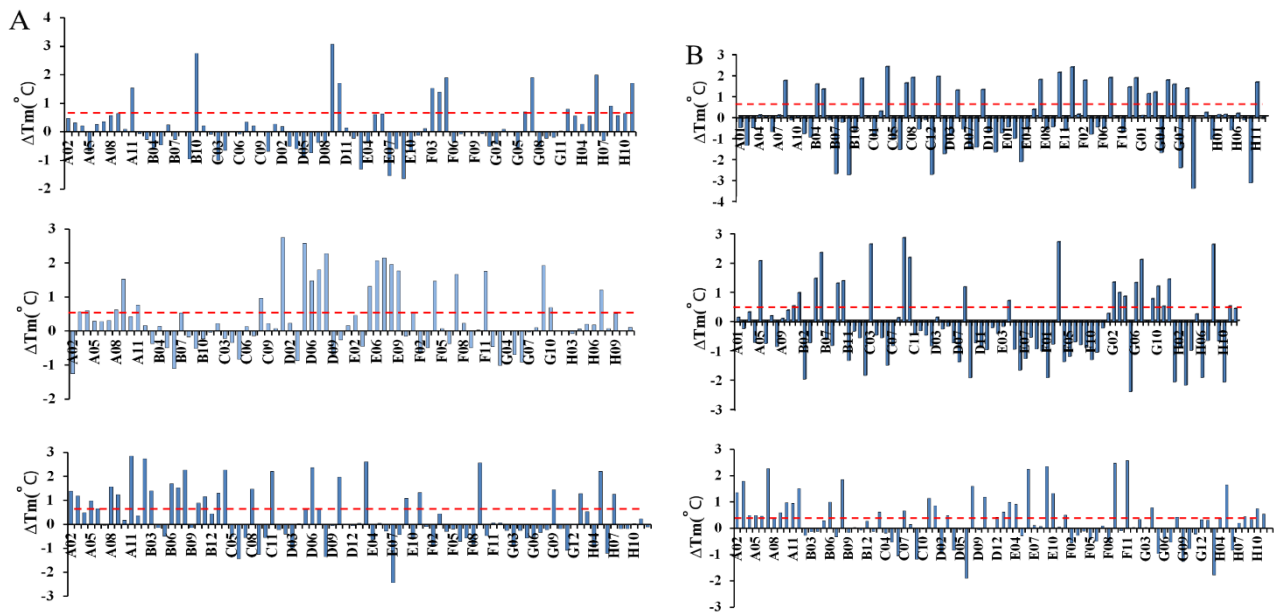
This work was supported by grants to C.B. from the Canadian Institutes of Health Research (CIHR MOP-84239)(<http://www.cihr-irsc.gc.ca/>), the NSERC-CREATE program on the Cellular Dynamics of Macromolecular Complexes (CDMC) (<http://www.nserc-crsng.gc.ca/>), a seed grant from Merck, Sharp and Dohme (<http://www.merck.ca/>), the Canada Foundation for Innovation (CFI)(<https://www.innovation.ca/>) and the Fonds de recherche du Québec-Santé (FRQ- S)(<http://www.frqs.gouv.qc.ca/>). We are grateful to Edward Ruediger and his colleagues at the medicinal chemistry platform at IRIC (Institut de recherche en immunologie et en oncologie (IRIC), Université de Montréal) for support with small molecule screening. We thank Dr. Aleksandr Sverzhinsky from the Department of Biochemistry and Molecular Medicine for helping us with analytical chromatography. We are thankful to Dr. Martin Schmeing and Dr. Asfarul Haque (Department of Biochemistry, McGill University) for helping us with data collection at the Facility of Electron Microscopy Research (FEMR) at McGill

University. Synchrotron X-ray data were collected at the Cornell High Energy Synchrotron Source (CHESS, MacCHESS beamline F1).

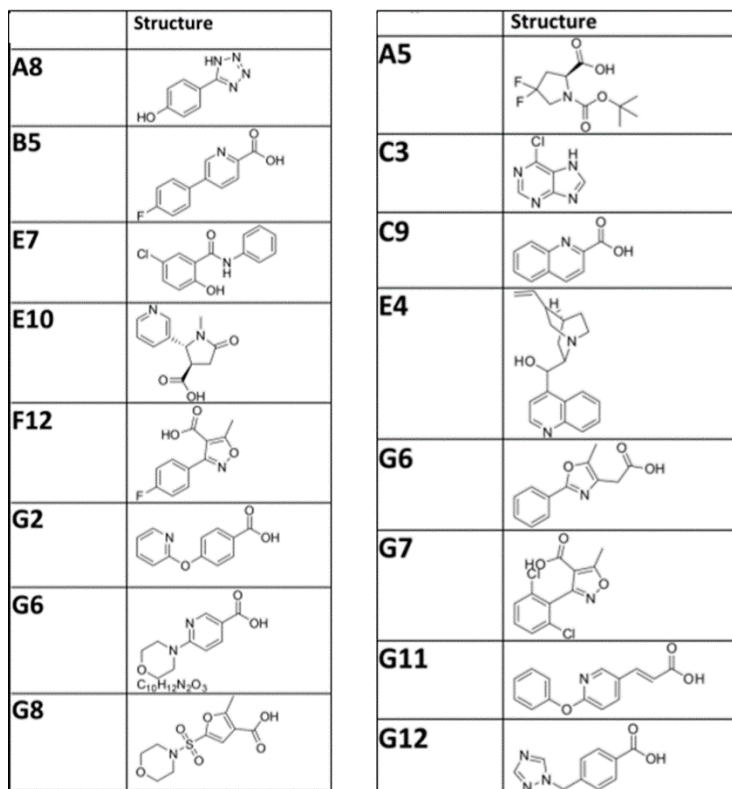
II.2.7 Supplementary content

Supplementary Table SII.2.1: Data collection and refinement statistics.

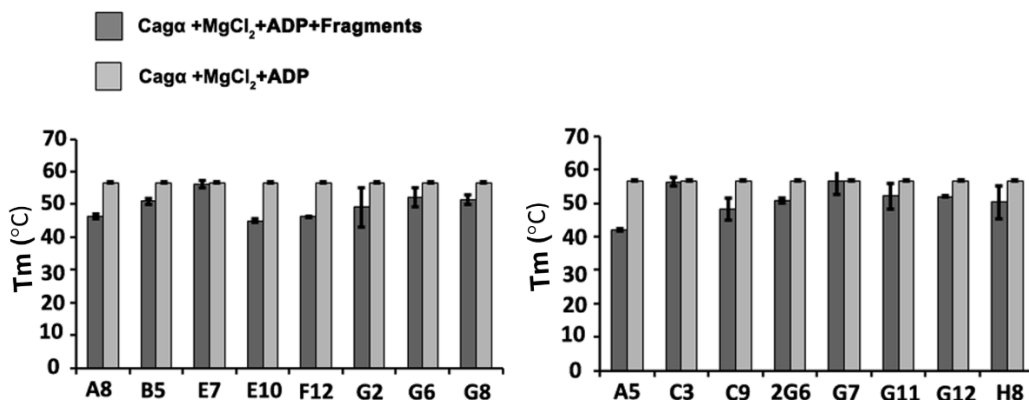
	Cagα-1G2
Resolution range	38.48 - 2.802 (2.902 - 2.802)
Space group	P 63 2 2
Unit cell	112.092 112.092 230.889 90 90 120
Total reflections	197211
Unique reflections	19828 (1836)
Multiplicity	10.1
Completeness (%)	90.44 (87.30)
Mean I/sigma(I)	2.33
Wilson B-factor	44.26
R-merge	0.10
R-pim	0.08
Reflections used in refinement	19787 (1835)
Reflections used for R-free	976 (92)
R-work	0.2570 (0.3280)
R-free	0.3266 (0.3991)
Protein residues	646
RMS(bonds)	0.013
RMS(angles)	1.62
Ramachandran favored (%)	93.15
Ramachandran allowed (%)	5.30
Ramachandran outliers (%)	1.56
Rotamer outliers (%)	6.50
Clashscore	3.87
Average B-factor	46.27
PDB	6BGE



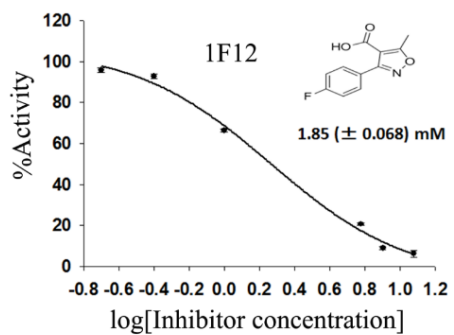
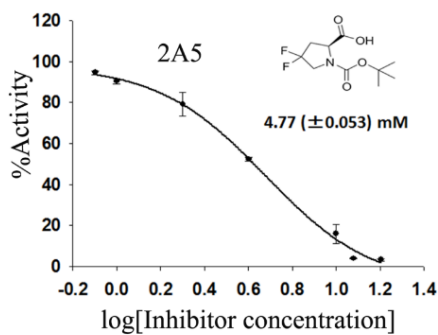
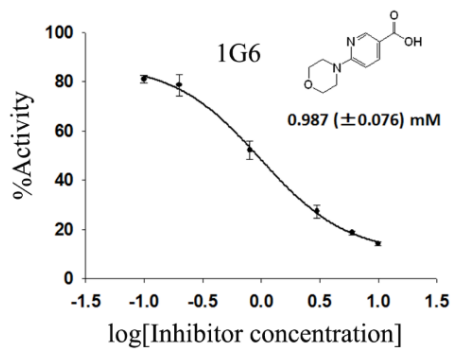
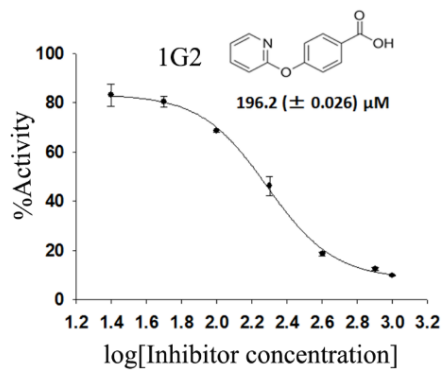
Supplementary Figure SII.2.1. Differential scanning fluorimetry screening of fragment libraries to identify Cag α binding molecules. Melting temperatures of Cag α in the presence of a library of 505 fragments. Differences of T_m between Cag α alone and in the presence of fragments are shown as (ΔT_m). Molecules are considered as stabilizing chosen if they change the T_m value higher than twice the standard deviation of reactions with Cag α alone (red line).



Supplementary Figure SII.2.2. Structures of 16 $Cag\alpha$ -stabilizing molecules identified by differential scanning fluorimetry.



Supplementary Figure SII.2.3. Changes of melting temperatures in the presence ADP and Mg^{2+} after co-incubation with 16 $Cag\alpha$ -stabilizing molecules. Melting temperatures of $Cag\alpha$ were determined in the presence of ADP and Mg^{2+} (light grey). Reduction of melting temperatures when experiments were conducted with ADP, Mg^{2+} and stabilizing fragments (dark grey). The data represent the results from three experiments.



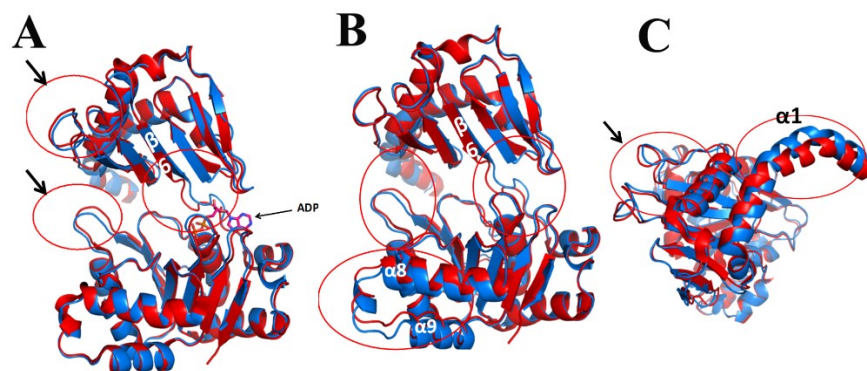
Supplementary Figure SII.2.4. Dose response curves of ATPase activity showing IC_{50} values in the presence of four molecules that inhibit the enzyme activity.

```

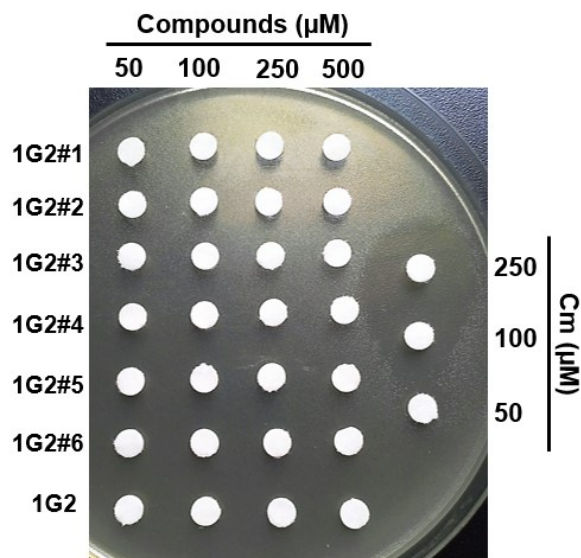
H.pylori_Cag alpha      MTEDRLSAEDKKFLE-VERALKEAALNPLRHATEELPGDPLKMENTITICYNGNKVWVVL 59
A.tumefaciens_VirB11  -----ME---VDPQLRILLKPILEWLDLDFRTEEVAINRRPGEAFVR 37
B.suis_VirB11         -----MMSNRSDFIIVPEAAVK---RAASVNFHLEPLRFLWLDLDFQITEVCVNRPGVEVFC 52
pKM101_traG          -----MTDAAFYQLGGLRREYLEDFTVFIRINCFQVEVICD 35
R388_TrwD            -----MSTVSKASPLSSGNRVG---KDQAVQQLLRPLDRFLDAPEVTELSICRPGVEVWK 52
                        : : * : * :
H.pylori_Cag alpha      KNNGEWQPFVDRDKAFSLRMLHFARCCASFKKKTIIDNYENPILSSNLANGERVQIVLS 119
A.tumefaciens_VirB11  Q-AGAFLKFFLPV-S---YDDLEDIAILAGALRKQDV-GPRNPLCATELPDGERLQICLP 91
B.suis_VirB11         R-ASAWKEYAVPNLD---YEHLISLGTATARFVDQDI-SDSRPVLSAILPMGERIQIVRP 107
pKM101_traG          T-FSGRRVVQNAIT---ADFIRNLAKSLVSS---NK-LTMQAINDVILPGGIRGVICLP 87
R388_TrwD            T-FEGWQVHEVPELT---EPFLQALITAIIVY---NG-VAPKSVNYVVLPGGQRGTTAQA 104
                        : : : * * * *
H.pylori_Cag alpha      PVTVNDETISISIRIPSKTIYPHSPFEEQGFYNLLD----- 155
A.tumefaciens_VirB11  FT-VPSGTVSLTIRRPSRVSLSLKEVSSRYDAPRWNQWKERK-----KRH 135
B.suis_VirB11         PA-CEHGTISVTRKPSFTTRILEDYAQQGFFKHVRPMSK-----SL 148
pKM101_traG          PA-VIDGTTAVAFRKLDAADKNLEQLTSEGIFSDCRKI TGSKQSLTDDDFL----- 138
R388_TrwD            PA-VIDGTLSPFLIRKHELIVKLEELDAEGAFDFTTDVSPNKPSAREANHYLTQDFTRL 163
* . * : : *
H.pylori_Cag alpha      -NKEQA-----ISAIKDGIAIGKNVIVCGGTSSGKTIYIKSIMEFIPKEERIISI 204
A.tumefaciens_VirB11  DQHDEAILRYYDNGDLEAFHACVVGRLTMLLCGPTSSGKTI MSKTLINAI PPQERLITI 195
B.suis_VirB11         TPPEQELLALKEAGDYMSFLRRAVQLERVI VAGETSSGKTI LMKALMQEIPFDQRLITI 208
pKM101_traG          -----KELHSSEKWPAPLQTAVEKRTI VICEGTS SSGKTI VLRALLKSLHKDERVITL 191
R388_TrwD            EPFEVELLKLKRDGTIREFLEKCVLYKRNIIAGKTS SSGKTI FARSLEIKVPPPEERIITI 223
: . : : : * * * * * : : : : : * * * *
H.pylori_Cag alpha      EDTFELVFKHKNYTLQFFGGN-----ITSADCLKSCLMRPDRRIIIGELRSSEAYD 256
A.tumefaciens_VirB11  EDTLELVIPH-ENHVRLLYSKNGA--GLGAVTAEHLLQASLRMRPDRILLGEIRDAAWA 252
B.suis_VirB11         EDVPELFLPDHPNHVHLFYPSEAKEENAPVTAATLLRSLRMRKPTRILLAE LRGGEAYD 268
pKM101_traG          EDVHEVTVDHVVEAVYMMYDAGK---IGPVSATDALRACMRLTPGRIIMTEL RDDAAWD 248
R388_TrwD            EDVHELFLPMHPNRVHMLYGYG----AGRVSADDECLAACMRQSPDRIFLAE LRGNEAWE 278
** . * : . : : : : : : : * : : * * * : * : * : *
H.pylori_Cag alpha      FYNVLCSGHGKTLTTLHAGSSEAFIRLANMSSNSAARNIKFSLIRGFKDLIDMIVHI 316
A.tumefaciens_VirB11  YLSEVVS GHPGSISTIHGANPVQGFKFLSLVKSSAQGASLEDRTLIDMLATAVDVIVPF 312
B.suis_VirB11         FINVAASGHGGSITSCHAGSCELTFFERLALMVLQNRQGRQLPYE IIRRLLYLVDDVVVHV 328
pKM101_traG          YLKALNTGHPGGVMSTHANSARDAFNRI GLLIKATPIGRMLDMSDIMRMLYSTIDVVVHM 308
R388_TrwD            YLNSLNTGHPGSITTTTHANNALQTFERCATLIKKS DVGRQLEMMIKLVLYTTIDVVLFF 338
: . : : * * : : * . : * : : : : : : : : * : : :
H.pylori_Cag alpha      NHH-----RQCDEFYIKHR----- 330
A.tumefaciens_VirB11  RA---HGDIEVGEIWLADARRRGETIGDLLNQQ 344
B.suis_VirB11         HNGVHDGTGSHISEVWYDPNTRKALS-LQHSSEKT- 361
pKM101_traG          E-----RKKIKIYFDPEYKMQCV-NGSL---- 331
R388_TrwD            K-----LKKLVEVFPYDPIFSKSKM-A----- 358
. * :

```

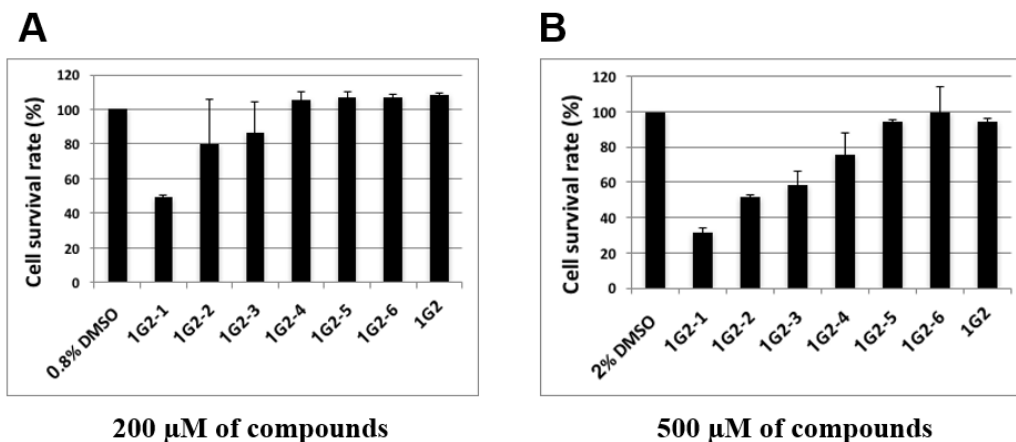
Supplementary Figure SII.2.5. Multiple sequence alignment of Cag α /VirB11 homologs. Selected Cag α homologs from the T4SS of *Agrobacterium tumefaciens* C58, *Brucella suis* 1330, plasmids pKM101 and R388 were aligned using clustal omega online tool (<https://www.ebi.ac.uk/>). The amino acids corresponding to the 1G2 binding site on Cag α are indicated in red and the ATPase active site is labelled green with a box.



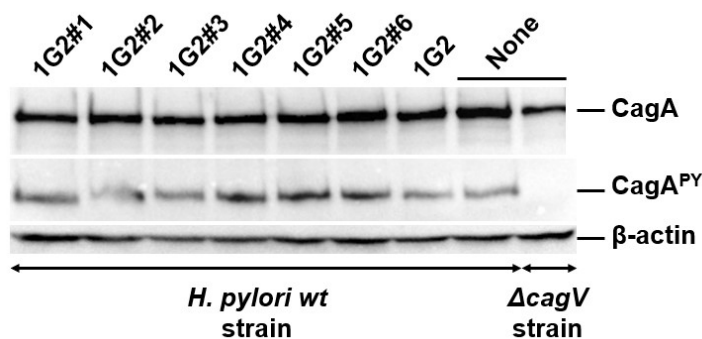
Supplementary Figure SII.2.6. Crystal structures of the Cag α -1G2 complex compared to previously determined structures of Cag α . a) Cag α -1G2 complex (blue) and Cag α -ADP complex (red), PDB ID: 1G6O; b) Cag α -1G2 complex (blue) and Cag α apoprotein structure (red), PDB ID: 1NLZ; c) view from the top of structure, Cag α -1G2 complex (blue) and Cag α apoprotein (red). Arrows and red circles indicate significant structural changes in different parts of the protein.



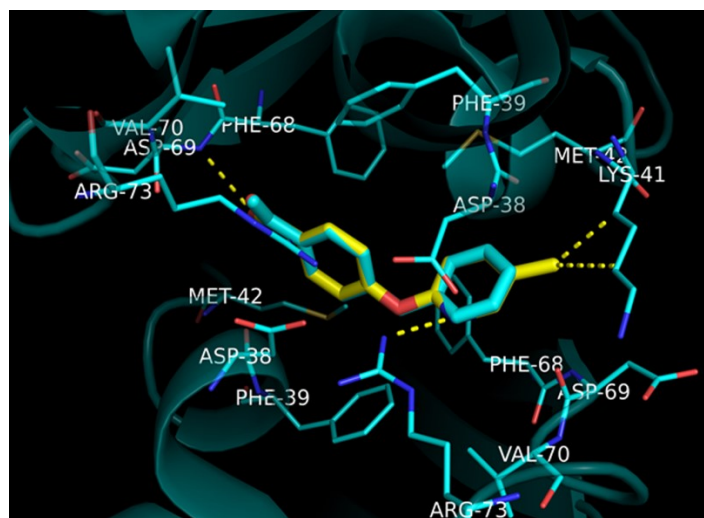
Supplementary Figure SII.2.7. Effects of molecule 1G2 and its derivatives on *H. pylori* growth. Molecule 1G2 and six derivatives (1G2-1 to 1G2-6) were spotted on Whatman paper disks at increasing concentrations (from 50 to 500 μ M) and placed on an agar plate inoculated with *H. pylori* 26695. Several concentrations of chloramphenicol (Cm) were tested as control. Growth was observed after 72 h of incubation, the halo around the Whatman paper disks with Cm indicate the inhibition of growth.



Supplementary Figure SII.2.8. Effects of molecule 1G2 and its derivatives on the viability of AGS cells. a) AGS cells were incubated with 0.8% DMSO or 200 μM of 1G2 and its derivatives for 24h. b) AGS cells were incubated with 2% DMSO or 500 μM of 1G2 and its derivatives for 24h. Cell survival was measured using the Cell Proliferation Reagent WST-1 kit and cell survival in the presence of DMSO was calculated as 100 %.



Supplementary Figure SII.2.9. CagA phosphorylation in the presence of molecule 1G2 and its derivatives. *H. pylori* strain 26692 wild type was pre-incubated with 1G2 and its derivatives, followed by infection of AGS cells (lanes 1-7); *H. pylori* wild type without added compound (lane 8) and *H. pylori* strain 26692 ΔcagV strain (lane 9) were used as controls. Western blotting was performed with anti-CagA (CagA), and anti-phosphotyrosine antibodies (CagA^{PY}), anti-actin antibody was used as loading control.



Supplementary Figure SII.2.10. Overlay of 1G2#4 modeled over the Cag α -1G2 complex. 1G2#4 represented in yellow stick as compare to 1G2 represented in blue color. Structural analysis shows that 1G2#4 may form additional hydrophobic interactions with the carbon chain of lysine.

II.2.8 References

1. Malfertheiner, P.; Link, A.; Selgrad, M., Helicobacter pylori: perspectives and time trends. *Nature reviews. Gastroenterology & hepatology* 2014, 11 (10), 628-638.
2. Herrero, R.; Parsonnet, J.; Greenberg, E. R., Prevention of gastric cancer. *Jama* 2014, 312 (12), 1197-8.
3. Plummer, M.; Franceschi, S.; Vignat, J.; Forman, D.; de Martel, C., Global burden of gastric cancer attributable to Helicobacter pylori. *International journal of cancer* 2015, 136(2), 487-90.
4. Gatta, L.; Vakil, N.; Vaira, D.; Scarpignato, C., Global eradication rates for Helicobacter pylori infection: systematic review and meta-analysis of sequential therapy. *Bmj* 2013, 347, f4587.
5. Ochman, H.; Lawrence, J. G.; Groisman, E. A., Lateral gene transfer and the nature of bacterial innovation. *Nature* 2000, 405 (6784), 299-304.
6. Grohmann, E.; Christie, P. J.; Waksman, G.; Backert, S., Type IV secretion in Gram-negative and Gram-positive bacteria. *Molecular microbiology* 2018, 107 (4), 455-471.
7. Segal, E. D.; Cha, J.; Lo, J.; Falkow, S.; Tompkins, L. S., Altered states: involvement of phosphorylated CagA in the induction of host cellular growth changes by Helicobacter pylori. *Proc Natl Acad Sci U S A* 1999, 96 (25), 14559-64.
8. Backert, S.; Ziska, E.; Brinkmann, V.; Zimny-Arndt, U.; Fauconnier, A.; Jungblut, P. R.; Naumann, M.; Meyer, T. F., Translocation of the Helicobacter pylori CagA protein in gastric epithelial cells by a type IV secretion apparatus. *Cellular microbiology* 2000, 2 (2), 155-64.
9. Fischer, W.; Püls, J.; Buhrdorf, R.; Gebert, B.; Odenbreit, S.; Haas, R., Systematic mutagenesis of the Helicobacter pylori cag pathogenicity island: essential genes for CagA translocation in host cells and induction of interleukin-8. *Mol. Microbiol.* 2001, 42, 1337-1348.
10. Selbach, M.; Moese, S.; Meyer, T. F.; Backert, S., Functional analysis of the Helicobacter pylori cag pathogenicity island reveals both VirD4-CagA-dependent and VirD4-CagA-independent mechanisms. *Infect. Immun.* 2002, 70, 665-671.
11. Tegtmeyer, N.; Neddermann, M.; Asche, C. I.; Backert, S., Subversion of host kinases: a key network in cellular signaling hijacked by Helicobacter pylori CagA. *Molecular microbiology* 2017, 105 (3), 358-372.
12. Naumann, M.; Sokolova, O.; Tegtmeyer, N.; Backert, S., Helicobacter pylori: A Paradigm Pathogen for Subverting Host Cell Signal Transmission. *Trends Microbiol* 2017, 25 (4), 316-328.
13. Gall, A.; Gaudet, R. G.; Gray-Owen, S. D.; Salama, N. R., TIFA Signaling in Gastric Epithelial Cells Initiates the cag Type 4 Secretion System-Dependent Innate Immune Response to Helicobacter pylori Infection. *MBio* 2017, 8 (4).

14. Krause, S.; Barcena, M.; Pansegrau, W.; Lurz, R.; Carazo, J. M.; Lanka, E., Sequence-related protein export NTPases encoded by the conjugative transfer region of RP4 and by the cag pathogenicity island of *Helicobacter pylori* share similar hexameric ring structures. *Proc Natl Acad Sci U S A* 2000, 97 (7), 3067-72.
15. Krause, S.; Pansegrau, W.; Lurz, R.; de la Cruz, F.; Lanka, E., Enzymology of type IV macromolecule secretion systems: the conjugative transfer regions of plasmids RP4 and R388 and the cag pathogenicity island of *Helicobacter pylori* encode structurally and functionally related nucleoside triphosphate hydrolases. *J Bacteriol* 2000, 182 (10), 2761-70.
16. Savvides, S. N.; Yeo, H. J.; Beck, M. R.; Blaesing, F.; Lurz, R.; Lanka, E.; Buhrdorf, R.; Fischer, W.; Haas, R.; Waksman, G., VirB11 ATPases are dynamic hexameric assemblies: new insights into bacterial type IV secretion. *EMBO J.* 2003, 22, 1969-1980.
17. Yeo, H. J.; Savvides, S. N.; Herr, A. B.; Lanka, E.; Waksman, G., Crystal structure of the hexameric traffic ATPase of the *Helicobacter pylori* type IV secretion system. *Mol. Cell* 2000, 6, 1461-1472.
18. Fujita, Y.; Krause, G.; Scheffner, M.; Zechner, D.; Leddy, H. E.; Behrens, J.; Sommer, T.; Birchmeier, W., Hakai, a c-Cbl-like protein, ubiquitinates and induces endocytosis of the E-cadherin complex. *Nat Cell Biol* 2002, 4 (3), 222-31.
19. Hare, S.; Bayliss, R.; Baron, C.; Waksman, G., A large domain swap in the VirB11 ATPase of *Brucella suis* leaves the hexameric assembly intact. *Journal of molecular biology* 2006, 360 (1), 56-66.
20. Brown, E. D.; Wright, G. D., Antibacterial drug discovery in the resistance era. *Nature* 2016, 529 (7586), 336-43.
21. Ruer, S.; Pinotsis, N.; Steadman, D.; Waksman, G.; Remaut, H., Virulence-targeted Antibacterials: Concept, Promise, and Susceptibility to Resistance Mechanisms. *Chem Biol Drug Des* 2015, 86 (4), 379-99.
22. Paschos, A.; den Hartigh, A.; Smith, M. A.; Atluri, V. L.; Sivanesan, D.; Tsolis, R. M.; Baron, C., An In Vivo High-Throughput Screening Approach Targeting the Type IV Secretion System Component VirB8 Identified Inhibitors of *Brucella abortus* 2308 Proliferation. *Infect Immun* 2011, 79 (3), 1033-43.
23. Smith, M. A.; Coincon, M.; Paschos, A.; Jolicoeur, B.; Lavalley, P.; Sygusch, J.; Baron, C., Identification of the Binding Site of *Brucella* VirB8 Interaction Inhibitors. *Chem Biol* 2012, 19 (8), 1041-8.
24. Casu, B.; Arya, T.; Bessette, B.; Baron, C., Fragment-based screening identifies novel targets for inhibitors of conjugative transfer of antimicrobial resistance by plasmid pKM101. *Sci Rep* 2017, 7 (1), 14907.
25. Casu, B.; Smart, J. P.; Hancock, M. A.; Smith, M.; Sygusch, J.; Baron, C., Structural analysis and inhibition of TraE from the pKM101 type IV secretion system. *J Biol. Chem.* 2016, 291, 23817-23829.

26. Shaffer, C. L.; Good, J. A.; Kumar, S.; Krishnan, K. S.; Gaddy, J. A.; Loh, J. T.; Chappell, J.; Almqvist, F.; Cover, T. L.; Hadjifrangiskou, M., Peptidomimetic Small Molecules Disrupt Type IV Secretion System Activity in Diverse Bacterial Pathogens. *MBio* 2016, 7 (2), e00221-16.
27. Ripoll-Rozada, J.; Garcia-Cazorla, Y.; Getino, M.; Machon, C.; Sanabria-Rios, D.; de la Cruz, F.; Cabezon, E.; Arechaga, I., Type IV traffic ATPase TrwD as molecular target to inhibit bacterial conjugation. *Molecular microbiology* 2016, 100 (5), 912-21.
28. Cabezon, E.; de la Cruz, F.; Arechaga, I., Conjugation Inhibitors and Their Potential Use to Prevent Dissemination of Antibiotic Resistance Genes in Bacteria. *Front Microbiol* 2017, 8, 2329.
29. Hilleringmann, M.; Pansegrau, W.; Doyle, M.; Kaufman, S.; MacKichan, M. L.; Gianfaldoni, C.; Ruggiero, P.; Covacci, A., Inhibitors of Helicobacter pylori ATPase Cagalpha block CagA transport and cag virulence. *Microbiology* 2006, 152 (Pt 10), 2919-30.
30. Sayer, J. R.; Wallden, K.; Pesnot, T.; Campbell, F.; Gane, P. J.; Simone, M.; Koss, H.; Buelens, F.; Boyle, T. P.; Selwood, D. L.; Waksman, G.; Tabor, A. B., 2- and 3-substituted imidazo[1,2-a]pyrazines as inhibitors of bacterial type IV secretion. *Bioorg Med Chem* 2014, 22 (22), 6459-70.
31. Mashalidis, E. H.; Sledz, P.; Lang, S.; Abell, C., A three-stage biophysical screening cascade for fragment-based drug discovery. *Nature protocols* 2013, 8 (11), 2309-24.
32. Casu, B.; Mary, C.; Sverzhinsky, A.; Fouillen, A.; Nanci, A.; Baron, C., The VirB8 homolog TraE from the plasmid pKM101 type IV secretion system interacts with TraD and forms a hexameric ring-like pore structure submitted to PNAS (accepted for review) 2018.
33. Sharifahmadian, M.; Arya, T.; Bessette, B.; Lecoq, L.; Ruediger, E.; Omichinski, J. G.; Baron, C., Monomer-to-dimer transition of Brucella suis type IV secretion system component VirB8 induces conformational changes. *The FEBS journal* 2017, 284 (8), 1218-1232.
34. Zambelloni, R.; Marquez, R.; Roe, A. J., Development of antivirulence compounds: a biochemical review. *Chem Biol Drug Des* 2015, 85 (1), 43-55.
35. Fischer, W.; Puls, J.; Buhrdorf, R.; Gebert, B.; Odenbreit, S.; Haas, R., Systematic mutagenesis of the Helicobacter pylori cag pathogenicity island: essential genes for CagA translocation in host cells and induction of interleukin-8. *Mol Microbiol* 2001, 42 (5), 1337-48.
36. Cogan, E. B.; Birrell, G. B.; Griffith, O. H., A robotics-based automated assay for inorganic and organic phosphates. *Anal Biochem* 1999, 271 (1), 29-35.
37. Otwinowski, Z.; Minor, W., Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 1997, 276, 307-26.

38. Murshudov, G. N.; Vagin, A. A.; Dodson, E. J., Refinement of macromolecular structures by the maximum-likelihood method. *Acta crystallographica. Section D, Biological crystallography* 1997, 53 (Pt 3), 240-55.
39. Emsley, P.; Cowtan, K., Coot: model-building tools for molecular graphics. *Acta crystallographica. Section D, Biological crystallography* 2004, 60 (Pt 12 Pt 1), 2126-32.
40. Adams, P. D.; Grosse-Kunstleve, R. W.; Hung, L. W.; Ioerger, T. R.; McCoy, A. J.; Moriarty, N. W.; Read, R. J.; Sacchettini, J. C.; Sauter, N. K.; Terwilliger, T. C., PHENIX: building new software for automated crystallographic structure determination. *Acta crystallographica. Section D, Biological crystallography* 2002, 58 (Pt 11), 1948-54.
41. Yuan, Q.; Carle, A.; Gao, C.; Sivanesan, D.; Aly, K. A.; Hoppner, C.; Krall, L.; Domke, N.; Baron, C., Identification of the VirB4-VirB8-VirB5-VirB2 pilus assembly sequence of type IV secretion systems. *J Biol Chem* 2005, 280 (28), 26349-59.

II.3 Analyse génétique et biochimique de souches d'*H. pylori* résistantes aux antibiotiques et inhibition de la virulence par la molécule 1G2

II.3.1 Présentation de l'article 3 et contribution des coauteurs

L'objet de cette étude consiste en une caractérisation génétique et biochimique de souches d'*H. pylori* présentant différents profils de résistances aux antibiotiques et isolées de patients au Québec, ainsi qu'à l'inhibition de leur virulence par la molécule 1G2 précédemment identifiée. Lors de cette étude, nous nous sommes plus particulièrement intéressés aux protéines encodées par le *cagPAI*. Sept souches cliniques, ainsi que les deux souches contrôles, furent identifiées *cag*-positives. Les motifs EPIYA et CM de CagA furent analysés, mettant en évidence une variabilité dans le domaine C-terminal de CagA. Des variations dans la séquence peptidique de CagL furent également observées entre les différentes souches, ainsi qu'une variabilité dans les niveaux de production des protéines Cag. Nous avons, par la suite, établi que les souches 3793, 3822 et 3930 induisaient les phénotypes les plus marqués sur les cellules humaines AGS, conduisant à des niveaux élevés de sécrétion d'interleukine-8 et à des modifications morphologiques. Finalement, l'efficacité de la molécule 1G2 fut testée sur les trois souches cliniques les plus virulentes (3793, 3822 et 3830), ainsi que sur la souche contrôle 26695, mettant en évidence que la molécule réduisait les niveaux de production d'interleukine-8. Néanmoins une baisse d'efficacité fut observée dans le cas de la souche 3822, qui pourrait s'expliquer du fait d'une variation identifiée au niveau du site de liaison de la molécule 1G2.

Dans le cadre de ce projet, j'ai réalisé l'ensemble des expériences et rédigé le manuscrit. Les souches d'*H. pylori* ont été obtenues par l'intermédiaire du Dre Gaudreau. Toutes les expériences ont été faites sous la supervision du Dr Baron qui a participé à la conception du projet, à l'analyse des résultats, à la correction du manuscrit.

Bien que nous ayons pris le parti de présenter ce travail sous la forme d'un article, ce manuscrit est toujours en préparation. Le travail n'est pas encore achevé, d'où l'absence de certains contrôles et de certaines expériences. Ce travail s'inscrit donc dans un projet à plus long terme.

Article 3

Genetic and biochemical analysis of antibiotic-resistant *H. pylori* strains isolated from patient biopsies shows variabilities among the strains

Flore Oudouhou, Christiane Gaudreau and Christian Baron

Department of Biochemistry and Molecular Medicine, Faculty of Medicine,
Université de Montréal, Québec, Canada

Keywords: *Helicobacter pylori*, *cagPAI*, CagA, CagL, Cag α , interleukin-8, antibiotic resistance, anti-virulence drug

Manuscript in preparation

II.3.2 Abstract

Helicobacter pylori resistance to antibiotics is a growing problem as it increasingly leads to treatment failure. While the bacterium is present worldwide, the severity of the clinical outcomes is highly dependent on the geographical origin and genetic characteristics of the strains. One of the major virulence factors identified in *H. pylori* is the *cag* pathogenicity island (*cagPAI*), which encodes a T4SS used by the bacterium to translocate effectors into human cells. Here, we investigated the genetic and biochemical variability among fifteen *H. pylori* antibiotic resistant strains that were isolated from patient biopsies in Quebec as well as two *H. pylori* control strains. Nine of them carried the *cagPAI*. We showed variability of the sequences of CagA and CagL, two proteins encoded by the *cagPAI*, and production levels among the Cag proteins. Three of the strains (3793, 3822 and 3830) induced the more profound phenotype changes in the host cells as they induce high levels of interleukin-8 secretion and morphological changes of epithelial gastric cells. We finally showed the efficiency of the previously characterized 1G2 molecule on the virulence of those strains.

II.3.3 Introduction

Helicobacter pylori is a human pathogen that colonizes the stomach of about half of the world's population. While the infection usually leads to asymptomatic gastritis, more severe outcomes such as peptic ulceration and adenocarcinoma arise in a subset of subjects [1, 2]. Different treatment regimens are established to eradicate the infection, but due to increasing resistance to antibiotics, failure of *H. pylori* eradication is becoming a growing problem [3]. The most virulent *H. pylori* strains carry the *cag* pathogenicity island (*cagPAI*), a 40 kb chromosomal region, and have higher potential to cause severe gastric diseases [4-6]. The *cagPAI* contains 28-30 genes in different strains that encode a type IV-secretion system (T4SS), a syringe-like macromolecular complex that spans across the double membrane of the bacterium and allows the injection of effectors directly into host cells. Indeed, several molecules were shown to be translocated through the *H. pylori* T4SS, among them the cytotoxin CagA, bacterial peptidoglycan, chromosomal DNA and recently heptose-1,7-bisphosphate (HBP), a metabolic precursor of lipopolysaccharide synthesis [7-10]. While all these effectors are known to

contribute to pro-inflammatory responses in epithelial gastric cells via activation of the NF- κ B transcription factor CagA is recognized as the major virulence factor.

CagA is a 120-145 kDa protein encoded by the *cagPAI*. It has been associated with increased risk of gastric cancer and is established as an oncoprotein [4, 6, 11]. Following bacterial attachment and CagA translocation into the epithelial cells, it can interact with more than 25 host cell proteins and hijacks cellular signaling pathways. CagA induces actin-based cytoskeletal rearrangements, cell motility and proliferation, affects cell polarity, cell adhesion, the cell cycle, the apoptosis pathway and it also induces inflammation [12]. The interaction of CagA with some of its partners depends on its phosphorylation status. Following its translocation, CagA undergoes phosphorylation by Src and Ab1 kinases on tyrosine sites of the EPIYA (Glu-Pro-Ile-Tyr-Ala) motifs that are present in the C-terminus [11, 13-15]. Based on the flanking sequences, four specific EPIYA motifs have been described. The EPIYA-A, -B and -C motifs are usually found in *H. pylori* isolates from Western countries while EPIYA-A, -B and -D are present in East Asian strains [16]. The EPIYA-C site occurs in different numbers, it is typically present in one to three copies and this is directly related to the levels of CagA tyrosine phosphorylation and binding to partners like SHP-2 (Src Homology 2 phosphatase) [16]. A larger number of EPIYA-C motifs and the presence of EPIYA-D are associated with increased virulence and carcinogenicity [11]. Within Western countries, CagA can carry a specific A/T polymorphism, called EPIYT motif, which presents an altered binding activity for phosphatidylinositol 3-kinase (PI3-kinase), influencing cancer risk [17]. In addition, the C-terminal domain includes another repeated sequence motif that was originally designated as the CagA multimerization (CM) motif enabling multimerization via this 16 amino-acids region [18]. Variations of five amino-acid residues are observed between Western and East Asian CagA CM motifs. Whereas Western species contain multiple CM motifs located within each EPIYA-C region plus one distal to the last EPIYA-C, East Asian CagA possesses a single CM motif that is located distally to the EPIYA-D segment.

The injection of effectors into the host cell cytoplasm requires bacterial adhesion to the gastric epithelium. CagL, located at the surface of the pilus, is one of the *cagPAI* encoded proteins proposed to contribute to this process. Indeed, CagL binds to many host integrins, such as the integrins α 5 β 1, α V β 3, α V β 5, α V β 6 and α V β 8, notably via the conserved RGD (Arg-Gly-Asp)

motif [19-21]. Its binding to host integrin can activate the NF- κ B pathway leading to the production of several pro-inflammatory cytokines. Interestingly, it has been shown that several variations in CagL amino-acid residues are associated with higher risk of gastric cancer, e.g. the CagL-Y58E59 variant that induces higher integrin α 5 β 1 expression levels and increases inflammation [22, 23]. Several other variations (Thr30; Asn101; Ala141; Glu142; Asn201; Ile234) are proposed to occur in higher rate in gastric cancer patients but their involvement requires further investigation [23].

In *H. pylori*, pilus biogenesis and translocation of effectors is likely energized by three ATPases [24-27]. Among them, Cag α is a VirB11-like ATPase. VirB11-like proteins are present in all T4SSs and they have been extensively characterized biochemically and structurally [28-32]. Because of its importance for T4SS functions, Cag α is an ideal target for compounds that specifically inhibit the T4SS. In our previous works, we identified and characterized the 1G2 molecule that binds to Cag α and reduces its ATPase activity *in vitro*. We also showed that the 1G2 molecule significantly reduces the production of proinflammatory cytokine secreted by AGS cells in response to *H. pylori* infection, suggesting the strong anti-virulence potential of the molecule [33].

Here, we describe the genetic and biochemical variability of antibiotic-resistant *H. pylori* strains isolated from patient biopsies in Quebec (Canada) and two *H. pylori* control strains. We defined which strains carried the *cagPAI* by targeting the *cagA* gene and found that nine strains were *cag*-positive. We then characterized the CagA EPIYA and CM motifs as well as the amino-acid variations in the CagL protein. Cag protein production levels were also analyzed showing variability among the strains. We then analyzed the morphological changes induced in response to the infection with the *H. pylori* strains and interleukin-8 secretion by epithelial gastric cells. Finally, we evaluated the efficiency of the 1G2 molecule on the virulence of the clinical strains..

II.3.4 Material and Methods

II.3.4.1 Bacterial strains, cell lines and culture conditions.

H. pylori strains 26695 (ATCC700392) and ATCC43504 were used as positive controls, and the Δ *cagV* (*hp0530*) mutant re-created in our laboratory, which has been previously described [24, 33], was used as negative control. The fourteen *H. pylori* clinical strains were isolated from patients at the CHUM (Centre Hospitalier de l'Université de Montréal, QC, Canada) (Supplementary Table SII.3.1). All the strains were cultivated on Columbia agar base (BD) containing 10% (v/v) horse serum (Wisent Inc.), vancomycin (10 µg/ml) and amphotericin B (10 µg/ml). Chloramphenicol (34 µg/ml) was added in case of the Δ *cagV* strain to select for the *cam* gene cassette used to disrupt the gene. For liquid cultures brain heart infusion (BHI) media (Oxoid) was supplemented with 10% fetal bovine serum (FBS) and appropriate antibiotics. Bacteria were cultivated at 37°C under microaerophilic conditions (5% oxygen, 10% CO₂). AGS cells were grown at 37°C in F12K media (Wisent Inc.) with 10% (v/v) FBS (Wisent Inc.) in a 5% CO₂ containing atmosphere.

II.3.4.2 Extraction of the bacterial genomic DNA

The extraction of the bacterial genomic DNA was performed using the GenElute Bacterial Genomic DNA kit (Sigma Aldrich), according to the instructions of the manufacturer.

II.3.4.3 Assessing the levels of Cag proteins produced in *H. pylori* strains

After 3-4 days of culture, bacteria were harvested from plates, washed with PBS and the quantity of bacteria was normalized using the O.D.₆₀₀ by lysis in corresponding amounts of Laemmli sample buffer. The production of proteins was assessed after separation by SDS-PAGE followed by western blotting with mouse polyclonal antiserum raised against CagA, CagL, CagV, Cag α or Cag δ (Abcam).

II.3.4.4 Imaging the AGS cell Hummingbird phenotype by interference contrast microscopy and quantification

AGS cells cultivated at 6×10^5 cells/well density in 6-well plates were infected with overnight cultures of *H. pylori* at a multiplicity of infection of 100:1. After 4 h of incubation under microaerophilic conditions, the media was removed and the wells were washed twice with cold

PBS. AGS cells were fixed with 2.5% glutaraldehyde and visualized with the Nikon Eclipse TE2000U microscope. The band intensities of expressed proteins were quantified using ImageLab software.

II.3.4.5 Assay for measuring IL-8 secretion by AGS epithelial cells

AGS cells cultivated at 6×10^5 cells/well density in 6-well plates were infected with overnight cultures of *H. pylori* at a multiplicity of infection (MOI) of 100:1. For the infection experiments in presence of the 1G2 molecule, and preceding the infection, overnight cultures of *H. pylori* were first pre-incubated with 1G2 for 1h. After 24 h of incubation under microaerophilic conditions, supernatants were sampled and centrifuged (15,000 g), before freezing at -80°C. The level of IL-8 in cell culture supernatants was determined by using a commercially available human IL-8 ELISA kit (Invitrogen).

II.3.5 Results

II.3.5.1 Analysis of the *cagPAI* in *H. pylori* strains from patient biopsies

We studied fourteen *H. pylori* strains from patient biopsies at the CHUM (Centre Hospitalier de l'Université de Montréal) presenting a resistance to clarithromycin and/or metronidazole, and two control strains obtained from the ATCC (26695 and ATCC 43504) (Supplementary Table SII.3.1). After extraction of bacterial genomic DNA, PCR amplification of the *cagA* and *cagL* genes were performed to determine the presence of the *cagPAI* (primers in Supplementary Table SII.3.2). The *vacA* gene, which encodes for the vacuolating cytotoxin and is present in virtually all *H. pylori* strains [34], was used as a control for PCR amplifications to confirm the identity of *H. pylori* and all strains were positive (data not shown). Among the sixteen *H. pylori* strains available in our lab nine strains, including the two controls, were *cagPAI* positive (Table II.3.1) and we observed variations in *cagA* genes size (Figure SII.3.1).

Table II.3.1. *cagPAI* positive and negative *H. pylori* strains

<i>cagPAI</i> positive strains	<i>cagPAI</i> negative strains
26695 (ATCC 700392)	3800
ATCC 43504	3811
3723	3834
3763	3836
3783	3864
3793	3855
3813	3859
3822	
3830	

II.3.5.2 Analysis of the variability of the CagA C-terminal domain

It has been reported that the CagA protein displays C-terminal variability, which involves different types and/or numbers of EPIYA repeat sequences and that diversity in CagA may lead to functional differences and distinct clinical outcomes [11, 17]. The *cagA* PCR products of the *cagPAI* positive strains were sequenced, followed by comparative analysis of the EPIYA motifs, CM types and A/T polymorphisms (Table II.3.2 and Supplementary Figures SII.3.2 and SII.3.3). While most of the strains presented a Western CagA type (EPIYA-ABC) with two CM motifs, one inside the EPIYA-C and one distal to the EPIYA-C segment, the ATCC 43504 strain possesses three repetitions of the EPIYA-C motif (EPIYA-ABCCC) and four CM motifs. Interestingly, the 3793 isolate carries both the Western CM motif (CM^W) (FPLKRHDKVDDLSKVG) and the East Asian CM motif (CM^{EA}) (FPLRRSAKVEDLSKVG). Moreover, two amino-acid variations were also observed in the CM^W motifs of the 3763 strain (FPLKKHDKVGDLSKVG). Only the 3830 strain presented an East Asian type (EPIYA-ABD) with one CM^{EA} motif distal to the EPIYA-C segment. Finally, The A/T polymorphism was observed in the EPIYA-B motifs of the strains 3763, 3783, 3822 and in the control strain 26695.

Table II.3.2. Characteristics of the CagA isoforms isolated from the *H. pylori* strains

<i>H. pylori</i> isolates	EPIYA motif type	EPIYA-B polymorphism	Number of CM motifs	CM motif type	CM motif I sequence	CM motif II sequence
26695	ABC	EPIYT	2	CM ^W	FPLKRHDKVDDLSKVG	FPLKRHDKVDDLSKVG
3723	ABC		2	CM ^W	*	*
3763	ABC	EPIYT	2	CM ^W	FPL <u>KK</u> HDKV <u>G</u> DLSKVG	FPLKRHDKV <u>G</u> DLSKVG
3783	ABC	EPIYT	2	CM ^W	*	*
3793	ABC		2	CM ^W /CM ^{EA}	*	FPL <u>RRS</u> AKV <u>E</u> DLSKVG
3813	ABC		2	CM ^W	FPLKR <u>H</u> SKVDDLSKVG	*
3822	ABC	EPIYT	2	CM ^W	*	*
3830	ABD		1	CM ^{EA}	*	FPL <u>RRS</u> AAV <u>N</u> DLSKVG
ATCC 43504	ABCCC		4	CM ^W	*	*

* : sequence identical to the control strain 26695 ; amino acid variation in the CM sequence are underlined ; CM^W: Western CM motif ; CM^{EA}: East Asian CM motif

II.3.5.3 Sequence variations of CagL

To examine the amino-acid sequences of CagL encoded by the *H. pylori* strains, we PCR-amplified *cagL* followed by sequencing and alignment to the sequence of CagL from *H. pylori* 26695 strain as a reference (Figure II.3.1). The alignment showed that only two strains, 3793 and 3822, did not contain the variation of Glu59, and that all of them contained the variations Asn201 and Ile234, previously reported to occur in higher rate in strains from gastric cancer patients. We also observed other sequence variations for which a correlation with gastric cancer initiation has not been described yet.

```

3793      MKTLVKNTIFSFLLLSVLMAEDITSGLKQLDSTYQETNQQALKNLDEIFSTTSPSANDKM
3822      MKTLVKNTILSFLLLSVLMAEDITSGLKQLDSTYQETNQQVLKNLDEIFSTTSPSANDKM
3723      MKTLMKNTIYSFLLLSVLMAEDITSGLKQLDSTYKETNQQVLKNLDEIFSTTSPSANNEI
43504     MKTLVKNTIMSFLLLSVLMAEDITSGLKQLDSTYKETNQQVLKNLDEIFSTTSPSANNEI
3813      MKTLVKNTIYSFLLLSVLMAEDITSGLKQLDNTYQETNQQVLKNLDEIFSTTSPSANNEI
3763      MKTLVKNTIYSFLLLSVLMAEDITSGLKQLDSTYKETNQQVLKNLDEIFSTTSPSANNEI
3830      MKTLVKNTISSFLLLSVLMAEDITSGLKQLDSTYQETNQQVLKNLDEIFSTTSPSANNEI
3783      MKTLVKNTISSFLLLSVLMAEDITSGLKQLDSTYQETNQQVLKNLDEIFSTTSPSANNEM
26695     MKTLVKNTISSFLLLSVLMAEDITSGLKQLDSTYQETNQQVLKNLDEIFSTTSPSANNEM
          ****:***: *****:*****:***:*****:*****:*****:***:
3793      GEEDALNIKKAAMALRGDLLALLKANFEANELFFISEDVIFKTYMSSPELLLTMYKINPLD
3822      GEEDALNIKKAAMALRGDLLALLKANFEANELFFISEDVIFKTYMSSPELLLTMYKINPLD
3723      GEEDALNIKKAAAIALRGDLLALLKANFEANELFFISEDVIFKTYMSSPELLLTMYKINPLD
43504     GQEEDALNIKKAAAIALRGDLLALLKANFEANELFFISEDVIFKTYMSSPELLLTMYKINPLD
3813      GQEEDALNIKKAAAIALRGDLLALLKANFEANELFFISEDVIFKTYMSSPELLLTMYKINPLD
3763      GQEEDALNIKKAAAIALRGDLLALLKANFEANELFFISEDVIFKTYMSSPELLLTMYKINPLD
3830      GXEDALNIKKAAAIALRGDLLALLKANFEANELFFISEDVIFKTYMSSPELLLTMYKINPLD
3783      GEEDALNIKKAAAIALRGDLLALLKANFEANELFFISEDVIFKTYMSSPELLLTMYKINPLD
26695     GEEDALNIKKAAAIALRGDLLALLKANFEANELFFISEDVIFKTYMSSPELLLTMYKINPLD
          * :*****:*****:*****:*****:*****:*****:*****:*****:
3793      QKTAEQQCISDKILVLYCGGKLEIEQEKQIRERLEASLKAYQSNIGGTASLIIASQTL
3822      QKTAEQQCISDKILVLYCGGKLEIEQEKQIRERLETSLKYQSNIGGTASLIIASQTL
3723      QKTAEQQCISDKVLVLYCEGKLEIEQEKQIRERLETSLKAYQSNIGGTASLIIASQTL
43504     QNTAEQQCISDKVLVLYCEGKLEIEQEKQIRERLETSLKYQSNIGGTASLIIASQTL
3813      QNTAEQQCISDKVLVLYCEGKLEIEQEKQIRERLETSLKAYQSNIGGTASLIIASQTL
3763      QNAAEQQCISDKVLVLYCEGKLEIEQEKQIRERLETSLKAYQSNIGGTASLIIASQTL
3830      QXTAEQQCISDKVLVLYCEGKLEIEQEKQIRERLETSLKAYQSNIGGTASLIIASQTL
3783      QNTAEQQCISDKVLVLYCEGKLEIEQEKQIRERLETSLKAYQSNIGGTASLIIASQTL
26695     QNTAEQQCISDKVLVLYCEGKLEIEQEKQIRERLETSLKAYQSNIGGTASLIIASQTL
          * :*****:*****:*****:*****:*****:*****:*****:*****:
3793      VESLKNKNFIKGIRKLMLAHNKVFLNYLEELDALERSLEQNKRQYLQERQSSKIIVK
3822      VESLKNKNFIKGIRKLMLAHNKVFLNYLEELDALERSLEQSKRQYLQERQSSKIIVK
3723      VESLKNKNXIKGIRKLMLAHNKVFLNYLEELDALERSLEQSKRQYLQERQSSKIIVX
43504     VESLKNKNFIKGIRKLMLAHNKVFLNYLEELDALERSLEQSKRQYLQERQSSKIIVK
3813      VESLKNKNFIKGIRKLMLAHNKIFLNYLEELDALERSLEQSKRQYLQERQSSKIIVK
3763      VESLKNKNFIKGIRKLMLAHNKVFLNYLEELDALERSLEQSKRQYLQERQSSKIIVK
3830      VESLXNKNFIKGIRKLMLAHNKVFLNYLEELDALERSLEQXKRQYLQERQSSKIIVK
3783      VESLKNKNFIKGIRKLMLAHNKVFLNYLEELDALERSLEQSKRQYLQERQSSKIIVK
26695     VESLKNKNFIKGIRKLMLAHNKVFLNYLEELDALERSLEQSKRQYLQERQSSKIIVK
          **** * ** *****:*****:*****:*****:*****:*****:

```

Figure II.3.1. Alignment of the CagL sequences from the *H. pylori* isolates. Variations that occur at a significantly higher rate in patients with gastric cancer are highlighted in red, all the other variations identified within the alignment are highlighted in yellow. The highly conserved RGD motif, involved in the interaction with the host cell integrin is indicated by a red box.

II.3.5.4 Detection of Cag proteins production

We next investigated the levels of several Cag proteins encoded by the *cagPAI*. CagA is the main T4SS-dependent virulence factor and its production levels may influence the severity of the clinical outcomes. While CagL is important for bacterial adhesion to host cells, Cag α is one of the three ATPases proposed to energize the *H. pylori* T4SS, and CagV and Cag δ proteins are both associated to the core complex and required for CagA translocation and the induction of the proinflammatory response [24]. After lysis of the bacteria, the protein production levels were assessed by western blots. Signals corresponding to the sizes of CagA, CagL, Cag α , CagV and Cag δ were detected (Figure II.3.2) and Cag production levels were quantified (Figure SII.3.4). While the experiments showed that CagA, Cag α and Cag δ production levels were quite similar

among the strains, variabilities were observed for CagL and CagV proteins production. The 3813 and 3822 strains produced high level of CagL, similarly to the two control strains (ATCC 43504 and 26695), and 3830 and 26695 control strain produced high level of CagV protein. Surprisingly, Cag proteins were not detected in the strain 3783, previously identified as *cagPAI*-positive.

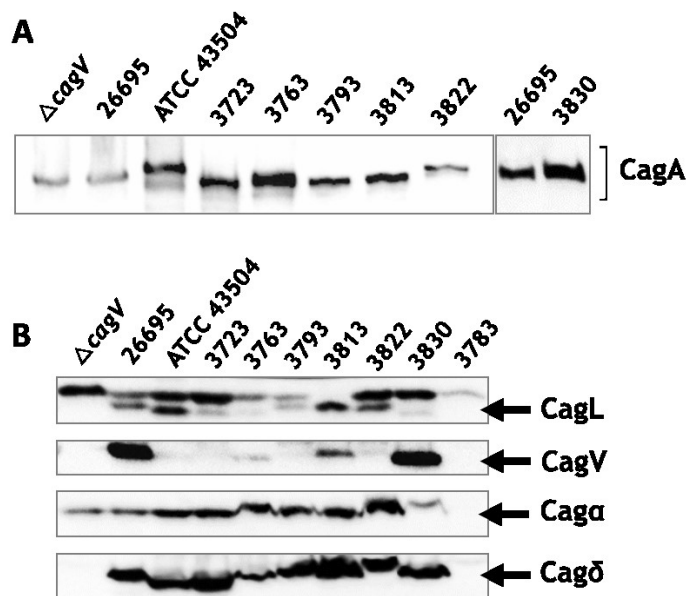


Figure II.3.2. Comparison of the levels of Cag proteins in *H. pylori* strains. Cell lysates of *H. pylori* strains were separated by SDS-PAGE, followed by western blotting with specific antisera. CagA (120-145 kDa), CagL (26 kDa), CagV (29 kDa), Cag α (37 kDa) and Cag δ (51 kDa) proteins were detected using specific antibodies. Equal amounts of total lysates were loaded and Ponceau staining were performed as loading controls for western blots (not shown). Experiments were reproduced 3 times.

II.3.5.5 Effect of infection with *H. pylori* on cytoskeletal changes

CagA translocation in epithelial cells induces dysregulation of the cellular signaling notably leading to the Hummingbird phenotype, characterized by spreading and elongation of the cells [41]. We analyzed the induction of the phenotype by infecting the AGS cells with the *H. pylori* strains, followed by interference contrast microscopy to monitor the cell shape and evaluated the proportion of cells displaying Hummingbird phenotype (Figure II.3.3). As predicted, no morphological changes were observed for the AGS cells alone nor for the cells incubated with

$\Delta cagV$ or 3783 strains. Incubation with positive controls 26695 and ATCC 43504 induced scattering of ~13% and ~14% of the cells, respectively. Similar proportion of cells arboring Hummingbird phenotype was observed when incubated with strain 3723, 3793 and 3813. Infection with the 3763 strain led to a lower number (below 5%) of cells displaying the Hummingbird phenotype. On the other hand, after infection with strain 3822 and 3830, we observed that a larger number of cells (~23% and ~29% respectively) displayed the characteristic morphological changes.

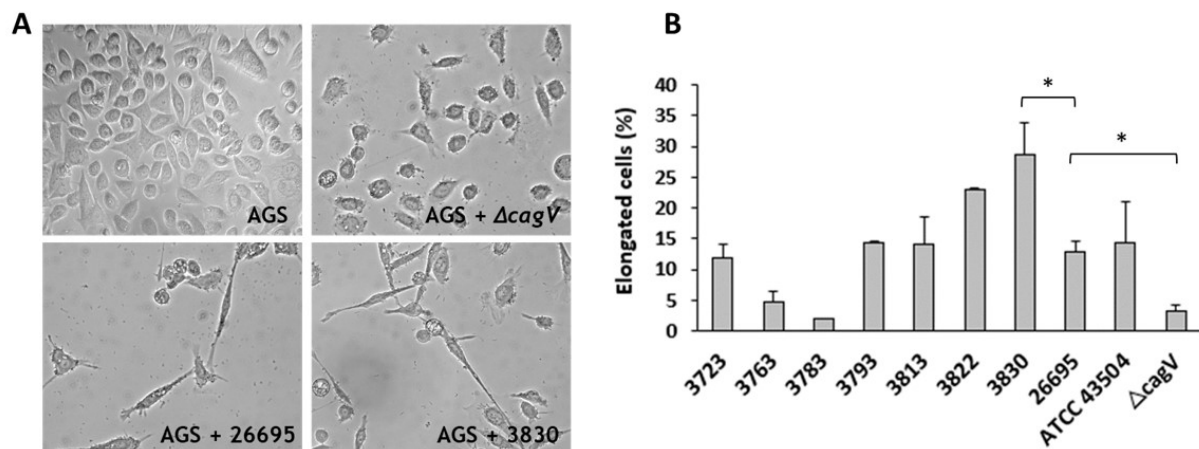


Figure II.3.3. AGS Hummingbird phenotype induced by *H. pylori* infection. (A) AGS cells were infected four hours with *H. pylori* strains at MOI 100. Dead cells were removed by washing and the adherent cells were fixed with 2.5 % glutaraldehyde and observed using interference contrast microscopy. Elongated cells profiles induced by the infection of AGS cells with $\Delta cagV$, 26695 and 3830 strains are illustrated. Each image represents the most representative field for each condition. (B) Elongated cells were quantified and expressed as a percentage of total cells. The data shown represent means and SDs of duplicate experiments with at least 500 cells for each infection condition. *: p<0.1, Student's T-test.

II.3.5.6 Analysis of IL-8 induction after co-cultivation with AGS cells

The infection of AGS cells with *H. pylori* is known to induce pro-inflammatory responses in epithelial gastric cells, leading to the production of chemokines and interleukins [36-40]. We analyzed the levels of IL-8 produced by the AGS cells in response to infection with the *H. pylori* strains. The T4SS-deficient $\Delta cagV$ strain was used as a negative control. We observed that infection with the 3783 strain did not lead to significant IL-8 secretion by the AGS cells and that the 3723 and 3763 strains induced lower levels than the *H. pylori* 26695 positive control. However, the 3793, 3822 and 3830 strains induce interleukin-8 production at equal amounts as *H. pylori* 26695 and at higher amounts than the control strain ATCC 43504 (Figure II.3.4).

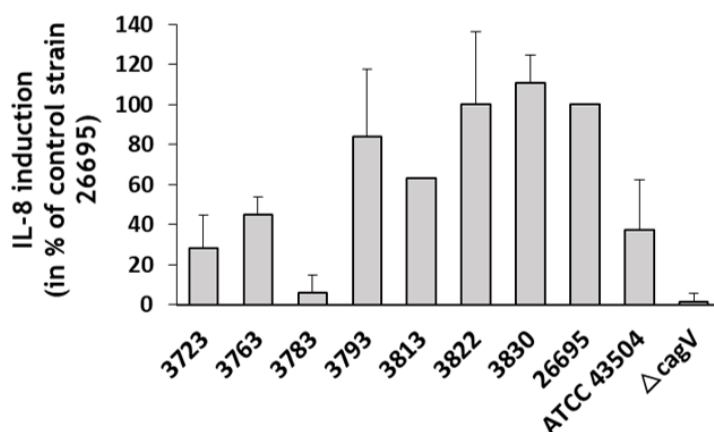


Figure II.3.4. Interleukin-8 production by AGS cells infected by the *H. pylori* strains. The AGS cells were infected for 24 h with the *H. pylori* isolates. The amounts of secreted IL-8 were measured by ELISA. The results are expressed as percent of those observed for the *H. pylori* 26695 control (n=4).

II.3.5.7 The anti-virulence 1G2 molecule is efficient on the clinical *H. pylori* strains

We previously described the 1G2 molecule that inhibits Cag α ATPase activity and significantly reduces the pro-inflammatory response induced by the infection with *H. pylori*, suggesting the strong anti-virulence potential of the molecule [33]. We thus undertook to evaluate its efficacy on highly virulent strains (3793, 3822 and 3830) displaying resistance to antibiotics. The 1G2 molecule was tested at 200 μ M concentration on the induction of IL-8 produced by AGS cells

upon co-cultivation with the *H. pylori* strains. We observed that the molecule significantly decreases the production of the pro-inflammatory cytokinin when AGS cells were infected with 3793, 3830 and 26695 control strains. However, a less drastic effect was observed on the pro-inflammatory response upon infection with the 3822 strain (Figure II.3.5).

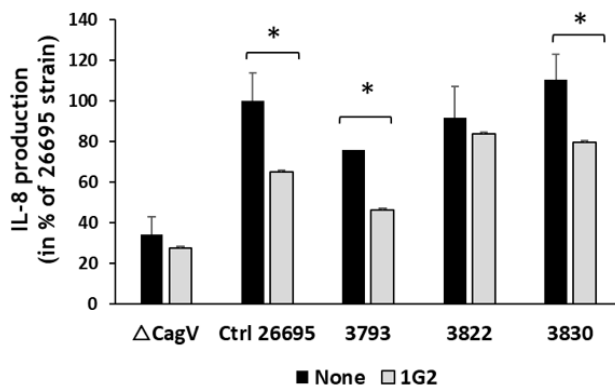


Figure II.3.5. Molecule 1G2 decreases interleukin-8 production by AGS cells upon co-cultivation with the *H. pylori* clinical strains. *H. pylori* strains were pre-incubated without or with 200 μ M of 1G2 molecule for 1h. AGS cells were then co-cultured with *H. pylori* strains overnight and IL-8 induction was measured by ELISA. The induction of IL-8 by the 26695 control strain incubated without 1G2 molecule was calculated as 100%. The $\Delta cagV$ strain was used as negative control. (n=2). *: p<0.1, Student's T-test.

II.3.5.8 Variation in the Cag α peptide sequence of the 3822 strain reduced 1G2 molecule efficiency

Molecule 1G2 binds to a hydrophobic pocket inside the Cag α protein structure created by the interaction between its N-terminal domains. Amino acids F68 and F39 make hydrophobic contacts with the two phenyl rings of the inhibitor, and R73 and D69 are the amino acids involved in forming a polar contact with 1G2 [33]. We hypothesized that the reduced effectiveness of the molecule 1G2 observed in case of infection with 3822 strain may be explained by a sequence variability in the Cag α isoform encoded by this particular strain. Sequence alignment of the Cag α proteins from 3822 and 26695 strains showed an amino-acid variation at position 73 in 3822 strain, from R73 to K73 (Figure II.3.6). As our previous work suggested that R73 interacts with the pyridine group of the second phenyl ring of 1G2

molecule, the residue variation observed in 3822 strain may lead to a lower affinity binding of the molecule to Cag α . This would explain the reduced efficacy of 1G2 observed on the pro-inflammatory response when 3822 strain was co-cultivated with AGS cells in presence of the molecule.

```

26695  MTEDRLSAEDKKFLEVERALKEAALNPLRHATEELFGDFIKMENITEICYNGNKVWVWLK
3822   MTEDRLSAEDKKFLEVERALKEAALNPLRHATEELFGDFIKMENITEICYNGNKVWVWLK
*****
26695  NNGEWQPFVDRDKAFSLRSLMHFARCCASFKKKTIDNYENPILSSNLANGERVQIVLSP
3822   NNGEWQPFVDRDKAFSLRSLMHFARCCASFKKKTIDNYENPILSSNLANGERVQIVLSP
*****
26695  VTVNDETISISIRIPSKTTYPHSFEEQGFYNLLDNKEQAISAIKDGIAIGKNVIVCGGT
3822   VTVNDETISISIRIPSKTTYPHSFEEQGFYNLLDNKEQAISAIKDGIAIGKNVIVCGGT
*****
26695  GSGKTTYIKSIMEFIPKEERIISIEDTEEIVFKHHKNYQLFFGGNITSADCLKSCLMR
3822   GSGKTTYIKSIMEFIPKEERIISIEDTEEIVFKHHKNYQLFFGGNITSADCLKSCLMR
*****
26695  PDRIILGELRSSEAYDFYNVLCSGHKGLTTLHAGSSEAFIRLANMSSNSAARNIKFE
3822   PDRIILGELRSSEAYDFYNVLCSGHKGLTTLHAGSSEAFIRLANMSSNSAARNIKFE
*****
26695  SLIEGFKDLIDMIVHINHKKQDEFYIKHR
3822   SLIEGFKDLIDIVHINHKKQDEFYIKHR
*****

```

Figure II.3.6. Sequence alignment of the Cag α isoforms from 26695 and 3822 strains shows variation in the 1G2 molecule binding site. The 1G2 molecule binding site is underlined in the red box. The K73 variation identified in the 3822 strain is indicated by a red arrow.

II.3.6 Discussion

The presence of a functional *cag* pathogenicity island (*cagPAI*) in *H. pylori* strains has been demonstrated to increase the risk for the development of more severe forms of gastric diseases [4-6, 11]. While the multidrug resistance in *H. pylori* strains is rising, it has become a particularly serious challenge to eradicate the bacterium. In this context, anti-virulence drugs, which focuses on disarming pathogenic bacteria by neutralizing their virulence factors, have emerge as therapeutic alternatives.

Here, we characterized *H. pylori* antibiotic resistant strains isolated from patient biopsies in Quebec, for whom antibiotic treatment has failed, and undertook to evaluate the effect of our previously identified 1G2 molecule on the inhibition of the pathogen virulence. We determined that nine out of the sixteen strains tested carried the *cagA* and *cagL* genes and were thus *cagPAI*

positive and investigated the virulence phenotype of those strains. Surprisingly, we observed that the *cagPAI* positive strain 3783 did not produce Cag proteins nor induce IL-8 production or Hummingbird phenotype, two commonly used assays for downstream T4SS function and *H. pylori* virulence. In contrast, the strains 3793, 3822 and 3830 induce a strong Hummingbird phenotype in AGS cells and high amounts of IL-8 production.

Analysis of the CagA C-terminal domain showed variability among the *H. pylori* strains. The 3830 strain carries an East Asian isoform of the CagA protein. Clinical studies have revealed that gastric cancer is more closely associated with East Asian CagA-producing *H. pylori* strains than with Western CagA-producing *H. pylori* strains. Those observations are in adequacy with our work that shows 3830 strain induces strong morphological changes and high pro-inflammatory response in AGS cells. Interestingly, we also showed that the 3793 strain carries both CM motifs (CM^W/CM^{EA}). The CM motif has been demonstrated to serve as a binding site for PAR1b, which interaction is highly influenced by the nature of the motif and its numbers of repetition [44]. Notably, the presence of CM^{EA} motif induces higher affinity to PAR1b. As PAR1b interaction to CagA leads to the disruption of cellular tight junctions, loss of polarity and microtubule disruption [45, 46], we suggest that the presence of the CM motifs combination may influence the virulence of the 3793 strain.

While the CagA protein is known to be one of the major virulence factors hijacking cellular signaling [12], bacterial adhesion to its target, notably mediated by CagL, induces the secretion of interleukin-8 from the gastric cells [42, 43]. While we hypothesized that diversity among CagL isoforms may influence the pathogenicity potential of the *H. pylori* strains, we could not make a correlation between the variations in CagL amino-acid residues and the phenotype profiles induced on the AGS cells. Notably, whereas the CagL-Glu59 variant is thought to be associated with more severe gastric diseases [22, 23, 35], we noticed that 3793 and 3822 strains were the only two strains that did not carry this variation. Cag proteins production levels were also monitored showing variabilities between the *H. pylori* strains. In particular, production levels of CagL and CagV proteins were variable and we observed that strain 3822 produced high level of CagL. Although accumulating evidence clearly points to a central role of CagA in gastric carcinogenesis, our work suggests that the contribution of other Cag proteins and *H.*

pylori virulence factors (e. g. VacA and outer membrane proteins OMPs) should be considered when evaluating the pathogenicity profiles of *H. pylori* strains.

Finally, we tested the anti-virulence potential of the previously characterized 1G2 molecule on the antibiotic resistant *H. pylori* strains. We showed that the molecule significantly decreases the pro-inflammatory response induced by *H. pylori* infection and its efficacy on both Western and East Asian strains. Chronic inflammation provoked upon *H. pylori* colonization is a major risk for the development of gastric cancer. Thus, the molecule 1G2, that could be used as a precursor for the synthesis of more potent molecules, has a strong therapeutic interest for the treatment of *H. pylori* infection.

II.3.7 Supplementary data

Table SII.3.1. *H. pylori* strains and resistance to antibiotics

<i>H. pylori</i> strains	Resistance	
	Metronidazole	Clarithromycin
3723	R	R
3763	R	R
3783	S	R
3793	R	R
3800	R	S
3811	R	R
3813	R	R
3822	R	R
3830	S	S
3834	R	R
3836	R	R
3846	S	S
3847	S	R
3855	R	S
3858	<i>und.</i>	<i>und.</i>
ATCC 43504	R	S
26695	S	S

R: resistance; S; sensitive; *und.* : undetermined

Table SII.3.2. Primers used for PCR amplification and sequencing

vacA-F1	ATGGAAATACAACAAACACACCGCAA
vacA-R1	CATTTTACTGCTAGTAATCCCTTCTGAA
cagL-F1	AATCTTGAAAAAAAAGCCGAC
cagL-R1	TAGCGTCATTAATCAAATAG
cagA-F1	TCAAATACACCAACGCCTCCAA
cagA-R1	CACATTATGCGCAACTATCTTATCA
cag α -F1	GAGCAAAAATAAGACTTATCTCACTTCTTT
cag α -R1	TAACCCTATGAAACTAAAAATAACAACCAA

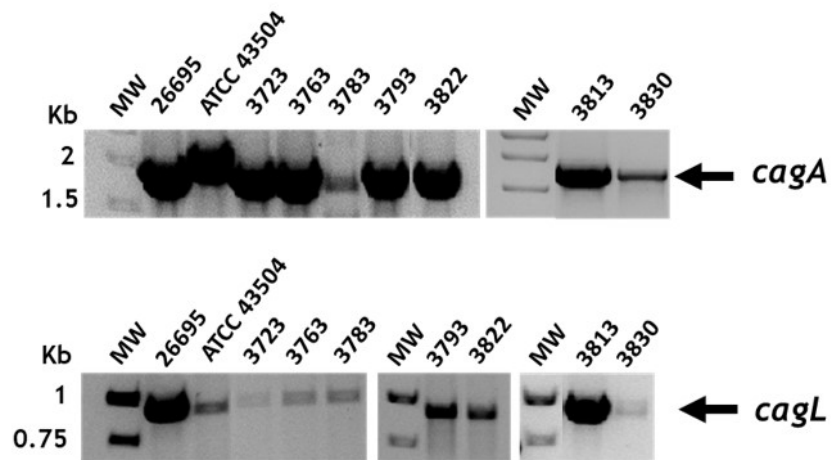


Figure SII.3.1. PCR amplification of *cagA* and *cagL* genes in *H. pylori* isolates. *cagA* and *cagL* genes PCR amplification were performed on genomic DNA isolated from the *H. pylori* strains followed by analysis on agarose gel. Kb : Kilobases ; MW : Molecular weight

```

3830      GLKXN-----EPIIYAQVNKKKAGQATSPEEPIYAQVAKKVS AKIDQLNEAT-SAINRKI
3763      GLKN-----EPIIYAKVNKKKTGQAASLEPIIYTQVAKKVN AKIDRLNQIA-SGLGDVG
3793      GLKNGKDKGPEEPIIYAQVNKKKTGQVASPEEPIYAQVAKKVTKKIDQLNQ AATSGFGGVG
3723      GLKNS-----TEPIIYAKVNKKKTGQAASPEEPIYAQVAKKVTA KIDQLNQAA-SGFGGVG
3813      GLKN-----EPIIYAQVNKKKTGQVASPEEPIYAQVAKKVN AKIDRLNQIA-SGLGSVG
3822      GLKXS-----TEPIIYAKVNKKKXGQVASPEEPIYTQVAKKVN AKIDRLNQIA-SGLGGVG
3783      GLKXS-----TEPIIYAKVNKKKTGQVASPEEPIYTQVAKKVN AKIDRLNQIA-SGLGGVG
26695     GLKNS-----TEPIIYAKVNKKKTGQVASPEEPIYTQVAKKVN AKIDRLNQIA-SGLGGVG
          ****:***** **.:* *****:*****. ***.**: : *...

3830      DRINKIASAGKVDGFSGAGRSASFEPIYATIDFDETNQAG-FPLR RSAAVNDLSKVGLS
3763      QATGFPLKHKDKVGDLSKVGLSASFEPIYATIE----DLGGPFPLKR HDKVDLSKVGLS
3793      Q-AGFPLKRHDKVDDLSKVG RSVSFEPIYATID----DLGGSFPLR SAKVEDLSKVGLS
3723      Q-AGFPLKRHDKVDDLSKVG RSVSFEPIYATID----DLGGPFPLKR HDKVDDLSKVGLS
3813      QAAGFPLKRHSKVDDLSKVG RSVSFEPIYATID----DLGGPFPLKR HDKVDDLSKVGLS
3822      QAAGFPLKRHDKVDDLSKVG RSVSFEPIYATID----DLGGPFPLKR HDKVDDLSKVGLS
3783      QAAGFPLKRHDKVDDLSKVG L SASFEPIYATID----DLGGPFPLKR HDKVDDLSKVG R
26695     QAAGFPLKRHDKVDDLSKVG L SASFEPIYATID----DLGGPFPLKR HDKVDDLSKVG R
          : . . . *...* * * *****: : * **.* * ***** *

```

Figure SII.3.2. Alignment of the sequences of CagA comprising the 3 EPIYA motifs. The EPIYA motifs are underlined in yellow boxes and CM motifs are underlined in red boxes, with variations inside the sequence highlighted in the same colors.

```

26695     GLKNSTEPIIYAKVNKKKTGQVASPEEPIYTQVAKKVN AKIDRLNQIASGLGGVQQAAG
ATCC43504 -MDSKTEPIIYAKVNKKKAGQVASPEEPIYAQVAKKVN AKIDRLNQ AASGLGGVQ-AG
          :...*****:*****:***** ***** **

26695     FPLKRHDKVDDLSKVG L SASFEPIYATIDDLGGPFPLKR HDKVDDLSKVG R-----
ATCC43504 FPLKRHDKVDDLSKVG RSVSFEPIYATIDDLGGPFPLKR HDKVDDLSKVG RSVSFEPIYA
          ***** * *****

26695     -----S
ATCC43504 TIDDLGGPFPLKRHDKVDDLSKVG RSVSFEPIYATIDDLGGPFPLKR HDKVDDLSKVG L
          *

```

Figure SII.3.3. Alignment between the sequences of CagA isoforms from *H. pylori* 26695 and ATCC 43504 strains. The EPIYA motifs are underlined in yellow boxes and CM motifs are underlined in red boxes. Alignment of the two CagA isoforms shows the presence of two extra EPYIA and CM motifs in the sequence of the ATCC 43504 strain.

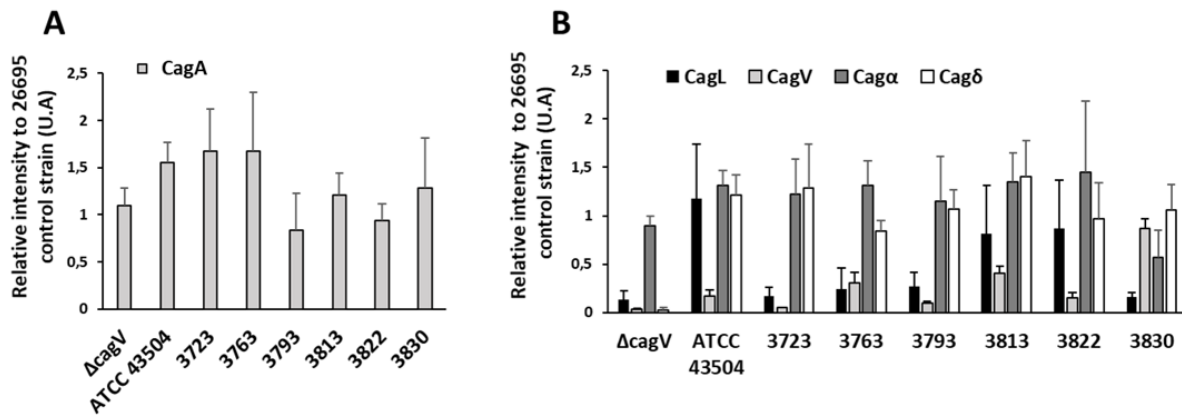


Figure SII.3.4. Comparison of production levels of Cag proteins in the *H. pylori* strains. (A) Comparison of the CagA production levels. (B) Comparison of the production levels of CagL, CagV, Cag α and Cag δ proteins in the *H. pylori* strains. Protein production levels were evaluated from western blot followed by quantification of band intensities of expressed proteins, using ImageLab software. Protein production levels are expressed as relative intensities to the 26695 control strain (n=3).

II.3.8 References

1. Covacci, A., et al., Helicobacter pylori virulence and genetic geography. *Science*, 1999. 284(5418): p. 1328-33.
2. Cover, T.L. and M.J. Blaser, Helicobacter pylori factors associated with disease. *Gastroenterology*, 1999. 117(1): p. 257-61.
3. Gatta, L., et al., Global eradication rates for Helicobacter pylori infection: systematic review and meta-analysis of sequential therapy. *BMJ*, 2013. 347: p. f4587.
4. Cover, T.L., Helicobacter pylori Diversity and Gastric Cancer Risk. *MBio*, 2016. 7(1): p. e01869-15.
5. Yong, X., et al., Helicobacter pylori virulence factor CagA promotes tumorigenesis of gastric cancer via multiple signaling pathways. *Cell Commun Signal*, 2015. 13: p. 30.
6. Hatakeyama, M., Helicobacter pylori CagA and gastric cancer: a paradigm for hit-and-run carcinogenesis. *Cell Host Microbe*, 2014. 15(3): p. 306-16.
7. Odenbreit, S., et al., Translocation of Helicobacter pylori CagA into gastric epithelial cells by type IV secretion. *Science*, 2000. 287(5457): p. 1497-500.
8. Viala, J., et al., Nod1 responds to peptidoglycan delivered by the Helicobacter pylori cag pathogenicity island. *Nat Immunol*, 2004. 5(11): p. 1166-74.
9. Varga, M.G., et al., Pathogenic Helicobacter pylori strains translocate DNA and activate TLR9 via the cancer-associated cag type IV secretion system. *Oncogene*, 2016. 35(48): p. 6262-6269.
10. Gall, A., et al., TIFA Signaling in Gastric Epithelial Cells Initiates the cag Type 4 Secretion System-Dependent Innate Immune Response to Helicobacter pylori Infection. *MBio*, 2017. 8(4).
11. Hatakeyama, M. and H. Higashi, Helicobacter pylori CagA: a new paradigm for bacterial carcinogenesis. *Cancer Sci*, 2005. 96(12): p. 835-43.
12. Tegtmeier, N., et al., Subversion of host kinases: a key network in cellular signaling hijacked by Helicobacter pylori CagA. *Mol Microbiol*, 2017. 105(3): p. 358-372.
13. Backert, S., et al., Phosphorylation of tyrosine 972 of the Helicobacter pylori CagA protein is essential for induction of a scattering phenotype in gastric epithelial cells. *Mol Microbiol*, 2001. 42(3): p. 631-44.
14. Selbach, M., et al., Src is the kinase of the Helicobacter pylori CagA protein in vitro and in vivo. *J Biol Chem*, 2002. 277(9): p. 6775-8.
15. Tammer, I., et al., Activation of Abl by Helicobacter pylori: a novel kinase for CagA and crucial mediator of host cell scattering. *Gastroenterology*, 2007. 132(4): p. 1309-19.
16. Higashi, H., et al., Biological activity of the Helicobacter pylori virulence factor CagA is determined by variation in the tyrosine phosphorylation sites. *Proc Natl Acad Sci U S A*, 2002. 99(22): p. 14428-33.

17. Zhang, X.S., et al., A specific A/T polymorphism in Western tyrosine phosphorylation B-motifs regulates *Helicobacter pylori* CagA epithelial cell interactions. *PLoS Pathog*, 2015. 11(2): p. e1004621.
18. Ren, S., et al., Structural basis and functional consequence of *Helicobacter pylori* CagA multimerization in cells. *J Biol Chem*, 2006. 281(43): p. 32344-52.
19. Wiedemann, T., et al., *Helicobacter pylori* CagL dependent induction of gastrin expression via a novel alpha V beta 5-integrin-integrin linked kinase signalling complex. *Gut*, 2012. 61(7): p. 986-96.
20. Barden, S. and H.H. Niemann, Adhesion of several cell lines to *Helicobacter pylori* CagL is mediated by integrin alphaVbeta6 via an RGD₂LXXL motif. *J Mol Biol*, 2015. 427(6 Pt B): p. 1304-1315.
21. Kwok, T., et al., *Helicobacter* exploits integrin for type IV secretion and kinase activation. *Nature*, 2007. 449(7164): p. 862-6.
22. Yeh, Y.C., et al., H. pylori cagL amino acid sequence polymorphism Y58E59 induces a corpus shift of gastric integrin alpha5beta1 related with gastric carcinogenesis. *Mol Carcinog*, 2011. 50(10): p. 751-9.
23. Ogawa, H., et al., Genetic variants of *Helicobacter pylori* type IV secretion system components CagL and CagI and their association with clinical outcomes. *Gut Pathog*, 2017. 9: p. 21.
24. Fischer, W., et al., Systematic mutagenesis of the *Helicobacter pylori* cag pathogenicity island: essential genes for CagA translocation in host cells and induction of interleukin-8. *Mol Microbiol*, 2001. 42(5): p. 1337-48.
25. Jurik, A., et al., The coupling protein Cagbeta and its interaction partner CagZ are required for type IV secretion of the *Helicobacter pylori* CagA protein. *Infect Immun*, 2010. 78(12): p. 5244-51.
26. Kutter, S., et al., Protein subassemblies of the *Helicobacter pylori* Cag type IV secretion system revealed by localization and interaction studies. *J Bacteriol*, 2008. 190(6): p. 2161-71.
27. Hilleringmann, M., et al., Inhibitors of *Helicobacter pylori* ATPase Cagalpha block CagA transport and cag virulence. *Microbiology*, 2006. 152(Pt 10): p. 2919-30.
28. Krause, S., et al., Sequence-related protein export NTPases encoded by the conjugative transfer region of RP4 and by the cag pathogenicity island of *Helicobacter pylori* share similar hexameric ring structures. *Proc Natl Acad Sci U S A*, 2000. 97(7): p. 3067-72.
29. Savvides, S.N., et al., VirB11 ATPases are dynamic hexameric assemblies: new insights into bacterial type IV secretion. *EMBO J*, 2003. 22(9): p. 1969-80.
30. Yeo, H.J., et al., Crystal structure of the hexameric traffic ATPase of the *Helicobacter pylori* type IV secretion system. *Mol Cell*, 2000. 6(6): p. 1461-72.
31. Machon, C., et al., TrwD, the hexameric traffic ATPase encoded by plasmid R388, induces membrane destabilization and hemifusion of lipid vesicles. *J Bacteriol*, 2002. 184(6): p. 1661-8.

32. Hare, S., et al., Identification, structure and mode of action of a new regulator of the *Helicobacter pylori* HP0525 ATPase. *EMBO J*, 2007. 26(23): p. 4926-34.
33. Arya, T., et al., Fragment-based screening identifies inhibitors of the ATPase activity and of hexamer formation of Cag α from the *Helicobacter pylori* type IV secretion system. *BioRxiv*, 2018.
34. Cover, T.L. and S.R. Blanke, *Helicobacter pylori* VacA, a paradigm for toxin multifunctionality. *Nat Rev Microbiol*, 2005. 3(4): p. 320-32.
35. Cherati, M.R., et al., *Helicobacter pylori* cagL amino acid polymorphism D58E59 pave the way toward peptic ulcer disease while N58E59 is associated with gastric cancer in north of Iran. *Microb Pathog*, 2017. 107: p. 413-418.
36. Backert, S., et al., The *Helicobacter pylori* Type IV Secretion System Encoded by the cag Pathogenicity Island: Architecture, Function, and Signaling, in *Type IV Secretion in Gram-Negative and Gram-Positive Bacteria*, S. Backert and E. Grohmann, Editors. 2017, *Springer International Publishing*. p. 188-209.
37. Backert, S. and T.F. Meyer, Type IV secretion systems and their effectors in bacterial pathogenesis. *Curr Opin Microbiol*, 2006. 9(2): p. 207-17.
38. Amieva, M.R., et al., Disruption of the epithelial apical-junctional complex by *Helicobacter pylori* CagA. *Science*, 2003. 300(5624): p. 1430-4.
39. Peek, R.M., Jr. and M.J. Blaser, *Helicobacter pylori* and gastrointestinal tract adenocarcinomas. *Nat Rev Cancer*, 2002. 2(1): p. 28-37.
40. Covacci, A. and R. Rappuoli, Tyrosine-phosphorylated bacterial proteins: Trojan horses for the host cell. *J Exp Med*, 2000. 191(4): p. 587-92.
41. Segal, E.D., et al., Altered states: involvement of phosphorylated CagA in the induction of host cellular growth changes by *Helicobacter pylori*. *Proc Natl Acad Sci U S A*, 1999. 96(25): p. 14559-64.
42. Rieder, G., et al., Role of adherence in interleukin-8 induction in *Helicobacter pylori*-associated gastritis. *Infect Immun*, 1997. 65(9): p. 3622-30.
43. Gorrell, R.J., et al., A novel NOD1- and CagA-independent pathway of interleukin-8 induction mediated by the *Helicobacter pylori* type IV secretion system. *Cell Microbiol*, 2013. 15(4): p. 554-70.
44. Nishikawa, H., et al., Impact of structural polymorphism for the *Helicobacter pylori* CagA oncoprotein on binding to polarity-regulating kinase PAR1b. *Sci Rep*, 2016. 6: p. 30031.
45. Saadat, I., et al., *Helicobacter pylori* CagA targets PAR1/MARK kinase to disrupt epithelial cell polarity. *Nature*, 2007. 447(7142): p. 330-3.
46. Zeaiter, Z., et al., Analysis of detergent-resistant membranes of *Helicobacter pylori* infected gastric adenocarcinoma cells reveals a role for MARK2/Par1b in CagA-mediated disruption of cellular polarity. *Cell Microbiol*, 2008. 10(3): p. 781-94.

Chapitre III : DISCUSSION

Au cours de mon doctorat, nous avons eu pour ambition de participer à l'amélioration de la compréhension de deux complexes macromoléculaires : 1) la machinerie eucaryote de biosynthèse et d'incorporation de la sélénocystéine dans les protéines et 2) le système de sécrétion de type 4 d'*H. pylori*, encodé par le *cagPAI*, un facteur de virulence identifié dans les souches les plus pathogènes. Dans cette optique, nous avons choisi de mettre en place une approche multidisciplinaire qui nous a permis d'étudier les interactions protéines-protéines entre les composants de la machinerie de biosynthèse et d'incorporation de la sélénocystéine et de résoudre la structure d'un des facteurs protéiques (SECp43) à basse résolution. Nous avons également entrepris d'identifier et de caractériser des inhibiteurs potentiels de la virulence d'*H. pylori*, de déterminer leur mode d'action *in vitro* et leur impact *in vivo*, en combinant des approches biochimiques et structurales. Finalement, nous avons étudié la variabilité génétique d'isolats bactériens d'*H. pylori* issus de biopsies gastriques de patients, présentant différents profils de résistance aux antibiotiques. En nous intéressant plus particulièrement à des protéines encodées par l'îlot de pathogénicité *cagPAI* (CagA, CagL et Cag α), nous avons tenté d'établir l'impact de cette variabilité sur la capacité de la bactérie à infecter les cellules AGS.

III.1 Identification de nouvelles interactions dans la machinerie Sec par BRET

La machinerie de biosynthèse et d'incorporation des sélénocystéines dans les protéines est une machinerie complexe gouvernée par des dynamiques d'interaction protéine-protéine et protéine-ARN, dans laquelle sont engagés : l'ARNt^{[Ser]^{Sec}} et l'ARNm de sélénoprotéines, comprenant l'élément SECIS, différents facteurs protéiques spécialisés et canoniques et le ribosome. Bien que les composants centraux de la machinerie soient à présent identifiés, de nombreux aspects mécanistiques nécessitent d'être clarifiés.

En effet, la séquence d'assemblage des composants et la dynamique d'interaction au sein de la machinerie ne sont pas résolues; il est également vraisemblable que toutes les interactions protéiques n'aient pas été élucidées. De plus, la structure des protéines impliquées dans la machinerie, les sites d'interaction et la stœchiométrie des protéines dans les complexes de

biosynthèse et d'incorporation des sélénocystéines ne sont pas résolus. Finalement, le ou les facteurs à l'origine du transport nucléocytoplasmique des complexes n'ont pas été identifiés.

Notre objectif a ainsi été d'optimiser une méthode *in vivo*, semi-quantitative, applicable à l'étude des interactions protéiques qui s'opèrent au sein de la machinerie de synthèse et d'incorporation des sélénocystéines, dans les cellules eucaryotes. Nous avons donc mis au point l'essai BRET, une technique d'analyse de la dynamique des interactions basée sur le transfert d'énergie entre une protéine donneuse d'énergie, couplée à une enzyme luminescente (*RLuc*), et une protéine receveuse, couplée à une protéine fluorescente (*GFP*), les deux protéines étant co-exprimées dans les cellules eucaryotes HEK293T (Fig. III.1).

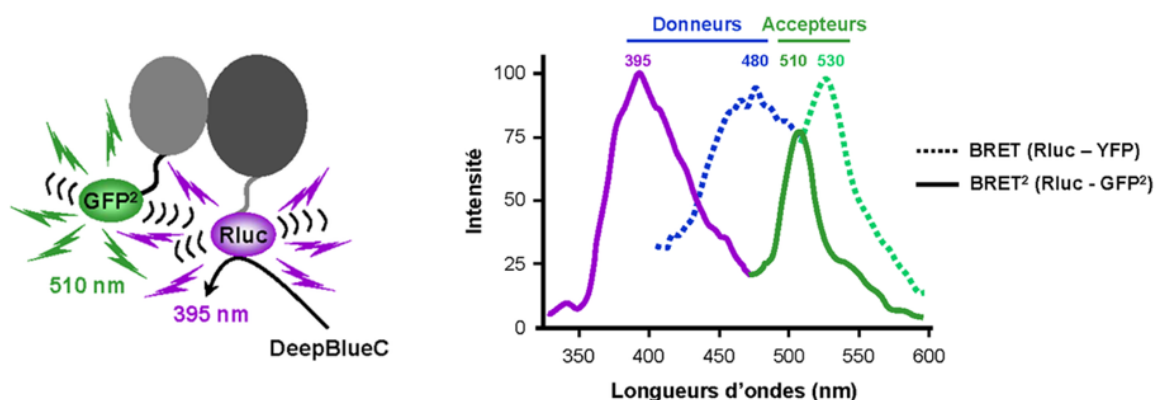


Figure III.1. Principe du BRET. Après ajout de son substrat (*DeepBlueC*), la *RLuc* émet de la luminescence à 395 nm, ce qui conduit à l'excitation de la *GFP* et à la réémission d'une fluorescence à 510 nm. Le transfert d'énergie est strictement dépendant de la distance entre les protéines donneuses et receveuses (< 10 nm) et de leur orientation. Inspiré de Hamdam et *al.*, 2006 [239].

Bien que notre intention fût de mener l'analyse d'interactions sur l'ensemble des composants de la machinerie de synthèse et d'incorporation de la Sec connus à ce jour, nos tentatives pour produire des protéines de fusion fonctionnelles avec les protéines EFSec et SBP2 ont échoué, et l'étude a été poursuivie en l'absence de ces constructions.

Dans un premier temps, l'essai BRET nous a permis de confirmer que plusieurs protéines impliquées dans les étapes de synthèse de la Sec (*SEPSECS* et *SEPHS1*) formaient des homomultimères tel qu'observé dans des études antérieures [43, 240]. Nous avons également

identifié d'autres facteurs (SECp43 et SEPHS2) dont la capacité à former des homomultimères *in vivo* n'avait pas été explorée (Tableau III.1). Nous avons, par la suite, mis en évidence des interactions jusqu'alors inconnues (SEPSECS-SEPHS2 et SEPHS1-SEPHS2), celles-ci ayant également été confirmées par co-immunoprécipitation à partir de lysats cellulaires (Tableau III.1). Bien que l'interaction entre SEPSECS et SEPHS2 n'avait jamais été démontrée, celle-ci est cohérente avec la fonction connue des deux protéines : SEPSH2 synthétise le donneur actif de sélénium ensuite utilisé par SEPSECS pour produire la Sec sur l'ARNt^{[Ser]Sec}. L'interaction de SEPSH2 avec son homologue SEPSH1 était, en revanche, inattendue. Le rôle de SEPHS1 dans la machinerie de synthèse et d'incorporation de la Sec n'est pas avéré, la déplétion de la protéine n'ayant, en effet, aucun impact sur la synthèse des sélénoprotéines. SEPSH1 interagit, cependant, avec plusieurs autres composants du système (SEPSECS, SECp43), ainsi la portée de cette interaction avec SEPHS2 nécessite d'être étudiée plus en détails.

Tableau III.1. Interactions identifiées dans cette étude et dans des travaux précédents

		INTERACTIONS IDENTIFIEES DANS NOTRE ETUDE	INTERACTIONS IDENTIFIEES DANS DES ETUDES ANTERIEURES
SEPSECS	SEPSECS	BRET	Structure cristallographique [43]
	SEPHS1	BRET	coIP [42]
	SEPHS2*	BRET ; coIP	
	SECp43	BRET	coIP [41, 42]
SEPHS1	SEPHS1	BRET	Structure cristallographique [240]
	SEPHS2*	BRET, coIP	
	SECp43	BRET	coIP [42]
	SBP2		coIP [42]
	EFSEC		coIP [42]
SEPHS2	SEPHS2*	BRET	
SECP43	SECp43*	BRET	
	SBP2		coIP [42]
	EFSEC		coIP [42]

(*) : interactions nouvellement identifiées

Dans notre étude, nous avons tiré profit de la technologie du BRET pour caractériser les interactions protéiques de façon semi-quantitative. Or, l'essai BRET est également adapté pour identifier, en temps réel et de façon précise, la localisation cellulaire et la dynamique spatiale des facteurs protéiques en microscopie à fluorescence. Des signaux d'export et de localisation nucléaire (NES et NLS) ont été prédits dans plusieurs facteurs protéiques de la machinerie (SBP2, EFSec, SECp43), mais les protéines responsables de l'import de l'ARNt^{[Ser]Sec} dans le noyau puis de son export une fois complexé à l'ARNm n'ont pas été déterminées. Ainsi, le BRET serait un outil efficace pour répondre à ces questions. Pour cela, différentes constructions de BRET seraient co-exprimées dans les cellules HEK293T afin de visualiser et suivre des changements de localisation cellulaire des protéines d'intérêt en temps réel par microscopie à fluorescence. De plus, le BRET permettrait d'établir la contribution de chaque protéine dans l'assemblage et le transport des complexes cytoplasmiques et nucléaires car, dépendant de la conformation et de l'orientation des protéines, il permet de détecter les changements d'interactions. Ainsi, il serait possible de tester différentes combinaisons de BRET lors desquelles l'interaction entre deux protéines (fusionnées aux protéines rapporteuses *RLuc* et *GFP*) seraient visualisés sous l'effet de l'induction de l'expression d'une protéine ou d'un effecteur tiers. L'importance, dans l'assemblage et le transport des complexes, de chaque protéine connue de la machinerie pourra ainsi être examinée, tandis que l'ajout d'ARNt^{[Ser]Sec} purifié, qui semble promouvoir l'interaction entre plusieurs composants, est également envisageable. Une alternative serait d'exploiter les propriétés spectrales des luciférases et protéines fluorescentes récemment développées, rendant possible l'utilisation concomitante de deux protéines rapporteuses fluorescentes excitables par une unique luciférase (*ex* : *EB-FP2* et *mAmetrine* émettent respectivement dans le bleu et le vert, dans le spectre visible). Cette configuration de BRET, mis au point et testée avec succès dans le cadre de l'étude des GPCRs (*G protein-coupled receptors*) [241], conduit à l'émission de deux signaux de fluorescence distincts, sans interférence dans les spectres d'émission, et permettrait d'étudier de façon simultanée l'interaction d'une protéine avec différents partenaires.

III.2 Analyse structurale du domaine N-terminal de SECp43 et identification de son site de dimérisation

La protéine SECp43 est une protéine importante pour la synthèse des sélénoprotéines. Elle semble, en effet, impliquée dans la méthylation de l'ARNt^{[Ser]^{Sec}}, étape requise pour la synthèse des sélénoprotéines liées au stress. Outre son interaction avec l'ARNt^{[Ser]^{Sec}}, SECp43 co-immunoprécipite avec la plupart des autres facteurs protéiques de la machinerie (SEPSECS, SEPHS1, EFSec, SBP2). Le BRET étant considéré comme une technique de mesure des interactions directes (le transfert d'énergie requiert une distance < 10 nm), nous avons ici mis en évidence que SECp43 interagissait de façon directe avec SEPSECS et avec SEPSH1, et que la protéine pouvait également former des homodimères.

Nous avons, par la suite, tenté de résoudre la structure tridimensionnelle de SECp43. À la suite de sa production dans la bactérie *E. coli* et d'une analyse par spectrométrie de masse, il fut, néanmoins, constaté que la région C-terminale de la protéine produite dans la bactérie était clivée. Nous avons donc décidé d'insérer un codon stop après le 180^e acide aminé afin de purifier une forme stable de la protéine. Cette région contient par ailleurs deux domaines de liaison à l'ARN (RRM1 et RRM2), précédemment résolus par RMN. L'utilisation du SAXS nous a permis de définir la structure globale du domaine N-terminal de la protéine, à savoir un domaine globulaire, et de reconstruire la liaison entre les deux domaines RRM en insérant les sept résidus absents dans les structures de RMN, par modélisation. Cette structure pourra servir de base pour des analyses futures des sites d'interaction de SECp43 avec ses protéines partenaires ainsi qu'avec l'ARNt^{[Ser]^{Sec}}. Afin d'obtenir une structure de la protéine à haute résolution, il aurait été intéressant de cristalliser SECp43. Cependant, les hautes concentrations en sels nécessaires à sa purification et sa stabilité (200 mM NaCl) ont rendu impossible sa cristallisation. La co-purification de SECp43 en complexe avec l'ARNt^{[Ser]^{Sec}}, en vue de l'obtention de co-cristaux, est également d'un grand intérêt. Néanmoins, les tentatives de purification de l'ARNt^{[Ser]^{Sec}}, menées par d'autres étudiants, au laboratoire, ont échoué.

Nous avons également entrepris de résoudre l'interface de dimérisation de SECp43. Pour cela, nous avons utilisé la technique de *phage display* qui nous a permis d'identifier une région d'intérêt composée de 7 résidus (LGSKPVR) dont la substitution en alanine altère (L163A,

G164A, S165S, R168A) ou inhibe (K166A et P167A) la formation de dimères. D'après la structure que nous proposons, cette région serait localisée à la surface de la protéine SECp43, dans une zone flexible du RRM2. Cette observation soulève de nouvelles questions, concernant notamment la stœchiométrie du complexe SECp43- ARNt^{[Ser]^{Sec}} ; il reste en effet à déterminer si un dimère de SECp43 coordonne un ou plusieurs ARNt^{[Ser]^{Sec}}. Les sites d'interaction avec les autres partenaires de SECp43 doivent également être examinés afin de déduire la séquence d'assemblage des différents complexes impliquant la protéine. Dans cette optique, nous avons par ailleurs entrepris d'identifier une zone potentielle d'interaction entre SECp43 et SEPSECS par *phage display*. La séquence peptidique identifiée a été localisée sur la structure de SEPSECS mais celle-ci n'a pas été vérifiée expérimentalement. Il serait d'autant plus intéressant de mener l'essai de *phage display* après immobilisation des partenaires connus de SECp43 afin de définir les sites d'interaction avec SECp43.

Malgré les nombreuses questions restées en suspens dans le cadre de ce projet, nous avons choisi de réorienter mes recherches sur le système plus largement étudié au sein du laboratoire du Dr Baron. La motivation de cette réorientation étant notamment de faciliter les collaborations avec les autres membres de l'équipe, de les faire bénéficier de mes expertises et de participer activement aux avancées faites dans le laboratoire. Ainsi, nous avons poursuivi nos recherches sur un second modèle de complexe macromoléculaire : le système de sécrétion de type 4 d'*H. pylori*.

III.3 Caractérisation d'inhibiteurs de l'activité ATPase de Caga d'*H. pylori* : 1G2, une molécule prometteuse

La bactérie *H. pylori* est un pathogène humain largement répandu dans la population mondiale, dont l'éradication devient problématique du fait de l'émergence de souches multirésistantes aux antibiotiques. Au sein du laboratoire du Dr Baron, nous avons pris le parti de développer et caractériser des molécules qui inhiberaient la virulence d'*H. pylori*, en ciblant de manière spécifique les protéines du CagSST4, un facteur majeur de virulence.

Dans le cadre d'un projet en collaboration avec plusieurs autres membres de l'équipe, j'ai ainsi mené une étude de caractérisation d'inhibiteurs potentiels de la virulence d'*H. pylori*, à la suite de l'infection de cellules eucaryotes par la bactérie. Ce projet, initié par Tarun Arya, a débuté

par le criblage d'une banque de 505 petites molécules organiques capables de stabiliser la structure de la protéine Cag α , encodée par le *cagPAI* d'*H. pylori*. Cette ATPase hexamérique, essentielle pour la translocation de CagA et pour l'induction d'une réponse inflammatoire dans la cellule hôte [144], est une cible idéale dans la recherche d'inhibiteurs de virulence de la bactérie puisqu'elle est nécessaire pour la fonctionnalité du SST4. D'autre part, son activité enzymatique peut aisément être évaluée et la structure cristallographique de Cag α a été déterminée par plusieurs groupes [160, 161]. Seize molécules capables d'interagir avec Cag α ont ainsi été identifiées dont quatre d'entre elles ont eu pour effet de réduire l'activité ATPase de Cag α , de façon non-compétitive. Parmi celles-ci, la molécule 1G2 semble induire un changement conformationnel de la structure de la protéine et inhiber la formation d'hexamères. D'autre part, Cag α a été co-cristallisée en présence de la molécule 1G2 et le site de liaison a ainsi pu être identifié. L'alignement des séquences des homologues Cag α /VirB11 a mis en évidence que ce site n'était pas conservé entre les homologues et suggère que la molécule 1G2 pourrait être un inhibiteur spécifique de Cag α . Mon objectif a, par la suite, été d'analyser les effets des molécules identifiées sur la fonctionnalité du SST4 d'*H. pylori in vivo*. Nous avons, dans premier temps, vérifié que les molécules n'étaient pas cytotoxiques, en testant à la fois la viabilité de la bactérie en présence des molécules d'intérêt et celle des cellules AGS, le modèle cellulaire eucaryote communément utilisé pour les essais d'infection par *H. pylori*. Après avoir conclu que les molécules n'étaient pas ou peu toxiques, nous avons procédé à des essais d'infection des cellules AGS par *H. pylori*, en incubant, au préalable, la bactérie avec les molécules. Tandis que l'effet d'inhibition des molécules sur la translocation de l'oncoprotéine CagA n'est pas significatif, la production d'interleukine-8 par les cellules AGS est, quant à elle, réduite de 50% en présence de la molécule 1G2. Ceci, suggère que la molécule 1G2 a un effet d'inhibition de la virulence de la bactérie *H. pylori*. À long terme, la chronicité de l'inflammation constitue un facteur de risque pour le développement de cancers gastriques. La molécule 1G2, qui réduit l'effet de l'infection par *H. pylori* sur la réponse pro-inflammatoire, présente donc un fort potentiel thérapeutique. Ainsi, nous présentons ici l'étude de caractérisation d'une molécule diminuant les fonctionnalités du SST4 pour laquelle un mode d'action est également proposé. Cette molécule pourra, par ailleurs, servir de base pour la conception de nouvelles molécules ayant un effet plus prononcé.

III.4 Rôle présumé de Cag α dans le SST4 d'*H. pylori*

Il fut précédemment proposé que l'injection de CagA dans le cytoplasme de la cellule hôte induisait une réponse inflammatoire [242], toutefois, la production d'interleukine-8 semble être largement influencée par la protéine CagL et par la translocation du peptidoglycane bactérien [157, 176]. En effet, la délétion de *cagA* n'inhibe pas la production d'interleukine-8 par les cellules AGS [144]. Les données de l'étude *in vivo*, qui montrent une réduction du niveau de production d'interleukine-8 sous l'effet de la molécule 1G2, alors que les niveaux de CagA phosphorylée ne sont pas affectés, corroborent donc ces observations, qui suggèrent l'importance de Cag α dans l'induction de la réponse pro-inflammatoire. Néanmoins, ces nouvelles données soulèvent plusieurs points.

Tout d'abord, nos données suggèrent que la molécule 1G2 n'inhibe que partiellement l'activité ATPase de Cag α ainsi que la formation d'hexamères. En effet, des études antérieures ont montré que la délétion du gène *cagA* avait pour effet d'abolir la translocation de CagA et la production d'interleukine-8 [144]. De ce fait, il est possible que, dans notre étude, l'inhibition induite sur Cag α ne soit pas suffisante pour bloquer la translocation de CagA. De façon intéressante, il fut également observé que le monomère de Cag α semblait conserver son activité ATPase mais que la conformation hexamérique était toutefois requise pour la translocation de CagA [161]. Il est à noter que ces observations ne sont pas contradictoires avec le mode d'action, ici, proposé de la molécule 1G2, qui réduit à la fois l'activité ATPase de Cag α et la formation d'hexamères.

Basé sur nos données et sur des études précédentes, il est possible de suggérer plusieurs modèles quant au rôle de Cag α et à la trajectoire empruntée par les effecteurs du CagSST4. Dans notre premier modèle, nous proposons que les mécanismes conduisant à l'injection de CagA et à la production d'interleukine-8 seraient distincts. Ce modèle se base notamment sur l'observation que l'hexamère de Cag α semble former un pore, dont la conformation ouverte, pourrait permettre le passage d'effecteurs du CagSST4 à travers la membrane interne [160]. Ainsi, Cag α pourrait spécifiquement être impliquée dans le transport du peptidoglycane, et donc, dans la production d'interleukine-8. Cette hypothèse est étayée par le fait que, tandis que la délétion de *cag β* , une autre ATPase du CagSST4 et homologue de VirD4 d'*A. tumefaciens*, abroge la translocation de CagA, celle-ci n'a pas d'impact sur la production d'interleukine-8 par les

cellules AGS [144]. Cag β serait ainsi responsable du recrutement de CagA et de son adressage à CagE [154]. L'ATPase CagE est, elle, indispensable pour la formation des pili, la translocation de CagA et l'induction d'interleukine-8. Basé sur son interaction avec Cag β et CagV et sur les similarités de séquences avec les protéines du plasmide de conjugaison R388, il est proposé que CagE appartienne au complexe protéique de la membrane interne [143, 243]. Ainsi CagE, qui permettrait également d'énergiser l'assemblage du pilus, fonctionnerait de pair avec Cag β pour permettre la translocation de CagA à travers le canal de la membrane interne, tandis que le peptidoglycane serait transporté à travers le pore formé par Cag α (Figure III.2.A). Néanmoins, ce modèle n'explique pas l'incapacité du mutant $\Delta cag\alpha$ à transloquer la protéine CagA [144]. Il est alors possible de suggérer que le phénotype induit par la molécule 1G2 soit moins prononcé que celui induit par la délétion du gène *cag α* , et que la présence de la protéine Cag α soit toutefois requise pour assurer la fonctionnalité du CagSST4. Dans notre modèle alternatif, l'ensemble des effecteurs seraient transloqués par le pore formé par la protéine Cag α , associée à CagE et au complexe de la membrane interne (Figure III.2B). Dans ce cas, comme nous le suggérons précédemment, et tel qu'observé dans nos travaux, l'inhibition induite par la molécule 1G2 sur Cag α ne serait pas suffisante pour empêcher la translocation de CagA. L'addition de concentrations plus élevées de molécule 1G2 permettrait de tester cette hypothèse. Finalement, dans notre troisième modèle, Cag α participerait à fournir l'énergie nécessaire à la translocation des effecteurs à travers le canal formé par CagE et par les composants du complexe de la membrane interne (Figure III.2C). Dans ce contexte, les ATPases Cag β et CagE pourraient agir de façon synergique pour compenser la baisse d'activité de Cag α , engendrée par l'addition du composé 1G2. Les deux ATPases fourniraient ainsi assez d'énergie au système et assureraient les étapes conduisant à la translocation de CagA.

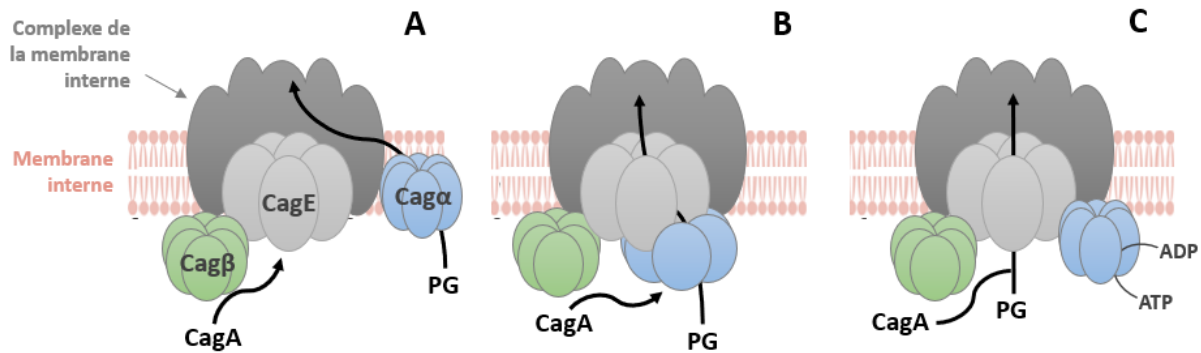


Figure III.2. Modèles illustrant les rôles hypothétiques de Cag α et des autres ATPases dans la translocation des effecteurs du CagSST4 à travers la membrane interne. (A) Le transfert de CagA est assuré par Cag β et CagE, tandis que le peptidoglycane (PG) est transloqué à travers le pore de Cag α . (B) CagA est recrutée par Cag β . L'ensemble des effecteurs sont transloqués à travers le canal de la membrane interne par l'intermédiaire de Cag α et CagE, toutes deux associées au complexe de la membrane interne. (C) Les effecteurs sont transloqués à travers CagE associée complexe de la membrane interne. Cag α produit l'énergie nécessaire à la translocation. Le complexe de la membrane interne comprend probablement les protéines CagV et CagW.

Le schéma actuel du CagSST4 d'*H. pylori* est basé, en partie, sur les homologies fonctionnelles et similarités structurales avec d'autres SST4 et la composition du complexe de la membrane interne n'est pas strictement établie. Dans le cas du SST4 du plasmide R388, les homologues de Cag β et CagE interagissent avec l'homologue de Cag α , dans une conformation hexamérique et indépendamment de son activité ATPase [244]. Ainsi, le modèle consensuel du CagSST4 repose, notamment, sur des interactions entre les ATPases et leur association au complexe de la membrane interne. Or, chez *H. pylori*, l'interaction de Cag α avec d'autres protéines du CagSST4 n'a pas été validée expérimentalement et son association avec le complexe de la membrane interne est hypothétique. De ce fait, il serait intéressant d'analyser le profil d'interaction entre Cag α et les autres protéines du CagSST4 et tout particulièrement avec Cag β et CagE. Si l'interaction avec une ou ces deux ATPases est avérée, l'ajout de la molécule 1G2 permettrait d'examiner, l'impact de l'altération de la formation d'hexamères de Cag α sur la formation du complexe. Au contraire, dans le cas où les interactions ne seraient pas détectées, le schéma actuel devra être réenvisagé.

D'autre part, de façon similaire à son homologue VirB11 d'*A. tumefaciens* [245], mais contrairement à ce qui fut proposé pour Cag α [246], il est également possible que Cag α soit directement impliquée dans la biogénèse des pili. Pour tester cette hypothèse, nous proposons d'étudier l'effet de la molécule 1G2 sur la formation des pili. Préalablement à l'infection des cellules AGS avec *H. pylori*, la bactérie sera incubée avec la molécule 1G2 et l'impact sur la biogénèse des pili sera visualisée par microscopie électronique. Pour induire un phénotype plus ou moins prononcé, différentes concentrations de la molécule pourront être testées.

Finalement, bien que notre objectif fût de caractériser le mode d'action et la cible de la molécule dans *H. pylori*, il est envisageable que la molécule 1G2 interfère avec d'autres protéines (CagSST4-dépendantes ou -indépendantes) qui interviendraient dans la production d'interleukine-8, ceci expliquant les phénotypes observés. Afin d'écarter cette hypothèse, la réponse pro-inflammatoire des cellules AGS pourrait être activée de façon spécifique par un composant purifié (ex : TNF- α), en absence d'*H. pylori* et en présence de la molécule 1G2, afin de déterminer si la molécule 1G2 affecte les niveaux de production d'interleukine-8.

En conclusion, l'utilisation de la molécule 1G2 dans la détermination de la fonction de Cag α dans le CagSST4 s'avère, ici, être un atout. En effet, sans strictement abolir son activité, 1G2 réduit l'activité de Cag α , ce qui permet, du même fait, de moduler son incidence sur le CagSST4. Alors que le CagSST4 reste fonctionnel et actif, il devient possible d'évaluer l'impact de la diminution d'activité de Cag α sur le système. Ainsi il est possible d'aborder l'exploration du système par une méthodologie originale de la biologie chimique. Toutefois, des données supplémentaires seront nécessaires pour établir la fonction exacte de Cag α dans le CagSST4.

III.5 Intérêt du développement d'inhibiteurs de virulence dans l'éradication d'*H. pylori*

Dans un contexte où les bactéries pathogènes résistantes aux antibiotiques tendent à se propager, le développement d'alternatives aux antibiotiques s'impose. L'utilisation inappropriée des antibiotiques et les stratégies mises en place par les bactéries pour lutter contre les mécanismes d'action des antibiotiques sont, en effet, à l'origine de l'émergence de bactéries multirésistantes. La bactérie *H. pylori* illustre bien cette évolution puisqu'au cours des 20 dernières années, les souches devenues résistantes aux traitements antibiotiques conventionnels se sont

considérablement multipliées. L'accélération du phénomène est d'autant plus inquiétante que la bactérie est largement répandue dans la population mondiale et qu'elle induit des pathologies sévères, dont l'adénocarcinome gastrique, à l'origine de 800 000 décès par an.

Les gènes responsables de la virulence des bactéries pathogènes peuvent être impliqués dans l'interaction hôte-pathogène, dans les processus d'invasion, de colonisation et de persistance, et dans la propagation des gènes de résistance aux antibiotiques. Ainsi, le ciblage des effecteurs de la virulence par des inhibiteurs est une stratégie innovante qui peut s'avérer efficace pour neutraliser et contrer la propagation des bactéries sans induire le développement de résistance. En effet, il est possible d'envisager que la pression de sélection résultante de l'utilisation d'inhibiteurs de virulence soit moins forte que lorsque les gènes constitutifs (domestiques) sont ciblés.

De plus, les inhibiteurs de virulence sont développés pour cibler une espèce bactérienne en particulier. De ce fait, contrairement aux antibiotiques qui manquent de spécificité, les molécules inhibitrices de virulence ne devraient pas avoir d'impact sur la survie des bactéries commensales, non pathogènes et bénéfiques pour l'organisme. Dans le cas d'*H. pylori*, nous proposons ainsi de cibler des protéines encodées par le *cagPAI*, un facteur avéré de la virulence de la bactérie, telle que la protéine Cag α .

D'autre part, il a été démontré que la bactérie *H. pylori* était intimement associée à l'Homme depuis plus de 60 000 ans [247]. Des données épidémiologiques récentes suggèrent même qu'une infection chronique pourrait être bénéfique pour l'hôte en conférant une protection contre les maladies gastro-œsophagiennes, l'asthme, certaines manifestations allergiques et des maladies inflammatoires de l'intestin (colite ulcéreuse, maladie de Crohn) (résumé dans la revue Backert et *al.* 2016 [186]). Dans ce contexte, neutraliser le potentiel pathogène de la bactérie, sans l'éradiquer, pourrait être une alternative intéressante. De plus, les inhibiteurs de virulence peuvent également être envisagés en supplément des traitements antibiotiques actuels pour assurer l'efficacité du traitement et garantir l'élimination de la bactérie.

III.6 Variabilité génétique et biochimique des isolats *H. pylori*, conséquences sur la virulence et inhibition de la virulence par 1G2

Dans le but d'examiner les variabilités entre des souches *H. pylori* directement isolées de patients infectés, nous avons, par la suite, initié un projet en collaboration avec Dre Christiane Gaudreau, dont le laboratoire est situé au CHUM (Centre Hospitalier de l'Université de Montréal). Ayant à notre disposition plusieurs souches d'*H. pylori*, issues de biopsies de patients, pour lesquelles les traitements antibiotiques entrepris avaient échoué, nous avons entrepris d'étudier les caractéristiques génétiques de ces différentes souches en nous intéressant plus particulièrement à des protéines associées à la virulence, et encodées par le *cagPAI*. Nous avons également établi les niveaux de production de ces protéines et mesuré l'influence de ces différentes souches sur la modification du phénotype des cellules épithéliales gastriques infectées. Puis, nous avons évalué l'efficacité de la molécule 1G2 sur l'inhibition de la virulence de ces souches.

Dans un premier temps, nous avons mis en évidence que neuf des isolats *H. pylori* comportaient l'îlot de pathogénicité *cag*. Nous avons, par la suite, constaté que trois souches (3793, 3822 et 3830) semblaient induire les effets les plus marqués sur le phénotype des cellules AGS. Les résultats de l'étude sont récapitulés dans le Tableau III.2.

Tableau III.2. Synthèse des données de caractérisation génétique et biochimique des souches d'*H. pylori*

Souche <i>H. pylori</i>	Type de motif EPIYA	Type de motif CM	Niveaux de CagL produits	Phénotype Hummingbird	Induction d'IL-8
26695	ABC	CM ^W	+++	++	+++
3723	ABC	CM ^W	+	++	+
3763	ABC	CM ^W	+	+	+
3783	ABC	CM ^W	-	-	-
3793	ABC	CM ^W /CM ^{EA}	+	++	+++
3813	ABC	CM ^W	+++	++	++
3822	ABC	CM ^W	+++	+++	+++
3830	ABD	CM ^{EA}	+	++++	+++
ATCC 43504	ABCCC	CM ^W	++++	++	+/++

- : nul; + : faible; ++ : moyen; +++ : élevé; ++++ : très élevé

L'analyse du domaine C-terminal de CagA a révélé que, tandis que la majorité des souches encodent une isoforme CagA de type occidental (EPIYA-ABC), la souche 3830 encodait pour un isoforme de type Asie de l'Est (EPIYA-ABD). De plus, la souche 3793 présentait une combinaison de motifs CM ; un motif CM occidental (CM^W) et un motif CM de type Asie de l'Est (CM^{EA}). Or, une fois dans la cellule hôte, les motifs EPIYA et CM présents dans le domaine C-terminal de CagA constituent les sites d'interaction avec les protéines de la cellule hôte. Plus précisément, le motif EPIYA-C/D est ciblé par la protéine SHP-2, impliquée dans l'activation de la voie des kinases MAP, dans la morphogénèse et dans la motilité cellulaire, tandis que le motif CM constitue notamment le site d'interaction des protéines de la famille PAR1 [248]. L'interaction au motif CM, qui inhibe l'activité de la protéine PAR1b, en particulier, conduit à la perte de polarité cellulaire, à la rupture des jonctions cellulaires et à la perturbation des microtubules et du cytosquelette. Or, il fut reporté que les motifs EPIYA-D et CM^{EA} présents dans l'isoforme CagA de type Asie de l'Est étaient liés par les protéines SHP-2 et Par1b, respectivement, avec une forte affinité [192, 249]. Ces observations pourraient ainsi expliquer les phénotypes marqués induits sur les cellules AGS à la suite de l'infection par les souches 3830 et 3793.

En revanche, une corrélation entre le génotype de la souche 3822, encodant une isoforme CagA similaire à celle de la souche contrôle 26695, et son haut niveau de virulence ne put pas être établie. Son effet prononcé sur le phénotype des AGS n'est pas non plus explicable par les niveaux de production des protéines Cag, bien que la souche 3822 produise des niveaux élevés de protéine CagL. Ces résultats suggèrent qu'afin de prédire le potentiel de virulence des souches *H. pylori*, d'autres facteurs doivent être pris en compte. D'autres protéines Cag nécessaires à la fonction du CagSST4 ainsi que les divers facteurs de virulence (SST4-dépendants et -indépendants) sécrétés par *H. pylori* sont ainsi à considérer. Ces résultats soulignent d'autant plus la complexité des mécanismes d'infection et de colonisation mises en œuvre par *H. pylori* et la difficulté d'identifier des marqueurs de virulence fiables.

Finalement, nous avons évalué l'efficacité de la molécule 1G2, précédemment identifiée, sur les trois souches cliniques d'*H. pylori* les plus virulentes (3793, 3822 et 3830). Nous avons ainsi pu constater que, comme dans le cas de la souche contrôle 26695, la molécule 1G2 réduisait les niveaux d'interleukine-8 produits par les cellules AGS sous l'effet de l'infection par les souches

d'*H. pylori*. L'efficacité de la molécule a, par ailleurs, été démontrée sur les souches de type occidental (souche 26695) ainsi que sur les souches de type Asie de l'Est (souche 3830), hautement pathogènes. De plus, son efficacité sur la diminution de la virulence de souches d'*H. pylori* résistantes aux antibiotiques souligne, d'autant plus, l'intérêt thérapeutique de cette molécule. Nous avons néanmoins constaté une efficacité amoindrie dans le cas de la souche 3822 qui présente une isoforme de Cag α . La variation du résidu en position 73 (R73K) qui, comme nos données le suggèrent, est impliqué dans la liaison de la molécule 1G2, semble affecter la fixation de la molécule 1G2, réduisant l'effet d'inhibition sur Cag α dans la souche 3822. Ces observations seront ainsi prises en considération lors du développement de dérivés de la molécule 1G2, afin d'augmenter la spécificité de la molécule.

III.7 Étude de la localisation de CagL par microscopie à fluorescence

La protéine CagL est une protéine essentielle du CagSST4, en effet, la délétion de *cagL* inhibe la translocation de CagA, l'induction d'interleukine-8 et la formation des pili [152, 153]. De plus, la localisation de CagL au bout du pilus permet son interaction avec les intégrines des cellules hôtes [152, 167-169, 174, 250]. Cette interaction permet l'adhésion de la bactérie aux cellules épithéliales gastriques et stimule la biogenèse des pili [148]. Il est également proposé que, du fait de sa similarité de séquence avec la protéine VirB5 d'*A. tumefaciens*, CagL constitue une piline mineure [152], mais cette hypothèse n'a pas été vérifiée expérimentalement. Tandis que les études de localisation de CagL sont généralement menées à un stade avancé d'infection (4h post-infection), sa localisation précédant l'infection ou à des stades précoces n'a pas été établie. Ainsi, nous émettons l'hypothèse que la protéine CagL pourrait se relocaliser au bout du pilus au cours de l'infection et proposons de mener une étude de localisation de la protéine à différents stades d'infection. Pour étudier sa localisation avec précision, nous exploiterons les avancées récentes en microscopie à fluorescence et emploierons le SIM (*Structured Illumination Microscopy*) dans un premier temps.

La formation du pilus étant induite sous l'effet du contact avec la cellule hôte, nous avons, tout d'abord, vérifié que la protéine CagL était produite dans la bactérie indépendamment du statut de l'infection. Nos données préliminaires confirment que la protéine est produite de façon

constitutive dans *H. pylori* (résultats non présentés). Par la suite, nous avons tenté de mener l'étude de localisation de CagL dans la bactérie *H. pylori* par immunofluorescence ; celle-ci n'a cependant pas abouti, les signaux détectés n'étant pas spécifiques. À présent, nous nous employons à générer la protéine CagL couplée à des protéines fluorescentes (mCherry, sfGFP, HaloTag). Pour cela nous avons mis en place deux stratégies. La première stratégie consiste à réintroduire la séquence nucléotidique codant pour CagL fusionnée à la protéine fluorescente à au locus initial du gène *cagL*, dans le chromosome, et ce, par recombinaison homologue à partir d'un plasmide. L'avantage de cette méthode est que CagL sera ainsi exprimée à des niveaux endogènes, sous le contrôle de son promoteur d'origine. Néanmoins, *cagL* est compris dans un opéron et sa séquence est chevauchante avec celle de *cagI*, ce qui complexifie sa réinsertion. La stratégie alternative consiste à exprimer *cagL*, sous le contrôle du promoteur *cagA*, à partir d'un plasmide, ce qui simplifiera sa production.

Cette étude de localisation permettra de contribuer à la compréhension du rôle de CagL dans le CagSST4 et dans la composition du pilus. L'étude est encore au stade préliminaire et s'inscrit dans un projet à plus long terme.

Conclusion

L'étude des complexes macromoléculaires, de leur composition, de leur assemblage et de leur dynamique est essentielle à la compréhension de leurs fonctions dans les cellules. Ces complexes et leurs composants représentent bien souvent des cibles d'intérêt en recherche clinique et pour le développement de traitements thérapeutiques.

La machinerie de synthèse et d'incorporation des sélénocystéines dans les protéines est gouvernée par des interactions entre facteurs protéiques et ARN. Les sélénoprotéines ainsi synthétisées interviennent dans de nombreuses voies métaboliques. L'altération de la machinerie de traduction ayant des conséquences sur le fonctionnement de l'organisme, il est important de poursuivre les efforts de recherche pour mieux appréhender cette machinerie et les facteurs qui la composent.

Du fait de l'émergence de souches d'*H. pylori* multirésistantes aux antibiotiques, il devient essentiel de développer des stratégies alternatives pour éradiquer la bactérie. Le CagSST4 est un facteur majeur de virulence encodé par la bactérie. En plus de son intérêt pour le développement d'inhibiteurs de virulence, l'étude de ses composants pourrait permettre de prédire la sévérité de l'infection et des pathologies associées.

Pour ces raisons, l'étude de ces deux complexes macromoléculaires furent une priorité de mon projet de doctorat.

Bibliographie

1. Rayman, M.P., *The importance of selenium to human health*. Lancet, 2000. **356**(9225): p. 233-41.
2. Hatfield, D.L., M.J. Berry, and V.N. Gladyshev, *Selenium: Its molecular biology and role in human health, Third Ed*. Springer, 2012.
3. Li, G.S., et al., *Keshan disease: an endemic cardiomyopathy in China*. Hum Pathol, 1985. **16**(6): p. 602-9.
4. Levander, O.A. and M.A. Beck, *Interacting nutritional and infectious etiologies of Keshan disease. Insights from coxsackie virus B-induced myocarditis in mice deficient in selenium or vitamin E*. Biol Trace Elem Res, 1997. **56**(1): p. 5-21.
5. Reeves, W.C., et al., *Reversible cardiomyopathy due to selenium deficiency*. JPEN J Parenter Enteral Nutr, 1989. **13**(6): p. 663-5.
6. Clark, L.C., et al., *Effects of selenium supplementation for cancer prevention in patients with carcinoma of the skin. A randomized controlled trial. Nutritional Prevention of Cancer Study Group*. JAMA, 1996. **276**(24): p. 1957-63.
7. Baum, M.K., et al., *High risk of HIV-related mortality is associated with selenium deficiency*. J Acquir Immune Defic Syndr Hum Retrovirol, 1997. **15**(5): p. 370-4.
8. Knekt, P., et al., *Is low selenium status a risk factor for lung cancer?* Am J Epidemiol, 1998. **148**(10): p. 975-82.
9. Combs, G.F., Jr. and W.P. Gray, *Chemopreventive agents: selenium*. Pharmacol Ther, 1998. **79**(3): p. 179-92.
10. Bellinger, F.P., et al., *Regulation and function of selenoproteins in human disease*. Biochem J, 2009. **422**(1): p. 11-22.
11. Vinceti, M., et al., *Adverse health effects of selenium in humans*. Rev Environ Health, 2001. **16**(4): p. 233-51.
12. Rahmanto, A.S. and M.J. Davies, *Selenium-containing amino acids as direct and indirect antioxidants*. IUBMB Life, 2012. **64**(11): p. 863-71.
13. Berry, M.J., et al., *Substitution of cysteine for selenocysteine in type I iodothyronine deiodinase reduces the catalytic efficiency of the protein but enhances its translation*. Endocrinology, 1992. **131**(4): p. 1848-52.
14. Zhong, L. and A. Holmgren, *Essential role of selenium in the catalytic activities of mammalian thioredoxin reductase revealed by characterization of recombinant enzymes with selenocysteine mutations*. J Biol Chem, 2000. **275**(24): p. 18121-8.
15. Xu, X.M., et al., *Targeted insertion of cysteine by decoding UGA codons with mammalian selenocysteine machinery*. Proc Natl Acad Sci U S A, 2010. **107**(50): p. 21430-4.
16. Taskov, K., et al., *Nematode selenoproteome: the use of the selenocysteine insertion system to decode one codon in an animal genome?* Nucleic Acids Res, 2005. **33**(7): p. 2227-38.
17. Gobler, C.J., et al., *The central role of selenium in the biochemistry and ecology of the harmful pelagophyte, Aureococcus anophagefferens*. ISME J, 2013. **7**(7): p. 1333-43.
18. Kryukov, G.V., et al., *Characterization of mammalian selenoproteomes*. Science, 2003. **300**(5624): p. 1439-43.

19. Lobanov, A.V., et al., *Evolutionary dynamics of eukaryotic selenoproteomes: large selenoproteomes may associate with aquatic life and small with terrestrial life*. Genome Biol, 2007. **8**(9): p. R198.
20. Lobanov, A.V., D.L. Hatfield, and V.N. Gladyshev, *Reduced reliance on the trace element selenium during evolution of mammals*. Genome Biol, 2008. **9**(3): p. R62.
21. Flohe, L., W.A. Gunzler, and H.H. Schock, *Glutathione peroxidase: a selenoenzyme*. FEBS Lett, 1973. **32**(1): p. 132-4.
22. Papp, L.V., et al., *From selenium to selenoproteins: synthesis, identity, and their role in human health*. Antioxid Redox Signal, 2007. **9**(7): p. 775-806.
23. Moghadaszadeh, B., et al., *Mutations in SEPNI cause congenital muscular dystrophy with spinal rigidity and restrictive respiratory syndrome*. Nat Genet, 2001. **29**(1): p. 17-8.
24. Rederstorff, M., A. Krol, and A. Lescure, *Understanding the importance of selenium and selenoproteins in muscle function*. Cell Mol Life Sci, 2006. **63**(1): p. 52-9.
25. Osawa, S., et al., *Recent evidence for evolution of the genetic code*. Microbiol Rev, 1992. **56**(1): p. 229-64.
26. Lee, B.J., et al., *Identification of a selenocysteyl-tRNA(Ser) in mammalian cells that recognizes the nonsense codon, UGA*. J Biol Chem, 1989. **264**(17): p. 9724-7.
27. Leinfelder, W., et al., *Gene for a novel tRNA species that accepts L-serine and cotranslationally inserts selenocysteine*. Nature, 1988. **331**(6158): p. 723-5.
28. Hatfield, D.L., et al., *Selenocysteine tRNA isoacceptors as central components in selenoprotein biosynthesis in eukaryotes*, in *Selenium in Biology and Human Health*, R.F. Burk, Editor. 1994, Springer New York. p. 25-44.
29. Bock, A., et al., *Selenoprotein synthesis: an expansion of the genetic code*. Trends Biochem Sci, 1991. **16**(12): p. 463-7.
30. Carlson, B.A., et al., *Mammalian and other eukaryotic tRNAs*, in *Selenium: Its Molecular Biology and Role in Human Health*, D. Hatfield, M.J. Berry, and V.N. Gladyshev, Editors. 2006, Springer US. p. 29-37
31. Allmang, C. and A. Krol, *Selenoprotein synthesis: UGA does not end the story*. Biochimie, 2006. **88**(11): p. 1561-71.
32. Bosl, M.R., et al., *Early embryonic lethality caused by targeted disruption of the mouse selenocysteine tRNA gene (Trsp)*. Proc Natl Acad Sci U S A, 1997. **94**(11): p. 5531-4.
33. Chittum, H.S., et al., *Selenocysteine tRNA[Ser]^{Sec} levels and selenium-dependent glutathione peroxidase activity in mouse embryonic stem cells heterozygous for a targeted mutation in the tRNA[Ser]^{Sec} gene*. Biochemistry, 1997. **36**(28): p. 8634-9.
34. Diamond, A.M., et al., *Dietary selenium affects methylation of the wobble nucleoside in the anticodon of selenocysteine tRNA([Ser]^{Sec})*. J Biol Chem, 1993. **268**(19): p. 14215-23.
35. Carlson, B.A., et al., *Selective rescue of selenoprotein expression in mice lacking a highly specialized methyl group in selenocysteine tRNA*. J Biol Chem, 2005. **280**(7): p. 5542-8.
36. Carlson, B.A., et al., *Selective restoration of the selenoprotein population in a mouse hepatocyte selenoproteinless background with different mutant selenocysteine tRNAs lacking Um34*. J Biol Chem, 2007. **282**(45): p. 32591-602.

37. Glass, R.S., et al., *Monoselenophosphate: synthesis, characterization, and identity with the prokaryotic biological selenium donor, compound SePX*. *Biochemistry*, 1993. **32**(47): p. 12555-9.
38. Forchhammer, K. and A. Bock, *Selenocysteine synthase from Escherichia coli. Analysis of the reaction sequence*. *J Biol Chem*, 1991. **266**(10): p. 6324-8.
39. Gelpi, C., E.J. Sontheimer, and J.L. Rodriguez-Sanchez, *Autoantibodies against a serine tRNA-protein complex implicated in cotranslational selenocysteine insertion*. *Proc Natl Acad Sci U S A*, 1992. **89**(20): p. 9739-43.
40. Kernebeck, T., A.W. Lohse, and J. Grotzinger, *A bioinformatical approach suggests the function of the autoimmune hepatitis target antigen soluble liver antigen/liver pancreas*. *Hepatology*, 2001. **34**(2): p. 230-3.
41. Xu, X.M., et al., *Evidence for direct roles of two additional factors, SECp43 and soluble liver antigen, in the selenoprotein synthesis machinery*. *J Biol Chem*, 2005. **280**(50): p. 41568-75.
42. Small-Howard, A., et al., *Supramolecular complexes mediate selenocysteine incorporation in vivo*. *Mol Cell Biol*, 2006. **26**(6): p. 2337-46.
43. Palioura, S., et al., *The human SepSecS-tRNA^{Sec} complex reveals the mechanism of selenocysteine formation*. *Science*, 2009. **325**(5938): p. 321-5.
44. Itoh, Y., et al., *Decameric Sela*tRNA^(Sec) ring structure reveals mechanism of bacterial selenocysteine formation*. *Science*, 2013. **340**(6128): p. 75-8.
45. Maenpaa, P.H. and M.R. Bernfield, *A specific hepatic transfer RNA for phosphoserine*. *Proc Natl Acad Sci U S A*, 1970. **67**(2): p. 688-95.
46. Carlson, B.A., et al., *Identification and characterization of phosphoseryl-tRNA^{[Ser]^{Sec} kinase}*. *Proc Natl Acad Sci U S A*, 2004. **101**(35): p. 12848-53.
47. Kim, I.Y. and T.C. Stadtman, *Selenophosphate synthetase: detection in extracts of rat tissues by immunoblot assay and partial purification of the enzyme from the archaean Methanococcus vannielii*. *Proc Natl Acad Sci U S A*, 1995. **92**(17): p. 7710-3.
48. Low, S.C., J.W. Harney, and M.J. Berry, *Cloning and functional characterization of human selenophosphate synthetase, an essential component of selenoprotein synthesis*. *J Biol Chem*, 1995. **270**(37): p. 21659-64.
49. Guimaraes, M.J., et al., *Identification of a novel selD homolog from eukaryotes, bacteria, and archaea: is there an autoregulatory mechanism in selenocysteine metabolism?* *Proc Natl Acad Sci U S A*, 1996. **93**(26): p. 15086-91.
50. Xu, X.M., et al., *Biosynthesis of selenocysteine on its tRNA in eukaryotes*. *PLoS Biol*, 2007. **5**(1): p. e4.
51. Tamura, T., et al., *Selenophosphate synthetase genes from lung adenocarcinoma cells: Sps1 for recycling L-selenocysteine and Sps2 for selenite assimilation*. *Proc Natl Acad Sci U S A*, 2004. **101**(46): p. 16162-7.
52. Labunskyy, V.M., D.L. Hatfield, and V.N. Gladyshev, *Selenoproteins: molecular pathways and physiological roles*. *Physiol Rev*, 2014. **94**(3): p. 739-77.
53. Low, S.C. and M.J. Berry, *Knowing when not to stop: selenocysteine incorporation in eukaryotes*. *Trends Biochem Sci*, 1996. **21**(6): p. 203-8.
54. Martin, G.W., 3rd, J.W. Harney, and M.J. Berry, *Selenocysteine incorporation in eukaryotes: insights into mechanism and efficiency from sequence, structure, and spacing proximity studies of the type I deiodinase SECIS element*. *RNA*, 1996. **2**(2): p. 171-82.

55. Bock, A., et al., *Selenium metabolism in prokaryotes*, in *Selenium: Its Molecular Biology and Role in Human Health*, D.L. Hatfield, M.J. Berry, and V.N. Gladyshev, Editors. 2006, Springer US. p. 9–28.
56. Walczak, R., et al., *A novel RNA structural motif in the selenocysteine insertion element of eukaryotic selenoprotein mRNAs*. RNA, 1996. **2**(4): p. 367-79.
57. Grundner-Culemann, E., et al., *Two distinct SECIS structures capable of directing selenocysteine incorporation in eukaryotes*. RNA, 1999. **5**(5): p. 625-35.
58. Korotkov, K.V., et al., *Mammalian selenoprotein in which selenocysteine (Sec) incorporation is supported by a new form of Sec insertion sequence element*. Mol Cell Biol, 2002. **22**(5): p. 1402-11.
59. Fletcher, J.E., et al., *The selenocysteine incorporation machinery: interactions between the SECIS RNA and the SECIS-binding protein SBP2*. RNA, 2001. **7**(10): p. 1442-53.
60. Latreche, L., et al., *Novel structural determinants in human SECIS elements modulate the translational recoding of UGA as selenocysteine*. Nucleic Acids Res, 2009. **37**(17): p. 5868-80.
61. Berry, M.J., et al., *Recognition of UGA as a selenocysteine codon in type I deiodinase requires sequences in the 3' untranslated region*. Nature, 1991. **353**(6341): p. 273-6.
62. Fagegaltier, D., et al., *Structural analysis of new local features in SECIS RNA hairpins*. Nucleic Acids Res, 2000. **28**(14): p. 2679-89.
63. Martin, G.W., 3rd, J.W. Harney, and M.J. Berry, *Functionality of mutations at conserved nucleotides in eukaryotic SECIS elements is determined by the identity of a single nonconserved nucleotide*. RNA, 1998. **4**(1): p. 65-73.
64. Berry, M.J., et al., *Functional characterization of the eukaryotic SECIS elements which direct selenocysteine insertion at UGA codons*. EMBO J, 1993. **12**(8): p. 3315-22.
65. Kryukov, G.V. and V.N. Gladyshev, *Selenium metabolism in zebrafish: multiplicity of selenoprotein genes and expression of a protein containing 17 selenocysteine residues*. Genes Cells, 2000. **5**(12): p. 1049-60.
66. Howard, M.T., et al., *A recoding element that stimulates decoding of UGA codons by Sec tRNA[Ser]Sec*. RNA, 2007. **13**(6): p. 912-20.
67. Copeland, P.R. and D.M. Driscoll, *Purification, redox sensitivity, and RNA binding properties of SECIS-binding protein 2, a protein involved in selenoprotein biosynthesis*. J Biol Chem, 1999. **274**(36): p. 25447-54.
68. Copeland, P.R., et al., *A novel RNA binding protein, SBP2, is required for the translation of mammalian selenoprotein mRNAs*. EMBO J, 2000. **19**(2): p. 306-14.
69. Lescure, A., et al., *cDNA cloning, expression pattern and RNA binding analysis of human selenocysteine insertion sequence (SECIS) binding protein 2*. Gene, 2002. **291**(1-2): p. 279-85.
70. Copeland, P.R., V.A. Stepanik, and D.M. Driscoll, *Insight into mammalian selenocysteine insertion: domain structure and ribosome binding properties of Sec insertion sequence binding protein 2*. Mol Cell Biol, 2001. **21**(5): p. 1491-8.
71. Papp, L.V., et al., *The redox state of SECIS binding protein 2 controls its localization and selenocysteine incorporation function*. Mol Cell Biol, 2006. **26**(13): p. 4895-910.
72. Seeher, S., et al., *Secisbp2 is essential for embryonic development and enhances selenoprotein expression*. Antioxid Redox Signal, 2014. **21**(6): p. 835-49.

73. Di Cosmo, C., et al., *Clinical and molecular characterization of a novel selenocysteine insertion sequence-binding protein 2 (SBP2) gene mutation (R128X)*. J Clin Endocrinol Metab, 2009. **94**(10): p. 4003-9.
74. Dumitrescu, A.M., et al., *Mutations in SECISBP2 result in abnormal thyroid hormone metabolism*. Nat Genet, 2005. **37**(11): p. 1247-52.
75. Donovan, J. and P.R. Copeland, *Threading the needle: getting selenocysteine into proteins*. Antioxid Redox Signal, 2010. **12**(7): p. 881-92.
76. Takeuchi, A., et al., *A short motif in Drosophila SECIS Binding Protein 2 provides differential binding affinity to SECIS RNA hairpins*. Nucleic Acids Res, 2009. **37**(7): p. 2126-41.
77. de Jesus, L.A., et al., *Nuclear assembly of UGA decoding complexes on selenoprotein mRNAs: a mechanism for eluding nonsense-mediated decay?* Mol Cell Biol, 2006. **26**(5): p. 1795-805.
78. Donovan, J., et al., *A novel protein domain induces high affinity selenocysteine insertion sequence binding and elongation factor recruitment*. J Biol Chem, 2008. **283**(50): p. 35129-39.
79. Caban, K., S.A. Kinzy, and P.R. Copeland, *The L7Ae RNA binding motif is a multifunctional domain required for the ribosome-dependent Sec incorporation activity of Sec insertion sequence binding protein 2*. Mol Cell Biol, 2007. **27**(18): p. 6350-60.
80. Gonzalez-Flores, J.N., et al., *The selenocysteine-specific elongation factor contains a novel and multi-functional domain*. J Biol Chem, 2012.
81. Donovan, J. and P.R. Copeland, *Evolutionary history of selenocysteine incorporation from the perspective of SECIS binding proteins*. BMC Evol Biol, 2009. **9**: p. 229.
82. Donovan, J. and P.R. Copeland, *Selenocysteine insertion sequence binding protein 2L is implicated as a novel post-transcriptional regulator of selenoprotein expression*. PLoS One, 2012. **7**(4): p. e35581.
83. Leibundgut, M., et al., *Selenocysteine tRNA-specific elongation factor SelB is a structural chimaera of elongation and initiation factors*. EMBO J, 2005. **24**(1): p. 11-22.
84. Fagegaltier, D., et al., *Characterization of mSelB, a novel mammalian elongation factor for selenoprotein translation*. EMBO J, 2000. **19**(17): p. 4796-805.
85. Tujebajeva, R.M., et al., *Decoding apparatus for eukaryotic selenocysteine insertion*. EMBO Rep, 2000. **1**(2): p. 158-63.
86. Itoh, Y., et al., *Crystal structure of human selenocysteine tRNA*. Nucleic Acids Res, 2009. **37**(18): p. 6259-68.
87. Macias, S., et al., *L30 binds the nascent RPL30 transcript to repress U2 snRNP recruitment*. Mol Cell, 2008. **30**(6): p. 732-42.
88. Chavatte, L., B.A. Brown, and D.M. Driscoll, *Ribosomal protein L30 is a component of the UGA-selenocysteine recoding machinery in eukaryotes*. Nat Struct Mol Biol, 2005. **12**(5): p. 408-16.
89. Ding, F. and P.J. Grabowski, *Identification of a protein component of a mammalian tRNA(Sec) complex implicated in the decoding of UGA as selenocysteine*. RNA, 1999. **5**(12): p. 1561-9.
90. Kinzy, S.A., K. Caban, and P.R. Copeland, *Characterization of the SECIS binding protein 2 complex required for the co-translational insertion of selenocysteine in mammals*. Nucleic Acids Res, 2005. **33**(16): p. 5172-80.

91. Schmidt, R.L. and M. Simonovic, *Synthesis and decoding of selenocysteine and human health*. Croat Med J, 2012. **53**(6): p. 535-50.
92. Agamy, O., et al., *Mutations disrupting selenocysteine formation cause progressive cerebello-cerebral atrophy*. Am J Hum Genet, 2010. **87**(4): p. 538-44.
93. Allamand, V., et al., *A single homozygous point mutation in a 3'untranslated region motif of selenoprotein N mRNA causes SEPNI-related myopathy*. EMBO Rep, 2006. **7**(4): p. 450-4.
94. Maiti, B., et al., *A mutation in the SEPNI selenocysteine redefinition element (SRE) reduces selenocysteine incorporation and leads to SEPNI-related myopathy*. Hum Mutat, 2009. **30**(3): p. 411-6.
95. Moriarty, P.M., C.C. Reddy, and L.E. Maquat, *Selenium deficiency reduces the abundance of mRNA for Se-dependent glutathione peroxidase 1 by a UGA-dependent mechanism likely to be nonsense codon-mediated decay of cytoplasmic mRNA*. Mol Cell Biol, 1998. **18**(5): p. 2932-9.
96. Fletcher, J.E., P.R. Copeland, and D.M. Driscoll, *Polysome distribution of phospholipid hydroperoxide glutathione peroxidase mRNA: evidence for a block in elongation at the UGA/selenocysteine codon*. RNA, 2000. **6**(11): p. 1573-84.
97. Martin, G.W., 3rd and M.J. Berry, *Selenocysteine codons decrease polysome association on endogenous selenoprotein mRNAs*. Genes Cells, 2001. **6**(2): p. 121-9.
98. Hatfield, D., et al., *Selenium induces changes in the selenocysteine tRNA[Ser]Sec population in mammalian cells*. Nucleic Acids Res, 1991. **19**(4): p. 939-43.
99. Jameson, R.R. and A.M. Diamond, *A regulatory role for Sec tRNA[Ser]Sec in selenoprotein synthesis*. RNA, 2004. **10**(7): p. 1142-52.
100. Budiman, M.E., et al., *Eukaryotic initiation factor 4a3 is a selenium-regulated RNA-binding protein that selectively inhibits selenocysteine incorporation*. Mol Cell, 2009. **35**(4): p. 479-89.
101. Wu, R., Q. Shen, and P.E. Newburger, *Recognition and binding of the human selenocysteine insertion sequence by nucleolin*. J Cell Biochem, 2000. **77**(3): p. 507-16.
102. Turanov, A.A., et al., *UGA codon position-dependent incorporation of selenocysteine into mammalian selenoproteins*. Nucleic Acids Res, 2013. **41**(14): p. 6952-9.
103. Nagy, E. and L.E. Maquat, *A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance*. Trends Biochem Sci, 1998. **23**(6): p. 198-9.
104. Shetty, S.P. and P.R. Copeland, *Selenocysteine incorporation: A trump card in the game of mRNA decay*. Biochimie, 2015. **114**: p. 97-101.
105. Gehring, N.H., et al., *Exon-junction complex components specify distinct routes of nonsense-mediated mRNA decay with differential cofactor requirements*. Mol Cell, 2005. **20**(1): p. 65-75.
106. Miniard, A.C., et al., *Nucleolin binds to a subset of selenoprotein mRNAs and regulates their expression*. Nucleic Acids Res, 2010. **38**(14): p. 4807-20.
107. Squires, J.E., et al., *SBP2 binding affinity is a major determinant in differential selenoprotein mRNA translation and sensitivity to nonsense-mediated decay*. Mol Cell Biol, 2007. **27**(22): p. 7848-55.
108. Holcombe, C., *Helicobacter pylori: the African enigma*. Gut, 1992. **33**(4): p. 429-31.

109. Breckan, R.K., et al., *The All-Age Prevalence of Helicobacter pylori Infection and Potential Transmission Routes. A Population-Based Study*. *Helicobacter*, 2016. **21**(6): p. 586-595.
110. Kusters, J.G., A.H. van Vliet, and E.J. Kuipers, *Pathogenesis of Helicobacter pylori infection*. *Clin Microbiol Rev*, 2006. **19**(3): p. 449-90.
111. Ford, A.C., et al., *Eradication therapy for peptic ulcer disease in Helicobacter pylori-positive people*. *Cochrane Database Syst Rev*, 2016. **4**: p. CD003840.
112. Malfertheiner, P., F.K. Chan, and K.E. McColl, *Peptic ulcer disease*. *Lancet*, 2009. **374**(9699): p. 1449-61.
113. Colquhoun, A., et al., *Global patterns of cardia and non-cardia gastric cancer incidence in 2012*. *Gut*, 2015. **64**(12): p. 1881-8.
114. Kuipers, E.J., *Review article: exploring the link between Helicobacter pylori and gastric cancer*. *Aliment Pharmacol Ther*, 1999. **13 Suppl 1**: p. 3-11.
115. Malfertheiner, P., L.C. Fry, and K. Monkemuller, *Can gastric cancer be prevented by Helicobacter pylori eradication?* *Best Pract Res Clin Gastroenterol*, 2006. **20**(4): p. 709-19.
116. Plummer, M., et al., *Global burden of gastric cancer attributable to Helicobacter pylori*. *Int J Cancer*, 2015. **136**(2): p. 487-90.
117. Genta, R.M., H.W. Hamner, and D.Y. Graham, *Gastric lymphoid follicles in Helicobacter pylori infection: frequency, distribution, and response to triple therapy*. *Hum Pathol*, 1993. **24**(6): p. 577-83.
118. Du, M.Q. and J.C. Atherton, *Molecular subtyping of gastric MALT lymphomas: implications for prognosis and management*. *Gut*, 2006. **55**(6): p. 886-93.
119. Weeks, D.L., et al., *A H⁺-gated urea channel: the link between Helicobacter pylori urease and gastric colonization*. *Science*, 2000. **287**(5452): p. 482-5.
120. Wen, Y., et al., *The HP0165-HP0166 two-component system (ArsRS) regulates acid-induced expression of HP1186 alpha-carbonic anhydrase in Helicobacter pylori by activating the pH-dependent promoter*. *J Bacteriol*, 2007. **189**(6): p. 2426-34.
121. Pflock, M., et al., *Acid-induced activation of the urease promoters is mediated directly by the ArsRS two-component system of Helicobacter pylori*. *Infect Immun*, 2005. **73**(10): p. 6437-45.
122. Pflock, M., et al., *Characterization of the ArsRS regulon of Helicobacter pylori, involved in acid adaptation*. *J Bacteriol*, 2006. **188**(10): p. 3449-62.
123. Basso, D., M. Plebani, and J.G. Kusters, *Pathogenesis of Helicobacter pylori infection*. *Helicobacter*, 2010. **15 Suppl 1**: p. 14-20.
124. Tsugawa, H., et al., *Reactive oxygen species-induced autophagic degradation of Helicobacter pylori CagA is specifically suppressed in cancer stem-like cells*. *Cell Host Microbe*, 2012. **12**(6): p. 764-77.
125. Oertli, M., et al., *Helicobacter pylori gamma-glutamyl transpeptidase and vacuolating cytotoxin promote gastric persistence and immune tolerance*. *Proc Natl Acad Sci U S A*, 2013. **110**(8): p. 3047-52.
126. Gebert, B., et al., *Helicobacter pylori vacuolating cytotoxin inhibits T lymphocyte activation*. *Science*, 2003. **301**(5636): p. 1099-102.
127. Pachathundikandi, S.K., et al., *Interplay of the Gastric Pathogen Helicobacter pylori with Toll-Like Receptors*. *Biomed Res Int*, 2015. **2015**: p. 192420.

128. Sugano, K., et al., *Kyoto global consensus report on Helicobacter pylori gastritis*. Gut, 2015. **64**(9): p. 1353-67.
129. Malfertheiner, P., et al., *Management of Helicobacter pylori infection--the Maastricht IV/ Florence Consensus Report*. Gut, 2012. **61**(5): p. 646-64.
130. Wu, W., Y. Yang, and G. Sun, *Recent Insights into Antibiotic Resistance in Helicobacter pylori Eradication*. Gastroenterol Res Pract, 2012. **2012**: p. 723183.
131. Costa, T.R., et al., *Secretion systems in Gram-negative bacteria: structural and mechanistic insights*. Nat Rev Microbiol, 2015. **13**(6): p. 343-59.
132. Fronzes, R., P.J. Christie, and G. Waksman, *The structural biology of type IV secretion systems*. Nat Rev Microbiol, 2009. **7**(10): p. 703-14.
133. Alvarez-Martinez, C.E. and P.J. Christie, *Biological diversity of prokaryotic type IV secretion systems*. Microbiol Mol Biol Rev, 2009. **73**(4): p. 775-808.
134. Dreiseikelmann, B., *Translocation of DNA across bacterial membranes*. Microbiol Rev, 1994. **58**(3): p. 293-316.
135. Christie, P.J., *Type IV secretion: the Agrobacterium VirB/D4 and related conjugation systems*. Biochim Biophys Acta, 2004. **1694**(1-3): p. 219-34.
136. Hofreuter, D., S. Odenbreit, and R. Haas, *Natural transformation competence in Helicobacter pylori is mediated by the basic components of a type IV secretion system*. Mol Microbiol, 2001. **41**(2): p. 379-91.
137. Hamilton, H.L., et al., *Neisseria gonorrhoeae secretes chromosomal DNA via a novel type IV secretion system*. Mol Microbiol, 2005. **55**(6): p. 1704-21.
138. Wallden, K., A. Rivera-Calzada, and G. Waksman, *Type IV secretion systems: versatility and diversity in function*. Cell Microbiol, 2010. **12**(9): p. 1203-12.
139. Odenbreit, S., et al., *Translocation of Helicobacter pylori CagA into gastric epithelial cells by type IV secretion*. Science, 2000. **287**(5457): p. 1497-500.
140. Cover, T.L., *Helicobacter pylori Diversity and Gastric Cancer Risk*. MBio, 2016. **7**(1): p. e01869-15.
141. Hatakeyama, M., *Helicobacter pylori CagA and gastric cancer: a paradigm for hit-and-run carcinogenesis*. Cell Host Microbe, 2014. **15**(3): p. 306-16.
142. Yong, X., et al., *Helicobacter pylori virulence factor CagA promotes tumorigenesis of gastric cancer via multiple signaling pathways*. Cell Commun Signal, 2015. **13**: p. 30.
143. Backert, S., et al., *The Helicobacter pylori Type IV Secretion System Encoded by the cag Pathogenicity Island: Architecture, Function, and Signaling*, in *Type IV Secretion in Gram-Negative and Gram-Positive Bacteria*, S. Backert and E. Grohmann, Editors. 2017, Springer International Publishing. p. 188-209.
144. Fischer, W., et al., *Systematic mutagenesis of the Helicobacter pylori cag pathogenicity island: essential genes for CagA translocation in host cells and induction of interleukin-8*. Mol Microbiol, 2001. **42**(5): p. 1337-48.
145. Frick-Cheng, A.E., et al., *Molecular and Structural Analysis of the Helicobacter pylori cag Type IV Secretion System Core Complex*. MBio, 2016. **7**(1): p. e02001-15.
146. Smart, J., et al., *Cag-delta (Cag3) protein from the Helicobacter pylori 26695 cag type IV secretion system forms ring-like supramolecular assemblies*. FEMS Microbiol Lett, 2017. **364**(1).
147. Low, H.H., et al., *Structure of a type IV secretion system*. Nature, 2014. **508**(7497): p. 550-553.

148. Rohde, M., et al., *A novel sheathed surface organelle of the Helicobacter pylori cag type IV secretion system*. Mol Microbiol, 2003. **49**(1): p. 219-34.
149. Chang, Y.W., et al., *In Vivo Structures of the Helicobacter pylori cag Type IV Secretion System*. Cell Rep, 2018. **23**(3): p. 673-681.
150. Andrzejewska, J., et al., *Characterization of the pilin ortholog of the Helicobacter pylori type IV cag pathogenicity apparatus, a surface-associated protein expressed during infection*. J Bacteriol, 2006. **188**(16): p. 5865-77.
151. Johnson, E.M., et al., *Genes required for assembly of pili associated with the Helicobacter pylori cag type IV secretion system*. Infect Immun, 2014. **82**(8): p. 3457-70.
152. Kwok, T., et al., *Helicobacter exploits integrin for type IV secretion and kinase activation*. Nature, 2007. **449**(7164): p. 862-6.
153. Shaffer, C.L., et al., *Helicobacter pylori exploits a unique repertoire of type IV secretion system components for pilus assembly at the bacteria-host cell interface*. PLoS Pathog, 2011. **7**(9): p. e1002237.
154. Fischer, W., *Assembly and molecular mode of action of the Helicobacter pylori Cag type IV secretion apparatus*. FEBS J, 2011. **278**(8): p. 1203-12.
155. Bonsor, D.A., et al., *Characterization of the translocation-competent complex between the Helicobacter pylori oncogenic protein CagA and the accessory protein CagF*. J Biol Chem, 2013. **288**(46): p. 32897-909.
156. Zhong, Q., et al., *Characterization of peptidoglycan hydrolase in Cag pathogenicity island of Helicobacter pylori*. Mol Biol Rep, 2011. **38**(1): p. 503-9.
157. Viala, J., et al., *Nod1 responds to peptidoglycan delivered by the Helicobacter pylori cag pathogenicity island*. Nat Immunol, 2004. **5**(11): p. 1166-74.
158. Varga, M.G., et al., *Pathogenic Helicobacter pylori strains translocate DNA and activate TLR9 via the cancer-associated cag type IV secretion system*. Oncogene, 2016. **35**(48): p. 6262-6269.
159. Gall, A., et al., *TIFA Signaling in Gastric Epithelial Cells Initiates the cag Type 4 Secretion System-Dependent Innate Immune Response to Helicobacter pylori Infection*. MBio, 2017. **8**(4).
160. Yeo, H.J., et al., *Crystal structure of the hexameric traffic ATPase of the Helicobacter pylori type IV secretion system*. Mol Cell, 2000. **6**(6): p. 1461-72.
161. Savvides, S.N., et al., *VirB11 ATPases are dynamic hexameric assemblies: new insights into bacterial type IV secretion*. EMBO J, 2003. **22**(9): p. 1969-80.
162. Terradot, L. and G. Waksman, *Architecture of the Helicobacter pylori Cag-type IV secretion system*. FEBS J, 2011. **278**(8): p. 1213-22.
163. Parsonnet, J., et al., *Risk for gastric cancer in people with CagA positive or CagA negative Helicobacter pylori infection*. Gut, 1997. **40**(3): p. 297-301.
164. Luo, B.H., C.V. Carman, and T.A. Springer, *Structural basis of integrin regulation and signaling*. Annu Rev Immunol, 2007. **25**: p. 619-47.
165. Stewart, P.L. and G.R. Nemerow, *Cell integrins: commonly used receptors for diverse viral pathogens*. Trends Microbiol, 2007. **15**(11): p. 500-7.
166. Hauck, C.R., M. Borisova, and P. Muenzner, *Exploitation of integrin function by pathogenic microbes*. Curr Opin Cell Biol, 2012. **24**(5): p. 637-44.
167. Jimenez-Soto, L.F., et al., *Helicobacter pylori type IV secretion apparatus exploits beta1 integrin in a novel RGD-independent manner*. PLoS Pathog, 2009. **5**(12): p. e1000684.

168. Wiedemann, T., et al., *Helicobacter pylori CagL dependent induction of gastrin expression via a novel alphavbeta5-integrin-integrin linked kinase signalling complex*. Gut, 2012. **61**(7): p. 986-96.
169. Barden, S. and H.H. Niemann, *Adhesion of several cell lines to Helicobacter pylori CagL is mediated by integrin alphaVbeta6 via an RGDLXXL motif*. J Mol Biol, 2015. **427**(6 Pt B): p. 1304-1315.
170. Barden, S., et al., *A helical RGD motif promoting cell adhesion: crystal structures of the Helicobacter pylori type IV secretion system pilus protein CagL*. Structure, 2013. **21**(11): p. 1931-41.
171. Barden, S., et al., *Structure of a three-dimensional domain-swapped dimer of the Helicobacter pylori type IV secretion system pilus protein CagL*. Acta Crystallogr D Biol Crystallogr, 2014. **70**(Pt 5): p. 1391-400.
172. Bonsor, D.A., et al., *Integrin engagement by the helical RGD motif of the Helicobacter pylori CagL protein is regulated by pH-induced displacement of a neighboring helix*. J Biol Chem, 2015. **290**(20): p. 12929-40.
173. Choi, J.M., et al., *Crystal structure of CagL from Helicobacter pylori K74 strain*. Biochem Biophys Res Commun, 2015. **460**(4): p. 964-70.
174. Conradi, J., et al., *An RGD helper sequence in CagL of Helicobacter pylori assists in interactions with integrins and injection of CagA*. Front Cell Infect Microbiol, 2012. **2**: p. 70.
175. Tegtmeier, N., et al., *A small fibronectin-mimicking protein from bacteria induces cell spreading and focal adhesion formation*. J Biol Chem, 2010. **285**(30): p. 23515-26.
176. Gorrell, R.J., et al., *A novel NOD1- and CagA-independent pathway of interleukin-8 induction mediated by the Helicobacter pylori type IV secretion system*. Cell Microbiol, 2013. **15**(4): p. 554-70.
177. Jimenez-Soto, L.F. and R. Haas, *The CagA toxin of Helicobacter pylori: abundant production but relatively low amount translocated*. Sci Rep, 2016. **6**: p. 23227.
178. Kaplan-Turkoz, B., et al., *Structural insights into Helicobacter pylori oncoprotein CagA interaction with beta1 integrin*. Proc Natl Acad Sci U S A, 2012. **109**(36): p. 14640-5.
179. Murata-Kamiya, N., et al., *Helicobacter pylori CagA interacts with E-cadherin and deregulates the beta-catenin signal that promotes intestinal transdifferentiation in gastric epithelial cells*. Oncogene, 2007. **26**(32): p. 4617-26.
180. Covacci, A. and R. Rappuoli, *Tyrosine-phosphorylated bacterial proteins: Trojan horses for the host cell*. J Exp Med, 2000. **191**(4): p. 587-92.
181. Backert, S., N. Tegtmeier, and M. Selbach, *The versatility of Helicobacter pylori CagA effector protein functions: The master key hypothesis*. Helicobacter, 2010. **15**(3): p. 163-76.
182. Lind, J., et al., *Systematic analysis of phosphotyrosine antibodies recognizing single phosphorylated EPIYA-motifs in CagA of Western-type Helicobacter pylori strains*. PLoS One, 2014. **9**(5): p. e96488.
183. Lind, J., et al., *Systematic analysis of phosphotyrosine antibodies recognizing single phosphorylated EPIYA-motifs in CagA of East Asian-type Helicobacter pylori strains*. BMC Microbiol, 2016. **16**(1): p. 201.
184. Amieva, M. and R.M. Peek, Jr., *Pathobiology of Helicobacter pylori-Induced Gastric Cancer*. Gastroenterology, 2016. **150**(1): p. 64-78.

185. Ren, S., et al., *Structural basis and functional consequence of Helicobacter pylori CagA multimerization in cells*. J Biol Chem, 2006. **281**(43): p. 32344-52.
186. N., T. and B. S., *Molecular Pathogenesis and Signal Transduction by Helicobacter pylori*. Springer, 2016. **400**.
187. Naumann, M., et al., *Helicobacter pylori: A Paradigm Pathogen for Subverting Host Cell Signal Transmission*. Trends Microbiol, 2017. **25**(4): p. 316-328.
188. Backert, S. and N. Tegtmeyer, *Type IV Secretion and Signal Transduction of Helicobacter pylori CagA through Interactions with Host Cell Receptors*. Toxins (Basel), 2017. **9**(4).
189. Backert, S. and M.J. Blaser, *The Role of CagA in the Gastric Biology of Helicobacter pylori*. Cancer Res, 2016. **76**(14): p. 4028-31.
190. Tegtmeyer, N., et al., *Subversion of host kinases: a key network in cellular signaling hijacked by Helicobacter pylori CagA*. Mol Microbiol, 2017. **105**(3): p. 358-372.
191. Mueller, D., et al., *c-Src and c-Abl kinases control hierarchic phosphorylation and function of the CagA effector protein in Western and East Asian Helicobacter pylori strains*. J Clin Invest, 2012. **122**(4): p. 1553-66.
192. Higashi, H., et al., *Biological activity of the Helicobacter pylori virulence factor CagA is determined by variation in the tyrosine phosphorylation sites*. Proc Natl Acad Sci U S A, 2002. **99**(22): p. 14428-33.
193. Selbach, M., et al., *Host cell interactome of tyrosine-phosphorylated bacterial proteins*. Cell Host Microbe, 2009. **5**(4): p. 397-403.
194. Franco, A.T., et al., *Activation of beta-catenin by carcinogenic Helicobacter pylori*. Proc Natl Acad Sci U S A, 2005. **102**(30): p. 10646-51.
195. Tsang, Y.H., et al., *Helicobacter pylori CagA targets gastric tumor suppressor RUNX3 for proteasome-mediated degradation*. Oncogene, 2010. **29**(41): p. 5643-50.
196. Franco, A.T., et al., *Regulation of gastric carcinogenesis by Helicobacter pylori virulence factors*. Cancer Res, 2008. **68**(2): p. 379-87.
197. Varona, C. and F. Mégrauda, *Infection à Helicobacter pylori et cancer gastrique*. Revue Francophone des Laboratoires, 2013(456): p. 67-76.
198. Leunk, R.D., et al., *Cytotoxic activity in broth-culture filtrates of Campylobacter pylori*. J Med Microbiol, 1988. **26**(2): p. 93-9.
199. Boquet, P. and V. Ricci, *Intoxication strategy of Helicobacter pylori VacA toxin*. Trends Microbiol, 2012. **20**(4): p. 165-74.
200. Cover, T.L. and S.R. Blanke, *Helicobacter pylori VacA, a paradigm for toxin multifunctionality*. Nat Rev Microbiol, 2005. **3**(4): p. 320-32.
201. Ricci, V., *Relationship between VacA Toxin and Host Cell Autophagy in Helicobacter pylori Infection of the Human Stomach: A Few Answers, Many Questions*. Toxins (Basel), 2016. **8**(7).
202. Utsch, C. and R. Haas, *VacA's Induction of VacA-Containing Vacuoles (VCVs) and Their Immunomodulatory Activities on Human T Cells*. Toxins (Basel), 2016. **8**(6).
203. Atherton, J.C., et al., *Mosaicism in vacuolating cytotoxin alleles of Helicobacter pylori. Association of specific vacA types with cytotoxin production and peptic ulceration*. J Biol Chem, 1995. **270**(30): p. 17771-7.
204. Foegeding, N.J., et al., *An Overview of Helicobacter pylori VacA Toxin Biology*. Toxins (Basel), 2016. **8**(6).

205. Rhead, J.L., et al., *A new Helicobacter pylori vacuolating cytotoxin determinant, the intermediate region, is associated with gastric cancer*. Gastroenterology, 2007. **133**(3): p. 926-36.
206. Argent, R.H., et al., *Functional association between the Helicobacter pylori virulence factors VacA and CagA*. J Med Microbiol, 2008. **57**(Pt 2): p. 145-50.
207. Ilver, D., et al., *Helicobacter pylori adhesin binding fucosylated histo-blood group antigens revealed by retagging*. Science, 1998. **279**(5349): p. 373-7.
208. Zambon, C.F., et al., *Helicobacter pylori babA2, cagA, and sl vacA genes work synergistically in causing intestinal metaplasia*. J Clin Pathol, 2003. **56**(4): p. 287-91.
209. Mahdavi, J., et al., *Helicobacter pylori SabA adhesin in persistent infection and chronic inflammation*. Science, 2002. **297**(5581): p. 573-8.
210. Yamaoka, Y., et al., *Search for putative virulence factors of Helicobacter pylori: the low-molecular-weight (33-35 K) antigen*. Dig Dis Sci, 1998. **43**(7): p. 1482-7.
211. Odenbreit, S., et al., *Genetic and functional characterization of the alpAB gene locus essential for the adhesion of Helicobacter pylori to human gastric tissue*. Mol Microbiol, 1999. **31**(5): p. 1537-48.
212. Javaheri, A., et al., *Helicobacter pylori adhesin HopQ engages in a virulence-enhancing interaction with human CEACAMs*. Nat Microbiol, 2016. **2**: p. 16189.
213. Belogolova, E., et al., *Helicobacter pylori outer membrane protein HopQ identified as a novel T4SS-associated virulence factor*. Cell Microbiol, 2013. **15**(11): p. 1896-912.
214. Lu, H., et al., *Duodenal ulcer promoting gene of Helicobacter pylori*. Gastroenterology, 2005. **128**(4): p. 833-48.
215. van Doorn, L.J., et al., *Clinical relevance of the cagA, vacA, and iceA status of Helicobacter pylori*. Gastroenterology, 1998. **115**(1): p. 58-66.
216. de Bernard, M. and M.M. D'Elia, *The immune modulating activity of the Helicobacter pylori HP-NAP: Friend or foe?* Toxicon, 2010. **56**(7): p. 1186-92.
217. Baron, C., *Antivirulence drugs to target bacterial secretion systems*. Curr Opin Microbiol, 2010. **13**(1): p. 100-5.
218. Finlay, B.B. and S. Falkow, *Common themes in microbial pathogenicity revisited*. Microbiol Mol Biol Rev, 1997. **61**(2): p. 136-69.
219. Marra, A., *Can virulence factors be viable antibacterial targets?* Expert Rev Anti Infect Ther, 2004. **2**(1): p. 61-72.
220. Lee, Y.M., F. Almqvist, and S.J. Hultgren, *Targeting virulence for antimicrobial chemotherapy*. Curr Opin Pharmacol, 2003. **3**(5): p. 513-9.
221. Mota, L.J. and G.R. Cornelis, *The bacterial injection kit: type III secretion systems*. Ann Med, 2005. **37**(4): p. 234-49.
222. Moraes, T.F., T. Spreter, and N.C. Strynadka, *Piecing together the type III injectisome of bacterial pathogens*. Curr Opin Struct Biol, 2008. **18**(2): p. 258-66.
223. Kauppi, A.M., et al., *Targeting bacterial virulence: inhibitors of type III secretion in Yersinia*. Chem Biol, 2003. **10**(3): p. 241-9.
224. Gauthier, A., et al., *Transcriptional inhibitor of virulence factors in enteropathogenic Escherichia coli*. Antimicrob Agents Chemother, 2005. **49**(10): p. 4101-9.
225. Dickey, S.W., G.Y.C. Cheung, and M. Otto, *Different drugs for bad bugs: antivirulence strategies in the age of antibiotic resistance*. Nat Rev Drug Discov, 2017. **16**(7): p. 457-471.

226. Smith, M.A., et al., *Identification of the binding site of Brucella VirB8 interaction inhibitors*. Chem Biol, 2012. **19**(8): p. 1041-8.
227. Paschos, A., et al., *An in vivo high-throughput screening approach targeting the type IV secretion system component VirB8 identified inhibitors of Brucella abortus 2308 proliferation*. Infect Immun, 2011. **79**(3): p. 1033-43.
228. Casu, B., et al., *Structural Analysis and Inhibition of TraE from the pKM101 Type IV Secretion System*. J Biol Chem, 2016. **291**(45): p. 23817-23829.
229. Shaffer, C.L., et al., *Peptidomimetic Small Molecules Disrupt Type IV Secretion System Activity in Diverse Bacterial Pathogens*. MBio, 2016. **7**(2): p. e00221-16.
230. Hilleringmann, M., et al., *Inhibitors of Helicobacter pylori ATPase CagAlpha block CagA transport and cag virulence*. Microbiology, 2006. **152**(Pt 10): p. 2919-30.
231. Sayer, J.R., et al., *2- and 3-substituted imidazo[1,2-a]pyrazines as inhibitors of bacterial type IV secretion*. Bioorg Med Chem, 2014. **22**(22): p. 6459-70.
232. Bjorkholm, B., et al., *Mutation frequency and biological cost of antibiotic resistance in Helicobacter pylori*. Proc Natl Acad Sci U S A, 2001. **98**(25): p. 14607-12.
233. Suerbaum, S., et al., *Free recombination within Helicobacter pylori*. Proc Natl Acad Sci U S A, 1998. **95**(21): p. 12619-24.
234. Falush, D., et al., *Traces of human migrations in Helicobacter pylori populations*. Science, 2003. **299**(5612): p. 1582-5.
235. Olbermann, P., et al., *A global overview of the genetic and functional diversity in the Helicobacter pylori cag pathogenicity island*. PLoS Genet, 2010. **6**(8): p. e1001069.
236. Yeh, Y.C., et al., *H. pylori cagL amino acid sequence polymorphism Y58E59 induces a corpus shift of gastric integrin alpha5beta1 related with gastric carcinogenesis*. Mol Carcinog, 2011. **50**(10): p. 751-9.
237. Ogawa, H., et al., *Genetic variants of Helicobacter pylori type IV secretion system components CagL and CagI and their association with clinical outcomes*. Gut Pathog, 2017. **9**: p. 21.
238. Hatakeyama, M. and H. Higashi, *Helicobacter pylori CagA: a new paradigm for bacterial carcinogenesis*. Cancer Sci, 2005. **96**(12): p. 835-43.
239. Hamdan, F.F., et al., *Monitoring protein-protein interactions in living cells by bioluminescence resonance energy transfer (BRET)*. Curr Protoc Neurosci, 2006. **Chapter 5**: p. Unit 5 23.
240. Wang, K.T., et al., *Crystal structures of catalytic intermediates of human selenophosphate synthetase I*. J Mol Biol, 2009. **390**(4): p. 747-59.
241. Breton, B., et al., *Multiplexing of multicolor bioluminescence resonance energy transfer*. Biophys J, 2010. **99**(12): p. 4037-46.
242. Brandt, S., et al., *NF-kappaB activation and potentiation of proinflammatory responses by the Helicobacter pylori CagA protein*. Proc Natl Acad Sci U S A, 2005. **102**(26): p. 9300-5.
243. Shariq, M., et al., *Biochemical Analysis of CagE: A VirB4 Homologue of Helicobacter pylori Cag-T4SS*. PLoS One, 2015. **10**(11): p. e0142606.
244. Ripoll-Rozada, J., et al., *Functional interactions of VirB11 traffic ATPases with VirB4 and VirD4 molecular motors in type IV secretion systems*. J Bacteriol, 2013. **195**(18): p. 4195-201.
245. Cascales, E. and P.J. Christie, *Agrobacterium VirB10, an ATP energy sensor required for type IV secretion*. Proc Natl Acad Sci U S A, 2004. **101**(49): p. 17228-33.

246. Tanaka, J., et al., *Structural definition on the surface of Helicobacter pylori type IV secretion apparatus*. Cell Microbiol, 2003. **5**(6): p. 395-404.
247. Linz, B., et al., *An African origin for the intimate association between humans and Helicobacter pylori*. Nature, 2007. **445**(7130): p. 915-918.
248. Saadat, I., et al., *Helicobacter pylori CagA targets PAR1/MARK kinase to disrupt epithelial cell polarity*. Nature, 2007. **447**(7142): p. 330-3.
249. Nishikawa, H., et al., *Impact of structural polymorphism for the Helicobacter pylori CagA oncoprotein on binding to polarity-regulating kinase PAR1b*. Sci Rep, 2016. **6**: p. 30031.
250. Koelblen, T., et al., *Molecular dissection of protein-protein interactions between integrin alpha5beta1 and the Helicobacter pylori Cag type IV secretion system*. FEBS J, 2017.

